Chapter II. Isoperimetric Problems

A. Abstract Derivatives

2026-01-30

For any real-valued function[al] $\Phi\colon V \to \mathbb{R}$ defined on a vector space V, let

$$\Phi'[\widehat{x};h] = \lim_{\lambda\to 0^+} \frac{\Phi[\widehat{x}+\lambda h]-\Phi[\widehat{x}]}{\lambda}, \qquad h \in V.$$

This is the directional derivative of Φ at the point $\widehat{x}$, in the direction h. The concept makes no reference at all to a topology on V; we bring this issue in later.

Examples. Note that when $V = \mathbb{R}^n$ and Φ is smooth, $\Phi'[x;h] = \nabla\Phi(x) \bullet h$. And when $V = PWS[a,b]$ and Λ is the usual variational integral, with $L = L(t,x,v)$ of class C^1, we have

$$\Lambda'[\widehat{x};h] = \int_a^b \left(\widehat{L}_x(r)h(r) + \widehat{L}_v(r)\dot{h}(r)\right)\,dr, \qquad h \in PWS[a,b].$$

When $\widehat{x} \in V$ is a point at which $\Phi'[\widehat{x};h]$ is defined for every h and the map $h \mapsto \Phi'[\widehat{x};h]$ is linear, we use the notation $D\Phi[\widehat{x}]$ for this map. We'll manipulate it like any other matrix or linear operator A, writing Ah instead of $A(h)$. So $\Phi'[\widehat{x};h] = D\Phi[\widehat{x}]h$.

It's safe to call $D\Phi[\widehat{x}]$ the Gâteaux derivative of Φ at $\widehat{x}$.

Digression. The terminology of Gâteaux derivatives is not completely standardized. Some authors require only $\Phi'[\widehat{x};-h] = -\Phi'[\widehat{x};h]$ for each h instead of demanding linearity, while others require not only linearity but also some mild form of continuity. These fine distinctions don't make a difference for us.

Various definitions of "differentiability" are available. The same linear map $D\Phi[\widehat{x}]$ appears in all of them. The terminology (Gâteaux, Fréchet, etc.) highlights the trustworthiness of that same operator as an approximation for the underlying function near $\widehat{x}$.

- Gâteaux differentiability describes reliable approximation independently for each given line through $\widehat{x}$:

$$\forall h \in V, \quad \Phi[\widehat{x}+\lambda h] = \Phi[\widehat{x}] + \lambda D\Phi[\widehat{x}]h + o(\lambda) \text{ as } \lambda \to 0.$$

- If V has a norm, we call Φ Fréchet differentiable at $\widehat{x}$ when the operator $D\Phi[\widehat{x}]$ gives an approximation that is uniform in the direction, i.e., when

$$\Phi[\widehat{x}+h] = \Phi[\widehat{x}] + D\Phi[\widehat{x}]h + o(\|h\|) \text{ as } \|h\| \to 0.$$

There are closely related considerations when we come to defining local minimizers in the Calculus of Variations. More on this later.

B. Local Surjectivity

2026-02-02

Theorem (Surjective Mapping). *Given an open set U in $\mathbb{R}^n$ and a C^1 mapping $G: U \to \mathbb{R}^n$, consider the Jacobian matrix at a point $u_0 \in U$:*

$$DG(u_0) = \begin{bmatrix} \dfrac{\partial g_1}{\partial u_1} & \dfrac{\partial g_1}{\partial u_2} & \cdots & \dfrac{\partial g_1}{\partial u_n} \\ \dfrac{\partial g_2}{\partial u_1} & \dfrac{\partial g_2}{\partial u_2} & \cdots & \dfrac{\partial g_2}{\partial u_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial g_n}{\partial u_1} & \dfrac{\partial g_n}{\partial u_2} & \cdots & \dfrac{\partial g_n}{\partial u_n} \end{bmatrix}_{u=u_0} .$$

If $DG(u_0)$ is invertible, then $G(u_0) \in \operatorname{int} G(U)$.

Proof. Introduce $F: \mathbb{R}^n \times U \to \mathbb{R}^n$ by defining

$$F(x, u) = G(u) - x, \qquad x \in \mathbb{R}^n,\ u \in U.$$

Let $x_0 = G(u_0)$. This F is C^1 and $D_u F(x_0, u_0) = DG(u_0)$ is invertible, so the Implicit Function Theorem provides an open set X containing x_0 and a C^1 function $\phi: X \to \mathbb{R}^n$ such that

(i) $0 = F(x, \phi(x)) = G(\phi(x)) - x$, $x \in X$, and

(ii) (a uniqueness condition we don't need).

Line (i) shows that every x in the open set X around x_0 lies in the range of G (indeed $x = G(\phi(x))$). That's the desired result. ////

Lemma. *Given a real vector space V and linear operators $M_1, \ldots, M_n: V \to \mathbb{R}$, TFAE:*

(a) *The given operators are linearly dependent, i.e., there exist $c_1, c_2, \ldots, c_n \in \mathbb{R}$, not all zero, such that*

$$c_1 M_1 + c_2 M_2 + \ldots + c_n M_n = 0.$$

(*On the right,* 0 *denotes the zero operator on V.*)

(b) *For each choice of $h_1, h_2, \ldots, h_n \in V$, the $n \times n$ matrix below is not invertible:*

$$\begin{bmatrix} M_1 h_1 & M_2 h_1 & M_3 h_1 & \cdots & M_n h_1 \\ M_1 h_2 & M_2 h_2 & M_3 h_2 & \cdots & M_n h_2 \\ M_1 h_3 & M_2 h_3 & M_3 h_3 & \cdots & M_n h_3 \\ \vdots & & & \ddots & \vdots \\ M_1 h_n & M_2 h_n & M_3 h_n & \cdots & M_n h_n \end{bmatrix}.$$

Proof. (a⇒b) For a collection of constants as described in (a), and any elements $h_1, \ldots, h_n$ of V, observe that

$$\begin{bmatrix} M_1h_1 & M_2h_1 & M_3h_1 & \cdots & M_nh_1 \\ M_1h_2 & M_2h_2 & M_3h_2 & \cdots & M_nh_2 \\ M_1h_3 & M_2h_3 & M_3h_3 & \cdots & M_nh_3 \\ \vdots & & & \ddots & \vdots \\ M_1h_n & M_2h_n & M_3h_n & \cdots & M_nh_n \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

In detail, row j of the indicated product equals

$$c_1M_1h_j + c_2M_2h_j + \cdots + c_nM_nh_j = (c_1M_1 + c_2M_2 + \cdots + c_nM_n)\, h_j = 0.$$

Since the column vector on the left side of this product is nonzero, the matrix involved must be singular.

(b⇒a) Let's use induction, recalling that a matrix is invertible if and only if its determinant is nonzero.

The case $n = 1$ is obvious, and it is enough to launch the generic proof.

Let's show case $n = 2$ explicitly to make the ideas clear. So assume

$$0 = \left| \begin{bmatrix} M_1h_1 & M_2h_1 \\ M_1h_2 & M_2h_2 \end{bmatrix} \right| = (M_1h_1)(M_2h_2) - (M_2h_1)(M_1h_2) \qquad \forall h_1, h_2 \in V.$$

The result is obvious if M_2 is itself the zero operator, since the choices $c_1 = 0$ and $c_2 = 1$ give the conclusion in that case. In the complementary case, there must exist some $h_2 \in V$ for which $M_2h_2 \neq 0$. Use this h_2 to define $c_1 = M_2h_2$ and $c_2 = -M_1h_2$. Then $c_1 \neq 0$ and the identity above becomes

$$0 = c_1(M_1h_1) + c_2(M_2h_1) = (c_1M_1 + c_2M_2)h_1 \qquad \forall h_1 \in V.$$

This is the desired result.

Now if the result is known for all dimensions up to and including $n-1$, imagine expanding the determinant in the statement by minors, along the first row:

$$0 = (M_1h_1)\Delta_1 + (M_2h_2)\Delta_2 + \cdots + (M_nh_1)\Delta_n.$$

Here each Δ_j is $(-1)^{j+1}$ times the determinant of a certain $(n-1) \times (n-1)$ submatrix. Note that each Δ_j involves only the $n-1$ inputs $h_2, \ldots, h_n$. Now suppose $\Delta_1 = 0$ for every possible combination of $h_2, \ldots, h_n$. Then by the $(n-1)$-dimensional case of our result (known to be true, by our induction hypothesis), the operators $M_2, \ldots, M_n$ must be linearly dependent. Of course that makes the larger set $M_1, \ldots, M_n$ linearly dependent also, so the result holds. In the complementary case, there must exist some choice of $h_2, \ldots, h_n$ for which $\Delta_1 \neq 0$. Use this particular choice to define $c_j = \Delta_j$ for each j, and then revisit the identity above:

$$\begin{aligned} 0 &= c_1(M_1h_1) + c_2(M_2h_1)\Delta_2 + \cdots + c_n(M_nh_1) \\ &= (c_1M_1 + c_2M_2 + \cdots + c_nM_n)h_1, \qquad h_1 \in V. \end{aligned}$$

This is the desired result. (Note that $c_1 \neq 0$.) ////

C. Lagrange Multipliers

2026-02-04

Now suppose V is a real vector space on which we have real-valued function[al]s $\Phi, \Gamma_1, \ldots, \Gamma_m$, and constants $\gamma_1 \ldots, \gamma_m$. Consider the constrained optimization problem

$$\min_{x\in V} \left\{\Phi[x] \,:\, \Gamma_j[x] = \gamma_j,\ j = 1, \ldots, m\right\}.$$

Suppose $\widehat{x}$ solves this. Let's show why the operators $D\Phi[\widehat{x}], D\Gamma_1[\widehat{x}], \ldots, D\Gamma_m[\widehat{x}]$ must be linearly dependent.

The Lemma above is key. Pick arbitrary $y, h_1, \ldots, h_m$ in V and define $G\colon \mathbb{R}^{1+m} \to \mathbb{R}^{1+m}$ via

$$G(r_0, r_1, \ldots, r_m) = \begin{bmatrix} \Phi[\widehat{x} + r_0 y + r_1 h_1 + \cdots + r_m h_m] \\ \Gamma_1[\widehat{x} + r_0 y + r_1 h_1 + \cdots + r_m h_m] \\ \vdots \\ \Gamma_m[\widehat{x} + r_0 y + r_1 h_1 + \cdots + r_m h_m] \end{bmatrix}.$$

Now $G(0) = (\Phi[\widehat{x}], \gamma_1, \ldots, \gamma_m)$. This is the limit of the sequence $(\Phi[\widehat{x}] - 1/k, \gamma_1, \ldots, \gamma_m)$, $k = 1, 2, \ldots$, and every point in this sequence lies outside the range of G. (Proof: Once the constraints are satisfied, it's impossible to push the value of the first component lower than $\Phi[\widehat{x}]$.) So in particular, $G(0)$ is not an interior point of $G(V)$. Thanks to the Local Surjection Theorem, the following matrix must fail to be invertible:

$$DG(0) = \begin{bmatrix} D\Phi[\widehat{x}]y & D\Phi[\widehat{x}]h_1 & D\Phi[\widehat{x}]h_2 & \cdots & D\Phi[\widehat{x}]h_m \\ D\Gamma_1[\widehat{x}]y & D\Gamma_1[\widehat{x}]h_1 & D\Gamma_1[\widehat{x}]h_2 & \cdots & D\Gamma_1[\widehat{x}]h_m \\ \vdots & & & & \\ D\Gamma_m[\widehat{x}]y & D\Gamma_m[\widehat{x}]h_1 & D\Gamma_m[\widehat{x}]h_2 & \cdots & D\Gamma_m[\widehat{x}]h_m \end{bmatrix}.$$

This happens for every choice of $y, h_1, \ldots, h_m$ in V, so by the Lemma above, the $m+1$ operators $D\Phi[\widehat{x}]$, $D\Gamma_1[\widehat{x}]$, ..., $D\Gamma_m[\widehat{x}]$ must be linearly dependent.

(How much smoothness do we need? To apply the Local Surjection Theorem requires the map $(r_0, \ldots, r_m) \mapsto \Gamma_j[\widehat{x} + r_0 y + \sum_j r_j h_j]$ to be C^1 near the origin of $\mathbb{R}^{1+m}$, for every choice of $y, h_1, \ldots, h_m$.)

Linear dependence requires existence of $c_0, c_1, \ldots, c_m$, not all zero, such that

$$\begin{aligned} 0 &= c_0 D\Phi[\widehat{x}] + c_1 D\Gamma_1[\widehat{x}] + \cdots + c_m D\Gamma_m[\widehat{x}] \\ &= D\left(c_0\Phi + c_1\Gamma_1 + \cdots + c_m\Gamma_m\right)[\widehat{x}]. \end{aligned}$$

The objective functional Φ is special, so its coefficient gets special attention. If $c_0 \neq 0$, we divide by it in the identity above and define $\lambda_j = c_j/c_0$ for each j. If $c_0 = 0$, we skip the division and define $\lambda_j = c_j$ for each j. The formal statement below condenses this two-way split.

Theorem (Lagrange Multipliers). *If $\widehat{x}$ achieves the minimum in the constrained problem above, then there exist $\lambda_0 \in \{0,1\}$ and $\lambda \in \mathbb{R}^m$, not both zero, such that*

$$0 = D\left(\lambda_0\Phi + \lambda_1\Gamma_1 + \cdots + \lambda_m\Gamma_m\right)[\widehat{x}].$$

Remarks. 1. Taking both $\lambda_0 = 0$ and $\lambda = 0$ in the conclusion above yields the correct-but-trivial statement $0 = 0$ for each and every $\widehat{x}$, so it is essential to stipulate "not both zero" to give the theorem any chance to be useful.

2. The possibility that $\lambda_0 = 0$ must be allowed to obtain a correct statement, but it usually signals a situation where there is something unusual about the problem formulation or the constraint structure. This situation is called "abnormal." In exploring a new problem, it's typical to start with a short exploration of the abnormal case: often this is easily shown to be a dead end, which guarantees that the normal situation will capture the desired solution. Examples appear below.

Example (Rayleigh Quotient). Suppose $V = \mathbb{R}^n$ and a symmetric matrix $A \in \mathbb{R}^{n\times n}$ is given. Here is a problem with just one constraint (so $m = 1$):

$$\min_{x\in\mathbb{R}^n}\left\{x^TAx \,:\, |x|^2 = 1\right\}.$$

Here $\Phi[x] = x^TAx$, $\Gamma[x] = |x|^2 = x^TIx$, $\gamma = 1$. For any $x, y \in \mathbb{R}^n$,

$$\begin{aligned}\Phi(x+y) - \Phi(x) &= (x+y)^TA(x+y) - x^TAx\\ &= \left(x^TAx + y^TAx + x^TAy + y^TAy\right) - x^TAx\\ &= 2x^TAy + y^TAy.\end{aligned}$$

So for any $h \in \mathbb{R}^n$,

$$\Phi'[x;h] = \lim_{r\to 0^+}\frac{\Phi[x+rh]-\Phi[x]}{r} = \lim_{r\to 0^+}\left(2x^TAh + rh^TAh\right) = 2x^TAh.$$

This is indeed a scalar-valued linear function of h, so (as an operator) $D\Phi[x] = 2x^TA$. For the special case $A = I$ we get $D\Gamma[x] = 2x^T$.

Now if $\widehat{x}$ gives the minimum, there exist $\lambda_0 \in \{0,1\}$ and $\lambda \in \mathbb{R}$, not both zero, such that

$$0 = D(\lambda_0\Phi + \lambda\Gamma)[\widehat{x}] = 2\lambda_0\widehat{x}^TA + 2\lambda\widehat{x}^T.$$

Take transposes (remember $A = A^T$) and cancel 2's to show that

$$\lambda_0 A\widehat{x} + \lambda\widehat{x} = 0.$$

Could $\lambda_0 = 0$ work? No, because "not both zero" then requires $\lambda \neq 0$, which entails $\widehat{x} = 0$, and that is incompatible with the constraint. So the problem is normal and we take $\lambda_0 = 1$. This shows that $\widehat{x}$ is an eigenvector of A, with eigenvalue $-\lambda$. The constraint requires $|\widehat{x}| = 1$. Consequently

$$\Phi[\widehat{x}] = (\widehat{x})^T A\widehat{x} = (\widehat{x})^T(-\lambda\widehat{x}) = -\lambda.$$

Geometrically, the Lagrange Multiplier Rule focusses the search for minimizers in this problem to unit eigenvectors for A. The absolute minimum for Φ on the unit sphere is achieved by the smallest eigenvalue. Every eigenvector for A is compatible with some solution of the Lagrange Multiplier setup, but the ones for other eigenvalues do not give the minimum. The largest eigenvalue gives the *maximum.* If there are eigenvalues strictly between the smallest and the largest, each of their eigenvectors gives neither a min nor a max. ////

Practice. Let A be a symmetric matrix of shape $n \times n$ and let $\widehat{u}$ be any eigenvector of A. Show that the minimum of $x^T Ax$ on the slice of the unit sphere $|x|^2 = 1$ defined by the orthogonality condition $\widehat{u}^T x = 0$ is again an eigenvalue of A.

Practice. Let A be a symmetric matrix of shape $n \times n$ and let $\widehat{u}_1, \ldots, \widehat{u}_k$ be linearly independent eigenvectors of A; assume $k < n$. Show that the minimum of $x^T Ax$ on the slice of the unit sphere $|x|^2 = 1$ defined by the orthogonality condition $\widehat{u}_j^T x = 0$ for $j = 1, 2, \ldots, k$ is again an eigenvalue of A.

Example. Here is an abnormal problem in $\mathbb{R}^2$:

$$\min\left\{\phi(s,t) = te^s \,:\, 0 = g(s,t) \stackrel{\text{def}}{=} s^2 + t^2\right\}.$$

The only point where $g = 0$ is $\widehat{x} = (0,0)$, so it must be the point that gives the minimum. Lagrange says there must be some λ_0, λ (not both zero) such that

$$\begin{aligned}(0,0) &= \nabla\left(\lambda_0\phi + \lambda g\right)(\widehat{x}) \\ &= \lambda_0(0,1) + \lambda(0,0) = (0,\lambda_0).\end{aligned}$$

This is correct when $\lambda_0 = 0$ (any λ then works), but it never holds with $\lambda_0 = 1$. ////

D. Applications in the Calculus of Variations

2026-02-06

The standard isoperimetric problem

$$\begin{aligned}\min\{\Lambda[x] := \int_a^b L(t,x(t),\dot{x}(t))\,dt \,&: x \in PWS[a,b],\ x(a) = A,\ x(b) = B, \\ &\int_a^b G_j(t,x(t),\dot{x}(t))\,dt = \gamma_j,\ j = 1,2,\ldots,m\}\end{aligned}$$

fits the pattern described abstractly above. We need only recognize the vector space X, affine subset S, subspace V, and functional Γ as follows:

$$\begin{aligned}S &= \{x \in PWS[a,b] \,:\, x(a) = A,\ x(b) = B\}, \\ V &= V_{II} = \{y \in PWS[a,b] \,:\, y(a) = 0 = y(b)\}, \\ \Gamma[x] &= \int_a^b G(t,x(t),\dot{x}(t))\,dt.\end{aligned}$$

The smoothness conditions on Λ and Γ will hold whenever the corresponding integrands L and G are of class C^1. The extremality theorem stated above concerns the operator

$$(\lambda_0\Lambda + \sum_{j=1}^{m} \lambda_j \Gamma_j)[x] = \int_a^b \left(\lambda_0 L(t, x(t), \dot{x}(t)) + \sum_{j=1}^{m} \lambda_j G_j(t, x(t), \dot{x}(t)) \right) dt.$$

Conclusion (3), that $D(\lambda_0\Lambda + \sum_j \lambda_j\Gamma_j)[\widehat{x}] = 0$, is equivalent to the statement that the arc $\widehat{x}$ obeys IEL for the integrand

$$\widetilde{L}(t, x, v) = \lambda_0 L(t, x, v) + \sum \lambda_j G_j(t, x, v).$$

The direct translation of the abstract theorem into the present context is as follows:

Theorem. *Suppose $\widehat{x}$ solves the isoperimetric variational problem stated above. If both L and G are C^1, then there must be constants $\lambda_0 \in \{0, 1\}$ and $\lambda \in \mathbb{R}$, not both zero, such that for some c,*

$$\lambda_0 \widehat{L}_v(t) + \lambda \widehat{G}_v(t) = c + \int_a^t \left(\lambda_0 \widehat{L}_x(r) + \lambda \widehat{G}_x(r) \right) dr \qquad t \in [a, b].$$

Remarks. 1. On regularity: Weierstrass/Hilbert applies to extremals for a given Lagrangian. Thus it applies if you use $\widetilde{L} = \lambda_0 L + \lambda_1 G_1 + \ldots + \lambda_m G_m$.

2. On the natural boundary conditions: Adding $\ell(x(a), x(b))$ to Λ and $g_j(x(a), x(b))$ to Γ_j is fully compatible with the developments above. In a problem where both endpoints are unconstrained, the minimizing $\widehat{x}$ will be an extremal for $\widetilde{L}$ as defined above, and also the endpoint function

$$\widetilde{\ell}(x, y) = \lambda_0 \ell(x, y) + \sum_j \lambda_j g_j(x, y)$$

will make

$$(p(a), -p(b)) = \nabla\widetilde{\ell}(\widehat{x}(a), \widehat{x}(b)), \quad \text{where} \quad p(t) = \widehat{\widetilde{L}}_v(t) \text{ e.a.} t \in [a, b].$$

An Abnormal Situation. In the isoperimetric problem

$$\min \left\{ \int_0^1 t^2 \dot{x}(t)^2 \, dt \, : \, x(0) = 0, \; x(1) = 1, \; \int_0^1 \sqrt{1 + \dot{x}(t)^2} \, dt = \sqrt{2} \right\},$$

every arc providing a global minimum must have corresponding constants $\lambda_0 \in \{0, 1\}$ and $\lambda \in \mathbb{R}$, not both zero, such that $\widehat{x}$ obeys (IEL) for

$$\widetilde{L} = \lambda_0 t^2 v^2 + \lambda \sqrt{1 + v^2}.$$

That is, since $\widetilde{L}_x = 0$, some constant c obeys

$$c = \widehat{L}_v(t) = 2\lambda_0 t^2 \dot{\widehat{x}}(t) + \lambda \frac{\dot{\widehat{x}}(t)}{\sqrt{1 + \dot{\widehat{x}}(t)^2}}.$$

Closer inspection of the constraints reveals there is only one admissible arc, namely $\widehat{x}(t) = t$. This arc must certainly give the minimum, and substitution in the relation above gives

$$2\lambda_0 t^2 + \frac{\lambda}{\sqrt{2}} = c \qquad \forall t \in [0,1].$$

This forces $\lambda_0 = 0$: the only form in which the conclusion of the theorem holds is the abnormal one. ////

F. The Catenary

We consider a flexible chain of length γ hanging between points (a, A) and (b, B) in a vertical plane. The shape of the chain will minimize the potential energy. Assuming constant linear density ρ, a segment of length ds will have mass $dm = \rho\, ds$ and gravitational potential energy $dU = (dm)(g)(x) = \rho g x \sqrt{dt^2 + dx^2}$. This leads to the problem

$$\begin{array}{ll} \text{minimize} & \Lambda[x] = \displaystyle\int_a^b x\sqrt{1 + \dot{x}(t)^2}\, dt \\ \text{over all} & x \in PWS[a,b] \\ \text{subject to} & \Gamma[x] = \displaystyle\int_a^b \sqrt{1 + \dot{x}(t)^2}\, dt = \gamma, \\ & x(a) = A,\ x(b) = B. \end{array}$$

2026-02-09

We look for admissible arcs that are extremal for $\widetilde{L} = (\lambda_0 x + \lambda)\sqrt{1 + v^2}$, with $\lambda_0 \in \{0, 1\}$ and λ not both zero. Now if $\lambda_0 = 0$, we know extremals for $\sqrt{1 + v^2}$ are straight lines. The straight line from (a, A) to (b, B) will be admissible exactly when its length exactly matches the available length, i.e., when

$$\gamma = \sqrt{(B - A)^2 + (b - a)^2}.$$

There is no meaningful choice available in this situation, but it is a valid optimum in a well-defined instance of the problem. Of course if the available length obeys $\gamma < \sqrt{(B - A)^2 + (b - a)^2}$ then there are no admissible arcs, so the problem has no solution. If the reverse inequality holds (strictly), there is a legitimate choice to make, and we will need $\lambda_0 = 1$ to do it. So we focus on $\widetilde{L} = (x + \lambda)\sqrt{1 + v^2}$. This has the special form $f(x)\sqrt{1 + v^2}$ considered in general some lectures ago. For an extremal $x()$, there must be a constant k such that

$$k(x(t) + \lambda) = \sqrt{1 + \dot{x}(t)^2}$$

Clearly $k = 0$ is not compatible, so the continuous function $x(t) + \lambda$ cannot change sign. It follows that $L_{vv}(t, x(t), v)$ never changes sign, so $x \in C^2$. Standard manipulations lead to

$$\dot{x}(t)^2 = k^2(x(t) + \lambda)^2 - 1.$$

On any interval where the sign of $\dot{x}$ is $\sigma \in \{-1, 1\}$,

$$\frac{dx}{\sqrt{k^2(x + \lambda)^2 - 1}} = \sigma\, dt.$$

Substitute $u = k(x + \lambda)$, $du = k\, dx$, to get

$$\begin{aligned} \frac{1}{k} \int \frac{du}{\sqrt{u^2 - 1}} &= \sigma t + \frac{C(\sigma)}{k} \\ \frac{1}{k} \log \left| u + \sqrt{u^2 - 1} \right| &= \sigma t + \frac{C(\sigma)}{k} \\ \log \left| u + \sqrt{u^2 - 1} \right| &= \sigma k t + C(\sigma). \end{aligned}$$

Take a chance on $u(t) > 1$, in which case

$$u + \sqrt{u^2 - 1} = e^{\sigma k t + C(\sigma)}.$$

This implies

$$\begin{aligned} u^2 - 1 = \left(e^{\sigma k t + C(\sigma)} - u \right)^2 &= e^{2(\sigma k t + C(\sigma))} - 2u e^{\sigma k t + C(\sigma)} + u^2 \\ 2u e^{\sigma k t + C(\sigma)} &= e^{2(\sigma k t + C(\sigma))} + 1 \\ u &= \frac{e^{\sigma k t + C(\sigma)} + e^{-\sigma k t - C(\sigma)}}{2} \end{aligned}$$

Note

$$\dot{u} = \frac{\sigma k e^{\sigma k t + C(\sigma)} - \sigma k e^{-\sigma k t - C(\sigma)}}{2}.$$

To make a smooth connection from an interval with $\sigma = +1$ to an interval with $\sigma = -1$ at some instant t, that transition point will have to make $\dot{u}(t) = 0$ in two ways:

$$\begin{aligned} 0 = e^{kt + C(1)} - e^{-kt - C(1)} &\implies kt + C(1) = 0 \\ 0 = -e^{-kt + C(-1)} + e^{kt - C(-1)} &\implies kt - C(-1) = 0 \end{aligned}$$

To reconcile these requires $C(-1) = -C(1)$. Now write $C = C(1)$ and summarize:

$$\begin{aligned} \dot{u} > 0 &\implies u = \frac{e^{kt + C} + e^{-kt - C}}{2}, \\ \dot{u} < 0 &\implies u = \frac{e^{-kt - C} + e^{kt + C}}{2}. \end{aligned}$$

Excellent: both expressions on the right are identical. So the single form below covers both cases at once:

$$u(t) = k(x(t) + \lambda) = \frac{e^{kt+C} + e^{-kt-C}}{2} = \cosh(kt + C).$$

We know x is convex (since $f(x) = x + \lambda$ is increasing, and we showed this in the general discussion), so we must have $k > 0$.

Now we have $x(t) = k^{-1}\cosh(kt + C) - \lambda$, with the three constants $k > 0$, C, and λ still to determine. The endpoint conditions give two equations, and the third comes from the arc length constraint. For the latter, we note that $\dot{x}(t) = \sinh(kt+C)$ and insist upon

$$\begin{aligned}\gamma &= \int_a^b \sqrt{1 + \sinh^2(kt + C)}\,dt = \int_a^b \sqrt{\cosh^2(kt + C)}\,dt \\ &= k^{-1}\sinh(kt + C)\Big|_a^b = k^{-1}\sinh(kb + C) - k^{-1}\sinh(ka + C).\end{aligned}$$

Rearranging the two endpoint equations by keeping one as-is and using the difference instead of the second leads to these 3 equations for the 3 unknowns k, C, and λ:

$$B - A = k^{-1}\cosh(kb + C) - k^{-1}\cosh(ka + C) \tag{1}$$

$$\gamma = k^{-1}\sinh(kb + C) - k^{-1}\sinh(ka + C) \tag{2}$$

$$\lambda = k^{-1}\cosh(ka + C) - A \ \left[= k^{-1}\cosh(kb + C) - B \right] \tag{3}$$

In this form, the constant λ does not appear in equations (1)–(2), so we can solve this pair of equations for k and C, and then recover λ from (3). Typically this calls for an approximate solution from the computer.

(Newton's Method for solving $F(p) = 0$ in $\mathbb{R}^n$ for a given function $F\colon\mathbb{R}^n \to \mathbb{R}^n$ is an effective choice.)

Let's try the case where $A = B$. This makes $B - A = 0$ in (1), leading to $\cosh(kb + C) = \cosh(ka + C)$. Since cosh is even, $k > 0$, and $a < b$, we must have

$$ka + C = -(kb + C), \quad \text{i.e.,} \quad 2C = -k(b + a), \quad \text{i.e.,} \quad C = -\frac{k(b + a)}{2}.$$

This reduces the remaining equations to

$$\gamma = k^{-1}\sinh(\tfrac{1}{2}k(b - a)) - k^{-1}\sinh(\tfrac{1}{2}k(a - b)) = 2k^{-1}\sinh(\tfrac{1}{2}k(b - a)) \tag{2}$$

$$\lambda = k^{-1}\cosh(\tfrac{1}{2}k(b - a)) - A \tag{3}$$

Given $\gamma > 0$, solve (2) for k, then use

$$x(t) = k^{-1}\left[\cosh(kt) - \cosh((\tfrac{1}{2}k(b - a))\right] + A, \qquad a \le t \le b.$$

In the even more special case where $A = B$ and $b = -a > 0$, we get

$$\gamma = 2k^{-1}\sinh(kb) \tag{2}$$

$$\lambda = k^{-1}\cosh(kb) - A \tag{3}$$

Given $\gamma > 0$, solve (2) for k, then use

$$x(t) = k^{-1}\left[\cosh(kt) - \cosh(kb)\right] + A, \qquad -b \le t \le b.$$

Note that the function $k \mapsto 2k^{-1}\sinh(kb)$ on the right side in (2) is even, increasing in the interval $k \ge 0$, and has the limit $2b$ as $k \to 0^+$. As anticipated above, there is no solution if $\gamma < 2b$ and some kind of degenerate solution when $\gamma = 2b$.

Note that the version of (2) arising in the symmetric case can be rearranged as

$$\frac{\gamma}{2b} = \frac{\sinh(kb)}{kb} = \frac{\sinh(z)}{z}, \qquad z = kb.$$

Thus a systematic method for inverting the function $z^{-1}\sinh(z)$ is all we would need to settle the symmetric case for any real $b > 0$.

G. Wishful Thinking

2026-02-11

"In the fields of observation, chance only favours the mind which is prepared ..."
– Louis Pasteur, 1854

Let X be a real vector space, with a subset S. Suppose functionals $\Lambda, \Gamma \colon S \to \mathbb{R}$ are given, together with a constant γ, and we face the optimization problem

$$\min_{x\in S}\left\{\Lambda[x] \,:\, \Gamma[x] = \gamma\right\}. \tag{P}$$

Here are two lines of reasoning that produce results that resemble the Lagrange Multiplier Rule. Their limitations are subtle, but important.

Optimistic Alternative 1.

Theorem. *Suppose $\widehat{x} \in S$ is admissible in (P). If there is some $\lambda \in \mathbb{R}$ such that*

$$x = \widehat{x} \text{ minimizes } \Lambda[x] + \lambda\Gamma[x] \tag{$*$}$$

then $\widehat{x}$ gives a minimum in (P).

Proof. For any $x \in S$ obeying $\Gamma[x] = \gamma$, line $(*)$ gives the central inequality below:

$$\begin{aligned}
\Lambda[x] &= \Lambda[x] + \lambda(\Gamma[x] - \gamma) \\
&= (\Lambda[x] + \lambda\Gamma[x]) - \lambda\gamma \\
&\ge (\Lambda[\widehat{x}] + \lambda\Gamma[\widehat{x}]) - \lambda\gamma \\
&= \Lambda[x] + \lambda(\Gamma[x] - \gamma) \\
&= \Lambda[\widehat{x}].
\end{aligned}$$

This is the desired result. ////

Implementation. If the set S is a shifted copy of some vector space X, and both Λ and Γ are Gâteaux differentiable on S, then every $\widehat{x}$ satisfying $(*)$ will be found among solutions of

$$0 = D(\Lambda + \lambda\Gamma)[\widehat{x}] \qquad \text{on } X. \tag{$**$}$$

So we look for points $\widehat{x}$ in S and constants $\lambda \in \mathbb{R}$ for which $\Gamma[\widehat{x}] = \gamma$ and $(**)$ holds. If we find one that actually minimizes $\widetilde{\Lambda} = \Lambda + \lambda\Gamma$, it is guaranteed to be a minimizer for (P).

Limitations. The Theorem above provides a simple test that, if passed, guarantees that an admissible point $\widehat{x}$ solves the given problem. But *there is no guarantee that the test will identify every solution.* The next example illustrates how the procedure just outlined can break down, leaving its user with a page filled with calculations having no obvious relevance to anything.

Example. Consider this problem with $S = X = \mathbb{R}^2$, where typical points have the form $x = (s, t)$:

$$\min\left\{\ell(s,t) \overset{\text{def}}{=} s + t \,:\, 1 = g(s,t) \overset{\text{def}}{=} st,\ s > 0\right\}.$$

The suggested procedure is to look for feasible points (s, t) where some λ makes $\nabla\widetilde{\ell} = 0$ for $\widetilde{\ell} = \ell + \lambda g = s + t + \lambda st$. Here

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \nabla\widetilde{\ell}(s,t) = \begin{bmatrix} 1 + \lambda t \\ 1 + \lambda s \end{bmatrix} \iff \lambda s = \lambda t = -1.$$

Clearly $\lambda = 0$ can't work, so we focus on points where $s = t$. Then the constraint forces $1 = st = s^2$, so $s = 1$, and our unique point of interest is $\widehat{x} = (1, 1)$, with corresponding multiplier $\lambda = -1$. It's easy to see geometrically, or by substituting $s = 1/t$ and using calculus, that $\widehat{x}$ really is this problem's global solution. However, the theorem above won't confirm this because $\widehat{x}$ **does not minimize** the function

$$(s,t) \mapsto s + t + \lambda st = s + t - st.$$

It's a critical point, but the inputs $s = 1 + r$ and $t = 1 + \sigma r$ give

$$s + t - st = (1 + r) + (1 + \sigma r) - (1 + r + \sigma r + \sigma r^2) = 1 - \sigma r^2.$$

Thus the point $\widehat{x}$ maximizes $\widetilde{\ell}$ along the lines where $\sigma > 0$ and minimizes $\widetilde{\ell}$ along the lines where $\sigma < 0$: $\widehat{x}$ is a *saddle point,* not a minimizer. So this simple problem has a solution, but the procedure outlined above won't suffice to find it. In fact, this process alone provides no firm reason for thinking the point $\widehat{x}$ is preferable to any other. ////

Optimistic Alternative 2. Extend the given problem (P) by defining the *value function*

$$V(\gamma) \overset{\text{def}}{=} \min\left\{\Lambda[x] \,:\, \Gamma[x] = \gamma\right\}. \tag{$P(\gamma)$}$$

Imagine starting with a nominal problem in which $\gamma = \widehat{\gamma}$ happens to have a minimizer $\widehat{x}$, and we are interested in what would change if the nominal value of the constraint was allowed to deviate from $\widehat{\gamma}$. The baseline situation is

$$\Lambda[\widehat{x}] = V(\widehat{\gamma}) = V(\Gamma[\widehat{x}]) \tag{1}$$

Now fix any y in X, and put $\gamma = \Gamma[y]$ into the definition above. Then the minimization problem sets up a contest between all x that satisfy $\Gamma[x] = \Gamma[y]$. Clearly one of the choices for x compatible with the constraints is our starting function y itself, so $V(\gamma) \leq \Lambda[y]$. (The input y is admissible, but there is no particular reason to expect it to actually provide a minimum.) Recalling the choice of γ, we have

$$V(\Gamma[y]) = V(\gamma) \leq \Lambda[y], \qquad \forall y \in X. \tag{2}$$

(The same reasoning applies to each fixed $y \in X$ independently, and that's expressed by the quantifiers in (2).)

Taken together, lines (1)–(2) imply

$$[0 =]\ \Lambda[\widehat{x}] - V(\Gamma[\widehat{x}]) \leq \Lambda[y] - V(\Gamma[y]), \qquad \forall y \in X.$$

That is, the function $y \mapsto (\Lambda - V \circ \Gamma)[y]$ has an unconstrained minimum over X at the point $y = \widehat{x}$. Therefore

$$0 = D(\Lambda - V \circ \Gamma)[\widehat{x}] = D\Lambda[\widehat{x}] - V'(\Gamma[\widehat{x}])D\Gamma[\widehat{x}].$$

Defining $\lambda = -V'(\widehat{\gamma})$, we have the same conclusion derived carefully in previous sections: *If $\widehat{x}$ is a minimizer in problem $P(\widehat{\gamma})$, then there exists some constant λ such that*

$$0 = D(\Lambda + \lambda\Gamma)[\widehat{x}].$$

Limitations. The weak point in the reasoning above is the *implicit assumption* that the function V is differentiable at $\widehat{\gamma}$. Independent effort using quite different methods is required to put a solid foundation under this approach. (Success is possible, however.)

Interpretation. In normal problems with a unique minimizer $\widehat{x}$ and a unique corresponding Lagrange multiplier λ, this alternative derivation suggests a new interpretation for the number λ. Namely, $-\lambda$ tells the rate of change of the objective value with respect to the constraint levels.

In Economics, problem $P(\gamma)$ might be faced by some business owner, whose goal is to minimize some financial loss (measured in dollars). Then the negative loss, $-V$, would be the owner's *profit*. Meanwhile, the number γ might represent the available amount of some commodity important in the business, perhaps in units of kg. Then $d(-V)/d\gamma$ tells the rate of change in the owner's profit as the available resource increases. Typically, having more resources drives profits up, so the multiplier $\lambda = d(-V)/d\gamma > 0$. The units of λ are *dollars per kilogram*, and the standard name for it in economics is the *shadow price*. It tells the value of resource γ to the business owner interested in problem $P(\gamma)$. The owner will compare λ with the cost of γ on the open market. If the market price is below λ, the owner can buy some γ and increase their profit by more than they paid; if the market price is above λ, selling some γ in the market will reduce profit from the optimization, but that loss will be more than covered by the profit from selling the raw material.

Summary: Problem-Solving Steps. Given an isoperimetric problem of the form above, ...

1. Find all admissible extremals for *a linear combination of the constraint integrands G_j alone*. These are the problem's "abnormal extremals," since they satisfy the conclusions of the main theorem with $\lambda_0 = 0$. In most problems this set of arcs is empty, or reveals that the set if competing arcs is somehow degenerate.
2. Find all admissible $\widehat{x}$ that, together with some constant $\lambda \in \mathbb{R}^m$, satisfy IEL for $\widetilde{L} = L + \sum_j \lambda_j G_j$.
3. If the problem has a solution, it is guaranteed to appear on the list of arcs identified in Steps 1 and 2. (We earned this knowledge through hard work in Sections B–C.)
4. For each pair $(\widehat{x}, \lambda)$ found in Step 2, take a chance: test whether the mapping $x \mapsto \Lambda + \lambda\Gamma$ is minimized over S by $\widehat{x}$. If so, you win: $\widehat{x}$ gives the minimum in the stated problem. (If not, it is still possible that $\widehat{x}$ is a minimizer ... but unfortunately the super-easy proof of Wishful Thinking Option 1 is not powerful enough to detect that.)