1 Convex sets

1 (2.1)**.** Let $C \subseteq \mathbb{R}^n$ be a convex set, with $x_1, \dots, x_k \in C$, and let $\theta_1, \dots, \theta_k \in \mathbb{R}$ satisfy $\theta_i \geq 0$, $\theta_1 + \cdots + \theta_k = 1$. Show that $\theta_1 x_1 + \cdots + \theta_k x_k \in C$.(The definition of convexity is that this holds for $k = 2$; you must show it for arbitrary k.) *Hint.* Use induction on k.

Proof.

- When $k = 2$, $\theta_i \geq 0, \theta_1 + \theta_2 = 1 \Longrightarrow \theta_1 x_1 + \theta_2 x_2 = \theta_1 x_1 + (1 - \theta_1) x_2 \in C$.
- If $k = n$, $\theta_i \geq 0, \theta_1 + \cdots + \theta_n = 1 \Longrightarrow \theta_1 x_1 + \cdots + \theta_n x_n \in C$ holds.
- Then $k = n + 1$, $\theta_i \geq 0, \theta_1 + \cdots + \theta_{n+1} = 1 \Longrightarrow \theta_1 x_1 + \cdots + \theta_n x_n + \theta_{n+1} x_{n+1} = (\theta_1 + \cdots + \theta_n) \frac{\theta_1 x_1 + \cdots + \theta_n x_n}{\theta_1 + \cdots + \theta_n} + \theta_{n+1} x_{n+1}$. $k = n, \theta_1 x_1 + \cdots + \theta_n x_n \in C$ holds, and $k = 2$ holds, so $\theta_1 x_1 + \cdots + \theta_n x_n + \theta_{n+1} x_{n+1} \in C$.
- so $\theta_1 x_1 + \cdots + \theta_k x_k \in C$ for arbitrary k.

□

2 (2.5)**.** What is the distance between two parallel hyperplanes $\left\{x \in \mathbb{R}^n | a^T x = b_1\right\}$ and $\left\{x \in \mathbb{R}^n |\ a^T x = b_2\right\}$?

Answer. The distance between the two hyperplanes is $\frac{|b_1 - b_2|}{\|a\|_2}$.

3 (2.11)**.** *Hyperbolic sets.* Show that the *Hyperbolic set* $\left\{x \in \mathbb{R}^2_+ | x_1 x_2 \geq 1\right\}$ is convex. As a generalization, show that $\left\{x \in \mathbb{R}^n_+ | \prod_{i=1}^n x_i \geq 1\right\}$ is convex. *Hint.* If $a, b \geq 0$ and $0 \leq \theta \leq 1$, then $a^\theta b^{1-\theta} \leq \theta a + (1 - \theta) b$.

Answer. (a) $x, y \in C$, then $z = \theta x + (1 - \theta) y$.

$$\begin{aligned} z_1 z_2 &= (\theta x_1 + (1-\theta) y_1)(\theta x_2 + (1-\theta) y_2) \\ &\geq x_1^\theta y_1^{1-\theta} \cdot x_2^\theta y_2^{1-\theta} \\ &= (x_1 x_2)^\theta (y_1 y_2)^{1-\theta} \\ &\geq 1 \end{aligned}$$

we get $z \in C$ and $\left\{x \in \mathbb{R}^2_+ | x_1 x_2 \geq 1\right\}$ is convex.

(b) $x, y \in C$, then $z = \theta x + (1 - \theta) y$.

$$\begin{aligned} z_1 z_2 &= \prod_{i=1}^n (\theta x_i + (1-\theta) y_i) \\ &\geq \prod_{i=1}^n x_i^\theta y_i^{1-\theta} \\ &\geq 1 \end{aligned}$$

we get $z \in C$ and $\left\{x \in \mathbb{R}_+^n \middle| \prod_{i=1}^n x_i \geq 1\right\}$ is convex.

4 (2.14)**.** *Erpanded and restricted sets.* Let $S \subseteq \mathbb{R}^n$, and let $\|\cdot\|$ be a norm on $\mathbb{R}^n$.

(a) For $a \geq 0$ we define S_a as $\{x \mid \operatorname{dist}(x, S) \leq a\}$, where $\operatorname{dist}(x, S) = \inf_{y \in S} \|x - y\|$. We refer to S_a as S *expanded* or *extended* by a. Show that if S is convex, then S_a is convex.

(b) For $a \geq 0$ we define $S_{-a} = \{x | B(x, a) \subseteq S\}$, where $B(x, a)$ is the ball (in the norm $\|\cdot\|$), centered at x, with radius a. We refer to S_{-a} as S *shrunk* or *restricted* by a, since S_{-a} consists of all points that are at least a distance a from $\mathbb{R}^n \backslash S$. Show that if S is convex, then S_{-a} is convex.

Proof. (a) $\forall x_1, x_2 \in S_a$, for $0 \leq \theta \leq 1$, $z = \theta x_1 + (1-\theta) x_2$

$$\begin{aligned}
\operatorname{dist}(z, S) &= \inf_{y \in S} \|z - y\| \\
&= \inf_{y_1, y_2 \in S} \|\theta x_1 + (1-\theta) x_2 - \theta y_1 - (1-\theta) y_2\| \\
&\leq \inf_{y_1, y_2 \in S} (\theta \|x_1 - y_1\| + (1-\theta) \|x_2 - y_2\|) \\
&= \theta \inf_{y_1 \in S} \|x_1 - y_1\| + (1-\theta) \inf_{y_2 \in S} \|x_2 - y_2\| \\
&\leq a
\end{aligned}$$

so $\forall x_1, x_2 \in S_a, z \in S_a$, S_a is convex.

(b) Consider $x_1, x_2 \in S_{-a}$, $\forall u$ with $\|u\| \leq a$,

$$x_1 + u \in S, \quad x_2 + u \in S$$

$\forall \theta \in [0, 1], \|u\| \leq a$,

$$z + u = \theta x_1 + (1-\theta) x_2 + u = \theta(x_1 + u) + (1-\theta)(x_2 + u) \in S$$

because S is convex. We conclude that $z \in S_{-a}$.

□

2 Convex functions

5 (3.1)**.** Suppose $f : \mathbb{R} \to \mathbb{R}$ is convex, and $a, b \in \operatorname{dom}(f)$ with $a < b$.

(a) Show that

$$f(x) \leq \frac{b-x}{b-a} f(a) + \frac{x-a}{b-a} f(b)$$

for all $x \in [a, b]$.

(b) Show that
$$\frac{f(x)-f(a)}{x-a} \le \frac{f(b)-f(a)}{b-a} \le \frac{f(b)-f(x)}{b-x}$$
for all $x \in (a,b)$. Draw a sketch that illustrates this inequality.

(c) Suppose f is differentiable. Use the result in (b) to show that
$$f'(a) \le \frac{f(b)-f(a)}{b-a} \le f'(b)$$
Note that these inequalities also follow from:
$$f(b) \ge f(a) + f'(a)(b-a)$$

(d) Suppose f is twice differentiable. Use the result in (c) to show that $f''(a) \ge 0$ and $f''(b) \ge 0$.

Proof.

(a) f is convex, so $f(\theta x_1 + (1-\theta)x_2) \le \theta f(x_1) + (1-\theta) f(x_2)$. When $x = \theta x_1 + (1-\theta)x_2$, $a = x_1, b = x_2$, we get $\theta = \frac{x_2 - x}{x_2 - x_1}$, so
$$f(x) \le \frac{b-x}{b-a} f(a) + \frac{x-a}{b-a} f(b)$$
for all $x \in [a,b]$

(b)
$$\begin{aligned}
f(x) &\le \frac{b-x}{b-a} f(a) + \frac{x-a}{b-a} f(b) \\
f(x) - f(a) &\le \frac{b-x}{b-a} f(a) + \frac{x-a}{b-a} f(b) - f(a) \\
\frac{f(x)-f(a)}{x-a} &\le \frac{f(b)-f(a)}{b-a}
\end{aligned}$$
So the left inequality holds. The inequality on the right is the same.
$$\frac{f(x)-f(a)}{x-a} \le \frac{f(b)-f(a)}{b-a} \le \frac{f(b)-f(x)}{b-x}$$
Geometrically, in figure 1 the inequalities mean that $k_{ax} < k_{ab} < k_{xb}$, k_{ab} means the slope of the line segment between $(a, f(a))$ and $(b, f(b))$.

□

6 (3.7)**.** Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is convex with $\mathrm{dom}(f) = \mathbb{R}^n$. and bounded above on $\mathbb{R}^n$. Show that f is constant.

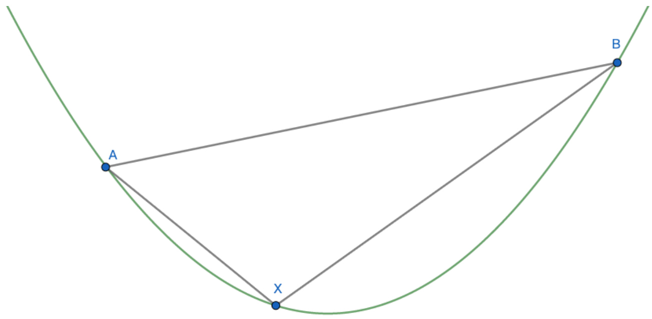

Figure 1: sketch that illustrates this inequality

Answer. Suppose f is not constant. $\exists x, y$, s.t. $f(x) < f(y)$.

$$g(t) = f(tx + (1-t)y)$$

is convex, $g(0) = f(y) > f(x) = g(1)$. We get

$$g(0) \leq \frac{t-1}{t} g(1) + \frac{1}{t} g(t)$$

for all $t > 1$, and

$$g(t) \geq tg(0) - (t-1)g(1) = g(1) + t(g(0) - g(1))$$

so g grows unboundedly as $t \to \infty$. This contradicts our assumption that f is bounded. So f is constant.

7 (3.16)**.** For each of the following functions determine whether it is convex, concave, quasiconvex, or quasiconcave.(consider only convexity and concavity)

(a) $f(x) = e^x - 1$ on $\mathbb{R}$.

(b) $f(x_1, x_2) = x_1 x_2$ on $\mathbb{R}^2_{++}$.

(c) $f(x_1, x_2) = 1/(x_1 x_2)$ on $\mathbb{R}^2_{++}$.

(d) $f(x_1, x_2) = x_1/x_2$ on $\mathbb{R}^2_{++}$.

(e) $f(x_1, x_2) = x_1^2/x_2$ on $\mathbb{R} \times \mathbb{R}_{++}$.

(f) $f(x_1, x_2) = x_1^\alpha x_2^{1-\alpha}$, where $0 \leq \alpha \leq 1$, on $\mathbb{R}^2_{++}$.

Answer.

(a) $f''(x) = e^x > 0$, so f is convex but not concave.

(b) $\nabla^2 f = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is neither positive semidefinite nor negative semidefinite, so f is neither convex nor concave.

(c) $\nabla^2 f = \frac{1}{x_1^3 x_2^3}\begin{pmatrix} 2x_2^2 & x_1x_2 \\ x_1x_2 & 2x_1^2 \end{pmatrix} \succeq 0$, so f is convex but not concave.

(d) $\nabla^2 f = \frac{1}{x_2^3}\begin{pmatrix} 0 & -x_2 \\ -x_2 & 2x_1 \end{pmatrix}$ is neither positive semidefinite nor negative semidefinite, so f is neither convex nor concave.

(e) $\nabla^2 f = \frac{2}{x_2^3}\begin{pmatrix} x_2^2 & -x_1x_2 \\ -x_1x_2 & x_1^2 \end{pmatrix} \succeq 0$, so f is convex but not concave.

(f)

$$\begin{aligned}
\nabla^2 f &= \alpha(1-\alpha)\begin{pmatrix} x_1^{\alpha-2}x_2^{1-\alpha} & -x_1^{\alpha-1}x_2^{-\alpha} \\ -x_1^{\alpha-1}x_2^{-\alpha} & -x_1^{\alpha}x_2^{-\alpha-1} \end{pmatrix} \\
&= \alpha(1-\alpha)x_1^{\alpha}x_2^{1-\alpha}\begin{pmatrix} -1/x_1^2 & 1/x_1x_2 \\ 1\ x_1x_2 & -1/x_2^2 \end{pmatrix} \\
&= -\alpha(1-\alpha)x_1^{\alpha}x_2^{1-\alpha}\begin{pmatrix} 1/x_1 \\ -1/x_2 \end{pmatrix}\begin{pmatrix} 1/x_1 & -1/x_2 \end{pmatrix} \\
&\preceq 0
\end{aligned}$$

so f is concave but not convex.

8 (3.18)**.** Adapt the proof of concavity of the log-determinant function in §3.1.5 to show the following.

(a) $f(X) = \operatorname{tr}(X^{-1})$ is convex on $\operatorname{dom}(f) = \mathbb{R}^n_{++}$.

(b) $f(X) = (\det X)^{1/n}$ is concave on $\operatorname{dom}(f) = \mathbb{R}^n_{++}$.

Answer.

(a) Define $g(t) = f(Z + tV)$, where $Z \succ 0$ and $V \in \mathbb{S}^n$.

$$\begin{aligned}
g(t) &= \operatorname{tr}\big((Z+tV)^{-1}\big) \\
&= \operatorname{tr}\Big(Z^{-1}(I + tZ^{-\frac{1}{2}}VZ^{-\frac{1}{2}})^{-1}\Big) \\
&= \operatorname{tr}\big(Z^{-1}Q(I+t\Lambda)^{-1}Q^t\big) \\
&= \operatorname{tr}\big(Q^tZ^{-1}Q(I+t\Lambda)^{-1}\big) \\
&= \sum_{i=1}^{n}(Q^tZ^{-1}Q)_{ii}(1+t\lambda_i)^{-1}
\end{aligned}$$

We express g as a positive weighted sum of convex functions $\frac{1}{1+t\lambda_i}$, hence it is convex.

(b) Define $g(t) = f(Z + tV)$, where $Z \succ 0$ and $V \in \mathbb{S}^n$.

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

where λ_i are the eigenvalues of $Z^{-\frac{1}{2}} V Z^{-\frac{1}{2}}$. We see that g is a concave function of t on $\{t \mid Z + tV \succ 0\}$, since $\det(Z) > 0$ and the geometric mean $(\prod_{i=1}^n x_i)^{\frac{1}{n}}$ is concave on $\mathbb{R}^n_{++}$.

9 (3.27)**.** *Diagonal elements of Cholesky factor.* Each $X \in \mathbf{S}^n_{++}$ has a unique *Cholesky factorization* $X = LL^T$, where L is lower triangular, with $L_{ii} > 0$. Show that L_{ii} is a concave function of X (with domain $\mathbf{S}^n_{++}$).

Hint. L_{ii} can be expressed as $L_{ii} = (w - z^t Y^{-1} z)^{1/2}$, where

$$\begin{bmatrix} Y & z \\ z^T & w \end{bmatrix}$$

is the leading $i \times i$ submatrix of X.

Answer. $f(z, Y) = z^t Y^{-1} z$ with $\mathrm{dom}(f) = \{(z, Y) | Y \succ 0\}$ is convex jointly in z and Y. Notice that

$$(z, Y, t) \in \mathrm{epi}(f) \quad \Longleftrightarrow \quad Y \succ 0, \quad \begin{bmatrix} Y & z \\ z^t & T \end{bmatrix} \succeq 0$$

so $\mathrm{epi}(f)$ is a convex set. Therefore, $w - z^t Y^{-1} z$ is a concave function of X. Since the squareroot is an increasing concave function, it follows from the composition rules that $l_{kk} = (w - z^t Y^{-1} z)^{\frac{1}{2}}$ is a concave function of X.

10 (3.31)**.** *Largest homogeneous und restimator.*Let f be a convex function. Define the function g as

$$g(x) = \inf_{\alpha > 0} \frac{f(\alpha x)}{\alpha}$$

(a) Show that g is homogeneous ($g(tx) = tg(x)$ for all $t \geq 0$).

(b) Show that g is the largest homogeneous underestimator of f: If h is homogeneous and $h(x) \leq f(x)$ for all x. then we have $h(x) \leq g(x)$ for all x.

(c) show that g is convex.

Answer.

(a) If $t = 0$, $g(tx) = g(0) = 0 = tg(x)$. If $t > 0$

$$g(tx) = \inf_{\alpha>0} \frac{f(\alpha tx)}{\alpha} = t \inf_{\alpha>0} \frac{f(\alpha tx)}{t\alpha} = tg(x).$$

so $\forall t \geq 0, g(tx) = tg(x)$.

(b) If h is a homogeneous underestimator, then

$$h(x) = \frac{h(\alpha x)}{\alpha} \leq \frac{f(\alpha x)}{\alpha}$$

for all $\alpha > 0$, so $h(x) \leq g(x)$.

(c) We can express g as

$$g(x) = \inf_{t>0} tf(x/t) = \inf_{t>0} h(x,t)$$

where h is the perspective function of f. We know h is convex, so g is convex.

11 (3.36)**.** (to do) Derive the conjugates of the following funetions.

(a) *Max function.* $f(x) = \max_{i=1,\dots,n} x_i$ on $\mathbb{R}^n$.

(b) *Saum of largest elements.* $f(x) = \sum_{i=1}^{r} x_{[i]}$ on $\mathbb{R}^n$.

(c) *Piecewise-linear function on* $\mathbb{R}$. $f(x) = \max_{i=1,\dots,m}(a_i x + b_i)$ on $\mathbb{R}$. You can assume that the a_i are sorted in increasing order, *i.e.*, $a_1 \leq \cdots \leq a_m$, and that none of the functions $a_i x + b_i$: is redundant, *i.e.*, for each k there is at least one x with $f(x) = a_k x + b_k$.

(d) *Power function.* $f(x) = x^p$ on $\mathbb{R}^n_{++}$.

(e) *Negative geometric mean.* $f(x) = -\left(\prod x_i\right)^{1/n}$ on $\mathbb{R}^n_{++}$.

(f) *Negative generalized logarithm for second-order cone.* $f(x,t) = -\log\left(t^2 - x^T x\right)$ on $\{(x,t) \in \mathbb{R}^n \times \mathbb{R} \mid \|x\|_2 < t\}$.

12 (3.37)**.** Show that the conjugate of $f(X) = \operatorname{tr}(X^{-1})$ with $\operatorname{dom}(f) = \mathbb{S}^n_{++}$ is given by

$$f^*(Y) = -2\operatorname{tr}(-Y)^{\frac{1}{2}}, \quad \operatorname{dom}(f^*) = -\mathbb{S}^n_{+}$$

Hint. The gradient of f is $\nabla f(X) = -X^{-2}$

Answer. Suppose Y has eigenvalue decomposition

$$Y = Q\Lambda Q^t = \sum_{i=1}^{n} \lambda_i q_i q_i^t$$

with $\lambda_1 > 0$. Let $X = Q\operatorname{diag}(t, 1, \ldots, 1)Q^t = tq_1q_1^t + \sum_{i=2}^n q_iq_i^t$. We have

$$\operatorname{tr}(XY) - \operatorname{tr}\left(X^{-1}\right) = t\lambda_1 + \sum_{i=2}^{n} \lambda_i - 1/t - (n-1)$$

which grows unboundedly as $t \to \infty$. Therefore $Y \notin \operatorname{dom}(f)^*$.

Next, assume $Y \preceq 0$. If $Y \prec 0$, we can find the maximum of

$$\operatorname{tr}(XY) - \operatorname{tr}\left(X^{-1}\right)$$

by setting the gradient equal to zero. We obtain $Y = -X^{-2}$, i.e., $X = (-Y)^{-1/2}$, and

$$f^*(Y) = -2\operatorname{tr}(-Y)^{1/2}$$

Finally we verify that this expression remains valid when $Y \preceq 0$, but Y is singular. This follows from the fact that conjugate functions are always closed, i.e., have closed epigraphs.

3 Convex optimization problems

13 (4.1)**.** Consider the optimization problem

$$\begin{cases} \text{minimize} & f_0(x_1, x_2) \\ \text{subject to} & 2x_1 + x_2 \geq 1 \\ & x_1 + 3x_2 \geq 1 \\ & x_1 \geq 0, \quad x_2 \geq 0 \end{cases}$$

Make a sketch of the feasible set. For each of the following objective functions, give the optimal set and the optimal value.

(a) $f_0(x_1, x_2) = x_1 + x_2$

(b) $f_0(x_1, x_2) = -x_1 - x_2$

(c) $f_0(x_1, x_2) = x_1$

(d) $f_0(x_1, x_2) = \max\{x_1, x_2\}$

(e) $f_0(x_1, x_2) = x_1^2 + 9x_2^2$

Answer. The feasible set is the convex hull of $(0, +\infty), (0, 1), (\frac{2}{5}, \frac{1}{5}), (1, 0), (+\infty, 0)$.

(a) $x^* = (\frac{2}{5}, \frac{1}{5})$

(b) Unbounded below.

(c) $X = \{(0, x_2) \mid x_2 \geq 1\}$

(d) $x^* = (\frac{1}{3}, \frac{1}{3})$

(e) $x^* = (\frac{1}{2}, \frac{1}{6})$

14 (4.2)**.** Consider the optimization problem

$$\text{minimize} \quad f_0(x) = -\sum_{i=1}^{m} \log\bigl(b_i - a_i^t x\bigr)$$

with domain $\operatorname{dom}(f_0) = \{x | Ax \prec b\}$, where $A \in \mathbb{R}^{m\times n}$(with rows a_i^t). We assume that $\operatorname{dom}(f_0)$ is nonempty.

Prove the following facts(which include the results quoted without proof on page 141).

(a) $\operatorname{dom}(f_0)$ is unbounded iff there exists a $v \neq 0$ with $Av \preceq 0$.

(b) f_0 is unbounded below iff there exists a v with $Av \preceq 0, Av \neq 0$.*Hint.*There exists a v such that $Av \preceq 0, Av \neq 0$ iff there exists no $z \succ 0$ such that $A^t z = 0$. This follows from the theorem of alternatives in example 2.21, page 50.

(c) If f_0 is bounded below then its minimum is attained, i.e., there exists an x that satisfies the optimality condition(4.23).

(d) The optimal set is affine: $X_{opt} = \{x^* + v \mid Av = 0\}$, where x^* is any optimal point.

Answer.

(a) If such a v exists, $\operatorname{dom}(f_0)$ is unbounded, since $x_0 + tv \in \operatorname{dom}(f_0)$ for all $t \geq 0$. Conversely, suppose x^k is a sequence of points in $\operatorname{dom}(f_0)$ with $\|x^k\|_2 \to \infty$. Define $v^k = x^k/\|x^k\|_2$. The sequence has a convergent subsequence because $\|v^k\|_2 = 1$ for all k. Let v be its limit. We have $\|v\|_2 = 1$ and, since $a_i^t v^k < b_i/\|x^k\|_2$ for all k, $a_i^t v \leq 0$. Therefore $Av \preceq 0$ and $v \neq 0$.

(b) If there exists such a v, then f_0 is unbounded below. Let j be an index with $a_j^t v < 0$. For $t \geq 0$,

$$\begin{aligned} f_0(x_0 + tv) &= -\sum_{i=1}^{m} \log\bigl(b_i - a_i^t x_0 - t a_i^t v\bigr) \\ &\leq -\sum_{i\neq j}^{m} \log\bigl(b_i - a_i^t x_0\bigr) - \log\bigl(b_j - a_j^t x_0 - t a_j^t v\bigr) \end{aligned}$$

and the righthand side decreases without bound as t increases.

Conversely, suppose f is unbounded below. Let x^k be a sequence with $b - Ax^k \succ 0$, and $f_0(x^k) \to -\infty$. By convexity,

$$f_0(x^k) \geq f_0(x_0) + \sum_{i=1}^{m} \frac{1}{b_i - a_i^t x_0} a_i^t (x^k - x_0) = f_0(x_0) + m - \sum_{i=1}^{m} \frac{b_i - a_i^t x^k}{b_i - a_i^t x_0}$$

so if $f_0(x^k) \to -\infty$, we must have $\max_i (b_i - a_i^t x^k) \to \infty$.

Suppose there exists a z with $z \succ 0$, $A^t z = 0$. Then

$$z^t b = z^t (b - Ax^k) \geq z_i \max_i (b_i - a_i^t x^k) \to \infty$$

We have reached a contradiction, and conclude that there is no such z. Using the theorem of alternatives, there must be a v with $Av \preceq 0, Av \neq 0$.

(c) We can assume that $\operatorname{rank}(A) = n$.

If $\operatorname{dom}(f_0)$ is bounded, then the result follows from the fact that the sublevel sets of f_0 are closed.

If $\operatorname{dom}(f_0)$ is unbounded, let v be a direction in which it is unbounded, i.e., $v \neq 0, Av \preceq 0$. Since $\operatorname{rank}(A) = 0$, we must have $Av \neq 0$, but this implies f_0 is unbounded. We conclude that if $\operatorname{rank}(A) = n$, then f_0 is bounded below iff its domain is bounded, and therefore its minimum is attained.

(d) We can limit $\operatorname{rank}(A) = n$. We show that f_0 has at most one optimal point. The Hessian of f_0 at x is

$$\nabla^2 f(x) = A^t \operatorname{diag}(d) A, \quad d_i = \frac{1}{(b_i - a_i^t x)^2}, \ i = 1, \dots, m$$

which is positive definite if $\operatorname{rank}(A) = n$, i.e., f_0 is strictly convex. Therefore the optimal point, if it exists, is unique.

15 (4.3)**.** Prove that $x^* = \left(1, \frac{1}{2}, -1\right)$ is optimal for the optimization problem

$$\begin{cases} \text{minimize} & \frac{1}{2} x^t P x + q^t x + r \\ \text{subject to} & -1 \leq x_i \leq 1, \quad i = 1, 2, 3 \end{cases}$$

where

$$P = \begin{bmatrix} 13 & 12 & -2 \\ 12 & 17 & 6 \\ -2 & 6 & 12 \end{bmatrix}, \quad q = \begin{bmatrix} -22.0 \\ -14.5 \\ 13.0 \end{bmatrix}, \quad r = 1$$

Answer. $\nabla f_0(x^*) = (-1, 0, 2)$. Therefore the optimality condition is that $\nabla f_0(x^*)^t (y - x) = -(y_1 - 1) + 2(y_2 + 1) \geq 0$ for all y satisfying $-1 \leq y_i \leq 1$.

16 (4.8)**.** Some simple LPs. Give an explicit solution of each of the following LPs.

(a) Minimizing a linear function over an affne set.

$$\begin{cases}\text{minimize} & c^t x\\ \text{subject to} & Ax = b\end{cases}$$

(b) Minimizing a linear function over a halfspace

$$\begin{cases}\text{minimize} & c^t x\\ \text{subject to} & a^t x \le b\end{cases}$$

where $a \neq 0$.

(c) Minimizing a linear function over a rectangle.

$$\begin{cases}\text{minimize} & c^t x\\ \text{subject to} & l \preceq x \preceq u\end{cases}$$

where l and u satisfy $l \preceq u$.

(d) Minimizing a linear function over the probability simplex.

$$\begin{cases}\text{minimize} & c^t x\\ \text{subject to} & \mathbf{1}^t x = 1, \quad x \succeq 0\end{cases}$$

What happens if the equality constraint is replaced by an inequality $\mathbf{1}^t x \le 1$? We can interpret this LP as a simple portfolio optimization problem. The vector x represents the allocation of our total budget over different assets, with x_i the fraction invested in asset i. The return of each investment is fixed and given by $-c_i$, so our total return (which we want to maximize) is $-c^t x$. If we replace the budget constraint $\mathbf{1}^t x = 1$ with an inequality $\mathbf{1}^t x \le 1$, we have the option of not investing a portion of the total budget.

(e) Minimizing a linear function over a unit box with a total budget constraint.

$$\begin{cases}\text{minimize} & c^t x\\ \text{subject to} & \mathbf{1}^t x = \alpha, \quad 0 \preceq x \preceq \mathbf{1}\end{cases}$$

where α is an integer between 0 and n. What happens if α is not an integer (but satisfies $0 \le \alpha \le n$)? What if we change the equality to an inequality $\mathbf{1}^t x \le \alpha$?

(f) Minimizing a linear function over a unit box with a weighted budget constraint.

$$\begin{cases}\text{minimize} & c^t x\\ \text{subject to} & d^t x = \alpha, \quad 0 \preceq x \preceq \mathbf{1}\end{cases}$$

with $d \succ 0$, and $0 \leq \alpha \leq \mathbf{1}^t d$.

Answer.

(a)
- If $b \notin \mathcal{R}(A)$, the optimal value is ∞.
- If c is orthogonal to the nullspace of A. $c = A^t\lambda + \hat{c}, A\hat{c} = 0$. So $c^t x = \lambda^t Ax + \hat{c}^t x = \lambda^t b$. The optimal value is $\lambda^t b$.
- If c is not in the range of $A^t(\hat{c} \neq 0)$, the problem is unbounded($p^* = -\infty$). To verify this, note that $x = x_0 - t\hat{c}$ is feasible for all t, $t \to \infty$, than target $\to$ unbounded.

In summary,

$$p^* = \begin{cases} +\infty & b \notin \mathcal{R}(A)\\ \lambda^t b & c = A^t\lambda \text{ for some } \lambda\\ -\infty & \text{otherwise}\end{cases}$$

(b) Let $c = a\lambda + \hat{c}$, with $a^t\hat{c} = 0$.

- If $\lambda > 0$, the problem is unbounded below. Choose $x = -ta$, and let $t \to \infty$:

$$c^t x = -tc^t a = -t\lambda a^t a \to -\infty$$

and

$$a^t x - b = -ta^t a - b \leq 0$$

for large t, so x is feasible for large t.

- If $\hat{c} \neq 0$, the problem is unbounded blow. Choose $x = ba - t\hat{c}$ and $t \to \infty$.
- If $c = a\lambda$ for some $\lambda \leq 0$, the optimal value is $c^t ab = \lambda b$

In summary, the optimal value is

$$*p = \begin{cases} \lambda b & c = a\lambda \text{ for some } \lambda \leq 0\\ -\infty & \text{otherwise}\end{cases}$$

(c) The optimal x_i^* minimizes $c_i x_i$ subject to the constraint $l_i \leq x_i \leq u_i$. If $c_i > 0$, then $x_i^* = l_i$; if $c_i < 0$, then $x_i^* = u_i$; if $c_i = 0$, then any x_i in the interval $[l_i, u_i]$ is optimal. Therefore, the optimal value of the problem is

$$p^* = l^t c^+ + u^t c^-$$

where $c_i^+ = \max\{c_i, 0\}$ and $c_i^- = \max\{-c_i, 0\}$.

(d) $c^t x = c_{\min}$, $x_i = 0$ if $c_i > c_{\min}$.

(e) Suppose c_i is sorted, the optimal value if $p^* = c_1 + c_2 + \cdots + c_{\lfloor \alpha \rfloor} + c_{1+\lfloor \alpha \rfloor}(\alpha - \lfloor \alpha \rfloor)$.

(f) Suppose that

$$\frac{c_1}{d_1} \leq \frac{c_2}{d_2} \leq \cdots \leq \frac{c_n}{d_n}$$

To minimize the objective, we choose

$$y_1 = d_1, \quad y_2 = d_2, \quad \cdots, \quad y_k = d_k$$

where $k = \max \{i \in \{1, \ldots, n\} \mid d_1 + \cdot + d_i \leq \alpha\}$ (and $k = 0$ if $d_1 > \alpha$). In trems of the original variables.

$$x_1 = \cdots = x_k = 1, \quad x_{k+1} = (\alpha - (d_1 + \cdots + d_k)) / d_{k+1}, \quad x_{k+2} = \cdots = x_n = 0$$

17 (4.9)**.** Square LP. Consider the LP

$$\begin{cases} \text{minimize} & c^t x \\ \text{subject to} & Ax \preceq b \end{cases}$$

with A square and nonsingular. Show that optimal value is given by

$$p^* = \begin{cases} c^t A^{-1} b & A^{-t} c \preceq 0 \\ -\infty & \text{otherwise} \end{cases}$$

Answer. $y = Ax$. We get

$$\begin{cases} \text{minimize} & c^t A^{-1} y \\ \text{subject to} & y \preceq b \end{cases}$$

If $A^{-t} c \preceq 0$, the optimal solution is $y = b$, with $p^* = c^t A^{-1} b$. Otherwise, the LP is unbounded below.

18 (4.15)**.** Relaration of Boolean LP. In a Boolean linear program, the variable x is constrained to have components equal to zero or one:

$$\begin{cases} \text{minimize} & c^t x \\ \text{subject to} & Ax \preceq b \\ & x_i \in \{0, 1\}, \quad i = 1, \ldots, n \end{cases} \tag{1}$$

In general, such problems are very difficult to solve, even though the feasible set is finite (containing at most 2^n points).

In a general method called relawation, the constraint that x_i be zero or one is replaced with the linear inequalities $0 \leq x_i \leq 1$:

$$\begin{cases} \text{minimize} & c^t x \\ \text{subject to} & Ax \preceq b \\ & 0 \leq x_i \leq 1, \quad i = 1, \ldots, n \end{cases} \tag{2}$$

We refer to this problem as the LP(1) relaration of the Boolean LP. The LP relaxation is far easier to solve than the original Boolean LP.

(a) Show that the optimal value of the LP relaxation(2) is a lower bound on the optimal value of the Boolean LP (1). What can you say about the Boolean LP if the LP relaxation is infeasible?

(b) It sometimes happens that the LP relaxation has a solution with $x_i \in \{0, 1\}$. What can you say in this case?

Answer.

(a) The feasible set of the relaxation includes the feasible set of the Boolean LP. It follows that the Boolean LP is infeasible if the relaxation is infeasible, and that the optimal value of the relaxation is less than or equal to the optimal value of the Boolean LP.

(b) The optimal solution of the relaxation is also optimal for the Boolean LP.

19 (4.21)**.** Some simple QCQPs. Give an explicit solution of each of the following QCQPs.

(a) Minimizing a linear function over an ellipsoid centered at the origin.

$$\begin{cases} \text{minimize} & c^t x \\ \text{subject to} & x^t A x \leq 1 \end{cases}$$

where $A \in \mathbb{S}^n_{++}$ and $c \neq 0$. What is the solution if the problem is not convex($A \notin \mathbb{S}^n_+$)?

(b) Minimizing a linear function over an ellipsoid.

$$\begin{cases} \text{minimize} & c^t x \\ \text{subject to} & (x - x_c)^t A (x - x_c) \leq 1 \end{cases}$$

where $A \in \mathbb{S}^n_{++}$ and $c \neq 0$.

(c) Minimizing a quadratic form over an ellipsoid centered at the origin.

$$\begin{cases} \text{minimize} & x^t B x \\ \text{subject to} & x^t A x \leq 1 \end{cases}$$

where $A \in \mathbb{S}^n_{++}$ and $B \in \mathbb{S}^n_{+}$. Also consider the nonconvex extension with $B \notin \mathbb{S}^n_{+}$.

Answer.

(a) If $A \succ 0$, the solution is

$$x^\star = -\frac{1}{\sqrt{c^T A^{-1} c}} A^{-1} c, \quad p^\star = -\left\|A^{-1/2} c\right\|_2 = -\sqrt{c^T A^{-1} c}$$

We make a change of variables $y = A^{1/2} x$, and write $\tilde{c} = A^{-1/2} c$. With this new variable the optimization problem becomes

$$\begin{cases} \text{minimize} & \tilde{c}^t y \\ \text{subject to} & y^t y \leq 1 \end{cases}$$

The answer is $y^* = -\tilde{c}/\|\tilde{c}\|_2$.

$$A = Q \operatorname{diag}(\lambda) Q^t = \sum_{i=1}^n \lambda_i q_i q_i^t$$

We define $y = Qx, b = Qc$, and express the problem as

$$\begin{cases} \text{minimize} & \sum_{i=1} b_i y_i \\ \text{subject to} & \sum_{i=1}^n \lambda_i y_i \leq 1 \end{cases}$$

If $\lambda_i > 0$ for all i, the problem reduces to the case we already discussed. Otherwise, we can distinguish several cases.

- $\lambda_n < 0$. The problem is unbounded below. Let $y_n \to \infty$, we can make any point feasible.
- $\lambda_n = 0$. If for some i, $b_i \neq 0$ and $\lambda_i = 0$, the problem is unbounded below.
- $\lambda_n = 0$, and $b_i = 0$ for all i with $\lambda_i = 0$. In this case we can reduce the problem to a smaller one with all $\lambda_i > 0$.

(b) $y = A^{1/2}(x - x_c), x = A^{-1/2} y + x_c$, and consider the problem

$$\begin{cases} \text{minimize} & c^t A^{-1/2} y + c^t x_c \\ \text{subject to} & y^t y \leq 1 \end{cases}$$

The solution is

$$y^\star = -\left(1/\left\|A^{-1/2} c\right\|_2\right) A^{-1/2} c, \quad x^\star = x_c - \left(1/\left\|A^{-1/2} c\right\|_2\right) A^{-1} c.$$

(c) If $B \succ 0$, then the optimal value is zero.

The smallest eigenvalue of $B \in \mathbb{S}^n$, can be characterized as

$$\lambda_{\min}(B) = \inf_{x^t x = 1} x^t B x$$

which equals to the problem

$$\begin{cases} \text{minimize} & y^t A^{-1/2} B A^{-1/2} y \\ \text{subject to} & y^t y \leq 1 \end{cases}$$

the optimal value is $\lambda_{\min}(A^{-1/2} B A^{-1/2})$

To summarize, the optimal value is

$$p^\star = \begin{cases} \lambda_{\min}\left(A^{-1/2} B A^{-1/2}\right) & \lambda_{\min}\left(A^{-1/2} B A^{-1/2}\right) \leq 0 \\ 0 & \text{otherwise} \end{cases}$$

20 (4.22)**.** Consider the QCQP

$$\begin{cases} \text{minimize} & \frac{1}{2} x^t P x + q^t x + r \\ \text{subject to} & x^t x \leq 1 \end{cases}$$

with $P \in \mathbb{S}^n_{++}$. Show that $x^* = -(P + \lambda I)^{-1} q$ where $\lambda = \max\left\{0, \bar{\lambda}\right\}$ and $\bar{\lambda}$ is the largest solution of the nonlinear equation

$$q^t (P + \lambda I)^{-2} q = 1$$

Answer. x is optimal if and only if $x^t x < 1, Px + q = 0 \quad \text{or} \quad x^t x = 1, Px + q = -\lambda x$ for some $\lambda \geq 0$. If $\|P^{-1} q\|_2 \leq 1$, it is optimal. Otherwise, $\|x\|_2 = 1 \quad \text{and} \quad (P + \lambda) x = -q$ for some $\lambda \geq 0$. $f(\lambda) = \|(P + \lambda)^{-1} q\|_2^2 = \sum_{i=1}^n \frac{q_i^2}{(\lambda + \lambda_i)^2}$ where $\lambda_i > 0$ are the eigenvalues of P. The nonlinear equation $f(\lambda) = 1$ has exactly one nonnegative solution $\bar{\lambda}$. The optimal solution is $x^* = -(P + \bar{\lambda} I)^{-1} q$.

4 Duality

21 (5.3)**.** Problems with one inequality constraint. Express the dual problem of

$$\begin{cases} \text{minimize} & c^t x \\ \text{subject to} & f(x) \leq 0 \end{cases}$$

with $c \neq 0$, in terms of the conjugate f^*. Explain why the problem you give is convex. We do not assume f is convex.

Answer. For $\lambda = 0, g(\lambda) = \inf c^t x = -\infty$. For $\lambda > 0$, $g(\lambda) = \inf(c^t x + \lambda f(x)) = \lambda \inf((c/\lambda)^t x + \lambda f(x) = -\lambda f_1^*(-c/\lambda)$. The dual problem is

$$\begin{cases} \text{minimize} & -\lambda f_1^*(-c/\lambda) \\ \text{subject to} & \lambda \geq 0 \end{cases}$$

22 (5.5)**.** Dual of general LP. Find the dual function of the LP

$$\begin{cases} \text{minimize} & c^t x \\ \text{subject to} & Gx \preceq h \\ & Ax = b \end{cases}$$

Give the dual problem, and make the implicit equality constraints explicit.

Answer.

(a) The Lagrangian is

$$L(x, \lambda, \mu) = \left(c^T + \lambda^T G + \mu^T A\right) x - h\lambda^T - \mu^T b$$

which is an affine function of x. It follows that the dual function is given by

$$g(\lambda, \mu) = \inf_x L(x, \lambda, \mu) = \begin{cases} -\lambda^t h - \mu^t b & c + G^t \lambda + A^t \mu = 0 \\ -\infty & \text{otherwise} \end{cases}$$

(b) The dual problem is

$$\begin{cases} \text{maximize} & g(\lambda, \mu) \\ \text{subject to} & \lambda \succeq 0 \end{cases}$$

After making the implicit constraints explicit, we obtain

$$\begin{cases} \text{maximize} & -\lambda^t h - \mu^t b \\ \text{subject to} & c + G^t \lambda + A^t \mu = 0 \\ & \lambda \geq 0 \end{cases}$$

23 (5.11)**.** Derive a dual problem for

$$\text{minimize} \quad \sum_{i=1}^{N} \|A_i x + b_i\|_2 + \frac{1}{2}\|x - x_0\|_2^2$$

The problem data are $A_i \in \mathbb{R}^{m_i \times n}$; $b_i \in \mathbb{R}^{m_i}$. and $x_0 \in \mathbb{R}^n$. First introduce new variables $y_i \in \mathbb{R}^{m_i}$ and equality constraints $y_i = A_i x + b_i$.

Answer. The Lagrangian is

$$L\left(x, z_1, \ldots, z_N\right)=\sum_{i=1}^N\left\|y_i\right\|_2+\frac{1}{2}\left\|x-x_0\right\|_2^2-\sum_{i=1}^N z_i^t\left(y_i-A_i x-b_i\right)$$

We frst minimize over y_i.

$$\inf_{y_i}\left(\left\|y_i\right\|_2+z_i^T y_i\right)= \begin{cases}0 & \left\|z_i\right\|_2 \leq 1 \\ -\infty & \text{otherwise}\end{cases}$$

We can minimize over x by setting the gradient with respect to x equal to zero.

$$x=x_0+\sum_{i=1}^N A_i^t z$$

Substituting in the Lagrangian gives the dual function

$$g\left(z_1, \ldots, z_N\right)= \begin{cases}\sum_{i=1}^N\left(A_i x_0+b_i\right)^t z_i-\frac{1}{2}\left\|\sum_{i=1}^N A_i^t z_i\right\|_2^2 & \left\|z_i\right\|_2 \leq 1, \quad i=1, \ldots, N \\ \text{otherwise}\end{cases}$$

The dual problem is

$$\begin{cases}\text{maximize} & \sum_{i=1}^N\left(A_i x_0+b_i\right)^t z_i-\frac{1}{2}\left\|\sum_{i=1}^N A_i^t z_i\right\|^2 \\ \text{subject to} & \left\|z_i\right\|_2 \leq 1, i=1, \ldots, N\end{cases}$$

24 (5.12)**.** Analytic centering. Derive a dual problem for

$$\text{minimize} \quad -\sum_{i=1}^m \log\left(b_i-a_i^t x\right)$$

with domain $\{x \mid a_i^t x<b_i, i=1, \ldots, m\}$. First introduce new variables y_i and equality constraints $y_i=b_i-a_i^t x$.

(The solution of this problem is called the analytic center of the linear inequalities $a_i^t x \leq b_i, i=1, \ldots, m$. Analytic centers have geometric applications (see §8.5.3), and play an important role in barrier methods (see chapter 11).)

Answer. We derive the dual of the problem

$$\begin{cases}\text{minimize} & -\sum_{i=1}^m \log y_i \\ \text{subject to} & y=b-A x\end{cases}$$

where $A \in \mathbb{R}^{m \times n}$ has a_i^t as its ith row. The Lagrangian is

$$L(x, y, \mu)=-\sum_{i=1}^m \log y_i+\mu^T(y-b+A x)$$

and the dual function is

$$g(\mu) = \inf_{x,y} \left(-\sum_{i=1}^{m} \log y_i + \mu^t (y - b + Ax) \right)$$

The term $\mu^t Ax$ is unbounded below as a function of x unless $A^t\mu = 0$. The terms in y are unbounded below if $\mu \not\succ$, and achieve their minimum for $y_i = 1/\mu_i$ otherwise. We therefore find the dual function

$$g(\mu) = \begin{cases} \sum_{i=1}^m \log \nu_i + m - b^T\mu & A^T\mu = 0, \quad \mu \succ 0 \\ -\infty & \text{otherwise} \end{cases}$$

and the dual problem

$$\begin{cases} \text{maximize} & \sum_{i=1}^m \log \mu_i - b^t\mu + m \\ \text{subject to} & A^t\mu = 0 \end{cases}$$

25 (5.21)**.** A conven problem in which strong duality fails. Consider the optimization problem

$$\begin{cases} \text{minimize} & e^{-x} \\ \text{subject to} & x^2/y \le 0 \end{cases}$$

with variables x and y, and domain $\mathcal{D} = \{(x, y) \mid y > 0\}$.

(a) Verify that this is a convex optimization problem. Find the optimal value.

(b) Give the Lagrange dual problem, and find the optimal solution λ^* and optimal value d^* of the dual problem. What is the optimal duality gap?

(c) Does Slater's condition hold for this problem?

(d) What is the optimal value $p^*(u)$ of the perturbed problem

$$\begin{cases} \text{minimize} & e^{-x} \\ \text{subject to} & x^2/y \le u \end{cases}$$

as a function of u? Verify that the global sensitivity inequality

$$p^*(u) \ge p^*(0) - \lambda^* u$$

does not hold.

Answer.

(a) $p^* = 1$.

(b) The Lagrangian is $L(x,y,\lambda) = e^{-x} + \lambda x^2/y$. The dual function if

$$g(\lambda) = \inf_{x,y>0}\left(e^{-x} + \lambda x^2/y\right) = \begin{cases} 0 & \lambda \geq 0 \\ -\infty & \lambda < 0 \end{cases}$$

so we can write the dual problem as

$$\begin{cases} \text{minimize} & 0 \\ \text{subject to} & \lambda \geq 0 \end{cases}$$

with optimal value $d^* = 0$. The optimal duality gap is $p^* - d^* = 1$.

(c) Slater's condition is not satisfied.

(d) $p^*(u) = 1$ if $u = 0, p^*(u) = 0$ if $u > 0$ and $p^*(u) = \infty$ if $u < 0$.

26 (5.26)**.** Consider the QCQP

$$\begin{cases} \text{minimize} & x_1^2 + x_2^2 \\ \text{subject to} & (x_1-1)^2 + (x_2-1)^2 \leq 1 \\ & (x_1-1)^2 + (x_2+1)^2 \leq 1 \end{cases}$$

with variable $x \in \mathbb{R}^2$.

(a) Sketch the feasible set and level sets of the objective. Find the optimal point x^* and optimal value p^*.

(b) Give the KKT conditions. Do there exist Lagrange multipliers λ_1^* and λ_2^* that prove that x^* is optimal?

(c) Derive and solve the Lagrange dual problem. Does strong duality hold?

Answer.

(a) The figure 2 shows the feasible set(the intersection of the two shaded disks) and some contour lines of the objective function. There is only one feasible point $(1,0)$, and we have $p^* = 1$.

(b) The KKT conditions are

$$\begin{gathered} (x_1-1)^2 + (x_2-1)^2 \leq 1, \quad (x_1-1)^2 + (x_2+1)^2 \leq 1, \\ \lambda_1 \geq 0, \quad \lambda_2 \geq 0 \\ 2x_1 + 2\lambda_1(x_1-1) + 2\lambda_2(x_1-1) = 0 \\ 2x_2 + 2\lambda_1(x_2-1) + 2\lambda_2(x_2+1) = 0 \\ \lambda_1\left((x_1-1)^2 + (x_2-1)^2 - 1\right) = \lambda_2\left((x_1-1)^2 + (x_2+1)^2 - 1\right) = 0. \end{gathered}$$

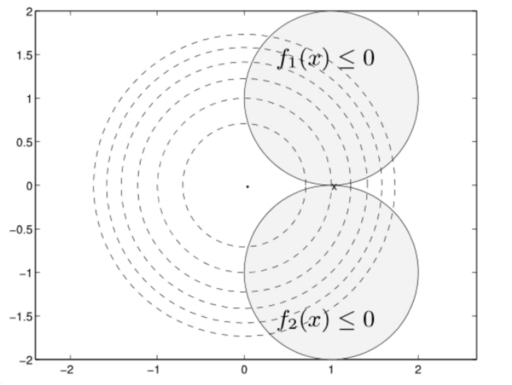

Figure 2: the feasible set

At $x = (1, 0)$, these conditions reduce to

$$\lambda_1 \geq 0, \quad \lambda_2 \geq 0, \quad 2 = 0, \quad -2\lambda_1 + 2\lambda_2 = 0$$

which have no solution.

(c) The Lagrange dual function is given by

$$g(\lambda_1, \lambda_2) = \inf_{x_1, x_2} L(x_1, x_2, \lambda_1, \lambda_2)$$

where

$$\begin{aligned}
& L(x_1, x_2, \lambda_1, \lambda_2) \\
=& x_1^2 + x_2^2 + \lambda_1 \left((x_1 - 1)^2 + (x_2 - 1)^2 - 1\right) + \lambda_2 \left((x_1 - 1)^2 + (x_2 + 1)^2 - 1\right) \\
=& (1 + \lambda_1 + \lambda_2) x_1^2 + (1 + \lambda_1 + \lambda_2) x_2^2 - 2(\lambda_1 + \lambda_2) x_1 - 2(\lambda_1 - \lambda_2) x_2 + \lambda_1 + \lambda_2
\end{aligned}$$

L reaches its minimum for

$$x_1 = \frac{\lambda_1 + \lambda_2}{1 + \lambda_1 + \lambda_2}, \quad x_2 = \frac{\lambda_1 - \lambda_2}{1 + \lambda_1 + \lambda_2}$$

and we find

$$g(\lambda_1, \lambda_2) = \begin{cases} -\frac{(\lambda_1+\lambda_2)^2 + (\lambda_1-\lambda_2)^2}{1+\lambda_1+\lambda_2} + \lambda_1 + \lambda_2 & 1 + \lambda_1 + \lambda_2 \geq 0 \\ -\infty & \text{otherwise} \end{cases}$$

Since g is symmetric, the optimum occurs with $\lambda_1 = \lambda_2$. The dual function then simplifies to

$$g(\lambda_1, \lambda_2) = \frac{2\lambda_1}{2\lambda_1 + 1}$$

We see that $g(\lambda_1, \lambda_2)$ tends to 1 as $\lambda_1 \to \infty$. We have $d^* = p^* = 1$, but the dual optimum is not attained.

In this example, the KKT conditions fail because the dual optimum is not attained.

27 (5.27)**.** Equality constrained least-squares. Consider the equality constrained least-squares problem

$$\begin{cases} \text{minimize} & \|Ax-b\|_2^2 \\ \text{subject to} & Gx = h \end{cases}$$

where $A \in \mathbb{R}^{m\times n}$ with $\operatorname{rank}(A) = n$, and $G \in \mathbb{R}^{p\times n}$ with $\operatorname{rank}(G) = p$.

Give the KKT conditions, and derive expressions for the primal solution x^* and the dual solution ν^*.

Answer.

(a) The Lagrangian is

$$\begin{aligned} L(x,\nu) &= \|Ax-b\|_2^2 + \nu^T(Gx-h) \\ &= x^T A^T A x + \left(G^T\nu - 2A^T b\right)^T x - \nu^T h, \end{aligned}$$

with minimizer $x = -(1/2)(A^tA)^{-1}(G^t\nu - 2A^tb)$. The dual function is

$$g(\nu) = -(1/4)\left(G^T\nu - 2A^Tb\right)^T\left(A^TA\right)^{-1}\left(G^T\nu - 2A^Tb\right) - \nu^T h$$

(b) The optimality conditions are

$$2A^t(Ax^* - b) + G^t\nu^* = 0, \quad Gx^* = h$$

(c) From the first equation,

$$x^\star = \left(A^tA\right)^{-1}\left(A^tb - (1/2)G^t\nu^\star\right)$$

Plugging this expression for x^* into the second equation gives

$$G\left(A^tA\right)^{-1}A^tb - (1/2)G\left(A^tA\right)^{-1}G^T\nu^\star = h$$

Substituting in the first expression gives an analytical expression for x^*.

28 (5.29)**.** The problem

$$\begin{cases} \text{minimize} & -3x_1^2 + x_2^2 + 2x_3^2 + 2\left(x_1+x_2+x_3\right) \\ \text{subject to} & x_1^2 + x_2^2 + x_3^2 = 1 \end{cases}$$

is a special case of (5.32), so strong duality holds even though the problem is not convex. Derive the KKT conditions. Find all solutions x, ν that satisfy the KKT conditions. Which pair corresponds to the optimum?

Answer.

(a) The KKT conditions are

$$x_1^2 + x_2^2 + x_3^2 = 1, \quad (-3+\nu)x_1 + 1 = 0, \quad (1+\nu)x_2 + 1 = 0, \quad (2+\nu)x_3 + 1 = 0$$

(b) From the KKT conditions $\nu \notin \{2, -1, 3\}$. We can therefore eliminate x and reduce the KKT conditions to a nonlinear equation in ν:

$$\frac{1}{(-3+\nu)} + \frac{1}{(1+\nu)} + \frac{1}{(2+\nu)} = 1$$

The lefthand side is plotted in the figure. There are four solutions:

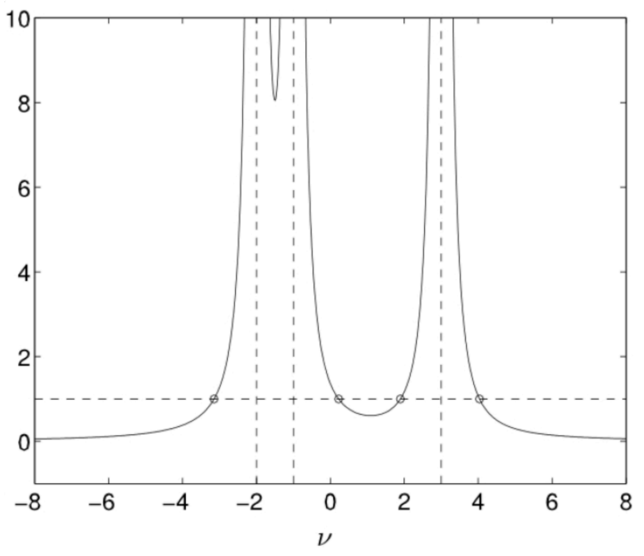

Figure 3: the lefthand side

$$\nu = -3.15, \quad \nu = 0.22, \quad \nu = 1.89, \quad \nu = 4.04$$

corresponding to

$$x = (0.16, 0.47, -0.87), \quad x = (0.36, -0.82, 0.45)$$
$$x = (0.90, -0.35, 0.26), \quad x = (-0.97, -0.20, 0.17)$$

(c) ν^* is the largest of the four values: $\nu^* = 4.0352$. This can be seen several ways. The simplest way is to compare the objective values of the four solutions x, which are

$$f_0(x) = 1.17, \quad f_0(x) = 0.67, \quad f_0(x) = -0.56, \quad f_0(x) = -4.70$$

We can also evaluate the dual objective at the four candidate values for ν. Finally we can note that we must have

$$\nabla^2 f_0(x^\star) + \nu^\star \nabla^2 f_1^\star(x^\star) \succeq 0$$

because x^* is a minimizer of $L(x, \nu^*)$. In other words

$$\begin{bmatrix} -3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} + \nu^* \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \succeq 0$$

and therefore $\nu^* \geq 3$.

29 (9.1)**.** Minimiing a quadratic function. Consider the problem of minimizing a quadratic function:

$$\text{minimize} \quad f(x) = \frac{1}{2}x^t P x + q^t x + r$$

where $P \in \mathbb{S}^n$(but we do not assume $P \succeq 0$).

(a) Show that if $P \not\succeq 0$, i.e., the objective function f is not convex, then the problem is unbounded below.

(b) Now suppose that $P \succeq 0$ (so the objective function is convex), but the optimality condition $Px^* = -q$ does not have a solution. Show that the problem is unbounded below.

Answer.

(a) If $P \not\succeq 0$, we can find v such that $v^t P v < 0$. With $x = tv$ we have

$$f(x) = \frac{1}{2}t^2(v^t P v) + t(q^t v) + r$$

which converges to $-\infty$ as t becomes large.

(b) This means $q \notin \mathcal{R}(P)$. Express q as $q = \tilde{q} + v$, where $\tilde{q}$ is the Euclidean projection of q onto $\mathcal{R}(P)$, and take $v = q - \tilde{q}$. This vector is nonzero and orthogonal to $\mathcal{R}(P)$, i.e., $v^t P v = 0$. It follows that for $x = tv$, we have

$$f(x) = tq^t v + r = t(\tilde{q} + v)^t v + r = t(v^t v) + r$$

which is unbounded below.

30 (9.2)**.** Minimizing a quadratic-over-linear fra tional function. Consider the problem of minimizing the function $f : \mathbb{R}^n \to \mathbb{R}$, defined as

$$f(x) = \frac{\|Ax - b\|_2^2}{c^t x + d}, \quad \mathrm{dom}(f) = \left\{x \mid c^T x + d > 0\right\}$$

We assume $\mathrm{rank}(A) = n$ and $b \notin \mathcal{R}(A)$.

(a) Show that f is closed.

(b) Show that the minimizer x^* of f is given by

$$x^* = x_1 + tx_2$$

where $x_1 = (A^tA)^{-1}A^tb, x_2 = (A^tA)^{-1}c$, and $t \in \mathbb{R}$ can be calculated by solving a quadratic equation.

Answer.

(a) Since $b \notin \mathcal{R}(A)$, the numerator is bounded below by a positive number $(\|Ax_{ls} - b\|_2^2)$. Therefore $f(x) \to \infty$ as x approaches the boundary of $\mathrm{dom}(f)$.

(b) The optimality conditions are

$$\begin{aligned}\nabla f(x) &= \frac{2}{c^Tx-d}A^T(Ax-b) - \frac{\|Ax-b\|_2^2}{(c^Tx-d)^2}c \\ &= \frac{2}{c^Tx-d}(x-x_1) - \frac{\|Ax-b\|_2^2}{(c^Tx-d)^2}x_2 \\ &= 0\end{aligned}$$

and $x = x_1 + tx_2$,

$$t = \frac{\|Ax-b\|_2^2}{2(c^Tx-d)} = \frac{\|Ax_1 + tAx_2 - b\|_2^2}{2(c^Tx_1 + tc^Tx_2 - d)}$$

In other words t must satisfy

$$\begin{aligned}2t^2c^Tx_2 + 2t\left(c^Tx_1 - d\right) &= t^2\|Ax_2\|_2^2 + 2t(Ax_1-b)^T Ax_2 + \|Ax_1-b\|_2^2 \\ &= t^2c^Tx_2 + \|Ax_1-b\|_2^2\end{aligned}$$

which reduces to a. quadratic equation

$$t^2c^Tx_2 + 2t\left(c^Tx_1 - d\right) - \|Ax_1-b\|_2^2 = 0$$

So that

$$\begin{aligned}c^T(x_1+tx_2) - d &= c^Tx_1 - d - \left(c^Tx_1-d\right) + \sqrt{(c^Tx_1-d)^2 + (c^Tx_2)\|Ax_1-b\|_2^2} \\ &= \sqrt{(c^Tx_1-d)^2 + (c^Tx_2)\|Ax_1-b\|_2^2} \\ &> 0.\end{aligned}$$

31 (9.3)**.** Initial point and sublevel set condition. Consider the function $f(x) = x_1^2 + x_2^2$ with domain $\mathrm{dom}(f) = \{(x_1, x_2) \mid x_1 > 1\}$.

(a) What is p^*?

(b) Draw the sublevel set $S = \{x \mid f(x) \le f(x^{(0)})\}$ for $x^{(0)} = (2,2)$. Is the sublevel set S closed? Is f strongly convex on S?

(c) What happens if we apply the gradient method with backtracking line search, starting at $x^{(0)}$? Does $f(x^{(k)})$ converge to p^*?

Answer.

(a) $p^* = \lim_{x \to (x,0)} f(x_1, x_2) = 1$

(b) No, the sublevel set is not closed. The points $(1 + 1/k, 1)$ are in the sublevel set for $k = 1, 2, \ldots,$ but the limit, $(1,1)$ is not.

(c) The algorithm gets stuck at $(1,1)$.

32 (9.5)**.** Backtracking line search. Suppose f is strongly convex with $mI \preceq \nabla^2 f(x) \preceq MI$. Let Δx be a descent direction at x. Show that the backtracking stopping condition holds for

$$0 < t \le -\frac{\nabla f(x)^t \Delta x}{M\|\Delta x\|_2^2}$$

. Use this to give an upper bound on the number of backtracking iterations.

Answer. The upper bound $\nabla^2 f(x) \preceq MI$ implies

$$f(x + t\Delta x) \le f(x) + t\nabla f(x)^T \Delta x + (M/2)t^2 \Delta x^T \Delta x$$

hence $f(x + t\Delta x) \le f(x) + \alpha t \nabla f(x)^t \Delta x$ if

$$t(1-\alpha)\nabla f(x)^T \Delta x + (M/2)t^2 \Delta x^T \Delta x \le 0$$

ie., the exit condition certainly holds if $0 \le t \le t_0$ with

$$t_0 = -2(1-\alpha)\frac{\nabla f(x)^T \Delta x}{M \Delta x^T \Delta x} \ge -\frac{\nabla f(x)^T \Delta x}{M \Delta x^T \Delta x}.$$

Assume $t_0 \le 1$. Then $\beta^k t \le t_0$ for $k \ge \log(1/t_0)/\log(1/\beta)$.

33 (9.9)**.** Newton decrement. Show that the Iewton decrement $\lambda(x)$ satisfies

$$\lambda(x) = \sup_{v^T \nabla^2 f(x) v = 1} \left(-v^T \nabla f(x)\right) = \sup_{v \ne 0} \frac{-v^T \nabla f(x)}{\left(v^T \nabla^2 f(x) v\right)^{1/2}}$$

Answer. The first expression follows from a a change of variables

$$w = \nabla^2 f(x)^{1/2} v, \quad v = \nabla^2 f(x)^{-1/2} w$$

and from

$$\sup_{\|w\|_2=1} -w^T \nabla^2 f(x)^{-1/2} \nabla f(x) = \left\| \nabla f(x)^{-1/2} \nabla f(x) \right\|_2 = \lambda(x).$$

The second expression follows immediately from the first.

34 (9.10)**.** The pure Neuton method. Newton's method with fixed step size $t = 1$ can diverge if the initial point is not close to x^*. In this problem we consider two examples.

(a) $f(x) = \log(e^x + e^{-x})$ has a unique minimizer $x^* = 0$. Run Newton's method with fixed step size $t = 1$, starting at $x^{(0)}$ and at $x^{(0)} = 1.1$.

(b) $f(x) = -\log x + x$ has a unique minimizer $x^* = 1$. Run Newton's method with fixed step size $t = 1$, starting at $x^{(0)} = 3$.

Plot f and f', and show the first few iterates.

Answer.

(a) $f(x) = \log(e^x + e^{-x})$ is a smooth convex function, with a unique minimum at the origin. The pure Newton method started at $x^{(0)} = 1$ produces the following sequence.

k	$x^{(k)}$	$f(x^{(k)}) - p^*$
1	$-8.134 \cdot 10^{-1}$	$4.338 \cdot 10^{-1}$
2	$4.094 \cdot 10^{-1}$	$2.997 \cdot 10^{-1}$
3	$-4.730 \cdot 10^{-2}$	$8.156 \cdot 10^{-2}$
4	$7.060 \cdot 10^{-5}$	$1.118 \cdot 10^{-3}$
5	$-2.346 \cdot 10^{-13}$	$2.492 \cdot 10^{-9}$

Started at $x^{(0)} = 1.1$, the method diverges.

k	$x^{(k)}$	$f(x^{(k)}) - p^*$
1	$-1.129 \cdot 10^{0}$	$5.120 \cdot 10^{-1}$
2	$1.234 \cdot 10^{0}$	$5.349 \cdot 10^{-1}$
3	$-1.695 \cdot 10^{0}$	$6.223 \cdot 10^{-1}$
4	$5.715 \cdot 10^{0}$	$1.035 \cdot 10^{0}$
5	$-2.302 \cdot 10^{4}$	$2.302 \cdot 10^{4}$

(b) $f(x) = -\log x + x$ is smooth and convex on $\mathrm{dom}(f) = \{x \mid x > 0\}$, with a unique minimizer at $x = 1$. The pure Newton method started at $x^{(0)} = 3$ gives as frst iterate

$$x^{(1)} = 3 - f'(3)/f''(3) = -3$$

which lies outside $\mathrm{dom}(f)$.