15.094, Spring 2014

Robust Modeling, Optimization and Computation

Midterm – Solutions

April 2, 2014

- This is a 3 hour exam.
- 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.
- 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!

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

$$\begin{aligned} Z_1 \quad = \quad \min_{\boldsymbol{x}\in\mathcal{X}} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}(\boldsymbol{\xi})\boldsymbol{x} \le \boldsymbol{b}(\boldsymbol{\xi}) \quad \forall \boldsymbol{\xi} \in \Xi, \end{aligned} \tag{1}$$

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 $\boldsymbol{b} : \mathbb{R}^k \to \mathbb{R}^m$.

   (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

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

   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

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

   where

$$\text{cone}(\Xi) := \left\{ \sum_{i=1}^{k} \lambda_i \boldsymbol{\xi}_i \ : \ \boldsymbol{\xi}_i \in \Xi,\ \lambda_i \ge 0,\ i \in \{1, ..., k\},\ k \in \mathbb{N} \right\}. \tag{2}$$

   Then $Z_3 > Z_1$.

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

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

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:

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathcal{X},\ \boldsymbol{y}(\cdot)} \quad & \boldsymbol{c}^\top \boldsymbol{x} + \max_{\boldsymbol{\xi}\in\Xi} \ \boldsymbol{d}^\top \boldsymbol{y}(\boldsymbol{\xi}) \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{\xi}) \le \boldsymbol{b}(\boldsymbol{\xi}) \quad \forall \boldsymbol{\xi} \in \Xi, \end{aligned}$$

where $\mathcal{X} \subseteq \mathbb{R}^{n_x}$, $\boldsymbol{c} \in \mathbb{R}^{n_x}$, $\boldsymbol{d} \in \mathbb{R}^{n_y}$, $\boldsymbol{A} \in \mathbb{R}^{m\times n_x}$, $\boldsymbol{B} \in \mathbb{R}^{m\times n_y}$, $\boldsymbol{y} : \mathbb{R}^k \to \mathbb{R}^{n_y}$, $\Xi := \{\boldsymbol{\xi} \in \mathbb{R}^k \ : \ \mathbf{e}^\top\boldsymbol{\xi} \le \Gamma,\ \boldsymbol{\xi} \ge \mathbf{0}\}$ 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) An optimal robust solution is always an interior point of the feasible region.

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

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b} \quad \forall \boldsymbol{A} \in \mathcal{U} \\ & \boldsymbol{x} \in \mathbb{R}^{n_r} \times \mathbb{Z}^{n_z}, \end{aligned} \tag{3}$$

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 (3) 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 (3).

Solution 1. Two points were given for answering True or False correctly. Three points were given for a correct justification.

1. (a) TRUE. Consider the following instance of Problem (1):

$$\begin{aligned} Z_1 \quad = \quad \min_{x\in[1,2]} \quad & x \\ \text{s.t.} \quad & x \le \xi \quad \forall \xi \in \Xi := \overline{\Xi} \cap \{0, 1\}, \end{aligned}$$

where $\overline{\Xi} := [0.1, 0.9]$. Then, $\Xi = \emptyset$ and thus, $Z_1 = 1$. On the other hand, $\overline{\Xi} \cap [0, 1] = [0.1, 0.9]$ and thus,

$$Z_2 = \min_{x\in[1,2]} \quad x \quad \text{s.t.} \quad x \le 0.1,$$

which is infeasible, yielding $Z_2 = +\infty$. We have thus constructed an example with $\frac{Z_2}{Z_1} = +\infty$.

(b) FALSE. Since linear functions are positive homogeneous of degree 1, for any fixed $\boldsymbol{\psi} \in \mathbb{R}^k$, we have

$$\begin{aligned} \boldsymbol{\psi}^\top \boldsymbol{\xi} \ge 0 \quad \forall \boldsymbol{\xi} \in \Xi \quad &\Leftrightarrow \quad \alpha \boldsymbol{\psi}^\top \boldsymbol{\xi} \ge 0 \quad \forall \boldsymbol{\xi} \in \Xi \text{ and } \alpha > 0 \\ &\Leftrightarrow \quad \boldsymbol{\psi}^\top \boldsymbol{\xi} \ge 0 \quad \forall \boldsymbol{\xi} \in \mathrm{cone}(\Xi). \end{aligned}$$

Thus, $Z_3 = Z_1$ always holds.

2. FALSE. (Counterexample from Ludovica Rizzo's exam) The minimax inequality states that, in the absence of convexity assumptions on $\mathcal{X}$ and Ξ, the converse inequality holds:

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

We now construct a counterexample to show that the inequality above may be strict. Let $\mathcal{X} := \{(x_1, x_2) \in \mathbb{R}^2_+ \ : \ x_1 + x_2 = 1\}$, $\Xi := \{(0,1)\} \cup \{(1,0)\}$ and $\boldsymbol{c}(\boldsymbol{\xi}) = \boldsymbol{\xi}$. Then,

$$\max_{\boldsymbol{\xi}\in\Xi} \boldsymbol{\xi}^\top \boldsymbol{x} = \max(x_1, x_2).$$

Therefore,

$$\min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{\xi}\in\Xi} \boldsymbol{\xi}^\top \boldsymbol{x} \ = \ \min_{\boldsymbol{x}\in\mathcal{X}} \max(x_1, x_2) = \frac{1}{2}.$$

On the other hand,

$$\min_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{\xi}^\top \boldsymbol{x} = \min_{x_1\in[0,1]} x_1(\xi_1 + \xi_2) - \xi_2 = \min(\xi_1, \xi_2),$$

yielding

$$\max_{\boldsymbol{\xi}\in\Xi} \min_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{\xi}^\top \boldsymbol{x} = \max_{\boldsymbol{\xi}\in\Xi} \min(\xi_1, \xi_2) = 0.$$

Thus,

$$\max_{\boldsymbol{\xi}\in\Xi} \min_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{\xi}^\top \boldsymbol{x} \ < \ \min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{\xi}\in\Xi} \boldsymbol{\xi}^\top \boldsymbol{x},$$

implying that the inequality in the statement does not hold without further assumptions on $\mathcal{X}$ and/or Ξ.

3. TRUE. Let $\mathcal{P}_a$ and $\mathcal{P}_s$ denote the problems on the left and on the right of the equality, respectively. Also, denote by Z_a and Z_s their respective optimal objective values.

For each $\boldsymbol{\xi} \in \Xi$, let $\boldsymbol{x}^\star(\boldsymbol{\xi}) \in \mathrm{argmin}_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{c}(\boldsymbol{\xi})^\top \boldsymbol{x}$. Then, $\boldsymbol{x}^\star(\boldsymbol{\xi}) \in \mathcal{X}\ \forall \boldsymbol{\xi} \in \Xi$. Thus, $\boldsymbol{x}^\star(\cdot)$ is feasible in $\mathcal{P}_a$

with objective value $\max_{\boldsymbol{\xi}\in\Xi} \boldsymbol{c}(\boldsymbol{\xi})^\top \boldsymbol{x}^\star(\boldsymbol{\xi}) = Z_s$. Therefore, $Z_a \leq Z_s$.

Suppose that $Z_a < Z_s$. Then, there exists $\tilde{\boldsymbol{x}}(\cdot)$ such that $\tilde{\boldsymbol{x}}(\boldsymbol{\xi}) \in \mathcal{X}\ \forall \boldsymbol{\xi} \in \Xi$ and

$$\max_{\boldsymbol{\xi}\in\Xi}\ \boldsymbol{c}(\boldsymbol{\xi})^\top \tilde{\boldsymbol{x}}(\boldsymbol{\xi})\ <\ \max_{\boldsymbol{\xi}\in\Xi}\ \min_{\boldsymbol{x}\in\mathcal{X}}\ \boldsymbol{c}(\boldsymbol{\xi})^\top \boldsymbol{x}(\boldsymbol{\xi})\ =\ \max_{\boldsymbol{\xi}\in\Xi}\ \boldsymbol{c}(\boldsymbol{\xi})^\top \boldsymbol{x}^\star(\boldsymbol{\xi}).$$

The above inequality in turn implies that there exists $\tilde{\boldsymbol{\xi}}$ such that

$$\boldsymbol{c}(\tilde{\boldsymbol{\xi}})^\top \tilde{\boldsymbol{x}}(\tilde{\boldsymbol{\xi}})\ < \boldsymbol{c}(\tilde{\boldsymbol{\xi}})^\top \boldsymbol{x}^\star(\tilde{\boldsymbol{\xi}}),$$

contradicting the fact that $\boldsymbol{x}^\star(\tilde{\boldsymbol{\xi}}) \in \operatorname{argmin}_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{c}(\tilde{\boldsymbol{\xi}})^\top \boldsymbol{x}$. Thus, we must have $Z_a = Z_s$.

4. FALSE. For example, if $\boldsymbol{B}$ and $\boldsymbol{d}$ are both zero, any robust optimal $\boldsymbol{x}$ and any constant $\boldsymbol{y}(\cdot)$ is optimal.

5. TRUE. Suppose $\boldsymbol{c}(\boldsymbol{\xi})^\top \boldsymbol{x}_1 \geq \boldsymbol{c}(\boldsymbol{\xi})^\top \boldsymbol{x}_2$ for all $\boldsymbol{\xi} \in \Xi$ and $\exists \boldsymbol{\xi}_1$ such that $\boldsymbol{c}(\boldsymbol{\xi}_1)^\top \boldsymbol{x}_1 > \boldsymbol{c}(\boldsymbol{\xi}_1)^\top \boldsymbol{x}_2$. Then, $\boldsymbol{x}_1$ is not a Pareto Robustly Optimal solution, which is a contradiction.

6. FALSE. With no other assumptions, any point of the feasible region may be an optimal robust solution. A counterexample is

$$\begin{aligned} \max_{x} \quad & x \\ \text{s.t.} \quad & 0 \leq x \leq 2 \\ & ux \geq 2 \quad \forall u \in [1, 2]. \end{aligned}$$

The feasible region of the problem without uncertain constraints is $[0, 2]$. The feasible set of the robust problem is $[1, 2]$ and $x = 2$ is an optimal robust solution. Thus, in this problem, there exists an optimal solution that is an endpoint (not an interior point) of the feasible interval.

7. FALSE. The problem may be infeasible, even if the algorithm returns an answer. For instance:

$$\begin{aligned} \max_{x} \quad & x \\ \text{s.t.} \quad & x \leq 1.5 \\ & ux \geq 1 \quad \forall u \in [0.9, 1.1] \\ & x \in \mathbb{Z} \end{aligned}$$

The linear relaxation of the above problem yields the robust solution $x^* = 1.5$, with no cuts generated. The branch-and-bound algorithm then returns $x^* = 1$. However, the original problem has no feasible integer solution.

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

Consider a cloud computing system over a finite planning horizon $\mathcal{T} := \{1, \ldots, T\}$. The system consists of a set $\mathcal{S} := \{1, \ldots, S\}$ of servers. Each server $s \in \mathcal{S}$ has I different kinds of resources indexed by $i \in \mathcal{I} := \{1, \ldots, I\}$ (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 $j \in \mathcal{J} := \{1, \ldots, J\}$. 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?

Solution 2. In addition to the decision variable introduced in the text, we let $x_{t,j,s}$ and $z_{t,j,s}$ denote the number of jobs of type j processed by and assigned to server s at time t, respectively. Also, we let $y^{\text{in}}_{t,j,s}$ ($y^{\text{out}}_{t,j,s}$) denote the number of jobs of type j reallocated to (from) server s at time t. Finally, we denote by $R_{t,i,s}$ the units of resource i available to server s at time t.

1. (a) The number of jobs of type j in the queue of server s at time t is given by:

$$q_{t,j,s} = q_{t-1,j,s} - x_{t-1,j,s} + y^{\text{in}}_{t,j,s} - y^{\text{out}}_{t,j,s} + z_{t,j,s},$$

where $q_{0,j,s} \in \mathbb{R}_+$ denotes the number of jobs of type j in the queue of server s at the beginning of the planning horizon.

(b) The objective of the problem is to minimize the combined costs for expanding the capacity, reallocating jobs, maintaining a queue and outsourcing. It is expressible as:

$$\text{minimize} \sum_{t,i,s} c^e_{t,i,s}(R_{t,i,s} - R_{t-1,i,s}) + \sum_{t,j,s} c_{t,j} q_{t,j,s} + \sum_{t,j,s} c^r_j y^{\text{out}}_{t,j,s} + \sum_{j,s} c^o_j q_{T,j,s},$$

where $R_{0,i,s} := R_{i,s}$. The resource constraints

$$\sum_j r_{i,j,s} x_{t,j,s} \leq R_{t,i,s} \quad \forall i, s, t$$

ensure that the resources needed to process $\{x_{t,j,s}\}_{j \in \mathcal{J}}$ jobs on server s at time t never exceed the server's resources. The reallocation constraints

$$\sum_s y^{\text{in}}_{t,j,s} = \sum_s y^{\text{out}}_{t,j,s} \quad \forall j, t$$

guarantee that every job leaving a server must be committed to another server. The allocation constraint

$$\sum_s z_{t,j,s} \geq d_{t,j} \quad \forall t, j \tag{4}$$

guarantees that all jobs are allocated to a server. The constraints

$$0 \leq R_{t-1,i,s} \leq R_{t,i,s} \quad \forall i, s, t$$

ensure that capacity cannot be decommissioned (sold). Finally, the constraints

$$q_{t,j,s},\ y^{\text{in}}_{t,j,s},\ y^{\text{out}}_{t,j,s},\ z_{t,j,s} \geq 0$$

ensure that the decisions are valid.

2. (a) There are multiple correct solutions for this question. It is natural to expect demand for the various types of jobs to be correlated. Thus, uncertainty sets that capture this correlation are given more points. Uncertainty sets that capture variability and correlation over time (e.g. day versus night) may be relevant depending on the frequency of arrivals (length of each time-interval), which is not given in the problem statement. We neglect it and assume that the job arrival process is stationary (most uncertainty sets discussed in class make this assumption, either implicitly or explicitly).

   We now proceed to construct a data-driven uncertainty set for $\boldsymbol{d} := (\boldsymbol{d}_1, \ldots, \boldsymbol{d}_T)$. Let $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ be the sample mean and sample covariance of the data. We let

$$\mathcal{U} := \bigtimes_{t \in \mathcal{T}} \mathcal{U}_t,$$

   where

$$\mathcal{U}_t := \left\{ \boldsymbol{d}_t \in \mathbb{R}^J : \|\boldsymbol{d}_t - \hat{\boldsymbol{\mu}}\| \leq \Gamma_1, \|\boldsymbol{\Sigma} - \hat{\boldsymbol{\Sigma}}\| \leq \Gamma_2, \begin{pmatrix} \frac{1-\epsilon}{\epsilon} & (\boldsymbol{d}_t - \boldsymbol{\mu})^T \\ \boldsymbol{d}_t - \boldsymbol{\mu} & \boldsymbol{\Sigma} \end{pmatrix} \succeq \boldsymbol{0} \right\}.$$

   We note that this set, which does not capture correlation over time may still be conservative.

   (b) The associated adaptive robust optimization problem can be obtained by modeling the decision variables for each time $t \in \mathcal{T}$ as functions of the history of observations $\{d_{\tau,j}\}_{\tau \leq t, j \in \mathcal{J}}$, enforcing that (ALL) constraints hold for all $\boldsymbol{d} \in \mathcal{U}$ and minimizing the objective in the worst-case realization of $\boldsymbol{d} \in \mathcal{U}$. This problem is severely computationally intractable as it optimizes over functional decisions and presents a continuum of constraints. We propose to obtain a conservative approximation by restricting the functional variables to be affine in the uncertain parameters. In the absence of integrality restrictions, this approximation is preferable over the finite adaptability one as the size of the corresponding robust counterpart will remain polynomial in the size of the original problem. Under this approximation and with the proposed uncertainty set, the RC is an SOCP of size polynomial in the size of the original problem and is thus efficiently solvable. IMPORTANT: If you modeled (4) as an equality constraint and restricted the functional variables to be constant or piecewise constant, the resulting robust counterpart will be infeasible! The same issue arises if you used job fractions rather than jobs numbers -resulting in a problem with random recourse- and did not model the queue constraint using an inequality!

(c) Possible alternatives would be to use e.g., (i) the CLT set, (ii) a data-driven box uncertainty set, or (iii) the "classical" ellipsoidal set. When compared to our proposed uncertainty set, (i) does not provide probabilistic guarantees at all, (ii) cannot capture correlation information, while (iii) does not provide probabilistic guarantees that depend on the number of samples (greater confidence with increasing number of samples). Nevertheless, a major drawback of our proposed uncertainty set as compared to (i) and (ii) is that the robust counterpart of the resource allocation problem will be an SOCP rather than an LP.

(d) The demand for cloud computing services will likely increase in the future. Unfortunately, our uncertainty set (which assumes stationarity of the job arrivals) is unable to identify any *trend* in the data. Ideally, one should "de-trend" the historical data first and then construct uncertainty sets around this trend. It is nevertheless unclear how to obtain elegant probabilistic guarantees when the data is not stationary.

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. 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

$$\|\boldsymbol{A}\|_F := \sqrt{\sum_{i,j} a_{ij}^2}$$

1. (8 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. (9 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 $\mathbf{e}_1 = (1, 0, \ldots, 0)$ and

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

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

3. (9 points) Show that the robust regression problem (7) is equivalent to (6).

4. (9 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 $\|\boldsymbol{A}\boldsymbol{x}\|_2 \leq \|\boldsymbol{A}\|_F \|\boldsymbol{x}\|_2$.

Solution 3. 1.

$$\begin{aligned} \|\boldsymbol{y} - ((\boldsymbol{X} + \boldsymbol{Q})\boldsymbol{\beta})\|_2 &\leq \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2 + \|\boldsymbol{Q}\boldsymbol{\beta}\|_2 \quad \text{(triangle inequality)} \\ &\leq \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2 + \|\boldsymbol{Q}\|_F \|\boldsymbol{\beta}\|_2 \quad \text{(hint)} \\ &\leq \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2 + \rho \|\boldsymbol{\beta}\|_2 \quad (\text{since } \boldsymbol{Q} \in \mathcal{U}) \end{aligned}$$

2. First we observe that for any vectors $\boldsymbol{u}, \boldsymbol{v}$, we have

$$\|\boldsymbol{u}\boldsymbol{v}^\top\|_F = \sqrt{\sum_{i,j}(u_i v_j)^2} = \sqrt{\left(\sum_i u_i^2\right)\left(\sum_j v_j^2\right)} = \|\boldsymbol{u}\|_2\|\boldsymbol{v}\|_2.$$

Also note that $\|\boldsymbol{f}(\boldsymbol{\beta})\|_2 \leq 1$ for any $\boldsymbol{\beta}$. Now

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

i.e. $\boldsymbol{Q}_0(\boldsymbol{\beta}) \in \mathcal{U}$.

3. Part 1 implies that the solution to (7) is no greater than the solution to (6); furthermore, they are equivalent if, for every $\boldsymbol{\beta}$, equality in Part 1 holds for some $\boldsymbol{Q} \in \mathcal{U}$. So in light of Part 2, it suffices for us to show that the equality holds for, in particular, $\boldsymbol{Q} = \boldsymbol{Q}_0(\boldsymbol{\beta})$.
   We consider only the non-trivial case where $\boldsymbol{\beta} \neq \mathbf{0}$. In this case, $\boldsymbol{f}(\boldsymbol{\beta}) = \dfrac{\boldsymbol{\beta}}{\|\boldsymbol{\beta}\|_2}$.
   Case 1: $\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} = 0$

$$\begin{aligned} \|\boldsymbol{y} - ((\boldsymbol{X} + \boldsymbol{Q})\boldsymbol{\beta})\|_2 \quad &= \|-\boldsymbol{Q}_0(\boldsymbol{\beta})\boldsymbol{\beta}\|_2 \\ &= \left\| \rho \mathbf{e}_1 \frac{\boldsymbol{\beta}^\top \boldsymbol{\beta}}{\|\boldsymbol{\beta}\|_2} \right\|_2 \\ &= \rho \|\boldsymbol{\beta}\|_2 \|\mathbf{e}_1\|_2 \\ &= \rho \|\boldsymbol{\beta}\|_2 \\ &= \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2 + \rho \|\boldsymbol{\beta}\|_2. \end{aligned}$$

   Case 2: $\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} \neq 0$

$$\begin{aligned} \|\boldsymbol{y} - ((\boldsymbol{X} + \boldsymbol{Q})\boldsymbol{\beta})\|_2 \quad &= \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Q}_0(\boldsymbol{\beta})\boldsymbol{\beta}\|_2 \\ &= \left\| \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} + \rho \frac{\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}}{\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2} \frac{\boldsymbol{\beta}^\top \boldsymbol{\beta}}{\|\boldsymbol{\beta}\|_2} \right\|_2 \\ &= \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2 \left(1 + \frac{\rho}{\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2} \|\boldsymbol{\beta}\|_2 \right) \\ &= \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2 + \rho \|\boldsymbol{\beta}\|_2. \end{aligned}$$

4. It is indeed efficiently solvable. Acceptable answers include observing that the problem is an unconstrained convex optimization problem (and can be thus be solved in closed form) or noting that it is equivalent to an SOCP (and thus solvable with standard interior-point techniques).