Problem correction: Assume that $Q$ is symmetric and invertible. (This correction has been made in the 2nd printing.)

Solution: We have

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

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

$$Ax^* = b \quad \text{and} \quad Qx^* + A'\lambda^* = 0. \quad (1)$$

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:

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

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

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

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

$$y'Qy > 0, \quad \forall \ y \neq 0, \ A y = 0.$$
Then, similar to the proof of Lemma 3.2.1, 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.,

\[
Q + cA'A > 0, \quad \forall \ c \geq \bar{c}.
\]

(2)

This can be shown similar to the proof of Lemma 3.2.1, pg. 298]. By duality theory, there is no duality gap for the preceding problem \( [q_c(\lambda^*) = f^*] \), and according to Example 3.4.3 from Section 3.4, the function \( q_c(\lambda) \) is quadratic in \( \lambda \), so that the second order Taylor’s expansion is exact for all \( \lambda \), i.e.,

\[
q_c(\lambda) = f^* + \nabla q_c(\lambda^*)' (\lambda - \lambda^*) + \frac{1}{2} (\lambda - \lambda^*)' \nabla^2 q_c(\lambda^*) (\lambda - \lambda^*), \quad \forall \ \lambda \in \mathbb{R}^m.
\]

(3)

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

\[
\nabla q_c(\lambda) = h(x(\lambda, c))
\]

\[
\nabla^2 q_c(\lambda) = -\nabla h(x(\lambda, c))' \left\{ \nabla^2 L_c(x(\lambda, c), \lambda) \right\}^{-1} \nabla h(x(\lambda, c)),
\]

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

\[
Qx + A'\lambda + cA'(Ax - b) = 0 \iff (Q + cA'A)x = cA'b - A'\lambda,
\]

so that

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

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

\[
\nabla q_c(\lambda) = h(x(\lambda, c)) = A(Q + cA'A)^{-1}(cA'b - A'\lambda) - b, \quad \forall \ c \geq \bar{c},
\]

(4)

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

\[
\nabla q_c(\lambda^*) = 0.
\]

(5)

Moreover, we have

\[
\nabla^2 q_c(\lambda) = -A(Q + cA'A)^{-1} A', \quad \forall \ \lambda \in \mathbb{R}^m,
\]

(6)

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

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

(a) We have

\[
\lambda^{k+1} = \lambda^k + c^k \nabla q_c(\lambda^k),
\]
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
\[
\nabla q_{c,k}(\lambda) = \nabla q_{c}(\lambda^*) + \nabla^2 q_{c}(\lambda^*)(\lambda - \lambda^*), \quad \forall \lambda \in \mathbb{R}^m,
\]
and by using Eqs. (5) and (6), we have
\[
\nabla q_{c}(\lambda) = -A(Q + cA'A)^{-1}A'\lambda - \lambda^*), \quad \forall \lambda \in \mathbb{R}^m,
\]
Therefore
\[
\lambda^{k+1} - \lambda^* = \lambda^k - \lambda^* - c^k A(Q + c^k A'A)^{-1}A'\lambda - \lambda^*\right)
\]
\[
= (I - c^k A(Q + c^k A'A)^{-1}A')(\lambda^k - \lambda^*),
\]
and by applying the results of Section 1.3, we obtain
\[
\|\lambda^{k+1} - \lambda^*\| \leq r^k \|\lambda^k - \lambda^*\|,
\]
where
\[
r^k = \max\{||1 - c^k E_{c,k}||, |1 - c^k e_{c,k}||\},
\]
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 \(\gamma\) 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

\[ r^k = \max_{1 \leq i \leq m} \left\{ 1 - \frac{c^k}{\gamma_i + c^k} \right\}. \]

(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

\[ \hat{c} = \max \left\{ 0, \max_{1 \leq i \leq m} \{-2\gamma_i\} \right\}. \]

Hence, the overall threshold value is

\[ c = \max\{\bar{c}, \hat{c}\}. \]

4.2.5 Using the results of Exercise 4.2.4, updating the multipliers with

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

implies

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

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

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

or

\[ \epsilon \leq \frac{\alpha^k}{\gamma_i + c^k} \leq 2 - \epsilon \quad (1) \]

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 \( i \).
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 \( B(x) = -\sum_{j=1}^{r} \ln(-g_j(x)) \). 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

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

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

\[ \mu_j^k > 0, \quad \forall \ j = 1, \ldots, r, \ \forall k, \quad (1) \]

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

Suppose that \( x^* \) is a limit point of the sequence \( \{ x^k \} \). Let \( \{ x^k \}_{k \in K} \) 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 \( \mu \): \( \mu_A \) is a vector with coordinates \( \mu_j \) for \( j \in A(x^*) \) and \( \mu_R \) is a vector with coordinates \( \mu_j \) for \( j \notin A(x^*) \). Then Eq. (2) is equivalent to

\[ \nabla f(x^k) + \nabla g_A(x^k) \mu_A^k + \nabla g_R(x^k) \mu_R^k = 0, \quad \forall k. \quad (3) \]

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

\[ \lim_{k \to \infty, k \in K} \mu_j^k = -\lim_{k \to \infty, k \in K} \frac{\epsilon_k}{g_j(x^k)} = 0, \quad \forall j \notin A(x^*), \]

i.e., \( \{ \mu_R^k \to 0 \}_{K} \). Therefore, by continuity of \( \nabla g_j \), we have

\[ \lim_{k \to \infty, k \in K} \nabla g_R(x^k) \mu_R^k = 0. \quad (4) \]

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 \( \nabla g_A(x^*) \nabla g_A(x^*)^T \) is invertible. Then, by continuity of \( \nabla g_j \), the
matrix $\nabla g_A(x^k)'\nabla g_A(x^k)$ is invertible for all sufficiently large $k \in K$. Premultiplying Eq. (3) by $(\nabla g_A(x^k)'\nabla g_A(x^k))^{-1}\nabla g_A(x^k)'$ gives

$$\mu^k_A = - (\nabla g_A(x^k)'\nabla g_A(x^k))^{-1}\nabla g_A(x^k)'(\nabla f(x^k) + \nabla g_R(x^k)\mu^k_R).$$

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

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

Define $\mu^*$ by $\mu^*_R = 0$ and

$$\mu^*_A = \lim_{k \to \infty, k \in K} \mu^k_A,$$

so that by letting $k \to \infty$ with $k \in K$, 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), $\mu^*$ must be nonnegative, so that $\mu^*$ 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 $\{\mu^k\}$ to corresponding Lagrange multipliers.

By Prop. 4.1.1, every limit point of $\{x^k\}$ is a global minimum of the original problem. Hence, for the convergence of $\{\mu^k\}$ to corresponding Lagrange multipliers, it is sufficient that every global minimum of the original problem is regular.

4.2.11 (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

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

According to Exercise 4.2.4(c), this method converges if $\alpha > \overline{\alpha}$, where the threshold value $\overline{\alpha}$ is

$$\overline{\alpha} = \begin{cases} 0 & \text{if } \overline{\zeta} \geq 0, \\ -2\overline{\zeta} & \text{if } \overline{\zeta} < 0, \end{cases}$$

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

$$(A(Q - \alpha A'A)^{-1}A')^{-1}.$$
To calculate $\zeta$, we use the matrix identity

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

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

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

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

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

from which

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

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

$$0 < \alpha \leq \bar{\xi}. \quad (3)$$

The condition (2) is written as

$$\alpha > 2(\alpha - \bar{\xi}) \quad \text{with} \quad \alpha > \bar{\xi},$$

or

$$\bar{\xi} < \alpha < 2\bar{\xi}. \quad (4)$$

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

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

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