Variational Calculus and Optimal Control

Minghao Mou

July 2023

Contents

1 Linear Space and Gateaux Variations

1.1 Real Linear Spaces

Example 1.1. The collection of real valued functions on a nonempty set S is a real linear space.

Example 1.2. the collection of all d-dimensional real vector valued functions on this set S.

Example 1.3. If continuity is definable on S, then $C(S)$ $(= C^0(S))$, the set of continuous real valued functions on S, will be a real linear space.

Example 1.4. For each open subset D of Euclidean space and each $m = 1, 2, ..., C^m(D)$, the set of functions on D having continuous partial derivatives of order $\leq m$, is a real linear space.

Example 1.5. If D is bounded with boundary ∂D, and $\overline{D} = D \bigcup \partial D$, then $C^m(\overline{D})$, the subset of $C^m(D) \bigcup C(\overline{D})$ consisting of those functions whose partial derivatives of order $\leq m$ each admit continuous extension to D, is a real linear space.

Example 1.6. $\mathcal{D} := \{y \in C[a, b] : y(a) = 0 \ \& \ y(b) = 1\}$ is not a linear space.

1.2 Functions from Linear Spaces

Example 1.7. When $f \in C([a, b] \times \mathbb{R}^2)$, then

$$F(y) := \int_a^b f(x, y(x), y'(x))dx$$

is defined on $\mathcal{Y} := C^1[a, b]$, since for each $y \in \mathcal{Y}$, the composite function

$$f[y(x)] = f(x, y(x), y'(x)) \in C[a, b].$$

However, if $f \in C([a, b] \times D)$ where D is a domain in $\mathbb{R}^2$, then F is defined only on a subset of

$$\mathcal{D}^\star := \{y \in C^1[a, b] : (y(x), y'(x)) \in D, \forall x \in [a, b]\}.$$

1.3 Fundamentals of Optimization

Lemma 1.8 (Optimality Condition). $y_0 \in \mathcal{D}$ minimizes J on $\mathcal{D}$ if and only if

$$J(y_0 + v) - J(y_0) \geq 0, \quad \forall y_0 + v \in \mathcal{D},$$

with equility holds if and only if $v = 0$.

Proposition 1.9. y_0 minimizes J on $\mathcal{D}$ (uniquely) if and only if for constants c_0 and $c \neq 0$, y_0 minimizes $c^2 J + c_0$ on $\mathcal{D}$ (uniquely).

1.3.1 Constrained Case

Proposition 1.10. If functions J and $G_1, G_2, ..., G_N$ are defined on $\mathcal{D}$, and for some constants $\lambda_1, \lambda_2, ..., \lambda_N$, y_0 minimizes $\tilde{J} := J + \lambda_1 G_1 + ... + \lambda_N G_N$ on $\mathcal{D}$ (uniquely), then y_0 minimizes J on $\mathcal{D}$ (uniquely) when further restricted on the set $G_{y_0} := \{y \in \mathcal{D} : G_j(y) = G_j(y_0), j = 1, 2, ..., N\}$.

Corollary 1.11. y_0 of Proposition 1.10 minimizes J on $\mathcal{D}$ (uniquely) when restricted to the set

$$G^{\star}_{y_0} := \{y \in \mathcal{D} : \lambda_j G_j(y) \leq \lambda_j G_j(y_0), j = 1, 2, ..., N\}.$$

Remark 1.12. These results illustrate an important principle: The solution to one minimization problem may also provide a solution for other problems.

Proposition 1.13. Suppose $f := f(x, y, z)$ and $g := g(x, y, z)$ are continuous on $[a, b] \times \mathbb{R}^2$ and there is a function $\lambda \in C[a, b]$, for which y_0 minimizes $\tilde{F}(y) := \int_a^b \tilde{f}[y(x)]dx$ on $\mathcal{D} \subset C^1[a, b]$ (uniquely) where $\tilde{f} := f + \lambda g$. Then y_0 minimizes $F(y) := \int_a^b f[y(x)]dx$ on $\mathcal{D}$ (uniquely) under the inequality constraint

$$\lambda(x) g[y(x)] \leq \lambda(x) g[y_0(x)], \quad \forall x \in [a, b]. \tag{1}$$

1.4 The Gateaux Variations

A decisive role in the optimization of a real valued function on a subset of $\mathbb{R}^n$ is played by its partial derivatives-or more generally by its directional derivatives-if they exist. When J is a real valued function on a subset of a linear space all, then it is not evident how to define its partial derivatives (unless all can be assigned a distinguished coordinate system). However, a definition for its directional derivatives is furnished by a straightforward generalization of that in $\mathbb{R}^n$:

Definition 1.14 (Gateaux Variation)**.** For $y, v \in \mathcal{Y}$:

$$\delta J(y; v) := \lim_{\epsilon \to 0} \frac{J(y + \epsilon v) - J(y)}{\epsilon} = \frac{\partial}{\partial \epsilon} J(y + \epsilon v)\Big|_{\epsilon=0}, \tag{2}$$

assuming that this limit exists, is called the *Gateaux variation* of J at y in the direction v.

Example 1.15. If $J = f \in C^1[a, b]$ and $Y, V \in \mathcal{Y} := \mathbb{R}^n$. Then,

$$\delta f(Y; V) = \lim_{\epsilon \to 0} \frac{f(Y + \epsilon V) - f(Y)}{\epsilon}, \tag{3}$$

is just the directional derivative of f when V is a unit vector. Thus we have that

$$\delta f(Y; V) = \nabla f(Y) \cdot V,$$

and this holds for all $V \in \mathcal{Y}$.

2 Minimization of Convex Functions

2.1 Convex Functions

Definition 2.1 (Convexity)**.** A real valued function J defined on a set $\mathcal{D}$ in a linear space $\mathcal{Y}$ is said to be (strictly) convex on $\mathcal{D}$ provided that when y and $y+v \in \mathcal{D}$ then $\delta J(y;v)$ is defined and $J(y+v) - J(y) \geq \delta f(y;v)$ (with equility holds if and only if $v = 0$).

Proposition 2.2. If J_1 and J_2 are convex functions on $\mathcal{D}$ then for each $c \in \mathbb{R}$, $c^2 J_1$ and $J_1 + J_2$ are also convex. Moreover, the latter is strictly convex if J_1 is strictly convex.

Proposition 2.3 (Unconstrained Optimality Condition)**.** If J is (strictly) convex on $\mathcal{D}$ then each $y_0 \in \mathcal{D}$ for which $\delta f(y_0; v) = 0, \forall y_0 + v \in \mathcal{D}$ minimizes J on $\mathcal{D}$.

2.2 Convex Integral Functions

If $f = f(x,y,z)$ and its partial derivatives f_y, f_z are defined and continuous on $[a,b] \times \mathbb{R}^2$, we know that the integral function

$$F(y) = \int_a^b f(x, y(x), y'(x))dx = \int_a^b f[y(x)]dx,$$

has $\forall y, v \in C^1[a,b]$, the variation

$$\delta F(y;v) := \int_a^b \left(f_y[y(x)]v(x) + f_z[y(x)]v'(x) \right) dx. \tag{4}$$

Convexity of F requires that $\forall y, y+v \in C^1[a,b]$

$$F(y+v) - F(y) \geq \delta F(y;\delta).$$

This would follow from the corresponding pointwise inequality between the integrands in the last expression; i.e. if for each $x \in (a,b)$:

$$f[y(x) + v(x)] - f[y(x)] \geq f_y[y(x)]v(x) + f_z[y(x)]v'(x), \tag{5}$$

or

$$\begin{aligned} f(x, y+v, z+w) - f(x,y,z) \geq f_y(x,y,z)v + f_z(x,y,z)w, \\ \forall (x,y,z), (x, y+v, z+w) \in (a,b) \times \mathbb{R}^2. \end{aligned} \tag{6}$$

The above inequality simply states that $f = f(x, y, z)$ is convex when x is held fixed. This restricted or partial convexity essential to our development is expressed and extended in the following which uses for illustration a function defined on a subset of $\mathbb{R}^3$.

Definition 2.4 (Strong Convexity)**.** $f(x, y, z)$ is said to be (strongly) convex on $S \subset \mathbb{R}^3$ if $f = f(x, y, z)$ and its partial derivatives f_y and f_z are defined and continuous on this set and there they satisfy the inequality:

$$\begin{aligned} f(x, y+v, z+w) - f(x, y, z) \geq f_y(x, y, z)v + f_z(x, y, z)w, \\ \forall (x, y, z), (x, y+v, z+w) \in S, \end{aligned}$$

(with equaility holds at (x, y, z) if and only if $v = 0$ or $w = 0$).

Theorem 2.5 (Solution to Convex Program)**.** Let $D \subset \mathbb{R}^2$ be a domain and for any given a_1, b_1, set

$$\mathcal{D} := \{y \in C^1[a, b] : y(a) = a_1, y(b) = b_1; (y(x), y'(x)) \in D\}.$$

If $f(\underline{x}, y, z)$ is (strongly) convex on $[a, b] \times D$, then

$$F(y) := \int_a^b f[y(x)]dx,$$

is (strictly) convex on $\mathcal{D}$. Hence each $y \in \mathcal{D}$ for which

$$\frac{\partial}{\partial x} f_z[y(x)] = f_y[y(x)] \tag{7}$$

on $[a, b]$, minimizes F on $\mathcal{D}$ (uniquely).

Proof. By (strong) convexity of $f(\underline{x}, y, z)$ one can show the (strict) convexity of $F(y)$. Note also that if $\frac{\partial}{\partial x} f_z[y(x)] = f_y[y(x)]$, then

$$\delta F(y; v) = f_z[y(x)]v(x)\Big|_a^b = 0, \quad \forall y, y+v \in \mathcal{D}.$$

Thus by Proposition 1.10, y such that $\frac{\partial}{\partial x} f_z[y(x)] = f_y[y(x)]$ minimizes $F(y)$ on $\mathcal{D}$. □

Theorem 2.6. Let I be an interval and set

$$\mathcal{D} := \{y \in C^1[a,b] : y(a) = a_1, y(b) = b_1; y'(x) \in I\}.$$

Then, if $f(\underline{x}, z)$ is (strongly) convex on $[a,b] \times I$, each $y \in \mathcal{D}$ which makes $f_z(x, y'(x)) = \text{const.}$ on (a,b) minimizes $F(y) := \int_a^b f(x, y'(x))dx$ on $\mathcal{D}$ (uniquely).

Proof. It directly follows from Theorem 2.5 and the fact that $\frac{\partial}{\partial x} f_z[y(x)] = f_y[y(x)] = 0$ implies $f_z(x, y'(x)) = \text{const.}$ $\square$

Corollary 2.7. If $f = f(z)$ is (strictly) convex on I, then $y_0(x) = \frac{b_1 - a_1}{b-a}(x - a) + a_1$ minimizes $F(y) := \int_a^b f(y'(x))dx$ on $\mathcal{D}$ (uniquely).

Free End-point Problem When we examine the proof of Theorem 2.5 we see that the end-point specification was used only to conclude that the constant $v^2(x) = 0$ and that $f_z[y(x)]v(x)\Big|_a^b = 0$. Hence these end-point conditions on y may be relaxed, if suitable compensation is made in $f_z[y(x)]$.

Proposition 2.8. Let D be a domain in $\mathbb{R}^2$ and suppose that $f(\underline{x}, y, z)$ is (strongly) convex on $[a,b] \times D$. Then each solution $y_0 \in \mathcal{D} := \{y \in C^1[a,b] : (y(x), y'(x)) \in D\}$ of the differential equation $\frac{\partial}{\partial x} f_z[y(x)] = f_y[y(x)]$ minimizes $F(y) := \int_a^b f[y(x)]dx$

1. on $\mathcal{D}^b := \{y \in \mathcal{D} : y(a) = y_0(a)\}$, if $f_z[y_0(b)] = 0$ (uniquely).
2. on $\mathcal{D}$, if $f_z[y_0(a)] = f_z[y_0(b)] = 0$, (uniquely within an additive constant).

2.3 (Strongly) Convex Functions

In order to apply the results of the previous section, we require a supply of functions which are (strongly) convex. In this section techniques for recognizing such convexity will be developed. We begin with the simpler case $f = f(x, z)$.

Proposition 2.9. If $f = f(x, z)$ and f_{zz} are continuous on $[a,b] \times I$ and for each $x \in [a,b]$, $f_{zz}(x, z) > 0$ (except possibly for a finite set of z values), then $f(\underline{x}, z)$ is strongly convex on $[a,b] \times I$.

Lemma 2.10. 1. The sum of a (strongly) convex function and one (or more) convex functions is again (strongly) convex.

2. The product of a (strongly) convex function $f(\underline{x}, y, z)$ by a continuous function $(p(x) > O)p(x) \geq 0$ is again (strongly) convex on the same set.

3. $f(\underline{x}, y, z) = \alpha(\underline{x}) + \beta(\underline{x})y + \gamma(\underline{x})z$ is (only) convex for any continuous functions α, β, and γ.

4. Each (strongly) convex function $f(\underline{x}, z)$ (or $f(\underline{x}, y)$ is also (strongly) convex when considered as a function $f(\underline{x}, y, z)$ on an appropriate set.

2.4 Minimization with Convex Constraints

Convexity may also be of advantage in establishing the minima of functions J that are constrained to the level sets of other functions G.

Theorem 2.11. If D is a domain in $\mathbb{R}^2$, such that for some constants $\lambda_j, j = 1, 2, ..., N$, $f(\underline{x}, y, z)$ and $\lambda_j g_j(\underline{x}, y, z)$ are convex on $[a, b] \times D$ (and at least one of these functions are strongly convex on this set), let

$$\tilde{f} := f + \sum_{j=1}^{N} \lambda_j g_j.$$

Then the solution y_0 of the differential equation

$$\frac{\partial}{\partial x}\tilde{f}_z[y(x)] = \tilde{f}_y[y(x)]$$

on (a, b), minimizes

$$F(y) := \int_a^b f[y(x)]dx$$

(uniquely) on

$$\mathcal{D} := \{y \in C^1[a, b] : y(a) = y_0(a), y(b) = y_0(b); (y(x), y'(x)) \in D\}$$

under the constraning relations

$$G_j(y) := \int_a^b g_j[y(x)]dx = G_j(y_0), \quad j = 1, 2, ..., N.$$

Proof. By construction $\tilde{f}(\underline{x}, y, z)$ is (strongly) convex on $[a,b] \times D$, so that by Theorem 2.5, y_0 minimizes

$$\tilde{F}(y) = \int_a^b \tilde{f}[y(x)]dx = F(y) + \sum_{j=1}^{N} G_j(y)$$

(uniquely) on $\mathcal{D}$. □

3 The Lemmas of Lagrange and du Bois-Reymond

Lemma 3.1 (du Bois-Reymond)**.** If $h \in C[a,b]$ and $\int_a^b h(x)v'(x)dx = 0, \forall v \in \mathcal{D}_0 := \{v \in C^1[a,b] : v(a) = v(b) = 0\}$, then $h =$ const. on $[a,b]$.

Proof. Define $v(x) := \int_a^x (h(t) - c)dx$, where $c = \frac{1}{b-a}\int_a^b h(t)dt$. We know that $v \in C^1[a,b]$, $v'(x) = h(x) - c$ on (a,b), and $v(a) = v(b) = 0$. Therefore, $v' \in \mathcal{D}_0$. From the hypothesis we have $0 \le \int_a^b (h(t)-c)^2 dt = \int_a^b h(x)v'(x)dx - cv(x)\big|_a^b = 0$. Therefore, $h \equiv c$. □

Proposition 3.2 (Generalized du Bois-Reymond)**.** If $h \in C[a,b]$ and for some $m = 1, 2, ...$

$$\int_a^b h(x)v^{(m)}(x)dx = 0, \quad \forall v \in \mathcal{D}_0,$$

where

$$\mathcal{D}_0 := \{v \in C^m[a,b] : v^{(k)}(a) = v^k(b) = 0, k = 1, 2, ..., m-1\},$$

then on $[a,b]$, h is a polynomial with degree $< m$.

Proposition 3.3. If $g, h \in C[a,b]$ and $\int_a^b [g(x)v(x) + h(x)v'(x)]dx = 0, \forall v \in \mathcal{D}_0 := \{v \in C^1[a,b] : v(a) = v(b) = 0\}$, then $h \in C^1[a,b]$ and $h' = g$.

Proof. Define $G(x) := \int_a^x g(t)dt$ for $x \in [a,b]$. Then $G \in C^1[a,b]$ and $G' = g$. Integration by parts yields:

$$0 = \int_a^b [g(x)v(x) + h(x)v'(x)]dx = \int_a^b [h(x) - G(x)]v'(x)dx + G(x)v(x)\Big|_a^b,$$

for all $v \in \mathcal{D}_0$. Therefore, by Lemma 3.1, $h(x) = G(x) +$ const. and thus $h \in C^1[a,b]$ and $h' = g$. □

Corollary 3.4. If $g \in C[a,b]$ and $\int_a^b g(x)v(x)dx = 0, \forall v \in \mathcal{D}_0 := \{v \in C^1[a,b] : v(a) = v(b) = 0\}$, then $g \equiv 0$ on $[a,b]$.

Lemma 3.5 (Lagrange)**.** if $g \in C[a,b]$ and for some $m = 0, 1, 2, ...,$

$$\int_a^b g(x)v(x)dx = 0,$$

for all $v \in \mathcal{D}_0 := \{v \in C^m[a,b] : v^{(k)}(a) = v^{(k)}(b) = 0, k = 0, 1, ..., m-1\}$, then $g \equiv 0$ on $[a,b]$.

Proposition 3.6. If $d = 2, 3, ...$ and for $G, H \in (C[a,b])^d$, we have

$$\int_a^b [G(x)V(x) + H(x)V'(x)]dx = 0,$$
$$\forall v \in \mathcal{D}_0 := \{V \in (C^1[a,b])^d : V(a) = V(b) = 0\},$$

then $H \in (C^1[a,b])^d$ and $H' = G$.

Corollary 3.7. When $H \in (C[a,b])^d$ and $\int_a^b H(x)V'(x)dx = 0, \forall v \in \mathcal{D}_0$, then $H(x) = \text{const.} \in \mathbb{R}^d$.

4 Local Extrema in Normed Linear Space

In $\mathbb{R}^n$, it is possible to give conditions which are necessary in order that a function f have a local extremal value on a subset D, expressed in terms of the vanishing of its gradient ∇f. In this chapter, we shall obtain analogous variational conditions which are necessary to characterize local extremal values of a function J on a subset $\mathcal{D}$ of a linear space $\mathcal{Y}$ supplied with a norm which assigns a "length" to each $y \in \mathcal{Y}$.

4.1 Norms for Linear Spaces

[1]: Section 5.1

4.2 Normed Linear Spaces: Convergence and Compactness

[1]: Section 5.2

4.3 Continuity

[1]: Section 5.3

Lemma 4.1 (Uniform Continuity)**.** If K is a compact set in a normed linear space $(\mathcal{Y}, \|\cdot\|)$, then a continuous function $F : K \to \mathbb{R}$ is uniformly continuous on K; i.e. given $\epsilon > 0$, $\exists \delta > 0$ such that $y, \tilde{y} \in K$ and $\|y - \tilde{y}\| < \delta \Rightarrow |F(y) - F(\tilde{y})| < \epsilon$.

Example 4.2. When $f \in C([a, b] \times \mathbb{R}^2)$, the function

$$F(y) := \int_a^b f[y(x)]dx = \int_a^b f(x, y(x), y'(x))dx$$

is defined $\forall y \in \mathcal{Y} := C^1[a, b]$ and is continuous with respect to the maximum norm.

Proposition 4.3. A continuous real-valued function J on a compact subset K of a normed linear space $(\mathcal{Y}, \|\cdot\|)$ assumes both maximum and minimum values at points in K, In particular, these values are finite.

4.4 (Local) Extremal Points

[1]: Section 5.4

4.5 Necessary Conditions: Admissible Directions

In minimizing a real valued function J over $\mathcal{D} \subset \mathcal{Y}$, where $(\mathcal{Y}, \|\cdot\|)$ is a normed linear space, it is natural to consider for each $y \in \mathcal{D}$ those directions $v \in \mathcal{Y}$ in which the restricted function $J|_{\mathcal{D}}$ admits variation at y; i.e., we wish to distinguish those directions v for which:

(1) $y + \epsilon v \in \mathcal{D}$ for any ϵ sufficiently small; and

(2) $\delta J(y; v)$ exists.

Such directions will be termed *admissible* at y for $\mathcal{D}$, or $\mathcal{D}$-admissible at y (for J). Observe that if v is $\mathcal{D}$-admissible at y, then so is each scalar multiple cv for $c \in \mathbb{R}$; 0 is always admissible.

Proposition 4.4 (First-Order Necessary Condition)**.**] In a normed linear space $(\mathcal{Y}, \|\cdot\|)$, if $y_0 \in \mathcal{D} \subset \mathcal{Y}$ is a (local) extremal point for a real-valued function J in $\mathcal{D}$, then $\delta J(y_0; v) = 0$ for all $\mathcal{D}$-admissible directions at y_0.

Our hope is that there will be "enough" admissible directions so that the condition $\delta J(y_0; v) = 0$ can determine y_0. Observe, though, that with this condition alone we cannot distinguish between a local maximum and a local minimum point-or between a local minimum point and a global minimum point. Moreover, as in $\mathbb{R}^d$, we must admit the possibility of stationary points (such as saddle points) which satisfy this condition but may be neither local maximum points nor local minimum points.

4.6 The Frechet Derivative

As we have seen, the Gateaux variation in a normed linear space is analogous to the directional derivative in $\mathbb{R}^n$. In particular, without further information, we cannot expect to use these variations to provide a good approximation to a function which has them-except, of course. in each separate direction. For this purpose in $\mathbb{R}^n$, we required that the function satisfy the stronger requirement of differentiability, and we shall simply lift the definition employed there, together with the associated terminology, to our normed linear space $(\mathcal{Y}, \|\cdot\|)$. In $\mathbb{R}^n$ with the Euclidean norm $|\cdot|$, a real valued function f is said to be differentiable at $y_0 \in \mathbb{R}^n$ provided that it is defined in a sphere $S(y_0)$ and there

$$f(y) = f(y_0) + l(y - y_0) + |y - y_0|\epsilon(y - y_0),$$

where $\epsilon(y - y_0)$ is a function with zero limit as $y - y_0 \to 0$, and l is the continuous linear function defined on $\mathbb{R}^n$ by $l(v) := \nabla f(y_0) \cdot v$.

Definition 4.5 ((Frechet) Differentiablity)**.** In a normed linear space $(\mathcal{Y}, \|\cdot\|)$, a real-valued function J is said to be (Frechet) differentiable at $y_0 \in \mathcal{Y}$ if J is defined in a sphere $S(y_0)$ and there exists a continuous linear function $L : \mathcal{Y} \to \mathbb{R}$ such that

$$J(y) = J(y_0) + L(y - y_0) + \|y - y_0\|\epsilon(y - y_0), \tag{8}$$

where $\epsilon(y - y_0)$ is a real-valued function which has zero limit as $y - y_0 \to 0$.

Proposition 4.6. If J is (Frechet) differentiable at y_0, then J has Gateaux variation $\delta J(y_0; v) = L(v)$ in each direction $v \in \mathcal{Y}$.

It follows that the linear function L of the definition is uniquely determined. It is denoted $J'(y_0)$ and called the *Frechet derivative* of J at y_0.

Proposition 4.7. In a normed linear space $(\mathcal{Y}, \|\cdot\|)$ if a real-valued function J is differentiable at $y_0 \in \mathcal{Y}$, then it is continuous at y_0.

As in $\mathbb{R}^n$, the converses of these propositions need not hold. Continuous functions are seldom differentiable. Moreover, if J admits the Gateaux variation $\delta J(y_0; v)$ in each direction $v \in \mathcal{Y}$, the resulting function of v may be neither linear nor continuous-and even these properties may not suffice for differentiability. Some additional conditions are required.

Theorem 4.8. In a normed linear space $(\mathcal{Y}, \|\cdot\|)$, if a real-valued function J has at each $y \in S_r(y_0)$ Gateaux variation $\delta J(y; v), \forall v \in \mathcal{Y}$ and

(a) $\delta J(y_0; v)$ is linear and continuous in v;

(b) $|\delta J(y; v) - \delta J(y_0; v)| \to 0$ as $y \to y_0$ uniformly for all $u \in B := \{u \in \mathcal{Y} : \|u\| = 1\}$.

Conditions (a) and (b) also imply a weak continuity of δJ at y_0 in the sense of the following:

Definition 4.9 (Weak Continuity)**.** In a normed linear space $(\mathcal{Y}, \|\cdot\|)$, the Gateaux variations $\delta J(y; v)$ of a real-valued function J are said to be *weakly continuous* at $y_0 \in \mathcal{Y}$ if for all $v \in \mathcal{Y}$, $\delta J(y; v) \to \delta J(y_0; v)$ as $y \to y_0$.

Proposition 4.10. When $f = f(x, y, z)$, f_y, and f_z are all in $C([a, b] \times \mathbb{R}^2)$ then $F(y) := \int_a^b f(x, y(x), y'(x))dx$ is differentiable and has weakly continuous variations at each $y_0 \in \mathcal{Y} := C^1[a, b]$ in the maximum norm.

4.7 Constrained Optimization and Lagrangian Multipliers

In this section we will develop the method of Lagrangian multipliers for characterizing the local extrema of a function J in a normed linear space when restricted to one or more level sets of other such functions. In this context, the level sets involved are called constraints, and the equations defining the sets are referred to as constraining relations.

To motivate the ensuing development, we consider first the problem of characterizing a (local) extremal point y_0 of a real valued function J in a normed linear space $(\mathcal{Y}, \|\cdot\|)$ when constrained to a level set of a real valued function G. If there is a pair of directions v, w for which there exist pairs of scalars $(\underline{r}, \underline{s})$ and $(\overline{r}, \overline{s})$ as small as we please such that $J(\underline{y}) < J(y_0) < J(\overline{y})$ while $G(\underline{y}) = G(y_0) = G(\overline{y})$, where $\underline{y} := y_0 + \underline{r}v + \underline{s}w$ and $\overline{y} := y_0 + \overline{r}v + \overline{s}w$, then y_0 cannot be a local extremal point.

We shall now assume that both J and G are defined in a neighborhood of y_0, and consider for fixed directions v, w the auxiliary functions

$$\rho = \mathcal{J}(r, s) := J(y_0 + rv + sw), \quad \sigma = \mathcal{G}(r, s) := G(y_0 + rv + sw),$$

which are defined in some two-dimensional neighborhood of the origin in $\mathbb{R}^2$. The pair of these functions $F := (\mathcal{J}, \mathcal{G})$ maps this neighborhood into $\mathbb{R}^2$ which contains the point $(\rho_0, \sigma_0) = (\mathcal{J}(0, 0), \mathcal{G}(0, 0)) = (J(y_0), G(y_0))$. If it also contains a full neighborhood of (ρ_0, σ_0), then there are preimage points $(\underline{r}, \underline{s})$ and $(\overline{r}, \overline{s})$ and associated $\underline{y}, \overline{y}$ for which $J(\underline{y}) < J(y_0) < J(\overline{y})$ while $G(\underline{y}) = G(y_0) = G(\overline{y})$.

Finally, to have $(\underline{r}, \underline{s}), (\overline{r}, \overline{s})$ as near $(0, 0)$ as we please we would require that each small neighborhood of $(0, 0)$ map onto a set which contains a full neighborhood of (ρ_0, σ_0). All of this is assured if the mapping F has an inverse defined in a neighborhood of (ρ_0, σ_0) which is continuous at (ρ_0, σ_0).

The simplest conditions which provide this continuous local inverse are well known, and form the content of the inverse function theorem which we state without proof.

Theorem 4.11 (Inverse Function Theorem)**.** For $X_0 \subset \mathbb{R}^d$ and $\tau > 0$, if a vector-valued function $F : S_\tau(X_0) \to \mathbb{R}^d$ has continuous first-order derivatives in each component with nonvanishing Jacobian determinant at X_0, then F provides a continuously invertible mapping between a neighborhood of X_0 and a region containing a full neighborhood of $F(X_0)$.

Proposition 4.12. In a normed linear space $(\mathcal{Y}, \|\cdot\|)$, if real-valued functions J and G are defined in a neighborhood of y_0 and have there in any pair of (fixed) directions v, w, Gateaux variations are continuous in this neighborhood and satisfy the Jacobian condition

$$\begin{vmatrix} \delta J(y_0; v) & \delta J(y_0; w) \\ \delta G(y_0; v) & \delta G(y_0; w) \end{vmatrix} \neq 0, \tag{9}$$

Then J cannot have a local extremal point at y_0 when constrained $G_{y_0} := \{y \in \mathcal{Y} : G(y) = G(y_0)\}$.

Proposition 4.13 (Lagrange First-order Necessary Condition)**.** In a normed linear space $(\mathcal{Y}, \|\cdot\|)$, if real-valued functions J and G are defined in a neighborhood of y_0, a local extremal point for J constrained to G_{y_0}, and have there weakly continuous Gateaux variations, then either

(a) $\delta G(y_0; w) = 0, \forall w \in \mathcal{Y}$; or

(b) There exists $\lambda \in \mathbb{R}$ such that $\delta(J + \lambda G)(y_0; v) = 0, \forall v \in \mathcal{Y}$.

Proposition 4.14 (Lagrange First-order Necessary Condition)**.** In a normed linear space $(\mathcal{Y}, \|\cdot\|)$, if real-valued functions $J, G_1, G_2, ..., G_N$ are defined in a neighborhood of y_0, have $y_0 \in \mathcal{Y}$ as one local extremal point for J constrained to $G_{y_0} := \{y \in \mathcal{Y} : G_j(y) = G_j(y_0), j = 1, 2, ..., N\}$, and have weak continuous Gateaux variations, then either

(a) $$\begin{vmatrix} \delta G_1(y_0; v_1) & \delta G_1(y_0; v_2) & \cdots & \delta G_1(y_0; v_N) \\ \delta G_2(y_0; v_1) & \delta G_2(y_0; v_2) & \cdots & \delta G_2(y_0; v_N) \\ \vdots & \vdots & \vdots & \vdots \\ \delta G_1(y_0; v_1) & \delta G_1(y_0; v_2) & \cdots & \delta G_1(y_0; v_N) \end{vmatrix} \neq 0, \forall v_j \in \mathcal{Y}, j = 1, 2, ..., N; \text{ or}$$

(b) there exist $\lambda_j \in \mathbb{R}, j = 1, 2, ..., N$ such that $\delta(f + \sum_{j=1}^{N} \lambda_j G_j)(y_0; v) = 0, \forall v \in \mathcal{Y}$.

5 The Euler-Lagrange Equations

A point y_0 is a (local) minimizing function for F on $\mathcal{D}$ if, from Proposition 4.4,

$$\delta F(y_0; v) = 0, \quad \forall \mathcal{D}\text{-admissible directions of } F \text{ at } y_0.$$

When f is sufficiently differentiable, there are enough such directions to infer that on (a, b), y_0 is a solution of the first and second equations of Euler-Lagrange.

In this chapter, only those conditions necessary for a local extremum are considered. It should be noted, however, that the initial investigators in these fields, often regarded a function which satisfied the necessary conditions as the extremal function sought, and the practice continues today in elementary treatments of the subject. Throughout this chapter, we shall supply the space $C^1[a, b]$ with the maximum norm $\|y\|_M := \max_x(|y(x)| + |y'(x)|)$.

5.1 The First Equation: Stationary Functions

For simplicity, suppose initially that the function $f = f(x, y, z)$, together with its derivatives f_y and f_z, is continuous on $[a, b] \times \mathbb{R}^2$. Then for each $y \in \mathcal{Y} := C^1[a, b]$:

$$F(y) := \int_a^b f(x, y(x), y'(x))dx = \int_a^b f[y(x)]dx$$

is defined. F has in each direction v the Gateaux variation

$$\delta F(y; v) = \int_a^b \left[f_y(x)v(x) + f_{y'}(x)v'(x)\right] dx, \tag{10}$$

where we use the compressed notation $f_y(x) := f_y[y(x)]$ and $f_{y'}(x) := f_z[y(x)]$.

Proposition 5.1. If $y \in \mathcal{Y}$ makes $\delta F(y; v) = 0, \forall v \in \mathcal{D}_0 := \{v \in \mathcal{Y} : v(a) = v(b) = 0\}$, then $f_{y'} \in C^1$, and

$$\frac{d}{dx} f_{y'}(x) = f_y(x), \quad x \in (a, b), \tag{11}$$

so that

$$\delta F(y; v) = f_{y'}(x)v(x)\Big|_a^b, \forall v \in \mathcal{Y}. \tag{12}$$

Proof. The first assertion is a direct consequence of Proposition 3.3 with $g(x) := f_y(x)$ and $h(x) := f_{y'}(x)$. The second assertion follows by noticing that the integrand is merely $(d/dx)\,[f_{y'}(x)v(x)]$. $\square$

Equation (11) is the first differential equation of Euler and Lagrange.

Definition 5.2 (Stationary Function)**.** Each C^1 function y which satisfies the differential equation (11) on some interval will be called a *stationary function* for f (of x, y, and y').

Observe that we do not require that a stationary function satisfy any particular boundary conditions, although in each problem, we might be interested only in those which meet given boundary conditions. Note that certain functions f with their derivatives f_y and f_z are defined only for a restricted class of functions y, so that variation of F at y can be performed only for a reduced class of v. As the preceding discussion shows, when y is stationary and meets the restrictions, then $\delta F(y; v) = 0, \forall v \in \mathcal{D}_0$ for which the variation at y is defined. However, there may also be *nonstationary functions* η which make $\delta F(\eta; v) = 0$ for the reduced class of v, and these may provide the true extremals.

5.2 Special Cases of the First Equation

Although every C^1 function y is stationary for $f(x, y, z) = z$ or yz, in general, it is difficult to find *any* solutions for the first equation. However, when one or more of the variables off is not present explicitly, then we can at least obtain a first integral of the differential equation. We shall analyze three such cases in this section.

Case I: $f = f(z)$**.** The first equation becomes $(d/dx)f_{y'}(x) = 0$ since $f_y(x) \equiv 0$. Therefore, $f_z(y'(x)) =$ const. and the stationary functions y have derivatives y' which lie in the level sets of f_z. In particular, the linear functions, for which $y' =$ const., must be stationary.

Case II: $f = f(x, z)$**.** Again we have $f_y(x, z) = 0$ so that the stationary condition is $f_z(x, y'(x)) =$ const.

Case III: $f = f(y, z)$**.** When $y \in C^2$, we have

$$\frac{d}{dx} f(y(x), y'(x)) = f_y(x) y'(x) + f_z(x) y''(x).$$

Upon substitution and cancellation we obtain

$$\begin{aligned} &\frac{d}{dx}\left[f(y(x), y'(x)) - y'(x) f_{y'}(x)\right] \\ &\quad = \frac{d}{dx} f(y(x), y'(x)) - y''(x) f_{y'}(x) - y'(x) \frac{d}{dx} f_{y'}(x) \\ &\quad = -y'(x)\left[\frac{d}{dx} f_{y'}(x) - f_y(x)\right]. \end{aligned} \tag{13}$$

When y is stationary, the right hand side vanishes. Thus on each interval of stationarity of y we have

$$f(x) - y'(x) f_{y'}(x) = \text{const.} \tag{14}$$

Conversely, if Equation (14) holds on an interval in which y' does not vanish, then y is stationary. In this case stationarity is characterized by (14) which is a first integral of the first equation (11).

5.3 The Second Equation

When $f = f(x, y, z)$ is C^1 and y is a C^1 solution of the first equation (11) on $[a, b]$, then integration gives

$$f_{y'}(x) = \int_a^x f_y(t) dt + \text{const.} \tag{15}$$

When $y \in C^2$, then

$$\begin{aligned} \frac{d}{dx} f(x, y(x), y'(x)) &= \frac{d}{dx} f(x) + f_y(x) y'(x) + f_{y'}(x) y''(x) \\ &= f_x(x) + \frac{d}{dx}\left(y'(x) f_{y'}(x)\right). \end{aligned}$$

Thus, $(d/dx)[f(x) - y'(x) f_{y'}(x)] = 0$, or

$$f(x) - y'(x) f_{y'}(x) = \int_a^x f_x(t) dt + c_0, \quad c_0 \in \mathbb{R}.$$

This equation resembles (14), the integral form of the first, and, moreover, it does not exhibit explicitly the C^2 requirement on y used in its derivation. Hence we can hope to obtain it directly. This is indeed possible (for extremal functions) but it is surprisingly complicated to do so.

Let $F(y) := \int_a^b f[y(x)]dx$ and $\mathcal{D} := \{y \in C^1[a,b] : y(a) = a_1, y(b) = b_1\}$.

Proposition 5.3. If $f \in C^1([a,b] \times \mathbb{R}^2)$ and $y_0 \in \mathcal{D}$ is a local extremal function for F on $\mathcal{D}$, then on $[a,b]$, y_0 satisfies the second Euler-Lagrange equation

$$f(x) - y'(x)f_{y'}(x) = \int_a^x f_x(t)dt + c_0, \tag{16}$$

for some constant c_0.

Observe that when $f = f(y,z)$, a local extremal function $y \in \mathcal{D}$, must also satisfy the equation $(d/dx)[f(x) - y'(x)f_{y'}(x)] = 0$ without additional smoothness assumptions.

Lemma 5.4. When y is only stationary, this proof does not yield the second equation unless y is C^2. However, if $f_z \in C^1$ then $y \in C^2$ when f_{zz} is non-vanishing.

5.4 Integral Constraints: Lagrange Multipliers

Consider the optimization problem

$$\min_y \; F(y) := \int_a^b f(x, y(x), y'(x))dx, \tag{17a}$$

$$\text{s.t. } G_i(y) := \int_a^b g_i(x, y(x), y'(x))dx = c_i, \quad c_i \in \mathbb{R}, i = 1, 2, ..., N, \tag{17b}$$

$$y \in \mathcal{D} := \{y \in C^1[a,b] : y(a) = a_1, y(b) = b_1\}. \tag{17c}$$

We can employ the method of Lagrangian multipliers (cf. Proposition 4.14) since, in general, the linearity and weak continuity of the variations δF and δG is assured by Proposition 4.10.

Theorem 5.5. Suppose that $f = f(x,y,z)$ and $g_i = g_i(x,y,z), i = 1, 2, ..., N$, together with their y and z partial derivatives, are continuous on $[a,b] \times \mathbb{R}^2$. Let y_0 be a local extremal function for the optimization problem (17), then either (a). The determinant $\left|\delta G_i(y_0; v_j)\right|_{i,j=1,2,\dots,N} = 0$, whenever $v_j \in \mathcal{D}_0 := \{v \in C^1[a,b] : v(a) = v(b) = 0\}, j = 1, 2, ..., N$; or

(b). $\exists \lambda_i \in \mathbb{R}, i = 1, 2, ..., N$ that makes y_0 stationary for the modified function $\tilde{f} := f + \lambda_i g_i$. Namely, y_0 solves

$$\frac{d}{dx}\tilde{f}_{y'}(x) = \tilde{f}_y(x) \quad \text{on } (a,b).$$

Proof. The proof directly follows from Proposition 4.14 and Proposition 5.1. □

5.5 Vector-valued Functions

6 Sufficient Conditions

As we have noted repeatedly, the equations of Euler-Lagrange are necessary but not sufficient to characterize a minimum value for the integral function

$$F(Y) := \int_a^b f(x, Y(x), Y'(x))dx = \int_a^b f[Y(x)]dx$$

on a set such as

$$\mathcal{D} := \{Y \in (C^1[a,b])^d : Y(a) = A, Y(b) = B\},$$

since they are only conditions for the stationarity of F. However, in the presence of (strong) convexity of $f(\underline{x}, Y, Z)$ these conditions do characterize (unique) minimization.

6.1 The Weierstrass Method

References

[1] Troutman, J. L. *Variational calculus and optimal control: optimization with elementary convexity.* Springer Science & Business Media, 2012.