Convex Optimization
Homework 5

Spring 1400
Due date: 10th of Ordibehesht

---

1. A set of n teams compete in a tournament. We model each team's ability by a number $a_j \in [0,1], j = 1,\dots,n$. When teams j and k play each other, the probability that team j wins is equal to $\mathbf{prob}(a_j - a_k + v > 0)$, where $v \sim \mathcal{N}(0,\sigma^2)$. You are given the outcome of m past games. These are organized as:
$$(j^{(i)}, k^{(i)}, k^{(i)}), \quad i = 1,\dots m$$
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, $\hat{a} \in \mathbb{R}^n$, given the outcomes, as a convex optimization problem. You will find the game incidence matrix $A \in \mathbb{R}^{m\times n}$, defined as
$$A_{il} = \begin{cases} y^{(i)} & l = j^{(i)} \\ -y^{(i)} & l = k^{(i)} \\ 0 & o.w., \end{cases}$$
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

```
A = sparse(1:m,train(:,1),train(:,3),m,n) + ...
 sparse(1:m,train(:,2),-train(:,3),m,n);
```

   (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.

2. We consider the measurement setup
$$y_i = f(a_i^T x + b_i + v_i), \quad i = 1,\dots,m$$
where $x \in \mathbb{R}^n$ is the vector to be estimated, $y_i \in \mathbb{R}$ are the measurements, $a_i \in \mathbb{R}^n$, $b_i \in \mathbb{R}$ are known, and v_i are IID noises with log-concave probability density. The function $f : \mathbb{R} \to \mathbb{R}$, which represents a measurement nonlinearity, is *not* known. However, it is known that $f(t) \in [l,u]$ for all t, where $0 < l < u$ are given.

   Explain how to use convex optimization to find a maximum likelihood estimate of x, as well as the function f. (This is an infinite-dimensional ML estimation problem, but you can be informal in your approach and explanation.)

3. (A specific instance of the previous problem.) We want to estimate a vector $x \in \mathbb{R}^n$, given some measurements

$$y_i = \phi(a_i^T x + v_i), \quad i = 1, \dots, m$$

Here $a_i \in \mathbb{R}^n$ are known, v_i are IID $\mathcal{N}(0, \sigma^2)$ random noises, and $\phi : \mathbb{R} \to \mathbb{R}$ is an unknown monotonic increasing function, known to satisfy

$$\alpha \leq \phi'(u) \leq \beta$$

for all u. (Here α and β are known positive constants, with $\alpha < \beta$.) We want to find a maximum likelihood estimate of x and ϕ, given y_i. (We also know a_i, σ, α, and β.)

This sounds like an infinite-dimensional problem, since one of the parameters we are estimating is a function. In fact, we only need to know the m numbers $z_i = \phi^{-1}(y_i), i = 1, \dots, m$. So by estimating ϕ we really mean estimating the m numbers $z_1, \dots, z_m$. (These numbers are not arbitrary; they must be consistent with the prior information $\alpha \leq \phi(u) \leq \beta$ for all u.)

(a) Explain how to find a maximum likelihood estimate of x and ϕ (i.e., $z_1, \dots, z_m$) using convex optimization.

(b) Carry out your method on the data given in `nonlin_meas_data.*`, which includes a matrix $A \in \mathbb{R}^{m \times n}$, with rows $a_1^T, ..., a_m^T$. Give $\hat{x}_{ml}$, the maximum likelihood estimate of x. Plot your estimated function $\hat{\phi}_{ml}$. (You can do this by plotting $(\hat{z}_{ml})_i$ versus y_i, with y_i on the vertical axis and $(\hat{z}_{ml})_i$ on the horizontal axis.)

Hint. You can assume the measurements are numbered so that y_i are sorted in nondecreasing order, i.e., $y_1 \leq y_2 \leq \ \leq y_m$. (The data given in the problem instance for part (b) is given in this order.)

4. A grayscale image is represented as an $m \times n$ matrix of intensities U^{orig}. You are given the values U_{ij}^{orig} , for $(i, j) \in \mathcal{K}$, where $\mathcal{K} \subset \{1, \dots, m\} \times \{1, \dots, n\}$. Your job is to *interpolate* the image, by guessing the missing values. The reconstructed image will be represented by $U \in \mathbb{R}^{m \times n}$, where U satisfies the interpolation conditions $U_{ij} = U_{ij}^{orig}$ for $(i, j) \in \mathcal{K}$.

The reconstruction is found by minimizing a roughness measure subject to the interpolation conditions. One common roughness measure is the ℓ_2 variation (squared),

$$\sum_{i=2}^{m} \sum_{j=1}^{n} (U_{ij} - U_{i-1,j})^2 + \sum_{i=1}^{m} \sum_{j=2}^{n} (U_{ij} - U_{i,j-1})^2$$

Another method minimizes instead the *total variation*,

$$\sum_{i=2}^{m} \sum_{j=1}^{n} |U_{ij} - U_{i-1,j}| + \sum_{i=1}^{m} \sum_{j=2}^{n} |U_{ij} - U_{i,j-1}|$$

Evidently both methods lead to convex optimization problems.

Carry out ℓ_2 and total variation interpolation on the problem instance with data given in `tv_img_interp.m`. This will define `m`, `n`, and matrices `Uorig` and `Known`. The matrix `Known` is $m \times n$, with (i, j) entry one if $(i, j) \in \mathcal{K}$, and zero otherwise. The mfile also has skeleton plotting code. (We give you the entire original image so you can compare your reconstruction to the original; obviously your solution cannot access U_{ij}^{orig} for $(i, j) \notin \mathcal{K}$.)

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

6. Formulate the following problem as a convex optimization problem. Find the minimum volume ellipsoid $\mathcal{E} = \{x \mid (x - x_0)^T A^{-1} (x - x_0) \leq 1\}$ that contains K given ellipsoids

$$\mathcal{E}_i = \{x \mid x^T A_i x + 2b_i^T x + c_i \leq 0\}, \quad i = 1, \dots, K.$$

Hint. See appendix B.

7. Consider the robust linear discrimination problem given in (8.23), page 424 of the book.

   (a) Show that the optimal value t^* is positive if and only if the two sets of points can be linearly separated. When the two sets of points can be linearly separated, show that the inequality $\|a\|_2 \leq 1$ is tight, i.e., we have $\|a^*\|_2 = 1$, for the optimal a^*.

   (b) Using the change of variables $\tilde{a} = a/t$, $\tilde{b} = b/t$, prove that the problem (8.23) is equivalent to the QP

$$\begin{array}{ll} \text{minimize} & \|\tilde{a}\|_2 \\ \text{subject to} & \tilde{a}^T x_i - \tilde{b} \geq 1, \quad i = 1, \ldots, N \\ & \tilde{a}^T y_i - \tilde{b} \leq -1, \quad i = 1, \ldots, M \end{array}$$

Good luck!