15.094, Spring 2014

Robust Modeling, Optimization and Computation

Midterm

April 2, 2014

- This is a 3 hour exam. The exam is designed so you do not have time pressure. We expect that most of you will finish in 2 hours.
- Please submit all your answers.
- You can use all the material in stellar (lecture notes, readings, homework solutions, recitations, helper material, etc.) your own notes, and your own homework solutions in either oral or in electronic form.
- You cannot communicate with others in either oral, written or electronic form.
- There are 3 problems with a total of 110 points. So, 10 points are extra credit.
- Please answer the questions robustly. Good luck!

Notation. For a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, we denote by $\text{trace}(\boldsymbol{A}) := \sum_{i=1}^{n} [\boldsymbol{A}]_{ii}$ the trace of $\boldsymbol{A}$ i.e., the sum of all diagonal elements of $\boldsymbol{A}$. For a set $\mathcal{U} \subseteq \mathbb{R}^k$, we denote by $\text{cone}(\mathcal{U})$ the conic hull of $\mathcal{U}$ i.e., the set of all conic combinations of points in $\mathcal{U}$. Thus,

$$\text{cone}(\mathcal{U}) := \left\{ \sum_{i=1}^{k} \lambda_i \boldsymbol{\xi}_i \;:\; \boldsymbol{\xi}_i \in \mathcal{U},\ \lambda_i \geq 0,\ i \in \{1, ..., k\},\ k \in \mathbb{N} \right\}. \tag{1}$$

Problem 1 (True/False)**.** (40 points)

For each of the following statements, indicate if the statement is True or False. In each case, provide a brief justification, sketch of a proof, or counterexample.

1. Consider the robust optimization problem

$$\begin{aligned} Z_1 \quad = \quad \min_{\boldsymbol{x} \in \mathcal{X}} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}(\boldsymbol{\xi})\boldsymbol{x} \leq \boldsymbol{b}(\boldsymbol{\xi}) \quad \forall \boldsymbol{\xi} \in \Xi, \end{aligned} \tag{2}$$

where $\boldsymbol{c} \in \mathbb{R}^n$, $\mathcal{X} \subseteq \mathbb{R}^n$, $\Xi \subseteq \mathbb{R}^k$, $\boldsymbol{A} : \mathbb{R}^k \to \mathbb{R}^{m \times n}$ and $\boldsymbol{b} : \mathbb{R}^k \to \mathbb{R}^m$. Note that beyond the apparent linearity assumptions, no convexity assumption is made on $\mathcal{X}$, Ξ, $\boldsymbol{A}(\boldsymbol{\xi})$ or $\boldsymbol{b}(\boldsymbol{\xi})$.

   (a) (5 points) Suppose $\boldsymbol{A}(\boldsymbol{\xi})$ and $\boldsymbol{b}(\boldsymbol{\xi})$ are affine in $\boldsymbol{\xi}$, and $\Xi := \overline{\Xi} \cap \{0,1\}^k$ for some convex set $\overline{\Xi} \subseteq \mathbb{R}^k$. Consider the problem

$$\begin{aligned} Z_2 \quad = \quad \min_{\boldsymbol{x} \in \mathcal{X}} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}(\boldsymbol{\xi})\boldsymbol{x} \leq \boldsymbol{b}(\boldsymbol{\xi}) \quad \forall \boldsymbol{\xi} \in \overline{\Xi} \cap [0,1]^k. \end{aligned}$$

   Then, the ratio $\frac{Z_2}{Z_1}$ can be arbitrarily large.

   (b) (5 points) Suppose $\boldsymbol{A}(\boldsymbol{\xi})$ and $\boldsymbol{b}(\boldsymbol{\xi})$ are linear in $\boldsymbol{\xi}$. Consider the problem

$$\begin{aligned} Z_3 \quad = \quad \min_{\boldsymbol{x} \in \mathcal{X}} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}(\boldsymbol{\xi})\boldsymbol{x} \leq \boldsymbol{b}(\boldsymbol{\xi}) \quad \forall \boldsymbol{\xi} \in \operatorname{cone}(\Xi), \end{aligned}$$

   where cone(Ξ) is defined in (1). Then $Z_3 > Z_1$.

2. (5 points) The following inequality is valid:

$$\min_{\boldsymbol{x} \in \mathcal{X}} \max_{\boldsymbol{\xi} \in \Xi} \boldsymbol{c}(\boldsymbol{\xi})^\top \boldsymbol{x} \ \leq \ \max_{\boldsymbol{\xi} \in \Xi} \min_{\boldsymbol{x} \in \mathcal{X}} \boldsymbol{c}(\boldsymbol{\xi})^\top \boldsymbol{x}.$$

3. (5 points) The following equality is valid:

$$\begin{aligned} \min_{\boldsymbol{x}(\cdot)} \quad & \max_{\boldsymbol{\xi} \in \Xi} \boldsymbol{c}(\boldsymbol{\xi})^\top \boldsymbol{x}(\boldsymbol{\xi}) \quad = \quad \max_{\boldsymbol{\xi} \in \Xi} \min_{\boldsymbol{x} \in \mathcal{X}} \boldsymbol{c}(\boldsymbol{\xi})^\top \boldsymbol{x}. \\ \text{s.t.} \quad & \boldsymbol{x}(\boldsymbol{\xi}) \in \mathcal{X} \ \ \forall \boldsymbol{\xi} \in \Xi \end{aligned}$$

4. (5 points) Consider the two stage robust optimization problem:

$$\begin{array}{cl} \min\limits_{\boldsymbol{x}\in\mathcal{X},\, \boldsymbol{y}(\cdot)} & \boldsymbol{c}^\top \boldsymbol{x} + \max\limits_{\boldsymbol{\xi}\in\Xi} \ \boldsymbol{d}^\top \boldsymbol{y}(\boldsymbol{\xi}) \\ \text{s.t.} & \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{\xi}) \leq \boldsymbol{b}(\boldsymbol{\xi}) \quad \forall \boldsymbol{\xi} \in \Xi, \end{array}$$

where $\mathcal{X} \subseteq \mathbb{R}^{n_x}$, $\boldsymbol{c} \in \mathbb{R}^{n_x}$, $\boldsymbol{d} \in \mathbb{R}^{n_y}$, $\boldsymbol{A} \in \mathbb{R}^{m\times n_x}$, $\boldsymbol{B} \in \mathbb{R}^{m\times n_y}$, $\boldsymbol{y} : \mathbb{R}^k \to \mathbb{R}^{n_y}$ $\Xi := \{\boldsymbol{\xi} \in \mathbb{R}^k \ : \ \mathbf{e}^\top \boldsymbol{\xi} \leq \Gamma,\ \boldsymbol{\xi} \geq \mathbf{0}\}$ and $\boldsymbol{b}$ is affine in $\boldsymbol{\xi}$. There does not exist an optimal $(\boldsymbol{x}, \boldsymbol{y}(\cdot))$ pair such that $\boldsymbol{y}(\cdot)$ is constant, i.e., $\boldsymbol{y}(\boldsymbol{\xi}) = \boldsymbol{y}\ \forall \boldsymbol{\xi}$.

5. (5 points) Let $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ be Pareto Robust Optimal solutions to the problem

$$\min_{\boldsymbol{x}\in\mathcal{X}} \quad \max_{\boldsymbol{\xi}\in\Xi} \ \boldsymbol{c}(\boldsymbol{\xi})^\top \boldsymbol{x}.$$

Suppose $\boldsymbol{c}(\boldsymbol{\xi}_1)^\top \boldsymbol{x}_1 > \boldsymbol{c}(\boldsymbol{\xi}_1)^\top \boldsymbol{x}_2$ for some $\boldsymbol{\xi}_1 \in \Xi$. Then there exists $\boldsymbol{\xi}_2 \in \Xi$ such that $\boldsymbol{c}(\boldsymbol{\xi}_2)^\top \boldsymbol{x}_1 < \boldsymbol{c}(\boldsymbol{\xi}_2)^\top \boldsymbol{x}_2$.

6. (5 points) Consider an adaptive robust optimization problem involving constraints of the form

$$\boldsymbol{a}(\boldsymbol{\xi})^\top \boldsymbol{y}(\boldsymbol{\xi}) \leq b(\boldsymbol{\xi}) \quad \forall \boldsymbol{\xi} \in \Xi, \tag{3}$$

where $\boldsymbol{a}(\boldsymbol{\xi}) \in \mathbb{R}^n$ is affine in $\boldsymbol{\xi}$, $\boldsymbol{y} : \mathbb{R}^k \to \mathbb{R}^n$ and $\Xi \subseteq \mathbb{R}^k$ is given by the intersection of concentric ellipsoids. If one restricts the functional decision variables $\boldsymbol{y}(\boldsymbol{\xi})$ to be affine in $\boldsymbol{\xi}$, then the adaptive problem is equivalent to a semi-definite optimization problem.

7. (5 points) Consider the robust mixed integer optimization problem

$$\begin{array}{cl} \min\limits_{\boldsymbol{x}} & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} \quad \forall \boldsymbol{A} \in \mathcal{U} \\ & \boldsymbol{x} \in \mathbb{R}^{n_r} \times \mathbb{Z}^{n_z}, \end{array} \tag{4}$$

where $\boldsymbol{c} \in \mathbb{R}^{n_r+n_z}$, $\boldsymbol{b} \in \mathbb{R}^m$ and $\mathcal{U} \subseteq \mathbb{R}^{m\times(n_r+n_z)}$. Robustifying the linear relaxation of (4) using the cutting plane algorithm and subsequently using the branch-and-bound algorithm to solve the resulting deterministic mixed-integer problem will always yield a feasible solution to (4).

Problem 2 (Resource Allocation on the Cloud)**.** (35 points)

Consider a cloud computing system over a finite planning horizon $\mathcal{T} := \{1, \dots, T\}$. The system consists of a set $\mathcal{S} := \{1, \dots, S\}$ of servers. Each server $s \in \mathcal{S}$ has I different kinds of resources indexed by $i \in \mathcal{I} := \{1, \dots, I\}$ (e.g., processing power, disk space, memory). We denote by $R_{i,s} \in \mathbb{R}_+$ the units (capacity) of resource $i \in \mathcal{I}$ available to server s during each period. Each of these resources may be expanded at any time $t \in \mathcal{T}$ at cost $c^{\mathrm{e}}_{t,i,s} \in \mathbb{R}_+$ per unit. We assume that expansion decisions take immediate effect.

There are J different types of jobs that can run on the servers. These are indexed by $j \in \mathcal{J} := \{1, \dots, J\}$. At the beginning of each period $t \in \mathcal{T}$, $d_{t,j} \in \mathbb{R}_+$ jobs of type $j \in \mathcal{J}$ arrive at a router which must allocate them among the available servers, effectively adding them to the server queue. Note that *not* all jobs of type j must be allocated to the same server. In fact, fractions of the same job may be processed by different servers. To process one unit of job type j during time-interval $[t, t+1)$, $t \in \mathcal{T}$, server s utilizes $r_{i,j,s}$ units of its resource i. Each server can process the jobs allocated to it in parallel, provided the resources required do not exceed its capacity.

At each time $t \in \mathcal{T}$, the cloud service provider incurs a cost $c_{t,j} \in \mathbb{R}_+$ per job of type $j \in \mathcal{J}$ in the queue of any one of the servers. Moreover, jobs may be dynamically reallocated from one server to another at a (job dependent) cost $c^{\mathrm{r}}_j \in \mathbb{R}_+$ per unit. Reallocation decisions also take immediate effect. At the end of the planning horizon, the cloud service provider must outsource any remaining jobs in the queue at unit cost $c^{\mathrm{o}}_j \in \mathbb{R}_+$, $j \in \mathcal{J}$.

Hint. Since job fractions can be allocated to each server, there is no need to consider integrality constraints for the number of jobs allocated to or processed by a server.

1. Suppose that the $d_{t,j}$, $j \in \mathcal{J}$, $t \in \mathcal{T}$ are deterministic:
   (a) (2 points) Formulate the expression for the number of jobs of type $j \in \mathcal{J}$ in the queue of server s at time $t \in \mathcal{T}$, denoted by $q_{t,j,s}$, in dependence of the queue length $q_{t-1,j,s}$ at $t-1$.
   (b) (8 points) Formulate the problem that the cloud service provider must solve to optimally allocate jobs (job fractions) to servers.
2. Suppose that $d_{t,j}$, $j \in \mathcal{J}$, $t \in \mathcal{T}$ are uncertain:
   (a) (5 points) You are given N historical realizations for the numbers of jobs that arrived in the past periods (these may be fractional). Propose an uncertainty set for $\{d_{t,j}\}_{j \in \mathcal{J}, t \in \mathcal{T}}$. State any assumptions you use.
   (b) (10 points) Formulate the associated adaptive robust optimization problem and propose a methodology for solving it (approximately). Discuss the benefits and drawbacks of your proposed solution approach (you do not need to derive the robust counterpart but you should discuss its complexity).

(c) (5 points) Comment on the benefits and drawbacks of your chosen uncertainty set as compared to possible alternatives. When answering this question, keep in mind your proposed solution to 2(b).

(d) (5 points) Do you believe that the demand for cloud computing services will change in the future? Do you think that the uncertainty set you proposed will accurately represent job demand uncertainty in the future?

Problem 3 (Robust Statistics)**.** (35 points)

In statistics, regression analysis is a statistical method concerned with estimating the relationships among variables. The most popular method for estimating the unknown parameters in a linear regression model is linear least squares. Given n observations $(y_i, \boldsymbol{x}_i) \in \mathbb{R} \times \mathbb{R}^m$, $i = 1, \ldots, n$, linear least squares is concerned with determining the optimal coefficient vector $\boldsymbol{\beta} \in \mathbb{R}^m$ which solves

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^m} \; \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\|_2 \,, \tag{5}$$

where

$$\boldsymbol{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}^\top \text{ and } \boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_n \end{bmatrix}^\top.$$

When the number of observations n is small, the linear least squares method is prone to over-fitting. In order to prevent over-fitting, the statistics and machine learning communities have proposed to introduce a regularization term in the objective of (5):

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^m} \; \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\|_2 + \rho\|\boldsymbol{\beta}\|_2, \tag{6}$$

yielding the so-called regularized linear least squares estimate which penalizes models with extreme parameter values. Here, $\rho \in \mathbb{R}_+$ is a constant typically tuned by cross-validation. In this problem, we investigate the relationship between robust optimization and regularization. Specifically, we will demonstrate that (6) is the robust counterpart of a problem where the matrix $\boldsymbol{X}$ is subject to uncertainty. We will guide you through the steps of the proof. Please note that each step is independent of the others so that you can get points even if you are unable to prove some of the statements.

Consider the robust regression problem

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^m} \; \max_{\boldsymbol{Q} \in \mathcal{U}} \; \|\boldsymbol{y} - ((\boldsymbol{X} + \boldsymbol{Q})\boldsymbol{\beta})\|_2 \,, \tag{7}$$

whose objective is to find the optimal regression coefficients $\boldsymbol{\beta}$ that minimize the ℓ_2-norm of the estimation error in the worst-case realization of $\boldsymbol{Q} \in \mathcal{U} \subseteq \mathbb{R}^{n \times m}$. Suppose the uncertainty set is expressible as

$$\mathcal{U} := \left\{ \boldsymbol{Q} \in \mathbb{R}^{n \times m} \;:\; \|\boldsymbol{Q}\|_F \leq \rho \right\},$$

where for any $l, k \in \mathbb{N}$ and $\boldsymbol{A} \in \mathbb{R}^{l \times k}$, $\|\boldsymbol{A}\|_F$ denotes the Frobenius norm of $\boldsymbol{A}$ defined through

$$\|\boldsymbol{A}\|_F := \sqrt{\text{trace}(\boldsymbol{A}^\top \boldsymbol{A})}.$$

1. (5 points) Show that for any $\boldsymbol{\beta} \in \mathbb{R}^m$ and $\boldsymbol{Q} \in \mathcal{U}$, it holds that

$$\|\boldsymbol{y} - ((\boldsymbol{X} + \boldsymbol{Q})\boldsymbol{\beta})\|_2 \leq \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2 + \rho\|\boldsymbol{\beta}\|_2.$$

2. (15 points) For $\boldsymbol{\beta} \in \mathbb{R}^m$, we define

$$\boldsymbol{Q}_0(\boldsymbol{\beta}) = \begin{cases} -\rho \mathbf{e}_1 [\boldsymbol{f}(\boldsymbol{\beta})]^\top & \text{if } \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} = 0 \\ -\rho \dfrac{\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}}{\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2} [\boldsymbol{f}(\boldsymbol{\beta})]^\top & \text{otherwise,} \end{cases}$$

   where

$$[\boldsymbol{f}(\boldsymbol{x})]_i = \begin{cases} \dfrac{x_i}{\|\boldsymbol{x}\|_2} & \text{if } \boldsymbol{x} \neq \mathbf{0}, \\ 0 & \text{otherwise.} \end{cases}$$

   Show that for any fixed $\boldsymbol{\beta} \in \mathbb{R}^m$, $\boldsymbol{Q}_0(\boldsymbol{\beta}) \in \mathcal{U}$.

3. (10 points) Show that the robust regression problem (7) is equivalent to (6).

4. (5 points) Is problem (6) efficiently solvable? Justify.

Hint. For any $l, k \in \mathbb{N}$, $\boldsymbol{A} \in \mathbb{R}^{l \times k}$ and $\boldsymbol{x} \in \mathbb{R}^k$, it holds that $\|\boldsymbol{A}\boldsymbol{x}\|_2 \leq \|\boldsymbol{A}\|_F \|\boldsymbol{x}\|_2$.