Corrections to Second Printing of *Numerical Optimization* (Last updated October 2, 2008)

To make it easier for the typesetter, we have included in many cases an explicit correction of the offending passages, marked by the tag "*corrected sentence*" or similar, so that he/she can cut-and-paste from the source (latex) file of this document into the latex file for the book. Note that in this document the labels of equations appearing in the *corrected sentences* may not print, or may print with wrong equation numbers, because they are using labels defined elsewhere in the book file. When inserted in the book file, they will produce the correct labels.

1. page 26, line 9. "+" should be "−" in this equation.

2. page 47, line −5. Replace g_k by ∇f_k in this formula.

3. page 49, Theorem 3.4. The theorem needs to be corrected, to include a slightly larger multiplier on the right-hand side of the inequality and to all the proviso that the inequality holds for all k sufficiently large. The corrected statement of the theorem is as follows:

   Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable, and that the iterates generated by the steepest-descent method with exact line searches converge to a point x^* where the Hessian matrix $\nabla^2 f(x^*)$ is positive definite. Let r be any scalar satisfying

   $$r \in \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}, 1 \right),$$

   where $\lambda_1 \leq \cdots \leq \lambda_n$ are the eigenvalues of $\nabla^2 f(x^*)$. Then for all k sufficiently large, we have

   $$f(x_{k+1}) - f(x^*) \leq r^2 [f(x_k) - f(x^*)].$$

4. pages 50-51. In Theorems 3.5 and 3.6 we only need to assume that f is twice continuously differentiable (not three times).

5. page 53, lines 1-5. This proof needs to be corrected so that we do not assume one of the things we are trying to prove, namely that $x_k \to x^*$. Replace the sentence "Since $\nabla^2 f(x^*)$ is nonsingular…" on line 1 by the following:

   Since $\nabla^2 f(x^*)$ is nonsingular, there is a radius $r > 0$ such that $\|\nabla^2 f_k^{-1}\| \leq 2\|\nabla^2 f(x^*)^{-1}\|$ for all x_k with $\|x_k - x^*\| \leq r$.

   Replace the sentence "Using this inequality…" on line 5 by the following:

   Choosing x_0 so that $\|x_0 - x^*\| \leq \min(r, 1/(2\tilde{L}))$, we can use this inequality inductively to deduce that the sequence converges to x^*, and the rate of convergence is quadratic.

6. 58: in the line before (3.44) it should read $\phi'(0) = \nabla f_{k-1}^T p_{k-1}$

7. page 68. The object $\bar{\Delta}$ is used in a different sense here from in the proof of Theorem 4.7 on pages 90-91. To avoid confusion, replace $\bar{\Delta}$ with $\hat{\Delta}$ on page 68, lines 2 and 10 of Algorithm 4.1, and on the line immediately following Algorithm 4.1.

8. page 68, line 7 of Algorithm 4.1 should read $\Delta_{k+1} = \frac{1}{4}\Delta_k$.

9. page 73, line 13. It should read: "...follows from Exercise 4.6."

10. page 71, line 16. before "to simplify the notation" insert ", and replace ∇f_k by g".

    Corrected phrase: so we drop the subscript "k" from the quantities Δ_k, p_k, and m_k, and replace ∇f_k by g to simplify the notation.

11. page 74, line 6. The sentence immediately after (4.14) should read:"This is a problem in two variables that is computationally inexpensive to solve. (After some algebraic manipulation one observes that it can be solved by finding the roots of a fourth-degree polynomial.)

12. page 75, line -7. There should be a "$-$" before r_{j+1}. The corrected formula is $d_{j+1} = -r_{j+1} + \beta_{j+1} d_j$.

13. page 80, Figure 4.5 and page 82, Figure 4.6. In both figures the $-\lambda^*$ along the horizontal axis should be λ^*.

14. page 87, last line. The right-hand side of this expression should be $m_k(-\Delta_k \nabla f_k / \|\nabla f_k\|) - f_k$. Corrected equation is:

$$m_k(p_k^{\mathrm{C}}) - m_k(0) \quad = \quad m_k(-\Delta_k \nabla f_k / \|\nabla f_k\|) - f_k$$

15. page 89, lines -7 to -5: The phrase "and that the level set ... is bounded." should be replaced by "and that f is bounded below on the level set ...." (The formula (4.38) remains the same.) The new text fragment is:

    and that f is bounded below on the level set

$$\{x \mid f(x) \leq f(x_0)\}. \tag{1}$$

16. page 90, equation (4.44). Delete "2" from the denominator. Corrected equation is

$$|\rho_k - 1| \leq \frac{\gamma \Delta_k (\beta \gamma \Delta_k / 2 + C_4(p_k))}{c_1 \epsilon \min(\Delta_k, \epsilon/\beta)}. \tag{2}$$

17. page 91, line 6. Delete 2 from the denominator of the middle expression and replace $\frac{1}{4}$ by $\frac{1}{2}$ in the rightmost expression. Corrected equation is

$$|\rho_k - 1| \leq \frac{\gamma \Delta_k c_1 \epsilon / (2\gamma)}{c_1 \epsilon \Delta_k} = \frac{1}{2}.$$

18. page 91, line 7. replace $\rho_k > \frac{3}{4}$ by $\rho_k > \frac{1}{4}$

19. page 109, line 10. $\in$ should be $\subset$.

20. page 115, display equation on line 19. "$\|x_{m+1} - x^*\|$" should be "$\|x_{m+1} - x^*\|_A$"
Corrected equation:
$$\|x_{m+1} - x^*\|_A \approx \epsilon\|x_0 - x^*\|_A.$$

21. page 118, equation (5.37), and also line 9. "$C^{-1}b$" should be "$C^{-T}b$" in both places.

22. page 118, line -4. Should read: "Set $p_0 = -y_0$, $k \leftarrow 0$."

23. page 119, line -5: it should read: "we will discuss it briefly in Chapter 6"

24. page 125, line 8. Missing "/" in the equation for $t(\xi)$. *Corrected equation:* $t(\xi) \stackrel{\text{def}}{=} (2\xi - 1)/(1 - \xi)$.

25. page 129, line 8. The "$\leq$" should be "$=$".

26. page 129, line 11, and also equation (5.65). In both places, χ_3 should be c_3.

27. page 139, line -5. Replace the phrase "we complete the first CG iteration, compute the new iterate $x^{(1)}$, and stop" by the phrase "we set $x^{(0)} = b = -\nabla f_k$ and return".

28. page 140, lines 7-8: Replace "p_k is the steepest descent direction $-\nabla f_k$" by "p_k is set to the steepest descent direction $-\nabla f_k$". Also, delete the sentence "This is the reason for choosing the initial estimate in the CG iteration as $x^{(0)} = 0$."

29. page 140, two lines before Algorithm 6.1 it should read: $\eta_k = \min(0.5, \sqrt{\|\nabla f_k\|})$ (i.e. gradient not Hessian).

30. page 140, line -6. The vector p should be p_k.

    Corrected sentence: $\nabla^2 f(x_k)p_k = -\nabla f_k$, starting from $x^{(0)} = 0$.

31. page 141, line 1: remove "minor" in "minor weakness". In line 11 remove "slightly".

32. page 148, lines 5 and 6 of Algorithm 6.5. These lines (Find index q ... and Interchange row ...) should be moved below the next line that starts with "**for** $j = 1, 2, \ldots, n$"

    8 lines below, the statement "if $j \leq n$" should read "if $j < n$".

33. page 170, line 2. Remove "set" after setting.

34. page 171, line after (7.16) and page 172, line 6. "Hessian" should be "Jacobian".

35. page 181, Figure 7.3. This figure should contain $p(6, 4) = e^2$ rather than $p(6, 4) = e$.

36. page 183, 4 lines after (7.34). "we saw in Chapter 9 ..." should be "we will see in Chapter 9 ...".

37. page 190, exercise 7.7. Should read: "expressing the intermediate derivatives ∇x_i, $i = 4, \ldots, 9$ in terms of quantities available at their parent nodes and then in terms of the independent variables x_1, x_2, x_3.

38. page 190, exercise 7.9. It should read: "for the evaluation point $x = (1, 2, \pi/2)^T$"

39. page 236, line -4. The terms $12x_3^2$ in both expressions should read $12x_3^2 - 4x_1$.

40. page 249, line 17. $i = 1, 2, q + 1$ should be $i = 1, 2, \ldots, q + 1$ (ellipsis missing).

41. page 269, line -11. The indices in the two summations should be j.

42. page 316, line -8. Replace $(-100, 0)$ by $(0, -100)$.

43. page 322, displayed formula on line -10. The denominator should contain an extra factor $\|\nabla c_1(x)\|$. Corrected formula reads:

$$d = -c_1(x) \frac{\nabla f(x)}{\|\nabla f(x)\| \|\nabla c_1(x)\|}.$$

44. page 329, formula (12.32). The fraction $\frac{1}{8}$ should be $\frac{1}{2}$. Corrected formula should read:

$$\min_x \left(x_1 - \frac{3}{2}\right)^2 + \left(x_2 - \frac{1}{2}\right)^4 \qquad \text{s.t.} \qquad \begin{bmatrix} 1 - x_1 - x_2 \\ 1 - x_1 + x_2 \\ 1 + x_1 - x_2 \\ 1 + x_1 + x_2 \end{bmatrix} \geq 0 \qquad (3)$$

45. page 334, equation (12.36) should read:

$$z_k = (-\sqrt{2}, 0)^T + (1/k)w.$$

46. page 336, line 12. The last word in the line should be "constraint" rather than "constrained". The line should read:

   One frequently used constraint qualification is the *linear independence constraint qualification* (LICQ) ...

47. page 339, line 14. The phrase should be "holds if $\nabla f(x^*)^T d \geq 0$ for all $d \in F_1$."

48. page 346, line 15. Immediately after the = sign, add "$f(x^*)+$". *Corrected equation:*

$$\begin{aligned} f(z_k) &\geq f(x^*) + \frac{1}{2}(z_k - x^*)^T \nabla_{xx} \mathcal{L}(x^*, \lambda^*)(z_k - x^*) + o(\|z_k - x^*\|^2) \\ &= f(x^*) + \frac{1}{2}\|z_k - x^*\|^2 d^T \nabla_{xx} \mathcal{L}(x^*, \lambda^*) d + o(\|z_k - x^*\|^2). \end{aligned}$$

49. page 354, 1 line below (12.75): x_j should be x_i.

   Corrected sentence: "(We simply take $x_i \equiv x^*$ in Definition **??**.)"

50. page 371, equation (13.14). The left-hand side of this equation should be A_p.

51. page 377, line 13. The "$-$" before the final term should be replaced by $+$.

    Corrected equation:

$$c^T x^+ = c_{\mathrm{B}}^T x_{\mathrm{B}} - (c_q - s_q) x_q^+ + c_q x_q^+ = c_{\mathrm{B}}^T x_{\mathrm{B}} + s_q x_q^+.$$

52. page 377, equation (13.24). The "$-$" before the final term should be replaced by $+$.

    Corrected equation:

$$c^T x^+ = c^T x + s_q x_q^+. \tag{4}$$

53. page 390, line 6. After "such that" append "$t_i > 0$, $t_l > 0$, and".

    Corrected passage: "we have a tie if there are two indices $i, l \in \mathcal{B}$ such that $t_i > 0$, $t_l > 0$, and"

54. page 391, line 7. Append "$t_i > 0$ for all $i \in \bar{\mathcal{B}}$ and"

    Corrected sentence: "Given a set of tied indices $\bar{\mathcal{B}} \subset \mathcal{B}$ with $t_i > 0$ for all $i \in \bar{\mathcal{B}}$ and"

55. page 391, lines 13-18. Replace these lines with the following material:

$$\frac{(B^{-1}E)_{i1}}{t_i} < \frac{(B^{-1}E)_{l1}}{t_l}, \qquad \text{for all } l \in \bar{\mathcal{B}},\ l \neq i,$$

    we choose i as the leaving index. Otherwise we retain in $\bar{\mathcal{B}}$ the indices i that tie for the smallest value of $(B^{-1}E)_{i1}/t_i$ and evaluate the second column of $B^{-1}E$. If just one value of $(B^{-1}E)_{i2}/t_i$ achieves the minimum, we choose it as the leaving index. Otherwise we consider coefficients $(B^{-1}E)_{ik}$ for successively higher k until just one candidate index remains.

56. page 406, equation (14.22). The second occurrence of $\alpha_{\mathrm{aff}}^{\mathrm{pri}}$ should be $\alpha_{\mathrm{aff}}^{\mathrm{dual}}$.

    Corrected equation:

$$\mu_{\mathrm{aff}} = (x + \alpha_{\mathrm{aff}}^{\mathrm{pri}} \Delta x^{\mathrm{aff}})^T (s + \alpha_{\mathrm{aff}}^{\mathrm{dual}} \Delta s^{\mathrm{aff}})/n.$$

57. page 406, line -3: delete the comma between "ΔX^{aff}" and "$\Delta S^{\mathrm{aff}} e$".

    Corrected fragment: $\Delta X^{\mathrm{aff}} \Delta S^{\mathrm{aff}} e$

58. page 407, equations (14.24a) and (14.24b). The algorithm is slightly better if the min with 1 is *not* performed.

    Corrected equations:

$$\begin{aligned} \alpha_{\max}^{\mathrm{pri}} &\stackrel{\mathrm{def}}{=} \min_{i:\Delta x_i < 0} -\frac{x_i^k}{\Delta x_i}, \\ \alpha_{\max}^{\mathrm{dual}} &\stackrel{\mathrm{def}}{=} \min_{i:\Delta s_i < 0} -\frac{s_i^k}{\Delta s_i}, \end{aligned}$$

59. page 416, lines 6 and 8. Remove "ω" from these equations.

60. page 435, below (15.25): remove the phrase ", $\|\cdot\|$ denotes the ℓ_2 norm,".

61. page 436, line 10: it should read $\mu < \mu^*$ instead of the other way round.

62. page 437, lines 7 and 8: Ψ should be replaced with ϕ.

63. page 444, lines 8 and 9. Replace "and that the constraints" by "so that the constraints"

64. page 454, 5-6 lines below (16.26d): Should read "active constraint gradients are linearly independent at the solution".

65. page 456, line -10: Should read: "First, linear dependence".

66. page 476, last line: add space after "to" .

67. page 483, equation (16.53). Replace "Δs" by "$\Delta\lambda$" and "$\Delta\lambda$" by "Δy". Also replace "ΛSe" by ΛYe.

    Corrected equation:

$$\begin{bmatrix} G & -A^T & 0 \\ A & 0 & -I \\ 0 & Y & \Lambda \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta\lambda \\ \Delta y \end{bmatrix} = \begin{bmatrix} -r_d \\ -r_b \\ -\Lambda Ye + \sigma\mu e \end{bmatrix}.$$

68. page 483, equation (16.54). Replace "Δs" by "$\Delta\lambda$"

69. page 484, lines 6-7. Replace all three occurrences of "s" in these lines by "λ". Also replace "S" by "Λ".

70. page 485, equation (16.56a): replace "$\max_{x,\lambda}$" by "$\max_\lambda \min_x$". Also, it should simply read (16.56) without the a.

71. page 487, exercise 16.10. Replace the first sentence by "Consider the problem (16.3a)-(16.3b), and assume that A has full row rank and that Z is the matrix whose columns span the null space of A." Also replace A by A^T in the last line of page 487.

72. page 487, last line. Replace A by A^T. Corrected inline equation reads "$GX^* + d = A^T\lambda^*$".

73. page 498, line 9. Replace "full column rank" by "full row rank".

74. page 499, 3 lines before (17.17). "ill conditioning of Q" should be "ill conditioning of $\nabla^2_{xx}Q$"

75. page 501.In figure 17.3 label the curves by value of μ.

76. page 515, line 8. Replace "and terminating when $\|\nabla_x \mathcal{L}_A(x, \lambda^k; \mu_k)\| \le \tau_k$" by "satisfying $\|\nabla_x \mathcal{L}_A(x_k, \lambda^k; \mu_k)\| \le \tau_k$". (This change is not necessary, but it improves clarity.)

77. page 515, add a new line after line 12: "Choose new tolerance $\tau_{k+1} \in (0, \tau_k)$;"

78. Page 517, Equation (17.53) should read ... $- \mu_k \lambda_i^k$

    i.e. μ is missing the subscript k. Same typo in (17.54) and (17.55)).

79. page 536, line 5. A comma is missing in $\mathcal{L}(x_k \lambda_k)$.

80. page 537, line 3. "A^T" should be "A_k^T".

81. page 537, line 6. "for for" should be "for".

82. page 542, line -1. Omit "$\alpha_k p_Z$".

83. page 543, line 3. Should read: "$p_k = Z_k p_{\mathrm{Z}} + Y_k p_{\mathrm{Y}}$. "

84. page 548, line 8. The term $D\phi(x_k; p_k)$ should read: "$D(\phi(x_k; \mu); p_k)$".

    The same correction should be made in page 552, line -3

85. page 551, line 7: the two "B" should be "M".

    Corrected passage: Set $M_{k+1} = M_k$;

86. page 560, line -6. A_k^T should be A_k.

    Corrected phrase: where Z_k is a basis for the null space of A_k.

87. page 574, Exercise 18.2. The starting point and solution are transposed. The two sentences should read: "Use the starting point $x_0 = (-1.8, 1.7, 1.9, -0.8, -0.8)$. The solution is approximately $x^* = (-1.71, 1.59, 1.82, -0.763, -0.763)$."

88. page 575, Exercise 18.10. Equation (18.69) should read

$$(I - A^T[AA^T]^{1}A)w$$

89. page 579. The definition of a cone should have $\alpha > 0$ instead of $\alpha \geq 0$. The corrected defintion should read as follows:

$$x \in \mathcal{F} \;\Rightarrow\; \alpha x \in \mathcal{F}, \;\; \text{for all } \alpha > 0. \tag{5}$$

90. page 579. The definition of affine hull in (A.3) and the example are wrong. The sentence containing the formula (A.3) and the sentence that follows should be replaced by the following.

    Corrected passage: An *affine set* in $\mathbb{R}^n$ is a the set of all vectors $x + y$, where x is given and y is any vector in a given subspace of $\mathbb{R}^n$. Given $\mathcal{F} \subset \mathbb{R}^n$, the *affine hull* of $\mathcal{F}$ (denoted by aff$\mathcal{F}$) is the smallest affine set containing $\mathcal{F}$. For instance, when $\mathcal{F}$ is the "ice-cream cone" defined as below by (**??**), we have aff$\mathcal{F} = \mathbb{R}^3$, while if $\mathcal{F}$ is the set of two isolated points $\mathcal{F} = \{(1,0,0), (0,2,0)\}$, we have

$$\text{aff}\mathcal{F} = \{(1,0,0) + \alpha(-1,2,0) \,|\, \text{for all } \alpha \in \mathbb{R}\}.$$

91. page 612: The year for reference [23] should be 1995.