15.094J: Robust Modeling, Optimization, Computation

Lecture 1: Why RO? Probability Theory and its Limitations

February 2015

Outline

Objectives Today

- Administration
- Why RO?
- What are the new primitives of uncertainty?
- What do robust solutions look like and what they offer?

Administration

- **Time**: Wednesdays, 9-12am.
- **Place**: E62-221.
- **Professor**: Dimitris Bertsimas, E40-147, tel. (617) 253-4223.
- **Office hours**: By appointment, e-mail: dbertsim@mit.edu. http://web.mit.edu/dbertsim/www/
- **TAs**:
  - Martin Copenhaver, E40-135; Tel: (617) 253-7412; Office hours: Tuesdays and Thursdays 1-2pm; e-mail: mcopen@mit.edu.
  - Nishanth Mundru, E40-130, Tel: (617) 253-7412; Office hours: Tuesdays and Thursdays 2-3pm; e-mail: nmundru@mit.edu
- **Recitation**: Mondays 1-2pm, E51-057

Administration

- **Text:** Research papers and class notes in https://stellar.mit.edu.
- **Recitations:** Background on optimization and probability, software for RO, computational aspects, examples and applications.
- **Course Requirements:** 30% problem sets, 30% midterm exam, and 40% final team project.
- **Background required:** Mathematical maturity. Knowledge of optimization and probability we will develop in recitations.

Why RO?

- Data Uncertainty
- Implementation Error
- Computational Limitations of Probability Theory in higher dimensions

Motivation: Data Uncertainty

- Tacoma Bridge
- Panama Canal
- A constraint from PILOT4 in the NETLIB library:

$$
\begin{gathered}
-15.79081 \cdot x_{826} - \textcolor{blue}{\mathit{8.598819}} \cdot x_{827} - 1.88789 \cdot x_{828} - 1.362417 \cdot x_{829} \\
-1.526049 \cdot x_{830} - 0.031883 \cdot x_{849} - 28.725555 \cdot x_{850} - 10.792065 \cdot x_{851} \\
-0.19004 \cdot x_{852} - 2.757176 \cdot x_{853} - 12.290832 \cdot x_{854} + 717.562256 \cdot x_{855} \\
-0.057865x \cdot x_{856} - 3.785417 \cdot x_{857} - 78.30661 \cdot x_{858} - 122.163055 \cdot x_{859} \\
-6.46609 \cdot x_{860} - 0.48371 \cdot x_{861} - 0.615264 \cdot x_{862} - 1.353783 \cdot x_{863} \\
-84.644257 \cdot x_{864} - 122.459045 \cdot x_{865} - 43.15593 \cdot x_{866} - 1.712592 \cdot x_{870} \\
-0.401597 \cdot x_{871} + \textcolor{red}{\mathit{1}} \cdot x_{880} - 0.946049 \cdot x_{898} - 0.946049 \cdot x_{916} \\
\geq 23.387405
\end{gathered}
$$

- Numbers such as *8.598819* are estimated and potentially inaccurate.
- Numbers such as *1* are probably certain.

Motivation: Data Uncertainty

Optimal solution

$$\begin{aligned} x^*_{826} &= 255.6112787181108, & x^*_{827} &= 6240.488912232100, \\ x^*_{828} &= 3624.613324098961, & x^*_{829} &= 18.20205065283259, \\ x^*_{849} &= 174397.0389573037, & x^*_{870} &= 14250.00176680900, \\ x^*_{871} &= 25910.00731692178, & x^*_{880} &= 104958.3199274139, \ldots \end{aligned}$$

Suppose the coefficients are only 0.1% inaccurate. Will above solution still be feasible?

Motivation: Data Uncertainty

Optimal solution

$$\begin{aligned} x^*_{826} &= 255.6112787181108, & x^*_{827} &= 6240.488912232100, \\ x^*_{828} &= 3624.613324098961, & x^*_{829} &= 18.2020506528259, \\ x^*_{849} &= 174397.0389573037, & x^*_{870} &= 14250.00176680900, \\ x^*_{871} &= 25910.00731692178, & x^*_{880} &= 104958.3199274139, \ldots \end{aligned}$$

Suppose the coefficients are only 0.1% inaccurate. Will above solution still be feasible?

Simulating with 1000 samples:

$$P(\text{constraint infeasible}) = 50\%,$$

$$\text{worst infeasibility} \ : \sum_{i=1}^{n} a_i x_i^* - b = -104.9, \quad \text{as opposed to} \ \geq 0.$$

Motivation: Data Uncertainty

Optimal solution

$$
\begin{aligned}
x^*_{826} &= 255.6112787181108, & x^*_{827} &= 6240.488912232100,\\
x^*_{828} &= 3624.613324098961, & x^*_{829} &= 18.2020506528325 9,\\
x^*_{849} &= 174397.0389573037, & x^*_{870} &= 14250.00176680900,\\
x^*_{871} &= 25910.00731692178, & x^*_{880} &= 104958.3199274139, \ldots
\end{aligned}
$$

Suppose the coefficients are only 0.1% inaccurate. Will above solution still be feasible?

Simulating with 1000 samples:

$$P(\text{constraint infeasible}) = 50\%,$$

$$\text{worst infeasibility} \ : \sum_{i=1}^{n} a_i x_i^* - b = -104.9, \quad \text{as opposed to} \ \geq 0.$$

From NETLIB Library:

- For 27/90 problems, solutions were *infeasible* for very small perturbations
- For 17/90 problems, solutions were infeasible by at least 50%.

Implementation Errors: Robust antenna design

- In a 2-D plane, antennae radiation patterns (i.e., the *radiation diagram*) vary with the incidence angle θ.
- Using a weighted array of linear antenna components, we can control the shape of the resulting diagram. Specifically, the diagram for a linear array of N elements spaced apart at distance d has the form

$$Z(\theta) = \sum_{l=0}^{N-1} w_l \exp\left(j\left(\frac{2\pi d}{\lambda}\right) l \sin\theta\right),$$

where w_l are the weights of the components (our design variables) and λ is the wavelength of the radiation.
- The weights w_l are generally complex-valued (i.e., we control both magnitude and phase).

Robust antenna design

- Often we wish to keep $Z(\theta)$ large in some angle range around a target angle θ_t and small outside this range.
- In other words, with Δ as the angular "beam-width," keep $Z(\theta)$ close to one (normalized) for $\theta \in [\theta_t - \Delta, \theta_t + \Delta]$, and close to zero for θ outside this range.
- Typical design problem: minimize the *uniform norm*, i.e., minimize the maximum deviation from the desired target pattern over all angles of interest. The *nominal design problem* then is

$$\text{minimize} \qquad \max_{\theta \in \Theta} |Z(\theta) - \delta(\theta)|,$$

  where $\Theta = \{\theta \mid |\theta - \theta_t| \geq \Delta\} \cup \{\theta_t\}$ and $\delta(\theta) = 1$ if $\theta = \theta_t$ and 0 otherwise.
- With complex design weights, this problem may be solved efficiently.

Implementation error

- In practice, however, we cannot implement the design weights with exact precision. One simple way to model this is to assume the actual weights $\tilde{w}_l$ are of the form

$$\tilde{w}_l \quad = \quad (1 + \xi_l) w_l,$$

  where w_l are the desired weights, and ξ are zero-mean random variables capturing the imperfect implementation precision.
- In general, these uncontrollable perturbations can have an adverse effect on the performance of the nominal design. The *robust design problem* wishes to protect against all such perturbations within some "reasonable" uncertainty set.

Robust formulation

- The specific uncertainty set we use is

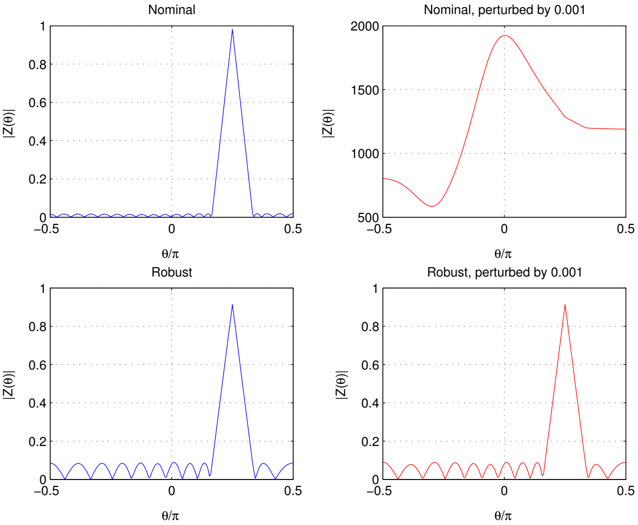
$$P_\gamma = \{\boldsymbol{\xi} \mid \|\boldsymbol{\xi}\| \leq \gamma\},$$

  parameterized by some protection level γ.
- The *robust design problem at level* γ, then, is

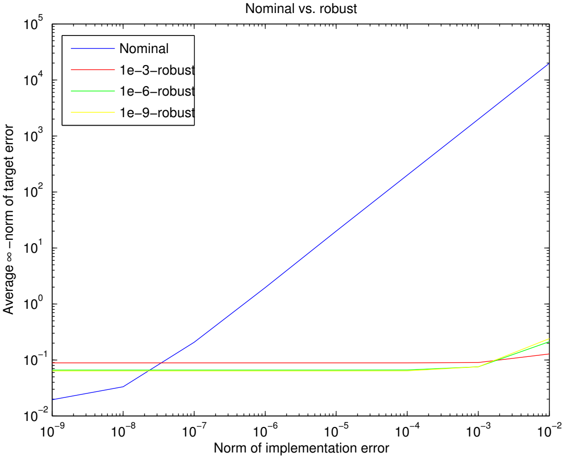
$$\text{minimize} \quad \max_{\theta \in \Theta, \boldsymbol{\xi} \in P_\gamma} |\tilde{Z}(\theta, \boldsymbol{\xi}) - \delta(\theta)|,$$

  where $\tilde{Z}(\theta, \boldsymbol{\xi})$ is the antenna diagram perturbed by implementation error $\boldsymbol{\xi}$.

$N = 20$, $d = 1.2\lambda$, $\theta_t = \pi/4$, and $\Delta = \pi/12$, $\gamma = 0.001$



Robust vs Nominal



Summary

- Even extremely small perturbations can have adverse effects on the performance of an antenna design which ignores these implementation errors.
- Robust solutions may be calculated efficiently. These solutions offer vastly superior protection in the face of random errors and do not perform much worse than the nominal designs in the absence of implementation errors.

Major applications of applied probability in the 20th century

Successes	**Challenges**
Queueing Theory: Erlang (1909) Modeling queues with exponential distributions in steady-state	Transient analysis, analysis of queueing networks with general distributions
Information Theory: Shannon (1948) Characterization of the capacity of a single user channel	Network Information theory
Market Design: Myerson (1981) Characterization of the optimal auction for single item and multiple buyers	Multiple items for buyers with budget constraints, and correlated valuations
Options Pricing: Black and Scholes (1973) Pricing options in low dimensions in complete markets	Pricing options in high dimensions in incomplete markets
Stochastic Optimization: Bellman (1957) Dynamic Programming, successful in low dimensions	Extending computation to high dimensions
Low Dimensions	*High Dimensions*

Philosophy / Motivation

- Performance analysis given that primitives are probability distributions does not extend in a computationally tractable way in *high dimensions*.
- Combining probability theory and optimization often leads to the *"Curse of dimensionality"*.
- Contrast with linear, convex, even discrete *optimization*.
- What was *Kolmogorov*'s intention?
- What was *Dantzig*'s intention?

Past Voices

"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)

"If a queue has an arrival process which cannot be well modeled by a Poisson process or one of its near relatives, it is likely to be difficult to fit any simple model, still less to analyze it effectively. So why do we insist on regarding the arrival times as random variables, quantities about which we can make sensible probabilistic statements? Would it not be better to accept that the arrivals form an irregular sequence, and carry out our calculations without positing a joint probability distribution over which that sequence can be averaged? " – J.F.C. Kingman (2009)

Where Probability Distributions Exist ?

- What is available in practice is **data**, not probability distributions.
- Probability distributions are not known in practice; they exist in *our imagination*.
- Thus, modeling with probability distributions is a *choice*. Is it the right one?
- But even if they are available, what happens if the future distribution is not the same as in the past?
- Examples: Hurricane Katrina and the Chernobyl nuclear disaster.

Example: A Capacity Expansion Problem

You are considering entering 4 new markets. You could build multiple factories (or none) in each market, but you have a finite budget of $500 M. The cost of building a factory differs by market (see below).

If you build a factory, you'll earn revenue over the long term, whose NPV is also represented below.

Market	1	2	3	4
Cost ($M / factory)	120	100	180	140
NPV Future Revenue ($B)	50	40	60	30

How many factories should you build in each market?

Example: A Capacity Expansion Problem (continued)

You might solve the following optimization problem:

$$\begin{array}{rl} \text{maximize} & 50x_1 + 40x_2 + 60x_3 + 30x_4 \\ \text{subject to} & 120x_1 + 100x_2 + 180x_3 + 140x_4 \leq 500 \\ & x_i \geq 0 \quad \forall i \\ & x_i \text{ integer.} \end{array}$$

Solution: $x_1^* = 4$, $x_2^* = x_3^* = x_4^* = 0$.

Example: A Capacity Expansion Problem (continued)

In reality, the upfront costs and the projected revenues may not be known exactly. Considering just the costs, suppose they are distributed according to a distribution $f(x)$.

Project	1	2	3	4
Expected Cost	120	100	180	140
St Dev of Cost	12	10	18	14

- How would you compute the probability that the above solution won't break the budget?
- What if the costs are correlated?
- How would you find the best solution which is feasible at least 95 % of the time?

Tractability of Stochastic Modeling

- Even for simple distributions and small problems, calculating the probability of violation of a constraint is often difficult.
- Simulation is only an approximation and may be very computationally expensive.
- Problems get worse as number of variables and constraints grow.
- Optimization in this environment is even harder.
- Another example: Options pricing in high dimensions.

Conclusions; Why RO?

- In my experience in the real world, *robustness* is often more important than *optimality*.
- *Motivation I:* Create solutions that are immune to implementation errors and data uncertainty.
- *Motivation II:* Develop a theory of performance analysis and optimization under uncertainty via optimization that is tractable in high dimensions.
- A remark on *Tractability:* We do not mean polynomial solvability. Rather the ability to solve problems of the size that we care in applications and within computational times appropriate for the application.

Objectives of the Class:

- Propose an alternative to stochastic *modeling* via RO to model uncertain phenomena.
- Develop RO as a tractable methodology for solving *optimization* problems under uncertainty.
- Cover a large number of *applications*.
- Expose you to large scale *computation* for RO.

Lectures

Lecture	Time	Topic	Readings
1	W, 2/04	Probability theory and its limitations	[1]
2	W, 2/04	The new primitives: Uncertainty sets	[1]
3	W, 2/11	Robust linear optimization I	[11]
4	W, 2/11	Robust linear optimization II	[29, 14]
5	W, 2/18	Robust mixed integer optimization	[30, 31]
6	W, 2/18	Robust convex optimization	[10, 32]
7	W, 2/25	Data driven robust optimization	[22]
8	W, 2/25	From data to decisions	[25]
9	W, 3/04	Adaptive multistage optimization I	[9, 23, 8]
10	W, 3/04	Adaptive multistage optimization II	[24, 15] [18, 17]
11	W, 3/11	Power of robust policies in adaptive optimiz.	[21, 20]
12	W, 3/11	Power of affine policies in adaptive optimiz.	[19, 23]
13	W, 3/18	RO in supply chains	[33]

Lectures

Lecture	Time	Topic	Readings
14	W, 3/18	RO in energy	[26]
	W, 3/25	Spring break	
	W, 4/01	Midterm	
15	W, 4/08	Robust portfolios	[35, 28]
16	W, 4/08	Robust options pricing	[5]
17	W, 4/15	RO and risk preferences	[13]
18	W, 4/15	Constructing utilities via RO	[27]
19	W, 4/22	Robust steady-state queueing theory	[7]
20	W, 4/22	Robust transient queueing theory	[6]
21	W, 4/29	Robust mechanism design	[4]
22	W, 4/29	RO in statistics	[34, 16]
23	W, 5/06	Robust Kalman filtering	[12]
24	W, 5/06	Robust information theory	[2, 3]
25	W, 5/13	Project presentations	
26	W, 5/13	Project presentations	

C. Bandi and D. Bertsimas.
Tractable stochastic analysis in high dimensions via robust optimization.
Mathematical Programming, Series B, 134(1):23–70, 2012.

C. Bandi and D. Bertsimas.
Channel coding via robust optimization, part i: The single channel case.
Submitted to *IEEE Transactions on Information Theory*, 2014.

C. Bandi and D. Bertsimas.
Channel coding via robust optimization, part ii: The multiple channel case.
Submitted to *IEEE Transactions on Information Theory*, 2014.

C. Bandi and D. Bertsimas.
Optimal design for multi-item auctions: A robust optimization approach.
Mathematics of Operations Research, 39(4):1012–1038, 2014.

C. Bandi and D. Bertsimas.
Robust option pricing - An ϵ-arbitrage approach.
European Journal of Operational Research, 239(3):842–853, 2014.

C. Bandi, D. Bertsimas, and N. Youssef.
Robust transient queueing theory.
Submitted to *Operations Research*, 2013.

C. Bandi, D. Bertsimas, and N. Youssef.
Robust queueing theory.
Operations Research, to appear, 2015.

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

A. Ben-Tal, A. Goryashko, A. Nemirovski, and J.P. Vial.
Retailer-supplier flexible commitments contracts: A robust optimization approach.
Manufacturing and Service Operations Management, 7(3):248–271, 2005.

A. Ben-Tal and A. Nemirovski.
Robust convex optimization.
Mathematics of Operations Research, 23(4):769–805, 1998.

A. Ben-Tal and A. Nemirovski.
Robust solutions of uncertain linear programs.
Operations Research Letters, 25(1):1–13, 1999.

D. Bertsimas and D. Brown.
Constrained stochastic LQC: A tractable approach.

IEEE Transactions on Automatic Control, 52:1826–1841, 2007.

D. Bertsimas and D. Brown.
Constructing uncertainty sets for robust linear optimization.
Operations Research, 57(6):1483–1495, 2009.

D. Bertsimas, D. Brown, and C. Caramanis.
Theory and applications of robust optimization.
SIAM Review, 53(3):464–501, 2011.

D. Bertsimas and C. Caramanis.
Finite adaptability in linear optimization.
IEEE Transations on Automatic Control, 55(12):2751–2766, 2010.

D. Bertsimas and M. Copenhaver.
Characterization of the equivalence of robustification and regularization in linear, median, and matrix regression.
submitted to *Annals of Statistics*, 2014.

D. Bertsimas and I. Dunning.
Multistage robust mixed integer optimization with adaptive partitions.
submitted to *Operations Research*, 2014.

D. Bertsimas and A. Georghiou.

Design of near optimal decision rules in multistage adaptive mixed-integer optimization.
Operations Research, to appear, 2015.

D. Bertsimas and V. Goyal.
On the power and limitations of affine policies in two-stage adaptive optimization.
Mathematical Programming, Series B, 134(2):491–531, 2011.

D. Bertsimas and V. Goyal.
On the approximability of adjustable robust convex optimization under uncertainty.
Mathematical Methods of Operations Research, 77(3):323–343, 2013.

D. Bertsimas, V. Goyal, and X.A. Sun.
A geometric characterization of the power of finite adaptability in multi-stage stochastic and adaptive optimization.
Mathematics of Operations Research, 36(1):24–54, 2011.

D. Bertsimas, V. Gupta, and N. Kallus.
Data-driven robust optimization.
Submitted to *Operations Research*, 2013.

D. Bertsimas, D. Iancu, and P. Parrilo.
Optimality of affine policies in multi-stage robust optimization.
Mathematics of Operations Research, 35(2):363–394, 2010.

D. Bertsimas, D. Iancu, and P. Parrilo.
A hierarchy of near-optimal policies for multi-stage adaptive optimization.
IEEE Transactions on Automatic Control, 56(12):2809–2824, 2011.

D. Bertsimas and N. Kallus.
From predictive to prescriptive analytics.
submitted to *Management Science*, 2015.

D. Bertsimas, E. Litvinov, X.A. Sun, J. Zhao, and T. Zheng.
Adaptive robust optimization for the security constrained unit commitment problem.
IEEE Transactions on Power Systems, 28(1):52–63, 2013.

D. Bertsimas and A. O'Hair.
Learning preferences under noise and loss aversion: An optimization approach.
Operations Research, 61(5):1190–1199, 2013.

D. Bertsimas and D. Pachamanova.

Robust multiperiod portfolio management in the presence of transaction costs.
Computers and Operations Research, 35(1):3–17, 2008.

D. Bertsimas, D. Pachamanova, and M. Sim.
Robust linear optimization under general norms.
Operations Research Letters, 32(6):510–516, 2004.

D. Bertsimas and M. Sim.
Robust discrete optimization and network flows.
Mathematical Programming, Series B, 98(1-3):49–71, 2003.

D. Bertsimas and M. Sim.
The price of robustness.
Operations Research, 52(1):35–53, 2004.

D. Bertsimas and M. Sim.
Tractable approximations to robust conic optimization problems.
Mathematical Programming, Series B, 107(1-2):5–36, 2006.

D. Bertsimas and A. Thiele.
A robust optimization approach to inventory theory.
Operations Research, 54:150–168, 2006.

L. El-Ghaoui and H. Lebret.
Robust solutions to least-square problems to uncertain data matrices.
SIAM Journal on Matrix Analysis and Applications, 18(4):1035–1064, 1997.

D. Goldfarb and G. Iyengar.
Robust portfolio selection problems.
Mathematics of Operations Research, 28(1):1–38, 2003.

15.094J: Robust Modeling, Optimization, Computation

Lectures 2: The New Primitives: Uncertainty Sets

February, 2015

Outline

1. What is RO?
2. Robust Modeling
   - Constructing Uncertainty Sets
   - Modeling Correlation Information
   - Typical sets
3. RO Insights

Linear Optimization

- **Nominal problem**

$$\begin{array}{rl} \text{maximize} & c_1x_1 + \ldots + c_nx_n \\ \text{subject to} & a_{11}x_1 + \ldots + a_{1n}x_n \leq b_1 \\ & \vdots \\ & a_{m1}x_1 + \ldots + a_{mn}x_n \leq b_m \\ & x_i \geq 0. \end{array}$$

- **Robust Problem**

$$\begin{array}{rll} \text{maximize} & c_1x_1 + \ldots + c_nx_n & \\ \text{subject to} & a_{11}x_1 + \ldots + a_{1n}x_n \leq b_1 & \\ & \vdots & \forall (a_{11}, \ldots, a_{mn}) \in \mathcal{U} \\ & a_{m1}x_1 + \ldots + a_{mn}x_n \leq b_m & \\ & x_i \geq 0. & \end{array}$$

- **What if $(c_1, \ldots, c_n, b_1, \ldots, b_m)$ are also uncertain?**

Robust modeling

- Replace probability distributions as primitives with *uncertainty sets*.
- Use *worst case* analysis, while bounding the power of nature - *Robust Optimization (RO).*
- Use *conclusions of probability theory*, and not its (Kolmogorov (1933)) axioms, to define uncertainty sets.

Constructing Uncertainty Sets

How do we construct uncertainty sets?

- We first suggest an approach based on the Central Limit Theorem by building on our previous example.
- Later, we'll see some more advanced approaches.

Constructing Uncertainty Sets: A First Try

Recall our previous example:

Project	1	2	3	4
Expected Cost	120	100	180	140
St Dev of Cost	12	10	18	14

- Although the cost of building a factory won't be exactly equal to its mean, we expect it to be close.
- Suggests Uncertainty set

$$-\Gamma \leq \frac{a_i - \overline{a}_i}{\sigma_i} \leq \Gamma.$$

Constructing Uncertainty Sets: A First Try (continued)

We might then solve:

$$\begin{array}{rl} \text{maximize} & 50x_1 + 40x_2 + 60x_3 + 30x_4 \\ \text{subject to} & a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 \leq 500 \text{ for all } \mathbf{a} \in \mathcal{U} \\ & x_i \text{ integer.} \end{array}$$

where $\mathbf{a} \in \mathcal{U}$ means

$$-\Gamma \leq \frac{a_1 - 120}{12} \leq \Gamma$$
$$-\Gamma \leq \frac{a_2 - 100}{10} \leq \Gamma$$
etc.

- Notice that we insist the solution be feasible for *any* value of the costs. (In particular, the *worst* values.)
- By varying the value of Γ, (e.g. $\Gamma = 2, 3$), we control level of robustness.

A Criticism of our first set $\mathcal{U}$

- A fair criticism of this approach is that it is unlikely that all the costs are at their worst value at the same time.
- Suggests we should pick a better uncertainty set.
- Suggestions?

Central Limit Theorem

- X_i: iid with mean μ and standard deviation σ.
- Recall our old friend, the CLT

$$\frac{X_1 + \ldots + X_n - n\mu}{\sigma \cdot \sqrt{n}} \sim N(0, 1).$$

- Intuitively, the CLT tells us that sums of random variables tend to be close to their mean.
- How might we use this to improve our uncertainty set?

Constructing Uncertainty Sets: CLT

- Uncertainty set

$$-\Gamma\sqrt{n} \leq \sum_{i=1}^{n} \frac{a_i - \bar{a}_i}{\sigma_i} \leq \Gamma\sqrt{n}$$

- Sum of variations from mean values is limited.
- Γ still controls the degree of robustness.
- For $\Gamma = 2$, $\mathbb{P}\left[\mathbf{a} \in \mathcal{U}\right] \sim 0.95$.
- For $\Gamma = 3$, $\mathbb{P}\left[\mathbf{a} \in \mathcal{U}\right] \sim 0.997$.

Later, we'll see some numerical examples of how the above sets affect solution quality.

Modeling Correlation Information

- Factor model : $\{\tilde{z}_i\}_{i=1,\ldots,n}$ depend on m factors $\tilde{f} = \left(\tilde{f}_1, \tilde{f}_2, \ldots, \tilde{f}_m\right)$

$$\tilde{z}_i = \mathbf{a}_i' \cdot \tilde{f} + \tilde{\epsilon}_i,$$

  $\{\tilde{\epsilon}_i\}$ are i.i.d.

- $$\mathcal{U}^{\text{Corr}} = \left\{ (z_1, \ldots, z_n) \;\middle|\; \begin{array}{ll} z_i = \displaystyle\sum_{j=1}^{m} a_{ij} f_j + \epsilon_i, & \forall i = 1, \ldots, n, \\ -\Gamma_f \leq \dfrac{\displaystyle\sum_{j=1}^{m} f_j - m \cdot \mu_f}{\sigma_f \cdot \sqrt{m}} \leq \Gamma_f, & \\ -\Gamma_\epsilon \leq \dfrac{\displaystyle\sum_{i=1}^{n} \epsilon_i - n \cdot \mu_\epsilon}{\sigma_\epsilon \cdot \sqrt{n}} \leq \Gamma_\epsilon. & \end{array} \right\}.$$

Typical Sets: Incorporating Distributional Information

- Shannon (1948) introduced the idea of Typical Sets:
- Property (a): A typical set has probability nearly 1.
- Property (b): All elements of typical set are nearly equiprobable.
- Given pdf $f(\cdot)$,

$$\mathcal{U}^{f-\text{Typical}} = \left\{ (z_1, \ldots, z_n) \left| -\Gamma \leq \frac{\sum_{i=1}^{n} \log f(z_i) - n \cdot \mu_f}{\sigma_f \cdot \sqrt{n}} \leq \Gamma. \right. \right\},$$

$$\mu_f = \int_{-\infty}^{\infty} f(x) \log f(x) \, \mathrm{d}x,$$
$$\sigma_f = \int_{-\infty}^{\infty} f(x) \left(\log f(x) - \mu_f\right)^2 \mathrm{d}x.$$

Theorem

- (a) $\mathbb{P}\left[\tilde{\mathbf{z}} \in \mathcal{U}^{\text{f-Typical}}\right] \to g(\Gamma) = 2\Phi(\Gamma) - 1$, as $n \to \infty$.
- (b) The conditional pdf $h(\tilde{\mathbf{z}}) = f(\tilde{\mathbf{z}} | \tilde{\mathbf{z}} \in \mathcal{U}^{\text{f-Typical}})$ satisfies:

$$\left| \frac{1}{n} \log h(\tilde{\mathbf{z}}) - \mu_f \right| \leq \epsilon_n,$$

  with $\epsilon_n \to 0$, as $n \to \infty$.
- $\tilde{u}_j = \log f\left(\tilde{z}_j\right)$, iid. Apply CLT: as $n \to \infty$,

$$\frac{\sum_{j=1}^{n} \tilde{u}_j - n\mu_f}{\sigma_f \cdot \sqrt{n}} \sim N\left(0, 1\right),$$

- Let $\tilde{\mathbf{z}} \in \mathcal{U}^{\text{f-Typical}}$.

$$h(\tilde{\mathbf{z}}) = f\left(z_1\right) f\left(z_2\right) \ldots f\left(z_n\right).$$

- Since $\tilde{\mathbf{z}} \in \mathcal{U}^{\text{f-Typical}}$, we have

$$\left| \frac{1}{n} \log h(\tilde{\mathbf{z}}) - \mu_f \right| = \left| \frac{1}{n} \sum_{j=1}^{n} \log f(z_j) - \mu_f \right| \leq \frac{\Gamma \cdot \sigma_f}{\sqrt{n}} \to 0,$$

Typical Sets

- $\tilde{z}_i \sim N(0, \sigma)$

$$\mathcal{U}_\epsilon^{\mathrm{G}} = \left\{ \mathbf{z} \,\middle|\, -\Gamma_\epsilon^{\mathrm{G}} \le \|\mathbf{z}\|^2 - n\sigma^2 \le \Gamma_\epsilon^{\mathrm{G}} \right\}.$$

- $\tilde{z}_i \sim Exp(\lambda)$

$$\mathcal{U}_\epsilon^{\mathrm{E}} = \left\{ \mathbf{z} \,\middle|\, \frac{n}{\lambda} - \frac{\sqrt{n}}{\lambda} \cdot \Gamma_\epsilon^{\mathrm{E}} \le \sum_{j=1}^{n} z_j \le \frac{n}{\lambda} + \frac{\sqrt{n}}{\lambda} \cdot \Gamma_\epsilon^{\mathrm{E}},\ \mathbf{z} \ge \mathbf{0} \right\}.$$

- $\tilde{z}_i \sim U[a, b]$

$$\mathcal{U}_\epsilon^{U} = \left\{ \mathbf{z} \,\middle|\, \begin{array}{c} n\dfrac{a+b}{2} - \Gamma_\epsilon^{U}\sqrt{n} \le \displaystyle\sum_{j=1}^{n} z_j \le n\dfrac{a+b}{2} + \Gamma_\epsilon^{U}\sqrt{n}, \\ a \le z_j \le b,\ j = 1, \ldots, n, \end{array} \right\}..$$

- $\tilde{z}_i \sim \mathrm{Bin}(p)$

$$\mathcal{U}_\epsilon^{B} = \left\{ \mathbf{z} \,\middle|\, \begin{array}{c} np - \Gamma_\epsilon^{B}\sqrt{n} \le \displaystyle\sum_{j=1}^{n} z_j \le np + \Gamma_\epsilon^{B}\sqrt{n}, \\ z_j \in \{0, 1\},\ j = 1, \ldots, n, \end{array} \right\}.$$

Insights

Recall our previous capacity expansion problem.

- Nominal problem

$$\begin{array}{ll} \text{maximize} & 50x_1 + 40x_2 + 60x_3 + 30x_4 \\ \text{subject to} & 120x_1 + 100x_2 + 180x_3 + 140x_4 \leq 500 \\ & x_i \text{ integer.} \end{array}$$

- Nominal Solution: $x_1^* = 4$, $x_2^* = x_3^* = x_4^* = 0$.

Market	1	2	3	4
Expected Cost	120	100	180	140
St Dev of Cost	12	10	18	14

We will compare the solution of the nominal problem, the problem with our naive uncertainty set, and our CLT based uncertainty set in terms of their feasibility and optimality.

Uncertainty Sets

- **Robust 1:**

$$\mathcal{U} = \left\{ a : -\Gamma \leq \frac{a_i - \overline{a}_i}{\sigma_i} \leq \Gamma \right\}$$

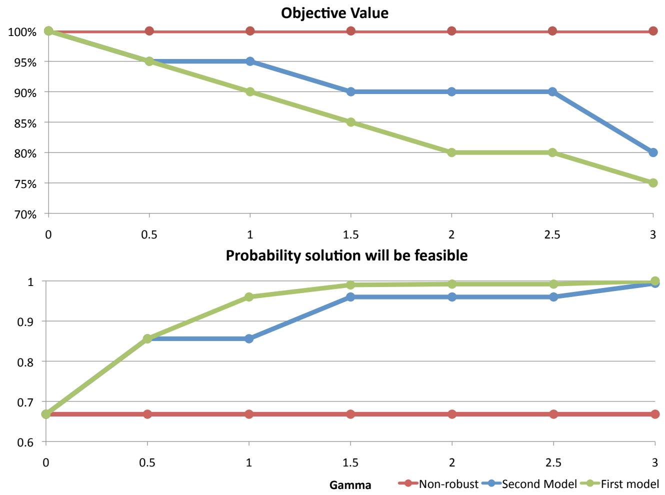
- **Robust 2:**

$$\mathcal{U} = \left\{ a : \sum_{i=1}^{n} \left| \frac{a_i - \overline{a}_i}{\sigma_i} \right| \leq \Gamma \sqrt{n} \right\}$$

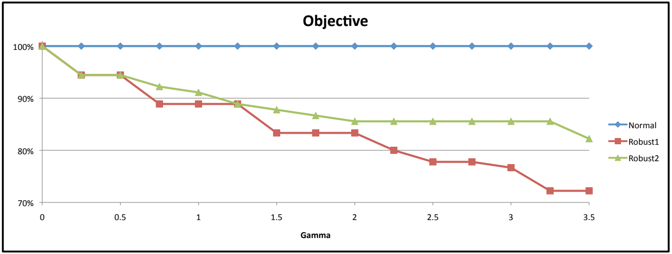
Robust Solution

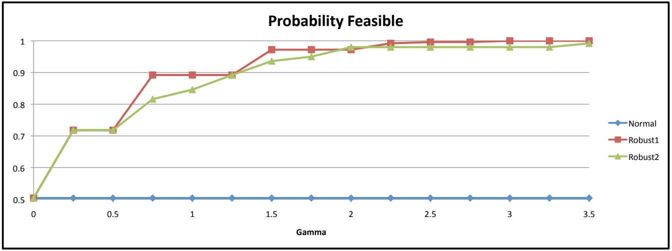
Problem	Γ	Objective	x_1	x_2	x_3	x_4	Unused Capital
Nominal	0	200	4	0	0	0	20
Robust 1	0.5	190	3	1	0	0	40
Robust 1	1	180	2	2	0	0	60
Robust 1	1.5	170	1	3	0	0	80
Robust 1	2	160	0	4	0	0	100
Robust 1	2.5	160	0	4	0	0	100
Robust 1	3	150	3	0	0	0	140
Robust 2	0.5	190	3	1	0	0	40
Robust 2	1	190	3	1	0	0	40
Robust 2	1.5	180	2	2	0	0	60
Robust 2	2	180	2	2	0	0	60
Robust 2	2.5	180	2	2	0	0	60
Robust 2	3	160	2	0	1	0	80

Tradeoff of robustness and optimality



10 variables





Modeling Demand

- We collected historical data: d_t. $t = 1, \ldots, T$.
- D_t, is future demand for day $t = 1, \ldots, n$.
- Compute $\mu = \dfrac{\sum_{t=1}^{T} d_t}{T}$.
- $\sigma^2 = \dfrac{\sum_{t=1}^{T}(d_t - \mu)^2}{T-1}$.
- $U = \{(D_1, \ldots, D_n) | \; -\Gamma \cdot \sigma \cdot \sqrt{n} \leq \sum_{t=1}^{n}(D_t - \mu) \leq \Gamma \cdot \sigma \cdot \sqrt{n}, \; |D_t - \mu| \leq \Gamma_1 \sigma\}$.

Insights

- Key intuition: Model uncertainty by conclusions of probability not its axioms.
- Often by sacrificing a bit of optimality, we can ensure feasibility for a large range of uncertainty.
- The price of robustness is often not great.

15.094J: Robust Modeling, Optimization, Computation

Lectures 3: Robust Linear Optimization I: Tractability

February 2015

Outline

1. RLO with Row-wise uncertainty
2. RLO with Row-wise Polyhedral Uncertainty
3. RLO with Row-wise Ellipsoidal uncertainty
4. RLO with General Polyhedral Uncertainty

Objectives Today

- Tractability of RLO
- Row-wise uncertainty
- General uncertainty

Row-wise Uncertainty

- Primitives: Uncertainty sets $\mathcal{U}_i$, $i = 1, \ldots, m$, $\mathbf{b}, c$ (known, WLOG).
- RLO with row-wise uncertainty:

$$\begin{array}{rll} \max & c'x & \\ \text{s.t.} & a_i'x \leq b_i & \forall a_i \in \mathcal{U}_i, \; i = 1, \ldots, m, \\ & x \geq \mathbf{0}. & \end{array}$$

- Note that the problem has infinitely many constraints.
- Reformulation:

$$\begin{array}{rl} \max & c'x \\ \text{s.t.} & \max_{a_i \in \mathcal{U}_i} a_i'x \leq b_i \\ & x \geq \mathbf{0}. \end{array}$$

- Note that the uncertainty for different constraints is independent.

Tractability

- Suppose that $\mathcal{U}_i$, $i = 1, \ldots, m$ are convex sets.
- Given an x, we can solve $i = 1, \ldots, m$:

$$\max_{a_i \in \mathcal{U}_i} a_i' x,$$

  efficiently.
- How should we solve the RLO problem?

Theoretical Tractability

- Solve the nominal problem; find x_0.
- Separation problem: Given an x_0, does there exist an $a_i \in \mathcal{U}_i$ that violates the constraint $a_i'x > b_i$?
- Solution: Solve $\max_{a_i \in \mathcal{U}_i} a_i'x$ and check whether

$$\max_{a_i \in \mathcal{U}_i} a_i'x \leq b_i.$$

- This shows that if $\mathcal{U}_i$ are convex, we can solve the separation problem in polynomial time, thus we can solve the RLO with convex uncertainty sets in polynomial time using the Ellipsoid method (see Chapter 8 of Bertsimas and Tsitsiklis [1997]).
- The key take away from this: Even though RLO has infinitely many constraints it is polynomially solvable.
- Question: How about practically solvable? The Ellipsoid method is not a practical algorithm.

Practical Tractability

- Solve the nominal problem; find x_0.
- Solve $\max_{a_i \in \mathcal{U}_i} a_i' x_0$, solution $\overline{a}_{i,0}$.
- Add the constraint $\overline{a}_{i,0}' x \leq b_i$ to the nominal problem
- Solve (the dual Simplex method is the right choice)

$$
\begin{aligned}
\max \quad & c'x \\
\text{s.t.} \quad & \overline{a}_{i,0}' x \leq b_i \\
& x \geq \mathbf{0}.
\end{aligned}
$$

- Find x_1; iterate.

Robust Counterpart-Polyhedral uncertainty

- $$\begin{array}{rl} \max & c'x \\ \text{s.t.} & \max_{a_i \in \mathcal{U}_i} a_i'x \le b_i. \\ & x \ge \mathbf{0}. \end{array}$$
- $\mathcal{U}_i = \{a_i |\ D_i a_i \le d_i\}$, $D_i :\ k_i \times n$.
- Consider the problem and its dual:

$$\begin{array}{rl} \max & a_i'x \\ \text{s.t.} & D_i a_i \le d_i \end{array} \qquad \begin{array}{rl} \min & p_i'd_i \\ \text{s.t.} & p_i'D_i = x' \\ & p_i \ge \mathbf{0}. \end{array}$$

Robust Counterpart continued

- RC becomes

$$
\begin{aligned}
\max_{x, p_i} \quad & c'x \\
\text{s.t.} \quad & p_i' d_i \leq b_i, \quad && i = 1, \ldots, m, \\
& p_i' D_i = x', \quad && i = 1, \ldots, m, \\
& p_i \geq \mathbf{0}, \quad && i = 1, \ldots, m, \\
& x \geq \mathbf{0}.
\end{aligned}
$$

- Original nominal problem: n variables, m constraints.
- Uncertainty dimension: k_i.
- Size of Robust Counterpart: $n + \sum_{i=1}^{m} k_i$, variables; $m + m \cdot n$ constraints.

Row-wise Ellipsoidal uncertainty

- RO:

$$\begin{aligned} \max \quad & c'x \\ \text{s.t.} \quad & \max_{a_i \in \mathcal{U}_i} a_i'x \leq b_i. \\ & x \geq \mathbf{0}. \end{aligned}$$

- $\mathcal{U}_i = \{a_i |\ a_i = \overline{a}_i + \mathbf{\Delta}_i' u_i,\ ||u_i||_2 \leq \rho\}$, $\mathbf{\Delta}_i :\ k_i \times n,\ u_i : k_i \times 1.$
- RC:

$$\begin{aligned} \max \quad & c'x \\ \text{s.t.} \quad & \overline{a}_i'x + \rho||\mathbf{\Delta}_i x||_2 \leq b_i, \qquad i = 1, \ldots, m. \\ & x \geq \mathbf{0}. \end{aligned}$$

- Second order cone problem, nearly as tractable as linear optimization.

Proof

- $Z^* = \max_{a \in \mathcal{U}} a'x = \overline{a}'x + \max_{||u|| \leq \rho} u'(\mathbf{\Delta} x)$
- Lagrangean dual:

$$Z(\lambda) = \overline{a}'x + \max \ u'(\mathbf{\Delta} x) - \lambda(u'u/2 - \rho^2/2).$$

- $u^* = \mathbf{\Delta} x/\lambda.$
-

$$Z(\lambda) = \overline{a}'x + \frac{1}{2}\left(\frac{||\mathbf{\Delta} x||^2}{\lambda} + \lambda\rho^2\right).$$

- For $\lambda \geq 0$, $Z^* \leq Z(\lambda)$ and strong duality: $Z^* = \min_{\lambda \geq 0} Z(\lambda)$.
- $\lambda^* = ||\mathbf{\Delta} x||/\rho.$
- $Z^* = \overline{a}'x + \rho||\mathbf{\Delta} x||.$

Robust Counterpart-General Norm uncertainty

- RO:

$$\begin{array}{ll} \max & c'x \\ \text{s.t.} & \max_{a_i \in \mathcal{U}_i} a_i'x \leq b_i. \\ & x \geq \mathbf{0}. \end{array}$$

- $U_i = \{a_i | \ a_i = \bar{a}_i + \mathbf{\Delta}_i' u_i, \ ||u_i|| \leq \rho\}$, $\mathbf{\Delta}_i : \ k_i \times n, \ u_i : k_i \times 1.$
- Dual norm:

$$||s||^* = \max_{\{||x|| \leq 1\}} |s'x|.$$

- The dual of the L_p-norm $||x||_p = \left(\sum_{j=1}^n |x_j|^p\right)^{1/p}$:
- $||s||^* = ||s||_q$ with $q = 1 + \dfrac{1}{p-1}$.
- The dual norm of the L_2 norm is L_2.
- The dual norm of the L_1 norm is the L_∞ norm.
- RC:

$$\begin{array}{lll} \max & c'x & \\ \text{s.t.} & \bar{a}_i'x + \rho||\mathbf{\Delta}_i x||^* \leq b_i, & i = 1, \ldots, m. \\ & x \geq \mathbf{0}. & \end{array}$$

General Polyhedral Uncertainty

- Define the operator $\text{vec}(A) := (a_1, a_2, \ldots, a_m)$ (vector concatenation of the rows of A transposed)
-
$$\begin{array}{rll} \max & c'x & \\ \text{s.t.} & \tilde{A}x \leq b, & \forall \tilde{A} \in \mathcal{U} \\ & x \in P. & \end{array}$$
- $\mathcal{U} = \{\tilde{A} \mid G \cdot \text{vec}(\tilde{A}) \leq d\}$,
- $G \in \Re^{l \times (m \cdot n)}$, $d \in \Re^{l \times 1}$, and $\text{vec}(\tilde{A}) \in \Re^{(m \cdot n) \times 1}$.

RC

- The RC is

$$\begin{aligned} \max \quad & c'x \\ \text{s.t.} \quad & p_i'G = x_i', \quad && i = 1, \dots, m \\ & p_i'd \le b_i, && i = 1, \dots, m \\ & p_i \ge \mathbf{0}, && i = 1, \dots, m \\ & x \in P, \end{aligned}$$

- $p_i \in \Re^{l \times 1}$.
- $x_i \in \Re^{(m \cdot n) \times 1}$, $i = 1, \dots, m$; x_i contains x in entries $(i-1) \cdot n + 1$ through $i \cdot n$, and zero everywhere else.

Proposition

- Suppose $\mathcal{U} \neq \emptyset$.
- A given $\hat{x}$ satisfies $\tilde{a}_i'\hat{x} \leq b_i$ for all $\tilde{A} \in \mathcal{U}$ if and only if there exists a vector $p_i \in \Re^{l \times 1}$ such that

$$\begin{array}{rcl} p_i'd & \leq & b_i \\ p_i'G & = & \hat{x}_i' \\ p_i & \geq & \mathbf{0} \end{array}$$

- $\hat{x}_i \in \Re^{(m \cdot n) \times 1}$ contains $\hat{x}$ in entries $(i-1) \cdot n + 1$ through $i \cdot n$, and zero everywhere else.

Proof

- Consider the primal-dual pair

$$\begin{array}{rl} \max_A & a_i'\hat{x} \\ \text{s.t.} & G \cdot \text{vec}(A) \leq d \end{array}$$

$$\begin{array}{rl} \min_{p_i} & p_i'd \\ \text{s.t.} & p_i'G = \hat{x}_i' \\ & p_i \geq \mathbf{0}. \end{array}$$

- Suppose that $\hat{x}$ satisfies $\tilde{a}_i'\hat{x} \leq b_i$ for all $\tilde{A} \in \mathcal{U}$.
- Then, $\max_A a_i'\hat{x} \leq b_i$.
- Then primal is feasible and bounded, and so is its dual.
- Thus, there exists a vector $p_i \in \Re^{(m \cdot n) \times 1}$ satisfying the dual constraints.
- By strong duality, the optimal objective function value of the dual equals $\max_A a_i'\hat{x}$ and is less than b_i.

Proof continued

- For the reverse, since $\mathcal{U} \neq \emptyset$, primal is feasible. Suppose there exists a vector $p_i \in \Re^{l \times 1}$ that satisfies the dual constraints.
- Since both problems are feasible, they must be bounded and their optimal objective function values must be equal.
- Then $\min_{p_i} p_i'd \leq p_i'd \leq b_i$.
- By strong duality, $\max_A a_i'\hat{x} = \min_{p_i} p_i'd \leq b_i$, and hence $\hat{x}$ satisfies $a_i'\hat{x} \leq b_i$ for all $\tilde{A} \in \mathcal{U}$.

RC

- RO:

$$\begin{array}{rll} \max & c'x & \\ \text{s.t.} & \tilde{A}x \leq b, & \forall \tilde{A} \in \mathcal{U} \\ & x \in P. & \end{array}$$

- $\mathcal{U} = \{\tilde{A} \mid G \cdot \text{vec}(\tilde{A}) \leq d\}$.
- The RC is

$$\begin{array}{rll} \max & c'x & \\ \text{s.t.} & p_i'G = x_i', & i = 1, \dots, m \\ & p_i'd \leq b_i, & i = 1, \dots, m \\ & p_i \geq \mathbf{0}, & i = 1, \dots, m \\ & x \in P. & \end{array}$$

General uncertainty sets under a general norm

- RO:

$$\begin{aligned} \max \quad & c'x \\ \text{s.t.} \quad & \tilde{A}x \leq b \\ & x \in P \\ & \forall \tilde{A} \in \mathcal{U} = \left\{ \tilde{A} \mid ||M(\text{vec}(\tilde{A}) - \text{vec}(\overline{A}))|| \leq \Delta \right\}. \end{aligned}$$

- RC:

$$\begin{aligned} \max \quad & c'x \\ \text{s.t.} \quad & \overline{a}_i x + \Delta ||M^{-1} x_i||^* \leq b_i, \quad i = 1, \ldots, m \\ & x \in P, \end{aligned}$$

- M invertible
- $x_i \in \Re^{(m \cdot n) \times 1}$ contains $x \in \Re^{n \times 1}$ in entries $(i-1) \cdot n + 1$ through $i \cdot n$, and 0 everywhere else.

Proof

- $y = \dfrac{M(\text{vec}(\tilde{A}) - \text{vec}(\overline{A}))}{\Delta}$.
- Then, $\mathcal{U} = \{y : \ ||y|| \leq 1\}$.

$$
\begin{aligned}
\max_{\{\tilde{A} \in \mathcal{U}\}} \left\{\tilde{a}_i{}'x\right\} &= \max_{\{\tilde{A} \in \mathcal{U}\}} \left\{(\text{vec}(\tilde{A}))'x_i\right\} \\
&= \max_{\{y: \ ||y|| \leq 1\}} \left\{(\text{vec}(\overline{A}))'x_i + \Delta(M^{-1}y)'x_i\right\} \\
&= \overline{a}_i'x + \Delta \max_{\{y|||y|| \leq 1\}} \left\{y'(M^{-1}x_i)\right\} \\
&= \overline{a}_i x + \Delta ||M^{-1}x_i||^*
\end{aligned}
$$

References

Dimitris Bertsimas and John Tsitsiklis. *Introduction to Linear Optimization*. Athena Scientific, 1997.

15.094J: Robust Modeling, Optimization, Computation

Lecture 4: RLO: Probabilistic Guarantees

February 2015

Outline

1. Guarantees for independent uncertainty
2. Guarantees for non-independent distributions
3. Philosophical Remarks

Objectives Today

- Probabilistic Guarantees for RLO
- Insights in selecting parameters

Row-wise Ellipsoidal uncertainty

- RO:

$$\begin{array}{rl} \max & c'x \\ \text{s.t.} & \max_{a_i \in \mathcal{U}_i} a_i'x \le b_i. \\ & x \ge \mathbf{0}. \end{array}$$

- $\mathcal{U}_i = \{a_i |\ a_i = \overline{a}_i + \mathbf{\Delta}_i' u_i,\ ||u_i||_2 \le \rho\}$, $\mathbf{\Delta}_i :\ k_i \times n,\ u_i : k_i \times 1.$
- RC:

$$\begin{array}{rll} \max & c'x & \\ \text{s.t.} & \overline{a}_i'x + \rho||\mathbf{\Delta}_i x||_2 \le b_i, & i = 1, \ldots, m. \\ & x \ge \mathbf{0}. & \end{array}$$

Probabilistic Guarantee

- Suppose u_i are independent, have zero mean and have support in $[-1, 1]$.
- Suppose that x satisfies $\overline{a}'x + \rho||\mathbf{\Delta}x|| \leq b$.
- Then

$$P(\tilde{a}'x > b) \leq e^{-\rho^2/2}.$$

- **Remarkable property**: Independent of the distributions of u (we do not even require identical distributions).
- How to select ρ: Suppose our tolerance for infeasibility is ϵ, that is $P(\tilde{a}'x > b) \leq \epsilon$.
- Use $\epsilon = e^{-\frac{\rho^2}{2}}$, select $\rho = \sqrt{2\log\left(\frac{1}{\epsilon}\right)}$.

ϵ	ρ
10^{-6}	5.25
10^{-5}	4.79
10^{-4}	4.29
10^{-3}	3.71
10^{-2}	3.03
10^{-1}	2.14

Proof from First Principles

- Let $X(\boldsymbol{\xi}) = w_0 + \sum_{i=1}^{k} w_i \xi_i$, where ξ_i are independent with zero mean and with support in $[-1, 1]$.
- Let $w = (w_1, \ldots, w_k)'$. We will first show that

$$P(X(\boldsymbol{\xi}) > 0) = P\left(w_0 + \sum_{i=1}^{k} w_i \xi_i > 0\right) \leq exp\left(-\frac{w_0^2}{2||w||^2}\right).$$

$$P(X(\boldsymbol{\xi}) > 0) = \int \chi(X(\boldsymbol{\xi}))\, dP(\boldsymbol{\xi}), \quad \chi(s) = \begin{cases} 0, & s \leq 0, \\ 1, & s > 0 \end{cases}$$

- Note that $\chi(s) \leq \gamma(s) = e^s$.
- Let $\alpha > 0$. Note also that $\chi(s) = \chi(\alpha \cdot s) \leq \gamma(\alpha \cdot s)$.
-

$$P(X(\boldsymbol{\xi}) > 0) \leq E[exp(\alpha w_0 + \sum_{i=1}^{k} \alpha w_i \xi_i)] = exp(\alpha w_0) \prod_{i=1}^{k} E[exp(\alpha w_i \xi_i)].$$

Proof continued

- For every random variable ξ with zero mean and support in $[-1,1]$

$$E[e^{t\xi}] \leq e^{t^2/2}.$$

- Let $f(s) = e^{ts} - \frac{e^t - e^{-t}}{2} s$.
- $f(s)$ convex in s. Maximum in $[-1,1]$ is at endpoint.
- $\max_{|s|\leq 1} f(s) = f(1) = f(-1) = \frac{e^t + e^{-t}}{2}$.

$$\begin{aligned} E[e^{t\xi}] &= \int f(s)\, dP(s) \quad \text{[zero mean]} \\ &\leq \max_{|s|\leq 1} f(s) \\ &= \frac{e^t + e^{-t}}{2} \\ &\leq e^{t^2/2} \quad \text{[Taylor series].} \end{aligned}$$

Proof continued

- For all $\alpha > 0$:

$$\begin{aligned} P(X(\boldsymbol{\xi}) > 0) &\leq exp(\alpha w_0) \prod_{i=1}^{k} E[exp(\alpha w_i \xi_i)] \\ &\leq exp\left(\alpha w_0 + \frac{\alpha^2}{2} \sum_{i=1}^{k} w_i^2\right). \end{aligned}$$

- Select α to minimize the upper bound.
-

$$P(X(\boldsymbol{\xi}) > 0) \leq \min_{\alpha > 0} exp\left(\alpha w_0 + \frac{\alpha^2}{2} ||w||^2\right).$$

- $\alpha^* = -w_0/||w||^2$.
- $P(X(\boldsymbol{\xi}) > 0) = P\left(w_0 + \sum_{i=1}^{k} w_i \xi_i > 0\right) \leq exp\left(-\frac{w_0^2}{2||w||^2}\right)$.

Proof of the key guarantee

- Suppose that x satisfies $\overline{a}'x + \rho||\mathbf{\Delta}x|| \leq b$.
- Then

$$P(\tilde{a}'x > b) = P(\overline{a}'x + u'\mathbf{\Delta}x > b) \leq P(u'\mathbf{\Delta}x > \rho||\mathbf{\Delta}x||).$$

- Select $w_0 = -\rho||\mathbf{\Delta}x||$ and $w = \mathbf{\Delta}x$, we obtain

$$P(\tilde{a}'x > b) \leq e^{-\frac{\rho^2}{2}}.$$

Guarantees for non-independent distributions

- RO:

$$
\begin{aligned}
\max \quad & c'x \\
\text{s.t.} \quad & \tilde{A}x \le b \\
& x \in P \\
& \forall \tilde{A} \in \mathcal{U} = \left\{ \tilde{A} \mid ||M(\text{vec}(\tilde{A}) - \text{vec}(\overline{A}))|| \le \Delta \right\}.
\end{aligned}
$$

- RC:

$$
\begin{aligned}
\max \quad & c'x \\
\text{s.t.} \quad & \overline{a}_i x + \Delta ||M^{-1} x_i||^* \le b_i, \quad i = 1, \dots, m \\
& x \in P,
\end{aligned}
$$

- $\text{vec}(\tilde{A}) \sim (\text{vec}(\overline{A}), \boldsymbol{\Sigma})$.
- Let $M = \boldsymbol{\Sigma}^{-\frac{1}{2}}$.

Probabilistic Guarantees

- $$P\left(\tilde{a}_i' x^* > b_i\right) \leq \frac{1}{1+\Delta^2\left(\frac{||\mathbf{\Sigma}^{\frac{1}{2}} x_i^*||^*}{\|\mathbf{\Sigma}^{\frac{1}{2}} x_i^*\|_2}\right)^2}.$$
- If L_p norm used in $\mathcal{U}$, then
$$P\left(\tilde{a}_i' x^* > b_i\right) \leq \frac{1}{1+\Delta^2 \min\left\{\frac{1}{p^2}, \frac{1}{n}\right\}}.$$
- If L_2 used in $\mathcal{U}$, then
$$P\left(\tilde{a}_i' x^* > b_i\right) \leq \frac{1}{1+\Delta^2}.$$
- Remark: Arbitrary Dependence structure.
- How to select Δ?

Proof

- Optimal robust solution x_i^* satisfies

$$(\mathrm{vec}(\overline{A}))' x_i^* + \Delta ||\mathbf{\Sigma}^{\frac{1}{2}} x_i^*||^* \leq b_i,$$

- Thus

$$P\left((\mathrm{vec}(\tilde{A}))' x_i^* > b_i\right) \leq P\left((\mathrm{vec}(\tilde{A}))' x_i^* \geq (\mathrm{vec}(\overline{A}))' x_i^* + ||\mathbf{\Sigma}^{\frac{1}{2}} x_i^*||^*\right).$$

- Bertsimas and Popescu: if S is a convex set, and $\tilde{X} \sim (\overline{X}, \mathbf{\Sigma})$, then

$$P\left(\tilde{X} \in S\right) \leq \frac{1}{1+d^2},$$

where

$$d^2 = \inf_{\tilde{X} \in S} \left(\tilde{X} - \overline{X}\right)' \mathbf{\Sigma}^{-1} \left(\tilde{X} - \overline{X}\right).$$

Proof continued

- $S_i = \left\{ \text{vec}(\tilde{A}) \mid (\text{vec}(\tilde{A}))' x_i \geq (\text{vec}(\overline{A}))' x_i + \Delta ||\boldsymbol{\Sigma}^{\frac{1}{2}} x_i||^* \right\}.$
- $d_i^2 = \inf_{\text{vec}(\tilde{A}) \in S_i} \left(\text{vec}(\tilde{A}) - \text{vec}(\overline{A}) \right)' \boldsymbol{\Sigma}^{-1} \left(\text{vec}(\tilde{A}) - \text{vec}(\overline{A}) \right).$
- Optimal solution (KKT):

$$\text{vec}(\overline{A}) + \Delta \left(\frac{||\boldsymbol{\Sigma}^{\frac{1}{2}} x_i||^*}{||\boldsymbol{\Sigma}^{\frac{1}{2}} x_i||_2} \right)^2 \boldsymbol{\Sigma} x_i,$$

- $$d^2 = \Delta^2 \left(\frac{||\boldsymbol{\Sigma}^{\frac{1}{2}} x_i||^*}{||\boldsymbol{\Sigma}^{\frac{1}{2}} x_i||_2} \right)^2.$$

On the interplay of probability and optimization

- Use Probability theorems to select parameters.
- Use optimization ideas to find best possible results in probability.
- In exercise we will explore other bounds.
- Use RO to solve problems under uncertainty computationally.

15.094J: Robust Modeling, Optimization, Computation

Lecture 5: Robust Mixed Integer Optimization

February 2015

Outline

1. RMIO: Tractability
2. RMIO: Probabilistic Guarantees
3. Robust 0-1 Optimization
4. Robust Network Flows

Row-wise Polyhedral Uncertainty

- Primitives: Uncertainty sets $\mathcal{U}_i$, $i = 1, \ldots, m$, b, c (known, WLOG).
- RLO with row-wise uncertainty:

$$\begin{array}{rl} \max & c'x \\ \text{s.t.} & a_i'x \leq b_i. \quad \forall a_i \in \mathcal{U}_i,\ i = 1, \ldots, m, \\ & x \geq \mathbf{0},\ x_i \in \mathcal{Z},\ i = 1, \ldots, k. \end{array}$$

- $\mathcal{U}_i = \{a_i |\ D_i a_i \leq d_i\}$, $D_i :\ k_i \times n$.
- RC

$$\begin{array}{rll} \max\limits_{x, p_i} & c'x & \\ \text{s.t.} & p_i' d_i \leq b_i, & i = 1, \ldots, m, \\ & p_i' D_i = x', & i = 1, \ldots, m, \\ & p_i \geq \mathbf{0}, & i = 1, \ldots, m, \\ & x \geq \mathbf{0},\ x_i \in \mathcal{Z},\ i = 1, \ldots, k. & \end{array}$$

- RMIO reduces to MIO.
- Same even if uncertainty is not row-wise.

Row-wise Ellipsoidal uncertainty

- RO:

$$\begin{array}{rl} \max & c'x \\ \text{s.t.} & \max_{a_i \in \mathcal{U}_i} a_i'x \leq b_i. \\ & x \geq \mathbf{0},\ x_i \in \mathcal{Z},\ i = 1, \ldots, k. \end{array}$$

- $\mathcal{U}_i = \{a_i |\ a_i = \overline{a}_i + \mathbf{\Delta}_i' u_i,\ ||u_i||_2 \leq \rho\}$, $\mathbf{\Delta}_i :\ k_i \times n,\ u_i : k_i \times 1.$
- RC:

$$\begin{array}{rl} \max & c'x \\ \text{s.t.} & \overline{a}_i'x + \rho||\mathbf{\Delta}_i x||_2 \leq b_i, \qquad i = 1, \ldots, m. \\ & x \geq \mathbf{0},\ x_i \in \mathcal{Z},\ i = 1, \ldots, k. \end{array}$$

- RMIO reduces to Mixed Integer Second order cone problem.
- Current versions of CPLEX and Gurobi support it, but more difficult than MIO.

Row-wise Budget of Uncertainty

- $$\begin{aligned} \text{minimize} \quad & \tilde{c}'x \\ \text{subject to} \quad & \tilde{A}x \le b \\ & x \ge \mathbf{0},\ x_i \in \mathcal{Z},\ i = 1, \ldots, k. \end{aligned}$$
- **Uncertainty for matrix** A: $a_{ij},\ j \in J_i$ is independent, symmetric and bounded random variable (but with unknown distribution) $\tilde{a}_{ij}, j \in J_i$ that takes values in $[a_{ij} - \hat{a}_{ij}, a_{ij} + \hat{a}_{ij}]$.
- **Uncertainty for cost vector** c: $c_j,\ j \in J_0$ takes values in $[c_j, c_j + d_j]$.

Budget of Uncertainty

- Consider an integer $\Gamma_i \in [0, |J_i|]$, $i = 0, 1, \ldots, m$.
- Γ_i adjusts the robustness of the proposed method against the level of conservativeness of the solution.
- Unlikely that all of the a_{ij}, $j \in J_i$ will change. We want to be protected against all cases that up to Γ_i of the a_{ij}'s are allowed to change.
- Nature will be restricted in its behavior, in that only a subset of the coefficients will change in order to adversely affect the solution.
- We will guarantee that if nature behaves like this then the robust solution will be feasible deterministically. Even if more than Γ_i change, then the robust solution will be feasible with very high probability.

RMIO

$$
\begin{aligned}
RMIO: \quad & \text{minimize} \quad c'x + \max_{\{S_0 |\ S_0 \subseteq J_0, |S_0| \le \Gamma_0\}} \left\{ \sum_{j \in S_0} d_j |x_j| \right\} \\
& \text{subject to} \quad \sum_j a_{ij} x_j + \max_{\{S_i |\ S_i \subseteq J_i, |S_i| \le \Gamma_i\}} \left\{ \sum_{j \in S_i} \hat{a}_{ij} |x_j| \right\} \le b_i, \quad \forall i \\
& \qquad\qquad\quad x \ge \mathbf{0},\ x_i \in \mathcal{Z},\ i = 1, \ldots, k.
\end{aligned}
$$

$$
\begin{aligned}
RC: \quad & \text{minimize} \quad c'x + z_0 \Gamma_0 + \sum_{j \in J_0} p_{0j} \\
& \text{subject to} \quad \sum_j a_{ij} x_j + z_i \Gamma_i + \sum_{j \in J_i} p_{ij} \le b_i & \forall i \\
& \qquad z_0 + p_{0j} \ge d_j y_j & \forall j \in J_0 \\
& \qquad z_i + p_{ij} \ge \hat{a}_{ij} y_j & \forall i \ne 0, j \in J_i \\
& \qquad p_{ij}, y_j, z_i \ge 0 & \forall i, j \in J_i \\
& \qquad -y_j \le x_j \le y_j & \forall j \\
& \qquad x \ge \mathbf{0}, x_i \in \mathcal{Z},\ i = 1, \ldots, k.
\end{aligned}
$$

Proof

- Given a vector x^*, we define:

$$\beta_i(x^*) = \max_{\{S_i | \; S_i \subseteq J_i, |S_i| = \Gamma_i\}} \left\{ \sum_{j \in S_i} \hat{a}_{ij} |x_j^*| \right\}.$$

- This equals to:

$$\begin{aligned} \beta_i(x^*) = \max \quad & \sum_{j \in J_i} \hat{a}_{ij} |x_j^*| z_{ij} \\ \text{s.t.} \quad & \sum_{j \in J_i} z_{ij} \le \Gamma_i \\ & 0 \le z_{ij} \le 1 \qquad \forall i, j \in J_i. \end{aligned}$$

- Dual:

$$\begin{aligned} \beta_i(x^*) = \min \quad & \sum_{j \in J_i} p_{ij} + \Gamma_i z_i \\ \text{s.t.} \quad & z_i + p_{ij} \ge \hat{a}_{ij} |x_j^*| \quad && \forall j \in J_i \\ & p_{ij} \ge 0 && \forall j \in J_i \\ & z_i \ge 0 && \forall i. \end{aligned}$$

Size

- Original Problem has n variables and m constraints
- RC has $2n + m + l$ variables, where $l = \sum_{i=0}^{m} |J_i|$ is the number of uncertain coefficients, and $2n + m + l$ constraints.
- Sparsity is preserved, attractive!

Probabilistic Guarantees

- x^* an optimal solution of RMIO.
- $\tilde{a}_{ij}, j \in J_i$ independent, symmetric and bounded random variables, support $[a_{ij} - \hat{a}_{ij}, a_{ij} + \hat{a}_{ij}]$.
- $$\Pr\left(\sum_j \tilde{a}_{ij} x_j^* > b_i\right) \leq \frac{1}{2^n}\left\{(1-\mu)\sum_{l=\lfloor\nu\rfloor}^{n}\binom{n}{l} + \mu \sum_{l=\lfloor\nu\rfloor+1}^{n}\binom{n}{l}\right\},$$
  $n = |J_i|$, $\nu = \frac{\Gamma_i + n}{2}$ and $\mu = \nu - \lfloor\nu\rfloor$; bound is tight.
- As $n \to \infty$
  $$\frac{1}{2^n}\left\{(1-\mu)\sum_{l=\lfloor\nu\rfloor}^{n}\binom{n}{l} + \mu \sum_{l=\lfloor\nu\rfloor+1}^{n}\binom{n}{l}\right\} \sim 1 - \Phi\left(\frac{\Gamma_i - 1}{\sqrt{n}}\right).$$

$\lvert J_i \rvert$	Γ_i
5	5
10	8.3565
100	24.263
200	33.899

Experimental Results

- Knapsack Problem

$$\begin{aligned} \text{maximize} \quad & \sum_{i \in N} c_i x_i \\ \text{subject to} \quad & \sum_{i \in N} w_i x_i \leq b \\ & x \in \{0,1\}^n. \end{aligned}$$

- $\tilde{w}_i$ independently distributed and follow symmetric distributions in $[w_i - \delta_i, w_i + \delta_i]$;
- c is not subject to data uncertainty.
- $|N| = 200$, $b = 4000$,
- w_i randomly chosen from $\{20, 21, \ldots, 29\}$.
- c_i randomly chosen from $\{16, 17, \ldots, 77\}$.
- $\delta_i = 0.1 w_i$.

Experimental Results. continued

Γ	Violation Probability	Optimal Value	Reduction
0	0.5	5592	0%
2.8	0.449	5585	0.13%
36.8	5.71×10^{-3}	5506	1.54%
82.0	5.04×10^{-9}	5408	3.29%
200	0	5283	5.50%

Robust 0-1 Optimization

- Nominal 0-1 optimization:

$$\begin{array}{ll} \text{minimize} & c'x \\ \text{subject to} & x \in X \subset \{0,1\}^n. \end{array}$$

- Reformulation:

$$\begin{array}{rll} Z^* = & \text{minimize} & c'x + \max\limits_{\{S|\ S \subseteq J, |S| \leq \Gamma\}} \sum\limits_{j \in S} d_j x_j \\ & \text{subject to} & x \in X, \end{array}$$

Contrast

- Other approaches to robustness are hard. Scenario based uncertainty:

$$\begin{aligned} \text{minimize} \quad & \max(c_1'x, c_2'x) \\ \text{subject to} \quad & x \in X. \end{aligned}$$

  is NP-hard for the shortest path problem.
- $d_1 \geq d_2 \geq \ldots \geq d_n$. Optimal robust solution is

$$Z^* = \min_{l=1,\ldots,n+1} d_l\Gamma + \min_{x \in X} \sum_{j=1}^{n} c_j x_j + \sum_{j=1}^{l} (d_j - d_l) x_j.$$

- Thus, if nominal problem is polynomially solvable the robust problem is also.

Proof

$$\text{Primal:} \quad Z^* = \min_{x \in X} c'x + \max \quad \sum_j d_j x_j u_j$$
$$\text{s.t.} \quad 0 \le u_j \le 1, \qquad \forall\, j$$
$$\sum_j u_j \le \Gamma$$

$$\text{Dual:} \quad Z^* = \min_{x \in X} c'x + \min \quad \theta\Gamma + \sum_j y_j$$
$$\text{s.t.} \quad y_j + \theta \ge d_j x_j, \qquad \forall\, j$$
$$y_j, \theta \ge 0$$

Proof, continued

- Solution: $y_j = \max(d_j x_j - \theta, 0)$
- $$Z^* = \min_{x \in X, \theta \geq 0} \theta\Gamma + \sum_j (c_j x_j + \max(d_j x_j - \theta, 0))$$
- Since $X \subset \{0,1\}^n$,
$$\max(d_j x_j - \theta, 0) = \max(d_j - \theta, 0)\ x_j$$
- $$Z^* = \min_{x \in X, \theta \geq 0} \theta\Gamma + \sum_j (c_j + \max(d_j - \theta, 0))\, x_j$$

Proof, continued

- $d_1 \geq d_2 \geq \ldots \geq d_n \geq d_{n+1} = 0$.
- For $d_l \geq \theta \geq d_{l+1}$,

$$\min_{x \in X, d_l \geq \theta \geq d_{l+1}} \theta\Gamma + \sum_{j=1}^{n} c_j x_j + \sum_{j=1}^{l} (d_j - \theta) x_j =$$

$$d_l\Gamma + \min_{x \in X} \sum_{j=1}^{n} c_j x_j + \sum_{j=1}^{l} (d_j - d_l) x_j = Z_l$$

- $$Z^* = \min_{l=1,\ldots,n+1} d_l\Gamma + \min_{x \in X} \sum_{j=1}^{n} c_j x_j + \sum_{j=1}^{l} (d_j - d_l) x_j.$$

Algorithm A

- **Input:** Vectors $c, d \in \Re^n_+$, an integer Γ, and a polynomial time algorithm that solves the problem $Z = \min c'x$ subject to $x \in X \subseteq \{0,1\}^n$ for all $c \geq \mathbf{0}$.
- **Output:** A solution $x^* \in X$ such that $x^* = \text{argmin}\left(c'x + \max_{\{S|\ S\subseteq J, |S|=\Gamma\}} \sum_{j\in S} d_j x_j\right)$.

Algorithm A, continued

$x^1 \leftarrow \arg\min\{c'x : x \in X\}$

FOR $l \in 2, \ldots, r$

    IF $d_l < d_{l-1}$

$$x^l \leftarrow \arg\min\{c'x + \sum_{j=1}^{l}(d_j - d_l)x_j : x \in X\}$$

$$Z_l \leftarrow c'x^l + \max_{\{S|\ S \subseteq J, |S| = \Gamma\}} \sum_{j \in S} d_j x_j^l$$

    ELSE

$$x^l \leftarrow x^{l-1}$$

$$Z_l \leftarrow Z_{l-1}$$

    END IF

END FOR

Algorithm A, continued

$$
\begin{aligned}
11 \quad &: \quad x^{r+1} \leftarrow \arg\min\{c'x + \sum_{j \in J} d_j x_j : x \in X\} \\
12 \quad &: \quad Z_{r+1} \leftarrow c'x^{r+1} + \max_{\{S|\ S \subseteq J, |S| = \Gamma\}} \sum_{j \in S} d_j x_j^{r+1} \\
13 \quad &: \quad \pi \leftarrow \arg\min\{Z_j : j \in J \cup \{r+1\}\} \\
14 \quad &: \quad Z^* = Z_\pi;\ x^* = x^\pi.
\end{aligned}
$$

Theorem

- Algorithm A correctly solves the robust 0-1 optimization problem.
- It requires at most $|J| + 1$ solutions of nominal problems. Thus, if the nominal problem is polynomially time solvable, then the robust 0-1 counterpart is also polynomially solvable.
- Robust minimum spanning tree, minimum assignment, minimum matching, shortest path and matroid intersection, are polynomially solvable.

Robust Approximation Algorithms

- If the nominal problem is α-approximable, is the robust counterpart also α-approximable?
- Use an α-approximate solution to

$$\min_{x \in X} \sum_{j=1}^{n} c_j x_j + \sum_{j=1}^{l} (d_j - d_l) x_j.$$

- Theorem: Overall algorithm is α-approximate.

Ellipsoidal Uncertainty

- $$\min_{x \in X} c'x + \max_{\tilde{s} \in D} \tilde{s}'x$$
- $D = \{s : \|\Sigma^{-1/2} s\|_2 \leq \Omega\}$
- Equivalent to:

$$\min_{x \in X} c'x + \Omega\sqrt{x'\Sigma x}$$

  Σ is the covariance matrix of the random cost coefficients: NP-hard
- D a polyhedron: NP-hard.

Uncorrelated uncertainty

- For $\boldsymbol{\Sigma} = diag(d_1^2, \ldots, d_n^2)$,

$$Z^* = \min_{x \in X} c'x + \Omega\sqrt{d'x}$$

  Complexity Open.

- Theorem: For $d_1 = \ldots = d_n = \sigma$,

$$Z^* = \min_{w=0,1,\ldots,n} Z(w),$$

$$Z(w) = \begin{cases} \min_{x \in X} & \left(c + \dfrac{\Omega\sigma}{2\sqrt{w}} e\right)' x + \dfrac{\Omega\sigma\sqrt{w}}{2} & w = 1, \ldots, n \\ \min_{x \in X} & (c + \Omega\sigma e)'x & w = 0. \end{cases}$$

Practical algorithm

- Until $\|x^{k+1} - x^k\| \leq \epsilon$, set $x^{k+1} := \arg\min_{y \in X} (c + \frac{\Omega}{2\sqrt{d' x^k}} d)' y$
- Output x^{k+1}
- Experimented on Shortest Path Problems, Uniform Matroid and Knapsack Problems, under randomly generated cost vectors in dimensions from 200 to 20,000.
- In 998 out of 1000 instances, optimal solution is found in solving less than 6 nominal problems!

Robust Network Flows

- Nominal

$$
\begin{aligned}
\min \quad & \sum_{(i,j)\in\mathcal{A}} c_{ij}x_{ij} \\
\text{s.t.} \quad & \sum_{\{j:(i,j)\in\mathcal{A}\}} x_{ij} - \sum_{\{j:(j,i)\in\mathcal{A}\}} x_{ji} = b_i \quad && \forall i \in \mathcal{N} \\
& 0 \le x_{ij} \le u_{ij} && \forall (i,j) \in \mathcal{A}.
\end{aligned}
$$

- X set of feasible solution flows.
- Robust

$$
\begin{aligned}
Z^* = \quad \min \quad & c'x + \max_{\{S|\ S\subseteq\mathcal{A},|S|\le\Gamma\}} \sum_{(i,j)\in S} d_{ij}x_{ij} \\
\text{s.t.} \quad & x \in X.
\end{aligned}
$$

Theorem

For any fixed $\Gamma \leq |\mathcal{A}|$ and every $\epsilon > 0$, we can find a solution $\hat{x} \in X$:

$$\hat{Z} = c'\hat{x} + \max_{\{S|\ S \subseteq \mathcal{A}, |S| \leq \Gamma\}} \sum_{(i,j) \in S} d_{ij}\hat{x}_{ij}$$

such that

$$Z^* \leq \hat{Z} \leq (1+\epsilon)Z^*$$

by solving $2\lceil \log_2(|\mathcal{A}|\overline{\theta}/\epsilon)\rceil + 3$ network flow problems, where $\overline{\theta} = \max\{u_{ij}d_{ij}\ : (i,j) \in \mathcal{A}\}$.

15.094J: Robust Modeling, Optimization, Computation

Lecture 6: Robust Convex Optimization

February 2015

Outline

1. Motivation
2. Robust Conic Optimization
3. Exact and Relaxed Robustness
4. Tractability
5. Probabilistic Guarantees
6. Conclusions

Motivation

- In earlier proposals (Ben-Tal and Nemirovski):
  - **(a)** RLOs become SOCPs
  - **(b)** Robust SOCPs become Semi-definite optimization problems (SDPs)
  - **(c)** Robust SDPs become NP-hard.
- In Contrast
  - **(a)** In Lecture 4, we have shown that RLO becomes LO.
  - **(b)** Today we show that Robust SOCPs stay SOCPs
  - **(c)** and Robust SDPs stay SDPs.
- RC inherits the complexity of the underlying deterministic problem.
- RC allows the user to control the tradeoff between robustness and optimality.
- RC is computationally tractable both practically and theoretically.

Nominal vs Robust

- Nominal

$$\begin{aligned} \max \quad & f_0(x, \tilde{D}_0) \\ \text{s.t.} \quad & f_i(x, \tilde{D}_i) \geq 0, \qquad i \in I \\ & x \in X \end{aligned}$$

- Exact Robust

$$\begin{aligned} \max \quad & \min_{D_\mathbf{0} \in \mathcal{U}_\mathbf{0}} f_0(x, D_0) \\ \text{s.t.} \quad & \min_{D_i \in \mathcal{U}_i} f_i(x, D_i) \geq 0, i \in I \\ & x \in X \end{aligned} \tag{1}$$

Uncertainty

- Data uncertainty

$$\tilde{D} = D^0 + \sum_{j \in N} \mathbf{\Delta} D^j \tilde{z}_j$$

- Uncertainty sets

$$\mathcal{U} = \left\{ D \mid \exists u \in \Re^{|N|} : D = D^0 + \sum_{j \in N} \mathbf{\Delta} D^j u_j, \|u\| \le \rho \right\}$$

Modeling power

Type	Constraint	D	$f(x, D)$
LO	$a'x \geq b$	(a, b)	$a'x - b$
QCQO	$\|Ax\|_2^2 + b'x + c \leq 0$	(A, b, c, d) $d^0 = 1/2, \Delta d^j = 0$	$\frac{d-(b'x+c)}{2}$ $-\sqrt{\|Ax\|_2^2 + \left(\frac{d+b'x+c}{2}\right)^2}$
SOCO(1)	$\|Ax + b\|_2 \leq c'x + d$	(A, b, c, d) $\boldsymbol{\Delta} c^j = \mathbf{0}, \Delta d^j = 0$	$c'x + d - \|Ax + b\|_2$
SOCO(2)	$\|Ax + b\|_2 \leq c'x + d$	(A, b, c, d)	$c'x + d - \|Ax + b\|_2$
SDO	$\sum_{j=1}^n A_i x_i - B \in S_+^m$	$(A_1, \ldots, A_n, B)$	$\lambda_{min}(\sum_{j=1}^n A_i x_i - B)$

Exact and Relaxed Robustness

- Exact Robustness (ER)

$$f\left(x, D^0 + \sum_{j\in N} \boldsymbol{\Delta} D^j u_j\right) \geq 0 \qquad \forall \|u\| \leq \rho.$$

- Relaxed Robustness (RR)

$$f(x, D^0) + \sum_{j\in N} \left\{ f(x, \boldsymbol{\Delta} D^j) v_j + f(x, -\boldsymbol{\Delta} D^j) w_j \right\} \geq 0$$

$$\forall (v, w) \in \Re_+^{|N|\times|N|} \; \|v + w\| \leq \rho.$$

Theorem

- Assumption 1: Norms satisfy $\|u\| = \|u^+\|$, $u_j^+ = |u_j|$. Examples L_p-norms.
- Assumption 2: f satisfies: $f(x, D)$ is concave in D for all $x \in \Re^n$, $f(x, kD) = kf(x, D)$, for all $k \geq 0$, D, $x \in \Re^n$,

(a) Under Assumption 1 and $f(x, A + B) = f(x, A) + f(x, B)$, ER and RR are equivalent.

(b) Under Assumptions 1 and 2, if x^* satisfies RR, it satisfies ER also.

Proof of part (a)

- Under linearity, RR becomes

$$f\left(x, D^0 + \sum_{j \in N} \mathbf{\Delta} D^j (v_j - w_j)\right) \geq 0 \qquad \forall \|v + w\| \leq \rho, \quad v, w \geq \mathbf{0},$$

- ER becomes

$$f\left(x, D^0 + \sum_{j \in N} \mathbf{\Delta} D^j r_j\right) \geq 0 \qquad \forall \|r\| \leq \rho.$$

- If x violates ER, there exists $r, \|r\| \leq \rho$ such that

$$f\left(x, D^0 + \sum_{j \in N} \mathbf{\Delta} D^j r_j\right) < 0.$$

- Let $v_j = \max\{r_j, 0\}$ and $w_j = -\min\{r_j, 0\}$.
- Clearly, $r = v - w$ and since $v_j + w_j = |r_j|$, $\|v + w\| = \|r\| \leq \rho$.
- x violates RR.

Proof of part (a), continued

- If x violates RR, then there exist $v, w \geq \mathbf{0}$ and $\|v + w\| \leq \rho$ such that

$$f\left(x, D^0 + \sum_{j \in N} \Delta D^j (v_j - w_j)\right) < 0.$$

- Let $r_j = v_j - w_j$ and we observe that $|r_j| \leq v_j + w_j$.
- For norms satisfying $\|u\| = \|u^+\|$, $u_j^+ = |u_j|$,

$$\|r\| = \|r^+\| \leq \|v + w\| \leq \rho,$$

and hence, x violates ER.

Proof of part (b)

- If x satisfies RR

$$f(x, D^0)+\sum_{j\in N}\left\{f(x, \boldsymbol{\Delta} D^j)v_j + f(x, -\boldsymbol{\Delta} D^j)w_j\right\} \geq 0,\ \forall\|v+w\| \leq \rho,\ \ v, w \geq \mathbf{0}.$$

- From concavity and homogeneity

$$f(x, A+B) \geq \frac{1}{2}f(x, 2A) + \frac{1}{2}f(x, 2B) = f(x, A) + f(x, B).$$

- Then

$$0 \leq f(x, D^0) + \sum_{j\in N}\left\{f(x, \boldsymbol{\Delta} D^j)v_j + f(x, -\boldsymbol{\Delta} D^j)w_j\right\} \leq$$

$$f(x, D^0 + \sum_{j\in N}\boldsymbol{\Delta} D^j(v_j - w_j))$$

for all $\|v + w\| \leq \rho, \quad v, w \geq \mathbf{0}$.

Proof of part (b), continued

- In part (a) we established that

$$f(x, D^0 + \sum_{j \in N} \mathbf{\Delta} D^j r_j) \geq 0 \qquad \forall \|r\| \leq \rho$$

is equivalent to

$$f(x, D^0 + \sum_{j \in N} \mathbf{\Delta} D^j (v_j - w_j)) \geq 0 \qquad \forall \|v + w\| \leq \rho, \quad v, w \geq \mathbf{0},$$

and thus x satisfies ER.

Tractability

RR is equivalent to

$$
\begin{aligned}
& f(x, D^0) \geq \rho y \\
& f(x, \mathbf{\Delta} D^j) + t_j \geq 0 \qquad \forall j \in N \\
& f(x, -\mathbf{\Delta} D^j) + t_j \geq 0 \quad \forall j \in N \\
& \|t\|^* \leq y \\
& y \in \Re,\ \ t \in \Re^{|N|}.
\end{aligned}
$$

Dual norm: $\|s\|^* = \max_{\|x\| \leq 1} s'x$.

Tractability, continued

(a) Under Assumptions 1 and 2, RR is equivalent to RR'

$$f(x, D^0) \geq \rho \|s\|^*,$$

where

$$s_j = \max\{-f(x, \mathbf{\Delta} D^j), -f(x, -\mathbf{\Delta} D^j)\}, \quad \forall j \in N.$$

(b) $f(x, D^0) \geq \rho \|s\|^*$, can be written as RR":

$$\begin{aligned} & f(x, D^0) \geq \rho y \\ & f(x, \mathbf{\Delta} D^j) + t_j \geq 0 \qquad \forall j \in N \\ & f(x, -\mathbf{\Delta} D^j) + t_j \geq 0 \quad \forall j \in N \\ & \|t\|^* \leq y \\ & y \in \Re, \;\; t \in \Re^{|N|}. \end{aligned}$$

Proof, part (a)

- We introduce the following problems:

$$\begin{aligned} z_1 = \max \quad & a'v + b'w \\ \text{s.t.} \quad & \|v + w\| \leq \rho \\ & v, w \geq \mathbf{0}, \end{aligned}$$

$$\begin{aligned} z_2 = \max \quad & \sum_{j \in N} \max\{a_j, b_j, 0\} r_j \\ \text{s.t.} \quad & \|r\| \leq \rho, \end{aligned}$$

  and show that $z_1 = z_2$.

- Suppose r^* is an optimal solution to z_2. For all $j \in N$, let

$$\begin{aligned} v_j = w_j = 0 \qquad & \text{if } \max\{a_j, b_j\} \leq 0 \\ v_j = |r_j^*|, w_j = 0 \quad & \text{if } a_j \geq b_j, a_j > 0 \\ w_j = |r_j^*|, v_j = 0 \quad & \text{if } b_j > a_j, b_j > 0. \end{aligned}$$

Proof part (a), continued

- Observe that $a_j v_j + b_j w_j \geq \max\{a_j, b_j, 0\} r_j^*$ and $w_j + v_j \leq |r_j^*|\ \forall j \in N$.
- If $v^+ \leq w^+$, $\|v\| \leq \|w\|$.
- Then $\|v + w\| \leq \|r^*\| \leq \rho$, and thus v, w are feasible in z_1 leading to

$$z_1 \geq \sum_{j \in N} (a_j v_j + b_j w_j) \geq \sum_{j \in N} \max\{a_j, b_j, 0\} r_j^* = z_2.$$

- Conversely, let v^*, w^* be an optimal solution to z_1.
- Let $r = v^* + w^*$. Clearly $\|r\| \leq \rho$ and observe that

$$r_j \max\{a_j, b_j, 0\} \geq a_j v_j^* + b_j w_j^* \quad \forall j \in N.$$

- Therefore, we have

$$z_2 \geq \sum_{j \in N} \max\{a_j, b_j, 0\} r_j \geq \sum_{j \in N} (a_j v_j^* + b_j w_j^*) = z_1,$$

leading to $z_1 = z_2$.

Proof part (a), continued

- $\mathcal{V} = \{(v, w) \in \Re_+^{|N| \times |N|} \ \|v + w\| \leq \rho\}$.
- Then,

$$
\begin{aligned}
& \min_{(v,w)\in\mathcal{V}} \sum_{j\in N} \left\{ f(x, \boldsymbol{\Delta} D^j) v_j + f(x, -\boldsymbol{\Delta} D^j) w_j \right\} \\
= & \ - \max_{(v,w)\in\mathcal{V}} \sum_{j\in N} \left\{ -f(x, \boldsymbol{\Delta} D^j) v_j - f(x, -\boldsymbol{\Delta} D^j) w_j \right\} \\
= & \ - \max_{\{\|r\|\leq\rho\}} \sum_{j\in N} \left\{ \max\{-f(x, \boldsymbol{\Delta} D^j), -f(x, -\boldsymbol{\Delta} D^j), 0\} r_j \right\}
\end{aligned}
$$

- Since $\|s\|^* = \max_{\|x\|\leq 1} s'x$, we obtain $\rho\|s\|^* = \max_{\|x\|\leq\rho} s'x$, i.e., RR' follows.
- Note that $s_j = \max\{-f(x, \boldsymbol{\Delta} D^j), -f(x, -\boldsymbol{\Delta} D^j)\} \geq 0$, since otherwise there exists an x such that $s_j < 0$, i.e., $f(x, \boldsymbol{\Delta} D^j) > 0$ and $f(x, -\boldsymbol{\Delta} D^j) > 0$. From Assumption 2 $f(x, \mathbf{0}) = 0$, contradicting the concavity of $f(x, D)$.

Proof, part (b)

- Suppose that x is feasible in RR'.
- Let $t = s$ and $y = \|s\|^*$,
- We can easily check that (x, t, y) are feasible in RR".
- Conversely, suppose, x is infeasible in RR', that is,

$$f(x, D^0) < \rho\|s\|^*.$$

- Since, $t_j \geq s_j = \max\{-f(x, \mathbf{\Delta} D^j), -f(x, -\mathbf{\Delta} D^j)\} \geq 0$
- We have $v^+ \leq w^+$, $\|v\|^* \leq \|w\|^*$.
- Thus, $\|t\|^* \geq \|s\|^*$, leading to

$$f(x, D^0) < \rho\|s\|^* \leq \rho\|t\|^* \leq \rho y,$$

i.e., x is infeasible in RR".

Dual norm

Norms	$\|u\|$	$\|t\|^* \leq y$
L_2	$\|u\|_2$	$\|t\|_2 \leq y$
L_1	$\|u\|_1$	$t_j \leq y, \forall j \in N$
L_∞	$\|u\|_\infty$	$\sum_{j\in N} t_j \leq y$
L_p	$\|u\|_p$	$\left(\sum_{j\in N} t_j^{\frac{q}{q-1}}\right)^{\frac{q-1}{q}} \leq y$
$L_2 \cap L_\infty$	$\max\{\|u\|_2, \rho\|u\|_\infty\}$	$\|s-t\|_2 +$ $\frac{1}{\rho}\sum_{j\in N} s_j \leq y,\ s \in \Re_+^{\|N\|}$
$L_1 \cap L_\infty$	$\max\{\frac{1}{\Gamma}\|u\|_1, \|u\|_\infty\}$	$\Gamma p + \sum_{j\in N} s_j \leq y$ $s_j + p \geq t_j,\ p \in \Re_+, s \in \Re_+^{\|N\|}$

Size

- Independent Pertubations
- Example

$$\begin{pmatrix} a_1 & a_2 \\ a_2 & a_3 \end{pmatrix} x_1 + \begin{pmatrix} a_4 & a_5 \\ a_5 & a_6 \end{pmatrix} x_2 \succeq \begin{pmatrix} a_7 & a_8 \\ a_8 & a_9 \end{pmatrix},$$

$\tilde{a}_i = a_i^0 + \Delta a_i \tilde{z}_i$.

- $f(x, \mathbf{\Delta} d^1) + t_1 \geq 0$ becomes

$$\lambda_{min}\left(\begin{pmatrix} \Delta a_1 & 0 \\ 0 & 0 \end{pmatrix} x_1 + \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} x_2 - \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}\right) + t_1 \geq 0,$$

as $t_1 \geq -\min\{\Delta a_1 x_1, 0\}$ or equivalently as linear constraints $t_1 \geq -\Delta a_1 x_1, t_1 \geq 0$.

Tractability

	L_∞	L_1	L_2	$L_2 \cap L_\infty$
Num. Vars.	$n+1$	1	1	$2\|N\|+1$
Num. linear Const.	$2n+1$	$2n+1$	0	$3\|N\|$
Num SOC Const.	0	0	1	1
LO	LO	LO	SOCO	SOCO
QCQO	SOCO	SOCO	SOCO	SOCO
SOCO(1)	SOCO	SOCO	SOCO	SOCO
SOCO(2)	SOCO	SOCO	SOCO	SOCO
SDO	SDO	SDO	SDO	SDO

Probabilistic Guarantees

If $\tilde{z} \sim \mathcal{N}(0, I)$, under the L_2 norm:

$$\mathrm{P}(f(x, \tilde{D}) < 0) \leq \frac{\sqrt{e}\rho}{\alpha} e^{\left(-\frac{\rho^2}{2\alpha^2}\right)}$$

Problem	α	ρ
LO	1	$O(\log(1/\epsilon))$
SOCO(1)	1	$O(\log(1/\epsilon))$
SOCO(2)	$\sqrt{2}$	$O(\log(1/\epsilon))$
QCQO	$\sqrt{2}$	$O(\log(1/\epsilon))$
SDO	$\sqrt{\log m}$	$O(\sqrt{\log m}\log(1/\epsilon))$

Conclusions

- Given a conic optimization problem, we proposed a robust counterpart of the same character as original, thus preserving computational tractability.
- Size of the proposed problem is very similar to original; depends on the norm we use; best results for L_2 norm.
- Probabilistic guarantee allows to select parameter controlling robustness and optimality.

15.094J: Robust Modeling, Optimization, Computation

Lecture 9: Adaptive Optimization

Outline

Central Problems of OR

George Dantzig: "Planning under uncertainty. This, I feel, is the real field that we should be all working in."

Problem	Current Theory	Proposal
Modelling under Uncertainty	Probability Theory	RO
Optimization under Uncertainty	DP	RO
Optimization over Time	DP	AO

DP: Dynamic Programming
RO: Robust Optimization
AO: Adaptive Optimization

Optimization over time and under uncertainty

- The current method proposed by Richard Bellman in 1953, and taught in first year courses around the world, consists of two ideas:
- Describe uncertainty using probability distributions.
- To decide what to do today, have a plan for every eventuality in the future.

Criticisms of current approach

- Probability distributions do not exist in practice, and stochastic models are by and large computationally intractable.
- How do humans take decisions?
- For example: In some of the most important decisions in life (to whom to marry, which career to follow, etc.) do you enumerate every eventuality?
- Moreover, DP in most cases is computationally intractable in dimensions 3 or higher.

An example of DP

- An Apple store needs to decide the ordering mechanism for iPhones 5.
- You need to decide **today** how many iPhones 5 u_1 to order.
- You also need to decide how many iPhones 5 u_t to order **at time** t.
- There is demand uncertainty d_t, we assume that *the probability distribution of d_t is known*.
- There are cost of ordering c_t, and costs $f(x_t)$ for keeping inventory x_t at time t. For example, $f(x_t) = h\max(x_t, 0) + p\max(-x_t, 0)$.
- TIme horizon is T, and salvage value at time T is s.

Solution method

- State: Inventory x_t, $t = 1, \dots, T$.
- Decision: Order u_t, $t = 1, \dots, T$.
- Uncertainty: Demand d_t, $t = 1, \dots, T$.
- Dynamics: $x_{t+1} = x_t + u_t - d_t$, x_1 known.
- Objective: $\min \sum_{t=1}^{T} (f(x_t) + c_t u_t)$.
- Bellman recursion:

$$J_T(x_T) = s \cdot x_T.$$

$$J_t(x_t) = \min_{u_t} E_{d_t}[c_t u_t + f(x_t) + J_{t+1}(x_t + u_t - d_t)], \quad t = T-1, \dots, 1.$$

- Key observation: In deciding what to do now u_1, we need to decide $u_t(x_t)$ for every inventory level in the future.
- For the Galleria store, $x_t \in \{0, \dots, 10,000\}$ and $T = 360$ (one year). So we need to calculate 3.6 million decisions to decide how many iPhones 5 to order today, $u_1(x_1)$.

Reflections

- Imagine also certain known unknowns: a) Samsung wins the appeal for patent infringement, b) the world enters a deeper recession.
- Imagine also certain unknown unknowns: A new Steve Jobs has been working in secrecy on a brand new technology on a voice recognition system, much superior to Siri, Google launches a brand new product that makes iPhones irrelevant, etc.
- Should we enumerate 3.6 million decisions?
- And what happens when instead of only the Galleria store we need to coordinate all Apple stores in New England that are served by the same distribution center?
- Then we need to enumerate of the order of 3.6^{10} million states in order to decide what to order from all the stores.

Adaptive Optimization

- Two time periods
- Data: c, d, A, uncertainty set $\mathcal{U}$.
- Timing: *Here and now decisions* x
- Then uncertainty B is observed.
- Finally: *Wait and see decisions* $y(B)$ are applied.

$$
\begin{aligned}
Z_{AO} = \max_x \quad & c'x + \min_{B\in\mathcal{U}} \max_{y(B)} d'y(B) \\
\text{s.t.} \quad & Ax + By(B) \le b, \qquad \forall B \in \mathcal{U} \\
& x, y(B) \ge \mathbf{0}
\end{aligned}
$$

- It is adaptive optimization as the $y(B)$ adapts to the data.
- We avoid the difficulty of probability theory, but it is still computationally intractable as it still calls for a plan for all eventualities B.

Robust Optimization

- Consider $y(B) = y$ for all B.
- Then we obtain RO

$$
\begin{aligned}
Z_{RO} = \max_x \quad & c'x + \min_{B \in \mathcal{U}} \max_y d'y \\
\text{s.t.} \quad & Ax + By \leq b, \qquad \forall B \in \mathcal{U} \\
& x, y \geq \mathbf{0}
\end{aligned}
$$

- In deciding x, we create a plan for the future y for all uncertainties B.
- RO computationally tractable if $\mathcal{U}$ is computationally tractable.

Finitely Adaptive Optimization

- Partition the uncertainty set in k convex subsets (cutting by hyperplanes for example) $\mathcal{U}_1, \ldots, \mathcal{U}_k$.
- Set

$$y(B) = \begin{cases} y_1, & \text{if } B \in \mathcal{U}_1 \\ y_2, & \text{if } B \in \mathcal{U}_2 \\ \vdots & \vdots \\ y_k, & \text{if } B \in \mathcal{U}_k \end{cases}$$

- Intuition: Aggregate uncertainty and find an aggregate adaptive plan imitating human thinking and planning.
- FAO (FAO=RO, if $k = 1$):

$$\begin{aligned} Z_{FAO} = \max_x \quad & c'x + \min_{i=1,\ldots,k} \max_{y_i \in \mathcal{U}_i} d'y_i \\ \text{s.t.} \quad & Ax + B_i y_i \leq b, \quad \forall B_i \in \mathcal{U}_i \\ & x, y_i \geq \mathbf{0}, \ i = 1, \ldots, k. \end{aligned}$$

- In deciding x, we create a few plans y_i if the uncertainty is in set $\mathcal{U}_i$.
- FAO computationally tractable if $\mathcal{U}_i$ are computationally tractable, like RO.
- Readily extends under MIO conditions.

Affinely Adaptive Optimization

- Set $y(B) = q + P\zeta$ where $\zeta = \mathrm{vec}(B)$.
- Let

$$\begin{aligned} Z_{AAO} = \max_{x} \quad & c'x + \min_{B\in\mathcal{U}} \max_{P,q} d'(q + P\zeta) \\ \text{s.t.} \quad & Ax + B(q + P\zeta) \leq b, \qquad \forall B \in \mathcal{U} \\ & q + P\zeta \geq \mathbf{0}, \qquad \forall B \in \mathcal{U} \\ & x \geq \mathbf{0}, \end{aligned}$$

- AAO computationally tractable if $\mathcal{U}$ is computationally tractable and we restrict P to semidefinite matrices, so that the constraint is convex.

LO Formulation of AAO

- AAO under right hand side uncertainty.
- Consider the two stage AO problem

$$
\begin{aligned}
\max_{\mathbf{x}} \quad & \mathbf{c}'\mathbf{x} + \min_{\mathbf{b}\in\mathcal{U}} \max_{\mathbf{y}(\mathbf{b})} \mathbf{d}'\mathbf{y}(\mathbf{b}) \\
\text{s.t.} \quad & \mathbf{Ax} + \mathbf{By}\left(\mathbf{b}\right) \leq \mathbf{b}, \quad \forall \mathbf{b} \in \mathcal{U}, \\
& \mathbf{y}\left(\mathbf{b}\right) \geq \mathbf{0}, \quad \forall \mathbf{b} \in \mathcal{U}, \\
& \mathbf{x} \geq \mathbf{0},
\end{aligned}
$$

where $\mathcal{U} = \{\mathbf{b} \,|\mathbf{Gb} \leq \mathbf{f}\,\}$.

- Suppose we restrict ourselves to recourse functions that are affine, that is,

$$
\mathbf{y}\left(\mathbf{b}\right) = \mathbf{Pb} + \mathbf{q}.
$$

LO Formulation of AAO, continued

- By substituting $\mathbf{y}(\mathbf{b}) = \mathbf{P}\mathbf{b} + \mathbf{q}$, we obtain the following static RO

$$
\begin{aligned}
\max_{\mathbf{x}} \quad & \mathbf{c}'\mathbf{x} + \min_{\mathbf{b}\in\mathcal{U}} \max_{\mathbf{P},\mathbf{q}} \mathbf{d}'(\mathbf{P}\mathbf{b} + \mathbf{q}) \\
\text{s.t.} \quad & \mathbf{A}\mathbf{x} + \mathbf{B}(\mathbf{P}\mathbf{b} + \mathbf{q}) \le \mathbf{b}, \quad \forall \mathbf{b} \in \mathcal{U}, \\
& \mathbf{P}\mathbf{b} + \mathbf{q} \ge 0, \quad \forall \mathbf{b} \in \mathcal{U}, \\
& \mathbf{x} \ge \mathbf{0},
\end{aligned}
$$

- By using the techniques introduced in previous lectures, this problem can be reformulated into a single linear optimization problem.
- In particular, let the matrices W, V be the dual variables introduced to model the constraints $\mathbf{A}\mathbf{y} + \mathbf{B}(\mathbf{P}\mathbf{b} + \mathbf{q}) \le \mathbf{b}$, $\mathbf{P}\mathbf{b} + \mathbf{q} \ge 0$, $\forall b \in \mathcal{U}$ respectively.

LO Formulation of AAO, continued

- The single linear optimization formulation is given by

$$
\begin{array}{ll}
\min\limits_{\mathbf{y},\mathbf{F},\mathbf{q},\eta,\mathbf{W},\mathbf{V},\mathbf{w}} & \mathbf{c}'\mathbf{y} + \eta \\
\text{s.t.} & \mathbf{f}'\mathbf{w} \leq \eta - \mathbf{d}'\mathbf{q}, \\
& \mathbf{G}'\mathbf{w} = \mathbf{P}'\mathbf{d}, \\
& \mathbf{W}'\mathbf{f} \leq \mathbf{A}\mathbf{y} + \mathbf{B}\mathbf{q}, \\
& \mathbf{G}'\mathbf{W} = \mathbf{I} - \mathbf{P}'\mathbf{B}', \\
& \mathbf{V}'\mathbf{f} \leq \mathbf{q}, \\
& \mathbf{G}'\mathbf{V} = -\mathbf{P}', \\
& \mathbf{y}, \mathbf{w}, \mathbf{V}, \mathbf{W} \geq \mathbf{0}.
\end{array}
$$

Supply Chain Management

- Single product, two echelon, multi-period supply chain.
- Inventories managed periodically over T time periods.
- Retailer : chooses commitments $w = (w_1, \ldots, w_T)$
  - These serve as forecasts for the supplier.
  - Helps the supplier determine its production capacity.
- At the beginning of period t
  - retailer has inventory x_t
  - orders a quantity q_t at a unit cost of c_t
  - Demand d_t is realized

Costs in the Model

- Therefore, the following direct costs are incurred
  - holding cost of $h_t \max[x_t + q_t - d_t, 0]$, h_t : unit holding cost
  - shortage cost of $p_t \max[d_t - x_t - q_t, 0]$, p_t : unit shortage cost
- Contractual costs incurred
  - penalty due to deviations between committed and actual orders:

    $$\alpha_t^+ \max[q_t - w_t, 0] + \alpha_t^- \max[w_t - q_t, 0]$$

    where α_t^+, α_t^- are unit penalties for positive and negative deviations.
  - penalties on deviations between successive commitments:

    $$\beta_t^+ \max[w_t - w_{t-1}, 0] + \beta_t^- \max[w_{t-1} - w_t, 0],$$

    where β_t^+, β_t^- are associated unit penalties.
- Inventory x_{T+1} left at the end has a unit salvage value of s.

Constraints

- Balance equations, that link the inventories, order quantities and realized demand.
- Upper and lower bounds.
- Nominal Optimization Problem

$$\min \quad \Big\{ -s\left[x_{T+1}\right]^{+} + \sum_{t=1}^{T} \Big[c_t q_t + h_t \left[x_{t+1}\right]^{+} + p_t \left[-x_{t+1}\right]^{+} + \alpha_t^{+} \left[q_t - w_t\right]^{+} + \alpha_t^{-} \left[w_t - q_t\right]^{+} + \beta_t^{+} \left[w_t - w_{t-1}\right]^{+} + \beta_t^{-} \left[w_{t-1} - w_t\right]^{+} \Big\}$$

$$\begin{aligned} \text{s.t.} \quad & x_{t+1} = x_t + q_t - d_t, \quad \forall t, \\ & L_t \leq q_t \leq U_t, \quad \forall t, \\ & \hat{L}_t \leq \sum_{\tau=1}^{t} q_\tau \leq \hat{U}_t, \quad \forall t. \end{aligned}$$

Optimization under Uncertain Demand

- For a given ordering policy, events preceding time t are determined by past demands.
- That is,

$$q_t = q_t\left(d^{t-1}\right),$$

where

$$d^{t-1} = (d_1, \ldots, d_{t-1}).$$

- On the other hand, $w = (w_1, \ldots, w_T)$ must be determined before any realization of demand data.
    - These are the "here and now" decisions.
- Let the demand vector $d^T = (d_1, \ldots, d_T)$ come from the uncertainty set

$$\mathcal{U}^T = \mathcal{U}_1 \times \mathcal{U}_2 \times \ldots \times \mathcal{U}_T,$$

where $\mathcal{U}_t$ is the uncertainty of demand d_t at period t.

Adaptive Robust Optimization Problem

- The Min-Max problem is given by

$$
\begin{aligned}
\min_{x_t(),q_t(),w_t} \quad & \Bigg\{ -s\left[x_{T+1}\left(d^T\right)\right]^+ + \sum_{t=1}^{T}\left[c_t q_t\left(d^{t-1}\right) + h_t\left[x_{t+1}\left(d^t\right)\right]^+ + p_t\left[-x_{t+1}\left(d^t\right)\right]^+\right. \\
& \qquad + \alpha_t^+\left[q_t\left(d^{t-1}\right) - w_t\right]^+ + \alpha_t^-\left[w_t - q_t\left(d^{t-1}\right)\right]^+ \\
& \qquad \left. + \beta_t^+\left[w_t - w_{t-1}\right]^+ + \beta_t^-\left[w_{t-1} - w_t\right]^+ \right\} \\
\text{s.t.} \quad & \forall d^t \in \mathcal{U}^t = \mathcal{U}_1 \times \mathcal{U}_2 \times \ldots \times \mathcal{U}_t,\ t = 1,\ldots,T: \\
& x_{t+1}\left(d^t\right) = x_t\left(d^{t-1}\right) + q_t\left(d^{t-1}\right) - d_t, \quad \forall t, \\
& L_t \le q_t\left(d^{t-1}\right) \le U_t, \quad \forall t, \\
& \hat{L}_t \le \sum_{\tau=1}^{t} q_\tau\left(d^{\tau-1}\right) \le \hat{U}_t, \quad \forall t.
\end{aligned}
$$

Difficult to Solve

- Solution using Dynamic Programming.
- Difficulties
  - the objective function is not smooth
- Even for simple polyhedral uncertainty sets, the problem is NP-hard in general.
- **Core difficulty**: The functional dependence of $q_t\left(d^{t-1}\right)$ is not known.
- How about affine functions?

Affine Adaptability leads to Tractability

- q_t is an affine function of realized demands, that is,

$$q_t = q_t^0 + \sum_{\tau=0}^{t-1} q_t^\tau d_\tau .$$

- With q_t being affine functions, this enforces the variables x_t to be affine too!

$$x_{t+1}\left(d^t\right) = x_{t+1}^0 + \sum_{\tau=1}^{t} x_{t+1}^\tau d_\tau .$$

- The problem now reduces to finding the parameters

$$\{q_t^\tau, x_t^\tau\} \ \forall \tau < t, \forall t = 1, \ldots, T$$

.

- Leads to a linear optimization formulation.

Performance of Fully Adaptive and Affine-adaptive solutions

- The fully adaptive problem is intractable (solved using Dynamic Programming), whereas the affinely adaptive problem can be reformulated as a linear optimization problem.
- In Ben-Tal et.al. [2005], experiments were performed on three kinds of datasets to compare the performance of Fully adaptive, affinely adaptaiveand static robust solutions.
- In almost all the cases, the affinely adaptive solution achieved the same objective as a fully adaptive solution.

Performance

Table 2 **Opt(Min-Max), AARC, and RC Solutions for Data Sets *A*12, *D*2, and *W*12 (in Parentheses: Excess Over the Opt(Min-Max) Solution)**

Data	Uncertainty (in %)	Opt(min-max)	AARC	RC
*D*2	10	40,750.0	40,750.0 (+0.0%)	40,750.0 (+0.0%)
	20	44,150.0	44,150.0 (+0.0%)	44,150.0 (+0.0%)
	30	47,550.0	47,550.0 (+0.0%)	47,550.0 (+0.0%)
	40	50,950.0	50,950.0 (+0.0%)	50,950.0 (+0.0%)
	50	54,350.0	54,350.0 (+0.0%)	54,350.0 (+0.0%)
	60	57,760.0	57,760.0 (+0.0%)	57,760.0 (+0.0%)
	70	61,170.0	61,170.0 (+0.0%)	61,170.0 (+0.0%)
*A*12	10	913.128	913.128 (+0.0%)	1,002.941 (+9.8%)
	20	1,397.440	1,397.440 (+0.0%)	1,397.440 (+0.0%)
	30	2,190.620	2,190.620 (+0.0%)	2,190.620 (+0.0%)
	40	3,087.540	3,087.540 (+0.0%)	3,087.540 (+0.0%)
	50	4,006.040	4,006.040 (+0.0%)	4,006.040 (+0.0%)
	60	4,934.680	4,934.680 (+0.0%)	4,934.680 (+0.0%)
	70	5,863.320	5,863.320 (+0.0%)	5,863.320 (+0.0%)
*W*12	10	13,531.8	13,531.8 (+0.0%)	15,033.4 (+11.1%)
	20	15,063.5	15,063.5 (+0.0%)	18,066.7 (+19.9%)
	30	16,595.3	16,595.3 (+0.0%)	21,100.0 (+27.1%)
	40	18,127.0	18,127.0 (+0.0%)	24,300.0 (+34.1%)
	50	19,658.7	19,658.7 (+0.0%)	27,500.0 (+39.9%)
	60	21,190.5	21,190.5 (+0.0%)	30,700.0 (+44.9%)
	70	22,722.2	22,722.2 (+0.0%)	33,960.0 (+49.5%)

Conclusions

- Philosophically replacing DP with FAO or AAO is closer to human decision making.
- FAO or AAO are tractable.
- High quality performance in an interesting application.

MULTISTAGE ROBUST MIXED INTEGER OPTIMIZATION WITH ADAPTIVE PARTITIONS

15.094J: Robust Modeling, Optimization, Computation

Iain Dunning

March 4, 2015

MIT Operations Research Center

OUTLINE

1. The problem & past work
2. Motivation
3. Algorithm
4. Bounds & extension to multistage
5. Computational results & JuMPeR implementation

Based on:

> Berstimas, Dunning. Multistage Robust Mixed Integer Optimization with Adaptive Partitions. Submitted to Operations Research.
> `http://www.optimization-online.org/DB_FILE/2014/11/4658.pdf`

> …the original problem that started my research is still outstanding - namely the problem of planning or scheduling dynamically over time, particularly planning dynamically under uncertainty. If such a problem could be successfully solved it could eventually through better planning contribute to the well-being and stability of the world. - George Dantzig, in "History of Mathematical Programming", 1991

- Planning decisions across time under uncertainty is at the core of operations research.
- Decisions can be
  - **continuous** - e.g. how much stock to order
  - **discrete** - e.g. whether to operate a coal-fired power plant

The difficulty arises from the **uncertainty** in our problem

- Must make modeling decision about how to represent it:
  - May have good short-term estimates of uncertainty, but long-term?
- Must model adaptability:
  - We need to decide some things **here-and-now**
  - But can **wait-and-see** for later decisions
- Must be **tractable**

Operations management: inventory control, supply chain flexibility, project management...

Industrial: electricity unit commitment, facility location/expansion, air traffic control...

Financial: portfolio construction, financial instruments...

OUR APPROACH

Take a **robust optimization** view of uncertainty

- Assume little about uncertainty
- Good evidence of tractability

Alternative would be a **stochastic optimization** view

- Multistage discrete decisions difficult
- Need distributions

$$
\begin{aligned}
z_{\text{full}} = \min_{x} \max_{\boldsymbol{\xi} \in \Xi} \quad & \sum_{t=1}^{T} c^t(\boldsymbol{\xi}) \cdot x^t\left(\boldsymbol{\xi}^1, \dots, \boldsymbol{\xi}^{t-1}\right) \\
\text{subject to} \quad & \sum_{t=1}^{T} A^t(\boldsymbol{\xi}) \cdot x^t\left(\boldsymbol{\xi}^1, \dots, \boldsymbol{\xi}^{t-1}\right) \le b(\boldsymbol{\xi}) \quad \forall \boldsymbol{\xi} = \left(\boldsymbol{\xi}^1, \dots, \boldsymbol{\xi}^T\right) \in \Xi \\
& x \in \mathcal{X}
\end{aligned}
$$

- T time stages, $t = 1$ is here-and-now
- Uncertain parameters $\boldsymbol{\xi}^t$ for each time t
  - Uncertainty set Ξ, captures correlation across time
- Adaptive decisions x^t for each time t
  - **Fully adaptive** because policy is arbitrary function of complete history
- Deterministic & integrality constraints $\mathcal{X}$

A HIERACHY OF ADAPTABILITY

One extreme: **static policy**

- Future decisisions cannot adapt - all here & now
- Very conservative, but very tractable

Other extreme: **fully adaptive policy**

- Generally intractable
- Some success for unit commitment problem (Berstimas et al 2013)

In-between: assume simpler policy

Linear decision rules, a.k.a. affine adaptability

- Applied to RO in (Ben-tal et. al. 2004)
- Good: problem class OK, simple, sometimes optimal
- Bad: no discrete recourse, changes problem structure, numerical problems
- Extensions: deflected linear decision rules (Chen et. al 2008), polynomial adaptability (Bertsimas et. al. 2010)

Piecewise linear decision rules

- Relatively new, (Bertsimas & Georghiou 2013, 2014)
- Piecewise linear for continuous decisions, piecewise constant for integer
- (2013) uses a cutting-plane method, scaling issues
- (2014) shows good results for multistage.

Finite adaptability

- Partition the uncertainty set, associate decision for each
- Effectively: piecewise constant policy, works well for discrete
- Preserves problem structure better than LDRs
- But can combine with LDRs to get piecewise LDRs!

How to pick the partitions?

- A priori, e.g. (Vayanos et. al. 2011)
- Fix number of partitions and optimize, e.g. (Bertsimas & Caramanis 2010), (Hanasusanto et. al. 2014)
- Optimizing directly results in very difficult MIO

Solve problem with static policy

Indentify good partitions heuristically

Solve partitioned problem

Identify more partitions, or stop

TWO-STAGE PROBLEMS

$$
\begin{aligned}
\min_{\mathbf{x},z}\ & z \\
\text{subject to}\quad & \mathbf{c}^1(\boldsymbol{\xi})\cdot\mathbf{x}^1 + \mathbf{c}^2(\boldsymbol{\xi})\cdot\mathbf{x}^2(\boldsymbol{\xi}) \le z \qquad \forall\boldsymbol{\xi}\in\Xi \\
& \mathbf{a}_i^1(\boldsymbol{\xi})\cdot\mathbf{x}^1 + \mathbf{a}_i^2(\boldsymbol{\xi})\cdot\mathbf{x}^2(\boldsymbol{\xi}) \le b_i(\boldsymbol{\xi})\forall\boldsymbol{\xi}\in\Xi,\ i\in\{1,\dots,m\} \\
& \mathbf{x}\in\mathcal{X},
\end{aligned}
$$

- Consider solving the static policy, continuous two-stage problem
- Using cutting-planes, we will add constraints until solution is feasible
- For every "uncertain constraint" we might add multiple cuts
- Each cut is associated with a value of $\boldsymbol{\xi}$
- If we remove all cuts that have slack $s > 0$, **solution will not change**
- Active constraints = active uncertain parameters, or "samples"

- Let $\hat{\Xi}$ be set of active uncertain parameters, 0 or 1 per uncertain constraint.
- Construct arbitrary partition of Ξ to create Ξ_1 and Ξ_2

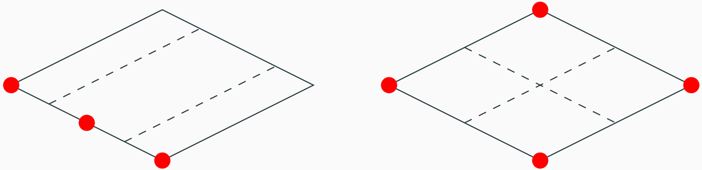
Theorem

If either $\hat{\Xi} \subseteq \Xi_1$ or $\hat{\Xi} \subseteq \Xi_2$, then there will be no improvement in the objective.

Proof.

If $\hat{\Xi} \subseteq \Xi_1$ then the solution associated with that partition can't be better than for Ξ, so overall solution is the same. $\square$

- We have to split the active uncertain parameters to improve the solution
- Given a set of N points, a **Voronoi diagram** defines a partition for each point such that each point in the partition is closer to that point than any other
- Use active uncertain parameters as the points



Partitions defined by hyperplanes:

$$\begin{aligned}
\Xi\left(\hat{\boldsymbol{\xi}}_{i}\right) &= \Xi \cap \left\{ \boldsymbol{\xi} \,\middle|\, \left\| \hat{\boldsymbol{\xi}}_{i} - \boldsymbol{\xi} \right\|_{2} \leq \left\| \hat{\boldsymbol{\xi}}_{j} - \boldsymbol{\xi} \right\|_{2}, \forall j \in I,\ i \neq j \right\} \\
&= \Xi \cap \left\{ \boldsymbol{\xi} \,\middle|\, \sum_{k} \left(\hat{\xi}_{i,k} - \xi_{k} \right)^{2} \leq \sum_{k} \left(\hat{\xi}_{j,k} - \xi_{k} \right)^{2}, \forall j \in I,\ i \neq j \right\} \\
&= \Xi \cap \left\{ \boldsymbol{\xi} \,\middle|\, \sum_{k} \left(\frac{\hat{\xi}_{i,k} - \hat{\xi}_{j,k}}{2} \right) \xi_{k} \geq \sum_{k} \left(\hat{\xi}_{i,k}^{2} - \hat{\xi}_{j,k}^{2} \right), \forall j \in I,\ i \neq j \right\},
\end{aligned}$$

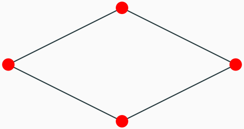
- Each partition defined by Ξ + $N - 1$ linear constraints
- Polyhedral Ξ gives polyhedral partitions
- Computational complexity of Voronoi diagrams bad if enumerating, but we don't need to

Solve $\rightarrow$ active parameters $\rightarrow$ partition $\rightarrow$ solve

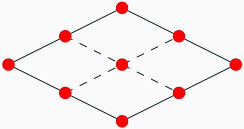
Two options for further partitioning:

- Collect new active parameters, and reconstruct partitions **non-nested**
- Associate active parameters with partitions, create **nested** partitions

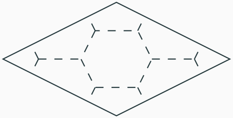
NON-NESTED



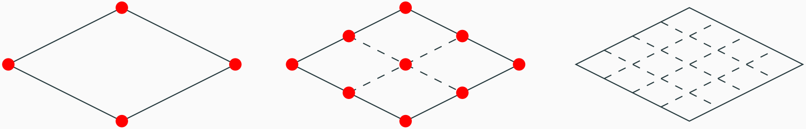
$\mathcal{F}^1 = \{\}$



$$\mathcal{F}^2 = \{(10, 10), (20, 5), (30, 10), (20, 15)\}$$



$$\mathcal{F}^3 = \{(10, 10), (15, 12.5), (20, 15), (25, 12.5), (30, 10), (20, 10), (15, 7.5)\ (20, 5)\ (25, 7.5)\}$$



$\mathcal{T}^1 = \text{root}$

$\mathcal{T}^2 = \text{root}$
- 10,10
- 20,15
- 30,10
- 20,5

$\mathcal{T}^3 = \text{root}$
- 10,10
  - 10,10
  - 15,12.5
  - 20,10
  - 15,7.5
- 20,15
  - ...
- 30,10
  - ...
- 20,5
  - 20,10
  - 25,12.5
  - 30,10
  - 25,7.5

$$\text{Siblings}\left(\hat{\boldsymbol{\xi}}\right) = \text{Children}\left(\text{Parent}\left(\hat{\boldsymbol{\xi}}\right)\right).$$

$$\begin{aligned}\Xi\left(\hat{\boldsymbol{\xi}}_{\mathrm{i}}\right) = &\left\{\boldsymbol{\xi} \,\middle|\, \left\|\hat{\boldsymbol{\xi}}_{\mathrm{i}} - \boldsymbol{\xi}\right\|_2 \leq \left\|\hat{\boldsymbol{\xi}}_{\mathrm{j}} - \boldsymbol{\xi}\right\|_2 \quad \forall \hat{\boldsymbol{\xi}}_{\mathrm{j}} \in \text{Sibls}\left(\hat{\boldsymbol{\xi}}_{\mathrm{i}}\right)\right\} \\ &\cap \left\{\boldsymbol{\xi} \,\middle|\, \left\|\text{Parent}\left(\hat{\boldsymbol{\xi}}_{\mathrm{i}}\right) - \boldsymbol{\xi}\right\|_2 \leq \left\|\hat{\boldsymbol{\xi}}_{\mathrm{j}} - \boldsymbol{\xi}\right\|_2 \quad \forall \hat{\boldsymbol{\xi}}_{\mathrm{j}} \in \text{Sibl}\left(\text{Parent}\left(\hat{\boldsymbol{\xi}}_{\mathrm{i}}\right)\right)\right\} \\ &\vdots \\ &\cap \Xi,\end{aligned}$$

EXAMPLE: TWO STAGE INVENTORY CONTROL

- Must order stock to meet future unknown demand
- Can order any amount **now**, cost \$50 per unit
- Realize demand, then can order
  - One bulk shipment of 25 units at \$60 per unit
  - Another bulk shipment of 25 units at \$75 per unit
- Holding costs of \$65 per unit

$$\begin{aligned} \min\ & z \\ \text{subject to}\ & 50x^1 + 65I^2(\xi) + 1500y_A^2(\xi) + 1875y_B^2(\xi) \leq z && \forall \xi \in \Xi \\ & I^2(\xi) \geq 0 && \forall \xi \in \Xi \\ & x^1 \geq 0 \\ & y_A^2(\xi),\ y_B^2(\xi) \in \{0, 1\} && \forall \xi \in \Xi \end{aligned}$$

$$I^2(\xi) = x^1 - \xi + 25y_A^2(\xi) + 25y_B^2(\xi), \qquad \Xi = \{\xi \,|\, 5 \leq \xi \leq 95\}$$

Solve static policy:

- $z = 10600$, $x^1 = 95$, $y_A^2 = 0$, $y_B^2 = 0$
- Worst cases are $\hat{\xi} = 5$ and $\hat{\xi} = 95$
- Create two partitions:
  - $\Xi\left(\hat{\xi} = 5\right) = \{\xi \,|5 \leq \xi \leq 50\}$
  - $\Xi\left(\hat{\xi} = 95\right) = \{\xi \,|50 \leq \xi \leq 95\}$

Solve with new partitions:

- $z = 7926$, $x^1 = 70$
  - $y_{A,1}^2 = 0$, $y_{B_1}^2 = 0$
  - $y_{A,2}^2 = 1$, $y_{B_2}^2 = 0$
- Worst cases are $\hat{\xi} = 5, \hat{\xi} = 50$ and $\hat{\xi} = 50, \hat{\xi} = 95$

Non-nested version:

- $\Xi\left(\hat{\xi}=5\right)=\{\xi\,|5\leq\xi\leq 35\}$
- $\Xi\left(\hat{\xi}=50\right)=\{\xi\,|35\leq\xi\leq 65\}$
- $\Xi\left(\hat{\xi}=95\right)=\{\xi\,|65\leq\xi\leq 95\}$
- $z=7575$, $x^1=45$, and

$$y_A^2(\xi)=\begin{cases}0, & 5\leq\xi<35,\\ 1, & 35\leq\xi\leq 95,\end{cases}$$

and

$$y_B^2(\xi)=\begin{cases}0, & 5\leq\xi<35,\\ 1, & 65\leq\xi\leq 95.\end{cases}$$

Nested partitions:

- Parent $\hat{\xi} = 5$:
  - $\Xi\left(\hat{\xi} = 5\right) = \{\xi \,| 5 \le \xi \le 27.5\}$
  - $\Xi\left(\hat{\xi} = 50\right) = \{\xi \,| 27.5 \le \xi \le 50\}$
- Parent $\hat{\xi} = 95$:
  - $\Xi\left(\hat{\xi} = 50\right) = \{\xi \,| 50 \le \xi \le 72.5\}$
  - $\Xi\left(\hat{\xi} = 95\right) = \{\xi \,| 72.5 \le \xi \le 95\}$
- $z = 7375$, $x^1 = 47.5$, and

$$y_A^2(\xi) = \begin{cases} 0, & 5 \le \xi < 27.5, \\ 1, & 27.5 \le \xi \le 95, \end{cases}$$

and

$$y_B^2(\xi) = \begin{cases} 0, & 5 \le \xi < 72.5, \\ 1, & 72.5 \le \xi \le 95. \end{cases}$$

Summary of results:

- Static: $z = 10600$
- First iteration: $z = 7926$
- Second iteration, non-nested: $z = 7575$
- Second iteration, nested: $z = 7375$
- Fully adaptive: $z = 7250$

Can easily incorporate linear decision rules

Make substitution

$$x^2(\xi) = F\xi + g,$$

for continuous decisions

Equivalent to piecewise affine once we partition, but with breaks heuristically determined, i.e.

$$x^2(\xi) = \begin{cases} F_1\xi + g_2, & \xi \in \Xi\left(\hat{\xi}_1\right), \\ F_2\xi + g_2, & \xi \in \Xi\left(\hat{\xi}_2\right), \\ \vdots & \vdots \end{cases}$$

There is an objective value associated with each partition

- Define active partition as partition that is binding overall objective
- Only look at uncertain parameters for active partitions
- e.g. in example, partition objectives are 5137.5, 6800, 5337.5, and 7375

We have been using uncertain parameters with minimum slack

- Could use only $s = 0$ uncertain parameters
- Situation dependent: may be hard to get 0 slack in many problems

MULTISTAGE PROBLEMS

One thing: must satisfy **non-anticipativity**

With affine, automatically satisfied

With finite adaptability, **easy to violate!**

Consider the same inventory problem as before, but $T = 3$

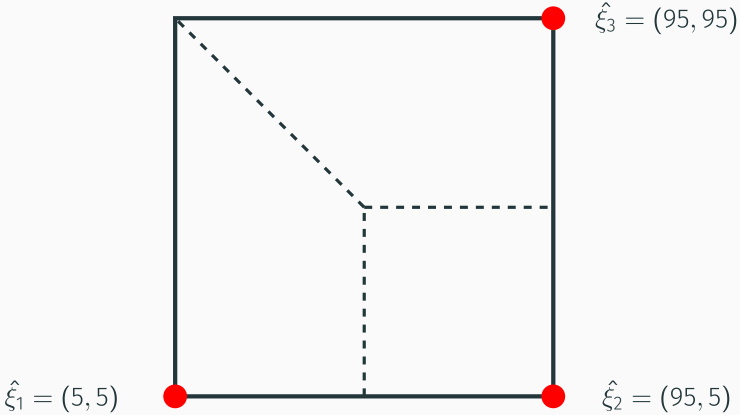
Uncertainty set

$$\Xi = \left\{5 \leq \xi^1, \xi^2 \leq 95\right\}$$

Active samples for static policy are

- $\hat{\boldsymbol{\xi}}_1 = (5, 5)$ (objective)
- $\hat{\boldsymbol{\xi}}_2 = (95, 5)$ (demand met, $t = 2$)
- $\hat{\boldsymbol{\xi}}_3 = (95, 95)$ (demand met, $t = 3$)

Construct partitions as before...


$\hat{\xi}_3 = (95, 95)$
$\hat{\xi}_1 = (5, 5)$
$\hat{\xi}_2 = (95, 5)$

Had to add $x_1^2 = x_3^2$ and $x_2^2 = x_3^2$, no adaptability left at $t = 2$

Solution: modify partitioning scheme to be aware of time

Goal: balance having minimal set of anticipativity constraints while getting most useful partitions.

For each pair $\hat{\boldsymbol{\xi}}_i$ and $\hat{\boldsymbol{\xi}}_j$ as before

1. Determine which components $\hat{\boldsymbol{\xi}}^t$ shared
2. Construct hyperplane using only components up to the first time stage they differ, i.e.

$$\Xi\left(\hat{\boldsymbol{\xi}}_i\right) = \Xi \cap \left\{ \boldsymbol{\xi} \,\middle|\, \left\|\hat{\boldsymbol{\xi}}_i^{t_{i,j}} - \boldsymbol{\xi}^{t_{i,j}}\right\|_2 \leq \left\|\hat{\boldsymbol{\xi}}_j^{t_{i,j}} - \boldsymbol{\xi}^{t_{i,j}}\right\|_2 \quad \forall \hat{\boldsymbol{\xi}}_j \in \mathcal{F}^k \right\}$$

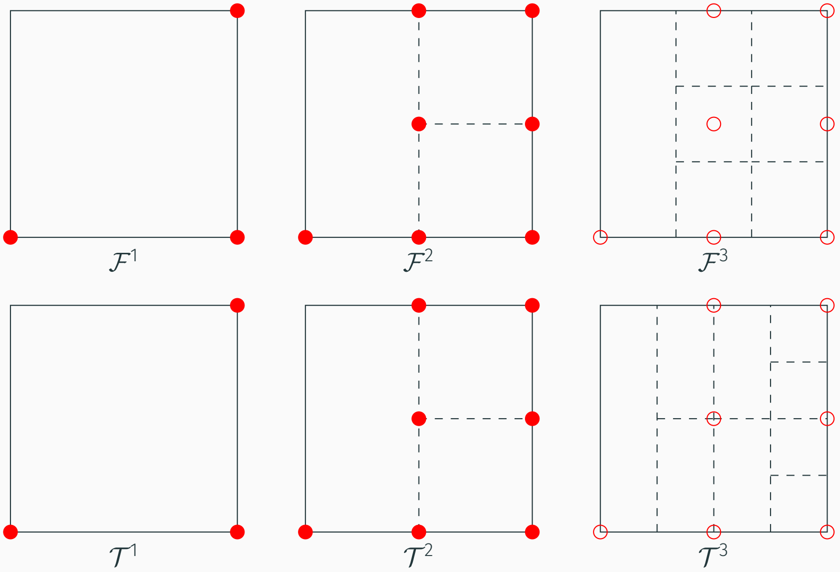
   where $t_{i,j}$ is the min t s.t. $\hat{\boldsymbol{\xi}}_i^t \neq \hat{\boldsymbol{\xi}}_j^t$
3. Proposition: sufficient to then enforce $x_i^t = x_j^t$ iff $\hat{\boldsymbol{\xi}}_i^{1,\ldots,t-1} = \hat{\boldsymbol{\xi}}_j^{1,\ldots,t-1}$ to ensure nonanticipativity

Apply same time-dependent partitioning rules, but nested

Can use similar rule to add non-anticipativity, but it is **overconservative**

Relatively cheap to check intersection of partitions, much less conservative

MULTISTAGE NESTED PARTITIONING



BOUNDS

Provide three types of bounds:

1. Lower bound on fully adaptive solution
2. Upper bound on subsequent iterations (monotonicity)
3. Lower bound on subsequent iterations

MIO branch-and-bound: have best integer UB, continuous relaxation LB

- Termination criteria e.g. $\frac{(\mathrm{UB}-\mathrm{LB})}{\mathrm{LB}}$

AMIO: UB is best approximation to fully adaptive, LB $=$?

- Termination criteria could be same e.g. $\frac{(\mathrm{UB}-\mathrm{LB})}{\mathrm{LB}}$

Proposition: the solution to

$$
\begin{aligned}
z_{\text{lower}}(\mathcal{A}) = \min_{\mathbf{x},z} \quad & z \\
\text{subject to} \quad & \sum_{t=1}^{T} \mathbf{c}^t\left(\hat{\boldsymbol{\xi}}_i\right) \cdot \mathbf{x}_i^t \leq z && \forall \hat{\boldsymbol{\xi}}_i \in \mathcal{A} \\
& \sum_{t=1}^{T} \mathbf{A}^t\left(\hat{\boldsymbol{\xi}}_i\right) \cdot \mathbf{x}_i^t \leq \mathbf{b}\left(\hat{\boldsymbol{\xi}}_i\right) && \forall \hat{\boldsymbol{\xi}}_i \in \mathcal{A} \\
& \mathbf{x}_i^t = \mathbf{x}_j^t && \forall \hat{\boldsymbol{\xi}}_i, \hat{\boldsymbol{\xi}}_j \in \mathcal{A} \text{ s.t. } \hat{\boldsymbol{\xi}}_i^{1,\ldots,t-1} = \hat{\boldsymbol{\xi}}_j^{1,\ldots,t-1} \\
& \mathbf{x} \in \mathcal{X},
\end{aligned}
$$

is a lower bound to the fully adaptive optimization problem.

Proof follows from the fact that $\mathcal{A}$ is a subset of Ξ, and we respect nonanticipativity. This is similar to the two-stage ”scenario based bound” in Hadjiyiannis et. al. 2011.

LOWER BOUND ON FULLY ADAPTIVE SOLUTION

Open question: can we do better?

This bound is very practical as we have these samples already “free”

Can improve bound by sampling more from uncertainty set, but to what end?

Getting a better bound with what we already know is key to progress

The non-nested variant **does not** decrease monotonically

The nested variant **does** decreases monotonically

Utility: can use solution from one iteration to warm start next iteration

LOWER BOUND ON SUBSEQUENT ITERATIONS

Use duality to estimate improvement that could be obtained by partitioning

Not directly applicable to AMIO, but can check relaxation

Estimate usefulness of partitioning further

COMPUTATIONAL EXPERIMENTS

Observation: spend more time, get better solutions

What is better? Gap? Improvement?

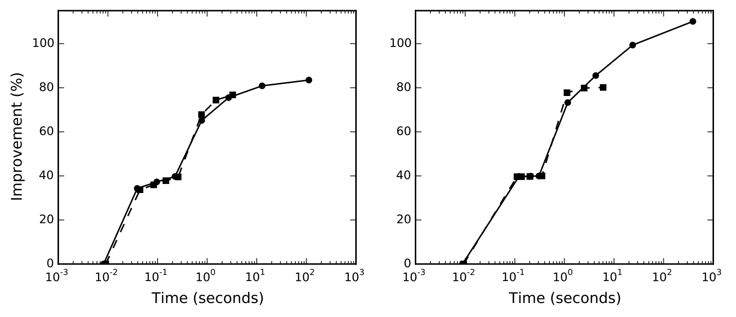
Upper and lower bounds shrink simultaneously

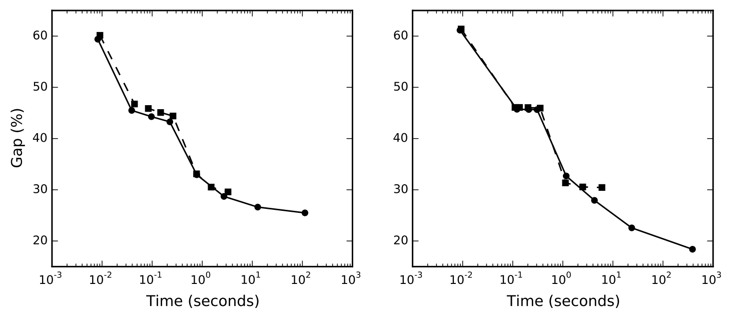
Can build projects now or later, but more expensive later.

$$
\begin{aligned}
\max_{z,\mathbf{x}} \quad & z \\
\text{subject to} \quad & r(\boldsymbol{\xi}) \cdot \left(x^1 + \theta x^2(\boldsymbol{\xi})\right) \geq z && \forall \boldsymbol{\xi} \in \Xi \\
& c(\boldsymbol{\xi}) \cdot \left(x^1 + x^2(\boldsymbol{\xi})\right) \leq B && \forall \boldsymbol{\xi} \in \Xi \\
& x^1 \in \{0,1\}^N \\
& x^2(\boldsymbol{\xi}) \in \{0,1\}^N && \forall \boldsymbol{\xi} \in \Xi,
\end{aligned}
$$

Shared uncertain factors induce structure across project revenues and costs.

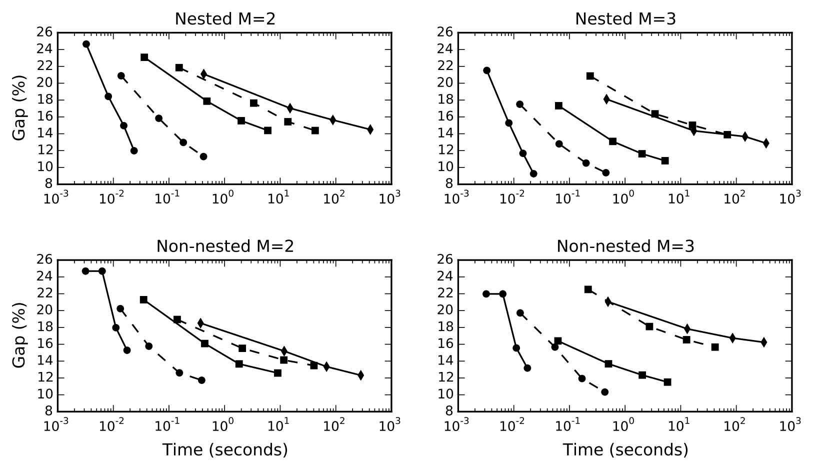
$N = 10$	Variant	Iteration 1	2	3	4	5	6	7	8
Total Time (s)	Non-nested	0.0	0.0	0.1	0.2	0.8	2.7	12.8	112.4
	Nested	0.0	0.0	0.1	0.1	0.2	0.7	1.5	3.2
Improvement (%)	Non-nested	0	34	37	40	65	76	81	83
	Nested	0	34	36	38	40	68	74	77
Gap (%)	Non-nested	62	48	46	45	34	29	27	34
	Nested	62	48	47	46	45	34	31	30


Improvement (%)
0
20
40
60
80
100
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
Time (seconds)
Time (seconds)

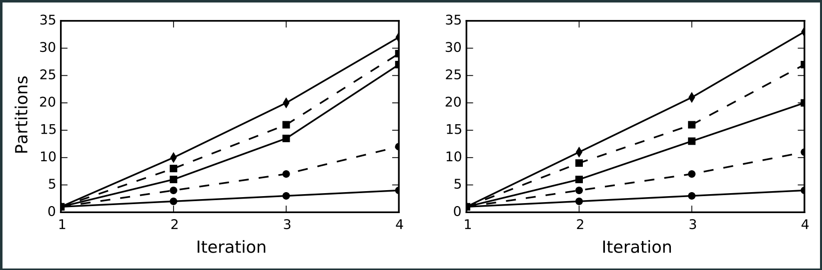

Gap (%)
20
30
40
50
60
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
Time (seconds)
Time (seconds)

Similar to working example before, but generalized to T stages:

$$
\begin{aligned}
\min_{\mathbf{x},\mathbf{y},\mathbf{I}} \quad & \sum_{t=2}^{T} \left(c_x x^{t-1} + c_h I^t + \sum_{m=1}^{M} c_m q_m y_m^t \right) && \\
\text{s.t} \quad & I^{t-1} + x^{t-1} + \sum_{,m=1}^{M} q_m y_m^t - \xi^t = I^t && \forall t \in \{2, \dots, T\} \\
& \sum_{s=1}^{t-1} x^s \leq \bar{x}_{tot,t} && \forall t \in \{2, \dots, T\} \\
& I^t \geq 0 && \forall t \in \{2, \dots, T\} \\
& x^{t-1} \geq 0 && \forall t \in \{2, \dots, T\} \\
& \mathbf{y}^t \in \{0,1\}^M, && \forall t \in \{2, \dots, T\}
\end{aligned}
$$



Nested M=2
Nested M=3
Non-nested M=2
Non-nested M=3
Gap (%)
Time (seconds)



Partitions
35
30
25
20
15
10
5
0
1
2
3
4
Iteration
35
30
25
20
15
10
5
0
1
2
3
4
Iteration

CONCLUSION

Finite adaptability with heuristically chosen partitions performs well

Have a lower bound, can trade off time for quality

Somewhat like branch-and-bound in spirit

FUTURE WORK

Smarter partitions

- Guess and improve?
- Use all active samples

Tighter integration into branch & bound

Better lower bounds

Class projects? Let's collaborate!

JuMPeR doesn’t yet have this implemented ”for free”

But can easily implement in JuMPeR!

Demo if time, post notebook otherwise

15.094J: Robust Modeling, Optimization and Computation

Lecture 11: Power of Robust Policies in Adaptive Optimization

Motivation

- RO is tractable
- But how much do we lose in performance?
- Is it worse for multistage optimization?

Stochastic Model

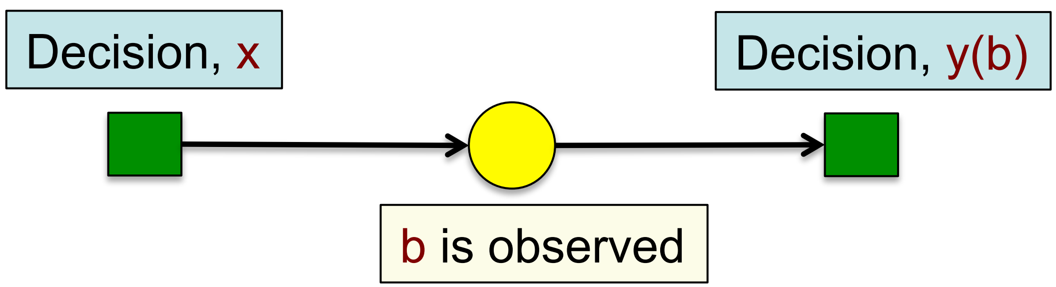
- **Two-stage Stochastic Optimization Model**

$$
\begin{aligned}
\textbf{zStoch} = \min c^T x + \mathbb{E}[d^T y(b)] & \\
Ax + By(b) &\geq b, \forall b \in \mathcal{U} \\
x &\in \mathbb{R}^n_+ \times \mathbb{Z}^p_+ \\
y(b) &\in \mathbb{R}^n_+ \times \mathbb{Z}^p_+
\end{aligned}
$$

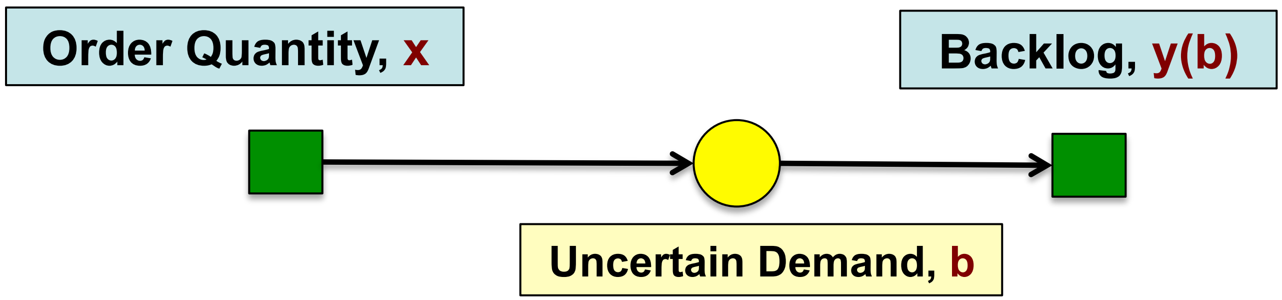
(Minimize **expected cost**)

(Uncertainty Set)

(Uncertain Right Hand Side)



Inventory Management



Minimize $\mathbb{E}_b\left[h(x-b)^+ \; + \; p(b-x)^+\right]$

$$\begin{aligned} \min \ \mathbb{E}_b\Big[\overbrace{h\big(x + y(b) - b\big)}^{\text{Holding Cost}} \;+\; \overbrace{p \cdot y(b)}^{\text{Backorder Penalty}}\Big] & \\ x + y(b) \;\geq\; & \; b, \ \forall\, b \\ x, y(b) \;\geq\; & \; 0 \end{aligned}$$

Stochastic Model

- **Two-stage Stochastic Optimization Model**

$$
\begin{aligned}
\textbf{zStoch} = \min c^T x + \mathbb{E}[d^T y(b)] & & \\
Ax + By(b) &\geq b, \ \forall b \in \mathcal{U} \\
x &\in \mathbb{R}^n_+ \times \mathbb{Z}^p_+ \\
y(b) &\in \mathbb{R}^n_+ \times \mathbb{Z}^p_+
\end{aligned}
$$

- **Computationally intractable in general**
- **Two-stage problem is #P-hard [Dyer and Stougie (2001)]**
- **Multi-stage problem is PSPACE-hard [Dyer and Stougie (2001)]**

Adaptive Optimization Model

- **Two-stage Adaptive Optimization Model**

$$
\begin{aligned}
\textbf{zAdapt} = \min c^T x + \boxed{\max_{b \in \mathcal{U}} \; d^T y(b)} & \quad \text{(Minimize worst-case cost)} \\
Ax + By(b) \;\geq\; b, \; \forall b \in \mathcal{U} & \quad \text{(Uncertainty Set)} \\
x \;\in\; \mathbb{R}^n_+ \times \mathbb{Z}^p_+ & \\
y(b) \;\in\; \mathbb{R}^n_+ \times \mathbb{Z}^p_+ &
\end{aligned}
$$

(Uncertain Right Hand Side)

- **Still computationally intractable in general**
- **Even approximating LO within an factor of O(log m) is NP-hard [Feige et al.'07]**

Robust Optimization Model

$$
\begin{aligned}
\mathbf{zRob} = \min c^T x + d^T y & \\
Ax + By & \geq b, \forall b \in \mathcal{U} \\
x & \in \mathbb{R}^n_+ \times \mathbb{Z}^p_+ \\
\boxed{y} & \in \mathbb{R}^n_+ \times \mathbb{Z}^p_+
\end{aligned}
$$

(Minimize Cost of a **static solution**)

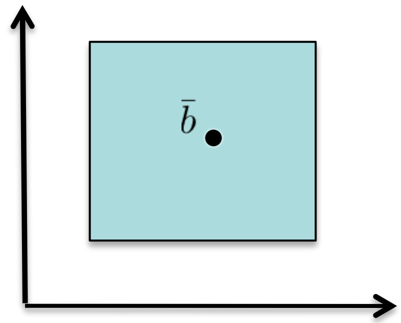
(Uncertainty Set)

(Uncertain Right Hand Side)

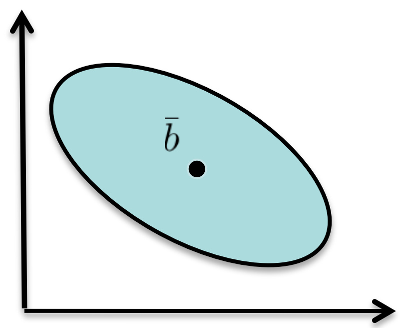
Solution **y** does **not depend on b**

- **Computationally tractable**
- **But does it give a highly conservative solution?**

Uncertainty Sets



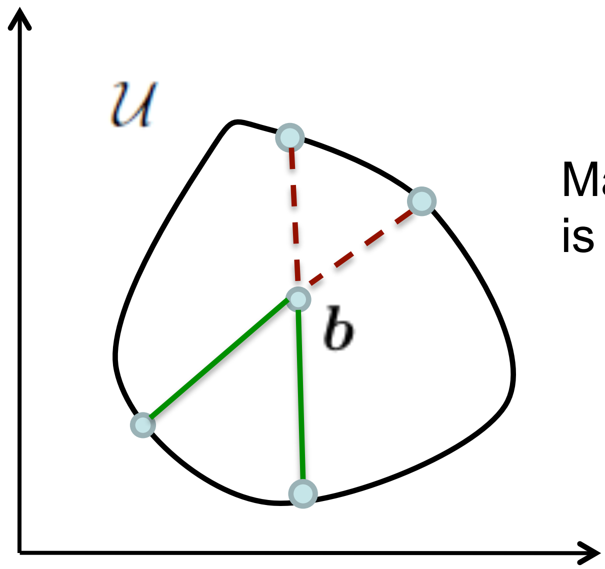
Hypercube : $||b - \bar{b}||_\infty \leq \beta$



Ellipsoid : $||\Sigma(b - \bar{b})||_2 \leq \beta$

Norm-Ball : $||\Sigma(b - \bar{b})||_p \leq \beta$

Symmetry of $\mathcal{U}$



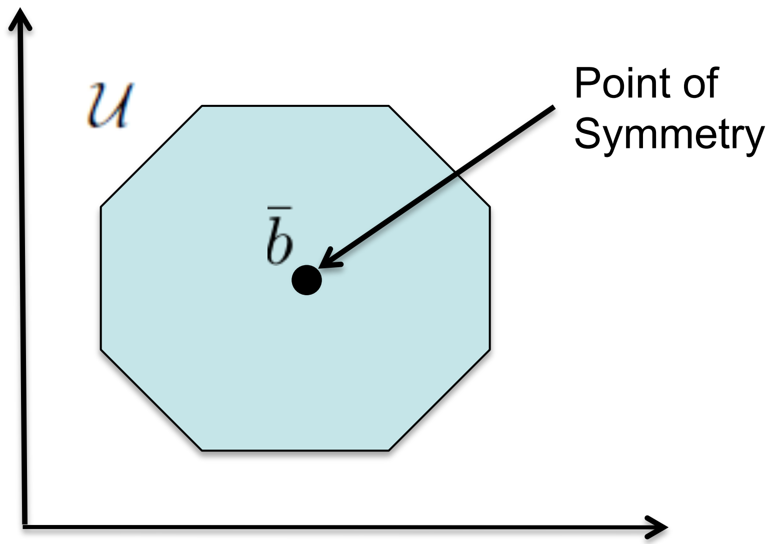
Maximum symmetry point
is **the point of symmetry of $\mathcal{U}$**

$\mathbf{sym}(b, \mathcal{U})$: minimum ratio of red and green segments

$$\mathbf{sym}(b, \mathcal{U}) = \max\{\alpha \mid b + \alpha \cdot (b - b') \in \mathcal{U},\ \forall b' \in \mathcal{U}\}$$

$$\mathbf{sym}(\mathcal{U}) = \max_{b \in \mathcal{U}} \ \mathbf{sym}(b, \mathcal{U})$$

Example (s=1)

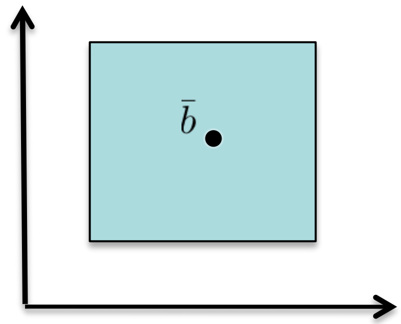


$$(\bar{b} - \delta) \in \mathcal{U} \Leftrightarrow (\bar{b} + \delta) \in \mathcal{U}, \ \forall \delta$$

$$\textsf{sym}(\mathcal{U}) = 1$$

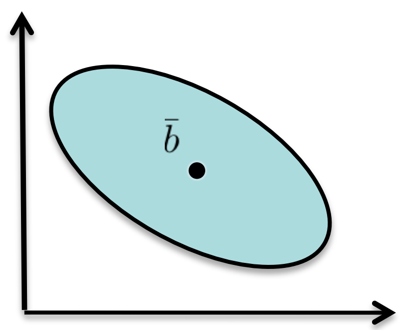
More Examples (s=1)

$\textbf{sym}(\mathcal{U}) = 1$



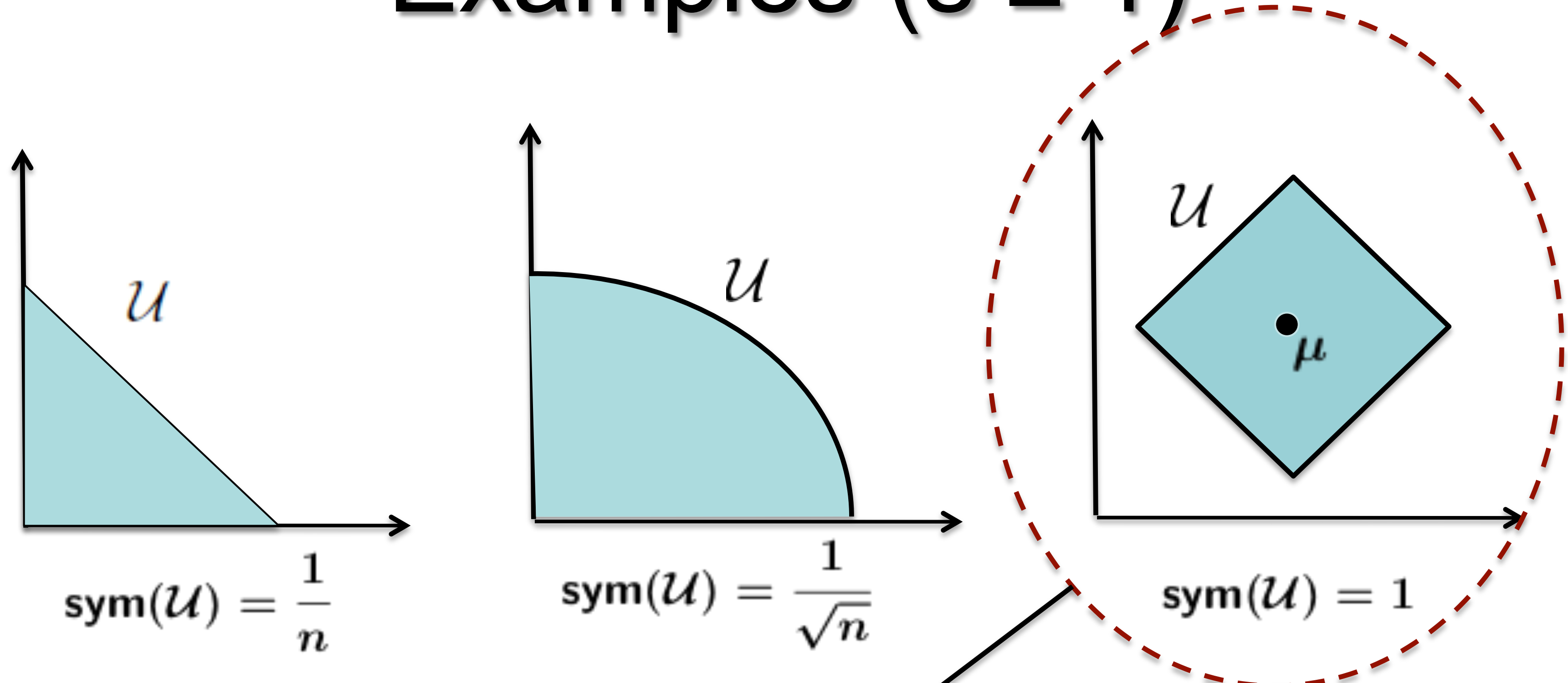
$\boxed{\textbf{Hypercube : } ||b - \bar{b}||_\infty \le \beta}$

$\textbf{sym}(\mathcal{U}) = 1$



$\boxed{\textbf{Ellipsoid : } ||\Sigma(b - \bar{b})||_2 \le \beta}$

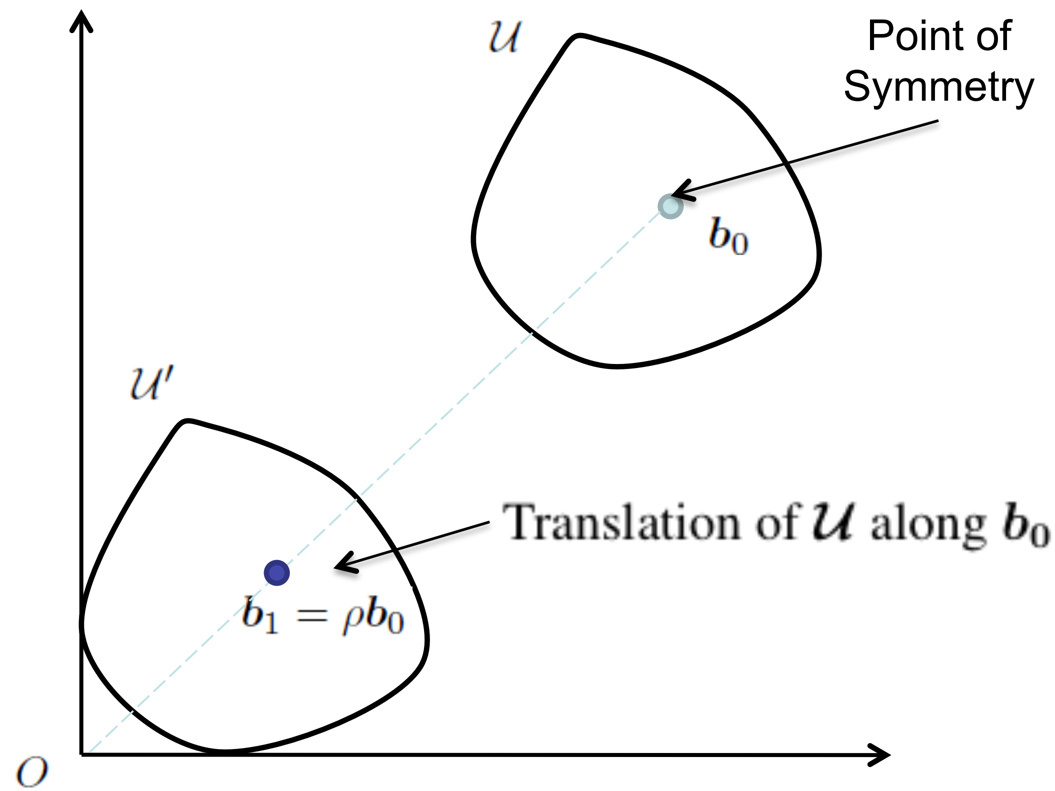
Examples (s ≤ 1)



$\mathrm{sym}(\mathcal{U}) = \frac{1}{n}$ $\quad$ $\mathrm{sym}(\mathcal{U}) = \frac{1}{\sqrt{n}}$ $\quad$ $\mathrm{sym}(\mathcal{U}) = 1$

$$\mathcal{U} = \left\{ b \in \mathbb{R}^n_+ : \left| \frac{\sum_{i \in S} b_i - |S|\mu}{\sqrt{|S|}} \right| \leq 2,\ \forall S \subseteq N := \{1, \ldots, n\} \right\}$$

Translation Factor of $\mathcal{U}$



$$\text{Translation factor of } \mathcal{U}, \ \rho(\boldsymbol{\mathcal{U}}) = \frac{||\boldsymbol{b_1}||}{||\boldsymbol{b_0}||}$$

Results: Robust Solutions

Stochastic (zStoch)

$$\begin{aligned} \min c^T x + \mathbb{E}[d^T y(b)] \\ Ax + By(b) &\geq b, \ \forall b \in \mathcal{U} \\ x, y(b) &\geq 0 \end{aligned}$$

Adaptive (zAdapt)

$$\begin{aligned} \min c^T x + \max_b d^T y(b) \\ Ax + By(b) &\geq b, \ \forall b \in \mathcal{U} \\ x, y(b) &\geq 0 \end{aligned}$$

Theorem 1 *Let* $\rho = \rho(\mathcal{U})$ *and* $s = \mathsf{sym}(\mathcal{U})$. *Then,*

$$\frac{z_{\mathsf{Rob}}}{z_{\mathsf{Stoch}}} \leq \left(1 + \frac{\rho}{s}\right)$$

- **Assumption**: **E[b] =** $\bar{b}$ where $\bar{b}$ is the **point of symmetry**

Our Results: Implications

Stochastic (zStoch)

$$\begin{aligned}\min c^T x + \mathbb{E}[d^T y(b)] \\ Ax + By(b) &\geq b, \ \forall b \in \mathcal{U} \\ x, y(b) &\geq 0\end{aligned}$$

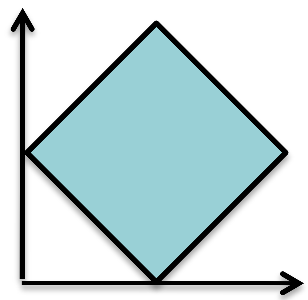
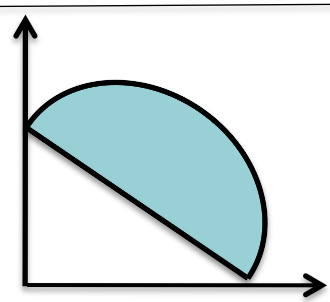
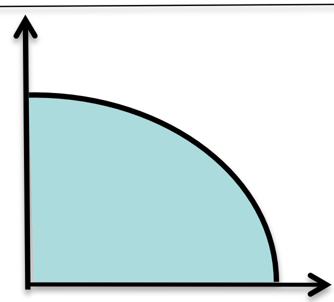
Adaptive (zAdapt)

$$\begin{aligned}\min c^T x + \max_b d^T y(b) \\ Ax + By(b) &\geq b, \ \forall b \in \mathcal{U} \\ x, y(b) &\geq 0\end{aligned}$$

Uncertainty Set (U) (RHS)	Stochasticity Gap zRob/zStoch	Adaptability Gap zRob/zAdapt
Symmetric (s=1)	$(1+\rho) \leq 2$	$(1+\rho) \leq 2$
General ($1/n < s \leq 1$)	$(1+\rho/s)$	$(1+\rho/s)$

- **Assumption**: **E[b] =** $\bar{b}$ where $\bar{b}$ is the **point of symmetry**

Bounds for different Sets

$\mathcal{U}(\rho = 1)$	**sym($\mathcal{U}$)**	**Stochasticity Gap**
	$\mathbf{1}$	$\mathbf{2}$
	$\dfrac{\mathbf{1}}{\sqrt{\mathbf{2}}}$	$(\mathbf{1} + \sqrt{\mathbf{2}})$
	$\dfrac{\mathbf{1}}{\sqrt{\boldsymbol{n}}}$	$(\mathbf{1} + \sqrt{\boldsymbol{n}})$

Integer Variables

Stochastic (zStoch)

$$
\begin{aligned}
\min c^T x + \mathbb{E}[d^T y(b)] & & \\
Ax + By(b) &\geq b, \ \forall b \in \mathcal{U} \\
x &\in \mathbb{R}^n_+ \times \mathbb{Z}^p_+ \\
y(b) &\in \mathbb{R}^n_+
\end{aligned}
$$

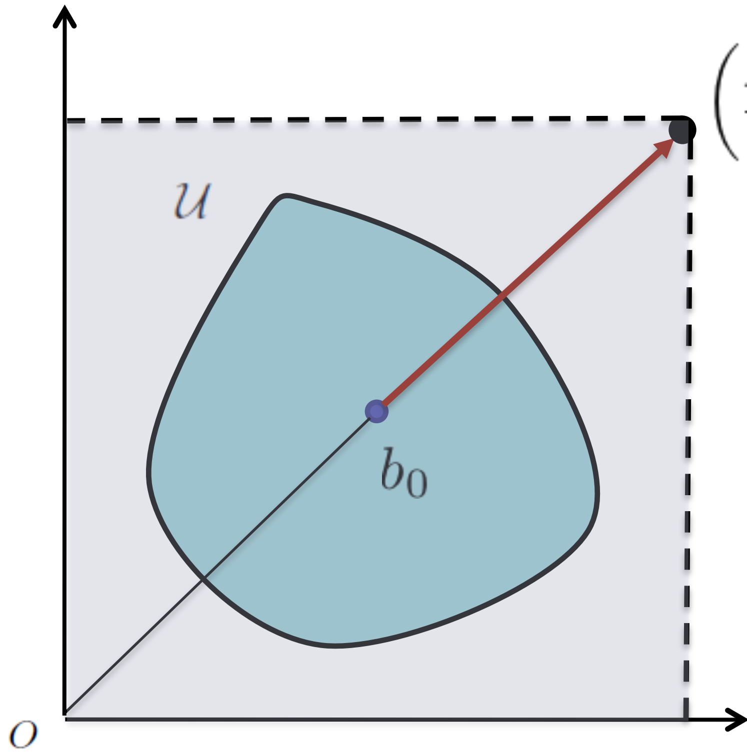
Adaptive (zAdapt)

$$
\begin{aligned}
\min c^T x + \max_b d^T y(b) & & \\
Ax + By(b) &\geq b, \ \forall b \in \mathcal{U} \\
x &\in \mathbb{R}^n_+ \times \mathbb{Z}^p_+ \\
y(b) &\in \mathbb{R}^n_+ \times \mathbb{Z}^p_+
\end{aligned}
$$

Uncertainty Set (U) (RHS)	Stochasticity Gap zRob/zStoch	Adaptability Gap zRob/zAdapt
Symmetric (s=1)	$\lceil (1+\rho) \rceil = 2$	$\lceil (1+\rho) \rceil = 2$
General ($1/n < s \leq 1$)	$\lceil (1+\rho/s) \rceil$	$\lceil (1+\rho/s) \rceil$

- **Assumption**: **E[b] =** $\bar{b}$ where $\bar{b}$ is the **point of symmetry**

Proof



$$\left(1+\frac{\rho}{s}\right) b_0 = \tau \cdot b_0$$

$$\boxed{b \leq \tau \cdot b_0, \ \forall b \in \mathcal{U}}$$

Optimal Stochastic Solution: $x^*, y^*(b), \ \forall b \in \mathcal{U}$

Feasible Static Solution: $(\tau x^*, \tau y^*(b_0))$

$$A(\tau x^*) + B(\tau y^*(b_0)) \geq \tau b_0 \geq b, \forall b \in \mathcal{U}$$

Cost Analysis

$$z_{\mathsf{Rob}} \leq \tau(c^T x^* + d^T y^*(b_0))$$

$$z_{\mathsf{Stoch}} = c^T x^* + \mathbb{E}_b[d^T y^*(b)]$$

$$Ax^* + By^*(b) \geq b$$

$$\mathbb{E}_b\big[Ax^* + By^*(b)\big] \geq \mathbb{E}_b[b]$$

$$Ax^* + B\mathbb{E}_b[y^*(b)] \geq b_0$$

$\mathbb{E}_b[y^*(b)]$ is a feasible solution for $b_0 \Rightarrow d^T y^*(b_0) \leq d^T \mathbb{E}_b[y^*(b)]$

$$z_{\mathsf{Rob}} \leq \tau \cdot z_{\mathsf{Stoch}} = \left(1 + \frac{\rho}{s}\right) z_{\mathsf{Stoch}}$$

Our Results: Cost, RHS uncertainty

Stochastic (zStoch)

$$\begin{aligned} \min c^T x + \mathbb{E}_{(b,d)}[d^T y(b,d)] & \\ Ax + By(b,d) &\geq b, \ \forall (b,d) \in \mathcal{U} \\ x, y(b,d) &\in \mathbb{R}^n_+ \end{aligned}$$

Adaptive (zAdapt)

$$\begin{aligned} \min c^T x + \max_{(b,d)} \ d^T y(b,d) & \\ Ax + By(b,d) &\geq b, \ \forall (b,d) \in \mathcal{U} \\ x, y(b,d) &\in \mathbb{R}^n_+ \end{aligned}$$

Uncertainty Set (U) (Cost and RHS)	Stochasticity Gap zRob/zStoch	Adaptability Gap zRob/zAdapt
Symmetric (s=1)		
General (1/n ≤ s ≤ 1)		

Assume: $\mathbf{E}_{b,d}[(b,d)] = (\bar{b},\bar{d})$ where $(\bar{b},\bar{d})$ is the **point of symmetry**

Our Results: Cost, RHS uncertainty

Stochastic (zStoch)

$$\begin{aligned} \min c^T x + \mathbb{E}_{(b,d)}[d^T y(b,d)] & \\ Ax + By(b,d) &\geq b, \ \forall (b,d) \in \mathcal{U} \\ x, y(b,d) &\in \mathbb{R}^n_+ \end{aligned}$$

Adaptive (zAdapt)

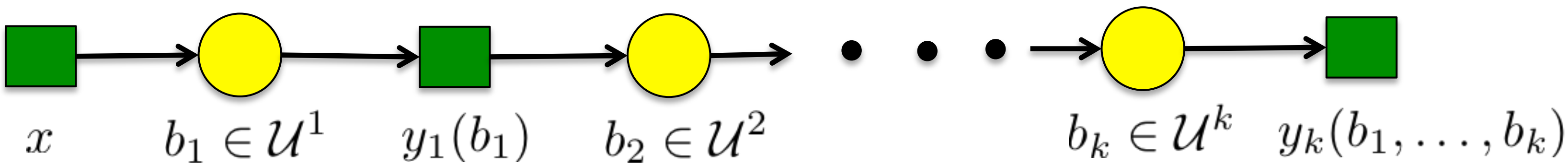
$$\begin{aligned} \min c^T x + \max_{(b,d)} \ d^T y(b,d) & \\ Ax + By(b,d) &\geq b, \ \forall (b,d) \in \mathcal{U} \\ x, y(b,d) &\in \mathbb{R}^n_+ \end{aligned}$$

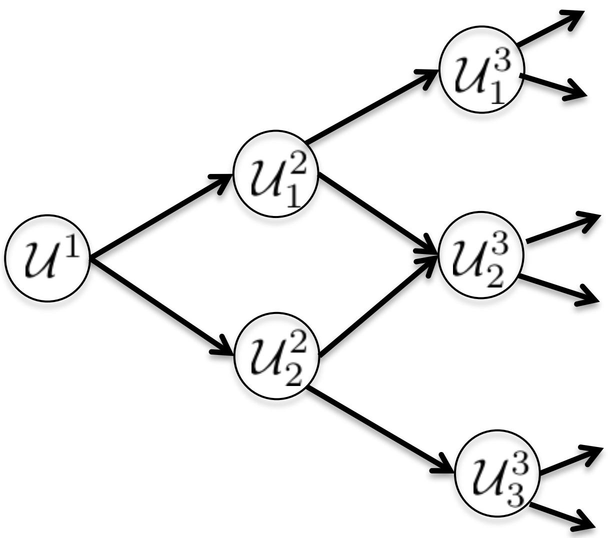
Uncertainty Set (U) (Cost and RHS)	Stochasticity Gap zRob/zStoch	Adaptability Gap zRob/zAdapt
Symmetric (s=1)	$\Omega(m)$	$(1+\rho)^2 \leq 4$
General ($1/n \leq s \leq 1$)	$\Omega(m)$	$(1+\rho/s)^2$

Assume: $E_{b,d}[(b,d)] = (\bar{b}, \bar{d})$ where $(\bar{b}, \bar{d})$ is the **point of symmetry**

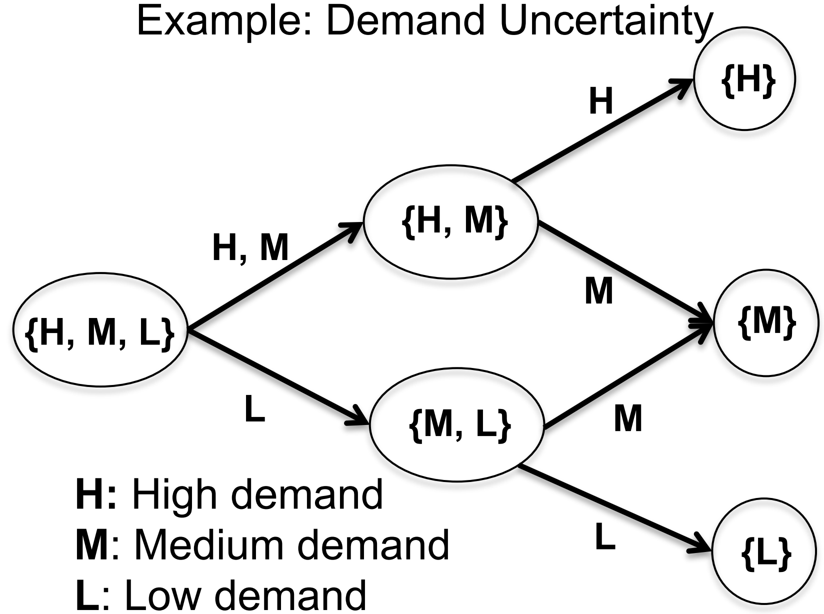
Multi-Stage Problems

Multi-Stage Stochastic Model





Example: Demand Uncertainty



Static solution is not a good approximation

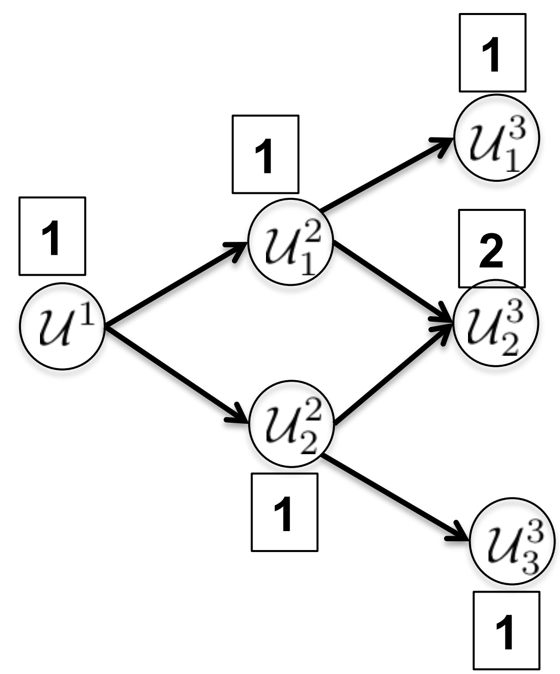
Fully-Adaptive Solution

- Requires optimal decision for each possible scenario
- Uncountable set of scenarios (typically)
- Suffers from the curse of dimensionality
- PSPACE hard to compute in general

Finitely Adaptive Solution

- Partition the scenarios into a small number of sets
- Compute a static solution for each set of scenarios in the partition
- Finite (small) number of solutions in each stage

- partition the scenarios according to the realized paths in the uncertainty network
- Number of paths is finite (small)
- In each stage k, compute a solution for each path from stage 1 to stage k
- For any path P, the solution is feasible for all possible parameter realizations on P



Performance of Finitely Adaptive Solution

Theorem 2 *Let* $\rho = \max \ \rho(\mathcal{U})$ *and* $s = \min \ \mathsf{sym}(\mathcal{U})$ *over all* $\mathcal{U}$. *Also, for all* $\mathcal{U}$, *let,*

$$\mathbb{E}[b] \geq \mathbf{b_0},$$

where $\mathbf{b_0}$ *is the point of symmetry of* $\mathcal{U}$. *Then,*

$$\textit{Cost of an optimal finitely adaptable solution} \leq \left(1 + \frac{\rho}{s}\right) z_{\mathsf{Stoch}}\ .$$

- **Finitely Adaptive** **solution is a** **good** **approximation of the multi-stage stochastic problem**
- **Performance bound = 2** **for uncertainty sets with symmetry = 1**

Bounds for Multi-stage Problems

Uncertainty Set (U) (RHS)	Stochasticity Gap zRob/zStoch	Adaptability Gap zRob/zAdapt
Symmetric (s=1)	$(1+\rho) \leq 2$	$(1+\rho) \leq 2$
General ($1/n< s \leq 1$)	$(1+\rho/s)$	$(1+\rho/s)$

Finitely Adaptive Solution: Multi-stage Adaptive Optimization Problem

Theorem 3 *Consider the adaptive problem with both rhs and cost uncertainty, i.e., both* $\mathbf{b}, \mathbf{d}$ *are uncertain. Let* $\rho = \max\ \rho(\mathcal{U})$ *and* $s = \min\ \mathsf{sym}(\mathcal{U})$ *over all* $\mathcal{U}$. *Then,*

$$\textit{Cost of an optimal finitely adaptable solution} \leq \left(1 + \frac{\rho}{s}\right)^2 z_{\mathsf{Adapt}}\ .$$

- **Finitely Adaptive solution is a good approximation of the multi-stage adaptive problem with both rhs and cost uncertainty**
- **Performance bound ≤ 4 for uncertainty sets with symmetry = 1**
- **Finitely adaptive solution is not a good approximation for the corresponding stochastic problem**

Conclusions

- Choose uncertainty sets carefully
- Criteria: Tractability and Symmetry
- Finite Adaptability, which humans heuristically use is **near optimal** if the uncertainty set is symmetric, that is reasonable known unknowns

15.094J: Robust Modeling, Optimization, Computation

Lecture 12: Affinely Adaptive Optimization

Outline

Motivation

- Affine policies have strong empirical performance.
- Under what circumstances affine policies are optimal?
- How suboptimal are they?
- How can we improve them?

Witnesses of robustness

- AO:

$$
\begin{aligned}
z_{Adapt}(\mathcal{U}) = \min \ & c^T x + \max_{b \in \mathcal{U}} d^T y(b) \\
& Ax + By(b) \geq b, \ \forall b \in \mathcal{U} \\
& x, y(b) \geq 0,
\end{aligned}
$$

- Suppose $x^*, y^*(b)$ for all $b \in \mathcal{U}$ is an optimal solution of AO, where the uncertainty set $\mathcal{U}$ is a polytope. Let $b^1, \ldots, b^K$ be the extreme points of $\mathcal{U}$. Then, the worst case cost is achieved at some extreme point, i.e.,

$$
\max_{b \in \mathcal{U}} d^T y^*(b) = \max_{j=1,\ldots,K} d^T y^*(b^j).
$$

Proof

- $\{b^1, \dots, b^K\} \subseteq \mathcal{U}$:

$$\max_{b \in \mathcal{U}} d^T y^*(b) \geq \max_{j=1,\dots,K} d^T y^*(b^j).$$

- For the sake of contradiction, suppose

$$\max_{b \in \mathcal{U}} d^T y^*(b) > \max_{j=1,\dots,K} d^T y^*(b^j).$$

  Let $\hat{b} = \text{argmax}\{d^T y^*(b) \mid b \in \mathcal{U}\}$, such that $\hat{b} \notin \{b^1, \dots, b^K\}$.

- Therefore,

$$d^T y^*(\hat{b}) > \max_{j=1,\dots,K} d^T y^*(b^j).$$

- Since $\hat{b} \in \mathcal{U}$, $\hat{b} = \sum_{j=1}^K \alpha_j \cdot b^j$, where $\alpha_j \geq 0$ for all $j = 1, \dots, K$ and $\alpha_1 + \dots + \alpha_K = 1$.

Proof, continued

- Consider the solution: $\hat{y}(\hat{b}) = \sum_{j=1}^{K} \alpha_j \cdot y^*(b^j)$.
- $\hat{y}(\hat{b})$ is feasible for $\hat{b}$ as,

$$Ax^* + B\hat{y}(\hat{b}) = A\left(\sum_{j=1}^{K} \alpha_j\right) x^* + B\left(\sum_{j=1}^{K} \alpha_j \cdot y^*(b^j)\right) =$$

$$\sum_{j=1}^{K} \alpha_j \cdot Ax^* + \sum_{j=1}^{K} \alpha_j \cdot By^*(b^j) = \sum_{j=1}^{K} \alpha_j \cdot \left(Ax^* + By^*(b^j)\right) \geq \sum_{j=1}^{K} \alpha_j \cdot b^j = \hat{b},$$

- Objective function value:

$$\begin{aligned}
d^T\hat{y}(\hat{b}) &= d^T\left(\sum_{j=1}^{K} \alpha_j \cdot y^*(b^j)\right) = \sum_{j=1}^{K} \alpha_j \cdot d^T y^*(b^j) \\
&\leq \sum_{j=1}^{K} \alpha_j \cdot \max\{d^T y^*(b^k) \mid k = 1, \ldots, K\} \\
&= \max\{d^T y^*(b^k) \mid k = 1, \ldots, K\} \\
&< d^T y^*(\hat{b}).
\end{aligned}$$

- This implies that $y^*(\hat{b})$ is not an optimal solution for $\hat{b}$; a contradiction.

Optimality of affine policies over the simplex

- For AO with
$$\mathcal{U} = \text{conv}(b^1, \ldots, b^{m+1}),$$
- $b^j \in \mathbb{R}^m_+$ for all $j = 1, \ldots, m$ such that $b^1, \ldots, b^{m+1}$ are affinely independent.
- Then, there is an optimal two-stage solution $\hat{x}, \hat{y}(b)$ for all $b \in \mathcal{U}$ such that $\hat{y}(b)$ is an affine function of b, i.e., for all $b \in \mathcal{U}$,
$$\hat{y}(b) = Pb + q,$$

Proof

- $x^*, y^*(b)$ optimal for AO.

$$Q = \left[\left(b^1 - b^{m+1}\right), \ldots, \left(b^m - b^{m+1}\right)\right]$$

$$Y = \left[\left(y^*(b^1) - y^*(b^{m+1})\right), \ldots, \left(y^*(b^m) - y^*(b^{m+1})\right)\right]$$

- Since $b^1, \ldots, b^{m+1}$ are affinely independent, $(b^1 - b^{m+1}), \ldots, (b^m - b^{m+1})$ are linearly independent.
- Q is a full-rank matrix and thus, invertible. For any $b \in \mathcal{U}$:

$$\hat{y}(b) = YQ^{-1}\left(b - b^{m+1}\right) + y^*(b^{m+1}).$$

- Since $b \in \mathcal{U}$, $b = \sum_{j=1}^{m+1} \alpha_j b^j$, where $\alpha_j \geq 0$ for all $j = 1, \ldots, m+1$ and $\alpha_1 + \ldots + \alpha_{m+1} = 1$.

Proof, continued

- We have

$$b = \sum_{j=1}^{m} \alpha_j b^j + \left(1 - \sum_{j=1}^{m} \alpha_j\right) b^{m+1} = \sum_{j=1}^{m} \alpha_j \left(b^j - b^{m+1}\right) + b^{m+1}$$

$$= Q \cdot \alpha + b^{m+1}, \ \alpha = (\alpha_1, \ldots, \alpha_m)^T$$

- Since Q is invertible, $Q^{-1}\left(b - b^{m+1}\right) = \alpha$, and thus

$$\begin{aligned} \hat{y}(b) &= Y \cdot \alpha + y^*(b^{m+1}) \\ &= \sum_{j=1}^{m} \alpha_j \left(y^*(b^j) - y^*(b^{m+1})\right) + y^*(b^{m+1}) \\ &= \sum_{j=1}^{m} \alpha_j y^*(b^j) + \left(1 - \sum_{j=1}^{m} \alpha_j\right) y^*(b^{m+1}) \\ &= \sum_{j=1}^{m+1} \alpha_j y^*(b^j) \end{aligned}$$

Proof, continued

- As before, $\hat{y}(b)$ is a feasible solution for all $b \in \mathcal{U}$.
- Since the worst case occurs at one of the extreme points of $\mathcal{U}$,

$$z_{Adapt}(\mathcal{U}) = \max_{b \in \mathcal{U}} \left(c^T x^* + d^T y^*(b)\right) = \max_{j=1,\ldots,m+1} \left(c^T x^* + d^T y^*(b^j)\right).$$

- Note that $\hat{y}(b^j) = y^*(b^j)$ for all $j = 1, \ldots, m+1$. Therefore,

$$\begin{aligned} \max_{b \in \mathcal{U}} \left(c^T x^* + d^T \hat{y}(b)\right) &= \max_{j=1,\ldots,m+1} \left(c^T x^* + d^T \hat{y}(b^j)\right) \\ &= \max_{j=1,\ldots,m+1} \left(c^T x^* + d^T y^*(b^j)\right) \\ &= z_{Adapt}(\mathcal{U}). \end{aligned}$$

Suboptimality of Affine Policies for Uncertainty Sets with $(m+2)$ Extreme Points

- Data $c = 0,\ d = (1, \dots, 1)',\ A = 0$, and for all $j = 1, \dots, m$

$$B_{ij} = \begin{cases} 1 & \text{if } i = j, \\ \frac{1}{\sqrt{m}} & \text{otherwise} \end{cases}$$

- $\mathcal{U} = \operatorname{conv}\left(\left\{b^0, b^1, \dots, b^{m+2}\right\}\right),\ b^0 = 0,\ b^j = e_j,\ \forall j = 1, \dots, m$

$$b^{m+1} = \left(\underbrace{\frac{1}{\sqrt{m}}, \dots, \frac{1}{\sqrt{m}}}_{m/2}, \underbrace{0, \dots, 0}_{m/2}\right),\ b^{m+2} = \left(\underbrace{0, \dots, 0}_{m/2}, \underbrace{\frac{1}{\sqrt{m}}, \dots, \frac{1}{\sqrt{m}}}_{m/2}\right)$$

- Given any $\delta > 0$, consider AO with data and uncertainty set $\mathcal{U}$ as above. Then,

$$z_{Aff}(\mathcal{U}) > (2 - \delta) \cdot z_{Adapt}(\mathcal{U}).$$

A Large Gap Example for Affine Policies

- Data $n_1 = n_2 = m$, $m^\delta > 200$, $c = 0$, $d = (1, \ldots, 1)^T$, $A = 0$,

$$B_{ij} = \begin{cases} 1 & \text{if } i = j, \\ \theta_0 & \text{otherwise} \end{cases}$$

- $\mathcal{U} = \text{conv}\left(\left\{b^0, b^1, \ldots, b^N\right\}\right)$, $\theta_0 = \frac{1}{m^{(1-\delta)/2}}$, $r = \lceil m^{1-\delta} \rceil$, $N = \binom{m}{r} + m + 2$ and

$$\begin{aligned} b^0 &= 0 \\ b^j &= e_j, \ \forall j = 1, \ldots, m \\ b^{m+1} &= \frac{1}{\sqrt{m}} \cdot e \\ b^{m+2} &= \theta_0 \cdot \left(\underbrace{1, \ldots, 1}_{r}, 0, \ldots, 0 \right), \end{aligned}$$

A Large Gap Example for Affine Policies, continued

- Exactly r coordinates are non-zero, each equal to θ_0.
- Extreme points b^j, $j \geq m+3$ are permutations of the non-zero coordinates of b^{m+2}.
- $\mathcal{U}$ has exactly $\binom{m}{r}$ extreme points of the form of b^{m+2}.
- All the non-zero extreme points of $\mathcal{U}$ are roughly on the boundary of the unit hypersphere centered at zero.
- Theorem: For the instance above with uncertainty set $\mathcal{U}$,

$$z_{Aff}(\mathcal{U}) = \Omega\left(m^{1/2-\delta}\right) \cdot z_{Adapt}(\mathcal{U}),$$

for any given $\delta > 0$.

Performance Guarantee for Affine Policies

- Consider AAO with $\mathcal{U} \subseteq \mathbb{R}^m_+$ convex, compact and full-dimensional and $A \geq 0$.
- Then
$$z_{Aff}(\mathcal{U}) \leq 3\sqrt{m} \cdot z_{Adapt}(\mathcal{U}),$$
- Worst case cost of an optimal affine policy is at most $3\sqrt{m}$ times the worst case cost of an optimal fully adaptable solution.
- In general,
$$z_{Aff}(\mathcal{U}) \leq 4\sqrt{m} \cdot z_{Adapt}(\mathcal{U}),$$
- Full characterization of AAO performance: $z_{Aff}(\mathcal{U}) = \Theta(\sqrt{m}) \cdot z_{Adapt}(\mathcal{U})$,
- Contrast with $z_{Rob}(\mathcal{U}) = \Theta(m) \cdot z_{Adapt}(\mathcal{U})$,

Single Echelon Case

- $x_{k+1} = x_k + u_k - w_k$
- x_k : inventory at period k
- w_k : unknown, bounded demands from customers, $w_k \in [\underline{w}_k, \overline{w}_k]$
- u_k : replenishment orders; no lead-time, but capacities, $u_k \in \left[L_k, U_k\right]$
- Linear ordering costs + any convex inventory cost $h_k(x_k)$

$$\mathcal{C}_k(u_k, x_k) = c_k \, u_k + h_k(x_k)$$

Single Echelon Case

- $x_{k+1} = x_k + u_k - w_k$
- x_k : inventory at period k
- w_k : unknown, bounded demands from customers, $w_k \in [\underline{w}_k, \overline{w}_k]$
- u_k : replenishment orders; no lead-time, but capacities, $u_k \in \left[L_k, U_k\right]$
- Linear ordering costs + any convex inventory cost $h_k(x_k)$

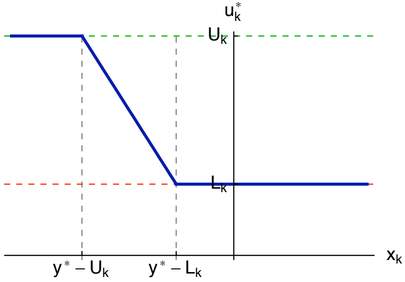
$$\mathcal{C}_k(u_k, x_k) = c_k\, u_k + h_k(x_k)$$

- Typical inventory example: holding and backlogging costs

$$h_k(x_k) = H_k \cdot \max(x_k, 0) + B_k \cdot \max(-x_k, 0)$$

Optimal Policies by Dynamic Programming

- (Modified) Base-stock policies optimal
  - Kasugai Kasegai (1960, 1961)



Optimality of Affine Policies in the Demands.

Theorem (Bertsimas, Iancu, Parrilo 2009a)

Ordering policies that are affine in the history of demands are optimal. In fact, for every time step $k = 1, \ldots, T$, the following quantities exist:

Optimality of Affine Policies in the Demands.

Theorem (Bertsimas, Iancu, Parrilo 2009a)

Ordering policies that are affine in the history of demands are optimal. In fact, for every time step $k = 1, \ldots, T$, the following quantities exist:

- *an affine ordering policy,* $u_k(w_{[k]}) \stackrel{def}{=} u_{k,0} + \sum_{t=1}^{k-1} u_{k,t} w_t$,
- *an affine inventory cost,* $z_{k+1}(w_{[k+1]}) \stackrel{def}{=} z_{k+1,0} + \sum_{t=1}^{k} z_{k+1,t} w_t$,

such that the following conditions are obeyed:

Optimality of Affine Policies in the Demands.

Theorem (Bertsimas, Iancu, Parrilo 2009a)

Ordering policies that are affine in the history of demands are optimal. In fact, for every time step $k = 1, \ldots, T$, the following quantities exist:

- *an affine ordering policy,* $u_k(w_{[k]}) \stackrel{def}{=} u_{k,0} + \sum_{t=1}^{k-1} u_{k,t} w_t$,
- *an affine inventory cost,* $z_{k+1}(w_{[k+1]}) \stackrel{def}{=} z_{k+1,0} + \sum_{t=1}^{k} z_{k+1,t} w_t$,

such that the following conditions are obeyed:

- $u_k(w_{[k]}) \in \big[L_k, U_k\big], \quad \forall\, w_{[k]}$
- $z_{k+1}(w_{[k+1]}) \geq h_{k+1}\Big(x_1 + \sum_{t=1}^{k}\big(u_t(w_{[t]}) - w_t\big)\Big), \quad \forall\, w_{[k+1]}$
- $$J_1^\star(x_1) = \max_{w_1,\ldots,w_k}\Big[\sum_{t=1}^{k}\big(c_t \cdot u_t(w_{[t]}) + z_t(w_{[t+1]})\big) + \\ + J_{k+1}^\star\Big(x_1 + \sum_{t=1}^{k}\big(u_t(w_{[t]}) - w_t\big)\Big)\Big]$$

Proof Outline. DP, Induction, Geometry.

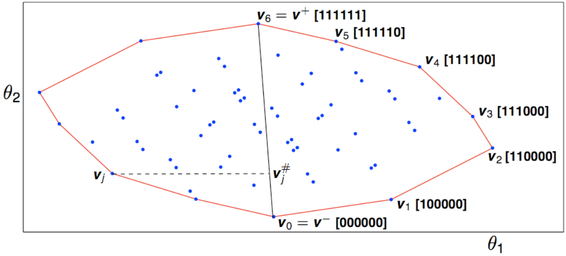
- Forward induction on k
- Assume true $1, \dots, k$. The problem for uncertainties at k is

$$J_{mM} = \max_{(\theta_1, \theta_2) \in \Theta} \left[\theta_1 + J^\star_{k+1}(\theta_2) \right]$$

Proof Outline. DP, Induction, Geometry.

- Forward induction on k
- Assume true $1, \ldots, k$. The problem for uncertainties at k is

$$J_{mM} = \max_{(\theta_1,\theta_2)\in\Theta} \left[\, \theta_1 + J^{\star}_{k+1}(\theta_2) \, \right]$$



Why Is This Relevant?

1. *Computational result*
   For piecewise affine costs (with m_k pieces), must solve a single LOP with $O\left(T^2 \cdot \max_k\{m_k\}\right)$ variables and constraints

2. *Insight*
   Decomposition of demand satisfaction by means of future orders,

3. *Tight existential result*
   E.g., such policies not optimal for $\sum_{t=1}^{k} u_t \in \left[\hat{L}_k, \hat{U}_k\right]$

Extensions : Supply Contracts, Service Level Constraints

- Supply contracts
  - Order bounds L_k, U_k not *fixed*, but part of contract
  - Retailer pays supplier $f(U) \geq 0$, and receives $g(L) \geq 0$ from supplier
  - Retailer decides L, U beforehand (time $k = 0$), and ordering policies u_k

Extensions : Supply Contracts, Service Level Constraints

- Supply contracts
  - Order bounds L_k, U_k not *fixed*, but part of contract
  - Retailer pays supplier $f(U) \geq 0$, and receives $g(L) \geq 0$ from supplier
  - Retailer decides L, U beforehand (time $k = 0$), and ordering policies u_k

Theorem

If f convex and g concave $\Rightarrow$ solve optimally by sub-gradient methods

If f, g also piecewise affine $\Rightarrow$ solve a single LOP

Extensions : Supply Contracts, Service Level Constraints

- Supply contracts
  - Order bounds L_k, U_k not *fixed*, but part of contract
  - Retailer pays supplier $f(U) \geq 0$, and receives $g(L) \geq 0$ from supplier
  - Retailer decides L, U beforehand (time $k = 0$), and ordering policies u_k

Theorem

If f convex and g concave $\Rightarrow$ solve optimally by sub-gradient methods

If f, g also piecewise affine $\Rightarrow$ solve a single LOP

- Can easily accommodate service-level constraints
  - Satisfy 90% of demand upon arrival
  - Never backlog more than P periods

General Multi-Echelon Problem

$$
\begin{aligned}
&\min_{u_1}\Big[\mathcal{C}_1(x_1,u_1)+\max_{w_1}\min_{u_2}\Big[\mathcal{C}_2(x_2,u_2)+\cdots+\max_{w_T}\mathcal{C}_{T+1}(x_{T+1})\Big]\ldots\Big]\Big],\\
&x_{k+1}=A_k\,x_k+B_k\,u_k-w_k,\\
&\quad f_k\ge D_k\,x_k+E_k\,u_k, \qquad\qquad k\in\{1,\ldots,T\}.
\end{aligned}
$$

- Affine policies *not* optimal

General Multi-Echelon Problem

$$\min_{u_1}\left[\mathcal{C}_1(x_1,u_1)+\max_{w_1}\min_{u_2}\left[\mathcal{C}_2(x_2,u_2)+\cdots+\max_{w_T}\mathcal{C}_{T+1}(x_{T+1})\right]\ldots\right],$$
$$x_{k+1}=A_k\,x_k+B_k\,u_k-w_k,$$
$$f_k\geq D_k\,x_k+E_k\,u_k,\qquad\qquad k\in\{1,\ldots,T\}.$$

- Affine policies *not* optimal
- Consider polynomial policies in $w_{[k]}\stackrel{\text{def}}{=}\left[w_1,w_2,\ldots,w_{k-1}\right]$

General Multi-Echelon Problem

$$\min_{u_1}\Big[\mathcal{C}_1(x_1,u_1)+\max_{w_1}\min_{u_2}\Big[\mathcal{C}_2(x_2,u_2)+\cdots+\max_{w_T}\mathcal{C}_{T+1}(x_{T+1})\Big]\ldots\Big]\Big],$$
$$x_{k+1}=A_k\,x_k+B_k\,u_k-w_k,$$
$$f_k\geq D_k\,x_k+E_k\,u_k,\qquad k\in\{1,\ldots,T\}.$$

- Affine policies *not* optimal
- Consider polynomial policies in $w_{[k]}\stackrel{\text{def}}{=}\left[w_1,w_2,\ldots,w_{k-1}\right]$
  - Example: degree $d=2$, $w_{[3]}=(w_1,w_2)$

$$u_3(w_{[3]})=\ell_0+\ell_1 w_1+\ell_2 w_2+\ell_{1,1}w_1^2+\ell_{1,2}w_1\,w_2+\ell_{2,2}w_2^2$$

Why Polynomials? [Bertsimas, Iancu, Parrilo 2009b]

1. Natural extension of affine case
2. Good approximation when optimal policies are continuous
3. Little burden on modeller : only choice of polynomial degree d
4. Can provide semidefinite programming relaxation
   - $T\left(\max_k r_k + \max_k m_k\right)$ SDP constraints, each of size $\binom{n_w\ T+d}{d}$
   - Solvable by interior-point methods
5. Degree d controls accuracy vs. computation trade-off

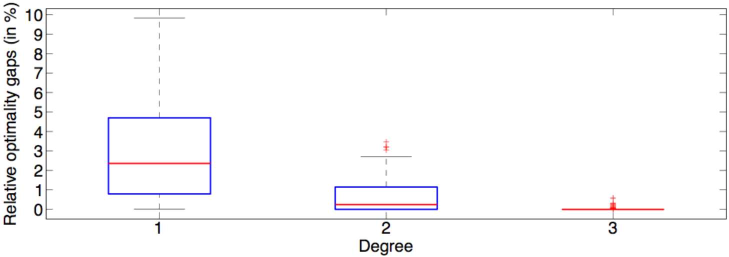
Single-echelon with Cumulative Orders

Single-echelon with Cumulative Orders

Relative optimality gaps (in %) for polynomial policies

	Degree $d = 1$					Degree $d = 2$					Degree $d = 3$				
T	avg	std	mdn	min	max	avg	std	mdn	min	max	avg	std	mdn	min	max
4	2.84	2.41	2.18	0.02	9.76	0.75	0.85	0.47	0.00	3.79	0.03	0.12	0.00	0.00	0.91
5	2.82	2.29	2.52	0.04	11.22	0.62	0.71	0.39	0.00	3.92	0.02	0.09	0.00	0.00	0.56
6	3.09	2.63	2.36	0.01	9.82	0.69	0.89	0.25	0.00	3.47	0.03	0.10	0.00	0.00	0.59
7	3.25	2.95	2.58	0.13	15.00	0.83	0.99	0.43	0.00	4.79	0.06	0.17	0.00	0.00	0.93
8	3.66	3.29	2.69	0.03	18.36	1.06	1.17	0.74	0.00	5.81	0.10	0.17	0.00	0.00	0.99
9	2.93	2.78	2.12	0.05	11.56	0.80	0.86	0.55	0.00	3.39	0.07	0.13	0.00	0.00	0.61
10	3.44	3.60	2.09	0.00	18.20	0.76	1.16	0.26	0.00	5.76	0.05	0.12	0.00	0.00	0.74

Polynomial policies for $T = 6$

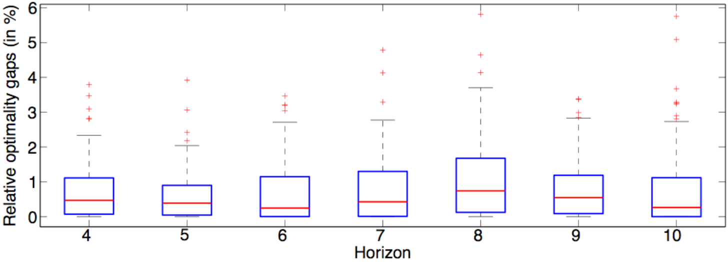


Single-echelon with Cumulative Orders

Relative optimality gaps (in %) for polynomial policies

	Degree $d = 1$					Degree $d = 2$					Degree $d = 3$				
T	avg	std	mdn	min	max	avg	std	mdn	min	max	avg	std	mdn	min	max
4	2.84	2.41	2.18	0.02	9.76	0.75	0.85	0.47	0.00	3.79	0.03	0.12	0.00	0.00	0.91
5	2.82	2.29	2.52	0.04	11.22	0.62	0.71	0.39	0.00	3.92	0.02	0.09	0.00	0.00	0.56
6	3.09	2.63	2.36	0.01	9.82	0.69	0.89	0.25	0.00	3.47	0.03	0.10	0.00	0.00	0.59
7	3.25	2.95	2.58	0.13	15.00	0.83	0.99	0.43	0.00	4.79	0.06	0.17	0.00	0.00	0.93
8	3.66	3.29	2.69	0.03	18.36	1.06	1.17	0.74	0.00	5.81	0.10	0.17	0.00	0.00	0.99
9	2.93	2.78	2.12	0.05	11.56	0.80	0.86	0.55	0.00	3.39	0.07	0.13	0.00	0.00	0.61
10	3.44	3.60	2.09	0.00	18.20	0.76	1.16	0.26	0.00	5.76	0.05	0.12	0.00	0.00	0.74

Performance of quadratic policies

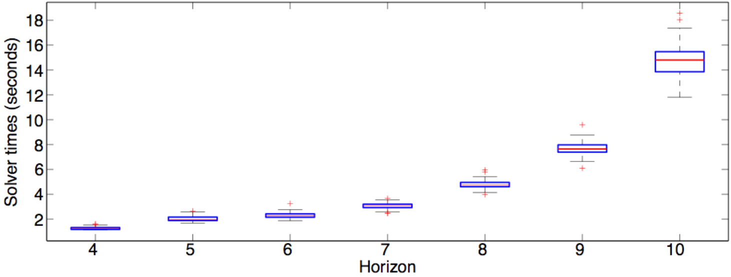


Single-echelon with Cumulative Orders

Relative optimality gaps (in %) for polynomial policies

	Degree $d = 1$					Degree $d = 2$					Degree $d = 3$				
T	avg	std	mdn	min	max	avg	std	mdn	min	max	avg	std	mdn	min	max
4	2.84	2.41	2.18	0.02	9.76	0.75	0.85	0.47	0.00	3.79	0.03	0.12	0.00	0.00	0.91
5	2.82	2.29	2.52	0.04	11.22	0.62	0.71	0.39	0.00	3.92	0.02	0.09	0.00	0.00	0.56
6	3.09	2.63	2.36	0.01	9.82	0.69	0.89	0.25	0.00	3.47	0.03	0.10	0.00	0.00	0.59
7	3.25	2.95	2.58	0.13	15.00	0.83	0.99	0.43	0.00	4.79	0.06	0.17	0.00	0.00	0.93
8	3.66	3.29	2.69	0.03	18.36	1.06	1.17	0.74	0.00	5.81	0.10	0.17	0.00	0.00	0.99
9	2.93	2.78	2.12	0.05	11.56	0.80	0.86	0.55	0.00	3.39	0.07	0.13	0.00	0.00	0.61
10	3.44	3.60	2.09	0.00	18.20	0.76	1.16	0.26	0.00	5.76	0.05	0.12	0.00	0.00	0.74

Solver times for quadratic policies



Serial Supply Chain

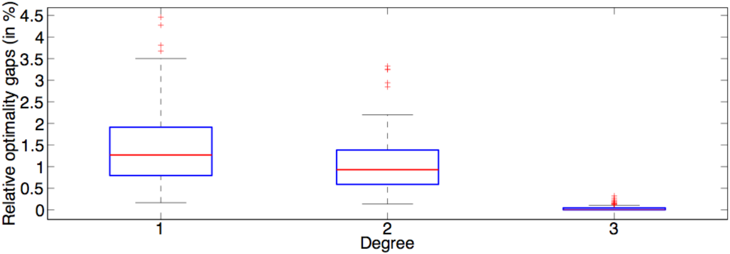
Serial supply chain

Serial Supply Chain

Relative gaps (in %) for the serial supply chain example

	Degree $d = 1$					Degree $d = 2$					Degree $d = 3$				
J	avg	std	mdn	min	max	avg	std	mdn	min	max	avg	std	mdn	min	max
2	1.87	1.48	1.47	0.00	8.27	1.38	1.16	1.11	0.00	6.48	0.06	0.14	0.01	0.00	0.96
3	1.47	0.89	1.27	0.16	4.46	1.08	0.68	0.93	0.14	3.33	0.04	0.06	0.00	0.00	0.32
4	1.14	2.46	0.70	0.05	24.63	0.67	0.53	0.53	0.01	2.10	0.04	0.07	0.00	0.00	0.38
5	0.35	0.37	0.21	0.03	1.85	0.27	0.32	0.15	0.00	1.59	0.02	0.03	0.00	0.00	0.15

Polynomial policies for $J = 3$ echelons.



Conclusions

- Demand-feedback policies for multi-period, multi-echelon problems

Conclusions

- Demand-feedback policies for multi-period, multi-echelon problems
- Single-echelon case:
  - Affine policies are optimal
  - Newsvendor costs $\Rightarrow$ a single LOP
  - Supply contracts - capacity pre-commitment problem

Conclusions

- Demand-feedback policies for multi-period, multi-echelon problems
- Single-echelon case:
  - Affine policies are optimal
  - Newsvendor costs $\Rightarrow$ a single LOP
  - Supply contracts - capacity pre-commitment problem
- Multi-echelon case:
  - Framework to compute polynomial policies - solve a single SDOP
  - Polynomial degree d controls performance-computation trade-off
  - Perform well in several inventory examples

15.094J: Robust Modeling, Optimization, Computation

Lecture 13: RO in Inventory Theory

Outline

1. Single station
2. Series systems
3. General Supply chains
4. Summary and Conclusions

Single station

- State x_k: stock available at the beginning of the kth period
- Control u_k: stock ordered at the beginning of the kth period
- Randomness w_k: demand during the kth period
- Dynamics: $x_{k+1} = x_k + u_k - w_k$
- Inventory Costs: $\max(hx_{k+1}, -px_{k+1})$
- Fixed costs: $cu_k + K1_{\{u_k>0\}}$.

Modeling Randomness

- $z_k = (w_k - \bar{w}_k)/\hat{w}_k \in [-1, 1]$.
- Uncertainty budget $\sum_{i=0}^{k} |z_i| \leq \Gamma_k$.
- Γ_k: budget of uncertainty controlling tradeoff between robustness and optimality.

The nominal model

- Goal is to solve:

$$\begin{aligned} \min \quad & \sum_{t=0}^{T-1} (c\, u_t + K\, 1_{\{u_t>0\}} + \max(h\, \overline{x}_{t+1}, -p\, \overline{x}_{t+1})) \\ \text{s.t.} \quad & u_t \geq 0 \ \forall t. \end{aligned}$$

- Can be formulated as a LO or MIO by replacing $\max(h\, \overline{x}_{t+1}, -p\, \overline{x}_{t+1})$ by new variable y_t
- Use closed-form expression $\overline{x}_{t+1} = x_0 + \sum_{s=0}^{t}(u_s - \overline{w}_s)$
- Model $1_{\{u_t>0\}}$ by $v_t \in \{0, 1\}$ with $0 \leq u_t \leq M\, v_t$.

The nominal model continued

$$\begin{aligned}
\min \quad & \sum_{t=0}^{T-1} (c\, u_t + K\, v_t + y_t) \\
\text{s.t.} \quad & y_t \geq h \left(x_0 + \sum_{s=0}^{t} (u_s - \overline{w}_s) \right), && \forall t, \\
& y_t \geq -p \left(x_0 + \sum_{s=0}^{t} (u_s - \overline{w}_s) \right), && \forall t, \\
& 0 \leq u_t \leq M\, v_t, \;\; v_t \in \{0, 1\}, && \forall t.
\end{aligned}$$

LO if no fixed costs, MIO if fixed costs.

The robust formulation

- Add uncertainty to the nominal model.
- Example: holding constraint $y_t \geq h\left(x_0 + \sum_{s=0}^{t}(u_s - \overline{w}_s)\right)$.
- Robust approach: at y_t and $u_0, \dots, u_t$ given, constraint must be feasible for any demand in the uncertainty set:

$$y_t \geq h\left(x_0 + \sum_{s=0}^{t}(u_s - \overline{w}_s - \widehat{w}_s\, z_s)\right)$$

$$\forall z \in Z = \left\{ |z_s| \leq 1 \ \forall s, \sum_{s=0}^{t} |z_s| \leq \Gamma_t \right\}.$$

The robust formulation, continued

In particular, it must be feasible for the demand yielding the greatest value of the right-hand side:

$$y_t \geq h\left(x_0 + \sum_{s=0}^{t}(u_s - \overline{w}_s\right) + \max_{z \in Z}(-h)\sum_{s=0}^{t}\widehat{w}_s\, z_s$$

Auxiliary problem:

$$\begin{aligned} \max \quad & \sum_{s=0}^{t} \widehat{w}_s \cdot (-z_s) \\ \text{s.t.} \quad & \sum_{s=0}^{t} |z_s| \leq \Gamma_t, \\ & |z_s| \leq 1, \ \forall s \leq t, \end{aligned} \qquad \Rightarrow \qquad \begin{aligned} \max \quad & \sum_{s=0}^{t} \widehat{w}_s z'_s \\ \text{s.t.} \quad & \sum_{s=0}^{t} z'_s \leq \Gamma_t, \\ & 0 \leq z'_s \leq 1, \ \forall s \leq t. \end{aligned}$$

The robust formulation, continued

By strong duality:

$$\begin{aligned}
\max \quad & \sum_{s=0}^{t} \widehat{w}_s z'_s & \Rightarrow \quad \min \quad & q_t \Gamma_t + \sum_{s=0}^{t} r_{st} \\
\text{s.t.} \quad & \sum_{s=0}^{t} z'_s \le \Gamma_t, & \text{s.t.} \quad & q_t + r_{st} \ge \widehat{w}_s, \ \forall s, \\
& 0 \le z'_s \le 1, \ \forall s \le t, & & q_t \ge 0, \ r_{st} \ge 0, \ \forall s \le t.
\end{aligned}$$

The holding constraint becomes:

$$y_t \ge h\left(x_0 + \sum_{s=0}^{t}(u_s - \overline{w}_s)\right) + h \cdot \min_{(q,r)\in Q}\left(q_t \Gamma_t + \sum_{s=0}^{t} r_{st}\right)$$

Enough to find (q, r) feasible: constraint is linear.

LO or MIO

$$
\begin{array}{lll}
\min & \displaystyle\sum_{t=0}^{T-1} (c\, u_t + K\, v_t + y_t) & \\
\text{s.t.} & y_t \geq h\left(x_0 + \displaystyle\sum_{s=0}^{t} (u_s - \overline{w}_s)\right) + h\, A_t, & \forall t, \\
& y_t \geq -p\left(x_0 + \displaystyle\sum_{s=0}^{t} (u_s - \overline{w}_s)\right) + p\, A_t, & \forall t, \\
& A_t = q_t\, \Gamma_t + \displaystyle\sum_{s=0}^{t} r_{st}, & \forall t, \\
& q_t + r_{st} \geq \widehat{w}_s,\;\; q_t \geq 0, r_{st} \geq 0, & \forall t,\;\; s \leq t, \\
& 0 \leq u_t \leq M v_t,\;\; v_t \in \{0, 1\}, & \forall t.
\end{array}
$$

Properties

- No fixed costs: Robust problem optimal ordering policy is also (S, S), or basestock, i.e., there exists a threshold sequence (S_k) such that, at each time period k, it is optimal to order $S_k - x_k$ if $x_k < S_k$ and 0, otherwise. S_k given in closed form.
- Fixed costs, optimal policy for robust problem is (s, S), i.e., there exists a threshold sequence (s_k, S_k) such that, at each time period k, it is optimal to order $S_k - x_k$ if $x_k < s_k$ and 0 otherwise, with $s_k \leq S_k$.
- Constrast to the stochastic case.

Budget of uncertainty

- Expected cost if distribution is known:

$$c \sum_{t=0}^{T-1} u_t + K \sum_{t=0}^{T-1} v_t + \sum_{t=0}^{T-1} E \max(h\, x_{t+1}, -p\, x_{t+1}).$$

- Assume that only first two moments are known. We want an upper bound on

$$E \max(h\, x_{t+1}, -p\, x_{t+1}) = h\, E(x_{t+1}) + (h+p)\, E \max(0, -x_{t+1})$$

Budget of uncertainty, continued

- Bertsimas and Popescu, 2001: $E\max(0, X-a)$

$$\leq \begin{cases} \frac{1}{2}\left(\mu - a + \sqrt{\sigma^2 + (\mu - a)^2}\right), & \text{if } a \geq \frac{\mu}{2} + \frac{\sigma^2}{2\,\mu}, \\ \frac{\mu}{\mu^2+\sigma^2}\left(\sigma^2 - \mu(a-\mu)\right), & \text{otherwise.} \end{cases}$$

- Bound is convex in a.
- Find the budgets of uncertainty that minimize the upper bound.

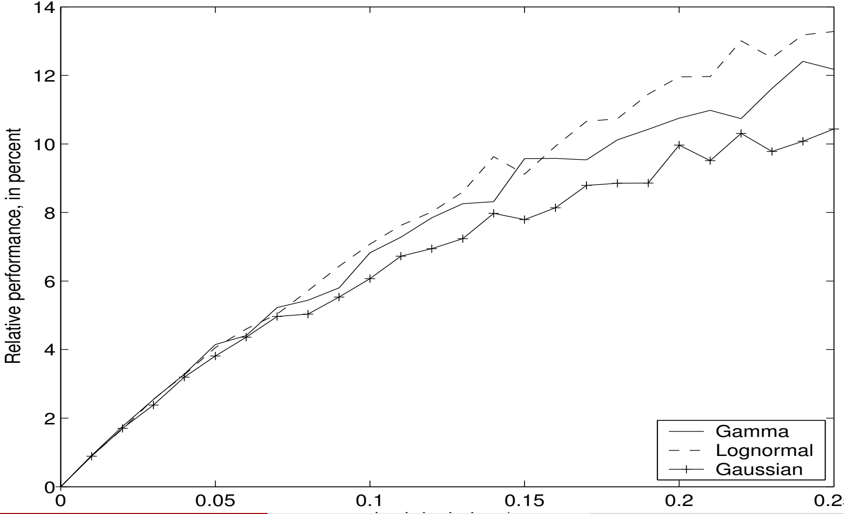
Example

- Goal: compare robust approach with dynamic programming when first two moments of distribution are known.
- Performance measure: $100 \cdot \dfrac{E(DP) - E(ROB)}{E(DP)}$.
- Questions:
  - Does the actual distribution (beyond first two moments) significantly affect performance?
  - What is the impact of the cost parameters?
  - What is the impact of DP assuming a wrong distribution?

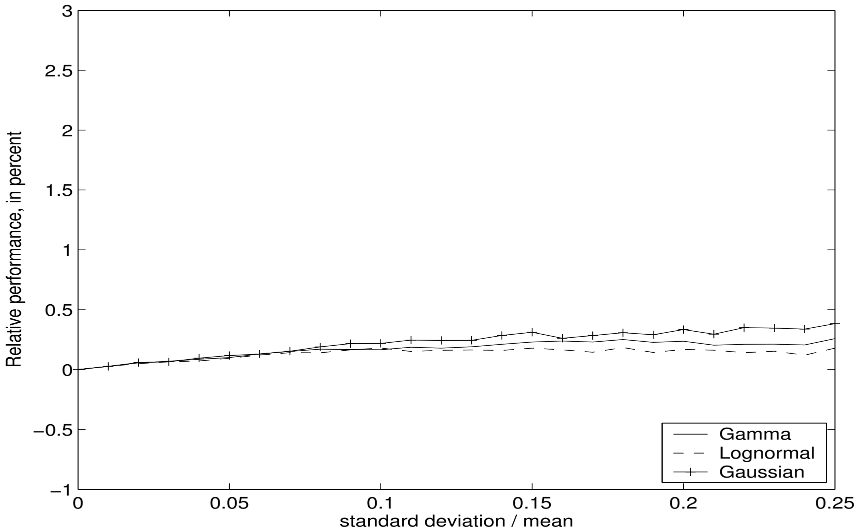
Impact of standard deviation

DP assumes binomial; actual distribution is different

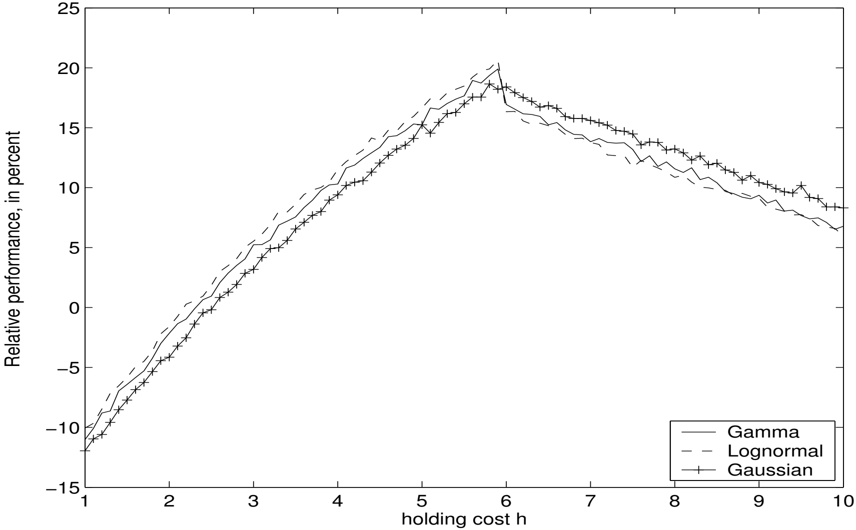
$(c = 1, h = 4, p = 6, \overline{w} = 100, \sigma = 20)$.



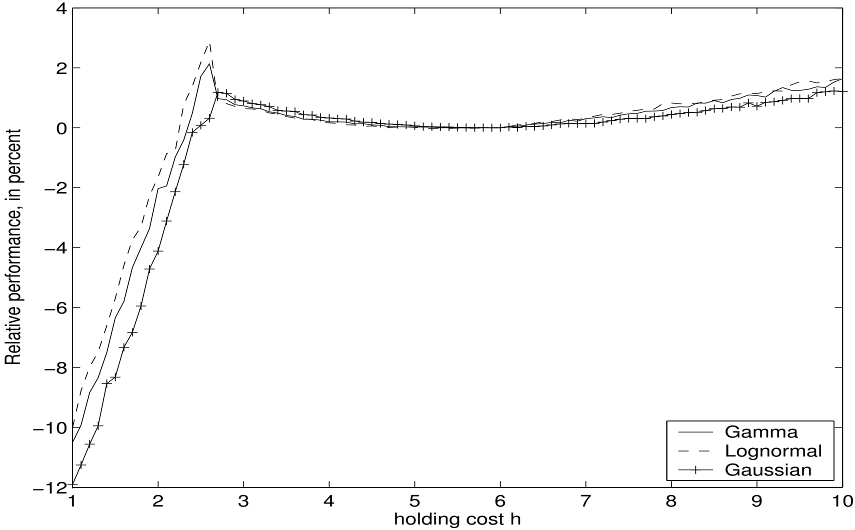
DP assumes almost right distribution (Gaussian)



DP assumes binomial; actual distribution is different



DP assumes right distribution (Gaussian)



Comments

- Robust approach leads to high-quality solutions,
- Outperforms dynamic programming for a wide range of parameters, in particular if assumption on distribution in DP is not accurate,
- It is robust to the actual demand distributions (beyond their first two moments).

Series systems

- Station $k+1$ supplies station k, and the demand at station k is the order made by station $k-1$. Station N is supplied by the outside world and the demand at station 1 is exogenous, subject to randomness. Stock at station k at time t is $I_k(t)$.
- Echelon k is stations 1 to $k \rightarrow X_k(t) = \sum_{j=1}^{k} I_j(t)$.
- Clark and Scarf, 1960: Optimal policy when costs are computed at the echelon level is basestock, when there are no fixed ordering costs except maybe for station N.
- Can the robust approach yield similar theoretical results?

The nominal model

$$
\begin{aligned}
\min \quad & \sum_{k=1}^{N} \sum_{t=0}^{T-1} \left(c_k\, U_k(t) + K_k\, 1_{\{U_k(t)>0\}} \right) \\
& \qquad + \sum_{k=1}^{N} \sum_{t=0}^{T-1} \max \left(h\, \overline{X}_k(t+1), -p\, \overline{X}_k(t+1) \right) \\
\text{s.t.} \quad & \overline{X}_k(t+1) = X_k(0) + \sum_{s=0}^{t} (U_k(s) - \overline{W}_k(s)), \ \forall k, t, \\
& U_k(t) \leq I_{k+1}(t), \ \forall k, t, \\
& U_k(t) \geq 0, \ \forall k, t.
\end{aligned}
$$

Again, LO or MIO.

The robust model

$$
\begin{aligned}
\min \quad & \sum_{k=1}^{N} \sum_{t=0}^{T-1} \left(c_k\, U_k(t) + K_k\, V_k(t) + Y_k(t) \right) \\
\text{s.t.} \quad & \overline{X}_k(t+1) = X_k(0) + \sum_{s=0}^{t} \left(U_k(s) - \overline{W}_k(s) \right), \ \forall k, t, \\
& A(t) = q(t)\, \Gamma(t) + \sum_{s=0}^{t} r(s,t), \ \forall t, \\
& Y_k(t) \geq h\left(\overline{X}_k(t+1) + A(t) \right), \ \forall k, t, \\
& Y_k(t) \geq p\left(-\overline{X}_k(t+1) + A(t) \right), \ \forall k, t, \\
& U_k(t) \leq \overline{X}_k(t+1) - \overline{X}_k(t), \ \forall k, t, \\
& q(t) + r(s,t) \geq \widehat{W}(t), \ q(t) \geq 0, \ r(s,t) \geq 0, \ \forall t, s \leq t, \\
& 0 \leq U_k(t) \leq M\, V_k(t), \ V_k(t) \in \{0, 1\}, \ \forall k, t.
\end{aligned}
$$

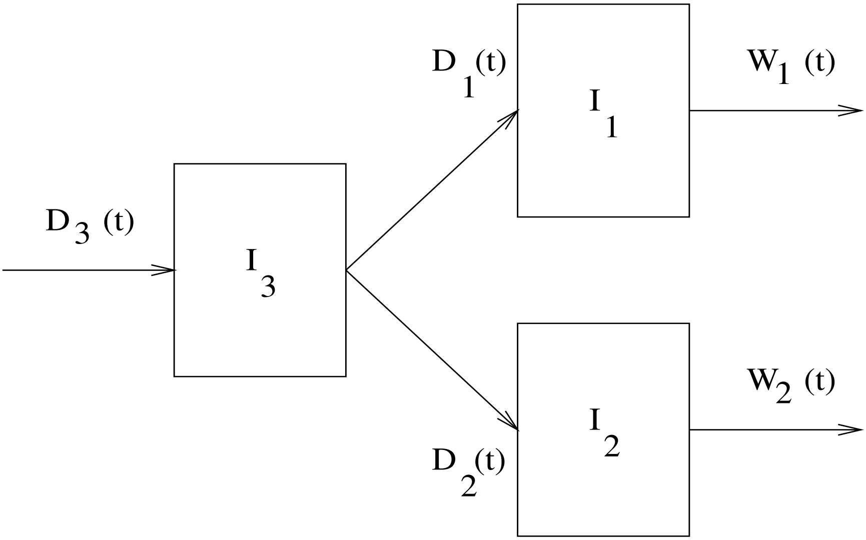
Properties

- The robust model of a series system remains of the same class as the nominal problem:
  - LO if no fixed costs,
  - MIO if fixed costs.
- Optimal robust policy is same as optimal nominal policy for modified demand: $$W'_k(t) = \overline{W}(t) + \frac{p_k - h_k}{p_k + h_k}(A(t) - A(t-1)).$$
- Optimal policy is basestock, parameters are computed by optimization, not DP.

General Supply chains

- **A new theoretical result for tree supply chains:**
  If no fixed costs, the optimal policy in the robust model is still basestock for modified cost parameters. [And we don't know what the optimal stochastic policy is.]
- **Tractability:**
  This approach is tractable for arbitrary supply chains as the robust problem remains an LOP if no fixed costs and a MIOP if fixed costs. [Complexity of the formulation does not increase with complexity of the network. Contrast with DP.]

Example

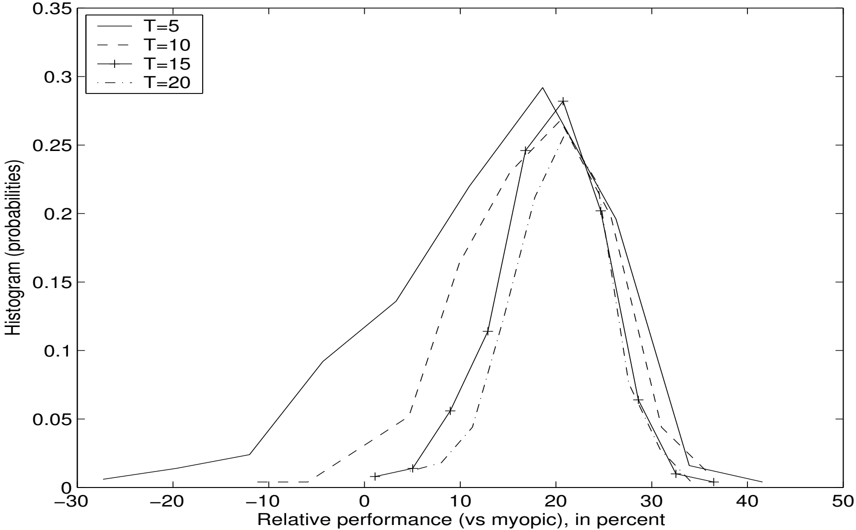


Example, continued

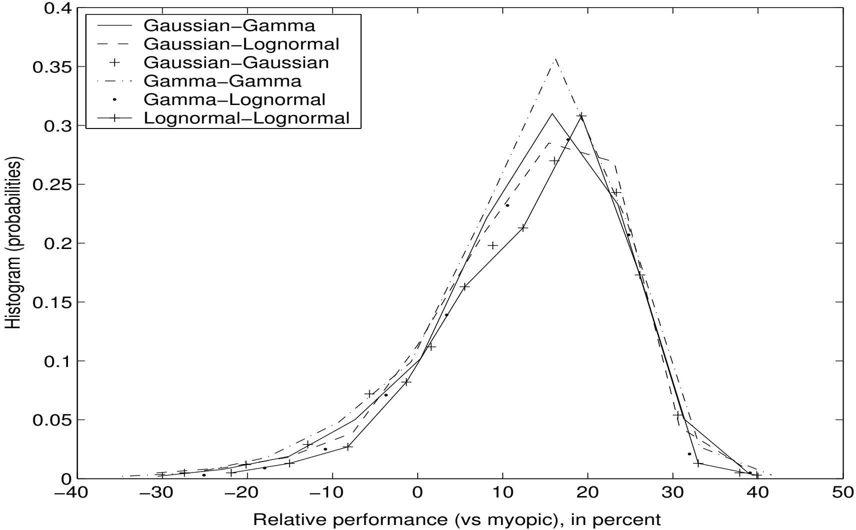
- Parameter selection is similar to single station.
- We will compare the robust approach to the myopic policy.
- Performance measure: histogram of $100 \cdot \frac{MYO - ROB}{MYO}$, with MYO (ROB) cost of myopic (robust) policy.
- Questions:
  - Role of time horizon in performance?
  - Role of distributions?

Impact of time horizon

Actual: Gamma, assumed: Gaussian distribution



Gaussian with $T = 5$



Comments

- Robust approach leads to high-quality solutions,
- Performs significantly better than myopic policies, in particular over many time periods, even when actual and assumed distributions are close,
- Is indeed robust to uncertainty on the distributions.

Summary and Conclusions

- Robust approach is numerically tractable even for large dimensions, without the curse of dimensionality.
- It offers qualitatively the same solutions as DP, when DP policies are known.
- Outperforms DP in computational experiments.
- Successfully applied approach to other problems.

15.094J Robust Modeling, Optimization and Computation

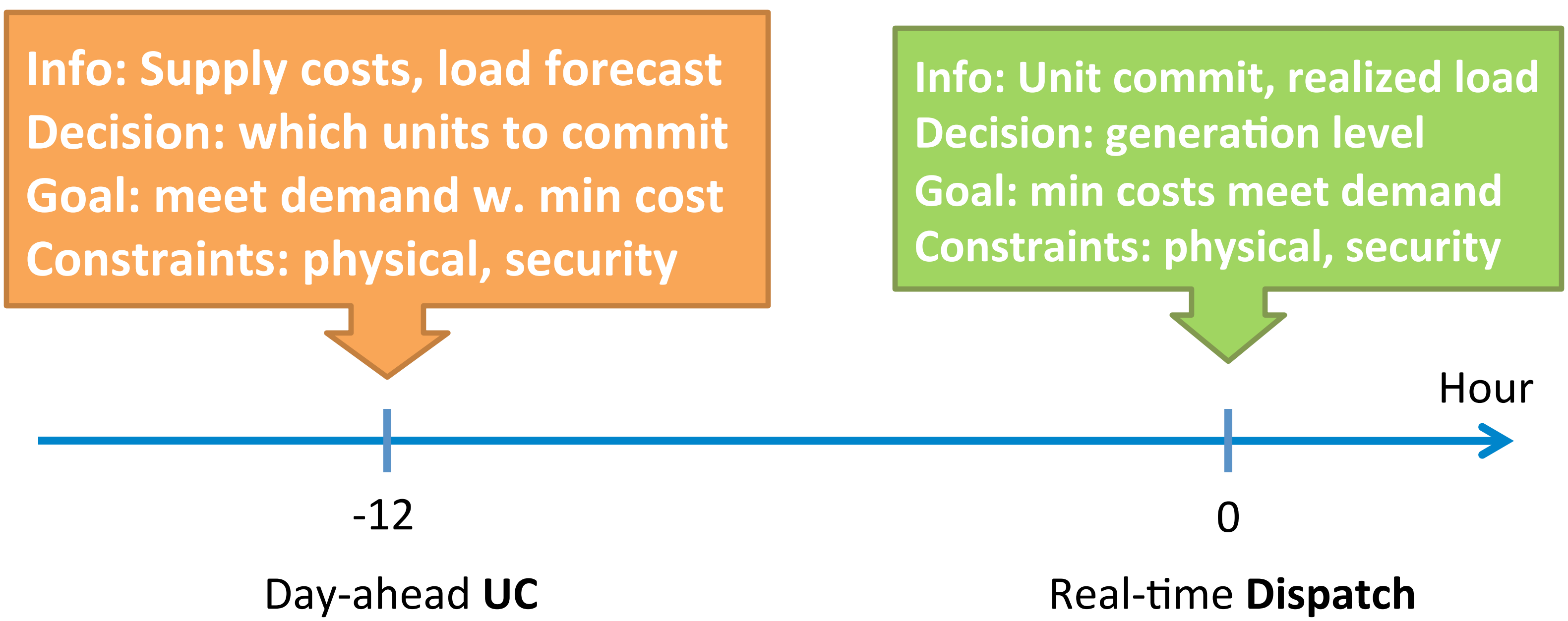
Lecture 14: RO in the unit commitment problem in electricity production

Outline

- Background
  - Unit commitment problem
- New challenges
  - Increasing uncertainty in supply/demand
- Adaptive Robust Optimization
- Conclusions

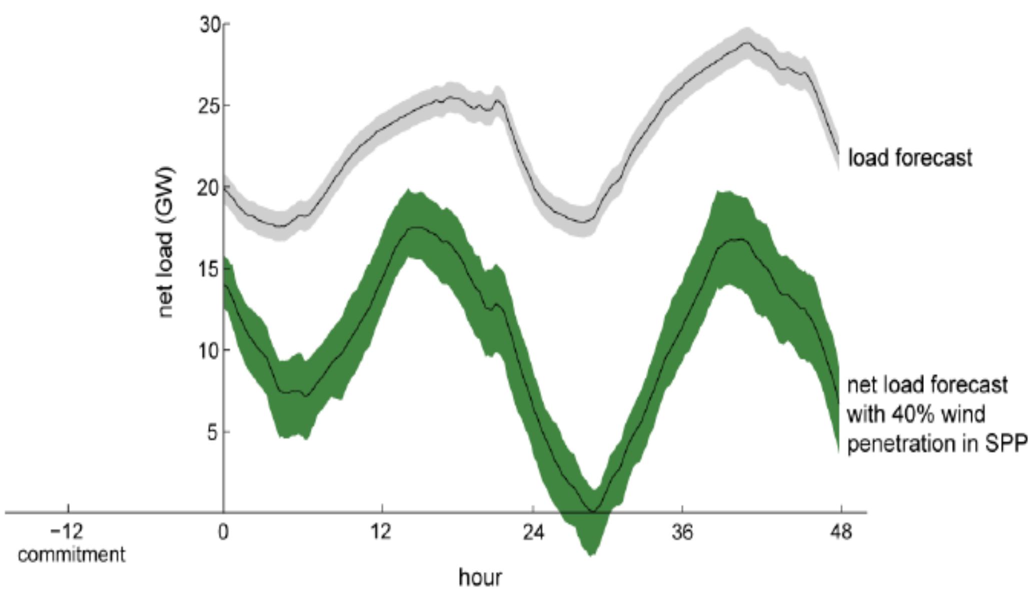
Electric Power System Operations

- Day-Ahead Decision Making: Unit Commitment
  - Generators must be committed before real-time operation (long startup time)



Challenges: Growing Uncertainty

- Supply uncertainty (Renewables like wind are exhibiting 40% annual growth)
- Demand uncertainty



Current Practice: Reserve Adjustment

- Deterministic Reserve adjustment approach

  Incorporating extra resources called reserve

  [Sen and Kothari 98] [Billinton and Fotuhi-Firuzabad 00]

Drawbacks:

1. Uncertainty not explicitly modeled
2. Both system and locational requirement are preset, heuristic, ad hoc

Current practice: MIO

Min $c'x + b'y$

s.t. $Fx<f$ (min-up/down times, start-up/shut-down)

$Hy<h$ (energy balance, reserve requirement and capacity ransmission limit, and ramping constraints)

$Ax+By<g$ (min-max generation capacity constraints)

$I_u\, y=d$

x binary (commitment variables)

$y>0$ (dispatch variables).

Stochastic Optimization

- Stochastic optimization approach

  Uncertainty modeled by distributions and scenarios

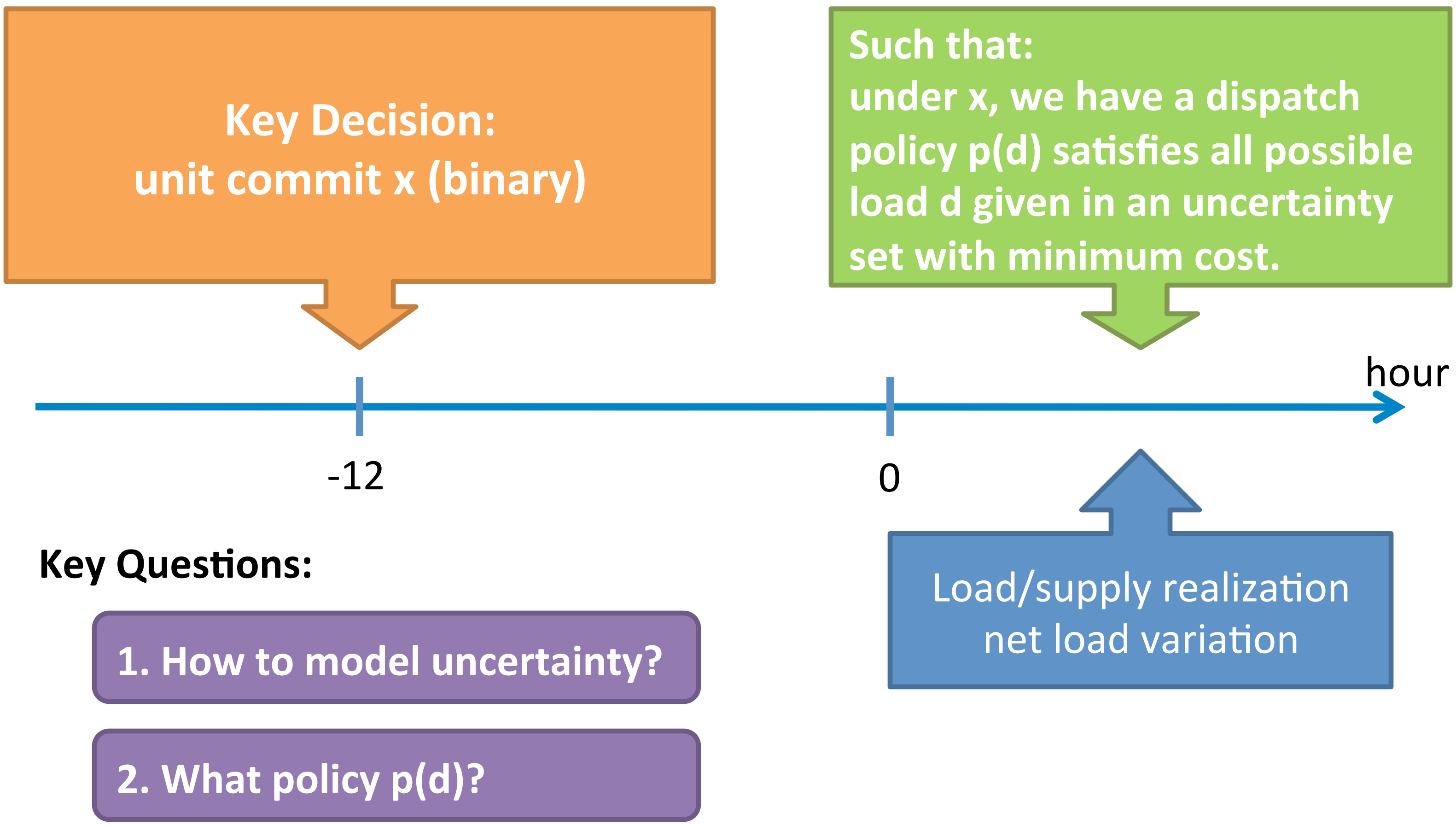
  [Takriti et. al. 96, 00] [Ozturk et. al. 05][Wong et. al. 07] [Wu et. al. 07]

Weakness:

1. Hard to select "right" scenarios in large systems
2. The huge number of scenarios needed make the problem intractable for large scale problems.

Adaptive Robust Optimization

- Two-stage robust optimization framework

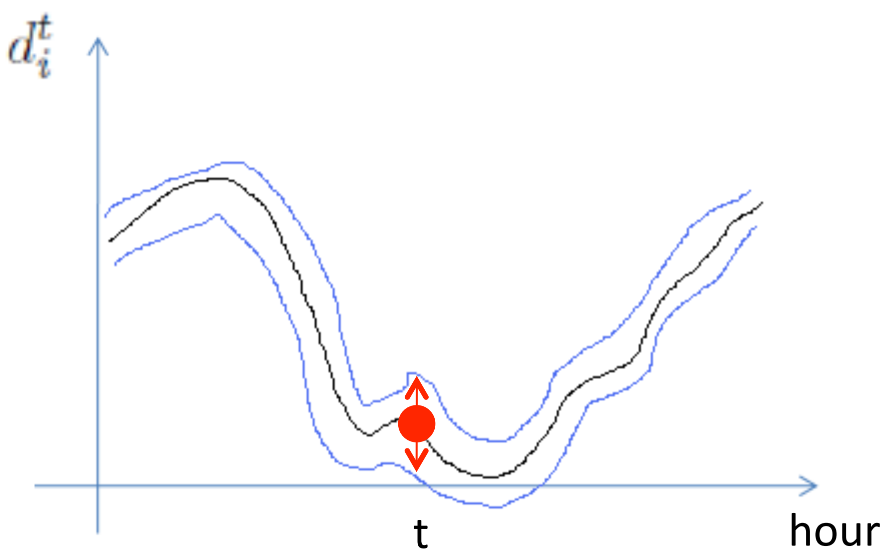


Model of Uncertainty

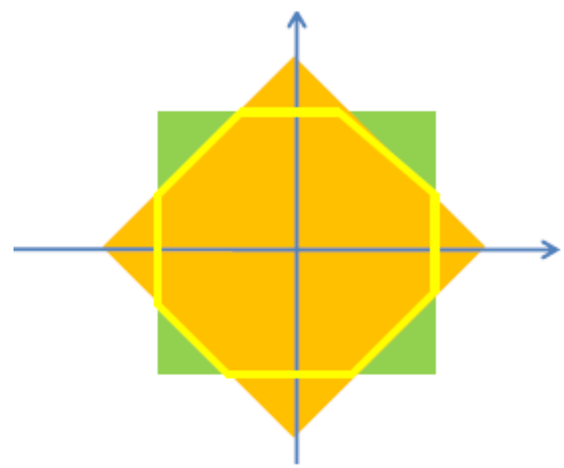
- Uncertainty model of net load variation

$$\mathcal{D}^t(\bar{\mathbf{d}}^t, \hat{\mathbf{d}}^t, \Delta^t) := \left\{ \mathbf{d}^t \in \mathbb{R}^{N_d} : \sum_{i \in N_d} \frac{|d_i^t - \bar{d}_i^t|}{\hat{d}_i^t} \leq \Delta^t, \right.$$

$$\left. d_i^t \in [\bar{d}_i^t - \hat{d}_i^t, \bar{d}_i^t + \hat{d}_i^t], \forall i \in N_d \right\}$$



$d_i^t \in [\bar{d}_i^t - \hat{d}_i^t, \bar{d}_i^t + \hat{d}_i^t]$



Budgeted

ARO

$$\text{Min}_{x,\, y(d)} \quad c'x + \max_{d \text{ in } D} \quad b'y(d)$$

s.t. $Fx<f$

$H\, y(d)<h$ for all d in D

$Ax+By(d) <g$ for all d in D

$I_u\, y(d)=d$ for all d in D

x binary

$y(d)>0$

$d=(d^t, t=1,\ldots,T)$

$D=D^1 \times D^2 \times \ldots \times D^T$

Re-formulation of ARO

$\text{Min}_x \ (c'x + \max_{d \text{ in } D} \min_{y \text{ in } O(y,d)} \ b'y)$

s.t. $Fx<f$, x binary

$O(y,d)=\{y: H\,y(d)<h,\ Ax+By(d) <g,\ I_u\, y(d)=d\}$

By duality $\min_{y \text{ in } O(y,d)} \ b'y$ is the same as

$S(x,d)=\max_{\lambda,\ \phi,\ \eta} \lambda'(Ax-g)-\phi'h+\eta'd$

s.t. $-\lambda'B-\phi'H+\eta'I_u = b'$

$\phi>0,\ \lambda>0,\ \eta$ free

Re-formulation of ARO

$R(x)=\max_{d \text{ in } D} \min_{y \text{ in } O(y,d)} b'y$ can be rewritten:

$R(x)=\max_{d,\lambda,\phi,\eta} \lambda'(Ax-g)-\phi'h+\eta'd$ bilinear problem

s.t. $-\lambda'B-\phi'H+\eta'I_u = b'$

d in D

$\phi>0, \lambda>0, \eta$ free

Benders Decomposition Algorithm

- Initialization: Get feasible x_0, solve bilinear problem $R(x_0)$ and obtain $d_1, \lambda_1, \phi_1, \eta_1$. Set k=1.
- Iteration k:
  - Step 1: Solve MIO: $\min_{x,a}\ c'x+a$

    $$\text{s.t. } a > \lambda_\iota'(Ax-g)-\phi_\iota'h+\eta_\iota'd \text{ , } i=1,\ldots,k$$

    $$Fx<f, \text{ x binary.}$$

    Let (x_k,a_k) optimal solution. Set $L=c'x_k+a_k$
  - Step 2: Solve $R(x_k)$ and obtain $d_{k+1}, \lambda_{k+1}, \phi_{k+1}, \eta_{k+1}$.

    Set $U=c'x_k+R(x_k)$
  - -- Step 3: If $U-L<\varepsilon$, stop and return x_k

    otherwise k=k+1 , go to Step 1.

Inner Problem: Solving R(x)

- Recall $R(x)=\max_{d, \lambda, \phi, \eta} \lambda'(Ax-g)-\phi'h+ \eta'd$

  s.t. $-\lambda'B-\phi'H+\eta'I_u = b'$

  d in D, $\phi>0$, $\lambda>0$, η free

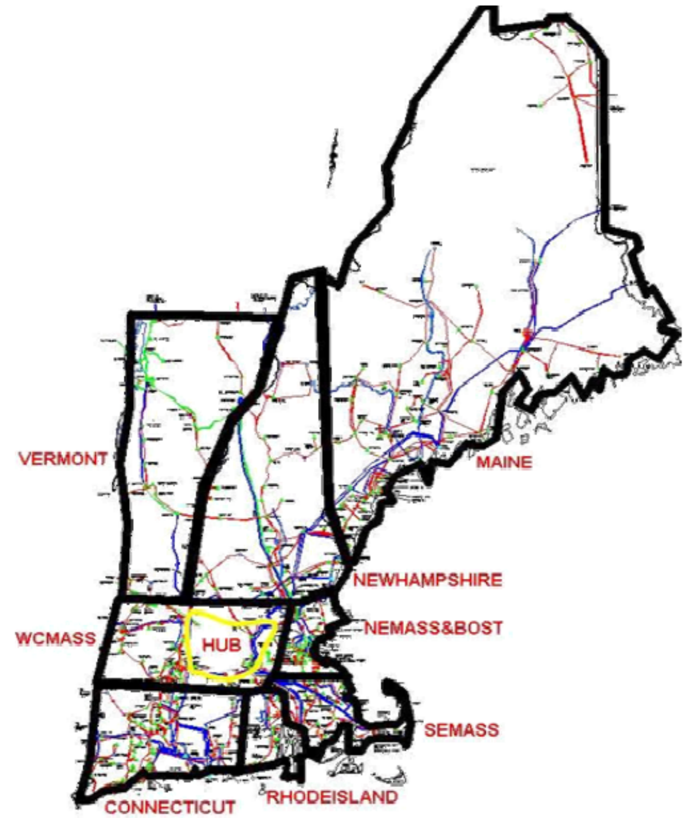
- Idea: Linearize: $\eta'd= \eta_j'd_j + (\eta - \eta_j)' d_j+ (d-d_j)' \eta_j$

- Algorithm converges to stationary point

- In practice 2-3 iterations are needed.

A Real-World Example: ISO-NE Power System

- 312 Generators
- 174 Loads
- 2816 Nodes
- 4 representative trans lines
- 24-hr data: gen/load/reserve
- Total gen cap: 31.4GW
- Total forecast load: 14.1GW

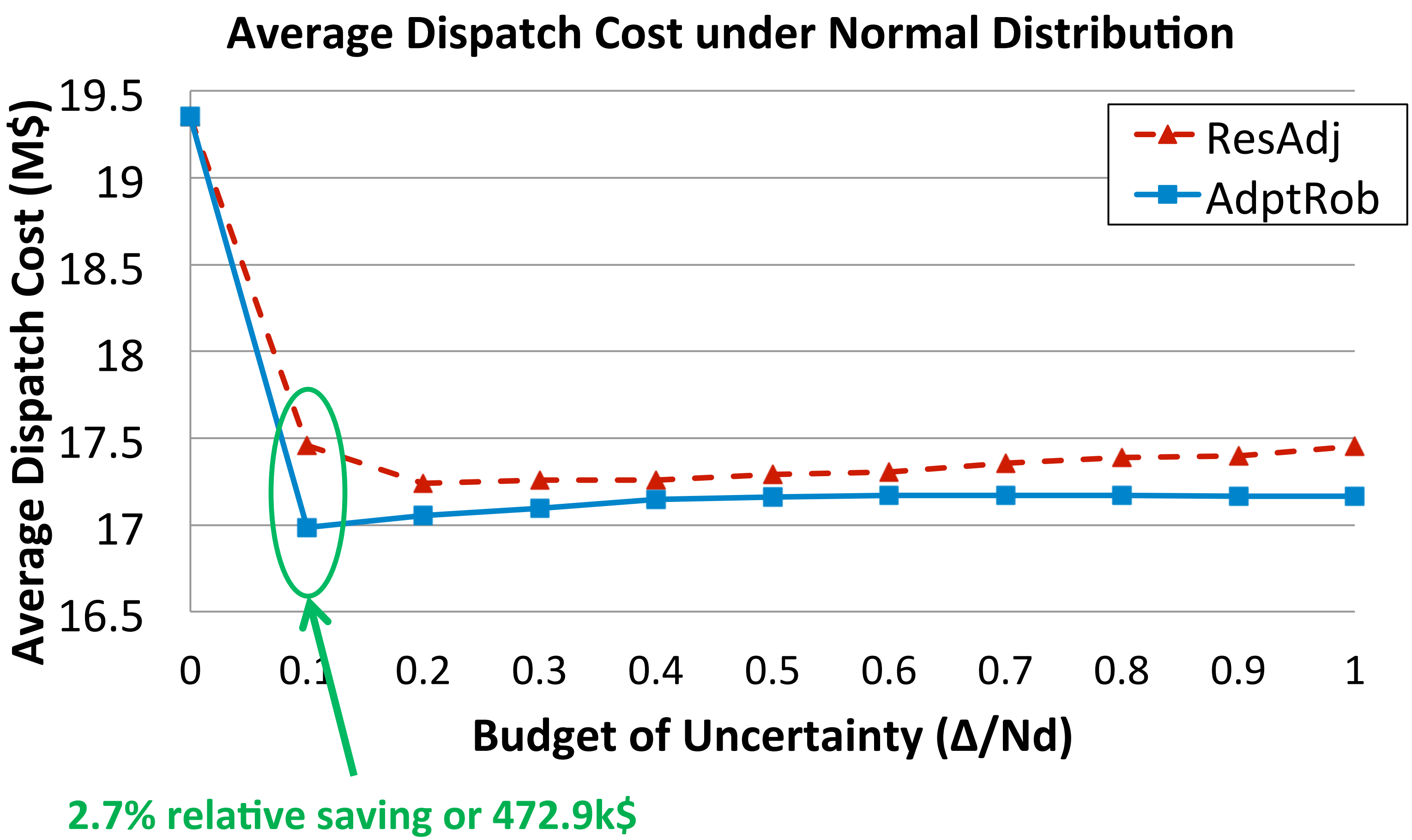


Computation Procedure and Measures

- Solve AdptRob and ResAdj UC solutions for Δ^t = 0, 0.1Nd,…,Nd for all t.
- Fix UC solutions, simulate dispatch over load samples
  - 1000 load samples from $[\bar{d}_i^t - \hat{d}_i^t, \bar{d}_i^t + \hat{d}_i^t]$
- Compute average dispatch cost and std.

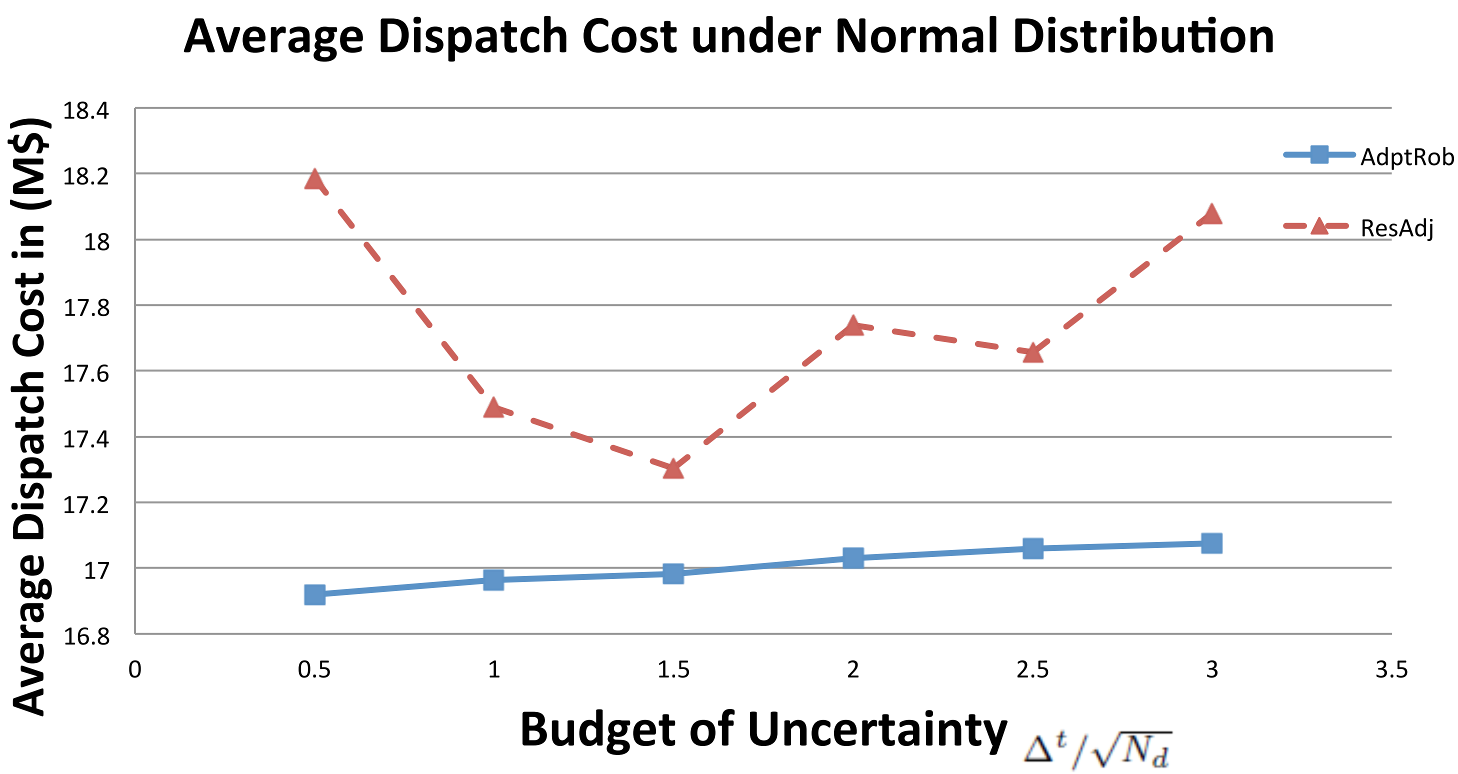
- Avg dispatch cost: Economic efficiency
- Standard deviation: Price and Operation Stability
- Robustness to distributions

Computational Results (I-a): Average dispatch cost



Avg Dispatch Cost Relative Saving := (ResAdj – AdptRob)/ResAdj **0.65% - 2.7%**

Design using Probability Law: Average dispatch cost



Design the uncertainty set
By probability law: CLT

$\Delta^t/\sqrt{N_d}$ vs Δ^t/N_d

Relative Saving: 1.86% to 6.96%
Absolute Saving: $321.2k to $1.27Million

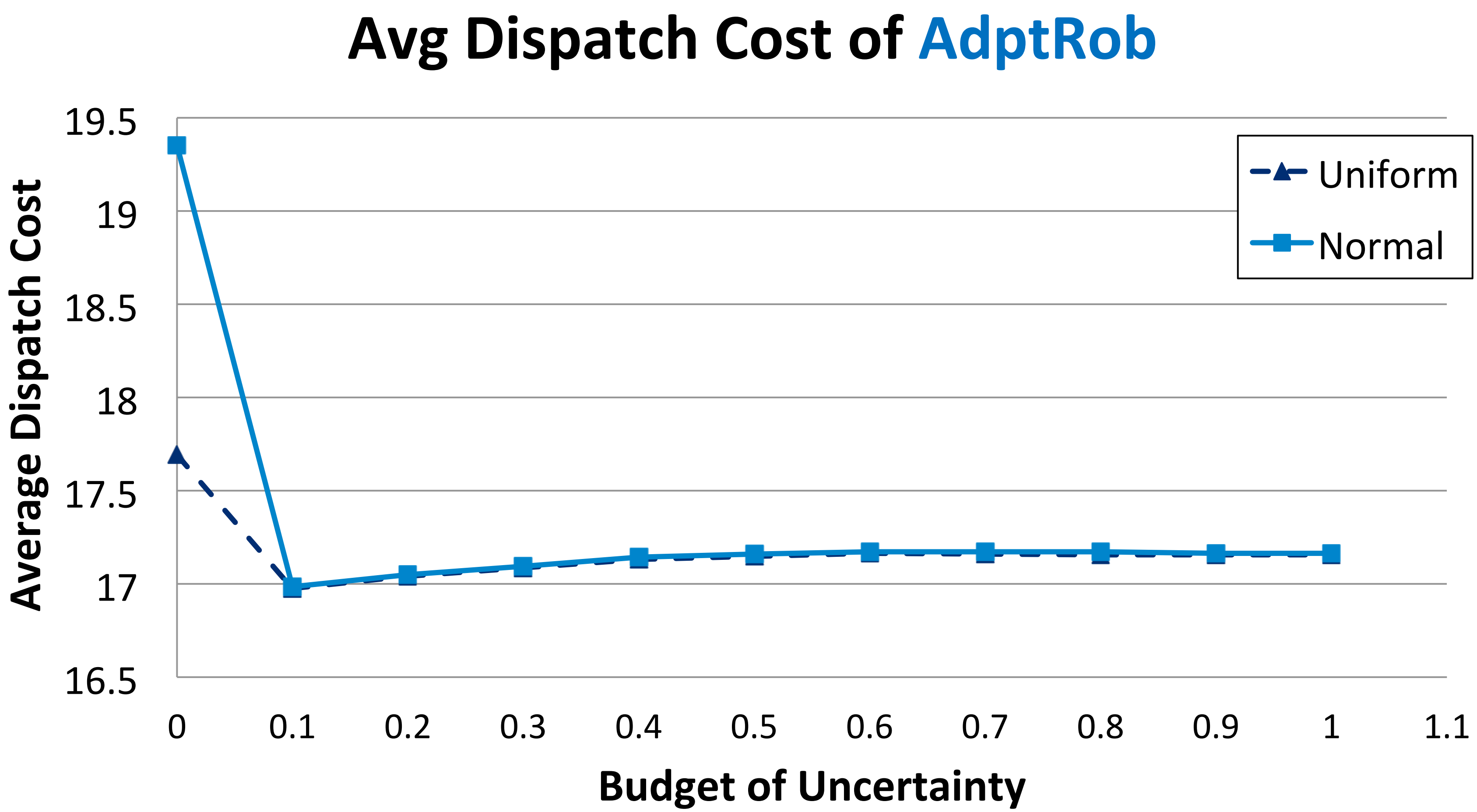
Computational Results (II): Volatility of Costs

Budget of Uncertainty	AdptRob Std disp cost ($k)	ResAdj Std disp cost ($k)	ResAdj/ AdptRob
0.1	47.5	687.5	**14.48**
0.2	46.4	687.5	**8.62**
0.3	45.4	377.8	**8.32**
0.4	44.2	366.7	**8.29**
0.5	44.1	377.2	**8.55**
0.6	44.0	370.9	**8.43**
0.7	44.0	377.1	**8.58**
0.8	43.9	370.7	**8.44**
0.9	43.9	357.9	**8.15**
1.0	43.9	361.0	**8.22**

Coeff Var: 44k/17.2M=0.25%
370k/17.3M=2.1%

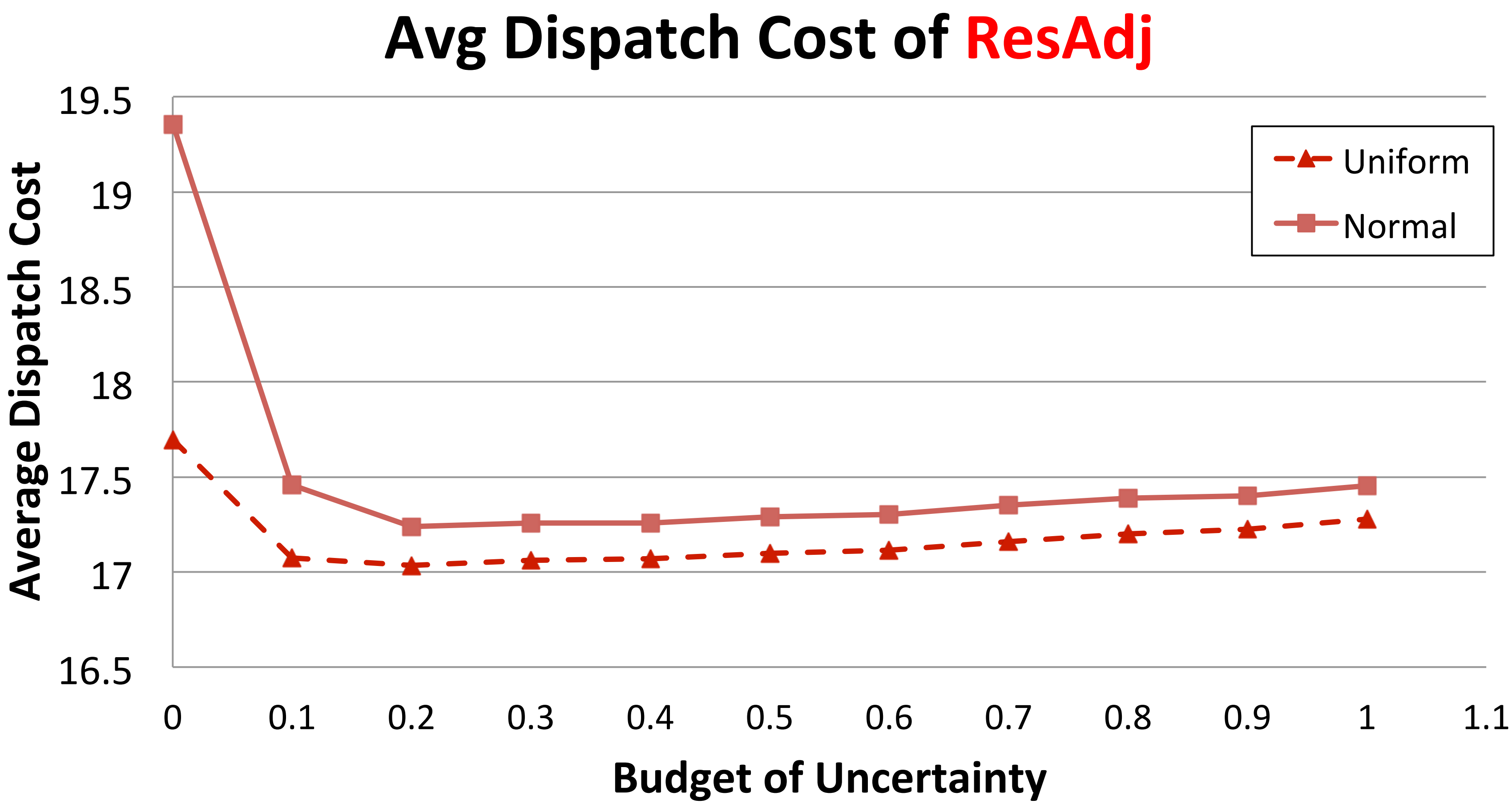
Significant reduction in cost volatility!

Computational Results (III): Robustness to Distribution



Relative difference: 0.0368% - 0.0920%
Absolute difference: $6.3k – $15.8k

Computational Results: Robustness to Distribution



Relative difference: 1.00% - 2.19%
Absolute difference: $174.4k – $382.2k

Concluding Remarks

- saves dispatch cost (6.92% $1.27M) → Economic Efficiency
- Significantly reduces cost volatility → Reduces Price & System Operation Volatility
- robust against load distributions → Data Driven Approach Demand Modeling

Reference: Adaptive Robust Optimization for Security Constrained Unit Commitment Problems, D. Bertsimas, E. Litvinov, A. Sun, J. Zhao, T. Zheng, *IEEE Transactions on Power Systems 2012*

15.094J: Robust Modeling, Optimization, Computation

Lecture 15: Robust Multi-period Portfolio Optimization

Outline

1. Multi-period portfolios under certainty
2. Robust Portfolios
3. Insights from Computations
4. Conclusions

Primitives

- M risky assets
- One riskless asset (asset 0)
- N trading periods, $t = 0, \ldots, N-1$.
- Goal is to manage the portfolio of assets in a manner that maximizes expected final wealth.
- Decision variables: x_t^m, $m = 0, 1, \ldots, M$, $t = 0, 1, \ldots, N$ dollar holdings at the beginning of time period t on asset m
- u_t^m dollar sells on asset m at time t
- v_t^m dollar buys on asset m at time t
- $c_{sell} u_t^m$ and $c_{buy} v_t^m$ transaction costs
- $\tilde{r}_t^m$ returns over $(t, t+1]$ for asset m
- r_t^0 for riskless asset known.

Dynamics

- Dynamics under certainty, $t = 1, \ldots, N$

$$
\begin{aligned}
x_t^m &= (1 + r_{t-1}^m)\left(x_{t-1}^m - u_{t-1}^m + v_{t-1}^m\right) \\
x_t^0 &= (1 + r_{t-1}^0)\left(x_{t-1}^0 + \sum_{m=1}^{M}(1 - c_{sell})u_{t-1}^m - \sum_{m=1}^{M}(1 + c_{buy})v_{t-1}^m\right)
\end{aligned}
$$

- Under known returns

$$
\begin{aligned}
\text{Max} \quad & \sum_{m=0}^{M} x_N^m \\
\text{s.t.} \quad & x_t^m = (1 + \tilde{r}_{t-1}^m)(x_{t-1}^m - u_{t-1}^m + v_{t-1}^m) \\
& x_t^0 = (1 + r_{t-1}^0)\left(x_{t-1}^0 + \sum_{m=1}^{M}(1 - c_{sell})u_{t-1}^m - \sum_{m=1}^{M}(1 + c_{buy})v_{t-1}^m\right) \\
& x_t^m \geq 0 \\
& u_t^m \geq 0,\ v_t^m \geq 0
\end{aligned}
$$

Initial Reformulation

$$
\begin{aligned}
R_0^m &= 1 \\
\tilde{R}_t^m &= (1+\tilde{r}_0^m)(1+\tilde{r}_1^m)\ldots(1+\tilde{r}_{t-1}^m) \\
\xi_t^m &= \frac{x_t^m}{\tilde{R}_t^m}, \quad \eta_t^m = \frac{u_t^m}{\tilde{R}_t^m}, \quad \zeta_t^m = \frac{v_t^m}{\tilde{R}_t^m}
\end{aligned}
$$

$$
\begin{aligned}
\text{Max} \quad & w \\
\text{s.t.} \quad & w \leq \sum_{m=1}^{M} \tilde{R}_N^m \xi_N^m + R_N^0 \xi_N^0 \\
& \xi_t^m = \xi_{t-1}^m - \eta_{t-1}^m + \zeta_{t-1}^m \\
& \xi_t^0 = \xi_{t-1}^0 + \sum_{m=1}^{M} (1-c_{sell}) \frac{\tilde{R}_{t-1}^m}{R_{t-1}^0} \eta_{t-1}^m - \sum_{m=1}^{M} (1+c_{buy}) \frac{\tilde{R}_{t-1}^m}{R_{t-1}^0} \zeta_{t-1}^m \\
& \xi_t^m,\ \eta_t^m,\ \zeta_t^m \geq 0.
\end{aligned}
$$

Final Reformulation

- Replace equalities by inequalities:

$$
\begin{aligned}
\text{Max} \quad & w \\
\text{s.t.} \quad & w \le \sum_{m=1}^{M} \tilde{R}_N^m \xi_N^m + R_N^0 \xi_N^0 \\
& \xi_t^m = \xi_{t-1}^m - \eta_{t-1}^m + \zeta_{t-1}^m \\
& \xi_t^0 \le \xi_{t-1}^0 + \sum_{m=1}^{M} (1 - c_{sell}) \frac{\tilde{R}_{t-1}^m}{R_{t-1}^0} \eta_{t-1}^m - \sum_{m=1}^{M} (1 + c_{buy}) \frac{\tilde{R}_{t-1}^m}{R_{t-1}^0} \zeta_{t-1}^m \\
& \xi_t^m,\ \eta_t^m,\ \zeta_t^m \ge 0.
\end{aligned}
$$

Modeling Uncertainty

- Uncertainty sets:

$$U_1 = \left\{ \begin{array}{l} ||\mathbf{\Sigma}_1^{-\frac{1}{2}} \left(\tilde{R}_1 - \overline{R}_1\right) || \leq \Delta \\ \underline{\delta}_2^m \tilde{R}_1^m \leq \tilde{R}_2^m \leq \overline{\delta}_2^m \tilde{R}_1^m, \;\; m = 1, \ldots, M \\ \vdots \\ \underline{\delta}_N^m \tilde{R}_{N-1}^m \leq \tilde{R}_N^m \leq \overline{\delta}_N^m \tilde{R}_{N-1}^m, \;\; m = 1, \ldots, M \end{array} \right\}$$

  where $\underline{\delta}_t^m$ and $\overline{\delta}_t^m$ are of the form $(1 + \underline{r}_t^m)$ and $(1 + \overline{r}_t^m)$, respectively.

- If data on the covariance matrices of future cumulative returns are available:

$$U_2 = \left\{ \; ||\mathbf{\Sigma}_t^{-\frac{1}{2}} \left(\tilde{R}_t - \overline{R}_t\right) || \leq \Delta_t, \;\; t = 1, \ldots, N \; \right\},$$

  where Δ_t are budgets of uncertainty.

Robust counterpart for U_1

$$
\begin{aligned}
\max \quad & w \\
\text{s.t.} \quad & w \le R_N^0 \xi_N^0 - \left(p^N - q^N\right)' \boldsymbol{\Sigma}_1^{-\frac{1}{2}} \overline{R}_1 - \frac{\Delta}{d} \cdot (u^N)' e \\
& \xi_t^m = \xi_{t-1}^m - \eta_{t-1}^m + \zeta_{t-1}^m \\
& \xi_{t+1}^0 \le \xi_t^0 - \left(p^t - q^t\right)' \boldsymbol{\Sigma}_1^{-\frac{1}{2}} \overline{R}_1 - \frac{\Delta}{d} \cdot (u^t)' e \\
& \xi_1^0 \le \xi_0^0 + \sum_{m=1}^{M} \left(1 - c_{sell}\right) \eta_0^m - \sum_{m=1}^{M} \left(1 + c_{buy}\right) \zeta_0^m \\
& \left(p^1 - q^1\right)' \boldsymbol{\Sigma}_1^{-\frac{1}{2}} + \begin{pmatrix} \underline{\delta}_2^1(1)\alpha_2^1(1) - \overline{\delta}_2^1(1)\beta_2^1(1) \\ \vdots \\ \underline{\delta}_2^M(1)\alpha_2^M(1) - \overline{\delta}_2^M(1)\beta_2^M(1) \end{pmatrix}' = \begin{pmatrix} -\frac{(1-c_{sell})}{R_1^0}\eta_1^1 + \frac{(1+c_{buy})}{R_1^0}\zeta_1^1 \\ \vdots \\ -\frac{(1-c_{sell})}{R_1^0}\eta_1^M + \frac{(1+c_{buy})}{R_1^0}\zeta_1^M \end{pmatrix} \\
& \left(p^t - q^t\right)' \boldsymbol{\Sigma}_1^{-\frac{1}{2}} + \begin{pmatrix} \underline{\delta}_2^1(t)\alpha_2^1(t) - \overline{\delta}_2^1(t)\beta_2^1(t) \\ \vdots \\ \underline{\delta}_2^M(t)\alpha_2^M(t) - \overline{\delta}_2^M(t)\beta_2^M(t) \end{pmatrix}' = \mathbf{0}'
\end{aligned}
$$

Robust counterpart for U_1, continued

$$
\begin{aligned}
&-p^t - q^t + u^t = \mathbf{0}\\
&\left((u^t)'e\right)\cdot e \geq d\cdot u^t\\
&-\alpha_\tau^m(t) + \beta_\tau^m(t) + \underline{\delta}_{\tau+1}^m \alpha_{\tau+1}^m(t) - \overline{\delta}_{\tau+1}^m \beta_{\tau+1}^m(t) = 0,\\
&-\alpha_t^m(t) + \beta_t^m(t) + \underline{\delta}_{t+1}^m \alpha_{t+1}^m(t) - \overline{\delta}_{t+1}^m \beta_{t+1}^m(t) = -\frac{(1-c_{sell})}{R_t^0}\eta_t^m + \frac{(1+c_{buy})}{R_t^0}\zeta_t^m,\\
&-\alpha_N^m(t) + \beta_N^m(t) = 0\\
&-\alpha_N^m(N) + \beta_N^m(N) = -\xi_N^m,\\
&\xi_t^m,\ \eta_t^m,\ \zeta_t^m,\ \alpha_\tau^m(t),\ \beta_\tau^m(t),\ p^t,\ q^t,\ u^t \geq 0,
\end{aligned}
$$

Robust counterpart for U_2

$$
\begin{array}{ll}
\max & w \\
\text{s.t.} & w \le \overline{R}_N' \boldsymbol{\xi}_N - \dfrac{\Delta_N}{d}\left(w^N\right)' e \\
& \xi_t^m = \xi_{t-1}^m - \eta_{t-1}^m + \zeta_{t-1}^m \\
& \xi_1^0 \le \xi_0^0 + \displaystyle\sum_{m=1}^{M} (1 - c_{sell})\,\eta_0^m - \sum_{m=1}^{M} (1 + c_{buy})\,\zeta_0^m \\
& \xi_{t+1}^0 - \xi_t^0 - \begin{pmatrix} -\frac{(1-c_{sell})}{R_t^0}\eta_t^1 + \frac{(1+c_{buy})}{R_t^0}\zeta_t^1 \\ \vdots \\ -\frac{(1-c_{sell})}{R_t^0}\eta_t^M + \frac{(1+c_{buy})}{R_t^0}\zeta_t^M \end{pmatrix}' \overline{R}_t + \dfrac{\Delta_t}{d}\left(w^t\right)' e \le 0 \\
& \left(2p^N - w^N\right)' = -\boldsymbol{\xi}_N' \boldsymbol{\Sigma}_N^{\frac{1}{2}} \\
& \left(2p^t - w^t\right)' = -\begin{pmatrix} -\frac{(1-c_{sell})}{R_t^0}\eta_t^1 + \frac{(1+c_{buy})}{R_t^0}\zeta_t^1 \\ \vdots \\ -\frac{(1-c_{sell})}{R_t^0}\eta_t^M + \frac{(1+c_{buy})}{R_t^0}\zeta_t^M \end{pmatrix}' \boldsymbol{\Sigma}_t^{\frac{1}{2}} \\
& \left(w^t\right)' e \ge d \cdot w^t \\
& w^t \ge p^t,\ t = 1, \ldots, N \\
& p^t,\ \xi_t^m,\ \eta_t^m,\ \zeta_t^m \ge \mathbf{0}
\end{array}
$$

Size

- The robust counterpart for U_1 has $2MN^2 + MN - 2M + N$ variables and $3MN^2 + 8MN - 3M + N + 2$ constraints
- For a portfolio of 500 stocks optimized over 6 time periods (e.g., rebalanced monthly for half a year): 38,006 variables and 76,508 constraints
- The robust counterpart for U_2 : $5MN - 2M + N$ variables and $8MN - 2M + 2N + 1$ constraints.
- For a portfolio of 500 stocks optimized over 6 time periods 14,006 variables and 23,016 constraints.

Approaches

1. Single Period Mean-Variance (MR1)
2. Single Period Robust using U_1 (SPR)
3. Multiperiod Robust with U_1 (MR1)
4. Multiperiod Robust with U_2 (MR2)
5. Multiperiod Nominal (MPN)

Questions

1. How do the different portfolio optimization approaches perform when single period returns are drawn from the same distribution at each time period, and when there is no noise, i.e., the simulated distributions have the same parameters as the ones used as input to the corresponding optimization problems?
2. How do the approaches perform when returns are drawn from distributions with different expected returns at every time period? Does the ability to "see ahead" help the multi-period approaches perform better?
3. How do the approaches perform when nature does not behave the way we expect probabilistically, e.g., when returns are drawn from different distributions than the ones specified at the beginning, or when the parameters of the distributions are perturbed?

Normal Single Period Returns

- $M = 3$ $\boldsymbol{\sigma} = [0.15, 0.20, 0.22]$, and their correlation matrix, which remains constant over time, is

$$Cor = \begin{bmatrix} 1 & 0.5 & 0.7 \\ 0.5 & 1 & -0.2 \\ 0.7 & -0.2 & 1 \end{bmatrix}.$$

  The single period riskless return is 0.025.
- $N = 5$ and transaction costs are 1% of the amount traded, .
- Experiments 1 and 2: $\underline{\delta}_t^m$ and $\overline{\delta}_t^m$ in U_2 are both set equal to the expected value of the corresponding single period asset return
- In Experiment 1, the returns are simulated from the same multivariate normal distribution at every time period.
- In Experiments 2-4, single period returns are simulated from multivariate normal distributions with different expected values at each time period
- In Experiment 2, the upper and lower bounds in the MR1 formulation are equal, and are set to the expected values of the single period returns.

Normal Single Period Returns, continued

- Experiment 3 is the same as Experiment 2, but the upper and lower bounds in the MR1 formulation are set to expected values $\pm 50\%$ of the standard deviation of the corresponding stock.
- Experiment 4 is the same with Experiment 3, but returns are simulated from a multivariate normal distribution whose expected values are lower than the expected values used by the four algorithms by 50% of the standard deviation for the corresponding single period return.

Asset Returns

Trading Period t	Expected Returns		
	Stock 1	Stock 2	Stock 3
0	0.080	0.090	0.120
1	0.075	0.080	0.110
2	0.080	0.100	0.070
3	0.080	0.110	0.090
4	0.080	0.105	0.110

Results over 1000 simulations

Exp. No.	Method	Mean	StdDev	Min	Max	Ratio	Prob (%)
1	SMV	0.0985	0.0674	-0.0982	0.3204	1.46	7.40
	SPR	0.0978	0.0676	-0.1052	0.3227	1.45	7.50
	MR1	0.0986	0.0506	-0.0081	0.3255	1.95	0.10
	MPN	0.0972	0.0962	-0.1855	0.4105	1.01	15.90
2	SMV	0.0892	0.0667	-0.1241	0.2987	1.34	9.30
	SPR	0.0883	0.0694	-0.1488	0.2989	1.27	9.80
	MR1	0.0905	0.0387	0.0170	0.2657	2.34	0.00
	MPN	0.0831	0.0949	-0.2280	0.3780	0.88	19.10
3	SMV	0.0876	0.0664	-0.1454	0.3113	1.32	8.90
	SPR	0.0876	0.0694	-0.1675	0.2899	1.26	9.40
	MR1	0.0902	0.0370	0.0211	0.2415	2.44	0.00
	MPN	0.0852	0.0971	-0.3109	0.3569	0.88	18.40
4	SMV	0.0781	0.0670	-0.1242	0.3202	1.17	11.80
	SPR	0.0775	0.0693	-0.1299	0.3006	1.12	13.40
	MR1	0.0876	0.0393	0.0176	0.2892	2.23	0.00
	MPN	0.0720	0.0951	-0.2177	0.3496	0.76	22.30

Discussion

- MR1 achieves better average return, probability of loss, and mean-to-standard deviation ratio
- Dominance in the mean-to-standard deviation ratio is particularly important, because it shows that the risk of the portfolio is decreased at no cost to the expected portfolio return.
- Standard deviation is appropriate as a measure of risk in these experiments, because the return distributions are symmetric (normal).
- MR1 also has better worst-case performance.
- In Experiment 3, where the the upper and lower bounds on future stock returns are set to be 50% of the corresponding returns' standard deviations away from the returns' expected values (as opposed to expected values, as in Experiment 2), MR1 becomes more conservative, and its worst case scenario performance improves.

Asymmetric (Lognormal) Single Period Returns

- 25 stocks over 5 time periods
- Returns given by factor model

$$\begin{aligned} \ln(1 + r_t^m) &= \boldsymbol{\Omega}_m' \left[\kappa \cdot e + \sigma \cdot \boldsymbol{\nu}_t \right], \; t = 0, 1, \ldots, N-1, m = 1, \ldots, M \\ \ln(1 + r_t^0) &= \kappa, \; t = 0, 1, \ldots, N-1, \end{aligned}$$

where $\{\boldsymbol{\nu}_0, \boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_{N-1}\}$ are independent K-dimensional Gaussian random vectors with zero mean and unit covariance matrix; $e \in \Re^K = (1, \ldots, 1)'$; $\boldsymbol{\Omega}_m \in \Re_+^K$ are fixed vectors; and $\kappa, \sigma > 0$ are fixed reals. Single period returns are therefore lognormal.

Experiments

- **Experiment 1**: The single period returns for each asset are assumed to be independent and identically distributed across time periods. The upper and lower bounds in MR1 are set to the expected values of the corresponding single period returns.
- **Experiment 2** is the same as Experiment 1, but the simulated returns are perturbed: after a realization of the vector $\boldsymbol{\nu}_t$ is obtained and the returns are computed, 10% of the value of each realized return is subtracted. The simulated values are therefore lower on average than the optimization problems "expect."
- **Experiment 3** is the same as Experiment 2, but the upper and lower bounds for each asset in the MR1 formulation are set to be 50% of the standard deviation of the corresponding asset.
- **Experiment 4**, the single period expected returns for the first time period are the same as the expected returns in Experiments 1-3, but the expected single period returns in later time periods are different. The MR1 and the MR2 "know" the expected returns more than one time period ahead. The upper and lower bounds for each asset in the MR1 formulation are set to the expected values of the single period returns.

Experiments, continued

- **Experiment 5** is the same as Experiment 4; however, 10% of the realized returns is subtracted in all simulations. The expected values for returns used as inputs in all optimization models, as well as the covariance matrices used in the formulation of MR2, are therefore not correct, so the models are misspecified. The upper and lower bounds for each asset return in the MR1 formulation are set to 50% of the standard deviation of the single period returns. The results of Experiment 5 are particularly important, because the setting of the experiment is the most realistic one.
- **Experiment 6** is the same as Experiment 1, but single period returns are drawn from a multivariate normal distribution (instead of a lognormal distribution) with the same expected value and single period covariance matrix as the single period lognormal distribution for returns.
- T, d-norm equal to $\sqrt{\text{Number of Stocks}}$ and $d = \text{Number of Stocks}$, and the values of λ and Δ change correspondingly.

Results

Exp. No.	Mhd	Mean	Std Dev	Ratio	Min	5th Per	50th Per	95th Per	Max	Prob (%)
1	SMV	0.1933	0.1396	1.39	-0.1983	-0.0255	0.1798	0.4283	0.7441	7.50
	SPR	0.1935	0.1535	1.26	-0.2285	-0.0387	0.1793	0.4637	0.8517	8.40
	MR1	0.2875	0.2338	1.23	-0.1866	-0.0130	0.2458	0.7497	1.4013	6.70
	MR2	0.3468	0.2598	1.34	-0.1135	0.0100	0.2964	0.8543	1.4306	4.70
	MPN	0.1805	0.1793	1.01	-0.2507	-0.0834	0.1660	0.5021	0.9159	15.90
2	SMV	0.0715	0.1252	0.57	-0.2199	-0.1176	0.0632	0.2951	0.5728	30.10
	SPR	0.0727	0.1359	0.53	-0.2689	-0.1289	0.0600	0.3152	0.6669	32.20
	MR1	0.1455	0.1971	0.74	-0.2386	-0.1121	0.1117	0.5098	1.3134	21.90
	MR2	0.1844	0.2058	0.90	-0.2320	-0.0889	0.1496	0.5493	1.0864	18.40
	MPN	0.0617	0.1598	0.39	-0.3128	-0.1739	0.0468	0.3491	0.6539	39.20
3	SMV	0.0709	0.1219	0.58	-0.2915	-0.1201	0.0657	0.2724	0.5364	29.90
	SPR	0.0702	0.1321	0.53	-0.2870	-0.1301	0.0634	0.3001	0.6871	31.20
	MR1	0.1357	0.1763	0.77	-0.2093	-0.0996	0.1102	0.4688	0.9867	22.60
	MR2	0.1774	0.2019	0.88	-0.2021	-0.0905	0.1494	0.5380	1.5102	17.60
	MPN	0.0553	0.1562	0.35	-0.3413	-0.1691	0.0435	0.3371	0.6691	38.40
4	SMV	0.1968	0.1623	1.21	-0.3010	-0.0647	0.1948	0.4738	0.6814	11.20
	SPR	0.1869	0.1839	1.02	-0.3858	-0.0973	0.1791	0.5112	0.7703	15.10
	MR1	0.6941	0.4544	1.53	-0.2022	0.0899	0.6129	1.5331	2.7689	1.90
	MR2	0.2872	0.2014	1.43	-0.1629	0.0009	0.2577	0.6504	1.5458	4.80
	MPN	0.1642	0.2069	0.79	-0.3493	-0.1578	0.1557	0.5184	0.9441	22.40
5	SMV	0.0733	0.1552	0.47	-0.3266	-0.1579	0.0588	0.3403	0.8522	33.20
	SPR	0.0623	0.1703	0.37	-0.3687	-0.1903	0.0460	0.3635	0.8555	38.90
	MR1	0.4212	0.3794	1.11	-0.2355	-0.0158	0.3404	1.0652	2.9457	6.60
	MR2	0.1426	0.1818	0.78	-0.2280	-0.0973	0.1151	0.4712	1.2849	21.20
	MPN	0.0502	0.1950	0.26	-0.4059	-0.2256	0.0306	0.3878	1.0548	43.70
6	SMV	0.2312	0.1376	1.68	-0.2354	-0.0046	0.2369	0.4454	0.5953	5.40
	SPR	0.2278	0.1831	1.24	-0.3578	-0.0721	0.2324	0.5214	0.7698	11.50
	MR1	0.3239	0.2934	1.10	-0.6136	-0.1254	0.3054	0.8267	1.3822	12.50
	MR2	0.4399	0.2956	1.49	-0.3519	-0.0042	0.4140	0.9790	1.5433	5.10
	MPN	0.1743	0.2794	0.62	-0.7380	-0.3052	0.1796	0.6384	1.1475	26.40

Pairwise Comparisons-% row policy beats column policy

	Exp. 1, $d = \sqrt{\text{Number of Stocks}}$					Exp. 1, $d =$ Number of Stocks				
	SMV	SPR	MR1	MR2	MPN	SMV	SPR	MR1	MR2	MPN
SMV		47.4	28.4	12.5	52.1		42.5	25.4	2.7	51.4
SPR	52.6		26.6	6.9	54.3	57.5		27.3	2.6	54.5
MR1	71.6	73.4		37.6	79.4	74.6	72.7		36.5	79.6
MR2	87.5	93.1	62.4		94.1	97.3	97.4	63.5		90.0

	Exp. 2, $d = \sqrt{\text{Number of Stocks}}$					Exp. 2, $d =$ Number of Stocks				
	SMV	SPR	MR1	MR2	MPN	SMV	SPR	MR1	MR2	MPN
SMV		45.0	28.9	12.6	52.5		41.3	29.1	5.5	53.5
SPR	55.0		26.7	6.5	54.1	58.7		29.7	5.5	57.2
MR1	71.1	73.3		37.6	79.7	70.9	70.3		33.5	80.1
MR2	87.4	93.5	62.4		91.2	94.5	94.5	66.5		87.4

	Exp. 3, $d = \sqrt{\text{Number of Stocks}}$					Exp. 3, $d =$ Number of Stocks				
	SMV	SPR	MR1	MR2	MPN	SMV	SPR	MR1	MR2	MPN
SMV		44.4	27.4	13.9	55.9		42.5	9.5	4.5	47.8
SPR	55.6		24.9	7.7	58.0	57.5		10.6	5.0	51.7
MR1	72.6	75.1		31.4	83.1	90.5	89.4		31.8	83.9
MR2	86.1	92.3	68.6		93.5	95.5	95.0	68.2		86.9

Pairwise Comparisons

	Exp. 4, $d = \sqrt{\text{Number of Stocks}}$					Exp. 4, $d = \text{Number of Stocks}$				
	SMV	SPR	MR1	MR2	MPN	SMV	SPR	MR1	MR2	MPN
SMV		58.7	1.2	31.9	56.9		50.1	0.0	31.2	54.9
SPR	41.3		0.5	31.4	53.5	49.9		0.1	31.6	55.7
MR1	98.8	99.5		88.1	96.9	100.0	99.9		88.0	96.0
MR2	68.1	68.6	11.9		73.7	68.8	68.4	12.0		71.7

	Exp. 5, $d = \sqrt{\text{Number of Stocks}}$					Exp. 5, $d = \text{Number of Stocks}$				
	SMV	SPR	MR1	MR2	MPN	SMV	SPR	MR1	MR2	MPN
SMV		59.4	0.9	35.0	56.3		52.9	0.0	33.8	55.6
SPR	40.6		0.0	32.7	53.5	47.1		0.2	35.1	56.0
MR1	99.1	100.0		85.3	93.2	100.0	99.8		82.9	91.2
MR2	65.0	67.3	14.7		71.6	66.2	64.9	17.1		67.7

	Exp. 6, $d = \sqrt{\text{Number of Stocks}}$					Exp. 6, $d = \text{Number of Stocks}$				
	SMV	SPR	MR1	MR2	MPN	SMV	SPR	MR1	MR2	MPN
SMV		48.6	35.6	16.6	59.7		46.9	34.4	7.0	58.2
SPR	51.4		31.8	9.1	58.9	53.1		35.0	6.4	62.2
MR1	64.4	68.2		30.0	81.3	65.6	65.0		25.4	81.4
MR2	83.4	90.9	70.0		93.7	93.0	93.6	74.6		92.6

Discussion

- MR1 and MR2 both have excellent performance in all experiments,
- MR2 tends to do better than MR1 when returns are identically distributed in all time periods (Experiments 1, 2, 3, and 6).
- MR1, on the other hand, performs extremely well when returns are not identically distributed, or when parameters in the model are misspecified, as is the case in Experiments 4 and 5. Moreover, it appears that its worst case performance can be improved by widening the bounds on returns in the formulation. This can be seen by comparing MR1's worst case performance in Experiments 2 and 3.

Conclusions

- Robust modeling: Systematic way for modeling.
- Use of cumulative returns.
- Tractability of RO in practice.

15.094J: Robust Modeling, Optimization, Computation

Lecture 16: Robust Option Pricing

Outline

What are options and why they are important?

- A call option on a stock of strike \$50 with maturity 3 months, gives the right to buy the stock 3 months from now at \$50. So, if the price at that time is \$90, then there is a profit of \$40.
- A put option gives the right to sell the stock.
- American versus European options.
- Call options are a widespread method of compensation for executives.
- Put options provide insurance.
- The derivatives industry: 10 trillion dollar industry

Basics

- An option is a contract defined on a set of *predetermined underlying securities*

$$\mathbf{S} = \{S_i\}_{i=1,\dots,M},$$

and is associated with a *payoff function*

$$P_f(\mathbf{S}, \mathbf{K}, T),$$

where $\mathbf{K}$ and T are a set of parameters specified in the contract.

Basics

- An option is a contract defined on a set of *predetermined underlying securities*
$$\mathbf{S} = \{S_i\}_{i=1,\dots,M},$$
and is associated with a *payoff function*
$$P_f(\mathbf{S}, \mathbf{K}, T),$$
where $\mathbf{K}$ and T are a set of parameters specified in the contract.

Example. European Call Option

The option holder has the right to buy a unit of stock at a price of K, at time T, even when the market price is S_T. The payoff, then, is
$$P_f(S, K, T) = \max(S_T - K, 0).$$

Dynamics

- Given a stock of price S_t

$$S_{t+1} = \begin{cases} u \cdot S_t & \text{with probability } q \\ d \cdot S_t & \text{with probability } 1-q \end{cases}$$

- Bond with price \$1

$$\$1 \rightarrow r, \quad r \text{ Riskless Return}$$

Payoff of an option

Consider an option (also called derivative security):

$$C_{t+1} = \begin{cases} C_u(S_t) & \text{with probability } q \\ C_d(S_t) & \text{with probability } 1-q \end{cases}$$

- What should the price of the option be?
- What should it depend on?

European call option

- $f(S) = \begin{cases} 0 & S \leq K \\ S - K & S > K \end{cases}$
- $C_u = C_u(S) = \max(u \cdot S - K, 0)$
- $C_d = C_d(S) = \max(d \cdot S - K, 0)$.

The Idea of no-arbitrage

- The story of the Princeton professor....
- Consider a portfolio that has x \$ worth of the stock and B \$ worth of the bond that pay off $B \cdot r$.
- Cost of portfolio: $x + B$
- Return of portfolio:

$$\begin{cases} x \cdot u + B{\cdot}r & \text{with probability } q \\ x \cdot d + B{\cdot}r & \text{with probability } 1 - q \end{cases}$$

Critical Idea

- Choose x, B so that we create the same payoff structure as the option, i.e.,

$$\begin{aligned} x \cdot u \; + \; B \cdot r &= C_u \\ x \cdot d \; + \; B \cdot r &= C_d \end{aligned}$$

- Solving the Linear System

$$x^* = \frac{C_u - C_d}{u - d}, \qquad B^* = \frac{u \cdot C_d - d \cdot C_u}{(u - d) r}$$

Critical Idea, continued

- What should the price of the option be?
- By No-Arbitrage

$$\underbrace{C}_{\substack{\text{Cost of} \\ \text{Option}}} = \underbrace{x^* + B^*}_{\text{Cost of Portfolio}}$$

- Since they have identical payoffs, they should have the same cost; otherwise there exists an arbitrage opportunity.

The Price

- By no arbitrage:

$$
\begin{aligned}
C &= \frac{C_u - C_d}{u-d} + \frac{u\cdot C_d - d\cdot C_u}{(u-d)r} \\
&= \frac{1}{r}\left[p\cdot C_u + (1-p)C_d\right] \quad , \quad p = \frac{r-d}{u-d}
\end{aligned}
$$

- Price of option is the present value of the expected payoff of the option, but not using the original probability q, but probability p, the risk neutral measure.

Multiple periods

- Consider now T periods
- Payoff of European call option

$$f(S_T) \;=\; (S_T - K)^+ \;=\; \max(S_T - K, 0)$$

- Price

$$C = \frac{1}{r^T} \sum_{n=0}^{T} \binom{T}{n} p^n \, (1-p)^{T-n} \left(u^n \, d^{T-n} \, S - K\right)^+$$

- Independent of probability q

Limitations

- Options depending on many securities problematic computationally.
- Modeling transaction costs, liquidity issues and market restrictions problematic conceptually and computationally.

The Model for stock returns

- Consider discrete time $\{0, 1, 2, \ldots, T\}$.
- Let $\tilde{r}_t^S$ be the return at t; i.e., the return from period $[t, t+1)$.
- Assuming $\{\tilde{r}_1^S, \tilde{r}_2^S, \ldots, \tilde{r}_\tau^S\}$ are independent, we have from the central limit theorem,

$$\frac{\sum_{i=1}^{\tau} \log\left(1 + \tilde{r}_i^S\right) - \tau \cdot \mu_{\log}}{\sigma_{\log} \cdot \sqrt{\tau}} \sim N(0, 1),$$

where $\mu_{\log}$, $\sigma_{\log}$ are mean and standard deviation of $\log\left(1 + \tilde{r}_i^S\right)$, respectively.

The Model for stock returns, continued

- The CLT motivates us to consider constraints of the form

$$\left| \frac{\log \widetilde{R}^S_\tau - \tau \cdot \mu_{\log}}{\sigma_{\log} \cdot \sqrt{\tau}} \right| \le \Gamma_\tau \qquad \forall \tau,$$

- $\widetilde{R}^S_\tau = \prod_{i=1}^{\tau}(1 + \widetilde{r}^S_i)$, is the cumulative return up to time τ and Γ_τ is some parameter.
- Other constraints can be based on the the assumption of a bounded support

$$\mu_r - \Gamma_\tau \sigma_r \le \frac{\widetilde{R}^S_\tau}{\widetilde{R}^S_{\tau-1}} \le \mu_r + \Gamma_\tau \sigma_r \qquad \forall \tau.$$

The Model for stock returns, continued

- Uncertainty set for stock returns:

$$\mathbb{U}^1 = \left\{ \widetilde{R}_t^S \;\middle|\; \begin{array}{ll} \underline{R_t^S} \leq \widetilde{R}_t^S \leq \overline{R_t^S}, & \forall t = 1 \dots T \\ \underline{r_t^S} \cdot \widetilde{R} \leq \widetilde{R}_t^S \leq \overline{r_t^S} \cdot \widetilde{R}_{t-1}^S, & \forall t = 1 \dots T \\ \underline{R_{t,\tau}^S} \cdot \widetilde{R}_\tau \leq \widetilde{R}_t \leq \overline{R_{t,\tau}^S} \cdot \widetilde{R}_\tau, & \forall \{(t,\tau) | \tau < t,\ t = 1 \dots T\} \end{array} \right\}$$

where $\underline{R_t^S} = e^{t \cdot \mu_{\log} - \Gamma \cdot \sqrt{t} \cdot \sigma_{\log}}$, $\overline{R_t^S} = e^{t \cdot \mu_{\log} + \Gamma \cdot \sqrt{t} \cdot \sigma_{\log}}$, $\underline{r_t^S} = \mu_r - \Gamma_t \cdot \sigma_r$,
$\overline{r_t^S} = \mu_r + \Gamma_t \cdot \sigma_r$,
$\underline{R_{t,\tau}^S} = (t - \tau) \cdot \mu_r - \Gamma_t \cdot \sigma_r \cdot \sqrt{t - \tau}$ and
$\overline{R_{t,\tau}^S} = (t - \tau) \cdot \mu_r + \Gamma_t \cdot \sigma_r \cdot \sqrt{t - \tau}$.

The idea of $\epsilon-$arbitrage

- Replicating portfolios and incomplete markets
  - Exact replication may not be possible.
- The idea of $\epsilon-$arbitrage.
  - Compute the best possible replicating portfolio, when the stock returns lie in an uncertainty set.
  - The resulting replication error stands for the $\epsilon-$arbitrage.
- ϵ can be seen as a measure of incompleteness of the market.

The Problem of Optimal Replication

- Given $P\left(S_T, K\right)$ is the payoff of the option.
- Define x_t^S and x_t^B are the amounts invested in the stock and the bond during the period $[t, t+1)$.
- W_T is the value of the replicating portfolio.

$$\begin{aligned} \min_{\left\{x_t^S, x_t^B, y_t\right\}} \max_{\left\{\widetilde{R}_t^S \in \mathbb{U}^1\right\}} \quad & \left|P\left(S_T, K\right) - W_T\right| \\ \text{s.t.} \quad & W_T = x_T^S + x_T^B \\ & x_t^S = \left(1 + \widetilde{r}_{t-1}^S\right)\left(x_{t-1}^S + y_{t-1}\right), \forall t = 1, \ldots, T, \\ & x_t^B = \left(1 + r_{t-1}^B\right)\left(x_{t-1}^B - y_{t-1}\right), \forall t = 1, \ldots, T, \end{aligned}$$

Robust Options Pricing

- European Call option: $P\left(\widetilde{S}, K\right) = \left(\widetilde{S}_T - K\right)^+$.
- The optimization problem

$$
\begin{aligned}
\min_{X}\ \max_{U^1} \quad & \left|\left(\widetilde{S}_T - K\right)^+ - W_T\right| \\
\text{s.t.} \quad & W_T = x_T^S + x_T^B, \\
& x_t^S = \left(1 + \widetilde{r}_{t-1}^S\right)\left(x_{t-1}^S + y_{t-1}\right), \forall t = 1 \ldots T, \\
& x_t^B = \left(1 + r_{t-1}^B\right)\left(x_{t-1}^B - y_{t-1}\right), \forall t = 1 \ldots T.
\end{aligned}
$$

Robust Options Pricing, continued

- Variable transformations :

$$\alpha_t^S = \frac{x_t^S}{R_t^S},\ \alpha_t^B = \frac{x_t^B}{R_t^B},\ \beta_t = \frac{y_t}{R_t^S},\ \text{where } \widetilde{R}_t^S = \prod_{i=0}^{t-1}\left(1+\widetilde{r}_i^S\right),\ \text{and } R_t^B = \prod_{i=0}^{t-1}\left(1+r_i^B\right).$$

- After substitution:

$$\begin{array}{ll} \displaystyle\min_{\{\alpha_t^S,\alpha_t^B,\beta_t\}} \max_{\{\tilde{R}_t^S \in \mathbb{U}^1\}} & \left|\left(S_0\widetilde{R}_T^S - K\right)^+ - \left(\widetilde{R}_T^S\alpha_T^S + R_T^B\alpha_T^B\right)\right| \\ \text{s.t.} & \alpha_t^S = \alpha_{t-1}^S + \beta_{t-1},\ \ \forall t = 1,\dots,T, \\ & \alpha_t^B = \alpha_{t-1}^B - \beta_{t-1}\dfrac{\widetilde{R}_{t-1}^S}{R_{t-1}^B},\ \forall t = 1,\dots,T. \end{array}$$

Robust Options Pricing, continued

- Substituting all intermediate α_t^B, α_t^S:

$$\min_{\{\alpha_t^S,\alpha_t^B,\beta_t\}} \max_{\{\tilde{R}_t^S \in \mathbb{U}^1\}} \left| \left(S_0 \widetilde{R}_T^S - K\right)^+ - \left(\alpha_0^S + \sum_{t=1}^{T} \beta_{t-1}\right) \widetilde{R}_T^S - \alpha_0^B R_T^B + \sum_{t=0}^{T-1} \beta_t \frac{\mathrm{R_T^B}}{\mathrm{R_t^B}} \widetilde{R}_t^S \right|.$$

- Inner problem:

$$\begin{aligned}
\min \quad & \kappa \\
\text{s.t.} \quad & \kappa \geq \left(S_0 \widetilde{R}_T^S - K\right)^+ - \left(\alpha_0^S + \sum_{t=1}^{T} \beta_{t-1}\right) \widetilde{R}_T^S - \alpha_0^B R_T^B + \sum_{t=0}^{T-1} \beta_t \frac{R_T^B}{R_t^B} \widetilde{R}_t^S, \quad \forall \tilde{R}_t^S \in \mathbb{U}^1, \\
& \kappa \geq -\left(\left(S_0 \widetilde{R}_T^S - K\right)^+ - \left(\alpha_0^S + \sum_{t=1}^{T} \beta_{t-1}\right) \widetilde{R}_T^S - \alpha_0^B R_T^B + \sum_{t=0}^{T-1} \beta_t \frac{R_T^B}{R_t^B} \widetilde{R}_t^S\right), \ \forall \tilde{R}_t^S \in \mathbb{U}^1.
\end{aligned}$$

Robust Options Pricing, continued

- Model the piecewise-linear function $\left(S_0 \widetilde{R}_T^S - K\right)^+$,
- We partition the uncertainty set $\mathbb{U}^1$

$$\mathbb{U}_a^1 = \mathbb{U}^1 \cap \left\{ \tilde{R}_T^S \geq \frac{K}{S_0} \right\},\ \mathbb{U}_b^1 = \mathbb{U}^1 \cap \left\{ \tilde{R}_T^S \leq \frac{K}{S_0} \right\}.$$

Robust Options Pricing, continued

- Using this partition, we obtain the following equivalent formulation

$$\min_{\{\alpha_0^S,\alpha_0^B,\beta_t\}} \epsilon$$

s.t.

$$\begin{aligned}
\epsilon \geq \quad & \left(S_0\widetilde{R}_T^S - K\right) - \left(\alpha_0^S + \sum_{t=0}^{T-1}\beta_t\right)\widetilde{R}_T^S - \alpha_0^B R_T^B + \sum_{t=0}^{T-1}\beta_t \frac{R_T^{\mathrm{B}}}{R_t^B}\widetilde{R}_t^S, && \forall \tilde{R}_t^S \in \mathbb{U}_a^1, \\
\epsilon \geq \quad & -\left(\left(S_0\widetilde{R}_T^S - K\right) - \left(\alpha_0^S + \sum_{t=0}^{T-1}\beta_t\right)\widetilde{R}_T^S - \alpha_0^B R_T^B + \sum_{t=0}^{T-1}\beta_t \frac{R_T^{\mathrm{B}}}{R_t^B}\widetilde{R}_t^S\right), && \forall \tilde{R}_t^S \in \mathbb{U}_a^1, \\
\epsilon \geq \quad & -\left(\alpha_0^S + \sum_{t=0}^{T-1}\beta_t\right)\widetilde{R}_T^S - \alpha_0^B R_T^B + \sum_{t=0}^{T-1}\beta_t \frac{R_T^{\mathrm{B}}}{R_t^B}\widetilde{R}_t^S, && \forall \tilde{R}_t^S \in \mathbb{U}_b^1, \\
\epsilon \geq \quad & -\left(-\left(\alpha_0^S + \sum_{t=0}^{T-1}\beta_t\right)\widetilde{R}_T^S - \alpha_0^B R_T^B + \sum_{t=0}^{T-1}\beta_t \frac{R_T^{\mathrm{B}}}{R_t^B}\widetilde{R}_t^S\right), && \forall \tilde{R}_t^S \in \mathbb{U}_b^1.
\end{aligned}$$

Robust Options Pricing, continued

- Can be converted to an equivalent linear optimization problem.
- Resulting size: $16T + 4$ decision variables and $4T + 4$ constraints.

Pricing Asian options

- Asian Call option: $P\left(\widetilde{S}, K\right) = \left(S_0 \widetilde{R}^S_{\text{ave}} - K\right)^+$, where $\widetilde{R}^S_{\text{ave}} = \sum_{t=1}^{T} \frac{\widetilde{R}^S_t}{T}$.
- Again leads to a linear formulation with size that scales linearly in T.

Pricing Lookback options

- $P\left(\widetilde{S},K\right)=\left(S_0\widetilde{R}^S_{\max}-K\right)^+,\ \widetilde{R}^S_{\max}=\max_{t=1\dots T}\left\{\widetilde{R}^S_t\right\}.$
- We obtain the following formulation

$$\min_{\{\alpha_0^S,\alpha_0^B,\beta_t\}} \quad \epsilon$$

s.t.

$$\forall k=1\dots T$$

$$\epsilon \geq \quad \left(S_0\tilde{R}^S_k-K\right)-\alpha_0^S\tilde{R}^S_T-\alpha_0^B R^B_T+\sum_{t=1}^{T}\beta_{t-1}\left(\frac{R^B_T}{R^B_t}\tilde{R}^S_t-\tilde{R}^S_T\right) \quad \forall \mathbb{U}^1_k\cap\left\{\tilde{R}^S_k\geq\frac{K}{S_0}\right\}$$

$$\epsilon \geq \quad -\left(S_0\tilde{R}^S_k-K\right)+\alpha_0^S\tilde{R}^S_T+\alpha_0^B R^B_T-\textstyle\sum_{t=1}^{T}\beta_{t-1}\left(\frac{R^B_T}{R^B_t}\tilde{R}^S_t-\tilde{R}^S_T\right) \quad \forall \mathbb{U}^1_k\cap\left\{\tilde{R}^S_k\geq\frac{K}{S_0}\right\}$$

$$\epsilon \geq \quad -\left(\alpha_0^S+\sum_{t=1}^{T}\beta_{t-1}\right)\tilde{R}^S_T-\alpha_0^B R^B_T+\sum_{t=1}^{T}\beta_{t-1}\frac{R^B_T}{R^B_t}\tilde{R}^S_t \quad \forall \mathbb{U}^1_k\cap\left\{\tilde{R}^S_k\leq\frac{K}{S_0}\right\}$$

$$\epsilon \geq \quad -\left(-\left(\alpha_0^S+\sum_{t=1}^{T}\beta_{t-1}\right)\tilde{R}^S_T-\alpha_0^B R^B_T+\sum_{t=1}^{T}\beta_{t-1}\frac{R^B_T}{R^B_t}\tilde{R}^S_t\right) \quad \forall \mathbb{U}^1_k\cap\left\{\tilde{R}^S_k\leq\frac{K}{S_0}\right\}$$

Pricing American Put options

- Can exercise the option at any time up to the time of exercise T.
- The payoff is then given by $P\left(\widetilde{S}, K\right) = \left(K - \widetilde{S}_\tau\right)$, where τ is the time of exercise.
  - τ depends on the utility of the option holder.
- In the absence of any information about the utility function of the option holder, we seek to find a replicating portfolio for all possible exercise policies.

Pricing American Put options

- Proceeding as before, we obtain the following formulation

$$\begin{aligned} \min_{\{\alpha_t^S,\alpha_t^B,\beta_t\}} \max_{\{\tilde{R}_t^S\}\in\mathbb{U}^1} \max_{\tau=1,\ldots,T} \quad & \left|\left(K - S_0\widetilde{R}_\tau\right)^+ - \left(\widetilde{R}_\tau^S\alpha_\tau^S + R_\tau^B\alpha_\tau^B\right)\right| \\ \text{s.t.} \quad & \alpha_t^S = \alpha_{t-1}^S + \beta_{t-1}, \quad \forall t = 1,\ldots,T, \\ & \alpha_t^B = \alpha_{t-1}^B - \beta_{t-1}\frac{\widetilde{R}_t^S}{R_t^B}, \ \forall t = 1,\ldots,T, \end{aligned}$$

  where τ is the time of exercise.
- The size of the resulting formulation will be quadratic in T.

Pricing Barrier options

- These options become inactive as soon as a condition($\mathbb{C}$) of the form $S_t \leq a$ or $S_t \geq b$ is reached.
- The problem of optimal replication then reduces to

$$\min_{\{x_t^S, x_t^B, y_t\}} \max_{\{\widetilde{R}_t^S \in \mathbb{U}^1 \cap \mathbb{C}\}} |P(S_T, K) - W_T|$$

  because if $\mathbb{C}$ is not satisfied the option ceases to exist and one need not worry about replicating its payoff.
- We obtain a linear formulation that scales linearly in T.

Pricing Multidimensional options

- Pricing options that depend on M underlying assets is difficult to price, for large M, using current methods because of
  - unavailability of an analytic solution and
  - the curse-of-dimensionality which prevents one from using dynamic programming.
- Proceeding as before, we seek to obtain the optimal solution of the following optimization problem

$$
\begin{aligned}
\min_{\{x_t^m, y_t^m\}} \quad & \max_{\{\widetilde{r}_t^m\} \in \mathbb{U}^M} \left| P_f\left(\{S^i\}_{i=1,\dots,M}, K\right) - W_T \right| \\
\text{s.t.} \quad & W_T = \sum_{m=0}^{M} x_T^m, \\
& x_t^m = \left(1 + \widetilde{r}_{t-1}^m\right) \cdot \left(x_{t-1}^m + y_{t-1}^m\right) \quad \forall t = 1, \dots, T, \ \forall m = 1, \dots, M, \\
& x_t^0 = \left(1 + \widetilde{r}_{t-1}^m\right) \cdot \left(x_{t-1}^0 - \sum_{m=1}^{M} y_{t-1}^m\right) \quad \forall t = 1, \dots, T.
\end{aligned}
$$

Multidimensional options

- $\mathbf{C}$: the covariance matrix of the single period returns.
- Uncertainty set $\mathbb{U}^M$

$$\mathbb{U}^M = \left\{ \widetilde{\mathbf{R}}_t \left| \begin{array}{ll} ||\mathbf{C}(\widetilde{\mathbf{R}}_1 - \overline{\mathbf{R}}_1)|| \leq \Gamma, & \\ \underline{r_t^m} \widetilde{R}_{t-1}^m \leq \widetilde{R}_t^m \leq \overline{r_t^m} \widetilde{R}_{t-1}^m & \forall t = 2, \dots, T, \ \forall m = 1, \dots, M, \\ \underline{R_t^m} \leq \widetilde{R}_t^m \leq \overline{R_t^m}, & \forall t = 1, \dots, T, \ \forall m = 1, \dots, M. \end{array} \right. \right\}.$$

- By choosing the ℓ_1 or ℓ_∞ we obtain LO formulations.

Computational Complexity

- Methodology scales polynomially with the dimension of the option and the discretization, unlike DP.

Option Type	European	Asian	Lookback	American	Index	American Index
Size	$O(T)$	$O(T)$	$O(T^2)$	$O(T^2)$	$O(M \cdot T)$	$O(M \cdot T^2)$

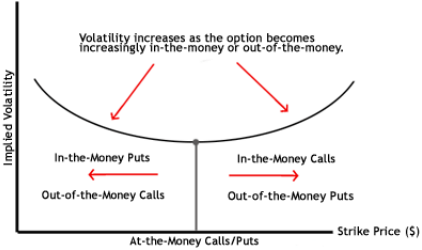
T : the number of time periods the option is written for.
M: the number of different assets required to define the option.
Size : Number of variables and constraints.

Modeling Flexibility

- Modeling many types of options - Barrier, Lookback etc.
- Modeling implied volatility smile
- Modeling transaction costs and other market restrictions, still obtaining LO formulations.

Modeling Implied Volatility smile

- *Implied volatility* of an option is simply that volatility that makes the Black-Scholes (BS) model price exactly equal to the observed market price.
    - Expect flat lines, if market participants use BS Model.
- Plotted across strike prices, they exhibit *smiles* or *smirks*.



Modeling the implied volatility smile

- Risk Aversion as an explanation for the Implied Volatility Smile.
- Lower value for Γ implies that the user is willing to take higher risk by ignoring the variability of stock prices.
- Higher value of Γ indicates that the user seeks a price that will allow him to replicate the payoff of the option for a larger range of stock prices.
- Historical evidence.
    - No smile was observed before the crash of 1987.
    - After the crash, the smile started appearing.

Modeling the implied volatility smile, continued

Risk Aversion as an explanation for the Implied Volatility Smile

- Empirical results indicate that a quadratic dependence of $\Gamma_{implied}$ with $\frac{K}{S_0}$ would be adequate to characterize the risk aversion of investor towards different strike prices.
- The following function is used to describe the relationship:

$$\Gamma(K) = \theta_0 + \theta_1 \frac{K - S_0}{S_0} + \theta_2 \left(\frac{K - S_0}{S_0} \right)^2, \ \theta_2 \geq 0.$$

- The quantity $\frac{K - S_0}{S_0}$ captures the distance between the strike and the spot price and is also called as *moneyness* in the literature.

Experiments setup

- We perform the following experiments:
  - Experiment 1 : Compare with actual market prices for European call options.
  - Experiment 2 : Compare with actual market prices for American put options.

Experiments setup

- All the experiments have a training stage and a testing stage.
    - In the training stage, we choose a random set of strike prices and calibrate (compute $\theta_0, \theta_1, \theta_2$) our model to it.
    - In the testing stage, we use our model to price the options for the remaining strikes.

Experiment 1

- The underlying security is Microsoft stock.
- The number of periods T = 18 weeks.
- The initial price of underlying security S_0= 21.4.
- Strike price of options K: ranges from 2.5 to 30.

Experiment 1

Out of Sample

No.	T	K/S	$\Gamma_{implied}$	Mkt Price	Model Price	Error
1	18	0.654	2.45	7.475	7.48	0.005
2	18	0.794	2.056	4.8	4.797	-0.003
3	18	0.888	1.75	3.25	3.232	-0.018
4	18	0.981	1.66	1.97	1.968	-0.002
5	18	1.028	1.65	1.47	1.462	-0.008
6	18	1.121	1.73	0.735	0.749	0.014

In Sample

No.	T	K/S	$\Gamma_{implied}$	Mkt Price	Model Price	Error
1	18	0.607	2.77	8.425	8.42	-0.005
2	18	0.701	2.25	6.55	6.561	0.011
3	18	0.841	1.921	4	3.984	-0.016
4	18	0.935	1.69	2.56	2.556	-0.004
5	18	1.285	2.19	0.155	0.152	-0.003

Experiment 2

- The underlying security is MSFT stock.
- The number of periods T = 25 weeks.
- The initial price of underlying security $S_0 = 24.8$.
- Strike price of options K: ranges from 7.5 to 50.

Experiment 2

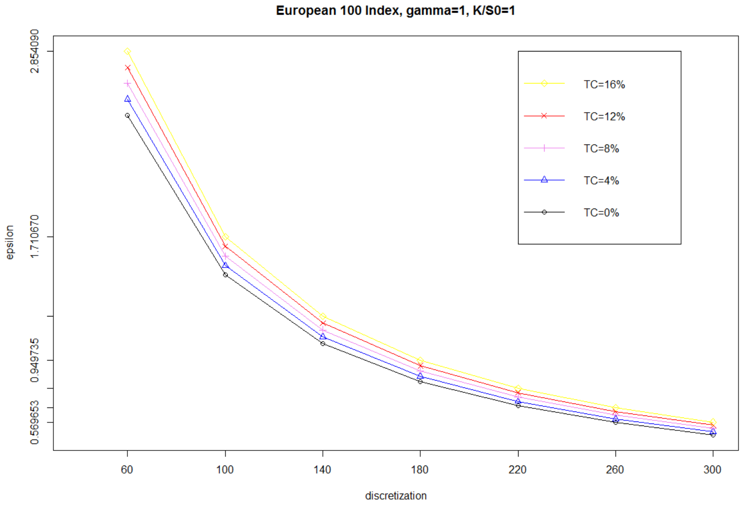
Out of Sample

No.	T	K/S	$\Gamma_{implied}$	Mkt Price	Model Price	Error
1	25	0.605	2.62	0.17	0.201	0.031
2	25	0.806	1.83	0.695	0.589	-0.106
3	25	0.968	1.6	1.895	1.764	-0.132
4	25	1.008	1.59	2.365	2.266	-0.099
5	25	1.21	1.9	5.85	5.939	0.089
6	25	1.411	2.87	10.5	10.703	0.203
7	25	1.815	7.7	20.45	20.303	-0.147

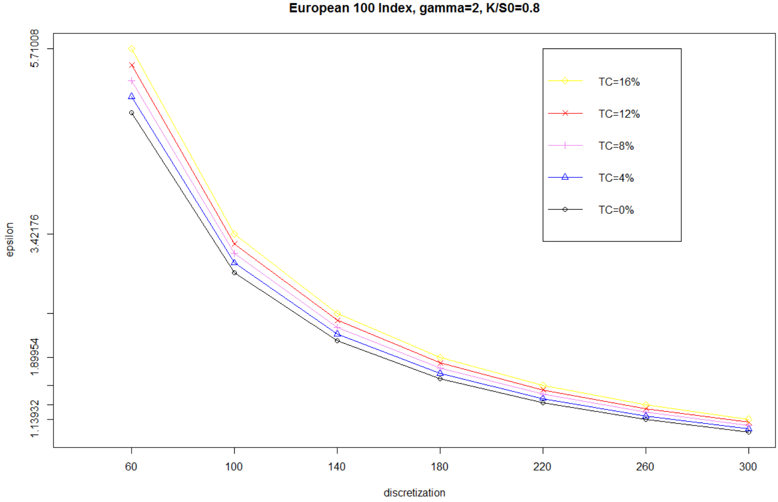
In Sample

No.	T	K/S	$\Gamma_{implied}$	Mkt Price	Model Price	Error
1	25	0.504	3.24	0.065	0.17	0.105
2	25	0.706	2.15	0.34	0.305	-0.035
3	25	0.766	1.94	0.525	0.442	-0.083
4	25	0.847	1.74	0.905	0.778	-0.127

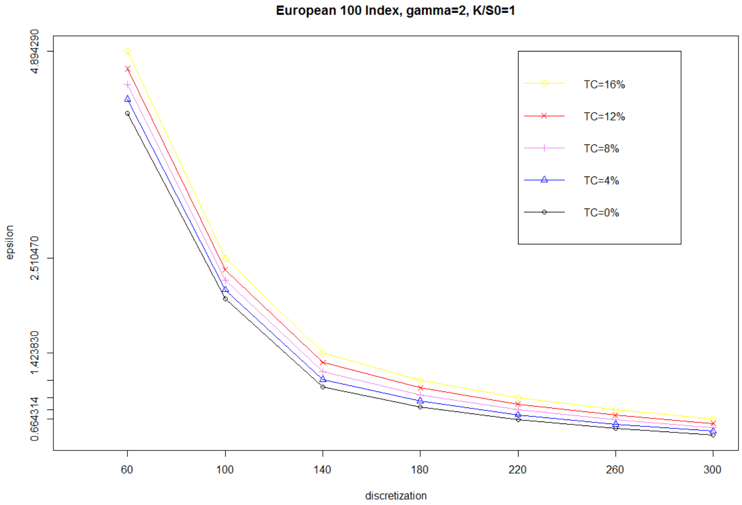
Dependence on Discretization and Transaction Costs



Dependence on Discretization and Transaction Costs



Dependence on Discretization and Transaction Costs



Conclusions

- Tractable approach to price options while accounting for the risk attitudes of the option writer.
- Approach scales polynomially (as opposed to exponentially) with the dimension of the original pricing problem.
- We provide a potential explanation for the phenomenon of the implied volatility smile, and support experimental evidence for the same.

15.094J: Robust Modeling, Optimization, Computation

Lecture 17: RO and Risk preferences

Outline

Primitives

- $\mathcal{X}$ set of all random variables distributed on $\mathbb{R}$.
- A **risk measure** is a functional $\mu : \mathcal{X} \to \mathbb{R}$.
- if $X_1, X_2 \in \mathcal{X}$, then X_1 is preferable (under μ) to X_2 if and only if $\mu(X_1) \leq \mu(X_2)$.
- Random variables with higher risk measures as less desirable, i.e., associated with *greater losses*.

Standard Deviation as a Measure of Risk

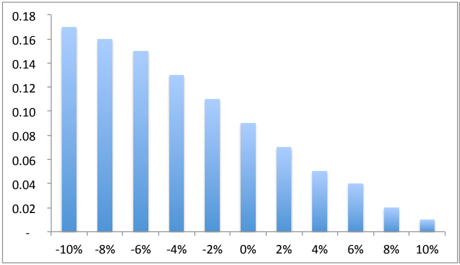
- Let R denote a random variable for the (annual) return on a risky portfolio.

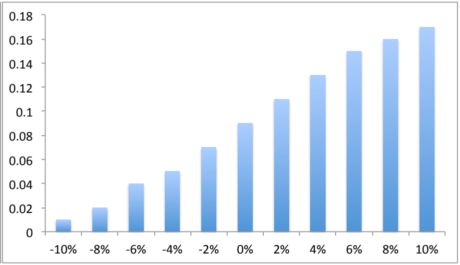
Year	Return
Year 1	12%
Year 2	-3%
Year 3	15%
Year 4	-7%
Year 5	8%

- $\mu = \frac{12-3+15-7+8}{5} = 5\%$.
- $\sigma^2 = \frac{1}{5}(12-5)^2 + \frac{1}{5}(-3-5)^2 + \ldots + \frac{1}{5}(8-5)^2 = 73.2$
- $\sigma = 8.55$.

A Criticism of Standard Deviation

Losses and Gains are symmetrically weighted.





These two distributions have the same standard deviation, but not the same “risk."

Value at Risk (VaR)

Given a time horizon, VaR at a given level α is a threshold value such that the probability that the loss on the portfolio over the time horizon exceeds this value is the given probability level.

Math: The α-quantile $q_\alpha(R) = \min\{r : P(R \leq r) \geq \alpha\} \quad 0 \leq \alpha \leq 1$.
The Value at Risk at level α: $VaR_\alpha(R) = -q_\alpha(R)$.

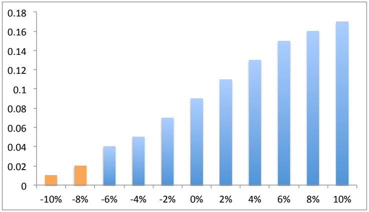


Figure : For $\alpha = 3\%$, $VaR_\alpha(R) = 8\%$.

Widely used by government regulators and banks.

Example

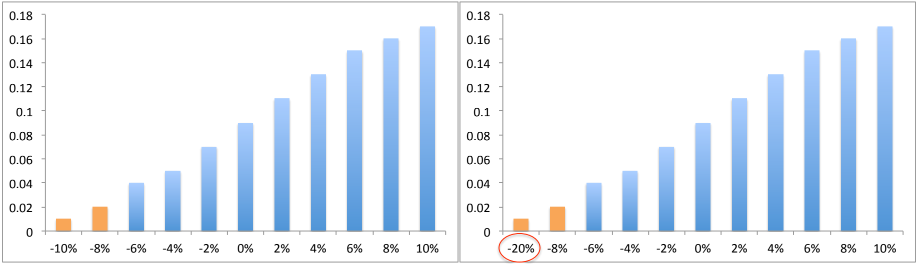
- Historical returns

Year	Return
Year 1	12%
Year 2	-3%
Year 3	15%
Year 4	-7%
Year 5	8%

- For $\alpha = 40\%$, $q_\alpha(R) = -3\%$ since $P(R \leq -3) = 0.4$ and $P(R < -3) < .4$.
- $VaR_\alpha(R) = -q_\alpha(R) = 3\%$.
- Note that if year 4 we lost 25%, still $VaR_\alpha = 3\%$.

Criticisms of Value at Risk

VaR is indifferent to severity of losses beyond the “critical” threshold.



Both distributions have same VaR_α for $\alpha = 3\%$, but different “risks."

Coherent Risk Measures

So what is an appropriate definition of risk?

P. Artzner, F. Delbaen, J. Eber, D. Heath (1999). "Coherent Measures of Risk". Mathematical Finance, 9, 3, 203–228.

$\rho(R)$ is a *coherent risk measure* if it satisfies the following four properties:

Monotonicity

If the return of Portfolio 2 is better than the return of Portfolio 1 under all scenarios, then **the risk** of Portfolio 2 is less than **the risk** of Portfolio 1.

Mathematically: If $R_1 \leq R_2$, then $\rho(R_1) \geq \rho(R_2)$.

Diversification

The risk of two portfolios together cannot be worse than the sum of their risks separately.

Mathematically: $\rho(R_1 + R_2) \leq \rho(R_1) + \rho(R_2)$.

Leverage

If you double your portfolio then you double your risk.

Mathematically: If $\alpha \geq 0$, then $\rho(\alpha R) = \alpha\rho(R)$.

Influence of Cash

If you add cash in your portfolio you reduce your risk by the amount the cash you add.

Mathematically: $\rho(R + \alpha) = \rho(R) - \alpha$.

Is Standard Deviation Coherent?

Consider $R_1 = 10\%$ deterministically,

$$R_2 = \begin{cases} 10\%, & \text{with probability } 1/2, \\ 30\% & \text{with probability } 1/2. \end{cases}$$

- $R_1 \leq R_2$ in all cases.
- $Stdev(R_1) = 0$.
- $Stdev(R_2) = \sqrt{\frac{1}{2}(10-20)^2 + \frac{1}{2}(30-20)^2} = 10\% > 0 = Stdev(R_1)$.
- **Standard Deviation violates monotonicity**; all other properties are satisfied.

Is Value at Risk Coherent?

- Return of first portfolio

$$R_1 = \begin{cases} 10\%, & \text{with probability } 0.6, \\ 3\%, & \text{with probability } 0.4. \end{cases}$$

- For $\alpha = 50\%$, $VaR_\alpha(R_1) = -10\%$, since $P(R_1 \leq 10\%) \geq .5$ and $P(R_1 < 10\%) < .5$.
- Return of second portfolio

$$R_2 = \begin{cases} 2\%, & \text{with probability } 0.6, \\ -10\%, & \text{with probability } 0.4. \end{cases}$$

- For $\alpha = 50\%$, $VaR_\alpha(R_2) = -2\%$, since $P(R_2 \leq 2\%) \geq .5$ and $P(R_2 < 2\%) < .5$.

Is Value at Risk Coherent? (continued)

- Then

$$R_1 + R_2 = \begin{cases} 12\%, & \text{with probability } 0.36, \\ 5\%, & \text{with probability } 0.24, \\ 0\%, & \text{with probability } 0.24, \\ -7\%, & \text{with probability } 0.16. \end{cases}$$

- For $\alpha = 50\%$, $VaR_\alpha(R_1 + R_2) = -5\%$, since $P(R_1 + R_2 \leq 5\%) \geq .5$ and $P(R_1 + R_2 < 5\%) < .5$.
- $VaR_\alpha(R_1 + R_2) > VaR_\alpha(R_1) + VaR_\alpha(R_2)$.
- **VaR does not satisfy the diversification property, i.e., it is not coherent.**
- It turns out that if Returns are normal, then VaR satisfies the diversification property.

Conditional Value at Risk

Natural question: What is the *expected* loss incurred in the α% worst cases in our portfolio? (Contrast to VaR_α)

Conditional Value at Risk:
$cVaR_\alpha$= Expected Loss when Loss is less than or equal to VaR_α.

$$cVaR_\alpha = -E[R \mid R \leq VaR_\alpha(R)].$$

- If R has a continuous distribution, $cVaR_\alpha$ is a coherent risk measure.

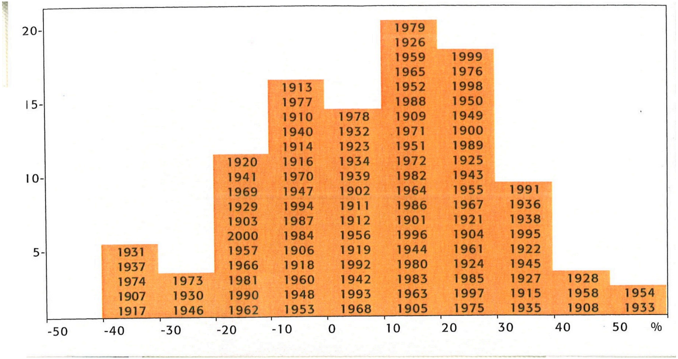
Example (revisited)

Year	Return
Year 1	12%
Year 2	-3%
Year 3	15%
Year 4	-7%
Year 5	8%

- Recall for $\alpha = 40\%$, $VaR_\alpha = 3\%$.
- $cVaR_\alpha = -\frac{1}{2}[-3\% - 7\%] = 5\%$
- $cVaR_\alpha = 5\%$.
- Expected loss in the 40% of the worst case years is 5%.

Real Returns of US Equities in 1900-2000

- Average Real Return: 6.7%
- Maximum annual loss: **Drawdown** $= CVaR = VaR = 40\%$.



Representation theorem for coherent risk measures

- Let $\mathcal{X}$ be the set of all random variables on $\mathbb{R}$ having support with cardinality N. A risk measure $\mu : \mathcal{X} \to \mathbb{R}$ is coherent if and only if there exists a family of probability measures $\mathcal{Q}$ that

$$\mu(X) \quad = \quad \sup_{Q \in \mathcal{Q}} \mathbb{E}_Q\left[X\right], \qquad \forall\, X \in \mathcal{X},$$

  where $\mathbb{E}_Q\left[X\right]$ denotes the expectation of the random variable X under the measure Q (as opposed to the measure of X itself).
- All coherent risk measures may be represented as the worst-case expected value over a family of "generalized scenarios."

Distorted probability measures and the Choquet integral

- Given a random variable $X \in \mathcal{X}$, we define: $S(x) = \mathbb{P}[X \geq x]$.
- A **distortion function** g is any non-decreasing function on $[0, 1]$ such that $g(0) = 0$ and $g(1) = 1$. The **distorted probability distribution** for a random variable $X \in \mathcal{X}$ given by $S(x)$ is the unique distribution defined by $S^*(x) = g(S(x))$.
- The **Choquet integral** of a random variable $X \in \mathcal{X}$ with respect to the distortion function g is defined as

$$\pi_g(X) = \int_0^\infty S^*(x)dx + \int_{-\infty}^0 [S^*(x) - 1]\, dx.$$

- The Choquet integral π_g satisfies monotonicity, translation invariance, and positive homogeneity. In addition, π_g satisfies subadditivity if and only if g is concave. Thus, π_g is coherent if and only if g is concave.

Value-at-risk

- For some $\alpha \in [0,1]$ we definte

$$g(u) = \begin{cases} 0, & \text{if } u < \alpha, \\ 1, & \text{otherwise.} \end{cases}$$

Then we have

$$\mu_g(X) = \int_0^\infty g(S(x))dx = \int_0^{S^{-1}(\alpha)} dx = S^{-1}(\alpha) = \inf\{x \mid \mathbb{P}[X \geq x] \leq \alpha\}$$

value-at-risk at level α, or $\mathrm{VaR}_\alpha(X)$.

- Note that g is not concave, which implies, that value-at-risk is *not* a coherent risk measure.

Conditional value at Risk

- We define $g(u) = \min(u/\alpha, 1)$ for some $\alpha \in [0,1]$. Then we have

$$\begin{aligned}
\mu_g(X) &= \int_0^\infty g(S(x))dx \\
&= \frac{1}{\alpha}\int_{S^{-1}(\alpha)}^{\infty} S(x)dx + \int_0^{S^{-1}(\alpha)} dx \\
&= \frac{x}{\alpha}S(x)\Big|_{S^{-1}(\alpha)}^{\infty} - \frac{1}{\alpha}\int_{S^{-1}(\alpha)}^{\infty} xdS(x) + S^{-1}(\alpha) \\
&= \mathbb{E}[X|X \geq S^{-1}(\alpha)] \\
&= \mathbb{E}[X|X \geq \mathrm{VaR}_\alpha(X)].
\end{aligned}$$

- Note that this is a coherent risk measure since g is concave. It is **conditional value at risk of** X **at level** α and denoted $CVAR_\alpha X$.

Comonotone Risk measures

- Is it true that all coherent risk measures can be represented as a Choquet integral under a concave distortion function?
- The set $A \subseteq \mathbb{R}^n$ is a **comonotonic set** if for all $x \in A$, $y \in A$, we have either $x \leq y$ or $y \leq x$.
- Clearly any one-dimensional set is comonotonic.
- A random variable $X = (X_1, \ldots, X_n)$ is **comonotonic** if its support $A \subseteq \mathbb{R}^n$ is a comonotonic set.
- An example of a comonotonic random variable is the joint payoff of a stock and a call option on the stock. Indeed, let S be the stock value at the exercise time, C be the call value, and K be the strike price. Then $C = \max(0, S - K)$. It is easy to see that any pair of payoffs (S_1, C_1), (S_2, C_2) satisfy either $S_1 \geq S_2$ and $C_1 \geq C_2$ or $S_1 \leq S_2$ and $C_1 \leq C_2$, and hence the support of the random variable (S, C) is comonotonic.

Comonotone Risk measures, continued

- A risk measure $\mu : \mathcal{X} \to \mathbb{R}$ is **comonotonically additive** if, for any comonotonic random variables X and Y, we have

$$\mu(X+Y) \quad = \quad \mu(X) + \mu(Y).$$

- If a coherent risk measure is comonotonically additive, we say the risk measure is **comonotone**.
- A risk measure $\mu : \mathcal{X} \to \mathbb{R}$ can be represented as the Choquet integral with a concave distortion function if and only if μ is comonotone.
- The subadditive property says we can do no worse by aggregating risk when dealing with a coherent risk measure;
- We will not benefit from diversifying risk when our risk measure is a Choquet integral and the underlying random variables are comonotonic.
- We can construct a coherent risk measure which violates comonotonic additivity.

Landscape of Risk Measures

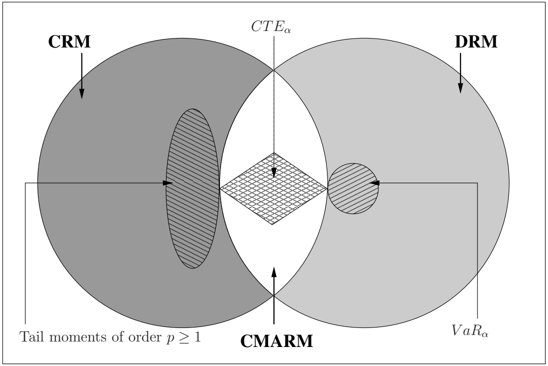


Figure : *Venn diagram of the risk measure universe. The box represents all functions $\mu : \mathcal{X} \to \mathbb{R}$. In bold are the three main classes of risk measures: coherent (CRM), distorted (DRM), and their intersection, comonotone (CMARM). Also illustrated are the specific subclasses CVAR$_\alpha$, VaR$_\alpha$, and tail moments of higher order.*

From coherent risk measures to convex uncertainty sets

- Consider $\tilde{a}'x \geq b$.
- Assumption A: The uncertain vector $\tilde{a}$ has support $\mathcal{A} = \{a_1, \ldots, a_N\}$ and distribution

$$\mathbb{P}[\tilde{a} = a] \quad = \quad \frac{1}{N} \sum_{j=1}^{N} \mathbf{1}(a_j = a),$$

- For a linear optimization problem with uncertain data $\tilde{a}$ and real number b, along with a risk measure μ, we define the **risk averse problem** to be

$$\begin{array}{ll} \text{minimize} & c'x \\ \text{subject to} & \mu(\tilde{a}'x) \leq b. \end{array}$$

Key connection

- If the risk measure μ is coherent and $\tilde{a}$ is distributed as in Assumption A, then the risk averse problem is equivalent to the robust optimization problem

$$\begin{array}{rll} \text{minimize} & c'x & \\ \text{subject to} & a'x \leq b & \forall\ a \in \mathcal{U}, \end{array}$$

where

$$\mathcal{U} = \mathrm{conv}\left(\left\{\sum_{i=1}^{N} q_i a_i \;\middle|\; q \in \mathcal{Q}\right\}\right) \subseteq \mathrm{conv}\left(\mathcal{A}\right),$$

and $\mathcal{Q}$ is the set of generating measures for μ

Proof

$$
\begin{aligned}
\mu(\tilde{a}'x) &= \sup_{q \in \mathcal{Q}} \mathbb{E}_q \left[\tilde{a}'x\right] \\
&= \sup_{q \in \mathcal{Q}} \sum_{i=1}^{N} (a_i'x) q_i \\
&= \sup_{q \in \mathcal{Q}} \left(\sum_{i=1}^{N} q_i a_i\right)' x \\
&= \sup_{a \in \tilde{\mathcal{U}}} a'x \\
&= \sup_{a \in \mathcal{U}} a'x,
\end{aligned}
$$

where $\tilde{\mathcal{U}} = \{\sum_{i=1}^{N} q_i a_i \mid q \in \mathcal{Q}\}$ and $\mathcal{U} = \text{conv}(\tilde{\mathcal{U}})$.

Summary

- The decision maker has some primitive risk measure μ
- If it is coherent, then, there is an *explicit* uncertainty set that should be used in the robust optimization framework.
- This uncertainty set is convex and its structure depends on the generating family $\mathcal{Q}$ for μ and the realizations $\mathcal{A}$ of $\tilde{a}$.

From comonotone measures to polyhedral uncertainty sets

- If $\mathcal{Q}$ is a finite set, then we have the following:

$$\sup_{Q \in \mathcal{Q}} \mathbb{E}_Q[X] = \max_{Q \in \mathcal{Q}} \mathbb{E}_Q[X] = \max_{Q \in \mathrm{conv}(\mathcal{Q})} \mathbb{E}_Q[X],$$

where $\mathrm{conv}(\mathcal{Q})$ denotes the convex hull of $\mathcal{Q}$.

- For a comonotone risk measure with distortion function g on a random variable Y with support $\{y_1, \ldots, y_N\}$ such that $\mathbb{P}[Y = y_i] = 1/N$, we have

$$\mu_g(Y) = \sum_{i=1}^{N} q_i y_{(i)},$$

where $y_{(i)}$ is the ith order statistic of Y, i.e., $y_{(1)} \leq \ldots \leq y_{(N)}$, and

$$q_i = g\left(\frac{N+1-i}{N}\right) - g\left(\frac{N-i}{N}\right).$$

Proof

WLOG $y_i = y_{(i)}$ for all $i \in \{1, \ldots, N\}$. Also $y_1 \geq 0$.

$$S_Y(y) = \begin{cases} 1, & \text{if } y < y_1, \\ \frac{N-i}{N}, & \text{if } y_i \leq y < y_{i+1},\ i = 1, \ldots, N-1, \\ 0, & \text{otherwise.} \end{cases}$$

$$\begin{aligned} \mu_g(Y) &= \int_0^{\infty} g\left(S_Y(y)\right) dy = \int_0^{y_1} g(1) dy + \sum_{i=1}^{N-1} \int_{y_i}^{y_{i+1}} g\left(\frac{N-i}{N}\right) dy \\ &= g(1) y_1 + \sum_{i=1}^{N-1} g\left(\frac{N-i}{N}\right)(y_{i+1} - y_i) \\ &= \sum_{i=1}^{N} \left(g\left(\frac{N-i+1}{N}\right) - g\left(\frac{N-i}{N}\right)\right) y_i \\ &= \sum_{i=1}^{N} q_i y_i = \sum_{i=1}^{N} q_i y_{(i)}. \end{aligned}$$

Connection with LO

In order to compute μ_g for a comonotone risk measure, we need an order statistic on the N possible values of the random variable Y

$$\sum_{i=1}^{N} q_i y_{(i)} = z^*_{q,y},$$

where $z^*_{q,y}$ is the optimal value of the LOP

$$\begin{array}{ll} \text{maximize} & \displaystyle\sum_{i=1}^{N} q_i \sum_{j=1}^{N} w_{ij} y_j \\ \text{subject to} & w \in W(N), \end{array}$$

$$W(N) = \left\{ w \in \mathbb{R}_+^{N^2} \;\middle|\; \sum_{i=1}^{N} w_{ij} = 1 \;\forall j \in \{1, \dots, N\},\; \sum_{j=1}^{N} w_{ij} = 1 \;\forall i \in \{1, \dots, N\} \right\}.$$

The generator $\mathcal{Q}$

$$\mathcal{Q} \quad = \quad \left\{ \sum_{j=1}^{N} w_{ij} q_i \;\middle|\; w \in W(N) \right\},$$

or, alternatively,

$$\mathcal{Q} \quad = \quad \{p \in \mathbb{R}^N \mid \exists\, \sigma \in \mathrm{S}_N \;:\; p_i = q_{\sigma(i)},\ \forall\, i \in \mathcal{N}\},$$

where S_N is the symmetric group on N elements.

Polyhedral uncertainty set

- For a risk averse problem with comonotone risk measure μ_g generated by a measure $q \in \Delta^N = \{p \in \mathbb{R}_+^N \mid e'p = 1\}$ and uncertain vector $\tilde{a}$ distributed as in Assumption A the risk averse problem is equivalent to

$$\begin{array}{rll} \text{minimize} & c'x & \\ \text{subject to} & a'x \leq b & \forall\, a \in \pi_q(\mathcal{A}). \end{array}$$

$$\pi_q(\mathcal{A}) = \operatorname{conv}\left(\left\{ \sum_{i=1}^{N} q_{\sigma(i)} a_i \;\middle|\; \sigma \in \mathrm{S}_N \right\}\right).$$

Comonotone measures and CVAR

- We will show that the set of all comonotone risk measures is finitely generated by the class of conditional value at risk measures.
- Note that as a comonotone risk measure produces generators with $q_i \leq q_{i+1}$ for all $i \in \{1, \ldots, N-1\}$, the space of possible generators q is

$$\tilde{\Delta}^N = \left\{ q \in \Delta^N \mid q_1 \leq \cdots \leq q_N \right\}.$$

Bijection

- There exists a bijection between $\tilde{\Delta}^N$ and the space of comonotone risk measures on random variables with a finite sample space of cardinality N.
- Clearly, any such comonotone risk measure defines a $q \in \tilde{\Delta}^N$ via $q_i = g\left(\frac{N+1-i}{N}\right) - g\left(\frac{N-i}{N}\right)$.
- Conversely, given any $q \in \tilde{\Delta}^N$, we may define a distortion function (on N values) as

$$\begin{aligned} g(0) &= 0, \\ g\left(\frac{i}{N}\right) &= \sum_{j=1}^{i} q_{N-j+1}, \qquad i = 1, \ldots, N. \end{aligned}$$

- One can easily verify that such a g satisfies:

$$\begin{aligned} g(1) &= 1, \\ g(i/N) &\geq g((i-1)/N) \qquad \forall\, i \in \mathcal{N}, \\ g(i/N) - g((i-1)/N) &\leq g((i-1)/N) - g((i-2)/N) \qquad \forall\, i \in \{2, \ldots, N\} \end{aligned}$$

- So g is a valid distortion function corresponding to a comonotone risk measure.

Key Connection

- Theorem: The restricted simplex $\tilde{\Delta}^N$ is generated by the N-member family

  $$\mathcal{G}_N = \{q \in \tilde{\Delta}^N \mid \exists\ k \in \mathcal{N}:\ q_i = 0\ \forall\ i \leq N-k,\ q_i = 1/k\ \forall\ i > N-k\},$$

  i.e., $\mathrm{conv}\,(\mathcal{G}_N) = \tilde{\Delta}^N$. Moreover, each $\hat{q} \in \mathcal{G}_N$ corresponds to the risk measure $CVAR_{i/N}\cdot$ for some $i \in \mathcal{N}$.
- Thus $CVAR_\alpha\cdot$ measures are fundamental.

Summary

- Given a coherent risk measure, we can define a (convex) uncertainty set..
- If further the risk measure is comonotone (related to $CVAR$), the uncertainty set is polyhedral.

15.094J: Robust Modeling, Optimization, Computation

Lecture 18: Constructing Utilities using RO

Learning Preferences

- Kahneman and Tversky [1979] proposed prospect theory as a psychologically more accurate description of preferences compared to expected utility theory
  - Human behavior is inconsistent
  - People are loss averse
  - Gain or loss of an extra unit has less impact (risk averse)

- Our Goal: algorithmize prospect theory to ask questions and compute preferences in a tractable manner

Adaptive Questionnaires

- Items are represented by vectors of attributes

$$\mathbf{x} = (x_1, x_2, \ldots, x_n)$$

- Assume a linear utility function

$$u(\mathbf{x}) = \mathbf{u}'\mathbf{x}, \qquad \mathbf{u} \in \mathcal{U}^0 = [-1, 1]^n.$$

- If question is between $\mathbf{x}$ and $\mathbf{y}$, either

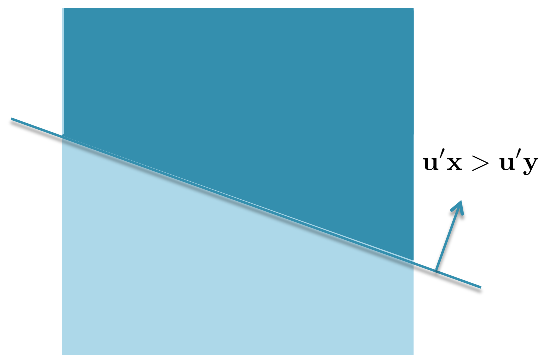
$$\mathbf{u}'\mathbf{x} > \mathbf{u}'\mathbf{y} \quad \text{or} \quad \mathbf{u}'\mathbf{x} < \mathbf{u}'\mathbf{y}$$

Adaptive Questionnaires

- Initial feasible space $\mathcal{U}^0 = [-1, 1]^n$

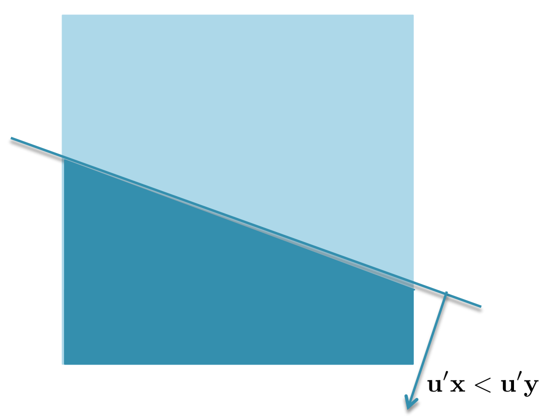
Adaptive Questionnaires

- Initial feasible space $\mathcal{U}^0 = [-1,1]^n$
- Each question will result in a linear inequality
  - $\mathbf{x}$ or $\mathbf{y}$?
    - If $\mathbf{x}$, $\mathbf{u}'\mathbf{x} > \mathbf{u}'\mathbf{y}$



Adaptive Questionnaires

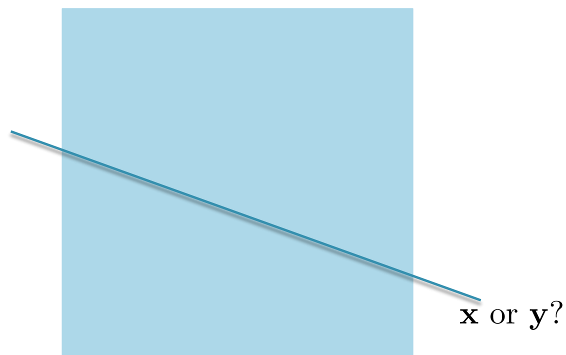
- Initial feasible space $\mathcal{U}^0 = [-1,1]^n$
- Each question will result in a linear inequality
  - $\mathbf{x}$ or $\mathbf{y}$?
    - If $\mathbf{x}$, $\mathbf{u}'\mathbf{x} > \mathbf{u}'\mathbf{y}$
    - If $\mathbf{y}$, $\mathbf{u}'\mathbf{x} < \mathbf{u}'\mathbf{y}$



Adaptive Questionnaires

- Initial feasible space $\mathcal{U}^0 = [-1, 1]^n$

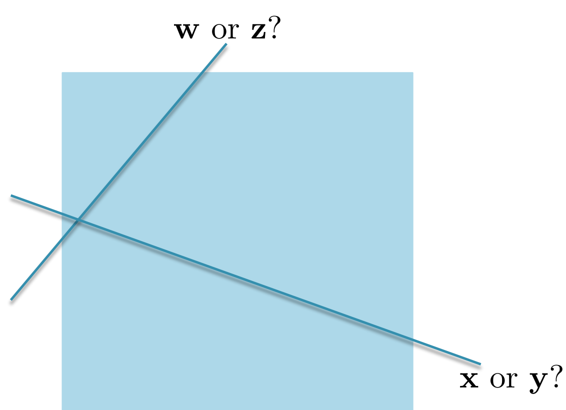
- Each question will result in a linear inequality
  - $\mathbf{x}$ or $\mathbf{y}$?
    - If $\mathbf{x}$, $\mathbf{u}'\mathbf{x} > \mathbf{u}'\mathbf{y}$
    - If $\mathbf{y}$, $\mathbf{u}'\mathbf{x} < \mathbf{u}'\mathbf{y}$



Adaptive Questionnaires

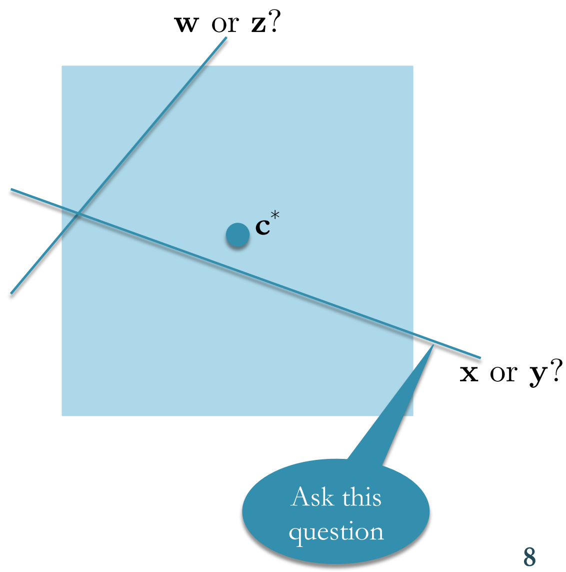
- Initial feasible space $\mathcal{U}^0 = [-1, 1]^n$

- Each question will result in a linear inequality
  - $\mathbf{x}$ or $\mathbf{y}$?
    - If $\mathbf{x}$, $\mathbf{u}'\mathbf{x} > \mathbf{u}'\mathbf{y}$
    - If $\mathbf{y}$, $\mathbf{u}'\mathbf{x} < \mathbf{u}'\mathbf{y}$
  - $\mathbf{w}$ or $\mathbf{z}$?
    - If $\mathbf{w}$, $\mathbf{u}'\mathbf{w} < \mathbf{u}'\mathbf{z}$
    - If $\mathbf{z}$, $\mathbf{u}'\mathbf{w} > \mathbf{u}'\mathbf{z}$



Adaptive Questionnaires

- Initial feasible space $\mathcal{U}^0 = [-1, 1]^n$
- Each question will result in a linear inequality
  - $\mathbf{x}$ or $\mathbf{y}$?
    - If $\mathbf{x}$, $\mathbf{u}'\mathbf{x} > \mathbf{u}'\mathbf{y}$
    - If $\mathbf{y}$, $\mathbf{u}'\mathbf{x} < \mathbf{u}'\mathbf{y}$
  - $\mathbf{w}$ or $\mathbf{z}$?
    - If $\mathbf{w}$, $\mathbf{u}'\mathbf{w} < \mathbf{u}'\mathbf{z}$
    - If $\mathbf{z}$, $\mathbf{u}'\mathbf{w} > \mathbf{u}'\mathbf{z}$
- Pick the question that cuts $\mathcal{U}^0$ closest to the analytic center $\mathbf{c}^*$



Adaptive Questionnaires

- Add inequality with small $\epsilon > 0$

$$\mathbf{u}'\mathbf{x} \geq \mathbf{u}'\mathbf{y} + \epsilon \qquad \text{or} \qquad \mathbf{u}'\mathbf{x} \leq \mathbf{u}'\mathbf{y} - \epsilon$$

- Compute analytic center of new polytope
- Select next question closest to analytic center
- Repeat until question limit or feasible space can not be reduced

Adaptive Questionnaires

- Introduced in marketing [Toubia et al. 2003, 2004]
- Results often suffer from response errors
  - Incorrect responses influence later questions
- We use integer and robust optimization

Self-Correcting Mechanism

- Accounts for inconsistencies and response errors
- Binary variable ϕ_i for each question i $(\mathbf{x}^i > \mathbf{y}^i)$

$$\mathbf{u}'(\mathbf{x}^i - \mathbf{y}^i) + (n + \epsilon)\phi_i \geq \epsilon$$

$$\mathbf{u}'(\mathbf{x}^i - \mathbf{y}^i) + (n + \epsilon)\phi_i \leq n$$

  - If $\phi_i = 0$, constraints are consistent with response
  - If $\phi_i = 1$, a response error or inconsistency is assumed

Compute the Analytic Center

- For the next question, compute the analytic center of

$$\begin{aligned}\mathcal{U}^k = \{\mathbf{u} \in \mathbb{R}^n \mid \; & \mathbf{u}'(\mathbf{x}^i - \mathbf{y}^i) + (\epsilon + n)\phi_i \geq \epsilon, \\ & \mathbf{u}'(\mathbf{x}^i - \mathbf{y}^i) + (\epsilon + n)\phi_i \leq n, \\ & -1 \leq u_j \leq 1, \\ & \sum_{i=1}^{k} \phi_k \leq \gamma k \\ & \phi_i \in \{0, 1\} \quad \}.\end{aligned}$$

Allow a small fraction of the constraints to "flip"

- After k questions, feasible utilities are given by the set $\mathcal{U}^k \subseteq \mathcal{U}^0$
  - $\mathcal{U}^k$ is the projection of a discrete set – union of polyhedra

Loss Averse Preferences

- Previously, the analytic center was taken to be the user's utility vector

- Instead, we solve

$$\max_{\mathbf{x} \in \mathbf{X}} \; \min_{\mathbf{u} \in \mathcal{U}^k} \mathbf{u}'\mathbf{x}$$

where the uncertainty set $\mathcal{U}^k$ is the outcome of a dynamic questionnaire

Robust Optimization with CVaR

- We want to balance robustness and optimality
- Maximize Conditional Value at Risk (CVaR)
  - Definition: CVaR at level α is the expected value of the worst $\alpha\%$ of the utilities in $\mathcal{U}^k$
  - Example:
    - Utilities of 5 different items: (-0.75, -0.25, 0.1, 0.3, 0.8)
      - CVaR(40%) = -0.50
    - Compare to VaR (α-quantile) and Standard Deviation
      - VaR(40%) = -0.25 Doesn't account for -0.75!
      - StdDev = 0.62 Increases for gains and loses!
- CVaR captures the **amount of losses**

Robust Optimization with CVaR

- Challenge: $\mathcal{U}^k$ is the union of polyhedra
  - Fix ϕ to the optimal values in the final computation of the analytic center

$$\Longrightarrow \mathcal{U}^k \text{ becomes a polyhedron}$$

- Challenge: CVaR of $\mathcal{U}^k$ is defined by an integral
  - Approximate $\mathcal{U}^k$ with random sampling
  - We use "Hit-and-Run," starting from $\mathbf{c}^*$ to sample N utility vectors

$$\{\mathbf{u}^1, \mathbf{u}^2, \ldots, \mathbf{u}^N\} \in \mathcal{U}^k$$

  - Polynomial time algorithm, fast in practice [Lovász, Vempala 2006]

Robust Optimization with CVaR

- Given $\{\mathbf{u}^1, \mathbf{u}^2, \ldots, \mathbf{u}^N\} \in \mathcal{U}^k$, solve

$$\begin{aligned} \max_{\mathbf{x} \in X} \quad \min_{\mathbf{y}} \quad & \frac{1}{\alpha N} \sum_{j=1}^{N} (\mathbf{u}_j' \mathbf{x}) y_j \\ \text{s.t.} \quad & \sum_{j=1}^{N} y_j = \alpha N \\ & 0 \leq y_j \leq 1 \end{aligned}$$

- Robust approach: $\alpha = \frac{1}{N}, \quad N \to \infty$

Empirical Evidence

- Experiment
  - Randomly select a "true" utility vector $\mathbf{u}^* \in \mathcal{U}^0$
  - Answer questions according to $\mathbf{u}^*$ with normally distributed noise $\zeta \sim N(0, \sigma)$

$$(\mathbf{u}^*)'(\mathbf{x}^i - \mathbf{y}^i) + \zeta \geq 0 \Longrightarrow \mathbf{x}^i > \mathbf{y}^i$$
$$(\mathbf{u}^*)'(\mathbf{x}^i - \mathbf{y}^i) + \zeta < 0 \Longrightarrow \mathbf{y}^i > \mathbf{x}^i$$

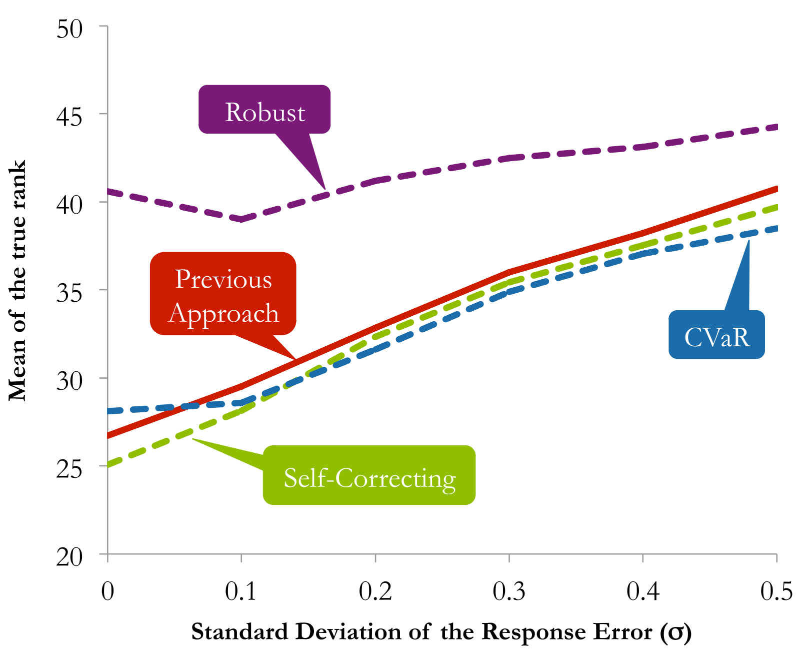
Empirical Evidence

- Rank all items according to $\mathbf{u}^*$ (from 1 to 102)

Item	1	2	3	4	5	6	7	8	9	10	…
Rank	32	3	10	55	2	8	15	4	27	80	…

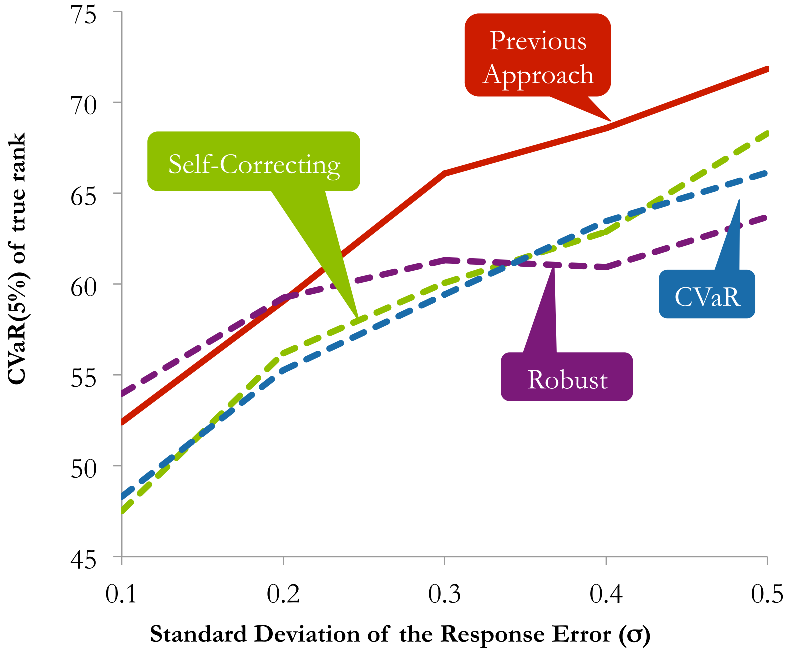
- Solve for the best 5 items according to our algorithm
  - We assume the user will be given choices
  - Example: Items 2, 5, 7, 8, 9
- Compute the average true rank of the solutions we found
  - Example: 10.2
  - Smaller is better

Average Value of the True Ranks



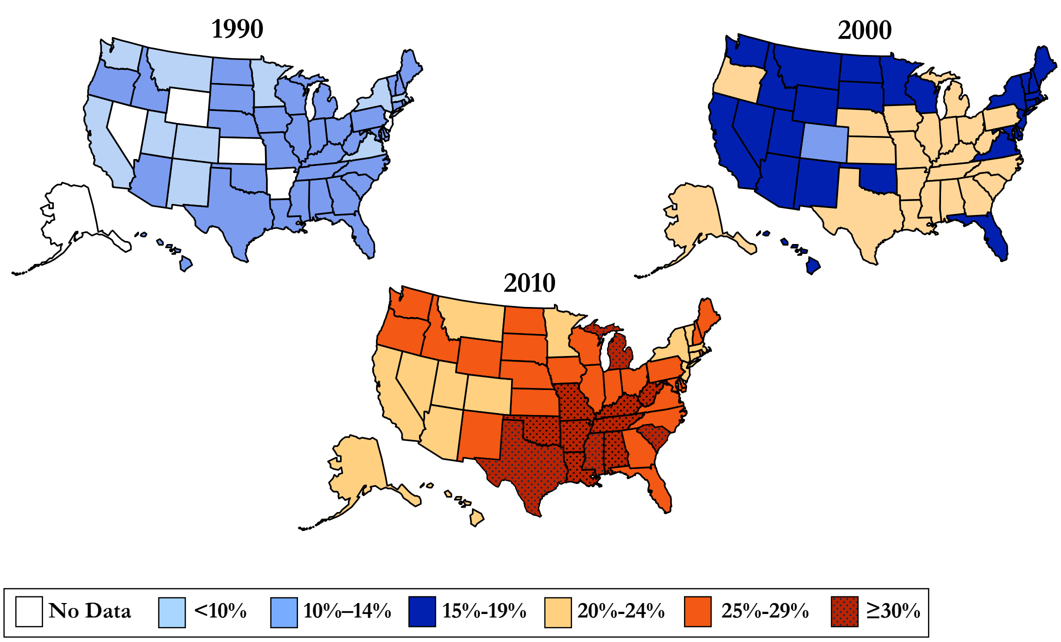
- Average over 10 users
- 15 questions
- $\gamma = 0.1$ for all methods
- $\alpha = 0.1$ for CVaR

CVaR(5%) of the True Ranks



- Average over 10 users
- 15 questions
- $\gamma = 0.1$ for all methods
- $\alpha = 0.1$ for CVaR

Obesity Trends Among U.S. Adults



Source: Behavioral Risk Factor Surveillance System, Centers for Disease Control and Prevention

Obesity and Diabetes in the US

- The incidence of type II diabetes is increasing with obesity
  - 25.8 million people in the US (8.3% of the population)
  - 35% of adults have prediabetes, 50% of adults 65+
- Leading causes of heart disease, stroke, blindness, amputation and kidney failure
- On average,
  - Obesity costs $1,429 more per person in medical costs each year
  - Diabetes costs more than twice as much per person

Diabetes and Dietary Adherence

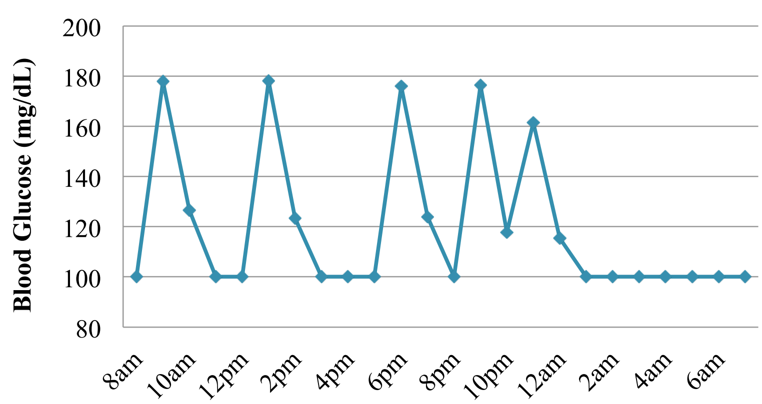
- "Patients with good diabetes self-care behaviors can attain excellent glycemic control…patients with diabetes are especially prone to substantial regimen adherence problems"
  - American Diabetes Association
- "There was substantial agreement by health professionals and patients alike that diet and diet-related issues constituted the most difficult problem faced by persons with diabetes and by health professionals caring for those persons."
  - Michigan Diabetes Research and Training Center
- "One way to improve dietary adherence rates in clinical practice may be to use a broad spectrum of diet options, to better match individual patient food preferences"
  - Dansinger et al., JAMA 2005

Our Approach

- We have developed a **personalized**, **comprehensive**, and **dynamic** system for diabetes management
- Learn food preferences to improve diet adherence
  - Self-Correcting Adaptive Questionnaires
  - A Robust CVaR Approach
- Model blood glucose dynamics
  - People respond differently to different foods
- Propose a daily food and exercise plan

Blood Glucose Dynamics

- Blood glucose level is measured in milligrams per decilitre (mg/dL)
- Blood glucose levels follow a trajectory (BG curve)

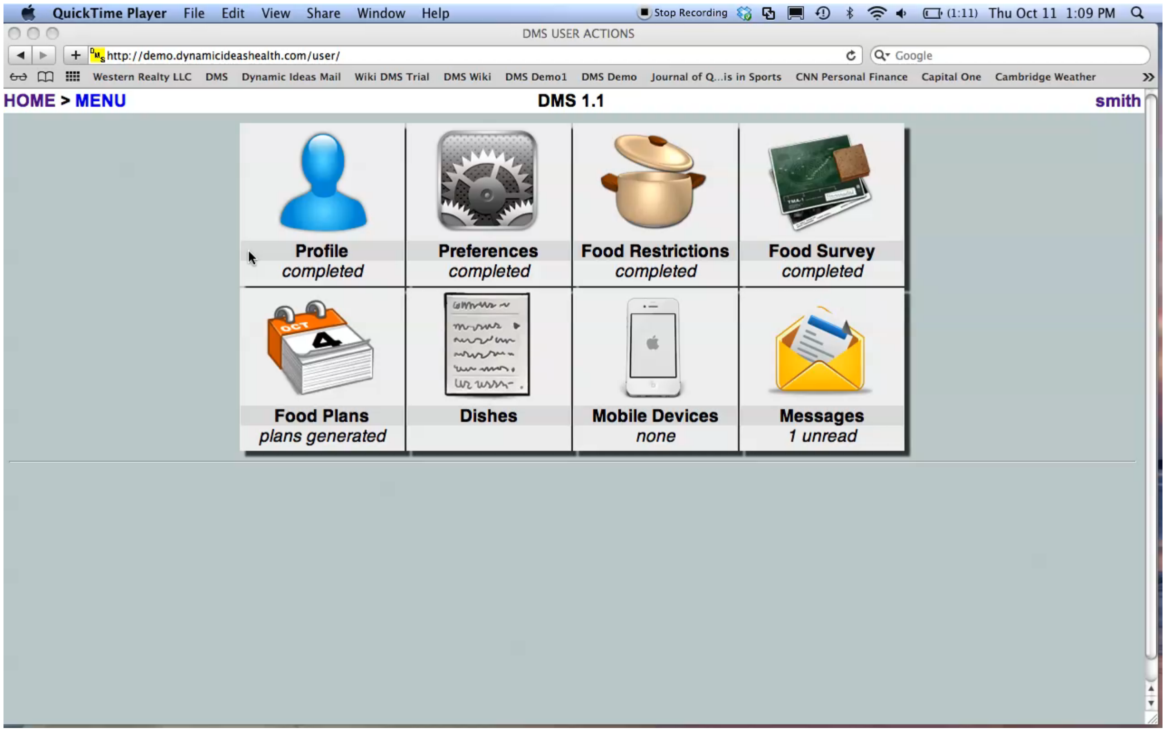


Blood Glucose Modeling

- We model the blood glucose levels as a function of
  - Fasting level
  - Food consumed
    - Classify foods into categories using the glycemic index
    - Measures the effects of carbohydrates on blood glucose levels
  - Exercise performed
- BG measurements can be used to learn the BG curve
  - Regression to update
  - Robust optimization to account for error

Plan Generation with MIO

- Maximize preferences
  - Penalties for high blood glucose levels, nutritional violations
- Bounds on max/min blood glucose levels
- Nutritional requirements
  - Calories, carbs, protein, fat, etc.
- Food group requirements
  - Fruits, vegetables, dairy, meat, starch
- "Appeal" constraints
  - Variety, timing, balanced meals



Summary

- Learn preferences consistent with human behavior
  - Self-Correcting Mechanism
  - A Robust CVaR approach

- An overall system to improve diabetes and diet management

15.094J: Robust Modeling, Optimization, Computation

Lecture 19: Robust Queueing Theory - Single Queue Analysis

Background

- There exists *thousands* of papers in queueing theory, since Erlang [1909].
- Under assumption of Poisson arrivals and Exponential service times, performance analysis is tractable.
- Departing from exponentiality, *steady-state* performance analysis problems become difficult or intractable.
  - *Analysis of $G/G/m$ queue still open*

    Formulated as a multi-dimensional problem in complex plane (Pollaczek [1957])

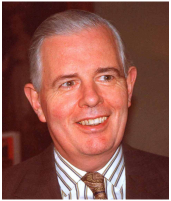
Background

- Moreover, steady-state does not always accurately portray the system's behavior.
  - *Transient behavior due to exogenous changes*, e.g., opening/closing/new control (manufacturing systems with frequent start-up periods)
  - *Slow convergence to steady-state*, e.g., due to heavy tails (internet traffic, call centers and data centers)
- Transient performance analysis problems are even more difficult and intractable.
  - *Analysis of Markovian queues is difficult*

    Use of special functions (Bessel, Hypergeometric)
    Lack of explicit generating functions (Gross & Harris [1974], Keilson [1979])
  - *Approximations and simulations due to lack of tractability*

    Numerical estimation for M/M/1 and M/D/1 queues (Mori [1976])
    Diffusion approximation of GI/GI/1 under heavy traffic (Newell [1971])

Background



"If a queue has an arrival process which cannot be well modeled by a Poisson process or one of its near relatives, it is likely to be difficult to fit any simple model, still less to analyze it effectively. So why do we insist on regarding the arrival times as random variables, quantities about which we can make sensible probabilistic statements? Would it not be better to accept that the arrivals form an irregular sequence, and carry out our calculations without positing a joint probability distribution over which that sequence can be averaged?" – J.F.C. Kingman [2009], Erlang Centennial.

Background

Non-Probabilistic Proposals

- Network Calculus
  - Models queueing primitives via deterministic arrival and service curves (Cruz [1991])
  - Leaky Bucket approach (Gallager and Parekh [1993,1994])
- Adversarial Queues
  - Stability analysis (Goel [1999], Borodin et. al. [2001], Gamarnik [2003])
- Worst-Case approach to performance analysis

Our Proposal and Contribution

- **Proposal:**
  - Replace *probability distributions* with *uncertainty sets* as primitives.
    - To construct uncertainty sets, use *conclusions* of probability theory.
  - Use *worst case analysis*, instead of *expected value analysis* while **bounding** the power of nature/adversary.
    - Optimization instead of Simulation

- **Contribution:**
  - Analysis of Multi-server queueing systems
  - Systems with heavy tailed arrivals and services
  - General networks of queues under steady-state regime (Lecture 20)

Constructing Uncertainty Sets

- We motivate our uncertainty set construction via probability limit laws.
  - Central Limit Theorem (CLT)

    Let $Y_1, Y_2, \ldots$ be a sequence of i.i.d. random variables, with mean μ and variance $\sigma^2 < \infty$, then

$$\frac{\sum_{i=1}^{n} Y_i - n\mu}{\sigma \cdot n^{1/2}} \sim \mathcal{N}(0,1).$$

  - Motivated by the CLT, we deterministically constrain $Y_1, \ldots, Y_n$ to satisfy

$$\mathcal{U} = \left\{ (Y_1, Y_2, \ldots, Y_n) \;\middle|\; \frac{\left| \sum_{i=k+1}^{n} Y_i - (n-k)\mu \right|}{(n-k)^{1/2}} \leq \Gamma, \ \forall k < n \right\}.$$

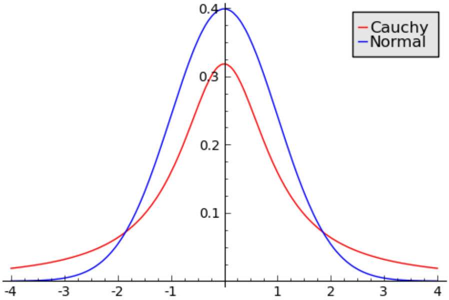
Heavy Tails matter!

- Cloud Computing and Data Centers
  - Heavy tails (Loboz et.al. [2012], Benson et.al. [2010])
  - Non-Poisson arrivals in computer usage (Peterson [1998-2003])
- Internet
  - Self similarity leading to heavy tailed processes (Willinger et al. [1998],Jelenkovic et al. [1997], Kumar et al. [2000])
- Call Centers
  - Heavy tailed arrivals and services (Barabasi [2005])

Modeling Heavy Tails

- To allow higher variability, we introduce a tail coefficient $\alpha \in (1, 2]$, such that

$$\mathcal{U} = \left\{ (Y_1, Y_2, \dots, Y_n) \;\middle|\; \frac{\left| \sum_{i=k+1}^{n} Y_i - (n-k)\mu \right|}{(n-k)^{1/\alpha}} \le \Gamma, \ \forall k < n \right\}.$$



Robust Queue Model

- The interarrival times belong to

$$\mathcal{U}^a = \left\{ (T_1, T_2, \dots, T_n) \left| \frac{\left| \sum_{i=k+1}^{n} T_i - \frac{(n-k)}{\lambda} \right|}{(n-k)^{1/\alpha_a}} \le \Gamma_a, \ \forall k \le n-1 \right. \right\}.$$

- The service times belong to

$$\mathcal{U}^s = \left\{ (X_1, X_2, \dots, X_n) \left| \frac{\left| \sum_{i=k+1}^{n} X_i - \frac{(n-k)}{\mu} \right|}{(n-k)^{1/\alpha_s}} \le \Gamma_s, \ \forall k \le n-1 \right. \right\}.$$

- λ : arrival rate, μ : service rate, Γ_a, Γ_s : variability parameters, α_a, α_s : tail coefficients.

Waiting Time in a Robust Single-Server Queue

- Constraining nature to obey the limit laws, we seek the highest waiting time

$$\widehat{W}_n = \max_{\mathbf{T}\in\mathcal{U}^a,\mathbf{X}\in\mathcal{U}^s} W_n.$$

Theorem

For an initially empty single-server queue with $\rho = \lambda/\mu < 1$, if $\{T_i\}_{i\geq 1} \in \mathcal{U}^a$, $\{X_i\}_{i\geq 1} \in \mathcal{U}^s$, $\alpha_a = \alpha_s = \alpha$, then the highest waiting time of the n^{th} customer can be characterized by

$$\widehat{W}_n = \begin{cases} (\Gamma_a + \Gamma_s)(n-1)^{1/\alpha} - \dfrac{1-\rho}{\lambda}(n-1) & \text{if } n \leq \widehat{n}_s, \\[2ex] \dfrac{\alpha - 1}{\alpha^{\alpha/(\alpha-1)}} \cdot \dfrac{\lambda^{1/(\alpha-1)}.(\Gamma_a + \Gamma_s)^{\alpha/(\alpha-1)}}{(1-\rho)^{1/(\alpha-1)}} & \text{if } n > \widehat{n}_s, \end{cases}$$

where the relaxation number

$$\widehat{n}_s = \left[\frac{\lambda(\Gamma_a + \Gamma_s)}{\alpha(1-\rho)}\right]^{\alpha/(\alpha-1)}$$

Waiting Time in a Robust Single-Server Queue (Proof)

Proof

- The waiting time of the n^{th} job can be expressed recursively in terms of the interarrival and service times using the Lindley recursion

$$W_n = \max\left(W_{n-1} + X_{n-1} - T_n, 0\right) = \max_{1\le j\le n-1}\left(\sum_{\ell=j}^{n-1} X_\ell - \sum_{\ell=j+1}^{n} T_\ell, 0\right).$$

- Thus, $\widehat{W}_n$ can be written as

$$\begin{aligned}\widehat{W}_n &= \max_{\mathbf{X}\in\mathcal{U}^s,\mathbf{T}\in\mathcal{U}^a}\ \max_{1\le j\le n-1}\left(\sum_{\ell=j}^{n-1} X_\ell - \sum_{\ell=j+1}^{n} T_\ell, 0\right)\\ &= \max_{1\le j\le n-1}\ \max_{\mathbf{X}\in\mathcal{U}^s,\mathbf{T}\in\mathcal{U}^a}\left(\sum_{\ell=j}^{n-1} X_\ell - \sum_{\ell=j+1}^{n} T_\ell, 0\right). \qquad (1)\end{aligned}$$

- The sums of the service times and interarrival times are bounded by

$$\sum_{\ell=j}^{n-1} X_\ell \le \frac{n-j}{\mu} + \Gamma_s (n-j)^{1/\alpha}, \quad \sum_{\ell=j+1}^{n} T_\ell \ge \frac{n-j}{\lambda} - \Gamma_a (n-j)^{1/\alpha}. \qquad (2)$$

Waiting Time in a Robust Single-Server Queue (Proof)

- Combining Eqs. (1) and (2), we obtain an one-dimensional concave maximization problem (since $1 < \alpha \leq 2$)

$$\max_{1 \leq j \leq n-1} \left\{ (\Gamma_a + \Gamma_s)(n-j)^{1/\alpha} - \frac{1-\rho}{\lambda}(n-j) \right\}.$$

- Making the transformation $x = n - j$, we obtain

$$\max_{1 \leq x \leq n-1} f(x) = \beta \cdot x^{1/\alpha} - \gamma \cdot x, \tag{3}$$

with $\beta = \Gamma_a + \Gamma_s$ and $\gamma = (1-\rho)/\lambda > 0$, given $\rho < 1$.

- The function $f(.)$ is a strictly concave function of x, monotonically increasing in x until

$$\hat{n}_s = \left(\frac{\beta}{\alpha\gamma} \right)^{\alpha/(\alpha-1)} = \left(\frac{\lambda(\Gamma_a + \Gamma_s)}{\alpha(1-\rho)} \right)^{\alpha/(\alpha-1)},$$

and monotonically decreasing afterwards.

Waiting Time in a Robust Single-Server Queue (Proof)

- We now examine the cases where
  (a) $\hat{n}_s > n-1$, i.e. $n \leq \hat{n}_s$:
  $f(.)$ is monotonically increasing on the interval $[0, n-1]$, and is therefore maximized at $x = n-1$ with optimal objective function

$$\beta(n-1)^{1/\alpha} - \gamma(n-1).$$

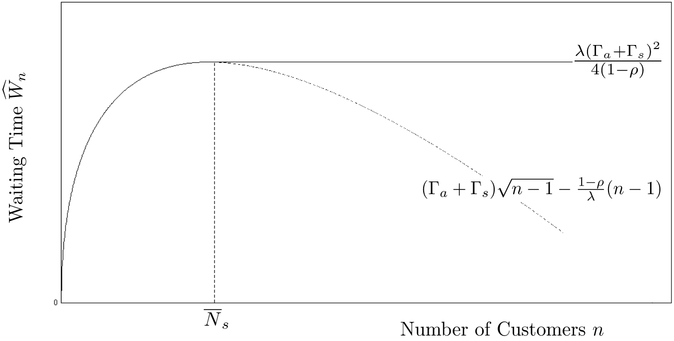
- (b) $\hat{n}_s \leq n-1$, i.e. $n > \hat{n}_s$:
  $\hat{n}_s \in [0, n-1]$, and hence $f(.)$ is maximized at $x = \hat{n}_s$ with optimal objective function

$$\frac{\alpha-1}{\alpha^{\alpha/(\alpha-1)}} \frac{\beta^{\alpha/(\alpha-1)}}{\gamma^{1/(\alpha-1)}}.$$

- The proof is completed by substituting (β, γ) by their respective values.

Insights: Transient and Steady-State Regimes

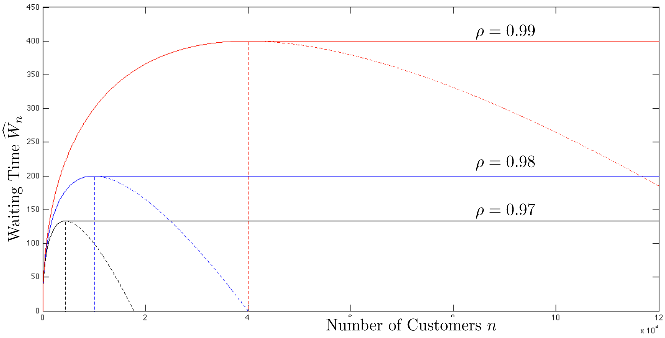
- Transient regime for $n \leq \hat{n}_s$
- Steady-state regime for $n > \hat{n}_s$



Insights: Behavior Under Heavy Traffic

- The higher the traffic intensity, the higher the waiting time and the longer it takes the queue to converge to steady-state.
- For light-tailed arrivals and services ($\alpha = 2$),

$$\widehat{W} \propto \frac{1}{1-\rho} \qquad \hat{n}_s \propto \frac{1}{(1-\rho)^2}$$



Insights: Behavior Under Heavy Tails

- For heavy tailed arrival and service distributions, the waiting time and relaxation number behave as

$$\widehat{W} \propto \frac{1}{(1-\rho)^{1/(\alpha-1)}} \qquad \widehat{n}_s \propto \frac{1}{(1-\rho)^{\alpha/(\alpha-1)}}.$$

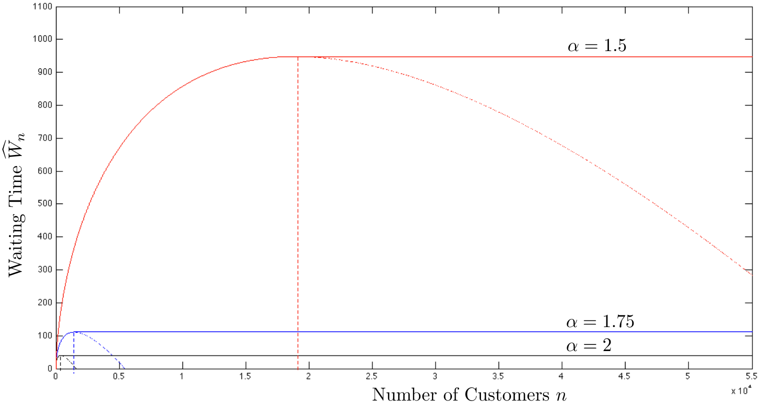
- For example, for $\alpha = 1.5$,

$$\widehat{W} \propto \frac{1}{(1-\rho)^2} \qquad \widehat{n}_s \propto \frac{1}{(1-\rho)^3}$$

which is qualitatively very different from approximating the system's behavior via light tailed processes (such as the Poisson process)

Insights: Behavior Under Heavy Tails

- The heavier the tails, the higher the waiting time and the queue takes much longer to converge to steady-state.



Relaxation Time in Robust Queues

- Relaxation Time
  - Time it takes for the waiting time to reach its steady-state value
  - Time until the arrival of the $\widehat{n}_s^{th}$ customer

$$\widehat{\tau}_s = \sum_{i=1}^{\widehat{n}_s} T_i = \lambda \cdot \left(\frac{\Gamma_a + \Gamma_s}{\alpha(1-\rho)} \right)^{\alpha/(\alpha-1)} + \mathcal{O}\left(\frac{1}{(1-\rho)^{1/(\alpha-1)}} \right)$$

Similarities with Probabilistic Queues

- Similar qualitative behavior for single-server queues as in probabilistic queueing theory.

Robust Approach ($\alpha = 2$)	Probabilistic Approach
$\widehat{W}_n = \begin{cases} (\Gamma_a + \Gamma_s)\sqrt{n} - \dfrac{1-\rho}{\lambda} n & \text{if } n \leq \widehat{n}_s \\ \dfrac{\lambda}{4} \cdot \dfrac{(\Gamma_a + \Gamma_s)^2}{(1-\rho)} & \text{if } n > \widehat{n}_s \end{cases}$	$\mathbb{E}[W_n] \leq \begin{cases} \dfrac{e}{2}\sqrt{\sigma_a^2 + \sigma_s^2}\sqrt{n} & \text{if } n \leq \overline{n}_s \\ \dfrac{\lambda}{2} \cdot \dfrac{\sigma_a^2 + \sigma_s^2}{(1-\rho)} & \text{if } n > \overline{n}_s \end{cases}$
$\widehat{n}_s = \dfrac{\lambda^2}{4} \cdot \dfrac{(\Gamma_a + \Gamma_s)^2}{(1-\rho)^2}$	$\overline{n}_s = \dfrac{\lambda^2}{e^2} \cdot \dfrac{\sigma_a^2 + \sigma_s^2}{(1-\rho)^2}$
$\widehat{\tau}_s \sim \dfrac{\lambda}{4} \cdot \dfrac{(\Gamma_a + \Gamma_s)^2}{(1-\rho)^2}$	$\overline{\tau}_s \sim \lambda \cdot \dfrac{\lambda\sigma_a^2 + \mu\sigma_s^2}{(1-\rho)^2}$

Extensions to Multiple Servers

- Consider a queue with m parallel servers and suppose we are interested in analyzing the performance measures of the queue for the n^{th} customer. Let

$$n = r + m \cdot v,$$

where r is the remainder of the division of n by m.

- We generalize our assumptions regarding the service times uncertainty set as follows.

$$\mathcal{U}_m^s = \left\{ (X_{m+r}, X_{2m+r}, \ldots, X_{vm+r}) \left| \frac{\left| \sum_{i=k+1}^{v} X_{im+r} - \frac{(v-k)}{\mu} \right|}{(v-k)^{1/\alpha_s}} \le \Gamma_s, \ \forall k \le n-1 \right. \right\},$$

where $0 \le r < m$, $1/\mu$ is the expected service time, $\Gamma_s$is a parameter that captures variability information and $1 < \alpha_s \le 2$ models possibly heavy-tailed probability distributions.

Waiting Time in a Robust Multi-Server Queue

Theorem

For an initially empty m-server queue with $\rho = \lambda/m\mu < 1$, if $\{T_i\}_{i\geq 1} \in \mathcal{U}^a$, $\{X_i\}_{i\geq 1} \in \mathcal{U}^s_m$, $\alpha_a = \alpha_s = \alpha$, and $n = r + m \cdot v$, then the highest waiting time

$$\widehat{W}_n = \begin{cases} (\Gamma_a + \Gamma_s/m^{1/\alpha})(n-r)^{1/\alpha} - \dfrac{1-\rho}{\lambda}(n-r) & \text{if } n \leq \overline{N}_m, \\[2ex] \dfrac{\alpha - 1}{\alpha^{\alpha/(\alpha-1)}} \cdot \dfrac{\lambda^{1/(\alpha-1)} . \left(\Gamma_a + \Gamma_s/m^{1/\alpha}\right)^{\alpha/(\alpha-1)}}{(1-\rho)^{1/(\alpha-1)}} & \text{if } n > \overline{N}_m, \end{cases}$$

where the relaxation number

$$\hat{n}_m = r + \left(\frac{\lambda(\Gamma_a + \Gamma_s/m^{1/\alpha})}{\alpha(1-\rho)}\right)^{\alpha/(\alpha-1)}.$$

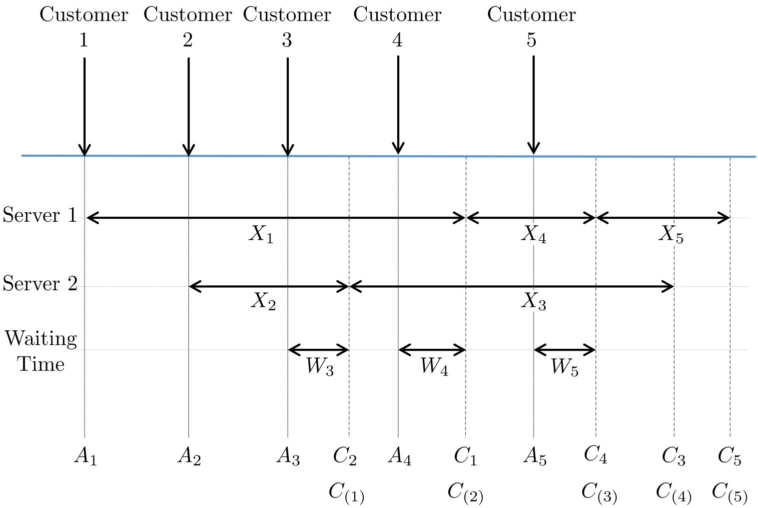
Waiting Time in a Robust Multi-Server Queue

Preliminaries

- Let A_n the arrival time of the n^{th} job where $A_n = \sum_{\ell=1}^{n} T_\ell$ for every n, and C_n the completion time of the n^{th} job, i.e., the time the n^{th} job leaves the system (including service).
- The central difficulty in analyzing probabilistic multi-server queues lies in the fact that overtaking may occur, i.e., the n^{th} departing job is not necessarily the n^{th} arriving job.
- To address this matter, we introduce the ordered sequence of completion times $C_{(1)} \leq C_{(2)} \leq \ldots \leq C_{(n)}$ and define D_n as the n^{th} interdeparture time given by $D_n = C_{(n)} - C_{(n-1)}$.

Waiting Time in a Robust Multi-Server Queue

Preliminaries



Waiting Time in a Robust Multi-Server Queue

Preliminaries

- Looking at the snapshot of the process for five jobs, the waiting times can be found as

$$W_1 = 0, \;\; W_2 = 0, \;\; W_3 = C_2 - A_3 = C_{(1)} - A_3,$$
$$W_4 = C_1 - A_4 = C_{(2)} - A_4, \;\; W_5 = C_4 - A_5 = C_{(3)} - A_5.$$

- By induction, we obtain the general expression of the n^{th} waiting time

$$W_n = \max\{C_{(n-m)} - A_n, 0\}. \tag{4}$$

- Note that

$$C_n = A_n + W_n + X_n = A_n + S_n, \tag{5}$$
$$C_n \geq C_{(n)}, \tag{6}$$
$$C_0 = 0 \;\; \text{and} \;\; C_r = A_r + X_r \;\; \text{for } 1 \leq r \leq m, \tag{7}$$

where $S_n = W_n + X_n$ denotes the sojourn time of the n^{th} job.

Waiting Time in a Robust Multi-Server Queue (Proof)

Proof

- By combining Eqs. (4), (5) and (6), we obtain

$$
\begin{aligned}
C_{(n-m)} &\leq \max\left\{C_{(n-2m)}, A_{n-m}\right\} + X_{n-m} \\
&\leq \max\left\{\max\left\{C_{(n-3m)}, A_{n-2m}\right\} + X_{n-2m}, A_{n-m}\right\} + X_{n-m}, \\
&\leq \max\left\{C_{(n-3m)} + X_{n-2m} + X_{n-m}, A_{n-2m} + X_{n-2m} + X_{n-m}, A_{n-m} + X_{n-m}\right\}.
\end{aligned}
$$

- Given that $n = vm + r$, $0 \leq r < m$,

$$
C_{(n-m)} \leq \max\left\{C_{(n-vm)} + \sum_{k=1}^{v-1} X_{n-km}\,,\; A_{n-(v-1)m} + \sum_{k=1}^{v-1} X_{n-km}\,,\; \ldots\,,\; A_{n-m} + X_{n-m}\right\}.
$$

- The n^{th} waiting time is therefore bounded by

$$
\begin{aligned}
W_n \leq \max\Bigg\{ & C_{(n-vm)} + \sum_{k=1}^{v-1} X_{n-km} - A_n\,,\; A_{n-(v-1)m} + \sum_{k=1}^{v-1} X_{n-km} - A_n\,,\; \ldots\,, \\
& A_{n-m} + X_{n-m} - A_n\,,\; 0\Bigg\}.
\end{aligned}
$$

Waiting Time in a Robust Multi-Server Queue (Proof)

- Note that $n - vm = r$ and $W_r = 0$ yielding $C_{(r)} \leq C_r = A_r + X_r$, for all $0 \leq r < m$. Then,

$$W_n \leq \max\left\{A_r + X_r + \sum_{k=1}^{v-1} X_{(v-k)m+r} - A_n\,,\; A_{m+r} + \sum_{k=1}^{v-1} X_{(v-k)\cdot m+r} - A_n\,,\; \ldots\,,\right.$$
$$\left. A_{n-m} + X_{n-m} - A_n\,,\; 0\right\}.$$

- Expressing the arrival times as $A_n = \sum_{\ell=1}^{n} T_\ell$ for every n, we obtain

$$W_n \leq \max\left\{\sum_{k=1}^{v} X_{(v-k)m+r} - \sum_{\ell=r+1}^{n} T_\ell\,,\; \sum_{k=1}^{v-1} X_{(v-k)m+r} - \sum_{\ell=m+r+1}^{n} T_\ell\,,\; \ldots\,,\right.$$
$$\left. X_{n-m} - \sum_{\ell=(v-1)m+r+1}^{n} T_\ell\,,\; 0\right\}.$$

Waiting Time in a Robust Multi-Server Queue (Proof)

- By substituting $\ell = v - k$, the above expression can be re-written as

$$W_n \leq \max_{0 \leq j \leq v-1} \left\{ \sum_{\ell=j}^{v-1} X_{\ell m+r} - \sum_{\ell=jm+r+1}^{vm+r} T_\ell \, , \, 0 \right\} . \tag{8}$$

  - Note that if we let $m = 1$ we recover the single-server case.
  - Note that the above bound is tight in the case where overtaking does not occur and jobs leave by order of their arrivals, i.e., $C_{(i)} = C_i, \; i \geq 1$.

- Since $\{X_i\}_{i\geq 1} \in \mathcal{U}_m^s$ and $\{T_i\}_{i\geq 1} \in \mathcal{U}^a$

$$\sum_{\ell=j}^{v-1} X_{\ell m+r} \leq \frac{v-j}{\mu} + \Gamma_s (v-j)^{1/\alpha} \; , \quad \sum_{\ell=jm+r+1}^{vm+r} T_\ell \geq \frac{m(v-j)}{\lambda} - m^{1/\alpha} \Gamma_a (v-j)^{1/\alpha} . \tag{9}$$

Waiting Time in a Robust Multi-Server Queue (Proof)

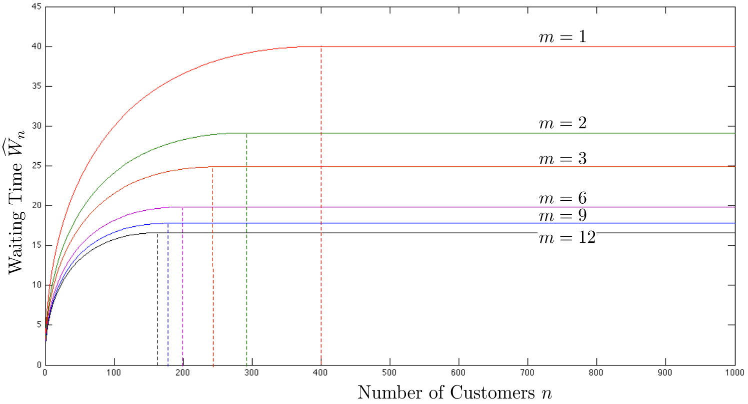
- Combining Eqs. (8) and (9), we obtain an one-dimensional concave maximization problem (since $1 < \alpha \leq 2$)

$$\widehat{W}_n = \max_{0 \leq j \leq \nu - 1} \left\{ \left(m^{1/\alpha} \Gamma_a + \Gamma_s \right) (\nu - j)^{1/\alpha} - \frac{m(1-\rho)}{\lambda} (\nu - j) \right\}.$$

- Note the similarity of this optimization problem with Eq. (3) presented for the single server case. The analysis follows from the proof of the single-server with $\beta = m^{1/\alpha} \Gamma_a + \Gamma_s$ and $\gamma = m(1-\rho)/\lambda$.

Insights: Behavior with Multiple Servers

- The higher the number of servers, the lower the waiting and relaxation times



Similarities with Probabilistic Queues

- Under steady-state, the highest waiting time for the light-tailed robust multi-server queue ($\alpha = 2$) is expressed as

$$\widehat{W} = \frac{\lambda}{4} \cdot \frac{\left(\Gamma_a + \Gamma_s/m^{1/2}\right)^2}{1-\rho}$$

- Similar qualitative insights as Kingman's bound for steady-state waiting time in $G/G/m$ queues

$$\mathbb{E}[W_n] \leq \frac{\lambda}{2} \cdot \frac{\sigma_a^2 + \sigma_s^2/m + (1/m - 1/m^2)/\mu^2}{1-\rho}$$

Summary and Conclusions

- Modeling queues via Uncertainty Sets
  - Captures heavy tails
  - Models multi-servers
- We obtain the following benefits
  - Tractability: Closed form expressions and tractable optimization problems.
  - Generalizability: Transient analysis and Multi server analysis.
- Next: Queueing Networks!

15.094J: Robust Modeling, Optimization, Computation

Lecture 20: Robust Queueing Theory - Queueing Networks Analysis

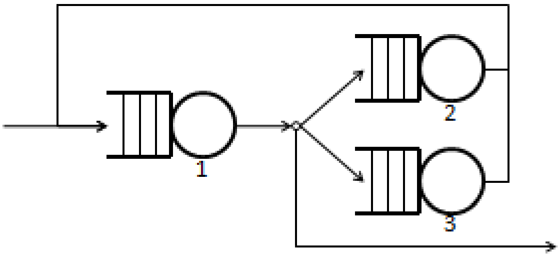
Background

- Under assumption of Poisson arrivals and Exponential service times, performance analysis is tractable.
  - Jackson networks [1957]
  - Kelly networks [1975]
- Departing from exponentiality, *steady-state* performance analysis problems become difficult or intractable.
  - *No tractable theory for networks of $G/G/m$ queues*

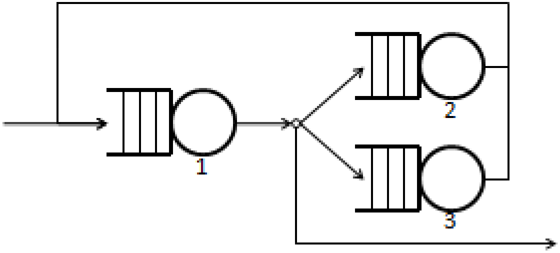
    Lack of Burke's theorem

    Approximations exist: QNA (Whitt [1983])

Queueing Network Analysis



- Need to understand
  1. Queueing Node Operator (Output of a Queue)
  2. Superposition Operator
  3. Thinning Operator



- Need to understand

1. Queueing Node Operator (Output of a Queue)
2. Superposition Operator
3. Thinning Operator

Robust Burke Theorem

- **Main Result**

"When the Arrival Process is a polytope and the Service Process is a polytope, the Departure Process is also a polytope identical to the arrival process polytope."

- **Implications**
  - Generalization of Burke's theorem beyond Markov arrival and service processes
  - Exact analysis of a network of queues can be carried away by regarding the network as a collection of single queues

Robust Burke Theorem

Theorem

If $\{T_i\}_{i\geq 1} \in \mathcal{U}^a$, $\{X_i\}_{i\geq 1} \in \mathcal{U}_m^s$, $\alpha_a = \alpha_s = \alpha$ and $\rho = \lambda/m\mu < 1$, then the interdeparture times $\{D_i\}_{i\geq 1}$ belong to the uncertainty set

$$\mathcal{U}^d = \left\{ (D_1, D_2, \ldots, D_n) \;\middle|\; \frac{\left| \displaystyle\sum_{i=k+1}^{n} D_i - \frac{n-k}{\lambda} \right|}{(n-k)^{1/\alpha}} \leq \Gamma_a + \mathcal{O}\left(\frac{1}{(n-k)^{1/\alpha}} \right), \; \forall k \leq n-1 \right\}.$$

Robust Burke Theorem (Proof)

Proof (single-server case)

- The n^{th} interdeparture time is expressed as

$$D_n = T_n + W_n - W_{n-1} + X_n - X_{n-1} = T_n + S_n - S_{n-1}.$$

- This implies

$$\sum_{i=k+1}^{n} D_i = \sum_{i=k+1}^{n} T_i + S_n - S_k. \tag{1}$$

- Since $S_\ell = W_\ell + X_\ell$, and by the Lindley recursion

$$W_\ell = \max_{1 \le j \le \ell-1} \left(\sum_{i=j}^{\ell-1} X_i - \sum_{i=j+1}^{\ell} T_i \right)$$

the sojourn time can be expressed as

$$S_\ell = \max_{1 \le j \le \ell-1} \left(\sum_{i=j}^{\ell} X_i - \sum_{i=j+1}^{\ell} T_i \right). \tag{2}$$

Robust Burke Theorem (Proof)

- The sum of interdeparture times can be written as

$$\sum_{i=k+1}^{n} T_i - S_k \leq \sum_{i=k+1}^{n} D_i \leq \sum_{i=k+1}^{n} T_i + S_n. \tag{3}$$

- We seek to minimize the left-hand-side and maximize the right-hand side of Eq. (3) over sets $\mathcal{U}^s$ and $\mathcal{U}^a$. Given our assumption on the interarrival times uncertainty set, we can bound

$$\frac{n-k}{\lambda} - \Gamma_a (n-k)^{1/\alpha} \leq \sum_{i=k+1}^{n} T_i \leq \frac{n-k}{\lambda} + \Gamma_a (n-k)^{1/\alpha}$$

Robust Burke Theorem (Proof)

- Bounding the sum of interdepartures, and dividing by $(n-k)^{1/\alpha}$, we obtain

$$-\Gamma_a - \frac{\widehat{S}_k}{(n-k)^{1/\alpha}} \leq \frac{\displaystyle\sum_{i=k+1}^{n} D_i - \frac{n-k}{\lambda}}{(n-k)^{1/\alpha}} \leq \Gamma_a + \frac{\widehat{S}_n}{(n-k)^{1/\alpha}}, \tag{4}$$

where

$$\widehat{S}_\ell = \max_{\mathbf{X}\in\mathcal{U}^s, \mathbf{T}\in\mathcal{U}^a} \left\{ \max_{1\leq j\leq \ell-1} \left(\sum_{i=j}^{\ell} X_i - \sum_{i=j+1}^{\ell} T_i \right) \right\}.$$

- Bounding the sum of service and interarrival times given uncertainty sets $\mathcal{U}^s$ and $\mathcal{U}^a$, *widehat*S_ℓ can therefore be expressed as

$$\widehat{S}_\ell = \max_{1\leq j\leq \ell-1} \left\{ \Gamma_a (\ell-j)^{1/\alpha} + \Gamma_s(\ell-j+1)^{1/\alpha} - (\ell-j)\,\frac{1-\rho}{\lambda} \right\} + \frac{1}{\mu}. \tag{5}$$

Robust Burke Theorem (Proof)

- The one-dimensional concave maximization problem in Eq. (5) is of the form

$$\begin{aligned}\max_{1\leq x\leq \ell-1} \beta\cdot x^{1/\alpha}+\delta\,(x+1)^{1/\alpha}-\gamma\cdot x \quad &\leq \quad \max_{1\leq x\leq \ell-1}(\beta+\delta)(x+1)^{1/\alpha}-\gamma(x+1)+\gamma,\\ &\leq \quad \frac{\alpha-1}{\alpha^{\alpha/(\alpha-1)}}\,\frac{(\beta+\delta)^{\alpha/(\alpha-1)}}{\gamma^{1/(\alpha-1)}}+\gamma, \qquad (6)\end{aligned}$$

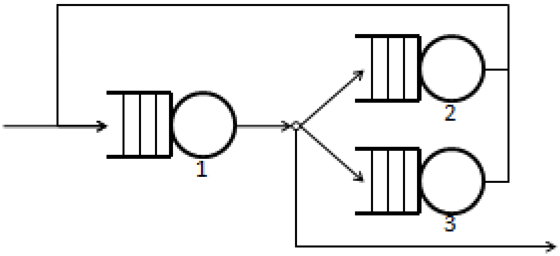
  where $\beta=\Gamma_a$, $\delta=\Gamma_s$, $\gamma=(1-\rho)/\lambda>0$, given $\rho<1$. Note that bound (6) is not tight unless $\ell\geq[(\beta+\delta)/\alpha\gamma]^{\alpha/(\alpha-1)}$.

- For $k<n$ where n is large, substituting β, δ and γ by their respective values yields

$$\begin{aligned}\frac{\widehat{S}_k}{(n-k)^{1/\alpha}}\leq\frac{\widehat{S}_n}{(n-k)^{1/\alpha}}\leq\frac{1}{(n-k)^{1/\alpha}}\left(\frac{\alpha-1}{\alpha^{\alpha/(\alpha-1)}}\cdot\frac{\lambda^{1/(\alpha-1)}\cdot(\Gamma_a+\Gamma_s)^{\alpha/(\alpha-1)}}{(1-\rho)^{1/(\alpha-1)}}+\frac{1}{\lambda}\right)\\ =\mathcal{O}\left(\frac{1}{(n-k)^{1/\alpha}}\right).\end{aligned}$$

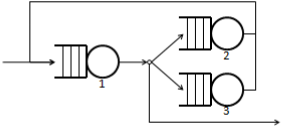
  Applying the above bounds to Eq. (4) completes the proof.

Queueing Network Analysis



- Need to understand
  1. Queueing Node Operator (Output of a Queue)
  2. Superposition Operator
  3. Thinning Operator

Superposition Operator



- Consider a queue j fed by m arrival processes.
- Let $\mathcal{U}_j^a$ denote the uncertainty set representing the interarrival times $\{T_i^j\}_{i \geq 1}$ from arrival process $j = 1, \ldots, J$.
- Denote the uncertainty set of the combined arrival process by $\mathcal{U}_{sup}^a$.
- Given the primitives $(\lambda_j, \Gamma_{a,j}, \alpha)$, $j = 1, \ldots, J$, we define the *superposition operator*

$$(\lambda_{sup}, \Gamma_{a,sup}, \alpha_{sup}) = Combine\Big\{ (\lambda_j, \Gamma_{a,j}, \alpha), j = 1, \ldots, J \Big\},$$

where $(\lambda_{sup}, \Gamma_{a,sup}, \alpha_{sup})$ characterize the merged arrival process $\{T_i\}_{i \geq 1}$.

Superposition Operator

Theorem (Superposition Operator)

The superposition of arrival processes characterized by the uncertainty sets

$$\mathcal{U}_j^a = \left\{ (T_1, \ldots, T_n) \;\middle|\; \frac{\left| \displaystyle\sum_{i=k+1}^{n} T_i - \frac{n-k}{\lambda_j} \right|}{(n-k)^{1/\alpha}} \leq \Gamma_{a,j} \,,\ \forall k \leq n-1 \right\}, \ j = 1, \ldots, J,$$

results in a merged arrival process characterized by the uncertainty set

$$\mathcal{U}_{sup}^a = \left\{ (T_1, \ldots, T_n) \;\middle|\; \frac{\left| \displaystyle\sum_{i=k+1}^{n} T_i - \frac{n-k}{\lambda_{sup}} \right|}{(n-k)^{1/\alpha}} \leq \Gamma_{a,sup} \,,\ \forall k \leq n-1 \right\},$$

where

$$\lambda_{sup} = \sum_{j=1}^{J} \lambda_j, \quad \alpha_{sup} = \alpha, \quad \Gamma_{a,sup} = \frac{\left(\displaystyle\sum_{j=1}^{J} \left(\lambda_j \Gamma_{a,j}\right)^{\alpha/(\alpha-1)} \right)^{(\alpha-1)/\alpha}}{\displaystyle\sum_{j=1}^{J} \lambda_j}.$$

Superposition Operator (Proof)

Proof

Consider the case of superposing two arrival processes, and then generalize the result through induction.

(a) Case where $J = 2$

- Let $\{T_i^j\}_{i\geq 1} \in \mathcal{U}_j^a$, $j = 1, 2$ with

$$\lambda_j \sum_{i=k_j+1}^{n_j} T_i^j \leq (n_j - k_j) + \lambda_j \Gamma_{a,j} (n_j - k_j)^{1/\alpha}, \; j = 1, 2.$$

- Summing over index $j = 1, 2$, we obtain

$$\sum_{i=k_1+1}^{n_1} \lambda_1 T_i^1 + \sum_{i=k_2+1}^{n_2} \lambda_2 T_i^2 \leq (n_1 - k_1 + n_2 - k_2) + \lambda_1 \Gamma_{a,1} (n_1 - k_1)^{1/\alpha} + \lambda_2 \Gamma_{a,2} (n_2 - k_2)^{1/\alpha} \tag{7}$$

Superposition Operator (Proof)

- Consider the time window T between the arrival of the k_1^{th} and the n_1^{th} jobs from arrival process 1, and assume that, within period T, the queue sees arrivals of jobs (k_2+1) up to (n_2-1) from arrival process 2

$$T = \sum_{i=k_1+1}^{n_1} T_i^1 \leq \sum_{i=k_2+1}^{n_2} T_i^2. \tag{8}$$

- During time window T, the queue receives a total of $(n_1 - k_1 + n_2 - k_2)$ jobs, with $(n_1 - k_1 + 1)$ arrivals detected from the first arrival process (including job k_1), and $(n_2 - k_2 - 1)$ arrivals from second arrival process.

- Therefore, period T can also be written in terms of the combined interarrival times $\{T_i\}_{i\geq 1}$ as

$$T = \sum_{i=k+1}^{n} T_i, \tag{9}$$

where $k = k_1 + k_2$ and $n = n_1 + n_2$.

Superposition Operator (Proof)

- Combining Eqs. (8) and (9) yields

$$(\lambda_1+\lambda_2)\sum_{i=k+1}^{n} T_i \leq \lambda_1 \sum_{i=k_1+1}^{n_1} T_i^{1} + \lambda_2 \sum_{i=k_2+1}^{n_2} T_i^{2}$$

which by Eq. (7) can be written as

$$(\lambda_1+\lambda_2)\sum_{i=k+1}^{n} T_i \leq (n-k) + \lambda_1 \Gamma_{a,1}\left(n_1-k_1\right)^{1/\alpha} + \lambda_2 \Gamma_{a,2}\left(n_2-k_2\right)^{1/\alpha}.$$

- Rearranging and dividing both sides by $(\lambda_1+\lambda_2)$ and $(n-k)^{1/\alpha}$, we obtain

$$\frac{\displaystyle\sum_{i=k+1}^{n} T_i - \frac{n-k}{\lambda_{sup}}}{(n-k)^{1/\alpha}} \leq \Gamma_{a,sup}(n,k),$$

where $\lambda_{sup}=\lambda_1+\lambda_2$, $\alpha_{sup}=\alpha$, and

$$\Gamma_{a,sup}(n,k) = \frac{\lambda_1}{\lambda_1+\lambda_2}\Gamma_{a,1}\left(\frac{n_1-k_1}{n_1-k_1+n_2-k_2}\right)^{1/\alpha} + \frac{\lambda_2}{\lambda_1+\lambda_2}\Gamma_{a,2}\left(\frac{n_2-k_2}{n_1-k_1+n_2-k_2}\right)^{1/\alpha}$$

Superposition Operator (Proof)

- By letting

$$x = \frac{n_1 - k_1}{n_1 - k_1 + n_2 - k_2}, \tag{10}$$

the maximum value that $\Gamma_{a,sup}(n,k)$ can achieve over the range of (n,k) can be determined by optimizing the following one-dimensional concave maximization problem over $x \in (0,1)$

$$\max_{x\in(0,1)} \left\{\beta x^{1/\alpha} + \delta\,(1-x)^{1/\alpha}\right\} = \left(\beta^{\alpha/(\alpha-1)} + \delta^{\alpha/(\alpha-1)}\right)^{(\alpha-1)/\alpha}, \tag{11}$$

where

$$\beta = \frac{\lambda_1}{\lambda_1+\lambda_2}\Gamma_{a,1}, \text{ and } \delta = \frac{\lambda_2}{\lambda_1+\lambda_2}\Gamma_{a,2}.$$

- Substituting β and δ by their respective values in Eq. (11) completes the proof for $m = 2$ with

$$\Gamma_{a,sup} = \frac{\left[(\lambda_1\Gamma_{a,1})^{\alpha/(\alpha-1)} + (\lambda_2\Gamma_{a,2})^{\alpha/(\alpha-1)}\right]^{(\alpha-1)/\alpha}}{\lambda_1+\lambda_2}.$$

We refer to this procedure of combining two arrival processes by the operator

$$(\lambda_{sup}, \Gamma_{a,sup}, \alpha_{sup}) = \mathit{Combine}\left\{(\lambda_1, \Gamma_{a,1}, \alpha), (\lambda_2, \Gamma_{a,2}, \alpha)\right\}.$$

Superposition Operator (Proof)

(b) Case for $J > 2$

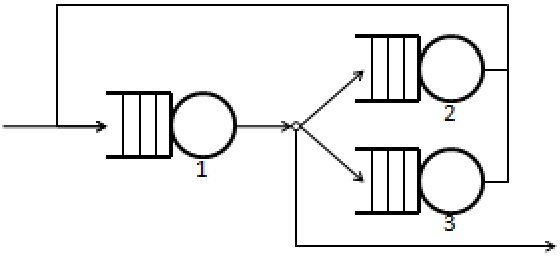
- Suppose that the arrivals to a queue come from arrival processes 1 through $(m-1)$. We assume that the combined arrival process belongs to the proposed uncertainty set, with

$$\overline{\lambda} = \sum_{j=1}^{m-1} \lambda_j \text{ and } \overline{\Gamma}_a = \frac{\left(\sum_{j=1}^{m-1} \left(\lambda_j \Gamma_{a,j}\right)^{\alpha/(\alpha-1)} \right)^{(\alpha-1)/\alpha}}{\overline{\lambda}}.$$

- Extending the proof to m sources can be easily done by repeating the procedure shown in part **(a)** through the operator

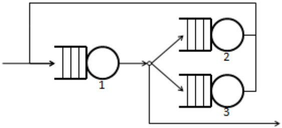
$$(\lambda_{sup}, \Gamma_{a,sup}, \alpha_{sup}) = Combine \left\{ \left(\overline{\lambda}, \overline{\Gamma}_a, \alpha\right), (\lambda_m, \Gamma_{a,m}, \alpha) \right\}.$$

Queueing Network Analysis



- Need to understand
  1. Queueing Node Operator (Output of a Queue)
  2. Superposition Operator
  3. Thinning Operator

Thinning Operator



- Consider an arrival process $\{T_i\}_{i\geq 1}$ in which a fraction f of arrivals are classified as type I and the remaining are classified as type II.
- Given the primitives (λ, Γ_a) at the original process and the fraction f, we define the *thinning operator*

$$(\lambda_{split}, \Gamma_{a,split}, \alpha) = Split\Big\{ (\lambda, \Gamma_a, \alpha), f \Big\}$$

where $(\lambda_{split}, \Gamma_{a,split}, \alpha)$ characterizes the thinned arrival process $\{T_i^{split}\}_{i\geq 1}$.

Thinning Operator

Theorem (Thinning Operator)

The thinned arrival process of a fraction f of arrivals belonging to $\mathcal{U}^a$ is described by the uncertainty set

$$\mathcal{U}^a_{split} = \left\{ (T_1^{split}, \ldots, T_n^{split}) \;\middle|\; \frac{\left| \displaystyle\sum_{i=k+1}^{n} T_i^{split} - \frac{n-k}{\lambda_{split}} \right|}{(n-k)^{1/\alpha}} \leq \Gamma_{a,split}, \quad \forall k \leq n-1 \right\}, \quad (12)$$

where $\lambda_{split} = \lambda \cdot f$ *and* $\Gamma_{a,split} = \Gamma_a \cdot \left(\dfrac{1}{f}\right)^{1/\alpha}$.

Overall Network analysis

Proof

- Consider an arrival process described by $\mathcal{U}^a$ and consider the time window between the k^{th} arrival and the n^{th} arrival. Suppose that a fraction f of these arrivals are type I arrivals, i.e., out of the total of $(n-k)$ arrivals excluding the k^{th} customer, $(n_{split} - k_{split})$ are type I arrivals, such that

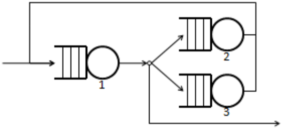
$$f = \frac{n_{split} - k_{split}}{n-k}.$$

- Let $\{T_i^{split}\}_{i\geq 1}$ denote the interarrival times in the thinned arrival process. Note that

$$\sum_{i=k_{split}+1}^{n_{split}} T_i^{split} \leq \sum_{i=k+1}^{n} T_i \leq \frac{n-k}{\lambda} + (n-k)^{1/\alpha}\,\Gamma_a.$$

with the first inequality is tight when the k^{th} and n^{th} customers are both classified as type I. The upper bound in Eq. (12) is obtained by substituting $(n-k)$ by $(n_{split} - k_{split})/f$. The lower bound is derived similarly, hence completing the proof.

Overall Network analysis



- Consider a network of J queues serving a single class of jobs. Each job enters the network through some queue j, and either leaves the network or departs towards another queue right after completion of his service.
- The primitive data in the queueing network are:
  (a) External arrival processes with parameters $(\lambda_j, \Gamma_{a,j}, \alpha_{a,j})$ that arrive to each node $j = 1, \ldots, J$.
  (b) Service processes with parameters $(\mu_j, \Gamma_{s,j}, \alpha_{s,j})$, and the number of servers m_j, $j = 1, \ldots, J$.
  (c) Routing matrix $\mathbf{F} = [f_{ij}]$, $i, j = 1, \ldots, J$, where f_{ij} denotes the fraction of jobs passing through queue i routed queue j. The fraction of jobs leaving the network from queue i is $1 - \sum_j f_{ij}$.

Overall Network analysis

Theorem

The behavior of a single class queueing network is equivalent to that of a collection of independent queues, with the arrival process to node j characterized by the uncertainty set

$$\mathcal{U}_j^a = \left\{ (T_1^j, \ldots, T_n^j) \;\middle|\; \frac{\left| \sum_{i=k+1}^{n} T_i^j - \frac{n-k}{\overline{\lambda}_j} \right|}{(n-k)^{1/\alpha}} \leq \overline{\Gamma}_{a,j}, \quad \forall k \leq n-1 \right\}, \;\; j = 1 \ldots, J,$$

where $\{\overline{\lambda}_1, \overline{\lambda}_2, \ldots, \overline{\lambda}_J\}$ and $\{\overline{\Gamma}_{a,1}, \overline{\Gamma}_{a,2}, \ldots, \overline{\Gamma}_{a,j}\}$ *satisfy the set of equations for all* $j = 1, \ldots, J$

$$\overline{\lambda}_j = \lambda_j + \sum_{i=1}^{J} \left(\overline{\lambda}_i f_{ij} \right), \tag{13}$$

$$\overline{\Gamma}_{a,j} = \frac{\left[\left(\lambda_j \cdot \Gamma_{a,j}\right)^{\alpha/(\alpha-1)} + \sum_{i=1}^{J} \left(\overline{\lambda}_i \cdot \overline{\Gamma}_{a,i} \right)^{\alpha/(\alpha-1)} \cdot f_{ij} \right]^{(\alpha-1)/\alpha}}{\overline{\lambda}_j}. \tag{14}$$

Overall Network analysis (Proof)

Proof

- Consider a queue j receiving jobs from
  - **(a)** external arrivals described by parameters $(\lambda_j, \Gamma_{a,j}, \alpha)$, and
  - **(b)** internal arrivals routed from queues $i, i = 1, \ldots, J$ resulting from splitting the effective departure process from queue i by f_{ij}. By the Robust Burke theorem, the effective departure process from queue i has the same form as the effective arrival process to queue i described by the parameters $\left(\overline{\lambda}_i, \overline{\Gamma}_{a,i}, \alpha\right)$.
- The effective arrival process to queue j can therefore be represented as

$$\left(\overline{\lambda}_j, \overline{\Gamma}_{a,j}, \alpha\right) = \mathit{Combine}\left\{ (\lambda_j, \Gamma_{a,j}, \alpha), \left(\mathit{Split}\left\{ \left(\overline{\lambda}_i, \overline{\Gamma}_{a,i}, \alpha\right), f_{ij}\right\}\right), i = 1, \ldots, J\right\} \quad (15)$$

- By the Splitting Operator, we substitute the split processes by their resulting parameters and obtain the superposition of $J+1$ arrival processes

$$\left(\overline{\lambda}_j, \overline{\Gamma}_{a,j}, \alpha\right) = \mathit{Combine}\left\{ (\lambda_j, \Gamma_{a,j}, \alpha), \left(f_{ij}\overline{\lambda}_i, \overline{\Gamma}_{a,i}\left(\frac{1}{f_{ij}}\right)^{1/\alpha}, \alpha\right), i = 1, \ldots, J\right\} \quad (16)$$

- Applying the Combine Operator yields Eqs. (13) and (14).

Overall Network analysis: Solving a Linear System

- Note that finding the overall network parameters $(\overline{\lambda}, \overline{\Gamma})$ amounts to solving a set of linear equations
- This could be achieved by defining

$$x_j = \left(\overline{\lambda}_j \overline{\Gamma}_{a,j}\right)^{\alpha/(\alpha-1)}$$

and rewriting the system as

$$\begin{cases} \overline{\lambda}_j = \lambda_j + \sum_{i=1}^{J} \left(\overline{\lambda}_i P_{ij}\right) & \forall j \\ x_j = \left(\lambda_j \Gamma_{a,j}\right)^{\alpha/(\alpha-1)} + \sum_{i=1}^{J} P_{ij} x_i & \forall j \end{cases}$$

Computational Results

Objectives

- Compare the performance of RQNA to the Queueing Network Analyzer (QNA) proposed by Whitt [1983] and simulations
- Investigate the relative performance of RQNA with respect to
    - system's network size and degree of feedback
    - maximum traffic intensity
    - diversity of external arrival distributions

Primitive Data

- Consider instances of stochastic queueing networks
- External arrivals $(\lambda_j, \sigma_{a,j}, \alpha_{a,j})$ and $c^2_{a,j} = \lambda^2_j \sigma^2_{a,j}$
- Service processes $(\mu_j, \sigma_{s,j}, \alpha_{s,j}, m_j)$ and $c^2_{s,j} = \mu^2_j \sigma^2_{s,j}$
- Routing matrix $\mathbf{P} = [P_{ij}]$

Precomputing Robust Variability Parameters

Similarly to QNA, we use simulation to construct the parameters (Γ_a, Γ_s)

- Consider a single queue with m servers characterized by $(\rho, \sigma_a, \sigma_s, \alpha_a, \alpha_s)$ and model

$$\Gamma_a = \sigma_a \text{ and } \Gamma_s = f(\rho, \sigma_a, \sigma_s, \alpha_a, \alpha_s).$$

- Motivated by Kingman's bound, we consider the functional form $f(\cdot)$

$$f(\rho, \sigma_s, \sigma_a, \alpha_a, \alpha_s) = \left(\theta_0 + \theta_1 \cdot \sigma_s^2/m + \theta_2 \cdot \sigma_a^2 \rho^2 m\right)^{(\alpha-1)/\alpha} - \sigma_a m^{(\alpha-1)/\alpha}$$

- For each service distribution, we run simulation over multiple instances of a single queue while varying parameters $(\rho, \sigma_a, \sigma_s, \alpha_a, \alpha_s)$ for different arrival distributions to compute corresponding coefficients $(\theta_0, \theta_1, \theta_2)$.

The RQNA Algorithm

ALGORITHM 1. **Robust Queueing Network Analyzer**

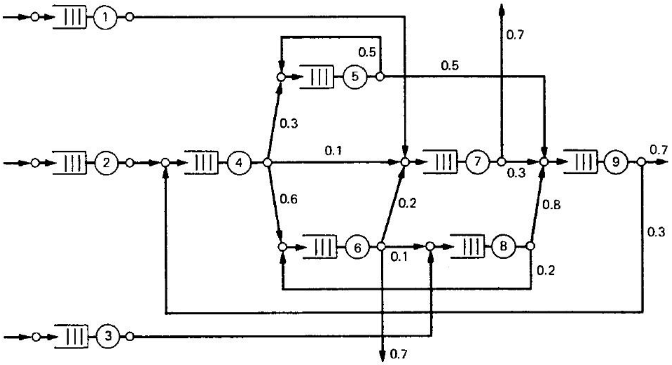
Input: External arrivals $(\lambda_j, \sigma_{a,j}, \alpha_{a,j})$, service parameters $(\mu_j, \sigma_{s,j}, \alpha_{s,j})$, and routing matrix $\mathbf{P} = [P_{ij}]$

Output: Waiting times $\widehat{W}$ at each node j, $j = 1, \ldots, J$.

1. For each external arrival process i in the network, set $\Gamma_{a,i} = \sigma_{a,i}$.
2. For each queue j in the network with parameters $(\mu_j, \sigma_{s,j}, \alpha_{s,j})$, compute
   - **(a)** the effective parameters $\left(\overline{\lambda}_j, \overline{\Gamma}_{a,j}, \overline{\alpha}_{a,j}\right)$ and set $\rho_j = \overline{\lambda}_j / \mu_j$,
   - **(b)** the variability parameter $\Gamma_{s,j} = f\left(\rho_j, \overline{\Gamma}_{a,j}, \sigma_{s,j}, \overline{\alpha}_{a,j}, \alpha_{s,j}\right)$,
   - **(c)** the waiting time $\widehat{W}$ at node j.

Performance of RQNA Compared to QNA and Simulation

- Kuehn Network with nine single-server queues
- Simulations run for normal and Pareto distributed service times



Performance of RQNA Compared to QNA and Simulation

- Kuehn Network with nine single-server queues
- Simulations run for normal and Pareto distributed service times

Case	Pareto Distribution		Normal Distribution	
$(c_{a,j}^2, c_{s,j}^2)$	QNA	RQNA	QNA	RQNA
$(0.25, 0)$	22.78	3.291	15.28	1.389
$(0.25, 1)$	18.48	-3.478	12.08	3.869
$(0.25, 4)$	20.13	-3.052	11.57	-3.882
$(1, 0)$	19.01	1.056	12.68	-3.797
$(1, 1)$	14.06	1.799	5.84	-2.555
$(1, 4)$	10.15	2.893	-10.45	-0.681
$(4, 0)$	21.82	-1.934	10.95	1.290
$(4, 1)$	23.71	-2.139	14.18	-3.508
$(4, 4)$	17.51	-2.974	11.55	1.671

Table : Single-Server Network: Sojourn time percent errors relative to simulation.

Performance of RQNA as a Function of Network Parameters

- Randomly generated networks of queues.
- Queues in the network are randomly assigned 3, 6, or 10 servers independently of each other.

Performance with respect to networks size and degree of feedback

% Loops/Nodes	n=10	n=15	n=20	n=25	n=30
0%	3.594	3.546	3.756	3.432	3.846
20%	3.696	4.014	4.02	4.392	4.452
35%	4.32	4.776	4.956	5.034	4.878
50%	4.95	4.806	5.358	5.67	6.192
70%	5.016	5.556	5.934	5.958	6.03

Table : Multi-Server Networks: Percent error versus network size and degree of feedback

Performance of RQNA as a Function of Network Parameters

- Randomly generated networks of queues.
- Queues in the network are randomly assigned 3, 6, or 10 servers independently of each other.

Performance with respect to traffic intensity and arrival distributions

# Arr. Dist.	$\rho = 0.95$	$\rho = 0.9$	$\rho = 0.8$	$\rho = 0.65$	$\rho = 0.5$
1	4.05	4.092	3.618	3.678	3.228
2	5.082	7.104	6.42	6.108	3.714
3	5.916	6.318	6.9	7.344	5.676
4	7.672	8.644	7.284	6.852	5.37

Table : Multi-Server Networks: Percent error versus traffic intensity and arrival distributions.

Summary of Computational Results

- RQNA produces results that are often significantly closer to simulated values compared to QNA.
- RQNA is somewhat sensitive to the degree of diversity of external arrival disributions
- RQNA is to a large extent insensitive to the
  - number of servers per queue
  - heavy-tailed nature of services
  - network size
  - traffic intensity

Summary and Conclusions

- Explored an alternative approach to model single-class queues by modeling primitive data through uncertainty sets
- Robust approach yields closed-form solutions for the waiting times in multi-server queues for heavy-tailed arrival and service processes operating under both transient and steady-state domains
- Analysis extends to arbitrary networks of queues through the key principle: (a) the departure from a queue, (b) the superposition, and (c) the thinning of arrival processes have the same uncertainty set representation as the original arrival processes

Summary and Conclusions

- Modeling queues via Uncertainty Sets
  - Capture heavy tails
  - Model multi-servers
- We obtain the following benefits
  - Tractability: Closed form expressions and tractable optimization problems.
  - Generalizability: Multi server analysis, Robust Burke theorem, Transient analysis, etc.
  - Accuracy: Computational results - errors within 8%.

15.094J: Robust Modeling, Optimization, Computation

Lecture 21: Robust Optimal Auctions

Outline

Mechanism Design

- Mechanism Design is an area in economics and game theory that has an engineering perspective.
- The goal is to design *economic mechanisms* or *incentives* to implement desired objectives (social or individual) in a strategic setting.
- Mechanism design has important applications in economics (e.g., design of voting procedures, markets, auctions), and more recently finds applications in E-commerce (ebay, ad-auctions).

Auction Theory

- Auction theory is part of Mechanism design theory.
- An auction is one of many ways that a seller can use to sell an object to potential buyers with unknown values.
- Participants: auctioneer, bidders.
- In an auction, the object is sold at a price determined by competition among buyers according to rules set by the seller (auction format), but the seller can use other methods.
- Auction Theory, extensive literature developed in Economics, and Computer Science (more recently).
- Two Nobel Prizes, Vickrey (1996) and Myerson (2007).

Example Auctions

- Open-outcry: ascending, descending
  - Ascending (English): Auctioneer announces ever increasing prices to solicit bids. Continues until only one person left in.
  - Descending (Dutch): Auctioneer announces decreasing prices until someone puts up their hand.
- Sealed-bid: Everyone puts bids in envelopes and gives to seller at the same time.
  - Two types: first-price, second-price
- Internet: EBay.com, Amazon.com, Liquidation.com
- Government: Treasury Bills, mineral rights (e.g. oil fields), assets (e.g. privatization), Electromagnetic spectrum
- Stock Market: IPOs, Opening Bell everyday
- Auctions are everywhere!

Optimal Auction Design

- Design an auction to maximize revenue of the auctioneer.
- Myerson [1981] characterized the optimal auction when
  - Buyers' valuations are sampled from independent probability distributions,
  - Buyers have no budget constraints.
- The optimal auction is a second price auction with a reserve:
  - Bidders submit their bids. If all the bids are less than the reserve, the auction is cancelled.
  - The highest bidder is allocated the item and is charged the second highest bid.

An example

- Inverted Jenny unique Plate Block sold for $3 million in a NY 2005 auction.
- A reserve was placed.
- The highest bidder won, and paid the second highest price.
- As per Myerson (1981), Nobel prize in Economics 2007.
- But happens if many stamps (part of a collection) are being auctioned?

Myerson Auction

- The reservation price is calculated by solving a non-linear equation

$$\frac{1-F(r)}{f(r)} = r,$$

  where $F(\cdot)$ is the cdf and $f(\cdot)$ is the pdf of the probability distribution.
- When distributions are not identical, then the reservation price varies with the bidder.
- Myerson auction not optimal for *correlated* valuations and when bidders have *budgets*.

Auctions in the real world

In the real world:

- Typical auctions involve multiple items.
- Bidders have budgets.
- The valuations are correlated.

In these situations, the overall problem is open.

- This is due to the multi-dimensional nature of the problem.
- Modeling using probability distributions leads to this analytical intractability.

Modeling uncertainty in valuations

- For each item $j \in \mathcal{M}$, we model the auctioneer's beliefs on valuations for item j using an uncertainty set $\mathcal{U}_j \in \mathbb{R}^n$.
- Example: Central Limit Theorem states that the normalized sum of random variables

$$\frac{S_n - n\mu}{\sigma \cdot \sqrt{n}}$$

is asymptotically standard normal.

$$\mathcal{U}_j^{\mathrm{CLT}} = \left\{ (v_{1j}, \ldots, v_{nj}) \,\middle|\, -\Gamma \le \frac{\sum_{i=1}^{n} v_{ij} - n \cdot \mu_j}{\sigma_j \cdot \sqrt{n}} \le \Gamma. \right\}$$

Modeling uncertainty in valuations

- Factor model : $\{\tilde{z}_i\}_{i=1,\dots,n}$ depend on m factors $\tilde{f} = \left(\tilde{f}_1, \tilde{f}_2, \dots, \tilde{f}_m\right)$

  $$\tilde{z}_i = \mathbf{a}_i' \cdot \tilde{f} + \tilde{\epsilon}_i,$$

  $\{\tilde{\epsilon}_i\}$ are i.i.d.

- $$\mathcal{U}^{\text{Corr}} = \left\{ (z_1, \dots, z_n) \;\middle|\; \begin{array}{ll} z_i = \displaystyle\sum_{j=1}^{m} a_{ij} f_j + \epsilon_i, & \forall i = 1, \dots, n, \\ -\Gamma_f \leq \dfrac{\displaystyle\sum_{j=1}^{m} f_j - m \cdot \mu_f}{\sigma_f \cdot \sqrt{m}} \leq \Gamma_f, & \\ -\Gamma_\epsilon \leq \dfrac{\displaystyle\sum_{i=1}^{n} \epsilon_i - n \cdot \mu_\epsilon}{\sigma_\epsilon \cdot \sqrt{n}} \leq \Gamma_\epsilon. & \end{array} \right\}.$$

Main Problem

- n buyers, indexed by $i \in \mathcal{N}$, are interested in buying a set of m items, indexed by $j \in \mathcal{M}$ sold by an auctioneer.
- Buyer $i \in \mathcal{N}$ has a valuation v_{ij} for item $j \in \mathcal{M}$, which is not known to the auctioneer, and beliefs modeled by uncertainty sets $\mathcal{U}_j$.
- Buyers are budget constrained with budgets $\{B_1, B_2, \ldots, B_n\}$.

Main Problem

- n buyers, indexed by $i \in \mathcal{N}$, are interested in buying a set of m items, indexed by $j \in \mathcal{M}$ sold by an auctioneer.
- Buyer $i \in \mathcal{N}$ has a valuation v_{ij} for item $j \in \mathcal{M}$, which is not known to the auctioneer, and beliefs modeled by uncertainty sets $\mathcal{U}_j$.
- Buyers are budget constrained with budgets $\{B_1, B_2, \ldots, B_n\}$.

Problem

Design an auction mechanism that is
(1) individually rational,
(2) budget feasible, and
(3) "worst case" optimal.

Optimization Problem Formulation

- $\mathbf{v} = (\mathbf{v}_j)_{j \in \mathcal{M}}$
- $\mathbf{v} \in \mathcal{U}$ refer to $\mathbf{v}_j \in \mathcal{U}_j$, $j \in \mathcal{M}$.
- $x_{ij}^{\mathbf{v}}$ is the fraction of item j allocated to buyer i
- $p_i^{\mathbf{v}}$ is the total payment charged to Buyer i, when the bid is $\mathbf{v}$.
- Properties :
  - (a) *Individual Rationality (IR)* : Buyers do not derive negative utility by participating in the auction.
  - (b) *Budget Feasibility (BF)* : Buyers are charged within their budget constraints.
  - (c) *Incentive Compatibility (IC)* : The total utility of the i^{th} buyer under truthful bidding is greater or equal to the total utility that Buyer i derives by bidding any other other bid vector $\mathbf{u}_i$.

Optimization Problem Formulation

- Call the optimization problem OPT.

$$
\begin{aligned}
Z^* = \max \quad & W \\
\text{s.t.} \quad & W - \sum_{i \in \mathcal{N}} p_i^{\mathbf{v}} \le 0, \;\; \forall \mathbf{v} \in \mathcal{U}, \\
& \sum_{i \in \mathcal{N}} x_{ij}^{\mathbf{v}} \le 1, \;\; \forall j \in \mathcal{M}, \forall \mathbf{v} \in \mathcal{U}, \\
\text{(IC)} \quad & \sum_{j \in \mathcal{M}} v_{ij} \cdot x_{ij}^{(\mathbf{u}_i, \mathbf{v}_{-i})} - p_i^{(\mathbf{u}_i, \mathbf{v}_{-i})} - \sum_{j \in \mathcal{M}} v_{ij} \cdot x_{ij}^{(\mathbf{v}_i, \mathbf{v}_{-i})} \\
& \quad + p_i^{(\mathbf{v}_i, \mathbf{v}_{-i})} \le 0, \;\; \forall (\mathbf{v}_i, \mathbf{v}_{-i}) \in \mathcal{U}, \forall (\mathbf{u}_i, \mathbf{v}_{-i}) \in \mathcal{U}, \forall i \in \mathcal{N}, \\
\text{(BF)} \quad & p_i^{\mathbf{v}} \le B_i, \;\; \forall i \in \mathcal{N}, \forall \mathbf{v} \in \mathcal{U}, \\
\text{(IR)} \quad & p_i^{\mathbf{v}} \le \sum_{j \in \mathcal{M}} v_{ij} \cdot x_{ij}^{\mathbf{v}}, \;\; \forall i \in \mathcal{N}, \forall \mathbf{v} \in \mathcal{U}, \\
& \mathbf{x}^{\mathbf{v}} \ge \mathbf{0}.
\end{aligned}
$$

- Z^* is the "worst case optimal" revenue that we intent to secure.

Robust Optimal Mechanism

- We characterize the mechanism that solves this optimization problem.
- Call it "Robust Optimal Mechanism (ROM)".
- Structure of ROM:
  - Compute Global Reserve R^*.
  - If the total bids result in realized revenue of less than or equal to R^*, then the auctioneer does not allocate the items.
  - Otherwise compute allocations and payments using a linear optimization problem.

Robust Optimal Mechanism

- *ROM* consists of Algorithms *ROM.a* and *ROM.b*.
- In *ROM.a*, which occurs prior to the realization of a specific bid vector $\mathbf{v}$, we compute the quantity R^*, which stands for the global reserve.
  - This involves a bilinear optimization problem.
- In *ROM.b*, when the bid vector $\mathbf{v}$ is realized, we calculate the allocation vector $\left\{a_{ij}^{\mathbf{v}}\right\}_{i\in\mathcal{N},j\in\mathcal{M}}$ and the payments $\{p_i^{\mathbf{v}}\}_{i\in\mathcal{N}}$ by solving linear optimization problems.

ROM.a

- **Input** : Uncertainty set $\mathcal{U}$, and budgets $B_1, \ldots, B_n$,
- **Output** : Global Reserve R^*.
- By solving the bilinear optimization problem, compute

$$\mathbf{R}^* = \min_{\mathbf{v}\in\mathcal{U}} \left\{ \begin{array}{cc} \max\limits_{\left(\{x_{ij}\}_{i\in\mathcal{N},j\in\mathcal{M}},\{r_i\}_{i\in\mathcal{N}}\right)} & \sum\limits_{i\in\mathcal{N}} r_i \\ \text{s.t.} & \sum\limits_{j\in\mathcal{M}} x_{ij}\cdot v_{ij} \le B_i,\ \forall i\in\mathcal{N}, \\ & r_i \le \sum\limits_{j\in\mathcal{M}} x_{ij}\cdot v_{ij},\ \forall i\in\mathcal{N}, \\ & \sum\limits_{i\in\mathcal{N}} x_{ij} \le 1,\ \forall j\in\mathcal{M}, \\ & \mathbf{x}\ge\mathbf{0}. \end{array} \right\}.$$

- Let $\mathbf{z}$ be the argmin of this bilinear optimization problem. Compute $\left(\left\{\xi_j^*\right\}_{j\in\mathcal{M}}, \left\{\eta_i^*\right\}_{i\in\mathcal{N}}, \left\{\theta_i^*\right\}_{i\in\mathcal{N}}\right)$ given by

$$\arg \left\{ \begin{array}{cc} \min\limits_{\{\xi_j,\eta_i,\theta_i\}} & \sum\limits_{j\in\mathcal{M}} \xi_j + \sum\limits_{i\in\mathcal{N}} \eta_i B_i \\ \text{s.t.} & \xi_j + z_{ij}\cdot\eta_i \ge z_{ij}\cdot\theta_i,\ \forall i\in\mathcal{N}, \forall j\in\mathcal{M}, \\ & \theta_i = 1,\ \forall i\in\mathcal{N}, \\ & \boldsymbol{\xi},\boldsymbol{\eta},\boldsymbol{\theta}\ge\mathbf{0}. \end{array} \right\}. \quad (1)$$

ROM.b

- **Input**: Bid vector $\mathbf{v} = \{v_{ij}\}_{i\in\mathcal{N}, j\in\mathcal{M}}$, reserve R^*.
- **Output**: Allocation vector $\left\{a_{ij}^{\mathbf{v}}\right\}_{i\in\mathcal{N}, j\in\mathcal{M}}$ and the payments $\{p_i^{\mathbf{v}}\}_{i\in\mathcal{N}}$.
- **Algorithm**:
  - If $\mathbf{v} \notin \mathcal{U}$, then do not allocate any item and charge zero, otherwise proceed to next step.
  - Solve the linear optimization problems:

$$\left(\left\{y_{ij}^{\mathbf{v}}\right\}_{i\in\mathcal{N}, j\in\mathcal{M}}, \{r_i^{\mathbf{v}}\}_{i\in\mathcal{N}}\right) = \arg\max_{(\mathbf{y},\mathbf{r})\in\mathcal{P}^{\mathbf{v}}} \sum_{i\in\mathcal{N}} \left(\sum_{j\in\mathcal{M}} y_{ij}\cdot v_{ij} - r_i\right), \quad (2)$$

$$\left(\left\{y_{ij,k}^{\mathbf{v}_{-k}}\right\}_{i\in\mathcal{N}, j\in\mathcal{M}}, \left\{r_{i,k}^{\mathbf{v}_{-k}}\right\}_{i\in\mathcal{N}}\right) = \arg\max_{(\mathbf{y},\mathbf{r})\in\mathcal{P}^{\mathbf{v}}} \sum_{i\in\mathcal{N}\setminus\{k\}} \left(\sum_{j\in\mathcal{M}} y_{ij}\cdot v_{ij} - r_i\right), \quad (3)$$

ROM.b contd...

- where

$$\mathcal{P}^{\mathbf{v}} = \left\{ \left(\{x_{ij}\}_{i\in\mathcal{N}, j\in\mathcal{M}}, \{r_i\}_{i\in\mathcal{N}} \right) \middle| \begin{array}{ll} \sum_{j\in\mathcal{M}} x_{ij} v_{ij} \leq B_i, & \forall i \in \mathcal{N}, \\ r_i \leq \sum_{j\in\mathcal{M}} x_{ij} v_{ij}, & \forall i \in \mathcal{N}, \\ \sum_{i\in\mathcal{N}} x_{ij} \leq 1, & \forall j \in \mathcal{M}, \\ \sum_{i\in\mathcal{N}} r_i \geq R^*, & \\ \mathbf{x} \geq \mathbf{0}. & \end{array} \right\}. \tag{4}$$

- Compute the allocation vector $\{a_k^{\mathbf{v}}\}_{k\in\mathcal{N}}$ and the payments $\{p_k^{\mathbf{v}}\}_{k\in\mathcal{N}}$ as follows

$$a_k^{\mathbf{v}} = y_k^{\mathbf{v}}, \tag{5}$$

$$p_k^{\mathbf{v}} = r_k^{\mathbf{v}} + \sum_{i\in\mathcal{N}\setminus\{k\}} \left(\sum_{j\in\mathcal{M}} y_{ij}^{\mathbf{v}_{-k}} \cdot v_{ij} - r_{i,k}^{\mathbf{v}_{-k}} \right) - \sum_{i\in\mathcal{N}\setminus\{k\}} \left(\sum_{j\in\mathcal{M}} y_{ij}^{\mathbf{v}} \cdot v_{ij} - r_i^{\mathbf{v}} \right). \tag{6}$$

Robust Optimal Mechanism

Theorem

ROM is the worst case optimal auction.

Proof.

Two steps:
(1) Show that the ROM.b leads to allocations and payments that lead to budget feasibility, individual rationality and incentive compatibility. That is, ***show that the allocations and payments are feasible to the primal optimization problem OPT.***
(2) Show that the revenue achieved is optimal by constructing a feasible solution to the dual of OPT that has the same revenue.
(3) By Strong Duality, the result follows. □

Proof (sample)

Budget Feasibility

- Suppose the buyers' response is to bid $\mathbf{v}^{\text{bid}}$.
- The payment charged to each buyer $k \in \mathcal{N}$ is given by

$$
\begin{aligned}
p_k^{\mathbf{v}^{\text{bid}}} &= r_k^{\mathbf{v}^{\text{bid}}} + \sum_{i \in \mathcal{N}\setminus\{k\}} \left(\sum_{j \in \mathcal{M}} y_{ij}^{\mathbf{v}_{-k}^{\text{bid}}} \cdot v_{ij}^{\text{bid}} - r_{i,k}^{\mathbf{v}_{-k}^{\text{bid}}} \right) - \sum_{i \in \mathcal{N}\setminus\{k\}} \left(\sum_{j \in \mathcal{M}} y_{ij}^{\mathbf{v}^{\text{bid}}} \cdot v_{ij}^{\text{bid}} - r_i^{\mathbf{v}^{\text{bid}}} \right) \\
&= r_k^{\mathbf{v}^{\text{bid}}} + \sum_{j \in \mathcal{M}} y_{kj}^{\mathbf{v}^{\text{bid}}} \cdot v_{ij}^{\text{bid}} - \sum_{j \in \mathcal{M}} y_{kj}^{\mathbf{v}^{\text{bid}}} \cdot v_{ij}^{\text{bid}} + \sum_{i \in \mathcal{N}\setminus\{k\}} \left(\sum_{j \in \mathcal{M}} y_{ij}^{\mathbf{v}_{-k}^{\text{bid}}} \cdot v_{ij}^{\text{bid}} - r_{i,k}^{\mathbf{v}_{-k}^{\text{bid}}} \right) \\
&\qquad - \sum_{i \in \mathcal{N}\setminus\{k\}} \left(\sum_{j \in \mathcal{M}} y_{ij}^{\mathbf{v}^{\text{bid}}} \cdot v_{ij}^{\text{bid}} - r_i^{\mathbf{v}^{\text{bid}}} \right) \qquad (7) \\
&= \sum_{j \in \mathcal{M}} y_{kj}^{\mathbf{v}^{\text{bid}}} \cdot v_{kj}^{\text{bid}} + \sum_{i \in \mathcal{N}\setminus\{k\}} \left(\sum_{j \in \mathcal{M}} y_{ij}^{\mathbf{v}_{-k}^{\text{bid}}} \cdot v_{ij}^{\text{bid}} - r_{i,k}^{\mathbf{v}_{-k}^{\text{bid}}} \right) - \sum_{i \in \mathcal{N}} \left(\sum_{j \in \mathcal{M}} y_{ij}^{\mathbf{v}^{\text{bid}}} \cdot v_{ij}^{\text{bid}} - r_i^{\mathbf{v}^{\text{bid}}} \right) \qquad (8) \\
&\le \sum_{j \in \mathcal{M}} y_{kj}^{\mathbf{v}^{\text{bid}}} \cdot v_{kj}^{\text{bid}} \qquad (9) \\
&\le B_i, \qquad (10)
\end{aligned}
$$

where (10) follows from (9) because $\left(\left\{ y_{ij}^{\mathbf{v}^{\text{bid}}} \right\}_{i \in \mathcal{N}, j \in \mathcal{M}} \right) \in \mathcal{P}^{\mathbf{v}^{\text{bid}}}$.

Proof (sample)

Worst Case Revenue

- If buyers bid their true valuation, then their utility is non-negative. If, however, buyers bid $\mathbf{v}^{\text{bid}} \notin \mathcal{U}$, then by Step 1 of *ROM.b*, their utility is zero. Therefore, if the buyers bid $\mathbf{v}^{\text{bid}}$, then $\mathbf{v}^{\text{bid}} \in \mathcal{U}$.
- From Step 2 of *ROM.b*, the payments $\left\{ r_i^{\mathbf{v}^{\text{bid}}} \right\}_{i \in \mathcal{N}}$ are feasible to (2) with $\mathbf{v} = \mathbf{v}^{\text{bid}}$. Since $\mathbf{v}^{\text{bid}} \in \mathcal{U}$, then from Step 1 of *ROM.a*,

$$\sum_{i=1}^{n} r_i^{\mathbf{v}^{\text{bid}}} \geq R^*. \tag{11}$$

Furthermore, we have

$$\begin{aligned} p_k^{\mathbf{v}^{\text{bid}}} &= r_k^{\mathbf{v}^{\text{bid}}} + \sum_{i \in \mathcal{N} \setminus \{k\}} \left(\sum_{j \in \mathcal{M}} y_{ij}^{\mathbf{v}_{-k}^{\text{bid}}} \cdot v_{ij}^{\text{bid}} - r_{i,k}^{\mathbf{v}_{-k}^{\text{bid}}} \right) - \sum_{i \in \mathcal{N} \setminus \{k\}} \left(\sum_{j \in \mathcal{M}} y_{ij}^{\mathbf{v}^{\text{bid}}} \cdot v_{ij}^{\text{bid}} - r_i^{\mathbf{v}^{\text{bid}}} \right) \\ &\geq r_k^{\mathbf{v}^{\text{bid}}}, \end{aligned}$$

- This implies that

$$\sum_{i=1}^{n} p_i^{\mathbf{v}^{\text{bid}}} \geq \sum_{i=1}^{n} r_i^{\mathbf{v}^{\text{bid}}} \geq R^*,$$

implying that the worst case revenue is at least R^*.

Proof (sample)

ROM achieves at least Z^*

- Consider the following relaxation of OPT, in which we eliminate the IC constraints:

$$
\begin{aligned}
Z_{\mathbf{1}}^* = \max \quad & W && (12)\\
\text{s.t.} \quad & W - \sum_{i \in \mathcal{N}} p_i^{\mathbf{v}} \leq 0, \quad \forall \mathbf{v} \in \mathcal{U}, \\
& \sum_{i \in \mathcal{N}} x_{ij}^{\mathbf{v}} \leq 1, \quad \forall j \in \mathcal{M}, \forall \mathbf{v} \in \mathcal{U}, \\
& p_i^{\mathbf{v}} \leq B_i, \quad \forall i \in \mathcal{N}, \forall \mathbf{v} \in \mathcal{U}, && (13)\\
& p_i^{\mathbf{v}} \leq \sum_{j \in \mathcal{M}} v_{ij} \cdot x_{ij}^{\mathbf{v}}, \quad \forall i \in \mathcal{N}, \forall \mathbf{v} \in \mathcal{U}, && (14)\\
& \mathbf{x}^{\mathbf{v}} \geq \mathbf{0}.
\end{aligned}
$$

The dual of (12) is as follows:

$$
\begin{aligned}
\min \quad & \sum_{\mathbf{v} \in \mathcal{U}} \left(\sum_{j=\mathbf{1}}^{m} \xi_{j,\mathbf{v}} + \sum_{i=\mathbf{1}}^{n} \eta_{i,\mathbf{v}} B_i \right) && (15)\\
\text{s.t.} \quad & \xi_{j,(\mathbf{v}_i,\mathbf{v}_{-i})} - v_{ij} \cdot \theta_{i,(\mathbf{v}_i,\mathbf{v}_{-i})} \geq 0, \quad \forall (\mathbf{v}_i, \mathbf{v}_{-i}) \in \mathcal{U}, \\
& \eta_{i,(\mathbf{v}_i,\mathbf{v}_{-i})} + \theta_{i,(\mathbf{v}_i,\mathbf{v}_{-i})} - \omega_{(\mathbf{v}_i,\mathbf{v}_{-i})} = 0, \quad \forall (\mathbf{v}_i, \mathbf{v}_{-i}) \in \mathcal{U}, \\
& \sum_{\mathbf{v} \in \mathcal{U}} \omega_{\mathbf{v}} = 1, \\
& \omega_{\mathbf{v}} \geq 0, \boldsymbol{\xi}_{\mathbf{v}} \geq \mathbf{0}, \boldsymbol{\eta}_{\mathbf{v}} \geq \mathbf{0}, \boldsymbol{\theta}_{\mathbf{v}} \geq \mathbf{0}.
\end{aligned}
$$

Proof (sample)

ROM achieves at least Z^*

- Since Problem (12) is obtained from OPT by eliminating the IC constraints, we have

$$Z_1^* \geq Z^*. \tag{16}$$

- Let $\mathbf{z}$ be an optimal solution in Step 1 of *ROM.a*. We next construct a feasible solution to the dual problem (15) with objective function equal to R^*. Let

$$\omega_{\mathbf{v}} = \begin{cases} 1, & \text{if } \mathbf{v} = \mathbf{z}, \\ 0, & \text{otherwise}, \end{cases}$$

$$\eta_{i,\mathbf{v}} = \begin{cases} \eta_i^*, & \text{if } \mathbf{v} = \mathbf{z}, \\ 0, & \text{otherwise}, \end{cases} \qquad \forall i \in \mathcal{N},$$

$$\xi_{j,\mathbf{v}} = \begin{cases} \xi_j^*, & \text{if } \mathbf{v} = \mathbf{z}, \\ 0, & \text{otherwise}, \end{cases} \qquad \forall j \in \mathcal{M},$$

$$\theta_{i,\mathbf{v}} = \begin{cases} 1 - \eta_i^*, & \text{if } \mathbf{v} = \mathbf{z}, \\ 0, & \text{otherwise}, \end{cases} \qquad \forall i \in \mathcal{N},$$

where $\left(\left\{\xi_j^*\right\}_{j\in\mathcal{M}}, \left\{\eta_i^*\right\}_{i\in\mathcal{N}}, \left\{\theta_i^*\right\}_{i\in\mathcal{N}}\right)$ were computed in **ROM.a**.

Proof (sample)

ROM achieves at least Z^*

- It is easy to verify that this is a dual feasible solution to Problem (15), with objective value given by

$$\sum_{j \in \mathcal{M}} \xi_j^* + \sum_{i \in \mathcal{N}} \eta_i^* B_i,$$

  which is equal to R^*.
- This leads to

$$R^* \geq Z_1^* \geq Z^*.$$

- This concludes the proof.

Summary

- Characterized the worst case optimal auction with budgets, for **any uncertainty set $\mathcal{U}$.**
- Can choose the uncertainty sets carefully, to include distributional information.
    - Capture arbitrary "risk measures" of the auctioneer.
- "Global reserve" structure allows auctioneer to ensure selling even the "bad" items.
- For the case of no budgets, the structure is the same as Myerson auction.
- *ROM* extends to uncertain budgets and indivisible items.

Single Item Auction without Budgets

- *ROM* is a second price auction with reserve price R^*.
- R^* is calculated by a linear optimization problem:

$$\begin{aligned} \min_{r,\mathbf{v}} \quad & r \\ \text{s.t.} \quad & r \geq v_i, \ \forall i \in \mathcal{N}, \\ & (v_1, v_2, \ldots, v_n) \in \mathcal{U}. \end{aligned}$$

Comparision with Myerson Auction

Computational Complexity

- *ROM* and the Myerson auction have the same structure, that of a second price auction with a reservation price.
- In Myerson auction, the reservation price is calculated by solving a non-linear equation
$$\frac{1 - F(r)}{f(r)} = r,$$
where $F(\cdot)$ is the cdf and $f(\cdot)$ is the pdf of the probability distribution.
- In *ROM*, the reservation price is calculated using a linear optimization problem.

Comparision with Myerson Auction

Robustness to Mis-specification

$$\text{Relative Revenue} = \frac{\text{ROM Revenue} - \text{Myerson Revenue}}{\text{Myerson Revenue}}$$

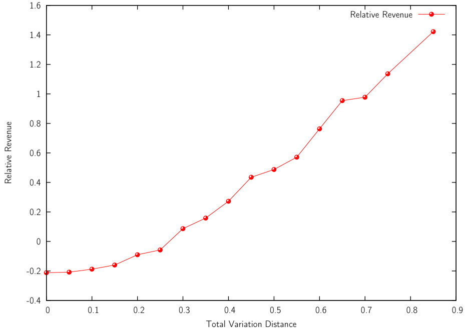


Figure : Robustness of *ROM-Si*.

Comparision with Myerson Auction

Capturing Correlation Information

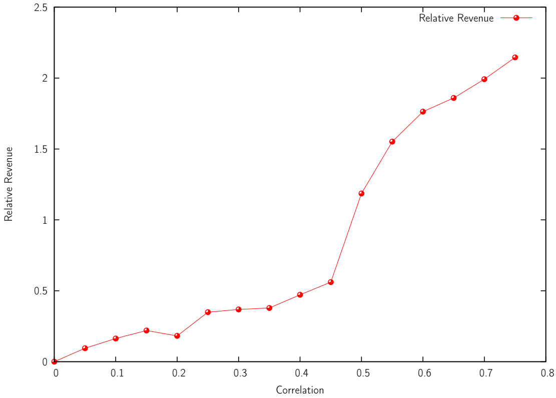


Figure : Effect of Correlations on the Revenue.

Conclusion

- Characterized the "worst case" optimal auction
- with budgeted bidders
    - for any uncertainty set $\mathcal{U}$
- Auction can be interpreted as a VCG auction with a "global reserve".
- Recovered the structure of Myerson Auction for bidders without budgets.
- Has benefits of robustness to mis-specification.
- Extensions: Can naturally model the risk attitudes of the auctioneer by modifying the uncertainty set.

15.094J: Robust Modeling, Optimization, Computation

Lecture 22: Robust Statistics

Outline

1. Robust Regression
2. Support Vector Machines
3. The impact of Robustness
4. Conclusions

Robust Regression

- Given data (y_i, x_i), $y_i \in \mathbb{R}$, $x_i \in \mathbb{R}^m$, $i = 1, 2, \ldots, n$.
- Robust L_p Regression optimization problem:

$$\min_{\boldsymbol{\beta},\beta_0} \max_{\boldsymbol{\Delta} X \in \mathcal{N}} \|y - (X + \boldsymbol{\Delta} X)\boldsymbol{\beta} - \beta_0 \mathbf{1}\|_p. \tag{1}$$

- $y = (y_1, \ldots, y_n)'$, $X = (x_1, \ldots, x_n)'$, and $\mathbf{1} = (1, \ldots, 1)'$
- $\mathcal{N}$ is the uncertainty set for $\boldsymbol{\Delta} X$.

Matrix Norms and Uncertainty Sets

- Norm $\| \bullet \|_{q,p}$ for an $n \times m$ matrix A:

$$\|A\|_{q,p} \equiv \sup_{x \in \mathbb{R}^m, x \neq \mathbf{0}} \frac{\|Ax\|_p}{\|x\|_q}, \;\; q, p \geq 1.$$

- Note that for any $x \in \mathbb{R}^m$ with $\|x\|_q = 1$, we have that $\|A\|_{q,p} = \|Ax\|_p$.
- $\mathcal{N}_1 = \{\mathbf{\Delta} X \in \mathbb{R}^{n \times m} \mid \|\mathbf{\Delta} X\|_{q,p} \leq \rho\}$.

Matrix Norms and Uncertainty Sets

- The p-Frobenius norm $\| \bullet \|_{p-F}$ of an $n \times m$ matrix A :

$$\|A\|_{p-F} \equiv \left(\sum_{i=1}^{n} \sum_{j=1}^{m} |A_{i,j}|^p \right)^{1/p} .$$

- For $p = 2$, we obtain the usual Frobenius norm.
- The dual norm of $\| \bullet \|_p$ is $\| \bullet \|_q$: $\frac{1}{p} + \frac{1}{q} = 1$.
- Thus, dual norm of $\| \bullet \|_p$ is $\| \bullet \|_{d(p)}$ with

$$d(p) = \frac{p}{p-1}, p \geq 1.$$

Note $d(1) = \infty$ and $d(\infty) = 1$.

- $\mathcal{N}_2 = \{\mathbf{\Delta} X \in \mathbb{R}^{n \times m} \mid \|\mathbf{\Delta} X\|_{p-F} \leq \rho\}$.

Equivalence of Robustification and Reguralization

- Under uncertainty set $\mathcal{N}_1$, Problem (1) is equivalent to problem

$$\min_{\boldsymbol{\beta},\beta_0} \| y - X\boldsymbol{\beta} - \beta_0\mathbf{1} \|_p + \rho \| \boldsymbol{\beta} \|_q .$$

- Under uncertainty set $\mathcal{N}_2$, Problem (1) is equivalent to problem

$$\min_{\boldsymbol{\beta},\beta_0} \| y - X\boldsymbol{\beta} - \beta_0\mathbf{1} \|_p + \rho \| \boldsymbol{\beta} \|_{d(p)},$$

Properties on matrix norms

- Definition:

$$[f(x,p)]_j = \begin{cases} sign(x_j)\left(\dfrac{|x_j|}{\| x \|_p}\right)^{p-1}, & x \neq \mathbf{0}, \\ 0, & x = \mathbf{0}, \end{cases} \quad j = 1, 2, \ldots, m,$$

where $sign(x) = \begin{cases} 1, & x \geq 0, \\ -1, & x < 0. \end{cases}$

- Proposition 1: (a) $[f(x,p)]'x = \|x\|_p$, (b) $\|f(x,p)\|_{d(p)} = 1$.
- Proposition 2: $\|A\|_{d(p),p} \leq \|A\|_{p-F}$.
- Proposition 3: For $u_1 \in \mathbb{R}^n$, $u_2 \in \mathbb{R}^m$, $p, q \geq 1$, $\|u_1 u_2'\|_{q,p} = \|u_1\|_p \|u_2\|_{d(q)}$.
- Proposition 4: For $u_1 \in \mathbb{R}^n$, $u_2 \in \mathbb{R}^m$, $p \geq 1$, $\|u_1 u_2'\|_{p-F} = \|u_1\|_p \|u_2\|_p$.

Proof of Equivalence of Reguralization and Robustification

$$\|y - (X + \boldsymbol{\Delta}X)\boldsymbol{\beta} - \beta_0\mathbf{1}\|_p \;=\; \|y - X\boldsymbol{\beta} - \beta_0\mathbf{1} - \boldsymbol{\Delta}X\boldsymbol{\beta}\|_p$$

$$\leq\; \|y - X\boldsymbol{\beta} - \beta_0\mathbf{1}\|_p \;+\; \|\boldsymbol{\Delta}X\boldsymbol{\beta}\|_p.$$

We obtain the bound

$$\|\boldsymbol{\Delta}X\boldsymbol{\beta}\|_p \;\leq\; \|\boldsymbol{\Delta}X\|_{q,p}\;\|\boldsymbol{\beta}\|_q,$$

Thus, for $\|\boldsymbol{\Delta}X\|_{q,p} \leq \rho$, $\|\boldsymbol{\Delta}X\boldsymbol{\beta}\|_p \;\leq\; \rho\|\boldsymbol{\beta}\|_q$, and for any $\boldsymbol{\Delta}X \in \mathcal{N}_1$,

$$\|y - (X + \boldsymbol{\Delta}X)\boldsymbol{\beta} - \beta_0\mathbf{1}\|_p \;\leq\; \|y - X\boldsymbol{\beta} - \beta_0\mathbf{1}\|_p \;+\; \rho\|\boldsymbol{\beta}\|_q.$$

Proof Continued

- Define

$$\boldsymbol{\Delta} X^o = \begin{cases} -\rho \dfrac{y - X\boldsymbol{\beta} - \beta_0 \mathbf{1}}{\|y - X\boldsymbol{\beta} - \beta_0 \mathbf{1}\|_p} [f(\boldsymbol{\beta}, q)]', & y - X\boldsymbol{\beta} - \beta_0 \mathbf{1} \neq \mathbf{0}, \\ -\rho u [f(\boldsymbol{\beta}, q)]', & y - X\boldsymbol{\beta} - \beta_0 \mathbf{1} = \mathbf{0}, \end{cases}$$

- where $f(x, p) \in \mathbb{R}^m$, $x \in \mathbb{R}^m$, $p \geq 1$, $u \in \mathbb{R}^n$, with $\|u\|_p = 1$.
- For $y - X\boldsymbol{\beta} - \beta_0 \mathbf{1} \neq \mathbf{0}$:

$$\|y - (X + \boldsymbol{\Delta} X^o)\boldsymbol{\beta} - \beta_0 \mathbf{1}\|_p = \|y - X\boldsymbol{\beta} - \beta_0 \mathbf{1} - \boldsymbol{\Delta} X^o \boldsymbol{\beta}\|_p$$

$$= \left\| y - X\boldsymbol{\beta} - \beta_0 \mathbf{1} + \rho \frac{y - X\boldsymbol{\beta} - \beta_0 \mathbf{1}}{\|y - X\boldsymbol{\beta} - \beta_0 \mathbf{1}\|_p} [f(\boldsymbol{\beta}, q)]' \boldsymbol{\beta} \right\|_p$$

Proof Continued

$$= \left\| (y - X\boldsymbol{\beta} - \beta_0\mathbf{1}) \left(1 + \frac{\rho\|\boldsymbol{\beta}\|_q}{\|y - X\boldsymbol{\beta} - \beta_0\mathbf{1}\|_p} \right) \right\|_p \quad ([f(\boldsymbol{\beta}, q)]'\boldsymbol{\beta} = \|\boldsymbol{\beta}\|_q)$$

$$= \|y - X\boldsymbol{\beta} - \beta_0\mathbf{1}\|_p + \rho\|\boldsymbol{\beta}\|_q.$$

Note that when $y - X\boldsymbol{\beta} - \beta_0\mathbf{1} = \mathbf{0}$, $\|y - (X + \boldsymbol{\Delta}X^o)\boldsymbol{\beta} - \beta_0\mathbf{1}\|_p =$ $\|y - X\boldsymbol{\beta} - \beta_0\mathbf{1}\|_p + \rho\|\boldsymbol{\beta}\|_q$ as well.

Proof Continued

- From Propositions 1, 3, we have that if $y - X\boldsymbol{\beta} - \beta_0\mathbf{1} \neq \mathbf{0}$,

$$\|\boldsymbol{\Delta}X^o\|_{q,p} = \rho \left\| \frac{y - X\boldsymbol{\beta} - \beta_0\mathbf{1}}{\|y - X\boldsymbol{\beta} - \beta_0\mathbf{1}\|_p} \right\|_p \|f(\boldsymbol{\beta}, q)\|_{d(q)} = \rho,$$

- and if $y - X\boldsymbol{\beta} - \beta_0\mathbf{1} = \mathbf{0}$,

$$\|\boldsymbol{\Delta}X^o\|_{q,p} = \rho\|u\|_p\|f(\boldsymbol{\beta}, q)\|_{d(q)} = \rho,$$

  and thus, $\boldsymbol{\Delta}X^o \in \mathcal{N}_1$.

- Hence

$$\max_{\boldsymbol{\Delta}X \in \mathcal{N}_1} \|y - (X + \boldsymbol{\Delta}X)\boldsymbol{\beta} - \beta_0\mathbf{1}\|_p \;=\; \|y - X + \boldsymbol{\beta} - \beta_0\mathbf{1}\|_p + \rho\|\boldsymbol{\beta}\|_q$$

Support Vector Machines

- Given categorical data (y_i, x_i), $y_i \in \{1, -1\}$, $x_i \in \mathbb{R}^m$, $i \in \{1, 2, \ldots, n\}$, we define the separation error $S(\boldsymbol{\beta}, \beta_0, y, X)$ of the hyperplane classifier $\boldsymbol{\beta}' x + \beta_0 = 0$, $x \in \mathbb{R}^m$, in space $\mathbb{R}^m$ by

$$S(\boldsymbol{\beta}, \beta_0, y, X) = \sum_{i=1}^{n} \max(0, 1 - y_i(\boldsymbol{\beta}' x_i + \beta_0)), \tag{2}$$

- The hyperplane which minimizes the separation error is the solution of the optimization problem

$$\min_{\boldsymbol{\beta}, \beta_0} S(\boldsymbol{\beta}, \beta_0, y, X), \tag{3}$$

which can be expressed as the linear optimization problem

$$\begin{array}{ll} \min\limits_{\boldsymbol{\beta}, \beta_0, \boldsymbol{\xi}} & \sum\limits_{i=1}^{n} \xi_i \\ \text{s.t.} & y_i(\boldsymbol{\beta}' x_i + \beta_0) \geq 1 - \xi_i, i \in \{1, 2, \ldots, n\} \\ & \xi_i \geq 0, i \in \{1, 2, \ldots, n\}. \end{array}$$

Robustification

- Consider uncertainty set

$$\mathcal{N}_3 = \left\{ \mathbf{\Delta} X \in \mathbb{R}^{n \times m} \mid \sum_{i=1}^{n} \|\mathbf{\Delta} x_i\|_p \leq \rho \right\}. \tag{4}$$

- The robust version of Problem (3):

$$\min_{\boldsymbol{\beta}, \beta_0} \max_{\mathbf{\Delta} X \in \mathcal{N}_3} S(\boldsymbol{\beta}, \beta_0, y, X + \mathbf{\Delta} X). \tag{5}$$

Robustification leads to Support Vector Machines

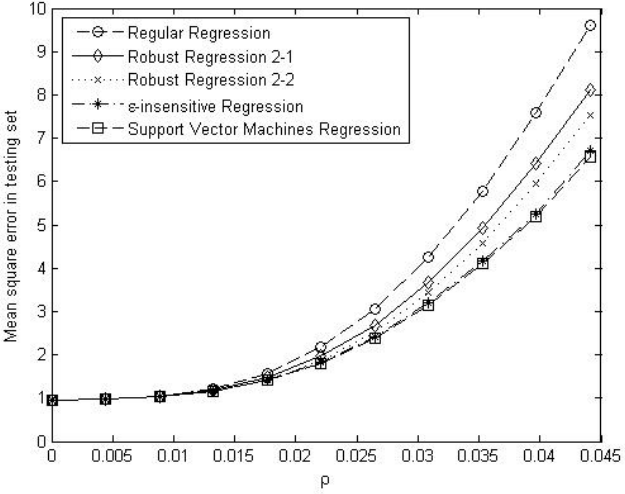
- Definition: The set of data (y_i, x_i), $i \in \{1, 2, \ldots, n\}$, is separable if there exists a hyperplane $\boldsymbol{\beta}' x + \beta_0 = 0$ in $\mathbb{R}^m$, such that for any $i \in \{1, 2, \ldots, n\}$, $y_i(\boldsymbol{\beta}' x_i + \beta_0) \geq 0$.
- Theorem: If the set of data (y_i, x_i), $i \in \{1, 2, \ldots, n\}$ is not separable, Problem (5) is equivalent to problem

$$
\begin{array}{ll}
\min\limits_{\boldsymbol{\beta}, \beta_0, \boldsymbol{\xi}} & \sum\limits_{i=1}^{n} \xi_i + \rho \parallel \boldsymbol{\beta} \parallel_{d(p)} \\
\text{s.t.} & \xi_i \geq 1 - y_i(\boldsymbol{\beta}' x_i + \beta_0), i \in \{1, 2, \ldots, n\} \\
& \xi_i \geq 0, i \in \{1, 2, \ldots, n\},
\end{array}
$$

The impact of Robustness

- $x_i \sim N(\mathbf{1}, 5I_3),\ i = 1, \ldots, 200.$
- $y_i = \boldsymbol{\beta}' x_i + \beta_0 + r$, $\beta_0 = 1$, $\boldsymbol{\beta} = (1, -3, 1)'$, $r \sim N(0, 1)$.
- Training set (50%), testing set (50%).
- $\Delta x_i \sim \rho N(0, \mathbf{1})$.
- The procedure was repeated 30 times and the average performance of each estimate was recorded.

The impact of Robustness



The impact of Robustness

- As ρ increases, the difference in the out-of-sample performance between the robust and the respective classical estimates increases, with the robust estimates always yielding better results.
- The robust regression estimate with $p = 2$ and $q = 2$ shows better performance than the robust regression estimate with $p = 2$ and $q = 1$.

Performance on Real Data

Data set	n	m
Abalone	4177	9
Auto MPG	392	8
Comp Hard	209	7
Concrete	1030	8
Housing	506	13
Space shuttle	23	4
WPBC	46	32

- Training (50%), validation (25%), and testing (25%).
- For each ρ, prediction error on validation set was measured, and ρ with highest performance on validation was used in testing.
- Prediction errors were averaged over the 30 partitions.

Mean square error

Data set	Regular	Rob 2-1	Rob 2-2	Supp vector
Abalone	5.7430	5.6345	5.5369	5.0483*
Auto MPG	18.7928	18.6981	18.5829	12.5251*
Comp Hard	2026.0024	1965.7531	1925.1250*	2348.2914
Concrete	132.4700	131.0820	129.3135	127.0923*
Forest Fires	5525.9994	4994.8077*	5266.3994	5229.5167
Housing	39.8084	39.4257	39.0716	24.6867*
Space shuttle	0.5323	0.5225*	0.5265	0.5501
WPBC	4723.0723	4630.1946	4489.2032	4410.4599*

Conclusions

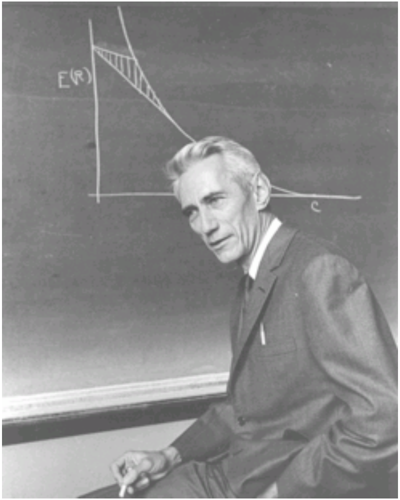
- Reguralization and Robustfication are Equivalent!
- Insights on the norms to use in reguirilization.
- Support vector, method developed 3 decades ago, is one of the strongest performing classification methods.
- Robustness provides insights on the reasons.

15.094J: Robust Modeling, Optimization, Computation

Lecture 23-24: Network Information Theory via Robust Optimization

Outline

Claude Shannon (1916-2001)



- BS, University of Michigan, 1936.
- SM, MIT, 1937.
- PhD, MIT, 1940.
- Institute of Advanced Study, Princeton, 1941.
- Bell Labs, 1942-1956.
- MIT, professor, 1956-1983.
- MIT, emeritus, 1983-2001.

Perspectives

- Shannon-1948 "A Mathematical Theory of Communication", started the field of Information theory.
- The paper is described as the Magna Carta of the modern information age.
- Like Einstein in 1906, 1915, Shannon was asking questions nobody else was asking at the time.

Citations in the Mathematical/Economics Sciences in the 20th century

- **Shannon-1948: 75,000 citations**.
- Keynes -1937: 25,000 citations.
- Kahneman-Tversky -1979: 35,000 citations.
- Metropolis -1953: 30,000 citations.
- von Neumann-Morgensten -1944: 25,000 citations.
- Kalman -1953: 20,000 citations.
- RSA -1978: 15,000 citations
- Karp -1972 : 9,000 citations
- Dantzig -1947: 8,000 citations
- Nash -1950: 5,000 citations

Information theory

Successes

- Characterization of the capacity of a single user channel, Shannon-1948.
- Closed form formula for the capacity of a variety of single user channels (example: additive Gaussian channel).
- Random coding with minimum-distance decoding achieves this capacity.

Challenges

- Network Information theory.
- Multi sender, multi receiver channels with interference.
- Understanding the limitations of wireless networks.

Research Objectives

- Use RO to attack stochastic systems in high dimensions.
- Connect RO and Information Theory to achieve tractability and address network information theory.
- Understand the tractability of the approach.
- Understand the limitations of the approach.

Robustness and Information Theory

- Typical sets introduced by Shannon can be viewed as uncertainty sets in RO.
- Decoding is a robustness property.

Typical Sets

Incorporating Distributional Information

- Shannon (1948) introduced the idea of Typical Sets:
- Property (a): A typical set has probability nearly 1.
- Property (b): All elements of typical set are nearly equiprobable.
- Given pdf $f(\cdot)$,

$$\mathcal{U}^{f-\text{Typical}} = \left\{ (z_1, \dots, z_n) \,\middle|\, -\Gamma \leq \frac{\sum_{i=1}^{n} \log f(z_i) - n \cdot \mu_{\log f}}{\sigma_{\log f} \cdot \sqrt{n}} \leq \Gamma. \right\},$$

$$\mu_{\log f} = \int_{-\infty}^{\infty} f(x) \log f(x) \, \mathrm{d}x,$$

$$\sigma_{\log f_j} = \int_{-\infty}^{\infty} f(x) \left(\log f(x) - \mu_{\log f} \right)^2 \mathrm{d}x.$$

Examples of Typical Sets

- $\tilde{z}_i \sim N(0, \sigma)$

$$\mathcal{U}_\epsilon^{\mathrm{G}} = \left\{ \mathbf{z} \,\middle|\, -\Gamma_\epsilon^{\mathrm{G}} \leq \|\mathbf{z}\|^2 - n\sigma^2 \leq \Gamma_\epsilon^{\mathrm{G}} \right\}.$$

- $\tilde{z}_i \sim Exp(\lambda)$

$$\mathcal{U}_\epsilon^{\mathrm{E}} = \left\{ \mathbf{z} \,\middle|\, \frac{n}{\lambda} - \frac{\sqrt{n}}{\lambda} \cdot \Gamma_\epsilon^{\mathrm{E}} \leq \sum_{j=1}^{n} z_j \leq \frac{n}{\lambda} + \frac{\sqrt{n}}{\lambda} \cdot \Gamma_\epsilon^{\mathrm{E}}, \; \mathbf{z} \geq \mathbf{0} \right\}.$$

- $\tilde{z}_i \sim U[a, b]$

$$\mathcal{U}_\epsilon^{U} = \left\{ \mathbf{z} \,\middle|\, \begin{array}{c} n\dfrac{a+b}{2} - \Gamma_\epsilon^{U}\sqrt{n} \leq \displaystyle\sum_{j=1}^{n} z_j \leq n\dfrac{a+b}{2} + \Gamma_\epsilon^{U}\sqrt{n}, \\ a \leq z_j \leq b, \; j = 1, \ldots, n, \end{array} \right\}..$$

- $\tilde{z}_i \sim \mathrm{Bin}(p)$

$$\mathcal{U}_\epsilon^{B} = \left\{ \mathbf{z} \,\middle|\, \begin{array}{c} np - \Gamma_\epsilon^{B}\sqrt{n} \leq \displaystyle\sum_{j=1}^{n} z_j \leq np + \Gamma_\epsilon^{B}\sqrt{n}, \\ z_j \in \{0, 1\}, \; j = 1, \ldots, n, \end{array} \right\}.$$

Single Use Gaussian Channel

- A transmitter wants to send M messages index by $i \in \mathcal{M}$.
- He codes the i^{th} message as a vector $\mathbf{x}_i \in \mathbb{R}^n$, and transmits it.
- The receiver, receives a vector $\mathbf{y}_i \in \mathbb{R}^n$ given by

$$\mathbf{y}_i = \mathbf{x}_i + \mathbf{z}_i,$$

  $\mathbf{z}_i$ is noise.
- Noise could be different than additive.

The Key Problem

- *Coding Problem:* Given power P, select $\mathbf{x}_i$ to represent the i^{th} message such that $||\mathbf{x}_i||^2 \leq nP$.
- *Decoding Problem:* Find a decoding function $g(\mathbf{y}_i)$ that maps $\mathbf{y}_i$ to one of the code words:

$$\frac{1}{M}\sum_{i=1}^{M}\mathbb{P}\left[g(\mathbf{y}_i) \neq i\right] \leq \epsilon_n,$$

with $\epsilon_n \to 0$, as $n \to \infty$.
- *Can we send M messages of length n with arbitrary small error ϵ_n?*

Key Insights

- Decoder

$$g(\mathbf{y}) = \arg\min_{i \in \mathcal{M}} \|\mathbf{y} - \mathbf{x}_i\|.$$

- Minimum distance decoding is a *robustness property*

$$\|\mathbf{x}_i + \mathbf{z} - \mathbf{x}_{i'}\| \geq \|\mathbf{z}\| \quad \forall \mathbf{z} \in \mathcal{U}_\epsilon^{\mathrm{G}}, \forall i, i' \neq i$$

$$\mathcal{U}_\epsilon^{\mathrm{G}} = \left\{ \mathbf{z} \,\middle|\, -\Gamma_\epsilon^{\mathrm{G}} \leq \|\mathbf{z}\|^2 - n\sigma^2 \leq \Gamma_\epsilon^{\mathrm{G}} \right\}.$$

- Capacity characterization

$$\begin{aligned} M_n^*(\epsilon) = \max \quad & M \\ \text{s.t.} \quad & \|\mathbf{x}_i + \mathbf{z} - \mathbf{x}_{i'}\| \geq \|\mathbf{z}\|, \quad \forall \mathbf{z} \in \mathcal{U}_\epsilon^{\mathrm{G}}, \forall i, i' \in \mathcal{M}, i' \neq i, \\ & \|\mathbf{x}_i\|^2 \leq nP, \quad \forall i \in \mathcal{M}. \end{aligned}$$

- Theorem (Shannon (1948))

$$\lim_{n\to\infty,\ \epsilon\to 0} \frac{\log_2 M_n^*(\epsilon)}{n} = \frac{1}{2} \cdot \log\left(1 + \frac{P}{\sigma^2}\right).$$

Optimal Coding

- *Inputs: R, P, n, σ, ϵ, ν*.
- Select γ_ϵ, $\mathbb{P}\left[\|\tilde{\mathbf{z}}_G\| \leq \gamma_\epsilon\right] \geq 1-\epsilon$, $\tilde{\mathbf{z}}_G \sim N\left(\mathbf{0}, \sigma \cdot I\right)$;
- $T = \left(\frac{1+\nu}{\zeta\nu} \cdot \frac{\gamma_\epsilon}{\sqrt{n}}\right)^n$, with $\zeta = \frac{\sigma}{\sqrt{n}} \cdot \Phi^{-1}\left(1 - O(\epsilon)\right)$,
- $M_0 = (1+\nu) \cdot \gamma_\epsilon$.;
- Let $\mathcal{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_T\}$ uniformly distributed on $\mathcal{S}_n(M_0) = \left\{\mathbf{z} \in \mathbb{R}^n \middle| \|\mathbf{z}\| = M_0\right\}$.
- Wyner (1967) developed methods to construct deterministic sequences to model uniformly distributed points.

Optimal Coding

- To count fraction of correct decoding:

$$v_{it} = \begin{cases} 1, & \text{if } \|\mathbf{x}_i + \mathbf{z}_t - \mathbf{x}_{i'}\| \geq \|\mathbf{z}_t\|, \forall i' \in \mathcal{M}, \\ 0, & \text{otherwise.} \end{cases}$$

- Encoding Algorithm

$$\begin{aligned} \|\mathbf{x}_i\|^2 &\leq nP, && \forall i \in \mathcal{M}, \\ \|\mathbf{x}_i - \mathbf{x}_k + \mathbf{z}_t\| + (1 - v_{it}) M_0 &\geq \|\mathbf{z}_t\|, && \forall t \in \mathcal{T}, \forall i, k \in \mathcal{M},\ k \neq i, \\ \sum_{t=1}^{T} v_{it} &\geq (1 - \epsilon) T, && \forall i \in \mathcal{M}, \\ v_{it} &\in \{0, 1\}, && \forall i \in \mathcal{M}, t \in \mathcal{T}, \end{aligned}$$

Reformulation

The set of quadratic, possibly non-convex, constraints

$$f_k(\mathbf{y}) = \mathbf{y}'\mathbf{A}_k\mathbf{y} + \mathbf{b}_k'\mathbf{y} + c_k \leq 0, \ \ \forall k \in \mathcal{K}.$$

is equivalent to the semidefinite optimization problem

$$\begin{gathered} \tilde{\mathbf{A}}_k \bullet \mathbf{Y} \leq 0, \ \ \forall k \in \mathcal{K}, \\ Y_{11} = 1, \ \mathbf{Y} \succeq \mathbf{0}, \ \operatorname{rank}(\mathbf{Y}) = 1, \end{gathered}$$

where

$$\mathbf{Y} = \begin{pmatrix} 1 \\ \mathbf{y} \end{pmatrix} (1, \mathbf{y}'), \ \ \tilde{\mathbf{A}}_k = \begin{pmatrix} c_k & \mathbf{b}_k \\ \mathbf{b}_k & \mathbf{A}_k \end{pmatrix}.$$

The overall optimization problem

$$
\begin{array}{rl}
\min & \operatorname{rank}(\mathbf{Y}) \\
\text{s.t.} & \mathbf{A}_i \bullet \mathbf{Y} \leq 0, \quad \forall i \in \mathcal{M}, \\
& \mathbf{B}_{ikt} \bullet \mathbf{Y} \leq 0, \quad \forall t \in \mathcal{T}, \forall i, k \in \mathcal{M},\ k \neq i, \\
& \mathbf{C}_i \bullet \mathbf{Y} \leq 0, \quad \forall i \in \mathcal{M}, \\
& \mathbf{D}_{it} \bullet \mathbf{Y} = 0, \quad \forall i \in \mathcal{M}, t \in \mathcal{T}, \\
& \mathbf{Y} \succeq \mathbf{0},
\end{array}
$$

Algorithm

- **Input :** R, P, σ, n, ν, ϵ.
- Solve the rank minimization SOP to compute r^*, codewords$\{\mathbf{x}_i\}_{i\in\mathcal{M}}$.
- If $r^* = 1$, then $R \in \mathcal{R}_n\left[P, \sigma, 2\epsilon\right]$.
- If $r^* \geq 2$, then $R \notin \mathcal{R}_n\left[P, \left(1+3\nu\right)\sigma, O(\epsilon)\right]$.
- As $n \to \infty$, $\epsilon \to 0$, $\nu \to 0$, the characterization of *the asymptotic capacity of the channel is tight.*

Remarks

- By solving a rank minimization SOP, we find the asymptotic capacity and the matching optimal code.
- *Similar RO problems for a variety of other type of channels:*
    - Binary symmetric channel,
    - Binary erasure channel,
    - Additive uniform noise channel.
    - Additive exponential channel.

Algorithm from Fazell, Hindi, Boyd –2003

- Solve the convex optimization problem

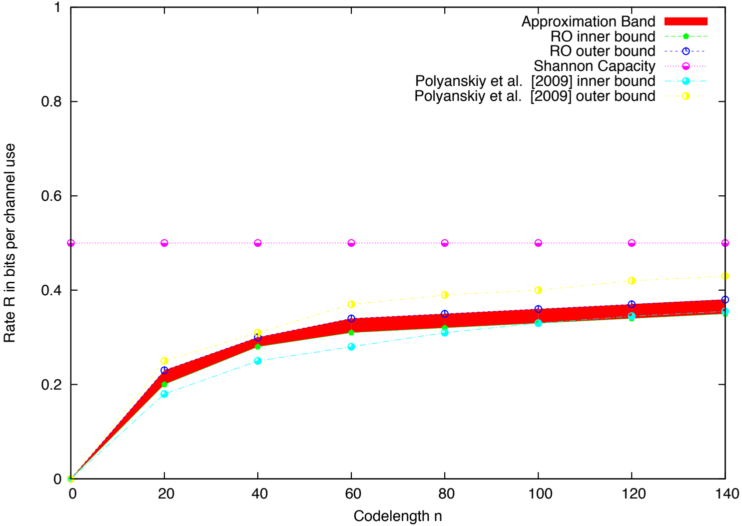
$$\begin{aligned} \min \quad & \mathrm{Tr}\,(\mathbf{X}) \\ \text{s.t.} \quad & \tilde{\mathbf{A}} \bullet \mathbf{X} \leq 0, \\ & \mathbf{X} \succeq 0, \end{aligned}$$

  and let $\mathbf{X}^0$ denote the optimal solution.

- For each iteration $k = 1, \ldots, K$, solve the optimization problem

$$\begin{aligned} \min \quad & \mathrm{Tr}\left(\left(\mathbf{X}^{k-1} + \delta I\right)^{-1} \mathbf{X}\right) \\ \text{s.t.} \quad & \tilde{\mathbf{A}} \bullet \mathbf{X} \leq 0, \\ & \mathbf{X} \succeq 0. \end{aligned}$$

Single User Gaussian Channel



Exponential single user channel

- Recall Typical set

$$\mathcal{U}_\epsilon(\lambda) = \left\{ \mathbf{z} \left| \frac{n}{\lambda} - \frac{\sqrt{n}}{\lambda} \cdot \Gamma_\epsilon^{\mathrm{E}} \le \sum_{j=1}^{n} z_j \le \frac{n}{\lambda} + \frac{\sqrt{n}}{\lambda} \cdot \Gamma_\epsilon^{\mathrm{E}},\ \mathbf{z} \ge \mathbf{0} \right. \right\}.$$

- MLE Decoder

$$\arg\min_{i \in \mathcal{B}(\mathbf{y})} \sum_{j=1}^{n} (y_j - x_{ij}),$$

where $\mathcal{B}(\mathbf{y}) = \left\{ i \in \mathcal{B} \middle| y_j \ge x_{ij},\ \forall j = 1, \ldots, n \right\}$.

- Generate $\mathbf{z}_t$ uniformly on $\mathcal{U}_\epsilon\left(\frac{\lambda}{1+2\nu}\right)$ to help us count the error probability.

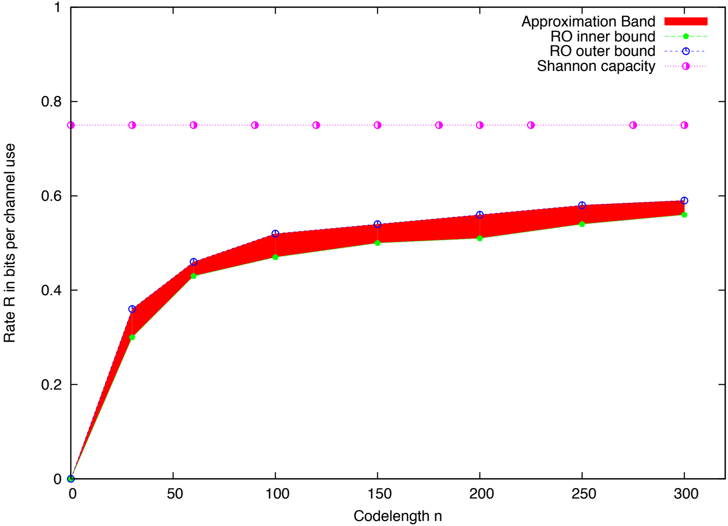
Algorithm

- Solve

$$\begin{aligned}
\max \quad & \sum_{i,k,t} v_{ikt} & \\
& \sum_{j=1}^{n} x_{ij} \leq nP, & \forall i = 1, \ldots, 2^{nR}, \\
& \sum_{j=1}^{n} x_{ij} + (2 - v_{it} - v_{ikt})\, M_0 \geq \sum_{j=1}^{n} x_{kj}, & \forall t, \forall i, k \neq i, \\
& x_{ij} + z_{tj} \geq x_{kj} - M_0 (1 - v_{ikt}), & \forall i, k, j, t, \\
& \sum_{t=1}^{T} v_{it} \geq (1 - \epsilon)\, T, & \forall i, \\
& v_{it}, v_{ikt} \in \{0, 1\}, & \forall i, k, t,
\end{aligned}$$

- If feasible, then $R \in \mathcal{R}_n^E [P, \lambda, 2\epsilon]$. Otherwise, then $R \notin \mathcal{R}_n^E \left[P, \dfrac{\lambda}{1 + 2\nu}, \epsilon\right]$.

Single User Exponential Channel



Two-User Gaussian Interference Channel

- Two transmitters, two receivers.
- User 1 selects a message i and transmits $\mathbf{x}_i^1$.
- User 2 selects a message k and transmits $\mathbf{x}_k^2$.
- The signal vectors $\mathbf{x}_i^1$ and $\mathbf{x}_k^2$ are power constrained, that is,

$$\left\|\mathbf{x}_i^1\right\|^2 \leq nP_1, \quad \left\|\mathbf{x}_i^2\right\|^2 \leq nP_2.$$

- The received signals $\mathbf{y}^1$, $\mathbf{y}^2$ are

$$\begin{aligned} \mathbf{y}^1 &= \mathbf{x}_i^1 + h_{12}\mathbf{x}_k^2 + \tilde{\mathbf{z}}_1, \\ \mathbf{y}^2 &= \mathbf{x}_k^2 + h_{21}\mathbf{x}_i^1 + \tilde{\mathbf{z}}_2, \end{aligned}$$

where h_{12}, h_{21} are interference parameters, and $\tilde{\mathbf{z}}_1$, $\tilde{\mathbf{z}}_2 \sim N(\mathbf{0}, \sigma \cdot I)$.

Decoding

Suppose the noise $\tilde{\mathbf{z}}$ is distributed **uniformly** in $\mathcal{B}_n(r) = \{\mathbf{z} \in \mathbb{R}^n \,|\, \|\mathbf{z}\| \leq r\}$. The maximum likelihood decoder for this channel is given by

$$i_1^* = \arg\max_{i \in \mathcal{M}^1} \left|\mathcal{B}_i^1\right|, \text{ where } \mathcal{B}_i^1 = \left\{k \in \mathcal{M}^2 \;:\; \left\|\mathbf{y}^1 - \left(\mathbf{x}_i^1 + h_{12}\mathbf{x}_k^2\right)\right\| \leq r\right\}$$

$$i_2^* = \arg\max_{i \in \mathcal{M}^2} \left|\mathcal{B}_i^2\right|, \text{ where } \mathcal{B}_i^2 = \left\{k \in \mathcal{M}^1 \;:\; \left\|\mathbf{y}^2 - \left(\mathbf{x}_i^2 + h_{21}\mathbf{x}_k^1\right)\right\| \leq r\right\}.$$

Parameters

- Parameter γ_ϵ: $\mathbb{P}\left[\|\tilde{\mathbf{z}}_G\| \leq \gamma_\epsilon\right] \geq 1-\epsilon$, $\tilde{\mathbf{z}}_G \sim N\left(\mathbf{0}, \sigma \cdot I\right)$.
- $T = \left(\frac{1+\nu}{\eta\nu} \cdot \frac{\gamma_\epsilon}{\sqrt{n}}\right)^n, \text{ with } \eta = \frac{\sigma}{\sqrt{n}} \cdot \Phi^{-1}\left(1-\epsilon^{1/4}\right),$
- $M_0 = (1+\nu) \cdot \gamma_\epsilon$.
- $\mathcal{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_T\}$ as before, $\|\mathbf{z}_t\| = M_0, \ t = 1, \ldots, T$.
- "Cushion" parameters: δ_1, δ_2, α_1, α_2 with the property that as $\nu \to 0$ and $n \to \infty$, $\delta_1, \delta_2 \to 0$ and $\alpha_1, \alpha_2 \to 1$.

Decoding

- For correct decoding we would require

$$\left|\left\{k' \in \mathcal{B}^2 \mid \left\|\mathbf{y}^1 - \mathbf{x}^1_{i_1^*} - h_{12}\mathbf{x}^2_k\right\|^2 \leq M_0^2\right\}\right| \geq$$

$$\left|\left\{k' \in \mathcal{B}^2 \mid \left\|\mathbf{y}^1 - \mathbf{x}^1_i - h_{12}\mathbf{x}^2_k\right\|^2 \leq M_0^2\right\}\right|, \forall i \in \mathcal{M}^1.$$

- We create a "cushion" δ_1, and instead require

$$\left|\left\{k' \in \mathcal{B}^2 \; : \; \left\|\mathbf{y}^1 - \mathbf{x}^1_{i_1^*} - h_{12}\mathbf{x}^2_k\right\|^2 \leq M_0^2 + \delta_1\right\}\right| \geq$$

$$\left|\left\{k' \in \mathcal{B}^2 \; : \; \left\|\mathbf{y}^1 - \mathbf{x}^1_i - h_{12}\mathbf{x}^2_k\right\|^2 \leq M_0^2 - \delta_1\right\}\right|, \forall i \in \mathcal{M}^1.$$

- Using this decoder, we want to ensure that the average probability of error is at most $\epsilon^{1/4}$, that is,

$$\frac{1}{M_2} \sum_{k \in \mathcal{M}^2} \mathbb{P}\left[g^1\left(\mathbf{y}^1\right) \neq i \,\middle|\, m^1 = i, m^2 = k\right] \leq \epsilon^{1/4}.$$

Variables

- "Counting" variables $\left\{v_i^1, v_{ik}^1, v_{ikt}^1\right\}_{i\in\mathcal{M}^1, k\in\mathcal{M}^2, t\in\mathcal{T}}$:

$$v_{ikt}^1 = \begin{cases} 1, & \text{if } \left|\mathcal{B}_{ikt,i'}^1\right| \le \left|\mathcal{B}_{ikt,i}^1\right|, \ \forall i' \in \mathcal{M}^1, \\ 0, & \text{otherwise,} \end{cases}$$

$$v_{ik}^1 = \begin{cases} 1, & \text{if } \sum_{t\in\mathcal{T}} v_{ikt}^1 \ge \left(1 - \epsilon^{1/4}\right) \cdot T, \\ 0, & \text{otherwise,} \end{cases}$$

$$v_i^1 = \begin{cases} 1, & \text{if } \sum_{k\in\mathcal{M}^2} v_{ik}^1 \ge \left(1 - \epsilon^{1/4}\right) \cdot M_2, \\ 0, & \text{otherwise,} \end{cases}$$

where

$$\mathcal{B}_{ikt,i'}^1 = \left\{k' \in \mathcal{B}^2 \ : \ \left\|\mathbf{x}_i^1 - \mathbf{x}_{i'}^1 + h_{12}\left(\mathbf{x}_k^2 - \mathbf{x}_{k'}^2\right) + \mathbf{z}_t\right\|^2 \le M_0^2 - \delta_1\right\},$$

$$\mathcal{B}_{ikt,i}^1 = \left\{k' \in \mathcal{B}^2 \ : \ \left\|h_{12}\left(\mathbf{x}_k^2 - \mathbf{x}_{k'}^2\right) + \mathbf{z}_t\right\|^2 \le M_0^2 + \delta_1\right\}.$$

More Variables

- $$v^1_{ii'kk't} = \begin{cases} 1, & \text{if } \left\| \mathbf{x}^1_i - \mathbf{x}^1_{i'} + h_{12}\left(\mathbf{x}^2_k - \mathbf{x}^2_{k'}\right) + \mathbf{z}_t \right\|^2 \leq M_0^2 - \delta_1, \\ 0, & \text{otherwise,} \end{cases}$$
$$v^1_{iikk't} = \begin{cases} 1, & \text{if } \left\| h_{12}\left(\mathbf{x}^2_k - \mathbf{x}^2_{k'}\right) + \mathbf{z}_t \right\|^2 \leq M_0^2 + \delta_1, \\ 0, & \text{otherwise.} \end{cases}$$
- The variables $\left\{ v^2_k, v^2_{ki}, v^2_{kit}, v^2_{kk'ii't} \right\}$ corresponding to User 2 are defined in a similar manner.

Algorithm: **Input** : n, R_1, R_2, σ, P_1, P_2, ϵ, ν.

$$\left\|\mathbf{x}_i^1\right\|^2 \leq nP_1,$$

$$\left\|\mathbf{x}_i^1 - \mathbf{x}_{i'}^1 + h_{12}\left(\mathbf{x}_k^2 - \mathbf{x}_{k'}^2\right) + \mathbf{z}_t\right\|^2 \leq M_0^2 - \delta_1 + \left(1 - v_{ii'kk't}^1\right) M_0^2,$$

$$\left\|h_{12}\left(\mathbf{x}_k^2 - \mathbf{x}_{k'}^2\right) + \mathbf{z}_t\right\|^2 \leq M_0^2 + \delta_1 + \left(1 - v_{iikk't}^1\right) M_0^2,$$

$$\sum_{k'=1}^{M_2} v_{iikk't}^1 \geq \sum_{k'=1}^{M_2} v_{ii'kk't}^1,$$

$$v_{ii'kk't}^1 \leq v_{ikt}^1, \quad v_{ikt}^1 \leq v_{ik}^1, \quad v_{ik}^1 \leq v_i^1,$$

$$\sum_{t=1}^{T} v_{ikt}^1 \geq \left(1 - \epsilon^{1/4}\right) \cdot T \cdot v_{ik}^1,$$

$$\sum_{k=1}^{M_2} v_{ik}^1 \geq \left(1 - \epsilon^{1/4}\right) \cdot M_2 \cdot v_i^1,$$

$$\sum_{i=1}^{M_1} v_i^1 \geq \left(1 - \epsilon^{1/4}\right) \cdot M_1,$$

Algorithm Continued

$$\left\|\mathbf{x}_k^2\right\|^2 \leq nP_2,$$

$$\left\|\mathbf{x}_k^2 - \mathbf{x}_{k'}^2 + h_{21}\left(\mathbf{x}_i^1 - \mathbf{x}_{i'}^1\right) + \mathbf{z}_t\right\|^2 \leq M_0^2 - \delta_2 + \left(1 - v_{kk'ii't}^2\right) M_0^2,$$

$$\left\|h_{21}\left(\mathbf{x}_i^1 - \mathbf{x}_{i'}^1\right) + \mathbf{z}_t\right\|^2 \leq M_0^2 + \delta_2 + \left(1 - v_{kkii't}^2\right) M_0^2,$$

$$\sum_{i'=1}^{M_1} v_{kkii't}^2 \geq \sum_{i'=1}^{M_1} v_{kk'ii't}^2,$$

$$v_{kk'ii't}^2 \leq v_{kit}^2, \quad v_{kit}^2 \leq v_{ki}^2, \quad v_{ki}^2 \leq v_k^2,$$

$$\sum_{t=1}^{T} v_{kit}^2 \geq \left(1 - \epsilon^{1/4}\right) \cdot T \cdot v_{ki}^2,$$

$$\sum_{i=1}^{M_1} v_{ki}^2 \geq \left(1 - \epsilon^{1/4}\right) \cdot M_1 \cdot v_k^2,$$

$$\sum_{k=1}^{M_2} v_k^2 \geq \left(1 - \epsilon^{1/4}\right) \cdot M_2,$$

Central Result

Capacity Region

- Reformulate as a rank minimization SOP

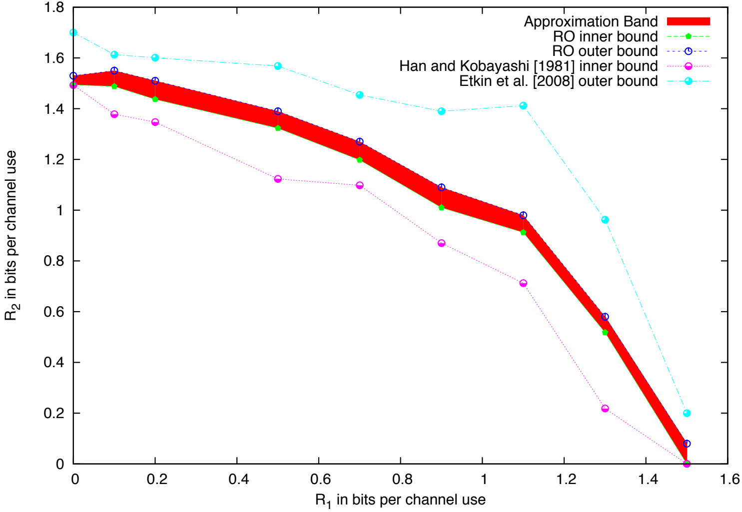
$$
\begin{aligned}
r^* = \min \quad & \operatorname{rank}(\mathbf{Y}) \\
\text{s.t.} \quad & \mathbf{A}_i \bullet \mathbf{Y} \leq 0, \\
& \mathbf{B}_i \bullet \mathbf{Y} = 0, \\
& \mathbf{Y} \succeq \mathbf{0}.
\end{aligned}
$$

- If $r^* = 1$, then $(R_1, R_2) \in \mathcal{R}_n^{\mathrm{IC}} \left[\alpha_1 P_1, \alpha_2 P_2, \frac{h_{12}}{\alpha_2}, \frac{h_{21}}{\alpha_1}, \sigma, O(\epsilon^{1/4}) \right]$,
- If $r^* \geq 2$, then $(R_1, R_2) \notin \mathcal{R}_n^{\mathrm{IC}} \left[P_1, P_2, h_{12}, h_{21}, (1+3\nu)\sigma, O(\epsilon) \right]$,
- Note that as $n \to \infty, \epsilon, \ \nu \to 0, \ \alpha_1, \alpha_2 \to 1$ and *the characterization of the asymptotic capacity is tight.*

Optimization problem as a function of the noise

Noise	Typical Set	Optimization Problem
Gaussian (independent)	Ball	*Rank minimization* with semidefinite constraints
Gaussian (correlated)	Ellipsoid	*Rank minimization* with semidefinite constraints
Exponential	Polyhedron	*Binary mixed linear* optimization problem
Uniform	Polyhedron	*Binary mixed linear* optimization problem
Binary symmetric noise	Polyhedron	*Binary optimization* problem

Two-user Gaussian Channel, $n = 60$



Extensions

- Multi-Cast channel.
- Multi-Access channel.
- Many transmitters, many receivers.

Conclusions

- The reason size becomes an issue is because of error probability guarantees.
- RO brings the power of *optimization* to the analysis of information theory.
- If underlying problem mixed binary, we can solve $n = 300 - 500$ and $M = 200,000$.
- If underlying problem SOP with rank constraints, we can solve $n = 100 - 150$ and $M = 100,000$.