

Route Choice in Hilly Terrain

Anthony Kay

Department of Mathematical Sciences, Loughborough University, Loughborough, U.K.

The orienteering route choice problem involves finding the fastest route between two given points, with running speed determined by various properties of the terrain. In this study, I consider only the effect of climbing or descending on running speed. If a runner's pace p (the reciprocal of speed) varies linearly with gradient m , the straight-line route always is fastest. However, a nonlinear formulation for $p(m)$, with $d^2p/dm^2 > 0$, will more accurately model runners' capabilities. As a result, critical gradients may exist for ascent and/or descent, such that optimal routes will never ascend or descend more steeply than the critical gradient. I review and propose several formulations for the pace function $p(m)$ and calculate their critical gradients. In principle, the Euler–Lagrange equation can be used to find optimal routes between arbitrary points on any topography where the height can be expressed as a smooth function of horizontal coordinates. I obtain first integrals of this equation for idealized landforms: hillsides with straight contours and axisymmetric hills. Next, optimal routes are computed for various combinations of start- and endpoints on these landforms based on various pace functions. These routes are classified as either subcritical or maximal steepness: The former ascends or descends less steeply than the critical gradient; the latter takes the line of steepest ascent where it is not steeper than the critical gradient, but follows a curve at the critical gradient where the slope is steeper. In some cases, the optimal route zigzags up or down a hill along sections of a critical-gradient curve.

Introduction

The sport of orienteering requires participants to determine the fastest route along a leg between two given points in a terrain that often is hilly or even mountainous. This is an example of a minimum-cost path problem: Find the path in two-dimensional space that minimizes the integral

$$t_{AB} = \int_A^B p ds, \quad (1)$$

where A and B are given points, p is a cost field (Miller and Bridwell 2009), and s is the distance along the path. In the application here, p is pace, the reciprocal of running speed (Scarf 2007). More precisely, p is defined as the time per unit horizontal distance, as shown on a map (as opposed to distance over sloping ground). A runner's pace is likely to be a function of three factors (Arnet 2009):

- the gradient at which the runner is climbing or descending, henceforth called *route gradient* and denoted m , with m being positive uphill;

Correspondence: Anthony Kay, Department of Mathematical Sciences, Loughborough University, Loughborough, Leicestershire, LE11 3TU, U.K.
e-mail: a.kay@lboro.ac.uk

Submitted: November 30, 2009. Revised version accepted: February 8, 2011.

- the slope of the terrain, denoted m_{\perp} hereafter; and
- the *runnability* of the terrain, which is the effect of vegetation or uneven ground on the runner's speed relative to that on a smooth surface.

Route gradient is related to terrain slope by

$$m = m_{\perp} \sin(\psi), \quad (2)$$

where ψ is the angle between the route and the contours. The route gradient is anisotropic (dependent on the direction of travel), whereas the other two factors are isotropic. Terrain slope is included separately from route gradient because it is slower to traverse horizontally across a cliff face than to run across a flat field.

Minimum cost path problems in geographical and economic applications often have been addressed by discretizing space. For the orienteering application, Hayes and Norman (1984) take this approach, imposing a 250 m grid on a complex area of mountain terrain and then using a dynamic programming algorithm to find the fastest route; pace is assumed to depend on route gradient and runnability in a rather crude way. Developments of the past 25 years in computer technology enabled Arnet (2009) to impose a 1 m grid on an admittedly smaller area of very complex terrain, accounting for all three preceding factors, and using Dijkstra's (1959) algorithm to optimize a route (or so it appears from Arnet's description, although he refers to a "minimum spanning tree"). Thus, the orienteering route choice problem appears to have been solved by the application of sufficient computing resources. This approach, however, does not yield much theoretical insight, and, furthermore, the least satisfactory aspect of Arnet's (2009) analysis is his treatment of route gradient.

In contrast, the purpose of this article is to gain a theoretical understanding of how route gradient alone affects route choice. For this purpose, I consider simple, idealized terrains, together with idealized representations of a runner's pace over those terrains. This simplification allows for a continuous representation of space, offering the prospect of closed-form mathematical solutions (the prevalence of discrete representations of space in geographical research being due to the complex nature of most real terrain). Scarf (2008) uses a variety of idealized terrain topographies when applying a route choice procedure based on comparing the actual shape of a hill with the calculated form of an isochronic hill on which all routes between two given points take the same time. Alexander (2000) determines minimum-cost routes on one idealized topography, a pyramidal hill, by direct calculation of a walker's energy cost. He also suggests the calculus of variations as a method with more general applicability, and that is indeed the classical method for route optimization in continuous space. Puu and Beckmann (1999) use the calculus of variations in urban transport applications, although they give solutions only for isotropic cost fields. In fact, the minimum cost path problem for an isotropic field formally is identical to geometrical optics, because Fermat's principle dictates that light travels between given points in an isotropic medium by the fastest route (Warntz 1957; Puu 1997; Kay 2006). Although Fermat's principle does not apply in anisotropic media, the calculus of variations is still valid and forms the basis of the analysis presented here.

I first consider two classes of idealized topography. Probably the simplest nontrivial case is a hillside with straight but unevenly spaced contours, so that the direction of steepest ascent is uniform (apart from reversals at a valley bottom or ridge top), and the terrain slope varies only in the direction up and down the hillside. Good approximations of this landform are common in the Pennine hills of northern England (Fig. 1a). For convenience, I henceforth refer to such hillsides as *Pennine slopes*. Second, I consider axisymmetric hills (i.e., with concentric circular contours), starting with conical hills (uniform slope) and then generalizing to cases where the slope varies with distance from the summit. Two British hills with approximate axisymmetry appear in Fig. 1b,c.

A runner's pace is idealized in several ways in this analysis. First, the isotropic factors of slope and runnability are ignored entirely, so that pace is a function only of route gradient: $p = p(m)$. This simplification is justified by the need to build a theoretical understanding one factor at a time. Second, Townshend, Worringham, and Stewart (2010) show that even when slope and runnability effects are eliminated (by

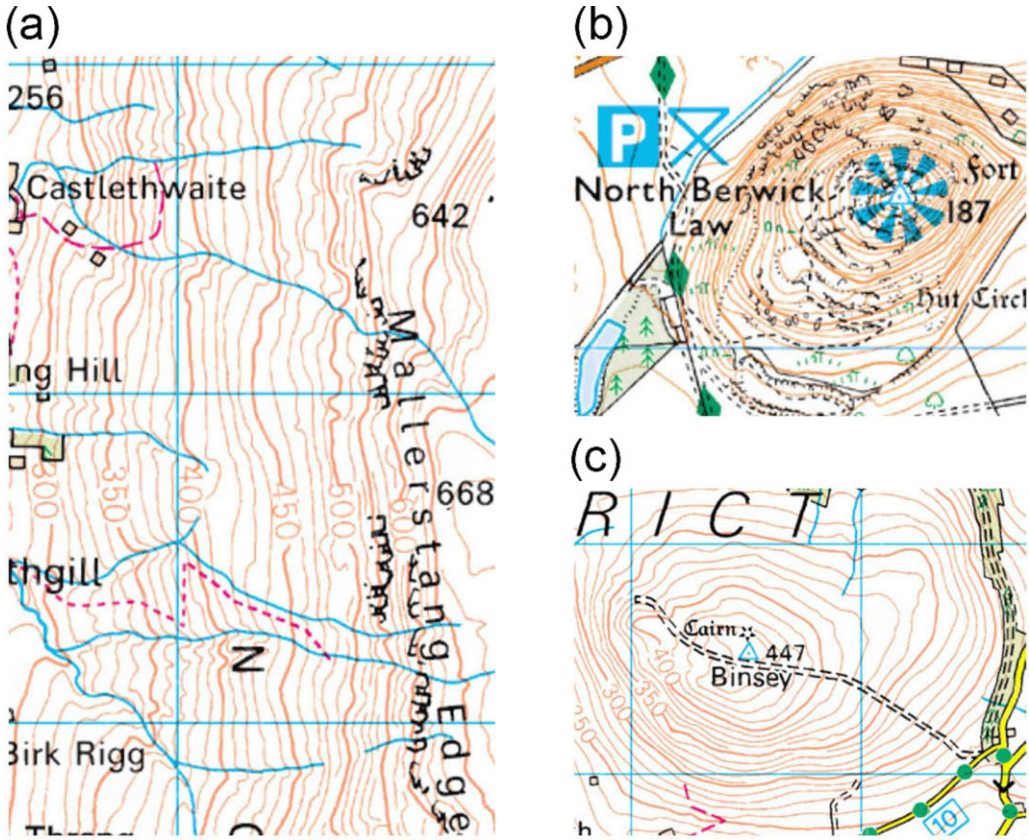


Figure 1. British locations approximating idealized landforms: (a) straight, unevenly spaced contours at Mallerstang in the North Pennines; (b) North Berwick Law, an axisymmetric volcanic plug in East Lothian, with $m_{\perp} \approx 0.5$; and (c) Binsey, an axisymmetric hill in Cumbria with $m_{\perp} \approx 0.2$. Map extracts are © Crown copyright Ordnance Survey. All rights reserved.

running along well-built tracks), a runner's pace depends not only on the current gradient, but also on that of the preceding section of route. For example, when the route levels out after an uphill section, a runner takes some distance to accelerate to his/her usual level-ground pace. I ignore this effect. Third, runners naturally adjust their gait to optimize speed or energy consumption (Alexander 2000), so the pace function $p(m)$ is likely to be continuous, but is probably not differentiable where a change of gait (running to walking) occurs; nevertheless, I take $p(m)$ to be at least twice differentiable at all gradients m .

Dependence of a runner's pace on route gradient

Given that there is no single pace function $p(m)$ that applies to all runners, one of the objectives here is to determine how route choice depends on an individual runner's pace function. The functional forms that I use for $p(m)$ are mostly taken from pace functions previously suggested by other authors for runners or walkers in hilly terrain; the numerical values of coefficients in those functions are mainly taken from my own examination of published experimental data and race results. I first consider a linear function, the well-known Naismith's rule, and its consequences for route choice. Critical gradients for both ascent and descent may exist for nonlinear $p(m)$, such that selection of a route that is steeper than the relevant critical gradient is never advantageous. I discuss this phenomenon in general terms before reviewing a variety of nonlinear

pace functions and calculating their critical gradients. Because uphill running receives considerably more attention in the literature than downhill running, I consider functions valid only for ascent before proposing functions applicable to both ascent and descent.

Naismith's rule: a linear pace function

Naismith's rule was first proposed in 1892 (Langmuir 1984): If the pace on level terrain is denoted by p_0 , then the rule may be written as

$$p = p_0(1 + \alpha m) \quad [m \geq 0]. \quad (3)$$

Scarf (1998) gives the value of the constant α as 7.92. Scarf's analysis of records from British fell races,¹ which generally include the same amount of descent as ascent, supports this value; however, the fitting of Naismith's rule to a selection of uphill-only race results here yields a value of 9.91. The discrepancy may be explained by noting that Naismith's rule does not explicitly account for descent: it can be taken either to apply only to routes with a start and finish at the same altitude or to imply that pace on a descent is the same as on level ground (Scarf 2008). The latter assumption, which may be formulated as

$$p = p_0(1 + \alpha mH(m)), \quad (4)$$

where H is the Heaviside unit-step function, was used by Arnet (2009). However, the transition from gentle uphill to gentle downhill running does not involve a change of gait, so the pace function is likely to be differentiable at $m = 0$. Thus, I remove the restriction $m \geq 0$ from Naismith's rule (3) and take it to be the linear approximation to a smooth nonlinear pace function, valid for sufficiently small (uphill or downhill) route gradients (Davey, Hayes, and Norman 1994). Substituting equation (3) into (1) and expressing differential distance in terms of differential height gain, $ds = dh/m$, yields

$$\begin{aligned} t_{AB} &= \int_A^B \frac{p_0}{m} dh + \int_A^B p_0 \alpha dh \\ &= p_0 \int_A^B ds + p_0 \alpha (h_B - h_A). \end{aligned} \quad (5)$$

Although Davey, Hayes, and Norman (1994) note this result, they do not observe its implication for general route choice problems: because the total height gain $h_B - h_A$ is fixed, t_{AB} is minimized by minimizing the distance $\int_A^B ds$. Hence, a straight-line route is always quickest when a linear pace function applies; no advantage exists in choosing a longer route at a gentler gradient. We may interpret this consequence as meaning that a straight-line route is quickest in any terrain where the gradient (uphill or downhill) is everywhere sufficiently gentle that a linear approximation to the true pace function $p(m)$ may be applied.

The critical gradient

For a runner with horizontal pace p on a gradient m , the rate at which height is gained is m/p . This quantity is maximized when

$$\frac{dp}{dm} = \frac{p}{m}, \quad (6)$$

provided that

$$\frac{d^2p}{dm^2} > 0, \quad (7)$$

which is the case for all the ensuing pace functions. Solutions of equation (6) may exist for either or both positive and negative values of m , as shown in Fig. 2. These solutions give the gradients at which a runner

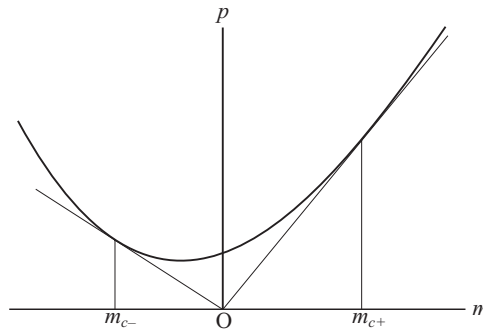


Figure 2. A notional pace function $p(m)$ with $d^2p/dm^2 > 0$. The rate of height gain/loss lm/p is maximized where a line from the origin to the pace-function curve has the least angle to the horizontal axis; that is, at gradients m_{c+} and m_{c-} where the line is tangential to the curve so that equation (6) is satisfied.

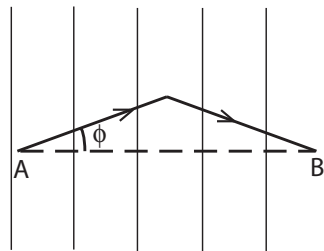


Figure 3. Optimal route from A to B on uniform slope with gradient $m_{\perp} > m_c$. The thin lines are contours.

with the given pace function can gain or lose height fastest: These are the uphill and downhill critical gradients, denoted m_{c+} and m_{c-} , respectively (or simply m_c when the discussion applies equally to ascent or descent). If the terrain is less steep than a critical gradient, a route in the direction of steepest ascent/descent maximizes a runner's rate of height gain/loss; but if the terrain slope m_{\perp} is greater than m_c , the rate of height gain/loss is maximized on a route at an angle

$$\phi = \cos^{-1}(m_c/m_{\perp}) \tag{8}$$

to the steepest ascent direction. As a simple application to a route choice problem with fixed start- and endpoints, consider a *direct ascent leg* (where the straight-line route from start- to endpoint is in the direction of steepest ascent) on a uniform Pennine slope with $m_{\perp} > m_c$. To minimize the time taken, a runner would need to use a zigzag route such as that shown in Fig. 3, with each section at angle ϕ to the straight-line route.

Davey, Hayes, and Norman (1994) appear to be the first to note the existence of a critical gradient and the need for zigzagging to minimize the time taken to climb a slope steeper than m_c , although their calculations apply to two particular choices of pace functions and are valid only for uphill running. Critical gradients may arise when optimizing the integral of any gradient-dependent cost function: By reploting earlier data for the metabolic cost of walking up and down various gradients, Minetti (1995) determines that the optimal gradient (for least metabolic cost) is around 0.25 for ascent, and -0.20 for descent; he also suggests zigzag routes on steeper hills. Rees (2004) makes a similar calculation to Davey, Hayes, and Norman (1994), but with a quadratic cost function that forces the critical gradients for ascent and descent to be of equal magnitude. However, he suggests that the parameter values in his cost function, and hence the

critical gradients, take different values depending on whether walking time or metabolic cost is minimized. Llobera and Sluckin (2007) do a thorough mathematical analysis of metabolic cost minimization on a slope, obtaining an equation of the same form as equation (6), and deriving a metabolic cost function from which they obtain values for uphill and downhill critical gradients.

The numerical values of m_{c+} from the preceding authors' work lie in the range 0.2 to 0.4, with values of m_{c-} generally being of smaller magnitude than m_{c+} . These values are consistent with a suggestion by Balstrøm (2002) that walkers avoid terrain with a slope in excess of 0.3. All of the preceding results, however, were obtained either from data for walking or from controlled experiments involving running on treadmills. A different picture emerges when we consider records from mountain races: A pace function derived by Scarf (2007) from results for British races yields a much higher value of m_{c+} , while my own analysis of records from uphill-only races worldwide suggests that no uphill critical gradient exists within the range of gradients encountered in races. This finding is consistent with observations that competitors in these races almost invariably choose a route directly up a slope, however steep. I speculate that this selection is because runners have the freedom to minimize their total cost (whether that be time taken or energy expended) by adjusting their gait (Alexander 2000) as well as by varying their route.

Although several of the preceding authors have sought support for their theory about critical gradients by examining mountain paths, the evidence on the ground is inconclusive and may reflect differences between walkers and race competitors. For example, on North Berwick Law (Fig. 1b), a prominent landmark with a slope generally around 0.5, the path to the summit spirals and zigzags at a gradient generally less than 0.2, somewhat gentler than any value suggested for m_c . In contrast, on the opposite side of the Firth of Forth is another volcanic plug, Largo Law, which has a path ascending almost directly to its southern top at a gradient in excess of 0.4 (see <http://www.geograph.org.uk/photo/12932>). This hill is not well known to walkers but has a well-established annual hill race. Furthermore, gorse growing on the lower slopes most likely discourages zigzagging: Runnability is an important factor in this case.

Nonlinear pace functions

I now briefly review the pace functions suggested by previous authors, and select a few functions to be used in the route choice calculations here. I do not claim that my selected functions are the most accurate representations of a runner's pace; rather, I seek a set of idealized pace functions that have some empirical support and can be used to illustrate the dependence of route choice on pace function. I also calculate critical gradients for each of the pace functions.

Scarf's function

Based on a regression analysis of British fell-running records, Scarf (1998, 2007) proposes a pace function that may be written as

$$p = p_0(1 + \alpha m)^\beta, \quad (9)$$

in which the best fit is obtained with $\beta \approx 1.14$ and $\alpha \approx 8.6$ for men, and $\beta \approx 1.16$ and $\alpha \approx 10.6$ for women. Scarf's function (9) is intended to take fatigue into account, and so p_0 is weakly dependent (by a $\beta - 1$ power law) on total distance to be covered. I ignore the very small effect that this dependence has on route choice between fixed points. Substituting equation (9) into (6), I find the uphill critical gradient to be

$$m_{c+} = \frac{1}{\alpha(\beta - 1)} \quad (10)$$

for Scarf's function; the numerical values for m_{c+} are 0.831 and 0.590, respectively, when α and β take the preceding values for men and women. I use the men's pace function

$$p = p_0(1 + 8.6m)^{1.14} \quad (11)$$

in a sample calculation for an uphill route choice problem.

Davey, Hayes, and Norman's exponential functions

Davey, Hayes, and Norman (1994) suggest two alternative formulas as good descriptions of data from their treadmill experiments about uphill running. Their formulas are for speed, v , along a sloping treadmill as a function of slope angle, and can be written in terms of pace $p = \sqrt{1+m^2}/v$ and gradient m as

$$p = p_0 \sqrt{1+m^2} e^{c \tan^{-1} m} \quad (12)$$

and

$$p = p_0 e^{k \tan^{-1} m}, \quad (13)$$

where c and k are empirical constants. The critical gradients for these pace functions are

$$m_{c+} = \frac{1}{c} \quad (14)$$

and

$$m_{c+} = \frac{1}{2} \left(k - \sqrt{k^2 - 4} \right), \quad (15)$$

respectively.

After reexamining the data from each of Davey, Hayes, and Norman's (1994) five experiments, I found that equation (13) achieves a slightly better fit than (12), but with $k \approx 3.0$ rather than Davey, Hayes, and Norman's value of $k = 3.5$. Therefore, my analysis uses

$$p = p_0 e^{3.0 \tan^{-1} m} \quad (16)$$

as a pace function for an uphill route choice example. The uphill critical gradient associated with equation (16) is $m_{c+} \approx 0.382$.

Quadratic functions

Rees (2004) proposes that a pace function of form

$$p = p_0(1 + am + bm^2) \quad (17)$$

should apply to walking both uphill and downhill, and Llobera and Sluckin (2007) also suggest a quadratic function for the metabolic cost of walking up and down slopes. However, this function forces the uphill and downhill critical gradients to be of equal magnitude:

$$m_{c\pm} = \pm b^{-1/2}. \quad (18)$$

Accordingly, I do not use it for route choice problems involving both ascent and descent. Nevertheless, I found that the quadratic function

$$p = p_0(1 + 5.75m + 12.33m^2) \quad (19)$$

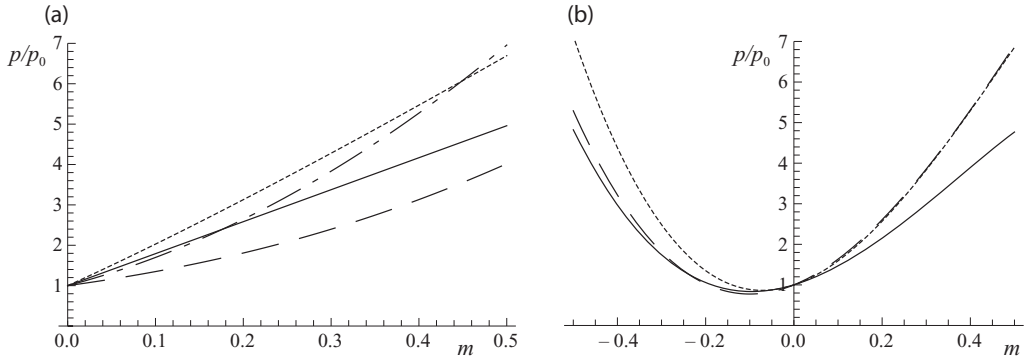


Figure 4. (a) Pace functions valid for uphill running only. Solid line: Naismith's rule (3) with $\alpha = 7.92$; dotted line: Searf's function (11); dashed line: Davey, Hayes, and Norman's (1994) function (16); dash-dotted line: quadratic function (19). (b) Pace functions valid for uphill and downhill. Solid line: CR (cubic, race records); dashed line: CS (cubic, skilled descender); dotted line: CT (cubic, timid descender).

achieves an excellent fit to the data of Minetti et al. (2002) for running uphill² and use it in uphill route choice examples. The critical gradient for equation (19) is $m_{c+} \approx 0.285$.

Graphs of pace functions valid only for uphill running appear in Fig. 4a. Linear approximations to these functions have widely differing values of the Naismith coefficient $[dp/dm]_{m=0}$. Searf (2007) discusses likely reasons for these discrepancies. Most important, both Naismith's rule and Searf's function are derived for routes starting and finishing at the same altitude, without separating the effects of descent from ascent; this probably explains their values of the Naismith coefficient being much higher than rules derived from experiments on uphill running.

Higher-degree polynomials

A polynomial function must have a degree of at least 3 to model the asymmetry between a runner's behavior uphill and downhill. Llobera and Sluckin (2007) argue that a quartic (degree 4) should be used to describe metabolic cost as a function of gradient. However, here the fitting of both cubic and quartic functions to various selections of data for runners' pace indicates that cubic functions of the form

$$p = p_0(1 + am + bm^2 + cm^3) \quad (20)$$

are more satisfactory. Next, I propose three such pace functions.

An Internet search for record times in uphill-only and downhill-only races, as well as uphill and downhill legs of other mountain races, yielded a dataset of pace versus gradient for 39 uphill cases (gradients $0.080 \leq m \leq 0.521$) and 7 downhill cases (gradients $-0.25 \leq m \leq -0.105$). Pace function CR (cubic, race records) is the following ordinary least-squares cubic fit to this dataset:

$$p = p_0(1 + 3.281m + 15.180m^2 - 13.343m^3), \quad (21)$$

for which $m_{c-} = -0.218$, with no uphill critical gradient.

Other studies suggest that an uphill critical gradient should exist, implying a pace function that yields both uphill and downhill critical gradients. Furthermore, considerably more diversity exists between individual runners' behaviors downhill than uphill: This feature may be observed in any British fell race, and also has been noted in the results of a controlled experiment by Townshend, Worringham, and Stewart (2010) with runners on comparatively gentle gradients (mean $|m| = 0.082$). Uphill running simply requires fitness, but downhill running also requires agility and even fearlessness. A measure of a runner's descending

skill is the ratio of uphill pace to downhill pace in a race that ascends a single mountain and descends by the same route. Two such races in Britain have published summit times as well as finishing times: Snowdon and Scafell Pike, with mean gradients of 0.125 and 0.25, respectively. I define a “skilled descender” and a “timid descender” as having uphill : downhill pace ratios 1.5 standard deviations above and below the mean for one year’s results from each of these races. Then, using the uphill pace data of Minetti et al. (2002) (for which m_{c+} certainly exists), I calculated notional paces for the skilled and timid descender at downhill gradients of -0.125 and -0.25 . Finally, I fit cubic functions to the Minetti et al. uphill data and the notional downhill paces for the skilled (CS) and timid (CT) descenders. These functions, together with their uphill and downhill critical gradients, are

$$\begin{aligned} \text{CS: } p &= p_0(1 + 4.400m + 20.394m^2 - 11.139m^3), \\ m_{c+} &= 0.262, m_{c-} = -0.201; \text{ and} \end{aligned} \quad (22)$$

$$\begin{aligned} \text{CT: } p &= p_0(1 + 3.676m + 23.981m^2 - 15.824m^3), \\ m_{c+} &= 0.249, m_{c-} = -0.183. \end{aligned} \quad (23)$$

The justification for the rather arbitrary procedure used to derive these formulas is the desire to obtain contrasting pace functions, idealized but with some evidential basis, for route choice studies.

Graphs of the pace functions CR, CS, and CT appear in Fig. 4b. They have minimum pace (maximum speed) at gradients of -0.096 , -0.100 , and -0.072 , respectively, consistent with the finding of Margaria (1938) (quoted by Minetti [1995]) that the minimum metabolic cost is at a downhill gradient of -10% . The critical gradients are rather sensitive to small details of the pace function: CS and CT were constructed to be almost identical uphill, but their values of m_{c+} differ by 0.013; CS and CR appear very similar downhill, but the value of m_{c-} for CS is about midway between those for CR and CT.

Route optimization by calculus of variations

The Euler–Lagrange equation and the critical gradient

I begin by imposing a general orthogonal coordinate system (ξ, η) on the plane, letting

$$f(\xi, \eta, \eta') = p \frac{ds}{d\xi}, \quad (24)$$

where primes denote derivatives with respect to ξ and total differentials indicate variations along a route. Equation (1) becomes

$$t_{AB} = \int_A^B p ds = \int_A^B f d\xi, \quad (25)$$

which is minimized for a route $\eta(\xi)$ satisfying the Euler–Lagrange equation

$$\frac{\partial f}{\partial \eta} - \frac{d}{d\xi} \left(\frac{\partial f}{\partial \eta'} \right) = 0. \quad (26)$$

For a coordinate ξ with a scale factor of unity, the quantity $ds/d\xi$ in equation (24) is the secant of the angle between a route and the local ξ coordinate curve.

The terrain height $h(\xi, \eta)$ is assumed to be a sufficiently differentiable function of position for all the following calculations to be valid. Its directional derivative along the route is the route gradient,

$$m = \frac{dh}{ds} = \frac{dh}{d\xi} \frac{d\xi}{ds}, \quad (27)$$

in which

$$\frac{dh}{d\xi} = \frac{\partial h}{\partial \xi} + \eta' \frac{\partial h}{\partial \eta}. \quad (28)$$

From equations (27) and (28) we may obtain

$$\frac{\partial m}{\partial \eta} = \frac{\partial}{\partial \eta} \left(\frac{dh}{d\xi} \right) \frac{d\xi}{ds} - m \frac{d\xi}{ds} \frac{\partial}{\partial \eta} \left(\frac{ds}{d\xi} \right), \quad (29)$$

and

$$\frac{\partial m}{\partial \eta'} = \frac{\partial h}{\partial \eta} \frac{d\xi}{ds} - m \frac{d\xi}{ds} \frac{\partial}{\partial \eta'} \left(\frac{ds}{d\xi} \right). \quad (30)$$

We can now calculate the derivatives required in equation (26) using (24), given that p is a function of m and using equations (29) and (30), where required. First,

$$\begin{aligned} \frac{\partial f}{\partial \eta} &= \frac{dp}{dm} \frac{\partial m}{\partial \eta} \frac{ds}{d\xi} + p \frac{\partial}{\partial \eta} \left(\frac{ds}{d\xi} \right) \\ &= \frac{dp}{dm} \frac{\partial}{\partial \eta} \left(\frac{dh}{d\xi} \right) + \left(p - m \frac{dp}{dm} \right) \frac{\partial}{\partial \eta} \left(\frac{ds}{d\xi} \right). \end{aligned} \quad (31)$$

Similarly,

$$\frac{\partial f}{\partial \eta'} = \frac{dp}{dm} \frac{\partial h}{\partial \eta} + \left(p - m \frac{dp}{dm} \right) \frac{\partial}{\partial \eta'} \left(\frac{ds}{d\xi} \right). \quad (32)$$

We now define the orthogonal coordinates with ξ in the direction of steepest ascent and η along the contours, so that $\partial h / \partial \eta \equiv 0$, and hence

$$\frac{\partial}{\partial \eta} \left(\frac{dh}{d\xi} \right) = 0. \quad (33)$$

The consequent vanishing of the first terms on the right-hand sides of equations (31) and (32) means that the Euler–Lagrange equation (26) is satisfied for a route on which m has a constant value such that

$$p - m \frac{dp}{dm} = 0. \quad (34)$$

Condition (34) is just a rearrangement of the critical gradient equation (6). Therefore, this method based on the calculus of variations is consistent with the critical-gradient theory developed by Davey, Hayes, and Norman (1994) and Llobera and Sluckin (2007), although I have not yet solved the fastest-route problem because I have taken no account of the locations of the start- and endpoints of a leg. If a terrain slope is less than m_c , a critical-gradient route is impossible; but even if a terrain is steeper than the critical gradient, solutions of the Euler–Lagrange equation between many pairs of points give optimal routes at subcritical gradients. If no smooth solution of the Euler–Lagrange equation can be found between a given pair of points, a route involving zigzagging at the critical gradient is indicated.

The Euler–Lagrange equation on Pennine slopes and axisymmetric hills

To satisfy the condition $\partial h / \partial \eta \equiv 0$, I use Cartesian coordinates $(\xi, \eta) = (x, y)$ with y along the contours on a Pennine slope, whereas I adopt polar coordinates $(\xi, \eta) = (r, \theta)$ on an axisymmetric hill. The pace p is independent of y or θ , and

$$\frac{ds}{dx} = \sqrt{1+y'^2} \quad \text{and} \quad \frac{ds}{dr} = \sqrt{1+r^2\theta'^2} \quad (35)$$

on the respective topographies. Thus, f as defined by equation (24) has no explicit dependence on y or θ , so that the Euler–Lagrange equation (26) has the first integrals

$$\frac{\partial f}{\partial y'} = C \quad \text{and} \quad \frac{\partial f}{\partial \theta'} = C \quad (36)$$

in the respective cases, with the constant C to be determined. The derivatives in equations (36) are evaluated by substituting equations (35) into (32), to yield

$$\frac{y'}{\sqrt{1+y'^2}} \left(p - m \frac{dp}{dm} \right) = C, \quad (37)$$

and

$$\frac{r^2\theta'}{\sqrt{1+r^2\theta'^2}} \left(p - m \frac{dp}{dm} \right) = C. \quad (38)$$

However, the coordinate system (ξ, η) should be set up so that η is a single-valued function of ξ along an optimal route. This certainly is true for (x, y) as previously defined on a Pennine slope, because an optimal route does not involve climbing such a slope and then coming back down farther along it. However, on an axisymmetric hill, an optimal route may climb to a point near the summit and then descend on the far side of the hill, so that θ is not a single-valued function of r ; but the radial coordinate $r(\theta)$ is single valued (except for direct ascent legs, treated separately below), so the identity

$$\theta' \equiv \frac{d\theta}{dr} = \left(\frac{dr}{d\theta} \right)^{-1} \equiv \frac{1}{r'} \quad (39)$$

is used to transform equation (38) to

$$\frac{r^2}{\sqrt{r^2 + r'^2}} \left(p - m \frac{dp}{dm} \right) = C. \quad (40)$$

Equations (37) and (40) are to be solved with boundary conditions given by the coordinates of the start- and endpoints,

$$y(x_A) = y_A \quad \text{and} \quad y(x_B) = y_B, \quad (41)$$

or

$$r(\theta_A) = r_A \quad \text{and} \quad r(\theta_B) = r_B. \quad (42)$$

The optimal route should be a straight line when the pace is given by Naismith's rule (3). Equation (3) yields

$$p - m \frac{dp}{dm} = p_0 \quad (43)$$

so that equations (37) and (40) reduce to

$$y' = \text{constant} \quad \text{and} \quad \frac{r^2 + r'^2}{r^4} = \text{constant}, \quad (44)$$

which indeed are equations of straight lines in the respective coordinate systems.

Some general remarks about solutions of the Euler–Lagrange equation

Satisfying the Euler–Lagrange equation (26) is a necessary but not sufficient condition for a minimum of $\int_A^B f d\xi$. Consider three examples that illustrate the care needed when interpreting solutions of the Euler–Lagrange equation.

First, consider a leg between points at the same height—that is, with $r_B = r_A$ —on an axisymmetric hill. The route along the circular contour connecting the points has

$$r = \text{constant}, \quad r' = 0, \quad m = 0, \quad p = p_0, \quad (45)$$

and thus satisfies equation (40) with $C = r_A p_0$; yet this contouring route gives neither a maximum nor a minimum of t_{AB} .

Second, consider a Pennine slope with uniform m_\perp (evenly spaced contours). Intuitively, the fastest route between any pair of points on such a slope should be a straight line. The straight-line route on a uniform slope has constant m , and hence constant p , and its equation is $y' = \text{constant}$, satisfying equation (37); this route certainly is the fastest for a runner with any realistic pace function, with $d^2 p/dm^2 \geq 0$. However, a hypothetical runner with $d^2 p/dm^2 < 0$ might find a different route faster. For example, for a leg ascending at 45° to the contours on a uniform slope m_\perp , the straight-line route gradient is $m = m_\perp/\sqrt{2}$; but a route that parallels the contours and then turns through a right angle to ascend directly would be faster than the straight-line route if

$$p(0) + p(m_\perp) < \sqrt{2} p\left(\frac{m_\perp}{\sqrt{2}}\right).$$

This situation is physically unrealistic but theoretically conceivable with $dp/dm > 0$ and $d^2 p/dm^2 < 0$: It would require the pace to be only slightly slower at route gradient m_\perp than at gradient $m_\perp/\sqrt{2}$, but much faster when contouring.

Third, consider direct ascent/descent legs. On a Pennine slope or an axisymmetric hill, such a leg has

$$y_B = y_A \quad \text{or} \quad \theta_B = \theta_A, \quad (46)$$

respectively. Solutions of (37) and (38) with $C = 0$ and satisfying the boundary conditions (46) are

$$y' = 0 \quad \text{and} \quad \theta' = 0. \quad (47)$$

These solutions are straight-line routes up the hills. If the terrain gradient m_\perp is less than the critical value m_c , the straight-line route gives the shortest time; but if $m_\perp > m_c$, the straight line yields a local maximum of t_{AB} . In the latter case, the solution of equations (37) and (38) with $C = 0$ that yields a minimum of t_{AB} is

$$p - m \frac{dp}{dm} = 0; \quad (48)$$

that is, a critical-gradient route (which needs to zigzag in order to satisfy the boundary conditions).

For general legs on Pennine slopes and axisymmetric hills, optimal solutions of equations (37) and (40) are classified as maximal-steepness routes if $C = 0$ or as subcritical routes if $C \neq 0$. From the preceding

discussion, a maximal-steepness route ascends/descends at the critical gradient where the terrain slope is steeper than the critical value but is perpendicular to the contours where the slope is less steep. A smooth join exists between a critical-gradient curve and a perpendicular line if the route crosses between regions of supercritical and subcritical slope, and zigzagging usually is required in a region of supercritical slope. Where a zigzag route is required, an infinite variety of such routes exists, each taking the same time but with different numbers and locations of the sharp changes of direction. Llobera and Sluckin (2007) obtain a unique (modulo chirality) optimal zigzag route by introducing a local optimization procedure involving an extra criterion that determines the direction taken at each step. I do not adopt such a procedure, but for purposes of illustration, the zigzag routes here have a single sharp change of direction.

Whereas maximal-steepness routes usually can be calculated analytically, subcritical routes require numerical methods. The value of C usually is unknown a priori, so an iterative “shooting” method is used. An initial estimate of C is made, for example, by evaluating the left-hand side of equation (37) or (40) for a route leaving point A in the direction of the straight line to B. Given the estimate of C , equation (37) or (40) may be numerically integrated forward in x or θ . On reaching $x = x_B$ or $\theta = \theta_B$, the value of y or r attained in general is not the correct value for the endpoint. The iteration proceeds by improving the estimate of C and numerically integrating for each such estimate until the computed route hits the endpoint to within some desired precision. Subcritical routes have a route gradient less than the critical value throughout, with larger values of C for routes at gentler gradients (we can take $C > 0$ without loss of generality).

Examples of optimal routes

Pennine slopes

A general Pennine slope has a variable slope $m_{\perp}(x)$, and all the features of interest can be brought out by considering the parabolic profile

$$h = \frac{1}{2}x^2, \quad m_{\perp} = x; \quad (49)$$

the axis $x = 0$ is a valley bottom of altitude $h = 0$, and the unit of distance is taken as the distance from the valley bottom to the contour where $m_{\perp} = 1$. The numerical equality of slope and x -coordinate means that the terrain is steeper than the uphill or downhill critical gradient where $x > m_{c+}$ or $x < m_{c-}$, respectively, for routes in the direction of increasing x . If a ridge rather than a valley exists along $x = 0$, the routes subsequently calculated are optimal in the reverse direction.

Comparison of routes for different pace functions

Consider legs on which the straight-line route from start- to endpoint would be at 45° to the contours. Optimal routes for three uphill-only pace functions are calculated on a leg starting in the valley bottom and ending at a point where the altitude is $h = 0.125$ and the slope is $m_{\perp} = 0.5$ (Fig. 5a), while for the cubic pace functions (valid both up- and downhill) I take start- and endpoints located where the slope is 0.5 on opposite sides of the valley (Fig. 5b). The slope of 0.5 is close to the steepest slope (apart from crags) in Fig. 1, and slightly more than the steepest gradient on which Minetti et al. (2002) measured runners' speeds.

Because d^2p/dm^2 is positive, extra distance at a shallower gradient is more than made up for by the benefit of avoiding very steep ascents. Thus, because the gradient is zero in the valley bottom on any route, the routes in Fig. 5 all curve so as to have less variation in gradient than on the straight-line route; furthermore, a route crossing the valley bottom must have an inflection point there (Fig. 5b). The greater the value of d^2p/dm^2 —that is, the greater the curvature of the pace function in Fig. 4—the more curved will be the optimal route shown in Fig. 5. The curvature of a route also keeps the route gradient below the critical value for the relevant pace function throughout each of the routes shown in Fig. 5a,b; for example, the route for the quadratic function (19) (for which $m_{c+} = 0.285$) in Fig. 5a approaches the endpoint at an angle of 26° to the contours, giving a subcritical gradient of 0.219, whereas the straight-line route has a gradient of $0.5 \cos 45^\circ \approx 0.35$ at the endpoint.

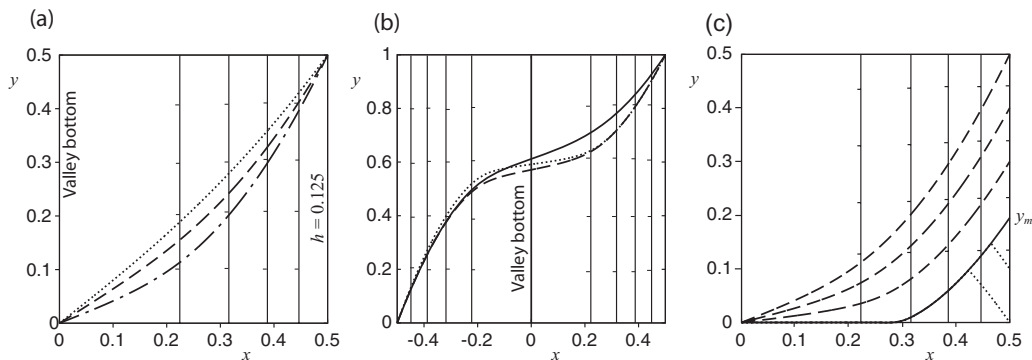


Figure 5. Fastest routes for runners with various pace functions on a slope of parabolic cross-section. (a) Routes from a point (0,0) in a valley bottom (altitude $h = 0$) to a point (0.5,0.5) at altitude $h = 0.125$, where the slope is 0.5: dotted line: Scarf's function (11); dashed line: Davey, Hayes, and Norman's (1994) function (16); dash-dotted line: quadratic function (19). (b) Routes between points where the slope is 0.5 on opposite sides of the valley: solid line: CR function (21); dashed line: CS function (22); dotted line: CT function (23). (c) Routes from point (0,0) in a valley bottom to various points (0.5, y_B) at altitude $h = 0.125$, for a runner with pace function (19): dashed curves climb at subcritical gradients to endpoints at $y_B = 0.3$, $y_B = 0.4$, and $y_B = 0.5$; solid curve is the limiting case, taking line of steepest ascent as far as $x = m_c = 0.285$, then climbing at the critical gradient to $y_B = y_m = 0.195$; dotted curves are upper sections of zigzag routes to $y_B = 0$ and $y_B = 0.1$. In all three panels, contours at intervals of 0.025 units are shown as thin lines with tags pointing downhill. CR, cubic, race records; CS, cubic, skilled descender; CT, cubic, timid descender.

Comparing Figs. 4b and 5b, pace functions that are similar for uphill or downhill running yield routes that are similar on their respective uphill or downhill sections. The timid descender needs to take the most curved route downhill, but the CR runner covers the greatest proportion of the cross-slope distance on the downhill section, because he/she is less averse to steep uphill running than are the other two runners.

Comparison of uphill routes to different destinations

We now consider routes from a valley bottom to a variety of endpoints on a contour where $m_{\perp} = 0.5$ for a runner with the quadratic pace function (19). Thus the startpoint has coordinates (0, 0), and the endpoints are at (0.5, y_B) with cross-slope distances in the range $0 \leq y_B \leq 0.5$. On the upper part of these legs (where $x > m_{c+}$), the slope is steeper than the critical gradient $m_{c+} = 0.285$.

For sufficiently large values of y_B (including the case of $y_B = 0.5$ already illustrated in Fig. 5a), subcritical optimal routes may be found, with C in equation (37) decreasing as y_B decreases. The limiting case as $C \rightarrow 0$ is a maximal-steepness route that ascends perpendicular to the contours (with $y' = 0$ in equation (37)) up to $x = m_{c+}$ and then follows a smooth curve at the critical gradient (with $p - mdp/dm = 0$ in (37)); its endpoint location is $y_B = y_m$, and a numerical value of y_m is calculated following equation (52). For $0 \leq y_B < y_m$, zigzagging is required. An optimal route with a single sharp change of direction may be constructed by following the limiting-case route until $y = (y_B + y_m)/2$, and then by reversing the cross-slope component of travel but continuing to ascend at the critical gradient (still satisfying equation (37) with $C = 0$). Examples of all the preceding route types are shown in Fig. 5c.

For a general Pennine slope, equation (2) for the route gradient becomes

$$m = \frac{m_{\perp}(x)}{\sqrt{1 + y'^2}}, \quad (50)$$

so that a critical-gradient route satisfies

$$\frac{dy}{dx} = \pm \sqrt{\left(\frac{m_{\perp}(x)}{m_c}\right)^2 - 1}. \quad (51)$$

On the parabolic slope (49), the solution of equation (51), starting from the critical-gradient contour $x = m_c$ at $y = y_0$, is

$$y - y_0 = \pm \frac{1}{2} \left\{ x \sqrt{\left(\frac{x}{m_c}\right)^2 - 1} - m_c \ln \left(\frac{x}{m_c} + \sqrt{\left(\frac{x}{m_c}\right)^2 - 1} \right) \right\}. \quad (52)$$

The previously described limiting-case route is given by equation (52) with $y_0 = 0$ and $m_c = 0.285$, yielding an endpoint at $y_m = 0.195$ when $x = 0.5$. The section of a zigzag route above the turning point is given by equation (52) with the negative sign and $y_0 = y_B + y_m$.

Similar routes may be calculated for downhill legs on Pennine slopes. For a leg that crosses a valley, starting at $x < -m_c$ and ending at $x > m_c$, zigzagging may be done on either or both the uphill and downhill sections, if required.

Conical hills

The simplest case of an axisymmetric hill is a conical hill, that is, a hill with a uniform slope m_{\perp} ; Fig. 1b shows North Berwick Law as a reasonable approximation to this idealized landform. Llobera and Sluckin (2007) refer to such a hill as “Mount Conicus” and only consider legs ending or starting at its summit. I consider this singular case first and then compute optimal routes for some more general example legs using the three cubic pace functions. No length scale exists on a conical hill; together with axisymmetry, this property allows us to arbitrarily fix the (r, θ) coordinates of one chosen point on the route (e.g., the start or the point of nearest approach to its summit).

Routes to and from a summit

In this case, any optimal route is of maximal-steepness type: a straight line to or from the summit if the slope is less than the respective uphill or downhill critical gradient, or a spiral ascending or descending at the critical gradient if the slope is steeper than this gradient. These routes were found by Llobera and Sluckin (2007), except that their local optimization algorithm yielded a downhill route zigzagging along sections of spiral with opposite chirality. Llobera and Sluckin (2007) did not indicate the exact nature of the spiral; it is a logarithmic spiral, because it has a constant angle $\phi = \cos^{-1}(m_c/m_{\perp})$ to the radial line from the summit. Fixing the start-point of an ascent or the endpoint of a descent at $(r = 1, \theta = 0)$, the equation of the spiral is

$$r = \exp(\pm b\theta) \quad (53)$$

(with the \pm sign allowing for spirals of either chirality), where

$$b = \frac{|m_c|}{\sqrt{m_{\perp}^2 - m_c^2}}. \quad (54)$$

The logarithmic spiral has the curious property of having a finite arc length but an infinite number of windings around the origin from any point on the spiral; because runners take discrete steps of nonzero length, however, they obviously do not circulate around the summit of a hill an infinite number of times!

Computing optimal routes on conical hills

For a route from start-point (r_A, θ_A) to endpoint (r_B, θ_B) , I write $\Delta\theta = \theta_B - \theta_A$ and may take $0 \leq \Delta\theta \leq \pi$ without loss of generality. If

$$m_{\perp} > |m_c| \quad \text{and} \quad \Delta\theta \leq \frac{1}{b} \left| \ln \frac{r_A}{r_B} \right| \quad (55)$$

(where b is found from equation (54) using the appropriate [uphill or downhill] critical gradient), a maximal-steepness route is optimal, because a more direct route is steeper than the critical-gradient logarithmic spiral. Exceptionally, if $r_B = r_A e^{b(2n\pi \pm \Delta\theta)}$ for some integer n , this optimal route is a logarithmic spiral with no zigzag. More generally, a zigzag route consisting of sections of logarithmic spiral is required. This route may be constructed with a single sharp change of direction at a point (r_i, θ_i) , where logarithmic spirals of opposite chiralities from the start- and endpoints intersect:

$$r_i = r_A \exp(b(\theta_i - \theta_A)) = r_B \exp(-b(\theta_i - \theta_B)), \quad (56)$$

in which the second equality yields the solution

$$\theta_i = \frac{1}{2} \left(\theta_A + \theta_B - \frac{1}{b} \ln \frac{r_A}{r_B} \right). \quad (57)$$

When equation (55) is not satisfied, the optimal route is subcritical and may involve only ascent or only descent, or may climb to a point of maximum altitude at some location (r_0, θ_0) and then descend to an endpoint. A route that climbs and then descends is tangential to the contours at (r_0, θ_0) , and the shooting method cannot be used to calculate such a route, because a route along a contour also satisfies equation (40), despite not being optimal (see section “Some general remarks about solutions of the Euler–Lagrange equation”). However, the tangential condition implies that

$$r' = 0, m = 0, p = p_0 \quad \text{at} \quad r = r_0, \quad (58)$$

which fixes

$$C = p_0 r_0 \quad (59)$$

in equation (40). To compute the route, I first use Taylor expansions about $r = r_0$, $\theta = \theta_0$, $r' = 0$, $m = 0$ in equation (40) to obtain

$$r - r_0 \sim \frac{1}{2} \frac{p_0 r_0}{p_0 + p_0'' m_{\perp}^2} (\theta - \theta_0)^2 \quad (60)$$

(where p_0'' is the value of $d^2 p/dm^2$ at $m = 0$) and then use this result to initiate a numerical integration both forward and backward from (r_0, θ_0) . The location (r_0, θ_0) is not known a priori; however, for a given pace function and a given value of m_{\perp} , every subcritical route collapses onto the same curve by a linear rescaling and a rotation. Thus, I do the calculation once for arbitrary (r_0, θ_0) ; then, when start- and endpoints of a leg are specified, I seek a pair of points with the required values of r_A/r_B and $\Delta\theta$ on the calculated curve and finally rescale and rotate to fit the start-point coordinates (r_A, θ_A) .

The subcritical curve is calculated with $p_0 = 1$, $r_0 = 1$, and $\theta_0 = \pi/2$, for the three cubic pace functions on conical hills with $m_{\perp} = 0.2$ and $m_{\perp} = 0.4$; the former hill is less steep than any of the critical gradients except for m_{c-} of CT, whereas the latter is steeper than all critical gradients except that no m_{c+} exists for CR. Graphs of the curves appear in Fig. 6, through limited azimuthal ranges for the sake of clarity. We may determine their asymptotic behavior as $r \rightarrow \infty$ as follows, applying the arguments separately to the uphill and downhill arms of the curves. If $m_{\perp} < m_{c+}$, the quantity $p - mdp/dm$ remains strictly positive for any direction of travel; in particular, this quantity approaches a positive constant as $r' \rightarrow \infty$ (i.e., as $m \rightarrow m_{\perp}$), so that a solution of equation (40) with $C > 0$ has

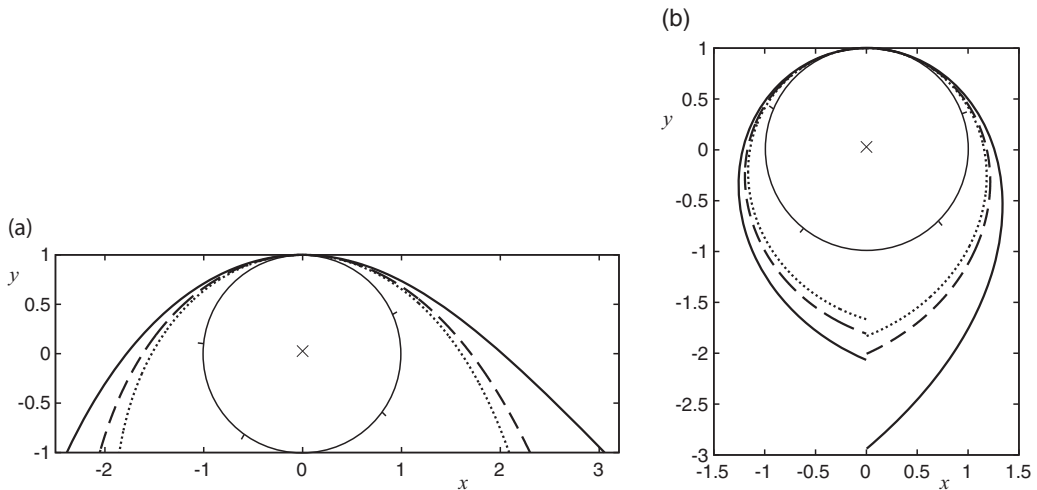


Figure 6. Subcritical curves on conical hills of slope (a) $m_{\perp} = 0.2$ and (b) $m_{\perp} = 0.4$ for pace functions CR (solid line), CS (dashed line), and CT (dotted line). The cross marks the hill summit, thin tagged circles are contours, and the direction of travel is anticlockwise (so the routes here are uphill on the right and downhill on the left). CR, cubic, race records; CS, cubic, skilled descender; CT, cubic, timid descender.

$$\frac{r^2}{\sqrt{r^2 + r'^2}} \rightarrow \text{constant as } r \rightarrow \infty \text{ (and } r' \rightarrow \infty). \quad (61)$$

Thus, the asymptote is a straight line ($r^2/\sqrt{r^2 + r'^2} = \text{constant}$) in the direction of steepest ascent ($r' = \infty$). Therefore, legs of any steepness possible on such a hill can be covered on the subcritical curve. In contrast, if $m_{\perp} > m_c$ with $C > 0$ in equation (40),

$$p - m \frac{dp}{dm} \rightarrow 0 \text{ as } r \rightarrow \infty; \quad (62)$$

thus, the asymptote is the critical-gradient spiral, and legs steeper than this cannot be covered on the subcritical curve. For instance, the CR curve for $m_{\perp} = 0.4$ in Fig. 6b has a straight line asymptote on its uphill arm because CR has no uphill critical gradient, but a spiral asymptote on its downhill arm because CR has $|m_c| = 0.218 < m_{\perp}$.

One unexpected feature of Fig. 6 is that the CS curves are closer to the CT curves than to the CR curves both uphill and downhill, whereas Fig. 4b suggests that optimal routes for CS and CR should be fairly similar downhill. However, equation (60) shows that the route is determined by p_0'' near the point where it is tangential to the contours. Now, p_0'' is proportional to the coefficient of m^2 in the cubic pace functions, and the value of this coefficient in CS is closer to that in CT than in CR.

Example legs on conical hills

I first consider legs on which the optimal routes involve only climbing or only descending, on a hill with slope $m_{\perp} = 0.4$. Fig. 7a illustrates two examples of the former, both involving a “quarter-turn” round the summit ($\Delta\theta = \pi/2$), but with the leg from A_2 starting twice as far down the hill as that from A_1 . All optimal routes from A_1 are subcritical, with the routes for pace functions CS and CT being almost identical, corresponding to the similarity between these pace functions for uphill running. On the steeper leg from A_2 , CS and CT require maximal-steepness zigzag routes, with the difference between the routes determined by the difference between their values of m_{c+} ; CR has no uphill critical gradient, and hence yields a more direct subcritical route.

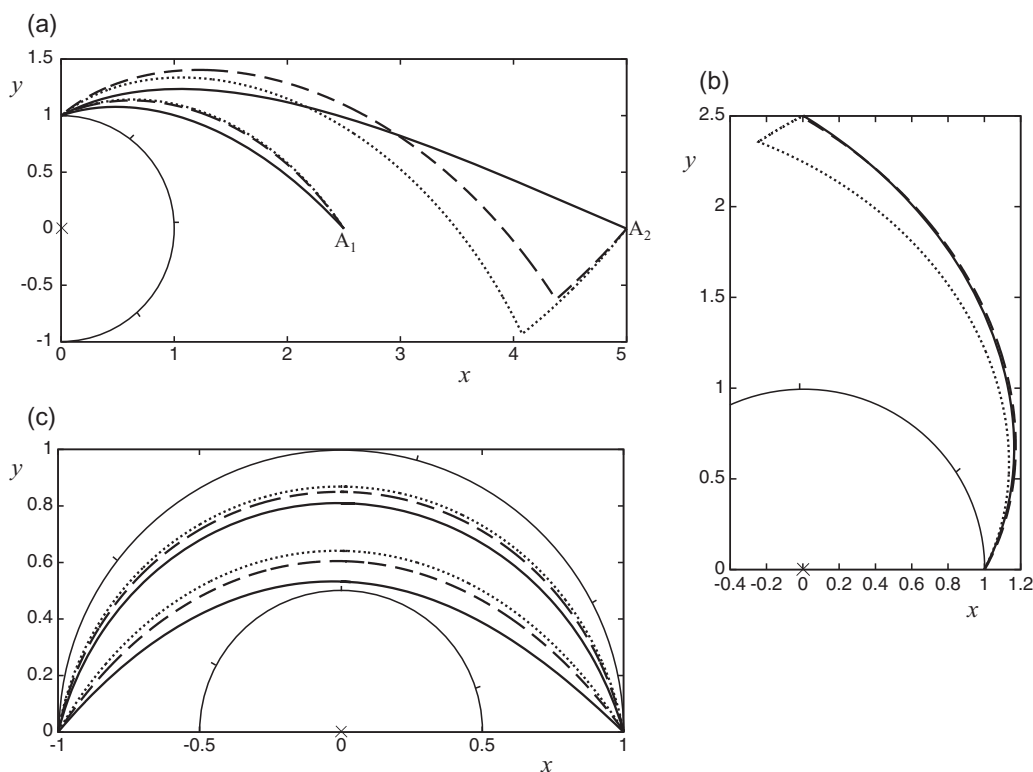


Figure 7. Optimal routes on conical hills, with pace functions CR (solid line), CS (dashed line), and CT (dotted line). (a) Legs starting at A_1 ($r = 2.5$, $\theta = 0$) and A_2 ($r = 5$, $\theta = 0$), with a common endpoint at ($r = 1$, $\theta = \pi/2$) on a hill with $m_{\perp} = 0.4$. (b) Leg from ($r = 1$, $\theta = 0$) to ($r = 2.5$, $\theta = \pi/2$) on a hill with $m_{\perp} = 0.4$. (c) Leg from ($r = 1$, $\theta = 0$) to ($r = 1$, $\theta = \pi$) on hills with $m_{\perp} = 0.2$ (the three routes that pass closer to the summit) and $m_{\perp} = 0.4$ (the outer three routes). The cross marks the hill summit, thin tagged circular arcs are contours, and the direction of travel is counterclockwise. CR, cubic, race records; CS, cubic, skilled descender; CT, cubic, timid descender.

The descent leg shown in Fig. 7b is the reverse of the leg from A_1 in Fig. 7a but is shown differently due to the convention of traveling counterclockwise around axisymmetric hills; and the descent routes are certainly not the reverse of the ascent routes. Downhill, the pace functions CR and CS are rather similar (Fig. 4b), and, correspondingly, the optimal routes for these pace functions are rather similar on the descent leg (although the CS route is of the maximal-steepness type, with a change of direction just before its endpoint).

I now turn to legs between points at the same altitude; that is, with $r_A = r_B$, for which the optimal route is subcritical for any pace function. In particular, consider the case of $\Delta\theta = \pi$, which is an orienteer's classic "over or around" route choice problem between points on opposite sides of a hill. On British hills, the choice often might be influenced by the presence of paths leading to the summit or sheep tracks that tend to contour round hills; but with a purely gradient-dependent pace function, the optimal route involves elements of both "over" and "around." It reaches its maximum altitude at some radial distance r_0 , where it is tangential to the contours, and may be characterized by the ratio $\rho = r_0/r_A$; the amount of ascent involved is $1 - \rho$, as a proportion of the altitude difference between the start-point and the summit. Fig. 7c furnishes graphs of optimal routes for the three cubic pace functions on conical hills of slopes $m_{\perp} = 0.4$ and $m_{\perp} = 0.2$. Unlike the uphill and downhill legs in Fig. 7a,b, optimal routes for a leg that requires equal amounts of ascent and descent are principally determined by the parameter p_0'' of the pace function, as described in relation to Fig. 6.

Fig. 8a portrays the radial distance ratio ρ for pace functions CR and CT and a range of hill slopes. With $m_{\perp} = 0.2$, the optimal routes already are “more around than over” ($\rho > 0.5$) for both pace functions (also see Fig. 7c), and ρ increases to more than 0.8 on a hill of slope $m_{\perp} = 0.4$. For gentler slopes ($m_{\perp} < 0.1$), ρ has an approximately quadratic dependence on m_{\perp} . To explain this, recall that optimal routes deviate from a straight line only because of the nonlinearity of pace functions (see section “Naismith’s rule: a linear pace function”); for small gradients m , the quadratic term dominates this nonlinearity of $p(m)$ in its Taylor expansion about $m = 0$, while ρ is a measure of the deviation of the optimal route from a straight line.

General axisymmetric hills

Now consider axisymmetric hills on which the slope m_{\perp} varies with distance r from the summit. We may classify such hills as either *domes*, on which $m_{\perp} \rightarrow 0$ as $r \rightarrow 0$, or *peaks*, with a nonzero slope as one approaches the summit; a route crossing the summit may be optimal on a dome but never on a peak. The slope may increase or decrease with distance from the summit or may be nonmonotonic; for example, a Gaussian profile $h \propto e^{-kr^2}$ might be a good approximation to the shape of some natural hills. Thus, several radii may exist for which the slope passes through the value m_{c+} or m_{c-} , dividing the hill into annuli where the slope is supercritical or subcritical for uphill or downhill running. Therefore, computations of optimal routes may be complicated, but the general principles established previously still apply:

- optimal routes may be of maximal-steepness or subcritical;
- a maximal-steepness route is perpendicular to the contours where the slope is below the critical value, but spirals (and possibly zigzags) at the critical gradient in a region of supercritical slope; and
- a subcritical route may be computed by the shooting method, but if it involves both ascent and descent, this calculation requires shooting in both directions from the point where it is tangential to the contours; the radial coordinate r_0 of this point is initially unknown, but the tangential condition still yields $C = p_0 r_0$ so that only one unknown parameter exists.

First, consider legs climbing to an endpoint at a summit. The optimal route has maximal-steepness: On a peak with $m_{\perp} > m_{c+}$ as $r \rightarrow 0$, this is a spiral around the summit (as on a steep conical hill), but on either a more gently sloping peak or a dome, the approach to the summit is straight. However, the slope may increase with radial distance, and if the start-point is in a region of supercritical slope, then the route starts by spiraling even if the final approach to the summit is straight. For example, consider a parabolic dome on which the slope is

$$m_{\perp} = r, \quad (63)$$

where we use a unit of distance similar to that for parabolic Pennine slopes, measuring the distance from the hill summit to the contour where the terrain gradient is unity. Given this numerical equality of slope and radial coordinate, uphill and downhill critical circles of radii m_{c+} and m_{c-} exist such that the slope is less than the respective critical value only within the circle. A maximal-steepness route on this dome is a straight line $\theta = \theta_c$ between the critical circle and the summit, and follows the critical-gradient spiral

$$\theta - \theta_c = \pm \left(\sqrt{\frac{r^2}{m_c^2} - 1} - \cos^{-1} \left(\frac{m_c}{r} \right) \right) \quad (64)$$

outside the critical circle; the arc length along this spiral from radius r to the critical circle is

$$s_c = \frac{r^2 - m_c^2}{2m_c}. \quad (65)$$

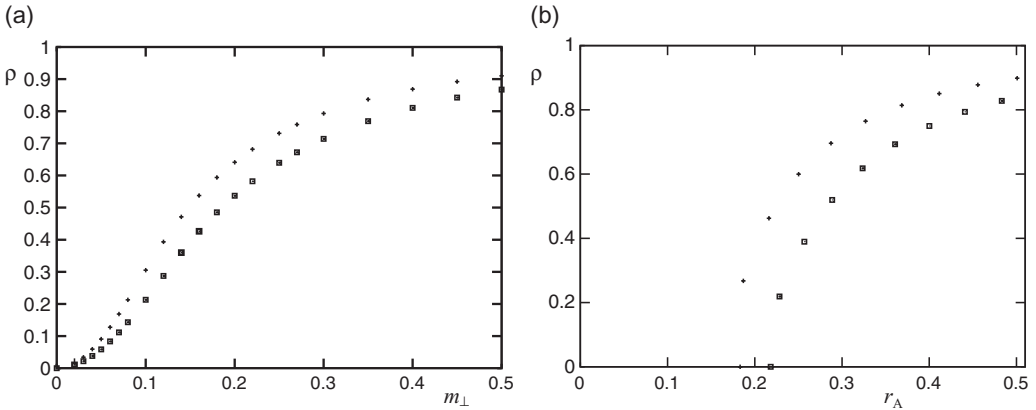


Figure 8. Ratio $\rho = r_0/r_A$, where r_0 is the distance of the nearest approach to a summit on an optimal route for legs from $(r = r_A, \theta = 0)$ to $(r = r_A, \theta = \pi)$: (a) as a function of slope m_{\perp} for conical hills; and (b) as a function of start/endpoint distance r_A on a parabolic dome. Pace functions are CR (squares) and CT (crosses). CR, cubic, race records; CT, cubic, timid descender.

Similar considerations apply to descent legs from a start-point at a summit; however, if the summit slope is subcritical, a runner may set off straight toward the endpoint and then needs to zigzag if that endpoint is in a region of supercritical slope. This difference between uphill and downhill behavior is similar to that noted by Llobera and Sluckin (2007).

Now consider “over or around” route choice legs, with $\Delta\theta = \pi$. On a peak, the optimal route always avoids the summit, but on a dome, the maximal-steepest route straight over the summit is fastest if it does not involve climbing or descending more steeply than the critical gradients; that is, if both $m_{\perp} < m_{c+}$ where $r \leq r_A$, and $m_{\perp} < |m_{c-}|$ where $r \leq r_B$. If these conditions are not satisfied, a subcritical route is required on a dome. I computed such routes for cases where $r_A = r_B$ on a parabolic dome: Fig. 8b graphs the ratio $\rho = r_0/r_A$ (where r_0 is the distance of nearest approach to the summit) in terms of r_A ; for comparison with Fig. 8a, note that the abscissa is also the slope at the start- and endpoints. For a start-point on steep ground (large r_A), the value of ρ is close to that on a conical hill with the same slope as at the present start-point: The steep ground induces runners to take a route at a small angle to the contours, so they never sample the gentler ground nearer the summit and behave much as they would do on a conical hill. In contrast, a start-point on more gently sloping ground induces runners to take a route closer to the summit, where the gradient is even more favorable and induces an even more direct route; hence, ρ undergoes a rapid decrease as the slope at the start-point decreases toward the lesser of the critical gradients.

Conclusions

I consider the problem of finding the fastest route between given points in hilly terrain, where the speed of travel depends only on the gradient along a route. The aim is to gain a theoretical understanding of the principles of route choice rather than to provide a recipe for solving specific route choice problems in real terrains. The latter requires a discrete representation of space and a computational algorithm such as those employed by Hayes and Norman (1984) or Arnet (2009); here one considers idealized topographies that can be represented by continuous functions. This representation facilitates analysis by the calculus of variations: First integrals of the Euler–Lagrange equation appear in equations (37) and (40), which are the basis of all my example calculations.

The pace functions here (specifying how pace depends on route gradient) must also be regarded as idealizations: They are an uneasy compromise between a desire to model realistically the available data on

runners' pace up and down hills (because the sport of orienteering motivates this work) and the need for a variety of simple pace functions with different properties to see how pace function affects route choice. A key property of pace functions is the possible existence of critical gradients for uphill and/or downhill running: These are the gradients at which a runner would maximize the rate at which height is gained or lost. The available empirical evidence is ambiguous as to whether runners have an uphill critical gradient within the range of gradients normally encountered in races, and, accordingly, one of the pace functions used in my example calculations has no uphill critical gradient. Data for downhill running are rather sparse but generally suggest a downhill critical gradient around 0.2.

A linear pace function always results in a straight-line route being optimal. Nonlinearity of $p(m)$, specifically with $d^2p/dm^2 > 0$, means that traversing two route sections at different gradients is slower than running the same total distance at the average gradient; thus, optimal routes tend to curve so as to reduce the variation of gradient along them. Larger values of d^2p/dm^2 tend to be associated with smaller values of critical gradients and a greater tendency to avoid steep climbs or descents.

Optimal routes are of two types, distinguished mathematically by the value of the constant of integration C in equations (37) and (40). Subcritical routes (with $C \neq 0$) ascend or descend less steeply than the respective critical gradient and do not take the line of direct ascent. Maximal-steepness routes (with $C = 0$) take the line of direct ascent where the slope is not steeper than the critical gradient, but on steeper slopes they curve so as to ascend or descend at the critical gradient. Where no smooth solution of the Euler–Lagrange equation exists between given start- and endpoints, a zigzag route consisting of sections of critical-gradient curve is required. Maximal-steepness routes depend only on the topography and the critical gradients, but subcritical routes may be influenced by the form of the pace function throughout the range of gradients between m_{c-} and m_{c+} . However, on legs involving ascent followed by descent (e.g., between points at the same altitude on opposite sides of an axisymmetric hill), the principal determinant appears to be the value of d^2p/dm^2 at $m = 0$.

This analysis potentially has much wider applicability than orienteering competitions: it applies to the minimization of the integral (1) whenever the cost function p depends only on route gradient (or on the directional derivative along the route of any smooth function of position). For example, a new road or railway may be routed to minimize fuel consumption of vehicles. While there may be less uncertainty over the cost function in this case than there is over runners' pace, the restriction of the methods used here to idealized landforms means that detailed planning would ultimately require computations with discretized geographical data.

Some scope exists to extend my analysis: Other topographies, such as an elliptical hill or a river valley with a nonzero gradient along the valley bottom, may be tractable, and dependence on isotropic factors (runnability and slope) could be included in pace functions. However, such increased complexity may yield diminishing returns in terms of theoretical understanding while still not providing accurate solutions to route choice problems on real terrain.

Acknowledgements

This work was first developed as an undergraduate project in 1978 when I was at St. Andrews University, studying mathematics but competing in orienteering events in my spare time. It was put aside, as the equations seemed intractable at the time; so I am grateful to Philip Scarf, whose recent articles again ignited my interest in the subject. I am also grateful to the reviewers whose comments on an earlier draft of this article helped me to refine it.

Notes

- Note for non-British readers: In certain parts of northern England, mountains are referred to as *fell*s (from an Old Norse word for mountain). A strong tradition exists of races up and down these fells, and this sport has become known as *fell running* in England. In Wales, it is more often called *mountain running*; in Scotland it is *hill running*, although the hills of Scotland are higher than the mountains of Wales.

- 2 Minetti et al. (2002) measure the metabolic cost of running on a wide range of uphill and downhill gradients; pace is proportional to metabolic cost if metabolic power is assumed constant. Calculations of pace made in this way are in accordance with uphill race records, but correlate very poorly with downhill race records, presumably because factors other than metabolism control downhill pace. Hence, Minetti et al.'s data furnish a meaningful pace function only for uphill running, and the coefficients in equation (19) are very different from those suggested by Rees (2004) and Llobera and Sluckin (2007).

References

- Alexander, R. M. (2000). "Walking and Running Strategies for Humans and Other Mammals." In *Biomechanics in Animal Behaviour*, 49–57, edited by P. Domenici and R. W. Blake. Oxford: BIOS Scientific.
- Arnet, F. (2009). "Arithmetical Route Analysis with Examples of the Long Final Courses of the World Orienteering Championships 2003 in Switzerland and 2005 in Japan." *Scientific Journal of Orienteering* 17, 4–21.
- Balström, T. (2002). "On Identifying the Most Time-Saving Walking Route in A Trackless Mountainous Terrain." *Danish Journal of Geography* 102, 51–8.
- Davey, R. C., M. Hayes, and J. M. Norman. (1994). "Running Uphill: An Experimental Result and Its Applications." *Journal of the Operational Research Society* 45, 25–9.
- Dijkstra, E. W. (1959). "A Note on Two Problems in Connection with Graphs." *Numerische Mathematik* 1, 269–71.
- Hayes, M., and J. M. Norman. (1984). "Dynamic Programming in Orienteering: Route Choice and the Siting of Controls." *Journal of the Operational Research Society* 35, 791–6.
- Kay, A. (2006). "Orienteering Route Choice: Running at the Speed of Light?" *Mathematics Today* 42, 138–43.
- Langmuir, E. (1984). *Mountaineering and Leadership*. Edinburgh: The Scottish Sports Council.
- Llobera, M., and T. J. Sluckin. (2007). "Zigzagging: Theoretical Insights on Climbing Strategies." *Journal of Theoretical Biology* 249, 206–17.
- Margaria, R. (1938). "Sulla Fisiologia E Specialmente Sul Consumo Energetico Della Marcia E Della Corsa A Varia Velocità Ed Inclinazione Del Terreno." *Atti Della Accademia Nazionale Dei Lincei* 7, 299–368.
- Miller, H. J., and S. A. Bridwell. (2009). "A Field-Based Theory for Time Geography." *Annals of the Association of American Geographers* 99, 49–75.
- Minetti, A. E. (1995). "Optimum Gradient of Mountain Paths." *Journal of Applied Physiology* 79, 1698–703.
- Minetti, A. E., C. Moia, G. S. Roi, D. Susta, and G. Ferretti. (2002). "Energy Cost of Walking and Running at Extreme Uphill and Downhill Slopes." *Journal of Applied Physiology* 93, 1039–46.
- Puu, T. (1997). *Mathematical Location and Land Use Theory: An Introduction*. Berlin: Springer.
- Puu, T., and M. Beckmann. (1999). "Continuous Space Modelling." In *Handbook of Transportation Science*, 269–310, edited by R. W. Hall. Boston: Kluwer.
- Rees, W. G. (2004). "Least-Cost Paths in Mountainous Terrain." *Computers and Geosciences* 30, 203–9.
- Scarf, P. (1998). "An Empirical Basis for Naismith's Rule." *Mathematics Today* 34, 149–51.
- Scarf, P. (2007). "Route Choice in Mountain Navigation, Naismith's Rule, and the Equivalence of Distance and Climb." *Journal of Sports Sciences* 25, 719–26.
- Scarf, P. (2008). "A Mathematical Excursion in the Isochronic Hills." *Mathematics Today* 44, 163–7.
- Townshend, A. D., C. J. Worringham, and I. B. Stewart. (2010). "Spontaneous Pacing during Overground Hill Running." *Medicine and Science in Sports and Exercise* 42, 160–9.
- Warntz, W. (1957). "Transportation, Social Physics, and the Law of Refraction." *The Professional Geographer* 9, 2–7.