Videos

- **4 (9/2/2020)** – Explanation of "necessary conditions does not *follow* from (1.27)" (around 1:00:00).
- **5 (9/8/2020)** – Question about uniqueness of first and second variation.
- **6 (9/10/2020)** – Preview discussion of Section 2.3.
- **7 (9/15/2020)** –

1 Introduction

1.2.1 – Unconstrained optimization

First-order necessary condition for optimality (stationary point, Fermat's theorem):

$$\nabla f(x^*) = 0.$$

Second-order necessary condition for optimality (positive semidefinite Hessian):

$$\nabla^2 f(x^*) \geq 0.$$

Second-order sufficient condition for optimality: If $f \in \mathcal{C}^2$ satisfies

$$\nabla f(x^*) = 0 \quad \text{ and } \quad \nabla^2 f(x^*) > 0,$$

on an interior point $x^* \in D$, then x^* is a strict local minimum of f.

Exercise 1.1

We say d is a feasible direction at x^*, if it is of unit length and there exists some $\delta > 0$ such that $x^* + \alpha d \in D$ for all $0 \leq \alpha \leq \delta$. As discussed after the exercise description, it is possible that $x^* \in D$, but there are no feasible directions. This shows that we need some restrictions on D in order for the statement to hold. We say that D is *locally star-shaped* if there exists some $\delta^* > 0$ such that every $x \in D \cap B(x^*, \delta^*)$ satisfies $x = x^* + \alpha d$ for some feasible direction d and $0 \leq \alpha < \delta^*$, see Figure 1. For every feasible direction d, we have

$$\begin{aligned} f(x^* + \alpha d) &= f(x^*) + \alpha \nabla f(x^*) \cdot d + \frac{1}{2} d^T \nabla^2 f(x^*) d\alpha^2 + o(a^2) \\ &\geq f(x^*) + \frac{1}{2} d^T \nabla^2 f(x^*) d\alpha^2 + o(a^2) \\ &> f(x^*), \end{aligned}$$

where the last inequality follows again by taking α sufficiently small. Recall that $o(\alpha^2)$ and the corresponding value of ϵ in the proof for the unconstrained case depend on the choice of d. When the set of all feasible d is compact, the Weierstrass Theorem again tells us that we can take the minimum of ϵ over all d. In that case, x^* is a local minimum.

When the set of all feasible directions is not not compact, I am not completely sure how to proceed. Assuming convexity of D will not help, because $D \cap B(x^*, \delta^*)$ can then still be an open star-shaped domain. I think that we should somehow use the continuity of f.

1.2.2 – Constrained optimization

Exercise 1.2

Consider the following example in dimension $n = 3$. Define

$$\begin{aligned} h_1(x, y, z) &= z, \\ h_2(x, y, z) &= z - y^2, \end{aligned}$$

such that D is the x-axis and we have

$$\nabla h_1(0) = \nabla h_2(0) = (0, 0, 1),$$

which shows that $x^* = 0$ is not a regular point. Now define $f(x, y, z) = x^2 + y$ such that $\nabla f(0) = (0, 1, 0) \perp (0, 0, 1)$, which shows that (1.24) cannot hold.

Augmented cost function

The *augmented cost* function is defined as

$$\ell(x, \lambda) := f(x) + \sum_{i=1}^{m} \lambda_i h_i(x),$$

with gradient given by

$$\nabla \ell(x, \lambda) = \begin{pmatrix} \ell_x(x, \lambda) \\ \ell_\lambda(x, \lambda) \end{pmatrix} = \begin{pmatrix} \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x) \\ h(x) \end{pmatrix}.$$

If x^* is a local constrained minimum of f and λ^* is the corresponding vector of Lagrange multipliers for which the first order necessary condition (1.24) holds, then $\nabla \ell(x^*, \lambda^*) = 0$.

On page 16, it is stated that "However, it does *not* follow that (1.27) is a necessary condition for x^* to be a constrained minimum of f." This is confusing me, because (1.27) does hold under the stated assumptions. Maybe the author wants to stress the fact that the regularity assumption is required.

1.3.2 – First variation and first-order necessary condition

Exercise 1.5

Let $\eta \in V$ be arbitrary and define g_η like in the text, so $g_\eta(\alpha) = J(y + \alpha\eta)$. We can differentiate it as a function of α, which yields

$$\begin{aligned} g'_\eta(\alpha) &= \frac{d}{d\alpha} \int_0^1 \phi(y(x) + \alpha\eta(x)) dx \\ &= \int_0^1 \frac{d}{d\alpha} \phi(y(x) + \alpha\eta(x)) dx \\ &= \int_0^1 \phi'(y(x) + \alpha\eta(x))\eta(x) dx. \end{aligned}$$

In the first step, we used Leibniz integral rule for differentiation under the integral sign, which is allowed because $(x, \alpha) \mapsto \phi(y(x) + \alpha\eta(x))$ is continuous in both variables. We obtain the desired result by evaluating $\delta J|_y(\eta) = g'_\eta(0)$.

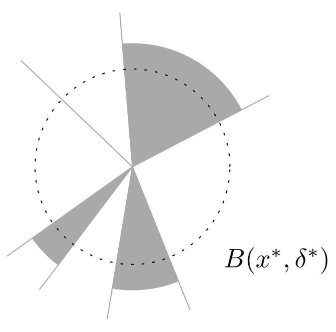

Figure 1: Domain D indicated in grey. Note that subsets of feasible directions can be either open or closed. A single isolated feasible direction is also drawn.

Exercise 1.6

First, observe that we can obtain an equivalent definition of the second variation, like the Gateaux derivative (1.33) was equivalent to the first variation, by using the definition of $o(\alpha^2)$ to rewrite the expansion (1.38) to

$$\delta^2 J|_y(\eta) = \lim_{\alpha\to 0} \frac{J(y+\alpha\eta) - J(y) - \delta J|_y(\eta)\alpha}{\alpha^2}.$$

By comparing this to the Taylor expansion of g_η around 0, given by

$$g_\eta(\alpha) = g_\eta(0) + \alpha g_\eta'(0) + \frac{\alpha^2}{2} g_\eta''(0) + o(\alpha^2),$$

which is equivalently stated as

$$\lim_{\alpha\to 0} \frac{g_\eta(\alpha) - g_\eta(0) - \alpha g_\eta'(0)}{\alpha^2} = \frac{1}{2} g_\eta''(0),$$

we conclude that

$$\delta^2 J|_y(\eta) = g_\eta''(0)/2.$$

Note this factor two difference because expansion (1.38) is not exactly analogous to the regular Taylor expansion. Because ϕ is twice differentiable, the map $(x,\alpha) \mapsto \phi'(y(x) + \alpha\eta(x))\eta(x)$ is continuous in both arguments, so we can again use Leibniz integral rule to compute

$$\begin{aligned} g_\eta''(\alpha) = \frac{d}{d\alpha} g_\eta'(\alpha) &= \frac{d}{d\alpha} \int_0^1 \phi'(y(x) + \alpha\eta(x))\eta(x)dx \\ &= \int_0^1 \frac{d}{d\alpha} \phi'(y(x) + \alpha\eta(x))\eta(x)dx \\ &= \int_0^1 \phi''(y(x) + \alpha\eta(x))\eta^2(x)dx, \end{aligned}$$

which we evaluate at zero to conclude

$$\delta^2 J|_y(\eta) = g_\eta''(0)/2 = \frac{1}{2}\int_0^1 \phi''(y(x))\eta^2(x)dx.$$

It is easily seen that this is indeed a quadratic form, by writing $\delta^2 J|_y(\eta) = B(\eta,\eta)/2$, with functional $B : V \times V \to \mathbb{R}$ defined as

$$B(\eta_1,\eta_2) = \int_0^1 \phi''(y(x))\eta_1(x)\eta_2(x)dx,$$

which is bilinear because $B(\eta_1,\eta_2) = B(\eta_2,\eta_1)$ and

$$\begin{aligned} B(\alpha\eta_1 + \beta\eta_2, \xi) &= \int_0^1 \phi''(y(x))(\alpha\eta_1(x) + \beta\eta_2(x))\xi(x)dx\alpha \\ &= \alpha\int_0^1 \phi''(y(x))\eta_1(x)\xi(x)dx + \beta\int_0^1 \phi''(y(x))\eta_2(x)\xi(x)dx \\ &= \alpha B(\eta_1,\xi) + \beta B(\eta_2,\xi). \end{aligned}$$

2 Calculus of Variations

Exercise 2.3

We define the *remainder* of L at some point $p \in [0,1]\times\mathbb{R}^2$ with some perturbation h as

$$R(p,h) = L(p+h) - L(p) - dL(p)\cdot h,$$

where $dL(p)$ is the differential of L at p. Whenever $dL(p)$ exists at p, Taylor's theorem for multivariable functions tells us that

$$\frac{R(p,h)}{\|h\|} \longrightarrow 0 \quad \text{as } h \to 0, \tag{1}$$

where we write $\|h\|$ for the usual Euclidean norm. Consider a sequence of functions (η_n) such that $\lim_{n\to\infty} \|\eta_n\|_1 = 0$ with respect to the 1-norm defined in (1.30). Now define $h_n(x) = (0, \eta_n(x), \eta_n'(x))$, then by Hölder's inequality, we have

$$\begin{aligned}\|h_n(x)\| &\le \|h_n(x)\|_1 \\ &\le \sup_{x\in[0,1]} |\eta_n(x)| + \sup_{x\in[0,1]} |\eta_n'(x)| \\ &= \|\eta_n\|_1,\end{aligned}$$

which means that we have $\lim_{n\to\infty} \|h_n(x)\| = 0$ for every $x \in [0,1]$. Now define $p(x) = (x, y(x), y'(x))$, then we have

$$\begin{aligned}\frac{1}{\|\eta_n\|_1}\int_a^b R(p(x),h_n(x))dx &\le (b-a)\sup_{x\in[0,1]} \frac{R(p(x),h_n(x))}{\|\eta_n\|_1} && \text{(2a)}\\ &\le (b-a)\sup_{x\in[0,1]} \underbrace{\frac{R(p(x),h_n(x))}{\|h_n(x)\|}}_{P_n(x):=}. && \text{(2b)}\end{aligned}$$

Let $\epsilon > 0$ be given, then for every $x \in [0,1]$ there is some $N(x)$ such that $P_n(x) < \epsilon$ for all $n \ge N(x)$. Since the interval $[0,1]$ is compact, the Weierstrass theorem gives us $N = \max_{x\in[0,1]} N(x)$ such that

$$\sup_{x\in[0,1]} P_n(x) < \epsilon \quad \text{for all } n \ge N.$$

This shows that the left-hand side of (2) goes to zero as $n \to \infty$. To conclude, observe that $\delta J|_y(\eta_n)$ defined by (2.14) satisfies the definition of (1.37), because we have shown that

$$J(y+\eta_n) - J(y) - \delta J|_y(\eta_n) = \int_a^b R(p(x),h_n(x))dx = o(\|\eta_n\|_1).$$

Also true for 0-norm?

For the 0-norm, the above proof does not go through, because it is not too difficult to construct a sequence (η_n) like y_2 in Figure 2.6 such that $\lim_{n\to\infty} \|\eta_n\|_0 = 0$, but which has $\sup_{x\in[0,1]} \eta_n'(x) \ge C$ for all n, with some constant $C > 0$. This means that $\|h_n(x)\|$ will not converge to zero as $n \to \infty$.