15.094J: Robust Modeling, Optimization, Computation

Lectures 11 & 12: Distributionally Robust Optimization (DRO)

Optimization

Optimization is about making good **decisions** in a **rigorous** way, often subject to constraints. Applications appear everywhere in science, mathematics and business:

- Managing a share portfolio
- Scheduling public transport
- Fitting a model to measured data
- Optimizing a supply chain
- Designing electronic circuits
- Choosing worker shift patterns
- Shaping aerodynamic components
- Recovering images from ray MRI data
- ...

Convex Optimization

$$\begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & x \in \mathrm{R}^d \\ & x \in X \end{aligned}$$

Any convex optimization problem can be written in the previous canonical form

The problem has several ingredients:

- The vector x collects the **decision variables**
- The space R^d is the **domain** of the decision variables
- The constraints set X describes **convex feasible region**
- The **objective function** endows a cost to each decision

Representing a Convex Optimization Problem

$$\begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & x \in \mathbb{R}^d \\ & \ell(x, p) \leq 0 \end{aligned}$$

Described by the following problem data:

- **Cost vector** c in $\mathbb{R}^d$
- **Constraint function** $\ell : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}$ (convex in the first argument)

$$X(p) = \left\{ x \in \mathbb{R}^d \ : \ \ell(x, p) \leq 0 \right\}$$

- **Parameter vector** p in $\mathbb{R}^n$

The parameter vector p must be estimated from historical data and is therefore almost invariable **uncertain**.

Convex Uncertain Optimization

Uncertain convex problem :

$$\begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & x \in \mathrm{R}^d \\ & \ell(x, \mathbf{p}) \leq 0 \end{aligned}$$

Notation: Uncertain quantities are denoted in bold

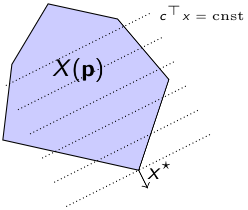
There are two settings:

1. Robust Setting : Take decision $x^\star$ **before** any $\mathbf{p}$ is observed
2. Adaptive Interpretation : Adapt decision $x^\star$ to uncertain value $\mathbf{p}$ **after** observing

Linear Uncertain Optimization

Uncertain linear problem :

$$\begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & x \in \mathbb{R}^d \\ & A(\mathbf{p})x \le b(\mathbf{p}) \end{aligned}$$



Assumption : **Technology matrix** $A(\mathbf{p})$ and **budget vector** $b(\mathbf{p})$ are affine functions.

Special Case of general convex optimization problem with **double affine** cost function $\ell(x, p) = \max_i \ell_i(x, p)$, with

$$\begin{aligned} \ell_i(x, \mathbf{p}) &= a_i(\mathbf{p})x - b_i(\mathbf{p}), \\ &= \tilde{a}_i(x)\mathbf{p} + \tilde{b}_i(x). \end{aligned}$$

Table of Contents

Describing Uncertainty

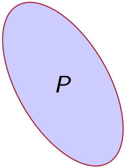
There are two fundamentally different approaches of how to describe an **uncertain quantity p**.

① **Set description**

- A set P describes all possible outcomes

$$\mathbf{p} \in P$$

- Directly related to *robust optimization*

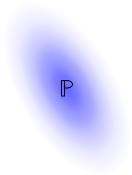


② **Probabilistic description**

- A distribution $\mathbb{P}$ describes all possible outcomes

$$\mathbf{p} \sim \mathbb{P}$$

- Directly related to *stochastic optimization*



Set Uncertainty

Uncertain Problem :

$$
\begin{aligned}
\min \quad & c^\top x \\
\text{s.t.} \quad & x \in \mathrm{R}^d \\
& \ell(x, \mathbf{p}) \leq 0,
\end{aligned}
$$

Robust optimization takes a **set description** perspective. A decision is only then feasible when feasible in the worst case

$$X = \cap_{p \in P} \left\{ x \in \mathrm{R}^d \ : \ \ell(x, p) \leq 0 \right\}.$$

Popular because

- **Convexity** of the feasible region is preserved
- Robust problem is as **tractable** as the problem $\max_{p \in P} \ \ell(x, p)$.

Robust Optimization

$$\begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & x \in \mathrm{R}^d \\ & \ell(x, p) \le 0, \quad \forall p \in P \end{aligned}$$

Exceptionally successful in a wide variety of challenging problems

- Engineering design
- Finance
- Machine learning
- Business analytics
- Systems operation and control
- ...

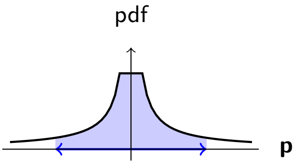
Set descriptions of uncertainty do not capture **distributional information**.

Distributional vs Adversarial Nature

Uncertain quantities sometimes have a **distributional**

- Stock prizes
- Mechanical loads
- Prediction errors
- ...

rather than an **adversarial nature**.



Set descriptions may lead to conservative decisions.

Probabilistic Uncertainty

Uncertain Problem :

$$\begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & x \in \mathrm{R}^d \\ & \ell(x, \mathbf{p}) \leq 0, \end{aligned}$$

Stochastic optimization takes a **probabilistic** perspective. A decision is only then feasible when feasible in expectation

$$X := \left\{ x \in \mathrm{R}^d \; : \; \mathbf{E}_{\mathbb{P}}\left[\ell(x, \mathbf{p})\right] \leq 0 \right\}.$$

The function ℓ should here be interpreted as how severely the decision x violates the constraints for a certain parameter realization p.

Constraint Violation Severity

$$X := \left\{ x \in \mathrm{R}^d \ : \ \mathbf{E}_{\mathbb{P}}\left[\ell(x, \mathbf{p})\right] \leq 0 \right\}.$$

Loss function ℓ is now treated as a **design parameter** measuring constraint **violation severity**.

Interesting particular cases

- For loss functions $\ell(x, p) \geq 0$, we get the **robust** constraint

$$X = \left\{ x \in \mathrm{R}^d \ : \ \cap_{p \in \operatorname{supp} \mathbb{P}} \left\{ x \in \mathrm{R}^d \ : \ \ell(x, p) \leq 0 \right\} \right\}.$$

- For loss functions $\ell(x, p) = \mathbb{1}\left\{x \notin X(p)\right\} - \alpha$, we get the **chance** constraint

$$X = \left\{ x \in \mathrm{R}^d \ : \ \mathbb{P}(x \notin X(\mathbf{p})) \leq \alpha \right\}$$

Stochastic Optimization

$$
\begin{aligned}
\min \quad & c^\top x \\
\text{s.t.} \quad & x \in \mathrm{R}^d \\
& \mathbf{E}_{\mathbb{P}}\left[\ell(x, \mathbf{p})\right] \leq 0
\end{aligned}
$$

Historically precedes robust optimization. Very **Limited success** in practical applications.

Two big hurdles to probabilistic approaches

1. Stochastic programming is often **not tractable**
2. Probability distributions are **never observed**

Stochastic Optimization – Tractability

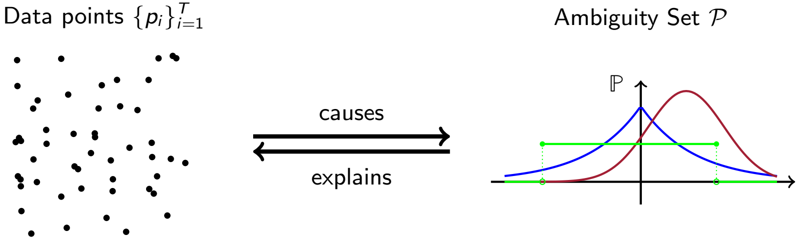
Feasibility of a fixed decision x requires

$$\mathbf{E}_{\mathbb{P}}\left[\ell(x, \mathbf{p})\right] \leq 0$$

- High-dimensional **integration is hard**

The problem of optimizing over a set for which feasibility is hard to check is generally even harder.

Probability distributions $\mathbb{P}$ are **never observed** directly. Rather in practice we have **data**.



To **estimate** a distribution $\mathbb{P}$ from data **exactly** would require an **infinite number** of samples.

Knight distinguishes two types of uncertainty

- **Risk :** Exposure to uncertain outcomes whose distribution is known
- **Ambiguity :** Exposure to uncertain outcomes without distributional information

Distributionally robust optimization attempts to marry stochastic and robust optimization.

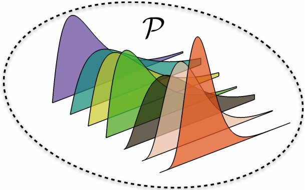
Uncertainty	No Risk	Risk
No Ambiguity	NO	SO
Ambiguity	RO	**DRO**

Knightian Uncertainty II

Knight distinguishes two types of uncertainty

- **Risk :** Exposure to uncertain outcomes whose distribution is known
- **Ambiguity :** Exposure to uncertain outcomes without distributional information

Distributionally robust optimization attempts to marry stochastic and robust optimization using an **ambiguity set** $\mathcal{P}$.



Knightian Uncertainty III

Uncertain problem :

$$\begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & x \in \mathrm{R}^d \\ & \ell(x, \mathbf{p}) \leq 0 \end{aligned}$$

DRO takes a **Knightian** perspective. A decision is only then feasible when feasible for any distribution in $\mathcal{P}$

$$X = \left\{ x \in \mathrm{R}^d \; : \; \mathbf{E}_{\mathbb{P}} \left[\ell(x, \mathbf{p}) \right] \leq 0, \quad \forall \mathbb{P} \in \mathcal{P} \right\}.$$

The best of both worlds ?

$$
\begin{aligned}
\min \quad & c^\top x \\
\text{s.t.} \quad & x \in \mathrm{R}^d \\
& \mathbf{E}_{\mathbb{P}}\left[\ell(x, \mathbf{p})\right] \leq 0, \quad \forall \mathbb{P} \in \mathcal{P}
\end{aligned}
$$

The ambiguity set $\mathcal{P}$ needs to be chosen well as to ensure **two objectives**

1. Uncertainty is well described
2. DRO is tractable

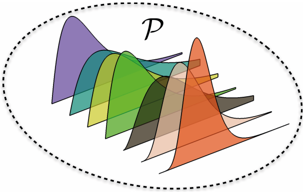
These two objectives can be **conflicting**.

Table of Contents

Ideal Ambiguity Sets

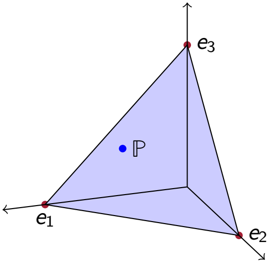
$$\mathbf{E}_{\mathbb{P}}\left[\ell(x, \mathbf{p})\right] \leq 0, \quad \forall \mathbb{P} \in \mathcal{P}$$

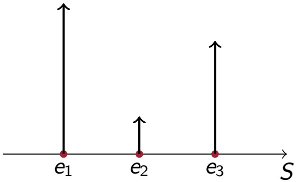
How should one construct an ambiguity set $\mathcal{P}$ for an uncertain parameter $\mathbf{p}$?



- The **ideal ambiguity set** is the smallest set of distributions such that $\mathbb{P}_{true} \in \mathcal{P}$.
- **Feasibility considerations** will demand the set $\mathcal{P}$ to have more structure.

Suppose **p** is completely uncertain besides taking value in Ω. The distribution of **p** can be any distribution in the **probability simplex** on Ω.





The probability simplex on Ω is in general

$$\mathcal{P}(\Omega) := \{\mathbb{P} \in \mathcal{M}_+(\Omega) \ : \ \mathbb{P}(\Omega) = 1\}.$$

For **finite sets** Ω the probability simplex is the **unit simplex**

$$\mathcal{P}(\Omega) := \left\{ \mathbb{P} \in \mathrm{R}^{|\Omega|} \ : \ \sum_{e \in S} \mathbb{P}(\mathrm{d}e) = 1,\ \mathbb{P}(e) \geq 0, \quad \forall e \in \Omega \right\}$$

which is a convex subset of $\mathrm{R}^{|\Omega|}$.

For **Borel measurable** subsets Ω of R^n, the probability simplex is a **Bauer simplex**

$$\mathcal{P}(\Omega) := \left\{ \mathbb{P} \in \mathcal{M}_+(\Omega) \ : \ \int_\Omega \mathbb{P}(\mathrm{d}e) = 1,\ \mathbb{P}(E) \geq 0, \quad \forall E \in \mathcal{B}(\Omega) \right\}$$

which is convex but not finite dimensional.

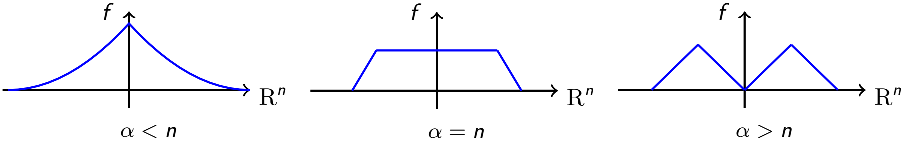
Structural information (unimodality, monotonicity) can be represented using simplices as well.

Simplicial Ambiguity Sets – Structured Distributions

α-**Unimodality :** If $\mathbb{P}$ on R^n has a continuous density function $f(e)$, then $\mathbb{P}$ is α-unimodal if

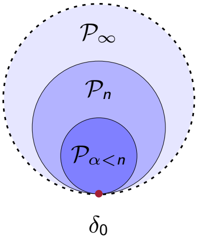
$$f(te)/t^{\alpha-n}, \quad \forall t > 0$$

is **non-increasing** in t for any $e \in \mathrm{R}^n$.



Unimodal simplex : $\mathcal{P}_\alpha(\mathrm{R}^n) = \{\mathbb{P} \in \mathcal{P}(\mathrm{R}^n) \ : \ \mathbb{P} \ \alpha\text{−unimodal}\}$
Special Cases:

- $\lim_{\alpha\to\infty} \mathcal{P}_\alpha(\mathrm{R}^n) = \mathcal{P}(\mathrm{R}^n)$
- $\mathcal{P}_0(\mathrm{R}^n) = \{\delta_0\}$

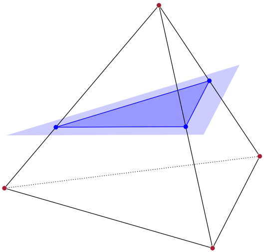


Moment Ambiguity Sets

Suppose **p** has **known moments**, that is,

$$\mathbf{E}_{\mathbb{P}}\left[g_j\right] = m_j, \quad \forall j.$$

Moments are linear functions of a probability distribution. **Moment sets are polytopic** subsets of the probability simplex.



For probability distributions on Ω in R^n we have

$$\mathbb{P} \in \left\{ \mathbb{P} \in \mathcal{P}(\Omega) \ : \ \int g_j(e)\mathbb{P}(\mathrm{d}e) = m_j, \ \ \forall j \right\}.$$

A common specific examples includes **known mean** μ and **known variance** Σ

$$\mathcal{P}(\mu, \Sigma) := \left\{ \mathbb{P} \in \mathcal{P}(\mathrm{R}^n) \ : \ \int e \, \mathbb{P}(\mathrm{d}e) = \mu, \ \int e \cdot e^\top \, \mathbb{P}(\mathrm{d}e) = \mu \cdot \mu^\top + \Sigma \right\}$$

which is an infinite dimensional polytope.

Observations

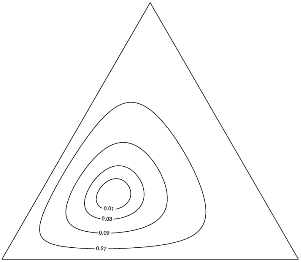
- The set $\mathcal{P}(\mu, \Sigma)$ is defined by n moment constraints for the mean and $n(n+1)/2$ moment constraints for the variance.
- Knowing the mean and second moment $S := \mu\mu^\top + \Sigma$ is equivalent.

Divergence Ambiguity Sets

Suppose **p** has a distribution which is not too different from a reference distribution $\hat{\mathbb{P}}$ as measured by some convex divergence metric d, i.e.

$$\mathcal{P}(\hat{\mathbb{P}}, r) := \left\{ \mathbb{P} \in \mathcal{P}(\Omega) \ : \ d(\mathbb{P}, \hat{\mathbb{P}}) \leq r \right\}$$

Divergence sets are convex **pseudo balls** in the probability simplex.



Very popular in the context of **sample data** $\{p_i\}_{i=1}^T$ in which case the **empirical distribution** is taken as a reference.

$$\hat{\mathbb{P}} = \frac{1}{T} \sum_{i=1}^{T} \delta_{p_i}$$

Divergence Ambiguity Sets II

Popular divergence metrics for finite dimensional S include

- **KL-divergence :**

$$\mathcal{P}_{\mathrm{KL}}(\hat{\mathbb{P}}, r) := \left\{ \mathbb{P} \in \mathcal{P}(\Omega) \ : \ \sum_{e \in S} \hat{\mathbb{P}}(e) \log\left(\frac{\hat{\mathbb{P}}(e)}{\mathbb{P}(e)}\right) \leq r \right\}$$

- **Pearson :**

$$\mathcal{P}_{\chi^2}(\hat{\mathbb{P}}, r) := \left\{ \mathbb{P} \in \mathcal{P}(\Omega) \ : \ \sum_{e \in S} \frac{(\mathbb{P}(e) - \hat{\mathbb{P}}(e))^2}{\hat{\mathbb{P}}(e)} \leq r \right\}$$

- **Kolmogorov-Smirnov :**

$$\mathcal{P}_{\mathrm{KS}}(\hat{\mathbb{P}}, r) := \left\{ \mathbb{P} \in \mathcal{P}(\Omega) \ : \ \max_{E \subseteq \Omega} \left| \mathbb{P}(E) - \hat{\mathbb{P}}(E) \right| \leq r \right\}$$

- **Wasserstein :**

$$\begin{aligned}\mathcal{P}_{\mathrm{WS}}(\hat{\mathbb{P}}, r) := \{ \mathbb{P} \in \mathcal{P}(\Omega) \ : \ \exists \mathbb{T} \in \mathcal{P}(\Omega \times \Omega), \ & \textstyle\sum_{e_1, e_2 \in S} \|e_1 - e_2\| \, \mathbb{T}(e_1, e_2) \leq r, \\ & \textstyle\sum_{e_1 \in S} \mathbb{T}(e_1, e_2) = \mathbb{P}(e_2), \quad \forall e_2 \in \Omega, \\ & \textstyle\sum_{e_2 \in S} \mathbb{T}(e_1, e_2) = \hat{\mathbb{P}}(e_1), \quad \forall e_1 \in \Omega \}\end{aligned}$$

The Feasibility Problem

$$\begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & x \in \mathrm{R}^d \\ & \mathbf{E}_{\mathbb{P}}\left[\ell(x, \mathbf{p})\right] \leq 0, \quad \forall \mathbb{P} \in \mathcal{P} \end{aligned}$$

Feasibility of a fixed decision $\bar{x}$ requires

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbf{E}_{\mathbb{P}}\left[\ell(x, \mathbf{p})\right] \leq 0$$

- Worst-case expectation problem
- Infinite dimensional (convex) optimization problem (when $\ell(\cdot, p)$ is convex)

Optimality requires robust counterpart of the convex set

$$X = \cap_{\mathbb{P} \in \mathcal{P}} \left\{ x \in \mathrm{R}^d \ : \ \mathbf{E}_{\mathbb{P}}\left[\ell(x, \mathbf{p})\right] \leq 0 \right\}$$

Table of Contents

Worst-case Expectation Problem

Problem : What is the worst-case expectation $\mathbf{E}_{\mathbb{P}}[\ell(x, \mathbf{p})]$ assuming only

- $\mathbf{p}$ takes values in a **finite set** Ω
- $\mathbf{p} \sim \mathbb{P}$ in convex ambiguity set $\mathcal{P}$

The problem results in an standard convex optimization problem

$$\begin{array}{rl} \text{inf / sup} & \sum_{e \in \Omega} \ell(x, e)\, \mathbb{P}(e) \\ \text{s.t.} & \mathbb{P} \in \mathrm{R}_+^{|\Omega|}, \\ & \sum_{e \in \Omega} \mathbb{P}(e) = 1, \\ & \mathbb{P} \in \mathcal{P} \end{array}$$

of dimension the cardinality of the set S.

Support function : the support function of a conve set $\mathcal{P}$ is given by

$$h_{\mathcal{P}}(g) := \sup \left\{ \sum_{e \in \Omega} g(e)\mathbb{P}(e) \ : \ \mathbb{P} \in \mathcal{P} \right\}$$

and is always a **convex function**.

Our convex reformulation has the **strong convex dual**

$$\begin{array}{lll} \min & h_{\mathcal{P}}(f - r\mathbb{1}) + r \\ \text{s.t.} & r \in \mathrm{R},\ f \in \mathrm{R}^{|\Omega|} \\ & f(e) \geq \ell(x, e), \quad \forall e \in \Omega \end{array} \quad = \quad \begin{array}{ll} \max & \sum_{i \in \Omega} \ell(x, e)\, \mathbb{P}(e) \\ \text{s.t.} & \mathbb{P} \in \mathrm{R}_+^{|\Omega|},\ \ \sum_{e \in \Omega} \mathbb{P}(e) = 1, \\ & \mathbb{P} \in \mathcal{P} \end{array}$$

whenever the *Slater condition* $\operatorname{int} \mathcal{P} \cap \mathcal{P}(\Omega) \neq \emptyset$ holds.

Distributionally robust optimization problem :

$$\begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & x \in \mathrm{R}^d \\ & \mathbf{E}_{\mathbb{P}}\left[\ell(x, \mathbf{p})\right] \leq 0, \quad \forall \mathbb{P} \in \mathcal{P} \end{aligned} \quad = \quad \begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & x \in \mathrm{R}^d,\ r \in \mathrm{R},\ f \in \mathrm{R}^{|\Omega|} \\ & h_{\mathcal{P}}(f - r\mathbb{1}) + r \leq 0 \\ & f(e) \geq \ell(x, e), \quad \forall e \in \Omega \end{aligned}$$

Exact convex reformulation if

- Constraint function $\ell(x, p)$ is convex in x for any p
- Sufficient condition : $\operatorname{int} \mathcal{P} \cap \mathcal{P}(S) \neq \emptyset$

Example : Worst-Case Probability Problem

Example : We consider $\mathbf{p}$ has a probability distribution on 1000 equidistant points on the interval $\Omega = [-1, 1]$. We assume the following prior information

- $\mathbf{E}_{\mathbb{P}}[\mathbf{p}] \in [-0.1, 0.1]$
- $\mathbf{E}_{\mathbb{P}}\left[\mathbf{p}^2\right] \in [0.5, 0.6]$
- $\mathbf{E}_{\mathbb{P}}\left[3\mathbf{p}^3 - 2\mathbf{p}\right] = -0.2$
- $\mathbb{P}(\mathbf{p} < 0) = 0.4$

What is the best information we have regarding

$$\mathbb{P}(\mathbf{p} \leq k) = \mathbf{E}_{\mathbb{P}}[\ell(x, \mathbf{p}) = \mathbb{1}\{\mathbf{p} \leq k\}]?$$

Example : Worst-Case Probability Problem II

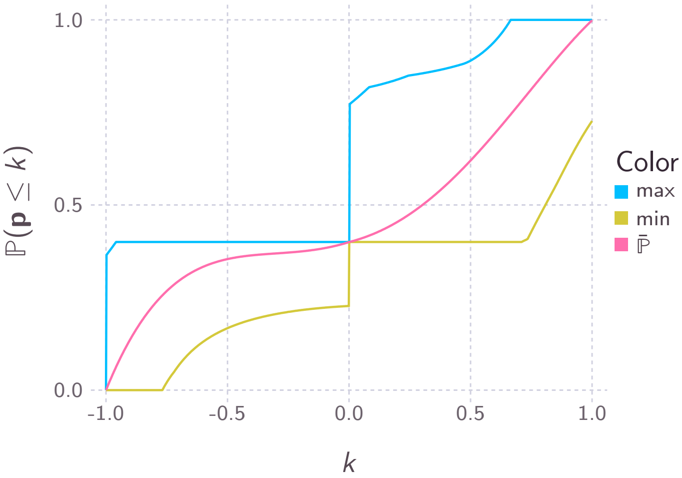


Table of Contents

Worst-case Expectation Problem

Problem : What is the worst-case expectation $\mathbf{E}_{\mathbb{P}}[\ell(x, \mathbf{p})]$ *assuming only*

- $\mathbf{p}$ takes values in R^n
- Mean : $\mathbf{E}_{\mathbb{P}}[\mathbf{p}] = \mu$
- Variance : $\mathbf{E}_{\mathbb{P}}\left[\mathbf{p}\mathbf{p}^\top\right] = \mu\mu^\top + \Sigma$

The problem results in an **infinite dimensional** convex optimization problem

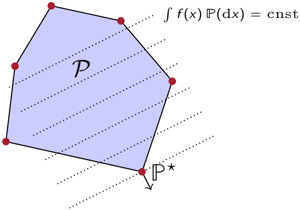
$$\begin{aligned} \inf / \sup \quad & \int \ell(x, e)\, \mathbb{P}(\mathrm{d}e) \\ \text{s.t.} \quad & \mathbb{P} \in \mathcal{P}(\mathrm{R}^n) \\ & \int e\, \mathbb{P}(\mathrm{d}e) = \mu, \\ & \int e \cdot e^\top\, \mathbb{P}(\mathrm{d}e) = \mu \cdot \mu^\top + \Sigma \end{aligned}$$

From infinite to finite dimensional optimization

Krein-Milman : For any compact convex set $\mathcal{P}$,

$$\sup_{\mathbb{P}\in\mathcal{P}} \int f(e)\,\mathbb{P}(\mathrm{d}e) = \sup_{\mathbb{P}\in \operatorname{ex}\mathcal{P}} \int f(e)\,\mathbb{P}(\mathrm{d}e).$$

Optimize over only over the **extreme points** of $\mathcal{P}$.



Two questions need to be answered such that we can optimize over $\mathcal{P}(\mu, \Sigma)$

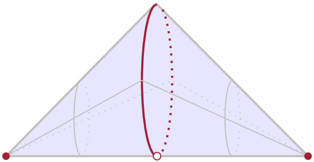
1. Can we identify its extreme points?
2. Can we optimize efficiently over them?

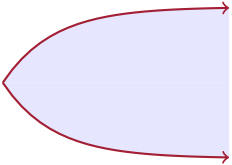
A point $\mathbb{P}$ is called an **extreme point** in a convex set $\mathcal{P}$ if it can not be written as the strict convex combination of two distinct points in the set $\mathcal{P}$.

$$\mathbb{P} \in \operatorname{ex} \mathcal{P} \iff !\exists \mathbb{P}_1, \mathbb{P}_2 \in \mathcal{P} : \quad \mathbb{P} = \lambda \mathbb{P}_1 + (1 - \lambda) \mathbb{P}_2$$

with $\mathbb{P}_1 \neq \mathbb{P}_2$ and $\lambda \in (0, 1)$.

Examples:

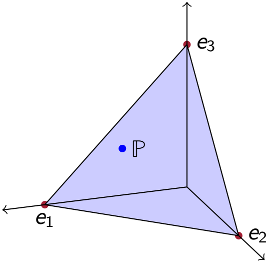




Step 1 : Extreme Distributions

The **Dirac distribution** δ_e is defined as the unique distribution satisfying

$$\forall E \in \mathcal{B}(\Omega): \quad \delta_e(E) = \begin{cases} 1 & \text{if } e \in E, \\ 0 & \text{Otherwise.} \end{cases}$$



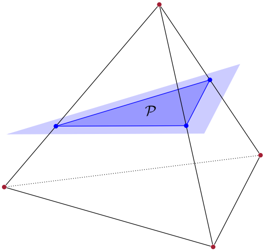
Theorem : The extreme distributions in the probability simplex on Ω are the Dirac distributions

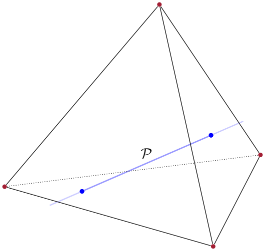
$$\operatorname{ex} \mathcal{P}(\Omega) = \{\delta_e \; : \; e \in \Omega\}$$

Step 1 : Extreme Distributions II

We are interested in the extreme points of moment ambiguity sets

$$\mathcal{P}(\mu, \Sigma) := \mathcal{P}(\mathrm{R}^n) \cap \left\{ \mathbb{P} \in M_+ \ : \ \int e \, \mathbb{P}(\mathrm{d}e) = \mu, \ \int e \cdot e^\top \, \mathbb{P}(\mathrm{d}e) = \mu \cdot \mu^\top + \Sigma \right\}$$





Theorem : the extreme points of a moment set are a finite convex combination of at most c extreme points δ_{e_k} of the probability simplex, i.e.

$$\mathrm{ex}\, \mathcal{P}(\mu, \Sigma) = \left\{ \sum_{k=1}^{c} \lambda_k \delta_{e_k} \ : \ \sum_{k=1}^{c} \lambda_k = 1, \ \lambda \in \mathrm{R}^c_+, \ e_k \in \mathrm{R}^n \right\} \cap \mathcal{P}(\mu, \Sigma)$$

Step 2 : Optimizing over extreme distributions

From **infinite** to **finite dimensional** optimization

$$\sup \left\{ \int \ell(x,e)\,\mathbb{P}(\mathrm{d}e) \ : \ \mathbb{P} \in \mathcal{P}(\mu,\Sigma) \right\} = \begin{array}{rl} \sup & \sum_k \lambda_k \ell(x, e_k) \\ \text{s.t.} & \lambda \geq 0, \ e_k \in \mathrm{R}^n \\ & \sum_k \lambda_k = 1, \ \sum_k \lambda_k e_k = \mu \\ & \sum_k \lambda_k e_k \cdot e_k^\top = \mu \cdot \mu^\top + \Sigma \end{array}$$

The problem of optimizing over the locations e_k and weights λ_k simultaneously remains hard in general.

Using clever reformulations, for **double affine** cost functions

$$\ell(x,p) = \max_i \ \tilde{a}_i(x)^\top p + \tilde{b}_i(x)$$

an exact **convex reformulation** is possible.

Main Theorem

Theorem : Semidefinite optimization formulation for $\ell(x,p) = \max_i \tilde{a}_i(x)^\top p + \tilde{b}_i(x)$

$$\begin{aligned} \sup_{\mathbb{P}\in\mathcal{P}(\mu,\Sigma)} \int \ell(x,e)\,\mathbb{P}(\mathrm{d}e) = \max \quad & \tilde{a}_i(x)^\top z_i + \tilde{b}_i(x)\lambda_i \\ \text{s.t.} \quad & \lambda \geq 0,\; z_i \in \mathrm{R}^n, \\ & \textstyle\sum_i \lambda_i = 1,\; \sum_i z_i = \mu, \\ & \textstyle\sum_i z_i \frac{1}{\lambda_i} z_i^\top \preceq \mu\mu^\top + \Sigma \end{aligned}$$

Observations :

- The constraints are equivalent to

$$\exists Z_i \in \mathrm{R}^{n\times n} : \; \sum_i \begin{pmatrix} Z_i & z_i \\ z_i^\top & \lambda_i \end{pmatrix} = \begin{pmatrix} S & \mu \\ \mu^\top & 1 \end{pmatrix}, \; \begin{pmatrix} Z_i & z_i \\ z_i^\top & \lambda_i \end{pmatrix} \succeq 0, \quad \forall i$$

  with second moment matrix $S = \mu\mu^\top + \Sigma$

The previous semidefinite program has the **strong convex dual**

$$
\begin{array}{llcll}
\min & \mathrm{Tr}(SP) + 2q^\top \mu + r & = & \max & \sum_i \tilde{a}_i(x) z_i + \tilde{b}_i(x)\lambda_i \\
\text{s.t.} & P \in \mathrm{R}^{n\times n},\ q \in \mathrm{R}^n,\ r \in \mathrm{R} & & \text{s.t.} & \lambda \geq 0,\ z_i \in \mathrm{R}^n, \\
& \begin{pmatrix} P & q \\ q^\top & r \end{pmatrix} \succeq \frac{1}{2} \begin{pmatrix} 0 & \tilde{a}_i(x) \\ \tilde{a}_i(x)^\top & 2\tilde{b}_i(x) \end{pmatrix},\ \forall i & & & \sum_i \lambda_i = 1,\ \sum_i z_i = \mu, \\
& & & & \sum_i z_i \frac{1}{\lambda_i} z_i^\top \preceq \mu\mu^\top + \Sigma
\end{array}
$$

whenever we have the *Slater condition*

$$
\mathrm{int}\, \mathcal{P}(\mu, \Sigma) \neq \emptyset \iff \Sigma \in \mathrm{S}^n_{++}
$$

Distributionally robust optimization problem :

$$
\begin{array}{lll}
\min & c^\top x \\
\text{s.t.} & x \in \mathrm{R}^d \\
& \mathbf{E}_{\mathbb{P}}\left[\ell(x, \mathbf{p})\right] \leq 0, \quad \forall \mathbb{P} \in \mathcal{P}(\mu, \Sigma)
\end{array}
\quad = \quad
\begin{array}{lll}
\min & c^\top x \\
\text{s.t.} & x \in \mathrm{R}^d,\ P \in \mathrm{R}^{d\times d},\ q \in \mathrm{R}^d,\ r \in \mathrm{R} \\
& \mathrm{Tr}(SP) + 2q^\top \mu + r \leq 0 \\
& \begin{pmatrix} P & q \\ q^\top & r \end{pmatrix} \succeq \frac{1}{2}\begin{pmatrix} 0 & \tilde{a}_i(x) \\ \tilde{a}_i(x)^\top & 2\tilde{b}_i(x) \end{pmatrix},\ \forall i
\end{array}
$$

Exact convex reformulation if

- Constraint function $\ell(x, p)$ is double affine.
- Sufficient condition : $\Sigma \in \mathrm{S}^n_{++}$

Table of Contents

Two-Stage Optimization

Two-stage optimization problem:

$$\min_{x \in \mathrm{R}^d} c^\top x + \sup_{\mathbb{P} \in \mathcal{P}} \mathbf{E}_{\mathbb{P}} \left[\min_y \left\{ d^\top y \ : \ A(\mathbf{p})x + By \leq b(\mathbf{p}) \right\} \right]$$

The problem contains three stages :

1. The vector x collects the **first-stage** decision variables
2. An uncertain variable $\mathbf{p} \sim \mathbb{P} \in \mathcal{P}$ is observed
3. The vector y collects the **second-stage** decision variables

Model problem for dynamic decision making.

Two-Stage Optimization II

The cost Q of the optimal second stage decision $y^\star$ depends on the initial decision x and an uncertain parameter p as

$$\begin{aligned} Q(x,p) &:= \min\left\{d^\top y \;:\; A(p)x + By \leq b(p)\right\} \\ &= \max\left\{\lambda^\top (A(p)x - b(p)) \;:\; \lambda \geq 0,\; d = -\lambda^\top B\right\} \end{aligned}$$

Two important remarks :

- The dual representation makes clear the **cost** $Q(x,p)$ **is convex** in x for any p.
- Using the vertices v_i of the polytope $\Lambda := \left\{\lambda \geq 0 \;:\; d = -\lambda^\top B\right\}$ we have

$$Q(x,p) = \max_i \; v_i^\top (A(p)x - b(p))$$

Two-Stage Optimization III

$$
\begin{aligned}
&\min_{x\in\mathbb{R}^d} c^\top x + \sup_{\mathbb{P}\in\mathcal{P}} \ \mathbf{E}_{\mathbb{P}}\left[\min_y \left\{ d^\top y \ : \ A(\mathbf{p})x + By \le b(\mathbf{p})\right\}\right] \\
&= \min_{x\in\mathbb{R}^d} c^\top x + \sup_{\mathbb{P}\in\mathcal{P}} \ \mathbf{E}_{\mathbb{P}}\left[Q(x,\mathbf{p})\right] \\
&= \min_{x\in\mathbb{R}^d} c^\top x + \beta(x)
\end{aligned}
$$

The problem is a **convex optimization** problem for any $\mathcal{P}$ as

$$
\beta(x) = \sup_{\mathbb{P}\in\mathcal{P}} \mathbf{E}_{\mathbb{P}}\left[Q(x,\mathbf{p})\right]
$$

is convex.

The **tractability** does depends on the ambiguity set $\mathcal{P}$.

Theorem : The problem is always tractable if $\mathcal{P} \subset \mathcal{P}(\Omega)$ with $|\Omega|$ finite.

$$\beta(x) = \sup\left\{\sum_{i\in\Omega} \mathbb{P}(i) Q(x,i) \ : \ \sum_{i\in\Omega} \mathbb{P}(i) = 1, \ \mathbb{P} \in \mathcal{P}\right\}$$
$$= \inf\left\{r + h_{\mathcal{P}}(q - r\mathbb{1}) \ : \ q_i \geq Q(x,i), \quad \forall i \in \Omega\right\}$$

Using the definition of the convex cost Q we can write

$$= \inf\left\{r + h_{\mathcal{P}}(q - r\mathbb{1}) \ : \ q_i \geq d^\top y_i, \ A(i)x + By_i \leq b(i) \quad \forall i \in \Omega\right\}$$

Observations :

- The number of constraints depends linearly on the cardinality of Ω.
- The type of optimization problem depends on the functional form of $h_{\mathcal{P}}$.

Tractability for second-order moment ambiguity sets

$$\mathcal{P}(\mu, \boldsymbol{\Sigma}) = \left\{ \mathbb{P} \in \mathcal{P}(\mathrm{R}^n) \ : \ \int x \, \mathbb{P}(\mathrm{d}x) = \mu, \ \int x x^\top \, \mathbb{P}(\mathrm{d}x) = \mu\mu^\top + \boldsymbol{\Sigma} \right\}$$

requires **assumptions** on the technology matrix and budget vector

- **Affine** technology matrix : $A(p) = A_0 + \sum_{j=1}^n p_j A_j$
- **Affine** budget vector : $b(p) = b_0 + \sum_{j=1}^n p_j b_j$

Using the extreme point representation of the cost function Q we have

$$\begin{aligned} \beta(x) &= \sup_{\mathbb{P} \in \mathcal{P}(\mu, \Sigma)} \mathbf{E}_{\mathbb{P}} \left[Q(x, \mathbf{p}) \right] \\ &= \sup_{\mathbb{P} \in \mathcal{P}(\mu, \Sigma)} \mathbf{E}_{\mathbb{P}} \left[\max_i v_i^\top (A(\mathbf{p})x - b(\mathbf{p})) \right] \end{aligned}$$

which has the form for which we have a **semidefinite representation**.

The function β can be written as

$$\begin{aligned}\beta(x) &= \sup_{\mathbb{P}\in\mathcal{P}(\mu,\Sigma)} \mathbf{E}_{\mathbb{P}}\left[\max_i v_i^\top (A(\mathbf{p})x - b(\mathbf{p}))\right] \\ &= \inf_{P,\, q,\, r} \left\{ \mathrm{Tr}(\Sigma P) + 2q^\top \mu + r \ :\ \begin{pmatrix} P & q \\ q^\top & r \end{pmatrix} \succeq \frac{1}{2} \begin{pmatrix} 0 & \tilde{a}_i(x) \\ \tilde{a}_i(x)^\top & 2\tilde{b}_i(x) \end{pmatrix},\ \forall i \right\}\end{aligned}$$

where $\tilde{a}_{ij}(x) = v_i^\top (A_j x - b_j)$ and $\tilde{b}_i = v_i^\top (A_0 x - b_0)$.

Observations :

- The number of constraints depends linearly on the number of extreme points of the polytope
$$\Lambda := \left\{ \lambda \geq 0 \ :\ d = -\lambda^\top B \right\}.$$
- The number of extreme points may be exponential in the number of constraints.

Equivalent adaptive optimization formulation

$$
\begin{aligned}
&\min_{x \in \mathrm{R}^d} c^\top x + \sup_{\mathbb{P} \in \mathcal{P}} \mathbf{E}_{\mathbb{P}} \left[\min_y \left\{ d^\top y \;:\; A(\mathbf{p})x + By \leq b(\mathbf{p}) \right\} \right] \\
= &\min_{x \in \mathrm{R}^d} c^\top x + \min_{y(\cdot)} \left\{ \sup_{\mathbb{P} \in \mathcal{P}} \mathbf{E}_{\mathbb{P}} \left[d^\top y(\mathbf{p}) \right] \;:\; A(p)x + By(p) \leq b(p), \;\; \forall p \in S \right\}
\end{aligned}
$$

where S the smallest set such that $\mathcal{P} \subseteq \mathcal{P}(S)$.

Adaptive optimization over

- A **static** decision x
- A decision **rule** $y(\cdot)$ adaptive to the uncertain parameter $\mathbf{p}$

Affine Decision Rules

For tractability, we restrict attention to second-stage decisions which are affine:

$$y(p) = y_r + Fp$$

- Vector y_r encodes the **nominal plan** for $p = 0$
- Matrix F describes how to **adapt** to $p \neq 0$

Affine Robust Counterpart :

$$\min_{x} \min_{y_r,\, F} \left\{ c^\top x + d^\top y_r + \sup_{\mathbb{P} \in \mathcal{P}} d^\top F\, \mathbf{E}_{\mathbb{P}}\left[\mathbf{p}\right] \;:\; A(p)x + By_r + BFp \leq b(p), \quad \forall p \in S \right\}$$

Observations

- Always tractable, but not exact.

Example – Multi-period Inventory Control

Consider a finite horizon, T period single product inventory control problem

- $\mathbf{d}_t$ for $t \in [T]$: Uncertain demands
- x_t for $t \in [T]$: Order quantities
- y_t for $t \in [T]$: Nett inventories

Inventory dynamics :

$$y_{t+1} = y_t + x_t - \mathbf{d}_t$$

Cost model :

- Marginal purchase cost in period t : c_t
- Marginal holding cost in period t : h_t
- Marginal backlogged cost in period t : b_t

Initial condition : No inventory $y_1 = 0$ and no purchase history $\mathbf{d}_0 = 0$

Example – Multi-period Inventory Control II

Minimize the worst-case expected total cost over the entire horizon :

$$
\begin{array}{ll}
\min & \sup_{\mathbb{P}\in\mathcal{P}} \mathbf{E}_{\mathbb{P}}\left[\sum_{t\in[T]} c_t x_t(\mathbf{d}) + v_t(\mathbf{d})\right] \\
\text{s.t.} & y_{t+1}(\mathbf{d}) = y_t(\mathbf{p}) + x_t(\mathbf{d}) - \mathbf{d}_t, \quad \forall t \in [T] \\
& v_t(\mathbf{d}) \geq h_t y_{t+1}(\mathbf{p}), \quad \forall t \in [T] \\
& v_t(\mathbf{d}) \geq -b_t y_{t+1}(\mathbf{p}), \quad \forall t \in [T] \\
& 0 \leq x_t(\mathbf{d}) \leq \bar{x}_t, \quad \forall t \in [T] \\
& x_t(\cdot),\ y_t(\cdot),\ v_t(\cdot) : \quad \mathrm{t-adaptable}
\end{array}
$$

The set of distributions we hedge against consists of

$$
\mathcal{P} := \left\{ \mathbb{P} \in \mathcal{P}(\mathrm{R}_+^T) \ : \quad \begin{array}{l} \mathbf{E}_{\mathbb{P}}\left[\mathbf{d}\right] = \mu \\ \mathbf{E}_{\mathbb{P}}\left[\sum_{t=s}^{T} (\mathbf{d_t} - \mu_t)^2\right] \leq \theta_{st}^2, \quad \forall s \leq t \end{array} \right\}
$$

with $\theta_{st}^2 := (t - 1 + 1)\sigma^2$.

Example – Multi-period Inventory Control III

Numerical Values

- Purchase cost $c_t = 0.1$
- Holding cost $h_t = 0.02$
- Backlog cost $b_t = 0.2$
- Maximum order quantity $\bar{x}_t = 260$

T	μ	σ^2	Lower Bound	Approx MM	Approx PCM
5	200	533.3	108.0	191.6	131.4
10	200	133.3	206.0	302.1	229.9
20	240	48.0	486.0	635.1	518.2
30	240	21.33	725.0	905.8	761.2