Optimization for Learning and Control

Errata

October 28, 2025

Chapter 2

- p. 13, 2.2.3, line -6: $\{\theta x + (1-\theta)y \mid \theta \in \mathbb{R}\} \subseteq \mathcal{A}$

- p. 21, eq. (2.27) and the following three lines: "$\mathbb{S}^n_+$" instead of "S^n_+" (4 instances).

- p. 22, first line: "$\mathcal{R}(B) \subsetneq \mathcal{R}(A)$" should be "$\mathcal{R}(B) \nsubseteq \mathcal{R}(A)$"

- p. 32, eq. (2.59): "$D \succeq 0$" instead of "$C \succeq 0$".

- p. 35, Exercise 2.2: "operator norm induced by the vector 2-norm" instead of "operator norm".

Chapter 3

- p. 45, in footnote 2 it reads "$\mathbb{Z}^n_+$ for some n" should be "$\mathbb{Z}_+$"

- p. 58, exercise 3.8: "Show that the for" should be "Show that for".

Chapter 4

- p. 65, eq. (4.8): "$\inf_z \{f_0(z) \mid z \in \mathcal{F} \cap B_2(x,r)\}$" instead of "$\inf_z \{f_0(z) \mid x \in \mathcal{F} \cap B_2(x,r)\}$".

- p. 68, line 5: $C \subseteq \mathbb{R}^n$ instead of $C \in \mathbb{R}^n$

- p. 70, line 1: "injective affine transformation" instead of "injective linear transformation".

- p. 70, first line of 4.2.2.4: "$K \subseteq \mathbb{R}^n$" instead of "$sK \in \mathbb{R}^n$".

- p. 70, line 6 of 4.2.2.4: "$K \cap (-K) = \{0\}$" instead of "$K \cap (-K) \in \{0\}$"

- p. 71, eq. (4.21) and 3 lines above (4.22): "$\mathbb{S}^n_+$" instead of "S^n_+".

- p. 72, middle: $a \succeq_K b,\ c \succeq_K d \quad \Rightarrow \quad a + c \succeq_K b + d$

- p. 75, ineq. below (4.33): $(\nabla f(x+tz) - \nabla f(x))^T(x+tz-x) \geq 0$.; "Dividing by t^2 and taking the limit..."

- p. 75, 7 lines above 4.3.2: "$\phi''(t) = (y-x)^T \nabla^2 f(x + t(y-x))(y-x)$" instead of "$\phi''(t) = (y-x)^T \nabla^2 f(x)(y-x)$".

- p. 76: "perspective of a function" instead of "perspective function of a function"

- p. 77, last line before 4.3.2.5: epi P_f instead of epi $P_f(x, t)$

- p. 78, 4.3.3.1: "All norms are convex functions" instead of "All norms are convex function".

- p. 81, Figure 4.16 (a): the text in the figure is partly cut off.

- p. 82, first line in 4.4.1.2: replace "it holds that" with "the subdifferential of their sum $f = f_1 + f_2$ satisfies"

- p. 86, 2 lines below (4.59): "$f_i(x) \le 0$" instead of "$f_i(x) \le 0$".

- p. 88: "$h : \mathbb{R}^n \times \mathbb{R}^q \to (-\infty, +\infty]$" instead of "$h : \mathbb{R}^n \times \mathbb{R}^q$".

- p. 89, Example 4.8: "$h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$" instead of "$h : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$", and all instances of $I_{\mathbb{R}^n_+}$ and $S_{\mathbb{R}^n_+}$ should be $I_{\mathbb{R}^m_+}$ and $S_{\mathbb{R}^m_+}$, respectively.

- p. 89, last line: relint $(\operatorname{dom} f) \cap$ relint $(\operatorname{dom} g) \ne \emptyset$ should be relint $A(\operatorname{dom} f) \cap$ relint $(\operatorname{dom} g) \ne \emptyset$

- p. 91, lines 1-2: "... Slater's condition implies the existence of primal and dual optimal points $x^\star$ and $(\lambda^\star, \mu^\star)$ with zero duality gap." should be "Slater's condition implies strong duality and that the dual optimal value is attained if the dual problem is feasible. However, Slater's condition does not guarantee that the primal optimal value is attained."

- p. 92, Exercise 4.5: "$x, v \in \mathbb{R}^n$" instead of "$v \in \mathbb{R}^n$".

- Exercise 4.8 reads "$f : X \to \mathbb{R}$" should be "$f : X \to \mathbb{R}_+$". Also, the displayed formula should be changed to:
$$\varphi\left(\int_X x f(x) dx\right) \le \int_X \varphi(x) f(x) dx$$

Chapter 5

- p. 97, eq. (5.10): $p \in \Delta^n$ instead of $x \in \Delta^n$

- p. 98, 5 lines below (5.13): "$\mathbb{S}^n_+$" instead of "S^n_+".

- p. 99, line 7: "Now, suppose $q \notin \mathcal{R}(P)$"

- Section 5.3.1, from the definition of g and below the sign in front of A should be changed. Line 4 from below: "$c + A^T \mu \in K^*$" instead of "$c - A^T \lambda \in K^*$"

- p. 100, eq. (5.19): "$| x \le 0$" instead of "$|, x \le 0$".

- p. 101: "log-sum-exp" instead of "log-sup-exp"

- Sec. 5.5, p. 106: "Partial Separability" instead of "Partially Separability"

- p. 109, line 17 from below reads "is can" should be "can"

- p. 103: rank(Z) instead of rank(X) in (5.24a); $\|A - Z\|_F^2$ instead of $\|A - X\|_F^2$ five lines below (5.24b); $\|u\|_2$ instead of $\|u\|$ four lines from bottom.

- p. 109, line 1: $V_k : \mathbb{R}^n \to \bar{\mathbb{R}}$ should be $V_k : \mathbb{R} \to \bar{\mathbb{R}}$

- p. 109, three lines below (5.36): "partial separability" instead of "partially separability"

- p. 110, line 2 from below the last $+$ should be $-$

- Exercise 5.1(a), an absolute value is missing in the epigraph formulation, and in the equivalent LP one constraint is missing.

Chapter 6

- p. 122, Algorithm 6.1: "initial step size $t > 0$" instead of "initial step size $t \geq 0$"

- p. 122, above Algorithm 6.2: "implies that $\phi(t) \to \infty$" should be "implies that $\phi(t) \to -\infty$"

- p. 122, Algorithm 6.2, 4 lines from below $\beta < \infty$ should be $u < \infty$

- p. 123, 3 lines above Alg. 6.3: "a basic trust-region method is" instead of "a basic trust-region is".

- p. 127, line 1: "is Lipschitz" instead of "it is Lipschitz".

- p. 129, §6.3.1, below first displayed equation: $\nabla^2 f(x) \succ 0$ instead of $\nabla f(x) \succ 0$

- p. 129, Alg. 1: "α_1" instead of "α" (2 instances).

- Sec. 6.3.2.2: "α_1" instead of "α" (5 instances).

- p. 131, line 4: $\frac{1}{2}\Delta x_{\mathrm{nt}}^T \nabla^2 f(x_k)\Delta x_{\mathrm{nt}}$ instead of $\Delta x_{\mathrm{nt}}^T \nabla^2 f(x_k)\Delta x_{\mathrm{nt}}$

- p. 142, inline equations with expression for $\nabla^2 f_0(x)$ (lines -10 and -6): "$f_i(x)\nabla^2 f_i(x)$" instead of "$f_i(x)^2\nabla^2 f_i(x)$".

- p. 145, below (6.63): "a realization of an estimator G_k of $\nabla f(x_k)$" instead of "a realization of an estimator of $G_k \in \mathbb{R}^n$ of $\nabla f(x_k)$"

- p. 145, 4 lines above 6.8.1: "in some iterations" instead of "in some iteration"

- p. 149, Example 6.6: "$g_k = \gamma x_k + \frac{1}{p} \dots$" should be "$g_k = \gamma x_k - \frac{1}{p} \dots$"

- p. 151, 6.8.4.1, line 7: "AdaGrad uses either $B_k = \gamma \operatorname{diag}(\operatorname{diag}(\hat{G}_k) + \varepsilon \mathbb{1})^{1/2}$" instead of "AdaGrad uses either $B_k = \gamma \operatorname{diag}((\operatorname{diag}(\hat{G}_k) + \varepsilon \mathbb{1})^{1/2}$"

- p. 159, line 8: "and $-\phi$ is a so-called generalized logarithm for K"

- p. 161, three lines from below:

$$g_\rho(\mu) = \inf_x L_\rho(x, \mu) = \inf_x \left\{ f_0(x) + \frac{\rho}{2}\|h(x) + \rho^{-1}\mu\|_2^2 \right\} - \frac{1}{2\rho}\|\mu\|_2^2.$$

- p. 163, definition of L_ρ: "$-\frac{1}{2\rho}\|\mu\|_2^2$" instead of "$-\frac{\mu}{\rho}\|\mu\|_2^2$".

- Exercise 6.10(e), p. 169: "ADMM" instead of "ADDM"

Chapter 7

- p. 183, line 10 from above, the comma between the parenthesis should be removed.

- p. 184, line 10 from below, dt missing after the square root.

- p. 184, line 9 from below, $x(t)$ should be removed.

- p. 186, line 6 from above, $T^\star$ should be T.

- p. 199, line 12 from below, $f(t) = u(t)^2$ should be $f(x,u) = u^2$.

- p. 200, line 8 from below, $h_{n+1}(y) = 0$ and $h_{N-2}(y) = 0$ should be $h_{N+1}(y) = 0$ and $h_{N+2}(y) = 0$.

- p. 203, line 10 from below, $(10, 3)$ should be $(10, -3)$.

Chapter 8

- p. 217, line 5 from below, $Ax - Bu$ should be $Ax + Bu$ in two places.

- p. 238, line 3 from below, $V_k(f(x, \mu_k(x)))$ should be $V_k(F(x, \mu_k(x)))$.

Chapter 9

- p. 255, in the displayed formula after (9.9), the second line, $g^T(x)$ should be replaced with $g(x)^T$

- p. 255, in the displayed formula after (9.9), the second last line, $g(x)$ should be replaced with $\mathbb{E}[\,Y\,|\,X = x\,]$

- p. 256, line 10 from above contains an extra $= x$ that should be removed.

- p. 265, line 8 from above reads "i[109] proposes" should be "In [109], Viterbi proposed"

Chapter 10

- p. 301, line 3 from above, x should be replaced with x_k, and "$k \in \mathbb{N}_K$" should be replaces by "$k \in \mathbb{N}_N$"

- p. 301, definition of X: replace $\beta^T(x_i)$ by $\beta(x_i)^T$

- p. 308, definition of $g(\lambda, \mu)$: "$\beta(x_k)^T \beta(x_l)$" instead of "$\beta^T(x_k), \beta(x_l)$"

- p. 313, eqn. (10.10): "$\Phi_i(x_{i-1})$" instead of "$\Phi_i(x)$"

Chapter 11

- p. 343, line 11 from above reads "means" should be "mean"

- Exercise 11.2, the B-matrix is wrong in the Matlab file.