Problem 1 (25 %) Second-Order Necessary Conditions

c In Theorem 2.3 the assumptions are that x^* is a local minimizer of f and that $\nabla^2 f$ is continuous in an open neighborhood of x^*. If this is true, then $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is positive semidefinite. Since $\nabla^2 f(x^*)$ is not necessarily positive definite, x^* does not have to be a unique (strict) local minimizer; there may be several points giving the same minimal function value. In Theorem 2.4, however, $\nabla^2 f(x^*)$ is positive definite which ensures that x^* is the only point where $\nabla f(x^*) = 0$ and the minimal function value $f(x^*)$ is achieved in a neighborhood around x^*.

Also note the fundamental difference between how the two theorems are formulated. For the second-order necessary conditions (Theorem 2.3), we already have a local minimizer (and some assumptions), and from that we can conclude something about the gradient and the Hessian. However, for the second-order sufficient conditions (Theorem 2.4) we check if three conditions are satisfied, and if they are, we can conclude that we have a strict local minimizer.

As an example of this, consider the two functions

$$f_1(x) = x_1^2 + 2x_2^2 \tag{1a}$$
$$f_2(x) = x_1^2 \tag{1b}$$

The functions are plotted in Figure 1 and Figure 2, respectively. For both of these functions $x^* = [0,\ 0]^\top$ is a local minimizer. The gradients and Hessians are

$$\nabla f_1(x) = \begin{bmatrix} 2x_1 \\ 4x_2 \end{bmatrix} \qquad \nabla^2 f_1(x) = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} > 0 \tag{2a}$$
$$\nabla f_2(x) = \begin{bmatrix} 2x_1 \\ 0 \end{bmatrix} \qquad \nabla^2 f_2(x) = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \geq 0 \tag{2b}$$

We see that $\nabla^2 f_1(x)$ is positive definite, while $\nabla^2 f_2(x)$ is positive semidefinite. For $f_1(x)$, any point in a neighborhood of x^* will give a function value larger than $f_1(x^*)$, which means that x^* is a strict minimizer for $f_1(x)$. For $f_2(x)$, on the other hand, any point x with $x_1 = 0$ gives the same function value as $f_2(x^*)$, which means that x^* is a non-strict minimizer for $f_2(x)$.

Problem 2 (40 %) The Newton Direction

a Differentiating the model function

$$m_k(p) := f_k + p^\top \nabla f_k + \frac{1}{2} p^\top \nabla^2 f_k p \approx f(x_k + p) \tag{3}$$

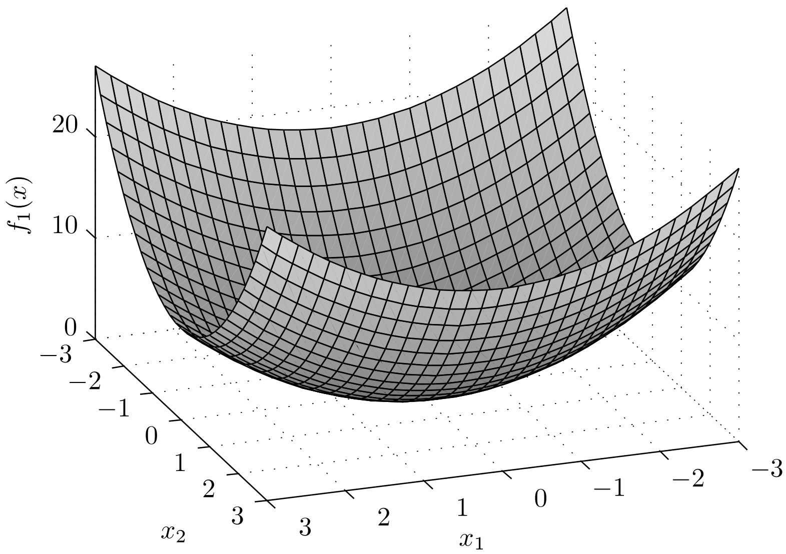

Figure 1: Surface plot of $f_1(x) = x_1^2 + 2x_2^2$.

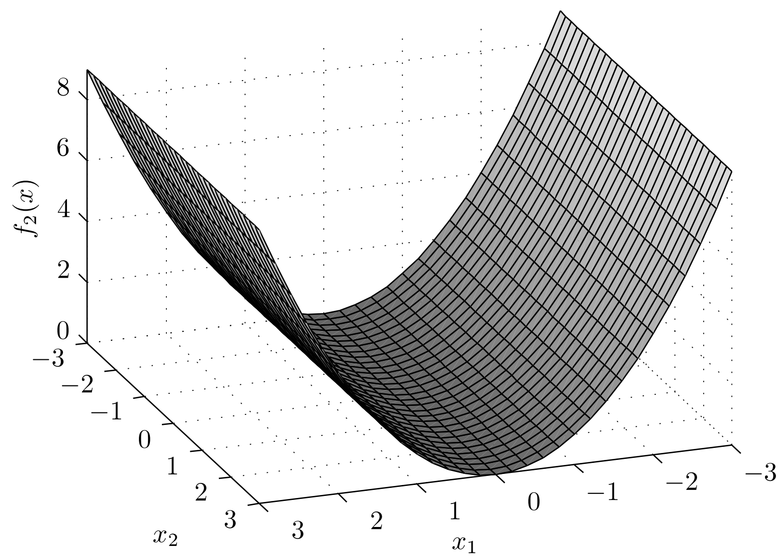

Figure 2: Surface plot of $f_2(x) = x_1^2$.

with respect to p and setting this to zero gives

$$\begin{aligned} m_k'(p) &= \nabla f_k + \frac{1}{2}(\nabla^2 f_k)^\top p + \frac{1}{2}\nabla^2 f_k p \\ &= \nabla f_k + \nabla^2 f_k p = 0 \end{aligned} \tag{4}$$

where we used the fact that the Hessian is symmetric (due to Clairaut's Theorem). Rearranging the last line of this equation (and assuming that the Hessian is nonsingular) gives

$$p = -(\nabla^2 f_k)^{-1} \nabla f_k \tag{5}$$

which is called the Newton direction.

b The rate of change in f in the direction p_k^N is given by $\nabla f_k^\top p_k^N$. We can write the change as

$$\nabla f_k^\top p_k^N = -\nabla f_k^\top (\nabla^2 f_k)^{-1} \nabla f_k \tag{6}$$

When $\nabla^2 f_k < 0$, then $(\nabla^2 f_k)^{-1} < 0$, and we have that the rate of change in f along p_k^N is $\nabla f_k^\top p_k^N > 0$. Hence, p_k^N is not a descent direction when $\nabla^2 f_k < 0$ is negative definite. When $\nabla^2 f_k$ is indefinite (both positive and negative eigenvalues), we cannot say whether p_k^N is a descent direction. When $(\nabla^2 f_k)^{-1} \leq 0$ or $(\nabla^2 f_k)^{-1} \geq 0$, p_k^N does not exist since the Hessian is singular and thereby not invertible.

c Given an unconstrained minimization problem with objective function

$$f(x) = \frac{1}{2} x^\top G x + x^\top c \tag{7}$$

with $G = G^\top > 0$ and $x \in \mathbb{R}^n$. The gradient of f is

$$\nabla f = \frac{1}{2} G x + \frac{1}{2} G^\top x + c = Gx + c \tag{8}$$

By convexity, the minimizer for the problem is then $x^* = -G^{-1}c$. The Hessian is

$$\nabla^2 f = G^\top = G \tag{9}$$

The Newton direction is then

$$p_k^N = -(\nabla^2 f_k)^{-1} \nabla f_k = -G^{-1}(Gx_k + c) = -x_k - G^{-1}c \tag{10}$$

Using this direction in the iteration sequence, we get that from any arbitrary point x_k, the next point x_{k+1} will be

$$x_{k+1} = x_k + p_k^N = x_k - x_k - G^{-1}c = -G^{-1}c = x^* \tag{11}$$

This means that regardless of starting point, the iteration sequence always converges to the optimum in one step.

d First, we note that the domain of f, $X = \{x \in \mathbb{R}^2 \mid x_1^2 + x_2^2 \leq 1\}$, is convex. To show that

$$f(x) = \frac{1}{2}x^\top G x + x^\top c, \qquad G = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad c = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{12}$$

is a convex function on X, we must show that for any two points x and y in S, the following holds:

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y), \qquad \text{for all } \alpha \in [0,\ 1] \tag{13a}$$

or equivalently

$$\Delta := f(\alpha x + (1-\alpha)y) - \alpha f(x) - (1-\alpha) f(y) \leq 0, \qquad \text{for all } \alpha \in [0,\ 1] \tag{13b}$$

Starting with the left-hand side of equation (13a), we have

$$\begin{aligned} f(\alpha x + (1-\alpha)y) =& \frac{1}{2}(\alpha x + (1-\alpha)y)^\top G(\alpha x + (1-\alpha)y) + (\alpha x + (1-\alpha)y)^\top c \\ =& \frac{1}{2}\alpha^2 x^\top G x + \frac{1}{2}\alpha(1-\alpha)x^\top G y \\ &+ \frac{1}{2}(1-\alpha)\alpha y^\top G x + \frac{1}{2}(1-\alpha)^2 y^\top G y + \alpha x^\top c + (1-\alpha)y^\top c \\ =& \frac{1}{2}\alpha^2 x^\top G x + (1-\alpha)\alpha x^\top G y + \frac{1}{2}(1-\alpha)^2 y^\top G y \\ &+ \alpha x^\top c + (1-\alpha) y^\top c \end{aligned} \tag{14}$$

while the right hand-side is

$$\alpha f(x) + (1-\alpha) f(y) = \alpha\frac{1}{2}x^\top G x + \alpha x^\top c + (1-\alpha)\frac{1}{2}y^\top G y + (1-\alpha)y^\top c \tag{15}$$

Now, subtracting equation (15) from equation (14) gives the form (13b). We see that both equations contain the terms $\alpha x^\top c + (1-\alpha)y^\top c$, which means that these terms disappear after subtraction. We are then left with

$$\Delta = \frac{1}{2}\alpha^2 x^\top G x + (1-\alpha)\alpha x^\top G y + \frac{1}{2}(1-\alpha)^2 y^\top G y - \alpha\frac{1}{2}x^\top G x - (1-\alpha)\frac{1}{2}y^\top G y \tag{16}$$

After some rearranging, we have

$$\begin{aligned} \Delta &= \frac{1}{2}\alpha(\alpha-1)x^\top G x + \frac{1}{2}\alpha(\alpha-1)y^\top G y + \alpha(1-\alpha)x^\top G y \\ &= -\frac{1}{2}\left(\alpha(1-\alpha)x^\top G x - 2\alpha(1-\alpha)x^\top G y + \alpha(1-\alpha)y^\top G y\right) \\ &= -\frac{1}{2}\alpha(1-\alpha)\left(x^\top G x - 2x^\top G y + y^\top G y\right) \\ &= -\frac{1}{2}\alpha(1-\alpha)(x-y)^\top G(x-y), \quad z := x - y \\ &= -\frac{1}{2}\alpha(1-\alpha)z^\top G z \end{aligned} \tag{17}$$

Since $\alpha \in [0,\ 1]$, we have that $\alpha \geq 0$ and $(1-\alpha) \geq 0$. From the definition of a positive definite matrix, we have that if G is positive definite, then $z^\top G z > 0$ for every nonzero z. The matrix G in this problem has the eigenvalues 3 and 1, and is therefore positive definite. Hence, we have $\Delta \leq 0$, and the function f is convex.

Problem 3 (35 %) The Rosenbrock Function

The Rosenbrock function is given by

$$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2 \tag{18}$$

The gradient is

$$\nabla f(x) = \begin{bmatrix} 200(x_2 - x_1^2)(-2x_1) - 2(1 - x_1) \\ 200(x_2 - x_1^2) \end{bmatrix} = \begin{bmatrix} -400(x_1 x_2 - x_1^3) + 2x_1 - 2 \\ 200(x_2 - x_1^2) \end{bmatrix} \tag{19}$$

whereas the Hessian is

$$\nabla^2 f(x) = \begin{bmatrix} -400x_2 + 1200x_1^2 + 2 & -400x_1 \\ -400x_1 & 200 \end{bmatrix} \tag{20}$$

To find local minimizers, we find the points where the gradient is zero:

$$\nabla f(x) = \begin{bmatrix} -400(x_1 x_2 - x_1^3) + 2x_1 - 2 \\ 200(x_2 - x_1^2) \end{bmatrix} = 0 \tag{21}$$

The second equation gives $x_2 = x_1^2$, which when substituted in the first equation gives

$$-400(x_1^3 - x_1^3) + 2x_1 - 2 = 0 \Rightarrow x_1 = 1 \Rightarrow x_2 = 1 \tag{22}$$

Hence, we have shown that $x^* = [1,\ 1]^\top$ is the only minimizer for the Rosenbrock function.

The Hessian at this point is

$$\nabla^2 f(x^*) = \begin{bmatrix} 802 & -400 \\ -400 & 200 \end{bmatrix} \tag{23}$$

A convenient way to check if this matrix is positive definite is to check if all leading principal minors of $\nabla^2 f(x^*)$ are positive. The first leading principal minor is $802 > 0$. The second leading principal minor is $802 \times 200 - (-400)^2 = 400 > 0$. Hence, the Hessian matrix is positive definite at x^*.