Going Beyond 1st and 2nd Moment Constraints in Ditributionally Robust Optimization

Erick Delage[1] Yinyu Ye[2]

[1]HEC Montréal, Montreal, Canada

[2]Stanford University, California

Chicago, August 26, 2009

Parameter Uncertainty in Optimization

Consider an optimization problem:

$$\underset{\mathbf{x}\in\mathcal{X}}{\text{minimize}} \qquad h(\mathbf{x},\xi)$$

- $h(\cdot,\cdot)$ is a profit function
- $\mathbf{x} \in \mathbb{R}^n$ is a vector of decision variables
- $\xi \in \mathbb{R}^m$ is a vector of parameters

Often in practice, some parameters cannot be pre-determined :

- Tomorrow's value of a stock
- Next year's demand for a flight

Limits of Stochastic Programming

Given that $\xi \sim G$, one option is to consider a stochastic program :

$$\text{(SP)} \qquad \underset{\mathbf{x}\in\mathcal{X}}{\text{minimize}} \ \mathbb{E}_G[h(\mathbf{x},\xi)]$$

Difficulty:

- Developing a reliable probabilistic model is hard !
- Solution is sensitive to the choice of probabilistic model

Distributionally Robust Optimization

Use information about the distribution to define a set $\mathcal{D}$, such that $F \in \mathcal{D}$, then consider the distributionally robust stochastic program:

$$(DRSP) \qquad \underset{\mathbf{x}\in\mathcal{X}}{\text{minimize}} \ \max_{F\in\mathcal{D}} \ \mathbb{E}_F[h(\mathbf{x},\xi)]$$

The conflict:

- The set $\mathcal{D}$ should capture available information for F
- There must exist a tractable algorithm that finds a solution

The Spectrum of Known Computational Difficulties

Given that the objective function is convex in $\mathbf{x}$ and "piecewise concave" in ξ, there are known tractable algorithms for:

- $\mathcal{D}$ fixes the support and mean [Dupacova (1987)]
- $\mathcal{D}$ fixes the mean and covariance matrix [Popescu (2007), Bertsimas *et al*. (2009), Natarajan *et al*. (2008), Delage *et al*. (2009)]
- $\mathcal{D}$ fixes the support and mean, and imposes "upper-bound" on covariance matrix [Delage *et al*. (2009)]

While the following are intractable forms [Bertsimas *et al*. (2005)] :

- $\mathcal{D}$ fixes the support, the mean and covariance matrix
- $\mathcal{D}$ fixes the d-th first moments with $d \geq 4$

Outline

Outline

1. Introduction
2. Extending the Spectrum of Tractable DRSP
3. Accounting for the d First Moments
4. Discussion & Conclusion

Weak Conditions on Objective Function

Let $h(\mathbf{x}, \xi) = \max_{k \in \{1,2,\ldots,K\}} h_k(\mathbf{x}, \xi)$ be such that for all k:

- $h_k(\mathbf{x}, \xi)$ is convex in $\mathbf{x}$
- $h_k(\mathbf{x}, \xi)$ is concave in ξ
- $h_k(\mathbf{x}, \xi)$'s value and "sub-gradients" are easily obtained

Examples of Objectives

- A two-stage stochastic linear program with cost uncertainty

$$\underset{\mathbf{x}}{\text{minimize}} \; \max_{F\in\mathcal{D}} \mathbb{E}_F[\min_{\mathbf{y}\in\mathcal{Y}(\mathbf{x})} c^\mathsf{T}\mathbf{x} + \xi^\mathsf{T}\mathbf{y}]$$

- An expected utility maximization problem with piecewise linear concave utility

$$\underset{\mathbf{x}}{\text{maximize}} \; \min_{F\in\mathcal{D}} \mathbb{E}_F[u(-h(\mathbf{x},\xi)))]$$

- An optimized certainty equivalent with piecewise linear concave utility [Bertsimas *et al.* (2009)]

$$\underset{\mathbf{x},t}{\text{maximize}} \; \min_{F\in\mathcal{D}} t + \mathbb{E}_F[u(-h(\mathbf{x},\xi)-t)]$$

Weak Conditions on $\mathcal{D}$

Let $\mathcal{D}_\Upsilon$ be a set of the form

$$\mathcal{D}_\Upsilon(\mathcal{S}, \psi(\cdot), \mathbf{b}, \mathcal{K}) = \left\{ F \in \mathcal{M} \,\middle|\, \begin{array}{l} \mathbb{P}_F(\xi \in \mathcal{S}) = 1 \\ \mathbf{z}^\mathsf{T} \left(\mathbf{b} - \mathbb{E}_F[\psi(\xi)]\right) \geq 0 \ , \ \forall \mathbf{z} \in \mathcal{K} \end{array} \right\}$$

such that:

1. The set $\mathcal{S}$ is closed, convex, and bounded
2. The set $\mathcal{K} \subseteq \mathbb{R}^p$ is a convex cone
3. $\forall \mathbf{z} \in \mathcal{K}$, the function $g(\mathbf{z}, \xi) = \mathbf{z}^\mathsf{T} \psi(\xi)$ is convex in ξ
4. One can in polynomial time :
    - evaluate the function $g(\mathbf{z}, \xi)$
    - find a sub-gradient of $g(\mathbf{z}, \xi)$ in ξ
    - call a separation oracle for $\mathbf{z} \in \mathcal{K}$

Existence of Efficient Algorithm

Theorem

If $h(\mathbf{x}, \xi)$ and $\mathcal{D}_\Upsilon(\mathcal{S}, \psi(\cdot), \mathbf{b}, \mathcal{K})$ satisfy our conditions, the distributionally robust problem:

$$\underset{\mathbf{x}\in\mathcal{X}}{\text{minimize}} \quad \max_{F\in\mathcal{D}_\Upsilon(\mathcal{S},\psi(\cdot),\mathbf{b},\mathcal{K})} \quad \mathbb{E}_F[h(\mathbf{x}, \xi)]$$

can be solved in polynomial time using a cutting plane algorithm.

Summary of Proof I: Verifying that $\mathcal{D}_\Upsilon$ is Non-Empty

- One finds a distribution that lies in the strict interior of $\mathcal{D}_\Upsilon$ by verifying that $t^* > 0$ for the problem:

$$\begin{aligned} \underset{F\in\mathcal{M},t\in\mathbb{R}}{\text{maximize}} \quad & t \\ \text{subject to} \quad & \mathbf{z}^\mathsf{T}(\mathbf{b} - \mathbb{E}_F[\psi(\xi)] - t\mathbf{r}) \geq 0\ ,\ \forall\, \mathbf{z} \in \mathcal{K} \\ & \mathbb{P}_F(\xi \in \mathcal{S}) = 1\ , \end{aligned}$$

  where $\mathbf{r}$ is any non-zero vector lying in the strict interior of $\mathcal{K}^*$.
- By Jensen's inequality, the feasibility problem is equivalent to the semi-infinite convex problem:

$$\begin{aligned} \underset{\mu\in\mathcal{S},t\in\mathbb{R}}{\text{maximize}} \quad & t \\ \text{subject to} \quad & \mathbf{z}^\mathsf{T}(\mathbf{b} - \psi(\mu) - t\mathbf{r}) \geq 0\ ,\ \forall\, \mathbf{z} \in \mathcal{K}\ , \end{aligned}$$

  which can be solved using the ellipsoid method.

Summary of Proof II: Solving the DRSP

- Given that $\exists F \in \mathcal{D}_\Upsilon$ that strictly satisfies all moment constraints, then one can solve the equivalent problem:

$$\begin{array}{ll} \underset{\mathbf{x},\ \mathbf{z}}{\text{minimize}} & \mathbf{b}^\mathsf{T}\mathbf{z} \\ \text{subject to} & h(\mathbf{x},\xi) - \mathbf{z}^\mathsf{T}\psi(\xi) \ \leq\ 0\ ,\ \forall\,\xi \in \mathcal{S} \\ & \mathbf{z} \in \mathcal{K} \end{array}$$

- After verifying $\mathbf{z} \in \mathcal{K}$, each step of the ellipsoid method involves solving the convex problem :

$$\underset{\xi\in\mathcal{S}}{\text{maximize}} \ \max_{k\in\{1,2,\ldots,K\}} \ h_k(\mathbf{x},\xi) - \mathbf{z}^\mathsf{T}\psi(\xi)$$

Outline

An Alternative to Fixing d First Moments

- No tractable algorithms exists for the DRSP with

$$\mathcal{D}(\mathbf{b}) = \left\{F \middle| \mathbb{E}_F[\psi^d(\xi)] = \mathbf{b}\right\} = \left\{F \middle| \mathbf{z}^\mathsf{T}\mathbb{E}_F[\psi^d(\xi)] = \mathbf{z}^\mathsf{T}\mathbf{b} \,,\ \forall\, \mathbf{z}\right\} \,,$$

where $d \geq 0$ and $\psi^d_\gamma(\xi) = \prod_{i=1}^n \xi_i^{\gamma_i}$, $\forall\, \gamma \in \mathbb{Z}^m$, $\sum_{i=1}^n \gamma_i \leq d$

- For cutting plane algorithm to work we need that :

$$\mathcal{D}_\Upsilon \supseteq \left\{F \middle| \mathbf{z}^\mathsf{T}\mathbb{E}_F[\psi^d(\xi)] \leq \mathbf{z}^\mathsf{T}\mathbf{b} \,,\ \forall\, \mathbf{z} \in \mathcal{K}_{\mathsf{cvx}}\right\} \supseteq \mathcal{D}(\mathbf{b}) \,,$$

where $\mathcal{K}_{\mathsf{cvx}} = \left\{\mathbf{z} \middle| \mathbf{z}^\mathsf{T}\psi^d(\xi) \text{ is convex polynomial}\right\}$

- Yet, verifying that $\mathbf{z} \in \mathcal{K}_{\mathsf{cvx}}$ is not easy:

$$\mathbf{z} \in \mathcal{K}_{\mathsf{cvx}} \ \Leftrightarrow\ \sum_{i,j} \zeta_i \zeta_j \frac{\partial^2 \mathbf{z}^\mathsf{T}\psi^d(\xi)}{\partial \xi_i \partial \xi_j} \geq 0 \,,\ \forall\, \xi,\ \zeta$$

Designing $\mathcal{D}_\Upsilon$ using Sum of Squares Theory (c.f. Rajwade (1993))

- We propose $\mathcal{D}_\Upsilon(\mathbb{R}^m, \psi^d(\cdot), \mathbf{b}, \mathcal{K}_{\text{sos}})$ where $\mathcal{K}_{\text{sos}} \subseteq \mathcal{K}_{\text{cvx}}$ is

$$\left\{ \mathbf{z} \middle| \exists Q \succeq 0,\ \forall \xi, \zeta \in \mathbb{R}^m,\ \sum_{i,j} \zeta_i \zeta_j \frac{\partial^2 \mathbf{z}^\mathsf{T} \psi^d(\xi)}{\partial \xi_i \partial \xi_j} = \|Q^{1/2}(\zeta \otimes \psi^{d/2}(\xi))\|^2 \right\}$$

- Membership of $\mathcal{K}_{\text{sos}}$ can be verified with an LMI:

$$\mathbf{z} \in \mathcal{K}_{\text{sos}} \Leftrightarrow A_Q Q = A_\mathbf{z} \mathbf{z} \text{ for some } Q \succeq 0$$

- If $m = 1$ or $d \leq 2$, then $\mathcal{K}_{\text{sos}} = \mathcal{K}_{\text{cvx}}$
- In general, F^* is more concentrated than $G \in \mathcal{D}(\mathbf{b})$ since $\forall\, \mu_0$

$$\exists \mathbf{z} \in \mathcal{K}_{\text{sos}}\ ,\ \|\xi - \mu_0\|^2 = \mathbf{z}^\mathsf{T} \psi^d(\xi)$$
$$\Rightarrow\ \mathbb{E}_{F^*}[\|\xi - \mu_0\|^2] \leq \mathbb{E}_G[\|\xi - \mu_0\|^2]\ ,\ \forall\, G \in \mathcal{D}(\mathbf{b})$$

Experiments with Portfolio Optimization

Consider the portfolio optimization model:

$$\underset{\mathbf{x}\in\mathcal{X}}{\text{maximize}} \quad \min_{F\in\mathcal{D}_\Upsilon(\mathbb{R}^m,\psi^d(\cdot),\mathbf{b},\mathcal{K}_{\mathsf{sos}})} \mathbb{E}_F[\min_k \alpha_k \xi^\mathsf{T}\mathbf{x} + \beta_k]$$

- There is an SDP, solved in $O(K^{1.5} n^{3.25d})$, which maximizes a lower bound for this problem (again exploiting SOS theory)

$$\mathbf{z}^\mathsf{T}\psi^d(\xi) - h(\mathbf{x},\xi) \geq 0 \;,\; \forall\,\xi$$

$$\Leftrightarrow\; \mathbf{z}^\mathsf{T}\psi^d(\xi) + \alpha_k\xi^\mathsf{T}\mathbf{x} + \beta_k \geq 0 \;,\; \forall\,\xi \;,\; \forall\,k$$

$$\Leftarrow\; \exists Q_k \succeq 0,\; \mathbf{z}^\mathsf{T}\psi^d(\xi) + \alpha_k\xi^\mathsf{T}\mathbf{x} + \beta_k = \|Q_k^{1/2}\psi^{d/2}(\xi)\|^2,\; \forall\,\xi,\; \forall\,k$$

$$\Leftrightarrow\; \exists Q_k \succeq 0 \;,\; A_Q Q_k = A_\mathbf{z}\mathbf{z} + \alpha_k A_\mathbf{x}\mathbf{x} + \beta_k A_\beta \;,\; \forall\,k$$

Experiments with Portfolio Optimization (II)

Consider the portfolio optimization model:

$$\underset{\mathbf{x}\in\mathcal{X}}{\text{maximize}} \min_{F\in\mathcal{D}_\Upsilon(\mathbb{R}^m,\psi^d(\cdot),\mathbf{b},\mathcal{K}_{\text{sos}})} \mathbb{E}_F[\min_k \alpha_k \xi^\mathsf{T}\mathbf{x} + \beta_k]$$

- Empirical evidence indicates that if G is known, this DRSP approximates closely DRSP with $\mathcal{D} = \{G\}$ as d increases

Method	CPU Time	G is Discrete (10'000 points)		
		Obj. Est. Error		Rel. Perf.
	sec.	μ	σ	μ
$\mathcal{D}_\Upsilon(d=2)$	0.6	-3000%	1500%	-0.06%
$\mathcal{D}_\Upsilon(d=4)$	3.3	-6%	4%	-0.006%
$\mathcal{D}_\Upsilon(d=6)$	24	-4%	4%	-0.006%
Stoch. Prog.	8	0%	0%	0%

Outline

1. Introduction
2. Extending the Spectrum of Tractable DRSP
3. Accounting for the d First Moments
4. Discussion & Conclusion

Discussion & Conclusion

- There is a rich family of sets $\mathcal{D}_\Upsilon$ that makes DRSP tractable
- The set $\mathcal{D}_\Upsilon(\mathcal{S}, \psi^d(\cdot), \mathbf{b}, \mathcal{K}_{sos})$ allows to account for information about higher moments of G
- We almost have in hand a complete framework for optimization under uncertainty:
  1. Identify the right objective given your risk attitude
  2. Gather information about the distribution of ξ
  3. Use info to construct an uncertainty set: $\mathbb{E}_G[\psi^d(\xi)] \in \mathcal{B}$
  4. Solve the DRSP :

$$\underset{\mathbf{x}\in\mathcal{X}}{\text{minimize}} \max_{\mathbf{b}\in\mathcal{B}} \max_{F\in\mathcal{D}_\Upsilon(\mathcal{S},\psi^d(\cdot),\mathbf{b},\mathcal{K}_{sos})} \mathbb{E}_F[h(\mathbf{x}, \xi)]$$

  5. As more information is gathered, $\mathcal{B} \to \{\mathbb{E}_G[\psi^d(\xi)]\}$, thus $F^* \to G$ approximately given enough computational power

Thank You!