

Nonlinear Programming
2nd Edition

Solutions Manual

Dimitri P. Bertsekas

Massachusetts Institute of Technology



Athena Scientific, Belmont, Massachusetts

NOTE

This solutions manual is continuously updated and improved. Portions of the manual, involving primarily theoretical exercises, have been posted on the internet at the book's www page

<http://www.athenasc.com/nonlinbook.html>

Many thanks are due to several people who have contributed solutions, and particularly to Angelia Nedic, Asuman Ozdaglar, and Cynara Wu.

Last Updated: May 2005

Solutions Chapter 1

SECTION 1.1

1.1.9 www

For any $x, y \in \mathcal{R}^n$, from the second order expansion (see Appendix A, Proposition A.23) we have

$$f(y) - f(x) = (y - x)' \nabla f(x) + \frac{1}{2} (y - x)' \nabla^2 f(z) (y - x), \quad (1)$$

where z is some point of the line segment joining x and y . Setting $x = 0$ in (1) and using the given property of f , it can be seen that f is coercive. Therefore, there exists $x^* \in \mathcal{R}^n$ such that $f(x^*) = \inf_{x \in \mathcal{R}^n} f(x)$ (see Proposition A.8 in Appendix A). The condition

$$m \|y\|^2 \leq y' \nabla^2 f(x) y, \quad \forall x, y \in \mathcal{R}^n,$$

is equivalent to strong convexity of f . Strong convexity guarantees that there is a unique global minimum x^* . By using the given property of f and the expansion (1), we obtain

$$(y - x)' \nabla f(x) + \frac{m}{2} \|y - x\|^2 \leq f(y) - f(x) \leq (y - x)' \nabla f(x) + \frac{M}{2} \|y - x\|^2.$$

Taking the minimum over $y \in \mathcal{R}^n$ in the expression above gives

$$\min_{y \in \mathcal{R}^n} \left((y - x)' \nabla f(x) + \frac{m}{2} \|y - x\|^2 \right) \leq f(x^*) - f(x) \leq \min_{y \in \mathcal{R}^n} \left((y - x)' \nabla f(x) + \frac{M}{2} \|y - x\|^2 \right).$$

Note that for any $a > 0$

$$\min_{y \in \mathcal{R}^n} \left((y - x)' \nabla f(x) + \frac{a}{2} \|y - x\|^2 \right) = -\frac{1}{2a} \|\nabla f(x)\|^2,$$

and the minimum is attained for $y = x - \frac{\nabla f(x)}{a}$. Using this relation for $a = m$ and $a = M$, we obtain

$$-\frac{1}{2m} \|\nabla f(x)\|^2 \leq f(x^*) - f(x) \leq -\frac{1}{2M} \|\nabla f(x)\|^2.$$

The first chain of inequalities follows from here. To show the second relation, use the expansion (1) at the point $x = x^*$, and note that $\nabla f(x^*) = 0$, so that

$$f(y) - f(x^*) = \frac{1}{2} (y - x^*)' \nabla^2 f(z) (y - x^*).$$

The rest follows immediately from here and the given property of the function f .

1.1.11 www

Since x^* is a nonsingular strict local minimum, we have that $\nabla^2 f(x^*) > 0$. The function f is twice continuously differentiable over $\mathbb{R}^n$, so that there exists a scalar $\delta > 0$ such that

$$\nabla^2 f(x) > 0, \quad \forall x, \quad \text{with } \|x - x^*\| \leq \delta.$$

This means that the function f is strictly convex over the open sphere $B(x^*, \delta)$ centered at x^* with radius δ . Then according to Proposition 1.1.2, x^* is the only stationary point of f in the sphere $B(x^*, \delta)$.

If f is not twice continuously differentiable, then x^* need not be an isolated stationary point. The example function f does not have the second derivative at $x = 0$. Note that $f(x) > 0$ for $x \neq 0$, and by definition $f(0) = 0$. Hence, $x^* = 0$ is the unique (singular) global minimum. The first derivative of $f(x)$ for $x \neq 0$ can be calculated as follows:

$$\begin{aligned} f'(x) &= 2x \left(\sqrt{2} - \sin \left(\frac{5\pi}{6} - \sqrt{3} \ln(x^2) \right) + \sqrt{3} \cos \left(\frac{5\pi}{6} - \sqrt{3} \ln(x^2) \right) \right) \\ &= 2x \left(\sqrt{2} - 2 \cos \frac{\pi}{3} \sin \left(\frac{5\pi}{6} - \sqrt{3} \ln(x^2) \right) + 2 \sin \frac{\pi}{3} \cos \left(\frac{5\pi}{6} - \sqrt{3} \ln(x^2) \right) \right) \\ &= 2x \left(\sqrt{2} + 2 \sin \left(\frac{\pi}{3} - \frac{5\pi}{6} + \sqrt{3} \ln(x^2) \right) \right) \\ &= 2x \left(\sqrt{2} - 2 \cos(2\sqrt{3} \ln x) \right). \end{aligned}$$

Solving $f'(x) = 0$, gives $x^k = e^{\frac{(1-8k)\pi}{8\sqrt{3}}}$ and $y^k = e^{\frac{-(1+8k)\pi}{8\sqrt{3}}}$ for k integer. The second derivative of $f(x)$, for $x \neq 0$, is given by

$$f''(x) = 2 \left(\sqrt{2} - 2 \cos(2\sqrt{3} \ln x) + 4\sqrt{3} \sin(2\sqrt{3} \ln x) \right).$$

Thus:

$$\begin{aligned} f''(x^k) &= 2 \left(\sqrt{2} - 2 \cos \frac{\pi}{4} + 4\sqrt{3} \sin \frac{\pi}{4} \right) \\ &= 2 \left(\sqrt{2} - 2 \frac{\sqrt{2}}{2} + 4\sqrt{3} \frac{\sqrt{2}}{2} \right) \\ &= 4\sqrt{6}. \end{aligned}$$

Similarly

$$\begin{aligned} f''(y^k) &= 2 \left(\sqrt{2} - 2 \cos \left(\frac{-\pi}{4} \right) + 4\sqrt{3} \sin \left(\frac{-\pi}{4} \right) \right) \\ &= 2 \left(\sqrt{2} - 2 \frac{\sqrt{2}}{2} - 4\sqrt{3} \frac{\sqrt{2}}{2} \right) \\ &= -4\sqrt{6}. \end{aligned}$$

Hence, $\{x^k \mid k \geq 0\}$ is a sequence of nonsingular local minima, which evidently converges to x^* , while $\{y^k \mid k \geq 0\}$ is a sequence of nonsingular local maxima converging to x^* .

1.1.12 www

(a) Let x^* be a strict local minimum of f . Then there is δ such that $f(x^*) < f(x)$ for all x in the closed sphere centered at x^* with radius δ . Take any local sequence $\{x^k\}$ that minimizes f , i.e. $\|x^k - x^*\| \leq \delta$ and $\lim_{k \rightarrow \infty} f(x^k) = f(x^*)$. Then there is a subsequence $\{x^{k_i}\}$ and the point $\bar{x}$ such that $x^{k_i} \rightarrow \bar{x}$ and $\|\bar{x} - x^*\| \leq \delta$. By continuity of f , we have

$$f(\bar{x}) = \lim_{i \rightarrow \infty} f(x^{k_i}) = f(x^*).$$

Since x^* is a strict local minimum, it follows that $\bar{x} = x^*$. This is true for any convergent subsequence of $\{x^k\}$, therefore $\{x^k\}$ converges to x^* , which means that x^* is locally stable. Next we will show that for a continuous function f every locally stable local minimum must be strict. Assume that this is not true, i.e., there is a local minimum x^* which is locally stable but is not strict. Then for any $\theta > 0$ there is a point $x^\theta \neq x^*$ such that

$$0 < \|x^\theta - x^*\| < \theta \quad \text{and} \quad f(x^\theta) = f(x^*). \quad (1)$$

Since x^* is a stable local minimum, there is a $\delta > 0$ such that $x^k \rightarrow x^*$ for all $\{x^k\}$ with

$$\lim_{k \rightarrow \infty} f(x^k) = f(x^*) \quad \text{and} \quad \|x^k - x^*\| < \delta. \quad (2)$$

For $\theta = \delta$ in (1), we can find a point $x^0 \neq x^*$ for which $0 < \|x^0 - x^*\| < \delta$ and $f(x^0) = f(x^*)$. Then, for $\theta = \frac{1}{2}\|x^0 - x^*\|$ in (1), we can find a point x^1 such that $0 < \|x^1 - x^*\| < \frac{1}{2}\|x^0 - x^*\|$ and $f(x^1) = f(x^*)$. Then, again, for $\theta = \frac{1}{2}\|x^1 - x^*\|$ in (1), we can find a point x^2 such that $0 < \|x^2 - x^*\| < \frac{1}{2}\|x^1 - x^*\|$ and $f(x^2) = f(x^*)$, and so on. In this way, we have constructed a sequence $\{x^k\}$ of distinct points such that $0 < \|x^k - x^*\| < \delta$, $f(x^k) = f(x^*)$ for all k , and $\lim_{k \rightarrow \infty} x^k = x^*$. Now, consider the sequence $\{y^k\}$ defined by

$$y^{2m} = x^m, \quad y^{2m+1} = x^0, \quad \forall m \geq 0.$$

Evidently, the sequence $\{y^k\}$ is contained in the sphere centered at x^* with the radius δ . Also we have that $f(y^k) = f(x^*)$, but $\{y^k\}$ does not converge to x^* . This contradicts the assumption that x^* is locally stable. Hence, x^* must be strict local minimum.

(b) Since x^* is a strict local minimum, we can find $\delta > 0$, such that $f(x) > f(x^*)$ for all $x \neq x^*$ with $\|x - x^*\| \leq \delta$. Then $\min_{\|x - x^*\| = \delta} f(x) = f^\delta > f(x^*)$. Let $G^\delta = \max_{\|x - x^*\| \leq \delta} |g(x)|$. Now, we have

$$f(x) - \epsilon G^\delta \leq f(x) + \epsilon g(x) \leq f(x) + \epsilon G^\delta, \quad \forall \epsilon > 0, \quad \forall x \quad \|x - x^*\| < \delta.$$

Choose ϵ^δ such that

$$f^\delta - \epsilon^\delta G^\delta > f(x^*) + \epsilon^\delta G^\delta,$$

and notice that for all $0 \leq \epsilon \leq \epsilon^\delta$ we have

$$f^\delta - \epsilon G^\delta > f(x^*) + \epsilon G^\delta.$$

Consider the level sets

$$L(\epsilon) = \{x \mid f(x) + \epsilon g(x) \leq f(x^*) + \epsilon G^\delta, \quad \|x - x^*\| \leq \delta\}, \quad 0 \leq \epsilon \leq \epsilon^\delta.$$

Note that

$$L(\epsilon^1) \subset L(\epsilon^2) \subset B(x^*, \delta), \quad \forall 0 \leq \epsilon^1 < \epsilon^2 \leq \epsilon^\delta, \quad (3)$$

where $B(x^*, \delta)$ is the open sphere centered at x^* with radius δ . The relation (3) means that the sequence $\{L(\epsilon)\}$ decreases as ϵ decreases. Observe that for any $\epsilon \geq 0$, the level set $L(\epsilon)$ is compact. Since x^* is strictly better than any other point $x \in B(x^*, \delta)$, and $x^* \in L(\epsilon)$ for all $0 \leq \epsilon \leq \epsilon^\delta$, we have

$$\bigcap_{0 \leq \epsilon \leq \epsilon^\delta} L(\epsilon) = \{x^*\}. \quad (4)$$

According to Weierstrass' theorem, the continuous function $f(x) + \epsilon g(x)$ attains its minimum on the compact set $L(\epsilon)$ at some point $x_\epsilon \in L(\epsilon)$. From (3) it follows that $x_\epsilon \in B(x^*, \delta)$ for any ϵ in the range $[0, \epsilon^\delta]$. Finally, since $x_\epsilon \in L(\epsilon)$, from (4) we see that $\lim_{\epsilon \rightarrow \infty} x_\epsilon = x^*$.

1.1.13 www

In the solution to the Exercise 1.1.12 we found the numbers $\delta > 0$ and $\epsilon^\delta > 0$ such that for all $\epsilon \in [0, \epsilon^\delta)$ the function $f(x) + \epsilon g(x)$ has a local minimum x_ϵ within the sphere $B(x^*, \delta) = \{x \mid \|x - x^*\| < \delta\}$. The Implicit Function Theorem can be applied to the continuously differentiable function $G(\epsilon, x) = \nabla f(x) + \epsilon \nabla g(x)$ for which $G(0, x^*) = 0$. Thus, there are an interval $[0, \epsilon_0)$, a number δ_0 and a continuously differentiable function $\phi : [0, \epsilon_0) \mapsto B(x^*, \delta_0)$ such that $\phi(\epsilon) = x'_\epsilon$ and

$$\nabla \phi(\epsilon) = -\nabla_\epsilon G(\epsilon, \phi(\epsilon)) (\nabla_x G(\epsilon, \phi(\epsilon)))^{-1}, \quad \forall \epsilon \in [0, \epsilon_0).$$

We may assume that ϵ_0 is small enough so that the first order expansion for $\phi(\epsilon)$ at $\epsilon = 0$ holds, namely

$$\phi(\epsilon) = \phi(0) + \epsilon \nabla \phi(0) + o(\epsilon), \quad \forall \epsilon \in [0, \epsilon_0). \quad (1)$$

It can be seen that $\nabla_x G(0, \phi(0)) = \nabla_x G(0, x^*) = \nabla^2 f(x^*)$, and $\nabla_\epsilon G(0, \phi(0)) = \nabla g(x^*)'$, which combined with $\phi(\epsilon) = x'_\epsilon$, $\phi(0) = (x^*)'$ and (1) gives the desired relation.

SECTION 1.2

1.2.5 www

(a) Given a bounded set A , let $r = \sup\{\|x\| \mid x \in A\}$ and $B = \{x \mid \|x\| \leq r\}$. Let $L = \max\{\|\nabla^2 f(x)\| \mid x \in B\}$, which is finite because a continuous function on a compact set is bounded. For any $x, y \in A$ we have

$$\nabla f(x) - \nabla f(y) = \int_0^1 \nabla^2 f(tx + (1-t)y)(x-y)dt.$$

Notice that $tx + (1-t)y \in B$, for all $t \in [0, 1]$. It follows that

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|,$$

as desired.

(b) The key idea is to show that x^k stays in the bounded set

$$A = \{x \mid f(x) \leq f(x^0)\}$$

and to use a stepsize α^k that depends on the constant L corresponding to this bounded set. Let

$$R = \max\{\|x\| \mid x \in A\},$$

$$G = \max\{\|\nabla f(x)\| \mid x \in A\},$$

and

$$B = \{x \mid \|x\| \leq R + 2G\}.$$

Using condition (i) in the exercise, there exists some constant L such that $\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$, for all $x, y \in B$. Suppose the stepsize α^k satisfies

$$0 < \epsilon \leq \alpha^k \leq (2 - \epsilon)\gamma^k \min\{1, 1/L\},$$

where

$$\gamma^k = \frac{|\nabla f(x^k)'d^k|}{\|d^k\|^2}.$$

Let $\beta^k = \alpha^k(\gamma^k - L\alpha^k/2)$, which can be seen to satisfy $\beta^k \geq \epsilon^2\gamma^k/2$ by our choice of α^k . We will, show by induction on k that with such a choice of stepsize, we have $x^k \in A$ and

$$f(x^{k+1}) \leq f(x^k) - \beta^k\|d^k\|^2, \tag{*}$$

for all $k \geq 0$.

To start the induction, we note that $x^0 \in A$, by the definition of A . Suppose that $x^k \in A$. By the definition of γ^k , we have

$$\gamma^k \|d^k\|^2 = |\nabla f(x^k)' d^k| \leq \|\nabla f(x^k)\| \cdot \|d^k\|.$$

Thus, $\|d^k\| \leq \|\nabla f(x^k)\|/\gamma^k \leq G/\gamma^k$. Hence,

$$\|x^k + \alpha^k d^k\| \leq \|x^k\| + \alpha^k G/\gamma^k \leq R + 2G,$$

which shows that $x^k + \alpha^k d^k \in B$. In order to prove Eq. (*), we now proceed as in the proof of Prop. 1.2.3. A difficulty arises because Prop. A.24 assumes that the inequality $\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$ holds for all x, y , whereas in this exercise this inequality holds only for $x, y \in B$. We thus essentially repeat the proof of Prop. A.24, to obtain

$$\begin{aligned} f(x^{k+1}) &= f(x^k + \alpha^k d^k) \\ &= \int_0^1 \alpha^k \nabla f(x^k + \tau \alpha^k d^k)' d^k d\tau \\ &\leq \alpha^k \nabla f(x^k)' d^k + \left| \int_0^1 \alpha^k \left(\nabla f(x^k + \alpha^k \tau d^k) - \nabla f(x^k) \right)' d^k d\tau \right| \quad (**) \\ &\leq \alpha^k \nabla f(x^k)' d^k + (\alpha^k)^2 \|d^k\|^2 \int_0^1 L \tau d\tau \\ &= \alpha^k \nabla f(x^k)' d^k + \frac{L(\alpha^k)^2}{2} \|d^k\|^2. \end{aligned}$$

We have used here the inequality

$$\|\nabla f(x^k + \alpha^k \tau d^k) - \nabla f(x^k)\| \leq \alpha^k L \tau \|d^k\|,$$

which holds because of our definition of L and because $x^k \in A \subset B$, $x^k + \alpha^k d^k \in B$ and (because of the convexity of B) $x^k + \alpha^k \tau d^k \in B$, for $\tau \in [0, 1]$.

Inequality (*) now follows from Eq. (**) as in the proof of Prop. 1.2.3. In particular, we have $f(x^{k+1}) \leq f(x^k) \leq f(x^0)$ and $x^{k+1} \in A$. This completes the induction. The remainder of the proof is the same as in Prop. 1.2.3.

1.2.10 www

We have

$$\nabla f(x) - \nabla f(x^*) = \int_0^1 \nabla^2 f(x^* + t(x - x^*)) (x - x^*) dt$$

and since

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

we obtain

$$(x - x^*)' \nabla f(x) = \int_0^1 (x - x^*)' \nabla^2 f(x^* + t(x - x^*)) (x - x^*) dt \geq m \int_0^1 \|x - x^*\|^2 dt.$$

Using the Cauchy-Schwartz inequality $(x - x^*)' \nabla f(x) \leq \|x - x^*\| \|\nabla f(x)\|$, we have

$$m \int_0^1 \|x - x^*\|^2 dt \leq \|x - x^*\| \|\nabla f(x)\|,$$

and

$$\|x - x^*\| \leq \frac{\|\nabla f(x)\|}{m}.$$

Now define for all scalars t ,

$$F(t) = f(x^* + t(x - x^*))$$

We have

$$F'(t) = (x - x^*)' \nabla f(x^* + t(x - x^*))$$

and

$$F''(t) = (x - x^*)' \nabla^2 f(x^* + t(x - x^*)) (x - x^*) \geq m \|x - x^*\|^2 \geq 0.$$

Thus F' is an increasing function, and $F'(1) \geq F'(t)$ for all $t \in [0, 1]$. Hence

$$\begin{aligned} f(x) - f(x^*) &= F(1) - F(0) = \int_0^1 F'(t) dt \\ &\leq F'(1) = (x - x^*)' \nabla f(x) \\ &\leq \|x - x^*\| \|\nabla f(x)\| \leq \frac{\|\nabla f(x)\|^2}{m}, \end{aligned}$$

where in the last step we used the result shown earlier.

1.2.11 www

Assume condition (i). The same reasoning as in proof of Prop. 1.2.1, can be used here to show that

$$0 \leq \nabla f(\bar{x})' \bar{p}, \tag{1}$$

where $\bar{x}$ is a limit point of $\{x^k\}$, namely $\{x^k\}_{k \in \bar{\mathcal{K}}} \longrightarrow \bar{x}$, and

$$p^k = \frac{d^k}{\|d^k\|}, \quad \{p^k\}_{k \in \bar{\mathcal{K}}} \rightarrow \bar{p}. \tag{2}$$

Since ∇f is continuous, we can write

$$\begin{aligned} \nabla f(\bar{x})' \bar{p} &= \lim_{k \rightarrow \infty, k \in \bar{\mathcal{K}}} \nabla f(x^k)' p^k \\ &= \liminf_{k \rightarrow \infty, k \in \bar{\mathcal{K}}} \nabla f(x^k)' p^k \\ &\leq \frac{\liminf_{k \rightarrow \infty, k \in \bar{\mathcal{K}}} \nabla f(x^k)' d^k}{\limsup_{k \rightarrow \infty, k \in \bar{\mathcal{K}}} \|d^k\|} < 0, \end{aligned}$$

which contradicts (1). The proof for the other choices of stepsize is the same as in Prop.1.2.1.

Assume condition (ii). Suppose that $\nabla f(x^k) \neq 0$ for all k . For the minimization rule we have

$$f(x^{k+1}) = \min_{\alpha \geq 0} f(x^k + \alpha d^k) = \min_{\theta \geq 0} f(x^k + \theta p^k), \quad (3)$$

for all k , where $p^k = \frac{d^k}{\|d^k\|}$. Note that

$$\nabla f(x^k)' p^k \leq -c \|\nabla f(x^k)\|, \quad \forall k. \quad (4)$$

Let $\hat{x}^{k+1} = x^k + \hat{\alpha}_k p^k$ be the iterate generated from x^k via the Armijo rule, with the corresponding stepsize $\hat{\alpha}_k$ and the descent direction p^k . Then from (3) and (4), it follows that

$$f(x^{k+1}) - f(x^k) \leq f(\hat{x}^{k+1}) - f(x^k) \leq \sigma \hat{\alpha}_k \nabla f(x^k)' p^k \leq -\sigma c \hat{\alpha}_k \|\nabla f(x^k)\|^2. \quad (5)$$

Hence, either $\{f(x^k)\}$ diverges to $-\infty$ or else it converges to some finite value. Suppose that $\{x^k\}_{k \in \mathcal{K}} \rightarrow \bar{x}$ and $\nabla f(\bar{x}) \neq 0$. Then, $\lim_{k \rightarrow \infty, k \in \mathcal{K}} f(x^k) = f(\bar{x})$, which combined with (5) implies that

$$\lim_{k \rightarrow \infty, k \in \mathcal{K}} \hat{\alpha}_k \|\nabla f(x^k)\|^2 = 0.$$

Since $\lim_{k \rightarrow \infty, k \in \mathcal{K}} \nabla f(x^k) = \nabla f(\bar{x}) \neq 0$, we must have $\lim_{k \rightarrow \infty, k \in \mathcal{K}} \hat{\alpha}_k = 0$. Without loss of generality, we may assume that $\lim_{k \rightarrow \infty, k \in \mathcal{K}} p^k = \bar{p}$. Now, we can use the same line of arguments as in the proof of the Prop. 1.2.1 to show that (1) holds. On the other hand, from (4) we have that

$$\lim_{k \rightarrow \infty, k \in \mathcal{K}} \nabla f(x^k)' p^k = \nabla f(\bar{x})' \bar{p} \leq -c \|\nabla f(\bar{x})\| < 0.$$

This contradicts (1), so that $\nabla f(\bar{x}) = 0$.

1.2.13 www

Consider the stepsize rule (i). From the Descent Lemma (cf. the proof of Prop. 1.2.3), we have for all k

$$f(x^{k+1}) \leq f(x^k) - \alpha^k \left(1 - \frac{\alpha^k L}{2}\right) \|\nabla f(x^k)\|^2.$$

From this relation, we obtain for any minimum x^* of f ,

$$f(x^*) \leq f(x^0) - \frac{\epsilon}{2} \sum_{k=0}^{\infty} \|\nabla f(x^k)\|^2.$$

It follows that $\nabla f(x^k) \rightarrow 0$, that $\{f(x^k)\}$ converges, and that $\sum_{k=0}^{\infty} \|\nabla f(x^k)\|^2 < \infty$, from which

$$\sum_{k=0}^{\infty} \|x^{k+1} - x^k\|^2 < \infty,$$