A noncalculus proof that Fermat's principle of least time implies the law of refraction

Harald Helfgott[a)]
Department of Mathematics, Princeton University, Princeton, New Jersey 08544

Michel Helfgott[b)]
Department of Mathematics, SUNY at Oswego, Oswego, New York 13126

(Received 21 March 2002; accepted 23 August 2002)

We provide an algebraic proof of the fact that Fermat's principle of least time implies Snell's law. This proof is closer to Fermat's original approach than the usual calculus-based developments of the subject.
[DOI: 10.1119/1.1514235]

The search for the law governing the refraction of light played an important role in the development of physics. Euclid knew the law of reflection of light for plane mirrors,[1] but despite the efforts of many scientists over the years, including giants such as Johannes Kepler, the law of refraction could not be found.[2] It appeared in print for the first time in Rene Descartes' *La Dioptrique* in 1637.[3] He found that $\sin\alpha/\sin\beta = K$, where α is the angle of incidence, β the angle of refraction, and K a constant that depends only on physical media that the ray of light goes through, e.g., air and water (see Fig. 1). Willebrord Snell already knew this law as early as 1621, and before him Thomas Hariot had made precise experiments about refraction arriving at the same conclusion.[4]

Descartes' approach to refraction led to the equality $K = \nu_2/\nu_1$, where ν_1 is the velocity of light in the first medium (henceforth taken to be air) and ν_2 is the velocity of light in the second medium (henceforth taken to be water).[5] Pierre de Fermat criticized Descartes' arguments and after strenuous effort was able to prove at around 1661 that if one were to accept the principle of least time, then $\sin\alpha/\sin\beta = \nu_1/\nu_2$.[6] That is, the value of K obtained by Fermat is exactly the inverse of the value obtained by Descartes. Note that Isaac Newton obtained the same law as Descartes.[7] In the seventeenth century the discrepancy could not be resolved because there was no means of determining accurate values for the velocities of light in air and in water. This determination was possible only in 1850 when Leon Foucault found very precise values for ν_1 and ν_2; it turned out that $\nu_2 < \nu_1$.[8] Because it is experimentally determined that $\alpha > \beta$, we can assert that $\sin\alpha/\sin\beta > 1$. Thus Fermat turned out to be right.

Fermat's argument is quite interesting but rather long. The tools he used foreshadow what would become the standard techniques of differential calculus.[9] The purpose of this note is to develop a noncalculus proof that differs from other noncalculus proofs[10] in that it is algebraic rather than geometrical in nature and does not use a limit process. We will compare the value of a given function at an assumed extremum with the value at a nearby point. This kind of argument can be used to obtain rigorous elementary solutions to many problems involving maxima and minima.[11]

Let us state with care what we want to prove: Suppose that given any two points, Q_1 (in air) and Q_2 (in water), a ray of light going from Q_1 to Q_2 does so touching the separation surface at a point R in such a way that the total time $Q_1R/\nu_1 + RQ_2/\nu_2$ is minimal. Then $\sin\alpha/\sin\beta = \nu_1/\nu_2$. The idea behind the proof is to disprove the two inequalities $\sin\alpha/\sin\beta < \nu_1/\nu_2$ and $\sin\alpha/\sin\beta > \nu_1/\nu_2$, so that necessarily $\sin\alpha/\sin\beta = \nu_1/\nu_2$.[12] We first assume that $\sin\alpha/\sin\beta < \nu_1/\nu_2$ and choose point P_1 (in air) and point P_2 (in water), lying on the path Q_1RQ_2 of minimal time, such that $P_1R = P_2R = r$ (see Fig. 2). This particular choice, a choice of Fermat, will simplify the calculations considerably. We define $x_1 = H_1R$ and $x_2 = RH_2$. Because we assumed that $\sin\alpha/\sin\beta < \nu_1/\nu_2$, we can conclude that $x_1/x_2 = (x_1/r)/(x_2/r) = \sin\alpha/\sin\beta < \nu_1/\nu_2$. So $(x_2/\nu_2) - (x_1/\nu_1) > 0$. Next select a positive δ such that $\delta < x_2$ and $\delta < 2((x_2/\nu_2) - (x_1/\nu_1))/((1/\nu_1) + (1/\nu_2))$. In the process of the proof, we will realize why we selected a positive δ satisfying these two conditions.

Let R' be the point on the surface of separation to the right of R, such that $RR' = \delta$. We will see that the ray of light would take less time going from P_1 to R' and then to P_2 instead of going from P_1 to R and then to P_2. We will have reached a contradiction, as we have taken the latter to be the path of minimal time. In other words, we will see that $\Delta < 0$, where

$$\Delta = \left(\frac{1}{\nu_1}P_1R' + \frac{1}{\nu_2}R'P_2\right) - \left(\frac{1}{\nu_1}P_1R + \frac{1}{\nu_2}RP_2\right). \qquad (1)$$

We note that

$$\begin{aligned}\Delta &= \frac{\sqrt{(x_1+\delta)^2 + h_1^2}}{\nu_1} + \frac{\sqrt{(x_2-\delta)^2 + h_2^2}}{\nu_2} - \left(\frac{r}{\nu_1} + \frac{r}{\nu_2}\right)\\ &= \frac{1}{\nu_1}(\sqrt{\delta^2 + 2\delta x_1 + r^2} - r) + \frac{1}{\nu_2}(\sqrt{\delta^2 - 2\delta x_2 + r^2} - r).\end{aligned} \qquad (2)$$

But $\sqrt{b+a} - \sqrt{a} < b/\sqrt{a}$ for any $a > 0$ and any b such that $b + a > 0$.[13] We let $a = r^2$ and $b = \delta^2 + 2\delta x_1$ or $b = \delta^2 - 2\delta x_2$, and obtain $\sqrt{\delta^2 + 2\delta x_1 + r^2} - r < (\delta^2 + 2\delta x_1)/r$ and $\sqrt{\delta^2 - 2\delta x_2 + r^2} - r < (\delta^2 - 2\delta x_2)/r$. So

$$\begin{aligned}\Delta &< \frac{1}{\nu_1}\left(\frac{\delta^2 + 2\delta x_1}{r}\right) + \frac{1}{\nu_2}\left(\frac{\delta^2 - 2\delta x_2}{r}\right)\\ &= \frac{\delta}{r}\left[\delta\left(\frac{1}{\nu_1} + \frac{1}{\nu_2}\right) + 2\left(\frac{x_1}{\nu_1} - \frac{x_2}{\nu_2}\right)\right].\end{aligned} \qquad (3)$$

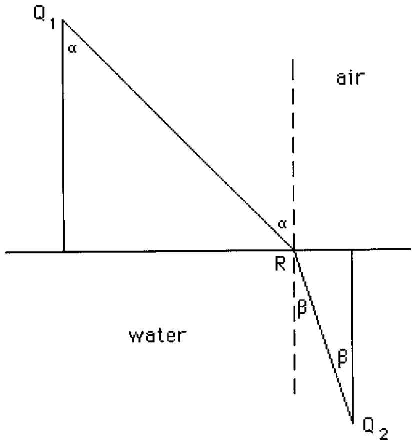

Fig. 1. Refraction of light.

We can easily see that $\delta((1/\nu_1)+(1/\nu_2))+2((x_1/\nu_1)-(x_2/\nu_2))<0$ if and only if $\delta<2((x_2/\nu_2)-(x_1/\nu_1))/((1/\nu_1)+(1/\nu_2))$. Hence, $\Delta<0$, and a contradiction has been reached.

To recapitulate, we began with the assumption that the ray strikes the surface at a point R such that $\sin\alpha/\sin\beta<\nu_1/\nu_2$. From this assumption we obtained the inequality $x_2/\nu_2>x_1/\nu_1$. We then chose a point R' at a distance δ to the right of R such that δ is less than both x_2 and the positive ratio $2((x_2/\nu_2)-(x_1/\nu_1))/((1/\nu_1)+(1/\nu_2))$. This choice of R' in turn led us to the conclusion that the time to traverse a route through R' is less than that through R, showing that any such R cannot be a path of minimum time.

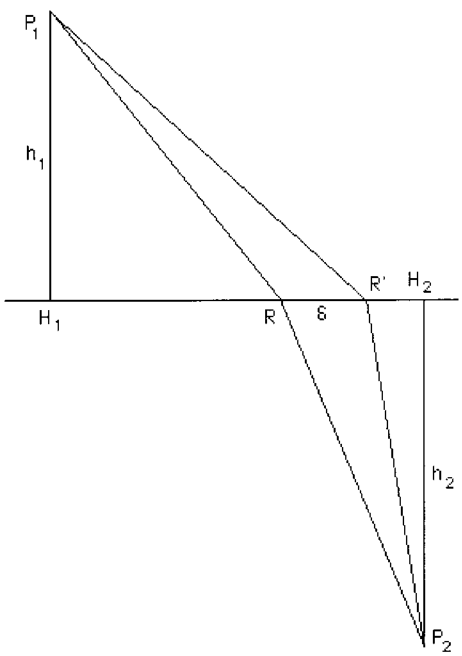

Fig. 2. The path of minimal time.

We can replace $\sin\alpha/\sin\beta>\nu_1/\nu_2$ by $\sin\alpha/\sin\beta<\nu_1/\nu_2$ by relabeling air and water as water and air, and hence Q_1, Q_2, ν_1, ν_2, α, β by Q_2, Q_1, ν_2, ν_1, β, α, respectively. Thus both inequalities lead to a contradiction, and consequently $\sin\alpha/\sin\beta=\nu_1/\nu_2$.

As the reader may have noted, a noncalculus proof of the fact that Fermat's principle implies Snell's law is bound to require some effort. The need to surmount these difficulties was one of the driving forces behind the early development of calculus by Gottfried Leibniz. In fact, the interest generated by the phenomenon of refraction led Leibniz to discuss it in 1684 in the first paper ever published on calculus.[14] The usual calculus-based proof is certainly impressive, and should be taught to undergraduates, but it does not convey a sense of the difficulties that Fermat had to surmount.[15]

It is pedagogically sound to start in high school with a quasiempirical approach to refraction.[16] In college, physics majors can compare a noncalculus and a calculus approach to Fermat's principle of least time in the context of discussing the phenomenon of refraction of light. A course on optics or the history of science would be the ideal environment for such a pursuit.

[a]Electronic mail: haraldh@princeton.edu
[b]Electronic mail: helfgott@oswego.edu

[1]A. I. Sabra, *Theories of Light* (Cambridge U.P., Cambridge, 1981), p. 69.
[2]M. M. Schiffer and L. Bowden, *The Role of Mathematics in Science* (The Mathematical Association of America, Washington, DC, 1984), pp. 82–84.
[3]W. B. Joyce and A. Joyce, "Descartes, Newton, and Snell's law," J. Opt. Soc. Am. **66** (1), 1–8 (1976).
[4]J. W. Shirley, "An early experimental determination of Snell's law," Am. J. Phys. **19**, 507–508 (1951).
[5]Reference 1, p. 111.
[6]Reference 1, pp. 149–150.
[7]Reference 2, pp. 91–95.
[8]M. V. Klein and T. E. Furtak, *Optics* (Wiley, New York, 1986), 2nd ed., p. 35.
[9]Reference 1, pp. 145–148.
[10]See, for instance, Ref. 2, pp. 84–91; A. Nussbaum, *Geometric Optics: An Introduction* (Addison–Wesley, Reading, MA, 1968), pp. 17–19; M. Golomb, "Elementary proofs for the equivalence of Fermat's principle and Snell's law," Am. Math. Monthly **71**, 542–543 (1964); J. W. Blaker, *Geometric Optics; The Matrix Theory* (Dekker, New York, 1971), pp. 5–6.
[11]H. Helfgott and M. Helfgott, "Maxima and minima before calculus," Pro Mathematica **12** (23-24), 135–158 (1998).
[12]This way of proving an equality was one of Archimedes' favorite tools. See W. Dunham, *Journey Through Genius* (Penguin, New York, 1990), p. 92.
[13]This inequality is valid because $(\sqrt{a+b}-\sqrt{a})(\sqrt{a+b}+\sqrt{a})=b+a-a=b$. Thus $\sqrt{a+b}-\sqrt{a}=b/(\sqrt{a+b}+\sqrt{a})<b/\sqrt{a}$.
[14]*The History of Mathematics: A Reader*, edited by J. Fauvel and J. Gray (MacMillan Education, London, 1987), pp. 432–433.
[15]See, for instance, G. F. Simmons, *Calculus with Analytic Geometry* (McGraw–Hill, New York, 1985), pp. 106–107.
[16]M. Helfgott, "Computer technologies and the phenomenon of refraction," Phys. Teach. **36** (5), 236–238 (1998).