Nonlinear Programming
3rd Edition

Theoretical Solutions Manual

Chapter 5

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 5

SECTION 5.2

5.2.4 (www)

We have

$$\begin{aligned} &\text{minimize} \quad f(x) = \frac{1}{2}x'Qx \\ &\text{subject to } Ax = b. \end{aligned}$$

Since x^* is an optimal solution of this problem with associated Lagrange multiplier λ^*, we have

$$Ax^* = b \qquad \text{and} \qquad Qx^* + A'\lambda^* = 0. \tag{1}$$

We also have

$$q_c(\lambda) = \min L_c(x, \lambda),$$

where

$$L_c(x, \lambda) = \frac{1}{2}x'Qx + \lambda'(Ax - b) + \frac{c}{2}||Ax - b||^2.$$

One way of showing that $q_c(\lambda)$ has the given form is to view $q_c(\lambda)$ as the dual of the penalized problem:

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2}x'Qx + \frac{c}{2}\|Ax - b\|^2 \\ &\text{subject to } Ax = b, \end{aligned}$$

which is a quadratic programming problem. Note that x^* is also a solution of this problem, so that the optimal value of the problem is f^*. Furthermore, by expanding the term $||Ax - b||^2$, the preceding problem is equivalent to

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2}x'(Q + cA'A)x + cb'Ax + \frac{1}{2}cb'b \\ &\text{subject to } Ax = b. \end{aligned}$$

Because x^* is the unique solution of the original problem, Q must be positive definite over the null space of A

$$y'Qy > 0, \qquad \forall\ y \neq 0,\ Ay = 0.$$

Then, similar to the proof of Lemma 5.2.1, it can be seen that there exists some positive scalar $\bar{c}$ such that $Q + cA'A$ is positive definite for all $c \geq \bar{c}$, i.e.,

$$Q + cA'A > 0, \qquad \forall\, c \geq \bar{c}. \tag{2}$$

(this can be shown similar to the proof of Lemma 5.2.1). By duality theory, there is no duality gap for the preceding problem [$q_c(\lambda^*) = f^*$], and according to Example 5.4.3 from Section 5.4, the function $q_c(\lambda)$ is quadratic in λ, so that the second order Taylor's expansion is exact for all λ, i.e.,

$$q_c(\lambda) = f^* + \nabla q_c(\lambda^*)'(\lambda - \lambda^*) + \frac{1}{2}(\lambda - \lambda^*)'\nabla^2 q_c(\lambda^*)'(\lambda - \lambda^*), \qquad \forall\, \lambda \in \Re^m. \tag{3}$$

We now need to calculate $\nabla q_c(\lambda^*)$ and $\nabla^2 q_c(\lambda^*)$. We have

$$\nabla q_c(\lambda) = h\big(x(\lambda, c)\big)$$

$$\nabla^2 q_c(\lambda) = -\nabla h\big(x(\lambda, c)\big)'\Big\{\nabla^2_{xx} L_c\big(x(\lambda, c), \lambda\big)\Big\}^{-1}\nabla h\big(x(\lambda, c)\big),$$

where $x(\lambda, c)$ minimizes $L_c(x, \lambda)$. To find $x(\lambda, c)$, we can solve $\nabla L_c(x, \lambda) = 0$, which yields

$$Qx + A'\lambda + cA'(Ax - b) = 0 \Leftrightarrow (Q + cA'A)x = cA'b - A'\lambda,$$

so that

$$x(\lambda, c) = (Q + cA'A)^{-1}(cA'b - A'\lambda), \qquad \forall\, c \geq \bar{c}$$

[$(Q + cA'A)^{-1}$ exists as implied by Eq. (2)]. Therefore

$$\nabla q_c(\lambda) = h\big(x(\lambda, c)\big) = A(Q + cA'A)^{-1}(cA'b - A'\lambda) - b, \qquad \forall\, c \geq \bar{c}, \tag{4}$$

from which by using Eq. (1), it can be seen that

$$\nabla q_c(\lambda^*) = 0. \tag{5}$$

Moreover, we have

$$\nabla^2 q_c(\lambda) = -A(Q + cA'A)^{-1}A', \qquad \forall\, \lambda \in \Re^m, \tag{6}$$

so that by using the preceding two relations in Eq. (3), we obtain

$$q_c(\lambda) = f^* - \frac{1}{2}(\lambda - \lambda^*)'A(Q + cA'A)^{-1}A'(\lambda - \lambda^*), \qquad \forall\, \lambda \in \Re^m, \quad \forall\, c \geq \bar{c}.$$

(a) We have

$$\lambda^{k+1} = \lambda^k + c^k \nabla q_{c^k}(\lambda^k),$$

so that

$$\lambda^{k+1} - \lambda^* = \lambda^k - \lambda^* + c^k \nabla q_{c^k}(\lambda^k).$$

We now express $\nabla q_{c^k}(\lambda^k)$ in an equivalent form. In what follows, we assume that $c^k \geq \bar{c}$ for all k, so that $\nabla q_{c^k}(\lambda)$ is linear for all k [cf. Eq. (4)]. By using the first order Taylor's expansion, we obtain

$$\nabla q_c(\lambda) = \nabla q_c(\lambda^*) + \nabla^2 q_c(\lambda^*)'(\lambda - \lambda^*), \qquad \forall\ \lambda \in \Re^m,$$

and by using Eqs. (5) and (6), we have

$$\nabla q_c(\lambda) = -A(Q + cA'A)^{-1}A'(\lambda - \lambda^*), \qquad \forall\ \lambda \in \Re^m,$$

Therefore

$$\begin{aligned} \lambda^{k+1} - \lambda^* &= \lambda^k - \lambda^* - c^k A(Q + c^k A'A)^{-1}A'(\lambda^k - \lambda^*) \\ &= \big(I - c^k A(Q + c^k A'A)^{-1}A'\big)(\lambda^k - \lambda^*), \end{aligned}$$

and by applying the results of Section 1.3, we obtain

$$\|\lambda^{k+1} - \lambda^*\| \leq r^k ||\lambda^k - \lambda^*||,$$

where

$$r^k = \max\big\{|1 - c^k E_{c^k}|, |1 - c^k e_{c^k}|\big\},$$

and E_c and e_c are the maximum and minimum eigenvalues of $A(Q + cA'A)^{-1}A'$.

(b) The matrix identity of Appendix A

$$(A + CBC')^{-1} = A^{-1} - A^{-1}C(B^{-1} + C'A^{-1}C)^{-1}C'A^{-1}$$

applied to $(Q + c_k A'A)^{-1}$ yields

$$(Q + c_k A'A)^{-1} = Q^{-1} - Q^{-1}A'\left(\frac{1}{c_k}I + AQ^{-1}A'\right)^{-1} AQ^{-1}$$

and so

$$A(Q + c_k A'A)^{-1}A' = AQ^{-1}A' - AQ^{-1}A'\left(\frac{1}{c_k}I + AQ^{-1}A'\right)^{-1} AQ^{-1}A'.$$

Let γ be an eigenvalue of $(AQ^{-1}A')^{-1}$. Using the facts that

$$\lambda = \{\text{eigenvalue of } A\} \Leftrightarrow \frac{1}{\lambda} = \{\text{eigenvalue of } A^{-1}\},$$

$$\lambda = \{\text{eigenvalue of } A\} \Leftrightarrow \lambda + c = \{\text{eigenvalue of } cI + A\},$$

we can see that

$$\frac{1}{\gamma} - \frac{1}{\gamma}\left(\frac{1}{c} + \frac{1}{\gamma}\right)^{-1}\frac{1}{\gamma} = \frac{1}{c + \gamma}$$

is an eigenvalue of

$$A(Q + cAA')^{-1}A'.$$

Thus

$$r^k = \max_{1\le i\le m} \left\{ \left| 1 - \frac{c^k}{\gamma_i + c^k} \right| \right\}.$$

(c) First, for the method to be defined we need $c^k \geq \bar{c}$ for all k sufficiently large. Second, for the method to converge, we need $r^k < 1$ for all k sufficiently large. Thus

$$\left| 1 - \frac{c}{\gamma_i + c} \right| < 1, \quad \forall\ i,$$

which is equivalent to

$$-2 < -\frac{c}{\gamma_i + c} < 0 \quad \text{or} \quad 0 < \frac{c}{\gamma_i + c} < 2.$$

Since $c > 0$, we must have $\gamma_i + c > 0$. Then solving the above inequality yields the threshold value

$$\hat{c} = \max\left\{ 0, \max_{1\le i\le m} \{-2\gamma_i\} \right\}.$$

Hence, the overall threshold value is

$$c = \max\{\bar{c}, \hat{c}\}.$$

5.2.5 (www)

Using the results of Exercise 5.2.4, updating the multipliers with

$$\lambda^{k+1} = \lambda^k + \alpha^k (Ax^k - b)$$

implies

$$\|\lambda^{k+1} - \lambda^*\| \leq \max_i \left\{ \left| 1 - \frac{\alpha^k}{\gamma_i + c^k} \right| \right\} \|\lambda^k - \lambda^*\|.$$

For the method to converge, we need for $k > \bar{k}$,

$$\left| 1 - \frac{\alpha^k}{\gamma_i + c^k} \right| \leq 1 - \epsilon, \quad \forall\ i,$$

or

$$\epsilon \leq \frac{\alpha^k}{\gamma_i + c^k} \leq 2 - \epsilon \tag{1}$$

for some $\epsilon > 0$. If Q is positive definite and $c^k = c$ for all k, we have $\gamma_i > 0$ for all i, and if $\delta \leq \alpha^k \leq 2c$, the condition (1) is satisfied for $\epsilon \leq \min\{\delta, 2\gamma_i\}/(c + \gamma_i)$ for all i.

5.2.9 (www)

In the logarithmic barrier method we have

$$x^k = \arg\min_{x \in S} \{ f(x) + \epsilon^k B(x) \},$$

where $S = \{x \in X \mid g_j(x) < 0,\ j = 1, \ldots, r\}$ and $B(x) = -\sum_{j=1}^r \ln\big(-g_j(x)\big)$. Assuming that f and g_j are continuously differentiable, x^k satisfies

$$\nabla f(x^k) + \epsilon^k \nabla B(x^k) = 0$$

or equivalently

$$\nabla f(x^k) - \sum_{j=1}^r \frac{\epsilon^k}{g_j(x^k)} \nabla g_j(x^k) = 0.$$

Define $\mu_j^k = -\frac{\epsilon^k}{g_j(x^k)}$ for all j and k. Then we have

$$\mu_j^k > 0, \qquad \forall\ j = 1, \ldots, r, \quad \forall k, \tag{1}$$

$$\nabla f(x^k) + \sum_{j=1}^r \mu_j^k \nabla g_j(x^k) = 0, \quad \forall\ k. \tag{2}$$

Suppose that x^* is a limit point of the sequence $\{x^k\}$. Let $\{x^k\}_{k \in \mathcal{K}}$ be a subsequence of $\{x^k\}$ converging to x^*, and let $A(x^*)$ be the index set of active constraints at x^*. Furthermore, for any x, let $\nabla g_A(x)$ be a matrix with columns $\nabla g_j(x)$ for $j \in A(x^*)$ and $\nabla g_R(x)$ be a matrix with columns $\nabla g_j(x)$ for $j \notin A(x^*)$. Similarly, we partition a vector μ: μ_A is a vector with coordinates μ_j for $j \in A(x^*)$ and μ_R is a vector with coordinates μ_j for $j \notin A(x^*)$. Then Eq. (2) is equivalent to

$$\nabla f(x^k) + \nabla g_A(x^k)\mu_A^k + \nabla g_R(x^k)\mu_R^k = 0, \qquad \forall\ k. \tag{3}$$

If $j \notin A(x^*)$, then $g_j(x^k) < -\delta$ for some positive scalar δ and for all large enough $k \in \mathcal{K}$, which guarantees the boundedness of the sequence $\{-1/g_j(x^k)\}_{\mathcal{K}}$. Since $\epsilon^k \to 0$, we have

$$\lim_{k \to \infty,\ k \in \mathcal{K}} \mu_j^k = -\lim_{k \to \infty,\ k \in \mathcal{K}} \frac{\epsilon^k}{g_j(x^k)} = 0, \qquad \forall\ j \notin A(x^*),$$

i.e., $\{\mu_R^k \to 0\}_{\mathcal{K}}$. Therefore, by continuity of ∇g_j, we have

$$\lim_{k \to \infty,\ k \in \mathcal{K}} \nabla g_R(x^k)\mu_R^k = 0. \tag{4}$$

Suppose now that x^* is a regular point, i.e., the gradients $\nabla g_j(x^*)$ for $j \in A(x^*)$ are linearly independent, so that the matrix $\nabla g_A(x^*)'\nabla g_A(x^*)$ is invertible. Then, by continuity of ∇g_j, the matrix $\nabla g_A(x^k)'\nabla g_A(x^k)$ is invertible for all sufficiently large $k \in \mathcal{K}$. Premultiplying Eq. (3) by $\big(\nabla g_A(x^k)'\nabla g_A(x^k)\big)^{-1}\nabla g_A(x^k)'$ gives

$$\mu_A^k = -\big(\nabla g_A(x^k)'\nabla g_A(x^k)\big)^{-1}\nabla g_A(x^k)'\big(\nabla f(x^k) + \nabla g_R(x^k)\mu_R^k\big).$$

By letting $k \to \infty$ over $k \in \mathcal{K}$, and by using the continuity of ∇f and ∇g_j and the relation (4), we obtain

$$\lim_{k\to\infty,\ k\in\mathcal{K}} \mu_A^k = -\big(\nabla g_A(x^*)'\nabla g_A(x^*)\big)^{-1}\nabla g_A(x^*)'\nabla f(x^*).$$

Define μ^* by $\mu_R^* = 0$ and

$$\mu_A^* = \lim_{k\to\infty,\ k\in\mathcal{K}} \mu_A^k,$$

so that by letting $k \to \infty$ with $k \in \mathcal{K}$, from Eq. (3) we have

$$\nabla f(x^*) + \nabla g_A(x^*)\mu_A^* + \nabla g_R(x^*)\mu_R^* = \nabla f(x^*) + \nabla g(x^*)\mu^* = 0.$$

In view of Eq. (1), μ^* must be nonnegative, so that μ^* is a Lagrange multiplier. Furthermore, assuming that x^* is a limit point of the sequence $\{x^k\}$, the regularity of x^* is sufficient to ensure the convergence of $\{\mu_j^k\}$ to corresponding Lagrange multipliers.

By Prop. 5.1.1, every limit point of $\{x^k\}$ is a global minimum of the original problem. Hence, for the convergence of $\{\mu_j^k\}$ to corresponding Lagrange multipliers, it is sufficient that every global minimum of the original problem is regular.

5.2.11 (www)

Consider first the case where f is quadratic, $f(x) = \frac{1}{2}x'Qx$ with Q positive definite and symmetric, and h is linear, $h(x) = Ax - b$, with A having full rank. Following the hint, the iteration $\lambda^{k+1} = \lambda^k + \alpha h(x^k)$ can be viewed as the method of multipliers for the problem

$$\begin{aligned} &\text{minimize} \ \ \tfrac{1}{2}x'Qx - \frac{\alpha}{2}\|Ax - b\|^2 \\ &\text{subject to} \ \ Ax - b = 0. \end{aligned}$$

According to Exercise 5.2.4(c), this method converges if $\alpha > \overline{\alpha}$, where the threshold value $\overline{\alpha}$ is

$$\overline{\alpha} = 0 \qquad \text{if} \qquad \overline{\zeta} \geq 0, \tag{1}$$

$$\overline{\alpha} = -2\zeta \qquad \text{if} \qquad \overline{\zeta} < 0, \tag{2}$$

where $\overline{\zeta}$ is the minimum eigenvalue of the matrix

$$\big(A(Q - \alpha A'A)^{-1}A'\big)^{-1}.$$

To calculate $\overline{\zeta}$, we use the matrix identity

$$\alpha A(Q - \alpha A'A)^{-1}A' = (I - \alpha AQ^{-1}A')^{-1} - I$$

of Section A.3 in Appendix A. If $\zeta_1, \ldots, \zeta_m$ are the eigenvalues of $\big(A(Q - \alpha A'A)^{-1}A'\big)^{-1}$, we have

$$\frac{\alpha}{\zeta_i} = \frac{1}{1 - \alpha \xi_i^{-1}} - 1.$$

where ξ_i are the eigenvalues of $(AQ^{-1}A')^{-1}$. This equation can be written as

$$\frac{\alpha}{\zeta_i} = \frac{\alpha}{\xi_i - \alpha},$$

from which

$$\zeta_i = \xi_i - \alpha.$$

Let $\overline{\xi} = \min\{\xi_1, \ldots, \xi_m\}$. Then the condition (1) is written as

$$0 < \alpha \le \overline{\xi}. \qquad (3)$$

The condition (2) is written as

$$\alpha > 2(\alpha - \overline{\xi}) \qquad \text{with} \qquad \alpha > \overline{\xi},$$

or

$$\overline{\xi} < \alpha < 2\overline{\xi}. \qquad (4)$$

Convergence is obtained under either condition (3) or (4), so we see that convergence is obtained for

$$0 < \alpha < 2\overline{\xi}.$$

In the case where f is nonquadratic and/or h is nonlinear, a local version of the above analysis applies.