Introduction to Robust Optimization

Dan Iancu

December 4, 2024

Outline for Today

1. Robust Optimization
   - Quick Recap
   - Calibrating Uncertainty Sets
   - Distributionally Robust Optimization
   - Connections with Other Areas
   - Some Important Caveats When Applying Robust Optimization

2. Dynamic Robust Optimization
   - Properly Writing a Robust DP
   - An Inventory Example
   - Tractable Approximations with Decision Rules
   - Some Practical Issues
   - Bellman Optimality
   - An Application in Monitoring

"Classical" Robust Optimization (RO)

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

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

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

The robust counterpart for $\boxed{(\bar{\mathbf{a}} + P\boldsymbol{z})^\intercal \boldsymbol{x} \leqslant b, \ \forall\, \boldsymbol{z} \in \mathcal{U}}$ is:

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

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

RC for **Linear Optimization** Problems with "Classical" Uncertainty Sets

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

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

- Several extensions
- Robust counterparts can be handled by large-scale modern solvers
- Already a lot of mileage in many practical problems:
  logistics and supply chain management, radiation therapy, scheduling, ...

RC for **Linear Optimization** Problems with "Classical" Uncertainty Sets

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

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

- How to pick parameters like ρ, Γ?
- How to build uncertainty sets (from data)?

How to Calibrate Uncertainty Sets?

For "structured" uncertainty sets, how should we pick parameters like ρ, Γ?

How to Calibrate Uncertainty Sets?

For "structured" uncertainty sets, how should we pick parameters like ρ, Γ?

Let's take a **probabilistic** view for a moment:

Suppose z_i *are really random, and we seek* ρ, Γ *to ensure* $\mathbb{P}\big[(\bar{a} + P\boldsymbol{z})^\intercal x \leqslant b\big]$ *is "large",* $\forall\, \mathbb{P}$

How to Calibrate Uncertainty Sets?

For "structured" uncertainty sets, how should we pick parameters like ρ, Γ?

Let's take a **probabilistic** view for a moment:
Suppose z_i *are really random, and we seek* ρ, Γ *to ensure* $\mathbb{P}\big[(\bar{a} + Pz)^\intercal x \leqslant b\big]$ *is "large",* $\forall\, \mathbb{P}$

Theorem (High Probability of Constraint Satisfaction)

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

How to Calibrate Uncertainty Sets?

For "structured" uncertainty sets, how should we pick parameters like ρ, Γ?

Let's take a **probabilistic** view for a moment:
Suppose z_i are really random, and we seek ρ, Γ to ensure $\mathbb{P}\big[(\bar{a} + Pz)^\intercal x \leqslant b\big]$ is "large", $\forall\, \mathbb{P}$

Theorem (High Probability of Constraint Satisfaction)

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

- *if x feasible for RC with $\mathcal{U}_{\text{box}} \stackrel{\text{def}}{=} \{z : \|z\|_\infty \leqslant 1\}$, then $\mathbb{P}\big[(\bar{a} + Pz)^\intercal x \leqslant b\big] = 1$.*
  (no need for independence or 0-mean)

How to Calibrate Uncertainty Sets?

For "structured" uncertainty sets, how should we pick parameters like ρ, Γ?

Let's take a **probabilistic** view for a moment:
Suppose z_i are really random, and we seek ρ, Γ to ensure $\mathbb{P}\big[(\bar{a} + Pz)^\intercal x \leqslant b\big]$ is "large", $\forall\, \mathbb{P}$

Theorem (High Probability of Constraint Satisfaction)

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

- *if x feasible for RC with* $\mathcal{U}_{\text{box}} \stackrel{\text{def}}{=} \{z : \|z\|_\infty \leqslant 1\}$, *then* $\mathbb{P}\big[(\bar{a} + Pz)^\intercal x \leqslant b\big] = 1$.
  (no need for independence or 0-mean)
- *if x feasible for RC with* $\mathcal{U}_{\text{ellipsoid}} \stackrel{\text{def}}{=} \{z : \|z\|_2 \leqslant \sqrt{2\ln(1/\epsilon)}\}$, *then*

$$\mathbb{P}\big[(\bar{a} + Pz)^\intercal x \leqslant b\big] \geqslant 1 - \epsilon.$$

How to Calibrate Uncertainty Sets?

For "structured" uncertainty sets, how should we pick parameters like ρ, Γ?

Let's take a **probabilistic** view for a moment:

Suppose z_i are really random, and we seek ρ, Γ to ensure $\mathbb{P}\big[(\bar{a} + Pz)^\intercal x \leqslant b\big]$ is "large", $\forall\, \mathbb{P}$

Theorem (High Probability of Constraint Satisfaction)

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

- *if x feasible for RC with* $\mathcal{U}_{\text{box}} \stackrel{\text{def}}{=} \{z : \|z\|_\infty \leqslant 1\}$, *then* $\mathbb{P}\big[(\bar{a} + Pz)^\intercal x \leqslant b\big] = 1$.
  (no need for independence or 0-mean)
- *if x feasible for RC with* $\mathcal{U}_{\text{ellipsoid}} \stackrel{\text{def}}{=} \{z : \|z\|_2 \leqslant \sqrt{2\ln(1/\epsilon)}\}$, *then*
  $$\mathbb{P}\big[(\bar{a} + Pz)^\intercal x \leqslant b\big] \geqslant 1 - \epsilon.$$
- *if x feasible for RC with* $\mathcal{U}_{\text{ellipsoid-box}} \stackrel{\text{def}}{=} \{z : \|z\|_2 \leqslant \sqrt{2\ln(1/\epsilon)}, \|z\|_\infty \leqslant 1\}$, *then*
  $$\mathbb{P}\big[(\bar{a} + Pz)^\intercal x \leqslant b\big] \geqslant 1 - \epsilon.$$

How to Calibrate Uncertainty Sets?

For "structured" uncertainty sets, how should we pick parameters like ρ, Γ?

Let's take a **probabilistic** view for a moment:

Suppose z_i are really random, and we seek ρ, Γ to ensure $\mathbb{P}\big[(\bar{a} + Pz)^\intercal x \leqslant b\big]$ is "large", $\forall\, \mathbb{P}$

Theorem (High Probability of Constraint Satisfaction)

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

- *if x feasible for RC with* $\mathcal{U}_{\text{box}} \overset{\text{def}}{=} \{z : \|z\|_\infty \leqslant 1\}$, *then* $\mathbb{P}\big[(\bar{a} + Pz)^\intercal x \leqslant b\big] = 1$.
  (no need for independence or 0-mean)
- *if x feasible for RC with* $\mathcal{U}_{\text{ellipsoid}} \overset{\text{def}}{=} \{z : \|z\|_2 \leqslant \sqrt{2\ln(1/\epsilon)}\}$, *then*
  $$\mathbb{P}\big[(\bar{a} + Pz)^\intercal x \leqslant b\big] \geqslant 1 - \epsilon.$$
- *if x feasible for RC with* $\mathcal{U}_{\text{ellipsoid-box}} \overset{\text{def}}{=} \{z : \|z\|_2 \leqslant \sqrt{2\ln(1/\epsilon)}, \|z\|_\infty \leqslant 1\}$, *then*
  $$\mathbb{P}\big[(\bar{a} + Pz)^\intercal x \leqslant b\big] \geqslant 1 - \epsilon.$$
- *if x feasible for RC with* $\mathcal{U}_{\text{budget}} = \{z \in \mathbb{R}^L : \|z\|_\infty \leqslant 1,\ \|z\|_1 \leqslant \Gamma = \sqrt{2\ln(1/\epsilon)}\sqrt{L}\}$,
  $$\mathbb{P}\big[(\bar{a} + Pz)^\intercal x \leqslant b\big] \geqslant 1 - \epsilon.$$

How to Calibrate Uncertainty Sets?

For "structured" uncertainty sets, how should we pick parameters like ρ, Γ?

Let's take a **probabilistic** view for a moment:

Suppose z_i are really random, and we seek ρ, Γ to ensure $\mathbb{P}\big[(\bar{a} + Pz)^\intercal x \leqslant b\big]$ is "large", $\forall\, \mathbb{P}$

Theorem (High Probability of Constraint Satisfaction)

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

- *if x feasible for RC with $\mathcal{U}_{\text{box}} \stackrel{\text{def}}{=} \{z : \|z\|_\infty \leqslant 1\}$, then $\mathbb{P}\big[(\bar{a} + Pz)^\intercal x \leqslant b\big] = 1$.*
  (no need for independence or 0-mean)
- *if x feasible for RC with $\mathcal{U}_{\text{ellipsoid}} \stackrel{\text{def}}{=} \{z : \|z\|_2 \leqslant \sqrt{2\ln(1/\epsilon)}\}$, then*
  $$\mathbb{P}\big[(\bar{a} + Pz)^\intercal x \leqslant b\big] \geqslant 1 - \epsilon.$$
- *if x feasible for RC with $\mathcal{U}_{\text{ellipsoid-box}} \stackrel{\text{def}}{=} \{z : \|z\|_2 \leqslant \sqrt{2\ln(1/\epsilon)}, \|z\|_\infty \leqslant 1\}$, then*
  $$\mathbb{P}\big[(\bar{a} + Pz)^\intercal x \leqslant b\big] \geqslant 1 - \epsilon.$$
- *if x feasible for RC with $\mathcal{U}_{\text{budget}} = \{z \in \mathbb{R}^L : \|z\|_\infty \leqslant 1,\ \|z\|_1 \leqslant \Gamma = \sqrt{2\ln(1/\epsilon)}\sqrt{L}\}$,*
  $$\mathbb{P}\big[(\bar{a} + Pz)^\intercal x \leqslant b\big] \geqslant 1 - \epsilon.$$

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

Another Quick Example: A 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 μ_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 so as to maximize worst-case value-at-risk at level $\epsilon = 0.5\%$:

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

Another Quick Example: A 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 μ_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 so as to maximize worst-case value-at-risk at level $\epsilon = 0.5\%$:

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

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

Another Quick Example: A 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 μ_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 so as to maximize worst-case value-at-risk at level $\epsilon = 0.5\%$:

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

- With $z_i \overset{\text{def}}{=} (r_i - \mu_i)/\sigma_i$, let's consider 3 uncertainty sets:
  1. $\mathcal{U}_{\text{box}} = \{z : \|z\|_\infty \leqslant 1\}$
  2. $\mathcal{U}_{\text{ellipsoid-box}} = \{z : \|z\|_\infty \leqslant 1, \|z\|_2 \leqslant \rho\}$, with $\rho = \sqrt{2\ln(1/\epsilon)} = 3.255$
  3. $\mathcal{U}_{\text{budget}} = \{z : \|z\|_\infty \leqslant 1, \|z\|_1 \leqslant \Gamma\}$ with $\Gamma = \sqrt{2\ln(1/\epsilon)}\sqrt{199} = 45.921$.
- Results:
  - $\mathcal{U}_{\text{box}}$: worst-case returns $r_i = \mu_i - \sigma_i$ 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%
- $\mathcal{U}_{\text{box}}$ can be quite conservative, a tiny bit of risk can go a long way...

Using Concentration Results to Model Uncertainty Sets

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

  "All epistemological value of the theory of probability is based on this: that large-scale random phenomena in their collective action create strict, non-random regularity" (Gnedenko and Kolmogorov, 1954)

Using Concentration Results to Model Uncertainty Sets

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

  "All epistemological value of the theory of probability is based on this: that large-scale random phenomena in their collective action create strict, non-random regularity" (Gnedenko and Kolmogorov, 1954)

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

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

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

Using Concentration Results to Model Uncertainty Sets

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

  "All epistemological value of the theory of probability is based on this: that large-scale random phenomena in their collective action create strict, non-random regularity" (Gnedenko and Kolmogorov, 1954)

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

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

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

- Many extensions possible
  - Modeling correlations through a factor model:
    $\mathcal{U}_{\text{corr}} \stackrel{\text{def}}{=} \left\{ \boldsymbol{x} : \boldsymbol{x} = P\boldsymbol{z} + \boldsymbol{\epsilon}, \left| \sum_{i=1}^{m} z_i - m\mu_y \right| \leqslant \Gamma \sigma_z \sqrt{m}, \left| \sum_{i=1}^{n} \epsilon_i \right| \leqslant \Gamma \sigma_\epsilon \sqrt{n}, \right\}.$
  - Using stable laws to model heavy-tailed cases where variance is undefined:
    $\mathcal{U}_{\text{HT}} \stackrel{\text{def}}{=} \left\{ (x_1, \dots, x_n) : \left| \sum_{i=1}^{n} x_i - n\mu \right| \leqslant \Gamma n^{1/\alpha} \right\}.$
  - 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) $\left| \frac{1}{n} \log f\left(\tilde{z} | \tilde{z} \in \mathcal{U}_{\text{typical}}\right) + H_f \right| \leqslant \varepsilon_n$

Using Concentration Results to Model Uncertainty Sets

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

  "All epistemological value of the theory of probability is based on this: that large-scale random phenomena in their collective action create strict, non-random regularity" (Gnedenko and Kolmogorov, 1954)

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

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

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

- Many extensions possible
  - Modeling correlations through a factor model:
    $\mathcal{U}_{\text{corr}} \stackrel{\text{def}}{=} \left\{\boldsymbol{x} : \boldsymbol{x} = P\boldsymbol{z} + \boldsymbol{\epsilon}, \left|\sum_{i=1}^{m} z_i - m\mu_y\right| \leqslant \Gamma\sigma_z\sqrt{m}, \left|\sum_{i=1}^{n} \epsilon_i\right| \leqslant \Gamma\sigma_\epsilon\sqrt{n}, \right\}.$
  - Using stable laws to model heavy-tailed cases where variance is undefined:
    $\mathcal{U}_{\text{HT}} \stackrel{\text{def}}{=} \left\{(x_1, \ldots, x_n) : \left|\sum_{i=1}^{n} x_i - n\mu\right| \leqslant \Gamma n^{1/\alpha}\right\}.$
  - 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) $\left|\frac{1}{n}\log f\left(\tilde{z} | \tilde{z} \in \mathcal{U}_{\text{typical}}\right) + H_f\right| \leqslant \varepsilon_n$

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

Using Hypothesis Tests to Model Uncertainty Sets

Another powerful idea: derive **data-driven** uncertainty sets from **hypothesis tests**

From Bertsimas, Gupta, Kallus (2017):

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	χ^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

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:

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

- 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**, so much of our earlier machinery (e.g., convex duality) can be applied if the set $\mathcal{P}$ is "well-behaved"
  - **Disadvantage:** Maximizing over continuous $\mathbb{P}$: -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:

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

- 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**, so much of our earlier machinery (e.g., convex duality) can be applied if the set $\mathcal{P}$ is "well-behaved"
  - **Disadvantage:** Maximizing over continuous $\mathbb{P}$: -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.)

Two Important Caveats When Working With Robust Models

An Example: A 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.

An Example: A 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:

An Example: A 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:

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

Parameters:
$\mathcal{T}$: discrete planning horizon, indexed by τ
$\mathcal{F}$: 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^s_{ij}: cost of shipping units from location i to j
$D_{j\tau}$: demand in period τ at location j.

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

An Example: A 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:

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

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

Baron et al. 2011 captured uncertain demands:

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

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

An Example: A 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:

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

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

Baron et al. 2011 captured uncertain demands:

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

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

An Example: A 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:

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

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

Baron et al. 2011 captured uncertain demands:

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

Step 3. Derive robust counterpart for the problem. Here, this will be a Conic Quadratic program.

Compare Two Models

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

$$
\begin{aligned}
\max_{X,I,Z,P} \quad & \sum_{\tau\in\mathcal{T}}\sum_{i\in\mathcal{F}}\sum_{j\in\mathcal{N}}(p-c^{s}_{ij})X_{ij\tau} - \sum_{\tau\in\mathcal{T}}\sum_{i\in\mathcal{F}} c_i P_{i\tau} - \sum_{i\in\mathcal{F}}(C_i Z_i - K_i I_i) \\
\text{s.t.} \quad & \sum_{i\in\mathcal{F}} X_{ij\tau} \leqslant D_{j\tau}, \quad j\in\mathcal{N},\ \tau\in\mathcal{T}, \\
& \sum_{j\in\mathcal{N}} X_{ij\tau} \leqslant P_{i\tau}, \quad i\in\mathcal{F},\ \tau\in\mathcal{T}, \\
& X_{ij\tau} \geqslant 0, \quad i\in\mathcal{F},\ j\in\mathcal{N},\ \tau\in\mathcal{T} \\
& P_{i\tau} \leqslant Z_i,\ \ Z_i \leqslant M\cdot I_i, \quad i\in\mathcal{F},\ \tau\in\mathcal{T} \\
& I \in \{0,1\}^{|\mathcal{F}|}, \qquad \text{(where } M \text{ is a large enough constant.)}
\end{aligned}
$$

Compare Two Models

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

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

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

$$\begin{aligned}
\max_{Y,I,Z,P} \quad & \sum_{\tau\in\mathcal{T}}\sum_{i\in\mathcal{F}}\sum_{j\in\mathcal{N}}(p-c^s_{ij})Y_{ij\tau}D_{j\tau} - \sum_{\tau\in\mathcal{T}}\sum_{i\in\mathcal{F}} c_i P_{i\tau} - \sum_{i\in\mathcal{F}}(C_i Z_i - K_i I_i) \\
\text{s.t.} \quad & \sum_{i\in\mathcal{F}} Y_{ij\tau} \leqslant 1, \quad j\in\mathcal{N},\ \tau\in\mathcal{T}, \\
& \sum_{j\in\mathcal{N}} Y_{ij\tau}D_{j\tau} \leqslant P_{i\tau}, \quad i\in\mathcal{F},\ \tau\in\mathcal{T}, \\
& Y_{ij\tau} \geqslant 0, \quad i\in\mathcal{F},\ j\in\mathcal{N},\ \tau\in\mathcal{T} \\
& P_{i\tau} \leqslant Z_i,\ Z_i \leqslant M\cdot I_i, \quad i\in\mathcal{F},\ \tau\in\mathcal{T} \\
& I\in\{0,1\}^{|\mathcal{F}|}, \quad \text{(where } M \text{ is a large enough constant.)}
\end{aligned}$$

Compare Two Models

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

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

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

$$\begin{aligned}
\max_{Y,I,Z,P} \quad & \sum_{\tau\in\mathcal{T}}\sum_{i\in\mathcal{F}}\sum_{j\in\mathcal{N}}(p-c^s_{ij})Y_{ij\tau}D_{j\tau} - \sum_{\tau\in\mathcal{T}}\sum_{i\in\mathcal{F}} c_i P_{i\tau} - \sum_{i\in\mathcal{F}}(C_i Z_i - K_i I_i) \\
\text{s.t.} \quad & \sum_{i\in\mathcal{F}} Y_{ij\tau} \leqslant 1, \quad j\in\mathcal{N},\ \tau\in\mathcal{T}, \\
& \sum_{j\in\mathcal{N}} Y_{ij\tau}D_{j\tau} \leqslant P_{i\tau}, \quad i\in\mathcal{F},\ \tau\in\mathcal{T}, \\
& Y_{ij\tau} \geqslant 0, \quad i\in\mathcal{F},\ j\in\mathcal{N},\ \tau\in\mathcal{T} \\
& P_{i\tau} \leqslant Z_i,\ Z_i \leqslant M\cdot I_i, \quad i\in\mathcal{F},\ \tau\in\mathcal{T} \\
& I\in\{0,1\}^{|\mathcal{F}|}, \qquad \text{(where } M \text{ is a large enough constant.)}
\end{aligned}$$

For fixed D, are these **deterministic/nominal** models **equivalent**?
Are their **robust counterparts** **equivalent**?

Compare Two Models

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

$$\begin{aligned}
\max_{X,I,Z,P} \quad & \sum_{\tau\in\mathcal{T}}\sum_{i\in\mathcal{F}}\sum_{j\in\mathcal{N}}(p-c^s_{ij})X_{ij\tau}-\sum_{\tau\in\mathcal{T}}\sum_{i\in\mathcal{F}}c_iP_{i\tau}-\sum_{i\in\mathcal{F}}(C_iZ_i-K_iI_i)\\
\text{s.t.} \quad & \sum_{i\in\mathcal{F}}X_{ij\tau}\leqslant D_{j\tau}, \quad j\in\mathcal{N},\ \tau\in\mathcal{T},\\
& \sum_{j\in\mathcal{N}}X_{ij\tau}\leqslant P_{i\tau}, \quad i\in\mathcal{F},\ \tau\in\mathcal{T},\\
& X_{ij\tau}\geqslant 0, \quad i\in\mathcal{F},\ j\in\mathcal{N},\ \tau\in\mathcal{T}\\
& P_{i\tau}\leqslant Z_i,\ \ Z_i\leqslant M\cdot I_i, \quad i\in\mathcal{F},\ \tau\in\mathcal{T}\\
& I\in\{0,1\}^{|\mathcal{F}|}, \qquad \text{(where } M \text{ is a large enough constant.)}
\end{aligned}$$

An equivalent **deterministic** model, with **decisions for fractions of demands** Y:

$$\begin{aligned}
\max_{Y,I,Z,P} \quad & \sum_{\tau\in\mathcal{T}}\sum_{i\in\mathcal{F}}\sum_{j\in\mathcal{N}}(p-c^s_{ij})Y_{ij\tau}D_{j\tau}-\sum_{\tau\in\mathcal{T}}\sum_{i\in\mathcal{F}}c_iP_{i\tau}-\sum_{i\in\mathcal{F}}(C_iZ_i-K_iI_i)\\
\text{s.t.} \quad & \sum_{i\in\mathcal{F}}Y_{ij\tau}\leqslant 1, \quad j\in\mathcal{N},\ \tau\in\mathcal{T},\\
& \sum_{j\in\mathcal{N}}Y_{ij\tau}D_{j\tau}\leqslant P_{i\tau}, \quad i\in\mathcal{F},\ \tau\in\mathcal{T},\\
& Y_{ij\tau}\geqslant 0, \quad i\in\mathcal{F},\ j\in\mathcal{N},\ \tau\in\mathcal{T}\\
& P_{i\tau}\leqslant Z_i,\ \ Z_i\leqslant M\cdot I_i, \quad i\in\mathcal{F},\ \tau\in\mathcal{T}\\
& I\in\{0,1\}^{|\mathcal{F}|}, \qquad \text{(where } M \text{ is a large enough constant.)}
\end{aligned}$$

The 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!

Another Caveat...

Are Robust Solutions **Pareto-Efficient**?

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

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

Are Robust Solutions **Pareto-Efficient**?

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

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

- Classical RO framework results in
  - Optimal value $J^\star_{\mathsf{RO}}$
  - Set of robustly optimal solutions

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

Set of **Robustly Optimal** Solutions

- $X^{\mathsf{RO}} = \{x \in \mathcal{X} : \exists\, y \geqslant 0 \text{ such that } D^\intercal y = x, \quad y^\intercal d \geqslant J^\star_{\mathsf{RO}}\}$
- $x \in X^{\mathsf{RO}}$ guarantees that no other solution exists with higher **worst-case** objective value $u^\intercal x$

Set of **Robustly Optimal** Solutions

- $X^{\mathsf{RO}} = \{x \in \mathcal{X} : \exists\, y \geqslant 0 \text{ such that } D^\intercal y = x, \quad y^\intercal d \geqslant J^\star_{\mathsf{RO}}\}$

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

- 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 (Iancu & Trichakis 2014)

$$\max_{x \in \mathcal{X}} \quad \min_{u \in \mathcal{U}} u^\intercal x \tag{1}$$

Definition

A solution x is called a **Pareto Robustly Optimal (PRO) solution** for Problem (1) 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

$$\begin{aligned} u^\intercal \bar{x} &\geqslant u^\intercal x, \quad \forall u \in \mathcal{U}, \quad \text{and} \\ \bar{u}^\intercal \bar{x} &> \bar{u}^\intercal x, \ \text{for some } \bar{u} \in \mathcal{U}. \end{aligned}$$

Pareto Robustly Optimal solutions (Iancu & Trichakis 2014)

$$\max_{x \in \mathcal{X}} \quad \min_{u \in \mathcal{U}} u^\intercal x \tag{1}$$

Definition

A solution x is called a **Pareto Robustly Optimal (PRO) solution** for Problem (1) if

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

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

$$\begin{aligned} u^\intercal \bar{x} &\geqslant u^\intercal x, \quad \forall u \in \mathcal{U}, \quad \text{and} \\ \bar{u}^\intercal \bar{x} &> \bar{u}^\intercal x, \ \text{for some } \bar{u} \in \mathcal{U}. \end{aligned}$$

- $X^{PRO} \subseteq X^{RO}$: 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 $\mathcal{X}^{\text{PRO}}$?
- Can we characterize $\mathcal{X}^{\text{PRO}}$?
  - Is it non-empty?
  - Is it convex?
  - When is $\mathcal{X}^{\text{PRO}} = \mathcal{X}^{\text{RO}}$?
- 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:*

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

Then, either

- $\mathcal{U}^* \stackrel{\text{def}}{=} \{y \in \mathbb{R}^n : y^\intercal u \geqslant 0, \forall\, u \in \mathcal{U}\}$ 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:*

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

Then, either

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

- $\mathcal{U}^* \stackrel{\text{def}}{=} \{y \in \mathbb{R}^n : y^\intercal u \geqslant 0, \forall\, u \in \mathcal{U}\}$ is the dual of $\mathcal{U}$

Finding PRO solutions

Theorem

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

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

Then, either

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

- $\mathcal{U}^* \overset{\text{def}}{=} \{y \in \mathbb{R}^n : y^\intercal u \geqslant 0, \forall u \in \mathcal{U}\}$ 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

Corollaries

- If $\bar{u} \in \mathrm{ri}(\mathcal{U})$, all optimal solutions to the problem below are PRO:

$$\begin{array}{ll} \text{maximize} & \bar{u}^\intercal x \\ \text{subject to} & x \in X^{\text{RO}} \end{array}$$

- If $0 \in \mathrm{ri}(\mathcal{U})$, then $X^{\text{PRO}} = X^{\text{RO}}$

- If $\bar{u} \in \mathrm{ri}(\mathcal{U})$, then $X^{\text{PRO}} = X^{\text{RO}}$ if and only if the optimal value of this LP is zero:

$$\begin{array}{ll} \text{maximize} & \bar{u}^\intercal y \\ \text{subject to} & x \in X^{\text{RO}} \\ & y \in \mathcal{U}^* \\ & x + y \in \mathcal{X} \end{array}$$

Optimizing over / Understanding X^{PRO}

- Secondary objective r: can we solve

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

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

- Secondary objective r: can we solve

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

Proposition

X^{PRO} *is not necessarily convex.*

- $\mathcal{X} = \{x \in \mathbb{R}^4_+ \; : \; x_1 \leqslant 1, \; x_2 + x_3 \leqslant 6, \; x_3 + x_4 \leqslant 5, \; x_2 + x_4 \leqslant 5\}$
- $\mathcal{U} = \text{conv}\Big(\big\{e_i, \; i \in \{1, \ldots, 4\}\big\}\Big)$
- $J^\star_{RO} = 1$, and $X^{RO} = \{x \in X \; : \; x \geqslant \mathbf{1}\}$
- $x^1 = \begin{bmatrix} 1 & 2 & 4 & 1 \end{bmatrix}^\intercal$, $x^2 = \begin{bmatrix} 1 & 4 & 2 & 1 \end{bmatrix}^\intercal \in X^{PRO}$
- $0.5\, x^1 + 0.5\, x^2$ is Pareto dominated by $\begin{bmatrix} 1 & 3 & 3 & 2 \end{bmatrix}^\intercal \in X^{RO}$.

Optimizing over / Understanding X^{PRO}

- Secondary objective r: can we solve

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

Proposition

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

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

What Are The Gains?

Example (Portfolio)

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

What Are The Gains?

Example (Portfolio)

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

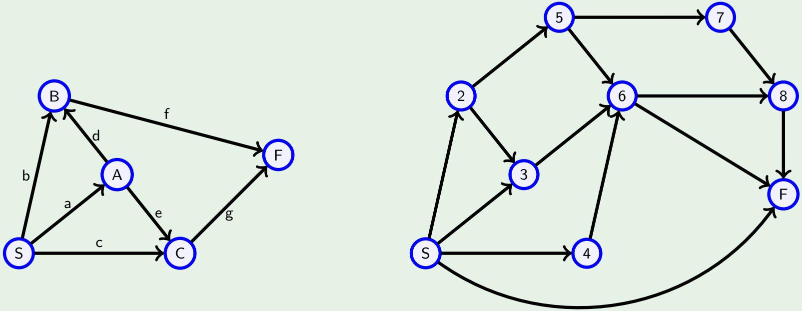
Example (Inventory)

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

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

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



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 $\tau_e = \tau_e^0 + \delta_e$

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

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

$$\begin{aligned} \tau_e &= \tau_e^0 + \delta_e - x_e \\ \mathbf{1}^\intercal x &\leqslant C \end{aligned}$$

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

Results – finance and inventory examples (10K instances)

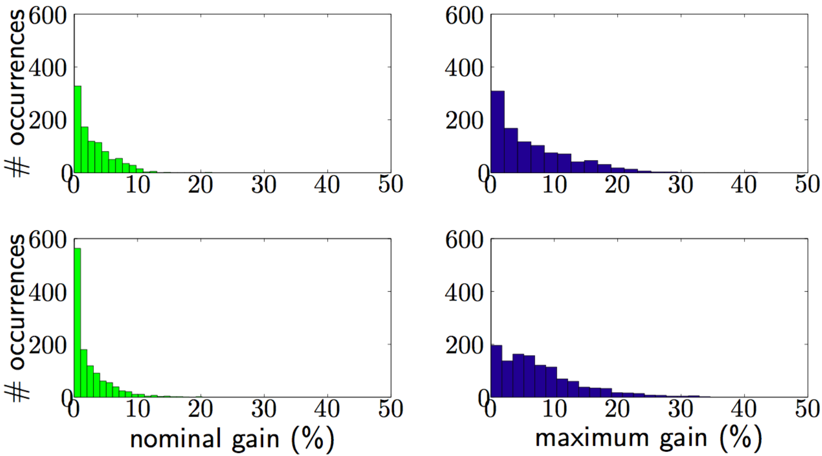
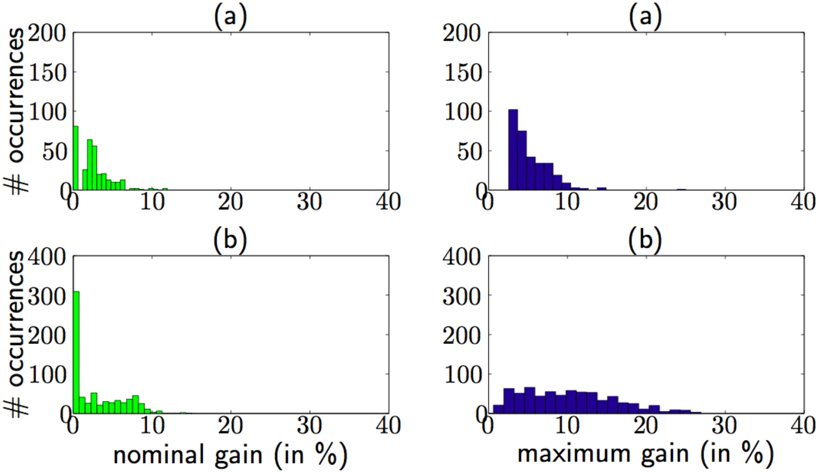


Figure: TOP: portfolio example. BOTTOM: inventory example. LEFT: performance gains in nominal scenario. RIGHT: maximal performance gains.

Results – two project management networks (10K instances)



Careful To Avoid Naïve Inefficiencies In Robust Models!

Dynamic Robust Optimization

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

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

Stochastic model:

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

Robust model:

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

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 $x, z \quad \to \quad$ find $y^\star(x, z) \quad \to \quad$ 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

Outline...

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. Applications

Consider the following *deterministic* inventory management problem:

$$\begin{aligned} \underset{\{x_t\}_{t=1}^T}{\text{minimize}} \quad & \sum_{t=1}^{T} \left(\overbrace{c_t x_t}^{\text{ordering cost}} + \overbrace{h_t (y_{t+1})^+}^{\text{holding cost}} + \overbrace{b_t (-y_{t+1})^+}^{\text{backlog cost}} \right) \\ \text{s.t.} \quad & y_{t+1} = y_t + x_t - d_t, \ \forall\, t, && \text{(Stock balance)} \\ & L_t \leqslant x_t \leqslant H_t, \ \forall\, t, && \text{(Min/max order size)} \\ & y_1 = a\,, && \text{(Initial stock level)} \end{aligned}$$

A simple motivating example

Consider the following *deterministic* inventory management problem:

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

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

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:

$$\begin{aligned}
\underset{\{x_t\}_{t=1}^{T}}{\text{minimize}} \quad & \sum_{t=1}^{T} \left(\overbrace{c_t x_t}^{\text{ordering cost}} + \overbrace{h_t (y_{t+1})^+}^{\text{holding cost}} + \overbrace{b_t (-y_{t+1})^+}^{\text{backlog cost}} \right) \\
\text{s.t.} \quad & y_{t+1} = y_t + x_t - d_t, \ \forall t, && \text{(Stock balance)} \\
& L_t \leqslant x_t \leqslant H_t, \ \forall t, && \text{(Min/max order size)} \\
& y_1 = a \,, && \text{(Initial stock level)}
\end{aligned}$$

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

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

Consider the following *deterministic* inventory management problem:

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

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

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

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

Ordering policies can depend on revealed demands:
$x_t(d_{[t-1]})$, 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:

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

where:

$$y_{t+1} := y_t + x_t - d_t$$

$$\mathcal{U}_t(d_{[t-1]}) := \left\{ d \in \mathbb{R} \, : \, \exists \, \xi \in \mathbb{R}^{T-t} \text{ such that } [d_{[t-1]};\ d\ ; \xi] \in \mathcal{U} \right\}$$

Robust Dynamic Programming Formulation

Our dynamic decision problem can also be written:

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

where:

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

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:

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

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:

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

Bellman Principle; Robust DP Recursions

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

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

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

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

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:

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

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:

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

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

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

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

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

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

Bellman Principle; Robust DP Recursions

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

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

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

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

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

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

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
    $$(\bar{a} + Pz)^\intercal x + d^\intercal y(z) \leqslant b, \quad \forall z \in \mathcal{U}$$
    where $y(z)$ is dynamically adjustable
  - A linear (affine) form $y = u + Vz$ would lead to the problem:
    $$\bar{a}^\intercal x + d^\intercal u + (P^\intercal x + V^\intercal d)^\intercal z \leqslant b, \quad \forall z \in \mathcal{U}.$$
    Constraint **linear** in decisions x, u, V and uncertainty z, so all previous results apply!

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

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

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

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

    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:

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

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:

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

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** $\mathcal{U}$?

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

Naïve Robustification

Consider a naïve robust optimization model:

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

Naïve Robustification

Consider a naïve robust optimization model:

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

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

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

Problem arises due to "$=$" constraint!

A less naïve robustification

Robustify an alternate linear programming formulation:

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

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:

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

A less naïve robustification

Robustify an alternate linear programming formulation:

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

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:

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

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

$$x_t\left(d_{[t-1]}\right) = x_t^0 + X_t d_{[t-1]}$$

$$s_t^+\left(d_{[t-1]}\right) = s_t^+ + S_t^+ d_{[t-1]}$$

$$s_t^-\left(d_{[t-1]}\right) = s_t^- + S_t^- d_{[t-1]}$$

Linear Decision Rules

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

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

- The **Robust Counterpart** problem becomes:

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

- **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), with **box** uncertainty

ρ (%)	**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), with **box** uncertainty

ρ (%)	**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 $x_t(\mathbf{d}_{[t-1]})$ and affine auxiliary variables $s_t^{+,-}(\mathbf{d}_{[t-1]})$ generate the optimal worst-case cost.*

Empirical Performance: Ben-Tal et al. ('04, '09), with **box** uncertainty

ρ (%)	**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* $x_t(\mathbf{d}_{[t-1]})$ *and affine auxiliary variables* $s_t^{+,-}(\mathbf{d}_{[t-1]})$ *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 an LP of $\mathcal{O}(m \cdot T^2)$.
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

A Critical Difference: Bellman Optimality in Stochastic and Robust Models

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

Stochastic model:

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

Robust model:

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

A Critical Difference: Bellman Optimality in Stochastic and Robust Models

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

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

A Critical Difference: Bellman Optimality in Stochastic and Robust Models

"Nature" reveals $z \quad \mapsto \quad$ DM choices $y(z)$

Stochastic model:

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

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?**

A Critical Difference: 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 $J^\star_{\text{rob}}$

A Critical Difference: Bellman Optimality in Stochastic and Robust Models

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

Stochastic model:

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

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^{wc} from the set

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

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

A Critical Difference: Bellman Optimality in Stochastic and Robust Models

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

Stochastic model:

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

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^{wc} from the set

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

    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
  - There are **infinitely many** worst-case optimal policies

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
   - I. and Trichakis [2014] discuss a potential fix

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

A Monitoring Problem

Disease Monitoring

Debt Monitoring

- **Significant uncertainty**
  - limited data to calibrate dynamic evolution
- **System can be monitored at finite number of times**
  - *e.g., healthcare*: testing requires office visit, expensive/invasive procedures
  - *e.g., (micro-)lending*: on-site visits, costly appraisals of collateral, etc.
  - **monitoring times must be chosen judiciously**
- **Complex, high-dimensional decision problem**
  - processes influence each other; monitoring / learning adds complexity

Robust Monitoring and Stopping [I., Trichakis, Yoon]

- Consider a system evolving over continuous time $[0, T]$
- State characterized by d processes, denoted $x(t) \in \mathbb{R}^d$
- A decision maker (DM) starts with initial information $x(0)$
- Can monitor the system at most n times, at $0 \leqslant t_1 \leqslant \cdots \leqslant t_n \leqslant T$. (let $t_0 \stackrel{\text{def}}{=} 0, t_{n+1} \stackrel{\text{def}}{=} T$)
- At each time t_p, the DM:
  - Observes the state $x_p \stackrel{\text{def}}{=} x(t_p)$
  - Updates information about possible future state values
  - Decides whether to stop or not
- When stopping at t, collect $g(t, x(t))$

Uncertainty Sets and Information Updating

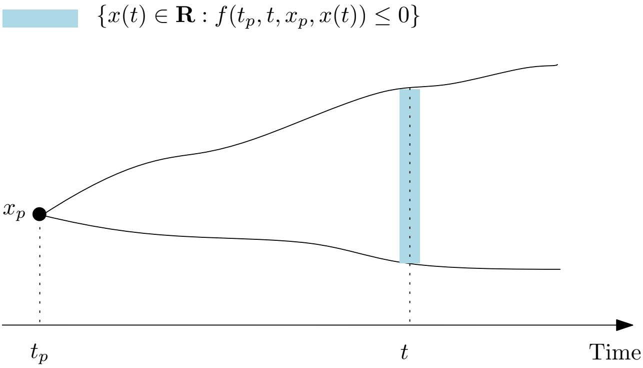
- An observation $\boldsymbol{x}_p$ at t_p imposes restrictions on future state values $\boldsymbol{x}(t)$ (for $t > t_p$) summarized through m constraints:

$$f(t_p, t, \boldsymbol{x}_p, \boldsymbol{x}(t)) \leqslant 0$$

Uncertainty Sets and Information Updating

- An observation x_p at t_p imposes restrictions on future state values $x(t)$ (for $t > t_p$) summarized through m constraints:

$$f(t_p, t, x_p, x(t)) \leqslant 0$$



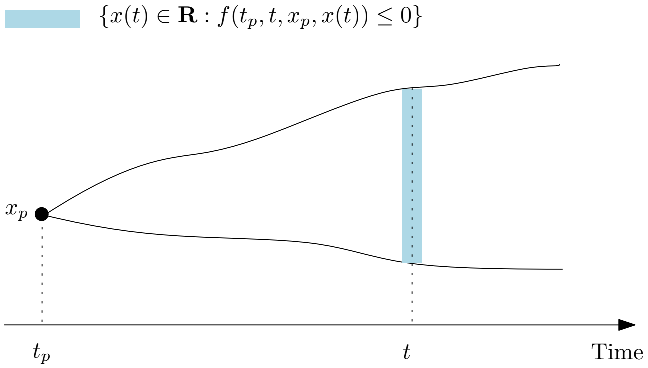
Case $d = 1$. Information about $x(t)$ acquired at $t_p < t$.

Uncertainty Sets and Information Updating

- An observation x_p at t_p imposes restrictions on future state values $x(t)$ (for $t > t_p$) summarized through m constraints:

$$f(t_p, t, x_p, x(t)) \leqslant 0$$

- $x(t)$ consistent with *all* restrictions from observations before t



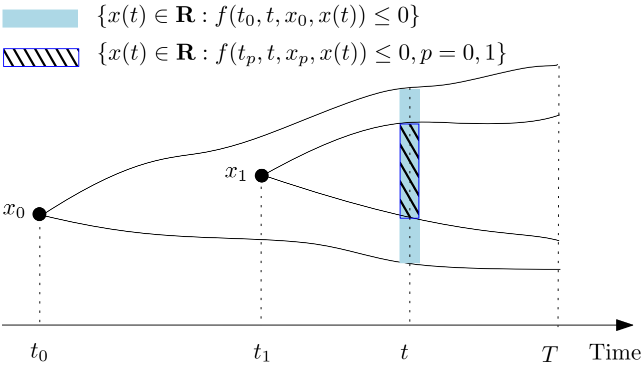
Case $d = 1$. Information about $x(t)$ acquired at $t_p < t$.

Uncertainty Sets and Information Updating

- An observation x_p at t_p imposes restrictions on future state values $x(t)$ (for $t > t_p$) summarized through m constraints:

$$f(t_p, t, x_p, x(t)) \leqslant 0$$

- $x(t)$ consistent with *all* restrictions from observations before t



Case $d = 1$. Information about $x(t)$ acquired at t_0 and t_1.

Uncertainty Sets and Information (Updating)

- Suppose DM committed to r monitoring times: $\mathbf{t}^{\{r\}} = [t_0, t_1, \ldots, t_r]$
- DM made $k \leqslant r$ observations so far: $\boldsymbol{x}^{\{k\}} = [\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_k]$.
- The future possible process values at times $t_{k+1}, \ldots, t_r, T$ lie in:

$$\mathcal{U}\left(\mathbf{t}^{\{r\}}, \boldsymbol{x}^{\{k\}}\right) = \Big\{[\boldsymbol{x}_{k+1}, \ldots, \boldsymbol{x}_r, \boldsymbol{x}_{n+1}] \in \mathbb{R}^{d \times (r-k+1)} : \\ f(t_p, t_q, \boldsymbol{x}_p, \boldsymbol{x}_q) \leqslant 0, \forall\, p, q \in \{0, 1, \ldots, r, n+1\}, p < q\Big\},$$

where $\mathbf{t}^{\{r\}} = [t_0, t_1, \ldots, t_r]$ and

- **Notation.** Let $\mathcal{U}_{k+1}$ = set of possible values for $\boldsymbol{x}_{k+1}$ (by projecting $\mathcal{U}$ above)

DM's Monitoring and Stopping Problem

- **Goal:** Find monitoring and stopping policy to maximize worst-case reward
  - Two versions, depending on choice of monitoring times

DM's Monitoring and Stopping Problem

- **Goal:** Find monitoring and stopping policy to maximize worst-case reward
  - Two versions, depending on choice of monitoring times
- **Static:** pick all times $t_1, \ldots, t_n$ at time $t_0 = 0$

DM's Monitoring and Stopping Problem

- **Goal:** Find monitoring and stopping policy to maximize worst-case reward
  - Two versions, depending on choice of monitoring times

- **Static:** pick all times $t_1, \ldots, t_n$ at time $t_0 = 0$
  - At time t_k, DM solves the problem:

$$V_k(\mathbf{t}^{\{n+1\}}, \boldsymbol{x}^{\{k\}}) = \max\Big(\underbrace{g\big(t_k, x_k\big)}_{\text{stop}}, \underbrace{\min_{x_{k+1} \in \mathcal{U}_{k+1}(\mathbf{t}^{\{n+1\}}, \boldsymbol{x}^{\{k\}})} V_{k+1}\big(\mathbf{t}^{\{n+1\}}, \boldsymbol{x}^{\{k+1\}}\big)}_{\text{continue}}\Big),$$

  - At time t_0, DM solves: $\mathbf{t}^S \in \arg\max_{\mathbf{t}^{\{n+1\}}} V_0(\mathbf{t}^{\{n+1\}}, \boldsymbol{x}^{\{0\}})$.

DM's Monitoring and Stopping Problem

- **Goal:** Find monitoring and stopping policy to maximize worst-case reward
  - Two versions, depending on choice of monitoring times
- **Static:** pick all times $t_1, \dots, t_n$ at time $t_0 = 0$
  - At time t_k, DM solves the problem:

$$V_k(\mathbf{t}^{\{n+1\}}, \mathbf{x}^{\{k\}}) = \max\Big(\underbrace{g(t_k, x_k)}_{\text{stop}}, \underbrace{\min_{x_{k+1} \in \mathcal{U}_{k+1}(\mathbf{t}^{\{n+1\}}, \mathbf{x}^{\{k\}})} V_{k+1}(\mathbf{t}^{\{n+1\}}, \mathbf{x}^{\{k+1\}})}_{\text{continue}}\Big),$$

  - At time t_0, DM solves: $\mathbf{t}^S \in \arg\max_{\mathbf{t}^{\{n+1\}}} V_0(\mathbf{t}^{\{n+1\}}, \mathbf{x}^{\{0\}})$.
- **Dynamic:** at time t_k, only pick next monitoring time t_{k+1}

DM's Monitoring and Stopping Problem

- **Goal:** Find monitoring and stopping policy to maximize worst-case reward
  - Two versions, depending on choice of monitoring times

- **Static:** pick all times $t_1, \dots, t_n$ at time $t_0 = 0$
  - At time t_k, DM solves the problem:

$$V_k(\mathbf{t}^{\{n+1\}}, \boldsymbol{x}^{\{k\}}) = \max\Big(\underbrace{g(t_k, \boldsymbol{x}_k)}_{\text{stop}}, \underbrace{\min_{\boldsymbol{x}_{k+1} \in \mathcal{U}_{k+1}(\mathbf{t}^{\{n+1\}}, \boldsymbol{x}^{\{k\}})} V_{k+1}(\mathbf{t}^{\{n+1\}}, \boldsymbol{x}^{\{k+1\}})}_{\text{continue}}\Big),$$

  - At time t_0, DM solves: $\mathbf{t}^S \in \arg\max_{\mathbf{t}^{\{n+1\}}} V_0(\mathbf{t}^{\{n+1\}}, \boldsymbol{x}^{\{0\}})$.

- **Dynamic:** at time t_k, only pick next monitoring time t_{k+1}
  - At time t_k, DM solves the problem:

$$J_k(\mathbf{t}^{\{k\}}, \boldsymbol{x}^{\{k\}}) = \max\Big(g(t_k, \boldsymbol{x}_k), \max_{t_{k+1} \in (t_k, T]} \min_{\boldsymbol{x}_{k+1} \in \mathcal{U}_{k+1}(\mathbf{t}^{\{k+1\}}, \boldsymbol{x}^{\{k\}})} J_{k+1}(\mathbf{t}^{\{k+1\}}, \boldsymbol{x}^{\{k+1\}})\Big),$$

  - DM seeks a **monitoring policy**: $\tau_k^D(\mathbf{t}^{\{k\}}, \boldsymbol{x}^{\{k\}})$.

Assumptions

Assumption (Monotonic Rewards)

$g(t, x)$ *component-wise monotonic in* x.

- Each state is either “good” or “bad”

Assumptions

Assumption (Increasing Rewards)

$g(t, \boldsymbol{x})$ *component-wise* **increasing** *in* $\boldsymbol{x}$.

Assumptions

Assumption (Increasing Rewards)

$g(t, \mathbf{x})$ *component-wise* **increasing** *in* $\mathbf{x}$.

Assumption (U-set Structure)

For any $0 \leqslant k \leqslant r \leqslant n$ *and given* $\mathbf{t}^{\{r\}}$ *and* $\mathbf{x}^{\{k\}}$,

i. **(Lattice)** $\mathcal{U}\left(\mathbf{t}^{\{r\}}, \mathbf{x}^{\{k\}}\right)$ *is a lattice;*

ii. **(Monotonicity)** $\mathcal{U}\left(\mathbf{t}^{\{r\}}, \mathbf{x}^{\{k\}}\right)$ *is increasing in* $\mathbf{x}^{\{k\}}$;

iii. **(Consistency)** $\mathcal{U}_{k+1}(\mathbf{t}^{\{r\}}, \mathbf{x}^{\{k\}}) = \mathcal{U}_{k+1}(\mathbf{t}^{\{r'\}}, \mathbf{x}^{\{k\}})$.

Assumptions

Assumption (Increasing Rewards)

$g(t, \mathbf{x})$ *component-wise* **increasing** *in* $\mathbf{x}$.

Assumption (U-set Structure)

For any $0 \leqslant k \leqslant r \leqslant n$ *and given* $\mathbf{t}^{\{r\}}$ *and* $\mathbf{x}^{\{k\}}$,

i. **(Lattice)** $\mathcal{U}\left(\mathbf{t}^{\{r\}}, \mathbf{x}^{\{k\}}\right)$ *is a lattice;*
ii. **(Monotonicity)** $\mathcal{U}\left(\mathbf{t}^{\{r\}}, \mathbf{x}^{\{k\}}\right)$ *is increasing in* $\mathbf{x}^{\{k\}}$*;*
iii. **(Consistency)** $\mathcal{U}_{k+1}(\mathbf{t}^{\{r\}}, \mathbf{x}^{\{k\}}) = \mathcal{U}_{k+1}(\mathbf{t}^{\{r'\}}, \mathbf{x}^{\{k\}})$.

- Lattice: technical requirement
- Monotonicity: better past $\rightarrow$ better future
- Consistency: future monitoring times $t_{k+2}, \ldots, t_r$ do not impact possible values for x_{k+1}

Examples

Example (Lattice with Cross-Constraints)

For $\mathcal{M} \subseteq \{1, \dots, d\}^2$, $\ell : \mathbb{R}^2 \to \mathbb{R}_-$ decreasing in its second argument, and $u : \mathbb{R}^2 \to \mathbb{R}_+$ increasing in its second argument:

$$\mathcal{U}\big(\mathbf{t}^{\{r\}}, \boldsymbol{x}^{\{k\}}\big) = \Big\{[\boldsymbol{x}_{k+1}, \dots, \boldsymbol{x}_r, \boldsymbol{x}_{n+1}] \in \mathbb{R}^{d\times(r-k+1)} : \\ x_p^m + \ell(t_p, t_q - t_p) \leqslant x_q^{m'} \leqslant x_p^m + u(t_p, t_q - t_p), \\ \forall\, (m, m') \in \mathcal{M},\ \forall p, q \in \{0, 1, \dots, r, n+1\},\ p < q\Big\}.$$

Examples

Example (Lattice with Cross-Constraints)

For $\mathcal{M} \subseteq \{1, \ldots, d\}^2$, $\ell : \mathbb{R}^2 \to \mathbb{R}_-$ decreasing in its second argument, and $u : \mathbb{R}^2 \to \mathbb{R}_+$ increasing in its second argument:

$$\mathcal{U}\left(\mathbf{t}^{\{r\}}, \mathbf{x}^{\{k\}}\right) = \Big\{[\mathbf{x}_{k+1}, \ldots, \mathbf{x}_r, \mathbf{x}_{n+1}] \in \mathbb{R}^{d\times(r-k+1)} : \\ x_p^m + \ell(t_p, t_q - t_p) \leqslant x_q^{m'} \leqslant x_p^m + u(t_p, t_q - t_p), \\ \forall\, (m, m') \in \mathcal{M},\ \forall p, q \in \{0, 1, \ldots, r, n+1\},\ p < q\Big\}.$$

Example (CLT-Budgeted Uncertainty Sets)

For $\Gamma > 0$, $\sigma > 0$, and μ,

$$\mathcal{U}(\mathbf{t}^{\{r\}}, \mathbf{x}^{\{k\}}) = \Big\{[x_{k+1}, \ldots, x_r, x_{n+1}] \in \mathbb{R}^{r-k+1} : \\ -\Gamma \leqslant \frac{x_q - x_p - (t_q - t_p)\mu}{\sqrt{t_q - t_p}\,\sigma} \leqslant \Gamma, \forall p, q \in \{0, \ldots, r, n+1\},\ p < q\Big\}.$$

Finding the Dynamic Monitoring Policy

Finding the Dynamic Monitoring Policy

Theorem (I., Trichakis, Yoon '18)

Under Assumption 1 and Assumption 2, static monitoring achieves **the same worst-case optimal reward** *as dynamic monitoring.*

Finding the Dynamic Monitoring Policy

Theorem (I., Trichakis, Yoon '18)

Under Assumption 1 and Assumption 2, static monitoring achieves **the same worst-case optimal reward** *as dynamic monitoring.*

- Can recover dynamic policy by repeatedly resolving for static policies
- Result extends to more general decision problem
  - When monitoring, DM can extract or inject values into processes

Finding the Dynamic Monitoring Policy

Theorem (I., Trichakis, Yoon '18)

Under Assumption 1 and Assumption 2, static monitoring achieves **the same worst-case optimal reward** *as dynamic monitoring.*

- Can recover dynamic policy by repeatedly resolving for static policies
- Result extends to more general decision problem
  - When monitoring, DM can extract or inject values into processes

Theorem (Solving Static Problem)

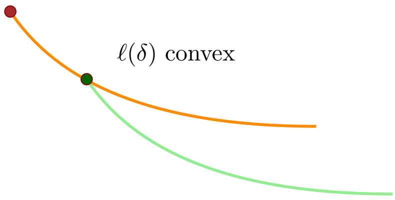
The worst-case optimal value under static monitoring can be obtained as:

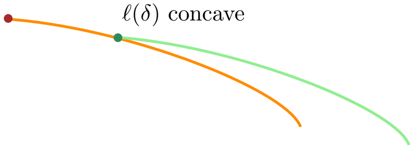
$$\max_{\mathbf{t}^{\{n+1\}}} \max_{k \in \{n, n+1\}} g\big(t_k, \underline{x}_k(\mathbf{t}^{\{n+1\}})\big).$$

- Without loss, can choose times to either stop at t_n or at T
- $\underline{x}_k(\mathbf{t}^{\{n+1\}})$ is the worst-case scenario (smallest state under $\mathbf{t}^{\{n+1\}}$)

Stationary Uncertainty Sets

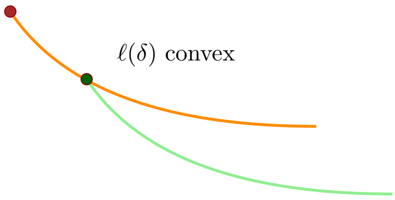
Consider uncertainty sets with $\ell(t, \delta) = \ell(\delta)$.

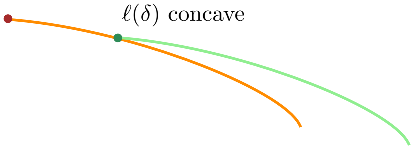




Stationary Uncertainty Sets

Consider uncertainty sets with $\ell(t, \delta) = \ell(\delta)$.

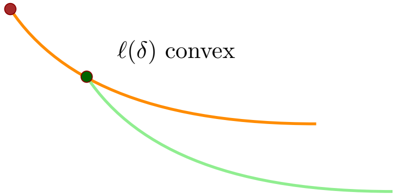


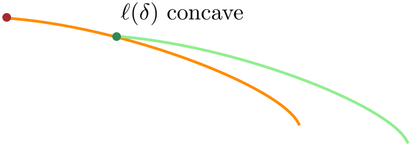


- Worst-case optimal to **stop at** t_1
  - $t_2, \ldots, t_n$ redundant
- Find t_1: solve 1D optimization

Stationary Uncertainty Sets

Consider uncertainty sets with $\ell(t, \delta) = \ell(\delta)$.





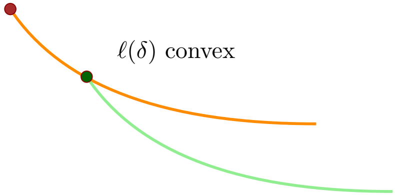
- Worst-case optimal to **stop at** t_1
  - $t_2, \ldots, t_n$ redundant
- Find t_1: solve 1D optimization

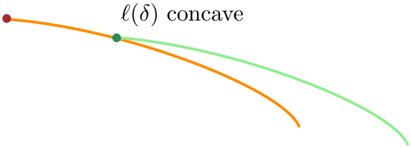
- Worst-case optimal to either:
  - Stop at T; monitor at $t_k^\star = \frac{kT}{n+1}$

    or
  - Stop at t_n; monitor at $t_k^\star \frac{k\,t_n}{n}$
- Find t_n: solve 1D optimization
- **Uniform-interval** monitoring optimal!

Stationary Uncertainty Sets

Consider uncertainty sets with $\ell(t, \delta) = \ell(\delta)$.





- Worst-case optimal to **stop at** t_1
  - $t_2, \ldots, t_n$ redundant
- Find t_1: solve 1D optimization

- Worst-case optimal to either:
  - Stop at T; monitor at $t_k^\star = \frac{k\,T}{n+1}$ or
  - Stop at t_n; monitor at $t_k^\star \frac{k\,t_n}{n}$
- Find t_n: solve 1D optimization
- **Uniform-interval** monitoring optimal!

- Paper also has additional results on general case ...
- ... and a case study in Cardiac Allograft Vasculopathy

Conclusions

- Robust decision making: a very relevant/realistic framework
  - *"When do we really have complete probabilistic descriptions?"*
- A powerful framework
  - **Flexible:** allows embedded various levels of information
  - **Tractable:** can solve many classes of problems (if suitable formulated)
  - **Not necessarily conservative** (if suitable formulated)
- Very useful (theory and practice)
  - theory: can be used to rationalize simple rules that work well
- Has a few "quirks"
  - Careful with formulating nominal model
  - Careful to avoid inefficiencies
- Specific areas in OM where it could be used more:
  - contracting in complex value chains (developing world, disruption risk, ...)
  - behavioral operations

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.