Solutions to Selected Problems in

NUMERICAL OPTIMIZATION

by J. Nocedal and S.J. Wright

Second Edition

Solution Manual Prepared by:

Frank Curtis
Long Hei
Gabriel López-Calva
Jorge Nocedal
Stephen J. Wright

Contents

1 Introduction

No exercises assigned.

2 Fundamentals of Unconstrained Optimization

Problem 2.1

$$
\begin{aligned}
\frac{\partial f}{\partial x_1} &= 100 \cdot 2(x_2 - x_1^2)(-2x_1) + 2(1 - x_1)(-1) \\
&= -400x_1(x_2 - x_1^2) - 2(1 - x_1) \\
\frac{\partial f}{\partial x_2} &= 200(x_2 - x_1^2) \\
\implies \nabla f(x) &= \begin{bmatrix} -400x_1(x_2 - x_1^2) - 2(1 - x_1) \\ 200(x_2 - x_1^2) \end{bmatrix} \\
\frac{\partial^2 f}{\partial x_1^2} &= -400[x_1(-2x_1) + (x_2 - x_1^2)(1)] + 2 = -400(x_2 - 3x_1^2) + 2 \\
\frac{\partial^2 f}{\partial x_2 \partial x_1} &= \frac{\partial^2 f}{\partial x_1 \partial x_2} = -400x_1 \\
\frac{\partial^2 f}{\partial x_2^2} &= 200 \\
\implies \nabla^2 f(x) &= \begin{bmatrix} -400(x_2 - 3x_1^2) + 2 & -400x_1 \\ -400x_1 & 200 \end{bmatrix}
\end{aligned}
$$

1. $\nabla f(x^*) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $x^* = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is the only solution to $\nabla f(x) = 0$

2. $\nabla^2 f(x^*) = \begin{bmatrix} 802 & -400 \\ -400 & 200 \end{bmatrix}$ is positive definite since $802 > 0,$ and $\det(\nabla^2 f(x^*)) = 802(200) - 400(400) > 0.$

3. $\nabla f(x)$ is continuous.

(1), (2), (3) imply that x^* is the only strict local minimizer of $f(x)$.

□

Problem 2.2

$$
\begin{aligned}
\frac{\partial f}{\partial x_1} &= 8 + 2x_1 \\
\frac{\partial f}{\partial x_2} &= 12 - 4x_2 \\
\Longrightarrow \nabla f(x) &= \begin{bmatrix} 8+2x_1 \\ 12-4x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\
& \text{One solution is} \quad x^* = \begin{pmatrix} -4 \\ 3 \end{pmatrix}. \\
& \text{This is the only point satisfying the first order necessary conditions.}
\end{aligned}
$$

$$
\nabla^2 f(x) = \begin{bmatrix} 2 & 0 \\ 0 & -4 \end{bmatrix} \quad \text{is not positive definite, since } \det(\nabla^2 f(x)) = -8 < 0.
$$

Therefore, x^* is **NOT** a minimizer. Consider $\min(-f(x))$. It is seen that $\nabla^2[-f(x)]$ is also not positive definite. Therefore x^* is **NOT** a maximizer. Thus x^* is a saddle point and only a stationary point.

$\square$

The contour lines of $f(x)$ are shown in Figure 1.

Problem 2.3

(1)

$$
\begin{aligned}
f_1(x) &= a^T x \\
&= \sum_{i=1}^{n} a_i x_i
\end{aligned}
$$

$$
\begin{aligned}
\nabla f_1(x) &= \begin{bmatrix} \frac{\partial f_1}{\partial x_1} \\ \cdots \\ \frac{\partial f_1}{\partial x_n} \end{bmatrix} = \begin{bmatrix} a_1 \\ \cdots \\ a_n \end{bmatrix} = a \\
\nabla^2 f_1(x) &= \begin{bmatrix} \frac{\partial^2 f_1}{\partial x_1^2} & \frac{\partial^2 f_1}{\partial x_2 \partial x_1} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} = \left[\frac{\partial^2 \sum_i a_i x_i}{\partial x_s \partial x_t} \right]_{\substack{s=1\cdots n \\ t=1\cdots n}} = 0
\end{aligned}
$$

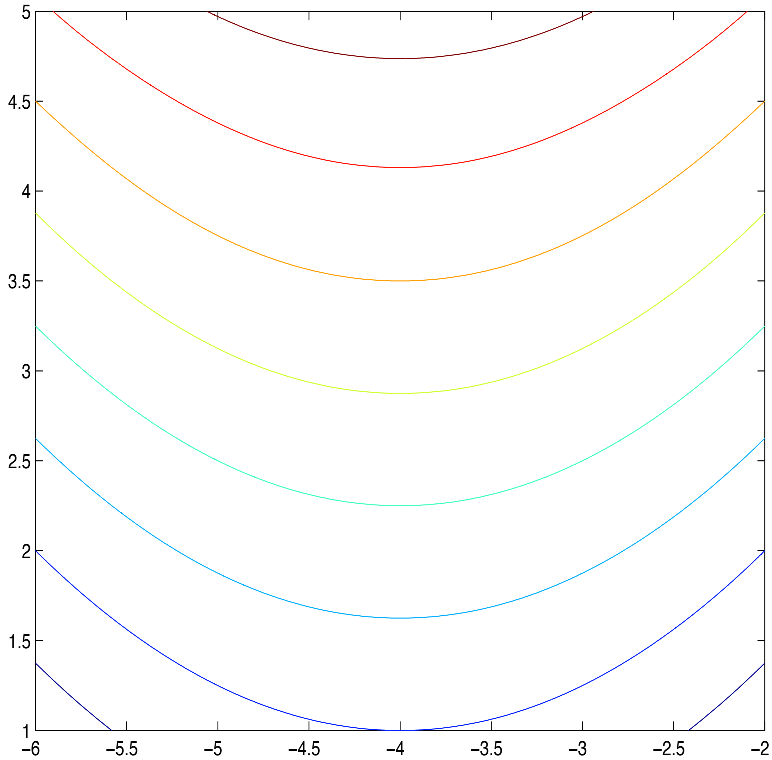

Figure 1: Contour lines of $f(x)$.

(2)

$$\begin{aligned}
f_2(x) &= x^T A x = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}x_i x_j \\
\nabla f_2(x) &= \left[\frac{\partial f_2}{\partial x_s}\right]_{s=1\cdots n} = \left[\sum_j A_{sj}x_j + \sum_i A_{is}x_i\right]_{s=1\cdots n} \\
&= \left[2\sum_{j=1}^{n} A_{sj}x_j\right]_{s=1\cdots n} \qquad \text{(since } A \text{ is symmetric)} \\
&= 2Ax
\end{aligned}$$

$$\begin{aligned}
\nabla^2 f_2(x) &= \left[\frac{\partial^2 f_2}{\partial x_s \partial x_t}\right]_{\substack{s=1\cdots n\\ t=1\cdots n}} = \left[\frac{\partial^2 \sum_i \sum_j A_{ij}x_i x_j}{\partial x_s \partial x_t}\right]_{\substack{s=1\cdots n\\ t=1\cdots n}} \\
&= \left[A_{st} + A_{ts}\right]_{\substack{s=1\cdots n\\ t=1\cdots n}} = 2A
\end{aligned}$$

$\square$

Problem 2.4

For any univariate function $f(x)$, we know that the second oder Taylor expansion is

$$f(x+\Delta x) = f(x) + f^{(1)}(x)\Delta x + \frac{1}{2}f^{(2)}(x+t\Delta x)\Delta x^2,$$

and the third order Taylor expansion is

$$f(x+\Delta x) = f(x) + f^{(1)}(x)\Delta x + \frac{1}{2}f^{(2)}(x)\Delta x^2 + \frac{1}{6}f^{(3)}(x+t\Delta x)\Delta x^3,$$

where $t \in (0,1)$.

For function $f_1(x) = \cos(1/x)$ and any nonzero point x, we know that

$$f_1^{(1)}(x) = \frac{1}{x^2}\sin\frac{1}{x}, \quad f_1^{(2)}(x) = -\frac{1}{x^4}\left(\cos\frac{1}{x} + 2x\sin\frac{1}{x}\right).$$

So the second order Taylor expansion for $f_1(x)$ is

$$\begin{aligned}\cos\tfrac{1}{x+\Delta x} = \;& \cos\tfrac{1}{x} + \left(\tfrac{1}{x^2}\sin\tfrac{1}{x}\right)\Delta x \\ & -\tfrac{1}{2(x+t\Delta x)^4}\left[\cos\tfrac{1}{x+t\Delta x} - 2(x+t\Delta x)\sin\tfrac{1}{x+t\Delta x}\right]\Delta x^2,\end{aligned}$$

where $t \in (0,1)$. Similarly, for $f_2(x) = \cos x$, we have

$$f_2^{(1)}(x) = -\sin x, \quad f_2^{(2)}(x) = -\cos x, \quad f_2^{(3)}(x) = \sin x.$$

Thus the third order Taylor expansion for $f_2(x)$ is

$$\cos(x+\Delta x) = \cos x - (\sin x)\Delta x - \frac{1}{2}(\cos x)\Delta x^2 + \frac{1}{6}[\sin(x+t\Delta x)]\Delta x^3,$$

where $t \in (0,1)$. When $x = 1$, we have

$$\cos(1+\Delta x) = \cos 1 - (\sin 1)\Delta x - \frac{1}{2}(\cos 1)\Delta x^2 + \frac{1}{6}[\sin(1+t\Delta x)]\Delta x^3,$$

where $t \in (0,1)$.

Problem 2.5

Using a trig identity we find that

$$f(x_k) = \left(1 + \frac{1}{2^k}\right)^2 (\cos^2 k + \sin^2 k) = \left(1 + \frac{1}{2^k}\right)^2,$$

from which it follows immediately that $f(x_{k+1}) < f(x_k)$.

Let θ be any point in $[0, 2\pi]$. We aim to show that the point $(\cos\theta, \sin\theta)$ on the unit circle is a limit point of $\{x_k\}$.

From the hint, we can identify a subsequence $\xi_{k_1}, \xi_{k_2}, \xi_{k_3}, \ldots$ such that $\lim_{j\to\infty} \xi_{k_j} = \theta$. Consider the subsequence $\{x_{k_j}\}_{j=1}^{\infty}$. We have

$$\begin{aligned}
\lim_{j\to\infty} x_{k_j} &= \lim_{j\to\infty} \left(1 + \frac{1}{2^k}\right) \begin{bmatrix} \cos k_j \\ \sin k_j \end{bmatrix} \\
&= \lim_{j\to\infty} \left(1 + \frac{1}{2^k}\right) \lim_{j\to\infty} \begin{bmatrix} \cos \xi_{k_j} \\ \sin \xi_{k_j} \end{bmatrix} \\
&= \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}.
\end{aligned}$$

Problem 2.6

We need to prove that "isolated local min" $\Rightarrow$ "strict local min." Equivalently, we prove the contrapositive: "not a strict local min" $\Rightarrow$ "not an isolated local min."

If x^* is not even a local min, then it is certainly not an isolated local min. So we suppose that x^* is a local min but that it is not strict. Let $\mathcal{N}$ be any nbd of x^* such that $f(x^*) \le f(x)$ for all $x \in \mathcal{N}$. Because x^* is not a strict local min, there is some other point $x_{\mathcal{N}} \in \mathcal{N}$ such that $f(x^*) = f(x_{\mathcal{N}})$. Hence $x_{\mathcal{N}}$ is also a local min of f in the neighborhood $\mathcal{N}$ that is different from x^*. Since we can do this for *every* neighborhood of x^* within which x^* is a local min, x^* cannot be an isolated local min.

Problem 2.8

Let S be the set of global minimizers of f. If S only has one element, then it is obviously a convex set. Otherwise for all $x, y \in S$ and $\alpha \in [0, 1]$,

$$f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha) f(y)$$

since f is convex. $f(x) = f(y)$ since x, y are both global minimizers. Therefore,

$$f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha) f(x) = f(x).$$

But since $f(x)$ is a global minimizing value, $f(x) \leq f(\alpha x + (1-\alpha)y)$. Therefore, $f(\alpha x + (1-\alpha y) = f(x)$ and hence $\alpha x + (1-\alpha)y \in S$. Thus S is a convex set. $\square$

Problem 2.9

$-\nabla f$ indicates steepest descent. $(p_k)\cdot(-\nabla f) = \|p_k\|\cdot\|\nabla f\|\cos\theta$. p_k is a descent direction if $-90^\circ < \theta < 90^\circ \iff \cos\theta > 0$.

$$\frac{p_k \cdot -\nabla f}{\|p_k\|\|\nabla f\|} = \cos\theta > 0 \qquad \iff \qquad p_k \cdot \nabla f < 0.$$

$$\nabla f = \begin{bmatrix} 2(x_1 + x_2^2) \\ 4x_2(x_1 + x_2^2) \end{bmatrix}$$

$$p_k \cdot \nabla f_k \Big|_{x=\begin{pmatrix}1\\0\end{pmatrix}} = \begin{pmatrix}-1\\1\end{pmatrix} \cdot \begin{pmatrix}2\\0\end{pmatrix} = -2 < 0$$

which implies that p_k is a descent direction.

$$p_k = \begin{pmatrix}-1\\1\end{pmatrix}, \qquad x = \begin{pmatrix}1\\0\end{pmatrix}$$

$$f(x_k + \alpha_k p_k) = f((1-\alpha, \alpha)^T) = ((1-\alpha) + \alpha^2)^2$$

$$\Longrightarrow \frac{d}{d\alpha} f(x_k + \alpha_k p_k) = 2(1 - \alpha + \alpha^2)(-1 + 2\alpha) = 0 \quad \text{only when } \alpha = \frac{1}{2}.$$

It is seen that $\frac{d^2}{d\alpha^2} f(x_k + \alpha_k p_k)\Big|_{\alpha=\frac{1}{2}} = 6(2\alpha^2 - 2\alpha + 1)\Big|_{\alpha=\frac{1}{2}} = 3 > 0$, so $\alpha = \frac{1}{2}$ is indeed a minimizer.

$\square$

Problem 2.10

Note first that

$$x_j = \sum_{i=1}^{n} S_{ji} z_i + s_j.$$

By the chain rule we have

$$\frac{\partial}{\partial z_i}\tilde{f}(z) = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}\frac{\partial x_j}{\partial z_i} = \sum_{j=1}^{n} S_{ji}\frac{\partial f}{\partial x_j} = \left[S^T \nabla f(x)\right]_i .$$

For the second derivatives, we apply the chain rule again:

$$\begin{aligned}
\frac{\partial^2}{\partial z_i \partial z_k}\tilde{f}(z) &= \frac{\partial}{\partial z_k}\sum_{j=1}^{n} S_{ji}\frac{\partial f(x)}{\partial x_j} \\
&= \sum_{j=1}^{n}\sum_{l=1}^{n} S_{ji}\frac{\partial^2 f(x)}{\partial x_j \partial x_l}\frac{\partial x_l}{\partial z_k} S_{lk} \\
&= \left[S^T \nabla^2 f(x) S\right]_{ki} .
\end{aligned}$$

Problem 2.13

$$x^* = 0$$

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \left|\frac{k}{k+1}\right| < 1 \qquad \text{and} \qquad \frac{k}{k+1} \to 1.$$

$$\text{For any } r \in (0,1), \exists\ k_0 \text{ such that } \ \forall\ k > k_0,\ \frac{k}{k+1} > r.$$

This implies x_k is **not** Q-linearly convergent. □

Problem 2.14

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^2} = \frac{(0.5)^{2^{k+1}}}{((0.5)^{2^k})^2} = \frac{(0.5)^{2^{k+1}}}{(0.5)^{2^{k+1}}} = 1 < \infty.$$

Hence the sequence is Q-quadratic.

Problem 2.15

$$x_k = \frac{1}{k!} \qquad x^* = \lim_{n\to\infty} x_k = 0$$

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \frac{k!}{(k+1)!} = \frac{1}{k+1} \xrightarrow{k\to\infty} 0.$$

This implies x_k is Q-superlinearly convergent.

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^2} = \frac{k!k!}{(k+1)!} = \frac{k!}{k+1} \longrightarrow \infty.$$

This implies x_k is **not** Q-quadratic convergent.

□

Problem 2.16

For k even, we have

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \frac{x_k/k}{x_k} = \frac{1}{k} \to 0,$$

while for k odd we have

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \frac{(1/4)^{2^k}}{x_{k-1}/k} = k\frac{(1/4)^{2^k}}{(1/4)^{2^{k-1}}} = k(1/4)^{2^{k-1}} \to 0,$$

Hence we have

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \to 0,$$

so the sequence is Q-superlinear. The sequence is not Q-quadratic because for k even we have

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^2} = \frac{x_k/k}{x_k^2} = \frac{1}{k}4^{2^k} \to \infty.$$

The sequence is however R-quadratic as it is majorized by the sequence $z_k = (0.5)^{2^k}$, $k = 1, 2, \ldots$. For even k, we obviously have

$$x_k = (0.25)^{2^k} < (0.5)^{2^k} = z_k,$$

while for k odd we have

$$x_k < x_{k-1} = (0.25)^{2^{k-1}} = ((0.25)^{1/2})^{2^k} = (0.5)^{2^k} = z_k.$$

A simple argument shows that z_k is Q-quadratic.

3 Line Search Methods

Problem 3.2

Graphical solution
We show that if c_1 is allowed to be greater than c_2, then we can find a function for which no steplengths $\alpha > 0$ satisfy the Wolfe conditions.

Consider the convex function depicted in Figure 2, and let us choose $c_1 = 0.99$.

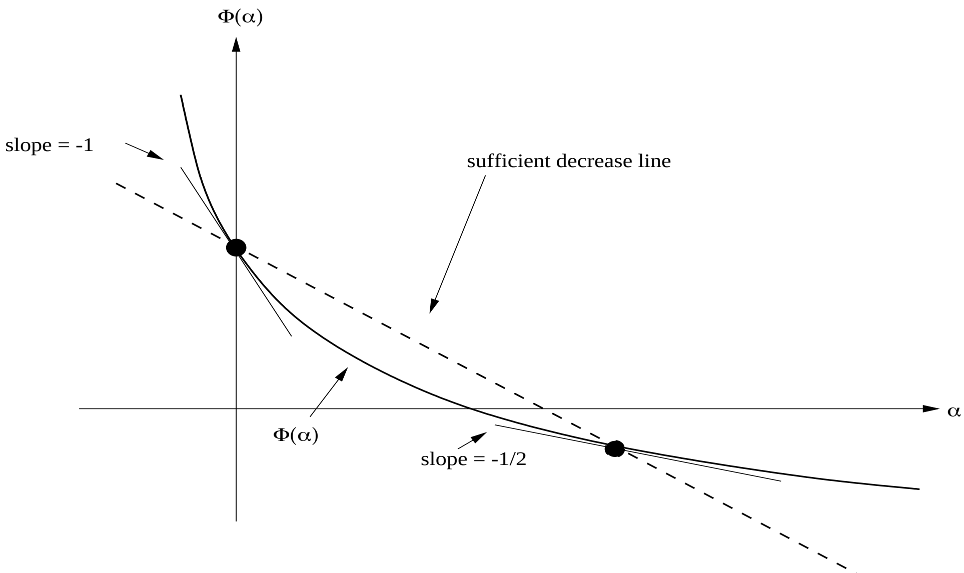

Figure 2: Convex function and sufficient decrease line

We observe that the sufficient decrease line intersects the function only once. Moreover for all points to the left of the intersection, we have

$$\phi'(\alpha) \leq -\frac{1}{2}.$$

Now suppose that we choose $c_2 = 0.1$ so that the curvature condition requires

$$\phi'(\alpha) \geq -0.1. \tag{1}$$

Then there are clearly no steplengths satisfying the inequality (1) for which the sufficient decrease condition holds.

$\square$

Problem 3.3

Suppose p is a descent direction and define

$$\phi(\alpha) = f(x + \alpha p), \qquad \alpha \geq 0.$$

Then any minimizer α^* of $\phi(\alpha)$ satisfies

$$\phi'(\alpha^*) = \nabla f(x + \alpha^* p)^T p = 0. \tag{2}$$

A strongly convex quadratic function has the form

$$f(x) = \frac{1}{2} x^T Q x + b^T x, \qquad Q > 0,$$

and hence

$$\nabla f(x) = Qx + b. \tag{3}$$

The one-dimensional minimizer is unique, and by Equation (2) satisfies

$$[Q(x + \alpha^* p) + b]^T p = 0.$$

Therefore

$$(Qx + b)^T p + \alpha^* p^T Q p = 0$$

which together with Equation (3) gives

$$\alpha^* = -\frac{(Qx + b)^T p}{p^T Q p} = -\frac{\nabla f(x)^T p}{p^T Q p}.$$

□

Problem 3.4

Let $f(x) = \frac{1}{2} x^T Q x + b^T x + d$, with Q positive definite. Let x_k be the current iterate and p_k a non-zero direction. Let $0 < c < \frac{1}{2}$.

The one-dimensional minimizer along $x_k + \alpha p_k$ is (see the previous exercise)

$$\alpha_k = -\frac{\nabla f_k^T p_k}{p_k^T Q p_k}$$

Direct substitution then yields

$$f(x_k) + (1 - c)\alpha_k \nabla f_k^T p_k = f(x_k) - \frac{(\nabla f_k^T p_k)^2}{p_k^T Q p_k} + c \frac{(\nabla f_k^T p_k)^2}{p_k^T Q p_k}$$

Now, since $\nabla f_k = Qx_k + b$, after some algebra we get

$$f(x_k + \alpha_k p_k) = f(x_k) - \frac{(\nabla f_k^T p_k)^2}{p_k^T Q p_k} + \frac{1}{2}\frac{(\nabla f_k^T p_k)^2}{p_k^T Q p_k},$$

from which the first inequality in the Goldstein conditions is evident. For the second inequality, we reduce similar terms in the previous expression to get

$$f(x_k + \alpha_k p_k) = f(x_k) - \frac{1}{2}\frac{(\nabla f_k^T p_k)^2}{p_k^T Q p_k},$$

which is smaller than

$$f(x_k) + c\alpha_k \nabla f_k^T p_k = f(x_k) - c\frac{(\nabla f_k^T p_k)^2}{p_k^T Q p_k}.$$

Hence the Goldstein conditions are satisfied.

Problem 3.5

First we have from (A.7)

$$\|x\| = \|B^{-1}Bx\| \leq \|B^{-1}\| \cdot \|Bx\|,$$

Therefore

$$\|Bx\| \geq \|x\|/\|B^{-1}\|$$

for any nonsingular matrix B.

For symmetric and positive definite matrix B, we have that the matrices $B^{1/2}$ and $B^{-1/2}$ exist and that $\|B^{1/2}\| = \|B\|^{1/2}$ and $\|B^{-1/2}\| = \|B^{-1}\|^{1/2}$. Thus, we have

$$\begin{aligned}
\cos\theta &= -\frac{\nabla f^T p}{\|\nabla f\| \cdot \|p\|} = \frac{p^T B p}{\|Bp\| \cdot \|p\|} \\
&\geq \frac{p^T B p}{\|B\| \cdot \|p\|^2} = \frac{p^T B^{1/2} B^{1/2} p}{\|B\| \cdot \|p\|^2} \\
&= \frac{\|B^{1/2}p\|^2}{\|B\| \cdot \|p\|^2} \geq \frac{\|p\|^2}{\|B^{-1/2}\|^2 \cdot \|B\| \cdot \|p\|^2} \\
&= \frac{1}{\|B^{-1}\| \cdot \|B\|} \geq \frac{1}{M}.
\end{aligned}$$

We can actually prove the stronger result that $\cos\theta \geq 1/M^{1/2}$. Defining $\tilde{p} = B^{1/2}p = -B^{-1/2}\nabla f$, we have

$$\begin{aligned}\cos\theta = \frac{p^T B p}{\|\nabla f\| \cdot \|p\|} &= \frac{\tilde{p}^T \tilde{p}}{\|B^{1/2}\tilde{p}\| \cdot \|B^{-1/2}\tilde{p}\|} \\ = \frac{\|\tilde{p}\|^2}{\|B^{1/2}\| \cdot \|\tilde{p}\| \cdot \|B^{-1/2}\| \cdot \|\tilde{p}\|} &= \frac{1}{\|B^{1/2}\| \cdot \|B^{-1/2}\|} \geq \frac{1}{M^{1/2}}.\end{aligned}$$

Problem 3.6

If $x_0 - x^*$ is parallel to an eigenvector of Q, then

$$\begin{aligned}\nabla f(x_0) &= Qx_0 - b = Qx_0 - Qx^* + Qx^* - b \\ &= Q(x_0 - x^*) + \nabla f(x^*) \\ &= \lambda(x_0 - x^*)\end{aligned}$$

for the corresponding eigenvalue λ. From here, it is easy to get

$$\begin{aligned}\nabla f_0^T \nabla f_0 &= \lambda^2 (x_0 - x^*)^T (x_0 - x^*), \\ \nabla f_0^T Q \nabla f_0 &= \lambda^3 (x_0 - x^*)^T (x_0 - x^*), \\ \nabla f_0^T Q^{-1} \nabla f_0 &= \lambda (x_0 - x^*)^T (x_0 - x^*).\end{aligned}$$

Direct substitution in equation (3.28) yields

$$\|x_1 - x^*\|_Q^2 = 0 \text{ or } x_1 = x^*.$$

Therefore the steepest descent method will find the solution in one step.

Problem 3.7

We drop subscripts on $\nabla f(x_k)$ for simplicity. We have

$$x_{k+1} = x_k - \alpha \nabla f,$$

so that

$$x_{k+1} - x^* = x_k - x^* - \alpha \nabla f,$$

By the definition of $\|\cdot\|_Q^2$, we have

$$\begin{aligned}\|x_{k+1} - x^*\|_Q^2 &= (x_{k+1} - x^*)^T Q (x_{k+1} - x^*) \\ &= (x_k - x^* - \alpha\nabla f)^T Q (x_k - x^* - \alpha \nabla f) \\ &= (x_k - x^*)^T Q (x_k - x^*) - 2\alpha \nabla f^T Q (x_k - x^*) + \alpha^2 \nabla f^T Q \nabla f \\ &= \|x_k - x^*\|_Q^2 - 2\alpha \nabla f^T Q (x_k - x^*) + \alpha^2 \nabla f^T Q \nabla f\end{aligned}$$

Hence, by substituting $\nabla f = Q(x_k - x^*)$ and $\alpha = \nabla f^T \nabla f / (\nabla f^T Q \nabla f)$, we obtain

$$\begin{aligned}
\|x_{k+1} - x^*\|_Q^2 &= \|x_k - x^*\|_Q^2 - 2\alpha \nabla f^T \nabla f + \alpha^2 \nabla f^T Q \nabla f \\
&= \|x_k - x^*\|_Q^2 - 2(\nabla f^T \nabla f)^2 / (\nabla f^T Q \nabla f) + (\nabla f^T \nabla f)^2 / (\nabla f^T Q \nabla f) \\
&= \|x_k - x^*\|_Q^2 - (\nabla f^T \nabla f)^2 / (\nabla f^T Q \nabla f) \\
&= \|x_k - x^*\|_Q^2 \left[1 - \frac{(\nabla f^T \nabla f)^2}{(\nabla f^T Q \nabla f) \|x_k - x^*\|_Q^2} \right] \\
&= \|x_k - x^*\|_Q^2 \left[1 - \frac{(\nabla f^T \nabla f)^2}{(\nabla f^T Q \nabla f)(\nabla f^T Q^{-1} \nabla f)} \right],
\end{aligned}$$

where we used

$$\|x_k - x^*\|_Q^2 = \nabla f^T Q^{-1} \nabla f$$

for the final equality.

Problem 3.8

We know that there exists an orthogonal matrix P such that

$$P^T Q P = \Lambda = \operatorname{diag}\{\lambda_1, \lambda_2, \cdots, \lambda_n\}.$$

So

$$P^T Q^{-1} P = (P^T Q P)^{-1} = \Lambda^{-1}.$$

Let $z = P^{-1} x$, then

$$\frac{(x^T x)^2}{(x^T Q x)(x^T Q^{-1} x)} = \frac{(z^T z)^2}{(z^T \Lambda z)(z^T \Lambda^{-1} z)} = \frac{(\sum_i z_i^2)^2}{(\sum_i \lambda_i z_i^2)(\sum_i \lambda_i^{-1} z_i^2)} = \frac{1}{\frac{\sum_i \lambda_i z_i^2}{\sum_i z_i^2} \cdot \frac{\sum_i \lambda_i^{-1} z_i^2}{\sum_i z_i^2}}.$$

Let $u_i = z_i^2 / \sum_i z_i^2$, then all u_i satisfy $0 \le u_i \le 1$ and $\sum_i u_i = 1$. Therefore

$$\frac{(x^T x)^2}{(x^T Q x)(x^T Q^{-1} x)} = \frac{1}{(\sum_i u_i \lambda_i)(\sum_i u_i \lambda_i^{-1})} = \frac{\phi(u)}{\psi(u)}, \tag{4}$$

where $\phi(u) = \frac{1}{\sum_i u_i \lambda_i}$ and $\psi(u) = \sum_i u_i \lambda_i^{-1}$.

Define function $f(\lambda) = \frac{1}{\lambda}$, and let $\bar{\lambda} = \sum_i u_i \lambda_i$. Note that $\bar{\lambda} \in [\lambda_1, \lambda_n]$. Then

$$\phi(u) = \frac{1}{\sum_i u_i \lambda_i} = f(\bar{\lambda}). \tag{5}$$

Let $h(\lambda)$ be the linear function fitting the data $(\lambda_1, \frac{1}{\lambda_1})$ and $(\lambda_n, \frac{1}{\lambda_n})$. We know that

$$h(\lambda) = \frac{1}{\lambda_n} + \frac{\frac{1}{\lambda_1} - \frac{1}{\lambda_n}}{\lambda_n - \lambda_1}(\lambda_n - \lambda).$$

Because f is convex, we know that $f(\lambda) \le h(\lambda)$ holds for all $\lambda \in [\lambda_1, \lambda_n]$. Thus

$$\psi(\lambda) = \sum_i u_i f(\lambda_i) \le \sum_i u_i h(\lambda_i) = h(\sum_i u_i \lambda_i) = h(\bar{\lambda}). \tag{6}$$

Combining (4), (5) and (6), we have

$$\begin{aligned}
\frac{(x^Tx)^2}{(x^TQx)(x^TQ^{-1}x)} &= \tfrac{\phi(u)}{\psi(u)} \ge \tfrac{f(\bar{\lambda})}{h(\bar{\lambda})} \ge \min_{\lambda_1 \le \lambda \le \lambda_n} \tfrac{f(\lambda)}{h(\lambda)} \quad (\text{since } \bar{\lambda} \in [\lambda_1, \lambda_n]) \\
&= \min_{\lambda_1 \le \lambda \le \lambda_n} \frac{\lambda^{-1}}{\frac{1}{\lambda_n} + \frac{\lambda_n - \lambda}{\lambda_1 \lambda_n}} \\
&= \lambda_1 \lambda_n \cdot \min_{\lambda_1 \le \lambda \le \lambda_n} \frac{1}{\lambda(\lambda_1 + \lambda_n - \lambda)} \\
&= \lambda_1 \lambda_n \cdot \frac{1}{\frac{\lambda_1 + \lambda_n}{2}(\lambda_1 + \lambda_n - \frac{\lambda_1 + \lambda_n}{2})} \quad (\text{since the minimum happens at } d = \tfrac{\lambda_1 + \lambda_n}{2}) \\
&= \frac{4\lambda_1\lambda_n}{(\lambda_1 + \lambda_n)^2}.
\end{aligned}$$

This completes the proof of the Kantorovich inequality.

Problem 3.13

Let $\phi_q(\alpha) = a\alpha^2 + b\alpha + c$. We get a, b and c from the interpolation conditions

$$\begin{aligned}
\phi_q(0) = \phi(0) &\Rightarrow c = \phi(0), \\
\phi_q'(0) = \phi'(0) &\Rightarrow b = \phi'(0), \\
\phi_q(\alpha_0) = \phi(\alpha_0) &\Rightarrow a = (\phi(\alpha_0) - \phi(0) - \phi'(0)\alpha_0)/\alpha_0^2.
\end{aligned}$$

This gives (3.57). The fact that α_0 does not satisfy the sufficient decrease condition implies

$$\begin{aligned}
0 &< \phi(\alpha_0) - \phi(0) - c_1\phi'(0)\alpha_0 \\
&< \phi(\alpha_0) - \phi(0) - \phi'(0)\alpha_0,
\end{aligned}$$

where the second inequality holds because $c_1 < 1$ and $\phi'(0) < 0$. From here, clearly, $a > 0$. Hence, ϕ_q is convex, with minimizer at

$$\alpha_1 = -\frac{\phi'(0)\alpha_0^2}{2\left[\phi(\alpha_0) - \phi(0) - \phi'(0)\alpha_0\right]}.$$

Now, note that

$$\begin{aligned} 0 &< (c_1 - 1)\phi'(0)\alpha_0 \\ &= \phi(0) + c_1\phi'(0)\alpha_0 - \phi(0) - \phi'(0)\alpha_0 \\ &< \phi(\alpha_0) - \phi(0) - \phi'(0)\alpha_0, \end{aligned}$$

where the last inequality follows from the violation of sufficient decrease at α_0. Using these relations, we get

$$\alpha_1 < -\frac{\phi'(0)\alpha_0^2}{2(c_1 - 1)\phi'(0)\alpha_0} = \frac{\alpha_0}{2(1 - c_1)}.$$

4 Trust-Region Methods

Problem 4.4

Since $\liminf \|g_k\| = 0$, we have by definition of the $\liminf$ that $v_i \to 0$, where the scalar nondecreasing sequence v_i is defined by $v_i = \inf_{k \geq i} \|g_k\|$. In fact, since $\{v_i\}$ is nonnegative and nondecreasing and $v_i \to 0$, we must have $v_i = 0$ for all i, that is,

$$\inf_{k \geq i} \|g_k\| = 0, \quad \text{for all } i.$$

Hence, for any $i = 1, 2, \ldots$, we can identify an index $j_i \geq i$ such that $\|g_{j_i}\| \leq 1/i$, so that

$$\lim_{i \to \infty} \|g_{j_i}\| = 0.$$

By eliminating repeated entries from $\{j_i\}_{i=1}^{\infty}$, we obtain an (infinite) subsequence $\mathcal{S}$ of such that $\lim_{i \in \mathcal{S}} \|g_i\| = 0$. Moreover, since the iterates $\{x_i\}_{i \in \mathcal{S}}$ are all confined to the bounded set $\mathcal{B}$, we can choose a further subsequence $\bar{\mathcal{S}}$ such that

$$\lim_{i \in \bar{\mathcal{S}}} x_i = x_\infty,$$

for some limit point x_∞. By continuity of g, we have $\|g(x_\infty)\| = 0$, so $g(x_\infty) = 0$, so we are done.

Problem 4.5

Note first that the scalar function of τ that we are trying to minimize is

$$\phi(\tau) \stackrel{\text{def}}{=} m_k(\tau p_k^{\text{S}}) = m_k(-\tau\Delta_k g_k/\|g_k\|) = f_k - \tau\Delta_k\|g_k\| + \frac{1}{2}\tau^2\Delta_k^2 g_k^T B_k g_k/\|g_k\|^2,$$