ECE 8823 (Convex Optimization), Spring 2017

Homework #6

Due Wednesday April 5, in class

Reading:
B&V Chapters 4, 6 and 7;
Boyd et al, "Distributed Optimization and Stat. Learning using ADMM," 2011

Along with 1, the assignment consists of 5 questions chosen in the following manner:
Problems 2, 3, and 6 are required.
Choose one of Problems 4 and 5.
Choose one of Problems 7 and 8.

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. Consider the general quadratic programming problem with linear constraints:

$$\min_{\boldsymbol{x}} \ \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{x} + \langle \boldsymbol{x}, \boldsymbol{q} \rangle \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}.$$

   We will assume that $\boldsymbol{P}$ is symmetric positive definite (and so has full rank).

   (a) Find the dual.

   (b) Show how the primal optimal solution $\boldsymbol{x}^\star$ can be computed from the dual optimal solution $\boldsymbol{\lambda}^\star$. (You should be able to convince yourself that both programs have unique solutions.)

   (c) Describe how we can use ADMM to solve the dual problem (in a non-distributed manner, to start). Why is ADMM easier on the dual than the primal?

   (d) Describe how we can write the dual as

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} g(\boldsymbol{\alpha}) + h(\boldsymbol{\beta}),$$

   where $g(\cdot)$ is **separable** (a sum of individual components of $\boldsymbol{\alpha}$) and $h(\cdot)$ is an indicator of a convex set which we can easily project onto. Write down explicitly how this projection operator works.

   (e) Using your results from part (d), describe how we can use distributed ADMM to solve the dual program.

3. B&V 4.8

4. B&V 4.12

5. B&V 4.14

6. B&V 6.3 (see note below)

7. B&V 7.1

8. B&V 7.2

A start on B&V 6.3(b)

Consider the problem where we have only two measurements/examples/rows-of-$\boldsymbol{A}$. We assume $a = 1$ just to keep things clean. We want to solve

$$\min_{\boldsymbol{x}} \; -\log\left(1 - (\langle \boldsymbol{x}, \boldsymbol{a}_1\rangle - b_1)^2\right) - \log\left(1 - (\langle \boldsymbol{x}, \boldsymbol{a}_2\rangle - b_2)^2\right) \quad \text{subject to} \quad \begin{aligned} \langle \boldsymbol{x}, \boldsymbol{a}_i\rangle - b_i - 1 &\leq 0 \\ -\langle \boldsymbol{x}, \boldsymbol{a}_i\rangle + b_i - 1 &\leq 0 \quad i = 1, 2 \end{aligned}$$

This is equivalent to

$$\max_{\boldsymbol{x},\boldsymbol{y}} \left(1 - y_1^2\right)\left(1 - y_2^2\right) \quad \text{subject to} \quad \begin{aligned} \boldsymbol{y} &= \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} \\ \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} - \mathbf{1} &\leq \mathbf{0} \\ -\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} - \mathbf{1} &\leq \mathbf{0} \end{aligned}$$

which is in turn equivalent to

$$\max_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{t}} \; t_1^2 t_2^2 \quad \text{subject to} \quad \begin{aligned} (1 - y_1^2) &\geq t_1^2 \\ (1 - y_2^2) &\geq t_2^2 \\ \boldsymbol{y} &= \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} \\ \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} - \mathbf{1} &\leq \mathbf{0} \\ -\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} - \mathbf{1} &\leq \mathbf{0} \end{aligned}$$

which finally is equivalent to

$$\max_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{t}} \; t_0 \quad \text{subject to} \quad \begin{aligned} t_1 t_2 &\geq t_0^2 \\ y_1^2 + t_1^2 &\leq 1 \\ y_2^2 + t_2^2 &\leq 1 \\ \boldsymbol{y} &= \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} \\ \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} - \mathbf{1} &\leq \mathbf{0} \\ -\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} - \mathbf{1} &\leq \mathbf{0} \end{aligned}$$

It is a fact that

$$t_1 t_2 \geq t_0^2 \quad \Leftrightarrow \quad \left\| \begin{bmatrix} 2t_0 \\ t_1 - t_2 \end{bmatrix} \right\|_2 \leq t_1 + t_2$$

(you must prove this fact yourself before you are allowed to use it). Thus the program is equivalent to the SOCP:

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{t}} \quad -t_0 \quad \text{subject to} \quad \left\| \begin{bmatrix} 2t_0 \\ t_1 - t_2 \end{bmatrix} \right\|_2 &\leq t_1 + t_2 \\ y_1^2 + t_1^2 &\leq 1 \\ y_2^2 + t_2^2 &\leq 1 \\ \boldsymbol{y} &= \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} \\ \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} - \mathbf{1} &\leq \mathbf{0} \\ -\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} - \mathbf{1} &\leq \mathbf{0} \end{aligned}$$