ECE 8823 (Convex Optimization), Spring 2017

Homework #1

Due Monday January 23, in class

Reading: B&V, Chapter 1. You might also want to skim Appendix A.

Please sign up for Piazza at https://piazza.com/gatech/spring2017/ece8823b.

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. In machine learning, it is common to encounter optimization programs of the form

$$\underset{\boldsymbol{x}\in\mathbb{R}^N, z\in\mathbb{R}}{\text{minimize}} \ \sum_{m=1}^{M} \max(\boldsymbol{a}_m^{\mathrm{T}}\boldsymbol{x}, z) + \tau\|\boldsymbol{x}\|_2^2.$$

Show how to express this program as a quadratic program with linear constraints. (Hint: $\max(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}, z) \leq u$ is the same as saying $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x} \leq u$ and $z \leq u$.)

3. A functional $f(\boldsymbol{x}) : \mathbb{R}^N \to \mathbb{R}$ is *concave* if for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$,

$$f(\theta\boldsymbol{x} + (1-\theta)\boldsymbol{y}) \ \geq \ \theta f(\boldsymbol{x}) + (1-\theta)f(\boldsymbol{y}), \quad \text{for all} \ \ 0 \leq \theta \leq 1.$$

Give a simple yet rigorous argument that

$$f(\boldsymbol{x}) \text{ is concave} \quad \Leftrightarrow \quad -f(\boldsymbol{x}) \text{ is convex.}$$

4. We use S_+^N to denote the set of $N \times N$ matrices that are symmetric and whose eigenvalues are non-negative.

   (a) Let $\lambda_{\min}(\boldsymbol{X})$ be a function that takes a symmetric matrix and returns the smallest eigenvalue (possibly negative) of $\boldsymbol{X}$. Show that $\lambda_{\min}(\boldsymbol{X})$ is concave.

   (b) Use your result from the previous section to show that S_+^N is a convex set.

   (c) Find a set of convex functions $f_1(\boldsymbol{X}), \ldots, f_M(\boldsymbol{X})$ that map *arbitrary* $N \times N$ matrices to scalars ($f_m(\boldsymbol{X}) : \mathbb{R}^{N\times N} \to \mathbb{R}$) and scalars $b_1, \ldots, b_M$ that specify S_+^N, meaning

$$\boldsymbol{X} \in S_+^N \quad \Leftrightarrow \quad f_m(\boldsymbol{X}) \leq b_m, \text{ for all } m = 1, \ldots, M.$$

   (Note that if f_m is linear, then f_m is both convex and concave, and so $f_m(\boldsymbol{X}) = b_m$ can be implemented using the pair of inequalities $f_m(\boldsymbol{X}) \leq b_m$ and $-f_m(\boldsymbol{X}) \leq -b_m$.)

Last updated 11:40, January 15, 2017

5. If $f(\boldsymbol{x}) : \mathbb{R}^N \to \mathbb{R}$ is linear, then the set $\{\boldsymbol{x} : f(\boldsymbol{x}) = c\}$, where $c \in \mathbb{R}$ is a hyperplane and is clearly convex. Can the set $\mathcal{C} = \{\boldsymbol{x} : f(\boldsymbol{x}) = c\}$ be convex if f is nonlinear? Prove it.

6. Recall that a norm is a functional $\|\cdot\| : \mathbb{R}^N \to \mathbb{R}$ which obeys

   - $\|\boldsymbol{x}\| \geq 0$ and $\|\boldsymbol{x}\| = 0$ if and only if $\boldsymbol{x} = \boldsymbol{0}$.
   - $\|a\boldsymbol{x}\| = |a|\,\|\boldsymbol{x}\|$ for all $\boldsymbol{x} \in \mathbb{R}^N$ and scalars $a \in \mathbb{R}$.
   - $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$.

   It should be clear for any valid norm, the set $\{\boldsymbol{x} \ : \ \|\boldsymbol{x}\| \leq r\}$ is convex for all r. Here you will show that something close to the converse is true. Let $\mathcal{C}$ be a closed convex set that is symmetric around the origin,

   $$\boldsymbol{x} \in \mathcal{C} \quad \Leftrightarrow \quad -\boldsymbol{x} \in \mathcal{C},$$

   has a non-empty interior, and is bounded (meaning that every point in $\mathcal{C}$ is within some Euclidean distance D of the origin, $\|\boldsymbol{x}\|_2 \leq D$ for all $\boldsymbol{x} \in \mathcal{C}$). Define the functional $g(\boldsymbol{x})$ as

   $$g(\boldsymbol{x}) = \inf\left\{t \ : \ t^{-1}\boldsymbol{x} \in \mathcal{C}, \ t > 0\right\}.$$

   (a) Show that $g(\boldsymbol{x})$ is a valid norm on $\mathbb{R}^N$.
   (b) Describe the unit ball of this norm, $\{\boldsymbol{x} \in \mathbb{R}^N \ : \ g(\boldsymbol{x}) \leq 1\}$.
   (c) *Optional:* Explain why this definition does not make any sense if $\mathcal{C}$ has an empty interior.
   (d) *Optional:* Suppose that $\mathcal{C}$ is not bounded. Why doesn't $g(\boldsymbol{x})$ define a norm?

7. Download and install the CVX package, http://cvxr.com/cvx/. Look through the CVX users guide. There is also a intro video on the website you can watch if you want.

   (a) Implement the Chebyshev approximation problem:

   $$\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_\infty$$

   You do not have to convert this to a linear program first; CVX know what to do with the ℓ_∞ norm. Run some experiments for randomly chosen $M \times N$ $\boldsymbol{A}$ and $\boldsymbol{y}$ (the problem is not really that interesting unless $M > N$). Let $\hat{\boldsymbol{r}} = \boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}}$ be the residual for the solution $\hat{\boldsymbol{x}}$. As you vary M and N (keeping $M > N$), how many terms in $\hat{\boldsymbol{r}}$ have maximum amplitude, $|r_m| = \|\hat{\boldsymbol{r}}\|_\infty$? (Start to think about how you might prove this; I might ask you to do it later.) Turn in the code and a plot of $\hat{\boldsymbol{r}}$ for an instance of your choosing.

   (b) Run some similar experiments for the ℓ_1 approximation problem:

   $$\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_1$$

   How does the number of non-zero entries in the solution residual $\hat{\boldsymbol{r}}$ vary as you change M and N (again keeping $M > N$)? (Start to think about how you might prove this; I might ask you to do it later.) Turn in the code and a plot of $\hat{\boldsymbol{r}}$ for an instance of your choosing.

(c) Run some similar experiments for the ℓ_∞ minimization program:

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_\infty \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},$$

where $\boldsymbol{A}$ is now underdetermined ($M < N$). As you vary M and N, how do the number of terms in $\hat{\boldsymbol{x}}$ that have maximum amplitude $\hat{x}_n = \|\hat{\boldsymbol{x}}\|_\infty$ change? (Start to think about how you might prove this; I might ask you to do it later.) Turn in the code and a plot of $\hat{\boldsymbol{x}}$ for an instance of your choosing.

(d) Now consider the ℓ_1 minimization program:

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},$$

where again $M < N$. How do the number of non-zero terms in the solution $\hat{\boldsymbol{x}}$ vary with M and N? (Start to think about how you might prove this; I might ask you to do it later.) Turn in the code and a plot of $\hat{\boldsymbol{x}}$ for an instance of your choosing.

Last updated 11:40, January 15, 2017