

Quadratic Gradient Descent Methods and Transversality Conditions

J. ARROYO and O. J. GARAY and J. MENCÍA

University of the Basque Country UPV/EHU

Department of Mathematics

Aptdo. 644, 48080 Bilbao

SPAIN

josujon.arroyo@ehu.es, oscarj.garay@ehu.es, jj.mencia@ehu.es

Abstract: Different gradient descent methods have been introduced in [1] to study a quite general family of variational problems under affine and isoperimetric constraints. Thus, in [1] gradient descent sequences are created by using linear, quadratic or cubic approximations to the gradient descent trajectories, and methods are numerically implemented in a computational platform (which we call XEL-platform). In this work, performance of the above quadratic gradient descent versions are analyzed under the influence of transversality constraints. We see that, within this context, transversality conditions can be dealt as isoperimetric constraints and, then, the XEL-platform can be used to localize minimizers in the spaces of curves determined by the prescribed constraints. Efficiency of the approach is analyzed by considering two very well known classical problems, the brachistochrone and closed planar elastica. In the first case, the effect of introducing an isoperimetric constraint is also considered and the estimated errors are shown to be numerically insignificant. In the second case, minimizers are well known (circles and eight-figure curves) and we see how the XEL-platform takes very distant curves (from the energy point of view) within the same homotopy class to the minimizer included in that class. It is also capable to detect local minimizers which may appear during the gradient descent from the initial curve towards the minimum.

Key-Words: Gradient descent, energy minimizers, isoperimetric conditions, transversality.

1 Introduction

In [1] we have developed gradient descent based methods of different type (linear, quadratic and cubic) to localize minima of an ample family of functionals defined on certain spaces of curves satisfying both affine and isoperimetric constraints. Numerical implementation of these methods have been integrated into a computational platform which we call XEL-platform. The general algorithm introduced in [1] applies equally well to minimizers with either fixed or free endpoints and when a finite number of isoperimetric constraints is allowed. Moreover, solutions provided by our algorithm are tested against well known explicit solutions of several variational problems have been checked for measurable errors. Some issues concerning the numerical procedure that might affect the final solution, such as the choice of the starting curve and the step tolerance in the isoperimetric constraints, have also been analyzed in [1].

The main purpose of this note is to show that the platform can also be used to analyze variational problems including not only affine and isoperimetric constraints but also a generalization of the former named transversality conditions.

We first describe the basic formalism which serves as base for the numerical treatment. Then, we apply our approach to finding the solutions of two classical variational problems, the brachistochrone and planar elastica, under the influence of affine, isoperimetric and transversality constraints. Treating transversality conditions as isoperimetric constraints allows us to use a quadratic type descent method implemented into the XEL-platform [1] to locate energy minimizers in both above problems which are reported graphically. The XEL-platform informs of local minima along the descent evolution which could be interpreted as a fake global minimizer if the allowed error is not sufficiently small. Whenever it is possible, results obtained here are numerically compared with known analytical solutions.

1.1 Lagrangians, constraints and higher order descent curves

If $H_0(I, \mathbb{R}^m) = L^2(I, \mathbb{R}^m)$ represents the set of square integrable functions from I to \mathbb{R}^m , where I is an interval $[a, b]$. we denote by $H_1(I, \mathbb{R}^m)$ the set of absolutely continuous maps $x : I \rightarrow \mathbb{R}^m$ such

that $\mathbf{x}^{(1)} \in H_0(I, \mathbb{R}^m)$, where $\mathbf{x}^{(1)}$ stands for the first derivative of the function. Finally, let $H_n(I, \mathbb{R}^m)$ denote the set of maps $\mathbf{x} : I \rightarrow \mathbb{R}^m$ such that $\mathbf{x}^{(k)} \in H_1(I, \mathbb{R}^m)$, $k \in \{0, \dots, n - 1\}$, $\mathbf{x}^{(k)}$ being the k -th derivative of \mathbf{x} . Then, $H_n(I, \mathbb{R}^m)$ is a Hilbert space with the following family of inner products [1]

$$\langle \mathbf{x}(t), \mathbf{y}(t) \rangle_{n,a,b} := \sum_{k=0}^{n-1} \langle \vec{\eta}_a^k * \mathbf{x}^{(k)}(a), \mathbf{y}^{(k)}(a) \rangle + \sum_{k=0}^{n-1} \langle \vec{\eta}_b^k * \mathbf{x}^{(k)}(b), \mathbf{y}^{(k)}(b) \rangle + \int_a^b \langle \mathbf{x}^{(n)}(t), \mathbf{y}^{(n)}(t) \rangle dt, \quad (1)$$

where $\langle \cdot, \cdot \rangle$ is the standard inner product in \mathbb{R}^m ,

$$\vec{\eta}_a^k * \mathbf{x}^{(k)}(a) = \left(\eta_{a,1}^k x_1^{(k)}(a), \dots, \eta_{a,m}^k x_m^{(k)}(a) \right),$$

$$\vec{\eta}_b^k * \mathbf{x}^{(k)}(b) = \left(\eta_{b,1}^k x_1^{(k)}(b), \dots, \eta_{b,m}^k x_m^{(k)}(b) \right),$$

$\eta_{a,j}^k \geq 0$, $\eta_{b,j}^k \geq 0$ and $\eta_{a,j}^k + \eta_{b,j}^k > 0$. These metrics are a generalization of metrics considered in [3]. [9], [11]. To simplify the notation, from now on the above inner product (1) and the space $H_n(I, \mathbb{R}^m)$ will be denoted simply by $\langle \cdot, \cdot \rangle_n$ and $X = H_n(I, \mathbb{R}^m)$, respectively.

We want to analyze the variational problem associated to a certain family of energy functionals $\mathcal{F} : X \rightarrow \mathbb{R}$ defined on X or on suitable subspaces of curves in X . We consider functionals of the form

$$\mathcal{F}(\mathbf{x}) = \int_a^b f(t, \mathbf{x}(a), \mathbf{x}(b), \mathbf{x}(t), \dots, \mathbf{x}^{(n-1)}(a), \mathbf{x}^{(n-1)}(b), \mathbf{x}^{(n-1)}, \mathbf{x}^{(n)}) dt, \quad (2)$$

where $\mathbf{x} = (x_j(t)) \in X$, $j = 1, \dots, m$ and $f : W \subset \mathbb{R}^{3nm+m+1} \rightarrow \mathbb{R}$ is a continuously differentiable functions defined on a sufficiently large domains W . We also assume that f satisfies suitable additional conditions which guarantee the (Gâteaux) differentiability of \mathcal{F} and the (local) convergence of the gradient steepest descent method [12].

Usually, one considers \mathcal{F} acting on subspaces of functions $\mathbf{x} = (x_1, \dots, x_m) \in X$ satisfying along with their derivatives $\mathbf{x}^{(i)} = (x_1^{(i)}, \dots, x_m^{(i)})$ given boundary conditions at the endpoints of the interval (they will be referred to as *affine constraints*). For instance, for a given $i \in \{0, 1, \dots, n - 1\}$, fix $\mathbf{p}^i = (p_1^i, \dots, p_m^i)$ and $\mathbf{q}^i = (q_1^i, \dots, q_m^i)$ be points in \mathbb{R}^m for $i \in \{0, 1, \dots, n - 1\}$. Then the following sub-

spaces

$$X_{a,j}^{(i)} = \left\{ \mathbf{y} : [a, b] \rightarrow \mathbb{R}^m; y_j^{(i)}(a) = p_j^i \right\},$$

$$X_{b,j}^{(i)} = \left\{ \mathbf{y} : [a, b] \rightarrow \mathbb{R}^m; y_j^{(i)}(b) = q_j^i \right\},$$

$$X_{a,b,j}^{(i)} = X_{a,j}^{(i)} \cap X_{b,j}^{(i)} = \left\{ \mathbf{y} : [a, b] \rightarrow \mathbb{R}^m; y_j^{(i)}(a) = p_j^i, y_j^{(i)}(b) = q_j^i \right\}, \quad (3)$$

for $j \in \{1, 2, \dots, m\}$ are affine subspaces of X . Denote by X_* any choice of either $X = H_n(I, \mathbb{R}^m)$ or of some intersection of any finite combination of the subspaces in (3). Then, X_* are closed submanifolds of X that can be considered as Hilbert manifolds whose corresponding tangent spaces (as submanifolds of X) at any point $\mathbf{x} \in X_*$ will be denoted by $T_{\mathbf{x}}X_* = \Omega_*$. These tangent spaces are finite intersections of the following linear subspaces for the corresponding $\{i, j\}$

$$\Omega_{a,j}^{(i)} = \left\{ \vec{v} : [a, b] \rightarrow \mathbb{R}^m; v_j^{(i)}(a) = 0 \right\},$$

$$\Omega_{b,j}^{(i)} = \left\{ \vec{v} : [a, b] \rightarrow \mathbb{R}^m; v_j^{(i)}(b) = 0 \right\},$$

$$\Omega_{a,b,j}^{(i)} = \Omega_{a,j}^{(i)} \cap \Omega_{b,j}^{(i)} = \left\{ \vec{v} : [a, b] \rightarrow \mathbb{R}^m; v_j^{(i)}(a) = v_j^{(i)}(b) = 0 \right\}. \quad (4)$$

Note that the linear subspaces of $X = T_{\mathbf{x}}X = \Omega$ given in (4) do not depend on the point $\mathbf{x} \in X_*$ and that the subspaces X_* are affine translations in X of the corresponding tangent spaces Ω_* . So endpoint (affine) constraints lead to spaces of functions which are not linear but they are affine spaces instead what causes minor computational additional difficulties.

In contrast, suppose that we are seeking functions which not only satisfy affine constraints but also verify extra restrictions (which will be called *isoperimetric restrictions*) of the form $\mathcal{G}(\mathbf{x}) = r$, $r \in \mathbb{R}$ where

$$\mathcal{G}(\mathbf{x}) = \int_a^b g(t, \mathbf{x}(a), \mathbf{x}(b), \mathbf{x}(t), \dots, \mathbf{x}^{(n-1)}(a), \mathbf{x}^{(n-1)}(b), \mathbf{x}^{(n-1)}, \mathbf{x}^{(n)}) dt, \quad (5)$$

and g is also a continuously differentiable function defined on sufficiently large domain of $\mathbb{R}^{3nm+m+1}$. Now any candidate to be a solution must lie in the hypersurface $X_{\mathcal{G}} = \mathcal{G}^{-1}(c) \subseteq X$, which is not an affine subspace. Of course, more that one isoperimetric restriction may appear at the same time, and then any solution to the variational problem must lie in $X_{\mathcal{G}} = X_{\mathcal{G}_1} \cap \dots \cap X_{\mathcal{G}_n}$.

Assuming the above functional (2) acting on X_* , the gradient of \mathcal{F} at \mathbf{x} is defined to be the unique $\nabla_* \mathcal{F}_{\mathbf{x}} \in \Omega_*$ that satisfies $\langle \nabla_* \mathcal{F}_{\mathbf{x}}, w \rangle_n = D\mathcal{F}_{\mathbf{x}}(w)$,

for all $w \in \Omega_*$. The existence and uniqueness of $\nabla_* \mathcal{F}_x$ is guaranteed by the Riesz representation theorem. Actually, as it has been proved in [1], for a functional $\mathcal{F}: X_* \rightarrow \mathbb{R}$ of the type (2) (satisfying suitable conditions), the gradient is given by

$$\nabla_* \mathcal{F}_x = (\nabla_* \mathcal{F}_{x_1}, \dots, \nabla_* \mathcal{F}_{x_m}) \in \Omega_*, \quad (6)$$

with

$$\nabla_* \mathcal{F}_{x_j} = \int_a^t \dots \int_a^t E_{x_j, n}^f dt + P_{x_j, 2n-1}(t), \quad (7)$$

where $E_{x_j, n}^f$ are defined recursively as

$$\begin{aligned} E_{x_j, 0}^f &= \frac{\partial f}{\partial x_j} = f_{x_j}, \\ E_{x_j, i}^f &= f_{x_j^{(i)}} - \int_a^t E_{x_j, i-1}^f ds \\ &= \sum_{k=0}^i (-1)^{i-k} \int_a^t \dots \int_a^t f_{x_j^{(j)}} ds. \end{aligned} \quad (8)$$

and $P_{x_j, 2n-1}(t) = \sum_{k=0}^{2n-1} c_{j,k} t^k$ are polynomials of degree $2n - 1$, whose coefficients $c_{j,k}$ are fully determined as the solutions of $2n \times 2n$ linear systems which depend on the concrete choice of the metric $\langle \cdot, \cdot \rangle_n$ and the affine boundary constraints.

Notice then that, while an extremal is a zero of the gradient for any choice of the metric (1), the gradient itself depend on the metric and, therefore the metric choice is crucial in our computations.

When \mathcal{F} is considered acting on a subspace satisfying additional isoperimetric constraints $X_{*, \mathcal{G}}$, the gradient of \mathcal{F} is the orthogonal projection of $\nabla_* \mathcal{F}_x$ onto the corresponding tangent space and computation of the gradient requires a more elaborated process. To simplify, we first assume that we are given an isoperimetric restriction \mathcal{G} as defined in (5). For this process to work we need to assume also that \mathcal{G} is *regular*, that is, $\nabla_* \mathcal{G}_x \neq 0$ at any point $x \in X_*$ where the functional \mathcal{F} is defined. Then Ω_* splits into the orthogonal sum or the line span by $\nabla_* \mathcal{G}_x$ and the tangent spaces $\Omega_{*, x, \mathcal{G}} = \{v \in \Omega_*; \langle \nabla_* \mathcal{G}_x, v \rangle = 0\}$ to the hypersurface $X_{*, \mathcal{G}} = \{x \in X_*, \mathcal{G}(x) = r\} \subset X_*$. The splitting depends on the point x and on the vector field $\nabla_* \mathcal{G}_x$ which spans the normal space to $\Omega_{*, x, \mathcal{G}}$. Denote now by $(\nabla_* \mathcal{F}_x)_{\mathcal{G}}$ the orthogonal projection of $\nabla_* \mathcal{F}_x$ onto the tangent space $\Omega_{*, x, \mathcal{G}}$, $\nabla_* \mathcal{F}_x = (\nabla_* \mathcal{F}_x)_{\mathcal{G}} + \lambda(x) \nabla_* \mathcal{G}_x$, for a certain smooth function $\lambda(x)$ on X_* , and the condition

$$D_x \mathcal{G}((\nabla_* \mathcal{F}_x)_{\mathcal{G}}) = 0$$

determines $\lambda(x)$, which is thus given by

$$\lambda(x) = \frac{\langle \nabla_* \mathcal{F}_x, \nabla_* \mathcal{G}_x \rangle}{\langle \nabla_* \mathcal{G}_x, \nabla_* \mathcal{G}_x \rangle}. \quad (9)$$

In case there are a finite number of isoperimetric constraints $\mathcal{G}_1, \dots, \mathcal{G}_h : X_* \rightarrow \mathbb{R}$. Then, the variational problem is defined on points $x \in X_*$ satisfying $\mathcal{G}_i(x) = r_i$ for $i \in \{1, \dots, h\}$, that is to say, the submanifold of points where the problem is defined, $X_{*, \mathcal{G}}$, and its tangent space, $\Omega_{*, x, \mathcal{G}}$, are, respectively

$$X_{*, \mathcal{G}} = X_{*, \mathcal{G}_1} \cap \dots \cap X_{*, \mathcal{G}_h},$$

$$\Omega_{*, x, \mathcal{G}} = \Omega_{*, x, \mathcal{G}_1} \cap \dots \cap \Omega_{*, x, \mathcal{G}_h}.$$

Thus, for any $x \in X_{*, \mathcal{G}}$ the projected gradient $(\nabla_* \mathcal{F}_x)_{\mathcal{G}}$ can be obtained from the decomposition

$$\begin{aligned} \nabla_* \mathcal{F}_x &= (\nabla_* \mathcal{F}_x)_{\mathcal{G}} + \lambda_1(x) \nabla_* \mathcal{G}_{1, x} + \dots \\ &\quad + \lambda_h(x) \nabla_* \mathcal{G}_{h, x}, \end{aligned} \quad (10)$$

for certain smooth functions $\{\lambda_i(x), i = 1, \dots, h\}$. Acting as previously in the single case, the functions $\lambda_i(x)$ can be obtained by solving a certain linear system (for more details see [1]).

On the other hand, one of the most common methods for minimization of \mathcal{F} is the gradient steepest descent method. Basically, the essence of this method is to analyze the behavior of the sequence $\{x_k, k \in \mathbb{N}\}$ of successive approximations for the local minimum points of \mathcal{F} given by the formula

$$x_{k+1} = x_k + t_k h_k, \quad k \in \mathbb{N},$$

where t_k is a sequence of positive numbers, the so called control parameters, which lie in a closed interval of the real line. In order to construct the sequence $\{x_k, k \in \mathbb{N}\}$, start with an arbitrary point $x_o \in X_*$ (where, of course, $\nabla_* \mathcal{F}_{x_o} \neq 0$), then, assuming that x_o, x_1, \dots, x_k have already been constructed, proceed by choosing a sequence $h_k \in \Omega_*$ such that $\langle \nabla_* \mathcal{F}_{x_k}, x_k \rangle < 0$ (usually, $h_k = -\nabla_* \mathcal{F}_{x_k}$) and then take $x_{k+1} = x_k + t_k h_k$. Assuming at the moment that only affine constraints are present, the functional (2) is defined on the space of functions X_* , an affine subspace of the main space X described previously. To solve the variational problem one seeks a function $x \in X_*$ that minimizes (2) among curves satisfying the boundary affine constraints. At the computational side of the problem, the implemented gradient descent algorithm creates a *functional decreasing sequence* $\{x_\nu\}$, $\nu \in \mathbb{N}$, into the X_* space itself, in such a way that

$$\mathcal{F}(x_0) > \mathcal{F}(x_1) > \dots > \mathcal{F}(x_\nu) > \dots \quad (11)$$

while $|\nabla_* \mathcal{F}_{x_\nu}| \rightarrow 0$, as $\nu \rightarrow \infty$. To start with, let $x_\nu \in X_*$ be a function satisfying the boundary constraints and take its gradient $\nabla_* \mathcal{F}_{x_\nu} \in \Omega_*$. The curve $\Gamma_\nu : \mathbb{R} \rightarrow X_*$ given by

$$\Gamma_\nu(\lambda) = x_\nu - \lambda \cdot \nabla_* \mathcal{F}_{x_\nu} \quad (12)$$

is then well defined. For $\lambda > 0$ this curve begins decreasing the functional, that is, $\mathcal{F}(\Gamma_\nu(\lambda))$ is a decreasing function for $\lambda > 0$ small enough. Hence, choosing an adequate $\lambda_\nu > 0$ and taking $x_{\nu+1} = \Gamma(\lambda_\nu)$, we have that $\mathcal{F}(x_\nu) > \mathcal{F}(x_{\nu+1})$ with $x_{\nu+1}$ satisfying the same affine boundary constraints as x_ν . Continuing in this manner, we obtain the decreasing sequence (11) and the way the $\{\lambda_\nu\}$ are chosen determines the way in which $|\nabla_* \mathcal{F}_{x_\nu}| \rightarrow 0$. Typically, a *recursion termination condition*, that is to say, a pre-determined stopping criterion, is established so that the algorithm is stopped when it is satisfied. Introduction of isoperimetric constraints at this point comes at a price, because now the space of functions $X_{*,\mathcal{G}}$, where the variational problem is defined, is not an affine subspace of X , and given $x_\nu \in X_{*,\mathcal{G}}$, the curve

$$\Gamma_\nu(\lambda) = x_\nu - \lambda \cdot (\nabla_* \mathcal{F}_{x_\nu})_{\mathcal{G}} \quad (13)$$

does not need to lie on $X_{*,\mathcal{G}}$. Thus, even though for small enough $\lambda > 0$, $\mathcal{F}(\Gamma_\nu(\lambda))$ is a decreasing function, the curve $\Gamma_\nu(\lambda)$ may be moving away from satisfying the isoperimetric constraint. Different procedures have been established in [1] to overcome this issue.

Although, for simplicity, the above arguments have been made for the traditional gradient descent linear model, we have proposed in [1] to explore new gradient descent curve models, within the scope of this type of variational problems. In fact, the standard linear model can be seen as the linear approximation to an ideal curve $\tilde{\Gamma} = \tilde{\Gamma}(\lambda)$ in $X_{*,\mathcal{G}}$, that passes through the point $\tilde{\Gamma}(0) = x_\nu$ with tangent vector $\tilde{\Gamma}'(0) = -(\nabla_* \mathcal{F}_{x_\nu})_{\mathcal{G}}$, descending towards the solution function of the variational problem. That is, the descent curve (13) would be nothing but

$$\Gamma_\nu(\lambda) = \tilde{\Gamma}(0) + \lambda \cdot \tilde{\Gamma}'(0) .$$

So, if there were a way to approximate the second derivative $\tilde{\Gamma}''(0)$, it would be reasonable to think that the new curve

$$\Gamma_{2,\nu}(\lambda) = \tilde{\Gamma}(0) + \lambda \cdot \tilde{\Gamma}'(0) + \frac{\lambda^2}{2} \tilde{\Gamma}''(0)$$

will improve the linear approximation. This approach have been considered in [1] not only for quadratic but also for cubic approximations of the descent curves.

Finally, a numerical method to locate minimizers of this general class of variational problems under both affine and isoperimetric constraints is implemented in [1] (see also, www.ikergeometry.org). The method is suitable for application to the energy functionals described previously, in particular it will be applicable to actions described in the next sections after some convenient adjustments.

1.2 Transversality constraints

Consider the isoperimetric constraint $\mathcal{G}(x) = r$, $r \in \mathbb{R}$, introduced in (5) and assume that the function g has the form

$$g(x(a), x(b), \dots, x^{(n-1)}(a), x^{(n-1)}(b)) . \quad (14)$$

Then, in this case the isoperimetric constraint determined by $\mathcal{G}(x)$ is actually a *transversality condition* given by $\mathcal{G}(x)$ with

$$\begin{aligned} \mathcal{G}(x) = & (b - a) \times \\ & \times g(x(a), x(b), \dots, x^{(n-1)}(a), x^{(n-1)}(b)) , \end{aligned} \quad (15)$$

that should be satisfied by a solution x at its end points. Thus, these kind of transversality conditions generalize the affine constraints and can be implemented using the isoperimetric constraint scheme. That is, they can be treated as isoperimetric constraints and what it has been proved in [1] for the latter applies to this case. In particular, the numerical procedure implemented in [1] under the XEL-platform to deal with order 1, 2 and 3 gradient descent methods, works equally well when dealing with transversality constraints. This approach is analyzed in the next two sections for some concrete well known lagrangian energies. For our analysis we use the quadratic (order 2) gradient descent method. This method is better suited for dealing with isoperimetric constraints than the linear one, being, at the same time, much less computational demanding than the cubic (order 3) method.

2 Brachystochrone

We first apply the XEL procedure to a version of the brachystochrone problem. As originally posed by Johann Bernoulli in 1696, the brachystochrone problem consists in finding the curve joining two given points (not vertically aligned) along which a bead of given mass falls under the influence of gravity in the minimum time. Newton, Leibniz, L'Hospital, and Jakob

and Johann Bernoulli showed that the solution is a cycloid, the curve traced out by a point on the rim of a rolling circle, [4]. Solutions of the classical Brachistochrone problem typically use techniques of calculus of variations, [6]. Here it will be solved by considering affine, isoperimetric and transversality constraints.

The shape of the wire will be given by the graph of a continuously differentiable function $y : I = [0, 1] \rightarrow \mathbb{R}$, $y \in H_1(I, \mathbb{R}^m)$, so we may think of the curve $\mathbf{x}(t) = (x(t), y(t))$ as a frictionless wire in a vertical plane (x, y) , with positive x -axis extending to the right and negative function y -axis extending downwards.

The descent time of a particle, assumed with an initial velocity $v_0 \neq 0$, along a curve $(x(t), y(t))$ joining the origin $(0, 0)$ and a point $(x(1), y(1))$, $y(1) < 0$, is given by

$$\mathcal{T}(\mathbf{x}) : = \frac{1}{\sqrt{2G}} \int_0^1 \frac{\sqrt{(x')^2 + (y')^2}}{\sqrt{v_0^2/(2G) - y(t)}} dt,$$

G being the gravitational force. Thus, the shape of the wire will be obtained by minimizing the above functional \mathcal{T} subject to certain boundary and transversality conditions. These can be complemented also with isoperimetric constraints that, in our case will come by prescribing the length of the curve

$$\mathcal{L}(\mathbf{x}) : = \int_0^1 \sqrt{(x')^2 + (y')^2} dt.$$

Putting a penalty on the length of the variation curves can be interpreted, by using a Lagrange multiplier argument, in terms of the minimization of the extended functional $\mathcal{T} - \lambda \mathcal{L}$. On the other hand, as pointed out in section 1.2, the isoperimetric constraints scheme introduced in [1] can be used to implement transversality conditions to be satisfied by the endpoints of a brachistochrone. For instance, suppose we want to obtain solutions to the brachistochrone problem for particles traveling from $(0, 0)$ to a point $\mathbf{x}(1) = (x(1), y(1))$ freely located on a suitable circumference, say, $(x - x_0)^2 + (y - y_0)^2 = r^2$. Then, we consider the Time and Length functionals, \mathcal{T} and \mathcal{L} respectively, acting on the space of functions

$$X_* = \{ \mathbf{x} = (x, y) : [0, 1] \rightarrow \mathbb{R}^2; \\ x(0) = 0, y(0) = 0 \},$$

whose tangent space is also $\Omega_* = X_*$.

The new transversality restriction specified by imposing the right endpoint to lie on a circle may be switched on by invoking another isoperimetric constraint, $\mathcal{G}_2(\mathbf{x}) = 0$, obtained by setting

$$g = (x(1) - x_0)^2 + (y(1) - y_0)^2 - r^2,$$

in (15), (x_0, y_0) and r being the circle center and radius, respectively.

Figure 1 shows the XEL output corresponding to the evolution under the gradient flow of a given initial curve (in black) to the brachistochrone curve (in blue) with free right endpoint on a circle, and assuming both no length constraint on the evolution curves (figure 1a) and an evolution by curves of the same length (figure 1b). In both cases the brachistochrone solution goes from the point $(0, 0)$ to the circle

$$(x - 0.85)^2 + (y + 1.15)^2 = 0.045.$$

The first curve of both evolutions is the same:

$$\mathbf{x}_0(t) = \left(x_0(t) = t, y_0(t) = -t^2 \sin^2 \left(\frac{3}{2} \pi t \right) \right).$$

Moreover, table 1 shows the corresponding numerical time in seconds along the solutions and the accumulated errors along both descents if non affine constraints are considered. A little explanation on what is meant by this is the following. Curves along the evolution show a small deviation from fulfilling the transversality condition (treated as isoperimetric constraint here), i.e., the right endpoint is slightly apart from the given circle. What the error row of table 1 contains is the accumulated error of all curves appearing in the evolution. As it is clear, the error is not numerically relevant.

\mathcal{T}	0.36756074	0.46297652
error	$4.0 \cdot 10^{-8}$	$6.0 \cdot 10^{-7}$

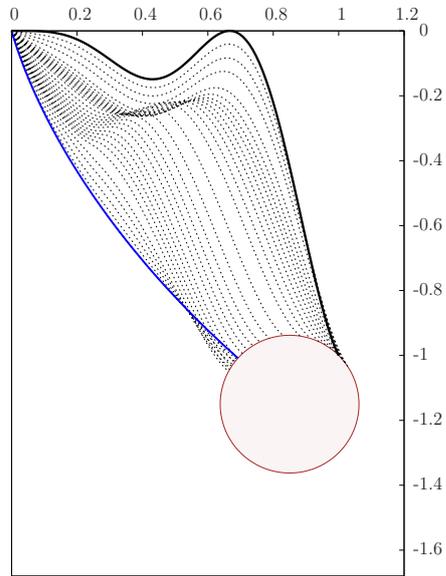
Table 1

To end this section a similar experiment is repeated when the left endpoint of the brachistochrone is $(0, 0)$ while its right endpoint is required to lie freely on a parable $y = m - n(x - x_0)^2$. Again both cases, minimizers among arbitrary length curves and minimizers when there is a length penalty on the variation curves are considered.

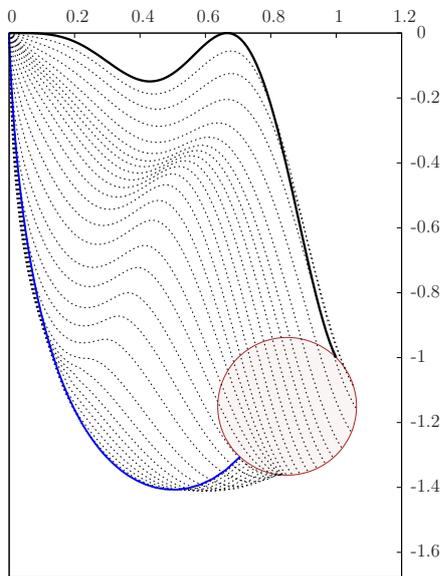
Figure 2 shows the brachistochrone curve (in blue) with free right endpoint on a certain parable assuming both no length constraint (figure 2a) and a length penalty for the evolution curves (figure 2b).

In this particular instance the brachistochrone solution goes from the point $(0, 0)$ to the parable

$$y = -0.96 - 4(x - 0.9)^2.$$

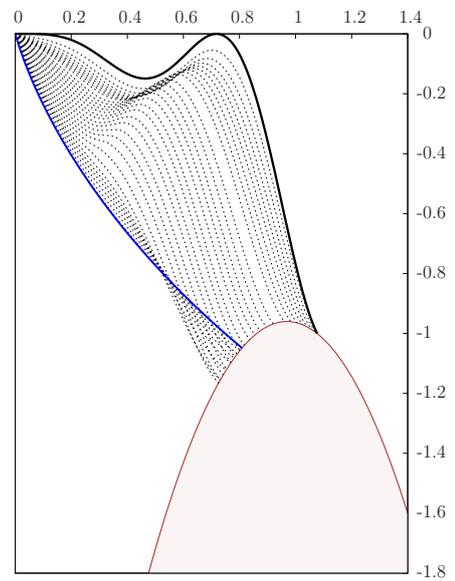


(a) With no penalty on the length of variation curves

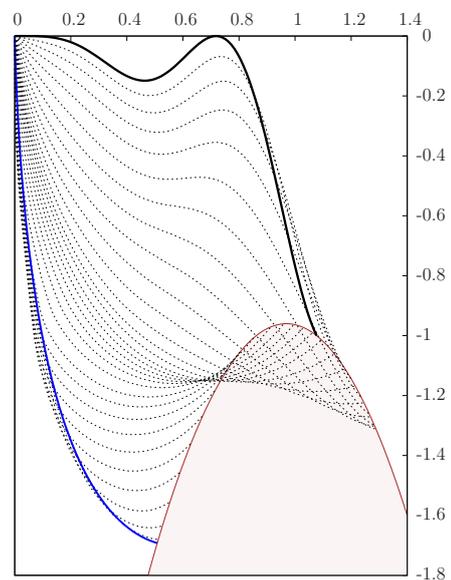


(b) With length penalty

Figure 1: Brachystochrone from $(0, 0)$ to a circle with and without length penalty.



(a) With no penalty on the length of variation curves



(b) With length penalty

Figure 2: Brachystochrone from $(0, 0)$ to a parabola with and without length penalty.

Also, the same curve as before x_0

$$x_0(t) = \left(x_0(t) = t, y_0(t) = -t^2 \sin^2 \left(\frac{3}{2} \pi t \right) \right)$$

has been used in both cases as initial step to initiate the descent method. Once more, the corresponding numerical time in seconds and the accumulated errors along the descents (with the same meaning as that of table 1) if non affine constraints are present have been collected in table 2.

\mathcal{T}	0.38226595	0.45807026
error	$8.0 \cdot 10^{-8}$	$8.0 \cdot 10^{-7}$

Table 2

3 Closed Elasticae in \mathbb{R}^2

Consider now another classical variational problem, the problem of minimizing the bending energy

$$\mathcal{F}(\gamma) = \int_{\gamma} \kappa^2(s) ds, \tag{16}$$

for planar curves γ joining two given points $p = (p_1, p_2)$ and $q = (q_1, q_2)$ in \mathbb{R}^2 , κ and s being the curvature and arc-length parameter of γ , respectively. In a letter to L. Euler in 1742, D. Bernoulli proposed minimizers of (16) as models for the elastic rods. Thus, curves minimizing (or more generally, critical curves of) (16) are called elastic curves or, simply, elasticae. Planar elastic curves were classified by L. Euler [5] in 1743 and since then, the variational problem associated to elastic curves, under different constraints and boundary conditions, has attracted the interest of many mathematicians [7]. In particular, their curvatures can be computed explicitly in terms of the Jacobi elliptic functions [8]. Then, there exist elasticae for which the curvature is always positive, called *orbitlike* or *non-inflectional* elasticae, and elasticae whose curvature changes sign, called *wavelike* or *inflectional* elasticae (for more details, see [8], [10]).

If we parametrize a plane curve by using the angle which makes with a fixed direction, $\theta(s)$, then the curvature is nothing but $\kappa(s) = \theta'(s)$ and the bending energy (16) can be written as

$$\mathcal{F}(\gamma) = \mathcal{F}(\theta) = \int_{\gamma} (\theta'(s))^2 ds. \tag{17}$$

Hence, up to plane congruences, the arc-length Cartesian parametrization of the curve γ is then obtained

by

$$\gamma(s) = \left(\int_{\gamma} \cos \theta(s), \int_{\gamma} \sin \theta(s) \right), \tag{18}$$

so the functionals defined by

$$\mathcal{G}_1(\gamma) = \mathcal{G}_1(\theta) = \int_{\gamma} \cos \theta(s) ds, \tag{19}$$

$$\mathcal{G}_2(\gamma) = \mathcal{G}_2(\theta) = \int_{\gamma} \sin \theta(s) ds. \tag{20}$$

will give us the isoperimetric constraints needed to ensure that the curves actually join the two given points p and q . Thus, the study of the variational problem associated to (16) is restricted to curves γ whose turning angle function $\theta(s)$ satisfies the equations

$$\mathcal{G}_1(\theta) = q_1 - p_1 \quad \text{and} \quad \mathcal{G}_2(\theta) = q_2 - p_2. \tag{21}$$

Angular constraints at the endpoints may be imposed by setting angles $\psi_0 = \angle(\vec{pq}, \vec{v})$ and $\psi_1 = \angle(\vec{pq}, \vec{w})$ at the points $p = \gamma(0)$ and $q = \gamma(L)$, respectively, in the above arc-length parametrization (18) of plane curves. These constraints are affine (boundary) conditions that determine both the affine space X_* and the vector space Ω_* where the variational problem is to be defined. According to section 1, we have that if the angles ψ_0 and ψ_1 have to remain fixed at endpoints the corresponding spaces are given by

$$\begin{aligned} X_* &= \{ \theta : [0, L] \rightarrow \mathbb{R}; \theta(0) = \psi_0, \\ &\quad \theta(L) = \psi_1 \text{ mod}(2\pi) \}, \\ \Omega_* &= \{ \vec{\phi} : [0, L] \rightarrow \mathbb{R}; \vec{\phi}(0) = 0, \vec{\phi}(L) = 0 \}. \end{aligned} \tag{22}$$

According to the notation introduced in section 1 the spaces X_* and the vector space Ω_* would correspond to $X_{a,b,j}^{(i)}$, $\Omega_{a,b,j}^{(i)}$, for $a = 0, b = L, i = 0$ and $j = 1$. The metric (1) defined on vector spaces Ω_* is given by

$$\langle \vec{\phi}, \vec{\varphi} \rangle = \vec{\phi}(0) \cdot \vec{\varphi}(0) + \int_0^L \vec{\phi}' \cdot \vec{\varphi}' ds. \tag{23}$$

Let us focus now on the problem of closed elastic curves. Closed elastica can be considered an special case of elastica with clamped points, that is, the variational problem is defined on the space of curves traveling from p to q , leaving p with direction \vec{v} and arriving at q with direction \vec{w} . On the other hand, it is well known [8], that the only planar closed elasticae are the figure-eight shaped wavelike elastica and the circle orbitlike elastica.

To work with XEL on closed elastic curves, we set $p = q$ and $\vec{w} = \vec{v}$ (that is $\psi_0 = \psi_1$), so, in order to guarantee that all curves during the gradient descent are closed, the experiments must take place in the space of curves parametrized by functions $\theta = \theta(s)$ satisfying the isoperimetric constraints $\mathcal{G}_1(\theta) = \mathcal{G}_2(\theta) = 0$, ($\mathcal{G}_1, \mathcal{G}_2$ are given in (21)), and the transversality condition

$$\theta(1) - \theta(0) = \int_0^1 \kappa(s) ds = 2\pi k,$$

for $k \in \mathbb{Z}$. Here, $k = 0$ stands for wavelike case while $k \neq 0$ implies orbitlike case.

In addition, we are going to check also that different evolutions created from far apart initial curves (meaning by this that they have very different elastic energy) of the same homotopy class, arrive to the same solution curve, as expected.

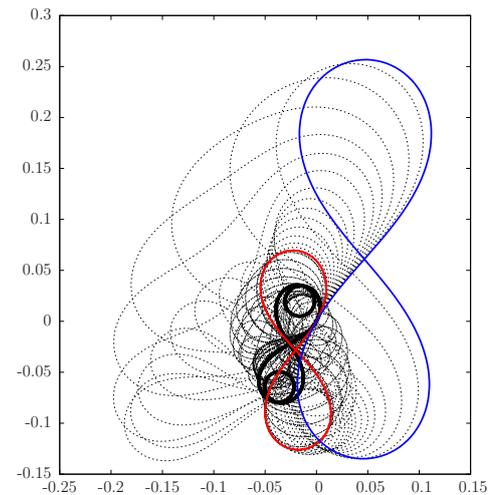
In fact, consider the following pairs of functions $\theta_0(s) = \tilde{\theta}(s)$ and $\theta_0(s) = \hat{\theta}(s)$ as initial curves to feed the gradient flow descent evolutions seeking for the figure-eight shaped elastica

$$\begin{aligned} \tilde{\theta}(s) &= \eta \sin(4\pi s) - 5 \cos(4\pi s), \\ \hat{\theta}(s) &= \eta \sin(64\pi s) - 5 \cos(64\pi s), \end{aligned} \tag{24}$$

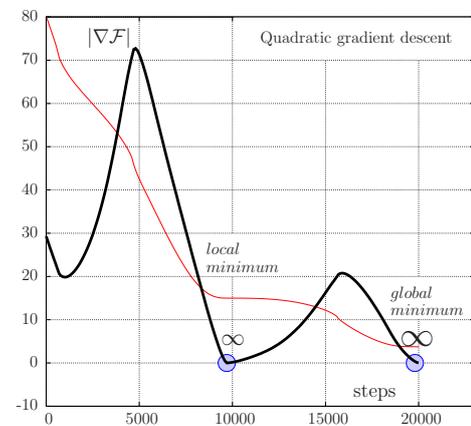
where $s \in [0, 1]$, $\tilde{\theta}(0) = \tilde{\theta}(1)$, $\hat{\theta}(0) = \hat{\theta}(1)$, and the parameter $\eta = 2.339073\dots$ has been computed to ensure that both curves satisfy the isoperimetric boundary conditions.

Step by step evolution of the initial curve $\theta_0 = \tilde{\theta}(s)$ is analyzed graphically in figure 3. The other initial function $\hat{\theta}(s)$ define a curve so knotty and its graphical evolution is so dense and we are not to make a clear picture of it. In spite of that, our numerical experiments show that in both cases the algorithm takes the two initial curves of (24) to the same global minimum: the eight figure elastica. Figure 3 shows the gradient flow evolution with initial curve $\theta_0 = \tilde{\theta}(s)$ (in black) to the eight figure elastica (in blue). The procedure stops once prescribed stopping conditions are fulfilled.

In this case, we see from to table 3 that, even though the bending energy of $\hat{\theta}(s)$ is more than two hundred times greater than that of $\tilde{\theta}$, the energies of the two solution curves reached by the algorithm taking them, respectively, as initial curves in the descent, are separated by less than $1.20 \cdot 10^{-7}$. On the other hand, in [2] we explicitly determine the curvature $\kappa(s)$ of the analytical wavelike elastic closed curve (figure eight curve) and obtain its elastic energy,



(a) The curve in red is an intermediate local minimums



(b) Schematic picture of the quadratic gradient descent evolution shown in (a). The black curve plots the gradient norm while the algorithm decreases the elastic energy (red curve) searching for its minimum.

Figure 3: Evolution of $\theta_0 = \tilde{\theta}$ curve (in black) under the gradient descent.

which is to be considered as reference value. Then, we see that the relative deviations between the numerical approximations energies and this value of reference are less than $5.00 \cdot 10^{-7}$. Taking into account also that, in both cases, once the descents are completed the deviations in the isoperimetric constraints are less than $6.00 \cdot 10^{-7}$, we see that the differences are not numerically relevant.

Moreover, along the curve evolution starting at $\tilde{\theta}$ the descent algorithm reports on the existence of an intermediate local minimum (colored in red in figure 3a). That is, in this case the first candidate for minimizer the algorithm reports on is, in fact, a local minimum which slows down the energy decreasing rate for a while. Then, the energy decreasing rate speeds up again until the final solution is reached. The rel-

ative deviation of the final energy with respect to the reference value is also less than $3.00 \cdot 10^{-7}$. This situation is outlined in figure 3b where the horizontal axis indicates the number of steps during the descending sequence. If we had chosen the cubic gradient descent evolution [1] instead of the quadratic one, the outlined picture would be roughly the same, but only a 20% of the number of steps would have been needed.

As for the circle case, we take as initial curves $\tilde{\theta}(s)$ and $\hat{\theta}(s)$ the functions given by

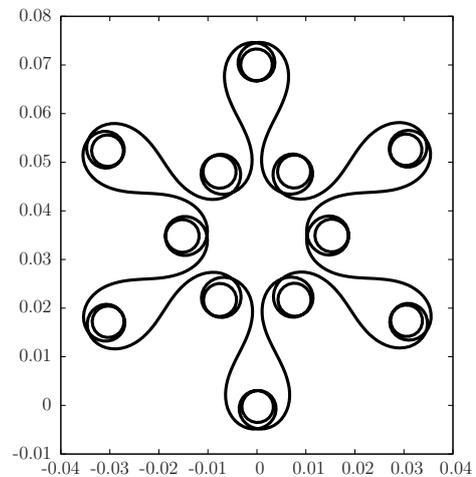
$$\begin{aligned} \tilde{\theta}(s) &= 2\pi s + \cos(12\pi s) + 8 \sin(12\pi s), \\ \hat{\theta}(s) &= 2\pi s + 12 \cos(12\pi s) + 24 \sin(24\pi s), \end{aligned} \quad (25)$$

where $s \in [0, 1]$ and $\tilde{\theta}(1) - \tilde{\theta}(0) = 2\pi$.

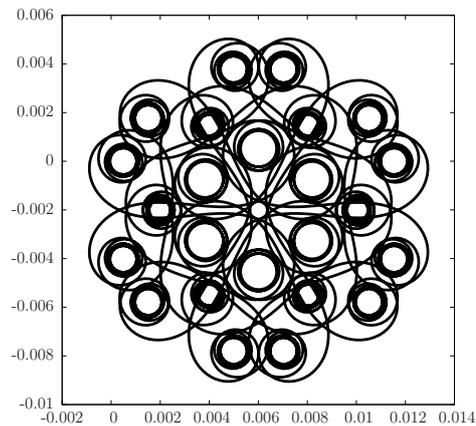
Again, step by step evolutions of the initial curve $\theta_0 = \tilde{\theta}(s)$ is analyzed graphically in figure 4. Once more, the initial function $\hat{\theta}(s)$ defines an extremely knotty curve so that we are not able to make a clear picture of its graphical evolution. Nevertheless, it can be seen that in both cases the algorithm takes the two initial curves of (25) to the global minimum in this case: the circular elastica. Figure 2b shows the evolution of $\tilde{\theta}$ (in black) towards to the global minimum: the circular elastica (in blue). The procedure stops once prescribed stopping conditions are fulfilled.

According to table 4 we see that in case of the circular elastica, the consequences are quite similar to those outlined above for the eight-figure elastica. For instance, we observe that along the main descent with prescribed affine constraints, the energies of the solution curves reached by the algorithm and the energy considered as reference value are separated by less than $6.00 \cdot 10^{-11}$. However, this time no intermediate local minimum shows up.

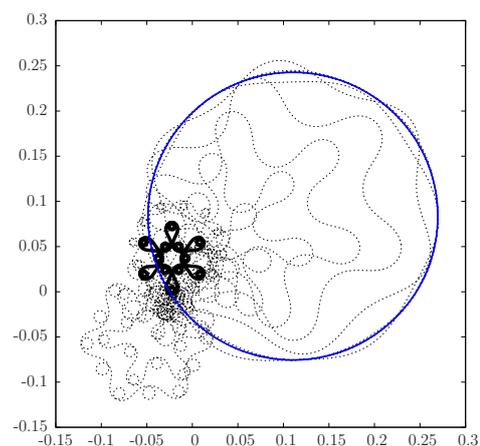
Finally, the above two evolution problems ending at the eight figure and circle, respectively, can be considered also under other affine constraints. For instance, we may assume that at least one of the tangent vectors (angles) at the endpoints is unprescribed. Due to the symmetry of the problem, the solution comes out immediately: in all cases, it would be the half of the figure-eight shaped elastica, while its energy would be that of a quarter of the figure-eight elastica bending energy. These facts are reflected in columns $\mathcal{F}(\theta_{last,1})$ and $\mathcal{F}(\theta_{last,2})$ of tables 3 and 4, respectively.



(a) Initial curve $\tilde{\theta}(s)$ given in (25) for a gradient descent looking for the circle.



(b) Initial curve $\hat{\theta}(s)$ given in (25).



(c) Evolution using $\tilde{\theta}$ as first step

Figure 4: The second curve is so curly that its evolution is too dense to take a clear picture, but it evolves toward the circle too. The numerical values are collected in table 4

θ_0	$\mathcal{F}(\theta_0)$	$\mathcal{F}(\theta_{last})$	$\mathcal{F}(\theta_{mid})$
$\hat{\theta}$	2406.0	112.439651	449.758531
$\tilde{\theta}$	615914	112.439597	-
1-1 ref.		112.439610	449.758438
θ_0	$\mathcal{F}(\theta_{last,1})$	$\mathcal{F}(\theta_{last,2})$	$\delta\mathcal{G}$
$\hat{\theta}$	28.109902	28.109902	$4.00 \cdot 10^{-8}$
$\tilde{\theta}$	28.109916	28.109906	$6.00 \cdot 10^{-7}$
1-1 ref.	28.109902	28.109902	

Table 3: Numerical summary of the experiments on the eight-shaped elasticae. $\mathcal{F}(\theta_0)$ stands for the bending energy of the initial curves in the descent; $\mathcal{F}(\theta_{last})$, $\mathcal{F}(\theta_{last,1})$ and $\mathcal{F}(\theta_{last,2})$ are the bending energies of the final curves for the three cases on affine boundary constraints; $\mathcal{F}(\theta_{mid})$ is the energy of the intermediate local minima (if any) that the descents report; and $\delta\mathcal{G}$ is the highest accumulated relative deviations of the isoperimetric constraints in each descent. The last row contains the reference values obtained in table 2 of [2].

θ_0	$\mathcal{F}(\theta_0)$	$\mathcal{F}(\theta_{last})$	$\mathcal{F}(\theta_{last,1})$
$\hat{\theta}$	46229	39.478418	28.109898
$\tilde{\theta}$	1739616	39.478418	28.109897
1-1 ref.	$4\pi^2 = 39.478417604\dots$		28.109902
θ_0	$\mathcal{F}(\theta_{last,2})$	$\delta\mathcal{G}$	-
$\hat{\theta}$	28.109898	$1.00 \cdot 10^{-7}$	-
$\tilde{\theta}$	28.109896	$2.20 \cdot 10^{-6}$	-
1-1 ref.	28.109902	-	-

Table 4: Summary of the experiments on the circular elastica with the same meanings for column headers as in table 3.

Acknowledgements: Research partially supported by MINECO-FEDER grant MTM2014-54804-P and Gobierno Vasco grant IT1094-16. Spain.

References:

- [1] J. Arroyo, O.J. Garay, J.J. Mencia and A. Pámpano, A gradient-descent method for Lagrangian densities depending on multiple derivatives, Preprint, 2018.
- [2] J. Arroyo, O.J. Garay and A. Pámpano, Boundary value problems for Euler-Bernoulli planar elastica. A solution construction procedure, Submitted.
- [3] H. Brézis, *Analyse fonctionnelle*, Mason Editeur, Paris, 1993.
- [4] W. Dunham, *Journey Through Genius*, Penguin Books, New York, 1991.
- [5] L. Euler, *Methodus inveniendi lineas curvas maximi minimive proprietate gaudentes, sive solutio problematis isoperimetrici lattissimo sensu accepti*, Bousquet, Lausannae et Genevae E 65A. O.O. Ser. I vol 24 1744.
- [6] I.M. Gelfand and S.V. Fomin, *Calculus of Variations*, Prentice-Hall Inc., Englewood Cliffs, NJ, 1963.
- [7] R. Levien, The elastica: a mathematical history. Technical Report No. UCB/EECS-2008-103, Univ. of Berkeley, <http://www.eecs.berkeley.edu/Pubs/TechRpts/2008/EECS-2008-103.html>.
- [8] J. Langer and D. A. Singer, The total squared curvature of closed curves, *J. Differ. Geom.* 20 (1984) 1-22.
- [9] A. Linnér, Gradients, preferred metrics and asymmetries. Preprint, 2001, <http://www.math.niu.edu/alinner/>
- [10] A.E. Love, *A treatise on the Mathematical theory of Elasticity*, Dover Publications, New York, 1944.
- [11] R.S. Palais, Morse theory on Hilbert manifolds, *Topology* 2 (1963) 299-340.
- [12] G. Smyrlis and V. Zisis, Local convergence of the steepest descent method in Hilbert spaces, *J. Math. Anal. Appl.* 300 (2004) 436-453.