Distributionally Robust Optimization with Moment Ambiguity Sets

Suhan Zhong

Texas A&M University

Joint with Jiawang Nie (UCSD), Liu Yang (XTU) and Guangming Zhou (XTU)

Decision with Uncertainty

Many decision problems are involved with uncertainties.

Example (The news vendor problem)

Suppose a company has to decide about order quantity x of a certain product to satisfy demand d.

$$\text{The cost per unit} = \begin{cases} c & \text{for initial order} \\ b & \text{for additional order} \\ h & \text{for holding (if not used)} \end{cases}$$

The total cost is

$$F(x, d) = \max\{(c - b)x + bd, (c + h)x - hd\}.$$

We want to find a best x to minimize $F(x, d)$.

Model with Uncertainty

In the news vendor problem, the demand d is usually unknown.
But we can make a clever decision based on...

- historic data;
- market or other companies behavior;
- the difference of cost per unit for c, b, h...

How to describe the uncertainty in a stochastic programming model?

Stochastic Programming Models

We can proceed the uncertainty as a *random variable* denoted as ξ.

The stochastic programming (SP) model is

$$\min_{x \in X} \mathbb{E}[F(x, \xi)].$$

It optimizes the total cost *on average*.

In SP model, we assume the distribution of ξ is known, or can be efficiently observed from historic data.

Robust Optimization Model

We may also make a safest choice by optimizing over the worst-case.

This gives the robust optimization (RO) model

$$\min_{x \in X} \max_{\xi \in S} F(x, \xi).$$

The uncertainty is assumed to be freely distributed in some sets.

This model is often computationally tractable, but may produce too pessimistic decisions.

A Combination of SP and RO

Combining stochastic programming and robust optimization model may give more reasonable decisions sometimes.

Some information of random variables may be well estimated:

- support of the random variable;
- mean value and the covariance;
- other descriptive statistics...

The distributionally robust optimization is a combination of SP and RO models.

DROM

The distributionally robust optimization of moments (DROM) is

$$\begin{cases} \min\limits_{x \in X} & f(x) \\ s.t. & \inf\limits_{\mu \in \mathcal{M}} \mathbb{E}_\mu[h(x,\xi)] \geq 0, \end{cases}$$

where $x \in X \subseteq \mathbb{R}^n$, $\xi \in \mathbb{R}^p$, $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}$.

The $\mathcal{M}$ is an *ambiguity set* that describes the uncertain measure μ of ξ.

$$\mathcal{M} = \left\{ \mu : \operatorname{supp}(\mu) \subseteq S,\, \mathbb{E}_\mu\Big(\underbrace{\begin{bmatrix} 1 & \xi_1 & \cdots & (\xi_p)^d \end{bmatrix}^T}_{[\xi]_d} \Big) \in Y \right\}.$$

We focus on DRO that is given by polynomials and moment ambiguity.

Application: portfolio selection model

Consider the portfolio selection model

$$\min_{x \in \Delta} \max_{\mu \in \mathcal{M}} \mathbb{E}_\mu[x^T r(\xi)],$$

where $r(\xi) \in \mathbb{R}[\xi]^n$ is a vector of loss functions, and

$$x \in \Delta := \{x \in \mathbb{R}^n \,|\, x \geq 0,\, x_1 + \cdots + x_n = 1\}.$$

Assume each $\mu \in \mathcal{M}$ is a probability measure. The model is equivalent to

$$\begin{cases} \min\limits_{(x_0, x)} & x_0 \\ s.t. & \inf\limits_{\mu \in \mathcal{M}} \mathbb{E}_\mu[x_0 - x^T r(\xi)] \geq 0, \\ & x \in \Delta,\, x_0 \in \mathbb{R}. \end{cases}$$

Reformulation of Expectation Constraint

Consider the worst-case expectation constraint

$$\inf_{\mu \in \mathcal{M}} \mathbb{E}_{\mu}[h(x,\xi)] \geq 0.$$

If $h(x,\xi)$ is linear in x, i.e.,

$$h(x,\xi) \; = \; \langle h(x), [\xi]_d \rangle, \quad h(x) = Ax + b,$$

then the above constraint is equivalent to

$$\boxed{\begin{array}{l} \langle h(x), y \rangle \geq 0, \\ \forall y \in K := cone(\{\mathbb{E}_{\mu}([\xi]_d) \,|\, \mu \in \mathcal{M}\}). \end{array}}$$

The K is based on the moment ambiguity set

$$\mathcal{M} = \{\mu : \mathrm{supp}(\mu) \subseteq S, \, \mathbb{E}_{\mu}([\xi]_d) \in Y\}.$$

Reformulation of DRO

By the dual relation,

$$\langle h(x), y \rangle \geq 0,\ \forall y \in K \quad \Leftrightarrow \quad h(x) \in K^*,$$

where K^* is the dual cone given as

$$K^* := \{w | \langle w, y \rangle \geq 0,\ \forall y \in K\}.$$

Therefore, the DRO is equivalent to

$$\begin{cases} \min\limits_{x \in X} & f(x) \\ s.t. & h(x) = Ax + b \in K^*. \end{cases}$$

When f is linear, the above is a linear conic optimization problem.

Expression of Cones

The cones K, K^* are hard to describe computationally. Recall that

$$\begin{aligned} K &= cone(\{\mathbb{E}_\mu([\xi]_d) : \mu \in \mathcal{M}\}) \\ &= cone(Y) \quad \cap \quad cone(\{\mathbb{E}_\mu([\xi]_d) : \text{supp}(\mu) \subseteq S\}) \end{aligned}$$

Under some general conditions,

$$K^* = Y^* + \mathscr{P}_d(S),$$

where Y^* is the dual cone of Y and $\mathscr{P}_d(S)$ is a polynomial cone

$$\mathscr{P}_d(S) = \{q \in \mathbb{R}[\xi]_d : q(\xi) \geq 0, \forall \xi \in S\}.$$

The K, K^* can be expressed or approximated by SDP constraints.

Example 1: Expression of $\overline{cone(Y)}$

Given the constraining moment set

$$Y = \left\{ \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_5 \end{bmatrix} \in \mathbb{R}^6 \;\middle|\; \begin{array}{c} 1 \le y_0 \le y_1 \le y_2 \le \\ y_3 \le y_4 \le y_5 \le 2 \end{array} \right\}.$$

We get the closure of conic hull of Y as follows

$$\overline{cone(Y)} = \left\{ \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_5 \end{bmatrix} \in \mathbb{R}^6 \;\middle|\; \begin{array}{c} t \le y_0 \le y_1 \le y_2 \le \\ y_3 \le y_4 \le y_5 \le 2t \\ \text{for some } t \ge 0 \end{array} \right\}.$$

Example 2: SDP Expression of K

Let $\xi \in \mathbb{R}^1$, $S = [a_1, a_2]$, $a_1 < a_2$ and

$$\mathcal{M} = \left\{ \mu \;\middle|\; \begin{array}{c} \text{supp}(\mu) = S, \\ 0 \leq \mathbb{E}_\mu(\xi^i) \leq 1,\ i = 0, 1, \ldots, 4 \end{array} \right\},$$

The cone K has an exact SDP expression:

$$(y_0, y_1, y_2, y_3, y_4) \geq 0, \quad \begin{bmatrix} y_0 & y_1 & y_2 \\ y_1 & y_2 & y_3 \\ y_2 & y_3 & y_4 \end{bmatrix} \succeq 0,$$

$$(a_1 + a_2) \begin{bmatrix} y_1 & y_2 \\ y_2 & y_3 \end{bmatrix} \succeq a_1 a_2 \begin{bmatrix} y_0 & y_1 \\ y_1 & y_2 \end{bmatrix} + \begin{bmatrix} y_2 & y_3 \\ y_3 & y_4 \end{bmatrix}.$$

Moment Approximation of DRO

Assume $X \subseteq \mathbb{R}^n$, $S \subseteq \mathbb{R}^p$ are given by polynomial tuples

$$X = \{x : c(x) \geq 0\}, \quad S = \{\xi : g(\xi) \geq 0\}.$$

Under some general assumption, DRO is equivalent to

$$\begin{cases} \min & f(x) \\ s.t. & c(x) \geq 0, \\ & h(x) \in \mathscr{P}_d(S) + Y^*. \end{cases}$$

The above optimization has

- Polynomial constraints $c(x) \geq 0$,
- Conic constraints $h(x) \in \mathscr{P}_d(S) + Y^*$.

We need *moment* and *SOS* techniques to build convex relaxations.

Some Notations

- Sum-of-squares (SOS) polynomial

$$f = f_1^2 + \cdots + f_k^2, \quad f_i \in \mathbb{R}[x],\ k \in \mathbb{N}.$$

$\Sigma[x]$ = set of all SOS polynomials.

- Quadratic module of $g = (g_1, \ldots, g_m)$ in ξ

$$\mathrm{QM}[g] \ := \ \Sigma[\xi] + g_1 \cdot \Sigma[\xi] + \cdots + g_m \cdot \Sigma[\xi].$$

- Localizing matrix:

$$vec(a)^T (L_c^{(k)}[w]) vec(b) = \langle c \cdot ab, w \rangle, \quad \forall a, b \in \mathbb{R}[\xi].$$

The moment matrix $M_k[w] := L_1^{(k)}[w]$.

Polynomial Optimization

$$\begin{cases} \min\limits_{x \in \mathbb{R}^n} & f(x) \\ s.t. & c(x) \geq 0. \end{cases}$$

It can be solved globally by the Moment-SOS hierarchy.

$$\begin{cases} \min & \langle f, w \rangle \\ s.t. & M_k[w] \succeq 0, \; L_c^{(k)}[w] \succeq 0, \\ & w \in \mathbb{R}^{\mathbb{N}^n_{2k}} \end{cases} \qquad \begin{cases} \max\limits_{\gamma \in \mathbb{R}^1} & \gamma \\ s.t. & f - \gamma \in \mathrm{QM}[c]_{2k}. \end{cases}$$

Convergence of Moment-SOS relaxations

• Under compactness/archimedeanness, Moment-SOS hierarchy has asymptotic convergence. (Lasserre 01)

• Under the archimedeanness and some optimality conditions, Moment-SOS hierarchy has finite convergence. (Nie 14)

Convex relxation of DRO

$$\begin{cases} \min & f(x) \\ s.t. & c(x) \geq 0, \\ & h(x) \in \mathscr{P}_d(S) + Y^*. \end{cases} \tag{1}$$

has the moment restriction (k is the relaxation order)

$$\begin{cases} \min & \langle f, w \rangle \\ s.t. & w \in \mathbb{R}^{\mathbb{N}^n_{2k}}, \quad w_0 = 1, \\ & M_k[w] \succeq 0, \quad L_c^{(k)}[w] \succeq 0, \\ & h(w_{e_1}, \ldots, w_{e_n}) \in \mathrm{QM}[g] + Y^*. \end{cases} \tag{2}$$

Theorem (Nie-Yang-Z.-Zhou)

If $f, -c$ are SOS-convex. w^* is a minimizer of (2) if and only if $(w^*_{e_1}, \ldots, w^*_{e_n})$ is a minimizer of (1), under some general assumptions.

The Dual Pair

Recall the moment restriction

$$\begin{cases} \min_{w} & \langle f, w\rangle \\ s.t. & w \in \mathbb{R}^{\mathbb{N}^n_{2k}}, \quad w_0 = 1, \\ & M_k[w] \succeq 0, \quad L_c^{(k)}[w] \succeq 0, \\ & h(w_{e_1}, \dots, w_{e_n}) \in \mathrm{QM}[g] + Y^*. \end{cases}$$

Its dual problem is the SOS relaxation

$$\begin{cases} \max_{\gamma, y} & \gamma - \langle b, y\rangle \\ s.t. & f(x) - y^T Ax - \gamma \in \mathrm{QM}[c], \\ & \gamma \in \mathbb{R},\ y \in cone(Y), \\ & M_k[y] \succeq 0,\ L_g^{(k)}[y] \succeq 0. \end{cases}$$

Convergence of Moment-SOS Relaxations

The dual pair form Moment-SOS relaxations of DROM.

Theorem (Nie-Yang-Z.-Zhou)

Suppose ξ is univariate and $S = [a_1, a_2]$. The Moment-SOS relaxation is tight at the lowest k.

Theorem (Nie-Yang-Z.-Zhou)

Let $w^{(k)}$ be the optimizer of kth order moment relaxation. Under some general conditions, $(w^{(k)}_{e_1}, \ldots, w^{(k)}_{e_n})$ converges to global optimizer of DRO.

The convergence is finite, under some extra optimality conditions.

Verify Finite Convergence

The kth order SOS relaxation of DROM is

$$\begin{cases} \max\limits_{\gamma, y} & \gamma - \langle b, y \rangle \\ s.t. & f(x) - y^T A x - \gamma \in \mathrm{QM}[c], \\ & \gamma \in \mathbb{R},\ y \in cone(Y), \\ & M_k[y] \succeq 0,\ L_g^{(k)}[y] \succeq 0. \end{cases}$$

Let $y^{(k)}$ be its maximizer. Then

$$\text{relaxation is tight} \quad \Leftrightarrow \quad y^{(k)} \text{ admits a measure } \mu \in \mathcal{M}$$

A sufficient condition is

$$r := \operatorname{rank} M_{d-d_0}[y^{(k)}] = \operatorname{rank} M_d[y^{(k)}].$$

Recover the Worst-Case Measure

Suppose the rank condition holds

$$r := \operatorname{rank} M_{d-d_0}[y^{(k)}] = \operatorname{rank} M_d[y^{(k)}].$$

It leads to a measure in $\mathcal{M}$, i.e.,

$$\mu = \theta_1 \delta_{u_1} + \cdots + \theta_r \delta_{u_r},$$

where each $\theta_i > 0$ and δ_{u_i} is the Dirac measure supported at u_i.

The above μ is the worst-case measure associated with the decision.

Construct Moment Ambiguity Set

Suppose $T = \{\xi^{(1)}, \ldots, \xi^{(N)}\}$ is a given sample set for ξ. One can randomly choose $T_1, \ldots, T_s \subseteq T$ such that each T_i contains $\lceil N/2 \rceil$ samples. Choose a smaller sample size s, say, $s = 5$. For a given degree d, choose the moment vectors l, $u \in \mathbb{R}^{\mathbb{N}^n_d}$ such that

$$l_\alpha = \min_{j=1,\ldots,s} \left\{ \frac{1}{|T_j|} \sum_{i \in T_j} (\xi^{(i)})^\alpha, \; \frac{1}{|T \setminus T_j|} \sum_{i \in T \setminus T_j} (\xi^{(i)})^\alpha \right\},$$

$$u_\alpha = \max_{j=1,\ldots,s} \left\{ \frac{1}{|T_j|} \sum_{i \in T_j} (\xi^{(i)})^\alpha, \; \frac{1}{|T \setminus T_j|} \sum_{i \in T \setminus T_j} (\xi^{(i)})^\alpha \right\}$$

for every power $\alpha \in \mathbb{N}^n_d$. The moment constraining set Y can be estimated as

$$Y = \{ y \in \mathbb{R}^{\mathbb{N}^n_d} : l \leq y \leq u \}.$$

Example 3: DRO with Univariate ξ

$$\begin{cases} \min\limits_{x \in \mathbb{R}^4} & -x_1 - 2x_2 - x_3 + 2x_4 \\ s.t. & \inf\limits_{\mu \in \mathcal{M}} \mathbb{E}_\mu[h(x,\xi)] \geq 0, \\ & x \geq 0,\ 1 - e^T x \geq 0, \end{cases}$$

where $e = (1,1,1,1)^T$, $S = [0,3]$,

$$\begin{aligned} h(x,\xi) = (x_4 - x_1 - 2)\xi^5 + (x_4 - 1)\xi^4 + (2x_1 + x_2 + x_4 + 1)\xi^3 \\ + (2x_1 - x_2 + x_4 - 1)\xi^2 + (2 - x_2 - x_3)\xi, \end{aligned}$$

$$Y = \left\{ y = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_5 \end{bmatrix} \in \mathbb{R}^6 \ \middle|\ \begin{array}{c} 1 \leq y_0 \leq y_1 \leq y_2 \leq \\ y_3 \leq y_4 \leq y_5 \leq 2 \end{array} \right\}.$$

Example 3 (Continued)

Note ξ is univariate. The Moment-SOS relaxation is tight at the initial relaxation order $k = 3$. We solve for

$$F^* \approx -0.0326, \quad x^* \approx (0.6775, 0.0000, 0.0000, 0.3225).$$

$$y^* \approx (0.9355, 0.9355, 0.9517, 1.0163, 1.2260, 1.8710).$$

The measure μ for achieving $y^* = \int [\xi]_5 \mathtt{d}\mu$ is

$$\mu = 0.9315\delta_{u_1} + 0.0040\delta_{u_2}, \quad u_1 \approx 0.9913, \quad u_2 \approx 3.0000.$$

Example 4: DRO with SOS-Convex Functions

$$\begin{cases} \min\limits_{x\in\mathbb{R}^3} & (x_1 - x_3 + x_1x_3)^2 + (2x_2 + 2x_1x_2 - x_3^2)^2 \\ s.t. & \inf\limits_{\mu\in\mathcal{M}} \mathbb{E}_\mu[h(x,\xi)] \geq 0, \\ & 1 - x_1^2 - x_2^2 - x_3^2 \geq 0,\ 3x_3 - x_1^2 - 2x_2^4 \geq 0, \end{cases}$$

where $S = \{\xi \in \mathbb{R}^2 : 1 - \xi_1^2 - \xi_2^2 \geq 0\}$,

$$h(x,\xi) = (1 - x_3)\xi_1^2\xi_2^2 + (x_1 - x_2 + x_3 - 1)\xi_1\xi_2^2 + \\ (x_1 + x_2 + x_3 + 1)\xi_2^2 + (x_1 - x_3)\xi_1^2 - \xi_2,$$

$$Y = \left\{ y \in \mathbb{R}^{\mathbb{N}_4^2} \;\middle|\; \begin{array}{c} y_{00} = 1,\ 0.1 \leq y_\alpha \leq 1\,(0 < |\alpha| \leq 4) \\ \begin{pmatrix} y_{20} & y_{11} & y_{30} & y_{12} \\ y_{11} & y_{02} & y_{21} & y_{03} \\ y_{30} & y_{21} & y_{40} & y_{22} \\ y_{12} & y_{03} & y_{22} & y_{04} \end{pmatrix} \preceq 2I_4 \end{array} \right\}.$$

Example 4 (Continued)

Note f and all $-c_i$ are SOS-convex. The Moment-SOS relaxation is tight at the initial relaxation order $k = 2$. We solve for

$$F^* \approx 0.0160, \quad x^* \approx (0.4060, 0.0800, 0.4706).$$

$$y^* \approx (0.3180, 0.2750, 0.1411, 0.2436, 0.1137, 0.0744, 0.2199, 0.0950, 0.0552, 0.0460, 0.2011, 0.0819, 0.0426, 0.0318, 0.0318).$$

The measure μ for achieving $y^* = \int [\xi]_4 \mathtt{d}\mu$

$$\mu = 0.2527\delta_{u_1} + 0.7473\delta_{u_2},$$

$$u_1 \approx (0.6325, 0.7745), \quad u_2 \approx (0.9434, 0.3317).$$

Example 5: Nonconvex DRO

$$\begin{cases} \min_{x \in \mathbb{R}^3} & x_1^4 - 2x_1^2 + 2x_2^3 + x_3^4 \\ s.t. & \inf_{\mu \in \mathcal{M}} \mathbb{E}_\mu[h(x,\xi)] \geq 0. \\ & x_1^2 + x_2^2 + x_3^2 - 1 \geq 0, \\ & 4 - x_1^2 - 2x_2^2 - x_3 \geq 0, \end{cases}$$

where $S = \{\xi \in \mathbb{R}^2 | g(\xi) := (\xi_1, \xi_2, 1 - e^T\xi) \geq 0\}$,

$$\begin{aligned} h(x,\xi) = (x_1 + x_2 + 1)\xi_2^4 + (3x_1 + x_2)\xi_1^2\xi_2 + \\ (x_1 + 2x_2 + x_3 + 1)\xi_1^3 + 2x_1 + x_2 - 2x_3, \end{aligned}$$

$$Y = \left\{ y \in \mathbb{R}^{\mathbb{N}_4^2} \;\middle|\; \begin{array}{l} y_{00} = 1,\ 0.2^i \leq y_{i0} \leq 0.6^i, \\ y_{i0} \geq 1.2 y_{0i},\ i = 1,2,3,4 \end{array} \right\}.$$

Example 5 (Continued)

The ξ is bivariate. The f and $-c_1$ are not convex. But Moment-SOS relaxations still have finite convergence at order $k = 3$, with

$$F^* \approx -7.0017.$$

The measure for achieving $y^* = \int [\xi]_4 \mathtt{d}\mu$ is

$$\mu = 0.0877\delta_{u_1} + 0.9123\delta_{u_2},$$

$$u_1 \approx (0.0000, 1.0000), \quad u_2 \approx (0.6139, 0.3861).$$

The $x^* = \pi(w^*) \approx (0.2692, -1.5454, -0.8493)$ is feasible for the DROM.

$$F^* - f(x^*) \approx 1.2204 \cdot 10^{-7}.$$

So F^*, x^* are optimal value and solution for the DRO.

Example 6: Portfolio Selection

Consider the portfolio selection model given earlier.

$$\min_{x \in \Delta_3} \max_{\mu \in \mathcal{M}} \mathbb{E}_\mu \left[x_1 r_1(\xi) + x_2 r_2(\xi) + x_3 r_3(\xi) \right],$$

where $\Delta_3 := \left\{ x \in \mathbb{R}^3 \,|\, x_1 + x_2 + x_3 = 1,\ x \geq 0 \right\}$.

Suppose $p = 3$, $S = [0, 1]^3$, and

$$Y = \left\{ y \in \mathbb{R}^{\mathbb{N}^3_3} \,|\, y_{000} = 1,\ 0.1 \leq y_\alpha \leq 1,\ |\alpha| \geq 1 \right\},$$

$$\begin{aligned} r_1(\xi) &= -1 + \xi_1 + \xi_1\xi_2 - \xi_1\xi_3 - 2\xi_1^3, \\ r_2(\xi) &= -1 - \xi_1\xi_2 + \xi_2^2 - \xi_2\xi_3 + \xi_2^3, \\ r_3(\xi) &= -1 + \xi_2\xi_3 - \xi_3^2 - \xi_3^3. \end{aligned}$$

Example 6 (Continued)

The Moment-SOS relaxation has finite convergence at the initial order $k = 2$. We solve for

$$F^* \approx -1.0136, \quad x^* \approx (0.1492, 0.3501, 0.5007).$$

The measure for achieving $y^* = \int [\xi]_4 \mathtt{d}\mu$ is

$$\mu = 0.5560\delta_{u_1} + 0.4440\delta_{u_2},$$

$$u_1 \approx (0.4911, -0.0000, 0.1905), \quad u_2 \approx (0.7538, 1.0000, 0.6005).$$

Future Work

There is much future work to do.

- How can we solve the DRO if $h(x, \xi)$ is not linear in x?
- How to find more convenient conditions for the finite convergence of our method?
- How can we make our method applicable to DRO not equipped with moment ambiguity?

Thank you very much!