Convex Optimization

Homework 6

Spring 1401
Due date: 30th of Ordibehesht

---

1. *Uniqueness of projection* Show that if $C \subseteq \mathbf{R}^n$ is nonempty, closed and convex, and the norm $\|.\|$ is strictly convex, then for every x_0 there is exactly one $x \in C$ closest to X_0. In other words the projection of X_0 on C is unique.

2. *Maximum volume rectangle inside a polyhedron.* Formulate the following problem as a convex optimization problem. Find the rectangle

$$\mathcal{R} = \{x \in \mathbf{R}^n \mid \ell \preceq x \preceq u\}$$

of maximum volume, enclosed in a polyhedron $\mathcal{P} = \{x \mid Ax \preceq b\}$. The variables are $\ell, u \in \mathbf{R}^n$.Your formulation should not involve an exponential number of constraints.

3. A set of n teams compete in a tournament. We model each team's ability by a number $a_j \in [0,1]$, $j = 1, ..., n$. When teams j and k play each other, the probability that team j wins is equal to $\mathbf{prob}(a_j - a_k + \nu > 0)$, where ν is a symmetric random variable with density

$$p(\nu) = \frac{2\sigma^{-1}}{(e^{\nu/\sigma} + e^{-\nu/\sigma})^2},$$

where σ controls the standard deviation of ν. For this question, you will likely find it useful that the cumulative distribution function (CDF) of ν is

$$F(t) = \int_{-\infty}^{t} p(\nu) d\nu = \frac{e^{t/\sigma}}{e^{t/\sigma} + e^{-t/\sigma}}.$$

You are given the outcome of m past games. These are organized as

$$(j^{(i)}, k^{(i)}, y^{(i)}), \qquad i = 1, ..., m,$$

meaning that game i was played between teams $j^{(i)}$ and $k^{(i)}$; $y^{(i)} = 1$ means that team $j^{(i)}$ won, while $y^{(i)} = -1$ means that team $k^{(i)}$ won. (We assume there are no ties.)

(a) Formulate the problem of finding the maximum likelihood estimate of team abilities, $\hat{a} \in \mathbf{R}^n$, given the outcomes, as a convex optimization problem. You will find the game incidence matrix $A \in \mathbf{R}^{m \times n}$, defined as

$$A_{il} = \begin{cases} y^{(i)} & l = j^{(i)} \\ -y^{(i)} & l = k^{(i)} \\ 0 & \textit{otherwise}, \end{cases}$$

useful.
The prior constraints $\hat{a} \in [0,1]$ should be included in the problem formulation. Also, we note that if a constant is added to all team abilities, there is no change in the probabilities of game outcomes. This means that $\hat{a}$ is determined only up to a constant, like a potential. But this doesn't affect the ML estimation problem, or any subsequent predictions made using the estimated parameters.

(b) Find $\hat{a}$ for the team data given in *team_data.m*, in the matrix *train*. (This matrix gives the outcomes for a tournament in which each team plays each other team once.) You may find the CVX function *log_normcdf* helpful for this problem.
You can form A using the commands

```
using SparseArrays;
A1 = sparse(1:m, train[:, 1], train[:,3], m, n);
A2 = sparse(1:m, train[:, 2], -train[:,3], m, n);
A = A1 + A2;
```

(c) Use the maximum likelihood estimate $\hat{a}$ found in part (b) to predict the outcomes of next year's tournament games, given in the matrix *test*, using $\hat{y}^{(i)} = \mathbf{sign}(\hat{a}_{j^{(i)}} - \hat{a}_{k^{(i)}})$. Compare these predictions with the actual outcomes, given in the third column of *test*. Give the fraction of correctly predicted outcomes.
The games played in *train* and *test* are the same, so another, simpler method for predicting the outcomes in *test* it to just assume the team that won last year's match will also win this year's match. Give the percentage of correctly predicted outcomes using this simple method.

4. *Internal rate of return for cash streams with a single initial investment.* We use the notation of example 3.34 in the textbook. Let $x \in \mathbf{R}^{n+1}$ be a cash flow over n periods, with x indexed from 0 to n, where the index denotes period number. We assume that $x_0 < 0$, $x_j \geq 0$ for $j = 1, 2, \ldots, n$, and $x_0 + \cdots + x_n > 0$. This means that there is an initial positive investment; thereafter, only payments are made, with the total of the payments exceeding the initial investment. (In the more general setting of example 3.34, we allow additional investments to be made after the initial investment.)

(a) Show that $\mathrm{IRR}(x)$ is quasilinear in this case.

(b) *Blending initial investment only streams.* Use the result in part (a) to show the following: Let $x^{(i)} \in \mathbf{R}^{n+1}$, $i = 1, \ldots, k$, be a set of k cash flows over n periods, each of which satisfies the conditions above. Let $\omega \in \mathbf{R}^k_+$, with $\mathbf{1}^T\omega = 1$, and consider the blended cash flow given by $x = \omega_1 x^{(1)} + \cdots + \omega_k x^{(k)}$. (We can think of this as investing a fraction ω_i in cash flow i.) Show that $\mathrm{IRR}(x) \leq \max_i \mathrm{IRR}(x^{(i)})$. Thus, blending a set of cash flows (with initial investment only) will not improve the IRR over the best individual IRR of the cash flows.

5. *Ellipsoidal peeling.* In this problem, you will implement an outlier identification technique using Lowner-John ellipsoids. Given a set of points $\mathcal{D} = \{x_1, \ldots, x_N\}$ in $\mathbf{R}^n$, the goal is to identify a set $\mathcal{O} \subseteq \mathcal{D}$ that are anomalous in some sense. Roughly speaking, we think of an outlier as a point that is far away from most of the points, so we would like the points in $\mathcal{D} \backslash \mathcal{O}$ to be relatively close together, and to be relatively far apart from the points in $\mathcal{O}$.

We describe a heuristic technique for identifying $\mathcal{O}$. We start with $\mathcal{O} = \emptyset$ and find the minimum volume (Lowner-John) ellipsoid $\mathcal{E}$ containing all $x_i \notin \mathcal{O}$ (which is all x_i in the first step). Each iteration, we flag (*i.e.* add to $\mathcal{O}$) the point that corresponds to the largest dual variable for the constraint $x_i \in \mathcal{E}$; this point will be one of the points on the boundary of $\mathcal{E}$, and intuitively, it will be the one for whom the constraint is 'most' binding. We then plot $\mathbf{vol}\ \mathcal{E}$ (on a log scale) versus $\mathbf{card}\ \mathcal{O}$ and hope that we see a sharp drop in the curve. We use the value of $\mathcal{O}$ after the drop.The hope is that after removing a relatively small number of points, the volume of the minimum volume ellipsoid containing the remaining points will be much smaller than the minimum volume ellipsoid for $\mathcal{D}$, which means the removed points are far away from the others.

For example, suppose we have 100 points that lie in the unit ball and 3 points with (Euclidean) norm 1000. Intuitively, it is clear that it is reasonable to consider the three large points outliers. The minimum volume ellipsoid of all 103 points will have very large volume. The three points will be the first ones removed, and as soon as they are, the volume of the ellipsoid ellipsoid will drop dramatically and be on the order of the volume of the unit ball.

Run 6 iterations of the algorithm on the data given in ellip_anomaly_data.m. plot $\mathbf{vol}\ \mathcal{E}$ (on a log scale) versus $\mathbf{card}\ \mathcal{O}$. In addition, on a single plot, plot all the ellipses found with the function ellipse_draw(A,b) along with the outliers (in red) and the remaining points (in blue).

Of course, we have chosen an example in $\mathbf{R}^2$ so the ellipses can be plotted, but one can detect outliers in $\mathbf{R}^2$ simply by inspection. In dimension much higher than 3, however, detecting outliers by plotting will become substantially more difficult, while the same algorithm can be used.

Note. In CVX, you should use det_rootn (which is SDP-representable and handled exactly) instead of log_det (which is handled using an inefficient iterative procedure).

6. **Optional**. *Planning production with uncertain demand.* You must order (nonnegative) amounts $r_1, \ldots, r_m$ of raw materials, which are needed to manufacture (nonnegative) quantities $q_1, \ldots, q_n$ of n different products. To manufacture one unit of product j requires at least A_{ij} units of raw material i, so we must have $r \succeq Aq$. (We will assume that A_{ij} are nonnegative.) The per-unit cost of the raw materials is given by $c \in \mathbf{R}^m_+$, so the total raw material cost is $c^T r$.

   The (nonnegative) demand for product j is denoted d_j; the number of units of product j sold is $s_j = \min\{q_j, d_j\}$. (When $q_j > d_j$, $q_j - d_j$ is the amount of product j produced, but not sold; when $d_j > q_j$, $d_j - q_j$ is the amount of unmet demand.) The revenue from selling the products is $p^T s$, where $p \in \mathbf{R}^n_+$ is the vector of product prices. The profit is $p^T s - c^T r$. (Both d and q are real vectors; their entries need not be integers.)

   You are given A, c and p. The product demand, however, is not known. Instead, a set of K possible demand vectors, $d^{(1)}, \ldots, d^{(K)}$, with associated probabilities $\pi_1, \ldots, \pi_K$, is given. (These satisfy $\mathbf{1}^T \pi = 1,\ \pi \succeq 0$.)

   You will explore two different optimization problems that arise in choosing r and q (the variables).

   I. Choose r and q ahead of time. You must choose r and q, knowing only the data listed above. (In other words, you must order the raw materials, and commit to producing the chosen quantities of products, before you know the product demand.) The objective is to maximize the expected profit.

   II. Choose r ahead of time, and q after d is known. You must choose r, knowing only the data listed above. Some time after you have chosen r, the demand will become known to you. This means that you will find out which of the K demand vectors is the true demand. Once you know this, you must choose the quantities to be manufactured. (In other words, you must order the raw materials before the product demand is known; but you can choose the mix of products to manufacture after you have learned the true product demand.) The objective is to maximize the expected profit.

   (a) Explain how to formulate each of these problems as a convex optimization problem. Clearly state what the variables are in the problem, what the constraints are, and describe the roles of any auxiliary variables or constraints you introduce.

   (b) Carry out the methods from part (a) on the problem instance with numerical data given in planning_data.m. This file will define A, D, K, c, m, n, p and pi. The K columns of D are the possible demand vectors. For both of the problems described above, give the optimal value of r, and the expected profit.

7. **Optional**. In circuit design, we must select the widths of a set of n components, given by the vector $\omega = (\omega_1, \ldots, \omega_n)$, which must satisfy width limits

$$W^{\min} \leq \omega_i \leq W^{\max}, \quad i = 1, \ldots, n,$$

   where $W^{\min}$ and $W^{\max}$ are given (positive) values. (You can assume there are no other constraints on ω.) The design is judged by three objectives, each of which we would like to be small: the circuit power $P(\omega)$, the circuit delay $D(\omega)$, and the total circuit area $A(\omega)$. These three objectives are (complicated) posynomial functions of ω. You do not know the functions P, D, or A. (That is, you do not know the coefficients or exponents in the posynomial expressions.) You do know a set of k designs, given by $\omega^{(1)}, \ldots, \omega^{(k)} \in \mathbf{R}^n$, and their associated objective values

$$P(\omega^{(j)}),\ D(\omega^{(j)}),\ A(\omega^{(j)}), \quad j = 1, \ldots, k.$$

   You can assume that these designs satisfy the width limits. The goal is to find a design w that satisfies the width limits, and the design specifications

$$P(\omega) \leq P_{\text{spec}},\ D(\omega) \leq D_{\text{spec}},\ A(\omega) \leq A_{\text{spec}}$$

   where P_{spec}, D_{spec}, and A_{spec} are given. Now consider the specific data given in blend_design_data.m. Give the following.

(a) A feasible design (*i.e.*, ω) that satisfies the specifications.

(b) A clear argument as to how you know that your design satisfies the specifications, even though you do not know the formulas for P, D, and A.

(c) Your method for finding w, including any code that you write.

Good Luck!