

Periodic Geodesics Generator

Anders Linnér

CONTENTS

- 1. Introduction
- 2. Specifics
- 3. Examples
- References

A method to generate periodic geodesics in arbitrary level surfaces is presented. The underlying algorithm resolves several technical complications posed by the constraints to stay in the surface and retain periodicity. The method exploits the “inverse” of the parallel transport equation and its “derivative.” This approach avoids most of the complications due to the intricate form of the geodesic curvature. The process flows any periodic curve in the surface along the negative gradient trajectory of the total squared geodesic curvature. The mathematical framework is that of an infinite-dimensional Riemannian manifold representing periodic curves of arbitrary length. The method is illustrated by an example in a sphere-like surface that is neither an ellipsoid nor a surface of revolution.

1. INTRODUCTION

1.1 Geodesics

The discovery of non-Euclidean geometries forced a reexamination of fundamental concepts such as straight lines. Although the length-minimizing property of a straight segment is important, for the purposes of this article a definition rooted in mechanics is more useful. Recall that a regular level-surface M in \mathbb{R}^3 has a well-defined normal direction at each point. A parameterized curve in the surface is considered “straight” whenever its acceleration is parallel to the normal at each point. In mathematics these curves are known as geodesics. A simple non-Euclidean example is given by the great circles in the unit sphere. Note that periodic geodesics, such as the great circles, have finite length $L > 0$.

1.2 Existence

The question regarding the existence of periodic geodesics has been the subject of much research. Possibly the earliest example is Jacobi’s [Jacobi 39] description of geodesics in the ellipsoid. These efforts have recently culminated with the works of Franks [Franks 92], Bangert [Bangert 93], and Hingston [Hingston 93]. Combined, they establish that there are infinitely many geometrically distinct periodic geodesics on any surface

2000 AMS Subject Classification: Primary 53C44, 58E10;
Secondary 53C22

Keywords: Curve straightening, geodesic, gradient trajectories,
steepest descent, periodic

that topologically is a sphere. Moreover, if $n(L)$ denotes the number of periodic geodesics of length no more than L on such a given surface, then $n(L)$ satisfies $\liminf n(L)/\log(L) > 0$ as L tends to infinity. It has proved to be a significant numerical challenge to produce even one example of a sphere-like surface that is neither a surface of revolution nor an ellipsoid. The purpose of the article is to present what the author believes are the first such “generic” examples and describe how they are generated. The main difficulties are posed by the periodic boundary conditions and the point-wise constraint to “stay in M .”

1.3 Convergence

The underlying method, which consists of an algorithm and a tested implementation, deals with arbitrary M , sphere-like or not. Each example generated by this method is the limit of a negative gradient trajectory of the “elastic energy.” Each periodic geodesic is a global minimum, but as $M = \mathbb{R}^2$ illustrates, there need not be any critical points and hence each trajectory diverges in this case. With this in mind, it is perhaps surprising that if the demand on periodicity is removed, then each trajectory converges [Linnér 03b] in any M . For fixed $\nu > 0$, Langer and Singer [Langer and Singer 87] established convergence when the term νL is added to the elastic energy. It is a significant open problem to determine if when $\nu = 0$ there exists a compact M with some divergent trajectory inside the space of periodic curves.

1.4 Curve-Shortening

The idea to systematically deform a given periodic curve in its free homotopy class as a way to produce a periodic geodesic goes back at least as far as Hadamard [Hadamard 98]. In manifolds with nontrivial fundamental group, it is natural to attempt to shrink the curve and minimize the length in its nontrivial homotopy class. The modern counterpart is the, much analyzed, curve-shortening flow. A noteworthy numerical simulation of this flow is based on the so-called level set method (Osher and Sethian). For $M = \mathbb{R}^2$ the curve-shortening flow is illustrated by an interactive applet at <http://math.berkeley.edu/~sethian/>. When the curve-shortening flow is extended to curves in sphere-like surfaces most initial curves are expected to shrink to a point. To combat this, Gage [Gage 90] introduced the area-preserving curve-shortening flow. As far as I know, there is no numerical simulation of this flow available at this time. Also observe that the choice of initial curve is nontrivial in this case. The enclosed area must be

equal to its counterpart for the limiting periodic geodesic. Curves with self-intersections are excluded in the curve-shortening process, and this is in stark contrast to the method presented here. Long periodic geodesics are of course expected to have many self-intersections.

1.5 Curve-Straightening

A completely different approach creates an infinite setting for Morse theory, and in this case the appropriate quantity is the integral of the square of the speed rather than the length by itself; see [Klingenberg 78]. The “energy,” considered by Klingenberg, still shrinks most initial curves to points in sphere-like surfaces. This behavior suggests replacing the speed with the curvature. The algebraic form of this “elastic” energy is of course much more complicated, especially on bent surfaces when the curvature is really the geodesic curvature. The flow along negative gradient trajectories of the elastic energy is known as curve-straightening. The rest of this article will show how to overcome several technical difficulties and present an implementation of curve straightening that produces periodic geodesics in sphere-like surfaces.

1.6 Parallel Transport

In the Euclidean plane it is often helpful to move a vector from one point to another without rotating the vector. This kind of move can be constrained so that the base of the vector is always on a given curve. The angle between the vector and the tangent of the curve is typically not constant and its derivative is a measure of the curvature of the given curve. This operation of “parallel transport along a curve” generalizes to tangent vectors of M and curves in M . Our intuition about this process is very limited as recognized already when M is the unit sphere. Specifically, the demand is that the derivative of the transported vector is parallel to the normal at each point, and the corresponding equation is referred to as the *parallel transport equation*. It follows that the tangent vector to a geodesic is indeed parallel transported along the curve. For general curves there is again an angle, the indicatrix, between the transported vector and the tangent to the curve. The derivative of the indicatrix is a measure of the curvature of the curve. The *geodesic curvature* is the derivative of the indicatrix divided by the length of the curve.

1.7 Curves via Their Indicatrix

Each curve in M is uniquely represented by its indicatrix as soon as the length and the initial point and direction are provided. A curve in this article is therefore given as

triple (θ, L, pi) , with θ in the Sobolev space $W_1^2[0, 1]$, the length L a positive real number, and pi an element of the unit tangent bundle SM . The fact that the elastic energy involves the square of the geodesic curvature dictates the choice of the Sobolev space. Observe that the curves are parameterized proportional to arc-length, and the length L is also the speed of the curve. With this indicatrix representation it is much more complicated to determine if a curve is periodic.

1.8 Inverted Parallel Transport

The core of the method in the present article is based on the inverse to the parallel transport equation. Roughly speaking, this inverse uses the geodesic curvature together with an initial point and direction and “recovers” the curve. The idea is to start with a periodic curve and reduce the integral of its total squared geodesic curvature by following the negative gradient without leaving the space of periodic curves. Since the inverse equation generates the final endpoint of the curve, it has the potential to facilitate the projection onto the space of periodic curves. The goal is to track the effect of a variation in the initial point as it manifests itself at the final point. An important, albeit disappointing, numerical insight is that the inverse equation is not capable of producing sufficiently accurate finite difference approximations of the variations at the final point to achieve this goal. To address this deficiency, it is necessary to bring in the “derivative” of the inverse to the parallel transport equation.

1.9 Derivative of the Inverse Equation and Gradients

The solutions of the inverse of the parallel transport equation depend smoothly on the initial conditions. It follows that the final point and its tangent have derivatives that depend on variations in the initial data. The “derivative” of the inverse equation is a second differential equation whose solutions produce, with desired numerical accuracy, the requested derivatives at the final point. Since the tangent space to the space of periodic curves is infinite-dimensional, it is nontrivial to convert these derivatives to a gradient. The approach used here is to engage a Hilbert basis in the space of indicatrices. To be precise, the space of all curves in M is parameterized by the tangent angle together with the length and an element of the unit tangent bundle. The latter provide the initial point and direction of the curve. The curves are parameterized so that the speed is equal to the length. The tangent angle is ultimately approximated by a finite degree polynomial by truncating the Hilbert ba-

sis expansion. There is of course freedom in the choice of basis, but the Gram-Schmidt sequence generated from x, x^2, \dots is adequate. Observe that the indicatrix of a periodic curve is not expected to be periodic.

1.10 Computational Experience

The software implementation of the outlined approach calls for a hybrid of both symbolic as well as numeric calculations. The ultimate goal is to draw the periodic geodesic, and therefore strong graphics support is an advantage. Mathematica is one system that offers these capabilities. The purpose of the present implementation is to demonstrate that the proposed method indeed is capable of producing examples of periodic geodesics. Many different numerical algorithms are asked to work in unison. The Hilbert basis expansions are of course truncated, and this cut-off, together with other parameters, must be carefully selected. It is premature to claim that the settings adapted so far are optimal. They are, however, adequate and, one might add, surprisingly demanding. Unfortunately, the computational speed inside Mathematica is far from satisfying and a future project is a complete conversion of the implementation to a fully compiled counterpart. In Mathematica’s defense, the substantial amount of experimentation and tweaking leading up to the present state of the code would have been far more cumbersome in a compiled environment.

1.11 Organization

The next section introduces the necessary notation to make the previous outline mathematically precise. Special attention is given to details that are easy to overlook but in fact are critical. The final section presents one specific example of a periodic geodesic on a sphere-like surface that is neither an ellipsoid nor a surface of revolution. The initial ideas for the method specialized and finalized here are found in Linnér [Linnér 91].

2. SPECIFICS

2.1 Notation

As a starting point let $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ be a given smooth function. The surface is given by $M = \{p \in \mathbb{R}^3 \mid f(p) = 0\}$, where it is assumed that $p \in M$ implies $\nabla f(p) \neq 0$. The unit normal to the surface is given by $n(p) = \nabla f(p) / |\nabla f(p)|$. The initial curve is some $\gamma : [0, 1] \rightarrow M$, where $|\gamma'(s)| = L > 0$ for all $s \in [0, 1]$ and L is the length of γ . The implementation facilitates this by automatically changing the parameter of any initial $\gamma(t)$, such that $\gamma'(t) \neq 0$, to this required form. Let

$\gamma_0 = \gamma(0)$ and $i_0 = \gamma'(0)/L$. The main example of this article is given by

$$f(x, y, z) = 2x^2 + 3y^2 + 5z^2 + x^3z + yz^3 - 1. \quad (2-1)$$

The coordinate plane $z = 0$ intersects this surface along the ellipse $2x^2 + 3y^2 = 1$. The initial curve $\gamma(t) = (\cos(2\pi t)/\sqrt{2}, \sin(2\pi t)/\sqrt{3}, 0)$ does not have constant speed, and the implementation calculates the new parameter s . The initial length L is expressible in terms of the complete elliptic integral E (Mathematica yields $4E(-1/2)/\sqrt{3}$). More generally, one can use a coordinate chart and “lift” a periodic curve from the plane to M , and then let the implementation reparameterize the lifted curve.

2.2 Geodesics

There are geodesics emanating in each direction from the initial point $(1/\sqrt{2}, 0, 0)$. The geodesic in the initial direction $(0, 2\pi/\sqrt{3}, 0)$ is different from γ . As the “geodesic ray” corresponding to these initial conditions is extended, it stays in a “strip” in the surface $M = f^{-1}(0)$. The unit speed geodesic “ray” extended 100 units generates the strip seen in Figure 1.

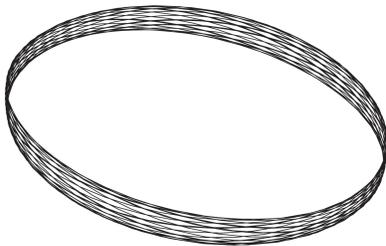


FIGURE 1. First geodesic ray in $2x^2 + 3y^2 + 5z^2 + x^3z + yz^3 = 1$.

As an example of a very different behavior, consider the direction given by rotating the previous direction “down” $\pi/4$ radians in the tangent plane. This time

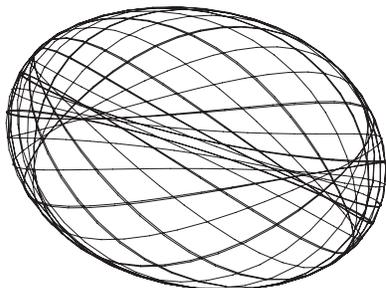


FIGURE 2. Second geodesic ray in $2x^2 + 3y^2 + 5z^2 + x^3z + yz^3 = 1$.

the unit speed geodesic extended 100 units is depicted in Figure 2.

It is doubtful that either of the two initial tangent directions produce a periodic geodesic; in fact, there is no guarantee any direction has this capability at this particular initial point.

2.3 Parallel Transport and Its Inverse

To proceed, parallel transport i_0 and get i defined on $[0, 1]$ such that $i(s) \in T_{\gamma(s)}M$ with $|i(s)| = 1$ and $i(0) = i_0$. Here, $T_{\gamma(s)}M$ is the tangent space to M at the point $\gamma(s)$. The parallel transport demand is expressed by $i'(s) = (i'(s) \cdot n(\gamma(s)))n(\gamma(s))$. Let $\theta : [0, 1] \rightarrow \mathbb{R}$ be the smooth angle between i and γ' with $\theta(0) = 0$. The parallel transport equation completely determines θ . Since $i(s) \cdot n(\gamma(s)) = 0$, it follows that $i'(s) \cdot n(\gamma(s)) = -i(s)[n(\gamma(s))]'$. It is clear that $|\nabla f(\gamma(s))|^2 = \nabla f(\gamma(s)) \cdot \nabla f(\gamma(s))$ implies that $|\nabla f(\gamma(s))| |\nabla f(\gamma(s))|' = \nabla f(\gamma(s)) \cdot \nabla f(\gamma(s))'$; hence $|\nabla f(\gamma(s))|' = n(\gamma(s)) \cdot Hf(\gamma(s))\gamma'(s)$ in terms of the Hessian Hf of second derivatives. Moreover, $|\nabla f(\gamma(s))| n(\gamma(s)) = \nabla f(\gamma(s))$ implies that

$$|\nabla f(\gamma(s))|' n(\gamma(s)) + |\nabla f(\gamma(s))| [n(\gamma(s))]' = Hf(\gamma(s))\gamma'(s), \quad (2-2)$$

and finally

$$i(s) \cdot [n(\gamma(s))]' = i(s) \cdot Hf(\gamma(s))\gamma'(s) / |\nabla f(\gamma(s))|. \quad (2-3)$$

The inverse parallel transport equation is the six-dimensional initial value problem

$$\begin{aligned} \gamma'(s) &= L(\cos \theta(s)i(s) + \sin \theta(s)n(\gamma(s)) \times i(s)) \\ i'(s) &= -(i(s) \cdot Hf(\gamma(s))\gamma'(s))n(\gamma(s)) / |\nabla f(\gamma(s))| \end{aligned} \quad (2-4)$$

with $\gamma(0) = \gamma_0$ and $i(0) = i_0$. Observe how θ , together with the length L and the initial data, determines both the parallel transport as well as the curve.

2.4 Unit Tangent Bundle

The proper setting for the parallel transport equation is the unit tangent bundle given by

$$SM = \{pi = (p, i) \in \mathbb{R}^3 \times \mathbb{R}^3 \mid f(p) = 0, i \in T_pM, |i| = 1\}. \quad (2-5)$$

This space is in fact a closed three-dimensional submanifold of $\mathbb{R}^3 \times \mathbb{R}^3$ expressed as $SM = g^{-1}(0, 0, 0)$ with $g :$

$\mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}^3$ given by $g(p, v) = (f(p), \nabla f(p) \cdot v, v \cdot v - 1)$. The derivative of g is represented by the matrix

$$dg(p, v) = \begin{bmatrix} \nabla f(p) & 0 \\ Hf(p)v & \nabla f(p) \\ 0 & 2v \end{bmatrix}. \quad (2-6)$$

The kernels of these maps restricted to SM provide the tangent bundle TSM , which has the following explicit basis:

$$\begin{aligned} &(i, (Hf(p)i \cdot i / |\nabla f(p)|)n(p)), \\ &(j, -(Hf(p)i \cdot j / |\nabla f(p)|)n(p)), (0, j), \end{aligned} \quad (2-7)$$

where $j = n(p) \times i$. In what follows, let $\{T_q\}_{q=1}^3$ be a normalized orthogonal basis for TSM .

2.5 Hilbert Space and Functional

In the present set-up the geodesic curvature is given by $\kappa = \theta'/L$. Geodesics are exactly the curves such that $\theta \equiv 0$. The goal is to straighten the initial curve as efficiently as possible without destroying periodicity. To avoid generating curves that shrink to a point with $L \rightarrow 0^+$, use the total squared curvature functional $\int \kappa^2$, also known as the elastic energy. In the Euclidean plane, the total squared curvature of a circle of radius r is given by $2\pi/r$, and the functional is unbounded as $r \rightarrow 0^+$. This behavior persists in general surfaces M . In terms of the indicatrix, the proper setting is given by the Sobolev space of absolutely continuous θ with θ' square-integrable. Denote this space by H_0 when it is assumed that $\theta(0) = 0$. Let \mathbb{R}^+ be all positive real numbers. The precise definition of the functional is given by $J : H_0 \times \mathbb{R}^+ \times SM \rightarrow \mathbb{R}$ such that

$$J(\theta, L, pi) = \frac{1}{2L} \int_0^1 (\theta'(s))^2 ds. \quad (2-8)$$

The half is added to simplify the derivative and the gradient formulas.

2.6 Periodic Curves

The space of periodic curves is represented by the closed infinite-dimensional submanifold

$$\Omega = \{(\theta, L, pi) \in H_0 \times \mathbb{R}^+ \times SM \mid \Lambda(\theta, L, pi) = pi\}, \quad (2-9)$$

where $\Lambda : H_0 \times \mathbb{R}^+ \times SM \rightarrow SM$ is defined by

$$\begin{aligned} \Lambda(\theta, L, pi) &= (\gamma(1), \gamma'(1)/L) = \\ &(\gamma(1), \cos \theta(1)i(1) + \sin \theta(1)n(\gamma(1)) \times i(1)), \end{aligned} \quad (2-10)$$

where $(\gamma(0), \gamma'(0)/L) = (\gamma(0), i(0)) = (\gamma_0, i_0) = pi$. The technical details verifying the space is a manifold are found in [Langer and Singer 87] in a slightly different and also more general setting. The gradient of J is most likely not tangent to Ω , so it is necessary to take its tangential part. To this end one first needs the derivative

$$D\Lambda(\theta, L, pi) : H_0 \times \mathbb{R} \times T_{pi}SM \rightarrow T_{(\gamma(1), \gamma'(1)/L)}SM. \quad (2-11)$$

2.7 Derivative of Inverse Parallel Transport

The derivative of the inverse parallel transport equation yields $d\gamma(1)$ and $di(1)$. As indicated in the introduction, the use of the inverse parallel transport equation to numerically estimate these derivatives is problematic if not futile. The main issue is the bound on the accuracy imposed by each choice of discrete derivative formula. This fixed bound is not necessarily within the numerical demands imposed by the choice of M and the initial curve. On the other hand, the derivative of the inverse parallel transport equation brings with it this desired numerical flexibility. A hefty price is paid in computational performance as one solves the six-dimensional

$$\begin{aligned} [d\gamma(s)]' &= (v_L/L)\gamma'(s) + v_\theta(s)n(p) \times \gamma'(s) \\ &+ L \cos(\theta(s))di(s) + L \sin(\theta(s))/|\nabla f(p)| \\ &\times [-n(p)Hf(p)d\gamma(s)n(p) \times i \\ &+ Hf(p)d\gamma(s) \times i + \nabla f(p) \times di(s)] \end{aligned}$$

$$\begin{aligned} [di(s)]' &= 2/|\nabla f(p)|(Hf(p)d\gamma(s) \cdot n(p)) \\ &\times (i \cdot Hf(p)\gamma'(s))n(p) \\ &- 1/|\nabla f(p)|(di(s) \cdot Hf(p)\gamma'(s))n(p) \\ &- 1/|\nabla f(p)|[i \cdot (d\gamma(s) \cdot DHf(p) \cdot \gamma'(s) \\ &+ Hf(p)[d\gamma(s)]'n(p) \\ &- 1/|\nabla f(p)|^2 (i \cdot Hf(p)\gamma'(s))Hf(p)d\gamma(s)]. \end{aligned} \quad (2-12)$$

The derivation of this system is in the same spirit as the one yielding the inverse to the transport equation. The tangent vector in $H_0 \times \mathbb{R} \times T_{pi}SM$ is written as (v_θ, v_L, v_{pi}) , and the six-dimensional initial conditions are given by $(d\gamma, di)(0) = \sum_{q=1}^3 v_{pi}(q)T_q$, where $v_{pi}(q)$ is the q th component of the three-dimensional v_{pi} . To save space here $\gamma(s)$ is written as p . There is also a term involving the three-tensor appearing as the derivative DHf of the Hessian Hf . Incidentally, Mathematica accepts tensors verbatim and also, thankfully, terms like $[d\gamma(s)]'$ on the right-hand side of the differential equation.

Putting things together,

$$\begin{aligned}
 D\Lambda(\theta, L, pi)(v_\theta, v_L, v_{pi}) = & \\
 (d\gamma(1), -\sin\theta(1)v_\theta(1)i(1) + \cos\theta(1)di(1) & \\
 + \cos\theta(1)v_\theta(1)n(\gamma(1)) \times i(1)) & \\
 + \sin\theta(1)((Hf(\gamma(1))d\gamma(1) \times i(1) & \\
 - (n(\gamma(1)) \cdot Hf(\gamma(1))d\gamma(1))n(\gamma(1)) & \\
 \times i(1))/|\nabla f(\gamma(1))| & \\
 + n(\gamma(1)) \times di(1)). &
 \end{aligned} \tag{2-13}$$

2.8 Projection

Using the Riemannian structure

$$\begin{aligned}
 \langle (v_\theta, v_L, v_{pi}), (w_\theta, w_L, w_{pi}) \rangle = & \\
 \int_0^1 v'_\theta(s)w'_\theta(s)ds + v_Lw_L + v_{pi} \cdot w_{pi}, & \tag{2-14}
 \end{aligned}$$

define the gradient $\nabla J(\theta, L, pi) \in H_0 \times \mathbb{R} \times T_{pi}SM$ by requiring the Gateaux derivative DJ to satisfy $\langle \nabla J(\theta, L, pi), v \rangle = DJ(\theta, L, pi)v$. Next, seek $\lambda \in T_{pi}SM$ so that $\nabla J(\theta, L, pi) - D\Lambda^T(pi)\lambda \in T_{(\theta, L, pi)}\Omega$. To determine the transpose of $D\Lambda$ it is necessary to have a normalized orthogonal Hilbert basis $\{P_l\}_{l=1}^\infty$ for H_0 . It turns out that the Gram-Schmidt basis generated from $1, x, x^2, \dots$ is the same as the antiderivatives of the Legendre polynomials, which are available in Mathematica. For $m = 1, 2, 3$, write

$$D\Lambda^T(pi)T_m = \left(\sum_{l=1}^\infty a_m^l P_l, b_m, \sum_{q=1}^3 c_m^q T_q \right), \tag{2-15}$$

and observe that

$$\begin{aligned}
 a_m^k = \langle D\Lambda^T(pi)T_m, (P_k, 0, 0) \rangle_{(\theta, L, pi)} = & \\
 \langle T_m, D\Lambda(\theta, L, pi)(P_k, 0, 0) \rangle_{pi}, & \tag{2-16}
 \end{aligned}$$

$$\begin{aligned}
 b_m = \langle D\Lambda^T(pi)T_m, (0, 1, 0) \rangle_{(\theta, L, pi)} = & \\
 \langle T_m, D\Lambda(\theta, L, pi)(0, 1, 0) \rangle_{pi}, & \tag{2-17}
 \end{aligned}$$

$$\begin{aligned}
 c_m^q = \langle D\Lambda^T(pi)T_m, (0, 0, T_q) \rangle_{(\theta, L, pi)} = & \\
 \langle T_m, D\Lambda(\theta, L, pi)(0, 0, T_q) \rangle_{pi}. & \tag{2-18}
 \end{aligned}$$

2.9 Preserved Periodicity

The requirement that the projected gradient is tangent to the space of periodic curves is expressed by

$$\nabla J(\theta, L, pi) - D\Lambda^T(\theta, L, pi)\lambda \in T_{(\theta, L, pi)}\Omega. \tag{2-19}$$

Observe that if $(\theta, L, pi) \in \Omega$, then $\Lambda(\theta, L, pi) = pi$. This leads to

$$\begin{aligned}
 D\Lambda(\theta, L, pi)[\nabla J(\theta, L, pi) - D\Lambda^T(pi)\lambda] = & \\
 \pi_3(\nabla J(\theta, L, pi) - D\Lambda^T(pi)\lambda), & \tag{2-20}
 \end{aligned}$$

where $\pi_3 : H_0 \times \mathbb{R} \times TSM \rightarrow TSM$ is given by $\pi_3(v_\theta, v_L, v_{pi}) = v_{pi}$. Since J does not depend on pi , the third component of ∇J vanishes, and the tangency condition simplifies to

$$\begin{aligned}
 (D\Lambda(\theta, L, pi)D\Lambda^T(pi) - \pi_3D\Lambda^T(pi))\lambda = & \\
 D\Lambda(\theta, L, pi)\nabla J(\theta, L, pi). & \tag{2-21}
 \end{aligned}$$

In terms of the Hilbert basis components, the left-hand side involves the crucial matrix $\Pi : T_{pi}SM \rightarrow T_{pi}SM$, given by the entries

$$\Pi_i^j = \sum_{l=1}^\infty a_j^l a_i^l + b_j b_i + \sum_{q=1}^3 c_j^q c_i^q - c_j^i \tag{2-22}$$

in row i and column j when the matrix is expressed with respect to $\{T_q\}_{q=1}^3$. The parameter λ is the solution of the three-dimensional linear system

$$\Pi\lambda = D\Lambda(\theta, L, pi)\nabla J(\theta, L, pi). \tag{2-23}$$

Apart from motivating the choice of Hilbert space, the only place where J enters is at this final step. The gradient depends on the choice of Riemannian structure, and with the current choice

$$\nabla J(\theta, L, pi) = \left(\frac{\theta(s)}{L}, -\frac{1}{2} \int_0^1 (\theta'(s)/L)^2 ds, 0 \right). \tag{2-24}$$

2.10 Approximate Expansions and Steepest Descent

All building blocks of the method are now in place. It is of course necessary to truncate the infinite series. Numerical experimentation indicates that replacing infinity with an integer in the range 5–15 is workable. At the low end of the range, there is the risk of insufficient accuracy. The high end of the range slows each step to taking maybe more than an hour (on a 1.2-GHz PC running Mathematica 5.0). Each step produces a tangent direction $v \in T_p\Omega$ such that the total squared curvature is reduced more rapidly in this direction than any other. The idea of a line search to find the optimal step size is less useful here since lines leave the manifold of periodic curves. More succinctly, since the manifold Ω is not flat, it is typical for the distance from $p + hv$ to Ω to

increase substantially as h grows. The key is therefore to select h so that $p + hv$ is “sufficiently close” to Ω and yet maintain a reduction of the energy at an acceptable rate. The implementation acts particularly carefully here by following a geodesic in SM emanating from $pi \in SM$ in the direction suggested by the third component of the projected negative gradient. Compare this with the traditional Euler step, which takes place in $T\Omega$, which in the third factor is TSM rather than SM . Experimentation is necessary to choose a suitable initial step size h and the subsequent interactive adjustments. As usual, it is useful to increase the step size to accelerate the speed of convergence as the limit is approached. There is considerable room for improvement in this particular area. The author developed an acceleration scheme based on the Adams-Bashford idea to help track divergent trajectories as they approach critical points at infinity. This scheme potentially adapts to the present manifold setting but is left for future versions; see the examples in [Linnér 03a].

3. EXAMPLES

The implementation reproduces accurately the explicit flow on curves in the Euclidean plane, and it finds periodic geodesics in a Torus of revolution. Returning to the case $f(x, y, z) = 2x^2 + 3y^2 + 5z^2 + x^3z + yz^3 - 1$, compare the initial ellipse in the xy -plane and the limit periodic geodesic in Figure 3.

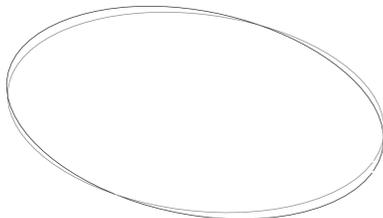


FIGURE 3. Initial ellipse and final periodic geodesic.

Here the darker curve is the periodic geodesic. The plots stop at $s = 0.998$ so that the reader can see the initial point and direction and compare them to the final point and direction. The z coordinate as a function of s is initially zero, and here its limit is given in Figure 4.

One way to look at the method is that it produces an initial point and direction with the property that the corresponding geodesic ray is in fact periodic. In the example the approximate initial point $(0.70925, -0.0010537, -0.0433068)$ and direction

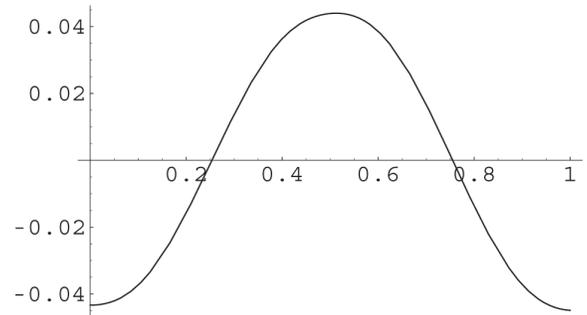


FIGURE 4. The z coordinate of the final periodic geodesic.

$(0.00230281, 0.999997, -0.000309739)$ emerge. The corresponding unit speed geodesic ray extended 100 units is shown in Figure 5.

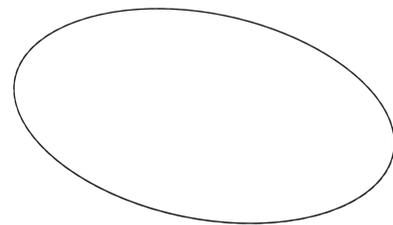


FIGURE 5. Geodesic ray of length 100.

This curve is “thicker” than the “true” limit. This is to be expected since the calculation was aborted somewhat prematurely; therefore, there is a slight inaccuracy in the calculated initial data. The method has been tested to the point where the total squared geodesic curvature of the “limit” is approximately machine precision. In such cases, no “thickening” is visible. Standard ellipsoids have also been considered. When $f(x, y, z) = x^2/a^2 + y^2/b^2 + z^2/c^2 - 1$, and the initial curve has constant $z \neq 0$, then there are ellipsoids where the flow tends to the geodesic with $z = 0$ through curves with constant z . A current project involves an attempt to resolve two contradicting claims found in the literature regarding the number of simple periodic geodesics on such ellipsoids.

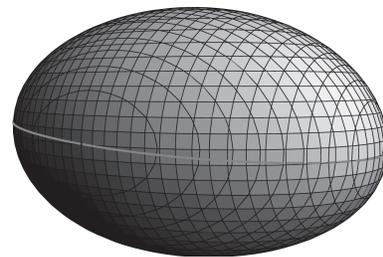


FIGURE 6. Periodic geodesic in $2x^2 + 3y^2 + 5z^2 + x^3z + yz^3 = 1$.

In conclusion, four distinct views of the periodic geodesic are presented in Figures 6–9. The initial curve, which is an ellipse, has approximate length 4.04554. The limit periodic geodesic has approximate length 4.0534. The Mathematica notebook is available at www.math.niu.edu/~alinner.

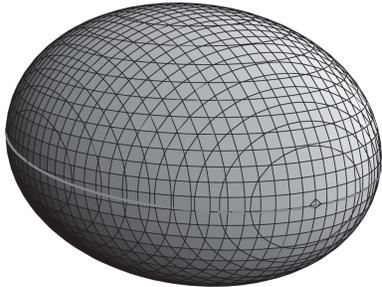


FIGURE 7. Periodic geodesic in $2x^2 + 3y^2 + 5z^2 + x^3z + yz^3 = 1$.

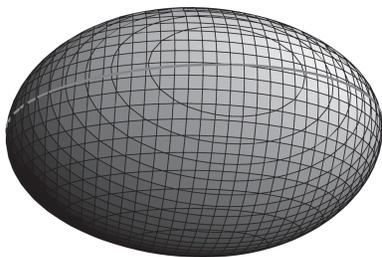


FIGURE 8. Periodic geodesic in $2x^2 + 3y^2 + 5z^2 + x^3z + yz^3 = 1$.

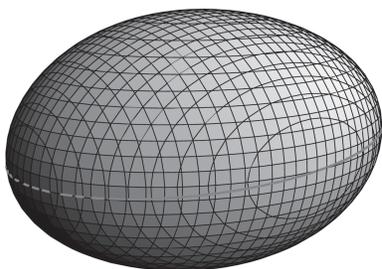


FIGURE 9. Periodic geodesic in $2x^2 + 3y^2 + 5z^2 + x^3z + yz^3 = 1$.

Anders Linnér, Department of Mathematical Sciences, Northern Illinois University, DeKalb, Illinois 60115
(alinner@math.niu.edu)

REFERENCES

- [Bangert 93] V. Bangert. “On the Existence of Closed Geodesics on Two-Spheres.” *Internat. J. Math.* 4:1 (1993), 1–10.
- [Franks 92] J. Franks. “Geodesics on S^2 and Periodic Points of Annulus Homeomorphisms.” *Invent. Math.* 108:2 (1992), 403–418.
- [Gage 90] M. Gage. “Deforming Curves on Convex Surfaces to Simple Closed Geodesics.” *Indiana Univ. Math. J.* 39 (1990), 1037–1059.
- [Hadamard 98] J. Hadamard. “Les surfaces à courbures opposées et leur lignes géodésiques.” *J. Math. Pures Appl.* (5) 4 (1898), 27–73.
- [Hingston 93] N. Hingston. “On the Growth of the Number Closed Geodesics on the Two-Sphere.” *International Research Notices* 9 (1993), 253–262.
- [Jacobi 39] C. G. J. Jacobi. “Note von der geodätischen Linie auf einem Ellipsoid und den verschiedenen Anwendungen einer merkwürdigen analytischen Substitution.” *Crelles J.* 19, 309–313, (1839).
- [Klingenberg 78] W. Klingenberg. *Lectures on Closed Geodesics*, Grundlehren Math. Wiss. 230. Berlin-Heidelberg-New York: Springer, 1978
- [Langer and Singer 87] J. Langer and D. A. Singer. “Curve-Straightening in Riemannian Manifolds.” *Ann. Global Anal. Geom.* 5 (1987), 133–150.
- [Linnér 91] A. Linnér. “Curve-Straightening in Closed Euclidean Submanifolds.” *Commun. Math. Phys.* 138 (1991), 33–49.
- [Linnér 03a] A. Linnér. “Symmetrized Curve-Straightening.” *Diff. Geo. Appl.* 18 (2003), 119–146.
- [Linnér 03b] A. Linnér. “Free Curve Straightening.” Preprint, 2003