Quantitative Risk Management Using Robust Optimization

Lecture Notes

Erick Delage

March 5, 2024

Acknowledgements

I would like to thank HEC Montréal and the CDM department of EPFL for supporting my ambitions of putting together learning material for a course on robust optimization. I would also like to thank all the students that participated to the course throughout the years. Their curiosity and dedication to learning the techniques of robust optimization have been a true inspiration for me. In particular, I would like to give special thanks to the following individuals who carefully reported a number of typos and needed clarifications: Masoud Chitsaz, Fan E, Matthieu Gruson, Aditya Malik, Saeed Marzban, Mehran Poursoltani, Lidiia Shchichko, and Quan Zhou. Finally, I would like to thank Chun Peng, Melvyn Sim, and Peng Xiong for their valuable help in developing the python examples and exercises.

Contents

Part I

Prologue

Chapter 1

Why a Surge of Interest in Robust Optimization?

1.1 A production problem example

Let's start our discussion with a simple example.

Example 1.1. :A simple production problem (See Example 1.1.1 in [10])

A company produces two kinds of drugs, DrugI and DrugII, containing a specific agent A, which is extracted from raw materials purchased on the market. The related production, cost, and resource data are given in the table below. The goal is to find the production plan that maximizes the profit of the company. Below the data for this problem is presented.

Table 1.1: Drug production data

Drug production data		
Parameter	Drug I	Drug II
Selling price, $ per 1000 packs	6,200	6,900
Content of agent A, g per 1000 packs	0.500	0.600
Manpower required, hours per 1000 packs	90.0	100.0
Equipment required, hours per 1000 packs	40.0	50.0
Operational costs, $ per 1000 packs	700	800

Table 1.2: Contents of raw materials

Raw material	Purchasing price, $ per kg	Content of agent A, g per kg
Raw I	100.00	0.01
Raw II	199.90	0.02

Table 1.3: Resources

Resources			
Budget, $	Manpower, hours	Equipment, hours	Capacity of raw materials, storage, kg
100,000	2,000	800	1,000

Actually, one can easily think of a linear programming model that could help identify a well motivated production plan. Take for instance the following one.

$$
\begin{array}{lll}
\underset{\mathrm{RI,RII,DI,DII}}{\text{maximize}} & 6200\mathrm{DI} + 6900\mathrm{DII} - (100\mathrm{RI} + 199.90\mathrm{RII} + 700\mathrm{DI} + 800\mathrm{DII}) & \\
\text{subject to} & \mathrm{RI} + \mathrm{RII} \leq 1000 & \text{(Storage)} \\
& 90\mathrm{DI} + 100\mathrm{DII} \leq 2000 & \text{(Manpower)} \\
& 40\mathrm{DI} + 50\mathrm{DII} \leq 800 & \text{(Equipment)} \\
& 100\mathrm{RI} + 199.9\mathrm{RII} + 700\mathrm{DI} + 800\mathrm{DII} \leq 100000 & \text{(Budget)} \\
& 0.01\mathrm{RI} + 0.02\mathrm{RII} - 0.5\mathrm{DI} - 0.6\mathrm{DII} \geq 0 & \text{(Agent A)} \\
& \mathrm{RI} \geq 0, \mathrm{RII} \geq 0, \mathrm{DI} \geq 0, \mathrm{DII} \geq 0, &
\end{array}
$$

where RI and RII are respectively the amount (in kg) of raw material of type 1 and 2 ordered, while DI and DII are respectively the amount (in 1000 packs) of drug 1 and 2 produced.

The solution of this mathematical model can be found easily and will suggest the following : order 438 kg of raw material 2 (no raw material 1) and produce 17 552 packs of drug 1 for a total profit of 8820$. Yet, we might realize that our solution is quite sensitive to the choice of parameters that we made. In particular, since the constraint (Agent A) is active at this solution (i.e. there are no extra amount of agent A after performing the production), a small perturbation of the coefficients that describe this constraint could make the current solution infeasible (i.e. impossible to implement). In particular, assume that after ordering the raw material and starting the production of some agent A, we realize that their was a 2% error in the estimation of the conversion rate of this raw material, namely that only 0.0196 g can be extracted per kg of raw material 2. This means that instead of producing 17 552 packs of drug 1, we will only be able to produce 17 201 packs. Hence, the profit will drop from 8820$ to 6889$ (i.e. a 22% drop). In fact, in practice the repercussions might be more severe if the company had for instance committed to the delivery of exactly 17 552 packs of drugs. Hypothetically, it might mean that a number of sick patients won't be able to pursue their drug treatment and will need to start over on a new one for which the side effects are unknown, etc.

This raises the question: "What could we have done better in order to protect ourselves from potential estimation errors? "

In fact, if we had considered that there was a chance that the conversion rates for raw materials 1 and 2 could potentially be off by 0.5% and 2% respectively. We could have instead solved the following linear program:

$$\begin{array}{lll}
\underset{\mathrm{RI,RII,DI,DII}}{\text{maximize}} & 6200\mathrm{DI} + 6900\mathrm{DII} - (100\mathrm{RI} + 199.90\mathrm{RII} + 700\mathrm{DI} + 800\mathrm{DII}) & \\
\text{subject to} & \mathrm{RI} + \mathrm{RII} \leq 1000 & \text{(Storage)} \\
& 90\mathrm{DI} + 100\mathrm{DII} \leq 2000 & \text{(Manpower)} \\
& 40\mathrm{DI} + 50\mathrm{DII} \leq 800 & \text{(Equipment)} \\
& 100\mathrm{RI} + 199.9\mathrm{RII} + 700\mathrm{DI} + 800\mathrm{DII} \leq 100000 & \text{(Budget)} \\
& 0.00995\mathrm{RI} + 0.0196\mathrm{RII} - 0.5\mathrm{DI} - 0.6\mathrm{DII} \geq 0 & \text{(Agent A)} \\
& \mathrm{RI} \geq 0, \mathrm{RII} \geq 0, \mathrm{DI} \geq 0, \mathrm{DII} \geq 0. &
\end{array}$$

This would have motivated using the production plan : order 878 kg of raw material 1 (no raw material 2) and produce 17 467 packs of drug 1 for a total profit of 8295$. This solution can be considered "immuned" to a respective 0.5% and 2% perturbation of the conversion rates of raw material 1 and 2. The immunization has to do with the guarantee that the production plan will be implementable and that it will generate this amount of profit no matter what the conversion rate ends up being, as long as it falls within the specified range.

1.2 A generalized need for robust optimization

In the above example, we observed that the solution of an optimization problem can quickly become infeasible (i.e. impossible to implement) when some parameters of the problem are slightly different than what had been initially planned. We also saw an example of modification that can be applied to the model to identify decisions that are "robust" to such perturbation. This observation is not merely a simple coincidence that appears on some oddly constructed models. In fact, in [13], the authors present an exhaustive study that confirmed that for many realistic decision problems :

> " quite small (just 0.1%) perturbations of "obviously uncertain" data coefficients can make the "nominal" optimal solution x^* heavily infeasible and thus practically meaningless. "

Their case study involved the NETLIB library, a library of about 90 realistic decision problems from different applications of operations research. The average number of decision variables in the library is around 2500 while they include around 1000 constraints on average. The authors considered that any coefficient that could not be described using a fraction p/q with $q = 1, ...100$ was most likely a coefficient that was subject to estimation error. Given such an "uncertain" coefficient $\tilde{a}_{ij}$ they considered that it might be perturbed such as $\tilde{a}_{ij} = (1+\epsilon\xi_{ij})a_{ij}$ where ξ_{ij} is distributed on $[-1, 1]$. Given a constraint $\sum_j \tilde{a}_{ij}x_{ij} \leq b_i$, they considered a solution x "unreliable" with respect to constraint i if there was more than a 2% chance that the constraint would be violated by a relative factor of more than 5%; in other words, that the ϵ-reliability index

$$\mathrm{Rel}^i_\epsilon(x) := \frac{\text{98th-Percentile}\left(\max\left(0,\ \sum_j \tilde{a}_{ij}x_{ij} - b_i\right)\right)}{\max(1, |b_i|)} \times 100\%$$

was above 5%.

Some of the results are summarized in the Table 2 of [13] reproduced below.

The authors then conclude that:

> " In applications of LP, there exists a real need of a technique capable of detecting cases when data uncertainty can heavily affect the quality of the nominal solution, and in these cases to generate a "reliable" solution, i.e. one which is immuned against uncertainty. "

1.3 A portfolio selection example

We follow with a second example which illustrates the need for robust optimization even in contexts where it is the objective function that is affected by uncertainty.

Example 1.2. : A simple portfolio optimization problem (See Portfolio Example in [31])

We consider a portfolio construction problem consisting of n stocks. We consider that stock i has future return r_i and are looking for the portfolio composition that will maximize total return on our investment. This can be done by solving the following linear program:

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{maximize}} & \displaystyle\sum_{i=1}^n r_i x_i \\ \text{subject to} & \displaystyle\sum_{i=1}^n x_i = 100\% \\ & x_i \geq 0\,,\ \forall\, i = 1, ..., n\ , \end{array}$$

where x_i is the proportion of the budget invested in stock i, the first constraint implies that we wish that all the budget be invested, and where $x_i \geq 0$ implies that we wish to avoid short selling a stock.

Actually, this is a fairly trivial problem to solve and it is usually a useless one to pose. It is trivial because the optimal solution is necessarily to invest all the budget in the stock which has largest final return r_i. It is useless because in an efficient financial market, it should only be possible to obtain a larger return with an investment by being exposed to larger risk. This fails to be captured by the model above given that the largest return is considered guaranteed.

Table 1.4: NETLIB problems with bad nominal solutions. "Nbad" refers to the number of unreliable constraints while "Index" refers to the largest relative violation, i.e. $\max_i \mathrm{Rel}_\epsilon^i(x^*)$ (in %)

Problem	Size	$\epsilon = 0.01\%$		$\epsilon = 0.1\%$		$\epsilon = 1\%$	
		Nbad	Index	Nbad	Index	Nbad	Index
80BAU3B	2263×9799	37	84	177	842	364	8420
25FV47	822×1571	14	16	28	162	35	1620
ADLITTLE	57×97			2	6	7	58
AFIRO	28×32			1	5	2	50
BNL2	2325×3489					24	34
BRANDY	221×249					1	5
CAPRI	272×353			10	39	14	390
CYCLE	1904×2857	2	110	5	1100	6	11 000
D2Q06C	2172×5167	107	1150	134	11 500	168	115 000
E226	224×282					2	15
FFFFF800	525×854					6	8
FINNIS	498×614	12	10	63	104	97	1040
GREENBEA	2393×5405	13	116	30	1160	37	11 600
KB2	44×41	5	27	6	268	10	2680
MAROS	847×1443	3	6	38	57	73	566
NESM	751×2923					37	20
PEROLD	626×1376	6	34	26	339	58	3390
PILOT	1442×3652	16	50	185	498	379	4980
PILOT4	411×1000	42	210 000	63	2 100 000	75	21 000 000
PILOT87	2031×4883	86	130	433	1300	990	13 000
PILOTJA	941×1988	4	46	20	463	59	4630
PILOTNOV	976×2172	4	69	13	694	47	6940
PILOTWE	723×2789	61	12 200	69	122 000	69	1 220 000
SCFXM1	331×457	1	95	3	946	11	9460
SCFXM2	661×914	2	95	6	946	21	9460
SCFXM3	91×1371	3	95	9	946	32	9460
SHARE1B	118×225	1	257	1	2570	1	25 700

In reality, we would need to consider that there is uncertainty about the return obtained from stock i. We call this uncertain return $\tilde{r}_i$ and characterize it using

$$\tilde{r}_i = \mu_i + \sigma_i z_i \ ,$$

where μ_i is the expected return, and σ_i describes the volatility of the return, finally z_i captures the source of the uncertainty about $\tilde{r}_i$. You may for example, think of each

z_i as being independently distributed according to a standard normal distribution.

We will consider the following, problem instance with $n = 150$:

$$\mu_i := 0.15 + i\frac{0.05}{150} \qquad \sigma_i := \frac{0.05}{450}\sqrt{2in(n+1)}\,, \qquad z_i \in [-1,\, 1]\,.$$

Note that μ_i increases with i but at the price of an increased risk portrayed by σ_i. Note also that our model assumes that each return $\tilde{r}_i$, lies in the following interval (see also table 1.5 for some examples).

$$\tilde{r}_i \in \left[0.15 + i\frac{0.05}{150} - \frac{0.05}{450}\sqrt{2in(n+1)},\ 0.15 + i\frac{0.05}{150} + \frac{0.05}{450}\sqrt{2in(n+1)}\right]$$

Table 1.5: Table of return intervals and expected return for different stocks in the model

Stock id	Return interval	Expected return
#1	[12.67%, 17.40%]	15.03%
#2	[11.72%, 18.41%]	15.06%
#3	[11.00%, 19.20%]	15.10%
#4	[10.40%, 19.86%]	15.13%
#5	[9.88%, 20.45%]	15.16%
...	...	...
#147	[−8.77%, 48.57%]	19.90%
#148	[−8.84%, 48.70%]	19.93%
#149	[−8.90%, 48.83%]	19.97%
#150	[−8.96%, 48.96%]	20.00%

We can also quantify our tolerance toward risk by assuming that at most Γ number of stocks will have their return diverge from their expected value. By controlling the value of Γ one can capture the idea that he is more or less risk averse. Setting $\Gamma = 0$ would justify investing in the stock with the highest estimated return no matter how risky it is, namely the stock with index 150 whose return might be between -8.96% and 48.96%. Conversely, by setting $\Gamma = 150$ the robust solution would simply invest all of the budget in the stock for which the lowest return achievable is the highest, namely stock #1 which has a worst-case return of 12.67%.

For illustrative purposes, let's assume we are worried of the case were up to four stock returns diverge from their expected value. Then, when investing in stock $i = 150$, we are at risk of loosing 8.96% of our investment although we expect to achieve 20% return. Alternatively, we could attempt to solve the following optimization problem that only worries about the worst-case return:

$$\begin{array}{lll} \underset{x\in\mathbb{R}^n}{\text{maximize}} & \displaystyle\min_{z:|z_i|\leq 1,\sum_i |z_i|\leq\Gamma} \sum_{i=1}^{n}(\mu_i+\sigma_i z_i)x_i \\ \text{subject to} & \displaystyle\sum_{i=1}^{n} x_i = 100\% \\ & x_i \geq 0\,,\,\forall\, i=1,...,n\;. \end{array}$$

We will later find out that the portfolio that is obtained using this model achieves an expected return of 18.62%, while it is guaranteed to generate a return above 17.38% as long as at most four stock returns diverge from there expected value.

1.4 Defining robust optimization

The two examples above illustrate perfectly what are the premises of robust optimization. The methodology is applied on a decision model for which we wish to identify solutions that are immuned with respect to the actual realization of some of the parameters within a given uncertainty region. In other words, let one be interested in an optimization model of the type:

$$\begin{array}{lll} (\text{Nominal problem}) & \underset{x}{\text{maximize}} & h(x,z) \\ & \text{subject to} & g_j(x,z)\leq 0\,,\,\forall\, j=1,...,J\;, \end{array}$$

where x is a vector of decision variables, z is a vector of parameters that affect the outcomes of our decision and which we might be uncertain about, $h(\cdot,\cdot)$ is a profit function, and $g_j(\cdot,\cdot)$ captures a constraint that might be affected by z. The robust optimization framework will suggest identifying solutions that are immuned to any realization that z might take in some uncertainty set $\mathcal{Z}$ by solving the following robust counterpart

$$\begin{array}{llll} (\text{Robust counterpart}) & \underset{x}{\text{maximize}} & \displaystyle\min_{z\in\mathcal{Z}}\; h(x,z) & (1.1\text{a}) \\ & \text{subject to} & g_j(x,z)\leq 0\,,\,\forall\, z\in\mathcal{Z},\;,\,\forall\, j=1,...,J & (1.1\text{b}) \end{array}$$

As was seen in the production problem, the modification to the constraints plays the role of identifying solutions that are implementable no matter what the realization of z is. Alternatively, the modification to the objective function plays the role of identifying solutions for which the worst-case profit is as high as possible.

One can also interpret the robust counterpart problem as robustifying the constraints of the nominal problem presented in epigraph form:.

$$\begin{array}{lll} (\text{Nominal in epigraph form}) & \underset{x,t}{\text{maximize}} & t \\ & \text{subject to} & t\leq h(x,z) \\ & & g_j(x,z)\leq 0\,,\,\forall\, j=1,...,J\;, \end{array}$$

for which the robust counterpart takes the form:

$$\text{(Robust counterpart in epigraph form)} \quad \begin{array}{ll} \underset{x,t}{\text{maximize}} & t \\ \text{subject to} & t \leq h(x,z)\,,\, \forall\, z \in \mathcal{Z} \\ & g_j(x,z) \leq 0\,,\, \forall\, z \in \mathcal{Z},\,,\, \forall\, j = 1,...,J\;. \end{array}$$

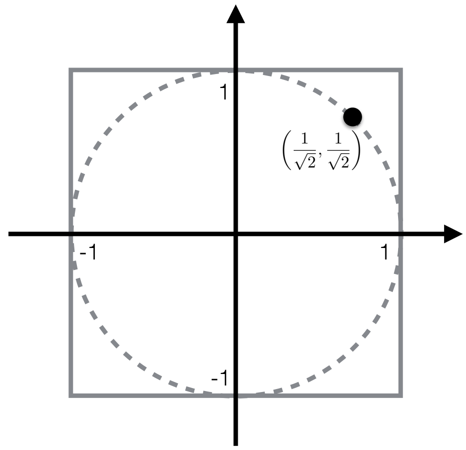
which is equivalent to the robust counterpart presented above. Through this derivation of the robust counterpart we can clearly observe that the robust counterpart maximizes an amount t that is guaranteed to be superseded by the true profit no matter what set of values the vector of parameters z takes.

1.5 A rise in popularity of the methodology

The idea of using a "maximin" model to identify good decisions is not a new one. It originally appeared in the work of Abraham Wald [49] who got his inspiration from the concept of zero-sum games in game theory. Indeed, he envisioned that decisions could be made assuming that nature was an adversarial player that would select the values of parameters after the decision was made. Later on, Allen L. Soyster introduces in [46] the idea of a robust constraint and applies it to linear programs. The idea did not initially gain much in popularity since it was considered too conservative. Indeed, A. L. Soyster was proposing that each parameter be defined in its own interval thus each would be allowed to take its respective worst-case value (similarly to what was presented in the production problem of section 1.1).

Actually, in the meantime, Herbert Scarf also proposed in [41] to employ a form of robust optimization which he called minimax stochastic programming as it was applied to evaluate the worst-case expected value of a function in cases where the distribution is unknown apart from its mean and variance statistics. He suggested optimizing this new measure and gave as example a simple newsvendor problem which he was able to solve analytically. Unfortunately, the approach was a little hard to generalize at the time to larger decision models.

One has to wait until the turn of the century to see a real enthusiasm catch on for the robust optimization method. Indeed, in their seminal paper [12], Ben-Tal and Nemirovski reintroduced the notion of robust optimization for convex optimization problems. The success of this paper lies in the fact that they were able to reduce the level of conservatism of the method by proposing the use of an ellipsoid as uncertainty set and to demonstrate that the robust counterpart model was often computationally tractable. Indeed, one can observe in the figure below that the ellipsoid (e.g. the circle in $\mathbb{R}^2$) does not allow all parameters to take on their worst-case value simultaneously.



The figure below presents the rise in popularity of the methodology between 1998 and 2015. In particular, the two graphs present a surprising growth of the annual number of new publications and new citations associated to this topic. We have estimated that these statistics indicate an average 20% yearly increase in productivity around this topic between 2005 and 2014. In addition, during this period, according to Proquest there has been 201 thesis and dissertation published and 715 abstracts using the keyword "robust optimization" at INFORMS annual meetings (127 actually in 2014).

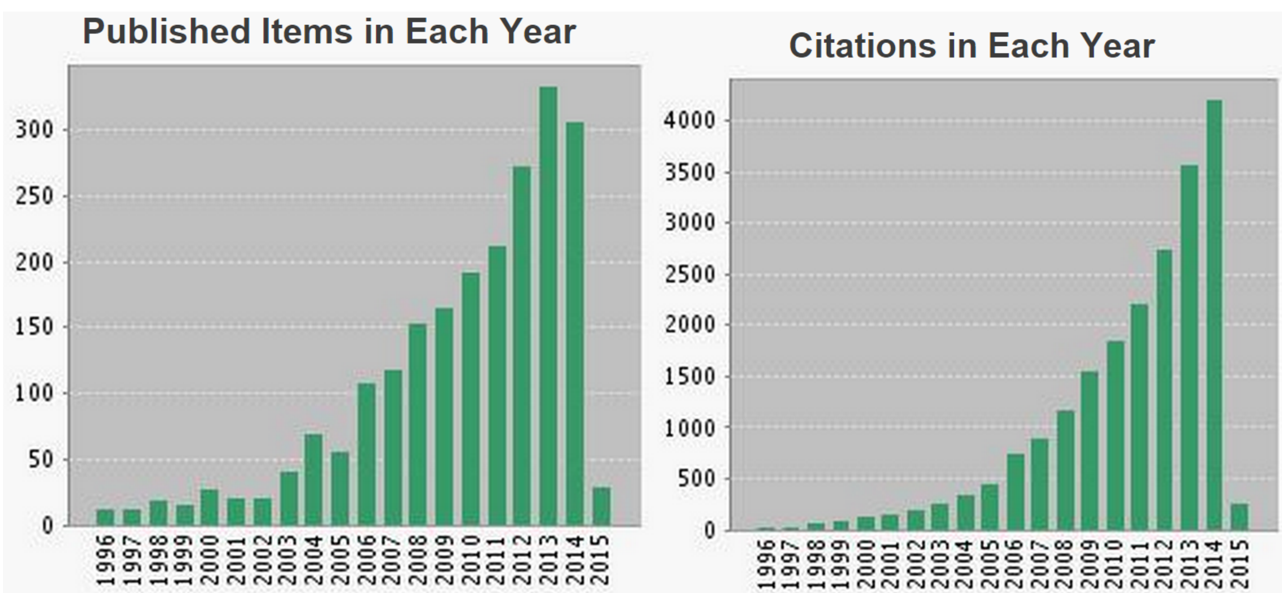


Figure 1.1: Rise in popularity of "robust optimization" in the scientific literature.

It is generally believed that the main reasons of the significant rebirth of robust optimization are the following:

- The discovery that if the structure of the uncertainty set is well chosen, then the robust counterpart model can be shown to reduce to an equivalent mathematical program that is not much harder to solve than the original problem.

- The development of more powerful computers and in particular of fast interior point methods that can be employed to solve convex optimization models that have richer structure than linear programs.

- Studies such as [19] which proposed simple uncertainty sets that could be used and which reinforced the connection with the stochastic programming literature. In particular, many now consider robust optimization as being a tractable approximation method that can address a number of stochastic programming models considered impossible to solve.

1.6 RSOME : A RO library for Python (and Matlab)

During this course, we will employ a Python library developed by Zhi Chen, Melvyn Sim and Peng Xiong in order to facilitate the application of robust optimization on decision problems that can be modelled using linear programming. The library was first introduced in [24] and all its documentation can be found on the official website (url).
https://xiongpengnus.github.io/rsome/about

The package can already be used to describe in algebraic form the details of a linear programming model. We will later show how to describe a robust optimization problem.

Looking back at our production problem one might wonder how to describe the nominal problem using "Robust and Stochastic Optimization Made Easy" (RSOME, pronounced "Aresome"). In fact, once the package is imported ("*from rsome import ro*") this is done using a set of simple commands that enables one to create a model ("*model = ro.Model*"), to define each decision vector ("*x=model.dvar*"), the objective function ("*model.max*"), each constraint ("*model.st*"), and then solve the model ("*model.solve()*") and retrieve the optimal value and solution ("*model.get()*" and "*x.get()*" respectively). The solver that is used by default when calling "*model.solve()*" is "*linprog()*" from the "*scipy.optimize*" package. One needs to know however that *linprog()* cannot solve robust optimization problems that employ more sophisticated uncertainty set (such as the ellipsoidal set). In this course, we will rather use Mosek which needs to be imported ("*from rsome import msk_solver as my_solver*") and called at the moment of solving ("*model.solve(my_solver)*").

Below you will find the models we developed for the nominal production planning and portfolio selection problems. The implementations are available in Google Colab.

Using RSOME to solve the nominal production problem (see Google Colab)

```
#Create model
model = ro.Model('production')

# Define decision variables
DrugI =model.dvar()                # Define a decision variable DrugI
DrugII =model.dvar()               # Define a decision variable DrugII
RawI =model.dvar()                 # Define a decision variable RawI
RawII =model.dvar()                # Define a decision variable RawII

#Objective to maximize the profit
model.max(6200*DrugI+6900*DrugII - (100*RawI+199.90*RawII+700*DrugI+800*DrugII))

# Storage constraint
model.st(RawI + RawII <= 1000)

# Manpower constraint
model.st(90*DrugI + 100*DrugII <= 2000)

# Equipment constraint
model.st(40*DrugI+50*DrugII <= 800)

# Budget constraint
model.st(100*RawI+199.9*RawII+700*DrugI+800*DrugII <= 100000)

# Constraint to have enough active agent A
model.st(0.01*RawI+0.02*RawII-0.5*DrugI-0.6*DrugII >= 0)

# Constraints that decision variables are non-negative
model.st(DrugI >= 0)                                 # DrugI is non-negative
model.st(DrugII >= 0)                                # DrugII is non-negative
model.st(RawI >= 0)                                  # RawI is non-negative
model.st(RawII >= 0)                                 # RawII is non-negative

# solve the model
model.solve(my_solver)
```

Using RSOME to solve the nominal portfolio selection (see Google Colab)

```
#Create portfolio model
model = ro.Model('portfolio')

#Portfolio weights
x = model.dvar(n)                                  # Fractions of investment

# Objective to maximize the return
model.max(mu@x.T)

# Constraint to invest all the wealth available
model.st(x.sum() == 1)                             # Summation of x is one
# Constraint that weights are positive
model.st(x >= 0)                                   # x is non-negative

model.solve(my_solver)

optobj_det = model.get() #get optimal objective value
xx_det    = x.get() #get optimal portfolio
```

In order to tap into the true purpose of RSOME, namely solving a robust optimization problem, one needs to define a vector of uncertain parameters ("*model.rvar*"), an uncertainty set (e.g. "$boxSet = (abs(z) \leq 1)$"), and implicate the uncertain vector in the objective function and constraints, while ensuring to link the uncertainty set in the objective function call ("*model.maxmin(...,boxSet)*") or the constraint ("*model.st((...).forall(boxSet))*").[1] See for example the robust portfolio selection problem below which employs a budgeted uncertainty set. Namely, that each z_i must be between -1 and 1 and that the sum of absolute deviation of the z vector from zero must be less than Γ.

[1]Note that with if only the *model.maxmin(...,boxSet)* command is used, then every objective function and constraint where the uncertain parameters appear will be internally replaced by its robust counterpart.

Using RSOME to solve the robust portfolio selection (see Google Colab)

```
# create a model
model = ro.Model('robustPortfolio')

# Declare uncertain parameters
z = model.rvar(n)  # Random variables
budgetSet = (z<=1, z>=-1,  #each parameter is between [-1, 1]
                  rso.norm(z,1)<=Gamma);   # Budget of uncertainty approach

# Portfolio weights
x=model.dvar(n);

# Objective to maximize the return
model.maxmin((mu+sigma*z)@x,budgetSet);
# Note that r is uncertain and depends in z.
# The objective should be interpreted as
# max{y : y<= (mu + sigma.*z)'x for z in uncertainty set}

# Constraint to invest all the wealth available
model.st(x.sum()==1);
# Constraint that weights are positive
model.st(x>=0);

# solve the model
model.solve(my_solver);
optobj_rob = model.get() #get optimal objective value
xx_rob   = x.get() #get optimal portfolio
```

When $\Gamma = 4$, one obtains with this model the robust portfolio that was proposed in the example of section 1.3 which guarantees a return above 17.38% as long as the return vector z satisfies the specified constraints. This protection is achieved at the price of a lower performance if the expected return values end up being reached. Specifically, this would mean a drop from 20% to 18.62% return, a difference of 1.38% which is often referred as the "price of robustness"[19].

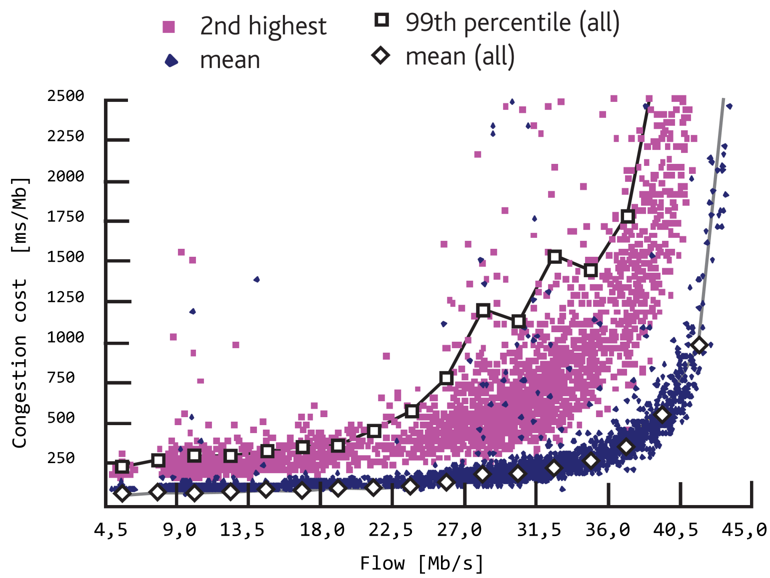
1.7 Applications

While one can find in [15] a number of references to interesting applications of robust optimization in statistics, supply chain management, integrated circuit design, antenna design, structural design, finance, etc. We describe briefly below some interesting applications that we have been exposed to.

1.7.1 Robust congestion minimization

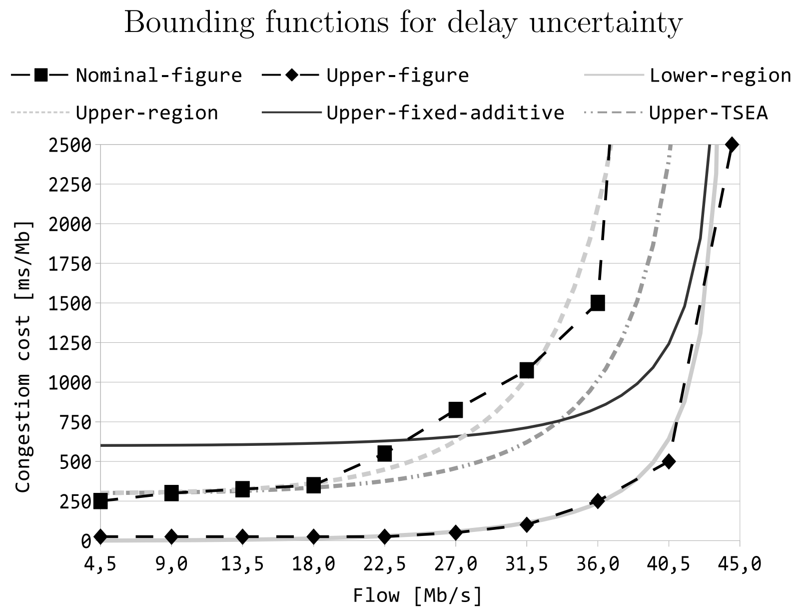
In [26], we considered the problem of routing packets on a telecommunication network with congestion. While in routing problems, delays of packets that travel on different links of such networks are typically modelled using M/M/1 queuing system, there are strong empirical evidence indicating that a number of factors cause the realized transmission time to often suffer substantial deviations from these theoretical estimates (see [29] as portrayed in the figure below).

Delay uncertainty on a link of a telecommunication network



Question: What is a good routing strategy to employ when one wishes to guarantee a certain level of Quality of Service (i.e. maximum packet delays) while what we know of the delay that might be encountered on each link given different amount of traffic takes the shape of historical observations that do not fit any predetermined theoretical model?

Our answer: We propose to encapsulate the uncertainty about the link delay between two convex functions, namely a nominal function $\hat{h}_i(x_i)$ (diamonds ♦) and an upper bounding function $h_i^+(x_i)$ (squares ■) as portrayed in the following figure.



We then replaced the nominal objective function calculating total weighted congestion

$$\underset{x\in\mathcal{X}}{\text{minimize}} \quad \sum_{i=1}^{n} \hat{h}_i(x_i) \ ,$$

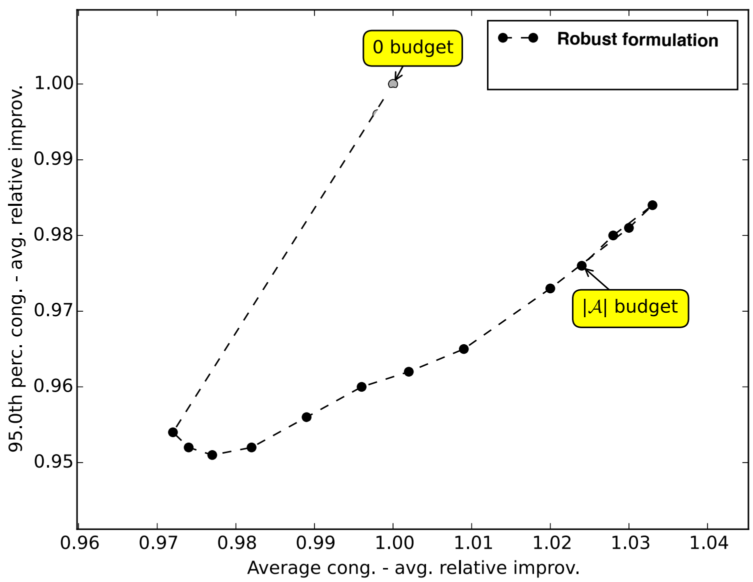
where $\mathcal{X}$ is the set of feasible routing strategy to satisfy the packet transmission demand, with the following robust counterpart

$$\underset{x\in\mathcal{X}}{\text{minimize}} \quad \max_{z\in\mathcal{Z}} \ \sum_{i=1}^{n} \hat{h}_i(x_i) + z_i(h_i^+(x_i) - \hat{h}_i(x_i)) \ ,$$

where $\mathcal{Z}$ is the set of all $z_i \in [0,1]$ such that $\sum_{i=1}^{n} z_i \leq \Gamma$. In other words, while we expect that some of the delay incurred might be as large as what is described by the upper bounding functions, we do not expect that more than Γ of these delay actually do.

With this solution scheme, RO has the potential to help Internet Service Providers (ISPs) to provide the requested *Quality of Service* (QoS) to incoming traffic. We experimented with an exhaustive set of realistic network management conditions in order to illustrate what type of trade-off can be achieved between the amount of total congestion that is expected versus the amount that is at risk of being achieved under less favourable conditions.

Trade-off between average (x-axis) and 95th percentile (y-axis) of total congestion

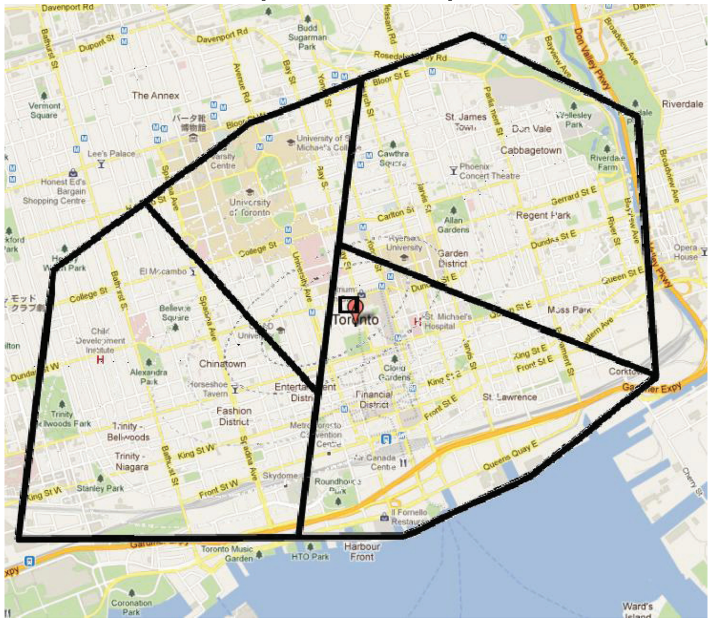


1.7.2 Robust Partitioning for Multi-Vehicle Routing

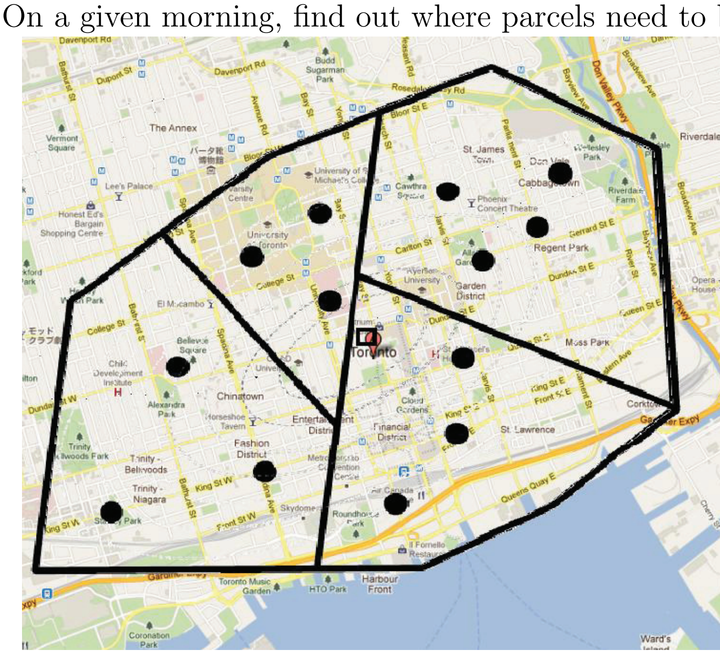
Consider being the operations director of a large parcel delivery service company. You need to divide a planar region into K subregions, each serviced by a different delivery vehicle, so that the total workload be most evenly distributed among the fleet.

Step #1: Identify the territory and where the parcel depot is located

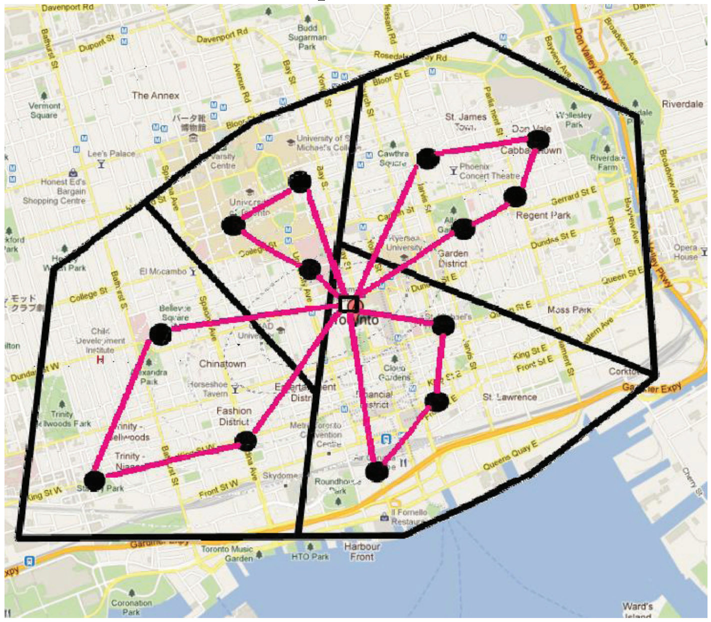
Step #2: Divide the territory into as many sectors as there are vehicle



Step #3: On a given morning, find out where parcels need to be delivered



Step #4: Each vehicle will deliver parcels to the locations in its own sector



Question: Given that you don't know yet where will the next batch of parcel need to be delivered but only know that some areas are more likely to receive parcels than others, how would you divide the space into sectors? Consider especially the fact that a driver will become frustrated if the sector he is assigned to always has a longer route to traverse than the sectors assigned to other drivers. Finally, how would you divide the territory if you did not have a "distribution" of parcel delivery locations but rather only some historical reports of where parcels were delivered at similar periods of time?

Our answer:

1. If we knew the distribution of parcel delivery locations, then we would probably need to divide the territory so that the largest workload among the different drivers is as small as possible

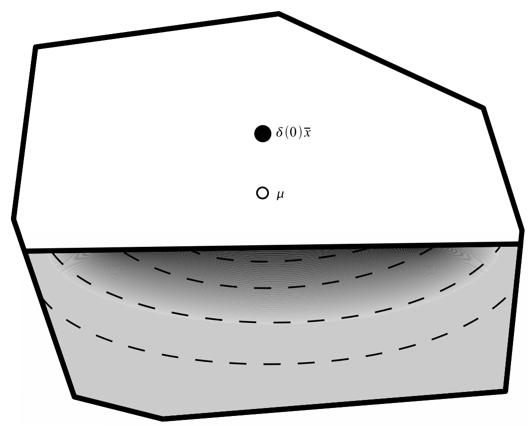
$$\underset{\{\mathcal{R}_1,\mathcal{R}_2,\ldots,\mathcal{R}_K\}}{\text{minimize}} \quad \max_i \;\; \mathbb{E}[\text{RouteLength}(\{\xi_1,\xi_2,\ldots,\xi_N\}\cap\mathcal{R}_i)] \;\; ,$$

2. Since we don't know the distribution, we instead use the historical data to identify a set $\mathcal{D}$ which we believe contains the true distribution of parcel delivery location.

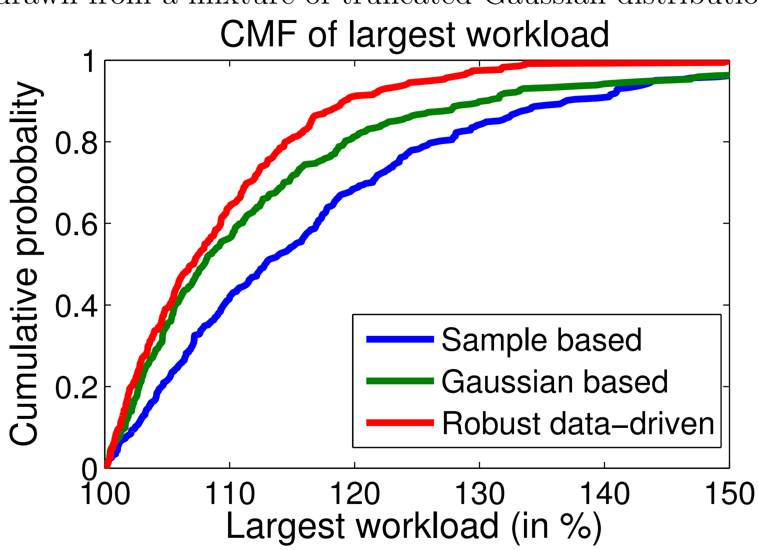
3. Given $\mathcal{D}$, we divide the territory so that the largest workload over the worst distribution of delivery locations is as small as possible

$$\underset{\{\mathcal{R}_1,\mathcal{R}_2,\ldots,\mathcal{R}_K\}}{\text{minimize}} \quad \sup_{F\in\mathcal{D}} \left\{\max_i \;\mathbb{E}[\text{RouteLength}(\{\xi_1,\xi_2,\ldots,\xi_N\}\cap\mathcal{R}_i)]\right\} \;\; ,$$

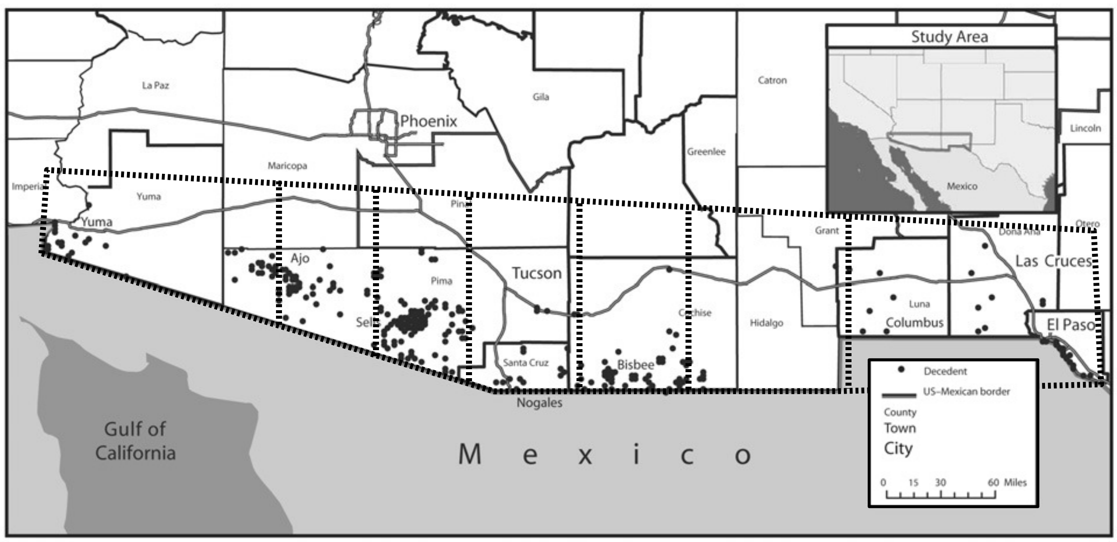
4. A side product is to characterize for any sector what is a worst-case distribution of demand locations



We simulated three partition schemes on a set of randomly generated parcel delivery problems where the territory needed to be divided into two regions and the delivery locations are drawn from a mixture of truncated Gaussian distributions.



We employed the same method to suggest how to divide the territory of USA-Mexico border among a set of seven border patrol cars.



1.7.3 Robust Aircraft Fleet Composition

In some joint work with The Boeing Cie we were asked to help develop a software that would allow airline companies to account for uncertain passenger demand at the moment of acquiring new aircraft. Note that these acquisitions are typically done 10 to 20 years ahead of schedule at a moment where many factors are still unknown to the airline company: e.g., passenger demand, fuel prices, etc. Yet, most airline companies sign these acquisition contracts based on a single scenario of what the future may be.

Question: Are these airline companies being neglectful in not considering a larger number of future scenarios for the different influencing factors?

Our answer: Not necessarily under two assumptions: 1) a reasonable model is used for what are the recourse actions that can be employed when the passenger demand is known and 2) the single scenario that is used is "well" chosen.

Here is our mathematical model (a stochastic mixed integer linear program) for this fleet composition problem:

$$(SP) \qquad \underset{\boldsymbol{x}}{\text{maximize}} \;\; \mathbb{E}[-\underbrace{\boldsymbol{o}^\mathsf{T}\boldsymbol{x}}_{\text{ownership cost}} + \underbrace{h(\boldsymbol{x}, \tilde{\boldsymbol{p}}, \tilde{\boldsymbol{c}}, \tilde{\boldsymbol{L}})}_{\text{future profits}}] \;\; ,$$

with $h(\boldsymbol{x}, \tilde{\boldsymbol{p}}, \tilde{\boldsymbol{c}}, \tilde{\boldsymbol{L}}) :=$

$$
\begin{array}{lll}
\max\limits_{z\geq 0, y\geq 0, w} & \sum\limits_k \Big(\sum\limits_i \overbrace{\tilde{p}_i^k w_i^k}^{\text{flight profit}} - \overbrace{\tilde{c}_k (z_k - x_k)^+}^{\text{rental cost}} + \overbrace{\tilde{L}_k (x_k - z_k)^+}^{\text{lease revenue}} \Big) & \\
\text{subject to} & w_i^k \in \{0,1\}\ ,\ \forall\, k, \forall\, i \quad \& \quad \sum\limits_k w_i^k = 1\ ,\ \forall\, i & \Big\}\ \text{Cover} \\
& y^k_{g\in \text{in}(v)} + \sum\limits_{i\in \text{arr}(v)} w_i^k = y^k_{g\in \text{out}(v)} + \sum\limits_{i\in \text{dep}(v)} w_i^k\ ,\ \forall\, k, \forall\, v & \Big\}\ \text{Balance} \\
& z_k = \sum\limits_{v\in\{v|\text{time}(v)=0\}} (y^k_{g\in \text{in}(v)} + \sum\limits_{i\in \text{arr}(v)} w_i^k)\ ,\quad \forall k & \Big\}\ \text{Count}
\end{array}
$$

where x captures how many aircraft of each type will initially be acquired, then z, y, and w are the possible recourse actions available once demand is known. First, w captures what type of aircraft serves each flight leg, y accounts for aircraft that are parked in airports between different flight, and z captures how many aircraft of each type is actually needed to serve the demand. Note that here we modelled the fact that the airline could both borrow some additional aircraft or lease out aircraft that it does not use.

This is in general a hard optimization problem to solve for the following reasons:

- One needs to identify and motivate a joint distribution model for the future flight leg profits achieved by each available aircraft.
- One needs to generate a large number of scenarios from the distribution model in order for the mathematical model to have finite dimensions.
- Current algorithms are inefficient at optimizing such stochastic programs because they involve integer decision variables.

What we realized is that if one has reliable information about the expected values $\bar{p} := \mathbb{E}[\tilde{\boldsymbol{p}}]$, $\bar{c} := \mathbb{E}[\tilde{\boldsymbol{c}}]$, and $\bar{L} := \mathbb{E}[\tilde{\boldsymbol{L}}]$, and wishes to be robust with respect to other properties of the distribution, then the robust optimization model

$$
(DRO) \qquad \underset{\boldsymbol{x}}{\text{maximize}} \ \min_{F\in\mathcal{D}} \mathbb{E}_F\,[- \underbrace{\boldsymbol{o}^\top \boldsymbol{x}}_{\text{ownership cost}} + \underbrace{h(\boldsymbol{x}, \tilde{\boldsymbol{p}}, \tilde{\boldsymbol{c}}, \tilde{\boldsymbol{L}})}_{\text{future profits}}\,]\ ,
$$

simply reduces to

$$
\begin{array}{lll}
\max\limits_{x,z\geq 0,y\geq 0,w} & -\underbrace{\boldsymbol{o}^{\mathsf{T}}\boldsymbol{x}}_{\text{ownership cost}} + \sum\limits_k \Big(\sum\limits_i \overbrace{\bar{p}_i^k w_i^k}^{\text{flight profit}} - \overbrace{\bar{c}_k (z_k - x_k)^+}^{\text{rental cost}} + \overbrace{\bar{L}_k (x_k - z_k)^+}^{\text{lease revenue}} \Big) & \\
\text{subject to} & w_i^k \in \{0,1\}\ ,\ \forall\, k, \forall\, i \quad \& \quad \sum\limits_k w_i^k = 1\ ,\ \forall\, i & \Big\}\ \text{Cover} \\
& y^k_{g\in \text{in}(v)} + \sum\limits_{i\in \text{arr}(v)} w_i^k = y^k_{g\in \text{out}(v)} + \sum\limits_{i\in \text{dep}(v)} w_i^k\ ,\ \forall\, k, \forall\, v & \Big\}\ \text{Balance} \\
& z_k = \sum\limits_{v\in\{v|\text{time}(v)=0\}} \big(y^k_{g\in \text{in}(v)} + \sum\limits_{i\in \text{arr}(v)} w_i^k\big)\ ,\quad \forall k & \Big\}\ \text{Count}
\end{array}
$$

We experimented with three test cases :

1. 3 types of aircraft, 84 flights, $\sigma_{\tilde{p}_i}/\mu_{\tilde{p}_i} \in [4\%, 53\%]$
2. 4 types of aircraft, 240 flights, $\sigma_{\tilde{p}_i}/\mu_{\tilde{p}_i} \in [2\%, 20\%]$
3. 13 types of aircraft, 535 flights, $\sigma_{\tilde{p}_i}/\mu_{\tilde{p}_i} \in [2\%, 58\%]$

Results:

Test	CPU Time		DRO sub-optimality
cases	DRO	SP	$\forall\, F \in \mathcal{D}$
#1	0.6 s	3 min	$< 6\%$
#2	1 s	14 min	$< 1\%$
#3	5 s	21 h	$< 7\%$

Our conclusion here was that in the three cases, it was wasteful to invest more than 7% of the expected profits in the development and resolution of models that would account for the uncertainty about factors such as passenger demand, fuel prices, etc.

Part II

Fundamental Theory

Chapter 2

Robust Counterpart of Linear Programs

In this chapter, we assume that the functions that need to be "robustified" are linear functions of both the decision variables and the vector of parameters. Namely, we investigate the robust counterpart model presented in problem (1.1) and repeated below

$$
\begin{array}{lll}
\text{(Robust counterpart)} & \underset{x}{\text{maximize}} & \min_{z\in\mathcal{Z}} \; h(x,z) \\
& \text{subject to} & g_j(x,z) \leq 0\,,\, \forall\, z \in \mathcal{Z},\,,\, \forall\, j = 1,...,J\;.
\end{array}
$$

and assume that the different functions that compose this model can be expressed as

$$
\begin{aligned}
h(x,z) &:= c(z)^T x + d(z) \\
g_j(x,z) &:= a_j(z)^T x - b_j(z)\;,
\end{aligned}
$$

where $x \in \mathbb{R}^n$, $z \in \mathbb{R}^m$, and where each function $c(z)$, $d(z)$, $a_j(z)$ and $b_j(z)$ is an affine function of z. In other words, it must be possible to describe each of these functions using the following forms

$$
\begin{array}{rcr}
c(z) := (P_0 z + p_0) & \& & d(z) = q_0^T z + r_0\,, \\
a_j(z) := (P_j z + p_j) & \& & b_j(z) = q_j^T z + r_j\,,
\end{array}
$$

for some $P_j \in \mathbb{R}^{n\times m}$, some $p_j \in \mathbb{R}^n$, some $q_j \in \mathbb{R}^m$ and some $r_j \in \mathbb{R}$.

The robust counterpart would then take the form:

$$
\begin{array}{lll}
\text{(LP-RC)} & \underset{x}{\text{maximize}} & \min_{z\in\mathcal{Z}} \; z^T P_0^T x + q_0^T z + p_0^T x + r_0 \qquad (2.1a) \\
& \text{subject to} & z^T P_j^T x + p_j^T x \leq q_j^T z + r_j\,,\, \forall\, z \in \mathcal{Z},\,,\, \forall\, j = 1,...,J \qquad (2.1b)
\end{array}
$$

This model is not amenable to readily available mathematical programming resolution software as it is not yet described in finite dimensional form. Indeed, each constraint indexed with j must be checked for all realization of z in $\mathcal{Z}$. Similarly, the objective function is not expressed in closed form; in order to evaluate it, one must search for the instance of z that achieves the minimum value.

For simplicity, we will start with a single robust constraint so that we have in hand a constraint that takes the form:

$$z^T P^T x + p^T x \leq q^T z + r\,,\ \forall\, z \in \mathcal{Z}$$

where we dropped the indexed notation for simplicity.

Let's initially look at the case where $P = I$, $p = a$, $q = 0$, $r = b$. This reduces to

$$(a + z)^T x \leq b\,,\ \forall\, z \in \mathcal{Z}\ , \tag{2.2}$$

which is perhaps the most famous version of a robust constraint.

The difficulty associated to treating this constraint is now entirely linked to the structure of $\mathcal{Z}$. Indeed, given a fixed x, we are asked to verify whether or not

$$\exists z \in \mathcal{Z}, z^T x > b - a^T x$$

if so then x would be infeasible. For a general uncertainty set $\mathcal{Z}$ (and in particular those that impose that z be integer), this question is known to be NP-hard (see NP-completeness of integer programming in [30]), meaning that we cannot expect to tackle problems where the vector of parameters would have a size larger than 10 or 20. On the other hand, if $\mathcal{Z}$ is simply a set of K scenarios for z, namely $z \in \{\bar{z}_1, \bar{z}_2, ..., \bar{z}_K\}$, then this verification is straightforward as shown in the following example.

Example 2.1. : Consider the case where $\mathcal{Z} := \{\bar{z}_1, \bar{z}_2, ..., \bar{z}_K\}$ and we wish to retrieve a tractable representation of the constraint:

$$z^T x \leq b - a^T x\,,\ \forall\, z \in \mathcal{Z}\ .$$

In this case, one simply need to check each member of $\mathcal{Z}$. Consequently, in this simple situation, the robust counterpart constraint (2.2) can be reformulated as:

$$\bar{z}_i^T x \leq b - a^T x\,,\ \forall i = 1, ..., K\ ,$$

or similarly

$$(a + \bar{z}_i)^T x \leq b\,,\ \forall i = 1, ..., K\ .$$

Note that in most practical contexts, we are interested in more than a finite set of scenarios (or if so it would be in a set of scenario of very large size). For this reason, we will first assume that $\mathcal{Z}$ is a bounded polyhedron and later work with convex uncertainty sets that are defined with a single convex inequality.

2.1 Polyhedral Uncertainty

In this section, we consider that uncertainty about the vector of parameters takes the form of a polyhedral set defined as follows.

Assumption 2.2. : The uncertainty set $\mathcal{Z}$ is a non-empty and bounded polyhedron that can be defined according to

$$\mathcal{Z} := \{z \in \mathbb{R}^m \,|\, w_i^T z \leq v_i \,,\, \forall\, i = 1, ..., s\}\,,$$

where for each $i = 1, ..., s$, we have that $w_i \in \mathbb{R}^m$ and $v_i \in \mathbb{R}$ capture a facet of the polyhedron through the expression $w_i^T z = v_i$. Moreover, since $\mathcal{Z}$ is non-empty, there must exist a $z_0 \in \mathcal{Z}$ and since it is bounded there must exist some $M > 0$ such that $\mathcal{Z} \subseteq \{z \in \mathbb{R}^m \,|\, -M \leq z \leq M\}$.

Under assumption 2.2, verifying whether a fixed x satisfies constraint (2.2) is equivalent to verifying whether the optimal value of the following LP is smaller or equal to $b - a^T x$.

$$\begin{aligned} \underset{z}{\text{maximize}} \quad & x^T z && \text{(2.3a)} \\ \text{subject to} \quad & Wz \leq v && \text{(2.3b)} \end{aligned}$$

where $W = [\; w_1 \;\; \ldots \;\; w_s \;]^T$ is the matrix in $\mathbb{R}^{s \times m}$ which rows are composed of each w_i.

Theorem 2.3. :*(LP Duality see Chapter 4 of [21]) Under assumption 2.2, the optimal value of linear program* (2.3) *is equal to the optimal value of the following dual problem*

$$\begin{aligned} \underset{\lambda}{\text{minimize}} \quad & v^T \lambda && \text{(2.4a)} \\ \text{subject to} \quad & W^T \lambda = x && \text{(2.4b)} \\ & \lambda \geq 0 && \text{(2.4c)} \end{aligned}$$

where $\lambda \in \mathbb{R}^s$. Moreover, problem (2.4) *has a feasible solution.*

Proof. Here is how one generally applies duality to replace a maximization problem with a minimization problem. Let us call Ψ the optimal value of problem (2.3). First, we will demonstrate how to obtain the dual problem which always achieves a larger value than Ψ, then we will employ Farkas lemma to guarantee that the two values are the same, and furthermore that the dual problem is feasible.

Step #1: Obtaining the dual problem Let's express a relaxed version of problem (2.3), where we have moved the constraints to the objective function:

$$\Upsilon(\lambda) := \max_{z} \;\; x^T z + \lambda^T (v - Wz)\;.$$

It is important to realize that as long as $\lambda \geq 0$ then $\Upsilon(\lambda) \geq \Psi$. This is the case because for any z that was feasible in problem (2.3) we will have

$$x^T z + \lambda^T (v - Wz) \geq x^T z \ ,$$

since $\lambda \geq 0$ and $v - Wz \geq 0$ for those z. Hence,

$$\max_{z \in \mathcal{Z}} \ x^T z \leq \max_{z \in \mathcal{Z}} \ x^T z + \lambda^T (v - Wz) \leq \max_{z} \ x^T z + \lambda^T (v - Wz) = \Upsilon(\lambda) \, .$$

The problem $\min_{\lambda \geq 0} \Upsilon(\lambda)$ therefore returns the lowest upper bound for Ψ. Yet, when studying more carefully the expression associated with $\Upsilon(\lambda)$, we can observe that

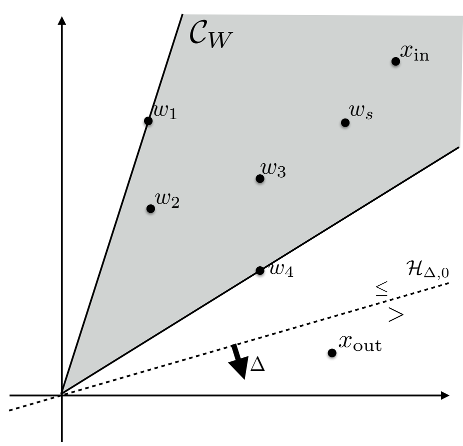
$$\Upsilon(\lambda) = \begin{cases} \lambda^T v & \text{if } x - W^T \lambda = 0 \\ \infty & \text{otherwise} \end{cases} \quad .$$

The problem $\min_{\lambda \geq 0} \Upsilon(\lambda)$ therefore reduces to problem (2.4).

Step #2: Introducing Farkas lemma The most critical step of this proof relies on Farkas lemma which states the following.

Lemma 2.4. : *Let W be a real $s \times m$ matrix and x be an m-dimensional vector. Then, exactly one of the following two statements is true:*

1. *There exists a $\lambda \in \mathbb{R}^s$ such that $W^T \lambda = x$ and $\lambda \geq 0$.*
2. *There exists a $\Delta \in \mathbb{R}^m$ such that $W\Delta \leq 0$ and $x^T \Delta > 0$.*



Let us consider the convex cone $\mathcal{C}_W := \{y \in \mathbb{R}^m \,|\, \exists \lambda \in \mathbb{R}^s, \ \lambda \geq 0, \ y = W^T \lambda\}$.[1] Geometrically, the cone $\mathcal{C}_W$ is the conic hull of the points in $\mathbb{R}^m$ defined by the rows of W, i.e. $\{ \ w_1 \quad w_2 \quad \ldots \quad w_s \ \}$. From this perspective, Condition 1 simply states that

[1] The cone $\mathcal{C}_W$ is convex since given any y_1 and y_2 in $\mathcal{C}_W$ (with their associated λ_1 and λ_2) and any $\theta \in [0, 1]$, we have that: $\theta\lambda_1 + (1-\theta)\lambda_2 \geq 0$ and $\theta y_1 + (1-\theta) y_2 = W^T(\theta\lambda_1 + (1-\theta)\lambda_2)$. So that, $\theta y_1 + (1-\theta) y_2 \in \mathcal{C}_W$.

$x \in \mathcal{C}_W$. Alternatively, Statement 2 appears more complex but can also be stated more simply. First note that it states that there exists a vector $\Delta \in \mathbb{R}^m$ such that the hyperplane $\mathcal{H}_{\Delta,0} := \{y \in \mathbb{R}^m \,|\, \Delta^T y = 0\}$ strictly separates x from the points $\{w_1, w_2, \ldots, w_s\}$: i.e.

$$\Delta^T x > 0 \;\;\&\;\; \Delta^T w_i \leq 0\,,\, \forall\, i = 1, \ldots, s\,.$$

Looking at the definition of our cone, Statement 2 is also equivalent to stating that $\mathcal{H}_{\Delta,0}$ strictly separates x from all of $\mathcal{C}_W$. Indeed, one can see that

$$W\Delta \leq 0 \;\;\Rightarrow\;\; \Delta^T w_i \leq 0 \,\forall\, i = 1, \ldots, s \;\;\Rightarrow\;\; \forall\, \lambda \geq 0,\, \Delta^T W \lambda = \sum_i \lambda_i \Delta^T w_i \leq 0\,.$$

It is therefore clear that both statements of Farkas lemma cannot be true simultaneously: x cannot be both a member of $\mathcal{C}_W$ and be separated from it by some hyperplane $\mathcal{H}_{\Delta,0}$. Yet, based on convex analysis it is also impossible that both statements are false. Indeed, the strict separating hyperplane theorem (see Appendix 10.1.1) states that it must be possible to strictly separate any point $x \notin \mathcal{C}_W$ from $\mathcal{C}_W$ using a hyperplane. We further argue that since $\mathcal{C}_W$ is a cone, it must always be possible to do so with a hyperplane such as $\mathcal{H}_{\Delta,0}$. To demonstrate this, let x be strictly separated using some $\mathcal{H}_{\Delta,\gamma} := \{y \in \mathbb{R}^m \,|\, \Delta^T y = \gamma\}$. It must first be the case that $\gamma \geq 0$ since $0 \in \mathcal{C}_W$ so that $\Delta^T 0 = 0 \leq \gamma$. Now let's assume that for all $y \in \mathcal{C}_W$ we would have that $\Delta^T y \leq 0$, well then it is clear that $\mathcal{H}_{\Delta,0}$ also strictly separates x from $\mathcal{C}_W$. In the case that there exists some $y_0 \in \mathcal{C}_W$ such that $\Delta^T y_0 > 0$ (thus implying that $\gamma > 0$) then the point $y_0' := (2\gamma/\Delta^T y_0) y_0$ should also be a member of $\mathcal{C}_W$. But, this leads to a contradiction since we would have

$$\Delta^T y_0' \leq \gamma < 2\gamma = (2\gamma/\Delta^T y_0)\Delta^T y_0 = \Delta^T y_0'\,.$$

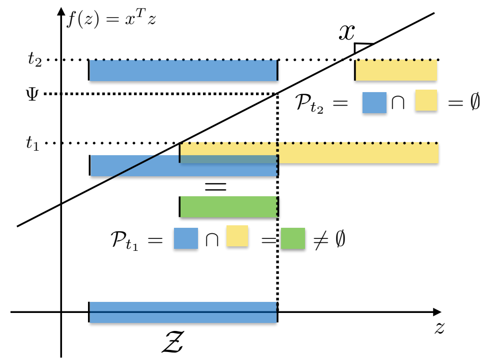
We conclude that either $x \in \mathcal{C}_W$ or there is a Δ such that $\mathcal{H}_{\Delta,0}$ strictly separates x from $\mathcal{C}_W$ which is equivalent to say that exactly one among Statement 1 or Statement 2 is true.

Step #3: Verifying strict duality when $0 \in \mathcal{Z}$ We first assume that the point 0 is a member of the uncertainty set $\mathcal{Z}$ which implies that $v \geq W0 = 0$. Now, let us for any fixed value t, consider the polyhedron $\mathcal{P}_t$ described as

$$\mathcal{P}_t := \{z \in \mathbb{R}^m \,|\, Wz \leq v \;\&\; x^T z \geq t\}\,.$$

Notice how any t for which $\mathcal{P}_t$ is empty provides an upper bound for Ψ while any t for which $\mathcal{P}_t$ is non-empty leads to a lower bound for Ψ, we must therefore have that

$$\Psi = \inf\{t \in \mathbb{R} \,|\, \mathcal{P}_t = \emptyset\}\,.$$



Yet based on Farkas lemma, exactly one of the following statement is true:

1. $\mathcal{P}_t \neq \emptyset$, i.e. there exists $z^+ \in \mathbb{R}^m$,$z^- \in \mathbb{R}^m$,$s \in \mathbb{R}$, and $y \in \mathbb{R}^m$ such that

$$\begin{bmatrix} x^T & -x^T & -1 & 0 \\ -W & W & 0 & -I \end{bmatrix} \begin{bmatrix} z^+ \\ z^- \\ s \\ y \end{bmatrix} = \begin{bmatrix} t \\ -v \end{bmatrix} \;\; \& \;\; \begin{bmatrix} z^+ \\ z^- \\ s \\ y \end{bmatrix} \geq 0$$

2. There exists $\gamma \in \mathbb{R}$ and $\lambda \in \mathbb{R}^s$, such that

$$\begin{bmatrix} x & -W^T \\ -x & W^T \\ -1 & 0 \\ 0 & -I \end{bmatrix} \begin{bmatrix} \gamma \\ \lambda \end{bmatrix} \leq 0 \;\; \& \;\; \begin{bmatrix} t & -v^T \end{bmatrix} \begin{bmatrix} \gamma \\ \lambda \end{bmatrix} > 0 \,.$$

Hence, the condition $\mathcal{P}_t = \emptyset$ can be equivalently replaced with the condition that there exists γ and λ such that

$$x\gamma - W^T\lambda = 0 \;\;\&\;\; \gamma \geq 0 \;\;\&\;\; \lambda \geq 0 \;\;\&\;\; t\gamma > v^T\lambda \,.$$

However, since we made the assumption that $0 \in \mathcal{Z}$ so that $v \geq 0$, these conditions can only be satisfied if $t\gamma > v^T\lambda \geq 0$ so that $\gamma \neq 0$. Because $\gamma > 0$, we can exploit the replacement of variable $\lambda' := \lambda/\gamma$ to get the equivalent condition

$$\exists \lambda' \geq 0,\; x = W^T\lambda' \;\;\&\;\; \lambda' \geq 0 \;\;\&\;\; t > v^T\lambda' \,.$$

We are left with the statement that:

$$\Psi = \inf\{t \in \mathbb{R} \,|\, \exists \lambda \in \mathbb{R}^s,\, \lambda \geq 0,\, x = W^T\lambda,\, t > v^T\lambda\} \,.$$

When studying closely the right-hand side of this equation, we realize that the infimum returns exactly the optimal value of problem (2.4).

Step #4: Verifying strict duality when $0 \notin \mathcal{Z}$ Now, let's generalize our previous conclusion to models where $0 \notin \mathcal{Z}$. This can be done by by reformulating problem (2.3) as

$$\Psi = \begin{array}{rl} \max_{\Delta} & x^T(z_0+\Delta) \\ \text{subject to} & W(z_0+\Delta) \leq v\,, \end{array}$$

where we let z be parametrized as $z := z_0 + \Delta$ with $\Delta \in \mathbb{R}^m$ and with z_0 as any feasible point in the set $\mathcal{Z}$. Next, some simple algebraic manipulations give us:

$$\Psi = x^T z_0 + \begin{array}{rl} \max_{\Delta} & x^T\Delta \\ \text{subject to} & W\Delta \leq v - Wz_0\,. \end{array}$$

In this form, it is clear that $\Delta := 0$ is feasible since the constraint then becomes $0 \leq v - Wz_0 \Leftrightarrow Wz_0 \leq v$ which is satisfied since $z_0 \in \mathcal{Z}$. We can therefore apply the result from Step #3 to obtain

$$\Psi = x^T z_0 + \begin{array}{rl} \min_{\lambda\in\mathbb{R}^s} & (v-Wz_0)^T\lambda \\ \text{subject to} & W^T\lambda = x \\ & \lambda \geq 0\,. \end{array}$$

By reintegrating the first term of the summation inside the minimization operation, we get an objective that looks like

$$(v-Wz_0)^T\lambda + x^T z_0 = v^T\lambda + (x - W^T\lambda)^T z_0 = v^T\lambda\,,$$

where the last equality is true for any feasible λ since those must satisfy $W^T\lambda = x$.

Step #5: Feasibility of problem (2.4) The final step of this proof consists in showing that Farkas lemma can once again be used to guarantee that the dual problem (2.4) is feasible. Well, indeed if it was not feasible then the lemma states that since condition 1 is not satisfied there must exist a Δ that satisfies $W\Delta \leq 0$ and $x^T\Delta > 0$. Looking back at the primal problem (2.3), one can now construct a solution $z_0 + \alpha\Delta$, with $\alpha \geq 0$ which necessarily satisfies the constraint

$$W(z_0+\alpha\Delta) = Wz_0 + \alpha W\Delta \leq Wz_0 \leq v$$

and allows to reach an arbitrarily large objective value

$$\lim_{\alpha\to\infty} x^T(z_0+\alpha\Delta) = \lim_{\alpha\to\infty} x^T z_0 + \alpha x^T\Delta = \infty\,,$$

since $x^T\Delta > 0$. But this is a contradiction since we assumed the polyhedron defined by $Wz \leq v$ was bounded thus that problem (2.3) cannot be unbounded. We can therefore conclude that the dual problem is feasible. In particular, there must exist a λ_0 that is feasible according to the dual problem (2.4).

This completes our proof of Theorem 2.3, namely that there are no gap between the optimal values of the primal and dual problems, and that the latter problem has a feasible solution.

□

Theorem 2.3 is important as it allows us to state that verifying the robust constraint (2.2) is equivalent to verifying whether the optimal value of problem (2.4) is lower or equal to $b - a^T x$. Yet, this verification is equivalent to identifying a feasible realization of λ for which $v^T\lambda \leq b - a^T x$. Hence, one should consider a λ that satisfies

$$\begin{aligned} & v^T\lambda \leq b - a^T x \\ & W^T\lambda = x \\ & \lambda \geq 0 \end{aligned}$$

to constitute a "certificate" that x actually satisfies

$$z^T x \leq b - a^T x \,,\, \forall\, z \in \mathcal{Z} \;.$$

This is easily verifiable, since for such a λ we have that

$$z^T x = z^T W^T \lambda \leq v^T \lambda \leq b - a^T x \quad \text{since } Wz \leq v \text{ and } \lambda \geq 0 \;.$$

The aspect that is more surprising is that searching through these types of certificates is sufficient, i.e. if no such λ certificate is found than x must be infeasible, in other words "it is not robust".

Example 2.5. : Consider the robust optimization problem:

$$\begin{aligned} \underset{x}{\text{maximize}} \quad & c^T x && (2.5a) \\ \text{subject to} \quad & (a+z)^T x \leq b \,,\, \forall\, z \in \mathcal{Z} && (2.5b) \\ & 0 \leq x \leq 1 \;, && (2.5c) \end{aligned}$$

where $\mathcal{Z}$ follows assumption 2.2. This problem is equivalent to solving

$$\begin{aligned} \underset{x,\lambda}{\text{maximize}} \quad & c^T x \\ \text{subject to} \quad & a^T x + v^T \lambda \leq b \\ & W^T \lambda = x \\ & \lambda \geq 0 \\ & 0 \leq x \leq 1 \;. \end{aligned}$$

In this problem, we are searching for both an x that achieves large objective value, and for a certificate λ that guarantees that x satisfies the robust constraint. Note that this problem has the same numerical structure as a problem in which we would consider z to be known, namely a linear program of slightly larger dimension.

In particular, say we are interested in the following "box" uncertainty set:

$$\mathcal{Z} := \{z \in \mathbb{R}^n \,|\, \bar{z}^- \leq z \leq \bar{z}^+\} ,$$

then we need to consider that

$$W = \begin{bmatrix} \boldsymbol{I} \\ -\boldsymbol{I} \end{bmatrix} \qquad , \qquad v = \begin{bmatrix} \bar{z}^+ \\ -\bar{z}^- \end{bmatrix} .$$

Hence, the reformulated problem will look like:

$$\begin{array}{ll} \underset{x,\lambda^+,\lambda^-}{\text{maximize}} & c^T x \\ \text{subject to} & a^T x + \bar{z}^{+T}\lambda^+ - \bar{z}^{-T}\lambda^- \leq b \\ & \lambda^+ - \lambda^- = x \\ & \lambda^+ \geq 0 \;\; \lambda^- \geq 0 \\ & 0 \leq x \leq 1 \,. \end{array}$$

This reformulation is implemented using RSOME in the following Colab file.

Remark 2.6. : Indeed, most modern application of robust optimization will involve solving a reformulation of the robust counterpart in which the decision vector x is optimized jointly with a set of certificates λ, issued from applying duality to each robust constraint. Each certificate that is returned guarantees that for the selected x the robust constraint associated to the certificate is satisfied.

Getting back at our robust linear program in general form, namely problem (2.1), the process of reformulating the problem would go as follows.

First, we would reformulate the problem in "epigraph" form (actually the hypograph form given that this is a maximization problem):

$$\begin{array}{ll} \underset{x,t}{\text{maximize}} & t \\ \text{subject to} & t - (z^T P_0^T x + q_0^T z + p_0^T x + r_0) \leq 0 \,,\, \forall\, z \in \mathcal{Z} \\ & z^T P_j^T x + p_j^T x \leq q_j^T z + r_j \,,\, \forall\, z \in \mathcal{Z}, \,,\, \forall\, j = 1, ..., J \,. \end{array}$$

For each of the constraint indexed by j, one would be interested in reformulating the following worst-case analysis.

$$\begin{array}{ll} \underset{z}{\text{maximize}} & x^T P_j z - q_j^T z \\ \text{subject to} & W z \leq v \end{array}$$

The dual problem would then take the form

$$\begin{array}{ll} \underset{\lambda}{\text{minimize}} & v^T\lambda \\ \text{subject to} & W^T\lambda = P_j^T x - q_j \\ & \lambda \geq 0 \end{array}$$

This requires us to identify a certificate $\lambda^{(j)}$ for each constraint indexed by j, that will satisfy the following properties

$$\begin{aligned} & p_j^T x + v^T\lambda^{(j)} \leq r_j \\ & W^T\lambda^{(j)} = P_j^T x - q_j \\ & \lambda^{(j)} \geq 0 \end{aligned}$$

After some manipulation of signs a similar conclusion is drawn for the epigraph constraint which is replaced with

$$\begin{aligned} & t + v^T\lambda^{(0)} - p_0^T x - r_0 \leq 0 \\ & W^T\lambda^{(0)} = -P_0^T x - q_0 \\ & \lambda^{(0)} \geq 0 \end{aligned}$$

Overall, we get the following theorem.

Theorem 2.7. : *The LP-RC problem, with a polyhedral $\mathcal{Z}$ described through $Wz \leq v$ (as in assumption 2.2), is equivalent to the following linear program*

$$\begin{array}{ll} \underset{x,\{\lambda^{(j)}\}_{j=0}^{J}}{\text{maximize}} & p_0^T x + r_0 - v^T\lambda^{(0)} \\ \text{subject to} & W^T\lambda^{(0)} = -P_0^T x - q_0 \\ & p_j^T x + v^T\lambda^{(j)} \leq r_j \,,\, \forall\, j = 1,\ldots,J \\ & W^T\lambda^{(j)} = P_j^T x - q_j \,,\, \forall\, j = 1,\ldots,J \\ & \lambda^{(j)} \geq 0 \,,\, \forall\, j = 0,\ldots,J \end{array}$$

where $\lambda^{(j)} \in \mathbb{R}^s$ are additional certificates that need to be optimized jointly with x.

2.2 General Uncertainty Sets

We now extend our discussion to the question of identifying tractable reformulations to problems that involve general convex uncertainty sets. To do so, for simplicity we look again at the special case of example 2.5 with an uncertainty set now defined as follows.

Assumption 2.8. : The uncertainty set $\mathcal{Z}$ is a bounded convex set defined by

$$\mathcal{Z} := \{z \in \mathbb{R}^m \,|\, f(z) \leq 0\,,\, Wz \leq v\}\,,$$

for some convex function $f(z)$. Moreover, there exists a realization $z_0 \in \mathcal{Z}$ that satisfies the nonlinear constraint strictly, namely $f(z_0) < 0$.

When we focus on the robust constraint we seek a way of validating for a fixed x the fact that the optimal value Ψ of the following optimization problem is smaller or equal to $b - a^T x$:

$$\begin{aligned} \Psi := \quad \max_{z} \quad & x^T z && \text{(2.6a)}\\ \text{subject to} \quad & f(z) \leq 0 && \text{(2.6b)}\\ & Wz \leq v\,. && \text{(2.6c)} \end{aligned}$$

Traditionally, in this case the dual problem is assembled based on Lagrangian duality (see Chapter 5 of [22]), which states that the optimal value of the problem above is equal to

$$\Psi = \max_{z} \min_{\gamma\geq 0, \lambda \geq 0} \mathcal{L}(z, \gamma, \lambda) := x^T z - \gamma f(z) + \lambda^T (v - Wz)\,.$$

The intuition behind this "constraint-free" reformulation is that for any fixed z, if z does not satisfy a constraint then the internal minimization problem can simply apply an arbitrarily large penalty through γ or λ so that $\mathcal{L}(z, \gamma, \lambda)$ reaches $-\infty$, which is considered equivalent to being infeasible.

As was the case for linear program, the optimal value of problem (2.6) or equivalently of the $\max\min \mathcal{L}(z, \gamma, \lambda)$ is bounded above by

$$\Upsilon^* := \min_{\gamma\geq 0, \lambda \geq 0} \max_{z} x^T z - \gamma f(z) + \lambda^T (v - Wz) \geq \Psi\,.$$

This can be again validated by considering that for any $\gamma \geq 0$ and $\lambda \geq 0$, evaluating $\mathcal{L}(z, \gamma, \lambda)$ at a feasible z will return a larger amount than $x^T z$. Note that we once again associate the notation Υ^* to the optimal value of the upper bounding problem.

We can simply manipulate the upper bounding problem to obtain

$$\Upsilon^* = \min_{\gamma\geq 0, \lambda \geq 0} v^T \lambda + \max_{z} \left((x - W^T\lambda)^T z - \gamma f(z)\right)\,,$$

which can be redefined as the optimal value of the following dual problem

$$\begin{aligned} \underset{\lambda,\gamma}{\text{minimize}} \quad & v^T\lambda + f_p^*(x - W^T\lambda, \gamma) && \text{(2.7a)}\\ \text{subject to} \quad & \lambda \geq 0 && \text{(2.7b)}\\ & \gamma \geq 0\,. && \text{(2.7c)} \end{aligned}$$

where $f_p^*(y, \gamma) = \max_z\ y^T z - \gamma f(z)$ and can somehow be interpreted as the adjoint (or perspective) function of the conjugate function of $f(z)$.

The question is now whether $\Upsilon^* = \Psi$ or not. To this end, we will employ a constraint qualification known as Slater's condition.

Lemma 2.9. :*(Strong duality see section 5.2.3 of [22]) Given a convex optimization problem of the form:*

$$\begin{aligned} \underset{z}{\text{maximize}} \quad & f_0(z) \\ \text{subject to} \quad & f_j(z) \leq 0\,,\ \forall\, j = 1, ..., J\ , \end{aligned}$$

where $f_0(z)$ is a concave function of z, $f_j(z)$ are affine functions of z for $j = 1, ..., k$, and $f_j(z)$ are convex functions of z for $j = k+1, ..., J$. If there exists some z_0 that satisfies the following conditions :

(Slater's condition $\quad f_j(z_0) \leq 0\,,\ \forall\, j = 1, \ldots, k \quad \& \quad f_j(z_0) < 0\,,\ \forall\, j = k+1, \ldots, J$

then the optimal value of the dual problem, which can be expressed as

$$\begin{aligned} \underset{\lambda}{\text{minimize}} \quad & \sup_z\,(f_0(z) - \sum_{j=1}^{J} \lambda_j f_j(z)) \\ \text{subject to} \quad & \lambda \geq 0\ , \end{aligned}$$

is equal to the optimal value of the primal problem.

One can employ the above lemma in the context of problem (2.6) where $Wz \leq v$ can be expressed as a series of constraints $w_i^T z \leq v_i$, and where there would be a single non-linear convex constraint $f(z) \leq 0$. In this case, the dual problem described in the theorem would be equivalent to problem (2.7). Moreover, the Slater's condition reduces to our assumption 2.8, namely that there exists a z_0 that satisfies $Wz \leq v$ and satisfies the nonlinear constraint with strict inequality, i.e. $f(z_0) < 0$.

To complete the tractable reformulation of the robust counterpart problem presented in example 2.5, one starts by replacing the robust constraint (2.5b) with its equivalent representation

$$\begin{aligned} \underset{x,\lambda,\gamma}{\text{maximize}} \quad & c^T x & \text{(2.8a)} \\ \text{subject to} \quad & v^T \lambda + f_p^*(x - W^T \lambda, \gamma) \leq b - a^T x & \text{(2.8b)} \\ & \lambda \geq 0, \ \ \gamma \geq 0 & \text{(2.8c)} \end{aligned}$$

Note that to consider the optimization model above tractable, one needs to have an analytical expression that evaluates $f_p^*(y, \gamma)$.

We generalize this "tractable" reformulation to the the general LP-RC model in the following theorem.

Theorem 2.10. : *The LP-RC problem with an uncertainty set $\mathcal{Z}$ that satisfies as-*

sumption 2.8 is equivalent to the following linear program

$$
\begin{aligned}
\underset{x,\{\lambda^{(j)}\}_{j=0}^{J},\{\gamma^{(j)}\}_{j=0}^{J}}{\text{maximize}} \quad & p_0^T x + r_0 - v^T\lambda^{(0)} - f_p^*(-P_0^T x - q_0 - W^T\lambda^{(0)}, \gamma^{(0)}) && (2.9a)\\
\text{subject to} \quad & v^T\lambda^{(j)} + f_p^*(P_j^T x + q_j - W^T\lambda^{(j)}, \gamma^{(j)}) + p_j^T x \le r_j\,,\ \forall\, j = 1,\ldots,J && (2.9b)\\
& \lambda^{(j)} \ge 0\,,\ \forall\, j = 0,\ldots,J && (2.9c)\\
& \gamma^{(j)} \ge 0\,,\ \forall\, j = 0,\ldots,J && (2.9d)\\
& && (2.9e)
\end{aligned}
$$

where $\gamma^{(j)} \in \mathbb{R}$ *and* $\lambda^{(j)} \in \mathbb{R}^s$ *are additional certificates that need to be optimized jointly with* x*.*

2.2.1 Example: Ellipsoidal Uncertainty

Let us assume that one is interested in the robust counterpart of example 2.5 with $\mathcal{Z}$ taking the following shape:

$$\mathcal{Z} := \{z \in \mathbb{R}^m \,|\, z^T\Sigma^{-1}z \le 1\}$$

with $\Sigma \in \mathbb{R}^{m\times m}$ such that $\Sigma \succ 0$, meaning that it is a positive definite matrix.

We are interested in identifying a tractable form for the reformulation presented in equation (2.8) in the specific context where $f(z) := z^T\Sigma^{-1}z - 1$, $W = 0$, and $v = 0$. Namely, we need to simplify the problem:

$$
\begin{aligned}
\underset{x,\gamma}{\text{maximize}} \quad & c^T x\\
\text{subject to} \quad & f_p^*(x,\gamma) \le b - a^T x\\
& \gamma \ge 0
\end{aligned}
$$

To do so, one needs to identify an analytical expression that evaluates $f_p^*(y,\gamma)$. This can be done by identifying a closed form solution of the problem

$$f_p^*(y,\gamma) := \max_z y^T z - \gamma(z^T\Sigma^{-1}z - 1)\,.$$

In the case that $\gamma = 0$, $f_p^*(y,0)$ reduces to the following function

$$f_p^*(y,0) := \begin{cases} 0 & \text{if } y = 0\\ \infty & \text{otherwise}\end{cases}$$

Thus the robust constraint (2.8b) is equivalent to

$$0 \le b - a^T x \qquad\qquad \& \qquad\qquad x = 0\,.$$

Otherwise, if $\gamma > 0$ then since the function that is maximized is concave and differentiable everywhere. One can identify the optimal solution by setting the first derivatives to zero.

$$\nabla_z\left(y^T z - \gamma(z^T\Sigma^{-1}z - 1)\right) = y - 2\gamma\Sigma^{-1}z^* = 0$$

Hence, $z^* = (2\gamma)^{-1}\Sigma y$, which indicates that

$$f_p^*(y,\gamma) = \frac{1}{2\gamma}y^T\Sigma y - \frac{1}{4\gamma}y^T\Sigma y + \gamma = \frac{1}{4\gamma}y^T\Sigma y + \gamma$$

Thus, the robust constraint (2.8b) can be reduced to

$$\frac{1}{4\gamma}x^T\Sigma x + \gamma \leq b - a^T x$$

Since the limit of this constraint as $\gamma \to 0$ is equivalent to the constraint we obtained for the case $\gamma = 0$, it is equivalent to state the reformulated optimization problem in terms of:

$$\begin{aligned} \underset{x,\gamma}{\text{maximize}} \quad & c^T x \\ \text{subject to} \quad & \frac{1}{4\gamma}x^T\Sigma x + \gamma \leq b - a^T x \\ & \gamma \geq 0 \;, \end{aligned}$$

where we interpret y^2/γ as equal to 0 when both $\gamma = 0$ and $y = 0$, and as equal to $infty$ when $y \neq \gamma = 0$ (also known as the recession function of y^2). Although this reformulation is a convex optimization problem, it is worth attempting to simplifying its form. To do so, we can observe that it might be possible to perform the optimization in γ analytically, considering that the problem is equivalent to

$$\begin{aligned} \underset{x}{\text{maximize}} \quad & c^T x \\ \text{subject to} \quad & \inf_{\gamma \geq 0} \frac{1}{4\gamma}x^T\Sigma x + \gamma \leq b - a^T x \;. \end{aligned}$$

Looking first at the case where $x = 0$, we get that $\gamma^* = 0$. Otherwise, we can set to zero the first derivative with respect to γ of $\frac{1}{4\gamma}x^T\Sigma x + \gamma$, an expression that is both convex and differentiable everywhere (with the exception of $\gamma = 0$ where it is infinite). We obtain that

$$-\frac{1}{4\gamma^2}x^T\Sigma x + 1 = 0$$

From which we conclude that $\gamma^* = (1/2)(x^T\Sigma x)^{1/2}$, which also retrieves the optimal γ when $x = 0$. After replacing this expression in our problem, we obtain

$$\underset{x}{\text{maximize}} \quad c^T x \tag{2.10a}$$

$$\text{subject to} \quad a^T x + (x^T\Sigma x)^{1/2} \leq b \;. \tag{2.10b}$$

This convex optimization problem is known as a second-order cone program for which there are a number of solvers readily available.

Remark 2.11. : It is worth highlighting the fact that reformulation presented in (2.10) could have been obtained by exploiting Cauchy-Schwartz inequality which states that $a^T b \leq \|a\|_2 \|b\|_2$ with equality achieved when a and b are colinear (i.e. parallel). Indeed, by this inequality we have that:

$$x^T z \;=\; x^T \Sigma^{1/2} \Sigma^{-1/2} z \;=\; (\Sigma^{1/2} x)^T (\Sigma^{-1/2} z) \;\leq\; \|\Sigma^{1/2} x\|_2 \|\Sigma^{-1/2} z\|_2 \;\leq\; \|\Sigma^{1/2} x\|_2 \,,$$

for all $z \in \mathcal{Z}$ since the constraint that describes this set imposes that $\|\Sigma^{-1/2} z\|_2^2 \leq 1$. Note that equality is achieved in the above expression when $\Sigma^{-1/2} z = \Sigma^{1/2} x / \|\Sigma^{1/2} x\|_2$ hence when $z := \Sigma x / \|\Sigma^{1/2} x\|_2 \in \mathcal{Z}$. In this case, we can confirm that

$$x^T z = x^T \Sigma x / \|\Sigma^{1/2} x\|_2 = \|\Sigma^{1/2} x\|_2 \,.$$

This implies that

$$\max_{z \in \mathcal{Z}} \; x^T z = \|\Sigma^{1/2} x\|_2 \,,$$

which confirms once again the result presented in equation (2.10).

2.3 Exercises

For each of the robust counterparts models presented below, derive a tractable linear programming reformulation and implement in RSOME (using Google Colab) both in its reduced and unreduced form. In your implementations, assume that $n = 150$ and that

$$c_i := 0.15 + i \frac{0.05}{n} \qquad a_i := 0 \,, \quad A_{ij} := \begin{cases} \frac{0.05}{450} \sqrt{2in(n+1)} & \text{if i=j} \\ 0 & \text{otherwise} \end{cases} \qquad b := 0.02 \,.$$

Also, for exercises 2.1 and 2.3 consider that each $\bar{z}_i \in \mathbb{R}^n$, $i = 1, \ldots, 150$ is composed as follows:

$$\bar{z}_i = \frac{0.05}{450} \sqrt{2in(n+1)} e_i \,,$$

where each e_i refers to the i-th column of the identity matrix, and that $\Gamma = 4$ in exercise 2.2 while $\alpha = 0.5$ in exercise 2.3.

Exercise 2.1. ***Convex Hull Set***
Derive a finite dimensional LP formulation for the robust counterpart of the following problem:

$$\begin{aligned} \text{maximize} \quad & c^T x \\ \text{subject to} \quad & (a + z)^T x \leq b \,, \, \forall \, z \in \mathcal{Z} \\ & 0 \leq x \leq 1 \\ & \sum_{i=1}^{n} x_i \leq 1 \,, \end{aligned}$$

where

$$\mathcal{Z} := \left\{ z \in \mathbb{R}^n \;\middle|\; \exists \theta \in \mathbb{R}^K,\, z = \sum_{i=1}^K \theta_i \bar{z}_i,\, \theta \geq 0,\, \sum_{i=1}^K \theta_i = 1 \right\} ,$$

for some predefined set of scenarios $\{\bar{z}_1, \bar{z}_2, \ldots, \bar{z}_K\}$?

Exercise 2.2. ***Budgeted Uncertainty Set***

Derive a finite dimensional LP formulation for the robust counterpart of the following problem:

$$\begin{array}{ll} \text{maximize} & c^T x \\ \text{subject to} & (a+z)^T x \leq b\,,\, \forall\, z \in \mathcal{Z}(\Gamma) \\ & 0 \leq x \leq 1 \\ & \sum_{i=1}^n x_i \leq 1\;, \end{array}$$

where $\mathcal{Z}(\Gamma)$ is the budgeted uncertainty set parametrized by $\Gamma \geq 0$ and defined as

$$\mathcal{Z}(\Gamma) := \left\{ z \in \mathbb{R}^m \;\middle|\; -1 \leq z \leq 1,\, \sum_i |z_i| \leq \Gamma \right\} ?$$

Note that to use the material presented in this section, you might need to employ the following equivalent representation:

$$\mathcal{Z}(\Gamma) := \left\{ z \in \mathbb{R}^m \;\middle|\; \exists s \in \mathbb{R}^m, -s_i \leq z_i \leq s_i \forall\, i = 1, \ldots, m,\;\; s \leq 1,\; \sum_i s_i \leq \Gamma \right\} .$$

or

$$\mathcal{Z}(\Gamma) := \left\{ z \in \mathbb{R}^m \;\middle|\; \exists \lambda^+, \lambda^- \mathbb{R}^m, z = \lambda^+ - \lambda^-,\; \lambda^+ \geq 0,\; \lambda^- \geq 0,\; \lambda^+ + \lambda^- \leq 1,\; \sum_i \lambda_i^+ + \lambda_i^- \leq \Gamma \right\} .$$

Exercise 2.3. ***CVaR Uncertainty Set***

Derive a finite dimensional LP formulation for the robust counterpart of the following problem:

$$\begin{array}{ll} \text{maximize} & c^T x \\ \text{subject to} & (a+z)^T x \leq b\,,\, \forall\, z \in \mathcal{Z} \\ & 0 \leq x \leq 1 \\ & \sum_{i=1}^n x_i \leq 1\;, \end{array}$$

where $\mathcal{Z}$ is the following uncertainty set parametrized with $\alpha \in]0,1]$ (the smaller the more robust) and defined as

$$\mathcal{Z} := \left\{ z \in \mathbb{R}^n \;\middle|\; \exists \theta \in \mathbb{R}^K,\, z = \sum_{i=1}^K \theta_i \bar{z}_i,\, \theta \geq 0,\, \sum_{i=1}^K \theta_i = 1,\, \theta \leq \frac{1}{K\alpha} \right\} ,$$

Chapter 3

Data-driven Uncertainty Set Design

The fact that the robust optimization perspective requires one to characterize uncertainty through the use of uncertainty sets is both a strength and a limitation of the method. On the positive side, the idea of an uncertainty set is somewhat easier to visualize than the notion of a distribution. As we indicated in the last chapter, it is also often easier to involve in an optimization process because of the property that there exists simply structured certificates that allows us to verify feasibility[1]. On the flip side, potentially because the field is relatively young, the main issue that is often faced by practitioners is the question of how to create an uncertainty set that captures accurately the knowledge and risks that are present.

Up to this date, the most documented methods of constructing uncertainty sets are motivated by the idea of an underlying distribution for the uncertain vector Z, and robustness can be perceived in terms of protecting the decision maker from scenarios that are drawn from such a distribution. In what follows, we explain important connections that can be made with chance constraints and to the concept of coherent risk measures.

3.1 Chance Constraint Approximation

In 1959, Charnes and Cooper in [23] introduced a concept that became very popular in the field of stochastic programming. Their idea was that when a constraint is affected by uncertainty, one should try to impose that the constraint be satisfied with high probability. Namely, a constraint of the type:

$$a(z)^T x \leq b(z)$$

should become the following "chance constraint"

$$\mathbb{P}(a(Z)^T x \leq b(Z)) \geq 1 - \epsilon$$

[1]In contrast, verifying the feasibility or the objective value of a function that involves a continuous distribution requires one to perform integration on high dimensional space which is generally hard to perform.

with $\epsilon > 0$ characterizing the amount of probability with which we are comfortable that the constraint might not be respected. The smaller ϵ is the more protection one obtains in term of ensuring that the constraint be met.

Another related concept that was introduced by financial institutions and is now a universal standard as made official in the Basel II Accord is the Value-at-Risk (VaR). Indeed, it is said that the value-at-risk of an uncertain expense is an amount for which we are assured with high probability that the expense will not reach. Alternatively, for an uncertain revenue, the VaR expresses (the negative of) an amount that we have high confidence the true revenue will surpass.

Definition 3.1. : Mathematically, for a revenue calculated as $c(z)^T x + d(z)$, it can be defined as

$$\mathrm{VaR}_{1-\epsilon}(c(Z)^T x + d(Z)) := -\sup\{y \in \mathbb{R} \,|\, \mathbb{P}(c(Z)^T x + d(Z) \geq y) \geq 1 - \epsilon\}$$

Hence, when maximizing an uncertain revenue one can formulate a VaR minimization problem as follows:

$$\begin{aligned} \underset{x,y}{\text{minimize}} \quad & -y \\ \text{subject to} \quad & \mathbb{P}(c(Z)^T x + d(Z) \geq y) \geq 1 - \epsilon \\ & x \in \mathcal{X} \,. \end{aligned}$$

One should note here that, given a candidate solution x, both the chance constraint and the value-at-risk formulation involve verifying whether a constraint is satisfied with high probability or not. Unfortunately, it is generally intractable to do so, unless the distribution has a very simple form (see section 3.1.2).

It is not surprising that the community has often interpreted robust constraints in terms of chance constraints. Indeed, the fact that a solution is said to be “robust” is understood by many as a statement about how likely it is to perform well. In fact, some might argue that a robust constraint should be interpreted as $\mathbb{P}(a(Z)^T x \leq b(Z)) = 100\%$. Yet, for many distributions assumptions, this statement is useless. Take for instance, the case where we would want $\mathbb{P}(Z^T x \leq 1) = 100\%$ under the hypothesis that the terms of Z be normally distributed. In this case, the constraint would plainly require that $x = 0$ since otherwise there is always some chance that the constraint would be violated. One must therefore be a bit more realistic and one option is to consider imposing that $\mathbb{P}(a(Z)^T x \leq b(Z)) \geq 1 - \epsilon$ for some $\epsilon > 0$.

The most straightforward way of approximating a chance constraint with a robust constraint is to calibrate the uncertainty set such that it is large enough to cover a set of realizations for Z that has more than $1-\epsilon$ chances of containing the true realization. In other words, one might employ robust optimization as suggested by the following theorem.

Theorem 3.2. : *Given some $\epsilon > 0$ and some random vector Z distributed according to F, let $\mathcal{Z}$ be a set such that*

$$\mathbb{P}(Z \in \mathcal{Z}) \geq 1 - \epsilon \,,$$

then one has the guarantee that any x satisfying the robust constraint

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

will also satisfy the following chance constraint

$$\mathbb{P}(a(Z)^T x \leq b(Z)) \geq 1 - \epsilon\ .$$

Proof. Let x satisfy the robust constraint

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

then the probability

$$\mathbb{P}(a(Z)^T x \leq b(Z)) \geq \mathbb{P}(Z \in \mathcal{Z}) \geq 1 - \epsilon\ .$$

□

Note that the theorem does not present an "if and only if" statement, which means that replacing a chance constraint by a robust constraint has the potential of (and has often the effect of) reducing the feasible region. We are left with the following guideline for constructing the uncertainty set involved in a robust optimization model.

Corollary 3.3. : *Given some $\epsilon > 0$ and some random vector Z distributed according to F, let $\mathcal{Z}$ be a set such that*

$$\mathbb{P}(Z \in \mathcal{Z}) \geq 1 - \epsilon\ ,$$

then the LP-RC optimization problem (2.1) *is a conservative approximation of the stochastic program*

$$\begin{array}{rl} \underset{x}{\text{minimize}} & VaR_{1-\epsilon}(Z^T P_0^T x + q_0^T Z + p_0^T x + r_0) \\ \text{subject to} & \mathbb{P}(Z^T P_j^T x + p_j^T x \leq q_j^T Z + r_j) \geq 1 - \epsilon\,,\ \forall\, j = 1, ..., J\ , \end{array}$$

where $VaR_{1-\epsilon}(\cdot)$ is as defined in definition 3.1. Specifically, by conservative approximation we mean that an optimal solution to the LP-RC problem will be feasible according to the above stochastic program where it will achieve an objective value that is lower than what was established by the LP-RC optimization model.

We present a few examples to illustrate how this result might be employed.

3.1.1 Example: Data-driven return vector uncertainty

You are given a set of historical monthly returns of 10 stocks for year 2000 - 2009, and are asked to approximate the following "value-at-risk" problem:

$$\begin{array}{ll}
\underset{x,y}{\text{minimize}} & -y \\
\text{subject to} & \mathbb{P}(r^T x \geq y) \geq 1-\epsilon \\
& \sum_{i=1}^n x_i = 1 \\
& x \geq 0\,,
\end{array}$$

where $\epsilon = 5\%$ and the distribution of r is considered as the empirical distribution of the monthly stock returns over the whole period of 2000-2009, in other words, any monthly return vector observed in this period is as likely to occur.

Question: Identify a robust optimization problem that can conservatively approximate this optimization model.

Our answer: Let's consider the following approximation to the value-at-risk problem described above:

$$\underset{x,y}{\text{minimize}} \quad -y \tag{3.1a}$$

$$\text{subject to} \quad r^T x \geq y\,,\ \forall\, r \in \mathcal{U} \tag{3.1b}$$

$$\sum_{i=1}^n x_i = 1 \tag{3.1c}$$

$$x \geq 0\,, \tag{3.1d}$$

where we will use the uncertainty set:

$$\mathcal{U}(r_0, \gamma) := \{r \in \mathbb{R}^{10} \,|\, \|r - r_0\|_2 \leq \gamma\}\,,$$

which would need to be parametrized such that $\mathbb{P}(r \in \mathcal{U}(r_0,\gamma)) \geq 1-\epsilon$. We choose to employ the most natural version of r_0, namely $r_0 = \mathbb{E}[r]$, which centers the uncertainty set at the expected return values. We are left with calibrating γ. This can be done by following the procedure:

1. For each monthly return r_k in the historical data:
   (a) Compute $\gamma_k := \|r_k - r_0\|_2$ which indicates how large γ needed to be for $r_k \in \mathcal{U}(r_0,\gamma)$.
2. Choose as γ the $(1-\epsilon)\times K$-th largest γ in the list $\{\gamma_k\}_{k=1}^K$ where K is the number of months in the study.

Our implementation in Python (see Google Colab) returns a policy which achieved a 95% value-at-risk guaranteed below 0.210 (i.e. awith 95% chances of loosing less than 21%), the robust policy actually achieved an in-sample VaR of 0.063 (i.e. loose less than 6.3%) for the historical data 2000-2009, and an out-of-sample VaR of 0.066 (i.e. lose less than 6.6%) in the years 2010-2014.

3.1.2 Example: Return vector uncertainty under normal distribution

Let's have a look at the reformulation for the robust counterpart presented in example 3.1.1. As presented in section 2.2.1, the tractable reformulation takes the shape:

$$\begin{aligned} \underset{x,y}{\text{minimize}} \quad & -y \\ \text{subject to} \quad & r_0^T x - \gamma \|x\|_2 \geq y \\ & \sum_{i=1}^n x_i = 1 \\ & x \geq 0\,. \end{aligned}$$

It is interesting to be aware of the following connection there is to make with the chance constrained problem that would be obtained if r was assumed to be normally distributed. Note that if r is normally distributed with mean r_0 and covariance matrix $\sigma^2 I$, then it is well known that $r^T x$ is also normally distributed with mean $r_0^T x$ and variance $\sigma^2 \|x\|_2^2$. Hence the constraint that states

$$\mathbb{P}(r^T x \geq y) \geq 1 - \epsilon$$

is equivalent to requiring that

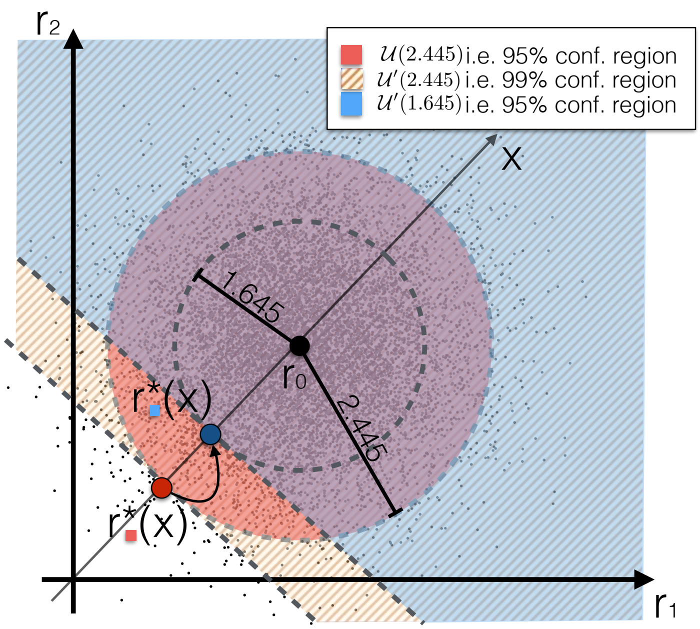
$$r_0^T x - \Phi^{-1}(1-\epsilon)\sigma \|x\|_2 \geq y\ ,$$

where $\Phi^{-1}(1-\epsilon)$ is the inverse function of the cumulative density function of the standard normal distribution.

This analysis allows us to conclude that when r is normally distributed, one can select γ such that the robust counterpart model present in (3.1) is exactly equivalent to the chance constraint. In particular, γ should be equal to $\Phi^{-1}(1-\epsilon)\sigma$. Unfortunately, when calibrating the size of the ellipsoidal set so that it contains the realized Z with some given probability, one would not obtain the optimal γ but a rather larger one, namely the calibrated γ would take a value of 2.445 instead of the "optimal" 1.645 when $\epsilon = 5\%$ and $\sigma = 1$.

This difference can be explained by the fact that for any given x, if $\inf_{r \in \mathcal{U}(r_0,\gamma)} r^T x \geq y$ then it is also the case that $\inf_{r \in \mathcal{U}'(r_0,\gamma)} r^T x \geq y$ where $\mathcal{U}'(r_0,\gamma) := \{r' \in \mathbb{R}^m \,|\, r'^T x \geq \inf_{r \in \mathcal{U}(r_0,\gamma)} r^T x\}$. In fact, $\mathcal{U}'(r_0,\gamma)$ is a half-space that has much larger volume and actually covers $\mathcal{U}(r_0,\gamma)$ thus providing a protection that is more than necessary (see the figure below for an illustration of this over-protection). This effect is often interpreted by practitioners as the over-conservatism of the robust optimization framework. In section 3.3, we will explain how one might search through the spectrum of sizes for an uncertainty set until he has identified a size for which the solution satisfies the level of protection that is needed. Such a search in this example could potentially identify $\gamma = 1.645$ as the right size to use.

Optimal 95% confidence region



3.1.3 Example: Data-driven expected return uncertainty

You are given a set of historical monthly returns of 10 stocks for years 2000 - 2009, and are asked to identify a sphere that will include the expected monthly return over the months of 2010. This uncertainty set is useful as it would allow to identify a portfolio of the ten stocks that maximizes the robust counterpart of the following model:

$$\begin{aligned} \underset{x}{\text{maximize}} \quad & \mu^T x \\ \text{subject to} \quad & \sum_{i=1}^{n} x_i = 1 \\ & x \geq 0\,, \end{aligned}$$

where μ is the vector of expected returns for each stock. One could estimate this vector μ using the average monthly return achieved in 2009; however, there are many reasons to believe that this estimate would be inaccurate and that we need to account for a confidence region.

The robust counterpart is therefore interesting and could be expressed as

$$\begin{array}{ll} \underset{x,y}{\text{maximize}} & y \\ \text{subject to} & \mu^T x \geq y \,, \forall \, \mu \in \mathcal{U} \end{array}$$

where $\mathcal{U}$ would capture the confidence region of μ. In particular, we will use the following form:

$$\mathcal{U}(\mu_0, \gamma) := \{\mu \in \mathbb{R}^{10} \mid \|\mu - \mu_0\|_2 \leq \gamma\} \,,$$

where μ_0 is the average monthly return achieved in 2009.

Question: How should we calibrate γ using the historical data from 2000-2009?

Our answer: In order to identify a γ such that we have high confidence that $\mu \in \mathcal{U}(\mu_0, \gamma)$, we need a distribution for μ. Unfortunately, confidence regions proposed from statistical theory are hard to employ here because :1) the number of samples used for estimation is small (12 samples), 2) we have good reasons to believe that the returns of 2010 are not distributed according the returns in 2009. For this reason, in [28] we employed a method that is based on "bootstrapping" to characterize what the joint distribution of (μ_0, μ) might be and choose γ such that, for some $\epsilon > 0$:

$$\mathbb{P}(\mu \in \mathcal{U}(\mu_0, \gamma)) \geq 1 - \epsilon \,.$$

Effectively, the idea was to quantify, based on the historical data, the likelihood that the average return in a given year is some distance away from the average return in the following year.

Here is the procedure that can achieve this purpose:

1. For any year k in the historical data:

   (a) Compute the average monthly return μ_k in year k

   (b) Compute the average monthly return μ_{k+1} in year $k+1$

   (c) Compute $\gamma_k := \|\mu_k - \mu_{k+1}\|_2$ which indicates how large γ needed to be in year k for $\mu_{k+1} \in \mathcal{U}(\mu_k, \gamma)$.

2. Choose as γ the $(1-\epsilon) \times K$-th largest γ in the list $\{\gamma_k\}_{k=1}^K$ where K is the number of years in the study.

The implementation that is presented in the Python code on Google Colab suggests using $\gamma = 0.367$. When we verified in 2010, we observed that $\|\mu - \mu_0\|_2$ had actually been equal to $0.094 < 0.367$, hence that $\mathcal{U}(\mu_0, 0.367)$ was indeed large enough.

3.2 Ambiguous Chance Constraint Approximation

It appears to be in [13] that the authors identify for the first time a connection between robust constraints and a form of chance constraint referred as "ambiguous chance constraint". Indeed, one of the main strengths of robust optimization is the idea that the approach is distribution-free. For this reason, it could be more interesting to relate the robust optimization formulation to a form of chance constraint that is distribution-free. Namely, many authors have considered the following assumption.

Assumption 3.4. : Let $Z \in \mathbb{R}^m$ be a random vector for which the distribution is not known, yet what is known of the random vector is that all Z_i's are independent from each other and that each of them is symmetrically distributed on the interval $[-1, 1]$.

Theorem 3.5. : *Given some $\epsilon > 0$ and some random vector Z that satisfies assumption 3.4, one has the guarantee that any x satisfying the robust constraint*

$$a(z)^T x \leq b(z)\,,\ \forall\, z \in \mathcal{Z}_{ell}(\gamma)\ ,$$

where

$$\mathcal{Z}_{ell}(\gamma) := \{z \in \mathbb{R}^m \mid \|z\|_2 \leq \gamma\}$$

and $\gamma := \sqrt{2\ln(1/\epsilon)}$ is guaranteed to satisfy the following chance constraint

$$\mathbb{P}(a(Z)^T x \leq b(Z)) \geq 1 - \epsilon\ .$$

even though the distribution of Z is not known.

In order to prove this theorem one must first be aware of the following lemma.

Lemma 3.6. : *Let $Z \in \mathbb{R}^m$ be a random vector with independent symmetrically distributed entries on the interval $[-1,\ 1]$ and let $a \in \mathbb{R}^m$ be such that $\|a\|_2 = 1$, then for every $\gamma > 0$, one has*

$$\mathbb{P}(a^T Z > \gamma) \leq \exp(-\gamma^2/2)\ .$$

Proof. The proof goes as follows. Since $\exp(\cdot)$ is a strictly increasing function, one can state that

$$\mathbb{P}(a^T Z > \gamma) \leq \mathbb{P}(a^T Z \geq \gamma) = P(\gamma a^T Z \geq \gamma^2) = P(\exp(\gamma a^T Z) \geq \exp(\gamma^2))\ .$$

But then, by Markov inequality which states that $\mathbb{P}(Y \geq \alpha) \leq \mathbb{E}[Y]/\alpha$ if Y is a positive random variable, we can infer that

$$\mathbb{P}(\exp(\gamma a^T Z) \geq \exp(\gamma^2)) \leq \frac{\mathbb{E}[\exp(\gamma a^T Z)]}{\exp(\gamma^2)} = \exp(-\gamma^2) \prod_j \mathbb{E}[\exp(\gamma a_j Z_j)]\ ,$$

where the last inequality follows from the fact that the Z_j are independent. The question is now to confirm that $\mathbb{E}[\exp(\gamma a_j Z_j)] \leq \exp(\gamma^2 a_j^2/2)$. If it was the case than we would know that

$$\begin{aligned} \exp(-\gamma^2)\prod_j \mathbb{E}[\exp(\gamma a_j Z_j)] &\leq \exp(-\gamma^2)\exp(\gamma^2\sum_j a_j^2/2) \\ &= \exp(-\gamma^2+\gamma^2/2) = \exp(-\gamma^2/2)\ , \end{aligned}$$

where we used the fact that $\|a\|_2^2 = 1$.

We are left with showing that $\mathbb{E}[\exp(\gamma a_j Z_j)] \leq \exp(\gamma^2 a_j^2/2)$. To do so, we first show that $\mathbb{E}[\exp(\gamma a_j Z_j)] \leq (1/2)(\exp(\gamma|a_j|) + \exp(-\gamma|a_j|))$ this is because among all symmetric distributions on the $[-1, 1]$ interval, the two-points distribution that puts half of the weight on each extremities leads to the highest expected value. By contradiction, assume that there is some mass p at $\delta < 1$ well then by symmetry there is necessary also some mass p at $-\delta$. Both masses could be pushed to the extremities to increase the expected value since, letting $g(\delta) := p(\exp(\gamma a_j\delta) + \exp(-\gamma a_j\delta))$ one can confirm that

$$g(\delta) = p(\exp(\gamma|a_j|\delta) + \exp(-\gamma|a_j|\delta))\ ,$$

and that

$$g'(\delta) = p\gamma|a_j|(\exp(\gamma|a_j|\delta) - \exp(-\gamma|a_j|\delta))\ ,$$

which is minimal when $\delta = 0$ so that $g'(\delta) \geq 0$. We can conclude that increasing δ increases $g(\delta)$.

The final step is to prove that $(1/2)(\exp(\gamma|a_j|)+\exp(-\gamma|a_j|)) \leq \exp(\gamma^2 a_j^2/2)$. This can be done by analysing the Taylor series expansion

$$\begin{aligned} (1/2)(\exp(\gamma|a_j|) + \exp(-\gamma|a_j|)) &= (1/2)(\sum_{n=0}^{\infty}\frac{(\gamma|a_j|)^n}{n!} + \sum_{n=0}^{\infty}\frac{(-\gamma|a_j|)^n}{n!}) \\ &= \sum_{n=0}^{\infty}\frac{(\gamma|a_j|)^{2n}}{(2n)!} \\ &\leq \sum_{n=0}^{\infty}\frac{(\gamma^2 a_j^2/2)^n}{n!} = \exp(\gamma^2 a_j^2/2)\ , \end{aligned}$$

where we used the fact that $(2n)! \geq 2^n n!$.[2] □

We are now able to prove theorem 3.5.

[2]Shown by induction, since it is obviously true for $n = 1$, we then look whether when it is true for n, it is also true for $n + 1$. Specifically,

$$(2(n+1))! = (2n+2)(2n+1)(2n)! \geq (2n+2)(2n+1)2^n n! = (2n+2)(2n+1)/(n+1)2^n(n+1)! \geq 2^{n+1}(n+1)!\ .$$

Proof. We first represent the constraint $a(z)^T x \leq b(z)$ in terms of the relation to z:

$$a(x)^T z \leq b(x)$$

where we overloaded the notation for $a(\cdot)$ and $b(\cdot)$ to represent the following

$$x^T P z - q^T z \leq r - p^T x \; ,$$

namely with:

$$a(x) := P^T x - q \qquad \& \qquad b(x) := r - p^T x \; .$$

With this new notation, we can express the robust constraint as

$$a(x)^T z \leq b(x) \, , \, \forall \, z \in \mathcal{Z} \; ,$$

which is equivalent to

$$\gamma \|a(x)\|_2 \leq b(x) \; .$$

We refer the reader to example 2.2.1 for some details about this reformulation.

Hence, any x that satisfies the robust constraint will be such that $\gamma\|a(x)\|_2 \leq b(x)$. Yet, we also now from lemma 3.6 that for a Z distributed according to a distribution that follows assumption 3.4, it must be the case that

$$\mathbb{P}((a(x)/\|a(x)\|_2)^T Z > \gamma) \leq \exp(-\gamma^2/2); .$$

In other words, it should be that

$$\mathbb{P}(a(x)^T Z \leq \gamma\|a(x)\|_2) \geq 1 - \exp(-\gamma^2/2); .$$

Therefore, for the robust x's, we have the guarantee that

$$\mathbb{P}(a(x)^T Z \leq b(x)) \geq \mathbb{P}(a(x)^T Z \leq \gamma\|a(x)\|_2) \geq 1 - \exp(-\gamma^2/2); ,$$

and can conclude that the chance constraint is necessarily satisfied since

$$\exp(-\gamma^2/2) = \exp(-\left(\sqrt{2\ln(1/\epsilon)}\right)^2/2) = \epsilon \; .$$

□

It might not come as a surprise that a similar result can be obtained for the ellipsoidal set intersected with the known support. We include below the details of this corollary which can also be found as Proposition 2.3.3 in [10].

Corollary 3.7. : *Given some $\epsilon > 0$ and some random vector Z that satisfies assumption 3.4, one has the guarantee that any x satisfying the robust constraint*

$$a(z)^T x \leq b(z) \, , \forall \, z \in \mathcal{Z}_{ell \cap box}(\gamma) \, ,$$

where

$$\mathcal{Z}_{ell \cap box}(\gamma) := \{z \in \mathbb{R}^m \,|\, z_i \in [-1, \, 1], \, \|z\|_2 \leq \gamma\}$$

and $\gamma := \sqrt{2\ln(1/\epsilon)}$ is guaranteed to satisfy the following chance constraint

$$\mathbb{P}(a(Z)^T x \leq b(Z)) \geq 1 - \epsilon \, .$$

even though the distribution of Z is not known.

Proof. First, following theorem 2.10 and the example 2.2.1, one can show that constraint

$$a(z)^T x \leq b(z) \, , \forall \, z \in \mathcal{Z}_{\text{ell}\cap\text{box}}(\gamma) \, ,$$

can be reformulated as

$$\exists v \in \mathbb{R}^m, \ \|v\|_1 + \gamma \|a(x) - v\|_2 \leq b(x) \tag{3.2}$$

where $v \in \mathbb{R}^n$ is an auxiliary decision vector while $a(x)$ and $b(x)$ are defined as in the proof of theorem 3.5. This being said, let the pair (x^*, v^*) be feasible with respect to this reformulated constraint. Then we must have that for any $z \in [-1, 1]^m$ that make the constraint $a(x^*)^T z \leq b(x^*)$ infeasible, the following also holds:

$$\begin{aligned}
a(x^*)^T z > b(x^*) \ &\Rightarrow \ (a(x^*) - v^*)^T z + v^{*T} z > b(x^*) \\
&\Rightarrow \ \|v^*\|_1 + (a(x^*) - v^*)^T z > b(x^*) && \text{since } \|z\|_\infty \leq 1 \\
&\Rightarrow \ b(x^*) - \gamma \|a(x^*) - v^*\|_2 + (a(x^*) - v^*)^T z > b(x^*) && \text{since } (x^*, v^*) \text{ satisfy (3.2)} \\
&\Rightarrow \ (a(x^*) - v^*)^T z > \gamma \|a(x^*) - v^*\|_2 \, .
\end{aligned}$$

This can directly be used to establish that

$$\begin{aligned}
\mathbb{P}(a(x^*)^T Z > b(x^*)) &\leq \mathbb{P}((a(x^*) - v^*)^T Z > \gamma \|a(x^*) - v^*\|_2) \\
&\leq \mathbb{P}\left(\frac{(a(x^*) - v^*)^T}{\|a(x^*) - v^*\|_2} Z > \gamma\right) \leq \exp(-\gamma^2/2) \, ,
\end{aligned}$$

where we exploited Lemma 3.6, or otherwise is trivially bounded above by zero when $a(x^*) = v^*$. The rest of the proof follows easily. □

Finally, making use of the above corollary, we present an analogous result for the budgeted uncertainty set which possibly contributed to its rise in popularity given that the reformulation associated to this set preserves the structural complexity of the constraint (i.e. a linear constraint is replaced with a set of linear constraints instead of a second order cone constraint).

Corollary 3.8. : *Given some $\epsilon > 0$ and some random vector Z that satisfies assumption 3.4, one has the guarantee that any x satisfying the robust constraint*

$$a(z)^T x \leq b(z)\,,\, \forall\, z \in \mathcal{Z}_{budg}(\Gamma)\ ,$$

where

$$\mathcal{Z}_{budg}(\Gamma) := \{z \in \mathbb{R}^m \,|\, z_i \in [-1,\, 1], \forall i,\ \|z\|_1 \leq \Gamma\}$$

and $\Gamma := \sqrt{2m\ln(1/\epsilon)}$ is guaranteed to satisfy the following chance constraint

$$\mathbb{P}(a(Z)^T x \leq b(Z)) \geq 1 - \epsilon\ .$$

even though the distribution of Z is not known.

Proof. This can be straightforwardly obtained by considering that a solution x^* that satisfies the robust constraint with budgeted uncertainty set $\mathcal{Z}_{\text{budg}}(\Gamma)$, also satisfies the robust constraint with the uncertainty set $\mathcal{Z}_{\text{ell}\cap\text{box}}(\Gamma/\sqrt{m})$. Hence, theorem 3.7 can be applied to indicate that the guarantee is provided as long as $\Gamma/\sqrt{m} = \sqrt{2\ln(1/\epsilon)}$. □

Note that in [19], the authors propose two other bounds methods for obtaining a smaller value for Γ while preserving the same probabilistic guarantees. Since the motivating arguments are a little more sophisticated, we leave to the reader to go read the details if he is curious. One can also find in chapter 2.2 of [10] a number of alternative hypothesis that can be made about the distribution in order for $\mathcal{Z}_{\text{ell}}(\gamma)$, $\mathcal{Z}_{\text{ell}\cap\text{box}}(\gamma)$, and $\mathcal{Z}_{\text{budg}}(\Gamma)$ to provide conservative (a.k.a. safe) approximations of the ambiguous chance constraint.

3.3 Risk-return Tradeoff Approximation

While, there are different ways of calibrating the parameters that express the size of the uncertainty set in order to approximate a certain chance-constraint that the decision maker has in mind, one should be aware that chance-constraints are themselves parametrized with ϵ that is potentially ambiguous since it captures the probability of failure that the decision maker (DM) is comfortable with. In fact, there are many situations in which this probability is not a firm value. Instead, the DM might be willing to accept a larger probability of failure as long as the performance is improved by an amount that is substantial enough. In other words, there is ambiguity regarding how much risk might be considered acceptable in order to achieve larger returns. In order to provide support for such DM, perhaps the goal of an optimization model is to present the DM with a set of solutions that each achieve different level of compromise between the performance that is expected and the risks of sub-performance. Practically speaking, this could imply the use of an optimization model that returns solution that are more or less conservative depending on some control parameter.

Let's take as example (again!) a portfolio selection problem. When investing in the stock market, we might be interested in maximizing the return of our investment while being worried that our investment leads to a loss. Yet, it might be unclear what type of trade-off we are willing to make between increased expected return and increased probability of loss. A model that is based on what was proposed by Markowitz [34] would propose solving the following optimization problem:

$$\begin{array}{ll} \text{maximize} & \mathbb{E}[r^T x] \\ \text{subject to} & \mathbb{P}(r^T x \geq 0) \geq 1 - \epsilon \\ & \sum_{i=1}^{n} x_i = 1 \\ & x \geq 0\ . \end{array}$$

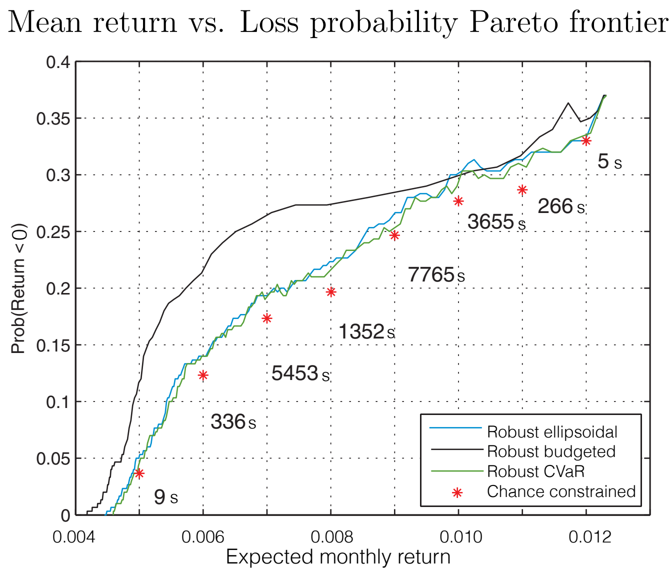
Now, given that the investor is not committed to a maximum probability of losses and is willing to be tempted by larger probabilities if the expected returns are significantly increased, then it becomes relevant to present to him the "Pareto" frontier of alternatives. In other words, we would identify a set of all portfolios that could be obtained by this optimization model as ϵ is adjusted from 0 to 100%. This is what was done by Bertsimas and Brown [15] while highlighting how robust optimization could be used to obtain portfolios that are highly competitive at a much reduced computational cost.

Namely, following the guidelines established in section 3.1, the robust approximation of this chance constraint would simply takes the form

$$\begin{array}{ll} \text{maximize} & \mu^T x \\ \text{subject to} & r^T x \geq 0\,,\ \forall\, r \in \mathcal{U} \\ & \sum_{i=1}^{n} x_i = 1 \\ & x \geq 0\ , \end{array}$$

for some well calibrated set $\mathcal{U}$. Whether the set $\mathcal{U}$ be an ellipsoidal set, a budgeted uncertainty set, or even a CVaR uncertainty set, each have a way of controlling how conservative (i.e. robust) the solution will be: γ, Γ, or α respectively.

The figure below (originally as Figure 1.1 in [15]) presents the Pareto frontier of performance pairs, i.e. (expected return, probability of loss) achieved by the stochastic programming model and its three robust approximations.



Note that in comparison on average the solutions of RO methods were obtained in 1 second.

3.4 Uncertainty Set Design based on Risk Measures

In [2], Artzner et al. introduce for the first time the notion of a family of risk measures that are rational to employ. He indicates that such measures ρ should satisfy the following properties when defined in terms of an uncertain income:

- Translation invariance : the risk of a position to which we add a guaranteed income is reduced by the amount of the income, i.e. $\rho(Y + c) = \rho(Y) - c$ when c is certain

- Subadditivity: the risk of the sum of risky positions should be lower than the sum of the risks, i.e. $\rho(X + Y) \leq \rho(X) + \rho(Y)$

- Positive homogeneity : if the consequences of a risky position are scaled by the same positive amount $\lambda \geq 0$, then the risk should be scaled by the same amount, i.e. $\rho(\lambda Y) = \lambda\rho(Y)$

- Monotonicity: A risky position that is guaranteed to return larger income than another risky position should be considered less risky, i.e. $X \geq Y \Rightarrow \rho(X) \leq \rho(Y)$.

- Relevance : if a risky position has the potential of leading to a loss, then the risk should be strictly positive, i.e. $X \leq 0 \& X \neq 0 \Rightarrow \rho(X) > 0$.

Based on these five axioms, the authors are able to demonstrate that the risk measure must be representable in the following form:

$$\rho(Y) := \sup_{F \in \mathcal{D}} \mathbb{E}_F[-Y] \,,$$

where $\mathcal{D}$ is a set of distributions for the random variable Y.

The family of coherent risk measure has caught a lot of attention since the financial crisis of 2008 as it was recognize that value-at-risk did not satisfy all of the mentioned axioms. Instead, there is now many arguments promoting the use of an alternative method for quantifying risk known as Conditional Value-at-Risk (CVaR). Intuitively, this new measure evaluates the expected value of the revenues under the scenarios that leads to the p% worst outcomes. For this reason, it obviously always overestimates risks when compared to the VaR, namely that $\mathrm{CVaR}_{1-\epsilon}(Y) \geq \mathrm{VaR}_{1-\epsilon}(Y)$. It is known to be a coherent risk measure and is now considered by many to be more reasonable to use than the VaR (see in particular a discussion on this topic in[4]). These considerations have led to an increase of the use of the CVaR in many disciplines such as healthcare, supply chain, network design, vehicle routing, energy, etc.

In [14], the authors actually identified an interesting connection between coherent risk measures and robust linear constraint. They actually established that there is a one to one correspondence between robust linear constraints and constraint on the risk measured by a coherent risk measure. In particular, they provide arguments for the following theorem.

Theorem 3.9. : *Given a coherent risk measure $\rho(\cdot)$, there always exists a convex uncertainty set $\mathcal{Z}$ such that the no risk constraint*

$$\rho(b(Z) - a(Z)^T x) \leq 0$$

is equivalent to imposing the robust constraint

$$a(z)^T x \leq b(z) \,,\, \forall\, z \in \mathcal{Z};.$$

The converse is also true.

Proof. Given any coherent risk measure $\rho(\cdot)$, we have mentioned that the representation theorem for this family of risk measure guarantees that the no risk constraint can be represented as

$$\sup_{F \in \mathcal{D}} \mathbb{E}_F \left[-\left(b(Z) - a(Z)^T x \right) \right] \leq 0 \,,$$

for some set of distribution $\mathcal{D}$. Since the revenue expression is a linear function of the random vector Z, one can obtain the following equivalent constraint

$$\sup_{F \in \mathcal{D}} \left[\begin{array}{cc} a(\mathbb{E}_F[Z])^T & b(\mathbb{E}_F[Z]) \end{array} \right] \left[\begin{array}{c} x \\ -1 \end{array} \right] \leq 0 \,.$$

This constraint can be reformulated in simpler terms as

$$\begin{bmatrix} a(z)^T & b(z) \end{bmatrix} \begin{bmatrix} x \\ -1 \end{bmatrix} \leq 0 \,,\, \forall\, z \in \mathcal{Z}' \,,$$

where

$$\mathcal{Z}' := \{z \in \mathbb{R}^m \,|\, \exists F \in \mathcal{D},\, z = \mathbb{E}_F[Z]\} \,,$$

which can further be reformulated as

$$a(z)^T x \leq b(z) \,,\, \forall\, z \in \mathcal{Z};.$$

where $\mathcal{Z} := \text{ConvexHull}(\mathcal{Z}')$.[3] It is clear that this last reformulation is exactly the robust constraint presented in the theorem.

The converse of this result can be obtained by observing that for any $\mathcal{Z}$, one can construct the following set of distributions

$$\mathcal{D} := \{F \,|\, \exists\, z \in \mathcal{Z},\, \mathbb{P}_F(Z = z) = 1\} \,,$$

where all distributions in $\mathcal{D}$ put all of their mass at a single $z \in \mathcal{Z}$, a.k.a. Dirac distributions, and considering the no risk constraint for the risk measure represented as $\rho(Y) := \sup_{F \in \mathcal{D}} \mathbb{E}[-Y]$. Indeed, we then have that:

$$\sup_{F \in \mathcal{D}} \mathbb{E}_F \left[- \left(b(Z) - a(Z)^T x \right) \right] = \sup_{z \in \mathcal{Z}} a(z)^T x - b(z)$$

so that $\rho(b(Z) - a(Z)^T x) \leq 0$ is equivalent to the robust constraint.

□

The above theorem is interesting as it says that robust constraints are not only useful to approximate some chance constraints or value-at-risk objectives but rather they can represent any risk attitude that can be characterized as a coherent risk measure. Alternatively, one can rest assured that when he imposes a robust linear constraint, he is employing an attitude towards risks that is justified and reasonable given that one considers that the axioms motivating the coherent risk measures are reasonable.

Remark 3.10. : It is important to note that the connection established between robust linear constraints and coherent risk measures has not yet been fully extended to nonlinear robust constraints. The only part of theorem 3.9 that holds more generally is the fact that any robust constraint can be interpreted as a coherent risk measure. The converse is not true, namely not all bounded risk constraint can be reformulated as a robust constraint. We refer curious readers to [18] for a discussion on potential generalizations. Yet, to this date the most accessible interpretation remains the idea of approximating chance-constraints.

[3]Indeed, given any z_1 and z_2 in $\mathcal{Z}'$, the fact that $a(z_i)^T x - b(z_i) \leq 0$ for $i = 1, 2$ implies that for all $\theta \in [0,\, 1]$, we have that:

$$a(\theta z_1 + (1-\theta) z_2)^T x - b(\theta z_1 + (1-\theta) z_2) = \theta(a(z_1)^T x - b(z_1)) + (1-\theta)(a(z_2)^T x - b(z_2)) \leq 0 \,.$$

More details can be found in Lemma 4.7.

3.4.1 The Conditional Value-at-Risk Measure

We now take a closer look at the conditional value-at-risk measure and demonstrate how it is related to the CVaR uncertainty set presented in chapter 2.

Mathematically, the most popular representation for the CVaR measure appeared in [39] and takes the following form when the random variable Y represents an uncertain revenue

$$\mathrm{CVaR}_{1-\epsilon}(Y) := \inf_t t + \frac{1}{\epsilon}\mathbb{E}[\max(0, -Y - t)] \; .$$

Intuitively, it is worth knowing that at optimum the value t^* will capture the value-at-risk for the given uncertain revenue so that

$$\begin{aligned}\mathrm{CVaR}_{1-\epsilon}(Y) &= \mathrm{VaR}_{1-\epsilon}(Y) + (1/\epsilon)\mathbb{E}[\max(0, -Y - \mathrm{VaR}_{1-\epsilon}(Y)] \\ &= \mathrm{VaR}_{1-\epsilon}(Y) + \mathbb{E}[-Y - \mathrm{VaR}_{1-\epsilon}(Y) \,|\, -Y \geq \mathrm{VaR}_{1-\epsilon}(Y)] \; .\end{aligned}$$

It is perhaps surprising that the most famous representation for CVaR does not take the shape of $\rho(Y) := \sup_{F\in\mathcal{D}} \mathbb{E}_F[-Y]$. The reason is potentially that the above representation is more intuitive. In any case, let's identify how the CVaR measure can be represented in the form $\rho(Y) := \sup_{F\in\mathcal{D}} \mathbb{E}_F[-Y]$. When the distribution is discrete, this can actually be obtained by employing linear programming duality. In particular, CVaR can be evaluated by solving the following linear program:

$$\begin{aligned}\underset{t,s}{\text{minimize}} \quad & t + \frac{1}{\epsilon}\sum_{i=1}^{K} p_i s_i \\ \text{subject to} \quad & s_i \geq -y_i - t \, , \; \forall\, i = 1, \ldots, K \\ & s_i \geq 0 \, , \; \forall\, i = 1, \ldots, K \; ,\end{aligned}$$

where $s \in \mathbb{R}^K$ while $p_i \in \mathbb{R}$ and $y_i \in \mathbb{R}$ represents respectively the probability and realized value of Y under each scenario $i = 1, \ldots, K$. Applying duality we obtain the following equivalent linear program:

$$\begin{aligned}\underset{\lambda}{\text{maximize}} \quad & -\lambda^T y \\ \text{subject to} \quad & \lambda \geq 0 \\ & \lambda_i \leq p_i/\epsilon \, , \; \forall\, i = 1, \ldots, K \\ & \sum_{i=1}^{K} \lambda_i = 1 \; .\end{aligned}$$

Hence, it takes the shape of $\sup_{F\in\mathcal{D}} \mathbb{E}_F[-Y]$ where $\mathcal{D} := \{F \,|\, \mathbb{P}_F(Y = y_i) \leq p_i/\epsilon \,\forall\, i = 1, \ldots, K, \, \sum_{i=1}^{K} \mathbb{P}_F(Y = y_i) = 1\}$.

Following theorem 3.9, we are now aware that given a random vector Z with discrete distribution described by $\{p_i, \bar{z}_i\}_{i=1}^{K}$, and any linear revenue function $b - a(Z)^T x$, using the set

$$\mathcal{Z}' := \{z \in \mathbb{R}^m \,|\, \exists F \in \mathcal{D}, \, z = \mathbb{E}_F[Z]\} \; ,$$

we are able to construct a robust linear constraint that is equivalent to

$$\mathrm{CVaR}_{1-\epsilon}(b(Z) - a(Z)^T x) \leq 0 \;.$$

Such a robust constraint takes the form

$$a(z)^T x \leq b(z) \,,\, \forall\, z \in \mathrm{ConvexHull}(\mathcal{Z}') \;.$$

Yet, by manipulating $\mathcal{Z}'$, we obtain

$$\begin{aligned}\mathcal{Z}' = &\; \{z \in \mathbb{R}^m \,|\, \exists F,\, \mathbb{P}_F(Y = y_i) \leq p_i/\epsilon,\, \sum_{i=1}^K \mathbb{P}_F(Y = y_i) = 1,\, z = \mathbb{E}_F[Z]\} \\ = &\; \{z \in \mathbb{R}^m \,|\, \exists \theta \in \mathbb{R}^K,\, \theta \geq 0,\, \theta_i \leq p_i/\epsilon,\, \sum_{i=1}^K \theta_i = 1,\, z = \sum_i \bar{z}_i \theta_i\} \;.\end{aligned}$$

Hence, $\mathcal{Z}'$ is convex and actually takes the shape of the CVaR uncertainty set encountered in chapter 2.

3.5 Exercises

3.5.1 Data-driven robust portfolio optimization

Consider the portfolio optimization problem studied in example 3.1.1. In Google Colab, you will find how we are able to manipulate the data in order to characterize the uncertainty in the return vector as a function of some Z primitive where each Z_i can be assumed to be independently and symmetrically over the $[-1, 1]$ interval. Namely, you can consider that

$$r = \mu + PZ \;.$$

You can answer all of the exercises below in Google Colab.

Exercise 3.1. ***Calibration of uncertainty sets using data***
For each of the three uncertainty sets below, calibrate the size parameter in order for $\mathcal{Z}$ to include 95% of the observed realization:

1. Budgeted uncertainty set, i.e. $\mathcal{Z} := \{z \in \mathbb{R}^m \,|\, z_i \in [-1,\, 1], \|z\|_1 \leq \Gamma\}$
2. Boxed ellipsoidal set, i.e. $\mathcal{Z} := \{z \in \mathbb{R}^m \,|\, z_i \in [-1,\, 1], \|z\|_2 \leq \gamma\}$

Note that we already provided the calibration scheme for the boxed ellipsoidal set in the Google Colab file, hence you are only asked to calibrate the budgeted set.

Exercise 3.2. ***Calibration of uncertainty sets using distribution hypothesis***
For each of the two uncertainty sets below, calibrate the size parameter in order for $\mathcal{Z}$ to be such that a robust linear constraint employing $\mathcal{Z}$ is guaranteed to return a solution that will satisfy the chance constraint with 95% probability as long as the distribution of Z satisfies assumption 3.4:

1. Budgeted uncertainty set, i.e. $\mathcal{Z} := \{z \in \mathbb{R}^m \,|\, z_i \in [-1,\, 1], , \|z\|_1 \leq \Gamma\}$
2. Boxed ellipsoidal set, i.e. $\mathcal{Z} := \{z \in \mathbb{R}^m \,|\, z_i \in [-1,\, 1], , \|z\|_2 \leq \gamma\}$

Exercise 3.3. ***Evaluation of performance***
Evaluate the performance of the portfolios obtained from the five robust optimization models below:

1. Robust optimization model that approximates the 95% Value-at-Risk model using the budgeted uncertainty set calibrated in question 3.1.
2. Robust optimization model that approximates the 95% Value-at-Risk model using the boxed ellipsoidal uncertainty set calibrated in question 3.1.
3. Robust optimization model that approximates the 95% Value-at-Risk model using a CVaR uncertainty set with parameter $\alpha = 0.05$
4. Robust optimization model that approximates the 95% Value at Risk model using the budgeted uncertainty set calibrated in question 3.2.
5. Robust optimization model that approximates the 95% Value-at-Risk model using the boxed ellipsoidal uncertainty set calibrated in question 3.2.

The performance of each portfolio obtained should be compared in terms of actual value at risk achieved under the following four sets of conditions:

1. The best value at risk achievable according to each optimization model
2. Empirical distributions of returns over the years 2000-2009
3. Empirical distributions of returns over the years 2010-2014
4. Distribution that assumes that each Z_i is i.i.d. and has 50% chance of achieving either extreme values of the interval $[-1, 1]$.

Discuss your findings.

3.5.2 Risk averse production problem

Consider a production problem as discussed in chapter 1.1 where both the conversion rates of raw materials and the profit generated by the assembled products are uncertain. This gives rise to the risk averse optimization problem:

$$
\begin{aligned}
\max_{x,y} \quad & \rho(\tilde{p}^T y - c^T x) && (3.3a)\\
\text{subject to} \quad & d^T y \leq z^T x,\ \forall z \in \mathcal{Z} && (3.3b)\\
& Ax + By \leq b\,, && (3.3c)\\
& x \geq 0,\ y \geq 0, && (3.3d)
\end{aligned}
$$

where $x \in \mathbb{R}^n$ denotes the orders (in kg) of raw material of types $i \in \{1, \dots, n\}$, $y \in \mathbb{R}^m$ denotes the number of boxes (of 1000 packs) of drugs of type $j = 1, \dots, m$, constraint (3.3c) captures K capacity constraints on the other resource needed (e.g. manpower, equipment, and budget) for production through $A \in \mathbb{R}^{K \times n}$ and $B \in \mathbb{R}^{K \times m}$ and $b \in \mathbb{R}^K$. Finally, the model handles uncertainty about raw materials through an uncertainty set defined as

$$\mathcal{Z} := \{z \in \mathbb{R}^n \,|\, \exists \Delta \in [-1,\, 1]^n,\, z_i = (1 + \Delta_i)\hat{z}_i,\, \forall i = 1, \dots, n,\, \sum_{i=1}^{n} \Delta_i \geq -\Gamma\sqrt{n}\}\,,$$

which captures an estimated conversion rate denoted by $\hat{z}$, and, since the individual relative rate perturbations are considered i.i.d. with mean zero, the fact that, because of the central limit theorem, we should have the sum of perturbations converging to a normal distribution with standard deviation proportional to $\sqrt{n}$ (see [3]). Furthermore, the profit uncertainty is based on a set of scenarios $\{\bar{p}_i\}_{i=1}^N$ with respective probabilities $\{\bar{q}_i\}_{i=1}^N$ and the risk aversion is modeled using the following risk measure: and

$$\rho(\tilde{p}^T y - c^T x) := \min_{q \in \mathcal{Q}} \sum_{i=1}^{N} q_i \bar{p}_i^T y - c^T x$$

with

$$\mathcal{Q} := \{q \in \mathbb{R}^N \mid \sum_{i=1}^{N} q_i = 1,\, \gamma^{-1}\bar{q}_i \leq q_i \leq \gamma\bar{q}_i,\, \forall i = 1, \dots, N\},$$

for some $\gamma \geq 1$.

Exercise 3.4. ***Reformulation for risk averse production problem***
Derive a finite dimensional LP formulation for problem (3.3) and implement the risk averse production problem using RSOME (using Google Colab) in its unreduced form first, then reduced form if you are brave enough.

Chapter 4

Adjustable Robust Linear Programming

There are many situations in which decisions need to be made and implemented at different points of time. Something peculiar happens in contexts where there is uncertainty as one needs to model the fact that later decisions might be able to exploit some information that was initially unavailable. Here is an illustrative example that highlights many of the difficulties that arise in the modeling stage.

4.1 Why worry about decision sequences?

Consider a simple inventory problem in which a retailer needs to order some goods in order to accommodate his customers while incurring the lowest ordering/holding/backlogging cost. In a deterministic setting, this might take the shape of the following convex optimization problem

$$\begin{array}{ll} \underset{x_t, y_t}{\text{minimize}} & \sum_t c_t x_t + h_t(y_{t+1})^+ + b_t(-y_{t+1})^+ \\ \text{subject to} & y_{t+1} = y_t + x_t - d_t \,,\, \forall\, t = 1, \ldots, T \\ & 0 \leq x_t \leq M \;, \end{array}$$

where $x_t \in \mathbb{R}$ captures the number of goods ordered at time t and received by time $t+1$, y_t is the number of goods in stock at the beginning of time t, with y_1 as the initial inventory, then d_t is the demand for the good between time t and $t+1$ and the cost parameters are c_t for ordering cost, h_t for holding (i.e. storage) cost, and b_t for backlog cost. Finally, the operation $(y)^+ := \max(0, y)$ is one that returns the value of y only if it is positive, otherwise returns 0.

It is well known that this problem can be reformulated as a linear program:

$$\begin{array}{ll}\underset{x_t,y_t,s_t^+,s_t^-}{\text{minimize}} & \sum_t c_t x_t + h_t s_t^+ + b_t s_t^- \\ \text{subject to} & y_{t+1} = y_t + x_t - d_t \,,\, \forall\, t = 1,\ldots,T \\ & s_t^+ \geq 0,\; s_t^- \geq 0 \\ & s_t^+ \geq y_{t+1} \\ & s_t^- \geq -y_{t+1} \\ & 0 \leq x_t \leq M \;,\end{array}$$

where $s_t^+ \in \mathbb{R}$ counts the number of storage spaces to pay for during period t while s_t^- counts the number of goods that are missing during period t.

A naïve approach that can be used to "robustify" this problem would be to simply state the robust counterpart as

$$\begin{array}{ll}\underset{x_t,y_t,s_t^+,s_t^-}{\text{minimize}} & \sup_{z\in\mathcal{Z}} \sum_t c_t(z) x_t + h_t(z) s_t^+ + b_t(z) s_t^- \\ \text{subject to} & y_{t+1} = y_t + x_t - d_t(z) \,,\, \forall\, z \in \mathcal{Z} \,,\, \forall\, t = 1,\ldots,T \\ & s_t^+ \geq 0,\; s_t^- \geq 0 \\ & s_t^+ \geq y_{t+1} \\ & s_t^- \geq -y_{t+1} \\ & 0 \leq x_t \leq M \;,\end{array}$$

where we simply robustified the objective function and each constraint that involved some uncertain parameters.

Based on what we know, it sounds fairly straightforward to obtain a tractable reformulation for this model when $\mathcal{Z}$ is polyhedral and $c_t(z)$, $h_t(z)$, $b_t(z)$, and $d(z)$ are affine functions of z. Unfortunately, there are obvious issues with this formulation. The first one that would be encountered is infeasibility of the constraint

$$y_{t+1} = y_t + x_t - d_t(z) \,,\, \forall\, z \in \mathcal{Z}$$

Indeed, the above constraint is equivalent to

$$y_{t+1} = y_t + x_t - \tilde{d} \,,\, \forall\, \tilde{d} \in \{\tilde{d} \in \mathbb{R} \,|\, \exists\, z \in \mathcal{Z}, \tilde{d} = d_t(z)\} \,.$$

Unless the implied uncertainty set for $\tilde{d}$ is an interval of zero length (i.e. no uncertainty about d_t) which is unlikely, this constraint cannot be satisfied. Take for instance any two possible demand values $\tilde{d}_1$ and $\tilde{d}_2$ that we would wish to be immuned to. This would require that the constraint is met for both then

$$\tilde{d}_1 = y_t + x_t - y_{t+1} = \tilde{d}_2 \,.$$

Yet, this is only possible if there is no uncertainty, i.e. $\tilde{d}_1 = \tilde{d}_2$ and otherwise necessarily leads to infeasibility.

A simple way of solving the above issue is to get rid of the equality constraint in the nominal problem before deriving the robust counterpart. This would lead to the deterministic model

$$\begin{aligned}
\underset{x_t,s_t^+,s_t^-}{\text{minimize}} \quad & \sum_t c_t x_t + h_t s_t^+ + b_t s_t^- && (4.1a)\\
\text{subject to} \quad & s_t^+ \geq 0,\ s_t^- \geq 0 && (4.1b)\\
& s_t^+ \geq y_1 + \sum_{t'=1}^{t} x_{t'} - d_{t'} && (4.1c)\\
& s_t^- \geq -y_1 + \sum_{t'=1}^{t} d_{t'} - x_{t'} && (4.1d)\\
& 0 \leq x_t \leq M \ , && (4.1e)
\end{aligned}$$

The robust counterpart of this model takes the form:

$$\begin{aligned}
\underset{x_t,s_t^+,s_t^-}{\text{minimize}} \quad & \sup_{z\in\mathcal{Z}} \sum_t c_t(z) x_t + h_t(z) s_t^+ + b_t(z) s_t^- && (4.2a)\\
\text{subject to} \quad & s_t^+ \geq 0,\ s_t^- \geq 0 && (4.2b)\\
& s_t^+ \geq y_1 + \sum_{t'=1}^{t} x_{t'} - d_{t'}(z)\ ,\ \forall\, z \in \mathcal{Z}\ ,\ \forall\, t = 1,\ldots,T && (4.2c)\\
& s_t^- \geq -y_1 + \sum_{t'=1}^{t} d_{t'}(z) - x_{t'}\ ,\ \forall\, z \in \mathcal{Z}\ ,\ \forall\, t = 1,\ldots,T && (4.2d)\\
& 0 \leq x_t \leq M \ . && (4.2e)
\end{aligned}$$

This is in fact similar to the model that was proposed in [20]. A solution to this model necessarily exists since one can set $x_t = 0$ for all t. Unfortunately, there are still two issues with this model. If these issues are not resolved before deriving the robust counterpart, this might mislead a practitioner to conclude that a robust solution is necessarily overly conservative.

First, there is the fact that this model makes an important assumption about what type of policy is used. Namely, it assumes that the policy for x_t is predefined at time $t = 1$ and never modified even though some information about earlier demand might be obtained as time progresses. This is fine if the objective is to model for instance a situation where all orders are made to suppliers at time $t = 1$ and the type of contract prevents the company to make future modification to these orders. In general, however it might be the case that at any point of time it is possible to exploit the information that was received about past demand in order to adjust the order that is about to be received. In this case, we need to consider that x_t is adjustable with respect to $\{d_{t'}\}_{t'=1}^{t-1}$. We will show how to do this shortly. Let's now have a look at the second issue.

The second issue is similar but much more subtle. To better understand it, let's have a look at a problem instance in which there is only one period of execution, and where $c_1 = 0.5$ while $h_1 = b_1 = 1$. In this context, we would be interested in robustifying

$$\begin{aligned} \underset{x_1}{\text{minimize}} \quad & 0.5x_1 + (y_1 + x_1 - d_1)^+ + (-y_1 - x_1 + d_1)^+ \\ \text{subject to} \quad & 0 \leq x_1 \leq 2 \,, \end{aligned}$$

with $y_1 = 0$ and $d_1 \in [0, 2]$. Note here that since the problem only has one stage of execution, it is unlikely that new information would be obtained by the time that the decision is implemented. The robust counterpart of this problem instance using the model described in problem (4.2) will take the form:

$$\begin{aligned} \underset{x_1, s_1^+, s_1^-}{\text{minimize}} \quad & 0.5x_1 + s_1^+ + s_1^- && (4.3\text{a}) \\ \text{subject to} \quad & s_1^+ \geq 0,\ s_1^- \geq 0 && (4.3\text{b}) \\ & s_1^+ \geq x_1 - d_1 \,,\ \forall\, d_1 \in [0, 2] && (4.3\text{c}) \\ & s_1^- \geq -x_1 + d_1 \,,\ \forall\, d_1 \in [0, 2] && (4.3\text{d}) \\ & 0 \leq x_1 \leq 2 \,. && (4.3\text{e}) \end{aligned}$$

One can easily verify that the optimal solution here suggests $x_1^* = 0$, ${s_1^+}^* = 0$ and ${s_1^-}^* = 2$ and an optimal value of 2. Indeed, it is the case that if $x_1 = 0$ the worst-case scenario would be that a demand of two units occurs and leads to a backlog cost of 2. However, is this truly the best that one can do to reduce worst-case inventory costs?

Think for instance about the solution $x_1^{**} = 1$ which would lead to two equivalent worst-case scenarios:

1. demand is 0 unit hence the total inventory cost would be 1.5: namely 0.5 in production cost, and 1 in holding cost

2. demand is 2 units hence the total inventory cost would be 1.5: namely 0.5 in production cost, and 1 in backlog cost

Overall, $x_1^{**} = 1$ leads to a worst-case total cost of 1.5 which is smaller than the worst-case cost of $x_1^* = 0$ which is 2. So, why did the robust counterpart model obtained from problem (4.2) not provide the best solution in terms of worst-case cost ?

The reason has to do with the fact that in problem (4.1), s_t^+ and s_t^- are not authentic decision variables but rather constitute a set of "auxiliary" decision variables that are employed by the linearisation scheme, which serves to evaluate the objective function. Indeed, the true robust counterpart takes the following form

$$\begin{aligned} \underset{x_1}{\text{minimize}} \quad & \sup_{d_1 \in [0, 2]} 0.5x_1 + (y_1 + x_1 - d_1)^+ + (-y_1 - x_1 + d_1)^+ \\ \text{subject to} \quad & 0 \leq x_1 \leq 2 \,. \end{aligned}$$

This model can in some sense be linearized but only as a two-stage problem:

$$\begin{array}{ll} \underset{x_1}{\text{minimize}} & \sup_{d_1 \in [0,2]} \; 0.5x_1 + h(x_1, d_1) \\ \text{subject to} & 0 \leq x_1 \leq 2\ . \end{array}$$

where

$$\begin{aligned} h(x_1, d_1) := \quad \min_{s_1^+, s_1^-} \quad & s_1^+ + s_1^- && (4.4) \\ \text{subject to} \quad & s_1^+ \geq 0,\ s_1^- \geq 0 && (4.5) \\ & s_1^+ \geq x_1 - d_1 && (4.6) \\ & s_1^- \geq -x_1 + d_1\ . && (4.7) \end{aligned}$$

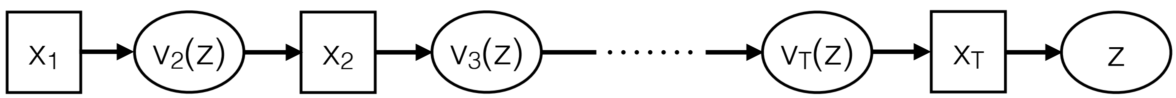
Note that in this linearized formulation the s_1^+ and s_1^- are allowed to depend on the instance of d_1 that is studied. Namely, with $x_1^{**} = 1$ they will take the values $s_1^+(d_1) := (1 - d_1)^+$ and $s_1^-(d_1) := (d_1 - 1)^+$. This is unlike the optimization problem (4.2) which proposed $x_1^* = 0$ and where the choice of s_1^+ and s_1^- needed to be fixed once before the realization of d_1 was known.

Conclusion: When robustifying a linear program that involves either decisions that are implemented (i.e. turned into an action) at different point of time (as x_t), or decision variables (called auxiliary decision variables) which only role in the mathematical program is to allow the computation of the objective value or the validation of a constraint (as s_t^+ and s_t^-), one must carefully identify the chronology of decisions and observations sequence and employ the adjustable robust counterpart framework. In particular, it is typically the case that auxiliary decision variables can be adjusted to the whole vector of uncertain parameters z.

4.2 The Adjustable Robust Counterpart Model

As seen in the above inventory problem, it is important before developing a robust optimization model to clearly layout the chronology of executions and observations as portrayed in the following diagram.

Chronology of executions x_i's, observations v_i's, and z



Note that in the above diagram, we represent decisions executed at time t as x_t, while observations made between time $t-1$ and t is represented by v_t (for "visual" evidence). The observation v_t is a function of z the underlying uncertainty that affects the decision problem as a whole. Finally, after the final decision is implemented, one can observe the realized uncertain vector z in its entirety in order to evaluate the objective function and assess whether all the constraint were met.

To be precise, while when there is no uncertainty a sequential decision problem might be easily described as

$$\begin{array}{ll} \underset{\{x_t\}_{t=1}^T}{\text{maximize}} & \sum_{t=1}^{T} c_t^T x_t + d \\ \text{subject to} & \sum_{t=1}^{T} a_{jt}^T x_t \leq b_j \,,\ \forall\, j = 1, \ldots, J, \end{array}$$

where each $x_t \in \mathbb{R}^n$ (without loss of generality), the situation is more complicated when uncertainty is inserted. Instead, one must consider the following multi-stage adjustable robust counterpart formulation

(Multi-Stage ARC)

$$\underset{x_1, \{x_t(\cdot)\}_{t=2}^T}{\text{maximize}} \quad \inf_{z \in \mathcal{Z}} \; c_1(z)^T x_1 + \sum_{t=2}^{T} c_t(z)^T x_t(v_t(z)) + d(z) \tag{4.8a}$$

$$\text{subject to} \quad a_{j1}(z)^T x_1 + \sum_{t=2}^{T} a_{jt}(z)^T x_t(v_t(z)) \leq b_j(z) \,,\ \forall\, z \in \mathcal{Z} \,,\ \forall\, j = 1, \ldots, J \tag{4.8b}$$

where $v_t : \mathbb{R}^m \to \mathbb{R}^\nu$ is a function that describes what is observed of z at time t, x_t is a mapping from the space of observations $\mathbb{R}^\nu$ to $\mathbb{R}^n$. The fact that each x_t is not a vector any more but rather a mapping is important as it enables the decision to react differently depending on the realized observation. Of course, this flexibility comes at the price of significant computational challenges.

Example 4.1. : Considering the general inventory problem presented in section 4.1, namely the following optimization problem

$$\begin{array}{ll} \underset{x_t, s_t^+, s_t^-}{\text{minimize}} & \sum_t c_t x_t + h_t s_t^+ + b_t s_t^- \\ \text{subject to} & s_t^+ \geq 0,\ s_t^- \geq 0 \\ & s_t^+ \geq y_1 + \sum_{t'=1}^{t} x_{t'} - d_{t'} \\ & s_t^- \geq -y_1 + \sum_{t'=1}^{t} d_{t'} - x_{t'} \\ & 0 \leq x_t \leq M \,, \end{array}$$

three questions clearly arise:

1. What is the source of uncertainty in this problem? Namely, the vector z which perfect knowledge would reduce the problem to a deterministic one where we can predict every outcome

2. What are the observations v_t that are made and could contribute to what decisions are executed? One would then need to describe how these observations are related to z.

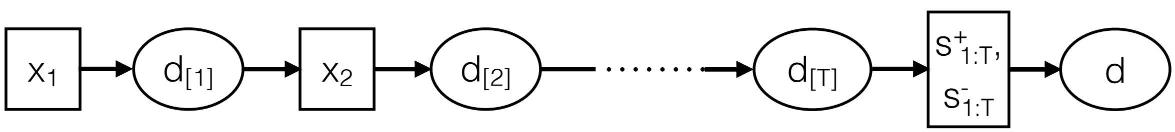
3. What is the chronology of each element of the problem: x_t's, y_t's, s_t's, and v_t's.

Let's assume that the uncertainty is limited to the vector of demand $d(z)$. One might consider that at each point of time, the inventory manager is able to observe all of the prior demand before making the order for the next period. In this case, we would have that

$$v_t(d) = d_{[t-1]} := [\ \boldsymbol{I}_{t-1} \quad \mathbf{0}_{t,T-t+1}\ ]d = [\ d_1 \quad d_2 \quad \cdots \quad d_{t-1}\ ]^T .$$

We are then left with defining the sequence of decision variables and observations

Chronology of executions and observations in inventory problem



Note that in the above chronology we made explicit that the s_t^+ and s_t^- variables are auxiliary variables that are adjustable with respect to the full uncertainty vector $d_{[T]} = d$. In fact, once $d_{[T]}$ is revealed the uncertainty can be considered reduced to zero, hence the presence of d in the chronology is somewhat artificial.

This being said we are left with the following multi-stage adjustable robust counterpart model:

$$\begin{aligned}
\underset{x_1,\{x_t(\cdot)\}_{t=2}^T,\{s_t^+(\cdot),s_t^-(\cdot)\}_{t=1}^T}{\text{minimize}} \quad & \sup_{d\in\mathcal{U}}\ c_1x_1 + \sum_t c_t x_t(d_{[t-1]}) + h_t s_t^+(d) + b_t s_t^-(d) \\
\text{subject to} \quad & s_t^+(d) \geq 0,\ s_t^-(d) \geq 0\,,\ \forall\, d\in\mathcal{U},\ \forall\, t \\
& s_t^+(d) \geq y_1 + \sum_{t'=1}^{t} x_{t'}(d_{[t'-1]}) - d_{t'}\,,\ \forall\, d\in\mathcal{U},\ \forall\, t \\
& s_t^-(d) \geq -y_1 + \sum_{t'=1}^{t} d_{t'} - x_{t'}(d_{[t'-1]})\,,\ \forall\, d\in\mathcal{U},\ \forall\, t \\
& 0 \leq x_t(d_{[t-1]}) \leq M\,,\ \forall\, d\in\mathcal{U},\ \forall\, t\,,
\end{aligned}$$

where $\mathcal{U}$ captures the set of potential demand vectors.

4.3 Time consistency issues

Consider a three stage inventory problem with an initial ordering cost of 1\$ per unit, and a larger second stage ordering cost of 4\$ per unit. We also assume that there are no holding cost and that backlog cost are only charged in the final stage at a cost of 10\$ per unit. Demand is expected to be of 1 unit for each time steps with a possible upward deviation of up to 1 unit. In order to control the level of conservatism of the solution, it is decided to use the budgeted uncertainty set with a budget of 1 (i.e. at most half of the future total demand deviation could occur). This gives rise to the following multi-stage ARC:

$$\begin{array}{ll}\underset{x_1,x_2(\cdot),s(\cdot)}{\text{minimize}} & \sup_{d\in\mathcal{U}} x_1 + 4x_2(d_1) + 10s(d) \\ \text{subject to} & s(d) \geq 0\,,\,\forall\, d\in\mathcal{U} \\ & s(d) \geq d_1 + d_2 - x_1 - x_2(d_1)\,,\,\forall\, d\in\mathcal{U} \\ & x_1 \geq 0 \\ & x_2(d_1) \geq 0\,,\,\forall\, d\in\mathcal{U}\ ,\end{array}$$

where $\mathcal{U} := \{d \in [0,\, 2]^2 \,|\, d_1 + d_2 \leq 3\}$. One can easily confirm that an optimal robust policy consists of ordering 3 units at time $t = 1$ and nothing at time $t = 2$. Under this policy, the worst-case total cost of 3\$ occurs for any realization of the pair (d_1, d_2) in $\mathcal{U}$. Intuitively, the policy is optimal since we wish to protect against the pair $(2, 1)$ which would require us to produce 3 units in order to avoid the large backlog cost, yet there is no reason to delay the purchase since the cost is lower at time $t = 1$.

The issue we wish to highlight here is the idea that the optimality of the policy that was identified relies entirely on the hypothesis that at time $t = 2$, the optimization problem that will be solved to select x_2 once d_1 is observed consists of the following:

$$\begin{array}{ll}\underset{x_2,s(\cdot)}{\text{minimize}} & \sup_{d_2\in\mathcal{U}_2(d_1)} x_1 + 4x_2 + 10s(d_2) \\ \text{subject to} & s(d_2) \geq 0\,,\,\forall\, d_2\in\mathcal{U}_2(d_1) \\ & s(d_2) \geq d_1 + d_2 - x_1 - x_2\,,\,\forall\, d_2\in\mathcal{U}_2(d_1) \\ & x_2 \geq 0\ ,\end{array}$$

where $\mathcal{U}_2(d_1) := \{d_2 \in \mathbb{R} \,|\, (d_1, d_2) \in \mathcal{U}\}$ captures the "slice" of $\mathcal{U}$ where d_1 is fixed to what was observed. In particular,

$$\begin{aligned}\mathcal{U}_2(0) &:= [0,\, 2] \\ \mathcal{U}_2(1) &:= [0,\, 2] \\ \mathcal{U}_2(2) &:= [0,\, 1]\ .\end{aligned}$$

Although this rule for updating the uncertainty set is implicitly assumed in the multistage ARC model, it may be more or less applicable from a modeling standpoint, depending on the particular application. We provide two examples in which this updating rule comes across as more or less realistic, and we comment on the potential pitfalls when this rule is violated.

- **Time-consistent situation:**Consider the owner of a coffee stand that is allowed to operate for one morning in the lobby of a hotel. The owner plans on selling coffee during the 7 a.m - 11 a.m period and possibly replenishing with fresh coffee at 9 a.m. Based on the hotel's occupancy level and his prior experience, he estimates that about 100 cups of coffee (one unit) might be purchased during the 7 a.m - 9 a.m interval, and about 100 cups (one unit) might be purchased during the 9 a.m - 11 a.m interval. He also considers it extremely unlikely that more than 300 cups of coffee (three units) would be needed in a single morning (e.g., since that happens to be the maximum number of guests at the hotel, and very few individuals buy two cups of coffee in the morning). This circumstance motivates an uncertainty set of the form $\mathcal{U}$, and it suggests that it may be reasonable to not order more coffee even after having sold 200 cups (two units) during the 7 a.m - 9 a.m interval.

- **Time-inconsistent situation:** Consider the same coffee stand owner that instead plans to move his stand at 9 a.m to a different nearby hotel that has similar occupancy. In this context, it might still be reasonable to initially assume when opening the stand at 7 a.m that no more than 300 cups of coffee (three units) would be needed the whole morning (possibly with the argument that if demands at the two hotels are independent, it would be unlikely that they are both significantly above their expected amounts). However, it seems unreasonable to assume that the sale of more than 200 cups (two units) in the first hotel by 9 a.m implies without any doubt that no more than 100 cups (one unit) are needed for customers at the second hotel. Instead, the owner may be tempted to believe that there might still be enough customers to sell up to 150 cups (1.5 units) in the second hotel and thus might make an order that departs from what his original optimal policy suggested.

The second scenario gives rise to time inconsistency in the sense that at time $t = 2$ the decision is actually taken with respect to a robust optimization model that uses an uncertainty set that is incoherent with $\mathcal{U}_2(d_1)$. In our example, this would be using $\tilde{\mathcal{U}}_2(2) := [0, 1.5]$ instead of $\mathcal{U}_2(2) := [0, 1]$. Looking back at the robust first-stage decision $x_1 = 3$, this would mean that if $d_1 = 2$ then the second ordering decision is made in order to minimize $\max_{d_2 \in \tilde{\mathcal{U}}_2(2)} \; 3 + 4x_2 + 10(3 + x_2 - 2 - d_2)^+$ which means an additional half unit will be ordered to avoid the excessive backlog cost. Under this scenario, the total cost ends up being $3+4\cdot 0.5 = 5$. Yet, it is clear that if the extra half unit had been purchased in the initial stage then the total cost would have been 3.5 for this scenario and always lower than this amount as long as the policy implemented at the second stage policy would be $x_2(d_1) := (d_1 + 2 - 3.5)^+$.

One can actually show that the decisions $x_1 = 4$ and $x_2(d_1) := (d_1 + 2 - 3.5)^+$. are optimal according to the bi-level problem:

$$\begin{aligned}
\underset{x_1, x_2(\cdot), s(\cdot)}{\text{minimize}} \quad & \sup_{d \in \mathcal{U}} \; x_1 + 4x_2(d_1) + 10s(d) \\
\text{subject to} \quad & s(d) \geq 0 \,, \; \forall\, d \in \mathcal{U} \\
& s(d) \geq d_1 + d_2 - x_1 - x_2(d_1) \,, \; \forall\, d \in \mathcal{U} \\
& x_1 \geq 0 \\
& x_2(d_1) \in \underset{x \geq 0}{\text{argmin}} \; \max_{d_2' \in \tilde{\mathcal{U}}_2(d_1)} \; c_2 x + 10(d_1 + d_2' - x_1 - x)^+ \,, \; \forall\, d_1 \in \mathcal{U}_1 \,,
\end{aligned}$$

where $\mathcal{U}_1 := \{d_1 \in \mathbb{R} \,|\, \exists d_2 \in \mathbb{R}, \, [d_1 \; d_2]^T \in \mathcal{U}\}$. This optimization model resolves time inconsistency in this example by making explicit the fact that the second stage decision must be coherent with respect to the uncertainty set which will be employed once we will observe d_1. Such bi-level optimization problems are only known to reduce to our multi-stage ARC when $\tilde{\mathcal{U}}_2(d_1) := \mathcal{U}_2(d_1) := \{d_2 \in \mathbb{R} \,|\, (d_1, d_2) \in \mathcal{U}\}$.

4.4 Difficulty of resolution

Some details about hardness of adjustable robust counterpart models.

Theorem 4.2. : *Solving problem* (4.8) *is NP-hard even when* $v_t(z) := z$ *and* $\mathcal{Z}$ *is polyhedral.*

Proof. This result is obtained by showing that the NP-complete 3-SAT problem can be reduced to verifying whether the optimal value of the following problem is greater or equal to zero:

$$\begin{aligned}
\underset{x(\cdot)}{\text{minimize}} \quad & \sup_{z \in [0,1]^m} \sum_{i=1}^{N} (x_i(z) - 1) & \text{(4.9a)} \\
\text{subject to} \quad & x_i(z) \geq a_{i,k}^T z + b_{i,k} \,, \; \forall\, z \in \mathcal{Z} \,, \; \forall\, i = 1, \ldots, N, \; \forall\, k = 1, \ldots, K, & \text{(4.9b)}
\end{aligned}$$

where $z \in \mathbb{R}^m$ is the uncertain vector, $x_i : \mathbb{R}^m \to \mathbb{R}$ is a second stage decision vector, and where $a_{i,k} \in \mathbb{R}^m$, and $b_{i,k} \in \mathbb{R}$ are known parameters of the model. Note that, under some easily identifiable conditions, the above problem is the adjustable robust counterpart of

$$\begin{aligned}
\underset{x \in \mathbb{R}^N}{\text{minimize}} \quad & \sum_{i=1}^{N} (x_i - 1) \\
\text{subject to} \quad & x_i \geq a_{i,k}^T z + b_{i,k} \,, \; \forall\, i = 1, \ldots, N, \; \forall\, k = 1, \ldots, K \,.
\end{aligned}$$

3-SAT problem: Let W be a collection of disjunctive clauses $W = \{w_1, w_2, ..., w_N\}$ on a finite set of variables $V = \{v_1, v_2, ..., v_m\}$ such that $|w_i| = 3\ \forall i \in \{1, ..., N\}$. Let each clause be of the form $w = v_i \vee v_j \vee \bar{v}_k$, where $\bar{v}$ is the negation of v. Is there a truth assignment for V that satisfies all the clauses in W?

Given an instance of the 3-SAT Problem, we can attempt to verify whether the optimal value of the following problem is larger or equal to 0

$$\max_{z} \quad \sum_{i=1}^{N} (h_i(z) - 1) \tag{4.10a}$$

$$\text{subject to} \quad 0 \leq z \leq 1\ , \tag{4.10b}$$

where $z \in \mathbb{R}^m$, and where $h_i(z) := \max\{z_{j_1}; z_{j_2}; 1 - z_{j_3}\}$ if the i-th clause is $w_i = v_{j_1} \vee v_{j_2} \vee \bar{v}_{j_3}$. It is straightforward to confirm that $\{z \in \mathbb{R}^n \,|\, 0 \leq z \leq 1\}$ is a polyhedron and that each $h_i(z)$ can be expressed as $h_i(z) := \max_k a_{i,k}^T \boldsymbol{z} + b_{i,k}$. Hence, we have that problem (4.10) can be expressed in the form of problem (4.9). Finally, we have that the answer to the 3-SAT problem is positive if and only if the optimal value of an instance of problem (4.9) achieves an optimal value greater or equal to 0. □

4.5 No value in delaying decisions

Actually, although the general prognostic of computational tractability of multi-stage ARC solutions is somewhat negative, there are a few circumstances where these solutions can be obtained easily. Among these circumstances are those where non-adjusted solutions are optimal thus making the multi-stage ARC model reduce to be equivalent to a model in which all decisions x_t are independent of the observations that were made, i.e. $x_t^*(v(z)) = x_t^*(v(z'))$ for all z and z' in $\mathcal{Z}$. Note that these are the situations in which the robust counterpart

$$(RC) \quad \underset{\{x_t\}_{t=1}^{T}}{\text{maximize}} \quad \inf_{z \in \mathcal{Z}} \sum_{t=1}^{T} c_t(z)^T x_t + d(z) \tag{4.11a}$$

$$\text{subject to} \quad \sum_{t=1}^{T} a_{jt}(z)^T x_t \leq b_j(z)\ ,\ \forall\, z \in \mathcal{Z}\ ,\ \forall\, j = 1, \ldots, J, \tag{4.11b}$$

can be used to obtain a fixed policy that actually achieves the same worst-case performance as an optimally adjusted policy. Here are the conditions that were established in [11] that can help identify such situations.

Theorem 4.3. : *The multi-stage ARC model presented in equation* (4.8) *is equivalent to the simpler robust counterpart model* (4.11) *where each $x_t \in \mathbb{R}^n$ is independent of z, when there exists a partition of the uncertain vector z as $z := [\ z_0 \quad z_1 \quad z_2 \quad \cdots \quad z_J\ ]^T$ such that*

1. *There exists nonempty convex compact sets* $\mathcal{Z}_j \subset \mathbb{R}^{\dim z_j}$ *such that*

$$\mathcal{Z} := \mathcal{Z}_0 \times \mathcal{Z}_1 \times \mathcal{Z}_2 \times \cdots \times \mathcal{Z}_J = \{z \,|\, \exists z_i \in \mathcal{Z}_i, \forall\, i,\, z = [\begin{array}{ccccc} z_0 & z_1 & z_2 & \cdots & z_J \end{array}]^T\}$$

2. *There exists some* $M > 0$ *such that any feasible* $\{x_t(\cdot)\}_{t=1}^T$ *is such that the condition* $\|x_t(v_t(z))\|_\infty \leq M$ *for all* $z \in \mathcal{Z}$ *is either explicitly or implicitly imposed*

3. *The objective function is a function of* z_0*, namely* $c_t(z) = c_t(z_0)$ *and* $d(z) = d(z_0)$

4. *The functions involved in defining each constraint* j *only depend on* z_j*, namely* $a_{jt}(z) = a_{jt}(z_j)$ *and* $b_{jt}(z) = b_{jt}(z_j)$

Note that the theorem above is slightly more general than what is presented in [11] as it accounts for multi-stage problems. We believe our proof is also simpler to follow as it exploits a famous theorem that originates from zero-sum games called Sion's minimax theorem.

Lemma 4.4. :*(Sion's minimax theorem [44]) Let* $\mathcal{X} \subset \mathbb{R}^n$ *be a convex set and* $\mathcal{Z} \in \mathbb{R}^m$ *be a compact convex set, and let* h *be a real-valued function on* $\mathcal{X} \times \mathcal{Z}$ *with*

1. $h(x, \cdot)$ *lower semicontinuous and quasi-convex on* $\mathcal{Z}$*,* $\forall\, x \in \mathcal{X}$

2. $h(\cdot, z)$ *upper semicontinuous and quasiconcave on* $\mathcal{X}$*,* $\forall\, z \in \mathcal{Z}$

then

$$\sup_{x \in \mathcal{X}} \min_{z \in \mathcal{Z}} \; h(x, z) = \min_{z \in \mathcal{Z}} \sup_{x \in \mathcal{X}} \; h(x, z) \; .$$

In particular, the conclusion is valid if instead of conditions 1 and 2, one can verify that $h(x, \cdot)$ *is convex on* $\mathcal{Z}$ *for all* $x \in \mathcal{X}$*, and* $h(\cdot, z)$ *is concave on* $\mathcal{X}$ *for all* $z \in \mathcal{Z}$*.*

We are now ready for the proof of theorem 4.3.

Proof. We will restrict our attention to the case where the multi-stage ARC model is feasible, otherwise the two problems are necessarily infeasible and thus equivalent.

Step #1: Full adjustability In this context, we start our proof by demonstrating this theorem in the case where $v_t(z) := z$ for all t (i.e. all decisions, even x_1, can use the information about the exact realization of z). In this case the multi-stage ARC model is presented as

$$\begin{array}{ll}
\underset{\{x_t(\cdot)\}_{t=1}^T}{\text{maximize}} & \inf_{z \in \mathcal{Z}} \displaystyle\sum_{t=1}^T c_t(z_0)^T x_t(z) + d(z_0) \\
\text{subject to} & \displaystyle\sum_{t=1}^T a_{jt}(z_j)^T x_t(z) \leq b_j(z_j) \,,\, \forall\, z \in \mathcal{Z} \,,\, \forall\, j = 1, \ldots, J \\
& \|x_t(z)\|_\infty \leq M \,,\, \forall\, t,\, \forall\, z \in \mathcal{Z}
\end{array}$$

where we made explicit the dependence of c_t, d, a_{jt}, and b_j on each of the members of $\{z_0, z_1, \cdots, z_J\}$. The optimal value of this model is equivalent to the optimal value (that we will call ψ) of the following robust two-stage problem

$$\psi := \min_{z\in\mathcal{Z}} \quad h(z) \ ,$$

where $h(z)$ is defined as

$$\begin{aligned} h(z) := \quad \max_{\{x_t\}_{t=1}^T} \quad & \sum_{t=1}^T c_t(z_0)^T x_t + d(z_0) \\ \text{subject to} \quad & \sum_{t=1}^T a_{jt}(z_j)^T x_t \le b_j(z_j) \ , \ \forall\, j = 1, \ldots, J \\ & \|x_t\|_\infty \le M \ , \ \forall\, t \ . \end{aligned}$$

By formulating the Lagrangian function of the inner maximization problem associated to $h(z)$, we obtain that

$$h(z) = \max_{\{x_t\}_{t=1}^T : \|x_t\|_\infty \le M, \forall t} \ \inf_{\lambda \ge 0} \sum_{t=1}^T c_t(z_0)^T x_t + d(z_0) + \sum_j \lambda_j (b_j(z_j) - \sum_t a_{jt}(z_j) x_t)$$

Referring to Sion's minimax theorem, we can verify that the Lagrangian function presented here is affine in both $\{x_t\}_{t=1}^T$ and λ, and that the feasible set for $\{x_t\}_{t=1}^T$ is compact. Hence, we can conclude that

$$h(z) = \inf_{\lambda\ge 0} \ \max_{\{x_t\}_{t=1}^T : \|x_t\|_\infty \le M, \forall t} \ \sum_{t=1}^T c_t(z_0)^T x_t + d(z_0) + \sum_j \lambda_j (b_j(z_j) - \sum_t a_{jt}(z_j) x_t) \ .$$

Since the multi-stage ARC model is feasible, it must be that $h(z)$ is finite hence that the infimum in λ is achieved[1]

$$h(z) = \min_{\lambda\ge 0} \ \max_{\{x_t\}_{t=1}^T : \|x_t\|_\infty \le M, \forall t} \ \sum_{t=1}^T c_t(z_0)^T x_t + d(z_0) + \sum_j \lambda_j (b_j(z_j) - \sum_t a_{jt}(z_j) x_t) \ .$$

Going once step back we get that

$$\psi = \min_{z\in\mathcal{Z}} \ h(z) = \min_{\lambda\ge 0} \min_{z\in\mathcal{Z}} \ \max_{\{x_t\}_{t=1}^T : \|x_t\|_\infty \le M, \forall t} \ \sum_{t=1}^T c_t(z_0)^T x_t + d(z_0) + \sum_j \lambda_j (b_j(z_j) - \sum_t a_{jt}(z_j) x_t) \ ,$$

since the order of two min operators can always be inverted. Now looking into the "$\min_z \ \max_x$" expression, we can once again apply Sion's minimax theorem to get that

$$\psi = \min_{\lambda\ge 0} \ \max_{\{x_t\}_{t=1}^T : \|x_t\|_\infty \le M, \forall t} \ \min_{z\in\mathcal{Z}} \ \sum_{t=1}^T c_t(z_0)^T x_t + d(z_0) + \sum_j \lambda_j (b_j(z_j) - \sum_t a_{jt}(z_j) x_t) \ .$$

[1] Notice that the infimum is taken over a lower semi-continuous function.

Yet, it just so happens that the minimization in z can decompose over the z_j's so that

$$\psi = \min_{\lambda \geq 0} \max_{\{x_t\}_{t=1}^T : \|x_t\|_\infty \leq M, \forall t} \sum_{t=1}^T \min_{z_0 \in \mathcal{Z}_0} c_t(z_0)^T x_t + d(z_0) + \sum_j \lambda_j \min_{z_j \in \mathcal{Z}_j} (b_j(z_j) - \sum_t a_{jt}(z_j) x_t) \ .$$

And as things always comes in threes, a third application of Sion's minimax will reveal that

$$\psi = \max_{\{x_t\}_{t=1}^T : \|x_t\|_\infty \leq M, \forall t} \min_{\lambda \geq 0} \sum_{t=1}^T \min_{z_0 \in \mathcal{Z}_0} c_t(z_0)^T x_t + d(z_0) + \sum_j \lambda_j \min_{z_j \in \mathcal{Z}_j} (b_j(z_j) - \sum_t a_{jt}(z_j) x_t) \ .$$

Note that this third application is made possible by the fact that the function that is implicated is affine in λ and concave in x_t. This would not be the case if the minimization in z did not decompose over independent minimizations in each z_j.

Observing that the expression we get for ψ takes the shape of a Lagrangian function, we can reformulate it as

$$\begin{aligned} \psi = \underset{\{x_t\}_{t=1}^T}{\text{maximize}} \quad & \min_{z_0 \in \mathcal{Z}_0} \sum_{t=1}^T c_t(z_0)^T x_t + d(z_0) \\ \text{subject to} \quad & \sum_{t=1}^T a_{jt}(z_j)^T x_t \leq b_j(z_j) \ , \ \forall\, z_j \in \mathcal{Z}_j \ , \ \forall\, j = 1, \ldots, J \\ & \|x_t\|_\infty \leq M \ , \ \forall\, t, \ \forall\, z \in \mathcal{Z} \end{aligned}$$

where the optimization now considers a set of decisions $\{x_t\}_{t=1}^T$ that is fixed prior to observing the realization of $z \in \mathcal{Z}$. This final optimization model clearly takes the shape of the simpler robust counterpart form (4.11).

Step #2: Partial adjustability In the case of partial adjustability, i.e. when some $v_t(z) :\neq z$, it is obviously possible (and possibly sub-optimal) to simply optimize over policies that do not adjust with respect to $v_t(z)$ and which would achieve

$$\begin{aligned} \underset{\{x_t\}_{t=1}^T}{\text{maximize}} \quad & \inf_{z \in \mathcal{Z}} \sum_{t=1}^T c_t(z)^T x_t + d(z) \\ \text{subject to} \quad & \sum_{t=1}^T a_{jt}(z)^T x_t \leq b_j(z) \ , \ \forall\, z \in \mathcal{Z} \ , \ \forall\, j = 1, \ldots, J \\ & \|x_t\|_\infty \leq M \ , \ \forall\, t, \ \forall\, z \in \mathcal{Z} \ . \end{aligned}$$

However, we just showed that the performance of this policy is exactly the same as the maximum performance that could be achieved if the policy was fully adjustable. It is therefore clear that partial adjustability cannot do better then that. This completes our proof. $\square$

Remark 4.5. : It is worth being aware that in [33] the authors have further studied what are conditions under which the adjustable and static solutions are equivalent. Furthermore, in [16] and some follow up work [17], the authors made further progresses in establishing conditions under which the solution of the non-adjustable robust counterpart model (4.11) can be considered to perform relatively well compared to a problem where decision variables are instead considered adjustable. In particular, they are able to identify general conditions under which the relative sub-optimality of such here-and-now decisions is bounded by a factor of two.

4.6 Exact solution methods for Robust Two-stage problems

In this section, we expose some algorithms that have been proposed to obtain exact solutions to two-stage adjustable robust counterpart model with "relatively complete" and "fixed" recourse. In particular, we are interested in obtaining the optimal "first-stage" decision x for the following problem:

$$\begin{array}{ll} \underset{x,y(\cdot)}{\text{maximize}} & \inf_{z\in\mathcal{Z}} c_1(z)^T x + c_2^T y(z) + d(z) \\ \text{subject to} & a_{j1}(z)^T x + a_{j2}^T y(z) \leq b_j(z)\,,\ \forall\, z \in \mathcal{Z}\,,\ \forall\, j = 1,\ldots,J \\ & x \in \mathcal{X}\ , \end{array}$$

where $x \in \mathbb{R}^n$ is the decision that needs to be initially implemented while $y \in \mathbb{R}^n$ is implemented once z is known. Also, one might note that the effect of the recourse decision variables $y(\cdot)$ is not affected by uncertainty (i.e. $c_2(z) := c_2$ and $a_{j2}(z) := a_{j2}$), a property commonly referred as "fixed recourse". For simplicity of exposure, it is also commonly assumed that the feasible set $\mathcal{X}$ is such that it guarantees that it is always possible to identify a recourse action y that will satisfy all the constraints, a property commonly referred as "relatively complete recourse". In other words,

$$\mathcal{X} \subseteq \{x \in \mathbb{R}^n \,|\, \forall\, z \in \mathcal{Z}, \exists y \in \mathbb{R}^n, a_{j1}(z)^T x + a_{j2}^T y \leq b_j(z)\,,\ \forall\, j = 1,\ldots,J\}\ .$$

In what follows we will mostly refer to the following representation of the model

$$(TSARC) \quad \underset{x\in\mathcal{X}}{\text{maximize}} \quad \inf_{z\in\mathcal{Z}}\ h(x,z)\ , \tag{4.12}$$

where we have that

$$\begin{array}{rl} h(x,z) := \max\limits_{y} & c_1(z)^T x + c_2^T y + d(z) \\ \text{subject to} & a_{j1}(z)^T x + a_{j2}^T y \leq b_j(z)\,,\ \forall\, j = 1,\ldots,J. \end{array}$$

This latter form makes explicit the fact that we are solely interested in an optimal first-stage decision x^* and its optimal worst-case total revenue $\inf_{z\in\mathcal{Z}}\ h(x^*,z)$. Note that once the realized z is known, it is possible to implement an optimal recourse policy simply by reoptimizing the second stage problem involved in evaluating $h(x^*, z)$.

Although, given the NP-hardness of the problem, there is in general no guarantee for either of the methods presented below to return an exactly optimal solution in a reasonable amount of time, some of these algorithms have been applied successfully to applications of practical sizes. We will later explore approximation schemes that are considered more tractable then these exact methods yet, before deploying such approximation schemes on large problems, it is usually interesting to confirm the quality of approximate solutions on small problem instances where exact solutions can be identified.

4.6.1 Vertex enumeration method

We first present the most straightforward way of solving two-stage robust counterpart problems in cases where it is possible to define $\mathcal{Z}$ as the convex hull of a certain number of points: $\mathcal{Z} := \text{ConvexHull}(\{\bar{z}_1, \bar{z}_2, \ldots, \bar{z}_K\})$.

Theorem 4.6. : *Assume that the uncertainty set $\mathcal{Z}$ is given as the convex hull of a finite set:*

$$\mathcal{Z} := \mathit{ConvexHull}(\{\bar{z}_1, \bar{z}_2, \ldots, \bar{z}_K\}) \, .$$

Then, the TSARC presented in problem (4.12) *is equivalent to*

$$\begin{aligned}
&\underset{x, \{y_k\}_{k=1}^K}{\text{maximize}} && \min_k \; c_1(\bar{z}_k)^T x + c_2^T y_k + d(\bar{z}_k) && (4.13a)\\
&\text{subject to} && a_{j1}(\bar{z}_k)^T x + a_{j2}^T y_k \le b_j(\bar{z}_k)\,,\ \forall\, k = 1, \ldots, K,\ \forall\, j = 1, \ldots, J && (4.13b)\\
& && x \in \mathcal{X} \, . && (4.13c)
\end{aligned}$$

To prove this theorem we will need to make use of the following lemma.

Lemma 4.7. : *Assume that the uncertainty set $\mathcal{Z}$ is given as the convex hull of a finite set :*

$$\mathcal{Z} := \mathit{ConvexHull}(\{\bar{z}_1, \bar{z}_2, \ldots, \bar{z}_K\}) \, .$$

Then, given a concave function $h(z)$ over $\mathcal{Z}$, the optimal value of $\min_{z \in \mathcal{Z}} h(z)$ *is equal to* $\min_{k \in \{1,2,\ldots,K\}} h(\bar{z}_k)$.

Proof. Let z^* be an optimal value of $\min_{z \in \mathcal{Z}}\ h(z)$, then since $z \in \mathcal{Z}$ it is necessarily the convex combination of the points in $\{\bar{z}_1, \bar{z}_2, \ldots, \bar{z}_K\}$. Namely, there must exist $\theta \in \mathbb{R}^K$, such that $\theta \ge 0$ and $\sum_k \theta_k = 1$, for which

$$z^* = \sum_k \bar{z}_k \theta_k \, .$$

Since this is the case, by the concavity of $h(z)$ it must also be that

$$\min_{z \in \mathcal{Z}} h(z) = h(z^*) = h(\sum_k \bar{z}_k \theta_k) \ge \sum_k \theta_k h(\bar{z}_k) \ge \min_k h(\bar{z}_k) \ge \min_{z \in \mathcal{Z}} h(z) \, ,$$

where we used, in order, Jensen's inequality for concave functions, the properties of θ, and the fact the each $\bar{z}_k$ is a member of $\mathcal{Z}$. Hence, we have that $\min_{z\in\mathcal{Z}} h(z) = \min_k h(\bar{z}_k)$. $\square$

We pursue with the proof of theorem 4.6.

Proof. For any fixed $x \in \mathcal{X}$ and $z \in \mathcal{Z}$, by duality of linear programs, $h(x, z)$ can be shown equal to the optimal value of

$$\begin{aligned} h(x,z) = \min_{\lambda} \quad & c_1(z)^T x + d(z) + \sum_j \lambda_j (b_j(z) - a_{j1}(z)^T x) \\ \text{subject to} \quad & c_2 = \sum_j a_{j2}\lambda_j \\ & \lambda \geq 0 \; , \end{aligned}$$

where $\lambda \in \mathbb{R}^J$. Strong duality applies here because of our assumption of relatively complete recourse which ensures that the maximization problem evaluated in $h(x, z)$ is feasible for any $x \in \mathcal{X}$ and $z \in \mathcal{Z}$. Presented in this form, we realize that $h(x, z)$ is actually a concave function of z, since it is the minimum of a set of affine functions, each indexed by a λ in some feasible set. Since $h(x, \cdot)$ is a concave function by lemma 4.7

$$\min_{z\in\mathcal{Z}} \; h(x, z) = \min_k h(x, \bar{z}_k) \; .$$

But then, in this form, we would only need to be robust with respect to the K scenarios for z which only imply the recourse action employed under those circumstances. This is how one obtains that problem (4.13) is equivalent to the TSARC model in terms of identifying x^* and its worst-case cost. $\square$

Note that in general the number of vertices of a convex polyhedron might be exponential with respect to the number of faces that describes it. In particular, the box uncertainty set in $\mathbb{R}^m$ is defined by $2m$ faces but involves 2^m vertices. This observations therefore limits significantly the use of the vertex enumeration method in order to solve the TSARC model of realistic sizes. Fortunately, in [53], the authors found a way to speed up the resolution of this enumeration method when the list of vertices has an exponential size.

4.6.2 Column-and-constraint generation method

In [53], the authors propose an iterative method that has the potential to identify a subset of vertices of $\mathcal{Z}$ for which a "reduced" vertex enumeration method can be employed and effectively return an optimal first-stage solution. In particular, the idea behind the method referred as "column-and-constraint generation" is to approximate problem (4.13) with

$$\begin{aligned}
\underset{x,s,\{y_k\}_{k=1}^{K'}}{\text{maximize}} \quad & s && \text{(4.14a)}\\
\text{subject to} \quad & s \leq c_1(\bar{z}'_k)^T x + c_2^T y_k + d(\bar{z}'_k)\,,\ \forall\, k = 1,\ldots,K' && \text{(4.14b)}\\
& a_{j1}(\bar{z}'_k)^T x + a_{j2}^T y_k \leq b_j(\bar{z}'_k)\,,\ \forall\, k = 1,\ldots,K',\ \forall\, j = 1,\ldots,J && \text{(4.14c)}\\
& x \in \mathcal{X}\ , && \text{(4.14d)}
\end{aligned}$$

where $\bar{z}'_k$ are member of $\mathcal{Z}'_v := \{\bar{z}'_1, \bar{z}'_2, \ldots, \bar{z}'_{K'}\}$, a set of reasonable size even in large practical problems.

Note first that if $\mathcal{Z}'_v$ is a subset of the vertices of $\mathcal{Z}$ then the optimal value of problem (4.14) (let's call it $\hat{s}$) is necessarily an upper bound to the optimal value (let's call it s^*) of the TSARC problem, i.e. $\hat{s} \geq s^*$. This is due to the fact that the TSARC problem is equivalent to

$$\begin{aligned}
\underset{x,s,\{y_k\}_{k=1}^{K}}{\text{maximize}} \quad & s && \text{(4.15a)}\\
\text{subject to} \quad & s \leq c_1(\bar{z}_k)^T x + c_2^T y_k + d(\bar{z}_k)\,,\ \forall\, k = 1,2,\ldots,K && \text{(4.15b)}\\
& a_{j1}(\bar{z}_k)^T x + a_{j2}^T y_k \leq b_j(\bar{z}_k)\,,\ \forall\, k = 1,2,\ldots,K,\ \forall\, j = 1,\ldots,J && \text{(4.15c)}\\
& x \in \mathcal{X}\ , && \text{(4.15d)}
\end{aligned}$$

where $\mathcal{Z}_v := \{\bar{z}_1, \bar{z}_2, \ldots, \bar{z}_K\}$ is the set of vertices of $\mathcal{Z}$. Since this problem involves all vertices which covers the subset $\mathcal{Z}'_v$ involved in problem (4.14) any of its optimal solution can be used to create an optimal solution for problem (4.14) which achieves the same objective value. Note that the opposite is not always true.

Now, given an optimal solution $(\hat{x}, \hat{s}, \{\hat{y}_k\}_{k=1}^{K'})$ to problem (4.14), let $\hat{z} := \text{argmin}_{z\in\mathcal{Z}}\, h(\hat{x}, z)$ then one can show that

- either $h(\hat{x}, \hat{z}) = \hat{s}$ which would indicate that $\hat{x}$ is optimal with respect to (4.15) since $s^* = \max_{x\in\mathcal{X}} \min_{z\in\mathcal{Z}} h(x, z) \geq \min_{z\in\mathcal{Z}} h(\hat{x}, z) = \hat{s} \geq s^*$
- or $h(\hat{x}, \hat{z}) < \hat{s}$ and $\hat{z}$ is a vertex of $\mathcal{Z}$ that is not a member of $\mathcal{Z}'_v$ which can be added to $\mathcal{Z}'_v$ to generate a tighter approximation.[2]

Based on this analysis, one can design the following procedure:

1. Take any $\hat{x} \in \mathcal{X}$
2. Identify $\hat{z} := \text{argmin}_{z\in\mathcal{Z}}\, h(\hat{x}, z)$ and construct $\mathcal{Z}'_v := \{\hat{z}\}$
3. Iterate until algorithm converged:
   (a) Solve problem (4.14) to obtain $\hat{x}$ and $\hat{s}$

[2]Note that theoretically, depending on the solution scheme, $\hat{z}$ might not be a vertex of $\mathcal{Z}$. However, whenever it is the case, there always exists an another optimal solution at a vertex.

(b) Identify $\hat{z} := \mathrm{argmin}_{z\in\mathcal{Z}} h(\hat{x}, z)$, if $h(\hat{x}, \hat{z}) = \hat{s}$ then the algorithm has converged, otherwise add $\hat{z}$ to $\mathcal{Z}'_v$ and iterate

This algorithm is guaranteed to converge in a finite number of iterations since all bounded polyhedron described by a finite number of linear constraints have a finite number of vertices thus the algorithm will converge after a number of iterations that is necessarily lesser or equal to the number of vertices of $\mathcal{Z}$. The hope is that in practice, one needs much less iterations to obtain an optimal solution.

The difficulty that remains to resolve is how to solve $\min_{z\in\mathcal{Z}} h(\hat{x}, z)$. This is an important step as it will most likely be the bottleneck of the algorithm. Hence, although we suggest a method below, one should invest special efforts in establishing for the application that interest him whether there is a more efficient procedure to do so. Our method will exploit what are known as the Karush-Kuhn-Tucker (KKT) conditions of optimality.

Lemma 4.8. :*(Karush-Kuhn-Tucker conditions, see section 5.5.3 of [22]) Given a convex optimization problem*

$$\begin{aligned} \underset{x}{\text{maximize}} \quad & f(x) \\ \text{subject to} \quad & g_j(x) \leq 0\,,\ \forall\, j = 1, 2, \ldots, J\,, \end{aligned}$$

where $x \in \mathbb{R}^n$, $f(x)$ is a concave differentiable function, $g_j(x)$ are convex differentiable functions for all $j = 1, 2, \ldots, J$. If this optimization problem satisfies strong duality, then any primal dual optimal solution pair must satisfy the following conditions

$$\begin{aligned} & g_j(\tilde{x}) \leq 0 \\ & \tilde{\lambda} \geq 0 \\ & \tilde{\lambda}_j g_j(\tilde{x}) = 0\,,\ \forall\, j = 1, 2, \ldots, J \\ & \nabla f(\tilde{x}) = \sum_j \tilde{\lambda}_j \nabla g_j(\tilde{x}) \end{aligned}$$

where $\nabla f(\tilde{x})$ refers to the gradient of $f(\cdot)$ at $\tilde{x}$ and similarly for $\nabla g_j(\tilde{x})$. Conversely, let $\tilde{x} \in \mathbb{R}^n$ and $\tilde{\lambda} \in \mathbb{R}^J$ be any points that satisfy the above conditions, then $\tilde{x}$ and $\tilde{\lambda}$ are primal and dual optimal, with zero duality gap.

In case of linear programming the KKT condition can be reformulated as follows.

Corollary 4.9. : *Given a linear programming problem*

$$\begin{aligned} \underset{y}{\text{maximize}} \quad & c^T y \\ \text{subject to} \quad & Ay \leq b\,, \end{aligned}$$

where $y \in \mathbb{R}^n$. If this optimization problem satisfies strong duality, then any primal dual optimal solution pair must satisfy the following conditions

$$\begin{aligned}
&A\tilde{y} \leq b \\
&\tilde{\lambda} \geq 0 \\
&\tilde{\lambda}_j(a_j^T\tilde{y} - b_j) = 0\,,\ \forall\, j = 1, 2, \ldots, J \\
&c = \sum_j \tilde{\lambda}_j a_j\,.
\end{aligned}$$

Conversely, let $\tilde{y} \in \mathbb{R}^n$ and $\tilde{\lambda} \in \mathbb{R}^J$ be any points that satisfy the above conditions, then $\tilde{y}$ and $\tilde{\lambda}$ are primal and dual optimal, with zero duality gap.

Hence, if we look at the optimization problem that evaluates $h(x, z)$ for some $x \in \mathcal{X}$ and some $z \in \mathcal{Z}$, since relatively complete recourse implies that this optimization model is feasible, this implies that the duality gap is zero hence that if the optimal value is finite[3] then it is achieved by a primal dual pair of variables that satisfy the KKT conditions above. In particular,

$$h(x, z) = c_1(z)^T x + c_2^T y + d(z)$$

for any pair (y, λ) such that

$$\begin{aligned}
&a_{j1}(z)^T x + a_{j2}^T y \leq b_j(z)\,,\ \forall\, j = 1, \ldots, J && \text{(4.16a)}\\
&\lambda \geq 0 && \text{(4.16b)}\\
&\lambda_j(a_{j1}(z)^T x + a_{j2}^T y - b_j(z)) = 0\,,\ \forall\, j = 1, \ldots, J && \text{(4.16c)}\\
&c_2 = \sum_j a_{j2}\lambda_j\,. && \text{(4.16d)}
\end{aligned}$$

In other words,

$$\begin{aligned}
\min_{z \in \mathcal{Z}} h(x, z) := \min_{z \in \mathcal{Z}, y, \lambda, u}\ & c_1(z)^T x + c_2^T y + d(z) \\
& a_{j1}(z)^T x + a_{j2}^T y \leq b_j(z)\,,\ \forall\, j = 1, \ldots, J \\
& \lambda \geq 0 \\
& \lambda_j \leq M u_j\,,\ \forall\, j = 1, \ldots, J \\
& b_j(z) - a_{j1}(z)^T x - a_{j2}^T y \leq M(1 - u_j)\,,\ \forall\, j = 1, \ldots, J \\
& c_2 = \sum_j a_{j2}\lambda_j \\
& u \in \{0, 1\}^J\,,
\end{aligned}$$

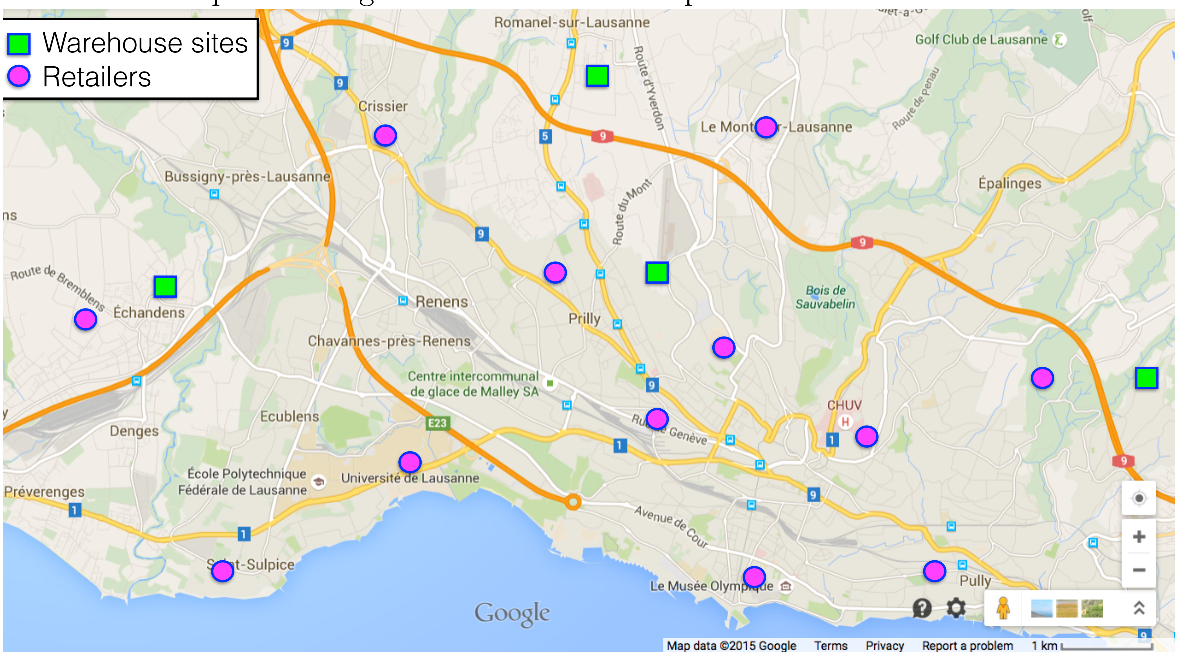
where M is some large positive constant, and where each u_j is a binary variable that was introduced to linearize the complementarity slackness conditions (4.16c) at the price of converting the problem into a mixed-integer linear program.

[3]This is necessarily the case for instance when the TSARC is known not to be unbounded, i.e. $\forall\, x \in \mathcal{X}$, $\exists\, z \in \mathcal{Z}$, $h(x, z) < \infty$.

4.7 Exercise: Facility location problem

In this exercise, you will implement the two solution schemes described above on a facility location-transportation problem. In particular, we consider a company that wishes to acquire some warehouses to produce the goods that will be distributed to the retailers. These retailers are located at a number of locations on a map and a number of candidate sites have already been selected.

Map indicating retailer locations and possible warehouse sites



For simplicity, we will assume that the maximum production that can be achieved at each site is already predetermined. This leads to a robust two-stage optimization problem in which the company must first decide which site to acquire, and later decide, once the local demand for goods is known, how many goods to produce from each sites and to be delivered to each retailers.

In particular, we will study the following model:

$$\begin{aligned}
\underset{x,y}{\text{maximize}} \quad & -\sum_{i=1}^{n} c_i x_i + \sum_{i=1}^{n}\sum_{j=1}^{m}(r_{ij} - d_{ij})y_{ij} && (4.17a)\\
\text{subject to} \quad & \sum_{i=1}^{n} y_{ij} \leq D_j \,, \; \forall\, j && (4.17b)\\
& \sum_{j=1}^{m} y_{ij} \leq P_i x_i \,, \; \forall\, i && (4.17c)\\
& y_{ij} \geq 0 \; \forall\, i, j && (4.17d)\\
& x \in \{0,1\}^n \,, && (4.17e)
\end{aligned}$$

where $x \in \{0,1\}^n$ describes which locations are acquired (i.e. $x_i = 1$ if location i is acquired), $y_{ij} \in \mathbb{R}$ describes how many goods are produced at warehouse i to satisfy the demand at location j. The objective function accounts for the fact that it costs c_i to acquire the warehouse at site i, a revenue of r_{ij} is obtained for delivering one unit of good to a customer at retailer location j from warehouse i, and d_{ij} accounts for the per unit cost of producing a good at warehouse i and transporting it to retailer j. We also account in this model for the fact that the production and delivery of goods must be such that we never hold more goods than there is demand D_j at a retailer site, and that each warehouse (if acquired) does not produce more than it is able to, namely P_i units. Tables 4.1, 4.2, and 4.3 below present the detailed parameter values. Note that all amount are in millions (i.e. installation cost is in million of dollars, capacity and demand is in million units of goods, while transportation cost and sale price are in dollars per units.

Table 4.1: Facility locations

	Installation cost	Capacity
Location #1	9.1	23
Location #2	8.0	168
Location #3	4.5	110
Location #4	2.1	295

Table 4.2: Retailer locations

	Nominal demand	Max deviation	Retail price
Retailer #1	24	18	2
Retailer #2	12	1	2
Retailer #3	18	14	2
Retailer #4	23	12	2
Retailer #5	24	13	2
Retailer #6	13	5	2
Retailer #7	11	6	2
Retailer #8	9	0	2
Retailer #9	18	4	2
Retailer #10	25	23	2
Retailer #11	25	21	2
Retailer #12	23	20	2

Table 4.3: Transportation costs from facility to retailers

	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12
#1	2.31	2.37	1.89	1.92	1.98	1.69	2.37	2.14	2.87	2.16	2.15	1.52
#2	1.88	2.36	2.02	2.77	1.17	1.45	3.64	1.45	1.83	1.80	1.74	2.42
#3	2.51	1.73	3.50	2.39	2.51	2.50	3.08	2.36	2.35	1.72	1.47	2.10
#4	1.71	2.99	1.40	0.96	1.79	1.81	1.89	2.01	2.28	1.71	2.98	2.66

The robust counterpart of this model will account for uncertainty about the demand vector D, and for the fact that each y_{ij} can be adapted to the overall demand D. Specifically, the adjustable robust optimization model takes the form:

$$
\begin{aligned}
\underset{x,y(\cdot)}{\text{maximize}} \quad & \inf_{z\in\mathcal{Z}(\Gamma)} -\sum_{i=1}^{n} c_i x_i + \sum_{i=1}^{n}\sum_{j=1}^{m}(r_{ij}-d_{ij})y_{ij}(z) && \text{(4.18a)}\\
\text{subject to} \quad & \sum_i y_{ij}(z) \le \bar{D}_j + \hat{D}_j z_j\,,\ \forall\, z\in\mathcal{Z}(\Gamma)\,,\ \forall\, j && \text{(4.18b)}\\
& \sum_j y_{ij}(z) \le P_i x_i\,,\ \forall\, z\in\mathcal{Z}(\Gamma)\,,\ \forall\, i && \text{(4.18c)}\\
& y_{ij}(z) \ge 0\,,\ \forall\, z\in\mathcal{Z}(\Gamma)\,,\ \forall\, i,j && \text{(4.18d)}\\
& x \in \{0,1\}^n\,, && \text{(4.18e)}
\end{aligned}
$$

where $\bar{D}$ is the nominal demand vector, $\hat{D}_j$ expresses what is the maximum deviation in demand one expects to see from the nominal amount at location j, and where $\mathcal{Z}(\Gamma)$ is the budgeted uncertainty set, in other words

$$\mathcal{Z}(\Gamma) := \left\{ z \in \mathbb{R}^m \,\middle|\, -1 \leq z \leq 1,\, \sum_{j=1}^{m} |z_j| \leq \Gamma \right\} .$$

Exercise 4.1. *Implementing vertex enumeration*
Solve with RSOME (using Google Colab) the robust two-stage optimization problem presented in problem (4.18) using vertex enumeration for the budgeted uncertainty set when $\Gamma = 1$ and $\Gamma = m$.

Exercise 4.2. *RC = multi-stage ARC under* $\Gamma = m$
Use theorem 4.3 to demonstrate that when $\Gamma = m$ problem (4.18) is equivalent to

$$\begin{array}{ll}
\underset{x,y}{\text{maximize}} & -\sum_{i=1}^{n} c_i x_i + \sum_{i=1}^{n}\sum_{j=1}^{m} (r_{ij} - d_{ij}) y_{ij} \\
\text{subject to} & \sum_{i=1}^{n} y_{ij} \leq \bar{D}_j - \hat{D}_j \,,\ \forall\, j \\
& \sum_{j=1}^{m} y_{ij} \leq P_i x_i \,,\ \forall\, i \\
& y_{ij} \geq 0 \ \forall\, i, j \\
& x \in \{0,1\}^n
\end{array}$$

Exercise 4.3. *Implementing column-and-constraint generation*
Solve with RSOME (using Google Colab) the robust two-stage optimization problem presented in problem (4.18) using column-and-constraint generation for the budgeted uncertainty set when $\Gamma = 4$.

Chapter 5

Value of Flexibility Using Tractable Decision Rules

Although the exact solution methods described in the previous chapter are valuable, they are confronted to two limitations. First, it is unclear how such exact methods might generalize to multi-stage problems as their implementation relies heavily on there being three level of optimization $x \rightarrow z \rightarrow y$. The second issue is that one actually has no guarantees that the exact algorithms will converge in a reasonable amount of time. For instances, in a computational study that was done in [1], the authors identified instances where x was fixed, $z \in \mathbb{R}^{64}$, $y \in \mathbb{R}^{64}$ where the optimal value could not be obtained in less than a full day on a powerful computer.

These issues highlight the need for some approximation schemes that can easily be applied on multi-stage problem and for which we can provide guarantees that solutions will be obtained in a reasonable amount of time (namely in an amount of time that grows polynomially in terms of the size of the problem) and with some guarantees in terms of worst-case performance. In what follows, we present the general idea behind an approximation scheme known to employ "affine decision rules". Intuitively, the conjecture behind this scheme is that the difficulty of resolution originates from the fact that each decision rule $x_t(\cdot)$, or $y(\cdot)$ in a two-stage problem, is very hard to optimize because of the complexity of the structure that we allow it to take when in fact a simpler structured decision rule might be almost as effective (see the illustration in figure 5.1).

5.1 Affine decision rules

The idea of using affine decision rules to approximate the multi-stage ARC problem (4.8) was first proposed in [11] for two-stage problems but can easily be generalized to the more general setting. We start by being more precise about what we mean by affine decision rules.

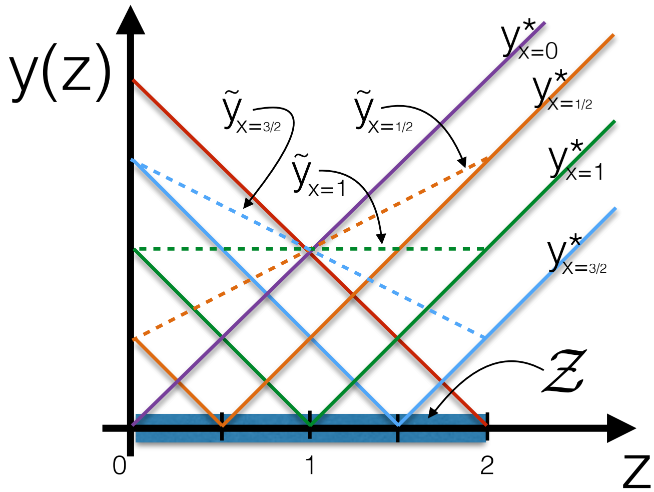


Figure 5.1: Example of decision rules for simple inventory problem from chapter 4.1. Note that the worst-case value over $\mathcal{Z}$ is the same whether we employ a complicated piecewise linear function y^* or a simpler affine function $\tilde{y}$.

Definition 5.1. : Let each observation mapping $v_t : \mathbb{R}^m \to \mathbb{R}^\nu$, the decision rule $x_t : \mathbb{R}^\nu \to \mathbb{R}^n$ is considered an affine decision rule if given any two observation vectors $\bar{v}_1 \in \mathbb{R}^\nu$ and $\bar{v}_2 \in \mathbb{R}^\nu$ and any $\lambda \in \mathbb{R}$, $x_t(\cdot)$ has the property that

$$x_t(\lambda\bar{v}_1 + (1-\lambda)\bar{v}_2) = \lambda x_t(\bar{v}_1) + (1-\lambda)x_t(\bar{v}_2) \; .$$

One can actually show that this is the case if and only if x_t can be represented as

$$x_t(\bar{v}) := x_t + X_t\bar{v}$$

for some $x_t \in \mathbb{R}^n$ and $X_t \in \mathbb{R}^{n\times\nu}$.

The approximation scheme known under the name of Affinely Adjustable Robust Counterpart (AARC) consists of replacing the multi-stage ARC model with an optimization model that will reduce its search to the space of affine decision rules. Namely, this scheme will seek the optimal solution to the following model:

$$(AARC) \tag{5.1a}$$

$$\underset{\{x_t\}_{t=1}^T, \{X_t\}_{t=2}^T}{\text{maximize}} \quad \inf_{z\in\mathcal{Z}} \; c_1(z)^T x_1 + \sum_{t=2}^T c_t(z)^T(x_t + X_t v_t(z)) + d(z) \tag{5.1b}$$

$$\text{subject to} \quad a_{j1}(z)^T x_1 + \sum_{t=2}^T a_{jt}(z)^T(x_t + X_t v_t(z)) \le b_j(z)\,,\; \forall\, z \in \mathcal{Z}\,,\; \forall\, j = 1, \ldots, J \tag{5.1c}$$

where each decision rule $x_t(\cdot)$ was replaced with its affine representation $x_t(\bar{v}) := x_t + X_t\bar{v}$, and where the optimization is now made over the finite dimensional space spanned by the set of decision vectors $x_t \in \mathbb{R}^n$ and decision matrices $X_t \in \mathbb{R}^{n\times\nu}$. However, it is still necessary to identify a tractable reformulation of this model where the worst-case analysis $z \in \mathcal{Z}$ is replaced by a joint optimization of decision variables and worst-case certificates. To do so, we will need to make the following assumption.

Assumption 5.2. : The multi-stage ARC model has fixed recourse and all observations are linear functions of z. In other words, mathematically we make the following assumptions

1. (**Fixed recourse**) For all $t = 2, \ldots, T$ and all $j = 1, \ldots, J$, the affine mappings $c_t(z)$ and $a_{jt}(z)$ are actually constant, i.e. $c_t(z) = c_t$ and $a_{jt}(z) = a_{jt}$.
2. (**Affine observations**) For all $t = 2, \ldots, T$, the observations $v_t(\cdot)$ can be described as $v_t(z) := V_t z$ for some $V_t \in \mathbb{R}^{\nu\times m}$.

Theorem 5.3. : *Let the multi-stage ARC model* (4.8) *satisfy assumption 5.2 and the uncertainty set $\mathcal{Z}$ be a bounded polyhedron that satisfies assumption 2.2, then the AARC model* (5.1) *can be described as the robust linear program:*

$$\underset{\{x_t\}_{t=1}^T,\{X_t\}_{t=2}^T}{\text{maximize}} \quad \inf_{z\in\mathcal{Z}} \; c_1(z)^T x_1 + \sum_{t=2}^T c_t^T(x_t + X_t V_t z) + d(z) \tag{5.2a}$$

$$\text{subject to} \quad a_{j1}(z)^T x_1 + \sum_{t=2}^T a_{jt}^T(x_t + X_t V_t z) \le b_j(z)\,,\; \forall\, z \in \mathcal{Z}\,,\; \forall\, j = 1, \ldots, J \tag{5.2b}$$

and can therefore be reformulated as a linear program with finite number of decisions and finite number of constraints. Furthermore, the optimal affine decision rules and optimal objective value obtained from the AARC model provide a "conservative" approximation of the multi-stage ARC model, i.e. the optimal affine decision rules are necessarily implementable in multi-stage ARC and achieve the approximated objective value, which thus provide a lower bound on the true optimal value of the multi-stage ARC.

Proof. The proof for this theorem is fairly straightforward. Problem (5.2) is simply obtained after replacing both the decision rule $x_t(\cdot)$ and the observation mapping $v_t(\cdot)$ with their affine and linear definition. One can then establish that the objective function and each constraint involve affine functions of the decision variables and of the perturbation z, hence that it is an instance of robust linear program for which we know how to obtain a tractable reformulation (see theorem 2.7).

Next, once the optimal solution is obtained in terms of $\{x_t^*\}_{t=1}^T$ and $\{X_t^*\}_{t=2}^T$ it is possible to construct a decision rule $x_t(\bar{v}) := x_t^* + X_t^* \bar{v}_t$ which will satisfy all constraints of the multi-stage ARC model and for which the objective value reduces to

$$\inf_{z\in\mathcal{Z}} \; c_1(z)^T x_1^* + \sum_{t=2}^T c_t^T(x_t^* + X_t^* V_t z) + d(z)\;,$$

which is exactly the optimal value of the AARC model. □

Example 5.4. : Looking back at the inventory problem presented in example 4.1, where we were trying to identify ordering strategies that are robust to demand uncertainty (as portrayed by $d \in \mathcal{U} \subset \mathbb{R}^m$), we recall formulating the multi-stage adjustable robust counterpart as follows:

$$\begin{array}{ll}\underset{x_1,\{x_t(\cdot)\}_{t=2}^T,\{s_t^+(\cdot),s_t^-(\cdot)\}_{t=1}^T}{\text{minimize}} & \sup_{d\in\mathcal{U}} c_1x_1 + \sum_t c_t x_t(d_{[t-1]}) + h_t s_t^+(d) + b_t s_t^-(d) \\ \text{subject to} & s_t^+(d) \geq 0,\ s_t^-(d) \geq 0\,,\ \forall\, d \in \mathcal{U},\ \forall\, t \\ & s_t^+(d) \geq y_1 + \sum_{t'=1}^{t} x_{t'}(d_{[t'-1]}) - d_{t'}\,,\ \forall\, d \in \mathcal{U},\ \forall\, t \\ & s_t^-(d) \geq -y_1 + \sum_{t'=1}^{t} d_{t'} - x_{t'}(d_{[t'-1]})\,,\ \forall\, d \in \mathcal{U},\ \forall\, t \\ & 0 \leq x_t(d_{[t-1]}) \leq M\,,\ \forall\, d \in \mathcal{U},\ \forall\, t\,. \end{array}$$

One can observe that this multi-stage ARC model satisfies assumption 5.2. Namely, that 1) the recourse is fixed as portrayed by the fact that all x_t are only multiplied to coefficients that are certain; and 2) that the observations are a linear function of the uncertain vector d. Indeed, regarding the latter, we can verify that

$$d_{[t-1]} = \begin{bmatrix} \boldsymbol{I}_{t-1} & \mathbf{0}_{t,T-t+1} \\ \mathbf{0}_{T-t\times t-1} & \mathbf{0}_{T-t\times T-t} \end{bmatrix} d = \begin{bmatrix} d_1 & \ldots & d_{t-1} & 0 & \ldots & 0 \end{bmatrix}^T ,$$

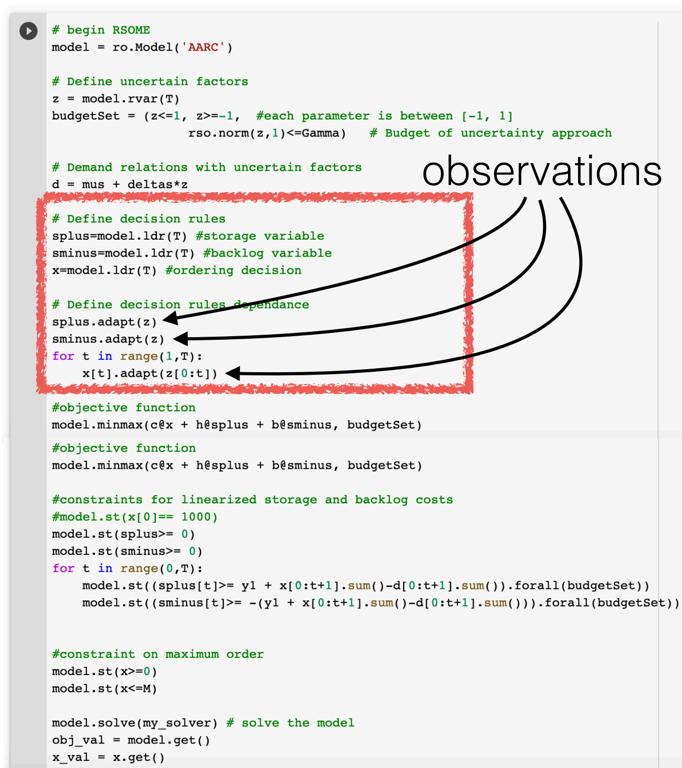
where we padded the unrevealed terms of d with zeros in order to be consistent with $V_t \in \mathbb{R}^{\nu\times m}$.

The affinely adjustable robust counterpart of this inventory model can be presented as

$$\begin{array}{ll}\underset{\substack{x_1,\{x_t,X_t\}_{t=2}^T,\\ \{s_t^+,S_t^+\}_{t=1}^T,\{s_t^-,S_t^-\}_{t=1}^T}}{\text{minimize}} & \sup_{d\in\mathcal{U}} c_1x_1 + \sum_t c_t(x_t + X_tV_td) + h_t(s_t^+ + S_t^+d) + b_t(s_t^- + S_t^-d) \\ \text{subject to} & s_t^+ + S_t^+d \geq 0,\ s_t^- + S_t^-d \geq 0\,,\ \forall\, d \in \mathcal{U},\ \forall\, t \\ & s_t^+ + S_t^+d \geq y_1 + \sum_{t'=1}^{t} x_{t'} + X_{t'}V_{t'}d - d_{t'}\,,\ \forall\, d \in \mathcal{U},\ \forall\, t \\ & s_t^- + S_t^-d \geq -y_1 + \sum_{t'=1}^{t} d_{t'} - (x_{t'} + X_{t'}V_{t'}d)\,,\ \forall\, d \in \mathcal{U},\ \forall\, t \\ & 0 \leq x_t + X_tV_td \leq M\,,\ \forall\, d \in \mathcal{U},\ \forall\, t\,, \end{array}$$

where each $X_t \in \mathbb{R}^{1\times m}$, $S_t^+ \in \mathbb{R}^{1\times m}$, $S_t^- \in \mathbb{R}^{1\times m}$, and where $V_t := \begin{bmatrix} \boldsymbol{I}_{t-1} & \mathbf{0}_{t,T-t+1} \\ \mathbf{0}_{T-t\times t-1} & \mathbf{0}_{T-t\times T-t} \end{bmatrix}$ such that $V_td = \begin{bmatrix} d_1 & \ldots & d_{t-1} & 0 & \ldots & 0 \end{bmatrix}$. Note that the size of this AARC model could be reduced by accounting for the fact that the observation vector v_t is smaller for smaller t. For simplicity of presentation, we choose to leave it this way as we understand that some terms of X_t will always be multiplied to zero and can therefore be set arbitrarily.

Figure 5.2 presents how this is implemented in Python using RSOME. Note how the linear decision rules are defined using the command "*splus=model.ldr(T)*", and how the dependence on the uncertain variables that are observed before implementing the decision are identified using "*splus.adapt(z)*".




```
# begin RSOME
model = ro.Model('AARC')

# Define uncertain factors
z = model.rvar(T)
budgetSet = (z<=1, z>=-1,  #each parameter is between [-1, 1]
                 rso.norm(z,1)<=Gamma)   # Budget of uncertainty approach

# Demand relations with uncertain factors
d = mus + deltas*z

# Define decision rules
splus=model.ldr(T) #storage variable
sminus=model.ldr(T) #backlog variable
x=model.ldr(T) #ordering decision

# Define decision rules dependance
splus.adapt(z)
sminus.adapt(z)
for t in range(1,T):
    x[t].adapt(z[0:t])

#objective function
model.minmax(c@x + h@splus + b@sminus, budgetSet)
#objective function
model.minmax(c@x + h@splus + b@sminus, budgetSet)

#constraints for linearized storage and backlog costs
#model.st(x[0]== 1000)
model.st(splus>= 0)
model.st(sminus>= 0)
for t in range(0,T):
    model.st((splus[t]>= y1 + x[0:t+1].sum()-d[0:t+1].sum()).forall(budgetSet))
    model.st((sminus[t]>= -(y1 + x[0:t+1].sum()-d[0:t+1].sum())).forall(budgetSet))


#constraint on maximum order
model.st(x>=0)
model.st(x<=M)

model.solve(my_solver) # solve the model
obj_val = model.get()
x_val = x.get()
```

Figure 5.2: Python code that implements AARC on the inventory model (see Google Colab).

Remark 5.5. : Note that the most famous example of linear observation mapping is simply the one that reveals at each period an additional subset of the terms in z. This is often referred as the property that "z is progressively revealed". Mathematically speaking, let z be composed of $T-1$ vectors $\{z_t\}_{t=1}^{T-1}$, with $z_t \in \mathbb{R}^{m'}$ such that $z := \left[\begin{array}{cccc} z_1^T & z_2^T & \dots & z_{T-1}^T \end{array}\right]$, we will say that z is "progressively revealed" if $V_t z = z_{[t-1]} = \left[\begin{array}{cccccc} z_1^T & z_2^T & \dots & z_{t-1}^T & 0 & \dots & 0 \end{array}\right]$. In this context, one can consider that the observation matrix $V_t \in \mathbb{R}^{\nu \times m}$, with $\nu := m$, is described as follows:

$$V_t := \left[\begin{array}{cc} \boldsymbol{I}_{(t-1)m' \times (t-1)m'} & \mathbf{0}_{(t-1)m' \times (T-t)m'} \\ \mathbf{0}_{(T-t)m' \times (t-1)m'} & \mathbf{0}_{(T-t)m' \times (T-t)m'} \end{array}\right].$$

In [25], the authors also discuss extensively the notion of "inexact revealed data" which refers to the idea that at each period of time it is not $z_{[t-1]}$ that is revealed but rather a measurement $\hat{z}_{[t-1]} \approx z_{[t-1]}$. This framework is also representable in terms of the multi-stage adjustable robust counterpart model by augmenting the uncertainty space $z \in \mathcal{Z}$ to $(\hat{z}, z) \in \mathcal{U}$ where $\mathcal{U}$ captures the relation between z and $\hat{z}$ (e.g. $\|z - \hat{z}\|_2 \leq \gamma$) and which projection over the z space is equal to $\mathcal{Z}$, namely $\mathcal{Z} = \{z \in \mathbb{R}^m \,|\, \exists \hat{z},\, (\hat{z}, z) \in \mathcal{U}\}$. It is then possible to model the progressively revealed measurements using

$$V_t := \begin{bmatrix} \boldsymbol{I}_{(t-1)m' \times (t-1)m'} & \mathbf{0}_{(t-1)m' \times (T-t)m'} & \mathbf{0}_{(t-1)m' \times m} \\ \mathbf{0}_{(T-t)m' \times (t-1)m'} & \mathbf{0}_{(T-t)m' \times (T-t)m'} & \mathbf{0}_{(T-t)m' \times m} \end{bmatrix},$$

such that $v_t([\hat{z}^T \; z^T]^T) := V_t[\hat{z}^T \; z^T]^T = \begin{bmatrix} \hat{z}_{[t-1]} & 0 & \ldots & 0 \end{bmatrix}$.

5.2 Piecewise affine decision rules through lifting the uncertainty set

Considering that the application of affine decision rules will generate sub-optimal policies compared to fully-adjustable ones, one might be tempted to investigate whether it is possible to obtain tighter conservative approximation by employing more sophisticated (yet tractable) decision rules, in particular nonlinear ones. We will see how this can actually be achieved for piecewise affine decision rules by employing affine decision rules on a lifted version of the uncertainty set.

Example 5.6. : In inventory problems, it is well-known that a base stock policy is a simple policy that can be very effective. Indeed, such a policy takes the form $x_t(y_t; \theta_t) := \max(0, \theta_t - y_t)$, where y_t captures the inventory at the very end of the previous time step, and θ_t captures the base stock level that needs to be reached before encountering the demand at step t. Given a set of base stock levels, once we replace y_t by its definition in terms of previous demand and let $y_1 = 0$, we obtain for instance that the base stock policy for $t = 2$ is represented as $x_2(d_1) := \max(0; \theta_2 - \theta_1 + d_1)$ while at step $t = 3$ we get $x_3(d_1, d_2) := \max(0; \theta_3 - \max(\theta_1 - d_1 - d_2\,;\, \theta_2 - d_2))$. Observe that these policies are not affine but rather piecewise affine in (d_1, d_2).

The above example motivates the use a class of piecewise affine policies described as

$$x_t(v_t(z)) := \bar{x}_t + \bar{X}_t v_t(z) + \sum_{k=1}^{\nu} \theta_{tk}^+ \max(0; v_{tk}(z)) + \theta_{tk}^- \max(0; -v_{tk}(z))$$

where $v_{tk}(z)$ captures the k-th observation in z. Without loss of generality we can assume that each $v_{tk}(z) = z_{i(k)}$ for some mapping $i : \mathbb{N} \to \mathbb{N}$. Under such assumption, a piecewise affine decision rule can be expressed as

$$x_t(v_t(z); \bar{x}_t, \bar{X}_t^+, \bar{X}_t^-) := \bar{x}_t + \bar{X}_t^+ V_t z^+ + \bar{X}_t^- V_t z^-,$$

where $z_i^+ := \max(0; z_i)$ and $z_i^- := \max(0; -z_i)$, and where we omit to include the adjustment X_tV_tz since it can be replicated as : $X_tV_tz^+ - X_tV_tz^-$. In this formulation, one can notice that the policy is actually affine with respect to the vector $[z^T\, {z^+}^T {z^-}^T]$. Hence, we can establish that optimizing such piecewise affine policies is equivalent to applying affine policies to the lifted uncertainty set

$$\mathcal{Z}' := \{(z, z^+, z^-) \in \mathbb{R}^{3m} \,|\, z \in \mathcal{Z},\; z_i^+ = \max(0; z_i),\; z_i^- = \max(0; -z_i),\; \forall\, i = 1, \ldots, m\}\,.$$

Specifically, we would be interested in solving the Lifted AARC model:

$(LAARC)$

$$\begin{aligned} \underset{\{x_t\}_{t=1}^T, \{X_t^+, X_t^-\}_{t=2}^T}{\text{maximize}} \quad & \inf_{(z,z^+,z^-)\in\mathcal{Z}'} c_1(z)^T x_1 + \sum_{t=2}^T c_t(z)^T(x_t + X_t^+V_tz^+ + X_t^-V_tz^-) + d(z) \\ \text{subject to} \quad & a_{j1}(z)^T x_1 + \sum_{t=2}^T a_{jt}(z)^T(x_t + X_t^+V_tz^+ + X_t^-V_tz^-) \le b_j(z)\,, \left\{ \begin{array}{l} \forall\,(z, z^+, z^-) \in \mathcal{Z}' \\ \forall\, j = 1, \ldots, J \end{array} \right. \end{aligned},$$

The difficulty that now arises is that $\mathcal{Z}'$ is not a convex polyhedron so that a constraint such that

$$a_{j1}(z)^T x_1 + \sum_{t=2}^T a_{jt}(z)^T(x_t + X_t^+V_tz^+ + X_t^-V_tz^-) \le b_j(z)\,, \forall\,(z, z^+, z^-) \in \mathcal{Z}'$$

cannot be reformulated through a direct application of duality theory for LPs.

One promising direction for tractable solution comes from considering the following theorem.

Theorem 5.7. : *Given that the decision problem has fixed recourse, the Lifted AARC model is equivalent to*

$$\underset{\{x_t\}_{t=1}^T, \{X_t^+, X_t^-\}_{t=2}^T}{\text{maximize}} \quad \inf_{(z,z^+,z^-)\in\mathcal{Z}''} c_1(z)^T x_1 + \sum_{t=2}^T c_t^T(x_t + X_t^+V_tz^+ + X_t^-V_tz^-) + d(z) \tag{5.3a}$$

$$\text{subject to} \quad a_{j1}(z)^T x_1 + \sum_{t=2}^T a_{jt}^T(x_t + X_t^+V_tz^+ + X_t^-V_tz^-) \le b_j(z)\,, \left\{ \begin{array}{l} \forall\,(z, z^+, z^-) \in \mathcal{Z}' \\ \forall\, j = 1, \ldots, J \end{array} \right. \tag{5.3b}$$

where $\mathcal{Z}'' := ConvexHull(\mathcal{Z}')$. Therefore, if the convex hull of $\mathcal{Z}'$ can be described with a finite number of linear constraints then the Lifted AARC model can be solved efficiently.

Proof. The proof simply relies on the fact that since $\mathcal{Z}'' \supseteq \mathcal{Z}'$, any feasible solution of problem (5.3) is necessarily feasible for the LAARC model. Alternatively, since the functions involved in each constraint are linear in (z, z^+, z^-), if we take a feasible solution to LAARC and verify feasibility in problem (5.3), it is necessarily the case that there is a worst-case realization for each constraint that occurs at one of the vertices of $\mathcal{Z}''$ which by construction were members of $\mathcal{Z}'$. This indicates that any feasible solution of LAARC is also feasible in problem (5.3). Hence, the feasible sets of both problems are equivalent. Furthermore, a similar argument, based on the linearity of the functions that are involved, can be used to establish that both objective functions are equivalent. We can thus conclude that the set of optimal solutions and the optimal value of both problems are therefore equivalent. □

This result informs us that one might be able to obtain a tractable reformulation of the LAARC model if he can identify a good representation for the convex hull of $\mathcal{Z}'$. In this regard, the following proposition might come in handy.

Proposition 5.8. : *Let $\mathcal{Z} \subseteq [-M, M]^m$ for some $M > 0$. Then, the uncertainty set $\mathcal{Z}'$ can be represented as*

$$\mathcal{Z}' = \left\{ (z, z^+, z^-) \in \mathbb{R}^{3m} \;\middle|\; \exists u^+ \in \{0,1\}^m,\, u^- \in \{0,1\}^m, \begin{array}{c} z \in \mathcal{Z} \\ z = z^+ - z^- \\ 0 \leq z^+ \leq Mu^+ \\ 0 \leq z^- \leq Mu^- \\ u^+ + u^- = 1 \end{array} \right\}.$$

Proof. One can confirm that given any $z \in \mathcal{Z}$, since for any $j = 1, \ldots, m$, we necessarily have that $u_j^+ + u_j^- = 1$, $u_j^+ \in \{0,1\}$, and $u_j^- \in \{0,1\}$, thus that either $z_j^+ > 0$ or $z_j^- > 0$. Hence, if $z_j > 0$ it is necessary that $z_j^+ = z_j$ and $z_j^- = 0$, while if $z_j < 0$ it is necessary that $z_j^+ = 0$ and $z_j^- = -z_j$. Finally, if $z_j = 0$ then the only option is for $z_j^+ = z_j^- = 0$. This is exactly the behaviour that is described by $\mathcal{Z}'$. □

This proposition is interesting for two reasons. First, by relaxing the binary constraints on u^+ and u^- we instantly obtain a tractable outer approximation of ConvexHull($\mathcal{Z}'$). In particular, this will be especially effective with the budgeted uncertainty set. Secondly, this representation of $\mathcal{Z}'$ provides us a way of performing the worst-case analysis of a fixed multi-stage piecewise affine policies.

Corollary 5.9. : *Given that the decision problem has fixed recourse, let $\{x_t\}_{t=1}^T$ and $\{X_t^+, X_t^-\}_{t=2}^T$ define a multi-stage piecewise affine policies. Then, one can verify the feasibility with respect to any j-th constraint of the LAARC model*

$$a_{j1}^T x_1 + \sum_{t=1}^{T} a_{jt}^T (x_t + X_t^+ V_t z^+ + X_t^- V_t z^-) \leq b_j(z)\,,\; \forall\, (z, z^+, z^-) \in \mathcal{Z}''$$

by solving the following mixed integer linear program

$$\begin{array}{ll} \underset{z,z^+,z^-,u^+,u^-}{\text{maximize}} & a_{j1}^T x_1 + \sum_{t=2}^{T} a_{jt}^T (x_t + X_t^+ V_t z^+ + X_t^- V_t z^-) - b_j(z) \\ \text{subject to} & z \in \mathcal{Z} \\ & z = z^+ - z^- \\ & 0 \leq z^+ \leq M u^+ \\ & 0 \leq z^- \leq M u^- \\ & u^+ + u^- = 1 \\ & u^+ \in \{0,1\}^m, \ u^- \in \{0,1\}^m \end{array}$$

to obtain (z^*, z^{+*}, z^{-*}) *and verifying that the optimal value is lower or equal to zero. In the situation that the multi-stage piecewise affine policy is infeasible, then the constraint*

$$a_{j1}^T x_1 + \sum_{t=2}^{T} a_{jt}^T (x_t + X_t^+ V_t z^{+*} + X_t^- V_t z^{-*}) \leq b_j(z^*)$$

separates the current multi-stage piecewise affine policies from the set of such policies that are feasible with respect to the j*-th constraint of the LAARC model.*

We now present a tractable representation of ConvexHull$(\mathcal{Z}')$ when $\mathcal{Z}$ is the budgeted uncertainty set.

Proposition 5.10. : *Let* $\mathcal{Z}$ *be the budgeted uncertainty set. Then the uncertainty set ConvexHull*$(\mathcal{Z}')$ *can be represented using the following tractable form*

$$\mathit{ConvexHull}(\mathcal{Z}') = \left\{ (z, z^+, z^-) \in \mathbb{R}^{3m} \;\middle|\; \begin{array}{c} z^+ + z^- \leq 1 \\ \sum_{i=1}^m z_i^+ + z_i^- \leq \Gamma \\ z = z^+ - z^- \\ 0 \leq z^+ \\ 0 \leq z^- \end{array} \right\}.$$

Proof. Let $\mathcal{Z}_2'$ be the set described in the proposition, i.e.

$$\mathcal{Z}_2' := \left\{ (z, z^+, z^-) \in \mathbb{R}^{3m} \;\middle|\; \begin{array}{c} z^+ + z^- \leq 1 \\ \sum_{i=1}^m z_i^+ + z_i^- \leq \Gamma \\ z = z^+ - z^- \\ 0 \leq z^+ \\ 0 \leq z^- \end{array} \right\}.$$

First, we work on simplifying the representation of $\mathcal{Z}'$ expressed in proposition 5.8

when using the budgeted uncertainty set. Indeed, we have that

$$\mathcal{Z}' = \left\{ (z, z^+, z^-) \in \mathbb{R}^{3m} \;\middle|\; \exists u^+ \in \{0,1\}^m,\ u^- \in \{0,1\}^m,\ \begin{array}{c} \|z\|_\infty \leq 1 \\ \|z\|_1 \leq \Gamma \\ z = z^+ - z^- \\ 0 \leq z^+ \leq u^+ \\ 0 \leq z^- \leq u^- \\ u^+ + u^- = 1 \end{array} \right\}$$

$$= \left\{ (z, z^+, z^-) \in \mathbb{R}^{3m} \;\middle|\; \exists u^+ \in \{0,1\}^m,\ u^- \in \{0,1\}^m,\ \begin{array}{c} z^+ + z^- \leq 1 \\ \sum_{i=1}^m z_i^+ + z_i^- \leq \Gamma \\ z = z^+ - z^- \\ 0 \leq z^+ \leq u^+ \\ 0 \leq z^- \leq u^- \\ u^+ + u^- = 1 \end{array} \right\} .$$

Based on this representation, it is clear that $\mathcal{Z}' \subseteq \mathcal{Z}_2'$ since $\mathcal{Z}'$ has additional constraint. It is also clear that ConvexHull($\mathcal{Z}'$) $\subseteq \mathcal{Z}_2'$ since $\mathcal{Z}_2'$ is convex. We are left with demonstrating that $\mathcal{Z}_2' \subseteq$ ConvexHull($\mathcal{Z}'$).

Given that a convex hull can be defined as the intersection of all the half spaces described by the supporting hyperplanes of the points that are covered, we can confirm that $\mathcal{Z}_2'$ is contained in any such constructed half spaces to conclude that $\mathcal{Z}_2' \subseteq$ ConvexHull($\mathcal{Z}'$). For any direction defined by (c, c^+, c^-) we can identify a supporting half space as all triplets (z, z^+, z^-) such that

$$c^T z^T + {c^+}^T z^+ + {c^-}^T z^- \leq \beta ,$$

with

$$\beta = \sup_{(z,z^+,z^-) \in \mathcal{Z}'} c^T z^T + {c^+}^T z^+ + {c^-}^T z^- .$$

To show that Z_2' is in this half space, we need to demonstrate that for all $c \in \mathbb{R}^m$, $c^+ \in \mathbb{R}^m$, and $c^- \in \mathbb{R}^m$,

$$\max_{(z,z^+,z^-) \in \mathcal{Z}_2'} \ c^T z^T + {c^+}^T z^+ + {c^-}^T z^- \ \leq \ \max_{(z,z^+,z^-) \in \mathcal{Z}'} \ c^T z^T + {c^+}^T z^+ + {c^-}^T z^- .$$

In other words, we should show that the optimal value of

$$\begin{array}{ll} \underset{z,z^+,z^-}{\text{maximize}} & c^T z^T + {c^+}^T z^+ + {c^-}^T z^- \\ \text{subject to} & z^+ + z^- \leq 1 \\ & \displaystyle\sum_{i=1}^m z_i^+ + z_i^- \leq \Gamma \\ & z = z^+ - z^- \\ & z^+ \geq 0,\ z^- \geq 0 , \end{array}$$

is always smaller or equal to the optimal value of

$$\begin{aligned}
\underset{z,z^+,z^-,u^+,u^-}{\text{maximize}} \quad & c^T z^T + {c^+}^T z^+ + {c^-}^T z^- \\
\text{subject to} \quad & z^+ + z^- \leq 1 \\
& \sum_{i=1}^{m} z_i^+ + z_i^- \leq \Gamma \\
& z = z^+ - z^- \\
& z^+ \geq 0,\ z^- \geq 0 \\
& u^+ \in \{0,1\}^m,\ u^- \in \{0,1\}^m \\
& 0 \leq z^+ \leq u^+ \\
& 0 \leq z^- \leq u^- \\
& u^+ + u^- = 1\ .
\end{aligned}$$

Yet, one can establish that any optimal solution (z, z^+, z^-) to the first problem can be used to construct a feasible solution to the second optimization problem that achieves the same objective value in the second problem. This is done with

$$\begin{aligned}
{u_j^+}' &:= \begin{cases} 1 & \text{if } z_j^+ > 0 \\ 0 & \text{otherwise} \end{cases} \\
{u_j^-}' &:= 1 - {u_j^+}' \\
{z_j^+}' &:= (z_j^+ + z_j^-){u_j^+}' \\
{z_j^-}' &:= (z_j^+ + z_j^-){u_j^-}' \\
z_j' &:= {z_j^+}' - {z_j^-}'\ .
\end{aligned}$$

The argument repose on realizing that any such optimal solution where both $z_j^+ > 0$ and $z_j^- > 0$ would necessarily have that $c_j + c_j^+ = -c_j + c_j^-$ so that it does not matter how the "weight" is distributed. Specifically,

$$\begin{aligned}
& c_j z_j' + c_j^+ {z_j^+}' + c_j^- {z_j^-}' = (c_j + c_j^+){z_j^+}' + c_j^- - c_j {z_j^-}' \\
&\quad = \begin{cases} (c_j + c_j^+)z_j^+ = c_j z_j + c_j^+ z_j^+ + c_j^- z_j^- & \text{if } z_j^- = 0 < z_j^+ \\ (c_j^- - c_j)z_j^- = c_j z_j + c_j^+ z_j^+ + c_j^- z_j^- & \text{if } z_j^+ = 0 < z_j^- \\ (c_j + c_j^+)(z_j^+ + z_j^-) = (c_j + c_j^+)z_j^+ + (-c_j + c_j^-)z_j^- & \text{if } 0 < z_j^- \text{ and } 0 < z_j^+ \end{cases} \\
&\quad = c_j z_j + c_j^+ z_j^+ + c_j^- z_j^-
\end{aligned}$$

□

Example 5.11. : Consider our inventory management problem in which we wish to optimize a policy that is piecewise affine with respect to the positive and negative deviations of each demand parameter. In particular, at time $t = 2$ we would like to design a policy that is parametrized as

$$x_2(d_1) := \bar{x}_2 + \bar{x}_2^+ \max(0; (d_1 - \bar{d}_1)/\hat{d}_1) + \bar{x}_2^- \max(0; (\bar{d}_1 - d_1)/\hat{d}_1)$$

such that we increase the order by $\bar{x}_2^+$ units per "normalized" unit of demand above the nominal amount, and decrease by $\bar{x}_2^-$ units per normalized unit below the nominal amount. This can be done by working in terms of (z, z^+, z^-) instead of d. Namely look for policies of the type:

$$x_2(d_1) := \bar{x}_2 + \bar{x}_2^+ \max(0; z_1) + \bar{x}_2^- \max(0; -z_1) \ .$$

An efficient representation for the convex hull of the lifted uncertainty set in terms of (z, z^+, z^-) was described in proposition 5.10 for the case where one is interested in using the budgeted uncertainty set.

This leads to the following Lifted AARC.

$$\begin{array}{ll}
\underset{\substack{x_1, \{x_t, X_t^+, , X_t^-\}_{t=2}^T, \\ \{r_t, R_t^+, R_t^-\}_{t=1}^T, \\ \{s_t, S_t^+, S_t^-\}_{t=1}^T}}{\text{minimize}} & \sup_{(z^+,z^-)\in\mathcal{Z}''} \ c_1 x_1 + \sum_t c_t(x_t + X_t^+ z_{[t-1]}^+ + X_t^- z_{[t-1]}^-) \\
& \quad +h_t(r_t + R_t^+ z^+ + R_t^- z^-) + b_t(s_t + S_t^+ z^+ + S_t^- z^-) \\
\text{subject to} & r_t + R_t^+ z^+ + R_t^- z^- \geq 0, \ s_t + S_t^+ z^+ + S_t^- z^- \geq 0 \,, \forall\, (z^+, z^-) \in \mathcal{Z}'', \ \forall t
\end{array}$$

$$r_t + R_t^+ z^+ + R_t^- z^- \geq y_1 + \sum_{t'=1}^{t} x_{t'} + X_{t'}^+ z_{[t'-1]}^+ + X_{t'}^- z_{[t'-1]}^- - d_{t'}(z^+, z^-) \,, \forall\, (z^+, z^-) \in \mathcal{Z}'', \ \forall t$$

$$s_t + S_t^+ z^+ + S_t^- z^- \geq -y_1 + \sum_{t'=1}^{t} d_{t'}(z^+, z^-) - (x_{t'} + X_{t'}^+ z_{[t'-1]}^+ + X_{t'}^- z_{[t'-1]}^-) \,, \forall\, (z^+, z^-) \in \mathcal{Z}'', \ \forall t$$

$$0 \leq x_t + X_t^+ z_{[t-1]}^+ + X_t^- z_{[t-1]}^- \leq M \,, \forall\, (z^+, z^-) \in \mathcal{Z}'', \ \forall t \,,$$

where $d_t(z^+, z^-) := \bar{d}_t + \hat{d}_t(z_t^+ - z_t^-)$ and where

$$\mathcal{Z}'' := \{(z^+, z^-) \in \mathbb{R}^{2T} \,|\, z^+ \geq 0, \ z^- \geq 0, \ z^+ + z^- \leq 1, \ \sum_{t=1}^{T} z_t^+ + z_t^- \leq \Gamma\}$$

is the representation of the convex hull identified in Proposition 5.10. This is implemented in RSOME using the following set of code.

5.3 Exercises : Facility Location Problem II

This set of exercises revisits the model discussed in the exercises of Section 4.

Exercise 5.1. ***Implementing Static RC***
Implement with RSOME (using Google Colab) a static RC approach to the following

```python
# begin RSOME
model = ro.Model('LAARC')

# Define uncertain factors
zplus = model.rvar(T)
zminus = model.rvar(T)
budgetSet = (zplus>=0, zminus>=0, zplus+zminus<=1,  #each parameter is between [-1, 1]
                  sum(zplus)+sum(zminus)<=Gamma)   # Budget of uncertainty approach

# Demand relations with uncertain factors
d = mus + deltas*(zplus-zminus)

# Define decision rules
splus=model.ldr(T) #storage variable
sminus=model.ldr(T) #backlog variable
x=model.ldr(T) #ordering decision

# Define decision rules dependance
splus.adapt(zplus)
splus.adapt(zminus)
sminus.adapt(zplus)
sminus.adapt(zminus)
for t in range(1,T):
    x[t].adapt(zplus[0:t])
    x[t].adapt(zminus[0:t])
#objective function
model.minmax(c@x + h@splus + b@sminus, budgetSet)

#constraints for linearized storage and backlog costs
#model.st(x[0]== 1000)
model.st(splus>= 0)
model.st(sminus>= 0)
for t in range(0,T):
    model.st((splus[t]>= y1 + x[0:t+1].sum()-d[0:t+1].sum()))
    model.st((sminus[t]>= -(y1 + x[0:t+1].sum()-d[0:t+1].sum())))


#constraint on maximum order
model.st(x>=0)
model.st(x<=M)

model.solve(my_solver) # solve the model
obj_val = model.get()
x_val = x.get()
```

Figure 5.3: Python code that implements LAARC with budgeted uncertainty set on the inventory model (see Google Colab).

facility location problem.

$$
\begin{aligned}
\underset{x,y}{\text{maximize}} \quad & -\sum_{i=1}^{n} c_i x_i + \sum_{i=1}^{n}\sum_{j=1}^{m}(r_{ij} - d_{ij}) y_{ij} \\
\text{subject to} \quad & \sum_{i=1}^{n} y_{ij} \leq D_j \,, \forall j = 1, \ldots, m \\
& \sum_{j=1}^{m} y_{ij} \leq P_i x_i \,, \forall i = 1, \ldots, n \\
& y \geq 0 \\
& x \in \{0,1\}^n
\end{aligned}
$$

when the D_j's are known to be in the set $\{D \in \mathbb{R}^m \,|\, \exists z \in \mathcal{Z},\, D_j = \bar{D}_j + \hat{D}_j z_j \,, \forall j = 1, \ldots, m\}$ with $\mathcal{Z}$ as the budgeted uncertainty set.

Exercise 5.2. ***Implementing AARC***
Implement with RSOME (using Google Colab) an AARC approach to the facility location problem presented in exercise 5.1.

Exercise 5.3. ***Implementing Lifted AARC***
Implement with RSOME (using Google Colab) an AARC approach to the facility location problem in exercise 5.1 after lifting the uncertainty to the space

$$(z, z^+, z^-) \in \{(z, z^+, z^-) \in \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^m \,|\, z = z^+ - z^-,\ z^+ \geq 0,\ z^- \geq 0,\ z^+ + z^- \leq 1,\ \sum_i z_i^+ + z_i^- \leq \Gamma\}\ ,$$

in order to produce policies which are piecewise affine such as $y_{ij}(z) := \bar{y}_{ij} + \sum_{k=1}^m \bar{Y}_{ijk}^+ \max(0; z_k) + \bar{Y}_{ijk}^- \max(0; -z_k)$.

Exercise 5.4. ***Comparison of Approximate Worst-case Bounds***
Compare (using Google Colab) the different optimal values obtained from the three approximation models (RC, AARC, and LAARC) on the robust facility location-transportation problem to the true worst-case value that can be achieved. In particular, compare these different approximate optimal worst-case values when the following budgets are used:

1. A budget of $\Gamma = 0$
2. A budget of $\Gamma = 1$
3. A budget of $\Gamma = 4$
4. A budget of $\Gamma = m - 1$

Discuss what you observe in these results.

Exercise 5.5. ***Comparison of Worst-case performance of Approximate robust policy***
Compare (using Google Colab) the quality of the approximate robust first-stage decision as follows. For each of them report what is the actual worst-case total profit generated by opening the selected facilities. Note that to get the actual worst-case profit, one should allow the optimal transportation y_{ij} decision to occur once the true demand is known. In particular, compare the worst-case performance of these first stage decisions when the following budgets are used:

1. A budget of $\Gamma = 0$
2. A budget of $\Gamma = 1$
3. A budget of $\Gamma = 4$
4. A budget of $\Gamma = m - 1$

Discuss what you observe in these results.

Chapter 6

Robust Nonlinear Programming

We focus on the reformulation of a robust constraint that involve non-linear functions. In particular, let's consider

$$g(x,z) \le 0\,,\,\forall\, z \in \mathcal{Z}\,, \tag{6.1}$$

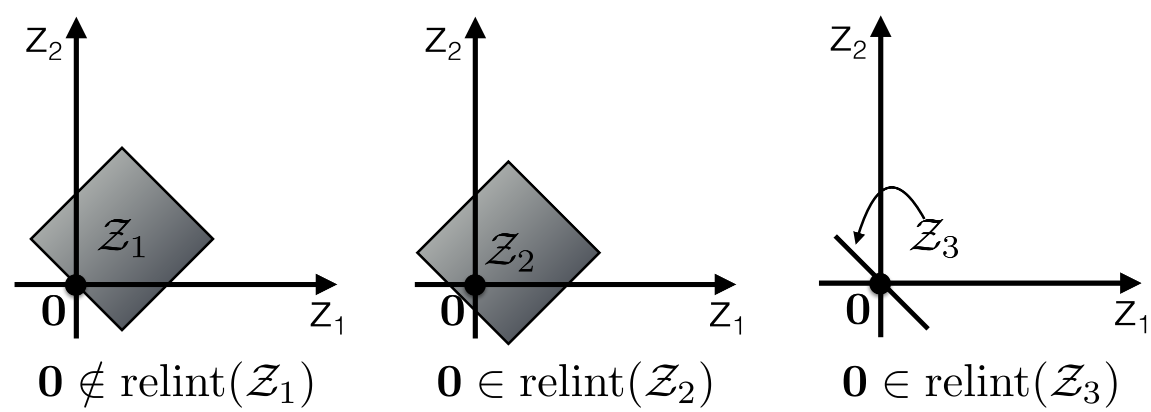
where $g(\cdot,\cdot)$ is a mapping defined over the convex domain $\mathcal{X}_g \times \mathcal{Z}_g$ (typically $\mathbb{R}^n \times \mathbb{R}^m$) with $\mathcal{X}_g \subseteq \mathbb{R}^n$ and $\mathcal{Z}_g \subseteq \mathbb{R}^m$. Furthermore, we will assume that $g(x,z)$ is convex in x for all $z \in \mathcal{Z}_g$ and concave in z for all $x \in \mathcal{X}_g$ while $\mathcal{Z} \subset \mathbb{R}^m$ is a given non-empty, convex and compact (i.e. closed and bounded) set. Finally, it will be assumed that there exists a vector $z_0 \in \mathbb{R}^m$ (possibly the nominal value for the parameters of $g(x,\cdot)$) such that z_0 is both in the relative interior of $\mathcal{Z}$ and in the relative interior of the domain of $g(x,\cdot)$, $\forall\, x \in \mathcal{X}$. This is a technical conditions that will be needed to apply duality theory.

Remark 6.1. : Specifically, $z_0 \in \mathrm{relint}(\mathcal{Z})$ means that there exists a ball centred at z_0 and of radius $\epsilon > 0$ which projection on the affine space spanned by $\mathcal{Z}$ is included in $\mathcal{Z}$. This translates as the following conditions since $\mathcal{Z}$ is assumed convex:

$$\exists \epsilon > 0, \forall z \in \mathcal{Z}, z_0 - \epsilon((z - z_0)/\|z - z_0\|_2) \in \mathcal{Z}\,.$$

Here is an illustration for $\mathcal{Z} \in \mathbb{R}^2$ when $z_0 = \mathbf{0}_m$.

Example of sets that include (or not) $\mathbf{0}$ in their relative interior



As seen in chapter 2, for any specific instance of constraint (6.1) it is possible to apply duality theory to reformulate the robust constraint in the form

$$\begin{aligned} &h(x, \lambda) \leq 0 \\ &\lambda \in \Lambda(x) \ , \end{aligned}$$

where λ is an additional variable that is used as a certificate that the robust constraint is met, and where $h(x, \lambda)$ would be a new convex function in terms of x and λ, and $\Lambda(x)$ is the feasible set for λ. Unfortunately, obtaining this reformulation through duality arguments is a tedious process that needs to be reapplied each time the uncertainty set or constraint function is modified. The good news is that in 2015, a group of researchers proposed a convenient way of systematically and efficiently reformulating robust non-linear constraints as discussed next.

6.1 The Fenchel Robust Counterpart

In [9], the authors present for the first time a method that can be used to obtain such a reformulation much more efficiently as it decomposes the dependence of the reformulation between $\mathcal{Z}$ and $g(\cdot, \cdot)$. In particular, here is the main theorem for what was called the Fenchel Robust Counterpart by A. Ben-Tal, D. den Hertog, and J.-P. Vial:

Theorem 6.2. : *The vector $x \in \mathcal{X}$ satisfies the robust constraint* (6.1) *if and only if there exists a $v \in \mathbb{R}^m$ that satisfies the following constraint:*

$$(FRC) \quad \delta^*(v|\mathcal{Z}) - g_*(x, v) \leq 0 \ , \tag{6.2}$$

where the support function δ^* is defined as

$$\delta^*(v|\mathcal{Z}) := \sup_{z \in \mathcal{Z}} z^T v$$

and the partial concave conjugate function g_* is defined as

$$g_*(x, v) := \inf_{z \in \mathcal{Z}_g} v^T z - g(x, z) \ .$$

Proof. We will limit our proof to establishing that the FRC constraint (6.2) is a conservative approximation of the robust constraint (6.1). To do this, we will reformulate the following expression

$$\psi := \max_{z \in \mathcal{Z}} \ g(x, z) \ ,$$

using the Lagrangian function on the following equivalent problem

$$\psi = \max_{z' \in \mathcal{Z}_g, \, z \in \mathcal{Z}, z' = z} \ g(x, z') \ .$$

Doing this we obtain

$$\psi = \max_{z' \in \mathcal{Z}_g,\, z \in \mathcal{Z}} \inf_v \; g(x, z') - v^T (z' - z) \,,$$

where $v \in \mathbb{R}^m$. Following a basic theory of sequential games, which appears in lemma 10.6, we can easily establish that

$$\psi \leq \inf_v \max_{z' \in \mathcal{Z}_g,\, z \in \mathcal{Z}} \; g(x, z') - v^T (z' - z) \;.$$

Yet, the right-hand side expression can be decomposed in two parts:

$$\psi \leq \inf_v \underbrace{\max_{z' \in \mathcal{Z}_g} \; g(x, z') - v^T z'}_{-g_*(x,v)} + \underbrace{\max_{z \in \mathcal{Z}} v^T z}_{\delta^*(v|\mathcal{Z})} \;.$$

Hence, imposing that

$$\exists v \in \mathbb{R}^m,\; \delta^*(v|\mathcal{Z}) - g_*(x, v) \leq 0 \;\Rightarrow\; g(x, z) \leq 0 \,,\, \forall\, z \in \mathcal{Z} \;.$$

To demonstrate that this condition is necessary, one needs some constraint qualification argument in order for strong duality to apply. This argument could easily come from Lagrangian duality if we assumed that $\mathcal{Z}_g$ was bounded (this would be without loss of generality in cases where $\mathcal{Z}$ would be bounded). In [9], the authors employ Fenchel duality to guarantee that

$$\psi = \delta^*(v|\mathcal{Z}) - g_*(x, v) \;.$$

We refer the reader to that article for more details. □

Example 6.3. : Consider the following robust optimization constraint:

$$p(x)^T z + s(x) - z^T P(x) z \leq 0 \,,\, \forall\, z \in \mathcal{Z} \;,$$

where $p : \mathbb{R}^n \to \mathbb{R}^m$ is an affine function of x, $s : \mathbb{R}^n \to \mathbb{R}$, and $P(x) : \mathbb{R}^n \to \mathbb{R}^{m \times m}$, and finally where

$$\mathcal{Z} := \{z \in \mathbb{R}^m \,|\, z^T Q z \leq r\} \;,$$

with $Q \in \mathbb{R}^{m \times m}$ a symmetric matrix and $r \in \mathbb{R}$.

After describing $g(x, z)$ as $g(x, z) := p(x)^T z + s(x) - z^T P(x) z$ and letting $z_0 = 0$, one needs to make the following assumptions in order to apply theorem 6.2:

- Impose that $x \in \mathcal{X}_g$ with $\mathcal{X}_g := \{x \,|\, P(x) \succeq 0\}$, namely that we have the guarantee that $P(x)$ is positive semi-definite in order to make $g(x, z)$ concave in z.

- Impose that $Q \succ 0$ and that $r > 0$, namely that Q is positive definite to ensure that $\mathcal{Z}$ is convex and bounded, and that $0 \in \mathrm{relint}(\mathcal{Z})$.

When applying theorem 6.2, we obtain that the constraint is equivalent to

$$\exists v \in \mathbb{R}^m \;\; \delta^*(v|\mathcal{Z}) - g_*(x, v) \leq 0 \,,$$

yet we still need to identify properly what form the two functions take.

Treating the uncertainty set We are interested in identifying $\delta^*(v|\mathcal{Z}) := \sup_{z\in\mathcal{Z}} z^T v$. To do so, let's apply Lagrangian duality:

$$\begin{aligned}\sup_{z\in\mathcal{Z}} z^T v &= \sup_{z} \inf_{\gamma\geq 0} z^T v - \gamma(z^T Q z - r)\\ &= \inf_{\gamma\geq 0} \sup_{z} z^T v - \gamma(z^T Q z - r)\ .\end{aligned}$$

Duality is strong here since Slater's condition applies considering that $\mathbf{0}_m^T Q \mathbf{0}_m = 0 < r$. Following what was done in example 2.2.1, since we established that

$$\sup_{z} y^T z - \gamma(z^T \Sigma^{-1} z - 1) = \frac{1}{4\gamma} y^T \Sigma y + \gamma\ ,$$

we can conclude that

$$\inf_{\gamma\geq 0} \sup_{z} z^T v - \gamma(z^T Q z - r) = \inf_{\gamma > 0} \frac{1}{4\gamma} v^T Q^{-1} v + r\gamma\ .$$

Further following the steps that we had taken in that example, we obtain that

$$\delta^*(v|\mathcal{Z}) = \sqrt{r}\sqrt{v^T Q^{-1} v}\ .$$

Treating the constraint function We are interested in identifying $g_*(x, v) := \inf_z v^T z + z^T P(x) z - p(x)^T z - s(x)$. We will actually reformulate this using a linear matrix inequality as follows:

$$\begin{aligned}g_*(x,v) &= \sup\{t \,|\, t \leq v^T z + z^T P(x) z - p(x)^T z - s(x)\,,\ \forall\, z\in\mathbb{R}^m\}\\ &= \sup\{t \,|\, v^T z + z^T P(x) z - p(x)^T z - s(x) - t \geq 0\,,\ \forall\, z\in\mathbb{R}^m\}\\ &= \sup_t \left\{ t \,\middle|\, \begin{bmatrix} z \\ 1 \end{bmatrix}^T \begin{bmatrix} P(x) & (v-p(x))/2 \\ (v-p(x))^T/2 & -s(x)-t \end{bmatrix} \begin{bmatrix} z \\ 1 \end{bmatrix} \geq 0\,,\ \forall\, z\in\mathbb{R}^m \right\}\\ &= \sup_t \left\{ t \,\middle|\, \begin{bmatrix} z \\ y \end{bmatrix}^T \begin{bmatrix} P(x) & (v-p(x))/2 \\ (v-p(x))^T/2 & -s(x)-t \end{bmatrix} \begin{bmatrix} z \\ y \end{bmatrix} \geq 0\,,\ \forall\, z\in\mathbb{R}^m,\ y\in\mathbb{R} \right\} \text{ (since } P(x) \succeq\\ &= \sup_t \left\{ t \,\middle|\, \begin{bmatrix} P(x) & (v-p(x))/2 \\ (v-p(x))^T/2 & -s(x)-t \end{bmatrix} \succeq 0 \right\}\end{aligned}$$

the latter inequality comes directly from the definition of a positive semi-definite matrix (see appendix 10.8).

Combining the two Now that we have two reformulations, we can reassemble the constraint:

$$\begin{aligned}&\sqrt{r}\sqrt{v^T Q^{-1} v} - t \leq 0\\ &\begin{bmatrix} P(x) & (v-p(x))/2 \\ (v-p(x))^T/2 & -s(x)-t \end{bmatrix} \succeq 0\ ,\end{aligned}$$

where $v \in \mathbb{R}^m$ and $t \in \mathbb{R}$ are additional decision variables that need to be optimized jointly with x.

In the above example, we can observe how the work of reformulating the constraint was divided into two steps. This is a clear advantage as it allows one to easily modify his model and recuperate the new formulation. One might for instance compare the solutions that are obtained using different types of uncertainty sets. In each case, the only modifications to the reformulated model would appear in the part that serves the purpose of evaluating $\delta^*(v|\mathcal{Z})$. As an example, the article establishes that in the case of using a polyhedron defined as $Bz \leq b$, then one requires an additional decision vector $\lambda \in \mathbb{R}^m$ and obtains the following constraints:

$$\begin{aligned}
& b^T\lambda - g_*(x, v) \leq 0 \\
& B^T\lambda = v \\
& \lambda \geq 0 \, ,
\end{aligned}$$

which translates in the context of our quadratic function to

$$\begin{aligned}
& b^T\lambda - t \leq 0 \\
& B^T\lambda = v \\
& \lambda \geq 0 \\
& \begin{bmatrix} P(x) & (v - p(x))/2 \\ (v - p(x))^T/2 & -s(x) - t \end{bmatrix} \succeq 0 \, .
\end{aligned}$$

The downside of employing the proposed Fenchel robust counterpart is that the conditions that need to be imposed are a bit more restrictive then needed to obtain a tractable reformulation of robust constraints. Indeed, by decomposing the influence of the uncertainty set and the constraint function some of the tractability of robust optimization is lost as illustrated in the following example.

Example 6.4. : Consider the following robust optimization constraint:

$$p(x)^T z + s(x) - z^T P(x) z \leq 0 \, , \, \forall \, z \in \mathcal{Z} \, ,$$

where $p : \mathbb{R}^n \to \mathbb{R}^m$ is an affine function of x, and so are $s : \mathbb{R}^n \to \mathbb{R}$, and $P(x) : \mathbb{R}^n \to \mathbb{R}^{m \times m}$, and finally where

$$\mathcal{Z} := \{z \in \mathbb{R}^m \,|\, z^T Q z \leq r\} \, ,$$

with $Q \in \mathbb{R}^{m \times m}$ a symmetric matrix and $r \in \mathbb{R}$. In fact, it is well known that this robust constraint has a tractable reformulation even in cases where there is no guarantee that $P(x)$ is positive semi-definite. Consider for instance imposing that there exists some $\lambda \geq 0$ for which

$$p(x)^T z + s(x) - z^T P(x) z + \lambda(r - z^T Q z) \leq 0 \, , \, \forall \, z \in \mathbb{R}^m \, .$$

It is clear that any x that satisfy this will satisfy the robust constraint since this constraint is stricter for $z \in \mathcal{Z}$. Yet, a famous version of the S-lemma (see theorem 2.2

in [38] attributed to [52]) guarantees that the two constraints are equivalent as long as there exists a z such that $z^T Qz < r$. Based on this lemma, it is therefore possible to reformulate the constraint as the following linear matrix inequality:

$$\begin{bmatrix} P(x) + \lambda Q & -p(x)/2 \\ -p(x)^T/2 & -s(x) - r\lambda \end{bmatrix} \succeq 0 \,.$$

This reformulation cannot be obtained using theorem 6.2 since here we allow $g(x, z)$ to be non-concave in z for some $x \in \mathcal{X}_g$.

We present below a list of useful theorems for employing the theory proposed in theorem 6.2 in order to obtain the Fenchel Robust Counterpart. The first two theorems identify a characterization of the support function for some simple polyhedra.

Theorem 6.5. : *If $\mathcal{Z} := \{z \in \mathbb{R}^m \,|\, 0 \le z \le 1,\, \sum_i z_i \le \rho\}$ with $\rho \ge 0$, then*

$$\begin{aligned} \delta^*(v|\mathcal{Z}) := \quad & \inf_{\omega \in \mathbb{R}^m, \lambda \in \mathbb{R}} && \sum_i \omega_i + \rho\lambda \\ & \text{subject to} && \lambda \ge v_i - \omega_i \,,\, \forall\, i \\ & && \lambda \ge 0,\ \omega \ge 0 \,. \end{aligned}$$

Hence, the robust counterpart takes the form:

$$\exists \omega \in \mathbb{R}^m, \lambda \in \mathbb{R}, v \in \mathbb{R}^m, \quad \left\{ \begin{array}{l} \sum_i \omega_i + \rho\lambda - g_*(x, v) \le 0 \\ \lambda \ge v_i - \omega_i \,,\, \forall\, i \\ \lambda \ge 0,\ \omega \ge 0 \end{array} \right. ,$$

where strong LP duality was applied.

Proof. Simply put

$$\begin{aligned} \delta^*(v|\mathcal{Z}) &= \sup_{z \in \mathcal{Z}} v^T z = \sup_z \inf_{\gamma \ge 0, \omega \ge 0, \lambda \ge 0} v^T z + \sum_i \gamma_i z_i + \sum_i \omega_i (1 - z_i) + \lambda(\rho - \sum_i z_i) \\ &= \inf_{\gamma \ge 0, \omega \ge 0, \lambda \ge 0} \sup_z v^T z + \sum_i \gamma_i z_i + \sum_i \omega_i (1 - z_i) + \lambda(\rho - \sum_i z_i) \\ &= \inf_{\gamma \ge 0, \omega \ge 0, \lambda \ge 0} \left\{ \begin{array}{cl} \sum_i \omega_i + \rho\lambda & \text{if } v_i + \gamma_i - \omega_i - \lambda = 0 \text{ for all } i \\ \infty & \text{otherwise} \end{array} \right. \\ &= \inf_{\lambda \ge 0, w \ge 0} \left\{ \begin{array}{cl} \sum_i \omega_i + \rho\lambda & \text{if } \lambda \ge v_i - \omega_i \text{ for all } i \\ \infty & \text{otherwise} \end{array} \right. , \end{aligned}$$

where again strong LP duality was applied. □

Theorem 6.6. : *If $\mathcal{Z} := \{z \in \mathbb{R}^m \,|\, z \ge 0,\, \sum_i z_i = 1\}$, then*

$$\begin{aligned} \delta^*(v|\mathcal{Z}) := \quad & \inf_{\lambda \in \mathbb{R}} && \lambda \\ & \text{subject to} && \lambda \ge v_i \,,\, \forall\, i \,. \end{aligned}$$

Hence, the robust counterpart takes the form:

$$\exists\lambda \in \mathbb{R}, v \in \mathbb{R}^m, \quad \begin{cases} \lambda - g_*(x, v) \leq 0 \\ \lambda \geq v_i \,, \forall\, i \end{cases} \quad .$$

Proof. Simply put

$$\begin{aligned} \delta^*(v|\mathcal{Z}) &= \sup_{z\in\mathcal{Z}} v^T z = \sup_z \inf_{\gamma,\lambda} v^T z + \sum_i \gamma_i z_i + \lambda(1 - \sum_i z_i) \\ &= \inf_{\gamma\geq 0,\lambda} \sup_z v^T z + \sum_i \gamma_i z_i + \lambda(1 - \sum_i z_i) \\ &= \inf_{\gamma\geq 0,\lambda} \begin{cases} \lambda & \text{if } v_i + \gamma_i - \lambda = 0 \text{ for all } i \\ \infty & \text{otherwise} \end{cases} \\ &= \inf_{\lambda} \begin{cases} \lambda & \text{if } \lambda \geq v_i \text{ for all } i \\ \infty & \text{otherwise} \end{cases} \quad . \end{aligned}$$

□

We now express a theorem that allows one to derive the support function of an uncertainty set known to be the affine projection of another set for which the support function is known.

Theorem 6.7. : *If $\mathcal{Z} \subset \mathbb{R}^m$ is an affine projection of $\mathcal{Z}_1 \subset \mathbb{R}^{m_1}$, namely that $\mathcal{Z} := \{z \in \mathbb{R}^m \,|\, \exists z' \in \mathcal{Z}_1,\, z = Az' + a_0\}$ for some $A \in \mathbb{R}^{m\times m_1}$ and $a_0 \in \mathbb{R}^m$, then $\delta^*(v|\mathcal{Z}) = a_0^T v + \delta^*(A^T v|\mathcal{Z}_1)$.*

Proof. Simply put

$$\delta^*(v|\mathcal{Z}) = \sup_{z\in\mathcal{Z}} v^T z = \sup_{z'\in\mathcal{Z}_1} v^T(Az' + a_0) = v^T a_0 + \delta^*(A^T v|\mathcal{Z}_1) \; .$$

□

We can now employ this relation to derive the support function of the popular budgeted uncertainty set.

Corollary 6.8. : *Consider using the budgeted uncertainty set $\mathcal{Z} := \{z \in \mathbb{R}^m \,|\, -1 \leq z \leq 1,\, \sum_i |z_i| \leq \Gamma\}$, then*

$$\begin{aligned} \delta^*(v|\mathcal{Z}) := \quad & \inf_{\omega^+\in\mathbb{R}^m,\omega^-\in\mathbb{R}^m,\lambda\in\mathbb{R}} && \sum_i \omega_i^+ + \sum_i \omega_i^- + \Gamma\lambda \\ & \text{subject to} && \lambda \geq v_i - \omega_i^+ \,, \forall\, i \\ & && \lambda \geq -v_i - \omega_i^- \,, \forall\, i \\ & && \omega \geq 0,\; \lambda \geq 0 \; . \end{aligned}$$

Hence, the robust counterpart takes the form:

$$\exists \omega^+ \in \mathbb{R}^m, \omega^- \in \mathbb{R}^m, \lambda \in \mathbb{R}, v \in \mathbb{R}^m, \quad \begin{cases} \sum_i \omega_i^+ + \sum_i \omega_i^- + \rho\lambda - g_*(x, v) \leq 0 \\ \lambda \geq v_i - \omega_i^+ , \forall\, i \\ \lambda \geq -v_i - \omega_i^- , \forall\, i \\ \omega^+ \geq 0,\ \omega^- \geq 0,\ \lambda \geq 0 \end{cases} .$$

Proof. First, observe how the budgeted uncertainty set can be described as the following projection $z = [\boldsymbol{I}\ \ -\boldsymbol{I}]z'$ for $z' \in \mathcal{Z}_1 := \{z \in \mathbb{R}^{2m} \,|\, 0 \leq z \leq 1,\, \sum_i z_i \leq \Gamma\}$. Then, based on theorems 6.5 and 6.7, we can conclude that

$$\delta^*(v|\mathcal{Z}) = \delta^*([\boldsymbol{I}\ \ -\boldsymbol{I}]^T v|\mathcal{Z}_1) \ .$$

Hence, we can state that the robust counterpart takes the form

$$\exists \omega^+ \in \mathbb{R}^m, \omega^- \in \mathbb{R}^m, \lambda \in \mathbb{R}, v \in \mathbb{R}^m, \quad \begin{cases} \sum_i \omega_i^+ + \sum_i \omega_i^- + \rho\lambda - g_*(x, v) \leq 0 \\ \lambda \geq v_i - \omega_i^+ , \forall\, i \\ \lambda \geq -v_i - \omega_i^- , \forall\, i \\ \omega^+ \geq 0,\ \omega^- \geq 0,\ \lambda \geq 0 \end{cases} .$$

□

Finally, we work out a theorem that allows us to easily manipulate known conjugate functions to obtain conjugate functions for functions that are obtained by affine transformation.

Theorem 6.9. : *If $g(x, z)$ is a positive affine mapping of $g'(x, z)$, namely that $g(x, z) := \alpha g'(x, z) + \beta$ for some $\alpha > 0$, then $g_*(x, v) = \alpha g'_*(x, v/\alpha) - \beta$.*

Proof. Simply put

$$\begin{aligned} g_*(x, v) &= \inf_{z \in \mathcal{Z}_g} v^T z - g(x, z) \\ &= \inf_{z \in \mathcal{Z}_g} v^T z - \alpha g'(x, z) - \beta \\ &= \alpha \left(\inf_{z \in \mathcal{Z}_g} (v/\alpha)^T z - g'(x, z) \right) - \beta \\ &= \alpha g'_*(x, v/\alpha) - \beta \ . \end{aligned}$$

□

6.2 Reference Tables from Ben-Tal et al. 2015

Table 6.1: Table of reformulations for uncertainty sets (Table 1 in [9])

Uncertainty region	$\mathcal{Z}$	Support function $\delta^*(v\|\mathcal{Z})$
Box	$\\|z\\|_\infty \leq \rho$	$\rho\\|v\\|_1$
Ball	$\\|z\\|_2 \leq \rho$	$\rho\\|v\\|_2$
Polyhedral	$b - Bz \geq 0$	$\inf_{w \geq 0: B^T w = v} b^T w$
Cone	$b - Bz \in C$	$\inf_{w \in C^*: B^T w = v} b^T w$
KL-Divergence	$\sum_l z_l \ln\left(\frac{z_l}{z_l^0}\right) \leq \rho$	$\inf_{u \geq 0} \sum_l z_l^0 u e^{(v_l/u)-1} + \rho u$
Geometric prog.	$\sum_i \alpha_i e^{(d_i)^T z} \leq \rho$	$\inf_{u \geq 0,\, w \geq 0: \sum_i d_i w_i = v} \sum_i \{w_i \ln\left(\frac{w_i}{\alpha_i u}\right) - w_i\} + \rho u$
Intersection	$\mathcal{Z} = \cap_i \mathcal{Z}_i$	$\inf_{\{w_i\}: \sum_i w_i = v} \sum_i \delta^*(w^i\|\mathcal{Z}_i)$
Example	$\mathcal{Z}_k = \{z\|\\|z\\|_k \leq \rho_k\}$ $k = 1, 2$	$\inf_{(w^1, w^2): w^1 + w^2 = v} \rho_1 \\|w^1\\|_\infty + \rho_2 \\|w^2\\|_2$
Minkowski sum	$\mathcal{Z} = \mathcal{Z}_1 + \cdots + \mathcal{Z}_K$	$\sum_i \delta^*(v\|\mathcal{Z}_i)$
Example	$\mathcal{Z}_1 = \{z \mid \\|z\\|_\infty \leq \rho_\infty\}$ $\mathcal{Z}_2 = \{z \mid \\|z\\|_2 \leq \rho_2\}$	$\rho_\infty \\|v\\|_1 + \rho_2 \\|v\\|_2$
Convex hull	$\mathcal{Z} = \mathrm{conv}(\mathcal{Z}_1, \ldots, \mathcal{Z}_K)$	$\max_i \delta^*(v\|\mathcal{Z}_i)$
Example	$\mathcal{Z}_1 = \{z \mid \\|z\\|_\infty \leq \rho_\infty\}$ $\mathcal{Z}_2 = \{z \mid \\|z - z^0\\|_2 \leq \rho_2\}$	$\max\{\rho_\infty \\|v\\|_1, (z^0)^T v + \rho_2 \\|v\\|\\|_2\}$

Table 6.2: Table of reformulations for constraint functions (Table 2 in [9])

Constraint function	$g(x,z)$	Partial concave conjugate $g_*(x,v)$
Linear in z	$z^T g(x)$	$\begin{cases} 0 & \text{if } v = g(x) \\ -\infty & \text{otherwise} \end{cases}$
Concave in z, separable in z and x	$g(z)^T x$	$\sup_{\{s^i\}_{i=1}^n : \sum_{i=1}^n s^i = v} \sum_i x_i (g_i)_*(s^i/x_i)$
Example	$-\sum_i \frac{1}{2}(z^T Q_i z) x_i$	$\sup_{\{s^i\}_{i=1}^n : \sum_{i=1}^n s^i = v} -\frac{1}{2}\sum_{i=1}^n \frac{(s^i)^T Q_i^{-1} s^i}{x_i}$
Sum of functions	$\sum_i g_i(x,z)$	$\sup_{\{s^i\}_{i=1}^n : \sum_i s^i = v} \sum_i (g_i)_*(x, s^i)$
Sum of separable functions	$\sum_i g_i(x, z_i)$	$\sum_{i=1}^n (g_i)_*(x, v_i)$
Example	$-\sum_{i=1}^m x_i^{z_i}$, $x_i > 1, 0 \leq z \leq 1$	$\begin{cases} \sum_{i=1}^m \left(\frac{v_i}{\ln x_i} \ln \frac{-v_i}{\ln x_i} - \frac{v_i}{\ln x_i} \right) & \text{if } v \leq 0 \\ -\infty & \text{otherwise} \end{cases}$

6.3 Exercises

Exercise 6.1. ***(Conditional Value at Risk Portfolio with uncertain probabilities)***

Consider the Conditional Value at Risk minimization problem:

$$\begin{aligned} \underset{x,t}{\text{minimize}} \quad & \text{CVaR}_{1-\epsilon}(r^T x) \\ & \sum_i x_i = 1 \\ & x \geq 0\,, \end{aligned}$$

where $x \in \mathbb{R}^n$, and the uncertainty about $r \in \mathbb{R}^n$ takes the form of a set of equiprobable scenarios $\{\bar{r}_k\}_{k=1}^K$. Now, let us be worried that the uniform distribution might not be the rights one for this problem. Instead we would like to minimize the worst-case CVaR that might be achieved under the following uncertainty set.

$$\mathcal{D}(\rho) := \{p \in \mathbb{R}^K \,|\, p \geq 0,\, \sum_k p_k = 1,\, \sum_{k=1}^K p_k \ln\left(\frac{p_k}{1/K}\right) \leq \rho\}\,.$$

Note that this set looks at distributions that diverge by at most ρ from the uniform distribution. In this context, the robust optimization problem might look like

$$\begin{aligned} \underset{x,t}{\text{minimize}} \quad & t && (6.3a) \\ & \text{CVaR}_{1-\epsilon}(r^T x; p) \leq t\,,\ \forall\, p \in \mathcal{D}(\rho) && (6.3b) \\ & \sum_i x_i = 1 && (6.3c) \\ & x \geq 0\,, && (6.3d) \end{aligned}$$

where $\text{CVaR}_{1-\epsilon}(r^T x; p) := \inf_s s + (1/\epsilon) \sum_{k=1}^K p_k \max(-\bar{r}_k^T x - s;\, 0)$ with $\{p \in \mathbb{R}^K \,|\, p \geq 0,\, \sum_k p_k = 1\}$ as domain which ensures the convexity in x and concavity in p.

Question: Derive a tractable reformulation of this problem as a convex optimization problem of finite dimension ?

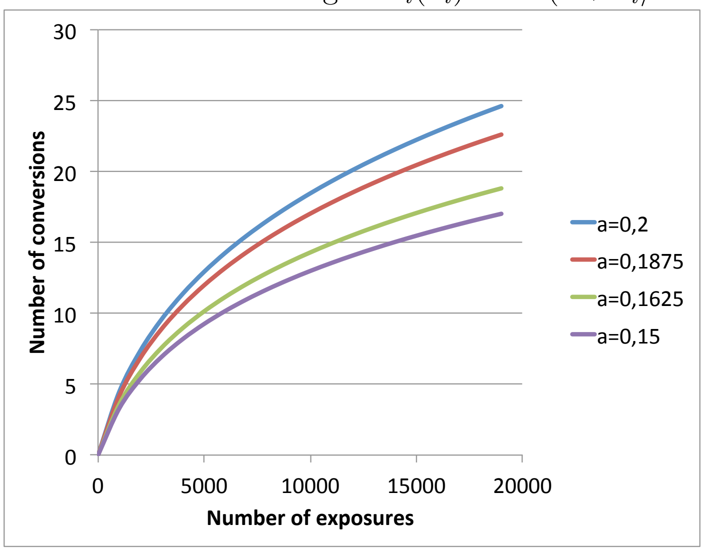
Exercise 6.2. ***(Planning an advertisement campaign with exposure rate uncertainty)***
Consider the problem of investing in an online ad campaign in order to do the promotion of a new app for the Ipad.

$$\begin{aligned} \underset{x}{\text{maximize}} \quad & \sum_i h_i(x_i) \\ \text{subject to} \quad & \sum_i p_i x_i \leq B \\ & x \geq 0\,, \end{aligned}$$

where $x \in \mathbb{R}^n$ identifies how many exposures per day an ad will have on each web site, for each website i, $h_i(\cdot)$ expresses an expected number of converted clients that originated from site i, and c_i is the cost of a single exposure on each website, and B is the total daily budget for the ad campaign.

In practice, it is particularly difficult to estimate the conversion function $h_i(\cdot)$ for each site. It is however expected that the conversion rate (i.e. the number of converted customer per additional ad) decreases as more ads are being displayed. For this reason, it is reasonable to think that the function $h_i(\cdot)$ would behave as $h_i(x_i) := c_i(1 + x_i/d_i)^{a_i} - c_i$ for some $0 < a_i \leq 1$, $c_i > 0$, and $d_i > 0$. Here are a few examples of such a parametric function.

Figure of the converted number of customers per ad displayed on a website according to $h_i(x_i) := 30(1 + x_i/1000)^{a_i} - 30$



Question 1: Present a tractable reformulation for the above problem considering that the uncertainty set for the vector $a \in \mathbb{R}^n$ is $\mathcal{U}_1 := \{a \in \mathbb{R}^n \,|\, \exists z \in \mathbb{R}^n, 0 \leq z \leq 1, \sum_i z_i \leq \Gamma,\, a_i = \bar{a}_i(1 - 0.25z_i)\,, \forall\, i\}$ where $\bar{a} \in [0,\, 1]^n$?

Question 2: How would the reformulation change if the following set was used instead:

$$\mathcal{U}_2 := \{a \in \mathbb{R}^n \,|\, \exists z \in \mathbb{R}^n, z \geq 0,\, \sum_i z_i \leq 1,\, \sum_i z_i \ln(z_i) \leq \rho,\, a_i = \bar{a}_i(1 - 0.25z_i)\,, \forall\, i\}\,?$$

(Hint: you can assume that the worst-case for a always occurs when $\sum_i z_i = 1$.)

Question 3: Solve the reformulation obtained in question #1 using a non-linear programming solver under conditions where $p := [0.110\ 0.085\ 0.090\ 0.080]^T$ and $B = 1$.

Exercise 6.3. ***(More robust non-linear reformulations)***

Consider the robust optimization problem:

$$\begin{aligned} \underset{x}{\text{maximize}} \quad & \min_{z \in \mathcal{Z}} \sum_i x_i \exp(z_i) \\ \text{subject to} \quad & \sum_i x_i \leq 1 \\ & x \geq 0 \ , \end{aligned}$$

where

$$\mathcal{Z} := \{ z \in \mathbb{R}^n \mid \exists v \in [-1,1]^n, w \in [-1,1], z = \mu + Q(v + 1 \cdot w), \|v\|_1 \leq \Gamma \} \ .$$

Question: Derive a tractable reformulation of this problem as a convex optimization problem of finite dimension ?

Part III

Advanced Methods

Chapter 7

Globalized Robust Counterpart

7.1 Controlled deterioration beyond $\mathcal{Z}$

Consider the following robust constraint :

$$g(x,z) \leq 0\,,\,\forall\, z \in \mathcal{Z}\;.$$

Although as we have seen, it is an interesting constraint to impose, in practice it might not be sufficient when there is a large amount of uncertainty. Indeed, in such cases, it is tempting to use a large uncertainty set $\mathcal{Z}$ to protect oneself from all potential realizations. On the other hand, such an approach might end up producing robust solutions that are overly conservative. If one instead reduces the size of $\mathcal{Z}$, the solutions are more opportunistic but there is automatically a loss in terms of guarantees that are associated to the proposed robust solution. This is the motivation for formulating a class of robust constraints that can guarantee feasibility over a set $\mathcal{Z}$ while limiting the degree of infeasibility when the realized vector of parameters lands outside of $\mathcal{Z}$, namely inside a larger set $\mathcal{Z}' \supset \mathcal{Z}$ with $\mathcal{Z}' \subseteq \mathbb{R}^m$. Here is the globalized robust counterpart constraint:

$$(GRC) \qquad g(x,z) \leq \psi(z)\,,\,\forall\, z \in \mathcal{Z}'\;,$$

where $\mathcal{Z}' \supset \mathcal{Z}$ captures the set of all possible realizations of z, and $\psi(z)$ is a non-negative convex function of z with the convex domain $\mathcal{Z}_\psi$ and designed such that $\psi(z) = 0$ for all $z \in \mathcal{Z}$. In words, the deterioration function $\psi(z)$ ensures that less than $\psi(z)$ excess "resources" will be needed for any $z \notin \mathcal{Z}$ that is in $\mathcal{Z}'$. (We refer the reader to [8] for similar derivations as presented in this section.)

Theorem 7.1. : *Given that $\mathcal{Z}'$ is bounded*[1] *and that there exists a z in the relative interior of $\mathcal{Z}_\psi$, of the domain of $g(x,\cdot)$ for all x and of $\mathcal{Z}'$, the GRC constraint is*

[1] Note that it is not clear to the author of these lecture notes what prevents the result from applying for general unbounded convex $\mathcal{Z}'$. A later version of these lecture notes might describe the more general result, otherwise the reformulation presented is still guaranteed to lead to a conservative approximation of the GRC constraint.

equivalent to

$$\exists\, w_1 \in \mathbb{R}^m, w_2 \in \mathbb{R}^m,\ \delta^*(w_1 + w_2|\mathcal{Z}') - g_*(x, w_1) - (-\psi)_*(w_2) \leq 0\ ,$$

where $(-\psi)_*(w)$ *is the conjugate function of* $(-\psi(\cdot))$*, namely* $(-\psi)_*(w) := \inf_{z\in\mathcal{Z}_\psi}\ w^T z + \psi(z)$.

Proof. This follows nearly directly from theorem 6.2 where we consider formulating the Fenchel robust counterpart of

$$h(x, z) \leq 0\,,\ \forall\, z \in \mathcal{Z}'\ .$$

with $h(x, z) := g(x, z) - \psi(z)$ and for which all the assumptions made by the theorem are respected. This FRC takes the form

$$\exists\, v \in \mathbb{R}^m,\ \delta^*(v|\mathcal{Z}') - h_*(x, v) \leq 0\ ,$$

where according to table 6.2 the expression for $h_*(x, v)$ can be reformulated as

$$h_*(x, v) = \sup_{(w_1,w_2)\in\mathbb{R}^m\times\mathbb{R}^m : w_1+w_2=v} g_*(x, w_1) + (-\psi)_*(w_2)\ .$$

This leads to the reformulation that is presented in our theorem. □

Corollary 7.2. : *Given that* $\mathcal{Z}'$ *is bounded (see footnote 1) and that there exists a* z *in the relative interior of* $\mathcal{Z}_\psi$*, of the domain of* $g(x,\cdot)$ *for all* x *and of* $\mathcal{Z}$ *and* $\mathcal{Z}'$*. Furthermore, let the deterioration function* $\psi(\cdot)$ *be defined as*

$$\psi(z) := \alpha\, \mathbf{dist}(z, \mathcal{Z})\,,$$

where $\mathbf{dist}(z, \mathcal{Z}) := \inf_{z'\in\mathcal{Z}} \|z' - z\|$ *for some norm* $\|\cdot\|$*, then the GRC constraint is equivalent to*

$$\exists\, w_1 \in \mathbb{R}^m, w_2 \in \mathbb{R}^m,\ \begin{cases} \delta^*(w_1|\mathcal{Z}') - g_*(x, w_1 + w_2) + \delta^*(w_2|\mathcal{Z}) \leq 0 \\ \|w_2\|_* \leq \alpha\ . \end{cases} ,$$

where $\|\cdot\|_*$ *is the dual norm of* $\|\cdot\|$*, namely* $\|w_2\|_* := \sup_{v:\|v\|\leq 1} v^T w_2$.

Finally, if $\mathcal{Z}' = \mathbb{R}^m$*, then the GRC constraint is equivalent to*

$$\exists\, w \in \mathbb{R}^m,\ \delta^*(w|\mathcal{Z}) - g_*(x, w) \leq 0\ \&\ \|w\|_* \leq \alpha\ ,$$

where we can recognize the original robust counterpart with a bound on the dual norm of w.

Proof. In order to apply theorem 7.1, one needs to identify $(-\psi)_*(w)$. We first have that

$$\begin{aligned}(-\psi)_*(w) &:= \inf_{z'} w^T z' + \psi(z') = \inf_{z'} w^T z' + \inf_{z\in\mathcal{Z}} \alpha\|z'-z\| \\ &= \inf_{z\in\mathcal{Z}}(\inf_{z'} w^T z' + \alpha\|z'-z\|) \\ &= \inf_{z\in\mathcal{Z}}(\inf_{\Delta} w^T(\Delta+z) + \alpha\|\Delta\|) \\ &= \inf_{z\in\mathcal{Z}} w^T z + (\inf_{\Delta} w^T\Delta + \alpha\|\Delta\|) \\ &= \begin{cases} -\delta^*(-w|\mathcal{Z}) & \text{if } \|w\|_* \leq \alpha \\ \infty & \text{otherwise.}\end{cases}\end{aligned}$$

where the last equality comes from

$$\begin{aligned}\inf_{\Delta} \lambda^T\Delta + \alpha\|\Delta\| &= \inf_{s\geq 0, w\in\mathbb{R}^m:\|w\|=s} \lambda^T w + \alpha s = \inf_{s\geq 0, v\in\mathbb{R}^m:\|v\|=1} s\lambda^T v + \alpha s \\ &= \inf_{s\geq 0} s\left(\inf_{v\in\mathbb{R}^m:\|v\|\leq 1} \lambda^T v\right) + \alpha s = \inf_{s\geq 0} s(-\|\lambda\|_*) + \alpha s \\ &= \begin{cases} 0 & \text{if } \|\lambda\|_* \leq \alpha \\ -\infty & \text{otherwise}\end{cases}\end{aligned}$$

where we used the definition of the dual norm $\|\lambda\|_* := \sup_{v\in\mathbb{R}^m:\|v\|\leq 1} \lambda^T v$.

Hence we obtain that the GRC constraint is equivalent to

$$\begin{aligned}&\exists\, w_1 \in \mathbb{R}^m, w_2 \in \mathbb{R}^m, \\ &\quad \delta^*(w_1 + w_2|\mathcal{Z}') - g_*(x, w_1) + \delta^*(-w_2|\mathcal{Z}) \leq 0 \\ &\quad \|w_2\|_* \leq \alpha\end{aligned}$$

which reduces to the following when replacing $w_1' := w_1 + w_2$ and $w_2' := -w_2$, so that $w_1 = w_1' + w_2'$:

$$\begin{aligned}&\exists\, w_1' \in \mathbb{R}^m, w_2' \in \mathbb{R}^m, \\ &\quad \delta^*(w_1'|\mathcal{Z}') - g_*(x, w_1' + w_2') + \delta^*(w_2'|\mathcal{Z}) \leq 0 \\ &\quad \|w_2'\|_* \leq \alpha\end{aligned}$$

In the case where $\mathcal{Z}' := \mathbb{R}^m$, then $\delta^*(w_1, \mathcal{Z}')$ reduces to

$$\delta^*(w_1, \mathcal{Z}') := \begin{cases} 0 & \text{if } w_1 = 0 \\ \infty & \text{otherwise}\end{cases}.$$

□

Remark 7.3. : Note that in the seminal paper of [7], the GRC paradigm was originally called "comprehensive robust counterpart", while in [6] the authors decided to rather use the expression "soft robust model" to refer to an analogous formulation. Furthermore, in the original paper and in [10], GRC is presented from the perspective of conic optimization as the idea of replacing the conic constraint:

$$G(x,z) \in \mathcal{K}\,,\, \forall\, z \in \mathcal{Z}\ ,$$

where $G(x,z)$ is a mapping from $\mathbb{R}^n \times \mathbb{R}^m$ to $\mathbb{R}^J$ and $\mathcal{K}$ is a cone, with the constraint

$$\mathbf{dist}(G(x,z'),\mathcal{K}) \leq \alpha\,\mathbf{dist}(z',\mathcal{Z})\,,\, \forall\, z' \in \mathcal{Z}'\ ,$$

which reflects the idea that we wish the vector $G(x,z')$ to be less than some distance away from being in the feasible set when z is some distance away from $\mathcal{Z}$. It is important to realize that although our definition of GRC is more general as we can employ $g(x,z) := \mathbf{dist}(G(x,z'),\mathcal{K})$ and $\psi(z) := \alpha\,\mathbf{dist}(z',\mathcal{Z})$, our concavity assumption for $g(x,z)$ does not cover this special case, except if the cone $\mathcal{K}$ is the negative orthant and $\mathbf{dist}(G(x,z'),\mathcal{K}) := \max_j G_j(x,z')$. We choose to focus on this GRC form because it appears more intuitive to apply and unifies somehow the notions of comprehensive and soft robustness. To demonstrate this in the later case, in [6] the authors study the application of our GRC constraint to stochastic programs with distribution ambiguity where they obtained a constraint of the form

$$\sup_{Q\in\mathcal{Q}(\epsilon)} \mathbb{E}_Q[g(x,Z)] \leq \epsilon\,,\, \forall\, \epsilon \in [0,\delta]\ ,$$

where Z is a random vector distributed according to Q, and $\mathcal{Q}(\epsilon)$ is a family of distribution sets parametrized by ϵ such that $\epsilon_1 > \epsilon_2 \rightarrow \mathcal{Q}(\epsilon_1) \supseteq \mathcal{Q}(\epsilon_2)$. In our context, this can be translated as

$$\mathbb{E}_Q[g(x,Z)] \leq \psi(Q)\,,\, \forall\, Q \in \mathcal{Q}(\delta)\ ,$$

where $\psi(Q) := \inf\{\epsilon \in [0,\delta]\,|\,Q \in \mathcal{Q}(\epsilon)\}$. The authors of [6] established a connection with convex risk measures which is straightforward to identify in this form as it is known that for any such measure $\rho(Y)$ applied on expenses, the constraint $\rho(g(x,z)) \leq 0$ can be expressed as

$$\mathbb{E}_Q[g(x,Z)] - \psi(Q) \leq 0\,,\, \forall\, Q \in \mathcal{M}\ ,$$

where $\psi(Q)$ is a convex function of the probability measure Q, and where $\mathcal{M}$ is the set of all probability measures on the measurable space $(\mathbb{R}^m,\mathcal{B})$ with $\mathcal{B}$ the σ-algebra on $\mathbb{R}^m$.

7.2 Examples

Both of the examples below come from [10] Exercise 3.1 and A. Nemirovski's lecture notes Exercise 4.3.

Example 7.4. : Production problem with bounded sensitivity to price drifts

"**Exercise 3.1 from [10]** Consider a situation as follows. A factory consumes n types of raw materials, coming from n different suppliers, to be decomposed into m pure components. The per unit content of component i in raw material j is $p_{ij} \geq 0$, and the necessary per month amount of component i is a given quantity $b_i \geq 0$. You need to make a long-term arrangement on the amounts of raw materials x_j coming every month from each of the suppliers, and these amounts should satisfy the system of linear constraints

$$Px \geq b, P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ p_{m1} & p_{n2} & \cdots & p_{mn} \end{bmatrix}.$$

The current per unit price of products j is c_j; this price, however, can vary in time, and from the history you know the volatilities $v_j \geq 0$ of the prices. How to choose x_j's in order to minimize the total cost of supply at the current prices, given an upper bound α on the sensitivity of the cost to possible future drifts in prices ? "

The GRC approach: One might express the deterministic model as:

$$\begin{aligned} \underset{x,t}{\text{minimize}} \quad & t \\ \text{subject to} \quad & t \geq c^T x \\ & Px \geq b \\ & x \geq 0 \,, \end{aligned}$$

where $x \in \mathbb{R}^n$ is the amount of raw material ordered from each supplier and $t \in \mathbb{R}$ captures the total cost of the order. Finally, the second constraint imposes that the minimum quantity of each pure component is satisfied.

Given that we wish to impose an upper bound of α on the sensitivity of the cost to possible future drifts in prices we will replace the first constraint with the following GRC.

$$t + \alpha\, \mathbf{dist}(z, \mathbf{0}_n) \geq \sum_{i=1}^{n} (c_i + v_i z_i) x_i \,, \, \forall\, z \in \mathbb{R}^n \,,$$

with $\mathbf{dist}(z_1, \mathbf{0}_n) := \|z_1\|_1$. Namely, this constraint can be interpreted as imposing that t^* captures a total cost that is known to only be perturbed by a factor of α of the total relative variation of prices when compared to their expected volatilities. According to corollary 7.2, since the distance measure is a norm, this GRC model is equivalent to the RC model to which we further impose that $\|\,\mathbf{diag}(v)x\|_1^* \leq \alpha$. In particular, we obtain:

$$\begin{array}{ll} \underset{x,t}{\text{minimize}} & t \\ \text{subject to} & t \geq c^T x \\ & v_i x_i \leq \alpha\,,\ \forall\, i = 1,\ldots,n \\ & Px \geq b \\ & x \geq 0\ , \end{array}$$

Example 7.5. : Linear estimator with controlled deterioration

"**Exercise 4.3 from A. Nemirovski's lecture notes** Consider the situation as follows: Unknown signal z known to belong to a given ball $B := \{z \in \mathbb{R}^n \,|\, z^T z \leq 1\}$ is observed according to the relation $y = Az + \xi$, where y is the observation, A is a given $m \times n$ sensing matrix, and ξ is an observation error. Given y, we want to recover a linear form $f^T z$ of z; here f is a given vector. The normal range of the observation error is $\mathcal{U} := \{\xi \in \mathbb{R}^m \,|\, \|\xi\|_2 \leq 1\}$. We are seeking for a linear estimate $\hat{f}(y) = g^T y$.

1. Formulate the problem of building the best, in the minimax sense (*i.e.*, with the minimal worst-case recovering error with respect to $x \in B$ and $\xi \in \mathcal{U}$), linear estimate as the RC of an uncertain LO problem and build tractable reformulation of this RC.
2. Formulate the problem of building a linear estimate with the worst-case, over signals z with $\|z\|_2 \leq 1+\rho_z$ and observation errors ξ with $\|\xi\|_2 \leq 1+\rho_\xi$, risk for all $\rho_z \geq 0$, $\rho_\xi \geq 0$, risk admitting the bound $\tau + \alpha_z\rho_z + \alpha_\xi\rho_\xi$ with given τ, α_x, and α_ξ; thus we want "desired performance" τ of the estimate in the normal range $B \times \mathcal{U}$ of $[z^T\ \xi^T]^T$ and "controlled deterioration of this performance" when z and/or ξ run out of their normal ranges. Build a tractable reformulation of this problem."

The GRC approach: One might answer question #1 using the following model

$$\begin{array}{ll} \underset{g,t}{\text{minimize}} & t \\ \text{subject to} & t \geq |g^T(Az+\xi) - f^T z|\,,\ \forall\,(z,\xi) : \|z\|_2 \leq 1,\ \|\xi\|_2 \leq 1\ , \end{array}$$

where $g \in \mathbb{R}^n$ is the linear estimator we are trying to identify and $t \in \mathbb{R}$ captures the largest error in estimation that can be observed if the realization of z and ξ land in their respective balls. This problem reduces to

$$\begin{array}{ll} \underset{g,t}{\text{minimize}} & t \\ \text{subject to} & t \geq \|g\|_2 + \|A^T g - f\|_2\ . \end{array}$$

When answering question #2, one needs modify the model in order to represent the guarantee that when z or ξ fall outside their respective balls, the estimation error does not deteriorate too much. Let $\text{Ball}_z(r) := \{z \,|\, \|z\|_2 \leq r\}$ and $\text{Ball}_\xi(r) := \{\xi \,|\, \|\xi\|_2 \leq r\}$

respectively be the two balls of radius r for z and ξ. We can formulate the following GRC model:

$$t + \alpha_z \,\mathbf{dist}(z, \mathrm{Ball}_z(1)) + \alpha_\xi \,\mathbf{dist}(\xi, \mathrm{Ball}_\xi(1)) \geq |g^T(Az + \xi) - f^T z| \,,\, \forall\, (z, \xi) \,,$$

where $\mathbf{dist}(z, \mathrm{Ball}_z(1)) = \inf_{z':\|z'\|_2 \leq 1} \|z - z'\|_2$ and $\mathbf{dist}(\xi, \mathrm{Ball}_\xi(1)) = \inf_{\xi':\|\xi'\| \leq 1} \|\xi - \xi'\|_2$. In this new model, we have the guarantee that for the robust linear estimator g^* that is returned, the estimation error will be lower than t^* if the norm of the perturbations are smaller than one, while the estimation error will not increase by more than a factor α_z of the distance from this ball that z ends up achieving, and similarly in terms of the deterioration for ξ's that would fall outside of $\mathrm{Ball}_\xi(1)$.

In order to identify a reduced form for this GRC model, we first divide the constraint in a set of two in order to expose concavity with respect to z and ξ.

$$\begin{aligned} &t + \alpha_z \,\mathbf{dist}(z, \mathrm{Ball}_z(1)) + \alpha_\xi \,\mathbf{dist}(\xi, \mathrm{Ball}_\xi(1)) \geq g^T(Az + \xi) - f^T z \,,\, \forall\, (z, \xi) \\ &t + \alpha_z \,\mathbf{dist}(z, \mathrm{Ball}_z(1)) + \alpha_\xi \,\mathbf{dist}(\xi, \mathrm{Ball}_\xi(1)) \geq -g^T(Az + \xi) + f^T z \,,\, \forall\, (z, \xi) \,. \end{aligned}$$

We will apply corollary 7.2 in two separate steps for each constraint first involving z then ξ. A first application of the theorem implies that for any fixed ξ, the first constraint above reduces to

$$\begin{aligned} &t + \alpha_\xi \,\mathbf{dist}(\xi, \mathrm{Ball}_\xi(1)) \geq g^T \xi + \|A^T g - f\|_2 \,,\, \forall\, \xi \\ &\|A^T g - f\|_2 \leq \alpha_z \end{aligned}$$

And a second application of the theorem for ξ makes the constraint further reduce to

$$\begin{aligned} &t \geq \|g^T\|_2 + \|A^T g - f\|_2 \\ &\|A^T g - f\|_2 \leq \alpha_z \\ &\|g\|_2 \leq \alpha_\xi \,. \end{aligned}$$

Actually when looking at the second set of constraint, we realize that it also reduces to the three constraints above.

We are left with:

$$\begin{aligned} \underset{g,t}{\text{minimize}} \quad & t \\ \text{subject to} \quad & t \geq \|g^T\|_2 + \|A^T g - f\|_2 \\ & \|A^T g - f\|_2 \leq \alpha_z \\ & \|g\|_2 \leq \alpha_\xi \,. \end{aligned}$$

7.3 Relation to Probabilistic Envelopes

In [51], the authors establish a connection between the application of GRC to linear constraints and what they call the imposition of probabilistic envelopes. Namely, they argue that in a context where some uncertain parameters are considered random (i.e. drawn from a distribution F), there are good reasons to be interested in a more extensive version of chance constraint which they call "probabilistic envelope constraint":

$$(PEC) \quad \mathbb{P}(a(Z)^T x \le b(Z) + s(\epsilon)) \ge 1 - \epsilon \,,\, \forall\, \epsilon \in]0,1] \,,$$

for some non-increasing function $s(\epsilon)$. In words, this constraint imposes a lower bounding envelope on the cumulative density function of $a(Z)^T x - b(Z)$.

Assumption 7.6. : The random vector Z has mean equal to zero and a covariance matrix $\Sigma \succ 0$. Moreover, the function $s(\epsilon)$ is a non-increasing convex function.

In the context described by this assumption, since the full knowledge of the distribution of Z might be unavailable, it is natural to consider instead the distributionally robust version of PEC:

$$(DRPEC) \quad \inf_{F \in \mathcal{D}(0,\Sigma)} \mathbb{P}_F(a(Z)^T x \le b(Z) + s(\epsilon)) \ge 1 - \epsilon \,,\, \forall\, \epsilon \in]0,1] \,,$$

where $\mathcal{D}(0,\Sigma)$ is the set of all distributions that have mean equal to zero and covariance matrix equal to Σ. Here is the flavour of the results that are presented in the paper.

Theorem 7.7. : *Given that assumption 7.6 is satisfied, constraint DRPEC is equivalent to the following GRC constraint:*

$$a(z)^T x \le b(z) + s\left(\frac{1}{1 + z^T \Sigma^{-1} z}\right) \,,\, \forall\, z \in \mathbb{R}^m \,.$$

Proof. We first present the probabilistic envelope constraint in terms of the random vector Z.

$$\inf_{F \in \mathcal{D}(0,\Sigma)} \mathbb{P}_F(a(x)^T Z \le b(x) + s(\epsilon)) \ge 1 - \epsilon \,,\, \forall\, \epsilon \in]0,1] \,.$$

This constraint is necessarily equivalent to:

$$\sup_{F \in \mathcal{D}(0,\Sigma)} \mathbb{P}_F(a(x)^T Z > b(x) + s(\epsilon)) \le \epsilon \,,\, \forall\, \epsilon \in]0,1] \,.$$

Then, based on [35], we will use the fact that

$$\begin{aligned} \sup_{F \in \mathcal{D}(0,\Sigma)} \mathbb{P}_F(a'^T Z > b') &= \sup_{F \in \mathcal{D}(0,\Sigma)} \mathbb{P}_F(a'^T Z \ge b') \\ &= \begin{cases} 1 & \text{if } b' < 0 \\ \frac{1}{1 + (b'/\|\Sigma^{1/2} a'\|_2)^2} & \text{if } b' \ge 0 \text{ and } a' \ne 0 \\ 0 & \text{otherwise (i.e. } b' \ge 0 \text{ and } a' = 0) \end{cases} \,. \end{aligned}$$

To reformulate the constraint, one can consider two cases. Either $a(x) \ne 0$ and we get that DRPEC is equivalent to

$$b(x) + s(\epsilon) \ge 0 \qquad \& \qquad \frac{1}{1 + ((b(x) + s(\epsilon)/\|\Sigma^{1/2} a(x)\|_2)^2} \le \epsilon \,.$$

or $a(x) = 0$ and then DRPEC is equivalent to $b(x) + s(\epsilon) \geq 0$. Yet, in both case we can simply impose:

$$b(x) + s(\epsilon) \geq \sqrt{\frac{1-\epsilon}{\epsilon}} \|\Sigma^{1/2} a(x)\|_2 \, , \, \forall \, \epsilon \in]0, 1]$$

Further processing this constraint we get a robust constraint in both z and ϵ:

$$a(x)^T z \leq b(x) + s(\epsilon) \, , \, \forall \, z : \|\Sigma^{-1/2} z\|_2 \leq \sqrt{\frac{1-\epsilon}{\epsilon}} \, , \, \forall \, \epsilon \in]0, 1] \, .$$

Furthermore, we can easily inverse the two enumerations over z and ϵ to obtain

$$a(x)^T z \leq b(x) + s(\epsilon) \, , \, \forall \, 0 < \epsilon \leq 1 : \sqrt{\frac{1-\epsilon}{\epsilon}} \geq \|\Sigma^{-1/2} z\|_2 \, , \, \forall \, z \, .$$

Yet, since $s(\epsilon)$ is non-decreasing, and the expression $(1-\epsilon)/\epsilon$ is decreasing, we know that the constraint is tightest when $\epsilon = 1/(1 + z^T \Sigma^{-1} z)$. This leads to the following GRC:

$$a(x)^T z \leq b(x) + s\left(\frac{1}{1 + z^T \Sigma^{-1} z}\right) \, , \, \forall \, z \, .$$

□

Remark 7.8. : It is worth mentioning that similar results can be obtained for the PEC when the distribution is known and that it has an ellipsoidal distribution. Namely, cases for which there exists a linear transformation of Z that make the super-level sets of the density be spheres centred at the mean. Some additional technical conditions are however required for the associated deterioration function $\psi(z)$ to be a convex function. See all the details in [51].

Chapter 8

Distributionally Robust Optimization

Distributionally robust optimization refers to a decision model that is based on stochastic programming but for which the knowledge of the distribution of Z is incomplete. In particular, one might consider the following optimization model

$$\underset{x\in\mathcal{X}}{\text{maximize}} \quad \mathbb{E}[g_0(x,Z)] \tag{8.1a}$$

$$\text{subject to} \quad \mathbb{E}[g_j(x,Z)] \leq b_j\,,\ \forall j=1,\ldots,J \tag{8.1b}$$

where $Z \in \mathbb{R}^m$ is a random vector, $g_0(x,z)$ is a profit function, each $g_j(x,z)$ captures a performance criterion that we wish to keep below b_j on average. Note that the expression $\mathbb{E}[g(x,Z)]$ is quite flexible in terms of what it can capture. For instance, it can capture the probability that a certain resource be depleted by using $g(x,z) := 0$ if $f(x,z) \leq d$ and $g(x,z) := 1$ otherwise, where $f(x,z)$ is the function that computes how many resources are used with the production plan x when z occurs. Indeed in this example,

$$\begin{aligned}\mathbb{E}[g(x,Z)] &= \mathbb{P}(f(x,Z)\leq d)\cdot\mathbb{E}[g(x,z)|f(x,z)\leq d] + \mathbb{P}(f(x,Z)>d)\cdot\mathbb{E}[g(x,z)|f(x,z)>d] \\ &= \mathbb{P}(f(x,Z)\leq d)\cdot 0 + \mathbb{P}(f(x,Z)>d)\cdot 1 = \mathbb{P}(f(x,Z)>d).\end{aligned}$$

The expression $\mathbb{E}[g_0(x,z)]$ can also be used to capture risk aversion through the expected utility theory when using $g_0(x,z) := u(f(x,z))$, where $f(x,z)$ is some revenue function while $u(\cdot)$ is a non-decreasing concave utility function (see [48] for more details about expected utility theory).

The DRO framework questions the classical assumption that the distribution of Z, which we call F, is exactly known. In practice, there is often many reasons to be doubtful about any particular choice for F. This is especially the case in situations where decision models are designed based on historical observations of the random vector Z. These observations could take the form of independent and identically distributed samples from F and would then typically be used to estimate the parameters of distribution form such as a normal distribution, Poisson distribution, Weibull distribution, etc. When the random vector is large, then it can easily be the case that there are many distribution models that could explain the data equally well. Selecting one of these in order to construct and solve a stochastic program such as (8.1) might give rise to what's called the "Optimizer's curse" (see [45]), i.e. identifying a solution that over exploits the described distribution model resulting in an optimistic bias about future performance which can lead to post-decision disappointment in out-of-sample tests. For this reason, the DRO paradigm suggests to drop the assumption of a known distribution F but to rather identify a distribution set $\mathcal{D}$ assumed to contain the true distribution. Confronted with such ambiguity about F, DRO follows an ambiguity aversion principle that replaces the stochastic program with

$$\underset{x \in \mathcal{X}}{\text{maximize}} \quad \inf_{F \in \mathcal{D}} \mathbb{E}_F[g_0(x, Z)] \tag{8.2a}$$

$$\text{subject to} \quad \mathbb{E}_F[g_j(x, Z)] \leq b_j\,,\ \forall\, F \in \mathcal{D},\ \forall j = 1, \ldots, J\,. \tag{8.2b}$$

Note that for each constraint, the DRO model will make sure that the expected value of $g_j(x, Z)$ is smaller or equal to b_j for all $F \in \mathcal{D}$, and will use as objective value the worst-case expected value of $g_0(x, Z)$ achieved by any distribution F in $\mathcal{D}$.

In this chapter, we will focus on the problem

$$\underset{x \in \mathcal{X}}{\text{minimize}} \quad \sup_{F \in \mathcal{D}} \mathbb{E}_F[h(x, \xi)]\,, \tag{8.3a}$$

where ξ is the random vector drawn from a distribution F, since it is the form that is most commonly used in the literature, but the results we obtain can easily be adapted to the model in (8.2) (e.g. consider $h_0(x, z) := -g_0(x, z)$).

8.1 Moment based models

In this type of approach, the random variable is assumed to have a continuous support and only a number of moments are known for the distribution F.

8.1.1 Mean and support models

Perhaps the most famous uncertainty set in this category is one that accounts for the mean and support of the distribution as follows:

$$\mathcal{D}(\mathcal{Z}, \mu) = \left\{ F \in \mathcal{M} \left| \begin{array}{l} \mathbb{P}_F(\xi \in \mathcal{Z}) = 1 \\ \mathbb{E}_F[\xi] = \mu \end{array} \right. \right\}\,,$$

where $\mathcal{M}$ is the set of all probability measures on the measurable space $(\mathbb{R}^m, \mathcal{B})$, where $\mathcal{B}$ is the Borel σ-algebra on $\mathbb{R}^m$, and where $\mathcal{Z} \subseteq \mathbb{R}^m$ is a Borel set (e.g. a closed set in $\mathbb{R}^m$).

This ambiguity set is actually the one that is used in probability inequalities such as Markov inequality which states that if ξ is a non-negative random variable with an expected value of μ then

$$\mathbb{P}(\xi \geq a) \leq \sup_{F \in \mathcal{D}([0,\infty[,\mu)} \mathbb{P}_F(\xi \geq a) = \mu/a\,.$$

The biggest conceptual challenge in dealing with this type of robust optimization model is the fact that, unlike previous robust optimization models that we encountered, nature does not control a finite dimensional vector but rather a function $F : \mathcal{B} \to \mathbb{R}^+$ (intuitively, $F : \mathbb{R}^m \to \mathbb{R}$ when F is a density function). Yet the analysis remains quite similar to what we have seen up to this point. For any fixed decision x, we would like to employ duality theory to reformulate the worst-case analysis problem as an infimum over a set of additional auxiliary variables in order to reintegrate the result of this analysis in the DRO model.

Looking more closely at the worst-case analysis problem, one might be able to recognize that it is actually a linear program but of semi-infinite dimension (infinite number of decision variables, and finite number of constraints). In particular, we are interested in

$$\begin{aligned}
\underset{F \in \mathcal{M}}{\text{maximize}} \quad & \int_{\mathcal{Z}} h(x,\xi) dF(\xi) && \text{(8.4a)}\\
\text{subject to} \quad & \int_{\mathcal{Z}} dF(\xi) = 1 && \text{(8.4b)}\\
& \int_{\mathcal{Z}} \xi dF(\xi) = \mu\,, && \text{(8.4c)}
\end{aligned}$$

where we further assume that $h(x, \cdot)$ is real-valued measurable in $(\mathbb{R}^m, \mathcal{B})$.

Based on the theory of semi-infinite conic programs (see Theorem 3.4 in [43]), one can establish that the following semi-infinite program is the dual problem and that strong duality applies as long as the ambiguity set $\mathcal{D}(\mathcal{Z}, \mu) \neq \emptyset$, i.e. that there exists an $F \in \mathcal{D}(\mathcal{Z}, \mu)$:

$$\begin{aligned}
\underset{r,q}{\text{minimize}} \quad & \mu^T q + r && \text{(8.5a)}\\
\text{subject to} \quad & z^T q + r \geq h(x,z)\,,\, \forall\, z \in \mathcal{Z}\,, && \text{(8.5b)}
\end{aligned}$$

where $r \in \mathbb{R}$ and $q \in \mathbb{R}^m$ are respectively the dual variables associated with constraints (8.4b) and (8.4c). One can for instance formulate the Lagrangian equation for this problem as:

$$\begin{aligned}
L(F, r, q) &:= \int_{\mathcal{Z}} h(x,\xi) dF(\xi) + r\left(1 - \int_{\mathcal{Z}} dF(\xi)\right) + q^T\left(\mu - \int_{\mathcal{Z}} \xi dF(\xi)\right)\\
&= r + \mu^T q + \int_{\mathcal{Z}} \left(h(x,\xi) - r - q^T \xi\right) dF(\xi)\,.
\end{aligned}$$

The dual problem is obtained through

$$\sup_F \inf_{r,q} L(F,r,q) \leq \inf_{r,q} \sup_F L(F,r,q) = \inf_{r,q} \begin{cases} r + \mu^T q & \text{if } h(x,z) - r - q^T z \leq 0\,,\ \forall\, z \in \mathcal{Z} \\ \infty & \text{otherwise} \end{cases} ,$$

with equality being met when $\mathcal{D}(\mathcal{Z},\mu) \neq \emptyset$.

Theorem 8.1. : *Let $\mathcal{D}(\mathcal{Z},\mu)$ be a distribution set for which there exists a feasible solution $F_0 \in \mathcal{D}(\mathcal{Z},\mu)$, then the moment problem* (8.4) *is equivalent to the following robust optimization problem:*

$$\underset{q}{\text{minimize}} \quad \sup_{z\in\mathcal{Z}} h(x,z) + (\mu - z)^T q\,. \tag{8.6}$$

It is worth presenting a little more intuition about this dual problem. Indeed, one can easily demonstrate that it provides an upper bound for the worst-case expectation of $h(x,\xi)$ by considering that for any pair (q,r) that allows the affine function $\hat{h}(x,\xi) := q^T\xi + r$ be a global over-estimator of $h(x,\xi)$, it must be that

$$\begin{aligned} h(x,z) \leq q^T z + r\,,\ \forall\, z \in \mathcal{Z} \;\Rightarrow\; \sup_{F\in\mathcal{D}(\mathcal{Z},\mu)} \mathbb{E}_F[h(x,\xi)] &\leq \sup_{F\in\mathcal{D}(\mathcal{Z},\mu)} \mathbb{E}_F[q^T\xi + r] \\ &= \sup_{F\in\mathcal{D}(\mathcal{Z},\mu)} q^T\mathbb{E}_F[\xi] + r \\ &= q^T\mu + r\ . \end{aligned}$$

Hence, the dual problem simply attempts to find the tightest affine global over-estimator of $h(x,\xi)$ for which it then becomes easy to evaluate the worst-case expectation (by the linearity property of expectation). As before, the strength of duality theory stands in establishing conditions under which this upper bound is tight.

Another valuable intuition that can be extracted from this dual problem builds upon the fact that this is a bounded (since the primal is feasible) linear program with a decision vector in $\mathbb{R}^{m+1}$. Indeed, it is well known that in a bounded finite dimensional linear program, only $m+1$ constraints are actually needed to identify any optimal solution (i.e. an optimal vertex of the polyhedron). Considering that this is also the case for semi-infinite linear programs such as problem (8.5), one is left with the conclusion that there exists a subset $\mathcal{Z}^* := \{z_1^*, z_2^*, \ldots, z_{m+1}^*\}$ of $\mathcal{Z}$ for which the following finite linear program returns the same optimal value as problem (8.5):

$$\underset{r,q}{\text{minimize}} \quad \mu^T q + r \tag{8.7a}$$

$$\text{subject to} \quad z^T q + r \geq h(x,z)\,,\ \forall\, z \in \mathcal{Z}^*\,. \tag{8.7b}$$

Since we assumed that $h(x,z)$ is real-valued, this LP is feasible, hence taking once more the dual of this problem we get that the following finite dimensional LP returns the same optimal value as (8.4):

$$\begin{aligned}
\underset{p\in\mathbb{R}^{m+1}}{\text{maximize}} \quad & \sum_{i=1}^{m+1} p_i h(x, z_i^*) && (8.8a)\\
\text{subject to} \quad & \sum_i p_i = 1 && (8.8b)\\
& \sum_{i=1}^{m+1} p_i z_i^* = \mu\,. && (8.8c)
\end{aligned}$$

The intuition we get from these arguments is that there always exists a worst-case distribution for problem (8.4) that is supported on at most $m+1$ points in $\mathcal{Z}$, although we do not know a priori which are these points and that these might depend on x. Actually, this intuition is confirmed by the following theorem which can be found as Lemma 3.1 in [43] but originally attributed to [40].

Theorem 8.2. : *Let $\mathcal{Z} \in \mathbb{R}^m$ be a Borel set, and F_0 be some feasible distribution according to $\mathcal{D}(\mathcal{Z}, \mu)$, then problem* (8.4) *is equivalent to the following finite dimensional optimization problem*

$$\begin{aligned}
\underset{p,\{z_i\}_{i=1}^{m+1}}{\text{maximize}} \quad & \sum_{i=1}^{m+1} p_i h(x, z_i) && (8.9a)\\
\text{subject to} \quad & \sum_{i=1}^{m+1} p_i = 1 \;\;\&\;\; p \geq 0 && (8.9b)\\
& \sum_{i=1}^{m+1} p_i z_i = \mu && (8.9c)\\
& z_i \in \mathcal{Z}\,,\ \forall\, i = 1, \ldots, m+1\,, && (8.9d)
\end{aligned}$$

where $p \in \mathbb{R}^{m+1}$ and each $z_i \in \mathbb{R}^m$.

Example 8.3. : Consider a certain step we did when proving Lemma 3.6 which required establishing a tight upper bound for $\mathbb{E}[\exp(\gamma a_i \xi)]$ knowing that the distribution of ξ is symmetric and supported on $[-1, 1]$. Based on this conditions, it is necessarily the case that $\mathbb{E}[\xi] = 0$ and one might instead study

$$\sup_{F\in\mathcal{D}([-1,1],0)} \mathbb{E}_F[\exp(\gamma a_i \xi)]\ .$$

Based on Theorem 8.2, we now know that the worst-case distribution of this problem is supported on two points since $m = 1$. Furthermore, we also know based on equation (8.6) that the bound is exactly equal to:

$$\sup_{F\in\mathcal{D}([-1,1],0)} \mathbb{E}_F[\exp(\gamma a_i \xi)] = \inf_q \sup_{z\in[-1,1]} \exp(\gamma a_i z) - zq\ .$$

Now since $\exp(\gamma a_i z) - zq$ is convex in z, it must be that the maximum is achieved at either $z = -1$ or $z = 1$, hence

$$\sup_{F \in \mathcal{D}([-1,1],0)} \mathbb{E}_F[\exp(\gamma a_i \xi)] = \inf_q \max(\exp(\gamma a_i) - q \ ; \ \exp(-\gamma a_i) + q) \ .$$

Looking now at the minimization in q we realize that the optimum is achieved at the intersection of both affine functions, namely when

$$\exp(\gamma a_i) - q = \exp(-\gamma a_i) + q \ .$$

This gives us $q^* = (1/2)(\exp(\gamma a_i) - \exp(-\gamma a_i))$ which can be reinserted in the expression above:

$$\sup_{F \in \mathcal{D}([-1,1],0)} \mathbb{E}_F[\exp(\gamma a_i \xi)] = \max(\exp(\gamma a_i) - q^* \ ; \ \exp(-\gamma a_i) + q^*) = (1/2)(\exp(\gamma a_i) + \exp(-\gamma a_i)) \ .$$

This is indeed what we had observed before as a bound. This bound is tight whether we look for symmetric distributions or distributions with expected values of 0 since in both case the worst-case distribution puts half of the weight at both -1 and 1.

Under the additional hypothesis that $\mathcal{Z}$ is a convex set and that $h(x, z) := \max_{k=1,\ldots,K} h_k(x, z)$ for some K with each $h_k(x, z)$ a concave function of z, i.e. that $h(x, z)$ is piecewise concave in z, then one can actually show that there always exists a worst-case distribution supported on at most K points, each of which lie on the region of $\mathbb{R}^m$ where the respective function $h_k(x, z)$ achieves a larger value than $h_{k'}(x, \xi)$ for all $k' \neq k$.

Theorem 8.4. : *When $\mathcal{Z}$ is a convex set and $h(x, z) := \max_{k=1,\ldots,K} h_k(x, z)$ for some K with each $h_k(x, z)$ a concave function of z, then problem (8.4) is equivalent to*

$$\begin{aligned}
\underset{p, \{z_k\}_{k=1}^K}{\text{maximize}} \quad & \sum_{k=1}^K p_k h_k(x, z_k) && (8.10a)\\
\text{subject to} \quad & \sum_{k=1}^K p_k = 1 \ , \ p \geq 0 && (8.10b)\\
& \sum_{k=1}^K p_k z_k = \mu && (8.10c)\\
& z_k \in \mathcal{Z} \ , \ \forall \, k = 1, \ldots, K \ . && (8.10d)
\end{aligned}$$

Proof. This proof is inspired by the proof of theorem 6.2 in [47] and decomposes as two steps. In the first step, we show that problem (8.10) is a conservative approximation of problem (8.4), namely that any of its feasible solution can be used to construct a feasible solution to problem (8.4) that achieves an objective value that is larger or equal to the objective value measured in (8.10). Secondly, we show that the same can be said of problem (8.4) being a conservative approximation of problem (8.10).

Step #1: Given any feasible solution to problem (8.10), one can construct the following candidate for problem (8.4).

$$p'_i := \begin{cases} p_k & \text{if } i \leq K \\ 0 & \text{otherwise} \end{cases} \qquad\qquad z'_i := \begin{cases} z_k & \text{if } i \leq K \\ z_1 & \text{otherwise} \end{cases} .$$

We start by verifying that this candidate is feasible in problem (8.4):

$$\begin{aligned}
&\sum_{i=1}^{m+1} p'_i = \sum_{k=1}^{K} p_k = 1 \\
&p'_i = p_k \geq 0\,,\ \forall\, i = 1, \ldots, K \\
&p'_i = 0 \geq 0\,,\ \forall\, i = K+1, \ldots, m+1 \\
&\sum_{i}^{m+1} p'_i z'_i = \sum_{k=1}^{K} p_k z_k = \mu \\
&z'_i = z_k \in \mathcal{Z}\,,\ \forall\, i = 1, \ldots, K \\
&z'_i = z_1 \in \mathcal{Z}\,,\ \forall\, i = K+1, \ldots, n\,.
\end{aligned}$$

We can then verify the second claim about the relation between the two objective values:

$$\sum_{i=1}^{m+1} p'_i h(x, z'_i) \ = \ \sum_{k=1}^{K} p_k h(x, z_k) \ \geq \ \sum_{k=1}^{K} p_k h_k(x, z_k)\,.$$

Step #2: We first define the following partition of $\mathcal{Z}$:

$$\mathcal{Z}_k := \{z \in \mathcal{Z} \,|\, h_k(x,z) \geq h_{k'}(x,z)\,,\ \forall\, k' < k \text{ and } h_k(x,z) > h_{k'}(x,z)\,,\ \forall\, k' > k\}\,.$$

Now, given any feasible solution to problem (8.4), we can construct the following candidate for problem (8.10):

$$p'_k := \sum_{i \in \mathcal{I}_k} p_i \qquad\qquad z'_k := \frac{1}{p'_k} \sum_{i \in \mathcal{I}_k} p_i z_i\,,$$

where $\mathcal{I}_k := \{i \in \{1, \cdots, m+1\} \,|\, z_i \in \mathcal{Z}_k\}$. We start by verify that this candidate is feasible:

$$\begin{aligned}
&\sum_{k=1}^K p'_k = \sum_{k=1}^K \sum_{i\in\mathcal{I}_k} p_i = \sum_i p_i = 1\\
&p'_k = \sum_{i\in\mathcal{I}_k} p_i \geq 0\,,\ \forall\, k = 1,\ldots,K\\
&\sum_{k=1}^K p'_k z'_k = \sum_{k=1}^K p'_k \left(\frac{1}{p'_k}\sum_{i\in\mathcal{I}_k} p_i z_i\right) = \sum_i p_i z_i = \mu\\
&z'_k = \left(\frac{1}{p'_k}\sum_{i\in\mathcal{I}_k} p_i z_i\right) \in \mathcal{Z}\,,\ \forall\, k = 1,\ldots,K\,,
\end{aligned}$$

where the last conditions follows from the fact that $\mathcal{Z}$ is a convex set. We can then verify the second claim about the relation between the two objective values:

$$\begin{aligned}
\sum_{k=1}^K p'_k h_k(x, z'_k) &= \sum_{k=1}^K p'_k h_k \left(x, \left(\frac{1}{p'_k}\sum_{i\in\mathcal{I}_k} p_i z_i\right)\right)\\
&\geq \sum_{k=1}^K p'_k \left(\frac{1}{p'_k}\sum_{i\in\mathcal{I}_k} p_i h_k(x, z_i)\right)\\
&= \sum_{k=1}^K \sum_{i\in\mathcal{I}_k} p_i h(x, z_i) \ = \ \sum_i p_i h(x, z_i)\,.
\end{aligned}$$

This completes our proof. □

One can actually simply exploit the result of theorem 8.4 to conclude that when $h(x, z)$ is a concave function of z, the worst-case distribution ends up being one that puts all of its mass at μ. Hence, the DRO problem reduces to a very trivial form.

Corollary 8.5. : *When $\mathcal{Z}$ is a convex set and $h(x, z)$ is a concave function of z, if $\mathcal{D}(\mathcal{Z}, \mu) \neq \emptyset$, then the DRO problem presented in* (8.3) *is equivalent to*

$$\underset{x\in\mathcal{X}}{\text{minimize}} \quad h(x, \mu)\,.$$

Now that we have built some much needed intuition about this dual reformulation and about the structure of worst-case distributions. We can turn ourself toward the reformulation of a DRO problem as in (8.3). Indeed, following the same steps as used for robust optimization, we get the following result.

Theorem 8.6. : *Let $\mathcal{D}(\mathcal{Z}, \mu)$ be a distribution set for which there exists a feasible solution $F_0 \in \mathcal{D}(\mathcal{Z}, \mu)$, the DRO problem presented in* (8.3) *is equivalent to the following robust optimization problem:*

$$\underset{x\in\mathcal{X}, q}{\text{minimize}} \qquad \sup_{z\in\mathcal{Z}} h(x, z) - z^T q + \mu^T q\,. \tag{8.11}$$

Moreover, the problem can be reformulated as follows when $\mathcal{Z}$ is a convex set and $h(x,z) := \max_k h_k(x,z)$ where each $h_k(x,z)$ is a concave function of z:

$$\begin{aligned} \underset{x\in\mathcal{X},q,\{v_k\}_k,t}{\text{minimize}} \quad & t + \mu^T q \\ \text{subject to} \quad & t \geq \delta^*(v_k \,|\, \mathcal{Z}) - h_*^k(x, v_k + q)\,,\, \forall\, k\,, \end{aligned}$$

where for each k, $v_k \in \mathbb{R}^m$, while $\delta^*(v|\mathcal{Z})$ is the support function of $\mathcal{Z}$ and $h_*^k(x,v)$ is the partial concave conjugate function of $h_k(x,z)$.

Proof. This follows naturally from reintegrating problem (8.5) in the DRO problem to obtain

$$\begin{aligned} \underset{x\in\mathcal{X},r,q}{\text{minimize}} \quad & \mu^T q + r \\ \text{subject to} \quad & z^T q + r \geq h(x,z)\,,\, \forall\, z \in \mathcal{Z}\,, \end{aligned}$$

and simply replacing r by the minimum it can achieve based on the robust constraint.

In the case, of a convex $\mathcal{Z}$ and a piecewise concave $h(x,z)$, then problem (8.11) is equivalent to

$$\underset{x\in\mathcal{X},q}{\text{minimize}} \quad \sup_{z\in\mathcal{Z}} \max_k h_k(x,z) + (\mu - z)^T q$$

which is itself equivalent to

$$\underset{x\in\mathcal{X},q}{\text{minimize}} \quad \max_k \sup_{z\in\mathcal{Z}} h_k(x,z) + (\mu - z)^T q$$

so that we can represent this problem in epigraph form:

$$\begin{aligned} \underset{x\in\mathcal{X},q,t}{\text{minimize}} \quad & t + \mu^T q \\ \text{subject to} \quad & t \geq \sup_{z\in\mathcal{Z}} h_k(x,z) - z^T q\,,\, \forall\, k\,. \end{aligned}$$

By exploiting Theorem 6.2 on each of the constraints, one obtains in each case

$$\exists v \in \mathbb{R}^m,\, t \geq \delta^*(v \,|\, \mathcal{Z}) - g_*(x,v)\,,$$

where $g_*^k(x,v)$ is the partial concave conjugate function of $g^k(x,z) := h_k(x,z) - q^T z$. The latter can be expanded to

$$g_*^k(x,v) := \inf_{z\in\mathcal{Z}_h} v^T z - h_k(x,z) + q^T z = \inf_{z\in\mathcal{Z}_h} (v+q)^T z - h_k(x,z) = h_*^k(x, v+q)\,.$$

This completes our proof. $\square$

8.1.2 Other moment functions

One might now wonder whether more sophisticated ambiguity sets can be used instead of $\mathcal{D}(\mathcal{Z},\mu)$. In fact, Wiesemann et al. [50] show how $\mathcal{D}(\mathcal{Z},\mu)$ can actually capture an extensive list of interesting ambiguity sets.

Example 8.7. : Consider that ξ is a random variable known to have a mean μ, and a variance of $\mathbb{E}[(\xi-\mu)^2]=\sigma^2$. This gives rise to the following DRO problem :

$$\underset{x\in\mathcal{X}}{\text{minimize}} \quad \sup_{F\in\mathcal{D}(\mu,\sigma^2)} \mathbb{E}_F[h(x,\xi)] \;,$$

where

$$\mathcal{D}(\mu,\sigma^2) := \{F\,|\,\mathbb{P}_F(\xi\in\mathbb{R})=1,\ \mathbb{E}_F[\xi]=\mu,\ \mathbb{E}_F[(\xi-\mu)^2]=\sigma^2\}\;.$$

Of course, this does not appear like a set that fits the description we have used until now. However, one can consider a lifted space where a DRO model that fits our assumptions will be exactly equivalent to this one. Namely, consider the lifting $\zeta = [\ \zeta_1 \quad \zeta_2\ ]^T := [\xi\ ,\ (\xi-\mu)^2]^T$ with the following support set

$$\mathcal{Z}' := \{z'\in\mathbb{R}^2\,|\,z_2'=(z_1'-\mu)^2\}\;.$$

One can actually show that the following DRO problem is equivalent to the first one:

$$\underset{x\in\mathcal{X}}{\text{minimize}} \quad \sup_{F'\in\mathcal{D}(\mathcal{Z}',[\mu\ ,\ \sigma^2]^T)} \mathbb{E}_{F'}[h(x,\zeta_1)]\;.$$

Indeed, for any random variable ξ with a feasible distribution $F\in\mathcal{D}(\mu,\sigma^2)$, one can construct the random variable $\zeta=(\xi,\ (\xi-\mu)^2)$ for which the distribution F' must lie in $\mathcal{D}(\mathcal{Z}',[\mu\ ,\ \sigma^2]^T)$ since $\mathbb{P}_F(\zeta\in\mathcal{Z}')=1$ and $\mathbb{E}_F[\zeta]=[\mu\ ,\ \sigma^2]^T$. Furthermore we have that $\mathbb{E}_F[h(x,\zeta_1)]=\mathbb{E}_F[h(x,\xi)]$ so that

$$\sup_{F\in\mathcal{D}(\mu,\sigma^2)} \mathbb{E}_F[h(x,\xi)] \leq \sup_{F'\in\mathcal{D}(\mathcal{Z}',[\mu\ ,\ \sigma^2]^T)} \mathbb{E}_{F'}[h(x,\zeta_1)]\;.$$

Alternatively, letting ζ be any random variable with a distribution $F'\in\mathcal{D}(\mathcal{Z}',[\mu\ ,\ \sigma^2]^T)$, we can simply consider the distribution of $\xi:=\zeta_1$ as being a member of $\mathcal{D}(\mu,\sigma^2)$ since $\mathbb{E}_{F'}[\xi]=\mathbb{E}_{F'}[\zeta_1]=\mu$ and $\mathbb{E}_{F'}[(\xi-\mu)^2]=\mathbb{E}_{F'}[(\zeta_1-\mu)^2]=\mathbb{E}_{F'}[\zeta_2]=\sigma^2$. Finally, we have that $\mathbb{E}_{F'}[h(x,\xi)]=\mathbb{E}_{F'}[h(x,\zeta_1)]$ so that

$$\sup_{F'\in\mathcal{D}(\mathcal{Z}',[\mu\ ,\ \sigma^2]^T)} \mathbb{E}_{F'}[h(x,\zeta_1)] \leq \sup_{F\in\mathcal{D}(\mu,\sigma^2)} \mathbb{E}_F[h(x,\xi)]\;.$$

Theorem 8.6 can therefore be used to reformulate the DRO model as:

$$\underset{x\in\mathcal{X},q\in\mathbb{R}^2}{\text{minimize}} \quad \sup_{z\in\mathcal{Z}'} h(x,z_1)+(\mu-z_1)q_1+(\sigma^2-z_2)q_2\;.$$

This is equivalent to

$$\underset{x\in\mathcal{X},q\in\mathbb{R}^2}{\text{minimize}}\ \sup_{z_1\in\mathbb{R}} h(x,z_1)+(\mu-z_1)q_1+(\sigma^2-(z_1-\mu)^2)q_2\ .$$

We now consider a special case where $h(x,z)$ is lower bounded by $-B$. Since $h(x,z)\geq -B$ for all $x\in\mathcal{X}$ and $z\in\mathbb{R}$, when $q_2<0$, we have that

$$\begin{aligned}\sup_{z_1\in\mathbb{R}} h(x,z)+(\mu-z_1)q_1+(\sigma^2-(z_1-\mu)^2)q_2 &\geq \sup_{z_1\in\mathbb{R}} -B+\mu q_1+\sigma^2 q_2 - z_1 q_1 - q_2(z_1-\mu)^2\\ &= \infty \text{ (as } z_1\to\infty)\ .\end{aligned}$$

Hence, it must be that $q_2\geq 0$ for the objective function to reach a value smaller than infinity. This being said, one can realize that the DRO is equivalent to

$$\begin{aligned}\underset{x\in\mathcal{X},q_1,q_2\geq 0,t}{\text{minimize}}\quad & t+\mu q_1+(\sigma^2-\mu^2)q_2\\ \text{subject to}\quad & t\geq \sup_{z_1\in\mathbb{R}} h_k(x,z)-(q_1-2q_2\mu)z_1-q_2 z_1^2\ ,\ \forall\, k\ ,\end{aligned}$$

when $h(x,z):=\max_k h_k(x,z)$. One might notice that this latter problem is a standard non-linear robust optimization problem which can be tackled using the methods discussed in Chapter 6 if each $h_k(x,z)$ is convex in x and concave in z.

Example 8.8. : Consider that one has information about the support $\mathcal{Z}$, the mean μ, and an upper bound on the second order moment matrix of the type $\mathbb{E}[\xi\xi^T]\preceq\Sigma$ where $A\preceq B$ refers to the fact that $B-A$ is positive semi-definite, i.e. $z^T(B-A)z\geq 0$ for all $z\in\mathbb{R}^m$. This gives rise to the following DRO problem :

$$\underset{x\in\mathcal{X}}{\text{minimize}}\ \sup_{F\in\mathcal{D}(\mathcal{Z},\mu,\Sigma)} \mathbb{E}_F[h(x,\xi)]\ ,$$

where

$$\mathcal{D}(\mathcal{Z},\mu,\Sigma):=\{F\,|\,\mathbb{P}(\xi\in\mathcal{Z})=1,\ \mathbb{E}[\xi]=\mu,\ \mathbb{E}[\xi\xi^T]\preceq\Sigma\}\ .$$

Of course, this does not appear like a set that fits the description we have used until now. However, one can consider a lifted space where a DRO model that fits our assumptions will be exactly equivalent to this one. Namely, consider the lifting $\zeta\in\mathbb{R}^m\times\mathbb{R}^{m\times m}$ with the following support set

$$\mathcal{Z}':=\{(z_1,Z_2)\in\mathbb{R}^m\times\mathcal{S}^{m\times m}\,|\,z_1\in\mathcal{Z},\ Z_2\succeq z_1z_1^T\}\ , \tag{8.12}$$

where $\mathcal{S}^{m\times m}$ is the space of all $m\times m$ symmetric matrices.

Similarly as was done in the previous example, one can consider the DRO model to be equivalent to

$$\underset{x\in\mathcal{X}}{\text{minimize}}\ \sup_{F\in\mathcal{D}(\mathcal{Z}',(\mu,\Sigma))} \mathbb{E}_F[h(x,\zeta_1)]\ .$$

In this context, since we are imposing $m+m(m+1)/2$ different moments (i.e. only the diagonal and lower triangle of Σ count as different moment constraints), we know that the worst-case distribution will be supported on $(1/2)(m+1)(m+2)$ points. Furthermore, we can reformulate the DRO as

$$\underset{x\in\mathcal{X},q,Q}{\text{minimize}} \quad \sup_{(z_1;Z_2):z_1\in\mathcal{Z},\, \mathcal{Z}_2\succeq z_1 z_1^T} h(x,z_1)+(\mu-z_1)^T q+(\Sigma-Z_2)\bullet Q\,,$$

where $A\bullet B := \sum_{ij} A_{ij}B_{ij}$ indicates the Frobenius inner product, and where $Q\in\mathcal{S}^{m\times m}$ and $Z_2\in\mathcal{S}^{m\times m}$ are symmetric matrices. This problem will reduce to the following if $h(x,z) := \max_k h_k(x,z)$:

$$\begin{aligned} \underset{x\in\mathcal{X},q,Q,r}{\text{minimize}} \quad & r+\mu^T q+\Sigma\bullet Q \\ \text{subject to} \quad & r\geq \sup_{(z_1;Z_2):z_1\in\mathcal{Z},\, \mathcal{Z}_2\succeq z_1 z_1^T} h_k(x,z_1)-z_1^T q+-Z_2\bullet Q\,,\ \forall\, k\,. \end{aligned}$$

Looking more closely at the $\sup_{Z_2\succeq z_1 z_1^T} -Z_2\bullet Q$ part of the constraint, we realize that if Q is not positive semi-definite, then the supremum can reach ∞ since Z_2 is unbounded above. We must therefore have that $Q\succeq 0$, for the same price we also get that the optimum is always achieved at $Z_2 = z_1 z_1^T$. The DRO therefore reduces to

$$\underset{x\in\mathcal{X},q,Q,r}{\text{minimize}} \quad r+\mu^T q+\Sigma\bullet Q \tag{8.13a}$$

$$\text{subject to} \quad r\geq \sup_{z_1\in\mathcal{Z}} h_k(x,z_1)-z_1^T q - z_1^T Q z_1\,,\ \forall\, k \tag{8.13b}$$

$$Q\succeq 0\,. \tag{8.13c}$$

Each constraint in this latter reformulation is a non-linear robust constraint with the right properties to be tackled by the theory presented in Chapter 6 when each $h_k(x,z)$ is convex in x and concave in z.

8.1.3 Accounting for moment uncertainty

When moments of distribution are estimated based on historical data, it is common to consider that the moments are not precisely known but are rather assumed to lie in some confidence region $\mathcal{U}$. In this context the distributionally robust model should consider as candidate worst-case distribution any distribution which mean lies in $\mathcal{U}$, thus giving rise to the following DRO model:

$$\underset{x\in\mathcal{X}}{\text{minimize}} \quad \sup_{\mu\in\mathcal{U},F\in\mathcal{D}(\mathcal{Z},\mu)} \mathbb{E}_F[h(x,z)]\,. \tag{8.14a}$$

Corollary 8.9. : *Let $\mathcal{D}(\mathcal{Z},\mu)$ be a distribution set and $\mathcal{U}\in\mathbb{R}^m$ be a bounded and convex uncertainty set for the moment vector μ. Given that for all $\mu\in\mathcal{U}$, there exists an $F\in\mathcal{D}(\mathcal{Z},\mu)$, the DRO problem presented in* (8.14) *is equivalent to the following robust optimization problem:*

$$\underset{x\in\mathcal{X},q}{\text{minimize}} \quad \sup_{z\in\mathcal{Z}} h(x,z)-z^T q+\delta^*(q\,|\,\mathcal{U})\,. \tag{8.15a}$$

Moreover, the problem can be reformulated as follows when $\mathcal{Z}$ is a convex set and $h(x,z) := \max_k h_k(x,z)$ where each $h_k(x,z)$ is a concave function:

$$\begin{aligned} \underset{x\in\mathcal{X},q,\{v_k\}_k,t}{\text{minimize}} \quad & t + \delta^*(q\,|\,\mathcal{U}) \\ \text{subject to} \quad & t \geq \delta^*(v_k\,|\,\mathcal{Z}) - h_*^k(x, v_k + q)\,,\, \forall\, k\,, \end{aligned}$$

where for each k, $v_k \in \mathbb{R}^m$, while $\delta^*(v|\mathcal{Z})$ is the support function of $\mathcal{Z}$ and $h_*^k(x,v)$ is the partial concave conjugate function of $h_k(x,z)$.

Proof. This result follows almost directly from theorems 8.1 and 8.6. From the former, we can establish that

$$\begin{aligned} \sup_{\mu\in\mathcal{U},F\in\mathcal{D}(\mathcal{Z},\mu)} \mathbb{E}_F[h(x,z)] &= \max_{\mu\in\mathcal{U}} \inf_q \sup_{z\in\mathcal{Z}} h(x,z) + (\mu - z)^T q \\ &= \max_{\mu\in\mathcal{U}} \inf_q \mu^T q + \sup_{z\in\mathcal{Z}} h(x,z) - z^T q \\ &= \inf_q \max_{\mu\in\mathcal{U}} \mu^T q + \sup_{z\in\mathcal{Z}} h(x,z) - z^T q \\ &= \inf_q \sup_{z\in\mathcal{Z}} \delta^*(q\,|\,\mathcal{U}) + h(x,z) - z^T q\,, \end{aligned}$$

where we employed Sion's minimax theorem which exploits the fact that $\mathcal{U}$ is bounded and convex, and that $\mu^T q + \sup_{z\in\mathcal{Z}} h(x,z) - z^T q$ is affine in μ and convex in q. Note that the introduction of the support function simply follows from its definition. The second reformulation follows from Theorem 8.6 which can exploit the fact that $h(x,z)$ would be piecewise concave in z. □

Example 8.10. : In [28], the authors explain how independently and identically distributed samples $\{\xi_i\}_{i=1}^M$ from F can be used to construct the following uncertainty set:

$$\mathcal{D}(\mathcal{Z},\hat{\mu},\hat{\Sigma},\gamma_1,\gamma_2) = \left\{ F \in \mathcal{M} \;\middle|\; \begin{array}{l} \mathbb{P}_F(\xi \in \mathcal{Z}) = 1 \\ (\mathbb{E}_F[\xi] - \hat{\mu})^T \hat{\Sigma}^{-1} (\mathbb{E}_F[\xi] - \hat{\mu}) \leq \gamma_1 \\ \mathbb{E}_F[(\xi - \hat{\mu})(\xi - \hat{\mu})^T] \preceq (1+\gamma_2)\hat{\Sigma} \end{array} \right\} \,,$$

where the second constraint imposes that the mean of ξ be located inside some confidence region described as an ellipsoid centered at the empirical estimate of the mean, $\hat{\mu}$, and with a shape that is defined through the empirical covariance matrix, $\hat{\Sigma}$. The third constraint is a bit more complicated to parse since it imposes a linear matrix inequality (in the form of an upper bound) on the centered second order moment matrix of ξ. Note that this is not exactly equivalent to imposing an LMI upper bound on the covariance matrix of ξ unless $\gamma_1 = 0$ (since then $\mathbb{E}_F[\xi] = \hat{\mu}$). Nevertheless, it allows one to control how far the realization might be from $\hat{\mu}$ on average.

One can recognize in this context that the DRO problem becomes:

$$\underset{x\in\mathcal{X}}{\text{minimize}} \;\; \max_{(\mu,\Sigma)\in\mathcal{U}} \;\; \sup_{F\in\mathcal{D}(\mathcal{Z},(\mu,\Sigma))} \mathbb{E}_F[h(x,\zeta)]\,,$$

where

$$\mathcal{U} := \left\{ (\mu, \Sigma) \in \mathbb{R}^m \times \mathcal{S}^{m\times m} \;\middle|\; \begin{array}{c} (\mu - \hat{\mu})^T \hat{\Sigma}^{-1} (\mu - \hat{\mu}) \leq \gamma_1 \\ \Sigma \preceq (1+\gamma_2)\hat{\Sigma} + \hat{\mu}\mu^T + \mu\hat{\mu}^T - \hat{\mu}\hat{\mu}^T \end{array} \right\}.$$

Based on these considerations, one can exploit corollary 8.9 and the results of Example 8.8 to claim that the DRO problem is equivalent to

$$\underset{x\in\mathcal{X},q}{\text{minimize}} \quad \sup_{z\in\mathcal{Z}'} h(x, z_1) - z_1^T q - Z_2 \bullet Q + \delta^*((q, Q) \,|\, \mathcal{U})$$

where $\mathcal{Z}'$ follows the definition in equation (8.12). Given that we already performed the analysis for $\sup_{z\in\mathcal{Z}'} h(x, z_1) - z_1^T q - Z_2 \bullet Q$, we are left with characterizing $\delta^*((q, Q) \,|\, \mathcal{U})$. To do so, we will describe the $\mathcal{U}$ set as the intersection of two sets:

$$\mathcal{U}_1 := \{(\mu, \Sigma) \in \mathbb{R}^m \times \mathcal{S}^{m\times m} \,|\, (\mu - \hat{\mu})^T \hat{\Sigma}^{-1} (\mu - \hat{\mu}) \leq \gamma_1\},$$

and

$$\mathcal{U}_2 := \{(\mu, \Sigma) \in \mathbb{R}^m \times \mathcal{S}^{m\times m} \,|\, \Sigma \preceq (1+\gamma_2)\hat{\Sigma} + \hat{\mu}\mu^T + \mu\hat{\mu}^T - \hat{\mu}\hat{\mu}^T\}.$$

In the first case, we have that $\mu := \hat{\mu} + \hat{\Sigma}^{1/2} w$ where $\|w\|_2 \leq \sqrt{\gamma_1}$. Based on Table 6.1 and theorem 6.7, we can use this fact to conclude that

$$\delta^*(q, Q \,|\, \mathcal{U}_1) := \hat{\mu}^T q + \sqrt{\gamma_1} \|\hat{\Sigma}^{1/2} q\|_2 + \mathbf{1}\{Q = 0\},$$

where $\mathbf{1}\{Q = 0\}$ is the indicator function that returns 0 if satisfied and ∞ otherwise.

In the second case, i.e. $\delta^*(q, Q \,|\, \mathcal{U}_2)$, we need to reformulate the following expression:

$$\begin{array}{ll} \underset{z_1\in\mathbb{R}^m, Z_2\in\mathcal{S}^{m\times m}}{\max} & Q \bullet Z_2 + q^T z_1 \\ \text{subject to} & Z_2 \preceq (1+\gamma_2)\hat{\Sigma} + \hat{\mu} z_1^T + z_1 \hat{\mu}^T - \hat{\mu}\hat{\mu}^T. \end{array}$$

After replacing $Z_2' := Z_2 - \hat{\mu} z_1^T - z_1 \hat{\mu}^T$, we get

$$\begin{array}{ll} \underset{z_1\in\mathbb{R}^m, Z_2'\in\mathcal{S}^{m\times m}}{\max} & Q \bullet Z_2' + (q + 2Q\hat{\mu})^T z_1 \\ \text{subject to} & Z_2' \preceq (1+\gamma_2)\hat{\Sigma} - \hat{\mu}\hat{\mu}^T, \end{array}$$

which can be optimized separately in z_1 and Z_2'. In the former case, the optimal value is unbounded unless $q = -2Q\hat{\mu}$ for which $(q + 2Q\hat{\mu})^T z_1 = 0$. In the latter case, the problem becomes unbounded if $Q \not\succeq 0$ since then their is a direction that can be used to create an eigenvector of Z_2' for which the eigenvalue is unbounded on the negative side yet leads to an arbitrary increase of the objective function. If $Q \succeq 0$, then the maximum is reached at $Z_2 = (1+\gamma_2)\hat{\Sigma} - \hat{\mu}\hat{\mu}^T$. Overall, we get

$$\delta^*(q, Q \,|\, \mathcal{U}_2) := ((1+\gamma_2)\hat{\Sigma} - \hat{\mu}\hat{\mu}^T) \bullet Q + \mathbf{1}\{q = -2\hat{\mu}^T Q\} + \mathbf{1}\{Q \succeq 0\}.$$

Taking these two results together and the rule for composing $\delta^*(z \mid \mathcal{Z}_1 \cap \mathcal{Z}_2)$, we obtain:

$$\begin{aligned}
\delta^*(q_1, Q_2) \;:=\; \min_{v_1, v_2 \in \mathbb{R}^m, V_1, V_2 \in \mathcal{S}^{m \times m}} \quad & ((1+\gamma_2)\hat{\Sigma} - \hat{\mu}\hat{\mu}^T) \bullet V_2 + \hat{\mu}^T v_1 + \sqrt{\gamma_1}\|\hat{\Sigma}^{1/2} v_1\|_2 \\
\text{subject to} \quad & V_1 = 0 \\
& v_2 = -2V_2\hat{\mu} \\
& V_2 \succeq 0 \\
& q_1 = v_1 + v_2 \;\;\&\;\; Q_2 = V_1 + V_2 \,.
\end{aligned}$$

Assembling this reduction with the model presented in (8.13), we get that the DRO model reduces to

$$\begin{aligned}
\underset{x \in \mathcal{X}, q, Q, r}{\text{minimize}} \quad & r + ((1+\gamma_2)\hat{\Sigma} + \hat{\mu}\hat{\mu}^T) \bullet Q + \hat{\mu}^T q + \sqrt{\gamma_1}\|\hat{\Sigma}^{1/2}(q + 2Q\hat{\mu})\|_2 \\
\text{subject to} \quad & r \geq \sup_{z_1 \in \mathcal{Z}} h_k(x, z_1) - z_1^T q - z_1^T Q z_1 \,, \forall\, k \\
& Q \succeq 0 \,.
\end{aligned}$$

8.2 Scenario-based models

An alternative to moment-based approaches consists in starting of with a set of scenarios $\mathcal{Z} := \{z^1, z^2, \ldots, z^K\}$ and to consider the following DRO problem

$$\underset{x \in \mathcal{X}}{\text{minimize}} \;\; \sup_{p \in \mathcal{U}} \sum_{k=1}^{K} p_k h(x, z^k) \,,$$

where $p \in \mathbb{R}^K$ is a vector describing the probability of obtaining each of the K scenarios for ξ while $\mathcal{U} \subseteq \{p \in \mathbb{R}^K \mid p \geq 0, \; \sum_{k=1}^K p_k = 1\}$ is the uncertainty set for the distribution. Such an uncertainty set can also be calibrated using historical data (in this case $\mathcal{Z}$ should contain all observed scenarios) but will never account for scenarios that were not listed in the finite set $\mathcal{Z}$. This can be a problem for the generalization power of the DRO model since in practice we usually don't expect historical scenarios to contain all possible realization of our parameters. The flip-side is that if $\mathcal{Z}$ does contain all possible samples, then most scenario-based methods will have the property that $\mathcal{U}(\{\xi^i\}_{i=1}^M)$, where each ξ^i is independently and identically distributed according to F, converges to $\{p^*\}$ where $p_k^* = \mathbb{P}_F(\xi = \xi^k)$ as M goes to infinity. This leads us to say that the data-driven scenario-based model is "consistent" since in the limit as more data is recovered, the data-driven solution will converge to the optimal solution of the stochastic program (8.1). A comprehensive summary of data-driven scenario based approach can be found in [5]. Note that while consistency is not a property of data-driven moment based models as the one presented in example 8.10, moment-based models have the advantage of considering a continuous space of possible realizations.

8.3 Wasserstein distance based models

In [36], the authors actually are the first to propose a general data-driven DRO formulation that can achieve the three most valued properties of such models:

1. Finite sample guarantee : The property that the optimal value of the DRO model is guaranteed with high probability to bound from above the expected cost when a finite number of i.i.d. realizations have been observed

2. Consistency : The property that the optimal solution will eventually converge to the optimal solution of the stochastic program (8.1) as more i.i.d. realizations are used to construct the distribution set $\mathcal{D}$.

3. Tractability : The DRO model can be solved using convex optimization algorithms for a large class of problems

We refer the reader to the full article for actual details. Interesting follow up results that involve applications such as classification and inverse optimization can be found in [42] and [37].

8.4 The power of randomized policy

Finally, it is worth mentioning an interesting recent result concerning DRO models that involve discrete (or integer) variables among the decision vector x. In [27], the authors have established that in such problems it is possible that in order to achieve the lowest possible worst-case expected cost, one is required to implement a decision that depends on the outcome of an independent randomization device. The two following example illustrate this counter-intuitive property of "mixed integer DRO" problems.

Example 8.11. :[The Ellsberg urn game] Consider an urn that contains an unknown number of red and blue balls in equal proportion. A player is asked to name one of the two colors and to draw a random ball from the urn. If the chosen ball has the stated color, the player incurs a penalty of \$1; otherwise, the player is rewarded \$1. One readily verifies that the player receives an expected reward of \$0 under either choice ("red" or "blue").

Assume now that the same game is played with an urn that contains red and blue balls, but neither the number of balls nor the proportions of their colors are known. Since the distribution of colors is completely ambiguous, the worst scenario is that no ball in the urn have the color named by the player, in which case the penalty of \$1 arises with certainty. Thus, a maximally ambiguity-averse player is indifferent between any pure strategy (i.e., naming red or blue) and paying a fixed amount of \$1. Assume instead that the player randomly names "red" (or "blue") with probability p (or $1 - p$), and that the (unknown) probability of the drawn ball being red (blue) is

q (or $1-q$). In that case, an ambiguity-averse player would be confronted with the following optimization problem.

$$\underset{p\in[0,1]}{\text{maximize}} \ \min_{q\in[0,1]} \ [pq+(1-p)(1-q)]\cdot(-\$1)+[p(1-q)+(1-p)q]\cdot\$1$$

The inner minimization problem has the parametric optimal solution $q^\star = 1$ if $p > 1/2$; $q^\star \in [0,1]$ if $p = 1/2$; $q^\star = 0$ if $p < 1/2$, which results in an objective value of $-2|p-1/2|$. Under the optimal choice $p^\star = 1/2$ (i.e., picking a color based on the throw of a fair coin), the player can completely suppress the ambiguity and receive the same expected reward of \$0 as in the first urn game where the proportions of colors are known. Hence to achieve the best worst-case expected profit, one must randomly choose between the color "red" and "blue".

Example 8.12. : A manager must implement one out of five candidate projects. The projects' net present values (NPVs) follow independent Bernoulli distributions that assign unknown probabilities p_i^L and p_i^H to the low and high NPV scenarios from Table 8.1. The scenarios are chosen such that projects with higher indices display a higher expected NPV as well as a higher dispersion, as indicated by the increasing spread between the associated two NPV scenarios.

Table 8.1: Low and high NPV scenarios (in \$1,000,000), as well as means of the five projects under nominal and ambiguous probabilities.

	Low NPV	High NPV	Mean	Mean if $p_i^L = 0.5$
Project 1	-0.6141	0.8500	$p_1^L\cdot(-0.6141)+p_1^H\cdot 0.8500$	0.1179
Project 2	-0.5471	1.9250	$p_2^L\cdot(-0.5471)+p_2^H\cdot 1.9250$	0.6889
Project 3	-0.3415	2.9500	$p_3^L\cdot(-0.3415)+p_3^H\cdot 2.9500$	1.3042
Project 4	-0.0750	3.9250	$p_4^L\cdot(-0.0750)+p_4^H\cdot 3.9250$	1.9250
Project 5	0.2168	4.8500	$p_5^L\cdot 0.2168+p_5^H\cdot 4.8500$	2.5334

Let's first assume that the probabilities are known to be such that $p_i^L = p_i^H = 0.5$. If the manager aims to maximize the expected NPV, then she selects project 5. The manager cannot improve upon this pure (i.e., deterministic) choice by making a random choice as long as the NPVs are independent of the randomization device.

$$\mathbb{E}[\sum_{i=1}^{5}\xi_i X_i] = \sum_{i=1}^{5}\mathbb{E}[\xi_i X_i] = \sum_{i=1}^{5}\mathbb{E}[\xi_i]\mathbb{E}[X_i] \le \max_{i=1,\dots,5}\mathbb{E}[\xi_i]\ ,$$

where ξ_i is the Bernoulli random variable capturing the NPV of project i, and X_i is the Bernoulli random variable representing the event that the random project selection ended up choosing project i.

Let's assume now that the probabilities p_i^L and p_i^H corresponding to the low and high NPV scenarios of the i-th project, respectively, are ambiguous, and that they are only known to satisfy

$$p_i^L = \frac{1}{2} + z_i \min\left\{0.3\mu_i,\ \frac{1}{2}\right\} \quad \text{and} \quad p_i^H = \frac{1}{2} - z_i \min\left\{0.3\mu_i,\ \frac{1}{2}\right\} \tag{8.16}$$

for some $z_i \in [-1,\, 1]$ with $\sum_{i=1}^5 |z_i| \leq 1$, where μ_i denotes project i's expected NPV under the nominal Bernoulli distribution from Table 8.1. Thus, possible deviations of the probabilities (p_i^L, p_i^H) from their nominal values $(1/2, 1/2)$ are proportional to the project's expected NPV under the nominal distribution. This gives rise to the following uncertainty set:

$$\mathcal{U} := \{(p^L, p^H) \in \mathbb{R}^5 \times \mathbb{R}^5 \,|\, \exists\, z \in [-1,\, 1]^5,\, p_i^L = 0.5 + \min(0.5, 0.3\mu_i) z_i,\, p^L + p^H = 1,\, \sum_{i=1}^5 |z_i| \leq 1\}\,.$$

Assume further that the manager maximizes the worst-case expected NPV over all distributions satisfying (8.16). In this case, the decision model takes the form of a DRO

$$\max_{x \in \{0,\,1\}^5 : \sum_{i=1}^5 x_i \leq 1} \quad \min_{F \in \mathcal{D}} \;\; \mathbb{E}_F[\sum_{i=1}^5 \xi_i x_i]\;.$$

One can find the solution of this DRO analytically by considering that if project j is chosen, then the worst-case analysis reduces to

$$\min_{F \in \mathcal{D}} \;\; \mathbb{E}_F[\xi_j] = (0.5 + \min(0.3\mu_j, 0.5))\xi_j^L + (0.5 - \min(0.3\mu_j, 0.5))\xi_j^H\;.$$

Hence, the worst-case expected value is evaluated to 0.066, 0.178, 0.016, -0.075, 0.2168 for projects 1 to 5 respectively. So that we are better choose project 5 with a worst-case expected value of 0.2168. Alternatively, we could try to identify a randomized strategy:

$$\max_{q \in [0,\,1]^5 : \sum_{i=1}^5 q_i = 1} \quad \min_{F \in \mathcal{D}} \;\; \mathbb{E}_{\xi \sim F, X \sim q}[\sum_{i=1}^5 \xi_i X_i]$$

Yet, we have that $\mathbb{E}_{\xi \sim F, X \sim q}[\sum_{i=1}^5 \xi_i X_i] = \sum_{i=1}^5 \mathbb{E}_{\xi \sim F}[\xi_i]\mathbb{E}_{x \sim q}[X_i] = \sum_i \mathbb{E}_F[\xi_i] q_i$. Hence the Randomized DRO reduces to

$$\max_{q \in [0,\,1]^5 : \sum_{i=1}^5 q_i = 1} \quad \min_{(p^L, p^H) \in \mathcal{U}} \;\; \sum_{i=1}^5 (p_i^L \cdot \xi_i^L + p_i^H \cdot \xi_i^H) q_i$$

The optimal randomized strategy chooses projects 3, 4, and 5 with probabilities 45.46%, 29.27%, and 25.27% respectively, yielding a worst-case expected value of 1.2111, which is more than five time larger than what any deterministic policy can achieve.

For implementation details, we refer the reader to the following Google Colab file.

8.5 Exercises

Consider the following DRO problem:

$$\underset{x \in \mathcal{X}}{\text{minimize}} \quad \sup_{F \in \mathcal{D}_1} \mathbb{E}_F[\max(-\frac{1}{2}\xi^T Q(x)\xi,\; x^T C \xi)]\,, \tag{8.17}$$

where $\mathcal{X} := \{x \in \mathbb{R}_+^n \,|\, Ax \leq b\}$ for some $A \in \mathbb{R}^{p \times n}$ and $b \in \mathbb{R}^p$, $Q(x) := \sum_{i=1}^n Q_i x_i$ with each $Q_i \in \mathbb{R}^{m \times m}$ such that $Q_i \succ 0$, and $C \in \mathbb{R}^{n \times m}$ such that each $C_{ij} \geq 0$.

Exercise 8.1. ***Mean-support DRO problem***
Derive an explicit finite dimensional representation for the DRO problem (8.17) when

$$\mathcal{D}_1 := \{F \,|\, \mathbb{P}_F(\xi \in \mathcal{Z}) \geq 1,\ \mathbb{E}_F[\xi] = \bar{\mu}\}\,,$$

where $\mathcal{Z} := \{z \in \mathbb{R}^m \,|\, Wz \leq v\}$, with $W \in \mathbb{R}^{p\times m}$, $v \in \mathbb{R}^p$, and $\bar{\mu} \in \mathbb{R}^n$.

Exercise 8.2. ***DRO with moment uncertainty***
Derive an explicit finite dimensional representation for problem (8.17) when the distribution ambiguity set takes the form:

$$\mathcal{D}_2(\Gamma) := \{F \,|\, \mathbb{P}_F(\xi \in \mathcal{Z}) = 1,\ \mathbb{E}_F[\xi] \geq \bar{\mu},\ \sum_i \mathbb{E}_F[\xi_i] - \bar{\mu}_i \leq \Gamma\}\,.$$

Exercise 8.3. ***Globalized DRO problem***
Derive an explicit finite dimensional representation for the following globalized distributionally robust optimization problem:

$$\begin{array}{ll} \underset{x\in\mathcal{X},t}{\text{minimize}} & t \\ \text{subject to} & \mathbb{E}_F[\max(-\frac{1}{2}\xi^T Q(x)\xi,\ x^T C\xi)] \leq \alpha\Gamma + t\,,\ \forall\, F \in \mathcal{D}_2(\Gamma)\,,\ \forall\, \Gamma \in [0, \bar{\Gamma}]\,, \end{array}$$

where $\alpha > 0$.

Chapter 9

Pareto Efficiency in Robust Optimization

In [32], the authors have exposed an interesting characteristic of optimal solutions to robust optimization problems.

Example 9.1. : Consider the following robust optimization model:

$$\underset{x\in[0,\,1]}{\text{maximize}} \quad \min_{z\in[0,1]} zx\,.$$

This robust optimization problem does not have a unique optimal solution. Indeed, the whole interval $x \in [0,\,1]$ achieves the same worst-case value of 0. When implementing this robust optimization model in RSOME (version rsome-0.0.6 with mosek-9.2.32, see Google Colab), one obtains that $x^* = 0$ which is indeed true but leaves one wondering why consider this value over any other value in [0, 1]. In fact, one can demonstrate that $x^{**} = 1$ performs as well as x^* in the worst-case but achieves a strictly better performance than x^* in terms of xz for any $z \in]0,\,1]$.

This example illustrates how robust optimization problems might have non-unique optimal solutions and raises the question of how to adequately select the solution that will be implemented. Note that when solving robust optimization models with a software package such as CPLEX, the choice of which solution among optimal ones is returned is completely arbitrary (hence out of our control).

The example also illustrates that while robust optimization provides some guarantees regarding the worst-case scenario, it does not provide any guarantees regarding how the solution behaves compared to others under non-worst-case scenarios. Actually, in [32], the authors argue that in the set of optimal robust solutions there are solutions that are clearly preferred to others (e.g. $x^{**} = 1$ is clearly preferred to $x^* = 0$ in example 9.1) since for those solutions there exists other optimal robust solutions that "strictly dominates" them, i.e. some other solution performs at least as well as them under all scenarios and strictly outperforms them under some scenarios.

Let's now make this discussion more formal. Consider the following robust optimization problem:

$$\underset{x\in\mathcal{X}}{\text{maximize}} \quad \min_{z\in\mathcal{Z}} \; h(x,z)\,.$$

Based on the above optimization model one can identify a set of optimal robust solutions as defined below.

Definition 9.2. : The set of all optimal robust solutions is defined as

$$\mathcal{X}_{\text{RO}} := \{x \in \mathcal{X} \mid \min_{z\in\mathcal{Z}} h(x,z) \geq \max_{x\in\mathcal{X}} \min_{z\in\mathcal{Z}} h(x,z)\}\,.$$

The question is : "Which of the solutions $x \in \mathcal{X}_{\text{RO}}$ should be used?"

The most common way of answering this question in the optimization community has been to employ a secondary objective $g(x)$ to discriminate among the different optimal solutions. For example, one natural choice of $g(x)$ might simply be $h(x,\hat{z})$ with $\hat{z}$ as the nominal scenario. Generally speaking, we would be looking for the optimal solution of:

$$\underset{x\in\mathcal{X}_{\text{RO}}}{\text{maximize}} \quad g(x)\,,$$

which can be characterized as

$$\begin{array}{ll} \underset{x\in\mathcal{X}}{\text{maximize}} & g(x) \\ \text{subject to} & h(x,z) \geq \gamma\,,\ \forall\, z \in \mathcal{Z}\,, \end{array}$$

where $\gamma := \max_{x\in\mathcal{X}} \ \min_{z\in\mathcal{Z}} \ \ h(x,z)$ is the best worst-case value achieved in the robust optimization problem. It is worth emphasizing that this problem is a convex optimization problem when $\mathcal{X}$ is convex, while $g(x)$ and $h(x,z)$ are concave functions in x. Furthermore, a Fenchel robust counterpart of the robust constraint can be obtained if $\mathcal{Z}$ is convex and $h(x,z)$ is convex in z.

In [32], the authors instead argue that some members of $\mathcal{X}_{\text{RO}}$ should be discarded before having to use a new criterion of selection. Indeed, remember that in example 9.1, it was clear to us that $x^{**} = 1$ was a better solution than $x^* = 0$. We next present the definitions of the type of solutions that are left when we discard from $\mathcal{X}_{\text{RO}}$ all solutions that have a similar property to $x^* = 0$.

Definition 9.3. : We will say that a solution x is Pareto robustly optimal (PRO) if it is robust optimal, i.e. $x \in \mathcal{X}_{\text{RO}}$, and if there exists no other $\bar{x} \in \mathcal{X}$ that Pareto dominates x, i.e. there exists no $\bar{x} \in \mathcal{X}$ for which the following two conditions hold:

1. $h(\bar{x},z) \geq h(x,z)$ for all $z \in \mathcal{Z}$
2. $h(\bar{x},z) > h(x,z)$ for some $z \in \mathcal{Z}$.

Moreover, one can define the set of all Pareto robustly optimal solutions as

$$\mathcal{X}_{\text{PRO}} := \{x \in \mathcal{X}_{\text{RO}} \,|\, \forall\, x' \in \{x' \in \mathcal{X} \,|\, h(x', z) \geq h(x, z)\,,\, \forall\, z \in \mathcal{Z}\},\, h(x', z) \leq h(x, z)\,,\, \forall\, z \in \mathcal{Z}\}\,.$$

Based on definition 9.3, it is clear that testing whether $x \in \mathcal{X}_{\text{PRO}}$ or not is equivalent to verifying whether $\hat{x} \in \mathcal{X}_{\text{RO}}$ and whether the optimal value of the following optimization model is lower or equal to zero:

$$\begin{aligned} \underset{x' \in \mathcal{X}, z \in \mathcal{Z}}{\text{maximize}} \quad & h(x', z) - h(x, z) && (9.1\text{a}) \\ \text{subject to} \quad & h(x', z') \geq h(x, z')\,,\, \forall\, z' \in \mathcal{Z}\,. && (9.1\text{b}) \end{aligned}$$

To simplify the analysis of this problem (see remark 9.12 for some direction of generalization), the authors of [32] assumed that the function $h(x, z)$ was affine in both x and z which we also assume for the rest of the chapter.

Assumption 9.4. : We assume that $\mathcal{X}$ and $\mathcal{Z}$ are convex sets, that $\mathcal{Z}$ contains a scenario $\bar{z}$ in its relative interior, and that $h(x, z)$ is affine in both x and z. In other words,

$$h(x, z) := c(z)^T x + d(z) \qquad \& \qquad h(x, z) := c'(x)^T z + d'(x)\,.$$

Perhaps, the most important result on this topic is an answer regarding the identification of a simple method to check whether an optimal robust solution is also Pareto robustly optimal.

Theorem 9.5. : *If assumption 9.4 is satisfied and $x \in \mathcal{X}_{RO}$, then, verifying whether $x \in \mathcal{X}_{PRO}$ is equivalent to verifying whether the optimal value of the following convex optimization problem is lower or equal to zero:*

$$\begin{aligned} \underset{\Delta}{\text{maximize}} \quad & c(\bar{z})^T \Delta && (9.2\text{a}) \\ \text{subject to} \quad & x + \Delta \in \mathcal{X} && (9.2\text{b}) \\ & c(z)^T \Delta \geq 0\,,\, \forall\, z \in \mathcal{Z}\,. && (9.2\text{c}) \end{aligned}$$

Proof. First, when $h(x, z) := c(z)^T x + d(z)$, problem (9.1) takes the form

$$\begin{aligned} \underset{x' \in \mathcal{X}, z \in \mathcal{Z}}{\text{maximize}} \quad & c(z)^T x' + d(z) - c(z)^T x - d(z) \\ \text{subject to} \quad & c(z')^T x' + d(z') \geq c(z')^T x + d(z')\,,\, \forall\, z' \in \mathcal{Z}\,, \end{aligned}$$

hence reduces to

$$\begin{aligned} \underset{x' \in \mathcal{X}, z \in \mathcal{Z}}{\text{maximize}} \quad & c(z)^T (x' - x) \\ \text{subject to} \quad & c(z')^T (x' - x) \geq 0\,,\, \forall\, z' \in \mathcal{Z}\,. \end{aligned}$$

One can then perform the variable replacement $\Delta := x' - x$ in order to get

$$\begin{array}{ll} \underset{\Delta, z\in\mathcal{Z}}{\text{maximize}} & c(z)^T\Delta \\ \text{subject to} & x+\Delta \in \mathcal{X} \\ & c(z')^T\Delta \geq 0\,,\ \forall\, z' \in \mathcal{Z}\,. \end{array}$$

We are left with the step of demonstrating equivalence between the two objective functions:

$$\max_{z\in\mathcal{Z}}\ c(z)^T\Delta \leq 0 \quad \Leftrightarrow \quad c(\bar{z})^T\Delta \leq 0\,.$$

Since $\bar{z} \in \mathcal{Z}$, it is first clear that

$$\max_{z\in\mathcal{Z}}\ c(z)^T\Delta \leq 0 \ \Rightarrow\ c(\bar{z})^T\Delta \leq \max_{z\in\mathcal{Z}}\ c(z)^T\Delta \leq 0\,.$$

The reverse direction needs a little more work. Let's assume that $c(\bar{z})^T\Delta \leq 0$ and take any $z \in \mathcal{Z}$. Since $\bar{z}$ is in the relative interior of the convex set $\mathcal{Z}$, as stated in remark 6.1 it must be that there exists an $\epsilon > 0$ such that $z' := \bar{z} - \epsilon((z-\bar{z})/\|z-\bar{z}\|_2) \in \mathcal{Z}$. In other words, there is a $z' \in \mathcal{Z}$ such that $\bar{z} := \theta z' + (1-\theta)z$ for some $\theta \in [0,\,1]$[1]. For this reason, we must have that

$$c(z)^T\Delta = c((1/(1-\theta))(\bar{z}-\theta z'))^T\Delta = (1/(1-\theta))(c(\bar{z})^T\Delta - \theta c(z')^T\Delta) \leq 0\,,$$

since to be feasible Δ must satisfy $c(z')^T\Delta \geq 0\,,\ \forall\, z' \in \mathcal{Z}$. Since such a conclusion can be drawn for any $z \in \mathcal{Z}$, we conclude that

$$c(\bar{z})^T\Delta \leq 0 \ \Rightarrow\ \sup_{z\in\mathcal{Z}}\ c(z)^T\Delta \leq 0\,.$$

□

The above theorem inspires us with both a method for identifying PRO solutions and a method for verifying whether $\mathcal{X}_{\text{RO}} = \mathcal{X}_{\text{PRO}}$.

Corollary 9.6. : *If assumption 9.4 is satisfied, let $\bar{z}$ be any scenario in the relative interior of $\mathcal{Z}$, then all optimal solutions to the problem* $\text{maximize}_{x\in\mathcal{X}_{RO}}\ c(\bar{z})^Tx + d(\bar{z})$ *are members of $\mathcal{X}_{PRO}$.*

It is worth mentioning that [32] also proves that all PRO solutions can be identified this way using different choices of vector $\bar{z}$ in the relative interior of $\mathcal{Z}$ (see proposition 1 in the article).

[1] Specifically, we would have that $\theta := \|z-\bar{z}\|_2/(\|z-\bar{z}\|_2+\epsilon)$

Proof. We can easily prove this by contradiction. Let x^* be an optimal solution of $\text{maximize}_{x\in\mathcal{X}_{\text{RO}}}\ c(\bar{z})^T x + d(\bar{z})$ such that $x^* \notin \mathcal{X}_{\text{PRO}}$. Based on Theorem 9.5 this implies that there exists a Δ^* for which the following conditions hold:

$$\begin{aligned}
&c(\bar{z})^T\Delta^* > 0\\
&x^* + \Delta^* \in \mathcal{X}\\
&c(z)^T\Delta^* \geq 0\,,\ \forall\, z \in \mathcal{Z}\,.
\end{aligned}$$

Yet, we can use Δ^* to create a solution x' that satisfies $x' \in \mathcal{X}_{\text{RO}}$, namely $x' := x^* + \Delta^*$. Indeed, $x' = x^* + \Delta^* \in \mathcal{X}$ by construction of Δ^*, and

$$\begin{aligned}
\min_{z\in\mathcal{Z}} c(z)^T x' + d(z) &= \min_{z\in\mathcal{Z}} c(z)^T(x^* + \Delta^*) + d(z) \geq \min_{z\in\mathcal{Z}} c(z)^T x^* + d(z) + \min_{z\in\mathcal{Z}} c(z)^T\Delta^*\\
&\geq \min_{z\in\mathcal{Z}} c(z)^T x^* + d(z) = \max_{x\in\mathcal{X}}\ \min_{z\in\mathcal{Z}} c(z)^T x + d(z)\,,
\end{aligned}$$

where the last equality comes from the fact that $x^* \in \mathcal{X}_{\text{RO}}$. The contradiction follows from the fact that since $c(\bar{z})^T\Delta^* > 0$, it must be that

$$c(\bar{z})^T x' + d(\bar{z}) = c(\bar{z})^T(x^* + \Delta^*) + d(\bar{z}) > c(\bar{z})^T x^* + d(\bar{z})\,.$$

This is impossible since x^* was supposed to be an optimal solution of $\text{maximize}_{x\in\mathcal{X}_{\text{RO}}}\ c(\bar{z})^T x + d(\bar{z})$. ☐

Corollary 9.7. : *If assumption 9.4 is satisfied, let $\bar{z}$ be any scenario in the relative interior of $\mathcal{Z}$, then $\mathcal{X}_{RO} = \mathcal{X}_{PRO}$ if and only if the optimal value of the following problem is zero:*

$$\begin{aligned}
\underset{x\in\mathcal{X}_{RO},\Delta}{\text{maximize}} \quad & c(\bar{z})^T\Delta && (9.3a)\\
\text{subject to} \quad & x + \Delta \in \mathcal{X} && (9.3b)\\
& c(z)^T\Delta \geq 0\,,\ \forall\, z \in \mathcal{Z}\,. && (9.3c)
\end{aligned}$$

Proof. We demonstrate that the first statement is false if and only if the second is false. First, say that $\mathcal{X}_{\text{RO}} \neq \mathcal{X}_{\text{PRO}}$, this implies that there is an $x \in \mathcal{X}_{\text{RO}}$ that is not PRO, however, theorem 9.5 states that there must exists a Δ such that $c(\bar{z})^T\Delta > 0$ in problem (9.2) which implies that the optimum of problem (9.3) must be strictly positive. Alternatively, let's assume that problem (9.3) has a non-zero optimal value. This first implies that the optimal value is strictly positive since $\Delta = 0$ is always a feasible solution. This implies that there exists a pair (x', Δ') such that $x' \in \mathcal{X}_{\text{RO}}$ and when evaluated in problem (9.2) is feasible and reaches a strictly positive objective value. Based on the same theorem, this implies that $x' \notin \mathcal{X}_{\text{PRO}}$. This concludes this proof. ☐

Given these interesting conclusions, one might now wonder if it is possible to perform optimization of a secondary objective $g(x)$ over the set of PRO solutions. Indeed, Iancu and Trichakis confirm that this is possible in the case that the problem is a linear program but requires the use of additional binary variables.

Theorem 9.8. : *If assumption 9.4 is satisfied and* $\mathcal{X} := \{x \in \mathbb{R}^n \,|\, Ax \leq b\}$*, with* $A \in \mathbb{R}^{p\times n}$ *and* $b \in \mathbb{R}^p$*, while* $\mathcal{Z} := \{z \in \mathbb{R}^m \,|\, Wz \leq v\}$*, then given any secondary objective function* $g(x)$*, the optimization problem* $\text{maximize}_{x\in\mathcal{X}_{PRO}}\ g(x)$ *is equivalent to the following optimization problem:*

$$\begin{array}{ll} \underset{x,\mu,\eta,\gamma,y}{\text{maximize}} & g(x) \\ \text{subject to} & c(\bar{z}) - A^T\mu + c_0\eta + \displaystyle\sum_{i=1}^{m} c_i\gamma_i = 0 \\ & W\gamma \leq v\eta \\ & 0 \leq b - Ax \leq M(1-y) \\ & 0 \leq \mu \leq My \\ & \eta \geq 0 \ \ , \ \ y \in \{0,\ 1\}^p \\ & x \in \mathcal{X}_{RO}\,, \end{array}$$

where $\mu \in \mathbb{R}^m$*,* $\eta \in \mathbb{R}$ *and* $\gamma \in \mathbb{R}^m$*, and where we let* $c(z) := c_0 + \sum_{i=1}^m c_i z_i$*.*

Proof. Based on theorem 9.5, we know that verifying the feasibility of $\mathcal{X}_{\text{PRO}}$ is equivalent to checking that the optimal value of problem (9.2) is smaller or equal to zero. After describing the set $\mathcal{X}$ and $\mathcal{Z}$ explicitly and deriving the reformulation for the robust constraint $c(z)^T\Delta \geq 0\,,\, \forall\, z \in \mathcal{Z}$, we get:

$$\begin{array}{ll} \underset{\Delta,\lambda}{\text{maximize}} & c(\bar{z})^T\Delta \\ \text{subject to} & A(x+\Delta) \leq b \\ & c_0^T\Delta \geq v^T\lambda \\ & W^T\lambda = -C^T\Delta \\ & \lambda \geq 0\,, \end{array}$$

where $C := [c_1\ c_2\ \ldots\ c_m]$.

By linear programming duality, we get that the optimal value is equal to the optimal value of the dual problem:

$$\begin{array}{ll} \underset{\mu,\eta,\gamma}{\text{minimize}} & (b - Ax)^T\mu \\ \text{subject to} & c(\bar{z}) - A^T\mu + c_0\eta + C\gamma = 0 \\ & W\gamma \leq v\eta \\ & \mu \geq 0,\ \eta \geq 0\,, \end{array}$$

where $\mu \in \mathbb{R}^p$ is the dual vector for the first constraint, while $\eta \in \mathbb{R}$ is the dual variable of the second constraint, and γ for the third. Hence, we have that the problem

$\text{maximize}_{x \in \mathcal{X}_{\text{PRO}}} \;\; g(x)$ is equivalent to

$$\begin{aligned} \underset{x,\mu,\eta,\gamma}{\text{maximize}} \quad & g(x) \\ \text{subject to} \quad & (b - Ax)^T \mu \leq 0 \\ & c(\bar{z}) - A^T \mu + c_0 \eta + C\gamma = 0 \\ & W\gamma \leq v\eta \\ & x \in \mathcal{X}_{\text{RO}} \\ & \mu \geq 0, \; \eta \geq 0 \,. \end{aligned}$$

Similarly as was done in chapter 4, we can linearize the bilinear constraint by exploiting the fact that since $Ax \leq b$ for all $x \in \mathcal{X}_{\text{RO}}$, and $\mu \geq 0$, the constraint is equivalent to

$$(b_i - a_i^T x)\mu_i = 0 \,, \; \forall \, i = 1, \ldots, p \,.$$

For each i, the resulting constraint can be linearized using a binary variable $y_i \in \{0, \; 1\}$.

$$\mu_i \leq My_i \quad \& \quad b_i - a_i^T x \leq M(1 - y_i) \,,$$

for some sufficiently large $M > 0$. □

It is worth emphasizing the fact that from an algorithmic perspective, robust optimization models benefit heavily from the fact that they are allowed to have optimal solutions that are not Pareto efficient. Indeed, in contrast with stochastic programming where typically all possible realizations of z must be used to confirm that a solution is optimal, robust optimization only requires a certificate of optimal performance for the worst-case scenario. In the following example, we illustrate this using a simple two-stage inventory problem.

Example 9.9. : Consider the following two-stage inventory problem:

$$\underset{x_1 \geq 0}{\text{minimize}} \; \max_{d \in [1, \, 3]} \; h_1(x_1) + \min_{x_2 \geq 0} cx_2 + b(d - x_1 - x_2)^+$$

where $x_1 \in \mathbb{R}$ is a first-stage production decision, $h_1(x_1)$ is some arbitrary first stage production cost function, d is the realized demand, x_2 is a recourse production decision that can adjust to the observed demand, and $cx_2 + b(d - x_1 - x_2)^+$ is the sum of second-stage production cost and backlog cost such that $c < b$. This problem can be represented as:

$$\underset{x_1 \geq 0, x_2 : [1, \, 3] \to \mathbb{R}_+}{\text{minimize}} \; \max_{d \in [1, \, 3]} \; h_1(x_1) + cx_2(d) + b(d - x_1 - x_2(d))^+ \,.$$

Let's assume that $x_1^* = 2$ is a minimizer of this problem. Since $c < b$ it is clear that if one minimizes the recourse problem:

$$\underset{x_2 \geq 0}{\text{minimize}} \, cx_2 + b(d - 2 - x_2)^+$$

one gets $x_2^*(d) := \max(0,\ d-2)$ which achieves a second stage cost of c in the worst-case (i.e. when $d = 3$). One can however realize that any policy in the set $\mathcal{X}_2^{**} := \{x_2 : [1,3] \to \mathbb{R}_+ \,|\, x_2^*(d) \leq x_2(d) \leq 1\,,\ \forall\, d \in [1,\ 3]\}$ also achieves this worst-case second stage cost since for all $x_2(\cdot) \in \mathcal{X}_2^{**}$ we have that

$$\max_{d\in[1,\ 3]}\ cx_2(d) + b(d-2-x_2(d))^+ = \max_{d\in[1,\ 3]}\ cx_2(d) + 0 \leq \max_{d\in[1,\ 3]}\ c\cdot 1 = c\,.$$

One example of a recourse policy in $\mathcal{X}_2^{**}$ takes the form of $x_2^{**}(d) = 1$ while another one might be $x_2^{***}(d) = (1/2)(d-1)$. We therefore can conclude that $\{(x_1^*, x_2^*(\cdot)),\ (x_1^*, x_2^{**}(\cdot)),\ (x_1^*, x_2^{***}(\cdot)),\ \} \in \mathcal{X}_{\mathrm{RO}}$ for this problem which is a good news since it means that there exist a static and affine decision rule that achieves optimality. On the other hand, it is clear that the static and affine decision rules are not Pareto efficiently robust solution, i.e. $\{(x_1^*, x_2^{**}(\cdot)),\ (x_1^*, x_2^{***}(\cdot)),\ \} \notin \mathcal{X}_{\mathrm{PRO}}$, since these solutions are strictly dominated by $(x_1^*, x_2^*(\cdot))$. Namely, in the case of the static solution $(x_1^*, x_2^{**}(\cdot))$, for all $d \in [1,\ 3]$ we can show that

$$h_1(x_1^*) + cx_2^*(d) + b(d - x_1^* - x_2^*(d))^+ \leq h_1(x_1^*) + c = h_1(x_1^*) + cx_2^{**}(d) + b(d - x_1^* - x_2^{**}(d))^+$$

while for all $d \in [1,\ 3[$:

$$h_1(x_1^*) + cx_2^*(d) + b(d - x_1^* - x_2^*(d))^+ = h_1(x_1^*) + cx_2^*(d) < h_1(x_1^*) + c = h_1(x_1^*) + cx_2^{**}(d) + b(d - x_1^* - x_2^{**}(d))^+\,.$$

As a final observation, one needs to be aware that, in multi-stage decision problems, while there are many situations in which static and affine policies are optimal robust solutions, these policies (actually optimal policies of any form) often prescribe recourse actions that are not optimal under the realized scenario for z (unless the realized z ends up being a worst-case one). This means that one should avoid implementing affine decision rules when future stages are reached, but instead should optimize an updated robust optimization model in a shrinking horizon fashion in order to identify an actual Pareto efficient decision for the current stage under the observed conditions of operation.

Remark 9.10. : The fact that robust optimal models might return solutions that are not Pareto robustly optimal is especially important given that in practice the performance of these solutions is often measured in terms of an expected value over a set of randomly generated scenarios for z. In this context, it is clear that if $x \notin \mathcal{X}_{\mathrm{PRO}}$, then the performance that will be measured will give the impression that a robust solution does not perform very well in a stochastic environment. Yet, if one instead makes the effort of recuperating an $x \in \mathcal{X}_{\mathrm{PRO}}$, this solution might achieve a much better performance in the context of randomly generated scenarios. As an example, we could think of evaluating the solution $x^* = 0$ from example 9.1 using a uniform distribution over $[0, 1]$. This would lead to an expected value $E[xz] = 0$ while the expected value of the PRO solution $x^{**} = 1$ achieves an expected value of 0.5 which is the best that can be achieved in this context.

Remark 9.11. : The distinction between $\mathcal{X}_{\text{RO}}$ and $\mathcal{X}_{\text{PRO}}$ will also raise some concerns when comparing the stochastic performance of robust solutions that are obtained from different conservative approximation schemes (say the AARC versus the LAARC models studied in chapter 5). Indeed, since the specific solution returned from each of these models is typically an arbitrary one among their respective $\mathcal{X}_{\text{RO}}$ sets, conclusions of the type "LAARC performs 50% better than AARC in terms of expected value on this problem instance" are somewhat misleading. Indeed, to be correct one should additionally explain how a unique solution was generated from each model. In case, where this fact remains unknown, only the worst-case performance over the uncertainty set $\mathcal{Z}$ can honestly be used to make a claim that an approximation method is better than some other.

Remark 9.12. : Finally, we wish to indicate some direction of generalization for theorem 9.5. Indeed, when $\mathcal{Z}$ is a convex set containing $\bar{z}$ in its relative interior, and when $h(x, z)$ is **affine** in z, we already can demonstrate the following reduction for problem 9.1.

$$\begin{array}{ll} \underset{x' \in \mathcal{X}}{\text{maximize}} & h(x', \bar{z}) - h(x, \bar{z}) \\ \text{subject to} & h(x', z') \geq h(x, z')\,,\ \forall\, z' \in \mathcal{Z}\,, \end{array}$$

which is a convex optimization problem if $h(x, z)$ is concave in x. The reduction is obtained using a similar argument as presented in the proof of theorem 9.5. Specifically, it relies on

$$\max_{z \in \mathcal{Z}}\ h(x', z) - h(x, z) \leq 0 \quad \Leftrightarrow \quad h(x', \bar{z}) - h(x, \bar{z}) \leq 0\,.$$

Yet, the steps to show this are exactly the same. First, we have:

$$\max_{z \in \mathcal{Z}}\ h(x', z) - h(x, z) \leq 0 \quad \Rightarrow \quad h(x', \bar{z}) - h(x, \bar{z}) \leq \max_{z \in \mathcal{Z}}\ h(x', z) - h(x, z) \leq 0\,.$$

While we also have

$$\begin{aligned} h(x', \bar{z}) - h(x, \bar{z}) \leq 0 \ \ &\Rightarrow\ \forall\, z \in \mathcal{Z}\,,\ \exists z' \in \mathcal{Z},\ ,\ \exists \theta \in]0, 1[\,,\ \bar{z} = \theta z' + (1 - \theta) z \\ \Rightarrow \ \ h(x', z) - h(x, z) &= h(x', (1/(1-\theta))(\bar{z} - \theta z')) - h(x, (1/(1-\theta))(\bar{z} - \theta z')) \\ &= (1/(1-\theta))(h(x', \bar{z}) - h(x, \bar{z}) - \theta(h(x', z') - h(x, z'))) \\ &\leq (1/(1-\theta))(h(x', \bar{z}) - h(x, \bar{z})) \leq 0\,. \end{aligned}$$

The reformulation that is obtained here leads us to believe that Pareto robustly optimal solution might also be found efficiently in scenario based distributionally robust optimization models, since $\sum_i p_i h(x, z^i)$ is affine in p.

Part IV

Supporting Material

Chapter 10

Appendix

10.1 Convex sets and convex functions

Definition 10.1. : A set $\mathcal{X} \subseteq \mathbb{R}^n$ is convex if for any two members $x_1 \in \mathcal{X}$ and $x_2 \in \mathcal{X}$, any "convex combination" of these two points is also a member of $\mathcal{X}$. Namely, for all $\theta \in [0, 1]$, we have that $\theta x_1 + (1 - \theta)x_2 \in \mathcal{X}$.

Definition 10.2. : A function $h : \mathcal{X} \to \mathbb{R}$, with $\mathcal{X} \subseteq \mathbb{R}^n$ as its domain, is said to be convex if its epigraph is a convex set. Namely, it is convex if and only if $\mathcal{X}$ is a convex set and that for any two members x_1 and x_2 of $\mathcal{X}$, and any convex combination $x_3 := \theta x_1 + (1 - \theta)x_2$, with $\theta \in [0, 1]$, of these two points, it is the case that $h(x_3) \leq \theta h(x_1) + (1 - \theta)h(x_2)$.

Definition 10.3. : A function $h : \mathcal{X} \to \mathbb{R}$, with $\mathcal{X} \subseteq \mathbb{R}^n$ as its domain, is said to be concave if the function $-h(x)$ is convex. Namely, it is concave if $\mathcal{X}$ is a convex set, and if for any two members x_1 and x_2 of $\mathcal{X}$, and any convex combination $x_3 := \theta x_1 + (1 - \theta)x_2$, with $\theta \in [0, 1]$, of these two points, it is the case that $h(x_3) \geq \theta h(x_1) + (1 - \theta)h(x_2)$.

Operations that preserve convexity of functions (see chapter 3.2 of [22] for more details):

- Addition of two convex function. Namely, if $g_1(x)$ and $g_2(x)$ are convex functions then $g_1(x) + g_2(x)$ is a convex function.

- Multiplying a convex function by a positive scalar. Namely, if $g(x)$ is convex then $\alpha g(x)$ is convex for any $\alpha \geq 0$.

- Taking the supremum over a set of convex functions. Namely, if $g(x, z)$ is convex for all $z \in \mathcal{Z}$ then $\sup_{z \in \mathcal{Z}} g(x, z)$ is convex

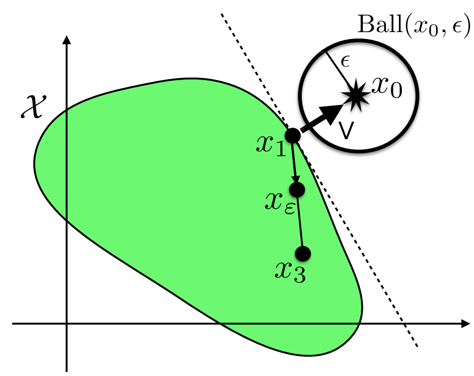
- Taking the infimum over a subset of variables for which the function is jointly convex. Namely, if $g(x, y)$ is jointly convex in x and y, and $\mathcal{X}$ is convex and non-empty, then $\inf_{x \in \mathcal{X}} g(x, y)$ is convex in y.

- Any composition of a convex function with an affine mapping. Namely, if $g : \mathbb{R}^n \to \mathbb{R}$ while $A \in \mathbb{R}^{n\times m}$ and $b \in \mathbb{R}^n$,then $g(Ax + b)$ is convex in x
- Some composition of convex and monotone functions. Namely, let $h : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ then $h(g(x))$ is convex in x if one of the conditions below apply:
  - $h(\cdot)$ is convex and nondecreasing and $g(\cdot)$ is convex
  - $h(\cdot)$ is convex and nonincreasing and $g(\cdot)$ is concave
- The perspective of a convex function. Namely, if $g(x)$ is convex, then $tg(x/t)$ is jointly convex in t and x as long as $t > 0$.

10.1.1 Strict separating hyperplane theorem

Theorem 10.4. :*(Strict separating hyperplane theorem) Let $\mathcal{X} \in \mathbb{R}^n$ be a closed convex set and $x_0 \notin \mathcal{X}$. Then there exists a hyperplane parametrized by $v \in \mathbb{R}^n$ and $b \in \mathbb{R}$ that strictly separates x_0 from $\mathcal{X}$. Namely,*

$$v^T x \le b, \forall x \in \mathcal{X} \qquad \& \qquad v^T x_0 > b$$



Proof. Since $\mathcal{X}$ is closed, it means that there exists some ball of radius $\epsilon > 0$ centered at x_0 which does not intersect with $\mathcal{X}$. Hence, when we try to identify the member of $\mathcal{X}$ that is closest to x_0, we will obtain a point x_1 such that $\|x_1 - x_0\|_2 \ge \epsilon$. To be clear, x_1 would be the solution of

$$x_1 := \operatorname*{argmin}_{x \in \mathcal{X}} \|x_1 - x_0\|_2 \, .$$

Let's construct the separating hyperplane $v := x_0 - x_1$ and $b := (x_0 - x_1)^T x_1$. For this hyperplane, we will show that both conditions are met.

First, again by contradiction, assume there exists some $x_3 \in \mathcal{X}$ such that $v^T x_3 > b$, in other words

$$(x_0 - x_1)^T x_3 > (x_0 - x_1)^T x_1 \;\Rightarrow\; (x_0 - x_1)^T (x_3 - x_1) \ge \varepsilon > 0 \, .$$

Well, then we will show that there is a point x_4 on the segment between x_1 and x_3, which is necessarily a member of $\mathcal{X}$ by convexity arguments, for which $\|x_4 - x_0\| < \|x_1 - x_0\|$ which contradicts the definition of x_1 as the closest point to x_0. Indeed, let's characterize the segment as any point generated with $x_1 + \theta(x_3 - x_1)$ for some $\theta \in [0, 1]$ and measure the squared distances to x_0 that can be achieved on this segment:

$$\begin{aligned}\|x_1 + \theta(x_3 - x_1) - x_0\|_2^2 &= \|(x_1 - x_0) + \theta(x_3 - x_1)\|_2^2 \\ &= \|x_1 - x_0\|_2^2 - 2\theta(x_0 - x_1)^T(x_3 - x_1) + \theta^2\|x_3 - x_1\|_2^2 \\ &\leq \|x_1 - x_0\|_2^2 - 2\theta\varepsilon + \theta^2\|x_3 - x_1\|_2^2 \\ &= \|x_1 - x_0\|_2^2 - \theta(2\varepsilon - \theta\|x_3 - x_1\|_2^2) < \|x_1 - x_0\|_2^2 ,\end{aligned}$$

where the last inequality is obtained by setting θ small enough in the interval $]0, \min(1,\ 2\varepsilon/\|x_3 - x_1\|_2^2)]$.

The second condition is straightforward :

$$\begin{aligned}v^T x_0 &= (x_0 - x_1)^T x_0 \geq (x_0 - x_1)^T x_0 - \|x_1 - x_0\|_2^2 + \epsilon^2 \\ &= (x_0 - x_1)^T x_0 - (x_1 - x_0)^T x_1 + (x_1 - x_0)^T x_0 + \epsilon^2 \\ &= (x_0 - x_1)^T x_0 + (x_0 - x_1)^T x_1 - (x_0 - x_1)^T x_0 + \epsilon^2 = (x_0 - x_1)^T x_1 + \epsilon^2 > b .\end{aligned}$$

□

Definition 10.5. : The subgradient of a convex function $g : \mathbb{R}^n \to \mathbb{R}$ at a point $x_0 \in \mathbb{R}^n$ is a vector $v \in \mathbb{R}^n$ such that

$$g(x) \geq g(x_0) + v^T(x - x_0) \,, \forall\, x \in \mathbf{dom}\, g \,.$$

The strict separating hyperplane theorem ensures that it always exists for convex function. In the special case of a differentiable function, it reduces to the gradient of the function.

10.2 Minimax theorems

Here is the most general yet weak version of a min max theorem.

Lemma 10.6. : *Let $\mathcal{X} \subseteq \mathbb{R}^n$ and $\mathcal{Z} \subseteq \mathbb{R}^m$, and given any function $g : \mathcal{X} \times \mathcal{Z} \to \mathbb{R}$, one has that*

$$\sup_{z \in \mathcal{Z}} \inf_{x \in \mathcal{X}} g(x, z) \leq \inf_{x \in \mathcal{X}} \sup_{z \in \mathcal{Z}} g(x, z) \,.$$

Proof.

$$\begin{aligned}&\inf_{x \in \mathcal{X}} g(x, z) \leq g(x', z) \,, \forall\, x' \in \mathcal{X},\ \forall\, z \in \mathcal{Z} \\ &\qquad \Rightarrow \sup_{z \in \mathcal{Z}} \inf_{x \in \mathcal{X}} g(x, z) \leq \sup_{z \in \mathcal{Z}} g(x', z) \,, \forall\, x' \in \mathcal{X} \\ &\qquad \Rightarrow \sup_{z \in \mathcal{Z}} \inf_{x \in \mathcal{X}} g(x, z) \leq \inf_{x' \in \mathcal{X}} \sup_{z \in \mathcal{Z}} g(x', z) \,.\end{aligned}$$

□

We follow with the most general version of a strong minimax theorem.

Lemma 10.7. : *(Sion's minimax theorem [44]) Let $\mathcal{X} \subset \mathbb{R}^n$ be a convex set and $\mathcal{Z} \in \mathbb{R}^m$ be a compact convex set, and let h be a real-valued function on $\mathcal{X} \times \mathcal{Z}$ with*

1. *$h(x, \cdot)$ lower semicontinuous and quasi-convex on $\mathcal{Z}$, $\forall\, x \in \mathcal{X}$*

2. *$h(\cdot, z)$ upper semicontinuous and quasiconcave on $\mathcal{X}$, $\forall\, z \in \mathcal{Z}$*

then

$$\sup_{x \in \mathcal{X}} \min_{z \in \mathcal{Z}} \; h(x, z) = \min_{z \in \mathcal{Z}} \sup_{x \in \mathcal{X}} \; h(x, z) \; .$$

In particular, the conclusion is valid if instead of conditions 1 and 2, one can verify that $h(x, \cdot)$ is convex on $\mathcal{Z}$ for all $x \in \mathcal{X}$, and $h(\cdot, z)$ is concave on $\mathcal{X}$ for all $z \in \mathcal{Z}$.

10.3 Conic Programming

Definition 10.8. : A symmetric matrix $Q \in \mathbb{R}^{n \times n}$ is positive semidefinite, referred as $Q \succeq 0$ if it satisfies the following condition

$$\forall\, x \in \mathbb{R}^n,\; x^T Q x \geq 0 \; .$$

Alternatively, this definition can be verified by confirming that all its eigenvalues are non-negative.

Definition 10.9. : A symmetric matrix $Q \in \mathbb{R}^{n \times n}$ is positive definite, referred as $Q \succ 0$ if it satisfies the following condition

$$\forall\, x \in \mathbb{R}^n,\; x^T Q x > 0 \; .$$

Alternatively, this definition can be verified by confirming that all its eigenvalues are strictly positive.

Solutions to Exercises

Solution to Exercise 2.1: One might consider that the constraint in this expression is equivalent to

$$(\sum_i \theta_i \bar{z}_i)^T x \leq b - a^T x\,, \forall\, \theta \in \mathcal{U}\ ,$$

where $\mathcal{U} := \{\theta \in \mathbb{R}^K \,|\, \theta \geq 0,\, \sum_i \theta_i = 1\}$. This constraint is further equivalent to

$$(\mathbb{Z}\theta)^T x \leq b - a^T x\,, \forall\, \theta \in \mathcal{U}\ ,$$

where $\mathbb{Z} := [\ \bar{z}_1 \quad \ldots \quad \bar{z}_K\ ]$. this is a form where theorem 2.7 can be applied, considering that $J = 1$, $p_0 := c$, $P_1 := \mathbb{Z}$, $r_1 = -1$, and W and v are as follow:

$$W := \begin{bmatrix} \mathbf{1}_K^T \\ -\mathbf{1}_K^T \\ -I \end{bmatrix} \qquad\qquad v := \begin{bmatrix} 1 \\ -1 \\ \mathbf{0}_K \end{bmatrix}\ .$$

This gives us the following LP

$$\begin{aligned} \underset{x,\mu_1,\mu_2,\lambda}{\text{maximize}} \quad & c^T x \\ \text{subject to} \quad & \mu_1 - \mu_2 \leq b - a^T x \\ & \mu_1 - \mu_2 - \lambda = \mathbb{Z}^T x \\ & \mu_1 \geq 0, \mu_2 \geq 0, \lambda \geq 0 \\ & 0 \leq x \leq 1 \\ & \sum_{i=1}^n x_i \leq 1\ , \end{aligned}$$

where $\mu_1 \in \mathbb{R}$, $\mu_2 \in \mathbb{R}$, and $\lambda \in \mathbb{R}^K$.

We can then replace the expression $\mu_1 - \mu_2$ with $\mu \in \mathbb{R}$ which leaves us with

$$\begin{aligned}
\underset{x,\mu,\lambda}{\text{maximize}} \quad & c^T x \\
\text{subject to} \quad & \mu \leq b - a^T x \\
& \mu - \lambda = \mathbb{Z}^T x \\
& \lambda \geq 0 \\
& 0 \leq x \leq 1 \\
& \sum_{i=1}^{n} x_i \leq 1 \ ,
\end{aligned}$$

Then, since λ is only involved in one equality constraint other than the non-negativity one, we can replace that equality constraint with

$$\mu \geq \mathbb{Z}^T x \ .$$

This leaves us with the final option of replacing the μ with the largest amount it can take which is $b - a^T x$. We get the constraint

$$b - a^T x \geq \mathbb{Z}^T x \ ,$$

which is equivalent to

$$\bar{z}_i^T x \leq b - a^T x \, , \, \forall \, i = 1, \ldots, K \ .$$

In conclusion, the reformulation reduces to

$$\begin{aligned}
\underset{x}{\text{maximize}} \quad & c^T x \\
\text{subject to} \quad & (a + \bar{z}_i)^T x \leq b \, , \, \forall \, i = 1, \ldots, K \\
& 0 \leq x \leq 1 \\
& \sum_{i=1}^{n} x_i \leq 1 \ ,
\end{aligned}$$

which is exactly saying that x must satisfy the constraint for each scenarios $\bar{z}_i$. Recall that this is what we did in example 2.1. We actually just proved a version of the well known principle of robust optimization which states that the robust constraint

$$g(x, z) \leq 0 \, , \, \forall \, z \in \mathcal{Z}$$

is equivalent to the robust constraint

$$g(x, z) \leq 0 \, , \, \forall \, z \in \text{ConvexHull}(\mathcal{Z}) \ ,$$

when $g(x, z)$ is affine with respect to z.

See Colab code for implementation in RSOME.

Solution to Exercise 2.2: In order to employ theorem 2.7, we need to describe the uncertainty set in the form $Wz \leq v$ which is not currently the case. Our first step will therefore be to raise the uncertainty space in $\mathbb{R}^{2m}$ as follow

$$\mathcal{Z}'(\Gamma) := \left\{ z' \in \mathbb{R}^{2m} \,\middle|\, \exists\, z \in \mathbb{R}^m,\, s \in \mathbb{R}^m,\, z' = [\begin{array}{cc} z^T & s^T \end{array}]^T,\, -s \leq z \leq s,\, s \leq 1,\, \sum_i s_i \leq \Gamma \right\} .$$

In this uncertainty space, the robust constraint is equivalent to

$$(a + [\begin{array}{cc} \boldsymbol{I} & 0 \end{array}] z')^T x \leq b \,,\, \forall\, z' \in \mathcal{Z}'(\Gamma) \,.$$

We can therefore consider that $J = 1$, $p_1 = a$, $P_1 = [\begin{array}{cc} \boldsymbol{I} & 0 \end{array}]$, $q_1 = 0$, $r_1 = b$, and that

$$W := \begin{bmatrix} -I & -I \\ I & -I \\ 0 & I \\ 0 & \mathbf{1}_m \end{bmatrix} \qquad\qquad v := \begin{bmatrix} \mathbf{0}_m \\ \mathbf{0}_m \\ \mathbf{1}_m \\ \Gamma \end{bmatrix}$$

Hence, the reduced robust counterpart takes the following form

$$\begin{array}{ll} \underset{x,\lambda_1,\lambda_2,\lambda_3,\lambda_4}{\text{maximize}} & c^T x \\ \text{subject to} & a^T x + \mathbf{1}_m^T \lambda_3 + \Gamma \lambda_4 \leq b \\ & x = \lambda_2 - \lambda_1 \\ & \lambda_3 + \lambda_4 = \lambda_1 + \lambda_2 \\ & \lambda_1 \geq 0 \,,\, \lambda_2 \geq 0 \,,\, \lambda_3 \geq 0 \,,\, \lambda_4 \geq 0 \\ & 0 \leq x \leq 1 \\ & \displaystyle\sum_{i=1}^n x_i \leq 1 \,, \end{array}$$

where $\lambda_1 \in \mathbb{R}^m$, $\lambda_2 \in \mathbb{R}^m$, $\lambda_3 \in \mathbb{R}^m$, and $\lambda_4 \in \mathbb{R}$.

See Colab code for implementation in RSOME.

Solution to Exercise 2.3: We employ a similar approach as in exercise 2.1, meaning that we first reformulate in terms of θ being the uncertain vector. This leads to the following robust counterpart:

$$\begin{array}{ll} \text{maximize} & c^T x \\ \text{subject to} & (a + \mathbb{Z}\theta)^T x \leq b \,,\, \forall\, \theta \in \mathcal{U} \\ & 0 \leq x \leq 1 \\ & \displaystyle\sum_{i=1}^n x_i \leq 1 \,, \end{array}$$

where $\mathcal{U} := \{\theta \in \mathbb{R}^K \,|\, \theta \geq 0,\, \sum_{i=1}^K \theta_i = 1,\, \theta \leq \frac{1}{K\alpha}\}$. To apply theorem 2.7, we need to characterize the different elements of the LP-RC. Specifically, we say that $q = a$, $P = \mathbb{Z}$, $r = b$, $p = 0$, and finally that

$$W := \begin{bmatrix} -I \\ I \\ \mathbf{1}_K^T \\ -\mathbf{1}_K^T \end{bmatrix} \qquad \& \qquad v := \begin{bmatrix} \mathbf{0}_K \\ \frac{1}{K\alpha}\mathbf{1}_K \\ 1 \\ -1 \end{bmatrix} .$$

This leads to the following LP reformulation

$$\begin{aligned} \underset{x,\lambda_1,\lambda_2,\lambda_3,\lambda_4}{\text{maximize}} \quad & c^T x \\ \text{subject to} \quad & a^T x + \frac{1}{K\alpha}\mathbf{1}_K^T \lambda_2 + \lambda_3 - \lambda_4 \leq b \\ & \mathbb{Z}^T x = -\lambda_1 + \lambda_2 + \mathbf{1}_K \lambda_3 - \mathbf{1}_K \lambda_4 \\ & \lambda_1 \geq 0\,, \lambda_2 \geq 0\,, \lambda_3 \geq 0\,, \lambda_4 \geq 0 \\ & 0 \leq x \leq 1 \\ & \sum_{i=1}^n x_i \leq 1\;, \end{aligned}$$

where $\lambda_1 \in \mathbb{R}^K$, $\lambda_2 \in \mathbb{R}^K$, $\lambda_3 \in \mathbb{R}$, and $\lambda_4 \in \mathbb{R}$. Given that λ_3 and λ_4 always appear in the expression $\lambda_3 - \lambda_4$, we can simply replace the expression with $s := \lambda_3 - \lambda_4$. Also, given that λ_1 is only involved in one constraint, it can be removed from the problem after replacing the equality constraint with an inequality in the appropriate direction. Overall, we get the following reduced LP:

$$\begin{aligned} \underset{x,\lambda_2,s}{\text{maximize}} \quad & c^T x \\ \text{subject to} \quad & a^T x + \frac{1}{K\alpha}\mathbf{1}_K^T \lambda_2 + s \leq b \\ & \mathbb{Z}^T x \leq \lambda_2 + \mathbf{1}_K s \\ & \lambda_2 \geq 0 \\ & 0 \leq x \leq 1 \\ & \sum_{i=1}^n x_i \leq 1\;, \end{aligned}$$

See Colab code for implementation in RSOME.

Solution to Exercise 3.1: See Google Colab.

Solution to Exercise 3.2: See Google Colab.

Solution to Exercise 3.3: See Google Colab.

Solution to Exercise 3.4: The reformulation takes the form:

$$\begin{array}{rl}\max\limits_{x,y,\lambda,\mu^+,\mu^-,t,\phi^+,\phi^-} & -\gamma\bar{q}^T\phi^+ + (1/\gamma)\bar{q}^T\phi^- + t - c@x \\ \text{s.t.} & \phi_i^+ - \phi_i^- = -\bar{p}_i^T y + t,\ \forall i = 1,\dots,N \\ & d^T y + 1^T\mu^+ + 1^T\mu^- + \Gamma\sqrt{n}\lambda \le \bar{z}^T x \\ & \mu_i^+ - \mu_i^- = -\bar{z}_i x_i + \lambda,\ \forall i = 1,\dots,n \\ & Ax + By \le b \\ & x \ge 0,\ y \ge 0 \\ & \mu^+ \ge 0,\ \mu^- \ge 0,\ \lambda \ge 0 \\ & \phi^+ \ge 0,\ \phi^- \ge 0\end{array}$$

Solution to Exercise 4.1: In the case of $\Gamma = 1$, there are $2m$ vertices to worry about:

$$\mathcal{Z}(1) = \text{ConvexHull}(\mathcal{Z}_v(1)),\ \mathcal{Z}_v(1) := \{-e_1, -e_2, \dots, -e_m, e_1, e_2, \dots, e_m\}\ .$$

In the case of $\Gamma = m$, the situation is more complicated as there are now 2^m vertices to take care of:

$$\mathcal{Z}(m) = \text{ConvexHull}(\mathcal{Z}_v(m)),\ \mathcal{Z}_v(m) := \{-1, 1\}^m\ .$$

In both cases, one can choose an order on the elements of $\mathcal{Z}_v(\Gamma)$ and consider that it is composed of a list $\mathcal{Z}_v(\Gamma) := \{\bar{z}_1, \dots, \bar{z}_K\}$. The reformulation then would look like

$$\begin{array}{rl}\underset{x,y(\cdot)}{\text{maximize}} & \min\limits_{k=1,\dots,K} -\sum\limits_{i=1}^n c_i x_i + \sum\limits_{i=1}^n\sum\limits_{j=1}^m (r_{ij} - d_{ij}) y_{ij}^k \\ \text{subject to} & \sum\limits_i y_{ij}^k \le \bar{D}_j + \hat{D}_j \bar{z}_{jk}\ ,\ \forall\, k\ ,\ \forall\, j \\ & \sum\limits_j y_{ij}^k \le P_i x_i\ ,\ \forall\, k\ ,\ \forall\, i \\ & y_{ij}^k \ge 0\ ,\ \forall\, k\ ,\ \forall\, i, j \\ & x \in \{0,1\}^n\ ,\end{array}$$

Solution to Exercise 4.2: Indeed looking at problem (4.17), we can recognize that the only constraint that is affected by uncertainty is the constraint

$$\sum_{i=1}^n y_{ij} \le \bar{D}_j + \hat{D} z_j\ ,\ \forall\, j\ ,$$

and that each of these constraint depend on a different z_j while $\mathcal{Z} := [-1,1] \times [-1,1] \times \cdots \times [-1,1]$ so that conditions 1, 3, and 4 are satisfied. Regarding condition 2, we can verify that $\|x\|_\infty \leq 1$, while each $|y_{ij}| \leq P_j \leq \|P\|_\infty$ so that if $M = max(1, \|P\|_\infty)$ the condition is satisfied. This means that the adjustable model is equivalent to the non-adjustable one, namely that it is equivalent to

$$\begin{aligned}
\underset{x,y}{\text{maximize}} \quad & -\sum_{i=1}^{n} c_i x_i + \sum_{i=1}^{n}\sum_{j=1}^{m}(r_{ij} - d_{ij})y_{ij} \\
\text{subject to} \quad & \sum_{i=1}^{n} y_{ij} \leq \bar{D}_j + \hat{D}_j z_j \,,\, \forall\, z \in \mathcal{Z}(m) \,,\, \forall\, j \\
& \sum_{j=1}^{m} y_{ij} \leq P_i x_i \,,\, \forall\, i \\
& y_{ij} \geq 0 \;\forall\, i, j \\
& x \in \{0,1\}^n \,.
\end{aligned}$$

Yet, we also know that for any choice of y_{ij} the worst-case scenario for the demand constraint would be that $z_j = -1$. This makes the constraint reduce to

$$\sum_{i=1}^{n} y_{ij} \leq \bar{D}_j - \hat{D}_j \,,\, \forall\, j$$

Solution to Exercise 4.3: In each iteration of the algorithm, we need to solve for some current $\mathcal{Z}'_v := \{\bar{z}_1, \ldots, \bar{z}_{K'}\}$

$$\begin{aligned}
\underset{x,y}{\text{maximize}} \quad & \min_{k=1,\ldots,K'} -\sum_{i=1}^{n} c_i x_i + \sum_{i=1}^{n}\sum_{j=1}^{m}(r_{ij} - d_{ij})y_{ij}^k \\
\text{subject to} \quad & \sum_{i} y_{ij}^k \leq \bar{D}_j + \hat{D}_j \bar{z}_{jk} \,,\, \forall\, k = 1, \ldots, K' \,,\, \forall\, j \\
& \sum_{j} y_{ij}^k \leq P_i x_i \,,\, \forall\, k = 1, \ldots, K' \,,\, \forall\, i \\
& y_{ij}^k \geq 0 \,,\, \forall\, k = 1, \ldots, K' \,,\, \forall\, i, j \\
& x \in \{0,1\}^n \,.
\end{aligned}$$

In order to obtain the new z' to add to the list Z'_v, one needs for the current version of the optimal x solve the following mixed integer program:

$$
\begin{array}{ll}
\underset{\substack{z \in \mathcal{Z}, y, \lambda, \gamma, \omega \\ \nu^\lambda, \nu^\gamma, \nu^\omega}}{\text{minimize}} & -\sum_{i=1}^{n} c_i x_i + \sum_{i=1}^{n} \sum_{j=1}^{m} (r_{ij} - d_{ij}) y_{ij} \\
\text{subject to} & r_{ij} - d_{ij} = \lambda_j + \gamma_i - \omega_{ij} \,,\, \forall\, i, j \\
& 0 \le \bar{D}_j + \hat{D}_j z_j - \sum_i y_{ij} \le M(1 - \nu_j^\lambda) \,,\, \forall\, j = 1, \ldots, m \\
& 0 \le P_i x_i - \sum_j y_{ij} \le M(1 - \nu_i^\gamma) \,,\, \forall\, i = 1, \ldots, n \\
& 0 \le y_{ij} \le M(1 - \nu_{ij}^\omega) \\
& 0 \le \lambda \le M\nu^\lambda \\
& 0 \le \gamma \le M\nu^\gamma \\
& 0 \le \omega \le M\nu^\omega \\
& \nu^\lambda \in \{0,1\}^m ,\, \nu^\gamma \in \{0,1\}^n \,, \nu^\omega \in \{0,1\}^{n \times m} \,,
\end{array}
$$

where $\lambda \in \mathbb{R}^m$ is the dual variable for the demand constraint, $\gamma \in \mathbb{R}^n$ is the dual variable for the max production constraint, and $\omega \in \mathbb{R}^{n \times m}$ is the dual variable for the non-negativity constraint.

When the budgeted uncertainty set is used, this program can be linearized to

$$
\begin{array}{ll}
\underset{\substack{z, w, y, \lambda, \gamma, \omega \\ \nu^\lambda, \nu^\gamma, \nu^\omega}}{\text{minimize}} & -\sum_{i=1}^{n} c_i x_i + \sum_{i=1}^{n} \sum_{j=1}^{m} (r_{ij} - d_{ij}) y_{ij} \\
& r_{ij} - d_{ij} = \lambda_j + \gamma_i - \omega_{ij} \,,\, \forall\, i, j \\
& 0 \le \bar{D}_j + \hat{D}_j z_j - \sum_i y_{ij} \le M(1 - \nu_j^\lambda) \,,\, \forall\, j = 1, \ldots, m \\
& 0 \le P_i x_i - \sum_j y_{ij} \le M(1 - \nu_i^\gamma) \,,\, \forall\, i = 1, \ldots, n \\
& 0 \le y_{ij} \le M(1 - \nu_{ij}^\omega) \\
& 0 \le \lambda \le M\nu^\lambda \\
& 0 \le \gamma \le M\nu^\gamma \\
& 0 \le \omega \le M\nu^\omega \\
& \nu^\lambda \in \{0,1\}^m ,\, \nu^\gamma \in \{0,1\}^n \,, \nu^\omega \in \{0,1\}^{n \times m} \\
& -w \le z \le w \\
& w \le 1 \\
& \sum_{j=1}^{m} w_j \le \Gamma \,.
\end{array}
$$

Solution to Exercise 6.1: We consider the following robust constraint

$$g(x,z) \leq t\,,\, \forall\, z \in \mathcal{D}(\rho)\;,$$

where $g(x,z) := \mathrm{CVaR}_{1-\epsilon}(r^T x; z)$.

First, we will work on the support function. We let $\mathcal{Z} := \mathcal{Z}_1 \cap \mathcal{Z}_2$ with $\mathcal{Z}_1$ the set discussed in theorem 6.6 and Z_2 be the KL-divergence set presented in table 6.1. Hence,

$$\begin{aligned}
\delta^*(v|\mathcal{Z}_1) &:= \min_{\lambda\in\mathbb{R}:\lambda\geq v} \lambda \\
\delta^*(v|\mathcal{Z}_2) &:= \min_{\mu\in\mathbb{R}:\mu\geq 0} \sum_{k=1}^{K} \frac{1}{K}\mu \exp(v_k/\mu - 1) + \rho\mu\;.
\end{aligned}$$

Based on the rule presented in table 6.1 for intersection of sets, this indicates us that support function for $\mathcal{Z}$ should be

$$\begin{aligned}
\delta^*(v|\mathcal{Z}) &= \min_{\lambda,\mu\geq 0, w^1, w^2:\lambda\geq w^1,\, w^1+w^2=v} \quad \lambda + \sum_{k=1}^{K}\frac{1}{K}\mu\exp(w_k^2/\mu - 1) + \rho\mu \\
&= \min_{\lambda,\mu\geq 0, w:\lambda\geq v-w} \quad \lambda + \sum_{k=1}^{K}\frac{1}{K}\mu\exp(w_k/\mu - 1) + \rho\mu\;.
\end{aligned}$$

Now, looking into $g_*(x,v)$, we start by laying out the detailed definition of this conjugate function:

$$\begin{aligned}
g_*(x,v) &:= \inf_{p\in\mathbb{R}^K:p\geq 0, \sum_k p_k=1} v^T p - \inf_s s + (1/(1-\epsilon)) \sum_k p_k \max(-\bar{r}_k^T x - s;\, 0) \\
&= \inf_{p\in\mathbb{R}^K:p\geq 0, \sum_k p_k=1} \sup_s v^T p - s - (1/(1-\epsilon)) \sum_k p_k \max(-\bar{r}_k^T x - s;\, 0) \\
&= \sup_s \inf_{p\in\mathbb{R}^K:p\geq 0, \sum_k p_k=1} v^T p - s - (1/(1-\epsilon)) \sum_k p_k \max(-\bar{r}_k^T x - s;\, 0) \\
&= \sup_s \min_{k=1,\ldots,K} v_k - s - (1/(1-\epsilon)) \max(-\bar{r}_k^T x - s;\, 0)\;,
\end{aligned}$$

where we exploited Sion's minimax theorem exploiting the fact that the feasible set for p is bounded, and where we realized that a search over the worst-case distribution is simply a search over the worst-case outcome.

In conclusion, we can state that the robust CVaR optimization model takes the form:

$$\begin{aligned} \underset{x,s,t,\lambda,\mu,v,w}{\text{minimize}} \quad & t \\ & \lambda + \sum_{k'} \frac{1}{K} \mu \exp(w_{k'}/\mu - 1) + \rho\mu - v_k + s + (1/(1-\epsilon)) \max(-\bar{r}_k^T x - s; 0) \leq t \,, \forall\, k \\ & \lambda \geq v - w \\ & \mu \geq 0 \\ & \sum_i x_i = 1 \\ & x \geq 0 \,, \end{aligned}$$

where $\lambda \in \mathbb{R}$, $\mu \in \mathbb{R}$, $v \in \mathbb{R}^K$, $w \in \mathbb{R}^K$, $s \in \mathbb{R}$, and $t \in \mathbb{R}$.

Solution to Exercise 6.2:
Question 1: The optimization problem can be reformulated as

$$\begin{aligned} \underset{x,y,t}{\text{maximize}} \quad & t \\ \text{subject to} \quad & t \leq \sum_i c_i y_i^{a_i} - c_i \,, \forall\, a \in \mathcal{U} \\ & y_i = 1 + x_i/d_i \\ & \sum_i p_i x_i \leq B \\ & x \geq 0 \,, \end{aligned}$$

We therefore wish to use theorem 6.2 in order to obtain a tractable form for :

$$t - \sum_i c_i y_i^{a_i} + c_i \leq 0 \,, \forall\, a \in \mathcal{U}$$

Considering first the function $g(y_i, a_i) := -\sum_i c_i y_i^{a_i}$, table 6.2 already tells us how to obtain the robust counterpart of $g'(y_i, a_i) := -y_i^{a_i}$, namely

$$g'_*(y_i, v_i) = \frac{v_i}{\ln(y_i)} \ln\left(\frac{-v_i}{\ln(y_i)}\right) - \frac{v_i}{\ln(y_i)} \,,$$

with the restriction that $v_i \leq 0$ otherwise the function evaluates to $-\infty$.

Using the help of theorem 6.9, we obtain that if $g''(y_i, a_i) := -c_i y_i^{a_i}$ then the conjugate function is

$$g''_*(y_i, v_i) = c_i \left(\frac{v_i/c_i}{\ln(y_i)} \ln\left(\frac{-v_i/c_i}{\ln(y_i)}\right) - \frac{v_i/c_i}{\ln(y_i)}\right) = \frac{v_i}{\ln(y_i)} \ln\left(\frac{-v_i/c_i}{\ln(y_i)}\right) - \frac{v_i}{\ln(y_i)} \,.$$

Finally, the table tells us how to handle sums of separable functions. Hence, we are left with the following robust counterpart:

$$\begin{aligned}
& t + \delta^*(v|\mathcal{U}) + \sum_i c_i - \sum_i \left(\frac{v_i}{\ln(y_i)} \ln\left(\frac{-v_i/c_i}{\ln(y_i)}\right) - \frac{v_i}{\ln(y_i)} \right) \leq 0 \\
& v \leq 0 \,.
\end{aligned}$$

Now, to obtain $\delta^*(v, |\mathcal{U}_1)$, we first note that $\mathcal{U}_1 = \bar{a} - 0.25\,\mathbf{diag}(\bar{a})\mathcal{Z}$, where $\mathbf{diag}(\bar{a}) = \sum_{i=1}^n \bar{a} e_i e_i^T$ (i.e. a matrix in $\mathbb{R}^{n\times n}$ with diagonal equal to $\bar{a}$) and with $\mathcal{Z}$ defined as the uncertainty set discussed in theorem 6.5. Hence,

$$\delta^*(v|\mathcal{Z}) := \min_{(\lambda,w)\in\mathbb{R}\times\mathbb{R}^m : \lambda \geq v - w\,,\, \lambda\geq 0,\, w\geq 0} \sum_i w_i + \Gamma\lambda \,.$$

Following theorem 6.7, we have that

$$\delta^*(v|\mathcal{U}) = \bar{a}^T v + \delta^*(-0.25\,\mathbf{diag}(\bar{a})v|\mathcal{Z}) \,,$$

hence, that

$$\delta^*(v|\mathcal{U}) = \min_{(\lambda,w)\in\mathbb{R}\times\mathbb{R}^m : \lambda \geq -0.25\,\mathbf{diag}(\bar{a})v - w\,,\, \lambda\geq 0,\, w\geq 0} \bar{a}^T v + \sum_i w_i + \Gamma\lambda \,.$$

Combining the two steps we get the tractable reformulation of the robust counterpart:

$$\begin{aligned}
\underset{x,y,t,\lambda,v,w}{\text{maximize}} \quad & t \\
\text{subject to} \quad & t + \bar{a}^T v + \sum_i w_i + \Gamma\lambda + \sum_i \left(c_i - \frac{v_i}{\ln(y_i)} \ln\left(\frac{-v_i/c_i}{\ln(y_i)}\right) + \frac{v_i}{\ln(y_i)} \right) \leq 0 \\
& \lambda \geq -0.25\,\mathbf{diag}(\bar{a})v - w \\
& v \leq 0 \\
& w \geq 0 \\
& \lambda \geq 0 \\
& y_i = 1 + x_i/d_i \\
& \sum_i p_i x_i \leq B \\
& x \geq 0 \,,
\end{aligned}$$

Question 2: To obtain $\delta^*(v, |\mathcal{U}_2)$, we first note that $\mathcal{U}_2 = \bar{a} - 0.25\,\mathbf{diag}(\bar{a})\mathcal{Z}$, where $\mathbf{diag}(\bar{a}) = \sum_{i=1}^n \bar{a} e_i e_i^T$ (i.e. a matrix in $\mathbb{R}^{n\times n}$ with diagonal equal to $\bar{a}$) and with $\mathcal{Z}$ defined as

$$\mathcal{Z} := \{z \in \mathbb{R}^n \,|\, z \geq 0,\, \sum_i z_i = 1,\, \sum_i z_i \ln(z_i) \leq \rho\} \,.$$

To "speed up" the analysis, we observe that $\mathcal{Z} := \mathcal{Z}_1 \cap \mathcal{Z}_2$ with $\mathcal{Z}_1$ the set discussed in theorem 6.6 and Z_2 be the KL-divergence set presented in table 6.1. Hence,

$$\delta^*(v|\mathcal{Z}_1) := \min_{\lambda\in\mathbb{R}:\lambda\geq v} \lambda$$
$$\delta^*(v|\mathcal{Z}_2) := \min_{\mu\in\mathbb{R}:\mu\geq 0} \sum_i \mu\exp(v_i/\mu - 1) + \rho\mu \; .$$

Based on the rule presented in table 6.1 for intersection of sets, this indicates us that the support function for $\mathcal{Z}$ should be

$$\delta^*(v|\mathcal{Z}) = \min_{\lambda,\mu\geq 0, w^1, w^2:\lambda\geq w^1,\, w^1+w^2=v} \lambda + \sum_i \mu\exp(w_i^2/\mu - 1) + \rho\mu \; .$$

Following theorem 6.7, we have that

$$\delta^*(v|\mathcal{U}_2) = \bar{a}^T v + \delta^*(-0.25\,\mathbf{diag}(\bar{a})v|\mathcal{Z}) \; ,$$

hence, that

$$\begin{aligned}\delta^*(v|\mathcal{U}_2) &= \min_{\lambda,\mu\geq 0, w^1, w^2:\lambda\geq w^1,\, w^1+w^2=-0.25\,\mathbf{diag}(\bar{a})v} \bar{a}^T v + \lambda + \sum_i \mu\exp(w_i^2/\mu - 1) + \rho\mu \\ &= \min_{\lambda,\mu\geq 0, w:\lambda\geq -0.25\,\mathbf{diag}(\bar{a})v - w} \bar{a}^T v + \lambda + \sum_i \mu\exp(w_i/\mu - 1) + \rho\mu \; .\end{aligned}$$

Combining the two steps we get the tractable reformulation of the robust counterpart:

$$\begin{aligned}\underset{x,y,t,s,\lambda,\mu,v,w}{\text{maximize}} \quad & t \\ \text{subject to} \quad & t + \bar{a}^T v + s + \sum_i \left(c_i - \frac{v_i}{\ln(y_i)} \ln\left(\frac{-v_i/c_i}{\ln(y_i)}\right) + \frac{v_i}{\ln(y_i)}\right) \leq 0 \\ & s \geq \lambda + \sum_i \mu\exp(w_i/\mu - 1) + \rho\mu \\ & \lambda \geq -0.25\,\mathbf{diag}(\bar{a})v - w \\ & v \leq 0 \\ & \mu \geq 0 \\ & y_i = 1 + x_i/d_i \\ & \sum_i p_i x_i \leq B \\ & x \geq 0 \; ,\end{aligned}$$

Question 3: Refer to Matlab implementation using YALMIP in "Ex6_2.m".

Solution to Exercise 6.3: We can reformulate this problem as

$$\begin{aligned} \underset{x,t}{\text{maximize}} \quad & t \\ \text{subject to} \quad & -\sum_i x_i \exp(z_i) \leq -t \,,\, z \in \mathcal{Z} \\ & \sum_i x_i \leq 1 \\ & x \geq 0 \,, \end{aligned}$$

We first look at the objective function $g(x,z) := -\sum_i x_i \exp(z_i)$. To find the partial concave conjugate, we start with $g'(x_i, z_i) := -x_i \exp(z_i)$. The partial concave conjugate of this function can be found as

$$g'_*(x_i, v_i) := \inf_{z_i} v_i z_i + x_i \exp(z_i) = v_i \ln(-v_i/x_i) - v_i \,,$$

as long as $v_i \leq 0$ otherwise the infimum goes to $-\infty$. Based on the sum of separable functions rules, we get

$$g_*(x,v) := \sum_i v_i \ln(-v_i/x_i) - v_i \,.$$

Next, we need to identify the support function of $\mathcal{Z}$. Yet, in the description, we see that it is the affine mapping of the sum of two sets : $\mathcal{Z} := \mu + Q(\mathcal{Z}_1 + \mathcal{Z}_2)$, where $\mathcal{Z}_1$ is the budgeted uncertainty set, while $\mathcal{Z}_2 := 1 \cdot [-1,1]$, an affine projection of the $[-1,1]$ interval. We therefore get:

$$\begin{aligned} \delta^*(v|[-1,1]) &:= \|v\|_1 \;\; \text{(based on support function of the box in } \mathbb{R}) \\ \delta^*(v|\mathcal{Z}_2) &:= \sum_i |v_i| \;\; \text{(based on theorem 6.7 and } \mathcal{Z}_2 := 1 \cdot [-1,1]) \\ \delta^*(v|\mathcal{Z}_1) &:= \min_{w^+ \geq 0, w^- \geq 0, \lambda \geq 0:\, \lambda \geq v - w^+,\, \lambda \geq -v - w^-} \sum_i w_i^+ + w_i^- + \Gamma\lambda \;\; \text{(based on corollary 6.8)} \\ \delta^*(v|\mathcal{Z}_1 + \mathcal{Z}_2) &:= \min_{w^+ \geq 0, w^- \geq 0, \lambda \geq 0: \lambda \geq v - w^+,\, \lambda \geq -v - w^-} \sum_i w_i^+ + w_i^- + \Gamma\lambda + \sum_i |v_i| \\ \delta^*(v|\mathcal{Z}) &:= \delta^*(v|\mu + Q(\mathcal{Z}_1 + \mathcal{Z}_2)) \\ &= \min_{w^+ \geq 0, w^- \geq 0, \lambda \geq 0: \lambda \geq Q^T v - w^+,\, \lambda \geq -Q^T v - w^-} \mu^T v + \sum_i w_i^+ + w_i^- + \Gamma\lambda + \sum_i |q_i^T v| \,, \end{aligned}$$

where q_i is the i-th column of Q. Note that the last two support function are obtained used the Minkowski sum rule from table 6.1 and theorem 6.7

Putting both of these analysis together we get:

$$\begin{array}{ll}
\underset{x,t,v,w^+,w^-,\lambda}{\text{maximize}} & t \\
\text{subject to} & \mu^T v + \sum_i w_i^+ + w_i^- + \Gamma\lambda + |\sum_i q_i^T v_i| - (\sum_i (v_i \ln(-v_i/x_i) - v_i)) \leq -t\,, \\
& \lambda \geq Q^T v - w^+ \\
& \lambda \geq -Q^T v - w^- \\
& w^+ \geq 0, w^- \geq 0, \lambda \geq 0 \\
& \sum_i x_i \leq 1 \\
& x \geq 0\ ,
\end{array}$$

Solution to Exercise 8.1: Since $\mathcal{D}_1 = \mathcal{D}(\mathcal{Z}, \bar{\mu})$, we can exploit theorem 8.6 to reformulate this DRO has

$$\begin{array}{ll}
\underset{x\in\mathcal{X},q,t,w^1,w^2}{\text{minimize}} & t \\
\text{subject to} & t \geq \delta^*(w^1|\mathcal{Z}) + \mu^T q - h_*^1(x, w^1 + q) \\
& t \geq \delta^*(w^2|\mathcal{Z}) + \mu^T q - h_*^2(x, w^2 + q)\,,
\end{array}$$

where $h^1(x,z) := -\frac{1}{2}\xi^T Q(x)\xi$ and $h^2(x,z) := x^T Cz$. We can first have a look at $\delta^*(v|\mathcal{Z})$ which is the support function of a polyhedron:

$$\delta^*(w|\mathcal{Z}) = \min_{\lambda\geq 0: W^T\lambda = w} v^T\lambda\,.$$

Next, we can obtain $h_*^2(x,w)$ using table 6.2 which tells us:

$$h_*^2(x,w) = \left\{ \begin{array}{ll} 0 & \text{if } w = C^T x \\ \infty & \text{otherwise} \end{array} \right. .$$

Finally, we can also exploit table 6.2 for $h_*^2(x,w)$ since $h^2(x,z) = -\sum_i z^T Q_i z x_i$:

$$h_*^1(x,w) = \sup_{s^1,s^2,\ldots,s^n:\sum_i s^i = w} -(1/2)\sum_i (1/x_i) s^i Q_i^{-1} s^i\,.$$

Assembling everything together, we get:

$$\begin{array}{ll}
\underset{x\in\mathcal{X},q,t,w^1,w^2,\lambda^1,\lambda^2,s^1,\ldots,s^n}{\text{minimize}} & t \\
\text{subject to} & t \geq v^T\lambda^1 + \mu^T q + (1/2)\sum_i (1/x_i) s^i Q_i^{-1} s^i \\
& \lambda^1 \geq 0\ ,\ W^T\lambda^1 = w^1 \\
& \sum_i s^i = w^1 + q \\
& t \geq v^T\lambda^2 + \mu^T q \\
& \lambda^2 \geq 0\ ,\ W^T\lambda^2 = w^2 \\
& w^2 + q = C^T x\,.
\end{array}$$

Solution to Exercise 8.2: We can easily remark that the DRO reduces to

$$\underset{x\in\mathcal{X}}{\text{minimize}} \quad \max_{\mu\in\mathcal{U}_1'(\Gamma)} \max_{F\in\mathcal{D}(\mathcal{Z},\mu)} \mathbb{E}_F[\max(-\frac{1}{2}\xi^T Q(x)\xi,\ x^T C\xi)]\,,$$

where

$$\mathcal{U}_1'(\Gamma) = \{\mu \,|\, \exists\Delta\in\mathcal{U}_1(\Gamma),\, \mu = \bar{\mu}+\Delta\}\,,$$

with

$$\mathcal{U}_1(\Gamma) := \{\Delta\in\mathbb{R}^m \,|\, \Delta\geq 0 \sum_{i=1}^m \Delta_i \leq \Gamma\}\,.$$

Based on corollary 8.9, one should understand that the DRO now becomes equivalent to :

$$\begin{array}{ll}
\underset{x\in\mathcal{X},q,t,w^1,w^2,\lambda^1,\lambda^2,s^1,\ldots,s^n}{\text{minimize}} & t+\delta^*(q|\mathcal{U}_1'(\Gamma)) \\
\text{subject to} & t \geq v^T\lambda^1 + (1/2)\sum_i (1/x_i)s^i Q_i^{-1}s^i \\
& \lambda^1\geq 0\,,\ W^T\lambda^1 = w^1 \\
& \sum_i s^i = w^1+q \\
& t\geq v^T\lambda^2 \\
& \lambda^2\geq 0\,,\ W^T\lambda^2 = w^2 \\
& w^2+q = C^T x\,.
\end{array}$$

where $\mathcal{U}_1'(\Gamma) = \{\mu \,|\, \exists\Delta\in\mathcal{U}_1(\Gamma),\, \mu=\bar{\mu}+\Delta\}$. Based on theorem 6.7, we have that $\delta^*(q|\mathcal{U}_1'(\Gamma)) = q^T\bar{\mu} + \delta^*(q|\mathcal{U}_1(\Gamma)$.

$$\delta^*(q|\mathcal{U}_1(\Gamma)) = \sup_{\Delta\in\mathcal{U}_1(\Gamma)} q^T\Delta = \min_{\omega\geq 0:\omega\geq q} \Gamma\omega\,.$$

Hence, we obtain:

$$\begin{array}{ll}
\underset{x\in\mathcal{X},q,t,w^1,w^2,\lambda^1,\lambda^2,s^1,\ldots,s^n,\omega}{\text{minimize}} & t+\bar{\mu}^T q + \Gamma\omega \\
\text{subject to} & \omega\geq 0\,,\ \omega\geq q \\
& t \geq v^T\lambda^1 + (1/2)\sum_i (1/x_i)s^i Q_i^{-1}s^i \\
& \lambda^1\geq 0\,,\ W^T\lambda^1 = w^1 \\
& \sum_i s^i = w^1+q \\
& t\geq v^T\lambda^2 \\
& \lambda^2\geq 0\,,\ W^T\lambda^2 = w^2 \\
& w^2+q = C^T x\,.
\end{array}$$

Solution to Exercise 8.3: The answer to this question is easier to reach if we realize that our previous efforts led us to:

$$\sup_{F\in\mathcal{D}_2(\Gamma)} \mathbb{E}_F[\max(-\frac{1}{2}\xi^T Q(x)\xi,\ x^T C\xi)] = \underset{(q,t,\omega)\in\mathcal{Q}}{\text{minimize}} \quad t + \bar{\mu}^T q + \Gamma\omega\,,$$

where

$$\mathcal{Q} := \left\{ (q,t,\omega) \;\middle|\; \exists w^1, w^2, \lambda^1, \lambda^2, s^1, \ldots, s^n, \omega\,,\; \begin{array}{l} \omega \geq 0\,,\ \omega \geq q \\ t \geq v^T\lambda^1 + (1/2)\sum_i (1/x_i) s^i Q_i^{-1} s^i \\ \lambda^1 \geq 0\,,\ W^T\lambda^1 = w^1 \\ \sum_i s^i = w^1 + q \\ t \geq v^T\lambda^2 \\ \lambda^2 \geq 0\,,\ W^T\lambda^2 = w^2 \\ w^2 + q = C^T x\,. \end{array} \right\}$$

Hence

$$\begin{aligned}
&\sup_{\Gamma\in[0,\bar{\Gamma}]} \sup_{F\in\mathcal{D}_2(\Gamma)} \mathbb{E}_F[\max(-\frac{1}{2}\xi^T Q(x)\xi,\ x^T C\xi)] - \alpha\Gamma \\
&\quad = \sup_{\Gamma\in[0,\bar{\Gamma}]} \min_{(q,t,\omega)\in\mathcal{Q}} \quad f(t,q,\omega,\Gamma) := t + \bar{\mu}^T q + \Gamma(\omega - \alpha) \\
&\quad = \min_{(q,t,\omega)\in\mathcal{Q}} \quad t + \bar{\mu}^T q + \sup_{\Gamma\in[0,\bar{\Gamma}]} \Gamma(\omega-\alpha) \\
&\quad = \min_{(q,t,\omega)\in\mathcal{Q}} \quad t + \bar{\mu}^T q + \max(0,\ \bar{\Gamma}(\omega-\alpha))\,,
\end{aligned}$$

where we first applied Sion's minimax theorem since Γ is in a bounded set and the function $f(t,q,\omega,\Gamma)$ is linear in both (t,q,ω) and Γ. The last equality comes from the fact that since $\Gamma(\omega-\alpha)$ is linear in Γ, then the supremum over Γ necessarily either occurs at $\Gamma = 0$ or $\Gamma = \bar{\Gamma}$.

We can finally reintegrate this expression inside the globalized distributionally robust optimization problem presented in this question:

$$
\begin{array}{ll}
\underset{x\in\mathcal{X},q,t_1,t_2,w^1,w^2,\lambda^1,\lambda^2,s^1,\dots,s^n,\omega}{\text{minimize}} & t_1 \\
\text{subject to} & t_2 + \bar{\mu}^T q \leq t_1 \\
& t_2 + \bar{\mu}^T q + \Gamma(\omega - \alpha) \leq t_1 \\
& \omega \geq 0 \ , \ \omega \geq q \\
& t_2 \geq v^T \lambda^1 + (1/2) \displaystyle\sum_i (1/x_i) s^i Q_i^{-1} s^i \\
& \lambda^1 \geq 0 \ , \ W^T \lambda^1 = w^1 \\
& \displaystyle\sum_i s^i = w^1 + q \\
& t_2 \geq v^T \lambda^2 \\
& \lambda^2 \geq 0 \ , \ W^T \lambda^2 = w^2 \\
& w^2 + q = C^T x \, .
\end{array}
$$

Bibliography

[1] A. Ardestani-Jaafari and Erick Delage. Robust optimization of sums of piecewise linear functions with application to inventory problems. *Operations Research*, 64(2):474–494, 2016.

[2] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.

[3] Chaithanya Bandi, Dimitris Bertsimas, and Nataly Youssef. Robust queueing theory. *Operations Research*, 63(3):676–700, 2015.

[4] Basel Committee on Banking Supervision. Consultative document : Fundamental review of the trading book. Technical report, Bank for International Settlements, 2012.

[5] G. Bayraksan and D. K. Love. *Data-Driven Stochastic Programming Using Phi-Divergences*, chapter 1, pages 1–19. 2015.

[6] A. Ben-Tal, D. Bertsimas, and D. B. Brown. A soft robust model for optimization under ambiguity. *Operations Research*, 58(4-part-2):1220–1234, 2010.

[7] A. Ben-Tal, S. Boyd, and A. Nemirovski. Extending scope of robust optimization: Comprehensive robust counterparts of uncertain problems. *Mathematical Programming*, 107(1-2):63–89, 2006.

[8] A. Ben-Tal, R. Brekelmans, D. den Hertog, and J.-P. Vial. Globalized robust optimization for nonlinear uncertain inequalities. *INFORMS Journal on Computing*, 29(2):350–366, 2017.

[9] A. Ben-Tal, D. den Hertog, and J.-P. Vial. Deriving robust counterparts of nonlinear uncertain inequalities. *Mathematical Programming*, 149(1-2):265–299–1495, 2015.

[10] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust Optimization.* Princeton Series in Applied Mathematics, 2009.

[11] A. Ben-Tal, A. Goryashko, E. Guslitzer, and A. Nemirovski. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99(2):351–376, 2004.

[12] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805, 1998.

[13] A. Ben-Tal and A. Nemirovski. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming, Series A*, 88(3):411–424, 2000.

[14] D. Bertsimas and D. B. Brown. Constructing uncertainty sets for robust linear optimization. *Operations Research*, 57(6):1483–1495, 2009.

[15] D. Bertsimas, D. B. Brown, and C. Caramanis. Theory and applications of robust optimization. *SIAM Review*, 53(3):464–501, 2011.

[16] D. Bertsimas and V. Goyal. On the power of robust solutions in two-stage stochastic and adaptive optimization problems. *Mathematics of Operations Research*, 35(2):284–305, 2010.

[17] D. Bertsimas, V. Goyal, and B. Lu. A tight characterization of the performance of static solutions in two-stage adjustable robust linear optimization. *Mathematical Programming*, 150(2):281–319, 2014.

[18] D. Bertsimas, V. Gupta, and N. Kallus. Data-driven robust optimization. *Mathematical Programming*, 2017.

[19] D. Bertsimas and M. Sim. The price of robustness. *Operations Research*, 52(1):35–53, 2004.

[20] D. Bertsimas and A. Thiele. A robust optimization approach to inventory theory. *Operations Research*, 54(1):150–168, 2006.

[21] D. Bertsimas and J. N. Tsitsiklis. *Introduction to Linear Optimization*, volume 1st edition. Athena Scientific, 1997.

[22] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[23] A. Charnes and W. W. Cooper. Chance constrained programming. *Management Science*, 6:73–79, 1959.

[24] Zhi Chen, Melvyn Sim, and Peng Xiong. Robust stochastic optimization made easy with rsome. *Management Science*, 2020.

[25] F.J.C.T. de Ruiter, A. Ben-Tal, R.C.M. Brekelmans, and D. den Hertog. Robust optimization of uncertain multistage inventory systems with inexact data in decision rules. *Computational Management Science*, 14(1):45–66, 2017.

[26] E. Delage, L. G. Gianoli, and B. Sanso. A practicable robust counterpart formulation for decomposable functions : A network congestion case study. *Operations Research*, 2018.

[27] E. Delage, D. Kuhn, and W. Wiesemann. Dice-sion making under uncertainty: When can a random decision reduce risk? Les cahiers du gerad g-2016-64, GERAD, 2016.

[28] E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010.

[29] C.A. Van Eijl. Capacity planning for carrier-scale ip networks. *BT Technology Journal*, 20(3):116–123, 2002.

[30] M. R. Garey and D. S. Johnson. *Computers and Intractability: A guide to the Theory of NP-Completeness.* W. H. Freeman and Company, 1979.

[31] J. Goh and M. Sim. Robust optimization made easy software. `http://www.robustopt.com`, 2011. Accessed: 2015-04-15.

[32] D. A. Iancu and N. Trichakis. Pareto efficiency in robust optimization. *Management Science*, 60(1):130–147, 2014.

[33] A. Marandi and D. den Hertog. When are static and adjustable robust optimization with constraint-wise uncertainty equivalent? *Mathematical Programming*, pages 1–14, 2017.

[34] H.M. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

[35] A. W. Marshall and I. Olkin. Multivariate chebyshev inequalities. *The Annals of Mathematical Statistics*, 31(4):1001–1014, 1960.

[36] P. Mohajerin Esfahani and D. Kuhn. Data-driven distributionally robust optimization using the wasserstein metric: performance guarantees and tractable reformulations. *Mathematical Programming*, 2017.

[37] P. Mohajerin Esfahani, S. Shafieezadeh Abadeh, G. Hanasusanto, and D. Kuhn. Data-driven inverse optimization with incomplete information. *Working draft*, pages 1–25, 2015.

[38] I. Pólik and T. Terlaky. A survey of the S-Lemma. *SIAM Review*, 49(3):371–418, 2007.

[39] R. T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–41, 2000.

[40] W. W. Rogosinsky. Moments of non-negative mass. In *Proceedings of Royal Society London Series A*, pages 1–27, 1958.

[41] H. Scarf. A min-max solution of an inventory problem. *Studies in The Mathematical Theory of Inventory and Production*, pages 201–209, 1958.

[42] S. Shafieezadeh Abadeh, P. Mohajerin Esfahani, and D. Kuhn. Distributionally robust logistic regression. In *NIPS*, 2015.

[43] A. Shapiro. On duality theory of conic linear problems. In M. A. Goberna and M. A. López, editors, *Semi-Infinite Programming: Recent Advances*, pages 135–165. Kluwer Academic Publishers, 2001.

[44] M. Sion. On general minimax theorems. *Pacific Journal of Mathematics*, 8(1):171–176, 1958.

[45] J. E. Smith and R. L. Winkler. The optimizer's curse: Skepticism and postdecision surprise in decision analysis. *Management Science*, 52(3):311–322, 2006.

[46] A. L. Soyster. Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research*, 21(5):1154–1157, 1973.

[47] B. Van Parys. *Distributionally robust control and optimization*. PhD thesis, ETH Z urich, 2015.

[48] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, NJ, 1944.

[49] A. Wald. Statistical decision functions which minimize the maximum risk. *The Annals of Mathematics*, 46(2):265–280, 1945.

[50] W. Wiesemann, D. Kuhn, and M. Sim. Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376, 2014.

[51] H. Xu, C. Caramanis, and S. Mannor. Optimization under probabilistic envelope constraints. *Operations Research*, 60(3):682–699, 2012.

[52] V.A. Yakubovich. S-procedure in nonlinear control theory. Technical report, Vestnik Leningrad. Univ. (in Russian), 1971.

[53] B. Zeng and L. Zhao. Solving two-stage robust optimization problems using a column-and-constraint generation method. *Operations Research Letters*, 41(5):457–461, 2013.