Distributionally robust optimization with sum-of-squares polynomial density functions and moment conditions

Krzysztof Postek (EUR)
Etienne de Klerk (Tilburg University & TU Delft)
Daniel Kuhn (EPFL)

Workshop on Distributionally Robust Optimization

Banff, 5th March 2018

Distributionally robust optimization

Consider an optimization problem where

- x is the decision variable
- z is an uncertain parameter with partly known probability distribution (measure) $\mu \in \mathcal{P}$ defined on a set $\mathbf{Z}$

$$\begin{aligned} \min_{x \in \mathbf{X}} \ & \sup_{\mu \in \mathcal{P}} \mathbb{E}_{\mu} f_0(x, z) \\ \text{s.t.} \ & \sup_{z \in \mathbf{Z}} f_j(x, z) \leq 0 \qquad j = 1, \ldots, J \end{aligned}$$

Distributionally robust optimization

- z is an uncertain parameter with partly known probability distribution (measure) $\mu \in \mathcal{P}$ defined on a set $\mathbf{Z}$

$$\sup_{\mu \in \mathcal{P}} \mathbb{E}_{\mu} f_0(x, z)$$

In this presentation, we only focus on the inner expectation-maximization problem, forget about x and set

$$f_0(x, z) = \phi_0(z)$$

Set of probability measures based on moments

We assume that $\mathcal{P} \subset \mathcal{M}$ is a family of measures defined on $\mathbf{Z}$ such that:

$$\mathbb{E}_\mu \phi_i(z) = b_i \qquad\qquad i = 1, \ldots, l$$

The expectation-maximization problem is:

$$\begin{aligned}
\max_{\mu \in \mathcal{M}} \ & \int_{\mathbf{Z}} \phi_0(z) d\mu \\
\text{s.t.} \ & \int_{\mathbf{Z}} 1 d\mu = 1 \\
& \int_{\mathbf{Z}} \phi_i(z) d\mu = b_i \qquad\qquad i = 1, \ldots, l
\end{aligned}$$

a.k.a. the Generalized Problem of Moments (GPM).

Example

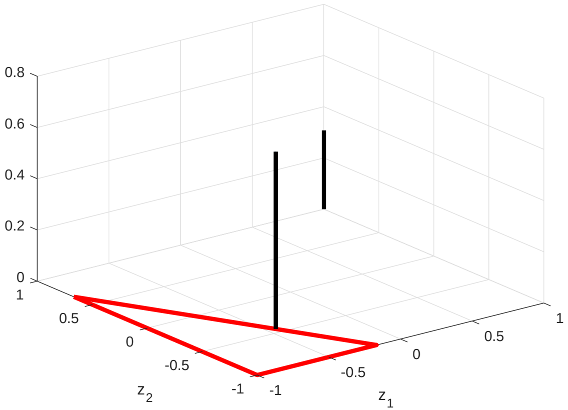
Consider $z = (z_1, z_2) \in [-1, 1]^2 = \mathbf{Z}$ such that

$$\int_{[-1,1]^2} z_1 d\mu = \int_{[-1,1]^2} z_2 d\mu = 0$$

Goal: evaluate the maximum probability $0.15z_1 + 0.075z_2 \leq -0.1$

$$\begin{aligned} \max_{\mu} \quad & \int_{[-1,1]^2} \mathbf{1}(\{(z_1, z_2) : \; 0.15z_1 + 0.075z_2 \leq -0.1\}) d\mu \\ \text{s.t.} \quad & \int_{[-1,1]^2} 1 d\mu = 1 \\ & \int_{[-1,1]^2} z_1 d\mu = \int_{[-1,1]^2} z_2 d\mu = 0 \end{aligned}$$

The worst-case distribution



$$\mu((-0.44, -0.44)) \approx 0.69, \;\; \mu((1, 1)) \approx 0.31$$

Discussion

- The worst-case distribution will always have at most $l + 1$ probability mass points (Rogosinsky, 1958)
- One does not expect this to be the case in many applications
- Therefore, distributionally robust optimization based on generalized moment problems can be over-conservative
- Need to model smooth probability density functions, e.g., polynomials

Using polynomials as smooth densities

$$\max_{h(z)} \int_{\mathbf{Z}} \phi_0(z)h(z)d\mu$$

$$\text{s.t.} \int_{\mathbf{Z}} h(z)d\mu = 1$$

$$\int_{\mathbf{Z}} \phi_i(z)h(z)d\mu = b_i \qquad i = 1, \ldots, I$$

where

- μ is some known reference measure (e.g. Lebesgue)
- $h(z)$ is a sum-of-squares (SOS) polynomial:

$$h(z) = \sum_{k=1}^{K} (a_i(z))^2,$$

where $a_i(z)$, $i = 1, \ldots, K$ are polynomials in z.

Forcing a polynomial to be SOS

Some notation:

- denote $z^\alpha = z_1^{\alpha_1} \cdot \ldots \cdot z_n^{\alpha_n}$
- define the set of all n-tuples of exponents of monomials of degree at most r:

$$N(n, r) = \left\{ \alpha \in \mathbb{N}^n : \sum_{i=1}^{n} \alpha_i \leq r \right\}$$

Forcing a polynomial to be SOS

Proposition

If a polynomial $h(z)$ of degree at most $2r$ can be written as

$$h(z) = \sum_{\alpha,\beta \in N(n,r)} H_{\alpha,\beta} z^\alpha z^\beta$$

$$= \begin{bmatrix} 1 \\ z_1 \\ z_2 \\ \vdots \\ z_n^r \end{bmatrix}^\top \begin{bmatrix} H_{1,1} & H_{1,2} & \cdots & H_{1,|N(n,r)|} \\ H_{2,1} & & & \\ \vdots & & \ddots & \\ & & & H_{|N(n,r)|,|N(n,r)|} \end{bmatrix} \begin{bmatrix} 1 \\ z_1 \\ z_2 \\ \vdots \\ z_n^r \end{bmatrix}$$

where $[H_{\alpha,\beta}]$ is a positive semidefinite matrix ($\forall y : \; y^\top H y \geq 0$), then $h(z)$ is an SOS polynomial.

SOS-based problem of moments

$$\sup_{H \succeq 0} \int_{\mathbf{Z}} \phi_0(z) \sum_{\alpha,\beta \in N(m,2r)} H_{\alpha,\beta} z^{\alpha+\beta} d\mu$$

$$\text{s.t.} \int_{\mathbf{Z}} \sum_{\alpha,\beta \in N(m,2r)} H_{\alpha,\beta} d\mu = 1$$

$$\int_{\mathbf{Z}} \phi_i(z) \sum_{\alpha,\beta \in N(m,2r)} H_{\alpha,\beta} z^{\alpha+\beta} d\mu = b_i, \qquad i = 1, \ldots, l,$$

equivalent to:

$$\sup_{H \succeq 0} \sum_{\alpha,\beta \in N(m,2r)} H_{\alpha,\beta} \int_{\mathbf{Z}} \phi_0(z) z^{\alpha+\beta} d\mu$$

$$\text{s.t.} \sum_{\alpha,\beta \in N(m,2r)} \int_{\mathbf{Z}} z^{\alpha+\beta} d\mu = 1$$

$$\sum_{\alpha,\beta \in N(m,2r)} \int_{\mathbf{Z}} \phi_i(z) z^{\alpha+\beta} d\mu = b_i, \qquad i = 1, \ldots, l,$$

Semidefinite programming form

This problem can be written as:

$$\begin{aligned} \max_{H \in \mathbb{S}^{|N(n,r)|}} \quad & \langle H, \Phi^0 \rangle \\ \text{s.t.} \quad & \langle H, E \rangle = 1 \\ & \langle H, \Phi^i \rangle = b_i \qquad i = 1, \ldots, I \\ & H \succeq 0 \end{aligned}$$

where $\langle A, B \rangle = \mathrm{Tr}(A^\top B)$ and where the matrices' entries are:

$$\Phi^0_{\alpha,\beta} = \int_{\mathbf{Z}} \phi_0(z) z^{\alpha+\beta} d\mu, \;\; E_{\alpha,\beta} = \int_{\mathbf{Z}} z^{\alpha+\beta} d\mu, \;\; \Phi^i_{\alpha,\beta} = \int_{\mathbf{Z}} \phi_i(z) z^{\alpha+\beta} d\mu$$

Our ability to compute these terms is crucial. Possible for several sets, e.g., when $\phi_0(z)$, $\phi_i(z)$ are polynomials.

Examples of known moments of monomials

Example

For the standard simplex, we have

$$\int_{\Delta_n} z^\alpha = \frac{\prod_{i=1}^n \alpha_i!}{(|\alpha| + n)!},$$

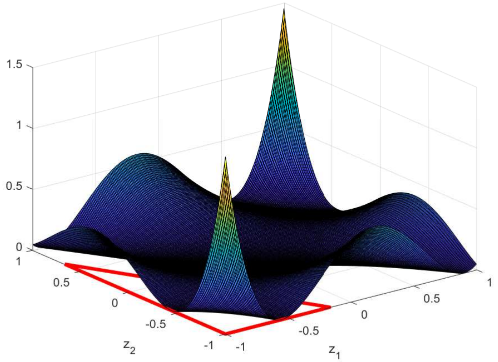
Example

For the hypercube $\mathcal{Q}_n$:

$$\int_{\mathcal{Q}_n} z^\alpha = \int_{\mathcal{Q}_n} x^\alpha dx = \prod_{i=1}^{n} \int_0^1 x_i^{\alpha_i} dx_i = \prod_{i=1}^{n} \frac{1}{\alpha_i + 1}.$$

Back to our example

Worst-case density obtained with polynomial degree $2r = 2$:



Worst-case probability: 0.3942 (compare with 0.6923)

Conjecture

As $r \to +\infty$, the optimal value of

$$\begin{aligned} \max_{H \in \mathbb{S}^{|N(n,r)|}} \; & \langle H, \Phi^0 \rangle \\ s.t. \; & \langle H, E \rangle = 1 \\ & \langle H, \Phi^i \rangle = b_i \qquad i = 1, \dots, l \\ & H \succeq 0 \end{aligned}$$

converges to the optimal value of

$$\begin{aligned} \max_{\mu} \; & \int_Z \phi_0(z) d\mu \\ s.t. \; & \int_{\mathbf{Z}} 1 d\mu = 1 \\ & \int_{\mathbf{Z}} \phi_i(z) d\mu = b_i \qquad i = 1, \dots, l. \end{aligned}$$

A reason behind the conjecture

For continuous $f(z)$ and convex $\mathbf{Z}$ the sequence of optimal values of

$$\min_{h(z)\in\Sigma_r(z)} \int_{\mathbf{Z}} f(z)h(z)d\mu$$
$$\text{s.t.} \int_{\mathbf{Z}} h(z)d\mu = 1.$$

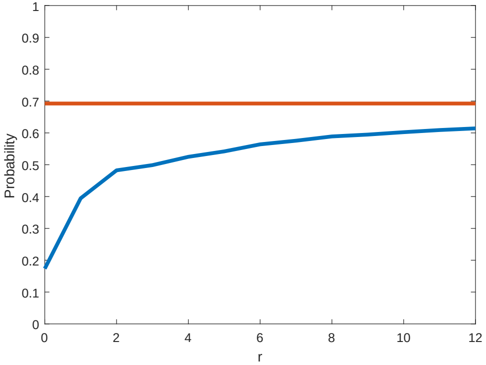
where Σ_r is the space of SOS polynomials of degree at most $2r$, converges (Lasserre, 2001) to:

$$\min_{z\in\mathbf{Z}} f(z).$$

as $r \to +\infty$.

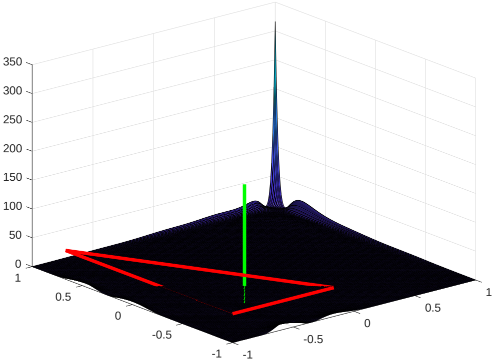
Theory - numerical investigation

r	Probability
0	0.1736
1	0.3946
2	0.4824
3	0.4988
4	0.5249
5	0.5419
6	0.5641
7	0.5755
8	0.5889
9	0.5947
10	0.6023
11	0.6090
12	0.6142
∞	0.6923



Back to our example

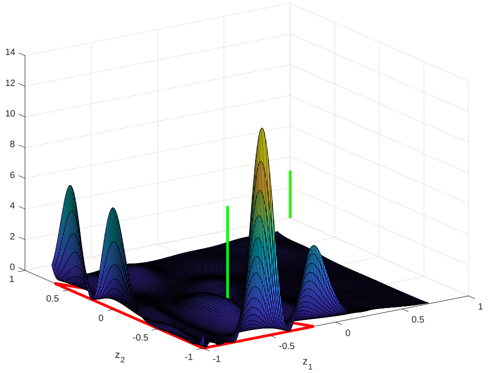
Worst-case density obtained with polynomial degree $2r = 24$:



Probability: 0.6142

Back to our example

Worst-case density obtained with polynomial degree $2r = 24$:



Probability: 0.6142

Computational heuristic

Instead of optimizing over a high-degree density $h(z)$ do:

1. Optimize a low-degree density polynomial $h_1(z)$.
2. Fix $\overline{h}_1(z)$, set the new probability density function as $\overline{h}_1(z)h_2(z)$, where $h_2(z)$ is the same degree as $\overline{h}_1(z)$, optimize over $h_2(z)$.
3. Fix $\overline{h}_2(z)$, set the new probability density function as $\overline{h}_1(z)\overline{h}_2(z)h_3(z)$, optimize over $h_3(z)$.
4. ...

We tested it also on several global optimization examples.

Conclusion

- we propose a new way of defining uncertain smooth probability measures
- the maximum expectation problem becomes an SDP
- proved (?) the convergence to the optimal value of a general problem of moments
- computational heuristic: modelling the polynomial density as a product of polynomial densities of smaller degree, optimized one after another

Thank you for your attention