Robust Modeling and Optimization Review

Tu Ni

AY2018/2019 Semester 2

1 Math Programming Basics

1.1 Linear Programming

1.2 Convex Optimization

Some properties of convex functions:

- If $f_1(\boldsymbol{x})$ and $f_2(\boldsymbol{x})$ are convex functions, and $a, b \geq 0$ then $f(\boldsymbol{x}) = af_1(\boldsymbol{x}) + bf_2(\boldsymbol{x})$ is a convex function
- If $f(\boldsymbol{x})$ is a convex function and $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{y} + \boldsymbol{b}$, then $g(\boldsymbol{y}) = f(\boldsymbol{A}\boldsymbol{y} + \boldsymbol{b})$ is a convex function
- If $f_i(\boldsymbol{x})$ is a convex function then $h(\boldsymbol{x}) = \max_i f_i(\boldsymbol{x})$ is a convex function

The notion of epigraph connects convex function and convex set:

$$\mathrm{epi} f = \{(\boldsymbol{x}, y) \mid \boldsymbol{x} \in \mathcal{S}, f(\boldsymbol{x}) \leq y\}$$

then $f : \mathcal{S} \mapsto \mathbb{R}$ is convex if and only if epi$f$ is convex. For instance, we can prove

$$g(\boldsymbol{x}) = \ln\left(\sum_{i=1}^{n} p_i \exp(b_i x_i + c_i)\right) \; p_i > 0$$

is a convex function by checking the epigraph.

The combination of cone and convex set leads to convex cone such as norm cone or second-order cone $\{(\boldsymbol{x}, t) : \|\boldsymbol{x}\| \leq t\}$.

Some operations that preserve convexity:

- Intersection

- Affine function, such as projection $\{\boldsymbol{x} \mid \exists \boldsymbol{y} : (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{C}\}$ where $\mathcal{C}$ is convex set; hyperbolic cone $\{\boldsymbol{x} : \|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}\|_2 \leq \boldsymbol{c}'\boldsymbol{x} + d\}$.
- Perspective, $\mathcal{C}$ convex $\Leftrightarrow \{(\boldsymbol{x}, t) : \boldsymbol{x}/t \in \mathcal{C}, t > 0\}$ convex

Example 1.1. *$f(\boldsymbol{x}, \boldsymbol{y}) : \mathbb{R}^m \times \mathbb{R}^n$ is jointly convex function and $\mathcal{C}$ is a convex set. Suppose*

$$g(\boldsymbol{x}) = \min_{\boldsymbol{y} \in \mathcal{C}} f(\boldsymbol{x}, \boldsymbol{y})$$

in which the minimization is always achievable, then $g(\boldsymbol{x})$ is also a convex function.

Proof. We first transform the function to a feasibility problem by epigraph, then apply the property of projection. □

Example 1.2. *Suppose $f(\boldsymbol{x})$ is convex on $\boldsymbol{x} \in \mathbb{R}^n$, then*

$$g(x, y) = y f(\boldsymbol{x}/y)$$

is a convex function on $y > 0$.

Consider a constrained optimization problem

$$\min f(x) \quad s.t. \ x \in P$$

where $P = \{x : g_j(x) \leq 0, j = 1, \ldots, p, h_i(x) = 0, i = 1, \ldots, m\}$, the KKT condition derives the optimal condition.

Theorem 1.1. *Necessary condition of KKT conditions*

- *$\bar{\boldsymbol{x}}$ is local minimum of P*
- *$\mathcal{I} = \{j : g_j(\bar{\boldsymbol{x}}) = 0\}$, set of tight constraints*
- *Constraint qualification condition* (CQC) : *The vectors $\nabla g_j(\bar{\boldsymbol{x}}), j \in \mathcal{I}$ and $\nabla h_i(\bar{\boldsymbol{x}}), i = 1, \ldots, m$ are linearly independent*

Then, there exists vector $(\boldsymbol{u}, \boldsymbol{v})$:

1. $\nabla f(\bar{\boldsymbol{x}}) + \sum_{j=1}^{p} u_j \nabla g_j(\bar{\boldsymbol{x}}) + \sum_{i=1}^{m} v_j \nabla h_j(\bar{\boldsymbol{x}}) = 0$
2. $\boldsymbol{u} \geq \boldsymbol{0}$
3. $u_j g_j(\bar{\boldsymbol{x}}) = 0, j = 1, \ldots, p$

To discuss the duality under convex optimization setup, we need to use generalized inequalities. For example, when dealing with linear constraints in LP, we have

$$\boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b} \Leftrightarrow \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} \geq \boldsymbol{0} \Leftrightarrow \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} \in \mathcal{K}$$

where

$$\mathcal{K} = \mathbb{R}_+^m$$

Similarly, we can extend it to cone constraints

$$\boldsymbol{Ax} \succeq_{\mathcal{K}} \boldsymbol{b} \Leftrightarrow \boldsymbol{Ax} - \boldsymbol{b} \succeq_{\mathcal{K}} \boldsymbol{0} \Leftrightarrow \boldsymbol{Ax} - \boldsymbol{b} \in \mathcal{K}$$

Frequently, we will face two special cones: second-order cones

$$\mathcal{L}^{n+1} = \left\{ (x_0, \underbrace{x_1, \ldots, x_n}_{\boldsymbol{x}}) \,|\, \|\boldsymbol{x}\|_2 \leq x_0 \right\}$$

and symmetric positive semi-definite cone

$$\mathcal{S}_+^n = \{\boldsymbol{X} | \boldsymbol{X} \text{ symmetric positive semidefinite matrix } \}$$

Then, we introduce dual cone to characterize the conic duality.

$$\boldsymbol{K}^* = \left\{\boldsymbol{y} : \boldsymbol{y}^\top \boldsymbol{x} \geq 0 \forall \boldsymbol{x} \in \boldsymbol{K}\right\}$$

Consider $\boldsymbol{K} \subset \mathbb{R}^n$ a non-empty set and the set $\boldsymbol{K}_* = \left\{\boldsymbol{z} \in \mathbb{R}^n | \boldsymbol{z}^\top \boldsymbol{x} \geq 0, \forall \boldsymbol{x} \in \boldsymbol{K}\right\}$. Then we have that

1. $\boldsymbol{K}_*$ is a closed convex cone
2. If int$\boldsymbol{K} \neq \emptyset$ then $\boldsymbol{K}_*$ is pointed
3. If $\boldsymbol{K}$ is a closed convex pointed cone, then int$\boldsymbol{K}_* \neq \emptyset$
4. If $\boldsymbol{K}$ is a closed convex cone, then so is $\boldsymbol{K}_*$ and $(\boldsymbol{K}_*)_* = \boldsymbol{K}$. The three most common cones are self-dual.

Now we present the primal and dual under conic inequality

$$\begin{array}{rl} Z_1 = \inf & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} & \boldsymbol{Ax} \succeq_{\mathcal{K}} \boldsymbol{b} \end{array} \qquad \begin{array}{rl} Z_2 = \sup & \boldsymbol{b}^\top \boldsymbol{y} \\ \text{s.t.} & \boldsymbol{Ay} = \boldsymbol{c} \\ & \boldsymbol{y} \succeq_{\mathcal{K}^*} \boldsymbol{0} \end{array}$$

Consider the conic problem (CP) and its dual problem (CD)

1. The dual to (CD) is equivalent to (CP)
2. For any $\boldsymbol{x}$ feasible for (CP) and $\boldsymbol{y}$ feasible for (CD) we have that $\boldsymbol{c}\top\boldsymbol{x} \geq \boldsymbol{b}\top\boldsymbol{y}$
3. If (CP) is bounded below and $\boldsymbol{Ax} - \boldsymbol{b} \in \text{int}\mathcal{K}$ for some $\boldsymbol{x}$, then (CD) is solvable and $Z_1 = Z_2$. (analogous result if (CD) is bounded above and strictly feasible).

4. If either (CP) or (CD) is bounded and strictly feasible; then any primal dual pair $(\boldsymbol{x}, \boldsymbol{y})$ is an optimal solution
   (a) if and only if $\boldsymbol{c}^\top \boldsymbol{x} = \boldsymbol{b} \top \boldsymbol{y}$
   (b) if and only if $\boldsymbol{y}^\top (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) = 0$

1.2.1 Second-order Cone Programming

The general model of SOCP is

$$\begin{array}{ll} \min & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{s.t.} & \|\boldsymbol{A}_i \boldsymbol{x} + \boldsymbol{b}_i\|_2 \leq \boldsymbol{c}_i^\top \boldsymbol{x} + d_i \quad i = 1, \ldots, N \end{array}$$

$$\boldsymbol{L}^{n+1} = \{(x_0, \boldsymbol{x}) | \|\boldsymbol{x}\|_2 \leq x_0\}$$

We can represent the constraints as

$$\|\boldsymbol{A}_i \boldsymbol{x} + \boldsymbol{b}_i\|_2 \leq \boldsymbol{c}_i^\top \boldsymbol{x} + d_i$$
$$\Updownarrow$$
$$\begin{bmatrix} \boldsymbol{c}_i^\top \boldsymbol{x} + d_i \\ \boldsymbol{A}_i \boldsymbol{x} + \boldsymbol{b}_i \end{bmatrix} \succeq \boldsymbol{L}^{n_i} \boldsymbol{0} \Leftrightarrow \begin{bmatrix} \boldsymbol{c}_i^\top \boldsymbol{x} \end{bmatrix} \succeq \boldsymbol{L}^{n_i} \begin{bmatrix} -d_i \\ -\boldsymbol{b}_i \end{bmatrix}$$

Moreover, an useful and common structure for second-order cone is

$$\boldsymbol{x}^\top \boldsymbol{x} \leq st(s, t \geq 0) \Leftrightarrow \left\| \begin{matrix} \boldsymbol{x} \\ (s-t)/2 \end{matrix} \right\|_2 \leq (s+t)/2 \Leftrightarrow \begin{bmatrix} (s+t)/2 \\ \boldsymbol{x} \\ (s-t)/2 \end{bmatrix} \succeq_{\boldsymbol{L}^{n+2}} 0$$

Many typical math programmings can be cast into SOCP

- Linear optimization: set $\boldsymbol{A} = \boldsymbol{0}, \boldsymbol{b} = \boldsymbol{0}$
- Convex quadratic: use the property that PSD $M = A'A$ and then transform the common structure

Recall the ambiguity set with generalized moments. If epigraph of $g(\boldsymbol{z})$ is second order cone representable, then we could replace $g(\boldsymbol{z})$ by $\boldsymbol{u}$ and add $g(\boldsymbol{z}) \leq \boldsymbol{u}$ to support. Similar idea is applicable in mean-variance ambiguity set.

1.2.2 Semidefinite Programming

The general model of SDP is

$$\begin{array}{ll} (P:) \min & \boldsymbol{C} \bullet \boldsymbol{X} \\ \text{s.t.} & \boldsymbol{A}_i \bullet \boldsymbol{X} = b_i i = 1, \ldots, m \\ & \boldsymbol{X} \succeq \boldsymbol{0} \end{array}$$

Where

$$\boldsymbol{X} \succeq \mathbf{0} \Leftrightarrow \boldsymbol{X} \in \mathcal{S}_{+}^{n}$$

$$\mathcal{S}_{+}^{n} = \{\boldsymbol{X} \in \mathbb{R}^{n\times n} : \boldsymbol{X}^{\top} = \boldsymbol{X}, \boldsymbol{X}\text{positive semidefinite}\}$$

Many typical math programmings can be cast into SDP

- Linear optimization: set $\boldsymbol{X}$ as the diagonal matrix and also represent $\boldsymbol{c}$, $\boldsymbol{A}_i$ as diagonal matrices.
- Convex quadratic: use the property of Schur complement, suppose $\boldsymbol{B} \succeq 0$

$$\begin{bmatrix} \boldsymbol{B} & \boldsymbol{C}' \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} \succeq \mathbf{0} \Leftrightarrow \boldsymbol{D} - \boldsymbol{C}\boldsymbol{B}^{-1}\boldsymbol{C}' \succeq \mathbf{0}$$

- Quadratic constraints in robust optimization: apply the S-lemma

$$\boldsymbol{z'}\boldsymbol{A}\boldsymbol{z} + 2\boldsymbol{b'}\boldsymbol{z} + c \geq 0 \quad \forall \boldsymbol{z} : \boldsymbol{z'}\boldsymbol{A}_1\boldsymbol{z} + 2\boldsymbol{b}_1'\boldsymbol{z} + c_1 \geq 0$$
$$\Updownarrow$$
$$\begin{bmatrix} c & \boldsymbol{b}' \\ \boldsymbol{b} & \boldsymbol{A} \end{bmatrix} - \boldsymbol{\tau} \begin{bmatrix} c_1 & \boldsymbol{b}_1' \\ \boldsymbol{b}_1 & \boldsymbol{A}_1 \end{bmatrix} \succeq \mathbf{0}, \tau \geq 0$$

- Positive polynomial: a polynomial $f(z)$ with even degree k is non-negative if and only if it is sum of squares, which means $f(z) = \boldsymbol{v}(z)'\boldsymbol{Q}\boldsymbol{v}(z)$ where $\boldsymbol{v}(z) = \left(1, z, z^2, \ldots, z^{k/2}\right)', \boldsymbol{Q} \succeq \mathbf{0}$ i.e., $\boldsymbol{Q} = \boldsymbol{L}'\boldsymbol{L}$

2 Satisficing & Robust Modeling

We first introduce uncertain LOP:

$$\begin{array}{rl} \min & c(z)'x \\ \text{s.t.} & A(z)x \geq b(z) \end{array}$$

where $A(z), b(z), c(z)$ are affine in z, and they can be generally expanded as:

$$A(z) = A^0 + \sum_{i\in[I_z]} A^i z_i$$
$$b(z) = b^0 + \sum_{i\in[I_z]} b^i z_i$$
$$c(z) = c^0 + \sum_{i\in[I_z]} c^i z_i$$

The impact of uncertainty may cause the objective is not attainable or even solution is not feasible.

2.1 Satisficing Modeling

The idea of satisficing modeling is to seek a good enough choice: specify a target objective and then find a solution that satisfy all constraints with high probability.

The first satisficing model is P-Model, which assumes a known distribution but generally leads to a non-linear optmization problem:

$$\begin{array}{ll} \max & \ln \mathbb{P}(\boldsymbol{A}(\tilde{\boldsymbol{z}})\boldsymbol{x} \geq \boldsymbol{b}(\tilde{\boldsymbol{z}})) \\ \text{s.t.} & \boldsymbol{x} \in \mathcal{X} \end{array}$$

Evaluating the objective is generally NP-hard because it's related to finding volume of polyhedron. Some models are tractable:

The second satisficing model is Chance Constrained Model, which is more common than P-Model:

$$\begin{array}{ll} \min & \boldsymbol{c}'\boldsymbol{x} \\ \text{s.t.} & \mathbb{P}[\boldsymbol{A}(\tilde{\boldsymbol{z}})\boldsymbol{x} \geq \boldsymbol{b}(\tilde{z})] \geq 1-\epsilon \\ & \boldsymbol{x} \in \mathcal{X} \end{array}$$

These two models are reasonable but hard to optimize, needless to say we do not know the distribution. In order to generalize the problem, let's introduce satisficing decision criterion.

First we define tolerance set as $\mathcal{T}(\boldsymbol{x}) = \{\boldsymbol{z} \in \mathcal{W} | \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} \geq \boldsymbol{b}(\boldsymbol{z})\}$. Then given a family of tolerance sets, $\mathcal{T}(x) \subseteq \mathcal{W}$, for $x \in \mathcal{X}$, a function $\nu : \mathbb{R}^N \mapsto \mathbb{R} \cup \{-\infty\}$ is a satisficing decision criterion if and only if it has the following two properties: For all $x, y \in \mathcal{X}$,

1. (Satisficing dominance) If $\mathcal{T}(y) \subseteq \mathcal{T}(x)$, then $\nu(x) \geq \nu(y)$
2. (Infeasibility) If $\mathcal{T}(x) = \emptyset$ then $\nu(x) = -\infty$

Theorem 2.1. *A function $\nu : \mathbb{R}^N \mapsto \mathbb{R} \cup \{-\infty\}$ is a satisficing decision criterion if and only if $\nu(x) = \max_{\alpha \in \mathcal{S}}\{\rho(\boldsymbol{\alpha}) | \boldsymbol{z} \in \mathcal{T}(\boldsymbol{x}), \forall \boldsymbol{z} \in \mathcal{U}(\boldsymbol{\alpha})\}$ for some function $\rho : \mathcal{S} \to \mathbb{R} \cup \{-\infty\}$ on domain $\mathcal{S} \subseteq \mathbb{R}^P$, for some P and for some family of nonempty uncertainty sets $\mathcal{U}(\boldsymbol{\alpha}) \subseteq \mathcal{W}$ defined for all $\boldsymbol{\alpha} \in \mathcal{S}$.*

With the help of theorem above, we can formulate a satisficing model as:

$$\begin{array}{lll} \max & \rho(\boldsymbol{\alpha}) & \\ \text{s.t.} & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} \geq \boldsymbol{b}(\boldsymbol{z}) & \forall \boldsymbol{z} \in \mathcal{U}(\boldsymbol{\alpha}) \\ & \boldsymbol{x} \in \mathcal{X} & \\ & \boldsymbol{\alpha} \in \mathcal{S} & \end{array}$$

Generally it's a difficult problem and we can look at a simple case:

Example 2.1. *Assume $\mathcal{S} = \mathbb{R}_+, \rho(\alpha) = \alpha$, and the uncertainty set $\mathcal{U}(\alpha)$ is nondecreasing in α (size of uncertainty set), then*

$$\nu_R(\boldsymbol{x}) = \max_{\alpha \geq 0}\{\alpha | A(\boldsymbol{z})\boldsymbol{x} \geq \boldsymbol{b}(\boldsymbol{z}) \quad \forall \boldsymbol{z} \in \mathcal{U}(\alpha)\}$$

which characterizes the largest possible size of uncertainty set for which x would remain feasible.

The example above is Robustness Optimization, which is somewhat equivalent to P-Model. They both determine the most robust solutions that can withstand all perturbations from the largest possible uncertainty set.

2.2 Robust Optimization

In many practical cases, we desire to minimize cost, while ensuring solution remains feasible for given size of uncertainty set. This idea leads to the development of robust optimization:

$$\begin{array}{ll} \min & \boldsymbol{c}'\boldsymbol{x} \\ \text{s.t.} & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} \geq \boldsymbol{b}(\boldsymbol{z}) \quad \forall \boldsymbol{z} \in \mathcal{U}(r) \\ & \boldsymbol{x} \in \mathcal{X} \end{array}$$

In modern robust optimization, usually we also have uncertainty for objective functions and thus we consider the *minmax* instead of *min*, which means optimizing the worst-case performance.

$$\begin{array}{ll} \min & \max_{z \in \mathcal{U}(r)} \boldsymbol{c}(\boldsymbol{z})'\boldsymbol{x} \\ \text{s.t.} & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} \geq b(\boldsymbol{z}) \quad \forall \boldsymbol{z} \in \mathcal{U}(r) \\ & \boldsymbol{x} \in \mathcal{X} \end{array}$$

Moreover, we care about

- Convex uncertainty set (tractable and conic representable)
- Adjustable size to vary conservativeness
- Important to relate meaning to the size r

It's crucial to find reasonable r to avoid over-conservative. Similarly, in the setting of Robustness Optimization, we need to set the cost budget, which can be somehow calculated by binary search.

3 Robust Linear Optimization

3.1 The Price of Robustness

First we will put robust optimization in linear condition to understand some key ideas, and then extend to more complicated conditions. We define the robust counterpart as:

$$\begin{array}{ll}\min & \max_{z\in\mathcal{U}(r)} \boldsymbol{c}(\boldsymbol{z})'\boldsymbol{x} \\ \text{s.t.} & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} \geq b(\boldsymbol{z}) \quad \forall \boldsymbol{z} \in \mathcal{U}(r) \\ & \boldsymbol{x} \in \mathcal{X}\end{array}$$

which is equivalent to

$$\begin{array}{ll}\min & t \\ \text{s.t.} & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} \geq b(\boldsymbol{z}) \quad \forall \boldsymbol{z} \in \mathcal{U}(r) \\ & \boldsymbol{c}(\boldsymbol{z})'\boldsymbol{x} \leq t \quad \forall \boldsymbol{z} \in \mathcal{U}(r) \\ & \boldsymbol{x} \in \mathcal{X}\end{array}$$

Since we consider linear case, constraints can be split in rows. Moreover, we know $A(z), b(z), c(z)$ can be expanded in terms of z_i, so we have

$$\begin{gathered}\boldsymbol{a}_m(\boldsymbol{z})^\top \boldsymbol{x} \geq b_m(\boldsymbol{z}) \\ \Updownarrow \\ \sum_{i\in[I_z]} \underbrace{\left(b_m^i - \boldsymbol{e}_m^\top \boldsymbol{A}^i \boldsymbol{x}\right)}_{=\boldsymbol{y}^m} z_i \leq \underbrace{\boldsymbol{e}_m^\top \boldsymbol{A}^0 \boldsymbol{x} - b_m^0}_{=t^m}\end{gathered}$$

Ignore index m for simplicity, focus on robust counterpart

$$\begin{gathered}\boldsymbol{y}^\top \boldsymbol{z} \leq t \quad \forall \boldsymbol{z} \in \mathcal{U}(r) \\ \Updownarrow \\ \max_{\boldsymbol{z}\in\mathcal{U}(r)} \boldsymbol{y}^\top \boldsymbol{z} \leq t\end{gathered}$$

We show some widely used uncertainty sets:

- Discrete scenarios with $\mathcal{U} = \left\{z^1, \ldots, z^S\right\}$, and it can be extended to convex combination
- Polyhedra with $\mathcal{U} = \{z | \exists u : Cz + Du \leq d\}$, equivalent to exponential extreme points
- Box with $\mathcal{U} = \{\boldsymbol{z} \mid \|\boldsymbol{z}\|_\infty \leq 1\}$
- 1-Norm with $\mathcal{U}(r) = \{\boldsymbol{z} \mid \|\boldsymbol{z}\|_1 \leq r\}$

Taking the dual of maximization problem in robust counterpart generates a minimization problem with the same optimal value because of strong duality.

What's more, in order to guarantee the minimum is less than some value, we only need to find a feasible solution. In this way, there is no need solving a sub-problem but building a linear programming problem as a whole.

$$
\begin{gathered}
\{(\boldsymbol{y}, t) | \boldsymbol{y}'\boldsymbol{z} \leq t \quad \forall \boldsymbol{z} \in \mathcal{U}\} \\
\Updownarrow \\
\left\{ (\boldsymbol{y}, t) \mid \begin{array}{c} \exists \boldsymbol{p} \geq \mathbf{0} \\ \boldsymbol{p}'\boldsymbol{d} \leq t \\ \boldsymbol{C}'\boldsymbol{p} = \boldsymbol{y} \\ \boldsymbol{D}'\boldsymbol{p} = \mathbf{0} \end{array} \right\}
\end{gathered}
$$

We introduce the notion "Budget of Uncertainty Set", which characterizes the price of robustness.

$$
\begin{aligned}
\mathcal{U}(r) &= \{\boldsymbol{z} | \|\boldsymbol{z}\|_\infty \leq 1, \|\boldsymbol{z}\|_1 \leq r\} \\
&= \left\{ \boldsymbol{z} \mid \begin{array}{c} \exists \boldsymbol{u} : \\ \boldsymbol{z} \leq \boldsymbol{u} \\ -\boldsymbol{z} \leq \boldsymbol{u} \\ \mathbf{0} \leq \boldsymbol{u} \leq \mathbf{1} \\ \mathbf{1}'\boldsymbol{u} \leq r \end{array} \right\}
\end{aligned}
$$

We know the optimal solution is from basic feasible solutions, so the parameter r above controls how many uncertain variables we allow to vary at the same time. In other words, it somehow removes the over-conservative worst case. Suppose data is symmetrically distributed, probability of every data to vary in the direction of violation is low. Even if uncertain data are correlated, as long as we could figure out underlying independent factors, the argument and result still hold.

Specifically, [1] discusses the problem by assuming $\boldsymbol{z}$ are independent, symmetric and bound random variables in $[-1, 1]$. It gives a probability guarantee of constraint violation:

$$
\mathbb{P}[\boldsymbol{y}'\tilde{\boldsymbol{z}} > t] \leq \frac{1}{2^{I_z}} \left\{ (1-\mu) \sum_{l=\lfloor \nu \rfloor}^{I_z} \begin{pmatrix} I_z \\ l \end{pmatrix} + \mu \sum_{l=\lfloor \nu \rfloor + 1}^{I_z} \begin{pmatrix} I_z \\ l \end{pmatrix} \right\} \approx 1 - \Phi\left(\frac{r-1}{\sqrt{n}}\right)
$$

For a fixed error bound, we can roughly get the relation of r and n as $r \approx C\sqrt{n}$. As the dimension of problem increases, the budget of uncertainty increases sub-linearly, which means we only need to sacrifice a little feasibility guarantee in exchange for near-optimality.

Let's explore the robust counterpart from another perspective. Given biaffine function $\phi : X \times Z \mapsto \mathbb{R}$, where X, Z are polyhedra. We have following two properties:

1. $\min_{\bar{\boldsymbol{z}} \in Z} \phi(\boldsymbol{x}, \bar{\boldsymbol{z}}) \leq \max_{\bar{\boldsymbol{x}} \in X} \phi(\bar{\boldsymbol{x}}, \boldsymbol{z})$
2. $\min_{\boldsymbol{z} \in Z} \max_{\boldsymbol{x} \in X} \phi(\boldsymbol{x}, \boldsymbol{z}) = \max_{\boldsymbol{x} \in X} \min_{\boldsymbol{z} \in Z} \phi(\boldsymbol{x}, \boldsymbol{z})$
3. $\max_{\boldsymbol{z} \in \cap Z_i} \boldsymbol{x}^\top \boldsymbol{z} = \min_{\boldsymbol{x}_i, i \in [I]} \left\{ \sum_{i \in [I]} \max_{\boldsymbol{z} \in Z_i} \boldsymbol{x}_i^\top \boldsymbol{z} \;\middle|\; \sum_{i \in [I]} \boldsymbol{x}_i = \boldsymbol{x} \right\}$

The third property is a direct application of first two, and we provide the proof sketch: First use $z_i \in Z_i, z_i = z$ to replace $z \in Z_i$, then we relax the equality constraints to construct *max-min* structure on LHS. Observing the biaffine function, it's valid to switch *max* and *min*. Finally re-arrange the RHS, and thus we conclude the infimum convolution for polyhedra.

With the knowledge above, we are able to deal with more complex uncertainty sets, such as the intersection of polyhedra. We still use the example from [1], and by the definition of dual norm, we have:

$$
\begin{gathered}
\boldsymbol{y}^\top \boldsymbol{z} \leq t \quad \forall \boldsymbol{z} \in \{\boldsymbol{z} \mid \|\boldsymbol{z}\|_\infty \leq 1, \|\boldsymbol{z}\|_1 \leq r\} \\
\Updownarrow \\
\max_{\boldsymbol{z}_1 : \|\boldsymbol{z}_1\|_\infty \leq 1} \boldsymbol{y}_1^\top \boldsymbol{z}_1 + \max_{\boldsymbol{z}_2 : \|\boldsymbol{z}_2\|_1 \leq r} \boldsymbol{y}_2^\top \boldsymbol{z}_2 \leq t \\
\boldsymbol{y}_1 + \boldsymbol{y}_2 = \boldsymbol{y} \\
\Updownarrow \\
\|\boldsymbol{y}_1\|_1 + r\|\boldsymbol{y}_2\|_\infty \leq t \\
\boldsymbol{y}_1 + \boldsymbol{y}_2 = \boldsymbol{y}
\end{gathered}
$$

It's straightforward to reformulate the problem and get a tractable LP:

$$
\begin{gathered}
\|\boldsymbol{y}_1\|_1 + r \|\boldsymbol{y}_2\|_\infty \leq t \\
\boldsymbol{y}_1 + \boldsymbol{y}_2 = \boldsymbol{y} \\
\Updownarrow \\
\|\boldsymbol{y}_1\|_1 + r \|\boldsymbol{y} - \boldsymbol{y}_1\|_\infty \leq t \\
\Updownarrow \\
1'\boldsymbol{s} + rt \leq t \\
t\mathbf{1} \geq \boldsymbol{y} - \boldsymbol{y}_1 \\
t\mathbf{1} \geq -\boldsymbol{y} + \boldsymbol{y}_1 \\
\boldsymbol{s} \geq \boldsymbol{y}_1 \\
\boldsymbol{s} \geq -\boldsymbol{y}_1
\end{gathered}
$$

Compared with the method in [1], this formulation requires more variables.

3.2 Robust Optimization Duality

Let's assume feasible and bounded uncertainty set, and revisit the original robust optimization:

$$
\begin{aligned}
\min_{\boldsymbol{x}} \quad & \max_{\boldsymbol{z}_0 \in Z_0} \boldsymbol{a}_0(\boldsymbol{z}_0)^\top \boldsymbol{x} \\
\text{s.t.} \quad & \max_{\boldsymbol{z}_m \in Z_m} \left\{ b_m(\boldsymbol{z}_m) - \boldsymbol{a}_m(\boldsymbol{z}_m)^\top \boldsymbol{x} \right\} \leq 0 \quad m \in [M]
\end{aligned}
$$

We treat the problem in a more straightforward manner, relaxing the constraints directly:

$$\min_{\boldsymbol{x}} \left\{ \max_{\boldsymbol{p} \geq \boldsymbol{0}, \boldsymbol{z}_m \in Z_m, m \in \{0,\ldots,M\}} \left\{ \boldsymbol{a}_0(\boldsymbol{z}_0)^\top \boldsymbol{x} + \sum_{m \in [M]} \left(b_m(\boldsymbol{z}_m) - \boldsymbol{a}_m(\boldsymbol{z}_m)^\top \boldsymbol{x} \right) p_m \right\} \right\}$$

All the maximum operators are aggregated so that we end up with *min-max* problem. Unfortunately, the objective function is not affine in $(p, \boldsymbol{z}_m)$ so we are not able to switch *min* and *max* directly.

For ease of exposition, we consider the following problem:

$$\max_{p \geq 0, \boldsymbol{z} \in Z} \left(b(\boldsymbol{z}) - \boldsymbol{a}(\boldsymbol{z})^\top \boldsymbol{x} \right) p$$

$$Z = \{ \boldsymbol{z} \in \mathbb{R}^I \mid \exists \boldsymbol{u} \ : \ \boldsymbol{C}\boldsymbol{z} + \boldsymbol{D}\boldsymbol{u} \leq \boldsymbol{d} \}$$

Then the perspective function plays an important role by change of variables: $\boldsymbol{z}$ to $\boldsymbol{z}/p$, $\boldsymbol{u}$ to $\boldsymbol{u}/p$ and solves the following problem instead:

$$\max_{p \geq 0, (p, \boldsymbol{z}) \in \bar{Z}} \left(b(\boldsymbol{z}/p) - \boldsymbol{a}(\boldsymbol{z}/p)^\top \boldsymbol{x} \right) p$$

$$\bar{Z} = \{ (p, \boldsymbol{z}) \in \mathbb{R}^{I+1} \mid \exists \boldsymbol{u} \ : \ \boldsymbol{C}\boldsymbol{z} + \boldsymbol{D}\boldsymbol{u} \leq \boldsymbol{d}p \}$$

In this way, we transform the objective function to a biaffine function and thus can apply the technique above. Finally, we construct the dual of robust optimization and strong duality holds:

$$\begin{array}{ll} \max & \displaystyle\sum_{m \in [M]} b_m(\boldsymbol{z}_m/p_m) p_m \\ \text{s.t.} & \displaystyle\sum_{m \in [M]} \boldsymbol{a}_m(\boldsymbol{z}_m/p_m) p_m = \boldsymbol{a}_0(\boldsymbol{z}_0) \\ & \boldsymbol{z}_0 \in Z_0 \\ & \boldsymbol{p} \geq \boldsymbol{0} \\ & (p_m, \boldsymbol{z}_m) \in \bar{Z}_m, m \in [M] \end{array}$$

Remark. A closer look at the perspective transformation leads to careful analysis:

$$(P1) \quad \begin{array}{ll} \max & pt + \boldsymbol{y}^\top \boldsymbol{z} p \\ \text{s.t.} & \boldsymbol{C}\boldsymbol{z} + \boldsymbol{D}\boldsymbol{u} \leq \boldsymbol{d} \\ & p \geq 0 \end{array}$$

versus

$$(P2) \quad \begin{array}{ll} \max & pt + \boldsymbol{y}^\top \boldsymbol{z} \\ \text{s.t.} & \boldsymbol{C}\boldsymbol{z} + \boldsymbol{D}\boldsymbol{u} \leq \boldsymbol{d}p \\ & p \geq 0 \end{array}$$

To verify the equivalence between $P1$ and $P2$, it's trivial for $p > 0$ but not $p = 0$. Specifically, if $p^* = 0$ in the optimal solution, we can argue that both of the corresponding objective values are either 0 or unbound at the same time.

Now we have the dual of robust optimization problem and we'd like to transform it back to get more intuition.

Lemma 3.1.

$$\bar{Z} = \{(p, \boldsymbol{z}) \in \mathbb{R}^{I+1} \mid \exists \boldsymbol{u} \ : \ \boldsymbol{C}\boldsymbol{z} + \boldsymbol{D}\boldsymbol{u} \leq \boldsymbol{d}p\}$$

If Z is bounded, there does not exist $\boldsymbol{z} \neq \boldsymbol{0}$ such that $(0, \boldsymbol{z}) \in \bar{Z}$.

Based on the lemma, for bounded polyhedra Z_m, we can develop following equivalence:

$$\begin{array}{ll} \max & \sum_{m\in[M]} b_m(\boldsymbol{z}_m/p_m)p_m \\ \text{s.t.} & \sum_{m\in[M]} \boldsymbol{a}_m(\boldsymbol{z}_m/p_m)p_m = \boldsymbol{a}_0(\boldsymbol{z}_0) \\ & \boldsymbol{p} \geq \boldsymbol{0} \\ & \boldsymbol{z}_0 \in Z_0 \\ & (p_m, \boldsymbol{z}_m) \in \bar{Z}_m, m \in [M] \end{array} \quad = \quad \begin{array}{ll} \max & \sum_{m\in[M]} b_m(\boldsymbol{z}_m)p_m \\ \text{s.t.} & \sum_{m\in[M]} \boldsymbol{a}_m(\boldsymbol{z}_m)p_m = \boldsymbol{a}_0(\boldsymbol{z}_0) \\ & \boldsymbol{p} \geq \boldsymbol{0} \\ & \boldsymbol{z}_0 \in Z_0 \\ & \boldsymbol{z}_m \in Z_m, m \in [M] \end{array}$$

On top of that, we reach the key conclusion: Primal Worst = Dual Best.

$$\begin{array}{lll} \min & \max_{\boldsymbol{z}_0\in Z_0} \boldsymbol{a}_0(\boldsymbol{z}_0)^\top \boldsymbol{x} & \\ \text{s.t.} & \boldsymbol{a}_m(\boldsymbol{z}_m)^\top \boldsymbol{x} \geq b_m(\boldsymbol{z}_m) & \forall \boldsymbol{z}_m \in Z_m, \\ & & m \in [M] \end{array} \quad = \quad \begin{array}{ll} \max & \sum_{m\in[M]} b_m(\boldsymbol{z}_m)p_m \\ \text{s.t.} & \sum_{m\in[M]} \boldsymbol{a}_m(\boldsymbol{z}_m)p_m = \boldsymbol{a}_0(\boldsymbol{z}_0) \\ & \boldsymbol{p} \geq \boldsymbol{0} \\ & \boldsymbol{z}_0 \in Z_0 \\ & \boldsymbol{z}_m \in Z_m, m \in [M] \end{array}$$

Actually, we can extend the result to unbounded polyhedra Z_m, which requires feasibility for some $\boldsymbol{p} > \boldsymbol{0}$. Then we can remove the case of $\boldsymbol{p} = \boldsymbol{0}$, and obtain the limit by convexity:

$$\begin{array}{ll} \max & \sum_{m\in[M]} b_m(\boldsymbol{z}_m/p_m)p_m \\ \text{s.t.} & \sum_{m\in[M]} \boldsymbol{a}_m(\boldsymbol{z}_m/p_m)p_m = \boldsymbol{a}_0(\boldsymbol{z}_0) \\ & \boldsymbol{p} \geq \boldsymbol{0} \\ & \boldsymbol{z}_0 \in Z_0 \\ & (p_m, \boldsymbol{z}_m) \in \bar{Z}_m, m \in [M] \end{array} \quad = \quad \begin{array}{ll} \sup & \sum_{m\in[M]} b_m(\boldsymbol{z}_m/p_m)p_m \\ \text{s.t.} & \sum_{m\in[M]} \boldsymbol{a}_m(\boldsymbol{z}_m/p_m)p_m = \boldsymbol{a}_0(\boldsymbol{z}_0) \\ & \boldsymbol{p} > \boldsymbol{0} \\ & \boldsymbol{z}_0 \in Z_0 \\ & (p_m, \boldsymbol{z}_m) \in \bar{Z}_m, m \in [M] \end{array}$$

Since we require that $\boldsymbol{p} > \boldsymbol{0}$ is feasible for the problem, the equivalence of perspective transformation is straightforward.

4 Distributionally Robust Optimization

4.1 Choice under Risk and Ambiguity

We first introduce the notion of choice under uncertainty that people always maximize their expected utility with following properties:

- Representation, $\mathbb{E}_{\mathbb{P}}\left[U(\tilde{r})\right]$
- $\tilde{\boldsymbol{r}} \sim \mathbb{P}$
- $\mathbb{E}_{\mathbb{P}}\left[\cdot\right]$ denotes taking expectation with respect to $\mathbb{P}$
- Utility function, $U : \mathbb{R} \mapsto \mathbb{R}$
- Increasing function, the higher the better
- Certainty equivalent, $\mathrm{CE}[\tilde{r}] = U^{-1}\left(\mathbb{E}_{\mathbb{P}}\left[U(\tilde{r})\right]\right)$

In decision-making process, people tend to be risk aversion, which means concave utility function and $U(\mathrm{CE}[\tilde{r}]) \leq U(E_{\mathbb{P}}[\tilde{r}])$. A more generalization of risk aversion in economics is convex preference, which says if $\tilde{r} \succeq \tilde{z}, \tilde{s} \succeq \tilde{z}$ then

$$\lambda\tilde{r} + (1-\lambda)\tilde{s} \succeq \tilde{z} \quad \forall \lambda \in [0,1]$$

Lemma 4.1. *Preference measure ρ is consistent with convex preference if and only if ρ is quasiconcave:*

$$\rho(\lambda\tilde{r} + (1-\lambda)\tilde{s}) \geq \min\{\rho(\tilde{r}), \rho(\tilde{s})\} \quad \forall \lambda \in [0,1]$$

We focus on concave utility and worst-case expectation over ambiguity set, and thus construct the objective

$$\inf_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[U(\tilde{r})\right]$$

which is quasi-concave, and also consistent with diverfication preference.

An alternative modeling approach is a convex combination of worst-case performance and best-case performance:

$$\rho_\alpha(\tilde{r}) = (1-\alpha)\inf_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[U(\tilde{r})\right] + \alpha \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[U(\tilde{r})\right]$$

but it does not satisfy convex preference in general.

4.2 Distributionally Robust Linear Optimization

We desire to optimize the worst-case performance and for ease of exposition, we consider minimize the worse-case dis-utility with the formulation:

$$\begin{array}{ll} \min & \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[U_0(\boldsymbol{a}_0(\tilde{\boldsymbol{z}})^\top \boldsymbol{x} + b_0(\tilde{\boldsymbol{z}}))\right] \\ \text{s.t.} & \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[U_m(\boldsymbol{a}_m(\boldsymbol{z})^\top \boldsymbol{x} + \ b_m(\tilde{\boldsymbol{z}}))\right] \le \tau_m \quad m \in [M] \\ & \boldsymbol{x} \in \mathcal{X} \end{array}$$

where $U_0, \ldots, U_M$ are concave utility functions. However, evaluating expected utility is typically hard even for simple utility functions, and tractability also highly depends on ambiguity set $\mathcal{F}$. We provide some tractable configurations of $\mathcal{F}$ to best of our knowledge:

1. Uncertain locations $\boldsymbol{z} \in \mathcal{Z}$
   - Ambiguity set

$$\mathcal{F} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^I) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{P}[\tilde{\boldsymbol{z}} \in \mathcal{Z}] = 1 \end{array} \right\}$$

   - It can be transformed to a robust optimization model without concern of utility

$$\sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[U(\boldsymbol{a}(\boldsymbol{z})^\top \boldsymbol{x} + b(\tilde{\boldsymbol{z}}))\right] \Leftrightarrow \max_{\boldsymbol{z}\in\mathcal{Z}} U(\boldsymbol{a}(\boldsymbol{z})^\top \boldsymbol{x} + b(\boldsymbol{z})) \le U(\tau)$$

$$\Updownarrow$$

$$\boldsymbol{a}(\boldsymbol{z})^\top \boldsymbol{x} + b(\boldsymbol{z}) \le \tau \quad \forall \boldsymbol{z} \in \mathcal{Z}$$

2. Known locations and scenarios $\tilde{s} \in [S]$ with ambiguous probabilities
   - Ambiguity set
   - Discrete scenarios $\tilde{s}$ that takes values in $[S]$, with $\mathbb{P}[\tilde{s} = s] = p_s$ and $\tilde{\boldsymbol{z}} = \hat{\boldsymbol{z}}^s$ when $\tilde{s} = s$, but probability $\boldsymbol{p}$ is uncertain
   - If polyhedral $\mathcal{P}$, say

$$\mathcal{P} = \left\{\boldsymbol{p} \mid \exists \boldsymbol{q} \ : \ \boldsymbol{R}\boldsymbol{p} + \boldsymbol{S}\boldsymbol{q} \le \boldsymbol{t}, \boldsymbol{1}^\top \boldsymbol{p} = 1, \boldsymbol{p} \ge 0\right\}$$

   Then we have

$$\sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[U(\boldsymbol{a}(\boldsymbol{z})^\top \boldsymbol{x} + b(\tilde{\boldsymbol{z}}))\right] \le \tau$$

$$\Updownarrow$$

$$\max_{\boldsymbol{p}\in\mathcal{P}} \sum_{s\in[S]} p_s U\left(\boldsymbol{a}(\hat{\boldsymbol{z}}^s)^\top \boldsymbol{x} + b(\hat{\boldsymbol{z}}^s)\right) \le \tau$$

and its robust counterpart

$$
\begin{aligned}
&\alpha + \boldsymbol{\beta}^\top \boldsymbol{t} \leq \tau \\
&\boldsymbol{R}^\top \boldsymbol{\beta} + \boldsymbol{1}^\top \alpha \geq \begin{pmatrix} U\left(\boldsymbol{a}(\hat{\boldsymbol{z}}^1)^\top \boldsymbol{x} + b(\hat{\boldsymbol{z}}^1)\right) \\ \vdots \\ U\left(\boldsymbol{a}(\hat{\boldsymbol{z}}^S)^\top \boldsymbol{x} + b(\hat{\boldsymbol{z}}^S)\right) \end{pmatrix} \\
&\boldsymbol{S}^\top \boldsymbol{\beta} = \boldsymbol{0} \\
&\boldsymbol{\beta} \geq \boldsymbol{0}, \alpha \in \mathbb{R}
\end{aligned}
$$

3. Known locations and scenarios with ambiguous probabilities and moments

   - Ambiguity set

$$
\mathcal{F} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^{I_z} \times [S]) \;\middle|\; \begin{array}{l} (\tilde{\boldsymbol{z}}, \tilde{s}) \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{g}(\tilde{\boldsymbol{z}})\right] = \boldsymbol{\sigma} \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} = \hat{\boldsymbol{z}}^s \mid \tilde{s} = s\right] = 1 \\ \mathbb{P}\left[\tilde{s} = s\right] = p_s \\ \boldsymbol{p} \in \mathcal{P}, \boldsymbol{\sigma} \in \mathcal{S} \end{array} \right\}
$$

   - On top of case 2, we have additional moments information such as means and covariance. In general, it can be $\mathbb{E}_{\mathbb{P}}\left[\boldsymbol{g}(\tilde{\boldsymbol{z}})\right] = \boldsymbol{\sigma}$, for some mapping function $\boldsymbol{g} : \mathbb{R}^{I_z} \mapsto \mathbb{R}^{I_u}$.
   - In the formulation, we also need moments constraint $\sum_{s \in [S]} p_s \boldsymbol{g}(\hat{\boldsymbol{z}}^s) = \boldsymbol{\sigma}$. If S is polyhedra as well as P, we could solve a linear programming. Consider the simplest case with linear utility and simplex probability support. Since the optimal solution must be BFS, we are able to get sparse probability distribution on worst-case solution.

4. Uncertain locations and scenarios with ambiguous probabilities

   - Ambiguity set

$$
\mathcal{F} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^{I_z} \times [S]) \;\middle|\; \begin{array}{l} (\tilde{\boldsymbol{z}}, \tilde{s}) \sim \mathbb{P} \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{Z}_s \mid \tilde{s} = s\right] = 1 \\ \mathbb{P}\left[\tilde{s} = s\right] = p_s \\ \boldsymbol{p} \in \mathcal{P} \end{array} \right\}
$$

   - Compared with case 2, under some scenario, we only know the support of $\tilde{z}$ instead of specific value, so we have to maximize over $\tilde{z} \in \ddagger_s$ in the formulation with linear utility

$$
\sup_{\mathbb{P} \in \mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{a}(\boldsymbol{z})^\top \boldsymbol{x} + b(\boldsymbol{z})\right] \Leftrightarrow \begin{array}{ll} \max & \displaystyle\sum_{s \in [S]} p_s \max_{\boldsymbol{z} \in \mathcal{Z}_s} \boldsymbol{a}(\boldsymbol{z})^\top \boldsymbol{x} + b(\boldsymbol{z}) \\ \text{s.t.} & \boldsymbol{p} \in \mathcal{P} \end{array}
$$

$$
\Updownarrow
$$

$$
\begin{array}{ll} \max & \displaystyle\sum_{s \in [S]} p_s \left(\boldsymbol{a}(\boldsymbol{z}_s)^\top \boldsymbol{x} + b(\boldsymbol{z}_s)\right) \\ \text{s.t.} & \boldsymbol{p} \in \mathcal{P} \\ & \boldsymbol{z}_s \in \mathcal{Z}_s, \forall s \in [S] \end{array}
$$

- Assume there exists $\boldsymbol{p} > 0, \boldsymbol{p} \in \mathcal{P}$, we can use perspective function to reformulate as linear programming. Similarly, the results can be extend to convex piecewise affine utility function.

5. Uncertain locations with ambiguous moments

- Ambiguity set

$$\mathcal{G} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^{I_z}) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{g}(\tilde{\boldsymbol{z}})\right] = \boldsymbol{\sigma} \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{Z}\right] = 1 \\ \text{for some } \boldsymbol{\sigma} \in \mathcal{S} \end{array} \right\}$$

- On top of case 1, we have additional moments information such as means and covariance. To explore some nice result, we first focus on mean such that $\boldsymbol{g}(\tilde{\boldsymbol{z}}) = \tilde{\boldsymbol{z}}$. By introducing probability measure, we have two equivalent formulations

$$\begin{array}{ll} \sup & \mathbb{E}_{\mathbb{P}}\left[U(\boldsymbol{a}(\tilde{\boldsymbol{z}})^\top \boldsymbol{x} + b(\tilde{\boldsymbol{z}}))\right] \\ \text{s.t.} & \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right] = \boldsymbol{\sigma} \\ & \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{Z}\right] = 1 \\ & \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^{I_z}), \tilde{\boldsymbol{z}} \sim \mathbb{P} \end{array} \quad \Leftrightarrow \quad \begin{array}{ll} \sup & \int_{\mathcal{Z}} U(\boldsymbol{a}(\boldsymbol{z})^\top \boldsymbol{x} + b(\boldsymbol{z}))\mathrm{d}\mu(\boldsymbol{z}) \\ \text{s.t.} & \int_{\mathcal{Z}} \boldsymbol{z}\mathrm{d}\mu(\boldsymbol{z}) = \boldsymbol{\sigma} \\ & \int_{\mathcal{Z}} \mathrm{d}\mu(\boldsymbol{z}) = 1 \\ & \mu \in \mathcal{M}_+(\mathbb{R}^{I_z}) \end{array}$$

Construct the dual with weak duality

$$\begin{array}{lll} Z_1 = & \sup & \mathbb{E}_{\mathbb{P}}\left[U(\boldsymbol{a}(\tilde{\boldsymbol{z}})^\top \boldsymbol{x} + b(\tilde{\boldsymbol{z}}))\right] \\ & \text{s.t.} & \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right] = \boldsymbol{\sigma} \\ & & \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{Z}\right] = 1 \\ & & \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^{I_z}), \tilde{\boldsymbol{z}} \sim \mathbb{P} \end{array} \qquad \begin{array}{lll} Z_2 = & \inf & \alpha + \boldsymbol{\beta}^\top \boldsymbol{\sigma} \\ & \text{s.t.} & \alpha + \boldsymbol{\beta}^\top \boldsymbol{z} \geq U(\boldsymbol{a}(\boldsymbol{z})^\top \boldsymbol{x} + b(\boldsymbol{z})) \quad \forall \boldsymbol{z} \in \mathcal{Z} \end{array}$$

In order to achieve strong duality, we use the robust optimization duality from previous section. Consider the case of convex piecewise affine utility function, we have

$$\begin{array}{lll} Z_2 = & \inf & \alpha + \boldsymbol{\beta}^\top \boldsymbol{\sigma} \\ & \text{s.t.} & \alpha + \boldsymbol{\beta}^\top \boldsymbol{z} \geq g_\ell(\boldsymbol{a}(\boldsymbol{z})^\top \boldsymbol{x} + b(\boldsymbol{z})) + h_\ell \quad \forall \boldsymbol{z} \in \mathcal{Z}, \ell \in [L] \end{array}$$

Recall: Primal Worst equals Dual Best

$$\begin{array}{lll} Z_3 = & \sup & \displaystyle\sum_{\ell \in [L]} \left(g_\ell(\boldsymbol{a}(\boldsymbol{z}_\ell)^\top \boldsymbol{x} + b(\boldsymbol{z}_\ell)) + h_\ell\right) p_\ell \\ & \text{s.t.} & \displaystyle\sum_{\ell \in [L]} p_\ell = 1 \\ & & \displaystyle\sum_{\ell \in [L]} \boldsymbol{z}_\ell p_\ell = \boldsymbol{\sigma} \\ & & \boldsymbol{p} > \boldsymbol{0} \\ & & \boldsymbol{z}_\ell \in \mathcal{Z} \ \forall \ell \end{array}$$

Remark. We need the assumption that σ in the interior of $\mathcal{Z}$ to guarantee $\boldsymbol{p} > \boldsymbol{0}$, and there exists a sequence of solutions $\boldsymbol{p}^k$, $\boldsymbol{z}_1^k, \ldots, \boldsymbol{z}_L^k$, for $k > 0$ that converge to the objective value Z_2, with $\mathbb{P}_k\left[\tilde{\boldsymbol{z}} = \boldsymbol{z}_\ell^k\right] = p_\ell^k \quad \forall \ell \in [L]$. Then we are able to prove the strong duality:

$$
\begin{aligned}
Z_1 &\leq Z_2 \\
&= \lim_{k\to\infty} \sum_{\ell\in[L]} \left(g_\ell(a(\boldsymbol{z}_\ell^k)^\top \boldsymbol{x} + b(\boldsymbol{z}_\ell^k)) + h_\ell\right) p_\ell^k \\
&\leq \lim_{k\to\infty} \sum_{\ell\in[L]} U\left(a(\boldsymbol{z}_\ell^k)^\top \boldsymbol{x} + b(\boldsymbol{z}_\ell^k)\right) p_\ell^k \\
&= \lim_{k\to\infty} \mathbb{E}_{\mathbb{P}_k}\left[U\left(a(\tilde{\boldsymbol{z}})^\top \boldsymbol{x} + b(\tilde{\boldsymbol{z}})\right)\right] \\
&\leq \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[U\left(a(\tilde{\boldsymbol{z}})^\top \boldsymbol{x} + b(\tilde{\boldsymbol{z}})\right)\right] \\
&= Z_1.
\end{aligned}
$$

- In many cases that the mean is not known but ambiguous with some support S, we can still derive the dual by switching *sup* and *inf*:

$$
\begin{array}{ll}
\sup & \mathbb{E}_{\mathbb{P}}\left[U(\boldsymbol{a}(\tilde{\boldsymbol{z}})^\top \boldsymbol{x} + b(\tilde{\boldsymbol{z}}))\right] \\
\text{s.t.} & \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right] = \boldsymbol{\sigma} \\
& \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{Z}\right] = 1 \\
& \boldsymbol{\sigma} \in \mathcal{S} \\
& \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^{I_z}), \tilde{\boldsymbol{z}} \sim \mathbb{P}
\end{array}
\quad \Leftrightarrow \quad
\begin{array}{ll}
\inf & \sup\limits_{\boldsymbol{\sigma}\in\mathcal{S}} \left\{\alpha + \boldsymbol{\beta}^\top \boldsymbol{\sigma}\right\} \\
\text{s.t.} & \alpha + \boldsymbol{\beta}^\top \boldsymbol{z} \geq U(\boldsymbol{a}(\boldsymbol{z})^\top \boldsymbol{x} + b(\boldsymbol{z})) \quad \forall \boldsymbol{z} \in \mathcal{Z}
\end{array}
$$

  Then it becomes a regular robust optimization problem.

- Consider more general moments beyond mean, we can introduce axillary random variables:

$$
\mathcal{F} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^{I_z} \times \mathbb{R}^{I_u}) \;\middle|\; \begin{array}{l} (\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{u}}\right] = \boldsymbol{\sigma} \\ \mathbb{P}\left[(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \bar{\mathcal{Z}}\right] = 1 \\ \text{for some } \boldsymbol{\sigma} \in \mathcal{S} \end{array} \right\}
$$

  where

$$
\bar{\mathcal{Z}} = \left\{(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \mathcal{Z} \times \mathbb{R}^{I_u} \mid \boldsymbol{u} = \boldsymbol{g}(\boldsymbol{z})\right\}
$$

  It can be a issue because $\bar{\mathcal{Z}}$ may not be a polyhedral or convex set. Polyderal moments can be used to approximate general moments, but may need exponential numbers so may not be practical. Examples like convex piece-wise affine

$$
g_i(\boldsymbol{z}) = \max_{k\in[K]} \left\{\boldsymbol{c}_i^{k\top} \boldsymbol{z} + d_i^k\right\} \quad \forall i \in [I_u]
$$

  or even representation via linear optimization

$$
g_i(\boldsymbol{z}) = \min\left\{\boldsymbol{c}_i^\top \boldsymbol{v} \mid \boldsymbol{D}_i \boldsymbol{v} + \boldsymbol{E}_i \boldsymbol{z} \geq \boldsymbol{f}_i\right\} \quad \forall i \in [I_u]
$$

- For convex generalized moments $\mathbb{E}_{\mathbb{P}}\left[\boldsymbol{g}(\tilde{\boldsymbol{z}})\right] \leq \boldsymbol{\sigma}$. its lifted ambiguity set will have corresponding support $\bar{\mathcal{Z}} = \left\{(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \mathcal{Z} \times \mathbb{R}^{I_u} \mid \boldsymbol{u} \geq \boldsymbol{g}(\boldsymbol{z})\right\}$. Fortunately, we can absorb lifted random variable $\tilde{\boldsymbol{u}}$ within $\tilde{\boldsymbol{z}}$ with adjusted support $\mathcal{Z}$ and $\mathcal{S}$. Hence, the ambiguity set can be generalized to the case of ambiguous mean.

6. Uncertain locations and scenarios with ambiguous probabilities and moments(Event-wise Ambiguity)

   - Ambiguity set

$$
\mathcal{F} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^{I_z} \times [S]) \;\middle|\; \begin{array}{ll} (\tilde{\boldsymbol{z}}, \tilde{s}) \sim \mathbb{P} & \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}} | \tilde{s} \in \mathcal{E}_k\right] = \boldsymbol{\sigma}_k & \forall k \in [K] \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{Z}_s | \tilde{s} = s\right] = 1 & \forall s \in [S] \\ \mathbb{P}\left[\tilde{s} = s\right] = p_s & \forall s \in [S] \\ \text{for some } \boldsymbol{p} \in \mathcal{P}, \boldsymbol{\sigma}_k \in \mathcal{Q}_k & \forall k \in [K] \end{array} \right\}
$$

   - In order to derive tractable reformulation, we assume $\mathcal{Z}_s$, $s \in [S]$, $\mathcal{Q}_k$, $k \in [K]$, $\mathcal{P}$ are polyhedra and $\mathcal{P} \subseteq \{\boldsymbol{p} > \boldsymbol{0}\}$. We first focus on fixed $\boldsymbol{p}, \boldsymbol{\sigma}_k$ and then optimize over them. The formulation

$$
\begin{array}{ll} \sup & \mathbb{E}_{\mathbb{P}}\left[U(\boldsymbol{a}(\tilde{\boldsymbol{z}})^\top \boldsymbol{x} + b(\tilde{\boldsymbol{z}}))\right] \\ \text{s.t.} & \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}} | \tilde{s} \in \mathcal{E}_k\right] = \boldsymbol{\sigma}_k \quad \forall k \in [K] \\ & \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{Z}_s | \tilde{s} = s\right] = 1 \quad \forall s \in [S] \\ & \mathbb{P}\left[\tilde{s} = s\right] = p_s \\ & \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^{I_z}), \tilde{\boldsymbol{z}} \sim \mathbb{P} \end{array}
$$

$$
\Updownarrow
$$

$$
\begin{array}{ll} \sup & \displaystyle\sum_{s \in [S]} \mathbb{E}_{\mathbb{P}_s}\left[U(\boldsymbol{a}(\tilde{\boldsymbol{z}})^\top \boldsymbol{x} + b(\tilde{\boldsymbol{z}}))\right] p_s \\ \text{s.t.} & \displaystyle\sum_{s \in \mathcal{E}_k} \mathbb{E}_{\mathbb{P}_s}\left[\tilde{\boldsymbol{z}}\right] p_s = q_k \boldsymbol{\sigma}_k \quad \forall k \in [K] \\ & \mathbb{P}_s\left[\tilde{\boldsymbol{z}} \in \mathcal{Z}_s\right] = 1 \quad \forall s \in [S] \\ & \mathbb{P}_s \in \mathcal{P}_0(\mathbb{R}^{I_z}), \tilde{\boldsymbol{z}} \sim \mathbb{P}_s \quad \forall s \in [S] \end{array}
$$

   where $q_k = \sum_{s \in \mathcal{E}_k} p_s$. Then we can derive its dual

$$
\begin{array}{ll} \inf & \displaystyle\sum_{s \in [S]} p_s \alpha_s + \sum_{k \in [K]} q_k \boldsymbol{\beta}_k^\top \boldsymbol{\sigma}_k \\ \text{s.t.} & \alpha_s + \displaystyle\sum_{k \in \mathcal{K}_s} \boldsymbol{\beta}_k^\top \boldsymbol{z} \geq U(\boldsymbol{a}(\boldsymbol{z})^\top \boldsymbol{x} + b(\boldsymbol{z})) \quad \forall \boldsymbol{z} \in \mathcal{Z}_s, s \in [S] \end{array}
$$

   where $\mathcal{K}_s = \{k \in [K] \mid s \in \mathcal{E}_k\}$.

Several extensions are developed to characterize practical issues:

- Mixture distribution: all the moments constraints are specific to each scenario

- Covariates as scenarios: using covariate information via regression
- K-means: partition the support set into clusters such that each set of cluster can be represented as polyhedral

Moreover, we'd like to highlight the Wasserstein ambiguity set. In the original definition, Wasserstein metric describes the cost of an optimal mass transportation plan:

$$d_W(\mathbb{P}, \mathbb{P}^\dagger) \triangleq \begin{array}{ll} \inf & \mathbb{E}_{\mathbb{P}}\left[\rho(\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{u}}^\dagger)\right] \\ \text{s.t.} & (\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{u}}^\dagger) \sim \mathbb{Q} \\ & \Pi_{\tilde{\boldsymbol{u}}}\mathbb{Q} = \mathbb{P} \\ & \Pi_{\tilde{\boldsymbol{u}}^\dagger}\mathbb{Q} = \mathbb{P}^\dagger \end{array}$$

with its corresponding ambiguity set

$$\mathcal{G}_W(\theta) = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathcal{U}) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{u}} \sim \mathbb{P} \\ d_W(\mathbb{P}, \mathbb{P}^\dagger) \leq \theta \end{array} \right\}$$

$$\mathbb{P}^\dagger\left[\tilde{\boldsymbol{u}}^\dagger = \hat{\boldsymbol{u}}_s\right] = 1/S,\ \forall s \in [S]$$

which can be reformulated as

$$\mathcal{F}_W(\theta) = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^{I_u} \times [S]\right) \;\middle|\; \begin{array}{ll} (\tilde{\boldsymbol{u}}, \tilde{s}) \sim \mathbb{P} & \\ \mathbb{E}_{\mathbb{P}}\left[\rho(\boldsymbol{u}, \hat{\boldsymbol{u}}_{\tilde{s}})\right] \leq \theta & \\ \mathbb{P}\left[\tilde{\boldsymbol{u}} \in \mathcal{U} \mid \tilde{s} = s\right] = 1 & \forall s \in [S] \\ \mathbb{P}\left[\tilde{s} = s\right] = \frac{1}{S} & \forall s \in [S] \end{array} \right\}$$

because of

$$\mathcal{G}_W(\theta) = \Pi_{\tilde{\boldsymbol{u}}}\mathcal{F}_W(\theta)$$

Finally, we can also characterize in the format of an event-wise ambiguity set

$$\mathcal{F}_W(\theta) = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^{I_u+1} \times [S]\right) \;\middle|\; \begin{array}{ll} ((\tilde{\boldsymbol{u}}, \tilde{v}), \tilde{s}) \sim \mathbb{P} & \\ \mathbb{E}_{\mathbb{P}}\left[\tilde{v} \mid \tilde{s} \in [S]\right] \leq \theta & \\ \mathbb{P}\left[(\tilde{\boldsymbol{u}}, \tilde{v}) \in \mathcal{Z}_s \mid \tilde{s} = s\right] = 1 & \forall s \in [S] \\ \mathbb{P}\left[\tilde{s} = s\right] = \frac{1}{S} & \forall s \in [S] \end{array} \right\},$$

where

$$\mathcal{Z}_s = \{(\boldsymbol{u}, v) \mid \boldsymbol{u} \in \mathcal{U}, v \geq \rho(\boldsymbol{u}, \hat{\boldsymbol{u}}_s)\}, s \in [S]$$

5 Risk Measures

In many cases, utility function is not a practical decision rule. Instead, people care about the minimal amount of cash which has to be added to a risky returns to make it acceptable, named as risk measure:

$$\mu[\tilde{r}] = \inf\{m \in \mathbb{R} \mid \tilde{r} + m \in \mathcal{A}\}$$

where $\mathcal{A}$ is an acceptance set of returns.

A functional $\mu : \mathcal{V} \mapsto \mathbb{R}$ is a (monetary) risk measure if and only if it satisfies:

- **Monotonicity:**
  For all $\tilde{r}, \tilde{s} \in \mathcal{V}$ such that $\tilde{r} \geq \tilde{s}$, then

$$\mu[\tilde{r}] \leq \mu[\tilde{s}].$$

  Note that $\tilde{r} \geq \tilde{s}$ here means state-wise dominance.

- **Translation Invariance:**
  For all $c \in \mathbb{R}$, $\mu[\tilde{r} + c] = \mu[\tilde{r}] - c$.

For instance, certainty equivalent with convex utility satisfies the conditions.

5.1 Value at Risk(VaR)

We define Value-at-Risk(VaR) as

$$\mathrm{VaR}_\epsilon[\tilde{r}] \triangleq \inf\left\{m \in \mathbb{R} \mid \mathbb{P}\left[\tilde{r} + m \geq 0\right] \geq 1 - \epsilon\right\}$$

with following properties

- It can be interpreted as the smallest capital requirement to make probability of augmented position high enough.
- It does not necessarily favor diversification.
- It is measured in three variables: the amount of potential loss, the probability of that amount of loss and the time frame.

If VaR is non-positive, it's equivalent to chance constraint.

$$\begin{gathered}\mathrm{VaR}_\epsilon[\boldsymbol{a}(\tilde{\boldsymbol{z}})'\boldsymbol{x} + b(\tilde{z})] \leq 0 \\ \Updownarrow \\ \mathbb{P}\left[\boldsymbol{a}(\tilde{\boldsymbol{z}})'\boldsymbol{x} + b(\tilde{z}) \geq 0\right] \geq 1 - \epsilon\end{gathered}$$

For this to be true, we need $\mathbb{P}\left[\tilde{r} + v \geq 0\right]$ with respect to v non-decreasing and right-continuous.

5.2 Conditional Value at Risk(CVaR)

We know VaR is generally difficult to solve because of its property, so we derive Conditional Value at Risk(CVaR), which takes an average of the losses exceeding the VaR value.

$$\mathrm{CVaR}^*_\epsilon[\tilde{r}] \triangleq \mathbb{E}_\mathbb{P}\left[-\tilde{r} \mid -\tilde{r} \geq \mathrm{VaR}_\epsilon(\tilde{r})\right] \geq \mathrm{VaR}_\epsilon(\tilde{r})$$

We first consider the discrete returns r_t, $t \in [T]$. Then if ϵT is an integer, we have

$$\begin{aligned}
\mathrm{CVaR}^*_\epsilon[\tilde{r}] &= \frac{1}{\epsilon T} \max_{\mathcal{S}\,:\ ,\mathcal{S}\subseteq[T],|\mathcal{S}|=\epsilon T} \sum_{t\in\mathcal{S}} -r_t \\
&= \frac{1}{\epsilon T} \max_{\boldsymbol{z}\in\{0,1\}^T,\boldsymbol{z}'\mathbf{1}=\epsilon T} \sum_{t\in[T]} -r_t z_t \\
&= \frac{1}{\epsilon T} \max_{\boldsymbol{z}\in[\mathbf{0},\mathbf{1}],\boldsymbol{z}'\mathbf{1}=\epsilon T} \sum_{t\in[T]} -r_t z_t
\end{aligned}$$

Taking the dual of that, we have

$$\begin{aligned}
\mathrm{CVaR}_\epsilon[\tilde{r}] = \quad &\min && s + \tfrac{1}{\epsilon T}\mathbf{1}'\boldsymbol{p} \\
&\text{s.t.} && s\mathbf{1} + \boldsymbol{p} \geq -\boldsymbol{r} \\
& && \boldsymbol{p} \geq \mathbf{0}
\end{aligned}$$

or equivalently

$$\mathrm{CVaR}_\epsilon[\tilde{r}] = \inf_s \left\{ s + \frac{1}{\epsilon}\frac{1}{T}\sum_{t\in[T]}(-r_t - s)^+ \right\}$$

where $(x)^+ = \max\{0, x\}$

From this expression, we can have a more common definition for CVaR in optimization:

$$\mathrm{CVaR}_\epsilon[\tilde{r}] \triangleq \inf_v \left\{ v + \frac{1}{\epsilon}\mathbb{E}_{\mathbb{P}}\left[(-\tilde{r} - v)^+\right] \right\}$$

It does not require ϵT as an integer and also works for continuous distribution of returns, so it's widely applied in robust optimization community.

Moreover, it provides the tightest upper bound for VaR:

$$\begin{aligned}
\mathrm{CVaR}_\epsilon[\tilde{r}] &= \inf_v \{v \mid \ \mathrm{CVaR}_\epsilon[\tilde{r}] \leq v\ \} \\
&= \inf_{v,s} \left\{v \mid (s-v) + \tfrac{1}{\epsilon}\mathbb{E}_{\mathbb{P}}\left[(-\tilde{r}-s)^+\right] \leq 0\right\} \\
&= \inf_{v,s} \left\{v \mid s + \tfrac{1}{\epsilon}\mathbb{E}_{\mathbb{P}}\left[(-\tilde{r}-v-s))^+)\right] \leq 0\ \right\} \\
&= \inf_{v,t\geq 0} \left\{v \mid \tfrac{1}{\epsilon}\mathbb{E}_{\mathbb{P}}\left[(-\tilde{r}-v+t))^+)\right] \leq t\ \right\} \\
&= \inf_{v,t>0} \left\{v \mid \ \mathbb{E}_{\mathbb{P}}\left[((-\tilde{r}-v)/t+1))^+)\right] \leq \epsilon\ \right\} \\
&\geq \inf_v \{v \mid \ \mathbb{P}\left[\tilde{r}+v \leq 0\right] \geq 1-\epsilon\ \} \\
&= \mathrm{VaR}_\epsilon[\tilde{r}]
\end{aligned}$$

The inequality comes from Markov inequality.

5.3 Convex Risk Measure

Recall that we mention diversification preference before, which is an important criteria. Here, we introduce convex risk measure as an equivalent interpretation.

Compared with normal risk measure, we require convexity that for all $\tilde{r}, \tilde{s} \in \mathcal{V}$, then

$$\mu(\lambda \tilde{r} + (1-\lambda)\tilde{s}) \leq \lambda \mu(\tilde{r}) + (1-\lambda)\mu(\tilde{s}).$$

We provide several examples:

- CVaR: $\mathrm{CVaR}_\epsilon[\tilde{r}] \triangleq \inf_v \left\{ v + \frac{1}{\epsilon}\mathbb{E}_\mathbb{P}\left[(-\tilde{r}-v)^+\right] \right\}$
- Worst-case CVaR: $\mathcal{F}\text{-}\mathrm{CVaR}_\epsilon[\tilde{r}] \triangleq \inf_v \left\{ v + \frac{1}{\epsilon}\sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_\mathbb{P}\left[(-\tilde{r}-v)^+\right] \right\}$
- Optimized Certainty Equivalent: $\mu^{\mathrm{OCE}}[\tilde{r}] = \inf_{v\in\mathbb{R}} \left\{ v + \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_\mathbb{P}\left[U(-\tilde{r}-v)\right] \right\}$
- Shortfall Risk Measure: $\mu^{\mathrm{SR}}[\tilde{r}] = \inf_{v\in\mathbb{R}} \left\{ v \mid \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_\mathbb{P}\left[U(-\tilde{r}-v)\right] \leq 0 \right\}$

5.4 Risk Measure under Ambiguity

From the view of robust optimization, we care about CVaR/VaR under ambiguity. For ease of exposition, risk measures are defined on losses instead of returns.

Theorem 5.1.

$$\mathcal{F}\text{-}\mathrm{VaR}_\epsilon[\boldsymbol{c}'\tilde{\boldsymbol{z}}] = \mathcal{F}\text{-}\mathrm{CVaR}_\epsilon[\boldsymbol{c}'\tilde{\boldsymbol{z}}]$$

for moment ambiguity set,

$$\mathcal{F} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^{I_z}) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_\mathbb{P}\left[\tilde{\boldsymbol{z}}\right] = \boldsymbol{\sigma} \\ \mathbb{P}\left[\tilde{\boldsymbol{z}} \in \mathcal{Z}\right] = 1 \end{array} \right\}$$

where

$$\mathcal{F}\text{-}\mathrm{VaR}_\epsilon[\tilde{s}] \triangleq \inf \left\{ v \in \mathbb{R} \mid \inf_{\mathbb{P}\in\mathcal{F}} \mathbb{P}\left[\tilde{s} \leq v\right] \geq 1-\epsilon \right\}$$

$$\mathcal{F}\text{-}\mathrm{CVaR}_\epsilon[\tilde{s}] \triangleq \inf_v \left\{ v + \frac{1}{\epsilon}\sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_\mathbb{P}\left[(\tilde{s}-v)^+\right] \right\}$$

Remark. The theorem does not hold for non linear disutility or ambiguity set with multiple scenarios.

Proof. We first change the representation such that

$$\mathcal{F}\text{-}\mathrm{VaR}_\epsilon[\boldsymbol{c}'\tilde{\boldsymbol{z}}] = \inf \left\{ v \in \mathbb{R} \mid \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{P}\left[\boldsymbol{c}'\tilde{\boldsymbol{z}} > v\right] \leq \epsilon \right\}$$

Then by the knowledge of last section, we have

$$\begin{array}{rll} \inf_{\mathbb{P}\in\mathcal{F}} \mathbb{P}\left[\boldsymbol{c}'\tilde{\boldsymbol{z}} > v\right] = & \inf & \alpha + \boldsymbol{\beta}'\boldsymbol{\sigma} \\ & \text{s.t.} & \alpha + \boldsymbol{\beta}'\boldsymbol{z} \geq 1 \quad \forall \boldsymbol{z} \in \mathcal{Z}, \boldsymbol{c}'\boldsymbol{z} > v \\ & & \alpha + \boldsymbol{\beta}'\boldsymbol{z} \geq 0 \quad \forall \boldsymbol{z} \in \mathcal{Z} \end{array}$$

Next we reformulate the last constraint by infimum convolution such that

$$
\begin{array}{rll}
\mathcal{F}\text{-VaR}_\epsilon[\boldsymbol{c}'\tilde{\boldsymbol{z}}] = & \inf & v \\
& \text{s.t.} & \alpha + \boldsymbol{\beta}'\boldsymbol{\sigma} \le \epsilon \\
& & \alpha + \boldsymbol{\beta}'\boldsymbol{z} \ge 0 \qquad \forall \boldsymbol{z} \in \mathcal{Z} \\
& & \alpha + \boldsymbol{\beta}'\boldsymbol{z} \ge 1 \qquad \forall \boldsymbol{z} \in \mathcal{Z}, \boldsymbol{c}'\boldsymbol{z} > v \\
= & \inf & v \\
& \text{s.t.} & \alpha + \boldsymbol{\beta}'\boldsymbol{\sigma} \le \epsilon \\
& & \alpha + \boldsymbol{\beta}'\boldsymbol{z} \ge 0 \qquad \forall \boldsymbol{z} \in \mathcal{Z} \\
& & \alpha + \inf\limits_{\boldsymbol{z} \in \mathcal{Z}} \{(\boldsymbol{\beta} - \boldsymbol{\beta}_2)'\boldsymbol{z}\} + \inf\limits_{\boldsymbol{z}:\boldsymbol{c}'\boldsymbol{z}>v} \{\boldsymbol{\beta}_2'\boldsymbol{z}\} \ge 1 \\
= & \inf & v \\
& \text{s.t.} & \alpha + \boldsymbol{\beta}'\boldsymbol{\sigma} \le \epsilon \\
& & \alpha + \boldsymbol{\beta}'\boldsymbol{z} \ge 0 \qquad \forall \boldsymbol{z} \in \mathcal{Z} \\
& & \alpha + \inf\limits_{\boldsymbol{z} \in \mathcal{Z}} \{(\boldsymbol{\beta} - \boldsymbol{c}p)'\boldsymbol{z}\} + vp \ge 1 \\
& & p > 0 \\
= & \inf & v \\
& \text{s.t.} & \alpha + \boldsymbol{\beta}'\boldsymbol{\sigma} \le \epsilon \\
& & \alpha + \boldsymbol{\beta}'\boldsymbol{z} \ge 0 \qquad \forall \boldsymbol{z} \in \mathcal{Z} \\
& & \alpha + \boldsymbol{\beta}'\boldsymbol{z} \ge 1 + \boldsymbol{c}'\boldsymbol{z}p - vp \qquad \forall \boldsymbol{z} \in \mathcal{Z} \\
& & p > 0 \\
= & \inf & v \\
& \text{s.t.} & \alpha + \boldsymbol{\beta}'\boldsymbol{\sigma} \le \epsilon/p \\
& & \alpha + \boldsymbol{\beta}'\boldsymbol{z} \ge 0 \qquad \forall \boldsymbol{z} \in \mathcal{Z} \\
& & \alpha + \boldsymbol{\beta}'\boldsymbol{z} \ge 1/p + \boldsymbol{c}'\boldsymbol{z} - v \qquad \forall \boldsymbol{z} \in \mathcal{Z} \\
& & p > 0
\end{array}
$$

The last reformulation is derived from perspective function over p.

Finally by replacing $\frac{1}{p} = r$ and $v = v + r$, we reach

$$
\begin{array}{rll}
\mathcal{F}\text{-VaR}_\epsilon[\boldsymbol{c}'\tilde{\boldsymbol{z}}] = & \inf & v + r \\
& \text{s.t.} & \frac{1}{\epsilon}(\alpha + \boldsymbol{\beta}'\boldsymbol{\sigma}) \le r \\
& & \alpha + \boldsymbol{\beta}'\boldsymbol{z} \ge 0 \qquad \forall \boldsymbol{z} \in \mathcal{Z} \\
& & \alpha + \boldsymbol{\beta}'\boldsymbol{z} \ \ge \boldsymbol{c}'\boldsymbol{z} \ - v \qquad \forall \boldsymbol{z} \in \mathcal{Z} \\
& & r \ge 0 \\
= & \inf & v + \frac{1}{\epsilon}(\alpha + \boldsymbol{\beta}'\boldsymbol{\sigma}) \\
& \text{s.t.} & \alpha + \boldsymbol{\beta}'\boldsymbol{z} \ge 0 \qquad \forall \boldsymbol{z} \in \mathcal{Z} \\
& & \alpha + \boldsymbol{\beta}'\boldsymbol{z} \ \ge \boldsymbol{c}'\boldsymbol{z} \ - v \qquad \forall \boldsymbol{z} \in \mathcal{Z} \\
= & \inf\limits_{v \in \mathbb{R}} & \left\{ v + \frac{1}{\epsilon} \sup\limits_{\mathbb{P} \in \mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[(\boldsymbol{c}'\tilde{\boldsymbol{z}} - v)^+\right] \right\}
\end{array}
$$

□

5.5 Satisficing Measures

Instead of setting parameter ϵ for VaR and CVaR, we may use target as an alternative risk measure. For instance, $\mathbb{P}[\tilde{r} \ge 0]$ denotes the probability of target attainment, which unfortunately does not favour diversification preference and can not be optimized in computation. Therefore, we need to introduce satisficing measure as a more rigorous and practical notion.

A functional $\rho : \mathcal{V} \mapsto [-\infty, \infty]$ is a satsificing measure if and only if

- Target satisficing: $\tilde{r} \geq 0 \Rightarrow \rho[\tilde{r}] = \infty$
- Target infeasibility: $\tilde{r} < 0 \Rightarrow \rho[\tilde{r}] = -\infty$
- Monotonicity: $\tilde{r} \geq \tilde{v} \Rightarrow \rho[\tilde{r}] \ \geq \rho[\tilde{v}]$
- Upper-semicontinuous: $\{\tilde{r} \in \mathcal{V} \mid \rho[\tilde{r}] \geq \alpha\}$ is closed for all α.

Theorem 5.2. *A functional $\rho : \mathcal{V} \mapsto [-\infty, \infty]$ is a satsificing measure if and only if*

$$\rho[\tilde{r}] = \sup\{\alpha \in \mathbb{R} \mid \mu_\alpha[\tilde{r}] \leq 0\}.$$

for some risk measure $\mu_\alpha : \mathcal{V} \mapsto [-\infty, \infty]$ satisfiying

- *Monotonicity: $\tilde{r} \geq \tilde{v} \Rightarrow \mu_\alpha[\tilde{r}] \leq \mu_\alpha[\tilde{v}]$*
- *Translation invariance: $\mu_\alpha[\tilde{r} + a] = \mu_\alpha[\tilde{r}] - a$*
- *Nondecreasing in $\alpha \in \mathbb{R}$*
- *Normalized: $\mu_\alpha[0] = 0$*
- *Lower-semicontinuity: $\{\tilde{r} \in \mathcal{V} \mid \mu_\alpha[\tilde{r}] \leq \tau\}$ is closed for all τ.*

Remark. For given ρ, $\mu_\alpha[\tilde{r}] = \inf\{\tau \ \in \mathbb{R} \mid \rho[\tilde{r} + \tau] \geq \alpha\}$

To be consistent with diversification preference, we also need satisfication measure to be quasi-concave, and corresponding risk measure to be convex.

We provide several examples of QSM:

- CVaR QSM: $\rho[\tilde{r}] = \sup\limits_{\alpha \in [0,1]} \{\alpha \mid \mathrm{CVaR}_{1-\alpha}[\tilde{r}] \leq 0\}$
  - $\rho_{\mathrm{CVaR}}[\tilde{r}] \leq \mathbb{P}\left[\tilde{r} \geq 0\right]$
  - The tightest QSM for probability measure we mentioned at the beginning of section
- Scale Invariant QSM: $\rho[\tilde{r}] = \sup_{\alpha,\beta} \left\{\alpha \in [0,1] \mid \beta \mathbb{E}_{\mathbb{P}}\left[U(\tilde{r}/\beta)\right] \geq \beta\alpha, \beta \geq 0\right\}$
  - $\rho[k\tilde{r}] = \rho[\tilde{r}]$ for all $k > 0$
  - Concave, nondecreasing utility, $U(\cdot)$ with $U(0) = 0$ and $U(r) < 0$ for all $r < 0$, $U(\infty) = 1$
  - $\rho[\tilde{r}] \leq \mathbb{P}\left[\tilde{r} \geq 0\right]$

- Generalized CVaR QSM, obtain CVaR QSM when $U(r) = \min\{r, 1\}$

- Scale Reciprocal QSM: $\rho[\tilde{r}] = \sup_{\alpha \geq 0} \{\alpha \mid \mathbb{E}_{\mathbb{P}}[U(-\alpha\tilde{r})] \leq 1\}$
  - Convex, nondecreasing, nonnegative disutility $U(\cdot)$ with $U(0) = 1$ and $U(r) > 1$ for all $r > 0$
  - $\rho[k\tilde{r}] = \frac{1}{k}\rho[\tilde{r}]$ for all $k > 0$
  - Probability Bound: for all $\theta > 0$, by Markov inequality

    $$\mathbb{P}[\tilde{r} < -\theta] = \mathbb{P}[U(-\alpha^*\tilde{r}) > U(\alpha^*\theta)] \leq \frac{\mathbb{E}_{\mathbb{P}}[U(-\alpha^*\tilde{r})]}{U(\alpha^*\theta)} \leq \frac{1}{U(\rho[\tilde{r}]\theta)}$$

    where $\rho[\tilde{r}] = \alpha^*$.
  - $U(\cdot)$ controls the probability bounds.

In order to optimize some decision variable $\boldsymbol{x}$ over QSM, we could apply binary search on ρ by iteratively solving small problem. Sometimes, according to QSM structure, it can be transformed to a single convex optimization problem.

6 Stochastic Programming

The basic structure of stochastic programming is two-stage linear programming problem in which we decide $\boldsymbol{x}$ first (Here-and-now), then wait for scenario to realize, and finally decide $\boldsymbol{y}$ (recourse) in response to scenario.

Mathematically, we induce first stage cost

$$\boldsymbol{c}'\boldsymbol{x} \quad s.t.\ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \geq 0$$

then have second stage cost

$$\boldsymbol{d}(s)'\boldsymbol{y} \quad s.t.\ \boldsymbol{T}(s)\boldsymbol{x} + \boldsymbol{Y}(s)\boldsymbol{y} \geq \boldsymbol{h}(\boldsymbol{s}), \boldsymbol{y} \geq 0$$

To solve stochastic programming, we rely on different underlying assumption for recourse. First we consider the fixed recourse, which means coefficients of recourse variables are constant.

$$\begin{array}{rll} f(\boldsymbol{x}, s) = & \min & \color{red}{\boldsymbol{d}'}\boldsymbol{y} \\ & \text{s.t.} & \boldsymbol{T}(s)\boldsymbol{x} + \color{red}{\boldsymbol{Y}}\boldsymbol{y} = \boldsymbol{h}(s) \\ & & \boldsymbol{y} \geq 0 \end{array}$$

Although the fixed recourse simplifies formulation, it's still possibly infeasible with given $\boldsymbol{x}$, so we introduce complete recourse to have feasibility requirement. For all $\boldsymbol{t}$ there exist $\boldsymbol{y}$ such that $\boldsymbol{Y}\boldsymbol{y} = \boldsymbol{t}, \boldsymbol{y} \geq \boldsymbol{0}$.

However, complete recourse is over-constrained because we only care about some $\boldsymbol{h}(s) - \boldsymbol{T}(s)\boldsymbol{x}$, instead of all $\boldsymbol{t}$. Hence, we come up with relatively complete recourse, which means for all $\boldsymbol{x} \in \mathcal{X}$, second stage problem would always be feasible for all scenarios. It's obvious that complete recourse implies relatively complete recourse, but converse is not true.

Typically, a stochastic programming can be formulated as

$$
\begin{aligned}
Z^* = \quad & \min \quad && \boldsymbol{c}'\boldsymbol{x} + \mathbb{E}_{\mathbb{P}}\left[f(\boldsymbol{x}, \tilde{s})\right] \\
& \text{s.t.} && \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\
& && \boldsymbol{x} \geq 0 \\
= \quad & \min && \boldsymbol{c}'\boldsymbol{x} + \sum_{s \in [S]} p_s \boldsymbol{d}(s)'\boldsymbol{y}(s) \\
& \text{s.t.} && \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\
& && \boldsymbol{T}(s)\boldsymbol{x} + \boldsymbol{Y}(s)\boldsymbol{y}(s) = \boldsymbol{h}(s) \quad && \forall s \in [S] \\
& && \boldsymbol{y}(s) \geq 0 && \forall s \in [S] \\
& && \boldsymbol{x} \geq 0
\end{aligned}
$$

Unfortunately, when encountering large amount of scenarios with high dimension $\boldsymbol{y}$, we have to deal with large number of decision variables and constraints at the same time. Consider the case of relatively complete recourse, by taking the dual second stage formulation, we remove the scenario uncertainty in the constraints and only need to enumerate over extreme points.

$$
\begin{aligned}
f(\boldsymbol{x}, s) = \quad & \min \quad && \boldsymbol{d}'\boldsymbol{y} \\
& \text{s.t.} && \boldsymbol{T}(s)\boldsymbol{x} + \boldsymbol{Y}\boldsymbol{y} = \boldsymbol{h}(s) \\
& && \boldsymbol{y} > 0 \\
= \quad & \max && (\boldsymbol{h}(s) - \boldsymbol{T}(s)\boldsymbol{x})'\boldsymbol{p} \\
& \text{s.t.} && \boldsymbol{Y}'\boldsymbol{p} \leq d \\
= \quad & \max_j && \{(\boldsymbol{h}(s) - \boldsymbol{T}(s)\boldsymbol{x})'\boldsymbol{p}_j\}
\end{aligned}
$$

Hence we get a reformulation with fewer variables but more constraints(can be exponential), which can be solved by Benders decomposition.

$$
\begin{aligned}
Z^* = \quad & \min \quad && \boldsymbol{c}^\top\boldsymbol{x} + \textstyle\sum_{s \in [S]} p_s f(\boldsymbol{x}, s) \\
& \text{s.t.} && \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\
& && \boldsymbol{x} \geq 0 \\
& \min && \boldsymbol{c}^\top\boldsymbol{x} + \textstyle\sum_{s \in [S]} p_s f_s \\
& \text{s.t.} && \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\
& && \boldsymbol{x} \geq 0 \\
& && (\boldsymbol{h}(s) - \boldsymbol{T}(s)\boldsymbol{x})^\top\boldsymbol{p}_j \leq f_s \quad]\forall j, s,
\end{aligned}
$$

Remark. The problem of stochastic programming is computational difficult only for two-stage, can be even harder for multi-stage.

7 Adaptive Robust Optimization

7.1 Static Recourse Adaption

Compared with stochastic programming, we care about the worst case objective of second stage over uncertain locations, instead of uncertain scenarios. We can formulate the problem as

$$\begin{array}{rll} Z^* = & \min & \boldsymbol{c}^\top \boldsymbol{x} + \sup\limits_{\boldsymbol{z}\in\mathcal{Z}} \left\{\boldsymbol{d}(\boldsymbol{z})^\top \boldsymbol{y}(\boldsymbol{z})\right\} \\ & \text{s.t.} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & & \boldsymbol{T}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{Y}(\boldsymbol{z})\boldsymbol{y}(\boldsymbol{z}) \geq \boldsymbol{h}(\boldsymbol{z}) \quad \forall \boldsymbol{z} \in \mathcal{Z} \\ & & \boldsymbol{x} \geq 0 \\ & & \boldsymbol{y} : \text{ measurable function} \end{array}$$

which is equivalent to

$$\begin{array}{rlll} Z^* = & \min & \boldsymbol{c}^\top \boldsymbol{x} + t & \\ & \text{s.t.} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} & \\ & & \boldsymbol{d}(\boldsymbol{z})^\top \boldsymbol{y}(\boldsymbol{z}) \leq t & \forall \boldsymbol{z} \in \mathcal{Z} \\ & & \boldsymbol{T}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{Y}(\boldsymbol{z})\boldsymbol{y}(\boldsymbol{z}) \geq \boldsymbol{h}(\boldsymbol{z}) & \forall \boldsymbol{z} \in \mathcal{Z} \\ & & \boldsymbol{x} \geq 0 & \\ & & \boldsymbol{y} : \text{ measurable function} & \end{array}$$

The structure is similar to regular robust optimization but has infinite number of variables along with infinite number of constraints. In this case, determining whether $\boldsymbol{x} \in \mathcal{X}$ is NP-hard so approximation is needed. One possible idea is to replace $\boldsymbol{y}$ by some function that leads to tractable formulation. However, we still want to use regular robust optimization for approximation, which denoted as Static Recourse Adaption.

$$\begin{array}{rll} Z_S^* = & \min & \boldsymbol{c}^\top \boldsymbol{x} + \sup\limits_{\boldsymbol{z}\in\mathcal{Z}} \{\boldsymbol{d}(\boldsymbol{z})^\top \boldsymbol{y}\} \\ & \text{s.t.} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & & \boldsymbol{T}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{Y}(\boldsymbol{z})\boldsymbol{y} \geq \boldsymbol{h}(\boldsymbol{z}) \quad \forall \boldsymbol{z} \in \mathcal{Z} \\ & & \boldsymbol{x} \geq 0, \\ & & \boldsymbol{y} \in \mathbb{R}^{I_y} \end{array}$$

Generally we have $Z_S^* \geq Z^*$, and in the worst case, the bound can be bad: $Z_S^* = \infty$, so we need to explore conditions when the bound is tight. Intuitively, if there exists worst case uncertain location, then the problem reduces to regular robust optimization surely.

$$\begin{aligned} \mathcal{X} &= \left\{ \boldsymbol{x} \;\middle|\; \begin{array}{l} \forall \boldsymbol{z} \in \mathcal{Z}, \exists \boldsymbol{y} : \\ \boldsymbol{T}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{Y}(\boldsymbol{z})\boldsymbol{y} \geq \boldsymbol{h}(\boldsymbol{z}) \end{array} \right\} \\ &= \left\{ \boldsymbol{x} \;\middle|\; \begin{array}{l} \exists \boldsymbol{y} : \\ \boldsymbol{T}(\boldsymbol{z}^\dagger)\boldsymbol{x} + \boldsymbol{Y}(\boldsymbol{z}^\dagger)\boldsymbol{y} \geq \boldsymbol{h}(\boldsymbol{z}^\dagger) \end{array} \right\} \end{aligned}$$

For example, $\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{d} + \boldsymbol{z}$, $\|\boldsymbol{z}\|_\infty \leq 1$.

We can also consider the complete recourse. In this case, if uncertainty set is bounded, then Z_S^* is finite if Z^* is finite. Moreover, we have following result

Theorem 7.1. *If there is only one recourse variable, then $Z_S^* = Z^*$ is finite.*

$$\begin{array}{rll} Z_S^* = & \min & \boldsymbol{c}^\top \boldsymbol{x} + dy \\ & \text{s.t.} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & & \boldsymbol{T}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{v}y \geq \boldsymbol{h}(\boldsymbol{z}) \quad \forall \boldsymbol{z} \in \mathcal{Z} \\ & & \boldsymbol{x} \geq 0 \\ & & y \in \mathbb{R} \end{array}$$

Proof. First, we observe that for the coefficient of recourse variable $\boldsymbol{v}$, its entries must share the same sign. W.L.O.G, we suppose the coefficient of recourse variable $\boldsymbol{v} = \mathbf{1}$. We also observe that $d \geq 0$, otherwise the problem is unbounded. Then we have

$$\begin{array}{rcl} y(\boldsymbol{z}) & \geq & [\boldsymbol{h}(\boldsymbol{z}) - \boldsymbol{T}(\boldsymbol{z})\boldsymbol{x}]_i, \forall i \in [I] \\ & = & \max_i [\boldsymbol{h}(\boldsymbol{z}) - \boldsymbol{T}(\boldsymbol{z})\boldsymbol{x}]_i \end{array}$$

The objective function of Z^* becomes

$$Z^* = \min \boldsymbol{c}'\boldsymbol{x} + d \sup_{\boldsymbol{z}} y(\boldsymbol{z}) = \min \boldsymbol{c}'\boldsymbol{x} + dy^*$$

where $y^* = \max_{\boldsymbol{z},i} [\boldsymbol{h}(\boldsymbol{z}) - \boldsymbol{T}(\boldsymbol{z})\boldsymbol{x}]_i$ Hence, $Z_S^* = Z^*$. □

7.2 Affine Recourse Adaption

We study the case when static recourse adaption is optimal, but with strong condition, then we consider more general cases in which Affine Recourse Adaption approximation is needed.

For simplicity, we can represent the problem in this way:

$$\begin{array}{rll} Z^* = & \min & \boldsymbol{c}^\top \boldsymbol{x} \\ & \text{s.t.} & \boldsymbol{x} \in \hat{\mathcal{X}} \\ & & \boldsymbol{x} \in \mathcal{X} \end{array}$$

where

$$\hat{\mathcal{X}} = \left\{ \boldsymbol{x} \;\middle|\; \begin{array}{l} \forall \boldsymbol{z} \in \mathcal{Z}, \exists \boldsymbol{y} : \\ \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{h}(\boldsymbol{z}) \end{array} \right\}$$

We can restrict $\boldsymbol{y}$ as affine function so that

$$\begin{array}{rll} Z_A^* = & \min & \boldsymbol{c}^\top \boldsymbol{x} \\ & \text{s.t.} & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}) \geq \boldsymbol{h}(\boldsymbol{z}) \quad \forall \boldsymbol{z} \in \mathcal{Z} \\ & & \boldsymbol{x} \in \mathcal{X} \\ & & \boldsymbol{y} \in \mathcal{L} \end{array}$$

where

$$\mathcal{L} = \left\{ \boldsymbol{y} : \mathbb{R}^{I_z} \mapsto \mathbb{R}^{J_y} \;\middle|\; \begin{array}{l} \boldsymbol{y}(\boldsymbol{z}) = \boldsymbol{y}^0 + \sum\limits_{i \in [I_z]} \boldsymbol{y}^i z_i \\ \text{for some } \boldsymbol{y}^0, \ldots, \boldsymbol{y}^{I_z} \in \mathbb{R}^{I_y} \end{array} \right\}$$

The explict formulation can be derived as

- express $\boldsymbol{h}(\boldsymbol{z}) - \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x}$ as $\boldsymbol{R}(\boldsymbol{x})\boldsymbol{z} + \boldsymbol{r}(\boldsymbol{x})$, where $\boldsymbol{R}(\boldsymbol{x})$ and $\boldsymbol{r}(\boldsymbol{x})$ are affine mapping of $\boldsymbol{x}$.
- express $\boldsymbol{y}(\boldsymbol{z}) = \boldsymbol{y}^0 + \boldsymbol{Y}\boldsymbol{z}$.

$$\begin{array}{rll} Z_A^* = & \min & \boldsymbol{c}^\top \boldsymbol{x} \\ & \text{s.t.} & \boldsymbol{B}\boldsymbol{y}^0 + \boldsymbol{B}\boldsymbol{Y}\boldsymbol{z} \geq \boldsymbol{R}(\boldsymbol{x})\boldsymbol{z} + \boldsymbol{r}(\boldsymbol{x}) \quad \forall \boldsymbol{z} \in \mathcal{Z} \\ & & \boldsymbol{x} \in \mathcal{X} \\ & & \boldsymbol{y}^0, \boldsymbol{Y} \text{ free variables} \end{array}$$

which is a classical robust optimization formulation and we omit the rest derivation.

In practice, ARA can be a good approximation:

Example 7.1.

$$\mathcal{X}_1 = \left\{ x \;\middle|\; \begin{array}{l} \exists y : \\ z \leq x + y \leq z \quad \forall z \in [0,1] \end{array} \right\}$$

$$\mathcal{X}_2 = \left\{ x \;\middle|\; \begin{array}{l} \forall z \in [0,1], \exists y : \\ z \leq x + y \leq z \end{array} \right\}$$

under affine recourse adaptation

$$\mathcal{X}_3 = \left\{ x \;\middle|\; \begin{array}{l} \exists y^0, y^1 : \\ z \leq x + y^0 + y^1 z \leq z \quad \forall z \in [0,1] \end{array} \right\}$$

we have

$$\mathcal{X}_3 = \mathcal{X}_2 = \mathbb{R}.$$

However, ARA can also be a bad approximation:

Example 7.2.

$$\mathcal{X}_1 = \left\{ x \;\middle|\; \begin{array}{l} \forall \boldsymbol{z} \in [-\mathbf{1}, \mathbf{1}], \exists y : \\ z_1 - z_2 \leq y + x \\ z_2 - z_1 \leq y + x \\ y + x \leq z_1 + z_2 + 2 \\ y + x \leq -z_1 - z_2 + 2 \end{array} \right\}$$

observe that

$$|z_1 - z_2| \leq -|z_1 + z_2| + 2 \quad \forall \boldsymbol{z} \in [-\mathbf{1}, \mathbf{1}]$$

for all $\boldsymbol{z} \in [-\mathbf{1}, \mathbf{1}]$*, we have*

$$y = |z_1 - z_2| - x$$

it's relative complete recourse, which means $\mathcal{X}_1 = \mathbb{R}$*.*

However, under the ARA, we have

$$\mathcal{X}_2 = \left\{ x \;\middle|\; \begin{array}{l} \exists y^0, y^1, y^2 : \forall \boldsymbol{z} \in [-\mathbf{1}, \mathbf{1}] \\ z_1 - z_2 \leq y^0 + y^1 z_1 + y^2 z_2 + x \\ z_2 - z_1 \leq y^0 + y^1 z_1 + y^2 z_2 + x \\ y^0 + y^1 z_1 + y^2 z_2 + x \leq z_1 + z_2 + 2 \\ y^0 + y^1 z_1 + y^2 z_2 + x \leq -z_1 - z_2 + 2 \end{array} \right\}$$

by checking the extreme point of $\boldsymbol{z}$*, we conclude* $\mathcal{X}_2 = \emptyset$*.*

From instance above, we know although sometimes ARA is a good approximation, it can be infeasible even in relatively complete recourse. Moreover, ARA only determines the first stage decision $\boldsymbol{x}$ and $\boldsymbol{y}$ will not be used for future decisions.

One way to enhance ARA is to better model the uncertainty set:

$$\mathcal{Z} = \{\boldsymbol{z} \mid \exists \boldsymbol{u} \ : \ \boldsymbol{C}\boldsymbol{z} + \boldsymbol{D}\boldsymbol{u} \leq \boldsymbol{d}\}$$

which can be lifted to

$$\bar{\mathcal{Z}} = \{(\boldsymbol{z}, \boldsymbol{u}) \mid \boldsymbol{C}\boldsymbol{z} + \boldsymbol{D}\boldsymbol{u} \leq \boldsymbol{d}\}$$

$$\mathcal{Z} = \Pi_{\boldsymbol{z}} \bar{\mathcal{Z}}$$

Then we define $\boldsymbol{y}$ as affine function not only to $\boldsymbol{z}$ but also to u

$$\begin{array}{rll} Z^*_{EA} = & \min & \boldsymbol{c}^\top \boldsymbol{x} \\ & \text{s.t.} & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}, \boldsymbol{u}) \geq \boldsymbol{h}(\boldsymbol{z}) \quad \forall (\boldsymbol{z}, \boldsymbol{u}) \in \bar{\mathcal{Z}} \\ & & \boldsymbol{x} \in \mathcal{X} \\ & & \boldsymbol{y} \in \bar{\mathcal{L}} \end{array}$$

where

$$\bar{\mathcal{L}} = \left\{ \boldsymbol{y} : \mathbb{R}^{I_z} \times \mathbb{R}^{I_u} \mapsto \mathbb{R}^{J_y} \;\middle|\; \begin{array}{l} \boldsymbol{y}(\boldsymbol{z}, \boldsymbol{u}) = \boldsymbol{y}^0 + \sum\limits_{i \in [I_z]} \boldsymbol{y}_z^i z_i + \sum\limits_{i \in [I_u]} \boldsymbol{y}_u^i u_i \\ \text{for some } \boldsymbol{y}^0, \boldsymbol{y}_z^1 \ldots, \boldsymbol{y}_z^{I_z}, \boldsymbol{y}_u^1 \ldots, \boldsymbol{y}_u^{I_u} \in \mathbb{R}^{I_y} \end{array} \right\}$$

Clearly, the approximation power is enhanced so we have

$$Z^*_S \geq Z^*_A \geq Z^*_{EA} \geq Z^*$$

Example 7.3. *Consider the uncertainty set*

$$\mathcal{Z} = \{\boldsymbol{z} \mid \boldsymbol{C}\boldsymbol{z} \leq \boldsymbol{d} \ \}$$

$$\hat{\mathcal{Z}} = \left\{(\boldsymbol{z}^+, \boldsymbol{z}^-) \middle| \boldsymbol{C}(\boldsymbol{z}^+ - \boldsymbol{z}^-) \leq \boldsymbol{d}, \boldsymbol{z}^+, \boldsymbol{z}^- \geq \boldsymbol{0}, {\boldsymbol{z}^+}'\boldsymbol{z}^- = 0 \right\}$$

since $\hat{\mathcal{Z}}$ is not convex, we derive the convex relaxation

$$\bar{\mathcal{Z}} = \left\{(\boldsymbol{z}^+, \boldsymbol{z}^-) \middle| \boldsymbol{C}(\boldsymbol{z}^+ - \boldsymbol{z}^-) \leq \boldsymbol{d}, \boldsymbol{z}^+, \boldsymbol{z}^- \geq \boldsymbol{0} \right\}$$

which leads to

$$\bar{\mathcal{L}} = \left\{ \boldsymbol{y} : \mathbb{R}^{2I_z} \mapsto \mathbb{R}^{J_y} \ \middle| \ \begin{array}{l} \boldsymbol{y}(\boldsymbol{z}^+, \boldsymbol{z}^-) = \boldsymbol{y}^0 + \sum\limits_{i \in [I_z]} \boldsymbol{y}_+^i z_i^+ + \sum\limits_{i \in [I_z]} \boldsymbol{y}_-^i z_i^- \\ \textit{for some } \boldsymbol{y}^0, \boldsymbol{y}_+^1 \ldots, \boldsymbol{y}_+^{I_z}, \boldsymbol{y}_-^1 \ldots, \boldsymbol{y}_-^{I_z} \in \mathbb{R}^{I_y} \end{array} \right\}$$

Another way to solve the issue is to remove recourse variables with hard constraints. Consider the constraint

$$\mathcal{X} = \{\boldsymbol{x} \mid \exists \boldsymbol{y}_1, \boldsymbol{y}_2 : \ \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}_1\boldsymbol{y}_1 + \boldsymbol{B}_2\boldsymbol{y}_2 \geq \boldsymbol{h}\}$$

from Fourier-Motzkin Elimination, there exists $\boldsymbol{U} \geq \boldsymbol{0}$ such that $\boldsymbol{U}\boldsymbol{B}_1 = \boldsymbol{0}$ and $\mathcal{X} = \{\boldsymbol{x} \mid \exists \boldsymbol{y}_2 : \ \boldsymbol{U}\boldsymbol{A}\boldsymbol{x} + \boldsymbol{U}\boldsymbol{B}_2\boldsymbol{y}_2 \geq \boldsymbol{U}\boldsymbol{h}\}$. Then we can apply ARA on new $\mathcal{X}$, which improves original ARA. In the other sense, we reduce more second-stage uncertainty but induce more constraints.

The last idea is led by LP duality. Focusing on the set $\hat{\mathcal{X}}$, we have

$$\hat{\mathcal{X}} = \left\{ \boldsymbol{x} \ \middle| \ \begin{array}{l} \forall \boldsymbol{z} \in \mathcal{Z}, \exists \boldsymbol{y} \ : \\ \boldsymbol{B}\boldsymbol{y} \ \geq \boldsymbol{R}(\boldsymbol{x})\boldsymbol{z} + \boldsymbol{r}(\boldsymbol{x}) \end{array} \right\},$$

which is equivalent to

$$\hat{\mathcal{X}} = \left\{ \boldsymbol{x} \middle| \sup_{\boldsymbol{z} \in \mathcal{Z}} f(\boldsymbol{x}, \boldsymbol{z}) \leq 0 \right\},$$

where

$$\begin{aligned} f(\boldsymbol{x}, \boldsymbol{z}) = \quad & \min \quad 0 \\ & \text{s.t.} \quad \boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{R}(\boldsymbol{x})\boldsymbol{z} + \boldsymbol{r}(\boldsymbol{x}) \\ & \qquad\ \ \boldsymbol{y} \ \text{ free} \\ = \quad & \max \quad (\boldsymbol{R}(\boldsymbol{x})\boldsymbol{z} + \boldsymbol{r}(\boldsymbol{x}))^\top \boldsymbol{w} \\ & \text{s.t.} \quad \boldsymbol{w} \in \mathcal{W} \end{aligned}$$

where

$$\mathcal{W} = \left\{\boldsymbol{w} \mid \boldsymbol{B}^\top \boldsymbol{w} = \boldsymbol{0}, \boldsymbol{w} \geq \boldsymbol{0}\right\}$$

ending up with

$$\begin{aligned} \hat{\mathcal{X}} \quad &= \quad \left\{ \boldsymbol{x} \middle| \sup_{\boldsymbol{z} \in \mathcal{Z}, \boldsymbol{w} \in \mathcal{W}} (\boldsymbol{R}(\boldsymbol{x})\boldsymbol{z} + \boldsymbol{r}(\boldsymbol{x}))^\top \boldsymbol{w} \leq 0 \right\} \\ &= \quad \left\{ \boldsymbol{x} \middle| \sup_{\boldsymbol{z} \in \mathcal{Z}, \boldsymbol{w} \in \mathcal{W}} (\boldsymbol{R}(\boldsymbol{x})\boldsymbol{z} + \boldsymbol{r}(\boldsymbol{x}))^\top \boldsymbol{w} \leq 0 \right\} \\ &= \quad \left\{ \boldsymbol{x} \middle| \sup_{\boldsymbol{w} \in \mathcal{W}} g(\boldsymbol{x}, \boldsymbol{w}) \leq 0 \right\} \end{aligned}$$

where

$$g(\boldsymbol{x},\boldsymbol{z}) = \begin{array}{ll} \max & (\boldsymbol{R}(\boldsymbol{x})\boldsymbol{z}+\boldsymbol{r}(\boldsymbol{x}))^\top \boldsymbol{w} \\ \text{s.t.} & \boldsymbol{z} \in \mathcal{Z}. \end{array}$$

Suppose $\mathcal{Z} = \{\boldsymbol{z} \mid \boldsymbol{D}\boldsymbol{z} \le \boldsymbol{\delta}, \boldsymbol{z} \ge \boldsymbol{0}\}$, then

$$g(\boldsymbol{x},\boldsymbol{z}) = \begin{array}{ll} \min & \boldsymbol{\delta'\lambda} + \boldsymbol{r(x)'w} \\ \text{s.t.} & \boldsymbol{D'\lambda} \ge \boldsymbol{R(x)'w} \\ & \boldsymbol{\lambda} \ge \boldsymbol{0} \end{array}$$

finally resulting

$$\begin{aligned} \hat{\mathcal{X}} &= \left\{ \boldsymbol{x} \,\middle|\, \sup_{\boldsymbol{w}\in\mathcal{W}} g(\boldsymbol{x},\boldsymbol{w}) \le 0 \right\} \\ &= \left\{ \boldsymbol{x} \,\middle|\, \begin{array}{l} \forall \boldsymbol{w} \in \mathcal{W}, \exists \boldsymbol{\lambda} \ge \boldsymbol{0} \;: \\ \boldsymbol{\delta}^\top \boldsymbol{\lambda} + \boldsymbol{r}(\boldsymbol{x})^\top \boldsymbol{w} \le 0 \\ \boldsymbol{D}^\top \boldsymbol{\lambda} \ge \boldsymbol{R}(\boldsymbol{x})^\top \boldsymbol{w} \end{array} \right\}. \end{aligned}$$

It is still a classical ARA problem, but if recourse problem is non-linear, we can only use dual ARA instead of primal ARA.

If encountering multi-period problem, we can simply divide all the periods into first-stage and other stages, then apply ARA to determine first-stage decision where all the other stages are concatenated.

7.3 Adaptive Distributional Robust Optimization

We can extend the result to adaptive distributional robust optimization and let's focus on uncertain locations with ambiguous moments.

$$Z^* = \begin{array}{lll} \min & \boldsymbol{c}^\top \boldsymbol{x} + \sup\limits_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{d}^\top \boldsymbol{y}(\tilde{\boldsymbol{z}})\right] & \\ \text{s.t.} & \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}) \ge \boldsymbol{h}(\boldsymbol{z}) & \forall \boldsymbol{z} \in \mathcal{Z} \\ & \boldsymbol{y} : \text{ measurable function} & \\ & \boldsymbol{x} \in \mathcal{X} & \end{array}$$

where

$$\mathcal{F} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^{I_z}) \,\middle|\, \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}] = \boldsymbol{\sigma} \\ \mathbb{P}[\tilde{\boldsymbol{z}} \in \mathcal{Z}] = 1 \\ \text{for some } \boldsymbol{\sigma} \in \mathcal{S} \end{array} \right\}$$

Alternatively, we can formulate as

$$Z^* = \begin{array}{ll} \min & \boldsymbol{c}^\top \boldsymbol{x} + \sup\limits_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{z}})] \\ \text{s.t.} & \boldsymbol{x} \in \mathcal{X} \end{array}$$

where

$$
\begin{aligned}
f(\boldsymbol{x}, \boldsymbol{z}) = \quad & \min \quad \boldsymbol{d}'\boldsymbol{y} \\
& \text{s.t.} \quad \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{h}(\boldsymbol{z}) \\
& \qquad\;\; \boldsymbol{y} \ \text{ free} \\
= \quad & \max \quad (\boldsymbol{h}(\boldsymbol{z}) - \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x})'\boldsymbol{w} \\
& \text{s.t.} \quad \boldsymbol{B}'\boldsymbol{w} = \boldsymbol{d} \\
& \qquad\;\; \boldsymbol{w} \geq \boldsymbol{0}.
\end{aligned}
$$

$f(\boldsymbol{x}, \boldsymbol{z})$ is piece-wise affine function so that we can apply previous technique to take the dual and finally get

$$
\begin{aligned}
Z^* = \quad & \min \quad \boldsymbol{c}^\top \boldsymbol{x} + \sup_{\mathbb{P} \in \mathcal{F}} \mathbb{E}_{\mathbb{P}} \left[\boldsymbol{d}^\top \boldsymbol{y}(\tilde{\boldsymbol{z}}) \right] \\
& \text{s.t.} \quad \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}) \geq \boldsymbol{h}(\boldsymbol{z}) && \forall \boldsymbol{z} \in \mathcal{Z} \\
& \qquad\;\; \boldsymbol{y} : \ \text{measurable function} \\
& \qquad\;\; \boldsymbol{x} \in \mathcal{X} \\
= \quad & \min \quad \boldsymbol{c}^\top \boldsymbol{x} + \sup_{\boldsymbol{\sigma} \in \mathcal{S}} \{\alpha + \boldsymbol{\beta}'\boldsymbol{\sigma}\} \\
& \text{s.t.} \quad \alpha + \boldsymbol{\beta}'\boldsymbol{z} \geq \boldsymbol{d}^\top \boldsymbol{y}(\boldsymbol{z}) && \forall \boldsymbol{z} \in \mathcal{Z} \\
& \qquad\;\; \boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{z}) \geq \boldsymbol{h}(\boldsymbol{z}) && \forall \boldsymbol{z} \in \mathcal{Z} \\
& \qquad\;\; \boldsymbol{y} : \ \text{measurable function} \\
& \qquad\;\; \boldsymbol{x} \in \mathcal{X}.
\end{aligned}
$$

Surprisingly, we reduce the problem to standard adaptive robust optimization and thus ARA can be applied.

8 Convex Robust Optimization

8.1 Min-Max Theorem in Convex Optimization

Theorem 8.1. *Let $\mathcal{X}$ be a compact and convex set, $\mathcal{Y}$ be a convex set. Given $f : \mathcal{X} \times \mathcal{Y} \mapsto \mathbb{R}$ with:*

$-f(\boldsymbol{x}, \cdot)$ upper semi-continuous and quasiconcave on $\mathcal{Y}, \forall \boldsymbol{x} \in \mathcal{X}$
$-f(\cdot, \boldsymbol{y})$ lower semi-continuous and quasiconvex on $\mathcal{X}, \forall \boldsymbol{y} \in \mathcal{Y}$

then we have

$$
\min_{\boldsymbol{x} \in \mathcal{X}} \max_{\boldsymbol{y} \in \mathcal{Y}} f(\boldsymbol{x}, \boldsymbol{y}) = \max_{\boldsymbol{y} \in \mathcal{Y}} \min_{\boldsymbol{x} \in \mathcal{X}} f(\boldsymbol{x}, \boldsymbol{y})
$$

Minimax saddle function is a special case with $f(\cdot, \boldsymbol{y})$ convex and $f(\boldsymbol{x}, \cdot)$ concave. We will show that minmax of saddle function can always transformed to minmax of biaffine function!! Results can also be extended to distributional ambiguity.

Suppose function $f(\boldsymbol{x}, \boldsymbol{z})$ (taking values in extend real line) is concave and upper-semi-continuous in $\boldsymbol{z}$ for all $\boldsymbol{x} \in \mathcal{X}$, then for all $\boldsymbol{x} \in \mathcal{X}$, we can construct the conjugate function

$$
-f(\boldsymbol{x}, \boldsymbol{z}) = \sup_{\boldsymbol{v}} \{\boldsymbol{z}'\boldsymbol{v} - g(\boldsymbol{x}, \boldsymbol{v})\}
$$

where

$$g(\boldsymbol{x}, \boldsymbol{v}) = \sup_{\boldsymbol{z}} \{\boldsymbol{z}'\boldsymbol{v} + f(\boldsymbol{x}, \boldsymbol{z})\}$$

which is jointly convex on $\boldsymbol{x}, \boldsymbol{v}$.

Then we can represent the robust optimization as

$$\inf_{\boldsymbol{x} \in \mathcal{X}} \sup_{\boldsymbol{z} \in \mathcal{Z}} f(\boldsymbol{x}, \boldsymbol{z}) \Leftrightarrow \inf_{\boldsymbol{x} \in \mathcal{X}} \sup_{\boldsymbol{z} \in \mathcal{Z}} \inf_{\boldsymbol{v}} \{-\boldsymbol{z}'\boldsymbol{v} + g(\boldsymbol{x}, \boldsymbol{v})\} \Leftrightarrow \inf_{(\boldsymbol{x}, \boldsymbol{v}, w) \in \bar{\mathcal{X}}} \sup_{\boldsymbol{z} \in \mathcal{Z}} \{-\boldsymbol{z}'\boldsymbol{v} + w\}$$

where

$$\bar{\mathcal{X}} = \{(\boldsymbol{x}, \boldsymbol{v}, w) \mid \boldsymbol{x} \in \mathcal{X}, g(\boldsymbol{x}, \boldsymbol{v}) \leq w\}.$$

8.2 Tractable Robust Counterpart

Let's explore the tractability of robust counterpart in convex representation. In order to derive solvable algorithm in polynomial time, we need to ensure the underlying feasibility problem can be solved efficiently. Therefore, the robust counterpart is tractable if and only if we could efficient check

- whether x is feasible in the robust counterpart
- if infeasible, isolate the scenarios that violate the robust counterpart

We first consider robust quadratic optimization, and thus we have the robust counterpart

$$\boldsymbol{x}'\boldsymbol{A}(\boldsymbol{z})'\boldsymbol{A}(\boldsymbol{z})\boldsymbol{x} + b(\boldsymbol{z})'\boldsymbol{x} + c(\boldsymbol{z}) \leq 0 \quad \forall \boldsymbol{z} \in \mathcal{W}$$

which is equivalent to

$$\max_{\boldsymbol{z} \in \mathcal{W}} \|\boldsymbol{y} + \boldsymbol{Y}\boldsymbol{z}\|_2^2 + v_0 + \boldsymbol{v}'\boldsymbol{z} \leq 0$$

Unfortunately, for general uncertainty set $\mathcal{W}$, the problem is intractable unless we have single ellipsoidal uncertainty set $\mathcal{E}(r) = \{\boldsymbol{z} | \|\boldsymbol{z}\|_2 \leq r\}$.

Lemma 8.2. *Suppose $\boldsymbol{z}'\boldsymbol{A}_1\boldsymbol{z} + 2\boldsymbol{b}_1'\boldsymbol{z} + c_1 > 0$ for some $\boldsymbol{z}$, then*

$$\underbrace{\boldsymbol{z}'\boldsymbol{A}_0\boldsymbol{z} + 2\boldsymbol{b}_0'\boldsymbol{z} + c_0}_{q_0(\boldsymbol{z})} \geq 0 \quad \forall \boldsymbol{z} : \underbrace{\boldsymbol{z}'\boldsymbol{A}_1\boldsymbol{z} + 2\boldsymbol{b}_1'\boldsymbol{z} + c_1}_{q_1(\boldsymbol{z})} \geq 0$$

if and only if there exists $\tau \geq 0$ such that

$$\begin{bmatrix} c_0 & \boldsymbol{b}_0' \\ \boldsymbol{b}_0 & \boldsymbol{A}_0 \end{bmatrix} - \tau \begin{bmatrix} c_1 & \boldsymbol{b}_1' \\ \boldsymbol{b}_1 & \boldsymbol{A}_1 \end{bmatrix} \succeq \boldsymbol{0}$$

Apply S-lemma above, we have $\exists \tau \geq 0$

$$\begin{bmatrix} -v_0 - \boldsymbol{y}'\boldsymbol{y} & -\left(\boldsymbol{Y}'\boldsymbol{y} + \frac{1}{2}\boldsymbol{v}\right)' \\ -\left(\boldsymbol{Y}'\boldsymbol{y} + \frac{1}{2}\boldsymbol{v}\right) & -\boldsymbol{Y}'\boldsymbol{Y} \end{bmatrix} - \tau \begin{bmatrix} r^2 & 0' \\ 0 & -\boldsymbol{I}_N \end{bmatrix} \succeq 0$$

which is equivalent to

$$\begin{bmatrix} -v_0 - \tau r^2 & -\frac{1}{2}\boldsymbol{v}' \\ -\frac{1}{2}v & \tau I_N \end{bmatrix} - \begin{bmatrix} \boldsymbol{y}' \\ \boldsymbol{Y}' \end{bmatrix} \boldsymbol{I}_m \begin{bmatrix} \boldsymbol{y}' \\ \boldsymbol{Y}' \end{bmatrix}' \succeq \mathbf{0}$$

Apply Schur complement, we have the SDP representation

$$\begin{bmatrix} \boldsymbol{I}_m & \boldsymbol{y} & \boldsymbol{Y} \\ \boldsymbol{y}' & -v_0 - \tau r^2 & -\frac{1}{2}\boldsymbol{v}' \\ \boldsymbol{Y}' & -\frac{1}{2}\boldsymbol{v} & \tau \boldsymbol{I}_N \end{bmatrix} \succeq \mathbf{0}$$

Similar to robust quadratic optimization, we can get tractable form for robust SOCP without RHS uncertainty as well.

9 Entropic Methods

We have tight and tractable results for DRO problems based on scenario-wise, moment based ambiguity sets. In practice, it's common that $\tilde{z}_j$ are independently distributed, so how can we deal with that in a better way?

9.1 Probability Inequality

Let's start with univariate random variable with known mean and variance

$$\mathcal{F} = \left\{ \mathbb{P} \mid \mathbb{E}_{\mathbb{P}}\left[\tilde{v}\right] = \mu, \mathbb{E}_{\mathbb{P}}\left[\tilde{v}^2\right] = \mu^2 + \sigma^2 \right\}$$

Consider the distribution ambiguity, by one-sided Chebyshev inequality, we have for all $\mathbb{P} \in \mathcal{F}$

$$\mathbb{P}\left[\tilde{v} > \mu + r\sigma\right] \leq \frac{1}{1 + r^2}$$

which is a tight bound such that the robust counterpart has a closed-form representation

$$\sup_{\mathbb{P} \in \mathcal{F}} \mathbb{P}\left[\tilde{v} > 0\right] = \begin{cases} 1 & \mu \geq 0 \\ \frac{1}{1 + (\mu/\sigma)^2} & \text{otherwise} \end{cases}$$

We also have similar result for multivariate from [2]. For all $\mathbb{P} \in \mathcal{F}$

$$\mathcal{F} = \left\{ \mathbb{P} \mid \mathbb{E}_{\mathbb{P}}\left[\tilde{\boldsymbol{z}}\right] = 0, \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{z}\boldsymbol{z}'\right] = \boldsymbol{\Sigma} \right\}$$

we have

$$\sup_{\mathbb{P} \in \mathcal{F}} \mathbb{P}\left[y_0 + \boldsymbol{y}'\tilde{\boldsymbol{z}} > 0\right] = \begin{cases} 1 & y_0 \geq 0 \\ \frac{1}{1 + \left(y_0 / \sqrt{\boldsymbol{y}'\boldsymbol{\Sigma}\boldsymbol{y}}\right)^2} & \text{otherwise} \end{cases}$$

In this way, we can represent the mean-variance chance constraint as

$$\sup_{\mathbb{P}\in\mathcal{F}} \mathbb{P}\left[y_0 + \boldsymbol{y}'\tilde{\boldsymbol{z}} > 0\right] \le \epsilon \Leftrightarrow y_0 + \sqrt{\frac{1-\epsilon}{\epsilon}}\sqrt{\boldsymbol{y}'\boldsymbol{\Sigma}\boldsymbol{y}} \le 0$$

which can be represented as standard robust optimization problem with associated uncertainty set

$$y_0 + \boldsymbol{y}'\boldsymbol{z} \le 0 \quad \forall \boldsymbol{z} \in \mathcal{U}(\epsilon)$$

where

$$\mathcal{U}(\epsilon) \triangleq \left\{ \boldsymbol{z} \middle| \left\| \boldsymbol{\Sigma}^{1/2}\boldsymbol{z} \right\|_2 \le \sqrt{\frac{1-\epsilon}{\epsilon}} \right\}$$

We'd like to know the property of this uncertainty set. Consider the case if the underlying distribution is normal $\tilde{\boldsymbol{z}} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$, then

$$y_0 + \Phi^{-1}(1-\epsilon)\sqrt{\boldsymbol{y}'\boldsymbol{\Sigma}\boldsymbol{y}} \le 0 \Leftrightarrow \mathbb{P}\left[y_0 + \boldsymbol{y}'\tilde{\boldsymbol{z}} > 0\right] < \epsilon$$

Consequently, the uncertainty set we derive before can be too conservative in this case, especially when ϵ is small.

9.2 Expected Excess

It's pretty usual to use CVaR to approximate chance constraints and we have to deal with expected excess

$$\sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[\left(y^0 + \boldsymbol{y}'\tilde{\boldsymbol{z}} - v\right)^+\right]$$

The tight result can be obtained by event-wise ambiguity set, but when $\tilde{z}_j$ are independently distributed, we can strengthen the bound despite it may not be tight.

Consider the moment generation functions $g_j(\theta) \triangleq \ln(\mathbb{E}_{\mathbb{P}}\left[\exp(\theta\tilde{z}_j)\right])$, which can be used to bound the expected excess as

$$\begin{array}{rl} & \mathbb{E}_{\mathbb{P}}\left[(y_0 + \boldsymbol{y}'\tilde{\boldsymbol{z}})^+\right] \\ \le & \inf_{\mu>0} \mu\mathbb{E}_{\mathbb{P}}\left[\exp\left((y_0 + \boldsymbol{y}'\tilde{\boldsymbol{z}})/\mu - 1\right)\right] \\ = & \inf_{\mu>0} \mu \exp\left(y_0/\mu - 1\right) \prod_{j=1}^N \mathbb{E}_{\mathbb{P}}\left[\exp\left(y_j\tilde{z}_j/\mu\right)\right] \\ = & \inf_{\mu>0} \mu \exp\left(y_0/\mu + \sum_{j=1}^N g_j\left(y_j/\mu\right) - 1\right) \end{array}$$

Unfortunately, calculating moment generation functions g_j is hard so we need approximation. Suppose $\tilde{z}_j$ has zero mean, then g_j is bounded by forward deviation and backward deviation as

$$g_j(\theta) \le \frac{\sigma_{fj}^2\theta^2}{2}, \forall\theta > 0; g_j(\theta) \le \frac{\sigma_{bj}^2\theta^2}{2}, \forall\theta < 0$$

where

$$\sigma_f = \sup_{\theta>0}\left\{\sqrt{2\frac{\ln(\mathbb{E}_{\mathbb{P}}\left[\exp(\theta\tilde{z})\right])}{\theta^2}}\right\} \geq \sigma, \sigma_b = \sup_{\theta>0}\left\{\sqrt{2\frac{\ln(\mathbb{E}_{\mathbb{P}}\left[\exp(-\theta\tilde{z})\right])}{\theta^2}}\right\} \geq \sigma$$

the equality holds for normal distribution.

Moreover, if $\tilde{z}$ is bounded and distributed in $[-\underline{z}, \overline{z}]$ $z, \overline{z} > 0$, then

$$\sigma_f(\tilde{z}) \leq \overline{\sigma}_f(\tilde{z}) = \frac{z+\overline{z}}{2}\sqrt{g\left(\frac{z-\overline{z}}{z+\overline{z}}\right)}$$

and

$$\sigma_b(\tilde{z}) \leq \overline{\sigma}_b(\tilde{z}) = \frac{z+\overline{z}}{2}\sqrt{g\left(\frac{\overline{z}-\underline{z}}{\overline{z}}\right)}$$

where

$$g(\mu) = 2\max_{s>0}\frac{\phi_\mu(s)-\mu s}{s^2}$$

and

$$\phi_\mu(s) = \ln\left(\frac{e^s+e^{-s}}{2}+\frac{e^s-e^{-s}}{2}\mu\right)$$

the bound is tight, which can be achieved by a two-point distribution.

Therefore, we can bound the expected excess as

$$\begin{aligned}
&\mathbb{E}_{\mathbb{P}}\left[(y_0+\boldsymbol{y}'\tilde{z})^+\right] \\
&\leq \inf_{\mu>0}\mu\exp\left(y_0/\mu+\sum_{j=1}^{N}g_j\left(y_j/\mu\right)-1\right) \\
&\leq \inf_{\mu>0}\mu\exp\left(y_0/\mu+\frac{\|\boldsymbol{u}\|_2^2}{2\mu^2}-1\right) \\
&= \pi\left(y_0,\boldsymbol{y}\right)
\end{aligned}$$

where

$$u_j = \max\left\{y_j p_j, -y_j q_j\right\}$$

Finally, we get the approximation of CVaR as

$$
\begin{aligned}
&\inf_v \left\{ v + \frac{1}{\epsilon} \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}} \left[(y_0 + \boldsymbol{y}'\tilde{\boldsymbol{z}} - v)^+ \right] \right\} \\
\leq &\min_v \left(v + \frac{\pi\left(y_0 - v, \boldsymbol{y}\right)}{\epsilon} \right) \\
= &y_0 + \min_v \left(v + \frac{\pi(-v, \boldsymbol{y})}{\epsilon} \right) \\
= &y_0 + \min_{v,\mu} \left(v + \frac{\frac{\mu}{e}\exp\left(\frac{-v}{\mu} + \frac{\|\boldsymbol{u}\|_2^2}{2\mu^2}\right)}{\epsilon} \right) \\
= &y_0 + \min_{\mu} \left(\frac{\|\boldsymbol{u}\|_2^2}{2\mu^2} - \mu \ln \epsilon \right) \\
= &y_0 + \sqrt{-2\ln\epsilon}\|\boldsymbol{u}\|_2
\end{aligned}
$$

Recall the case of normal distribution, we can construct the uncertainty set by letting

$$
\mathcal{F} = \left\{ \mathbb{P} \mid \tilde{\boldsymbol{z}} = \boldsymbol{\Sigma}^{1/2}\tilde{\boldsymbol{\zeta}}, \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{\zeta}\right] = \boldsymbol{0}, \zeta_i \text{ independent}, \sigma_{fi}, \sigma_{bi} \leq 1 \right\}
$$

and thus

$$
y_0 + \sqrt{-2\ln(\epsilon)}\sqrt{\boldsymbol{y}'\boldsymbol{\Sigma}\boldsymbol{y}} \leq 0 \Rightarrow \mathbb{P}\left[y_0 + \boldsymbol{y}'\tilde{\boldsymbol{z}} > 0\right] < \epsilon
$$

which has less conservative ambiguity.

9.3 Adaptive Distributional Robust Optimization

We consider the application in adaptive distributional robust optimization

$$
\begin{array}{ll}
\min & \boldsymbol{c}'\boldsymbol{x} + \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{d}'\boldsymbol{y}(\tilde{\boldsymbol{z}})\right] \\
\text{s.t.} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\
& \boldsymbol{T}(\tilde{\boldsymbol{z}})\boldsymbol{x} + \boldsymbol{Y}\boldsymbol{y}(\tilde{\boldsymbol{z}}) = \boldsymbol{h}(\tilde{\boldsymbol{z}}) \\
& \boldsymbol{x} \geq \boldsymbol{0} \\
& y_i(\tilde{\boldsymbol{z}}) \geq 0, \forall i
\end{array}
$$

We focus on the recourse $\boldsymbol{Y}$, and define

$$
\begin{array}{rll}
\overline{d}_j = & \min & \boldsymbol{d}'\boldsymbol{y} \\
& \text{s.t.} & \boldsymbol{Y}\boldsymbol{y} = \boldsymbol{0} \\
& & y_j = 1 \\
& & \boldsymbol{y} \geq \boldsymbol{0}
\end{array}
$$

for all $j \in \mathcal{J}$ if the problem is feasible(which is for sure under complete recourse). Then $\overline{d}_j \geq 0$ and finite with optimum solutions $\overline{\boldsymbol{y}}^j$.

Given the knowledge above, we can introduce deflected linear decision rule as

$$\boldsymbol{y}(\tilde{\boldsymbol{z}}) = \boldsymbol{y}^0 + \sum_{j\in[I_z]} \boldsymbol{y}^j \tilde{z}_j + \sum_{j\in\mathcal{J}} \overline{\boldsymbol{y}}^j \left(y_j^0 + \boldsymbol{y}_j' \tilde{\boldsymbol{z}}\right)^-$$

We need to check the feasibility of the problem:

first, since $\boldsymbol{Y}\overline{\boldsymbol{y}}^j = \boldsymbol{0}$, we only need following condition for each uncertainty variable

$$\begin{gathered} \boldsymbol{T}_j \boldsymbol{x} + \boldsymbol{Y}\boldsymbol{y}_j = \boldsymbol{h}_j \quad \forall j = 0, \dots, N \\ \Downarrow \\ \boldsymbol{T}(\tilde{\boldsymbol{z}})\boldsymbol{x} + \boldsymbol{Y}\boldsymbol{y}(\tilde{\boldsymbol{z}}) = \boldsymbol{h}(\tilde{\boldsymbol{z}}) \end{gathered}$$

second, we have to consider those $i \notin \mathcal{J}$

$$\begin{gathered} y_i^0 + \sum_{j\in[I_z]} y_i^j z_j \geq 0 \quad \forall z \in \mathcal{Z} \\ \Downarrow \\ y_i(\tilde{\boldsymbol{z}}) \geq 0 \end{gathered}$$

third, the objective function is also changed

$$\sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[\boldsymbol{d}'\boldsymbol{y}(\tilde{\boldsymbol{z}})\right] \leq \boldsymbol{d}'\boldsymbol{y}^0 + \sum_{i\in F} \overline{d}_i \sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[\left(-y_i^0 - \boldsymbol{y}_i'\tilde{\boldsymbol{z}}\right)^+\right]$$

finally we can use entropic expected excess for reformulation

$$\begin{aligned} Z_{DLDR} = \quad \min \quad & \boldsymbol{c}'\boldsymbol{x} + \boldsymbol{d}'\boldsymbol{y}^0 + \sum_{j\in\mathcal{J}} \overline{d}_j \pi\left(-y_j^0, -\boldsymbol{y}_j\right) \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{T}_i\boldsymbol{x} + \boldsymbol{Y}\boldsymbol{y}^i = \boldsymbol{h}_i \quad \forall i \in [I_z] \cup \{0\} \\ & y_i^0 + \textstyle\sum_{j=1}^N y_i^j z_j \geq 0 \quad \forall \boldsymbol{z} \in \mathcal{Z}, \forall i \notin \mathcal{J} \\ & \boldsymbol{x} \geq 0 \end{aligned}$$

10 Robust Combinatorial Optimization

References

[1] Dimitris Bertsimas and Melvyn Sim. "The price of robustness". In: *Operations research* 52.1 (2004), pp. 35–53.

[2] Ioana Popescu. "Robust mean-covariance solutions for stochastic optimization". In: *Operations Research* 55.1 (2007), pp. 98–112.