Nonlinear Programming
3rd Edition

Theoretical Solutions Manual

Chapter 2

Dimitri P. Bertsekas

Massachusetts Institute of Technology

Athena Scientific, Belmont, Massachusetts

NOTE

This manual contains solutions of the theoretical problems, marked in the book by (www) It is continuously updated and improved, and it is 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 David Brown, Angelia Nedic, Asuman Ozdaglar, Cynara Wu.

Last Updated: May 2016

Solutions Chapter 2

SECTION 2.1

2.1.3 (www)

We have that

$$f(x^{k+1}) \le \max_i \left(1 + \lambda_i P^k(\lambda_i)\right)^2 f(x^0), \tag{1}$$

for any polynomial P^k of degree k and any k, where $\{\lambda_i\}$ is the set of the eigenvalues of Q. Chose P^k such that

$$1 + \lambda P^k(\lambda) = \frac{(z_1 - \lambda)}{z_1} \cdot \frac{(z_2 - \lambda)}{z_2} \cdots \frac{(z_k - \lambda)}{z_k}.$$

Define $I_j = [z_j - \delta_j, z_j + \delta_j]$ for $j = 1, \ldots, k$. Since $\lambda_i \in I_j$ for some j, we have

$$\left(1 + \lambda_i P^k(\lambda_i)\right)^2 \le \max_{\lambda \in I_j} \left(1 + \lambda P^k(\lambda)\right)^2.$$

Hence

$$\max_i \left(1 + \lambda_i P^k(\lambda_i)\right)^2 \le \max_{1 \le j \le k} \max_{\lambda \in I_j} \left(1 + \lambda P^k(\lambda)\right)^2. \tag{2}$$

For any j and $\lambda \in I_j$ we have

$$\begin{aligned}\left(1 + \lambda P^k(\lambda)\right)^2 &= \frac{(z_1 - \lambda)^2}{z_1^2} \cdot \frac{(z_2 - \lambda)^2}{z_2^2} \cdots \frac{(z_k - \lambda)^2}{z_k^2} \\ &\le \frac{(z_j + \delta_j - z_1)^2 (z_j + \delta_j - z_2)^2 \cdots (z_j + \delta_j - z_{j-1})^2 \delta_j^2}{z_1^2 \cdots z_j^2}.\end{aligned}$$

Here we used the fact that $\lambda \in I_j$ implies $\lambda < z_l$ for $l = j+1, \ldots, k$, and therefore $\frac{(z_l - \lambda)^2}{z_l^2} \le 1$ for all $l = j+1, \ldots, k$. Thus, from (2) we obtain

$$\max_i \left(1 + \lambda_i P^k(\lambda_i)\right)^2 \le R, \tag{3}$$

where

$$R = \left\{ \frac{\delta_1^2}{z_1^2}, \frac{\delta_2^2 (z_2 + \delta_2 - z_1)^2}{z_1^2 z_2^2}, \ldots, \frac{\delta_k^2 (z_k + \delta_k - z_1)^2 \cdots (z_k + \delta_k - z_{k-1})^2}{z_1^2 z_2^1 \cdots z_k^2} \right\}.$$

The desired estimate follows from (1) and (3).

2.1.4 (www)

It suffices to show that the subspace spanned by $g^0, g^1, \ldots, g^{k-1}$ is the same as the subspace spanned by $g^0, Qg^0, \ldots, Q^{k-1}g^0$, for $k = 1, \ldots, n$. We will prove this by induction. Clearly, for $k = 1$ the statement is true. Assume it is true for $k - 1 < n - 1$, i.e.

$$\text{span}\{g^0, g^1, \ldots, g^{k-1}\} = \text{span}\{g^0, Qg^0, \ldots, Q^{k-1}g^0\},$$

where $\text{span}\{v^0, \ldots, v^l\}$ denotes the subspace spanned by the vectors $v^0, \ldots, v^l$. Assume that $g^k \neq 0$ (i.e. $x^k \neq x^*$). Since $g^k = \nabla f(x^k)$ and x^k minimizes f over the manifold $x^0 + \text{span}\{g^0, g^1, \ldots, g^{k-1}\}$, from our assumption we have that

$$g^k = Qx^k - b = Q\left(x^0 + \sum_{i=0}^{k-1} \xi_i Q^i g^0\right) - b = Qx^0 - b + \sum_{i=0}^{k-1} \xi_i Q^{i+1} g^0.$$

The fact that $g^0 = Qx^0 - b$ yields

$$g^k = g^0 + \xi_0 Qg^0 + \xi_1 Q^2 g^0 + \ldots + \xi_{k-2} Q^{k-1} g^0 + \xi_{k-1} Q^k g^0. \tag{1}$$

If $\xi_{k-1} = 0$, then from (1) and the inductive hypothesis it follows that

$$g^k \in \text{span}\{g^0, g^1, \ldots, g^{k-1}\}. \tag{2}$$

We know that g^k is orthogonal to $g^0, \ldots, g^{k-1}$. Therefore (2) is possible only if $g^k = 0$ which contradicts our assumption. Hence, $\xi_{k-1} \neq 0$. If $Q^k g^0 \in \text{span}\{g^0, Qg^0, \ldots, Q^{k-1}g^0\}$, then (1) and our inductive hypothesis again imply (2) which is not possible. Thus the vectors $g^0, Qg^0, \ldots, Q^{k-1}g^0, Q^k g^0$ are linearly independent. This combined with (1) and linear independence of the vectors $g^0, \ldots, g^{k-1}, g^k$ implies that

$$\text{span}\{g^0, g^1, \ldots, g^{k-1}, g^k\} = \text{span}\{g^0, Qg^0, \ldots, Q^{k-1}g^0, Q^k g^0\},$$

which completes the proof.

2.1.5 (www)

Let x^k be the sequence generated by the conjugate gradient method, and let d^k be the sequence of the corresponding Q-conjugate directions. We know that x^{k+1} minimizes f over

$$x^0 + \text{span}\ \{d^0, d^1, \ldots, d^k\}.$$

Let $\tilde{x}^k$ be the sequence generated by the method described in the exercise. In particular, $\tilde{x}^1$ is generated from x^0 by steepest descent and line minimization, and for $k \geq 1$, $\tilde{x}^{k+1}$ minimizes f over the two-dimensional linear manifold

$$\tilde{x}^k + \text{span}\,\{\tilde{g}^k \text{ and } \tilde{x}^k - \tilde{x}^{k-1}\},$$

where $\tilde{g}^k = \nabla f(\tilde{x}^k)$. We will show by induction that $x^k = \tilde{x}^k$ for all $k \geq 1$.

Indeed, we have by construction $x^1 = \tilde{x}^1$. Suppose that $x^i = \tilde{x}^i$ for $i = 1, \ldots, k$. We will show that $x^{k+1} = \tilde{x}^{k+1}$. We have that $\tilde{g}^k$ is equal to $g^k = \beta^k d^{k-1} - d^k$ so it belongs to the subspace spanned by d^{k-1} and d^k. Also $\tilde{x}^k - \tilde{x}^{k-1}$ is equal to $x^k - x^{k-1} = \alpha^{k-1} d^{k-1}$. Thus

$$\text{span}\,\{\tilde{g}^k \text{ and } \tilde{x}^k - \tilde{x}^{k-1}\} = \text{span}\,\{d^{k-1} \text{ and } d^k\}.$$

Observe that x^k belongs to

$$x^0 + \text{span}\,\{d^0, d^1, \ldots, d^{k-1}\},$$

so

$$x^0 + \text{span}\,\{d^0, d^1, \ldots, d^{k-1}\} \supset x^k + \text{span}\,\{d^{k-1} \text{ and } d^k\} \supset x^k + \text{span}\,\{d^k\}.$$

The vector x^{k+1} minimizes f over the linear manifold on the left-hand side above, and also over the linear manifold on the right-hand side above (by the definition of a conjugate direction method). Moreover, $\tilde{x}^{k+1}$ minimizes f over the linear manifold in the middle above. Hence $x^{k+1} = \tilde{x}^{k+1}$.

2.1.6 (PARTAN) (www)

Suppose that $x^1, \ldots, x^k$ have been generated by the method of Exercise 1.6.5, which by the result of that exercise, is equivalent to the conjugate gradient method. Let y^k and x^{k+1} be generated by the two line searches given in the exercise.

By the definition of the congugate gradient method, x^k minimizes f over

$$x^0 + \text{span}\,\{g^0, g^1, \ldots, g^{k-1}\},$$

so that

$$g^k \perp \text{span}\,\{g^0, g^1, \ldots, g^{k-1}\},$$

and in particular

$$g^k \perp g^{k-1}. \qquad (1)$$

Also, since y^k is the vector that minimizes f over the line $y_\alpha = x^k - \alpha g^k$, $\alpha \geq 0$, we have

$$g^k \perp \nabla f(y^k). \tag{2}$$

Any vector on the line passing through x^{k-1} and y^k has the form

$$y = \alpha x^{k-1} + (1-\alpha)y^k, \qquad \alpha \in \Re,$$

and the gradient of f at such a vector has the form

$$\begin{aligned} \nabla f\big(\alpha x^{k-1} + (1-\alpha)y^k\big) &= Q\big(\alpha x^{k-1} + (1-\alpha)y^k\big) - b \\ &= \alpha(Qx^{k-1} - b) + (1-\alpha)(Qy^k - b) \\ &= \alpha g^{k-1} + (1-\alpha)\nabla f(y^k). \end{aligned} \tag{3}$$

From Eqs. (1)-(3), it follows that g^k is orthogonal to the gradient $\nabla f(y)$ of any vector y on the line passing through x^{k-1} and y^k.

In particular, for the vector x^{k+1} that minimizes f over this line, we have that $\nabla f(x^{k+1})$ is orthogonal to g^k. Furthermore, because x^{k+1} minimizes f over the line passing through x^{k-1} and y^k, $\nabla f(x^{k+1})$ is orthogonal to $y^k - x^{k-1}$. Thus, $\nabla f(x^{k+1})$ is orthogonal to

$$\text{span}\,\{g^k, y^k - x^{k-1}\},$$

and hence also to

$$\text{span}\,\{g^k, x^k - x^{k-1}\},$$

since x^{k-1}, x^k, and y^k form a triangle whose side connecting x^k and y^k is proportional to g^k. Thus x^{k+1} minimizes f over

$$x^k + \text{span}\,\{g^k, x^k - x^{k-1}\},$$

and it is equal to the one generated by the algorithm of Exercise 1.6.5.

2.1.7 (www)

The objective is to minimize over $\Re^n$, the positive semidefinite quadratic function

$$f(x) = \frac{1}{2}x'Qx + b'x.$$

The value of x^k following the kth iteration is

$$x^k = \arg\min\left\{ f(x) | x = x^0 + \sum_{i=1}^{k-1} \gamma^i d^i, \gamma^i \in \Re \right\} = \arg\min\left\{ f(x) | x = x^0 + \sum_{i=1}^{k-1} \delta^i g^i, \delta^i \in \Re \right\},$$

where d^i are the conjugate directions, and g^i are the gradient vectors. At the beginning of the $(k+1)$st iteration, there are two possibilities:

(1) $g^k = 0$: In this case, x^k is the global minimum since $f(x)$ is a convex function.

(2) $g^k \neq 0$: In this case, a new conjugate direction d^k is generated. Here, we also have two possibilities:

  (a) A minimum is attained along the direction d^k and defines x^{k+1}.

  (b) A minimum along the direction d^k does not exist. This occurs if there exists a direction d in the manifold spanned by $d^0, \ldots, d^k$ such that $d'Qd = 0$ and $b'd \neq 0$. The problem in this case has no solution.

If the problem has no solution (which occurs if there is some vector d such that $d'Qd = 0$ but $b'd \neq 0$), the algorithm will terminate because the line minimization problem along such a direction d is unbounded from below.

If the problem has infinitely many solutions (which will happen if there is some vector d such that $d'Qd = 0$ and $b'd = 0$), then the algorithm will proceed as if the matrix Q were positive definite, i.e. it will find one of the solutions (case 1 occurs).

However, in both situations the algorithm will terminate in at most m steps, where m is the rank of the matrix Q, because the manifold

$$\{x \in \Re^n | x = x^0 + \sum_{i=0}^{k-1} \gamma^i d^i, \gamma^i \in \Re\}$$

will not expand for $k > m$.

2.1.8 (www)

Let S_1 and S_2 be the subspaces with $S_1 \cap S_2$ being a proper subspace of $\Re^n$ (i.e. a subspace of $\Re^n$ other than $\{0\}$ and $\Re^n$ itself). Suppose that the subspace $S_1 \cap S_2$ is spanned by linearly independent vectors v_k, $k \in K \subseteq \{1, 2, \ldots, n\}$. Assume that x^1 and x^2 minimize the given quadratic function f over the manifolds M_1 and M_2 that are parallel to subspaces S_1 and S_2, respectively, i.e.

$$x^1 = \arg\min_{x \in M_1} f(x) \quad \text{and} \quad x^2 = \arg\min_{x \in M_2} f(x)$$

where $M_1 = y^1 + S_1$, $M_2 = y^2 + S_2$, with some vectors $y^1, y^2 \in \Re^n$. Assume also that $x^1 \neq x^2$. Without loss of generality we may assume that $f(x^2) > f(x^1)$. Since $x^2 \notin M_1$, the vectors $x^2 - x^1$

and $\{v_k \mid k \in K\}$ are linearly independent. From the definition of x^1 and x^2 we have that

$$\frac{d}{dt} f(x^1 + tv^k)\Big|_{t=0} = 0 \qquad \text{and} \qquad \frac{d}{dt} f(x^2 + tv^k)\Big|_{t=0} = 0,$$

for any v^k. When this is written out, we get

$$x^{1\prime} Qv^k - b'v^k = 0 \qquad \text{and} \qquad x^{2\prime} Qv^k - b'v^k = 0.$$

Subtraction of the above two equalities yields

$$(x^1 - x^2)' Qv^k = 0, \qquad \forall\, k \in K.$$

Hence, $x^1 - x^2$ is Q-conjugate to all vectors in the intersection $S_1 \cap S_2$. We can use this property to construct a conjugate direction method that does not evaluate gradients and uses only line minimizations in the following way.

Initialization: Choose any direction d^1 and points y^1 and z^1 such that $M_1^1 = y^1 + \text{span}\{d^1\}$, $M_2^1 = z^1 + \text{span}\{d^1\}$, $M_1^1 \neq M_2^1$. Let $d^2 = x_1^1 - x_1^2$, where $x_1^i = \arg\min_{x \in M_i^1} f(x)$ for $i = 1, 2$.

Generating new conjugate direction: Suppose that Q-conjugate directions $d^1, d^2, \ldots, d^k$, $k < n$ have been generated. Let $M_1^k = y^k + \text{span}\{d^1, \ldots d^k\}$ and $x_k^1 = \arg\min_{x \in M_1^k} f(x)$. If x_k^1 is not optimal there is a point z^k such that $f(z^k) < f(x_k^1)$. Starting from point z^k we again search in the directions $d^1, d^2, \ldots, d^k$ obtaining a point x_k^2 which minimizes f over the manifold M_2^k generated by z^k and $d^1, d^2, \ldots, d^k$. Since $f(x_k^2) \leq f(z^k)$, we have

$$f(x_k^2) < f(x_k^1).$$

As both x_k^1 and x_k^2 minimize f over the manifolds that are parallel to $\text{span}\{d^1, \ldots, d^k\}$, setting $d^{k+1} = x_k^2 - x_k^1$ we have that $d^1, \ldots, d^k, d^{k+1}$ are Q-conjugate directions (here we have used the established property).

In this procedure it is important to have a step which given a nonoptimal point x generates a point y for which $f(y) < f(x)$. If x is an optimal solution then the step must indicate this fact. Simply, the step must first determine whether x is optimal, and if x is not optimal, it must find a better point. A typical example of such a step is one iteration of the cyclic coordinate descent method, which avoids calculation of derivatives.

SECTION 2.2

2.2.1 (www)

The proof is by induction. Suppose the relation $D^k q^i = p^i$ holds for all k and $i \leq k-1$. The relation $D^{k+1} q^i = p^i$ also holds for $i = k$ because of the following calculation

$$D^{k+1} q^k = D^k q^k + \frac{y^k y^{k\prime} q^k}{q^{k\prime} y^k} = D^k q^k + y^k = D^k q^k + (p^k - D^k q^k) = p^k.$$

For $i < k$, we have, using the induction hypothesis $D^k q^i = p^i$,

$$D^{k+1} q^i = D^k q^i + \frac{y^k (p^k - D^k q^k)' q^i}{q^{k\prime} y^k} = p^i + \frac{y^k (p^{k\prime} q^i - q^{k\prime} p^i)}{q^{k\prime} y^k}.$$

Since $p^{k\prime} q^i = p^{k\prime} Q p^i = q^{k\prime} p^i$, the second term in the right-hand side vanishes and we have $D^{k+1} q^i = p^i$. This completes the proof.

To show that $(D^n)^{-1} = Q$, note that from the equation $D^{k+1} q^i = p^i$, we have

$$D^n = \begin{bmatrix} p^0 & \cdots & p^{n-1} \end{bmatrix} \begin{bmatrix} q^0 & \cdots & q^{n-1} \end{bmatrix}^{-1}, \qquad (*)$$

while from the equation $Qp^i = Q(x^{i+1} - x^i) = (Qx^{i+1} - b) - (Qx^i - b) = \nabla f(x^{i+1}) - \nabla f(x^i) = q^i$, we have

$$Q \begin{bmatrix} p^0 & \cdots & p^{n-1} \end{bmatrix} = \begin{bmatrix} q^0 & \cdots & q^{n-1} \end{bmatrix},$$

or equivalently

$$Q = \begin{bmatrix} q^0 & \cdots & q^{n-1} \end{bmatrix} \begin{bmatrix} p^0 & \cdots & p^{n-1} \end{bmatrix}^{-1}. \qquad (**)$$

(Note here that the matrix $\begin{bmatrix} p^0 & \cdots & p^{n-1} \end{bmatrix}$ is invertible, since both Q and $\begin{bmatrix} q^0 & \cdots & q^{n-1} \end{bmatrix}$ are invertible by assumption.) By comparing Eqs. (*) and (**), it follows that $(D^n)^{-1} = Q$.

2.2.2 (www)

For simplicity, we drop superscripts. The BFGS update is given by

$$\begin{aligned} \bar{D} &= D + \frac{pp'}{p'q} - \frac{Dqq'D}{q'Dq} + q'Dq \left(\frac{p}{p'q} - \frac{Dq}{q'Dq} \right) \left(\frac{p}{p'q} - \frac{Dq}{q'Dq} \right)' \\ &= D + \frac{pp'}{p'q} - \frac{Dqq'D}{q'Dq} + q'Dq \left(\frac{pp'}{(p'q)^2} - \frac{Dqp' + pq'D}{(p'q)(q'Dq)} + \frac{Dqq'D}{(q'Dq)^2} \right). \\ &= D + \left(1 + \frac{q'Dq}{p'q} \right) \frac{pp'}{p'q} - \frac{Dqp' + pq'D}{p'q} \end{aligned}$$

2.2.3 (www)

(a) For simplicity, we drop superscripts. Let $V = I - \rho q p'$, where $\rho = 1/(q'p)$. We have

$$\begin{aligned} V'DV + \rho pp' &= (I - \rho qp')' D (I - \rho qp') + \rho pp' \\ &= D - \rho(Dqp' + pq'D) + \rho^2 pq'Dqp' + \rho pp' \\ &= D - \frac{Dqp' + pq'D}{q'p} + \frac{(q'Dq)(pp')}{(q'p)^2} + \frac{pp'}{q'p} \\ &= D + \left(1 + \frac{q'Dq}{p'q}\right)\frac{pp'}{p'q} - \frac{Dqp' + pq'D}{p'q} \end{aligned}$$

and the result now follows using the alternative BFGS update formula of Exercise 1.7.2.

(b) We have, by using repeatedly the update formula for D of part (a),

$$\begin{aligned} D^k &= V^{k-1\prime} D^{k-1} V^{k-1} + \rho^{k-1} p^{k-1} p^{k-1\prime} \\ &= V^{k-1\prime} V^{k-2\prime} D^{k-2} V^{k-2} V^{k-1} + \rho^{k-2} V^{k-1\prime} p^{k-2} p^{k-2\prime} V^{k-1} + \rho^{k-1} p^{k-1} p^{k-1\prime}, \end{aligned}$$

and proceeding similarly,

$$\begin{aligned} D^k = {} & V^{k-1\prime} V^{k-2\prime} \cdots V^{0\prime} D^0 V^0 \cdots V^{k-2} V^{k-1} \\ & + \rho^0 V^{k-1\prime} \cdots V^{1\prime} p^0 p^{0\prime} V^1 \cdots V^{k-1} \\ & + \rho^1 V^{k-1\prime} \cdots V^{2\prime} p^1 p^{1\prime} V^2 \cdots V^{k-1} \\ & + \cdots \\ & + \rho^{k-2} V^{k-1\prime} p^{k-2} p^{k-2\prime} V^{k-1} \\ & + \rho^{k-1} p^{k-1} p^{k-1\prime} \end{aligned}$$

Thus to calculate the direction $-D^k \nabla f(x^k)$, we need only to store D^0 and the past vectors p^i, q^i, $i = 0, 1, \ldots, k-1$, and to perform the matrix-vector multiplications needed using the above formula for D^k. Note that multiplication of a matrix V^i or $V^{i\prime}$ with any vector is relatively simple. It requires only two vector operations: one inner product, and one vector addition.

2.2.4 (www)

Suppose that D is updated by the DFP formula and H is updated by the BFGS formula. Thus the update formulas are

$$\bar{D} = D + \frac{pp'}{p'q} - \frac{Dqq'D}{q'Dq},$$

$$\bar{H} = H + \left(1 + \frac{p'Hp}{q'p}\right)\frac{qq'}{q'p} - \frac{Hpq' + qp'H}{q'p}.$$

If we assume that HD is equal to the identity I, and form the product $\bar{H}\bar{D}$ using the above formulas, we can verify with a straightforward calculation that $\bar{H}\bar{D}$ is equal to I. Thus if the initial H and D are inverses of each other, the above updating formulas will generate (at each step) matrices that are inverses of each other.

2.2.5 (www)

(a) By pre- and postmultiplying the DFP update formula

$$\bar{D} = D + \frac{pp'}{p'q} - \frac{Dqq'D}{q'Dq},$$

with $Q^{1/2}$, we obtain

$$Q^{1/2}\bar{D}Q^{1/2} = Q^{1/2}DQ^{1/2} + \frac{Q^{1/2}pp'Q^{1/2}}{p'q} - \frac{Q^{1/2}Dqq'DQ^{1/2}}{q'Dq}.$$

Let

$$\bar{R} = Q^{1/2}\bar{D}Q^{1/2}, \qquad R = Q^{1/2}DQ^{1/2},$$

$$r = Q^{1/2}p, \qquad q = Qp = Q^{1/2}r.$$

Then the DFP formula is written as

$$\bar{R} = R + \frac{rr'}{r'r} - \frac{Rrr'R}{r'Rr}.$$

Consider the matrix

$$P = R - \frac{Rrr'R}{r'Rr}.$$

From the interlocking eigenvalues lemma, the eigenvalues $\mu_1, \ldots, \mu_n$ satisfy

$$\mu_1 \le \lambda_1 \le \mu_2 \le \cdots \le \mu_n \le \lambda_n,$$

where $\lambda_1, \ldots \lambda_n$ are the eigenvalues of R. We have $Pr = 0$, so 0 is an eigenvalue of P and r is a corresponding eigenvector. Hence, since $\lambda_1 > 0$, we have $\mu_1 = 0$. Consider the matrix

$$\bar{R} = P + \frac{rr'}{r'r}.$$

We have $\bar{R}r = r$, so 1 is an eigenvalue of $\bar{R}$. The other eigenvalues are the eigenvalues $\mu_2, \ldots, \mu_n$ of P, since their corresponding eigenvectors $e_2, \ldots, e_n$ are orthogonal to r, so that

$$\bar{R}e_i = Pe_i = \mu_i e_i, \qquad i = 2, \ldots, n.$$

(b) We have

$$\lambda_1 \le \frac{r'Rr}{r'r} \le \lambda_n,$$

so if we multiply the matrix R with $r'r/r'Rr$, its eigenvalue range shifts so that it contains 1. Since

$$\frac{r'r}{r'Rr} = \frac{p'Qp}{p'Q^{1/2}RQ^{1/2}p} = \frac{p'q}{q'Q^{-1/2}RQ^{-1/2}q} = \frac{p'q}{q'Dq},$$

multiplication of R by $r'r/r'Rr$ is equivalent to multiplication of D by $p'q/q'Dq$.

(c) In the case of the BFGS update

$$\bar{D} = D + \left(1 + \frac{q'Dq}{p'q}\right)\frac{pp'}{p'q} - \frac{Dqp' + pq'D}{p'q},$$

(cf. Exercise 1.7.2) we again pre- and postmultiply with $Q^{1/2}$. We obtain

$$\bar{R} = R + \left(1 + \frac{r'Rr}{r'r}\right)\frac{rr'}{r'r} - \frac{Rrr' + rr'R}{r'r},$$

and an analysis similar to the ones in parts (a) and (b) goes through.

2.2.6 (www)

(a) We use induction. Assume that the method coincides with the conjugate gradient method up to iteration k. For simplicity, denote for all k,

$$g^k = \nabla f(x^k).$$

We have, using the facts $p^{k'}g^{k+1} = 0$ and $p^k = \alpha^k d^k$,

$$\begin{aligned} d^{k+1} &= -D^{k+1}g^{k+1} \\ &= -\left(I + \left(1 + \frac{q^{k'}q^k}{p^{k'}q^k}\right)\frac{p^kp^{k'}}{p^{k'}q^k} - \frac{q^kp^{k'} + p^kq^{k'}}{p^{k'}q^k}\right)g^{k+1} \\ &= -g^{k+1} + \frac{p^kq^{k'}g^{k+1}}{p^{k'}q^k} \\ &= -g^{k+1} + \frac{(g^{k+1} - g^k)'g^{k+1}}{d^{k'}q^k}d^k. \end{aligned}$$

The argument given at the end of the proof of Prop. 1.6.1 shows that this formula is the same as the conjugate gradient formula.

(b) Use a scaling argument, whereby we work in the transformed coordinate system $y = D^{-1/2}x$, where the matrix D becomes the identity.