# Module 2: solutions to textbook exercises

**4.3** Let $f : \mathbb{R}_+ \to \mathbb{R}$ be defined as $f(x) = x \ln x$ where we use the convention that $0 \ \ln 0 = 0$.

(a) Show that $f$ is a convex function.

(b) Show that $g : \mathbb{R}^n_+ \to \mathbb{R}$ defined as $g(x) = \sum_{i=1}^n f(x_i)$ is a convex function.

(c) Derive the conjugate function of $f$.

**Solution**

(a) We have $f'(x) = 1 + \ln x$ and $f''(x) = 1/x > 0$ for $x \in \mathbb{R}_{++}$, and hence $f$ is convex on $\mathbb{R}_{++}$. To see that $f$ is convex on $\mathbb{R}_+$, we note that for $x = 0$ and $y > 0$, we have for all $\theta \in [0, 1)$,

$$\begin{aligned} f(\theta x + (1-\theta)y) &= f((1-\theta)y) = (1-\theta)y\ln(y) + (1-\theta)y\ln(1-\theta) \\ &\le (1-\theta)f(y) = (1-\theta)y\ln(y). \end{aligned}$$

(b) The function $g$ is a sum of convex functions, and hence it is a convex function.

(c) We have that

$$f^*(y) = \sup_{x \ge 0}\{yx - x\ln x\}.$$

The derivative of $h(x) = yx - x\ln x$ is given by $h'(x) = y - \ln x - 1$, and hence $x = e^{y-1}$ maximizes $h$ and

$$f^*(y) = e^{y-1}(y - (y-1)) = e^{y-1}.$$

**4.4** Show that the function $f : \mathbb{R}^n \to \mathbb{R}$ defined as

$$f(x) = \ln(e^{x_1} + \cdots + e^{x_n})$$

is convex.

**Solution** Expressing the function as $f(x) = \ln(\mathbb{1}^T z(x))$ with $z(x) = (e^{x_1}, \dots, e^{x_n})$ and applying the chain rule yields (*cf.* Example 2.4)

$$\frac{\partial f}{\partial x^T} = \frac{1}{\mathbb{1}^T z(x)}\mathbb{1}^T \operatorname{diag}(z(x)) = \frac{z(x)^T}{\mathbb{1}^T z(x)},$$

and hence

$$\nabla f(x) = \frac{z(x)}{\mathbb{1}^T z(x)}.$$

Applying the chain rule once more, we obtain the Hessian

$$\begin{aligned} \nabla^2 f(x) &= \frac{1}{\mathbb{1}^T z(x)}\operatorname{diag}(z(x)) - z(x)\frac{1}{(\mathbb{1}^T z(x))^2}\mathbb{1}^T \operatorname{diag}(z(x)) \\ &= \operatorname{diag}(\nabla f(x)) - \nabla f(x)\nabla f(x)^T. \end{aligned}$$

To show that $f$ is convex, we will show that $\nabla^2 f(x)$ is positive semidefinite for all $x$. To this end, note that $\nabla f(x) \in \mathbb{R}^n_{++}$ and $\mathbb{1}^T \nabla f(x) = 1$ for all $x \in \mathbb{R}^n$, and hence $\nabla^2 f(x)\mathbb{1} = 0$, which implies that the Hessian is singular. Moreover, the Hessian satisfies the identity

$$\nabla^2 f(x) = \nabla^2 f(x)\operatorname{diag}(\nabla f(x))^{-1}\nabla^2 f(x),$$

and this implies that for all $v \in \mathbb{R}^n$,

$$\begin{aligned} v^T \nabla^2 f(x) v &= v^T \nabla^2 f(x) \operatorname{diag}(\nabla f(x))^{-1} \nabla^2 f(x) v \\ &= \| \operatorname{diag}(\nabla f(x))^{-1/2} \nabla^2 f(x) v \|_2^2 \\ &\geq 0. \end{aligned}$$

In other words, $\nabla^2 f$ is positive semidefinite, and hence $f$ is convex.

Alternatively, we can show that

$$v^T(\operatorname{diag}(\nabla f(x)) - \nabla f(x) \nabla f(x)^T) v = v^T \operatorname{diag}(\nabla f(x)) v - (\nabla f(x)^T v)^2 \geq 0,$$

using the Cauchy-Schwarz inequality. Indeed, letting $y = \operatorname{diag}(\nabla f(x))^{1/2} v$ and $w = \operatorname{diag}(\nabla f(x))^{1/2} \mathbb{1}$, we find that

$$v^T(\operatorname{diag}(\nabla f(x)) - \nabla f(x) \nabla f(x)^T) v = \|y\|_2^2 - (w^T y)^2.$$

By the Cauchy-Schwarz inequality, it holds that $|w^T y| \leq \|w\|_2 \|y\|_2$, and noting that $\|w\|_2 = \mathbb{1}^T \nabla f(x) = 1$, we can conclude that $\|y\|_2^2 - (w^T y)^2 \geq 0$.

**4.5** Let $g : \mathbb{R} \to \mathbb{R}$ be defined as $g(t) = f(x + vt)$ where $f : \mathbb{R}^n \to \mathbb{R}$ and $v \in \mathbb{R}^n$ are given, and $\operatorname{dom} g = \{t \mid x + tv \in \operatorname{dom} f\}$.

(a) Show that $f$ is a convex function if and only if $g$ is a convex function for all $x \in \operatorname{dom} f$ and $v \in \mathbb{R}^n$.

(b) Show that $f : \mathbb{S}^n \to \mathbb{R}$ defined as $f(X) = -\ln \det X$ with $\operatorname{dom} f = \mathbb{S}^n_{++}$ is a convex function. *Hint:* Show that $g(t) = -\ln \det(X + Vt)$ is convex for all $X \in \mathbb{S}^n_{++}$ and $V \in \mathbb{S}^n$.

**Solution**

(a) The function $g$ is convex if and only if for all $s, t \in \operatorname{dom} g$ and for $\theta \in [0, 1]$ it holds that $f(x + v(\theta s + (1-\theta)t)) \leq \theta f(x + vs) + (1-\theta) f(x + vt)$. With $z = x + vs$ and $y = x + vt$ this can also be written as $f(\theta z + (1-\theta)y) \leq \theta f(z) + (1-\theta) f(y)$. Hence, it follows that if $f$ is convex, then so is $g$. Since any $z, y \in \operatorname{dom} f$ can be written as above with, *e.g.*, $x = z$, $v = y - z$, $s = 0$ and $t = 1$, the converse also holds.

(b) We have

$$\begin{aligned} g(t) = -\ln \det(X + Vt) &= -\ln \det(X^{1/2}(I + X^{-1/2} V X^{-1/2} t) X^{1/2}) \\ &= -\sum_{i=1}^n \ln(1 + \lambda_i t) - \ln \det X \end{aligned}$$

where $\lambda_i$, $i \in \mathbb{N}_n$ are the eigenvalues of $X^{-1/2} V X^{-1/2}$. Hence

$$g'(t) = -\sum_{i=1}^n \frac{\lambda_i}{1 + \lambda_i t}; \qquad g''(t) = \sum_{i=1}^n \frac{\lambda_i^2}{(1 + \lambda_i t)^2} \geq 0$$

and therefore $f$ is convex.

**4.7** Show that $f : \mathbb{S}^n_{++} \to \mathbb{R}_+$ defined as $f(X) = y^T X^{-1} y$ is a convex function.
*Hint:* Use the Schur complement formula (2.58) to show that $\operatorname{epi} f$ is a convex set.

**Solution** By the Schur complement formula, it holds that

$$\begin{aligned}\operatorname{epi} f &= \{(X,t) \in \mathbb{S}_{++} \times \mathbb{R}_+ \mid f(X) \leq t\} \\ &= \left\{(X,t) \in \mathbb{S}_{++} \times \mathbb{R}_+ \;\middle|\; \begin{bmatrix} t & y^T \\ y & X \end{bmatrix} \in \mathbb{S}_+^{n+1} \right\}.\end{aligned}$$

Since $\mathbb{S}_+^{n+1}$ is a convex cone and the $(n+1) \times (n+1)$ matrix is an affine function of $t$ and $X$, it follows that $\operatorname{epi} f$ is a convex set, and hence $f$ is a convex function.

**4.12** Derive the dual norm of the quadratic norm $\|\cdot\|_P : \mathbb{R}^n \to \mathbb{R}_+$ defined as $\|x\|_P = \sqrt{x^T P x}$ for some $P \in \mathbb{S}_{++}^n$.

**Solution** Using the definition of the dual norm, we find that the dual norm of $\|x\| = \|x\|_P$ can be expressed as

$$\|y\|_* = \sup_{\|x\| \leq 1} y^T x = \sup_{\|u\|_2 \leq 1} y^T P^{-1/2} u = \|P^{-1/2} y\|_2 = \|y\|_{P^{-1}}.$$