Homework 1

Benjamin Noland

1. (Boyd & Vandenberghe, Exercise 2.2) Let $C \subseteq \mathbb{R}^n$ be a convex set, and let $L \subseteq \mathbb{R}^n$ be a line. Then L is also a convex set. Since set intersection preserves convexity, $L \cap C$ is therefore a convex set.

   Conversely, let $C \subset \mathbb{R}^n$ be an arbitrary set, and suppose that for every line $L \subseteq \mathbb{R}^n$, $L \cap C$ is convex. Let $x_1, x_2 \in C$ and $0 \leq \theta \leq 1$. Let

   $$L = \{\alpha x_1 + (1 - \alpha)x_2 : \alpha \in \mathbb{R}\}.$$

   Then $L \cap C$ is convex by assumption. Thus, since $x_1, x_2 \in L \cap C$, $\theta x_1 + (1-\theta)x_2 \in L \cap C$, and so $\theta x_1 + (1 - \theta)x_2 \in C$ in particular. Thus C is convex.

   The corresponding result for affine sets requires a short lemma: if $S_1, S_2 \subseteq \mathbb{R}^n$ are affine sets, then $S_1 \cap S_2$ is affine. To prove this, let $x_1, x_2 \in S_1 \cap S_2$, and let $\theta \in \mathbb{R}$. Then $x_1, x_2 \in S_i$ $(i = 1, 2)$ in particular. Thus, since each of the sets S_1 and S_2 are affine, $\theta x_1 + (1 - \theta)x_2 \in S_i$ $(i = 1, 2)$. Thus $\theta x_1 + (1 - \theta)x_2 \in S_1 \cap S_2$, so that $S_1 \cap S_2$ is affine.

   Now for the main result. Let $C \subseteq \mathbb{R}^n$ be an affine set, and let $L \subseteq \mathbb{R}^n$ be a line. Then L is also an affine set. Thus $L \cap C$ is affine by the lemma above.

   Conversely, let $C \subseteq \mathbb{R}^n$ be an arbitrary set, and suppose that for every line $L \subseteq \mathbb{R}^n$, $L \cap C$ is affine. Let $x_1, x_2 \in C$ and $\theta \in \mathbb{R}$. Let

   $$L = \{\alpha x_1 + (1 - \alpha)x_2 : \alpha \in \mathbb{R}\}$$

   Then $L \cap C$ is affine by assumption. Thus since $x_1, x_2 \in L \cap C$, $\theta x_1 + (1-\theta)x_2 \in L \cap C$. In particular, $\theta x_1 + (1 - \theta)x_2 \in C$, so that C is affine.

2. (Boyd & Vandenberghe, Exercise 2.3) Let $C \subseteq \mathbb{R}^n$ be closed and midpoint convex. Let $x_1, x_2 \in C$ and $0 \leq \theta \leq 1$. I claim that for every $n \in \mathbb{N}$, the set

   $$P_n = \left\{\frac{m}{2^n}x_1 + \left(1 - \frac{m}{2^n}\right) : m \in \mathbb{Z}, 0 \leq m \leq 2^n\right\}$$

   is contained in C. I will proceed by induction on n. When $n = 1$, we have

   $$P_1 = \left\{x_1, \frac{1}{2}(x_1 + x_2), x_2\right\} \subseteq C,$$

since $(x_1 + x_2)/2 \in C$ because C is midpoint convex. This establishes the base case. Now let $n > 1$ and suppose that $P_n \subseteq C$. Let $k \in \mathbb{Z}$, with $0 \le k \le 2^{n+1}$. If k is of the form $k = 2m$ for some $m \in \mathbb{Z}$, $0 \le m \le 2^n$, then

$$\frac{k}{2^{n+1}} x_1 + \left(1 - \frac{k}{2^{n+1}}\right) x_2 = \frac{m}{2^n} x_1 + \left(1 - \frac{m}{2^n}\right) x_2 \in P_n.$$

So suppose k is of the form $k = 2m + 1$ for some $m \in \mathbb{Z}$, $0 \le m < 2^n$. Then by midpoint convexity of C,

$$\begin{aligned} C \ni \frac{1}{2}\left[\frac{m}{2^n} x_1 + \left(1 - \frac{m}{2^n}\right) x_2\right] &+ \frac{1}{2}\left[\frac{m+1}{2^n} x_1 + \left(1 - \frac{m+1}{2^n}\right) x_2\right] \\ &= \frac{2m+1}{2^{n+1}} x_1 + \left(1 - \frac{2m+1}{2^{n+1}}\right) x_2 \\ &= \frac{k}{2^{n+1}} x_1 + \left(1 - \frac{k}{2^{n+1}}\right) x_2. \end{aligned}$$

Thus $P_{n+1} \subseteq C$. Hence $P_n \subseteq C$ for every $n \in \mathbb{N}$ by induction.

Next, define a sequence $\{y_n\}$ in C as follows:

$$y_n = \frac{m}{2^n} x_1 + \left(1 - \frac{m}{2^n}\right) x_2 \quad \text{for every } n \in \mathbb{N}$$

where

$$m = \max\left\{k \in \mathbb{Z} : 0 \le k < 2^n, \frac{k}{2^n} \le \theta\right\}.$$

Thus, for every $n \in \mathbb{N}$,

$$\frac{m}{2^n} \le \theta \le \frac{m+1}{2^n},$$

or equivalently (by rearranging this inequality),

$$\left|\theta - \frac{m}{2^n}\right| \le \frac{1}{2^n}.$$

Now let $y = \theta x_1 + (1 - \theta) x_2$. I claim that $y_n \to y$ as $n \to \infty$. We have

$$\begin{aligned} \|y_n - y\|_2 &= \left\| \left[\frac{m}{2^n} x_1 + \left(1 - \frac{m}{2^n}\right) x_2\right] - [\theta x_1 + (1 - \theta) x_2] \right\|_2 \\ &= \left\| \left(\frac{m}{2^n} - \theta\right) x_1 - \left(\frac{m}{2^n} - \theta\right) x_2 \right\|_2 \\ &= \left|\frac{m}{2^n} - \theta\right| \|x_1 - x_2\|_2 \\ &\le \frac{1}{2^n} \|x_1 - x_2\|_2 \to 0 \text{ as } n \to \infty. \end{aligned}$$

Thus $\|y_n - y\|_2 \to 0$ as $n \to \infty$, so that $y_n \to y$ as $n \to \infty$, as claimed. Thus, since C is closed, $y \in C$. This shows that C is convex.

3. (Boyd & Vandenberghe, Exercise 2.10)

   (a) Let $x, v \in \mathbb{R}^n$ and define the line

$$L = \{x + \theta v : \theta \in \mathbb{R}\}.$$

It suffices to show that $L \cap C$ is convex. Let $\theta \in \mathbb{R}$. Then

$$\begin{aligned}(x{+}\theta v)^T & A(x + \theta v) + b^T(x + \theta v) + c \\ &= x^T Ax + 2\theta x^T Av + \theta^2 v^T Av + b^T x + \theta b^T v + c \\ &= \alpha\theta^2 + \beta\theta + \gamma,\end{aligned}$$

where

$$\begin{aligned}\alpha &= v^T Av \\ \beta &= 2x^T Av + b^T v \\ \gamma &= x^T Ax + b^T x + c.\end{aligned}$$

Thus we can write

$$L \cap C = \{x + \theta v : \theta \in \mathbb{R}, \alpha\theta^2 + \beta\theta + \gamma \leq 0\}.$$

Next, define the set

$$S = \{\theta \in \mathbb{R} : \alpha\theta^2 + \beta\theta + \gamma \leq 0\}.$$

Since $A \succeq 0$, $\alpha = v^T Av \geq 0$, and so $\alpha\theta^2 + \beta\theta + \gamma$ is a (possibly degenerate) upward-curving parabola. If the parabola has no real roots, then $S = \emptyset$, which is trivially convex. If $\alpha \neq 0$, then it has roots $\theta_1 \leq \theta_2$ (not necessarily distinct), and $S = [\theta_1, \theta_2]$, which is a convex set. If $\alpha = 0$ (i.e., the degenerate case), then S reduces to

$$S = \{\theta \in \mathbb{R} : \beta\theta + \gamma \in (-\infty, 0]\},$$

i.e., the preimage of the convex set $(-\infty, 0]$ under the affine function $\theta \mapsto \beta\theta + \gamma$, and thus is convex.

Finally, define an affine function $f : \mathbb{R} \to \mathbb{R}^n$ by $f(\theta) = x + \theta v$. Then we can write $L \cap C = f(S)$. This shows that $L \cap C$ is convex, and therefore C is convex.

The converse, however, does not hold. To see this, consider the set

$$C = \{x \in \mathbb{R} : -x^2 - 1 \leq 0\} = \{x \in \mathbb{R} : x^2 + 1 \geq 0\} = \mathbb{R},$$

which is convex, but $A = -1 < 0$.

4. (Boyd & Vandenberghe, Exercise 2.12)

(a) A slab can be expressed as the intersection of two halfspaces:

$$\{x \in \mathbb{R}^n : \alpha \leq a^T x \leq \beta\} = \{x \in \mathbb{R}^n : a^T x \geq \alpha\} \cap \{x \in \mathbb{R}^n : a^T x \leq \beta\}.$$

Thus, since halfspaces are convex and set intersection preserves convexity, a slab is convex.

(b) Let

$$R = \{x \in \mathbb{R}^n : \alpha_i \leq x_i \leq \beta_i, 1 \leq i \leq n\}$$

denote a rectangle. Let $x, y \in R$ and $0 \leq \theta \leq 1$. Then for every $1 \leq i \leq n$,

$$\begin{aligned} \theta x_i + (1-\theta) y_i &\geq \theta \alpha_i + (1-\theta)\alpha_i = \alpha_i \\ \theta x_i + (1-\theta) y_i &\leq \theta \beta_i + (1-\theta)\beta_i = \beta_i. \end{aligned}$$

Thus $\theta x + (1-\theta) y \in R$, so that R is convex.

(c) A wedge can be expressed as the intersection of two halfspaces:

$$\{x \in \mathbb{R}^n : a_1^T x \leq b_1, a_2^T x \leq b_2\} = \{x \in \mathbb{R}^n : a_1^T x \leq b_1\} \cap \{x \in \mathbb{R}^n : a_2^T x \leq b_2\}$$

Thus, since halfspaces are convex and set intersection preserves convexity, a wedge is convex.

(d) This set can be expressed as the intersection

$$C = \{x \in \mathbb{R}^n : \|x - x_0\|_2 \leq \|x - y\|_2 \text{ for every } y \in S\} = \bigcap_{y \in S} C_y,$$

where for every $y \in S$, the set C_y is defined as

$$C_y = \{x \in \mathbb{R}^n : \|x - x_0\|_2 \leq \|x - y\|_2\}.$$

Let $y \in S$. Then $x \in C_y$ if and only if

$$(x - x_0)^T (x - x_0) \leq (x - y)^T (x - y),$$

or equivalently (after some algebra),

$$(2y - x_0)^T x \leq y^T y - x_0^T x_0.$$

Thus

$$C_y = \{x \in \mathbb{R}^n : (2y - x_0)^T x \leq y^T y - x_0^T x_0\}.$$

If $2y - x_0 \neq 0$, then C_y is a halfspace, hence convex. If $2y - x_0 = 0$, then $x_0^T x_0 = 4y^T y$, and so $C_y = \emptyset$ if $y = 0$, and $C_y = \mathbb{R}$ if $y \neq 0$. In either case, C_y is convex.

Therefore, since C_y is convex for every $y \in S$, the intersection $C = \bigcap_{y \in S} C_y$ is also convex.

5. (Boyd & Vandenberghe, Exercise 2.21) Let $A \subseteq \mathbb{R}^{n+1}$ denote the set in question. Let $(a_1, b_1), (a_2, b_2) \in A$, and let $\theta_1, \theta_2 \geq 0$. Then

$$\theta_1(a_1, b_1) + \theta_2(a_2, b_2) = (\theta_1 a_1 + \theta_2 a_2, \theta_1 b_1 + \theta b_2).$$

Thus, for every $x \in C$,

$$(\theta_1 a_1 + \theta_2 a_2)^T x = \theta_1 a_1^T x + \theta_2 a_2^T x \leq \theta_1 b_1 + \theta_2 b_2,$$

and for every $x \in D$,

$$(\theta_1 a_1 + \theta_2 a_2)^T x = \theta_1 a_1^T x + \theta_2 a_2^T x \geq \theta_1 b_1 + \theta_2 b_2.$$

Thus $\theta_1(a_1, b_1) + \theta_2(a_2, b_2) \in A$, so that A is a convex cone. In particular, if there is no hyperplane separating C and D, then for every $(a, b) \in \mathbb{R}^{n+1}$ with

$$\begin{aligned} a^T x &\leq b \quad \text{for every } x \in C \\ a^T x &\geq b \quad \text{for every } x \in D \end{aligned}$$

we must have $a = 0$, since otherwise $\{x \in \mathbb{R}^n : a^T x = b\}$ would be a hyperplane separating C and D. Hence $0 \leq b$ and $0 \geq b$, so that $b = 0$. It therefore follows that A is the singleton $A = \{0\}$ in this case.

6. (Convexity of $x \mapsto e^{ax}$ on $\mathbb{R}$) First, note that the domain $\mathbb{R}$ is convex. In addition, since $x \mapsto e^{ax}$ is twice differentiable, we can compute

$$\frac{d^2}{dx^2} e^{ax} = a^2 e^{ax} \geq 0 \quad \text{for any } x \in \mathbb{R}.$$

Thus $x \mapsto e^{ax}$ is convex.

7. (Concavity of $x \mapsto \log x$ on $\mathbb{R}_{++}$) First, note that the domain $\mathbb{R}_{++} = (0, \infty)$, being an interval in $\mathbb{R}$, is convex. In addition, since $x \mapsto \log x$ is twice differentiable, we can compute

$$\frac{d^2}{dx^2} \log x = -\frac{1}{x^2} \leq 0 \quad \text{for any } x \in \mathbb{R}_{++}.$$

Thus $x \mapsto \log x$ is concave.

8. (Convexity of $X \mapsto -\log\det X$ on S^n_{++}) Let $f : S^n \to \mathbb{R}$, with $\operatorname{dom} f = S^n_{++}$, be defined by $f(X) = \log\det X$. Let L be a line with $L \cap S^n_{++} \neq \emptyset$, say

$$L = \{Z + tV : t \in \mathbb{R}\},$$

where $Z, V \in S^n$. Let $g : \mathbb{R} \to \mathbb{R}$, with

$$\operatorname{dom} g = L \cap S^n_{++} = \{t \in \mathbb{R} : Z + tV \succ 0\},$$

denote the restriction of f to L; that is, $g(t) = f(Z + tV)$. It suffices to show that g is a convex function. We can assume without loss of generality that $Z \succ 0$, since because $\operatorname{dom} g \neq \emptyset$, we can always choose a basepoint $Z' = Z + t'V \in S^n_{++}$ for some $t' \in \operatorname{dom} g$. Note that $\operatorname{dom} g$ is simply the preimage of S^n_{++} under the continuous affine function $t \mapsto Z + tV$. Thus, since S^n_{++} is an open subset of S^n, $\operatorname{dom} g$ is an open, convex subset of $\mathbb{R}$, hence an open interval.

Since $Z \succ 0$, Z has a unique matrix square root $Z^{1/2}$. Thus we can write

$$\begin{aligned}
g(t) &= \log\det(Z + tV) \\
&= \log\det[Z^{1/2}(I + tZ^{-1/2}VZ^{-1/2})Z^{1/2}] \\
&= \log[\det(Z^{1/2})\det(I + tZ^{-1/2}VZ^{-1/2})\det(Z^{1/2})] \\
&= \log[\det(Z)\det(I + tZ^{-1/2}VZ^{-1/2})] \\
&= \log\det Z + \log\det(I + tZ^{-1/2}VZ^{-1/2}).
\end{aligned}$$

Note that

$$\begin{aligned}
\det(I + tZ^{-1/2}VZ^{-1/2}) &= t^n \det\left(\frac{1}{t} - (-Z^{-1/2}VZ^{-1/2})\right) \\
&= t^n \prod_{i=1}^{n}\left(\frac{1}{t} - (-\lambda_i)\right) \\
&= \prod_{i=1}^{n}(1 + t\lambda_i),
\end{aligned}$$

where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of $Z^{-1/2}VZ^{-1/2}$. Thus,

$$\begin{aligned}
g(t) &= \log\det Z + \log\det(I + tZ^{-1/2}VZ^{-1/2}) \\
&= \log\det Z + \log\prod_{i=1}^{n}(1 + t\lambda_i) \\
&= \log\det Z + \sum_{i=1}^{n}\log(1 + t\lambda_i).
\end{aligned}$$

Taking derivates, we find that

$$g'(t) = \sum_{i=1}^{n}\frac{\lambda_i}{1 + t\lambda_i} \quad \text{and} \quad g''(t) = \sum_{i=1}^{n} -\frac{\lambda_i^2}{(1 + t\lambda_i)^2} \leq 0$$

for every $t \in \operatorname{dom} g$. Since $\operatorname{dom} g$ is open, we therefore conclude that g, and hence f, is concave. In particular, $-\log\det X$ is convex on S^n_{++}.

9. (Boyd & Vandenberghe, Exercise 3.11) Since f is convex and differentiable, we have

$$\begin{aligned}
f(y) &\geq f(x) + \nabla f(x)^T(y - x) \\
f(x) &\geq f(y) + \nabla f(y)^T(x - y)
\end{aligned}$$

for every $x, y \in \operatorname{dom} f$. Thus, negating both sides of the second inequality, we find

$$-f(x) \leq -f(y) + \nabla f(y)^T(y - x),$$

and adding this to the first inequality above, we get

$$\nabla f(y)^T(y - x) \geq \nabla f(x)^T(y - x),$$

or equivalently,

$$(\nabla f(y) - \nabla f(x))^T(y - x) \geq 0.$$

Therefore ∇f is monotone.

The converse, however, does not hold. As an example, consider the function $\psi : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$\psi(x) = \begin{pmatrix} x_1 \\ x_1 + x_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

Note that for any $x \in \mathbb{R}^2$,

$$\begin{aligned} x^T \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} x &= x_1^2 + x_2^2 + x_1 x_2 \\ &= x_1^2 + x_2^2 + \frac{1}{2}x_1 x_2 + \frac{1}{2}x_1 x_2 \\ &= x^T \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix} x \geq 0, \end{aligned}$$

since the matrix in the last inequality is positive definite (this can be easily verified using Sylvester's criterion). It follows immediately that ψ is monotone, since for any $x, y \in \mathbb{R}^2$,

$$(\psi(x) - \psi(y))^T(x - y) = (x - y)^T \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix} (x - y) \geq 0.$$

But if we had $\psi = \nabla f$ for some differentiable function $f : \mathbb{R}^2 \to \mathbb{R}$, we would necessarily have

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial^2 f}{\partial x_2 \partial x_1} = 0 \quad \text{and} \quad \frac{\partial^2 f}{\partial x_1 \partial x_2} = 1,$$

which is impossible.

10. (Boyd & Vandenberghe, Exercise 3.18)

  (a) As in problem 8, let $g : \mathbb{R} \to \mathbb{R}$ denote the restriction of f to the line

  $$L = \{Z + tV : t \in \mathbb{R}\},$$

  where as before we can assume without loss of generality that $Z \succ 0$. Thus Z has a unique matrix square root $Z^{1/2}$. Furthermore, the matrix $Z^{-1/2}VZ^{-1/2}$ is real symmetric, and can therefore be written as $Z^{-1/2}VZ^{-1/2} = Q^T\Lambda Q$, where Q is an orthogonal matrix and $\Lambda = \mathrm{diag}\{\lambda_1, \ldots, \lambda_n\}$, where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $Z^{-1/2}VZ^{-1/2}$. We can therefore write

$$\begin{aligned}
g(t) &= f(Z+tV) \\
&= \mathrm{Tr}[(Z+tV)^{-1}] \\
&= \mathrm{Tr}[Z^{-1/2}(I+tZ^{-1/2}VZ^{-1/2})^{-1}Z^{-1/2}] \\
&= \mathrm{Tr}[Z^{-1}(I+tZ^{-1/2}VZ^{-1/2})^{-1}] \\
&= \mathrm{Tr}[Z^{-1}(I+tQ^T\Lambda Q)^{-1}] \\
&= \mathrm{Tr}[Z^{-1}(Q^T+t\Lambda Q)^{-1}Q] \\
&= \mathrm{Tr}[Z^{-1}Q^T(I+t\Lambda)^{-1}Q] \\
&= \mathrm{Tr}[QZ^{-1}Q^T(I+t\Lambda)^{-1}] \\
&= \sum_{i=1}^{n}[QZ^{-1}Q^T]_{ii}\frac{1}{1+t\lambda_i}.
\end{aligned}$$

Each of the functions $t \mapsto 1/(1+t\lambda_i)$ is convex on $\mathrm{dom}\, g$, and it therefore follows that g itself is convex. Thus f is convex.

(b) As in problem 8, let $g : \mathbb{R} \to \mathbb{R}$ denote the restriction of f to the line

$$L = \{Z + tV : t \in \mathbb{R}\},$$

where as before we can assume without loss of generality that $Z \succ 0$. Thus Z has a unique matrix square root $Z^{1/2}$. We can therefore write

$$\begin{aligned}
g(t) &= f(Z+tV) \\
&= [\det(Z+tV)]^{1/n} \\
&= [\det[Z^{1/2}(I+tZ^{-1/2}VZ^{-1/2})Z^{1/2}]]^{1/n} \\
&= [\det Z \det(I+tZ^{-1/2}VZ^{-1/2})]^{1/n} \\
&= (\det Z)^{1/n}\left[\prod_{i=1}^{n}(1+t\lambda_i)\right]^{1/n},
\end{aligned}$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $Z^{-1/2}VZ^{-1/2}$. Since $Z \succ 0$, $\det Z > 0$. In addition, the geometric mean function $x \mapsto (\prod_{i=1}^{n} x_i)^{1/n}$ is concave on $\mathbb{R}^n_{++}$. It therefore follows that g is concave. Thus f is concave as well.