EE236A Linear Programming

Exercises

Prof. L. Vandenberghe
Electrical Engineering Department
University of California, Los Angeles

Winter Quarter 2012-2013

1 Linear inequalities, halfspaces, polyhedra

Exercise 1. When does one halfspace contain another? Give conditions under which

$$\{x \mid a^T x \le b\} \subseteq \{x \mid \tilde{a}^T x \le \tilde{b}\}$$

($a \ne 0$, $\tilde{a} \ne 0$). Also find the conditions under which the two halfspaces are equal.

Exercise 2. What is the distance between the two parallel hyperplanes $\{x \in \mathbf{R}^n \mid a^T x = b_1\}$ and $\{x \in \mathbf{R}^n \mid a^T x = b_2\}$?

Exercise 3. Consider a waveform

$$s(x,t) = f(t - a^T x)$$

where t denotes time, x denotes position in $\mathbf{R}^3$, $f : \mathbf{R} \to \mathbf{R}$ is a given function, and $a \in \mathbf{R}^3$ is a given nonzero vector. The surfaces defined by

$$t - a^T x = \text{constant}$$

are called wavefronts. What is the velocity (expressed as a function of a) with which wavefronts propagate? As an example, consider a sinusoidal plane wave $s(x,t) = \sin(\omega t - k^T x)$.

Exercise 4. Which of the following sets S are polyhedra? If possible, express S in inequality form, *i.e.*, give matrices A and b such that $S = \{x \mid Ax \le b\}$.

(a) $S = \{y_1 a_1 + y_2 a_2 \mid -1 \le y_1 \le 1, -1 \le y_2 \le 1\}$ for given $a_1, a_2 \in \mathbf{R}^n$.

(b) $S = \{x \in \mathbf{R}^n \mid x \succeq 0,\ \mathbf{1}^T x = 1,\ \sum_{i=1}^n x_i a_i = b_1,\ \sum_{i=1}^n x_i a_i^2 = b_2\}$, where $a_i \in \mathbf{R}$ ($i = 1, \ldots, n$), $b_1 \in \mathbf{R}$, and $b_2 \in \mathbf{R}$ are given.

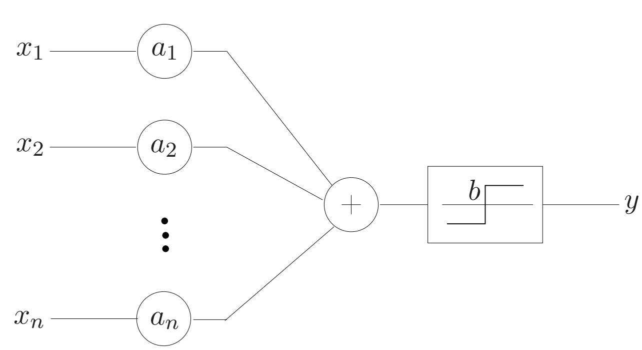
(c) $S = \{x \in \mathbf{R}^n \mid x \succeq 0,\ x^T y \le 1 \text{ for all } y \text{ with } \|y\| = 1\}$.

(d) $S = \{x \in \mathbf{R}^n \mid x \succeq 0,\ x^T y \le 1 \text{ for all } y \text{ with } \sum_i |y_i| = 1\}$.

(e) $S = \{x \in \mathbf{R}^n \mid \|x - x_0\| \le \|x - x_1\|\}$ where $x_0, x_1 \in \mathbf{R}^n$ are given. S is the the set of points that are closer to x_0 than to x_1.

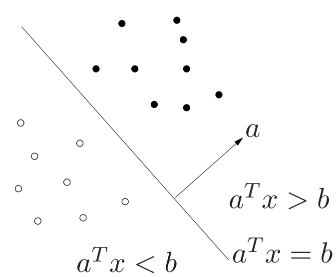
(f) $S = \{x \in \mathbf{R}^n \mid \|x - x_0\| \le \|x - x_i\|,\ i = 1, \ldots, K\}$ where $x_0, \ldots, x_K \in \mathbf{R}^n$ are given. S is the set of points that are closer to x_0 than to the other x_i.

Exercise 5. *Linear and piecewise-linear classification.* The figure shows a block diagram of a linear classification algorithm.

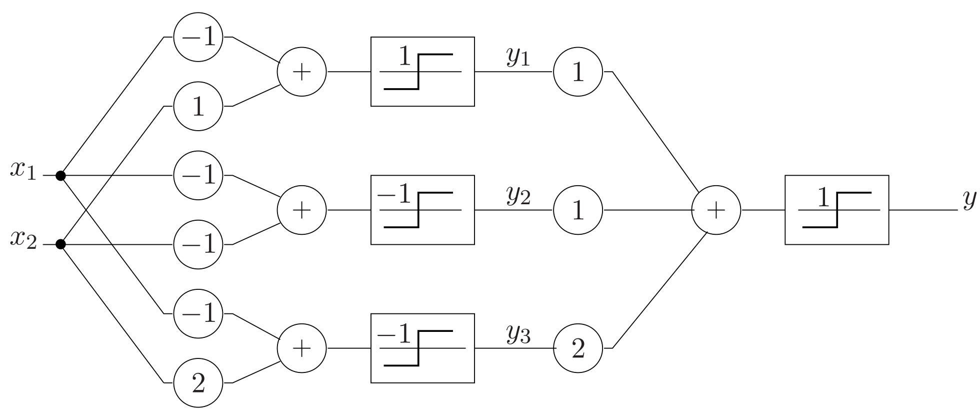


The classifier has n inputs x_i. These inputs are first multiplied with coefficients a_i and added. The result $a^T x = \sum_{i=1}^n a_i x_i$ is then compared with a threshold b. If $a^T x \geq b$, the output of the classifier is $y = 1$; if $a^T x < b$, the output is $y = -1$.

The algorithm can be interpreted geometrically as follows. The set defined by $a^T x = b$ is a hyperplane with normal vector a. This hyperplane divides $\mathbf{R}^n$ in two open halfspaces: one halfspace where $a^T x > b$, and another halfspace where $a^T x < b$. The output of the classifier is $y = 1$ or $y = -1$ depending on the halfspace in which x lies. If $a^T x = b$, we arbitrarily assign $+1$ to the output. This is illustrated below.



By combining linear classifiers, we can build classifiers that divide $\mathbf{R}^n$ in more complicated regions than halfspaces. In the block diagram below we combine four linear classifiers. The first three take the same input $x \in \mathbf{R}^2$. Their outputs y_1, y_2, and y_3 are the inputs to the fourth classifier.



Make a sketch of the region of input vectors in $\mathbf{R}^2$ for which the output y is equal to 1.

Exercise 6. *Measurement with bounded errors.* A series of K measurements $y_1, \ldots, y_K \in \mathbf{R}$, are taken in order to estimate an unknown vector $x \in \mathbf{R}^q$. The measurements are related to the unknown vector x by $y_i = a_i^T x + v_i$, where v_i is a measurement noise that satisfies $|v_i| \leq \alpha$ but is otherwise unknown. The vectors a_i and the measurement noise bound α are known. Let X denote the set of vectors x that are consistent with the observations $y_1, \ldots, y_K$, *i.e.*, the set of x that could have resulted in the measured values of y_i. Show that X is a polyhedron.

Now we examine what happens when the measurements are occasionally in error, *i.e.*, for a few i we have no relation between x and y_i. More precisely suppose that I_{fault} is a subset of $\{1, \ldots, K\}$, and that $y_i = a_i^T x + v_i$ with $|v_i| \leq \alpha$ (as above) for $i \notin I_{\text{fault}}$, but for $i \in I_{\text{fault}}$, there is no relation between x and y_i. The set I_{fault} is the set of times of the faulty measurements.

Suppose you know that I_{fault} has at most J elements, *i.e.*, out of K measurements, at most J are faulty. You do not know I_{fault}; you know only a bound on its cardinality (size). Is X (the set of x consistent with the measurements) a polyhedron for $J > 0$?

2 Some simple linear programs

Exercise 7. Consider the LP

$$\begin{array}{ll} \text{minimize} & c_1x_1 + c_2x_2 + c_3x_3 \\ \text{subject to} & x_1 + x_2 \geq 1 \\ & x_1 + 2x_2 \leq 3 \\ & x_1 \geq 0 \\ & x_2 \geq 0 \\ & x_3 \geq 0. \end{array}$$

Give the optimal value and the optimal set for the following values of c: $c = (-1, 0, 1)$, $c = (0, 1, 0)$, $c = (0, 0, -1)$.

Exercise 8. For each of the following LPs, express the optimal value and the optimal solution in terms of the problem parameters (c, k, d, α, d_1, d_2). If the optimal solution is not unique, it is sufficient to give one optimal solution.

(a)

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & 0 \leq x \leq \mathbf{1} \end{array}$$

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

(b)

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & -1 \leq \mathbf{1}^T x \leq 1 \end{array}$$

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

(c)

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & 0 \leq x_1 \leq x_2 \leq \cdots \leq x_n \leq 1. \end{array}$$

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

(d)

$$\begin{array}{ll} \text{maximize} & c^T x \\ \text{subject to} & \mathbf{1}^T x = k \\ & 0 \leq x \leq \mathbf{1} \end{array}$$

with variable $x \in \mathbf{R}^n$. k is an integer with $1 \leq k \leq n$.

(e)

$$\begin{array}{ll}\text{maximize} & c^T x \\ \text{subject to} & \mathbf{1}^T x \le k \\ & 0 \le x \le \mathbf{1}\end{array}$$

with variable $x \in \mathbf{R}^n$. k is an integer with $1 \le k \le n$.

(f)

$$\begin{array}{ll}\text{maximize} & c^T x \\ \text{subject to} & -y \le x \le y \\ & \mathbf{1}^T y = k \\ & y \le \mathbf{1}\end{array}$$

with variables $x \in \mathbf{R}^n$ and $y \in \mathbf{R}^n$. k is an integer with $1 \le k \le n$.

(g)

$$\begin{array}{ll}\text{maximize} & c^T x \\ \text{subject to} & d^T x = \alpha \\ & 0 \le x \le \mathbf{1}\end{array}$$

with variable $x \in \mathbf{R}^n$. α and the components of d are positive.

(h)

$$\begin{array}{ll}\text{minimize} & \mathbf{1}^T u + \mathbf{1}^T v \\ \text{subject to} & u - v = c \\ & u \ge 0, \ v \ge 0\end{array}$$

with variables $u \in \mathbf{R}^n$ and $v \in \mathbf{R}^n$.

(i)

$$\begin{array}{ll}\text{minimize} & d_1^T u - d_2^T v \\ \text{subject to} & u - v = c \\ & u \ge 0, \ v \ge 0\end{array}$$

with variables $u \in \mathbf{R}^n$ and $v \in \mathbf{R}^n$. We assume that $d_1 \ge d_2$.

Exercise 9. *An optimal control problem with an analytical solution.* We consider the problem of maximizing a linear function of the final state of a linear system, subject to bounds on the inputs:

$$\begin{array}{ll}\text{maximize} & d^T x(N) \\ \text{subject to} & |u(t)| \le U, \quad t = 0, \ldots, N-1 \\ & \sum_{t=0}^{N-1} |u(t)| \le \alpha,\end{array} \tag{1}$$

where x and u are related via the recursion

$$x(t+1) = Ax(t) + Bu(t), \quad x(0) = 0,$$

and the problem data are $d \in \mathbf{R}^n$, $U, \alpha \in \mathbf{R}$, $A \in \mathbf{R}^{n \times n}$ and $B \in \mathbf{R}^n$. The variables are the input sequence $u(0), \ldots, u(N-1)$.

(a) Express (1) as an LP.

(b) Formulate a simple algorithm for solving this LP. (It can be solved very easily, without using a general LP code.) *Hint.* The problem is a variation on exercise 8, parts (a),(b),(c).

(c) Apply your method to the matrices

$$A = \begin{bmatrix} 9.9007\,10^{-1} & 9.9340\,10^{-3} & -9.4523\,10^{-3} & 9.4523\,10^{-3} \\ 9.9340\,10^{-2} & 9.0066\,10^{-1} & 9.4523\,10^{-2} & -9.4523\,10^{-2} \\ 9.9502\,10^{-2} & 4.9793\,10^{-4} & 9.9952\,10^{-1} & 4.8172\,10^{-4} \\ 4.9793\,10^{-3} & 9.5021\,10^{-2} & 4.8172\,10^{-3} & 9.9518\,10^{-1} \end{bmatrix}, \tag{2}$$

$$B = \begin{bmatrix} 9.9502\,10^{-2} \\ 4.9793\,10^{-3} \\ 4.9834\,10^{-3} \\ 1.6617\,10^{-4} \end{bmatrix}. \tag{3}$$

(You can download these matrices by executing the matlab file `ex9data.m` which can be found on the class webpage. The calling sequence is `[A,b] = ex9data`.) Use

$$d = (0, 0, 1, -1), \qquad N = 100, \qquad U = 2, \qquad \alpha = 161.$$

Plot the optimal input and the resulting sequences $x_3(t)$ and $x_4(t)$.

Remark. This model was derived as follows. We consider a system described by two second-order equations

$$\begin{aligned} m_1 \ddot{v}_1(t) &= -K(v_1(t) - v_2(t)) - D(\dot{v}_1(t) - \dot{v}_2(t)) + u(t) \\ m_2 \ddot{v}_2(t) &= K(v_1(t) - v_2(t)) + D(\dot{v}_1(t) - \dot{v}_2(t)). \end{aligned}$$

These equations describe the motion of two masses m_1 and m_2 with positions $v_1 \in \mathbf{R}$ and $v_2 \in \mathbf{R}$, respectively, and connected by a spring with spring constant K and a damper with constant D. An external force u is applied to the first mass. We use the values

$$m_1 = m_2 = 1, \qquad K = 1, \qquad D = 0.1,$$

so the state equations are

$$\begin{bmatrix} \ddot{v}_1(t) \\ \ddot{v}_2(t) \\ \dot{v}_1(t) \\ \dot{v}_2(t) \end{bmatrix} = \begin{bmatrix} -0.1 & 0.1 & -1.0 & 1.0 \\ 0.1 & -0.1 & 1.0 & -1.0 \\ 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 \end{bmatrix} \begin{bmatrix} \dot{v}_1(t) \\ \dot{v}_2(t) \\ v_1(t) \\ v_2(t) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} u(t).$$

We discretize the system by considering inputs u that are piecewise constant with sampling interval $T = 0.1$, *i.e.*, we assume u is constant in the intervals $[0.1k, 0.1(k+1))$, for $k = 0, 1, 2, \ldots$. It can be shown that the discretized state equations are

$$z((k+1)T) = Az(kT) + Bu(kT), \quad k \in \mathbf{Z}, \tag{4}$$

where $z(t) = (\dot{v}_1(t), \dot{v}_2(t), v_1(t), v_2(t))$, and A and B given by (2) and (3).

Using the cost function $d^T x(N)$ with $d = (0, 0, 1, -1)$ means that we try to maximize the distance between the two masses after N time steps.

Exercise 10. *Power allocation problem with analytical solution.* Consider a system of n transmitters and n receivers. The ith transmitter transmits with power x_i, $i = 1, \dots, n$. The vector x is the variable in this problem. The path gain from each transmitter j to each receiver i is denoted A_{ij} and is assumed to be known. (Obviously, $A_{ij} \geq 0$, so the matrix A is elementwise nonnegative. We also assume that $A_{ii} > 0$.) The signal received by each receiver i consists of three parts: the desired signal, arriving from transmitter i with power $A_{ii}x_i$, the interfering signal, arriving from the other transmitters with power $\sum_{j \neq i} A_{ij}x_j$, and noise v_i (v_i is positive and known). We are interested in allocating the powers x_i in such a way that the signal to noise plus interference ratio (SNIR) at each of the receivers exceeds a level α. (Thus α is the minimum acceptable SNIR for the receivers; a typical value might be around $\alpha = 3$.) In other words, we want to find $x \geq 0$ such that for $i = 1, \dots, n$

$$A_{ii}x_i \geq \alpha \left(\sum_{j \neq i} A_{ij}x_j + v_i \right).$$

Equivalently, the vector x has to satisfy the set of linear inequalities

$$x \geq 0, \qquad Bx \geq \alpha v \tag{5}$$

where $B \in \mathbf{R}^{n \times n}$ is defined as

$$B_{ii} = A_{ii}, \qquad B_{ij} = -\alpha A_{ij}, \quad j \neq i.$$

(a) Suppose you are given a desired level of α, so the right hand side αv in (5) is a known positive vector. Show that (5) is feasible if and only if B is invertible and $z = B^{-1}\mathbf{1} \geq 0$ ($\mathbf{1}$ is the vector with all components 1). Show how to construct a feasible power allocation x from z.

(b) Show how to find the largest possible SNIR, *i.e.*, how to maximize α subject to the existence of a feasible power allocation.

Hint. You can refer to the following result from linear algebra. Let $T \in \mathbf{R}^{n \times n}$ be a matrix with nonnegative elements, and $s \in \mathbf{R}$. Then the following statements are equivalent:

(a) There exists an $x \geq 0$ with $(sI - T)x > 0$.

(b) $sI - T$ is nonsingular and the matrix $(sI - T)^{-1}$ has nonnegative elements.

(c) $s > \max_i |\lambda_i(T)|$ where $\lambda_i(T)$ ($i = 1, \dots, n$) are the eigenvalues of T. The quantity $\rho(T) = \max_i |\lambda_i(T)|$ is called the *spectral radius* of T. It is a complicated but readily computed function of T.

(For such s, the matrix $sI - T$ is called a *nonsingular M-matrix.*)

Remark. This problem gives an analytic solution to a very special form of transmitter power allocation problem. Specifically, there are exactly as many transmitters as receivers, and no power limits on the transmitters. One consequence is that the receiver noises v_i play no role at all in the solution — just crank up all the transmitters to overpower the noises!

3 Geometry of linear programming

Exercise 11. (a) Is $\tilde{x} = (1, 1, 1, 1)$ an extreme point of the polyhedron $\mathcal{P}$ defined by the linear inequalities

$$\begin{bmatrix} -1 & -6 & 1 & 3 \\ -1 & -2 & 7 & 1 \\ 0 & 3 & -10 & -1 \\ -6 & -11 & -2 & 12 \\ 1 & 6 & -1 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \leq \begin{bmatrix} -3 \\ 5 \\ -8 \\ -7 \\ 4 \end{bmatrix}?$$

If it is, find a vector c such that $\tilde{x}$ is the unique minimizer of $c^T x$ over $\mathcal{P}$.

(b) Same question for the polyhedron defined by the inequalities

$$\begin{bmatrix} 0 & -5 & -2 & -5 \\ -7 & -7 & -2 & -2 \\ -4 & -4 & -7 & -7 \\ -8 & -3 & -3 & -4 \\ -4 & -4 & 2 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \leq \begin{bmatrix} -12 \\ -17 \\ -22 \\ -18 \\ -8 \end{bmatrix}$$

and the equality $8x_1 - 7x_2 - 10x_3 - 11x_4 = -20$.

Feel free to use MATLAB (in particular the `rank` command).

Exercise 12. We define a polyhedron

$$\mathcal{P} = \{x \in \mathbf{R}^5 \mid Ax = b, \ -\mathbf{1} \leq x \leq \mathbf{1}\},$$

with

$$A = \begin{bmatrix} 0 & 1 & 1 & 1 & -2 \\ 0 & -1 & 1 & -1 & 0 \\ 2 & 0 & 1 & 0 & 1 \end{bmatrix}, \qquad b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

The following three vectors x are in $\mathcal{P}$:

(a) $\hat{x} = (1, -1/2, 0, -1/2, -1)$

(b) $\hat{x} = (0, 0, 1, 0, 0)$

(c) $\hat{x} = (0, 1, 1, -1, 0)$.

Are these vectors extreme points of $\mathcal{P}$? For each $\hat{x}$, if it is an extreme point, give a vector c for which $\hat{x}$ is the unique solution of the optimization problem

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & -\mathbf{1} \leq x \leq \mathbf{1}. \end{array}$$

Exercise 13. An $n \times n$ matrix X is called *doubly stochastic* if

$$X_{ij} \geq 0, \quad i, j = 1, \ldots, n, \qquad \sum_{i=1}^{n} X_{ij} = 1, \quad j = 1, \ldots, n, \qquad \sum_{j=1}^{n} X_{ij} = 1, \quad i = 1, \ldots, n,$$

In words, the entries of X are nonnegative, and its row and column sums are equal to one.

(a) What are the extreme points of the set of doubly stochastic matrices? How many extreme points are there? Explain your answer.

(b) What is the optimal value of the LP

$$\begin{array}{ll} \text{maximize} & a^T X b \\ \text{subject to} & X \text{ doubly stochastic,} \end{array}$$

with X as variable, for a general vector $b \in \mathbf{R}^n$ and each of the following choices of a?

- $a = (1, 0, 0, \dots, 0)$.
- $a = (1, 1, 0, \dots, 0)$.
- $a = (1, -1, 0, \dots, 0)$.

Exercise 14. *Carathéodory's theorem.* A point of the form $\theta_1 v_1 + \cdots + \theta_m v_m$, where $\theta_1 + \cdots + \theta_m = 1$ and $\theta_i \geq 0$, $i = 1, \dots, m$, is called a *convex combination* of $v_1, \dots, v_m$. Suppose x is a convex combination of points $v_1, \dots, v_m$ in $\mathbf{R}^n$. Show that x is a convex combination of a subset of $r \leq n+1$ of the points $v_1, \dots, v_m$. In other words, show that x can be expressed as

$$x = \hat{\theta}_1 v_1 + \cdots + \hat{\theta}_m v_m,$$

where $\hat{\theta}_i \geq 0$, $\sum_{i=1}^m \hat{\theta}_i = 1$, and at most $n+1$ of the coefficients $\hat{\theta}_i$ are nonzero.

4 Formulating problems as LPs

Exercise 15. Formulate the following problems as LPs:

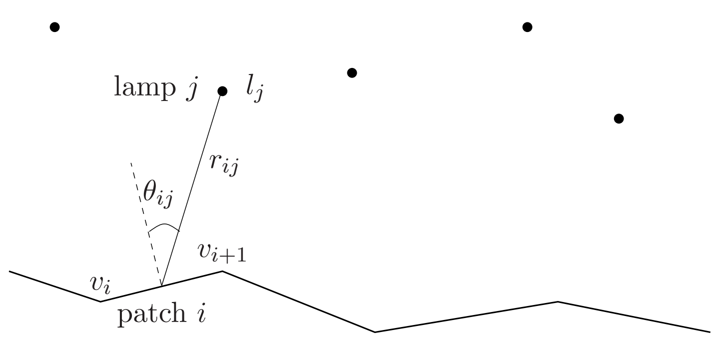
(a) minimize $\|Ax - b\|_1$ subject to $\|x\|_\infty \leq 1$.

(b) minimize $\|x\|_1$ subject to $\|Ax - b\|_\infty \leq 1$.

(c) minimize $\|Ax - b\|_1 + \|x\|_\infty$.

In each problem, $A \in \mathbf{R}^{m \times n}$ and $b \in \mathbf{R}^m$ are given, and $x \in \mathbf{R}^n$ is the optimization variable.

Exercise 16. *An illumination problem.* We consider an illumination system of m lamps, at positions $l_1, \dots, l_m \in \mathbf{R}^2$, illuminating n flat patches.



The patches are line segments; the ith patch is given by $[v_i, v_{i+1}]$ where $v_1, \ldots, v_{n+1} \in \mathbf{R}^2$. The variables in the problem are the lamp powers $p_1, \ldots, p_m$, which can vary between 0 and 1.

The illumination at (the midpoint of) patch i is denoted I_i. We will use a simple model for the illumination:

$$I_i = \sum_{j=1}^{m} a_{ij} p_j, \qquad a_{ij} = r_{ij}^{-2} \max\{\cos\theta_{ij}, 0\}, \tag{6}$$

where r_{ij} denotes the distance between lamp j and the midpoint of patch i, and θ_{ij} denotes the angle between the upward normal of patch i and the vector from the midpoint of patch i to lamp j, as shown in the figure. This model takes into account "self-shading" (*i.e.*, the fact that a patch is illuminated only by lamps in the halfspace it faces) but not shading of one patch caused by another. Of course we could use a more complex illumination model, including shading and even reflections. This just changes the matrix relating the lamp powers to the patch illumination levels.

The problem is to determine lamp powers that make the illumination levels close to a given desired illumination level I_{des}, subject to the power limits $0 \le p_i \le 1$.

(a) Suppose we use the maximum deviation

$$\phi(p) = \max_{k=1,\ldots,n} |I_k - I_{\text{des}}|$$

as a measure for the deviation from the desired illumination level. Formulate the illumination problem using this criterion as a linear programming problem.

(b) There are several suboptimal approaches based on weighted least-squares. We consider two examples.

i. *Saturated least-squares.* We can solve the least-squares problem

$$\text{minimize} \quad \sum_{k=1}^{n} (I_k - I_{\text{des}})^2$$

ignoring the constraints. If the solution is not feasible, we saturate it, *i.e.*, set $p_j := 0$ if $p_j \le 0$ and $p_j := 1$ if $p_j \ge 1$.

Download the MATLAB file `ex16data.m` from the class webpage and generate problem data by `[A,Ides] = ex16data`. (The elements of A are the coefficients a_{ij} in (6).) Compute a feasible p using this first method, and calculate $\phi(p)$.

Note. Use the 'backslash' operator `\` to solve least-squares problem in MATLAB: `x=A\b` computes the solution of

$$\text{minimize} \quad \|Ax - b\|^2.$$

Try '`help slash`' for details.

ii. *Weighted least-squares.* We consider another least-squares problem:

$$\text{minimize} \quad \sum_{k=1}^{n} (I_k - I_{\text{des}})^2 + \mu \sum_{i=1}^{m} (p_i - 0.5)^2,$$

where $\mu \ge 0$ is used to attach a cost to a deviation of the powers from the value 0.5, which lies in the middle of the power limits. For large enough μ, the solution of this problem will satisfy $0 \le p_i \le 1$, *i.e.*, be feasible for the original problem.

Explain how you solve this problem in MATLAB. For the problem data generated by `ex16data.m`, find the smallest μ such that p becomes feasible, and evaluate $\phi(p)$.

(c) Using the same data as in part (b), solve the LP you derived in part (a). Compare the solution with the solutions you obtained using the (weighted) least-squares methods of part (b).

Exercise 17. In exercise 16, we encountered the problem

$$\begin{array}{ll} \text{minimize} & \max_{k=1,\ldots,n} |a_k^T p - I_{\text{des}}| \\ \text{subject to} & 0 \leq p \leq \mathbf{1} \end{array} \tag{7}$$

(with variables p). We have seen that this is readily cast as an LP.

In (7) we use the maximum of the *absolute deviations* $|I_k - I_{\text{des}}|$ to measure the difference from the desired intensity. Suppose we prefer to use the *relative deviations* instead, where the relative deviation is defined as

$$\max\{I_k/I_{\text{des}}, I_{\text{des}}/I_k\} - 1 = \begin{cases} (I_k - I_{\text{des}})/I_{\text{des}} & \text{if } I_k \geq I_{\text{des}} \\ (I_{\text{des}} - I_k)/I_k & \text{if } I_k \leq I_{\text{des}}. \end{cases}$$

This leads us to the following formulation:

$$\begin{array}{ll} \text{minimize} & \max_{k=1,\ldots,n} \max\{\, a_k^T p/I_{\text{des}},\ I_{\text{des}}/(a_k^T p)\,\} \\ \text{subject to} & 0 \leq p \leq \mathbf{1} \\ & a_k^T p > 0, \quad k = 1,\ldots,n. \end{array} \tag{8}$$

Explain how you can solve this using linear programming (*i.e.*, by solving one or more LPs).

Exercise 18. Download the file `ex18data.m` from the class website and execute it in MATLAB using the command `[t,y] = ex18data`. This will generate two vectors t, $y \in \mathbf{R}^{42}$. We are interested in fitting a linear function $f(t) = \alpha + \beta t$ through the points (t_i, y_i), *i.e.*, we want to select α and β such that $f(t_i) \approx y_i$, $i = 1,\ldots,42$.

We can calculate α and β by optimizing the following three criteria.

(a) *Least-squares:* select α and β by minimizing

$$\sum_{i=1}^{42} (y_i - \alpha - \beta t_i)^2.$$

(Note that the recommended method for solving a least-squares problem

$$\text{minimize} \quad \|Ax - b\|^2$$

in MATLAB is `x = A\b`.)

(b) ℓ_1*-norm approximation:* select α and β by minimizing

$$\sum_{i=1}^{42} |y_i - \alpha - \beta t_i|.$$

(c) ℓ_∞*-norm approximation:* select α and β by minimizing

$$\max_{i=1,\ldots,42} |y_i - \alpha - \beta t_i|.$$

Find the optimal values of α and β for each the three optimization criteria. This yields three linear functions $f_{\rm ls}(t)$, $f_{\ell_1}(t)$, $f_{\ell_\infty}(t)$. Plot the 42 data points, and the three functions f. What do you observe?

Exercise 19. This exercise is concerned with a problem that has applications in VLSI. The problem is to find the optimal positions of n cells or modules on an integrated circuit. More specifically, the variables are the coordinates x_i, y_i, $i = 1, \ldots, n$, of the n cells. The cells must be placed in a square $C = \{(x, y) \mid -1 \leq x \leq 1, -1 \leq y \leq 1\}$.

Each cell has several terminals, which are connected to terminals on other cells, or to input/output (I/O) terminals on the perimeter of C. The positions of the I/O terminals are known and fixed.

The connections between the cells are specified as follows. We are given a matrix $A \in \mathbf{R}^{N\times n}$ and two vectors $b_x \in \mathbf{R}^N$, $b_y \in \mathbf{R}^N$. Each row of A and each component of b_x and b_y describe a connection between two terminals. For each $i = 1, \ldots, N$, we can distinguish two possibilities, depending on whether row i of A describes a connection between two cells, or between a cell and an I/O terminal.

- If row i describes a connection between two cells j and k (with $j < k$), then

$$a_{il} = \begin{cases} 1 & \text{if } l = j \\ -1 & \text{if } l = k \\ 0 & \text{otherwise} \end{cases}, \qquad b_{x,i} = 0, \qquad b_{y,i} = 0.$$

  In other words, we have

$$a_i^T x - b_{x,i} = x_j - x_k, \qquad a_i^T y - b_{y,i} = y_j - y_k.$$

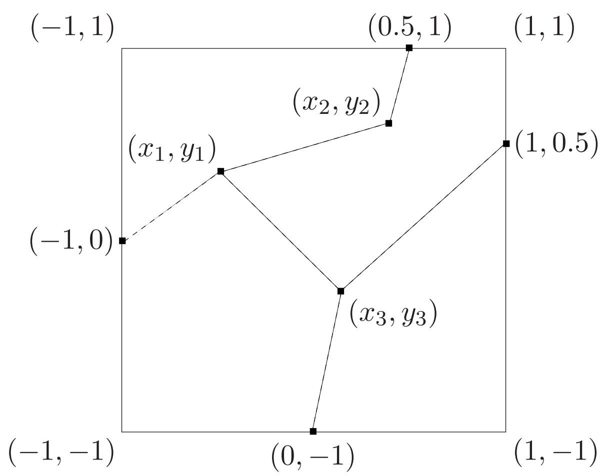
- If row i describes a connection between a cell j and an I/O terminal with coordinates $(\bar{x}, \bar{y})$, then

$$a_{il} = \begin{cases} 1 & \text{if } l = j \\ 0 & \text{otherwise} \end{cases}, \qquad b_{i,x} = \bar{x}, \qquad b_{i,y} = \bar{y},$$

  so we have

$$a_i^T x - b_{x,i} = x_j - \bar{x}, \qquad a_i^T y - b_{y,i} = y_j - \bar{y}.$$

The figure illustrates this notation for an example with $n = 3$, $N = 6$.



For this example, A, b_x and b_y given by

$$A = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \qquad b_x = \begin{bmatrix} 0.0 \\ 0.0 \\ -1.0 \\ 0.5 \\ 0.0 \\ 1.0 \end{bmatrix}, \qquad b_y = \begin{bmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 1.0 \\ -1.0 \\ 0.5 \end{bmatrix}.$$

The problem we consider is to determine the coordinates (x_i, y_i) that minimize some measure of the total wirelength of the connections. We can formulate different variations.

(a) Suppose we use the Euclidean distance between terminals to measure the length of a connection, and that we minimize the sum of the squares of the connection lengths. In other words we determine x and y by solving

$$\text{minimize} \quad \sum_{i=1}^{N} \left((a_i^T x - b_{x,i})^2 + (a_i^T y - b_{y,i})^2 \right)$$

or, in matrix notation,

$$\text{minimize} \quad \|Ax - b_x\|^2 + \|Ay - b_y\|^2. \tag{9}$$

The variables are $x \in \mathbf{R}^n$ and $y \in \mathbf{R}^n$. (Note that we don't have to add the constraints $-1 \le x_i \le 1$ and $-1 \le y_i \le 1$ explicitly, since a solution with a cell outside C can never be optimal.) Since the two terms in (9) are independent, the solution can be obtained by solving two least-squares problems, one to determine x, and one to determine y. Equivalently, we can solve two sets of linear equations

$$(A^T A)x = A^T b_x, \qquad (A^T A)y = A^T b_y.$$

(b) A second and more realistic choice is to use the Manhattan distance between two connected terminals as a measure for the length of the connection, *i.e.*, to consider the optimization problem

$$\text{minimize} \quad \sum_{i=1}^{N} \left(|a_i^T x - b_{x,i}| + |a_i^T y - b_{y,i}| \right).$$

In matrix notation, this can be written as

$$\text{minimize} \quad \|Ax - b_x\|_1 + \|Ay - b_y\|_1.$$

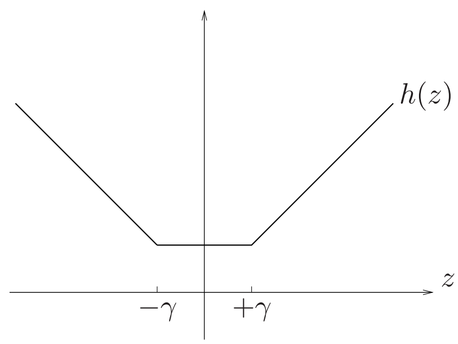
(c) As a third variation, suppose we measure the length of a connection between two terminals by the Manhattan distance between the two points, as in (b) , but instead of minimizing the sum of the lengths, we minimize the maximum length, *i.e.*, we solve

$$\text{minimize} \quad \max_{i=1,\ldots,N} \left(|a_i^T x - b_{x,i}| + |a_i^T y - b_{y,i}| \right).$$

(d) Finally, we can consider the problem

$$\text{minimize} \quad \sum_{i=1}^{N} \Big(h(a_i^T x - b_{x,i}) + h(a_i^T y - b_{y,i}) \Big)$$

where h is a piecewise-linear function defined as $h(z) = \max\{z, -z, \gamma\}$ and γ is a given positive constant. The function h is plotted below.



Give LP formulations for problems (b), (c) and (d). You may introduce new variables, but you must explain clearly why your formulation and the original problem are equivalent.

Numerical example. We compare the solutions obtained from the four variations for a small example. For simplicity, we consider a one-dimensional version of the problem, *i.e.*, the variables are $x \in \mathbf{R}^n$, and the goal is to place the cells on the interval $[-1, 1]$. We also drop the subscript in b_x. The four formulations of the one-dimensional placement problem are the following.

(a) ℓ_2-*placement:* minimize $\|Ax - b\|^2 = \sum_i (a_i^T x - b_i)^2$.

(b) ℓ_1-*placement:* minimize $\|Ax - b\|_1 = \sum_i |a_i^T x - b_i|$.

(c) ℓ_∞-*placement:* minimize $\|Ax - b\|_\infty = \max_i |a_i^T x - b_i|$.

(d) ℓ_1-*placement with 'dead zone':* minimize $\sum_i h(a_i^T x - b_i)$. We use a value $\gamma = 0.02$.

To generate the data, download the file `ex19data.m` from the class webpage. The command `[A,b] = ex19data('large')` generates a problem with 100 cells and 300 connections; `[A,b] = ex18data('small')` generates a problem with with 50 cells and 150 connections. You can choose either problem.

A few remarks and suggestions:

- Use the 'backslash' operator, `x = A\b`, to solve a least-squares problem

$$\text{minimize} \quad \|Ax - b\|^2.$$

  An alternative is `x = (A'*A)\(A'*b)`, which is somewhat faster but less accurate.
- The other three problems require an LP solver.

Compare the solutions obtained by the four methods.

- Plot a histogram of the n positions x_i for each solution (using the `hist` command).
- Also plot a histogram of the connnection lengths $|a_i^T x - b_i|$.

- Compute the total wire length $\sum_i |a_i^T x - b_i|$ for each of the four solutions.
- Compute the length of the longest connection $\max_i |a_i^T x - b_i|$ for each of the four solutions.
- So far we have assumed that the cells have zero width. In practice we have to take overlap between cells into account. Assume that two cells i and j overlap when $|x_i - x_j| \leq 0.01$. For each of the four solutions, calculate how many pairs of cells overlap. You can express the overlap as a percentage of the total number $n(n-1)/2$ of pairs of cells.

Are the results what you expect? Which of the four solutions would you prefer if the most important criteria are total wirelength $\sum_i |a_i^T x - b_i|$ and overlap?

Exercise 20. Suppose you are given two sets of points $\{v^1, v^2, \ldots, v^K\}$ and $\{w^1, w^2, \ldots, w^L\}$ in $\mathbf{R}^n$. Can you formulate the following two problems as LP feasibility problems?

(a) Determine a hyperplane that separates the two sets, *i.e.*, find $a \in \mathbf{R}^n$ and $b \in \mathbf{R}$ with $a \neq 0$ such that

$$a^T v^i \leq b, \quad i = 1, \ldots, K, \qquad a^T w^i \geq b, \quad i = 1, \ldots, L.$$

Note that we require $a \neq 0$, so you have to make sure your method does not return the trivial solution $a = 0$, $b = 0$. You can assume that the matrices

$$\begin{bmatrix} v^1 & v^2 & \cdots & v^K \\ 1 & 1 & \cdots & 1 \end{bmatrix}, \qquad \begin{bmatrix} w^1 & w^2 & \cdots & w^L \\ 1 & 1 & \cdots & 1 \end{bmatrix}$$

have rank $n+1$.

(b) Determine a sphere separating the two sets of points, *i.e.*, find $x_c \in \mathbf{R}^n$, $R \geq 0$ such that

$$(v^i - x_c)^T (v^i - x_c) \leq R^2, \quad i = 1, \ldots, K, \qquad (w^i - x_c)^T (w^i - x_c) \geq R^2, \quad i = 1, \ldots, L.$$

(x_c is the center of the sphere; R is its radius.)

Exercise 21. Download the file `ex21data.m` from the class website and run it in MATLAB using the command `[X,Y] = ex21data(id)`, where `id` is your student ID number (a nine-digit integer). This will create two matrices $X \in \mathbf{R}^{4\times 100}$ and $Y \in \mathbf{R}^{4\times 100}$. Let x_i and y_i be the ith columns of X and Y, respectively.

(a) Verify (prove) that it is impossible to strictly separate the points x_i from the points y_i by a hyperplane. In other words, show that there exist no $a \in \mathbf{R}^4$ and $b \in \mathbf{R}$ such that

$$a^T x_i + b \leq -1, \quad i = 1, \ldots, 100, \qquad a^T y_i + b \geq 1, \quad i = 1, \ldots, 100.$$

(b) Find a quadratic function that strictly separates the two sets, *i.e.*, find $A = A^T \in \mathbf{R}^{4\times 4}$, $b \in \mathbf{R}^4$, $c \in \mathbf{R}$, such that

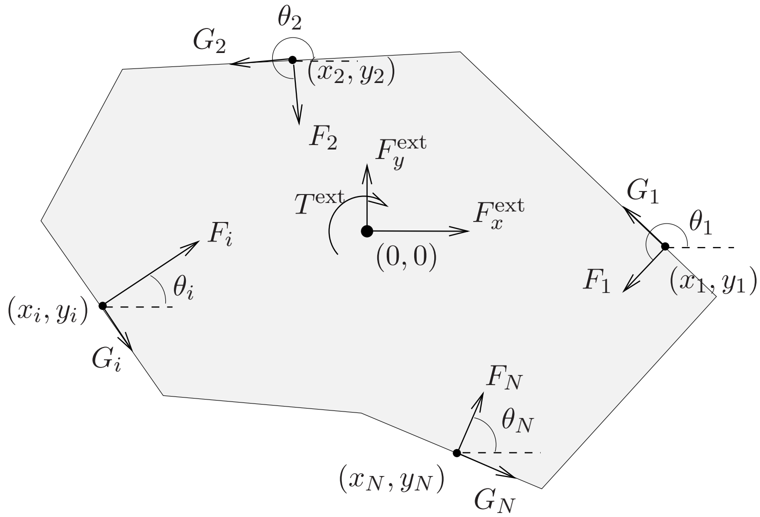
$$x_i^T A x_i + b^T x_i + c \leq -1, \quad i = 1, \ldots, 100, \qquad y_i^T A y_i + b^T y_i + c \geq 1, \quad i = 1, \ldots, 100.$$

(c) It may be impossible to find a hyperplane that strictly separates the two sets, but we can try to find a hyperplane that separates as many of the points as possible. Formulate a heuristic (*i.e.*, suboptimal method), based on solving a single LP, for finding $a \in \mathbf{R}^4$ and $b \in \mathbf{R}$ that minimize the number of misclassified points. We consider x_i as misclassified if $a^T x_i + b > -1$, and y_i as misclassified if $a^T y_i + b < 1$.

Describe and justify your method, and test it on the problem data.

Exercise 22. *Robot grasp problem with static friction.*

We consider a rigid object held by N robot fingers. For simplicity we assume that the object and all forces acting on it lie in a plane.



The fingers make contact with the object at points (x_i, y_i), $i = 1, \ldots, N$. (Although it does not matter, you can assume that the origin $(0,0)$ is at the center of gravity of the object.) Each finger applies a force with magnitude F_i on the object, in a direction normal to the surface at that contact point, and pointing towards the object. The horizontal component of the ith contact force is equal to $F_i \cos\theta_i$, and the vertical component is $F_i \sin\theta_i$, where θ_i is the angle between the inward pointing normal to the surface and a horizontal line.

At each contact point there is a friction force G_i which is tangential to the surface. The horizontal component is $G_i \sin\theta_i$ and the vertical component is $-G_i \cos\theta_i$. The orientation of the friction force is arbitrary (*i.e.*, G_i can be positive or negative), but its magnitude $|G_i|$ cannot exceed μF_i, where $\mu \geq 0$ is a given constant (the friction coefficient).

Finally, there are several external forces and torques that act on the object. We can replace those external forces by equivalent horizontal and vertical forces F_x^{ext} and F_y^{ext} at the origin, and an equivalent torque T^{ext}. These two external forces and the external torque are given.

The static equilibrium of the object is characterized by the following three equations:

$$\sum_{i=1}^{N} (F_i \cos\theta_i + G_i \sin\theta_i) + F_x^{\text{ext}} = 0 \tag{10}$$

(the horizontal forces add up to zero),

$$\sum_{i=1}^{N} (F_i \sin\theta_i - G_i \cos\theta_i) + F_y^{\text{ext}} = 0 \tag{11}$$

(the vertical forces add up to zero),

$$\sum_{i=1}^{N}((F_i\cos\theta_i + G_i\sin\theta_i)y_i - (F_i\sin\theta_i - G_i\cos\theta_i)x_i) + T^{\mathrm{ext}} = 0 \tag{12}$$

(the total torque is zero). As mentioned above, we assume the friction model can be expressed as a set of inequalities

$$|G_i| \le \mu F_i, \quad i = 1, \ldots, N. \tag{13}$$

If we had no friction, then $N = 3$ fingers would in general be sufficient to hold the object, and we could find the forces F_i by solving the three linear equations (10)-(12) for the variables F_i. If there is friction, or $N > 3$, we have more unkown forces than equilibrium equations, so the system of equations is underdetermined. We can then take advantage of the additional degrees of freedom to find a set of forces F_i that are 'small'. Express the following two problems as LPs.

(a) Find the set of forces F_i that minimizes $\sum_{i=1}^{N} F_i$ subject to the constraint that the object is in equilibrium. More precisely, the constraint is that there exist friction forces G_i that, together with F_i, satisfy (10)-(13).

(b) Find a set of forces F_i that minimizes $\max_{i=1,\ldots,N} F_i$ subject to the constraint that the object is in equilibrium.

Which of these two problems do you expect will have a solution with a larger number of F_i's equal to zero?

Exercise 23. *Linear programming in decision theory.* Suppose we have a choice of p available actions $a \in \{1, \ldots, p\}$, and each action has a certain cost (which can be positive, negative or zero). The costs depend on the value of an unknown parameter $\theta \in \{1, \ldots, m\}$, and are specified in the form of a *loss matrix* $L \in \mathbf{R}^{m\times p}$, with L_{ij} equal to the cost of action $a = j$ when $\theta = i$.

We do not know θ, but we can observe a random variable x with a distribution that depends on θ. We will assume that x is a discrete random variable with values in $\{1, 2, \ldots, n\}$, so we can represent its distribution, for the m possible values of θ, by a matrix $P \in \mathbf{R}^{n\times m}$ with

$$P_{ki} = \mathbf{prob}(x = k \,|\theta = i).$$

A *strategy* is a rule for selecting an action a based on the observed value of x. A *pure* or *deterministic* strategy assigns to each of the possible observations a unique action a. A pure stratey can be represented by a matrix $T \in \mathbf{R}^{p\times n}$, with

$$T_{jk} = \begin{cases} 1 & \text{action } j \text{ is selected when } x = k \text{ is observed} \\ 0 & \text{otherwise.} \end{cases}$$

Note that each column of a pure strategy matrix T contains exactly one entry equal to one, and the other entries are zero. We can therefore enumerate all possible pure strategies by enumerating the 0-1 matrices with this property.

As a generalization, we can consider *mixed* or *randomized strategies*. In a mixed strategy we select an action randomly, using a distribution that depends on the observed x. A mixed strategy is represented by a matrix $T \in \mathbf{R}^{p\times n}$, with

$$T_{jk} = \mathbf{prob}(a = j \mid x = k).$$

The entries of a mixed strategy matrix T are nonnegative and have column sums equal to one:

$$T_{jk} \geq 0, \quad j = 1, \ldots, p, \quad k = 1, \ldots, n, \qquad \mathbf{1}^T T = \mathbf{1}^T.$$

A pure strategy is a special case of a mixed strategy with all the entries T_{jk} equal to zero or one.

Now suppose the value of θ is i and we apply the strategy T. Then the expected loss is given by

$$\sum_{k=1}^{n} \sum_{j=1}^{p} L_{ij} T_{jk} P_{ki} = (LTP)_{ii}.$$

The diagonal elements of the matrix LTP are the expected losses for the different values of $\theta = 1, \ldots, m$. We consider two popular definitions of an optimal mixed strategy, based on minimizing a function of the expected losses.

(a) *Minimax strategies.* A minimax strategy minimizes the maximum of the expected losses: the matrix T is computed by solving

$$\begin{array}{ll} \text{minimize} & \max_{i=1,\ldots,m} (LTP)_{ii} \\ \text{subject to} & T_{jk} \geq 0, \quad j = 1, \ldots, p, \quad k = 1, \ldots, n \\ & \mathbf{1}^T T = \mathbf{1}^T. \end{array}$$

The variables are the pn entries of T. Express this problem as a linear program.

(b) *Bayes strategies.* Assume that the parameter θ itself is random with a known distribution $q_i = \mathbf{prob}(\theta = i)$. The Bayes strategy minimizes the *average expected loss*, where the average is taken over θ. The matrix T of a Bayes strategy is the optimal solution of the problem

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{m} q_i (LTP)_{ii} \\ \text{subject to} & T_{jk} \geq 0, \quad j = 1, \ldots, p, \quad k = 1, \ldots, n \\ & \mathbf{1}^T T = \mathbf{1}^T. \end{array}$$

This is a linear program in the pn variables T_{jk}. Formulate a simple algorithm for solving this LP. Show that it is always possible to find an optimal Bayes strategy that is a pure strategy.

Hint. First note that each column of the optimal T can be determined independently of the other columns. Then reduce the optimization problem over column k of T to one of the simple LPs in Exercise 8.

(c) As a simple numerical example, we consider a quality control system in a factory. The products that are examined can be in one of two conditions ($m = 2$): $\theta = 1$ means the product is defective; $\theta = 2$ means the product works properly. To examine the quality of a product we use an automated measurement system that rates the product on a scale of 1 to 4. This rating is the observed variable x: $n = 4$ and $x \in \{1, 2, 3, 4\}$. We have calibrated the system to find the probabilities $P_{ij} = \mathbf{prob}(x = i \mid \theta = j)$ of producing a rating $x = i$ when the state of the product is $\theta = j$. The matrix P is

$$P = \begin{bmatrix} 0.7 & 0.0 \\ 0.2 & 0.1 \\ 0.05 & 0.1 \\ 0.05 & 0.8 \end{bmatrix}.$$

We have a choice of three possible actions ($p = 3$): $a = 1$ means we accept the product and forward it to be sold; $a = 2$ means we subject it to a manual inspection to determine whether it is defective or not; $a = 3$ means we discard the product. The loss matrix is

$$L = \begin{bmatrix} 10 & 3 & 1 \\ 0 & 2 & 6 \end{bmatrix}.$$

Thus, for example, selling a defective product costs us \$10; discarding a good product costs \$6, et cetera.

i. Compute the minimax strategy for this L and P (using an LP solver). Is the minimax strategy a pure strategy?
ii. Compute the Bayes strategy for $q = (0.2, 0.8)$ (using an LP solver or the simple algorithm formulated in part 2).
iii. Enumerate all ($3^4 = 81$) possible pure strategies T (in MATLAB), and plot the expected losses $((LTP)_{11}, (LTP)_{22})$ of each of these strategies in a plane.
iv. On the same graph, show the losses for the minimax strategy and the Bayes strategy computed in parts (a) and (b).
v. Suppose we let q vary over all possible prior distributions (all vectors with $q_1 + q_2 = 1$, $q_1 \geq 0$, $q_2 \geq 0$). Indicate on the graph the expected losses $((LTP)_{11}, (LTP)_{22})$ of the corresponding Bayes strategies.

Exercise 24. *Robust linear programming.*

(a) Let $x \in \mathbf{R}^n$ be a given vector. Prove that $x^T y \leq \|x\|_1$ for all y with $\|y\|_\infty \leq 1$. Is the inequality tight, *i.e.*, does there exist a y that satisfies $\|y\|_\infty \leq 1$ and $x^T y = \|x\|_1$?

(b) Consider the set of linear inequalities

$$a_i^T x \leq b_i, \quad i = 1, \ldots, m. \tag{14}$$

Suppose you don't know the coefficients a_i exactly. Instead you are given nominal values $\bar{a}_i$, and you know that the actual coefficient vectors satisfy

$$\|a_i - \bar{a}_i\|_\infty \leq \rho$$

for a given $\rho > 0$. In other words the actual coefficients a_{ij} can be anywhere in the intervals $[\bar{a}_{ij} - \rho, \bar{a}_{ij} + \rho]$, or equivalently, each vector a_i can lie anywhere in a rectangle with corners $\bar{a}_i + v$ where $v \in \{-\rho, \rho\}^n$ (*i.e.*, v has components ρ or $-\rho$).

The set of inequalities (14) must be satisfied for all possible values of a_i, *i.e.*, we replace (14) with the constraints

$$a_i^T x \leq b_i \text{ for all } a_i \in \{\bar{a}_i + v \mid \|v\|_\infty \leq \rho\} \text{ and for } i = 1, \ldots, m. \tag{15}$$

A straightforward but very inefficient way to express this constraint is to enumerate the 2^n corners of the rectangle of possible values a_i and to require that

$$\bar{a}_i^T x + v^T x \leq b_i \text{ for all } v \in \{-\rho, \rho\}^n \text{ and for } i = 1, \ldots, m.$$

This is a system of $m2^n$ inequalities.

Use the result in (a) to show that (15) is in fact equivalent to the much more compact set of nonlinear inequalities

$$\bar{a}_i^T x + \rho\|x\|_1 \leq b_i, \quad i = 1, \ldots, m. \tag{16}$$

(c) Consider the LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \leq b_i, \quad i = 1, \ldots, m. \end{array}$$

Again we are interested in situations where the coefficient vectors a_i are uncertain, but satisfy bounds $\|a_i - \bar{a}_i\|_\infty \leq \rho$ for given $\bar{a}_i$ and ρ. We want to minimize $c^T x$ subject to the constraint that the inequalities $a_i^T x \leq b_i$ are satisfied for *all* possible values of a_i. We call this a *robust LP*:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \leq b_i \text{ for all } a_i \in \{\bar{a}_i + v \mid \|v\|_\infty \leq \rho\} \text{ and for } i = 1, \ldots, m. \end{array} \tag{17}$$

It follows from (b) that we can express this problem as a nonlinear optimization problem

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \bar{a}_i^T x + \rho\|x\|_1 \leq b_i, \quad i = 1, \ldots, m. \end{array} \tag{18}$$

Express (18) as an LP.

Solving (18) is a *worst-case* approach to dealing with uncertainty in the data. If $x^\star$ is the optimal solution of (18), then for any *specific* value of a_i, it may be possible to find feasible x with a lower objective value than $x^\star$. However such an x would be infeasible for some other value of a_i.

Exercise 25. *Robust Chebyshev approximation.*

In a similar way as in the previous problem, we can consider Chebyshev approximation problems

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

in which $A \in \mathbf{R}^{m\times n}$ is uncertain. Suppose we can characterize the uncertainty as follows. The values of A depend on parameters $u \in \mathbf{R}^p$, which are unknown but satisfy $\|u\|_\infty \leq \rho$. Each row vector a_i can be written as $a_i = \bar{a}_i + B_i u$ where $\bar{a}_i \in \mathbf{R}^n$ and $B_i \in \mathbf{R}^{n\times p}$ are given. In the *robust Chebyshev approximation* we minimize the worst-case value of $\|Ax - b\|_\infty$. This problem can be written as

$$\text{minimize} \quad \max_{\|u\|_\infty \leq \rho} \max_{i=1,\ldots,m} \left|(\bar{a}_i + B_i u)^T x - b_i\right|. \tag{19}$$

Show that (19) is equivalent to

$$\text{minimize} \quad \max_{i=1,\ldots,m}(|\bar{a}_i^T x - b_i| + \rho\|B_i^T x\|_1). \tag{20}$$

To prove this you can use the results from exercise 24. There is also a fairly straightforward direct proof. Express (20) as an LP.

Exercise 26. Describe how you would use linear programming to solve the following problem. You are given an LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \leq b \end{array} \tag{21}$$

in which the coefficients of $A \in \mathbf{R}^{m\times n}$ are uncertain. Each coefficient A_{ij} can take arbitrary values in the interval

$$[\bar{A}_{ij} - \Delta A_{ij}, \bar{A}_{ij} + \Delta A_{ij}],$$

where $\bar{A}_{ij}$ and ΔA_{ij} are given with $\Delta A_{ij} \geq 0$. The optimization variable x in (21) must be feasible for all possible values of A. In other words, we want to solve

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \leq b \text{ for all } A \in \mathcal{A} \end{array}$$

where $\mathcal{A} \subseteq \mathbf{R}^{m \times n}$ is the set

$$\mathcal{A} = \{A \in \mathbf{R}^{m \times n} \mid \bar{A}_{ij} - \Delta A_{ij} \leq A_{ij} \leq \bar{A}_{ij} + \Delta A_{ij},\ i = 1, \ldots, m,\ j = 1, \ldots, n\}.$$

If you know more than one solution method, you should give the most efficient one.

Exercise 27. Optimization problems with uncertain data sometimes involve two sets of variables that can be selected in two stages. When the first set of variables is chosen, the problem data are uncertain. The second set of variables, however, can be selected *after* the actual values of the parameters have become known.

As an example, we consider two-stage robust formulations of the Chebyshev approximation problem

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

with variables $x \in \mathbf{R}^n$ and $y \in \mathbf{R}^p$. The problem parameters A, B, b are uncertain, and we model the uncertainty by assuming that there are m possible scenarios (or instances of the problem). In scenario k, the values of A, B, b are A_k, B_k, b_k.

In the two-stage setting we first select x before the scenario is known; then we choose y after learning the actual value of k. The optimal choice of y in the second stage is the value that minimizes $\|A_k x + B_k y + b_k\|_\infty$, for given x, A_k, B_k, b_k. We denote by $f_k(x)$ the optimal value of this second-stage optimization problem for scenario k:

$$f_k(x) = \min_y \|A_k x + B_k y + b_k\|_\infty, \quad k = 1, \ldots, m.$$

(a) We can minimize the worst-case objective by solving the optimization problem

$$\text{minimize} \quad \max_{k=1,\ldots,m} f_k(x)$$

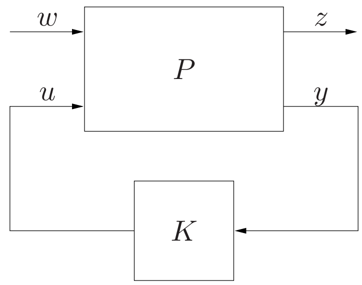
with x as variable. Formulate this problem as an LP.

(b) If we know the probability distribution of the scenarios we can also minimize the expected cost, by solving

$$\text{minimize} \quad \sum_{k=1}^{m} \pi_k f_k(x)$$

with x as variable. The coefficient $\pi_k \geq 0$ is the probability that (A, B, b) is equal to (A_k, B_k, b_k). Formulate this problem as an LP.

Exercise 28. *Feedback design for a static linear system.* In this problem we use linear programming to design a linear feedback controller for a static linear system. (The method extends to dynamical systems but we will not consider the extension here.) The figure shows the system and the controller.



The elements of the vector $w \in \mathbf{R}^{n_w}$ are called the *exogeneous inputs*, $z \in \mathbf{R}^{n_z}$ are the *critical outputs*, $y \in \mathbf{R}^{n_y}$ are the *sensed outputs*, and $u \in \mathbf{R}^{n_u}$ are the *actuator inputs*. These vectors are related as

$$\begin{aligned} z &= P_{zw}w + P_{zu}u \\ y &= P_{yw}w + P_{yu}u, \end{aligned} \tag{22}$$

where the matrices P_{zu}, P_{zw}, P_{yu}, P_{yw} are given.

The controller feeds back the sensed outputs y to the actuator inputs u. The relation is

$$u = Ky \tag{23}$$

where $K \in \mathbf{R}^{n_u \times n_y}$. The matrix K will be the design variable.

Assuming $I - P_{yu}$ is invertible, we can eliminate y from the second equation in (22). We have

$$y = (I - P_{yu}K)^{-1}P_{yw}w$$

and substituting in the first equation we can write $z = Hw$ with

$$H = P_{zw} + P_{zu}K(I - P_{yu}K)^{-1}P_{yw}. \tag{24}$$

The matrix H is a complicated nonlinear function of K.

Suppose that the signals w are disturbances or noises acting on the system, and that they can take any values with $\|w\|_\infty \leq \rho$ for some given ρ. We would like to choose K so that the effect of the disturbances w on the output z is minimized, *i.e.*, we would like z to be as close as possible to zero, regardless of the values of w. Specifically, if we use the infinity norm $\|z\|_\infty$ to measure the size of z, we are interested in determining K by solving the optimization problem

$$\text{minimize} \quad \max_{\|w\|_\infty \leq \rho} \|Hw\|_\infty, \tag{25}$$

where H depends on the variable K through the formula (24).

(a) We first derive an explicit expression for the objective function in (25). Show that

$$\max_{\|w\|_\infty \leq \rho} \|Hw\|_\infty = \rho \max_{i=1,\ldots,n_z} \sum_{j=1,\ldots,n_w} |h_{ij}|$$

where h_{ij} are the elements of H. Up to the constant ρ, this is the maximum row sum of H: for each row of H we calculate the sum of the absolute values of its elements; we then select the largest of these row sums.

(b) Using this expression, we can reformulate problem (25) as

$$\text{minimize} \quad \rho \max_{i=1,\ldots,n_z} \sum_{j=1,\ldots,n_w} |h_{ij}|, \tag{26}$$

where h_{ij} depends on the variable K through the formula (24). Formulate (25) as an LP. *Hint.* Use a change of variables

$$Q = K(I - P_{yu}K)^{-1},$$

and optimize over $Q \in \mathbf{R}^{n_u \times n_y}$ instead of K. You may assume that $I+QP_{yu}$ is invertible, so the transformation is invertible: we can find K from Q as $K = (I + QP_{yu})^{-1}Q$.

Exercise 29. Formulate the following problem as an LP. Find the largest ball

$$\mathcal{B}(x_c, R) = \{x \mid \|x - x_c\| \leq R\}$$

enclosed in a given polyhedron

$$\mathcal{P} = \{x \mid a_i^T x \leq b_i, \ \ i = 1, \ldots, m\}.$$

In other words, express the problem

$$\begin{array}{ll} \text{maximize} & R \\ \text{subject to} & \mathcal{B}(x_c, R) \subseteq \mathcal{P} \end{array}$$

as an LP. The problem variables are the center $x_c \in \mathbf{R}^n$ and the radius R of the ball.

Exercise 30. Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be two polyhedra described as

$$\mathcal{P}_1 = \{x \mid Ax \leq b\}, \qquad \mathcal{P}_2 = \{x \mid -\mathbf{1} \leq Cx \leq \mathbf{1}\},$$

where $A \in \mathbf{R}^{m\times n}$, $C \in \mathbf{R}^{p\times n}$, and $b \in \mathbf{R}^m$. The polyhedron $\mathcal{P}_2$ is symmetric about the origin, *i.e.*, if $x \in \mathcal{P}_2$, then $-x \in \mathcal{P}_2$. We say the origin is the *center* of $\mathcal{P}_2$.

For $t > 0$ and $x_c \in \mathbf{R}^n$, we use the notation $t\mathcal{P}_2 + x_c$ to denote the polyhedron

$$t\mathcal{P}_2 + x_c = \{tx + x_c \mid x \in \mathcal{P}_2\},$$

which is obtained by first scaling $\mathcal{P}_2$ by a factor t about the origin, and then translating its center to x_c.

Explain how you would solve the following two problems using linear programming. If you know different formulations, you should choose the most efficient method.

(a) Find the largest polyhedron $t\mathcal{P}_2 + x_c$ enclosed in $\mathcal{P}_1$, *i.e.*,

$$\begin{array}{ll} \text{maximize} & t \\ \text{subject to} & t\mathcal{P}_2 + x_c \subseteq \mathcal{P}_1. \end{array}$$

(b) Find the smallest polyhedron $t\mathcal{P}_2 + x_c$ containing $\mathcal{P}_1$, *i.e.*,

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & \mathcal{P}_1 \subseteq t\mathcal{P}_2 + x_c. \end{array}$$

In both problems the variables are $t \in \mathbf{R}$ and $x_c \in \mathbf{R}^n$.

Exercise 31. Consider the linear system of exercise 9, equation (4). We study two optimal control problems. In both problems we assume the system is initially at rest at the origin, *i.e.*, $z(0) = 0$.

(a) In the first problem we want to determine the most efficient input sequence $u(kT)$, $k = 0, \ldots, 79$, that brings the system to state $(0, 0, 10, 10)$ in 80 time periods (*i.e.*, at $t = 8$ the two masses should be at rest at position $v_1 = v_2 = 10$). We assume the cost (*e.g.*, fuel consumption) of the input signal u is proportional to $\sum_k |u(kT)|$. We also impose the constraint that the amplitude of the input must not exceed 2. This leads us to the following problem:

$$\begin{array}{ll} \text{minimize} & \sum_{k=0}^{79} |u(kT)| \\ \text{subject to} & z(80T) = (0, 0, 10, 10) \\ & |u(kT)| \leq 2, \quad k = 0, \ldots, 79. \end{array} \tag{27}$$

The state z and the input u are related by (4) with $z(0) = 0$. The variables in (27) are $u(0)$, $u(T)$, $\ldots$, $u(79T)$.

(b) In the second problem we want to bring the system to the state $(0, 0, 10, 10)$ as quickly as possible, subject to the limit on the magnitude of u:

$$\begin{array}{ll} \text{minimize} & N \\ \text{subject to} & z(NT) = (0, 0, 10, 10) \\ & |u(kT)| \leq 2, \quad k = 0, \ldots, N-1. \end{array}$$

The variables are $N \in \mathbf{Z}$, and $u(0)$, $u(T)$, $\ldots$, $u(N-1)T$.

Solve these two problems numerically. Plot the input u and the positions v_1, v_2 as functions of time.

Exercise 32. We consider a linear dynamical system with state $x(t) \in \mathbf{R}^n$, $t = 0, \ldots, N$, and actuator or input signal $u(t) \in \mathbf{R}$, for $t = 0, \ldots, N-1$. The dynamics of the system is given by the linear recurrence

$$x(t+1) = Ax(t) + bu(t), \quad t = 0, \ldots, N-1,$$

where $A \in \mathbf{R}^{n \times n}$ and $b \in \mathbf{R}^n$ are given. We assume that the initial state is zero, *i.e.*, $x(0) = 0$.

The *minimum fuel optimal control problem* is to choose the inputs $u(0), \ldots, u(N-1)$ so as to minimize the total fuel consumed, which is given by

$$F = \sum_{t=0}^{N-1} f(u(t)),$$

subject to the constraint that $x(N) = x_{\text{des}}$, where N is the (given) time horizon, and $x_{\text{des}} \in \mathbf{R}^n$ is the (given) final or target state. The function $f : \mathbf{R} \to \mathbf{R}$ is the *fuel use map* for the actuator, which gives the amount of fuel used as a function of the actuator signal amplitude. In this problem we use

$$f(a) = \begin{cases} |a| & |a| \leq 1 \\ 2|a| - 1 & |a| > 1. \end{cases}$$

This means that fuel use is proportional to the absolute value of the actuator signal, for actuator signals between -1 and 1; for larger actuator signals the marginal fuel efficiency is half.

(a) Formulate the minimum fuel optimal control problem as an LP.

(b) Solve the following instance of the problem:

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 0.95 \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ 0.1 \end{bmatrix}, \quad x(0) = (0, 0), \quad x_{\rm des} = (10, 0), \quad N = 20.$$

We can interpret the system as a simple model of a vehicle moving in one dimension. The state dimension is $n = 2$, with $x_1(t)$ denoting the position of the vehicle at time t, and $x_2(t)$ giving its velocity. The initial state is $(0, 0)$, which corresponds to the vehicle at rest at position 0; the final state is $x_{\rm des} = (10, 0)$, which corresponds to the vehicle being at rest at position 10. Roughly speaking, this means that the actuator input affects the velocity, which in turn affects the position. The coefficient $A_{22} = 0.95$ means that velocity decays by 5% in one sample period, if no actuator signal is applied.

Plot the input signal $u(t)$ for $t = 0, \ldots, 19$, and the position and velocity (*i.e.*, $x_1(t)$ and $x_2(t)$) for $t = 0, \ldots, 20$.

Exercise 33. *Approximating a matrix in infinity norm.* The infinity (induced) norm of a matrix $A \in \mathbf{R}^{m\times n}$, denoted $\|A\|_{\infty,i}$, is defined as

$$\|A\|_{\infty,i} = \max_{i=1,\ldots,m} \sum_{j=1}^{n} |a_{ij}|.$$

The infinity norm gives the maximum ratio of the infinity norm of Ax to the infinity norm of x:

$$\|A\|_{\infty,i} = \max_{x\neq 0} \frac{\|Ax\|_\infty}{\|x\|_\infty}.$$

This norm is sometimes called the max-row-sum norm, for obvious reasons.

Consider the problem of approximating a matrix, in the max-row-sum norm, by a linear combination of other matrices. That is, we are given $k + 1$ matrices $A_0, \ldots, A_k \in \mathbf{R}^{m\times n}$, and need to find $x \in \mathbf{R}^k$ that minimizes

$$\|A_0 + x_1A_1 + \cdots + x_kA_k\|_{\infty,i}.$$

Express this problem as a linear program. Explain the significance of any extra variables in your LP. Carefully explain why your LP formulation solves this problem, *e.g.*, what is the relation between the feasible set for your LP and this problem?

Exercise 34. We are given p matrices $A_i \in \mathbf{R}^{n\times n}$, and we would like to find a single matrix $X \in \mathbf{R}^{n\times n}$ that we can use as an approximate right-inverse for each matrix A_i, *i.e.*, we would like to have

$$A_iX \approx I, \quad i = 1, \ldots, p.$$

We can do this by solving the following optimization problem with X as variable:

$$\text{minimize} \quad \max_{i=1,\ldots,p} \|I - A_iX\|_\infty. \tag{28}$$

Here $\|H\|_\infty$ is the 'infinity-norm' or 'max-row-sum norm' of a matrix H, defined as

$$\|H\|_\infty = \max_{i=1,\ldots,m} \sum_{j=1}^{n} |H_{ij}|,$$

if $H \in \mathbf{R}^{m\times n}$.

Express the problem (28) as an LP. You don't have to reduce the LP to a canonical form, as long as you are clear about what the variables are, what the meaning is of any auxiliary variables that you introduce, and why the LP is equivalent to the problem (28).

Exercise 35. Explain how you would use linear programming to solve the following optimization problems.

(a) Given $A \in \mathbf{R}^{m\times n}$, $b \in \mathbf{R}^m$,

$$\text{minimize} \ \sum_{i=1}^{m} \max\{0, a_i^T x + b_i\}.$$

The variable is $x \in \mathbf{R}^n$.

(b) Given $A \in \mathbf{R}^{m\times n}$, $b \in \mathbf{R}^m$,

$$\begin{array}{ll} \text{minimize} & \max_{i=1,\ldots,m} \max\{a_i^T x + b_i, 1/(a_i^T x + b_i)\} \\ \text{subject to} & Ax + b > 0. \end{array}$$

The variable is $x \in \mathbf{R}^n$.

(c) Given m numbers $a_1, a_2, \ldots, a_m \in \mathbf{R}$, and two vectors $l, u \in \mathbf{R}^m$, find the polynomial $f(t) = c_0 + c_1 t + \cdots + c_n t^n$ of lowest degree that satisfies the bounds

$$l_i \le f(a_i) \le u_i, \quad i = 1, \ldots, m.$$

The variables in the problem are the coefficients c_i of the polynomial.

(d) Given $p+1$ matrices $A_0, A_1, \ldots, A_p \in \mathbf{R}^{m\times n}$, find the vector $x \in \mathbf{R}^p$ that minimizes

$$\max_{\|y\|_1=1} \|(A_0 + x_1 A_1 + \cdots + x_p A_p)y\|_1.$$

Exercise 36. Suppose you are given an infeasible set of linear inequalities

$$a_i^T x \le b_i, \quad i = 1, \ldots, m,$$

and you are asked to find an x that satisfies many of the inequalities (ideally, as many as possible). Of course, the exact solution of this problem is difficult and requires combinatorial or integer optimization techniques, so you should concentrate on heuristic or sub-optimal methods. More specifically, you are asked to formulate a heuristic method based on solving a single LP.

Test the method on the example problem in the file `ex36data.m` available on the class webpage. (The MATLAB command `[A,b] = ex36data` generates a sparse matrix $A \in \mathbf{R}^{100\times 50}$ and a vector $b \in \mathbf{R}^{100}$, that define an infeasible set of linear inequalities.) To count the number of inequalities satisfied by x, you can use the MATLAB command

```
length(find(b-A*x > -1e-5)).
```

Exercise 37. Consider the linear-fractional program

$$\begin{array}{ll} \text{minimize} & (c^T x + \gamma)/(d^T x + \delta) \\ \text{subject to} & Ax \leq b, \end{array} \qquad (29)$$

where $A \in \mathbf{R}^{m\times n}$, $b \in \mathbf{R}^m$, $c, d \in \mathbf{R}^n$, and $\gamma, \delta \in \mathbf{R}$. We assume that the polyhedron

$$\mathcal{P} = \{x \in \mathbf{R}^n \mid Ax \leq b\}$$

is bounded and that $d^T x + \delta > 0$ for all $x \in \mathcal{P}$.

Show that you can solve (29) by solving the LP

$$\begin{array}{ll} \text{minimize} & c^T y + \gamma z \\ \text{subject to} & Ay - zb \leq 0 \\ & d^T y + \delta z = 1 \\ & z \geq 0 \end{array} \qquad (30)$$

in the variables $y \in \mathbf{R}^n$ and $z \in \mathbf{R}$. More precisely, suppose $\hat{y}$ and $\hat{z}$ are a solution of (30). Show that $\hat{z} > 0$ and that $\hat{y}/\hat{z}$ solves (29).

Exercise 38. Consider the problem

$$\begin{array}{ll} \text{minimize} & \|Ax - b\|_1/(c^T x + d) \\ \text{subject to} & \|x\|_\infty \leq 1, \end{array}$$

where $A \in \mathbf{R}^{m\times n}$, $b \in \mathbf{R}^m$, $c \in \mathbf{R}^n$, and $d \in \mathbf{R}$. We assume that $d > \|c\|_1$.

(a) Formulate this problem as a linear-fractional program.

(b) Show that $d > \|c\|_1$ implies that $c^T x + d > 0$ for all feasible x.

(c) Show that the problem is equivalent to the convex optimization problem

$$\begin{array}{ll} \text{minimize} & \|Ay - bt\|_1 \\ \text{subject to} & \|y\|_\infty \leq t \\ & c^T y + dt = 1, \end{array} \qquad (31)$$

with variables $y \in \mathbf{R}^n$, $t \in \mathbf{R}$.

(d) Formulate problem (31) as an LP.

Exercise 39. Explain how you would solve the following problem using linear programming. You are given two sets of points in $\mathbf{R}^n$:

$$S_1 = \{x_1, \ldots, x_N\}, \qquad S_2 = \{y_1, \ldots, y_M\}.$$

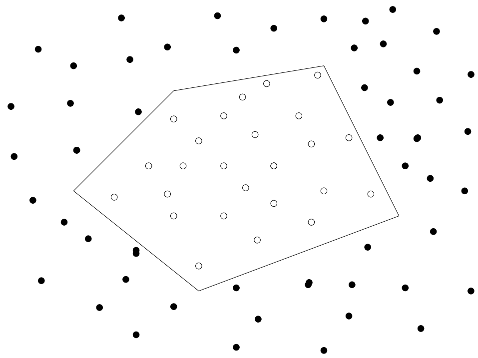
You are asked to find a polyhedron

$$\mathcal{P} = \{x \mid a_i^T x \leq b_i,\ i = 1, \ldots, m\}$$

that contains the points in S_1 in its interior, and does not contain any of the points in S_2:

$$S_1 \subseteq \{x \mid a_i^T x < b_i,\ i = 1, \ldots, m\}, \qquad S_2 \subseteq \{x \mid a_i^T x > b_i \text{ for at least one } i\} = \mathbf{R}^n \setminus \mathcal{P}.$$

An example is shown in the figure, with the points in S_1 shown as open circles and the points in S_2 as filled circles.



You can assume that the two sets are separable in the way described. Your solution method should return a_i and b_i, $i = 1, \ldots, m$, given the sets S_1 and S_2. The number of inequalities m is not specified, but it should not exceed $M + N$. You are allowed to solve one or more LPs or LP feasibility problems. The method should be efficient, *i.e.*, the dimensions of the LPs you solve should not be exponential as a function of N and M.

Exercise 40. Explain how you would solve the following problem using linear programming. Given two polyhedra

$$\mathcal{P}_1 = \{x \mid Ax \leq b\}, \qquad \mathcal{P}_2 = \{x \mid Cx \leq d\},$$

prove that $\mathcal{P}_1 \subseteq \mathcal{P}_2$, or find a point in $\mathcal{P}_1$ that is not in $\mathcal{P}_2$. The matrices $A \in \mathbf{R}^{m\times n}$ and $C \in \mathbf{R}^{p\times n}$, and the vectors $b \in \mathbf{R}^m$ and $d \in \mathbf{R}^p$ are given.

If you know several solution methods, give the most efficient one.

Exercise 41. Formulate the following problem as an LP:

$$\begin{array}{ll} \text{maximize} & \sum_{j=1}^{n} r_j(x_j) \\ \text{subject to} & \sum_{j=1}^{n} A_{ij}x_j \leq c_i^{\max}, \quad i = 1, \ldots, m \\ & x_j \geq 0, \quad j = 1, \ldots, n. \end{array} \tag{32}$$

The functions r_j are defined as

$$r_j(u) = \begin{cases} p_j u & 0 \leq u \leq q_j \\ p_j q_j + p_j^{\text{disc}}(u - q_j) & u \geq q_j, \end{cases} \tag{33}$$

where $p_j > 0$, $q_j > 0$ and $0 < p_j^{\text{disc}} < p_j$. The variables in the problem are x_j, $j = 1, \ldots, n$. The parameters A_{ij}, $c_i^{\max}$, p_j, q_j and p_j^{disc} are given.

The variables x_j in the problem represent activity levels (for example, production levels for different products manufactured by a company). These activities consume m resources, which are limited. Activity j consumes $A_{ij}x_j$ of resource i. (Ordinarily we have $A_{ij} \geq 0$, *i.e.*, activity j consumes resource i. But we allow the possibility that $A_{ij} < 0$, which means that activity j actually *generates* resource i as a by-product.) The total resource consumption is additive, so the total of resource i consumed is $c_i = \sum_{j=1}^{n} A_{ij}x_j$. Each resource consumption is limited: we must have $c_i \leq c_i^{\max}$, where $c_i^{\max}$ are given.

Activity j generates revenue $r_j(x_j)$, given by the expression (33). In this definition $p_j > 0$ is the basic price, $q_j > 0$ is the quantity discount level, and p_j^{disc} is the quantity discount price, for (the product of) activity j. We have $0 < p_j^{\mathrm{disc}} < p_j$. The total revenue is the sum of the revenues associated with each activity, *i.e.*, $\sum_{j=1}^n r_j(x_j)$. The goal in (32) is to choose activity levels that maximize the total revenue while respecting the resource limits.

5 Duality

Exercise 42. The main result of linear programming duality is that the optimal value of the LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \leq b \end{array}$$

is equal to the optimal value of the LP

$$\begin{array}{ll} \text{maximize} & -b^T z \\ \text{subject to} & A^T z + c = 0 \\ & z \geq 0, \end{array}$$

except when they are both infeasible. Give an example in which both problems are infeasible.

Exercise 43. Consider the LP

$$\begin{array}{ll} \text{minimize} & 47x_1 + 93x_2 + 17x_3 - 93x_4 \\ \text{subject to} & \begin{bmatrix} -1 & -6 & 1 & 3 \\ -1 & -2 & 7 & 1 \\ 0 & 3 & -10 & -1 \\ -6 & -11 & -2 & 12 \\ 1 & 6 & -1 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \leq \begin{bmatrix} -3 \\ 5 \\ -8 \\ -7 \\ 4 \end{bmatrix}. \end{array}$$

Prove, without using any LP code, that $x = (1, 1, 1, 1)$ is optimal.

Exercise 44. Consider the polyhedron

$$\mathcal{P} = \{x \in \mathbf{R}^4 \mid Ax \leq b, \;\; Cx = d\}$$

where

$$A = \begin{bmatrix} -1 & -1 & -3 & -4 \\ -4 & -2 & -2 & -9 \\ -8 & -2 & 0 & -5 \\ 0 & -6 & -7 & -4 \end{bmatrix}, \qquad b = \begin{bmatrix} -8 \\ -17 \\ -15 \\ -17 \end{bmatrix}$$

and

$$C = \begin{bmatrix} 13 & 11 & 12 & 22 \end{bmatrix}, \qquad d = 58.$$

(a) Prove that $\hat{x} = (1, 1, 1, 1)$ is an extreme point of $\mathcal{P}$.

(b) Prove that $\hat{x}$ is optimal for the LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \leq b \\ & Cx = d \end{array}$$

with $c = (59, 39, 38, 85)$.

(c) Is $\hat{x}$ the only optimal point? If not, describe the entire optimal set.

You can use any software, but you have to justify your answers analytically.

Exercise 45. Consider the LP

$$\begin{array}{ll} \text{minimize} & 47x_1 + 93x_2 + 17x_3 - 93x_4 \\ \text{subject to} & \begin{bmatrix} -1 & -6 & 1 & 3 \\ -1 & -2 & 7 & 1 \\ 0 & 3 & -10 & -1 \\ -6 & -11 & -2 & 12 \\ 1 & 6 & -1 & -3 \\ 11 & 1 & -1 & -8 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \leq \begin{bmatrix} -3 \\ 5 \\ -8 \\ -7 \\ 4 \\ 5 \end{bmatrix} + \epsilon \begin{bmatrix} 1 \\ -3 \\ 13 \\ 46 \\ -2 \\ -75 \end{bmatrix}. \end{array} \tag{34}$$

where $\epsilon \in \mathbf{R}$ is a parameter. For $\epsilon = 0$, this is the LP of exercise 43, with one extra inequality (the sixth inequality). This inequality is inactive at $\hat{x} = (1, 1, 1, 1)$, so $\hat{x}$ is also the optimal solution for (34) when $\epsilon = 0$.

(a) Determine the range of values of ϵ for which the first four constraints are active at the optimum.

(b) Give an explicit expression for the optimal primal solution, the optimal dual solution, and the optimal value, within the range of ϵ you determined in part (a). (If for some value of ϵ the optimal points are not unique, it is sufficient to give one optimal point.)

Exercise 46. Consider the parametrized primal and dual LPs

$$\begin{array}{ll} \text{minimize} & (c + \epsilon d)^T x \\ \text{subject to} & Ax \leq b, \end{array} \qquad \begin{array}{ll} \text{maximize} & -b^T z \\ \text{subject to} & A^T z + c + \epsilon d = 0 \\ & z \geq 0 \end{array}$$

where

$$A = \begin{bmatrix} 2 & 3 & 5 & -4 \\ 2 & -1 & -3 & 4 \\ -2 & -1 & 3 & 1 \\ -4 & 2 & 4 & -2 \\ 2 & -3 & -9 & 1 \end{bmatrix}, \qquad b = \begin{bmatrix} 6 \\ 2 \\ 1 \\ 0 \\ -8 \end{bmatrix},$$

$$c = (8, -32, -66, 14), \quad d = (-16, -6, -2, 3).$$

(a) Prove that $x^\star = (1, 1, 1, 1)$ and $z^\star = (9, 9, 4, 9, 0)$ are optimal when $\epsilon = 0$.

(b) How does $p^\star(\epsilon)$ vary as a function of ϵ around $\epsilon = 0$? Give an explicit expression for $p^\star(\epsilon)$, and specify the interval in which it is valid.

(c) Also give an explicit expression for the primal and dual optimal solutions for values of ϵ around $\epsilon = 0$.

Remark: The problem is similar to the sensitivity problem discussed in the lecture notes. Here we consider the case where c is subject to a perturbation, while b is fixed, so you have to develop the 'dual' of the derivation in the lecture notes.

Exercise 47. Consider the pair of primal and dual LPs

$$\begin{array}{ll} \text{minimize} & (c+\epsilon d)^T x \\ \text{subject to} & Ax \leq b + \epsilon f, \end{array} \qquad \begin{array}{ll} \text{maximize} & -(b+\epsilon f)^T z \\ \text{subject to} & A^T z + c + \epsilon d = 0 \\ & z \geq 0 \end{array}$$

where

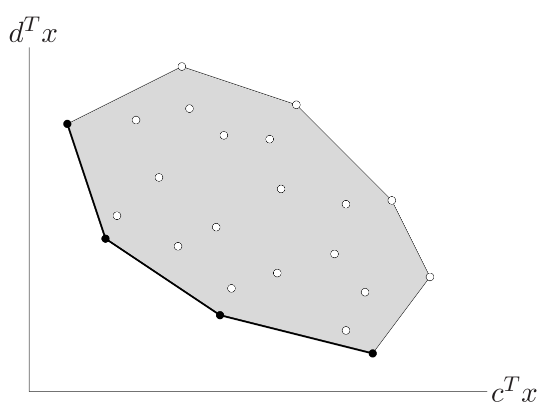
$$A = \begin{bmatrix} -4 & 12 & -2 & 1 \\ -17 & 12 & 7 & 11 \\ 1 & 0 & -6 & 1 \\ 3 & 3 & 22 & -1 \\ -11 & 2 & -1 & -8 \end{bmatrix}, \quad b = \begin{bmatrix} 8 \\ 13 \\ -4 \\ 27 \\ -18 \end{bmatrix}, \quad c = \begin{bmatrix} 49 \\ -34 \\ -50 \\ -5 \end{bmatrix}, \quad d = \begin{bmatrix} 3 \\ 8 \\ 21 \\ 25 \end{bmatrix}, \quad f = \begin{bmatrix} 6 \\ 15 \\ -13 \\ 48 \\ 8 \end{bmatrix}$$

and ϵ is a parameter.

(a) Prove that $x^\star = (1,1,1,1)$ is optimal when $\epsilon = 0$, by constructing a dual optimal point $z^\star$ that has the same objective value as $x^\star$. Are there any other primal or dual optimal solutions?

(b) Express the optimal value $p^\star(\epsilon)$ as a continuous function of ϵ on an interval that contains $\epsilon = 0$. Specify the interval in which your expression is valid. Also give explicit expressions for the primal and dual solutions as a function of ϵ over the same interval.

Exercise 48. In some applications we are interested in minimizing *two* cost functions, $c^T x$ and $d^T x$, over a polyhedron $\mathcal{P} = \{x \mid Ax \leq b\}$. For general c and d, the two objectives are competing, *i.e.*, it is not possible to minimize them simultaneously, and there exists a trade-off between them. The problem can be visualized as in the figure below.



The shaded region is the set of pairs $(c^T x, d^T x)$ for all possible $x \in \mathcal{P}$. The circles are the values $(c^T x, d^T x)$ at the extreme points of $\mathcal{P}$. The lower part of the boundary, shown as a heavy line, is called the *trade-off curve*. Points $(c^T\hat{x}, d^T\hat{x})$ on this curve are efficient in the following sense: it is not possible to improve both objectives by choosing a different feasible x.

Suppose $(c^T\hat{x}, d^T\hat{x})$ is a breakpoint of the trade-off curve, where $\hat{x}$ is a nondegenerate extreme point of $\mathcal{P}$. Explain how the left and right derivatives of the trade-off curve at this breakpoint can be computed.

Hint. Compute the largest and smallest values of γ such that $\hat{x}$ is optimal for the LP

$$\begin{array}{ll} \text{minimize} & d^T x + \gamma c^T x \\ \text{subject to} & Ax \leq b. \end{array}$$

Exercise 49. Consider the ℓ_1-norm minimization problem

$$\text{minimize} \quad \|Ax + b + \epsilon d\|_1$$

with

$$A = \begin{bmatrix} -2 & 7 & 1 \\ -5 & -1 & 3 \\ -7 & 3 & -5 \\ -1 & 4 & -4 \\ 1 & 5 & 5 \\ 2 & -5 & -1 \end{bmatrix}, \qquad b = \begin{bmatrix} -4 \\ 3 \\ 9 \\ 0 \\ -11 \\ 5 \end{bmatrix}, \qquad d = \begin{bmatrix} -10 \\ -13 \\ -27 \\ -10 \\ -7 \\ 14 \end{bmatrix}.$$

(a) Suppose $\epsilon = 0$. Prove, without using any LP code, that $x^\star = \mathbf{1}$ is optimal. Are there any other optimal points?

(b) Give an explicit formula for the optimal value as a function of ϵ for small positive and negative values of ϵ. What are the values of ϵ for which your expression is valid?

Exercise 50. Consider the following optimization problem in x:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \|Ax + b\|_1 \leq 1 \end{array} \tag{35}$$

where $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, $c \in \mathbf{R}^n$.

(a) Formulate this problem as a an LP in inequality form and explain why your LP formulation is equivalent to problem (35).

(b) Derive the dual LP, and show that it is equivalent to the problem

$$\begin{array}{ll} \text{maximize} & b^T z - \|z\|_\infty \\ \text{subject to} & A^T z + c = 0. \end{array}$$

What is the relation between the optimal z and the optimal variables in the dual LP?

(c) Give a direct argument (*i.e.*, not quoting any results from LP duality) that whenever x is primal feasible (*i.e.*, $\|Ax + b\|_1 \leq 1$) and z is dual feasible (*i.e.*, $A^T z + c = 0$), we have

$$c^T x \geq b^T z - \|z\|_\infty.$$

Exercise 51. *Lower bounds in Chebyshev approximation from least-squares.* Consider the Chebyshev approximation problem

$$\text{minimize} \quad \|Ax - b\|_\infty \tag{36}$$

where $A \in \mathbf{R}^{m \times n}$ $(m \geq n)$ and $\mathbf{rank}\, A = n$. Let x_{cheb} denote an optimal point for the Chebyshev approximation problem (there may be multiple optimal points; x_{cheb} denotes one of them).

The Chebyshev problem has no closed-form solution, but the corresponding least-squares problem does. We denote the least-squares solution x_{ls} as

$$x_{\text{ls}} = \text{argmin}\, \|Ax - b\| = (A^T A)^{-1} A^T b.$$

The question we address is the following. Suppose that for a particular A and b you have computed the least-squares solution x_{ls} (but not x_{cheb}). How suboptimal is x_{ls} for the Chebyshev problem? In other words, how much larger is $\|Ax_{\text{ls}} - b\|_\infty$ than $\|Ax_{\text{cheb}} - b\|_\infty$? To answer this question, we need a lower bound on $\|Ax_{\text{cheb}} - b\|_\infty$.

(a) Prove the lower bound

$$\|Ax_{\text{cheb}} - b\|_\infty \geq \frac{1}{\sqrt{m}} \|Ax_{\text{ls}} - b\|_\infty,$$

using the fact that for all $y \in \mathbf{R}^m$,

$$\frac{1}{\sqrt{m}} \|y\| \leq \|y\|_\infty \leq \|y\|.$$

(b) In the duality lecture we derived the following dual for (36):

$$\begin{array}{ll} \text{maximize} & b^T z \\ \text{subject to} & A^T z = 0 \\ & \|z\|_1 \leq 1. \end{array} \tag{37}$$

We can use this dual problem to improve the lower bound obtained in (a).

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

  $$\hat{z} = -r_{\text{ls}}/\|r_{\text{ls}}\|_1, \qquad \tilde{z} = r_{\text{ls}}/\|r_{\text{ls}}\|_1,$$

  are both feasible in (37).
- By duality $b^T \hat{z}$ and $b^T \tilde{z}$ are lower bounds for $\|Ax_{\text{cheb}} - b\|_\infty$. Which is the better bound? How does it compare with the bound obtained in part (a) above?

One application is as follows. You need to solve the Chebyshev approximation problem, but only within, say, 10%. You first solve the least-squares problem (which can be done faster), and then use the bound from part (b) to see if it can guarantee a maximum 10% error. If it can, great; otherwise solve the Chebyshev problem (by slower methods).

Exercise 52. A matrix $A \in \mathbf{R}^{(mp)\times n}$ and a vector $b \in \mathbf{R}^{mp}$ are partitioned in m blocks of p rows:

$$A = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_m \end{bmatrix}, \qquad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix},$$

with $A_k \in \mathbf{R}^{p\times n}$, $b_k \in \mathbf{R}^p$.

(a) Express the optimization problem

$$\text{minimize} \quad \sum_{k=1}^{m} \|A_k x - b_k\|_\infty \tag{38}$$

as an LP.

(b) Suppose $\mathbf{rank}(A) = n$ and $Ax_{\rm ls} - b \neq 0$, where $x_{\rm ls}$ is the solution of the least-squares problem

$$\text{minimize} \quad \|Ax - b\|^2.$$

Show that the optimal value of (38) is bounded below by

$$\frac{\sum_{k=1}^m \|r_k\|^2}{\max_{k=1,\ldots,m} \|r_k\|_1},$$

where $r_k = A_k x_{\rm ls} - b_k$ for $k = 1, \ldots, m$.

Exercise 53. Let x be a real-valued random variable which takes values in $\{a_1, a_2, \ldots, a_n\}$ where $0 < a_1 < a_2 < \cdots < a_n$, and $\mathbf{prob}(x = a_i) = p_i$. Obviously p satisfies $\sum_{i=1}^n p_i = 1$ and $p_i \geq 0$ for $i = 1, \ldots, n$.

(a) Consider the problem of determining the probability distribution that maximizes $\mathbf{prob}(x \geq \alpha)$ subject to the constraint $\mathbf{E}\, x = b$, *i.e.*,

$$\begin{array}{ll} \text{maximize} & \mathbf{prob}(x \geq \alpha) \\ \text{subject to} & \mathbf{E}\, x = b, \end{array} \tag{39}$$

where α and b are given ($a_1 < \alpha < a_n$, and $a_1 \leq b \leq a_n$). The variable in problem (39) is the probability distribution, *i.e.*, the vector $p \in \mathbf{R}^n$. Write (39) as an LP.

(b) Take the dual of the LP in (a), and show that it is can be reformulated as

$$\begin{array}{ll} \text{minimize} & \lambda b + \nu \\ \text{subject to} & \lambda a_i + \nu \geq 0 \text{ for all } a_i < \alpha \\ & \lambda a_i + \nu \geq 1 \text{ for all } a_i \geq \alpha. \end{array}$$

The variables are λ and ν. Give a graphical interpretation of this problem, by interpreting λ and ν as coefficients of an affine function $f(x) = \lambda x + \nu$. Show that the optimal value is equal to

$$\begin{cases} (b - a_1)/(\bar{a} - a_1) & b \leq \bar{a} \\ 1 & b \geq \bar{a}, \end{cases}$$

where $\bar{a} = \min\{a_i \mid a_i \geq \alpha\}$. Also give the optimal values of λ and ν.

(c) From the dual solution, determine the distribution p that solves the problem in (a).

Exercise 54. *The max-flow min-cut theorem.* Consider the maximum flow problem with nonnegative arc flows:

$$\begin{array}{ll} \text{maximize} & t \\ \text{subject to} & Ax = te \\ & 0 \leq x \leq c. \end{array} \tag{40}$$

Here $e = (1, 0, \ldots, 0, -1) \in \mathbf{R}^m$, $A \in \mathbf{R}^{m \times n}$ is the node-arc incidence matrix of a directed graph with m nodes and n arcs, and $c \in \mathbf{R}^n$ is a vector of positive arc capacities. The variables are $t \in \mathbf{R}$ and $x \in \mathbf{R}^n$. In this problem we have an external supply of t at node 1 (the 'source' node) and $-t$ at node m (the 'target' node), and we maximize t subject to the balance equations and the arc capacity constraints.

A *cut* separating nodes 1 and m is a set of nodes that contains node 1 and does not contain node m, *i.e.*, $S \subset \{1, \ldots, m\}$ with $1 \in S$ and $m \notin S$. The *capacity* of the cut is defined as

$$C(S) = \sum_{k \in A(S)} c_k,$$

where $A(S)$ is the set of arcs that start at a node in S and end at a node outside S. The problem of finding the cut with the minimum capacity is called the *minimum cut problem.* In this exercise we show that the solution of the minimum cut problem (with positive weights c) is provided by the dual of the maximum flow problem (40).

(a) Let $p^\star$ be the optimal value of the maximum flow problem (40). Show that

$$p^\star \leq C(S) \tag{41}$$

for all cuts S that separate nodes 1 and m.

(b) Derive the dual problem of (40), and show that it can be expressed as

$$\begin{array}{ll} \text{minimize} & c^T v \\ \text{subject to} & A^T y \leq v \\ & y_1 - y_m = 1 \\ & v \geq 0. \end{array} \tag{42}$$

The variables are $v \in \mathbf{R}^n$ and $y \in \mathbf{R}^m$.

Suppose x and t are optimal in (40), and y and v are optimal in (42). Define the cut

$$\tilde{S} = \{i \mid y_i \geq y_1\}.$$

Use the complementary slackness conditions for (40) and (42) to show that

$$x_k = c_k$$

if arc k starts at a node in $\tilde{S}$ and ends at a node outside $\tilde{S}$, and that

$$x_k = 0$$

if arc k starts at a node outside $\tilde{S}$ and ends at a node in $\tilde{S}$. Conclude that

$$p^\star = C(\tilde{S}).$$

Combined with the result of part 1, this proves that $\tilde{S}$ is a minimum-capacity cut.

Exercise 55. A project consisting of n different tasks can be represented as a directed graph with n arcs and m nodes. The arcs represent the tasks. The nodes represent precedence relations: If arc k starts at node i and arc j ends at node i, then task k cannot start before task j is completed. Node 1 only has outgoing arcs. These arcs represent tasks that can start immediately and in parallel. Node m only has incoming arcs. When the tasks represented by these arcs are completed, the entire project is completed.

We are interested in computing an optimal schedule, *i.e.*, in assigning an optimal start time and a duration to each task. The variables in the problem are defined as follows.

- y_k is the duration of task k, for $k = 1, \ldots, n$. The variables y_k must satisfy the constraints $\alpha_k \leq y_k \leq \beta_k$. We also assume that the cost of completing task k in time y_k is given by $c_k(\beta_k - y_k)$. This means there is no cost if we we use the maximum allowable time β_k to complete the task, but we have to pay if we want the task finished more quickly.
- v_j is an upper bound on the completion times of all tasks associated with arcs that end at node j. These variables must satisfy the relations

$$v_j \geq v_i + y_k \quad \text{if arc } k \text{ starts at node } i \text{ and ends at node } j.$$

Our goal is to minimize the sum of the completion time of the entire project, which is given by $v_m - v_1$, and the total cost $\sum_k c_k(\beta_k - y_k)$. The problem can be formulated as an LP

$$\begin{array}{ll} \text{minimize} & -e^T v + c^T(\beta - y) \\ \text{subject to} & A^T v + y \leq 0 \\ & \alpha \leq y \leq \beta, \end{array}$$

where $e = (1, 0, \ldots, 0, -1)$ and A is the node-arc incidence matrix of the graph. The variables are $v \in \mathbf{R}^m$, $y \in \mathbf{R}^n$.

(a) Derive the dual of this LP.

(b) Interpret the dual problem as a minimum cost network flow problem with nonlinear cost, *i.e.*, a problem of the form

$$\begin{array}{ll} \text{minimize} & \sum_{k=1}^{n} f_k(x_k) \\ \text{subject to} & Ax = e \\ & x \geq 0, \end{array}$$

where f_k is a nonlinear function.

Exercise 56. This problem is a variation on the illumination problem of exercise 16. In part (a) of exercise 16 we formulated the problem

$$\begin{array}{ll} \text{minimize} & \max_{k=1,\ldots,n} |a_k^T p - I_{\text{des}}| \\ \text{subject to} & 0 \leq p \leq \mathbf{1} \end{array}$$

as the following LP in $p \in \mathbf{R}^m$ and an auxiliary variable w:

$$\begin{array}{ll} \text{minimize} & w \\ \text{subject to} & -w \leq a_k^T p - I_{\text{des}} \leq w, \quad k = 1, \ldots, n \\ & 0 \leq p \leq \mathbf{1}. \end{array} \tag{43}$$

Now suppose we add the following constraint on the lamp powers p: no more than half the total power $\sum_{i=1}^m p_i$ is in any subset of r lamps (where r is a given integer with $0 < r < m$). The idea is to avoid solutions where all the power is concentrated in very few lamps. Mathematically, the constraint can be expressed as

$$\sum_{i=1}^{r} p_{[i]} \leq 0.5 \sum_{i=1}^{m} p_i \tag{44}$$

where $p_{[i]}$ is the ith largest component of p. We would like to add this constraint to the LP (43). However the left-hand side of (44) is a complicated nonlinear function of p.

We can write the constraint (44) as a set of linear inequalities by enumerating all subsets $\{i_1, \ldots, i_r\} \subseteq \{1, \ldots, m\}$ with r different elements, and adding an inequality

$$\sum_{k=1}^{r} p_{i_k} \leq 0.5 \sum_{i=1}^{m} p_i$$

for each subset. Equivalently, we express (44) as

$$s^T p \leq 0.5 \sum_{i=1}^{m} p_i \text{ for all } s \in \{0,1\}^m \text{ with } \textstyle\sum_{i=1}^m s_i = r.$$

This yields a set of $\left(\begin{array}{c} m \\ r \end{array} \right)$ linear inequalities in p.

We can use LP duality to derive a much more compact representation. We will prove that (44) can be expressed as the set of $1 + 2m$ linear inequalities

$$rt + \sum_{i=1}^{m} x_i \leq 0.5 \sum_{i=1}^{m} p_i, \qquad p_i \leq t + x_i, \quad i = 1, \ldots, m, \qquad x \geq 0 \tag{45}$$

in $p \in \mathbf{R}^m$, and auxiliary variables $x \in \mathbf{R}^m$ and $t \in \mathbf{R}$.

(a) Given a vector $p \in \mathbf{R}^m$, show that the sum of its r largest elements (*i.e.*, $p_{[1]} + \cdots + p_{[r]}$) is equal to the optimal value of the LP (in the variables $y \in \mathbf{R}^m$)

$$\begin{array}{ll} \text{maximize} & p^T y \\ \text{subject to} & 0 \leq y \leq \mathbf{1} \\ & \mathbf{1}^T y = r. \end{array} \tag{46}$$

(b) Derive the dual of the LP (46). Show that it can be written as

$$\begin{array}{ll} \text{minimize} & rt + \mathbf{1}^T x \\ \text{subject to} & t\mathbf{1} + x \geq p \\ & x \geq 0, \end{array} \tag{47}$$

where the variables are $t \in \mathbf{R}$ and $x \in \mathbf{R}^m$. By duality the LP (47) has the same optimal value as (46), *i.e.*, $p_{[1]} + \cdots + p_{[r]}$.

It is now clear that the optimal value of (47) is less than $0.5\sum_i^m p_i$ if and only if there is a feasible solution t, x in (47) with $rt + \mathbf{1}^T x \leq 0.5\sum_i^m p_i$. In other words, p satisfies the constraint (44) if and only if the set of linear inequalities (45) in x and t are feasible. To include the nonlinear constraint (44) in (43), we can add the inequalities (45), which yields

$$\begin{array}{ll}
\text{minimize} & w \\
\text{subject to} & -w \leq a_k^T p - I_{\text{des}} \leq w, \quad k = 1, \ldots, n \\
& 0 \leq p \leq \mathbf{1} \\
& rt + \mathbf{1}^T x \leq 0.5 \;\; \mathbf{1}^T p \\
& p \leq t\mathbf{1} + x \\
& x \geq 0.
\end{array}$$

This is an LP with $2m + 2$ variables p, x, w, t, and $2n + 4m + 1$ constraints.

Exercise 57. In this problem we derive a linear programming formulation for the following variation on ℓ_∞- and ℓ_1-approximation: given $A \in \mathbf{R}^{m\times n}$, $b \in \mathbf{R}^m$, and an integer k with $1 \leq k \leq m$,

$$\text{minimize} \quad \sum_{i=1}^{k} |Ax - b|_{[i]}. \tag{48}$$

The notation $z_{[i]}$ denotes the ith largest component of $z \in \mathbf{R}^m$, and $|z|_{[i]}$ denotes the ith largest component of the vector $|z| = (|z_1|, |z_2|, \ldots, |z_m|) \in \mathbf{R}^m$. In other words in (48) we minimize the sum of the k largest residuals $|a_i^T x - b_i|$. For $k = 1$, this is the ℓ_∞-problem; for $k = m$, it is the ℓ_1-problem.

Problem (48) can be written as

$$\text{minimize} \quad \max_{1 \leq i_1 < i_2 < \cdots < i_k \leq m} \sum_{j=1}^{k} |a_{i_j}^T x - b_{i_j}|,$$

or as the following LP in x and t:

$$\begin{array}{ll}
\text{minimize} & t \\
\text{subject to} & s^T(Ax - b) \leq t \quad \text{for all } s \in \{-1, 0, 1\}^m, \ \|s\|_1 = k.
\end{array}$$

Here we enumerate all vectors s with components -1, 0 or $+1$, and with exactly k nonzero elements. This yields an LP with $2^k \begin{pmatrix} m \\ k \end{pmatrix}$ linear inequalities.

We now use LP duality to derive a more compact formulation.

(a) We have seen that for $c \in \mathbf{R}^m$ and $1 \leq k \leq n$, the the optimal value of the LP

$$\begin{array}{ll}
\text{maximize} & c^T v \\
\text{subject to} & -y \leq v \leq y \\
& \mathbf{1}^T y = k \\
& y \leq \mathbf{1}
\end{array} \tag{49}$$

is equal to $|c|_{[1]} + \cdots + |c|_{[k]}$. Take the dual of the LP (49) and show that it can be simplified as

$$\begin{array}{ll}
\text{minimize} & kt + \mathbf{1}^T z \\
\text{subject to} & -t\mathbf{1} - z \leq c \leq t\mathbf{1} + z \\
& z \geq 0
\end{array} \tag{50}$$

with variables $t \in \mathbf{R}$ and $z \in \mathbf{R}^m$. By duality the optimal values of (50) and (49) are equal.

(b) Now apply this result to $c = Ax - b$. From part (a), we know that the optimal value of the LP

$$\begin{array}{ll} \text{minimize} & kt + \mathbf{1}^T z \\ \text{subject to} & -t\mathbf{1} - z \leq Ax - b \leq t\mathbf{1} + z \\ & z \geq 0, \end{array} \tag{51}$$

with variables $t \in \mathbf{R}$, $z \in \mathbf{R}^m$ is equal to $\sum_{i=1}^k |Ax - b|_{[i]}$. Note that the constraints (51) are linear in x, so we can simultaneously optimize over x, *i.e.*, solve it as an LP with variables x, t and z. This way we can solve problem (48) by solving an LP with $m+n+1$ variables and $3m$ inequalities.

Exercise 58. *A portfolio optimization problem.* We consider a portfolio optimization problem with n assets or stocks held over one period. The variable x_i will denote the amount of asset i held at the beginning of (and throughout) the period, and p_i will denote the price change of asset i over the period, so the return is $r = p^T x$. The optimization variable is the portfolio vector $x \in \mathbf{R}^n$, which has to satisfy $x_i \geq 0$ and $\sum_{i=1}^n x_i \leq 1$ (unit total budget).

If p is exactly known, the optimal allocation is to invest the entire budget in the asset with the highest return, *i.e.*, if $p_j = \max_i p_i$, we choose $x_j = 1$, and $x_i = 0$ for $i \neq j$. However, this choice is obviously very sensitive to uncertainty in p. We can add various constraints to make the investment more robust against variations in p.

We can impose a diversity constraint that prevents us from allocating the entire budget in a very small number of assets. For example, we can require that no more than, say, 90% of the total budget is invested in any 5% of the assets. We can express this constraint as

$$\sum_{i=1}^{\lfloor n/20 \rfloor} x_{[i]} \leq 0.9$$

where $x_{[i]}$, $i = 1, \ldots, n$, are the values x_i sorted in decreasing order, and $\lfloor n/20 \rfloor$ is the largest integer smaller than or equal to $n/20$.

In addition, we can model the uncertainty in p by specifying a set $\mathcal{P}$ of possible values, and require that the investment maximizes the return in the worst-case scenario. The resulting problem is:

$$\begin{array}{ll} \text{maximize} & \min_{p \in \mathcal{P}} p^T x \\ \text{subject to} & \mathbf{1}^T x \leq 1, \quad x \geq 0, \quad \sum_{i=1}^{\lfloor n/20 \rfloor} x_{[i]} \leq 0.9. \end{array} \tag{52}$$

For each of the following sets $\mathcal{P}$, can you express problem (52) as an LP?

(a) $\mathcal{P} = \{p^{(1)}, \ldots, p^{(K)}\}$, where $p^{(i)} \in \mathbf{R}^n$ are given. This means we consider a finite number of possible scenarios.

(b) $\mathcal{P} = \{\bar{p} + By \mid \|y\|_\infty \leq 1\}$ where $\bar{p} \in \mathbf{R}^n$ and $B \in \mathbf{R}^{n \times m}$ are given. We can interpret $\bar{p}$ as the expected value of p, and $y \in \mathbf{R}^m$ as uncertain parameters that determine the actual values of p.

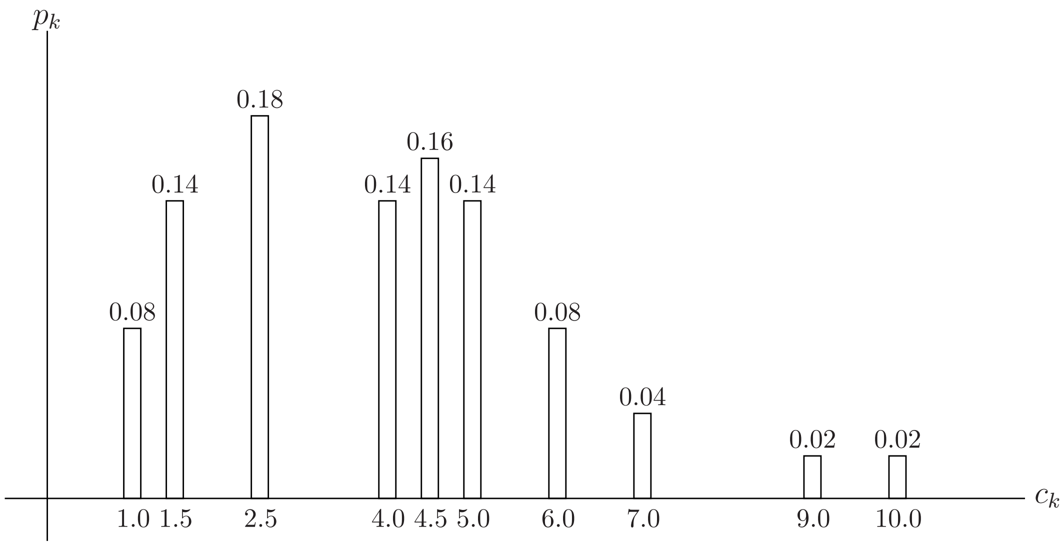
(c) $\mathcal{P} = \{\bar{p} + y \mid By \leq d\}$ where $\bar{p} \in \mathbf{R}^n$, $B \in \mathbf{R}^{r \times m}$, and $d \in \mathbf{R}^r$ are given. Here we consider a polyhedron of possible value of p. (We assume that $\mathcal{P}$ is nonempty.)

You may introduce new variables and constraints, but you must clearly explain why your formulation is equivalent to (52). If you know more than one solution, you should choose the most compact formulation, *i.e.*, involving the smallest number of variables and constraints.

Exercise 59. Let v be a discrete random variable with possible values $c_1, \ldots, c_n$, and distribution $p_k = \mathbf{prob}(v = c_k)$, $k = 1, \ldots, n$. The β-*quantile* of v, where $0 < \beta < 1$, is defined as

$$q_\beta = \min\{\alpha \mid \mathbf{prob}(v \leq \alpha) \geq \beta\}.$$

For example, the 0.9-quantile of the distribution shown in the figure is $q_{0.9} = 6.0$.



A related quantity is

$$f_\beta = \frac{1}{1-\beta} \sum_{c_k > q_\beta} p_k c_k + \left(1 - \frac{1}{1-\beta} \sum_{c_i > q_\beta} p_i\right) q_\beta.$$

If $\sum_{c_i > q_\beta} p_i = 1 - \beta$ (and the second term vanishes), this is the conditional expected value of v, given that v is greater than q_β. Roughly speaking, f_β is the mean of the tail of the distribution above the β-quantile. In the example of the figure,

$$f_{0.9} = \frac{0.02 \cdot 6.0 + 0.04 \cdot 7.0 + 0.02 \cdot 9.0 + 0.02 \cdot 10.0}{0.1} = 7.8.$$

We consider optimization problems in which the values of c_k depend linearly on some optimization variable x. We will formulate the problem of minimizing f_β, subject to linear constraints on x, as a linear program.

(a) Show that the optimal value of the LP

$$\begin{array}{ll} \text{maximize} & c^T y \\ \text{subject to} & 0 \leq y \leq (1-\beta)^{-1} p \\ & \mathbf{1}^T y = 1, \end{array} \qquad (53)$$

with variable $y \in \mathbf{R}^n$, is equal to f_β. The parameters c, p and β are given, with $p > 0$, $\mathbf{1}^T p = 1$, and $0 < \beta < 1$.

(b) Write the LP (53) in inequality form, derive its dual, and show that the dual is equivalent to the piecewise-linear minimization problem

$$\text{minimize} \quad t + \frac{1}{1-\beta}\sum_{k=1}^{n} p_k \max\{0, c_k - t\}, \tag{54}$$

with a single scalar variable t. It follows from duality theory and the result in part 1 that the optimal value of (54) is equal to f_β.

(c) Now suppose $c_k = a_k^T x$, where $x \in \mathbf{R}^m$ is an optimization variable and a_k is given, so $q_\beta(x)$ and $f_\beta(x)$ both depend on x. Use the result in part 2 to express the problem

$$\begin{array}{ll} \text{minimize} & f_\beta(x) \\ \text{subject to} & Fx \preceq g, \end{array}$$

with variable x, as an LP.

As an application, we consider a portfolio optimization problem with m assets or stocks held over a period of time. We represent the portfolio by a vector $x = (x_1, x_2, \ldots, x_m)$, with x_k the amount invested in asset k during the investment period. We denote by r the vector of returns for the m assets over the period, so the total return on the portfolio is $r^T x$. The loss (negative return) is denoted $v = -r^T x$.

We model r as a discrete random variable, with possible values $-a_1, \ldots, -a_n$, and distribution

$$p_k = \mathbf{prob}(r = -a_k), \quad k = 1, \ldots, n.$$

The loss of the portfolio $v = -r^T x$ is therefore a random variable with possible values $c_k = a_k^T x$, $k = 1, \ldots, n$, and distribution p.

In this context, the β-quantile $q_\beta(x)$ is called the *value-at-risk* of the portfolio, and $f_\beta(x)$ is called the *conditional value-at-risk*. If we take β close to one, both functions are meaningful measures of the risk of the portfolio x. The result of part 3 implies that we can minimize $f_\beta(x)$, subject to linear constraints in x, via linear programming. For example, we can minimize the risk (expressed as $f_\beta(x)$), subject to an upper bound on the expected loss (*i.e.*, a lower bound on the expected return), by solving

$$\begin{array}{ll} \text{minimize} & f_\beta(x) \\ \text{subject to} & \sum\limits_k p_k a_k^T x \leq R \\ & \mathbf{1}^T x = 1 \\ & x \succeq 0. \end{array}$$

Exercise 60. *A generalized linear-fractional problem.*
Consider the problem

$$\begin{array}{ll} \text{minimize} & \|Ax - b\|_1 / (c^T x + d) \\ \text{subject to} & \|x\|_\infty \leq 1 \end{array} \tag{55}$$

where $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, $c \in \mathbf{R}^n$ and $d \in \mathbf{R}$ are given. We assume that $d > \|c\|_1$. As a consequence, $c^T x + d > 0$ for all feasible x.

Remark: There are several correct answers to part (a), including a method based on solving one single LP.

(a) Explain how you would solve this problem using linear programming. If you know more than one method, you should give the simplest one.

(b) Prove that the following problem provides lower bounds on the optimal value of (55):

$$\begin{array}{ll} \text{maximize} & \lambda \\ \text{subject to} & \|A^T z + \lambda c\|_1 \leq b^T z - \lambda d \\ & \|z\|_\infty \leq 1. \end{array} \tag{56}$$

The variables are $z \in \mathbf{R}^m$ and $\lambda \in \mathbf{R}$.

(c) Use linear programming duality to show that the optimal values of (56) and (55) are in fact equal.

Exercise 61. Consider the problem

$$\text{minimize} \quad \sum_{i=1}^m h(a_i^T x - b_i) \tag{57}$$

where h is the function

$$h(z) = \begin{cases} 0 & |z| \leq 1 \\ |z| - 1 & |z| > 1 \end{cases}$$

and (as usual) $x \in \mathbf{R}^n$ is the variable, and $a_1, \ldots, a_m \in \mathbf{R}^n$ and $b \in \mathbf{R}^m$ are given. Note that this problem can be thought of as a sort of hybrid between ℓ_1- and ℓ_∞-approximation, since there is no cost for residuals smaller than one, and a linearly growing cost for residuals larger than one.

Express (57) as an LP, derive its dual, and simplify it as much as you can.

Let $x_{\rm ls}$ denote the solution of the least-squares problem

$$\text{minimize} \quad \sum_{i=1}^m (a_i^T x - b_i)^2,$$

and let $r_{\rm ls}$ denote the residual $r_{\rm ls} = Ax_{\rm ls} - b$. We assume A has rank n, so the least-squares solution is unique and given by

$$x_{\rm ls} = (A^T A)^{-1} A^T b.$$

The least-squares residual $r_{\rm ls}$ satisfies

$$A^T r_{\rm ls} = 0.$$

Show how to construct from $x_{\rm ls}$ and $r_{\rm ls}$ a feasible solution for the dual of (57), and hence a lower bound for its optimal value $p^\star$. Compare your lower bound with the trivial lower bound $p^\star \geq 0$. Is it always better, or only in certain cases?

Exercise 62. *Self-dual homogeneous LP formulation.*

(a) Consider the LP

$$\begin{array}{ll} \text{minimize} & f_1^T u + f_2^T v \\ \text{subject to} & M_{11} u + M_{12} v \leq f_1 \\ & -M_{12}^T u + M_{22} v = f_2 \\ & u \geq 0 \end{array} \tag{58}$$

in the variables $u \in \mathbf{R}^p$ and $v \in \mathbf{R}^q$. The problem data are the vectors $f_1 \in \mathbf{R}^p$, $f_2 \in \mathbf{R}^q$, and the matrices $M_{11} \in \mathbf{R}^{p\times p}$, $M_{12} \in \mathbf{R}^{p\times q}$, and $M_{22} \in \mathbf{R}^{q\times q}$.

Show that if M_{11} and M_{22} are *skew-symmetric*, *i.e.*,

$$M_{11}^T = -M_{11}, \qquad M_{22}^T = -M_{22},$$

then the dual of the LP (58) can be expressed as

$$\begin{array}{ll} \text{maximize} & -f_1^T w - f_2^T y \\ \text{subject to} & M_{11}w + M_{12}y \leq f_1 \\ & -M_{12}^T w + M_{22}y = f_2 \\ & w \geq 0, \end{array} \tag{59}$$

with variables $w \in \mathbf{R}^p$ and $y \in \mathbf{R}^q$.

Note that the dual problem is essentially the same as the primal problem. Therefore if u, v are primal optimal, then $w = u$, $y = v$ are optimal in the dual problem. We say that the LP (58) with skew-symmetric M_{11} and M_{22} is *self-dual*.

(b) Write down the optimality conditions for problem (58). Use the observation we made in part (a) to show that the optimality conditions can be simplified as follows: u, v are optimal for (58) if and only if

$$\begin{array}{c} M_{11}u + M_{12}v \leq f_1 \\ -M_{12}^T u + M_{22}v = f_2 \\ u \geq 0 \\ u^T(f_1 - M_{11}u - M_{12}v) = 0. \end{array}$$

In other words, u, v must be feasible in (58), and the nonnegative vectors u and

$$s = f_1 - M_{11}u - M_{12}v$$

must satisfy the complementarity condition $u^T s = 0$.

It can be shown that if (58) is feasible, then it has an optimal solution that is *strictly complementary*, *i.e.*,

$$u + s > 0.$$

(In other words, for each k either $s_k = 0$ or $u_k = 0$, but not both.)

(c) Consider the LP

$$\begin{array}{ll} \text{minimize} & 0 \\ \text{subject to} & b^T\tilde{z} + c^T\tilde{x} \leq 0 \\ & -b\tilde{t} + A\tilde{x} \leq 0 \\ & A^T\tilde{z} + c\tilde{t} = 0 \\ & \tilde{z} \geq 0, \quad \tilde{t} \geq 0 \end{array} \tag{60}$$

with variables $\tilde{x} \in \mathbf{R}^n$, $\tilde{z} \in \mathbf{R}^m$, and $\tilde{t} \in \mathbf{R}$. Show that this problem is self-dual.

Use the result in part (b) to prove that (60) has an optimal solution that satisfies

$$\tilde{t}(c^T\tilde{x} + b^T\tilde{z}) = 0$$

and

$$\tilde{t} - (c^T\tilde{x} + b^T\tilde{z}) > 0.$$

Suppose we have computed an optimal solution with these properties. We can distinguish the following cases.

- $\tilde{t} > 0$. Show that $x = \tilde{x}/\tilde{t}$, $z = \tilde{z}/\tilde{t}$ are optimal for the pair of primal and dual LPs

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \leq b \end{array} \tag{61}$$

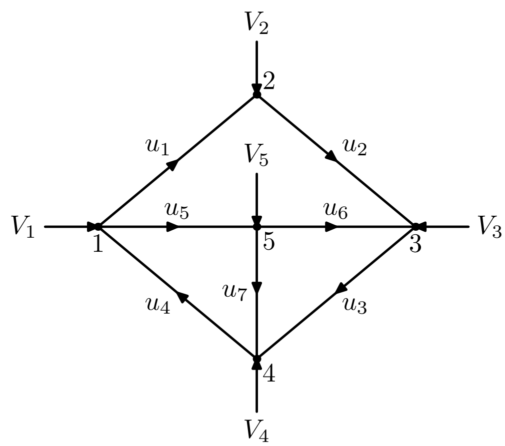
  and

$$\begin{array}{ll} \text{maximize} & -b^T z \\ \text{subject to} & A^T z + c = 0 \\ & z \geq 0. \end{array} \tag{62}$$

- $\tilde{t} = 0$ and $c^T \tilde{x} < 0$. Show that the dual LP (62) is infeasible.
- $\tilde{t} = 0$ and $b^T \tilde{z} < 0$. Show that the primal LP (62) is infeasible.

This result has an important practical ramification. It implies that we do not have to use a two-phase approach to solve the LP (61) (*i.e.*, a phase-I to find a feasible point, followed by a phase-II to minimize $c^T x$ starting at the feasible point). We can solve the LP (61) and its dual, or detect primal or dual infeasibility, by solving one single, feasible LP (60). The LP (60) is much larger than (61), but it can be shown that the cost of solving it is not much higher if one takes advantage of the symmetry in the constraints.

Exercise 63. We consider a network flow problem on the simple network shown below.



Here u_1, …, $u_7 \in \mathbf{R}$ denote the *flows* or traffic along links $1, \ldots, 7$ in the direction indicated by the arrow. (Thus, $u_1 = 1$ means a traffic flow of one unit in the direction of the arrow on link 1, *i.e.*, from node 1 to node 2.) $V_1, \ldots, V_5 \in \mathbf{R}$ denote the external inputs (or outputs if $V_i < 0$) to the network. We assume that the net flow into the network is zero, *i.e.*, $\sum_{i=1}^5 V_i = 0$.

Conservation of traffic flow states that at each node, the total flow entering the node is zero. For example, for node 1, this means that $V_1 - u_1 + u_4 - u_5 = 0$. This gives one equation per node, so we have 5 traffic conservation equations, for the nodes $1, \ldots, 5$, respectively. (In fact, the equations are redundant since they sum to zero, so you could leave one, *e.g.*, for node 5, out. However, to answer the questions below, it is easier to keep all five equations.)

The cost of a flow pattern u is given by $\sum_i c_i |u_i|$, where $c_i > 0$ is the tariff on link i. In addition to the tariff, each link also has a maximum possible traffic level or link capacity: $|u_i| \leq U_i$.

(a) Express the problem of finding the minimum cost flow as an LP in inequality form, for the network shown above.

(b) Solve the LP from part (a) for the specific costs, capacities, and inputs

$$c = (2, 2, 2, 1, 1, 1, 1), \qquad V = (1, 1, 0.5, 0.5, -3), \qquad U = (0.5, 0.5, 0.1, 0.5, 1, 1, 1).$$

Find the optimal dual variables as well.

(c) Suppose we can increase the capacity of one link by a small fixed amount, say, 0.1. Which one should we choose, and why? (You're not allowed to solve new LPs to answer this!) For the link you pick, increase its capacity by 0.1, and then solve the resulting LP exactly. Compare the resulting cost with the cost predicted from the optimal dual variables of the original problem. Can you explain the answer?

(d) Now suppose we have the possibility to increase or reduce two of the five external inputs by a small amount, say, 0.1. To keep $\sum_i V_i = 0$, the changes in the two inputs must be equal in absolute value and opposite in sign. For example, we can increae V_1 by 0.1, and decrease V_4 by 0.1. Which two inputs should we modify, and why? (Again, you're not allowed to solve new LPs!) For the inputs you pick, change the value (increase or decrease, depending on which will result in a smaller cost) by 0.1, and then solve the resulting LP exactly. Compare the result with the one predicted from the optimal dual variables of the original problem.

Exercise 64. Let $P \in \mathbf{R}^{n\times n}$ be a matrix with the following two properties:

- all elements of P are nonnegative: $p_{ij} \geq 0$ for $i = 1, \ldots, n$ and $j = 1, \ldots, n$
- the columns of P sum to one: $\sum_{i=1}^{n} p_{ij} = 1$ for $j = 1, \ldots, n$.

Show that there exists a $y \in \mathbf{R}^n$ such that

$$Py = y, \qquad y \geq 0, \qquad \sum_{i=1}^{n} y_i = 1.$$

Remark. This result has the following application. We can interpret P as the transition probability matrix of a Markov chain with n states: if $s(t)$ is the state at time t (*i.e.*, $s(t)$ is a random variable taking values in $\{1, \ldots, n\}$), then p_{ij} is defined as

$$p_{ij} = \mathbf{prob}(s(t+1) = i \mid s(t) = j).$$

Let $y(t) \in \mathbf{R}^n$ be the probability distribution of the state at time t, *i.e.*,

$$y_i(t) = \mathbf{prob}(s(t) = i).$$

Then the distribution at time $t+1$ is given by $y(t+1) = Py(t)$.

The result in this problem states that a finite state Markov chain always has an equilibrium distribution y.

Exercise 65. *Arbitrage and theorems of alternatives.*

Consider an event (for example, a sports game, political elections, the evolution of the stock-market over a certain period) with m possible outcomes. Suppose that n wagers on the outcome are possible. If we bet an amount x_j on wager j, and the outcome of the event is

i, then our return is equal to $r_{ij}x_j$ (this amount does not include the stake, *i.e.*, we pay x_j initially, and receive $(1+r_{ij})x_j$ if the outcome of the event is i, so $r_{ij}x_j$ is the net gain). We allow the bets x_j to be positive, negative, or zero. The interpretation of a negative bet is as follows. If $x_j < 0$, then initially we *receive* an amount of money $|x_j|$, with an obligation to *pay* $(1+r_{ij})|x_j|$ if outcome i occurs. In that case, we lose $r_{ij}|x_j|$, *i.e.*, our net gain is $r_{ij}x_j$ (a negative number).

We call the matrix $R \in \mathbf{R}^{m\times n}$ with elements r_{ij} the *return matrix*. A *betting strategy* is a vector $x \in \mathbf{R}^n$, with as components x_j the amounts we bet on each wager. If we use a betting strategy x, our total return in the event of outcome i is equal to $\sum_{j=1}^n r_{ij}x_j$, *i.e.*, the ith component of the vector Rx.

(a) *The arbitrage theorem.* Suppose you are given a return matrix R. Prove the following theorem: there is a betting strategy $x \in \mathbf{R}^n$ for which

$$Rx > 0 \tag{63}$$

if and only if there exists no vector $p \in \mathbf{R}^m$ that satisfies

$$R^T p = 0, \qquad p \succeq 0, \qquad p \neq 0. \tag{64}$$

We can interpret this theorem as follows. If $Rx > 0$, then the betting strategy x guarantees a positive return for all possible outcomes, *i.e.*, it is a sure-win betting scheme. In economics, we say there is an *arbitrage opportunity.*

If we normalize the vector p in (64) so that $\mathbf{1}^T p = 1$, we can interpret it as a probability vector on the outcomes. The condition $R^T p = 0$ means that the expected return

$$\mathbf{E}\, Rx = p^T Rx = 0$$

for all betting strategies. We can therefore rephrase the arbitrage theorem as follows. There is no sure-win betting strategy (or arbitrage opportunity) if and only if there is a probability vector on the outcomes that makes all bets fair (*i.e.*, the expected gain is zero).

(b) *Options pricing.* The arbitrage theorem is used in mathematical finance to determine prices of contracts. As a simple example, suppose we can invest in two assets: a stock and an option. The current unit price of the stock is S. The price $\bar{S}$ of the stock at the end of the investment period is unknown, but it will be either $\bar{S} = Su$ or $\bar{S} = Sd$, where $u > 1$ and $d < 1$ are given numbers. In other words the price either goes up by a factor u, or down by a factor d. If the current interest rate over the investment period is r, then the present value of the stock price $\bar{S}$ at the end of the period is equal to $\bar{S}/(1+r)$, and our unit return is

$$\frac{Su}{1+r} - S = S\frac{u-1-r}{1+r}$$

if the stock goes up, and

$$\frac{Sd}{1+r} - S = S\frac{d-1-r}{1+r}$$

if the stock goes down.

We can also buy options, at a unit price of C. An option gives us the right to purchase one stock at a fixed price K at the end of the period. Whether we exercise the option or not, depends on the price of the stock at the end of the period. If the stock price $\bar{S}$ at the end of the period is greater than K, we exercise the option, buy the stock and sell it immediately, so we receive an amount $\bar{S} - K$. If the stock price $\bar{S}$ is less than K, we do not exercise the option and receive nothing. Combining both cases, we can say that the value of the option at the end of the period is $\max\{0, \bar{S} - K\}$, and the present value is $\max\{0, \bar{S} - K\}/(1 + r)$. If we pay a price C per option, then our return is

$$\frac{1}{1+r}\max\{0, \bar{S} - K\} - C$$

per option.

We can summarize the situation with the return matrix

$$R = \left[\begin{array}{cc} (u-1-r)/(1+r) & (\max\{0, Su-K\})/((1+r)C) - 1 \\ (d-1-r)/(1+r) & (\max\{0, Sd-K\})/((1+r)C) - 1 \end{array}\right].$$

The elements of the first row are the (present values of the) returns in the event that the stock price goes up. The second row are the returns in the event that the stock price goes down. The first column gives the returns per unit investment in the stock. The second column gives the returns per unit investment in the option.

In this simple example the arbitrage theorem allows us to determine the price of the option, given the other information S, K, u, d, and r. Show that if there is no arbitrage, then the price of the option C must be equal to

$$\frac{1}{1+r}(p\max\{0, Su-K\} + (1-p)\max\{0, Sd-K\})$$

where

$$p = \frac{1+r-d}{u-d}.$$

Exercise 66. We consider a network with m nodes and n directed arcs. Suppose we can apply labels $y_r \in \mathbf{R}$, $r = 1, \ldots, m$, to the nodes in such a way that

$$y_r \geq y_s \quad \text{if there is an arc from node } r \text{ to node } s. \tag{65}$$

It is clear that this implies that if $y_i < y_j$, then there exists no directed path from node i to node j. (If we follow a directed path from node i to j, we encounter only nodes with labels less than or equal to y_i. Therefore $y_j \leq y_i$.)

Prove the converse: if there is no directed path from node i to j, then there exists a labeling of the nodes that satisfies (65) and $y_i < y_j$.

Exercise 67. The projection of a point $x_0 \in \mathbf{R}^n$ on a polyhedron $\mathcal{P} = \{x \mid Ax \leq b\}$, in the ℓ_∞-norm, is defined as the solution of the optimization problem

$$\begin{array}{ll} \text{minimize} & \|x - x_0\|_\infty \\ \text{subject to} & Ax \leq b. \end{array}$$

The variable is $x \in \mathbf{R}^n$. We assume that $\mathcal{P}$ is nonempty.

(a) Formulate this problem as an LP.

(b) Derive the dual problem, and simplify it as much as you can.

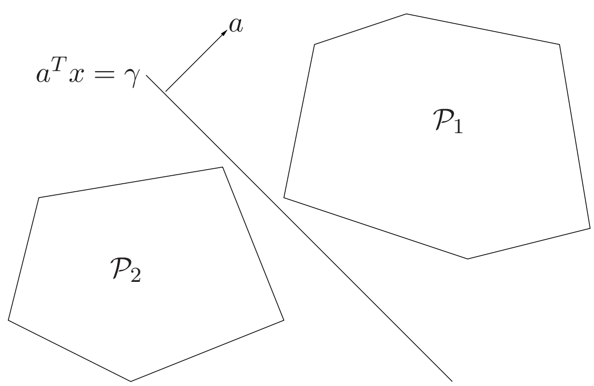
(c) Show that if $x_0 \notin \mathcal{P}$, then a hyperplane that separates x_0 from $\mathcal{P}$ can be constructed from the optimal solution of the dual problem.

Exercise 68. Describe a method for constructing a hyperplane that separates two given polyhedra

$$\mathcal{P}_1 = \{x \in \mathbf{R}^n | Ax \leq b\}, \qquad \mathcal{P}_2 = \{x \in \mathbf{R}^n | Cx \leq d\}.$$

Your method must return a vector $a \in \mathbf{R}^n$ and a scalar γ such that

$$a^T x > \gamma \text{ for all } x \in \mathcal{P}_1, \qquad a^T x < \gamma \text{ for all } x \in \mathcal{P}_2.$$



You can assume that $\mathcal{P}_1$ and $\mathcal{P}_2$ do not intersect. If you know several methods, you should give the most efficient one.

Exercise 69. Suppose the feasible set of the LP

$$\begin{array}{ll} \text{maximize} & b^T z \\ \text{subject to} & A^T z \leq c \end{array} \tag{66}$$

is nonempty and bounded, with $\|z\|_\infty < \mu$ for all feasible z. Show that any optimal solution of the problem

$$\begin{array}{ll} \text{minimize} & c^T x + \mu \|Ax - b\|_1 \\ \text{subject to} & x \geq 0 \end{array}$$

is also an optimal solution of the LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0, \end{array} \tag{67}$$

which is the dual of problem (66).

Exercise 70. An alternative to the phase-I/phase-II method for solving the LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \leq b, \end{array} \tag{68}$$

is the "big-M"-method, in which we solve the auxiliary problem

$$\begin{array}{ll} \text{minimize} & c^T x + Mt \\ \text{subject to} & Ax \leq b + t\mathbf{1} \\ & t \geq 0. \end{array} \tag{69}$$

$M > 0$ is a parameter and t is an auxiliary variable. Note that this auxiliary problem has obvious feasible points, for example, $x = 0$, $t \geq \max\{0, -\min_i b_i\}$.

(a) Derive the dual LP of (69).

(b) Prove the following property. If $M > \mathbf{1}^T z^\star$, where $z^\star$ is an optimal solution of the dual of (68), then the optimal t in (69) is zero, and therefore the optimal x in (69) is also an optimal solution of (68).

Exercise 71. *Robust linear programming with polyhedral uncertainty.* Consider the robust LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \max_{a \in \mathcal{P}_i} a^T x \leq b_i, \quad i = 1, \ldots, m, \end{array}$$

with variable $x \in \mathbf{R}^n$, where $\mathcal{P}_i = \{a \mid C_i a \leq d_i\}$. The problem data are $c \in \mathbf{R}^n$, $C_i \in \mathbf{R}^{m_i \times n}$, $d_i \in \mathbf{R}^{m_i}$, and $b \in \mathbf{R}^m$. We assume the polyhedra $\mathcal{P}_i$ are nonempty.

Show that this problem is equivalent to the LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & d_i^T z_i \leq b_i, \quad i = 1, \ldots, m \\ & C_i^T z_i = x, \quad i = 1, \ldots, m \\ & z_i \geq 0, \quad i = 1, \ldots, m \end{array}$$

with variables $x \in \mathbf{R}^n$ and $z_i \in \mathbf{R}^{m_i}$, $i = 1, \ldots, m$. *Hint.* Find the dual of the problem of maximizing $a_i^T x$ over $a_i \in \mathcal{P}_i$ (with variable a_i).

Exercise 72. *Strict complementarity.* We consider an LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \leq b, \end{array}$$

with $A \in \mathbf{R}^{m \times n}$, and its dual

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

We assume the optimal value is finite. From duality theory we know that any primal optimal $x^\star$ and any dual optimal $z^\star$ satisfy the complementary slackness conditions

$$z_i^\star (b_i - a_i^T x^\star) = 0, \quad i = 1, \ldots, m.$$

In other words, for each i, we have $z_i^\star = 0$, or $a_i^T x^\star = b_i$, or both.

In this problem you are asked to show that there exists at least one primal-dual optimal pair $x^\star$, $z^\star$ that satisfies

$$z_i^\star (b_i - a_i^T x^\star) = 0, \qquad z_i^\star + (b_i - a_i^T x^\star) > 0,$$

for all i. This is called a *strictly complementary* pair. In a strictly complementary pair, we have for each i, either $z_i^\star = 0$, or $a_i^T x^\star = b_i$, but not both.

To prove the result, suppose $x^\star$, $z^\star$ are optimal but not strictly complementary, and

$$\begin{array}{lll} a_i^T x^\star = b_i, & z_i^\star = 0, & i = 1, \ldots, M \\ a_i^T x^\star = b_i, & z_i^\star > 0, & i = M+1, \ldots, N \\ a_i^T x^\star < b_i, & z_i^\star = 0, & i = N+1, \ldots, m \end{array}$$

with $M > 1$. In other words, $m - M$ entries of $b - Ax^\star$ and $z^\star$ are strictly complementary; for the other entries we have zero in both vectors.

(a) Use Farkas' lemma to show that the following two sets of inequalities/equalities are strong alternatives:

- There exists a $v \in \mathbf{R}^n$ such that

$$\begin{array}{rcll} a_1^T v & < & 0 & \\ a_i^T v & \leq & 0, & i = 2, \ldots, M \\ a_i^T v & = & 0, & i = M+1, \ldots, N. \end{array} \tag{70}$$

- There exists a $w \in \mathbf{R}^{N-1}$ such that

$$a_1 + \sum_{i=1}^{N-1} w_i a_{i+1} = 0, \qquad w_i \geq 0, \quad i = 1, \ldots, M-1. \tag{71}$$

(b) Assume the first alternative holds, and v satisfies (70). Show that there exists a primal optimal solution $\tilde{x}$ with

$$\begin{array}{rcll} a_1^T \tilde{x} & < & b_1 & \\ a_i^T \tilde{x} & \leq & b_i, & i = 2, \ldots, M \\ a_i^T \tilde{x} & = & b_i, & i = M+1, \ldots, N \\ a_i^T \tilde{x} & < & b_i, & i = N+1, \ldots, m. \end{array}$$

(c) Assume the second alternative holds, and w satisfies (71). Show that there exists a dual optimal $\tilde{z}$ with

$$\begin{array}{rcll} \tilde{z}_1 & > & 0 & \\ \tilde{z}_i & \geq & 0, & i = 2, \ldots, M \\ \tilde{z}_i & > & 0, & i = M+1, \ldots, N \\ \tilde{z}_i & = & 0, & i = N+1, \ldots, m. \end{array}$$

(d) Combine (b) and (c) to show that there exists a primal-dual optimal pair x, z, for which $b - Ax$ and z have at most $\tilde{M}$ common zeros, where $\tilde{M} < M$. If $\tilde{M} = 0$, x, z are strictly complementary and optimal, and we are done. Otherwise, we apply the argument given above, with $x^\star$, $z^\star$ replaced by x, z, to show the existence of a strictly complementary pair of optimal solutions with less than $\tilde{M}$ common zeros in $b - Ax$ and z. Repeating the argument eventually gives a strictly complementary pair.

Exercise 73. Prove the following result. If the feasible set of a linear program

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \leq b \end{array}$$

is nonempty and bounded, then the feasible set of the corresponding dual problem

$$\begin{array}{ll} \text{maximize} & -b^T z \\ \text{subject to} & A^T z + c = 0 \\ & z \geq 0 \end{array}$$

is nonempty and *unbounded*.

Exercise 74. We are given $M + N$ polyhedra described by sets of linear inequalities

$$P_i = \{x \in \mathbf{R}^n \mid A_i x \leq b_i\}, \quad i = 1, \ldots, M + N.$$

We define two sets $S = P_1 \cup P_2 \cup \cdots \cup P_M$ and $T = P_{M+1} \cup P_{M+2} \cup \cdots \cup P_{M+N}$.

(a) Explain how you can use linear programming to solve the following problem. Find a vector c and a scalar d such that

$$c^T x + d \leq -1 \quad \text{for } x \in S, \qquad c^T x + d \geq 1 \quad \text{for } x \in T \tag{72}$$

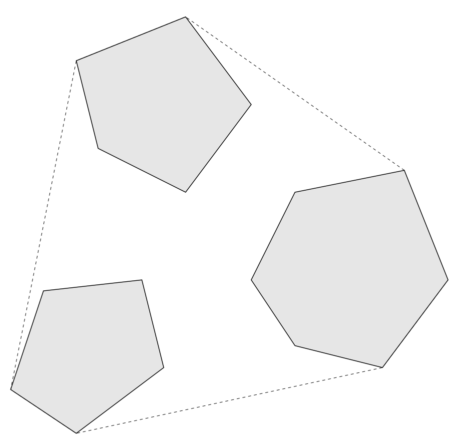
or show that no such c and d exist. Geometrically, the problem is to construct a hyperplane that strictly separates the polyhedra P_1, ..., P_M from the polyhedra P_{M+1}, ..., P_{M+N}.

If you know several methods, give the most efficient one. In particular, you should avoid methods based on enumerating extreme points, and methods that involve linear programs with dimensions that grow quadratically (or faster) with M or N.

(b) The convex hull of a set S, denoted $\mathbf{conv}\, S$, is defined as the set of all convex combinations of points in S:

$$\mathbf{conv}\, S = \{\theta_1 v_1 + \cdots + \theta_m v_m \mid \theta_1 + \cdots + \theta_m = 1,\ v_i \in S,\ \theta_i \geq 0,\ i = 1, \ldots, m\}.$$

The convex hull of the shaded set S the figure is the polyhedron enclosed by the dashed lines.



Show that if no separating hyperplane exists between S and T (*i.e.*, there exists no c and d that satisfy (72)), then the convex hulls $\mathbf{conv}\, S$ and $\mathbf{conv}\, T$ intersect.

6 The simplex method

Exercise 75. Solve the following linear program using the simplex algorithm with Bland's pivoting rule. Start the algorithm at the extreme point $x = (2, 2, 0)$, with active set $I = \{3, 4, 5\}$.

$$\begin{array}{ll}
\text{mininimize} & x_1 + x_2 - x_3 \\
\text{subject to} & \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \leq \begin{bmatrix} 0 \\ 0 \\ 0 \\ 2 \\ 2 \\ 2 \\ 4 \end{bmatrix}.
\end{array}$$

Exercise 76. Use the simplex method to solve the following LP:

$$\begin{array}{ll}
\text{minimize} & -24x_1 + 396x_2 - 8x_3 - 28x_4 - 10x_5 \\
\text{subject to} & \begin{bmatrix} 12 & 4 & 1 & -19 & 7 \\ 6 & -7 & 18 & -1 & -13 \\ 1 & 17 & 3 & 18 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 12 \\ 6 \\ 1 \end{bmatrix} \\
& x \geq 0.
\end{array}$$

Start with the initial basis $\{1, 2, 3\}$, and use Bland's rule to make pivot selections. Also compute the dual optimal point from the results of the algorithm.

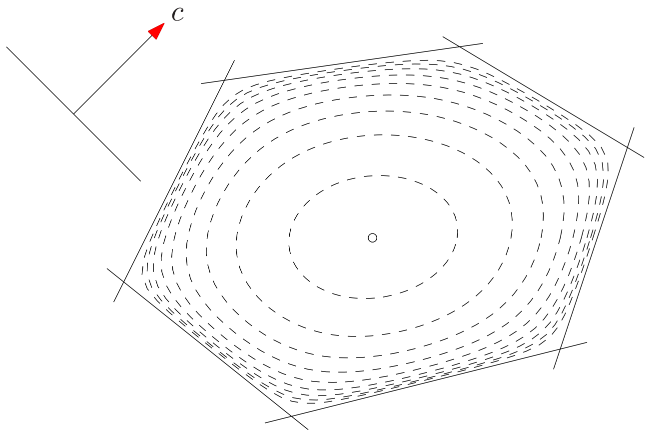
7 Interior-point methods

Exercise 77. The figure shows the feasible set of an LP

$$\begin{array}{ll}
\text{minimize} & c^T x \\
\text{subject to} & a_i^T x \leq b_i, \quad i = 1, \ldots, 6
\end{array}$$

with two variables and six constraints. Also shown are the cost vector c, the analytic center, and a few contour lines of the logarithmic barrier function

$$\phi(x) = -\sum_{i=1}^{m} \log(b_i - a_i^T x).$$



Sketch the central path as accurately as possible. Explain your answer.

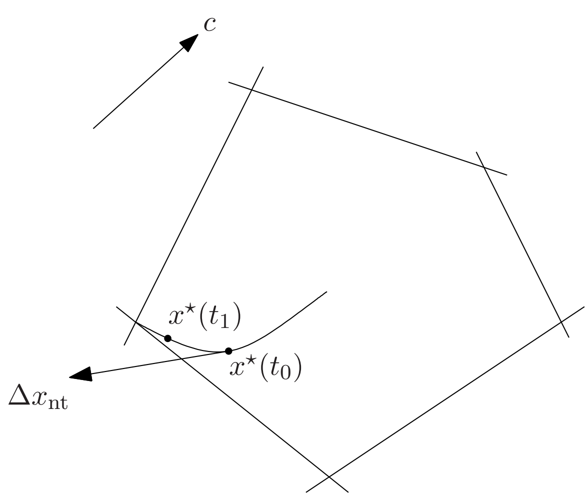
Exercise 78. Let $x^\star(t_0)$ be a point on the central path of the LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \le b, \end{array}$$

with $t_0 > 0$. We assume that A is $m \times n$ with $\mathbf{rank}(A) = n$. Define $\Delta x_{\rm nt}$ as the Newton step at $x^\star(t_0)$ for the function

$$t_1 c^T x - \sum_{i=1}^m \log(b_i - a_i^T x),$$

where a_i^T denotes the ith row of A, and $t_1 > t_0$. Show that $\Delta x_{\rm nt}$ is tangent to the central path at $x^\star(t_0)$.



Hint. Find an expression for the tangent direction $\Delta x_{\rm tg} = dx^\star(t_0)/dt$, and show that $\Delta x_{\rm nt}$ is a positive multiple of $\Delta x_{\rm tg}$.

Exercise 79. In the lecture on barrier methods, we noted that a point $x^*(t)$ on the central path yields a dual feasible point

$$z_i^*(t) = \frac{1}{t(b_i - a_i^T x^*(t))}, \quad i = 1, \ldots, m. \tag{73}$$

In this problem we examine what happens when $x^*(t)$ is calculated only approximately.

Suppose x is strictly feasible and v is the Newton step at x for the function

$$tc^T x + \phi(x) = tc^T x - \sum_{i=1}^{m} \log(b - a_i^T x).$$

Let $d \in \mathbf{R}^m$ be defined as $d_i = 1/(b_i - a_i^T x)$, $i = 1, \ldots, m$. Show that if

$$\lambda(x) = \| \mathbf{diag}(d) A v \| \leq 1,$$

then the vector

$$z = (d + \mathbf{diag}(d)^2 A v)/t$$

is dual feasible. Note that z reduces to (73) if $x = x^*(t)$ (and hence $v = 0$).

This observation is useful in a practical implementation of the barrier method. In practice, Newton's method provides an approximation of the central point $x^*(t)$, which means that the point (73) is not quite dual feasible, and a stopping criterion based on the corresponding dual bound is not quite accurate. The results derived above imply that even though $x^*(t)$ is not exactly centered, we can still obtain a dual feasible point, and use a completely rigorous stopping criterion.

Exercise 80. Let $\mathcal{P}$ be a polyhedron described by a set of linear inequalities:

$$\mathcal{P} = \{x \in \mathbf{R}^n \mid Ax \leq b\},$$

where $A \in \mathbf{R}^{m \times n}$ and $b \in \mathbf{R}^m$. Let ϕ denote the logarithmic barrier function

$$\phi(x) = -\sum_{i=1}^{m} \log(b_i - a_i^T x).$$

(a) Suppose $\widehat{x}$ is strictly feasible. Show that

$$(x - \widehat{x})^T \nabla^2 \phi(\widehat{x})(x - \widehat{x}) \leq 1 \quad \Longrightarrow \quad Ax \leq b,$$

where $\nabla^2 \phi(\widehat{x})$ is the Hessian of ϕ at $\widehat{x}$. Geometrically, this means that the set

$$\mathcal{E}_{\text{inner}} = \{x \mid (x - \widehat{x})^T \nabla^2 \phi(\widehat{x})(x - \widehat{x}) \leq 1\},$$

which is an ellipsoid centered at $\widehat{x}$, is enclosed in the polyhedron $\mathcal{P}$.

(b) Suppose $\widehat{x}$ is the analytic center of the inequalities $Ax < b$. Show that

$$Ax \leq b \quad \Longrightarrow \quad (x - \widehat{x})^T \nabla^2 \phi(\widehat{x})(x - \widehat{x}) \leq m(m-1).$$

In other words, the ellipsoid

$$\mathcal{E}_{\text{outer}} = \{x \mid (x - \widehat{x})^T \nabla^2 \phi(\widehat{x})(x - \widehat{x}) \leq m(m-1)\}$$

contains the polyhedron $\mathcal{P}$.

Exercise 81. Let $\hat{x}$ be the analytic center of a set of linear inequalities

$$a_k^T x \le b_k, \quad k = 1, \ldots, m.$$

Show that the kth inequality is redundant (*i.e.*, it can be deleted without changing the feasible set) if

$$b_k - a_k^T \hat{x} \ge m \sqrt{a_k^T H^{-1} a_k}$$

where H is defined as

$$H = \sum_{k=1}^{m} \frac{1}{(b_k - a_k^T \hat{x})^2} a_k a_k^T.$$

Exercise 82. The analytic center of a set of linear inequalities $Ax \le b$ depends not only on the geometry of the feasible set, but also on the representation (*i.e.*, A and b). For example, adding redundant inequalities does not change the polyhedron, but it moves the analytic center. In fact, by adding redundant inequalities you can make *any* strictly feasible point the analytic center, as you will show in this problem.

Suppose that $A \in \mathbf{R}^{m \times n}$ and $b \in \mathbf{R}^m$ define a bounded polyhedron

$$\mathcal{P} = \{x \mid Ax \le b\}$$

and that $x^\star$ satisfies $Ax^\star < b$. Show that there exist $c \in \mathbf{R}^n$, $\gamma \in \mathbf{R}$, and a positive integer q, such that

(a) $\mathcal{P}$ is the solution set of the $m + q$ inequalities

$$\begin{array}{l} Ax \le b \\ \left.\begin{array}{l} c^T x \le \gamma \\ c^T x \le \gamma \\ \quad \vdots \\ c^T x \le \gamma \end{array}\right\} q \text{ copies.} \end{array} \tag{74}$$

(b) $x^\star$ is the analytic center of the set of linear inequalities given in (74).

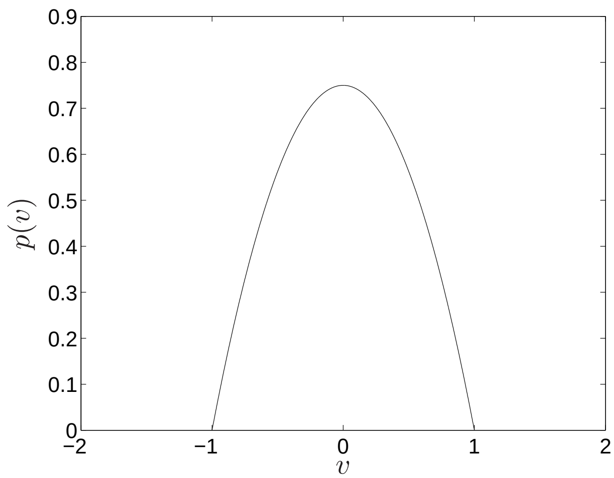
Exercise 83. *Maximum-likelihood estimation with parabolic noise density.* We consider the linear measurement model

$$y_i = a_i^T x + v_i, \quad i = 1, \ldots, m.$$

The vector $x \in \mathbf{R}^n$ is a vector of parameters to be estimated, $y_i \in \mathbf{R}$ are the measured or observed quantities, and v_i are the measurement errors or noise. The vectors $a_i \in \mathbf{R}^n$ are given. We assume that the measurement errors v_i are independent and identically distributed with a parabolic density function

$$p(v) = \begin{cases} (3/4)(1 - v^2) & |v| \le 1 \\ 0 & \text{otherwise} \end{cases}$$

(shown below).



Let $\bar{x}$ be the maximum-likelihood (ML) estimate based on the observed values y, *i.e.*,

$$\bar{x} = \underset{x}{\operatorname{argmax}} \left(\sum_{i=1}^{m} \log \left(1 - (y_i - a_i^T x)^2 \right) + m \log(3/4) \right).$$

Show that the true value of x satisfies

$$(x - \bar{x})^T H (x - \bar{x}) \leq 4m^2$$

where

$$H = 2 \sum_{i=1}^{m} \frac{1 + (y_i - a_i^T \bar{x})^2}{(1 - (y_i - a_i^T \bar{x})^2)^2} a_i a_i^T.$$

Exercise 84. *Potential reduction algorithm.* Consider the LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \leq b \end{array}$$

with $A \in \mathbf{R}^{m \times n}$. We assume that $\mathbf{rank}\, A = n$, that the problem is strictly feasible, and that the optimal value $p^\star$ is finite.

For $l < p^\star$ and $q > m$, we define the *potential function*

$$\varphi_{\text{pot}}(x) = q \log(c^T x - l) - \sum_{i=1}^{m} \log(b_i - a_i^T x).$$

The function φ_{pot} is defined for all strictly feasible x, and although it is not a convex function, it can be shown that it has a unique minimizer. We denote the minimizer as $x^\star_{\text{pot}}(l)$:

$$x^\star_{\text{pot}}(l) = \underset{Ax<b}{\operatorname{argmin}} \left(q \log(c^T x - l) - \sum_{i=1}^{m} \log(b_i - a_i^T x) \right).$$

(a) Show that $x^\star_{\text{pot}}(l)$ lies on the central path, *i.e.*, it is the minimizer of the function

$$tc^T x - \sum_{i=1}^{m} \log(b_i - a_i^T x)$$

for some value of t.

(b) Prove that the following algorithm converges and that it returns a suboptimal x with $c^T x - p^\star < \epsilon$.

> **given** $l < p^\star$, tolerance $\epsilon > 0$, $q > m$
> **repeat** $\{$
> 1. $x := x^\star_{\mathrm{pot}}(l)$
> 2. **if** $\frac{m(c^T x - l)}{q} < \epsilon$, **return**$(x)$
> 3. $l := \frac{q-m}{q} c^T x + \frac{m}{q} l$
>
> $\}$

Exercise 85. Consider the following variation on the barrier method for solving the LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \leq b_i, \quad i = 1, \ldots, m. \end{array}$$

We assume we are given a strictly feasible $\hat{x}$ (*i.e.*, $a_i^T \hat{x} < b_i$ for $i = 1, \ldots, m$), a strictly dual feasible $\hat{z}$ ($A^T \hat{z} + c = 0$, $\hat{z} > 0$), and a positive scalar ρ with $0 < \rho < 1$.

> **initialize:** $x = \hat{x}$, $w_i = (b_i - a_i^T \hat{x}) \hat{z}_i$, $i = 1, \ldots, m$
> **repeat:**
> 1. $x := \mathrm{argmin}_y (c^T y - \sum_{i=1}^m w_i \log(b_i - a_i^T y))$
> 2. $w := \rho w$

Give an estimate or a bound on the number of (outer) iterations required to reach an accuracy $c^T x - p^\star \leq \epsilon$.

Exercise 86. *The inverse barrier.* The inverse barrier of a set of linear inequalities

$$a_i^T x \leq b_i, \quad i = 1, \ldots, m,$$

is the function ψ, defined as

$$\psi(x) = \sum_{i=1}^m \frac{1}{b_i - a_i^T x}$$

for strictly feasible x. It can be shown that ψ is convex and differentiable on the set of strictly feasible points, and that $\psi(x)$ tends to infinity as x approaches the boundary of the feasible set.

Suppose $\hat{x}$ is strictly feasible and minimizes

$$c^T x + \psi(x).$$

Show that you construct from $\hat{x}$ a dual feasible point for the LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \leq b_i, \quad i = 1, \ldots, m. \end{array}$$

Exercise 87. Assume the primal and dual LPs

$$
\begin{array}{llll}
\text{(P)} & \text{minimize} & c^T x & \\
& \text{subject to} & Ax \leq b &
\end{array}
\qquad
\begin{array}{lll}
\text{(D)} & \text{maximize} & -b^T z \\
& \text{subject to} & A^T z + c = 0 \\
& & z \geq 0
\end{array}
$$

are strictly feasible. Let $\{x(t) \mid t > 0\}$ be the central path and define

$$
s(t) = b - Ax(t), \qquad z(t) = \frac{1}{t} \begin{bmatrix} 1/s_1(t) \\ 1/s_2(t) \\ \vdots \\ 1/s_m(t) \end{bmatrix}.
$$

(a) Suppose x^*, z^* are optimal for the primal and dual LPs, and define $s^* = b - Ax^*$. (If there are multiple optimal points, x^*, z^* denote an arbitrary pair of optimal points.) Show that

$$
z(t)^T s^* + s(t)^T z^* = \frac{m}{t}
$$

for all $t > 0$. From the definition of $z(t)$, this implies that

$$
\sum_{k=1}^{m} \frac{s_k^*}{s_k(t)} + \sum_{k=1}^{m} \frac{z_k^*}{z_k(t)} = m. \tag{75}
$$

(b) As t goes to infinity, the central path converges to the optimal points

$$
x_{\rm c}^* = \lim_{t\to\infty} x(t), \qquad s_{\rm c}^* = b - Ax_{\rm c}^* = \lim_{t\to\infty} s(t), \qquad z_{\rm c}^* = \lim_{t\to\infty} z(t).
$$

Define $I = \{k \mid s_{{\rm c},k}^* = 0\}$, the set of active constraints at $x_{\rm c}^*$. Apply (75) to $s^* = s_{\rm c}^*$, $z^* = z_{\rm c}^*$ to get

$$
\sum_{k\notin I} \frac{s_{{\rm c},k}^*}{s_k(t)} + \sum_{k\in I} \frac{z_{{\rm c},k}^*}{z_k(t)} = m.
$$

Use this to show that $z_{{\rm c},k}^* > 0$ for $k \in I$. This proves that the central path converges to a strictly complementary solution, *i.e.*, $s_{\rm c}^* + z_{\rm c}^* > 0$.

(c) The primal optimal set is the set of all x that are feasible and satisfy complementary slackness with $z_{\rm c}^*$:

$$
X_{\rm opt} = \{x \mid a_k^T x = b_k,\ k \in I,\ a_k^T x \leq b_k,\ k \notin I\}.
$$

Let x^* be an arbitrary primal optimal point. Show that

$$
\prod_{k\notin I} (b_k - a_k^T x^*) \leq \prod_{k\notin I} (b_k - a_k^T x_{\rm c}^*).
$$

Hint. Use the arithmetic-geometric mean inequality

$$
\left(\prod_{k=1}^{m} y_k \right)^{1/m} \leq \frac{1}{m} \sum_{k=1}^{m} y_k
$$

for nonnegative vectors $y \in \mathbf{R}^m$.

Exercise 88. The most expensive step in one iteration of an interior-point method for an LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \leq b \end{array}$$

is the solution of a set of linear equations of the form

$$A^T D A \, \Delta x = y, \tag{76}$$

where D is a positive diagonal matrix, the right-hand side y is a given vector, and Δx is the unknown. The values of D and y depend on the method used and on the current iterate, and are not important for our purposes here. For example, the Newton equation in the barrier methodm

$$\nabla^2 \phi(x) v = -tc - \nabla \phi(x),$$

is of the form (76). In the primal-dual method, we have to solve two sets of linear equations of the form (76) with $D = X^{-1} Z$.

It is often possible to speed up the algorithm significantly by taking advantage of special structure of the matrix A when solving the equations (76).

Consider the following three optimization problems that we encountered before in this course.

- *ℓ_1-minimization*:

  $$\text{minimize } \|Pu + q\|_1$$

  ($P \in \mathbf{R}^{r \times s}$ and $q \in \mathbf{R}^r$ are given; $u \in \mathbf{R}^s$ is the variable).
- *Constrained ℓ_1-minimization*:

  $$\begin{array}{ll} \text{minimize} & \|Pu + q\|_1 \\ \text{subject to} & -\mathbf{1} \leq u \leq \mathbf{1} \end{array}$$

  ($P \in \mathbf{R}^{r \times s}$ and $q \in \mathbf{R}^r$ are given; $u \in \mathbf{R}^s$ is the variable).
- *Robust linear programming* (see exercise 24):

  $$\begin{array}{ll} \text{minimize} & w^T u \\ \text{subject to} & Pu + \|u\|_1 \mathbf{1} \leq q \end{array}$$

  ($P \in \mathbf{R}^{r \times s}$, $q \in \mathbf{R}^r$, and $w \in \mathbf{R}^s$ are given; $u \in \mathbf{R}^s$ is the variable).

For each of these three problems, answer the following questions.

(a) Express the problem as an LP in inequality form. Give the matrix A, and the number of variables and constraints.

(b) What is the cost of solving (76) for the matrix A you obtained in part (a), if you do not use any special structure in A (knowing that the cost of solving a dense symmetric positive definite set of n linear equations in n variables is $(1/3)n^3$ operations, and the cost of a matrix-matrix multiplication $A^T A$, with $A \in \mathbf{R}^{m \times n}$, is mn^2 operations)?

(c) Work out the product $A^T DA$ (assuming D is a given positive diagonal matrix). Can you give an efficient method for solving (76) that uses the structure in the equations? What is the cost of your method (*i.e.*, the approximate number of operations when r and s are large) as a function of the dimensions r and s ?
Hint. Try to reduce the problem to solving a set of s linear equations in s variables, followed by a number of simple operations.

For the third problem, you can use the following formula for the inverse of a matrix $H + yy^T$, where y is a vector:

$$(H + yy^T)^{-1} = H^{-1} - \frac{1}{1 + y^T H^{-1} y} H^{-1} yy^T H^{-1}.$$

Exercise 89. In this problem you are asked to write a MATLAB code for the ℓ_1-approximation problem

$$\text{minimize} \quad \|Pu + q\|_1, \tag{77}$$

where $P = \mathbf{R}^{r \times s}$ and $q \in \mathbf{R}^r$. The calling sequence for the code is `u = l1(P,q)`. On exit, it must guarantee a relative accuracy of 10^{-6} or an absolute accuracy of 10^{-8}, *i.e.*, the code can terminate if

$$\|Pu + q\|_1 - p^\star \leq 10^{-6} \cdot p^\star$$

or

$$\|Pu + q\|_1 - p^\star \leq 10^{-8},$$

where $p^\star$ is the optimal value of (77). You may assume that P has full rank ($\mathbf{rank}\, P = s$).

We will solve the problem using Mehrotra's method as described in applied to the LP

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T v \\ \text{subject to} & \left[\begin{array}{cc} P & -I \\ -P & -I \end{array}\right] \left[\begin{array}{c} u \\ v \end{array}\right] \leq \left[\begin{array}{c} -q \\ q \end{array}\right]. \end{array} \tag{78}$$

We will take advantage of the structure in the problem to improve the efficiency.

(a) *Initialization.* Mehrotra's method can be started at infeasible primal and dual points. However good feasible starting points for the LP (78) are readily available from the solution u_{ls} of the least-squares problem

$$\text{minimize} \quad \|Pu + q\|$$

(in MATLAB: `u = -P\q`). As primal starting point we can use $u = u_{\text{ls}}$, and choose v so that we have strict feasibility in (78). To find a strictly feasible point for the dual of (78), we note that $P^T P u_{\text{ls}} = -P^T q$ and therefore the least-squares residual $r_{\text{ls}} = P u_{\text{ls}} + q$ satisfies

$$P^T r_{\text{ls}} = 0.$$

This property can be used to construct a strictly feasible point for the dual of (78). You should try to find a dual starting point that provides a positive lower bound on $p^\star$, *i.e.*, a lower bound that is better than the trivial lower bound $p^\star \geq 0$.

Since the starting points are strictly feasible, all iterates in the algorithm will remain strictly feasible, and we don't have to worry about testing the deviation from feasibility in the convergence criteria.

(b) As we have seen, the most expensive part of an iteration in Mehrotra's method is the solution of two sets of equations of the form

$$A^T X^{-1} Z A \Delta x = r_1 \tag{79}$$

where X and Z are positive diagonal matrices that change at each iteration. One of the two equations is needed to determine the affine-scaling direction; the other equation (with a different right-hand side) is used to compute the combined centering-corrector step. In our application, (79) has $r+s$ equations in $r+s$ variables, since

$$A = \begin{bmatrix} P & -I \\ -P & -I \end{bmatrix}, \qquad \Delta x = \begin{bmatrix} \Delta u \\ \Delta v \end{bmatrix}.$$

By exploiting the special structure of A, show that you can solve systems of the form (79) by solving a smaller system of the form

$$P^T D P \Delta u = r_2, \tag{80}$$

followed by a number of inexpensive operations. In (80) D is an appropriately chosen positive diagonal matrix. This observation is important, since it means that the cost of one iteration reduces to the cost of solving two systems of size $s \times s$ (as opposed to $(r+s) \times (r+s)$). In other words, although we have introduced r new variables to express (77) as an LP, the extra cost of introducing these variables is marginal.

(c) Test your code on randomly generated P and q. Plot the duality gap (on a logarithmic scale) versus the iteration number for a few examples and include a typical plot with your solutions.

Exercise 90. This problem is similar to the previous problem, but instead we consider the constrained ℓ_1-approximation problem

$$\begin{array}{ll} \text{minimize} & \|Pu+q\|_1 \\ \text{subject to} & -\mathbf{1} \preceq u \preceq \mathbf{1} \end{array} \tag{81}$$

where $P = \mathbf{R}^{r \times s}$ and $q \in \mathbf{R}^r$. The calling sequence for the code is `u = cl1(P,q)`. On exit, it must guarantee a relative accuracy of 10^{-5} or an absolute accuracy of 10^{-8}, *i.e.*, the code can terminate if

$$\|Pu+q\|_1 - p^\star \le 10^{-5} \cdot p^\star$$

or

$$\|Pu+q\|_1 - p^\star \le 10^{-8},$$

where $p^\star$ is the optimal value of (81). You may assume that P has full rank (**rank** $P = s$).

We will solve the problem using Mehrotra's method as described in applied to the LP

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T v \\ \text{subject to} & \begin{bmatrix} P & -I \\ -P & -I \\ I & 0 \\ -I & 0 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \preceq \begin{bmatrix} -q \\ q \\ \mathbf{1} \\ \mathbf{1} \end{bmatrix}. \end{array} \tag{82}$$

We will take advantage of the structure in the problem to improve the efficiency.

(a) *Initialization.* For this problem it is easy to determine strictly feasible primal and dual points at which the algorithm can be started. This has the advantage that all iterates in the algorithm will remain strictly feasible, and we don't have to worry about testing the deviation from feasibility in the convergence criteria.

As primal starting point, we can simply take $u = 0$, and a vector v that satisfies $v_i > |(Pu + q)_i|$, $i = 1, \ldots, r$. What would you choose as dual starting point?

(b) As we have seen, the most expensive part of an iteration in Mehrotra's method is the solution of two sets of equations of the form

$$A^T X^{-1} Z A \Delta x = r_1 \tag{83}$$

where X and Z are positive diagonal matrices that change at each iteration. One of the two equations is needed to determine the affine-scaling direction; the other equation (with a different right-hand side) is used to compute the combined centering-corrector step. In our application, (83) has $r + s$ equations in $r + s$ variables, since

$$A = \begin{bmatrix} P & -I \\ -P & -I \\ I & 0 \\ -I & 0 \end{bmatrix}, \qquad \Delta x = \begin{bmatrix} \Delta u \\ \Delta v \end{bmatrix}.$$

By exploiting the special structure of A, show that you can solve systems of the form (83) by solving a smaller system of the form

$$(P^T \tilde{D} P + \hat{D}) \Delta u = r_2, \tag{84}$$

followed by a number of inexpensive operations. The matrices $\tilde{D}$ and $\hat{D}$ in (84) are appropriately chosen positive diagonal matrices.

This observation is important, since the cost of solving (84) is roughly equal to the cost of solving the least-squares problem

$$\text{minimize} \quad \|Pu + q\|.$$

Since the interior-point method converges in very few iterations (typically less than 10), this allows us to conclude that the cost of solving (81) is roughly equal to the cost of 10 least-squares problems of the same dimension, in spite of the fact that we introduced r new variables to cast the problem as an LP.

(c) Test your code on randomly generated P and q. Plot the duality gap (on a logarithmic scale) versus the iteration number for a few examples and include a typical plot with your solutions.

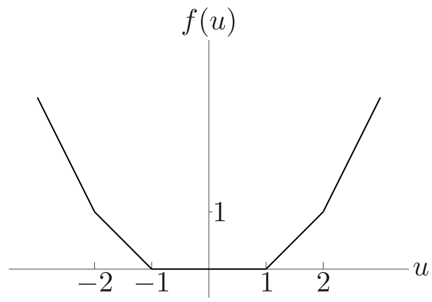
Exercise 91. Consider the optimization problem

$$\text{minimize} \quad \sum_{i=1}^{m} f(a_i^T x - b_i)$$

where

$$f(u) = \begin{cases} 0 & |u| \leq 1 \\ |u| - 1 & 1 \leq |u| \leq 2 \\ 2|u| - 3 & |u| \geq 2. \end{cases}$$

The function f is shown below.



The problem data are $a_i \in \mathbf{R}^n$ and $b_i \in \mathbf{R}$.

(a) Formulate this problem as an LP in inequality form

$$\begin{array}{ll} \text{minimize} & \bar{c}^T \bar{x} \\ \text{subject to} & \bar{A}\bar{x} \leq \bar{b}. \end{array} \tag{85}$$

Carefully explain why the two problems are equivalent, and what the meaning is of any auxiliary variables you introduce.

(b) Describe an efficient method for solving the equations

$$\bar{A}^T D \bar{A} \Delta \bar{x} = r$$

that arise in each iteration of Mehrotra's method applied to the LP (85). Here D is a given diagonal matrix with positive diagonal elements, and r is a given vector.

Compare the cost of your method with the cost of solving the least-squares problem

$$\text{minimize} \quad \sum_{i=1}^{m} (a_i^T x - b_i)^2.$$

Exercise 92. The most time consuming step in a primal-dual interior-point method for solving an LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \leq b \end{array}$$

is the solution of linear equations of the form

$$\left[\begin{array}{ccc} 0 & A & I \\ A^T & 0 & 0 \\ X & 0 & Z \end{array}\right] \left[\begin{array}{c} \Delta z \\ \Delta x \\ \Delta s \end{array}\right] = \left[\begin{array}{c} r_1 \\ r_2 \\ r_3 \end{array}\right],$$

where X and Z are positive diagonal matrices. After eliminating Δs from the last equation we obtain

$$\left[\begin{array}{cc} -D & A \\ A^T & 0 \end{array}\right] \left[\begin{array}{c} \Delta z \\ \Delta x \end{array}\right] = \left[\begin{array}{c} d \\ f \end{array}\right]$$

where $D = XZ^{-1}$, $d = r_1 - Z^{-1} r_3$, $f = r_2$.

Describe an efficient method for solving this equation for an LP of the form

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Px \leq q \\ & -\mathbf{1} \leq x \leq \mathbf{1}, \end{array}$$

where $P \in \mathbf{R}^{m \times n}$ is a dense matrix. Distinguish two cases: $m \gg n$ and $m \ll n$.

Exercise 93. A network is described as a directed graph with m arcs or links. The network supports n flows, with nonnegative rates $x_1, \ldots, x_n$. Each flow moves along a fixed, or pre-determined, path or route in the network, from a source node to a destination node. Each link can support multiple flows, and the total traffic on a link is the sum of the rates of the flows that travel over it. The total traffic on link i can be expressed as $(Ax)_i$, where $A \in \mathbf{R}^{m \times n}$ is the flow-link incidence matrix defined as

$$A_{ij} = \begin{cases} 1 & \text{flow } j \text{ passes through link } i \\ 0 & \text{otherwise.} \end{cases}$$

Usually each path passes through only a small fraction of the total number of links, so the matrix A is sparse.

Each link has a positive capacity, which is the maximum total traffic it can handle. These link capacity constraints can be expressed as $Ax \preceq b$, where b_i is the capacity of link i.

We consider the network rate optimization problem

$$\begin{array}{ll} \text{maximize} & f_1(x_1) + \cdots + f_n(x_n) \\ \text{subject to} & Ax \preceq b \\ & x \succeq 0, \end{array}$$

where

$$f_k(x_k) = \begin{cases} x_k & x_k \leq c_k \\ (x_k + c_k)/2 & x_k \geq c_k, \end{cases}$$

and $c_k > 0$ is given. In this problem we choose feasible flow rates x_k that maximize a utility function $\sum_k f_k(x_k)$.

(a) Express the network rate optimization problem as a linear program in inequality form.

(b) Derive the dual problem and show that it is equivalent to

$$\begin{array}{ll} \text{minimize} & b^T z + g_1(a_1^T z) + \cdots + g_n(a_n^T z) \\ \text{subject to} & A^T z \succeq (1/2)\mathbf{1} \\ & z \succeq 0 \end{array}$$

with variables $z \in \mathbf{R}^m$, where a_k is the kth column of A and

$$g_k(y) = \begin{cases} (1-y)c_k & y \leq 1 \\ 0 & y \geq 1. \end{cases}$$

(c) Suppose you are asked to write a custom implementation of the primal-dual interior-point method for the linear program in part 1. Give an efficient method for solving the linear equations that arise in each iteration of the algorithm. Justify your method, assuming that m and n are very large, and that the matrix $A^T A$ is sparse.