L. El Ghaoui, S. Sojoudi

Homework Assignment #2

Due date: 9/26/19 before class on Gradescope. Please LaTeX your homework solution and submit an electronic version.

Exercise 1 (Boolean least-squares) Consider the following problem, known as *Boolean Least Squares*:

$$\phi = \min_x \|Ax - b\|_2^2 \ : \ x_i \in \{-1, 1\}, \quad i = 1, \ldots, n.$$

Here, the variable is $x \in \mathbb{R}^n$, where $A \in \mathbb{R}^{m,n}$ and $b \in \mathbb{R}^m$ are given. This is a basic problem arising, for instance, in digital communications. A brute force solution is to check all 2^n possible values of x, which is usually impractical.

1. Show that the problem is equivalent to

$$\begin{aligned} \phi = \min_{X,x} \quad & \operatorname{trace}(A^\top AX) - 2b^\top Ax + b^\top b \\ \text{s.t.} \quad & X = xx^\top, \\ & X_{ii} = 1, \quad i = 1, \ldots, n, \end{aligned}$$

in the variables $X = X^\top \in \mathbb{R}^{n,n}$ and $x \in \mathbb{R}^n$.

- The constraint $X = xx^\top$, i.e., the set of rank-1 matrices is not convex, therefore the problem is still hard. However, an efficient approximation can be obtained by relaxing this constraint to $X \succeq xx^\top$, obtaining

$$\begin{aligned} \phi \geq \phi_{\text{sdp}} = \min_X \quad & \operatorname{trace}(A^\top AX) - 2b^\top Ax + b^\top b \\ \text{s.t.} \quad & \begin{bmatrix} X & x \\ x^\top & 1 \end{bmatrix} \succeq 0, \\ & X_{ii} = 1, \quad i = 1, \ldots, n. \end{aligned}$$

  The relaxation produces a lower-bound to the original problem. Once that is done, an approximate solution to the original problem can be obtained by rounding the solution: $x_{\text{sdp}} = \operatorname{sgn}(x^*)$, where x^* is the optimal solution of the semidefinite relaxation.

- Another approximation method is to relax the non-convex constraints $x_i \in \{-1, 1\}$ to convex interval constraints $-1 \leq x_i \leq 1$ for all i, which can be written $\|x\|_\infty \leq 1$. Therefore a different lower bound is given by:

$$\phi \geq \phi_{\text{int}} \doteq \min \|Ax - b\|_2^2 \ : \ \|x\|_\infty \leq 1.$$

  Once that problem is solved, we can round the solution by $x_{\text{int}} = \operatorname{sgn}(x^*)$ and compare the original objective value $\|Ax_{\text{int}} - b\|_2^2$.

2. Which one of ϕ_{sdp} and ϕ_{int} produces the closest approximation to ϕ? Justify carefully your answer.

3. Use now 100 independent realizations with normally distributed data, $A \in \mathbb{R}^{10,10}$ (independent entries with mean zero) and $b \in \mathbb{R}^{10}$ (independent entries with mean 1). Plot and compare the histograms of $\|Ax_{\text{sdp}} - b\|_2^2$ of part 1, $\|Ax_{\text{int}} - b\|_2^2$ of part 1, and the objective corresponding to a naïve method $\|Ax_{\text{ls}} - b\|_2^2$, where $x_{\text{ls}} = \text{sgn}\left((A^\top A)^{-1} A^\top b\right)$ is the rounded ordinary Least Squares solution. Briefly discuss accuracy and computation time (in seconds) of the three methods.

4. Assume that, for some problem instance, the optimal solution (x, X) found via the SDP approximation is such that x belongs to the original non-convex constraint set $\{x : x_i \in \{-1, 1\}, \ \ i = 1, \ldots, n\}$. What can you say about the SDP approximation in that case?

Exercise 2 (Regularization for noisy data) Consider a least-squares problem

$$\min_x \ \|Ax - y\|_2^2,$$

in which the data matrix $A \in \mathbb{R}^{m,n}$ is noisy. Our specific noise model assumes that each row $a_i^\top \in \mathbb{R}^n$ has the form $a_i = \hat{a}_i + u_i$, where the noise vector $u_i \in \mathbb{R}^n$ has zero mean and covariance matrix $\sigma^2 I_n$, with σ a measure of the size of the noise. Therefore, now the matrix A is a function of the uncertain vector $u = (u_1, \ldots, u_n)$, which we denote by $A(u)$. We will write $\hat{A}$ to denote the matrix with rows $\hat{a}_i^\top$, $i = 1, \ldots, m$. We replace the original problem with

$$\min_x \ \mathbb{E}_u\{\|A(u)x - y\|_2^2\},$$

where $\mathbb{E}_u$ denotes the expected value with respect to the random variable u. Show that this problem can be written as

$$\min_x \ \|\hat{A}x - y\|_2^2 + \lambda \|x\|_2^2,$$

where $\lambda \geq 0$ is some regularization parameter, which you will determine. That is, regularized least-squares can be interpreted as a way to take into account uncertainties in the matrix A, in the expected value sense. *Hint:* compute the expected value of $((\hat{a}_i + u_i)^\top x - y_i)^2$, for a specific row index i.

Exercise 3 (Robust QP) Consider the QP

$$\min_x \ \frac{1}{2} x^\top P x + q^\top x + r \ : \ Ax \leq b$$

For simplicity we assume that only teh matrix P is subject to errors and the other parameters are exactly known. The *robust* QP writes

$$\min_x \max_{P \in \mathcal{E}} \ \frac{1}{2} x^\top P x + q^\top x + r \ : \ Ax \leq b$$

where $\mathcal{E}$ is the set of possible matrices P. Below, we consider different sets $\mathcal{E}$. For each $\mathcal{E}$, express the robust QP as a convex problem. If the problem can be expressed in standard form (QP, QCQP, SOCP, SDP), say so.

1. $\mathcal{E} = \{P_1, \ldots, P_k\}$ – a finite set of matrices where each matrix is PSD.

2. A set specified by a nominal value $P_0 \in \mathbb{S}^n_+$ plus a bound on the eigenvalues of the deviation $P - P_0$

$$\mathcal{E} = \{P \in \mathbb{S}^n_+ \mid -\gamma I \preceq P - P_0 \preceq \gamma I\}$$

   where $\gamma \in \mathbb{R}$.

3. An ellipsoid of matrices

$$\mathcal{E} = \left\{P_0 + \sum_{i=1}^{K} P_i u_i \mid \|u\|_2 \leq 1\right\}$$

   where each P_i is PSD.

Exercise 4 (Univariate LASSO) Consider the problem

$$\min_{x \in \mathbb{R}} f(x) \doteq \frac{1}{2}\|ax - y\|_2^2 + \lambda |x|,$$

where $\lambda \geq 0$, $a \in \mathbb{R}^m$, $y \in \mathbb{R}^m$ are given, and $x \in \mathbb{R}$ is a scalar variable.

Assume that $y \neq 0$ and $a \neq 0$, (since otherwise the optimal solution of this problem is simply $x = 0$). Prove that the optimal solution of this problem is

$$x^* = \begin{cases} 0 & \text{if } |a^\top y| \leq \lambda \\ x_{\mathrm{ls}} - \mathrm{sgn}(x_{\mathrm{ls}}) \frac{\lambda}{\|a\|_2^2} & \text{if } |a^\top y| > \lambda, \end{cases}$$

where

$$x_{\mathrm{ls}} \doteq \frac{a^\top y}{\|a\|_2^2}$$

corresponds to the solution of the problem for $\lambda = 0$.