Exercises

Exercises

Basic definitions

5.1 A simple example. Consider the optimization problem

\[
\begin{align*}
\text{minimize} & \quad x^2 + 1 \\
\text{subject to} & \quad (x - 2)(x - 4) \leq 0,
\end{align*}
\]

with variable \( x \in \mathbb{R} \).

(a) Analysis of primal problem. Give the feasible set, the optimal value, and the optimal solution.

(b) Lagrangian and dual function. Plot the objective \( x^2 + 1 \) versus \( x \). On the same plot, show the feasible set, optimal point and value, and plot the Lagrangian \( L(x, \lambda) \) versus \( x \) for a few positive values of \( \lambda \). Verify the lower bound property \( (p^* \geq \inf_x L(x, \lambda) \) for \( \lambda \geq 0 \). Derive and sketch the Lagrange dual function \( g \).

(c) Lagrange dual problem. State the dual problem, and verify that it is a concave maximization problem. Find the dual optimal value and dual optimal solution \( \lambda^* \). Does strong duality hold?

(d) Sensitivity analysis. Let \( p^*(u) \) denote the optimal value of the problem

\[
\begin{align*}
\text{minimize} & \quad x^2 + 1 \\
\text{subject to} & \quad (x - 2)(x - 4) \leq u,
\end{align*}
\]

as a function of the parameter \( u \). Plot \( p^*(u) \). Verify that \( dp^*(0)/du = -\lambda^* \).

Solution.

(a) The feasible set is the interval \([2, 4]\). The (unique) optimal point is \( x^* = 2 \), and the optimal value is \( p^* = 5 \).

The plot shows \( f_0 \) and \( f_1 \).

(b) The Lagrangian is

\[
L(x, \lambda) = (1 + \lambda)x^2 - 6\lambda x + (1 + 8\lambda).
\]

The plot shows the Lagrangian \( L(x, \lambda) = f_0 + \lambda f_1 \) as a function of \( x \) for different values of \( \lambda \geq 0 \). Note that the minimum value of \( L(x, \lambda) \) over \( x \) (i.e., \( g(\lambda) \)) is always less than \( p^* \). It increases as \( \lambda \) varies from 0 toward 2, reaches its maximum at \( \lambda = 2 \), and then decreases again as \( \lambda \) increases above 2. We have equality \( p^* = g(\lambda) \) for \( \lambda = 2 \).
For $\lambda > -1$, the Lagrangian reaches its minimum at $\hat{x} = 3\lambda/(1 + \lambda)$. For $\lambda \leq -1$ it is unbounded below. Thus

$$g(\lambda) = \begin{cases} -9\lambda^2/(1 + \lambda) + 1 + 8\lambda & \lambda > -1 \\ -\infty & \lambda \leq -1 \end{cases}$$

which is plotted below.

We can verify that the dual function is concave, that its value is equal to $p^* = 5$ for $\lambda = 2$, and less than $p^*$ for other values of $\lambda$.

(c) The Lagrange dual problem is

$$\begin{align*}
\text{maximize} & \quad -9\lambda^2/(1 + \lambda) + 1 + 8\lambda \\
\text{subject to} & \quad \lambda \geq 0.
\end{align*}$$

The dual optimum occurs at $\lambda = 2$, with $d^* = 5$. So for this example we can directly observe that strong duality holds (as it must — Slater’s constraint qualification is satisfied).

(d) The perturbed problem is infeasible for $u < -1$, since $\inf_x(x^2 - 6x + 8) = -1$. For $u \geq -1$, the feasible set is the interval

$$[3 - \sqrt{1+u}, 3 + \sqrt{1+u}]$$

given by the two roots of $x^2 - 6x + 8 = u$. For $-1 \leq u \leq 8$ the optimum is $x^*(u) = 3 - \sqrt{1+u}$. For $u \geq 8$, the optimum is the unconstrained minimum of $f_0$, ...
Exercises

\[ i.e., x^*(u) = 0. \] In summary,

\[ p^*(u) = \begin{cases} \infty & u < -1 \\ 11 + u - 6 \sqrt{1 + u} & -1 \leq u \leq 8 \\ 1 & u \geq 8. \end{cases} \]

The figure shows the optimal value function \( p^*(u) \) and its epigraph.

Finally, we note that \( p^*(u) \) is a differentiable function of \( u \), and that

\[ \frac{dp^*(0)}{du} = -2 = -\lambda^*. \]

5.2 Weak duality for unbounded and infeasible problems. The weak duality inequality, \( d^* \leq p^* \), clearly holds when \( d^* = -\infty \) or \( p^* = \infty \). Show that it holds in the other two cases as well: If \( p^* = -\infty \), then we must have \( d^* = -\infty \), and also, if \( d^* = \infty \), then we must have \( p^* = \infty \).

Solution.

(a) \( p^* = -\infty \). The primal problem is unbounded, \( i.e., \) there exist feasible \( x \) with arbitrarily small values of \( f_0(x) \). This means that

\[ L(x, \lambda) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \]

is unbounded below for all \( \lambda \geq 0 \), \( i.e., \) \( g(\lambda) = -\infty \) for \( \lambda \geq 0 \). Therefore the dual problem is infeasible (\( d^* = -\infty \)).

(b) \( d^* = \infty \). The dual problem is unbounded above. This is only possible if the primal problem is infeasible. If it were feasible, with \( f_i(\tilde{x}) \leq 0 \) for \( i = 1, \ldots, m \), then for all \( \lambda \geq 0 \),

\[ g(\lambda) = \inf(f_0(x) + \sum_{i} \lambda_i f_i(x)) \leq f_0(\tilde{x}) + \sum_{i} \lambda_i f_i(\tilde{x}), \]

so the dual problem is bounded above.

5.3 Problems with one inequality constraint. Express the dual problem of

\[ \begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad f(x) \leq 0,
\end{align*} \]
Exercises

(c) Defining $z = -\lambda w$, we obtain the equivalent problem

$$\begin{align*}
\text{maximize} & \quad -b^T z \\
\text{subject to} & \quad A^T z + c = 0 \\
& \quad z \geq 0.
\end{align*}$$

This is the dual of the original LP.

5.5 Dual of general LP. Find the dual function of the LP

$$\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad Gx \leq h \\
& \quad Ax = b.
\end{align*}$$

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

Solution.

(a) The Lagrangian is

$$L(x, \lambda, \nu) = c^T x + \lambda^T (Gx - h) + \nu^T (Ax - b)$$

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

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

(b) The dual problem is

$$\begin{align*}
\text{maximize} & \quad g(\lambda, \nu) \\
\text{subject to} & \quad \lambda \geq 0.
\end{align*}$$

After making the implicit constraints explicit, we obtain

$$\begin{align*}
\text{maximize} & \quad -\lambda^T h - \nu^T b \\
\text{subject to} & \quad c + G^T \lambda + A^T \nu = 0 \\
& \quad \lambda \geq 0.
\end{align*}$$

5.6 Lower bounds in Chebyshev approximation from least-squares. Consider the Chebyshev or $\ell_\infty$-norm approximation problem

$$\text{minimize} \quad \|Ax - b\|_\infty,$$

where $A \in \mathbb{R}^{m \times n}$ and $\text{rank} A = n$. Let $x_{ch}$ denote an optimal solution (there may be multiple optimal solutions; $x_{ch}$ denotes one of them).

The Chebyshev problem has no closed-form solution, but the corresponding least-squares problem does. Define

$$x_{ls} = \arg\min_x \|Ax - b\|_2 = (A^T A)^{-1} A^T b.$$ 

We address the following question. Suppose that for a particular $A$ and $b$ we have computed the least-squares solution $x_{ls}$ (but not $x_{ch}$). How suboptimal is $x_{ls}$ for the Chebyshev problem? In other words, how much larger is $\|Ax_{ls} - b\|_\infty$ than $\|Ax_{ch} - b\|_\infty$?

(a) Prove the lower bound

$$\|Ax_{ls} - b\|_\infty \leq \sqrt{m} \|Ax_{ch} - b\|_\infty,$$

using the fact that for all $z \in \mathbb{R}^m$,

$$\frac{1}{\sqrt{m}} \|z\|_2 \leq \|z\|_\infty \leq \|z\|_2.$$
(b) In example 5.6 (page 254) we derived a dual for the general norm approximation problem. Applying the results to the $\ell_\infty$-norm (and its dual norm, the $\ell_1$-norm), we can state the following dual for the Chebyshev approximation problem:

\[
\begin{align*}
\text{maximize} & \quad b^T \nu \\
\text{subject to} & \quad \|\nu\|_1 \leq 1 \\
& \quad A^T \nu = 0.
\end{align*}
\] (5.104)

Any feasible $\nu$ corresponds to a lower bound $b^T \nu$ on $\|Ax_{\text{cheb}} - b\|_\infty$.

Denote the least-squares residual as $r_{ls} = b - Ax_{ls}$. Assuming $r_{ls} \neq 0$, show that

\[
\hat{\nu} = -r_{ls}/\|r_{ls}\|_1, \quad \tilde{\nu} = r_{ls}/\|r_{ls}\|_1,
\]

are both feasible in (5.104). By duality $b^T \hat{\nu}$ and $b^T \tilde{\nu}$ are lower bounds on $\|Ax_{\text{cheb}} - b\|_\infty$. Which is the better bound? How do these bounds compare with the bound derived in part (a)?

Solution.

(a) Simple manipulation yields

\[
\|Ax_{\text{cheb}} - b\|_\infty \geq \frac{1}{\sqrt{m}} \|Ax_{\text{cheb}} - b\|_2 \geq \frac{1}{\sqrt{m}} \|Ax_{ls} - b\|_2 \geq \frac{1}{\sqrt{m}} \|Ax_{ls} - b\|_\infty.
\]

(b) From the expression $x_{ls} = (A^T A)^{-1} A^T b$ we note that

\[
A^T r_{ls} = A^T (b - A(A^T A)^{-1} b) = A^T b - A^T b = 0.
\]

Therefore $A^T \hat{\nu} = 0$ and $A^T \tilde{\nu} = 0$. Obviously we also have $\|\hat{\nu}\|_1 = 1$ and $\|\tilde{\nu}\|_1 = 1$, so $\hat{\nu}$ and $\tilde{\nu}$ are dual feasible.

We can write the dual objective value at $\hat{\nu}$ as

\[
b^T \hat{\nu} = -\frac{b^T r_{ls}}{\|r_{ls}\|_1} = \frac{(Ax_{ls} - b)^T r_{ls}}{\|r_{ls}\|_1} = \frac{\|r_{ls}\|_2^2}{\|r_{ls}\|_1}.
\]

and, similarly,

\[
b^T \tilde{\nu} = \frac{\|r_{ls}\|_2^2}{\|r_{ls}\|_1}.
\]

Therefore $\hat{\nu}$ gives a better bound than $\tilde{\nu}$.

Finally, to show that the resulting lower bound is better than the bound in part (a), we have to verify that

\[
\frac{\|r_{ls}\|_2^2}{\|r_{ls}\|_1} \geq \frac{1}{\sqrt{m}} \|r_{ls}\|_\infty.
\]

This follows from the inequalities

\[
\|x\|_1 \leq \sqrt{m}\|x\|_2, \quad \|x\|_\infty \leq \|x\|_2
\]

which hold for general $x \in \mathbb{R}^m$.

5.7 Piecewise-linear minimization. We consider the convex piecewise-linear minimization problem

\[
\begin{align*}
\text{minimize} & \quad \max_{i=1, \ldots, m}(a_i^T x + b_i) \\
\end{align*}
\] (5.105)

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

Exercises

5.10 Optimal experiment design. The following problems arise in experiment design (see §7.5).

(a) \textit{D-optimal design.}

\begin{align*}
\text{minimize} & \quad \log \det \left( \sum_{i=1}^{p} x_i v_i v_i^T \right)^{-1} \\
\text{subject to} & \quad x \geq 0, \quad 1^T x = 1.
\end{align*}

(b) \textit{A-optimal design.}

\begin{align*}
\text{minimize} & \quad \text{tr} \left( \sum_{i=1}^{p} x_i v_i v_i^T \right)^{-1} \\
\text{subject to} & \quad x \geq 0, \quad 1^T x = 1.
\end{align*}

The domain of both problems is \( \{x \mid \sum_{i=1}^{p} x_i v_i v_i^T > 0\} \). The variable is \( x \in \mathbb{R}^p \); the vectors \( v_1, \ldots, v_p \in \mathbb{R}^n \) are given.

Derive dual problems by first introducing a new variable \( X \in \mathbb{S}^n \) and an equality constraint \( X = \sum_{i=1}^{p} x_i v_i v_i^T \), and then applying Lagrange duality. Simplify the dual problems as much as you can.

Solution.

(a) \textit{D-optimal design.}

\begin{align*}
\text{minimize} & \quad \log \det(X^{-1}) \\
\text{subject to} & \quad X = \sum_{i=1}^{p} x_i v_i v_i^T \\
& \quad x \geq 0, \quad 1^T x = 1.
\end{align*}

The Lagrangian is

\[
L(x, Z, z, \nu) = \log \det(X^{-1}) + \text{tr}(ZX) - \sum_{i=1}^{p} x_i v_i^T Z v_i - z^T x + \nu(1^T x - 1)
\]

\[
= \log \det(X^{-1}) + \text{tr}(ZX) + \sum_{i=1}^{p} x_i (-v_i^T Z v_i - z_i + \nu) - \nu.
\]

The minimum over \( x_i \) is bounded below only if \( \nu - v_i^T Z v_i = z_i \). Setting the gradient with respect to \( X \) equal to zero gives \( X^{-1} = Z \). We obtain the dual function

\[
g(Z, z) = \begin{cases} 
\log \det Z + n - \nu & \nu - v_i^T Z v_i = z_i, \quad i = 1, \ldots, p \\
-\infty & \text{otherwise}.
\end{cases}
\]

The dual problem is

\[
\text{maximize} \quad \log \det Z + n - \nu \\
\text{subject to} \quad v_i^T Z v_i \leq \nu, \quad i = 1, \ldots, p,
\]

with domain \( \mathbb{S}^n_{++} \times \mathbb{R} \). We can eliminate \( \nu \) by first making a change of variables \( W = (1/\nu)Z \), which gives

\[
\text{maximize} \quad \log \det W + n + n \log \nu - \nu \\
\text{subject to} \quad v_i^T W v_i \leq 1, \quad i = 1, \ldots, p.
\]

Finally, we note that we can easily optimize \( n \log \nu - \nu \) over \( \nu \). The optimum is \( \nu = n \), and substituting gives

\[
\text{maximize} \quad \log \det W + n \log n \\
\text{subject to} \quad v_i^T W v_i \leq 1, \quad i = 1, \ldots, p.
\]
(b) $A$-optimal design.

minimize $\text{tr}(X^{-1})$
subject to $X = \left( \sum_{i=1}^{p} x_i v_i v_i^T \right)^{-1}$
$x \succeq 0,$ $1^T x = 1.$

The Lagrangian is

$$L(X, Z, z) = \text{tr}(X^{-1}) + \text{tr}(Z X) - \sum_{i=1}^{p} x_i v_i^T Z v_i - z^T x + \nu(1^T x - 1)$$

$$= \text{tr}(X^{-1}) + \text{tr}(Z X) + \sum_{i=1}^{p} x_i (-v_i^T Z v_i - z_i + \nu) - \nu.$$ 

The minimum over $x$ is unbounded below unless $v_i^T Z v_i + z_i = \nu$. The minimum over $X$ can be found by setting the gradient equal to zero: $X = Z$, or $X = Z^{-1/2}$ if $Z \succ 0$, which gives

$$\inf_{X \succ 0} (\text{tr}(X^{-1}) + \text{tr}(Z X)) = \begin{cases} 2 \text{tr}(Z^{1/2}) & Z \succeq 0 \\ -\infty & \text{otherwise}. \end{cases}$$

The dual function is

$$g(Z, z, \nu) = \begin{cases} -\nu + 2 \text{tr}(Z^{1/2}) & Z \succeq 0, \quad v_i^T Z v_i + z_i = \nu \\ -\infty & \text{otherwise}. \end{cases}$$

The dual problem is

maximize $-\nu + 2 \text{tr}(Z^{1/2})$
subject to $v_i^T Z v_i \leq nu, \quad i = 1, \ldots, p$
$Z \succeq 0.$

As a first simplification, we define $W = (1/\nu) Z$, and write the problem as

maximize $-\nu + 2 \sqrt{\nu} \text{tr}(W^{1/2})$
subject to $v_i^T W v_i \leq 1, \quad i = 1, \ldots, p$
$W \succeq 0.$

By optimizing over $\nu > 0$, we obtain

maximize $(\text{tr}(W^{1/2}))^2$
subject to $v_i^T W v_i \leq 1, \quad i = 1, \ldots, p$
$W \succeq 0.$

5.11 Derive a dual problem for

minimize $\sum_{i=1}^{N} ||A_i x + b_i||_2 + (1/2)||x - x_0||_2^2.$

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 $y_i = A_i x + b_i$.

Solution. The Lagrangian is

$$L(x, z_1, \ldots, z_N) = \sum_{i=1}^{N} ||y_i||_2 + \frac{1}{2} ||x - x_0||_2^2 - \sum_{i=1}^{N} z_i (y_i - A_i x - b_i).$$
5.13 Lagrangian relaxation of Boolean LP. A Boolean linear program is an optimization problem of the form
\[
\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad Ax \preceq b \\
& \quad x_i \in \{0, 1\}, \quad i = 1, \ldots, n,
\end{align*}
\]
and is, in general, very difficult to solve. In exercise 4.15 we studied the LP relaxation of this problem,
\[
\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad Ax \preceq b \\
& \quad 0 \leq x_i \leq 1, \quad i = 1, \ldots, n,
\end{align*}
\]
which is far easier to solve, and gives a lower bound on the optimal value of the Boolean LP. In this problem we derive another lower bound for the Boolean LP, and work out the relation between the two lower bounds.

(a) **Lagrangian relaxation.** The Boolean LP can be reformulated as the problem
\[
\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad Ax \preceq b \\
& \quad x_i(1 - x_i) = 0, \quad i = 1, \ldots, n,
\end{align*}
\]
which has quadratic equality constraints. Find the Lagrange dual of this problem. The optimal value of the dual problem (which is convex) gives a lower bound on the optimal value of the Boolean LP. This method of finding a lower bound on the optimal value is called **Lagrangian relaxation.**

(b) Show that the lower bound obtained via Lagrangian relaxation, and via the LP relaxation (5.107), are the same. *Hint. Derive the dual of the LP relaxation (5.107).*

**Solution.**

(a) The Lagrangian is
\[
L(x, \mu, \nu) = c^T x + \mu^T (Ax - b) - \nu^T x + x^T \text{diag}(\nu) x
\]
Minimizing over $x$ gives the dual function
\[
g(\mu, \nu) = \begin{cases} 
-\frac{b^T \mu}{\nu} - \frac{(1/4) \sum_{i=1}^n (c_i + a_i^T \mu - \nu_i)^2}{\nu_i} & \nu \geq 0 \\
-\infty & \text{otherwise}
\end{cases}
\]
where $a_i$ is the $i$th column of $A$, and we adopt the convention that $a^2/0 = \infty$ if $a \neq 0$, and $a^2/0 = 0$ if $a = 0$.

The resulting dual problem is
\[
\begin{align*}
\text{maximize} & \quad -b^T \mu - \frac{(1/4) \sum_{i=1}^n (c_i + a_i^T \mu - \nu_i)^2}{\nu_i} \\
\text{subject to} & \quad \nu \geq 0.
\end{align*}
\]
In order to simplify this dual, we optimize analytically over $\nu$, by noting that
\[
\sup_{\nu_i \geq 0} \left( -\frac{(c_i + a_i^T \mu - \nu_i)^2}{\nu_i} \right) = \begin{cases} 
(c_i + a_i^T \mu) & c_i + a_i^T \mu \leq 0 \\
0 & c_i + a_i^T \mu \geq 0
\end{cases} = \min\{0, (c_i + a_i^T \mu)\}.
\]
This allows us to eliminate $\nu$ from the dual problem, and simplify it as
\[
\begin{align*}
\text{maximize} & \quad -b^T \mu + \sum_{i=1}^n \min\{0, c_i + a_i^T \mu\} \\
\text{subject to} & \quad \mu \geq 0.
\end{align*}
\]
(b) We follow the hint. The Lagrangian and dual function of the LP relaxation are

\[ L(x, u, v, w) = c^T x + u^T (Ax - b) - v^T x + w^T (x - 1) \]

\[ = (c + A^T u - v + w)^T x - b^T u - 1^T w \]

\[ g(u, v, w) = \begin{cases} 
-b^T u - 1^T w & A^T u - v + w + c = 0 \\
-\infty & \text{otherwise.} \end{cases} \]

The dual problem is

\[
\begin{align*}
\text{maximize} & \quad -b^T u - 1^T w \\
\text{subject to} & \quad A^T u - v + w + c = 0 \\
& \quad u \geq 0, v \geq 0, w \geq 0,
\end{align*}
\]

which is equivalent to the Lagrange relaxation problem derived above. We conclude that the two relaxations give the same value.

5.14 A penalty method for equality constraints. We consider the problem

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad Ax = b,
\end{align*}
\]

where \( f_0 : \mathbb{R}^n \to \mathbb{R} \) is convex and differentiable, and \( A \in \mathbb{R}^{m \times n} \) with \( \text{rank} A = m \).

In a quadratic penalty method, we form an auxiliary function

\[ \phi(x) = f(x) + \alpha \|Ax - b\|_2^2, \]

where \( \alpha > 0 \) is a parameter. This auxiliary function consists of the objective plus the penalty term \( \alpha \|Ax - b\|_2^2 \). The idea is that a minimizer of the auxiliary function, \( \tilde{x} \), should be an approximate solution of the original problem. Intuition suggests that the larger the penalty weight \( \alpha \), the better the approximation \( \tilde{x} \) to a solution of the original problem.

Suppose \( \tilde{x} \) is a minimizer of \( \phi \). Show how to find, from \( \tilde{x} \), a dual feasible point for (5.108). Find the corresponding lower bound on the optimal value of (5.108).

**Solution.** If \( \tilde{x} \) minimizes \( \phi \), then

\[ \nabla f_0(\tilde{x}) + 2\alpha A^T (A\tilde{x} - b) = 0. \]

Therefore \( \tilde{x} \) is also a minimizer of

\[ f_0(x) + \nu^T (Ax - b) \]

where \( \nu = 2\alpha (A\tilde{x} - b) \). Therefore \( \nu \) is dual feasible with

\[ g(\nu) = \inf_{\tilde{x}} (f_0(x) + \nu^T (Ax - b)) \]

\[ = f_0(\tilde{x}) + 2\alpha \|A\tilde{x} - b\|_2^2. \]

Therefore,

\[ f_0(x) \geq f_0(\tilde{x}) + 2\alpha \|A\tilde{x} - b\|_2^2 \]

for all \( x \) that satisfy \( Ax = b \).

5.15 Consider the problem

\[
\begin{align*}
\text{minimize} & \quad f_0(x) \\
\text{subject to} & \quad f_i(x) \leq 0, \quad i = 1, \ldots, m,
\end{align*}
\]

(5.109)
Exercises

(c) minimize $x^T \Sigma x$
subject to $\begin{align*}
\mathcal{P}^T x &\geq r_{\text{min}} \\
1^T x &= 1, \\
x &\geq 0,
\end{align*}$
$[n/20] t + 1^T u \leq 0.9$
$\lambda 1 + u \geq 0$
$u \geq 0,$
with variables $x, u, t, v.$

5.20 Dual of channel capacity problem. Derive a dual for the problem

\[
\begin{align*}
\text{minimize} \quad & -c^T x + \sum_{i=1}^m y_i \log y_i \\
\text{subject to} \quad & P x = y \\
& x \geq 0, \\
& 1^T x = 1,
\end{align*}
\]

where $P \in \mathbb{R}^{m \times n}$ has nonnegative elements, and its columns add up to one (i.e., $P^T 1 = 1$). The variables are $x \in \mathbb{R}^n, y \in \mathbb{R}^m$. (For $c_j = \sum_{i=1}^n p_{ij} \log p_{ij}$, the optimal value is, up to a factor $\log 2$, the negative of the capacity of a discrete memoryless channel with channel transition probability matrix $P$; see exercise 4.57.) Simplify the dual problem as much as possible.

Solution. The Lagrangian is
\[
L(x, y, \lambda, \nu, z) = -c^T x + \sum_{i=1}^m y_i \log y_i - \lambda^T x + \nu(1^T x - 1) + z^T (Px - y)
\]
\[
= (-c - \lambda + \nu 1 + P^T z)^T x + \sum_{i=1}^m y_i \log y_i - z^T y - \nu.
\]
The minimum over $x$ is bounded below if and only if
\[
-c - \lambda + \nu 1 + P^T z = 0.
\]
To minimize over $y$, we set the derivative with respect to $y_i$ equal to zero, which gives $\log y_i + 1 - z_i = 0$, and conclude that
\[
\inf_{y_i \geq 0} (y_i \log y_i - z_i y_i) = -e^{z_i - 1}.
\]
The dual function is
\[
g(\lambda, \nu, z) = \begin{cases} 
-\sum_{i=1}^m e^{z_i - 1} - \nu & -c - \lambda + \nu 1 + P^T z = 0 \\
-\infty & \text{otherwise.}
\end{cases}
\]
The dual problem is
\[
\text{maximize} \quad -\sum_{i=1}^m \exp(z_i - 1) - \nu \\
\text{subject to} \quad P^T z - c + \nu 1 \geq 0.
\]
This can be simplified by introducing a variable $w = z + \nu 1$ (and using the fact that $1 = P^T 1$), which gives
\[
\text{maximize} \quad -\sum_{i=1}^m \exp(w_i - \nu - 1) - \nu \\
\text{subject to} \quad P^T w \geq c.
\]
Finally we can easily maximize the objective function over $\nu$ by setting the derivative equal to zero (the optimal value is $\nu = -\log(\sum_i e^{w_i - w_1})$, which leads to
\[
\text{maximize} \quad -\log(\sum_{i=1}^m \exp w_i) - 1 \\
\text{subject to} \quad P^T w \geq c.
\]
This is a geometric program, in convex form, with linear inequality constraints (i.e., monomial inequality constraints in the associated geometric program).
Exercises

There are four solutions:
\[ \begin{align*}
\nu &= -3.15, & \nu &= 0.22, & \nu &= 1.89, & \nu &= 4.04,
\end{align*} \]
corresponding to
\[ \begin{align*}
x &= (0.16, 0.47, -0.87), & x &= (0.36, -0.82, 0.45),
& x &= (0.90, -0.35, 0.26), & x &= (-0.97, -0.20, 0.17).
\end{align*} \]
(c) \( \nu^* \) 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
\[ \begin{align*}
f_0(x) &= 1.17, & f_0(x) &= 0.67, & f_0(x) &= -0.56, & f_0(x) &= -4.70.
\end{align*} \]
We can also evaluate the dual objective at the four candidate values for \( \nu \). Finally we can note that we must have
\[ \begin{align*}
\nabla^2 f_0(x^*) + \nu^* \nabla^2 f_1(x^*) \succeq 0,
\end{align*} \]
because \( x^* \) is a minimizer of \( L(x, \nu^*) \). In other words
\[ \begin{align*}
\begin{bmatrix}
-3 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 2
\end{bmatrix} + \nu^* \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix} \succeq 0,
\end{align*} \]
and therefore \( \nu^* \geq 3. \)

5.30 Derive the KKT conditions for the problem

\[ \begin{align*}
\text{minimize} & \quad \text{tr } X - \log \det X \\
\text{subject to} & \quad Xs = y,
\end{align*} \]

with variable \( X \in S^n \) and domain \( S^n_{++} \). \( y \in \mathbb{R}^n \) and \( s \in \mathbb{R}^n \) are given, with \( s^T y = 1 \). Verify that the optimal solution is given by
\[ \begin{align*}
X^* &= I + yy^T - \frac{1}{s^T s} ss^T.
\end{align*} \]

Solution. We introduce a Lagrange multiplier \( z \in \mathbb{R}^n \) for the equality constraint. The KKT optimality conditions are:
\[ \begin{align*}
X > 0, & \quad Xs = y, & X^{-1} &= I + \frac{1}{2} (zs^T + sz^T). \quad (5.30.A)
\end{align*} \]
We first determine \( z \) from the condition \( Xs = y \). Multiplying the gradient equation on the right with \( y \) gives
\[ s = X^{-1} y = y + \frac{1}{2} (z + (z^T y) s). \quad (5.30.B) \]
By taking the inner product with \( y \) on both sides and simplifying, we get \( z^T y = 1 - y^T y \). Substituting in (5.30.B) we get
\[ z = -2y + (1 + y^T y) s, \]
and substituting this expression for \( z \) in (5.30.A) gives
\[ \begin{align*}
X^{-1} &= I + \frac{1}{2} (-2ys^T - 2sy^T + 2(1 + y^T y) ss^T) \\
&= I + (1 + y^T y) ss^T - ys^T - sy^T.
\end{align*} \]
Finally we verify that this is the inverse of the matrix $X^*$ given above:

\[
(I + (1 + y^T y)ss^T - ys^T - sy^T)X^* = (I + yy^T - (1/s^T s)ss^T) + (1 + y^T y)(ss^T + sy^T - ss^T) - (ys^T + yy^T - ys^T) - (sy^T + (y^T y) sy^T - (1/s^T s)ss^T)
\]

\[
= I.
\]

To complete the solution, we prove that $X^* > 0$. An easy way to see this is to note that

\[
X^* = I + yy^T - \frac{ss^T}{s^T s} = \left(I + \frac{ys^T}{\|s\|_2} \frac{ss^T}{s^T s} \right) \left(I + \frac{ys^T}{\|s\|_2} - \frac{ss^T}{s^T s}\right)^T.
\]

5.31 Supporting hyperplane interpretation of KKT conditions. Consider a convex problem with no equality constraints,

\[
\text{minimize } f_0(x) \\
\text{subject to } f_i(x) \leq 0, \quad i = 1, \ldots, m.
\]

Assume that $x^* \in \mathbb{R}^n$ and $\lambda^* \in \mathbb{R}^m$ satisfy the KKT conditions

\[
f_i(x^*) \leq 0, \quad i = 1, \ldots, m
\]

\[
\lambda^*_i \geq 0, \quad i = 1, \ldots, m
\]

\[
\nabla f_0(x^*) + \sum_{i=1}^m \lambda^*_i \nabla f_i(x^*) = 0.
\]

Show that

\[
\nabla f_0(x^*)^T (x - x^*) \geq 0
\]

for all feasible $x$. In other words the KKT conditions imply the simple optimality criterion of §4.2.3.

Solution. Suppose $x$ is feasible. Since $f_i$ are convex and $f_i(x) \leq 0$ we have

\[
0 \geq f_i(x) \geq f_i(x^*) + \nabla f_i(x^*)^T (x - x^*), \quad i = 1, \ldots, m.
\]

Using $\lambda^*_i \geq 0$, we conclude that

\[
0 \geq \sum_{i=1}^m \lambda^*_i \left(f_i(x^*) + \nabla f_i(x^*)^T (x - x^*)\right)
\]

\[
= \sum_{i=1}^m \lambda^*_i f_i(x^*) + \sum_{i=1}^m \lambda^*_i \nabla f_i(x^*)^T (x - x^*)
\]

\[
= -\nabla f_0(x^*)^T (x - x^*).
\]

In the last line, we use the complementary slackness condition $\lambda^*_i f_i(x^*) = 0$, and the last KKT condition. This shows that $\nabla f_0(x^*)^T (x - x^*) \geq 0$, i.e., $\nabla f_0(x^*)$ defines a supporting hyperplane to the feasible set at $x^*$.

Perturbation and sensitivity analysis

5.32 Optimal value of perturbed problem. Let $f_0, f_1, \ldots, f_m : \mathbb{R}^n \to \mathbb{R}$ be convex. Show that the function

\[
p^*(u, v) = \inf \{ f_0(x) | \exists x \in \mathcal{D}, \ f_i(x) \leq u_i, \ i = 1, \ldots, m, \ Ax - b = v \}
\]