Corrections to First Printing of Nocedal-Wright

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 24: equation above (2.15): index of the Hessian should be k, not $k+1$.

   Corrected sentence: When x_k and x_{k+1} lie in a region near the solution x^*, within which ∇f is positive definite, the final term in this expansion is eventually dominated by the $\nabla^2 f_k(x_{k+1} - x_k)$ term, and we can write

2. page 25: There are some errors in the sentence that includes (2.20).

   Corrected sentence: In fact, the equivalent formula for (**??**), applied to the inverse approximation $H_k \stackrel{\rm def}{=} B_k^{-1}$, is

3. page 26: There are spaces before two commas in the last paragraph before the new subsection (descent , Newton ,)

4. page 37: first and second lines: italicize *inexact* instead of line search.

5. page 46. We need to introduce a new equation since some of the references to (3.20) are not correct.

   Corrected Paragraph: For some algorithms, such as conjugate gradient methods, we will not be able to prove the limit (**??**), but only the weaker result

   $$\liminf_{k \to \infty} \|\nabla f_k\| = 0. \tag{1}$$

   In other words, just a subsequence of the gradient norms $\|\nabla f_{k_j}\|$ converges to zero, rather than the whole sequence (see Appendix **??**). This result too can be proved by using Zoutendijk's condition (**??**), but instead of a constructive proof, we outline a proof by contradiction. Suppose that (1) does not hold, so that the gradients remain bounded away from zero, that is, there exists $\gamma > 0$ such that

   $$\|\nabla f_k\| \geq \gamma, \quad \text{for all } k \text{ sufficiently large.} \tag{2}$$

   Then from (**??**) we conclude that

   $$\cos\theta_k \to 0, \tag{3}$$

   that is, the entire sequence $\{\cos\theta_k\}$ converges to 0. To establish (1), therefore, it is enough to show that a subsequence $\{\cos\theta_{k_j}\}$ is bounded away from zero. We will use this strategy in Chapter ?? to study the convergence of nonlinear conjugate gradient methods.

By applying this proof technique, we can prove global convergence in the sense of (??) or (1) for a general class of algorithms.

6. page 49: Last sentence before new section: Theorem 3.4 (not 3.3).

   Corrected Sentence: For example, if $\kappa(Q) = 800$, $f(x_1) = 1$ and $f(x^*) = 0$, Theorem ?? suggests that the function value will still be about 0.08 after one thousand iterations of the steepest-descent method.

7. page 78, line 12. Add the words "p^* is feasible and " after the words "if and only if".

   Corrected phrase: if and only if p^* is feasible and there is a scalar $\lambda \geq 0$ such that the following conditions are satisfied:

8. page 79, equation (4.21). Minus signs are missing.

   Corrected equation:

$$p(\lambda) = -Q(\Lambda + \lambda I)^{-1} Q^T g = -\sum_{j=1}^{n} \frac{q_j^T g}{\lambda_j + \lambda} q_j, \tag{4}$$

9. page 90, last line. $\Delta_k \geq \bar{\Delta}$ should be $\Delta_k \leq \bar{\Delta}$.

   Corrected sentence: We now derive a bound on the right-hand-side that holds for all sufficiently small values of Δ_k, that is, for all $\Delta_k \leq \bar{\Delta}$, where $\bar{\Delta}$ is to be determined.

10. page 91, lines 7-10. $\rho_k > \frac{1}{2}$ should be $\rho_k > \frac{3}{4}$, there is a missing parenthesis after $\frac{1}{4}$, and the $\leq$ in the displayed equation should be $\geq$.

    Corrected passage: Therefore, $\rho_k > \frac{3}{4}$, and so by the workings of Algorithm ??, we have $\Delta_{k+1} \geq \Delta_k$ whenever Δ_k falls below the threshold $\bar{\Delta}$. It follows that reduction of Δ_k (by a factor of $\frac{1}{4}$) can occur in our algorithm only if

$$\Delta_k \geq \bar{\Delta},$$

11. page 118. The last term of equation (5.37) should be $(C^{-T}b)^T\hat{x}$ not $(C^{-1}b)^T\hat{x}$. Likewise for the equation two lines later.

    Corrected lines: The quadratic ϕ defined by (??) is transformed accordingly to

$$\hat{\phi}(\hat{x}) = \tfrac{1}{2}\hat{x}^T(C^{-T}AC^{-1})\hat{x} - (C^{-T}b)^T\hat{x}. \tag{5}$$

If we use Algorithm ?? to minimize $\hat{\phi}$ or, equivalently to solve the linear system

$$(C^{-T}AC^{-1})\hat{x} = C^{-T}b,$$

12. page 166, line 2. Insert "a composition of" before the words "elementary arithmetic operations".

    Corrected sentence: This technique takes the view that the computer code for evaluating the function can be broken down into a composition of elementary arithmetic operations, to which the chain rule (one of the basic rules of calculus) can be applied.

13. page 171, line 15 and line 21. The reference to (7.18) should be replaced by a reference to (7.9).

    Corrected version of line 15: By substituting (**??**) into (**??**), we obtain

    Corrected version of lines 20-21: A similar argument shows that the nonzero elements of the fourth column of the Hessian can be estimated by substituting (**??**) into (**??**); we obtain eventually that

14. page 176, line 2 of Section 7.2. "exact" should be "analytic".

    Corrected sentence: Automatic differentiation is the generic name for techniques that use the computational representation of a function to produce analytic values for the derivatives.

15. page 176, second paragraph of Section 7.2. Delete the following sentence altogether: "Similarly, each line of a computer program usually contains just a few arithmetic operations."

16. page 180, line 15-16. Delete the following sentence altogether: "(When node j is *not* a child of node i, we have $\partial x_j/\partial x_i = 0$, and so the corresponding term need not appear in the summation.)"

17. page 180, lines -4 and -6 (that is, fourth and sixth lines from the bottom of the page). For consistency, replace these two instances of $\frac{\partial x_j}{\partial x_i}$ by $\partial x_j/\partial x_i$.

    Corrected sentences: During the forward sweep—the evaluation of f—we not only calculate the values of each variable x_i, but we also calculate and store the numerical values of each partial derivative $\partial x_j/\partial x_i$. Each of these partial derivatives is associated with a particular arc of the computational graph. The numerical values of $\partial x_j/\partial x_i$ computed during the forward sweep are then used in the formula (**??**) during the reverse sweep.

18. page 181, line 7. "Node 12" should be replaced by "Node 9".

    Corrected line: Node 9 is the child of nodes 3 and 8, so we use formula

19. page 208, line 8. We need $0 \le \phi_k \le 1$, not just $\phi_k \ge 0$.

    Corrected Sentence: Also, since BFGS and DFP updating preserve positive definiteness of the Hessian approximations when $s_k^T y_k > 0$, this relation implies that the same property will hold for the Broyden family if $0 \le \phi_k \le 1$.

20. page 220, line -3. It should read $\det(I + xy^T) = 1 + y^T x$.

    Corrected Sentence: First show that $\det(I + xy^T) = 1 + y^T x$, where x and y are n-vectors.

21. page 221. Part (c) of 8.9: The reference should be to (8.45) rather than (8.43).

    Corrected sentence: Now use this relation to establish (**??**).

22. page 316, line 25. In this displayed equation, the final part $k = 0, \pm 1, \pm 2, \ldots$ should be $k = \pm 1, \pm 3, \pm 5, \ldots$.

    Corrected display equation:

$$(x_1, x_2) = (k\pi, -1), \quad \text{for} \quad k = \pm 1, \pm 3, \pm 5, \ldots.$$

23. page 357. Delete the material on lines 6 through 29 inclusive, and replace with the following material:

    Inserted material:

    We return to our claim that the set N defined by (**??**) is a closed set. This fact is needed in the proof of Lemma **??** to ensure that the solution of the projection subproblem (**??**) is well-defined. Note that N has the following form:

$$N = \{s \,|\, s = A\lambda, \; C\lambda \geq 0\}, \tag{6}$$

    for appropriately defined matrices A and C.

    We outline a proof of closedness of N for the case in which LICQ holds (Definition **??**), that is, the matrix A has full column rank. It suffices to show that whenever $\{s_k\}$ is a sequence of vectors satisfying $s_k \to s^*$, then $s^* \in N$. Because A has full column rank, for each s_k there is a unique vector λ_k such that $s_k = A\lambda_k$; in fact, we have $\lambda_k = (A^T A)^{-1} A^T s_k$. Because $s_k \in N$, we must have $C\lambda_k \geq 0$, and since $s_k \to s^*$, we also have that

$$\lim_k \lambda_k = \lim_k (A^T A)^{-1} A^T s_k = (A^T A)^{-1} A^T s^* \stackrel{\text{def}}{=} \lambda^*.$$

    The inequality $C\lambda^* \geq 0$ is a consequence of $C\lambda_k \geq 0$ for all k, whereas $s^* = A\lambda^*$ follows from the facts that $s_k - A\lambda_k = 0$ for all k, and $s_k - A\lambda_k \to s^* - A\lambda^*$ Therefore we have identified a vector λ^* such that $s^* = A\lambda^*$ and $C\lambda^* \geq 0$, so $s^* \in N$, as claimed.

    There is a more general proof, using an argument due to Mangasarian and Schumaker [?] and appealing to Theorem **??** (iii), that does not require LICQ. We omit the details here.

24. page 370, lines 2-4. In this displayed equation, the expression $b^T \Delta\pi$ which appears in four places should be replaced by $\Delta b^T \pi$ in all cases.

    Corrected display equation:

$$\begin{aligned} c^T \Delta x &= (b + \Delta b)^T \Delta\pi + \Delta b^T \pi \\ &= (x + \Delta x)^T A^T \Delta\pi + \Delta b^T \pi \\ &= -(x + \Delta x)^T \Delta s + \Delta b^T \pi = \Delta b^T \pi. \end{aligned}$$

25. page 371, Theorem 13.2. Add a third item to this theorem: "(iii) If (13.1) is feasible and bounded, then it has an optimal solution."

    Corrected Theorem statement:

    (i) If there is a feasible point for (**??**), then there is a basic feasible point.

    (ii) If (**??**) has solutions, then at least one such solution is a basic optimal point.

    (iii) If (**??**) is feasible and bounded, then it has an optimal solution.

26. page 372, Insert a new paragraph just before the end of the proof to read as follows:

    Inserted paragraph: The final statement (iii) is a consequence of finite termination of the simplex method. We comment on the latter property in the next section.

27. page 372, first paragraph in the section headed "Vertices of the Feasible Polytope". Replace the passage "an intersection of half-spaces. The" by "and the".

    Corrected paragraph: The feasible set defined by the linear constraints is a polytope, and the *vertices* of this polytope are the points that do not lie on a straight line between two other points in the set. Geometrically, they are easily recognizable; see Figure **??**.

28. page 373, line 3. Replace "We" by "we", and insert the phrase "for $i = m+1, m+2, \ldots, n$" at the end of this sentence.

    Corrected sentence: Because of (**??**) and the fact that α and $1-\alpha$ are both positive, we must have $y_i = z_i = 0$ for $i = m+1, m+2, \ldots, n$.

29. page 374, first two paragraphs of Section 13.3. A new sentence needs to be added, and two phrases inserted. It is easiest to state the new version of these paragraphs, which is as follows:

    As we just described, all iterates of the simplex method are basic feasible points for (**??**) and therefore vertices of the feasible polytope. Most steps consist of a move from one vertex to an adjacent one for which the set of basic indices $\mathcal{B}(x)$ differs in exactly one component. On most steps (but not all), the value of the primal objective function $c^T x$ is decreased. Another type of step occurs when the problem is unbounded: The step is an edge along which the objective function is reduced, and along which we can move infinitely far without ever reaching a vertex.

    The major issue at each simplex iteration is to decide which index to change in the basis set $\mathcal{B}$. Unless the step is a direction of unboundedness, one index must be removed from $\mathcal{B}$ and replaced by another from outside $\mathcal{B}$. We can get some insight into how this decision is made by looking again at the KKT conditions (**??**) to see how they relate to the algorithm.

30. page 375, lines -5 and -6. In this bulleted item, replace "the component with index p" by "corresponding to x_p", and add the following phrase before the semicolon: "or determining that no such component exists (the unbounded case)"

Corrected item: keep increasing x_q until one of the components of x_{B} (corresponding to x_p, say) is driven to zero, or determining that no such component exists (the unbounded case);

31. page 377, statement of Theorem 13.4. After "nondegenerate", add the words "and bounded".

    Corrected statement: Provided that the linear program (**??**) is nondegenerate and bounded, the simplex method terminates at a basic optimal point.

32. page 378, add a sentence *after* the end of the proof, which happens at line 4: "Note that this result gives us a proof of Theorem 13.2 (iii) for the nondegenerate case."

    Inserted sentence: Note that this result gives us a proof of Theorem **??**(iii) for the nondegenerate case.

33. page 378, between lines 16 and 17: Add two extra lines (details below). On line 17, after "denote the index" add the phrase "of the basic variable", rearrange lines 17 and 18, and delete the parenthesized phrase "(p is the leaving index)".

    Corrected passage from the Procedure 13.1 (verbatim, should be ready to be copied into the latex source file):

```
\>\> Solve $Bt = A_q$ for $t$; \\
\>\> {\bf if} $t \le 0$ \\
\>\>\> {\bf STOP}; (* problem is unbounded *) \\
\>\> Calculate $x_q^+ = \min_{i \, | \, t_i>0} \, (\xB)_i / t_i$,
and use $p$ to denote the index of the basic \\
\>\>\> variable for which this minimum is achieved;\index{Simplex method!leaving ind
```

34. page 393, exercise 13.4. Reword this exercise by inserting the following phrase after the word "then": "the matrix B in (13.13) is singular, and therefore"

    Corrected exercise: Show that when $m \le n$ and the rows of A are linearly dependent in (**??**), then the matrix B in (**??**) is singular, and therefore there are no basic feasible points.

35. page 408, equation (4.28a). Add a minus sign before $S^{-1}Xr_c$.

    Corrected equation:

$$AD^2A^T\Delta\lambda = -r_b + A(-S^{-1}Xr_c + x - \sigma\mu S^{-1}e), \tag{7a}$$

$$\Delta s = -r_c - A^T\Delta\lambda, \tag{7b}$$

$$\Delta x = -x + \sigma\mu S^{-1}e - S^{-1}X\Delta s. \tag{7c}$$