GRADIENTS, PREFERRED METRICS AND ASYMMETRIES

ANDERS LINNÉR

ABSTRACT. Assume the curves x_τ flow according to $\partial x_\tau / \partial \tau = \Phi(x_\tau)$. Evolutions that only use *local* data, such as curve shortening, yield partial differential equations. In the present paper *global* information is permitted and Φ involves integrals of x_τ. Specifically, $\Phi = -\nabla F$ where $F : \Omega \to \mathbb{R}$ is given by $F(x) = \int f(t, x, \dot{x})\, dt$. The elements of Ω correspond to curves subject to point-wise and isoperimetric constraints. Ideally Ω is an infinite-dimensional smooth 'surface-like' submanifold of some suitable Hilbert space X. The choice of inner product in X profoundly affects the formulas for the gradient and also properties of x_τ such as smoothness and symmetry. The gradient formulas are derived without the help of any Hilbert basis in three different inner products. The equation $\nabla F(x) = 0$ is shown to be more discriminating than the Euler-Lagrange equation. The main contribution is the identification of two fundamental quantities that radically simplify and streamline all the gradient expressions. One of the three metrics retains both smoothness and symmetry. Interestingly, the gradient in this metric fails to commute with the derivative operator. There is essentially only one type of metric whose gradient commutes with the derivative, but its flow can destroy symmetry and ruin smoothness. There is a non-standard metric of this type whose flow does preserve symmetry. The symbolic form of the gradient is not symmetric with respect to external symmetries and it is possible to exploit this 'asymmetry' and choose the metric so that the form of the gradients is greatly simplified.

1. Introduction.

1.1 Motivation. The investigation presented here is motivated by a phenomenon observed in the geometric curve-straightening flow, which loosely speaking is the negative gradient flow of the total squared curvature. If the flow takes place in the space of closed planar curves of rotation number *one*, then it is shown in [**2**] that the flow preserves several symmetries. In 1987 the author completed a numerical simulation of the flow designed to handle curves that lack symmetry and also curves of *arbitrary* rotation number. In the case when the rotation number is *zero* there is a certain double figure eight, which has a vertical axis of symmetry, and this axis of symmetry rotates under the flow; see [**3**]. Based on these examples there was little reason to doubt that the curve-straightening flow preserves symmetries. In 1995 the author extended the algorithm to allow for curves of variable length and non-periodic boundary conditions. During this process a considerable effort was needlessly spent in an attempt to remove a perceived flaw in the implementation leading to a loss of symmetry. As it turned out that the flaw was actually a feature. The curve-straightening flow does in fact not necessarily preserve symmetries of reflection in the case of non-closed curves. It is proved rigorously in **[5]** that the semi-circle, which flows towards its own diameter, will not retain its symmetry of reflection. Both Robert Bryant and Joel Langer have independently asked the author about the reason for this. The short, and admittedly unsatisfying, answer is that the failure to preserve symmetries is due to the choice of inner product. The purpose of this paper and its companion [6] "Symmetrized Curve-Straightening" is to give the complete answer.

1.2 Comparison with the Calculus of Variations. Consider functionals of the form $F(x) = \int f\left(t, x(t), \dot{x}(t)\right) dt$. In the classical calculus of variations the fundamental necessary condition for an arbitrary variation to vanish is given by the Euler-Lagrange equation $\frac{d}{dt} f_{\dot{x}}\left(t, x(t), \dot{x}(t)\right) = f_x\left(t, x(t), \dot{x}(t)\right)$ (E-L). In the simplest problem of the calculus of variations the endpoints $x(a)$ and $x(b)$ are assumed fixed. If there is sufficient smoothness, then (E-L) is a second order ordinary differential equation. In the simplest problem the fixed endpoints serve to determine the two constants of integration expected in the solution of (E-L). When one or both of the endpoints are free there is a need for more necessary conditions. The classical theory derives the so-called natural boundary conditions to cover these particular cases. To deal with isoperimetric constraints of the form $G(x) = \int g\left(t, x(t), \dot{x}(t)\right) dt = C$, the classical theory establishes some version of the Lagrange multiplier theorem. Superficially it suffices to replace f by $f + \lambda g$ in (E-L), and then use the constraint as an additional condition to account for the unknown real number λ. The hope is to solve (E-L) and get a one-parameter family of solutions x_λ and then only consider λ's such that $G(x_\lambda) = C$.

1.3 Steepest descent. It is shown below how the method of steepest descent, by following the trajectory in the negative gradient direction, may be used to solve the general isoperimetric problem. The method applies equally well to the case of fixed or free endpoints. Any number of isoperimetric constraints is allowed as long as the number is finite. The method is generally applicable in that it suffices to give the data of the problem, such as partial derivatives $\left(f_x, f_{\dot{x}}, g_x, g_{\dot{x}}\right)$ as well as an initial function x with derivative $\dot{x}$. Once $x, \dot{x}$ are given, the formulas developed involve nothing more than integrals and algebraic operations. This is an important advantage when a numerical algorithm is devised since numerical integration is better behaved than numerical differentiation. The domain of the unconstrained functions is the set of absolutely continuous functions with the square of the derivative Lebesgue-integrable. Given an inner product $\langle , \rangle$, the gradient and the directional derivative of a functional F are paired dually by $\langle \nabla F(x), v \rangle = DF(x)v$, where $\nabla F(x)$ is in the primal space and $DF(x)$ is in the dual. The gradient ∇F depends on the choice of metric $\langle , \rangle$.

1.4 Gradients without a basis. In finite dimensions the gradient is given as the column matrix of the first partial derivatives. Meanwhile, the row matrix represents the directional derivative, which is the transpose of the gradient. In an infinite dimensional separable Hilbert spaces this kind of relationship is available as soon as a basis is chosen. If $\{e_i\}_{i=1}^{\infty}$ is an orthonormal basis, then $DF(x)e_i$ is the 'partial' derivative. The use of a basis to carry out the calculation of a gradient is sometimes a necessity; see [**4**]. The integral and derivative operations offer opportunities, not available in the finite dimensional case, to define functionals without the use of a basis. It is exciting to report that although the gradients depend on the choice of metric, there is no need for a basis to determine the gradients of the functionals considered in this paper. Introduce the quantities

$$E_x^f(t) = f_{\dot{x}}\left(t, x(t), \dot{x}(t)\right) - \int_a^t f_x\left(s, x(s), \dot{x}(s)\right) ds \text{, and } W_x^f = \int_a^b f_x\left(t, x(t), \dot{x}(t)\right) dt \, .$$

In Theorem 1 it is shown that

$$\nabla_I F(x) = \int_a^t E_x^f(s) ds + \left(W_x^f - \beta \int_a^b E_x^f(t) dt \right) \frac{\alpha(t-a)+1}{\alpha\beta(b-a)+\alpha+\beta},$$

$$\nabla_{II} F(x) = \int_a^t E_x^f(s) \cosh(t-s) ds + \left(W_x^f - \int_a^b E_x^f(s) \sinh(b-s) ds \right) \frac{\cosh(t-a)}{\sinh(b-a)}, \text{ and}$$

$$\nabla_{III} F(x) = \begin{cases} \displaystyle\int_p^t E_x^f(s) ds + W_x^f, & a \le t \le p \\ \displaystyle\int_p^t E_x^f(s) ds + W_x^f\,(t-p+1) & p \le t \le b \end{cases}.$$

The statement 'vanishing gradient' is independent of the metric in the sense that if the gradient is zero in one metric, then the gradient is zero in all metrics. It is shown in Theorem 2 that the equation $\nabla F(x) = 0$ encapsulates both (E-L) and the natural boundary conditions. When both the endpoints are fixed, the gradients are given in Theorem 3.

1.5 Gradients in the space of derivatives. Integration behaves better than differentiation numerically. The function x is in general only known explicitly at the start of the descent. At subsequent flow-times only a discrete representation of x is available, and this causes numerical problems when $\dot{x}$ is computed. To circumvent these difficulties it is shown how to introduce a scalar field on the space of derivatives and avoid a numerical approximation of $\dot{x}$. In this case the gradient flow takes place in the space of derivatives. There is a natural lift of the flow to the space of functions. The relationship between the gradient in the space of derivatives and the lifted gradient is analyzed in Section 5. An important insight is that in general the gradient operator and the derivative operator fail to commute. In fact, essentially only the metric $\langle , \rangle_{III}$ commutes with the projection. Incidentally, this metric with $p = a$ is indeed the same metric used in the traditional curve-straightening flow, and it is known to destroy symmetries of reflection; see [**5**]. It is a natural idea to change the metric in order to combat this behavior. The analysis of the related problem also sheds light on the source of the hyperbolic functions in the gradient with respect to the metric $\langle,\rangle_{II}$.

1.6 Isoperimetric constraints and the values of the multipliers. To illustrate how to deal with isoperimetric constraints it is shown in Theorem 4 that in the case of $\langle,\rangle_I$, with $\alpha = 1, \beta = 0$, and $a = 0, b = 1$, it must be that

$$\lambda(x) = \frac{\left(\int_0^1 E_x^f(t)dt + 2W_x^f\right)W_x^g + \int_0^1 E_x^g(t)\left(E_x^f(t) + W_x^f\right)dt}{\left(\int_0^1 E_x^g(t)dt + 2W_x^g\right)W_x^g + \int_0^1 E_x^g(t)\left(E_x^g(t) + W_x^g\right)dt}.$$

The projected gradient used in the steepest descent is given by $\nabla_I^\pi F(x) = \nabla_I F(x) - \lambda \nabla_I G(x)$. In the formula it is not assumed that $\nabla_I^\pi F(x) = 0$, but if this is the case, then $\lambda(x)$ agrees with the classical Lagrange multiplier. Similar formulas hold for other metrics.

1.7 Simplifications and asymmetries. The derivations of the various gradients follow a fixed pattern, which is applicable to other metrics. As mentioned above, the geometric properties of functions along the gradient trajectories are in general affected by a change in the metric. As in the case of the curve-straightening flow, it can be that certain symmetries, which are not preserved by the flow using $\langle,\rangle_I$ with $\alpha = 1$ and $\beta = 0$, are preserved with a different choice of metric. A part from being 'natural', in the sense that the gradient operator and the derivative projection commute, the choice of $\alpha = 1$ and $\beta = 0$ does have other advantages. For instance, the gradient is given by the comparatively simple formula

$$\nabla_I F(x) = \int_a^t E_x^f(s)ds + W_x^f(t+1).$$

There is a very important simplification when there is no explicit dependence on x in f so that $f_x = 0$. In this case $E_x^f(t) = f_{\dot{x}}(t, x(t), \dot{x}(t))$ and $W_x^f = 0$. Note also that in this case the 'discontinuity' disappears in the metric $\langle,\rangle_{III}$. The issues of 'space' symmetries in the context of the curve-straightening flow are examined in detail in the companion paper 'Symmetrized Curve-straightening'; see [6]. The issues of 'external symmetries' are also interesting. In Remark 2 following Theorem 4 it is illustrated how the 'symbolic form' of the gradient is not 'preserved' when the roles of the endpoints are interchanged. As a result it is sometimes possible to eliminate the quantity W_x^f from the gradient formula by changing the metric.

2. The Unconstrained Case

2.1 Definitions and the gradient. Let X be the space of absolutely continuous functions $x:[a,b] \to R$. Let $\dot{x}$ denote the derivative, which exists almost everywhere, and define the vector space $H = \left\{ x \in X \,\middle|\, \dot{x} \in L^2[a,b] \right\}$. Let $w \in H$ and $v \in H$. Three examples of inner products are considered. The first inner product is given by

$$\langle w,v\rangle_I = \alpha w(a)v(a) + \beta w(b)v(b) + \int_a^b \dot{w}(t)\dot{v}(t)dt ,$$

where $\alpha \geq 0$, $\beta \geq 0$ and $\alpha + \beta > 0$. The second inner product is given by

$$\langle w,v\rangle_{II} = \int_a^b w(t)v(t)dt + \int_a^b \dot{w}(t)\dot{v}(t)dt .$$

As a comparison, note that when $\alpha = \beta = (b-a)/2$, one gets the trapezoid approximation to $\int_a^b w(t)v(t)dt$.

Finally, the third inner product is given by

$$\langle w,v\rangle_{III} = w(p)v(p) + \int_a^b \dot{w}(t)\dot{v}(t)dt ,$$

where it is assumed that $a \leq p \leq b$. Note that if $p = a$ or $p = b$, then $\langle,\rangle_{III}$ is subsumed by $\langle,\rangle_I$. The cases $a < p < b$ are not standard. Given any continuously differentiable functional $F : H \to \mathbb{R}$, the directional derivative $DF(x) : H \to \mathbb{R}$ and the gradient $\nabla F(x) \in H$ satisfy $\langle \nabla F(x), v\rangle_H = DF(x)v$. Here $\langle,\rangle_H$ is the inner product in H. To save space write ch for the hyperbolic cosine cosh and sh for sinh .

Theorem 1. *Let $f : \mathbb{R}^3 \to \mathbb{R}$ be continuously differentiable. Denote the partial derivatives with respect to the second and third argument by f_x, and $f_{\dot{x}}$. Consider only functions $x \in H$ such that $f(t,x(t),\dot{x}(t))$, $f_x(t,x(t),\dot{x}(t))$, $f_{\dot{x}}(t,x(t),\dot{x}(t)) \in L^1[a,b]$. Put*

$$E_x^f(t) = f_{\dot{x}}(t,x(t),\dot{x}(t)) - \int_a^t f_x(s,x(s),\dot{x}(s))\,ds , \text{ and } W_x^f = \int_a^b f_x(t,x(t),\dot{x}(t))\,dt .$$

The following is true for $F(x) = \int_a^b f(t,x(t),\dot{x}(t))\,dt$

(I) The gradient with respect to

$$\langle w,v\rangle_I = \alpha w(a)v(a) + \beta w(b)v(b) + \int_a^b \dot{w}(t)\dot{v}(t)dt$$

is given by

$$\nabla_I F(x) = \int_a^t E_x^f(s)ds + \left(W_x^f - \beta \int_a^b E_x^f(s)ds \right) \frac{\alpha(t-a)+1}{\alpha\beta(b-a)+\alpha+\beta} .$$

(II) The gradient with respect to

$$\langle w,v\rangle_{II} = \int_a^b w(t)v(t)dt + \int_a^b \dot{w}(t)\dot{v}(t)dt$$

is given by

$$\nabla_{II} F(x) = \int_a^t E_x^f(s)\,\mathrm{ch}(t-s)ds + \left(W_x^f - \int_a^b E_x^f(s)\,\mathrm{sh}(b-s)ds \right) \frac{\mathrm{ch}(t-a)}{\mathrm{sh}(b-a)} .$$

(III) The gradient with respect to

$$\langle w,v\rangle_{III} = w(p)v(p) + \int_a^b \dot{w}(t)\dot{v}(t)dt$$

is given by

$$\nabla_{III}F(x) = \begin{cases} \int_p^t E_x^f(s)ds + W_x^f, & a \le t \le p \\ \int_p^t E_x^f(s)ds + W_x^f\,(t-p+1) & p \le t \le b \end{cases}.$$

Proof. It suffices to verify the relation $\langle \nabla F(x), v\rangle = DF(x)v$ in the three cases. To be brief the most difficult case is illustrated here. The other two are left to the reader. First, a standard computation using integration by parts shows that

$$DF(x)v = \int_a^b \big(f_x\,(t,x(t),\dot{x}(t))\,v(t) + f_{\dot{x}}\,(t,x(t),\dot{x}(t))\,\dot{v}(t)\big)dt = v(b)\int_a^b f_x\,(t,x(t),\dot{x}(t))\,dt +$$
$$\int_a^b \left(f_{\dot{x}}\,(t,x(t),\dot{x}(t)) - \int_a^t f_x\,(s,x(s),\dot{x}(s))\,ds \right)\dot{v}(t)dt = W_x^f v(b) + \int_a^b E_x^f(t)\dot{v}(t)dt.$$

When the metric is $\langle,\rangle_{II}$, the first main term of the second integral is rewritten with the help of Leibniz' rule

$$\big\langle \nabla_{II}F(x),v\big\rangle_{II} = \int_a^b \left(\int_a^t E_x^f(s)\,\mathrm{ch}(t-s)ds + \left(W_x^f - \int_a^b E_x^f(s)\,\mathrm{sh}(b-s)ds \right)\frac{\mathrm{ch}(t-a)}{\mathrm{sh}(b-a)} \right) v(t)dt +$$
$$\int_a^b \left(\left(E_x^f(t) + \int_a^t E_x^f(s)\,\mathrm{sh}(t-s)ds \right) + \left(W_x^f - \int_a^b E_x^f(s)\,\mathrm{sh}(b-s)ds \right)\frac{\mathrm{sh}(t-a)}{\mathrm{sh}(b-a)} \right)\dot{v}(t)dt.$$

Integrate the second main term of the second integral by parts

$$\int_a^b \left(W_x^f - \int_a^b E_x^f(s)\,\mathrm{sh}(b-s)ds \right)\frac{\mathrm{sh}(t-a)}{\mathrm{sh}(b-a)}\,dv(t) =$$
$$\left(W_x^f - \int_a^b E_x^f(s)\,\mathrm{sh}(b-s)ds \right)v(b) - \int_a^b \left(W_x^f - \int_a^b E_x^f(s)\,\mathrm{sh}(b-s)ds \right)\frac{\mathrm{ch}(t-a)}{\mathrm{sh}(b-a)}\,v(t)dt.$$

Also integrate the first main term of the first integral by parts again using Leibniz' rule

$$\int_a^b \left(\int_a^t E_x^f(s)\,\mathrm{ch}(t-s)ds \right) v(t)dt = \int_a^b v(t)d\left(\int_a^t E_x^f(s)\,\mathrm{sh}(t-s)ds \right) =$$
$$v(b)\int_a^b E_x^f(s)\,\mathrm{sh}(b-s)ds - \int_a^b \left(\int_a^t E_x^f(s)\,\mathrm{sh}(t-s)ds \right)\dot{v}(t)dt$$

Once both of these computations are combined, it follows that

$$\big\langle \nabla_{II}F(x),v\big\rangle_{II} = W_x^f v(b) + \int_a^b E_x^f(t)\dot{v}(t)dt = DF(x)v\,.$$

Remark 1. The quantity E_x^f introduced here is the negative of the anti-derivative of the so-called

Euler operator; see page 16 in [1]. In Mechanics the quantity W_x^f is the work. It is very difficult to keep the calculations manageable without the help of these two quantities. The author repeatedly fell in to the trap of overzealous simplifications, which then hides the cleaner structure put forth in the theorem.

Remark 2. The inner product

$$\langle w,v\rangle = w(a)v(a) + \int_a^b \dot{w}(t)\dot{v}(t)dt\,,$$

used in [2], [3], [4] and [5], is a standard Sobolev inner product. Note that $\langle,\rangle = \langle,\rangle_I$ with $\alpha = 1$ and $\beta = 0$, leads to the much simpler formula

$$\nabla_I F(x) = \int_a^t E_x^f(s)ds + W_x^f\,(t-a+1)\,.$$

If the integrand f of the functional F has no explicit dependence on x so that $f_x = 0$, then

$$\nabla_I F(x) = \int_a^t E_x^f(s)ds = \int_a^t f_{\dot{x}}\,(s,x(s),\dot{x}(s))\,ds\,,$$

since $W_x^f = 0$ in this case. Moreover, $\nabla_I F(x)\big|_{t=a} = 0$ so the initial point $x(a)$ is kept fixed.

Remark 3. In general the endpoints are not expected to remain fixed under a positive or negative gradient flow. For instance, if $\alpha = 1$ and $\beta = 0$, then $\nabla_I F(x)\big|_{t=a} = W_x^f$.

Remark 4. The deductive proof of Theorem 1 is of limited use since it gives little indication as to how the gradients were determined in the first place. To explain the techniques behind the computation of the gradients the derivations of $\nabla_{II} F(x)$ is given with all the details.

2.2 Derivation of the gradients. The proof of Theorem 1 uses an adaptation of a result known as the duBois-Reymond lemma. This Lemma is used in several other places in this paper.

Lemma (duBois-Reymond). *Suppose* $g \in L^2[a,b]$ *and* $\int_a^b g(t)\dot{v}(t)dt = 0$ *for all* $v \in H$ *such that* $v(a) = v(b) = 0$ *, then there exists a constant* C *such that* $\| g(t) - C \|_{L^2} = 0$ *.*

Proof. Given g , define

$$C = \frac{1}{b-a}\int_a^b g(u)du\,.$$

Put

$$v(t) = \int_a^t (\,g(s) - C\,)\,ds\,.$$

It follows that $v \in H$ as well as $v(a) = v(b) = 0$ and $\dot{v}(t) = g(t) - C$. See Theorem 7.20 page 148 of for this particular direction of the fundamental theorem in the context of absolutely continuous function. From this conclude that

$$\int_a^b (\,\dot{v}(t) + C\,)\dot{v}(t)dt = 0\,.$$

This in turn implies that

$$\int_a^b \dot{v}^2(t)dt = 0\,,$$

and hence $\| g(t) - C \|_{L^2} = 0$.

Remark 1. As usual, the notation $g(t) = C$ is used in the sense of L^2 (even on subintervals of $[a,b]$ when needed.)

Remark 2. There are other versions of the duBois-Reymond's available. For $g \in L^1$ and v smooth see page 15 in [1]. The present version is specifically adapted to the framework in place here. Interestingly, the proofs have very little in common.

Next look at the derivation in the case $\langle , \rangle_{II}$, which is by far the most complicated case. First

$$\langle w, v\rangle_{II} = \int_a^b w(t)v(t)dt + \int_a^b \dot{w}(t)\dot{v}(t)dt = v(b)\int_a^b w(t)dt + \int_a^b \left(\dot{w}(t) - \int_a^t w(s)ds\right)\dot{v}(t)dt\,.$$

When $w = \nabla_{II}F(x)$, then $\left\langle \nabla_{II}F(x), v\right\rangle_{II} = DF(x)v$ leads to

$$\left(\int_a^b w(t)dt - W_x^f\right)v(b) + \int_a^b \left(\left(\dot{w}(t) - \int_a^t w(s)ds\right) - E_x^f(t)\right)\dot{v}(t)dt = 0\,.$$

Since this is true for all $v \in H$, it is true for all v's such that $v(a) = v(b) = 0$. Apply the Lemma and conclude that there is a constant C such that

$$\dot{w}(t) - \int_a^t w(s)ds = E_x^f(t) + C$$

Next choose $v(t) = 1$ to show that $\int_a^b w(t)dt = W_x^f$, and then $v(t) = t$ to see that $C = 0$. Let

$$W(t) = \int_a^t w(s)ds$$

so that $W(a) = 0$ and $W(b) = W_x^f$. The transformed equation is $\ddot{W}(t) - W(t) = E_x^f(t)$. This is a non-homogeneous linear equation. The solutions have the form $W(t) = W_h(t) + W_p(t)$ where $W_h(t) = A\,\mathrm{ch}(t-a) + B\,\mathrm{sh}(t-a)$ is the general solution of the homogeneous problem. To find a particular solution W_p, use the method of variations of parameters. To this end let $W_p(t) = A(t)\,\mathrm{ch}(t-a) + B(t)\,\mathrm{sh}(t-a)$, so that

$$\dot{W}_p(t) = \dot{A}(t)\,\mathrm{ch}(t-a) + A(t)\,\mathrm{sh}(t-a) + \dot{B}(t)\,\mathrm{sh}(t-a) + B(t)\,\mathrm{ch}(t-a)\,.$$

Impose the condition $\dot{A}(t)\,\mathrm{ch}(t-a) + \dot{B}(t)\,\mathrm{sh}(t-a) = 0$, so that

$$\ddot{W}_p(t) = \dot{A}(t)\,\mathrm{sh}(t-a) + A(t)\,\mathrm{ch}(t-a) + \dot{B}(t)\,\mathrm{ch}(t-a) + B(t)\,\mathrm{sh}(t-a)\,.$$

This leads to the system

$$\begin{bmatrix} \mathrm{ch}(t-a) & \mathrm{sh}(t-a) \\ \mathrm{sh}(t-a) & \mathrm{ch}(t-a) \end{bmatrix}\begin{bmatrix} \dot{A}(t) \\ \dot{B}(t) \end{bmatrix} = \begin{bmatrix} 0 \\ E_x^f(t) \end{bmatrix},$$

with solution

$$\begin{bmatrix} \dot{A}(t) \\ \dot{B}(t) \end{bmatrix} = \begin{bmatrix} \operatorname{ch}(t-a) & -\operatorname{sh}(t-a) \\ -\operatorname{sh}(t-a) & \operatorname{ch}(t-a) \end{bmatrix} \begin{bmatrix} 0 \\ E_x^f(t) \end{bmatrix} = \begin{bmatrix} -E_x^f(t)\operatorname{sh}(t) \\ E_x^f(t)\operatorname{ch}(t) \end{bmatrix}.$$

A particular solution is given by

$$\begin{aligned} W_p(t) &= -\operatorname{ch}(t-a)\int_a^t E_x^f(s)\operatorname{sh}(s-a)ds + \operatorname{sh}(t-a)\int_a^t E_x^f(s)\operatorname{ch}(s-a)ds \\ &= \int_a^t E_x^f(s)\operatorname{sh}(t-s)ds \end{aligned}.$$

This particular solution combined with the general homogeneous solution yields

$$W(t) = A\operatorname{ch}(t-a) + B\operatorname{sh}(t-a) + \int_a^t E_x^f(s)\operatorname{sh}(t-s)ds\,.$$

Since $W(a)=0$ it follows that $A=0$, and $W(t) = B\operatorname{sh}(t-a) + \int_a^t E_x^f(s)\operatorname{sh}(t-s)ds$. At the other endpoint $W(b) = B\operatorname{sh}(b-a) + \int_a^b E_x^f(s)\operatorname{sh}(b-s)ds = W_x^f$.

Finally, Leibniz' rule yields

$$\nabla_{II}F(x) = w(t) = \dot{W}(t) = B\operatorname{ch}(t-a) + \int_a^t E_x^f(s)\operatorname{ch}(t-s)ds\,.$$

Hence,

$$\nabla_{II}F(x) = \int_a^t E_x^f(s)\operatorname{ch}(t-s)ds + \left(W_x^f - \int_a^b E_x^f(s)\operatorname{sh}(b-s)ds\right)\frac{\operatorname{ch}(t-a)}{\operatorname{sh}(b-a)}.$$

2.3 The Calculus of Variations. The purpose of the next Theorem is to explain how the Euler-Lagrange equation and the 'natural' boundary conditions appear in the context of gradients. First note that if the gradient is zero at a point with respect to one metric, then it is zero with respect to all metrics because

$$\begin{aligned} &\nabla_I F(\hat{x}) = 0 \Leftrightarrow \left\langle \nabla_I F(\hat{x}), v \right\rangle_I = 0 \ \forall v \in H \Leftrightarrow \\ &\quad DF(\hat{x})v = 0 \ \forall v \in H \Leftrightarrow \left\langle \nabla_{II} F(\hat{x}), v \right\rangle_{II} = 0 \ \forall v \in H \Leftrightarrow \nabla_{II} F(\hat{x}) = 0 \end{aligned}$$

A function where the gradient is zero is called a critical point. In order to compare with things with the classical setup it is assumed that the solution is C^2. The proof of the following theorem illustrates how it is possible for some properties to be readily established using one particular metric, and yet the same properties may well be obscured in a different metric.

Theorem 2. *If $\hat{x} \in C^2$ is a critical point, then $E_{\hat{x}}^f(t) = 0$ for all $t \in [a,b]$, and $W_{\hat{x}}^f = 0$. The statement*

$$\frac{d}{dt}E_{\hat{x}}^f(t) = 0\,,$$

with equality everywhere, is the Euler-Lagrange equation. The statements $E_{\hat{x}}^f(a) = E_{\hat{x}}^f(b) = 0$, correspond to the natural boundary conditions

$$f_{\dot{\hat{x}}}\left(a,\hat{x}(a),\dot{\hat{x}}(a)\right) = f_{\dot{\hat{x}}}\left(b,\hat{x}(b),\dot{\hat{x}}(b)\right) = 0\,.$$

Proof. Consider the gradient formula in the metric $\langle,\rangle_{III}$. Set $t = p$ in the formula for the gradient and conclude that $W^f_{\hat{x}} = 0$. The derivative of the gradient formula reveals that $E^f_{\hat{x}}(t) = 0$ for all $t \in [a,b]$. Next note that $f_{\dot{\hat{x}}}\left(a,\hat{x}(a),\dot{\hat{x}}(a)\right) = E^f_{\hat{x}}(a) = 0$. Finally, since

$$\frac{d}{dt}E^f_{\hat{x}}(t) = 0$$

implies

$$\frac{d}{dt}f_{\dot{x}}\left(t,\hat{x}(t),\dot{\hat{x}}(t)\right) = f_x\left(t,\hat{x}(t),\dot{\hat{x}}(t)\right),$$

it is also true that

$$f_{\dot{x}}\left(b,\hat{x}(b),\dot{\hat{x}}(b)\right) - f_{\dot{x}}\left(a,\hat{x}(a),\dot{\hat{x}}(a)\right) = \int_a^b f_x\left(t,\hat{x}(t),\dot{\hat{x}}(t)\right)dt = W^f_{\hat{x}} = 0\,,$$

and hence $f_{\dot{x}}\left(b,\hat{x}(b),\dot{\hat{x}}(b)\right) = 0$.

Remark 1: The statement $\nabla F(\hat{x}) = 0$ incorporates the Euler-Lagrange equation as well as the necessary conditions given by the natural boundary conditions. It follows that the equation $\nabla F(\hat{x}) = 0$ in general is more selective than the Euler-Lagrange equation.

Remark 2: It is straightforward to prove the Theorem 2 using the gradient formula of the metric $\langle,\rangle_I$. It suffices to evaluate the gradient at the two endpoints. In the case of the metric $\langle,\rangle_{II}$ things are more complicated.

3. Constrained Endpoints

3.1 The 'evaluation at a point' functional.

Write ct for coth.

Theorem 3. *Let $\tau \in [a,b]$ and $x_\tau \in \mathbb{R}$ be given. Define $\phi_\tau : H \to \mathbb{R}$ by $\phi_\tau(x) = x(\tau) - x_\tau$, then*

$$\nabla_I\phi_\tau(x) = \begin{cases} \dfrac{\beta(b-\tau)+1}{\alpha\beta(b-a)+\alpha+\beta}(\alpha(t-a)+1) & a \le t \le \tau \\ \dfrac{\alpha(\tau-a)+1}{\alpha\beta(b-a)+\alpha+\beta}(\beta(b-t)+1) & \tau \le t \le b \end{cases},$$

$$\nabla_{II}\phi_\tau(x) = \begin{cases} (\mathrm{ct}(b-a)\,\mathrm{ch}(\tau-a) - \mathrm{sh}(\tau-a))\,\mathrm{ch}(t-a) & a \le t \le \tau \\ \mathrm{ch}(\tau-a)(\mathrm{ch}(b-a)\,\mathrm{ch}(t-a) - \mathrm{sh}(t-a)) & \tau \le t \le b \end{cases},$$

and

$$\nabla_{III}\phi_\tau(x) = \begin{cases} 1+p-\tau & a \le t \le \tau \\ 1+p-t & \tau \le t \le p \\ 1 & p \le t \le b \end{cases}, \qquad \nabla_{III}\phi_\tau(x) = \begin{cases} 1 & a \le t \le p \\ 1+t-p & p \le t \le \tau \\ 1+\tau-p & \tau \le t \le b. \end{cases}$$

Proof. The technique is illustrated by the most difficult case. Suppose the metric is $\langle,\rangle_{II}$, then

$$\left\langle \nabla_{II}\phi_\tau(x), v\right\rangle_{II} = \int_a^\tau (\operatorname{ct}(b-a)\operatorname{ch}(\tau-a)-\operatorname{sh}(\tau-a))\operatorname{ch}(t-a)v(t)dt + \int_\tau^b \operatorname{ch}(\tau-a)(\operatorname{ct}(b-a)\operatorname{ch}(t-a)-\operatorname{sh}(t-a))v(t)dt + \int_a^\tau (\operatorname{ct}(b-a)\operatorname{ch}(\tau-a)-\operatorname{sh}(\tau-a))\operatorname{sh}(t-a)\dot v(t)dt + \int_\tau^b \operatorname{ch}(\tau-a)(\operatorname{ct}(b-a)\operatorname{sh}(t-a)-\operatorname{ch}(t-a))\dot v(t)dt$$

Integrate the last two terms by parts

$$\int_a^\tau (\operatorname{ct}(b-a)\operatorname{ch}(\tau-a)-\operatorname{sh}(\tau-a))\operatorname{sh}(t-a)dv = (\operatorname{ct}(b-a)\operatorname{ch}(\tau-a)-\operatorname{sh}(\tau-a))\operatorname{sh}(\tau-a)v(\tau) - \int_a^\tau (\operatorname{ct}(b-a)\operatorname{ch}(\tau-a)-\operatorname{sh}(\tau-a))\operatorname{ch}(t-a)v(t)dt$$

and

$$\int_\tau^b \operatorname{ch}(\tau-a)(\operatorname{ct}(b-a)\operatorname{sh}(t-a)-\operatorname{ch}(t-a))dv = -\operatorname{ch}(\tau-a)(\operatorname{ct}(b-a)\operatorname{sh}(\tau-a)-\operatorname{ch}(\tau-a))v(\tau) - \int_\tau^b \operatorname{ch}(\tau-a)(\operatorname{ct}(b-a)\operatorname{ch}(t-a)-\operatorname{sh}(t-a))v(t)dt$$

This leads to

$$\left\langle \nabla_{II}\phi_\tau(x), v\right\rangle_{II} = (\operatorname{ct}(b-a)\operatorname{ch}(\tau-a)-\operatorname{sh}(\tau-a))\operatorname{sh}(\tau-a)v(\tau) - \operatorname{ch}(\tau-a)(\operatorname{ct}(b-a)\operatorname{sh}(\tau-a)-\operatorname{ch}(\tau-a))v(\tau) = v(\tau) = D\phi_\tau(x)v$$

Remark 1. Note that the gradient vector fields do not depend on x so these are constant vector fields on H.

Remark 2. If $\tau = a$, then

$$\nabla_I\phi_a(x) = \frac{\beta(b-t)+1}{\alpha\beta(b-a)+\alpha+\beta},$$

$$\nabla_{II}\phi_a(x) = \operatorname{ct}(b-a)\operatorname{ch}(t-a)-\operatorname{sh}(t-a), \text{ and}$$

$$\nabla_{III}\phi_a(x) = \begin{cases} 1+p-t & a \le t \le p \\ 1 & p \le t \le b \end{cases}.$$

If $\tau = b$, then

$$\nabla_I\phi_b(x) = \frac{\alpha(t-a)+1}{\alpha\beta(b-a)+\alpha+\beta},$$

$$\nabla_{II}\phi_b(x) = \frac{\operatorname{ch}(t-a)}{\operatorname{sh}(b-a)}, \text{ and}$$

$$\nabla_{III}\phi_b(x) = \begin{cases} 1 & a \le t \le p \\ 1+t-p & p \le t \le b \end{cases}.$$

All of these cases yield constant vector fields. When the metric is $\langle,\rangle_I$ or $\langle,\rangle_{II}$, the vectors in the vector field are smooth (C^∞) elements in H . Only if $p=a$ or $p=b$ is the same true in the metric $\langle,\rangle_{III}$.

3.2 Derivations. Consider the derivation when the metric is $\langle,\rangle_{II}$ and let $w=\nabla_{II}\phi_\tau(x)$. The following holds for all $v\in H$

$$v(b)\int_a^b w(t)dt+\int_a^b\left(\dot w(t)-\int_a^t w(s)ds\right)\dot v(t)dt=v(\tau)\,.$$

Let v be a nonzero constant function and conclude that

$$\int_a^b w(t)dt=1\,.$$

For v 's such that $\dot v(t)=0$ for all $t\in[\tau,b]$ and with $v(\tau)=v(b)=0$ it is true that

$$\int_a^\tau\left(\dot w(t)-\int_a^t w(s)ds\right)\dot v(t)dt=0\,.$$

An application of the duBois-Reymond lemma gives a constant $C_a^\tau\in\mathbb{R}$ such that

$$\dot w(t)-\int_a^t w(s)ds=C_a^\tau \text{ for } t\in[a,\tau]\,.$$

From this deduce that

$$v(b)+C_a^\tau v(\tau)-C_a^\tau v(a)+\int_\tau^b\left(\dot w(t)-\int_a^t w(s)ds\right)\dot v(t)dt=v(\tau)\,.$$

If $v(a)=v(\tau)=v(b)=0$, then

$$\int_\tau^b\left(\dot w(t)-\int_a^t w(s)ds\right)\dot v(t)dt=0\,,$$

and this time

$$\dot w(t)-\int_a^t w(s)ds=C_\tau^b \text{ for } t\in[\tau,b]$$

so that $v(b)+C_a^\tau v(\tau)-C_a^\tau v(a)+C_\tau^b v(b)-C_\tau^b v(\tau)=v(\tau)$. Choose $v(t)=(\tau-t)(b-t)$ and conclude that either $C_a^\tau=0$ or $\tau=a$. Choose $v(t)=(\tau-t)(t-a)$ and conclude that either $C_\tau^b=-1$ or $\tau=b$. The cases $\tau=a$ and $\tau=b$ follow from the generic case and are ignored here. Thus,

$$\dot w(t)-\int_a^t w(s)ds=\begin{cases}0 & a\le t<\tau\\ -1 & \tau<t\le b\end{cases}.$$

Let $W(t)=\int_a^t w(s)ds$ so that $W(a)=0$ and $W(b)=1$. The equations turn into

$$\ddot W(t)-W(t)=\begin{cases}0 & a\le t<\tau\\ -1 & \tau<t\le b\end{cases},$$

with solution

$$W(t)=\begin{cases} A_a^\tau \operatorname{ch}(t-a)+B_a^\tau \operatorname{sh}(t-a) & a\le t<\tau \\ A_\tau^b \operatorname{ch}(t-a)+B_\tau^b \operatorname{sh}(t-a)+1 & \tau<t\le b \end{cases},$$

such that $W(a)=A_a^\tau=0$ and $W(b)=A_\tau^b \operatorname{ch}(b-a)+B_\tau^b \operatorname{sh}(b-a)+1=1$.
Smoothness implies $B_a^\tau \operatorname{sh}(\tau-a)=A_\tau^b \operatorname{ch}(\tau-a)+B_\tau^b \operatorname{sh}(\tau-a)+1$ and $B_a^\tau \operatorname{ch}(\tau-a)=A_\tau^b \operatorname{sh}(\tau-a)+B_\tau^b \operatorname{ch}(\tau-a)$. In matrix form this is given by

$$\begin{bmatrix} \operatorname{sh}(\tau-a) & -\operatorname{ch}(\tau-a) & -\operatorname{sh}(\tau-a) \\ \operatorname{ch}(\tau-a) & -\operatorname{sh}(\tau-a) & -\operatorname{ch}(\tau-a) \\ 0 & \operatorname{ch}(b-a) & \operatorname{sh}(b-a) \end{bmatrix}\begin{bmatrix} B_a^\tau \\ A_\tau^b \\ B_\tau^b \end{bmatrix}=\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \text{ or}$$

$$\begin{bmatrix} sh(\tau-a)ch(\tau-a) & -ch^2(\tau-a) & -sh(\tau-a)ch(\tau-a) \\ sh(\tau-a)ch(\tau-a) & -sh^2(\tau-a) & -sh(\tau-a)ch(\tau-a) \\ 0 & ch(b-a) & sh(b-a) \end{bmatrix}\begin{bmatrix} B_a^\tau \\ A_\tau^b \\ B_\tau^b \end{bmatrix}=\begin{bmatrix} ch(\tau-a) \\ 0 \\ 0 \end{bmatrix}, \text{ or}$$

$$\begin{bmatrix} 1 & -\operatorname{ct}(\tau-a) & -1 \\ 0 & 1 & 0 \\ 0 & \operatorname{ch}(b-a) & \operatorname{sh}(b-a) \end{bmatrix}\begin{bmatrix} B_a^\tau \\ A_\tau^b \\ B_\tau^b \end{bmatrix}=\begin{bmatrix} \operatorname{csch}(\tau-a) \\ -\operatorname{ch}(\tau-a) \\ 0 \end{bmatrix}.$$

It follows that $B_a^\tau=\operatorname{ch}(\tau-a)\operatorname{ct}(b-a)-\operatorname{sh}(\tau-a)$, and $A_\tau^b=-\cosh(\tau-a)$ $B_\tau^b=\operatorname{ch}(\tau-a)\operatorname{ct}(b-a)$ so that

$$W(t)=\begin{cases} (\operatorname{ch}(\tau-a)\operatorname{ct}(b-a)-\operatorname{sh}(\tau-a))\operatorname{sh}(t-a) & a\le t\le\tau \\ -\operatorname{ch}(\tau-a)\operatorname{ch}(t-a)+\operatorname{ch}(\tau-a)\operatorname{ct}(b-a)\operatorname{sh}(t-a)+1 & \tau\le t\le b \end{cases}.$$

Finally,

$$\begin{aligned} \nabla_{II}\phi_\tau(x) &= w(t)=\dot W(t) \\ &= \begin{cases} (\operatorname{ch}(\tau-a)\operatorname{ct}(b-a)-\operatorname{sh}(\tau-a))\operatorname{ch}(t-a) & a\le t\le\tau\,. \\ -\operatorname{ch}(\tau-a)\operatorname{sh}(t-a)+\operatorname{ch}(\tau-a)\operatorname{ct}(b-a)\operatorname{ch}(t-a) & \tau\le t\le b \end{cases} \end{aligned}$$

3.3 Projection onto the space of functions with fixed endpoints. Let ∇^π denote the projected gradient onto the affine subspace $\left\{x\in H \,\middle|\, x(a)=x_a,\, x(b)=x_b\right\}$, where H is equipped with the metric $\langle,\rangle_I$, $\langle,\rangle_{II}$ and $\langle,\rangle_{III}$ respectively.

Theorem 4. *Suppose*

$$F(x)=\int_a^b f(t,x(t),\dot x(t))\,dt\,,\ \textit{and}\ \ E_x^f(t)=f_{\dot x}(t,x(t),\dot x(t))-\int_a^t f_x(s,x(s),\dot x(s))\,ds\,.$$

The projected gradients are given by

$$\nabla_I^\pi F(x)=\int_a^t E_x^f(s)ds-\int_a^b E_x^f(s)ds\frac{t-a}{b-a}\,,$$

$$\nabla_{II}^\pi F(x)=\int_a^t E_x^f(s)\operatorname{ch}(t-s)ds-\int_a^b E_x^f(s)\operatorname{ch}(b-s)dt\frac{\operatorname{sh}(t-a)}{\operatorname{sh}(b-a)}, \text{ and}$$

$$\nabla^{\pi}_{III}F(x)=\begin{cases}\int_a^t E_x^f(s)ds-\int_a^b E_x^f(s)ds\dfrac{t-a}{(1+p-a)(1+b-p)-1} & a\le t\le p\\ \int_a^t E_x^f(s)ds-\int_a^b E_x^f(s)ds\dfrac{(1+p-a)(1+t-p)-1}{(1+p-a)(1+b-p)-1} & p\le t\le b\end{cases}.$$

Proof. The technique of the proof is illustrated in the case when the gradient is $\langle,\rangle_I$.
The projected gradient may be written as $\nabla^{\pi}_I F(x)=\nabla_I F(x)-\lambda_a\nabla_I\phi_a-\lambda_b\nabla_I\phi_b$. It is also true that $D\phi_a\left(\nabla^{\pi}_I F(x)\right)=0$ and $D\phi_b\left(\nabla^{\pi}_I F(x)\right)=0$. Together the two equations form a linear system that determines λ_a and λ_b. Specifically,

$$\begin{bmatrix}D\phi_a(\nabla_I\phi_a) & D\phi_a(\nabla_I\phi_b)\\ D\phi_b(\nabla_I\phi_a) & D\phi_b(\nabla_I\phi_b)\end{bmatrix}\begin{bmatrix}\lambda_a\\ \lambda_b\end{bmatrix}=\begin{bmatrix}D\phi_a(\nabla_I F(x))\\ D\phi_b(\nabla_I F(x))\end{bmatrix},$$

which is equal to

$$\frac{1}{\alpha\beta(b-a)+\alpha+\beta}\begin{bmatrix}\beta(b-a)+1 & 1\\ 1 & \alpha(b-a)+1\end{bmatrix}\begin{bmatrix}\lambda_a\\ \lambda_b\end{bmatrix}=\begin{bmatrix}\nabla_I F(x)|_a\\ \nabla_I F(x)|_b\end{bmatrix}.$$

The solution is given by

$$\begin{bmatrix}\lambda_a\\ \lambda_b\end{bmatrix}=\begin{bmatrix}\alpha+\dfrac{1}{(b-a)} & -\dfrac{1}{(b-a)}\\ -\dfrac{1}{(b-a)} & \beta+\dfrac{1}{(b-a)}\end{bmatrix}\begin{bmatrix}\nabla_I F(x)|_a\\ \nabla_I F(x)|_b\end{bmatrix}.$$

With

$$W_x^f=\int_a^b f_x(t,x(t),\dot{x}(t))\,dt$$

this rewrites as

$$\lambda_a=\frac{\alpha+\dfrac{1}{(b-a)}}{\alpha\beta(b-a)+\alpha+\beta}\left(W_x^f-\beta\int_a^b E_x^f(s)ds\right)-\frac{\dfrac{1}{(b-a)}}{\alpha\beta(b-a)+\alpha+\beta}\left(\alpha\int_a^b E_x^f(s)ds+(\alpha(b-a)+1)W_x^f\right)$$

$$\lambda_b=-\frac{\dfrac{1}{(b-a)}}{\alpha\beta(b-a)+\alpha+\beta}\left(W_x^f-\beta\int_a^b E_x^f(s)ds\right)+\frac{\beta+\dfrac{1}{(b-a)}}{\alpha\beta(b-a)+\alpha+\beta}\left(\alpha\int_a^b E_x^f(s)ds+(\alpha(b-a)+1)W_x^f\right)$$

An impressive simplification produces

$$\lambda_a=-\frac{1}{b-a}\int_a^b E_x^f(s)ds\,,$$

$$\lambda_b=\frac{1}{b-a}\int_a^b E_x^f(s)ds+W_x^f\,,$$

$$\nabla^{\pi}_I F(x)=\int_a^t E_x^f(s)ds-\frac{\beta}{\alpha\beta(b-a)+\alpha+\beta}\int_a^b E_x^f(s)ds(\alpha(t-a)+1)+\frac{1}{b-a}\int_a^b E_x^f(s)ds\frac{\alpha a+\beta b-(\alpha+\beta)t}{\alpha\beta(b-a)+\alpha+\beta}$$

Another massive simplification yields

$$\nabla^{\pi}_I F(x)=\int_a^t E_x^f(s)ds-\int_a^b E_x^f(s)ds\frac{t-a}{b-a}\,.$$

Remark 1. Note the absence of W_x^f in the expression for the projected gradients. It is also obvious that $\nabla^\pi F(x)\big|_{t=a} = \nabla^\pi F(x)\big|_{t=b} = 0$.

Remark 2. From Remark 1 following Theorem 1 it is seen that if only the initial point $x(a)$ is kept fixed and $f_x = 0$, then the projected gradient is given by

$$\nabla_I^{\pi,a} F(x) = \int_a^t f_{\dot{x}}\left(s, x(s), \dot{x}(s)\right) ds ,$$

provided $\beta = 0$. In general

$$\nabla_I^{\pi,a} F(x) = \int_a^t E_x^f(s) ds + \frac{W_x^f - \beta \int_a^b E_x^f(s) ds}{\alpha\beta(b-a) + \alpha + \beta}\left(\alpha(t-a) + 1 - \frac{\beta(b-t)+1}{\beta(b-a)+1}\right).$$

The general cases when only $x(b)$ is kept fixed are given by

$$\nabla_I^{\pi,b} F(x) = \int_a^t E_x^f(s) ds - \frac{\alpha(t-a)+1}{\alpha(b-a)+1}\int_a^b E_x^f(s) ds .$$

It is somewhat surprising that both W_x^f and β are absent in this last formula. Another peculiarity is that if $\alpha = 1$ and $\beta = 0$, then

$$\nabla_I^{\pi,a} F(x) = \int_a^t E_x^f(s) ds + W_x^f (t-a) \text{, and}$$

$$\nabla_I^{\pi,b} F(x) = \int_a^t E_x^f(s) ds - \frac{(t-a)+1}{(b-a)+1}\int_a^b E_x^f(s) ds ,$$

but on the other hand if $\alpha = 0$ and $\beta = 1$, then

$$\nabla_I^{\pi,a} F(x) = \int_a^t E_x^f(s) ds + \left(W_x^f - \beta\int_a^b E_x^f(s) ds\right)\left(1 - \frac{(b-t)+1}{(b-a)+1}\right) \text{, and}$$

$$\nabla_I^{\pi,b} F(x) = -\int_t^b E_x^f(s) ds .$$

In the last four gradient formulas, the first and fourth are possible to interchange, and also the second and the third. The fourth and the second formulas are preferred both theoretically and numerically.

4. ISOPERIMETRIC CONSTRAINTS

4.1 One isoperimetric constraint. Assume the isoperimetric constraint is given by a functional $G(x) = \int g\left(t, x(t), \dot{x}(t)\right) dt$, where g satisfies the same technical assumptions as f of Theorem 1. Consider only the functions $x \in H$ subject to the conditions of Theorem 1. The constraint $G(x) = 0$ is classically known as an isoperimetric constraint. Assume regularity so that $G(x) = 0$ implies $DG(x)$ is onto and hence $\nabla G(x) \neq 0$. With this assumption H splits into a one-dimensional normal space and the tangent space $T_x\left(G^{-1}(0)\right) = \left\{ v \in H \;\middle|\; DG(x)v = 0 \right\}$. The splitting depends on x and the nonzero gradient $\nabla G(x)$. Note that $\nabla G(x)$ depends on the metric and it spans the normal space. The projection of a gradient onto the tangent space is denoted by $\nabla^{\pi,G}$. In Section 2 and 3 the concept of tangent space is not needed because in Section 2, $T_x H = H$, and in Section 3 $T_x\left(\phi_\tau^{-1}(0)\right) = \left\{ v \in H \;\middle|\; v(\tau) = 0 \right\}$, so in both cases the tangent space is independent of x. When the regularity assumption is satisfied, then for each metric there is a unique scalar field λ defined on $G^{-1}(0)$ such that $\nabla F(x) = \nabla^{\pi,G} F(x) + \lambda(x) \nabla G(x)$. Here $\lambda(x)$ is determined by the condition

$$DG(x)\nabla^{\pi,G} F(x) = 0 .$$

Remark 1. The condition $DG(x)\nabla^{\pi,G}F(x) = 0$ determines $\lambda(x)$. In fact the formula is given by

$$\lambda(x) = \frac{DG(x)\nabla F(x)}{DG(x)\nabla G(x)} .$$

To use this formula, recall the first line in the proof of Theorem 1,

$$DG(x)v = W_x^g v(b) + \int_a^b E_x^g(t)\dot{v}(t)dt .$$

When the metric is $\langle , \rangle_I$, use

$$\nabla_I F(x)(b) = \left(\alpha \int_a^b E_x^f(s)ds + W_x^f \right) \frac{1}{\alpha\beta(b-a) + \alpha + \beta} \text{, and}$$

$$\frac{d}{dt}\nabla_I F(x) = E_x^f(t) + \left(W_x^f - \beta \int_a^b E_x^f(s)ds \right) \frac{\alpha}{\alpha\beta(b-a) + \alpha + \beta} .$$

Remark 2. In Section 1.6 $\lambda(x)$ is given in the special case of $\langle , \rangle_I$ with $\alpha = 1, \beta = 0$ and $a = 0, b = 1$.

4.2 Multiple constraints. This section deals with the case when one of the endpoints or both of the endpoints are fixed and there is an isoperimetric constraint, or when there are several isoperimetric constraints and possibly fixed endpoints. First is necessary to check that the constrained subsets intersect transversally. To guarantee that this is indeed the case, assume that all the gradients of the constraining functionals are linearly independent. The next thing to decide is whether to deal with one constraint at a time or several constraints at a time or maybe all constraints at once. The following intuitive lemma implies that each way of dealing with the constraints is theoretically equivalent.

Projection Lemma. *Let H be a Hilbert space. Consider two closed linear subspaces V and W. Let $P_V^H : H \to V$, $P_W^H : H \to W$, $P_{V\cap W}^H : H \to V \cap W$, $P_{V\cap W}^V : V \to V \cap W$, and $P_{V\cap W}^W : W \to V \cap W$ denote the various projections. The following is always true*

$$P_{V\cap W}^H = P_{V\cap W}^V \circ P_V^H = P_{V\cap W}^W \circ P_W^H .$$

Proof. Given $x \in H$ there is a unique decomposition $x = x_V + x_V^{\perp_H}$ such that $x_V \in V$, and $P_V^H x_V^{\perp_H} = 0$. Similarly write $x = x_{V\cap W} + x_{V\cap W}^{\perp_H}$. With this notation it follows that

$$P_{V\cap W}^V \circ P_V^H(x) = P_{V\cap W}^V\left(x_V\right) = \left(x_V\right)_{V\cap W} .$$

It is also true that

$$x = x_V + x_V^{\perp_H} = (x_V)_{V\cap W} + (x_V)_{V\cap W}^{\perp_V} + x_V^{\perp_H} .$$

Now suppose that $y \in V \cap W$ is arbitrary. Since

$$\left\langle (x_V)_{V\cap W}^{\perp_V} + x_V^{\perp_H}, y \right\rangle_H = \left\langle (x_V)_{V\cap W}^{\perp_V}, y \right\rangle_H + \left\langle x_V^{\perp_H}, y \right\rangle_H = \left\langle (x_V)_{V\cap W}^{\perp_V}, y \right\rangle_V + 0 = 0 ,$$

it follows that

$$(x_V)_{V\cap W}^{\perp_V} + x_V^{\perp_H} \in (V\cap W)^{\perp_H} .$$

Now

$$x = (x_V)_{V\cap W} + \left((x_V)_{V\cap W}^{\perp_V} + x_V^{\perp_H} \right)$$

expresses x as a sum of a vector in $V \cap W$, and a vector in $(V\cap W)^{\perp_H}$. By uniqueness it must be that $(x_V)_{V\cap W} = x_{V\cap W}$. It follows that $P_{V\cap W}^H = P_{V\cap W}^V \circ P_V^H$. Obviously $P_{V\cap W}^H = P_{V\cap W}^W \circ P_W^H$, and hence $P_{V\cap W}^V \circ P_V^H = P_{V\cap W}^W \circ P_W^H$.

Remark 1. With the help of the projection lemma the formula for $\lambda(x)$ may be applied to gradients that are themselves the result of projections. This is particularly useful in the presence of constrained endpoints.

Remark 2. If all the constraints are dealt with at once, or at least all of the isoperimetric constraints when there are several constraints, then $\lambda(x)$ is a vector with the same number of components as the number of constraints. The formula for $\lambda(x)$ this time involves the inverse of a matrix. For theoretical purposes this may be the preferred approach but for computations some consideration should be given the other possibilities.

5. ASSOCIATED VARIATIONAL PROBLEMS

5.1 A scalar field in the space of derivatives. In order to define the associated variational problem it is assumed that the anti-derivatives are 'normalized' so that $x(p) = x_p$ for some given value $x_p \in \mathbb{R}$, and $a \le p \le b$. Define $H_{x_p} = \left\{ x \in H \;\middle|\; x(p) = x_p \right\}$, and let $F : H_{x_p} \to \mathbb{R}$ be given by

$$F(x) = \int_a^b f(t, x(t), \dot{x}(t))\, dt ,$$

where f satisfies the same assumptions as in Theorem 1. The associated functional is given by $\overline{F} : L^2[a,b] \to \mathbb{R}$ where

$$\overline{F}(v) = \int_a^b f\left(t, x_p + \int_p^t v(s)ds, v(t) \right) dt .$$

Note that if $v(t) = \dot{x}(t)$, then $\overline{F}(v) = F(x)$. Let

$$\overline{x}_v(t) = x_p + \int_p^t v(s)ds .$$

The characteristic function with $a \le \alpha < \beta \le b$ is given by

$$\chi_{[\alpha,\beta]}(t) = \begin{cases} 0 & a \le t < \alpha \\ 1 & \alpha \le t < \beta \\ 0 & \beta \le t \le b \end{cases}.$$

The techniques of the proof of Theorem 1 may be used to prove that the gradient vector field in $L^2[a,b]$ is given by $\nabla \overline{F}(v) = E^f_{\overline{x}_v} + W^f_{\overline{x}_v} \chi_{[p,b]}$.

Remark. As seen in this formula, the associated problem yields gradients that are free of derivatives of the function v. This fact is important when numerical algorithms are developed because it is well known that integration is more forgiving than differentiation. The flow along the gradient trajectories of F is rarely known explicitly so only discrete approximations of x are generally available. This means $\dot{x}$ is also only available as a discrete approximation. In contrast, the flow of $\overline{F}$ yields discrete approximations of v and hence discrete approximations of the anti-derivative $\overline{x}_v$. The operation $v \mapsto \overline{x}_v$ is better behaved numerically than the operation $x \mapsto \dot{x}$.

5.2 Preferred metrics. The process of differentiation yields a projection from H to $L^2[a,b]$, which restricts to H_{x_p}. There is also a 'lift' from $L^2[a,b]$ to $TH_{x_p} = \left\{ x \in H \,\middle|\, x(p) = 0 \right\}$ given by

$$v \mapsto \int_p^t v(s)ds .$$

A comparison with $\nabla_I^{\pi,a} F$ from Remark 2 of Section 3.3 shows that if $\beta = 0$, then $\nabla_I^{\pi,a} F$ projects onto $\nabla \overline{F}$ when $p = a$. Conversely, $\nabla \overline{F}$ lifts to $\nabla_I^{\pi,a} F$ when $\beta = 0$. Look at the following diagram

$$\begin{array}{ccc} x & \to & \nabla F(x) \\ \downarrow & & \\ \dot{x} & \to & \nabla_{L^2} \overline{F}(\dot{x}) \end{array}.$$

The upper half depends on the metric in H. The question is if this diagram may be completed to form a commutative diagram.

Proposition. *Given $x_p \in \mathbb{R}$, let $H_{x_p} = \left\{ x \in H \,\middle|\, x(p) = x_p \right\}$. Suppose $F : H_{x_p} \to \mathbb{R}$ and $\overline{F} : L^2[a,b] \to \mathbb{R}$ are given by*

$$F(x) = \int_a^b f(t, x(t), \dot{x}(t))\, dt \text{, and}$$

$$\overline{F}(v) = \int_a^b f\left(t, x_p + \int_p^t v(s)ds, v(t) \right) dt .$$

Denote the metric in H by $\langle , \rangle_H$ and let $TH_{x_p} = \left\{ x \in H \,\middle|\, x(p) = 0 \right\}$ have the induced metric. Suppose the form of $\langle , \rangle_H$ is given by $\langle \upsilon, \omega \rangle_H = \langle \upsilon, \omega \rangle_0 + \langle \dot{\upsilon}, \dot{\omega} \rangle_{L^2}$. Consider all f and x such that both $\nabla_{L^2} \overline{F}(\dot{x})$ and $\nabla_H F(x)$ exist. $\nabla_{L^2} \overline{F}(\dot{x}) = \frac{d}{dt} \nabla_H F(x)$ if and only if $\left\langle \nabla_H F(x), \sigma \right\rangle_0$ is a positive constant times $\nabla_H F(x)(p)\sigma(p)$ for all $\sigma \in TH_{x_p}$.

Proof. Start with an arbitrary $\sigma \in TH_{x_p}$ so that $\sigma(p) = 0$. Compute the two directional derivatives and note that they are related by $DF(x)\sigma = D\overline{F}(\dot{x})\dot{\sigma}$. It follows that

$$\left\langle \nabla_{L^2}\overline{F}(\dot{x}),\dot{\sigma}\right\rangle_{L^2} = \left\langle \nabla_H F(x),\sigma\right\rangle_H = \left\langle \nabla_H F(x),\sigma\right\rangle_0 + \left\langle \frac{d}{dt}\nabla_H F(x),\dot{\sigma}\right\rangle_{L^2}.$$

If $\left\langle \nabla_H F(x),\sigma\right\rangle_0$ is a positive constant times $\nabla_H F(x)(p)\sigma(p)$, then $\left\langle \nabla_H F(x),\sigma\right\rangle_0 = 0$, and

$$\nabla_{L^2}\overline{F}(\dot{x}) = \frac{d}{dt}\nabla_H F(x)$$

since $\dot{\sigma}$ is an arbitrary element in $L^2[a,b]$. Conversely, if $\left\langle \nabla_H F(x),\sigma\right\rangle_0$ is not a positive constant times $\nabla_H F(x)(p)\sigma(p)$ for some $\sigma \in TH_{x_p}$, then

$$\left\langle \nabla_{L^2}\overline{F}(\dot{x}) - \frac{d}{dt}\nabla_H F(x),\dot{\sigma}\right\rangle_{L^2} \neq 0$$

for this σ and hence $\nabla_{L^2}\overline{F}(\dot{x}) \neq \frac{d}{dt}\nabla_H F(x)$.

Remark. In light of this Proposition it is reasonable to refer to $\langle,\rangle_{III}$ as the 'preferred' metric. The positive constant alluded to in the Proposition is 1 in this case.

5.3 The associated problem and the metric $\langle,\rangle_{II}$. It is still true that $DF(x)\sigma = D\overline{F}(\dot{x})\dot{\sigma}$, so it follows that $\langle \nabla F(x),\sigma\rangle_{II} = \left\langle \nabla\overline{F}(\dot{x}),\dot{\sigma}\right\rangle_{L^2}$. In general, consider a pair of functions $\alpha,\beta \in TH_{x_a}$ such that $\langle \alpha,\sigma\rangle_{II} = \left\langle \dot{\beta},\dot{\sigma}\right\rangle_{L^2}$ for all $\sigma \in TH_{x_a}$. Use integration by parts and rewrite

$$\int_a^b \alpha(t)\sigma(t)dt + \int_a^b \dot{\alpha}(t)\dot{\sigma}(t)dt = \int_a^b \dot{\beta}(t)\dot{\sigma}(t)dt$$

in the form

$$\int_a^b \left\{ \int_t^b \alpha(s)ds + \dot{\alpha}(t) - \dot{\beta}(t) \right\} \dot{\sigma}(t)dt$$

An application of the duBois-Reymond's lemma shows that

$$\int_t^b \alpha(s)ds + \dot{\alpha}(t) - \dot{\beta}(t) = c$$

for some $c \in R$. The difference $\dot{\alpha} - \dot{\beta}$ must be twice differentiable and

$$\frac{d}{dt}\left(\dot{\alpha} - \dot{\beta}\right) = \alpha .$$

If α and β are twice differentiable, then $\ddot{\alpha} - \alpha = \ddot{\beta}$. When $\ddot{\beta}$ is known this is a non-homogenous linear ordinary differential equation. Note that the homogenous solution may be expressed in terms of hyperbolic functions so this analysis yields a better understanding of the presence of hyperbolic functions in $\nabla_{II}F$. Finally, as mentioned in the introduction, at least one 'natural' metric is known to break space symmetries along gradient trajectories; see [5]. It is proved in [6] that the metrics $\langle,\rangle_{II}$ and $\langle,\rangle_{III}$, with $p = (a+b)/2$, both preserve symmetries. Gradients in metric $\langle,\rangle_{II}$ do not commute with the derivative operator, but smoothness is maintained. Metric $\langle,\rangle_{III}$ is preferred but smoothness is sometimes destroyed as seen in the following example.

Example: Suppose $p = 1/2$ in the metric $\langle,\rangle_{III}$. Consider the smooth unconstrained functional

$$F(x) = \int_0^1 \cos(x(t))\,dt\,.$$

The initial function is given by $x_0(t) = \pi/2$. The global minimum value of F is -1, and this value is attained at the 'nearby' function given by $x_\infty(t) = \pi$. Figure 1 shows snapshots of the negative gradient trajectory with initial point x_0 and limit x_∞. The gradient flow slows down as the limit is approached. Increasing amounts of flow-time separates the curves indicated. Smoothness is lost instantly during the descent. Symmetry of reflection is retained. To summarize, nice properties of the gradient flow come at a cost. Only metric $\langle , \rangle_{II}$ yields gradients capable of preserving symmetries and maintaining smoothness. Hyperbolic functions complicate the gradient formulas and the gradient will not commute with the derivative operator. If computational speed or peaceful coexistence with the derivative operator is important, then either symmetry preservation or smoothness must be sacrificed.

References

[1] Buttazzo, G., Giaquinta, M. and Hildebrandt, S., *One-dimensional Variational Problems*, Oxford Lecture Series in Mathematics and its Applications, Oxford Uinversity Press 1998

[2] Linnér, A., *Some properties of the curve straightening flow in the plane*, Trans. Amer. Math. Soc. **314** (1989), 605-617

[3] Linnér, A., *Steepest descent used as a tool to find critical points of* $\int k^2$ *defined on curves in the plane with arbitrary boundary conditions*, Geometric Analysis and Computer Graphics, MSRI Publications, Vol. 17, Springer-Verlag, 1991, 127-138

[4] Linnér, A., *Curve-straightening in closed Euclidean submanifolds,* Commun. Math. Phys. **138** (1991), 33-49

[5] Linnér, A., *Curve-straightening and the Palais-Smale condition,* Trans. Amer. Math. Soc. **350** (1998), 3743-3765

[6] Linnér, A., *Symmetrized Curve-straightening,* Preprint (to appear in Differential Geometry and its Applications)

[7] Rudin, W., *Real and Complex Analysis, third edition*, 1987 McGraw-Hill

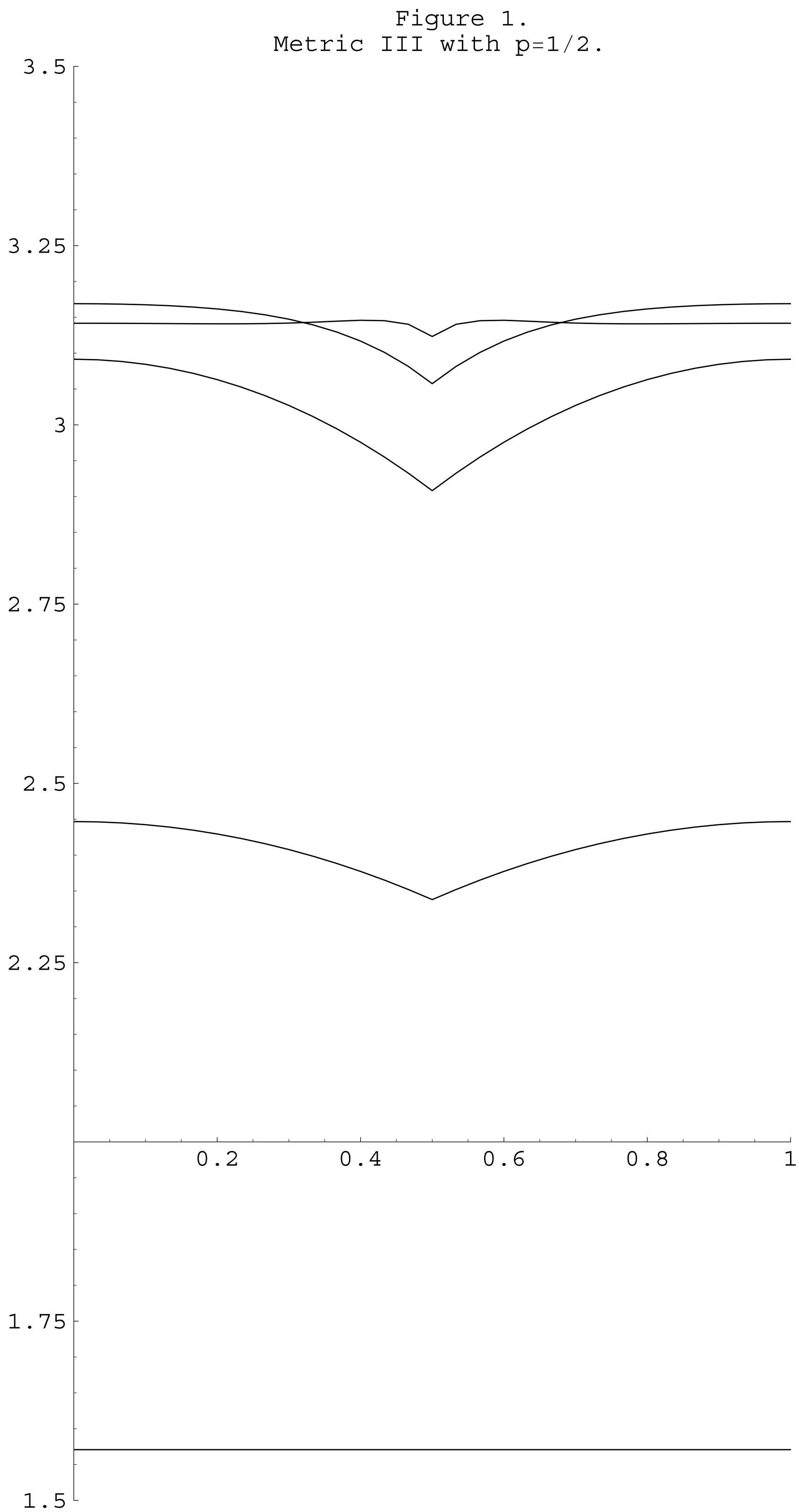
Figure 1.
Metric III with p=1/2.
3.5
3.25
3
2.75
2.5
2.25
1.75
1.5
0.2
0.4
0.6
0.8
1