ECE 8823 (Convex Optimization), Spring 2015

Homework #2

Due Monday February 6, in class

Reading: Boyd and Vandenberghe, Chapters 2 and 3

1. Using you class notes, prepare a 1-2 paragraph summary of what we talked about in class in the last week. I do not want just a bulleted list of topics, I want you to use complete sentences and establish context (Why is what we have learned relevant? How does it connect with other things you have learned here or in other classes?). The more insight you give, the better.

2. Let $\boldsymbol{X}_0$ be an $N \times N$ symmetric matrix with eigenvalue decomposition

$$\boldsymbol{X}_0 = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{T}} = \sum_{n=1}^{N} \lambda_n \boldsymbol{v}_n \boldsymbol{v}_n^{\mathrm{T}}.$$

Show that the closest symmetric positive semidefinite matrix to $\boldsymbol{X}_0$ can be calculated simply by truncating the terms with negative eigenvalues in the expression above:

$$\arg\min_{\boldsymbol{X}\in S_+^N} \|\boldsymbol{X}_0 - \boldsymbol{X}\|_F = \sum_{n=1}^{N} \max(\lambda_n, 0)\boldsymbol{v}_n\boldsymbol{v}_n^{\mathrm{T}} =: \hat{\boldsymbol{X}}.$$

Do this by establishing the "obtuseness" property for $\hat{\boldsymbol{X}}$,

$$\langle \boldsymbol{X} - \hat{\boldsymbol{X}}, \boldsymbol{X}_0 - \hat{\boldsymbol{X}} \rangle \leq 0 \quad \text{for all } \boldsymbol{X} \in S_+^N.$$

(Recall that the inner product between two matrices is

$$\langle \boldsymbol{X}, \boldsymbol{Y} \rangle = \sum_{n,\ell} X_{n,\ell} Y_{n,\ell} = \mathrm{trace}(\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{X}).$$

It is easy to see that $\langle \boldsymbol{X}, \boldsymbol{X} \rangle = \|\boldsymbol{X}\|_F^2 = \sum_{n,\ell} X_{n,\ell}^2$. It is always useful to remember that if one of the matrices in the inner product is rank-1, then you can re-write the inner product as a quadratic form; for any $\boldsymbol{v} \in \mathbb{R}^N$, we have $\langle \boldsymbol{X}, \boldsymbol{v}\boldsymbol{v}^{\mathrm{T}} \rangle = \boldsymbol{v}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{v}$.)

3. Let $\mathcal{C}$ and $\mathcal{D}$ be closed convex sets in $\mathbb{R}^N$, with $\mathcal{C} \subset \mathcal{D}$.

   (a) Does $P_{\mathcal{C}}(P_{\mathcal{D}}(\boldsymbol{x}_0)) = P_{\mathcal{C}}(\boldsymbol{x}_0)$? Prove that it is true or provide a counter example.

   (b) Does your answer change if $\mathcal{D}$ is a subspace?

   (c) Given an arbitrary $N \times N$ matrix $\boldsymbol{X}_0$, how do you compute the closest symmetric matrix to $\boldsymbol{X}_0$?

   (d) Given an arbitrary $N \times N$ matrix $\boldsymbol{X}_0$, how do you compute the closest symmetric postive semidefinite matrix to $\boldsymbol{X}_0$?

Last updated 10:15, January 27, 2017

4. Let f_i be the function that takes a symmetric positive definite matrix $\boldsymbol{X}$ and returns the ith entry along the diagonal of the inverse of $\boldsymbol{X}$:

$$f_i(\boldsymbol{X}) = \boldsymbol{X}^{-1}[i,i], \quad \boldsymbol{X} \in S^N_{++}.$$

Show that f_i is convex.

5. (a) Let $f_1(\boldsymbol{x})$ and $f_2(\boldsymbol{x})$ be convex functions on $\mathbb{R}^N$. Show that

$$f(\boldsymbol{x}) = \max\{f_1(\boldsymbol{x}), f_2(\boldsymbol{x})\}$$

is convex.

(b) Illustrate the above in $\mathbb{R}^1$ by making a sketch with affine functions $f_1(x) = a_1x + b_1$ and $f_2(x) = a_2x + b_2$. You may choose a_1, b_1, a_2, b_2 to your liking.

(c) Is it necessarily true that

$$f(\boldsymbol{x}) = \min\{f_1(\boldsymbol{x}), f_2(\boldsymbol{x})\}, \quad f_1, f_2 \text{ convex},$$

is convex? Sketch an example in $\mathbb{R}^1$ that supports your argument.

(d) Let $\{f_\tau(\boldsymbol{x})\}$ be a collection of convex functions on $\mathbb{R}^N$ that are indexed by $\tau \in \mathcal{T}$ — this set could be finite, countable (e.g. $\mathcal{T} = \mathbb{N} = \{1, 2, 3, 4, \ldots\}$), or uncountable (e.g. $\mathcal{T} = [a, b]$, an interval of the real line). Show that

$$f(\boldsymbol{x}) = \sup_{\tau \in \mathcal{T}} f_\tau(\boldsymbol{x})$$

is convex. (If you did part (a) cleanly, this extension will be straightforward.)

(e) Show that the conjugate function

$$f^*(\boldsymbol{a}) = \sup_{\boldsymbol{x} \in \mathcal{C}} \left[\langle \boldsymbol{x}, \boldsymbol{a} \rangle - f(\boldsymbol{x})\right], \quad \mathcal{C} = \operatorname{dom} f,$$

is convex, and that this is true whether or not f is convex.

6. You have a large amount of money M that you are going to gamble on a horse race. You want to be smart about it, though.

There are N horses running in the race. You will divide up your money to place a bet of x_i on each of them. Clearly,

$$\sum_{i=1}^{N} x_i = M.$$

As with any parimutuel betting scenario, if horse i wins, the payout to you is proportional to the amount you bet on horse i versus what everybody else (the "public") bet on horse i. If you wager x_i on horse i and the public wagers s_i then

$$\begin{aligned}\text{payout if horse } i \text{ wins} &= C \cdot (\text{total amount of money bet on all horses}) \cdot \frac{x_i}{x_i + s_i} \\ &= C \cdot \left(M + \sum_{i=1}^{N} s_i\right) \frac{x_i}{x_i + s_i}.\end{aligned}$$

The constant C above is less than 1, and represents the fact that the track takes a cut of all the bets (the "vigorish" is $1 - C$). A typical value of C might be 0.8 or 0.9.

The reason you are betting is that you have two pieces of key knowledge about this race. First, you know the *actual* probability p_i that horse i will win. Second, you know s_i, the amount that the public will end up placing on horse i.

(a) With your knowledge of the probabilities $\boldsymbol{p}$ and public money $\boldsymbol{s}_i$, write down a convex optimization program answer will tell you how much to bet on each horse to maximize your expected return. (In the end, you should be maximizing a concave function over a concave set.)

(b) Using Fenchel duality, show how this expected payout can be computed by solving an optimization program in one variable. (Hint: look at the resource allocation example in the notes.) All of the relevant functions are given to you here, so you can (and should) compute their conjugates explicitly.

(c) Show how the primal optimal solution (the best x_i) can be recovered from the (single variable) dual solution.

(d) Here are the track odds right before closing[1]:

1. Effinex	15-1
2. Frosted	5-1
3. Keen Ice	15-1
4. California Chrome	1-1
5. Win the Space	30-1
6. Melatonin	12-1
7. War Story	30-1
8. Shaman Ghost	20-1
9. Hoppertunity	15-1
10. Arrogate	5-2

Saying horse i has track odds A-B simply means that[2]

$$\frac{\sum_{i=1}^{N} s_i}{s_i} = \frac{A}{B}.$$

You know that the true probabilities are

1. Effinex	0.05
2. Frosted	0.15
3. Keen Ice	0.05
4. California Chrome	0.15
5. Win the Space	0.05
6. Melatonin	0.15
7. War Story	0.05
8. Shaman Ghost	0.05
9. Hoppertunity	0.05
10. Arrogate	0.25

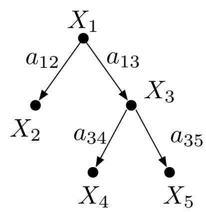
[1] This is from the 2016 Breeders Cup

[2] Another way to interpret this is that the public (or the market) is indicating that they think that the probability that horse i wins is $B/(B + A)$.

The public has wagered a total of \$1 million on this race. You have \$500,000. The vig is 10%. How much do you bet on each horse?

(e) Arrogate won the race[3]. How did you make out?

7. Consider the following graph:



The nodes correspond to Gaussian random variables and the edges between the nodes describe their relationship. Specifically, from a set of iid Gaussian random variables $E_n \sim \text{Normal}(0, 1)$, we generate

$$\begin{aligned} X_1 &= E_1 \\ X_2 &= a_{12}X_1 + E_2 \\ X_3 &= a_{13}X_1 + E_3 \\ X_4 &= a_{34}X_3 + E_4 \\ X_5 &= a_{35}X_3 + E_5. \end{aligned}$$

Let $\boldsymbol{X} = \begin{bmatrix} X_1 & X_2 & X_3 & X_4 & X_5 \end{bmatrix}^{\mathrm{T}}$ be the random vector consisting of the collection of these random variables. We can generate K independent realizations of $\boldsymbol{X}$ in MATLAB using

```
X = zeros(5,K);
for kk = 1:K
    E = randn(5,1);
    X(1,kk) = E(1);
    X(2,kk) = a12*X(1,kk) + E(2);
    X(3,kk) = a13*X(1,kk) + E(3);
    X(4,kk) = a34*X(3,kk) + E(4);
    X(5,kk) = a35*X(3,kk) + E(5);
end
```

None of these random variables is independent of the others. That is, the covariance matrix $\boldsymbol{R} = \mathrm{E}[\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}]$ is non-zero everywhere (assuming, of course, that the specified $a_{i,j} \neq 0$). But it should be clear that pairs of nodes *without* edges between them are conditionally independent (given the values at the other nodes). So, for example:

$$X_2 | (X_1, X_4, X_5) \text{ is independent of } X_3 | (X_1, X_4, X_5)$$

(And in fact in this case, X_2 and X_3 are independent given just the values of X_1.) It is a fact that if two entries X_i, X_j in a Gaussian random vector are conditionally

[3] That actually happened.

independent, then the corresponding entries in the *inverse* covariance matrix are zero; $S_{i,j} = S_{j,i} = 0$, where $\boldsymbol{S} = \boldsymbol{R}^{-1}$.

For this problem, we will fix the values $a_{12} = 2,\ a_{13} = -1,\ a_{34} = 1/2,\ a_{35} = -0.1$.

(a) Compute by hand the covariance matrix $\boldsymbol{R}$. Verify (using MATLAB) that its inverse is zero in the appropriate places.

(b) In MATLAB, generate $K = 1000$ realizations of $\boldsymbol{X}$. Compute the standard sample covariance

$$\hat{\boldsymbol{R}} = \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}.$$

Describe how close it is to the true covariance matrix both quantitatively (using a reasonable metric of your choosing) and qualitatively.

(c) Now suppose that we know the *structure* of the graph above, but not the values of $a_{i,j}$. Recall (from our first set of notes) that the inverse of the sample covariance is the solution to the convex program

$$\min_{\boldsymbol{X}\in S^5_+} \quad -\log\det \boldsymbol{X} + \operatorname{trace}(\boldsymbol{X}\hat{\boldsymbol{R}}).$$

We can exploit our knowledge of the graph structure above to help us get a better estimate of the covariance matrix from these same $K = 1000$ samples. Let $\mathcal{I} = \{(i,j) :$ there is no edge between nodes i and $j\}$. Now estimate the convariance matrix by using CVX to solve

$$\min_{\boldsymbol{X}} \quad -\log\det \boldsymbol{X} + \operatorname{trace}(\boldsymbol{X}\hat{\boldsymbol{R}}), \quad \text{subject to} \quad \boldsymbol{X} \in S^5_+, \quad X_{i,j} = 0,\ (i,j) \in \mathcal{I}.$$

and taking the inverse of your answer. Describe how close it is to the true covariance matrix both quantitatively and qualitatively.

(CVX handles the semidefinite constraint very naturally; see the section on Set Membership in http://web.cvxr.com/cvx/doc/. Also, the log det function is built into CVX: http://web.cvxr.com/cvx/doc/funcref.html#built-in-functions.)

(d) *Optional*: Show that if X_i, X_j are conditionally independent, then $\boldsymbol{R}^{-1}_{i,j} = 0$. You can do this by combining the expression for the conditional covariance matrix for a Gaussian random vector and the Schur complement for $\boldsymbol{R}$ (see Appendix C.4 of Boyd).

8. *(Optional)* Read page 83 of Boyd and Vandenberghe, and then read appendix A.3.3. Show that if $f(\boldsymbol{x})$ is a closed, convex function[4] then the conjugate of the conjugate of f is equal to f:

$$f^{**}(\boldsymbol{x}) = f(\boldsymbol{x}).$$

[4] A function is closed if its epigraph is a closed set — see the aforementioned appendix.