**ECE 8823 (Convex Optimization), Spring 2017**

**Homework #3**

**Due Wednesday February 15, in class**

**Reading: B& V, Chapters 3 and 9.**

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 $S^N$ denote the space of $N \times N$ real-valued symmetric matrices with inner product $\langle \boldsymbol{X}, \boldsymbol{Y} \rangle = \text{trace}(\boldsymbol{XY})$. Of course this inner product obeys the standard Cauchy-Schwarz inequality:
$$\langle \boldsymbol{X}, \boldsymbol{Y} \rangle \;\leq\; \|\boldsymbol{X}\|_F \, \|\boldsymbol{Y}\|_F, \tag{1}$$
with equality if and only if $\boldsymbol{Y} = \alpha \boldsymbol{X}$. If $\boldsymbol{\lambda} : S^N \rightarrow \mathbb{R}^N$ is the map from a symmetric matrix to its eigenvalues sorted in descending order,
$$\boldsymbol{\lambda}(\boldsymbol{X}) = \begin{bmatrix} \lambda_1(\boldsymbol{X}) \\ \lambda_2(\boldsymbol{X}) \\ \vdots \\ \lambda_N(\boldsymbol{X}) \end{bmatrix}, \quad \lambda_1(\boldsymbol{X}) \geq \lambda_2(\boldsymbol{X}) \geq \cdots \geq \lambda_(\boldsymbol{X}).$$
then we can re-write (1) as
$$\langle \boldsymbol{X}, \boldsymbol{Y} \rangle \;\leq\; \left( \sum_{n=1}^{N} \lambda_n^2(\boldsymbol{X}) \right)^{1/2} \left( \sum_{n=1}^{N} \lambda_n^2(\boldsymbol{Y}) \right)^{1/2},$$
since the Frobenius norm of a symmetric matrix is the same as the $\ell_2$ norm of its eigenvalues.

   There is a more refined inequality for symmetric matrices, called the **Fan inequality**, which says
$$\langle \boldsymbol{X}, \boldsymbol{Y} \rangle \;\leq\; \langle \boldsymbol{\lambda}(\boldsymbol{X}), \boldsymbol{\lambda}(\boldsymbol{Y}) \rangle,$$
with equality if and only if $\boldsymbol{X}$ and $\boldsymbol{Y}$ are diagonalized with the same ordering by the same orthonormal eigenvectors $\{\boldsymbol{v}_n\}$,
$$\boldsymbol{X} = \sum_{n=1}^{N} \lambda_n(\boldsymbol{X}) \boldsymbol{v}_n \boldsymbol{v}_n^{\mathrm{T}}, \quad \boldsymbol{Y} = \sum_{n=1}^{N} \lambda_n(\boldsymbol{Y}) \boldsymbol{v}_n \boldsymbol{v}_n^{\mathrm{T}}.$$
   Note that this requires not only that $\boldsymbol{X}$ and $\boldsymbol{Y}$ have the same eigenvectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N$, but also that $\boldsymbol{v}_n$ corresponds to the $n$th largest eigenvalue in both cases. Obviously, I am telling you all of this as it will be useful for solving the following:

(a) Let $\mathcal{C}$ be the set of symmetric matrices with operator norm less than 1:

$$\mathcal{C} = \{\boldsymbol{X} \in S^N \;:\; \|\boldsymbol{X}\| \leq 1\}, \quad \|\boldsymbol{X}\| = \max_{\|\boldsymbol{v}\|_2=1} \|\boldsymbol{X}\boldsymbol{v}\|_2.$$

Calculate the support functional

$$h(\boldsymbol{A}) = \sup_{\boldsymbol{X}\in\mathcal{C}} \langle \boldsymbol{X}, \boldsymbol{A}\rangle.$$

(b) Given a $\boldsymbol{X}_0 \in S^N$, describe how to solve the following closest point problem:

$$\underset{\boldsymbol{X}\in S^N}{\text{minimize}} \;\|\boldsymbol{X} - \boldsymbol{X}_0\|_F \quad \text{subject to} \quad \|\boldsymbol{X}\| \leq 1.$$

You might want to start by arguing that for the dual

$$\underset{\|\boldsymbol{A}\|_F\leq 1}{\text{maximize}} \;\langle \boldsymbol{X}_0, \boldsymbol{A}\rangle - h(\boldsymbol{A}),$$

you only need to take the maxima over symmetric $\boldsymbol{A}$.

(c) Define $\|\boldsymbol{X}\|_* = \sum_{n=1}^N |\lambda_n(\boldsymbol{X})|$. Given a $\boldsymbol{X}_0 \in S^N$, describe how to solve the following closest point problem:

$$\underset{\boldsymbol{X}\in S^N}{\text{minimize}} \;\|\boldsymbol{X} - \boldsymbol{X}_0\|_F \quad \text{subject to} \quad \|\boldsymbol{X}\|_* \leq 1.$$

(You will want to revisit the supplementary notes about projecting onto the $\ell_1$ ball.)

(d) Use the Fan inequality to show that the solution to the nonconvex problem

$$\underset{\boldsymbol{X}\in S^N}{\text{minimize}} \;\|\boldsymbol{X} - \boldsymbol{X}_0\|_F \quad \text{subject to} \quad \text{rank}(\boldsymbol{X}) = r,$$

is given by the truncated eigenvalue expansion for $\boldsymbol{X}_0$.

(e) Discuss the similarities between the solutions in part (c) and part (d). What are the errors in both cases?

(f) *Optional (and not easy):* Prove the Fan inequality.

3. A *sublevel set* for a functional $f : \mathbb{R}^N \to \mathbb{R}$ is

$$\mathcal{S}(f, \beta) = \{\boldsymbol{x} \in \mathbb{R}^N \;:\; f(\boldsymbol{x}) \leq \beta\}.$$

(a) Show that if $f$ is a convex function, then all of its sublevel sets are convex.

(b) Show that the converse is not true by providing a counter-example (a function which has convex sublevel sets but is not convex).

4. Suppose I tell you that $\boldsymbol{X}$ is a random vector with a covariance matrix of the form

$$\mathrm{E}[\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}] = \begin{bmatrix} ? & -1 & 5 & -4 & -4 \\ -1 & ? & 4 & 5 & 0 \\ 5 & 4 & ? & -1 & -2 \\ -4 & 5 & -1 & ? & 1 \\ -4 & 0 & -2 & 1 & ? \end{bmatrix}$$

Find the diagonal for this matrix that minimizes the maximum of the variances of the $X_i$.

5. (B& V 9.1) Consider the problem of minimizing a quadratic function

$$\min_{\boldsymbol{x}} \quad f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{x} + \boldsymbol{q}^{\mathrm{T}}\boldsymbol{x} + r,$$

where $\boldsymbol{P} \in S^N$ (i.e. $\boldsymbol{P}$ is symmetric but not necessarily postive semidefinite).

(a) Show that if $\boldsymbol{P} \not\succeq \mathbf{0}$ (which means $f$ is not convex), then the problem is unbounded below (and hence has no minimizer).

(b) Now suppose that $\boldsymbol{P} \succeq \mathbf{0}$ (which means $f$ is convex). Show that if there is no $\boldsymbol{x}_0$ such that $\boldsymbol{P}\boldsymbol{x}_0 = -\boldsymbol{q}$, then $f$ is unbounded below.

6. (B& V 9.5) Recall that when using backtracking to select a step size to move from $\boldsymbol{x}_0$ in direction $\boldsymbol{d}$, we start with $t = 1$, and then iterative decrease $t$ by factor of $\beta < 1$ until

$$f(\boldsymbol{x}_0 + t\boldsymbol{d}) - f(\boldsymbol{x}_0) \;\leq\; \alpha t\langle \boldsymbol{d}, \nabla f(\boldsymbol{x}_0)\rangle, \tag{2}$$

where $0 < \alpha < 1/2$ is some user-defined parameter. Show that if $f$ is strongly convex,

$$m\mathbf{I} \preceq \nabla^2 f(\boldsymbol{x}) \preceq M\mathbf{I}, \quad \text{for all} \;\; \boldsymbol{x}\mathbb{R}^N,$$

and $\boldsymbol{d}$ is a descent direction, then the stopping condition holds when

$$t \;\leq\; -\frac{\langle \boldsymbol{d}, \nabla f(\boldsymbol{x}_0)\rangle}{M\|\boldsymbol{d}\|_2^2}.$$

Use this to derive an upper bound on the number of backtracking iterations.

7. We have presented gradient descent as a basic method for solving smooth unconstrained problems. In this problem we will explore its use in solving nonsmooth constrained problems, specifically linear programs.

(a) Implement gradient descent: write a MATLAB function

```
function xstar = gd(f, gradf, x0, tol)
```

that takes a function handle for evaluating $f : \mathbb{R}^N \to \mathbb{R}$, a function handle for evaluating $\nabla f : \mathbb{R}^N \to \mathbb{R}^N$, a starting point $\boldsymbol{x}_0 \in \mathbb{R}^N$, and a tolerance so that you terminate when $\|\nabla f(\boldsymbol{x}^{(k)})\|_2^2 \leq \text{tol}$. For now, you can hard-code the backtracking parameters as $\alpha = 0.001$ and $\beta = 0.8$. You might also want to cap the total number of iteration at something reasonable (say $10,000$).

(b) Consider the linear program in $\mathbb{R}^2$,

$$\min_{\boldsymbol{x}} \; \langle \boldsymbol{x}, \boldsymbol{c}\rangle \quad \text{subject to} \quad \langle \boldsymbol{x}, \boldsymbol{a}_m\rangle \leq b_m$$

where

$$\boldsymbol{c} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{a}_m = \begin{bmatrix} \cos(m\pi/3) \\ \sin(m\pi/3) \end{bmatrix}, \quad b_m = 1, \quad m = 1, 2, \ldots, 6.$$

Sketch the feasible region in the plane — the region where all six linear constraints hold. Using MATLAB, create and image of $\langle \boldsymbol{x}, \boldsymbol{c}\rangle$ over the feasible region. Where is the minimizer $\boldsymbol{x}^\star$?

(c) Now consider the smooth, unconstrained problem

$$\min_{\boldsymbol{x}} \langle \boldsymbol{x}, \boldsymbol{c} \rangle - \frac{1}{\tau} \sum_{m=1}^{6} \log(b_m - \langle \boldsymbol{x}, \boldsymbol{a}_m \rangle).$$

Using MATLAB, create an image of the functional over the feasible region for $\tau = 1$. What is happening near the boundary?

(d) Solve the program above using gradient descent for $\tau = 1, 10, 100, 500, 1000$ with `tol=1e-4` and starting at the origin, $\boldsymbol{x}_0 = \mathbf{0}$. Make a figure of the feasible region, then put an 'x' where your solution landed for these different $\tau$. Note how many iterations it took for each $\tau$.

(e) Do the same, but use the solution for $\tau = 1$ as the starting point for $\tau = 10$ then use that solution for $\tau = 100$, etc. How many total iterations does it take to get the answer for $\tau = 1000$?