Numerical Optimization

Theory

Composing derivatives

Theorem 7.1-1

Let $f:\Omega \subset X \to Y$ be a function differentiable at a point $a \in \Omega$ and let $g:\Omega' \subset Y \to Z$ be a function differentiable at the point $b = f(a) \in \Omega'$. Suppose that $f(\Omega) \subset \Omega'$. Then the composite function

$$h = gf:\Omega \subset X \to Z$$

is differentiable at the point $a \in \Omega$ and

$$h'(a) = g'(b)f'(a).$$

For the special case of functions

$$f:\Omega \subset \mathbb{R}^n \to \mathbb{R}^m, \quad g:\Omega' \subset \mathbb{R}^m \to \mathbb{R}^l,$$

there corresponds to the composition of the derivatives $g'(b)$ and $f'(a)$ the multiplication of the derivative matrices of the functions concerned:

$$\begin{pmatrix} \partial_1 h_1(a) & \cdots & \partial_n h_1(a) \\ \vdots & & \vdots \\ \partial_1 h_l(a) & \cdots & \partial_n h_l(a) \end{pmatrix} = \begin{pmatrix} \partial_1 g_1(b) & \cdots & \partial_m g_1(b) \\ \vdots & & \vdots \\ \partial_1 g_l(b) & \cdots & \partial_m g_l(b) \end{pmatrix} \begin{pmatrix} \partial_1 f_1(a) & \cdots & \partial_n f_1(a) \\ \vdots & & \vdots \\ \partial_1 f_m(a) & \cdots & \partial_n f_m(a) \end{pmatrix},$$

which can also be written as

$$\partial_j h_i(a) = \sum_{k=1}^{m} \partial_k g_i(b) \partial_j f_k(a), \quad 1 \leqslant i \leqslant l,\ 1 \leqslant j \leqslant n. \qquad \square$$

Mean value theorem

Theorem 7.1-2 (Mean value theorem)
Let $f:\Omega \subset X \to Y$ and let a and b be two points of Ω such that the segment $[a, b]$ lies within Ω. Suppose that the function f is continuous at every point of the closed segment $[a, b]$ and differentiable at every point of the open segment $]a, b[$. Then

$$\| f(b) - f(a) \|_Y \leqslant \sup_{x \in]a,b[} \| f'(x) \|_{\mathscr{L}(X;Y)} \| b - a \|_X. \qquad \square$$

Implicit function theorem

Theorem 7.1-3 (Implicit function theorem)
Let $\varphi:\Omega \subset X_1 \times X_2 \to Y$ be a function continuously differentiable in Ω and let $(a_1, a_2) \in \Omega$, $b \in Y$ be points such that

$$\varphi(a_1, a_2) = b, \quad \partial_2\varphi(a_1, a_2) \in \mathrm{Isom}(X_2; Y).$$

Suppose that the space X_2 is complete.

Then there exist an open subset $O_1 \subset X_1$, an open subset $O_2 \subset X_2$ and a continuous function, called the implicit function,

$$f: O_1 \subset X_1 \to X_2,$$

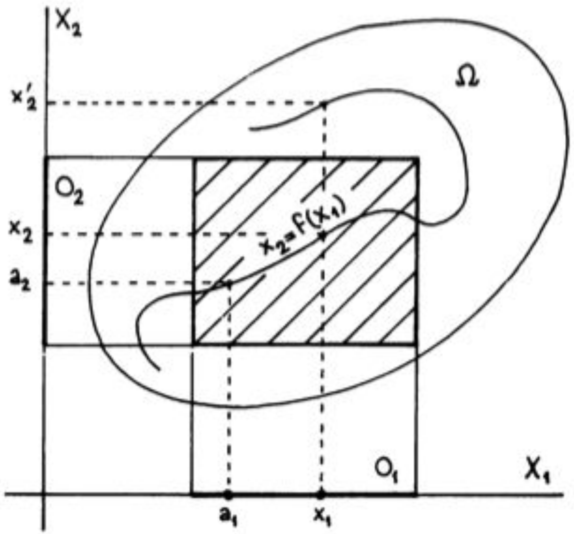
such that $(a_1, a_2) \in O_1 \times O_2 \subset \Omega$ and (cf. figure 7.1-1)

$$\{(x_1, x_2) \in O_1 \times O_2 : \varphi(x_1, x_2) = b\} = \{(x_1, x_2) \in O_1 \times X_2 : x_2 = f(x_1)\}.$$

Moreover, the function f is differentiable at the point a_1 and

$$f'(a_1) = -\{\partial_2\varphi(a_1, a_2)\}^{-1}\partial_1\varphi(a_1, a_2). \qquad \square$$

Figure 7.1-1



Taylor formulae for differentiable functions

Theorem 7.1-4 **(*Taylor's formulae for differentiable functions*)**
Let $f:\Omega \subset X \to Y$ and let $[a, a+h]$ be any closed segment lying within Ω.

(1) *If f is differentiable at a, then*

$$f(a+h) = f(a) + f'(a)h + \|h\|\varepsilon(h), \quad \lim_{h \to 0} \varepsilon(h) = 0.$$

(2) Mean value theorem: *if $f \in \mathscr{C}^0(\Omega)$ and f is differentiable over $]a, a+h[$, then*

$$\|f(a+h) - f(a)\| \leqslant \sup_{x \in]a, a+h[} \|f'(x)\| \, \|h\|.$$

(3) The Taylor–Maclaurin formula: *if $f = \mathscr{C}^0(\Omega)$, f is differentiable over $]a, a+h[$ and $Y = \mathbb{R}$, then*

$$f(a+h) = f(a) + f'(a+\theta h)h, \quad 0 < \theta < 1.$$

(4) Taylor's formula with integral remainder: *if $f \in \mathscr{C}^1(\Omega)$ and Y is a complete vector space, then*

$$f(a+h) = f(a) + \int_0^1 \{f'(a+th)h\} \, \mathrm{d}t.$$

Taylor's formulae for twice differentiable functions

Theorem 7.1-5 **(*Taylor's formulae for twice differentiable functions*)** *Let $f : \Omega \subset X \to Y$ and let $[a, a+h]$ be any closed segment lying within Ω.*

(1) The Taylor–Young formula: *if f is differentiable in Ω and twice differentiable at a, then*

$$f(a+h) = f(a) + f'(a)h + \tfrac{1}{2} f''(a)(h, h) + \|h\|^2 \varepsilon(h), \quad \lim_{h \to 0} \varepsilon(h) = 0.$$

(2) The generalised mean value theorem: *if $f \in \mathscr{C}^1(\Omega)$ and f is twice differentiable over $]a, a+h[$, then*

$$\|f(a+h) - f(a) - f'(a)h\| \leqslant \tfrac{1}{2} \sup_{x \in]a, a+h[} \|f''(x)\| \, \|h\|^2.$$

(3) The Taylor–Maclaurin formula: *if $f \in \mathscr{C}^1(\Omega)$, f is twice differentiable over $]a, a+h[$ and $Y = \mathbb{R}$, then*

$$f(a+h) = f(a) + f'(a)h + \tfrac{1}{2} f''(a + \theta h)(h, h), \quad 0 < \theta < 1.$$

(4) Taylor's formula with integral remainder: *if $f \in \mathscr{C}^2(\Omega)$ and Y is a complete vector space, then*

$$f(a+h) = f(a) + f'(a)h + \int_0^1 (1-t)\{f''(a+th)(h, h)\}\, dt.$$

Necessary condition for a relative extremum

***Theorem 7.2-1* (Necessary condition for a relative extremum)**
Let Ω be an open subset of a normed vector space V and $J:\Omega \subset V \to \mathbb{R}$ a function. If the function J has a relative extremum at a point $u \in \Omega$ and if it is differentiable at this point, then

$$J'(u) = 0.$$

Necessary condition for a constrained relative extremum

***Theorem 7.2-2** (Necessary condition for a constrained relative extremum)*

Let Ω be an open subset of a product $V_1 \times V_2$ of normed vector spaces, the space V_1 being complete, let $\varphi: \Omega \to V_2$ be a function of class $\mathscr{C}^1$ over Ω and let $u = (u_1, u_2)$ be a point of the set

$$U = \{(v_1, v_2) \in \Omega: \quad \varphi(v_1, v_2) = 0\} \subset \Omega$$

at which

$$\partial_2 \varphi(u_1, u_2) \in \mathrm{Isom}(V_2).$$

Let $J: \Omega \to \mathbb{R}$ be a function differentiable at u. If the function J has a relative extremum at u with respect to the set U, then there exists an element $\Lambda(u) \in \mathscr{L}(V_2; \mathbb{R})$ such that

$$J'(u) + \Lambda(u)\varphi'(u) = 0.$$

Necessary condition for a constrained relative extremum (bis)

***Theorem 7.2-3* (*Necessary condition for a constrained relative extremum*)**
Let Ω be an open subset of $\mathbb{R}^n$, let $\varphi_i : \Omega \to \mathbb{R}$, $1 \leqslant i \leqslant m$, be functions of class $\mathscr{C}^1$ over Ω and let u be a point of the set

$$U = \{v \in \Omega: \quad \varphi_i(v) = 0, \quad 1 \leqslant i \leqslant m\} \subset \Omega,$$

at which the derivatives $\varphi_i'(u) \in \mathscr{L}(\mathbb{R}^n; \mathbb{R})$, $1 \leqslant i \leqslant m$, are linearly independent.
Let $J : \Omega \to \mathbb{R}$ be a function differentiable at u. If the function J has a relative extremum at u with respect to the set U, there exist m numbers $\lambda_i(u)$, $1 \leqslant i \leqslant m$, uniquely defined, such that

$$J'(u) + \lambda_1(u)\varphi_1'(u) + \cdots + \lambda_m(u)\varphi_m'(u) = 0.$$

Necessary condition for a relative minimum

Theorem 7.3-1 (**Necessary condition for a relative minimum**)
Let Ω be an open subset of a normed vector space V and $J: \Omega \subset V \to \mathbb{R}$ a function differentiable in Ω and twice differentiable at the point $u \in \Omega$. If the function J has a relative minimum at u, then

$$J''(u)(w, w) \geqslant 0 \quad \text{for every } w \in V.$$

Sufficient conditions for a relative minimum

Theorem 7.3-2 (***Sufficient conditions for a relative minimum***)
Let Ω be an open subset of a normed vector space V, u a point of Ω and $J:\Omega \subset V \to \mathbb{R}$ a function differentiable in Ω such that $J'(u) = 0$.

(1) *If the function J is twice differentiable at u and if there exists a number α such that*

$$\alpha > 0 \quad \text{and} \quad J''(u)(w, w) \geqslant \alpha \| w \|^2 \quad \text{for every } w \in V,$$

then the function J has a strict relative minimum at u.

(2) *If the function J is twice differentiable in Ω and if there exists a ball $B \subset \Omega$ centred at u such that*

$$J''(v)(w, w) \geqslant 0 \quad \text{for every } v \in B, w \in V,$$

then the function J has a relative minimum at u.

Necessary condition for a relative minimum over a convex set

***Theorem 7.4-1** (Necessary condition for a relative minimum over a convex set)*

Let $J : \Omega \subset \mathbb{R}$ be a function defined over an open subset Ω of a normed vector space V and let U be a convex subset of Ω. If the function J is differentiable at a point $u \in U$ and if it has a relative minimum at u with respect to the set U, then

$$J'(u)(v-u) \geqslant 0 \quad \text{for every } v \in U.$$

Convexity and first derivative

***Theorem 7.4-2* (*Convexity and first derivative*)**
Let $J: \Omega \subset V \to \mathbb{R}$ be a function differentiable in an open set Ω of a normed vector space V and let U be a convex subset of Ω.

(1) *The function J is convex over U if and only if*

$$J(v) \geqslant J(u) + J'(u)(v-u) \quad \text{for every } u, v \in U.$$

(2) *The function J is strictly convex over U if and only if*

$$J(v) > J(u) + J'(u)(v-u) \quad \text{for every } u, v \in U, \quad u \neq v.$$

Convexity and second derivative

Theorem 7.4-3 (Convexity and second derivative)
Let $J:\Omega \subset V \to \mathbb{R}$ be a function twice differentiable in an open subset Ω of a normed vector space V and let U be a convex subset of Ω.

(1) *The function J is convex over U if and only if*

$$J''(u)(v-u, v-u) \geqslant 0 \quad \text{for every } u, v \in U.$$

(2) *If*

$$J''(u)(v-u, v-u) > 0 \quad \text{for every } u, v \in U, \quad u \neq v,$$

then the function J is strictly convex over U.

Minima of convex functions

Theorem 7.4-4

Let U be a convex subset of a normed vector space V.

(1) *If a convex function $J: U \subset V \to \mathbb{R}$ has a relative minimum at a point of U, it has, in fact, a minimum there, that is to say, with respect to the entire set U.*

(2) *A strictly convex function $J: U \subset V \to \mathbb{R}$ has at most one minimum, and that minimum is strict.*

(3) *Let $J: \Omega \subset V \to \mathbb{R}$ be a convex function defined over an open subset Ω of V containing U and differentiable at a point $u \in U$. Then the function J has a minimum at u with respect to the set U if and only if*

$$J'(u)(v-u) \geqslant 0 \quad \text{for every } v \in U.$$

(4) *If the set U is open, the preceding condition is equivalent to Euler's equation $J'(u)=0$.*

Conditions for Newton method convergence

Theorem 7.5-1

Assume that the space X is complete and that the function $f: \Omega \subset X \to Y$ is differentiable in the open set Ω. Assume, too, that there exist three constants r, M and β such that

$$r > 0 \quad and \quad B \stackrel{\text{def}}{=} \{x \in X : \|x - x_0\| \leqslant r\} \subset \Omega,$$

(1)
$$\sup_{k \geqslant 0} \sup_{x \in B} \|A_k^{-1}(x)\|_{\mathscr{L}(X;Y)} \leqslant M,$$

(2)
$$\sup_{k \geqslant 0} \sup_{x, x' \in B} \|f'(x) - A_k(x')\|_{\mathscr{L}(X;Y)} \leqslant \frac{\beta}{M}, \quad \text{and} \quad \beta < 1,$$

(3)
$$\|f(x_0)\| \leqslant \frac{r}{M}(1 - \beta).$$

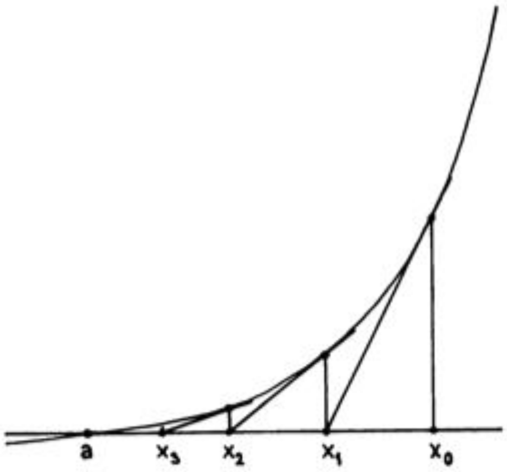
Then the sequence $(x_k)_{k \geqslant 0}$ defined by

$$x_{k+1} = x_k - A_k^{-1}(x_{k'}) f(x_k), \quad k \geqslant k' \geqslant 0,$$

is entirely contained within the ball B and converges to a zero of f, which is the only zero of f in the ball B. Lastly, the convergence is geometric:

$$\|x_k - a\| \leqslant \frac{\|x_1 - x_0\|}{1 - \beta} \beta^k.$$

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}, \quad k \geqslant 0,$$



Conditions for Newton method convergence Bis

Theorem 7.5-2

Suppose that the space X is complete and that the function $f:\Omega\subset X\to Y$ is continuously differentiable in the open set Ω. Let a be a point of Ω such that

$$\begin{cases} f(a)=0, \quad A \overset{\text{def}}{=} f'(a)\in \mathrm{Isom}\,(X;Y), \\ \sup\limits_{k\geq 0} \|A_k - A\|_{\mathscr{L}(X;Y)} \leqslant \dfrac{\lambda}{\|A^{-1}\|_{\mathscr{L}(Y;X)}}, \quad \text{and} \quad \lambda<\frac{1}{2}. \end{cases}$$

Then there exists a closed ball B centred at a such that, for every point $x_0\in B$, the sequence $(x_k)_{k\geqslant 0}$ defined by

$$x_{k+1} = x_k - A_k^{-1} f(x_k), \quad k\geqslant 0,$$

is entirely contained within B and converges to a point a, which is the only zero of f within the ball B. Lastly, the convergence is geometric: there exists a number β such that

$$\beta<1 \quad \textit{and} \quad \|x_k - a\| \leqslant \beta^k \|x_0 - a\|, \quad k\geqslant 0.$$

Newton method convergence for finding null derivatives

Theorem 7.5-3

Let Ω be an open set of a complete space V and let $J:\Omega \subset V \to \mathbb{R}$ be a function which is twice differentiable in the open set Ω. Suppose, moreover, that there exist three constants r, M and β such that

$$r > 0 \quad \text{and} \quad B \overset{\text{def}}{=} \{v \in V : \|v - u_0\| \leqslant r\} \subset \Omega,$$

$$A_k(v) \in \operatorname{Isom}(V; V') \quad \text{for every } v \in B \quad \text{and}$$

$$\sup_{k \geqslant 0} \sup_{v \in B} \|A_k^{-1}(v)\|_{\mathscr{L}(V';V)} \leqslant M,$$

$$\sup_{k \geqslant 0} \sup_{v, v' \in B} \|J''(v) - A_k(v')\|_{\mathscr{L}(V;V')} \leqslant \frac{\beta}{M}, \quad \text{and} \quad \beta < 1,$$

$$\|J'(u_0)\|_{V'} \leqslant \frac{r}{M}(1 - \beta).$$

Then the sequence $(u_k)_{k \geqslant 0}$ defined by

$$u_{k+1} = u_k - A_{k'}^{-1}(u_{k'})J'(u_k), \quad k \geqslant k' \geqslant 0,$$

is entirely contained within the ball B and converges to a zero of J', which is the only zero of J' in the ball B. Lastly, the convergence is geometric.

Newton method convergence for finding null derivatives (bis)

Theorem 7.5-4

Let Ω be an open subset of a complete space V and let $J:\Omega \subset V \to \mathbb{R}$ be a function which is twice continuously differentiable in Ω. Moreover, let u be a point of Ω such that

$$\begin{cases} J'(u) = 0, \quad J''(u) \in \mathrm{Isom}\,\mathscr{L}(V; V'), \\ \sup\limits_{k \geqslant 0} \| A_k - J''(u) \|_{\mathscr{L}(V; V')} \leqslant \dfrac{\lambda}{\|(J''(u))^{-1}\|_{\mathscr{L}(V'; V)}} \quad \text{and} \quad \lambda < \dfrac{1}{2}. \end{cases}$$

Then there exists a closed ball $B \subset V$ with centre at u such that, for every point $u_0 \in B$, the sequence $(u_k)_{k \geqslant 0}$ defined by

$$u_{k+1} = u_k - A_k^{-1} J'(u_k), \quad k \geqslant 0,$$

is entirely contained in B and converges to the point u, which is, moreover, the only zero of J' in the ball B. Lastly, the convergence is geometric.