

# 15.094 Homework 2

Due: 03/17/2021

## 1 General Convex Uncertainty Sets

### 1.1

Let

$$h(z) = \sum_{i=1}^r \frac{\alpha_i}{p_i} |(d^i)^T z - \beta_i|^{p_i},$$

where  $\alpha_i > 0$  and  $p_i > 1$ , for all  $i$ .

Derive the robust counterpart of the constraint

$$(\bar{a} + Pz)^T x \leq b, \quad \forall z \in \mathcal{Z} = \{z : h(z) \leq \Gamma\}.$$

(Hint: You may use the results from the textbook without proof).

### 1.2

The so-called entropy uncertainty set is defined as

$$\mathcal{Z} = \left\{ z \in \mathbb{R}^L \mid \|z\|_\infty \leq 1, \sum_{i=1}^L \{(1 + z_i) \log(1 + z_i) + (1 - z_i) \log(1 - z_i)\} \leq \beta \right\},$$

where  $\log$  denotes the natural logarithm or logarithm in base  $e$ .

Derive the robust counterpart of the constraint

$$(\bar{a} + Pz)^T x \leq b, \quad \forall z \in \mathcal{Z}$$

## 2 Cloud computing system

(Please use the code provided in `HW2.CloudComputing.jl/ipynb` for this question. Data generation functions, cost vectors and plotting tools are provided.)

Consider a cloud computing system over a finite time horizon  $t \in \{1, \dots, T\}$ , with a set of servers,  $s \in \{1, \dots, S\}$ . To lower verbosity, we will use  $S$  interchangeably to refer to set  $\{1, \dots, S\}$ . Each server  $s \in S$  has a finite amount of resources indexed by  $i \in \{1, \dots, I\}$ .

$r_{i,s} \in \mathbb{R}^+$ ,  $\forall i \in I, s \in S$  is the fixed capacity of resource  $i$  at server  $s$ . It costs a constant  $F_i$  to maintain one unit of resource  $i$  at server  $s$  in each time period  $t$ . The optimization problem is to find the cost-minimizing value of fixed resources  $r_{i,s}$  subject to uncertain demand.

The demand is represented in time period  $t$  with  $D_{i,s,t}$  jobs arriving at the facility, requesting resource  $i$  from server  $s$  to complete. We must never fail to satisfy demand, since folks need their Netflix! There are three ways to complete each job:

- Serve  $D_{i,s,t}$  in the current server using fixed capacity  $r_{i,s}$ .
- Serve  $D_{i,s,t}$  in the current server by temporarily expanding  $r_{i,s}$  of server  $s$  in time  $t$  by  $e_{i,s,t}$ , while incurring an additional cost of  $V_i$  per unit.
- Transfer some or part of the job ( $u_{i,s_1,s_2,t}$ ) from server  $s_1$  to server  $s_2$  with spare capacity in resource  $i$ , costing  $C_i$  for a transfer from server  $s_1$  to  $s_2$ .

There are some additional constraints on the system. Expansions  $e_{i,s,t}$  result in increases in the temperature of each server. Our temperature surrogate is  $h_{s,t}$ , which is the three time step moving average of expansions in all resources (moving average of  $\sum_{i \in I} e_{i,s,t}$ ) for each server  $s$  at each time  $t$ .  $h_{s,t}$  cannot exceed 1. Assume that the temperature of each server is zero at all time periods before  $t = 1$ .

To keep it simple, we will assume that servers may complete fractional jobs, and capacities of resources are continuously variable.

## 2.1

Please formulate the nominal linear optimization problem, minimizing the sum of fixed, expansion and reallocation costs subject to demand and temperature constraints.

## 2.2

Use JuMP/JuMPeR to solve the nominal capacity allocation problem, using the `generate_data` function provided, with  $I = 6$  resources,  $S = 5$  servers, over  $T = 25$  periods. Please include the plot for matrix  $r$  (your fixed capacity allocation), your optimal cost, the time-statistics plots of expansions  $e$ , as well as the temperature plot. (Please use the plotting functions provided for your convenience. There should be four plots total.) Examine the temperature plot. Does your system anticipate peaks? How can you tell?

## 2.3

Solve the nominal problem, but without the ability to transfer jobs ( $u_{i,s_1,s_2,t} = 0, \forall i, s_1, s_2, t$ ). Provide the same four plots as above. How does this affect the cost and the allocation of  $r$  and  $e$ ? What is the intuition behind it? As far as

the system temperature, can the system anticipate peak demand? How is the system temperature response different compared to the problem with transfers?

## 2.4

Write the robust version of the capacity allocation problem by adding uncertainty  $d$  to demands  $(D_{i,s,t} + d_{i,s,t})$  in your optimization problem. The uncertainty in demand is decoupled for each server; assume that uncertain variables  $d_{1:I,s,1:T}$  for each server  $s$  lie in a budget uncertainty with bounded support  $[-1, 1]$  so that we are guaranteed to be feasible for %95 of uncertain outcomes. Please provide the form of your uncertainty sets as well as computations for the safety factors.

## 2.5

Solve the RO problem in JuMPeR/JuMP with and without job transfers. (The robust counterpart will take a lot longer to solve.) Please provide plots of  $r$  and the temperature, as well as the costs. (There should be four plots). How do your  $r$ , cost and the temperature plots change? How are the robust solutions similar or different? What do the results say about the ability of server farms to buffer variable demand?

## 2.6

Include the source code for the nominal and robust formulations in your problem set.

As always, please come to office hours if you have any questions or concerns.