TMA4180
Solutions to recommended exercises in Chapter 2 of Nocedal and Wright

Exercise 2.1

The gradient and Hessian of $f(x)$ are calculated directly from their definitions:

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 100 \cdot 2(x_2 - x_1^2)(-2x_1) + 2(1-x_1)(-1) \\ 100 \cdot 2(x_2 - x_1^2) \cdot 1 + 0 \end{bmatrix}$$
$$= \begin{bmatrix} -400(x_2 - x_1^2)x_1 - 2(1-x_1) \\ 200(x_2 - x_1^2) \end{bmatrix}.$$

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

Observe that $\nabla^2 f(x)$ is continuous for all $x \in \mathbb{R}^2$. Candidates for local minimizer are points x^* s.t. $\nabla f(x^*) = 0$. This gives the system of two equations,

$$-400(x_2 - x_1^2)x_1 - 2(1-x_1) = 0,$$
$$200(x_2 - x_1^2) = 0,$$

whose only solution is $x^* = (1,1)^T$. By theorem 2.2 (First-Order Necessary Conditions), this is the only candidate for a minimum. The Hessian at x^* is

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

which is positive definite because

$$z^T \nabla^2 f(x^*) z = 802z_1^2 - 800z_1z_2 + 200z_2^2 = 2z_1^2 + 200(2z_1 - z_2)^2 > 0, \quad \forall z = (z_1, z_2)^T \neq 0.$$

Thus, by Theorem 2.4 (Second-Order Sufficient Conditions), x^* is a local minimizer of f.

Exercise 2.2

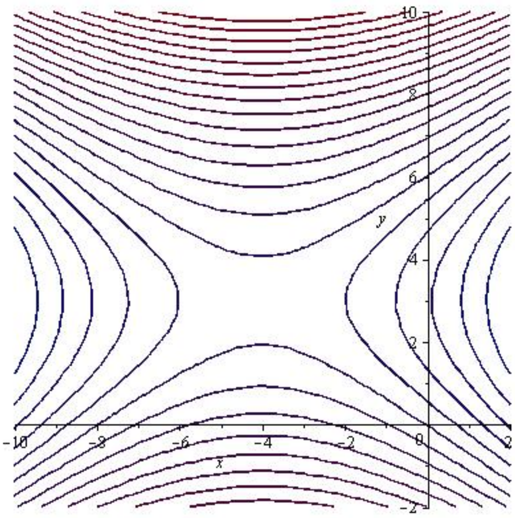
A point x^* is stationary if $\nabla f(x^*) = 0$. Here,

$$\nabla f(x^*) = \begin{bmatrix} 8 + 2x_1^* \\ 12 - 4x_2^* \end{bmatrix} = 0 \quad \Rightarrow x^* = (-4, 3)^T \tag{1}$$

The Hessian,

$$\nabla^2 f(x) = \begin{bmatrix} 8 & 0 \\ 0 & -4 \end{bmatrix}, \tag{2}$$

is indefinite[1] $\forall x \in \mathbb{R}^2$, and thus x^* is a saddle. The contour lines are seen in the figure below.



Exercise 2.3

We do the calculations here very thoroughly as the results are used over and over again in this course.

$$f_1(x) = a^T x = a_1 x_1 + a_2 x_2 + \ldots + a_n x_n,$$

$$\nabla f_1(x) = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = a,$$

[1]A diagonal matrix has eigenvalues equal to its diagonal elements, and a matrix is indefinite if it has both positive and negative eigenvalues.

$$\nabla^2 f_1(x) = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix} = 0, \qquad (\nabla f_1(x) \text{ is a constant vector.})$$

For f_2, let a_{ij} denote the jth entry of row i of A. Then

$$\begin{aligned} f_2(x) = x^T A x = x^T &\begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n \end{bmatrix} \\ = &\; a_{11}x_1^2 \quad + a_{12}x_2x_1 + \ldots + a_{1n}x_nx_1 + \\ &\; a_{21}x_1x_2 \; + a_{22}x_2^2 \quad + \ldots + a_{2n}x_nx_2 + \\ &\; \ldots + \\ &\; a_{n1}x_1x_n + a_{n2}x_2x_n + \ldots + a_{nn}x_n^2 \\ = &\sum_{i=1}^{n} a_{ii}x_i^2 + 2 \sum_{i,j=1, i \neq j}^{n} a_{ij}x_ix_j \end{aligned}$$

On the last line we use that A is symmetric, i.e., $a_{ij} = a_{ji}$. Next, we calculate the gradient and Hessian of f_2,

$$\nabla f_2(x) = \begin{bmatrix} 2a_{11}x_1 + 2a_{12}x_2 + \ldots + 2a_{1n}x_n \\ \vdots \\ 2a_{n1}x_1 + 2a_{n2}x_2 + \ldots + 2a_{nn}x_n \end{bmatrix} = 2Ax,$$

$$\nabla^2 f_2(x) = \begin{bmatrix} 2a_{11} & 2a_{12} & \cdots & 2a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ 2a_{n1} & 2a_{n2} & \cdots & 2a_{nn} \end{bmatrix} = 2A.$$

Hence, for a quadratic function $f(x) = x^T A x + a^T x + c$ with $A \in \mathbb{R}^{n \times n}$ symmetric, $a \in \mathbb{R}^n$ and $c \in \mathbb{R}$, $\nabla f(x) = 2Ax + a$ and $\nabla^2 f(x) = 2A$.

Exercise 2.6

Let x^* be an isolated local minimizer. Then there exists a neighborhood of x^*, $\mathcal{N}$, s.t. x^* is the only local minimizer in $\mathcal{N}$, i.e.,

$$f(x^*) \leq f(x), \quad \forall x \in \mathcal{N}.$$

Since x^* is the only point in $\mathcal{N}$ that has this property, $f(x) = f(x^*)$ only when $x = x^*$. Thus, the following must hold,

$$f(x^*) < f(x), \quad \forall x \in \mathcal{N} \setminus \{x^*\}.$$

Hence, x^* is a strict local minimizer by the definition.

Exercise 2.8

f is convex on a set S if

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y), \quad \forall x, y \in S,\ \forall \alpha \in [0,1]. \tag{3}$$

Define the set of global minimizers,

$$\Omega = \{\bar{x} \in S | f(\bar{x}) \leq f(x),\ \forall x \in S\}.$$

To show that Ω is a convex set, we need to prove that

$$\alpha\bar{x} + (1-\alpha)\bar{y} \in \Omega, \quad \forall \bar{x}, \bar{y} \in \Omega,\ \forall \alpha \in [0,1].$$

Since $\Omega \subset S$, we have from (3) that

$$\begin{aligned} f(\alpha\bar{x} + (1-\alpha)\bar{y}) &\leq \alpha f(\bar{x}) + (1-\alpha)f(\bar{y}) \\ &\leq^{(*)} \alpha f(x) + (1-\alpha)f(x) = f(x), \quad \forall \bar{x}, \bar{y} \in \Omega,\ \forall \alpha \in [0,1]. \end{aligned}$$

Hence $\alpha\bar{x} + (1-\alpha)\bar{y} \in \Omega$.

(*) We use that for $\bar{y} \in \Omega,\ f(\bar{y}) \leq f(x),\ \forall x \in S$.