15.094 Homework 3 Solutions

Due: 03/31/2021

1 Sparsity

In this exercise we examine what happens to sparsity properties of an LO under robustification. This is important as our ability to solve large scale LO problems within within an acceptable time horizon depends on the sparsity of the problems.

(a) Consider the constraint $a^T x \leq b,\ \forall a \in \mathcal{U}$, where $\mathcal{U} = \{a :\ \|a - \overline{a}\|_\infty \leq \Gamma\}$ (here $\overline{a} \in \mathbb{R}^n$ and $\Gamma > 0$ are given. Formulate the RC as a finite number of linear inequality constraints.

Solution

The given constraint is equivalent to $(\overline{a}+z)^T x \leq b \quad \forall z : \|z\|_\infty \leq \Gamma$. Using what we have learned in class this can be written as $\overline{a}^T x + \Gamma\|x\|_1 \leq b$. Using auxiliary variables u_i which represent $|x_i|$, this can be written as:

$$\bar{a}^T x + \Gamma \sum_{i=1}^{n} u_i \leq b$$

$$u_i \geq x_i,\ \forall i = 1, \ldots, n$$

$$u_i \geq -x_i,\ \forall i = 1, \ldots, n$$

(b) Comment on the sparsity of the new representation in Part (a). If $\overline{a}$ has mostly zero entries (in which case we would call the constraint $\overline{a}^T x \leq b$ "sparse"), are the constraints in the new representation of the uncertain constraint still sparse? Please be as specific as possible.

Solution The first constraint now contains n more non-zero entries, resulting in a less sparse constraint. Additionally, we have added n auxiliary variables to the problem. However the rest of the constraints contain only two non-zero entries, which are very sparse.

(c) Repeat Parts (a) and (b) with a different uncertainty set:

$$\mathcal{U} = \{a : \|a - \bar{a}\|_1 \le \Gamma\}.$$

Solution

Likewise, we can easily derive that the RC is $\bar{a}^T x + \Gamma\|x\|_\infty \le b$. Introducing a single auxiliary variable v which represents $\|x\|_\infty$,

$$\bar{a}^T x + \Gamma v \le b$$
$$v \ge x_i, \ \forall i = 1, \ldots, n$$
$$v \ge -x_i, \ \forall i = 1, \ldots, n$$

The first constraint is still sparse, as the number of non-zero entries increases by one. We have only added 1 auxiliary variable to the problem. The rest of the constraints are again sparse, as they involve two non-zero entries.

(d) What differences, if any, did you observe between Parts (b) and (c)? Offer an explanation for why there are (or are not) differences. Does this say anything about how to choose uncertainty sets?

Solution

In part (b), robustification results in $2n$ sparse constraints and 1 dense constraint, with n new variables. In part (c), robustification results in $2n+1$ sparse constraints, with the addition of one auxiliary variable. Thus, simply from the scope of sparsity and problem size, (c) is a better choice. This demonstrates that if you wish to preserve certain nice property of the original model (in this case, sparsity), the choice of uncertainty set might matter.

(e) (Extra Credit) You can test the time performance of the two different uncertain formulations on the robust counterpart of Problem 2.5 (the robust unlimited cash case), by varying the uncertainty type between the box uncertainty and 1-norm uncertainty, for a choice of $\Gamma = 1$. Please report the solution time of the two problems. Are the results what you expect? (Note: Make sure to only time the `solve` step, since the model creation time can be significant. Furthermore, make sure that you `solve` another problem before you do your time comparison, because of the compile time of JuMP/JuMPeR/Gurobi.)

2 World Food Programme, Syria Case Study

For this problem, we will be considering a simplified example of the multicommodity flow problem that was posed in Lecture 6 on the World Food Programme. The nominal problem is provided in `HW3_WFP.jl`, with the required data in the Homework3 folder (and we will be covering it in Recitation 5). The principles of the problem are as follows:

- The objective of initial problem is to minimize total cost while supplying enough food for all people for the given day.
- We have a set of demand nodes (N_D).
- We have data `pc`, which designates the prices of goods available at potential supply nodes. The supply nodes are separated into a set of local nodes N_L, regional nodes N_R and international nodes N_I.
- We have data `hc` which designates the edges along which commodities can flow, with the costs in the units of USD per metric ton.
- We design rations over all available commodities, to satisfy nutritional constraints. For simplicity, we assume that the rations are the same for all demand nodes, and that all people receive the same rations.

Please note that there are some important unit changes in the constraints. Procurement and transportation costs are in USD/ton. Nutritional requirements per person are in their own values. Nutrient contents of commodities are per 100g of the commodities. The rations are in kgs per person.

We have defined a baseline model and the following variables for your convenience:

- $\texttt{F}_{\text{source, sink, commodity}}$: how many tons of a commodity travel from source node to sink node.
- $\texttt{procurement}_{\text{node, commodity}}$: how many tons of a commodity are procured in the node.
- $\texttt{delivery}_{\text{node, commodity}}$: how many tons of commodity are delivered to a node.
- $\texttt{transportation}_{\text{source, sink}}$: how many tons of all commodities travel from source node to sink node.
- $\texttt{ration\_pp}_{\text{commodity}}$: how many kg of a commodity are in one person's rations.
- $\texttt{nutrients\_pp}_{\text{nutrient}}$: how much of a nutrient is in one person's rations.
- `transportation_cost`: sum of all transportation costs in USD.

- `procurement_cost`: sum of all procurement costs in USD.

In this section, you will be consider a series of case studies over the WFP problem, with and without uncertainty, to improve our hypothetical implementation of the program.

Performance Assessment

To be able to assess the performance of our allocation under different constraints and levels of robustness, we will consider the performance metrics below.

- Probabilistic guarantee of our resource allocation (0.5 if zero uncertainty, ≥ 0.5 with robustness).
- Total cost in USD.
- Ratio of procurement cost to total cost.
- Ratio of transportation cost to total cost.
- Ratio of international procurement cost as a fraction of total procurement cost.
- A slack variable on the constraint for nutrients per person. At 1, the nutrient constraint should be satisfied exactly, while values between 0 and 1 the nutritional constraints should be satisfied by the fraction.
- Total cost per person fed in USD. If we satisfy demand fractionally with a nutrient slack, please count each person by that fraction.
- Number of individual purchases (i.e. nonzero `procurement` variables).
- Number of active transport edges (i.e. nonzero `transportation` variables).

Some of these have not been included yet, so we expect you to formulate the variables/post-processing as needed. We advise you to design an efficient method to report the results of these objectives, since we will rely on these repeatedly.

2.1

Please solve the nominal cost-minimizing problem that satisfies nutritional needs fully, and report the results with respect to variables defined in Performance Assessment. Furthermore, comment on the food mix, as well as nutrients per person. (Optional: try turning off the diet constraints and see how the rations change.)

Solution

Please check out `HW3_WFP_Solutions.jl` for the code that replicates the solutions. The solution of the nominal is below.

PP	TotalCost	Proc/Total	Trans/Total	Intl/TotalProc	Slack	Cost/Person	NumActiveProc	NumActiveTrans
Float64	Float64	Float64	Float64	Float64	Float64	Float64	Int64	Int64
0.5	7915.7	0.604	0.396	0.306	1.0	0.1028	14	7

Figure 1: Performance metrics for optimization problem with no uncertainty and unlimited budget.

PP	TotalCost	Proc/Total	Trans/Total	Intl/TotalProc	Slack	Cost/Person	NumActiveProc	NumActiveTrans
Float64	Float64	Float64	Float64	Float64	Float64	Float64	Int64	Int64
0.5	6000.0	0.604	0.396	0.306	0.758	0.1028	14	7

Figure 2: Performance metrics for optimization problem with no uncertainty and limited budget.

If you look at the actual composition of rations and nutrients, you see that using a relatively sparse set of commodities (8 out of 25) and procurements (14/187), we can satisfy the nutritional requirements exactly.

Optional component: If you turn off the diet constraints, you will see that the fat content of the optimal rations is unacceptably high.

2.2

Unfortunately, in reality we only have \$6000 dollars for the day, to be able to satisfy as much demand as possible. Come up with a way to change the problem so that you can satisfy the budget constraints and feed as many people as possible. (Hint: use your slack variable.) Explain your method and and the changes to your model. Report your results with respect to variables in Performance Assessment. How does our strategy with limited resources change relatively to our strategy with unlimited resources?

Solution:

Using the hint, the new objective of the problem is to maximize the slack variable, subject to a budget constraint. This allows you to fulfill as much of the demand as possible, while not going above budget. You can see the relevant changes in `HW3_WFP_Solutions.jl`.

The results look quite similar. You adopt the same procurement and transportation strategy, with the caveat that you buy less of each item since you have a limited budget

2.3

We would like to protect against uncertainty in commodity costs. The international commodity prices (i.e. cost of goods bought in international nodes N_I) are independent, normally distributed with a standard deviation equal to 5% of the nominal price. The price of a commodity in a regional or local node is also normally distributed around the nominal price, but is positively correlated to the prices of other commodities in that node. The cost perturbation is as follows, for commodity c in local/regional node n:

$$\Delta(\mathrm{Cost_{n,c}}) = 0.35\mathrm{Cost_{n,c}} z_{\mathrm{n,c}} + 0.05 \sum_{\mathrm{k \in \{commodities\ in\ n\} \setminus c}} \mathrm{Cost_{n,k}} z_{\mathrm{n,k}}$$

where z is the uncertain variable in the ellipsoidal uncertainty set. Please come up with your P matrix (hint: it's square), model this with a joint ellipsoidal uncertainty set, and report your formulation here. Provide a method to compute the safety factor of your uncertainty set for a given probabilistic guarantee between 1 and 0.5. Provide intuition about why the uncertainty may be correlated.

Solution:

As per the hint in Discussions, it is actually even easier to think elementwise about $P^T x$ than about the matrix P, since the robust counterpart of the constraint involves a $\Gamma||P^T x||_2$ safety term. Note the original procurement cost constraint:

```
@constraint(m, procurement_cost >= sum(r[:Price] * procurement[r.A,
r.Food] for r in eachrow(pc)))
```

This is a simple sum of products of nominal prices and quantities.

$$\text{Procurement Cost} \geq \sum_{N,C} \text{Cost}_{n,c} \times \mathrm{Quantity_{n,c}}$$

Thus we can think about the perturbation in cost in an elementwise fashion, where $\Delta(\mathrm{Cost_{n,c}})$ describes the perturbation of price of commodity c at node n. Thus the coefficients of $x_{\mathrm{n,c}}$ below describe the terms of the in the $P^T x$ vector:

$$\begin{aligned}
\Delta(\mathrm{Cost_{n,c}})\mathrm{Quantity_{n,c}} &= 0.05\mathrm{Cost_{n,c}}\mathrm{Quantity_{n,c}} && \forall\ c,\ n \in N_I \\
\Delta(\mathrm{Cost_{n,c}})\mathrm{Quantity_{n,c}} &= 0.35\mathrm{Cost_{n,c}}\mathrm{Quantity_{n,c}} + \\
&\quad 0.05 \sum_{\mathrm{k \in \{commodities\ in\ n\} \setminus c}} \mathrm{Cost_{n,k}}\mathrm{Quantity_{n,k}} && \forall\ c,\ n \in \{N \setminus N_I\}
\end{aligned}$$

You get the robust counterpart of the above, with z lying in an ellipsoidal set, by expressing it as the following:

$$\text{Procurement Cost} \geq \sum_{N,C} \text{Cost}_{n,c} \times \mathrm{Quantity_{n,c}} + \Gamma||\Delta(\mathrm{Cost_{N,C}})||_2$$

To get the value of coefficient Γ, you can use the inverse of the cumulative distribution function of the normal distribution at a given probabilistic protection, since we assume that our uncertainty is normally distributed. This you can obtain using the `quantile` function on the normal distribution in

`Distributions.jl`. Please check out the code for the exact implementation in JuMP.

As for why the uncertainty might be correlated in this fashion, there are a few compelling answers. International nodes likely have relatively low, uncorrelated uncertainty since they receive their goods from all over the world, with more stable pricing. The price shock due to a drought in Libya on international commodities coming from Gaziantep will be damped by the availability of the same commodities from around the world. Furthermore, we don't expect too much cross-commodity correlation, eg. the global price of rice is likely only marginally correlated with the global price of milk.

As for regional/local prices, the reasoning about the correlation is the inverse. Since these nodes receive goods only locally, adverse events will tend to influence the supply, and thus the price, of all goods in the given region. It's likely that a bad chickpea harvest in Aleppo is related to local phenomena such as adverse weather or pests, which will tend to influence all produce from Aleppo.

2.4

Solve the robust problem above over a range of safety factors that protect against [50, 75, 85, 90, 92, 95, 96, 97, 98, 99]% of uncertain outcomes. We encourage you to formulate this RO problem in JuMP using the robust counterpart instead of using JuMPeR, to aid your understanding. But either method is sufficient.

Tabulate the metrics from Performance Assessment for all safety factors and evaluate the results. How does our procurement and transportation strategy evolve as we protect against more uncertainty? How might practical considerations affect your choice of uncertainty protection?

(Note: Please relax the required equality constraints to be able to use the robust counterpart, and watch out for numerical precision of your solution.)

Solution:

The hint is getting at having to relax constraints to do with transportation cost (i.e. transportation cost should be greater than rather than equal to). The numerical precision issues should be obvious when you observe procurements with very small (1e-9) values in your solution. There are easily "trimmed" by only accepting values above a tolerance. We use the recommended 1e-3 value, with the caveat that we will almost certainly trim away any procurement and transportation of salt, since it is obtained in such small quantities. However, we will ignore this issue, since these costs are very small, and the same takeaways will apply. The results are given below.

As we protect against more uncertainty, we see a few distinct trends. The most obvious are that our slack value decreases and our cost/person increases, meaning we are able to satisfy a lower proportion of the demand. However, you will notice that the behavior of the other variables is quite unmonotonic, which implies that there is a change of basis, eg. a discrete change in whether or not a certain commodity is in the rations.

PP Float64	TotalCost Float64	Proc/Total Float64	Trans/Total Float64	Intl/TotalProc Float64	Slack Float64	Cost/Person Float64	NumActiveProc Int64	NumActiveTrans Int64
0.5	6000.0	0.604	0.396	0.306	0.758	0.1028	14	7
0.75	6000.0	0.538	0.462	0.504	0.72	0.10826	12	8
0.85	6000.0	0.544	0.456	0.493	0.705	0.11049	13	8
0.9	6000.0	0.548	0.452	0.485	0.697	0.11185	13	8
0.92	6000.0	0.55	0.45	0.48	0.693	0.11252	13	8
0.95	6000.0	0.549	0.451	0.488	0.685	0.1138	15	9
0.96	6000.0	0.549	0.451	0.492	0.682	0.11433	15	9
0.97	6000.0	0.55	0.45	0.488	0.678	0.11498	15	9
0.98	6000.0	0.55	0.45	0.488	0.673	0.11583	17	8
0.99	6000.0	0.55	0.45	0.534	0.665	0.1171	16	9
0.995	6000.0	0.549	0.451	0.596	0.66	0.11805	16	9

Figure 3: Performance table with limited resources.

We also see that the relative cost of robustness is low. Even in very uncertain probabilities (0.99 probability protection), we can provide around 67% of the required demand, and 88% percent of the demand we could provide with no adverse uncertain outcomes.

Given the results, two obvious practical considerations that might affect your choice of uncertainty protections is how many transports you have, and how many procurements you are willing to plan. If you only had manpower for 15 procurements and 9 transports, you could still use those resources effectively to protect against 97% of uncertain outcomes, but not 98%. Otherwise, the table would allow you to make decisions about how to allocate other scarce resources.

These aren't the only practical considerations. Some others may be that you don't want to spend too much money on transportation versus procurement, or that you want to keep the level of international procurement low to provide more benefit to the local economy.

2.5

Let's assume again that we have unlimited cash. Solve the robust cost-minimizing robust problem, and generate the Performance Assessment matrix. What do the results imply about how we should apply our procurement and transportation policies when the uncertainty is realized (i.e. when we are about to purchase and prices do not match what we expect)? A qualitative answer is sufficient, but please provide reasoning using the above data and the optimization formulation.

Solution:

The performance table is given below:

Interestingly, the solutions with limited and unlimited resources are exactly the same. This has some great practical considerations.

Let's assume we want to protect against 97% of uncertain outcomes. We arrive at Homs and we have budget to buy and transport milk as a result of our optimization. But we also know how much milk we would buy and transport if we wanted to feed everyone, since we have solved our problem without budget constraints. Thus, when we arrive at the market and the price is different than what we expect, we can solve a simple linear equation to figure out how much

PP Float64	TotalCost Float64	Proc/Total Float64	Trans/Total Float64	Intl/TotalProc Float64	Slack Float64	Cost/Person Float64	NumActiveProc Int64	NumActiveTrans Int64
0.5	7915.7	0.604	0.396	0.306	1.0	0.1028	14	7
0.75	8335.7	0.538	0.462	0.504	1.0	0.10826	12	8
0.85	8507.6	0.544	0.456	0.493	1.0	0.11049	13	8
0.9	8612.5	0.548	0.452	0.485	1.0	0.11185	13	8
0.92	8664.2	0.55	0.45	0.48	1.0	0.11252	13	8
0.95	8762.4	0.549	0.451	0.488	1.0	0.1138	15	9
0.96	8803.7	0.549	0.451	0.492	1.0	0.11433	15	9
0.97	8853.7	0.55	0.45	0.488	1.0	0.11498	15	9
0.98	8919.2	0.55	0.45	0.488	1.0	0.11583	17	8
0.99	9017.0	0.55	0.45	0.534	1.0	0.1171	16	9
0.995	9089.8	0.549	0.451	0.596	1.0	0.11805	16	9

Figure 4: Performance table with unlimited resources.

we can afford without going over the budget constraints. (Obviously, we would never want to buy more than the amount designated by the unlimited-cash solution.) Once we have decided on a procurement strategy, budget doesn't affect the kinds of goods bought, only their quantity.

If we are even more clever, we would think jointly about the budget for buying and transporting all goods from the procurement nodes. But at that point, you might as well solve the optimization problem in real time, without any cost uncertainty!

2.6

Please include your robust optimization code. (No need to include data processing and solver printouts.)

Solution:

Again, please check out the code if you need to clarify anything, and come see us at office hours!