Assignment 8

Problem 1: Second-order necessary conditions

a) If x^* is a local minimizer of $f(x)$ and $\nabla^2 f(x)$ exists and is continuous in an open neighborhood of x^*, then

$$\nabla f(x^*) = 0$$

$$\nabla^2 f(x^*) \geq 0$$

b) If the theorem was not true we would be able to move along a direction p in the vacinity of x^* such that $f(x^* + tp)$ would be smaller than $f(x^*)$. Hence x^* would nat be a local minimizer. Therefore the conditions stated in theorem 2.3 must hold for a local minimizer x^*.

c) Theorem 2.3 only states that $\nabla^2 f(x^*) \geq 0$, while theorem 2.4 states that $\nabla^2 f(x^*) > 0$ for a local minimizer x^*.

Hence theorem 2.3 also includes the case where $\nabla^2 f(x^*) = 0$ meaning that one could move within the vaccinity of x^* along a path p, i.e $\tilde{x} = x^* + tp$ with the following result:

$$\begin{aligned} f(\tilde{x}) &= f(x^* + tp) \\ &\approx f(x^*) + tp^T \nabla f(x^*) \\ &\quad + \frac{1}{2} t^2 p^T \nabla^2 f(x^*) p^T \\ &= f(x^*) \end{aligned}$$

Hence theorem 2.3 would allow for multiple local minimizers, while theorem 2.4 would only allow for one.

Problem 2: The Newton direction

$$m_k(p) = f_k + p^T \nabla f_k + \frac{1}{2} p^T \nabla^2 f_k p$$
$$\approx f(x_k + p)$$

a) $$\frac{dm_k}{dp}(p) = \nabla f_k + \nabla^2 f_k p$$

$$\frac{dm_k}{dp}(p) = 0$$

$$\Longrightarrow \quad \nabla^2 f_k p = -\nabla f_k$$

Assuming $\nabla^2 f_k > 0$

$\Longrightarrow$ $(\nabla^2 f_k)^{-1}$ exists

$$\Longrightarrow \quad \underline{\underline{p = -(\nabla^2 f_k)^{-1} \nabla f_k}}$$

b) $\nabla^2 f_k < 0$

$\Longrightarrow$ $(\nabla^2 f_k)^{-1}$ exists

$$\frac{dm_k}{dp}(p) = \nabla f_k + \nabla^2 f_k p$$

$$\frac{dm_k}{dp}(p) = 0 \implies \nabla f_k + \nabla^2 f_k p = 0$$

$$\nabla^2 f_k < 0 \iff (\nabla^2 f_k)^{-1} < 0$$

$$\implies \nabla f_k^T p = -\nabla f_k^T (\nabla^2 f_k)^{-1} \nabla f_k > 0$$

$\nabla f_k^T p$ is equal to the change in the objective function as the result of a step along p. Since $\nabla f_k^T p > 0$ it means that the objective function will increase along p. Hence p is not a descent direction, but an ascend direction.

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

$G = G^T > 0$

$x_{k+1} = x_k + p_k^N$

$$x_{k+1} = x_k + \left(-(\nabla^2 f_k)^{-1} \nabla f_k\right)$$

$$\nabla f = Gx + c$$

$$\nabla^2 f = G$$

$$\Longrightarrow x_{k+1} = x_k - G^{-1}(Gx_k + c)$$

$$x_{k+1} = x_k - x_k - G^{-1}c$$

$$x_{k+1} = -G^{-1}c$$

$$f(x_{k+1}) = \frac{1}{2} x_{k+1}^T G x_{k+1} + x_{k+1}^T c$$

$$= \frac{1}{2}(-G^{-1}c)^T G(-G^{-1}c) + (-G^{-1}c)^T c$$

$$= \frac{1}{2} c^T (G^{-1})^T G G^{-1} c - c^T (G^{-1})^T c$$

$$G = G^T \longrightarrow G^{-1} = (G^{-1})^T.$$

$$\Longrightarrow f(x_{k+1}) = \frac{1}{2} c^T G^{-1} c - c^T G^{-1} c$$

$$= -\frac{1}{2} c^T G^{-1} c$$

$$\nabla f(x_{k+1}) = \nabla f(-G^{-1}c)$$

$$= G\cdot(-G^{-1}c) + c$$

$$= -c + c = 0$$

$$\nabla^2 f(x_{k+1}) = G^T = G > 0$$

By theorem 2.4 x_{k+1} is a strict minimizer of f.

d) $f(x) = \frac{1}{2} x^T G x + x^T c$

$x \in X$, $X = \{x \in R^2 \mid x_1^2 + x_2^2 \leq 1\}$

$G = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$, $c = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

$f(x)$ is a convex function

$\Updownarrow$

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

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

$$= \frac{1}{2}\alpha x^T G \alpha x + \frac{1}{2}\alpha x^T G (1-\alpha)y$$
$$+ \frac{1}{2}(1-\alpha)y^T G (1-\alpha)y$$
$$+ \frac{1}{2}(1-\alpha)y^T G \alpha x$$
$$+ \alpha x^T c$$
$$+ (1-\alpha)y^T c$$

$$= \alpha^2 \frac{1}{2} x^T G x + \alpha x^T c$$
$$+ (1-\alpha)^2 \frac{1}{2} y^T G y + (1-\alpha)y^T c$$
$$+ \alpha(1-\alpha) x^T G y$$

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

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

$$= \alpha^2 \tfrac{1}{2} x^T G x + \cancel{\alpha x^T c} + (1-\alpha)^2 \tfrac{1}{2} y^T G y^T + \cancel{(1-\alpha) y^T c}$$

$$+ \alpha(1-\alpha) x^T G y$$

$$- \alpha \tfrac{1}{2} x^T G x - \cancel{\alpha x^T c} - (1-\alpha) y^T G y - \cancel{(1-\alpha) y^T c}$$

$$= \tfrac{1}{2}\alpha^2 x^T G x + \tfrac{1}{2}(1-\alpha)^2 y^T G y$$

$$+ \alpha(1-\alpha) x^T G y$$

$$- \alpha \tfrac{1}{2} x^T G x - (1-\alpha) y^T G y$$

$$= \tfrac{1}{2}\Big[\alpha^2 x^T G x + (1-\alpha)^2 y^T G y$$

$$+ 2\alpha(1-\alpha) x^T G y$$

$$- \alpha x^T G x - (1-\alpha) y^T G y\Big]$$

$$= \tfrac{1}{2}\Big[\alpha(\alpha-1) x^T G x + (1-\alpha)((\cancel{1}-\alpha)-\cancel{1}) y^T G y$$

$$+ 2\alpha(1-\alpha) x^T G y\Big]$$

$$= \tfrac{1}{2}\Big[\alpha(\alpha-1) x^T G x - (1-\alpha)\alpha\, y^T G y$$

$$+ 2\alpha(1-\alpha) x^T G y\Big]$$

$$= \tfrac{1}{2}\alpha(1-\alpha)\left[-y^T G y - x^T G x + 2 x^T G y\right]$$

$$= -\frac{1}{2}\alpha(1-\alpha)\left[y^T G y + x^T G x - 2x^T G y \right]$$

$$= -\frac{1}{2}\alpha(1-\alpha)\left[(y-x)^T G (y-x) \right]$$

$$\leq 0$$

$$\Longrightarrow f(\alpha x + (1-\alpha) y) \leq \alpha f(x) + (1-\alpha) f(y)$$

$\Longleftrightarrow$ f is a convex function

$$X = \{ x \in \mathbb{R}^2 \mid x_1^2 + x_2^2 \leq 1 \}$$

By inspetion X is a convex domain as it is the unit circle.

$$\left. \begin{array}{l} f \text{ is convex} \\ X \text{ is convex} \end{array} \right\} \Longrightarrow \text{The problem is convex}$$

Problem 3 : The Rosenbrock function

$$f(x) = 100\,(x_2 - x_1^2)^2 + (1 - x_1)^2$$

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix}$$

$$= \begin{bmatrix} 200\,(x_2 - x_1^2)(-2x_1) + 2\,(1 - x_1)\cdot(-1) \\ 200\,(x_2 - x_1^2)\cdot 1 \end{bmatrix}$$

$$= \begin{bmatrix} -400\,(x_2 - x_1^2)x_1 + 2\,(x_1 - 1) \\ 200\,(x_2 - x_1^2) \end{bmatrix}$$

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix}$$

$$\frac{\partial^2 f}{\partial x_1^2} = \frac{\partial}{\partial x_1}\left(-400x_1(x_2 - x_1^2) + 2(x_1 - 1)\right)$$

$$= -400 \cdot (x_2 - x_1^2) - 400x_1(-2x_1)$$

$$+ 2$$

$$= -400(x_2 - x_1^2 - 2x_1^2) + 2$$

$$= -400(x_2 - 3x_1^2) + 2$$

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial}{\partial x_2}\left(-400x_1(x_2 - x_1^2) + 2(x_1 - 1)\right)$$

$$= -400x_1$$

$$\frac{\partial^2 f}{\partial x_2 \partial x_1} = \frac{\partial}{\partial x_1}\left(200(x_2 - x_1^2)\right)$$

$$= 200 \cdot (-2x_1) = -400x_1$$

$$\frac{\partial^2 f}{\partial x_2^2} = \frac{\partial}{\partial x_2}\left(200\left(x_2 - x_1^2\right)\right)$$

$$= 200$$

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