A variant to Naismith’s problem with
application to path planning *

Erik I. Verriest *

School of ECE, Georgia Institute of Technology, Atlanta, GA
30306-0250 USA (Tel: 404-894-2949; e-mail:
erik.verriest@ece.gatech.edu).

Abstract: Naismith obtained a set of empirical rules for the time required to move through a terrain. In this paper we solve the problem of minimization the transit time between two points on a given terrain. We give an interpretation of Naismith’s rule which leads to a very elegant geometric construction of the optimal solution. Indeed, there is some ambiguity in the interpretation of Naismith’s rule. We first solve the problem for a conical mountain, then generalize for a terrain with arbitrary topography. We conclude with a discussion of the relative merit of our variant with respect to the known solution. In particular, we show that the difference of these interpretations amounts to less than 10% in the worst case, thus justifying the use of this simple solution. This problem is a paradigm for navigation of autonomous vehicles in heterogenous terrain. It may prove usefull for path planning for robotic rovers as used for instance on the surface of Mars.

Keywords: Optimal control, path planning, autonomous vehicles

1. INTRODUCTION

Naismith’s rule allows walkers to compute the time for their journeys. The time is given by allowing a walking speed of 4 km/hr, but adding an extra minute for each 10 m of ascent.

In 1996, N. MacKinnon posed the following problem in the American Mathematical Monthly [5].

A conical mountain has base radius 1650 m and vertical height 520 m. Points A and B are diametrically opposite at the base of the mountain. How should a path be constructed between A and B on the surface of the mountain which minimizes the time taken to walk from A to B?

A solution appeared in [6]. In this paper we propose an alternate solution, to a different interpretation of the same problem, and show that this solution has a simpler implementation of the optimal control. This problem has adaptations to different terrain models. Whereas here we shall consider only a uniform terrain but consider different effects for up and down sloped terrain, similar effects depending on the type of terrain (e.g., sand, mud or pebbles) and or different types of vegetation are solved in a similar way. One application of interest is in path planning for Mars rover surface operations [3].

We point out that Naismith’s rule is a rule of thumb that helps in the planning of a walking or hiking expedition by calculating how long it will take to walk the route, including ascents. The rule was devised by William Naismith, a Scottish mountaineer, in 1882 [10]. The fact that great books such as [8] do not mention the rule, adds to its mystique and misinterpretations.

2. SETTING THE IDEAS: TRAVERSING A CONE SHAPED MOUNTAIN

2.1 Parameter reduction

This is a typical optimal control problem, solvable with the Pontryagin minimum principle [1]. Allowing for differences of conditioning between individuals, I shall slightly generalize the problem: Let $v$ be the horizontal speed and $\mu$ the additional time required for ascent per unit vertical distance. Thus for an ascent of $dz$ over an elementary distance (projected onto the horizontal plane), $ds$, a time $dt = ds/v + \mu dz$ is required, while the corresponding descent will only take $dt = ds/v$ time units. The problem is one of minimizing elapsed time in hiking along a path $\Gamma$,

$$T_{\Gamma} = \int_{\text{ascent}} \left( \frac{ds}{v} \right) + \int_{\text{descent}} \left( \frac{ds}{v} + \mu dz \right).$$ (1)

Introducing the indicator function $I(\cdot)$ with $I(u) = 1$ if $u > 0$ and $I(u) = 0$ for $u \leq 0$, the elapsed time can be written as

$$T_{\Gamma} = \int_{\Gamma} \left( \frac{ds}{v} + \mu I(dz)dz \right).$$ (2)

2.2 Geometry and parameter reduction

We now specialize the generalized set-up for the geometry at hand: Let the conical mountain have base radius $R$ and height $h$. For this conical mountain, it is nice to use cylindrical coordinates $(z, \theta, \rho)$, with the top at $\rho = 0$. 

* This work was supported in part by a RIM Grant (Georgia Institute of Technology).
In these coordinates the conical surface is given by the equation (Figure 1):

\[ z(\rho, \theta) = \left(1 - \frac{\rho}{R}\right) h \]  

(3)

for \( \rho < R \) and 0 else. For \( z = z(\rho, \theta) \), the elementary distance (projected onto the horizontal plane) along the path is

\[ ds^2 = d\rho^2 + \rho^2 d\theta^2 = \left[ \left(\frac{d\rho}{d\theta}\right)^2 + \rho^2 \right] d\theta^2. \]  

(4)

Let

\[ \frac{d\rho}{d\theta} = \tan \psi \]  

(5)

where \( \psi \) is the heading with respect to the circular path, centered at O, i.e., the complement of the angle between the trajectory and the radius vector (Figure 2).

Denote for simplicity

\[ \rho = Rr \]  

(6)

the ‘control’.

From (3), we get \( z_\rho = 0 \) and \( z_\theta = -\frac{h}{R} \). Hence

\[ dz = -\frac{h}{R} d\rho = -\frac{h}{R} \frac{d\rho}{d\theta} d\theta = -hu d\theta. \]  

(7)

Note that \( u < 0 \) and \( u > 0 \) correspond respectively with ascent and descent. Consequently, using (4),(6) and (7) in (2), the elapsed time along a path \( \Gamma \) is equal to

\[ T_\Gamma = \frac{1}{v} \int_\Gamma \left( \sqrt{R^2u^2 + \rho^2} - \mu vh (-u) \right) d\theta \]  

(8)

Introducing the single dimensionless parameter

\[ \mu_0 = \frac{vh}{R}, \]  

(9)

and normalizing by letting \( \rho = Rr \), the problem reduces to the minimization of the dimensionless quantity

\[ \tau_\Gamma = \int_\Gamma \left( \sqrt{u^2 + r^2} - \mu_0 I(-u) \right) d\theta \]  

(10)

along a path \( \Gamma \) with dynamical constraint

\[ \frac{dr}{d\theta} = u \]  

(11)

and specified initial and final coordinates.

Note that he minimum time is retrieved from \( \tau \) by

\[ T_\Gamma = \frac{R}{v} \]  

(12)

2.3 Characterization of the optimal path

The Hamiltonian for the above problem (10)-(11) is obtained by adjoining the constraint (11) with the Lagrange multiplier (or co-state) function \( \lambda(\theta) \) to the integrand in (10)

\[ H = \sqrt{u^2 + r^2} + [\lambda - \mu_0 I(-u)]u. \]  

(13)

The Pontryagin minimum principle postulates that the optimal solution is obtained by locally (i.e. for each fixed \( r \) and \( \lambda \)) minimizing this Hamiltonian with respect to the control \( u \). As function of \( u \), the Hamiltonian has the asymptotes \( H_+(u) = (\lambda + 1)u \) for \( u \to \infty \) and \( H_-(u) = (\lambda - \mu_0 - 1)u \) for \( u \to -\infty \). Clearly, a minimum of \( H \) exists if and only if

\[ \begin{cases} \lambda - \mu_0 - 1 < 0 \\ \lambda + 1 > 0 \end{cases}, \]  

(14)

or,

\[ -1 < \lambda < \mu + 1 \]  

(15)

Note also that the derivative of the Hamiltonian with respect to \( u \) is

\[ H_u = \frac{u}{\sqrt{u^2 + r^2}} + \lambda \] for \( u > 0 \),

(16)

\[ H_u = \frac{u}{\sqrt{u^2 + r^2}} - \lambda - \mu_0 \] for \( u < 0 \).

(17)

(18)

Consider now the following three cases:

Case 1: \( \lambda \leq 0 \). Then also \( \lambda - \mu_0 < 0 \). For \( u < 0 \), \( H_u < 0 \), while \( H_u = 0 \) for

\[ \frac{u}{\sqrt{u^2 + r^2}} = -\lambda > 0. \]  

(19)

Hence a unique minimum of \( H \) exists in \( u \geq 0 \) if also \( \lambda \geq -1 \) and is attained for

\[ u^* = -\frac{\lambda r}{\sqrt{1 - \lambda^2}} \geq 0, \]  

(20)

(the superscript * denotes optimality) at which

\[ H^* = r \sqrt{1 - \lambda^2}. \]  

(21)

Case 2: \( \lambda > \mu_0 > 0 \). Since now \( \lambda - \mu_0 > 0 \), for \( u > 0 \), it follows that \( H_u > 0 \), and a unique minimum of \( H \) exists in \( u \leq 0 \) if also \( \lambda - \mu_0 < 1 \). It is determined by

\[ \frac{u}{\sqrt{u^2 + r^2}} = -\lambda + \mu_0 < 0. \]  

(22)

This gives

\[ u^* = \frac{(\lambda - \mu_0)r}{\sqrt{1 - (\lambda - \mu_0)^2}} \leq 0, \]  

(23)

at which

\[ H^* = r \sqrt{1 - (\lambda - \mu_0)^2}. \]  

(24)
Case 3: $0 < \lambda < \mu_0 > 0$. For $u > 0$, we get $\mathcal{H}_u > 0$, and for $u < 0$, $\mathcal{H}_u < 0$. Hence the unique minimizer of $\mathcal{H}$ is $u^* = 0$. With this:

$$\mathcal{H}' = r. \quad (25)$$

Dynamically speaking, what happens in these three cases? The co-state satisfies (using the dot-notation for the derivative with respect to $\theta$):

$$\dot{\lambda} = -\mathcal{H}_r = -\frac{r}{\sqrt{u^2 + r^2}} = -\frac{1}{\sqrt{1 + \frac{u^2}{r^2}}}. \quad (26)$$

In case 1, (i.e. $\lambda(\theta_0) \leq 0$) this yields

$$\dot{\lambda} = -\sqrt{1 - \lambda^2}. \quad (27)$$

using (20). The above differential equation in $\lambda$ can be readily integrated

$$d\theta = -d\arcsin(\lambda), \quad (28)$$

to yield

$$\lambda(\theta) = \lambda(\theta_0) \cos(\theta - \theta_0) - \sqrt{1 - \lambda(\theta_0)^2} \sin(\theta - \theta_0). \quad (29)$$

Hence $\lambda(\theta)$ is periodic with period $2\pi$. Eventually, $\lambda(\theta)$ crosses the zero level. This occurs for $\theta_1 = \theta_0 + \arctan\frac{\lambda(\theta_0)}{\sqrt{1 - \lambda(\theta_0)^2}} = \theta_0 + \arcsin(\lambda(\theta_0)) < \theta_0 + \frac{\pi}{2}$. At that time the above dynamic equation is no longer valid.

Likewise, if $\lambda(\theta_0) > \mu_0$, then (23) in (26) yields

$$d\theta = -d\arcsin(\lambda(\theta) - \mu_0) \quad (30)$$

and integrating:

$$\lambda(\theta) = \mu_0 + (\lambda(\theta_0) - \mu_0) \cos(\theta - \theta_0) + \sqrt{1 - (\lambda(\theta_0) - \mu_0)^2} \sin(\theta - \theta_0). \quad (31)$$

Thus also in this case is $\lambda(\theta)$ periodic with period $2\pi$. Eventually, $\lambda(\theta)$ crosses the level $\mu_0$, (for some $\theta_1' = \theta_0 + \arctan\frac{\lambda(\theta_0) - \mu_0}{\sqrt{1 - (\lambda(\theta_0) - \mu_0)^2}} = \theta_0 + \arcsin(\lambda(\theta_0) - \mu_0) < \theta_0 + \frac{\pi}{2}$).

Finally the third case $0 < \lambda(\theta_0) < \mu_0$, which yields $u^* = 0$, leads to

$$\dot{\lambda} = -1, \quad (32)$$

so that eventually $\lambda(\theta)$ crosses the zero level (at $\theta_1'' = \theta_0 + \lambda(\theta_0)$).

We conclude thus that if $\lambda(\theta_0) > \mu_0$, $u^*$ remains negative (i.e., the path is ascending) until some $\theta_1 = \theta_0 + \arcsin(\lambda(\theta_0) - \mu_0) < \theta_0 + \frac{\pi}{2}$. At this point $\lambda(\theta_1) = \mu_0$, so that $u^* = 0$, i.e., the path follows a level circle, a contouring arc, while $\lambda$ must decrease further, eventually reaching 0 (when $\theta = \theta_2$). At that point $\lambda(\theta_2) = 0$ and $u^* > 0$, making the path descending. The path remains descending for at most an angular increase of $\theta_3 - \theta_2 = \arcsin(\lambda(\theta_2)) = \pi$. Hence, at most one ascent and descent can occur along the optimal path.

Furthermore, the ascending and descending paths are such that their projection onto the horizontal plane is a straight line. This is easily seen as follows: Since he Hamiltonian is not an explicit function of $\theta$, $\mathcal{H}$ must be a constant of the motion. In terms of this constant $\mathcal{H}$ we get from (20) and (21) upon elimination of $\lambda$

$$u = r\sqrt{\frac{r^2}{\mathcal{H}^2} - 1}. \quad (33)$$

Similarly, (23) and (24) yield

$$u = -r\sqrt{\frac{r^2}{\mathcal{H}^2} - 1}. \quad (34)$$

With the state equation (11), this leads to the equation for the descending and ascending parts in polar coordinates

$$\dot{r} = \pm r\sqrt{\frac{r^2}{\mathcal{H}^2} - 1} \quad (35)$$

with $+$ for the descending and $-$ for ascending parts respectively. Integrating, yields

$$\theta - \theta_0 = \pm \left[ \arcsin\frac{\mathcal{H}}{r} - \arcsin\frac{\mathcal{H}}{r_0} \right], \quad (36)$$

from which after some inversions we get respectively

$$r = \frac{r_0}{\cos(\theta - \theta_0) \pm \sqrt{\frac{r^2}{\mathcal{H}^2} - 1} \sin(\theta - \theta_0)}. \quad (37)$$

These are the equations of two straight lines through the point $(r_0, \theta_0)$ in polar coordinates.

At this point we reach the conclusion that the optimal path consists of an ascent with fixed bearing (the projection on the horizontal plane is a straight line), a contour at fixed level, or equivalently fixed radius, and a descent, again with fixed bearing.

### 2.4 Optimization as a parameter optimization problem

Based on the conclusion reached at the end of the previous section, consider a potentially optimal path APQB in Figure 3. Let the contour part of the path have radius $r$. The performance index (10), evaluated along this path is $\mathcal{X}P + \mathcal{P}Q + \mathcal{Q}B + \mu_0(1 - r)$.}

![Fig. 3. Potential optimal path](image_url)

Since $r$ is fixed, it is immediately clear that an optimal path will be symmetrical about the line at $\theta = \frac{\pi}{2}$. Hence it suffices to optimize the path APR. We get from the geometry,

$$\mathcal{X}P = \sqrt{r^2 - 2r\cos(\alpha)} + 1 \quad (38)$$

$$\mathcal{P}R = \left(\frac{\pi}{2} - \alpha\right)r \quad (39)$$
Thus minimizing $AP + PR$ or
\[ \min_{\alpha} \left[ \sqrt{r^2 - 2r \cos(\alpha) + 1} + \left( \frac{\pi}{2} - \alpha \right) r \right] \] (40) gives the solution
\[ \cos \alpha = r. \] (41)
This implies that $AP = \sqrt{1 - r^2}$ and that the line $AP$ is tangent to the circle with radius $r$, or equivalently $AOP$ is a right triangle, and $\alpha = \psi_0$, the initial heading with respect to a circular path (Figure 4).

Fig. 4. Circular path

At this point we narrowed down all potentially optimal paths to a one-parameter set, parameterized by $r$. A final optimization over $r$ is required to yield the minimum time path. With $r = \cos \alpha$, the performance index (10) is
\[
\tau = 2(\sqrt{AP} + PR) + \mu_0 (1 - r)
= 2\sqrt{1 - r^2} + [\pi - 2 \arccos r]r + \mu_0 (1 - r). \] (42)
This is minimal for
\[ r^* = \sin \frac{\mu_0}{2}. \] (43)
The value of this local minimum is
\[ \tau^*(\mu_0) = \mu_0 + 2 \cos \frac{\mu_0}{2}. \] (44)
Comparing to the circum navigation time (contouring the mountain at level $0$), $\tau_{CN} = \pi$, we see that (44) fails to give the global optimum solution if $\mu_0 > \pi$. Thus the global solution is:
\[
r^*(\mu_0) = \begin{cases} 
\sin \frac{\mu_0}{2} & \text{if } \mu_0 < \pi \\
1 & \text{else}
\end{cases}
\] (45)
\[
\tau^*(\mu_0) = \begin{cases} 
\mu_0 + \cos \frac{\mu_0}{2} & \text{if } \mu_0 < \pi \\
\pi & \text{else}
\end{cases}
\] (46)
\[
\psi^*(\mu_0) = \begin{cases} 
\frac{\pi}{2} & \text{if } \mu_0 < \pi \\
\frac{\mu_0}{2} & \text{if } \mu_0 < \pi \\
0 & \text{else}
\end{cases}
\] (47)
(48)
These functions are plotted in Figures 5, 6 and 7.

The solution for the specified numerical values is:
\[
\mu_0 = \frac{\mu h}{R} = \frac{1}{10} \cdot \frac{4000}{1650} \cdot 520 = \frac{208}{99} = 2.10101010\ldots
\]
\[\tau^*(\mu_0) = 3.095,\]
and thus by (12)
\[T^* = \tau^* \frac{R}{v} = 76.608 \text{ min}\]
The angle of departure with respect to the level line is
\[\psi^*_0 = \frac{\pi}{2} - \frac{\mu_0}{2} = 0.52029 \text{ rad} = 29.81^\circ.\]
The mountain needs to be ascended to a height of
\[z = \left[ 1 - \sin\left( \frac{\mu_0}{2} \right) \right] h = 68.809 \text{ m}.\]

Remark: Obviously there are two symmetrical solutions, one passing the mountain top on the right, one passing on the left. While mathematically equivalent, in case a Tibetan Buddhist mani stone is situated near the summit, according to Buddhist protocol, the mani wall should be passed on the left.

3. GENERAL SOLUTION

In my previous solution, based on optimal control theory, it was found that the optimal trajectory consisted of a straight line, ascending to a height $h$, then contouring at that level, followed by a descent, again in straight line (i.e., as projected onto the horizontal). The optimum height $h$ depends on the relative times required to cover a horizontal distance and to ascend. (See my earlier solution). Here I give a simple geometric argument to show optimality in the case of a (locally) convex mountain (or a concave pit). The given problem is a special case of this.

Consider thus a mountain and two points $A$ and $B$ such that in a convex domain $D$, containing these points the level sets, restricted to $D$, are convex. (Figure 5). Assume
time for ascending from its lowest level to B. Also any path to the right of the straight line path AB, requires time for an initial ascent, and covers also more horizontal distance.

Now let us assume that the optimal path ascends to level \( h_1 \). Let \( P \) be point where the level is visited the first time, and \( Q \) the point where it is visited for the last time. The shortest time path from A to P is along the straight line AB. The required time is \( T_M = \frac{100}{\cos \alpha} + 100 \tan \alpha \). Likewise, the shortest path from Q to B is along the straight line QB, requiring a time \( \frac{2000}{\cos \alpha} + 100 \tan \alpha \). Next, the optimal path P to Q is along the level line at \( h_1 \), since a path going to the right would contradict that \( h_1 \) is the highest level (by convexity), while a path to the left would again cover a larger horizontal distance and require a final ascent back to Q, thus requiring more time also.

Finally, it is easily seen that P and Q must respectively be the points of tangency of the tangents to the level line at \( h_1 \), respectively passing through A and B. Indeed, for any \( P_1 \) closer to A than P, the horizontal distance traversed to reach P is \( AP_1 + \int_{P_1}^{P} ds \), which exceeds the distance for the direct path: \( AP + \int_{P}^{P} ds \).

The optimal level \( h_1 \) is determined by a parameter optimization, and depends on the shape of the level line and the parameters \( v \) and \( \mu \).

If A and B are at different levels, with \( h_A < h_B \), then a similar reasoning shows that the optimal path must be contained in the domain bounded by AB, the level line at \( h_B \), and AR, where AR is tangent to the level line at \( h_B \) (Figure 6). An optimal maximal height \( h_1 > h_B \) can be determined, for which the optimal path would be APQB, with AP and QB tangent to the level line at height \( h_1 \).

If \( h_A > h_B \), then the optimal path must be inside the domain ASB where BS is tangent to the level line at \( h_A \). Again there is an optimal \( h_1 > h_A \) and the corresponding optimal path is APQB, with AP and BQ tangent to the level line at \( h_1 \) (Figure 3).

4. DISCUSSION

For a sloped terrain, one may consider a ‘walking speed’ of 4km/hr as measured on the slope (and not projected onto the horizontal as was assumed in the above solution). This leads to a different solution from the one given in [6]. This effectively adds an additional time due to walking along the longer incline for the uphill. For this interpretation the performance index (1) is replaced by

\[
T = \int_{\text{ascent}} \left( \frac{\sqrt{dz^2 + dx^2}}{v} + \mu dz \right) + \int_{\text{descent}} \left( \frac{\sqrt{dz^2 + dx^2}}{v} + \mu dz \right).
\] (49)

As reported, the corresponding solution may be found by unrolling the cone onto the plane. Instead, the interpretation we have given to problem lets one simply look at the projection as given on a topographical map. We feel that this information is simpler to use. Indeed, the necessary time can easily be estimated in the general case from the information provided on a topographic map. While the cone can be unrolled, and the optimal inferred from it graphically, this surely will no longer hold for the case of a dome shaped mountain, which cannot be flattened. On the other hand, the presented solution in Section 3, remains simple to implement.

If one compares the solution to both problems (1) and (49), for the cone shaped mountain then the solution in [6] gives an optimal time of 76.71037 minutes, and this path reaches an altitude of 62.72045 m. The solution given in Section 2 gives an optimal time of 76.608 min and the altitude reached is 68.809 m. How different are these two interpretations of the optimization? Assuming that ‘walking’ changes to ‘rockclimbing’ once the slope is such that the person standing straight on his feet touches the slope in front with the outstretched hand, it follows that for a typical person (me) the corresponding slope angle is about \( \alpha = 69.878 \) degree (See Figure 8). The climb of a slope at that angle for a horizontal distance of 100 m requires in MacKinnon’s interpretation of Naismith’s rule a time \( t_M = \frac{100}{\cos \alpha} + 100 \tan \alpha \frac{1}{10} = 31.65 \) minutes.
Fig. 8. Limit: walking - climbing

With the model of Section 2, this time is
\[ t = \frac{60}{4000} + \frac{100 \tan \alpha}{10} = 28.79 \text{ minutes}. \]
In this extreme case the two estimates differ by less than 10%.

Not much has actually been published about Naismith’s rule, and therefore it is difficult to guess which interpretation Naismith had in mind. Somewhat amusing exercises in modeling are given in [2, 9]. The first seems to incline towards MacKinnon’s interpretation, the second is ambiguous. Some websites [11, 12] interpret the rules. The second of these corroborates our interpretation, assuming that a walker can maintain a speed of 5 km/hr on level ground, but that 1 hour needs to be added for every 600 meter of ascent. Either way, if one optimizes the path for climbing to the top of a conical mountain in minimum time, the path is the same: straight up. Well seasoned climbers know very well that this may not be the path that minimizes metabolic effort [7]. An optimal slope exists for the latter, so that one rather zigzags up towards the top [4]. Let us close in citing one of the authors’s comments, namely that Naismith was an optimist and it is proposed to add 50 % to his estimates.

REFERENCES

[12] www.geog.leeds.ac.uk/pgrads/s.fritz/gisril98.html