Exercises

1. **Magnitude least-squares.** The magnitude least-squares problem has the form

   \[ \min_w \| |X^T w| - y \|_2, \]

   where \( X \in \mathbb{R}^{n \times m} \), \( y \in \mathbb{R}^m \) are given, and \( |z| \) stands for the vector formed with the magnitude of the elements of vector \( z \). This problem arises in many applications, for example in when recovering a signal or image from the intensity measurements of its Fourier transform, as in X-ray crystallography. (Here, we only work with real numbers, although the problem is often posed in a complex-valued setting, for which the problem is referred to as the phase retrieval problem.) Derive an SDP approximation to the problem. Find its dual. *Hint:* Express the problem as a quadratic one:

   \[ \min_{w,v} \| v - y \|_2^2 : v_i^2 = (X^T w)_i^2, \quad i = 1, \ldots, m, \]

   and use rank relaxation.

2. **Sparse Gaussian models.** We are given data points \( x_1, \ldots, x_n \in \mathbb{R}^p \), which we assume are generated by a multivariate Gaussian. The empirical covariance matrix \( C \) is given by

   \[ C = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{x})(x_i - \hat{x})^T, \quad \hat{x} := \frac{1}{n} \sum_{i=1}^n x_i. \]

   This formula can be obtained by assuming that the data points are generated by a Gaussian \((\hat{x}, C)\), and solving the maximum-likelihood problem, which turns out to have the form

   \[ \max_{X>0} \log \det X - \text{Tr} CX, \]

   At optimum, we have \( X = C^{-1} \) if \( C \) is invertible; otherwise, the maximum-likelihood problem is not well-posed. The matrix \( X = C^{-1} \) is called the precision matrix. It can be shown that \( X_{ij} = 0 \) implies that variables \( i, j \) are conditionally independent.

   In many applications, we’d like to fit a *sparse* Gaussian model to data, in the sense that many pairs \((i, j)\) are conditionally independent. This leads to a modified version of the above problem:

   \[ \max_X \log \det X - \text{Tr} X C - \rho \| X \|_1 \]
where \( \rho > 0 \) is a parameter, and \( \|X\|_1 \) contains the sum of the absolute values of all the elements of \( X \).

(a) Form a dual to the problem.

(b) Show that for every \( \rho > 0 \), the optimal solution to the penalized problem is unique, and bounded as follows: \( \alpha I \preceq X \preceq \beta I \), where

\[
\alpha := \frac{1}{\lambda_{\max}(C) + \rho p}, \quad \beta := \frac{p}{\rho},
\]

where \( \lambda_{\max}(C) \) is the largest eigenvalue of \( C \).

Solution:

(a) We can write the problem as

\[
\max_{X \succ 0} \min_{\|U\|_\infty \leq \rho} \log \det X + \text{Tr} X (C + U),
\]

where \( \|U\|_\infty \) denotes the maximal absolute value of the entries of \( U \). This corresponds to seeking an estimate with maximal worst-case likelihood, over all component-wise bounded additive perturbations \( C + U \) of the empirical covariance matrix \( C \).

We can obtain the dual by exchanging the max and the min:

\[
\min_U \{- \log \det(C + U) - p : \|U\|_\infty \leq \rho, \ C + U \succ 0\} \quad (15.1)
\]

The diagonal elements of an optimal \( U \) are simply \( \hat{U}_{ii} = \rho \). The corresponding covariance matrix estimate is \( \hat{C} := C + \hat{U} \). Since the above dual problem has a compact feasible set, the primal and dual problems are equivalent, that is, strong duality holds. The optimality conditions relate the primal and dual solutions by \( \hat{C} X = I \).

(b) An optimal \( X \) satisfies \( X = (C + U)^{-1} \), where \( \|U\|_\infty \leq \rho \). To establish a lower bound on the eigenvalues of \( X \), we need an upper bound on those of \( C + U \). We have, from convexity and positive homogeneity of the largest eigenvalue function:

\[
\lambda_{\max}(C + U) \leq \lambda_{\max}(C) + \lambda_{\max}(U).
\]

Since \( \|U\|_\infty \leq \rho \), we have

\[
\lambda_{\max}(U) \leq \|U\| = \max_{\xi: \|\xi\|_2 \leq 1} \|U \xi\|_2 \leq pp,
\]

In the above, we have used that, for every \( i \in \{1, \ldots, p\} \), the \( i \)-th component of \( U \xi \) is of the form \( u^T \xi \) where \( \|u\|_\infty \leq \rho \), hence its absolute value is bounded above by \( \rho \). We obtain \( X \succeq \alpha I \), where \( \alpha = 1/(\lambda_{\max}(C) + pp) \).
Likewise, we can show that $X$ is bounded above. Indeed, at optimum, the primal-dual gap is zero:

$$0 = -\log \det(C + U) - p - \log \det X + \langle C, X \rangle + \rho \|X\|_1$$

$$= -p + \langle C, X \rangle + \rho \|X\|_1,$$

where we have used $(C + U)X = I$. Since $C, X$ are both positive semi-definite, we obtain

$$\|X\| \leq \|X\|_F \leq \|X\|_1 \leq \frac{p}{\rho}I,$$

as claimed.

3. The unconstrained optimization problem

$$q^*(\rho) := \sum_{k=1}^{m} \sqrt{\rho^2 + (a_k^T x - b_k)^2},$$

(15.2)

where $\rho$ is a small positive constant, is sometimes used as a smooth approximation of the $\ell_1$-minimization problem

$$p^* := \text{minimize } \|Ax - b\|_1.$$

($A \in \mathbb{R}^{m \times n}$ is the matrix with rows $a_k^T$.) In this problem, we investigate the following issue: how small should $\rho$ be, in order to guarantee that $|\phi(\rho) - p^*| \leq \epsilon$?

**Solution:**

(a) Problem (15.2) is equivalent to

$$\text{minimize } \sum_{k=1}^{m} \sqrt{\rho^2 + y_k^2}$$

subject to $y = Ax - b$,

with variables $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$. Derive the Lagrange dual.

(b) Let $\hat{x}$ be the optimal solution of (15.2) and define $\hat{y} = A\hat{x} - b$ for $k = 1, \ldots, m$. Let $p^*$ be the optimal value of the $\ell_1$-minimization problem. Show that

$$\|\hat{y}\|_1 \geq p^* \geq \sum_{k=1}^{m} \frac{y_k^2}{\sqrt{\rho^2 + y_k^2}},$$

and that this implies

$$q^*(\rho) \geq p^* \geq q^*(\rho) - m\rho$$

where $q^*(\rho)$ is the optimal value of (15.2).
Solution.

(a) The Lagrangian is

\[ L(x, y, z) = \sum_{k=1}^{m} \sqrt{\rho^2 + y_k^2} + z^T (y - Ax + b). \]

This is bounded below as a function of \( x \) only if \( A^T z = 0 \). We find the minimum over \( y_k \) by setting the derivative equal to zero:

\[ z_k = -\frac{y_k}{\sqrt{\rho^2 + y_k^2}}. \]

This is solvable only if \( |z_k| < 1 \), with solution

\[ y_k = -\rho \frac{z_k}{\sqrt{1 - z_k^2}}, \quad \inf_y \left( \sqrt{\rho^2 + y_k^2} + z_k y_k \right) = \rho \sqrt{1 - z_k^2}. \]

If \( z_k = 1 \), the infimum of \( (\rho^2 + y_k^2)^{1/2} + z_k y_k \) is reached asymptotically as \( y_k \to -\infty \), and if \( z_k = -1 \), the infimum of is reached as \( y_k \to \infty \). Therefore the dual problem is

\[
\begin{align*}
\text{maximize} & \quad b^T z + \rho \sum_{k=1}^{m} \sqrt{1 - z_k^2} \\
\text{subject to} & \quad A^T z = 0 \\
\end{align*}
\]

(15.3)

(b) The first inequality is an immediate consequence of the definition of \( p^* \). The second inequality can be derived from the optimality conditions of the smooth problem,

\[ \hat{z}_k = \frac{-\hat{y}_k}{\sqrt{\rho^2 + \hat{y}_k^2}}, \quad A^T \hat{z} = 0, \quad \hat{y} = A\hat{x} - b, \]

and the dual of the \( \ell_1 \) problem

\[
\begin{align*}
\text{maximize} & \quad -b^T z \\
\text{subject to} & \quad A^T z = 0 \\
\end{align*}
\]

\[ \|z\|_{\infty} \leq 1. \]

If we plug in \( \hat{z} \) in the dual of the \( \ell_1 \)-approximation problem we find the lower bound

\[
p^* \geq -b^T \hat{z} = (A\hat{x} - b)^T \hat{z} = -\hat{y}^T z = \sum_{k=1}^{m} \hat{y}_k^2 / \sqrt{\rho^2 + \hat{y}_k^2}. \]
The inequality $q^*(\rho) \geq p^*$ follows from

$$\|\hat{y}\|_1 \leq \sum_k \sqrt{\rho^2 + (a_k^T \hat{x} - b_k)^2} = q^*(\rho).$$

The second inequality $q^*(\rho) - m \rho$ follows from

$$\sum_{k=1}^m \frac{\hat{y}_k^2}{\sqrt{\rho^2 + \hat{y}_k^2}} = \sum_{k=1}^m \sqrt{\rho^2 + \hat{y}_k^2} - \rho^2 \sum_{k=1}^m \frac{1}{\sqrt{\rho^2 + \hat{y}_k^2}}$$

$$= q^*(\rho) - \rho^2 \sum_{k=1}^m \frac{1}{\sqrt{\rho^2 + \hat{y}_k^2}}$$

$$\geq q^*(\rho) - m \rho.$$