Nonlinear Optimization

by Andrzej Ruszczyński

Errata

Below is the list of typos in the 1st printing, which remained in the 4th printing. They are all absolutely inconsequential.

You have the 4th printing if you see “10 9 8 7 6 5 4” on the bottom of the copyright page.

Fourth Printing

page 34, line 1: Replace "Theorem 2.32" by "Corollary 2.29"

page 61, lines 18 and 22: Replace the set $\left\{ \begin{bmatrix} g \\ -1 \end{bmatrix} : g \in \partial\psi(0) \right\}$ with

cone $\left\{ \begin{bmatrix} g \\ -1 \end{bmatrix} : g \in \partial\psi(0) \right\}$

page 62, line -4: Replace “subdifferentiable” by “continuous”.

page 63, line 7: Replace “Lemma 2.75” by “Theorem 2.74”.

page 65, line 3: Replace the formula with

$$\|y - \hat{z}\|_{\diamond} \geq f(y) \geq f(x) + \langle g, y - x \rangle.$$

($f(x)$ was missing on the right hand side).

page 65, line -1: The right hand side should read $= \left\{ \frac{x - \Pi_Z(x)}{\|x - \Pi_Z(x)\|} \right\}$.

page 66, line 5: The right hand side should read $= \frac{x - \Pi_Z(x)}{\|x - \Pi_Z(x)\|}$.

page 67, line -2: Replace $\overline{\mathbb{R}}$ by $\mathbb{R}$.

page 104, line -7: Replace $\|x_{\mathrm{R}}(\tau) - x_0\|$ by $x_{\mathrm{R}}(\tau) - x_0$.

page 105, line 6: Replace $i \in I^0(x_0)$ by $i = 1, \ldots, m$.

page 153, line 7: $g_i(\hat{x}(b)) < 0$ should be $g_i(\hat{x}(b)) < b_i$.

page 156, line 12: replace $x_2 R_3$ by $x_3 R_3$.

page 157, line 15: remove the second “will”.

page 162, line -7: “page 145” should be “page 29”.

page 195, line 12: $b \in v^*(-\lambda)$ should be $b \in \partial v^*(-\lambda)$.

page 230, line -7: “Theorem 5.11” should be “Theorem 5.7”.

page 231, line -14: “Theorem 5.29” should be “Theorem 5.8”.

page 234, line -1: There should be minus sign before $\langle \nabla f(x^k), \left[\nabla^2 f(x^k)\right]^{-1} \nabla f(x^k) \rangle$.

page 244, line -4 and -1: There should be no minus sign in the formula for α_k.

page 245, line 2: There should be no minus sign in front of $\|\nabla f(x^{k-1})\|^2$.

page 247, line -6: There should be coefficient $\frac{1}{2}$ front of the first term on the rhs.

page 252, line -7: There should be coefficient $\frac{1}{2}$ in front of the first term on the rhs.

page 282, line -3: $9x_1 + 16x_2$ should be $9x_1 + 16|x_2|$.

page 282, line -1: $\frac{9}{16}$ should be $\frac{9^2}{16^2}$.

page 283, line 5: $\frac{9}{16}$ should be $\frac{9^2}{16^2}$.

page 314, line -12: “Theorem 4.79” should be “Theorem 4.33”.

page 319, line 11: “Lemma 6.44” should be “Lemma 6.15”.

page 327, line -1: The last line should read: $A_k d^k = -g^0(x^k)$.

page 338, line -4: In (6.85) the second line should be $ZY\mathbb{1} = \sigma\mathbb{1}$.

page 397, line 19: (7.93) should be (7.92)

Second Printing

Below is the list of errors in the 1st printing, which remained in the 2nd printing.

You have the 2nd printing if you see “10 9 8 7 6 5 4 3 2” on the bottom of the copyright page.

page 61, lines 11–28: Replace the text by the following:

LEMMA 2.75. *If a convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is subdifferentiable at a point x, then for every d*

$$\liminf_{h\to d} f'(x;h) = \sup_{g\in\partial f(x)} \langle g, d\rangle.$$

Proof. Consider the function $\psi(d) = f'(x;d)$. As $f(x+d) \ge f(x) + \psi(d)$, we have $\partial\psi(0) \subset \partial f(x)$. The function $\psi(\cdot)$ is convex and positively homogeneous, and thus its epigraph is a convex cone. A vector $g \in \partial\psi(0)$ if and only if $\alpha \ge \langle g, d\rangle$ for all $(d,\alpha) \in \mathrm{epi}(\psi)$. This is the same as

$$\mathrm{cone}\left\{\begin{bmatrix} g \\ -1 \end{bmatrix} : g \in \partial\psi(0)\right\} = \left[\mathrm{epi}(\psi)\right]^\circ = \left[\overline{\mathrm{epi}(\psi)}\right]^\circ.$$

In the last equation we used Lemma 2.25(ii). The cone $\overline{\mathrm{epi}(\psi)}$ is the epigraph of the convex positively homogeneous function $\varphi(d) = \liminf_{h\to d} f'(x;h)$. Applying Theorem 2.27 to the last displayed equation we conclude that

$$\mathrm{epi}(\varphi) = \left[\mathrm{epi}(\varphi)\right]^{\circ\circ} = \left[\mathrm{cone}\left\{\begin{bmatrix} g \\ -1 \end{bmatrix} : g \in \partial\psi(0)\right\}\right]^\circ.$$

Thus $(d, \alpha) \in \operatorname{epi}(\varphi)$ if and only if $\langle g, d\rangle \le \alpha$ for all $g \in \partial\psi(0)$, so

$$\varphi(d) = \sup_{g\in\partial\psi(0)} \langle g, d\rangle \le \sup_{g\in\partial f(x)} \langle g, d\rangle.$$

On the other hand, by Definition 2.72 for all $g \in \partial f(x)$ we have

$$f'(x; h) = \lim_{\tau\downarrow 0} \frac{1}{\tau}\big[f(x+\tau h) - f(x)\big] \ge \langle g, h\rangle.$$

Letting $h \to d$ we get $\varphi(d) \ge \sup_{g\in\partial f(x)}\langle g, d\rangle$. The postulated equation is true, and we have additionally established that $\partial f(x) = \partial\psi(0)$. □

pages 68–70: The proof of Theorem 2.85 as presented in the book is correct only in the interior of the domain of f. The general proof should be the following:

Proof. Consider the directional derivatives

$$\psi(d) = f'(x; d), \qquad \psi_1(d) = f_1'(x; d), \qquad \psi_2(d) = f_2'(x; d),$$

as functions of the direction d with values in $\overline{\mathbb{R}}$. They are convex and positively homogeneous, and thus their epigraphs are convex cones. In the proof of Lemma 2.75 we established that $\partial f(x) = \partial\psi(0)$, $\partial f_1(x) = \partial\psi_1(0)$, and $\partial f_2(x) = \partial\psi_2(0)$. We have to prove that $\partial\psi(0) = \partial\psi_1(0) + \partial\psi_2(0)$.

Define the following two cones:

$$C_1 = \left\{ \begin{bmatrix} d \\ \alpha_1 \\ \alpha_2 \end{bmatrix} \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} : \begin{bmatrix} d \\ \alpha_1 \end{bmatrix} \in \operatorname{epi}(\psi_1) \right\},$$

$$C_2 = \left\{ \begin{bmatrix} d \\ \alpha_1 \\ \alpha_2 \end{bmatrix} \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} : \begin{bmatrix} d \\ \alpha_2 \end{bmatrix} \in \operatorname{epi}(\psi_2) \right\}.$$

We also define the linear operator $A : \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}^n \times \mathbb{R}$ as follows:

$$A \begin{bmatrix} d \\ \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} d \\ \alpha_1 + \alpha_2 \end{bmatrix}.$$

Then

$$\operatorname{epi}(\psi) = A(C_1 \cap C_2).$$

In the proof of Lemma 2.75 we also established that

$$\left\{ \begin{bmatrix} g \\ -1 \end{bmatrix} : g \in \partial\psi(0) \right\} = \big[\operatorname{epi}(\psi)\big]^\circ.$$

It remains to calculate the polar cone to $\operatorname{epi}(\psi)$. We have

$$\big[\operatorname{epi}(\psi)\big]^\circ = \left\{ \begin{bmatrix} g \\ \beta \end{bmatrix} \in \mathbb{R}^n \times \mathbb{R} : A^T \begin{bmatrix} g \\ \beta \end{bmatrix} \in \big(C_1 \cap C_2\big)^\circ \right\}.$$

Choose $\alpha_i > f_i'(x; x_0 - x)$, $i = 1, 2$. Then the point $(x_0 - x, \alpha_1, \alpha_2)$ is in $\operatorname{int}(C_1) \cap C_2$ and we can apply Theorem 2.35 to conclude that

$$
\begin{aligned}
&\big(C_1 \cap C_2\big)^\circ = C_1^\circ + C_2^\circ = \\
&\left\{ \begin{bmatrix} g_1 \\ -1 \\ 0 \end{bmatrix} \in \mathbb{R}^{n+2} : g_1 \in \partial\psi_1(0) \right\} + \left\{ \begin{bmatrix} g_2 \\ 0 \\ -1 \end{bmatrix} \in \mathbb{R}^{n+2} : g_2 \in \partial\psi_2(0) \right\} \\
&\qquad\qquad = \left\{ \begin{bmatrix} g_1 + g_2 \\ -1 \\ -1 \end{bmatrix} \in \mathbb{R}^{n+2} : g_1 \in \partial\psi_1(0),\ g_2 \in \partial\psi_2(0) \right\}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\big[\operatorname{epi}(\psi)\big]^\circ \\
&= \left\{ \begin{bmatrix} g \\ \beta \end{bmatrix} \in \mathbb{R}^{n+1} : A^T \begin{bmatrix} g \\ \beta \end{bmatrix} = \begin{bmatrix} g_1 + g_2 \\ -1 \\ -1 \end{bmatrix} : g_1 \in \partial\psi_1(0),\ g_2 \in \partial\psi_2(0) \right\} \\
&= \left\{ \begin{bmatrix} g_1 + g_2 \\ -1 \end{bmatrix} \in \mathbb{R}^{n+1} : g_1 \in \partial\psi_1(0),\ g_2 \in \partial\psi_2(0) \right\}.
\end{aligned}
$$

Thus

$$\partial\psi(0) = \partial\psi_1(0) + \partial\psi_2(0),$$

as required. □

page 107, line 13: Replace "$=$" by "$\geq$".

page 107, lines 19–29: Replace the text by the following:
By Lemma 2.75, if d is an element of the set on the right hand side, then

$$\liminf_{h \to d} g'(x_0; h) = \sup_{s \in \partial g(x_0)} \langle s, d \rangle \leq 0.$$

We can thus find a sequence of directions $h^k \to d$ and scalars $\delta_k \downarrow 0$ such that $g'(x_0; h^k) \leq \delta_k$. For sufficiently large k we construct the directions

$$d^k = (1 - \alpha_k)h^k + \alpha_k(x_S - x_0), \quad \alpha_k \in [0, 1],$$

where x_S is the Slater point. Suppose $\alpha_k > \delta_k/(\delta_k - g(x_S))$. We can estimate the directional derivative as follows:

$$
\begin{aligned}
g'(x_0; d^k) &\leq (1 - \alpha_k)g'(x_0; h^k) + \alpha_k g'(x_0; x_S - x_0) \\
&\leq (1 - \alpha_k)\delta_k + \alpha_k g(x_S) < 0.
\end{aligned}
$$

As $\delta_k \downarrow 0$, for all sufficiently large k it is possible to choose α_k satisfying these conditions, and such that $\alpha_k \to 0$, as $k \to \infty$. Then $d^k \to d$.

First Printing

Below is the list of errors in the 1st printing, which were corrected in the 2nd printing. You have the 1st printing if you see "10 9 8 7 6 5 4 3 2 1" on the bottom of the copyright page.

page 12, line -3: Replace $\|F_{k-1} + G\|$ by $\|F_{n-1} + G\|$

page 12, line -2: Replace "work" by "energy"

page 12, line -1: Replace the formula by $f_2(F) = \displaystyle\sum_{k=1}^{n-1} \|F_k\|^2$

page 25, line -9: Replace "2.16" by "2.17".

page 29, line -1: Replace $\{A^T\lambda : \lambda \in C^\circ\}$ by $\overline{\{A^T\lambda : \lambda \in C^\circ\}}$.

page 30, line 5: Replace $\{A^T\lambda : \lambda \in C^\circ\}$ by $\overline{\{A^T\lambda : \lambda \in C^\circ\}}$.

page 35, line -5: Replace $\{A^T\lambda : \lambda \in K_2^\circ\}$ by $\overline{\{A^T\lambda : \lambda \in K_2^\circ\}}$.

page 36, lines 10–23: Replace the text with the following:

$$z = \lim_{j\to\infty} A^T\lambda^j, \quad \lambda^j \in K_2^\circ.$$

It follows that there exists a sequence $e^j \to 0$ such that $-A^T\lambda^j - e^j \in K_1^\circ$. Thus, for every $x \in K_1$ we obtain $-\langle x, A^T\lambda^j + e^j\rangle \le 0$, that is,

$$\langle Ax, \lambda^j\rangle + \langle x, e^j\rangle \ge 0.$$

As $\lambda^j \in K_2^\circ$, we also have $\langle y, \lambda^j\rangle \le 0$ for all $y \in K_2$. Hence

$$\langle Ax - y, \lambda^j\rangle + \langle x, e^j\rangle \ge 0 \quad \text{for all} \quad x \in K_1 \quad \text{and all} \quad y \in K_2.$$

Suppose that $\|\lambda^j\| \ge \delta > 0$ for an infinite set of indices $j \in \mathcal{J}$. Dividing by $\|\lambda^j\|$ and passing to the limit (over a sub-subsequence, if necessary), we obtain

$$\langle Ax - y, \lambda\rangle \ge 0 \quad \text{for all} \quad x \in K_1 \quad \text{and all} \quad y \in K_2,$$

where λ is an accumulation point of $\lambda^j/\|\lambda^j\|$ for $j \in \mathcal{J}$. Clearly, $\|\lambda\| = 1$. Since the set of all $Ax - y$, where $x \in K_1$ and $y \in K_2$, contains a neighborhood of 0 by assumption, we obtain a contradiction. Thus we must have $\lambda^j \to 0$, as $j \to \infty$. This yields $z = 0$ and $w = 0$, a contradiction again. Consequently, $\alpha > 0$.

Dividing (2.13) by α and rearranging terms we obtain

$$v = w/\alpha + \lim_{j\to\infty} A^T\big(\lambda^j/\alpha\big), \quad w \in K_1^\circ, \quad \lambda^j \in K_2^\circ.$$

Thus $v \in K_1^\circ + D^\circ$, as required.

page 36, line -2: Replace “is not needed” by “and the closure operation in K° are not needed.”

page 49, line -5: Replace “=” by “$\le$”.

page 60, line -3: “Lemma 2.36” should be “Lemma 2.70”.

page 79, line 8: Replace “$=$” by “$\stackrel{\triangle}{=}$”. Replace $\frac{s}{\alpha}$ by $\frac{s}{\gamma}$.

page 85, problem 2.7: Replace K° by $(K_1 \cap \cdots \cap K_m)^\circ$.

page 97, line -11: Remove "closed".

page 97, line -8: Replace "Let Y_i, $i \in I$, be the given sets and let" by "Let Y_i, $i \in I$, be the given sets. Suppose they are compact and let"

page 98, line 14: Add "If the sets Y_i are not compact, we construct their compact subsets satisfying the assumptions. We define a finite set S by choosing one element in each intersection of $n+1$ sets Y_i, and we replace each Y_i by $\operatorname{conv}(Y_i \cap S)$."

page 102, line 6: Replace y_0 by $g(x_0)$.

page 106, line 16: “Lemma 3.10” should be “Lemma 3.16”.

page 106, line -3: Replace $\{[g'(x_0)]^T\lambda : \lambda \in [T_{Y_0}(g(x_0))]^\circ\}$ by $\overline{\{[g'(x_0)]^T\lambda : \lambda \in [T_{Y_0}(g(x_0))]^\circ\}}$.

page 143, line 7: In formula (3.68) in the middle row replace $\frac{\tau^2}{2}$ by $\frac{1}{2}$

page 167, line 9: Replace “Theorem 3.33” by “Theorem 3.34”.

page 167, line 11: Replace “Theorem 3.46” by “Theorem 3.33”.

page 167, line -9: replace “Theorem 3.46” by “Theorem 3.33”.

page 188, line -8: In example 4.23, line 3, exchange i and j.

page 197, line 10: Remove ϱ from the first sum and $\hat{\mu}_i$ from the second sum.

page 198, line 14: Replace $\nabla^2 f(\hat{x})$ by $\nabla^2_{xx} L(\hat{x}, \hat{\mu})$.

page 201, line 4: Replace “$\hat{z}_i = -g_i(\hat{x})$ ” by “ $(\hat{z}_i)^2 = -g_i(\hat{x})$.”

page 205, line -4: Replace “$\ge$” by “$\le$”.

page 206, line 3: replace $x_j \in \{0, 1\}$ by $x_j \in \{-1, 1\}$.

page 206, lines 4–5: Replace “and is not assumed to be positive semidefinite (otherwise 0 is an optimal solution)” by “. Such Problems are used to classify n objects into two groups, with c_{ij} representing dissimilarity of objects i and j.”

page 221, line 14: $f(x^{k+1}) - f(x^k)$ should be $f(x^k) - f(x^{k+1})$.

page 228, line -1: $\|g\|^2$ should be $\|g\|^4$ (twice).

page 229, line 7: $\|g\|^2$ should be $\|g\|^4$.

page 229, line 15: $\|g\|^2$ should be $\|g\|^4$.

page 241, line -4: "Theorems 5.29 and 5.33" should be "Theorems 5.8 and 5.10"

page 244, line -6: "an" should be "and"

page 329, line -8: Delete one "$(d^k, \hat{\lambda}^k, \hat{\mu}^k)$".

page 332, line -8: In (6.79) replace $g(x) = z$ by $g(x) = -z$.

page 332, line -7: Replace $g(x_S)$ by $-g(x_S)$.

page 332, line -6: Replace $g(x_S)$ by $-g(x_S)$.

page 333, line 1: Replace $z = g(\bar{x})$ by $z = -g(\bar{x})$.

page 333, line 8: Replace $g_i(x)$ by $-g_i(x)$.

page 334, line 3: In (6.81) replace $g(x) = z$ by $g(x) = -z$.

page 344, line 10: Replace $\begin{bmatrix} 0 \\ -4 \end{bmatrix}$ by $\begin{bmatrix} 1 \\ -3 \end{bmatrix}$.

page 412, line 11: Replace $Y = \overline{\operatorname{int} Y}$ by $Y \subset \overline{\operatorname{int} Y}$.

page 442, line -10: Replace the text by: "Mathematics of Operations Research **31** (2006), 433–452."

Many thanks to Tsvetan Asamov, Matthias Claus, Kristian Debrabant, Jonathan Eckstein, Agust Egilsson, John Gubner, Jean-Baptiste Hiriart-Urruty, Robert Jenssen, Dosung Lee, Stephen Robinson, Matthew Saltzman, Shankar Subramanian, Andrew Swann, Xiao-Hui Wu, Hong-Kun Xu, and my Rutgers and Princeton students for pointing to me these inaccuracies.

If you find any errors in the book which are not listed here, please mail me at `rusz@business.rutgers.edu`