# Lecture 19 : Dynamic Robust Optimization
# +
# Distributionall Robust Optimization

December 3, 2025

# Recall "Classical" Robust Optimization (RO)

- Only information about unknowns $z$: they belong to an **uncertainty set** $\mathcal{U}$
- Solve the following optimization problem:

$$(P) \qquad \begin{aligned} & \inf_{x} \sup_{z \in \mathcal{U}} C(x, z) \\ & \text{s.t. } f_i(x, z) \leq 0, \forall z \in \mathcal{U}, \forall i \in I \end{aligned}$$

- This model has **infinitely many** constraints
- W.l.o.g., we can consider uncertainty only in the constraints
- Each and every constraint must satisfied: $f_i(x, z) \leq 0, \forall z \in \mathcal{U}$
- How to reformulate this as a **finite-dimensional, tractable** optimization problem, a.k.a. the **robust counterpart**?

# "Classical" Uncertainty Sets

The robust counterpart for $\boxed{(\bar{a} + Pz)^{\mathsf{T}} x \leq b, \ \forall\, z \in \mathcal{U}}$ is:

# "Classical" Uncertainty Sets

The robust counterpart for $\boxed{(\bar{a} + Pz)^\top x \le b,\ \forall z \in \mathcal{U}}$ is:

| U-set | $\mathcal{U}$ | Robust Counterpart | Tractability |
|---|---|---|---|
| Box | $\|z\|_\infty \le \rho$ | $\bar{a}^\top x + \rho \|P^\top x\|_1 \le b$ | LO |
| Ellipsoidal | $\|z\|_2 \le \rho$ | $\bar{a}^\top x + \rho \|P^\top x\|_2 \le b$ | CQO |
| Polyhedral | $Dz \le d$ | $\exists y : \begin{cases} \bar{a}^\top x + d^\top y \le b \\ D^\top y = P^\top x \\ y \ge 0 \end{cases}$ | LO |
| Budget | $\begin{cases} \|z\|_\infty \le \rho \\ \|z\|_1 \le \Gamma \end{cases}$ | $\exists y : \bar{a}^\top x + \rho \|y\|_1 + \Gamma \|P^\top x - y\|_\infty \le b$ | LO |
| Convex | $h_k(z) \le 0$ | $\exists \{w_k, u_k\}_{k \in K} : \begin{cases} a^\top x + \sum_k u_k h_k^* \left( \frac{w_k}{u_k} \right) \le b \\ \sum_k w_k = P^\top x \\ u \ge 0 \end{cases}$ | Conv. Opt. |

# "Classical" Uncertainty Sets

The robust counterpart for $\boxed{(\bar{a} + Pz)^\top x \le b,\ \forall\, z \in \mathcal{U}}$ is:

| U-set | $\mathcal{U}$ | Robust Counterpart | Tractability |
|---|---|---|---|
| Box | $\|z\|_\infty \le \rho$ | $\bar{a}^\top x + \rho \|P^\top x\|_1 \le b$ | LO |
| Ellipsoidal | $\|z\|_2 \le \rho$ | $\bar{a}^\top x + \rho \|P^\top x\|_2 \le b$ | CQO |
| Polyhedral | $Dz \le d$ | $\exists y \;:\; \begin{cases} \bar{a}^\top x + d^\top y \le b \\ D^\top y = P^\top x \\ y \ge 0 \end{cases}$ | LO |
| Budget | $\begin{cases} \|z\|_\infty \le \rho \\ \|z\|_1 \le \Gamma \end{cases}$ | $\exists y \;:\; \bar{a}^\top x + \rho \|y\|_1 + \Gamma \|P^\top x - y\|_\infty \le b$ | LO |
| Convex | $h_k(z) \le 0$ | $\exists \{w_k, u_k\}_{k \in K} \;:\; \begin{cases} a^\top x + \sum_k u_k h_k^* \left(\frac{w_k}{u_k}\right) \le b \\ \sum_k w_k = P^\top x \\ u \ge 0 \end{cases}$ | Conv. Opt. |

- Several extensions
- Robust counterparts can be handled by large-scale modern solvers
- Enough for many practical problems

# Two Important Caveats for Robust Models

# Example: Facility Location Problem (Baron et al. 2011)

*Need to decide where to open facilities, how much capacity to install, and how to assign customer demands over a future planning horizon, in order to maximize profits.*

# Example: Facility Location Problem (Baron et al. 2011)

*Need to decide where to open facilities, how much capacity to install, and how to assign customer demands over a future planning horizon, in order to maximize profits.*

**Step 1.** Start with a **deterministic** model formulation:

# Example: Facility Location Problem (Baron et al. 2011)

*Need to decide where to open facilities, how much capacity to install, and how to assign customer demands over a future planning horizon, in order to maximize profits.*

**Step 1.** Start with a **deterministic** model formulation:

$$
\begin{aligned}
\max_{X,I,Z,P} \quad & \sum_{\tau \in \mathcal{T}} \sum_{i \in \mathcal{F}} \sum_{j \in \mathcal{N}} (p - c_{ij}^s) X_{ij\tau} - \sum_{\tau \in \mathcal{T}} \sum_{i \in \mathcal{F}} c_i P_{i\tau} - \sum_{i \in \mathcal{F}} (C_i Z_i + K_i I_i) \\
\text{subject to} \quad & \sum_{i \in \mathcal{F}} X_{ij\tau} \leq D_{j\tau}, \qquad j \in \mathcal{N},\ \tau \in \mathcal{T}, \\
& \sum_{j \in \mathcal{N}} X_{ij\tau} \leq P_{i\tau}, \qquad i \in \mathcal{F},\ \tau \in \mathcal{T}, \\
& P_{i\tau} \leq Z_i,\ \ Z_i \leq M \cdot I_i, \qquad i \in \mathcal{F},\ \tau \in \mathcal{T} \quad (M \text{ is a large constant}) \\
& X \geq 0, \quad I \in \{0,1\}^{|\mathcal{F}|}
\end{aligned}
$$

**Parameters**

$\mathcal{T}$: discrete planning horizon, indexed by $\tau$
$\mathcal{F}$: potential facility locations, indexed by $i$
$\mathcal{N}$: demand node locations, indexed by $j$
$p$: unit price of goods
$c_i$: cost per unit of production at facility $i$
$C_i$: cost per unit of capacity for facility $i$
$K_i$: cost of opening a facility at location $i$
$c_{ij}^s$: cost of shipping units from $i$ to $j$
$D_{j\tau}$: demand in period $\tau$ at location $j$

**Decision variables**

$X_{ij\tau}$: quantity of demand $j$ in period $\tau$ satisfied by $i$
$P_{i\tau}$: quantity produced at facility $i$ in period $\tau$
$I_i$: whether facility $i$ is open (0/1)
$Z_i$: capacity of facility $i$ if open

# Example: Facility Location Problem (Baron et al. 2011)

*Need to decide where to open facilities, how much capacity to install, and how to assign customer demands over a future planning horizon, in order to maximize profits.*

**Step 1.** Start with a **deterministic** model formulation:

$$
\begin{aligned}
\max_{X,I,Z,P} \quad & \sum_{\tau \in \mathcal{T}} \sum_{i \in \mathcal{F}} \sum_{j \in \mathcal{N}} (p - c_{ij}^s) X_{ij\tau} - \sum_{\tau \in \mathcal{T}} \sum_{i \in \mathcal{F}} c_i P_{i\tau} - \sum_{i \in \mathcal{F}} (C_i Z_i + K_i I_i) \\
\text{subject to} \quad & \sum_{i \in \mathcal{F}} X_{ij\tau} \leq D_{j\tau}, \qquad j \in \mathcal{N},\ \tau \in \mathcal{T}, \\
& \sum_{j \in \mathcal{N}} X_{ij\tau} \leq P_{i\tau}, \qquad i \in \mathcal{F},\ \tau \in \mathcal{T}, \\
& P_{i\tau} \leq Z_i,\ \ Z_i \leq M \cdot I_i, \qquad i \in \mathcal{F},\ \tau \in \mathcal{T} \quad (M \text{ is a large constant}) \\
& X \geq 0, \quad I \in \{0,1\}^{|\mathcal{F}|}
\end{aligned}
$$

**Step 2.** Identify all uncertain parameters and model the uncertainty set $\mathcal{U}$.

Baron et al. 2011 captured uncertain demands with an ellipsoidal uncertainty set:

$$
\mathcal{U} = \left\{ D \in \mathbb{R}^{|\mathcal{N}| \cdot |\mathcal{T}|} \;\middle|\; \sum_{j \in \mathcal{N}} \sum_{t \in \mathcal{T}} \left( \frac{D_{jt} - \bar{D}_{jt}}{\epsilon_t \bar{D}_{jt}} \right)^2 \leq \rho^2 \right\},
$$

$\{\bar{D}_{jt}\}_{j \in \mathcal{N}; t \in \mathcal{T}}$ are "nominal" demands, $\epsilon_t$ is allowed deviation (%), $\rho$ is the size of the ellipsoid

# Example: Facility Location Problem (Baron et al. 2011)

*Need to decide where to open facilities, how much capacity to install, and how to assign customer demands over a future planning horizon, in order to maximize profits.*

**Step 1.** Start with a **deterministic** model formulation:

$$\begin{aligned}
\max_{X,I,Z,P} \quad & \sum_{\tau\in\mathcal{T}}\sum_{i\in\mathcal{F}}\sum_{j\in\mathcal{N}}(p - c_{ij}^s)X_{ij\tau} - \sum_{\tau\in\mathcal{T}}\sum_{i\in\mathcal{F}} c_i P_{i\tau} - \sum_{i\in\mathcal{F}}(C_i Z_i + K_i I_i) \\
\text{subject to} \quad & \sum_{i\in\mathcal{F}} X_{ij\tau} \le D_{j\tau}, \quad j\in\mathcal{N},\ \tau\in\mathcal{T}, \\
& \sum_{j\in\mathcal{N}} X_{ij\tau} \le P_{i\tau}, \quad i\in\mathcal{F},\ \tau\in\mathcal{T}, \\
& P_{i\tau} \le Z_i,\ \ Z_i \le M\cdot I_i, \quad i\in\mathcal{F},\ \tau\in\mathcal{T} \quad (M \text{ is a large constant}) \\
& X \ge 0, \quad I \in \{0,1\}^{|\mathcal{F}|}
\end{aligned}$$

**Step 2.** Identify all uncertain parameters and model the uncertainty set $\mathcal{U}$.
Baron et al. 2011 captured uncertain demands with an ellipsoidal uncertainty set:

$$\mathcal{U} = \left\{ D \in \mathbb{R}^{|\mathcal{N}|\cdot|\mathcal{T}|} \ \middle|\ \sum_{j\in\mathcal{N}}\sum_{t\in\mathcal{T}}\left(\frac{D_{jt} - \bar{D}_{jt}}{\epsilon_t \bar{D}_{jt}}\right)^2 \le \rho^2 \right\},$$

Equivalently, can write $D_{jt} = \bar{D}_{jt}(1 + \epsilon_t \cdot z_{jt})$, where $z \in \mathcal{U} = \{z \in \mathbb{R}^{|\mathcal{N}|\cdot|\mathcal{T}|} \ :\ \|z\|_2 \le \rho\}$

# Example: Facility Location Problem (Baron et al. 2011)

*Need to decide where to open facilities, how much capacity to install, and how to assign customer demands over a future planning horizon, in order to maximize profits.*

**Step 1.** Start with a **deterministic** model formulation:

$$\begin{aligned}
\max_{X,I,Z,P} \quad & \sum_{\tau\in\mathcal{T}}\sum_{i\in\mathcal{F}}\sum_{j\in\mathcal{N}}(p-c^s_{ij})X_{ij\tau} - \sum_{\tau\in\mathcal{T}}\sum_{i\in\mathcal{F}} c_i P_{i\tau} - \sum_{i\in\mathcal{F}}(C_i Z_i + K_i I_i) \\
\text{subject to} \quad & \sum_{i\in\mathcal{F}} X_{ij\tau} \le D_{j\tau}, \qquad j\in\mathcal{N},\ \tau\in\mathcal{T}, \\
& \sum_{j\in\mathcal{N}} X_{ij\tau} \le P_{i\tau}, \qquad i\in\mathcal{F},\ \tau\in\mathcal{T}, \\
& P_{i\tau}\le Z_i,\ \ Z_i \le M\cdot I_i, \qquad i\in\mathcal{F},\ \tau\in\mathcal{T} \quad (M \text{ is a large constant}) \\
& X\ge 0,\ \ I\in\{0,1\}^{|\mathcal{F}|}
\end{aligned}$$

**Step 2.** Identify all uncertain parameters and model the uncertainty set $\mathcal{U}$.
Baron et al. 2011 captured uncertain demands with an ellipsoidal uncertainty set:

$$\mathcal{U} = \left\{ D\in\mathbb{R}^{|\mathcal{N}|\cdot|\mathcal{T}|} \;\middle|\; \sum_{j\in\mathcal{N}}\sum_{t\in\mathcal{T}}\left(\frac{D_{jt}-\bar{D}_{jt}}{\epsilon_t \bar{D}_{jt}}\right)^2 \le \rho^2 \right\},$$

**Step 3.** Derive robust counterpart for the problem. Here, a Conic Quadratic program.

# Compare Two Models

Our initial model, with **decisions for quantities** $X$:

$$
\begin{aligned}
\max_{X,I,Z,P} \quad & \sum_{\tau\in\mathcal{T}}\sum_{i\in\mathcal{F}}\sum_{j\in\mathcal{N}}(p-c^s_{ij})X_{ij\tau} - \sum_{\tau\in\mathcal{T}}\sum_{i\in\mathcal{F}} c_i P_{i\tau} - \sum_{i\in\mathcal{F}}(C_i Z_i + K_i I_i) \\
\text{subject to} \quad & \sum_{i\in\mathcal{F}} X_{ij\tau} \le D_{j\tau}, \qquad j\in\mathcal{N},\ \tau\in\mathcal{T}, \\
& \sum_{j\in\mathcal{N}} X_{ij\tau} \le P_{i\tau}, \qquad i\in\mathcal{F},\ \tau\in\mathcal{T}, \\
& P_{i\tau} \le Z_i,\ \ Z_i \le M\cdot I_i, \qquad i\in\mathcal{F},\ \tau\in\mathcal{T} \quad (M \text{ is a large constant}) \\
& X\ge 0,\ \ I\in\{0,1\}^{|\mathcal{F}|}
\end{aligned}
$$

# Compare Two Models

Our initial model, with **decisions for quantities** $X$:

$$\begin{aligned}
\max_{X,I,Z,P} \quad & \sum_{\tau\in\mathcal{T}}\sum_{i\in\mathcal{F}}\sum_{j\in\mathcal{N}}(p-c_{ij}^{s})X_{ij\tau}-\sum_{\tau\in\mathcal{T}}\sum_{i\in\mathcal{F}}c_iP_{i\tau}-\sum_{i\in\mathcal{F}}(C_iZ_i+K_iI_i)\\
\text{subject to} \quad & \sum_{i\in\mathcal{F}}X_{ij\tau}\le D_{j\tau}, \quad j\in\mathcal{N},\ \tau\in\mathcal{T},\\
& \sum_{j\in\mathcal{N}}X_{ij\tau}\le P_{i\tau}, \quad i\in\mathcal{F},\ \tau\in\mathcal{T},\\
& P_{i\tau}\le Z_i,\ \ Z_i\le M\cdot I_i, \quad i\in\mathcal{F},\ \tau\in\mathcal{T} \quad (M \text{ is a large constant})\\
& X\ge 0,\ \ I\in\{0,1\}^{|\mathcal{F}|}
\end{aligned}$$

Another model, with **decisions for fractions of demands** $Y$:

$$\begin{aligned}
\max_{Y,I,Z,P} \quad & \sum_{\tau\in\mathcal{T}}\sum_{i\in\mathcal{F}}\sum_{j\in\mathcal{N}}(p-c_{ij}^{s})Y_{ij\tau}D_{j\tau}-\sum_{\tau\in\mathcal{T}}\sum_{i\in\mathcal{F}}c_iP_{i\tau}-\sum_{i\in\mathcal{F}}(C_iZ_i-K_iI_i)\\
\text{subject to} \quad & \sum_{i\in\mathcal{F}}Y_{ij\tau}\le 1, \quad j\in\mathcal{N},\ \tau\in\mathcal{T},\\
& \sum_{j\in\mathcal{N}}Y_{ij\tau}D_{j\tau}\le P_{i\tau}, \quad i\in\mathcal{F},\ \tau\in\mathcal{T},\\
& P_{i\tau}\le Z_i,\ \ Z_i\le M\cdot I_i, \quad i\in\mathcal{F},\ \tau\in\mathcal{T}\\
& Y\ge 0,\ \ I\in\{0,1\}^{|\mathcal{F}|}
\end{aligned} \tag{1}$$

# Compare Two Models

Our initial model, with **decisions for quantities** $X$:

$$\begin{aligned}
\max_{X,I,Z,P} \quad & \sum_{\tau\in\mathcal{T}}\sum_{i\in\mathcal{F}}\sum_{j\in\mathcal{N}}(p-c^s_{ij})X_{ij\tau} - \sum_{\tau\in\mathcal{T}}\sum_{i\in\mathcal{F}} c_i P_{i\tau} - \sum_{i\in\mathcal{F}}(C_i Z_i + K_i I_i) \\
\text{subject to} \quad & \sum_{i\in\mathcal{F}} X_{ij\tau} \le D_{j\tau}, \quad j\in\mathcal{N},\ \tau\in\mathcal{T}, \\
& \sum_{j\in\mathcal{N}} X_{ij\tau} \le P_{i\tau}, \quad i\in\mathcal{F},\ \tau\in\mathcal{T}, \\
& P_{i\tau} \le Z_i,\ \ Z_i \le M\cdot I_i, \quad i\in\mathcal{F},\ \tau\in\mathcal{T} \quad (M \text{ is a large constant}) \\
& X \ge 0,\ \ I\in\{0,1\}^{|\mathcal{F}|}
\end{aligned}$$

Another model, with **decisions for fractions of demands** $Y$:

$$\begin{aligned}
\max_{Y,I,Z,P} \quad & \sum_{\tau\in\mathcal{T}}\sum_{i\in\mathcal{F}}\sum_{j\in\mathcal{N}}(p-c^s_{ij})Y_{ij\tau}D_{j\tau} - \sum_{\tau\in\mathcal{T}}\sum_{i\in\mathcal{F}} c_i P_{i\tau} - \sum_{i\in\mathcal{F}}(C_i Z_i - K_i I_i) \\
\text{subject to} \quad & \sum_{i\in\mathcal{F}} Y_{ij\tau} \le 1, \quad j\in\mathcal{N},\ \tau\in\mathcal{T}, \\
& \sum_{j\in\mathcal{N}} Y_{ij\tau}D_{j\tau} \le P_{i\tau}, \quad i\in\mathcal{F},\ \tau\in\mathcal{T}, \\
& P_{i\tau} \le Z_i,\ \ Z_i \le M\cdot I_i, \quad i\in\mathcal{F},\ \tau\in\mathcal{T} \\
& Y \ge 0,\ \ I\in\{0,1\}^{|\mathcal{F}|}
\end{aligned} \tag{1}$$

- For fixed $D$, these **deterministic/nominal** models **are equivalent**

# Compare Two Models

Our initial model, with **decisions for quantities** $X$:

$$\begin{aligned}
\max_{X,I,Z,P} \quad & \sum_{\tau\in\mathcal{T}}\sum_{i\in\mathcal{F}}\sum_{j\in\mathcal{N}}(p-c^s_{ij})X_{ij\tau} - \sum_{\tau\in\mathcal{T}}\sum_{i\in\mathcal{F}} c_i P_{i\tau} - \sum_{i\in\mathcal{F}}(C_i Z_i + K_i I_i) \\
\text{subject to} \quad & \sum_{i\in\mathcal{F}} X_{ij\tau} \le D_{j\tau}, \quad j\in\mathcal{N},\ \tau\in\mathcal{T}, \\
& \sum_{j\in\mathcal{N}} X_{ij\tau} \le P_{i\tau}, \quad i\in\mathcal{F},\ \tau\in\mathcal{T}, \\
& P_{i\tau} \le Z_i,\ \ Z_i \le M\cdot I_i, \quad i\in\mathcal{F},\ \tau\in\mathcal{T} \quad (M \text{ is a large constant}) \\
& X \ge 0,\ \ I\in\{0,1\}^{|\mathcal{F}|}
\end{aligned}$$

Another model, with **decisions for fractions of demands** $Y$:

$$\begin{aligned}
\max_{Y,I,Z,P} \quad & \sum_{\tau\in\mathcal{T}}\sum_{i\in\mathcal{F}}\sum_{j\in\mathcal{N}}(p-c^s_{ij})Y_{ij\tau}D_{j\tau} - \sum_{\tau\in\mathcal{T}}\sum_{i\in\mathcal{F}} c_i P_{i\tau} - \sum_{i\in\mathcal{F}}(C_i Z_i - K_i I_i) \\
\text{subject to} \quad & \sum_{i\in\mathcal{F}} Y_{ij\tau} \le 1, \quad j\in\mathcal{N},\ \tau\in\mathcal{T}, \\
& \sum_{j\in\mathcal{N}} Y_{ij\tau}D_{j\tau} \le P_{i\tau}, \quad i\in\mathcal{F},\ \tau\in\mathcal{T}, \\
& P_{i\tau} \le Z_i,\ \ Z_i \le M\cdot I_i, \quad i\in\mathcal{F},\ \tau\in\mathcal{T} \\
& Y \ge 0,\ \ I\in\{0,1\}^{|\mathcal{F}|}
\end{aligned} \tag{1}$$

- For fixed $D$, these **deterministic/nominal** models **are equivalent**
- But their **robust counterparts are not equivalent**!
  - The feasible set in the second formulation is **larger**

# Compare Two Models

Our initial model, with **decisions for quantities** $X$:

$$\begin{aligned}
\max_{X,I,Z,P} \quad & \sum_{\tau\in\mathcal{T}}\sum_{i\in\mathcal{F}}\sum_{j\in\mathcal{N}}(p-c^s_{ij})X_{ij\tau} - \sum_{\tau\in\mathcal{T}}\sum_{i\in\mathcal{F}} c_i P_{i\tau} - \sum_{i\in\mathcal{F}}(C_i Z_i + K_i I_i) \\
\text{subject to} \quad & \sum_{i\in\mathcal{F}} X_{ij\tau} \le D_{j\tau}, \quad j\in\mathcal{N},\ \tau\in\mathcal{T}, \\
& \sum_{j\in\mathcal{N}} X_{ij\tau} \le P_{i\tau}, \quad i\in\mathcal{F},\ \tau\in\mathcal{T}, \\
& P_{i\tau} \le Z_i,\ \ Z_i \le M\cdot I_i, \quad i\in\mathcal{F},\ \tau\in\mathcal{T} \quad (M \text{ is a large constant}) \\
& X \ge 0,\ \ I\in\{0,1\}^{|\mathcal{F}|}
\end{aligned}$$

Another model, with **decisions for fractions of demands** $Y$:

$$\begin{aligned}
\max_{Y,I,Z,P} \quad & \sum_{\tau\in\mathcal{T}}\sum_{i\in\mathcal{F}}\sum_{j\in\mathcal{N}}(p-c^s_{ij})Y_{ij\tau}D_{j\tau} - \sum_{\tau\in\mathcal{T}}\sum_{i\in\mathcal{F}} c_i P_{i\tau} - \sum_{i\in\mathcal{F}}(C_i Z_i - K_i I_i) \\
\text{subject to} \quad & \sum_{i\in\mathcal{F}} Y_{ij\tau} \le 1, \quad j\in\mathcal{N},\ \tau\in\mathcal{T}, \\
& \sum_{j\in\mathcal{N}} Y_{ij\tau}D_{j\tau} \le P_{i\tau}, \quad i\in\mathcal{F},\ \tau\in\mathcal{T}, \\
& P_{i\tau} \le Z_i,\ \ Z_i \le M\cdot I_i, \quad i\in\mathcal{F},\ \tau\in\mathcal{T} \\
& Y \ge 0,\ \ I\in\{0,1\}^{|\mathcal{F}|}
\end{aligned} \tag{2}$$

- Reason: true formulation allows choosing $X$ (and $Z$) after observing $D$:

$$\max_{I,Z}\min_{D_{j,1}}\max_{X_{i,j,1},P_{i,1}}\min_{D_{j,2}}\max_{X_{i,j,2},P_{i,2}}\cdots$$

- Second formulation implements ordering quantities that **depend on demand**!

The **robust counterparts** of **equivalent** deterministic models **may be different**!

You should always try to allow your formulation
to be as flexible as possible!

# Dynamic Decisions and Robust Dynamic Optimization

# Dynamic (Robust) Optimization

$x$ chosen $\quad \mapsto \quad z$ revealed $\quad \mapsto \quad y(x, z)$ chosen

# Dynamic (Robust) Optimization

$x$ chosen $\quad \mapsto \quad z$ revealed $\quad \mapsto \quad y(x, z)$ chosen

Stochastic model:

$$\min_{x} \; \mathbb{E}_{z}\left[ \min_{y(x,z)} f(x, y, z) \right]$$

Robust model:

$$\min_{x} \max_{z \in \mathcal{U}} \min_{y(x,z)} \; f(x, y, z)$$

# Dynamic (Robust) Optimization

$$x \text{ chosen} \quad \mapsto \quad z \text{ revealed} \quad \mapsto \quad y(x, z) \text{ chosen}$$

Stochastic model:

$$\min_{x} \ \mathbb{E}_{z}\left[\min_{y(x,z)} f(x, y, z)\right]$$

Robust model:

$$\min_{x} \max_{z \in \mathcal{U}} \min_{y(x,z)} \ f(x, y, z)$$

- Solve problems via Dynamic Programming:
  - Given $x, z \quad \rightarrow \quad$ find $y^\star(x, z) \quad \rightarrow \quad$ find $x^\star$
  - Bellman principle: $y^\star$ optimal for any given $x, z$

# Dynamic (Robust) Optimization

$$x \text{ chosen} \quad \mapsto \quad z \text{ revealed} \quad \mapsto \quad y(x, z) \text{ chosen}$$

Stochastic model:

$$\min_{x} \; \mathbb{E}_z \left[ \min_{y(x,z)} f(x, y, z) \right]$$

Robust model:

$$\min_{x} \max_{z \in \mathcal{U}} \min_{y(x,z)} \; f(x, y, z)$$

- Solve problems via Dynamic Programming:
  - Given $x, z$ → find $y^\star(x, z)$ → find $x^\star$
  - Bellman principle: $y^\star$ optimal for any given $x, z$

1. Properly writing a **robust** DP
2. Tractable approximations with decision rules
3. A subtle point: is Bellman optimality really **necessary**?
   - If not, what to replace it with?
   - Why is this relevant?
4. Some applications

# A simple motivating example

Consider the following *deterministic* inventory management problem:

$$\underset{\{x_t\}_{t=1}^{T}}{\text{minimize}} \quad \sum_{t=1}^{T} \left( \overbrace{c_t x_t}^{\text{ordering cost}} + \overbrace{h_t (y_{t+1})^+}^{\text{holding cost}} + \overbrace{b_t (-y_{t+1})^+}^{\text{backlog cost}} \right)$$

$$\begin{aligned} \text{s.t.} \quad & y_{t+1} = y_t + x_t - d_t,\ \forall\, t, && \text{(Stock balance)} \\ & L_t \le x_t \le H_t,\ \forall\, t, && \text{(Min/max order size)} \\ & y_1 = a\,, && \text{(Initial stock level)} \end{aligned}$$

# A simple motivating example

Consider the following *deterministic* inventory management problem:

$$\underset{\{x_t\}_{t=1}^{T}}{\text{minimize}} \sum_{t=1}^{T} \left( \overbrace{c_t x_t}^{\text{ordering cost}} + \overbrace{h_t (y_{t+1})^+}^{\text{holding cost}} + \overbrace{b_t (-y_{t+1})^+}^{\text{backlog cost}} \right)$$

$$\begin{aligned} \text{s.t.} \quad & y_{t+1} = y_t + x_t - d_t, \ \forall\, t, && \text{(Stock balance)} \\ & L_t \le x_t \le H_t, \ \forall\, t, && \text{(Min/max order size)} \\ & y_1 = a\,, && \text{(Initial stock level)} \end{aligned}$$

where

- $x_t$ is number of goods ordered at time $t$ and received at $t+1$
- $y_t$ is number of goods in stock at beginning of time $t$
- $d_t$ is demand during period $t$
- $a$ is the initial inventory

# A simple motivating example

Consider the following *deterministic* inventory management problem:

$$\underset{\{x_t\}_{t=1}^{T}}{\text{minimize}} \sum_{t=1}^{T} \left( \overbrace{c_t x_t}^{\text{ordering cost}} + \overbrace{h_t (y_{t+1})^+}^{\text{holding cost}} + \overbrace{b_t (-y_{t+1})^+}^{\text{backlog cost}} \right)$$

$$\begin{aligned} \text{s.t.} \quad & y_{t+1} = y_t + x_t - d_t, \ \forall\, t, && \text{(Stock balance)} \\ & L_t \le x_t \le H_t, \ \forall\, t, && \text{(Min/max order size)} \\ & y_1 = a\,, && \text{(Initial stock level)} \end{aligned}$$

What if future demands known to reside in an **uncertainty set** $\mathcal{U}$?

$$d := (d_1, d_2, \ldots, d_T) \in \mathcal{U} \subseteq \mathbb{R}^T$$

# A simple motivating example

Consider the following *deterministic* inventory management problem:

$$\underset{\{x_t\}_{t=1}^{T}}{\text{minimize}} \quad \sum_{t=1}^{T} \left( \overbrace{c_t x_t}^{\text{ordering cost}} + \overbrace{h_t (y_{t+1})^+}^{\text{holding cost}} + \overbrace{b_t (-y_{t+1})^+}^{\text{backlog cost}} \right)$$

$$\begin{aligned} \text{s.t.} \quad & y_{t+1} = y_t + x_t - d_t, \ \forall\, t, && \text{(Stock balance)} \\ & L_t \le x_t \le H_t, \ \forall\, t, && \text{(Min/max order size)} \\ & y_1 = a\,, && \text{(Initial stock level)} \end{aligned}$$

What if future demands known to reside in **uncertainty set** $\mathcal{U}$?

$$d := (d_1, d_2, \ldots, d_T) \in \mathcal{U} \subseteq \mathbb{R}^T$$

Ordering policies can depend on revealed demands:

$$x_t(d_{[t-1]}), \ \text{where } d_{[t-1]} := (d_1, d_2, \ldots, d_{t-1}) \in \mathbb{R}^{t-1}.$$

# Robust Dynamic Programming Formulation

Our dynamic decision problem can also be written:

$$\min_{L_1 \le x_1 \le H_1} \Big[ c_1 x_1 + \max_{d_1 \in \mathcal{U}_1(\emptyset)} \Big[ h_1(y_2)^+ + b_1(-y_2)^+$$
$$+ \min_{L_2 \le x_2 \le H_2} \Big[ c_2 x_2 + \max_{d_2 \in \mathcal{U}_2(d_1)} \Big[ h_2(y_3)^+ + b_2(-y_3)^+ + \dots$$
$$+ \min_{L_T \le x_T \le H_T} \Big[ c_T x_T + \max_{d_T \in \mathcal{U}_T(d_{[T-1]})} [h_T(y_{T+1})^+ + b_T(-y_{T+1})^+] \Big] \dots \Big]$$

where:

$$y_{t+1} := y_t + x_t - d_t$$
$$\mathcal{U}_t(d_{[t-1]}) := \Big\{ d \in \mathbb{R} \, : \, \exists \, z \in \mathbb{R}^{T-t} \text{ such that } [d_{[t-1]};\ d\ ; z] \in \mathcal{U} \Big\}$$

# Robust Dynamic Programming Formulation

Our dynamic decision problem can also be written:

$$\min_{L_1 \le x_1 \le H_1} \Big[ c_1 x_1 + \max_{d_1 \in \mathcal{U}_1(\emptyset)} \Big[ h_1 (y_2)^+ + b_1(-y_2)^+$$
$$+ \min_{L_2 \le x_2 \le H_2} \Big[ c_2 x_2 + \max_{d_2 \in \mathcal{U}_2(d_1)} \Big[ h_2 (y_3)^+ + b_2(-y_3)^+ + \ldots$$
$$+ \min_{L_T \le x_T \le H_T} \Big[ c_T x_T + \max_{d_T \in \mathcal{U}_T(d_{[T-1]})} [ h_T (y_{T+1})^+ + b_T(-y_{T+1})^+ ] \Big] \ldots \Big]$$

where:

$$y_{t+1} := y_t + x_t - d_t$$
$$\mathcal{U}_t(d_{[t-1]}) := \Big\{ d \in \mathbb{R} \, : \, \exists\, z \in \mathbb{R}^{T-t} \text{ such that } [d_{[t-1]};\ d\ ; z] \in \mathcal{U} \Big\}$$

1. Nested min-max problems
2. Explicit rule for “conditioning”: *projection* of uncertainty set

# Bellman Principle; Robust DP Recursions

- The **state** of the system at time $t$:

$$S_t := \left[y_t;\ d_{[t-1]}\right] = \left[y_t;\ d_1\ d_2;\ \ldots\ ; d_{t-1}\right] \in \mathbb{R}^{\mathsf{T}}$$

# Bellman Principle; Robust DP Recursions

- The **state** of the system at time $t$:

$$S_t := \left[y_t;\ d_{[t-1]}\right] = \left[y_t;\ d_1\ d_2;\ \ldots\ ; d_{t-1}\right] \in \mathbb{R}^{T}$$

- Value function $J_t^\star(S_t)$ given by:

$$J_t^\star(S_t) = \min_{L_t \le x_t \le H_t} \left[ c_t x_t + \max_{d_t \in \mathcal{U}_t(d_{[t-1]})} \left[ h_t(y_{t+1})^+ + b_t(-y_{t+1})^+ + J_{t+1}^\star(S_{t+1}) \right] \right]$$

# Bellman Principle; Robust DP Recursions

- The **state** of the system at time $t$:

$$S_t := \left[y_t;\ d_{[t-1]}\right] = \left[y_t;\ d_1\ d_2;\ \ldots\ ; d_{t-1}\right] \in \mathbb{R}^{\mathsf{T}}$$

- Value function $J_t^\star(S_t)$ given by:

$$J_t^\star(S_t) = \min_{L_t \le x_t \le H_t} \left[ c_t x_t + \max_{d_t \in \mathcal{U}_t(d_{[t-1]})} \left[ h_t(y_{t+1})^+ + b_t(-y_{t+1})^+ + J_{t+1}^\star(S_{t+1}) \right] \right]$$

Observations:

1. General $\mathcal{U} \longrightarrow$ high-dimensional $S_t \longrightarrow$ **curse of dimensionality**

# Bellman Principle; Robust DP Recursions

- The **state** of the system at time $t$:

$$S_t := \left[y_t;\ d_{[t-1]}\right] = \left[y_t;\ d_1\ d_2;\ \ldots\ ; d_{t-1}\right] \in \mathbb{R}^T$$

- Value function $J_t^\star(S_t)$ given by:

$$J_t^\star(S_t) = \min_{L_t \le x_t \le H_t} \left[ c_t x_t + \max_{d_t \in \mathcal{U}_t(d_{[t-1]})} \left[ h_t(y_{t+1})^+ + b_t(-y_{t+1})^+ + J_{t+1}^\star(S_{t+1}) \right] \right]$$

Observations:

1. General $\mathcal{U} \longrightarrow$ high-dimensional $S_t \longrightarrow$ **curse of dimensionality**
2. When $\mathcal{U}$ has special structure, can reduce state space

# Bellman Principle; Robust DP Recursions

- The **state** of the system at time $t$:

$$S_t := \left[y_t;\ d_{[t-1]}\right] = \left[y_t;\ d_1\ d_2;\ \ldots\ ; d_{t-1}\right] \in \mathbb{R}^{\mathsf{T}}$$

- Value function $J_t^\star(S_t)$ given by:

$$J_t^\star(S_t) = \min_{L_t \le x_t \le H_t} \left[ c_t x_t + \max_{d_t \in \mathcal{U}_t(d_{[t-1]})} \left[ h_t(y_{t+1})^+ + b_t(-y_{t+1})^+ + J_{t+1}^\star(S_{t+1}) \right] \right]$$

Observations:

1. General $\mathcal{U} \longrightarrow$ high-dimensional $S_t \longrightarrow$ **curse of dimensionality**
2. When $\mathcal{U}$ has special structure, can reduce state space

$$\mathcal{U}_{\text{box}} = \left\{ d\ :\ \underline{d}_t \le d_t \le \bar{d}_t] \right\} \quad \rightarrow \quad S_t = y_t$$

$$\mathcal{U}_{\text{budget}} = \left\{ d : \exists\, z, \|z\|_\infty \le 1, \|z\|_1 \le \Gamma, d_t = \bar{d}_t + \hat{d}_t z_t \right\} \rightarrow S_t = \left[ y_t, \sum_{\tau=1}^{t-1} |z_\tau| \right]^{\mathsf{T}}$$

# Bellman Principle; Robust DP Recursions

- The **state** of the system at time $t$:

$$S_t := \left[y_t;\ d_{[t-1]}\right] = \left[y_t;\ d_1\ d_2;\ \ldots\ ; d_{t-1}\right] \in \mathbb{R}^T$$

- Value function $J_t^\star(S_t)$ given by:

$$J_t^\star(S_t) = \min_{L_t \le x_t \le H_t} \left[ c_t x_t + \max_{d_t \in \mathcal{U}_t(d_{[t-1]})} \left[ h_t(y_{t+1})^+ + b_t(-y_{t+1})^+ + J_{t+1}^\star(S_{t+1}) \right] \right]$$

Observations:

1. General $\mathcal{U} \longrightarrow$ high-dimensional $S_t \longrightarrow$ **curse of dimensionality**
2. When $\mathcal{U}$ has special structure, can reduce state space

- Reduce computational burden
- Prove structural results, comparative statics

$$x_t^\star(y) = \min\left(H_t, \max(L_t,\ \theta_t - y)\right) \qquad \text{(modified) } \textit{base-stock policy}$$

# Tractable Approximations Via Decision Rules

Back to our basic dynamic robust model:

$$\min_{x} \max_{z \in \mathcal{U}} \min_{y(z)} \; f(x, y, z)$$

- Finding Bellman-optimal rules $y^{\star}(z)$ generally intractable

# Tractable Approximations Via Decision Rules

Back to our basic dynamic robust model:

$$\min_{x} \max_{z \in \mathcal{U}} \min_{y(z)} \; f(x, y, z)$$

- Finding Bellman-optimal rules $y^\star(z)$ generally intractable
- Pragmatic idea: let's focus on some "simple" decision rules that we can compute
- For instance, with a **static** $y(z) = y$, could just apply all our previous machinery

# Tractable Approximations Via Decision Rules

Back to our basic dynamic robust model:

$$\min_{x} \max_{z \in \mathcal{U}} \min_{y(z)} \; f(x, y, z)$$

- Finding Bellman-optimal rules $y^\star(z)$ generally intractable
- Pragmatic idea: let's focus on some "simple" decision rules that we can compute
- For instance, with a **static** $y(z) = y$, could just apply all our previous machinery
- Ben-Tal et. al: **Linear Decision Rules**
  - Suppose we have a constraint

    $$(\bar{a} + Pz)^{\mathsf{T}} x + d^{\mathsf{T}} y(z) \le b, \quad \forall\, z \in \mathcal{U}$$

    where $y(z)$ is dynamically adjustable
  - A linear (affine) form $y = u + Vz$ would lead to the problem:

    $$\bar{a}^{\mathsf{T}} x + d^{\mathsf{T}} u + (P^{\mathsf{T}} x + V^{\mathsf{T}} d)^{\mathsf{T}} z \le b, \quad \forall z \in \mathcal{U}.$$

    Constraint **linear** in decisions $x, u, V$ and uncertainty $z$, so all previous results apply!

# Tractable Approximations Via Decision Rules

Back to our basic dynamic robust model:

$$\min_{x} \max_{z \in \mathcal{U}} \min_{y(z)} \ f(x, y, z)$$

- Finding Bellman-optimal rules $y^\star(z)$ generally intractable
- Pragmatic idea: let's focus on some "simple" decision rules that we can compute
- For instance, with a **static** $y(z) = y$, could just apply all our previous machinery
- Ben-Tal et. al: **Linear Decision Rules**
  - Suppose we have a constraint

    $$(\bar{a} + Pz)^\mathsf{T} x + d^\mathsf{T} y(z) \le b, \quad \forall\, z \in \mathcal{U}$$

    where $y(z)$ is dynamically adjustable
  - A linear (affine) form $y = u + Vz$ would lead to the problem:

    $$\bar{a}^\mathsf{T} x + d^\mathsf{T} u + (P^\mathsf{T} x + V^\mathsf{T} d)^\mathsf{T} z \le b, \quad \forall z \in \mathcal{U}.$$

    Constraint **linear** in decisions $x, u, V$ and uncertainty $z$, so all previous results apply!
- So how to apply these **static** or **linear** rules in a real problem?

# Implementation and Potential Pitfalls

Recall our inventory problem. The **deterministic** version can be reformulated as an LP:

$$
\begin{aligned}
\underset{x_t, y_t, s_t^+, s_t^-}{\text{minimize}} \quad & \sum_{t=1}^{T} \left( c_t x_t + h_t s_t^+ + b_t s_t^- \right) \\
\text{s.t.} \quad & s_t^+ \geq 0,\; s_t^- \geq 0 \,,\; \forall\, t, \\
& s_t^+ \geq y_{t+1} \,,\; \forall\, t, \\
& s_t^- \geq -y_{t+1} \,,\; \forall\, t, \\
& y_{t+1} = y_t + x_t - d_t \,,\; \forall\, t, \\
& L_t \leq x_t \leq H_t \,,\; \forall\, t,
\end{aligned}
$$

where

- $s_t^+$ : physical inventory held at end of period $t$
- $s_t^-$ : backlogged customer demand at end of period $t$

# Implementation and Potential Pitfalls

Recall our inventory problem. The **deterministic** version can be reformulated as an LP:

$$
\begin{aligned}
\underset{x_t, y_t, s_t^+, s_t^-}{\text{minimize}} \quad & \sum_{t=1}^{T} \left( c_t x_t + h_t s_t^+ + b_t s_t^- \right) \\
\text{s.t.} \quad & s_t^+ \geq 0,\; s_t^- \geq 0 \,,\; \forall\, t, \\
& s_t^+ \geq y_{t+1} \,,\; \forall\, t, \\
& s_t^- \geq -y_{t+1} \,,\; \forall\, t, \\
& y_{t+1} = y_t + x_t - d_t \,,\; \forall\, t, \\
& L_t \leq x_t \leq H_t \,,\; \forall\, t,
\end{aligned}
$$

where

- $s_t^+$ : physical inventory held at end of period $t$
- $s_t^-$ : backlogged customer demand at end of period $t$

What if demand known to reside in an **uncertainty set** $\mathcal{U}$?

$$d := (d_1, d_2, \ldots, d_T) \in \mathcal{U} \subseteq \mathbb{R}^T$$

# Naïve Robustification

Consider a naïve robust optimization model:

$$
\begin{aligned}
\underset{x_t, y_t, s_t^+, s_t^-}{\text{minimize}} \quad & \sum_{t=1}^{T} \left( c_t x_t + h_t s_t^+ + b_t s_t^- \right) \\
\text{s.t.} \quad & s_t^+ \geq 0,\ s_t^- \geq 0\,,\ \forall\, t \\
& s_t^+ \geq y_{t+1}\,,\ \forall\, t \\
& s_t^- \geq -y_{t+1}\,,\ \forall\, t \\
& y_{t+1} = y_t + x_t - d_t\,,\ \forall\, t,\ \forall\, d \in \mathcal{U} \\
& L_t \leq x_t \leq H_t\,,\ \forall\, t
\end{aligned}
$$

# Naïve Robustification

Consider a naïve robust optimization model:

$$
\begin{aligned}
\underset{x_t, y_t, s_t^+, s_t^-}{\text{minimize}} \quad & \sum_{t=1}^{T} \left( c_t x_t + h_t s_t^+ + b_t s_t^- \right) \\
\text{s.t.} \quad & s_t^+ \geq 0,\ s_t^- \geq 0\,,\ \forall\, t \\
& s_t^+ \geq y_{t+1}\,,\ \forall\, t \\
& s_t^- \geq -y_{t+1}\,,\ \forall\, t \\
& y_{t+1} = y_t + x_t - d_t\,,\ \forall\, t,\ \forall\, d \in \mathcal{U} \\
& L_t \leq x_t \leq H_t\,,\ \forall\, t
\end{aligned}
$$

Unfortunately, this is **infeasible** even when $\mathcal{U} = \{d^{(1)}, d^{(2)}\}$:

$$
\left\{ \begin{array}{l} y_{t+1} = y_t + x_t - d_t^{(1)} \\ y_{t+1} = y_t + x_t - d_t^{(2)} \end{array} \right\} \Rightarrow\ d_t^{(1)} = d_t^{(2)}
$$

Problem arises due to "=" constraint!

# A less naïve robustification

Robustify an alternate linear programming formulation:

$$
\begin{aligned}
\underset{x_t, s_t^+, s_t^-}{\text{minimize}} \quad & \sum_t \left( c_t x_t + h_t s_t^+ + b_t s_t^- \right) \\
\text{s.t.} \quad & s_t^+ \geq 0,\ s_t^- \geq 0,\ \forall t, \\
& s_t^+ \geq y_1 + \sum_{t'=1}^{T} (x_{t'} - d_{t'}),\ \forall t, \\
& s_t^- \geq -y_1 + \sum_{t'=1}^{T} (d_{t'} - x_{t'}),\ \forall t, \\
& L_t \leq x_t \leq H_t,\ \ \forall t\ ,
\end{aligned}
$$

where we simply replace $y_{t+1} := y_1 + \sum_{t'=1}^{T} (x_{t'} - d_{t'})$.

# A less naïve robustification

Robustify an alternate linear programming formulation:

$$
\begin{aligned}
\underset{x_t, s_t^+, s_t^-}{\text{minimize}} \quad & \sum_{t=1}^{T} \left( c_t x_t + h_t s_t^+ + b_t s_t^- \right) \\
\text{s.t.} \quad & s_t^+ \geq 0,\ s_t^- \geq 0,\ \forall t, \\
& s_t^+ \geq y_1 + \sum_{t'=1}^{T} (x_{t'} - d_{t'}),\ \forall t,\ \forall\, d \in \mathcal{U} \\
& s_t^- \geq -y_1 + \sum_{t'=1}^{T} (d_{t'} - x_{t'}),\ \forall t,\ \forall\, d \in \mathcal{U} \\
& L_t \leq x_t \leq H_t, \forall\, t\ .
\end{aligned}
$$

# A less naïve robustification

Robustify an alternate linear programming formulation:

$$
\begin{aligned}
\underset{x_t, s_t^+, s_t^-}{\text{minimize}} \quad & \sum_{t=1}^{T} \left( c_t x_t + h_t s_t^+ + b_t s_t^- \right) \\
\text{s.t.} \quad & s_t^+ \geq 0,\ s_t^- \geq 0,\ \forall t, \\
& s_t^+ \geq y_1 + \sum_{t'=1}^{T} (x_{t'} - d_{t'}),\ \forall t,\ \forall\, d \in \mathcal{U} \\
& s_t^- \geq -y_1 + \sum_{t'=1}^{T} (d_{t'} - x_{t'}),\ \forall t,\ \forall\, d \in \mathcal{U} \\
& L_t \leq x_t \leq H_t, \forall\, t\ .
\end{aligned}
$$

**Q:** If orders $x_t$ are **static** (i.e., fixed $t = 0$), should $(s_t^+, s_t^-)$ also be static?

# A less naïve robustification

Robustify an alternate linear programming formulation:

$$
\begin{aligned}
\underset{x_t, s_t^+, s_t^-}{\text{minimize}} \quad & \sum_{t=1}^{T} \left( c_t x_t + h_t s_t^+ + b_t s_t^- \right) \\
\text{s.t.} \quad & s_t^+ \geq 0,\ s_t^- \geq 0,\ \forall t, \\
& s_t^+ \geq y_1 + \sum_{t'=1}^{T} (x_{t'} - d_{t'}),\ \forall t,\ \forall\, d \in \mathcal{U} \\
& s_t^- \geq -y_1 + \sum_{t'=1}^{T} (d_{t'} - x_{t'}),\ \forall t,\ \forall\, d \in \mathcal{U} \\
& L_t \leq x_t \leq H_t, \forall\, t\ .
\end{aligned}
$$

**Q:** If orders $x_t$ are **static** (i.e., fixed $t = 0$), should $(s_t^+, s_t^-)$ also be static?

**A:** No, that would be **unnecessarily conservative**!
**Auxiliary (i.e., "reformulation") variables should be fully adjustable, even under static "implementable" decisions.**

# Linear Decision Rules

- Take both **ordering policies** and **auxiliary variables** to depend *linearly* on demands

$$
\begin{aligned}
x_t\left(d_{[t-1]}\right) &= x_t^0 + X_t d_{[t-1]} \\
s_t^+\left(d_{[t-1]}\right) &= s_t^+ + S_t^+ d_{[t-1]} \\
s_t^-\left(d_{[t-1]}\right) &= s_t^- + S_t^- d_{[t-1]}
\end{aligned}
$$

# Linear Decision Rules

- Take both **ordering policies** and **auxiliary variables** to depend *linearly* on demands

$$
\begin{aligned}
x_t\left(d_{[t-1]}\right) &= x_t^0 + X_t d_{[t-1]} \\
s_t^+\left(d_{[t-1]}\right) &= s_t^+ + S_t^+ d_{[t-1]} \\
s_t^-\left(d_{[t-1]}\right) &= s_t^- + S_t^- d_{[t-1]}
\end{aligned}
$$

- The **Robust Counterpart** problem becomes:

$$
\begin{aligned}
\min_{\mathcal{X}} \quad & \max_{d \in \mathcal{U}} \sum_{t=1}^{T} c_t \cdot (x_t^0 + X_t d) + h_t \cdot (s_t^+ + S_t^+ d) + b_t \cdot (s_t^- + S_t^- d) \\
\text{s.t.} \quad & s_t^+ + S_t^+ d \geq 0, \quad s_t^- + S_t^- d \geq 0, \ \forall\, d \in \mathcal{U} \\
& s_t^+ + S_t^+ d \geq y_1 + \sum_{\tau=1}^{T} (x_\tau^0 + X_\tau d_{[\tau-1]} - d_\tau), \ \forall\, d \in \mathcal{U}, \\
& s_t^- + S_t^- d \geq -y_1 - \sum_{\tau=1}^{T} (x_\tau^0 + X_\tau d_{[\tau-1]} - d_\tau), \ \forall\, d \in \mathcal{U}, \\
& L_t \leq x_t + X_t d \leq H_t \,, \ \forall\, d \in \mathcal{U},
\end{aligned}
$$

- **Decision variables:** coefficients $\mathcal{X} = \left\{x_t^0, X_t, s_t^+, S_t^+, s_t^-, S_t^-\right\}_{t=1}^{T}$
- **Two** layers of sub-optimality: **policies** *and* **auxiliary variables**; any good?

# Empirical Performance: Ben-Tal et al. ('04, '09)

| $\rho$ (%) | **OPT** | **Linear (Gap)** | **Static (Gap)** |
|---|---|---|---|
| 10 | 13531.8 | 13531.8 (+0.0%) | 15033.4 (+11.1%) |
| 20 | 15063.5 | 15063.5 (+0.0%) | 18066.7 (+19.9%) |
| 30 | 16595.3 | 16595.3 (+0.0%) | 21100.0 (+27.1%) |
| 40 | 18127.0 | 18127.0 (+0.0%) | 24300.0 (+34.1%) |
| 50 | 19658.7 | 19658.7 (+0.0%) | 27500.0 (+39.9%) |
| 60 | 21190.5 | 21190.5 (+0.0%) | 30700.0 (+44.9%) |
| 70 | 22722.2 | 22722.2 (+0.0%) | 33960.0 (+49.5%) |

# Empirical Performance: Ben-Tal et al. ('04, '09)

| $\rho$ (%) | **OPT** | **Linear (Gap)** | **Static (Gap)** |
|---|---|---|---|
| 10 | 13531.8 | 13531.8 (+0.0%) | 15033.4 (+11.1%) |
| 20 | 15063.5 | 15063.5 (+0.0%) | 18066.7 (+19.9%) |
| 30 | 16595.3 | 16595.3 (+0.0%) | 21100.0 (+27.1%) |
| 40 | 18127.0 | 18127.0 (+0.0%) | 24300.0 (+34.1%) |
| 50 | 19658.7 | 19658.7 (+0.0%) | 27500.0 (+39.9%) |
| 60 | 21190.5 | 21190.5 (+0.0%) | 30700.0 (+44.9%) |
| 70 | 22722.2 | 22722.2 (+0.0%) | 33960.0 (+49.5%) |

Theorem ( Bertsimas, I., Parrilo 2010, I., Sharma & Sviridenko 2013 )

*For any* **convex** *order costs* $c_t(\cdot)$ *and inventory costs* $h_t(\cdot)$, *affine orders* $x_t(d_{[t-1]})$ *and affine auxiliary variables* $s_t^{+,-}(d_{[t-1]})$ *generate the optimal worst-case cost.*

# Empirical Performance: Ben-Tal et al. ('04, '09)

| $\rho$ (%) | **OPT** | **Linear (Gap)** | **Static (Gap)** |
|---|---|---|---|
| 10 | 13531.8 | 13531.8 (+0.0%) | 15033.4 (+11.1%) |
| 20 | 15063.5 | 15063.5 (+0.0%) | 18066.7 (+19.9%) |
| 30 | 16595.3 | 16595.3 (+0.0%) | 21100.0 (+27.1%) |
| 40 | 18127.0 | 18127.0 (+0.0%) | 24300.0 (+34.1%) |
| 50 | 19658.7 | 19658.7 (+0.0%) | 27500.0 (+39.9%) |
| 60 | 21190.5 | 21190.5 (+0.0%) | 30700.0 (+44.9%) |
| 70 | 22722.2 | 22722.2 (+0.0%) | 33960.0 (+49.5%) |

Theorem ( Bertsimas, I., Parrilo 2010, I., Sharma & Sviridenko 2013 )

*For any* **convex** *order costs $c_t(\cdot)$ and inventory costs $h_t(\cdot)$, affine orders $x_t(d_{[t-1]})$ and affine auxiliary variables $s_t^{+,-}(d_{[t-1]})$ generate the optimal worst-case cost.*

**Why is this relevant?**

1. *Insight*: orders only depend on **backlogged** demand
2. *Computational*: if $c_t, h_t$ piecewise affine ($m$ pieces), must solve $\mathcal{O}(m \cdot T^2)$ LP
3. *Extensions*: can embed decisions at $t = 0$ (e.g., capacities, order pre-commitments)
4. *Robust dynamic* **critically different** *from stochastic dynamic*
   - Stochastic model with complete $\mathbb{P}$ requires "complex" policies; affine very suboptimal
   - Robust model admits a very "simple" class of optimal policies

# Bellman Optimality in Stochastic and Robust Models

"Nature" reveals $z$ $\mapsto$ DM chooses $y(z)$

Stochastic model:

$$J^\star_{\text{sto}} = \mathbb{E}_z\left[\min_{y(z)} \; f(y, z)\right]$$

Robust model:

$$J^\star_{\text{rob}} = \max_{z \in \mathcal{U}} \min_{y(z)} \; f(y, z)$$

# Bellman Optimality in Stochastic and Robust Models

"Nature" reveals $z$ $\mapsto$ DM chooses $y(z)$

Stochastic model:

$$J^\star_{\text{sto}} = \mathbb{E}_z \left[ \min_{y(z)} \; f(y, z) \right]$$

Robust model:

$$J^\star_{\text{rob}} = \max_{z \in \mathcal{U}} \min_{y(z)} \; f(y, z)$$

- Solve problems via Dynamic Programming:
  - Given $z$, find $y^\star(z) \in \arg\min_y f(y, z)$
  - Bellman principle: $y^\star(z)$ optimal for any $z$

# Bellman Optimality in Stochastic and Robust Models

"Nature" reveals $z$ $\mapsto$ DM chooses $y(z)$

Stochastic model:

$$J^\star_{\text{sto}} = \mathbb{E}_z\left[\min_{y(z)} \; f(y,z)\right]$$

Robust model:

$$J^\star_{\text{rob}} = \max_{z\in\mathcal{U}} \min_{y(z)} \; f(y,z)$$

- Solve problems via Dynamic Programming:
  - Given $z$, find $y^\star(z) \in \arg\min_y f(y,z)$
  - Bellman principle: $y^\star(z)$ optimal for any $z$

**Question:** Is Bellman optimality for $y$ really **necessary?**

# Bellman Optimality in Stochastic and Robust Models

"Nature" reveals $z$ $\mapsto$ DM chooses $y(z)$

Stochastic model:

$$J^\star_{\text{sto}} = \mathbb{E}_z \left[ \min_{y(z)} f(y,z) \right]$$

Robust model:

$$J^\star_{\text{rob}} = \max_{z \in \mathcal{U}} \min_{y(z)} f(y,z)$$

- Solve problems via Dynamic Programming:
  - Given $z$, find $y^\star(z) \in \arg\min_y f(y,z)$
  - Bellman principle: $y^\star(z)$ optimal for any $z$

**Question:** Is Bellman optimality for $y$ really **necessary?**

- For **stochastic model**, Bellman-optimally **necessary** to obtain $J^\star_{\text{sto}}$
- For **robust model**, Bellman-optimally **sufficient**, but **not necessary** to obtain $J^\star_{\text{rob}}$

# Bellman Optimality in Stochastic and Robust Models

"Nature" reveals $z$ $\mapsto$ DM chooses $y(z)$

Stochastic model:

$$J^\star_{\text{sto}} = \mathbb{E}_z \left[ \min_{y(z)} \; f(y, z) \right]$$

Robust model:

$$J^\star_{\text{rob}} = \max_{z \in \mathcal{U}} \min_{y(z)} \; f(y, z)$$

- Solve problems via Dynamic Programming:
  - Given $z$, find $y^\star(z) \in \arg\min_y f(y, z)$
  - Bellman principle: $y^\star(z)$ optimal for any $z$

**Question:** Is Bellman optimality for $y$ really **necessary?**

- For **stochastic model**, Bellman-optimally **necessary** to obtain $J^\star_{\text{sto}}$
- For **robust model**, Bellman-optimally **sufficient**, but **not necessary** to obtain $J^\star_{\text{rob}}$
  - Any policy $y^{\text{wc}}$ from the set

$$\mathcal{Y}^{\text{wc}} := \left\{ y : \mathcal{U} \to \mathbb{R}^m \; : \; f\big(y(z), z\big) \le J^\star_{\text{rob}}, \; \forall z \in \mathcal{U} \right\}.$$

will be "optimal" in the robust problem, i.e., $\max_{z \in \mathcal{U}} f(y^{\text{wc}}(z), z) = J^\star_{\text{rob}}$

# Bellman Optimality in Stochastic and Robust Models

$$\text{"Nature" reveals } z \quad \mapsto \quad \text{DM chooses } y(z)$$

Stochastic model:

$$J^\star_{\text{sto}} = \mathbb{E}_z \left[ \min_{y(z)} \; f(y, z) \right]$$

Robust model:

$$J^\star_{\text{rob}} = \max_{z \in \mathcal{U}} \min_{y(z)} \; f(y, z)$$

- Solve problems via Dynamic Programming:
  - Given $z$, find $y^\star(z) \in \arg\min_y f(y, z)$
  - Bellman principle: $y^\star(z)$ optimal for any $z$

**Question:** Is Bellman optimality for $y$ really **necessary?**

- For **stochastic model**, Bellman-optimally **necessary** to obtain $J^\star_{\text{sto}}$
- For **robust model**, Bellman-optimally **sufficient**, but **not necessary** to obtain $J^\star_{\text{rob}}$
  - Any policy $y^{\text{wc}}$ from the set

$$\mathcal{Y}^{\text{wc}} := \left\{ y : \mathcal{U} \to \mathbb{R}^m \;:\; f\big(y(z), z\big) \le J^\star_{\text{rob}}, \; \forall z \in \mathcal{U} \right\}.$$

    will be "optimal" in the robust problem, i.e., $\max_{z \in \mathcal{U}} f(y^{\text{wc}}(z), z) = J^\star_{\text{rob}}$
- The set of **worst-case optimal policies** $\mathcal{Y}^{\text{wc}}$ is non-empty and degenerate

# Implications for Robust Dynamic Models

1. Bellman optimality **not necessary**; **worst-case optimality** necessary
   - Introduces *degeneracy* in policies/decisions

2. This degeneracy is typical for robust multi-stage problems
   ("If adversary does not play optimally, you don't have to, either...")

3. Critically different from stochastic problems

4. A blessing: may allow finding policies with **simple structure**
   - e.g., affine...

5. A curse: may yield Pareto inefficiencies in the decision process

6. Worst-case optimal policies **must be implemented with resolving**

# Another Caveat...

# Are Robust Solutions “Efficient”?

$$\max_{x \in \mathcal{X}} \min_{u \in \mathcal{U}} u^{\mathsf{T}} x$$

- Feasible set of solutions $\mathcal{X} = \left\{x \in \mathbb{R}^n : Ax \leq b\right\}$
- Uncertainty set of objective coefficients $\mathcal{U} = \{u \in \mathbb{R}^n : Du \geq d\}$

# Are Robust Solutions "Efficient"?

$$\max_{x \in \mathcal{X}} \quad \min_{u \in \mathcal{U}} \; u^\intercal x$$

- Feasible set of solutions $\mathcal{X} = \left\{x \in \mathbb{R}^n : Ax \leq b\right\}$

- Uncertainty set of objective coefficients $\mathcal{U} = \{u \in \mathbb{R}^n : Du \geq d\}$

- Classical RO framework finds the **optimal value** $J^\star_{\mathbf{RO}}$ and a point $x$ from the **set of robustly optimal solutions** $x \in X^{\mathrm{RO}}$:

$$X^{\mathrm{RO}} = \left\{x \in \mathcal{X} : \exists\, y \geq 0 \text{ such that } D^\intercal y = x, \quad y^\intercal d \geq J^\star_{\mathrm{RO}}\right\}$$

- $x \in X^{\mathrm{RO}} \Rightarrow$ no other solution exists with higher **worst-case** objective value $u^\intercal x$

# Are Robust Solutions "Efficient"?

$$\max_{x \in \mathcal{X}} \quad \min_{u \in \mathcal{U}} u^\top x$$

- Feasible set of solutions $\mathcal{X} = \left\{ x \in \mathbb{R}^n : Ax \leq b \right\}$
- Uncertainty set of objective coefficients $\mathcal{U} = \{ u \in \mathbb{R}^n : Du \geq d \}$
- Classical RO framework finds the **optimal value** $J^\star_{\mathbf{RO}}$ and a point $x$ from the **set of robustly optimal solutions** $x \in X^{\mathrm{RO}}$:

$$X^{\mathrm{RO}} = \left\{ x \in \mathcal{X} : \exists\, y \geq 0 \text{ such that } D^\top y = x, \quad y^\top d \geq J^\star_{\mathrm{RO}} \right\}$$

- $x \in X^{\mathrm{RO}} \Rightarrow$ no other solution exists with higher **worst-case** objective value $u^\top x$
- *What if an uncertainty scenario materializes that does not correspond to the worst-case?*
- *Are there any guarantees that no other solution $\bar{x}$ exists that, apart from protecting us from worst-case scenarios, also performs better overall, under all possible uncertainty realizations?*

# Pareto Robustly Optimal solutions (I. & Trichakis 2014)

$$\max_{x \in \mathcal{X}} \quad \min_{u \in \mathcal{U}} u^\mathsf{T} x \tag{3}$$

**Definition**

A solution $x$ is called a **Pareto Robustly Optimal (PRO) solution** for Problem (3) if

(a) it is robustly optimal, i.e., $x \in X^{\mathrm{RO}}$, and

(b) there is no $\bar{x} \in \mathcal{X}$ such that

$$u^\mathsf{T}\bar{x} \geq u^\mathsf{T}x, \quad \forall u \in \mathcal{U}, \quad \text{and}$$
$$\bar{u}^\mathsf{T}\bar{x} > \bar{u}^\mathsf{T}x, \quad \text{for some } \bar{u} \in \mathcal{U}.$$

# Pareto Robustly Optimal solutions (I. & Trichakis 2014)

$$\max_{x \in \mathcal{X}} \quad \min_{u \in \mathcal{U}} u^{\mathsf{T}} x \tag{3}$$

**Definition**

A solution $x$ is called a **Pareto Robustly Optimal (PRO) solution** for Problem (3) if

(a) it is robustly optimal, i.e., $x \in X^{\mathrm{RO}}$, and

(b) there is no $\bar{x} \in \mathcal{X}$ such that

$$u^{\mathsf{T}}\bar{x} \geq u^{\mathsf{T}}x, \quad \forall u \in \mathcal{U}, \quad \text{and}$$
$$\bar{u}^{\mathsf{T}}\bar{x} > \bar{u}^{\mathsf{T}}x, \quad \text{for some } \bar{u} \in \mathcal{U}.$$

- $X^{\mathrm{PRO}} \subseteq X^{\mathrm{RO}}$: set of all PRO solutions

# Some questions

- Given a RO solution, is it also PRO?
- How can one find a PRO solution?
- Can we optimize over $X^{\mathrm{PRO}}$?
- Can we characterize $X^{\mathrm{PRO}}$?
  - Is it non-empty?
  - Is it convex?
  - When is $X^{\mathrm{PRO}} = X^{\mathrm{RO}}$?
- How does the notion generalize in other RO formulations?

# Finding PRO solutions

**Theorem**

*Given a solution* $x \in X^{\mathrm{RO}}$ *and an arbitrary point* $\bar{p} \in \mathrm{ri}(\mathcal{U})$, *consider the following linear optimization problem:*

$$\begin{array}{ll} \text{maximize} & \bar{p}^{\mathsf{T}} y \\ \text{subject to} & y \in \mathcal{U}^* \\ & x + y \in \mathcal{X}. \end{array}$$

*Then, either*

- $\mathcal{U}^* := \{y \in \mathbb{R}^n : y^{\mathsf{T}} u \geq 0, \forall\, u \in \mathcal{U}\}$ is the dual of $\mathcal{U}$

# Finding PRO solutions

**Theorem**

*Given a solution $x \in X^{\mathrm{RO}}$ and an arbitrary point $\bar{p} \in \mathrm{ri}(\mathcal{U})$, consider the following linear optimization problem:*

$$\begin{array}{ll} \text{maximize} & \bar{p}^{\mathsf{T}} y \\ \text{subject to} & y \in \mathcal{U}^* \\ & x + y \in \mathcal{X}. \end{array}$$

*Then, either*

- *the optimal value is zero and $x \in X^{\mathrm{PRO}}$, or*

- $\mathcal{U}^* := \{y \in \mathbb{R}^n : y^{\mathsf{T}} u \geq 0, \forall\, u \in \mathcal{U}\}$ is the dual of $\mathcal{U}$

# Finding PRO solutions

**Theorem**

*Given a solution $x \in X^{\mathrm{RO}}$ and an arbitrary point $\bar{p} \in \mathrm{ri}(\mathcal{U})$, consider the following linear optimization problem:*

$$\begin{array}{ll} \text{maximize} & \bar{p}^{\mathsf{T}} y \\ \text{subject to} & y \in \mathcal{U}^* \\ & x + y \in \mathcal{X}. \end{array}$$

*Then, either*

- *the optimal value is zero and $x \in X^{\mathrm{PRO}}$, or*
- *the optimal value is strictly positive and $\bar{x} = x + y^\star \in X^{\mathrm{PRO}}$, for any optimal $y^\star$.*

- $\mathcal{U}^* := \{y \in \mathbb{R}^n : y^{\mathsf{T}} u \geq 0, \forall\, u \in \mathcal{U}\}$ is the dual of $\mathcal{U}$

# Remarks

- Finding a point $\bar{u} \in \text{ri}(\mathcal{U})$ can be done efficiently using LP techniques
- Testing whether $x \in X^{\text{RO}}$ is no harder than solving the classical RO problem in this setting
- Finding a PRO solution $x \in X^{\text{PRO}}$ is no harder than solving the classical RO problem in this setting

# Optimizing Over / Understanding the Set $X^{\mathrm{PRO}}$

- Secondary objective $r$: can we solve

$$\begin{array}{ll} \text{maximize} & r^{\mathsf{T}} x \\ \text{subject to} & x \in X^{\mathrm{PRO}} ? \end{array}$$

- Interesting case: $X^{\mathrm{RO}} \neq X^{\mathrm{PRO}}$

# Optimizing Over / Understanding the Set $X^{\mathrm{PRO}}$

- Secondary objective $r$: can we solve

$$\begin{array}{ll} \text{maximize} & r^\mathsf{T} x \\ \text{subject to} & x \in X^{\mathrm{PRO}}? \end{array}$$

- Interesting case: $X^{\mathrm{RO}} \neq X^{\mathrm{PRO}}$

**Proposition**
$X^{\mathrm{PRO}}$ *is not necessarily convex.*

**Proposition**
*If* $X^{\mathrm{RO}} \neq X^{\mathrm{PRO}}$, *then* $X^{\mathrm{PRO}} \cap \mathrm{ri}(X^{\mathrm{RO}}) = \emptyset$.

- Whether solution to nominal RO is PRO depends on algorithm used for solving LP
- Simplex **better for RO problems** than interior point methods

# What Are The Gains?

## Example (Portfolio)

- $n + 1$ assets, with returns $r_i$
- $r_i = \mu_i + \sigma_i \zeta_i$, $i = 1, \ldots, n$, $r_{n+1} = \mu_{n+1}$
- $\zeta$ unknown, $U = \{\zeta \in \mathbb{R}^n : -\mathbf{1} \leq \zeta \leq \mathbf{1}, \mathbf{1}^\mathsf{T}\zeta = 0\}$
- Objective: select weights $x$ to maximize worst-case portfolio return

# What Are The Gains?

## Example (Portfolio)

- $n+1$ assets, with returns $r_i$
- $r_i = \mu_i + \sigma_i \zeta_i$, $i = 1, \dots, n$, $r_{n+1} = \mu_{n+1}$
- $\zeta$ unknown, $U = \{\zeta \in \mathbb{R}^n : -\mathbf{1} \le \zeta \le \mathbf{1}, \mathbf{1}^\mathsf{T}\zeta = 0\}$
- Objective: select weights $x$ to maximize worst-case portfolio return

# What Are The Gains?

## Example (Portfolio)

- $n+1$ assets, with returns $r_i$
- $r_i = \mu_i + \sigma_i \zeta_i$, $i = 1, \ldots, n$, $r_{n+1} = \mu_{n+1}$
- $\zeta$ unknown, $U = \{\zeta \in \mathbb{R}^n : -\mathbf{1} \leq \zeta \leq \mathbf{1},\ \mathbf{1}^\mathsf{T}\zeta = 0\}$
- Objective: select weights $x$ to maximize worst-case portfolio return

## Example (Inventory)

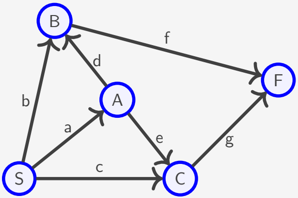
- One warehouse, $N$ retailers where uncertain demand is realized
- Transportation, holding costs and profit margins differ for each retailer
- Demand driven by market factors $d_i = d_i^0 + q_i^\mathsf{T} z, \quad i = 1, \ldots, N$
- Market factors $z$ are uncertain

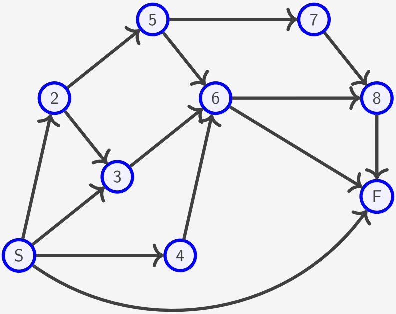
$$z \in \mathcal{U} = \{z \in \mathbb{R}^N : -b \cdot \mathbf{1} \leq z \leq b \cdot \mathbf{1},\ -B \leq \mathbf{1}^\mathsf{T} z \leq B\}$$

# Numerical experiments

## Example (Project management)

- A PERT diagram given by directed, acyclic graph $G = (\mathcal{N}, \mathcal{E})$
- $\mathcal{N}$ are project events, $\mathcal{E}$ are project activities / tasks



# Numerical experiments

Example (Project management)

- A PERT diagram given by directed, acyclic graph $G = (\mathcal{N}, \mathcal{E})$
- $\mathcal{N}$ are project events, $\mathcal{E}$ are project activities / tasks
- Task $e \in \mathcal{E}$ has uncertain duration $\tau_e = \tau_e^0 + \delta_e$

$$\delta \in \mathcal{U} := \left\{ \delta \in \mathbb{R}_+^{|\mathcal{E}|} \ : \ \delta \leq b \cdot \mathbf{1}, \ \ \mathbf{1}^\mathsf{T} \delta_e \leq B \right\}$$

- Task $e \in \mathcal{E}$ can be expedited by allocating a budgeted resource $x_e$

$$\tau_e = \tau_e^0 + \delta_e - x_e$$
$$\mathbf{1}^\mathsf{T} x \leq C$$

- Goal: find resource allocation $x$ to minimize worst-case completion time

# Results – finance and inventory examples (10K instances)

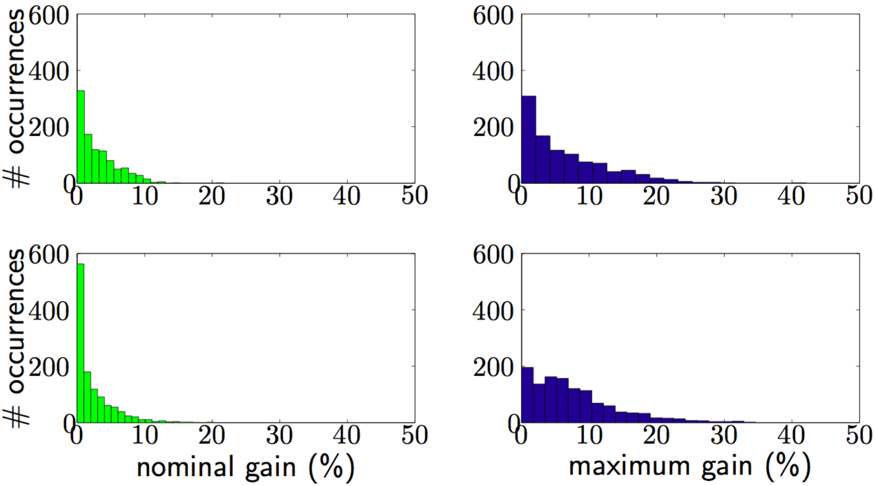
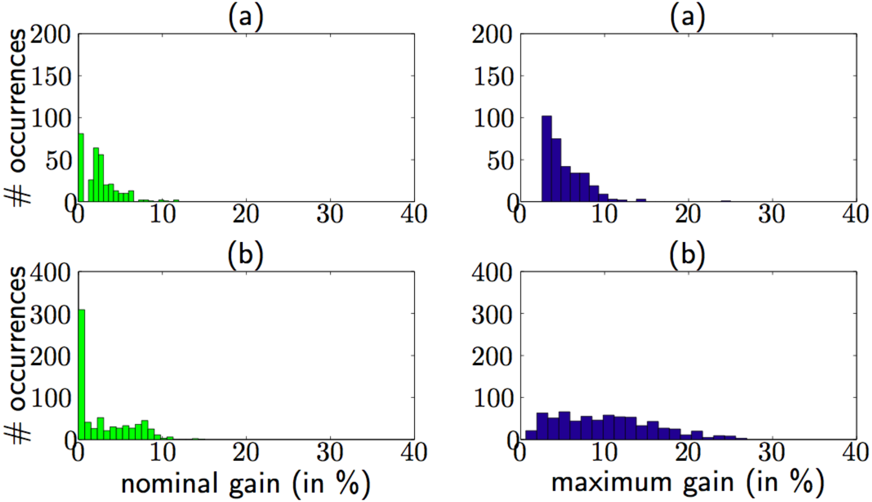
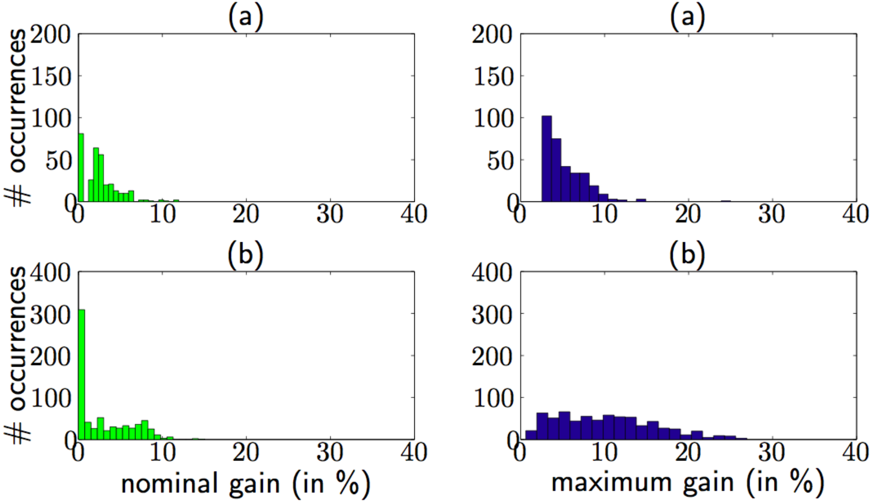



Figure: TOP: portfolio example. BOTTOM: inventory example. LEFT: performance gains in nominal scenario. RIGHT: maximal performance gains.

# Results – two project management networks



# Results – two project management networks




**Careful To Avoid Naïve Inefficiencies In Robust Models!**

# "Classical" Uncertainty Sets

The robust counterpart for $(\bar{a} + Pz)^{\mathsf{T}} x \leq b, \ \forall z \in \mathcal{U}$ is:

| U-set | $\mathcal{U}$ | Robust Counterpart | Tractability |
|---|---|---|---|
| Box | $\|z\|_\infty \leq \rho$ | $\bar{a}^{\mathsf{T}} x + \rho \|P^{\mathsf{T}} x\|_1 \leq b$ | LO |
| Ellipsoidal | $\|z\|_2 \leq \rho$ | $\bar{a}^{\mathsf{T}} x + \rho \|P^{\mathsf{T}} x\|_2 \leq b$ | CQO |
| Polyhedral | $Dz \leq d$ | $\exists y : \begin{cases} \bar{a}^{\mathsf{T}} x + d^{\mathsf{T}} y \leq b \\ D^{\mathsf{T}} y = P^{\mathsf{T}} x \\ y \geq 0 \end{cases}$ | LO |
| Budget | $\begin{cases} \|z\|_\infty \leq \rho \\ \|z\|_1 \leq \Gamma \end{cases}$ | $\exists y : \bar{a}^{\mathsf{T}} x + \rho \|y\|_1 + \Gamma \|P^{\mathsf{T}} x - y\|_\infty \leq b$ | LO |
| Convex | $h_k(z) \leq 0$ | $\exists \{w_k, u_k\}_{k \in K} : \begin{cases} a^{\mathsf{T}} x + \sum_k u_k h_k^* \left( \frac{w_k}{u_k} \right) \leq b \\ \sum_k w_k = P^{\mathsf{T}} x \\ u \geq 0 \end{cases}$ | Conv. Opt. |

**How to construct uncertainty sets?**
**How to pick parameters like $\rho, \Gamma$?**

# How to Calibrate Uncertainty Sets?

- Take a **probabilistic** view: $z_i$ are random; true distribution $\mathbb{P}$ only known to satisfy $\mathbb{P} \in \mathcal{P}$
- We seek **uncertainty sets** $\mathcal{U}$ to get **high probability of constraint satisfaction**:

$$x \text{ satisfies } (\bar{a} + Pz)^{\mathsf{T}} x \leq b, \forall z \in \mathcal{U} \quad \Rightarrow \quad \mathbb{P}\left[(\bar{a} + Pz)^{\mathsf{T}} x \leq b\right] \text{ is "large" } \forall \mathbb{P} \in \mathcal{P}$$

# How to Calibrate Uncertainty Sets?

- Take a **probabilistic** view: $z_i$ are random; true distribution $\mathbb{P}$ only known to satisfy $\mathbb{P} \in \mathcal{P}$
- We seek **uncertainty sets** $\mathcal{U}$ to get **high probability of constraint satisfaction**:

$$x \text{ satisfies } (\bar{a} + Pz)^{\mathsf{T}} x \leq b, \forall z \in \mathcal{U} \quad \Rightarrow \quad \mathbb{P}\left[(\bar{a} + Pz)^{\mathsf{T}} x \leq b\right] \text{ is "large" } \forall \mathbb{P} \in \mathcal{P}$$

# How to Calibrate Uncertainty Sets?

- Take a **probabilistic** view: $z_i$ are random; true distribution $\mathbb{P}$ only known to satisfy $\mathbb{P} \in \mathcal{P}$
- We seek **uncertainty sets** $\mathcal{U}$ to get **high probability of constraint satisfaction**:

$$x \text{ satisfies } (\bar{a} + Pz)^{\mathsf{T}} x \leq b, \forall z \in \mathcal{U} \quad \Rightarrow \quad \mathbb{P}\left[(\bar{a} + Pz)^{\mathsf{T}} x \leq b\right] \text{ is ``large'' } \forall \mathbb{P} \in \mathcal{P}$$

**Theorem**

*Suppose* $z_i$ *are* **independent r.v. with mean 0 and support on** $[-1, 1]$. *Then:*

# How to Calibrate Uncertainty Sets?

- Take a **probabilistic** view: $z_i$ are random; true distribution $\mathbb{P}$ only known to satisfy $\mathbb{P} \in \mathcal{P}$
- We seek **uncertainty sets** $\mathcal{U}$ to get **high probability of constraint satisfaction**:

$$x \text{ satisfies } (\bar{a} + Pz)^{\mathsf{T}} x \le b, \forall z \in \mathcal{U} \quad \Rightarrow \quad \mathbb{P}\left[(\bar{a} + Pz)^{\mathsf{T}} x \le b\right] \text{ is ``large'' } \forall \mathbb{P} \in \mathcal{P}$$

### Theorem

*Suppose $z_i$ are* **independent r.v. with mean 0 and support on** $[-1, 1]$*. Then:*

- $\mathcal{U}_{\text{box}} := \{z : \|z\|_\infty \le 1\} \;\Rightarrow\; \mathbb{P}\left[(\bar{a} + Pz)^{\mathsf{T}} x \le b\right] = 1.$ *(no need for independence or 0-mean)*

# How to Calibrate Uncertainty Sets?

- Take a **probabilistic** view: $z_i$ are random; true distribution $\mathbb{P}$ only known to satisfy $\mathbb{P} \in \mathcal{P}$
- We seek **uncertainty sets** $\mathcal{U}$ to get **high probability of constraint satisfaction**:

$$x \text{ satisfies } (\bar{a} + Pz)^{\mathsf{T}} x \leq b, \forall z \in \mathcal{U} \quad \Rightarrow \quad \mathbb{P}\big[(\bar{a} + Pz)^{\mathsf{T}} x \leq b\big] \text{ is "large" } \forall \mathbb{P} \in \mathcal{P}$$

**Theorem**

*Suppose $z_i$ are* **independent r.v. with mean 0 and support on** $[-1, 1]$*. Then:*

- $\mathcal{U}_{\text{box}} := \{z : \|z\|_\infty \leq 1\} \;\Rightarrow\; \mathbb{P}\big[(\bar{a} + Pz)^{\mathsf{T}} x \leq b\big] = 1$. *(no need for independence or 0-mean)*
- $\mathcal{U}_{\text{ellipsoid}} := \{z : \|z\|_2 \leq \sqrt{2 \ln(1/\epsilon)}\} \;\Rightarrow\; \mathbb{P}\big[(\bar{a} + Pz)^{\mathsf{T}} x \leq b\big] \geq 1 - \epsilon$.

# How to Calibrate Uncertainty Sets?

- Take a **probabilistic** view: $z_i$ are random; true distribution $\mathbb{P}$ only known to satisfy $\mathbb{P} \in \mathcal{P}$
- We seek **uncertainty sets** $\mathcal{U}$ to get **high probability of constraint satisfaction**:

$$x \text{ satisfies } (\bar{a} + Pz)^{\mathsf{T}} x \leq b, \forall z \in \mathcal{U} \quad \Rightarrow \quad \mathbb{P}\left[(\bar{a} + Pz)^{\mathsf{T}} x \leq b\right] \text{ is "large" } \forall \mathbb{P} \in \mathcal{P}$$

**Theorem**

*Suppose $z_i$ are* **independent r.v. with mean 0 and support on** $[-1, 1]$. *Then:*

- $\mathcal{U}_{\text{box}} := \{z \,:\, \|z\|_\infty \leq 1\} \;\Rightarrow\; \mathbb{P}\left[(\bar{a} + Pz)^{\mathsf{T}} x \leq b\right] = 1$. *(no need for independence or 0-mean)*
- $\mathcal{U}_{\text{ellipsoid}} := \{z \,:\, \|z\|_2 \leq \sqrt{2\ln(1/\epsilon)}\} \;\Rightarrow\; \mathbb{P}\left[(\bar{a} + Pz)^{\mathsf{T}} x \leq b\right] \geq 1 - \epsilon$.
- $\mathcal{U}_{\text{ellipsoid-box}} := \{z \,:\, \|z\|_2 \leq \sqrt{2\ln(1/\epsilon)}, \|z\|_\infty \leq 1\} \;\Rightarrow\; \mathbb{P}\left[(\bar{a} + Pz)^{\mathsf{T}} x \leq b\right] \geq 1 - \epsilon$.

# How to Calibrate Uncertainty Sets?

- Take a **probabilistic** view: $z_i$ are random; true distribution $\mathbb{P}$ only known to satisfy $\mathbb{P} \in \mathcal{P}$
- We seek **uncertainty sets** $\mathcal{U}$ to get **high probability of constraint satisfaction**:

$$x \text{ satisfies } (\bar{a} + Pz)^\mathsf{T} x \le b, \forall z \in \mathcal{U} \quad \Rightarrow \quad \mathbb{P}\big[(\bar{a} + Pz)^\mathsf{T} x \le b\big] \text{ is ``large'' } \forall \mathbb{P} \in \mathcal{P}$$

**Theorem**

*Suppose $z_i$ are* **independent r.v. with mean 0 and support on** $[-1, 1]$. *Then:*

- $\mathcal{U}_{\text{box}} := \{z : \|z\|_\infty \le 1\} \;\Rightarrow\; \mathbb{P}\big[(\bar{a} + Pz)^\mathsf{T} x \le b\big] = 1$. *(no need for independence or 0-mean)*
- $\mathcal{U}_{\text{ellipsoid}} := \{z : \|z\|_2 \le \sqrt{2\ln(1/\epsilon)}\} \;\Rightarrow\; \mathbb{P}\big[(\bar{a} + Pz)^\mathsf{T} x \le b\big] \ge 1 - \epsilon$.
- $\mathcal{U}_{\text{ellipsoid-box}} := \{z : \|z\|_2 \le \sqrt{2\ln(1/\epsilon)}, \|z\|_\infty \le 1\} \;\Rightarrow\; \mathbb{P}\big[(\bar{a} + Pz)^\mathsf{T} x \le b\big] \ge 1 - \epsilon$.
- $\mathcal{U}_{\text{budget}} = \{z \in \mathbb{R}^L : \|z\|_\infty \le 1, \|z\|_1 \le \Gamma = \sqrt{2\ln(1/\epsilon)}\sqrt{L}\} \;\Rightarrow\; \mathbb{P}\big[(\bar{a} + Pz)^\mathsf{T} x \le b\big] \ge 1 - \epsilon$.

- Some probabilistic information allows controlling conservatism: **useful in applications!**
- The budget $\Gamma$ depends on the dimension of $z$ ($L$), whereas $\rho$ does not!
- Proofs based on concentration inequalities

# Example: Portfolio Problem (Ben-Tal and Nemirovski)

- 200 risky assets; asset # 200 is cash, with yearly return $r_{200} = 5\%$ and zero risk
- Yearly returns $r_i$ are **independent r.v.** with values in $[\mu_i - \sigma_i, \mu_i + \sigma_i]$ and means $\mu_i$:

$$\mu_i = 1.05 + 0.3\frac{(200-i)}{199}, \quad \sigma_i = 0.05 + 0.6\frac{(200-i)}{199}, \quad i = 1, \ldots, 199.$$

- Goal: distribute \$1 to maximize worst-case value-at-risk at level $\epsilon = 0.5\%$:

$$\max_{x,t}\left\{ t \ : \ \mathbb{P}\left[\sum_{i=1}^{199} r_i x_i + r_{200}x_{200} \geq t\right] \geq 1-\epsilon, \ \forall\, \mathbb{P}, \ \sum_{i=1}^{200} x_i = 1, \ x \geq 0 \right\},$$

# Example: Portfolio Problem (Ben-Tal and Nemirovski)

- 200 risky assets; asset # 200 is cash, with yearly return $r_{200} = 5\%$ and zero risk
- Yearly returns $r_i$ are **independent r.v.** with values in $[\mu_i - \sigma_i, \mu_i + \sigma_i]$ and means $\mu_i$:

$$\mu_i = 1.05 + 0.3\frac{(200-i)}{199}, \quad \sigma_i = 0.05 + 0.6\frac{(200-i)}{199}, \quad i = 1, \dots, 199.$$

- Goal: distribute \$1 to maximize worst-case value-at-risk at level $\epsilon = 0.5\%$:

$$\max_{x,t}\left\{t \ : \ \mathbb{P}\left[\sum_{i=1}^{199} r_i x_i + r_{200}x_{200} \geq t\right] \geq 1-\epsilon, \ \forall \mathbb{P}, \ \sum_{i=1}^{200} x_i = 1, \ x \geq 0\right\},$$

- With $z_i := (r_i - \mu_i)/\sigma_i$, let's consider 3 uncertainty sets:
  1. $\mathcal{U}_{\text{box}} = \{z : \|z\|_\infty \leq 1\}$
  2. $\mathcal{U}_{\text{ellipsoid-box}} = \{z : \|z\|_\infty \leq 1, \|z\|_2 \leq \rho\}$, with $\rho = \sqrt{2\ln(1/\epsilon)} = 3.255$
  3. $\mathcal{U}_{\text{budget}} = \{z : \|z\|_\infty \leq 1, \|z\|_1 \leq \Gamma\}$ with $\Gamma = \sqrt{2\ln(1/\epsilon)}\sqrt{199} = 45.921$.

# Example: Portfolio Problem (Ben-Tal and Nemirovski)

- 200 risky assets; asset # 200 is cash, with yearly return $r_{200} = 5\%$ and zero risk
- Yearly returns $r_i$ are **independent r.v.** with values in $[\mu_i - \sigma_i, \mu_i + \sigma_i]$ and means $\mu_i$:

$$\mu_i = 1.05 + 0.3\frac{(200-i)}{199}, \quad \sigma_i = 0.05 + 0.6\frac{(200-i)}{199}, \quad i = 1, \ldots, 199.$$

- Goal: distribute \$1 to maximize worst-case value-at-risk at level $\epsilon = 0.5\%$:

$$\max_{x,t}\left\{ t \ : \ \mathbb{P}\left[\sum_{i=1}^{199} r_i x_i + r_{200}x_{200} \geq t\right] \geq 1-\epsilon, \ \forall \mathbb{P}, \ \sum_{i=1}^{200} x_i = 1, \ x \geq 0 \right\},$$

- With $z_i := (r_i - \mu_i)/\sigma_i$, let's consider 3 uncertainty sets:
  1. $\mathcal{U}_{\text{box}} = \{z : \|z\|_\infty \leq 1\}$
  2. $\mathcal{U}_{\text{ellipsoid-box}} = \{z : \|z\|_\infty \leq 1, \|z\|_2 \leq \rho\}$, with $\rho = \sqrt{2\ln(1/\epsilon)} = 3.255$
  3. $\mathcal{U}_{\text{budget}} = \{z : \|z\|_\infty \leq 1, \|z\|_1 \leq \Gamma\}$ with $\Gamma = \sqrt{2\ln(1/\epsilon)}\sqrt{199} = 45.921.$
- **Results:**
  - $\mathcal{U}_{\text{box}}$: worst-case returns yield less than risk-free return of 5%, so optimal to keep all money in cash; robust optimal return 1.05, risk 0
  - $\mathcal{U}_{\text{ellipsoid-box}}$: robust optimal value is 1.12, risk 0.5%
  - $\mathcal{U}_{\text{budget}}$: robust optimal value is 1.10, risk 0.5%
- **Allowing a tiny bit of risk can go a long way...**

# Using Concentration Results to Model Uncertainty Sets

- Bertsimas & Bandi: use the **implications** of the **Central Limit Theorem**

# Using Concentration Results to Model Uncertainty Sets

- Bertsimas & Bandi: use the **implications** of the **Central Limit Theorem**

- Suppose we have uncertainties $\{X_i\}_{i=1}^n$, each with mean $\mu$, standard deviation $\sigma$

$$\mathcal{U}_{\text{CLT}} := \left\{ (x_1, \ldots, x_n) \, : \, \left| \sum_{i=1}^n x_i - n\mu \right| \leq \Gamma \sigma \sqrt{n} \right\}.$$

  - If $n$ large and CLT premises hold, $\Gamma = 2$ (3) would give 95% (99%) coverage

# Using Concentration Results to Model Uncertainty Sets

- Bertsimas & Bandi: use the **implications** of the **Central Limit Theorem**

- Suppose we have uncertainties $\{X_i\}_{i=1}^n$, each with mean $\mu$, standard deviation $\sigma$

$$\mathcal{U}_{\text{CLT}} := \left\{ (x_1, \ldots, x_n) \,:\, \left| \sum_{i=1}^{n} x_i - n\mu \right| \leq \Gamma \sigma \sqrt{n} \right\}.$$

  - If $n$ large and CLT premises hold, $\Gamma = 2$ (3) would give 95% (99%) coverage

- Many extensions possible
  - Modeling correlations through a factor model:
    $\mathcal{U}_{\text{corr}} := \left\{ x \,:\, x = Pz + \epsilon,\ \left| \sum_{i=1}^{m} z_i - m\mu_y \right| \leq \Gamma \sigma_z \sqrt{m},\ \left| \sum_{i=1}^{n} \epsilon_i \right| \leq \Gamma \sigma_\epsilon \sqrt{n} \right\}$
  - Using stable laws to model heavy-tailed cases where variance is undefined:
    $\mathcal{U}_{\text{HT}} := \left\{ (x_1, \ldots, x_n) \,:\, \left| \sum_{i=1}^{n} x_i - n\mu \right| \leq \Gamma n^{1/\alpha} \right\}$
  - Constructing typical sets: if $H_f$ is the (Shannon) entropy of $f$,
    (i) $\mathbb{P}[\tilde{z} \in \mathcal{U}_{\text{typical}}] \to 1$, (ii) $\left| \frac{1}{n} \log f\left(\tilde{z} | \tilde{z} \in \mathcal{U}_{\text{typical}}\right) + H_f \right| \leq \epsilon_n$

# Using Concentration Results to Model Uncertainty Sets

- Bertsimas & Bandi: use the **implications** of the **Central Limit Theorem**
- Suppose we have uncertainties $\{X_i\}_{i=1}^n$, each with mean $\mu$, standard deviation $\sigma$

$$\mathcal{U}_{\text{CLT}} := \left\{(x_1, \ldots, x_n) \, : \, \left|\sum_{i=1}^n x_i - n\mu\right| \leq \Gamma\sigma\sqrt{n}\right\}.$$

  - If $n$ large and CLT premises hold, $\Gamma = 2$ (3) would give 95% (99%) coverage

- Many extensions possible
  - Modeling correlations through a factor model:
    $\mathcal{U}_{\text{corr}} := \left\{x \, : \, x = Pz + \epsilon, \, \left|\sum_{i=1}^m z_i - m\mu_y\right| \leq \Gamma\sigma_z\sqrt{m}, \, \left|\sum_{i=1}^n \epsilon_i\right| \leq \Gamma\sigma_\epsilon\sqrt{n}\right\}$
  - Using stable laws to model heavy-tailed cases where variance is undefined:
    $\mathcal{U}_{\text{HT}} := \left\{(x_1, \ldots, x_n) \, : \, \left|\sum_{i=1}^n x_i - n\mu\right| \leq \Gamma n^{1/\alpha}\right\}$
  - Constructing typical sets: if $H_f$ is the (Shannon) entropy of $f$,
    (i) $\mathbb{P}[\tilde{z} \in \mathcal{U}_{\text{typical}}] \to 1$, (ii) $\left|\frac{1}{n}\log f\left(\tilde{z} | \tilde{z} \in \mathcal{U}_{\text{typical}}\right) + H_f\right| \leq \epsilon_n$

- Bertsimas & Bandi used these to derive **robust equivalents** for several classical queueing theory and information theory results

# Distributionally Robust Optimization

# Distributionally Robust Modeling

- We saw that embedding some more probabilistic information can help

# Distributionally Robust Modeling

- We saw that embedding some more probabilistic information can help
- Let's change the paradigm slightly:
  - we know the probability distribution for $\tilde{z}$ belongs to an *ambiguity set*: $\mathbb{P} \in \mathcal{P}$
  - we model $\mathcal{P}$ and are interested in robust expected constraint satisfaction:

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[f(x, \tilde{z})] \leq b$$

- Now, the adversary is choosing $\mathbb{P}$, instead of $z$
  - **Advantage:** $\mathbb{E}_{\mathbb{P}}[f(x, \tilde{z})]$ as an expression of $\mathbb{P}$ is **always linear**
  - If $\mathbb{P}$ has discrete, finite support: much of our earlier machinery (e.g., convex duality) can be applied if the set $\mathcal{P}$ is "well-behaved"
  - Continuous $\mathbb{P}$: $\infty$-dimensional optimization

# Distributionally Robust Modeling

- We saw that embedding some more probabilistic information can help
- Let's change the paradigm slightly:
  - we know the probability distribution for $\tilde{z}$ belongs to an *ambiguity set*: $\mathbb{P} \in \mathcal{P}$
  - we model $\mathcal{P}$ and are interested in robust expected constraint satisfaction:

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}}[f(x, \tilde{z})] \leq b$$

- Now, the adversary is choosing $\mathbb{P}$, instead of $z$
  - **Advantage:** $\mathbb{E}_{\mathbb{P}}[f(x, \tilde{z})]$ as an expression of $\mathbb{P}$ is **always linear**
  - If $\mathbb{P}$ has discrete, finite support: much of our earlier machinery (e.g., convex duality) can be applied if the set $\mathcal{P}$ is "well-behaved"
  - Continuous $\mathbb{P}$: $\infty$-dimensional optimization
- Very old idea, dating to the 1950s (Scarf 1958, Zackova 1966)
- Kuhn, Shafiee, Wiesemann (2024): tutorial on state-of-the-art. Can model:
  - known (**bounds** on) moments, e.g., means, covariance matrix, higher order
  - known (**bounds on**) quantiles (e.g., median) or spread statistics
  - multiple confidence regions
  - distance from a nominal distribution (Kullback-Leibler, Wasserstein, etc.)

# Esfahani and Kuhn (2015)

**Baseline problem.** Single-stage stochastic program:

$$J^\star := \inf_{x \in X} \mathbb{E}_{\mathbb{P}}\left[h(x, z)\right]$$

- $x \in X \subseteq \mathbb{R}^n$ is the decision,
- $z \in \mathcal{U} \subseteq \mathbb{R}^m$ is a random vector,
- $\mathbb{P}$ (distribution of $z$) is *unknown*.

**Data.** We have i.i.d. samples $\widehat{\mathcal{U}}_N := \{z_1, \ldots, z_N\}$ and form the empirical distribution:

$$\widehat{\mathbb{P}}_N := \frac{1}{N} \sum_{i=1}^{N} \delta_{z_i}.$$

# Esfahani and Kuhn (2015)

**Baseline problem.** Single-stage stochastic program:

$$J^\star := \inf_{x \in X} \mathbb{E}_{\mathbb{P}}\big[h(x, z)\big]$$

- $x \in X \subseteq \mathbb{R}^n$ is the decision,
- $z \in \mathcal{U} \subseteq \mathbb{R}^m$ is a random vector,
- $\mathbb{P}$ (distribution of $z$) is *unknown*.

**Data.** We have i.i.d. samples $\widehat{\mathcal{U}}_N := \{z_1, \ldots, z_N\}$ and form the empirical distribution:

$$\widehat{\mathbb{P}}_N := \frac{1}{N} \sum_{i=1}^{N} \delta_{z_i}.$$

Classical solution method: **Sample Average Approximation (SAA)**

$$J_{\mathrm{SAA}} := \inf_{x \in X} \mathbb{E}_{\widehat{\mathbb{P}}_N}[h(x, z)] = \inf_{x \in X} \frac{1}{N} \sum_{i=1}^{N} h(x, z_i).$$

# Esfahani and Kuhn (2015)

**Baseline problem.** Single-stage stochastic program:

$$J^\star := \inf_{x \in X} \mathbb{E}_{\mathbb{P}}\big[h(x, z)\big]$$

- $x \in X \subseteq \mathbb{R}^n$ is the decision,
- $z \in \mathcal{U} \subseteq \mathbb{R}^m$ is a random vector,
- $\mathbb{P}$ (distribution of $z$) is *unknown*.

**Data.** We have i.i.d. samples $\widehat{\mathcal{U}}_N := \{z_1, \ldots, z_N\}$ and form the empirical distribution:

$$\widehat{\mathbb{P}}_N := \frac{1}{N} \sum_{i=1}^{N} \delta_{z_i}.$$

Classical solution method: **Sample Average Approximation (SAA)**

$$J_{\mathrm{SAA}} := \inf_{x \in X} \mathbb{E}_{\widehat{\mathbb{P}}_N}[h(x, z)] = \inf_{x \in X} \frac{1}{N} \sum_{i=1}^{N} h(x, z_i).$$

SAA is asymptotically consistent, but for small $N$ it can:

- *overfit* the data ("optimizer's curse")
- give poor out-of-sample performance

# Wasserstein Metric and Ambiguity Sets

**Wasserstein distance.** Let $\mathcal{M}(\mathcal{U})$ be the set of all distributions supported on $\mathcal{U}$. For $\mathbb{Q}_1, \mathbb{Q}_2 \in \mathcal{M}(\mathcal{U})$,

$$d_{\mathbb{W}}(\mathbb{Q}_1, \mathbb{Q}_2) := \inf_{\pi \in \Pi} \int_{\mathcal{U}^2} \|\xi_1 - \xi_2\| \, d\pi(\xi_1, \xi_2)$$

- $\Pi$ is **the set of all couplings of** $\mathbb{Q}_1$ **and** $\mathbb{Q}_2$, i.e., joint distributions of $\xi_1$ and $\xi_2$ with marginals given by $\mathbb{Q}_1$ and $\mathbb{Q}_2$, respectively
- $\|\cdot\|$ is any norm. More popular choices: $\|\cdot\|_\infty, \|\cdot\|_1, \|\cdot\|_2$

# Wasserstein Metric and Ambiguity Sets

**Wasserstein distance.** Let $\mathcal{M}(\mathcal{U})$ be the set of all distributions supported on $\mathcal{U}$. For $\mathbb{Q}_1, \mathbb{Q}_2 \in \mathcal{M}(\mathcal{U})$,

$$d_{\mathbb{W}}(\mathbb{Q}_1, \mathbb{Q}_2) := \inf_{\pi \in \Pi} \int_{\mathcal{U}^2} \|\xi_1 - \xi_2\| \, d\pi(\xi_1, \xi_2)$$

- $\Pi$ is **the set of all couplings of** $\mathbb{Q}_1$ **and** $\mathbb{Q}_2$, i.e., joint distributions of $\xi_1$ and $\xi_2$ with marginals given by $\mathbb{Q}_1$ and $\mathbb{Q}_2$, respectively
- $\|\cdot\|$ is any norm. More popular choices: $\|\cdot\|_\infty, \|\cdot\|_1, \|\cdot\|_2$

**Wasserstein ambiguity set (ball).**

$$\mathbb{B}_\epsilon(\widehat{\mathbb{P}}_N) := \left\{ \mathbb{Q} \in \mathcal{M}(\mathcal{U}) \ : \ d_{\mathbb{W}}(\mathbb{Q}, \widehat{\mathbb{P}}_N) \le \epsilon \right\}.$$

- Centered at the empirical distribution $\widehat{\mathbb{P}}_N$.
- Radius $\epsilon$ controls *conservatism*.
- Contains both discrete and continuous distributions close to $\widehat{\mathbb{P}}_N$.

# Wasserstein DRO Formulation

**Distributionally robust objective.** For fixed decision $x$, worst-case expected cost is:

$$\sup_{\mathbb{P} \in \mathbb{B}_\epsilon(\widehat{\mathbb{P}}_N)} \mathbb{E}_{\mathbb{P}}\big[h(x, z)\big].$$

**Data-driven distributionally robust optimization:**

$$J_N(\epsilon) := \inf_{x \in X} \sup_{\mathbb{P} \in \mathbb{B}_\epsilon(\widehat{\mathbb{P}}_N)} \mathbb{E}_{\mathbb{P}}\big[h(x, z)\big].$$

Interpretation:

- Take all distributions $\mathbb{P}$ within distance $\epsilon$ of the data-driven $\widehat{\mathbb{P}}_N$.
- Optimize against the *most adversarial* such distribution.

Goal:

- Choose $\epsilon$ and solve $J_N(\epsilon)$ so that
  - we get *good out-of-sample performance*, and
  - we retain *finite-sample* and *asymptotic* guarantees.

# Measure Concentration and Choice of Radius

Assume a **light-tail condition** on $\mathbb{P}$:

$$\mathbb{E}_{\mathbb{P}}\big[\exp(\|z\|^a)\big] < \infty \quad \text{for some } a > 1.$$

Then a measure concentration result (Fournier–Guillin) implies: for some $c_1, c_2 > 0$,

$$\mathbb{P}^N\big[d_W(\mathbb{P}, \widehat{\mathbb{P}}_N) \geq \epsilon\big] \;\leq\; \begin{cases} c_1 \exp\big(-c_2 N \epsilon^{\max\{m,2\}}\big), & \epsilon \leq 1, \\ c_1 \exp\big(-c_2 N \epsilon^a\big), & \epsilon > 1. \end{cases}$$

For a prescribed significance level $\beta \in (0,1)$, we can choose a radius $\epsilon_N(\beta)$ such that

$$\mathbb{P}^N\big[d_W(\mathbb{P}, \widehat{\mathbb{P}}_N) \leq \epsilon_N(\beta)\big] \;\geq\; 1 - \beta.$$

**Interpretation:** with probability at least $1 - \beta$, the *true* distribution $\mathbb{P}$ lies inside the Wasserstein ball $\mathbb{B}_{\epsilon_N(\beta)}(\widehat{\mathbb{P}}_N)$.

# Finite-sample Performance Guarantee

Fix $\beta \in (0,1)$ and choose $\epsilon = \epsilon_N(\beta)$ as in the concentration bound

Let $x_N$ be an optimizer of the DRO problem

$$J_N := \inf_{x \in X} \sup_{\mathbb{P} \in \mathbb{B}_{\epsilon_N(\beta)}(\widehat{\mathbb{P}}_N)} \mathbb{E}_{\mathbb{P}}[h(x, z)].$$

# Finite-sample Performance Guarantee

Fix $\beta \in (0,1)$ and choose $\epsilon = \epsilon_N(\beta)$ as in the concentration bound

Let $x_N$ be an optimizer of the DRO problem

$$J_N := \inf_{x \in X} \sup_{\mathbb{P} \in \mathbb{B}_{\epsilon_N(\beta)}(\widehat{\mathbb{P}}_N)} \mathbb{E}_{\mathbb{P}}[h(x, z)].$$

Then, with probability at least $1 - \beta$ (over the sampling of $\widehat{\mathcal{U}}_N$),

$$\mathbb{E}_{\mathbb{P}}\left[h(x_N, z)\right] \; \leq \; J_N.$$

**So:**

- $J_N$ is an **upper confidence bound on the out-of-sample cost of** $x_N$ **valid with confidence level** $1 - \beta$
- We can also get **asymptotic consistency**: as $\beta_N \to 0$, by choosing $\epsilon_N = \epsilon_N(\beta_N)$, we get $J_N \to J^\star$ almost surely, so the finite-sample Wasserstein DRO asymptotically recovers the true stochastic program

# Convex Reformulations

Focus on Nature's Problem, i.e., the *inner* worst-case expectation for a fixed $x$:

$$(\mathrm{NP}) \quad \sup_{\mathbb{P} \in \mathbb{B}_\epsilon(\widehat{\mathbb{P}}_N)} \mathbb{E}_{\mathbb{P}}[\ell(z)]$$

**Assumptions:** the support $\mathcal{U}$ of $z$ is convex and closed and the loss function $\ell$ is:

$$\ell(z) = \max_{k \leq K} \ell_k(z),$$

where each $\ell_k$ is proper, **concave**, and upper semicontinuous

# Convex Reformulations

Focus on Nature's Problem, i.e., the *inner* worst-case expectation for a fixed $x$:

$$\text{(NP)} \quad \sup_{\mathbb{P}\in\mathbb{B}_\epsilon(\widehat{\mathbb{P}}_N)} \mathbb{E}_{\mathbb{P}}[\ell(z)]$$

**Assumptions:** the support $\mathcal{U}$ of $z$ is convex and closed and the loss function $\ell$ is:

$$\ell(z) = \max_{k\leq K} \ell_k(z),$$

where each $\ell_k$ is proper, **concave**, and upper semicontinuous

**Key result.** The optimal value of (NP) equals the optimal value of:

$$\begin{aligned} \min_{\lambda,\, s_i,\, z_{ik}} \quad & \lambda\varepsilon + \frac{1}{N}\sum_{i=1}^{N} s_i \\ \text{s.t.} \quad & [-\ell_k + \chi_{\mathcal{U}}]^*(z_{ik}) - \langle z_{ik}, \xi_i\rangle \ \leq \ s_i, \quad \forall i,k, \\ & \|z_{ik}\|_* \ \leq \ \lambda, \quad \forall i,k, \end{aligned}$$

where $\chi_{\mathcal{U}}$ is the indicator function of $\mathcal{U}$, $[f]^*$ is the Fenchel conjugate of $f$, and $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$.

- **Enough to solve a finite-dimensional convex problem**
- $\ell_k$ **linear,** $\mathcal{U}$ **polyhedral, 1- or** $\infty$**-norm in** $d_{\mathbb{W}}(\cdot,\cdot)$ $\Rightarrow$ **finite-dimensional LP**

# Using Hypothesis Tests to Model Uncertainty Sets

Bertsimas, Gupta, Kallus ('17): **data-driven** uncertainty sets from **hypothesis tests**

**Table 1** Summary of data-driven uncertainty sets proposed in this paper. SOC, EC and LMI denote second-order cone representable sets, exponential cone representable sets, and linear matrix inequalities, respectively

| Assumptions on $\mathbb{P}^*$ | Hypothesis test | Geometric description | Eqs. | Inner problem |
|---|---|---|---|---|
| Discrete support | $\chi^2$-test | SOC | (13, 15) | |
| Discrete support | G-test | Polyhedral* | (13, 16) | |
| Independent marginals | KS Test | Polyhedral* | (21) | Line search |
| Independent marginals | K Test | Polyhedral* | (76) | Line search |
| Independent marginals | CvM Test | SOC* | (76, 69) | |
| Independent marginals | W Test | SOC* | (76, 70) | |
| Independent marginals | AD Test | EC | (76, 71) | |
| Independent marginals | Chen et al. [23] | SOC | (27) | Closed-form |
| None | Marginal Samples | Box | (31) | Closed-form |
| None | Linear Convex Ordering | Polyhedron | (34) | |
| None | Shawe-Taylor and Cristianini [46] | SOC | (39) | Closed-form |
| None | Delage and Ye [25] | LMI | (41) | |

The additional "*" notation indicates a set of the above type with one additional, relative entropy constraint. $KS$, $K$, $CvM$, $W$, and $AD$ denote the Kolmogorov–Smirnov, Kuiper, Cramer-von Mises, Watson and Anderson-Darling goodness of fit tests, respectively. In some cases, we can identify a worst-case realization of $\mathbf{u}$ in (1) for bi-affine $f$ and a candidate $\mathbf{x}$ with a specialized algorithm. In these cases, the column "Inner Problem" roughly describes this algorithm

# References I

A. Ben-Tal, A. Goryashko, E. Guslitzer, and A. Nemirovski. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99(2):351–376, 2004.

F. de Ruiter, A. Ben-Tal, R. Brekelmans, and D. den Hertog. Adjustable robust optimization with decision rules based on inexact revealed data. Center discussion paper series no. 2014-003, CentER, 2014.

A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming*. MPS / SIAM Series on Optimization. SIAM, 2009.