The Brachistochrone Problem over Surfaces

Pablo V. Negrón–Marrero*
Department of Mathematics
University of Puerto Rico
Humacao, PR 00791-4300

Abstract

We consider the problem of finding curves of minimum time of descent, joining two given points over a given frictionless surface and under the influence of gravity. We discuss the existence and minimality of extremals for the corresponding time functional, and find explicit solutions in certain special cases, including a closed form solution for the problem on an inclined plane. A discussion of numerical methods for computing these minimizers is given with several numerical examples for which explicit solutions are not known.

Keywords: brachistochrone, parameterized surface, calculus of variations, Euler–Lagrange equations

1 Introduction

The brachistochrone problem consists of finding the curve, joining two (non–vertical) given points, along which a bead of given mass falls without friction and under the influence of gravity, in the minimum time. This problem was first posed by Johann Bernoulli in 1696 and solved that same year by Newton, Leibniz, the Bernoulli brothers Johann and Jacob, and de L'Hôpital. The solution of the brachistochrone problem was pivotal to the development of the now very important branch of analysis called the calculus of variations. Since then variations of the brachistochrone problem, some of them including friction or variable force fields, have been presented in many books and papers (see e.g. [1], [3], [8], [11], [13], [14], [16], and [18]).

In this paper we discuss a variation of this problem that is hardly ever considered, namely the problem of finding curves of minimum time of descent over a given frictionless surface and under the influence of gravity. In [8] the problem over parametrized surfaces is considered but for orthogonal parametrizations. Since for a general surface, these

*pablo.negron1@upr.edu

orthogonal parametrizations are only possible locally, the explicit solutions given in [8] in terms of certain integrals, are only local. In this paper we do not assume that the parametrization of the surface is orthogonal. However as it turns out the existence of extremals and their minimality can only be guaranteed locally. We mention here as well the work in [19] where the surface instead of been described parametrically, it is constructed or approximated using splines applied to USGS elevation data.

By using conservation of energy, we do a derivation of the *time integral* functional describing the total time of descent along a given curve over a surface. We consider parametrized surfaces which are the image of a function $\mathbf{\Phi} : D \to \mathbb{R}^3$ where $D \subset \mathbb{R}^2$. Curves over the surface are thus given by a composition of $\mathbf{\Phi}$ together with a curve $\boldsymbol{\psi}$ over D. From the time functional we can formally derive the Euler–Lagrange equations for $\boldsymbol{\psi}$. These equations together with the requirement that the curve over the original surface joins two given points on the surface, yields a nonlinear boundary value problem (BVP) characterizing the extremals for the time integral functional. However, the resulting differential equation does not satisfy any of the standard requirements (c.f. [12]) for the existence of solutions of this BVP. Nonetheless we can invoke a result from Picard [17] to get that our BVP has a solution which is unique provided that the two points that the curve is to connect, are "sufficiently" close. It is interesting to note that our problem reduces to the one for geodesics when the constant for the acceleration due to gravity is set to zero. However, even for the geodesic problem, the results about the existence of extremals are still local.

Our problem (and the one for geodesics as well) is what is called a "parameter form problem" of the calculus of variations (cf. [7, Chap. 2, Sec. 10]). Exploiting this structure and with a local result for extremals like the one described in the previous paragraph, Bliss [2] shows how to construct an exact field for a parameter form problem, and that at least locally, the extremals are weak minimizers (in the C^1 norm) of the original functional. This result holds as a special case for our problem as well.

This paper is mostly expository. Our main contribution is to put together on a single reference different results concerning the existence and minimality of extremals, explicit solutions in special cases, and a discussion of numerical methods for computing the desired minimizers. Numerical schemes for the calculations of the minimizing extremals in this problem are usually overlooked. This is not a trivial matter, requiring the combination of several numerical techniques for the efficient calculation of these minimizers.

2 Motion over a surface

Consider a surface $S \subset \mathbb{R}^3$ with parametrization $\mathbf{\Phi} : D \to \mathbb{R}^3$ given by

$$(x, y, z) = \mathbf{\Phi}(u, v) = (\hat{x}(u, v), \hat{y}(u, v), \hat{z}(u, v))^T \quad (u, v) \in D, \tag{1}$$

where $D \subset \mathbb{R}^2$ is open and $\mathbf{\Phi} \in C^2$. We assume that there is a constant gravitational field pointing in the negative z direction, with a constant g of acceleration.

Given a curve $\boldsymbol{\psi}$ on D, where

$$\boldsymbol{\psi}(\tau) = (\hat{u}(\tau), \hat{v}(\tau))^T, \quad \tau \in [\tau_1, \tau_2],$$

a curve $\boldsymbol{\sigma}$ over S is thus given by

$$\boldsymbol{\sigma}(\tau) = \boldsymbol{\Phi}(\boldsymbol{\psi}(\tau)) = \boldsymbol{\Phi}(\hat{u}(\tau), \hat{v}(\tau)), \quad \tau \in [\tau_1, \tau_2].$$

From this we get that

$$\boldsymbol{\sigma}'(\tau) = \mathrm{D}\boldsymbol{\Phi}(\boldsymbol{\psi}(\tau))\boldsymbol{\psi}'(\tau),$$

where $\mathrm{D}\boldsymbol{\Phi}(u, v) = [\boldsymbol{\Phi}_u(u, v), \boldsymbol{\Phi}_v(u, v)]$. The element of arc length over $\boldsymbol{\sigma}$ is thus given by

$$\mathrm{d}s = \|\boldsymbol{\sigma}'(\tau)\| \, \mathrm{d}\tau = \|\mathrm{D}\boldsymbol{\Phi}(\boldsymbol{\psi}(\tau))\boldsymbol{\psi}'(\tau)\| \, \mathrm{d}\tau.$$

We let $v(\tau)$ be the speed of a particle moving over the surface at the point $\boldsymbol{\sigma}(\tau)$. Conservation of energy now implies that

$$\frac{1}{2}mv(\tau)^2 + mg\hat{z}(\boldsymbol{\psi}(\tau)) = \frac{1}{2}mv_0^2 + mgz_a,$$

where v_0 is the initial speed and z_a the initial height. Thus

$$v(\tau) = \sqrt{\alpha_0 - 2g\hat{z}(\boldsymbol{\psi}(\tau))}, \tag{2}$$

with $\alpha_0 = v_0^2 + 2gz_a$. It follows now that the total time of descend of a particle of mass m along the curve $\boldsymbol{\sigma}$ over S is given by

$$T[\boldsymbol{\psi}] = \int_{\tau_1}^{\tau_2} \frac{\|\mathrm{D}\boldsymbol{\Phi}(\boldsymbol{\psi}(\tau))\boldsymbol{\psi}'(\tau)\|}{\sqrt{\alpha_0 - 2g\hat{z}(\boldsymbol{\psi}(\tau))}} \mathrm{d}\tau. \tag{3}$$

We seek to minimize $T[\cdot]$ over the set of admissible[1] curves $\boldsymbol{\psi}$ given by

$$\begin{aligned}\mathcal{A} = \big\{\boldsymbol{\psi} \in C^1[\tau_1, \tau_2] \,:\, &\boldsymbol{\psi}(\tau) \in D \;\forall \tau \in [\tau_1, \tau_2],\; \boldsymbol{\Phi}(\boldsymbol{\psi}(\tau_1)) = \mathbf{a},\; \boldsymbol{\Phi}(\boldsymbol{\psi}(\tau_2)) = \mathbf{b}, \\ &\alpha_0 - 2g\hat{z}(\boldsymbol{\psi}(\tau)) > 0 \;\forall \tau \in [\tau_1, \tau_2],\; \boldsymbol{\psi}'(\tau) \neq \mathbf{0} \;\forall \tau \in [\tau_1, \tau_2]\big\},\end{aligned} \tag{4}$$

where $\mathbf{a}, \mathbf{b} \in S$. It follows from the results in [4] or [7], that the value of the time integral (3) is independent of the parametrization of the curve $\boldsymbol{\psi}$. That is, if $\boldsymbol{\psi}$ is re-parametrized via a one to one and onto re-parametrization, the value of $T[\boldsymbol{\psi}]$ does not change.

[1] The set of admissible functions could be defined for functions with piecewise continuous first derivative. However as we will see in Proposition 3.3, the extremals of the functional (3) must have continuous first derivatives.

3 Existence and minimizing properties of extremals

If we let

$$f(\mathbf{y},\mathbf{z}) = \frac{\|\mathrm{D}\mathbf{\Phi}(\mathbf{y})\mathbf{z}\|}{\sqrt{\alpha_0 - 2g\hat{z}(\mathbf{y})}}, \quad \mathbf{y},\mathbf{z} \in \Omega, \tag{5}$$

where

$$\Omega = \left\{(\mathbf{y},\mathbf{z}) \in D \times \mathbb{R}^2 \,:\, \alpha_0 - 2g\hat{z}(\mathbf{y}) > 0,\ \mathbf{z} \neq \mathbf{0}\right\}, \tag{6}$$

then

$$\begin{aligned} f_{\mathbf{z}}(\mathbf{y},\mathbf{z}) &= \frac{\mathrm{D}\mathbf{\Phi}(\mathbf{y})^T\mathrm{D}\mathbf{\Phi}(\mathbf{y})\mathbf{z}}{\|\mathrm{D}\mathbf{\Phi}(\mathbf{y})\mathbf{z}\|\,\sqrt{\alpha_0 - 2g\hat{z}(\mathbf{y})}}, & (7)\\ f_{\mathbf{y}}(\mathbf{y},\mathbf{z}) &= g\,\frac{\|\mathrm{D}\mathbf{\Phi}(\mathbf{y})\mathbf{z}\|}{(\alpha_0 - 2g\hat{z}(\mathbf{y}))^{\frac{3}{2}}}\,\vec{\nabla}\hat{z}(\mathbf{y}) + \frac{(\mathbf{z}\mathbf{z}^T) : (\mathrm{D}\mathbf{\Phi}(\mathbf{y})^T\mathrm{D}^2\mathbf{\Phi}(\mathbf{y}))}{\|\mathrm{D}\mathbf{\Phi}(\mathbf{y})\mathbf{z}\|\,\sqrt{\alpha_0 - 2g\hat{z}(\mathbf{y})}}, & (8) \end{aligned}$$

where

$$(\mathbf{z}\mathbf{z}^T) : (\mathrm{D}\mathbf{\Phi}(\mathbf{y})^T\mathrm{D}^2\mathbf{\Phi}(\mathbf{y})) = [\mathbf{z}^T(\mathrm{D}\mathbf{\Phi}(\mathbf{y})^T\mathrm{D}\mathbf{\Phi}_u(\mathbf{y}))\mathbf{z}, \mathbf{z}^T(\mathrm{D}\mathbf{\Phi}(\mathbf{y})^T\mathrm{D}\mathbf{\Phi}_v(\mathbf{y}))\mathbf{z}]^T.$$

The Euler–Lagrange equations for the functional $T[\cdot]$ are now given by

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\left[f_{\mathbf{z}}(\boldsymbol{\psi}(\tau),\boldsymbol{\psi}'(\tau))\right] = f_{\mathbf{y}}(\boldsymbol{\psi}(\tau),\boldsymbol{\psi}'(\tau)), \tag{9}$$

or in expanded form:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\tau}\left[\frac{B(\boldsymbol{\psi}(\tau))\boldsymbol{\psi}'(\tau)}{\Lambda_1(\boldsymbol{\psi}(\tau),\boldsymbol{\psi}'(\tau))\Lambda_2(\boldsymbol{\psi}(\tau))}\right] &= g\,\frac{\Lambda_1(\boldsymbol{\psi}(\tau),\boldsymbol{\psi}'(\tau))}{\Lambda_2^3(\boldsymbol{\psi}(\tau))}\,\vec{\nabla}\hat{z}(\boldsymbol{\psi}(\tau))\\ &\quad + \frac{(\boldsymbol{\psi}'(\tau)\boldsymbol{\psi}'(\tau)^T) : (\mathrm{D}\mathbf{\Phi}(\boldsymbol{\psi}(\tau))^T\mathrm{D}^2\mathbf{\Phi}(\boldsymbol{\psi}(\tau)))}{\Lambda_1(\boldsymbol{\psi}(\tau),\boldsymbol{\psi}'(\tau))\Lambda_2(\boldsymbol{\psi}(\tau))}, \quad (10) \end{aligned}$$

where

$$\begin{aligned} B(\boldsymbol{\psi}(\tau)) &= \mathrm{D}\mathbf{\Phi}(\boldsymbol{\psi}(\tau))^T\mathrm{D}\mathbf{\Phi}(\boldsymbol{\psi}(\tau)),\\ \Lambda_1(\boldsymbol{\psi}(\tau),\boldsymbol{\psi}'(\tau)) &= \|\mathrm{D}\mathbf{\Phi}(\boldsymbol{\psi}(\tau))\boldsymbol{\psi}'(\tau)\| = \sqrt{\boldsymbol{\psi}'(\tau)^T B(\boldsymbol{\psi}(\tau))\boldsymbol{\psi}'(\tau)},\\ \Lambda_2(\boldsymbol{\psi}(\tau)) &= \sqrt{\alpha_0 - 2g\hat{z}(\boldsymbol{\psi}(\tau))}. \end{aligned}$$

Note that B is the metric tensor of the surface S.

We now study the matrix $f_{\mathbf{z}\mathbf{z}}$, the hessian matriz of f with respect to the $\mathbf{z}$ variable.

Proposition 3.1. *Let* $\mathrm{D}\mathbf{\Phi}(\mathbf{y})$ *be of full rank for all* $\mathbf{y} \in D$*. Then for all* $(\mathbf{y},\mathbf{z}) \in \Omega$*, the matrix* $f_{\mathbf{z}\mathbf{z}}(\mathbf{y},\mathbf{z})$ *is positive semi–definite with null space span*$\{\mathbf{z}\}$*. Moreover, the matrix*

$$M(\mathbf{y},\mathbf{z}) = \begin{pmatrix} f_{\mathbf{z}\mathbf{z}}(\mathbf{y},\mathbf{z}) & \mathbf{z}\\ \mathbf{z}^T & 0 \end{pmatrix}.$$

is nonsingular.

Proof: Setting $A = \mathrm{D}\mathbf{\Phi}(\mathbf{y})$ and $B = A^T A$, is easy to get that

$$f_{\mathbf{zz}}(\mathbf{y},\mathbf{z}) = \frac{1}{\|A\mathbf{z}\|^3 \sqrt{\alpha_0 - 2g\hat{z}(\mathbf{y})}} \left[\|A\mathbf{z}\|^2 B - B\mathbf{z}\mathbf{z}^T B \right].$$

Let

$$\langle \mathbf{p}, \mathbf{w} \rangle_B = \mathbf{p}^T B \mathbf{w}.$$

Since B is positive definite, this an inner product in $\mathbb{R}^2$ with induced norm $\|\mathbf{p}\|_B = \sqrt{\langle \mathbf{p}, \mathbf{p} \rangle_B}$. Noting that $\|A\mathbf{z}\| = \|\mathbf{z}\|_B$, we get that

$$\mathbf{w}^T f_{\mathbf{zz}}(\mathbf{y},\mathbf{z})\mathbf{w} = \frac{1}{\|A\mathbf{z}\|^3 \sqrt{\alpha_0 - 2g\hat{z}(\mathbf{y})}} \left[\|\mathbf{z}\|_B^2 \|\mathbf{w}\|_B^2 - \langle \mathbf{z}, \mathbf{w} \rangle_B^2 \right].$$

But by the Cauchy–Schwartz inequality,

$$\|\mathbf{z}\|_B^2 \|\mathbf{w}\|_B^2 - \langle \mathbf{z}, \mathbf{w} \rangle_B^2 \geq 0,$$

with equality if and only if $\mathbf{w}$ and $\mathbf{z}$ are proportional. Hence $f_{\mathbf{zz}}(\mathbf{y},\mathbf{z})$ is positive semi–definite with $N(f_{\mathbf{zz}}(\mathbf{y},\mathbf{z})) = \mathrm{span}\,\{\mathbf{z}\}$.

To show that $M(\mathbf{y},\mathbf{z})$ is nonsingular, assume that

$$M(\mathbf{y},\mathbf{z}) \begin{pmatrix} \mathbf{w} \\ \alpha \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ 0 \end{pmatrix}.$$

Then

$$f_{\mathbf{zz}}(\mathbf{y},\mathbf{z})\mathbf{w} + \alpha\mathbf{z} = \mathbf{0}, \quad \mathbf{z}^T\mathbf{w} = 0.$$

The first of these two equations implies that $\alpha\mathbf{z} \in \mathrm{Range}(f_{\mathbf{zz}}(\mathbf{y},\mathbf{z})) = \mathrm{span}\,\{\mathbf{z}\}^\perp$. Thus $\mathbf{z}^T(\alpha\mathbf{z}) = 0$, which implies that $\alpha = 0$. But then $f_{\mathbf{zz}}(\mathbf{y},\mathbf{z})\mathbf{w} = \mathbf{0}$, which implies $\mathbf{w} \in N(f_{\mathbf{zz}}(\mathbf{y},\mathbf{z})) = \mathrm{span}\,\{\mathbf{z}\}$, that is $\mathbf{w} = \beta\mathbf{z}$. Now since $\mathbf{z}^T\mathbf{w} = 0$, we get $\beta = 0$, and thus that $\mathbf{w} = \mathbf{0}$. Hence $M(\mathbf{y},\mathbf{z})$ is nonsingular. □

Remark 3.2. It follows from Proposition 3.1 that if $\mathbf{p} \in \mathrm{Range}(f_{\mathbf{zz}}(\mathbf{y},\mathbf{z}))$, then the system

$$M(\mathbf{y},\mathbf{z}) \begin{pmatrix} \mathbf{w} \\ \gamma \end{pmatrix} = \begin{pmatrix} \mathbf{p} \\ 0 \end{pmatrix},$$

has a unique solution $(\mathbf{w},\gamma)$ with $\gamma = 0$ and $\mathbf{w}$ satisfying

$$f_{\mathbf{zz}}(\mathbf{y},\mathbf{z})\mathbf{w} = \mathbf{p}, \quad \mathbf{z}^T\mathbf{w} = 0.$$

Thus $\mathbf{w} \in \mathrm{Range}(f_{\mathbf{zz}}(\mathbf{y},\mathbf{z})) = \mathrm{span}\,\{\mathbf{z}\}^\perp$. If we let $\mathbf{z} = (z_1, z_2)^T$, then $\mathrm{Range}(f_{\mathbf{zz}}(\mathbf{y},\mathbf{z})) = \mathrm{span}\left\{\mathbf{z}^\perp\right\}$ where $\mathbf{z}^\perp = (z_2, -z_1)^T$. Moreover $\mathbf{z}^\perp$ is an eigenvector of $f_{\mathbf{zz}}(\mathbf{y},\mathbf{z})$ corresponding to the eigenvalue $f_{z_1z_1}(\mathbf{y},\mathbf{z}) + f_{z_2z_2}(\mathbf{y},\mathbf{z}) > 0$. Hence

$$\mathbf{w} = \frac{\mathbf{p}}{f_{z_1z_1}(\mathbf{y},\mathbf{z}) + f_{z_2z_2}(\mathbf{y},\mathbf{z})}.$$

We record here, for future reference, the expressions for the second order partial derivatives of the function f. With $A = \mathrm{D}\boldsymbol{\Phi}(\mathbf{y})$ and $B = A^T A$ we have:

$$
\begin{aligned}
f_{z_l z_p} &= \frac{1}{\Lambda_1^3 \Lambda_2} \left[\Lambda_1^2 B_{lp} - (B_{lj} z_j)(B_{pj} z_j) \right], && \text{(11a)}\\
f_{z_l y_p} &= \frac{1}{\Lambda_1^3 \Lambda_2^3} \left[\Lambda_1^2 \Lambda_2^2 \frac{\partial B_{lj}}{\partial y_p} z_j + g \Lambda_1^2 (B_{lj} z_j) \frac{\partial \hat{z}}{\partial y_p} - \frac{1}{2} \Lambda_2^2 (B_{lj} z_j) \left(\frac{\partial B_{ij}}{\partial y_p} z_i z_j \right) \right], && \text{(11b)}
\end{aligned}
$$

where

$$
\begin{aligned}
\Lambda_1 &= \Lambda_1(\mathbf{y}, \mathbf{z}) = \|A\mathbf{z}\| = \sqrt{\mathbf{z}^T B \mathbf{z}},\\
\Lambda_2 &= \Lambda_2(\mathbf{y}) = \sqrt{\alpha_0 - 2g\hat{z}(\mathbf{y})}.
\end{aligned}
$$

We note that since $B_{ij} = A_{ki} A_{kj}$, then

$$
\frac{\partial B_{ij}}{\partial y_p} = A_{ki} \frac{\partial A_{kj}}{\partial y_p} + \frac{\partial A_{ki}}{\partial y_p} A_{kj} = \left[(\mathrm{D}\boldsymbol{\Phi})^T \mathrm{D}\boldsymbol{\Phi}_p + (\mathrm{D}\boldsymbol{\Phi}_p)^T \mathrm{D}\boldsymbol{\Phi} \right]_{ij},
$$

where $\boldsymbol{\Phi}_p = \frac{\partial \boldsymbol{\Phi}}{\partial y_p}$.

It follows from (11a) that

$$
\frac{f_{z_1 z_1}}{z_2^2} = \frac{f_{z_2 z_2}}{z_1^2} = -\frac{f_{z_1 z_2}}{z_1 z_2} = f_1,
$$

where $f_1 = f_1(\mathbf{y}, \mathbf{z})$ is given by:

$$
f_1 = \frac{\det B}{\Lambda_1^3 \Lambda_2}.
$$

Using this we get that the differential equation (9) (or (14a)) is equivalent to the following scalar equation:

$$
(\psi_1'' \psi_2' - \psi_2'' \psi_1') f_1 = f_{z_2 y_1} - f_{z_1 y_2}. \tag{12}
$$

Moreover, since $f_1 > 0$ over the set Ω (cf. (6)), we get from [4, Page 126] the following result establishing that extremals of (3) can not have jump discontinuities in their first derivative.

Proposition 3.3. *Let $\mathrm{D}\boldsymbol{\Phi}(\mathbf{y})$ be of full rank for all $\mathbf{y} \in D$. Then if $(\mathbf{c}_1, \mathbf{c}_2) \in \Omega$, any solution $\boldsymbol{\psi}$ of (9) such that $(\boldsymbol{\psi}(\tau_1), \boldsymbol{\psi}'(\tau_1)) = (\mathbf{c}_1, \mathbf{c}_2)$ for some τ_1, has $\boldsymbol{\psi}'$ continuous for as long as $(\boldsymbol{\psi}(\tau), \boldsymbol{\psi}'(\tau))$ remains in Ω.*

From (5) and (7) we get that:

$$
f_{\mathbf{z}}(\mathbf{y}, \mathbf{z}) \cdot \mathbf{z} = f(\mathbf{y}, \mathbf{z}).
$$

Upon differentiating this with respect to $\mathbf{y}$ and omitting arguments, we get

$$
[\mathrm{D}_{\mathbf{y}}(f_{\mathbf{z}})]^T \mathbf{z} = f_{\mathbf{y}}, \tag{13}
$$

where $\mathrm{D}_{\mathbf{y}}(f_{\mathbf{z}}) = (f_{z_i y_j})$. Using this we can now show the following:

Proposition 3.4. *Suppose that $\boldsymbol{\psi}$ is a C^2 solution of the Euler–Lagrange equation* (9)*. Then under the condition that $\boldsymbol{\psi}' \cdot \boldsymbol{\psi}'' = 0$, we have that*

$$\boldsymbol{\psi}''(\tau) = \left[\frac{f_{z_2y_1} - f_{z_1y_2}}{f_{z_1z_1} + f_{z_2z_2}}\right] (\psi_2'(\tau), -\psi_1'(\tau))^T,$$

where the arguments of the derivatives of f are $(\boldsymbol{\psi}(\tau), \boldsymbol{\psi}'(\tau))$. In particular for any $(\mathbf{c}_1, \mathbf{c}_2) \in \Omega$, this equation has a solution satisfying $\boldsymbol{\psi}(0) = \mathbf{c}_1$, $\boldsymbol{\psi}'(0) = \mathbf{c}_2$ which is unique and exists on a maximal interval J such that $(\boldsymbol{\psi}(\tau), \boldsymbol{\psi}'(\tau)) \in \Omega$ for all $\tau \in J$.

Proof: Since $\boldsymbol{\psi}$ is C^2, it follows upon expanding (9) and using (13), that

$$\mathrm{D}_{\mathbf{y}}(f_{\mathbf{z}})\boldsymbol{\psi}'(\tau) + f_{\mathbf{zz}}\boldsymbol{\psi}''(\tau) = [\mathrm{D}_{\mathbf{y}}(f_{\mathbf{z}})]^T \boldsymbol{\psi}'(\tau),$$

which upon rearrangement leads to

$$f_{\mathbf{zz}}\boldsymbol{\psi}''(\tau) = ([\mathrm{D}_{\mathbf{y}}(f_{\mathbf{z}})]^T - \mathrm{D}_{\mathbf{y}}(f_{\mathbf{z}}))\boldsymbol{\psi}'(\tau).$$

But

$$([\mathrm{D}_{\mathbf{y}}(f_{\mathbf{z}})]^T - \mathrm{D}_{\mathbf{y}}(f_{\mathbf{z}}))\boldsymbol{\psi}'(\tau) = (f_{z_2y_1} - f_{z_1y_2})\,(\psi_2'(\tau), -\psi_1'(\tau))^T.$$

The result now follows from Proposition 3.1 and Remark 3.2. The last part of the statement follows from the standard existence and uniqueness theorem for initial value problems. □

This result by itself does not guarantee the existence of extremals for our problem satisfying the boundary conditions in $\mathcal{A}$ (cf. (4)). However we can use a result of Picard [17] and its modification by Bliss [2] to parameter form problems of the calculus of variations, to get that if $\mathbf{b}$ is sufficiently close to $\mathbf{a}$, this boundary value problem has a unique solution. In the statement of the following theorem, the parameter τ is taken to be that of arc length along the curve $\boldsymbol{\psi}(\cdot)$.

Theorem 3.5. *Let $\boldsymbol{\Phi} \in C^2(D)$ be 1–1 and with $\mathrm{D}\boldsymbol{\Phi}(\mathbf{y})$ of full rank for all $\mathbf{y} \in D$. Then if $\mathbf{b}$ is sufficiently close to $\mathbf{a}$, there exists $\tau^* > 0$ such that the boundary value problem*

$$\boldsymbol{\psi}''(\tau) = \left[\frac{f_{z_2y_1} - f_{z_1y_2}}{f_{z_1z_1} + f_{z_2z_2}}\right] (\psi_2'(\tau), -\psi_1'(\tau))^T, \tag{14a}$$

$$\boldsymbol{\Phi}(\boldsymbol{\psi}(0)) = \mathbf{a}, \quad \boldsymbol{\Phi}(\boldsymbol{\psi}(\tau^*)) = \mathbf{b}, \tag{14b}$$

has a solution which is unique and such that $\|\boldsymbol{\psi}'(\tau)\| = 1$ and $(\boldsymbol{\psi}(\tau), \boldsymbol{\psi}'(\tau)) \in \Omega$ for $0 \leq \tau \leq \tau^*$.

We can now use the results in [2] together with Theorem 3.5 to show that one can construct an exact field of extremals for the functional (3) and that at least locally, the extremal in Theorem 3.5 is a weak minimizer (in the C^1 norm) for (3) over $\mathcal{A}$ whenever $\mathbf{b}$ is sufficiently close to $\mathbf{a}$.

Remark 3.6. Another consequence of Proposition 3.1 is that the system (9), or equivalently (10), reduces to a single scalar equation (cf. equation (12)). Thus an additional condition is required to determine the solution extremal. That is precisely what is done in Theorem 3.5 where the extremal is parametrized by arc length which imposes the condition that $\boldsymbol{\psi}' \cdot \boldsymbol{\psi}' = 1$. More generally, one could impose the condition that $\boldsymbol{\psi}' \cdot \boldsymbol{\psi}' = c$ along the curve, for some positive constant c. The parameter c plays a particular role in the numerical schemes for computing extremals that we discuss in Section 5.

4 Special cases

We now consider several special cases of parameterized surfaces, some of which first integrals of (10) can be obtained by inspection.

4.1 Surface of revolution

We consider first the case of a surface of revolution where the axis of rotation is the z axis. Such a surface can be parametrized, with $a_v, b_v \in \mathbb{R} \cup \{-\infty, \infty\}$ and $a_v < b_v$, by $\boldsymbol{\Phi} : [0, 2\pi) \times (a_v, b_v) \to \mathbb{R}^3$ where:

$$\boldsymbol{\Phi}(u, v) = [p(v)\cos(u), p(v)\sin(u), q(v)]^T, \tag{15}$$

where p and q are smooth functions with p positive. It follows now, omitting arguments, that

$$\begin{aligned}
\mathrm{D}\boldsymbol{\Phi}^T \mathrm{D}\boldsymbol{\Phi} &= \begin{bmatrix} p(v)^2 & 0 \\ 0 & (p'(v))^2 + (q'(v))^2 \end{bmatrix}, \\
\mathrm{D}\boldsymbol{\Phi}^T \mathrm{D}\boldsymbol{\Phi}_u &= \begin{bmatrix} 0 & p(v)p'(v) \\ -p(v)p'(v) & 0 \end{bmatrix}, \\
\mathrm{D}\boldsymbol{\Phi}^T \mathrm{D}\boldsymbol{\Phi}_v &= \begin{bmatrix} p(v)p'(v) & 0 \\ 0 & p''(v)p'(v) + q''(v)q'(v) \end{bmatrix}.
\end{aligned}$$

It follows now that (15) represents a set orthogonal coordinates over the surface S, and that equations (10) have as a first integral that

$$p(\hat{v}(\tau))^2 \hat{u}'(\tau) = c_1 \Lambda_1(\tau) \Lambda_2(\tau), \tag{16}$$

for some constant c_1 and with

$$\begin{aligned}
\Lambda_1(\tau) &= \sqrt{p(\hat{v}(\tau))^2 \hat{u}'(\tau)^2 + (p'(\hat{v}(\tau))^2 + q'(\hat{v}(\tau))^2)\hat{v}'(\tau)^2}, && \text{(17a)} \\
\Lambda_2(\tau) &= \sqrt{\alpha_0 - 2gq(\hat{v}(\tau))}. && \text{(17b)}
\end{aligned}$$

From Proposition 3.3 we get that the constant c_1 must have the same value all along the solution curve. Note that $c_1 = 0$ if and only if $\hat{u}'(\tau) = 0$ for all τ, and $\mathrm{sign}(\hat{u}') = \mathrm{sign}(c_1)$

when $c_1 \neq 0$. If $c_1 \neq 0$, then (16) gives a differential equation for $\hat{v}$ as a function of u, which in principle could be solved either explicitly or numerically, up to the determination of the constant c_1 and another constant of integration. When $c_1 = 0$, since $\hat{u}'(\tau) = 0$ for all τ, we have that both of the given points $\mathbf{a}$ and $\mathbf{b}$ on the surface, must belong to the same meridional section. In this case as well, it is easy to check that the second equation in (9) is automatically satisfied for any function $\hat{v}$, and thus imposes no condition. This is because the system (9) is always under–determined (cf. Proposition 3.1) and thus requires an extra equation or condition. This extra condition can be chosen (cf. Remark 3.6) as $\boldsymbol{\psi}' \cdot \boldsymbol{\psi}' = \text{const}$. Using this we get that if $c_1 = 0$, then we must have $\hat{u} = \text{const}$ and $\hat{v}' = \text{const}$.

4.1.1 A sphere

We now consider the case of a sphere of radius a with centre at the origin, without the poles. Thus we take in (15)

$$p(v) = a\cos(v), \quad q(v) = a\sin(v), \quad (a_v, b_v) = \left(-\frac{\pi}{2}, \frac{\pi}{2}\right). \tag{18}$$

It follows now that (16) and (17) reduce respectively to

$$a^2\cos^2(\hat{v}(\tau))\hat{u}'(\tau) = c_1\Lambda_1(\tau)\Lambda_2(\tau). \tag{19}$$

and

$$\begin{aligned} \Lambda_1(\tau) &= a\sqrt{\cos^2(\hat{v}(\tau))\hat{u}'(\tau)^2 + \hat{v}'(\tau)^2}, \\ \Lambda_2(\tau) &= \sqrt{\alpha_0 - 2ag\sin(\hat{v}(\tau))}. \end{aligned}$$

Assuming $c_1 \neq 0$, we must have that $\hat{u}'(\tau) \neq 0$ for all τ. Thus $\hat{v}(\cdot)$ can be expressed as a function of $\hat{u}$. In this case and writing u instead of $\hat{u}$, we get after squaring both sides of (19) that:

$$\hat{v}'(u)^2 = \cos^2(\hat{v})\left[\frac{a^2\cos^2(\hat{v}) - c_1^2\Lambda_2^2(\hat{v})}{c_1^2\Lambda_2^2(\hat{v})}\right], \tag{20}$$

where the prime denotes now derivative with respect to u and

$$\Lambda_2(v) = \sqrt{\alpha_0 - 2ag\sin(v)}.$$

After taking square roots on both sides of (20) and separating variables, we are led to the following expression:

$$u = \pm\int \sec(\hat{v})\left[\frac{c_1^2\Lambda_2^2(\hat{v})}{a^2\cos^2(\hat{v}) - c_1^2\Lambda_2^2(\hat{v})}\right]^{\frac{1}{2}} \mathrm{d}\hat{v} + c_2. \tag{21}$$

The constants c_1 and c_2, in principle, could be determined now from the condition that the curve over the sphere joins the given points $\mathbf{a}$ and $\mathbf{b}$. However, formula (21) is of

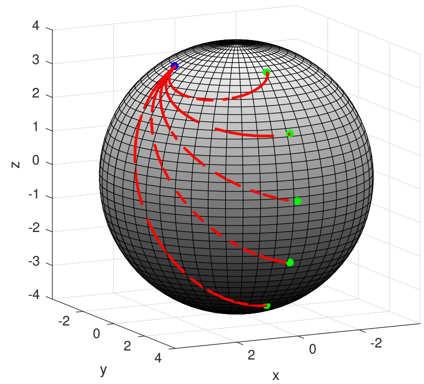

Figure 1: Several brachistochrone curves over a sphere.

little use in numerical calculations because the sign in this formula can change along the curve, at points unknown before hand. Still the formula could be useful in those cases in which the integral can be computed explicitly.

In Figure 1 we show a numerical simulation (see Section 5) of various brachistochrone curves (in red) over a sphere of radius $a = 4$, with $g = 9.8$ and $v_0 = 0.1$ (MKS system). The initial point (in blue) for all curves is $\mathbf{\Phi}(0, \frac{\pi}{3})$, where $\mathbf{\Phi}$ is given as in (18), and the final points (in green) are $\mathbf{\Phi}(\frac{\pi}{2}, v)$ with $v = -\frac{\pi}{3}, -0.5, 0, 0.5, \frac{\pi}{3}$. The corresponding values for the constant c_1 in (19) as well as the minimum times of descent and arc length for each curve, are given in the following table:

v	c_1	time of descent	length of curve
$-\frac{\pi}{3}$	0.16969	1.7357	10.0002
-0.5	0.33303	1.6963	8.7419
0	0.47778	1.6508	7.262
0.5	0.63612	1.63	5.7427
$\frac{\pi}{3}$	0.80386	2.1586	4.8678

Note that the time of descent for the brachistochrone with $v = \frac{\pi}{3}$ is the largest even though it is the shortest curve. This due to the fact that this curve, been close to horizontal, then the effect due to the acceleration of gravity is less significant on the motion.

For the case of geodesics on the sphere where $g = 0$, the integral in (21) can be computed explicitly yielding that

$$u = \pm \sin^{-1}(\beta \tan(\hat{v})) + c_2, \tag{22}$$

where

$$\beta = \frac{K}{\sqrt{1 - K^2}}, \quad K = \frac{|c_1|\, v_0}{a}.$$

4.1.2 Right circular cylinders

For any $a > 0$, we take now in (15)

$$p(v) = a, \quad q(v) = v, \quad (a_v, b_v) = (-\infty, \infty). \tag{23}$$

Equations (16) and (17) reduce now to

$$a^2 \hat{u}'(\tau) = c_1 \Lambda_1(\tau) \Lambda_2(\hat{v}(\tau)), \tag{24}$$

for some constant c_1, and where

$$\Lambda_1(\tau) = \sqrt{a^2(\hat{u}'(\tau))^2 + (\hat{v}'(\tau))^2}, \quad \Lambda_2(\hat{v}(\tau)) = \sqrt{\alpha_0 - 2g\hat{v}(\tau))}.$$

Provided $c_1 \neq 0$ this leads to the separable differential equation

$$\frac{\mathrm{d}\hat{v}}{\mathrm{d}u} = \pm a \left[\frac{a^2 - c_1^2 \Lambda_2^2(\hat{v})}{c_1^2 \Lambda_2^2(\hat{v})} \right]^{\frac{1}{2}},$$

which has general solution

$$u = \pm \frac{a}{2gc_1^2} \left[\frac{1}{a^2} \sqrt{c_1^2 \Lambda_2^2(\hat{v})(a^2 - c_1^2 \Lambda_2^2(\hat{v}))} + \sin^{-1} \left(\frac{1}{a} \sqrt{a^2 - c_1^2 \Lambda_2^2(\hat{v})} \right) \right] + c_2.$$

Introducing the parameter θ such that

$$\frac{\mathrm{d}\hat{v}}{\mathrm{d}u} = a \tan\theta,$$

we get that

$$\hat{u}(\theta) = \pm \frac{a}{4gc_1^2}[2\theta + \sin 2\theta] + c_2, \quad \hat{v}(\theta) = \frac{1}{2gc_1^2}[c_1^2 \alpha_0 - a^2 \cos^2\theta].$$

It is interesting to note that these equations represent a cycloid in the uv plane, while the time integral minimizing curve is given by

$$\boldsymbol{\sigma}(\theta) = (a\cos\hat{u}(\theta), a\sin\hat{u}(\theta), \hat{v}(\theta))^T.$$

4.1.3 One–sheet circular hyperboloid

For $a > 0$ and $c > 0$, we take in (15)

$$p(v) = a\cosh(v), \quad q(v) = c\sinh(v), \quad (a_v, b_v) = (-\infty, \infty). \tag{25}$$

In this case, equations (16) and (17) reduce now to

$$a^2 \cosh^2(\hat{v}) \hat{u}'(\tau) = c_1 \Lambda_1(\tau) \Lambda_2(\hat{v}(\tau)), \tag{26}$$

for some constant c_1, and where

$$\begin{aligned}\Lambda_1(\tau) &= \sqrt{a^2\cosh^2(\hat v)(\hat u'(\tau))^2 + (a^2\sinh^2(\hat v) + c^2\cosh^2(\hat v))(\hat v'(\tau))^2},\\ \Lambda_2(\hat v(\tau)) &= \sqrt{\alpha_0 - 2gc\sinh(\hat v(\tau))}.\end{aligned}$$

Provided $c_1 \neq 0$ this leads to the separable differential equation

$$\frac{\mathrm{d}\hat v}{\mathrm{d}\hat u} = \pm\left[\frac{a^2\cosh^2(\hat v)}{a^2\sinh^2(\hat v) + c^2\cosh^2(\hat v)}\,\frac{a^2\cosh^2(\hat v) - c_1^2\Lambda_2^2(\hat v)}{c_1^2\Lambda_2^2(\hat v)}\right]^{\frac{1}{2}}.$$

This equation is not integrable in closed form even in the special case in which $a = c$ and $g = 0$! In Example 5.1 we present some numerical simulations for this surface.

4.2 An explicit surface

Let $\hat z : D \to \mathbb{R}$, where $D \subset \mathbb{R}^2$ is open, be a C^2 function. We consider the surface S parametrized by:

$$\mathbf{\Phi}(u,v) = (u, v, \hat z(u,v))^T, \quad (u,v) \in D. \tag{27}$$

We have now (omitting arguments) that

$$\begin{aligned} B &= \mathrm{D}\mathbf{\Phi}^T\mathrm{D}\mathbf{\Phi} = \begin{bmatrix} 1+\hat z_u^2 & \hat z_u\hat z_v \\ \hat z_u\hat z_v & 1+\hat z_v^2\end{bmatrix},\\ \mathrm{D}\mathbf{\Phi}^T\mathrm{D}\mathbf{\Phi}_u &= \begin{bmatrix} \hat z_u\hat z_{uu} & \hat z_u\hat z_{uv} \\ \hat z_v\hat z_{uu} & \hat z_v\hat z_{uv}\end{bmatrix}, \quad \mathrm{D}\mathbf{\Phi}^T\mathrm{D}\mathbf{\Phi}_v = \begin{bmatrix} \hat z_u\hat z_{uv} & \hat z_u\hat z_{vv} \\ \hat z_v\hat z_{uv} & \hat z_v\hat z_{vv}\end{bmatrix}.\end{aligned}$$

We record here for latter reference the coefficient of the right hand side of (14a). Using (5) and (11), a lengthy but otherwise elementary computation yields that in this case,

$$\frac{f_{z_2y_1} - f_{z_1y_2}}{f_{z_1z_1} + f_{z_2z_2}} = \frac{(\hat z_{uu}(\hat u')^2 + 2\hat z_{uv}\hat u'\hat v' + \hat z_{vv}(\hat v')^2)\Lambda_2^2(\tau) - g\Lambda_1^2(\tau)}{\Lambda_2^2(\tau)(1+\hat z_u^2+\hat z_v^2)((\hat u')^2+(\hat v')^2)}\,(\hat z_v\hat u' - \hat z_u\hat v'), \tag{28}$$

where

$$\begin{aligned}\Lambda_1(\tau) &= \sqrt{(1+\hat z_u^2)(\hat u')^2 + 2\hat z_u\hat z_v\hat u'\hat v' + (1+\hat z_v^2)(\hat v')^2},\\ \Lambda_2(\tau) &= \sqrt{\alpha_0 - g\hat z(\hat u,\hat v)}.\end{aligned}$$

Note that the numerator of the fraction on the right hand side of (28) is the quadratic form of the matrix $\Lambda_2^2 H_{\hat z}(\hat u,\hat v) - gB(\hat u,\hat v)$ where $H_{\hat z}$ is the hessian matrix of $\hat z$.

In general is not obvious how to get first integrals of (10) in this case. We now consider two special cases in which this can be done.

4.2.1 Generalized cylinders

We consider the case in which $\hat{z}$ in (27) depends only on u. It is easy to check that in this case equation (10) yields the first integral:

$$\hat{v}'(\tau) = c_1 \Lambda_1(\tau)\Lambda_2(\hat{u}(\tau)), \tag{29}$$

where

$$\Lambda_1(\tau) = \sqrt{(1+\hat{z}_u^2)(\hat{u}'(\tau))^2 + (\hat{v}'(\tau))^2}, \quad \Lambda_2(\hat{u}(\tau)) = \sqrt{\alpha_0 - 2g\hat{z}(\hat{u}(\tau))}.$$

If $c_1 = 0$ then $\hat{v}$ must be constant and there are extremals if and only if the points $\mathbf{a}, \mathbf{b}$ in the boundary conditions have the same v coordinate. In the case $c_1 \neq 0$, then $\hat{v}'(\tau) \neq 0$, and after squaring we can write (29) as

$$\left[1 - c_1^2\Lambda_2^2(\hat{u}(\tau))\right](\hat{v}'(\tau))^2 = c_1^2\Lambda_2^2(\hat{u}(\tau))(1+\hat{z}_u^2)(\hat{u}'(\tau))^2.$$

Thus provided that $c_1^2\Lambda_2^2(\hat{u}(\tau)) < 1$, we can take $\hat{v}$ as the parameter, and writing v instead of $\hat{v}$ we get from the equation above that

$$\frac{\mathrm{d}\hat{u}}{\mathrm{d}v} = \pm\left[\frac{1 - c_1^2\Lambda_2^2(\hat{u})}{c_1^2\Lambda_2^2(\hat{u})(1+\hat{z}_u^2)}\right]^{\frac{1}{2}}.$$

Since $\hat{z}_u$ is only a function of u, this is a separable equation with general solution

$$v = \pm\int\left[\frac{c_1^2\Lambda_2^2(\hat{u})(1+\hat{z}_u^2)}{1 - c_1^2\Lambda_2^2(\hat{u})}\right]^{\frac{1}{2}} \mathrm{d}\hat{u} + c_2, \tag{30}$$

where the constants c_1 and c_2 are chosen to match the boundary conditions.

4.2.2 Non vertical planes

As a further special case of (27), we take the surface to be a non vertical plane which without lost of generality we can take it to be given by the equation $z = \hat{z}(u) = au$ and $D = \mathbb{R}^2$. Equation (30) now simplifies to

$$v = \pm\sqrt{1+a^2}\int\left[\frac{c_1^2\Lambda_2^2(\hat{u})}{1 - c_1^2\Lambda_2^2(\hat{u})}\right]^{\frac{1}{2}} \mathrm{d}\hat{u} + c_2,$$

where $\Lambda_2(u) = \sqrt{\alpha_0 - 2agu}$. This integral can be computed explicitly yielding that

$$v = \pm\frac{\sqrt{1+a^2}}{2agc_1^2}\left[\sqrt{c_1^2\Lambda_2^2(\hat{u})(1 - c_1^2\Lambda_2^2(\hat{u}))} + \sin^{-1}\left(\sqrt{1 - c_1^2\Lambda_2^2(\hat{u})}\right)\right] + c_2.$$

If we introduce the parameter θ by the requirement that

$$\frac{\mathrm{d}\hat{u}}{\mathrm{d}v} = \frac{1}{\sqrt{1+a^2}}\tan\theta,$$

then it follows from (29) that

$$c_1^2\Lambda_2^2(\hat{u}) = \cos^2\theta, \quad 1 - c_1^2\Lambda_2^2(\hat{u}) = \sin^2\theta,$$

and that

$$\hat{u}(\theta) = \frac{1}{2agc_1^2}[c_1^2\alpha_0 - \cos^2\theta], \quad \hat{v}(\theta) = \pm\frac{\sqrt{1+a^2}}{4agc_1^2}[2\theta + \sin 2\theta] + c_2.$$

These equations are the ones for a cycloid on the uv plane. In this case, they represent the projection on the xy plane of the time integral minimizing curve given by

$$\boldsymbol{\sigma}(\theta) = \boldsymbol{\Phi}(\boldsymbol{\psi}(\theta)) = (\hat{u}(\theta), \hat{v}(\theta), a\hat{u}(\theta))^T.$$

4.3 Vertical plane

This case is that of the classical brachistochrone problem. Without loss of generality we parametrized a vertical plane by:

$$\boldsymbol{\Phi}(u,v) = (0,u,v)^T, \quad (u,v) \in \mathbb{R}^2.$$

Thus in this case we have that

$$\begin{aligned} B &= \mathrm{D}\boldsymbol{\Phi}^T\mathrm{D}\boldsymbol{\Phi} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \\ \mathrm{D}\boldsymbol{\Phi}^T\mathrm{D}\boldsymbol{\Phi}_u &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathrm{D}\boldsymbol{\Phi}^T\mathrm{D}\boldsymbol{\Phi}_v = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}. \end{aligned}$$

It is easy to check that in this case equation (10) yields the first integral:

$$\hat{u}'(\tau) = c_1\Lambda_1(\tau)\Lambda_2(\hat{v}(\tau)), \tag{31}$$

where

$$\Lambda_1(\tau) = \sqrt{(\hat{u}'(\tau))^2 + (\hat{v}'(\tau))^2}, \quad \Lambda_2(\hat{v}(\tau)) = \sqrt{\alpha_0 - 2g\hat{v}(\tau)}.$$

Provided that $c_1 \neq 0$ we have now that

$$\frac{\mathrm{d}\hat{v}}{\mathrm{d}u} = \pm\left[\frac{1 - c_1^2\Lambda_2^2(\hat{v})}{c_1^2\Lambda_2^2(\hat{v})}\right]^{\frac{1}{2}}.$$

The general solution of this equation is given by

$$u = \pm\frac{1}{2gc_1^2}\left[\sqrt{c_1^2\Lambda_2^2(\hat{v})(1 - c_1^2\Lambda_2^2(\hat{v}))} + \sin^{-1}\left(\sqrt{1 - c_1^2\Lambda_2^2(\hat{v})}\right)\right] + c_2.$$

If we introduce the parameter θ by the requirement that

$$\frac{\mathrm{d}\hat{v}}{\mathrm{d}u} = \tan\theta,$$

then it follows that

$$\hat{u}(\theta) = \pm\frac{1}{4gc_1^2}[2\theta + \sin 2\theta] + c_2, \quad \hat{v}(\theta) = \frac{1}{2gc_1^2}[c_1^2\alpha_0 - \cos^2\theta].$$

These equations correspond to those of a *cycloid* on the yz plane.

5 Numerical results

In this section we describe the numerical scheme that we use for computing approximations of the minimizers of the time integral (3). The procedure consists of two stages: first we do a direct minimization of a suitable discretization of (3), and then use the computed approximate minimizer as a starting point in a shooting method to solve (14). Thus, the direct minimization works as a *predictor step* while the shooting method takes the role of a *corrector step*.

Minimization of the time integral: the predictor step

We assume that the parametrization is chosen so that $[\tau_1, \tau_2] = [0, 1]$. Let $h = 1/n$, $n \geq 1$, and define

$$\tau_j = jh, \quad 0 \leq j \leq n, \quad \tau_{j-\frac{1}{2}} = \frac{\tau_{j-1} + \tau_j}{2} = \left(j - \frac{1}{2}\right) h, \quad 1 \leq j \leq n.$$

We let $\boldsymbol{\psi}(\tau) = (\hat{u}(\tau), \hat{v}(\tau))^T$ and denote by $\boldsymbol{\psi}_j = (u_j, v_j)^T$ an approximation of $\boldsymbol{\psi}(\tau_j) = (\hat{u}(\tau_j), \hat{v}(\tau_j))^T$, $0 \leq j \leq n$. The *discretized* curve is given by the matrix:

$$\boldsymbol{\psi}^h = (\boldsymbol{\psi}_0, \boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_n) \in \mathbb{R}^{2\times(n+1)}. \tag{32}$$

Using the boundary conditions in (4) we set

$$\boldsymbol{\psi}_0 = \boldsymbol{\alpha} \equiv (\hat{u}(0), \hat{v}(0))^T, \quad \boldsymbol{\psi}_n = \boldsymbol{\beta} \equiv (\hat{u}(1), \hat{v}(1))^T,$$

where $\mathbf{\Phi}(\hat{u}(0), \hat{v}(0)) = \mathbf{a}$ and $\mathbf{\Phi}(\hat{u}(1), \hat{v}(1)) = \mathbf{b}$.

We introduce the approximations:

$$\boldsymbol{\psi}(\tau_{j-\frac{1}{2}}) \approx \frac{\boldsymbol{\psi}_{j-1} + \boldsymbol{\psi}_j}{2} = \boldsymbol{\psi}_{j-\frac{1}{2}}, \tag{33a}$$

$$\boldsymbol{\psi}'(\tau_{j-\frac{1}{2}}) \approx \frac{\boldsymbol{\psi}_j - \boldsymbol{\psi}_{j-1}}{h} = \delta\boldsymbol{\psi}_{j-\frac{1}{2}}. \tag{33b}$$

The mid–point rule for approximating integrals applied to (3) and (5), gives that

$$T[\boldsymbol{\psi}] \approx h \sum_{j=1}^{n} f\left(\boldsymbol{\psi}(\tau_{j-\frac{1}{2}}), \boldsymbol{\psi}'(\tau_{j-\frac{1}{2}})\right).$$

Using the approximations (33) we get the *discretized time integral*:

$$T_h[\boldsymbol{\psi}^h] = h \sum_{j=1}^{n} f\left(\boldsymbol{\psi}_{j-\frac{1}{2}}, \delta\boldsymbol{\psi}_{j-\frac{1}{2}}\right). \tag{34}$$

Note that $T_h[\boldsymbol{\psi}^h]$ is a function of $(\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \ldots, \boldsymbol{\psi}_{n-1})$.

The condition $\boldsymbol{\psi}' \cdot \boldsymbol{\psi}'' = 0$ present in Proposition 3.4 and Theorem 3.5 implies that $\boldsymbol{\psi}' \cdot \boldsymbol{\psi}' = c$ for some positive constant c. The constant c is unknown and characterizes the parametrization of the curve $\boldsymbol{\psi}$. By allowing this constant to be unknown is that we can take the interval for the parametrization to be $[0,1]$. The constraint $\boldsymbol{\psi}' \cdot \boldsymbol{\psi}' = c$ is equivalent to

$$\int_0^1 \left(\boldsymbol{\psi}'(\tau) \cdot \boldsymbol{\psi}'(\tau) - c\right)^2 \mathrm{d}\tau = 0.$$

Motivated by this we can state our discretized problem as:

$$\min_{\mathcal{A}_{h,c}} T_h[\boldsymbol{\psi}^h],$$

where

$$\mathcal{A}_{h,c} = \Big\{ (\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \ldots, \boldsymbol{\psi}_{n-1}) \in \mathbb{R}^{2\times(n-1)} : \alpha_0 - 2g\hat{z}(\boldsymbol{\psi}_{j-\frac{1}{2}}) > 0,\ 1 \le j \le n,$$
$$\boldsymbol{\psi}_j \in D,\ 1 \le j \le n-1,\ h\sum_{j=1}^{n}\left(\left\|\delta\boldsymbol{\psi}_{j-\frac{1}{2}}\right\|^2 - c\right)^2 = 0 \Big\}.$$

We use a penalty method to approximate solutions of this problem. That is, for $\mu > 0$ sufficiently large, we define

$$F(\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \ldots, \boldsymbol{\psi}_{n-1}, c) = T_h[\boldsymbol{\psi}^h] + \frac{1}{2}\mu h \sum_{j=1}^{n}\left(\left\|\delta\boldsymbol{\psi}_{j-\frac{1}{2}}\right\|^2 - c\right)^2. \tag{35}$$

We now solve

$$\min_{\mathcal{A}_h} F(\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \ldots, \boldsymbol{\psi}_{n-1}, c),$$

with

$$\mathcal{A}_h = \Big\{ (\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \ldots, \boldsymbol{\psi}_{n-1}, c) \in \mathbb{R}^{2\times(n-1)} \times (0,\infty) :$$
$$\alpha_0 - 2g\hat{z}(\boldsymbol{\psi}_{j-\frac{1}{2}}) > 0,\ 1 \le j \le n,\ \boldsymbol{\psi}_j \in D,\ 1 \le j \le n-1 \Big\}.$$

Approximate solutions of this problem are computed using a *gradient flow* iteration (cf. [15]) on the $\boldsymbol{\psi}_j$'s variables, together with a simple steepest descent iteration on c. To implement these we need the partial derivatives of F. Since for a given j, $1 \le j \le n-1$, only the j–th and $(j+1)$–th terms in (34) and (35) depend on $\boldsymbol{\psi}_j$, the partial derivatives with respect to the $\boldsymbol{\psi}_j$'s are given by:

$$\frac{\partial F}{\partial \boldsymbol{\psi}_j} = -\left[f_{\mathbf{z}}^{j+\frac{1}{2}} - f_{\mathbf{z}}^{j-\frac{1}{2}}\right] + \frac{h}{2}\left[f_{\mathbf{y}}^{j-\frac{1}{2}} + f_{\mathbf{y}}^{j+\frac{1}{2}}\right] \tag{36}$$

$$-2\mu\left[\left(\left\|\delta\boldsymbol{\psi}_{j+\frac{1}{2}}\right\|^2 - c\right)\delta\boldsymbol{\psi}_{j+\frac{1}{2}} - \left(\left\|\delta\boldsymbol{\psi}_{j-\frac{1}{2}}\right\|^2 - c\right)\delta\boldsymbol{\psi}_{j-\frac{1}{2}}\right], \tag{37}$$

for $1 \leq j \leq n-1$, where $f_{\mathbf{z}}^{j-\frac{1}{2}} = f_{\mathbf{z}}\left(\boldsymbol{\psi}_{j-\frac{1}{2}}, \delta\boldsymbol{\psi}_{j-\frac{1}{2}}\right)$, etc., and $f_{\mathbf{z}}, f_{\mathbf{y}}$ are given by (7) and (8) respectively. Also

$$\frac{\partial F}{\partial c} = -\mu h \sum_{j=1}^{n} \left(\left\| \delta\boldsymbol{\psi}_{j-\frac{1}{2}} \right\|^2 - c \right). \tag{38}$$

In our calculations we kept the penalization parameter fixed but it could be updated by some of the usual penalization techniques (cf. [6]).

The shooting method: the corrector step

We let $\mathbf{G}(\mathbf{y}, \mathbf{z})$ represent the right hand side function in (14a). For any given vector $\mathbf{v}$ we let $\boldsymbol{\psi}(\cdot; \mathbf{v})$ be the unique solution of the initial value problem

$$\boldsymbol{\psi}''(\tau) = \mathbf{G}(\boldsymbol{\psi}(\tau), \boldsymbol{\psi}'(\tau)), \quad \tau > 0 \tag{39a}$$

$$\boldsymbol{\psi}(0) = \boldsymbol{\alpha}, \quad \boldsymbol{\psi}'(0) = \mathbf{v}. \tag{39b}$$

Assuming that $\boldsymbol{\psi}(\cdot; \mathbf{v})$ exists over $[0, 1]$, we seek a vector $\mathbf{v}^*$ such that $\boldsymbol{\psi}(1; \mathbf{v}^*) = \boldsymbol{\beta}$. If we let

$$\mathbf{g}(\mathbf{v}) = \boldsymbol{\psi}(1; \mathbf{v}) - \boldsymbol{\beta},$$

then we are looking for a solution of the 2×2 nonlinear system of equations $\mathbf{g}(\mathbf{v}) = \mathbf{0}$. We compute approximate solutions of this system using Newton's method. If we Let

$$\mathbf{W}(\tau; \mathbf{v}) = \frac{\partial}{\partial \mathbf{v}} \boldsymbol{\psi}(\tau; \mathbf{v}),$$

then we get, after differentiating in (39) with respect to $\mathbf{v}$ and omitting the dependence on $\mathbf{v}$, that

$$\mathbf{W}''(\tau) = \mathbf{G}_{\mathbf{y}}(\boldsymbol{\psi}(\tau), \boldsymbol{\psi}'(\tau))\mathbf{W}(\tau) + \mathbf{G}_{\mathbf{z}}(\boldsymbol{\psi}(\tau), \boldsymbol{\psi}'(\tau))\mathbf{W}'(\tau), \quad \tau > 0 \tag{40a}$$

$$\mathbf{W}(0) = \mathbf{O}, \quad \mathbf{W}'(0) = \mathbf{I}, \tag{40b}$$

where $\mathbf{O}$ and $\mathbf{I}$ are the zero and identity 2×2 matrices respectively. Newton's method for the solution of $\mathbf{g}(\mathbf{v}) = \mathbf{0}$ is now given by the iteration:

$$\mathbf{v}_{j+1} = \mathbf{v}_j - \mathbf{W}(1; \mathbf{v}_j)^{-1}\mathbf{g}(\mathbf{v}_j), \quad j = 0, 1, \ldots,$$

with $\mathbf{v}_0$ given by the approximate derivative at $\tau = 0$ of the computed approximate minimizer in the predictor step.

For the next set of numerical examples we used the following values for some of the numerical parameters:

$$g = 9.8, \quad v_0 = 0.1, \quad n = 100, \quad \mu = 50,$$

the units of g and v_0 been in the MKS system. We should mention that the derivatives of the functions f and $\mathbf{G}$ appearing in (37) and (40a) where computed symbolically using MATLAB$^{\text{TM}}$ as well as the implementation of the numerical schemes described in this section.

Example 5.1. We consider first the circular hyperboloid of one sheet of Subsection 4.1.3 with $a = 1$ and $c = 2$. The initial and final points over the surface are given respectively by $\mathbf{\Psi}(0, 1.75)$ and $\mathbf{\Psi}(1.5, 1.75)$. Note that both points are at the same height. We show in Figure 2 the computed brachistochrone (in red) and the corresponding geodesic curve (with $g = 0$) (in magenta). (The initial and final points are shown in blue and green respectively.) In the following table we show the value of the time integral (3) and the length of each of the resulting curves (over the surface):

Curve	value of time integral	length of curve
brachistochrone	1.6194	5.159
geodesic	43.7404	4.374

As expected the geodesic is the shortest curve but it is travelled at a constant speed of v_0 resulting in the large value of the time integral. On the other hand, the brachistochrone is longer but the speed along the curve varies according to (2) resulting in the faster time to cover the curve.

In Figure 3 we show the computed minimizing curve $\boldsymbol{\psi}(\tau) = (\hat{u}(\tau), \hat{v}(\tau))$, $\tau \in [0, 1]$ on the uv plane. Figure 4 shows the norm of $\boldsymbol{\psi}'(\tau)$ for $\tau \in [0, 1]$ consistent with the constraint that $\boldsymbol{\psi}'(\tau) \cdot \boldsymbol{\psi}'(\tau) = \text{const}$.

Example 5.2. For our next example we consider a hyperbolic paraboloid given by the parametrization

$$\mathbf{\Psi}(u, v) = (u, v, u^2 - v^2)^T, \quad (u, v) \in \mathbb{R}^2.$$

The initial and final points over the surface are given respectively by $\mathbf{\Psi}(1, -1)$ and $\mathbf{\Psi}(-1, -1)$ again both points at the same height. We show in Figure 5 the computed brachistochrone (in red) and the corresponding geodesic curve (with $g = 0$) (in magenta). In the following table we show the value of the time integral (3) and the length of each of the resulting curves (over the surface):

Curve	value of time integral	length of curve
brachistochrone	1.1311	2.6425
geodesic	23.2221	2.3222

In Figure 6 we show the computed minimizing curve $\boldsymbol{\psi}(\tau) = (\hat{u}(\tau), \hat{v}(\tau))$, $\tau \in [0, 1]$ on the uv plane, while Figure 7 shows the norm of $\boldsymbol{\psi}'(\tau)$ for $\tau \in [0, 1]$, again consistent with the constraint that $\boldsymbol{\psi}'(\tau) \cdot \boldsymbol{\psi}'(\tau) = \text{const}$.

6 Final comments

The parameter g in the BVP (14) could be used as a continuation parameter to prove existence of solutions for this problem without the "local" condition of closeness for the two points in the surface. The most obvious choice would be to continue from $g = 0$ which reduces our problem to that for geodesics. However, even for the case $g = 0$, there is no "global" result on the existence of solutions for the resulting boundary value problem.

Instead of working with parametrized surfaces, one could work with implicit surfaces defined by a scalar equation of x, y, z. In this case the surface S is given by

$$S = \{(x, y, z) \,:\, H(x, y, z) = 0\}\,, \tag{41}$$

with a smooth H. The time integral is now given by:

$$T[\boldsymbol{\sigma}] = \int_{\tau_1}^{\tau_2} \frac{\|\boldsymbol{\sigma}'(\tau)\| \,\mathrm{d}\tau}{\sqrt{\alpha_0 - 2g\hat{z}(\tau)}}, \tag{42}$$

where $\boldsymbol{\sigma}(\tau) = (\hat{x}(\tau), \hat{y}(\tau), \hat{z}(\tau))$, for $\tau \in [\tau_1, \tau_2]$ belongs to the (admissible) set:

$$\mathcal{B} = \big\{\boldsymbol{\sigma} \in C^1[\tau_1, \tau_2] \,:\, \boldsymbol{\sigma}'(\tau) \neq \mathbf{0},\ \alpha_0 - 2g\hat{z}(\tau) > 0,\ \text{and } H(\boldsymbol{\sigma}(\tau)) = 0\ \forall \tau \in [\tau_1, \tau_2], \\ \boldsymbol{\sigma}(\tau_1) = \mathbf{a},\ \boldsymbol{\sigma}(\tau_2) = \mathbf{b}\big\}, \tag{43}$$

with $\mathbf{a}, \mathbf{b} \in S$. The Euler–Lagrange equations for this problem are given by

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\tau}\left[\frac{1}{\sqrt{\alpha_0 - 2g\hat{z}(\tau)}} \frac{\boldsymbol{\sigma}'(\tau)}{\|\boldsymbol{\sigma}'(\tau)\|}\right] &= \frac{g}{(\alpha_0 - 2g\hat{z}(\tau))^{\frac{3}{2}}}\,\mathbf{e}_3 - \lambda(\tau)\vec{\nabla} H(\boldsymbol{\sigma}(\tau)), \\ H(\boldsymbol{\sigma}(\tau)) &= 0, \quad \boldsymbol{\sigma}(\tau_1) = \mathbf{a}, \quad \boldsymbol{\sigma}(\tau_2) = \mathbf{b}, \end{aligned}$$

where $\mathbf{e}_3 = (0, 0, 1)^T$ and $\lambda \in C[\tau_1, \tau_2]$ is the Lagrange multiplier corresponding to the constraint of belonging to the surface S. Although the differential equation above appears to be simpler than (10), we are looking for solutions over the surface S together with the determination of the Lagrange multiplier λ. We refer to [5] for a treatment of the brachistochrone problem over an implicit, including friction. Also we refer to [9], [10] and the references there in, for further details on existence of solutions and numerical methods for initial value problems over surfaces.

References

[1] N. Ashby, W. E. Brittin, W. F. Love, and W. Wyss. Brachistochrone with Coulomb Friction. *American Journal of Physics*, 43(10):902–906, 1975.

[2] G. A. Bliss. An existence theorem for a differential equation of the second order, with an application to the calculus of variations. *Trans. Amer. Math. Soc.*, 5, 113–125, 1904.

[3] G. A. Bliss. Calculus of Variations. *The Mathematical Association of America*, The Open Court Publishing Co., Chicago, Illinois, 1925.

[4] O. Bolza. Lectures on the Calculus of Variations. Chelsea Publishing Company, 1904.

[5] V. Čović and M. Vesković. Brachistochrone on a surface with Coulomb friction. *International Journal of Non-Linear Mechanics*, 43, 437–450, 2008.

[6] A. V. Fiacco and G. P. McCormick. Nonlinear Programming: Sequential Unconstrained Minimization Techniques. John Wiley & Sons, New York, N.Y., 1968. Reprinted by SIAM Publications, 1990.

[7] I. M. Gelfand and S. V. Fomin. Calculus of Variations. Prentice–Hall, Inc., Englewood Cliffs, N.J., 1963. Reprinted by Dover Publications, Inc., Mineola, N.Y., 2000.

[8] J. Gemmer, M. Nolan, and R. Umble. Generalizations of the Brachistochrone Problem. *Pi Mu Epsilon Journal*, 13(4), 207–218, 2011.

[9] Hairer, E., *Geometric Integration of Ordinary Differential Equations on Manifolds*, BIT Numerical Mathematics, 41, pp. 996–1007, 2001.

[10] Hairer, E., Solving Differential Equations on Manifolds, Université de Genève, Lecture Notes, June 2011.

[11] L. Hawn and T. Kiser. Exploring the Brachistochrone Problem. *The American Mathematical Monthly*, 102(4):328–336, 1995.

[12] H. B. Keller. Numerical Methods for Two Point Boundary Value Problems. Dover Publications, Inc., New York, 1992.

[13] S. C. Lipp. Brachistochrone with Coulomb Friction. *SIAM J. Control Optim.*, 35(2):562–584, 1997.

[14] P. V. Negrón–Marrero. On the specification of initial conditions in the Classical Brachistochrone Problem: The Compatibility Condition. Technical Report, Univ. of Puerto Rico at Humacao, (http://mate.uprh.edu/~urmaa/reports), 2005.

[15] Neuberger, J. W., Sobolev Gradients and Differential Equations, *Lecture Notes in Math.* 1670, Springer, Berlin, 1997.

[16] A. S. Parnovsky. Some Generalizations of Brachistochrone Problem. *Acta Physica Polonica*, A93 Supplement, S-55, 1998.

[17] E. Picard. Traite d'analyse. Vol. 3, Edition 2, 1908.

[18] G. J. Tee. Isochrones and Brachistochrones. *Neural, Parallel and Scientific Computations, Dynamic Publishers Inc.*, 7:311-342, 1999.

[19] N. C. Wright, M. P. Hennessey, and C. Shakiban. Brachistochrone on a 2D Curved Surface Using Optimal Control and Spline-Based DEM Map Data. *ASME 2012 5th Annual Dynamic Systems and Control Conference joint with the JSME 2012 11th Motion and Vibration Conference*, pp. 703–709, 2012.

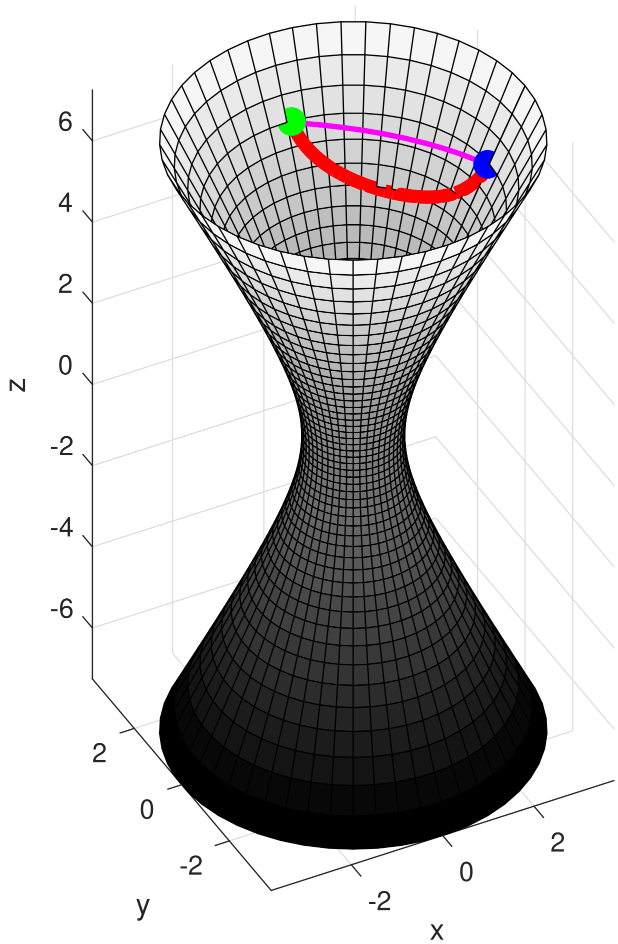

Figure 2: Computed brachistochrone (in red) and the corresponding geodesic curve (in magenta) over a one sheet circular hyperboloid.

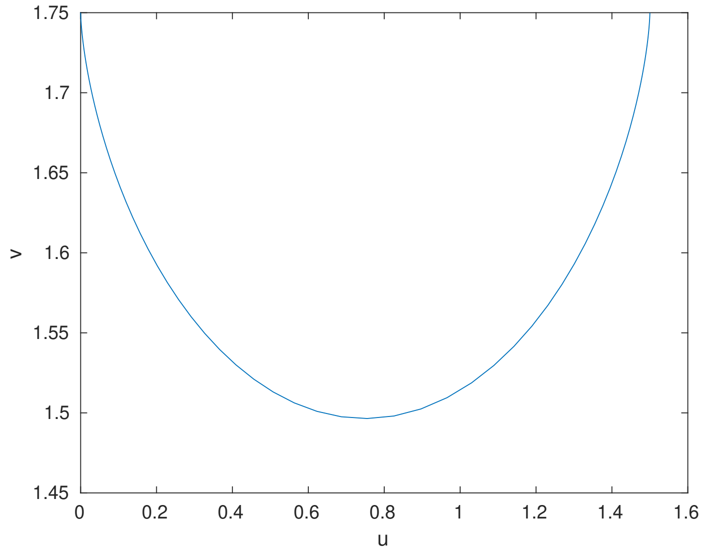

Figure 3: Computed parametrized curve $\boldsymbol{\psi}$ on the uv plane for the one sheet circular hyperboloid.

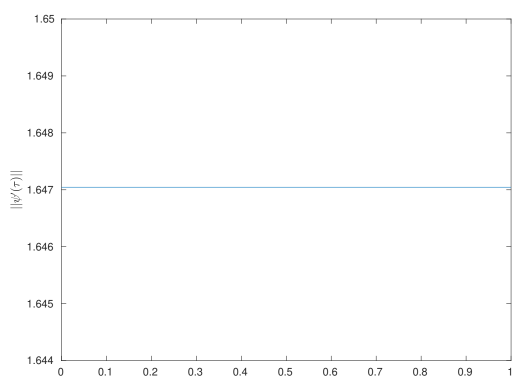

Figure 4: Norm of the computed $\boldsymbol{\psi}'(\tau)$ for the one sheet circular hyperboloid.

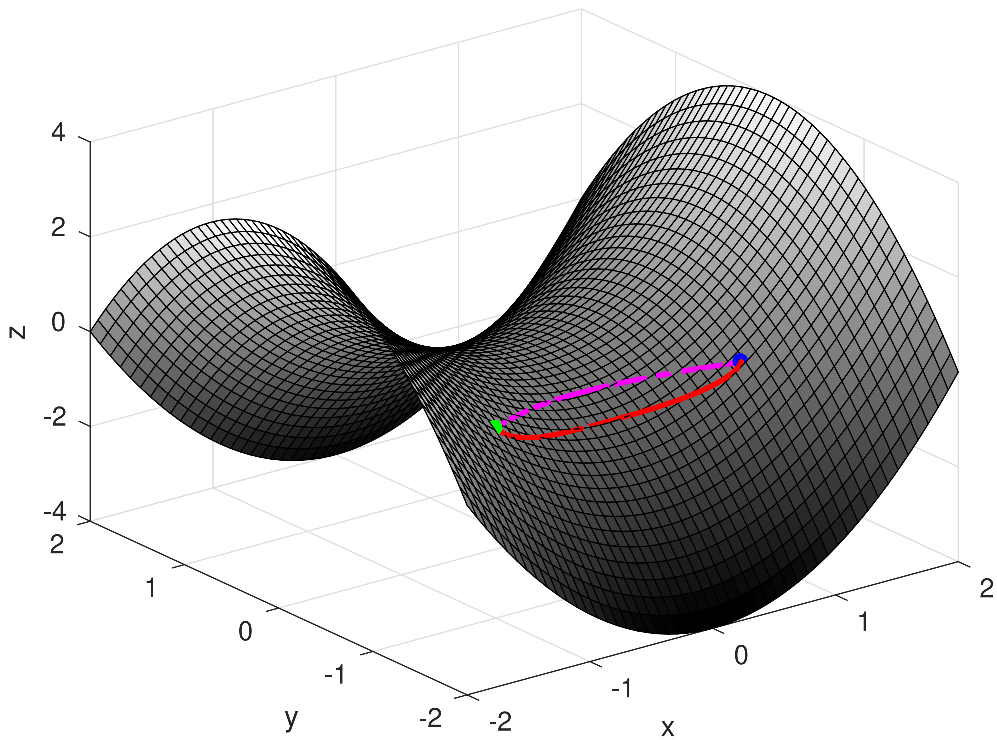

Figure 5: Computed brachistochrone (in red) and the corresponding geodesic curve (in magenta) over a hyperbolic paraboloid.

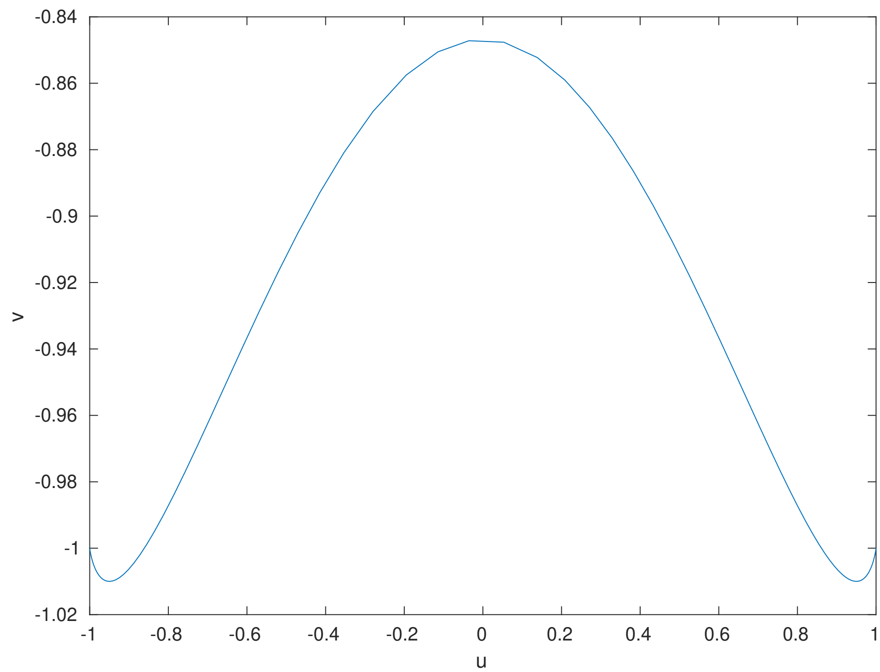

Figure 6: Computed parametrized curve $\boldsymbol{\psi}$ on the uv plane for the hyperbolic paraboloid.

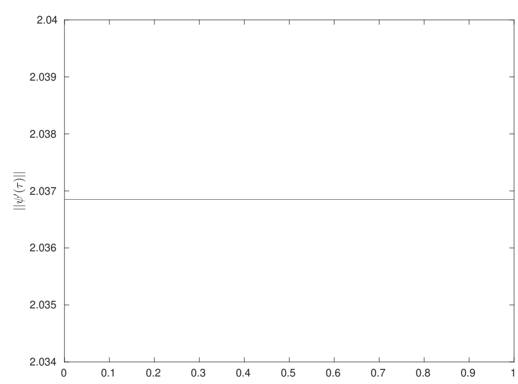

Figure 7: Norm of the computed $\boldsymbol{\psi}'(\tau)$ for the hyperbolic paraboloid.