geophysical monograph series

NUMBER 22

Basic Geophysics

Enders A. Robinson and Dean Clark

Foreword by Tijmen Jan Moser

SEG SOCIETY OF EXPLORATION GEOPHYSICISTS

GEOPHYSICAL MONOGRAPH SERIES

NUMBER 22

BASIC GEOPHYSICS

Enders A. Robinson

Dean Clark

with a foreword by
managing and volume editor
Tijmen Jan Moser

SOCIETY OF EXPLORATION
GEOPHYSICISTS

ISBN 978-0-931830-56-3 (Series)
ISBN 978-1-56080-345-4 (Volume)

Library of Congress Control Number: 2017952740

Society of Exploration Geophysicists
8801 S. Yale Ave., Ste. 500
Tulsa, OK 74137-3575

Published 2017
Printed in the United States of America

This book is dedicated to

Gregory W. Randolph, M.D.

The Claire and John Bertucci Endowed Chair at
Harvard Medical School, Massachusetts Eye and Ear Infirmary,
and Massachusetts General Hospital

and

Giuseppe Barbesino, M.D.

Massachusetts General Hospital

This page has been intentionally left blank

Contents

Preface

Come forth into the light of things
Let Nature be your teacher.
— William Wordsworth (1770 – 1850)

Chapters 1 through 4 in this book are composed of articles published by the authors in the journal of the Society of Exploration Geophysicists, *The Leading Edge* (*TLE*). Chapters 5 through 8 are taken from their unpublished manuscript, *Waves of Discovery*. James T. Robertson, member of the *TLE* Editorial Board, writes, "The intent is to cogently present the body of seismic theory that underlies modern exploration seismology in a format that transfers understanding to the audience. Some readers will be very familiar with many of the topics; others will be expert in only a few. I suspect that most of us ought to know more about these subjects than we really do. If each reader can find an insight or two that were not previously appreciated, this book would accomplish its purpose."

An important mode of thinking is visual thinking. Visualization is especially useful in solving problems where shapes, forms, or patterns are concerned. Visual thinking is used constantly by geophysicists, from the layout of a seismic survey to the final interpretation. Geophysicists use visual imagery in the analysis of seismic data and in the synthesis of the structure of the underground earth. They are called upon to exercise an acute sense of visual perception in the production of intricate sequences of images. This process ultimately holds the key to unlocking the secrets of the heterogeneities of the solid earth. Geophysicists also must make use of conceptualization in those instances when the desired results are not at all obvious.

Three kinds of visual imagery are necessary for effective visual thinking: (1) perceptual imagery, (2) mental imagery, and (3) graphic imagery. The first, *perceptual imagery*, is sensory experience of the physical world. It is what you see with your eyes and record in your brain. You do not record everything you observe. One reason is an over-saturation of input. People tend to see better those things which are more important, or more unusual, or more easily-recorded.

The second, *mental imagery,* is constructed in the mind and utilizes stored information recorded from perceptual imagery. Mental imagery is important when you examine the computer-generated images formed from actual seismic data. The clarity of mental images depends upon several factors. First, it depends on seeing the perceived geological objects. Second, it depends upon the image-reproduction mechanism in the brain. There are individual variations in the ability to imagine things visually, as there is always a range of answers. Try visualizing a series of seismic objects and see if you can determine a pattern in your own imaging ability. Visual images can be consciously enhanced. Are you better at visualizing two-dimensional objects than three-dimensional? Are you better at small structures than large ones? Where do you see the image, in front of your eyes or deep in your mind? Visual imaging ability depends not only on the ability to form images, but also upon the supply of pertinent imagery that is stored in the mind.

The third, *graphic imagery*, is the art of constructing physical images in order to convey specific information. To take advantage of visual thinking ability, this third type of visualization is necessary. Computer graphics allows the recording, storage, manipulation, and communication of images to fill the pictures generated in the imagination. Graphic imagery falls into two categories. One category includes pictures formed to communicate with others, and the other category pictures made for oneself.

Leonardo da Vinci was one of the greatest masters of the art of graphic imagery and his mythic stature is well deserved. Some 4000 pages of his drawings and writings survive, which included a vast array of subjects such as anatomy, inventions, landscapes, mythology, and cartography. An essential paradox of his drawings was that they portray nature more closely than had ever been done before, yet every part of them seems to be infused with all of the baffling eccentricities of Leonardo.

In contrast to Leonardo, Christiaan Huygens (1629 – 1695) produced unadorned artwork that precisely conveyed physical and mathematical knowledge. He invented the first accurate clock and the first astronomical telescope. His work included early telescopic studies of the rings of Saturn and the discovery of its moon Titan, the determination of the speed of light, and major studies of mechanics, optics, and probability. He was the most influential proponent of the wave theory of light.

After devoting a great deal of conscious thought to the enigmatic structure of the propagation of waves, it has been said that Huygens discovered his wavelet construction while watching the ripples on a canal in Holland. Wavelets were gamboling before his eyes. While the smaller groups continued to propagate in the background, his mental eye, rendered more acute by repeated visions of this kind, could distinguish the larger structure of their manifold action, in circular rows fitting together. The wavelets, all twining and twisting in their motion, formed the wavefront that whirled before his eyes. The result of this vision was Huygens' brilliant insight that a wavefront is formed by all of the wavelets emanating from the previous wavefront.

This principle, ever since called Huygens' principle, is fundamental and serves as a foundation of geophysical processing. Seismology uses elastic waves generated by natural earthquakes as well as controlled sources to probe the internal structure of the earth. How simple it is to look at wave motion on the surface of the water as compared to wave motion involved in a seismic survey. Geophysicists must use powerful computers to follow the seismic waves traveling through the three-dimensional earth. Digital signal processing originated in exploration geophysics, but now it is used everywhere to obtain a better estimate of the original form of a distorted signal. Signal restoration often can achieve amazingly good results. Throughout, the basic building block is the Huygens' wavelet.

The figures in this book will help you to see and understand seismic signals better. You can exercise your perceptual ability by looking at the figures and redrawing them. Such an activity requires imagining shape and destination in your mind's eye. Not only Huygens, but also Galileo, Descartes, Fermat, Newton, Hamilton, Green, Faraday, and Maxwell were masters of this art, and their mastery allows us to build upon their successes. We are indebted to the past, and upon it we chart our portion in the continuing course of the growth and development of this profession.

We want to express our sincere thanks to Susan Stamm, Books Manager for the Society of Exploration Geophysicists. In early discussions, she provided the inspiration for this book. Her diligence and hard work made this book a reality. We appreciate all that she has done from the beginning to the end.

We are most grateful for the honor of having the distinguished Dr. Tijmen Jan Moser as the editor of this book. His scientific insight and acumen were invaluable. He contributed the following foreword, in which he picks out classical drawings of Huygens that portray the essence of reflection seismology. The drawings in various guises permeate this book.

This page has been intentionally left blank

Foreword

Classical physics redux

Geophysics is the study of the physics of the earth, its surface, interior, and surrounding environment in space. An important part of geophysics is seismology. Seismic waves are vibrations that travel through the earth's interior or along its surface. Seismology is the study of how seismic waves can be used to determine the interior structure of the earth. Earthquake seismology is concerned with the structure of the entire earth, whereas exploration seismology deals with only the upper few kilometers of the earth's crust. In earthquake seismology, the source is passive, because the signals are earthquakes, which naturally occur within the remote medium (the subterranean earth). In exploration seismology, the source is active, because the signals are generated intentionally in the accessible medium (the surface or near-surface of the earth).

Exploration seismic data can be used to analyze potential petroleum reservoirs and mineral deposits, locate groundwater, find archaeological relics, determine the thickness of glaciers and soils, and assess sites for environmental remediation. Seismic exploration of the upper crust has yielded an unparalleled source of knowledge of the geologic history of the earth. Like the 1895 novel, *The Time Machine,* by H. G. Wells, which has a machine that allows its operator to travel back and forth through time, the upper crust is a kind of time machine. The surface corresponds to present time, and going down in depth is equivalent to traveling back in time. The subsurface layers provide a remarkable history of the earth and its plant and animal life.

At the forefront of every scientific discipline, there are those driven by a passion to know what is most essential about their subject. Any penetrating study requires a great deal of patience and determination. It is a difficult course, but often the most rewarding. Scientists who know the foundations of their subject are better equipped to tackle new problems and achieve worthwhile results. Mathematicians need to have command of the mathematics involved. Geologists must be conversant with the geologic principles at play. Physicists must draw upon their knowledge of encompassing physical principles. Computer scientists must be familiar with the intricacies of computer hardware and software. Geophysicists make use of all of the above sciences, as well as the many geophysical principles involved. Their foremost tool is the traveling seismic wave. Seismic wave propagation is a source of three-dimensional information, as opposed to potential methods, which provide two-dimensional information, and well logging, which is essentially one-dimensional. The dimensionality of seismic wave propagation can be expanded to five by adding offset and azimuth. Multicomponent seismic data even further enrich our information of the elastic properties of the earth's interior. Even so, the inversion of seismic data into interpretable information remains a monumental challenge.

Waves of one form or another can be found in amazingly different situations. A wave is a traveling disturbance. Ocean waves travel for thousands of kilometers through the water. Seismic waves travel through the earth. Sound waves travel through the air to our ears, where we process the disturbances and interpret them. Much of the current understanding of wave motion has come from the study of acoustics. Ancient Greek philosophers, many of whom were interested in music, hypothesize a connection between waves and sound, and that vibrations, or disturbances, must be responsible for sounds.

Pythagoras (c 580 BC – c 500 BC) has been credited as the founder of the science of acoustics which, in his time, was the study of the physical aspects of music, and he believed that sound was composed of vibrations. Scientific theories of wave propagation became more prominent in the 17th century with Galileo Galilei (1564 – 1642) publishing a clear statement of the connection between vibrating bodies and the sounds they produce. Robert Boyle (1627 – 1691), in a classic experiment in 1660, proved that sound cannot travel through a vacuum, thereby showing that it needs a medium in which to propagate. Christiaan Huygens (1629 – 1695) wrote *Nouveau Cycle Harmonique*, a work that demonstrates his extensive knowledge of the theory of music; the result of his investigations was the discovery of a better tuning system in which the octave is divided into 31 steps.

Under the tremendous influence of Fermat and Newton, the prevailing theory in the 17th century was that light propagates as a stream of particles. In *Traité de la Lumière* (1690), Huygens made one of the great contributions to theoretical physics when he postulated that light was a wave, supporting this postulate with Huygens' principle. Huygens' wave theory of light, however, was in direct opposition to the prevailing theory, so it generally was disregarded at the time. Yet, Huygens' principle for the propagation of waves proved to be useful in explaining many phenomena which geophysicists encounter in their daily practice: reflection, refraction, diffraction, anisotropy, caustics, and multipathing.

Huygens solved a number of problems in optics and mechanics. More than a decade before Newton published the three laws of motion, Huygens solved the problem of centrifugal force. In so doing, Huygens twice differentiated a vector-valued function and made use of the yet-to-be-published Newton's second law. Another important achievement of Huygens was his investigation of evolvents, a concept that he introduced. Evolvents are connected with another investigation of Huygens, namely the theory of wavefronts. In using his Huygens' principle, Huygens discovered that, in certain cases, singularities could appear on the wavefront. Singular points were important in his investigation of the correspondence between waves and rays. In this regard, Huygens made use of methods that are now attributed to the calculus of variations and Hamiltonian mechanics.

In *Traité de la Lumière* (1690), Huygens presents a drawing that illustrates his principle for the propagation of waves—it appears in this book as Figure 12 of Chapter 3. This seminal book contains four additional classical drawings, shown in Figure 1, which in one form or another appear in practically every book on geophysics and optics. In Figure 1, the top-left

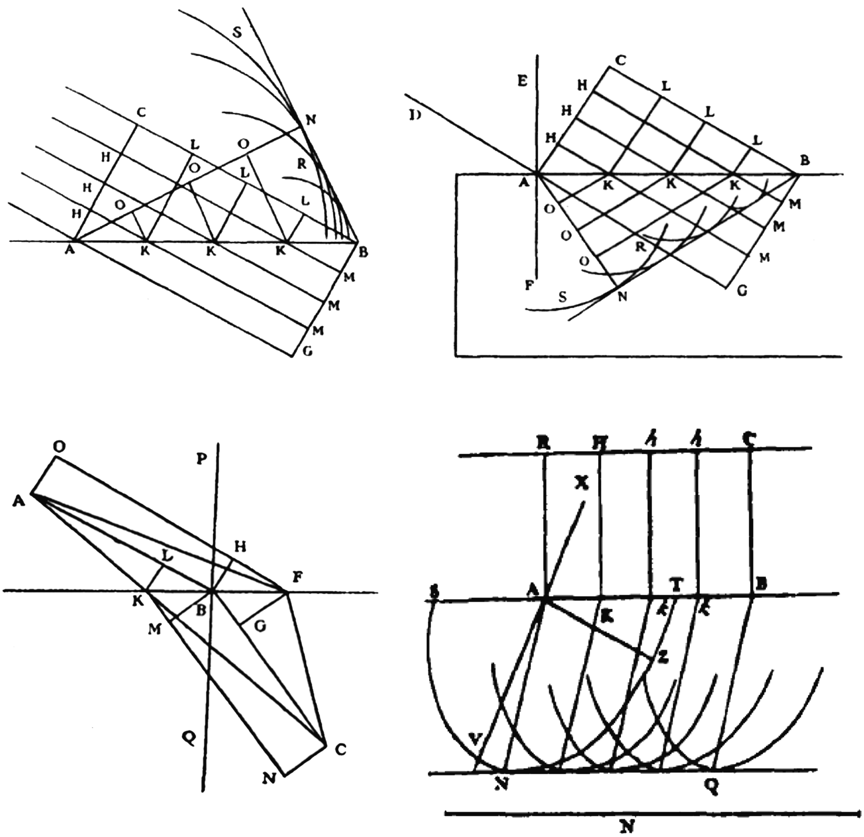


Figure 1. Four classical drawings of Huygens from *Traité de la Lumière* (1690).

diagram illustrates how the process of reflection is explained by using Huygens' principle—it appears in simplified form in this book as Figure 32 of Chapter 1. The top-right diagram illustrates how the process of refraction is explained by using Huygens' principle—it appears in simplified form in this book as Figure 33 of Chapter 1. The bottom-left diagram illustrates Fermat's principle—it appears in simplified form as Figure 23 of Chapter 1. The bottom-right diagram illustrates how anisotropy on Icelandic crystal is explained by using Huygens' principle—it appears in simplified form as Figure 2 of Chapter 4. A sixth figure, appearing here as Figure 2, must be mentioned as well. This classical figure illustrates how the formation of a caustic and triplicated wave at a circular boundary is explained by using Huygens' principle. The envelope of the elementary waves at the incident wavefront A-F-E generate the caustic line N-b-e-h-C and the triplicated wavefronts a-b-c, d-e-f, and g-h-k. Indeed, Huygens' principle is central in most seismic imaging/migration

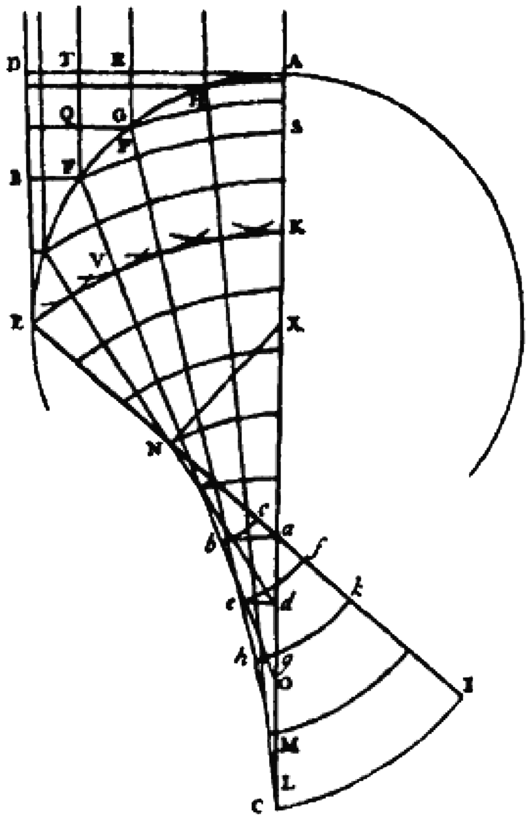


Figure 2. Another classical drawings of Huygens from *Traité de la Lumière* (1690).

techniques, making it not an exaggeration to call Huygens the first great geophysicist.

In the 18th century, Jean Le Rond d'Alembert (1717 – 1783) derived the wave equation, one of the first main partial differential equations of physics. The wave equation provided a thorough and general mathematical description of waves and laid the foundations for the study and description of wave phenomena. A good understanding of the wave equation has been a great advantage in geophysics. For instance, the superposition principle stated that individual waves can travel independently and add up when meeting each other, which has been instrumental in understanding the propagation of seismic waves and their use in seismic imaging.

In the 17th century, the foundations of classical physics were established and it gradually became a mature scientific discipline that described the natural world with remarkable accuracy. In the 1860s, James Clerk Maxwell (1831 – 1879) extended the framework of classical physics to take into account electrical and magnetic forces. By the end of the 19th century, there was a growing sense that theoretical physics would soon be complete. It was suggested the underlying principles were firmly established and all that remained were details of determining some numbers to more decimal places. Of course, shortcomings were evident, but there was a general feeling that these were mere details that could be resolved easily.

Within the first decade of the 20th century, however, everything changed. As anticipated, the shortcomings were addressed promptly, but they proved anything but minor. They ignited a revolution, and required a fundamental rewriting of nature's laws. Classical conceptions, which for hundreds of years expressed the intuitive sense of the world, were overthrown. Quantum mechanics and general relativity were developed. The result, modern physics, became absolutely necessary for the understanding of the micro-world and the macro-world.

It is true that geophysics is concerned primarily with classical concepts. Apart from some well logging tools and some concepts in wave propagation, such as scattering, and inversion, such as uncertainty, quantum mechanics does not play a major role in geophysics. In the same way, relativity does not play an important role, perhaps apart from some sophisticated concepts in inversion. Still, the fact that theoretical physicists were taken so much by surprise by the revolution caused by quantum mechanics and relativity ought to be a lesson for geophysicists never to think that there is a single black-box solution to all geophysical problems.

Throughout this time, classical physics remained valid, but the belief persisted that there was little left for classical physicists to discover and they became occupied with filling in small details. As the age of technology became ascendant in the last half of the 20th century, however, the emergence of the electronic digital computer allowed scientists to compute numerical solutions for problems of vast magnitude, something that was never dreamed possible. Classical physics was once more in demand.

The impossible takes a little longer

In 18th century France, Charles-Alexandre de Calonne was the Minister of Finance to King Louis XVI. In a commanding tone, Queen Marie Antoinette demanded that the Minister do something that was extremely difficult. His reply was: "Madame, si c'est possible, c'est fait. Impossible? Cela se fera." (Madame, if it is possible, it is done. Impossible? It will be done).

In the 1950s, good images of the subterranean earth were very difficult to make. Despite unassailable difficulties, geophysics was the first science to enlist computers for digital signal processing. Geophysicists originated extensive methods of seismic acquisition and devised new mathematical algorithms that allowed the construction of better and better subsurface geologic images. A particular challenge in geophysics was that data typically are incomplete and inconsistent at the same time. Subsurface complexities showed a fractal character and existed at all scales. Adequate parametrization of a model of the earth and forward modeling, the prediction of the seismic wave response of a given earth model, were important problems to which no universal solution might exist. Inversion of seismic data recorded at the earth's surface was known to be ill-posed in the sense of Hadamard (1865 – 1963). Therefore, geophysicists did not attack the whole problem but instead judiciously separated the overall problem into component parts that could be handled by the available computers. The main thrust of their work was with two-dimensional analyses of the primary reflections of compressional body waves. Three- and four-component acquisition and the effects of shear waves, surface waves, refracted waves, diffracted waves, and anisotropy were left to special studies. By the last decade of the 20th century, computers and geophysical instrumentation had advanced to such proportions that three-dimensional images of the subsurface became commonplace. Visualization made a similar development starting from relatively simple two-dimensional seismic sections, to full graphical display of three-dimensional volumes, and even higher dimensions for pre-stack seismic volumes. Geophysicists, like the renowned minister Charles-Alexandre de Calonne, could respond, "If it is possible, it is done."

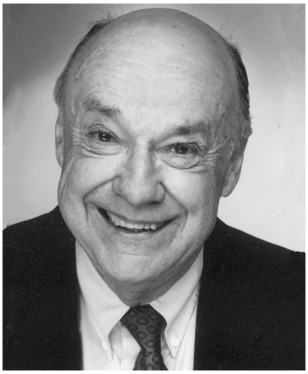
Seismic images have continued to improve. Even so, extremely detailed images of the subterranean earth are impossible to make, and the ultimate dream of unlimited image resolution is still beyond the horizon. To achieve this objective, the geophysical industry has assembled elaborate systems of active remote sensing. The great developments in computers and instrumentation have made possible the recording and analysis of great amounts of accurate scientific observations. For acquisition, seismic signals are dispersed from thousands of sending points. Returns from each of these signals are received at many receiving points, which can number in the hundreds of thousands. For processing, the most powerful super computers yet constructed are employed. Intricate computer programs are based upon advanced physics and mathematics. Developments such as quantum computing are still ahead of us, but may dramatically change the way geophysicists analyze data and extract information on the earth's interior from them.

Today is a most exciting time for geophysicists. Never before have such a massive array of sophisticated equipment and powerful computers been available to any scientific endeavor. What is most important is a good knowledge of geophysical principles. A geophysicist today has the opportunity to do what before would be considered impossible. The response of geophysicists is that of Charles-Alexandre: "Impossible? It will be done."

Tijmen Jan Moser

This page has been intentionally left blank

About the Authors



Enders A. Robinson is professor emeritus of geophysics at Columbia University in the Maurice Ewing and J. Lamar Worzel Chair. He received a B.S. in mathematics in 1950, an M.S. in economics in 1952, and a Ph.D. in geophysics in 1954, all from Massachusetts Institute of Technology. As a research assistant in the mathematics department at MIT in 1950, Robinson was assigned to seismic research. Paper-and-pencil mathematics on the analytic solution of differential equations was expected. Instead, Robinson digitized the seismic records and processed them on the MIT Whirlwind digital computer. The success of digital signal processing led to the formation of the MIT Geophysical Analysis Group in 1952 with Robinson as director. Almost the entire geophysical exploration industry participated in this digital enterprise. In 1965, Robinson and six colleagues formed Digicon, one of the first companies to do commercial digital seismic processing. In 1996, Digicon and Veritas combined to form VeritasDGC, which combined with CGG in 2007.

With Sven Treitel, Robinson received the SEG award for best paper in GEOPHYSICS in 1964, the SEG Reginald Fessenden Award in 1969, and the Conrad Schlumberger Award from the European Association of Exploration Geophysicists, also in 1969. In 1983, Robinson was made an honorary member of SEG. In 1984, he received the Donald G. Fink Prize Award from the Institute of Electrical and Electronic Engineers. In 1988, he was elected to membership in the National Academy of Engineering. He received the SEG Maurice Ewing Medal and the SEG award for best paper in GEOPHYSICS in 2001, the Blaise Pascal Medal for Science and Technology from the European Academy of Sciences in 2003, and the Desiderius Erasmus Award from the European Association of Geoscientists and Engineers in 2010. Robinson is the author of 20 books and the coauthor of 13.

Dean Clark joined the publications department at SEG in 1981 as associate editor of *The Leading Edge* (*TLE*). He became *TLE*'s editor in 1984 and served in that capacity until retiring in 2013. He served as a U.S. Army officer from 1966 to 1969. He then joined the sports department of the *Tulsa World* newspaper where he spent 12 years as a reporter and columnist. Clark has written one hundred scientific and literary articles covering all phases of exploration geophysics including the history of geophysics, biographies of leading geophysicists, expositions of current developments and new trends, and mathematical tutorials. He also has written short stories and short plays based on characters in the Sherlock Holmes stories of Sir Arthur Conan Doyle. Six of his plays have been performed in New York during the annual January gathering of the Baker Street Irregulars to celebrate Holmes' birthday.

Clark is a graduate of Phillips Exeter Academy and the University of Oklahoma, where he earned a bachelor of science degree in journalism in 1966. He is a founding member of the Afghanistan Perceivers of Oklahoma and a member of Circulo Holmes, the Sherlock Holmes club in Barcelona, Spain.

Chapter 1

Classical Geophysics

Pythagoras and Archimedes

Recently an acquaintance, who shall be called Jones to protect the guilty, was mathematically immortal for a few minutes. He discovered a corollary to the Pythagorean theorem and was immediately overwhelmed with thoughts about being forever paired with Pythagoras in textbooks. These were quickly followed by the idea, which he found delicious, of being eternally cursed by all 5th grade students the night before final arithmetic exams.

The Pythagorean theorem (Figure 1) states that, in a right-angled triangle, the square on the hypotenuse (the long side) is equal to the sum of squares on the two legs (the other two sides). Jones's involvement with Pythagoras began when he was asked, by the son of a neighbor, to help create an 8th grade science project. Jones suggested they attempt to devise a visual proof of the Pythagorean theorem because it is quite possibly the most important relationship in mathematics.

Jones began with right triangle *abc* and constructed a square on *c*, the hypotenuse. Then he added three more triangles around the square to form the construction displayed in Figure 2. It can be shown mathematically that the four triangles are identical; thus, the large box is a square because all four sides are length $a + b$ and all of the angles are right angles. The area of the square, of course, can be expressed as

$$(a + b)^2 \ .$$

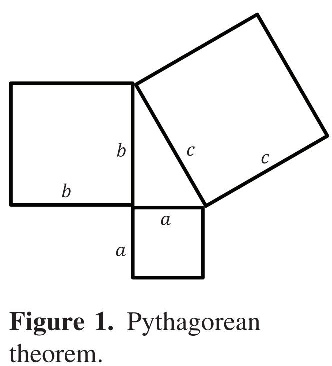


Figure 1. Pythagorean theorem.

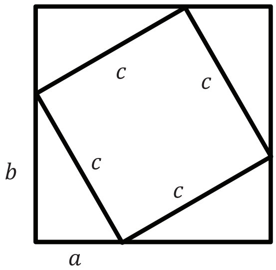


Figure 2. Pythagoras-type proof.

However, by inspection, we see the area of the large square also equals the small square plus the four triangles, or

$$c^2 + 4\left(\frac{ab}{2}\right) .$$

Because the areas are equal, we can equate them mathematically,

$$(a + b)^2 = c^2 + 4\left(\frac{ab}{2}\right) .$$

After both sides are expanded, the equation reads

$$a^2 + 2ab + b^2 = c^2 + 2ab .$$

The ab terms can be subtracted from both sides, leaving the familiar

$$a^2 + b^2 = c^2 ,$$

which is what Jones wanted to prove.

Even so, Jones could not leave well enough alone. That night, just before going to bed, he began to fiddle with the triangles. This time, however, he put them inside the square of the hypotenuse, resulting in the construction displayed in Figure 3. It immediately can be seen that the area of the square of the hypotenuse equals the sum of the areas of the four identical triangles plus the area of the small square, with side $b - a$, in the center. Thus,

$$c^2 = 4\left(\frac{ab}{2}\right) + (b - a)^2 = 2ab + (b - a)^2 .$$

After writing down the formula, Jones was suddenly struck by what the terms meant: c^2 (i.e., the square of the hypotenuse) equaled $2ab$ (i.e., twice the product of the two sides) plus $(b - a)^2$ (i.e., the square of their difference). IMMORTALITY! Pythagoras, after 2500 years, had been extended. Just to be sure, Jones pulled several ungainly numbers out of the air and plugged them into his formula. All worked. There was no doubt that Pythagoras and Jones would be forever linked *à la* Damon and Pythias, Castor and Pollux, etc. Then, alas, disillusionment. When Jones expanded the right side of his equation, the ab terms cancelled out and all that was left was

$$c^2 = a^2 + b^2 .$$

Jones really shouldn't feel dejected. He has not, to be sure, made a distinguished contribution to mathematics, but he has joined a long, and rather formidable, group of people who have discovered and rediscovered the Pythagorean theorem over the centuries.

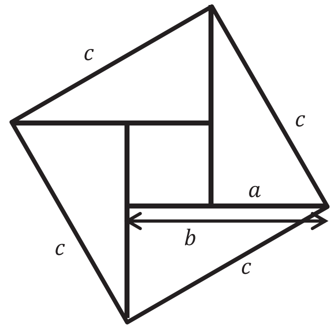


Figure 3. Jones' proof.

Although little is known about Pythagoras himself, many stories have come down to us, most of which are, of course, apocryphal. Pythagoras (c 580 BC – c 500 BC) is a Greek philosopher who, after extensive traveling, settled in the Greek seaport of Crotona, located in southern Italy, where he founded the famous Pythagorean School, an academy for the study of philosophy, mathematics, and natural science.

Tradition ascribes to Pythagoras the discovery of the theorem on the right triangle that now universally bears his name. However, the first Jones (our name for any person who independently discovered the theorem or provided a new proof) is a Babylonian of Hammurabi's time who lived a thousand years earlier than Pythagoras. The old Babylonian cuneiform text BM85196 contains the following problem.

A stick of length 30 stands against a wall. The upper end slips down a distance 6. How far has the lower end moved? The solution of the problem can be found by means of the Pythagorean theorem as follows. The hypotenuse of the right triangle is 30 and the vertical leg is $30 - 6 = 24$. The square of 30 is 900. The square of 24 is 576. We perform the subtraction $900 - 576 = 324$. The answer is $\sqrt{324} = 18$.

There is much conjecture as to the proof that Pythagoras himself gave, but it is generally believed that it was a dissection-type proof—that is, a geometric version of the proof which our own modern Jones gave. On the left of Figure 4, we reproduce Figure 2. We see that the large square (i.e., the square with side $a + b$) is dissected into five pieces, namely the square on the hypotenuse and four right triangles congruent to the given triangle. Let us now dissect the square with side $a + b$ differently, as shown on the right of Figure 4. We see that the square is now divided into six pieces, namely the two squares on the legs and four right triangles congruent to the given triangle. By subtracting the four congruent triangles from both sides of Figure 4, we see that the square of the hypotenuse (on the left) is equal to the sum of the squares of the legs (on the right).

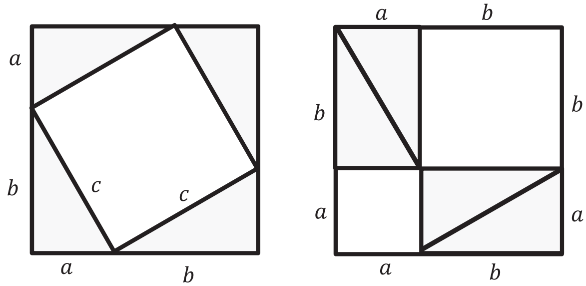


Figure 4. Dissection-type proof.

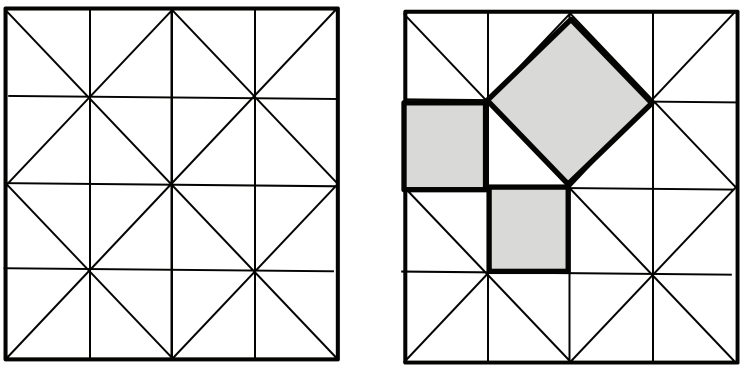


Figure 5. Diagram of Socrates.

Another Jones is found in China during the Han period (202 BC – AD 220). The illustration shown on the left in Figure 4 appears in the *Chou Pei*, a Chinese manuscript that dates back to this period but is believed to contain much older mathematical material. Although the manuscript contains no actual proof of the theorem, it does mention the right triangle with integral sides, 3, 4, and 5.

The discovery also is claimed for India because of the *Apastamba-Sulba-Sutra*, the date of which is put at least as early as the 5th or 4th century BC but whose material is considered much older than the book itself.

The most famous proof of all is that given by Euclid (Book 1, Proposition 47). As many will recall from high school geometry, Euclid's proof is somewhat complicated, making use of various construction lines and a long chain of deductive steps. We will not burden the reader with the proof, but instead give the diagram (Figure 5) used by Socrates (in Plato's *Meno*) to convince a slave boy of the truth of the theorem. This proof concerns a special case, and does not prove the theorem, but it does give a maximum of intuitive insight.

Another Jones is the Hindu mathematician and astronomer Bhaskara who flourished around AD 1150. Bhaskara's dissection proof of the Pythagorean theorem is the geometric version of the second proof of our own Jones. On the left in Figure 6, we reproduce Figure 3, which shows the square of the hypotenuse cut into four congruent right triangles plus a square with its side equal to the difference of the legs on the given triangle. We rearrange these pieces on the right in Figure 6. Then we carefully draw the dotted vertical line. On the left of the dotted vertical line lies the square of the little leg. On the right lies the square of the big leg. Thus, we see that the sum of the squares on the two legs is equal to the square of the hypotenuse. Bhaskara draws the figure but offers no explanation except the word, "Behold." (Our own Jones, earlier in this section, supplies the algebraic proof.)

These examples are ample evidence that the (probably misnamed) Pythagorean theorem has intrigued professional and amateur mathematicians throughout recorded history. Many of the proofs cited are ingenious, but our admiration for the cleverness of their creators must be tempered by the realization that they, like our acquaintance Jones, know the answer before starting. How different are the circumstances facing the original Jones who, by definition, did not. Although the theorem ultimately boils down to a very simple equation, the invaluable relationship between the three sides of a right triangle is a complex concept—not an intuitively obvious mathematical idea such as the equation for the area of a rectangle.

The original Jones had to discover and then prove this complicated idea without algebra and using an awkward incomplete number system. It was, by any measure, a brilliant accomplishment. Because of the primitive state of mathematics at the time, it is probable the theorem's discoverer was not the traditional desk-bound mathematician but a practical man-of-affairs who gradually recognized the pattern during his daily routine. This almost certainly involved intimate association with the earth, the measure of which (as the name implies) led to the evolution of geometry. It is not

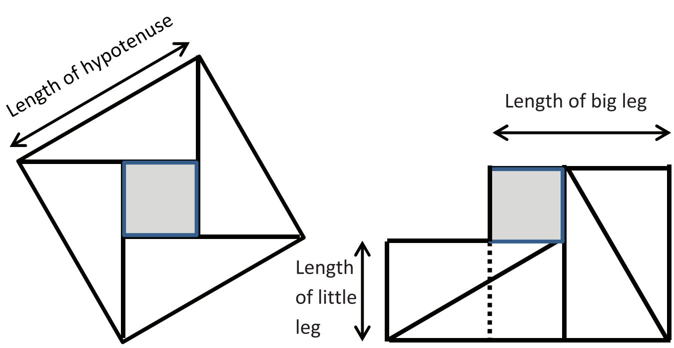


Figure 6. Dissection proof of Bhaskara.

farfetched to speculate that the theorem's discoverer was regularly employed in tasks which required accurate measurements of the earth.

We concede that this, like much involving Pythagoras and the theorem, is conjecture; but it seems highly probable and thus the original Jones, whatever else he was, is also among the first and greatest geophysicists.

We end with a challenge to all potential Joneses. In Figure 7, there are: (left) a **given** large square showing a bold small square within, (middle) the **given** bold small square by itself, and (right) the large square cut into four pieces. The problem is to rearrange the given bold small square and the four pieces into one square. If the two given squares are placed along the legs of a right triangle, the constructed square can be placed on the hypotenuse. The problem illustrates the Pythagorean theorem.

Geophysics in the early 1960s became the first science to experience the digital revolution which today is universal, from high-definition television to the latest developments in medicine. But the digital revolution did not result from new knowledge; it resulted from new technology, namely the digital computer. If the ancient Greeks had possessed this technology, mathematicians might never have developed one of their most beautiful creations—calculus.

Eudoxus of Cnidos, who lived from 408 BC to 355 BC, studied with Archytas of Tarentum, a follower of Pythagoras. Eudoxus visited Athens where he attended lectures by Plato, spent over a year in Egypt where he studied astronomy with the priests at Heliopolis, and then traveled to Cyzicus in the Propontus where he established a school which proved very popular and where he is believed to have introduced the *method of exhaustion*.

The most famous application of the method of exhaustion was to approximate the area or volume of a figure by adding the areas or volumes of a sequence of inscribed figures within the given figure. The successive

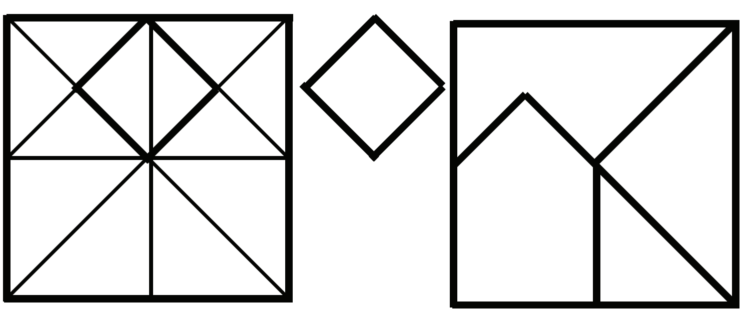


Figure 7. A challenge.

inscribed figures, in the limit, exhaust the given figure. Around 225 BC, Archimedes of Syracuse used the method of exhaustion to determine areas and volumes of geometric figures. In other words, Archimedes almost invented integral calculus about 1900 years before Newton and Leibniz.

Archimedes uses the method of exhaustion to determine one of the most important formulas in mathematics, the area of a circle. In Figure 8, a regular polygon (i.e., a polygon with n equal sides) is inscribed in a circle. The polygon is composed of n triangles, one triangle for each side of the polygon and the area of the polygon is n times the area of one of the triangles:

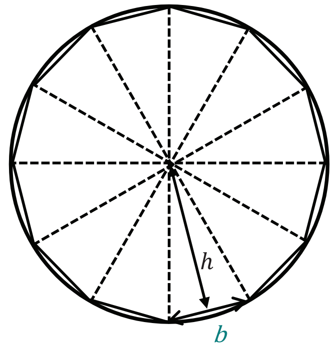


Figure 8. Polygon with twelve equal sides inscribed in a circle.

$$A = n(\text{area of triangle}) = n\frac{bh}{2} \ , \tag{1}$$

where b and h are, respectively, the base and height of one of the triangles. This equation yields a discrete approximation for the area of a circle and can approximate the true value to any degree of accuracy if n is made large enough. It is, in effect, a type of discrete algorithm that is required for programming a mathematical operation on a digital computer. If digital computers had been available in the time of Archimedes, it is conceivable that his efforts would have stopped here. However, he continues to analyze equation 1 and rewrites it as

$$A = \frac{h(nb)}{2} \ . \tag{2}$$

The term nb is the perimeter of the polygon. As n increases and b simultaneously decreases (an interplay that is at the heart of integral calculus), the area of the polygon approaches the area of the circle. The perimeter becomes closer and closer to the circumference c of the circle, and the height h of the triangle approaches the radius r of the circle. Thus, Archimedes makes the intuitive leap that the area of the polygon would, as it approximated a circle to an increasing degree, approach the value:

$$A = \frac{rc}{2} \ . \tag{3}$$

This expression obtains the area of the circle as one-half of the product of the radius and circumference. The constant π is the ratio of the circumference c to the diameter $2r$ of the circle and substituting this definition into equation 3 yields the familiar closed expression for the area of the circle

$$A = \pi r^2 \ . \tag{4}$$

In summary, the approximation given by equation 1 approaches the true value given by equation 4 in the limit, as n tends to infinity and b tends to 0. Because equation 4 provides the exact value, it would appear more useful. However, in order to use equation 4, it is necessary to know the numerical value of π and, until Archimedes, it was not known how to compute the value of this fundamental constant. Empirical estimates of the time, using methods such as wrapping a cord around a cylindrical object or rolling a circular object along a ruler, range from 3 to 3.2. Archimedes solves the problem with another clever application of the theory of exhaustion. He observes that the perimeter of a regular polygon inscribed in a circle is smaller than the circumference of the circle and, similarly, the perimeter of a regular polygon circumscribed about a circle is larger than the circumference of the circle. Consequently, Archimedes realizes that he could calculate lower and upper limits on the circumference of the circle, from which he could estimate the value of π, using equation 1 to determine the areas of these two polygons. The greater the number of sides of the polygons, the closer the limits would be. Archimedes computes this for polygons of 96 sides and finds that

$$3\frac{10}{71} < \pi < 3\ \frac{1}{7} \ . \tag{5}$$

In addition to integral calculus, Archimedes also anticipates differential calculus. His method of finding the tangent to his spiral (the so-called Archimedes spiral) is an important early "application" of differential calculus. However, the Greeks fail to recognize the connection between area problems and tangency problems—now known (as its name implies) as the monumentally important fundamental theorem of calculus. This connection is not discovered until the 17th century.

Can the Archimedes approach be applied to modern geophysics? Indeed, it can determine the path of a seismic ray. In the olden days, a common method of hand migration uses a wavefront chart, which is a graph showing wavefronts and raypaths for an assumed vertical distribution of velocity $V(z)$. The raypaths are found by applying Snell's law. Each raypath is characterized by Snell's parameter

$$p = \frac{\sin i_0}{V_0} \ .$$

In this equation, i_0 is the angle of incidence and V_0 is the velocity at the horizontal datum plane $z_0 = 0$. We can construct the raypath chart by:

1. dividing the medium into a large number of thin beds, each with constant velocity
2. for bed n, letting the thickness (the vertical displacement of the ray) be Δz_n
3. letting the horizontal displacement of the ray be Δx_n
4. letting the traveltime be Δt_n
5. letting i_n denote the angle of incidence.

Based on Figure 9, we quickly come up with the discrete equations that would be used in a digital computer to obtain the raypaths and the wavefronts:

$$\frac{\sin i_n}{V_n} = \frac{\sin i_0}{V_0} = p \ ,$$

$$V_n = V(z_n)\Delta x_n = \Delta z_n \tan i_n \ , \tag{6}$$

$$\Delta t_n = \frac{\Delta z_n}{V_n} \cos i_n \ .$$

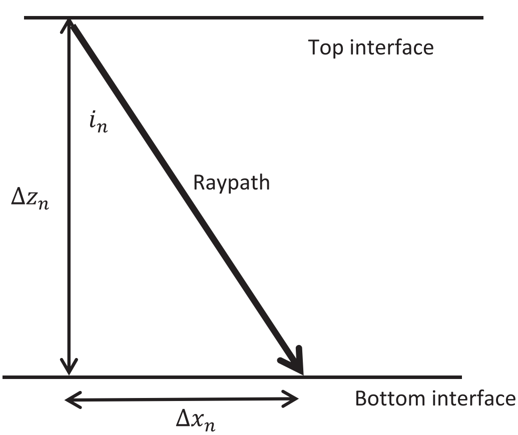


Figure 9. Raypath in thin bed.

However, in order to obtain a continuous solution, we divide the last two equations by Δz_n and take the limit as n approaches infinity:

$$\begin{aligned} \frac{\sin i}{V} &= \frac{\sin i_0}{V_0} = p \ , \\ V &= V(z) \ , \\ \frac{dx}{dz} &= \tan i \ , \\ \frac{dt}{dz} &= \frac{1}{V \cos i} \ . \end{aligned} \tag{7}$$

Integration then produces the values for x and t:

$$\begin{aligned} x &= \int_0^z \tan i \, dz = \int_0^z \frac{pV}{\sqrt{1-(pV)^2}} \, dz \ , \\ t &= \int_0^z \frac{1}{V \cos i} \, dz = \int_0^z \frac{1}{V\sqrt{1-(pV)^2}} \, dz \ . \end{aligned} \tag{8}$$

Equations 8 will generate an analytic solution for the raypaths and wavefronts. In isotropic media, raypaths are always perpendicular to the corresponding wavefronts. In the case of constant velocity, the wavefronts are concentric circles and the raypaths are straight lines radiating from the center.

But the applied seismologist rarely has a case with a constant velocity. In many situations, the velocity function is the linear expression

$$V = V_0 + az \ ,$$

where V_0 and a are constants. In this case, the wavefronts are circles, but they are no longer concentric. The raypaths are circles orthogonal to the wavefront circles. Figure 10 is a wavefront chart showing the circular wavefronts and the associated circular raypaths. The chart acts as an analog computer. For each pair of values t and $\sin i_0$ measured on the reflection seismogram, one pair of values of x and z (which gives the underground position of the reflector) can be read from the chart.

When the velocity functions are more complex than a linear expression, it is very difficult or impossible to obtain a closed-form solution by integration. As a result, numerical analysis has to be used to approximate the integrals. There was no easy way to compute the wavefront charts

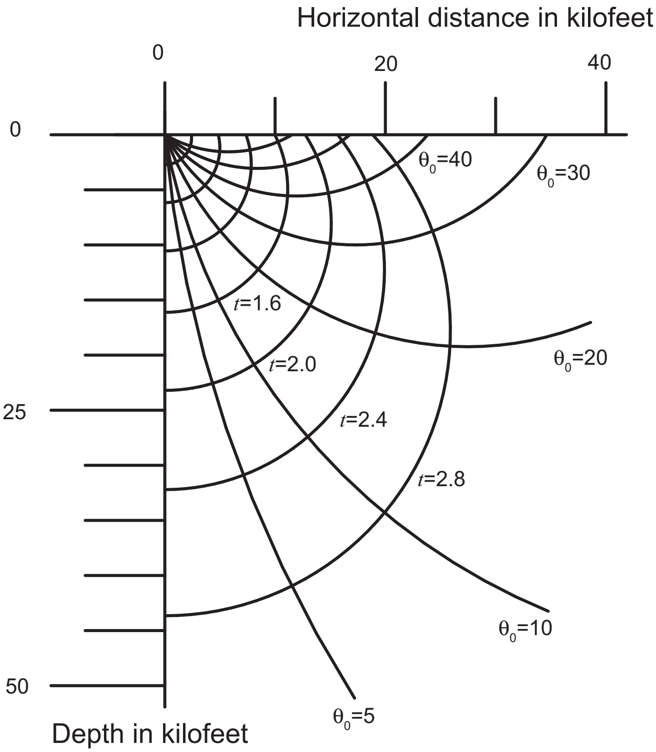


Figure 10. Wavefront chart for a linear increase of velocity with depth.

for more complex velocity functions before the advent of the digital computer. With the lightning speed of the digital computer in the 21st century, however, such approximations are accomplished very quickly and incredibly accurately. The analog computers represented by the wavefront charts are no longer needed for seismic migration.

In a sense, mathematics has come full circle. Archimedes is the pioneer who took computing from discrete approximations to exact analytic expressions. His work is the beginning of the journey that has led to the invention of calculus and, over the next few centuries, mathematicians spent considerable effort looking for "closed form" solutions to increasingly complicated integrals.

Even so, for most of the interesting problems in contemporary geophysics, the differential equations and/or integral equations, although beautiful to see, are too difficult or unwieldy to solve analytically. The digital revolution takes us back (albeit inside the memory of the digital computer) to numerical algorithms that use the type of discrete approximations as originated by Archimedes. The more things change, the more they stay the same.

Heron and Snell

Heron of Alexandria (c AD 10 – c AD 70), a Greek mathematician who taught at the Museum in Alexandria, wrote on the measurement of geometric figures and apparently had a flair for invention that was centuries before his time.

Heron is credited with the first documented steam engine, the *aeolipile* or "wind ball" (Figure 11), which consists of a hollow sphere mounted on a boiler by two pipes. The mounting allows the sphere to rotate and, as the water boils, steam rises through the pipes into the sphere, ultimately escaping from the sphere through two bent outlet tubes (canted nozzles). The escaping steam produces a rotary motion—a principle similar to that used in jet propulsion. The steam engines of the 18th century are based, in part, on this design.

Heron also developed instruments for measuring distances and roads and is credited with the formula that gives the area of a triangle from its sides. If a, b, and c are the lengths of the sides of a triangle, then the semiperimeter is

$$s = (a + b + c)/2 \ .$$

Heron's formula states the area of the triangle is

$$\sqrt{s(s-a)(s-b)(s-c)} \ .$$

This formula is commonly used in land surveying.

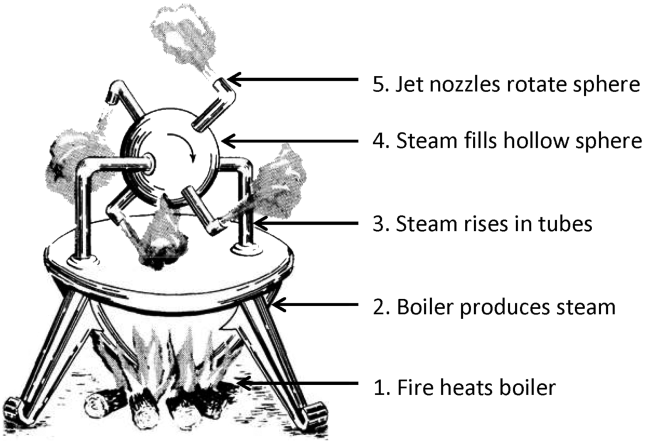


Figure 11. Heron's steam engine, the *aeolipile*.

Heron's inventions are notable for another modern concept, self-regulating feedback control systems. For example, one invention is a self-filling wine bowl which has a hidden float valve that automatically senses the level of wine in a bowl. Heron perhaps anticipates rotary drilling in his description of a machine, called the Cheirobalistra, which has a refined screw-cutting technique that would enable it to bore a tunnel through a mountain. Heron devised the odometer, which is an instrument for measuring the distance traveled by a wheeled vehicle. The pace is the length of two single steps (i.e., left to right plus right to left). A mile consists of 1000 paces. In English units the pace is 5.28 feet, so the English mile is 1000 paces or 5280 feet. Suppose the circumference of a chariot wheel is 2.5 paces. The wheel turns 400 times in one mile. For each revolution of the wheel, a pin engages a 400-toothed cogwheel. The cogwheel would make one complete revolution for every mile. This cogwheel engages another gear that has holes along its circumference, where pebbles are located that drop one by one into a box. To know the number of miles driven, the charioteer would count the number of pebbles in the box.

Heron's most important contribution (at least to anyone using reflection seismology) is that he established the law of reflection—that the angle of incidence is equal to the angle of reflection.

According to Aristotle, there were three kinds of motion: rectilinear, circular, and mixed. The four elements of the sublunar world tend to move in straight lines: earth downward, fire upward, water and air falling in between. Aether, on the other hand, naturally moved in circles. Then it was Gallileo's work in the 17th century that began our understanding of modern science. As Newton wrote in *Principia Mathematica* (1687), "By the first two Laws [of motion], Galileo discovered . . . that the motion of projectiles was in the curve of a parabola." Until Galileo, people generally accepted Aristotle's dictum that most motion appeared to be in either straight lines or circles.

Following this reasoning, the straight line is the shortest (i.e., minimum length) route between two points. The minor arc of a great circle is the shortest (i.e., minimum length) path connecting two points on the surface of a sphere. In contradistinction, the other way around the great circle is the longest (i.e., maximum length) path between the two points. This example illustrates how the solution of an extreme problem can obtain either the minimum or maximum case.

Heron, however, was the first to make use of an extreme principle, namely minimization. In his *Catoptrica*, Heron showed that the path taken by a ray of light reflected from a plane mirror is shorter than any other reflected path that might be drawn between the source and point of observation. Heron then proved that, in a mirror, the angle of reflection is equal to the angle of incidence. In other words, if the angles differed from each other, the distance that the light traveled would not be the least possible. Heron, along with everyone else until the modern scientific revolution, thought that light travels with infinite speed and can cover any distance in no time. For this reason, it was natural for him to use the idea of minimum distance, for the idea of minimum time would not have occurred to him.

The bending of visual images of objects partially submersed in water has been noted since antiquity. In modern terminology, when a ray of light passes from one medium to another, it is bent (refracted). In Figure 12, θ_i is the angle that the incident ray makes with the normal (at the boundary of two layers) and θ_r is the angle that the refracted ray makes with the normal. In this example, light is traveling from a less dense medium to a more dense medium. In such a case, the ray bends toward the normal.

The ancient Alexandrian astronomers intuitively realized the need to correct for atmospheric refraction in computing times of rising and setting heavenly bodies. Ptolemy, slightly younger than Heron, made measurements of the angles of incidence and refraction for the passage of light from air to water and attempted to find a mathematical relationship between the two angles. However, this problem was more difficult than the reflection relationship and Ptolemy was unsuccessful. In fact, progress was not made until 1611 when Johannes Kepler (1571 – 1630) showed that, for any given pair of materials, the ratio θ_i/θ_r is (approximately) fixed for small angles; that is,

$$\frac{\theta_i}{\theta_r} \approx n \text{ when } \theta_i \text{ and } \theta_r \text{ are small} .$$

A contemporary of Kepler, Willebrord Snell (1580 – 1626) was a Dutch astronomer and mathematician. He received a Master of Arts degree from

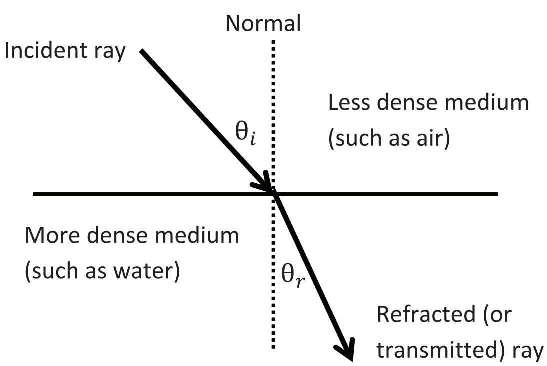


Figure 12. Refraction of light waves. The ray is bent toward the normal when it passes from a less dense to a more dense medium.

the University of Leiden in 1608, and he succeeded his father as professor of mathematics at that University in 1613. In 1617, Willebrord Snell published *Eratosthenes Batavus*, which contains his methods for measuring the Earth. He was the originator of triangulation, which is now the universally employed method in surveying and mapping large areas. To make measurements, he travelled quite widely in the Netherlands. As a baseline, he used the distance from his house to the local church spire. Then he devised a system of triangles which allowed him to determine the distance between the towns of Alkmaar and Bergen-op-Zoom, which was 130 km. His measurements were surprisingly accurate, allowing him to deduce a good value for the radius of the Earth. Snell also published several works on astronomy which contained data from his own observations.

In 1621, Snell discovered the law of refraction now known as Snell's law. Because he did not publish this result before his death in 1626, his law was not widely known at the time. Fortunately in 1703, Christiaan Huygens published Snell's law in *Dioptrica*. Snell's law stated that the ratio $\sin\theta_i/\sin\theta_r$ is fixed for any given pair of materials; that is, *Snell's law of refraction* was

$$\frac{\sin\theta_i}{\sin\theta_r} = n\ .$$

The constant n, called the *relative index of refraction* (or the *relative refractive index*), was characteristic of the two media.

Table 1 shows that the relative refractive index for air to water is 1.33 or $4/3$. Note that, for small θ, we have $\sin\theta \approx \theta$. This explains Kepler's version of the law:

$$n = \frac{\sin\theta_i}{\sin\theta_r} \approx \frac{\theta_i}{\theta_r} \quad \text{for small angles}\ .$$

Of course, as we all know, Snell's law is basic to understanding the paths taken by seismic waves. Seismic waves are mechanical waves, as are sound waves and water waves. For such waves, the refraction is in the opposite direction as that of light, as shown in Figure 13.

Both Kepler and Snell, along with everyone else of their day, assume the velocity of light is either infinite or else the velocity is so great that it might as well be infinite. Nowhere in their thinking does either of them make use of the velocity of light. In fact, they have no need to refer to the velocity of light at all. Their results are established by purely empirical findings such as shown in Table 1. The observational results are the data given in

Table 1. Refraction of light from air to water.

1. Incident angle θ_i in degrees (in air)	**2. Refraction angle θ_r in degrees (in water)**	**3. Kepler's ratio *n* of incident angle over refraction angle**	**4. Snell's ratio *n* of sine of incident angle over sine of refraction angle**
0	0		
10	7.5	1.333	1.33
20	14.9	1.343	1.33
30	22.1	1.357	1.33
40	28.9	1.384	1.33
50	35.2	1.420	1.33
60	40.6	1.478	1.33
70	45.0	1.556	1.33
80	47.8	1.674	1.33

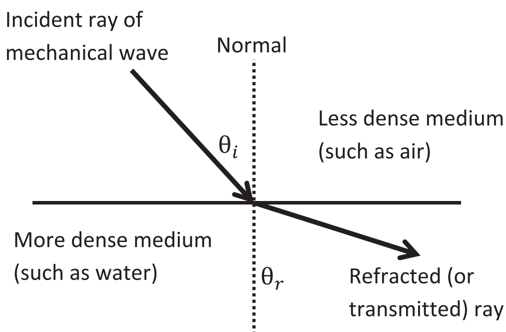


Figure 13. Refraction of mechanical waves. The ray is bent away from the normal (dotted line) when it passes from a less dense to a more dense medium.

columns 1 and 2 in the table. Kepler's result is confirmed by column 3. It shows that θ_i/θ_r is almost constant for θ_i less than about 20°. From the same data, Snell finds that $\sin\theta_i/\sin\theta_r$ (which is given in column 4) is constant for all θ_i.

Figure 14 represents in principle an experiment made by Snell. Project a pencil beam of light from air to water. If the beam is normal to the interface, then no bending occurs. If the beam is oblique to the interface, a bend is produced. The angle of the beam in the air is the angle θ_i of incidence. The amount of the bend varies with the angle of incidence. The angle of the beam in the water is the angle θ_r of refraction. In the case of refraction, as in the case of reflection, the path of a light ray is reversible. If instead the path of light is from water to air, the path would be the same. Observations of the

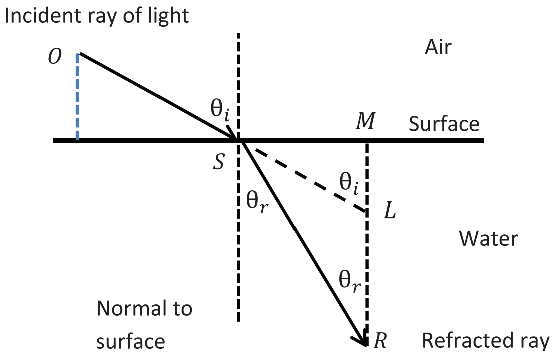


Figure 14. Refraction of light waves. The ray is bent towards the normal (dashed line) when it passes from a less dense to a more dense medium.

two angles θ_i and θ_r would be measured for various pairs of materials; for example, air–water, air–glass, water–glass. For any given pair of materials, the ratio of the sines of the two angles is constant, independent of the angle of incidence. Snell's law is an empirical law.

It is important to note that the law of refraction exists in two versions. One version is Snell's version. It is an *empirical law* which was inferred by careful study of observational data. The other version is the *theoretical law* which is mathematically derived from basic principles. The first great effort toward the theoretical law is by Descartes, but he was doomed to failure by the prevalent misunderstanding of the nature of light. The second great effort is by Fermat, who succeeded in establishing the theoretical law. In so doing, he makes the first critical leap toward the formulation of the principle of least action, which remains central in modern physics and mathematics. The third great effort is by Huygens, who succeeded in establishing the wave theory of light. He is the first to calculate the actual speed of light. Huygens' theoretical proof of Snell's law is the one given in every textbook.

Let us say a few words about René Descartes (1596 – 1650), whom we will discuss in the next section. In addition to being acclaimed as the father of modern philosophy and having his ideas garner the strict attention of Europe for nearly 100 years, Descartes is celebrated for constructing a whole mechanistic scheme of the cosmos. It is important to realize that Descartes worked in the early part of the 17th century. His era is that of Johannes Kepler (1571 – 1630) and Galileo Galilei (1564 – 1642). Kepler is renowned for his *Astronomia nova*, published in 1609. This book records Kepler's discovery of the first two of his three laws of planetary motion. Galileo is celebrated for his *Two New Sciences*, published in 1638 by Peter Elsevier in Leiden. This work contains the basis for all future work in dynamics.

Published in 1637 in Leiden, Descartes' *Discourse on Method* is a significant work in modern philosophy, and it was essential to the development of modern science. The book includes three appendices, *Optics, Meteorology*, and *Geometry*. In *Geometry*, Descartes lays the foundations of an entirely new area of mathematics: analytic geometry and the Cartesian coordinate system. Analytic geometry is the bridge between algebra and geometry. It is the branch of mathematics that addresses geometric properties using algebraic operations and notation to locate points within a coordinate system.

In *Optics*, Descartes struggles to find a mathematical model of light. Descartes assumes that light represents a force of motion that has a directional quantity. He derives the laws of reflection and refraction by setting up a model based on an analogy that compares the motion of a tennis ball to a ray of light. Like others of his time, he assumes the speed of propagation of light is instantaneous and does not understand that light travels as a wave. As a result, his analogy is doomed at the outset. It must be remembered that Descartes worked at the same time as Galileo. It is the very beginning of modern science. Descartes mathematically derives the same result as the empirical findings of Snell; namely, in refraction, the ratio of the sines is constant. Descartes' model on the refraction of light, although not physically valid, has permanent value because it is the first concrete example of mathematical model building. Since then, the Cartesian method of model building has been a mainstay of science. Descartes makes clear that a model can never provide an impeccable demonstration in the sense of Euclidean geometry, but it is a heuristic device that is used to discover laws, such as that of refraction, which can later be confirmed in experience. In another part of the *Optics*, Descartes explains the workings of the eyeball. He then elaborates a vision of the biological mechanisms that can be explained by the supposed laws of a mechanistic physics. Descartes' ideas are more relevant today than they ever could have been in his lifetime.

With quantum physics, we know much more about light than Descartes could ever have known in the early 17th century. By the late 17th century, there is a division. Newton adhers to the particle theory of light; Huygens adhers to the wave theory of light. The advent of the electromagnetic waves of James Clerk Maxwell in the 19th century spells the doom of Newton's particle theory of light. With Maxwell, light is no longer a stream of material corpuscles. Light is an electromagnetic wave, traveling at the speed of light and transporting energy from the source to receiver. The frequency of the wave determines its color. Infrared light has a lower frequency than visible red light. Visible red light has a lower frequency than visible blue light. Visible blue light has a lower frequency than ultraviolet light.

The amplitude of a wave is the height of its crests. The intensity of a mechanical wave is proportional to the square to the magnitude of its amplitude. Maxwell's electromagnetic theory says that the energy of an electromagnetic wave also depends upon its intensity, similar to a mechanical wave. According to Maxwell, the energy of the wave does not depend upon its frequency. As a result, red light and blue light of the same amplitude would have exactly the same energy. But what happens in a glowing fire as its fuel expires? Shakespeare writes:

In me thou see'st the glowing of such fire,
That on the ashes of his youth doth lie,
As the death-bed, whereon it must expire,
Consum'd with that which it was nourish'd by.

A blacksmith with his powerful bellows can make the coals in his forge burn blue-white. As the fire dies, the coals glow redder and redder. What happens to the blue light as the fire cools? As the fire cools, its total energy becomes less by a certain amount. According to Maxwell, the energy of each component wave would become less by the same amount. Suppose the total energy is reduced by 50% because of cooling. The blue energy would be reduced by 50%, and the red energy would be reduced by 50%. In fact, by the reasoning of Maxwell, the fading coals should display the same mixture of colors at all temperatures.

Maxwell's analysis, however, does not agree with reality. It is well known that hot blue-white coals become red as the fire cools. It is for that reason that Max Planck suspected something was amiss with the theory of electromagnetism. What could be done? In 1900, Max Planck proposes a drastic solution to this problem. He suggests that light is composed of particles that move in a wavelike manner. These particles of a light wave are called *photons*.

Let us describe a photon. Each photon has an energy that increases in proportion to the frequency of the light wave. Because blue light has a higher frequency than red light, it takes more energy to activate blue light than red light. Let us look at this statement by way of analogy. Suppose that a powerful medieval country has an army of 10,000 red foot soldiers and 2000 blue knights. Over the years, the country becomes much weaker, so it cannot field such a formidable army. What would the new army be like? The obvious answer would be that everything is reduced by one-half, such as 5000 red foot soldiers and 1000 blue knights. That is the answer provided by Maxwell's theory. Yet that is not what happens in actuality. The reduced army might still have 6000 soldiers, but they would be all red foot soldiers and no blue knights at all. Why?

The reason is simple. A blue knight needs a horse, weapons, and armor. A red foot soldier needs bows and arrows. It takes a certain threshold of energy to forge steel to make swords and armor. The weakened county would not be able to meet that threshold. However, the country still would have enough energy to field a reduced number of red foot soldiers.

How does this apply to photons? Let us go back to the cooling fire. According to Planck, the energy of each photon depends upon its frequency. A blue photon has a high frequency, and thus a high energy content. It requires a high threshold of energy for the fire to activate a blue photon. When the fire cools, its total energy decreases. The more energetic blue photons become more difficult to excite. At some point, the cooling fire would not be able to meet the blue threshold at all. This explains why blue light is absent in the dying coals in the forge. The lower-energy red photons can be excited more easily at lower temperatures, as evidenced by the preponderance of red coals.

Let us now introduce the fundamental equation of quantum physics. According to quantum physics, electromagnetic energy can be emitted only in quantized form. In other words, the energy is a multiple of an elementary unit $E = hf$. In this equation, E is the energy of a photon, h is a fundamental constant, and f is the frequency. The intensity of the light is simply a measure of the total number N of photons in the light wave. The wave will have the total energy $E_{\text{total}} = Nhf$. In other words, the total energy of the wave equals the number N of photons times the constant h times the frequency f of each photon.

Let us summarize the contribution of Planck. In effect, Planck shows that Maxwell's wave theory does not account for all of the properties of light. The Maxwell theory predicts that the energy of a light wave depends only on its intensity, not on its frequency. However, experiments show that the energy imparted by light to atoms depends on the light's frequency, not on its intensity. For example, some chemical reactions are provoked only by light of a frequency higher than a certain threshold. Light of a frequency lower than the threshold, no matter how intense, does not initiate the reaction.

In the way of the analogy, suppose an army of 1000 red archers attack a castle. A thousand arrows cannot break down the walls. Increase the intensity of the red archers to 10,000. Ten thousand arrows cannot break down the walls. The destructive power does not depend on intensity of the red archers. One blue cannon, however, can break down a wall. It is not the intensity but the type of weapon that matters. A red archer corresponds to a low-frequency wave and the blue cannon to a high-frequency wave. The blue high-frequency wave carries more energy that the red low-frequency wave.

Through the analysis of the thermally produced light, Max Planck is credited with inventing the fundamental constant h that defines quantum mechanics. It is called Planck's constant. In fact, Planck's constant and the speed of light are two of the most important physical constants.

Let us now outline the modern theory of how light behaves. According to Einstein, light can travel only at one speed; it is the speed c of light in a vacuum which is approximately 300,000 km/s. Under no circumstances can light speed up or slow down. Hence, light travels as fast in glass as it does in a vacuum. But light goes slower in glass than in air. There seems to be a contradiction. How does light travel in a material substance? It is similar to the way that any other wave is transported, namely, by particle-to-particle interaction.

An electromagnetic wave (e.g., a light wave) is produced by a vibrating electric charge. As the wave moves through the vacuum of empty space, it travels at a speed c. When the photon hits a particle of matter, the energy is absorbed. If conditions are right, the absorbed energy is reemitted in the form of light. The new light wave has the same frequency as the original wave. The new light wave travels at a speed of c through the vacuum between atoms until it hits a neighboring particle. The energy is absorbed by this new particle and again the energy is reemitted in the form of a new wave. It is in this way that the energy is transported from particle to particle through the medium. Every photon travels between the interatomic vacuum at speed c. However, because of the time delay in the process of absorption and reemission, the net speed of light v in the medium is reduced. For example, for visible light the refractive index of glass is typically around 1.5, meaning that light in glass travels at $c/1.5 \approx 200{,}000$ km/s.

Now we come to an enigma that held everyone in the dark until the light of Maxwell. From the time of Pythagoras, people have been familiar with mechanical waves as exemplified by sound. Sound travels faster in a dense substance, such as steel, than in a sparse substance, such as air. Maxwell shows that light is an electromagnetic wave, which is quite different from a mechanical wave. Light travels slower in a dense substance, such as steel, than in a sparse substance, such as air. When rays pass obliquely from one transparent body to another, they are deflected so they make a larger angle to the normal for the body with the greater speed. As shown in Figure 15, refracted sound (having greater speed) bends from the normal and refracted light (having lesser speed) bends toward the normal. In conclusion, the contrast in velocity is what determines diffraction, not

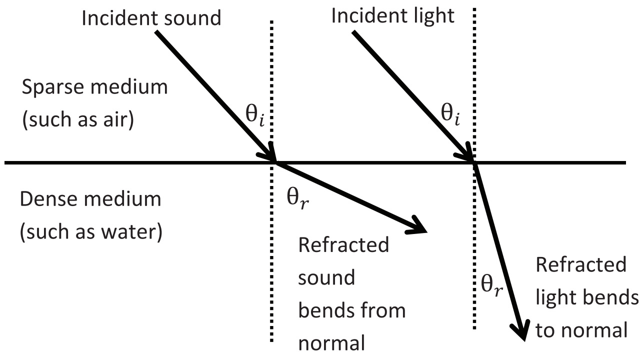


Figure 15. (left) Refraction of sound versus (right) refraction of light.

the contrast in density. For electromagnetic waves (such as light), the velocity is always less in the denser medium, whereas for mechanical waves (such as seismic waves), the velocity is generally greater in the denser medium.

Descartes as a geophysicist

Not many years ago, a few miles outside of Tulsa, it is said there was an Oklahoma farmer who had a horse with remarkable mathematical ability. The animal quickly mastered arithmetic and elementary algebra and then took on geometry and trigonometry with little difficulty. However, shortly after beginning his first lesson in analytic geometry, the horse began balking and kicking in his stall. The farmer racked his brain for an explanation. Finally, it came: “You should never put Descartes (i.e., the cart) before the horse.”

That probably apocryphal story is evidence that today Rene Descartes’ scientific reputation rests primarily on his mathematical contributions, chiefly the system of coordinate axes. (It is known as the Cartesian coordinate system because Descartes, in the style of the time, often used a Latinized version of his name—Renatus Cartesius—in signing his writings.) The value of this discovery, which led directly to analytic geometry and probably at least indirectly to most modern mathematics, overshadows Descartes’ importance in many fields, including geophysics which unquestionably owes him a great debt. If Descartes’ work had been limited solely to his geophysical and meteorological investigations and speculations, he would today be viewed as a great pioneer. Incredibly, that also could be said of him in an astounding number of intellectually demanding disciplines. Historians of science list mathematics, physics, astronomy, anatomy, physiology, psychology, metaphysics, epistemology, ethics, and theology—none of which represents his primary interest.

His overwhelming interest, of course, is philosophy. Often he is called the father of modern philosophy and he is unquestionably one of its towering figures. Even without his many impressive scientific contributions, Descartes' importance to science would be immense because of his pivotal role in ending the centuries-long domination of both science and philosophy by Aristotle. This revolution is very possibly the critical step in generating the era of rapid intellectual progress, particularly in science, the world subsequently has enjoyed.

Descartes was born in the village of La Haye (near Tours and now known as La Haye-Descartes), France in 1596, the son of a lawyer who apparently always regretted that his obviously brilliant offspring had failed to devote his great intellect to his own profession. Descartes traveled extensively from 1616 to 1628 and then settled in Holland where he lived for the next 21 years (although he changed his residence about 20 times) and did nearly all of his famous work. In 1649, he accepted a generous financial offer from Queen Christina of Sweden. Unfortunately, the eccentric queen demanded that she receive her philosophy instruction at 5 A.M. Descartes, a late riser all his life, promptly caught pneumonia and did not survive the harsh Swedish winter. He died in February 1650, only four months after arriving.

Published in 1637 is Descartes' most famous book and certainly one of the most remarkable of all time, *Discourse on the Method of Rightly Conducting the Reason and Seeking Truth in the Sciences* (usually called simply *Discourse on Method*). Among its many distinctions is its general regard as a masterpiece of French prose. This is the book which contains the most famous, and probably most quoted, sentence in philosophy: "I think therefore I am."

The book's impact on many fields is immense but it is incalculably important in the evolution of the modern scientific method. In *Discourse*, Descartes argues that scientists should avoid vague notions and try to describe the world by mathematical equations. The results of such research, he asserts, would be knowledge that would have practical applications of great benefit to society. Descartes provides examples of the way his method works in three appendices which describe discoveries he personally had made using the system. These appendices, particularly the third, are today more famous than the body of the text itself and they contain a number of ideas that are central to geophysical exploration.

The first appendix, *Optics*, contains the first publication of the law of refraction. This seems strange because, as all geophysicists know, the law is named after Willebrord Snell who had made the discovery in 1621 but did not publish it. Although Descartes was accused of appropriating the idea without giving proper credit to Snell (who died more than a decade before *Discourse's* publication) and the controversy has been lasting, opinion currently seems to be that Descartes discovered the law independently. Many call it the *Snell-Descartes law of refraction.* In *Optics*, Descartes also discusses lenses and optical instruments, describes the functioning of the eye and several of its malfunctions, and presents a preliminary version of an enormously important scientific concept—the wave theory of light.

The second appendix, *Meteorology*, is the first modern treatment of that subject. Descartes examines rain, wind, and clouds and provides a correct explanation for the rainbow. He argues against the concept that heat is an invisible fluid, correctly concluding that it is a form of internal motion.

The third appendix, *Geometry*, presents the foundation for the mathematical discipline now called analytic geometry. A number of legends describe the initial hint which prompted Descartes toward this discovery. One says that the inspiration came while he was watching a fly crawl about on the ceiling near the corner of his room, from which it strikes him that the path of the fly could be described by a mathematical relation connecting the fly's distances from the two adjacent walls. This concept is revolutionary because it links geometry, the form of mathematical reasoning dominant in the West from the earliest times, to algebra, which had been brought to an advanced state in the East. The union makes it possible to transform geometric problems into ones expressible in algebraic terms. Analytic geometry is the mathematical advance that prepared the way for the invention of calculus about 30 years later.

Of course, the use of coordinate systems of axes is essential to nearly all geophysical surveys. The familiar seismic sections are two-dimensional graphs with distance on the horizontal axis and time on the vertical; the desired results of processing and interpretation are three-dimensional maps of blocks of the earth's crust. Descartes has several other geophysical ideas which have been negated as scientific knowledge has grown but which remain fascinating because of their ingenuity. For example, attraction, the basic property of magnetism, has been known since very early times wherever lodestone was present. Lodestone, a magnetized piece of magnetite, attracts not only other pieces of lodestone but pieces of ordinary iron as well. After the magnetic compass arrived in the West from China in the 13th century, various ideas of magnetism began appearing, but Descartes is the first to propose a scientific theory to account for the phenomenon.

Descartes believed that space was filled with vortexes, inside of which were some screw-like particles. These particles had left-handed or right-handed threads, depending upon from which pole of the vortex they derived. The particles entered through the north and south poles of the earth and screwed into holes or pores in its iron or lodestone interior, thus giving rise to the earth's magnetism. Lodestone on the earth's surface, in order to offer the least resistance to the circulation of these particles, lined up its holes in the correct direction. Although this concept did not stand the test of time, the explanation was ingenious enough to account for all magnetic phenomena known in 1644. (It was only after the experimental work in electricity and magnetism in the 19th and 20th centuries that really satisfactory explanations became possible.) Thus, Descartes may be considered one of the founders of the geophysical discipline of terrestrial magnetism.

Although early cultures speculated on the creation of the universe and the beginning of the world, they could do little more than use imagination and poetry. Scientific reasoning was not an evolving field. Descartes, in the *Principles of Philosophy* published in 1644, was the first person to attempt a scientific explanation. It has long been superseded but remains interesting, not only for historical value, but because it contains life cycles for stars, a striking similarity to modern cosmogonies. At that time, the implications of the entire universe, as we know it, were not appreciated. The main concern of astronomy was with the comets and planets. Within his universe, Descartes assumed that everything was arranged as a collection of vortexes. At the center of each vortex of whirling matter was a star. As time passed, a skin grew over each star—the spots on our own sun representing the beginning of the process. In due course, the skin would stop the star from shining and then the vortex would collapse. The dead star could become either a comet moving from one vortex to another or a planet staying in another vortex moving in orbit around the star in the center. At the time, Descartes' system appeared a brilliant synthesis for it was the first new physical scheme of the universe since that of Aristotle. While Descartes' theory became untenable with the publication of Sir Isaac Newton's concepts in 1687, it was not until the middle of the next century that the Newtonian universe fully replaced Descartes' vortexes in the European scientific community.

The Dane Niels Stensen (known also by the Latinized Nicolaus Steno) published *Prodomus (The Messenger)* in 1671, which usually is considered the advent of scientific geology. Stensen was concerned with the stratigraphy of the earth's outer crust. However, as early as 1644, Descartes had made some conjectures about the earth's internal constitution. He, again in *Principles of Philosophy*, said that the earth was once molten (as he believed the sun to be) and that, although the earth had now cooled and shrunk on the outside, its inner core was still very hot. He believed that a layer of material surrounded the incandescent central core. In the 20th century, geophysics partially repaid its debt to Descartes when earthquake seismology confirmed the existence of the mantle and showed that his model essentially was correct.

Descartes assumed that the universe was filled with matter which, due to some initial motion, had settled down into a system of vortices which carry the sun, the stars, the planets, and comets in their paths. The Cartesian system seemed self-evident because no one could admit the idea that the great masses of planets were suspended in empty space and that they were held in their orbits by an invisible influence. In contrast, Newton's theory proposed that gravitation acts instantaneously, regardless of distance. In other words, Newton assumed that the universe was governed by action at a distance. Newton supposed that a change in intrinsic properties of one system induced a change in the intrinsic properties of a distant system without any intervening process that carried this influence contiguously in space and time.

Descartes also plays a role in the science of mechanics, ironically helping prepare the way for Newton to overturn the Cartesian cosmogony. The first coherent body of laws of motion is conceived by Aristotle; but, as the Dark Ages ended in Europe, dissatisfaction arose concerning Aristotle's explanation of the motion of an arrow. At the start of its motion, the arrow is obviously driven forward by the bowstring. But what keeps it moving after this? According to Aristotle, as soon as the arrow leaves the bow, air would rush in to take up the space that the arrow has occupied. Then this air would act as a force to push the arrow along its course and thereby cause the motion to continue. Once this force is spent, the arrow would seek its natural place and fall to the ground, unless it first reaches the target.

There were obvious flaws in this theory (mainly because it was known that air offered resistance to the movement of bodies) but the overwhelming stature of Aristotle deterred serious questioning for centuries. Roger Bacon expressed doubts in the 13th century but the first steps toward a new explanation had to wait for Galileo (1564 – 1642). Descartes followed up Galileo's work and claimed that a body will continue its motion indefinitely at a fixed velocity and in the same direction unless it suffers a collision. This statement, now known as Newton's first law of motion (the law of inertia), first appeared in *Principles of Philosophy.* The idea was demonstrated by an illustration (Figure 16) from Descartes' *Epistolae*, posthumously published in 1668. It showed that a tennis ball falls to the ground in a smooth parabolic path after being hit.

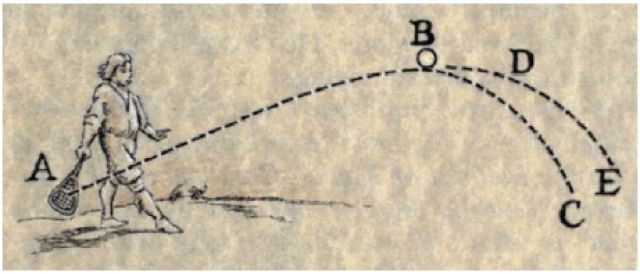


Figure 16. Tennis ball illustration from Descartes' *Epistolae* (1668).

There is one additional scientific development, of immense importance in exploration, with which Descartes was probably at least tangentially associated. During his years in Holland, he became a friend of the distinguished diplomat and poet, Constantijn Huygens, and was an occasional visitor to his home. There he met Huygens' son, Christiaan. Descartes is known to have read and been impressed by the younger Huygens' teenage efforts in geometry. Christiaan Huygens, of course, became one of the greatest of all scientists, ranking second in esteem in his lifetime only to Newton himself. Among other major scientific accomplishments, Huygens was the author of the wave theory of propagation, with Huygens' principle underlining the very foundation of reflection seismology. The concepts of mass, weight, momentum, force, and work were finally clarified in Huygens' treatment of the phenomena of impact, centripetal force, and the pendulum.

Could Huygens' interest in science—particularly in the study of light—have been spurred by his early contact with Descartes? We will never know for certain, but it is difficult to imagine there was no influence at all. It is possible, perhaps even probable, that something Descartes said during their conversations gestated within Huygens for years before emerging as the wave theory. The wave theory of light does not make much of an impression initially, primarily because the unassailable Newton favored particles. Huygens' principle would lay dormant for more than a century before being resurrected by the brilliant experiments of Thomas Young in 1803. In 1816, it is Augustin-Jean Fresnel who shows that Huygens' principle, together with his own principle of interference, could explain both the rectilinear propagation of light and also diffraction effects. Since then, the wave theory has become a key idea in physics. Ultimately, the recognition of the fundamental duality existing between the propagation of particles and waves has become the very foundation of modern physics.

In 1663, Huygens was elected to the Royal Society of London, which had been founded in 1660. In 1666, at the invitation of King Louis XVI, Huygens moved to Paris as the foremost member the newly formed

French Academy of Sciences. The parallel between the works of Huygens and Newton was noteworthy. Working within the same 17th century milieu, they both encountered some of the same problems. For example, the wave theory of light of Huygens and the particle theory of Newton finally came together in the realm of quantum physics. No other worker in that decisive era of science approached the level or scope of Huygens and Newton.

The publication priority of the law of refraction remains with Descartes in his *Optics* (1637) in which he states his conclusion but presents no experimental verification. Descartes' discourse on refraction is interesting in itself—even if it is significantly out of date—because it opens the door to the physical understanding of light. Remember that Descartes regards light as a force of motion that has a directional quantity. In other words, light is a ray with length and direction.

Descartes first considers the law of reflection by harking back to Heron's analogy between the motion of a ball (which has finite velocity) and light (which, according to Descartes, has infinite velocity). He first analyzes the motion of the ball upon bouncing from a hard surface (Figure 17). The ball, being impelled from *A* toward *B*, meets the surface of the ground at point *B* and bounces.

In order to predict the direction that the ball must go next, Descartes describes a circle centered at *B* and passing through *A*. Because the ball moves at a constant speed, the time it takes to travel from *B* to another point on the circumference of the circle must be the same time it takes the ball to travel from *A* to *B*. Descartes draws three straight lines *AC*, *HB*, and *FE* (each perpendicular to *CE* and such that *CB* equals *BE*).

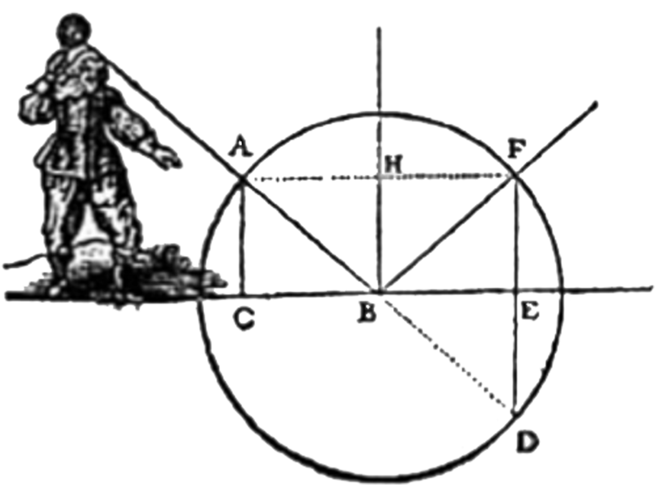


Figure 17. Descartes' law of reflection.

Then he says that the time it takes the ball to advance to the right from *A* to *B* must be equal to the time it takes the ball to advance to the right from *B* to some point on the line *FE*. The ball cannot be simultaneously at a point on the line *FE* and at a point on the circumference of the circle unless it is either at point *D* or point *F* because these are the only points where the circle and line intersect. Because the ground prevents the ball from continuing toward *D*, it must go to *F*.

Descartes then makes the analogy of the paths of the ball to rays of light. He asserts that you can easily see how reflection occurs. If a ray coming from point *A* falls to point *B* on a flat mirror, it is reflected toward *F* in such a manner that the angle of reflection *FBH* is equal to the angle of incidence *ABH*.

Next Descartes treats refraction and again uses the tennis ball analogy, as illustrated in Figure 18. Redrawing and labeling the illustration produces the diagram shown in Figure 19.

Descartes supposes that a ball impelled from *A* to *B* does not meet an unyielding surface at the *CBE* plane but rather a cloth. Because the cloth is weak and loosely woven, the ball ruptures it. Descartes then continues his description, but we will put his words into equations. Both *AB* and *BI* are radii of the same circle. Denote the radius by *R*; that is,

$$R = AB = BI \ .$$

(This is another example of Descartes' invention of the use of coordinate systems in mathematics.) Denote the segment *CB* by d_1. Let θ_1 denote

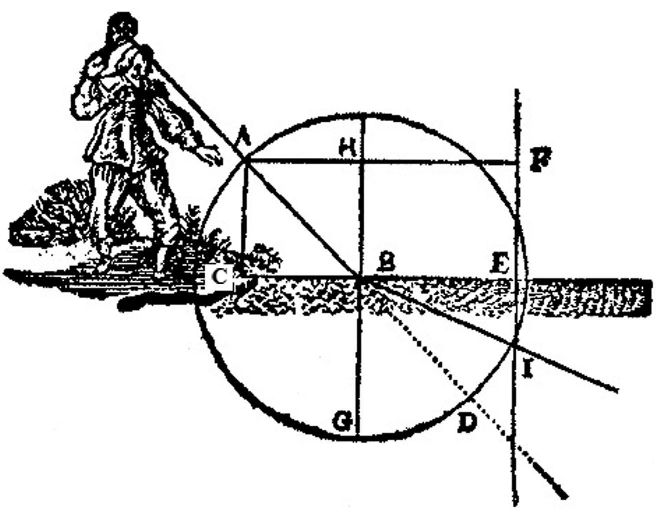


Figure 18. Descartes' law of refraction (*C* clarified for reference).

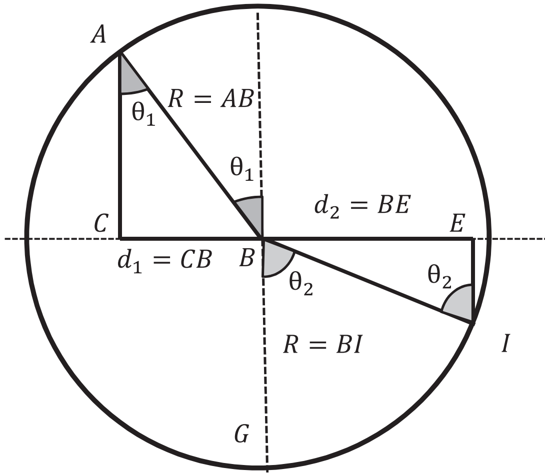


Figure 19. A redrawn and relabeled representation of Descartes' law of refraction shown in Figure 18.

the angle of incidence *CAB*. Thus, $d_1 = R \sin \theta_1$. Denote the segment *BE* by d_2. Let θ_2 denote the angle of refraction *BIE*. Thus, $d_2 = R \sin \theta_2$. Descartes wishes to find d_2 and hence θ_2. As to horizontal component, he faces two choices:

$$\text{Choice (1): } \frac{d_2}{v_1} = \frac{d_1}{v_2} \ , \quad \text{Choice (2): } \frac{d_2}{v_2} = \frac{d_1}{v_1} \ . \tag{9}$$

As to component choice, Descartes chooses Choice (1). Descartes makes an analogy with a tennis ball. As to velocity, he faces two choices:

$$\text{Choice (1): } v_2 > v_1 \ , \quad \text{Choice (2): } v_2 < v_1 \ . \tag{10}$$

Nobody knows anything about light. Descartes, in talking about a tennis ball, in effect is talking about a particle of light. Descartes writes, "The harder and firmer are the small particles of a transparent body, the more easily do they allow the light to pass." In other words, it is easier for light to pass through water than air. Thus, (according to Descartes), the velocity v_2 through water is greater that the velocity v_1 through air. Thus, on the v_2 choice, Descartes chooses Choice (1). With these choices, Descartes obtains

$$\frac{R \sin \theta_2}{v_1} = \frac{R \sin \theta_1}{v_2} \ . \tag{11}$$

As a result, the relative refractive index is

$$n = \frac{\sin \theta_1}{\sin \theta_2} = \frac{v_2}{v_1} > 1 \ . \tag{12}$$

By theoretical means alone, Descartes justifies Snell's empirical law of refraction, which for air to water is

$$n = \frac{\sin\theta_1}{\sin\theta_2} > 1 \ . \tag{13}$$

(In the same vein, Newton in 1678 finds theoretical means of verifying Kepler's empirical laws on the motion of planets.) Descartes faces two choices on each account and, in hindsight, we can see that Descartes is wrong on both accounts. Fortunately, for Descartes at least, two wrongs can make a right.

Now let us look at how we can redeem the derivation of Descartes. In regard to the horizontal component, he should have taken the other choice; namely,

$$\frac{d_2}{v_2} = \frac{d_1}{v_1} \ . \tag{14}$$

In regard to the velocity, he should have taken the other choice; namely $v_2 < v_1$. In other words, light slows down in water. Equation 14 becomes

$$\frac{R\sin\theta_2}{v_2} = \frac{R\sin\theta_1}{v_1} \ . \tag{15}$$

As a result, the relative refractive index is

$$n = \frac{\sin\theta_1}{\sin\theta_2} = \frac{v_1}{v_2} > 1 \ . \tag{16}$$

Thus, the mathematics is corrected easily. Now we must turn our attention to the model. It turns out that the diagram of Descartes is correct. How could it not be correct? It is drawn by the greatest geometer since Euclid. The analytic geometry of Descartes changes the whole scheme of mathematics and makes possible infinitesimal analysis. In order to explain this, we must first look ahead to the work of Huygens in 1678.

Huygens sets everything right. Descartes uses a tennis ball as an analogy for light. In other words, Descartes thinks of light as a particle. Huygens, in 1678, proposes that light travels as a wave. What does that imply? In ancient times, rowing vessels, especially galleys, are used extensively in the Mediterranean for naval warfare and trade. Galleys have advantages over sailing ships. They are easier to maneuver, capable of short bursts of speed, and able to move independently of the wind. If the starboard oarsmen and the port oarsmen each row at the same speed, the galley travels in a straight line.

If both sides slow down together, the galley maintains its direction. If the starboard oarsmen slow down as directed but the port oarsmen are late in slowing down, the galley veers to starboard. A light wave in air slows down when it hits water. If the light hits the water at normal incidence, the light wave slows down but continues in the same direction. If the light hits the water at oblique incidence, however, then one side of the wavefront hits the water sooner than the other side. Like the galley, the light changes its direction. The passage of light cannot be explained by a single ray. At least two closely spaced rays must be used. In an isotropic medium, the wavefronts are orthogonal to the rays. The distance between two rays along a wavefront is a measure of the width of the wavefront.

Figure 20 is Huygens' model (as also will be shown in Figure 34, in the subsequent section on Huygens' model). The same model is given by Fermat (as will be shown in Figure 24). Interface *R* separates air on top and water below. There are two raypaths and two wavefronts. We must now show how these qualities are placed upon Descartes' model.

Figure 21 is the essence of Descartes' model. The left semicircle represents air; the right semicircle represents water. The bent raypath composed of d_1 in air and d_2 in water is made straight along the line *CBE* in Descartes' model. The interface *R* appears twice in Descartes' model. The upper *R* represents the air side; the lower *R* represents the water side.

Transit maps are found in the transit vehicles and at the platforms. They help people efficiently use the public transport system, including the stations which function as interchanges between lines. Unlike conventional maps, transit maps usually are not geographically accurate—instead they use straight lines and fixed angles, and often illustrate a fixed distance between stations, compressing those in the outer area of the system and expanding those close to the center. In effect, Descartes' diagram is a transit map of Huygens' diagram. In Descartes' transit map, the raypaths

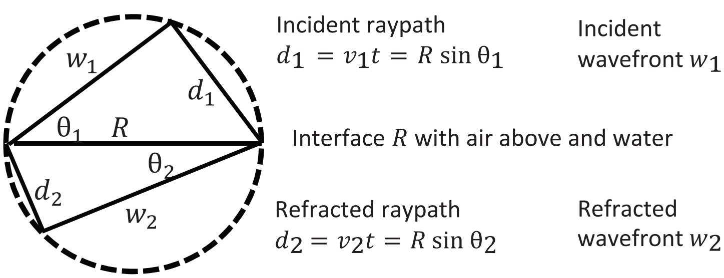


Figure 20. Huygens' model of Snell's law.

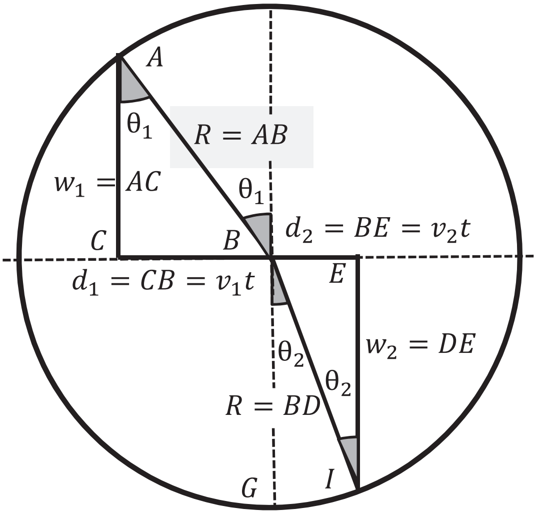


Figure 21. Descartes' law of refraction as a transit map showing rays and wavefronts.

are horizontal lines and the wavefronts are vertical lines which, if extended, would make a perfect orthogonal Cartesian grid.

The vertical lines *AC* and *EI* represent wavefronts. The bent ray is represented by the straight line *CE*, where *B* is the point of bending. The two sides of the interface *R* are represented by the bent line *ABI*, where *B* is the point of separation of the air side from the water side. Triangle *CAB* represents what happens in the upper medium (air). Line *AB* is the interface, and angle θ_1 is the angle between the wavefront and the interface. The distance d_1 is the distance that the wavefront travels in time t. Thus,

$$d_1 = v_1 t = R \sin \theta_1 \ . \tag{17}$$

Triangle *BIE* represents what happens in the lower medium (water). Line *BI* is the same interface, but on the water side. Angle θ_2 is the angle between the wavefront and the interface. The distance d_2 is the distance that the wavefront travels in time t. Thus,

$$d_2 = v_2 t = R \sin \theta_2 \ . \tag{18}$$

If we divide equation 17 by equation 18, we obtain the Huygens' equation for refraction

$$n = \frac{\sin \theta_1}{\sin \theta_2} = \frac{v_1}{v_2} \ . \tag{19}$$

Descartes freed geometry from the shackles of diagrams and opened up the realm of geometric objects defined on a coordinate system. Descartes immediately used his analytic geometry to build a mathematical model that would explain refraction. Because Descartes (like everyone else of his time) did not know how light traveled, his model was in error. Francis Bacon (1561 – 1626), a contemporary of Descartes, wrote, "Truth comes out of error more readily than out of confusion." The validly of Bacon's statement was revealed by the fact that Fermat and Huygens, using analytic geometry, were able to correct Descartes' error and devise the correct model for the refraction of light. Analytic geometry almost completely replaced diagrams with mathematical equations. The last great proof using diagrams only was Newton's proof of the law of gravity in 1689 (as will be shown in our subsequent section on Newton).

Descartes' Mathematical Problem. As a challenge to the reader, we present one of the problems posed by Descartes in *Geometry*. The problem is to find the roots of the equation $z^2 - az + b^2 = 0$. Descartes does it with the diagram in Figure 17. He makes NL equal to $a/2$ and LM equal to b. Then he draws MQR parallel to LN, and with N as a center describes a circle through L cutting MQR at the points Q and R. The problem is to show that the two roots are MQ and MR. A second challenge is to explain what happens to the roots when the line MR is tangent to the circle and when it does not meet the circle.

Descartes' Solution. From high school algebra, we recall that the solutions of the quadratic equation of the form

$$Az^2 + Bz + C = 0$$

are given by the formula

$$z = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A} .$$

In Descartes' problem, we have $A = 1,\ B = -a$, $C = b^2$, so in terms of the given quantities the two solutions are

$$z = \left(\frac{a}{2}\right) \pm \sqrt{\left(\frac{a}{2}\right)^2 - b^2} .$$

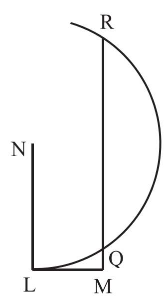


Figure 22. Construction of Descartes.

Now turn to Figure 22. Draw NP parallel to LM. Thus, NP also is equal to b. Also draw line NQ, which is a radius of the circle, so NQ is equal to

NL, which by construction is equal to $a/2$. Next use the Pythagorean theorem, which says that

$$PR = PQ = \sqrt{\left(\frac{a}{2}\right)^2 - b^2}\ .$$

Finally note that PM equals $a/2$. Therefore, in terms of the figure, the two algebraic solutions can be written $z = PM + PR$ and $z = PM - PQ$, or in other words the two solutions are MR and MQ. When line MR is tangent to the circle, both roots are equal, and when line MR does not meet the circle, the roots are complex.

Fermat and the principle of least time

Descartes indeed has a role in the discovery of the law of refraction (generally credited to Willebrord Snell), but Descartes' work has one major flaw—it essentially puts things exactly backwards! More specifically, light travels faster in air than in water, but Descartes assumes the opposite. Moving forward from this "mistake," however, we have Fermat. Born in Beaumont-de-Lomagne, France, Pierre de Fermat (1601 – 1665) devises a principle, now known as Fermat's principle, which forms the basis of our understanding of the propagation of seismic waves through the earth.

In his famous (perhaps notorious) "Last Theorem," Fermat stated—actually scribbled in the margin of a book—that he had found how the fundamental equation

$$x^n + y^n = z^n \quad (n > 2)$$

did not have a solution in whole numbers and that the proof was simple but he did not have room to write it in the available space. For three centuries, mathematicians failed, despite great effort, to confirm his statement, until 1994 when Andrew Wiles published a proof for Fermat's Last Theorem. With more than 150 pages, it was described as one of the highest achievements of number theory. He had accomplished what was considered by many as impossible. Fermat invented analytic geometry independently of Descartes (and extended it to three dimensions instead of stopping at two), apparently intuited the basic principles of differential calculus before Newton, was at least a co-founder (with Pascal and Huygens) of probability theory, and inaugurated the modern "theory of numbers."

It took many years for Fermat to receive credit for much of his amazing body of work because he was an amateur mathematician—most of his formal education and his entire career were in law—so he devoted only his spare time to the subject. He did not publish in normal mathematical channels of communication; many of his ideas were circulated only in letters to friends and these were not published until 1679, 14 years after his death near Toulouse, France.

Fermat also contributes to physical theory, most importantly with Fermat's principle, also known as the principle of least time. This concept has had tremendous influence on the development of physical thought in and beyond the study of classical optics (to which Fermat first applied it). As previously stated, it is absolutely indispensable to our current concept of seismic wave propagation. A variety of natural phenomena exhibit what might be called the minimum principle, or its twin the maximum principle. These principles find expression in certain geometric statements. (For example, a straight line is the shortest distance between two points on a plane or a circle encloses the largest area of all closed curves of equal length on a plane.) Many of these examples are known to the ancients. One story says that the Phoenician princess Dido obtained a grant from a North African chief, the grant being for as much land as she could enclose in an ox hide. Dido cut the hide into long, thin strips; tied the ends together; and staked out the rectangular area upon which Carthage was built. She could have obtained even more land if she had laid out the city in the shape of a circle.

As shown in our previous section, Heron of Alexandria is believed to be the first to apply an extreme principle to light. He obtains the *law of reflection*, namely, the angle of incidence is equal to the angle of reflection. The related law of refraction, which is basic to geophysical exploration, is studied experimentally by the Greek-Egyptian astronomer Ptolemy in the first century AD; however, it is not formulated for another 1500 years. Johannes Kepler, in his study of optics in the early 17th century, develops many hypotheses—some shrewd and close to the mark—concerning the refraction of light, but nothing of the first magnitude results from this research.

The *law of refraction* is discovered some years later by Snell and, independently, by Descartes. Snell formulates it around 1621, after many years of experimentation as well as the study of Kepler's *Ad Vitellionem Paralipomena* (1604) and Risner's *Optica* (1606), both of which quote lbn al-Haytham and Witelo.

Pierre de Fermat, a contemporary of Descartes, contested this result because he believed that light must travel slower (not faster) in a denser medium. Fermat had studied the works of the ancient mathematicians in the original Greek and, consequently, was inspired to derive the law of refraction by using Heron's assumption that light always prefers the shortest possible path. However, it was immediately obvious to him that it could not be the shortest path, as Heron reasoned for reflection, because the shortest path between a point above water and a point below water would be a straight line, and there would be no refraction (i.e., no bending) at all. Explaining this difference caused Fermat to make one of the most important conceptual leaps in the history of physics. His intuition told him that light moves slower in the denser medium. Thus, instead of assuming that light travels along a path that minimizes distance, Fermat reasoned that light travels along the path that minimizes time. This became *Fermat's Principle of Least Time*.

Fermat's conceptual leap was astounding because he had to assume

1. the speed of light is finite (which had not yet been demonstrated),
2. light has a fixed characteristic speed in each substance, and
3. counterintuitively the speed is slower in denser media (which is opposite to the view of Descartes and counter to the physics of sound).

Furthermore, because calculus had not yet been invented, Fermat also has to do pioneering work in mathematics to derive the law of refraction based upon his principle of least time. His method of minimization involves ingenuity that is very close to differential calculus. Figure 23 shows two neighboring paths, labeled old path and new path, from point *A* to point *B*. The essential component is that Fermat uses two rays, which are critical to understanding that light travels as a wave not as a particle.

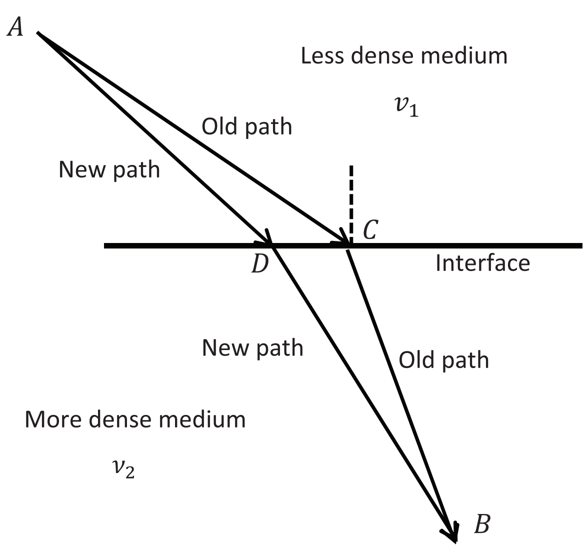


Figure 23. Derivation of Fermat's law of refraction. The raypath *ADB* is a slight displacement from raypath *ACB*.

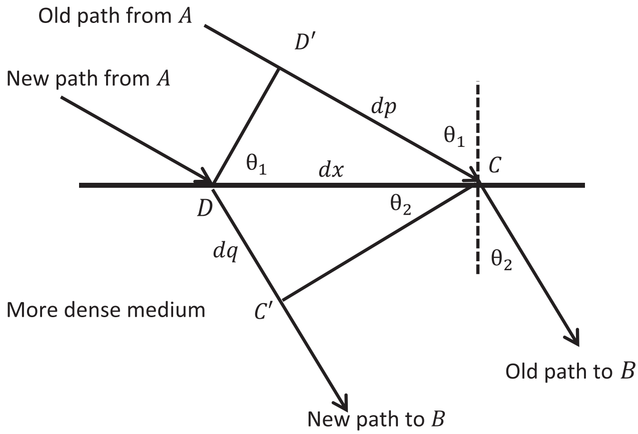


Figure 24. Enlargement of Figure 14 near the interface.

In Figure 24, note the perpendicular DD' to the old path and the perpendicular CC' to the new path. The figure shows the mathematics near the interface when DD' is considered infinitesimally small. The upper segment AD of the new path is shorter (by the amount dp) than the upper segment AC of the old path. In compensation, the lower segment DB of the new path is longer (by the amount dq) than the lower segment CB of the old path. The new path "gains" the time it would take to go the distance dp, but "loses" the time it would take to go the distance dq. Time is distance divided by velocity. Thus, the net time difference dt is $dp/v_1 - dq/v_2$. Some basic trigonometry reveals that $dp = dx \sin\theta_1$ and $dq = dx \sin\theta_2$. As a result, the net time difference can be written

$$dt = dx\,\frac{\sin\theta_1}{v_1} - dx\,\frac{\sin\theta_2}{v_2} \quad .$$

This can be rearranged as

$$\frac{dt}{dx} = \frac{\sin\theta_1}{v_1} - \frac{\sin\theta_2}{v_2} \quad .$$

The shortest possible time for light to travel from A to B would occur when the difference between the two "legs" of the path is 0. Setting dt/dx to 0 immediately leads to

$$\frac{\sin\theta_1}{v_1} = \frac{\sin\theta_2}{v_2} \quad ,$$

which can be rearranged as Fermat's law of refraction

$$\frac{\sin\theta_1}{\sin\theta_2} = \frac{\nu_1}{\nu_2} \ .$$

Because Fermat assumes that $\nu_2 < \nu_1$, then θ_2 must be smaller than θ_2 and, in accordance with Snell, a light ray must bend toward the normal when passing into a denser medium.

Fermat published his findings in 1650 but his derivation was not recognized for several years. In fact, Fermat ignited a major scientific controversy because he had not provided a physical explanation for his assumption that the light ray would automatically follow a path of least time. Only a direct measurement of the speed of light in two media with different refractive indices would suffice. This could not be done at the time and indeed was not possible until the middle of the 19th century.

Without this knowledge about the velocities involved, it could not be determined whose derivation was correct and the greatest physicists of the time (or any time) disagreed. Hooke and Newton were among those who believed that light travels faster in denser media. Huygens agreed with Fermat that light travels slower in denser media and used this to derive the law of refraction based on his wave-theory of light. Perhaps the clinching confirmation that Fermat was correct came some years later when his findings could be established by calculus.

Fermat was very critical of Descartes' work. He was convinced that there was little merit in Descartes' proof of the law of refraction and, in fact, he came to regard Descartes as something of a fraud. As a result, Fermat undertook his own derivation of the law of refraction as a point of honor. He embarked on a course of research which eliminated the bouncing ball analogy. Like Descartes, he reached back to Heron of Alexandria, not for bouncing balls, but for the idea of minimum path. However, he altered it by postulating that the path of a light ray connecting two fixed points was the one for which the time of transit, not the length, was a minimum. His own words were: "Je reconnois premièrement la vérité de ce principe, que la nature agit toujours par les voies les plus courtes."

After several years of hard work, Fermat arrived at the law of refraction. Fermat was able to give an expression for the constant n in Snell's law,

$$\frac{\sin\theta_i}{\sin\theta_r} = n \ .$$

Fermat found that this constant was the ratio of the velocity of the incident light to the velocity of the refracted light. Fermat's equation was

$$n = \frac{v_i}{v_r} \ .$$

If the incident light was in air, then we could take $v_i \approx c$. In such a case, Fermat's equation became

$$n \approx \frac{c}{v_r} \ .$$

This equation confirmed that light must travel more slowly in the dense medium than in a sparse medium. Fermat's equation was as astonishing as Maxwell's equation for the velocity of given by

$$c = \frac{1}{\sqrt{\varepsilon\mu}} \ .$$

Here ε and μ were the dielectric and permeability constants, respectively.

Descartes' partisans, however, refused to accept Fermat's proof and argued that his physical assumption regarding least time was faulty. Fermat had little chance against what was then the scientific establishment and he made a less than valiant defense of his discovery, one that would in time be accepted as one of the most basic physical laws. Fermat's last word on this matter was expressed in a letter to one of the principal followers of Descartes:

> I do not pretend nor have I ever pretended to be Nature's private confidant. She has obscure and secret ways that I have never pretended to penetrate. I have only offered her a little geometrical assistance on the subject of refraction if it should be needed. But since you assure me, Monsieur, that her affairs are in order without it and that she is content to follow the course that M. Descartes has prescribed, I surrender to you my supposed physical conquest and content myself with the purely geometrical problem.

One should not judge Fermat too harshly by this "surrender." The theories of Descartes were so entrenched that nearly 100 years passed before European science dropped them for the model of Isaac Newton. Time, though, was on Fermat's side and finally vindicated him, at least scientifically, long after his death.

Let us now give the modern demonstration of Fermat's theorem using calculus. Consider Figure 25. The traveltime between A and I, by way

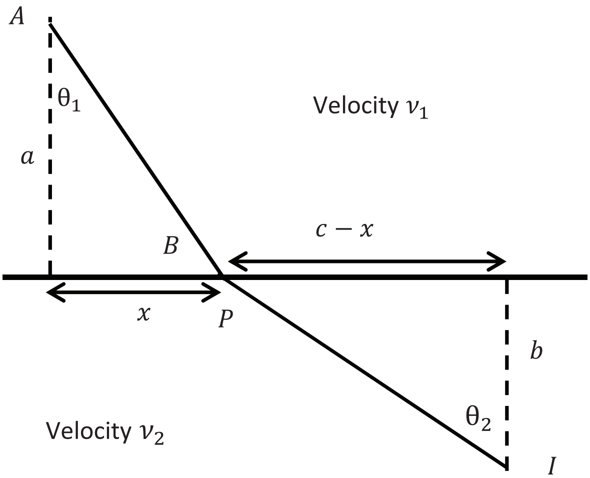


Figure 25. Derivation of Snell's law.

of point P, is

$$t = \frac{(a^2 + x^2)^{1/2}}{v_1} + \frac{(b^2 + (c - x)^2)^{1/2}}{v_2}$$

and, according to the fundamental theory of maxima and minima in differential calculus, the minimum value of t can be obtained by taking the derivative of this equation and setting it equal to zero. The derivative of t with respect to x is

$$\frac{x}{v_1(a^2 + x^2)^{1/2}} + \frac{c - x}{v_2(b^2 + (c - x)^2)^{1/2}} \ .$$

And, if this is made equal to zero, the result is Snell's law, because

$$\frac{x}{(a^2 + x^2)^{1/2}} = \sin\theta_1 \ ,$$

$$\frac{c - x}{(b^2 + (c - x)^2)^{1/2}} = \sin\theta_2 \ .$$

Let us now review Fermat's principle, which may be used to find the raypaths. The principle states that the path of a ray between any two points P_1 and P_2 is such that the time required to traverse the path will be a minimum. This principle is valid for media with varying velocities or for the case in which the ray passes from one medium to a second with a different velocity.

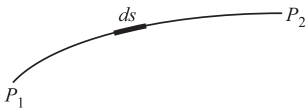


Figure 26. The element ds on a path from one point to another.

Consider an arbitrary path connecting two points P_1 and P_2 as shown in Figure 26. We need to find an expression for the time required to pass along the path from P_1 to P_2. The time needed to traverse a portion of length ds is ds/v, where v is the velocity at the point where ds is situated. The total time is then

$$t = \int_{P_1}^{P_2} \frac{ds}{v} \ .$$

This integral is a line integral. It will have different values for different paths connecting the end points P_1 and P_2. Now suppose we choose a path only slightly displaced from the one shown in Figure 25. The difference of time of traversal for the two paths will, in general, be of the same order of magnitude as that of a quantity determining the displacement of the paths. Fermat's principle states that, if we have the correct path, the difference of time of traversal of it and a neighboring path will be of the order of magnitude of the square or higher powers of the quantity measuring the displacement of the two paths. This is analogous to the statement in ordinary calculus that a function $f(x)$ has a minimum (or maximum) at the point $z = a$. Suppose we wish the value of this function at a point x close to $x = a$. Then by Taylor's theorem we can write

$$f(x) = f(a) + \left(\frac{df}{dx}\right)\bigg|_{x=a} (x-a) + \frac{1}{2}\left(\frac{d^2f}{dx^2}\right)\bigg|_{x=a} (x-a)^2 + \cdots \ .$$

If the function $f(x)$ has a minimum at $x = a$, then $df/dx = 0$ for $x = a$. In such a case, we have

$$f(x) - f(a) = \frac{1}{2}\left(\frac{d^2f}{dx^2}\right)\bigg|_{x=a} (x-a)^2 + \cdots \ .$$

Thus, the difference in the value of the function at x and at a is of the order of magnitude of $(x-a)^2$. On the other hand, if there is no minimum at $x = a$, $f(x) - f(a)$ is of the order of magnitude of $(x-a)$.

Huygens' principle

According to a story, perhaps apocryphal, the great Dutch scientist Christiaan Huygens (1629 – 1695) dropped a stone into the canal next to his house, intently observed the circular wave that moved out across the water surface, and—like Pythagoras and Plato—intuited that a world of perfection was behind the imperfect visible world and this perfect world was constructed of perfect mathematical and geometric formulations. Of course, the waves that Huygens supposedly observed as a boy were never perfectly circular, but his mind held a clear understanding of a perfect circle and it can be argued that, in the spirit of Plato, Huygens spent his life in uncovering the massively important role played by the circle in science.

The Netherlands was a great seafaring nation, and Huygens' first scientific/technological contributions involved improvements to the two most important navigational tools of the 17th century, the telescope and the clock. Huygens, helped by his brother Constantijn, designed a telescope that was far superior to contemporary devices and, in March 1655, he discovered Saturn's moon, Titan. He also was able to explain the curious extension of Saturn, which had intrigued astronomers since Galileo first observed it in 1610: Saturn was encircled by a ring, thin and flat, nowhere touching, inclined to the ecliptic.

In that period, the central problem of navigation was determining longitude. Longitude could, in effect, be measured by time: if the difference in local time at two points was known, the longitudinal distance between them could be computed. However, in the first part of the 17th century, this was not a practical option because the existing mechanical clocks were not sufficiently accurate. Galileo had discovered that a pendulum could be used as a frequency-determining device for a clock (and many consider him the father of this scientific breakthrough), but he never built such a clock. Huygens was the true inventor of the pendulum clock in which the escapement counts the swings and a driving weight provides the push. In effect, the escapement was a feedback regulator that controlled the speed of this type of mechanical clock. Huygens produced his first clock in December 1656, and it was much more accurate than contemporary clocks. Pendulum clocks became the most accurate clocks in the world for the next 300 years.

It is not an overstatement to contend that this is one of the great technological breakthroughs in history. The invention by Huygens of the first accurate clock can be considered the beginning of the modern world, which is based on science and technology, because it permitted much more sophisticated experiments and detailed measurements.

Huygens was able to construct such a clock because of his investigations of the mathematics of the circle, and this would lead him to additional discoveries that are of major importance in modern geophysics. Galileo believed that a pendulum is isochronic; in other words, that the period of a pendulum does not depend on the amplitude of its swing. Huygens, via mathematics, found that a pendulum swinging through the arc of the circle is not isochronic—it only appears isochronic when the length of the arc is quite short relative to the length of the pendulum. This property would give a clock with a long pendulum an advantage over a clock with a short pendulum; however, the pendulums of the early clocks were kept short and light to minimize the amount of energy needed to keep them in motion. As a result, the early pendulum clocks had very wide pendulum swings, which decreased their accuracy.

Huygens had to solve the so-called *tautochrone* problem in order to construct a perfect clock. More specifically, Huygens had to find the curve down which a mass will slide under the influence of gravity in the same amount of time, regardless of its starting point. Huygens used geometrical methods because calculus had not yet been invented. He showed that the required curve was a cycloid, instead of the circular arc of a pendulum's swing, and therefore that conventional pendulums are not isochronous.

Our analysis will make use of calculus. A *simple harmonic oscillator* is an oscillator that is neither driven nor damped. It consists of a mass m, which experiences a single force F, which pulls the mass in the direction of the point $x = 0$ and depends only on the mass's position x and a positive constant k. When displaced from its equilibrium position, the restoring force is

$$F = -kx \ .$$

Newton's second law $F = ma$ becomes

$$-kx = m\frac{d^2x}{dx^2} \ .$$

Solving this differential equation, we find that the motion is described by the sinusoidal function

$$x(t) = A\cos(\omega t + \varphi) \ ,$$

where the angular frequency ω is given by

$$\omega = \sqrt{k/m} \ .$$

Such sinusoidal motion is known as *simple harmonic motion*. It consists of sinusoidal oscillations about the equilibrium point, with constant amplitude A and constant frequency ω. The motion is periodic, repeating itself in a sinusoidal fashion. Its period is $T = 2\pi/\omega$, which is the time for a single oscillation. Its cyclic frequency is $f = 1/T$, which is the number of cycles per unit time. The position at a given time t also depends on the phase φ, which determines the starting point on the sinusoidal wave. The period and frequency are determined by the size of the mass m and the force constant k, while the amplitude and phase are determined by the starting position and velocity.

A simple pendulum exhibits approximately simple harmonic motion under the conditions of no damping and small amplitude. Assuming no damping, the differential equation governing a simple pendulum is

$$\frac{d^2\theta}{dt^2} + \frac{g}{\ell}\sin\theta = 0 \ ,$$

where g is acceleration due to gravity, ℓ is the length of the pendulum, and θ is the angular displacement. If the maximum displacement of the pendulum is small, we can use the approximation $\sin\theta \approx \theta$. The differential equation approximately reduces to

$$\frac{d^2\theta}{dt^2} + \frac{g}{\ell}\theta = 0 \ .$$

The solution to this equation is given by

$$\theta(t) = \theta_0 \cos\left(\sqrt{\frac{g}{\ell}}t + \varphi\right) \ ,$$

where θ_0 is the largest angle attained by the pendulum. The period, the time for one complete oscillation, is given by the expression

$$T = 2\pi\sqrt{\frac{g}{\ell}} \ ,$$

which is a good approximation of the actual period only when θ_0 is small. Huygens seeks to build a clock that will tell the correct time with no such limitation as the smallness of θ_0.

Isochronous (from the Greek "equal" and "time") pertains to processes that require timing coordination to be successful. A sequence of events is isochronous if the events occur regularly, or at equal time intervals.

Isochrony is important in clocks. A property of simple harmonic motion is that the period T does not depend on the amplitude. Thus, simple harmonic motion is *isochronous*. In other words, it takes the same time to make the same sinusoidal oscillation regardless of the amplitude of the oscillation. The theory of Fourier series and integrals rests upon this property.

For example, see Figure 27. A mechanical clock is *isochronous* if it runs at the same rate regardless of changes in its drive force, so that it maintains the same periodic time T as its mainspring unwinds. As we have just seen, a conventional pendulum is not isochronous. However, if the maximum angle θ_0 of swing is small, then a conventional pendulum is nearly isochronous. Huygens reasons that this limitation is because the conventional pendulum swings in the arc of a circle. As a result, he designs a clock whose pendulum swings in the arc of a rolling circle. A rolling circle (a.k.a. cycloid) is the curve that is generated by a point on the circumference of a circle as it rolls along a straight line.

The cycloidal clock of Huygens provides the correct time, regardless of the size of the largest angle attained by its pendulum. In other words, the cycloidal pendulum is a simple harmonic oscillator, unlike the usual pendulum. In *Horologium Oscillatorium* (1673), Huygens provides a complete mathematical description of his cycloidal clock. Its pendulum is forced to swing in an arc of a cycloid. Huygens accomplishes this by suspending the pendulum (composed of a bob on a wire string) at the cusp of the evolute of the cycloid. Such a cycloidal pendulum is isochronous, regardless of amplitude. The cycloidal clock is extremely accurate, but unfortunately the movement causes an excessive amount of friction.

Meanwhile, Robert Hooke, another person with a prominent role in establishing the fundamentals of geophysics, invented the anchor escapement for a pendulum clock. The anchor escapement required a smaller angle of swing than the angle required by the escapements of the early pendulum clocks. Pendulum clocks became so accurate that the cycloidal clock quickly became passé. However, Huygens made one more great

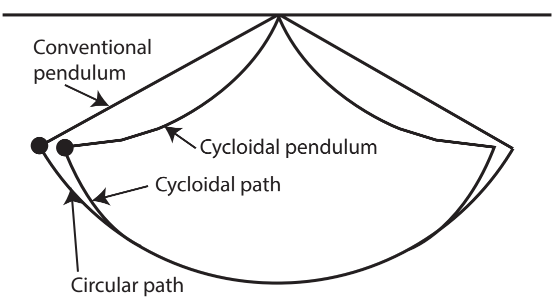


Figure 27. The cycloidal pendulum is a simple harmonic oscillator, whereas the conventional pendulum is not.

contribution to the measurement of time when, in 1675, he built a chronometer that used a balance wheel and a spiral spring instead of a pendulum. Balance wheels and spiral springs were the basis for almost all watches until the invention of the quartz crystal oscillator in the 20th century.

Although the cycloidal clock was of practical use for only a limited time, it is historically important because it can be viewed as the first successful design of an intricate apparatus based on higher mathematics. Heron of Alexandria and Leonardo da Vinci use mechanical principles in order to design their inventions, but the mathematics involved essentially is simply Euclidean geometry which dates from about 300 BC. The introduction of higher mathematics to accomplish mechanical design gives Huygens the distinction of being the father of modern technology.

In *Traité de la Lumière* (1690), Huygens presents one of the great contributions to theoretical physics when he postulates that light is a wave. His postulate is supported with a principle (or construction) known ever since as Huygens' principle. The two key ideas of Huygens' principle are secondary waves and the enveloping mechanism. Huygens demonstrates his construction with diagrams such as those displayed in Figure 28. The left diagram shows how a spherical wavefront propagates. The right diagram shows how a plane wavefront propagates. In other words, in the spirit of

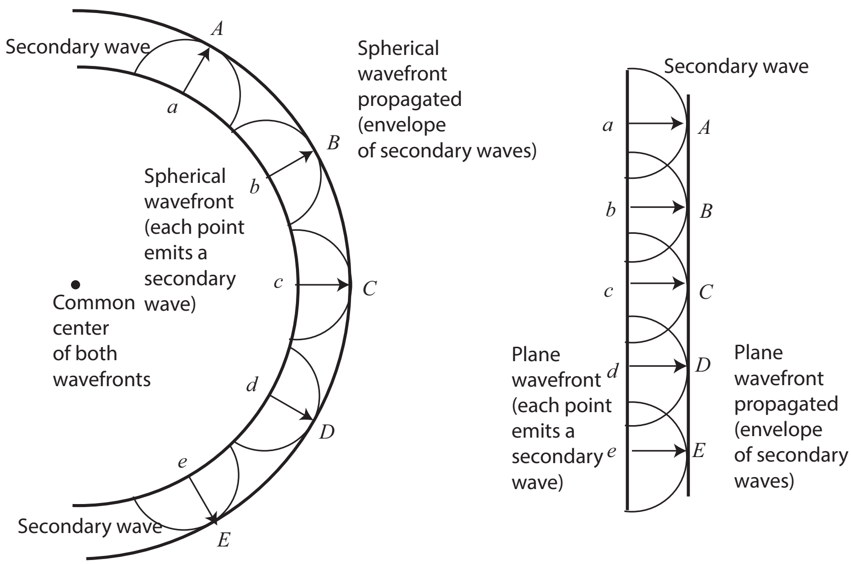


Figure 28. An envelope of secondary waves produces a new wavefront.

Plato, Huygens uses the perfect world of circles to explain the intricacies of wave motion.

In 1921, at Belle Isle, Oklahoma, USA, J. C. Karcher is the first person to record a seismic reflection line. Figure 29 shows Karcher's 1921 migration diagram of the Viola interface as given by an envelope of circular arcs. Each time a geophysicist records a prestack depth migration, they are using a method based upon the fundamental concept known as Huygens' principle.

Huygens' principle may be summarized as follows. Given a wavefront at a given instant of time, each point on the wavefront emits a spherical secondary wave (see Figure 30). The secondary waves are in phase with the original wavefront and propagate outward with the same speed. The secondary waves constructively interfere and their envelope forms a new wavefront. In the same manner, the envelope of secondary waves from this new wavefront produces the next wavefront.

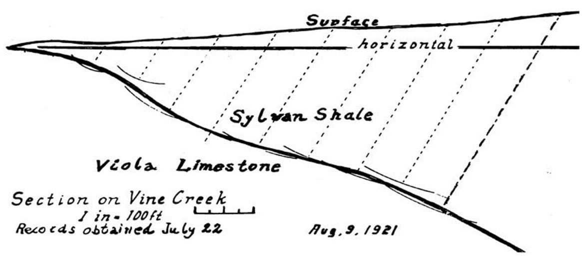


Figure 29. Viola interface as given by an envelope of circular arcs.

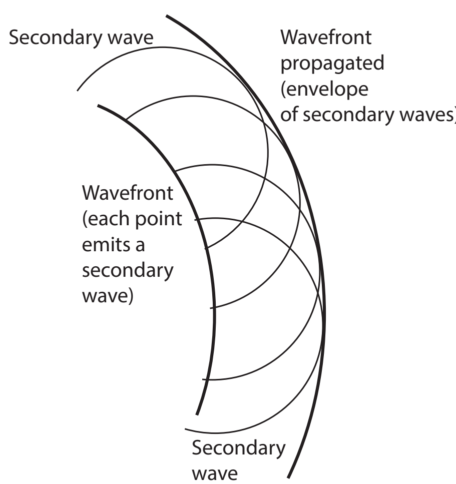


Figure 30. Envelope of secondary waves forms a new wavefront.

Huygens imagines that this process repeats itself as the wave propagates. If the medium is homogeneous and isotropic, the spherical secondary waves may be constructed with finite radii. On the other hand, if the medium is inhomogeneous, the secondary waves will have infinitesimal radii, and the magnitudes of the radii will depend on the wave velocity of the medium at the respective centers of the secondary waves. Calculus is needed to deal effectively with these infinitesimals. In inhomogeneous media, the secondary waves will not be circular anymore, so to consider them as circular one needs infinitesimal calculus.

A plane wave is a wave for which quantities vary only with time and with the distance along a fixed direction. The direction vector is denoted by the unit vector $\boldsymbol{u}$ where $|\boldsymbol{u}| = 1$. The plane wave propagates in the direction of $\boldsymbol{u}$ with velocity v. Mathematically a plane wave must be of infinite extent in order to propagate as a plane wave. However, many waves are approximately plane waves in a localized region of space. A plane wave of constant frequency has wavefronts (surfaces of constant phase) that are parallel planes of constant peak-to-peak amplitude normal to the direction vector $\boldsymbol{u}$.

Geometrical seismology is based upon the concept of a ray. The constant-phase surfaces can be described by constructing the normals to these surfaces (the wave normals or rays), and one can follow the motion of these surfaces by moving along the directions of these rays. This mode of description evidently is possible for waves other than plane waves. For example, in the case of spherical waves in a homogeneous medium, the rays consist of straight lines radiating in all directions from a common point, and the surfaces of constant phase are concentric spherical surfaces. If the direction vector is directed outward from the center, one speaks of a diverging wave. If the direction vector is directed toward the center, one speaks of a converging wave. In order to follow the motion of a limited portion of a wave surface, we can construct a bundle (or pencil) of rays through this portion of the surface. Such a pencil of rays is called a beam, and for plane waves the beam consists of a parallel bundle of rays. For a spherical portion of a wave surface, the rays diverge from (or converge on) a point F, known as the focal point of the pencil. Such a pencil is called *stigmatic* (see Figure 31). The portion of a wave surface which has different radii of curvature in two mutually orthogonal directions form an *astigmatic* pencil and do not

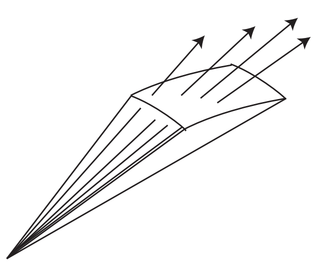


Figure 31. A divergent pencil of rays.

pass through a common point. In the case of the propagation of waves in a homogeneous medium, the rays always are straight lines, and the wave surfaces do not change shape as the wave propagates. If, however, the velocity varies from point to point, the rays will be curved lines, and the wave surfaces will not maintain an unaltered shape.

There is an advantage gained by describing wave motion in terms of rays. Under what condition is it advantageous to use rays instead of waves? Let us consider what happens when we try to form a narrow beam from a plane wave by allowing the plane wave to fall on a screen, in which there is a hole, placed perpendicular to the direction of propagation. The waves emerging from the hole, in general, will not form a section of a plane wave with a parallel bundle of rays, but will spread out generally in all directions. If the wavelengths of the waves are very small compared to the linear dimensions of the aperture, however, this spreading effect (a.k.a. diffraction) becomes extremely small. Such is the case of light waves, whose wavelengths are very small compared with the dimensions of ordinary objects. In such cases, we can neglect diffraction effects and treat the bundle of light rays emerging from the aperture as strictly parallel. Similarly, an obstacle placed in the path of a beam of light casts a sharply defined geometrical shadow. The laws of optics and optical systems, to the approximation in which one can neglect typical wave effects such as diffraction and interference, comprise the subject of geometrical optics.

Geometrical optics, or ray optics, describes light in terms of rays. Of course, a ray is an abstraction that approximates the path along which light propagates. Light rays propagate in straight paths as they travel in a homogeneous medium. A ray will split into a reflected ray and a refracted ray at the interface between two layers. A ray follows curved paths in a medium in which the velocity changes. The province of geometrical optics does not account for diffraction or interference. Rays are useful when the wavelength is small compared to the size of structures with which the light interacts. Unfortunately the wavelength of a typical seismic wave is never small compared to the size of geologic structures of interest. Despite this limitation, the use of the geometrical theory of seismic rays is useful and productive.

A number of phenomena cannot be described by geometrical seismology, but can be described by wave theory. The principle of superposition is fundamental to the application of wave theory. This principle is based upon the fact that the various wave trains which, in their totality, compose a beam may be considered as being mutually independent so we may treat the elementary waves as if each exists alone. As a result, the behavior of the beam as a whole may be computed as the sum of the effects of the elementary waves. This is characteristic of all wave motion governed by the wave equation, which is a linear equation.

Huygens' principle makes it possible to follow the propagation of waves. Huygens' principle may be described in this manner. Suppose that we know the shape of one of the constant-phase surfaces of a wave, e.g., one of the crests of the wave, at some instant of time. We can find the shape of this wave surface at a later time Δt by considering each point on the original wave surface as a source of secondary spherical waves which diverge from these points. Construct spheres at each point of the original wavefront. Make the radius of each of these spheres equal to $v\Delta t$, where v is the velocity of the wave. Then the envelope of these spherical secondary waves yields the shape of this wavefront at a time Δt later.

Huygens uses his principle in order to find the direction of reflected rays. As illustrated in Figure 32, line *AB* is the incident ray and *BF* is the reflected ray. The rays in a plane wave are parallel. In particular, ray *MC* is parallel to ray *AB*. Ray *CQ* is parallel to ray *BF*. The angle of incidence is $\theta_1 = \angle NBA$. The angle of reflection is $\theta_k = \angle NBF$. The velocity in the medium is v_1.

A wavefront is perpendicular to the rays. Lines *AM* and *BP* are wavefronts of the incident plane wave. Ray *AB* strikes the interface sooner than ray MC. More specifically, ray *MC* must travel an extra time t, in which it travels the extra distance $PC = v_1 t$. As soon as ray *AB* hits the interface at *B*, it initiates a Huygens circular wavelet with the center at *B*. At time t, this emergent wavelet has radius $v_1 t$. As soon as ray *MC* hits the interface at *C*, it initiates a Huygens' circular wavelet with the center at *C*.

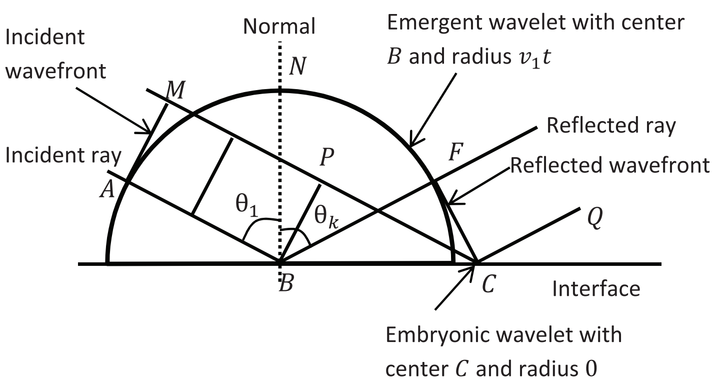


Figure 32. Reflected wavefront *CF* is an envelope of the emergent wavelet and the embryonic wavelet.

At time t, this wavelet is embryonic for it has radius 0. In other words, this embryonic wavelet is simply point C. The envelope of the emergent wavelet and the embryonic wavelet is the reflected wavefront CF, where CF is the tangent from point C to the emergent wavelet.

Triangle $\triangle CBP$ and triangle $\triangle BCF$ are each right triangles. They have a common hypotenuse BC. Recall that the extra distance is $PC = v_1 t$ and that the emergent wavelet has radius $BF = v_1 t$. Thus, $PC = BF$. Thus, right triangle $\triangle CPB$ is congruent to right triangle $\triangle BCF$. From plane geometry,

$$\measuredangle CBP = \theta_1 \quad \text{and} \quad \measuredangle BCF = \theta_k \ .$$

Thus, the angle θ_1 of incidence is equal to the angle θ_k of refection. This is Heron's law of reflection.

Moving to Figure 33, let an incident wave pass from the upper medium (incident medium) with velocity v_1 to the lower medium (refracting medium) with different velocity v_2. In our case, we assume that

$$v_2 < v_1 \ .$$

Line AB is the incident ray. Let θ_i be the angle of incidence. Huygens uses his principle in order to find the direction of the refracted wave. It takes the same time t to go from P to C as to go from B to G. Thus, $PC = v_1 t$ and $BG = v_2 t$. A wavefront is perpendicular to the rays. Lines AM and BP are wavefronts of the incident wave. Ray AB strikes the interface

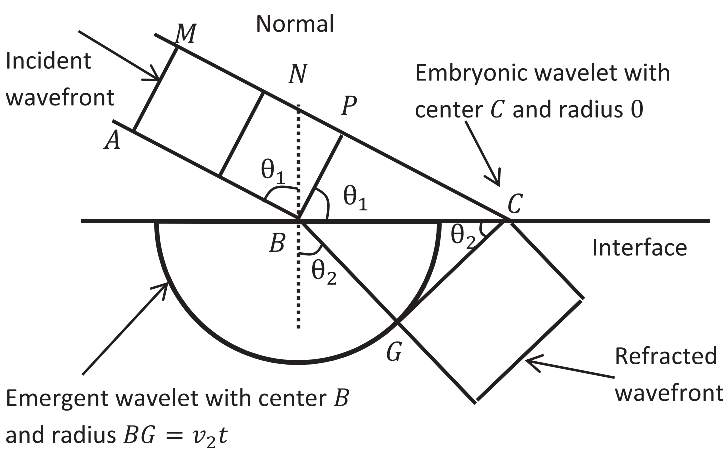


Figure 33. Refracted wavefront GF is an envelope of the emergent wavelet and the embryonic wavelet.

sooner than ray MC. More specifically, ray MC must travel the extra time t, during which it travels the extra distance $PC = v_1 t$.

As soon as ray AB hits the interface at B, it initiates a Huygens' circular wavelet (in the lower medium) with the center at B. At time t, this emergent wavelet has radius $v_2 t$. As soon as ray MC hits the interface at C, it initiates a Huygens' circular wavelet (in the lower medium) with the center at C. At time t, this wavelet is embryonic for it has radius 0. In other words, this embryonic wavelet is simply point C. The envelope of the emergent wavelet and the embryonic wavelet is the refracted wavefront CG, where CG is the tangent from point C to the emergent wavelet. Line BG is the refracted ray and θ_2 is the angle of refraction.

Triangle $\triangle CBP$ and triangle $\triangle BCG$ are each right triangles. They have a common hypotenuse BC. Recall that the extra distance is $PC = v_1 t$ and that the emergent wavelet has radius $BG = v_2 t$. Thus,

$$t = \frac{PC}{v_1} = \frac{BG}{v_2} \ .$$

From plane geometry,

$$\measuredangle CBP = \theta_1 \quad \text{and} \quad \measuredangle BCG = \theta_2 \ .$$

Thus,

$$PC = BC \sin\theta_1 \quad \text{and} \quad BG = BC \sin\theta_2 \ .$$

Substituting for BC and BG, we have

$$v_1 t = BC \sin\theta_1 \quad \text{and} \quad v_2 t = BC \sin\theta_2 \ .$$

If we divide these two equations, we obtain

$$\frac{v_1}{v_2} = \frac{\sin\theta_1}{\sin\theta_2} \ .$$

This is Snell's law of refraction.

In Figure 34, we see that Snell's law reduces to two right triangles with a common hypotenuse R.

The formulation of Huygens' principle with spherical wavelets has an obvious defect, because it produces a wavefront traveling backward as well as one traveling forward. Figure 35 depicts a spherical wavefront AA emitted by a source S, the forward traveling wavefront BB at later time Δt, and also the backward traveling wavefront CC also at later time Δt, as given by Huygens' construction. The wavefront traveling backward does not exist.

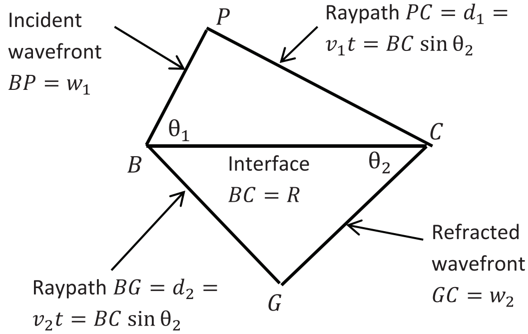


Figure 34. Huygens' essence of Snell's law.

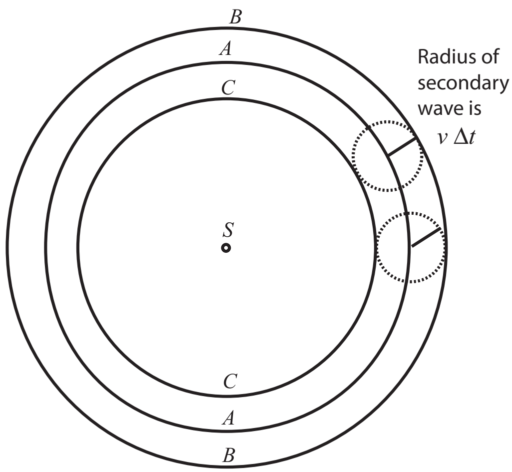


Figure 35. Original wavefront *AA*, backward traveling wavefront *CC*, and forward traveling wavefront *BB*.

To complete the analysis, we must know how the amplitudes of these secondary waves vary with direction. It turns out that this variation of amplitude with direction of propagation is complicated. This problem is solved by Fresnel and Kirchhoff. Their more advanced formulation of Huygens' principle states that we can obtain the wavefield at any point by first considering each point on any closed surface (it may be taken as a wavefront for convenience) as a source of secondary waves and then superposing the effects of these secondary waves at the point in question. However, these secondary waves have different amplitudes in different directions. If we take dependence of amplitude and phase into consideration, then the wavefront traveling backward turns out to have zero amplitude. We will not give a rigorous formulation of the Fresnel–Kirchhoff formulation of Huygens' principle here, but instead we will give an approximate treatment.

The principle of superposition of waves states that, when two or more waves are incident on the same point, the total displacement at that point is equal to the vector sum of the displacements of the individual waves. If a crest of a wave meets a crest of another wave of the same frequency at the same point, then the magnitude of the displacement is the sum of the individual magnitudes; this is known as constructive interference. If a crest of one wave meets a trough of another wave, then the magnitude of the displacements is equal to the difference in the individual magnitudes; this is known as destructive interference.

Interference and diffraction are terms that describe a wave interacting with something that changes its amplitude, such as another wave. There is no important physical difference between interference and diffraction. It is a question of usage. When there are only a few sources, e.g., two, we call it interference, but with a large number of sources, we call it diffraction. Interference is a phenomenon in which two waves superimpose to form a resultant wave of greater or lower amplitude. If one causes two (or more) beams from two separate portions of the wavefront to recombine, the resulting variations of intensity with position are termed interference effects.

Diffraction refers to various phenomena that occur when a wave encounters an obstacle. In classical physics, the diffraction phenomenon is described as the apparent bending of waves around small obstacles and the spreading out of waves past small openings. Diffraction takes into consideration the interference effects caused by a limitation of the cross section of a wavefront. The extended form of Huygens' construction allows diffractions. Such diffractions appear in directions other than that of the incident wave. The diffractions occur in such directions because the mutual cancellation by destructive interference of the secondary waves is not complete.

When the cross section of a beam is limited by allowing the wave to pass through an opaque screen containing one or more apertures, the distribution of intensity in the transmitted beam as observed on another screen is called a diffraction pattern. If the diffracting screen (or obstacle) is placed between source and observing screen, and no lenses or mirrors are employed, the resulting phenomenon is called Fresnel diffraction. In general, both source and observing screen are at finite distances from the diffracting screen. On the other hand, if the incident wave is a plane wave, and if the diffracted waves are observed on a distant screen, then the resulting pattern is known as a Fraunhofer diffraction pattern. Fundamentally, both types of diffraction are only different aspects of the same basic phenomenon and are explicable in terms of Huygens' principle.

Let us examine Huygens' principle more closely. We will use Figure 36 twice. In the figure, some sort of obstacle (as indicated by the opaque screen) is inserted between O and P. In the first instance, we will suppose that the

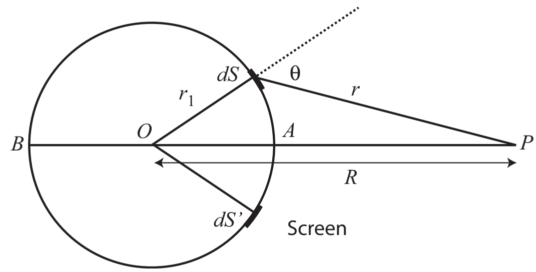


Figure 36. A wavefront encountering a screen and the absence of the screen.

screen is absent. In the second instance, we will suppose that the screen is present.

First instance (the screen is absent). Consider a spherical wave diverging from a source O, and suppose we wish to compute the amplitude of this wave at a point P, which lies at a distance R from the source O. First, we construct a spherical surface (wavefront) of radius $r_1 < R$ with its center at O. On this wavefront, we must consider each element of area dS as the source of secondary waves which in their totality combine at P to produce the resultant wave motion at P. The relative phases of the secondary waves arriving at P may be obtained by observing that it takes a time r/v for a disturbance at dS to reach P. Thus, the relative phases of these waves are given by $2\pi r/\lambda$. What are the relative amplitudes of these waves? The relative amplitudes are proportional to dS and the area of the elementary source on the wavefront, and they are inversely proportional to r. In addition, the relative amplitudes depend on the angle θ in the form $(1 + \cos\theta)/2$, where θ is the angle between the normal to the spherical surface and r, so that $\cos\theta$ varies from $+1$ at A to -1 at B. This so-called obliquity factor $(1 + \cos\theta)/2$ varies from $+1$ at A to 0 at B. Thus, the obliquity factor eliminates the wavefront traveling backward in the elementary Huygens' construction. Let us look at the wavefront traveling forward. Without the screen, we have an unobstructed wave. In such a case, the secondary waves from all elements of area dS mutually destroy each other by interference at P, except for those originating in a very small region around the point A; hence, the effect is almost the same as if the light traveled in a straight line from O to P.

Second instance (the screen is present). The screen blocks the lower hemisphere of the wavefront. Let dS' on the lower hemisphere be the counterpart of dS on the upper hemisphere. Now the mutual cancellation of the secondary waves from dS and dS' cannot occur, because the screen prevents the

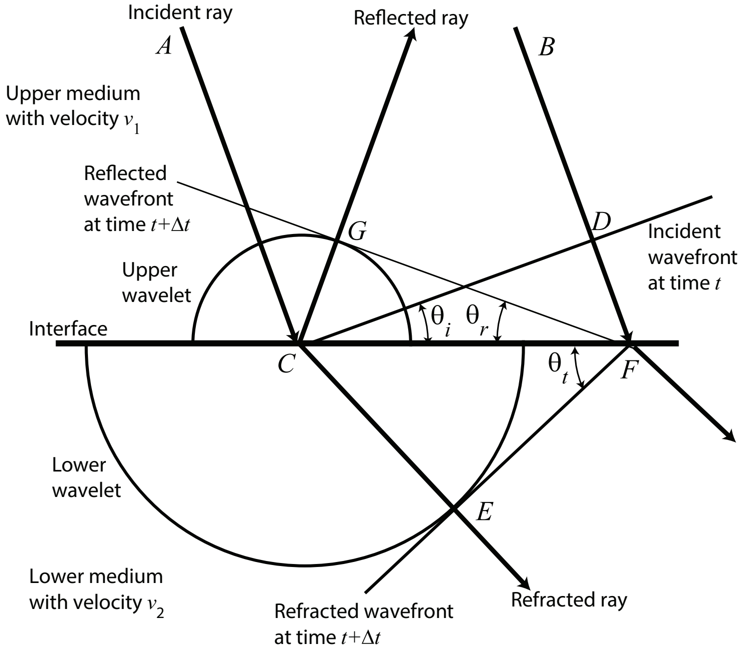


Figure 37. Downgoing plane wavefront *CD*, upgoing reflected wavefront *GF*, and downgoing refracted wavefront *EF*.

waves from dS' from reaching P. This gives rise, then, to diffractions, and one may obtain even larger intensities at P than without the screen.

Huygens uses his wave theory to finally establish the laws of refection and refraction which, particularly the latter, had been sought since ancient times. In Figure 37, a plane interface separates the upper medium from the lower medium. Assume the wave velocity in the lower medium is greater than in the upper medium, i.e., $v_2 > v_1$. A downgoing plane wavefront *CD* in the upper medium is slantwise incident on the interface. As each point on the wavefront arrives at the interface, it behaves according to Huygens' principle and emits two secondary waves, one upward and the other downward. There are two envelopes—the envelope that produces the upgoing reflected wavefront *GF* and the envelope that produces the downgoing refracted wavefront *EF*.

The law of reflection and refraction can be derived by analyzing the part of the incident wavefront that lies between rays *AC* and *BDF* as point *C* contacts the interface *CF*. Let Δt represent the time increment needed for the wave to travel from *D* to *F*, which means that $DF = v_1 \Delta t$. The two secondary waves emanating from point C are different because the secondary wave above the interface is a semicircle with radius $v_1\Delta t$ and the secondary wave below the interface is a semicircle with radius $v_2\Delta t$. The envelopes at time $t + \Delta t$ are given by the tangent lines FG and FE.

The angle θ_i between incident wavefront CD and interface CF is the angle of incidence. The angle θ_r between reflected wavefront FG and interface CF is the angle of reflection. The angle θ_t between refracted wavefront and interface CF is the angle of refraction. These angles are part of three right triangles (CFD, CFG, and CFE) which have the common hypotenuse CF. Thus, the sines of these three angles have a common denominator, that is

$$\sin\theta_i = \frac{DF}{CF}, \quad \sin\theta_r = \frac{CG}{CF}, \quad \sin\theta_t = \frac{CE}{CF} \ .$$

Also, $CF = v_1\Delta t$ and $= v_2\Delta t$, which means that the above equations can be written

$$\sin\theta_i = \frac{v_1\Delta t}{CF}, \quad \sin\theta_r = \frac{v_1\Delta t}{CF}, \quad \sin\theta_t = \frac{v_2\Delta t}{CF} \ .$$

Because $\sin\theta_i = \sin\theta_r$, it follows that the incidence angle θ_i is equal to the reflection angle θ_r. It also follows that

$$\frac{\sin\theta_i}{\sin\theta_t} = \frac{\left(\dfrac{v_1\Delta t}{CF}\right)}{\left(\dfrac{v_2\Delta t}{CF}\right)} = \frac{v_1}{v_2} \ ,$$

which is Snell's law—one of the mathematical underpinnings of seismology and a marvelous proof of a relationship that had been sought for centuries. Snell's law also can be derived by using Fermat's principle.

Thus, the wavefront method of Huygens correctly, and with a dramatic simplicity and elegance, generates the laws of reflection and refraction. The work of Huygens on the telescope and the clock would secure his place in the annals of exploration geophysics. Huygens also is known for studying anisotropy (Icelandic spar), caustics, and polarization. His principle of constructing wave motion by the use of secondary waves appears in all textbooks on physics. Among all of his other accomplishments, Huygens is one of the great pioneers of exploration geophysics.

Isaac Newton and the birth of geophysics

Because Newton's achievement can be considered the starting point of modern geophysics and planetary science, this section takes a slight detour to discuss some of its impact on what geoscientists do today.

Isaac Newton (1643 – 1727) composed *Principia* during 1685 and 1686. Its first edition was published on 5 July 1687. In the book, Newton established the mathematical form of the theory of gravity. The practical application of Newton's law of gravity arrived pretty quickly (very quickly, considering the long time that it took for "technology transfer" in that era). In 1735 – 1745, the French Academy sent expeditions to Lapland and Peru on which Pierre Bouguer established gravitational relationships such as the variation of gravity with elevation, the horizontal attraction caused by mountains, and the density of the earth.

Today, gravity surveying typically is accomplished with a gravimeter, an instrument consisting of a weight attached to a spring that stretches or contracts according to an increase or decrease in gravity. It is designed to measure differences in gravity acceleration rather than absolute magnitudes. Gravimeters used in geophysical surveys are capable of detecting differences in the earth's gravitational field to one part in one hundred million. We will now present Newton's geometrical solution to the inverse square law of gravitation as given in *Principia*.

Part A. Properties of conic sections

Circles, ellipses, parabolas, and hyperbolas are called *conic sections*. All of the conic sections can be characterized as follows. A conic is either a circle or the locus of a point which moves so that the ratio of its absolute distance from a given point (a focus) to its absolute distance from a given line (a directrix) is a positive constant e (the eccentricity). We can use this characterization to find the polar equation of a conic of eccentricity e with one focus at the origin O and with the line $x = k$ as the directrix (where $k > 0$). As shown in the notation of Figure 38, the focus-directrix property is $PO = e\,PD$. Because the origin and focus are the same, we have $PO = r$. Also $\cos(\pi - \theta) = -\cos\theta$. Thus,

$$PD = EC = EO + OC = -r\cos\theta + k \ ,$$

and the focus-directrix property becomes

$$r = e(-r\cos\theta + k) \ ,$$

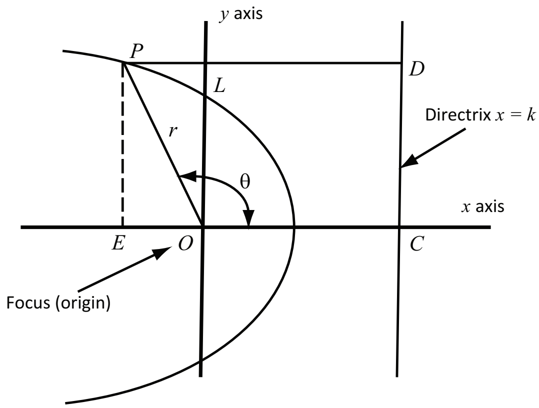


Figure 38. Illustration of focus and directrix.

which yields the curve

$$r(\theta) = \frac{ke}{1 + e\cos\theta} \quad (e > 0, k > 0) \ , \tag{20}$$

which also can be written as

$$\frac{1}{r(\theta)} = \frac{1}{ke} + \frac{1}{k}\cos\theta \ . \tag{21}$$

Here $r(\theta)$ indicates the function with value r at θ. Let us define the *semi-latus-rectum* ℓ as

$$\ell = r\left(\frac{\pi}{2}\right) = ke \ . \tag{22}$$

The focus is at origin O. The latus-rectum 2ℓ is the breadth of the curve at the focus. An apsis is an extreme point in an object's orbit. In other words, an apsis is a point of either least distance or greatest distance on the curve from the focus O. For elliptic orbits about a larger body, there are two apsides, named with the prefixes peri- (meaning "near") and ap- or apo- (meaning "away from"). For example, the Sun is at the focus of the elliptical path of a planet orbiting it. The point of least distance in the elliptical path is the perihelion, and the point of greatest distance is the aphelion. In equation 22, the minimum radius occurs when $\theta = 0$, i.e., $r(0)$ is the periapsis radius. The maximum radius occurs when $\theta = \pi$, i.e., $r(\pi)$ is the apoapsis radius. The average of the periapsis and apoapsis radii is

$$a = \frac{r(0) + r(\pi)}{2} = \frac{ke}{2(1+e)} + \frac{ke}{2(1-e)} = \frac{ke}{1-e^2} \ . \tag{23}$$

The chord passing through the foci is called the major axis of the conic, and the length of this chord is $2|a|$. Thus, we call a the semi-major-axis. From equations 22 and 23, we have

$$\ell = a(1 - e^2) \ . \tag{24}$$

We now define c as

$$c = a - r(0) = a - \frac{ke}{1+e} = a - \frac{a(1-e^2)}{1+e} = ae \ , \tag{25}$$

which yields the eccentricity as

$$e = \frac{c}{a} \ . \tag{26}$$

The distance between the foci is $2|c|$. The parameters ℓ, a, and c represent geometrical dimensions common to all conic sections.

When $e = 1$, the curve is a parabola. Because the parabola is the borderline between open and closed curves, the second focus is at infinity, so $a = \infty$, $c = \infty$, and equations 24 and 26 do not hold. Equations 24 and 26, however, do hold for any conic except a parabola. The special case $e = 0$ represents a circle; then the foci are at the same point, so $\ell = a$ and $c = 0$. When $0 < e < 1$, the curve is an ellipse, and $a > 0$, $c > 0$. When $1 < e < \infty$, the curve is a hyperbola, and $a < 0$, $c < 0$. For all conics, the semi-latus-rectum ℓ is positive.

Part B. Apollonius equation for the ellipse

At this juncture, we would like to say a few words about the *ellipse* and consider Figure 39. An ellipse has two perpendicular axes about which the ellipse is symmetric. Due to this symmetry, these axes intersect at the center of the ellipse (denoted by C). The larger of these two axes is called the major axis. The smaller of these two axes is called the minor axis. The semi-major axis (a.k.a. major radius) is denoted by a. The semi-minor axis (a.k.a. minor radius) is denoted by b. The chord of an ellipse that is perpendicular to the major axis and passes through either one of its two foci is the latus rectum of the ellipse. The length of each latus rectum is $2b^2/a$. Accordingly, the semi-latus-rectum is $\ell = b^2/a$.

The ancient Greek geometer and astronomer Apollonius of Perga is credited with developing the theory of conic sections. An ellipse is a compressed circle. A circle of radius a is pushed down so that it becomes an ellipse with major radius a and minor radius b, where $b < a$. The

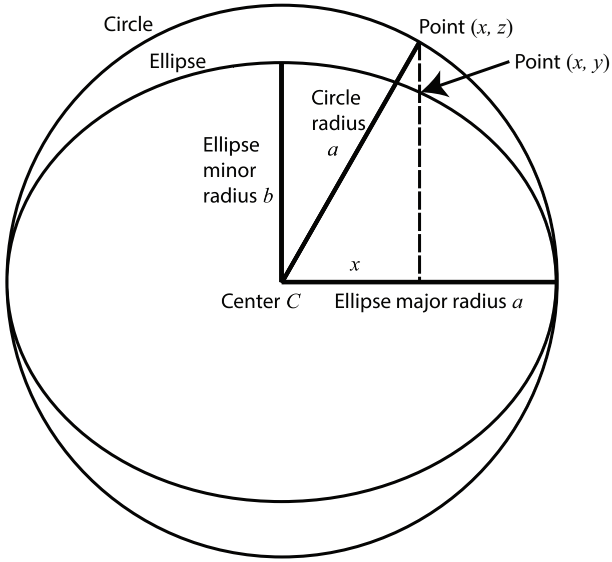


Figure 39. Ellipse as a circle that is compressed.

coordinates of the circle are denoted by (x, z). Thus,

$$z = \sqrt{a^2 - x^2} \ .$$

The coordinates of the ellipse are denoted by (x, y). Apollonius finds the equation for y in this way. First, Apollonius writes the proportion

$$z{:}a \ \text{ as } \ y{:}b \ .$$

This proportionality can be written as the equation

$$\frac{z}{a} = \frac{y}{b} \ .$$

Next, Apollonius takes the square of each side of the equation. The result is the *Apollonius equation for the ellipse*,

$$\frac{a^2 - x^2}{a^2} = \frac{y^2}{b^2} \ ,$$

which becomes the familiar equation for an ellipse, namely,

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \ .$$

Part C. Bounding parallelogram and conjugate diameters

A diameter is a chord (such as GP) that passes through the center C of the ellipse (see Figure 40). For an ellipse, two diameters (such as GP and KD) are conjugate if and only if the tangent line to the ellipse at an endpoint of one diameter is parallel to the other diameter. Each pair of conjugate diameters of an ellipse has a corresponding tangent parallelogram, sometimes called a bounding parallelogram. A bounding parallelogram is skewed compared to a bounding rectangle. It is well known that all bounding parallelograms for a given ellipse have the same area as the bounding rectangle.

In particular, it follows that the rectangle with sides a and b has the same area as the parallelogram with sides a_1 and b_1 (see Figure 41). The area of the rectangle is ab, whereas the area of the parallelogram is $b_1\,a_0$, where a_0 is the normal to the line DK from point P. Thus, we have

$$ab = b_1\,a_0 \ ,$$

which yields

$$\frac{a}{a_0} = \frac{b_1}{b} \ .$$

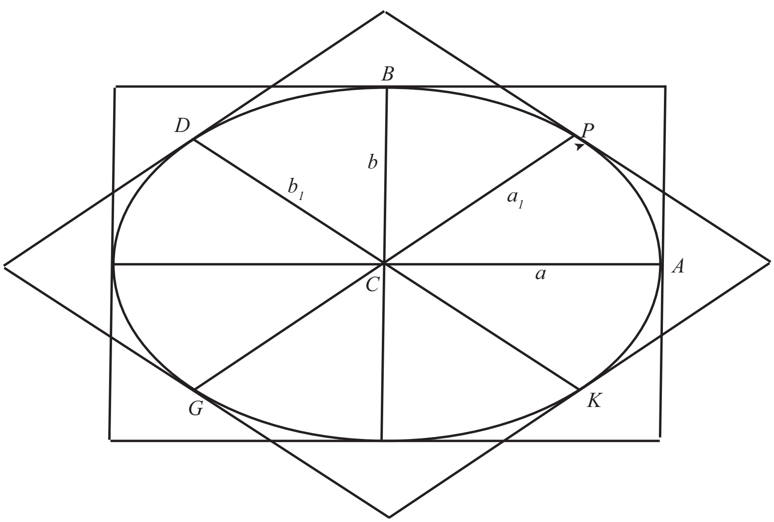


Figure 40. The area of the bounding rectangle (with perpendicular sides $2a$ and $2b$) is equal to area of any bounding parallelogram (such as the parallelogram with oblique sides $2a_1$ and $2b_1$).

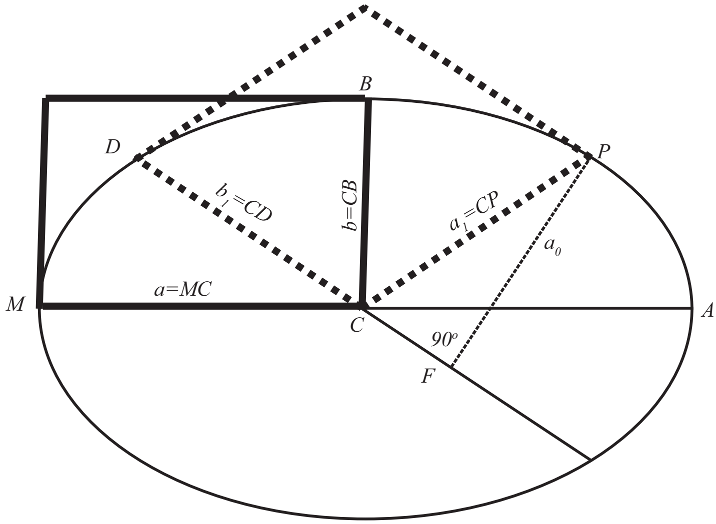


Figure 41. The area of the rectangle (bold solid lines) equals the area of the parallelogram (bold dashed lines).

Part D. Apollonius equation in the oblique coordinate system

Considering Figure 42, an ellipse with major radius a and minor radius b has conjugate radii a_1 and b_1. An *alias transformation* is a transformation in which the coordinate system is changed, leaving points in the original coordinate system "fixed" while changing their representation in the new coordinate system. A rectangular coordinate system (a.k.a. Cartesian coordinate system) is one whose axes are perpendicular. Let the original coordinate system be rectangular where the coordinates of point Q are (x, y) and where

$$x = CM \quad \text{and} \quad y = MQ \ .$$

An oblique coordinate system is one whose axes are not perpendicular. Let the axes of the new coordinate system be conjugate diameters of the ellipse. It is apparent that the new coordinate system is oblique where the coordinates of point Q are (x_1, y_1) and where

$$x_1 = CV \quad \text{and} \quad y_1 = VQ \ .$$

The Apollonius equation for the ellipse in the rectangular coordinate system is

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \quad \text{or} \quad \frac{(a+x)(a-x)}{a^2} = \frac{y^2}{b^2} \ . \tag{27}$$

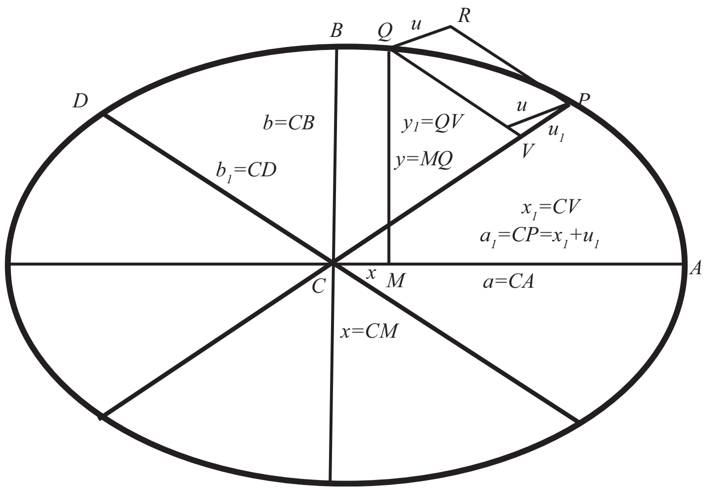


Figure 42. Major radius $a = CA$ and minor radius $b = CB$ form rectangular axes. Conjugate radii $a_1 = CP$ and $b_1 = CD$ form oblique axes. Point Q has rectangular coordinates (x, y) and oblique coordinates (x_1, y_1). In other words, there are two different designations for the same point Q. The two line segments labeled u are equal and parallel.

The Apollonius equation for the ellipse in the oblique coordinate system is

$$\frac{x_1^2}{a_1^2} + \frac{y_1^2}{b_1^2} = 1 \quad \text{or} \quad \frac{(a_1 + x_1)(a_1 - x_1)}{a_1^2} = \frac{y_1^2}{b_1^2} \ . \tag{28}$$

Next define u_1 as

$$u_1 = VP \ .$$

Because $a_1 = CP$, $x_1 = CV$, it follows that

$$a_1 = x_1 + u_1 \ .$$

Thus,

$$a_1 - x_1 = u_1 \quad \text{and} \quad a_1 + x_1 = 2a_1 - u_1 \ .$$

The *Apollonius equation* 28 in the oblique coordinate system becomes

$$\frac{y_1^2}{b_1^2} = \frac{u_1(2a_1 - u_1)}{a_1^2} \ . \tag{29}$$

Part E. Line *EP* has same length as semi-major axis *a*

An ellipse is the set of all points such that the sum of the distances from two fixed points (i.e., the foci S and H) is constant. In other words, as illustrated in Figure 43, the distance from one focus S to any point P on the ellipse plus the distance from that same point P to the other focus H is always the same number for a particular ellipse. In fact, the fixed distance is given by $2a$, where a is the major radius CA. Thus, we may write

$$SP + PH = 2a \ .$$

In terms of components, this equation becomes

$$SE + EI + IP + PH = 2a \ .$$

Let PR be tangent to the ellipse at point P. The lines HI and CE are drawn parallel to the tangent PR. Because the foci are equally distant from the center, we have $SC = CH$. It follows that $SE = EI$.

In addition, it is demonstrated that P is the point of reflection of the ray IPH. By Heron's law of reflection, it follows that $IP = PH$. Thus,

$$EI + EI + IP + IP = 2a$$

or

$$EI + IP = EP = a \ .$$

This equation establishes that line EP has same length as semi-major axis CA.

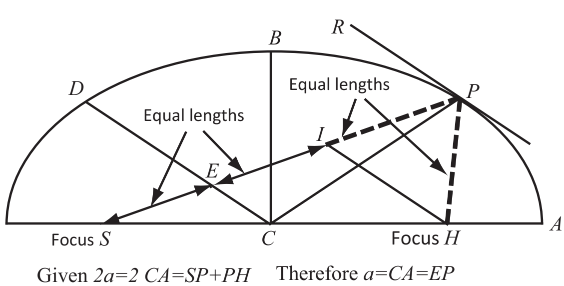


Figure 43. In the ellipse, $SE = EI$ and $IP = PH$, so $EP = CA$.

Part F. Galileo and the path of a projectile

Modern science began with Galileo. His lasting achievement was the overturning of the overwhelming influence of Aristotle on physics. A particular roadblock to progress was Aristotle's concept of the motion of a projectile. According to Aristotle, the motion of a projectile was composed of thrust and descent. The upward thrust was produced by the source (such as the string on a bow or gunpowder), and the descent was produced by the natural tendency of an arrow or cannonball to fall to the ground. Because (as deemed by the conventional wisdom) the two motions could not occur simultaneously, it was necessary for the upward thrust to be completed before the projectile could fall. Thus, a projectile would follow a straight-line path until the upward thrust was consumed; then it would fall straight down. Aristotle had no notion about the theory of gravity.

Aristotle's incorrect theory of the motion of a projectile dominated theoretical physics for centuries. According to Aristotle, a gunner should aim straight at the target. However, practical military men knew this procedure was incorrect, so they would aim above the target. Galileo took the decisive step to correct Aristotle's false description. First, he developed the theory of freely falling bodies based on his realization that a force directly affects the vertical velocity of a freely falling object. Through experimentation, Galileo found this force causes the object to gain equal increments in vertical velocity over equal intervals of time. Thus, the force of gravity produced a constant acceleration g. Galileo expressed this phenomenon mathematically by writing the vertical velocity as $v = gt$, where t denotes time. This equation held whatever the weight of the object might be. The weight of the object represented the force of gravity and could be written as mg, where m denoted mass. Galileo devised a clever method to integrate the velocity with respect to time. The method yielded Galileo's celebrated equation for the fallen distance,

$$u = \frac{gt^2}{2} \quad . \tag{30}$$

Generally, it is agreed that this work of Galileo marks the beginning of physical science as we know it today. However, we should also say the contemporaneous work of Descartes in analytic geometry marks the beginning of mathematical science as we know it today. In this regard, Huygens and Newton use the results of both Galileo and Descartes to advance mathematics and physics.

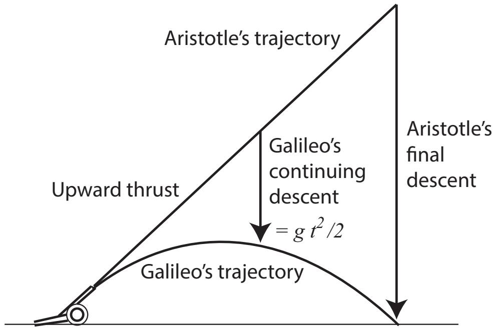


Figure 44. Galileo: Parabolic trajectory with continuing descent $u = gt^2/2$, which gives $g = 2u/t^2$.

Galileo used the Italian word *gravita*, which was translated as *gravity* in English. Next, Galileo applied this result to the calculation of ballistic trajectories with the revolutionary concept that Aristotle's thrust component and Aristotle's descent component could both act at the same time, resulting in a parabolic trajectory (Figure 44). Newton stated in Book 1 of the *Principia*: "Galileo discovered that the descent of bodies varied as the square of time, and that the motion of projectiles was in the curve of a parabola." However, Newton neglected to say that Descartes discovered the way to connect the equation of the parabola (analytic) to the picture of the parabola (geometry).

Part G. Newton's *Proposition XL Problem VI*

Newton puts everything into one diagram, which is shown in Figure 45. All of the upper-case letters on the diagram represent points. We will use lower-case letters to label line segments. In *Principia*, Newton writes:

> *SECTION III. Of the motion of bodies in eccentric conic sections.*
>
> *PROPOSITION XL PROBLEM VI.*
>
> *If a body revolves in an ellipse, it is required to find the law of the centripetal force tending to the focus of the ellipse.*
>
> *Let S be the focus of the ellipse. Draw SP cutting the diameter DK of the ellipsis in E, and the ordinate QV in X; and complete the parallelogram QXPR. It is evident that EP is equal to the greater semi-axis.*

Note that we have already established this result in Part E. Newton will now use this result.

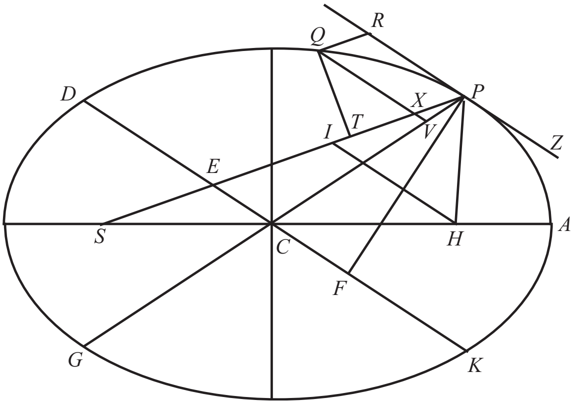


Figure 45. The diagram as given in the *Principia*.

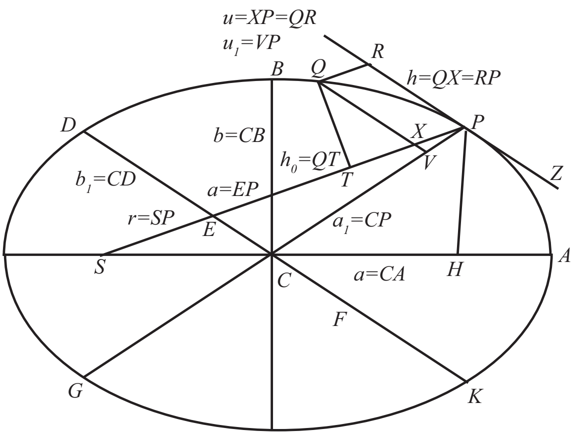


Figure 46. The diagram as given in the *Principia* with lower-case letters added to indicate segments.

Next, as shown in Figure 46, because triangles *EPC* and *XPV* are similar, we have

$$\frac{PX}{PV} = \frac{PE}{PC}$$

and, as also will appear as equation 34 in Part G,

$$\frac{u}{u_1} = \frac{a}{a_1} \ .$$

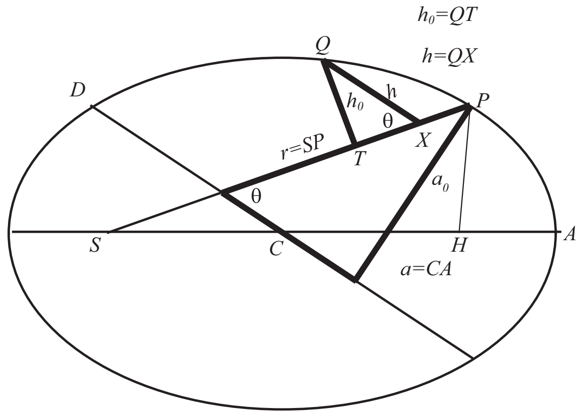


Figure 47. Diagram to illustrate the similar triangles.

Then, as shown in Figure 47, because right triangles *QTX* and *PFE* are similar, we have

$$QX/QT = PE/PF \ ,$$

which is

$$\frac{h}{h_0} = \frac{a}{a_0} \ .$$

Part C also contains

$$\frac{a}{a_0} = \frac{b_1}{b} \ .$$

Thus, we have

$$\frac{h}{h_0} = \frac{b_1}{b} \ .$$

If we square the elements of this equation, we obtain the following equation, which also will appear as equation 35 in Part G,

$$\frac{h^2}{h_0^2} = \frac{b_1^2}{b^2} \ .$$

Part H. Newton's treatment of a planet as a projectile

Newton's elliptical trajectory *PQ* has tangential component $PR = h$ and descending component $RQ = u$, as shown in Figure 48. According to

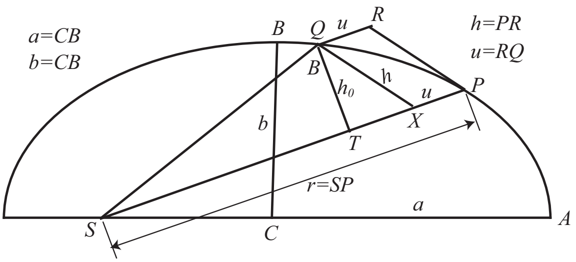


Figure 48. Newton's elliptical trajectory.

Galileo, the descending component is $u = gt^2/2$. According to Kepler, the time for the planet to travel from P to Q is $t = rh_0/2$.

The attracting body is the Sun S. The orbiting body is the planet P. Newton uses Galileo's idea by taking the novel approach of treating a planet as a projectile. Newton reasons that the planet is falling continually toward the sun (see Figure 48). If there is no gravitational force, the planet would travel along the tangent line and travel from P to R in the time interval t (à la the thinking of Aristotle). But there is gravity, so the planet descends the distance u in the time interval. If the points P and Q are close together, then the acceleration of gravity is (approximately) constant. Such is the case in Earth-bound satellites. In time t, the planet travels along the elliptic arc from P to Q, and this arc can be approximated by Galileo's parabola. Line SP has length r. Note that RQ is parallel to SP. Line QX is parallel to the tangent RP. According to Galileo, the descending component is given by equation 30, which is $u = (gt^2)/2$.

At about the same time as Galileo, Johannes Kepler empirically discovers three very important laws. Kepler's first law says that a planet, such as the Earth, travels in an elliptic orbit around the Sun. The Sun S is at one focus of the ellipse. Nothing is at the other focus. The perihelion is the point in the orbit of a planet at which it is closest to the Sun. The aphelion is the point at which it is furthest from the Sun. These two points lie at opposite ends of the major axis. The planet's speed changes along its orbit. The planet moves fastest at the perihelion and slowest at the aphelion.

Kepler's second law states that the radial line SP joining the Sun to a planet sweeps out equal areas during equal intervals of time. In other words, the second law says that the time it takes for the planet to travel from point P to point Q is proportional to the area of the sector PSQ that the planet sweeps out in that time. Again see Figure 48. There is no simple formula for the area of a pie wedge cut from an ellipse, but, if we consider a very short time interval, the area of the sector is very nearly that of triangle PSQ. The area of the triangle is one-half the product of the base r and height h_0. According to Kepler's second law, the equation for time is

$$t = \frac{rh_0}{2} \ . \tag{31}$$

Newton combines Galileo's equation 30 and Kepler's equation 31. The result obtains Newton's planetary gravitational acceleration.

$$g = \frac{2u}{t^2} = \frac{2u}{\left(\frac{rh_0}{2}\right)^2} = \frac{8u}{r^2 h_0^2} = \left(\frac{8u}{h_0^2}\right)\frac{1}{r^2} \ . \tag{32}$$

Unlike Galileo's earthy g, Newton's planetary g is not constant but varies as the planet travels along its orbit. Newton wants to know what happens at point P when Q approaches P. In other words, Newton wants to find the limit of the quotient $8u/h_0^2$ as $P \to Q$. Next we will show what Newton does.

The Apollonius equation 28 for the ellipse in the oblique coordinate system may be written as

$$\frac{u_1(2a_1 - u_1)}{y_1^2} = \frac{a_1^2}{b_1^2} \ . \tag{33}$$

Newton then turns to the geometry of the ellipse. He uses the two equations given in Part G, namely,

$$\frac{u}{u_1} = \frac{a}{a_1} \tag{34}$$

and

$$\frac{h^2}{h_0^2} = \frac{b_1^2}{b^2} \ . \tag{35}$$

Multiplying equations 33, 34, and 35, the result is

$$\frac{u_1(2a_1 - u_1)}{y_1^2}\ \frac{u}{u_1}\ \frac{h^2}{h_0^2} = \frac{a_1^2}{b_1^2}\ \frac{a}{a_1}\ \frac{b_1^2}{b^2} \ . \tag{36}$$

The left side of equation 36 is

$$\frac{u_1(2a_1 - u_1)}{y_1^2}\ \frac{u}{u_1}\ \frac{h^2}{h_0^2} = (2a_1 - u_1)\frac{h^2}{y_1^2}\ \frac{u}{h_0^2} \ . \tag{37}$$

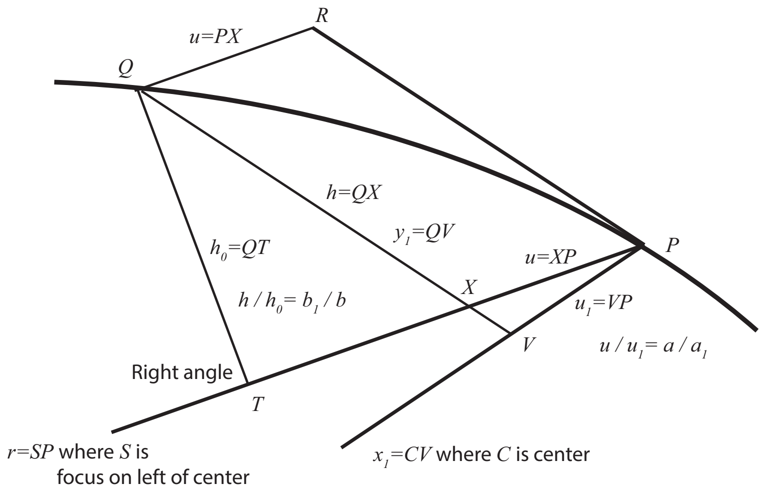


Figure 49. Enlargement of the elliptical path *PQ*. Parallelogram *PRQX* is that shown in Figure 48. Line segment $u = XP$ is along line $r = SP$ from Sun to planet. Line segment $u_1 = VP$ (as shown in Figure 42) is along major conjugate radius $a_1 = CP$.

In Figure 49, line segment u_1 is on the x_1 axis, and line segment h is parallel to the y_1 axis. At this point, Newton uses a rudimentary form of calculus. He is careful to take limits along one or the other of these two oblique axes. In Figure 49, as point Q approaches point P, two things happen:

(a) $VP \to 0$; thus $u_1 \to 0$, so $(2a_1 - u_1) \to 2a_1$.
(b) $XV \to 0$; thus $h \to y_1$.

As a result, the right side of equation 37 approaches a limit given by

$$(2a_1 - u_1)\frac{h^2}{y_1^2}\,\frac{u}{h_0^2} \to 2a_1\,\frac{u}{h_0^2}\ . \tag{38}$$

We replace the left side of equation 38 by the right side of 36 and obtain

$$\frac{2a_1 u}{h_0^2} = \frac{a_1^2}{b_1^2}\,\frac{a}{a_1}\,\frac{b_1^2}{b^2}\ .$$

We see that a_1 and b_1 both disappear and the result is

$$\frac{2u}{h_0^2} = \frac{a}{b^2}\ . \tag{39}$$

The term on the right is constant for any given ellipse. The semi-latus-rectum of the ellipse is

$$\ell = \frac{b^2}{a} \ .$$

so equation 39 becomes

$$\frac{2u}{h_0^2} = \frac{1}{\ell} \ . \tag{40}$$

Newton's planetary gravitational acceleration in equation 32 becomes

$$g = \frac{8u}{h_0^2}\frac{1}{r^2} = \frac{4}{\ell}\frac{1}{r^2} \ . \tag{41}$$

Thus, the planetary gravitational acceleration varies as the inverse square of the radial distance r from Sun to planet.

The first correct account of centrifugal force is given by Huygens in 1662. Huygens introduces a physical principle, which Einstein later calls the principle of equivalence. It states that a gravitational acceleration in a small region of space may be simulated by an appropriate motion of the frame of reference. Using Galileo's equation $u = (gt^2)/2$ and the principle of equivalence, Huygens describes centrifugal force. Huygens uses basic concepts of the theory of motion, including: the proportionality between an object's weight and mass; the relation between force, mass, and acceleration; and the equality of action and reaction.

Although the contribution of Robert Hooke (1635 – 1708) to the theory of gravity often is overlooked, now it is known he proposes that planetary motion is simultaneously composed of two motions: one a thrust (due to inertia) along the tangent line and the other a descent (caused by gravity) toward the central body (the Sun). Hooke also proposes that the force exerted by the Sun varies inversely as the square of the radial distance from the Sun to the planet. However, Hooke is unable to mathematically prove his ideas. As a consequence, Hooke, in a series of letters in 1679, sends his propositions to a mathematics professor at Cambridge University, Isaac Newton, and changes history (and how it is that Hooke claims Newton obtains the inverse square law from him).

If m is the mass of the planet, the gravitational force is given by $F = mg$. Thus, we have obtained Newton's solution to Hooke's assertion—the gravitational force of the Sun on a planet varies inversely as the square of the distance r from the Sun to the planet. If the Sun exerts a gravitational force on the Earth, then the Earth should exert the same force on the Sun. Thus, Newton arrives at the law of universal gravitation—two objects attract each other with a force proportional to the product of their masses and inversely proportional to the square of the distance between them. The symbol G designates the constant of proportionality.

The universal law says the following. Every point mass attracts every single other point mass by a force pointing along the line intersecting both points. The force is proportional to the product of the two masses and inversely proportional to the square of the distance between them. The equation is

$$F = G\frac{m_1 m_2}{r^2} \ .$$

In this equation,

F is the force between the masses,
G is the gravitational constant (6.673×10^{11} N$(\mathrm{m/kg})^2$),
m_1 is the first mass,
m_2 is the second mass, and
r is the distance between the centers of the masses.

Newton is unable to determine of the value of G. However, another result by Hooke provides the means to solve this problem. In 1676, Hooke publishes this important concept (now known as *Hooke's law*) in the form of one word, namely, *CEIINOSSITTUU*. This fabricated word is an anagram. In 1679, Hooke reveals its meaning, given by the four Latin words *ut tensio, sic uis*, which are usually translated into English as, "strain is proportional to stress." A torsion balance measures the value of an unknown stress by the amount of strain produced. In 1798, Henry Cavendish (1731 – 1810) constructs a torsion balance that is sufficiently accurate to measure the gravitational force between two lead balls. The resulting value of G is very close to the value accepted today. Using the accepted value of G, the mass of the earth is obtained as 6×10^{24} kg. Today, Cavendish is remembered as the person who weighed the Earth.

Newton's solution of the law of gravity, which we have just given, makes use of Euclidean geometry. Newton is in possession of his own version of calculus (which he calls the method of fluents and fluxions) at the time. However, *Principia* contains no explicit use of calculus (setting aside a small amount of material that Newton adds in later editions when trying to solidify his priority of invention). Instead of calculus, Newton employs such reasoning as "ultimate ratios," but this reasoning had already been employed in ancient times, notably by Archimedes. Newton's reluctance to use calculus is puzzling. Newton uses uncanny ingenuity in his mathematical arguments, producing again and again a clever geometric demonstration of a fact that today would be done by calculus. England did not teach anything but Newton's geometrical and fluxional methods for more than a century. In contrast, on the continent, mathematicians generally discard Newton's fluxional calculus in favor of the calculus of Leibniz, which spurs the development of 18th century mathematics.

In this section, we have given Newton's geometrical solution of the inverse square law of gravitation. Instead of speculating as to why Newton does not provide a calculus solution based upon his fluents and fluxions, we will now present such a solution. What follows is what might have been.

If a body is in motion, the polar coordinates $r(t)$ and $\theta(t)$ specifying its position are known as *fluents*. As the body moves, the fluents continuously change. The rates (with respect to time t) at which they change are known as velocities or *fluxions*. The fluxions are denoted by

$$v_r = \dot{r} \ ,$$

$$v_\theta = r \ \dot{\theta} \ .$$

Newton develops algorithms for calculating fluxions. For the gravity problem, he would need *fluxions of fluxions*, which represent accelerations or *forces*. According to Kepler, the Earth revolves about the Sun in an elliptical orbit with the Sun at one focus. It is assumed that the Earth is held in this course by a force exerted upon the Earth by the Sun. It is assumed that all of the force is exerted in the direction from the Earth to the Sun. Such a force is called a central force. The central force is

$$\text{central force} = a_r = \ddot{r} - r \, \dot{\theta}^2 \ . \tag{42}$$

Because no sideway force is exerted, it follows that a_θ must be zero; that is,

$$\text{sideways force} = a_\theta = r \ \ddot{\theta} + 2 \ \dot{r} \, \dot{\theta} = 0 \ . \tag{43}$$

This expression for a_θ can be written as

$$\text{sideways force} = a_\theta = \frac{1}{r} \ (\text{fluxion of } r^2 \, \dot{\theta}) = 0 \ .$$

The above equation yields

$$\text{fluxion of } r^2\,\dot{\theta} = 0 \ .$$

If the fluxion is zero, then the fluent $r^2\,\dot{\theta}$ must be constant. The constant is denoted by C, so

$$r^2\,\dot{\theta} = C \ . \tag{44}$$

The equation for the ellipse is given by equation 21, where e is the eccentricity and k is the directrix:

$$\frac{1}{r} = \frac{1}{ke} + \frac{1}{k}\cos\theta \ . \tag{45}$$

Taking the fluxion of each side of equation 21, the result is

$$-\frac{1}{r^2}\dot{r} = -\frac{1}{k}(\sin\theta)\dot{\theta} \ .$$

Using equation 44 to replace $\dot{\theta}$, the result is

$$k\dot{r} = (\sin\theta)C \ .$$

Now taking the fluxion of each side of this equation, the result is

$$k\ddot{r} = (\cos\theta)\ \dot{\theta}\ C \ .$$

Using equation 44 to replace $\dot{\theta}$, the result is

$$\ddot{r} = \frac{C^2}{k\,r^2}\cos\theta \ . \tag{46}$$

Using equation 46 to replace $\ddot{r}$ and using equation 44 to replace $\dot{\theta}$ in the central force 42, the result is

$$\text{central force} = a_r = \ddot{r} - r\ \dot{\theta}^2 = \frac{C^2}{k\,r^2}\cos\theta - r\left(\frac{C}{r^2}\right)^2 \ .$$

Thus,

$$\text{central force} = \frac{C^2}{k\,r^2}\left(\cos\theta - \frac{k}{r}\right) \ .$$

Equation 21 for the ellipse may be written as

$$\cos\theta - \frac{k}{r} = -\frac{1}{e} \ .$$

Therefore,

$$\text{central force} = \frac{C^2}{k\,r^2}\left(-\frac{1}{e}\right) = -\left(\frac{C^2}{ke}\right)\frac{1}{r^2} \ . \qquad (47)$$

Thus, the central force is an inverse square law, which is to be proven.

At the time of Newton's work, the vortex theory of planetary motion as given by René Descartes (1596 – 1650) holds sway. Believing in the impossibility of the vacuum, Descartes claims that matter fills all space. Matter consists of ordinary matter (which we see), ether (which we cannot see), and light. Motion results from the impact of particle upon particle. (As geophysicists, we know that seismic wave motion is caused by the serial action of each particle of rock impacting the next particle in the path of the wave.) Movement in Descartes' universe tends to create a swirl or vortex. The solar system is an ethereal vortex with the sun at the center and the planets swirling around it. A subsidiary vortex would carry the moons around a planet. The universe consists of such vortices, each with a star as center. The vortices fit together like an aggregation of soap bubbles.

Just as we understand seismic waves, we can understand the mechanical universe of Descartes. But then Newton arrives in 1687 with a theory that posits the mysterious, occult quality known as gravity. In time, Newton' theory would prevail. A major reason is that Newton's formula (as derived here) provides the accuracy and predictive scope to explain geophysical and astrophysical phenomena. We accept that gravity exerts a force through the vacuum of space, but we cannot comprehend the mechanism of action at a distance. In other words, we can grasp what gravity does, but we are unable to fathom how. We can understand the universe only up to an insurmountable bound. Beyond that bound, we cannot go. Voltaire (*Letters on the English, Letter XV, On Attraction*, ca. 1778) writes:

> *Vortices may be called an occult quality because their existence was never proved. Attraction, on the contrary, is a real thing because its effects are demonstrated, and the proportions of it are calculated. The cause of this cause is among the Arcana of the Almighty.*
>
> *Procedes huc, et non amplius. (Thus far shalt thou go, and no farther.)*

Newton's geometric proof, just given, is intricate and complicated. In effect, Newton's proof brings together Apollonius, Galileo, and Kepler. Reviewing Newton's work, however, continental mathematicians and astronomers have difficulties understanding his arcane geometrical approach which uses quantities that become vanishingly small. Moreover, Newton makes no attempt at any systematic use of calculus. Rather, it is European mathematicians familiar with the calculus of Gottfried Wilhelm Leibniz who translate Newton's mathematical language into Leibniz's language. Newton's proof is generally considered the greatest scientific achievement in the history of mankind, and it is the subsequent contributions, and mathematics, of others that pave the way for further progress.

This page has been intentionally left blank

Chapter 2

Wave Equation

Seismic ray direction

When Pythagoras was about 18 years old, he went to the island of Lesbos where he worked and learned from Anaximander and Thales of Miletus. Thales had visited Egypt and he recommended that Pythagoras go to Egypt. Pythagoras arrived there around 547 BC when he was in his early 20s. He stayed in Egypt for 21 years and learned a variety of things, probably including some geometry that is routinely credited to the Greeks.

The evidence for this is imposing, literally—the Great Pyramid of Khufu, one of the largest buildings ever constructed. Erected in at least 2500 BC and ranked as the world's tallest building for the next 43 centuries, its volume staggers the imagination. According to the *Encyclopedia Britannica*, St. Peter's in Rome, the cathedrals in Florence and Milan, and Westminster Abbey and St. Paul's in London could all fit inside at the same time. The precision of this massive construction is impressive even by modern standards. Although thousands of presumably unskilled laborers were involved, and the tools they used were very limited, the lengths of the sides of its base all lie within a band of 20 cm. The lengths are

West side: 230.36 m
North side: 230.25 m
East side: 230.39 m
South side: 230.45 m

When originally constructed in ancient times, the surface of the Great Pyramid was covered with highly polished Tura limestone blocks, serving as outer casing stones. Carefully interlocked, these outer stones gave a uniform smooth surface to all four sides. As a result, the entire structure shone brilliant white in the light of the reflected sun. The casing stones were removed in AD 1300 for various building projects in nearby Cairo. They were removed, except for a few along the base of the pyramid. Perhaps if the casing stones were still in place, the sides would be even more accurate than they are today.

This precision, on such a vast scale, suggests sophisticated mathematical knowledge. Analysis of the geometry of the King's Chamber (a rectangular parallelepiped) inside the Great Pyramid is further confirmation. The perfectly aligned walls, ceiling, and floor are composed of large polished blocks of pink Aswan granite. These blocks are smooth to within one-quarter of a millimeter over 6 meter lengths. This granite is so hard that experts say it would be extremely difficult to reach this accuracy today with the best machining equipment. The chamber's dimensions are length 1048 cm, width 524 cm, and height 586 cm. Thus, modern dimensions lead to the conclusion that the floor is designed to be a double square, because the length is exactly twice the width.

Is this the intent of the ancient Egyptian architects? Yes, according to Isaac Newton who establishes to general assent that the floor of the King's Chamber is indeed 10×20 royal cubits. With the royal cubit as the distance parameter, the dimensions of the chamber are length $a = 20$, width $b = 10$, and height $h = 11.18$ (see Figure 1). For the volume to equal a double cube, the height would have to equal the width. But it does not—it exceeds the width by 1.18 royal cubits. Why? The builders apparently have taken extreme care to make the floor a double square, so it is doubtful they simply picked an arbitrary height. And, in fact, they did not,

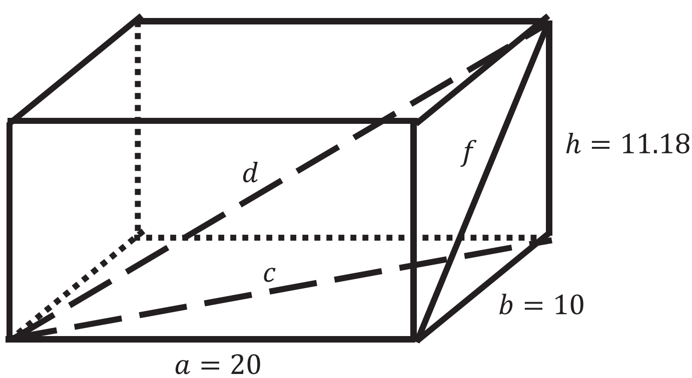


Figure 1. The King's Chamber of the Great Pyramid.

although it will require some arithmetic to answer that question and confirm that the Egyptians were very advanced in some areas of mathematics that weren't "finalized" for several thousand years.

Observe that the floor diagonal is

$$c = \sqrt{20^2 + 10^2} = 10\sqrt{5} = 22.36 \ .$$

Thus, the chamber is designed so that its height,

$$h = 5\sqrt{5} = 11.18 \ ,$$

is one half the length of the floor diagonal. This, at least mathematically, relates the awkward-looking parameter 11.18 to another dimension but, again, we ask why? Observe that the small side of the room has base $b = 10$ and height $h = 5\sqrt{5}$. As a result, the face diagonal of the small side is

$$f = \sqrt{100 + 125} = 15 \ .$$

The room diagonal d is the hypotenuse of the right triangle with sides $f = 15$ and $a = 20$. The result is that

$$d = \sqrt{225 + 400} = 25$$

and that the triangle with sides

$$f = 15, \ a = 20, \ d = 25$$

is a 3:4:5 right triangle. The harmonic proportion of the room shows the intimate relationship between 1:2:3:4:5 and strongly implies that the designers of the King's Chamber knew the basic properties of right triangles at least 2000 years before Pythagoras.

The fascinating mathematics of the King's Chamber can be used to illustrate some mathematics of importance in applied geophysics, even how the direction of a seismic ray can be determined. But first a word about vectors. A vector is specified by magnitude and direction. For example, sailing orders might be given by the vector expressed as 2 km, ENE, which means sail 2 km east–northeast. When a grid is imposed, a vector may be specified by coordinates. If east is the x axis and north is the y axis, then ENE represents an angle of $\pi/8$ radians. (There are $\pi/2$ radians in 90°, so there are $\pi/8$ radians in 90°/4.) Thus, the x coordinate is $2\cos\pi/8$ and the y coordinate is $2\sin\pi/8$, and the above sailing order in this grid system would be the vector (1.74661, 0.97435). This means sail 1.74661 km east and then sail 0.97435 km north.

Two vectors can be multiplied in various ways. The most common form is the dot product, which is simply the sum of the products of the respective coordinates. The dot product of the vectors (1,2) and (3,4) is

$$(1 \times 3) + (2 \times 4) = 3 + 8 = 11 \ .$$

The first step in our application to geophysics is to impose a (x, y, t) coordinate system on the lower wedge of the chamber (see Figure 2). In terms of a traditional two-dimensional seismic survey, the x-axis would be the horizontal axis, the y-axis would be the vertical axis, the t-axis would be the traveltime axis, and the plane $ABFE$ would be the traveltime surface $t(x, y)$. Lines of constant time on the traveltime surface, which represents the dependence of traveltime t on the horizontal and vertical coordinates x and y, appear on the (x, y) plane as contour lines (or level lines) which represent wavefronts. For example, line AB represents the wavefront for traveltime of zero and line CD represents the wavefront for traveltime of FD.

One of the fundamental concepts of calculus is that slope is the rate at which an ordinate of a point on a line changes with respect to a change in the abscissa. In other words, the slope is the tangent of the angle of inclination of the line. The angle of inclination α of line AB is zero, so its slope is $\tan \alpha = 0$. The angle of inclination β of line BF is not zero, and its slope is

$$\tan \beta = FD/DB = h/b = 11.18/10 = 1.118 \ .$$

The partial derivatives of a function give the rate of change of a function in the directions parallel to the coordinate axes. Thus, the partial

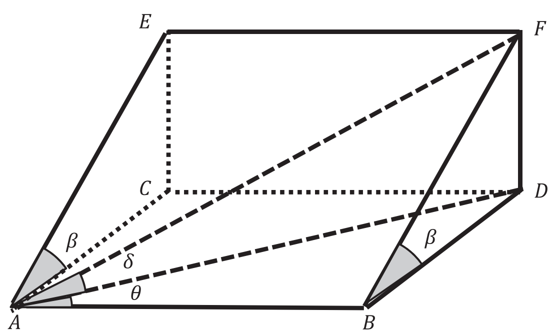


Figure 2. The lower wedge of the King's Chamber.

derivatives of $t(x, y)$ are

$$\frac{\partial t}{\partial x} = \tan \alpha = 0 \ ,$$

$$\frac{\partial t}{\partial y} = \tan \beta = 1.118 \ .$$

However, what is really needed is the rate of change of traveltime t in an arbitrary direction—a directional derivative. In this case, if the specified direction is that of the floor diagonal AD, the directional derivative is the slope $\tan \delta$ of the room diagonal; i.e.,

$$\tan \delta = \frac{FD}{AD} = \frac{h}{c} = \frac{11.18}{22.36} = 0.5 \ .$$

In developing a general formula for the directional derivative, we see that the height FD equals the rise in the elevation of the surface along the line AB plus the rise in the elevation of the surface along the line BD. In other words,

$$FD = AB \tan \alpha + BD \tan \beta \ ,$$

which is

$$FD = AB \frac{\partial t}{\partial x} + BD \frac{\partial t}{\partial y} \ .$$

Therefore, the directional derivative is

$$\tan \delta = FD/AD = \left(\frac{AB}{AD}\right) \frac{\partial t}{\partial x} + \left(\frac{BD}{AD}\right) \frac{\partial t}{\partial y} \ .$$

If θ is the angle that the floor diagonal makes with the x-axis, then

$$\frac{AB}{AD} = \cos \theta \quad \text{and} \quad \frac{BD}{AD} = \sin \theta \ ,$$

and the directional derivative

$$\tan \delta = \cos \theta \frac{\partial t}{\partial x} + \sin \theta \frac{\partial t}{\partial y} \ ,$$

which also can be written as the dot product of two vectors,

$$\tan\delta = (\cos\theta,\ \sin\theta)\cdot\left(\frac{\partial t}{\partial x},\frac{\partial t}{\partial y}\right)\ .$$

The first vector on the right side is the unit vector in the required direction, namely,

$$\boldsymbol{u} = (\cos\theta,\ \sin\theta)\ .$$

The second vector is called the gradient of t; that is,

$$\operatorname{grad} t = \left(\frac{\partial t}{\partial x},\frac{\partial t}{\partial y}\right)\ .$$

Consequently, the directional derivative can be written more concisely as the dot product

$$\tan\delta = \boldsymbol{u}\cdot\operatorname{grad} t\ .$$

Thus, we have the important result

$$\text{directional derivative} = \boldsymbol{u}\cdot\operatorname{grad} t\ .$$

The "robustness" of this definition of directional derivative is demonstrated by analyzing the King's Chamber if it were rebuilt as indicated in Figure 3. The traveltime surface is the plane *AGHE*. This more complicated design has a rise in elevation along the horizontal axis and along the vertical axis.

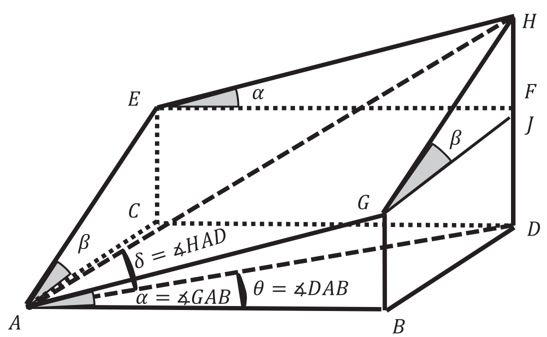


Figure 3. The reconstructed King's Chamber.

The rise in the elevation of the surface is

$$DH = BG + JH = AB \tan \alpha + BD \tan \beta \ ,$$

which, as before, provides the required formula

$$\tan \delta = \frac{DH}{AD} = \left(\frac{AB}{AD}\right) \tan \alpha + \left(\frac{BD}{AD}\right) \tan \beta = u \cdot \operatorname{grad} t \ .$$

Today's "modern" view of differentiability in two dimensions, x and y, allows a function $t(x, y)$ of two variables to be differentiated at a point (x_0, y_0) if the surface it defines in (x, y, t) space looks (in the limit) like a plane near the point. The plane that $t(x, y)$ resembles is called the *tangent plane*, and thus the formulas derived from the right angles in the King's Chamber apply for any differentiable function. The directional derivative is illustrated in Figure 4.

The modern, full version of Fermat's principle states that time must at a stationary value. By a stationary value, we mean a value for which the slope of $f(x)$ versus x is zero. Equivalently, a stationary value is one for which the function is either at a maximum or a minimum, or else it occurs at an inflection point where the (horizontal) tangent is zero (for example, see Figure 5).

For something to think about, convince yourself that the directional derivative is greatest in the direction of the gradient. (In other words, rays move in the direction of the gradient of the traveltime.) In this chapter, we derive the equation

$$\text{directional derivative} = \boldsymbol{u} \cdot \operatorname{grad} t \ ,$$

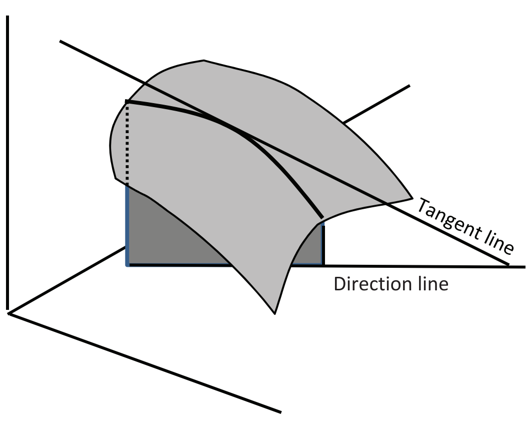


Figure 4. The directional derivative is the slope of the tangent line to the curve obtained by intersecting the surface with the vertical plane though the direction line.

Figure 5. The three examples of stationary values.

where $\boldsymbol{u}$ is the unit direction vector and grad t is the gradient of the traveltime. Use this equation. The key to understanding difficult mathematics is to break down the problem into elementary components.

The eikonal equation and Pythagoras

The computation of traveltimes from a velocity function is required in many seismic processing and modeling schemes—in particular Kirchhoff depth migration and related methods. One of the most popular methods uses the eikonal equation, whose derivation goes back to the work of Pythagoras about 2500 years ago.

Pythagoras, the Greek philosopher and mathematician, was born on the small island of Samos in the Aegean Sea about 570 BC, a century before the golden age of classical Greece. He founded a philosophical and religious society in Croton on the east coast of the tip of Italy about 532 BC. His closest followers lived permanently within the gates of the society, had no personal possessions, were vegetarians, and followed a strict code of secrecy.

Pythagoras, of course, is best known now for the Pythagorean theorem in geometry—one of the most famous and important of all mathematical formulas. However, additional mathematical discoveries also are attributed to Pythagoras, or rather more generally to the Pythagoreans. A central belief of Pythagoras and his followers is that "everything is number," which they describe as a quantity that could be expressed as a ratio of two integers (i.e., a rational number). The Pythagoreans demonstrate that pitch could be represented as a simple ratio of the lengths of equally tight strings. The Pythagoreans also are credited with, by using their own theorems, disproving their own belief that all numbers are rational. This is the famous demonstration that the square root of two (as given by the length of the hypotenuse of an equilateral right triangle with each leg equaling one) must be irrational. (According to legend, the Pythagoreans attempted to keep this discovery a secret, because it invalidates their entire raison d'etre, and the man who developed the proof was killed.)

The Pythagoreans knew that the sum of the angles of a triangle is equal to two right angles and the generalization that the interior angles of a polygon with n sides have a sum of $2n - 4$ right angles and the sum of its exterior angles equals four right angles. The Pythagoreans knew of the five regular solids. It was thought that Pythagoras himself knew how to construct the first three, but it was unlikely that he would have known how to construct the other two.

Furthermore, the Pythagoreans were among the greatest astronomers of their times. They taught that the Earth was a sphere at the center of the universe, recognized that the orbit of the Moon was inclined to the equator of the Earth, and were among the first to realize that Venus the evening star was the same celestial body as the morning star.

The famous Pythagorean theorem has a close relative, here termed the "secret" Pythagorean theorem, that is the basis of the modern differential eikonal equation. Figure 6 shows the right triangle *BAO* with horizontal leg *OA*, vertical leg *OB*, hypotenuse *AB*, and altitude *OC*. The Pythagorean theorem for this right triangle states that

$$OA^2 + OB^2 = AB^2$$

or, in the familiar wording, the sum of the squares of the legs of a right triangle equals the square of the hypotenuse. The "secret" Pythagorean theorem says that the sum of squares of the reciprocals of the legs of a right triangle equals the square of the reciprocal of the altitude:

$$OA^{-2} + OB^{-2} = OC^{-2} \ .$$

This strikes many, at first glance, as counterintuitive, but the proof is straightforward. Draw line *OD* collinear with *OC* and with length that is the reciprocal of the length of *OC* (i.e., $OD = 1/OC$). Drop line *DE* perpendicular to the horizontal. Then $DE = OD \sin\theta$. But $\sin\theta = OC/OB$.

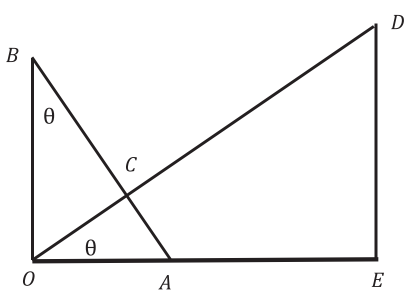


Figure 6. Right triangle *BAO* with horizontal leg *OA*, vertical leg *OB*, hypotenuse *AB*, and altitude *OC*.

The key insight is to recall that, by definition, $(OD)(OC) = 1$, and, therefore,

$$DE = OD\left(\frac{OC}{OB}\right) = \frac{1}{OB}$$

or, in words, the length of DE is the reciprocal of the length of OB. Similarly, the length of OE is the reciprocal of the length of OA (i.e., $OE = OA - 1$). Thus, the "regular" Pythagorean theorem for right triangle ODE, i.e.,

$$OE^2 + DE^2 = OD^2$$

is the secret Pythagorean theorem for right triangle BAO, i.e.,

$$\left(\frac{1}{OA}\right)^2 + \left(\frac{1}{OB}\right)^2 = \left(\frac{1}{OC}\right)^2 .$$

For example, if $OA = 0.5000$, $OB = 0.8660$, and $OC = 0.4330$, their respective reciprocals are $OE = 2$, $DE = 1.1547$, and $OD = 2.3094$.

So, how is this secret Pythagorean theorem used in modern applied geophysics? In Figure 7, assume O is a point on a wavefront at a given instant of time. Let line BCA be the wavefront at subsequent time increment Δt. The distance OA is the horizontal space increment Δx traveled by the wavefront in time increment Δt. Similarly, distance OB is the vertical space increment Δy traveled by the wavefront in the same time increment. Because the wavefront is normal to the raypath, distance OC is the space increment Δs traveled along the raypath by the wavefront in time increment Δt.

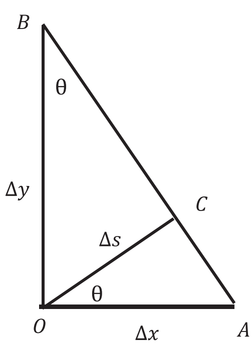


Figure 7. Wavefront *BCA* at subsequent time increment *OC*.

The secret Pythagorean theorem in terms of legs Δx and Δy and altitude Δs is

$$\left(\frac{1}{\Delta x}\right)^2 + \left(\frac{1}{\Delta y}\right)^2 = \left(\frac{1}{\Delta s}\right)^2 .$$

If we multiply this equation by $(\Delta t)^2$, we obtain

$$\left(\frac{\Delta t}{\Delta x}\right)^2 + \left(\frac{\Delta t}{\Delta y}\right)^2 = \left(\frac{\Delta t}{\Delta s}\right)^2 ,$$

and suddenly we are practically face-to-face with the very useful geophysical concept of "slowness," which is defined as the reciprocal of velocity.

The apparent speed along the horizontal direction is $\Delta x/\Delta t$. Thus, the apparent slowness along the horizontal direction is $\Delta t/\Delta x$. Similarly, the apparent speed along the vertical direction is $\Delta y/\Delta t$ and the apparent slowness along the vertical direction is $\Delta t/\Delta y$. The actual speed along the raypath direction is $\Delta s/\Delta t$, and thus the actual slowness along the raypath direction is $\Delta t/\Delta s$.

Therefore, the above equation can be interpreted as saying the sum of squares of the apparent slownesses in the coordinate directions equals the square of the actual slowness in the raypath direction. Because the actual slowness in the raypath direction is the reciprocal of seismic velocity v, the equation may be written

$$\left(\frac{\Delta t}{\Delta x}\right)^2 + \left(\frac{\Delta t}{\Delta y}\right)^2 = \left(\frac{1}{v}\right)^2 .$$

In the limit, as the increments become smaller, this equation becomes a differential equation, the so-called eikonal equation. If we add the third spatial dimension, the eikonal equation is

$$\left(\frac{\partial t}{\partial x}\right)^2 + \left(\frac{\partial t}{\partial y}\right)^2 + \left(\frac{\partial t}{\partial z}\right)^2 = \left(\frac{1}{v}\right)^2 .$$

In this equation, the function $t(x, y, z)$ is the traveltime (also called the eikonal) from the source to the point (x, y, z) and $1/v(x, y, z)$ is the slowness (or reciprocal velocity) at that point. This is practically applied by using the eikonal equation to describe traveltime propagation in an isotropic medium when the velocity function is known at all points in space and, as an initial condition, the source or a particular wavefront is specified. Furthermore, one must choose one of the two branches of the solutions (either the wave going from the source or the wave going to the source). The eikonal equation then yields the traveltime field $t(x, y, z)$ in a heterogeneous medium, as required for migration and other seismic processing needs.

The Pythagorean theorem has many proofs, several of which we detail in the first section of Chapter 1. The most famous is Proposition 47 from Book 1 of Euclid's *Elements*. We are grateful to Chris Liner for providing us with another proof of the secret Pythagorean theorem which, like its more famous relative, can be evaluated from several directions.

The Pythagorean theorem establishes that

$$a^2 + b^2 = c^2 .$$

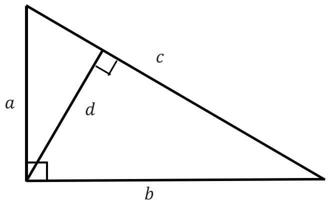


Figure 8. Right triangle with sides a, b, and c.

Furthermore,

$$\frac{ab}{2} = \frac{cd}{2} \ ,$$

because both equal the area of the same triangle (see Figure 8). The latter can be rearranged as

$$c = \frac{ab}{d} \ .$$

Squaring both sides and then substituting for c^2 leaves

$$a^2 + b^2 = \left(\frac{ab}{d}\right)^2 \ .$$

Finally, multiplying both sides by $1/(a^2b^2)$ produces the secret Pythagorean theorem

$$\left(\frac{1}{a}\right)^2 + \left(\frac{1}{b}\right)^2 = \left(\frac{1}{d}\right)^2 \ .$$

Michael Faraday and the eikonal equation

It will probably surprise some to see the name of Michael Faraday (1791 – 1867) because he is rarely, if ever, associated with seismology or mathematics. Some of Faraday's concepts, however, provide an effective analog for some fundamentals of exploration seismology.

As has often been told, Faraday metamorphosed, in a remarkably short time, from a barely literate apprentice bookbinder into one of the world's most influential scientists. His ingenious experiments yielded some of the most significant principles and inventions in scientific history. He developed the first dynamo (in the form of a copper disk rotated between the poles of a permanent magnet), which was the precursor of modern dynamos and generators. He (and independently Joseph Henry) discovered electromagnetic induction, and a vast industry resulted from this work. During the 1830s, he laid the foundations of classical electromagnetic field theory, later fully developed mathematically by James Clerk Maxwell, and his concept of a field based on lines of force became a fundamental principle of modern theoretical physics. Faraday made so many discoveries during his 40-year career that the 1981 edition of the *Encyclopedia Britannica* contained six separate entries on the results of his work in addition to a long biography. The combination of the amazing life and astonishing discoveries prompted the famous novelist Aldous Huxley to write: "Even if I could be Shakespeare, I think I should still choose to be Faraday."

Static or frictional electricity was known to the ancients. Objects became charged in two ways, called positive and negative, governed by the rule that like charges repel and unlike charges attract. A field of fixed charges, an electrostatic field, was described by two physical quantities, the field strength vector $\boldsymbol{E}$ and the potential ϕ.

The field strength vector is defined as $\boldsymbol{E} = \boldsymbol{F}/q$, where q is the test charge and $\boldsymbol{F}$ is the force acting on this charge at the given point in the field. Faraday proposes the use of lines of force as a means to visualize an electric field. A line of force is a straight or curved line whose tangent at each point coincides with the direction of the field strength vector. Lines of force begin at positive charges and end at negative charges. The concept of a potential is now almost universally regarded as representing a powerful advance in knowledge but, as will be discussed later, that was not always true.

In the electrical case, the potential ϕ at a point is the work per unit charge that would be necessary to carry a positive test charge from infinity to that point. The work is then said to be stored in the field as potential energy. Because potential is a measure of work, potential is not directional. The potential at a point is represented by a real number (i.e., a scalar), while the force acting on a test body would be a vector. How is the concept of a potential used? The answer is that $\boldsymbol{E}$ can be obtained from the potential by simply taking the gradient (usually designated by grad). Like a sheep in wolf's clothing, a gradient is a derivative in vector's clothing. The basic equation, which says that the negative gradient of the potential gives the electric field, is

$$-\mathrm{grad}\,\phi = \boldsymbol{E}\ .$$

Any problem with specified charges can be solved by first computing the potential, and then using this equation to obtain the field. There is a physical significance to this equation. Because $\boldsymbol{E}$ is the gradient of a scalar potential, it follows from the mathematics of vector calculus that the curl of $\boldsymbol{E}$ must vanish. This condition represents a fundamental property of electrostatics; namely, the electric field in electrostatics is curl-free. This equation can be used to obtain a geometrical description of an electric field. First, the field lines (that is, lines that are always tangent to the electric field vector) are drawn. Next, the surfaces of equal potential are drawn. Because the electric field vector is the negative gradient of the potential, it follows that $\boldsymbol{E}$ is perpendicular to the equipotential surface. If $\boldsymbol{E}$ is not, then $\boldsymbol{E}$ would have a component in the surface and, in turn, that component would be changing in the surface, which means that the surface would not be equipotential. Consequently, to avoid this contradiction, the lines of force must be always perpendicular to the equipotential surface (Figure 9).

In common usage, the quotient of the change in elevation divided by the change in distance often is called the *gradient*. For example, if the hill rises 1 m in a horizontal distance of 5 m, the (common-usage) gradient would be 1 over 5. In mathematical language, this (common-usage) gradient is actually the directional derivative in the direction of the horizontal separation. In other words, when surveyors speak of gradient, they often mean directional derivative. In order for the directional derivative to be a gradient (in the mathematical sense), the direction of the horizontal separation must be in the steepest direction. It is useful to keep this distinction between directional derivative and gradient. The gradient is a two-dimensional vector that points in the direction of the greatest slope. The dot product of a gradient vector and an arbitrary direction vector gives the rate of change of the function in that direction.

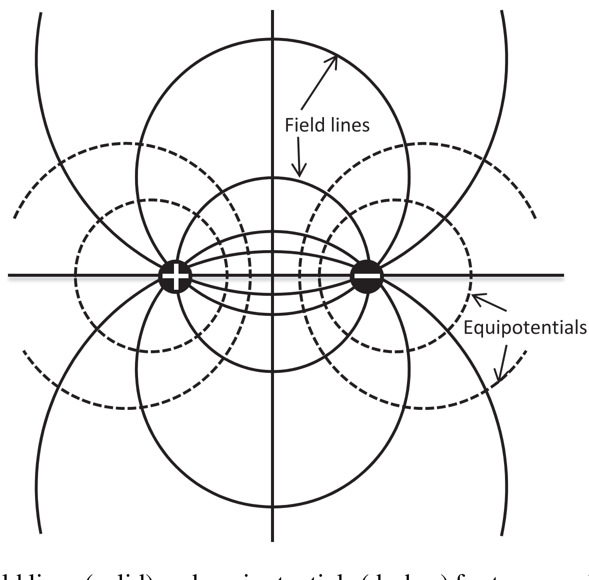


Figure 9. Field lines (solid) and equipotentials (dashes) for two equal and opposite point charges. The positive charge is an isolated source, and the negative charge is an isolated sink. The equipotential surfaces are everywhere perpendicular to the field lines.

First developed to solve problems of gravitational attraction, the concept of potential can be understood in that context. In fact, a simple gravitational analogy is helpful in explaining potential. Over a small region, gravity may be taken as uniform and parallel (simply stated, down). We do work in carrying an object up a hill. This work is stored as potential energy, and it can be recovered by descending in any way we choose. An imaginary terrain, with the accompanying topographic map, can be used to visualize a potential function.

Topographic maps provide information about elevation of the surface above sea level by contour lines. Each point on a contour line has the same elevation, so a contour line represents an equipotential curve. A set of contour lines tells the trained interpreter the shape of the terrain. Hills are represented by concentric loops. A valley is an elongated depression in the landscape formed by the action of water or carved out by glaciers. Valley bottoms appear as "U" or "V" shaped contour lines with their closed end pointing toward higher elevation. Steep slopes have closely spaced contour lines; gentle slopes have widely spaced contour lines. The contour interval is the elevation difference between adjacent contour lines. Using our gravitational field analogy, the contour lines on a topographic map are lines of constant elevation above sea level and hence of constant gravitational potential energy. If we let a ball roll down a mountain, the ball rolls down a path perpendicular to the contour lines—i.e., down the steepest descent or negative gradient. So, if we could measure the contour lines before releasing the ball, we could predict the path it would follow down the mountain. The downward path is the curve of steepest slope or negative gradient.

The gradient in three dimensions is an important concept as well.

This brings us (fortunately or unfortunately depending upon individual predilection) back to the mathematics and the link between Faraday's theory about potential fields and exploration seismology. Vectors are those quantities, often denoted in bold face, that have a magnitude and direction. A fairly detailed description of the basic mathematical manipulation of vectors can be found in Sheriff's *Encyclopedic Dictionary of Applied Geophysics* (2002). The key point for this section is that, in a Cartesian coordinate system, a two-dimensional vector may be represented as the ordered pair of real numbers. If $\boldsymbol{i}$ represents the unit vector in the x direction, and $\boldsymbol{j}$ the unit vector in the y direction, then $\boldsymbol{r} = x\boldsymbol{i} + y\boldsymbol{j}$. The magnitude of the vector is

$$|\boldsymbol{r}| = r = \sqrt{x^2 + y^2} \; .$$

Let the two vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ have magnitudes $|\boldsymbol{v}| = v$ and $|\boldsymbol{w}| = w$. The dot product of the two vectors is defined as the scalar value

$$\boldsymbol{v} \cdot \boldsymbol{w} = |\boldsymbol{v}|\,|\boldsymbol{w}| \cos\theta = vw \cos\theta \ ,$$

where θ is the angle between $\boldsymbol{v}$ and $\boldsymbol{w}$. Two nonzero vectors are orthogonal (perpendicular) if and only if their dot product is zero.

What is the counterpart of potential in seismic theory? We assume, for simplicity, an isotropic medium. The traveltime function $t(x, y)$ is analogous to the potential. This function is a scalar function that represents the traveltime surface as a function of x and y. The x axis is the horizontal axis and the y axis is the vertical (depth) axis. The gradient (or gradient vector field) of a scalar function t is denoted ∇t where ∇ (the nabla symbol) denotes the vector differential operator called del. In the case of two Cartesian coordinates, the operator del is

$$\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right) .$$

The notation grad t also is commonly used for the gradient. More specifically, the gradient of $t(x, y)$ is the vector function whose first component is the partial derivative of t with respect to x and whose second component is the partial derivative of t with respect to y. In other words, the gradient of t is

$$\operatorname{grad} t = \nabla t = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right) t = \left(\frac{\partial t}{\partial x}, \frac{\partial t}{\partial y}\right) .$$

Let $u = (\cos\theta,\ \sin\theta)$ be the unit vector in a given direction. The directional derivative is the dot product

$$\tan\delta = \boldsymbol{u} \cdot \operatorname{grad}\ t \ ,$$

where δ is the angle of elevation. An important application of this expression for the directional derivative is in finding the direction of steepest increase (or decrease) of the function $t(x, y)$. The directional derivative is greatest for the direction in which the dot product of $\boldsymbol{u}$ with the gradient vector is greatest, which is exactly when $\boldsymbol{u}$ points in the same direction as the gradient. Another application is in finding the direction of no increase (or decrease) of the function $t(x, y)$. The directional derivative is zero for the direction in which the dot product of $\boldsymbol{u}$ with the gradient vector is zero, which is exactly when $\boldsymbol{u}$ points in the same direction as the contour curve. The value of the function neither decreases nor increases along a contour curve. The gradient vector is perpendicular to the contour curve and the magnitude of the gradient indicates the steepness of the slope.

What is the counterpart of line of force in seismic theory? If the potential $t(x, y)$ is the seismic traveltime function, then its contour curves represent wavefronts. A vector field is a rule that assigns a vector, in our case the gradient, $\operatorname{grad} t(x, y)$, to each point. In visualizing a vector field, we imagine there is a vector extending from each point. Thus, the vector field associates a direction to each point. If the seismic energy (for the most part) moves in such a manner that its direction at any point coincides with the direction of the gradient at that point, then the curve traced out is called a *seismic ray*. A ray corresponds to the flow line (or the line of force) in other disciplines.

A vector function can be used to represent a space curve. Here, for simplicity, we use only two dimensions, but a third dimension always can be added. Let $s(t)$ denote arc length measured along the curve. Suppose the curve is defined by the equations $x = x(s)$ and $y = y(s)$. This curve can be represented by the vector $\boldsymbol{r}(s) = [x(s), y(s)]$. The vector extends from the origin to the point on the curve. As s increases, the tip of the vector traces out the curve. The arclength s of the curve is defined by $ds^2 = dx^2 + dy^2$. From $\boldsymbol{r} = (x, y)$, we have

$$d\boldsymbol{r} = (dx, dy) \ .$$

Thus,

$$|d\boldsymbol{r}|^2 = dx^2 + dy^2 \ ,$$

so

$$|d\boldsymbol{r}| = ds \ .$$

The derivative vector $\boldsymbol{\tau}$ is

$$\boldsymbol{\tau} = \frac{d\boldsymbol{r}}{ds} = \left(\frac{dx}{ds}, \frac{dy}{ds}\right) = \frac{dx}{ds}\boldsymbol{i} + \frac{dy}{ds}\boldsymbol{j} \ .$$

The derivative vector points in the direction of movement and thus is tangent to the curve. The magnitude of the derivative vector $\boldsymbol{\tau}$ is

$$|\boldsymbol{\tau}| = \left|\frac{d\boldsymbol{r}}{ds}\right| = \frac{|d\boldsymbol{r}|}{ds} = \frac{ds}{ds} = 1 \ .$$

Because $|\boldsymbol{\tau}| = 1$, it follows that $\boldsymbol{\tau}$ is the unit tangent vector to the space curve. The requirement that the unit tangent $\boldsymbol{\tau}$ and the gradient have the same direction is

$$\operatorname{grad} t(x, y) = n(x, y)\ \boldsymbol{\tau}(x, y)$$

where $n(x, y)$ is some scalar-valued function. The flow lines of the vector field $n(x, y)\,\boldsymbol{\tau}(x, y)$ are the same as the flow lines of $\operatorname{grad} t(x, y)$, because the scalar-valued function $n(x, y)$ cannot affect the direction. The above equation $\operatorname{grad} t = n\,\boldsymbol{\tau}$ can be written as

$$\left(\frac{\partial t}{\partial x}, \frac{\partial t}{\partial y}\right) = n\,\boldsymbol{\tau}\ .$$

If we take the squared magnitude of the above equation, we obtain

$$\left(\frac{\partial t}{\partial x}\right)^2 + \left(\frac{\partial t}{\partial y}\right)^2 = n^2\,|\boldsymbol{\tau}|^2 = n^2\ .$$

Here, we use the fact that $\boldsymbol{\tau}$ is a unit vector. The previous section on the eikonal equation provides that equation as

$$\left(\frac{\partial t}{\partial x}\right)^2 + \left(\frac{\partial t}{\partial y}\right)^2 = \frac{1}{v^2}\ .$$

Thus, the scalar-valued function is the reciprocal of seismic velocity; that is,

$$n(x, y) = \frac{1}{v(x, y)}\ .$$

Strictly, velocity is a vector, but in seismology the term velocity usually refers to speed, that is, the propagation rate of a seismic wave without implying direction. Another word for speed is swiftness. As a result, n is called slowness. The vector eikonal equation is

$$\operatorname{grad} t = n\boldsymbol{\tau}\ ,$$

where t is the seismic wavefront, n is the slowness, and $\boldsymbol{\tau}$ is the unit tangent to the seismic ray. The left side involves the wavefront, the right side involves the ray, and the connecting scalar is the slowness. At shallow depths, the slowness is large, so the gradient is large and the wavefronts are closely spaced. At large depths, the slowness is small, so the gradient is small and the wavefronts are widely spaced. In the seismic case, instead of equipotential surfaces, we have wavefronts, and instead of lines of force, we have seismic rays, but the mathematics are the same.

An *isochron* is a line or curve on a map connecting points of identical travel time. An equivalent concept is a *contour line*, which is a line joining points of equal elevation on a map. In three dimensions, the isochron or contour line becomes a *level surface*. In a field of force, a level surface is a surface perpendicular to all lines of force. In other words, a level surface is an *equipotential* (i.e., having the same potential).

The mathematical development given here may seem rather straightforward and, indeed, rather simplistic to those with basic training in vector calculus. But this was not always the case. Faraday's initial presentation of lines of force, on 3 April 1846, was greeted with ridicule by his most eminent contemporaries. One went so far as to publish the suggestion that the self-taught Faraday, who had virtually no education in or knowledge of mathematics, should leave theoretical physics to the properly trained.

This is a depressingly common scenario in science. Great theoretical breakthroughs are routinely dismissed out of hand by the contemporary experts, only to be later adopted by younger scientists and rather quickly proven. The opposite unfortunately is also true, as exemplified by the tenacity of pseudo-science and common misconceptions. Followers of, for example, the flat-earth theory are using precisely the above arguments to make their case ("mainstream science ignores us").

Although Faraday did not gain fame in his lifetime, he certainly did in the aftermath and now holds a prominent position in the history of science. James Clerk Maxwell, 40 years younger than Faraday, published the first of his three famous papers on electromagnetism in 1855, the year that Faraday retired, and Maxwell had his entire electromagnetic theory worked out in less than a decade. Galileo and Huygens laid the foundations for Newton, as exemplified by the three laws of motion, which inaugurated classical physics. Faraday and Gauss laid the foundations for Maxwell, as exemplified by Maxwell's equations, which inaugurated modern physics.

Other examples of initially scorned ideas, of more than minimal import to geophysics, are Fourier analysis and plate tectonics. The very eminent geophysicist J. Tuzo Wilson told one of the authors of this book (RDC) that he had been "roundly booed at the AAPG convention" after making a presentation in support of plate tectonics in the early 1960s.

This initial resistance by great scientists to some of the greatest scientific breakthroughs is perhaps a bigger mystery than those which Faraday unraveled. A recent entertaining, and well written, account of Faraday's remarkable life and career is *The Electric Life of Michael Faraday*, by Alan Hirshfield (2006). In *Maxwell on the Electromagnetic Field*, Thomas

K. Simpson (1997) provides an almost line-by-line analysis of Maxwell's 1855, 1861, and 1864 papers which develop the mathematical foundation of the theory that began with Faraday's work in the 1830s. Simpson also provides extensive background on the lives and accomplishments of Faraday, Maxwell, and Lord Kelvin (who did much of the initial work on the mathematics used by Maxwell). The book has extensive illustrations and commentaries designed to assist readers with limited knowledge of mathematics and physics. It is also quite witty in places, which would have delighted Maxwell, whose fondness for clever wordplay was well known during his lifetime.

Pressure and particle velocity

Before we address the specifics of seismic exploration, it is helpful to look into the work of Jean-Baptiste le Rond d'Alembert (1717 – 1783), the discoverer of a principle now universally known as the *d'Alembert principle*. It is the dynamic analogue to the principle of virtual work for applied forces in a static system. If the negative terms in accelerations are recognized as inertial forces, the d'Alembert principle is: *The total virtual work of the impressed forces plus the inertial forces vanishes for reversible displacements.* Essentially, the d'Alembert principle is a restatement of Newton's second law of motion. It says that the second law may be viewed as a balance between real and fictitious inertial reaction forces.

In 1744, d'Alembert published his *Traité de l'Équilibre et du Mouvement des Fluides*, in which he applies his principle to fluids; this led to partial differential equations which he was then unable to solve. In 1745, he developed the part of the subject which dealt with the motion of air in his *Théorie Générale des Vents*, and this again led him to partial differential equations. A second edition, in 1746, was dedicated to Frederick the Great of Prussia, and procured an invitation to Berlin and the offer of a pension; he declined the former, but subsequently, after some pressing, pocketed his pride and the latter. In 1747, d'Alembert applied differential calculus to the problem of a vibrating string, and arrived at the *wave equation* (in which he assumed that $c = 1$):

$$\frac{\partial^2 u(x,t)}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 u(x,t)}{\partial t^2} \quad . \tag{1}$$

He then showed that the wave equation has the solution

$$u(x,t) = f(x+t) + g(x-t) \ , \tag{2}$$

where f and g are arbitrary functions. This solution was published in the transactions of the Berlin Academy for 1747.

The proof begins by saying that, if $\partial u/\partial x$ be denoted by p and $\partial u/\partial t$ by q, then we have the exact differential

$$du = p\,dx + q\,dt \ .$$

In these quantities, the wave equation 1 becomes

$$\frac{\partial q}{\partial t} = \frac{\partial p}{\partial x} \ .$$

Therefore, $p\,dt + q\,dx$ is also an exact differential. Denote it by dv, so

$$dv = p\,dt + q\,dx \ .$$

Hence,

$$du + dv = p\,dx + q\,dt + p\,dt + q\,dx = (p+q)(dx+dt)$$

and

$$du - dv = p\,dx + q\,dt - p\,dt - q\,dx = (p-q)(dx-dt) \ .$$

Thus, $u + v$ must be a function of $+t$, and $u - v$ must be a function of $x - t$. We may therefore put

$$u + v = 2f(x+t)$$

and

$$u - v = 2g(x-t) \ .$$

By adding, we can eliminate the v terms, leaving

$$u(x,t) = f(x+t) + g(x-t) \ ,$$

which is the d'Alembert solution shown in equation 2.

Euler took up the matter and showed that, for the wave equation 1 with $c \neq 1$, the general solution is

$$u(x,t) = f(x+ct) + g(x-ct) \ .$$

The chief remaining contributions of d'Alembert to mathematics were on physical astronomy, especially on the precession of the equinoxes, and on variations in the obliquity of the ecliptic. These were collected in his

Système du Monde, published in three volumes in 1754. The originator of the d'Alembert principle must be regarded as not only brilliant but also inspiring.

Now let us return to seismic exploration. The subsurface rock layers transmit and reflect seismic waves. Seismic waves travel with a velocity dependent on the nature of the rock itself. When a seismic wave traveling in one rock meets an interface with another rock, the wave is partially reflected back into the first rock and partially transmitted into the second rock. A seismic source at or near the surface transmits seismic energy into the sedimentary layers, and because of reflections from the various interfaces some of this energy arrives back at the surface over time. The various arrival times of these reflected events convey intelligence about the subsurface rock layers.

A primary reflection is a reflection that travels directly down to the interface, and then directly back up to the surface. A multiple reflection is a reflection that bounces back and forth among various interfaces as it proceeds on its trip. The reflected waves, both primaries and multiples, are detected on the surface by receivers. The purpose of digital seismic processing is to transform the raw data into computer-generated images of the subsurface geological structures.

On land, the source of energy may be a dynamite explosion, thumping produced by lifting and dropping a heavy weight, or vibrations introduced by a vibrator coupled with the earth. The seismic source is generated at or near the earth's surface. Preplaced geophones on the ground run in a line or on a grid, and the signals picked up by these detectors are digitized and recorded. On marine surveys, a source is actuated every few seconds as the vessel moves over a predetermined course. The air gun is the predominant energy source at sea. The seismic signals are picked up by hydrophones embedded in a cable trailing the vessel below the water's surface, and, as in land surveys, the data are digitized and recorded.

The three main types of seismic waves are:

a) P-waves (i.e., primary or compressional waves) are longitudinal waves that represent particle motion toward and away from the source. They travel within the earth's interior, so they are classified as body waves. Their propagation velocity ranges from about 2 km/sec to 14 km/sec depending upon the transmission medium.
b) S-waves (i.e., secondary or shear waves) are transverse waves that represent particle motion at right angles to the direction of propagation. They travel within the earth's interior, so they are classified as body waves. Because they are transverse waves, they do not travel through

liquid, such as the water layer in exploration at sea. Their propagation velocity is about 0.6 times that of P-waves.

c) Surface waves, such as Love waves and Rayleigh waves, are classified as ground roll and considered as noise on the recorded exploration data.

Love waves are transverse horizontal waves that represent particle motion at right angles to the source direction and in a horizontal plane. They travel within layers on the earth's surface, so they are classified as surface waves. Their propagation velocity is in the range of 3 km/sec to 4.5 km/sec.

Rayleigh waves represent particle motion in an elliptical orbit in a vertical plane through the axis of propagation (the so-called *plane of propagation*). They travel along the earth's surface, so they are classified as surface waves. Their propagation velocity is in the range of 2.5 km/sec to 4 km/sec.

Elastic waves transfer energy in a solid from one point to another by the movements of small particles. The particles themselves are not transported, but vibrate with small displacements about their respective equilibrium positions. Their movements propagate energy in the form of a wave. A basic assumption is that the displacements never violate the elastic limit of the solid. The transfer of energy by an elastic wave in a solid, such as rock or steel, involves many millions of particles. The dimensions of the elastic wave, such as its amplitude and wavelength, are large relative to the dimensions of the vibrations of the particles. In fluids, such propagating waves are called *acoustic waves*. Our speech is transported by acoustic waves in the air; in this case, the particles are air molecules. The elastic waves that propagate in the solid earth, as well as the acoustic waves that propagate in the sea and other bodies of water, are known as *seismic waves*.

Most seismic interpretation is conducted in terms of models involving P-wave propagation through a sequence of rock formations. It is commonplace to make the assumption that each formation is homogeneous and isotropic. A seismic wave, as well as other types of waves, represents an interchange between kinetic and potential energy. The kinetic energy comes from the physical motion of the particles, and the potential energy comes from their relative positions with respect to the inter-particle elastic (or restoring) forces. The rest position of each particle can be taken as the origin of the coordinate system describing that particle. The particle vibrates around its rest position. At any instant, we can, in principle, measure the particle's velocity and also, if we wish, the particle's acceleration. In addition, we can measure the degree of compression of the particles in the form of a pressure or stress. In the case of a homogeneous isotropic solid, physical theory tells us that we need only two quantities for a complete specification of the wave motion resulting from the particle motion. In seismic work, it is convenient to take for these two quantities the *particle velocity* and the *pressure*.

The simplest wave form is a sinusoidal curve. These are known as sinusoidal waves, or by the older terminology as simple harmonic waves. By Fourier's theorem, any periodic signal can be synthesized by a superposition of sinusoidal waves. A sinusoidal wave is a periodic function, that is, a wave train that involves a succession of crests and troughs of the same shape. When we do not have sinusoidal waves, but instead have some complicated wave motion (which is usually the case), then we must perform a Fourier analysis to determine the spectral distribution (or spectrum) of the wavenumbers and frequencies present.

In seismic exploration, innumerable situations occur in which we can picture a single isolated pulse of a seismic disturbance as propagating from one place to another. Pulses of this sort can be set up by taking a stretched string and producing in it a local deformation, by pulling one end and then holding it still. The subsequent pulse travels at a constant speed. At any instant, only a limited region of the string is disturbed. The regions before and after are quiescent. The pulse (a.k.a. wavelet) travels in this way until it reaches the far end of the string, at which point it is reflected. As long as the pulse continues uninterrupted, however, it preserves the same shape. How can we relate the behavior of a wavelet to what we have already learned of sinusoidal waves? The answer is obtained by Fourier analysis, but in terms of the Fourier integral instead of a Fourier series.

One way to look at the wave motion is to take a slice through the earth and take a "snapshot" at one instant of time. We see waveforms, and the spatial distances between peaks are indicative of the wavelengths λ. As shown in Figure 10, the horizontal axis measures distance from the shot. This figure shows that, for a fixed instant of time, the high amplitude has traveled the farthest.

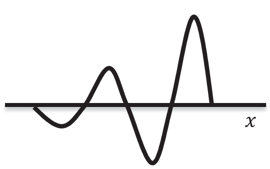


Figure 10. Wavelet at one instant of time as a function of distance.

Another way to look at waves is to stay at one point within the earth and record the wave motion over time. Again we see waveforms, and the time distances between peaks are indicative of the periods T. As shown in Figure 11, in this case the horizontal axis measures vt, where v is a constant velocity. Thus, the horizontal axis has the dimensions of distance. The wavelet shown appears as the reverse of the wave profile shown in Figure 10. This figure shows that, for a fixed spatial point, the high amplitude appears first.

The reciprocal of the wavelength is called the wavenumber κ; the reciprocal of the period is called the frequency f. Often we work with sines and cosines, so it is convenient to use instead the radian wavenumber k and radian frequency ω, which are respectively

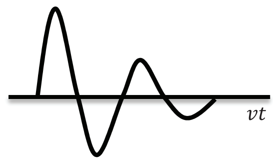


Figure 11. Wavelet at one point of space as a function of time.

$$k = 2\pi\kappa = \frac{2\pi}{\lambda}\ , \quad \omega = 2\pi f = \frac{2\pi}{T}\ .$$

Frequently in exploration seismology, we are concerned with P-wave propagation in layers of homogeneous, isotropic rocks. Such a rock may be described by two physical parameters: density ρ and velocity v of propagation. The velocity v is called by various names, such as the *wave velocity*, the *propagation velocity*, or the *material velocity*. Most often, it simply is called *velocity*. These parameters, ρ and v, can be measured in an oil well by borehole instruments.

We emphasize the distinction between particle velocity (abbreviated p. v.) and wave velocity. Wave velocity (more precisely, group velocity) represents the speed at which seismic energy is transported through the body of the rock. Typical wave velocities in rock formations are on the order of thousands of meters per second. Particle velocity represents the speed of the very small rock particles about their stationary equilibrium positions. A typical particle velocity is expressed in terms of millionths of a meter per second. To conceptualize some idea of particle velocity, we can consider the base-plate of the mechanical vibrator used in the vibroseis method. The displacement of the base-plate is no more than 1 or 2 cm, so the particle velocity imparted to the earth at a 10 Hz frequency is on the order of 1 m/s. This velocity is entirely different, both in concept and in magnitude, from the wave velocity of the waves leaving the base-plate. The wave velocity would be on the order of 1000 m/s, or more. The geophones placed on the surface of the ground pick up and record the particle velocity, which, when amplified, appears as the seismic trace. These received signals have much smaller particle velocities than the source signals. The direct wave from a dynamite source might have a particle velocity of 0.001 m/s, whereas the wave corresponding to a deep reflection might have a particle velocity of 10^{-8} m/s.

We have characterized the physical properties of a homogeneous isotropic rock formation by its density and velocity. Another important parameter is the *acoustic impedance* (a.k.a. *characteristic impedance*) Z. It is defined as the product of density ρ times velocity v; that is, the acoustic impedance is $Z = \rho v$. To convey a physical feeling for the acoustic impedance, Nigel Anstey calls it acoustic hardness. This is not exactly the same as what the geologist refers to as hardness, but it is still a useful characterization. Suppose, for example, that we have soft shale. Such a rock would have a low velocity v, so its acoustic impedance, or hardness, would be small. Hard limestone, however, would have a high velocity, and so its acoustic impedance, or hardness, would be large.

An analogy might be helpful. The best known one is the electric analogy. In electrical work, we have current (amps), voltage (volts), and resistance (ohms). Ohm's law states:

$$\text{voltage} = +(\text{current})(\text{resistance}) \ .$$

Thus, for a given voltage, we would have a large current in a material of low resistance (a conductor, such as copper) and a small current in a material of high resistance (a dielectric, such as wood). We now make the analogy:

$$\text{particle velocity} \sim \text{current} \ ,$$

$$\text{pressure} \sim \text{voltage} \ ,$$

$$\text{acoustic impedance} \sim \text{resistance} \ .$$

Corresponding to Ohm's law then, we have the equation

$$\text{pressure} = +(\text{particle velocity})(\text{acoustic impedance}) \ .$$

Thus, for a given pressure, we would have a large particle velocity in a material of low acoustic impedance (soft shale) and a small particle velocity in a material of high acoustic impedance (hard limestone).

Just as an electric wave may be described in terms of either its voltage or its current, so a seismic wave may be described in terms of either its particle velocity or its pressure. A seismic trace is a graph of amplitude versus time. In marine work, the hydrophone measures pressure, so the amplitude of a marine seismic trace is in terms of pressure. In land work, the geophone measures particle velocity, so the amplitude of a land seismic trace is in terms of particle velocity.

The objective of seismic exploration is to delineate the subsurface of the earth. An important development in the acquisition of seismic data is the use of dual sensors. A *dual sensor* is composed of a hydrophone and a geophone.

The output of a dual signal is *pressure* as measured by the hydrophone and *particle velocity* (p. v.) as measured by the geophone.

The definition of the *Fresnel reflection coefficient* requires the consideration of both the particle velocity and pressure. In seismic work, the reflection coefficient of an interface is found by requiring that

a) particle velocity be continuous across the interface and
b) pressure be continuous across the interface.

Downgoing and upgoing waves transport energy to and from reflecting horizons. Particle velocity and pressure determine the partition of energy at each horizon. Thus, in seismic exploration, there is interplay of two dualities, namely the duality of the upgoing and downgoing waves and the duality of particle velocity and pressure. Seismic exploration is concerned with traveling seismic waves. However, a traveling wave is never recorded as such in seismic acquisition. The signal recorded by a geophone is the sum of the downgoing and upgoing particle-velocity waves. Given the geophone data alone, the separation of the component downgoing and upgoing waves cannot be done unless some type of traveling-wave assumption is made. The signal recorded by a hydrophone is the sum of the downgoing and upgoing pressure waves. Given the hydrophone data alone, the separation of the component downgoing and upgoing waves cannot be done unless a traveling-wave assumption is made.

The wavefield is the sum of two components. One component is the downgoing wave and the other component is the upgoing wave. Furthermore, there are two wave equations of interest, one is the wave equation for particle velocity and the other is the wave equation for pressure. The numbers defining the layers increase with increasing depth. Denote the density, or mass per unit volume, of layer i by ρ_i. Denote the propagation velocity in layer i by v_i. The product $Z_i = \rho_i v_i$ is the acoustic impedance of layer i. In layer i, denote particle velocity (p. v.) by V_i, pressure by p_i. Let D_i denote the downgoing component of particle velocity and let U_i denote the upgoing component of particle velocity. Similarly, let d_i denote the downgoing component of pressure and let u_i denote the upgoing component of pressure. Thus, we have the two equations

$$V_i = D_i + U_i \quad \text{and} \quad p_i = d_i + u_i \;, \tag{3}$$

where the first equation is for particle velocity (p. v.) and the second equation is for pressure.

There are two different conventions used in addressing pressure waves and particle-velocity waves. One is the Anstey (1977) convention and the other is the Berkhout (1987) convention. Each has its merits, but we shall use the Berkhout convention. The two types of traveling waves are pressure waves and particle velocity (p. v.) waves.

In the **Anstey Convention** (not used in this book): downgoing pressure and p. v. waves are 180° out-of-phase, and upgoing pressure and p. v. waves are in-phase. The equations are

$$d_i = -Z_i D_i \quad \text{and} \quad u_i = Z_i U_i \ .$$

In the **Berkhout Convention** (used in this book): downgoing pressure and p. v. waves are in-phase, and upgoing pressure and p. v. waves are 180° out-of-phase. The equations are

$$d_i = Z_i D_i \quad \text{and} \quad u_i = -Z_i U_i \ . \tag{4}$$

The first equation in expression 4 says that the downgoing pressure wave has the same polarity as the downgoing particle-velocity wave and that the two are related by a scale factor given by the acoustic impedance. The second equation in expression 4 says that the upgoing pressure wave has the opposite polarity as the upgoing particle-velocity wave and that the two also are related by the same scale factor.

Let us give an example of a traveling-wave assumption. Figure 12 shows an event. Looking at the event, can we tell whether the event is upgoing or downgoing? The answer is no. The geophone signal by itself is not enough to determine whether the event is upgoing or downgoing. Suppose, however, that we make the assumption that the event is a primary reflection coming from depth. Such an event is upgoing.

Next Figure 13 represents the use of a dual sensor, so that both the geophone and the hydrophone signal are recorded. Neither the geophone signal by itself, nor the hydrophone signal by itself, is enough to determine whether the event is upgoing or downgoing. However, the two signals are out of phase, so (by the Berkhout convention used in this book) it can be concluded that the event is upgoing. This result is the essence of the so-called d'Alembert equations (Robinson, 1999).

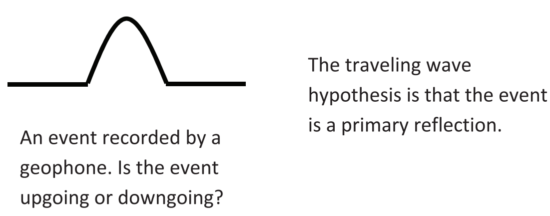
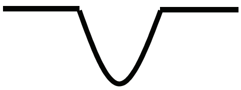


Figure 12. An event that requires a traveling-wave assumption.

An event recorded by a geophone. Is the event upgoing or downgoing?



The same event recorded by a hydrophone. Is the event upgoing or downgoing?

Figure 13. An event that does not require a traveling-wave assumption.

Land surveys with geophones cannot be tied to marine surveys with hydrophones without the linkage provided by a traveling-wave assumption. In fact, deconvolution, migration, and nearly every other method of seismic processing require representations of traveling waves. The net result is that seismic processing simply cannot be done with conventionally recorded data unless one is willing to make traveling-wave assumptions. Does conventional seismic processing face a serious impasse if one is unwilling to make such assumptions? The answer is that all traveling-wave assumptions can be avoided. Over the past 30 years, remarkable advances have been made in seismic instrumentation. One such advance is the dual sensor. It yields an output of both *pressure* as measured by a hydrophone and *particle velocity* (p. v.) as measured by a geophone. In such a case, the downgoing and the upgoing traveling waves can be computed directly from the data by means of the d'Alembert equations.

In terms of the electrical analogy we mentioned previously, seismic particle velocity corresponds to electrical current, seismic pressure corresponds to electrical voltage, and the acoustic impedance of the rock corresponds to resistance. The equation that pressure is equal to the product of particle velocity and characteristic impedance corresponds to the well known equation that voltage is equal to the product of current and resistance. In order to install an electric wire, an electrician must know both current and voltage. For the same reason, a geophysicist should know both particle velocity and pressure.

Let the dual sensor be placed at the top of layer i. The dual sensor measures both particle velocity disturbance and pressure disturbance at that depth. Equations 3 and 4 may be solved for the downgoing and upgoing particle-velocity waves as follows. Substituting

$$d_i = Z_i D_i \quad \text{and} \quad u_i = -Z_i U_i \quad \text{into } p_i = d_i + u_i,$$

we obtain

$$p_i = Z_i D_i - Z_i U_i.$$

Thus, equations 3 become

$$V_i = D_i + U_i \quad \text{and} \quad p_i = Z_i D_i - Z_i U_i \ .$$

These equations yield the particle velocity p. v. form of the d'Alembert equations,

$$D_i = \frac{V_i + p_i/Z_i}{2} \quad \text{and} \quad U_i = \frac{V_i - p_i/Z_i}{2} \ . \tag{5}$$

The d'Alembert equations are fundamental. For a dual sensor planted in a given layer, the geophone records the particle-velocity trace V_i and the hydrophone records the pressure trace p_i. If the acoustic impedance of that layer is known, the d'Alembert equations 5 can be used to find the downgoing component D_i and the upgoing component U_i of the particle-velocity trace. In the d'Alembert equations 5, the inputs are the particle velocity V_i, pressure p_i, and acoustic impedance Z_i, all measured at the receiver. The outputs are the downgoing wave D_i and the upgoing wave U_i, both occurring at the receiver. In this form of the d'Alembert equations, the downgoing wave and the upgoing wave are each in terms of particle velocity (p. v.).

In this book, we address the particle-velocity components D_i and U_i. We shall not make use of the corresponding d'Alembert equations for the two components d_i and u_i of the pressure trace, which are

$$d_i = \frac{Z_i V_i + p_i}{2} \quad \text{and} \quad u_i = \frac{-Z_i V_i + p_i}{2} \ . \tag{6}$$

It is important to have standards that are followed. Because so many seismic sections are traded and sold, and because one geographic area may require the use of geophysical data recorded and processed by several companies, it is important that everyone follow the same conventions. One such convention is polarity, which refers to which direction is used as the positive direction for seismic amplitude. When two seismic lines intersect, we want the reflection observed on the near traces (i.e., the small offset traces) to be identical. For large offsets, the same event may be due to reflections from different depth points because of the dips of the reflecting interfaces. Thus, we should not expect far traces to have identical reflections. If one company displays positive amplitudes in one direction and another company in the opposite direction, a great deal of confusion can result. One convention is:

1. A positive seismic signal is defined as a positive acoustic pressure on a hydrophone or a downward motion on a geophone.
2. A positive seismic signal produces a positive number on tape, a positive deflection (upswing) on a monitor record, and a peak (black area) on a seismic section.

Now we will consider an explosion in a shot hole buried below the surface of the earth or, alternatively, an explosion below the surface of the water. As shown in Figure 14, the resulting downgoing signal will have a positive particle velocity and positive pressure. The upgoing signal from the explosion will have a negative particle velocity but a positive pressure.

Consider the process of the reflection of the downward-traveling seismic pulse. Once the seismic pulse leaves the immediate area of the shot, we can assume that the seismic pulse, now of a definite form, propagates through the earth and reflects from rock interfaces in accordance with the linear laws of elastic wave propagation.

Next we want to consider essentially a one-dimensional problem, in which we treat only variations along the depth axis z, which points straight down into the earth. Thus, we disregard the generally spherical form of the wavefront and treat the problem as that of a plane wave traveling vertically downward in a direction perpendicular to the flat-lying (horizontal) rock strata. Suppose at some instant a wave is traveling in a rock layer of density ρ_1 and seismic wave velocity v_1, and then the wave encounters a different rock layer of density ρ_2 and velocity v_2. The result is that a portion of the energy in the wave will be reflected at the interface and the remainder will be transmitted. This splitting of the incident wave into a reflected wave and a transmitted wave at the interface is caused by the abrupt change in rock density and/or velocity. For the case of normal incidence (which we are addressing here), the reflected and transmitted waves have identical shape and breadth as the incident wave, but they differ in

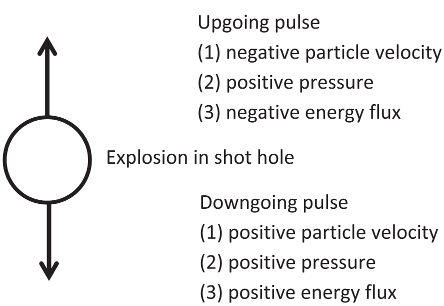


Figure 14. Explosion in a buried shot hole.

amplitude. The ratio of the amplitude of the reflected wave to that of the incident wave is termed the *reflection coefficient*. Similarly, the ratio of the amplitude of the transmitted wave to that of the incident wave is called the *transmission coefficient*. However, the reflection coefficient defined for a wave in which the amplitude is measured in terms of particle velocity is different from the reflection coefficient in terms of pressure. The same statement also holds for the transmission coefficient. Let us now find this difference.

Suppose that a wave propagating in shale comes to an interface with a limestone layer. Because the rocks do not rupture at the interface, the particle velocity on both sides of the interface must be the same; that is, at the interface, the particle velocity on the shale side must be the same at all times as the particle velocity on the limestone side. We express this physical fact in mathematical terms by the statement that particle velocity must be continuous at an interface. Similarly, pressure must be continuous at an interface.

An explosion produces a large downgoing pulse, and we assign to it a positive amplitude in the case of particle velocity and a positive amplitude in the case of pressure. Suppose the explosion is set off in soft shale ($Z_1 = \rho_1 v_1$ is small). The source pulse travels down and hits hard limestone ($Z_2 = \rho_2 v_2$ is large). Referring to Figure 15, the incident pulse has a fixed amount of positive energy flux, given by the product of its positive particle velocity times its positive pressure. Because the soft shale has a low impedance, the particle velocity is relatively large in magnitude, and its pressure is correspondingly small. At the interface, some of the incident energy is transmitted into the limestone and the rest is reflected back into the shale. The total particle velocity must be the same on both sides of the interface. The hard limestone, because of its higher impedance, cannot take up all of the particle velocity present in the soft shale. Hence, only some of the particle velocity will be transmitted into the limestone. To keep the bookkeeping straight (i.e., continuity of particle velocity), a reflected wave must be thrown upward, back into the shale. The particle velocity of this reflected wave must be a negative quantity, so as to cancel out some of the incident

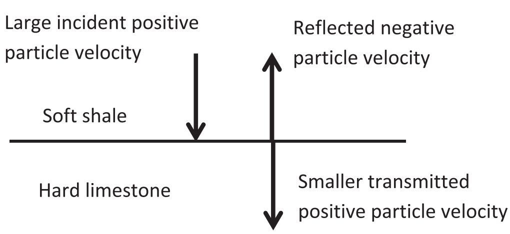


Figure 15. Continuity of particle velocity at a soft-to-hard interface. The arrows indicate the raypath direction, i.e., the direction of travel.

positive particle velocity; that is, we have the bookkeeping entry (where p. v. is particle velocity)

$$\text{(large incident positive p. v.)} + \text{(reflected negative p. v.)}$$
$$= \text{(smaller transmitted positive p. v.)}\ .$$

Thus, the particle velocity reflection coefficient is negative, and the particle velocity transmission coefficient is positive.

Next let us look at pressure and refer to Figure 16. Because the soft shale has a low impedance, the downgoing incident pressure is relatively small. By continuity, the pressure must be the same on both sides of the interface. The limestone is hard and can support more pressure than the soft shale. Thus, the amplitude of the transmitted pressure wave is greater than the amplitude of the incident pressure wave. To keep the bookkeeping straight (i.e., continuity of pressure), a reflected wave must be thrown upward, back into the shale. The pressure of this reflected wave must be a positive quantity, so as to augment the incident pressure; that is, we have the bookkeeping entry

$$\text{(small incident positive pressure)} + \text{(reflected positive pressure)}$$
$$= \text{(larger transmitted positive pressure)}\ .$$

Thus, the pressure reflection coefficient is positive, and the pressure transmission coefficient is also positive.

Remember that we are only going to consider particle velocity (p. v.) traveling waves and not pressure traveling waves. Otherwise we would get bogged down in too much notation to retain the reader's interest. The air is layer 1. The ground surface is interface 1. Below is layer 2. In general, interface i is between layer i above and layer $i+1$ below. The density and propagation velocity of a given layer carry the number of the layer as a subscript. The Fresnel coefficients carry the number of the interface as a subscript. Suppose a downgoing particle-velocity unit impulse in

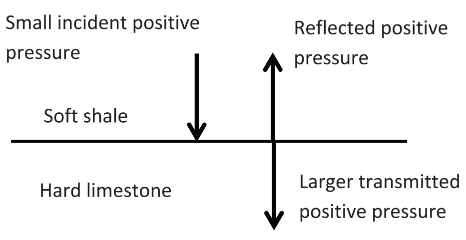
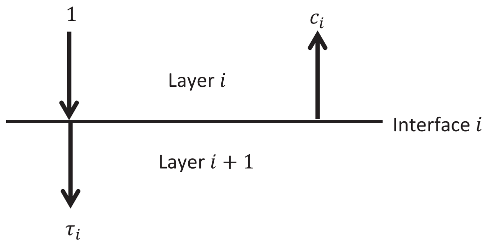


Figure 16. Continuity of pressure at a soft-to-hard interface. The arrows indicate the raypath direction.

Figure 17. The Fresnel reflection coefficient and the transmission coefficient.



layer i strikes the interface i. By definition, the resulting upgoing particle-velocity impulse in layer i is the particle-velocity Fresnel reflection coefficient r_i. Also, by definition, the resulting downgoing particle-velocity impulse in layer $i+1$ is the particle-velocity Fresnel transmission coefficient τ_1.

Figure 17 shows one interface, e.g., interface i, in isolation. Layer i lies above the interface and layer $i+1$ below. When a downgoing wave strikes the interface, part of the energy is reflected back into the same layer and the rest of the energy is transmitted into the next layer. Both the particle velocity and the pressure of the wave motion must be continuous across the interface, as expressed by the equations

$$D_i + U_i = D_{i+1} \quad \text{and} \quad d_i + u_i = d_{i+1} \ . \tag{7}$$

If equations 4 are used, then equations 7 may be written as

$$D_i + U_i = D_{i+1} \quad \text{and} \quad Z_i D_i - Z_i U_i = Z_{i+1} D_{i+1} \ . \tag{8}$$

The particle-velocity Fresnel reflection coefficient c_i and the particle-velocity Fresnel transmission coefficient τ_i are defined as

$$c_i = \frac{U_i}{D_i} \quad \text{and} \quad \tau_i = \frac{D_{i+1}}{D_i} \ . \tag{9}$$

The solution of equations 8 and 9 obtains the Fresnel coefficients in terms of the acoustic impedances of the two layers; that is,

$$c_i = \frac{Z_i - Z_{i+1}}{Z_i + Z_{i+1}} \quad \text{and} \quad \tau_i = \frac{2Z_i}{Z_i + Z_{i+1}} = 1 + c_i \ . \tag{10a}$$

The reflection coefficient and the transmission coefficient as given by equation 10a are called Fresnel coefficients because they only address the interface between the two layers in question, and not the entire system which can have many layers. Each Fresnel coefficient carries the number of the interface as a subscript. The Fresnel coefficients in equation 10a are the coefficients for an incident downgoing particle-velocity wave striking the interface from above. It is apparaent that the reflection coefficient must be less than one in magnitude. The corresponding Fresnel coefficients for a downgoing incident pressure wave can be obtained in a similar manner. The pressure Fresnel reflection coefficients are of reverse sign to the particle-velocity Fresnel reflection coefficients.

Consider next an incident upgoing particle-velocity wave in layer $i+1$ striking interface i from below. The resulting Fresnel reflection and transmission coefficients carry a prime, and are given by

$$c_i' = -c_i \quad \text{and} \quad \tau_i' = 1 + c_i' = 1 - c_i = \frac{Z_{i+1}}{Z_i}\,\tau_i = \frac{2Z_{i+1}}{Z_i + Z_{i+1}} \ . \qquad (10b)$$

The two-way transmission coefficient through interface i (i.e., passage in one direction followed by passage in the other direction) is

$$\tau_i\,\tau_i' = (1 + c_i)(1 - c_i) = 1 - c_i^2 \ . \qquad (11)$$

In the 17th century, Galileo, Huygens, and Newton formulate the foundations of mechanics. The definition of velocity is an essential element in that formulation. Velocity is a ratio. It is distance over time or, in the limit, instantaneous distance over instantaneous time. In the late 19th century, the French mathematician Henri Poincare reaches the conclusion that the velocity of light is the ultimate signal velocity in the universe. No signal can ever be transmitted at a velocity greater than that of light (300,000 km/sec in round numbers). Thus, no physical object, whether a subatomic particle or an entire galaxy, can travel at a speed greater than 300,000 km/sec. In astronomy, it is usual to measure distances in terms of the velocity of light; for example, one light-year is the distance which light travels in one year. Similarly, one light-second is the distance light travels in one second. In other words, one light-second is the distance 300,000 km. Let us now use the light-second as our distance unit. Light has a velocity of one light-second per second. A light-second is called a *natural unit* because the velocity of light, a basic physical quantity, has the value of unity. In terms of natural units, Poincare's conclusion is that no physical velocity can be greater than one in magnitude—that is, greater than the speed of light.

What are the implications of the conclusion that no physical velocity can be greater than one? Suppose we point a telescope in a given direction at the sky. We see two galaxies E and B. By means of the Doppler redshift, we can measure the velocity $v_E = 0.5$ with which E is receding from us and the velocity $v_B = 0.7$ with which B is receding from us. These velocities are given in natural units (in which the velocity of light is one). What is the velocity v_F at which galaxy B is moving away from E? Einstein says that the simple subtraction formula

$$v_F = v_B - v_E = 0.7 - 0.5 = 0.2$$

does not work. Instead we must use the Einstein subtraction formula

$$v_F = \frac{v_B - v_E}{1 - v_B v_E} = \frac{0.7 - 0.5}{1 - 0.7 \times 0.5} = 0.3077\ . \tag{12}$$

Now let us apply this thinking to geophysics. Choose a system with n interfaces, numbered

$$1, 2, 3, \ldots, \alpha - 1, \alpha, \alpha + 1, \ldots, n\ . \tag{13}$$

The Fresnel reflection coefficients of the interfaces are given by $c_1, c_2, \ldots, c_n$. Interface 1 is the surface of the ground or the surface of the water as the case may be. The lowermost interface is interface n. All of the material below interface n is referred to as basement rock. The basement is one-way in that it accepts downgoing energy but returns no reflected energy.

In Figure 18, the source is in layer α, which is the layer between interface $\alpha - 1$ and interface α. Layer α is called the source layer. Partition the given system into two component systems, denoted by A and B. System A contains all of the interfaces above the source. System B contains all of the interfaces below the source and above the basement rock. System A has the sequence of reflection coefficients $c_1, c_2, \ldots, c_{\alpha-1}$. System B has the sequence of reflection coefficients $c_\alpha, c_{\alpha+1}, \ldots, c_n$.

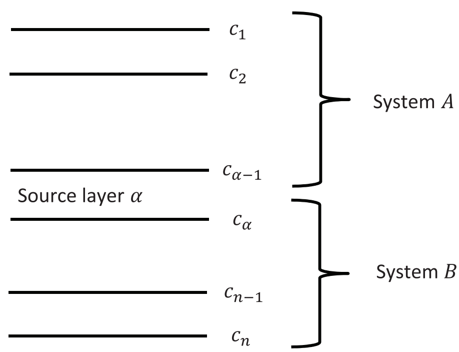


Figure 18. The two component systems separated by the source layer.

When excited, the source produces a downgoing particle-velocity signal S as well as an upgoing particle-velocity signal $-S$. System A has the system reflection coefficient R'_A for waves striking the system from below. System B has the system reflection coefficient R_B for waves striking the system from above.

In Figure 19, the receiver (a dual sensor) is in layer β. Assume that the receiver is strictly below the source, so $\beta > \alpha$. Recall that system B contains all of the interfaces below the source; that is, system B has the series of Fresnel reflection coefficients $c_\alpha, c_{\alpha+1}, \ldots, c_n$. Partition system B into two component systems, where the upper system E is comprised of the interfaces between source and receiver, and the lower system F is comprised of the interfaces between receiver and basement rock. Thus, system E has Fresnel reflection coefficients $c_\alpha, c_{\alpha+1}, \ldots, c_{\beta-1}$ and system F has Fresnel reflection coefficients $c_\beta, c_{\beta+1}, \ldots, c_n$.

As we have seen, system B has the system reflection coefficient R_B. Let system E have the system reflection coefficients R_E for waves striking the system from above and R'_E for waves striking the system from below. System F has the system reflection coefficient R_F for waves striking the system from above.

From the dual sensor data, we can use the d'Alembert equations 5 to compute the downgoing wave D_β and upgoing wave U_β in the receiver layer. They are related by

$$U_\beta = R_F D_\beta \ . \tag{14}$$

Equation 14 may be written as

$$R_F = \frac{U_\beta}{D_\beta} \ . \tag{15}$$

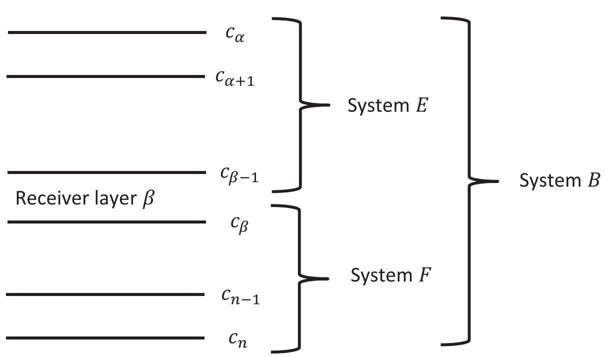


Figure 19. The two component systems comprising system B with the separation at the receiver layer.

Einstein deconvolution uses this equation to compute R_F. We contend that R_F as computed contains only information from the isolated system F; that is, R_F contains no information at all from the overlying systems A and E. We claim that we have stripped away the influence of all of the layers overlying the receiver layer β. The source wavelet and all of the reverberations and ghosts from these upper layers are entirely gone.

System B has the system reflection coefficient R_B. The coefficient R_B involves only the Fresnel coefficients in

$$\text{sequence}\, B = c_\alpha, c_{\alpha+1}, \ldots, c_{\beta-1}, c_\beta, c_{\beta+1}, \ldots, c_n \ .$$

System E has the system reflection coefficient R_E. This coefficient R_E involves only the Fresnel coefficients in

$$\text{sequence}\, E = c_\alpha, c_{\alpha+1}, \ldots, c_{\beta-1} \ .$$

System F has the system reflection coefficient R_F. The reflection coefficient R_F contains only the Fresnel coefficients in

$$\text{sequence}\, F = c_\beta, c_{\beta+1}, \ldots, c_n \ .$$

Certainly,

$$\text{sequence}\, F = \text{sequence}\, B - \text{sequence}\, E \ .$$

Can we conclude that

$$R_F = R_B - R_E \ ?$$

The answer is no! We must use the Einstein subtraction formula

$$R_F = \frac{R_B - R_E}{\Omega_E + R_B R_E'} \ . \tag{16}$$

The novelty of this expression rests in the appearance of the all-pass system Ω_E, which involves only the sequence E. An all-pass system adjusts the phase without changing the amplitude spectra. This all-pass system takes the place of the "1" appearing in the Einstein subtraction formula 12 for velocities.

System F, namely, the system between the receiver and the basement rock, contains the reflection coefficients of interest in exploration. The near-surface reflection coefficients, that is, those of the interfaces above the receiver, give rise to the reverberations and ghosts that we want to eliminate. The dual sensor measures both the particle velocity signal V_β and the pressure signal p_β in layer β. In addition, the acoustic impedance $p_\beta C_\beta$ of the receiver layer must be obtained or estimated.

The Einstein deconvolution method can be described in two steps. The first step is to convert the particle velocity and pressure signals into the downgoing wave D_β and the upgoing wave U_β in layer β. In order to complete this step, the d'Alembert equations 5 are used. The second step is to deconvolve the upgoing wave by the downgoing wave as shown by equation 15. The result of Einstein deconvolution is the unit-impulse reflection response R_F of system F.

In summary, the deconvolution of the upgoing wave U_β by the downgoing wave D_β yields the unit-impulse reflection response R_F of the subsystem below the receiver. The Einstein deconvolution process strips away the multiples and ghosts caused by the upper system. It should be emphasized that the Einstein deconvolution process also strips away the unknown source signature wavelet S.

Let us now look at the mechanisms for ghosts and reverberations in more detail. Robinson (1966) provides two geophysical processing methods described in terms of the mathematics of the Z-transform (more precisely, the generating function). For example, the generating function (or Z-transform) of the digital signal $h_0, h_1, h_0, \ldots$ is defined as

$$H(Z) = h_0 + h_1 Z + h_2 Z^2 + \cdots \ .$$

On the other hand, electrical engineers use the z-transform

$$H(z) = h_0 + h_1 z^{-1} + h_2 z^{-2} + \cdots \ ,$$

which we never use.

One method is the method for the elimination of seismic ghost reflections and the other method is the method for the elimination of water-confined reverberations. The methods are strictly valid only for flat horizontal interfaces and for vertical incidence. The ghost-producing filter is given by

$$G = 1 - c'_A Z^n \ , \tag{17}$$

where c'_A is a constant (of magnitude less than unity) representing the Fresnel reflection coefficient of the overlying interface for an upgoing incident wave, and n is a constant (assumed to be an integer) representing the time delay of the ghost with respect to the primary.

The water reverberation problem in marine seismic operations may be described as follows. The water–air interface is a strong reflector. Let the water-bottom interface also be a strong reflector. In such a case, the water layer approximates a non-attenuating medium bounded by two strong reflecting interfaces and hence represents an energy trap. A seismic pulse generated in this energy trap will be successively reflected between the two interfaces. Consequently, reflections from deep horizons below the water layer will be obscured by the water reverberations.

Let c_B be the Fresnel reflection coefficient for a downgoing incident particle-velocity signal striking the water bottom, and let c'_A be the Fresnel reflection coefficient for an upgoing incident particle-velocity signal striking the water surface. The two-way traveltime in the water layer is n time units. Let the source be a downgoing unit spike. The reverberation-producing filter is given by

$$\Lambda = \frac{1}{1 - c_B c'_A Z^n} \ . \tag{18}$$

This expression also holds for an upgoing unit spike source.

The ghost-producing filter G for systems can be obtained from equation 17 by replacing $c'_A Z^n$ with R'_A. The result is

$$G = 1 - R'_A \ . \tag{19}$$

The source produces reverberations in the source layer, that is, in layer α. The mathematical structure is the same as given for the case of reverberations between two interfaces. However, now system coefficients must be used instead of Fresnel coefficients. Thus, the system reflection coefficient R'_A is used for the layers above the source. Likewise, the system reflection coefficient R_B is used for the layers below the source. The reverberation-producing filter can be obtained from equation 18 by replacing $c_B c'_A Z^n$ with $R_B R'_A$. Thus, the reverberation-producing filter for this case is

$$\Lambda = \frac{1}{1 - R_B R'_A} \ . \tag{20}$$

The downgoing particle-velocity wave D_α in the source layer is composed of the source signature S, the ghost G, and the reverberation Λ. Thus,

$$D_\alpha = SG\,\Lambda = \frac{S(1 - R'_A)}{1 - R_B R'_A} = S\Lambda - SR'_A\,\Lambda \ . \tag{21}$$

The direct upgoing wave in the source layer due to the source is

$$U_\alpha^{\text{direct}} = -S \ . \tag{22}$$

The downgoing wave D_α is reflected from System B. The reflection coefficient is R_B. The upgoing reflected wave appears as $R_B D_\alpha$ in the source layer. This reflected wave is given by

$$U_\alpha^{\text{reflected}} = R_B D_\alpha = S R_B G \Lambda \ . \tag{23}$$

The entire upgoing wave is the sum of the direct upgoing wave, shown in equation 22, plus the reflected upgoing wave, shown in equation 23; that is,

$$U_\alpha = U_\alpha^{\text{direct}} + U_\alpha^{\text{reflected}} = -S + S R_B G \Lambda \ . \tag{24}$$

See Figure 20. The above expression becomes

$$\begin{aligned} U_\alpha &= -S + S R_B (1 - R'_A)\Lambda = -S(1 + R_B R'_A \Lambda) + S R_B \Lambda \\ &= -S \frac{1 - R_B R'_A + R_B R'_A}{1 - R_B R'_A} + S R_B \Lambda = -S\Lambda + S R_B \Lambda \ . \end{aligned}$$

As geophysicists began to study the earth by using quantitative physical methods, they visualized the underground in modest and idealistic terms. The source and the receivers were on or very close to the surface. The source produced a downgoing wave and the receiver recorded an upgoing wave. This ideal model did not include effects such as reverberations and ghost reflections, about which geophysicists had little knowledge. However, as time went on, geophysicists became aware of the great difficulties presented by these effects. With dual sensors and seismic processing, reverberations and ghost reflections above a buried receiver could be stripped away. The resulting seismic data approximated the ideal model of the past, with the datum no longer being at the surface of the ground, but at the depth of the receiver.

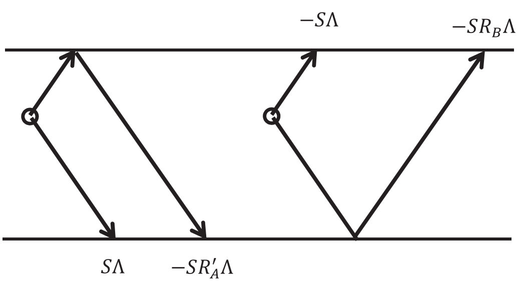


Figure 20. (left) Components of D_α and (right) components of U_α.

Let us now discuss *seismic inversion by dynamic deconvolution* (Robinson, 1975). The Lorentz transform (used in relativity theory) can be written as

$$
\begin{aligned}
t_2 &= \frac{1}{(1-c_1^2)^{1/2}}(t_1 - c_1 x_1) \ , \\
x_2 &= \frac{1}{(1-c_1^2)^{1/2}}(-c_1 t_1 + x_1) \ .
\end{aligned}
\tag{25}
$$

In the Lorentz equations:

a) The variables t_1 and x_1 are, respectively, the time and space coordinates of an event in frame 1.
b) The variables t_2 and x_2 are, respectively, the time and space coordinates of the event in frame 2.
c) The constant c_1 (where $|c_1| < 1$) is the velocity (in natural units such that the velocity of light is unity) between the two frames.

The Lorentz transformation is a consequence of the invariance of the interval between events. By direct substitution, it can be shown that the coordinates of two events satisfy the equation

$$t_2^2 - x_2^2 = t_1^2 - x_1^2 \ .$$

Now we want to find the relationship between the waves in the seismic layered model, shown in Figure 21. As usual, all of the waves are digitized with unit time spacing. For a downward wave incident on interface i, the reflection coefficient is denoted by c_i. Then, for an upgoing wave incident on interface i, the reflection coefficient is $-c_i$. We will assume that the amplitudes of the seismic waves are measured in units such that the squared amplitude is proportional to energy. Then, for either downgoing

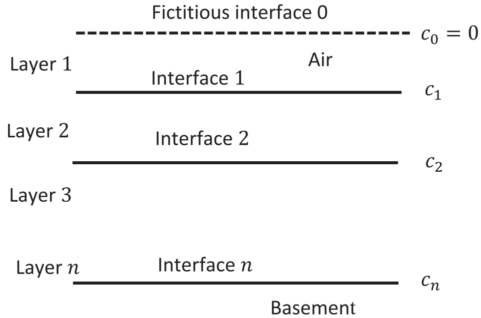


Figure 21. The layered system. It has n interfaces, with air above interface 1 and basement below interface n. The seismic one-way traveltime between adjacent interfaces is 0.5 time unit, so the two-way traveltime is one time unit.

or upgoing waves, the transmission coefficient through interface i is

$$\tau_i = +\sqrt{1 - c_i^2} \ .$$

Instead of the conventional treatment, we will try to put this relationship into a more general setting. We know that the waves in each layer obey their respective wave equation.

Referring to Figure 22, let $D_1(Z)$ and $U_1(Z)$ be, respectively, the generating functions of the downgoing wave and the upgoing wave at the top of layer 1. Let $D_2(Z)$ and $U_2(Z)$ be the corresponding functions for layer 2. The symbol Z represents a unit delay operator. Thus, $Z^{1/2}$ represents a delay of one-half time unit, and $Z^{-1/2}$ represents an advance of one-half time unit. We want to move the waves at the top of layer 1 down to the interface which is at the bottom of layer 1. The downgoing wave takes one-half time unit to travel down, so it is delayed by one-half time unit. On the other hand, the upgoing wave must be depropagated back in time by one-half time unit, in order for it to go backwards to the interface from which it came. In other words, the upgoing wave must be advanced by one-half time unit. Thus, at the bottom of layer 1, the generating functions are

$$Z^{1/2}D_1(Z), \quad Z^{-1/2}U_1(Z) \ .$$

Just across the interface, the generating functions are

$$D_2(Z), \quad U_2(Z) \ .$$

We now draw the following analogy between the layered model and the special theory of relativity. Layer 1 corresponds to frame 1, and layer 2 to frame 2. The reflection coefficient c_1 of the interface corresponds to the velocity c_1 between the two frames. Let $Z^{1/2}\,D_1(Z)$ correspond to t_1,

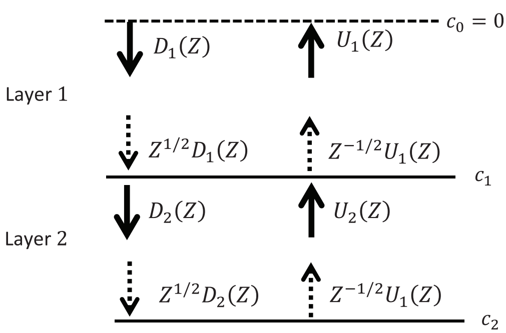


Figure 22. The layered system. The wave motion for a layer (solid arrow) is always measured at the top of the layer. The dotted arrows show the wave motion at the bottoms of the layers.

$Z^{-1/2}\,U_1(Z)$ correspond to x_1, $D_2(Z)$ correspond to t_2, and $U_2(Z)$ correspond to x_2. Under this correspondence, the Lorentz transformation becomes

$$\begin{aligned} D_2(Z) &= \frac{1}{(1-c_1^2)^{1/2}}(Z^{1/2}D_1(Z) - c_1 Z^{-1/2}U_1(Z)) \ , \\ U_2(Z) &= \frac{1}{(1-c_1^2)^{1/2}}(-c_1 Z^{1/2}D_1(Z) + Z^{-1/2}U_1(Z)) \ . \end{aligned} \tag{26}$$

The Lorentz equations 26 describe reflection and refraction at an interface. In matrix terms, these Lorentz equations are

$$\begin{bmatrix} D_2 \\ U_2 \end{bmatrix} = \frac{Z^{-1/2}}{\sqrt{1-c_1^2}} \begin{bmatrix} Z & -c_1 \\ -c_1 Z & 1 \end{bmatrix} \begin{bmatrix} D_1 \\ U_1 \end{bmatrix} .$$

We have shown that, when we cross an interface, the wave motion must be related by the Lorentz transformation. Even as the velocity c_1 (in natural units) between the two frames must physically be less than one in magnitude, so we also know that a physical reflection coefficient c_1 must be less than one in magnitude. (Here we are not including the case of perfect reflection, in which case c_1 is plus one or else minus one). The Lorentz transform for layered media is a consequence of the invariance of the *net downgoing energy* in each of the layers.

Turning to Figure 23, now let us establish the Einstein subtraction formula by means of the Lorentz transformation. Let the source be the unit impulse $D_1(Z) = 1$. Let the reflection seismogram be $U_1(Z) = R_1(Z)$. Then

$$\begin{bmatrix} D_2 \\ U_2 \end{bmatrix} = \frac{Z^{-1/2}}{\sqrt{1-c_1^2}} \begin{bmatrix} Z & -c_1 \\ -c_1 Z & 1 \end{bmatrix} \begin{bmatrix} 1 \\ R_1 \end{bmatrix} ,$$

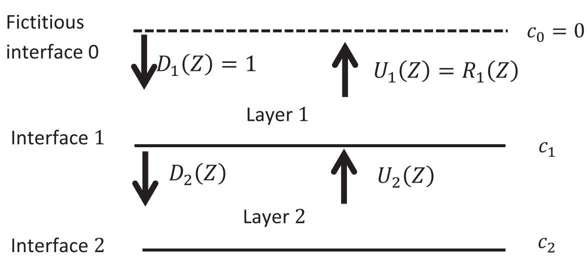


Figure 23. Source is the unit impulse $D_1(Z) = 1$. Reflection seismogram is $U_1(Z) = R_1(Z)$.

which gives

$$\sqrt{1-c_1^2}\,Z^{1/2}D_2 = Z - c_1R_1 \ ,$$

$$\sqrt{1-c_1^2}\,Z^{1/2}U_2 = -c_1Z + R_1 \ .$$

The reflection seismogram R_2 is the result obtained by deconvolving the upgoing wave U_2 by the downgoing wave D_2; that is,

$$R_2 = \frac{U_2}{D_2} = \frac{R_1 - c_1Z}{Z - R_1c_1} \ . \tag{27}$$

This equation is the *Einstein subtraction formula*.

Let us now give the *dynamic deconvolution* computational scheme for the inversion of a reflection seismogram. System 1 has n interfaces, numbered $1, 2, 3, \ldots, n$. The Fresnel reflection coefficients of the interfaces are given by $c_1, c_2, \ldots, c_n$. Interface 1 is the surface of ground or water. The lowermost interface is interface n. Below interface n is basement rock, which accepts downgoing waves but returns no reflected waves.

As shown in Figure 24, the field-recorded reflection seismogram $r_1, r_2, r_3, \ldots$ occurs in layer 1 (the air) just below fictitious interface 0. The generating function of the reflection seismogram is

$$\begin{aligned} R_1(Z) &= r_1Z + r_2Z^2 + r_3Z^3 + \cdots \\ &= c_1Z + (\text{terms in higher powers of } Z) \ . \end{aligned}$$

We know $R_1(Z)$ must have this form, because c_1Z represents the first bounce off interface 1. No multiple reflections can appear at the time of the first bounce.

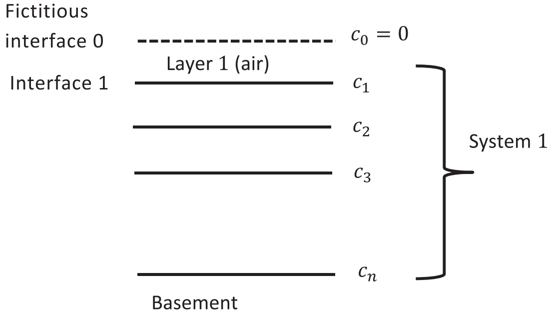


Figure 24. System 1 has n interfaces, with air above interface 1 and basement below interface n. The reflection seismogram $r_1, r_2, r_3, \ldots$ is measured just below the fictitious interface 0, which is one-half time unit above interface 1.

Remove interface 1. Now the top interface is interface 2. The resulting system is called system 2, which has $n-1$ interfaces, numbered $2, 3, \ldots, n$. The Fresnel reflection coefficients of the interfaces are $c_2, \ldots, c_n$. Interface 2 is now the surface of ground or water. Interface n is the lowermost interface. Basement rock is below interface n.

Then, as shown in Figure 25, the reflection seismogram for system 2 is represented by the generating function

$$R_2(Z) = c_2 Z + (\text{terms in higher powers of } Z) \ .$$

Here, $c_2 Z$ represents the first bounce.

Next, remove interface 2. We are left with system 3 with $n-2$ interfaces, numbered $3, \ldots, n$. The resulting reflection seismogram has generating function

$$R_3(Z) = c_3 Z + (\text{terms in higher powers of } Z) \ ,$$

where $c_3 Z$ represents the first bounce.

Thus, we arrive at a suite of reflection seismograms $(k = 1, 2, 3, \ldots)$ with generating functions

$$R_k(Z) = c_k Z + (\text{terms in higher powers of } Z) \ ,$$

where $c_k Z$ represents the first bounce off interface k. Therefore, we can make the following important conclusion. Given the reflection seismogram for layer k, we immediately can find the reflection coefficient c_k for layer k, because c_k is simply the first coefficient appearing in the seismogram.

We will now give the inversion algorithm of *dynamic deconvolution*, given the reflection seismogram R_1. This seismogram is the one physically recorded in the field. The problem is to find the sequence $c_1, c_2, \ldots, c_n$ of reflection coefficients and the suite of reflection seismograms $R_2, R_3, \ldots, R_n$.

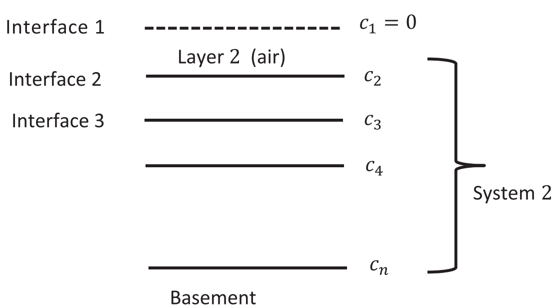


Figure 25. System 2 has $n-1$ interfaces, with air above interface 2 and basement below interface n. Now the missing interface 1 plays the role of the fictitious interface.

The *Einstein subtraction formula* is given by equation 27, which is

$$R_2 = \frac{R_1 - c_1 Z}{Z - R_1 c_1} \ . \tag{28}$$

We have in our possession R_1 (i.e., the field-recorded seismogram). In step 1, we find c_1 as the first bounce of R_1, and then we compute R_2 by the Einstein addition formula 28. So ends step 1.

The next Einstein subtraction formula is

$$R_3 = \frac{R_2 - c_2 Z}{Z - R_2 c_2} \ . \tag{29}$$

In step 2, we find c_2 as the first bounce of R_2, and then we use the Einstein subtraction formula 29 to find R_3. So ends step 2.

In step 3, we find c_3 as the first bounce of R_3, and then we find R_4 as

$$R_4 = \frac{R_3 - c_3 Z}{Z - R_3 c_3} \ .$$

So ends step 3.

Thus, given R_1, we can perform the entire deconvolution and obtain the sequence of reflection coefficients $c_1, c_2, c_3, \ldots$ and the suite of reflection seismograms $R_2, R_3, R_4, \ldots$. From the sequence of reflection coefficients, we can compute the impedance function of the earth.

There are many other features of dynamic deconvolution which make it attractive, such as the option of the determination of the reflection coefficients in the reverse order $c_n, c_{n-1}, \ldots, c_2, c_1$ (Robinson, 1984). This reverse order can be useful, because, in many cases, the most harmful noise appears at the beginning of the reflection seismogram. In such a case, it is better to work backwards in time on the reflection seismogram, leaving its noisy beginning to the end of the computations.

Further insight on dynamic deconvolution can be obtained by examining the role played by the multiple reflections.

Moving to Figure 26, for interface 1, the Fresnel reflection coefficient is c_1 and the two-way Fresnel transmission coefficient is $(1 - c_1^2)$. For interface 2, the Fresnel reflection coefficient is c_2. System 1 is composed of interface 1, and all of the interfaces below and no interface above. System 2 is composed of interface 2, and all of the interfaces below and no interface above. The reflection seismogram for system 1 is R_1 and the reflection seismogram for system 2 is R_2.

Then, referring to Figure 27, all wave motion is measured at a point just below the interface. The unit source pulse is at A. The observed reflection seismogram is composed of pulses at $C, F, I, L, \ldots$. All multiples encounter

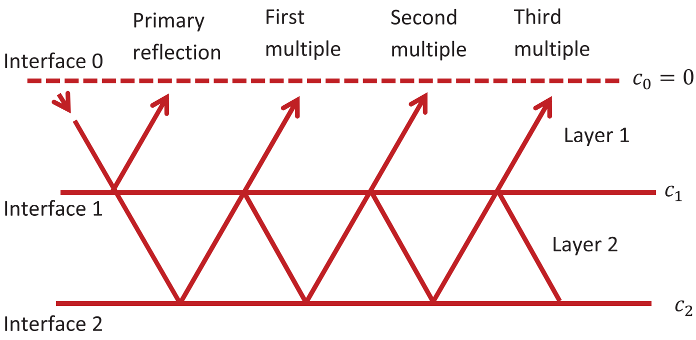


Figure 26. The refection seismogram is measured just below the fictitious interface 0. The unit source (illustrated by the small downgoing arrow) also occurs at the same place.

a two-way transmission though interface 1. The first multiple has one jog in layer 2. The second multiple has two jogs in layer 2. The third multiple has three jogs in layer 2, etc.

Per Figure 27, the constituent pulses of the observed reflection seismogram are given by the factors in the second column of the following table.

Reflection	**Constituent Pulse**
Primary reflection = ABC	Two-way travel through layer 1: Z Reflection upward at B: c_1
First multiple reflection = $ABDEF$	Two-way travel through layer 1: Z Two-way transmission through interface 1: $(1 - c_1^2)$ Reflection upward at D: R_2
Second multiple reflection = $ABDEGHI$	Two-way travel through layer 1: Z Two-way transmission through interface 1: $(1 - c_1^2)$ Reflection upward at D: R_2 Reflection downward at E: $-c_1$ Reflection upward at G: R_2
Third multiple reflection = $ABDEGHJKL$	Contributes another factor $-c_1R_2$ to above
Fourth multiple reflection	Contributes another factor $-c_1R_2$ to above
...	...

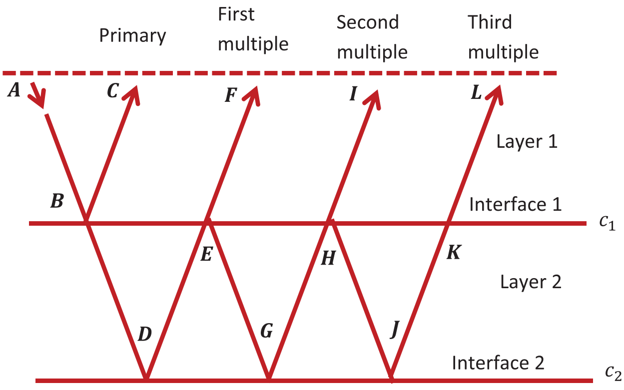


Figure 27. The same diagram as in Figure 26 but now with points A, B, C, ... labeled.

The reflection seismogram for system 1 is

$$R_1 = \text{primary} + \text{first multiple} + \text{second multiple} + \text{third multiple} + \text{fourth multiple} + \cdots .$$

If we multiply the various factors together, the above table yields

Reflection	**Result**
Primary reflection	$c_1 Z$
First multiple reflection	$(1 - c_1^2) R_2 Z$
Second multiple reflection	$(1 - c_1^2) R_2 Z (-c_1 R_2)$
Third multiple reflection	$(1 - c_1^2) R_2 Z (-c_1 R_2)^2$
Fourth multiple reflection	$(1 - c_1^2) R_2 Z (-c_1 R_2)^3$
...	...

Thus,

$$R_1 = c_1 Z + (1 - c_1^2) R_2 Z + (1 - c_1^2) R_2 Z (-c_1 R_2) + (1 - c_1^2) R_2 Z (-c_1 R_2)^2 + \cdots ,$$

which may be written as

$$R_1 = c_1 Z + (1 - c_1^2) R_2 Z [1 + (-c_1 R_2) + (-c_1 R_2)^2 + \cdots] .$$

Thus,

$$R_1 = c_1 Z + \frac{(1 - c_1^2) R_2 Z}{1 + c_1 R_2} \ .$$

Combining terms, we obtain the Einstein addition formula

$$R_1 = \frac{c_1 Z + R_2 Z}{1 + c_1 R_2} \ .$$

If we solve this equation for R_2, we again obtain the *Einstein subtraction formula*

$$R_2 = \frac{R_1 - c_1 Z}{Z - R_1 c_1} \ . \tag{30}$$

This equation is the same as equation 27. Knowledge of the Lorentz transform and the Einstein subtraction formula provides valuable insight on the behavior of seismic waves in the sedimentary earth system.

Chapter 3

Elasticity

The wave equation

The foundation of seismology is the theory of wave motion, a complicated concept that is still—after centuries of experiments and speculations by many of the very greatest scientists—an area of active research in many disciplines. Even simple forms of wave motion are difficult to describe verbally; but, ironically, the simplest type of wave is remarkably easy to describe (and subsequently analyze) mathematically.

This is one of those areas where, in the words of Nobel Prize physicist Steven Weinberg, mathematics has a "spooky" correlation to the physical world. Although some naturally occurring crystals have perfect geometric shapes, right triangles are a purely mathematical concept. They exist outside of our ordinary experience of the physical world. Have you ever found a rock in your back yard or on a field trip that is in the shape of a perfect right triangle? Yet we remember from elementary trigonometry (the mathematical analysis of the properties of triangles) that the graph of the sine function perfectly represents certain periodic motions, such as the (small) oscillations of a pendulum. This type of sinusoidal motion is called *simple harmonic motion*.

The pure sine curve, $u = \sin x$, is quite restricted. The value of u never can be greater than 1 or less than -1 and x must traverse a distance of 2π radians before one cycle of motion is completed. These limitations are, however, not serious. The sine function is easily tailored to represent any regularly repeating motion no matter what its height/depth (or amplitude), its frequency of oscillation, or its value when it crosses a "starting" point (often the $x = 0$ line). Such an all-purpose sine function can be written, supposing u to be the disturbance caused by the motion, as

$$u = A \sin 2\pi\left(\frac{x}{\lambda} - \frac{t}{T}\right) ,$$

where x is distance and t is time.

The number A (chosen to be positive) represents the amplitude; the distance between consecutive crests is λ (called the *wavelength*); the quantity T is the period or the time it takes the wave to complete one cycle. The crest of the wave moves a distance λ in time T. Because λ is a distance and T a time, the quotient λ/T equals the wave's velocity—almost always expressed simply as v. Therefore, if x is fixed and t is allowed to vary, a wave crest sweeps past the fixed point with a propagation velocity given by v.

There are other useful ways in which we can write a sine function to represent wave motion. Instead of wavelength and period, we can use wavenumber k and frequency ω where

$$k = \frac{2\pi}{\lambda} \quad \text{and} \quad \omega = \frac{2\pi}{T} ,$$

and then the sinusoidal wave may be expressed as

$$u = A \sin(kx - \omega t) ,$$

which represents a simple harmonic progressive wave. We also can write this curve as

$$u = A \sin k(x - vt) ,$$

because $v = \lambda/T = \omega/k$.

The quantity ω (which is expressed in units of radians per second) represents angular frequency. It is related to cyclical frequency (expressed in Hertz) by the equation $\omega = 2\pi f$. Likewise, the quantity k (expressed in radians per meter) is the angular wavenumber, related to cyclical wavenumber κ by the equation $k = 2\pi\kappa$. In the expression $x - vt$, substitute $x + vt'$ for x, and $t + t'$ for t. The result is

$$x - vt = (x + vt') - v(t + t') = x - vt + vt' - vt' = x - vt .$$

Thus, $x - vt$ reproduces itself when x becomes $x + vt'$ and t becomes $t + t'$. Therefore, any function of $x - vt$ can be said to represent a propagating wave.

An effective way to illustrate this is to imagine a taut string lying on the x axis. If the string is displaced perpendicular to the x axis, then the shape of the resulting curve can be written $u = f(x)$. If the displacements alter in such a way that the pulse travels with velocity v in the positive x direction without change of shape, the equation representing the pulse at any time t will still be $u = f(x)$, provided that we move the origin a distance vt in the positive direction (see Figure 1). In reference to the old origin, the equation of the pulse will have x replaced by $x - vt$ or $u = f(x - vt)$.

Therefore, we can state that this is the general equation of a wave with constant shape traveling in the positive direction with velocity v. Furthermore, every wave of this type must be expressible in this form. Similarly, a wave going in the opposite direction (i.e., negative x) is represented by a function $u = g(x + vt)$.

Pythagoras discovered that the pitch of a sound from a plucked string depends upon the string's length, and that harmonious sounds are given off by strings whose lengths are in the ratio of whole numbers. However, significant additional progress was impossible until the invention of calculus, more than 2000 years later, which permitted the English mathematician Brook Taylor to make the first productive attempt at the quantification of wave motion.

Consider a stretched string (such as a violin string) with initial shape $f(x)$. According to basic differential calculus, the slope of the tangent line at any point represents the rate of change of the function with respect to x. This rate of change is the first derivative of f with respect to x; in turn, the rate of change of the slope (or second derivative of f with respect to x) represents the curvature of the function. A basic assumption is that the displacements never violate the elastic limit of the string.

Now consider the motion of any particular point on the string. We have seen that the traveling wave can be represented as $f(x - vt)$, which, at the point $x = 0$, becomes merely $f(-vt)$. That point is moving up and down approximately at right angles to the x axis. Returning to basic calculus recall that the point's up-and-down velocity is given by the first derivative

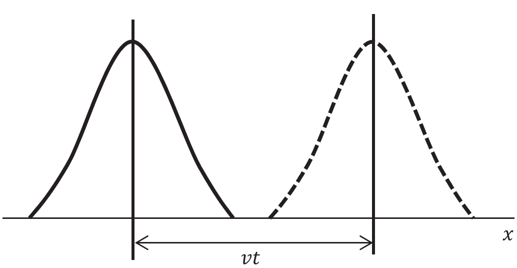


Figure 1. Illustration of how $f(x - vt)$ reproduces itself.

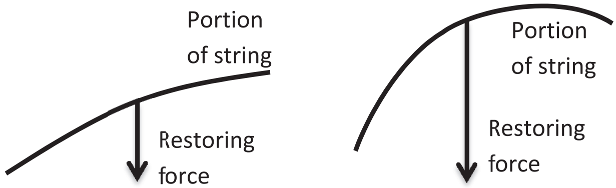


Figure 2. Force is proportional to the curvature of the string.

of f with respect to t and the point's acceleration is given by the second derivative of f with respect to t.

When the string is in its equilibrium position (horizontal along the x axis), there is no net vertical force acting on any point on the string. However, when the string is curved, the tension in the string exerts a restoring force which attempts to move it back to its equilibrium position. The more the curvature, the greater is this restoring force (Figure 2). Taylor notes this feature and reasons that the restoring force is proportional to curvature and, as a student of Sir Isaac Newton, he knew that force is proportional to acceleration. Thus, he writes the equation

$$\text{curvature} = \alpha \text{ acceleration} \ ,$$

where α is a constant of proportionality. Taylor could not fully develop the properties of this equation because he had no knowledge of partial derivatives. But after their invention, his speculation is confirmed. It also happens that the constant of proportionality is $1/v^2$. In modern mathematical notation, this equation is written

$$\frac{\partial^2 u}{\partial x^2} = \frac{1}{v^2} \frac{\partial^2 u}{\partial t^2}$$

and it is known as the one-dimensional wave equation. When generalized to three dimensions, it governs the melodies of Pythagoras, the propagation of seismic energy through the earth, and many other kinds of wave motion. The three-dimensional wave equation is

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = \frac{1}{v^2} \frac{\partial^2 u}{\partial t^2} \ .$$

If we now go back to one of our original equations for simple harmonic motion,

$$u = A \sin(kx - \omega t) \ ,$$

and take second partial derivatives with respect to x and t, and then substitute these second derivatives into the one-dimensional wave equation, we obtain

$$-A\,k^2 \sin(kx - \omega t) = \frac{1}{v^2}(-A\omega^2)\sin(kx - \omega t)\ .$$

After canceling common factors, we discover that

$$k^2 = \frac{\omega^2}{v^2}\ ,$$

which is called the *dispersion equation* for the one-dimensional wave equation. It relates wavenumber and frequency. When extended to three dimensions, this equation becomes

$$k_x^2 + k_y^2 + k_z^2 = \frac{\omega^2}{v^2}\ ,$$

where $k = \sqrt{k_x^2 + k_y^2 + k_z^2}$ is the wavenumber and k_x, k_y, k_z are the wavenumber components in the three coordinate directions.

The wave equation is an equation in space and time coordinates (x, y, z, and t), whereas the related dispersion equation is an equation in wavenumber and frequency coordinates (k_x, k_y, k_z, and ω). Geophysicists usually want to look at the data in the familiar time–space display; but often it is advantageous to transform the data into the frequency–wavenumber domain for computer processing. The basis for the transformation from one domain to the other is the mathematical operation known as the Fourier transform.

The use of frequency–wavenumber analysis is one of the powerful tools of seismic data processing. Measurement of wavenumber as a function of frequency provides a reliable means of separating and measuring the various velocity components on a seismic section. Other important seismic processing operations (including dip moveout as well as migration) can make use of ω, k analysis. All of these processing methods are tied physically to the dispersion equation, which in turn follows from Taylor's inspired insight (in 1715)—spatial curvature is proportional to temporal acceleration—that led to the original formulation of the wave equation.

A wave at a boundary

"Give me a place to stand," Archimedes said, "a lever and a rock, and I will move the Earth." He very likely was the first scientist to speculate about moving the Earth, but modern exploration geophysicists were probably the first to make such a daring idea the very underpinning of their systematic investigations. For, indeed, moving the Earth—admittedly on a very small scale, nothing like what Archimedes had in mind—is precisely what we must do to generate the seismic waves we need to bring us information about the subsurface.

The part of the wave in which geophysicists are most interested is the wavefront; i.e., a surface over which the phase of the traveling wave disturbance is the same. If we know the position of a wavefront at a certain time, we can find its subsequent position at any later time by means of one of the most elegant and fundamental concepts in seismology—Huygens' principle.

The great Dutch scientist Christiaan Huygens (1629 – 1695) was one of the most formidable intellects who followed in the immediate wake of Galileo. Huygens had first-rank talents in many disciplines—astronomy, optics, and mathematics in particular. His contributions ranged from the most far-reaching theoretical speculations to extremely important advances in contemporary technology. Among the latter were the perfection of the pendulum clock and innovations in lens grinding (which led to his discovery of the rings around Saturn). Among the former were some of the basic concepts in the theory of probability and the famous principle of wave motion which bears his name.

To explain Huygens' principle, let us recall the three-dimensional wave equation. In the case of a homogeneous isotropic medium (meaning the wave velocity is the same at all places and in all directions), the wave equation yields a particularly simple and beautiful solution: the spherical wave. Not surprisingly, a spherical wave is one in which the successive wavefronts of wave motion emanating from a point source are concentric spheres centered on the source. If we slice through the three-dimensional physical reality to produce a two-dimensional cross-section, the spheres will become circles (Figure 3).

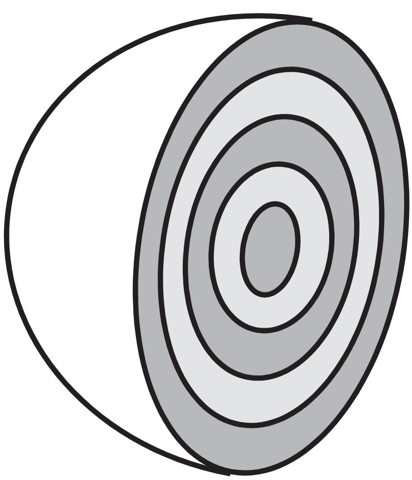


Figure 3. Spherical wave.

Huygens' principle dates from 1690 and thus significantly precedes the development of the wave equation, but Huygens' insight was so profound that his principle actually makes use of the so-called Green's function in the subsequently confirming mathematics. The basis of Huygens' concept is that each point along a wavefront may be viewed as a point source that produces a secondary spherical wavelet (known as the Huygens' wavelet), which propagates away from the point in all directions. In Figure 4, curve AB represents the instantaneous position of a wavefront. Let v be the wave velocity. In order to find the location of the wavefront after an interval of time Δt, we draw many secondary spheres of radius $v\Delta t$, all with centers on wavefront AB. A surface tangent to all of the secondary spheres is called an *envelope*. The envelope in the direction of propagation is the new wavefront $A'B'$.

This technique, known as Huygens' construction, is one of the most elegant products of classical physics. But, as usual, the physical reality is more complicated than the theory, because the logical conclusion of Huygens' theory is that two, not just one, disturbances would be propagated—one on each side of the wavefront and traveling in opposite directions. However, it is common knowledge that disturbance occurs only in the "forward" direction of wave propagation and that the theoretical

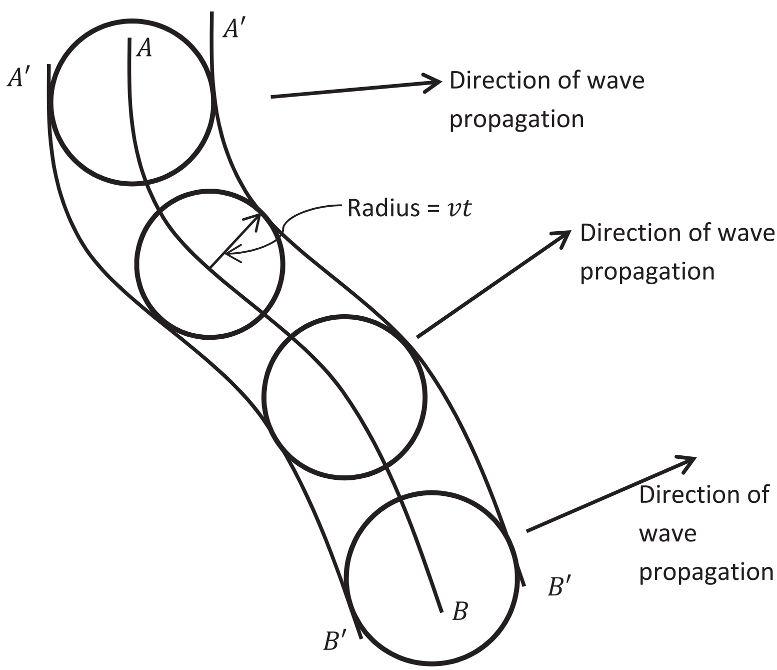


Figure 4. Huygens' construction.

"backward" disturbance, predicted by Huygens' principle and shown by dashed line $A''B''$ in Figure 4, does not appear. Why not?

It took nearly 200 years to obtain a completely satisfactory solution to the mystery. Fresnel took a major step toward clearing up this thorny question when he proposed, in 1826, that the backward wave did not occur because of destructive interference effects. (Fresnel only introduced an artificial obliquity factor to achieve this; Kirchhoff completed the theory.) Forty years later, Kirchhoff developed an integral solution of the wave equation which shows that the secondary wavelets from the point sources on the wavefront do destroy each other by mutual interference except on the forward propagating wavefront $A'B'$.

Huygens' principle is particularly valuable when we want a graphical explanation of the three fundamental ways in which the direction of a propagating wave can change: reflection, refraction, and diffraction (Figure 5).

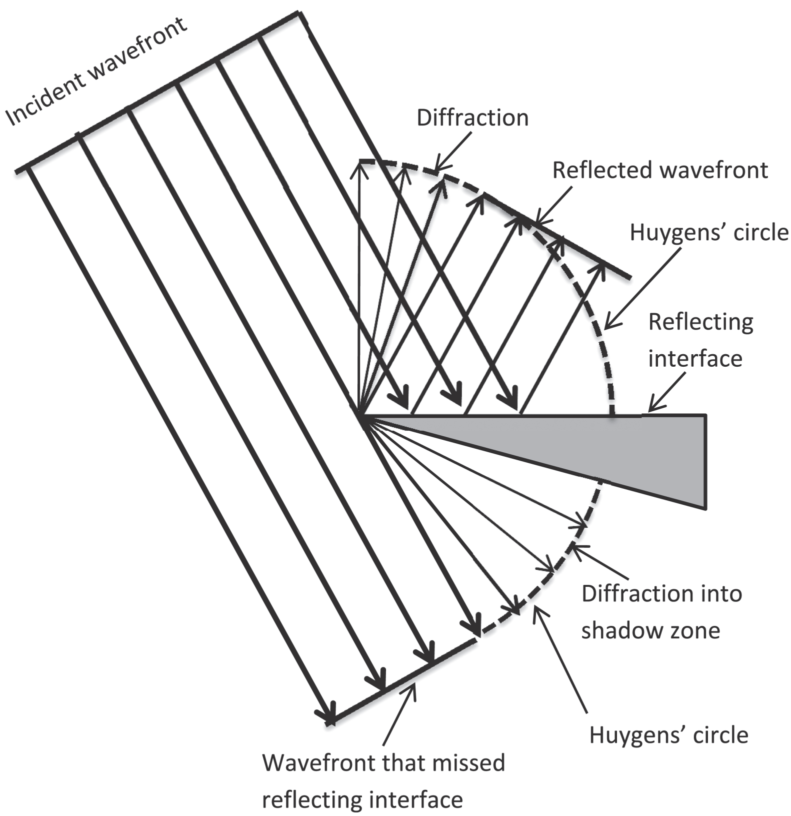


Figure 5. A wave at a boundary.

In most cases, when a wave arrives at a boundary between two different media, part of the wave's energy is reflected back into the original medium and part of it is transmitted into the new medium but at a different direction (i.e., refracted). The study of reflection and refraction is facilitated by the use of rays. In an isotropic medium, a ray is a line that is everywhere perpendicular to the successive wavefronts. While seismic energy does not travel only along raypaths, the greater part of the energy does indeed follow them. (Some seismic energy would reach a point by diffraction even if the raypaths between the point and the energy source were blocked.) Raypaths, therefore, constitute a very useful method of studying wave propagation. This discipline is called *geometrical seismology* or *geometrical acoustics*. It is the geophysical counterpart of the wellknown subject of geometrical optics.

There is an interesting facet of geometrical seismology which does not appear in textbooks on geometrical optics. Because light occurs at such high frequencies and travels at such a high speed, the waveforms of light are not measured as a function of time. As a result, geometrical optics only deals with the spatial paths of light rays. In seismic work, however, we routinely measure waveforms, and the seismic section is the resulting space–time representation. In geometrical seismology, therefore, we must consider space–time paths in addition to purely spatial paths. This space–time feature adds a whole new dimension to the classic ray theory as found in books on geometrical optics.

Let us initially apply Huygens' principle to a wavefront that is completely reflected from a plane surface and see how it specifies the direction and curvature of the reflected wavefront. Figure 6 illustrates the reflection of a spherical wave. The surface ABC represents the hypothetical position of the wavefront if the plane reflecting surface were not present. When the wave reaches P (the nearest point of the reflecting plane to the source S), the point P becomes the origin of a secondary wavelet. At immediately succeeding instants, the adjacent points to the left and right of point P are struck by the incident wavefront, and, in turn, they also become sources of secondary wavelets. The totality of the secondary wavelets emitted by the successive points on the reflecting plane has the spherical surface $AB'C$ for its envelope. We notice that the reflected wavefront appears to come from some point S' behind the reflecting plane; in fact, it's a simple matter to show that S' is the mirror image of the source point S.

Now let's use Huygens' principle to derive the direction of a reflected ray that is not perpendicular to the reflecting surface. In Figure 7, θ is the angle of incidence (by definition, the angle that the incident ray makes with the normal to the reflecting interface). The corresponding angle of reflection is denoted θ'. The incident wavefront falls along the line AB,

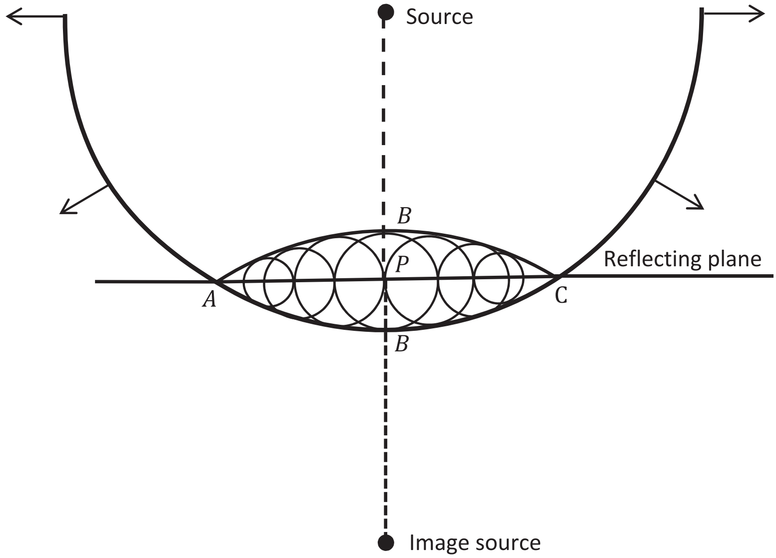


Figure 6. Reflection of a spherical wave.

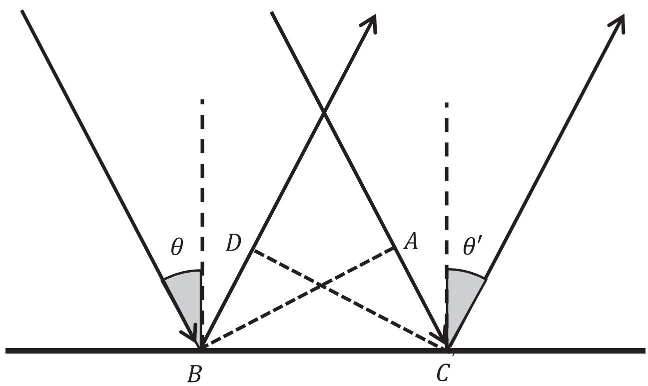


Figure 7. Law of reflection.

which is shown at the instant point *B* strikes the reflecting surface. Some of the energy in the incident wave at point *B* will reflect toward point *D*. As the incident wavefront continues toward the interface, each point between *B* and *C* will successively serve as a point source for secondary wavelets which will propagate back into the original medium. When the incident wavefront reaches point *C*, the reflected wavefront will be on line *DC*.

The speed of the reflected wave is the same as that of the incident wave, making the length of line *AB* equal to that of line *CD*. Trigonometry then can be used to establish one of the cornerstones of seismic theory; i.e., that

$\theta = \theta'$, or *the angle of incidence is equal to the angle of reflection*. This is called the *law of reflection*.

Of course, in most cases, all of the energy will not be reflected from the interface, but some will be transmitted into the second medium. When the wave (not at vertical incidence) enters another medium with a different speed, it will change direction—or undergo refraction. Figure 8a illustrates Huygens' construction of a refracted wave.

We suppose that the velocity v_2 of the wave in the lower medium is greater than velocity v_1 of the wave in the upper medium. When the incident wave reaches the interface at *B*, the wavelet radiating from *B* into the lower medium travels faster than the wavelet moving from *A* toward *C*. Wavelets from successive portions of the wavefront entering the lower medium will have longer radii in a given time interval than those traveling in the upper medium. Thus, the refracted ray will bend away from the vertical; the angle of refraction θ_2 will be greater than the angle of incidence.

This concept is illustrated further in Figure 8b, which shows the refraction construction from Figure 8a overlain by the hypothetical propagation of wavefront *AB* as it would have occurred in the absence of interface *BC*. Figure 8b makes it apparent that, if the velocity in the lower medium is greater, the angle of refraction must be greater than the angle of incidence. Indeed, one of the basic theorems of Euclidean geometry makes it apparent that θ_1 could equal θ_2 only if wavefront *AB* maintained an identical direction of propagation as it passed through interface *BC*.

Now refer to Figure 8a again. In the time Δt that it takes a wavelet to travel from point *A* to point *C*, a wavelet in the lower medium will travel from point *B* to point *D*. Because distance equals velocity times time, we

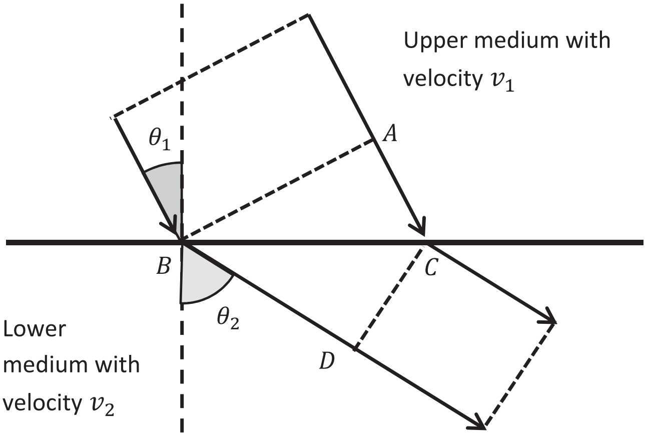


Figure 8a. Snell's law of refraction.

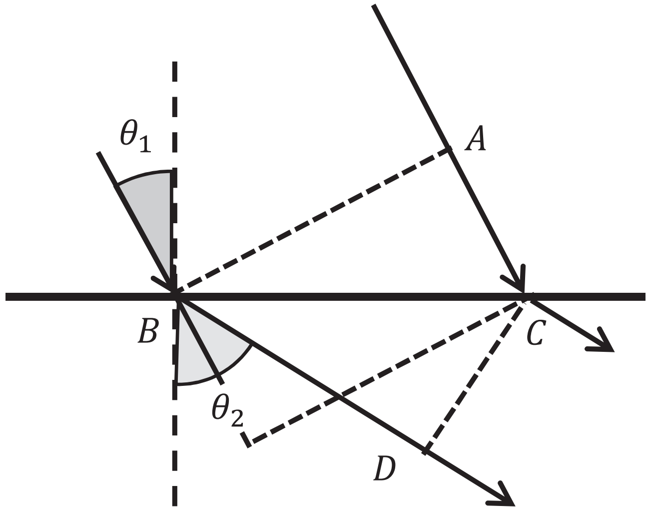


Figure 8b. Refraction construction overlain by hypothetical propagation of wavefront *AB*.

immediately see that

$$AC = v_1 \Delta t \quad \text{and} \quad BD = v_2 \Delta t \ .$$

Because triangles *ABC* and *BCD* are both right triangles and, respectively, contain angles θ_1 and θ_2, we see that

$$\sin \theta_1 = \frac{AC}{BC} \quad \text{and} \quad \sin \theta_2 = \frac{BD}{BC} \ .$$

Substituting for *AC* and *BD*, respectively, we obtain

$$\sin \theta_1 = \frac{v_1 \Delta t}{BC} \quad \text{and} \quad \sin \theta_2 = \frac{v_2 \Delta t}{BC} \ .$$

Eliminating $\Delta t / BC$, we obtain

$$\frac{\sin \theta_1}{v_1} = \frac{\sin \theta_2}{v_2} \ ,$$

which is known as the *law of refraction* or *Snell's law*.

Although this proof of Snell's law seems remarkably simple to modern eyes, it somehow eluded mathematicians—unlike the related law of reflection which was known at least as far back as the Greeks—until relatively recent times. Snell, reportedly after years of work, discovered the correct version of the law in 1621, but he did not publish it. Fellow Dutchman

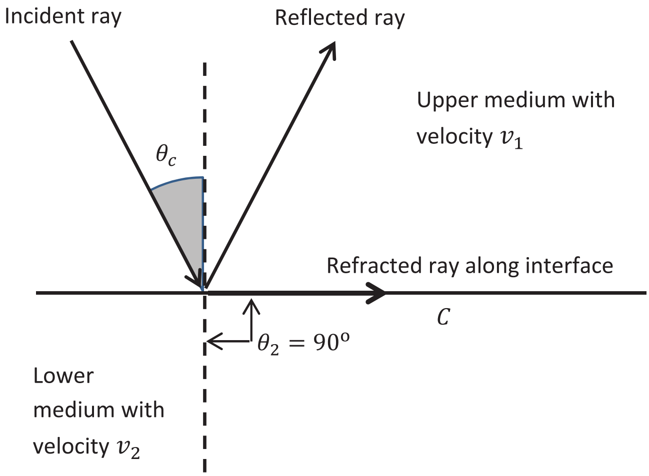


Figure 9. Critical angle θ_c.

Huygens, born three years after Snell's death, saw this work—which subsequently disappeared—and it was due to Huygens' efforts that Snell was given his deserved scientific immortality. In addition, Snell had a second profound effect on geophysical exploration; by developing a method of determining distances using trigonometric triangulation, he became one of the most important pioneers of scientific mapmaking.

In refraction from a medium with lower velocity to a medium with higher velocity (the usual case in seismic exploration), the rays turn away from the normal. So, in the case of

$$v_1 \sin \theta_2 > v_2 \ ,$$

what happens? Snell's law predicts that $\sin \theta_1 > 1$ which is an impossibility. However, the theory does not predict impossibility. Instead, it predicts that such refraction is impossible. In fact, when $\theta_1 = \theta_c$, with θ_c defined by $\sin \theta_c = v_1/v_2$, Snell's law says that $\sin \theta_c = 1$ and, therefore, $\theta_c = 90$; such an angle of incidence θ_c is called the *critical angle*, and the refracted ray grazes the interface (see Figure 9).

For angles of incidence greater than the critical angle, there is total internal reflection; i.e., all of the energy is reflected at the boundary back into the slower medium. The exploration technique called *amplitude with offset* utilizes this property. Because total reflection occurs at angles greater than the critical angle, reflected waves that originate from distant sources will have greater amplitude than those from sources that are "inside" the critical angle.

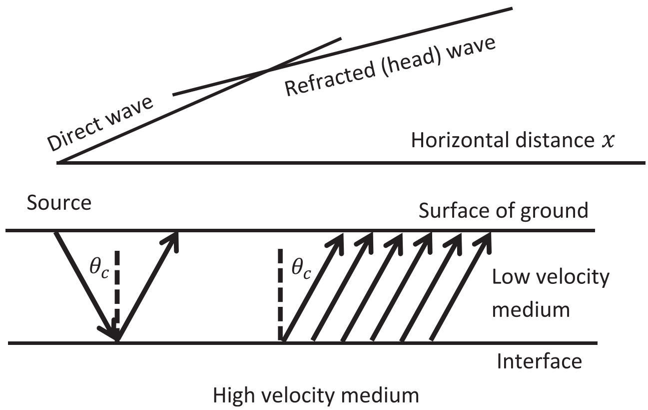


Figure 10. Head wave in a refraction survey.

The wave which travels along the interface, after the incident wave strikes at the critical angle, is known as the *head wave* (see Figure 10). As the head wave travels along the interface, it continuously feeds energy back up into the slower medium. This escaping energy leaves the interface at the same angle, the critical angle. Naturally, this escaping energy can be detected at the surface by geophones, and this is the fundamental principle of seismic refraction prospecting, the most important method of geophysical exploration for petroleum in the 1920s. It is still used but long ago yielded its dominant position in exploration to the reflection method.

When waves pass around an obstacle or through an aperture, they tend to curl around the edges so that the shadow of the obstacle on the downstream side is not sharply defined. This aspect of wave behavior is called *diffraction*. Diffracted sound waves can be heard around corners, and water waves entering a harbor spread into the area behind the breakwater. The amount of diffraction can be qualitatively determined by the ratio of the linear dimensions of the obstacle to the wavelength λ. If the obstacle has a length approximately equal to λ, then the amount of diffraction is large (i.e., many waves easily bend around the obstacle), so that the notion of a shadow zone becomes meaningless. In fact, a shadow zone of full darkness exists only in the limit of zero wavelength. For example, on a piano a note at middle C (264 Hz) has a wavelength of 1.3 m. That is comparable to room dimensions, so a sound from around a corner is audible. However, if the obstacle has dimensions very much greater than λ, the diffraction into the hidden region becomes negligible and the shadow of the obstacle is relatively sharp. Geophysicists are confronted with diffracted events every day. The most comprehensive treatment of seismic diffraction with an abundance of clear and instructive figures is given in *Classical and Modern*

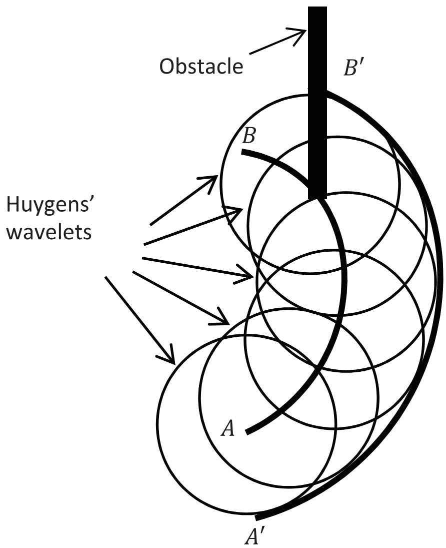


Figure 11. Huygens' construction of a diffracted wavefront showing how the wavefront $A'B'$ curls in behind the obstacle.

Diffraction Theory and in *Seismic Diffraction* (Klem-Musatov et al., 2016a, 2016b).

Diffraction allows us to account for the penetration of wave motion into regions forbidden by geometrical seismology. The law of reflection and Snell's law are not required to hold for diffracted rays. However, Huygens' principle does remain in force. Its explanation of the movement of diffracted waves around an obstacle is shown in Figure 11.

In fact, this use of Huygens' principle, in particular in its embodiment in Kirchhoff's solution of the wave equation, might be interpreted as basic. In this sense, a reflection may be thought of as the interference result of diffraction from points lying on the reflector. In other words, each source of a secondary wavelet may be considered a diffraction source. Huygens' construction represents the resulting interference pattern. Although each individual diffraction point on the reflecting surface does not obey the law of reflection (because the diffracted rays travel in all directions), the resulting envelope does yield a wavefront obeying the law of reflection.

In conclusion, we hesitantly offer a conjecture about Huygens' principle, hesitantly because some seismologists might consider it anathema. Huygens' abilities are not limited to science. An accomplished artist, an example of his art is reproduced in Figure 12. Appearing in the original publication of Huygens' *Traité de la Lumière*, it represents his own graphical conception of Huygens' principle. Could it be that such an

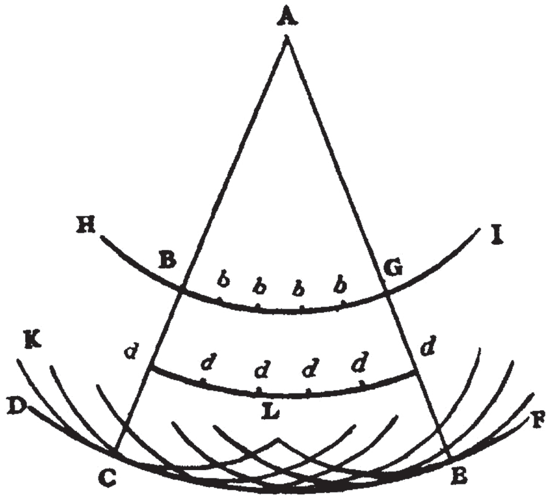


Figure 12. Huygens' principle as given in *Traité de la Lumière* (1690).

elegantly beautiful explanation for an ever-present natural phenomenon resulted more from Huygens' artistic sensibilities than from rigorous scientific analysis?

Stress and strain

Elasticity is the property that enables a fluid or solid body to resist change in size and shape when an external force is applied and to return to its original size and shape when the force is removed. This concept is a major building block in seismology because it is the elastic properties of rocks which allow seismic waves to propagate through the earth.

The theory of elasticity is one of the major achievements of classical physics. Its architects include many of the colossal scientific figures of the 17th, 18th, and 19th centuries, among whom are Robert Hooke, Gottfried Wilhelm Leibniz, James Bernoulli and his nephew Daniel Bernoulli, Leonhard Euler, Thomas Young, Charles Augustin Coulomb, Augustin Louis Cauchy, Claude Louis Marie Navier, Simeon Denis Poisson, and George Gabriel Stokes. The framework, which they and others constructed to quantify and analyze elasticity, derives from the basic concepts of classical mechanics known as stress and strain and the mathematical linkage known as Hooke's law.

Stresses are forces per unit area that are transmitted through a material, i.e., forces exerted by one part of a body on a neighboring part. Stresses that act perpendicularly to a surface are normal stresses;

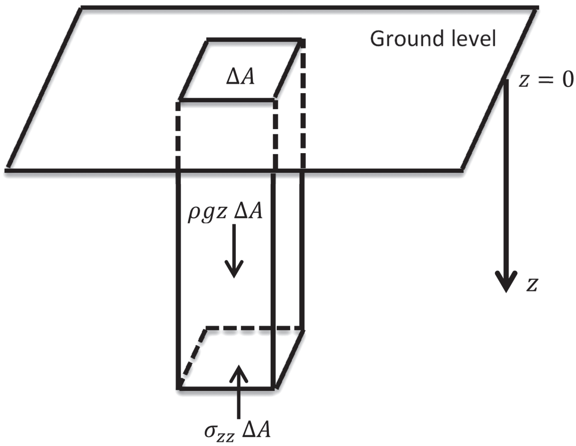


Figure 13. Normal stress.

those that act parallel to it are shear stresses. As an example, consider the force that acts at the base of a column of rock at depth z (beneath the ground level) to support the column (see Figure 13). The weight of the column of cross-sectional area ΔA is

$$\rho g z \, \Delta A \ ,$$

where ρ is constant density and g is acceleration of gravity. This weight must be balanced by an upward surface force

$$\sigma_{zz} \, \Delta A$$

distributed on the horizontal surface element of area ΔA at depth z. Here, we assume that there are no vertical forces on the lateral surfaces of the column. Thus, the quantity σ_{zz} is the surface force per unit area acting perpendicularly to the horizontal surface ΔA; that is, σ_{zz} is a normal stress. In equilibrium, the opposing forces must be the same, so

$$\sigma_{zz} = \rho g z \ .$$

This normal stress, due to the weight of the overlying rock or overburden, is known as the *lithostatic stress*.

Vertical subsurface areas also receive normal stresses. The normal stress acting in the x direction on a plane perpendicular to the x direction is σ_{xx}. The horizontal normal stress components σ_{xx} and σ_{yy} can include large-scale

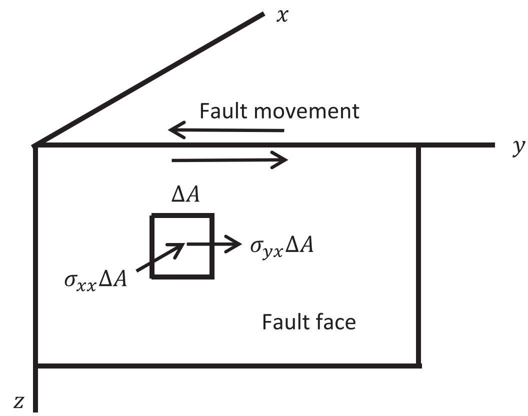


Figure 14. Strike-slip fault.

tectonic forces, in which case

$$\sigma_{xx} \neq \sigma_{yy} \neq \sigma_{zz} \ .$$

However, there are instances in global geophysics in which rock has been heated to sufficiently high temperatures or was initially sufficiently weak so that the three normal stresses are each equal to the weight of the overburden. When the three stresses are equal, they are referred to as the pressure. This balance between pressure and the weight of the overburden is called a *lithostatic state of stress*. Likewise, hydrostatic equilibrium can exist in the sea, where pressure forces are exerted equally in all directions and pressure increases linearly with depth.

Of course, forces also can act parallel to an area. Consider the forces acting on the element of area ΔA lying in the plane of a strike-slip fault (see Figure 14). The normal compressive force $\sigma_{xx}\Delta A$ acting on the fault face is a consequence of both the weight of the overburden and the tectonic forces which push the two sides of the fault together. The tangential or shear force $\sigma_{yx}\Delta A$ is the frictional resistance that opposes the tectonic forces, driving the left lateral motion of the fault.

Both types of stress are involved in Figure 15, which is a model of a zone of continental collision in which (as often happens) a thin sheet of crystalline rock is overthrust upon adjacent continental rocks by means of a low-angle thrust fault. The thrust sheet has been put in position as a result of horizontal tectonic forces. If the influence of gravity is neglected, the total horizontal tectonic force F_T due to horizontal tectonic stress σ_{xx} is

$$\sigma_{xx}\, HW \ ,$$

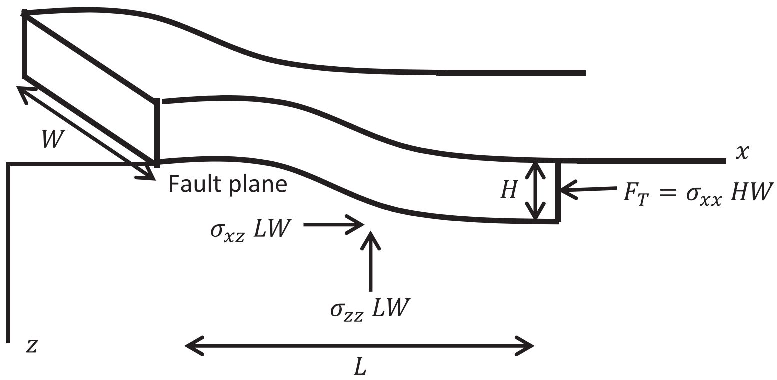


Figure 15. Zone of continental collision.

where H is the thickness of the thrust sheet and W is the width of the sheet. The total resisting shear stress F_R is

$$\sigma_{xz}\,LW\ ,$$

where L is the length of the thrust fault. Often the shear stress σ_{xz} is proportional to the normal stress pressing the surfaces together. In such cases,

$$\sigma_{xz} = c\,\sigma_{zz}\ ,$$

where σ_{zz} is the vertical normal stress acting on the base of the thrust sheet and the constant c is the coefficient of friction. If we assume σ_{zz} has the lithostatic value

$$\sigma_{zz} = \rho g H\ ,$$

then by setting $F_T = F_R$, we find that

$$\sigma_{xx} = c\rho g L\ .$$

This quantity is the tectonic stress required to emplace a thrust sheet of length L.

The double subscript notation is necessary, because the stress on a surface element in a solid body is not, in general, normal (perpendicular) to that surface, but impacts the surface element at an angle. However, the stress can be described by separating it into normal and tangential components by the use of appropriate coordinate axes. This is illustrated in Figure 16, where three mutually perpendicular axes (the traditional x, y, z)

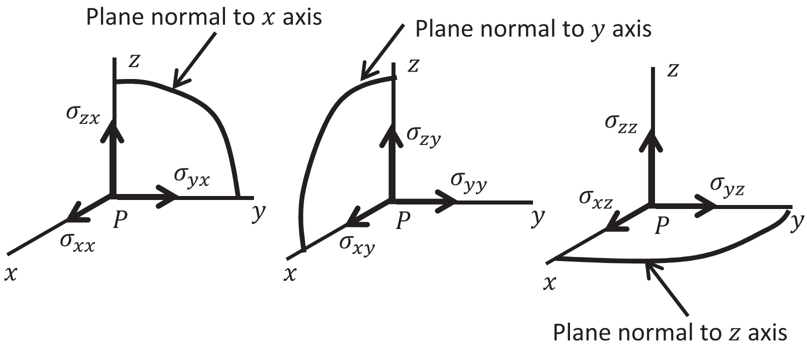


Figure 16. Stresses acting on the three planes normal to the three axes.

are oriented at point P. The stresses acting on the three planes normal to the three axes, and which pass through P, are indicated. This is one of the most conventional notations for stress. The symbol σ indicates stress; the first subscript refers to the direction of the force component and the second subscript to the direction of the normal to the element of area. Thus, a stress notation in which both subscripts are identical, such as σ_{xx}, represents a normal stress. A stress notation with differing subscripts indicates a shear stress.

There are, in theory, as shown in Figure 16, three normal stresses and six shear stresses. For practical purposes, however, there are only three independent shear stresses, because

$$\sigma_{xy} = \sigma_{yx}, \quad \sigma_{xz} = \sigma_{zx}, \quad \sigma_{yz} = \sigma_{zy} \ .$$

These equalities must hold, because there can be no net torque on the small cube, otherwise it would be spinning.

Strain quantifies the deformation or distortion that a body undergoes due to the application of external forces. There are two basic categories: normal strain and shear strain. The conventional notation for strain is the symbol ε and two subscripts. Thus, ε_{xx} represents a normal strain and ε_{xy} represents a shear strain.

A geologic example illustrating normal and shear strain is presented in Figure 17. Figure 17 (left) shows a family of lines that could model a geologic cross section of flat horizontal beds. Figure 17 (right) shows the same model after a shearing distortion. The line AB initially is normal to the family of lines (or parallel beds); but, after folding, this line makes the angle ψ with the normal. The angle ψ is called the *angular shear* and its

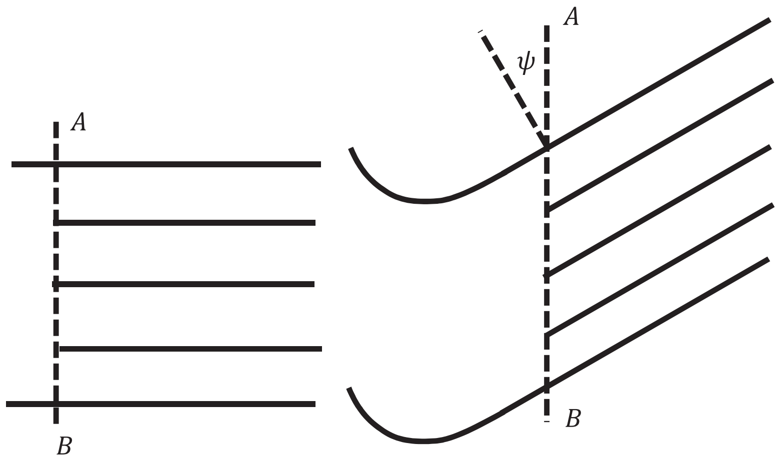


Figure 17. (left) Flat horizontal beds and (right) the same beds after folding.

tangent is defined as the *shear strain*. The ratio of the increased length of each line in Figure 17 (right) to its original length in Figure 17 (left) is defined as the *normal strain*.

Because stress and strain are both expressed as ratios, they often are confused. However, there are two important differences between them: stresses are related to force (in fact, one working definition is that stress is a measurement of a material's internal resistance to an external force), whereas strains deal strictly with configuration; and stresses specify conditions at a particular instant, whereas strains compare conditions at two different times.

For a more formal mathematical development of normal strain, let us look at a simple situation, called homogeneous strain, in which the strain is consistent throughout the rock. Suppose that the block of rock in Figure 18 has been stretched uniformly so that its dimension in the x-direction has changed. The movement u_x of a rock grain initially at x is proportional to x. The proportionality constant is $\Delta L/L$, where L is the original length of the block and ΔL is the change in its length. Thus,

$$u_x = \frac{\Delta L}{L} x \ .$$

If the strain is not uniform, the proportionality constant will vary from place to place; in this case, the proportionality factor is denoted as a kind of local $\Delta L/L$ which, when the displacement is very small (nearly always the

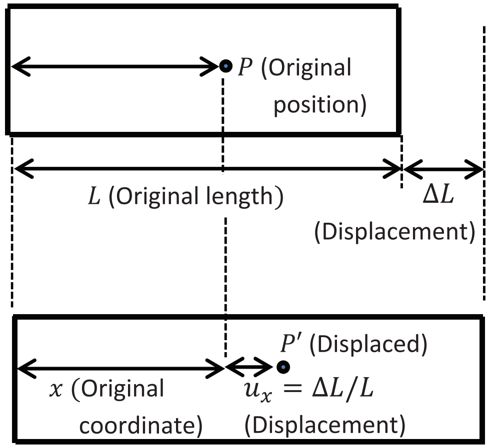


Figure 18. (top) Original block of rock and (bottom) block stretched uniformly in x-direction.

situation in geophysical exploration), can be expressed as

$$\varepsilon_{xx} = \frac{\partial u_x}{\partial x} \ .$$

(Obviously, some mildly complicated calculus is involved. The particulars can be found in many standard textbooks.)

The number ε_{xx}, called the normal strain, describes the amount of stretching in the x-direction. In general, there also is stretching in the y and z directions, and those amounts, ε_{yy} and ε_{zz}, are similarly established and defined as the normal strains in the y and z directions.

To develop a mathematical description of shear strain, we again begin with the homogeneous case by isolating a small cube (see Figure 19) in undisturbed rock. When the rock is deformed, the cube is transformed so that its initially rectangular cross section has become a parallelogram. If the strain is symmetric with respect to x and y, the total angle of shear is composed of two equal parts, $\theta/2$. From Figure 19 (right), we see that the x displacement u_x is proportional to the y coordinate, namely,

$$u_x = [\tan(\theta/2)]\, y \ .$$

When an angle is small, it is approximately equal to its tangent, making

$$u_x = (\theta/2)y \quad \text{and} \quad u_y = (\theta/2)x \ .$$

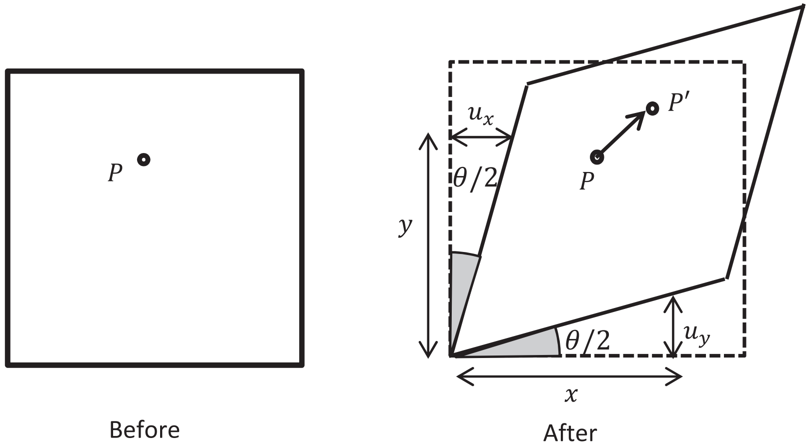


Figure 19. Both angles rotate inward toward the diagonal.

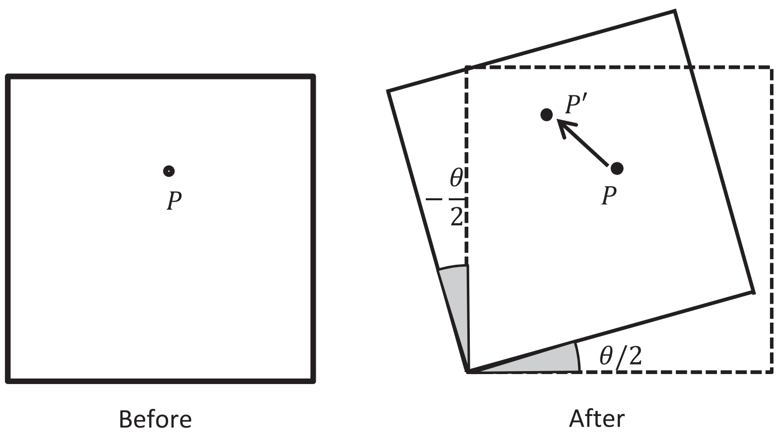


Figure 20. Both angles rotate in the same direction.

This permits the shear strains ε_{xy} and ε_{yx} to be defined as $\theta/2$. The displacement formulas then become

$$u_x = \varepsilon_{xy} y \quad \text{and} \quad u_y = \varepsilon_{yx} x \ .$$

Now assume a slightly altered situation, as shown in Figure 20. Instead of having both angles rotate inward toward the diagonal, let both move in the same direction, so that

$$u_x = -\left(\frac{\theta}{2}\right)y \quad \text{and} \quad u_y = \left(\frac{\theta}{2}\right)x \, .$$

In this case, the cube is simply rotated through the angle $\theta/2$. There is no distortion so, by definition, there is no strain. Thus, we must make certain that our mathematical description of strain eliminates pure rotations such as this. This can be achieved by defining shear strain as the arithmetic average of the two angles, thus (because small angles are approximately equal to their tangents, and tangents are equal to derivatives),

$$\varepsilon_{xy} = \varepsilon_{yx} = \left(\frac{1}{2}\right)\left(\frac{\partial u_x}{\partial y} + \frac{\partial u_y}{\partial x}\right) \, .$$

Nine strains (three normal and six shear) can exist in a three-dimensional body, but, by extending the reasoning which establishes that $\varepsilon_{xy} = \varepsilon_{yx}$, it follows that only three shear strains are independent.

As noted previously, because stress and strain both are expressed as ratios, they often are confused. The two important differences between them are as follows. Stresses are related to force (e.g., stress is a measurement of a material's internal resistance to an external force), whereas strains deal strictly with configuration; and stresses specify conditions at a particular instant, whereas strains compare conditions at two different times. This is emphasized because the "state of strain at a particular instant" is commonly discussed. This means the strain relating the body at that instant to some earlier shape, which is almost always the original undeformed shape. On the other hand, the stress condition at any particular instant can be specified completely by the force distribution at that instant; nothing has to be known about prior force distribution.

Rocks are among the many substances that can be considered in the context of small, perfectly elastic deformations. Thus, we can assume that elastic theory in general and Hooke's law (the mathematical relationship of stress and strain) in particular are applicable to the forces in and deformations of the earth. This is quite fortunate, because the essence of Hooke's law, stress is proportional to strain, is linear—a rarity among physical phenomena (99% of which are nonlinear) and a great convenience mathematically (99% of the known techniques are linear).

Seismic waves occur when the equilibrium of the particles in the earth is disturbed. Hooke's law can be combined with Newton's law of motion to analyze the wave motion. These computational manipulations, and the physical inferences that can be extracted from them, will be the themes of the following sections.

Hooke's law

In the first section of this chapter, *The Wave Equation*, the central topic is traveling waves on a string. These waves propagate along the string, i.e., travel from one end to the other (and back). However, each point on the string vibrates at right angles to the direction of wave motion. This type of wave, in which particle motion is perpendicular to wave motion, is called a *transverse wave*.

A sound wave can travel through any material medium (gases, liquids, solids). In the fluid case (gas or liquid), sound waves are called longitudinal waves because the particles of the medium vibrate along the direction of motion. This is in contrast to a transverse wave. The displacements that occur as a result of sound waves in air involve the longitudinal displacements of individual molecules from their equilibrium positions. This results in a series of high- and low-pressure regions called, respectively, *condensations* and *rarefaction*s. If the source of the sound waves vibrates sinusoidally (such as the diaphragm of a loudspeaker), the pressure variations also will be sinusoidal. The term sound wave includes audible waves (waves with frequencies within the range of sensitivity of the human ear, typically from 20 to 20,000 Hz), infrasonic waves (frequencies below the audible range), and ultrasonic waves (frequencies above the audible range).

A sound wave that travels through a solid usually is called an *elastic or seismic wave*. It can be either longitudinal or transverse. The purpose of this section is to review the theoretical reasoning and some of the mathematics which explain why both types of wave motion can propagate in a solid medium.

Wave motion occurs in many areas, not just those included in physics textbooks. For example, the ever-present business cycle (of which, in recent years, geophysicists have been only too aware) represents wave motion in the economic sphere. What's the difference between physical wave motion and other types of waves? The answer is given by the English mathematician Brook Taylor in 1713. He finds that the velocity of the traveling mechanical wave on a string depends on the string's physical properties (i.e., its tension and density). A basic assumption is that the displacements of the string never violate the elastic limit of the string. Taylor finds that the wave velocity depends upon the physics. This is the type of linkage that makes physics a science. There is no such linkage in economics where wave motion also predominates. But wave motion in geophysics has physical linkage, so also in acoustics, optics, electronics, quantum mechanics, and every other branch of physics involving wave motion.

Linkage is the key concept in geophysics. We must link the observed seismic waves with the physical characteristics of the subterranean rocks. In order to prepare for this task, it is necessary to consider the fundamental physical mechanisms involved. Taylor's expression for the speed v of transverse waves on a string is

$$v = \sqrt{\frac{T}{\rho}} \, ,$$

where T is the tension in the string and ρ is the linear density of the string.

Now consider a sound wave in the air. The compressed and rarefacted regions of the air composing the wave correspond to variations in the normal value of the air pressure. If the air has a bulk modulus B and an equilibrium density ρ, then the speed of sound in the air is

$$v = \sqrt{\frac{B}{\rho}} \, .$$

Bulk modulus is defined as the ratio of the change in pressure p to the resulting fractional change in volume V; typically this is written

$$B = -\frac{\Delta p}{\left(\frac{\Delta V}{V}\right)} \, .$$

Note that B must always be positive, because an increase in pressure, which means Δp is positive, results in a decrease in volume. Hence, $\Delta p/\Delta V$ is always negative, and multiplying it by -1 will ensure that the entire expression is always positive. The symbol Δ indicates the *increment of a quantity* in the above equation. The symbol Δ also is used to indicate a completely different quantity, called *dilatation*, but it should be clear from the context which usage is meant.

It is interesting to compare the two expressions for velocity. In both cases, the speed of the wave depends upon an elastic property of the medium (T or B) and on an inertial property ρ of the medium. In fact, the speed of all mechanical waves can be determined by an expression of the general form

$$v = \sqrt{\frac{\text{elastic property}}{\text{inertial property}}} \, .$$

Various elastic constants can be used to describe a solid medium, but only two independent ones are required to specify the elastic properties. For theoretical studies, Lamé's constants often are used. They are the pair λ and μ, where μ usually is referred to as the shear modulus.

So how are Lamé's constants determined? To answer this question, we must return to the early years of the modern scientific revolution and the very interesting gentleman after whom Hooke's law is named. Robert Hooke (1635 – 1703) is most remembered for the simple formula that stress is proportional to strain, but, in retrospect, that discovery seems rather incidental among an incredible number of profound speculations and solid accomplishments in a large number of disciplines.

Hooke developed theories about the wave motion of light and the inverse-square nature of gravitational force that anticipated Huygens and Newton. He predicted steam engines, named the basic unit of biology the cell, speculated that matter was composed of atoms, discovered the hair-spring which made small chronometers possible, was the first to state that matter expands when heated, was an ingenious instrument designer and experimentalist, and a gifted microscopist (the drawings in his famous book, *Micrographia*, are admired for their artistic merit as well as their scientific observations). Finally, he was a skilled architect, playing a key role in the rebuilding of London after the Great Fire of 1666. (Bethlehem Royal Hospital, which gave the term bedlam to the language, was designed by Hooke.)

Although he devoted little time to the earth sciences, some rank Hooke second only to Nicholaus Steno among the geologists of the era. Hooke studied fossils, earthquakes, the structure of crystals, and probably discovered independently (and perhaps prior to Steno) the law of constancy of interfacial angles. He speculated about evolution two centuries before Darwin and proposed a dynamic earth model three centuries before the plate tectonics revolution.

Hooke accomplished all of this by age 42. He lived another 25 years but did little more of importance, spending most of the time arguing over scientific priority—particularly with his greatest contemporary, and bitterest enemy, Sir Isaac Newton. It has been suggested that the extraordinarily nasty relationship between Hooke and Newton was a major factor in the latter's nervous breakdown in 1692. However, Newton recovered and his ideas dominated European science, to the detriment of the reputations of Hooke and others, for the next two centuries. Hooke's discovery of the relationship between stress and strain probably dates from the 1660s, but he did not publish it until 1678. The mathematics of the time did not allow him to develop the principle into the sophisticated form that makes it an invaluable tool in physics. This was done by others, primarily Cauchy (1823) and Stokes (1845), who produced the modern theory that is studied today.

Hooke's scientific work combined the generosity of the new age and the parsimony of the past. He was active in promoting schemes for cooperative endeavor, such as weather records, and he was lavish with fruitful suggestions. Hooke, however, also had an anxiety to wear the laurels of priority. He was the last well known author to use the time-honored device of securing priority by preliminary announcement in an anagram. He first gave Hooke's law as the anagram *ceiiinosssttuv* two years before he disclosed the solution *ut tensio sic vis* in a published description of the experimental evidence supporting the law.

Hooke's law says that the pulling power of a stretched string is proportional to the displacement. This is the case of a body subjected to deformation in a single direction. Hooke's law simply says that, for a linear body, such as a string, normal stress σ_{xx} is proportional to normal strain e_{xx}, that is,

$$\sigma_{xx} = Ee_{xx} \ .$$

The proportionality constant is called Young's modulus (named after English scientist Thomas Young, whose versatility rivals that of Hooke, but that's another and later story). For most materials, Young's modulus is of the order of a megabar (10^{12} dynes/cm^2).

Even so, this simple form of Hooke's law does not hold in three dimensions. The concept known as *dilatation* must be developed before the 3D form of the law can be derived. The stresses and strains in a three-dimensional body are described in the previous section. There are three normal stresses σ_{xx}, σ_{yy}, σ_{zz} and six shear stresses. However, there are only three independent shear stresses, because

$$\sigma_{xy} = \sigma_{yx}, \quad \sigma_{xz} = \sigma_{zx}, \quad \sigma_{yz} = \sigma_{zy} \ .$$

The same reasoning shows that there are three normal strains and six shear strains but only three independent shear strains.

The development of a mathematical foundation for seismology requires that the nine stress components be linked to the nine strain components for each small piece of rock. The 3D form of Hooke's law provides this relationship and gives us a vital insight into the nature of wave motion in the earth (or any other solid body). If it is assumed that the rock is isotropic, i.e., noncrystalline so that there are no preferred directions, the stress components must be related to the strain components in a way that does not depend on the coordinate directions. This means the σ_{xy} and e_{xy} must be related in only one form,

$$\sigma_{xy} = (\text{constant})\ e_{xy}\ .$$

This constant is defined to be the Lamé shear modulus cited earlier. Actually, for mathematical convenience, this constant usually is denoted by 2μ, which produces

$$\sigma_{xy} = 2\mu\ e_{xy}, \quad \sigma_{xz} = 2\mu\ e_{xz}, \quad \sigma_{yz} = 2\mu\ e_{yz}\ ,$$

the mathematical form of the relationship between shear stresses and shear strains.

An obvious initial assumption is that this form would hold for the relationship between normal stress and strain (i.e., $\sigma_{xx} = 2\mu\ e_{xx}$). But this is not correct, because normal strains result in a change in area (2D) or volume (3D). This involves dilatation, the basic concept which brought mathematical immortality to Gottfried Wilhelm Leibniz (1646 – 1716) as co-inventor of calculus. Although Leibniz sometimes is called the most brilliant intellect in an age when genius was fairly common, he is practically forgotten in the final years of his life. And, when he initially is resurrected, it is as the prime target in Voltaire's merciless satire, *Candide* (1758).

Ultimately, however, Leibniz reassumes an honored place in this history of science because his dilatation concept—not the fluxions of Newton—became the foundation of modern calculus. The Leibniz theory is superior, because it introduces multiple variables at the outset. And, because it is based on area instead of Newton's tangent lines, the fundamental tie with integration (also based on area) becomes more apparent.

Consider the rectangle (dark portion) with sides u and v in Figure 21. The product uv represents the area, and the symbolism $d(uv)$, called the differential of uv, represents an incremental increase in the area (or, in the physics of Leibniz, a small pulsation). The light portion of Figure 21 represents the differential $d(uv)$.

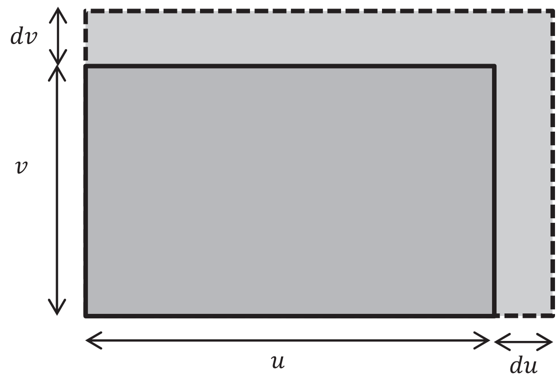


Figure 21. Area (dark portion) and its incremental increase (light portion).

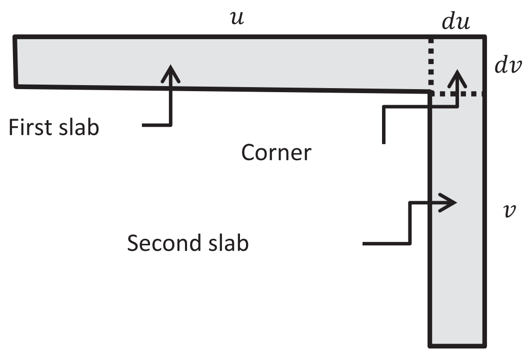


Figure 22. The sum of three elements.

Figure 22 shows that the differential is the simple sum of three elements: the two slabs *udv* and *vdu* plus the tiny section in the corner, *du dv*, or

$$d(uv) = udv + vdu + du\,dv\ .$$

Leibniz, more or less by intuition, eliminates the last term, thus deriving the fundamental equation of differential calculus. It is the masterstroke of genius, although it was not appreciated at the time. The formal proof is written by Augustin-Louis Cauchy more than a century later. This often is cited as the date that rigor (some say rigor mortis) became an integral part of mathematics.

After elimination of the last term, the percentage increase in area can be obtained by dividing the differential by the original area,

$$\frac{d(uv)}{uv} = \frac{dv}{v} + \frac{du}{u}\ .$$

Analysis of this equation shows that the percentage increase in area is the sum of the percentage increases of the two sides (which is a very welcome discovery, because it introduces linearity into the calculations by replacing a product with a sum). The obvious counterpart for this equation in elasticity theory is

$$\Delta = e_{xx} + e_{yy} \ ,$$

because the definition of normal strain is the ratio of change in length to the original length. Extension to 3D is straightforward,

$$\Delta = e_{xx} + e_{yy} + e_{zz} \ ,$$

which says that the dilatation (the ratio of the increase in volume of a small piece of rock to the original volume) is equal to the sum of the three normal strains.

Adding the dilatation to Hooke's law results in a formula for normal stress with the basic form

$$\sigma_{xx} = 2\mu e_{xx} + (\text{constant})\ \Delta \ .$$

This new constant is defined as the other previously mentioned Lamé constant. Hooke's law for the three normal stresses thus reads

$$\sigma_{xx} = 2\mu e_{xx} + \lambda\Delta \ ,$$

$$\sigma_{yy} = 2\mu e_{yy} + \lambda\Delta \ ,$$

$$\sigma_{zz} = 2\mu e_{zz} + \lambda\Delta \ .$$

The next sections will explain how the two types of seismic wave motion, longitudinal and shear, are extracted from Hooke's law.

Cartesian fields of dilatation

Throughout history, there has always been a struggle in the human mind between the discrete and the continuous. This difference in viewpoint climaxed in the 17th century because of a fundamental conflict in the cosmologies of Descartes and Newton. The 17th century was arguably the century of scientific genius: Kepler, Napier, Bacon, Galileo, Descartes, Fermat, Pascal, Huygens, Hooke, Boyle, Leeuwenhoek, Newton, and Leibniz.

Descartes and Newton stand apart from this pantheon of scientific immortals. The former was the key figure in establishing modern science; the latter endowed it with concepts that were so consistently successful that science soon gained the position of respect—if not outright awe—that it has held ever since.

Both Descartes and Newton saw the necessity of introducing mathematics into physics; and both are ranked with the greatest mathematicians, even though both treated mathematics as a tool to further physical investigations and not as an end in itself.

Their major difference was in style. Descartes gave total freedom to his imagination which usually outran contemporary experimental results and mathematical knowledge. Newton operated at the other extreme, exercising iron control over his scientific intellect. This was perfectly exemplified in his declaration, *Hypotheses non fingo*, usually translated, “I do not form hypotheses.” Newton was referring to the distinction between unfounded hypotheses and experimental evidence. In *Principia* (Book 3, Rule 3), Newton wrote, “We are certainly not to relinquish the evidence of experiments for the sake of dreams and vain fictions of our own devising.”

The major difference in the systems proposed by Newton and Descartes was caused by the controversial concept of “action at a distance.” Newtonian science conceded that interaction between discrete and separated particles must be occurring. An example—indeed, the fundamental example upon which his entire system rests—was gravitational attraction; another was electrical attraction between two charged particles. Cartesian science, conversely, did not allow the possibility of discrete particles interacting through empty space; instead, it filled all space with an ethereal substance which acted on bodies (as a chip of wood is carried about by an eddy in a pool). In this system, the force of gravity on a planet was but a manifestation of a much different reality—the sweeping of the planet through space by a Cartesian vortex.

Contemporary debate over these contrasting views was intense and was put into stark relief by Voltaire when he wrote in *Lettres Anglaises*, “A Frenchman who arrives in London finds a great change in philosophy, as everything else. He left the world full, he finds it empty. In Paris one sees the universe composed of vortices of subtle matter. In London one sees none of this.”

The debate lasted almost a century, but Newtonian physics, about 1740, ultimately gained universal acceptance in Europe’s scientific community. However, this situation was not permanent. Descartes’ ideas reemerged in a different guise in the middle of the 19th century. The researches of Michael Faraday and James Clerk Maxwell on electromagnetism led to experimental results which could only be explained by the introduction of electromagnetic fields.

The fields, like Cartesian vortices, fill all space. From a physical point of view, the concept of a field is necessary in our attempt to visualize how one body affects another. Instead of assuming that separated bodies can interact without anything transpiring through the finite distance between them, we imagine that an object creates (or is surrounded by) a field. Any other object which contacts this field is affected by it. The type of field is determined by the physical characteristics of the object from which it emanates.

The notion of a field is extremely important in geophysics. Consider the top of a stove with one burner turned on. The heat flows from the burner to other points on the top of the stove. The temperature distribution on the stove top represents a field, and the temperature at any point on the stove top is a function of the spatial coordinates x, y and the time coordinate t. Because temperature is a scalar quantity (possessing magnitude but not direction), the set of all temperatures forms a scalar field. The temperature distribution in the solid earth represents a scalar field with three spatial coordinates and a time coordinate. A familiar scalar field is a topographic map, the field being the elevation of the surface of the earth above mean sea level.

The topographic map is also a convenient way to introduce an important concept, one that is familiar to everyone who has climbed a mountain. This is the gradient, abbreviated grad. The gradient is the scalar field manifestation of the mathematical operation known as differentiation in calculus. As a climber goes up the mountain, he experiences the rate of change of elevation. The gradient is defined as the vector (because it has both magnitude and direction) which points in the steepest direction (i.e., the direction in which the rate of change is greatest). The magnitude of this vector is the rate of change of the elevation along this path. Thus, if the climber continually follows the gradient, he will take the shortest path to the top of the mountain.

An important property of the gradient is that its direction is always perpendicular to the contour curve. This is extremely useful because it means that the gradient is related to the contour curve in a way that is completely independent of any particular coordinate system. For example, if one topographic map is made with respect to true north and another with respect to magnetic north, the gradient on each is the same.

The map of the gradients at all of the points forms a vector field (because, of course, each gradient is a vector). Geophysics routinely addresses many different types of vector fields, such as the flow of water in the ocean, gravity fields, electromagnetic fields, and seismic wavefields.

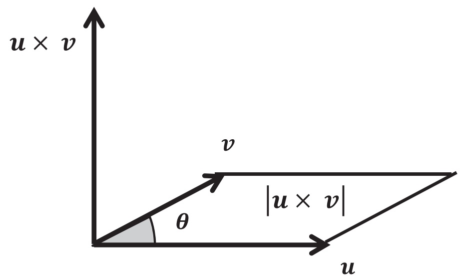


Figure 23. The cross product.

A physical quantity is called a vector if and only if (1) it has a numerical magnitude, (2) it has a direction in space, and (3) it obeys the parallelogram rule for addition. Three examples of vectors are displacement, velocity, and acceleration of a particle. In a Cartesian (x, y, z) frame, the three base vectors are $\boldsymbol{i}, \boldsymbol{j}$, and $\boldsymbol{k}$. They point in the x, y, z directions, respectively. The magnitude of each base vector is one. A vector is depicted by a pointer having its direction and magnitude. If the vector pertains to point P, the tail of the vector is placed at point P. A vector $\boldsymbol{u}$ is denoted by a boldfaced letter and its magnitude by the corresponding italic letter u; that is, $u = |\boldsymbol{u}|$. A three-dimensional vector field $\boldsymbol{u}(\boldsymbol{r})$ is a function of position

$$\boldsymbol{r} = x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k} = (x, y, z) \ .$$

Let $\boldsymbol{u} = u_1\boldsymbol{i} + u_2\boldsymbol{j} + u_3\boldsymbol{k}$ and $\boldsymbol{v} = v_1\boldsymbol{i} + v_2\boldsymbol{j} + v_3\boldsymbol{k}$ be two vectors (see Figure 23). Let the angle between them be θ. The *dot product* (or scalar product or inner product) is defined as

$$\boldsymbol{u}\cdot\boldsymbol{v} = u_1v_1 + u_2v_2 + u_3v_3 = uv\cos\theta \ .$$

Unless $\boldsymbol{u}$ and $\boldsymbol{v}$ are parallel, they determine a plane. The cross product is perpendicular to the plane defined by the two vectors. Let θ be the angle between the two vectors. Let $\boldsymbol{n}$ be the unit vector perpendicular to this plane and pointing in the right-side direction. Then, the *cross product* (or vector product) of $\boldsymbol{u}$ and $\boldsymbol{v}$ in that order is defined as

$$\boldsymbol{u}\times\boldsymbol{v} = \boldsymbol{n}\,uv\sin\theta = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} \ .$$

The magnitude of the cross product is

$$|\boldsymbol{u}\times\boldsymbol{v}| = uv\sin\theta \ .$$

Thus, the cross product provides the area of the parallelogram spanned by the two vectors.

Taking the two vectors, we can write every combination of components in the grid shown below.

	v_1	v_2	v_3
u_1	u_1v_1	u_1v_2	u_1v_3
u_2	u_2v_1	u_2v_2	u_2v_3
u_3	u_3v_1	u_3v_2	u_3v_3

The elements of the dot product are along the diagonal. The elements of the cross product are off the diagonal. The dot product measures similarity because it only accumulates auto-interactions. The cross product measures cross-interactions.

Define the vector ∇ given by

$$\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right) .$$

For a scalar function ϕ, the gradient is defined as the vector

$$\nabla\phi \equiv \text{grad } \phi \ = \left(\frac{\partial\phi}{\partial x}, \frac{\partial\phi}{\partial y}, \frac{\partial\phi}{\partial z}\right) .$$

The differential $d\boldsymbol{r}$ is

$$d\boldsymbol{r} = (dx, dy, dz) .$$

We have

$$\nabla\phi \cdot d\boldsymbol{r} = \frac{\partial\phi}{\partial x}dx + \frac{\partial\phi}{\partial y}dy + \frac{\partial\phi}{\partial z}dz = d\phi .$$

Let $ds = |d\boldsymbol{r}|$. Thus, $d\boldsymbol{r}/ds$ is a unit vector, which is called the unit tangent vector. The above equation produces

$$\frac{d\phi}{ds} = \nabla\phi \cdot \left(\frac{d\boldsymbol{r}}{ds}\right) .$$

Because $d\boldsymbol{r}/ds$ is a unit vector, it follows that $\nabla\phi \cdot (d\boldsymbol{r}/ds)$ is the component of $\nabla\phi$ in the direction of $d\boldsymbol{r}$. Thus, the gradient of ϕ is a vector whose component in any direction is the derivative of ϕ in that direction. At a given point, the component $d\phi/ds$ will have its greatest value when $d\boldsymbol{r}$ is in the same direction as the gradient. Thus, the gradient extends in the direction in which the derivative of ϕ has its maximum value.

Through each point P in the vector field, there passes a surface given by $\phi(x, y, z) =$ constant. On this surface,

$$\nabla\phi \cdot d\boldsymbol{r} = d\phi = 0 \ .$$

This result shows that the gradient is perpendicular to every vector $d\boldsymbol{r}$ on the surface at point P. Consequently, the gradient is normal to the surface.

Corresponding to the dot product and the cross product are the divergence and curl, respectively. (Note: European authors often use the notation rot $\boldsymbol{u}$ instead of the notation curl $\boldsymbol{u}$ used here.) For a vector $\boldsymbol{u} = (u_1, u_2, u_3)$, the divergence and curl are, respectively,

$$\nabla \cdot \boldsymbol{u} \equiv \operatorname{div} \boldsymbol{u} = \frac{\partial u_1}{\partial x} + \frac{\partial u_2}{\partial y} + \frac{\partial u_3}{\partial z} \ ,$$

$$\nabla \times \boldsymbol{u} \equiv \operatorname{curl} \boldsymbol{u} = \begin{vmatrix} i & j & k \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ u_1 & u_2 & u_3 \end{vmatrix} \ .$$

The normal and tangential components of a vector may be defined with respect to either a line or a surface. Usually, we will make use of a mixture; that is, the unit vector $\boldsymbol{n}$ that is normal to a surface, and the unit vector $\boldsymbol{t}$ that is tangent to a line.

Let $\boldsymbol{u}$ be a vector field and let C be a curve in three-dimensional space. The position vector $\boldsymbol{r}$ is defined as the vector from the origin O to a point P on the curve. The *differential* vector $d\boldsymbol{r}$ is called the *differential distance vector* at point P. The length of the differential vector $d\boldsymbol{r}$ is the differential arc length ds along the curve at P. The differential vector $d\boldsymbol{r}$ also gives the direction of the curve at P. The unit tangent vector $\boldsymbol{t}$ to the curve at point P is specified by

$$d\boldsymbol{r} = \frac{d\boldsymbol{r}}{ds} ds = \boldsymbol{t}\, ds \ .$$

In what follows, we will utilize the scalar differential

$$\boldsymbol{u} \cdot d\boldsymbol{r} = (\boldsymbol{u} \cdot \boldsymbol{t})\ ds \ .$$

Let θ denote the angle between $\boldsymbol{u}$ and $\boldsymbol{t}$. Because $|\boldsymbol{t}| = \mathbf{1}$, this differential may be written as

$$\boldsymbol{u} \cdot d\boldsymbol{r} = |\boldsymbol{u}||\boldsymbol{t}| \cos\theta\, ds = u_t\, ds \ .$$

In other words, the differential $\boldsymbol{u} \cdot d\boldsymbol{r}$ is equal to the product of the component u_t of $\boldsymbol{u}$ in the direction of the curve at P and the differential length ds.

The *line integral* of $\boldsymbol{u}$ taken along the curve between two specified end points P_0 and P_1 is defined as

$$\int_{P_0}^{P_1} \boldsymbol{u} \cdot d\boldsymbol{r} = \int_{P_0}^{P_1} (\boldsymbol{u} \cdot \boldsymbol{t})\, ds \ .$$

Line integral is a generic term which applies to any vector field, but the name *work integral* actually derives from the special case in which the field is a force field. For the moment, let $\boldsymbol{u}$ represent force. The work done by the force on the specified path from points P_0 to P_1 is the work integral

$$W = \int_{P_0}^{P_1} \boldsymbol{u} \cdot d\boldsymbol{r} \ .$$

The work integral evaluates the integral of the component of $\boldsymbol{u}$ along the line. This usage of the word "line" includes a curved line as well as a straight line. Because $d\boldsymbol{r} = \boldsymbol{t}\, ds$, we see that

$$W = \int_{P_0}^{P_1} \boldsymbol{u} \cdot \boldsymbol{t}\, ds \ .$$

Because $|\boldsymbol{t}| = 1$, the dot product is

$$\boldsymbol{u} \cdot \boldsymbol{t} = |\boldsymbol{u}|\, |\boldsymbol{t}| \cos\theta = |\boldsymbol{u}| \cos\theta = u_t \ .$$

Here, θ is the angle between the two vectors and u_t is the component of $\boldsymbol{u}$ on the tangent to the line. Thus, the work integral is

$$W = \int_{P_0}^{P_1} u_t\, ds \ .$$

The line integral is a summation of contributions along any arbitrary curve (not excluding a closed curve) in a vector field. Each contribution is the product of the vector component in the direction of the path multiplied by the path length. When the vector represents force, the line integral is called a work integral because it gives the total amount of work performed in traveling along that curve. But the usefulness of the line integral is by no means confined to cases where the vector field is a force field. For example, in fluid dynamics, the vector may be velocity (so it is a velocity field). If the curve is closed, the work integral actually defines the *circulation* around that curve. The work integral is a scalar quantity because work is a scalar.

The value of the work integral depends, in general, on the curve C but not on the coordinate system used in the evaluation.

Generally, the value of the line integral between the given points depends upon the path chosen. If two different paths join the same end points, the line integral along one path does not have to be the same as the line integral along the other path. However, suppose that $\boldsymbol{u}$ is the gradient of a function ϕ; that is, $\boldsymbol{u} = \nabla\phi$. Then,

$$\boldsymbol{u} \cdot d\boldsymbol{r} = d\boldsymbol{r} \cdot \boldsymbol{u} = d\boldsymbol{r} \cdot \nabla\phi = \boldsymbol{d}\boldsymbol{\phi} \ .$$

Thus, the line integral becomes

$$\int_{P_0}^{P_1} \boldsymbol{u} \cdot d\boldsymbol{r} = \int_{P_0}^{P_1} d\phi = \phi(P_1) - \phi(P_0) \ .$$

This shows that, if $\boldsymbol{u} = \nabla\phi$, then the line integral depends only upon the end points. In such a case, the line integral about a closed path is zero. Conversely, if the line integral about a closed path is zero for every closed path, then it can be shown that there is a function ϕ such that $\boldsymbol{u} = \nabla\phi$. The line integral around a closed path is called the *circulation*. Thus, the circulation is zero if and only if $\boldsymbol{u} = \nabla\phi$. Such a function ϕ is called the potential. Path independence of the line integral is equivalent to the vector field being *conservative*. It is an identity in vector calculus that, for any scalar function ϕ, it is true that $\nabla \times \nabla\phi = 0$. A field with zero curl is called *irrotational*. Thus, a conservative vector field is irrotational.

Using Figure 24, let γ be the angle between the unit normal vector $\boldsymbol{n}$ and a vector $\boldsymbol{u}$. The normal component of $\boldsymbol{u}$ relative to the surface S is

$$u_n = \boldsymbol{u} \cdot \boldsymbol{n} = u \cos\gamma \ .$$

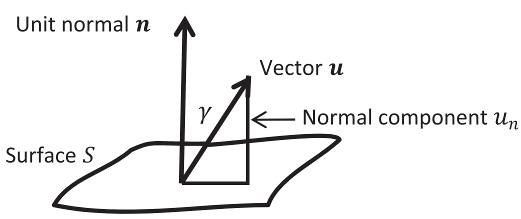


Figure 24. Normal component relative to the surface.

Next, using Figure 25, let θ be the angle between the unit tangent $\boldsymbol{t}$ and a vector $\boldsymbol{u}$. The tangential component of $\boldsymbol{u}$ relative to line C is

$$u_t = \boldsymbol{u} \cdot \boldsymbol{t} = u \cos\theta \ .$$

Two concepts from integral calculus—the flux integral and the work integral—are important. The remainder of this section defines the flux integral and its relationship to divergence. The next section will define the line integral and its relationship to curl. Ultimately, these concepts will be integrated into a discussion of Hooke's law to explain seismic wave propagation.

Moving to Figure 26, consider a small surface, such as a postage stamp of area dS. The word "flux" designates the flow of a field through a particular area. If the vector cuts through the stamp at an angle, the flux is defined as the area of the stamp times the component of the vector in the direction of the normal. In the extreme case, the vector is parallel to the stamp, so it does not cut through the stamp at all. In this case, the normal component is zero, and thus the flux is zero.

A school of fish is swimming along a flow line. If a net is placed perpendicular to the flow line, the most fish will be caught. It the net is placed oblique to the flow line, less fish will be caught. If the net is placed along

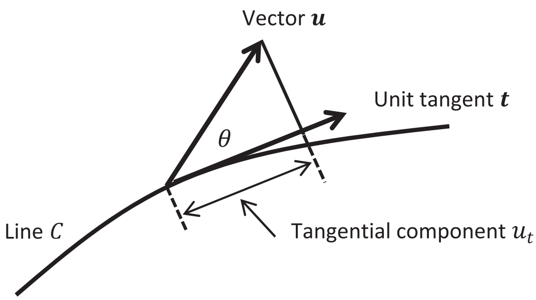


Figure 25. Tangential component relative to the line.

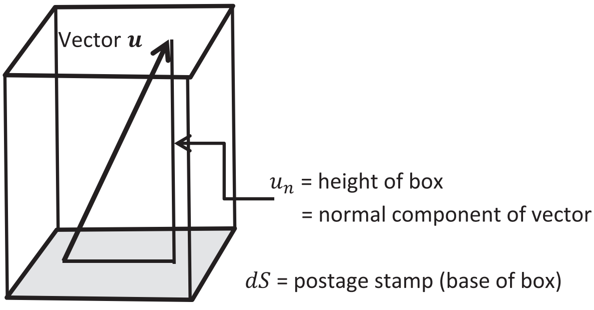


Figure 26. Case in which a vector cuts the postage stamp at an angle. Flux equals volume of box.

the flow line, no fish will be caught. However, suppose that the mesh of the net is so large that the fish pass right through the net. Instead of saying the number of fish caught, we say the number of fish that pass though the net. This number is called the *flux*. Flux is the amount of a quantity passing through a surface. The flux depends on the strength of the field, the size of the surface, and the orientation of the surface relative to the direction of the field. In other words, the flux depends upon the number of fish going along the flow line, the size of the net, and the orientation of the net relative to the direction of the flow line.

Let $\boldsymbol{u}$ be a vector field. Let $\boldsymbol{n}$ be a unit normal vector to the surface. Let the area of the surface be dS. The flux through the surface is determined by the component of $\boldsymbol{u}$ that is in the direction of $\boldsymbol{n}$, i.e., by $u_n = \boldsymbol{u} \cdot \boldsymbol{n} = u \cos \theta$. Thus, the flux is

$$\text{flux} = u \cos \theta \, dS = u_n \, dS \; .$$

The flux integral is a straightforward extension of this basic definition. Take any arbitrary surface, such as a geologic interface that has been bent and deformed, and divide its total area into a lot of contiguous postage stamps. The flux integral of a vector field through this interface is the summation of the fluxes for each postage stamp (as the limit of the area of each stamp becomes infinitesimal).

A surface S is closed if it is the entire boundary of a three-dimensional region. A sphere is a perfectly round geometrical object in three-dimensional space that has the surface of a completely round ball. In other words, the sphere is the surface, not the ball inside. A sphere and an egg shell are examples of a closed surface. A field line crossing S from inside to outside is said to be *leaving* the volume. A field line crossing S from outside to inside is said to be *entering* the volume. If N_1 lines are leaving and N_2 lines are entering, then $N_1 - N_2$ is called the net number of lines *emerging* from the volume. For convenience, we drop the word "net" used in this context. We have the statement

$$\text{outward flux of } u = \text{number of lines of } u \text{ emerging} \; .$$

The equation is

$$\text{outward flux of } u = \iint_S u_n \, dS = \iint_S \boldsymbol{u} \cdot \boldsymbol{n} \, dS \; .$$

In summary, if the surface is closed around a region, then the flux integral over the surface area yields the total flux of the field emerging out through the surface.

Divergence is the rate of flux expansion (positive divergence) or flux contraction (negative divergence). A positive divergence at a point means that the point is a *source*. A negative divergence means that the point is a *sink*. A divergence of zero means there is neither expansion nor contraction. The divergence of a vector is the ratio of the flux of a *closed surface* divided by the volume contained by the surface (as the limit of the volume becomes infinitesimal); that is,

$$\text{divergence of vector} = \frac{\text{flux}}{\text{volume}} \text{ as volume} \to 0 \ .$$

Divergence is a scalar quantity because both flux and volume are scalars. Let S be the surface and V be the region. In mathematical terms, the above word equation is

$$\operatorname{div} \boldsymbol{u} = \lim_{V \to 0} \frac{1}{V} \iint_S \boldsymbol{u} \cdot \boldsymbol{n} \, dS \ .$$

A very important result is developed straightforwardly from this definition. Assume that many small contiguous cells are crowded together inside a large egg. According to the definition just given, we see each cell satisfies the relationship

$$\text{flux} = \text{divergence} \times \text{volume} \ .$$

Now, add these equations over all of the infinitesimally-small interior cells that compose the volume inside the closed surface. The summation of the left side (i.e., the flux) is called the *flux integral*. This integral gives the outward flux through the closed surface because all of the interior fluxes cancel. The flux integral is

$$\iint_S \boldsymbol{u} \cdot \boldsymbol{n} \, dS \ .$$

The summation of the right side (i.e., divergence times volume) is called the *volume integral*. The volume integral is

$$\iint_V \operatorname{div} \boldsymbol{u} \, dV \ .$$

If we equate the above two integrals, we obtain

$$\iint_S \boldsymbol{u} \cdot \boldsymbol{n}\, dS = \iint_V \operatorname{div} \boldsymbol{u}\, dV$$

or

$$\iint_S u_n S = \iint_V \operatorname{div} \boldsymbol{u}\, dV \ .$$

This result is called the *divergence theorem* or *the flux theorem*. It also is known as *Gauss's law* or as a form of *Green's theorem*. This link between area and volume is of great importance in applied mathematics.

Figure 27 illustrates the divergence theorem, which is as follows. The surface integral of normal component times surface element (of closed surface) is equal to the volume integral of divergence times volume element (of region within the surface).

This result also maneuvers us into a position to adapt the fundamental idea of Descartes—a wavefield pervading all space—to seismology.

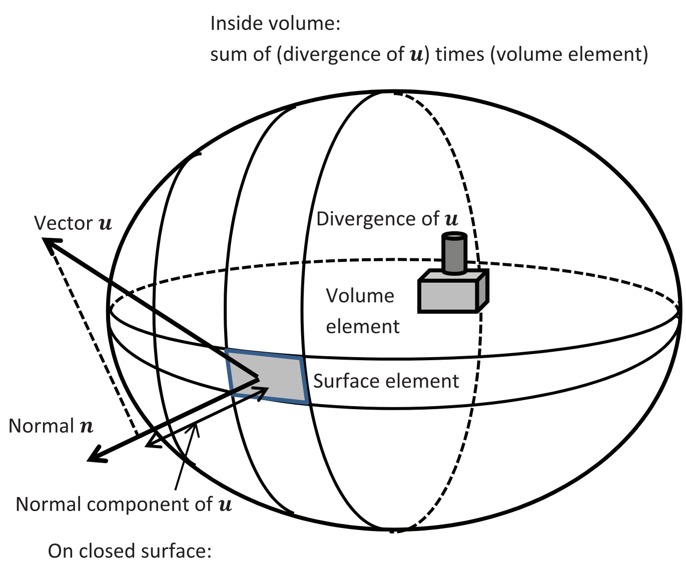


Figure 27. The divergence theorem illustrated—both sums (actually integrals) are the same.

We commonly think of a seismic wavefield extending through the subsurface rock layers. Its simplest form would occur in an unbounded elastic solid that is homogeneous (the same at all points) and isotropic (the same in all directions). In cases where the stresses on this solid are not in equilibrium, wave motion can result.

Unbalanced stresses cause a small particle to oscillate about its equilibrium position. This displacement is a vector (usually designated $\boldsymbol{u}$) which has length equal to the amount of movement and direction equal to the direction of displacement. Both quantities vary as time varies. The vector field $\boldsymbol{u}$ at all points and all times makes up the seismic wavefield, a direct application of the Cartesian description of nature.

Divergence is the vector form of the calculus concept of *dilatation*, and this relationship is a key to the vector treatment of wave motion. Consider the 1D case shown in Figure 28. All movement occurs along the x axis. The coordinate of point P in its undisplaced position is x, and its displaced coordinate is $x + u$. Thus, the quantity u is the displacement. In order to define the strain at point P, we must consider how its position relative to adjacent points has changed. Point Q is very close to P and has coordinate $x + \Delta x$ in its undisplaced position. Q's displacement is $u + \Delta u$. The Δu can be identified as the flux issuing from the 1D "volume" Δx, so the divergence of u can be written as

$$\operatorname{div} \boldsymbol{u} = \frac{\text{flux}}{\text{volume}} = \frac{\Delta u}{\Delta x} \ ,$$

but the last expression is the normal strain ε_{xx}. Therefore,

$$\operatorname{div} \boldsymbol{u} = \varepsilon_{xx} \ .$$

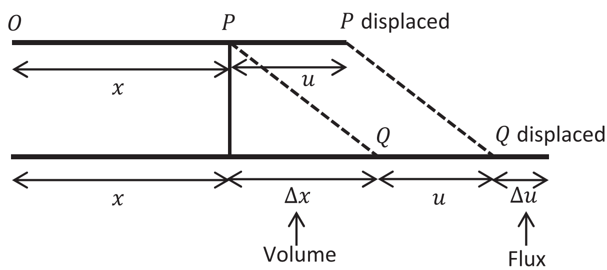


Figure 28. Displacements of points P and Q. Both of the above lines should be superimposed because P and Q lie on the same line, but the diagram is drawn this way for clarity.

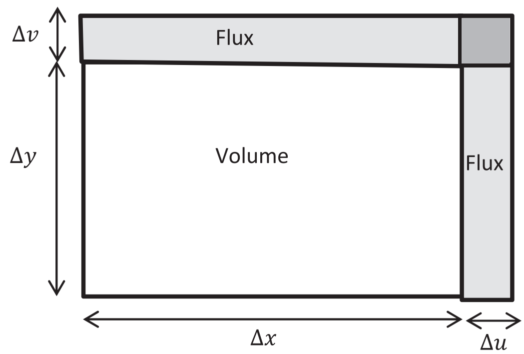


Figure 29. Flux and volume in the 2D case. Remember that Leibniz says we can forget the smallest area (upper right rectangle) as it is an infinitesimal of higher order.

The 2D case is shown in Figure 29. The flux is the dotted area, making the total flux

$$\Delta y\, \Delta u + \Delta x\, \Delta v \ .$$

The 2D "volume" is $\Delta x\, \Delta y$. Thus, the divergence of vector $\boldsymbol{u}$ (with components u and v) is

$$\operatorname{div} \boldsymbol{u} = \frac{\text{flux}}{\text{volume}} \ ,$$

which is

$$\operatorname{div} \boldsymbol{u} = \frac{\Delta y\, \Delta u + \Delta x\, \Delta v}{\Delta x \Delta y} = \frac{\Delta u}{\Delta x} + \frac{\Delta v}{\Delta y} \ .$$

However, the two ratios on the right are, respectively, the normal strains ε_{xx} and ε_{yy}. The sum of these strains is, by definition, the *dilatation* Δ and is equivalent to the divergence of the displacement vector $\boldsymbol{u}$. Similar reasoning extends the result to three dimensions, that is,

$$\operatorname{div} \boldsymbol{u} = \varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz} = \Delta \ .$$

In other words, div u gives the net amount by which a small box is being alternately compressed and stretched as a particle oscillates. Divergence, then, represents motion in the normal directions and none in the tangential directions. In the forthcoming section on equations of motion, we will show that, if this small box is in the earth, this dilatation propagates as wave motion. The resulting waves are known as compressional waves or longitudinal waves or P waves, and they are the most important type of wave used in seismic exploration for petroleum.

This is a stunningly practical resurrection of the Cartesian vortices, a concept that once seemed permanently exiled from respectable scientific discourse. Voltaire, in his preface to the French translation of Newton's *Principia*, writes: "If there were still somebody absurd enough to defend subtle and twisted (screw-formed) matter. . .that produces gravity, one would say: this man is a Cartesian; if he should believe in monads, one would say he is a Leibnizian. But there are no Newtonians, as there are no Euclideans. It is the privilege of error to give its name to a sect."

Voltaire's harsh judgment did not meet the test of time. Today, Euclidean geometry is but a sect of the more general non-Euclidean geometry, and Newtonian physics is but a sect of quantum physics. And, Newton's gravitational attraction through empty space is but a sect of Einstein's theory of gravitation (also known as general relativity). In Einstein's theory, space itself is curved by twisted and screw-formed matter, and this curvature is what causes gravitational attraction. This makes Einstein, in Voltaire's definition, a Cartesian.

Cartesian fields of rotation

The protracted struggle between the systems of Descartes and Newton transformed both of them into symbolic figures. One, Newton, embodied the ideal of modern and successful science, firmly based upon experimental data which is subjected to precise mathematical treatment. The other, Descartes, symbolized a reactionary and fallacious attempt to subject science to metaphysics, disregarding experiments and, indeed, replacing them by fantastic and unprovable hypotheses about the behavior of matter. The latter view was, of course, the one propagated by the Newtonians. The Cartesians stated the situation somewhat differently.

Rene Descartes (1596 – 1650) was an eminent French scientist and philosopher. His Cartesian system established the ideal of mathematical certitude in metaphysical demonstrations and, by brushing aside the then familiar scholastic subtleties, introduced modern philosophy and science of thought. Sir Isaac Newton (1642 – 1727) was a revered English natural philosopher and mathematician. He has been credited with the invention of calculus, the formulation of the laws of motion and the law of universal gravitation, and the discovery of the spectrum of light.

The Cartesians recognized the great superiority of Newtonian precision as compared to the Cartesian cosmology. But they rejected outright Newtonian attraction because they saw instantaneous action-at-distance as an occult quality or, even worse, as magic or miracle. They did not admit the existence of a perfectly void space, that is, the existence of "nothing" through which gravitational attraction was supposed to act. Descartes' teachings denied the existence of a void or vacuum and held that spatial extension and matter were identical.

Voltaire (*Letters,* c. 1779) wrote: "Geometry, which Descartes had, in a sense, created was a good guide and would have shown him a safe path to physics. But at the end he abandoned this guide and delivered himself to the spirit of the system. From then on, his philosophy became nothing more than ingenious romance. He created a world that existed only in his imagination and filled it with vortices of subtle matter, the speed of which some people even calculated."

The last comment was a jibe at Huygens, who was the first to calculate the actual speed of light. Voltaire's attack was probably prompted by Huygens' non-acceptance of Newton's gravitational theory. Huygens thought any theory that did not include a mechanical explanation was fundamentally flawed. His own ideas about gravity were based on Cartesian vortices.

Despite the disdain with which Voltaire and other Newtonians treated Cartesian vortices, the idea was not so ridiculous as their sneering suggests. Huygens was not the only great name who accepted it and attempted to extend the theory. Varignon, Leibniz, Kant, and Laplace were all influenced by the concept of cosmic vortices.

The wholesale shift of Descartes' thinking back into the forefront of physics began with the work of Leonhard Euler. This most prolific mathematician of all time developed the mechanics of fluid media, thereby providing an idealized mathematical model of the transmission of action in a continuous medium. His mathematics enabled those who thought of energy as something propagated across space to interpret this concept in a precise manner. Euler's work formed the basis of a new metaphor—the field. A field can be understood roughly as a region of space in which each point is characterized by quantities (scalar or vector) that are functions of the space coordinates x, y, z and time t.

Faraday provides the experimental evidence suggesting the existence of the field in the first half of the 19th century. Maxwell, shortly afterward, unifies light, electricity, and magnetism with a single formulation—the electromagnetic field—that is now a cornerstone of physical theory. In electromagnetism, the easiest circumstance to treat is the case in which nothing depends upon time—the static case. All charges are permanently fixed in space; or, if they do move, they move as a steady flow in a circuit (so both the charge density ρ and the current density $\boldsymbol{j}$ are constant). In this case, the electromagnetic field breaks into two fields that are not interconnected. The electric field $\boldsymbol{E}$ and the magnetic field $\boldsymbol{B}$ are distinct phenomena as long as charges and currents are static.

The electrostatic field, a vector field represented by a vector $\boldsymbol{E}$ which does not change with time, is due to electric charges represented by the charge density ρ. If a small region of space contains positive charges, then the field vectors $\boldsymbol{E}$ will emanate from that source; and if a small region contains negative charges, the field vectors will flow to that sink. Therefore, the source points of the field are those for which the divergence is greater than zero, and the sink points are those for which the divergence is less than zero. [Figure 55 in Chapter 2 shows field lines (solid) and equipotentials (dashes) for two equal and opposite point charges. The positive charge is an isolated source, and the negative charge is an isolated sink.]

Mathematically, the divergence of the electrostatic field $\boldsymbol{E}$ is

$$\operatorname{div} \boldsymbol{E} = \frac{\rho}{\epsilon_0} ,$$

where the sources are represented by the charge density ρ, and ϵ_0 is the electric constant. This formula is known as *Gauss's law* and is one of Maxwell's four equations. It also relates to the modern theory of seismic wave propagation (to which we shall return after the introduction of a complementary concept).

If the density of an electric current (designated j) through a wire is constant, a static magnetic field $\boldsymbol{B}$ will surround the wire (Figure 30), making loops around the currents. There is an electric charge, but there is no magnetic charge. Magnetic fields do not diverge from a source so, in every case,

$$\operatorname{div} \boldsymbol{B} = 0 .$$

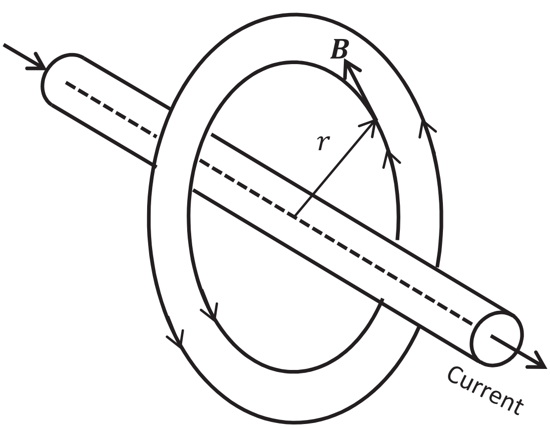


Figure 30. The magnetic field (only two lines illustrated) outside of an infinitely long straight wire carrying a constant electric current.

However, the magnetic field does have a rotation or curl that is proportional to the current density; that is,

$$\operatorname{curl} \boldsymbol{B} = \mu_0 \boldsymbol{j} \ ,$$

where μ_0 is the magnetic constant.

Curl is a concept which also can be applied to any vector field. An electrostatic field has sources but no rotation, so in every case

$$\operatorname{curl} \boldsymbol{E} = 0 \ .$$

To summarize these concepts—which are of major importance in geophysics—an electrostatic field $\boldsymbol{E}$ may be described as a vector field with a given divergence and zero curl, while a magneto-static field $\boldsymbol{B}$ has a given curl and zero divergence.

Curl is one of two basic differential operators used on vector fields (the other is divergence). Curl is best explained via the so-called *work integral*. In physics, the quantity known as "work" is defined as the product of the force component in the direction of a path and the length of the path. When the path is straight, this definition can be applied directly. If the path is curved, it is more complicated. The computation must take into account that the force may vary in both magnitude and direction, and that the path followed also may change in direction. The curve must be broken up into many small straight-line segments; then the work along each segment is computed and all of these are summed. This summation becomes the work integral (in the limit when each segment becomes infinitesimal). Let $\boldsymbol{u}$ be the field vector, let $\boldsymbol{r}$ be the vector from the origin to a point on the line, let ds be the differential of arc length, and $\boldsymbol{t}$ be the unit tangent vector of the line. The differential $d\boldsymbol{r}$ is defined by the equation

$$\frac{d\boldsymbol{r}}{ds} = \boldsymbol{t} \ .$$

Now, let $\boldsymbol{u}$ represent any vector field. Then more frequently, the work integral is called a *line integral*.

Let C_1 and C_2 be two closed curves with a common segment (Figure 31). Let the line integral be taken around each curve so that the common segment is traversed in opposite directions. Then, the contribution to the line integral for each curve is equal and opposite for the common segment. Thus, the sum of the two line integrals is equal to the single line integral over the closed curve which consists of the two original curves less their common segment.

Let us say a few words before we give the formal definitions of curl. Let us choose u to represent the velocity field of water, which we take to be incompressible. The water is swirling. Let us put a small paddle wheel in the water (Figure 32), which is free to rotate about its axis. The small wheel would start to spin because the impinging water would exert a net torque on the paddles. Its angular velocity will vary depending upon the location of the wheel and the positioning of its axis.

If the wheel turns, then the field has curl at that point. If it does not turn, then the field has zero curl at that point. If a field has zero curl

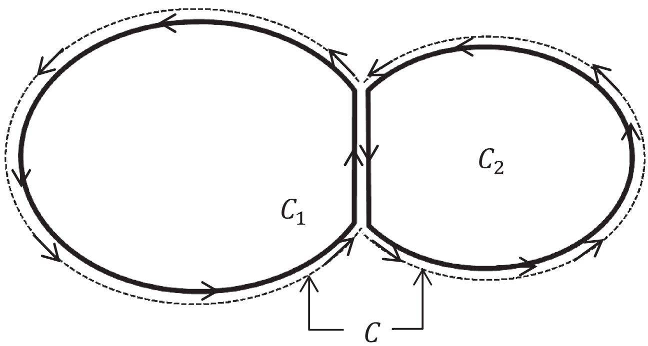


Figure 31. The coalescence of two closed curves C_1 and C_2 into a single closed curve C. The contributions to the work integral along the common segment cancel out because of the opposite senses of circulation along this segment.

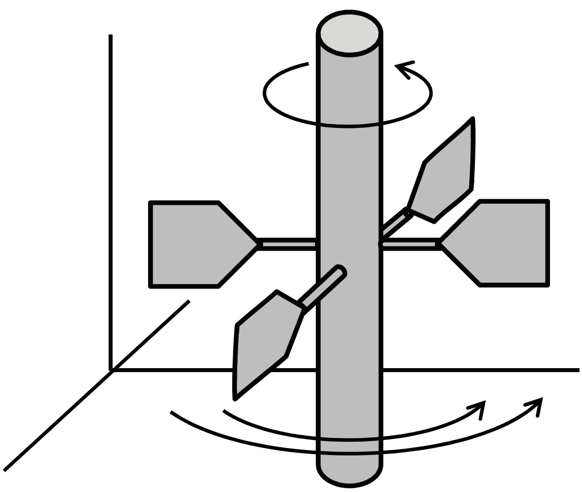


Figure 32. A small paddle wheel rotating in circularly rotating water.

everywhere, then the field is called *irrotational*. If the paddle wheel turns, the water is pushing harder on one side than the other, making it rotate. The greater is the rotation, then the bigger is the curl. For a conservative field, the work needed to move from point A to point B, along any path, is the same. Gravity is an example of a conservative field. The energy yielded by falling is the same as the energy required for lifting. Conservative fields have zero curl. If a field has non-zero curl, then it is not conservative.

The curl is a vector. It has magnitude and direction. The magnitude represents the amount of rotation at a point. How do you find the direction? You must orient the axis of the paddle wheel so that you achieve maximum rotation. In other words, the direction of the curl is the direction that results in the most rotation.

The line integral provides the means for defining the curl (see Figure 33). We take a point P within a vector field $\boldsymbol{u}$. The curl of $\boldsymbol{u}$ is also a vector, which is denoted by curl $\boldsymbol{u}$. At point P, we take a fixed unit vector $\boldsymbol{n}$ and consider any small surface element A with this fixed vector as the normal. This small surface element is bounded by a small closed curve C. Denote the angle between curl $\boldsymbol{u}$ and $\boldsymbol{n}$ by θ. The normal component (i.e., the component in the direction $\boldsymbol{n}$) of curl $\boldsymbol{u}$ is the scalar quantity given by

$$\mathrm{curl}_n\,\boldsymbol{u} = (\mathrm{curl}\,\boldsymbol{u}) \cdot \boldsymbol{n} = |\mathrm{curl}\,\boldsymbol{u}|\,|\boldsymbol{n}|\cos\theta = |\mathrm{curl}\,\boldsymbol{u}|\cos\theta\ .$$

The circulation represents how much the field pushes in traversing around the path. Accordingly, the circulation is the line integral of a vector $\boldsymbol{u}$ around the path. Let the path be the circumference of the small

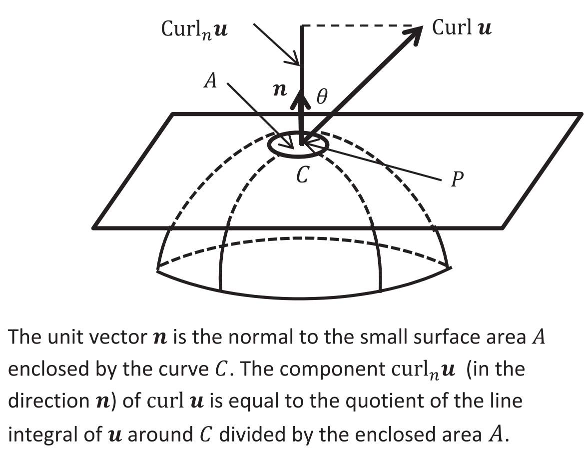


Figure 33. Geometry used to define the curl.

closed curve C. Thus, the circulation is

$$\text{circulation around the closed curve } C = \int_C u(\boldsymbol{r})d\boldsymbol{r} \ .$$

The normal component $\text{curl}_n\, \boldsymbol{u}$ is the projection of curl $\boldsymbol{u}$ onto the normal vector $\boldsymbol{n}$. It is important to remember that $\text{curl}_n\, \boldsymbol{u}$ is a scalar. The normal component is given by

$$\text{curl}_n\, \boldsymbol{u} = \frac{\text{circulation around the closed curve } C}{\text{area } A \text{ within closed curve } C} \ .$$

This definition holds in the limit when the area A becomes infinitesimally small. Mathematically, the above equation is

$$\text{curl}_n\, \boldsymbol{u} = \lim_{A \to 0} \frac{1}{A} \int_C \boldsymbol{u} \cdot \boldsymbol{t}\, ds \ .$$

The integrand is the tangential component to the vector; that is,

$$\boldsymbol{u} \cdot \boldsymbol{t} = |\boldsymbol{u}|\, |\boldsymbol{t}| \cos\theta = |\boldsymbol{u}| \cos\theta = u_t \ .$$

Thus, we may write

$$\text{curl}_n\, \boldsymbol{u} = \lim_{A \to 0} \frac{1}{A} \int_C u_t\, ds \ .$$

Now, we wish to give a heuristic proof of a very important relationship, called Stokes' theorem.

Using Figure 34, we subdivide the surface S into a set of small surface elements, each approximately rectangular. We will prove the theorem for

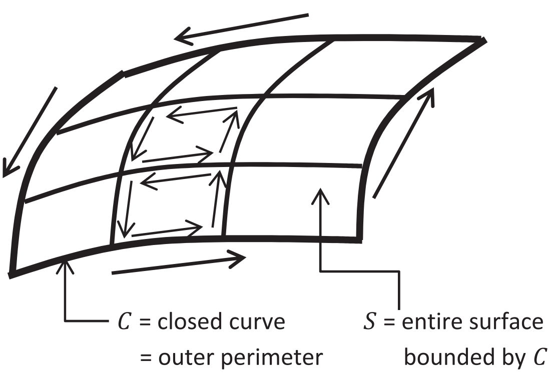


Figure 34. Surface S composed of a set of small rectangular elements.

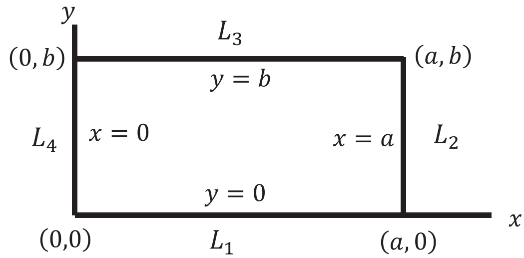


Figure 35. A separate small rectangle.

any of these separate small rectangles. This suffices to prove the theorem in general, because we can sum over all of the small rectangles. The sum of the surface integrals over the separate rectangles equals the surface integral over the entire surface. Because the line integrals over interior boundaries cancel in pairs, it follows that the sum of all of the line integrals equals the line integral around C.

Then, using Figure 35 to establish Stokes' theorem for a small rectangular area, we choose the coordinate axes so that the x and y axes are along the sides of the rectangle. The z axis is in the direction of the unit normal vector $\boldsymbol{n}$. Thus, $\boldsymbol{n} = \boldsymbol{k}$, where $\boldsymbol{k}$ is the unit vector in the z direction. Define the four line integrals for the four sides of the rectangle as

$$L_1 = \int_0^a u_1(x, 0)dx \ ,$$

$$L_2 = \int_0^b u_2(a, y)dy \ ,$$

$$L_3 = \int_a^0 u_1(x, b)dx \ ,$$

$$L_4 = \int_b^0 u_2(0, y)dy \ .$$

Then, the line integral around the entire small rectangle is

$$I = \int_C \boldsymbol{u} \cdot \boldsymbol{t}\, ds = \int_C u_t\, ds = L_1 + L_2 + L_3 + L_4 \ .$$

Next, we wish to evaluate the surface integral

$$J = \iint_S (\operatorname{curl} \boldsymbol{u}) \cdot \boldsymbol{n}\, dS \ .$$

The curl is

$$\operatorname{curl}\boldsymbol{u} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ u_1 & u_2 & u_3 \end{vmatrix} .$$

Because $\boldsymbol{n} = \boldsymbol{k}$, it follows that

$$(\operatorname{curl}\boldsymbol{u}) \cdot \boldsymbol{n} = (\operatorname{curl}\boldsymbol{u}) \cdot \boldsymbol{k} = \frac{\partial u_2}{\partial x} - \frac{\partial u_1}{\partial y} .$$

Therefore, the surface integral equals

$$J = \iint_S (\operatorname{curl}\boldsymbol{u}) \cdot \boldsymbol{n}\, dS = \int_0^b \int_0^a \left(\frac{\partial u_2}{\partial x} - \frac{\partial u_1}{\partial y} \right) dx\, dy .$$

We now split the surface integral into the difference of two integrals, choosing the order of integration differently in the two cases. We have

$$J = \int_0^b \int_0^a \frac{\partial u_2}{\partial x}\, dx\, dy - \int_0^a \int_0^b \frac{\partial u_1}{\partial y}\, dy\, dx .$$

This may be written as

$$J = \int_0^b \left[\int_0^a \frac{\partial u_2}{\partial x} dx \right] dy - \int_0^a \left[\int_0^b \frac{\partial u_1}{\partial y} dy \right] dx .$$

We integrate the terms in the square brackets to obtain

$$\left[\int_0^a \frac{\partial u_2}{\partial x} dx \right] = u_2(a, y) - u_2(0, y) ,$$

$$\left[\int_0^b \frac{\partial u_1}{\partial y} dy \right] = u_1(x, b) - u_1(x, 0) .$$

Thus, the surface integral becomes

$$J = \int_0^b [u_2(a, y) - u_2(0, y)] dy - \int_0^a [u_1(x, b) - u_1(x, 0)] dx .$$

This can be written as

$$J = \int_0^b u_2(a, y)dy - \int_0^b u_2(0, y)dy - \int_0^a u_1(x, b)dx + \int_0^a u_1(x, 0)dx \ .$$

This yields

$$J = \int_0^b u_2(a, y)dy + \int_b^0 u_2(0, y)dy + \int_a^0 u_1(x, b)dx + \int_0^a u_1(x, 0)dx \ .$$

This shows that the surface integral is

$$J = L_2 + L_4 + L_3 + L_1 \ .$$

Thus, we may conclude that the line integral I equals surface integral J; that is,

$$\int_C \boldsymbol{u} \cdot \boldsymbol{t} \, ds = \iint_S (\operatorname{curl} \boldsymbol{u}) \cdot \boldsymbol{n} \, dS \ .$$

In this equation, C refers to a closed curve in space, S refers to the surface that is bounded by the curve, ds refers to an element of arc length on the curve, and dS refers to an element of area on the surface. This result is known as *Stokes' theorem*, which can be stated as: *the line integral of* $\boldsymbol{u} \cdot \boldsymbol{t}$ *around a closed curve is equal to the surface integral of* $(\mathit{curl}\, \boldsymbol{u}) \cdot \boldsymbol{n}$ *over the area enclosed by the curve.*

We recognize $(\operatorname{curl} \boldsymbol{u}) \cdot \boldsymbol{n}$ as the normal component $\operatorname{curl}_n \boldsymbol{u}$. We recognize $\boldsymbol{u} \cdot \boldsymbol{t}$ as the tangential component u_t. Thus, Stokes' theorem also may be written as

$$\oint_C u_t \, ds = \iint_S \operatorname{curl}_n \boldsymbol{u} \, dS \ .$$

This equation states that the line integral of the tangential component u_t around the closed curve C equals the surface integral of the normal component $\operatorname{curl}_n \boldsymbol{u}$ over the bounded surface S.

Let us summarize. In the previous section, we have a closed surface S (like an unbroken egg shell) enclosing a volume V (like the inside of the egg). The *divergence theorem* is

$$\oiint_S \boldsymbol{u} \cdot \boldsymbol{n} \, dS = \iiint_V \operatorname{div} \boldsymbol{u} \, dV \ .$$

In this section, we have a closed curve C (like the perimeter of a blanket) enclosing a surface S (like the blanket). *Stokes' theorem* (which might well be called the *curl theorem*) is, as shown in Figure 36,

$$\oint_C \boldsymbol{u} \cdot \boldsymbol{t}\, ds = \iint_S \operatorname{curl} \boldsymbol{u} \cdot \boldsymbol{n}\, dS \ .$$

The preceding discussion may leave the reader with a feeling of not knowing exactly what the curl of a vector is. The fact that the curl has something to do with the work integral around a closed path suggests that the curl somehow describes Cartesian vortices which are rotating, swirling, or curling around. An item taken from fluid motion will help make these impressions clearer. (This example takes us back to the fundamental work of Euler on fluid mechanics.)

Suppose that a bucket of water is rotated about its axis at a constant angular velocity ω. This velocity vector points in the direction of the axis and has magnitude denoted by ω. At equilibrium, all of the water is rotating at this angular velocity, so that each small volume of water is undergoing circular motion (Figure 37). Consider a small element of volume at

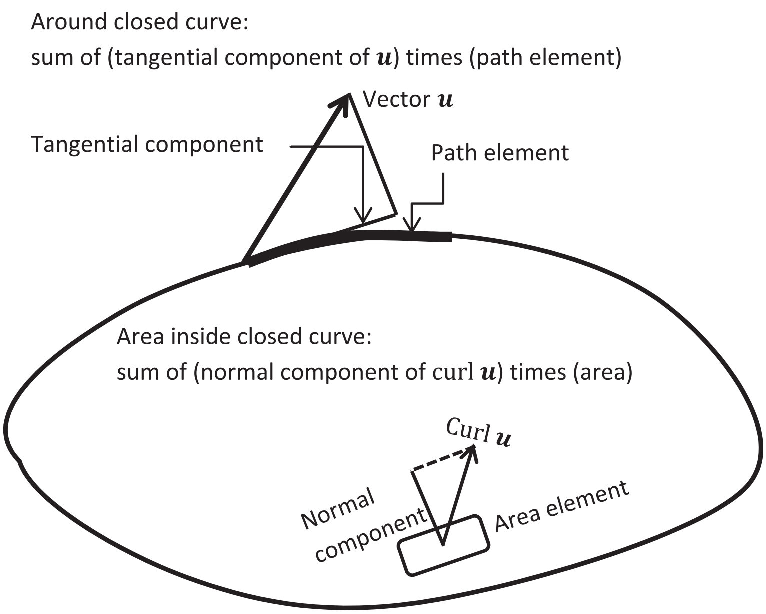


Figure 36. Stokes' theorem illustrated. The two sums (actually integrals) are equal.

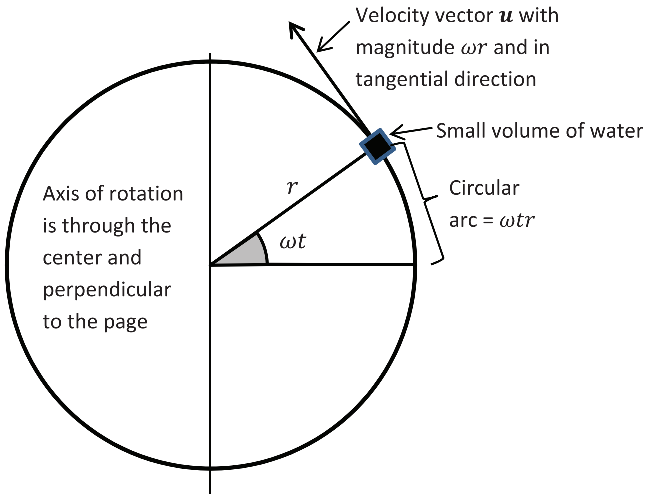


Figure 37. A small volume element of water. The element is rotating in a circle of radius r. The angular velocity vector $\boldsymbol{\omega}$ points in the direction of the axis of rotation, and the vector has magnitude denoted by ω.

radius $\boldsymbol{r}$. In time t, this small volume traverses a circular arc of angle ωt, so the circular distance traversed is ωtr. Thus, the magnitude of its velocity is the distance ωtr divided by t, or ωt. The direction of this velocity is the tangential direction. Denote the velocity vector with this magnitude and direction as $\boldsymbol{u}$.

Now, let us find the work integral of the velocity around this circle. Because the velocity vector is tangential to the circular path, the full magnitude ωr of the velocity is used to compute the work integral. The length of the circular path is $2\pi r$. Thus, the value of the work integral is magnitude times distance; that is, the value is $2\pi r^2\ \omega$.

Because all of the water in the bucket is rotating at a constant angular velocity, we would expect the curl of the velocity vector to be the same everywhere. This is indeed the case; that is,

$$\operatorname{curl} \boldsymbol{u} = \text{constant} \ .$$

Moreover, we would expect that the curl would point in the same direction as the axis of rotation, and this is also the case.

Now, let us consider the flux of the curl across the given circle of radius r. This flux integral is equal to the summation of the curl times the small area elements composing the circle. Because the curl is a constant, this summation reduces to simply the magnitude of the curl times the area of the circle,

$$\text{flux of curl}\,\boldsymbol{u} = |\text{curl}\,\boldsymbol{u}|(nr^2) \ .$$

Stokes' theorem says that the work integral and flux are equal; that is,

$$2\pi r^2 \boldsymbol{\omega} = \pi r^2 \,|\text{curl}\,\boldsymbol{u}| \ .$$

Because both $\boldsymbol{\omega}$ and curl $\boldsymbol{u}$ point in the same direction, this result gives (upon cancellation of πr^2)

$$\boldsymbol{\omega} = \frac{1}{2}\text{curl}\,\boldsymbol{u} \ . \qquad (1)$$

The curl of any vector $\boldsymbol{u}$ represents the tangential motion of the vector $\boldsymbol{u}$. Traditionally, this tangential motion is called the *rotation vector*. More specifically, the rotation vector $\boldsymbol{\omega}$ is defined by means of equation 1. The curl records the direction and magnitude of the maximum circulation at a given point.

The vector concepts of div and curl are discussed because they are an effective way to define a seismic wavefield. Particle motion is described by a displacement vector $\boldsymbol{u}$, which represents the oscillation of a tiny rock particle. The wavefield allows two kinds of motion—motion in the longitudinal direction (the stretching and shrinking of the particle) and motion in the transverse direction. Longitudinal motion is given by dilatation

$$\Delta = \text{div}\,\boldsymbol{u}$$

and rotational motion by the rotation

$$\boldsymbol{\omega} = \frac{1}{2}\text{curl}\,\boldsymbol{u} \ .$$

When these concepts are combined with Hooke's law (described in a previous section), it can be shown that dilatation Δ and rotation $\boldsymbol{\omega}$ can propagate separately, as distinct seismic waveforms.

Equations of motion

Oceanographers have long said that our planet should be called Water, not Earth, because dry land represents less than a quarter of planetary surface. Applying this same line of reasoning to the structure of all matter, we should call—according to the Newtonian perspective—the whole universe Vacuum. The stars are but tiny isles in the ocean of interstellar "near vacuum." Even the individual atoms making up matter are nothing but small nuclei with surrounding electrons enmeshed in a sea of vacuum.

But today, our view of "vacuum" is that of the Cartesians. This omnipresent medium, called "vacuum" since ancient times, is by no means emptiness or nothing. Vacuum influences everything it surrounds. Every experiment in elementary particle physics demonstrates the interaction of subatomic particles with one another and with the vacuum.

Johann Wolfgang von Goethe, whose great literary eminence has all but eclipsed his (not insignificant) scientific contributions, writes in Faust:

> Let us fathom it, whatever may befall,
> In this, thy Nothing, may I find my All.

Goethe's metaphor encompasses an insight of the great quantum physicist Paul Dirac, who realized that some physical objects reveal themselves only occasionally. An unexcited atom in a minimum-energy state does not radiate and, consequently, remains unobservable if not subjected to any action. Each elementary particle is but a manifestation of its own sea. The particle is unobservable until its sea is acted upon in a definite way. When a quantum of light gets into this "Dirac sea," the sea can eject out of itself an electron of negative energy. A multitude of conclusions have followed from Dirac's insight, including the discovery of the positron and other antiparticles.

During the past half century, the Dirac sea has turned into the ocean known as physical vacuum, and Dirac himself once said that the problem of describing vacuum was the primary one facing physicists. The currently prevailing description, the concept of fields pervading all space, has evolved from the once-derided Cartesian vortices. And, among the many scientific ideas dependent on this view is the mathematical foundation of the theory of seismic wave propagation.

In previous sections, we have discussed the mathematics of stress and strain for a medium in static equilibrium. This section will elaborate on that foundation to illustrate why wave motion can result when the stresses on a solid are not in equilibrium.

More than one mathematical technique can be used to support this assumption. In this section, we will use the tool commonly known as vector analysis which in large part is the creation of one of the most remarkable figures in modern science, J. Willard Gibbs.

Gibbs now is regularly cited as the greatest scientist ever produced in the United States. In fact, shortly after his death, historian Henry Adams called

Gibbs "the greatest of all Americans, judged by his rank in science." However, during his lifetime (1839 – 1901), he was all but unknown to U.S. scientific leaders. This resulted from a combination of curious circumstances: his own casual interest in recognition (e.g., he never joined the American Physical Society); a teaching style that was accessible to only the brightest graduate students; much more interest, by the contemporary U.S. scientific establishment, in immediately useful ideas rather than highly theoretical work (e.g., the immense fame of Gibbs' contemporaries Bell and Edison); and mathematics far beyond the abilities of most who read the lightly regarded *Transactions of the Connecticut Academy of Sciences* in which he published (in the 1870s) his first important papers. James Clerk Maxwell did recognize the importance of Gibbs' ideas; Maxwell personally made a physical model based on a Gibbs' concept and sent it to him. However, Maxwell died before he could convince other European scientists of the value of Gibbs' thought. It was not until the last few years of his life that Gibbs received richly deserved honors (notably the Copley Medal of the Royal Society) from the scientific community.

A greater honor came half a century later. Albert Einstein, shortly before his death, was asked to name the most powerful thinkers he had known. "Lorentz," Einstein answered without hesitation. Then, after some reflection, he added: "I never met Willard Gibbs; perhaps, had I done so, I might have placed him beside Lorentz."

The rigorous mathematics that Gibbs developed for thermodynamics and statistical mechanics are his greatest scientific achievements. This section, though, is built around still another Gibbs concept of monumental importance—vector analysis. Gibbs developed its present form more than a century ago—in *Elements of Vector Analysis*, first printed in 1881. Astonishingly, considering the surgical precision with which vector analysis can treat many complex physical situations, this idea was greeted with considerable open hostility. One reviewer called it "a hermaphrodite monster, compounded of the notations of Hamilton and Grassman." However, its usefulness was unanimously recognized by the turn of the century. Today, vector analysis so thoroughly permeates scientific literature that mathematician/historian Edna E. Kramer called this concept "indispensable to every serious student of physics."

Vectors are a most convenient means for mathematically describing and analyzing seismic waves.

Consider an unbounded elastic solid that is homogeneous (same at all points) and isotropic (same in all directions). If a disturbance passes through the material, the displacement of a small particle whose equilibrium position is the point $P(x, y, z)$ can be specified at any instant as a vector

$\boldsymbol{u}(x, y, z, t)$. This vector has its origin at the point of equilibrium and points in the direction of particle displacement. The length of the vector gives the amount of displacement. As time varies, the length and direction of the vector alter to represent the oscillation of a small particle about its equilibrium point.

Adapting basic vector concepts to the mathematical theory of elasticity requires the introduction of *vector fields*.

A *physical field* is a quantity which depends upon position in space. The simplest possible physical field is a scalar field; i.e., a field which is characterized at each point by a single number. (A good example in geophysics is the potential field representing the force of gravity. This field does not change with time. There are scalar fields which do change with time. Consider material, such as the solid earth, that has been heated at some places and cooled at others. The temperature of the body varies from point-to-point in a complicated way, and will be a function not only of position but also of time. This is an example of what is known as a time-dependent scalar field.)

In geophysics, *scalar fields* are depicted by means of contours, which are imaginary surfaces (in 3D) or lines (in 2D) drawn through all points for which the field has the same value. Contour lines originated, of course, on maps where they connect points with the same elevation. But contour lines also can be used in other areas, such as on a temperature field (where they are called *isothermal surfaces or isotherms*).

Vector fields, in contrast, are fields in which a vector is attached to each point in space. The flow of heat in the earth is an example. If the temperature is high at one place and low at another, there is a flow of heat from the warmer place to the colder. Thus, heat flow is a quantity which has direction. A scalar is not a sufficient mathematical description. However, heat flow can be represented at each point by a vector. This vector varies with both position and time. Its magnitude indicates the amount of heat flowing at any point at the designated time, and its direction indicates the direction of the flow.

When fields vary with time, the variation can be obtained by taking the derivative (or rate of change) with respect to time. Finding the rate of change with respect to position is trickier because there are three coordinates instead of one. Even so, it can be neatly handled by the concept of the *gradient* (which is defined as a vector which gives the rate of change of a field with respect to position). The *gradient vector* (abbreviated grad) is perpendicular to the contour line at which it originates, points uphill in the steepest direction, and possesses magnitude equal to the rate of change in that direction. The gradient has two particularly valuable properties—it indicates the direction and amount of the greatest rate of change at any particular point, and it is independent of any system of coordinate axes.

Another key concept in vector fields is that of the divergence of a vector (abbreviated div). Divergence, unlike gradient, is a scalar. To obtain a physical interpretation of divergence, consider a small box in space that is subjected to elastic wave motion. The displacement of a small particle (e.g., a micro-grain of sand) in that small box is denoted by the vector $\boldsymbol{u}$. As this particle oscillates, each of the six faces of the small box will undergo normal strain. The divergence of vector $\boldsymbol{u}$ is equal to the sum of the normal strains in the three coordinate directions. That is, div $\boldsymbol{u}$ indicates the net amount by which the small box is being alternatively stretched and compressed (in the normal directions) as the particles oscillate.

The dilatation Δ is obtained by the equation

$$\Delta = \operatorname{div} \boldsymbol{u} = \nabla \cdot \mathbf{u} = \nabla \cdot (u_x, u_y, u_z) = \frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y} + \frac{\partial u_z}{\partial z} \ .$$

Indeed, the divergence (denoted by div $\boldsymbol{u}$) and the dilatation (denoted by Δ) are merely different symbolizations for the same scalar.

The last concept that must be introduced is the rotation $\boldsymbol{\omega}$ of vector $\boldsymbol{u}$. The rotation is obtained by the equation

$$\boldsymbol{\omega} = 0.5 \operatorname{curl} \boldsymbol{u} \ .$$

The rotation $\boldsymbol{\omega}$ is also a vector. Consider a particle that is undergoing only shear strain, no normal strain; thus, this particle will be performing some kind of rotating motion in a plane at right angles to the direction of propagation.

These concepts—gradient, divergence, and curl—can be combined with ideas presented in previous sections to derive mathematical expressions which describe wave motion in elastic solids.

When a body is in equilibrium position, no net force is acting on any point in the body. When the body is deformed, however, the stresses (due to internal elastic forces of the body itself) exert restoring force which attempts to reestablish equilibrium. The more the distortion, the greater is the restoring force. As we have seen, the English mathematician Brook Taylor discovered that restoring force is proportional to spatial curvature in the case of a vibrating string. The greater the curvature, the greater is the restoring force. The following two identities and the two Laplacians will be used in our development.

First identity	divergence of curl of a vector is zero: $\mathrm{div}\ \mathrm{curl}\,\boldsymbol{u} = 0$ or $\nabla \cdot (\nabla \times \boldsymbol{u}) = 0$
Second identity	curl of gradient of scalar is zero: $\mathrm{curl}\ \mathrm{grad}\, f = 0$ or $\nabla \times (\nabla f) = 0$
Scalar Laplacian: used for dilatational wave equation	$\nabla^2 \Delta = \mathrm{div}\,\mathrm{grad}\,\Delta$ or $\nabla^2 \Delta = \nabla \cdot (\nabla \Delta)$
Vector Laplacian: used for rotational wave equation	$\nabla^2 \boldsymbol{\omega} = \nabla(\nabla \cdot \boldsymbol{\omega}) - \nabla \times (\nabla \times \boldsymbol{\omega})$ or $\nabla^2 \boldsymbol{\omega} = \nabla(\nabla \cdot \boldsymbol{\omega}) - \nabla \times (\nabla \times \boldsymbol{\omega})$

But what is spatial curvature in the case of a solid body? It can be represented by second derivatives of the particle displacement $\boldsymbol{u}$ with respect to the spatial coordinates (x, y, z). The gradient, divergence, and curl are each first derivatives; i.e., each represents a slope of one form or another. Second derivatives represent curvature, and there are only three independent second derivatives:

second derivative (i)	$\mathrm{grad}\ \mathrm{div}\,\boldsymbol{u}$	or	$\nabla\nabla \cdot \boldsymbol{u}$
second derivative (ii)	$\mathrm{curl}\ \mathrm{curl}\,\boldsymbol{u}$	or	$\nabla \times \nabla \times \boldsymbol{u}$
second derivative (iii)	$\mathrm{div}\ \mathrm{curl}\,\boldsymbol{u}$	or	$\nabla \cdot \nabla \times \boldsymbol{u}$

The quantity $\mathrm{div}\,\boldsymbol{u}$ (or dilatation) represents the change in volume of a small region (due to compression and stretching, the internal stresses). Thus, the second derivative (i), $\mathrm{grad}\ \mathrm{div}\,\boldsymbol{u}$, is the vector field which represents the directional slope of the dilatation.

Equation 1 in the previous section can be written as

$$\mathrm{curl}\,\boldsymbol{u} = 2\boldsymbol{\omega}\ . \tag{2}$$

Thus, the second derivative (ii) is

$$\mathrm{curl}\ \mathrm{curl}\,\boldsymbol{u} = \mathrm{curl}\,2\boldsymbol{\omega} = 2\,\mathrm{curl}\,\boldsymbol{\omega}\ . \tag{3}$$

Because $\mathrm{curl}\,\boldsymbol{u}$ involves no change in volume, the divergence of curl $\boldsymbol{u}$ is zero. Thus, the second derivative (iii) is

$$\mathrm{div}\ \mathrm{curl}\,\boldsymbol{u} = \mathrm{div}(2\boldsymbol{\omega}) = 2\,\mathrm{div}(\boldsymbol{\omega}) = 0\ . \tag{4}$$

Thus, the second derivative (iii) is not a consideration. We will make use of second derivative (i) and second derivative (ii); namely, $\mathrm{grad}\ \mathrm{div}\,\boldsymbol{u}$ and $\mathrm{curl}\ \mathrm{curl}\,\boldsymbol{u}$.

Now, returning to the reasoning of Taylor (force per unit of volume is proportional to spatial curvature), the relationship can be written as

$$\text{force density} = (\text{first constant})\,\text{grad div}\,\boldsymbol{u} + (\text{second constant})\,\text{curl curl}\,\boldsymbol{u} \ .$$

The result is that the first constant is $\lambda + 2\mu$ and the second constant is $-\mu$, where λ and μ are Lamé's constants. This determination of the unknown constants in the equation is from a direct application of Hooke's law. Thus, the above equation becomes

$$\text{force density} = (\lambda + 2\mu)\,\text{grad div}\,\boldsymbol{u} - \mu\,\text{curl curl}\,\boldsymbol{u} \ .$$

However, Newton's second law of motion equates force to the product of mass and acceleration. Thus, the force density is a product of mass density (denoted by ρ) and acceleration (written as a second derivative with respect to time). The above equation becomes

$$\rho\frac{\partial^2\boldsymbol{u}}{\partial t^2} = (\lambda + 2\mu)\,\text{grad div}\,\boldsymbol{u} - \mu\,\text{curl curl}\,\boldsymbol{u} \ , \tag{5a}$$

which in alternative notation is

$$\rho\frac{\partial^2\boldsymbol{u}}{\partial t^2} = (\lambda + 2\mu)\nabla(\nabla\cdot\boldsymbol{u}) - \mu\,\nabla\times(\nabla\times\boldsymbol{u}) \ . \tag{5b}$$

In seismology, this equation is known as the *equation of motion*. We will now show that the equation of motion leads to two different wave equations.

The equation of motion will be written in another form. The first term on the right involves div $\boldsymbol{u}$, which is the dilatation Δ; that is,

$$\nabla\cdot\boldsymbol{u} \equiv \Delta \ .$$

The second term on the right involves curl $\boldsymbol{u}$, which by equation 2 is

$$\text{curl}\,\boldsymbol{u} = 2\boldsymbol{\omega} \ .$$

By making these substitutions, the equation of motion 5 can be written as

$$\rho\frac{\partial^2\boldsymbol{u}}{\partial t^2} = (\lambda + 2\mu)\,\text{grad}\,\Delta - 2\mu\,\text{curl}\,\boldsymbol{\omega} \ , \tag{6a}$$

which in alternative notation is

$$\rho \frac{\partial^2 \boldsymbol{u}}{\partial t^2} = (\lambda + 2\mu)\nabla\Delta - 2\mu\, \nabla \times \boldsymbol{\omega} \ . \tag{6b}$$

The equation of motion explicitly displays the dilatation Δ and the rotation $\boldsymbol{\omega}$.

The dilatation $\Delta = \operatorname{div} \boldsymbol{u}$ represents a measure of change in volume of a small region without any rotation. Thus,

$$\operatorname{curl} \Delta = 0 \quad \text{or} \quad \nabla \times \Delta = 0 \ . \tag{7}$$

The rotation $\boldsymbol{\omega} = 0.5 \operatorname{curl} \boldsymbol{u}$ represents movement of a small region without any change in volume. Thus,

$$\operatorname{div} \boldsymbol{\omega} = 0 \quad \text{or} \quad \nabla \cdot \boldsymbol{\omega} = 0 \ . \tag{8}$$

We will keep helpful expressions 7 and 8 in mind.

Wave equation for dilatational waves (aka compressional waves). Take the divergence of the equation of motion in expression 6a. The term involving $\boldsymbol{\omega}$ vanishes, leaving

$$\rho \frac{\partial^2 \operatorname{div} \boldsymbol{u}}{\partial t^2} = (\lambda + 2\mu)\ \operatorname{div} \operatorname{grad} \Delta \ . \tag{9}$$

The scalar Laplacian is $\nabla^2\Delta = \operatorname{div} \operatorname{grad} \Delta$. In essence, the operator div grad is encountered so often that it has its own symbol, ∇^2, and its own name, the Laplacian. By using this symbol, the above equation becomes

$$\rho \frac{\partial^2 \Delta}{\partial t^2} = (\lambda + 2\mu)\nabla^2\Delta \ .$$

If we define α as

$$\alpha = \sqrt{\frac{\lambda + 2\mu}{\rho}} \ ,$$

the above equation becomes the wave equation

$$\frac{1}{\alpha^2} \frac{\partial^2 \boldsymbol{\Delta}}{\partial t^2} = \nabla^2\Delta \ . \tag{10}$$

In rectangular coordinates, this wave equation is

$$\frac{1}{\alpha^2}\frac{\partial^2 \Delta}{\partial t^2} = \frac{\partial^2 \Delta}{\partial x^2} + \frac{\partial^2 \Delta}{\partial y^2} + \frac{\partial^2 \Delta}{\partial z^2} \ . \tag{11}$$

This three-dimensional wave equation says that the dilatation Δ propagates with wave velocity α. These dilatational waves also are known as P-waves, compressional waves, or longitudinal waves.

Wave equation for rotational waves (aka shear waves). Take the curl of the equation of motion in expression 6a. The term involving Δ vanishes, leaving

$$\rho \frac{\partial^2 \mathrm{curl}\, \boldsymbol{u}}{\partial t^2} = -2\mu\, \mathrm{curl\ curl}\, \boldsymbol{\omega} \ . \tag{12}$$

The vector Laplacian is

$$\nabla^2 \boldsymbol{\omega} = \nabla(\nabla \cdot \boldsymbol{\omega}) - \mathrm{curl\ curl}\, \boldsymbol{\omega} \ ,$$

which yields

$$\mathrm{curl\ curl}\, \boldsymbol{\omega} = \nabla(\nabla \cdot \boldsymbol{\omega}) - \nabla^2 \boldsymbol{\omega} \ . \tag{13}$$

Equation 8 says that $\nabla \cdot \boldsymbol{\omega} = 0$. Thus, equation 13 becomes

$$\mathrm{curl\ curl}\, \boldsymbol{\omega} = -\nabla^2 \boldsymbol{\omega} \ .$$

Equation 2 says that $\mathrm{curl}\, \boldsymbol{u} = 2\boldsymbol{\omega}$. Thus, equation 12 becomes

$$\rho \frac{\partial^2 (2\boldsymbol{\omega})}{\partial t^2} = -2\mu(-\nabla^2 \boldsymbol{\omega}) \ ,$$

which is

$$\frac{\rho}{\mu}\frac{\partial^2 \omega}{\partial t^2} = \nabla^2 \omega \ . \tag{14}$$

If we define β as

$$\beta = \sqrt{\frac{\mu}{\rho}} \ ,$$

then equation 14 (in rectangular coordinates) becomes

$$\frac{1}{\beta^2}\frac{\partial^2 \omega}{\partial t^2} = \frac{\partial^2 \omega}{\partial x^2} + \frac{\partial^2 \omega}{\partial y^2} + \frac{\partial^2 \omega}{\partial z^2} \ . \tag{15}$$

This again is recognized as the three-dimensional wave equation, and the interpretation is that the rotation ω propagates with wave velocity β. These rotational waves also are known as S-waves, shear waves, or transverse waves.

The bottom line is that modern theory states there are two types of seismic body waves. One type results from the normal pulsations due to the compressional stretching/shrinking of a continuum of small rock particles; these generate P-waves traveling at velocity α. The second type results from tangential pulsations due to the rotation of a continuum of small rock particles; these generate S-waves traveling at velocity β.

We have completed the formidable task of deriving the equation of motion and thereby obtaining the wave equations for compressional waves and for shear waves. Readers who have followed the argument can feel a sense of accomplishment, for this material usually is covered in a graduate course in seismology.

Now, we want to find out what happens when a wave strikes an interface between two elastic media (see Figure 38). The various known constants are

$$\alpha_1 = \text{P-wave velocity in medium 1} \ ,$$

$$\beta_1 = \text{S-wave velocity in medium 1} \ ,$$

$$\alpha_2 = \text{P-wave velocity in medium 2} \ ,$$

$$\beta_2 = \text{S-wave velocity in medium 2} \ ,$$

$$\rho_1 = \text{density in medium 1} \ ,$$

$$\rho_2 = \text{density in medium 2} \ .$$

We assume that a plane P-wave with amplitude 1 is incident on the interface. The angles of reflection and transmission are given by Snell's law

$$\frac{\sin\phi_1}{\alpha_1} = \frac{\sin\psi_1}{\beta_1} = \frac{\sin\phi_2}{\alpha_2} = \frac{\sin\psi_2}{\beta_2} \ .$$

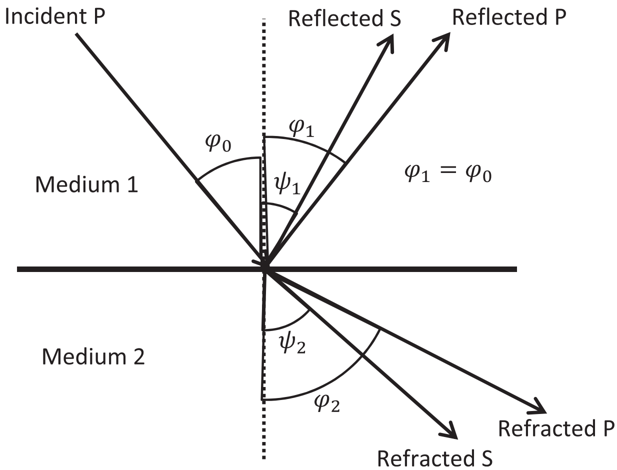


Figure 38. Incident P-wave and resulting reflected and refracted P- and S-waves.

In this equation,

$$\phi_1 = \text{angle of reflected P-wave in medium 1} \ ,$$

$$\psi_1 = \text{angle of refracted S-wave in medium 1} \ ,$$

$$\phi_2 = \text{angle of reflected P-wave in medium 2} \ ,$$

$$\psi_2 = \text{angle of refracted S-wave in medium 2} \ .$$

At the interface, the stresses and the displacements must be continuous. In other words, they cannot change abruptly at the interface. For example, if the normal stress was not the same on each side of the interface, there would be a net force that would produce an acceleration that would prevent equilibrium. Thus, the normal stress must be continuous at the interface. The same reasoning holds for the tangential stress. Also, the normal component of the displacement must be continuous at the interface. Otherwise, one medium either would separate from the other or else would penetrate the other. Likewise, the tangential component of the displacement must be continuous at the interface. Otherwise, one medium would slide over the other.

Karl Zoeppritz uses the continuity of normal and tangential stresses as well as the continuity of normal and tangential displacements in order to derive four equations. In this way, he obtains the dependence of the amplitudes of reflection and transmission coefficients upon the angle of incidence. Let

$$A_1 = \text{displacement of reflected P-wave in medium 1} \ ,$$

$$B_1 = \text{displacement of reflected S-wave in medium 1} \ ,$$

$$A_2 = \text{displacement of reflected P-wave in medium 2} ,$$

$$B_2 = \text{displacement of refracted S-wave in medium 2} .$$

The Zoeppritz equations can be displayed in matrix form as

$$Mx = y ,$$

where

$$M = \begin{pmatrix} \cos\phi_1 & \frac{\alpha_1}{\beta_1}\sin\psi_1 & \frac{\alpha_1}{\alpha_2}\cos\phi_2 & -\frac{\alpha_1}{\beta_2}\sin\psi_2 \\ -\sin\phi_1 & \frac{\alpha_1}{\beta_1}\cos\psi_1 & \frac{\alpha_1}{\alpha_2}\sin\phi_2 & \frac{\alpha_1}{\beta_2}\cos\psi_2 \\ -\cos 2\psi_1 & -\sin 2\psi_1 & \frac{\rho_2}{\rho_1}\cos 2\psi_2 & -\frac{\rho_2}{\rho_1}\sin 2\psi_2 \\ \sin 2\phi_1 & -\frac{\alpha_1^2}{\beta_1^2}\cos 2\psi_1 & \frac{\rho_2}{\rho_1}\frac{\beta_2^2}{\beta_1^2}\frac{\alpha_1^2}{\alpha_2^2}\sin 2\phi_2 & \frac{\rho_2}{\rho_1}\frac{\alpha_1^2}{\beta_1^2}\cos 2\psi_2 \end{pmatrix} ,$$

$$x = \begin{pmatrix} A_1 \\ A_2 \\ B_1 \\ B_2 \end{pmatrix}, \quad y = \begin{pmatrix} \cos\phi_1 \\ \sin\phi_1 \\ \cos 2\psi_1 \\ \sin 2\phi_1 \end{pmatrix} .$$

The Zoeppritz equations can be solved for the four unknowns, namely,

1. the reflected compressional-wave amplitude A_1,
2. the reflected shear-wave amplitude B_1,
3. the transmitted compressional-wave amplitude A_2,
4. the transmitted shear-wave amplitude B_2.

The solution is

$$x = M^{-1} y .$$

We shall discuss the practical use of an approximate solution of the Zoeppritz equations for P-to-P reflection amplitudes to derive the AVO attributes. The study of the dependence of amplitude versus offset between source and receiver is called AVO analysis. It can be employed to estimate a rock's fluid content, porosity, density, and velocity, as well as to give shear wave information. Typically, amplitude decreases with offset, due to geometrical spreading, attenuation, and other factors. An

AVO anomaly is commonly expressed as increasing AVO in a sedimentary section. Often this is where the hydrocarbon reservoir is "softer" (lower acoustic impedance) than the surrounding shales. An AVO anomaly also can include examples where amplitude with offset falls at lower rates than the surrounding reflective events.

AVO is a tool to discover new hydrocarbon reservoirs and to define the extent and composition of existing hydrocarbon reservoirs. An increasing AVO curve is normally more pronounced in oil-bearing sediments and especially in gas-bearing sediments. A primary purpose of AVO is to detect hydrocarbon-filled sedimentary traps.

An important point to remember is that the existence of abnormal (rising or falling) amplitude anomalies sometimes can be caused by other factors, such as alternative lithologies and residual hydrocarbons in a breached gas column. Modeling of the petrophysical properties and a good understanding of the sedimentary succession is paramount for successful hydrocarbon detection using AVO. Not all oil and gas fields are associated with an obvious AVO anomaly and AVO analysis is by no means a failsafe method for gas and oil exploration.

In exploration seismology, we are interested in the angle dependency of the P to P reflections given by the coefficient A_1. The object is to estimate the elastic parameters of reservoir rocks from reflection amplitudes and relate these parameters to reservoir fluids.

The exact expression for A_1 derived from the solution of the Zoeppritz equations is complicated and not intuitive in terms of its practical use for inferring petrophysical properties of reservoir rocks. Instead, we shall use the approximation provided by Aki and Richards (1980) as the starting point for deriving a series of practical AVO equations. Now that we only need to deal with the P to P reflection amplitude A_1, we shall switch to the conventional notation by replacing A_1 with $R(\theta)$ as the angle-dependent reflection amplitude for AVO analysis.

The computer can easily solve the Zoeppritz equations for the reflected and refracted P-wave and S-wave amplitudes A_1, B_1, A_2, B_2. The most useful component is the angle-dependency of the P to P reflections given by the coefficient A_1. It can be used to estimate the elastic parameters of reservoir rocks from the observed reflection amplitudes. Then these parameters can be related to reservoir oil hydrocarbons.

Shuey recast P to P reflection amplitude in successive ranges of angle of incidence and arranged the P to P equation in those terms. His work led to practical developments in AVO analysis. By assuming that changes in elastic properties of rocks across the layer boundary are small and propagation angles are within the subscritical range, he approximated the exact expression for $R(\theta)$ by the approximate equation

$$R(\theta) \approx \left[\frac{1}{2}\left(\frac{\Delta\alpha}{\alpha}+\frac{\Delta\rho}{\rho}\right)\right]+\left[\frac{1}{2}\frac{\Delta\alpha}{\alpha}-\frac{\beta^2}{\alpha^2}\frac{\Delta\beta}{\beta}-2\frac{\beta^2}{\alpha^2}\frac{\Delta\rho}{\rho}\right]\sin^2\theta + \left[\frac{1}{2}\frac{\Delta\alpha}{\alpha}\right](\tan^2\theta-\sin^2\theta)\ .$$

This equation is known as Shuey's three-term AVO equation. In the equation,

$$\alpha = \text{average P-wave velocity} = (\alpha_1+\alpha_2)/2\ ,$$

$$\Delta\alpha = \text{difference of P-wave velocity} = (\alpha_2-\alpha_1)\ ,$$

$$\beta = \text{average S-wave velocity} = (\beta_1+\beta_2)/2\ ,$$

$$\Delta\beta = \text{difference of S-wave velocity} = (\beta_2-\beta_1)\ ,$$

$$\rho = \text{average density} = (\rho_1+\rho_2)/2\ ,$$

$$\Delta\rho = \text{difference of density} = (\rho_2-\rho_1)\ ,$$

$$\theta = \text{average of the P-wave incidence and transmission angles} = (\varphi_1+\varphi_2)/2\ .$$

The geophysicist observes changes in reflection amplitudes as a function of angle of incidence. The quantity $\Delta\alpha/\alpha$ describes the fractional change in P-wave velocity across the layer boundary and hence may be referred to as the P-wave reflectivity. The quantity $\Delta\beta/\beta$ describes the fractional change in the S-wave velocity across the layer boundary and hence may be referred to as the S-wave reflectivity. The quantity $\Delta\rho/\rho$ describes the fractional change in density across the layer boundary.

In other words, we wish to estimate the P-wave reflectivity $\Delta\alpha/\alpha$, the S-wave reflectivity $\Delta\beta/\beta$, and the fractional change in density $\Delta\rho/\rho$. We make these estimates from the data provided by the observed angle-dependent reflection amplitudes.

A matter of concern is that the *modeled* reflection amplitudes are a *function of angle of incidence*. However, the *observed* reflection amplitudes are available from CMP data as a *function of offset*. A need then arises either to transform the model equation for the reflection amplitudes from angle to offset coordinates or to actually transform the CMP data from offset to angle coordinates. While the first approach is theoretically appealing, the practical schemes are based on the latter approach.

The elastic property that is most directly related to angular dependence of reflection coefficient $R(\theta)$ is Poisson's ratio σ.

In Shuey's three-term AVO equation, the first term,

$$R_P = \left[\frac{1}{2}\left(\frac{\Delta\alpha}{\alpha} + \frac{\Delta\rho}{\rho}\right)\right] ,$$

is the reflection amplitude at normal incidence. At intermediate angles ($0 < \theta < 30^\circ$), the third term may be dropped, thus leading to a two-term approximation

$$R(\theta) = R_P + G \sin^2 \theta ,$$

where

$$G = \frac{1}{2}\frac{\Delta\alpha}{\alpha} - \frac{\beta^2}{\alpha^2}\frac{\Delta\beta}{\beta} - 2\frac{\beta^2}{\alpha^2}\frac{\Delta\rho}{\rho} \ .$$

In this section, we have seen how amplitude versus offset (AVO) describes the dependency of seismic amplitude with source-and-receiver offset. By means of AVO analyses, information as to fluid content, porosity, density, seismic velocity, shear wave properties, and hydrocarbon indicators can be estimated.

This page has been intentionally left blank

Chapter 4

Rays, Anisotropy, and Maxwell's Equations

Ray equation

Geometrical optics, geometrical acoustics, and geometrical seismology all are terms which describe the limiting case of wave theory when the wavelength is small compared with the linear dimensions of the region occupied by the wavefield. That is, geometric wave theory deals with methods which can be used when the wavelength is small compared to the scale of dimensions of the physical region in space.

There is a special type of vector called a *unit vector*. A unit vector has magnitude 1, with no units. A unit vector is represented by a letter marked with a circumflex, which appears above the letter. It often is referred to as "hat." For example, $\hat{\boldsymbol{t}}$ (read "t-hat") is a unit vector. In other places in this book, we have not used the "hat" notation for the three unit vectors $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$. Here, we must, because we want to use the symbol $\boldsymbol{k}$ for the wavenumber vector.

The three unit vectors for Cartesian coordinates are:

1. the unit vector $\hat{\boldsymbol{i}}$ pointing in the $+x$ direction,
2. the unit vector $\hat{\boldsymbol{j}}$ pointing in the $+y$ direction,
3. the unit vector $\hat{\boldsymbol{k}}$ pointing in the $+z$ direction.

Any vector can be expressed in terms of unit vectors. For example, the vector (x, y, z) can be expressed as $x\hat{\boldsymbol{i}} + y\hat{\boldsymbol{j}} + z\hat{\boldsymbol{k}}$.

We start with the Helmholtz equation

$$\nabla^2 u + k^2 u = 0 \ , \tag{1}$$

where $\boldsymbol{u}$ is the wavefield and k is the wavenumber. Both $\boldsymbol{u}$ and k depend upon the spatial coordinates and frequency ω. For simplicity, we consider the two-dimensional case with coordinates x and z, but everything can be easily extended to the case of three dimensions x, y, z. The wavenumber can be written as

$$k = \frac{\omega}{c} \ .$$

In this equation, the velocity c may depend upon x and z. As usual, a point with coordinates x and z can be represented by a vector $\boldsymbol{r}$, which we can write in the two equivalent forms given by

$$r = (x, z) \quad \text{or} \quad \boldsymbol{r} = x\hat{\boldsymbol{i}} + z\hat{\boldsymbol{k}} \ .$$

The wavefield $\boldsymbol{u}$ is a function of position and frequency, so we write u as

$$u = \boldsymbol{u}(\boldsymbol{r}, \omega) \ .$$

If c depends upon position ω, the medium is called *inhomogeneous*. If c depends on the direction of the wave, the medium is called *anisotropic*. At this point, let us emphasize that we only deal with *isotropic* media, so c does not depend upon the direction of the wave. However, we allow inhomogeneous media, so generally we write c as $c(\boldsymbol{r})$.

In the case of a homogeneous isotropic medium, the wave velocity c is a constant. The resulting solution of the Helmholtz equation is

$$u(\boldsymbol{r}, \omega) = A(\omega) \exp(-i\boldsymbol{k} \cdot \boldsymbol{r}) \ .$$

In this equation, the vector

$$\boldsymbol{k} = k_x\hat{\boldsymbol{i}} + k_z\hat{\boldsymbol{k}}$$

is the wavenumber vector and the amplitude $A(\omega)$ is a function of ω. The surfaces $-\boldsymbol{k} \cdot \boldsymbol{r} =$ constant are plane surfaces of constant phase.

The basic approximation of geometric wave theory begins with an attempt to find a solution of the Helmholtz equation of the form

$$u(\boldsymbol{r}, \omega) = A(\boldsymbol{r}, \omega) \exp(-iA(\boldsymbol{r}, \omega)) \ .$$

In this equation, the function $\phi(\boldsymbol{r})$, which we also can write as $\phi(x, z)$, is called the *phase*. As such, this equation is not an approximation, as any function can be written in such a form. The approximation lies in finding suitable expressions for the amplitude $A(\boldsymbol{r}, \omega)$ and phase $A(\boldsymbol{r}, \omega)$, which are valid in the case of small wavelengths.

Now, we want to substitute the above expression for $\boldsymbol{u}$ into the Helmholtz equation. Let us first evaluate the individual terms in the Helmholtz equation separately, and then we will put it all together. The first term is

$$\nabla^2 u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial z^2} \ .$$

Let us evaluate the second partial derivative with respect to x. We obtain the expression

$$\begin{aligned}\frac{\partial^2 u}{\partial x^2} &= \frac{\partial^2}{\partial x^2}(Ae^{i\phi}) = A\frac{\partial^2}{\partial x^2}e^{i\phi} + 2\frac{\partial A}{\partial x}\left(\frac{\partial}{\partial x}e^{i\phi}\right) + e^{i\phi}\frac{\partial^2 A}{\partial x^2} \\ &= A\frac{\partial}{\partial x}\left(i\frac{\partial}{\partial x}e^{i\phi}\right) + 2\frac{\partial A}{\partial x}\left(i\frac{\partial}{\partial x}e^{i\phi}\right) + e^{i\phi}\frac{\partial^2 A}{\partial x^2} \\ &= \left[iA\frac{\partial^2\phi}{\partial x^2} + i^2 A\left(\frac{\partial\phi}{\partial x}\right)^2 + 2i\frac{\partial A}{\partial x}\frac{\partial\phi}{\partial x} + \frac{\partial^2 A}{\partial x^2}\right]e^{i\phi} \ .\end{aligned}$$

A similar expression holds for the second partial derivative with respect to z. Adding these two expressions, we obtain

$$\begin{aligned}\nabla^2 u = &\left\{ iA\left[\frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial z^2}\right] - A\left[\left(\frac{\partial\phi}{\partial x}\right)^2 + \left(\frac{\partial\phi}{\partial x}\right)^2\right]\right. \\ &\left. + 2i\left[\frac{\partial A}{\partial x}\frac{\partial\phi}{\partial x} + \frac{\partial A}{\partial z}\frac{\partial\phi}{\partial z}\right] + \left[\frac{\partial^2 A}{\partial x^2} + \frac{\partial^2 A}{\partial z^2}\right]\right\}e^{i\phi} \ ,\end{aligned}$$

which is

$$\nabla^2 u = \{iA\nabla^2\phi - A(\nabla\phi)^2 + 2i(\nabla A \cdot \nabla\phi) + \nabla^2 A\}e^{i\phi} \ .$$

In this equation, $\nabla\phi$ is the gradient of ϕ, ∇A is the gradient of A, $\nabla A \cdot \nabla\phi$ is the dot product of ∇A and $\nabla\phi$, and $(\nabla\phi)^2$ is the dot product of $\nabla\phi$ and $\nabla\phi$. Thus, we have evaluated the first term in the Helmholtz equation. The second term is easier. It is simply

$$k^2 u = k^2 A\, e^{i\phi} \ .$$

Putting both terms together, we obtain the Helmholtz equation in the form

$$[iA\nabla^2\phi - A(\nabla\phi)^2 + 2i(\nabla A \cdot \nabla\phi) + \nabla^2 A]\, e^{i\phi} = 0 \ ,$$

which yields

$$-A[(\nabla\phi)^2 - k^2] + i[2\nabla A \cdot \nabla\phi + A\nabla^2\phi] + \nabla^2 A = 0 \ . \qquad (2)$$

We call this expression the *phase form of the Helmholtz equation*.

Thus far, we have only written Helmholtz equation 1 in a different form, namely, the phase form given by equation 2. Now, we are ready for the basic approximation that will simplify the Helmholtz equation. This approximation says that the wavelength λ is regarded as small compared with the linear dimensions of the region occupied by the wavefield. In the limiting case of a plane wave, the phase is

$$\phi = -\boldsymbol{k} \cdot \boldsymbol{r} = -k\sigma \ .$$

In this equation,

$$k = |\boldsymbol{k}|$$

and σ is the projection of $\boldsymbol{r}$ on the direction of the vector $\boldsymbol{k}$. Because $k = 2\pi/\lambda$, we have

$$\phi = -\frac{2\pi\sigma}{\lambda} \ .$$

This expression for ϕ in the case of plane waves motivates us to use a similar expression, namely,

$$\phi = \frac{\chi}{\lambda}$$

in the general case. Thus, our basic approximation involves using an expression for ϕ containing λ explicitly. For $\phi = \chi/\lambda$, we have

$$\nabla\phi = \frac{1}{\lambda}\nabla\chi \ .$$

Thus, the phase form (equation 2) of Helmholtz equation 1 becomes

$$-\frac{A}{\lambda^2}[\nabla\chi - (2\pi)^2] + \frac{i}{\lambda}[2\nabla A \cdot \nabla\phi + A\nabla^2\chi] + \nabla^2 A = 0 \ .$$

This expression has three terms. If wavelength λ is small, then it follows that $1/\lambda^2$ is very large and $1/\lambda$ is large. Thus, the most significant term is the first, the next most significant is the second term, and the third term is insignificant. Each of these terms, being of different orders of magnitude, must be set equal to zero separately. If we return now to the notation ϕ, the first two terms give, respectively, the so-called *eikonal equation*

$$(\nabla\phi)^2 - k^2 = 0 \tag{3}$$

and the so-called *geometric spreading equation*

$$2\nabla A \cdot \nabla\phi + A\nabla^2\phi = 0 \ .$$

These two equations result from the small wavelength assumption and are the two basic equations for geometric optics, geometric acoustics, and geometric seismology.

Next, we want to address eikonal equation 3, which explicitly can be written as

$$\left(\frac{\partial\phi}{\partial x}\right)^2 + \left(\frac{\partial\phi}{\partial z}\right)^2 = 0 \ . \tag{4}$$

Thus, the eikonal equation is an inhomogeneous, nonlinear, first-order partial differential equation. Given the wavenumber

$$k(\boldsymbol{r}) = \frac{\omega}{c(\boldsymbol{r})} \ ,$$

we must solve this equation for the phase $\phi(\boldsymbol{r})$. Knowledge of the phase function $\phi(\boldsymbol{r})$ lets us visualize the way waves are propagated in an inhomogeneous medium. The surface defined by those values of $\boldsymbol{r}$ for which $\phi(\boldsymbol{r}, \omega)$ is constant is a surface of constant phase, i.e., a *wavefront*. The propagation of the wavefronts provides a pictorial representation of the propagation of the wave. In the case of plane waves, the wavefronts are

$$\phi = -\boldsymbol{k} \cdot \boldsymbol{r} = \text{constant} \ ,$$

which are planes. In an inhomogeneous medium, the wavefronts have a curved shape instead of being planes.

If we operate with

$$\nabla = \hat{\boldsymbol{i}}\frac{\partial}{\partial x} + \hat{\boldsymbol{k}}\frac{\partial}{\partial z} \tag{14}$$

on a scalar function ϕ, we obtain a vector which is called the gradient of ϕ, i.e.,

$$\text{grad}\,\phi = \nabla\phi = \hat{\boldsymbol{i}}\frac{\partial\phi}{\partial x} + \hat{\boldsymbol{k}}\frac{\partial\phi}{\partial z} \ .$$

Now, let us show that the vector $\nabla\phi$ is normal (i.e., perpendicular) to the surface = constant. Take an incremental displacement

$$d\boldsymbol{r} = \hat{\boldsymbol{\imath}}\,dx + \boldsymbol{k}\,d\hat{z}$$

on the surface. Because ϕ does not change anywhere on this surface, it follows that ϕ does not change in the displacement $d\boldsymbol{r}$. That is,

$$d\phi = \frac{\partial\phi}{\partial x}dx + \frac{\partial\phi}{\partial z}dz = (\nabla\phi)\cdot d\boldsymbol{r} = 0\ ,$$

which states that $\nabla\phi$ is perpendicular to any vector which lies in the plane tangential to the surface

$$\phi = \text{constant}\ .$$

Thus, we have proven our assertion that the gradient is normal to the surface.

In our work, of course, the surface $\phi =$ constant is the wavefront. In effect, the function ϕ represents the local phase. The small wavelength assumption means that the wavenumber vector

$$\boldsymbol{k} = (k_x, k_z) = k_x\hat{\boldsymbol{\imath}} + k_z\hat{\boldsymbol{k}}$$

varies only gradually (i.e., by just a small fraction of its magnitude $k = 2\pi/\lambda$) in one wavelength λ. The phase ϕ shows a decrease with x at a rate equal to k_x, which is the x-component of the local wavenumber. Similarly, ϕ shows a decrease with z at a rate equal to k_z. Thus, we can write

$$\frac{\partial\phi}{\partial x} = -k_x, \quad \frac{\partial\phi}{\partial z} = -k_z\ .$$

These two equations ensure that locally the wave is nearly of sinusoidal form

$$\exp -i(k_x\,x + k_z\,z)\ ,$$

which is required in the small wavelength assumption. These two equations can be combined into the one equation

$$\frac{\partial\phi}{\partial x}\hat{\boldsymbol{\imath}} + \frac{\partial\phi}{\partial z}\hat{\boldsymbol{k}}_z = -(k_x\hat{\boldsymbol{\imath}} + k_z\hat{\boldsymbol{k}})\ ,$$

which is

$$\nabla\phi = -\boldsymbol{k} \ .$$

This is an important result; in words, the gradient of the wavefront is the negative of the wavenumber vector. Thus, the wavenumber vector $\boldsymbol{k}$ is normal to the wavefront.

A *ray* is defined as a path along which energy propagates. In the case of an isotropic medium, the energy propagates in the direction of the wavenumber vector $\boldsymbol{k}$. As we have previously stated, we assume throughout that the medium is isotropic. Because the wavenumber vector $\boldsymbol{k}$ is normal to the wavefront, it follows that a ray is a continuous curve that is drawn perpendicular to all surfaces of constant phase.

Before we solve the eikonal equation for the phase function ϕ, let us first find an equation for the rays that does not involve ϕ. Huygens' principle is basically a statement of how physical waves propagate. Huygens' principle asserts that a wavefront (i.e., a surface of constant phase) must proceed along a ray with speed c, where c is the wave velocity. Wave velocity in an inhomogeneous medium depends upon position $\boldsymbol{r}$, so we write $c(\boldsymbol{r})$. If $\boldsymbol{r}$ denotes the position on the ray, then the velocity of the wavefront is $d\boldsymbol{r}/dt$. The ray is in the direction of $\boldsymbol{k}$, so a unit vector in the direction of the ray is $\boldsymbol{k}/|\boldsymbol{k}|$, which is $\boldsymbol{k}/k$. Thus, Huygens' principle is

$$\frac{d\boldsymbol{r}}{dt} = c(\boldsymbol{r})\frac{\boldsymbol{k}}{k} \ .$$

That is, the new wavefront at $\boldsymbol{r} + d\boldsymbol{r}$ is obtained from the old wavefront at $\boldsymbol{r}$ by considering each point $\boldsymbol{r}$ as a source of new waves, each traveling with the velocity $c(\boldsymbol{r})$ at that point. The constructive waves are the waves that travel on the ray, and they travel a distance $dr = c(\boldsymbol{r})dt$ in the ray direction (i.e., in the direction of the unit vector $\boldsymbol{k}/k$).

Let $d\sigma$ be the differential of arc length along the ray. Then,

$$d\sigma = c(\boldsymbol{r})dt \ ,$$

so Huygens' principle also can be written as

$$\frac{d\boldsymbol{r}}{d\sigma} = \frac{\boldsymbol{k}}{k} \ . \tag{5}$$

This equation states that, if $\boldsymbol{r}$ is the coordinate of a point on the ray, and σ is arc length along the ray, then the unit tangent vector $d\boldsymbol{r}/d\sigma$ to the ray is equal to the unit vector $\boldsymbol{k}/k$. As we have shown in this chapter,

$$\boldsymbol{k} = -\nabla\phi \ .$$

Thus, equation 5 for the coordinate $\boldsymbol{r}$ of the ray becomes the *ray equation*

$$k\frac{d\boldsymbol{r}}{d\sigma} = -\nabla\phi \ .$$

Recall that we want to find the equation for rays that does not involve ϕ. Thus, we must eliminate ϕ from the above equation. Let us write down the two components of the *ray equation*, namely,

$$k\frac{dx}{d\sigma} = -\frac{\partial\phi}{\partial x} \quad \text{and} \quad k\frac{dz}{d\sigma} = -\frac{\partial\phi}{\partial z} \ . \tag{6}$$

We now differentiate the first of the components of the ray equation to obtain

$$\frac{d}{d\sigma}\left(k\frac{dx}{d\sigma}\right) = -\frac{d}{d\sigma}\left(\frac{\partial\phi}{\partial x}\right) = -\left[\frac{\partial^2\phi}{\partial x^2}\frac{dx}{d\sigma} + \frac{\partial^2\phi}{\partial z\partial x}\frac{dz}{d\sigma}\right] .$$

We then substitute for $dx/d\sigma$ and $dz/d\sigma$ in this equation, making use of the expressions for $dx/d\sigma$ and $dz/d\sigma$ from the two components of the ray equation shown in equation 6. Thus, we obtain

$$\frac{d}{d\sigma}\left(k\frac{dx}{d\sigma}\right) = -\left[\frac{\partial^2\phi}{\partial x^2}\left(-\frac{1}{k}\frac{\partial\phi}{\partial x}\right) + \frac{\partial^2\phi}{\partial z\,\partial x}\left(-\frac{1}{k}\frac{\partial\phi}{\partial z}\right)\right] ,$$

which is

$$\frac{d}{d\sigma}\left(k\frac{dx}{d\sigma}\right) = \frac{1}{2k}\frac{\partial}{\partial x}\left[\left(\frac{\partial\phi}{\partial x}\right)^2 + \left(\frac{\partial\phi}{\partial z}\right)^2\right] .$$

But, from eikonal equation 4, the expression in square brackets is k^2. Thus, we have

$$\frac{d}{d\sigma}\left(k\frac{dx}{d\sigma}\right) = \frac{1}{2k}\frac{\partial}{\partial x}[k^2] \ .$$

This equation is

$$\frac{d}{d\sigma}\left(k\frac{dx}{d\sigma}\right)=\frac{\partial k}{\partial x}\ .$$

Similarly, we can obtain

$$\frac{d}{d\sigma}\left(k\frac{dz}{d\sigma}\right)=\frac{\partial k}{\partial z}\ .$$

Combining the above two equations, we find

$$\frac{d}{d\sigma}\left[k\frac{d}{d\sigma}(x\hat{\boldsymbol{i}}+z\hat{\boldsymbol{k}})\right]=\left(\hat{\boldsymbol{i}}\frac{\partial}{\partial x}+\hat{\boldsymbol{k}}\frac{\partial}{\partial z}\right)k$$

or

$$\frac{d}{d\sigma}\left(k\frac{d\boldsymbol{r}}{d\sigma}\right)=\nabla k\ . \tag{7}$$

This is the desired *ray equation* for the position $\boldsymbol{r}$ on the ray. This is a differential equation for the position vector $\boldsymbol{r}$ as a function of arc length σ along the ray. This equation only requires the value of the wavenumber

$$k(\boldsymbol{r})=\frac{\omega}{c(\boldsymbol{r})}$$

as a function of $\boldsymbol{r}$. Solution of the ray equation provides the equation of the ray expressed in terms of the vector position

$$\boldsymbol{r}=(x,\,y)$$

on the ray as a function $\boldsymbol{r}(\sigma)$ of the arc length σ along the ray.

In the preceding paragraph, we obtain the ray equation 7. Its solution yields the curve $\boldsymbol{r}(\sigma)$, which expresses a point $\boldsymbol{r}$ on the ray as a function of arc length along the ray. We now want to derive the equation for the phase ϕ.

As we know, the surface of constant phase is a wavefront, and the ray curves are orthogonal to the wavefront. If $\boldsymbol{r}(\sigma)$ is a point on the ray, then the wavefront through this point is the surface of constant phase $\phi(\boldsymbol{r}(\sigma))$. That is, we can characterize the wavefront by the value of the arc length σ at the point on the ray where the wavefront cuts the ray. Thus, we can simply write the phase as $\phi(\sigma)$. The derivative of the phase with respect to arc length is $d\phi(\sigma)/d\sigma$. We now evoke the chain rule to obtain

$$\frac{d\phi(\sigma)}{d\sigma} = \frac{d\phi}{dx}\frac{dx}{d\sigma} + \frac{d\phi}{dz}\frac{dz}{d\sigma} \ .$$

We then substitute for $dx/d\sigma$ and $dz/d\sigma$ the values

$$\frac{dx}{d\sigma} = -\frac{1}{k}\frac{\partial\phi}{\partial x}, \quad \frac{dz}{d\sigma} = -\frac{1}{k}\frac{\partial\phi}{\partial z}$$

obtained from the ray equation components shown in equation 6. Thus, we obtain

$$\frac{d\phi(\sigma)}{d\sigma} = -\frac{1}{k}\left[\left(\frac{\partial\phi}{\partial x}\right)^2 + \left(\frac{\partial\phi}{\partial z}\right)^2\right] .$$

Referring to eikonal equation 4, we see that the expression in brackets is equal to k^2. Thus, we have

$$\frac{d\phi(\sigma)}{d\sigma} = -k \ .$$

This differential equation can be integrated directly to obtain the following expression for the phase:

$$\phi(\sigma) = \phi(\sigma_0) - \int_{\sigma_0}^{\sigma} k(\sigma_1)\, d\sigma_1 \ .$$

Because the parameter σ measures distance along the ray path, we see that the value of the phase ϕ at σ is equal to the initial value $\phi(\sigma_0)$ plus the integral of the wavenumber $k(\sigma_1)$ along the ray from $\sigma_1 = \sigma_0$ to $\sigma_1 = \sigma$.

Discovery of anisotropy

Iceland spar was pre-1845 nomenclature for a particularly beautiful type of transparent calcite discovered on the east coast of Iceland in the middle of the 17th century. Almost immediately, it became the focus of one of the most famous scientific investigations of that or any time. The results profoundly affected both physics and geophysics from that point forward. In fact, for more than a century, a curious optical property of Iceland spar was a dominating concern at the highest levels of science because of differing explanations by the world's two most esteemed virtuosi, Isaac Newton (1642 – 1727) and Christiaan Huygens (1629 – 1695).

In the 17th century, virtuoso had a much different meaning than the current implication of spectacular musical dexterity. Then, it connoted someone with a great talent for discovering and demonstrating scientific truth. Most often, this meant the virtuoso combined the attributes of theorist and experimentalist plus instrument designer and builder—a striking parallel to the demands placed upon the pioneer exploration geophysicists of the early 20th century.

For most of his adult life, Huygens reigned practically unchallenged (and this was a period when science overflowed with superlative intellects) as the supreme virtuoso in European intellectual circles. Before he turned 30, and less than 20 years after Galileo's death, Huygens had made immeasurably important extensions of Galileo's fundamental discoveries in astronomy and properties of the pendulum.

Using a telescope of his design (equipped with lenses manufactured via a Huygens-conceived technique that remained state-of-the-art until well into the 20th century), Huygens discovered the rings around Saturn. His interest in astronomy led to a concern for accurate measurement of time, which in turn led to the development of the pendulum (or grandfather's) clock. The pendulum clock increased the accuracy of clocks from about 15 minutes per day to a few seconds per day and launched the modern era of precision timekeeping. Isaac Asimov summarized the value of the pendulum clock dramatically: "It is difficult to see how physics could have advanced much further without such an invention."

Despite recurring health problems, Huygens continued to contribute prodigiously and importantly in a wide variety of disciplines throughout his life. His interests ranged from the most immediately useful to the outer boundaries of imagination. His final book, *Casmotheoros*, speculated about the existence of life on other planets. He was one of the founders of the mathematical theory of probability. He wrote the first book on this subject and another did not appear for many years. It also was in Huygens' work that the product of mass and the square of velocity, a relationship of fundamental importance, appeared for the first time.

The extent of Huygens' scientific production was put into perspective by the effort required by his homeland, The Netherlands, to publish a comprehensive edition of his work. The project went into high gear in 1885 and needed more than 60 years to complete. The 22 volumes in the *Oeuvres Completes* have been cited as the best publication of the work of any scientist. A summary of Huygens' work occupies more than 13 double-column pages in Scribner's *Dictionary of Scientific Biograph*y and it omits at least one important contribution—his speculation that the earth is not a perfect sphere but flattened at the poles.

Geophysicists are, of course, most familiar with Huygens because he is considered the founder of the wave theory of light, much of which also applies to the propagation of acoustic waves. Wave propagation is a complicated process. The mathematics of this subject is still an active field of experimentation and research, not only in geophysics but in virtually all physical sciences. Well before the introduction of the wave equation, however, Huygens devises a procedure—one that remains in accordance with modern views—to give an approximate description of wave propagation. This is the very familiar *Huygens' principle* in which a wavefront at any given instant is considered to be composed of many separate point sources. Each source radiates a secondary wavelet and the superposition of all of the secondary wavelets gives the wave motion at a later time. Specifically, the new wavefront is the envelope of the secondary wavelets in the direction of propagation.

Christiaan Huygens was determined to get a better view of the heavens. He invented and constructed the first powerful astronomical telescopes. One of his telescopes had a focal length of 210 feet. He developed the first compound eyepiece for a telescope by using multiple lenses. In 1656, Huygens devised a new method in order to find the radius of the earth's orbit, and found it to be equal to 12,000 earth diameters, a value which was remarkably close to the true value of 11,728, a distance called one astronomical unit (denoted by the symbol AU).

The French Royal Academy recruited Huygens and took advantage of his inventions so as to make the newly formed French Royal Observatory the most advanced in the world. His inventions made possible extremely accurate observations of the timing of the eclipses of the satellites of planet Jupiter. Huygens took up residence in Paris in May 1666. The French king gave Huygens a generous stipend and a prestigious apartment in the Academy building.

During his years in Paris, Huygens was a mainstay of the French Academy. He was recognized as one of the greatest mathematicians in Europe. He gained fame for his physical theories and inventions. He developed laws of motion before Newton. In 1669, Huygens sent to the *Journal des Sçavans* in France a brief communication that included very clear statements of three of the most fundamental laws of nature: the conservation of momentum, the conservation of kinetic energy in collisions, and the center-of-mass law.

The French wanted to map the world. In order to determine longitude, they needed an accurate chronometer. Unfortunately, a pendulum clock could not serve as a chronometer because of the jouncing it gets on a ship or in a carriage. The position of the satellites of Jupiter could serve as a natural chronometer. Ships did not provide a stable platform for telescopes, so the satellite chronometer would only work well on land. At the French Royal Observatory, Jean-Dominique Cassini laboriously compiled tables of the eclipses of Jupiter's satellites for use in the determination of longitude.

Because of ill health, Huygens took a leave from the French Academy in June 1676 and returned to his native Holland. At the Academy on 22 August 1676, Cassini reported that light arrived from Jupiter's nearest satellite, named Io, with a delay, so that it took between 20 and 22 minutes to cross a distance equal to the diameter of the annual orbit of the earth. (By definition, the diameter of the annual orbit of the earth is 2 astronomical units, i.e., 2 AU.) On 7 December 1676, the *Journal des Sçavans* published in French the Roemer report. It described the work of Cassini's assistant, Ole Roemer, and reported the delay as 22 minutes. (Note that the Roemer report chose the upper value of the range "20 to 22 minutes" given earlier by Cassini.) This result was all that Cassini and Roemer needed, because it showed that such delays did exist and thereby must be carefully observed and included in the compiled tables of the eclipses. (Actually, the reported delay of 22 minutes was in error—to the nearest minute, the correct value is 17 minutes.)

In Holland, Huygens continued his study of light waves, and he solved the double-refraction problem, which was exhibited by Icelandic spar. On 16 September 1677, Huygens received his copy of the 25 July 1677 issue of the English journal *Philosophical Transactions of the Royal Society*. This issue contained an English translation of the Roemer report that had appeared in the 7 December 1676 issue of the *Journal des Sçavans*.

For many years, Huygens had advocated the finiteness of the velocity of light, for it was essential to the workings of Huygens' principle. On learning about Cassini's report of the delay of $x = 22$ minutes due to traveling the distance $X = 2AU$, Huygens was more than delighted. He was ecstatic. He now had distance (2 AU) and time (22 minutes) for the same thing. They were related by the velocity c of light; that is, Huygens used the relationship

$$X = cx \ .$$

In so doing, Huygens invented the concept of light-minute; namely, the light-minute is the distance traveled by light in one minute. Thus, the distance 2 AU is 22 light-minutes.

Because the Earth moves in its orbit so much faster than Jupiter moves in its orbit, we may consider the two planets as moving apart for half of the year and moving together for the other half of the year. Let us use symbols to describe the argument that Huygens made in *Traité de la Lumière*. For convenience, we begin with the Earth at the point in its orbit farthest from Jupiter. It takes one-half year (denoted by t) for the Earth to go away from the point nearest to Jupiter. It then takes one-half year for the Earth to come back. We will use an analogy with a journey on an airplane. The clock loses five hours when you go away by airplane from New York to London, but you gain five hours when you come back from London to New York. By analogy, the clock loses 22 minutes (denoted by $-x$) when the Earth goes away from farthest to nearest point, but the clock gains 22 minutes (denoted by $+x$) when the Earth comes back. Thus, going away takes $t-x$, but coming back takes $t+x$. In scientific terms, the sending period is

$$s = t - x \ .$$

and the receiving period is

$$r = t + x \ .$$

To Huygens, it does not matter whether Jupiter stands still and the Earth moves, or whether the Earth stands still and Jupiter moves, or if it is a combination of both. In other words, absolute velocities do not matter. Only the relative velocity between the two planets matters.

Huygens projects s onto r in two steps. Because of the use of relative velocity, the same projection factor k (otherwise known as the optical Doppler factor) applies in each step. The intermediate projection is called the *proper time*, which is denoted by p. Thus, we have

$$ks = p \quad \text{and} \quad kp = r \ .$$

The above two equations give

$$k(ks) = r \quad \text{so} \quad k = \sqrt{\frac{r}{s}} \ .$$

The relative velocity (in terms of the velocity of light) is

$$v = \frac{x}{t} \ .$$

Thus, the optical Doppler factor is

$$k = \frac{p}{s} = \frac{r}{p} = \sqrt{\frac{r}{s}} = \sqrt{\frac{t+x}{t-x}} = \sqrt{\frac{1+(x/t)}{1-(x/t)}} = \sqrt{\frac{1+v}{1-v}} \ .$$

The optical Doppler effect is appropriate for electromagnetic waves. It is the counterpart to the acoustical Doppler effect, which is appropriate for mechanical waves.

Huygens computed the numerical value the velocity of light. As we have seen, $X = cx$, so the velocity of light is

$$c = \frac{X}{x} = \frac{2\,\mathrm{AU}}{22\ \text{minutes}} \ .$$

As we have noted, Huygens, in 1656, obtained the value of the AU as 12,000 diameters of the earth. The modern value of the AU is 11,728 diameters of the earth, but Huygens' value of 12,000 was as close as could be expected in the mid-17th century. In 1672, Cassini in France and Richer in Cayenne calculated the AU. They had access to the most advanced instruments, and the French government paid for observations in Paris and French Guiana. Cassini obtained the value of 11,000 earth diameters for the AU. Cassini published his result of 11,000 in two papers without any reference to Huygens. Huygens deferred to Cassini and used Cassini's value of 11,000 to calculate the speed of light.

Huygens used the system of measurement that was legal in France until 1812, and he took the diameter of the Earth as 2865 leagues. A league was 2282 toises. Thus, Huygens gave the diameter of the Earth as

$$\text{earth diameter} = 2865 \times 2282 = 6{,}537{,}930 \text{ toises} \ .$$

Let us now convert this value to metric system. One toise (i.e., 1 fathom) is exactly 6 pieds (i.e., 6 feet). The value 6 pieds is about 1.949 meters. One meter is 0.001 kilometer. Thus, the earth's diameter, according to Huygens, is the value

$$\text{earth diameter} = 6{,}537{,}930 \times 1.949 \times 0.001 = 12{,}742 \text{ kilometers} \ .$$

According to *Wikipedia*, the average diameter of the earth, which is referred to in common usage, is 12,756 km, a difference from Huygens' value of 14 km.

The diameter of the earth's orbit is 2 AU. According to Cassini, light takes 22 minutes to travel that distance. Also according to Cassini, 1 AU is 11,000 earth diameters. Huygens, in using these values, obtains the velocity c of light as

$$c = \frac{2\,\mathrm{AU}}{22\ \text{minutes}} = \frac{(22{,}000\ \text{earth diameters})}{22\ \text{minutes}} = \frac{1000\ \text{earth diameters}}{\text{minute}} \ .$$

As Huygens wrote, this makes the speed of light equal to 1000 earth diameters per minute. In other words, the speed of light is 1000/60, which is 16.67 earth diameters per second. An earth diameter is 12,742 kilometers. Thus, the velocity of light is

$$c = \frac{1000 \times 12{,}742}{60} \approx 212{,}000 \text{ kilometers per second} .$$

Thus, Huygens computes the value of light as 212,000 kilometers per second, which is short of the correct value of 300,000 kilometers per second. There are two reasons for this incorrect value.

1. The delay of 22 minutes given by Cassini is in error. Actually the delay (rounded to the nearest minute) is 17 minutes.
2. The AU value of the 11,000 earth diameters given by Cassini is in error. The actual value of the AU is closer to Huygens' value of 12,000 earth diameters.

If Huygens had used these values (which he did not), he would have obtained the reconstituted velocity

$$c = \frac{2\text{ AU}}{17\text{ minutes}} = \frac{(24{,}000\text{ earth diameters})}{17\text{ minutes}} = \frac{1412\text{ earth diameters}}{\text{minute}} ,$$

which is

$$c = \frac{1412 \times 12{,}742}{60} \approx 299{,}862 \text{ kilometers per second} .$$

The modern value is 299,792 km per second, from which Huygens' reconstituted value for the speed of light was off by 70 km per second. The world credited Huygens with the first calculation of the speed of light. The world did not comprehend until the 20th century that Huygens had originated the optical Doppler effect. Neither Cassini nor Roemer ever tried to calculate the velocity of light from the delay time. In fact, it never occurred to them to do so. But why should it have occurred to them? They had solved the age-old problem of whether the propagation of light was instantaneous or not. They found that it was not.

The concept of a ray goes back to ancient times. The ancient Greeks thought that light rays emanate from the eye and intercept objects, which are thereby seen. The ancient Greeks also thought that the speed with which light rays emerge from the eye is very high, most likely infinite. One reason was that an observer can open his eyes and immediately see the distant stars. For Euclid, a ray of light was a discrete physical object. The concept of ray underwent many modifications over the centuries, but the ray was always regarded to be the physical manifestation of light. Some would say that a ray was a line; others would say it was a narrow beam. In either case, the ray by itself had some type of physical reality. The ray represented the basis for the understanding of light. In the 17th century, the particle theory of light was sanctioned. A ray was like a string of particles, in which each particle communicates its motion only to the next particle in line. Only Ignace Pardies and Christiaan Huygens did not give a physical reality to a ray. They both reduced the ray to being nothing more than a mathematical construction. Huygens' *Traité de la Lumière* was published in 1678. In this masterpiece, Huygens went further than Pardies, in that he introduced a fundament principle (Huygens' principle) that depicted light as a traveling wave generated by a continuous succession of secondary waves (a.k.a., wavelets).

Christiaan Huygens relegated the ray to a mathematical existence. Huygens believed that a wavelet was generated around each particle with the particle at the center. These wavelets reconstituted at any moment the wavefront of light. If the wavelets were spherical, then the rays would be orthogonal to the wavefront. The outstanding characteristic of Huygens' construction was that the wavelets need not be spherical. If the wavelets were not spherical, then the rays would not be orthogonal to the wavefront. Instead, they would be oblique to the wavefront.

The Vikings sailed without the use of astrolabes, magnetic compasses, or any other known device. They could find their way despite clouds, fog, and the long northern summer twilights. Viking legends speak of the use of glowing "sunstones" to find the position of the sun on cloudy days and even when the sun was just below the horizon. It is believed that these "sunstones" were Iceland spar, a transparent variety of calcite, or crystallized calcium carbonate.

Iceland spar splits light into two polarized images. The images have different brightness depending on the polarization. Sunlight is polarized when it enters the earth's atmosphere. The Iceland spar must be oriented until two images have the same relative brightness. In such a position, the Iceland spar is aligned to the sun.

In 1669, Erasmus Bartholin (also known by the Latinized name, Bartholinus), doctor of medicine and professor of mathematics at the University of Copenhagen, received a piece of Iceland spar and noted the remarkable optical phenomenon which he called double refraction. In his text on the geometry of crystals (*Experimenta crystalli Islandici disdiaclastici quibus mira & insolita refractio detegitur*), Bartholin wrote:

> Greatly prized by all men is the diamond, but he who prefers the knowledge of unusual phenomena will have no less joy in a new sort of body, recently brought to us from Iceland, which is perhaps one of the greatest wonders that nature has produced. As my investigation of this crystal proceeded, there showed itself a wonderful and extraordinary phenomenon: objects which are looked at through a crystal do not show, as in the case of other transparent bodies, a single refracted image but they appear double.

If the crystal is rotated, one of the dots remains stationary but the other (following the crystal's motion) appears to move in a circle around the first. The rays forming the fixed dot behave as if they merely passed through a plate of glass. Bartholin calls them the ordinary (or *o*) waves. The waves coming from the other dot, which behave unusually, are known as the extraordinary (or *e*) waves. As might be expected, the *o* waves correspond to the case in which the medium is isotropic and, therefore, the Huygens' wavelets are spheres. But calcite is *anisotropic*, and in such a medium Huygens reasoned there are two wavelets for each point source: one is spherical (corresponding to the *o* waves) and the other is a spheroid (corresponding to the *e* waves). A spheroid, or ellipsoid of revolution, is a quadric surface obtained by rotating an ellipse about one of its principal axes—in other words, an ellipsoid with two equal semi-diameters.

Iceland spar provides Huygens with an ideal model upon which to test his principle. A spherical wavelet spreads with constant velocity in all directions; a spheroidal wavelet spreads with a velocity which varies with the direction of the ray. (A spheroid is a three-dimensional geometric surface generated by rotating an ellipse about one of its axes.) In the case of calcite, the velocities of both types of secondary wavelets are identical in the direction of the optic axis, but, in other directions, the velocity of the spheroidal wavelet exceeds that of the spherical wavelet. The spheroidal wavelet spreads at its greatest velocity in a plane perpendicular to the optic axis. In other directions, the velocity of the spheroidal wavelet ranges between the lower limit of its velocity along the optic axis and the upper limit of its velocity normal to the optic axis.

A ray represents an energy flow. As shown in Figure 1, a ray for the ordinary wavelet goes from the origin to a point on the sphere (which represents the wavefront of the *o* wavelet). A ray for the extraordinary wavelet goes from the origin to a point on the ellipsoid (which represents

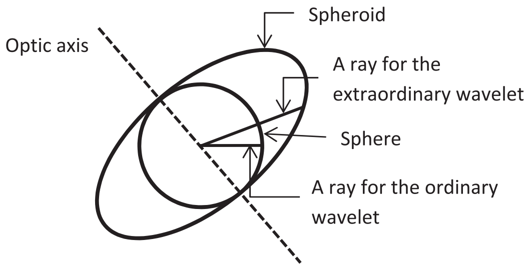


Figure 1. Rays for the ordinary wavelet and for the extraordinary wavelet.

the wavefront of the *e* wavelet). The length of a ray is proportional to the velocity in the ray's direction. Obviously, ray velocity is the same in all directions for the spherical wavelet; but, for the ellipsoidal wavelet, the velocity is minimum for a ray along the optic axis and maximum for one perpendicular to the optic axis.

In 1690 (although the work had been completed and presented to the French Academy in 1678), Huygens published his *Traité de la Lumière*, in which he explains the double refraction of Iceland spar. His argument is diagrammed in Figure 2. Suppose that line AA' represents the wavefront at a given time. Each point on AA' acts as a source of secondary wavelets. The envelope BB' of the spherical wavelets locates the wavefront of the *o* wave at a later time. The envelope CC' of the spheroidal wavelets gives the wavefront of the *e* wave at the same later time. Energy travels along rays, so the energy of the *o* wave is moving vertically down. However, as can be seen in the figure, the energy of the *e* wave is moving at angle θ to the vertical.

Both wavefronts, however, are parallel to the horizontal; hence, the normal vector to each wavefront points straight down. In the case of the *o* wave, the ray and the normal to the front both point in the same direction (vertically down, as shown in Figure 2). In the case of the *e* wave, the ray and the normal point in different directions. The ray points at angle θ; the normal points straight down. When something is viewed through a transparent anisotropic medium, such as calcite, two images are seen. One is due to *o* waves (which are traveling in the same direction as the wavefront) and the other due to the *e* waves (which are traveling in a different direction). It is a difficult concept to grasp because it occurs rarely in everyday experience. If ocean waves worked like this (they do not), it would be possible for two surfboarders starting at point A in Figure 2 to approach the shore on different azimuths. The one riding the *o* wave would come in along line *AB*. The other riding the *e* wave would approach along line *AC*.

The elegant simplicity of Huygens' explanation of double refraction gave the wave theory of light broad intuitive appeal. More specifically,

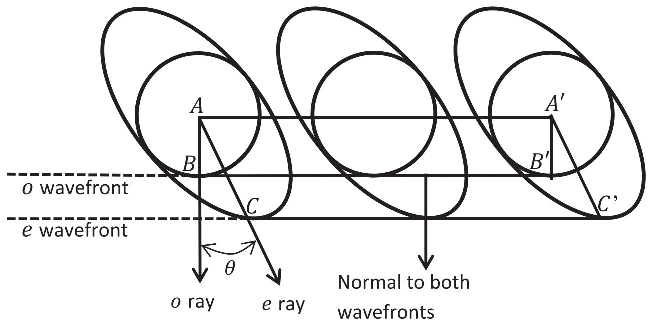


Figure 2. Huygens' explanation of double refraction.

Huygens showed that double refraction could be explained by wave theory. It could not be explained by ray theory. No one, however, took serious notice of this accomplishment of Huygens. The chief critic was none other than Sir Isaac Newton. Newton was not satisfied with Huygens' reasoning and continued to support the particle theory along with everyone else. By the middle of the 1690s, Newton had become the glamor name in science because of his work, *Philosophiae Naturalis Principia Mathematica* (a.k.a. the *Principia*), published in 1687 and often cited as the greatest scientific book ever written.

Huygens was ill much of the last five years of his life (1691 – 1695) and left behind no co-workers or supporters capable of vigorously defending his views. Newton's reputation, on the other hand, grew to almost mythic stature. By the time of this death in 1727, Newton's prestige in the scientific community was probably as high as any scientist's prestige has ever been during the scientist's own lifetime. The practically unassailable stature of Newton virtually prevented any interest in the wave theory for the rest of the 18th century.

During the 18th century, there was never a battle between the proponents of particle theory versus wave theory. Huygens' wave theory was simply ignored. In 1803, however, it was revived by Thomas Young (known to geophysicists via Young's modulus), whose brilliant experiments produced results explainable by waves but not by particles. Also during this time, Etienne-Louis Malus, using highly accurate devices, established new standards for experimental work in optics. In 1807, Malus started with a mathematical account of Huygens' construction and found excellent agreement between the theory and his experimental findings. A turning point came in 1808, when Pierre-Simon de Laplace announced that Malus had empirically confirmed Huygens' construction for double refraction. Augustin Fresnel extended Young's work and, assisted by a suggestion from Young that light waves were transverse (as opposed to Huygens' view that they were longitudinal), finally gave the complete explanation for the double refraction of Iceland spar—the crystal splits the light into two plane-polarized beams. Wave theory led physics for the rest of the 19th century.

Then, in the 20th century, quantum mechanics becomes prevalent. The photon is the smallest unit of light or other electromagnetic energy. It has no mass and no electric charge. The photon behaves both as a wave and as a particle. The photon displays wave-like phenomena, such as diffraction and interference. For example, a single photon passing through a double slit exhibits interference phenomena, but only if no measurement is made at the slit. However, the photon is not a short pulse of electromagnetic radiation. Instead, the photon seems to be a point-like particle, because it is absorbed or emitted as a whole by arbitrarily small systems, such as an atomic nucleus or the point-like electron. Newton is correct in the sense that light is a particle (photon). Huygens is correct in the sense that a photon does not travel as a particle travels, but instead travels spread out as a wave. A photon cannot be at rest; it must always be in motion, traveling as a wave at one speed, and one speed only, the speed of light in vacuum.

In summary, a most important milestone was reached when Huygens demonstrated, in *Traité de la Lumière*, that only his principle, nothing else, could determine the direction of the extraordinary ray in Iceland spar, thereby confirming the wave theory of light. Huygens was the first to show a candle transmitting waves of light (see Figure 3).

Huygens' optical Doppler effect has been a mainstay of radar and astronomy. In the 20th century, it was noticed that the light spectra of distant galaxies are shifted toward the red end of the spectrum. The optical Doppler effect says that, if an astronomical object is moving away from Earth, its light will be shifted to longer (red) wavelengths. In 1929,

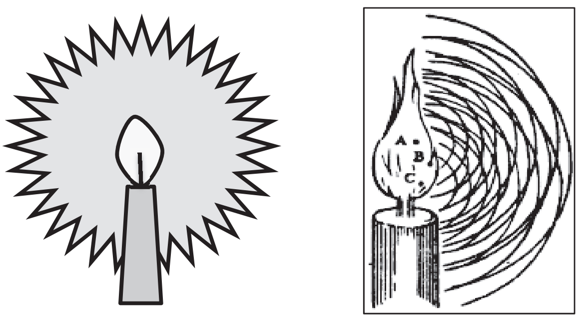


Figure 3. Candles were always drawn emitting rays (left) until Huygens, in 1678, drew a picture that changed the world: a candle transmitting waves (right).

astronomer Edwin Hubble compared the galaxies' spectra with their distances, and showed that the amount of "red shift" is proportional to distance. The most obvious explanation for the "red shift" is that the galaxies are receding from Earth and each other, and the farther the galaxy, the faster the recession. If all galaxies are flying apart at high speed, the entire universe must have been concentrated in a single point at some time in the past (the big bang). Another important use of the optical Doppler effect is in the discovery of extra solar planets and brown dwarfs.

Newton ended his scientific researches soon after Huygens' passing in 1695. Newton published *Opticks* in 1704, but the experiments had been done many years earlier. Newton began a second career in 1696 as warden and then master of the Royal Mint, a post that he held until his death in 1727. Newton reorganized the Mint, bought new equipment, and used his alchemical knowledge of metallurgy. Newton instituted the shift to a strict gold standard, which served English commerce well for more than 200 years.

Maxwell's equations

Let us now discuss the wave equation. The variable x is distance and t is time. The function p is pressure and the function V is particle velocity. The density, or mass per unit volume, is denoted by ρ.

First, we want to find the wave equation for V. From Newton's second law, it is possible to directly derive the following first-order partial differential equation:

$$\frac{\partial p}{\partial x} = -\rho \frac{\partial V}{\partial t} \ . \tag{8}$$

A symmetric first-order partial differential equation can be obtained from Hooke's law:

$$\frac{\partial p}{\partial t} = -K \frac{\partial V}{\partial x} \ , \tag{9}$$

where K is Young's modulus.

If we take $\partial/\partial t$ of equation 8 and $\partial/\partial x$ of equation 9, we obtain

$$\frac{\partial}{\partial t}\frac{\partial p}{\partial x} = -\rho \frac{\partial}{\partial t}\frac{\partial V}{\partial t} \ ,$$

$$\frac{\partial}{\partial x}\frac{\partial p}{\partial t} = -K \frac{\partial}{\partial x}\frac{\partial V}{\partial x} \ ,$$

which are

$$\frac{\partial^2 p}{\partial t\,\partial x} = -\rho\frac{\partial^2 V}{\partial t^2} \ , \tag{10}$$

$$\frac{\partial^2 p}{\partial x\,\partial t} = -K\frac{\partial^2 V}{\partial x^2} \ . \tag{11}$$

The left side of equation 10 is the same as the left side of equation 11. Therefore, their right sides are equal, which gives the wave equation for V as

$$-K\frac{\partial^2 V}{\partial x^2} = -\rho\frac{\partial^2 V}{\partial t^2} \ , \tag{12}$$

which is

$$\frac{\partial^2 V}{\partial x^2} = \frac{1}{\left(\frac{K}{\rho}\right)}\frac{\partial^2 V}{\partial t^2} \ . \tag{13}$$

Next, we want to find the wave equation for p. It follows that, by taking $\partial/\partial x$ in equation 8 and $\partial/\partial t$ in equation 9, we obtain

$$\frac{\partial}{\partial x}\frac{\partial p}{\partial x} = -\rho\frac{\partial}{\partial x}\frac{\partial V}{\partial t} \ ,$$

$$\frac{\partial}{\partial t}\frac{\partial p}{\partial t} = -K\frac{\partial}{\partial t}\frac{\partial V}{\partial x} \ ,$$

which are

$$\frac{\partial^2 p}{\partial x^2} = -\rho\frac{\partial^2 V}{\partial x\partial t} \ , \tag{14}$$

$$\frac{\partial^2 p}{\partial t^2} = -K\frac{\partial^2 V}{\partial t\partial x} \ . \tag{15}$$

If we solve equations 14 and 15, we obtain the wave equation for p given by

$$-\frac{1}{\rho}\frac{\partial^2 p}{\partial x^2} = -\frac{1}{K}\frac{\partial^2 p}{\partial t^2} \ , \tag{16}$$

which is

$$\frac{\partial^2 p}{\partial x^2} = \frac{1}{\left(\frac{K}{\rho}\right)} \frac{\partial^2 p}{\partial t^2} \ . \tag{17}$$

We see that the two wave equations 13 and 17 have the same form. The constant (K/ρ) is the square of the propagation velocity v; that is, $v = \sqrt{K/\rho}$.

For non-dissipative and non-dispersive electromagnetic waves, Maxwell's equations yield

$$\nabla \times \boldsymbol{E} = -\mu \frac{\partial \boldsymbol{H}}{\partial t} \ , \quad \nabla \times \boldsymbol{H} = -\varepsilon \frac{\partial \boldsymbol{E}}{\partial t} \ ,$$

where $\boldsymbol{E}$ and $\boldsymbol{H}$ are the electric and magnetic fields and ε and μ are the dielectric and permeability constants, respectively. For horizontal polarization of $\boldsymbol{E} = (E_x, 0, 0)$ and $\boldsymbol{H} = (\mathbf{0}, H_y, 0)$, Maxwell's equations give the symmetric first-order partial differential equations

$$\frac{\partial E_x}{\partial z} = -\mu \frac{\partial H_y}{\partial t}, \quad \frac{\partial E_x}{\partial t} = -\frac{1}{\varepsilon} \frac{\partial H_y}{\partial z} \ .$$

By cross differentiation, as before, it is apparent that E_x and H_y satisfy the wave equations

$$\frac{\partial^2 E_x}{\partial z^2} = \frac{1}{c^2} \frac{\partial^2 E_x}{\partial t^2}, \quad \frac{\partial^2 H_y}{\partial z^2} = \frac{1}{c^2} \frac{\partial^2 H_y}{\partial t^2} \ ,$$

where the propagation velocity c of light obeys the relation $c = 1/\sqrt{\varepsilon\mu}$.

Back in the day, geophysics visualized the underground in modest and idealistic terms. The source and the receivers were on or very close to the surface. The source produced a downgoing wave and the receiver recorded an upgoing wave. This ideal model did not include effects such as reverberations and ghost reflections, about which geophysicists had little knowledge. However, as time passed, geophysicists became aware of the great difficulties presented by these effects.

Now, with dual sensors and seismic processing, reverberations and ghost reflections above a buried receiver can be stripped away. The resulting seismic data approximates the ideal model of the past, with the datum no longer being at the surface of the ground, but at the depth of the receiver.

Chapter 5

Three-dimensional Wave Equation

Wave equation

Many important physical systems can be understood as manifestations of wave phenomena. It is, therefore, appropriate to begin our discussion with wave motion. The scalar wave equation governs the wave motion for many physical phenomena. A significant property of waves is that they carry energy over time from one spatial point to another. That is, waves are nature's way of transporting energy. In studying wave motion, the independent variables are x, y, z, t. The first three represent spatial coordinates and the fourth represents time. The dependent variable u represents the disturbance; i.e., the quantity undergoing wave motion. Here, we assume that u is a scalar and denote it by $u(x, y, z, t)$. The three spatial dimensional coordinates are x, y, z. The homogeneous form of the three-dimensional wave equation is

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} - \frac{1}{v^2}\frac{\partial^2 u}{\partial t^2} = 0 \ . \tag{1}$$

The three-dimensional wave equation also can be written as

$$\nabla^2 u = \frac{1}{v^2}\frac{\partial^2 u}{\partial t^2} \ . \tag{2}$$

Here, ∇ stands for the vector differential operator del whose coefficients in a Cartesian coordinate system are

$$\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right) .$$

The symbol ∇^2 is an abbreviation for the inner product or dot product of ∇ with itself:

$$\nabla^2 = \nabla \cdot \nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right) \cdot \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)$$
$$= \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \ .$$

In the wave equation 1, the constant v represents the velocity of the traveling waves. The equation can be derived from first principles. However, we will assume the existence of this equation as one of the basic equations of mathematical physics, and then continue from there. Let us look closely at the wave equation. On the left is the sum of the three second partial derivatives with respect to each of the space variables. Also on the left is the negative of the second partial derivative with respect to time divided by the quantity v^2. At this point in our discussion, we assume that the quantity v is a positive constant. The right side of the above wave equation is zero, so it is a homogeneous partial differential equation.

In the two-dimensional case of x, y, the wave equation reduces to its two-dimensional form,

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} - \frac{1}{v^2}\frac{\partial^2 u}{\partial t^2} = 0 \ . \tag{3}$$

In the one-dimensional case of x, the wave equation reduces to its one-dimensional form,

$$\frac{\partial^2 u}{\partial x^2} - \frac{1}{v^2}\frac{\partial^2 u}{\partial t^2} = 0 \ . \tag{4}$$

Now, let us make an important observation about solutions of the wave equation, namely, that the sum of two solutions is also a solution. More specifically, if u_1 and u_2 are solutions, so is $c_1 u_1 + c_2 u_2$, where c_1 and c_2 are arbitrary constants. This behavior is known as the *superposition principle*. The one-dimensional wave equation is the easiest wave equation with which to work. Surprisingly, the next easiest is the three-dimensional equation, and the most difficult is the two-dimensional equation. With the advent of collecting three-dimensional seismic data, the three-dimensional wave equation has assumed its rightful place in seismic analysis.

One-dimensional wave equation

Jean-le-Rond d'Alembert (1717–1783) was born in Paris. He was abandoned by his mother on the steps of the little church of St. Jean-le-Rond, which then nestled under the great porch of Notre Dame. He was taken to the parish commissary, who, following the usual practice in such cases, gave him the Christian name of Jean-le-Rond; it is not known by what authority he subsequently assumed the right to prefix *de* to his name. He was boarded by the parish with the wife of a glazier who lived near the cathedral, and here he found a real home, though a humble one. His schooling allowed him to obtain a mathematical education. His essay in 1738 on integral calculus and another essay in 1740 on "ducks and drakes" or ricochets attracted attention. He was elected as a member of the French Academy in 1740. Nearly all of his major mathematical works were produced during the years 1743 to 1754. The first of these was his *Traité de Dynamique*, published in 1743, in which he enunciated the principle known by his name, namely, that the "internal forces of inertia" (that is, the forces which resist acceleration) must be equal and opposite to the forces which produce the acceleration. This may be inferred from Newton's second reading of his third law of motion, but the full consequences had not been realized previously. This principle enabled mathematicians to obtain the differential equations of motion of any rigid system.

In 1744, d'Alembert published his *Traité de l'Équilibre et du Mouvement des Fluides*, in which he applies his principle to fluids; this led to partial differential equations which he was then unable to solve. In 1745, he developed that part of the subject which dealt with the motion of air in his *Théorie Générale des Vents*, and this again led him to partial differential equations. A second edition, in 1746, was dedicated to Frederick the Great of Prussia, and procured an invitation to Berlin and the offer of a pension; he declined the former, but subsequently, after some pressing, pocketed his pride and the latter. In 1747, he applied differential calculus to the problem of a vibrating string and arrived at the one-dimensional *wave equation.* D'Alembert showed that the general solution of the one-dimensional wave equation takes the form

$$u(x,\, t) = f\left(t - \frac{x}{v}\right) + g\left(t + \frac{x}{v}\right) , \tag{5}$$

where $f(s)$ and $g(s)$ are arbitrary functions of the variable s. Because of the superposition principle, we can satisfy ourselves with d'Alembert's solution by showing that the functions f and g separately satisfy the wave equation. Because the proofs are similar, we will only conduct the one for f.

We take the partial derivative of f two times, the first time with respect to t and the second time with respect to x:

$$\frac{\partial f}{\partial t} = f'\left(t - \frac{x}{v}\right), \quad \frac{\partial f}{\partial x} = -\frac{1}{v} f'\left(t - \tfrac{x}{v}\right) ,$$

where $f'(s)$ is an abbreviation for df/ds and s represents the argument $s = t - x/v$. We differentiate again, and obtain

$$\frac{\partial^2 f}{\partial t^2} = f''\left(t - \frac{x}{v}\right), \quad \frac{\partial^2 f}{\partial x^2} = \frac{1}{v^2} f''\left(t - \frac{x}{v}\right) ,$$

where $f''(s)$ is an abbreviation for $d^2 f/ds^2$. These expressions can be directly combined to give the one-dimensional wave equation 3. The proof for g follows along the same lines.

The waveform f at time t and point x is

$$\text{waveform } A = f\left(t - \frac{x}{v}\right) .$$

Let $\Delta x = v\Delta t$. The waveform f at the later time $t + \Delta t$ and point $x + \Delta x$ is

$$\text{waveform } B = f\left(t + \Delta t - \frac{x + \Delta x}{v}\right) .$$

We write waveform B as

$$\text{waveform } B = f\left(t + \Delta t - \frac{x}{v} - \Delta t\right) = f\left(t - \frac{x}{v}\right) .$$

This equation shows that waveform B is the same as waveform A. Assume that velocity v is positive. Then, point $x + v\Delta t$ lies to the right of point x at a distance $v\Delta t$. Thus, we see that the undistorted waveform f moves at velocity v and goes to the right. A similar argument shows that the undistorted waveform g moves at velocity $-v$ and thus goes to the left.

In conclusion, the two terms in the general solution shown in equation 13 have the following interpretation. The term $f(t - x/v)$ represents a waveform moving at velocity v with no change in size or shape. The term $g(t + x/v)$ represents a waveform moving at velocity $-v$ with no change in size or shape.

Sinusoidal waves

A sense is a physiological capacity of organisms that provides data for perception. Aristotle (384 BC – 322 BC) gives the traditional classification of the five sense organs: sight, smell, taste, touch, and hearing. The eye is the organ of vision. The spectrum of light to which the eye is sensitive varies from the color red (composed of lower electromagnetic frequencies of light) to the color violet (composed of higher electromagnetic frequencies of light). In other words, the eyes sense frequencies of electromagnetic waves of light. The ear is the organ of hearing. The human ear can perceive frequencies of sound from 16 cycles per second, which is a very deep bass, to 28,000 cycles per second, which is a very high pitch. In other words, the ears sense frequencies of mechanical waves of sound. In order to understand the meaning of frequency, we must appeal to the concept of sinusoidal waves. A sine wave or sinusoid is a mathematical curve that describes a smooth repetitive oscillation. It is named after the function sine, of which it is the graph. Fourier shows that any motion can be expressed in terms of a sum of sinusoidal oscillations.

Let us consider a sinusoidal wave. A cycle is defined as one complete performance of a vibration, electric oscillation, current alternation, or other periodic process. For example, one cycle is marked by two successive peaks in the wave motion. The wave consists of cycle after cycle, with no end in sight. We can look at a wave in time t. The time duration of one cycle is called the period T. Let time t be measured in seconds. To summarize,

$$t = \text{number of seconds in question} \ ,$$
$$T = \text{number of seconds per cycle} \ ,$$
$$\frac{1}{T} = \text{cycles per second} \ .$$

Thus,

$$\frac{t}{T} = \frac{\text{number of seconds in question}}{\text{number of seconds per cycle}}$$
$$= \text{number of cycles in question} \ .$$

The angular frequency ω is defined as

$$\omega = \frac{2\pi}{T} = \text{radians per cycle/number of seconds per cycle}$$
$$= \text{radians per second} \ .$$

The argument of a sine function must be in radians. There are 2π radians per cycle. The argument is

$$\frac{2\pi t}{T} = \omega t = \text{radians per cycle} \times \text{number of cycles in question}$$
$$= \text{radians in question} .$$

A sinusoidal function would be

$$\cos \omega t = \cos \frac{2\pi t}{T} .$$

Instead of time, we can use distance. We can look at a wave in distance x. The spatial duration of one cycle is called the wavelength λ. Let time x be measured in meters. To summarize,

$$x = \text{number of meters in question} ,$$
$$\lambda = \text{number of meters per cycle} ,$$
$$\frac{1}{\lambda} = \text{cycles per meter} .$$

Thus,

$$\frac{x}{\lambda} = \frac{\text{number of meters in question}}{\text{number of meters per cycle}} = \text{number of cycles in question} .$$

The wavenumber k is defined as

$$k = \frac{2\pi}{\lambda} = \text{radians per cycle/number of seconds per cycle}$$
$$= \text{radians per second} .$$

The argument of a sine function must be in radians. There are 2π radians per cycle. The argument is

$$\frac{2\pi x}{\lambda} = kx = \text{radians per cycle} \times \text{number of cycles in question}$$
$$= \text{radians in question} .$$

A sinusoidal function would be

$$\cos kx = \cos \frac{2\pi x}{\lambda} . \tag{6}$$

To summarize, the period T of a sinusoidal wave is the interval of time over which the shape of the wave repeats. The period represents the crest-to-crest or trough-to-trough duration in seconds at a given distance. The wavelength λ of a sinusoidal wave is the interval of distance over which the shape of the wave repeats. The wavelength represents the crest-to-crest or trough-to-trough length in meters at a given time (see Figure 1).

What is the relationship between wavelength and period of a sinusoidal wave? The relationship is: the wave moves a distance of one wavelength λ in a time of one period T. Velocity is distance divided by time. Thus,

$$v = \text{velocity} = \frac{\text{distance}}{\text{time}} = \frac{\text{wavelength}}{\text{period}} = \frac{\lambda}{T} \quad . \tag{7}$$

The angular frequency ω is

$$\omega = \frac{2\pi}{T} \quad . \tag{8}$$

The wavenumber k is

$$k = \frac{2\pi}{\lambda} \quad . \tag{9}$$

As a result, velocity is given by

$$v = \frac{\lambda}{T} = \frac{\omega}{k} \quad . \tag{10}$$

The period is given by

$$T = \frac{2\pi}{\omega} \quad . \tag{11}$$

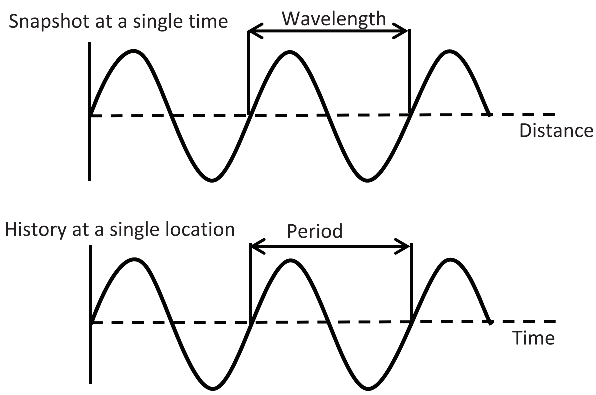


Figure 1. Wavelength and period.

As a result, we have

$$\lambda = vT = v\frac{2\pi}{\omega} = \frac{2\pi v}{\omega} \ . \tag{12}$$

A sinusoidal traveling wave (in one dimension x) may be represented by the equation

$$u(x,\, t) = A\cos\left(2\pi\left(\frac{t}{T} - \frac{x}{\lambda}\right) + \theta\right) = A\cos\phi \ . \tag{13}$$

Here, the constant A is amplitude of the wave, T is period, λ is wavelength, and θ is a constant phase angle. In the context of communication waveforms, the time-variant angle,

$$\phi = 2\pi\left(\frac{t}{T} - \frac{x}{\lambda}\right) + \theta = \left(\frac{2\pi t}{T} - \frac{2\pi x}{\lambda}\right) + \theta \ ,$$

is referred to as the *instantaneous phase* of the wave at x and t. The instantaneous phase also can be written as

$$\phi = (\omega t - kx) + \theta$$

or as

$$\phi = \omega\left(t - \frac{k}{\omega}x\right) + \theta = \omega\left(t - \frac{x}{v}\right) + \theta \ . \tag{14}$$

Here, ω is angular frequency, k is wavenumber, v is velocity, and θ is a constant phase angle. Thus, the sinusoidal traveling wave, shown in equation 13, may be represented by the equation

$$u(x,\, t) = A\cos\left(\omega\left(t - \frac{x}{v}\right) + \theta\right) \ . \tag{15}$$

The sinusoidal wave $u(x,\, t)$ satisfies our one-dimensional wave equation 3.

Plane waves

Two types of wave are of special interest in the three-dimensional case. One is the plane wave, which is treated in this section. The other is the spherical wave, which will be treated in the next section. Suppose you form a line of soldiers side by side, all facing in a certain direction, such as northeast. They all march at the same speed in that direction. What you see is the whole line moving in that direction. It is the same with waves.

In two dimensions, a line wave is a line composed of the same waveform. The line wave moves at a given speed in a certain direction. In three dimensions, a plane wave is a plane composed of the same waveform. The plane wave moves at a given speed in a certain direction.

A plane wave propagates along an arbitrary direction specified by the unit vector

$$\boldsymbol{u} = (\alpha,\ \beta,\ \gamma)\ .$$

Here, we must be careful to distinguish between the boldface $\boldsymbol{u}$ for unit vector and the italic u for the quantity undergoing wave motion. By definition, the unit vector has a magnitude of one; that is,

$$1 = |\boldsymbol{u}| = \sqrt{\alpha^2 + \beta^2 + \gamma^2}\ .$$

Let $\boldsymbol{r}$ be the vector

$$\boldsymbol{r} = (x,\ y,\ z)\ .$$

Its length is

$$r = |\boldsymbol{r}| = \sqrt{x^2 + y^2 + z^2}\ .$$

A plane wave traveling in the $\boldsymbol{u}$ direction is

$$u(\boldsymbol{r},\ t) = f\left(t - \frac{\boldsymbol{r}\cdot\boldsymbol{u}}{v}\right) = f\left(t - \frac{\alpha x + \beta y + \gamma z}{v}\right)\ . \tag{16}$$

Let us verify that equation 16 is a solution of the three-dimensional wave equation. Because

$$\frac{\partial f}{\partial x} = -\frac{\alpha}{v}f' \quad \text{and} \quad \frac{\partial f'}{\partial x} = -\frac{\alpha}{v}f''\ ,$$

it follows that

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}\left(-\frac{\alpha}{v}f'\right) = -\frac{\alpha}{v}\frac{\partial f'}{\partial x} = -\frac{\alpha}{v}\left(-\frac{\alpha}{v}f''\right) = \frac{\alpha^2}{v^2}f''\ . \tag{17}$$

Thus,

$$\frac{\partial^2 f}{\partial x^2} = \frac{\alpha^2}{v^2}f''\ .$$

A similar result holds for y and for z; namely,

$$\frac{\partial^2 f}{\partial y^2} = \frac{\beta^2}{v^2} f'' \ ,$$

$$\frac{\partial^2 f}{\partial z^2} = \frac{\gamma^2}{v^2} f'' \ .$$

We have

$$\nabla^2 f = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = \frac{\alpha^2 + \beta^2 + \gamma^2}{v^2} f'' = \frac{1}{v^2} f'' = \frac{1}{v^2}\frac{\partial^2 f}{\partial t^2} \ . \qquad (18)$$

This equation verifies that equation 16 is a solution of the three-dimensional wave equation.

If we interpret what equation 16 means, we see that a constant value of the disturbance u occurs at a given value of t for all values of the position vector $\boldsymbol{r}$ obeying $\boldsymbol{r} \cdot \boldsymbol{u} =$ constant. This condition describes a plane perpendicular to $\boldsymbol{u}$. As the value of $\boldsymbol{r} \cdot \boldsymbol{u}$ increases, the plane wave moves in the $\boldsymbol{u}$ direction. The corresponding expression for a plane wave moving in the $-\boldsymbol{u}$ direction is

$$u(\boldsymbol{r}, t) = g\left(t + \frac{\boldsymbol{r} \cdot \boldsymbol{u}}{v}\right) \ . \qquad (19)$$

The propagation vector (or angular wavenumber vector) $\boldsymbol{k}$ points in the propagation direction.

Let us look at the sinusoidal plane wave

$$u(\boldsymbol{r}, t) = \cos\left[\omega\left(t - \frac{\boldsymbol{r} \cdot \boldsymbol{u}}{v}\right) + \theta\right] \ . \qquad (20)$$

The propagation vector is the vector given by

$$\boldsymbol{k} = \frac{\omega}{v}\boldsymbol{u} \ .$$

Let k_x, k_y, k_z be the three components of $\boldsymbol{k}$ along the coordinate axes. We have

$$\boldsymbol{k} = (k_x, k_y, k_z) \ .$$

Suppose α, β, γ are the three components of the unit vector $\boldsymbol{u}$. We have

$$\boldsymbol{u} = (\alpha, \beta, \gamma) = \alpha \boldsymbol{i_x} + \beta \boldsymbol{i_y} + k_z \boldsymbol{i_z} \ , \qquad (21)$$

which gives

$$(k_x, k_y, k_z) = \left(\frac{\omega}{v}\alpha,\ \frac{\omega}{v}\beta,\ \frac{\omega}{v}\gamma\right) . \tag{22}$$

The magnitude of $\boldsymbol{k}$ is

$$k = |\boldsymbol{k}| = \frac{\omega}{v} . \tag{23}$$

The magnitude k is called the *angular wave number*. The wavelength

$$\lambda = \frac{2\pi v}{\omega} = \frac{2\pi}{k} \tag{24}$$

represents the crest-to-crest or trough-to-trough spatial distance between waves at a given time. This distance is measured in the direction of wave propagation, i.e., the direction of $\boldsymbol{k}$. The crest-to-crest or trough-to-trough distance between waves at a given time depends upon the direction. If the distance is measured in the x direction, we obtain from equation 22 the x-angular wave number

$$k_x = \frac{\omega\alpha}{v} ,$$

from which we obtain the x-angular wavelength

$$\lambda_x = \frac{\lambda}{\alpha} = \frac{2\pi}{k\alpha} = \frac{2\pi v}{\omega\alpha} = \frac{2\pi}{k_x} . \tag{25}$$

Similarly, the y-wavelength and z-wavelengths are

$$\lambda_y = \frac{\lambda}{\beta} = \frac{2\pi}{k_y}, \quad \lambda_z = \frac{\lambda}{\gamma} = \frac{2\pi}{k_z} . \tag{26}$$

The period of the sinusoidal wave is

$$T = \frac{2\pi}{\omega} . \tag{27}$$

From equation 24, we see that the velocity v is

$$v = \frac{\omega\lambda}{2\pi} = \frac{\lambda}{T} , \tag{28}$$

which says that a crest moves a distance λ in the $\boldsymbol{k}$ direction in one period T. In the x-direction, the crest moves a distance λ_x in one period T, so the apparent velocity in the x direction is

$$v_x = \frac{\lambda_x}{T} = \frac{\lambda}{T\alpha} = \frac{v}{\alpha} \ . \tag{29}$$

This apparent velocity is greater than the wave velocity v, unless $\alpha = 1$, in which case it is equal to v. Similarly, the apparent velocities in the y and z directions are

$$v_y = \frac{v}{\beta}, \quad v_z = \frac{v}{\gamma} \ . \tag{30}$$

Spherical waves

Of all three-dimensional waves, only the plane wave (sinusoidal or not) moves through space with an unchanging profile. Thus, the idea of a wave as a propagating disturbance whose profile is unaltered is not generally true in the three-dimensional case as it is in the one-dimensional case. However, there is another special type of three-dimensional wave which retains its shape but not its amplitude as it propagates. This wave is the spherical wave. For a spherical wave, we make the assumption that the function $u(\boldsymbol{r}t)$ has spherical symmetry about the origin; namely, the assumption that $u(\boldsymbol{r}t) = u(rt)$. The scalar r is the length of the vector $\boldsymbol{r}$; that is,

$$r = |\boldsymbol{r}| = \sqrt{x^2 + y^2 + z^2} \ . \tag{31}$$

If we throw a stone into a pool, the surface ripples that emanate from the point of impact spread out in two-dimensional circular waves. Extending this concept to three dimensions, we envision a very small pulsating sphere (i.e., an approximate point source) surrounded by a fluid. As the source expands and contracts, it generates pressure variations which propagate outward as spherical waves.

An idealized point source is one for which the radiation emanating from it streams out radially, uniformly in all directions. The source is said to be isotropic and the resulting wavefronts are again concentric spheres which increase in diameter as they expand out into the surrounding space. The obvious symmetry of the wavefronts suggests that it might be more convenient to describe them in terms of spherical polar coordinates (see Figure 2).

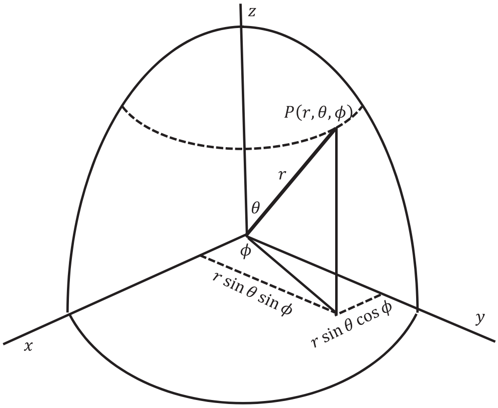


Figure 2. Spherical polar coordinates.

In this representation, the Laplacian operator is

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{r^2 \sin\theta}\,\frac{\partial}{\partial\theta}\left(\sin\theta\,\frac{\partial}{\partial\theta}\right) + \frac{1}{r^2 \sin^2\theta}\frac{\partial^2}{\partial\phi^2}\ , \tag{32}$$

where r, θ, ϕ are defined, respectively, by

$$x = r\sin\theta\cos\phi, \quad y = r\sin\theta\sin\phi, \quad z = r\cos\phi\ . \tag{33}$$

Remember that we are looking for a description of spherical waves, waves which are spherically symmetric. In other words, waves that are characterized by the fact that they do not depend on θ and ϕ. In such a case, we can use the simplified expression

$$u(\boldsymbol{r}) = u(r,\ \theta,\ \phi) = u(r)\ . \tag{34}$$

Then the Laplacian of $u(r)$ is simply

$$\nabla^2 u(r) = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial u}{\partial r}\right)\ . \tag{35}$$

This result also can be obtained in an alternate way. We start with the Cartesian form of the Laplacian, operate on the spherically symmetric wave function $u(r)$, and convert each term to polar coordinates. Examining the x dependence, we know by spherical symmetry that

$$u(\boldsymbol{r}) = u(r)\ .$$

We have

$$\frac{\partial u}{\partial x} = \frac{\partial u}{\partial r}\ \frac{\partial r}{\partial x}, \qquad \frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 u}{\partial r^2}\ \left(\frac{\partial r}{\partial x}\right)^2 + \frac{\partial u}{\partial r}\frac{\partial^2 r}{\partial x^2}\ . \tag{36}$$

Using

$$x^2 + y^2 + z^2 = r^2\ ,$$

we have

$$\frac{\partial r}{\partial x} = \frac{x}{r}, \qquad \frac{\partial^2 r}{\partial x^2} = \frac{1}{r}\ \frac{\partial}{\partial x}(x) + x\frac{\partial}{\partial x}\left(\frac{1}{r}\right) = \frac{1}{r}\ \left(1 - \frac{x^2}{r^2}\right)\ .$$

Thus, we have

$$\frac{\partial^2 u}{\partial x^2} = \frac{x^2}{r^2}\frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\ \left(1 - \frac{x^2}{r^2}\right)\frac{\partial u}{\partial r} \tag{37}$$

and also the two similar expressions

$$\frac{\partial^2 u}{\partial y^2} = \frac{y^2}{r^2}\frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\left(1 - \frac{y^2}{r^2}\right)\frac{\partial u}{\partial r}\ , \tag{38}$$

$$\frac{\partial^2 u}{\partial z^2} = \frac{z^2}{r^2}\frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\left(1 - \frac{z^2}{r^2}\right)\frac{\partial u}{\partial r}\ . \tag{39}$$

Adding equations 37, 38, and 39, we obtain the Laplacian of $u(r)$ as simply

$$\nabla^2 u(r) = \frac{\partial^2 u}{\partial r^2} + \frac{2}{r}\frac{\partial u}{\partial r}\ . \tag{40}$$

Because

$$\frac{\partial^2 u}{\partial r^2} + \frac{2}{r}\frac{\partial u}{\partial r} = \frac{1}{r}\left[\frac{\partial u}{\partial r} + \left(\frac{\partial u}{\partial r} + r\frac{\partial^2 u}{\partial r^2}\right)\right] = \frac{1}{r}\left[\frac{\partial u}{\partial r} + \frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right)\right] = \frac{1}{r}\ \frac{\partial^2 (ru)}{\partial r^2}\ ,$$

equation 40 also can be expressed as

$$\nabla^2 u(r) = \frac{1}{r}\ \frac{\partial^2 (ru)}{\partial r^2}\ . \tag{41}$$

In the case of spherical waves, the three-dimensional wave equation 2 becomes

$$\nabla^2 u(r) = \frac{1}{v^2}\frac{\partial^2 u(r)}{\partial t^2} \; . \tag{42}$$

Using equation 41, wave equation 42 can be written as

$$\frac{1}{r}\frac{\partial^2 (ru)}{\partial r^2} = \frac{1}{v^2}\;\frac{\partial^2 u}{\partial t^2} \; .$$

Multiplying both sides by r, we obtain

$$\frac{\partial^2 (ru)}{\partial r^2} = \frac{1}{v^2}\;\frac{\partial^2 (ru)}{\partial t^2} \; . \tag{43}$$

Notice that this expression is now just the one-dimensional wave equation where the space variable is r. The wave function is the product ru. The solution of equation 43 then is simply

$$ru(r,\, t) = f(r - vt)$$

or

$$u(r,\, t) = \frac{f(r - vt)}{r} \; . \tag{44}$$

This represents a spherical wave progressing radially outward from the origin at a constant speed v, and having an arbitrary functional form f, but attenuated as $1/r$. Another solution is

$$u(r,\, t) = \frac{g(r + vt)}{r} \; . \tag{45}$$

In this case, the wave converges toward the origin. A special case of the general solution,

$$u(r,\, t) = c_1\frac{f(r - vt)}{r} + c_2\frac{g(r + vt)}{r} \; , \tag{46}$$

is the sinusoidal spherical wave

$$u(r,\, t) = \frac{A}{r}\cos k(r \mp vt) \; . \tag{47}$$

The constant A is called the source strength. At any fixed value of time, this represents a cluster of concentric spheres filling all space. Each wavefront, or surface of constant phase, is given by $kr = constant$. Notice that the amplitude of the spherical wave in equation 47 is the function A/r, where the term r^{-1} serves as an attenuation factor. The attenuation factor is a direct consequence of energy conservation. Often, the attenuation factor goes under the name of geometric spreading. That is, unlike the plane wave, a spherical wave decreases in amplitude, thereby changing its profile, as it expands and moves out from the origin. As a spherical wavefront propagates out, its radius increases. Far enough away from the source, a small area of the wavefront will closely resemble a portion of a plane wave (see Figure 3).

If we visualize a pencil (very narrow beam) of rays spreading out radially from the source point, we can see that the intensity (i.e., the flow of energy across unit area in unit time) decreases as the area increases. This is the wellknown *inverse square law*. Because amplitude is proportional to the square root of intensity, the amplitude of a spherical wave decays as the inverse first power of distance, as given by equation 45. Thus, the decay, or spherical spreading relation

$$\frac{1}{r} = \frac{1}{vt} \ , \tag{48}$$

is applicable to a constant-velocity medium, for then the rays remain radial and the wavefront stays spherical. In the real earth, however, the actual velocity-depth relations result in refraction caused by bending or curving of the rays, so that the wavefronts no longer remain spherical. In the usual case, a greater decay results.

The effect of geometric spreading is the dominant effect in producing the observed decay of the amplitude of the raw seismic traces as time increases. In comparison with the other complexities observed on a seismogram, the phenomenon of geometric spreading is easy to handle. It is a definite effect which is observed, it is a major effect, and it is easy to compensate. All we need to do is multiply each value of the seismic trace by a factor proportional to the travel distance r.

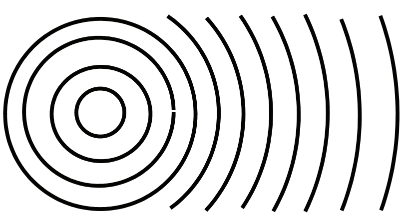


Figure 3. The radius of the spherical wavefront increases as it propagates away from its center.

Green's function

George Green (1793 – 1841) was born in Nottinghamshire and lived there for most of his life. His father was a baker who built and owned a brick windmill used to grind grain. Green began working daily in his father's bakery at the age of about five. He had to move sacks of grain and adjust the sails of the windmill. Every month, the heavy millstones had to be replaced or repaired. Air inside the mill was laden with finely granulated particles, harmful to health. Green was self-taught. He received only about one year of formal schooling as a child, between the ages of eight and nine. When Green was thirsty for knowledge, he became a member of the Nottingham Subscription Library. This membership gave Green the resources of books from which he could learn mathematics. In 1828, Green published *An Essay on the Application of Mathematical Analysis to the Theories of Electricity and Magnetism*. He published it privately at his own expense, because no established journal would publish anything by a person without a formal education. His *Essay* was purchased by about 50 people, mostly friends who could not understand any of the mathematics. His *Essay* introduced several significant advances, among them Green's theorem, potential functions, and Green's functions, all of which are fundamental to mathematical physics. Members of the Nottingham Subscription Library suggested that Green obtain a university education. Sir Edward Bromhead, one of the library's subscribers, encouraged Green to apply to Cambridge. Accordingly, Green turned to learning the Greek and Latin languages, as required for admission. In 1832, at age of nearly 40, Green was admitted as an undergraduate at Cambridge. He won the first-year mathematical prize. He graduated in 1838 as the fourth highest scoring student in his graduating class (while mathematician James Joseph Sylvester was the second highest). The Cambridge Philosophical Society, on the basis of Green's *Essay* and three other publications, elected Green as a fellow. In the next two years (1838 – 1840), Green published an additional six papers with applications to hydrodynamics, sound, and optics. In 1840, Green became ill, and he died a year later at age 47.

Green was the first person to create a mathematical theory of electricity and magnetism, and his theory formed the foundation for the subsequent work of James Clerk Maxwell, William Thomson, and others. Green's work on potential theory ran parallel to that of Carl Friedrich Gauss. Green's work was not recognized during his lifetime. In 1845, however, William Thomson (then aged 21), later known as Lord Kelvin, discovered Green's work and disseminated it for future mathematicians. Green's work on the motion of waves anticipated the WKBJ approximation of quantum mechanics. Green's efforts on light-waves and the properties of the ether produced what is now known as the Cauchy-Green tensor. Green's theorem and Green's functions became important tools in quantum mechanics, classical mechanics, and signal processing.

In those cases, there are many advantages of the wavefield being able to be described by a linear equation. In such a representation, the effect of independent inputs is additive. Thus, the effect due to a complicated input can be analyzed by expressing the complicated input as a superposition of simple inputs. Once the effect of the simple input is known, then it follows from the property of linear filters that the effect of the complicated input can be found. Consider the filter shown in Figure 4. Let us confine our attention to time filters. The symbol t denotes time.

Let G represent a filter (or operator) that transforms the input signal $q(t)$ into the output signal $u(t)$. In symbols,

$$Gq(t) = u(t) \ .$$

The filter is called linear if it satisfies:

1. the *superposition property*—namely, the input $q_1(t) + q_2(t)$ gives the output $u_1(t) + u_2(t)$; and
2. the *multiplicative property*—namely, the input $c\ q(t)$ gives the output $c\ u(t)$, where c is a constant.

A linear filter G is *time invariant* if it has the property

$$Gq(t - \tau) = u(t - \tau) \ .$$

In other words, if the input $q(t)$ is shifted by τ time units, then the output $u(t)$ also is shifted by τ time units.

The Dirac delta function $\delta(t)$ is equal to zero everywhere on the t axis except at $t = 0$. At zero, the delta function is a spike that is infinitely high and infinitesimally thin. However, the total area under the spike is equal to one. The most valuable property of the delta function is its sifting property; namely,

$$\int_{-\infty}^{\infty} q(\tau)\ \delta(t - \tau) d\tau = q(t) \ .$$

Figure 4. The filter or operator G transforms the input $q(t)$ into the output $u(t)$.

Now, let us now introduce the concept of *Green's function*, named after George Green. This concept applies to linear differential equations (both ordinary and partial). The concept applies to linear filters (time filters, space filters, and space–time filters). The three-dimensional wave equation 2 can be written as

$$\nabla^2 u - \frac{1}{v^2}\frac{\partial^2 u}{\partial t^2} = 0 \ .$$

This equation is a *homogeneous* partial differential equation because its right side is zero. In solving this differential equation, we obtain the output u. There is no input involved. The *inhomogeneous* equation is

$$\nabla^2 u - \frac{1}{v^2}\frac{\partial^2 u}{\partial t^2} = q \ .$$

In this equation, there is both input q and output u. The inhomogeneous equation also can be written as

$$\left(\nabla^2 - \frac{1}{v^2}\frac{\partial^2}{\partial t^2}\right) u = q \ .$$

The expression in parentheses is a differential operator, which we denote by L. Thus, the inhomogeneous equation is

$$Lu = q \ .$$

This equation says that the operator L acting on the output u is equal to the input q. Define G to be the inverse of L; that is,

$$G = L^{-1} \ .$$

We have

$$u = L^{-1} q \quad \text{or} \quad u = Gq \ .$$

This equation says that the operator G acting on the input q is equal to the output u.

For simplicity, we will describe Green's function for a time-invariant linear filter G, as shown in Figure 5. Let the input be the delta function $\delta(t - \tau)$. Green's function $g(t)$ is defined as the resulting output. In symbols,

$$G\delta(t) = g(t) \ .$$

Figure 5. The filter or operator G transforms delta function $\delta(t)$ into Green's function $g(t)$.

Because the filter is time-invariant, we have

$$G\delta(t - \tau) = g(t - \tau) \ .$$

Because filter G is linear, it satisfies the multiplicative property

$$G[q(\tau)\delta(t - \tau)] = q(\tau)g(t - \tau) \ .$$

Here, $q(\tau)$ is an arbitrary constant, defined to be the value of a time function $q(t)$ at some fixed time τ. The superposition property says that we may integrate over all possible values of τ. As a result, we obtain

$$\int_{-\infty}^{\infty} G[q(\tau)\ \delta(t - \tau)d\tau] = \int_{-\infty}^{\infty} q(\tau)g(t - \tau)d\tau \ .$$

Because G operates on t, not τ, we can take G outside of the integral. Thus, we have

$$G\left[\int_{-\infty}^{\infty} q(\tau)\ \delta(t - \tau)d\tau\right] = \int_{-\infty}^{\infty} q(\tau)g(t - \tau)d\tau \ .$$

We can apply the sifting property to the expression within the brackets. The result is

$$G[q(t)] = \int_{-\infty}^{\infty} q(\tau)g(t - \tau)d\tau \ .$$

As we know, the output is

$$Gq(t) = u(t) \ .$$

The final result is

$$u(t) = \int_{-\infty}^{\infty} q(\tau)g(t - \tau)d\tau \ .$$

The above equation is called the *superposition integral*. It states that the output $u(t)$ of a time-invariant linear filter with input $q(t)$ is obtained by (1) finding Green's function $g(t)$ and then (2) using the superposition integral to determine the output $u(t)$. In reflection seismology, the superposition integral is known as the *convolution integral*, where the input $q(\tau)$ is the reflectivity and the output $u(t)$ is the seismic trace. Green's function $g(t)$ is the seismic wavelet and $L = G^{-1}$ is the deconvolution operator.

In electrical engineering, the Green's function is called the impulse response function. In optics, the Green's function is referred to as the point spread function of a spatially-varying optical filter. In physics and wave propagation, the Green's function represents the effect of a linear wave at point r and time t resulting from an impulsive source applied at point r_0 at time t_0. Although the idea behind the Green's function is the same, the terminology is different for different fields of science. Regardless of the area of application, the fundamental fact is that the Green's function is the response of a system to an impulsive excitation.

Green's functions are used in physics, specifically in quantum field theory, aerodynamics, aeroacoustics, electrodynamics, statistical field theory, and Born scattering theory. In quantum field theory, Green's functions take the roles of propagators. Green's functions are used widely in electrodynamics and quantum field theory, where the relevant differential operators are often difficult or impossible to solve exactly but can be solved in a perturbative manner using Green's functions. In field theory contexts, the Green's function often is called the propagator or two-point correlation function because it is related to the probability of measuring a field at one point given that it is sourced at a different point. Green's functions are powerful tools for obtaining relatively simple and general solutions of basic problems, such as scattering and bound-level information. The bound-level treatment gives a clear physical understanding of questions such as superconductivity, the Kondo effect, and, to a lesser degree, disorder-induced localization.

Next, we would like to make the wave equation easier to understand, and to do so we will use Green's function. Let us introduce the source function, or input, $q(x, y, z, t)$. Accordingly, the inhomogeneous form of the wave equation 1 is

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} - \frac{1}{v^2}\frac{\partial^2 t}{\partial z^2} = q(x, y, z, t) \ . \tag{49}$$

We interpret equation 49 as a linear system with input q and output u, as shown in Figure 6. In addressing input–output systems, the key factor is the

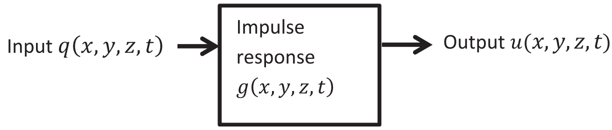


Figure 6. Green's function g transforms input q into output u.

determination of the impulse response function. In effect, it is Christiaan Huygens who discovers this impulse response function, at least empirically, as 17th-century mathematics was not developed enough to write down its mathematical form. The study of wave propagation in three dimensions can be made quite complicated. Huygens' contribution is discovering how to make it beautifully simple. He observes that a point source radiates a spherically symmetric wave-like disturbance, or wavelet. A point source at the origin can be represented by the product $\delta(x)\delta(y)\delta(z)$ of three Dirac delta functions. If, in addition, the time function at the source is an impulse or spike $\delta(t)$, then an impulsive point source would have the form

$$\delta(x)\delta(y)\delta(z)\delta(t)\ . \tag{50}$$

We now let the impulsive point source 50 be the input $q(x, y, z, t)$. The resulting output, by definition, is the impulse response function. Instead of the terminology "impulse response function," physicists use the expression "Green's function." Without proof, we write down the Green's function, which in the present case is

$$g(x, y, z, t) = \frac{\delta\left(t - \frac{r}{v}\right)}{-4\pi r}\ , \tag{51}$$

where $t > 0$ and r is defined as

$$r = \sqrt{x^2 + y^2 + z^2}\ . \tag{52}$$

We see that r is the length of the vector from the origin to the point (x, y, z). The quantity r/v is the travel time for the impulsive disturbance to travel from the origin to the point (x, y, z). The delta function $\delta(t - r/v)$ represents a spike disturbance on the spherical shell with the center at the origin and radius $r = vt$. This result is in conformity with Huygens, in that the impulsive point source has produced an impulsive spherically symmetric wavelet. The divisor -4π in equation 51 is a constant which comes from the mathematical derivation, while the divisor r in equation 51 represents spherical divergence.

Thus, we have stated that the Green's function for the three-dimensional wave equation is an impulsive spherical shell (with a spherical divergence factor). The equation for this spherical shell is $r = vt$ or

$$x^2 + y^2 + z^2 = v^2t^2 \ . \tag{53}$$

In four dimensions (x, y, z, t), this is the equation for a cone with the vertex at the source point (0, 0, 0, 0). Thus, a spike at the source results in a spike at all points on this cone, and zero everywhere else. We cannot draw a four-dimensional figure, but, if we suppress in our minds the y coordinate, this fundamental cone looks like the one depicted in Figure 7.

Because we now know the impulse response function, we can exhibit the input–output relationship as

$$\text{input } q(x,\, y,\, z,\, t) \text{ convolved with } \frac{\delta\left(t - \frac{r}{v}\right)}{-4\pi r} \text{ gives output } u(x,\, y,\, z,\, t) \ .$$

This convolutional relationship between input and output may be written as

$$\text{output} = \left(\frac{\delta\left(t - \frac{r}{v}\right)}{-4\pi r}\right) **** \ (\text{input}) \ ,$$

where the four asterisks indicate convolution with respect to the four independent variables. Four-dimensional convolution might seem a little formidable at first, but, as we will see, no new principles beyond one-dimensional convolution are needed.

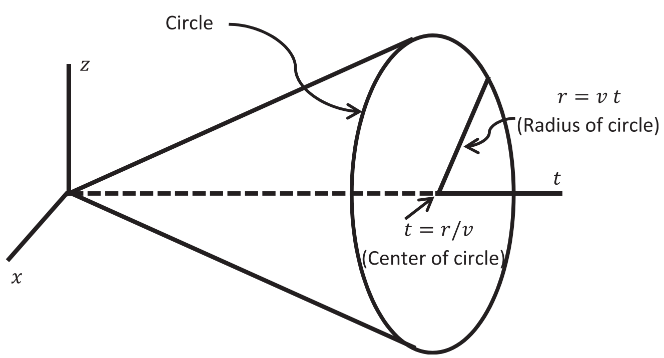


Figure 7. Fundamental cone describing spherical wave propagation. The circle represents the three-dimensional spherical shell at time $t = r/v$. The disturbance is a spike (with the spherical divergence factor) at all points on this cone, and zero elsewhere.

As an example, consider an input given by the point source wavelet $\delta(x)\delta(y)\delta(z)f(t)$, where $f(t)$ is some transient wave shape. The output $u(x, y, z, t)$ is given by the convolution equation

$$u(x, y, z, t) = (\text{Green's function}) **** (\delta(x)\delta(y)\delta(z)f(t)) \ . \tag{54}$$

The Green's function in equation 54 is

$$g(x, y, z, t) = \frac{\delta\left(t - \frac{1}{v}\sqrt{x^2 + y^2 + z^2}\right)}{-4\pi\sqrt{x^2 + y^2 + z^2}} \ . \tag{55}$$

Thus, equation 54 becomes (with each integral running from $-\infty$ to ∞)

$$u = \int d\xi \int d\eta \int d\zeta \int d\tau \frac{\delta\left(\tau - \frac{\sqrt{\xi^2 + \eta^2 + \zeta^2}}{v}\right)}{-4\pi\sqrt{\xi^2 + \eta^2 + \zeta^2}} \times \delta(x - \xi)\delta(y - \eta)\delta(z - \zeta)f(t - \tau) \ .$$

We now use the sifting property of the delta function for each of the first three integrals to obtain

$$u(x, y, z, t) = \int_{-\infty}^{\infty} d\tau \ \frac{\delta\left(\tau - \frac{\sqrt{x^2 + y^2 + z^2}}{v}\right)}{-4\pi\sqrt{x^2 + y^2 + z^2}} f(t - \tau) \ . \tag{56}$$

The radius is $r = \sqrt{x^2 + y^2 + z^2}$. Equation 56 becomes

$$u(x, y, z, t) = \int_{-\infty}^{\infty} d\tau \ \frac{\delta\left(\tau - \frac{r}{v}\right)}{-4\pi r} f(t - \tau) \ . \tag{57}$$

If we use the sifting property of the delta function, then equation 57 becomes

$$u(x, y, z, t) = \frac{f(t - r/v)}{-4\pi r} \ . \tag{58}$$

In most books, this fundamental result is obtained from first principles. In keeping with our input–output theme, however, we have obtained this result from the Green's function. We have written down the impulse response function, or, in other words, the Green's function, for the three-dimensional wave equation. As we have seen, it is an impulsive spherical shell (the Huygens' spherical wavelet) expanding outward from the impulsive point source as time increases.

Dispersion equation

In exponential form, the pure sinusoidal plane wave 20 is represented by

$$\begin{aligned} u(x,\, t) &= \exp\left[i\left(\omega t - \frac{\boldsymbol{u}\cdot\boldsymbol{r}}{v}\right)\right] = \exp[i(\omega t - \boldsymbol{k}\cdot\boldsymbol{r})] \\ &= \exp[i(\omega t - k_x x - k_y x - k_z x)] = \exp[i\phi] \ . \end{aligned} \tag{59}$$

The phase is

$$\phi = \omega t - k_x x - k_y y - k_z z \ . \tag{60}$$

Let us now substitute the sinusoidal wave 59 into the wave equation 1. First, we evaluate the derivatives

$$\begin{aligned} \frac{\partial^2 u}{\partial x^2} &= (-ik_x)^2 e^{i\phi} = -k_x^2 e^{i\phi} \ , \\ \frac{\partial^2 u}{\partial y^2} &= (ik_y)^2 e^{i\phi} = -k_y^2 e^{i\phi} \ , \\ \frac{\partial^2 u}{\partial t^2} &= (i\omega)^2 e^{i\phi} = -\omega^2 e^{i\phi} \ . \end{aligned}$$

Then, we substitute the above values into wave equation 1 and thus obtain

$$-k_x^2 - k_y^2 - k_z^2 - \frac{1}{v^2}(-\omega^2) = 0 \ ,$$

which is

$$k_x^2 + k_y^2 + k_z^2 = \frac{\omega^2}{v^2} \ . \tag{61}$$

This equation is called the dispersion equation for the three-dimensional wave equation. The dispersion equation for the two-dimensional wave equation 3 is

$$k_x^2 + k_z^2 = \frac{\omega^2}{v^2} \ . \tag{62}$$

The dispersion equation for the one-dimensional wave equation 4 is

$$k_z^2 = \frac{\omega^2}{v^2} \ . \tag{63}$$

Now, let us look at the one-dimensional case. For a given value of v, the frequency ω and the wavenumber k_z must satisfy the dispersion equation 63. If we solve for ω in terms of k_z, we obtain

$$\omega = \pm v k_z \ . \tag{64}$$

Thus, in the case of the one-dimensional wave equation with constant velocity v, the frequency ω is proportional to the wavenumber k_z. The constant of proportionality is either $+v$ or $-v$. The choice $+v$ means that the wave is

$$\exp i\phi = \exp i(\omega t - k_z z) = \exp i k_z (vt - z) \ . \tag{65}$$

This wave is propagating in the positive z direction. The choice $-v$ means that the wave is propagating in the negative z direction.

Because a crest of a sinusoidal wave moves with velocity v in the direction of the propagation vector $\boldsymbol{k}$, it is called a *traveling wave*. We now come to an important feature of wave motion, namely, the phenomenon of evanescent waves. As an example, let us solve the dispersion equation 61 for k_z. We obtain

$$k_z = \pm \sqrt{\frac{\omega^2}{v^2} - k_x^2 - k_y^2} \ . \tag{66}$$

Two cases may occur, namely,

$$k_x^2 + k_y^2 \leq \frac{\omega^2}{v^2} \tag{67}$$

and

$$k_x^2 + k_y^2 > \frac{\omega^2}{v^2} \ . \tag{68}$$

In the first case, the quantity k_z is real, whereas in the second case the quantity k_z is imaginary. In our discussion to this point, we have tacitly assumed that the first case holds, in which case the crests of the waves travel in either the plus or minus z-direction. In the second case, define κ_z by $k_z = -i\kappa_z$, so

$$\exp i\phi = \exp i(\omega t - k_x x - k_y y + i\kappa_z z) = \exp i(\omega t - k_x x - k_y y) \exp -\kappa_z z \ .$$

Thus, in the z-direction, the wave is not sinusoidal but exponential. There are no crests, and the wave does not propagate in the z-direction, but instead the amplitude is attenuated exponentially in the z-direction. Such a wave is called an *evanescent wave* (or *exponential wave*), as opposed to the traveling wave (or sinusoidal wave) of the first case. The propagation vector of this evanescent wave has the form

$$\boldsymbol{k} = (k_x,\, k_y, -i\kappa_z)\ , \tag{69}$$

which has no real component in the z-direction. Note that in the case of the one-dimensional wave equation, the dispersion relation 64 cannot yield imaginary values of k_z, so evanescent waves cannot occur.

Group velocity

The dispersion relation of a wave equation gives a functional relationship between the components of the $\boldsymbol{k}$ vector and the angular frequency ω. In one dimension, we can plot ω as a function of k; that is, we can plot $\omega(k)$. In a plot of ω versus k, the slope of the straight line from the origin to some point on this curve, such as point $(\omega,\, k)$, is the phase velocity $v_p = \omega/k$ at this point. The tangent to the curve at some point ω is the group velocity $v_g = d\omega/dk$ at that point. Let us give two examples.

Example 1. The first example is that of a seismic wave in a homogeneous isotropic medium with material (rock) velocity v. The velocity v does not depend upon the frequency. Thus, the plot of $\omega(k)$ is simply a straight line through the origin; that is, $\omega = vk$. The phase and group velocities are

$$v_p = \frac{\omega}{k} = v, \quad v_g = \frac{d\omega}{dk} = \frac{d(vk)}{dk} = v\ . \tag{70}$$

We see that both the phase and the group velocities are equal to the material velocity v. Because we usually regard a rock layer as homogeneous and isotropic, this case is important in the analysis of seismic records.

Example 2. In this example, an observer on the surface of the earth measures only the horizontal component of a wave. Suppose, however, that the wave is two-dimensional, yet the surface observer is unaware of the depth dimension. The dispersion relation (equation 62) is

$$\omega^2 = v^2 k_x^2 + v^2 k_z^2\ ,$$

where the constant v is the rock velocity. The surface observer sees this dispersion relation as

$$\omega^2 = v^2k^2 + \omega_0^2 \ ,$$

as our k_x is his k, and our $v^2k_z^2$ is his fixed constant ω_0^2. His plot of frequency versus wavenumber is

$$\omega(k) = \sqrt{\omega_0^2 + v^2k^2} \ , \tag{71}$$

so his phase velocity is

$$v_p = \frac{\omega}{k} = \sqrt{\frac{\omega_0^2}{k^2} + v^2} \ , \tag{72}$$

which always exceeds the rock velocity v. His group velocity is

$$v_g = \frac{d\omega}{dk} = \frac{v^2k}{\sqrt{\omega_0^2 + v^2k^2}} = \frac{v^2}{v_p} = \frac{v}{v_p}v \ , \tag{73}$$

which is always less than the rock velocity, so $v_g < v$. Because the energy and the information travel at the group velocity v_g, the observer is unable to transmit intelligence at a velocity greater than the rock velocity. The observer cannot transmit a wave of any frequency ω that is less than ω_0. For, in that case, $\omega - \omega_0^2 = v^2k^2$ is negative, so k is imaginary. It is for this reason that the observer calls ω_0 the cutoff frequency; the observer sees the earth as a low cut (i.e., high pass) frequency filter.

The wavenumber vector $\boldsymbol{k}$ plays a fundamental role in the study of wave motion. The magnitude k of $\boldsymbol{k}$ is called the wavenumber. A sinusoidal plane wave (equation 59) has the form (with $\boldsymbol{k}$ and ω constant):

$$\exp i(\omega t - \boldsymbol{k} \cdot \boldsymbol{r}) \ . \tag{74}$$

The quantity

$$\phi = \omega t - \boldsymbol{k} \cdot \boldsymbol{r} \tag{75}$$

is the phase. At a fixed time, a wavefront is defined as a surface of constant phase. Thus, for this plane wave, the wavefront is the plane $\boldsymbol{k} \cdot \boldsymbol{r} =$ constant, and is normal to the vector $\boldsymbol{k}$. A crest of the wave occurs whenever the phase is a multiple of 2π; i.e., whenever

$$\phi = 0, \pm 2\pi, \pm 4\pi, \pm 6\pi, \ldots \ .$$

Let us measure distance in the direction of vector $\boldsymbol{k}$ and define the distance λ by $\lambda = 2\pi/k$. In this distance (for a fixed time), the phase changes by $-k\lambda = -2\pi$. Thus, the distance λ measures the distance (in the direction of $\boldsymbol{k}$) between crests at a fixed time t, and hence λ is the wavelength. The wavenumber is equal to the number of wavelengths in a distance of 2π in the direction of $\boldsymbol{k}$.

As time increases, each wavefront propagates in the direction of $\boldsymbol{k}$. Let us consider the wavefront defined by the crest $\phi = 0$. Another wavefront, of course, would serve equally as well. In the time span from $t = 0$ to $t = t$, this wavefront travels a distance r_k in the direction of $\boldsymbol{k}$, and we can write

$$\phi = 0 = \omega t - kr_k \ .$$

Thus, the velocity v of the wavefront is

$$v = \frac{r_k}{t} = \frac{\omega}{k} \ . \tag{76}$$

The velocity

$$v = \omega/k$$

is called the *phase velocity* v_p (see equation 70). It is the velocity of the wavefront in the direction of the wavenumber vector $\boldsymbol{k}$. The reciprocal q of the phase velocity gives the time for the wavefront to move a unit distance in the direction of $\boldsymbol{k}$. Because

$$q = \frac{1}{v} = \frac{k}{\omega} \tag{77}$$

represents time per unit distance, it is called the *slowness*.

Now, we would like to define an important concept, the *slowness vector*. The phase velocity v cannot be readily converted into a vector with physically meaningful components; however, slowness $1/v$ can be converted as follows. Because $1/v = k/\omega$, we define the slowness vector $\boldsymbol{q}$ as

$$\boldsymbol{q} = \frac{\boldsymbol{k}}{\omega} \ . \tag{78}$$

The magnitude of $\boldsymbol{q}$ is the slowness, that is,

$$q = \frac{k}{\omega} \ .$$

We have generalized the expression $v = \omega/k$ to show dependence upon the vector $\boldsymbol{k}$ rather than on the scalar k.

At this point, it would be helpful to express some of the foregoing vectors in terms of their components. Let α, β, γ be the direction cosines of the $\boldsymbol{k}$ vector. Then, we have

$$\boldsymbol{r} = (x,\, y,\, z)\ , \quad \boldsymbol{k} = (k_x,\, k_y,\, k_z) = k(\alpha,\, \beta,\, \gamma)\ , \tag{79}$$

so the phase is

$$\phi = \omega t - k_x x - k_y y - k_z z = \omega t - k(\alpha x + \beta y + \gamma z)\ . \tag{80}$$

The slowness vector is

$$\boldsymbol{q} = (q_x,\, q_y,\, q_z) = \left(\frac{k_x}{\omega},\, \frac{k_y}{\omega},\, \frac{k_z}{\omega}\right) = \frac{k}{\omega}(\alpha,\, \beta,\, \gamma) = \frac{1}{v}(\alpha,\, \beta,\, \gamma)\ . \tag{81}$$

The components of $\boldsymbol{q}$ are the horizontal slowness for x, the horizontal slowness for y, and the vertical slowness for z. Their reciprocals

$$v_x = \frac{1}{q_x}, \quad v_y = \frac{1}{q_y}, \quad v_z = \frac{1}{q_z} \tag{82}$$

are the x, y, z phase velocities, respectively. However, v_x, v_y, v_z is each greater than or equal to v, and hence in no way can be considered as the components of a velocity vector. Instead, their reciprocals are the components of the slowness vector.

The wavenumber $\boldsymbol{k} = \omega \boldsymbol{q}$ depends upon frequency ω. Let us plot $\boldsymbol{k} = (k_x,\, k_y,\, k_z)$ in three-dimensional space, with k_x, k_y, k_z plotted on the orthogonal axes. If we fix frequency at a given value ω, we see that the tip of vector $\boldsymbol{k}$ defines a surface called the *wavenumber surface*. This surface can be designated by

$$\omega(k_x,\, k_y,\, k_z) = \text{constant}\ . \tag{83}$$

We recall that the group velocity in the case of one dimension is defined as

$$v_g = \frac{d\omega}{dk}\ .$$

Strictly by analogy, we define the group velocity in the case of three dimensions as the vector

$$\boldsymbol{v}_g = \frac{d\omega}{d\boldsymbol{k}}\ . \tag{84}$$

Now, we must explain what we mean by this operation. Consider the vector

$$\boldsymbol{v}_g = \frac{d\omega}{d\boldsymbol{k}} = \left(\frac{\partial\omega}{\partial k_x}, \frac{\partial\omega}{\partial k_y}, \frac{\partial\omega}{\partial k_z}\right) . \tag{85}$$

It is natural to call $\partial\omega/\partial k_x$ the horizontal group velocity (as x is a horizontal direction) and similarly for $\partial\omega/\partial k_y$. Likewise, let us call $\partial\omega/\partial k_z$ the vertical group velocity (as z is the vertical direction). Because

$$\boldsymbol{q} = \frac{\boldsymbol{k}}{\omega} \quad \text{and} \quad \frac{\partial \boldsymbol{q}}{\partial k} = \frac{1}{\omega} ,$$

the group velocity vector from equation 85 also can be written as

$$\boldsymbol{v}_g = \frac{\partial \boldsymbol{q}}{\partial k}\frac{\partial\omega}{\partial \boldsymbol{q}} = \frac{1}{\omega}\frac{\partial\omega}{\partial \boldsymbol{q}} = \frac{1}{\omega}\left(\frac{\partial\omega}{\partial q_x}, \frac{\partial\omega}{\partial q_y}, \frac{\partial\omega}{\partial q_z}\right) . \tag{86}$$

Equations 85 and 86 show that the group velocity vector can be written as

$$\boldsymbol{v}_g = \nabla_k \omega = \frac{1}{\omega}\nabla_q \omega , \tag{87}$$

where ∇_k and ∇_q are the gradient operators with respect to the wavenumber vector $\boldsymbol{k}$ and the slowness vector $\boldsymbol{q}$, respectively. That is, ∇_k and ∇_q are defined as the vector operators given by

$$\nabla_k = \left(\frac{\partial}{\partial k_x}, \frac{\partial}{\partial k_y}, \frac{\partial}{\partial k_z}\right), \quad \nabla_q = \left(\frac{\partial}{\partial q_x}, \frac{\partial}{\partial q_y}, \frac{\partial}{\partial q_z}\right) . \tag{88}$$

From vector analysis, we know that the gradient to a surface is normal to this surface. Thus, the gradient $\nabla_k\omega$ is normal to the wavenumber surface $\omega(k_x, k_y, k_z) =$ constant. Because $\boldsymbol{v}_g$ is equal to this gradient, it follows that the group velocity vector is normal to the wavenumber surface (see Figure 8). Thus, we have found an important interpretation of group velocity; namely, that the group velocity can be expressed geometrically in terms of the wavenumber surface.

An isotropic medium is a medium in which the velocity $v(x, y, z)$ depends only upon the location of the point (x, y, z) and not upon the direction in which the velocity is measured. For such a medium, the slowness $q(x, y, z) = 1/v(x, y, z)$ is, therefore, also independent of direction. Because $\boldsymbol{k} = \omega\boldsymbol{q}$, it follows that, for an isotropic medium, the wavenumber surface $\omega(k_x, k_y, k_z)$ is a sphere. A geometrical property of a sphere is that

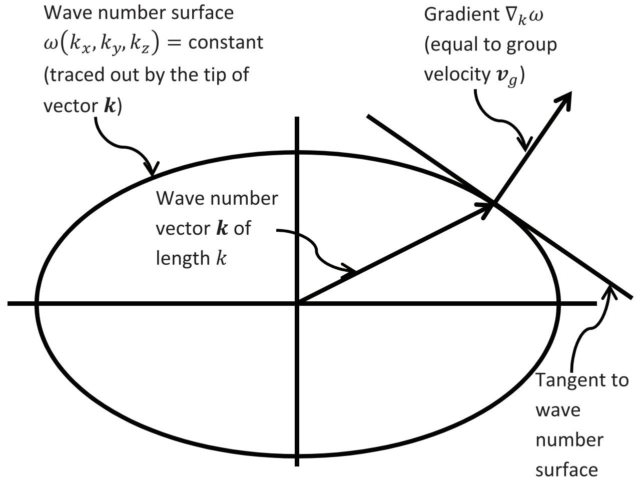


Figure 8. Anisotropic medium. The group velocity vector is normal to the wavenumber surface, but generally does not point in the direction of the propagation vector. (Note only two dimensions are shown in this diagram.)

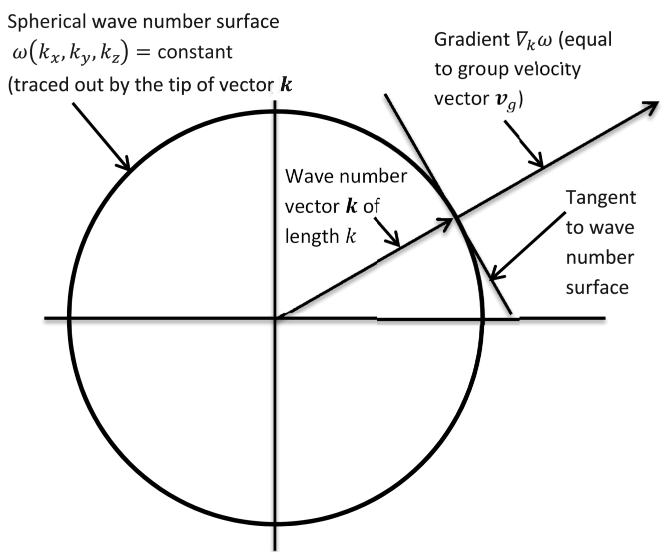


Figure 9. Isotropic medium. The group velocity vector is normal to the wave number surface, which is a sphere, and hence points in the direction of the propagation vector. (Note only two dimensions are shown in this diagram.)

the normal to the surface of a sphere is in the same direction as the radius vector. For the wavenumber surface, the normal is $\boldsymbol{v}_g$ and the radius is $\boldsymbol{k}$. Thus, we have established a fundamental theorem, namely: *For an isotropic medium, the group velocity vector* $\boldsymbol{v}_g$ *points in the same direction as the propagation vector* $\boldsymbol{k}$ (see Figure 9). Because intelligence travels with the group velocity, we see that intelligence in an isotropic medium travels in the propagation direction $\boldsymbol{k}$. The situation is more complicated in the case of an anisotropic medium (Figure 8). In that case, the group velocity vector $\boldsymbol{v}_g$ is no longer in the direction of $\boldsymbol{k}$.

This page has been intentionally left blank

Chapter 6

Ray Tracing and Seismic Modeling

Hamilton's equations

The concept of a ray is very useful. It is a line drawn in space that corresponds to the direction of the flow of radiant energy. Rays are a geometric idealization; they have no width. As such, a ray is a mathematical device rather than a physical entity. In practice, we can produce very narrow beams or pencils (as, for example, a laser beam), and we might imagine a ray to be the unattainable limit on the narrowness of such a beam. Similar to the lines of geometry, rays are a convenient fiction. They exist in the real world as a beam of light; and beams have width. In the same way, we may think of seismic rays as idealized beams in the direction of the flow of seismic energy.

A medium is *homogeneous* if its properties are not a function of position, i.e., if its properties do not vary from point to point. In other words, a homogeneous medium is identical throughout. Otherwise, the medium is said to be *inhomogeneous* or, alternatively, *heterogeneous*. If a homogeneous medium were cut into pieces, then every piece would be identical. If subsurface rocks found in the earth were figuratively cut into pieces, differences generally would be observed. Thus, seismology has to deal with inhomogeneous media.

A medium is *isotropic* if its properties (such as density and Young's modulus) do not depend upon a particular direction. If these properties are direction dependent, then we can say that the medium is *anisotropic*. More specifically, a material is said to be anisotropic if the value of a measurement of a rock property varies with direction.

Seismic anisotropy can be defined as the dependence of velocity upon direction. There are many different types of anisotropy. Of particular interest are the types that possess an axis of rotational invariance; namely, if the formation is rotated about such an axis, then the material is still indistinguishable from what it was before. Two cases of seismic anisotropy are of special interest: one is vertical transverse isotropy (VTI) and the other is horizontal transverse isotropy (HTI). Vertical transverse isotropy has a vertical axis of rotational invariance. This type of anisotropy is associated with layering and shale and is found where gravity is the dominant factor. Horizontal transverse isotropy, otherwise known as azimuthal anisotropy, has a horizontal axis of rotational invariance. This type of anisotropy is associated with cracks and fractures and is found where regional stress is the dominant factor.

Glass is an example of a homogeneous isotropic medium. Iceland spar is an example of a homogeneous anisotropic medium. In glass, the ray is orthogonal to wavefronts. In Iceland spar, there are two paths of light, known as the *ordinary ray* and the *extraordinary ray*. The ordinary ray is orthogonal to wavefronts. The extraordinary ray is not orthogonal to wavefronts.

Rays in a homogeneous medium are straight lines. Rays in an inhomogeneous medium are curved lines. The curved line does not have to be smooth. For example, if the inhomogeneous medium is composed of different homogeneous layers, the ray will be straight through each individual layer but will bend (by refraction) at each interface. Rays in an isotropic medium are orthogonal to wavefronts; more specifically, the rays are perpendicular to the wavefronts at every point of intersection. Accordingly, in an isotropic medium, a ray is parallel to the propagation vector $\boldsymbol{k}$. Except in special cases such as the ordinary ray in Iceland spar, rays in an anisotropic medium are not orthogonal to the wavefronts.

Anisotropy differs from the rock property called heterogeneity in that anisotropy is the variation in vectorial values with direction at a point while heterogeneity is the variation in scalar or vectorial values between two or more points. Table 1 describes the behavior of rays and wavefronts for various media.

Often, light can be thought of as rays. A water wave as seen on a pond appears as a set of moving wavefronts. If a stone is thrown into a pond, the crests (wavefronts) form a pattern of concentric circles. The energy of the disturbance travels outward radially from the center. That is, the energy is propagated along rays at right angles to the wavefront. If we carefully watch the wave motion, we observe that the longest waves appear at the outside of the expanding pattern of concentric circles. As we watch the progress of one of these outside crests, we will suddenly see them disappear. This is not an illusion, and the next crest coming from behind also

Table 1. Behavior of rays and wavefronts for varying media.

	Homogeneous	**Inhomogeneous**
Isotropic	Rays are straight lines. Wavefronts are orthogonal to rays. Example: glass	Rays are curved lines. Wavefronts are orthogonal to rays. Example: typical earth model
Anisotropic	Rays are straight lines. Wavefronts generally are not orthogonal to rays. Example: Iceland spar	Rays are curved lines. Wavefronts generally not orthogonal to rays. Example: advanced earth model

disappears. More and more crests keep coming from behind and disappear at the outside edge. On the inside of the ring, new crests keep appearing from the now calmed central water. The reason for this phenomenon is as follows. The wave packet represented by the concentric rings moves outward at the group velocity. The group velocity is the velocity at which the energy of the disturbance propagates outward. The crests, however, move at the phase velocity. Examining the physics, it can be shown that the group velocity is less than the phase velocity. Thus, the crest of each wave moves faster than the wave packet, and the crests move forward with respect to the concentric pattern. Once a crest reaches the outside of the pattern, it cannot go any further, because no energy has yet arrived there, and so the crest simply vanishes.

In many physical problems, we can find the dispersion relation for the wave equation governing the wave motion. In symbols, we can write such a dispersion relation as the mathematical equation

$$\omega = \omega(\boldsymbol{k}, \boldsymbol{r}) \ , \tag{1}$$

where ω is the frequency, $\boldsymbol{r}$ is the position vector, and $\boldsymbol{k}$ is the wavenumber vector. We may think of $\omega(\boldsymbol{r}, \boldsymbol{k})$ as a surface in six-dimensional space. For a fixed value of $\boldsymbol{r}$, the subsurface is the wavenumber surface in a three-dimensional space $\boldsymbol{k}$. As we have seen in equations 84 and 87 in Chapter 5, the gradient to this three-dimensional surface is the group velocity

$$\nabla_k \omega = \frac{\partial \omega}{\partial \boldsymbol{k}} = \boldsymbol{v_g} \ . \tag{2}$$

For a fixed value of $\boldsymbol{k}$, the subsurface is a surface in three-dimensional space $\boldsymbol{r}$. The gradient to this three-dimensional surface is

$$\nabla_r \omega = \frac{\partial \omega}{\partial \boldsymbol{r}} \ . \tag{3}$$

These two gradients play an interesting role in the Hamilton theory.

Example. Consider a two-dimensional vertically stratified medium:

$$\boldsymbol{r} = (x, z), \quad \boldsymbol{k} = (k_x, k_z), \quad \boldsymbol{v}(x, z) = \boldsymbol{v}(\boldsymbol{z}) \ . \tag{4}$$

By definition, the velocity depends only upon the depth coordinate z, and not on the horizontal coordinate x. The dispersion relation is

$$\omega(\boldsymbol{k}, \boldsymbol{r}) = v(z)\sqrt{k_x^2 + k_z^2} = v(z)\,k \ . \tag{5}$$

Thus, ω is a function of the form $\omega(k_x, k_z, z)$. The group velocity vector is

$$\boldsymbol{v_g} = \left(\frac{\partial \omega}{\partial k_x}, \frac{\partial \omega}{\partial k_z}\right) = \left(\frac{k_x v}{k}, \frac{k_z v}{k}\right) = (\alpha v(z),\ \gamma v(z)) \ , \tag{6}$$

where wavenumber k is $\sqrt{k_x^2 + k_z^2}$, and where $\alpha = \sin\theta, \gamma = \cos\theta$ are the direction cosines of the $\boldsymbol{k}$ vector.

As shown in Figure 1, θ is the angle that $\boldsymbol{k}$ makes with the z axis. Because the group velocity is independent of frequency ω, there is no dispersion. The other gradient vector is, by equation 5,

$$\frac{\partial \omega}{\partial \boldsymbol{r}} = \left(\frac{\partial \omega}{\partial x}, \frac{\partial \omega}{\partial z}\right) = (0,\ v'(z)k) \ . \tag{7}$$

This vector has a zero component in the x direction, so that variations can only take place in the z-direction, as we would expect for a vertically stratified medium.

Let us now develop the Hamilton wave-particle duality. We start our treatment with the dispersion equation $\omega(\boldsymbol{k}, \boldsymbol{r})$. The basic assumption we make for the medium is that frequency and wavenumber vary slowly. That is, we assume (1) that the frequency ω does not change greatly in one oscillation period, and (2) that the wavenumber vector $\boldsymbol{k}$ does not change much in magnitude and direction over a distance of one wavelength.

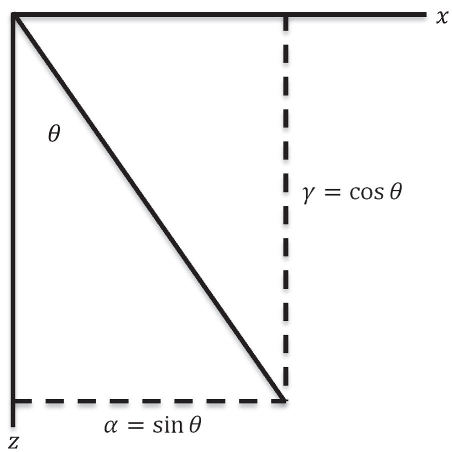


Figure 1. Two-dimensional propagation.

We recall that a plane wave in a homogeneous medium can be written in the form

$$\exp i\phi = \exp i(\omega t - \boldsymbol{k} \cdot \boldsymbol{r}) \ , \tag{8}$$

where ω and $\boldsymbol{k}$ are constant. The quantity

$$\phi = \omega t - \boldsymbol{k} \cdot \boldsymbol{r} \tag{9}$$

is called the phase. For such a plane wave, we see that

$$d\phi = \omega\, dt - \boldsymbol{k} \cdot d\boldsymbol{r} = \left(\omega - \boldsymbol{k} \cdot \frac{d\boldsymbol{r}}{dt}\right) dt \ , \tag{10}$$

so we can write the phase as

$$\phi = \int d\phi = \int \left(\omega - \boldsymbol{k} \cdot \frac{d\boldsymbol{r}}{dt}\right) dt \ . \tag{11}$$

We now make use of our basic assumption of slow variation of ω and $\boldsymbol{k}$. That is, ω and $\boldsymbol{k}$ are approximately constant with respect to period and wavelength, respectively. Thus, we assume that, for our slowly varying medium, the same equation holds, namely, the equation

$$\phi = \int_{t_0}^{t_1} \left[\omega(\boldsymbol{k}, \boldsymbol{r}) - \boldsymbol{k} \cdot \frac{d\boldsymbol{r}}{dt} \right] dt \ . \tag{12}$$

In physical terms, the quantity within the parentheses in equation 12 is called the *Hamiltonian* and the phase ϕ is called the action. According to *Hamilton's principle of least action*, the required solution for the ray path is found by minimizing the action; i.e., by minimizing the above integral. The magnitude of the integral depends upon the mathematical function chosen for the ray path $\boldsymbol{r}(t)$ as a function of t. In order to examine the relationship between the action (i.e., phase) ϕ and the function $\boldsymbol{r}(t)$, it is convenient to calculate the change in ϕ for the transition from some arbitrary function $\boldsymbol{r}(t)$ to another infinitely close, but also arbitrary, function $\boldsymbol{r_1}(t)$.

Figure 2 shows two such conceivable paths, where time t is plotted along the abscissa and $\boldsymbol{r}$ is schematically plotted on the ordinate. We assume that all such paths pass through the same points $\boldsymbol{r_0} = \boldsymbol{r}(t)$ and $\boldsymbol{r_1} = \boldsymbol{r}(t)$ at the initial time t_0 and final time t_1, respectively. The vertical distance between two such paths at some instant of time is called the variation of $\boldsymbol{r}$ and denoted by $\delta\boldsymbol{r}$. At the end points t_0 and t_1, of course $\delta\boldsymbol{r} = \boldsymbol{0}$ because all paths coincide at these points by assumption. The reason that the symbol δ is used is that we want to make clear the difference between the variation δ and the differential d. The differential is taken for the same path at various instants of time, whereas the variation is taken for the same instant of time between different paths.

The variation in the action (phase) is given by

$$\delta\phi = \delta \int_{t_0}^{t_1} \left[\omega(\boldsymbol{k}, \boldsymbol{r}) - \boldsymbol{k} \cdot \frac{d\boldsymbol{r}}{dt} \right] dt \ , \tag{13}$$

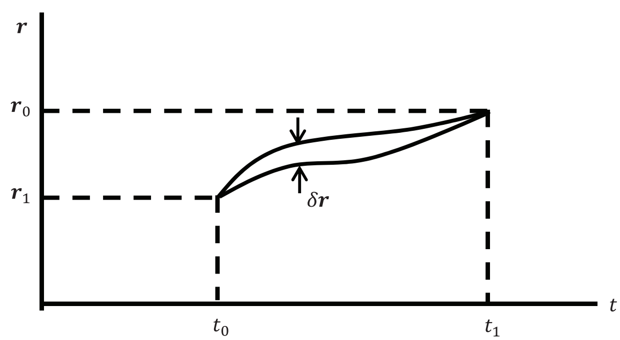


Figure 2. Schematic illustration of the variation $\delta\boldsymbol{r}$.

which is

$$\delta\phi = \delta \int_{t_0}^{t_1} \left[\frac{\partial \omega}{\partial \boldsymbol{k}} \cdot \delta \boldsymbol{k} + \frac{\partial \omega}{\partial \boldsymbol{r}} \cdot \delta \boldsymbol{r} - \boldsymbol{k} \cdot \delta \left(\frac{d\boldsymbol{r}}{dt} \right) \right] dt \ . \tag{14}$$

In equation 14, the two partial derivatives of ω are gradients; i.e.,

$$\frac{\partial \omega}{\partial \boldsymbol{k}} = \left(\frac{\partial \omega}{\partial k_x}, \frac{\partial \omega}{\partial k_y}, \frac{\partial \omega}{\partial k_z} \right) = \nabla_k \omega = \text{gradient of } \omega \text{ with respect to } \boldsymbol{k} \ , \tag{15}$$

$$\frac{\partial \omega}{\partial \boldsymbol{r}} = \left(\frac{\partial \omega}{\partial x}, \frac{\partial \omega}{\partial y}, \frac{\partial \omega}{\partial z} \right) = \nabla_r \omega = \text{gradient of } \omega \text{ with respect to } \boldsymbol{r} \ . \tag{16}$$

The last term inside the parentheses in equation 13 can be integrated by parts. Doing so, we obtain

$$\int_{t_0}^{t_1} \left[-\boldsymbol{k} \cdot \delta \frac{d\boldsymbol{r}}{dt} \right] dt = [-\boldsymbol{k} \cdot \delta \boldsymbol{r}]_{t_0}^{t_1} + \int_{t_0}^{t_1} \left[\frac{d\boldsymbol{k}}{dt} \cdot \delta \boldsymbol{r} \right] dt \ . \tag{17}$$

Because all curves $\boldsymbol{r}(t)$ pass through the same endpoints (as previously stated), the integrated part becomes zero. Thus, the variation in the phase is

$$\delta\phi = \int_{t_0}^{t_1} \left[\delta \boldsymbol{k} \cdot \left(-\frac{d\boldsymbol{r}}{dt} + \frac{\partial \omega}{\partial \boldsymbol{k}} \right) + \delta \boldsymbol{r} \cdot \left(\frac{d\boldsymbol{k}}{dt} + \frac{\partial \omega}{\partial \boldsymbol{r}} \right) \right] dt \ . \tag{18}$$

The independent variables are now $\boldsymbol{k}$ and $\boldsymbol{r}$. The variations $\delta\boldsymbol{k}$ and $\delta\boldsymbol{r}$ are completely arbitrary. Thus, for $\delta\phi$ to be zero, each of the following two equations must be satisfied:

$$\frac{d\boldsymbol{r}}{dt} = \frac{\partial \omega}{\partial \boldsymbol{k}} \quad \text{and} \quad \frac{d\boldsymbol{k}}{dt} = -\frac{\partial \omega}{\partial \boldsymbol{r}} \ . \tag{19}$$

These two equations are called *Hamilton's equations*.

Now, we will show that the frequency ω stays constant along a ray path. The rate of change of frequency $\omega(\boldsymbol{k}, \boldsymbol{r})$ with time for arbitrary rates of change of $\boldsymbol{k}$, $\boldsymbol{r}$ is the total derivative

$$\frac{d\omega}{dt} = \frac{\partial \omega}{\partial \boldsymbol{k}} \cdot \frac{\partial \boldsymbol{k}}{\partial t} + \frac{\partial \omega}{\partial \boldsymbol{r}} \cdot \frac{\partial \boldsymbol{r}}{\partial t} \ . \tag{20}$$

Substituting Hamilton's equations 19 into the right side of equation 20, we obtain zero. Thus, for the rates of changes along the ray, as given by

Hamilton's equations, the derivative $d\omega/dt$ is zero. Hence, frequency ω remains constant along a ray.

Let us now state Hamilton's duality between waves and particles. Let the coordinate vector $\boldsymbol{r}$ represent the coordinates of both the wave and the particle. Then, the wavenumber vector $\boldsymbol{k}$ of the wave corresponds to the momentum vector of the particle. The frequency ω of the wave corresponds to the energy of the particle. As we have said, the group velocity

$$\frac{d\boldsymbol{r}}{dt} = \frac{\partial \omega}{\partial \boldsymbol{k}} \ , \tag{21}$$

with which the wave packet travels, corresponds to the velocity with which the particle travels along the ray (see equation 19). Although the wave particle duality is established by Sir William Hamilton in 1825, the physical significance of this duality is not understood until 1924, when Prince Louis de Broglie suggests that an electron has a dual character (that is, an electron is a particle with laws of motion that are wave-like in character). This wave particle dualism for the electron matches the wave particle dualism for the photon as developed by Arthur Compton in 1923.

Ray tracing

If we solve Hamilton's equation, we can trace the path of the ray. This procedure is called *ray tracing*. Let us give some examples.

Example 1. In a constant velocity medium, we have

$$\omega(\boldsymbol{k}, \boldsymbol{r}) = kv = \sqrt{k_x^2 + k_z^2}\, v \ . \tag{22}$$

The component form of Hamilton's equations give

$$\frac{dx}{dt} = \frac{\partial \omega}{\partial k_x} = \frac{k_x v}{k} \quad \text{and} \quad \frac{dz}{dt} = \frac{\partial \omega}{\partial k_z} = \frac{k_z v}{k} \tag{23}$$

and

$$\frac{dk_x}{dt} = -\frac{\partial \omega}{\partial x} = 0 \quad \text{and} \quad \frac{dk_z}{dt} = \frac{\partial \omega}{\partial z} = 0 \ . \tag{24}$$

Equations 24 say that k_x and k_z are constant. Thus,

$$k = \sqrt{k_x^2 + k_z^2} = \text{constant} \ .$$

Integrating equations 23, we obtain the *equations of the ray*

$$x = \frac{k_x}{k} vt + x_0 \quad \text{and} \quad z = \frac{k_z}{k} vt + z_0 \ , \tag{25}$$

where x_0 and z_0 are the values of x and z, respectively, at $t = 0$.

Thus, the ray is in the direction of the constant propagation vector (k_x, k_z). Because k is the magnitude of the propagation vector, we see that we have a plane wave moving with velocity v.

Example 2. Let us now consider a stratified medium with the dispersion relation

$$\omega(\boldsymbol{k}, \boldsymbol{r}) = kv(z) = \sqrt{k_x^2 + k_z^2}\, v(z) \ . \tag{26}$$

In this case, Hamilton's equations are the four equations

$$\frac{dx}{dt} = \frac{\partial \omega}{\partial k_x} = \frac{k_x v(z)}{k} \quad \text{and} \quad \frac{dz}{dt} = \frac{\partial \omega}{\partial k_z} = \frac{k_z v(z)}{k} \tag{27}$$

and

$$\frac{dk_x}{dt} = -\frac{\partial \omega}{\partial x} = 0 \quad \text{and} \quad \frac{dk_z}{dt} = -\frac{\partial \omega}{\partial z} = kv'^{(z)} \ . \tag{28}$$

Dividing dx/dt by dz/dt in equation 27, we obtain

$$\frac{dx}{dz} = \frac{\frac{dx}{dt}}{\frac{dz}{dt}} = \frac{\frac{\partial \omega}{\partial k_x}}{\frac{\partial \omega}{\partial k_z}} = \frac{k_x}{k_z} \ . \tag{29}$$

Thus, the rays are in the direction of the wavenumber vector $\boldsymbol{k}$. The first equation in 28 gives $k_x =$ constant. We now use our previous general result that frequency ω is constant along a ray. Thus, we see that $k_x/\omega =$ constant along a ray in a stratified medium. If we call this constant p_x, then we use equation 26 and have

$$p_x = \frac{k_x}{\omega} = \frac{k_x}{kv(z)} = \text{constant} \tag{30}$$

along a ray. Referring to Figure 3, we next define $\theta(z)$ as the angle between the ray and the vertical. That is, $\sin\theta = k_x/k$ and $\cos\theta = k_z/k$.

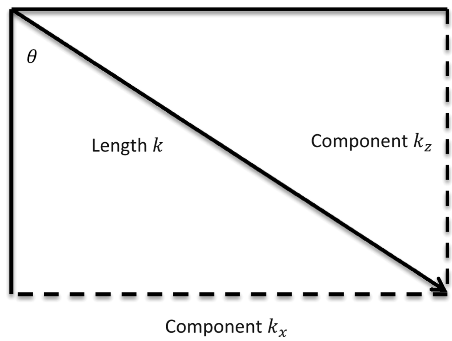


Figure 3. Definition of the angle $\boldsymbol{\theta}$.

Thus, equation 30 is seen to be Snell's law

$$p_x = \frac{\sin \theta}{v(z)} = \text{constant} \ . \tag{31}$$

Using Snell's law, as shown in equation 31, we have

$$\frac{k_x}{k} = \sin \theta(z) = p_x v(z) \quad \text{and} \quad \frac{k_z}{k} = \cos \theta(z) = \sqrt{1 - p_x^2\, v^2(z)} \ . \tag{32}$$

We will now integrate the equation

$$\frac{dx}{dz} = \frac{k_x}{k_z} = \frac{p_x v(z)}{\sqrt{1 - p_x^2\, v^2(z)}} \ . \tag{33}$$

Doing so, we obtain the so-called horizontal distance equation

$$x = \int_0^z \frac{p_x v(z) dz}{\sqrt{1 - p_x^2\, v^2(z)}} \ . \tag{34}$$

Also, we can integrate the second Hamilton equation shown in equation 27,

$$\frac{dz}{dt} = \frac{k_z v(z)}{k} = \sqrt{1 - p_x^2\, v^2(z)}\, v(z) \ , \tag{35}$$

to obtain the so-called time equation

$$t = \int_0^z \frac{dz}{v(z)\sqrt{1 - p_x^2\, v^2(z)}} \ . \tag{36}$$

Time equation 36 together with horizontal distance equation 34 compose the so-called time–distance relationship for the ray. That is, equations 36 and 34 give the so-called *time–distance curve* as a function of the parameter p_x of waves originating at $z = 0$ and traveling to depth $z =$ constant.

As displayed in Figure 4, each point on the time–distance curve is determined by a particular value of p_x. Because each value of p_x designates one ray, the time–distance curve summarizes information for all of the rays reaching depth z.

We can write the horizontal distance equation in another form. The first of the Hamilton equations 27 is

$$\frac{dx}{dt} = \frac{k_x v(z)}{k} = v(z) \sin \theta(z) \quad \text{(along a ray)} . \tag{37}$$

However, Snell's equation is $\sin \theta(z) = p_x v(z)$. Thus, equation 37 is

$$\frac{dx}{dt} = p_x v^2(z) \quad \text{(along a ray)} , \tag{38}$$

so the required *alternative form of the horizontal distance equation* is

$$x = \int_0^t p_x v^2(z) dt = p_x \int_0^t v^2(z) dt . \tag{39}$$

Remember that Hamilton's equations 27 and 28 apply for paths along a ray. A perfect example is equation 38 above.

On the other hand, the time–distance curve applies to all rays (each characterized by a value of p_x) which extend to a given depth $z =$ constant. Now, we want to show that

$$\frac{dx}{dt} = \frac{1}{p_x} \quad \text{(along the time–distance curve)} . \tag{40}$$

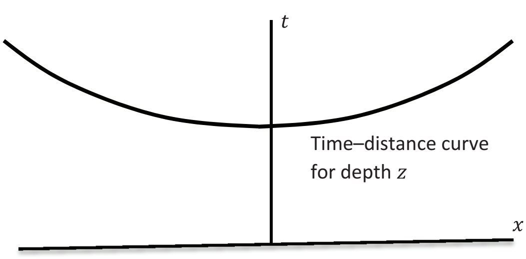


Figure 4. Time–distance curve for a given value of vertical coordinate z.

The geometric argument, as shown in Figure 5, can be used. From the diagram, we see that

$$\frac{dx}{v\,dt} = \frac{1}{\sin\theta} \ . \tag{41}$$

Using Snell's law $p_x = \sin\theta/v$, we obtain the required result, equation 40. Thus, the slope of the time–distance curve is equal to Snell's parameter p_x; that is,

$$\frac{dt}{dx} = p_x \text{ (along the time–distance curve)} \ . \tag{42}$$

The significance of the parameter p_x is realized by taking $z = 0$ in Snell's equation,

$$p_x = \frac{\sin\theta(0)}{v(0)} \ . \tag{43}$$

Equation 43 shows that p_x is proportional to the sine of the incidence angle $\theta(0)$ of the ray at the surface $z = 0$.

In the case of constant velocity, i.e., $v(z) = V = constant$, the time–distance curve is the hyperbola

$$V^2t^2 - x^2 = z^2 = \text{constant} \ . \tag{44}$$

In the case of a stratified medium with velocity function $v(z)$, generally the shape of the time-distance curve will resemble a hyperbola (see Figure 6). Suppose a spike at time $t = 0$ propagates outward from a point source at $z = 0$. At some depth (or, alternatively, height) $z = constant$, we measure the arrival time t of the spike as a function of the horizontal coordinate x. The result is the time–distance curve, such as the one shown in Figure 5. Suppose we now fit the time–distance curve to the hyperbola

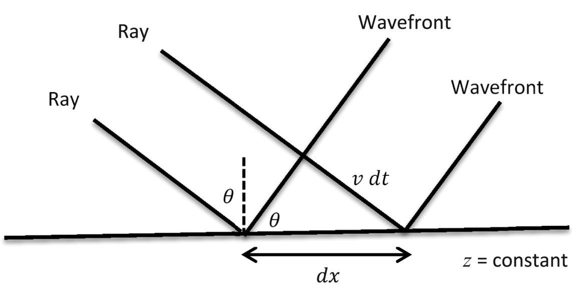


Figure 5. The horizontal distance between two adjacent rays is ***dx***.

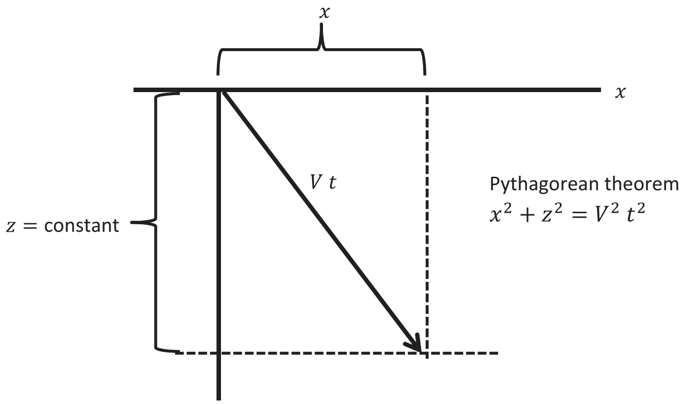


Figure 6. Geometry for determining the time–distance hyperbola in a medium of constant velocity $\boldsymbol{v}$.

equation 44. The constant velocity V is determined so as to yield the best fit. The question is—how does the fitted constant velocity V relate to the variable velocity function $v(z)$?

If we differentiate equation 44 for the hyperbola, we obtain

$$2V^2 t\,dt - 2x\,dx = 0 \;, \tag{45}$$

so

$$\frac{dt}{dx} = \frac{x}{tV^2} \quad \text{(on the hyperbola)} \;. \tag{46}$$

On the time–distance curve, however, we have $dt/dx = p_x$. Thus, we set the two derivatives equal to obtain

$$p_x = \frac{x}{tV^2} \;. \tag{47}$$

Thus, the required value of V^2 is equal to $x/p_x t$. If we solve equation 47 for V and make use of equation 39, we obtain

$$V^2 = \frac{x}{p_x t} = \frac{1}{t}\int_0^t v^2(z)dt \;. \tag{48}$$

Equation 48 says that V^2 is the mean-square value of $v(z)$ along the ray path. Thus, if we fit the time–distance curve by a hyperbola, the velocity V

that we obtain is the root mean square (RMS) velocity given by

$$V = V_{\mathrm{RMS}} = \sqrt{\frac{1}{t}\int_0^t v^2(z)dt} \ . \tag{49}$$

In words, we can say that, if we need an average velocity to characterize the stratified medium to some depth z, an excellent choice is the root mean square velocity.

Eikonal equation

In the general case for a spatially varying velocity function $v(x, y, z) = v(\boldsymbol{r})$, we can always solve Hamilton's equations numerically or graphically. The procedure is analogous to that employed in the mechanics of a particle moving in a three-dimensional potential field, and where acceleration is a function of position. In order to gain insight, let us recast Hamilton's equations in a slightly different form. We assume an isotropic medium. Again we start from the first principles, and proceed in an intuitive way. The seismic traveltime field $t(\boldsymbol{r})$ can be defined as the value of the traveltime from a convenient reference wave surface S_0 to an arbitrary point P with position vector $\boldsymbol{r}$. It is implicitly understood that P can be reached by a ray from S_0.

See Figures 7 and 8. Denote the point on S_0 at the foot of the ray by P_0 with position vector $\boldsymbol{r}_0$. Then, the seismic traveltime field is

$$t(\boldsymbol{r}) = \int_{\boldsymbol{r}_0}^{\boldsymbol{r}} p\, ds \ , \tag{50}$$

where $p(\boldsymbol{r}) = 1/v(\boldsymbol{r})$ is the slowness and ds is an increment of path length along the given ray. It is understood, of course, that the path of integration is along the given ray.

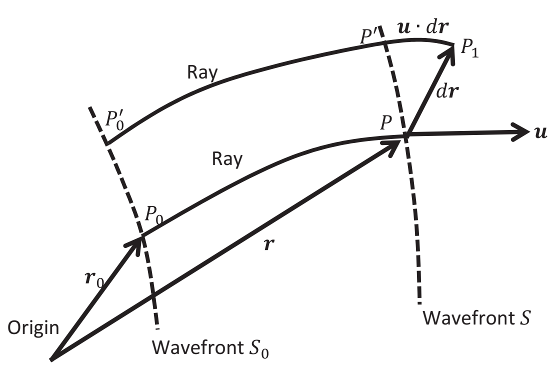


Figure 7. Two wavefronts and two rays.

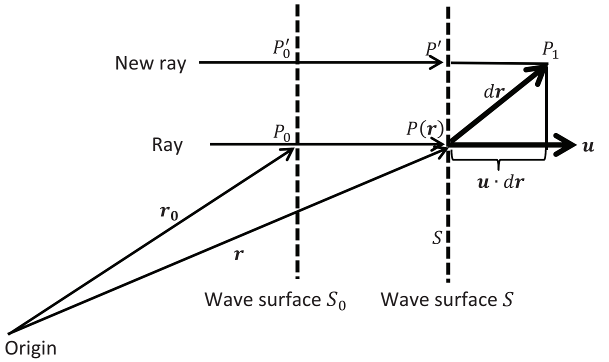


Figure 8. An oversimplified diagram showing the essence of Figure 7. In an isotropic medium, the rays and wave surfaces are orthogonal. For infinitesimal displacements, the rays are infinitesimally parallel and the wave surfaces also are infinitesimally parallel.

Suppose that a seismic event at time zero is represented by the reference wavefront S_0. Then, at another time, $t =$ constant, the seismic event will be represented by the wavefront S. Wavefront S consists of all points that can be reached in the time interval $t =$ constant by rays starting on S_0. The equation defining S is the locus of positions $\boldsymbol{r}$ satisfying

$$t(\boldsymbol{r}) = t = \text{constant} \; . \tag{51}$$

Suppose now that we consider another point P_1, which is close to P. The distance from P to P_1 is $d\boldsymbol{r}$. What is the corresponding change in the traveltime as $\boldsymbol{r}$ moves from P to P_1? First, let us construct the ray that goes from the reference surface S_0 to P_1. We call this ray $P_0'P'P_1'$, where the intermediate point lies on the surface S. The distances $P_0'P'$ and P_0P are approximately the same, as they are distances between the two wave surfaces S and S_0. Thus, the change in traveltime is due primarily to the path length $|P'P_1|$. The traveltime increment is

$$dt = p|P'P_1| = p\,\boldsymbol{u} \cdot d\boldsymbol{r} \; , \tag{52}$$

where $\boldsymbol{u}$ is the *unit tangent vector* at P tangent to the original ray P_0P. From differential calculus, we know that, in general, dt is

$$dt = \frac{\partial t}{\partial x}dx + \frac{\partial t}{\partial y}dy + \frac{\partial t}{\partial z}dz = \nabla t \cdot d\boldsymbol{r} \; . \tag{53}$$

It follows, therefore, that the gradient of the seismic traveltime function is

$$\nabla t = p\boldsymbol{u} \tag{54}$$

or

$$\left(\frac{\partial t}{\partial x}, \frac{\partial t}{\partial y}, \frac{\partial t}{\partial z}\right) = (p\alpha, p\beta, p\gamma) \ . \tag{55}$$

The ray cuts the wavefront at P. This equation says that, at P, the gradient of the wavefront is equal to the slowness p times the tangent to the ray. Therefore, the gradient and the tangent both point in the same direction. Because the gradient is orthogonal to the wavefront, and the tangent is along the ray, it follows that the ray is orthogonal to the wavefront. The absolute value of equation 54 gives a result known as the *eikonal equation*

$$|\nabla t| = p \ . \tag{56}$$

Because $|\boldsymbol{u}| = \mathbf{1}$, the eikonal equation, in square form, is

$$\nabla t \cdot \nabla t = p^2$$

or

$$\left(\frac{\partial t}{\partial x}\right)^2 + \left(\frac{\partial t}{\partial y}\right)^2 + \left(\frac{\partial t}{\partial z}\right)^2 = \frac{1}{v^2(x, y, z)} \ . \tag{57}$$

Example. Let us consider a two-dimensional stratified medium, with horizontal coordinate x and depth z, and with velocity function $v(z)$. The surface of the earth is the line $z = 0$. The standard time–distance curve which depicts a seismic event received on the surface of the earth is, in the present notation,

$$t(x, z = 0) = t(x, 0) \ . \tag{58}$$

Previously, simply $t(x)$ is used for this function. Let the emergence angle of the ray be θ. Then, the unit tangent vector to the ray is

$$u(\alpha, \beta) = (\sin\theta, \ \cos\theta) \ , \tag{59}$$

where $\alpha = \sin\theta$ and $\beta = \cos\theta$ are the direction cosines of $\boldsymbol{u}$. The eikonal equation is (because $p = 1/v$)

$$\left(\frac{\partial t}{\partial x}, \frac{\partial t}{\partial z}\right) = \left(\frac{\sin\theta}{v}, \frac{\cos\theta}{v}\right) . \tag{60}$$

The first component,

$$\frac{\partial t}{\partial x} = \frac{\sin\theta(z)}{v(z)} = p_x , \tag{61}$$

states that the derivative of the time−distance curve is equal to the Snell parameter p_x.

Ray equations

The seismic ray at any given point follows the direction of the gradient of the traveltime field $t(\boldsymbol{r})$. As before, let $\boldsymbol{u}$ be the unit vector along the ray. The ray in general will follow a curved path, and $\boldsymbol{u}$ will be the tangent to this curved raypath. We now want to derive an equation that will tell us how $\boldsymbol{u}$ changes along the curved raypath. We write the vector $\boldsymbol{u}$ as

$$\boldsymbol{u} = u(\alpha, \beta, \gamma), \quad \text{where } \alpha^2 + \beta^2 + \gamma^2 = 1 . \tag{62}$$

In this section, the position vector $\boldsymbol{r}$ always represents a point on a specific raypath, and not any arbitrary point in space. As time increases, $\boldsymbol{r}$ traces out the particular raypath in question. Now, let us find how a general function of position $f(\boldsymbol{r})$ will change along the raypath curve. For a general displacement $d\boldsymbol{r}$, calculus dictates

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz = \nabla f \cdot d\boldsymbol{r} . \tag{63}$$

Because, by assumption, $d\boldsymbol{r}$ is along the ray-path curve, it has length ds. Thus, we can write

$$d\boldsymbol{r} = \boldsymbol{u}ds , \tag{64}$$

where $\boldsymbol{u}$ is the unit tangent vector to the raypath curve. Hence, we obtain

$$df = (\nabla f \cdot \boldsymbol{u})\, ds , \tag{65}$$

so the directional derivative of f along the raypath curve is

$$\frac{df}{ds} = (\nabla f \cdot \boldsymbol{u}) . \tag{66}$$

Because the curve is a raypath curve, eikonal equation 54 ($\nabla t = p\boldsymbol{u}$) holds, so

$$\frac{df}{ds} = \nabla f \cdot \frac{\nabla t}{p} = \frac{\nabla f \cdot \nabla t}{\boldsymbol{p}} \ . \tag{67}$$

We now make a special choice for f; namely, $f = \alpha / v = p\alpha$. From equation 55, we have

$$f = p\alpha = \frac{\partial t}{\partial x} \ . \tag{68}$$

Thus, equation 67 gives

$$\frac{d}{ds}\,(p\alpha) = \frac{d}{ds}\frac{\partial t}{\partial x} = \frac{1}{p}(\nabla t \cdot \nabla)\frac{\partial t}{\partial x} \ . \tag{69}$$

Now we use algebra. The dot product in the expression on the right can be written out in full, and the terms collected to give

$$\frac{d}{ds}\,(p\alpha) = \frac{1}{2p}\frac{\partial}{\partial x}\left[\left(\frac{\partial t}{\partial x}\right)^2 + \left(\frac{\partial t}{\partial y}\right)^2 + \left(\frac{\partial t}{\partial z}\right)^2\right] . \tag{70}$$

The square form of eikonal equation 57 says that the expression in brackets is p^2. Thus,

$$\frac{d}{ds}\,(p\alpha) = \frac{1}{2p}\frac{\partial (p^2)}{\partial x} = \frac{\partial p}{\partial x} \ . \tag{71}$$

Similar equations hold for the y and z components. Thus, we are led to the vector equation

$$\frac{d}{ds}\,(p\boldsymbol{u}) = \frac{\partial p}{\partial \boldsymbol{r}} \ . \tag{72}$$

Equation 72, together with equation 64, which we write as

$$\frac{d\boldsymbol{r}}{ds} = \boldsymbol{u} \ , \tag{73}$$

are called the *ray equations*. In summary, the two ray equations 73 and 72 are

$$\frac{d\boldsymbol{r}}{ds} = \boldsymbol{u} \quad \text{and} \quad \frac{d(p\boldsymbol{u})}{ds} = \frac{\partial p}{\partial \boldsymbol{r}} \ . \tag{74}$$

Reduction of Hamilton's equations to the ray equations

The frequency ω, wavenumber k, and velocity v are related by

$$\omega = kv \ . \tag{75}$$

An increment ds of distance s along a raypath and an increment ds of traveltime t are related by

$$ds = v\,dt \ . \tag{76}$$

Thus, Hamilton's equation

$$\frac{d\boldsymbol{r}}{dt} = \frac{\partial \omega}{\partial \boldsymbol{k}} \tag{77}$$

becomes

$$v\,\frac{d\boldsymbol{r}}{ds} = \frac{\partial (kv)}{\partial \boldsymbol{k}} \tag{78}$$

because

$$\frac{d\boldsymbol{r}}{dt} = \frac{d\boldsymbol{r}}{ds}\frac{ds}{dt} = v\,\frac{d\boldsymbol{r}}{ds} \ .$$

The velocity $v(x, y, z)$ is a function of x, y, z, whereas k is considered as the function

$$k = \sqrt{k_x^2 + k_y^2 + k_z^2} \ . \tag{79}$$

Thus, equation 78 becomes

$$v\,\frac{d\boldsymbol{r}}{ds} = v\,\frac{dk}{d\boldsymbol{k}} \ . \tag{80}$$

The gradient of k with respect to the variables k_x, k_y, k_z is the vector

$$\frac{dk}{d\boldsymbol{k}} = \left(\frac{\partial k}{\partial k_x}, \frac{\partial k}{\partial k_y}, \frac{\partial k}{\partial k_z}\right) = \left(\frac{k_x}{k}, \frac{k_y}{k}, \frac{k_z}{k}\right) \ . \tag{81}$$

We recognize this vector as the unit vector in the $\boldsymbol{k}$ direction. This unit vector, as we have seen, is the unit tangent vector to the raypath curve; that is,

$$\frac{dk}{d\boldsymbol{k}} = \frac{\boldsymbol{k}}{k} = \boldsymbol{u} = (\alpha, \beta, \gamma), \quad \text{where } \alpha^2 + \beta^2 + \gamma^2 = 1 \ . \tag{82}$$

Canceling v from each side of equation 80, we obtain

$$\frac{d\boldsymbol{r}}{ds} = \boldsymbol{u} \ . \tag{83}$$

This equation is the *first ray equation*. Thus, we have shown that the first Hamilton equation 77 is the same as the first ray equation 83.

Next, consider the second Hamilton equation

$$\frac{d\boldsymbol{k}}{dt} = -\frac{\partial\omega}{\partial\boldsymbol{r}} \ . \tag{84}$$

We can write this equation as

$$v\frac{d\boldsymbol{k}}{ds} = -\frac{\partial(kv)}{\partial\boldsymbol{r}} \ , \tag{85}$$

which is

$$v\frac{d\boldsymbol{k}}{ds} = -k\frac{\partial v}{\partial\boldsymbol{r}} \ . \tag{86}$$

Now, we use the equation

$$\boldsymbol{k} = k\boldsymbol{u} = \frac{\omega}{v}\boldsymbol{u} = \omega p\boldsymbol{u} \ , \tag{87}$$

where $p = 1/v$ is the slowness. The second Hamilton equation 84, with the aid of equation 86, becomes

$$p^{-1}\ \frac{d(\omega p\,\boldsymbol{u})}{ds} = -\omega p\frac{\partial(p^{-1})}{\partial\boldsymbol{r}} \ ,$$

which is

$$p^{-1}\,\omega\frac{d(p\,\boldsymbol{u})}{ds} = \omega p^{-1}\frac{\partial p}{\partial\boldsymbol{r}}$$

or

$$\frac{d(p\boldsymbol{u})}{ds} = \frac{\partial p}{\partial \boldsymbol{r}} \ . \tag{88}$$

This equation is the *second ray equation*. Thus, we have shown that the second Hamilton equation 84 is the same as the second ray equation 88.

Numerical ray tracing

Let us now consider the general case in which we have a spatially varying velocity function $v(x, y, z) = v(\boldsymbol{r})$. This velocity function represents a velocity field. For a fixed constant v, the equation $v(\boldsymbol{r}) = v$ specifies those positions $\boldsymbol{r}$ which have this fixed value of v. The locus of such positions makes up an *isovelocity surface*. The gradient

$$\nabla v(\boldsymbol{r}) = \left(\frac{\partial v}{\partial x}, \frac{\partial v}{\partial y}, \frac{\partial v}{\partial z}\right) \tag{89}$$

is normal to the isovelocity surface and points in the direction of the greatest increase in velocity. Similarly, the equation $p(\boldsymbol{r}) = p$ for a fixed value of slowness p specifies an *isoslowness surface*. The gradient

$$\nabla p(\boldsymbol{r}) = \left(\frac{\partial p}{\partial x}, \frac{\partial p}{\partial y}, \frac{\partial p}{\partial z}\right)$$

is normal to the isoslowness surface and points in the direction of greatest increase in slowness. The isovelocity and isoslowness surfaces coincide, and

$$\nabla v = -p^{-2}\, \nabla p \ , \tag{90}$$

so the respective gradients point in opposite directions, as we would expect.

Now that we have discussed the mathematics, let us talk about how it is applied. Like flying an airplane, expert boat handling takes both knowledge and practice. Any twin engine boat is capable of very precise control. If you put one engine in forward and the other in reverse, the boat will pivot on its axis. You can steer the boat without touching the wheel. Put one hand on each of the throttles, with both engines in forward. You control the boat by using only the throttles. You steer the boat to the left by decreasing the speed on the left throttle and increasing the speed on the right throttle. You steer the boat to the right by doing the opposite.

A seismic ray makes its way through the slowness field. As the wavefront progresses in time, the raypath is bent according to the slowness field. For example, suppose we have a stratified earth in which the slowness decreases with depth. A vertical raypath will not bend, as it is pulled equally in all lateral directions. However, a non-vertical ray will drag on its slow side, so it will steer around the curve away from the vertical and bend toward the horizontal. This is the case of a diving wave, whose raypath eventually curves enough to reach the earth's surface again. Certainly, the slowness field, together with the initial direction of the ray, determines the entire raypath. Except in special cases, we must determine such raypaths numerically.

Assume we know the slowness function $p(\boldsymbol{r})$ and the ray direction $\boldsymbol{u}_1$ at point $\boldsymbol{r}_1$. Considering Figure 9, we now want to give an algorithm for finding the ray direction $\boldsymbol{u}_2$ at point $\boldsymbol{r}_2$. We choose a small, but finite, change in path length Δs. Then, we use the first ray equation, which as we recall is

$$\frac{d\boldsymbol{r}}{ds} = \boldsymbol{u} \ ,$$

to compute the change $\Delta\boldsymbol{r} = \boldsymbol{r}_2 - \boldsymbol{r}_1$. The required approximation is

$$\Delta\boldsymbol{r} = \boldsymbol{u}_1 \Delta\boldsymbol{s}$$

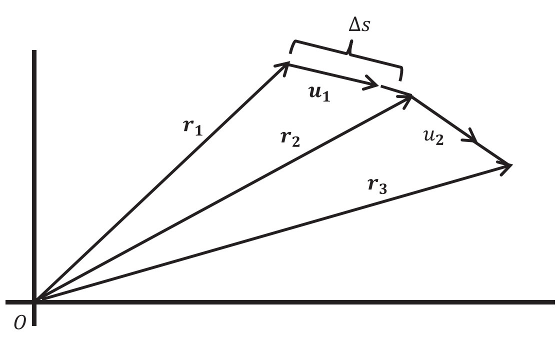
or

$$\boldsymbol{r}_2 = \boldsymbol{r}_1 + \Delta\boldsymbol{r} = \boldsymbol{r}_1 + \boldsymbol{u}_1 \Delta\boldsymbol{s} \ .$$

Thus, we have found the first desired quantity $\boldsymbol{r}_2$. Next, we use the second ray equation, which we recall is

$$\frac{d(p\boldsymbol{u})}{ds} = \nabla p \ ,$$

Figure 9. Graphical integration of the equation of a ray. The slowness is increasing in the direction indicated.



in the form

$$d(p\boldsymbol{u}) = \nabla p \, ds \ .$$

The required approximation is

$$\Delta(p\boldsymbol{u}) = \nabla p \, \Delta s$$

or

$$p(r_2)\boldsymbol{u}_2 - p(r_1)\boldsymbol{u}_1 = \nabla p \, \Delta s \ .$$

For accuracy, ∇p may be evaluated by differentiating the known function $p(\boldsymbol{r})$ midway between $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$. Thus, the desired $\boldsymbol{u}_2$ is given as

$$\boldsymbol{u}_2 = \frac{p(r_1)}{p(r_2)}\boldsymbol{u}_1 + \frac{\Delta s}{p(r_2)}\nabla p \ .$$

Note that the vector $\boldsymbol{u}_1$ is pulled in the direction of ∇p in forming $\boldsymbol{u}_2$. This is because $+\Delta p$ points in the direction of increasing slowness. That is, the ray drags on the slow side, and so it is bent in the direction of increasing slowness. The special case of no bending occurs when $\boldsymbol{u}_1$ and ∇p are parallel. As we have seen, a vertical wave in a stratified medium is an example of such a special case. Thus, we have found how to advance the wave along the ray by an incremental raypath distance. We can repeat the algorithm to advance the wave by any desired distance.

This page has been intentionally left blank

Chapter 7

Reflection, Refraction, and Diffraction

Explanation of terms

In seismic exploration, primarily we are concerned with traveling waves, and the three most important terms which describe wave propagation are *reflection*, *refraction*, and *diffraction*. All of these terms imply change in the direction of the traveling wave due to some interface or obstacle. When a wave is reflected from an interface, part of the energy is thrown back and does not pass through the interface. When a wave is refracted at an interface, part of the energy passes through the interface but the direction of propagation generally is changed. When a wave is diffracted by an obstacle or aperture, the wave passes around the obstacle or through the aperture by bending its direction of propagation and modifying its behavior.

The important thing to remember is that reflection and refraction are defined in terms of least-time raypaths, whereas diffraction is a phenomenon which does not have to satisfy the least-time criterion. As a result, we can formulate the law of reflection (angle of incidence equals angle of reflection) and the law of refraction (Snell's law). The phenomena of reflection and refraction form the basis of geometric raypath theory. In contrast, diffraction is not a part of geometric ray path theory, so instead we must determine the theoretical properties of diffraction by means of the wave equation. The Kirchhoff solution of the wave equation allows us to find approximations that represent diffracted events in simple cases. In even slightly more complex cases, no analytic solutions for diffraction are known, so we must rely on numerical solutions of the wave equation. The difficulty of such solutions makes us realize how fortunate we are that so many important properties of wave propagation can be explained in terms of the geometric ray theory of reflections and refractions. Despite the fact that diffraction is not governed by geometric raypath theory, we can still use raypaths to visualize point diffraction phenomena. The reason is that, in such cases, the diffraction occurs at a point at which the laws of reflection and refraction break down; once we are away from that point, however, we are in well-behaved media where geometric raypath theory is perfectly valid.

In the treatment of seismic wave propagation, it is convenient to confine attention to the lines along which the waves travel. Such lines are called rays, and the theory of wave propagation in these terms is called *seismic ray theory* or *geometrical seismology*. It is possible to obtain quantitative descriptions of many properties of seismic propagation in terms of geometric seismology alone, but such descriptions must be considered as only approximations to a true description, and hence limited in their application. There is a large class of phenomena that can be understood only in terms of the wave properties of seismic disturbances. These interference and diffraction phenomena form the subject matter of this chapter; they are general properties of waves. Thus, here we are discussing material which can be categorized as physical seismology, the study of the physical nature of seismic propagation as elastic waves.

The term interference is used in its most general sense to mean any effect due to the superposition of two waves at a given point in space. We can show that interference effects broadly defined account for reflection and refraction and also for diffraction, which is defined as the bending of waves around obstacles. The term interference, in its classic but more restricted sense, refers to variations of the total intensity of the seismic disturbance due to the superposition of two or more independent, interfering wave trains propagating in straight lines. Although there is no essential difference between interference and diffraction, the term diffraction is usually reserved for description of waves spreading around an obstacle or through an aperture. In such cases, the "interference" is due to the superposition of secondary wavelets originating at different points of the same wavefront. The obstacle or aperture acts as a spatial filter, selecting only certain portions of the original wavefront which will compose the diffracted wavefront.

Christiaan Huygens shows that the propagation of a wavefront can be attributed to the superposition of spherical secondary wavelets emitted from all parts of the wavefront. Furthermore, in general, anything that prevents the emission of secondary wavelets from one or more parts of the wavefront will alter the nature and direction of succeeding wavefronts. As we have seen, the term *diffraction* designates those interference effects caused by the presence of an aperture or obstacle in the path of a seismic wave. Such obstructions allow only certain portions of the incident wavefront to propagate by the emission of Huygens' secondary wavelets. One of the results is that the seismic waves deviate from straight-line propagation. Thus, diffraction often is said to involve the bending of waves around obstacles. Because diffraction involves the propagation of a single wavefront as affected by obstacles in its path, we may say that diffraction represents the interference of a wave with itself. Diffraction can be explained by the mutual interference of the Huygens' wavelets that are emitted from those portions of the wavefront not obstructed by the obstacle.

Imagine parallel seismic rays incident upon a narrow slit in the rock formation, passing through the slit, and illuminating a plane interface. If the plane interface is very close to the slit, a sharp image of the aperture is formed on the interface. As the interface is moved farther away from the slit, fringes of varying illumination begin to appear at the edges of the image. These fringes represent the diffraction pattern. The study of such patterns is important in the understanding of the seismic events we see on the record sections.

Huygens' principle

A wavefront is a surface over which a wave disturbance has a constant phase. As an illustration, consider a small portion of a spherical wavefront emanating from a monochromatic point source S in a homogeneous medium. Clearly, if the radius of the wavefront at a given time is r, at some later time t it will simply be $r + vt$, where v is the velocity of the wave. But suppose instead that the wave passes through a non-uniform sheet of material, so that the wavefront itself is distorted. How can we determine its new form? Or for that matter, what will it look like at some later time if it is allowed to continue unobstructed?

The major step toward the solution of this problem appears publically in 1690 in the book by Christiaan Huygens, *Traité de la Lumière*, which he had presented to the French Academy twelve years earlier. In it, he enunciates what has since become known as *Huygens' principle.* The principle states that *every point on a primary wavefront serves as the source of spherical secondary wavelets such that the primary wavefront at some later time is the envelope of these wavelets*. In more advanced discussions, the envelope prescription is abandoned, and instead the superposition of the wavelets is described in detail by means of the Kirchhoff integral. The Kirchhoff approach also gives rise to the inclination factor which is of considerable interest. This inclination factor explains why it is valid to neglect backradiation in the Huygens' construction so that the waves propagate only in the forward direction.

The search for new inventions often leads directly to mathematical discoveries. A famous example is the *Horologium Oscillatorium* of Christiaan Huygens (1673), where the search for better timepieces for the more accurate determination of longitude at sea leads not only to the pendulum clock but also to important geometrical theorems. Huygens' ability to bring the disciplines of mathematics, mechanics, and optics to bear on his interest in astronomy enables him to design, construct, and operate a telescope with which he discovers the fourth satellite of Saturn. Eminent as a physicist as well as an astronomer, Huygens establishes the wave theory of light. It is intriguing that many of the basic principles of optics are predicated on the wave theory of light and yet are completely independent of the exact nature of that wave. It is for this reason that Huygens' principle can serve to describe not only light waves but wave motion in other disciplines as well.

As we have seen, a wavefront is a surface over which the wave disturbance has a constant phase. Returning to Huygen's principle in *Traité de la Lumière* (1690): every point on a primary wavefront serves as the source of spherical secondary wavelets. These secondary wavelets advance with speed and frequency equal to that of the primary wave at each point in space. The primary wavefront at some later time is the envelope of these secondary wavelets. If the medium is homogeneous, the spherical secondary wavelets may be constructed with finite radii. On the other hand, if the medium is inhomogeneous, the wavelets will have infinitesimal radii, and the magnitudes of the radii will depend on the wave velocity of the medium at the respective centers of the wavelets.

Figure 1 shows a wavefront, as well as a number of spherical secondary wavelets, which during time *dt* have propagated out to a radius of *v dt*. The wave velocity *v* in an inhomogeneous medium depends upon the position of the center of the secondary wavelet. The envelope of all of these wavelets is the advanced primary wavefront.

The description of Huygens' principle, which we have just given, is a physical one. In the words of Huygens (1690): "The result is that around each particle there arises a wavelet of which this particle is center."

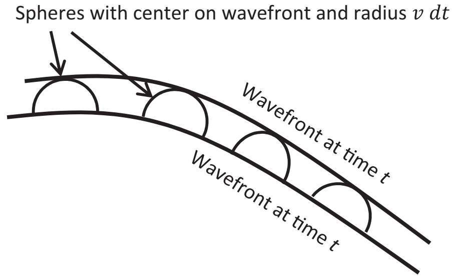


Figure 1. The propagation of a wavefront according to Huygens' principle of 1690. The advanced wavefront is the envelope of the spheres.

However, if we go back to Huygens' earlier work, *Horologium Oscillatorium* (1673), another description of Huygens' principle is possible. Interestingly, this earlier description fits in more readily with the systems approach of the electrical engineer and the time series analyst. It is well known that the appearance of *Geometria Indivisibilibus Continuorum* of Bonaventura Cavalieri (1635) stimulated many mathematicians of that time to study problems involving infinitesimals. Referring to Figure 2, consider the tangent problem, i.e., *the problem of finding a tangent to a given curve at a given point*, which at the time began to take a prominent place beside the ancient problems of finding volumes and centers of gravity. In this search, Huygens follows Euclid's method of geometrical reasoning, and he establishes himself as one of the great 17th century mathematicians. If we follow Huygens' reasoning, his famous principle also may be described as follows. *At every point on the wavefront, construct a sphere of radius v dt tangent to the wavefront. The locus of the centers of these tangent spheres is the advanced wavefront*.

When comparing Figures 1 and 2, we see that the two statements of Huygens' principle, that is, the principles of 1690 and 1673, are equivalent to the first order. Because the Huygens' principle of 1690 is treated in every physics book, we will spend more time with the principle of 1673. For simplicity, we will deal with a homogeneous medium so the spherical wavelets may be constructed with finite radii all of equal magnitude $v\Delta t$, where v is the constant wave velocity and Δt is the time increment between wavefronts. For pedagogical reasons, we make use of only two spatial dimensions, so we can use circles instead of spheres in our drawings. However, for physical correctness, readers should make the transition in their minds to three spatial dimensions. Let us now look at the spherical secondary wavelet, which, again for pedagogical reasons, we will call Huygens' circle. For propagation in the $+z$-direction, we only require a semicircle. Such a Huygens' semicircle is shown in Figure 3.

This semicircle (of radius $v\Delta t$) is the response at time Δt of an impulsive point source at time 0. In engineering terms, the Huygens' semicircle

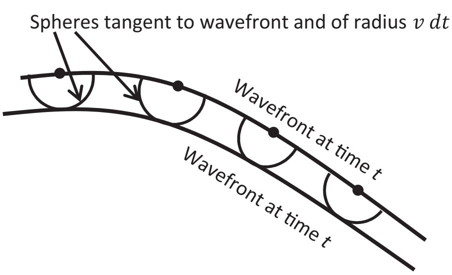


Figure 2. The propagation of a wavefront according to Huygens' principle of 1673. The advanced wavefront is the locus of the centers of the tangent spheres.

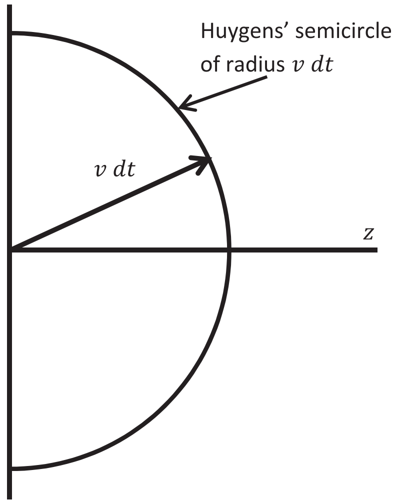


Figure 3. Huygens' semicircle for propagation in the z-direction.

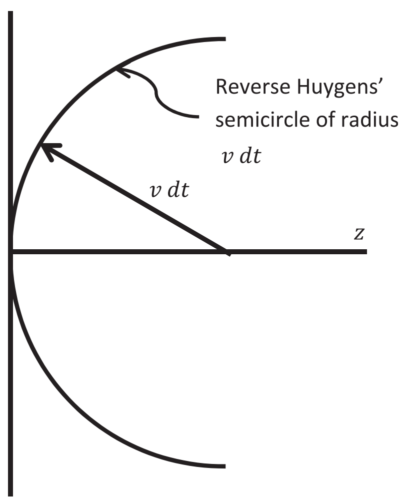


Figure 4. Reverse Huygens' semicircle for propagation in the z-direction.

is the impulse response (i.e., Green's function) for right-traveling waves. Also in engineering terms, specifically the relation that the output is the convolution of the input with the impulse-response (i.e., Green's function), we can state: the wavefront at time $t + \Delta t$ is the convolution of the wavefront at time t with the Huygens' semicircle. As is well known in engineering, the operation of convolution with a given function is equivalent to the operation of correlation with the reverse of the given function. The reverse of the Huygens' semicircle is shown in Figure 4.

When two curves cross, they contact each other only at one point, namely, the point of intersection. When two curves are tangent, they contact each other at one point, namely, the point of tangency, but the neighboring points are also very close together. For this reason, a point of tangency represents a higher-order contact between two curves than a straightforward point of intersection. The practical application of this geometrical property is that, when we correlate two curves, there is a significant output only when the two curves are tangent and not when they are merely crossing each other. When we correlate the input wavefront with the reverse Huygens' semicircle, we hold the position of the wavefront fixed, and let the reverse Huygens' semicircle sweep out the entire plane. However, significant output will occur only when the wavefront and the reverse Huygens' semicircle are tangent. Such places of tangency are, indeed, precisely as the tangency points in Huygens' principle of 1673, as we have already illustrated in Figure 2. The outputs occur at the centers of the reverse Huygens' semicircles. Thus, the locus of these centers make up the output wavefront. In engineering terms, we can state the *Huygens' principle of 1673* as follows. *The output wavefront is the convolution of the input wavefront with the Huygens' semicircle.* Equivalently, we may state that the output wavefront is the correlation of the input wavefront with the reverse Huygens' semicircle.

Up to this point, we have dealt with only the space aspects of Huygens' principle. In order to form a basis for digital space–time processing, we must introduce the time aspects. The Huygens' circle represents the spatial manifestation of an event. The spike-like event occurs at time zero and spreads out in expanding concentric Huygens' circles as time increases.

Let v be the (constant) wave velocity. As apparent from Figure 5, the equation for the Huygens' circle at time t is

$$x^2 + z^2 = (vt)^2 \ . \tag{1}$$

Looking to Figure 6, consider a space–time manifestation on the fixed plane $z =$ constant. Rearranging equation 1, we obtain

$$(vt)^2 - x^2 = z^2 \ , \tag{2}$$

which is a hyperbola for a fixed value of z.

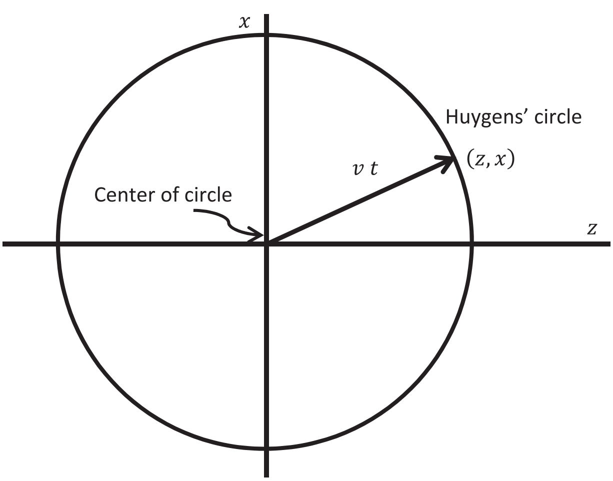


Figure 5. Huygens' circle for a fixed value of time t.

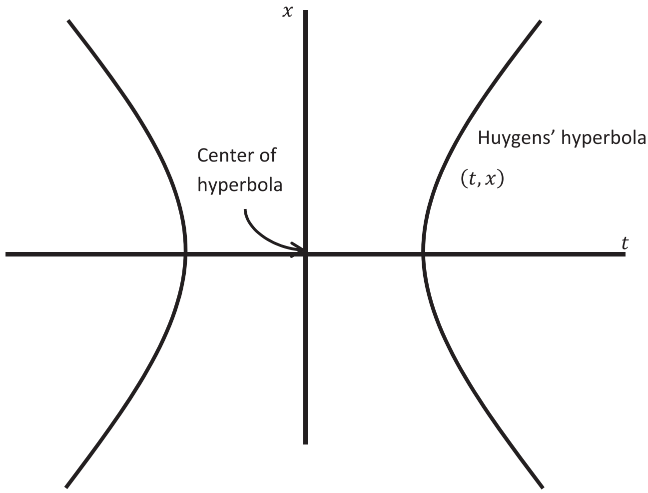


Figure 6. Huygens' hyperbola for a fixed value of depth z.

For propagation in the $+z$ direction, we want the right half of the circle (i.e., the Huygens' semicircle). For propagation in the $+t$ direction, we want the right branch of the hyperbola (i.e., the Huygens' semi-hyperbola). Suppose that we have a space–time event measured on plane z. To find the corresponding space–time event on plane $z + \Delta z$, we use the following counterpart of Huygens' principle: *The output space–time event is the convolution of the input space–time event with the Huygens' semihyperbola*. This Huygens' semi-hyperbola is the one for vertical propagation distance $+\Delta z$.

See Figure 7, which illustrates Huygens' space–time principle. The input space–time event is the fixed semi-hyperbola corresponding to vertical propagation distance z. The convolution is carried out by reversing the Huygens' semi-hyperbola for vertical distance Δz and sliding it around the entire plane. There is significant output only when the two semi-hyperbolas are tangent, and the output occurs at the center point of the reversed sliding hyperbola. We see that the output is the semi-hyperbola corresponding to vertical propagation distance $z + \Delta z$.

Now, we want to show that integration of the eikonal equation is equivalent to application of Huygens' principle. Suppose the wavefront at time t is S. The wavefront S is the locus formed by the tips of the position vectors $\boldsymbol{r}$ for which $t(\boldsymbol{r}) = t$. What is the wavefront at $t + \Delta t$? This new wavefront can be obtained from the old one by advancing each point in the direction of the normal to S, i.e., in the direction of the ray, by a distance Δs. Over this distance, $t(\boldsymbol{r})$ will change by the amount

$$\Delta t(\boldsymbol{r}) = |\nabla t|\ \Delta s = p\Delta s\ . \tag{3}$$

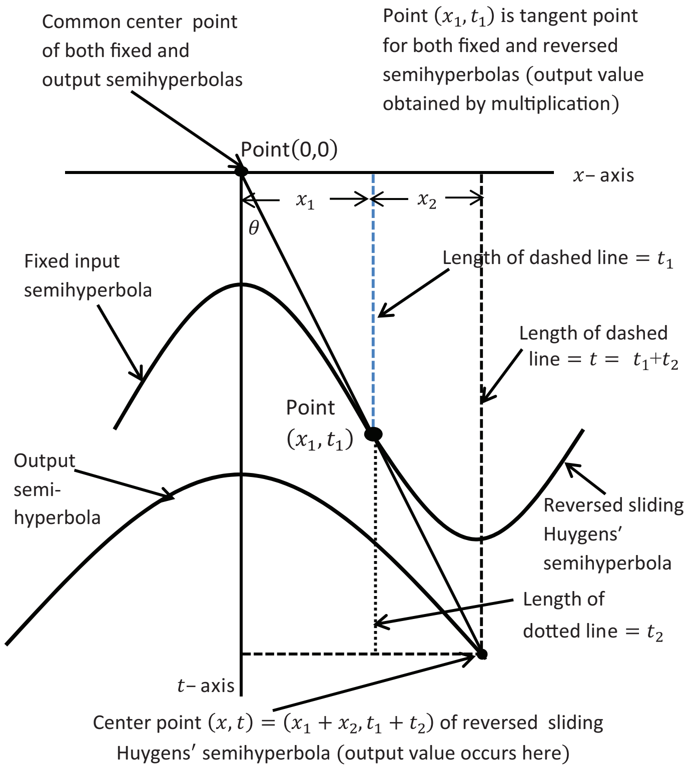


Figure 7. Huygens' space–time principle.

The first equality follows from the property which says that the gradient gives the rate of change in the direction of the normal. The second equality follows from the eikonal equation. From the equation defining *S*, namely, the equation $t(\boldsymbol{r}) = t$, we know that

$$\Delta t(\boldsymbol{r}) = \Delta t \ . \tag{4}$$

Thus, Δs must satisfy

$$p\Delta s = \Delta t$$

or

$$\Delta s = v\Delta t \ . \tag{5}$$

Thus, the new wavefront is obtained by advancing in the direction of the gradient (that is, along the ray) by an amount $v\Delta t$. Thus, the new wavefront is equivalent to the surface obtained from the envelope of all forward-moving spherical wavelets of radius $v\Delta t$ originating on the old wavefront. In fact, each such wavelet touches the envelope at the same place as the ray does. This entire discussion requires that our increments be infinitesimal, unless the velocity is constant.

Reflection and refraction

Next, we would like to consider a number of phenomena related to the propagation of waves and their interaction with material media. In particular, we shall study the characteristics of waves as they progress through various substances, crossing interfaces, while being reflected and refracted in the process. Many of the basic principles of wave motion are predicated on the wave aspects of the phenomenon and are yet completely independent of the exact nature of that wave. As we shall see, the most important such principle is Huygens' principle.

Suppose the wave impinges on the interface separating two different media. As we know, a portion of the incident flux density will be diverted back in the form of a reflected wave, while the remainder will be transmitted across the boundary as a refracted wave. We now seek to determine the general principles governing, or at least describing, propagation, reflection, and refraction. Suppose that we have a monochromatic plane wave incident on the smooth interface separating two different media. We can determine the wave's behavior using Huygens' construction.

In Figure 8, the angles θ_i, θ_r, and θ_t are the angles of incidence, reflection, and transmission (or refraction), respectively. The application of Huygens' principle gives

$$\frac{\sin\theta_i}{v_i} = \frac{\sin\theta_r}{v_r} = \frac{\sin\theta_t}{v_t} \quad . \tag{6}$$

It then follows from the first two terms that the *angle of incidence equals the angle of reflection;* that is

$$\theta_i = \theta_r \ . \tag{7}$$

This law of reflection first appears in *Catoptrics*, a book which is purported to have been written by Euclid. The first and last terms yield

$$\frac{\sin\theta_i}{\sin\theta_r} = \frac{v_i}{v_r} \quad . \tag{8}$$

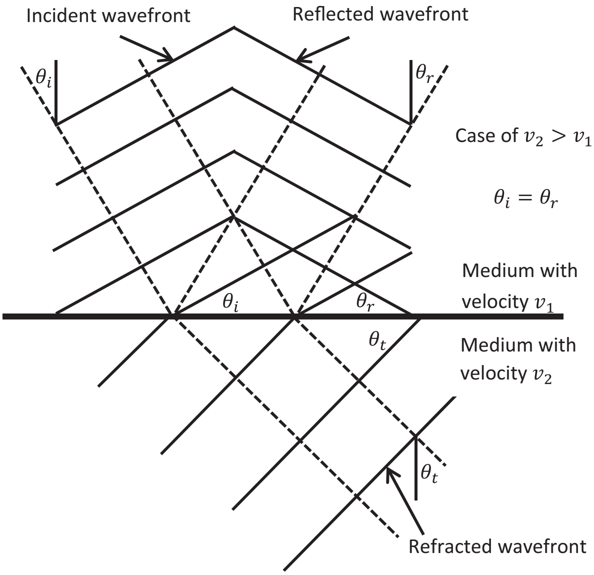


Figure 8. Angle $\boldsymbol{\theta_i}$ of incidence, angle $\boldsymbol{\theta_r}$ of reflection, and angle $\boldsymbol{\theta_t}$ of transmission (or refraction).

This is the *law of refraction*. On the basis of some very fine observations, Claudius Ptolemy of Alexandria finds the expression

$$\frac{\theta_i}{\theta_r} = \frac{v_i}{v_r} \;, \tag{9}$$

which is approximately correct for small angles. Kepler very nearly derives the law of refraction in his book, *Supplements to Vitello*, in 1604. Unfortunately, he is misled by some erroneous data compiled earlier by Vitello (ca. 1270). Finally, the correct relationship seems to have been determined independently by Snell at the University of Leyden and the French mathematician, Descartes. The law of refraction generally is referred to as *Snell's law*.

Fermat's principle

The laws of reflection and refraction, and indeed many aspects of the manner in which waves propagate in general, can be understood by means of Fermat's principle. This principle provides an insightful and highly useful way of appreciating and anticipating the behavior of waves.

Hero of Alexandria first set forth a variational principle. In his formulation of the law of reflection, he asserted that the path actually taken by light in going from some point *S* to a point *P* via a reflecting surface was the shortest possible one. For more than 1500 years, Hero's observation stood alone. Then, in 1657, Fermat propounded his principle of least time, which encompassed both reflection and refraction. Because a beam of light traversing an interface does not take a straight line or minimum spatial path between a point in the incident medium and one in the transmitting medium, Fermat reformulated Hero's statement to read: *The actual path between two points taken by a beam of light is the one which is traversed in the least time.*

The original statement of Fermat's principle of least time as given above is in need of some modification. To that end, recall that, if we have a function, such as $f(x)$, then we can determine the specific value of the variable x, which makes $f(x)$ have a stationary value, by setting $df/dx = 0$ and solving for x. By a stationary value, we mean a value for which the slope of $f(x)$ versus x is zero. Equivalently, a stationary value is one where the function has a maximum, a minimum, or a point of inflection having a horizontal tangent.

Fermat's principle in its modern form reads: *A ray in going from point* S *to point* P *must traverse a path for which the transit time is stationary with respect to variations of that path.* In other words, the transit time for the true trajectory will equal, to a first approximation, the transit times of paths immediately adjacent to it. And so there will be many curves neighboring the actual one, which would take very nearly the same time for the ray to traverse. This latter point makes it possible to begin to understand how a ray manages to be so clever in its meanderings. Suppose that we have a beam of wave motion advancing through a homogeneous isotropic medium so that a ray passes from points S to P. Particles within the material are driven by the incident disturbance and they reradiate in all directions. Quite generally, wavelets originating in the immediate vicinity of a stationary path will arrive at P by routes which differ only slightly and will therefore reinforce each other. Wavelets taking other paths will arrive at P out of phase and therefore tend to cancel each other. That being the case, energy will effectively propagate along that ray from S to P, which satisfies Fermat's principle.

The transit time for a ray need not always be a minimum. A nonminimum situation is depicted in Figure 9, which shows a segment of a three-dimensional ellipsoidal surface. If the source S and the observer P are at the foci of the ellipsoid, then, by definition, the length SQP will be constant regardless of where on the perimeter Q happens to be. It is also a geometrical property of the ellipse that $\theta_i = \theta_r$ for any location of Q. All transit times from S to P via a reflection are therefore precisely equal. None is a minimum, and the transit time is clearly stationary with

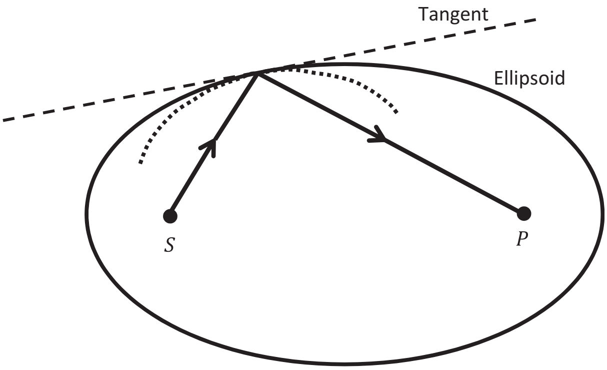


Figure 9. Segment of a three-dimensional ellipsoidal surface.

respect to variations. Rays leaving *S* and striking the surface will arrive at the focus *P*. From another viewpoint, we can say that radiant energy emitted by *S* will be scattered by the ellipsoidal surface such that the wavelets will substantially reinforce each other only at *P* where they have traveled the same distance and have the same phase. In any case, if there is a plane interface tangent to the ellipse at *Q*, the exact same path *SQP* traversed by a ray will then be a relative minimum. At the other extreme, if the mirrored surface conforms to a curve lying within the ellipse, like the dashed one shown, that same ray along *SQP* will now negotiate a relative maximum transit time. This is true even though other unused paths (where $\theta_i \neq \theta_r$) would actually be shorter (i.e., apart from inadmissible curved paths). Thus, in all cases, the rays travel a path with a stationary transit time in accordance with Fermat's principle. Note that, because the principle speaks only about the path and not the direction along it, a ray going from *P* to *S* will trace out the same route as one from *S* to *P*. This is the very useful *principle of reciprocity*.

Suppose that we have a stratified material composed of ***m*** layers, each having a different wave velocity as in Figure 10. Then the transit time from *S* to *P* will be

$$t = \frac{s_1}{v_1} + \frac{s_2}{v_2} + \cdots + \frac{s_m}{v_m}$$

or

$$t = \sum_{i=1}^{m} \frac{s_i}{v_i} \ , \tag{10}$$

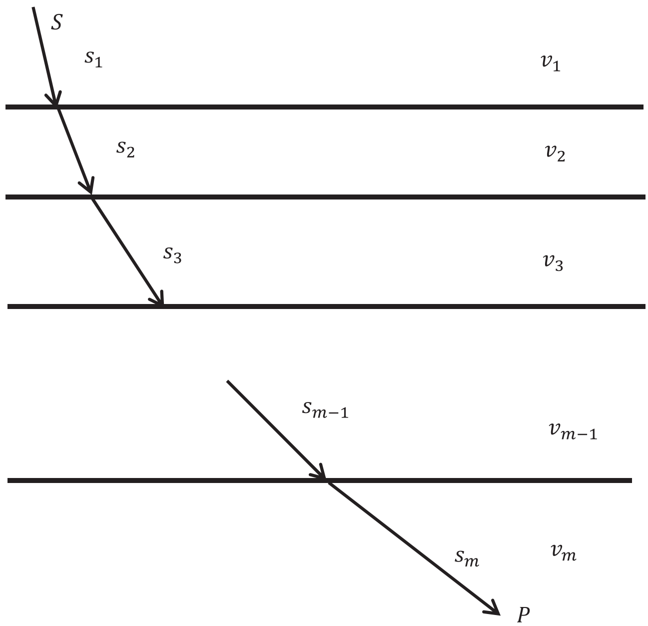


Figure 10. Stratified material composed of ***m*** layers, each having a different wave velocity.

where s_i and v_i are the path length and speed associated with the ith contribution. For an inhomogeneous medium where v is a continuous function of position, the summation must be changed to an integral

$$t = \int_S^P \frac{ds}{v(s)} \; . \tag{11}$$

Fermat's principle states that, going from points S to P, the ray paths traverse the route having the smallest transit time.

Thus far, we have merely stated Huygens' principle without any justification or proof of its validity. In 1814, Fresnel successfully further analyzes Huygens' principle; then in 1882, Kirchhoff shows that the Huygens' principle is a direct consequence of the wave equation, thereby putting it on a firm mathematical base. That there was a need for a bit of reformulation of the principle is evident from the figures, where we deceptively only drew hemispherical wavelets. Had we drawn them as spheres, there would be a back wave moving toward the source; something which is not observed. This difficulty is addressed theoretically by Fresnel and Kirchhoff, and we will present this theory in the last chapter. For the present, we merely use the forward waves when applying Huygens' construction.

Resolution and the Fresnel zone

A modeling program is used to compute the seismic data section that would result from a given configuration of geologic layers. It is a useful tool in seismic interpretation for relating the events displayed on seismic records to the underground structure. A close fit of the data generated by the model to the data actually observed in nature, however, does not mean that the model structure is the same as the earth's structure. Other models might give approximately the same data. When we change the parameters of the model so that the generated data conforms to the observed data, good results do not necessarily mean that we have achieved the true model. Despite this inherent limitation, models still represent one of the best approaches to solving the seismic inverse problem.

The shape of a reflector influences the type of reflection seen on the seismic section. The curvature of the reflector serves to focus or defocus the reflected energy.

As shown in Figure 11, curvature that is concave upward tends to concentrate the reflected energy, thus causing an amplitude increase on the reflected event. A curvature that is convex upward tends to spread the reflected energy, thus causing an amplitude decrease. Resolution refers to the ability to detect separate features that lie close together and may produce events that overlap. We can speak of time resolution and spatial resolution. Our ability to resolve two features depends upon the distance between them in comparison with the wavelengths of the seismic waves with which we illuminate them. Thus, shorter wavelengths have higher resolving power. Although the only way to get smaller wavelengths is to increase the frequency, there are several practical reasons why we do not

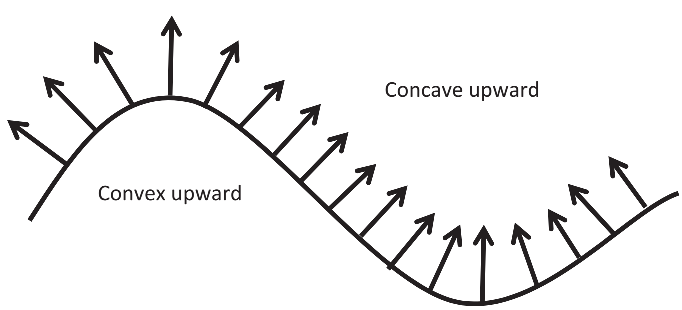


Figure 11. Convex upward curvature spreads reflected energy causing decrease in amplitude. Concave upward curvature concentrates reflected energy causing increase in amplitude.

record higher frequencies on the seismic records. In order to filter out horizontally traveling energy, such as the surface waves known as ground roll, we combine (mix) the outputs of a group (array) of geophones to produce a single seismic trace. This ground mix attenuates high frequencies and thus decreases seismic resolution. Without ground mix, however, the deep seismic reflections would be lost in the noise represented by the surface waves. Without adequate resolution, we would not be able to detect the presence of thin beds or small wedges and pinchouts. Such features require vertical resolving power.

Spatial resolving power must be a consideration too. We need to know how large a structure must be in order to be seen on seismic data. Spatial resolving power involves the concept of the Fresnel zone. From the ray point of view, seismic energy travels from the source to the reflector along the raypath. Each ray behaves in a way irrespective of adjacent rays. According to this interpretation, a reflection involves only a point on the reflecting interface. Although this ray-theoretical approach is often useful, it is more realistic to think physically in terms of wavefronts, the in-phase location of a disturbance moving outward from the source. Detectors buried within the earth would observe a wavefront as it passes. As a wavefront reaches a reflector, part of it will be reflected. Consider a spherical wavefront and a plane reflecting interface. Suppose that the seismic wavelet consists of approximately 1.5 cycles. As a result, the disturbance will continue for a region behind the leading-edge wavefront. Consider the leading-edge wavefront when the part of the disturbance, which is a quarter wavelength behind the wavefront, is tangent to the reflector.

As shown in Figure 12, the portion of the reflector between points of contact with the wavefront is the area which, in effect, produces the reflection. This area is called the first Fresnel zone. If we take into consideration two-way traveltime, we see that the energy from the periphery of the first Fresnel zone will reach a detector at the source location one-half wavelength later than energy from the center of the first Fresnel zone. Thus, all of the energy reflected from the first Fresnel zone will arrive at the detector within one-half wavelength, and will therefore interfere constructively. If the center point (i.e., the reflecting point) is removed, as by cutting a small hole in the interface, the reflection will be observed nonetheless. The concept of Fresnel zone shows that an area instead of a point on the reflector produces a reflection. This distinction is the essence of understanding spatial resolution. In summary, a Fresnel zone is that portion of a reflecting interface such that all reflectors from that portion will arrive back at the source point within a half-cycle of each other, thereby producing

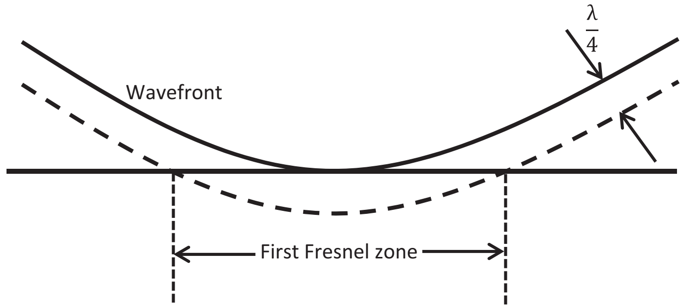


Figure 12. Cross section of the first Fresnel zone is circular. Subsequent Fresnel zones are annular in cross section, and concentric with the first. Odd-numbered Fresnel zones have relatively intense field strengths, whereas even-numbered Fresnel zones are nulls.

constructive interference. The reflection from the outer edge of the first Fresnel zone gains one-half wavelength (one-quarter wavelength on the downward path plus one-quarter wavelength on the upward path) in comparison to the reflection from the center. Thus, the edge reflection will arrive one-half wavelength after the center reflection. All reflections within the zone are sandwiched between these two limits and add constructively. The first Fresnel zone is surrounded by a second Fresnel zone which is a ring. The reflected energy from within this ring is delayed by one-half cycle to one cycle. The third Fresnel zone is a ring which surrounds the second. The reflected energy from the third Fresnel zone is delayed by one cycle to one and one-half cycle. Thus, the Fresnel zone concept can be extended to all higher orders. The higher zones will nearly cancel each other, however, so the net effect is that of the first zone. The dimension of the first Fresnel zone is the important factor in finding what part of the reflecting interface produces a seismic reflection. Because wavelength depends upon frequency, the dimensions of this zone depend upon the actual frequencies present in the seismic wavelet. Because a wavelet is composed of different frequency components, first Fresnel zones of different size are operative in the reflection of a seismic wavelet. Usually, however, we consider only the first Fresnel zone associated with the dominant frequency present in the wavelet. The wavelet is smaller for a high-frequency wave, and thus the zone is smaller. Suppose a wavelet has two dominant frequencies, one high and the other low. A small zone is responsible for the reflection of the high-frequency component, whereas a large zone is responsible for the reflection of the low-frequency component. If a change occurs in the reflecting interface, it would affect the low-frequency component more than the high-frequency component, and the result would be a change in the wave shape of the reflected wave.

As an example, consider a plane reflecting interface at a depth of 5000 ft and an average velocity of 10,000 ft/sec. The first Fresnel zone has a radius of 793 ft for a 40 Hz component, and 1125 ft for a 20 Hz component. Here, we have assumed that the wavefront is spherical. The radius of the Fresnel zone depends upon the curvature of the wavefront. For spherical waves originating at a surface source point, the curvature of the wavefront will decrease as the depth of the reflecting interface increases. Thus, spatial resolving power deteriorates with depth. A deep feature must have a larger areal extent in order to produce the same effect as a smaller feature at less depth.

Diffraction

So far, we have considered only some of the simplest applications of the wave equation in the presence of matter. Problems in which the acoustic impedance is described by a continuous function of position cannot be solved in general. There are no specific variations of the acoustic impedance of broad general interest for which an integral of the wave equation exists. Exact solutions, even in situations where the boundaries are only slightly more complicated than those we have considered, are rare. For example, there is no exact solution for the propagation of a wave in the presence of an infinite plane in which there is a hole of arbitrary shape. Where an exact solution exists, as it does for such special geometries as a semi-infinite half plane or an infinite circular cylinder, we can discern some general aspects of the solution.

Next, considering Figure 13, let a plane be drawn from the source to the edge of the boundary of the half plane, or two planes from the source tangent to the infinite cylinder; the regions behind these planes are called *shadow regions*.

The exact solutions indicate that there are variations of intensity of the wave in the vicinity of the shadow boundary and that there are alternations of light and darkness (in the case of light waves) that are appreciable over distances of the order of ten times the wavelength of light in the shadow regions. This phenomenon, known as diffraction, is a typical wave phenomenon. We think of light as traveling in straight lines, which by and large it does. Because of this, an object in the path of light casts shadows. The straighter the line that light travels, the sharper the shadow. Upon close examination, however, even the sharpest shadow is blurred slightly at the edges. This is because there is a slight bending of light around the edges

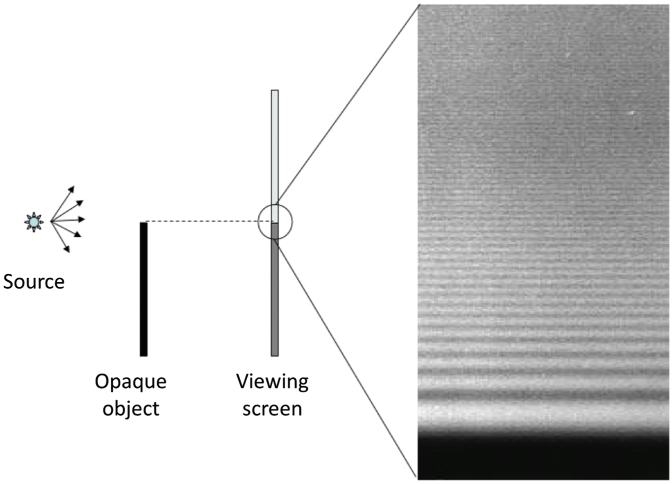


Figure 13. Diffraction occurs within distances of a few wavelengths from the boundary of the shadow.

of the object. This bending of light (or any wave motion) around corners is called *diffraction*.

Waves of any kind superimposed on one another produce a resulting wave that is different than either wave alone. The superposition of waves and the resulting interference is fundamental to the study of diffraction. Recall that longitudinal sound waves interfere to produce beats. In a similar way, the interference of transverse light waves produces colors. Constructive and destructive interference is reviewed in Figure 14.

Interference of water waves is a common sight. A negative consequence of seismic wave propagation is the limitation it imposes on seismic exploration. The events due to small objects become less and less well defined as the size of the object approaches the wavelength of the seismic waves illuminating it. If the object is smaller than a single wavelength, then careful data collection, data processing, and data interpretation are required. Even so, no amount of theory can defeat this fundamental diffraction limit. Approximation methods for the solution of the wave equation with variable material parameters and for situations involving complicated boundaries are available. Some of these methods have been designed to exhibit diffraction. We will be particularly interested in the Fresnel approximation. The phenomenon known as diffraction plays a role of the utmost importance in the branches of physics and engineering that address wave propagation. In this chapter, we consider some of the foundations of scalar diffraction theory. The theory discussed here is sufficiently general to be applied in other fields, such as acoustic-wave propagation, radio-wave propagation, and physical optics. To fully understand the principle of seismic imaging

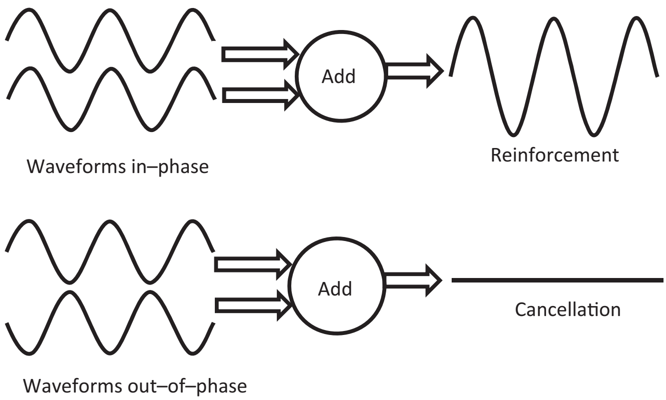


Figure 14. Constructive and destructive interference.

and seismic data processing systems, it is essential that diffraction and the limitations it imposes on system performance be appreciated.

The term diffraction can be defined as any deviation of rays from rectilinear paths which cannot be interpreted as reflection or refraction. The classic description of the diffraction of light rays is as follows. An aperture in an opaque screen is illuminated by a light source, and the light intensity is observed across a plane some distance behind the screen. The rectilinear theory of light propagation predicts that the shadow behind the screen should be well defined, with sharp borders. Observations show, however, that transition from light to shadow is gradual rather than abrupt. If the quality of the light source is good, then we would observe more striking results, such as the presence of light and dark fringes extending far into the geometrical shadow of the screen. Such effects cannot be satisfactorily explained by a strict ray theory of light, which requires rectilinear propagation of light rays in the absence of reflection and refraction. The initial step in the evolution of a theory that would explain such effects is made by the first proponent of the wave theory of light, Christiaan Huygens, in the year 1690. Huygens expresses an intuitive conviction that, if each point on the wavefront of a light disturbance is considered to be a new source of a *secondary spherical disturbance*, then the wavefront at any later instant can be found by constructing the *envelope* of the secondary wavelets.

When a wave passes a point in a medium, the resulting disturbance of the medium at that point is itself a source of a new wave motion of the same frequency (Huygens' principle). Consider waves of water incident from the left upon a barrier with a narrow opening. First, suppose the opening is small compared to the wavelength of the wave. While the waves are incident upon the opening, the water sloshing up and down in the opening acts as a "point" source of new waves. As a result, concentric-shaped waves are produced on the other side of the barrier. As the opening is widened, the waves produced on the other side no longer emanate from a point-like source, and the resulting waves are less circular in shape. When the opening is very wide compared to the wavelength of the waves, the waves from the left side simply pass through unobstructed with only slight diffraction at the edges. In the case of light waves, this slight diffraction blurs the edges of a shadow. If the source of light causing the shadow of an object is very small (preferably a point), fringes of brightness and darkness can be seen.

The amount of diffraction depends on the wavelength of the wave. Radio waves are very long, ranging from 176 to 560 meters for the standard AM broadcast band. As a result, AM radio waves readily bend around objects that might otherwise obstruct them. The radio waves of the FM band are much shorter, ranging from 2.8 to 3.4 meters, and do not bend as much as AM waves. This is one of the reasons that FM reception is often poor in localities where AM comes in loud and clear. Diffraction, therefore, aids radio reception.

Seismic waves, which have wavelengths as large as a few hundred meters, bend around pinchouts, disconformities, and other geologic structures to various extents. The equation is

$$\text{wavelength} \times \text{frequency} = \text{velocity} \ .$$

For example, a wavelength of 10 m for a seismic wave of 100 Hz frequency traveling in the shallow earth (e.g., the upper 10 m in depth) would have a velocity of 1000 m/s. A wavelength of 250 m for a seismic wave of 20 Hz frequency traveling in the deep earth (e.g., depth of 5000 m) would have a velocity of 5000 m/s.

Diffraction curves

Although seismic reflection traveltime profiles or sections often give a remarkably direct indication of the structural configuration of the layers of sedimentary rock lying beneath the earth's surface, they can be misleading and often are difficult to interpret geologically even when record quality (event continuity and the reflection signal-to-background noise ratio) is excellent. For years, it has been the task of trained and experienced exploration seismologists to interpret the data, sort out and correlate the signal—"pick the record"—and plot a series of points or a line, representative of these correlations, on a separate section. When the structure is complex, showing large deviations from horizontal layering, the subsurface position of the reflecting point on the interface does not lie under the shot-receiver but is displaced to one side or the other of this point. The direction and magnitude of this updip displacement depends on the direction and magnitude of the inclination or dip of the strata, bedding plane, or seismic interface. In our analysis here, we address two spatial dimensions only, namely, a horizontal distance x and depth z. The variable x can take on any value from $-\infty$ to $+\infty$, but the depth z must be positive. Also, time t must be positive. Because z and t must be positive, there is a certain duality between them. There is nothing in our methods that cannot be extended to three spatial dimensions; the extension is straightforward and involves no new principles. In this discussion, we will make use of the stacked seismic record section which (approximately) gives coincident source-receiver geometry, as shown in Figure 15.

Although the travelpaths of energy between surface source locations and positions along the true reflector surface may be quite complex, we do know that upward and downward legs must be identical, and that the travelpath (a raypath) strikes the reflector surface at right angles. This last fact follows from the law of reflection (incidence angle of 0° equals reflection angle of 0°).

The wave equation describes the motion of the waves generated by a physical experiment. The stacked record section, however, does not correspond to a wavefield resulting from any single experiment. There are many sources excited sequentially, but the record section gives the appearance that all of the sources are activated simultaneously. As a result, a theoretical physical experiment is hypothesized to justify the use of the wave equation to operate on the wave motion appearing on the stacked record section. This theoretical physical experiment (Loewenthal et al., 1976) is called the exploding reflectors model. The theoretical experiment is the following. The receivers are located on the surface of the ground; the sources are not at the surface but are distributed within the earth. More specifically, along every reflector surface, the sources are positioned

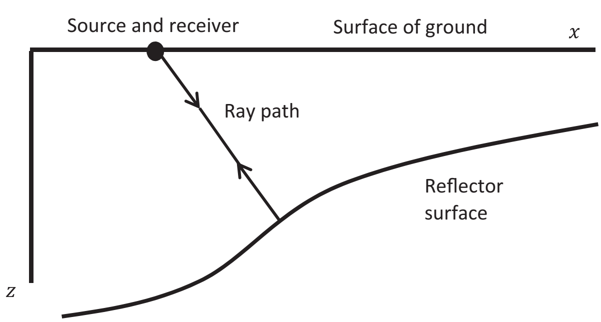


Figure 15. Raypaths for coincident source-receiver locations.

with the strengths that are proportional to the reflection coefficients, and all sources are activated at the same instant $t = 0$.

We concern ourselves with upward traveling waves only, that is, waves traveling upward toward the surface $z = 0$. We ignore all multiple reflections, and ignore all transmission effects at the interfaces. As we know, a record section involves the two-way traveltime from the surface source to the reflector, and back to the surface receiver. In our theoretical experiment, however, we are only concerned with one-way traveltime from a reflector source to the receiver. As a result, we must convert our record section from two-way traveltime to one-way traveltime. This conversion can be accomplished simply by dividing our stacked record section time scale by two. Now, seismic record generation can be stated in the following terms. The configuration of exploding reflectors is considered as an initial condition for a wavefield governed by the wave equation. Take the upward traveling waves at the exploding reflectors, run the time clock starting at time $t = 0$ so as to propagate the upgoing waves forward in time to the surface detector positions. This forward propagation to time t can be viewed as the mechanism that generates the stacked record section (i.e., section with coincident source-receiver pairs).

Let us summarize our development to this point. In conventional processing, stacking produces a zero-offset (i.e., coincident source-receiver) section, and this section forms the starting point for our discussion of diffraction. We will describe the diffraction process in detail. But before introducing diffraction, we want to discuss some special cases.

The first example is the case of a horizontal plane reflector (see Figure 16). With a coincident source-receiver, the seismic waves propagate vertically to the reflector from the source-receiver point and then back again to the same point. The true depth is denoted by z. The time points appearing on the seismic record (in the case of coincident source and receiver) represent the two-way traveltime t, that is, the time down to the reflector plus the (same) time back up to the surface. The one-way traveltime $t/2$ is obtained by dividing the two-way traveltime by two. We may convert record times t into apparent depths $vt/2$ by multiplying the one-way times $t/2$ by the constant velocity v. Because the raypaths are vertical, the interface appearing on the stacked section (called the *apparent interface*) is the same as the geometric interface (called the *true interface*).

Let us now consider the case for which the true interface is flat but sloping, as shown in Figure 17. Because we are concerned only with raypaths that are normally incident on the true interface, the true interface is tangent to the incident wavefront. The wavefront can be drawn at each source point corresponding to half of the measured reflection time. We

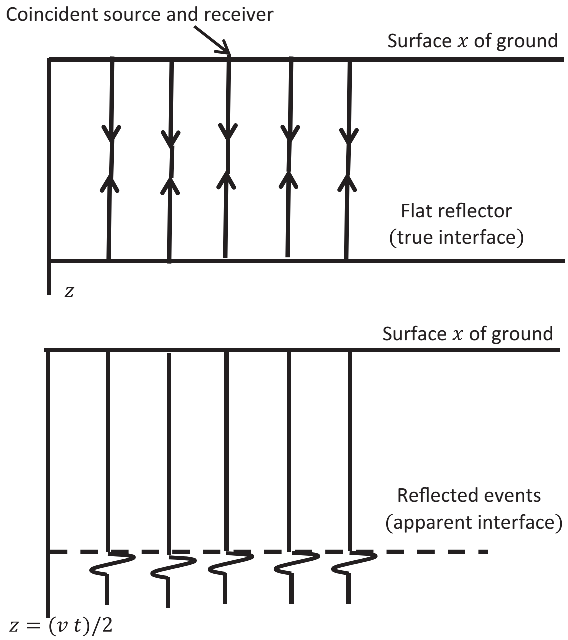


Figure 16. (top) Flat (i.e., horizontal) reflector and (bottom) resulting stacked section (with time converted to depth). In this case (i.e., the case of horizontal interfaces), the true depth z is the same as the apparent depth $vt/2$.

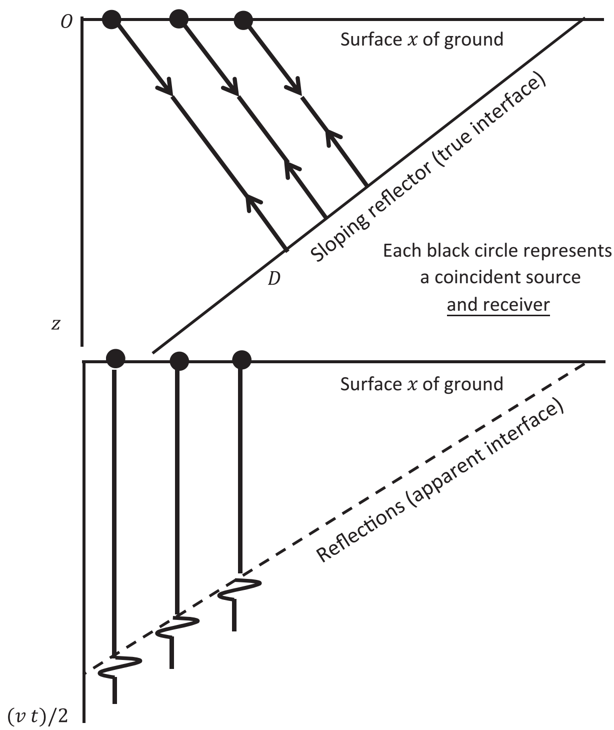


Figure 17. (top) Dipping reflector and (bottom) resulting stacked section (with time converted to depth). In this case (i.e., the case of a sloping interface), the true depth z is not the same as the apparent depth $vt/2$.

recall our assumption that the receiver is at the same point as the source point. Thus, the true interface is the envelope of these wavefronts. Because of the constant velocity assumption, the wavefronts are arcs of circles.

Figure 18 depicts the case of a sloping reflector in a three-dimensional figure, with axes x, $vt/2$, z. The event E is due to depth point D. The distances OE and OD are the same. We define α as the angle that the apparent interface makes with the horizontal axis (i.e., the x-axis), and β as the angle that the true interface makes with the x-axis. Then, from the figure, we see that

$$\tan\alpha = \frac{OE}{OP} \quad \text{and} \quad \sin\beta = \frac{OD}{OP} \ .$$

Because $OE = OD$, we have

$$\tan\alpha = \sin\beta \ . \tag{12}$$

In the constant velocity case, the geometry of wave propagation may be described elegantly by the curves known as *conies* or *conic sections*. These curves arise from the intersection of a plane with a cone of revolution. The intersection formed by a plane that cuts every element of such a cone is an ellipse. In the situation where the plane is perpendicular to the axis of the cone, the section is a circle. The intersection formed by a plane which cuts both napes of the cone is a hyperbola. In the particular case where the plane is parallel to an element of the cone, the section is a parabola. The conic sections can be defined metrically as follows. The ellipse is the locus of points, the sum of whose distances from two fixed points, called the *foci*, is constant. The hyperbola is the locus of points, the difference of whose distances from two fixed points, the foci, is constant. The parabola

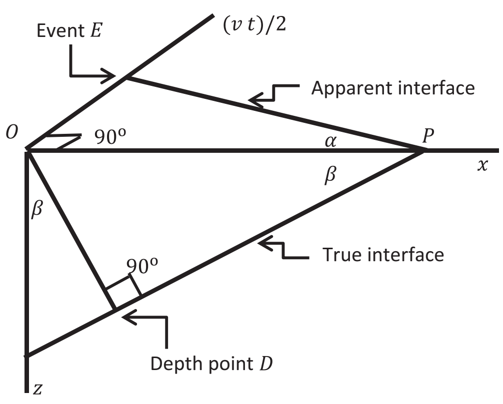


Figure 18. Dipping reflector in space–time diagram.

is the locus of points equally distant from a fixed point, the focus, and a fixed line, the directrix.

In seismic work, we are concerned with sources and receivers at the surface of the earth and reflectors at depth. A certain type of idealization deserves special study, namely, the case of a *point reflector*, also called a *point diffractor*. When such a diffraction point is illuminated by a surface source, it is assumed that the diffraction point acts as a secondary source and hence sets off outgoing wave motion in all directions. If t represents the two-way traveltime from a surface point to the diffraction point and back to the same surface point, then we must use the one-way time $t/2$ as the time when we regard the diffraction point as the secondary source. Thus, the one-way equal traveltime locus from this secondary source is described by the equation

$$x^2 + z^2 = \left(\frac{vt}{2}\right)^2 . \tag{13}$$

If z is the fixed depth of the diffraction point, then z is a constant in the above equation, and thus it becomes the equation for a hyperbola

$$x^2 - \left(\frac{vt}{2}\right)^2 = \text{constant} . \tag{14}$$

The seismic energy recorded at the surface is a function of the horizontal coordinate x and the two-way time t. On the recorded seismic section, shown in Figure 19, the diffraction point P produces energy which lies on the indicated hyperbola. As a result, such hyperbolas are called *diffraction curves*. The important fact to remember is that seismic sections are a function of x

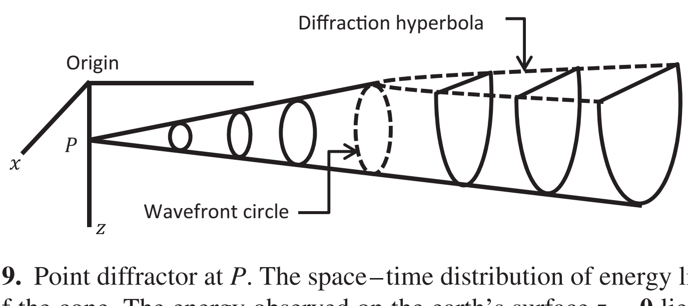


Figure 19. Point diffractor at P. The space–time distribution of energy lies on the surface of the cone. The energy observed on the earth's surface $z = \mathbf{0}$ lies on the diffraction hyperbola, whereas at any fixed instant of time the energy lies on a wavefront circle.

and t, and on such sections we can see diffraction curves as energy falling on a hyperbola.

Diffractions differ from true reflections. A shot produces energy that travels down to a diffraction point. Diffractions represent energy that returns from the diffraction point without obeying the reflection law (incidence angle equals reflection angle). Consider an example of a sharp diffracting edge at a fault. The diffraction events can be seen along the edges.

Suppose that we could see a cross-section of the earth; that is, a section that is a function of x and z. If we could take a snapshot of the wave motion at a fixed time t, then the wave motion due to the diffraction point would appear as the circle

$$x^2 + z^2 = \text{constant} . \tag{15}$$

As a result, such circles are called wavefronts. As time t increases, the wavefronts are an expanding set of concentric circles with the diffraction point as center (see Figure 18).

Instead of using a three-dimensional drawing with x, z, and $vt/2$ coordinates, let us now use a two-dimensional drawing with an x axis and another axis which depicts both z and $vt/2$. As depicted in Figure 20, both the z and $vt/2$ coordinates appear on the same vertical axis. The true interface is plotted with respect to the z-axis, while the apparent interface is plotted with respect to the $vt/2$ axis.

Let us now plot the wavefront curve and the diffraction curve on the same diagram. As shown in Figure 21, we observe that the wavefront

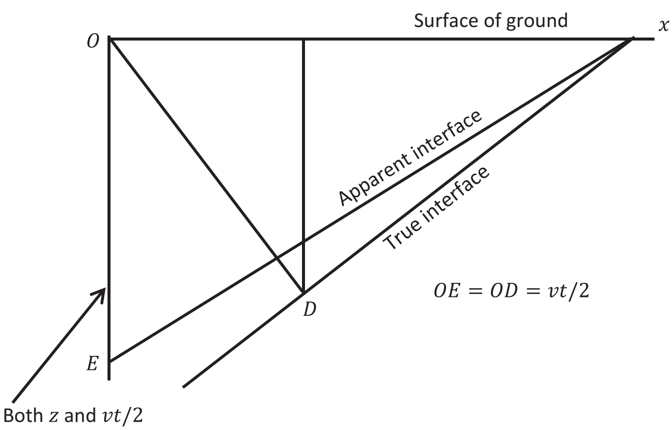


Figure 20. True interface (the reflector) and apparent interface (the reflected seismic event).

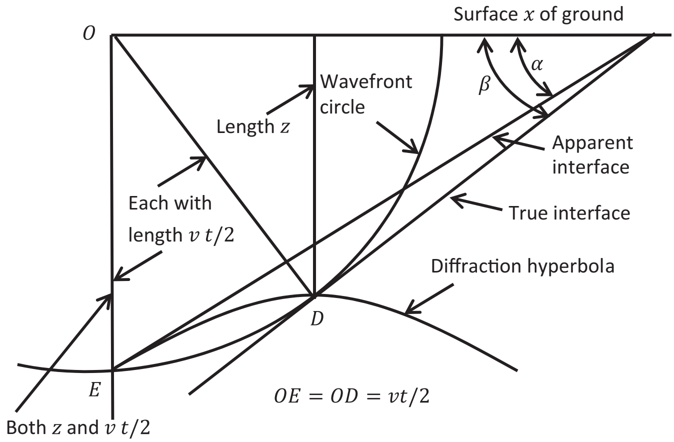


Figure 21. The wavefront circle and the diffraction curve intersect at two points E and D.

curve is the semicircle

$$x^2 + z^2 = \left(\frac{vt}{2}\right)^2 = \text{constant} \ , \tag{16}$$

while the diffraction curve is the semi-hyperbola

$$x^2 - \left(\frac{vt}{2}\right)^2 = z^2 = \text{constant} \ . \tag{17}$$

We see that the wavefront semicircle with center (0, 0) and the diffraction semi-hyperbola with apex (x, z) intersect at two points, namely, the apparent reflection point $E = (0, vt/2)$ and the true reflection point $D = (x, z)$. This fact may be directly verified by substituting each of these points into the respective equations for the semicircle and the semi-hyperbola.

The wavefront curve is defined as the snapshot that shows the wave energy originating from a point source. The true interface is the envelope of the wavefront curves corresponding to sources at the surface. This true interface is tangent to the wavefront curves, as shown in Figure 22.

The diffraction curve is defined as the locus of events given by the reflections originating from a point reflector. The apparent interface is the envelope of the diffraction curves corresponding to each of the points on the true interface. The apparent interface is tangent to the diffraction curves, as shown in Figure 23.

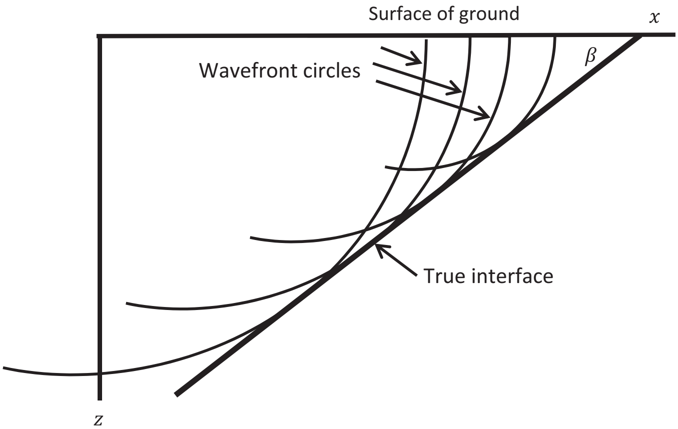


Figure 22. The true interface (the reflector) is tangent to the wavefront circles.

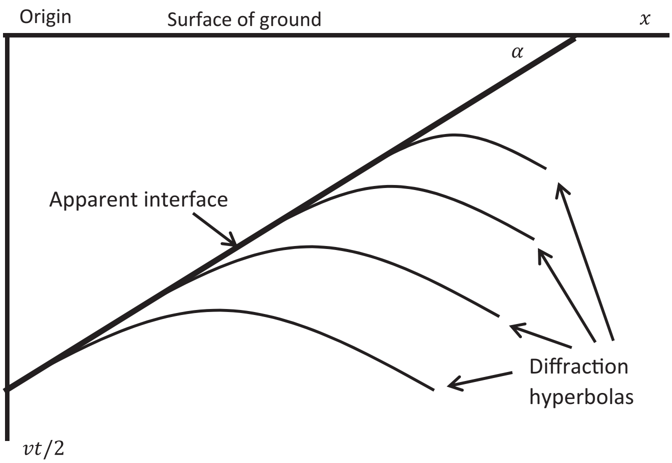


Figure 23. The apparent interface (the reflected seismic event) is tangent to the diffraction curves.

Let us consider the (x, z)-plane for non-negative values of z. We let z represent depth into the ground, so $z = 0$ represents the surface. First, we consider the wavefront that is due to a source at the surface point $(x_0, 0)$, and that has traveled a distance $vt/2$ to the depth point (x, z). This wavefront is the semicircle in (x, z) space, shown in Figure 24. The equation for the circle is

$$(x - x_0)^2 + z^2 = \left(\frac{vt}{2}\right)^2 = \text{constant} . \tag{18}$$

Next, we consider the diffraction curve that is due to a point diffractor at (x, z). As shown in Figure 25, let the source and receiver be at $(x_0, 0)$. The distance from $(x_0, 0)$ to the point diffractor is $vt/2$, where v is the wave

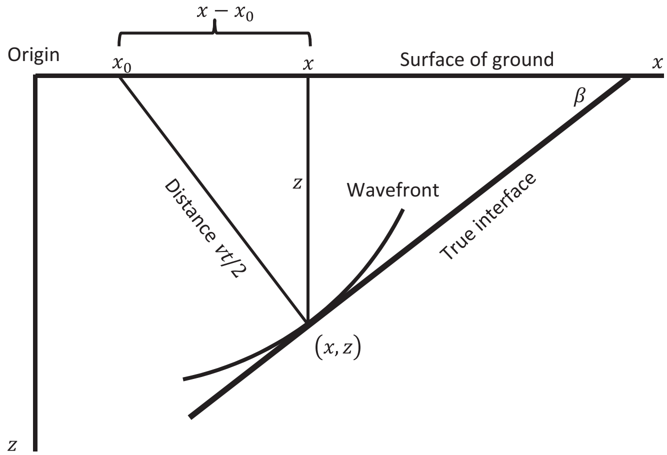


Figure 24. The wavefront due to a source at the point $(x_0,\ 0)$ on the surface of the earth.

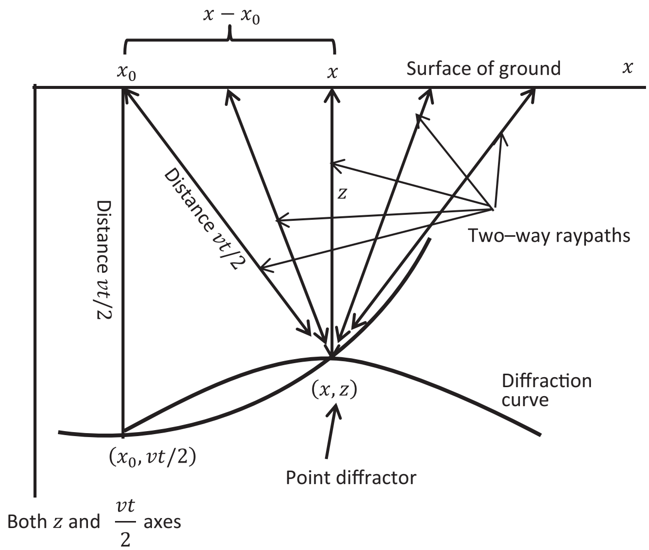


Figure 25. The diffraction curve due to a point diffractor at (x, z).

velocity and $t/2$ is the one-way time. Thus,

$$\frac{vt}{2} = \sqrt{(x - x_0)^2 + z^2}\ . \tag{19}$$

This distance is plotted directly under the source-receiver point $(x_0, 0)$; that is, the event corresponding to the point reflector (x, z) is

$$\left(x_0, \frac{vt}{2}\right) = \left(x_0, \sqrt{(x - x_0)^2 + z^2}\right)\ . \tag{20}$$

We recall that the diffraction curve is defined as the locus of events given by reflections from the point reflector for all source-receiver points $(x_0, 0)$. Hence, the diffraction curve is the locus given by equation 19, which is the semi-hyperbola in $(x, vt/2)$ space for a fixed depth z given by

$$\left(\frac{vt}{2}\right)^2 - (x - x_0)^2 = z^2 = \text{constant} . \tag{21}$$

We recall that the stacked section represents a zero-offset section. That is, the stacked section depicts events which have the same source and receiver position. Figure 26 depicts a point diffractor. Each line from the surface to the point diffractor represents a two-way travel path. The length of such a line is $vt/2$, where t is the two-way traveltime.

On the stacked section, the event that corresponds to the point diffractor falls at a distance of $vt/2$ directly under the source-receiver point, as shown in Figure 27.

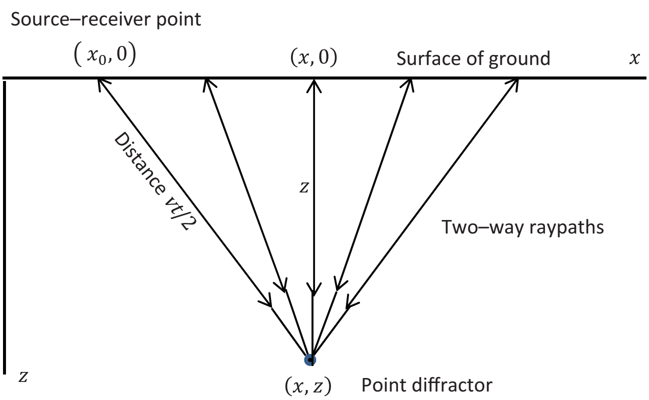


Figure 26. Point diffractor in the earth.

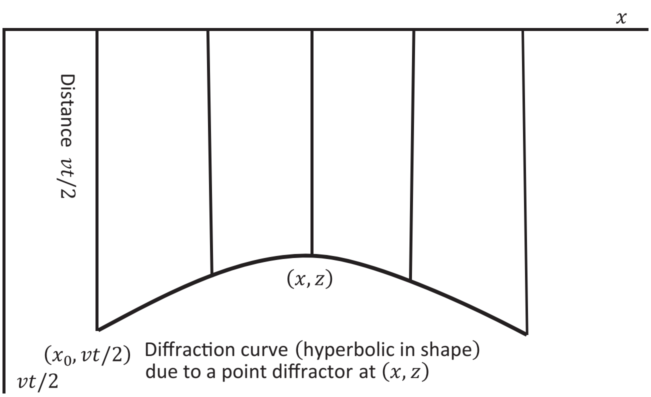


Figure 27. Diffraction curve on coincident source-receiver seismic section.

Diffractions appearing on seismic sections

In the foregoing section, we have learned:

1. A point diffractor in the subsurface produces a hyperbolic-shaped event (the diffraction curve) on the coincident source-receiver seismic section.
2. A straight-line reflecting interface with dip angle β produces a straight-line event with dip angle α on the coincident source-receiver seismic section. Both of these straight lines are assumed to be infinitely long, both cut the horizontal x-axis at the same point, and their dip angles are related by the equation $\tan\alpha = \sin\beta$.

In this section, as in the foregoing one, we are dealing with two-dimensional earth cross-sections (lateral coordinate x and depth coordinate z) and the corresponding two-dimensional seismic sections (same lateral coordinate x and time coordinate t). Now, however, we want to consider more practical cases than infinitely long rectilinear interfaces and point diffractors. The understanding of the diffraction process is basic to the comprehension of seismic phenomenon. For example, one portion of a reflector may possess a sharp boundary at which the reflection coefficient suddenly changes. We would want to be able to tell where that boundary point occurs. We will see that a discontinuity in the reflection capability of an interface does not mean a discontinuity in the seismic events which result. In fact, abrupt discontinuities in seismic events are not present on the seismic section. In studying the effects of discontinuities in the geologic structure on the resulting seismic records, resort usually must be made to numerical calculation. The mathematics that control such calculations are governed by Huygens' principle in its mathematical form as the Kirchhoff integral solution of the wave equation, or by various finite-difference or finite-element approximations to the wave equation, or by some transformation technique such as with the Fourier transform or the Radon transform (i.e., slant stacking).

Let us consider the case of a flat reflecting interface that terminates at an edge (see Figure 28). Suppose that a plane wavefront is incident on the reflector. A reflected plane wave will result, with angle of reflection equal to angle of incidence. However, the reflected plane wavefront will not terminate abruptly, but blends into a circular portion with center at the termination (edge) of the reflector. As a result, energy is diffracted into regions which are not in a position that is proper for a reflection. The energy that does not obey the law of reflection is organized by the geometry of the reflector (in this case a terminating flat interface) and makes up the diffracted

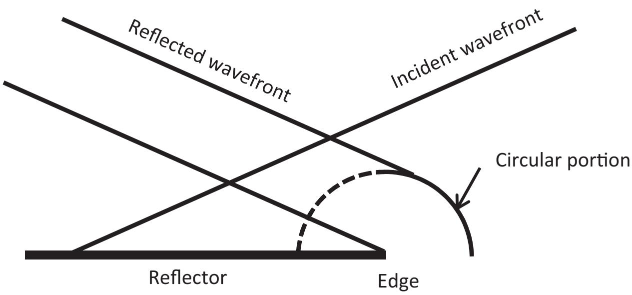


Figure 28. Reflector with edge, incident wavefront, reflected wavefront, and circular portion with center at edge.

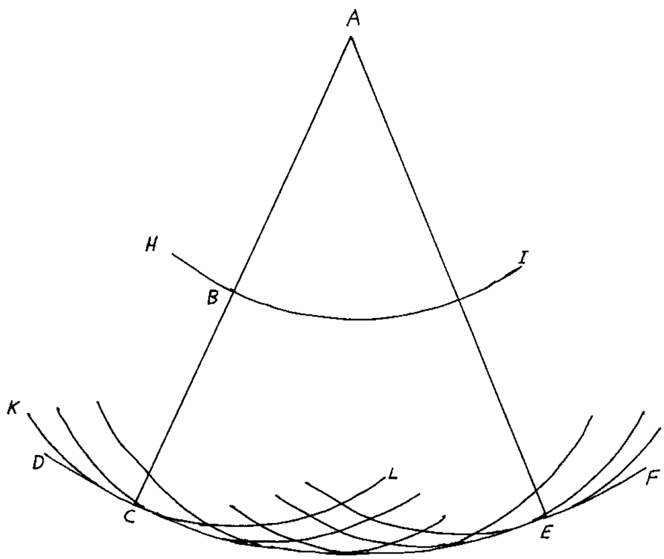


Figure 29. Wavelets as drawn by Christiaan Huygens.

wave. Huygens' principle explains how waves are propagated. It states that the motion of a particle of matter affects all of the surrounding particles. Each particle acts as a new elementary source and exerts an effect on all of the surrounding particles. A particle is linked to surrounding particles by elastic forces. The vibratory motion of a particle about its center of gravity changes the distances to the surrounding particles, and hence the elastic forces on them. Thus, these neighboring particles begin to vibrate about their respective centers of gravity.

To continue this story, let us quote from *Traité de la Lumière*, by Christiaan Huygens (1690), and consider Figure 29. (Note that Huygens is discussing the propagation of light, but his argument is also suitable to elastic waves in rock which we are discussing.)

> Each particle of the matter in which a wave proceeds not only communicates its motion to the next particle to it, which is on the straight line drawn from the luminous point, but to all the others which touch it and which oppose its motion. The result is that around this particle there arises a wave of which this particle is the center. So if *DCF* is a wave coming from the luminous point *A*, which is its center, the particle *B*, one of those which is within the sphere *DCF*, will set up its particular wave *KCL* which will touch the wave *DCF* in *C* at the same instant that the principle wave coming from the point *A* has reached *DCF*. It is plain that the only point of wave *KCL* which will touch the wave *DCF* is the point *C* which is on the straight line *AB*. In the same way the other particles contained within the sphere *DCF* will each have made its own wave. But each of these waves can be only infinitely feeble compared to the wave *DCF*, to the composition of which all the other contribute by that part of their surface which is the most distant from the center *A*. We see further that the wavefront *DCF* is determined by the extreme limit of motion which has gone out from the point *A* in a certain period of time, there being no motion beyond this wave. All the properties of light and those which pertain to reflection and refraction are explained fundamentally in this manner.

At the top of Figure 30, we see a flat half-plane reflecting interface; at the bottom, we see the resulting coincident source-receiver seismic section. This section shows the reflection from the half-plane (the flat event) and the diffraction from the edge of the plane (the hyperbola). The reason for the diffraction can be understood in terms of Huygens' principle.

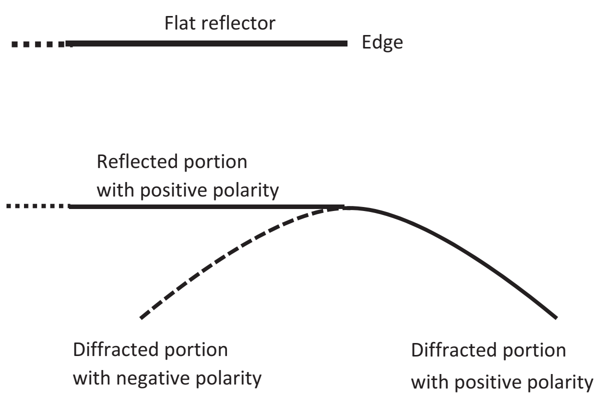


Figure 30. (top) Flat reflector with edge and (bottom) resulting seismic event composed of a reflected portion and a diffraction curve.

The particles near the edge of the reflecting plane radiate energy in all directions, and the portions of this energy with lateral components do not cancel, as would be the case if the reflecting plane were infinite in extent. Some of the important features of diffractions may be seen by examining this figure. The diffraction event is tangent to the reflection event, and the continuity of the entire event is maintained at the edge point where the reflection blends into the diffraction. There is no abrupt phase break to indicate the termination of the reflectors. Amplitude and wave shape are smooth and continuous. The amplitude of the reflection, however, decreases before the end of the reflector is reached, and is only half as strong at the end of the reflector as it is along the body of the reflector. The energy which disappears from the reflection as the edge is reached instead appears in the diffracted event. The two arms of the hyperbolic diffraction curve are interesting. The right arm, which represents the continuation of the reflector, has the same polarity as the reflected event. The left arm of the diffraction curve, which appears under the reflected event, has the opposite polarity as the reflected event. The magnitude of the amplitudes on the two diffraction arms at points equidistant from the apex are the same.

Consider the following thought experiment. Suppose we have an infinitely long flat reflector. As we know, it will produce an infinitely long reflection showing no diffraction. Suppose now we cut it, and consider only the left-side portion. We obtain a left-side reflection together with a diffraction hyperbola. If we consider only the right-side portion, we obtain a right-side reflection together with a diffraction hyperbola of the opposite polarity. The original infinitely long reflection must be the sum of the effects of the two half-planes. The two diffraction curves have the same amplitude but opposite polarity. As a result, they cancel each other completely. Thus, no diffraction would result at the junction of the

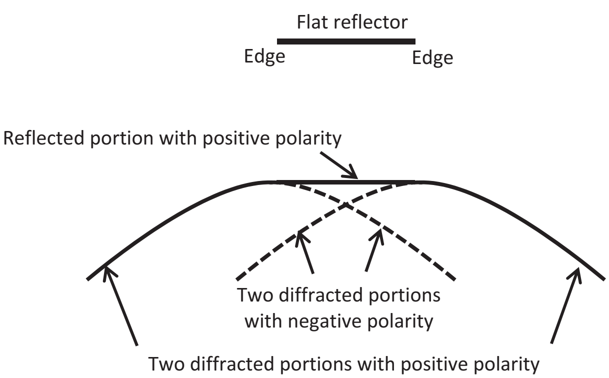


Figure 31. (top) Finite flat interface (two edges) and (bottom) resulting seismic event composed of a reflected portion and two diffraction curves.

component parts. At the junction, half of the energy is contributed by the left half-plane and the other half by the right half-plane (see Trorey, 1977).

In summary, the sign of a diffracted wave must change in going from one side of an edge to the other. For example, Figure 31 shows a strip with diffractions at each edge. Each branch (right and left) from each edge follows the same hyperbolic-like curve in time and distance, but, on the left-side edge, the diffraction is positive to the left and negative to the right. The opposite occurs on the right-side edge. The reflection itself is assumed to be positive, so that the two diffraction branches extending away from the reflection have the same sign as the reflection. Moreover, the diffraction causes the composite reflection and diffraction precisely at the edge to have only 50% of the amplitude of the main body of the reflection.

For those many readers who wish to probe deeper into the fundamentals of seismic diffraction theory and to see its many practical applications, reference is made to *Classical and Modern Diffraction Theory* and *Seismic Diffraction* (Klem-Musatov et al., 2016a, 2016b). Whereas the second volume is replete with working examples of situations encountered in seismic exploration, the first volume delves into the historical development of the subject. Here the reader can follow one of the most remarkable developments in the history of science as put forth in the papers: *A physico-mathematical treatise on light, colors and the rainbow* (Grimaldi, 1665); *Treatise on light* (Huygens, 1690); *Experiments and calculations relative to physical optics* (Young, 1803); *Memoir on the diffraction of light* (Fresnel, 1818); *Mathematical analysis to the theories of electricity and magnetism* (Green, 1828); *Theory of air vibration in pipes with open ends* (Helmholtz, 1858), *On the ray theory of light* (Kirchhoff, 1462), and *On the passage of waves through apertures in plane screens, and allied problems* (Rayleigh, 1697).

Chapter 8

Migration by the WKBJ Method

Geometrical optics and physical optics

The idiom, "seeing is believing," means that only physical or concrete evidence is convincing. When we look at the ocean, we see waves, not rays. Yet, we do see rays of light coming in through cracks in a wall. Light is composed of packets of energy called photons. However, light travels as a wave, not as a particle. But unlike sound waves or water waves, light waves do not need matter or material to carry its energy. Unlike sound waves which are mechanical waves that can only travel through a solid, a liquid, or a gas, light waves are electromagnetic waves that can travel through a vacuum.

Light waves travel out from their source in straight lines called *rays*. Rays do not curve around corners; so, when they hit an opaque object (one that does not allow light to pass through it), they are blocked from reaching the other side of that object. We see a dark shadow (geometrical shadow) in the area from which light is blocked. Even so, some light (diffracted light) does appear in the shadow zone. Diffraction is the bending of light around the corners of an obstacle into the region of geometrical shadow. Diffractions can be explained by wave theory. In classical physics, the diffraction phenomenon is described as the interference of waves according to the Huygens–Fresnel principle. These characteristic behaviors are exhibited when a wave encounters an obstacle or a slit that is comparable in size to its wavelength. Similar effects occur when a light wave travels through a medium with a varying refractive index, or when a sound wave travels through a medium with varying acoustic impedance. Diffraction occurs with all waves, including sound waves, water waves, and electromagnetic waves. While diffraction occurs whenever propagating waves encounter such changes, its effects are generally most pronounced for waves whose wavelength is roughly comparable to the dimensions of the diffracting object or slit.

Geometric optics, or ray optics, deals with light rays or "beams" of light. Light rays are conceptual lines along which the bulk of luminous energy is propagated. Geometric optics is applicable to the study of prisms, lenses, mirrors, and optical devices such as microscopes, telescopes, and cameras. In such cases, the wavelike properties of light become insignificant as the objects are very large as compared to the wavelength of light. Geometric optics is applicable if the diffractive effects are negligibly small. Geometric optics is the limiting case of physical optics when the wavelength tends to zero. In terms of frequency, geometric optics is the limiting case when the frequency tends to infinity. In geometric optics, the wave nature of light and the associated diffraction phenomena are not taken into consideration. In an optically isotropic medium, the light rays are orthogonal to the wave surfaces and directed toward the outward normals to these surfaces. Geometric optics can explain the phenomena of reflection and refraction. The field of geometric optics is already well developed by the time of Isaac Newton. The grinding of lenses to assist vision is already common by 1600, and Galileo understands how to make the first astronomical telescope in 1610. Isaac Newton even begins to understand colors, although he does not consider light as wavelike.

Fermat's principle states that the actual path of the propagation of light from point A to point B is such that the time required to traverse this path is extremal with respect to that required for any other conceivable path between these points. Fermat's principle follows from the Huygens–Fresnel principle under the condition that the wavelength of the light is infinitely small. It is the most general principle of geometric optics, one from which all of the fundamental laws of geometric optics can be derived. For instance, the law of rectilinear propagation of light in an optically homogeneous medium follows from Fermat's principle. The laws of the reflection and refraction of light also follow from Fermat's principle. In addition, the principle of optical reversibility is a consequence of Fermat's principle. Suppose a ray in medium 1 is incident to the boundary with medium 2 at an angle θ_i. The ray is refracted and enters medium 2 at angle θ_r. This principle states that then a ray from medium 2 falling on the boundary with medium 1 at angle θ_r will enter medium 1 after refraction at angle θ_i.

Maxwell's electromagnetic theory puts the wave theory of light on firm footing. Physical optics, or wave optics, is the branch of optics which studies interference, diffraction, polarization, and other phenomena for which the ray approximation of geometric optics is not valid. Physical optics takes into consideration the wave-like properties of light. It develops more advanced concepts on the basis of the Huygens–Fresnel principle. It has its origins in Young's double slit experiment and consequently with interference of light which is a characteristic of waves. Physical optics deals with diffraction which is noticeable only when the obstacle's size is of the order of the wavelength of light. The subject matter of physical optics develops after the 1801 discovery by Thomas Young that light has wave properties. In this context, physical optics is an intermediate method between geometric optics, which ignores wave effects, and full wave electromagnetism, which is a precise theory. The word "physical" means that it is more physical than geometric or ray optics and not that it is an exact physical theory.

The diffraction of light is observed in media with sharply defined inhomogeneities (for example, holes in opaque screens, the boundaries of opaque bodies, etc.). In its narrower sense, diffraction refers to deviations from the laws of geometric optics, such as the bending of light around small opaque obstacles. A strict mathematical solution of diffraction problems, based on the wave equation and having boundary conditions depending on the nature of the obstacle, is exceptionally difficult. It usually becomes necessary to resort to approximate methods.

Huygens' principle (in the form as given by Huygens) states that the position of the front of a traveling wave can be represented at any instant of time as the envelope of all secondary wavelets. The sources of the secondary wavelets are points reached by the front of the primary wave in the preceding instant of time. The principle explains the laws of the reflection and refraction of light; however, it cannot explain diffraction phenomena. Its more advanced form is called the Huygens–Fresnel principle. Consider a source of a light wave and an arbitrary closed surface surrounding the source. The Huygens–Fresnel principle states that, at any point located outside of the closed surface, the primary wave is the result of the superposition of secondary wavelets which are emitted by elementary sources distributed continuously along the closed surface. In other words, outside the closed surface, the primary wave can be replaced by secondary wavelets which interfere upon superposition. The closed surface is usually made to coincide with one of the wave surfaces of the primary wave so that the initial phases of all secondary wavelets are the same.

Migration is a term used in reflection seismology to describe the process of moving the recorded reflection events to their correct spatial positions by backward projection (a.k.a. depropagation). Migration has gone through three distinct phases in the history of the exploration seismic method; namely,

Phase 1: 1921–1965. Mechanical migration.

Phase 2: 1965–1995. Computer migration, generally limited to two spatial dimensions because of severe limitations in computer power and data acquisition.

Phase 3: 1995–present. Computer migration, generally done in three spatial dimensions due to great advances in computer power and data acquisition.

Phase 1. The first reflection seismic survey recognizes that reflections can occur in any direction (bearing) and hence the reflecting points do not necessarily lie vertically beneath observation points on the surface of the earth. J. C. Karcher interprets his earliest reflection data at Belle Isle, Oklahoma in 1921 (see Figure 1). He assumes a constant velocity, swings arcs around the surface observation points, and draws the interface as the envelope of these arcs. Such a process of transforming the observed seismic events into a map of the reflecting horizons is the process of migration. Migration refers to the movement of an observed event to its true spatial position. The pre-computer methods of migration all involve manual (visual) picking of reflection events before migrating.

As the years passed, a variety of numerical, geometric, and mechanical schemes were devised to carry out the migration of seismic data. Wavefront and raypath charts were used, which allowed one to handle various velocity functions. Such charts could be constructed for arbitrary vertical and lateral distributions of velocity. However, except in special cases, velocity functions that varied only in the vertical direction were used, that is, velocity functions of the form $v(z)$. This development culminated in the work of Hagedoorn (1954) on the conceptual aspect of migration.

Phase 2. In the early 1960s, the exploration seismic industry underwent a digital revolution. Data were collected in digital form and extensive computer processing was performed to put the data in their final form. One of the

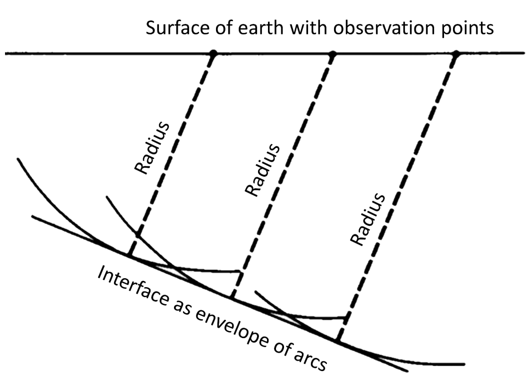


Figure 1. The Karcher migration scheme for the first seismic reflection survey in 1921.

processes that had to be computer programmed was migration. A major computational breakthrough occurred when various companies realized that migration could be accomplished by the use of existing stacking programs. The process of stacking made use of a level-layered model, that is, a model in which all of the interfaces are flat and horizontal. Geophysicists called such a model a layer-cake model.

One of the first steps in seismic processing is to collect or gather all traces whose respective source positions and receiver positions are centered around a common midpoint. Such a collection is called a *common midpoint gather* (CMP). Under a level-layered model, the curve of traveltime (for a reflection from a given level interface) versus horizontal offset (i.e., the horizontal distance from source to receiver) is hyperbolic in shape. Empirical hyperbolas are fitted to the observed hyperbolic-shaped events; the result is that an empirical velocity function $v(z)$ can be computed. Also, all of the traces on the common midpoint gather are summed along these hyperbolas, and the result is called the *stacked trace* for that midpoint. The stacked trace is plotted at the coordinate of that midpoint, and the collection of many of these stacked traces for a sequence of midpoints is called the *stacked section*.

It is the stacked section that forms the starting point for the conventional migration schemes used in Phase 2. Now we come to the important computational breakthrough. Hyperbolic-shaped events are observed on stacked sections, and they are recognized as diffractions from subsurface diffracting points. By using the velocity function in conjunction with the layer-cake model, a theoretical diffraction curve can be computed for each subsurface point. By adding the values of the traces that fall on each curve, we collapse the energy to the apex, which corresponds to the position of the diffraction point. Because an interface may be considered as a sequence of closely spaced diffraction points, this diffraction stack method represents a method of migration.

Diffraction stack migration was developed independently by many different oil and geophysical companies in the 1960s. The computational breakthrough was in place in the 1960s, and technical advances soon followed. It was assumed that the CMP stacks represent data with either primary reflections or diffractions, but no reverberations or other types of multiple reflections. Various migration schemes based upon CMP stacked data were developed. They were designated as methods of migration after stack (i.e., poststack migration). Depth migration was based on schemes that allowed velocity functions of the form $v(x, z)$, where the lateral dimension is x and the depth direction is z. However, most of the poststack schemes usually used a velocity function $v(z)$ that depended only on depth. Such a case was called time migration. Poststack time migration usually provided good results when the dip was small. In order to alleviate problems of greater dips, poststack migration was used in conjunction with dip moveout (DMO). Poststack migration was the standard until about 1995.

Phase 3. In the 1990s, large-scale parallel processing made an appearance and data collection became much less expensive due to advances in instrumentation. As a result, three-dimensional velocity functions became cost effective. Depth migration based upon velocity functions of the form $v(x, y, z)$ came into widespread use. Three-dimensional seismic methods became a common exploration and production tool.

The terms geometric acoustics and physical acoustics have the same meaning as the corresponding terms in optics. The same applies to geometric seismology and physical seismology. It might be said that the history of seismic migration follows the history of optics. Phase 1 corresponds to the drawings of rays found on the stone carvings of ancient Egyptians. Phase 2 corresponds to the rays used in geometric optics. Phase 3 corresponds to waves used in physical optics. Although depth migration methods have generally replaced time migration methods, a firm understanding of the time migration technique remains important for comprehending the processes we use and the reasons behind them.

Exploding reflectors model

The paper by Loewenthal et al. (1976) introduces the exploding reflector model that represents the foundation of time migration techniques. In the subsequent years, many other algorithms for migration have been implemented. Some of the early work is that of Bardan (1980), Berkhout (1980), Bleistein and Cohen (1982), Claerbout (1976), Clayton and Stolt (1981), Gazdag (1978), Hubral (1977), McMechan (1983), Robinson (1986), Schneider (1978), Stolt (1978), Weglein (1982), and Yilmaz (1978). This list represents only a sample of the early papers in migration; many hundreds of papers have now been written.

The time migration methods can be described as various ways of implementing the classical WKBJ approximation, named after Wentzel (1926), Kramers (1926), Brillouin (1926), and Jeffreys (1924). The WKBJ method is a classic approximation for problems involving the propagation of waves through an inhomogeneous medium. It is used for obtaining an approximation to the solutions of the one-dimensional time-independent Schrödinger equation, valid when the wavelength of the solution varies slowly with position. The WKBJ method has a long history. In 1838, George Green publishes *On the Motion of Waves in a Variable Canal of Small Width and Depth.* This work of Green anticipates the WKBJ approximation. In 1924, Sir Harold Jeffreys develops the method for use in classical physics. In 1926, Wentzel, Kramers, and Brillouin also develop the method and apply it to quantum mechanics. The exploding reflectors hypothesis depends on an inherent (but, in fact, incidental) assumption that the amplitude of the seismic pulse is invariant as it is transmitted through the earth layers. The work of Gray (1984) shows that this assumption is indeed a consequence of the WKBJ approximation, and so is valid only in those situations where the WKBJ approximation is applicable. The phase term which makes up the essential element of time migration (as well as various other types of migration methods) is the WKBJ phase correction factor, so time migration methods, such as the conventional versions of Kirchhoff migration, finite-difference migration, and frequency-wavenumber (f-k) migration, are not general wave-equation methods but are simply aspects of the WKBJ approximation.

In seismic acquisition, each source is recorded at a number of geophone locations and each geophone location is used to record from a number of source locations. After correcting these data for statics, normal moveout, and DMO (because dipping reflections do not involve a common reflecting point), they are added (stacked) to provide a common-midpoint section that approximates the traces that would be recorded by a coincident source and receiver at each location. Stacking attenuates random effects and reduces the effect of events whose dependence on offset is different from that of primary reflections. When a stacked section is migrated, a migration scheme must be used that is applicable to data recorded with a coincident source and receiver (zero-offset). A zero-offset section could be recorded by moving a single source and a single receiver along the line with no separation between them. The recorded energy follows raypaths that are normal incidence to reflecting interfaces. Because such a recording geometry cannot be realized in practice, the following alternative is employed to produce the same seismic section. Imagine exploding sources that are located along the reflecting interfaces. Also, consider one receiver located on the surface at each CMP location along the line. The sources explode in unison and send out waves that propagate upward. The waves are recorded by the receivers at the surface. The earth model described by this experiment is referred to as the *exploding reflectors model* (Loewenthal et al., 1976).

The seismic section that results from the exploding reflectors model is largely equivalent to the zero-offset section, with one important distinction. The zero-offset section is recorded as two-way traveltime (from source to reflection point to receiver), while the exploding reflectors model is recorded as one-way traveltime (from the reflection point at which the source is located to the receiver).

Virtually all poststack migration methods, including time migration, are based upon the exploding reflectors model. The input data for the migration process consist of a CMP (common midpoint) stacked seismic section. The CMP section can be defined by its amplitude as a function of one-way traveltime t and horizontal coordinate x along a straight surface line. In the usual treatments of poststack migration methods, the effects of the migration on multiple reflections as well as on reflections from locations outside of the vertical plane of the earth cross-section are not discussed. Finally, an inherent (but incidental) assumption in the exploding reflectors model is that the basic seismic pulse is a simple impulse and that its amplitude is invariant as it is transmitted through the earth layers. The purpose of this section is to discuss this inherent assumption in connection with time migration.

In the mathematical approximations, there has always been a question of how to treat the transmission effects. As we have described previously, the inherent assumption in the exploding reflectors hypothesis is that the amplitude of a seismic pulse is unchanged as the pulse is transmitted through the layers. In other words, the inherent assumption in time migration is that there is no transmission loss due to the inhomogeneous earth. In Phase 2, much effort is spent on trying to calculate an accurate amplitude factor. The basic difficulty is that the two-dimensional (x, z)-plane is but a slice of the physical reality of three dimensions (x, y, z). Because so much can happen out of the two-dimensional plane, it is essentially impossible to determine an accurate amplitude factor with two-dimensional data. The solution would have to wait until the use of three-dimensional data in Phase 3.

Therefore, not only in phase but also in amplitude, the conventional theory of time migration rests on the WKBJ method of approximation. However, the question of the applicability of WKBJ theory to the migration of actual seismic data from different geological environments (e.g., slowly varying earth structures, etc.) is a major field of research activity.

Migration

The purpose of reflection seismology is to determine the structure of the subsurface from seismic traces recorded at the surface. The recorded seismic data are subjected to various processing operations by means of digital computers in order to transform the data into a valid picture of cross-sections of the earth which can be interpreted in geological terms. Migration is one of the last operations to be performed on the data. The reason is that many migration methods require that multiple reflections and surface waves be removed, or at least severely attenuated. In addition, migration requires an adequate velocity function of the subsurface. The use of receiver arrays, and the processing operations of CMP stacking and predictive deconvolution, usually provide adequate removal of multiples and surface waves. For any given midpoint, a velocity search procedure provides a velocity function $v(z)$ as a function of depth z. The resulting CMP stacked section is an approximation to a section which would be obtained if each source and its corresponding receiver were at the same point. Thus, in Phase 2, it is the usual practice to perform migration on the CMP stacked sections, making use of the velocity functions obtained by the velocity search procedures.

The definition of terms in migration is important; therefore, at the outset, we give the following generally accepted definitions.

Prestack migration: Any migration process that starts with the unstacked traces as input data.

Poststack migration: Any migration process that starts with the CMP stack as input data. (The material in this chapter is restricted to this type of migration.) It is assumed that the CMP section corresponds to a hypothetical zero-offset (i.e., a coincident source-receiver) experiment in which the exploding reflectors model holds.

Time migration: Poststack migration (as above) in which the depth-point image is put at the minimum of the reflection time. Time migration is strictly valid only for a horizontally stratified medium, for example, one with constant density and vertical velocity variations only.

Depth migration: Any migration (prestack or poststack) that takes into consideration geologic structure that is not necessarily horizontally stratified; in other words, all migration methods that are not time-migration schemes.

WKBJ migration: The WKBJ method is not limited to any seismic processing scheme in general, or to any migration scheme in particular. Even so, in this chapter we only discuss the WKBJ method in relation to time migration and its attendant exploding reflectors hypothesis. However, the WKBJ method is valuable in tracking raypaths in situations that do not involve CMP stacking and exploding reflectors. The WKBJ method is an important tool for geophysicists.

The word migration refers to the movement, or migration, of the observed events on the stacked section to their true spatial positions. The input data to a poststack migration program would consist of a CMP stacked section. Provided that the stacking operation has been successful, the primary reflections on this section approximate what would have been generated in a hypothetical survey composed of a series of source-and-receiver combinations discretely spaced at some constant increment along the horizontal seismic exploration line. Because, in this hypothetical survey, each source and the corresponding receiver is at the same point on the surface of the earth, the offset (or distance) between each source and its receiver is zero. As a result, the CMP stacked section often is described as a zero-offset section. Because, in actuality, the seismic sources are set off sequentially and not simultaneously, the zero-offset data do not represent any wavefield resulting from a single seismic experiment. As a result, Loewenthal et al. (1976) introduce a hypothetical physical experiment to provide an intuitive picture of zero-offset migration: the *exploding reflector model*. In this model, the energy sources are not at the surface of the earth, but they are distributed along the subsurface reflecting interfaces (the reflectors). In other words, the reflectors are represented by buried sources, which are all activated at the same time $t = 0$. Therefore, in the exploding reflectors model, we need to be concerned only with upward traveling waves.

Because the actual CMP section involves two-way traveltime (the time from the surface point to the reflector plus the time back up on the same raypath to the same surface point), seismic time needs to be converted to one-way traveltime when using the exploding reflector model. In this discussion, whenever we are dealing with the exploding reflector hypothesis, we will assume that the conversion has been made so that the variable t represents one-way time. The conversion is completed simply by dividing the two-way time by 2. In seismic data processing centers in practice, however, the time scale of CMP sections is kept unchanged, and instead the velocity of wave propagation is divided by a factor of 2.

Under the exploding reflectors model, the migration of CMP data can be described as the depropagation of the primary reflections recorded at the surface of the earth back in time and down into the earth to their time origin $t = 0$. There is an inherent (but, in fact, incidental) assumption that the basic seismic wavelet is a simple pulse and that its amplitude is invariant as it is transmitted through the layers of the earth.

The CMP stacked section may be considered as a wavefield measured at the surface of the earth. Given the approximate velocity variations within the earth as given by the velocity function, the migration process downward continues this wavefield into the subsurface and thereby elucidates the sources of the reflected and diffracted seismic events. Therefore, migration is the inverse process in which the recorded seismic waves are depropagated (with time running backward) to the corresponding reflector locations. In the process of seismic data acquisition, the upward traveling waves are recorded at the surface of the earth. In migration, these recorded waves, in the form of the stacked section, are used either as boundary conditions or initial conditions for a wavefield governed by the wave equation. Migration is the inverse propagation (or depropagation) process, which pushes these upgoing waves back into the earth in reverse time in order to arrive at the reflector locations.

In Phase 2, there are many implementations of time migration methods. In this chapter, however, we show that all time migration methods are compatible with the amplitude and phase terms of the WKBJ approximation (or, in other words, geometrical acoustics), which has a long history in physics, mathematics, and engineering as well as in geophysics. This categorization of time migration techniques as being consistent with the WKBJ method (i.e., geometric acoustics) has value in that it brings together a large segment of recent research in exploration geophysics with the classical methods used in applied physics, and thus adds to the unity of science.

Wave depropagation simplified

Migration is a term used in reflection seismology to describe the process of moving the recorded reflection events to their correct spatial positions. Migration can be done by depropagating (a.k.a. backtracking) seismic waves. Let us now give a simple example that explains the mathematics.

Ocean waves have wavelengths comparable to the seismic waves used in petroleum exploration, but the velocity of ocean waves is much smaller so that they can be observed easily. Let us imagine a long straight beach which we take as the x-axis. We let the z-axis point directly seaward, with $z = 0$ corresponding to the beach line. For this simple example, we still suppose that the ocean waves are sinusoidal with frequency ω, velocity v, and direction of travel θ, all fixed. The angle θ is measured with respect to the z-axis. We now suppose that someone on a ship at sea radios us on the beach that a large-crested wave has passed the ship at time $t = 0$. We observe that same wave hitting the beach at time $t = t_0$. The question is: what is the location (range and bearing) of the ship?

The part of the question as to the range is easy. The range is $R = vt_0$, so the ship can be anywhere on a circle of radius R and center at our position on the beach. Let us now make our analogy with oil exploration. We think of the beach as the surface of the ground, and think of the ocean as the subsurface geologic rock structure. The boat with unknown position corresponds to an unknown oil reservoir we are exploring. The ocean waves correspond to the seismic waves. By auxiliary means, we can find the seismic wave velocity v, and we can measure the arrival time t_0 (in this case, one-way time from depth to the surface of the earth) of the seismic wavelet due to the oil reservoir. Thus, we immediately can determine the range $R = vt_0$ of the oil reservoir.

We thus know the range of the ship or the oil reservoir. With no other information, we cannot determine its bearing, so the bearing angle could be anywhere from -90° to 90°. The average value could be zero, so we could guess that the ship or oil reservoir has made an angle of zero with the z-axis; that is, the ship or reservoir is at right angles to the beach or earth's surface from our observation position. An unmigrated seismic record makes this assumption. It puts the cause of each event directly under the position on the earth's surface where this event is observed. In other words, an unmigrated seismic record always draws each bearing as straight down into the earth.

The reasoning for performing the data processing operation of migration is to compute the true bearing for each event, and then put the cause of the event at the computed range in the direction of the computed bearing. Of course, if all geologic rock layers are flat and horizontal, then indeed the seismic waves from an exploding reflector model go straight up (each with a bearing angle of zero degrees). In such a case, an unmigrated seismic section serves perfectly well. It also serves well provided the dips of the layers are all small and random. However, when there are many small dips all going in the same direction, or when there are some large dips as in the overthrust belt of the Rocky Mountains, or a combination of both, then an unmigrated seismic record section does not serve exploration well.

In order to find the bearing of the ship or the oil reservoir, we must measure some additional quantities. In Figure 2, we see the necessary relationships. On the beach, we need a timepiece and a measuring stick, and we must take measurements at two or more stations on shore. With the timepiece, we measure the time between two consecutive crests of the wave at a given station. This time measurement is the wave period T. With the meter stick, we measure the distance between two stations where adjacent crests hit the beach at the same instant. This distance measurement is horizontal (x-coordinate) wavelength λ_x.

The word "bearing" refers to direction, especially angular direction measured from one position to another using geographical or celestial reference lines. With the two measurements T and λ_x, we can determine the angle θ of bearing as follows. First, we must determine the wavelength of the wave. The wavelength λ is given by $\lambda = vT$; that is, the wavelength is

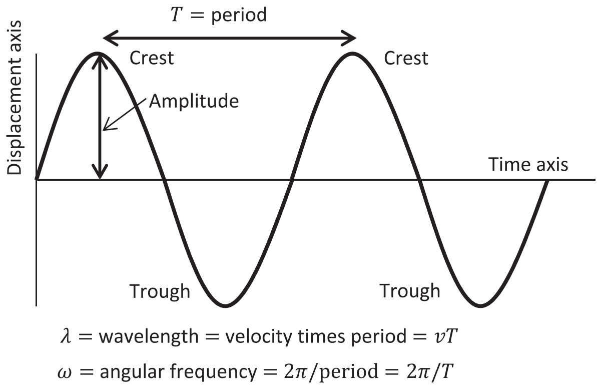


Figure 2. Sinusoidal plane wave.

equal to the distance a crest travels during the elapsed time of one period. The angle θ of bearing is given by

$$\sin\theta = \frac{\lambda}{\lambda_x} \ . \tag{1}$$

The same principle applies to all migration schemes. In effect, we find the bearing angle θ and then backtrack along this bearing by letting time run backwards from the arrival time t_0 to the source time 0. When we reach time 0, we know we have reached the source, and the total distance that we have backtracked is equal to the range $R = vt_0$. Thus, we locate the source of the event shown on the seismic record section. This process is seismic migration, and it involves the depropagation of the seismic waves observed at the surface of the earth.

Migration involves many seismic waves coming from different directions and various arrival times. In order to make things mathematically tractable, we appeal to the power of the Fourier transform. When we do things in the frequency domain instead of the time domain, we use spatial frequencies (also called *wavenumbers*) instead of wavelengths, and we use temporal frequencies (simply called *frequencies*) instead of wave periods. The well known relationships are

$$k = \frac{2\pi}{\lambda}, \quad k_x = \frac{2\pi}{\lambda_x}, \quad \omega = \frac{2\pi}{T} \ , \tag{2}$$

where k is called the wavenumber, k_x the horizontal wave number, and ω the frequency. We also can define the vertical wavenumber k_z by the equation

$$k_z^2 = k^2 - k_x^2 \ . \tag{3}$$

This equation says that k_x and k_z are the sides of a right triangle with hypotenuse k. The wavenumber k is equal to

$$k = \frac{2\pi}{\lambda} = \frac{2\pi}{vT} = \frac{\omega}{v} \ . \tag{4}$$

We define the propagation vector $\boldsymbol{k}$ pointing in the direction of the wave (i.e., $\boldsymbol{k}$ makes an angle θ with the z-axis) where $\boldsymbol{k}$ has length k and components

$$k_x = k \sin\theta, \quad k_z = k\cos\theta \ . \tag{5}$$

The seismic disturbance (wave motion) at any point (x, z) at any time t may be denoted by the symbol $u(x, z, t)$. The surface of the earth is given by depth $z = 0$, so the wave motion which we measure at the receivers on the ground is $u(x, 0, t)$. We can compute the two-dimensional Fourier transform of the observed wave motion $u(x, 0, t)$ with respect to x and t to obtain the surface wavefield spectrum

$$U(k_x, 0, \omega) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(x, 0, t)\exp[-i(\omega t - k_x x)]dx\, dt \ . \tag{6}$$

Our purpose is to take the wave motion associated with the sinusoidal wave characterized by k_x, ω, and then depropagate this sinusoidal wave in the direction θ as determined by

$$\sin\theta = \frac{k_x}{k} = \frac{k_x v}{\omega} \ .$$

The depropagation terminates when we reach the distance given by t_0.

The implementation of this depropagation scheme in the frequency domain is done as follows. The travel time t has differential

$$dt = \frac{\partial t}{\partial x}dx + \frac{\partial t}{\partial z}dz \ .$$

The derivative dx/dt is the horizontal apparent velocity, so

$$\lambda_x = \frac{dx}{dt}T \ .$$

As a result, we have

$$\frac{dt}{dx} = \frac{T}{\lambda_x} = \frac{2\pi}{\lambda_x}\frac{T}{2\pi} = \frac{k_x}{\omega} \ .$$

Likewise,

$$\frac{dt}{dz} = \frac{k_z}{\omega} \ .$$

Thus, the time differential is

$$dt = \frac{k_x}{\omega} dx + \frac{k_z}{\omega} dz \ . \tag{7}$$

When we depropagate by a time span of t_0, we let time run backward. This means, in engineering terms, that we must introduce a time advance of t_0. A time-advance operator is the pure phase-shift system given in the frequency domain by $\exp(i\omega t_0)$. Let the depropagation path be from the receiver point (x, $z = 0$) on the surface of the earth to the source point at a depth given by the point ($x = 0$, z). Here, we assume that time t_0 is the one-way time from source to receiver. Thus, the time advance is equal to

$$\int_0^{t_0} dt = \frac{1}{\omega}\int_x^0 k_x \, dx + \frac{1}{\omega}\int_0^z k_z \, dz \ ,$$

which (for the constant-velocity medium treated here) is

$$\omega t_0 = -k_x x + k_z z \ .$$

The observed seismic wave motion has sinusoidal component

$$U(k_x, 0, \omega) \ \exp(i\omega t) \ .$$

We multiply this component by the phase-shift (pure advance) filter $\exp(i\omega t_0)$ to obtain

$$U(k_x, 0, \omega) \ \exp[i\omega(t + t_0)] = U(k_x, 0, \omega) \ \exp[i(\omega t - k_x x + k_z z)] \ .$$

This expression gives the depropagating sinusoidal wave. We thus integrate this expression over k_x and ω to obtain the depropagating wave.

We remember that k_z is not an independent variable, but is given by the positive square root

$$k_z = +\sqrt{k^2 - k_x^2} = +\sqrt{\left(\frac{\omega}{v}\right)^2 - k_x^2} \ . \tag{8}$$

In fact, it is this link of k_z to k_x and ω that makes the operation of depropagation possible. Thus, the required equation for the *depropagating wave* is the integral

$$u(x, z, t) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U(k_x, 0, \omega) \exp\left[i\sqrt{\left(\frac{\omega}{v}\right)^2 - k_x^2}\, z \right] \times \exp[i(\omega t - k_x x)] dk_x \, d\omega \ . \tag{9}$$

This integral is the inverse Fourier transform of

$$U(k_x, 0, \omega) \exp\left[i\sqrt{\left(\frac{\omega}{v}\right)^2 - k_x^2}\, z \right] .$$

Thus, depropagation is achieved by multiplying the surface wavefield spectrum by the filter

$$\exp\left[i\sqrt{\left(\frac{\omega}{v}\right)^2 - k_x^2}\, z \right] . \tag{10}$$

It is for this reason that this filter is called the *depropagation* (or *migration*) filter.

Thus, we have found the wavefield $u(x, z, t)$ at an arbitrary space–time point (x, z, t) by depropagation. In other words, we have found the correct bearings. This first step of the depropagation process is called *wavefield reconstruction*. The second step of the depropagation process involves stopping at the correct range. We recall that the signal originated at the source at time $t = 0$. Thus, we set $t = 0$ in equation 9 for the depropagating wave in order to obtain the final answer; namely,

$$u(x, z, 0) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U(k_x, 0, \omega) \exp\left[i\sqrt{\left(\frac{\omega}{v}\right)^2 - k_x^2}\, z \right] \times \exp[-ik_x x] dk_x \, d\omega \ . \tag{11}$$

This second step is called *imaging*, as it gives the required sources.

To this point in our discussion, we have assumed that the velocity v is constant. A typical geophysical assumption, however, is the stratified earth assumption in which we assume that v varies in the depth direction but not in the horizontal direction, so we write $v(z)$ indicating that the velocity is a function of depth. In the stratified case, the depropagation operator 10 becomes

$$\exp\left[i \int_0^z k_z(z)\, dz \right] = \exp\left[i \int_0^z \sqrt{\left(\frac{\omega}{v(z)}\right)^2 - k_x^2}\, dz \right] . \tag{12}$$

The depropagation method given above is known as *frequency-wavenumber* (or *f-k*) *migration*.

Time migration

Wave depropagation, or migration, as used in seismic reflection exploration, requires that surface waves and multiple reflections be adequately attenuated, and, in addition, that an adequate velocity function of the subsurface be supplied by other means. Seismic arrays, CMP stacking, and deconvolution usually attenuate surface waves and multiple reflections. Velocity analysis methods provide a velocity function $v(z)$ for each midpoint analyzed, and any given velocity function is assumed to hold in a certain horizontal range surrounding the midpoint in question. Moreover, the CMP stacked section is an approximation to a source-receiver coincident section (zero-offset section), so that the exploding reflector hypothesis can be used. Thus, it is common practice to migrate CMP stacked sections into the reflectivity function of the subsurface. The following analysis illustrates the mathematical steps that are used to justify time migration.

As usual, we let x represent the coordinate along the surface of the earth, and z represent the coordinate of depth into the earth, with $z = 0$ denoting the surface and z being measured positively down. As often done in theoretical studies in seismic exploration, we shall assume that density ρ is constant. Also in this chapter, we consider only two spatial dimensions x and z. The wavefield $u(x, z, t)$ satisfies the two-dimensional scalar wave equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial z^2} - \frac{1}{v^2}\frac{\partial^2 u}{\partial t^2} = 0 . \tag{13}$$

The theory of time migration is based on the *stratified earth assumption*, namely, that there is no lateral velocity variation, so the velocity v is a function $v(z)$ of depth only. The wavenumber function

$$k(z) = \frac{\omega}{v(z)} \tag{14}$$

characterizes the stratified earth assumption. Because variation only occurs in the z-direction, we are motivated to take the Fourier transform of the wave disturbance $u(x, z, t)$ with respect to each variable except z. Thus, we form the Fourier transform with respect to x and t, namely,

$$U(k_x, z, \omega) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(x, z, t) \exp[-i(\omega t - k_x x)] dx\, dt \ . \tag{15}$$

In the wave equation, the second partial derivatives with respect to each of the variables x, z, and t occur. Our intention is to take the Fourier transform of the wave equation, so we need the Fourier transforms of these derivatives. The z-derivative is easy, as we can take the $\partial^2/\partial z^2$ outside the integral sign. Thus, the Fourier transform (F.T.) is

$$\text{F.T. of } \frac{\partial^2 u}{\partial z^2} = \frac{\partial^2 U}{\partial z^2} \ .$$

The other two derivatives can be found by the well known property of the Fourier transform, namely, that

$$\text{F.T. of } \frac{\partial^2 u}{\partial x^2} = (-ik_x)^2\, U$$

and

$$\text{F.T. of } \frac{\partial^2 u}{\partial t^2} = \frac{(i\omega)^2}{v^2}\, U \ .$$

Therefore, the Fourier transform of wave equation 13 is

$$(-ik_x)^2\, U + \frac{\partial^2 U}{\partial z^2} = \frac{(i\omega)^2}{v^2}\, U \ . \tag{16}$$

Because variables x and t have been eliminated, the partial derivative with respect to z becomes a full derivative, and equation 16 becomes

$$\frac{d^2 U}{dz^2} = \left(\frac{\omega^2}{v^2} - k_x^2 \right) U = 0 \ . \tag{17}$$

Because $v(z)$ varies with z, this equation is a second-order ordinary differential equation with a variable coefficient. There is no exact solution, but an approximate solution can be found as follows.

The coefficient in ordinary differential equation 17 is denoted by $k_z^2(z)$; that is,

$$k_z^2(z) = \frac{\omega^2}{v^2(z)} - k_x^2 = k^2(z) - k_x^2 \ . \tag{18}$$

Both ω and k_x are constants in this equation. One of two cases must occur, namely, $k_z^2(z)$ is either positive or negative. The case of negative $k_z^2(z)$ produces evanescent waves, whereas the case of positive $k_z^2(z)$ produces traveling waves. Here, we will only treat the positive case, so we can define $k_z(z)$ as the positive square root given by

$$k_z(z) = +\sqrt{k^2(z) - k_x^2} \ . \tag{19}$$

We can write ordinary differential equation 17 as

$$\frac{d^2U}{dz^2} + \left(\frac{\omega}{v}\right)^2 R^2\, U = 0 \ , \tag{20}$$

where R is defined as the positive square root

$$R = +\sqrt{1 - \left(\frac{vk_x}{\omega}\right)^2} \tag{21}$$

and $k_z = (\omega/v)R$. Under the assumption that $v(z)$ varies with z, it follows that $R(z)$ also varies with z, and in such a case there is no closed-form solution of differential equation 20. An approximate solution can be found, however, by assuming that $v(z)$ and hence $R(z)$ are constants, and thus the resulting constant-coefficient differential equation 20 can be factored to obtain

$$\left(\frac{d}{dz} + i\frac{\omega}{v}R\right)\left(\frac{d}{dz} - i\frac{\omega}{v}R\right) U = 0 \ .$$

This factorization yields the two first-order differential equations

$$\frac{dU}{dz} + i\frac{\omega}{v}RU = 0 \text{ (downgoing equation) }, \tag{22}$$

$$\frac{dU}{dz} - i\frac{\omega}{v}RU = 0 \text{ (upgoing equation) }. \tag{23}$$

Now, the assumption that $v(z)$ and $R(z)$ are constants is dropped, but the two first-order equations 22 and 23 are retained. Thus, in the given approximate method, the scalar wave equation is replaced by two first-order equations 22 and 23, the first of which is called the *downgoing (one-way) wave equation*, and the second the *upgoing (one-way) wave equation*. Neither of these two equations admits multiple reflections. Because the full wave equation is a second-order equation, two boundary conditions are required for its solution, such as the values of U and dU/dz on a surface. In seismic practice, however, we do not record dU/dz. The use of the full wave equation produces multiple reflections, which complicates the migration problem to a great extent. There are many ingenious methods that use the full wave equation. On the other hand, the one-way equations do provide a viable way to do seismic migration. A one-way equation requires only a single boundary condition, it has stable numerical solutions, and it can produce no multiple reflections.

In summary, ordinary differential equation 20 with a variable coefficient has no closed solution. As an approximation, the coefficient is assumed constant, then the equation is factored, and downgoing equation 22 and upgoing equation 23 result. Then, the approximation consists of letting the coefficient again be variable in these two equations. This procedure gives a crude WKBJ approximation (i.e., the phase term but not the amplitude term).

Let us consider upgoing (one-way) wave equation 23, which we write as

$$\frac{dU}{dz} - ik_z(z)\, U = 0 \ , \tag{24}$$

where k_z is defined by

$$k_z(z) = \frac{\omega}{v}R = \frac{\omega}{v(z)}\sqrt{1 - v(z)\frac{k_x}{\omega}} \ . \tag{25}$$

This one-way equation is a first-order linear equation, a type of equation considered near the beginning of every book on differential equations.

In these text books, it is shown that the solution is of the form

$$U = A \exp\left[i \int k_z(z)\, dz \right] , \tag{26}$$

where A is a constant (with respect to z). We can easily verify that this form is indeed a solution by substituting it into the differential equation. We obtain the expression

$$\frac{d}{dz}\left\{ A \exp\left[i \int k_z(z)\, dz \right] \right\} - i\, k_z(z) A \exp\left[i \int k_z(z)\, dz \right] .$$

Carrying out the differentiation, this expression becomes

$$A \exp\left[i \int k_z(z)\, dz \right] \frac{d}{dz}\left[i \int k_z(z)\, dz \right] - i\, k_z(z) A \exp\left[i \int k_z(z)\, dz \right] .$$

Because

$$\frac{d}{dz}\left[i \int k_z(z)\, dz \right] = k_z(z) ,$$

the expression is zero. Thus, we have verified that the solution is of the form given above.

The constant A in the solution is determined by an initial condition. In the classic migration problem, we assume that we have an exploding-reflectors source given by $u(x, z, t = 0)$ which produces upgoing waves that appear on the earth's surface as the seismic section $u(x, z = 0, t)$. Given that the seismic section is known, the migration problem consists of finding the exploding-reflectors source. Let us now solve the classic migration problem using the one-way upgoing wave equation.

We compute the Fourier transform of the seismic section, that is, we compute

$$U(k_x, 0, \omega) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(x, 0, t) \exp[-i(\omega t - k_x x)] dx\, dt .$$

We then use $U(k_x, 0, \omega)$ as the boundary condition in the solution of the differential equation. Thus, the constant A is $U(k_x, 0, \omega)$ and the solution becomes

$$U(k_x, z, \omega) = U(k_x, 0, \omega)\ \exp\left[i \int_0^z k_z(z)\, dz \right] . \tag{27}$$

We now take the inverse Fourier transform

$$u(x, z, t) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U(k_x, z, \omega) \exp[i(\omega t - k_x x)] dk_x\, d\omega \ ,$$

which, upon using the above solution 27 of the differential equation, is

$$u(x, z, t) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U(k_x, 0, \omega)\ \exp\left\{ i \left[\omega t - k_x x + \int_0^z k_z(z)\, dz \right] \right\} dk_x\, d\omega \ . \tag{28}$$

Equation 28 represents the required wavefield reconstruction. The required exploding-reflectors source is obtained by setting $t = 0$ in equation 28; that is,

$$u(x, z, 0) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U(k_x, 0, \omega) \exp\left\{ i \left[-k_x x + \int_0^z k_z(z)\, dz \right] \right\} dk_x\, d\omega \ . \tag{29}$$

Equation 29 represents the required imaging. For computational purposes, we rearrange this double integral as

$$u(x, z, 0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} U(k_x, 0, \omega)\ \exp\left[i \int_0^z k_z(z)\, dz \right] d\omega \right\}$$
$$\times\ \exp(-ik_x x) dk_x \ . \tag{30}$$

We see that the expression in curly brackets merely involves integrating $U(k_x, z, \omega)$, given by equation 27, over ω. Then, we only have to take the single inverse Fourier transform with respect to k_x to obtain the required $u(x, z, 0)$.

In the case of constant velocity v, the vertical wavenumber $k_z(z)$ reduces to the constant

$$k_z(z) = \sqrt{\left(\frac{\omega}{v}\right)^2 - k_x^2} \ .$$

In this case, we have

$$\int_0^z k_z(z)dz = k_z \int_0^z dz = k_z z \ .$$

Hence, equation 28 reduces to equation 9 and equation 29 reduces to equation 11. Thus, we have obtained the same result as we obtained in our simplified treatment of wave depropagation.

WKBJ migration

The following interpretation can be given to the WKBJ approximation. The WKBJ approximation for a wave in an inhomogeneous space represents the primary wave traveling by refractions through the medium. The entire wave motion is composed of all of the internal reflections and refractions within the medium. This complete wave motion can be represented by an infinite series, called the Bremmer (1951) series, each term of which represents a wave that is produced by a particular number of reflections inside the medium. The Bremmer series is only valid for large values of k_z. With this model, the WKBJ approximation is the first term in the series, that is, it is the primary wave that has suffered no internal reflections. In the language of reflection seismology, the WKBJ wave is the primary, i.e., the WKBJ wave is not a multiple. The WKBJ wave represents the high-frequency limit in which the velocity changes are too small to produce reflections; in this limit, there is no provision for transmission loss.

In all of the typical migration processes used in the seismic industry, it is assumed that all of the multiples have been removed in advance by stacking and deconvolution. As a result, each of the usual industrial migration processes deals only with traveling primary waves. In this sense, all industrial migration processes have a relationship to the WKBJ method (i.e., geometric acoustics). Thus, unlike predictive deconvolution, which was developed explicitly for exploration seismology, migration does not involve a new model but is simply a straightforward application of classical wave-propagation techniques.

Much of the seismic data processing currently in use rests on a basic assumption known as the *stratified earth hypothesis*. This hypothesis states that, for the purposes of wave propagation and wave depropagation, the velocity function depends only upon the depth coordinate z. Thus, we can write the velocity function as $v(z)$. This assumption means that the velocity is the same on any horizontal plane, so the earth is considered a stratification of these horizontal planes. The velocity function $v(z)$ can either vary continuously in z or in discrete steps. For computer processing, however, some discrete approximation is always made, so the earth appears as a sequence of horizontal layers each of finite thickness. Such a model of discrete flat horizontal layers is known as the layer-cake model.

Let us now describe the application of the WKBJ method to seismic time migration. As we will show, the result for the phase change is exactly the same result as that given in the previous section, *Time Migration*. In the derivation which follows, we use the classic treatment of the WKBJ method. When $k_z(z)$ is a constant, the solution of differential equation 17 is

$$U = A \exp(\pm i k_z z) \ , \tag{31}$$

where A is a constant and the choice of sign in the exponent depends upon the direction of the wave. With the motivation of equation 31, let us write the solution of equation 17 in the form

$$U = A(z) \exp[i\theta(z)] \ , \tag{32}$$

in the case when $k_z(z)$ is a variable but changes sufficiently slowly. Equation 32 is the typical high-frequency assumption; decoupling of upward and downward propagating waves is assumed as in one-way equations 22 and 23. The functions $A(z)$ and $d\theta/dz$ also will be slowly varying, as we see when we compare equations 32 and 31.

Substituting equation 32 into equation 17, we obtain

$$A'' + [2i\theta' A' + i\theta'' A] + [k_z^2(z) - (\theta')^2]\, A = 0 \ ,$$

where the prime represents differentiation with respect to z. Let d be the distance over which the function $A(z)$ and $\theta(z)$ vary significantly. Sufficiently slow variation means that d is much greater than the wave length λ. The order of magnitude of the derivatives can be estimated by

$$A' \approx \frac{A}{d}, \quad A'' \approx \frac{A'}{d}, \quad \theta'' = \frac{\theta'}{d} \ . \tag{33}$$

Also,

$$k_z \approx k = \frac{2\pi}{\lambda} \ .$$

Using these relations, the terms in equation 33 have the estimates

$$A'' \approx \frac{A}{d^2} \ , \tag{34}$$

$$[2i\theta' A' + i\theta'' A] \approx 3j \, \frac{\theta' A}{d} \ , \tag{35}$$

$$[k_z^2(z) - (\theta')^2] \, A \approx \frac{4\pi^2}{\lambda^2} A \ . \tag{36}$$

Note that, because θ' is itself slowly varying, $(\theta')^2$ is a second-order term. Because, for large d, term 34 is small in comparison with terms 35 and 36, it can be neglected. However, because terms 35 and 36 have different orders of smallness for large d, they must be set equal to zero separately. Setting the left side of equation 36 equal to zero, we obtain the *eikonal equation*

$$(\theta')^2 = k_z^2(z) \ ,$$

which has the solution

$$\theta' = \pm k_z(z) \ . \tag{37}$$

Let z_s denote the depth of the pulse source and z_r denote the depth of the received pulse. For a downgoing pulse, we would have $z_s < z_r$, whereas, for an upgoing pulse, we would have $z_s > z_r$. If we integrate equation 37 from z_s to z_r, we obtain

$$\theta = \pm \int_{z_s}^{z_r} k_z(z) \, dz \ . \tag{38}$$

Substituting $\pm k_z$ for θ' in the left side of equation 35 and setting the result equal to zero, we obtain the equation, known as the *first transport equation*,

$$\pm [2k_z A' + k_z' A] = 0 \ .$$

This equation can be written as

$$\frac{A'}{A} = -\frac{1}{2}\frac{k_z'}{k_z} \ . \tag{39}$$

If we integrate equation 39 from z_s to z_r, we obtain

$$\log\frac{A(z_r)}{A(z_s)} = -\frac{1}{2}\log\frac{k_z(z_r)}{k_z(z_s)} \ ,$$

which is

$$\frac{A(z_r)}{A(z_s)} = \sqrt{\frac{k_z(z_s)}{k_z(z_r)}} \ . \tag{40}$$

If we normalize the pulse by requiring $A(z_s) = 1$, then the complete solution of equation 17 is obtained by substituting equations 38 and 40 into equation 32. The result (with $k_z(z) > 0$) is

$$U = \sqrt{\frac{k_z(z_s)}{k_z(z_r)}}\ \exp\left[\pm i\int_{z_s}^{z_r} k_z(z)\,dz\right] \ . \tag{41}$$

In order to restore wave motion, we must multiply equation 41 by the factor

$$\exp[(\omega t - k_x x)] \ .$$

Then, if the positive sign in the exponent of equation 41 is chosen, we would have a wave propagation in the negative z-direction (i.e., upgoing waves); whereas, if the negative sign is chosen, we would have a wave propagating in the positive z-direction (i.e., downgoing waves). These two waves (i.e., upgoing and downgoing) propagate through the medium independently of one another (as, in advance, assumed by equation 32), as there are no reflections within the present approximation. This approximation is known as the *WKBJ approximation*. It also is called the *geometric optics approximation* in optics and the *geometric acoustics approximation* in acoustics. The term

$$\int_{z_s}^{z_r} k_z(z)\,dz$$

in the exponent of equation 41 is the phase change as the wave travels from an arbitrary point z_s to the point of observation z_r. The z-dependence of the wave amplitude in equation 41 is given by the factor, called the *WKBJ amplitude factor*,

$$\sqrt{\frac{k_z(z_s)}{k_z(z_r)}} \, . \tag{42}$$

The WKBJ amplitude factor is an important result of the WKBJ solution. However, the amplitude factor is neglected in many of the conventional methods of poststack seismic migration, so we will not consider it further here.

Under the exploding reflector hypothesis, the source is considered to be at the subsurface interface at depth z and the receiver on the surface at depth $z = 0$. Thus, we have upgoing wave motion, so we choose the positive sign in the exponent of equation 41. Equation 41, with $z_s = 0$ and $z_r = 0$, with the positive sign in the exponent and without the amplitude factor, becomes (with $k_z(z) > 0$)

$$U_{\mathrm{WKBJ}} = A \exp\left[i \int_z^0 k_z(z)\, dz \right]$$

or

$$U_{\mathrm{WKBJ}} = A \exp\left[-i \int_0^z k_z(z)\, dz \right] . \tag{43}$$

We call this expression the *WKBJ operator* for the propagation of upgoing waves under the exploding reflector hypothesis. Because $u(x, z, t)$ denotes the wavefield, we see that $u(x, 0, t)$ denotes the wavefield at the surface $z = 0$ of the earth. The Fourier transform of $u(x, z, t)$ with respect to x and t is $U(k_x, z, w)$. We have not transformed it with respect to z because the wave velocity $v(z)$ varies in the z-direction. The inverse Fourier transform is then

$$u(x, z, t) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U(k_x, z, \omega) \; \exp\{i[\omega t - k_x x]\} dk_x \, d\omega \, .$$

Now, suppose we know the upgoing wavefield $u(x, 0, t)$ at the surface $z = 0$ of the earth, and suppose that we wish to use the WKBJ method to find the upgoing wavefield $u(x, z, t)$ depropagated to depth z. We proceed as follows. First, we compute the Fourier transform $U(k_x, 0, \omega)$ of the surface wavefield $u(x, 0, t)$. Then, we multiply this surface Fourier transform by the inverse of the WKBJ filter U_{WKBJ} given by equation 43. This inverse is simply the reciprocal, namely,

$$U_{\mathrm{WKBJ}}^{-1} = \frac{1}{U_{\mathrm{WKBJ}}} = \exp\left[i \int_0^z k_z(z)\, dz \right] , \tag{44}$$

where

$$k_z(z) = +\sqrt{\left(\frac{\omega}{v(z)}\right)^2 - k_x^2} \; . \tag{45}$$

Thus, the result of multiplying $U(k_x, 0, \omega)$ with U_{WKBJ}^{-1} is

$$U(k_x, z, \omega) = U(k_x, 0, \omega)\ \exp\left[i \int_0^z k_z(z)\, dz \right] . \tag{46}$$

The inverse WKBJ operator, as given by equation 44, is an inverse all-pass filter in the region where $k_z(z)$ is real. Here, we use the term "all-pass" to designate a causal filter that produces a pure phase shift, and the term "inverse all-pass" to designate an anticausal pure phase-shift filter. Now, the required solution is obtained by taking the inverse Fourier transform of equation 46; that is, the depropagated wave motion at depth z, according to the seismic WKBJ approximation, is

$$u(x, z, t) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U(k_x, 0, \omega) \exp\left[i \int_0^z k_z(z)\, dz \right] \exp\{ i[\omega t - k_x x]\} dk_x\, d\omega \, . \tag{47}$$

This is the WKBJ equation (without the amplitude factor) for obtaining the *upgoing wavefield depropagated to depth* z *from the observed upgoing wavefield at the surface*. Because equation 47 is the same as equation 28 in the *Time Migration* section, we therefore see that WKBJ migration as given here is the same as time migration, as we wished to show. The same type of argument also can be applied to other types of migration methods to show their relationship to the WKBJ method (i.e., geometric acoustics).

In a uniform medium, the phase of the depropagated wave is

$$\omega t - k_x x + k_z z \; .$$

In a stratified medium, the phase is

$$\omega t - k_x x + \int_0^z k_z(z)\, dz \ ,$$

as the phase is simply the number of cycles undergone by the wave along its depropagation path from the surface $z = 0$ to depth z in the earth.

Time migration is based on the assumption that there is no lateral velocity variation, so the velocity $v(z)$ is a function of depth only. Depth migration is based upon three-dimensional velocity functions $v(x, y, z)$. In this book, we have applied the WKBJ method only to velocity functions of the form $v(z)$. Even so, the WKBJ method can be applied to more general velocity functions $v(x, y, z)$ as well.

This page has been intentionally left blank

References

Aki, K., and P. G. Richards, 1980, Quantitative seismology, Theory and methods, v. 1. Chapter 5: W. H. Freeman and Co.

Anstey, N. A., 1977, Seismic interpretation: The physical aspects: International Human Resources Development Corp. Press.

Bartholinus, E., 1669, Experimenta crystalli Islandici disdiaclastici quibus mira & insolita refractio detegitur (Experiments with the double refracting Iceland crystal which led to the discovery of a marvelous and strange refraction).

Bardan, V., 1980, Despre Migrarea Sectiunilor Seismice Prin Utilizarea Ecuatiei Scalare a Undelor, St. Cerc. Geol., Geofiz., Geogr: Geofizica, **18**, 41–59.

Berkhout, A. J., 1980, Seismic migration: Imaging of acoustic energy by wavefield extrapolation: Elsevier.

Berkhout, A. J., 1987, Applied seismic wave theory: Elsevier.

Bleistein, N., and J. K. Cohen, 1982, The velocity inversion problem – Present status, new directions: Geophysics, **47**, no. 11, 1497–1511, http://dx.doi.org/10.1190/1.1441300.

Bremmer, H., 1951, The WKB approximation as the first term of a geometrical-optical series: Symposium on the theory of electromagnetic waves, Interscience Publishers, 169–179.

Brillouin, L., 1926, La mécanique ondulatoire de Schrödinger: une méthode générale de resolution par approximations successives: Comptes Rendus de l'Academie des Sciences, **183**, 24–26.

Claerbout, J., 1976, Fundamentals of geophysical data processing: McGraw-Hill.

Clayton, R. W., and R. H. Stolt, 1981, A Born-WBKJ inversion method for acoustic reflection data: Geophysics, **46**, no. 11, 1559–1567, http://dx.doi.org/10.1190/1.1441162.

Gazdag, J., 1978, Wave equation migration with the phase shift method: Geophysics, **43**, no. 7, 1342–1351, http://dx.doi.org/10.1190/1.1440899.

Gray, S. H., 1984, A problem of discrete approximations to an arbitrarily varying one-dimensional seismic model: Geophysical Journal of the Royal Astronomical Society, **78**, no. 2, 431–437, http://dx.doi.org/10.1111/j.1365-246X.1984.tb01958.x.

Hagedoorn, J. C., 1954, A process of seismic reflection interpretation: Geophysical Prospecting, **2**, no. 2, 85–127, http://dx.doi.org/10.1111/j.1365-2478.1954.tb01281.x.

Hirshfield, A., 2006, The electric life of Michael Faraday: Walker Books.

Hubral, P., 1977, Time migration, some ray theoretical aspects: Geophysical Prospecting, **25**, no. 4, 738–745, http://dx.doi.org/10.1111/j.1365-2478.1977.tb01200.x.

Jeffreys, H., 1924, On certain approximate solutions of linear differential equations of the second order: Proceedings of the London Mathematical Society, **23**, Part 2, 428–436.

Klem-Musatov, K., H. Hoeber, T. Moser, and M. Pelissier, eds., 2016a, Classical and Modern Diffraction Theory: SEG.

Klem-Musatov, K., H. Hoeber, T. Moser, and M. Pelissier, eds., 2016b, Seismic Diffraction: SEG.

Kramers, H. A., 1926, Wellenmechanik und halbzahlige Quantisierung: Zeitschrift für Physik, **39**, no. 10–11, 828–840, http://dx.doi.org/10.1007/BF01451751.

Loewenthal, D., L. Lu, R. Roberson, and J. Sherwood, 1976, The wave equation applied to migration: Geophysical Prospecting, **24**, no. 2, 380–399, http://dx.doi.org/10.1111/j.1365-2478.1976.tb00934.x.

McMechan, G. A., 1983, Migration by extrapolation of time dependent boundary values: Geophysical Prospecting, **31**, no. 3, 413–420, http://dx.doi.org/10.1111/j.1365-2478.1983.tb01060.x.

Robinson, E. A., 1966, Multichannel *z*-transforms and minimum-delay: Geophysics, **31**, no. 3, 482–500, http://dx.doi.org/10.1190/1.1439788.

Robinson, E. A., 1975, Dynamic predictive deconvolution: Geophysical Prospecting, **23**, no. 4, 779–797, http://dx.doi.org/10.1111/j.1365-2478.1975.tb01558.x.

Robinson, E. A., 1984, Seismic inversion and deconvolution, Part A. Classical methods. Pergamon (An imprint of Elsevier).

Robinson, E. A., 1986, Migration of seismic data by the WKBJ method: Proceedings of the IEEE, **74**, no. 3, 428–439, http://dx.doi.org/10.1109/PROC.1986.13484.

Robinson, E. A., 1999, Seismic inversion and deconvolution: Part B. Dual sensor technology, Pergamon (An imprint of Elsevier).

Schneider, W. A., 1978, Integral formulation for migration in two and three dimensions: Geophysics, **43**, no. 1, 49–76, http://dx.doi.org/10.1190/1.1440828.

Simpson, T. K., 1997, Maxwell on the electromagnetic field: A guided study: Rutgers University Press.

Sheriff, R. E., 2002, Encyclopedic dictionary of applied geophysics, 4th ed.: SEG.

Stolt, R., 1978, Migration by Fourier transform: Geophysics, **43**, no. 1, 23–48, http://dx.doi.org/10.1190/1.1440826.

Trorey, A. W., 1977, Diffractions for arbitrary source-receiver locations: Geophysics, **42**, no. 6, 1177–1182, http://dx.doi.org/10.1190/1.1440782.

Weglein, A. B., 1982, Multidimensional seismic analysis, migration and inversion: Geoexploration, **20**, no. 1–2, 47–60, http://dx.doi.org/10.1016/0016-7142(82)90006-0.

Wentzel, G., 1926, Eine Verallgemeinerung der Quantenbedingungen für die Zwecke der Wellenmechanik: Zeitschrift für Physik, **38**, no. 6–7, 518–529, http://dx.doi.org/10.1007/BF01397171.

Yilmaz, Ö., 1979, Prestack partial migration, Ph.D. thesis, Stanford University.

This page has been intentionally left blank

Index

A

B

C

D

I

N

O

P

Q

R

Y

Z