Route Choice in Hilly Terrain

ANTHONY KAY

Department of Mathematical Sciences, Loughborough University,
Loughborough, Leicestershire, LE11 3TU

Tel: +44 1509 222878

Fax: +44 1509 223969

Email: A.Kay@Lboro.ac.uk

Abstract

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 we consider only the effect of climbing or descent on running speed. If a runner's pace p (the reciprocal of running speed) varies linearly with gradient m, the straight-line route is always fastest. This may be a reasonable approximation at gentle gradients, but various studies have suggested that a nonlinear formulation for $p(m)$, with $\mathrm{d}^2p/\mathrm{d}m^2 > 0$, is more appropriate. As a result, there is are critical gradients for both ascent and descent, such that optimal routes will never climb or descend more steeply than the critical gradient. Thus it may be necessary to zigzag when climbing or descending a hill where the slope is steeper than the runner's critical gradient.

In general, 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. Results are presented for idealised landforms: a hillside with straight contours and various axisymmetric hills. If the slope varies, the general tendency is for optimal routes to curve so that the variation of the gradient along the route is less than on a straight-line route.

Results based on different formulations for pace as a function of gradient are compared.

1 Introduction

The sport of orienteering requires participants to determine the fastest route along a leg between two given points in terrain which is often hilly or even mountainous. This is an example of a *minimum cost path* problem: find the path in 2-dimensional space which minimises the integral

$$t_{\mathrm{AB}} = \int_{\mathrm{A}}^{\mathrm{B}} p \, \mathrm{d}s \tag{1}$$

where A and B are given points, p is a *cost field* (Miller & Bridwell, 2009) and s is distance. In our application, 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 the map (as opposed to distance over the 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 positive uphill;
- The gradient of the terrain, denoted $m_\perp$ hereafter;
- The *runnability* of the terrain, which is the effect of vegetation or uneven ground in reducing the runner's speed relative to that on a smooth path.

Route gradient is related to terrain gradient by

$$m = m_\perp \sin \psi \,, \tag{2}$$

where ψ is the angle between the route and the contours. The route gradient is anisotropic (dependent on the direction of travel), while the other two factors are isotropic. Terrain gradient is included separately from route gradient, since it is slower to traverse horizontally across a cliff face than to run across a flat field. Nevertheless, in this paper we are concerned only with the effect of ascent and descent on route choice, and we shall ignore the two isotropic factors: thus $p = p(m)$. This will still be a reasonable first approximation for many mountain navigation events, as well as giving better insight into the effect of a single factor.

Minimum cost path problems in geographical and economic applications have often been addressed by discretising space. For the orienteering application, Hayes & Norman (1984) have taken this approach, imposing a 250-metre grid on a complex area of mountain terrain and using a dynamic programming algorithm to find the fastest route; pace was taken to depend on gradient and runnability in a rather crude way. Twenty-five years of subsequent development in computer power have enabled Arnet (2009) to impose a 1-metre grid on an admittedly smaller area of very complex terrain, taking account of all three factors described above, and using Dijkstra's algorithm to optimise the path (or so it appears from Arnet's description, although he refers to a "minimum spanning tree"). It may thus appear that the orienteering route choice problem has been solved by the application of sufficient computing resources; but this approach does not yield much theoretical insight, and furthermore the least satisfactory aspect of the analysis by Arnet (2009) is his treatment of route gradient.

In urban transport applications, Puu & Beckmann (1999) and Miller & Bridwell (2009) argue for the advantages of treating space as continuous. If continuous modelling is useful in this context, even though cities may naturally be considered as networks of discrete nodes and edges, it is certainly to be favoured in the setting of an open hillside. Scarf (2008) has developed a decision-making 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. However, the classical method for route optimisation in continuous space is to solve the Euler-Lagrange equation in the calculus of variations. Puu & Beckmann (1999) base their analysis on this, although they only give solutions for isotropic cost fields. In fact, the minimum cost path problem for an isotropic field is formally identical to geometrical optics, since Fermat's Principle dictates that light travels between given points in an isotropic medium by the fastest route (Warntz, 1957; Kay, 2006).

Although Fermat's Principle does not apply in anisotropic media, the Euler-Lagrange equation is still valid, and will form the basis of our analysis. Following Scarf (2008), we shall consider idealised topographies that allow progress by analytical methods and give good insight into the principles of route choice. Probably the simplest non-trivial 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 gradient varies only in the direction up and down the slope. Good approximations to such slopes are common in the Pennine hills of Northern England: see Figure 1. For convenience we shall henceforth refer to such hillsides as *Pennine slopes*. Another simple case, sharing the attribute of having terrain gradient independent of one spatial coordinate, is an axisymmetric hill, i.e. with concentric circular contours. This case was considered by Scarf (2008), who suggested Kirk Fell and Middle Fell in the English Lake District as examples of such hills; however, Binsey is a better Cumbrian example, while the volcanic plugs of Eastern Scotland provide some of the most nearly axisymmetric hills in Britain: see Figure 2. We shall examine optimal routes on a conical hill, i.e. with uniform terrain gradient, before generalising to cases where the terrain gradient varies with distance from the summit.

The fastest route will obviously depend crucially on the assumed form of the *pace function*, the dependence of pace on route gradient. Thus in Section 2 of this paper we shall briefly review the literature on runners' pace as a function of gradient, and propose some plausible formulations for the pace function. A feature of realistic pace functions is that there are critical gradients for both ascent and descent, such that it is never advantageous to choose a route that is steeper than the relevant critical gradient. This phenomenon is discussed in section 3. The main mathematical development begins in Section 4, where it is shown how the Euler-Lagrange equation yields the critical gradient; simplified forms of the Euler-Lagrange equation are then derived for Pennine slopes and

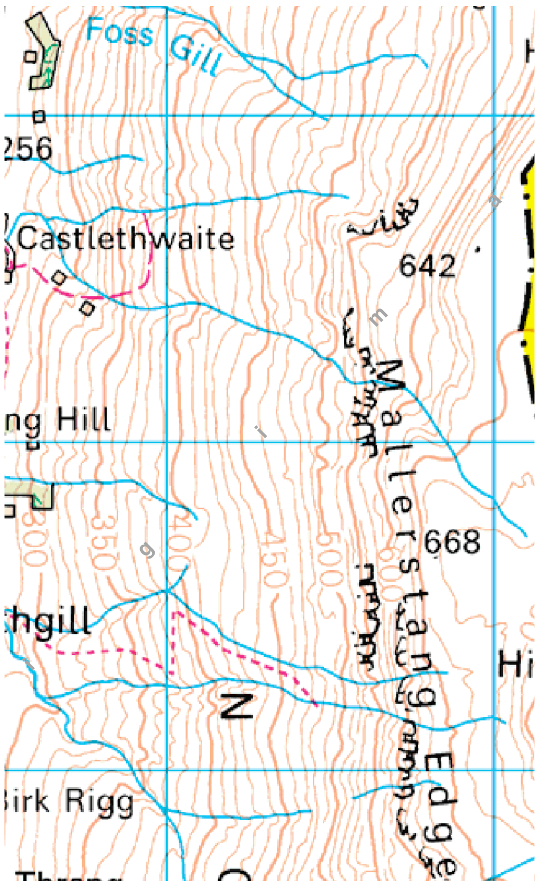

Figure 1: The slope below Mallerstang Edge in the North Pennines, with (approximately!) parallel but unevenly spaced contours. Gradients measured due west of the 668-metre spot height range from below 0.2 (between the 400- and 450-metre contours) to more than 0.5 (between the 500- and 550-metre contours). Map extract is © Crown copyright Ordnance Survey. All rights reserved.

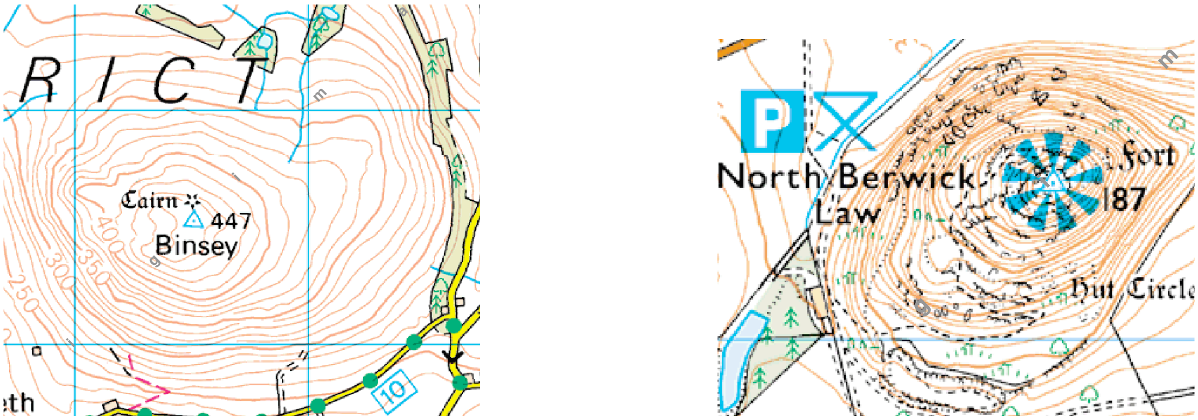

Figure 2: Examples of hills with approximately circular contours. Left: Binsey (Cumbria), with gradients generally less than 0.2. Right: North Berwick Law (East Lothian), a volcanic plug with gradients around 0.5. Map extracts are © Crown copyright Ordnance Survey. All rights reserved.

axisymmetric hills. The next three sections contain detailed calculations of optimal routes on Pennine slopes, conical hills and other axisymmetric hills, respectively. We draw some conclusions and make suggestions for further work in Section 8.

2 Dependence of Runner's Pace on Route Gradient

A variety of formulae has been suggested for the pace function $p(m)$. We shall first discuss the well-known Naismith's Rule, which was first proposed in 1892 (Langmuir, 1984), before considering other rules based on experimental evidence on uphill running. Finally we formulate empirical rules intended to be valid for both uphill and downhill running.

2.1 Naismith's Rule and a refinement thereof

If the pace on level terrain is denoted p_0, then Naismith's Rule may be written

$$p = p_0(1 + \alpha m) \qquad [m \geq 0]; \tag{3}$$

Scarf (1998) gives the value of the constant α as 7.92. As explained by Scarf (2008), the rule does not explicitly account for descent: it can either be taken 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. The latter assumption, which may be formulated as

$$p = p_0(1 + \alpha m H(m)) \tag{4}$$

where H is the Heaviside unit-step function, was used by Arnet (2009). However, it is more reasonable to suppose that the pace function should be differentiable for all m. We then remove the restriction $m \geq 0$ from Naismith's rule (3) and take it to be the linear approximation to the true pace function, valid for sufficiently small (uphill or downhill) route gradients (Davey, Hayes & Norman (1994), herefater DHN). Substituting (3) into (1) and expressing differential distance in terms of differential height gain, $\mathrm{d}s = \mathrm{d}h/m$, we obtain

$$\begin{aligned} t_{\mathrm{AB}} &= \int_{\mathrm{A}}^{\mathrm{B}} \frac{p_0}{m}\,\mathrm{d}h + \int_{\mathrm{A}}^{\mathrm{B}} p_0\alpha\,\mathrm{d}h \\ &= p_0 \int_{\mathrm{A}}^{\mathrm{B}} \mathrm{d}s + p_0\alpha(h_{\mathrm{B}} - h_{\mathrm{A}})\,. \end{aligned} \tag{5}$$

Although this result was noted by DHN, they did not observe its implication for general route choice problems: since the total height gain $h_{\mathrm{B}} - h_{\mathrm{A}}$ is fixed, t_{AB} is minimised by minimising the distance $\int_{\mathrm{A}}^{\mathrm{B}} \mathrm{d}s$. Hence the straight-line route is always quickest when Naismith's Rule applies; there is no advantage in choosing a longer route at a gentler gradient. We may interpret this as meaning that the 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.

Scarf (1998, 2007) has introduced a refinement of Naismith's rule, which may be written as

$$p = p_0(1 + \alpha m)^{\beta} . \tag{6}$$

This was based on a regression analysis of fell-running records (Scarf, 2007) in which the best fit is obtained with $\beta \approx 1.14$ and $\alpha \approx 8.6$ for men but $\beta \approx 1.16$ and $\alpha \approx 10.6$ for women. Scarf's Rule (6) is intended to take account of fatigue, and so p_0 is now weakly dependent (by a $\beta - 1$ power law) on total distance to be covered. We shall ignore the very small effect that this will have on route choice between fixed points. The main deficiency of Scarf's rule is that, like the original Naismith's rule, it does not account for descent separately from ascent.

2.2 Other uphill pace functions

Alternative formulae, expressed in terms of the angle of slope σ rather than the gradient $m = \tan\sigma$, have been proposed by DHN on the basis of treadmill experiments. They suggest that their data fit either of the formulae

$$p \cos\sigma = p_0 \, \mathrm{e}^{k\sigma} \tag{7}$$

or

$$p = p_0 \, \mathrm{e}^{k\sigma} \tag{8}$$

where $p\cos\sigma$ in the first formula is the pace over the sloping ground. We have re-examined the data from each of DHN's five experiments, using the correlation coefficient obtained in a regression analysis to indicate goodness of fit to Naismith's and Scarf's rules as well as the exponential laws (7) and (8). We find that (8) gives the best fit, but with $k \approx 3.0$ rather than the value of 3.5 given by DHN. Naismith's rule is worst, and Scarf's Rule is usually less good than exponential laws. We shall henceforth refer to (8), in the form

$$p = p_0 \, \mathrm{e}^{k \tan^{-1} m} , \tag{9}$$

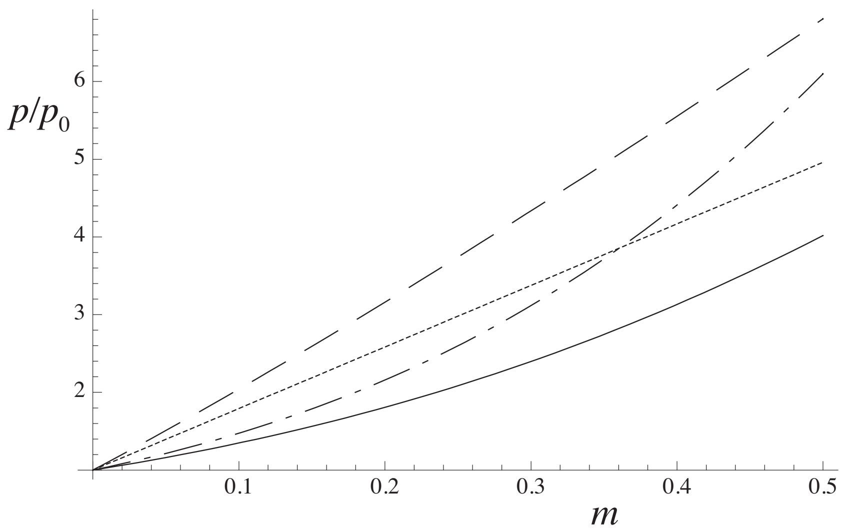

Figure 3: Pace as a function of uphill gradient. Dotted line: Naismith's Rule (3) with $\alpha = 7.92$. Dashed line: Scarf's Rule (6) with $\alpha = 8.6, \beta = 1.15$. Solid line and dash-dotted line: DHN's Rule (9) with $k = 3.0$ and $k = 3.9$ respectively.

as DHN's Rule. This form of pace function is supported by the results of Minetti et al (2002), who did treadmill experiments both uphill and downhill, using a larger number of experimental subjects than DHN. Using the measured energy cost of running, they derived theoretical maximum running speeds on gradients varying from 0.45 to -0.45. Their uphill data give a good fit to DHN's rule (9), but with $k \approx 3.9$, and also show good agreement with race winners' speeds in several uphill races.

Figure 3 is a comparison of the linear Naismith's Rule, Scarf's power-law rule and the exponential DHN rule with $k = 3.0$ or $k = 3.9$, for gradients $0 \leq m \leq 0.5$. It is clear that when we take the linear approximations to the various nonlinear pace functions, they have widely differing values of the "Naismith coefficient" $[\mathrm{d}p/\mathrm{d}m]_{m=0}$. We have not investigated this issue in detail, but there are two likely factors contributing to the discrepancies. Firstly, the failure of Naismith's and Scarf's rules to account for descent may contribute to their much higher values of the Naismith coefficient, compared with rules derived from experiments on uphill running. Secondly, running in controlled conditions on a treadmill may allow a faster pace when going uphill than under race conditions in difficult terrain (Norman, 2004). It has been suggested by Norman (2004) that on mountain terrain the DHN Rule should be used with k equal to the constant α in Naismith's Rule; but with $\alpha = 7.92$ this yields $p/p_0 \approx 39$ when $m = 0.5$, which seems unreasonably slow.

2.3 Pace functions valid for both ascent and descent

Uncertainty over the correct form of pace function for downhill running is even greater than for uphill. In particular, Minetti et al (2002) found that in a sample of downhill races the actual running speeds are considerably slower (by a factor of around 3) than the theoretical maximum: a runner's metabolism is not the limiting factor when running downhill. It seems clear that running speed should attain a maximum at some downhill gradient, probably in the range -0.1 to -0.2 but varying considerably according to the skills of the individual runner. Beyond this consideration, our choice of pace function will be dictated at least partly by mathematical convenience: we require $p(m)$ to be differentiable for all m, including at $m = 0$. So piecewise linear functions such as (4) and the more sophisticated but discontinuous (!) functions $p(m)$ proposed by Langmuir (1984) and Hayes & Norman (1984) for negative m cannot be used.

Rees (2004) has suggested a quadratic formula

$$p = a + bm + cm^2 \tag{10}$$

applicable to both uphill and downhill. His value of b is much smaller than a or c (all coefficients being positive), so that this gives maximum speed on very gentle downward slopes, or even on level ground if we take Rees' suggestion that $b = 0$ is sufficiently accurate. Llobera & Sluckin (2007) (hereafter LS) also proposed a quadratic formula to relate metabolic cost to gradient, but then showed that a quartic would give a more realistic fit to the data. Both Rees and LS based their formulae on data for walkers, so they may not be applicable to competitive runners. Nevertheless, and notwithstanding the observation of Minetti et al (2002) that downhill running speeds depend on factors other than metabolism, a quartic does seem to be suitable to represent pace as a function of gradient for both uphill and downhill running. Such a curve can be fitted to any of the uphill pace functions proposed above, and can be adjusted to account for the great diversity of runners' behaviour on steep downhill slopes: observers of the British fell-running scene will be aware that some runners appear to descend hills exponentially fast (as predicted by DHN's Rule beyond its domain of validity!), while others (including the present author) tend to be more cautious. We have chosen to fit quartic functions to DHN's Rule with $k = 3.9$, using the following criteria: (i) the value of $p(m)$ at $m = 0.5$ (the steepest gradient considered) must be the same in the quartic and in DHN's rule; (ii) the critical gradient (see Section 3 below) must be the same for the two rules. We then choose arbitrarily the negative value of m at which the minimum of pace (maximum speed) occurs: this gradient of fastest pace is taken to be $m = -0.2$ and $m = -0.15$ for two quartics, representing respectively a runner who is skilled at descending steep terrain, and a more timid runner. We now have three items of data, yielding an underdetermined system of equations for the four coefficients in the quartic

$$p = p_0(1 + am + bm^2 + cm^3 + dm^4)\,; \tag{11}$$

the free parameter is then used to adjust the form of $p(m)$ for downhill gradients so as to emphasise the difference between the skilled and timid descenders. The resulting quartics are

$$p = p_0(1 + 3.635m + 9.441m^2 + 3.332m^3 + 8.084m^4)\,, \tag{12}$$

$$p = p_0(1 + 3.460m + 10.663m^2 - 1.152m^3 + 13.567m^4) \tag{13}$$

respectively for the skilled and timid descenders: see Figure 4.

It may be argued that we should use a quartic formula for all our subsequent calculations. However, for some legs involving only uphill running we shall make comparisons

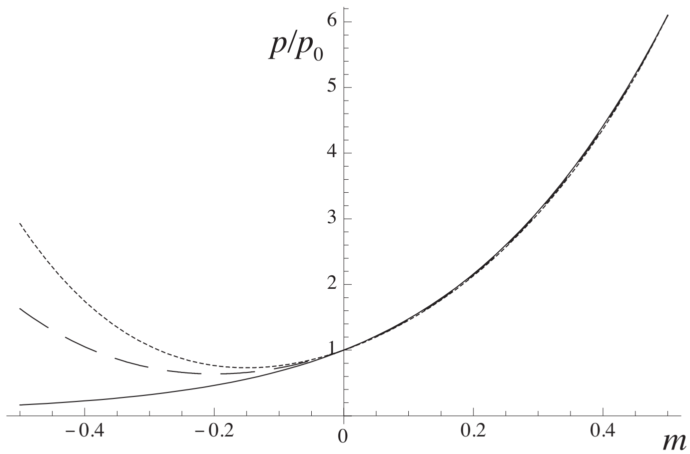

Figure 4: Pace as a function of gradient. Solid line: DHN's Rule (9) with $k = 3.9$. Dashed line: quartic fit (12) for "skilled descender". Dotted line: quartic fit (13) for "timid descender".

of optimal routes calculated assuming each of the nonlinear uphill rules shown in Figure 3.

3 The critical gradient

Consider a *direct ascent leg*, i.e. with the straight-line route between start and end point being perpendicular to the contours, on a Pennine slope of uniform gradient. The height difference h_{AB} between start and end points is fixed, and the time taken on the straight-line route is

$$t_{\mathrm{AB}} = h_{\mathrm{AB}} \frac{p}{m} \,. \tag{14}$$

This is minimised when

$$\frac{\mathrm{d}p}{\mathrm{d}m} = \frac{p}{m} \,. \tag{15}$$

The solution of (15) for a given pace function $p(m)$ gives a *critical gradient* m_c which is optimal for ascending. If the terrain gradient $m_\perp$ is less than m_c, the straight-line

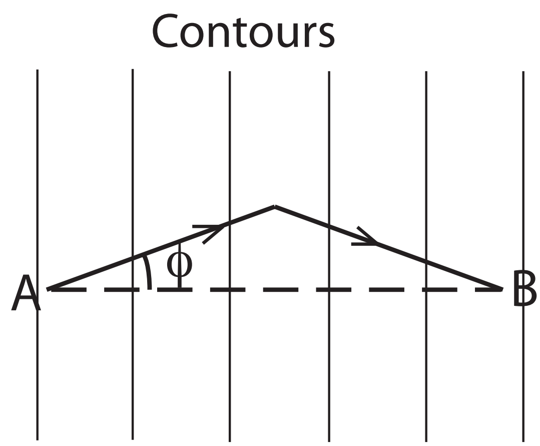

Figure 5: Optimal route from A to B on uniform slope with gradient $m_\perp > m_c$.

route will be fastest on a direct ascent leg; but if $m_\perp > m_c$, a runner will minimise t_{AB} by finding a route that ascends at the critical gradient. In general, such a route will not be a smooth curve and so would not be obtained in a numerical integration of the Euler-Lagrange equation as described in Section 4 below. For instance, on a uniform slope with $m_\perp > m_c$ an optimal route to a point directly above the starting point could be as shown in Figure 5, with two sections each at an angle $\phi = \cos^{-1}(m_c/m_\perp)$ to the straight-line route (although a zig-zag route with any number of sections at this angle to the direct route would be equally fast).

DHN appear to have been the first to observe the existence of a critical gradient. They calculated the critical slope angle σ_c for their exponential pace functions (7) and (8); for (8) they found that

$$\sin 2\sigma_c = \frac{2}{k}\,, \tag{16}$$

which in terms of gradient is

$$m_c = \frac{1}{2}\left(k - \sqrt{k^2-4}\right)\,. \tag{17}$$

Critical gradients also exist for other nonlinear pace functions. Applying (15) to Scarf's

Rule (6), we find a critical gradient

$$m_c = \frac{1}{\alpha(\beta - 1)}, \tag{18}$$

while our quartic pace functions were constructed to have the same critical gradient as DHN's rule with $k = 3.9$.

The critical gradient phenomenon was analysed in detail by LS. Although they sought to minimise the total metabolic cost to a walker rather than the time taken, the structure of their problem is essentially the same as ours. Indeed, despite their rather different mathematical approach, they arrived at an equation of identical form to (15) for the critical gradient. Their dissipation function, like our pace functions, is a function of gradient with positive first and second derivatives when the gradient is positive. Pace and metabolic dissipation both increase with increasing gradient: a runner is slowed down ($\mathrm{d}p/\mathrm{d}m > 0$) and has to expend more energy on a steeper hill. Moreover, these effects become more severe on very steep hills: the runner's retardation and extra energy expenditure per unit increase in gradient become greater as the slope increases ($\mathrm{d}^2p/\mathrm{d}m^2 > 0$). It is this positive curvature of the pace function (Figure 3 above) and the dissipation function (Figure 3 in LS) that yields a critical gradient; a less curved function, e.g. Scarf's Rule, will tend to yield a higher value of critical gradient. LS also find a critical gradient for downhill walking, since their dissipation function reaches a minimum for a certain negative gradient and then increases with continued positive curvature as the downhill gradient steepens further. This will also apply to our quartic pace functions, and we shall denote a downhill critical gradient as m_{c-}.

The numerical values of uphill and downhill critical gradients, above which it is never optimal to climb or descend, will be a crucial feature in determining optimal routes. LS suggest $m_c = \tan 16° \approx 0.287$ for walkers. Scarf's rule with $\alpha = 8.6$ and $\beta = 1.15$ yields the very high value $m_c \approx 0.775$ from (18), due to the very weak curvature of the function (6). DHN found $m_c = 1/2.8 \approx 0.357$ using the rule (7). Our reanalysis of their data using rule (9) gives a slightly higher value $m_c \approx 0.382$ if we take $k = 3.0$ in (17); however, with $k = 3.9$ as obtained from the data of Minetti et al (2002) for "elite athletes practicing endurance mountain racing" we obtain $m_c \approx 0.276$, which is similar to LS's value and is the value used in constructing our quartic pace functions. [Note that a runner obeying the DHN rule (9) with $k \leq 2.0$, i.e. who is slowed considerably less than an average runner by steep gradients, has no critical gradient according to (17), and so will find that the straight-line route is always best on a direct ascent leg.]

For downhill walking, LS find $m_{c-} = \tan 12.4° \approx 0.220$, whereas our rather arbitrary extrapolations of pace functions to downhill gradients yield the higher values of 0.320 for (12) (skilled descender) and 0.266 for (13) (timid descender); only the latter case agrees with LS's observation that $|m_{c-}| < m_c$. Nevertheless, higher values are consistent with the observed behaviour of many British fell runners.

Our numerical values of critical gradients lend some support to the suggestion by Balstrøm (2002) that walkers will avoid terrain with a gradient in excess of 0.3. LS suggest that human trails on hills should take optimal routes, in particular zigzagging where the gradient is greater than a critical value. Evidence from British hills is ambiguous: on North Berwick Law, with a terrain gradient in excess of 0.6 on its steepest side, the path to the summit zigzags at a gradient generally less than 0.2 (see Figure 2), somewhat gentler than any of the critical values listed above. In contrast Largo Law, a volcanic plug on the opposite side of the Firth of Forth from North Berwick Law, has a path ascending almost directly to its south top at a gradient in excess of 0.4: see figure 6. The hill is not well known to walkers, but has a well established annual hill race. The path (which is not shown on Ordnance Survey maps) may have been worn by runners, who have a greater propensity to take direct routes than walkers; the gorse on the lower slopes has probably also discouraged zigzagging.

Figure 6: Largo Law from the south, showing the path to the summit. Image Copyright Richard Webb (downloaded from http://www.geograph.org.uk/photo/12932, 19 October 2009).

4 Optimisation by Calculus of Variations

4.1 The Euler-Lagrange equation and the critical gradient

We begin by imposing a general orthogonal coordinate system (ξ, η) on the plane. We then set

$$f(\xi, \eta, \eta') = p \, \frac{\mathrm{d}s}{\mathrm{d}\xi} \, , \tag{19}$$

where primes denotes derivatives with respect to ξ and total differentials indicate variations along the route. Then (1) becomes

$$t_{\mathrm{AB}} = \int_{\mathrm{A}}^{\mathrm{B}} p \, \mathrm{d}s = \int_{\mathrm{A}}^{\mathrm{B}} f \, \mathrm{d}\xi \, , \tag{20}$$

which is minimised on a route $\eta(\xi)$ satisfying the Euler-Lagrange equation

$$\frac{\partial f}{\partial \eta} - \frac{\mathrm{d}}{\mathrm{d}\xi} \left(\frac{\partial f}{\partial \eta'} \right) = 0 \, . \tag{21}$$

Note that for a coordinate ξ with a scale factor of unity, the quantity $\mathrm{d}s/\mathrm{d}\xi$ in (19) is the secant of the angle between the 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:

$$m = \frac{\mathrm{d}h}{\mathrm{d}s} = \frac{\mathrm{d}h}{\mathrm{d}\xi}\,\frac{\mathrm{d}\xi}{\mathrm{d}s}\,, \tag{22}$$

in which

$$\frac{\mathrm{d}h}{\mathrm{d}\xi} = \frac{\partial h}{\partial \xi} + \eta'\,\frac{\partial h}{\partial \eta}\,. \tag{23}$$

From (22) and (23) we may obtain

$$\frac{\partial m}{\partial \eta} = \frac{\partial}{\partial \eta}\left(\frac{\mathrm{d}h}{\mathrm{d}\xi}\right)\frac{\mathrm{d}\xi}{\mathrm{d}s} - m\,\frac{\mathrm{d}\xi}{\mathrm{d}s}\,\frac{\partial}{\partial \eta}\left(\frac{\mathrm{d}s}{\mathrm{d}\xi}\right)\,, \tag{24}$$

$$\frac{\partial m}{\partial \eta'} = \frac{\partial h}{\partial \eta}\,\frac{\mathrm{d}\xi}{\mathrm{d}s} - m\,\frac{\mathrm{d}\xi}{\mathrm{d}s}\,\frac{\partial}{\partial \eta'}\left(\frac{\mathrm{d}s}{\mathrm{d}\xi}\right)\,. \tag{25}$$

We can now calculate the derivatives required in (21) using (19), noting that p is a function of m and using the results (24) and (25) where required. Firstly,

$$\begin{aligned}\frac{\partial f}{\partial \eta} &= \frac{\mathrm{d}p}{\mathrm{d}m}\,\frac{\partial m}{\partial \eta}\,\frac{\mathrm{d}s}{\mathrm{d}\xi} + p\,\frac{\partial}{\partial \eta}\left(\frac{\mathrm{d}s}{\mathrm{d}\xi}\right)\\ &= \frac{\mathrm{d}p}{\mathrm{d}m}\,\frac{\partial}{\partial \eta}\left(\frac{\mathrm{d}h}{\mathrm{d}\xi}\right) + \left(p - m\,\frac{\mathrm{d}p}{\mathrm{d}m}\right)\frac{\partial}{\partial \eta}\left(\frac{\mathrm{d}s}{\mathrm{d}\xi}\right)\,. \qquad (26)\end{aligned}$$

Similarly,

$$\frac{\partial f}{\partial \eta'} = \frac{\mathrm{d}p}{\mathrm{d}m}\,\frac{\partial h}{\partial \eta} + \left(p - m\,\frac{\mathrm{d}p}{\mathrm{d}m}\right)\frac{\partial}{\partial \eta'}\left(\frac{\mathrm{d}s}{\mathrm{d}\xi}\right)\,. \tag{27}$$

We now define our 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

$$\frac{\partial}{\partial \eta}\left(\frac{\mathrm{d}h}{\mathrm{d}\xi}\right) = 0\,. \tag{28}$$

The consequent vanishing of the first terms on the right-hand sides of (26) and (27) means that the Euler-Lagrange equation (21) will be satisfied on a route on which m has a constant value such that

$$p - m\,\frac{\mathrm{d}p}{\mathrm{d}m} = 0\,. \tag{29}$$

The condition (29) is just a rearrangement of the critical gradient equation (15). We have shown that our method based on the calculus of variations is consistent with the critical-gradient theory developed by DHN and LS, but we have not yet solved the fastest-route problem since we have taken no account of the locations of the start and end points of a leg. If the terrain gradient is below the critical values, a critical-gradient route will obviously be impossible; but even if the terrain is steeper than critical, solutions of the Euler-Lagrange equation between many pairs of points will give optimal routes at sub-critical gradients. If no 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.

4.2 The Euler-Lagrange equation on Pennine slopes and axisymmetric hills

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

$$\frac{\mathrm{d}s}{\mathrm{d}x} = \sqrt{1 + y'^2} \qquad \text{and} \qquad \frac{\mathrm{d}s}{\mathrm{d}r} = \sqrt{1 + r^2\theta'^2} \tag{30}$$

on the respective topographies. Thus f as defined by (19) has no explicit dependence on y or θ, so that Euler's equation (21) has the first integrals

$$\frac{\partial f}{\partial y'} = C \qquad \text{and} \qquad \frac{\partial f}{\partial \theta'} = C \tag{31}$$

in the respective cases, with the constant C to be determined. Using (30) in (27), these equations become

$$\frac{y'}{\sqrt{1 + y'^2}} \left(p - m \, \frac{\mathrm{d}p}{\mathrm{d}m} \right) = C \tag{32}$$

and

$$\frac{r^2\theta'}{\sqrt{1 + r^2\theta'^2}} \left(p - m \, \frac{\mathrm{d}p}{\mathrm{d}m} \right) = C \, . \tag{33}$$

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 will certainly be true for (x, y) as defined above on a Pennine slope, since an optimal route will not involve climbing such a slope and then coming back down further 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 would not a single-valued function of r. On the other hand, $r(\theta)$ will be single-valued (except for direct ascent legs, treated separately below), so we note that

$$\theta' \equiv \frac{\mathrm{d}\theta}{\mathrm{d}r} = \left(\frac{\mathrm{d}r}{\mathrm{d}\theta}\right)^{-1} \equiv \frac{1}{r'}\,, \tag{34}$$

which transforms (33) to

$$\frac{r^2}{\sqrt{r^2 + r'^2}}\left(p - m\,\frac{\mathrm{d}p}{\mathrm{d}m}\right) = C\,. \tag{35}$$

Equations (32) and (35) are to be solved with boundary conditions given by the coordinates of the start and end points,

$$y(x_{\mathrm{A}}) = y_{\mathrm{A}} \quad \text{and} \quad y(x_{\mathrm{B}}) = y_{\mathrm{B}} \tag{36}$$

or

$$r(\theta_{\mathrm{A}}) = r_{\mathrm{A}} \quad \text{and} \quad r(\theta_{\mathrm{B}}) = r_{\mathrm{B}}\,. \tag{37}$$

A quick check on (32) and (35) may be made by noting that the optimal route should be a straight line when the pace is given by Naismith's rule (3). But (3) yields

$$p - m\,\frac{\mathrm{d}p}{\mathrm{d}m} = p_0 \tag{38}$$

so that (32) and (35) reduce to

$$y' = \text{ constant} \qquad \text{and} \qquad \frac{r^2 + r'^2}{r^4} = \text{ constant}, \tag{39}$$

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

In some cases detailed below, the value of C can be deduced *a priori* so that numerical integration of (32) or (35) is straightforward, albeit requiring a numerical evaluation of

y' or r' from (32) or (35) at each step in the integration. However, in many cases C is not known and an iterative "shooting" method is required. An initial estimate of C is made, for instance by evaluating the left-hand side of (32) or (35) for a route leaving point A in the direction of the straight line to B. Given the estimate of C, (32) or (35) may be integrated forward in x or θ; on reaching $x = x_{\mathrm{B}}$ or $\theta = \theta_{\mathrm{B}}$, the value of y or r attained will not in general be the correct value for the end-point. The iteration proceeds by improving the estimate of C and numerically integrating for each such estimate until the computed route hits the end-point to within the desired precision.

4.3 Some general remarks on solutions of the Euler-Lagrange equation

Satisfying the Euler-Lagrange equation (21) is a necessary but not sufficient condition for a minimum of $\int_{\mathrm{A}}^{\mathrm{B}} f \, \mathrm{d}\xi$. We 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, i.e. with $r_{\mathrm{B}} = r_{\mathrm{A}}$, on an axisymmetric hill. The route along the circular contour connecting the points has

$$r = \text{constant}, \quad r' = 0, \quad m = 0, \quad p = p_0, \tag{40}$$

and so satisfies (35) with $C = r_{\mathrm{A}} p_0$; yet this contouring route gives neither a maximum nor a minimum of t_{AB}.

Secondly, consider a Pennine slope with uniform terrain gradient. 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 $y' = \text{constant}$, so (32) is indeed satisfied; and this route certainly is the fastest for a runner with any realistic pace function, with $\mathrm{d}^2p/\mathrm{d}m^2 \geq 0$. However, a hypothetical runner with $\mathrm{d}^2p/\mathrm{d}m^2 < 0$ might find a different route faster. For instance, for a leg ascending at 45° to the contours on a uniform slope of terrain gradient $m_\perp$, the straight-line route gradient is $m = m_\perp/\sqrt{2}$; but a route which goes parallel to 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 is physically unrealistic but theoretically conceivable with $\mathrm{d}p/\mathrm{d}m > 0$ and $\mathrm{d}^2p/\mathrm{d}m^2 < 0$: 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.

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

$$y_\mathrm{B} = y_\mathrm{A} \qquad \text{or} \qquad \theta_\mathrm{B} = \theta_\mathrm{A} \tag{41}$$

respectively. Solutions of (32) and (33) with $C = 0$ and satisfying the boundary conditions (41) are

$$y' = 0 \qquad \text{and} \qquad \theta' = 0. \tag{42}$$

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 will indeed give the shortest time; but if $m_\perp > m_c$ the straight line will in fact yield a local *maximum* of t_AB. In the latter case, the solution of (32) and (33) with $C = 0$ which yields a minimum of t_AB is

$$p - m\,\frac{\mathrm{d}p}{\mathrm{d}m} = 0, \tag{43}$$

i.e. a critical-gradient route (which will need to zigzag in order to satisfy the boundary conditions).

Where a zigzagging critical-gradient route is indicated, there is an infinite variety of such routes, each taking the same time but with different numbers and locations of the sharp changes of direction. LS obtain a unique (modulo chirality) optimal zigzag route by introducing a local optimisation procedure involving an extra criterion which determines the direction taken at each step. We do not adopt such a procedure, but for purposes of illustration our zigzag routes will have a single sharp change of direction. This arbitrary criterion makes them unique modulo chirality.

5 Examples of optimal routes on Pennine slopes

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

$$h = \frac{1}{2}x^2, \qquad m_\perp = x\ : \tag{44}$$

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 the terrain gradient is unity. The numerical equality of terrain gradient 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. Note that if there is a ridge rather than a valley along $x = 0$, the routes calculated below are optimal in the reverse direction.

5.1 Uphill-only legs

We consider legs up a parabolic slope, from a start-point $(x_\mathrm{A}, y_\mathrm{A}) = (0, 0)$ in the valley bottom to end-points at $x_\mathrm{B} = 0.5$ and with various values of y_B, the along-slope displacement. The terrain gradient of 0.5 at the end-points is close to the steepest gradient (apart from crags) in Figures 1 and 2, and slightly more than the steepest gradient on which Minetti et al (2002) measured runners' speeds. We first compare optimal routes found for several of the nonlinear uphill pace functions discussed in Section 2.2, and we then consider in detail how the critical gradient affects route choice on a parabolic slope.

Figure 7 shows fastest routes with three different pace functions on a leg with end-point at $y_\mathrm{B} = 0.5$, so that the straight-line route would be at 45° to the contours. All the routes curve in such a way as to make the gradient along the route more nearly constant than on the straight-line route. This is because of the positive values of $\mathrm{d}^2p/\mathrm{d}m^2$, which mean that extra distance at a shallower gradient will be more than made up for by the benefit of avoiding very steep ascents. The greater the value of $\mathrm{d}^2p/\mathrm{d}m^2$, i.e. the greater the curvature of the pace function in Figure 3, the more curved will be the optimal route shown in Figure 7. The curvature also keeps the route gradient below the critical value

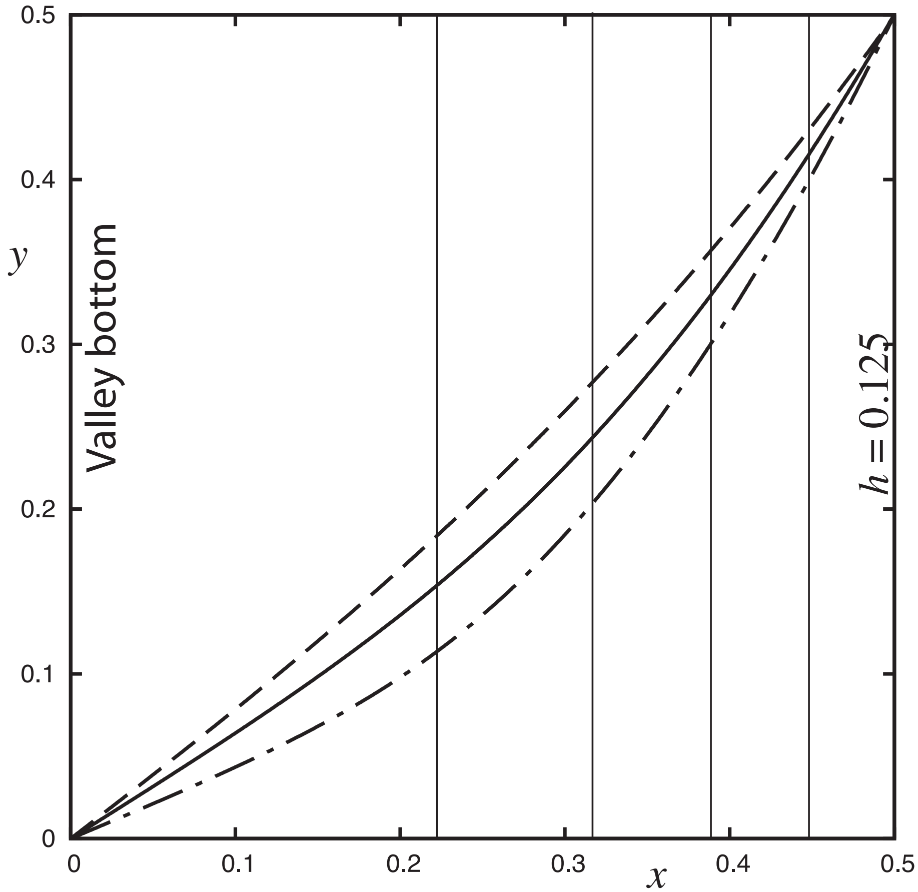

Figure 7: Fastest routes on a diagonal leg up a parabolic slope, from a point $(0, 0)$ in a valley bottom (altitude $h = 0$) to a point $(0.5, 0.5)$ at altitude $h = 0.125$, where the terrain gradient is 0.5. Routes are shown for runners with the three nonlinear pace functions plotted in Figure 3. Dashed line: Scarf's Rule (6) with $\alpha = 8.6, \beta = 1.15$. Solid line and dash-dotted line: DHN's Rule (9) with $k = 3.0$ and $k = 3.9$ respectively. The thin lines are contours at intervals of 0.025 units.

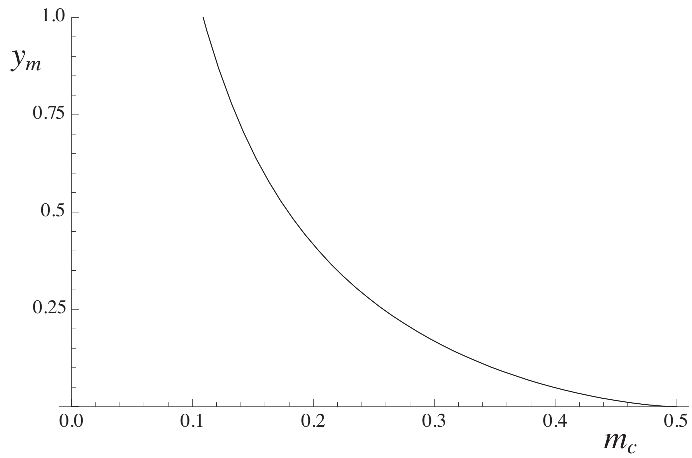

Figure 8: Minimum value of y_{B} for which the fastest route to an end-point at $x_{\mathrm{B}} = 0.5$ on the parabolic slope (44) does not require zigzagging, as a function of runner's critical gradient m_c.

for the relevant pace function throughout each of the routes shown in Figure 7: the straight-line route would have a gradient of $0.5 \cos 45° \approx 0.35$ at the end-point, but the route for DHN's Rule with $k = 3.9$ (for which $m_c = 0.276$) approaches the end-point at an angle of 26° to the contours, giving a gradient of 0.219

On a direct ascent leg to an end-point where the terrain gradient is above the critical value (i.e. with $x_{\mathrm{B}} > m_c$ on the parabolic slope defined by (44)), a runner can ascend perpendicular to the contours while $x \leq m_c$, but above the contour $x = m_c$ he must zigzag upwards at the critical gradient $m = m_c$. On a Pennine slope, equation (2) for the route gradient becomes

$$m = \frac{m_\perp(x)}{\sqrt{1 + y'^2}}\,, \tag{45}$$

so that a critical-gradient route satisfies

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \pm\sqrt{\left(\frac{m_\perp(x)}{m_c}\right)^2 - 1}\,. \tag{46}$$

On the parabolic slope (44) the solution of (46) starting from the critical-gradient contour

$x = m_c$ at $y = y_0$ is

$$y - y_0 = \pm\frac{1}{2}\left\{x\sqrt{\left(\frac{x}{m_c}\right)^2 - 1} - m_c \ln\left(\frac{x}{m_c} + \sqrt{\left(\frac{x}{m_c}\right)^2 - 1}\right)\right\}. \tag{47}$$

Setting $x = x_\mathrm{B} > m_c$ and $y_0 = 0$, (47) can be interpreted as giving the minimum along-slope displacement y_m of an endpoint that can be reached without zigzagging from a start-point $x_\mathrm{A} \le m_c$, as a function of the end-point upslope coordinate x_B. Alternatively, for a fixed end-point altitude, equation (47) gives the dependence of y_m on a runner's critical gradient, and this dependence is plotted for $x_\mathrm{B} = 0.5$ in Figure 8. For instance, with $m_c = 0.276$, we find that $y_m = 0.212$ at $x_\mathrm{B} = 0.5$; so a runner whose pace function is DHN's rule with $k = 3.9$, on a leg starting at $(x_\mathrm{A} \le m_c, y_\mathrm{A} = 0)$ and ending at the contour $x = 0.5$, must zigzag if the end-point is at $|y_\mathrm{B}| < 0.212$. This is shown in Figure 9, where fastest routes from the valley bottom at $(0, 0)$ to various endpoints on the contour $x = 0.5$ are plotted. The solid line shows the route which ascends directly to the critical-gradient contour and then follows the curve (47) to an end-point at $y_\mathrm{B} = y_m = 0.212$: since

$$y - y_0 \sim \frac{2\sqrt{2}}{3\sqrt{m_c}}(x - m_c)^{3/2} + O((x - m_c)^{5/2}) \quad \text{as } x \searrow m_c \tag{48}$$

(from (47)), the direction of travel varies continuously at $x = m_c$ but the curvature of the route is discontinuous at this point. Dashed lines show zigzag routes to $y_\mathrm{B} = 0$ (the direct ascent leg) and $y_\mathrm{B} = 0.1$, with a single turning point (sharp change in direction) at some location (x_t, y_t). From the chiral symmetry indicated by the $\pm$ sign in (47),

$$y_t = \frac{1}{2}(y_\mathrm{B} + y_m), \tag{49}$$

and x_t can then be obtained from numerical solution of (47); this is plotted for $m_c = 0.276$ and $x_\mathrm{B} = 0.5$ in Figure 10. The equation of the curve beyond the turning point is given by (47) with the minus sign and $y_0 = y_\mathrm{B} + y_m$. These routes which ascend directly where the terrain gradient is sub-critical and then along curves satisfying (47) on super-critical terrain are solutions of (32) with $C = 0$, but with $y' = 0$ while $x < m_c$ whereas

(29) applies where $x > m_c$. The dotted curves in Figure 9 are solutions of (32) with positive values of C, giving fastest routes which climb at route gradients below critical to various end-points with along-slope displacements $y_{\mathrm{B}} > y_m$. These were obtained numerically by the shooting method described in Section 4.2 (whereas any attempt to use numerical integration with $C = 0$ to obtain the complete solid curve in Figure 9 will fail because the numerical scheme will not pick up the solution (43), although that curve can be obtained using a very small but positive value of C). Note that when $y_{\mathrm{B}} < y_m$ the route is determined entirely by the value of the critical gradient, whereas when $y_{\mathrm{B}} > y_m$ the form of the pace function $p(m)$ throughout a range of gradients $0 \le m \le m_{\max}$ (for some $m_{\max} < m_c$) is relevant.

5.2 Legs involving downhill running

We first consider downhill-only legs, starting where the terrain gradient is 0.5 and ending in the valley bottom. Such legs are the reverse of the uphill-only legs considered above. However, we adopt the convention that x always increases along a leg so, noting the symmetry of the parabolic terrain profile (44), we take the start point to be at $(x_{\mathrm{A}}, y_{\mathrm{A}}) = (-0.5, 0)$ and the end-points to be at $x_{\mathrm{B}} = 0$ with variable y_{B}. We use the quartic pace functions (12) and (13), representing a "skilled descender" and a "timid descender", respectively. Figure 11 shows optimal routes to an end-point at $y_{\mathrm{B}} = 0.5$, the reverse of the 45° uphill leg in Figure 7; but since the pace functions are not symmetric with respect to uphill and downhill running, the optimal downhill routes are not the reverse of optimal uphill routes. Comparing routes for the two pace functions, the timid descender needs to take a route at a gentler gradient (a more acute angle to the contours) on the steeper part of the slope, and so travels a longer distance.

The critical gradient phenomenon applies to downhill as well as uphill running, so when $|x_{\mathrm{A}}| > m_{c-}$ there is again a minimum value of y_{B} for which no zigzagging is required: this minimum along-slope displacement y_{m-} can again be found from Figure 8. Calculations of zigzagging optimal routes when $y_{\mathrm{B}} < y_{m-}$ proceed exactly as for the

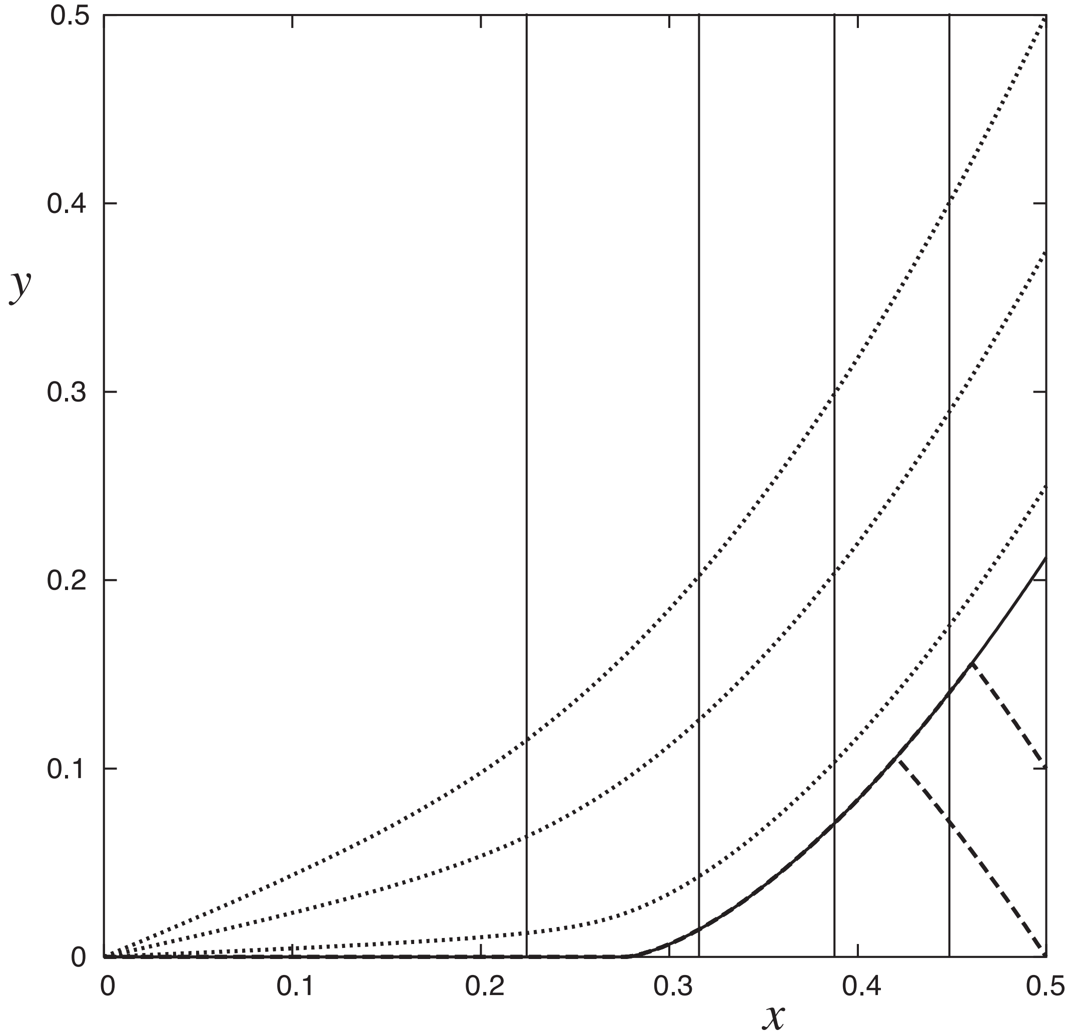

Figure 9: Fastest routes up a parabolic slope, from a point $(0, 0)$ in a valley bottom (height $h = 0$) to various points $(0.5, y_{\mathrm{B}})$ at altitude $h = 0.125$, assuming the pace function to be DHN's Rule with $k = 3.9$. Solid curve is fastest route to $y_{\mathrm{B}} = 0.212$, the smallest value of y_{B} for which for which zigzagging is not required. Dashed curves (coinciding with the solid curve before the sharp change in direction) are zigzag routes to $y_{\mathrm{B}} = 0$ and $y_{\mathrm{B}} = 0.1$. Dotted curves are routes to $y_{\mathrm{B}} = 0.25$, $y_{\mathrm{B}} = 0.375$ and $y_{\mathrm{B}} = 0.5$. Thin lines are contours at intervals of 0.025 units.

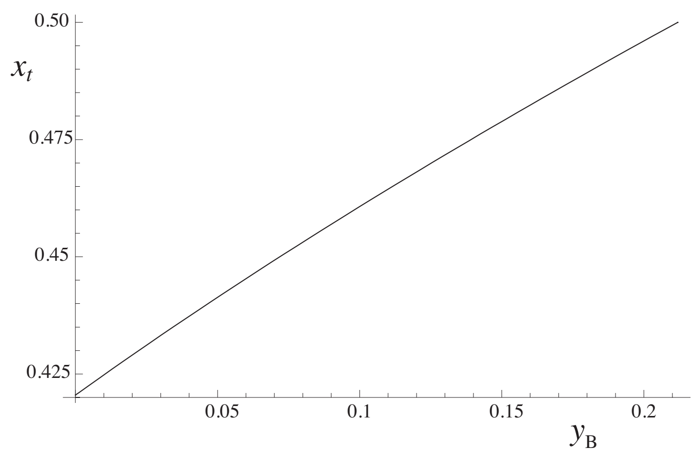

Figure 10: x-coordinate of turning point on legs with $x_B = 0.5$ and $y_B < y_m$ on the parabolic slope (44), for a runner with critical gradient $m_c = 0.276$.

uphill case, since the route then only depends on the critical gradient. For the limiting case $y_B = y_{m-}$, routes are shown on Figure 11 for the two quartic pace functions: the timid descender has a smaller critical downhill gradient and hence a larger value of y_{m-}, as given by Figure 8. Zigzagging will be necessary for a wider range of y_B values for the timid descender than for the skilled descender.

Now consider legs across the valley, starting downhill from the point $(x_A, y_A) = (-0.5, 0)$ and finishing uphill at $x_B = 0.5$, so that the terrain gradient is 0.5 at both the start and end points. The asymmetry of the quartic pace functions with respect to downhill and uphill running ensures asymmetric optimal routes with respect to the valley bottom. Optimal routes to an end-point at $y_B = 1.0$, i.e. with the straight-line route at 45° to the contours as it was for uphill-only and downhill-only examples considered above, are shown in Figure 12 for the two quartic pace functions. Because the more timid descender needs to travel at a gentler gradient on the downhill part of the route, he is forced to climb more steeply (which is slower) on the uphill part.

One feature apparent on Figure 12 is that optimal routes have an inflection point where they cross the valley bottom, with a minimum of y' (a maximum of the angle

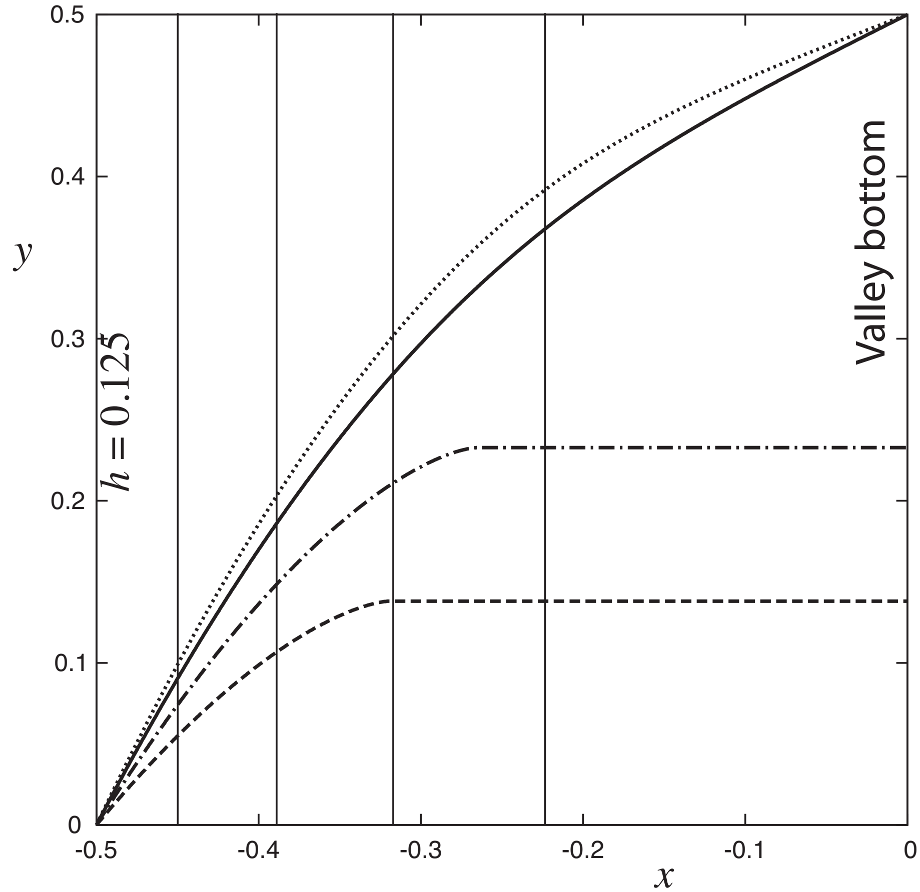

Figure 11: Optimal routes down a slope with a parabolic profile, from a start-point where the perpendicular gradient is 0.5 to end-points in the valley bottom. Solid and dotted curves are fastest routes on a leg at 45° to the contours, for skilled and timid descenders (pace functions (12) and (13)), respectively. Dashed and dash-dotted curves are routes with the minimum along-slope displacement for which zigzagging is not required by these respective runners. Thin lines are contours at intervals of 0.025 units.

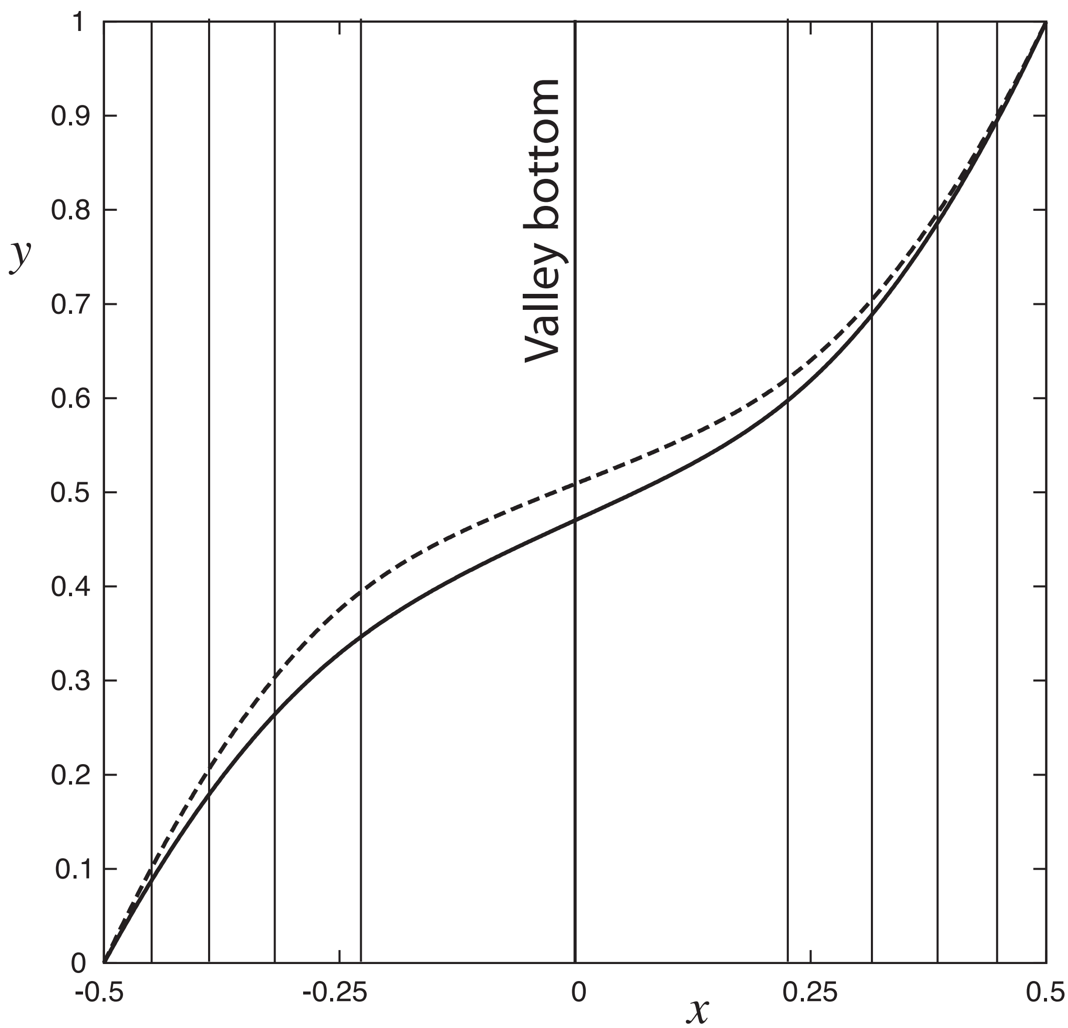

Figure 12: Optimal routes across a valley with a parabolic profile, between points on opposite sides of the valley where the perpendicular gradient is 0.5. The solid and dashed curves are fastest for pace function (12) ("skilled descender") and (13) ("timid descender") respectively. Thin lines are contours at intervals of 0.025 units.

between the route and the contours). At the valley bottom, $m = 0$ for any direction of travel, and a Maclaurin expansion yields

$$p - m\,\frac{\mathrm{d}p}{\mathrm{d}m} \sim p_0 - \frac{1}{2}\,p_0''m^2 + \dots \tag{50}$$

where p_0'' is the value of $\mathrm{d}^2p/\mathrm{d}m^2$ at $m = 0$. All our nonlinear pace functions have $p_0'' > 0$; thus $(p - m\,\mathrm{d}p/\mathrm{d}m)$ has a local maximum at $m = 0$, so from (32) y' must have a local minimum.

The critical gradient phenomenon comes into play on both the downhill and uphill sections of a cross-valley leg. If both $x_\mathrm{A} < -m_{c-}$ and $x_\mathrm{B} > m_c$, zigzagging will be required if the along-slope displacement satisfies $y_\mathrm{B} < y_{m-} + y_m$. It is then immaterial whether the zigzagging is done while travelling downhill and/or uphill; route sections will be given by (47) (with m_c replaced by m_{c-} for the downhill section, and with appropriate values of y_0 inserted) where the terrain gradient is steeper than critical, and will be perpendicular to the contours where the terrain gradient is sub-critical. In the limiting case where $y_\mathrm{B} = y_{m-} + y_m$, fastest routes can be constructed by concatenating the routes to $y_\mathrm{B} = y_{m-}$ in Figure 11 with a route parallel to the solid curve in Figure 9, since by construction both quartic pace functions have the same critical uphill gradient as the pace function used in the latter diagram.

6 Optimal routes on a conical hill

6.1 Routes to and from the summit

The simplest case of a hill with concentric circular contours is a conical hill, i.e. with a uniform value of terrain gradient $m_\perp$; Figure 2 shows North Berwick Law as a reasonable approximation to this idealised landform. LS refer to such a hill as “Mount Conicus”, and only consider routes ending or starting at the summit. If the terrain gradient is less than the respective uphill or downhill critical gradient, the optimum route to or from the summit is a straight line. However, if the terrain gradient is greater than critical, the fastest route is a spiral ascending or descending at the critical gradient (LS's local optimisation algorithm indicates that the downhill route should zigzag, with successive sections of the zigzag formed from sections of the spiral with opposite chirality; but our procedure would not prefer such a route over a continuous spiral). LS do not indicate the exact nature of the spiral; it is in fact a logarithmic spiral, since it has a constant angle $\phi = \cos^{-1}(m_c/m_\perp)$ to the radial line from the summit. For ascending, its equation is

$$r = r_{\mathrm{A}} \exp\left(\pm b_+(\theta - \theta_{\mathrm{A}})\right) \tag{51}$$

for a spiral starting at the point $(r = r_{\mathrm{A}}, \theta = \theta_{\mathrm{A}})$ (with the $\pm$ sign allowing for spirals of either chirality), where

$$b_+ = \frac{m_c}{\sqrt{m_\perp^2 - m_c^2}}\,; \tag{52}$$

for descending, (51) has b_+ replaced by b_-, which is defined by replacing m_c with m_{c-} in (52). 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; but since runners take discrete steps of non-zero length, they will obviously not circulate around the summit of a hill an infinite number of times!

6.2 General routes on steep hills

Routes to and from the summit are a singular case. We now consider more general routes on a conical hill with $m_\perp > \max\{m_c, m_{c-}\}$. Given start and end points with coordinates $(r_{\mathrm{A}}, \theta_{\mathrm{A}})$ and $(r_{\mathrm{B}}, \theta_{\mathrm{B}})$ respectively, the form of the fastest route can only depend on the azimuthal displacement $\Delta\theta \equiv \theta_{\mathrm{B}} - \theta_{\mathrm{A}}$ and the radius ratio $R \equiv r_{\mathrm{B}}/r_{\mathrm{A}}$, since there is no length scale if $m_\perp$ is uniform. We may take $0 \le \Delta\theta \le \pi$ without loss of generality. Depending on the values of R and $\Delta\theta$, the optimal route will be one of three types, designated "logarithmic spiral", "tangential" and "intermediate".

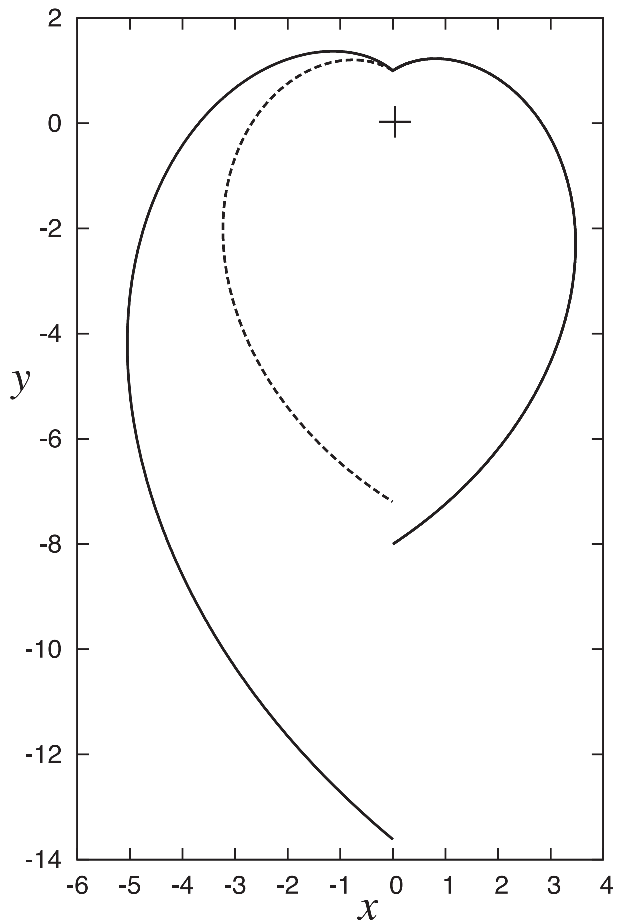

Figure 13: Sections of logarithmic spiral routes for ascent and descent at the critical gradients for pace functions (12) (solid line) and (13) (dashed line) on a conical hill with gradient $m_{\perp} = 0.5$. The direction of travel is anticlockwise, so the curve on the right side is for ascent while the curves on the left side are for descent. The cross marks the summit of the hill.

6.2.1 Logarithmic spiral routes

If

$$R < e^{-b_+ \Delta\theta} \quad \text{or} \quad R > e^{b_- \Delta\theta}\,, \tag{53}$$

the fastest route from A to B will involve zigzagging along sections of logarithmic spiral in order to avoid climbing or descending at a gradient steeper than critical. [Exceptionally, if $R = e^{b_\pm(2n\pi \pm \Delta\theta)}$ for some integer n, a logarithmic spiral route with no zigzag but involving more than one half circuit around the summit will be possible.] Without loss of generality, we take the point $(r = 1, \theta = \pi/2)$ to be the end-point of any ascending leg and the start-point of any descending leg. Logarithmic spiral routes climbing and descending to and from this point at the critical gradients are shown in Figure 13 for the quartic pace functions (12) and (13) on a conical hill with $m_\perp = 0.5$ (fairly typical of North Berwick Law). Since both pace functions are constructed to have the same uphill critical gradient $m_c = 0.276$, there is a single curve on the right side of the figure showing a spiral ascending through an azimuthal range of π, whereas the left side shows spirals descending at the critical gradients of 0.320 and 0.266 for the "skilled descender" and "timid descender", respectively. If the start-point $(r_\mathrm{A} = 1/R, \theta_\mathrm{A} = \pi/2 - \Delta\theta)$ for an ascent or the end-point $(r_\mathrm{B} = R, \theta_\mathrm{B} = \pi/2 + \Delta\theta)$ for a descent lies outside the appropriate spiral section, the respective condition in (53) will be satisfied and a zigzag route will be required. Such a route is shown in figure 14 for an ascent leg with $R = 1/5, \Delta\theta = \pi/2$. On such a zigzag route with a single sharp change of direction, the turning point is where logarithmic spirals of opposite chiralities from A and B cross; so its coordinates (r_t, θ_t) are given by

$$r_t = r_\mathrm{A} \exp\left(b_\pm(\theta_t - \theta_\mathrm{A})\right) = r_\mathrm{B} \exp\left(-b_\pm(\theta_t - \theta_\mathrm{B})\right) \tag{54}$$

in which the second equality yields the solution

$$\theta_t = \frac{1}{2}\left(\theta_\mathrm{A} + \theta_\mathrm{B} - \frac{1}{b_\pm} \ln \frac{r_\mathrm{A}}{r_\mathrm{B}}\right)\,. \tag{55}$$

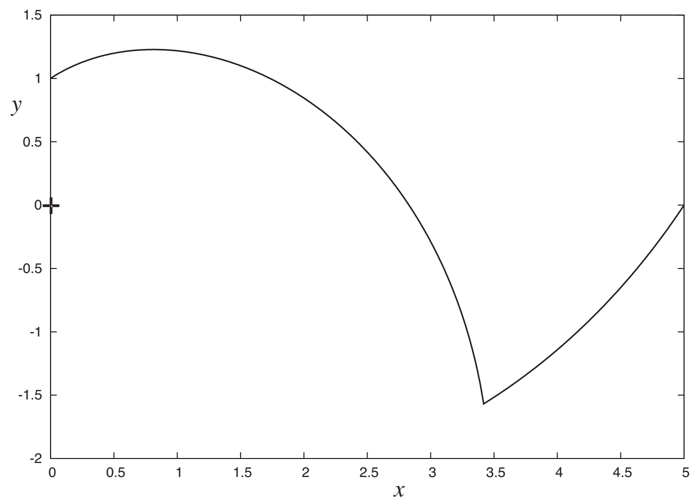

Figure 14: An optimal route zigzagging along logarithmic spiral sections from $(r_{\mathrm{A}} = 5, \theta_{\mathrm{A}} = 0)$ to $(r_{\mathrm{B}} = 1, \theta_{\mathrm{B}} = \pi/2)$ on a hill with $m_{\perp} = 0.5$, for a pace function with $m_c = 0.276$. The turning point is at $r_t \approx 3.761, \theta_t \approx -0.430$ radians, calculated using (54) and (55).

6.2.2 Tangential routes

When $e^{-b_+ \Delta\theta} < R < e^{b_- \Delta\theta}$, it is still possible that the fastest route may involve only ascent or only descent; but many routes across conical hills will involve ascent followed by descent. The latter kind of route will reach a maximum altitude at some minimum radial distance r_0, and there is no reason to suppose that the optimal route will have a sharp change of direction at this point. Thus the route will be tangential to the contours at its highest point:

$$r' = 0,\ m = 0,\ p = p_0 \quad \text{at} \quad r = r_0, \tag{56}$$

which fixes

$$C = p_0 r_0 \tag{57}$$

in (35). To compute this tangential route, we first expand about $r = r_0, \theta = \theta_0, r' = 0$ in (35) to obtain

$$r - r_0 \sim \frac{1}{2} \frac{p_0 r_0}{p_0 + p_0'' m_\perp^2} (\theta - \theta_0)^2 \tag{58}$$

and then use this to initiate a numerical integration both forwards and backwards from (r_0, θ_0); this avoids computational difficulties in trying to integrate through the tangential point. [The difficulties arise because a route along a contour, i.e. with $r' = 0, m = 0, p = p_0$ on the entire route, also satisfies (35) even though it is certainly not an optimal route: see section 4.3 above.] For any given pace function and value of $m_\perp$, we can normalise $p_0 = 1$ and $r_0 = 1$ and hence collapse all tangential optimal routes onto a single curve. This is shown for the quartic pace functions (12) and (13) on a hill with $m_\perp = 0.5$ in Figure 15, with the tangential point at $\theta_0 = \pi/2$. There is a small difference between the ascending sections of the routes for the two pace functions, as the functions do not exactly match when $m > 0$ (despite appearances in figure 4); but the major difference is in the descending sections where the "timid descender" stays closer to the contours.

Given the values of R and $\Delta\theta$ for a leg on a conical hill, we can seek a pair of points with the required R and $\Delta\theta$ on the tangential curve (noting that the direction of travel is

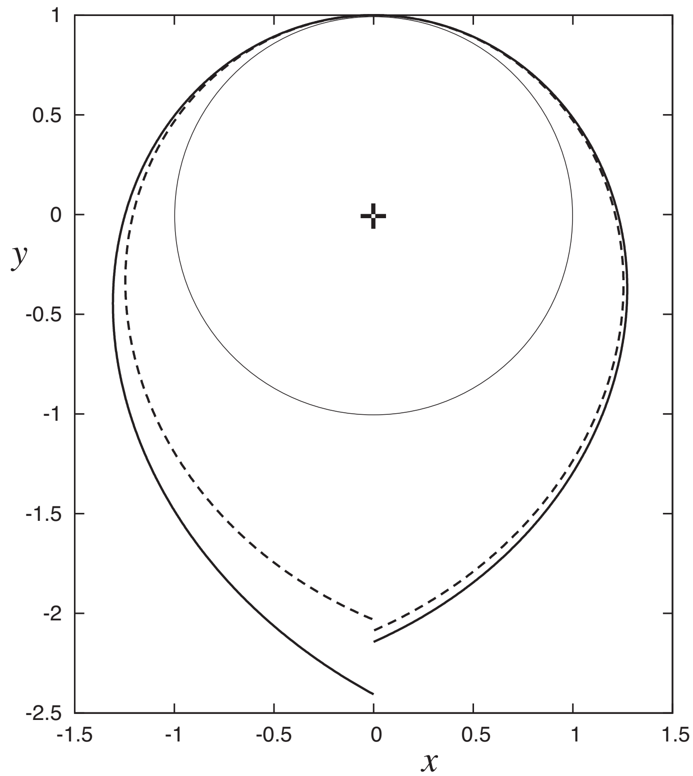

Figure 15: Optimal routes, tangential to a contour (shown by a thin circle) at ($r_0 = 1, \theta_0 = \pi/2$) on a conical hill with $m_\perp = 0.5$, for pace functions (12) (solid line) and (13) (dashed line). The direction of travel is anticlockwise. The cross marks the hill summit.

anticlockwise); if successful, the tangential curve is the fastest route for that leg. Every optimal route involving both ascent and descent will be found in this way, as well as some which involve only climbing or only descent; but in the latter case there is no *a priori* way of determining whether the optimal route between two points will be along a tangential curve. In practice, finding the pair of points on the tangential route with the required R and $\Delta\theta$ will involve trawling through the numerical data from which figure 15 is plotted. A case of particular interest is where the start and end points are at the same altitude, i.e. $R = 1$: we may ask how high up the hill does the optimal route climb between such points for any given azimuthal separation. Figure 16 shows r_0/r_{A}, the distance from the summit at the highest point of the tangential route, as a proportion of the distance at the start and end points; this is plotted against $\Delta\theta$ for the two quartic pace functions on a hill with $m_\perp = 0.5$. The amount of climbing and descent on this optimal route is $1 - r_0/r_{\mathrm{A}}$ as a proportion of the altitude difference between the start/end points and the summit. Figure 16 shows that on this rather steep hill even our "skilled descender" should never do more than 20% of the climb to the summit, and the "timid descender" will stay even closer to the contours – although in reality, routes round such a steep hill in the British uplands would often depend on the presence of sheep tracks.

6.2.3 Intermediate routes

There will be some $(R, \Delta\theta)$ combinations which cannot be found on the tangential route, but also do not satisfy the criteria (53) for the optimum route to be zigzagging along logarithmic spiral sections. In these intermediate cases, the optimum route will consist either of only ascent (with $R < 1$) or only descent (with $R > 1$). These cases are the only ones on a conical hill for which the fastest route needs to be computed by the shooting method, with a sequence of iterations homing in on the correct value of C in (35). As an example, consider points on opposite sides of a conical hill, i.e. with $\Delta\theta = \pi$, with one point four times as far from the summit as the other; so the ascent route between

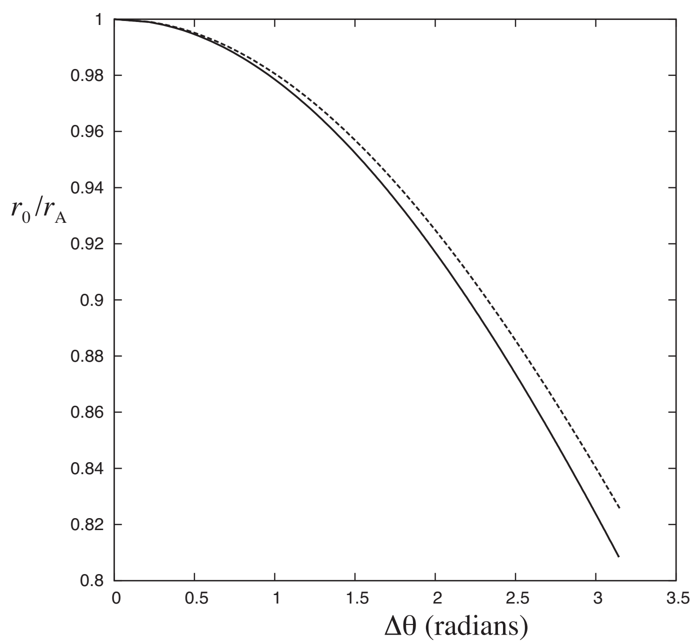

Figure 16: Distance of nearest approach to the summit, as a proportion of radial distance at start point, for tangential routes to an end point at the same radial distance as the start on a conical hill with $m_\perp = 0.5$. Pace functions are (12) (solid line) and (13) (dashed line).

the points has $R = 1/4$, while the descent route has $R = 4$. If $m_\perp = 0.5$, both the uphill and downhill routes between these points will be of the intermediate type for both of the quartic pace functions. These routes are plotted in Figure 17, in which we may note in particular the appearance of spiralling into and out of the higher point (also seen in Figure 13 for logarithmic spiral routes): this must happen if the gradient along the route is not to change sign, but the idea of approaching or leaving the higher point "round the back" would be counter-intuitive to most orienteers.

6.3 Routes on gentler hills

If the terrain gradient is less than both the uphill and downhill critical gradients, the optimal route between any start and end points will be along a tangential route, i.e. with $C = p_0 r_0$ in (35). This is apparent from Figure 18 which shows the tangential routes on a conical hill with $m_\perp = 0.2$ for both of our quartic pace functions. These routes asymptote to straight lines at large distances from the summit, in contrast to the tangential routes on the steep hill (Figure 15). When $m_\perp$ is less than the critical gradients, the quantity $p - m\,\mathrm{d}p/\mathrm{d}m$ remains strictly positive for any direction of travel; in particular, this quantity approaches a positive constant as $r' \to \infty$, so that a solution of (35) has

$$\frac{r^2}{\sqrt{r^2 + r'^2}} \to \text{ constant as } r \to \infty \text{ (and } r' \to \infty), \tag{59}$$

which is the equation of the asymptotic straight line. So a pair of start and end points with arbitrarily large or small R and arbitrarily small $\Delta\theta$ (i.e. requiring the steepest climbing or descent possible on the given hill) can be found on the nearly straight sections of the curve in Figure 18.

Considering routes between points at the same altitude, Figure 19 shows r_0/r_A plotted against $\Delta\theta$ for a hill with $m_\perp = 0.2$, which may be compared with the equivalent plot for a steeper hill in Figure 16. On the gentler hill, optimal routes between points on opposite sides of the hill will climb more than half way to the summit, even for the "timid descender".

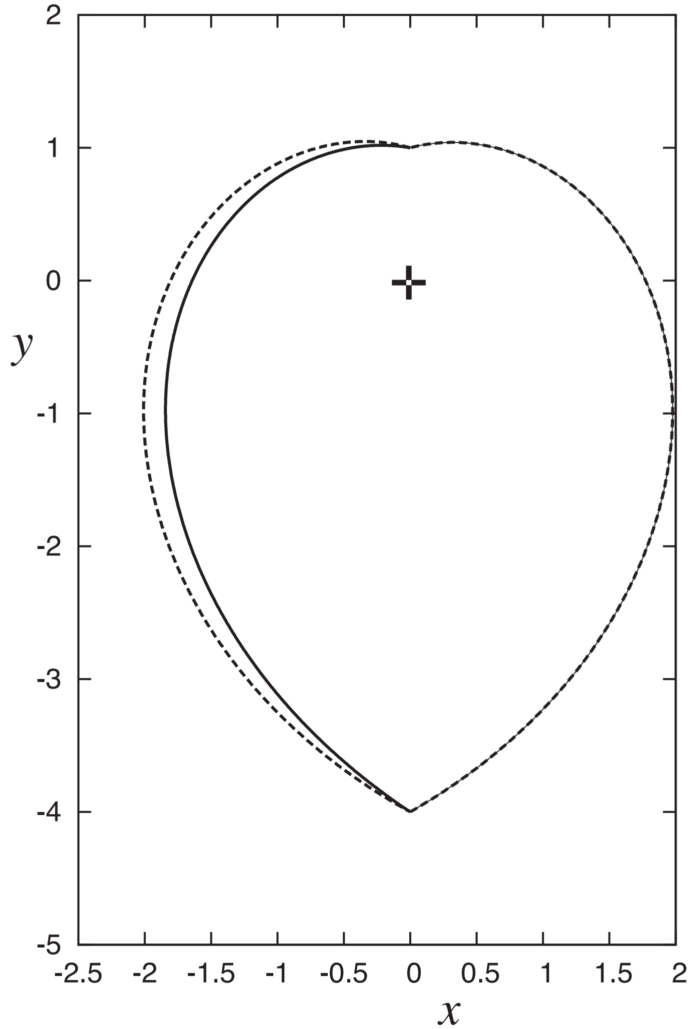

Figure 17: Optimal routes of intermediate type between points on opposite sides of a conical hill with $m_{\perp} = 0.5$, for pace functions (12) (solid line) and (13) (dashed line). The direction of travel is anticlockwise, so the ascent routes are on the right of the figure (with the routes for the two pace functions being indistinguishable), while the descent routes are on the left. The cross marks the hill summit.

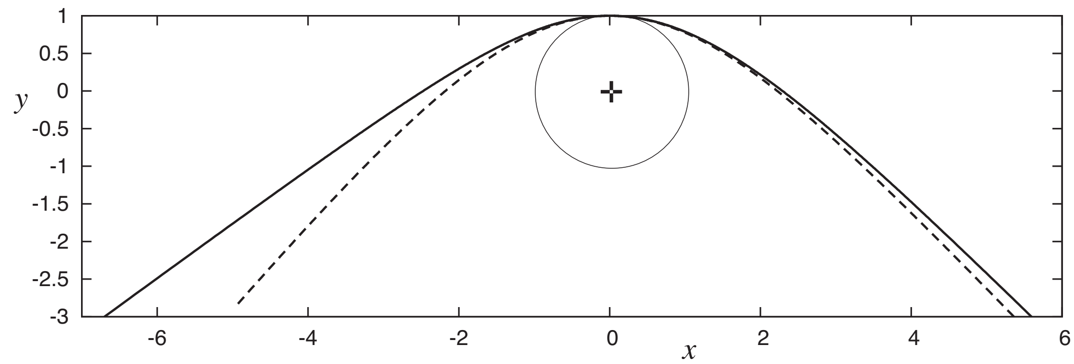

Figure 18: Optimal routes, tangential to a contour (shown by a thin circle) at ($r_0 = 1, \theta_0 = \pi/2$) on a conical hill with $m_\perp = 0.2$, for pace functions (12) (solid line) and (13) (dashed line). The direction of travel is from right to left. The cross marks the hill summit.

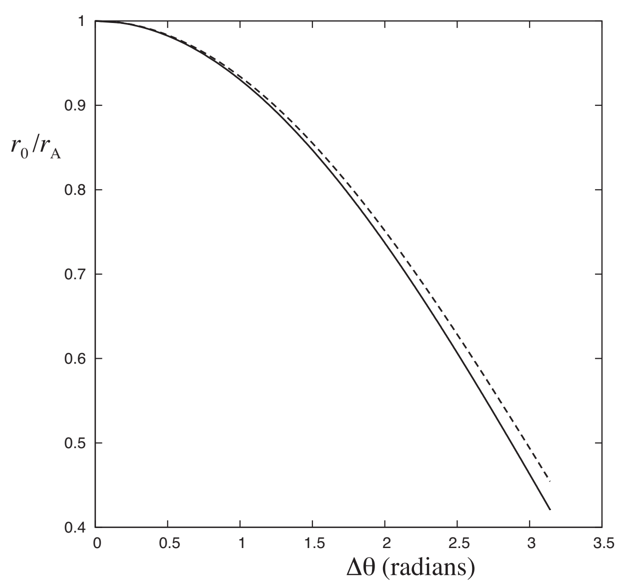

Figure 19: As figure 16, but for a hill with $m_\perp = 0.2$.

On some hills, $m_\perp$ may lie between the uphill and downhill critical gradients. For instance, with pace function (12), a hill with $m_\perp = 0.3$ is steeper than the uphill critical gradient $m_c = 0.276$ but less steep than the downhill critical gradient $m_c = 0.320$. So in this example, an optimal route involving any downhill running will certainly be along a tangential route, but some uphill legs may be along sections of logarithmic spiral or along the intermediate type of route.

Orienteers often refer to "over or round" route choices when faced with a leg between points on opposite sides of a hill. Rather than this dichotomy, we have presented route choice as a continuous optimisation problem. However, the "over or round" choice may be characterised by considering start and end points at the same altitude, at opposite sides of a hill, i.e. with $R = 1, \Delta\theta = \pi$. For the "skilled descender" pace function (12) and a range of conical hill gradients up to $m_\perp = 0.5$ we have computed the closest approach to the summit on the optimal route, expressed as r_0/r_A: results are plotted as dots in Figure 20. As $m_\perp$ increases from zero, r_0/r_A increases very slowly at first: it is the deviation from linearity of the pace function that causes optimal routes to be curved, and on a hill with very gentle gradients any route must remain within the nearly-linear regime so that the optimal route is close to the straight-line route through the summit. However, for hills steeper than $m_\perp = 0.1$, the tendency for the optimal route to go round rather than over the hill increases rapidly with increasing $m_\perp$, and with $m_\perp = 0.25$ the optimal route is already "more round than over ": the route does not go nearer than halfway to the summit.

7 Optimal routes on other axisymmetric hills

We now consider axisymmetric hills on which the terrain gradient $m_\perp$ varies with distance r from the summit. Such a hill may be *domed*, with level ground at the summit ($m_\perp = 0$ at $r = 0$) and $m_\perp$ increasing monotonically with r; or it may be *peaked*, with the gradient decreasing monotonically from a maximum value at the summit; or the

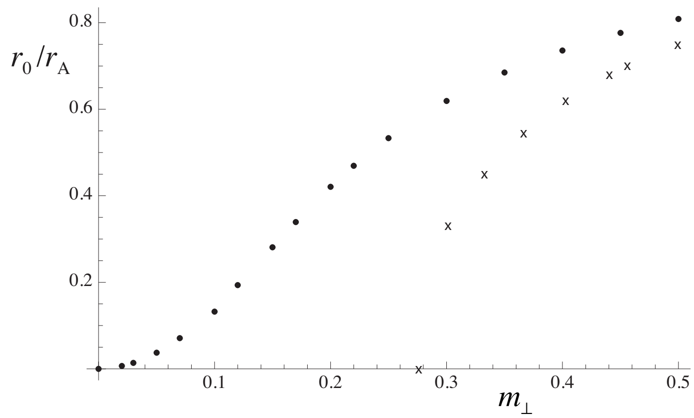

Figure 20: Distance of nearest approach to the summit, as a proportion of radial distance at start point, for optimal (tangential) routes between points at the same altitude on opposite sides of a conical hill (dots) and a parabolic domed hill (crosses), as a function of hill steepness $m_\perp$ (measured at the start and end points in the case of the domed hill), for pace function (12).

variation may be non-monotonic, e.g. a Gaussian profile $h \propto e^{-kr^2}$ might be a good approximation to the shape of some natural hills. We shall only give a detailed analysis for an example of a domed hill profile, noting that route choice optimisation on other axisymmetric hills follows the general principles derived for conical and domed hills and for Pennine slopes of variable gradient. We use the "skilled descender" pace function (12) for all the calculations below.

7.1 The parabolic domed hill

We consider a domed hill of parabolic cross-section, with terrain gradient given by

$$m_\perp = r, \tag{60}$$

where we have chosen the unit of distance similarly to that for parabolic Pennine slopes, as the distance from the hill summit to the contour where the terrain gradient is unity.

We can refer to uphill and downhill critical circles on this hill, such that only within the respective critical circle is the gradient less than the critical value for ascent or descent. Due to the numerical equality of terrain gradient and radial coordinate, these circles have respective radii m_c and m_{c-} and are centred on the summit.

A route which is tangential to the contours at its highest point $r = r_0$ will always have the constant C in (35) given by (57). We can set $p_0 = 1$, so that $C = r_0$. [Note that, unlike on the conical hill where there was no length scale, we will not be able to normalise the radial coordinate to unity at the start/end-point or point of closest approach to the summit; so tangential routes for a given pace function will not collapse onto a single curve on the parabolic domed hill.] The optimal route between a start point within the uphill critical circle and an end point within the downhill critical circle will always be along a tangential route, but the iterative method will be required to determine the value of C (or r_0) for the route between any particular pair of points. Even with the start and/or end point outside the critical circles, it will often be possible to find a tangential route between the points by the iterative method. For example, we have calculated tangential routes between points at equal distances from the summit (i.e. $r_{\mathrm{A}} = r_{\mathrm{B}}$) on opposite sides of the hill ($\Delta\theta = \pi$): the route with $r_{\mathrm{A}} = r_{\mathrm{B}} = 0.5$ is shown by a solid curve in Figure 21. Its radius of closest approach to the summit, as a proportion of the radius at the start and end points, is $r_0/r_{\mathrm{A}} = 0.751$, not much less than the value of 0.809 for a conical hill with the same gradient (0.5) as found at our present start and end points. Because a steep local gradient at a start-point will tend to induce a route at a small angle to the contours, the runner on the domed hill does not sample the gentler gradients much nearer the summit, so behaves much as he would do on a conical hill. However, with start and end points located closer to the summit where the gradient is gentler, an optimal route on a conical hill with that gradient would pass closer to the summit, so a runner on the domed hill would sample even more favourable gradients which would in turn encourage him to take an even more direct route near the summit. So on the domed hill, the distance of closest approach to the summit decreases faster with decreasing gradient (measured at the start and end points) than on the conical hill: see Figure 20. Whereas a route straight over the summit is never optimal on a conical hill with any non-zero gradient, Figure 20 shows that it becomes optimal on the domed hill when the start and end points have gradient ≤ 0.276, the lesser of the two critical gradient values for the pace function (12). Indeed the straight route would be optimal between points on opposite sides of the hill as long as the start and end points are within the respective uphill and downhill critical circles, in the same way that a straight line route normal to the contours is optimal between two points on opposite sides of a valley with parabolic cross-section if both start and end points are below the level of the appropriate critical gradients.

On a conical hill, tangential routes have a fixed non-zero value of C in (35) (once the length-scale has been fixed); they are distinct from critical-gradient spiral routes, which have $C = 0$, and there is also an intermediate class of route. In contrast, on a domed hill, tangential routes exist with all $C \geq 0$. As $C \to 0$ the route smoothly approaches the form shown in Figure 22, in which the straight section over the summit between the critical circles satisfies (35) with $C = 0$ because $r' = \pm\infty$. This joins smoothly onto the curved sections (of either chirality), which climb (on the right of Figure 22) or descend (on the left) at the respective critical gradients. These critical-gradient routes satisfy (35) with $C = 0$ because $p - m\mathrm{d}p/\mathrm{d}m = 0$, and have the equation

$$\theta - \theta_c = \pm\left(\sqrt{\frac{r^2}{m_c^2} - 1} - \cos^{-1}\left(\frac{m_c}{r}\right)\right) \tag{61}$$

if they reach the critical circle $r = m_c$ at azimuthal location $\theta = \theta_c$.

A runner on any leg starting or ending at the summit will take a route with $C = 0$ as shown in Figure 22. Going uphill from a start point where $r > m_c$, he will go along the curve (61) until reaching $r = m_c$, and then take the straight line to the summit. The same applies in reverse to a runner descending from the summit to a point outside the downhill critical circle, but there may be more practical difficulties in this case. Before starting, a runner would need to calculate an offset from the straight-line route in order

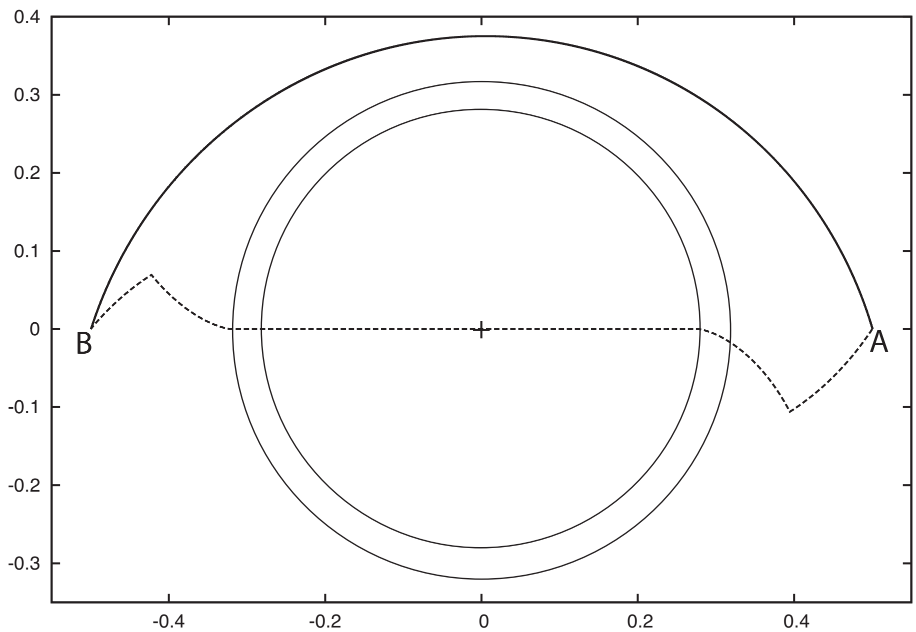

Figure 21: Locally optimal routes between points on opposite sides of a parabolic domed hill, for a runner with pace function (12). Start point A and end point B are situated where the terrain gradient is 0.5. The solid line is a tangential route (which is globally optimal in this case). The dashed line is a route over the summit, requiring zigzagging from A to the uphill critical circle (radius 0.276) and from the downhill critical circle (radius 0.320) to B.

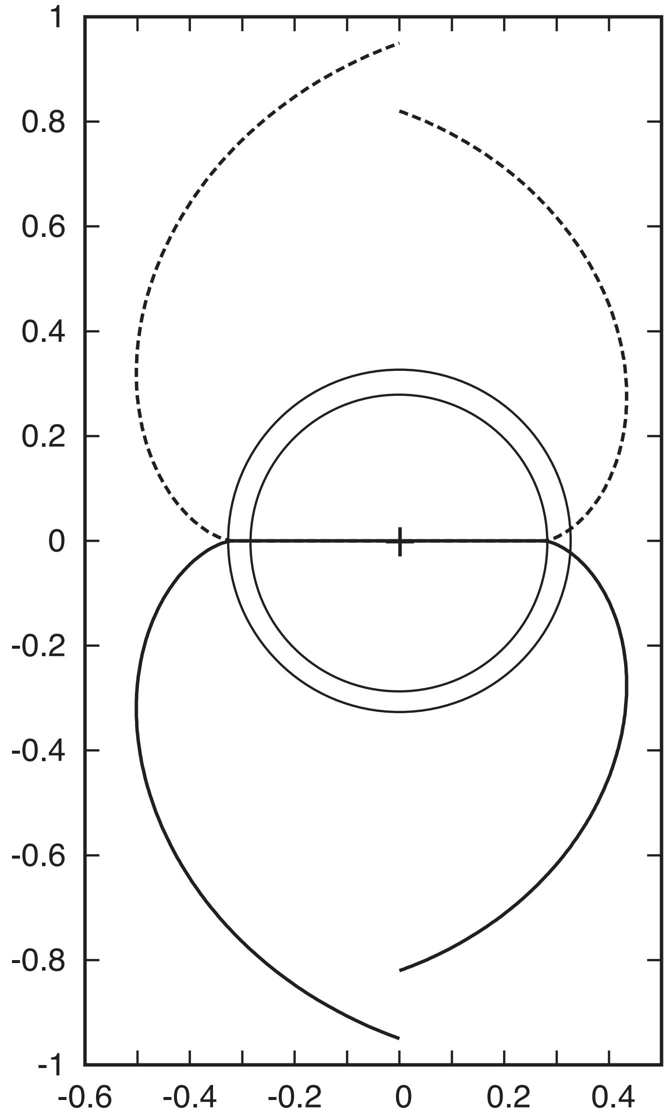

Figure 22: Routes with $C = 0$ on a parabolic domed hill, for a runner with pace function (12). The routes climb at the uphill critical gradient (on the right) to the uphill critical circle (radius 0.276), then pass straight over the summit to the downhill critical circle (radius 0.320), and finally descend at the downhill critical gradient. Solid and dashed lines indicate routes of opposite chiralities.

to arrive at the critical circle at a point where the critical-gradient descent curve takes him to the given end-point. Alternatively, he could head straight for the end-point and, after reaching $r = m_{c-}$, zigzag along at least two sections of curves of the form (61) with opposite chiralities (indicated by the $\pm$ sign) and different values of θ_c. This difference between uphill and downhill behaviour is similar to that noted by LS.

A zigzag route will always be required if the start and end points are on the same side of the hill ($\Delta\theta$ fairly small, but not necessarily zero), with either or both outside the relevant critical circle, and where it is not possible to connect the points by a smooth route with $C = 0$ or by a less steep tangential route. For instance, consider the case with $\Delta\theta = 0$, $r_\mathrm{A} > m_c$ and $r_\mathrm{B} < m_c$. We may take $\theta_\mathrm{A} = \theta_\mathrm{B} = 0$, and the optimal route certainly climbs along the straight, radial route $\theta = 0$ above the uphill critical circle, where $m_c > r > r_\mathrm{B}$. Below the critical circle, we seek an optimal zigzag route consisting of two sections, with a turning point (r_t, θ_t) to be found. The upper section of the zigzag satisfies (61) with $\theta_c = 0$ and a minus sign on the right-hand side; the lower section satisfies (61) with $\theta_c = 2\theta_t$ (by chiral symmetry) and a plus sign. Since this lower section also passes through the point $(r = r_\mathrm{A}, \theta = 0)$, we have

$$-2\theta_t = \sqrt{\frac{r_\mathrm{A}^2}{m_c^2} - 1} - \cos^{-1}\left(\frac{m_c}{r_\mathrm{A}}\right) \tag{62}$$

and r_t can then be found by numerical solution of (61) at $\theta = \theta_t$. A zigzag route calculated in this way for $r_\mathrm{A} = 0.5$ can be seen on the right of Figure 21, with a similar downhill zigzag on the left. However, this figure shows these zigzags joined by a straight route over the summit, showing that a locally optimal, zigzagging route with $C = 0$ exists between points on opposite sides of a domed hill, in addition to the tangential route also shown on this figure. To resolve this "over or round" dichotomy and determine which of these routes is globally optimal, it is necessary to compute the time taken along each of the routes. Numerical integration of (1) is required for the tangential route. For the "over" route, the time taken along the straight section with gradient $m = \pm m_\perp = \pm r$ is

$$\int_0^{m_c} p(r)\,\mathrm{d}r + \int_0^{m_{c-}} p(-r)\,\mathrm{d}r\,; \tag{63}$$

to this we need to add times taken along uphill and/or downhill critical-gradient curves. These times are simply the arc lengths multiplied by the value of pace at the critical gradient, where the arc length between the critical radius $r = m_c$ and some greater radius $r = r_f$ along a critical gradient curve (61) (or zigzagging along any number of sections of such curves) is found to be

$$s_{cf} = \frac{r_f^2 - m_c^2}{2m_c} \,. \tag{64}$$

We have computed runners' times for cases with start and end points at the same radial distance on opposite sides of the hill ($r_\mathrm{A} = r_\mathrm{B}$ and $\Delta\theta = \pi$) and with the pace function (12). The "over" and "round" routes are equally fast when $r_\mathrm{A} \approx 0.445$, with the "over" route being faster when r_A is smaller than this value and the "round" route being faster when r_A is larger. A less skilled descender, e.g. with pace function (13), would find the "round" route faster with start and end point positions at smaller radii than 0.445.

7.2 Other axisymmetric hills

On a domed hill, optimal routes will tend to go closer to the summit than on a conical hill, given start and end points located where the hill has the same gradient. A peaked hill will produce the opposite tendency, with optimal routes keeping further away from the summit. Consider a peaked hill with gradient steeper than both the uphill and downhill critical values at the summit. There will be uphill and downhill critical circles, outside which the gradient becomes less steep than the respective critical values. If the start or end point is within the appropriate critical circle, spiralling and/or zigzagging may be needed. As on a Pennine slope when climbing from the valley bottom to a point where the gradient is steeper than critical (see Figure 9), it may be necessary to use a route which climbs or descends perpendicular to the contours outside the critical circle, and then zigzags at the critical gradient within this circle. On the other hand, for many choices of start and end point, a tangential route will be optimal. As on the conical hill, a straight-line route over the summit will never be optimal, even when the gradient at the summit is less than the critical values.

A hill with a Gaussian or similar profile appears domed from the summit out to some radius of maximum gradient, but the gradient decreases beyond this radius. Route choice problems are likely to be particularly interesting if the maximum gradient is greater than critical, so that there are annuli of super-critical gradients for climbing and descending. Nevertheless, the route choice possibilities are likely to be similar to those for a domed hill, although the details will be more intricate. For instance, a route between points at opposite sides of the hill and outside the super-critical annuli might start by climbing towards the summit, then zigzag upwards through the uphill super-critical annulus, then straight to the summit; it would then continue downhill reversing the pattern of uphill running.

8 Discussion

We have considered the problem of finding the fastest route between given points in hilly terrain, where the speed of travel depends only on the gradient along the route. Even with this simplifying assumption and the restriction to a few idealised landforms, there is a rich variety of solutions.

The choice of route depends crucially on a runner's "pace function" (variation of pace p with route gradient m) and the associated critical gradients. The data on such pace functions are sparse and sometimes contradictory. In any case, individual runners will behave differently in this respect, especially when running downhill where the pace is dictated by factors other than the runner's metabolism. Hence we have proposed two quartic pace functions which we believe to be plausible. We chose to fit these functions to the uphill data of Minetti et al (2002), not only because these data appear to be among the best available, but also because a pace function with a large curvature ($\mathrm{d}^2p/\mathrm{d}m^2$) makes it easier to illustrate many of the interesting features of our analysis. Naismith's rule, as used in the from (4) by Scarf (2008) and Arnet (2009), would appear to be unsatisfactory for fastest-route studies in steep terrain, since it does not give rise to a critical gradient. Most of the evidence that we have found suggests that critical gradients for both ascent and descent should not be too far from 0.3.

Our analysis is based on the Euler-Lagrange equation in the calculus of variations, which in general yields somewhat intractable second-order boundary-value problems. Our idealised topographies have reduced the equations to first order, with an undetermined constant of integration C. This constant is zero when the optimal route involves sections of ascent or descent at the critical gradient or normal to the contours (where the terrain gradient is sub-critical), and analytical solutions are then available. In other cases, a straightforward numerical integration is required, with the value of C either determined by the insight that a route which climbs and then descends (on an axisymmetric hill) must be tangential to the contours at its highest point, or found by an iterative procedure.

An important principle is that where a leg involves only climbing or only descent, it will curve so as to reduce the variation of gradient along the route. This is a result of pace functions having $\mathrm{d}^2p/\mathrm{d}m^2 > 0$, which means that traversing two route sections at different gradients will always be slower than running the same total distance at the average gradient. Larger values of $\mathrm{d}^2p/\mathrm{d}m^2$ tend to be associated with smaller values of critical gradients and a greater tendency to avoid steep climbs or descents.

Orienteers tend to think in terms of a dichotomous “over or round” route choice between points on opposite sides of a hill. In contrast, by modelling route choice as a continuous optimisation problem, we generally propose a route intermediate between contouring and going straight over the summit. Nevertheless, in the case of domed hills, we have found that a dichotomy does sometimes appear: a route over the summit and a route going round the hill with a small element of climbing and descent are both locally optimal, and the dichotomy can only be resolved by direct calculation of times along the two routes.

The analysis could be modified to include isotropic elements (runnability and terrain gradient) in the pace function, although this would invalidate the simple results relating to critical gradient. It could also be extended to slightly more complicated geometries: an elliptical hill may be tractable, as would a river valley with a non-zero gradient along the valley bottom. For a large area of complex terrain, a computational method based on a discretisation of the terrain would certainly be required to optimise routes, but the present study could still provide useful input to such computations. For the dynamic programming approach used by Hayes & Norman (1984) on a relatively coarse grid, our results could be used to provide improved estimates of travel times between adjacent grid points, since the terrain within each grid square is relatively simple in form. This would not be relevant to the computations of Arnet (2009) in which the grid size is comparable to a runner's step length, but our results do highlight the importance of using a nonlinear pace function in such studies.

Possibly the greatest limitation to the utility of our results is the uncertainty over runners' pace functions. Norman (2004) has commented on the difficulties of obtaining data that truly represent the pace of runners on hills. In any case, pace functions vary considerably between individual runners, especially downhill (J.A. Easterbrook, personal communication). On the other hand, pace functions (or other cost functions, such as fuel consumption) will be easier to measure and more consistent for road and railway vehicles, so our analysis may be more useful when planning road or railway routes – although there are likely to be other environmental constraints on these in hilly terrain.

Acknowledgements

This work was first attempted as an undergraduate project in 1978, while I was at St. Andrews University, studying mathematics but competing in orienteering events in my spare time. It was laid aside, as the equations seemed intractable at the time; so I am grateful to Philip Scarf, whose recent articles have re-ignited my interest in the subject.

References

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.

Balstrøm T (2002) On identifying the most time-saving walking route in a trackless mountainous terrain. *Danish Journal of Geography* **102**, 51 – 58.

Davey RC, Hayes M & Norman JM (1994) Running uphill: an experimental result and its applications. *J. Opl. Res. Soc.* **45**, 25 – 29.

Hayes M & Norman JM (1984) Dynamic programming in orienteering: route choice and the siting of controls. *J. Opl. Res. Soc.* **35**, 791 – 796.

Kay A (2006) Orienteering route choice: running at the speed of light? *Mathematics Today* **42**, 138 – 143.

Langmuir E (1984) *Mountaincraft and Leadership.* The Scottish Sports Council, Edinburgh.

Llobera M & Sluckin TJ (2007) Zigzagging: Theoretical insights on climbing strategies. *J. Theor. Biol.* **249**, 206 – 217.

Miller HJ & Bridwell SA (2009) A field-based theory for time geography. *Annals of the Association of American Geographers* **99**, 49 – 75.

Minetti AE, Moia C, Roi GS, Susta D & Ferretti G (2002) Energy cost of walking and running at extreme uphill and downhill slopes. *J. Appl. Physiol.* **93**, 1039 – 1046.

Norman JM (2004) Running uphill: energy needs and Naismith's rule. *J. Opl. Res. Soc.* **55**, 308 – 311.

Puu T & Beckmann M (1999) Continuous space modelling. In *Handbook of Transportation Science*, ed. RW Hall, 269 – 310. Kluwer, Boston.

Rees WG (2004) Least-cost paths in mountainous terrain. *Computers & Geosciences* **30**, 203 – 209.

Scarf P (1998) An empirical basis for Naismith's rule. *Mathematics Today* **34**, 149 – 151.

Scarf P (2007) Route choice in mountain navigation, Naismith's rule, and the equivalence of distance and climb. *Journal of Sports Sciences* **25**, 719 – 726.

Scarf P (2008) A mathematical excursion in the isochronic hlls. *Mathematics Today* **44**, 163 – 167.

Warntz W (1957) Transportation, social physics, and the law of refraction. *The Professional Geographer*, **9**, 2 – 7.