SAND2016-2897C

Surrogate-based model for optimization under uncertainty

Jianqiang Cheng, Richard L.-Y. Chen, Habib Najm, Ali Pinar, Cosmin Safta, and Jean-Paul Watson

Sandia National Laboratories
Livermore, CA and Albuquerque, NM

SIAM UQ Lausanne, Switzerland
April 5-8, 2016

Outlines

Outlines

Outlines

Optimization under uncertainty

We consider an optimization problem under uncertainty as follows:

$$\min_{\mathbf{x}\in X} Q(\mathbf{x}, \xi) \tag{1}$$

- $\mathbf{X} \subset \mathbb{R}^n$ is a deterministic set of feasible solutions
- $\xi \in \mathcal{S} \subset \mathbb{R}^m$ with a distribution F where $\mathcal{S}$ is the support of ξ
- $Q(\mathbf{x}, \xi)$ is a cost function in $\mathbf{x}$ that depends on a random vector ξ

Optimization under uncertainty

Robust optimization

When only the support $\mathcal{S}$ is known, its worst-case problem is:

$$(RO) \quad \min_{\mathbf{x} \in X} \max_{\xi \in \mathcal{S}} Q(\mathbf{x}, \xi) \tag{2}$$

Optimization under uncertainty

Robust optimization

When only the support $\mathcal{S}$ is known, its worst-case problem is:

$$(RO) \quad \min_{\mathbf{x} \in X} \max_{\xi \in \mathcal{S}} Q(\mathbf{x}, \xi) \tag{2}$$

Stochastic optimization

When the distribution F is known, one popular model is as follows:

$$(SP) \quad \min_{\mathbf{x} \in X} \mathbb{E}_F[Q(\mathbf{x}, \xi)] \tag{3}$$

Optimization under uncertainty

Distributionally robust optimization

When the distribution F is partially known, such as support, first and second moments, we have anther optimization problem:

$$(DRO) \quad \min_{\mathbf{x}\in X} \max_{F\in\mathcal{D}} \mathbb{E}_F[Q(\mathbf{x},\xi)] \tag{4}$$

- $\mathcal{D}$ is an ambiguity set of F that encompasses the partial information on F, such as support, first and second moments.

Surrogate model

Why surrogate modeling?

- For some applications, the evaluation of the objective and constraint functions is computationally expensive

Surrogate model

Why surrogate modeling?

- For some applications, the evaluation of the objective and constraint functions is computationally expensive

Idea

- Constructing **approximation** models, known as surrogate models, that mimic the behavior of the target model as closely as possible while being **computationally cheaper** to evaluate

Polynomial model

Polynomials

A general polynomial with degree $d \in Z_+$:

$$H(\xi) = \sum_{|\boldsymbol{a}| \leq d} h_{\boldsymbol{a}} \xi^{\boldsymbol{a}} \tag{5}$$

- $h_{\boldsymbol{a}}$ is the coefficient and $\boldsymbol{a} \in \mathbb{Z}_+^m$
- $\xi^{\boldsymbol{a}} = \xi_1^{a_1} \xi_2^{a_2} \dots \xi_m^{a_m}$
- $|\boldsymbol{a}| = \sum_{i=1}^{m} a_i$

Distributionally robust optimization

We recall the targeted distributionally robust optimization problems as follows:

$$\min_{\mathbf{x}\in X} \max_{F\in\mathcal{D}} \mathbb{E}_F[Q(\mathbf{x},\xi)] \tag{6}$$

Distributionally robust optimization

We recall the targeted distributionally robust optimization problems as follows:

$$\min_{\mathbf{x}\in X} \max_{F\in\mathcal{D}} \mathbb{E}_F[Q(\mathbf{x},\xi)] \tag{6}$$

$$\mathcal{G}(\mathbf{x}) = \max_{F\in\mathcal{D}} \mathbb{E}_F[H(\mathbf{x},\xi)] \tag{7a}$$

$$\text{s.t.} \quad Q(\mathbf{x},\xi) \leq H(\mathbf{x},\xi), \quad \forall \xi \in \mathcal{S} \tag{7b}$$

- $H(\mathbf{x},\xi)$ strongly dominates function $Q(\mathbf{x},\xi)$ on the support $\mathcal{S}$
- $\max_{F\in\mathcal{D}} \mathbb{E}_F[Q(\mathbf{x},\xi)] \leq \mathcal{G}(\mathbf{x})$
- $\min_{\mathbf{x}\in X} \mathcal{G}(\mathbf{x})$ provides an upper bound for problem (10)

Surrogate model

We focus on polynomial surrogate models with degree $d \in Z_+$ for function $Q(\mathbf{x}, \xi)$, i.e.,

$$H(\mathbf{x}, \xi) = \sum_{|\boldsymbol{a}| \leq d} h_{\boldsymbol{a}}(x) \xi^{\boldsymbol{a}} \tag{8}$$

- Zero-degree polynomial surrogate: $H(\mathbf{x}, \xi) = h_0(\mathbf{x})$
- First-degree polynomial surrogate: $H(\mathbf{x}, \xi) = h_0(\mathbf{x}) + h_1(\mathbf{x})^T \xi$
- Second-degree polynomial surrogate: $H(\mathbf{x}, \xi) = h_0(\mathbf{x}) + h_1(\mathbf{x})^T \xi + \xi^T h_2(x) \xi$

where $h_0(\mathbf{x}) \in R$, $h_1(\mathbf{x})$ is a vector and $h_2(\mathbf{x})$ is a matrix.

Robust optimization

When $d = 0$, the surrogate model becomes

$$\min_{\mathbf{x}\in X} \quad h_0(x) \tag{9a}$$

$$\text{s.t.} \quad Q(\mathbf{x}, \xi) \leq h_0(x), \quad \forall \xi \in \mathcal{S} \tag{9b}$$

which is the robust variant of problem (3).
Remark: it also shows that: the RO problem is more conservative than the DRO problem.

Distributionally robust optimization

We recall the distributionally robust variant of problem (3)

$$\min_{\mathbf{x}\in X} \max_{F\in\mathcal{D}} \mathbb{E}_F[Q(\mathbf{x},\xi)] \tag{10}$$

where $\mathcal{D}$ is an ambiguity set of F.

Distributionally robust optimization

Assumption 1

The distributional uncertainty set accounts for information about the convex support $\mathcal{S}$, mean μ in the strict interior of $\mathcal{S}$, and an upper bound $\Sigma \succ 0$ on the covariance matrix of the random vector ξ

$$\mathcal{D}(\mathcal{S}, \mu, \Sigma) = \left\{ F \,\middle|\, \begin{array}{l} \mathbb{P}(\xi \in \mathcal{S}) = 1 \\ \mathbb{E}_F[\xi] = \mu \\ \mathbb{E}_F[(\xi - \mu)(\xi - \mu)^T] \preceq \Sigma \end{array} \right\}. \tag{11}$$

Distributionally robust optimization

Theorem

Under Assumption 1, if $Q(\mathbf{x}, \xi)$ is F-integrable for any $F \in \mathcal{D}$ and $H(\mathbf{x}, \xi)$ is convex in ξ for any $\mathbf{x} \in \mathbf{X}$, then problem (7) with the surrogate model of second-degree polynomial is equivalent to the distributionally robust optimization, i.e.,
$\min_{\mathbf{x} \in X} \max_{F \in \mathcal{D}} \mathbb{E}_F[Q(\mathbf{x}, \xi)]$.

Sketch of Proof

- Second-degree polynomial,
  $H(\mathbf{x}, \xi) = h_0(\mathbf{x}) + h_1(\mathbf{x})^T \xi + \xi^T h_2(x) \xi$

Sketch of Proof

- Second-degree polynomial,
  $H(\mathbf{x}, \xi) = h_0(\mathbf{x}) + h_1(\mathbf{x})^T \xi + \xi^T h_2(x) \xi$
- Since $H(\mathbf{x}, \xi)$ is convex in ξ for any $\mathbf{x} \in \mathbf{X}$, we have $h_2(x) \succeq 0$.

Sketch of Proof

- Second-degree polynomial,
  $H(\mathbf{x}, \xi) = h_0(\mathbf{x}) + h_1(\mathbf{x})^T \xi + \xi^T h_2(x) \xi$
- Since $H(\mathbf{x}, \xi)$ is convex in ξ for any $\mathbf{x} \in \mathbf{X}$, we have $h_2(x) \succeq 0$.
- Assumption 1 implies that $\mathbb{E}_F \xi\xi^T \preceq \Sigma + \mu\mu^T$

Sketch of Proof

- Second-degree polynomial, $H(\mathbf{x}, \xi) = h_0(\mathbf{x}) + h_1(\mathbf{x})^T \xi + \xi^T h_2(x) \xi$
- Since $H(\mathbf{x}, \xi)$ is convex in ξ for any $\mathbf{x} \in \mathbf{X}$, we have $h_2(x) \succeq 0$.
- Assumption 1 implies that $\mathbb{E}_F \xi\xi^T \preceq \Sigma + \mu\mu^T$
- The subsequent formulation of problem (7) is

$$\min_{\mathbf{x}, h_0(\mathbf{x}), h_1(\mathbf{x}), h_2(\mathbf{x})} \quad h_0(\mathbf{x}) + h_1(\mathbf{x})^T \mu + (\Sigma + \mu\mu^T) \bullet h_2(\mathbf{x}) \tag{12a}$$

$$\text{s.t.} \quad Q(\mathbf{x}, \xi) \leq h_0(\mathbf{x}) + h_1(\mathbf{x})^T \xi + \xi^T h_2(x) \xi, \quad \forall \xi \in \mathcal{S} \tag{12b}$$

$$h_2(x) \succeq 0, h_0(\mathbf{x}) \in \mathbb{R}, h_1(\mathbf{x}) \in \mathbb{R}^m, \tag{12c}$$

$$\mathbf{x} \in \mathbf{X} \tag{12d}$$

where "$\bullet$" is the inner product defined by $A \bullet B = \sum_{i,j} A_{ij} B_{ij}$.

Sketch of Proof

With the results of Lemma 1 in [1], problem (10) is equivalent to the following problem:

$$\begin{aligned}
\min_{\mathbf{x},r,v,V} \quad & r + v^T\mu + (\Sigma + \mu\mu^T) \bullet V && \text{(13a)}\\
\text{s.t.} \quad & Q(\mathbf{x},\xi) \leq r + v^T\xi + \xi^T V\xi, \quad \forall \xi \in \mathcal{S} && \text{(13b)}\\
& V \succeq 0, r \in \mathbb{R}, v \in \mathbb{R}^m && \text{(13c)}\\
& \mathbf{x} \in \mathbf{X} && \text{(13d)}
\end{aligned}$$

It is easy to find that problem (12) is equivalent to problem (13).

Distributionally robust optimization

We consider a more general ambiguity set, the same as that of [1]:

Assumption 2

The distributional uncertainty set accounts for information about the convex support $\mathcal{S}$, mean μ in the strict interior of $\mathcal{S}$, and an upper bound $\Sigma \succ 0$ on the covariance matrix of the random vector ξ

$$\mathcal{D}(\mathcal{S},\mu,\Sigma) = \left\{ F \left| \begin{array}{l} \mathbb{P}(\xi \in \mathcal{S}) = 1 \\ (\mathbb{E}_F[\xi] - \mu)^T \Sigma^{-1} (\mathbb{E}_F[\xi] - \mu) \le \rho_1 \\ \mathbb{E}_F[(\xi-\mu)(\xi-\mu)^T] \preceq \rho_2 \Sigma \end{array} \right. \right\} . \quad (14)$$

where $\rho_1, \rho_2 \ge 0$.

Distributionally robust optimization

Accordingly,the corresponding polynomial surrogate problem is:

$$\min_{\mathbf{x}\in\mathbf{X},h_0(\mathbf{x}),h_1(\mathbf{x}),h_2(\mathbf{x})\succeq 0} \quad \max_{\mathbb{E}_F[\xi],\mathbb{E}_F[\xi\xi^T]} h_0(\mathbf{x})+$$

$$h_1(\mathbf{x})^T\mathbb{E}_F[\xi]+\mathbb{E}_F[\xi\xi^T]\bullet h_2(\mathbf{x}) \tag{15a}$$

$$\text{s.t.} \quad Q(\mathbf{x},\xi)\leq h_0(\mathbf{x})+h_1(\mathbf{x})^T\xi+\xi^T h_2(x)\xi, \quad \forall\xi\in\mathcal{S} \tag{15b}$$

$$(\mathbb{E}_F[\xi]-\mu)^T\Sigma^{-1}(\mathbb{E}_F[\xi]-\mu)\leq\rho_1 \tag{15c}$$

$$\mathbb{E}_F[(\xi-\mu)(\xi-\mu)^T]\preceq\rho_2\Sigma \tag{15d}$$

Theorem

Under Assumption 2, if $Q(\mathbf{x}, \xi)$ is F-integrable for any $F \in \mathcal{D}$, then the corresponding polynomial surrogate problem, i.e, problem (15) is equivalent to the distributionally robust optimization, i.e., $\min_{\mathbf{x} \in X} \max_{F \in \mathcal{D}} \mathbb{E}_F[Q(\mathbf{x}, \xi)]$.

Sketch of Proof

- Given $\mathbf{x} \in \mathbf{X}$, $h_0(\mathbf{x})$, $h_1(\mathbf{x})$ and $h_2(\mathbf{x}) \succeq 0$, the inner maximum problem of problem (15) is equivalent to the following form:

$$\max_{\mathbf{p},\boldsymbol{P}} \quad h_0(\mathbf{x}) + h_1(\mathbf{x})^T\mathbf{p} + \boldsymbol{P} \bullet h_2(\mathbf{x}) \tag{16a}$$

$$\text{s.t.} \quad (\mathbf{p}-\mu)^T\Sigma^{-1}(\mathbf{p}-\mu) \leq \rho_1 \tag{16b}$$

$$\mathbf{p}\mathbf{p}^T \preceq \boldsymbol{P} \preceq \rho_2\Sigma + \mathbf{p}^T\mu + \mu\mathbf{p}^T - \mu\mu^T \tag{16c}$$

where $\mathbf{p}$ and $\boldsymbol{P}$ are substitutes for the variables $\mathbb{E}_F[\xi]$ and $\mathbb{E}_F[\xi\xi^T]$ respectively
the first inequality of constraint (16c) results from the fact $\mathbb{E}_F[\xi]\mathbb{E}_F[\xi^T] \preceq \mathbb{E}_F[\xi\xi^T]$.

Sketch of Proof

- Since $h_2(\mathbf{x}) \succeq 0$, $\boldsymbol{P}^* = \rho_2\Sigma + \mathbf{p}^T\mu + \mu\mathbf{p}^T - \mu\mu^T$ is a valid optimal solution since it maximizes the objective.

Sketch of Proof

- Since $h_2(\mathbf{x}) \succeq 0$, $\boldsymbol{P}^* = \rho_2\Sigma + \mathbf{p}^T\mu + \mu\mathbf{p}^T - \mu\mu^T$ is a valid optimal solution since it maximizes the objective.
- Then after replacing $\mathbf{q} = \Sigma^{\frac{-1}{2}}(\mathbf{p} - \mu)$, problem (16) is simplified to

$$\max_{\mathbf{q}} \quad h_0(\mathbf{x}) + h_1(\mathbf{x})^T\mu + \mu^T h_2(\mathbf{x})\mu + \rho_2\Sigma \bullet h_2(\mathbf{x}) + [h_1(\mathbf{x})^T + 2\mu^T h_2(\mathbf{x})]\Sigma^{\frac{1}{2}}\mathbf{q} \tag{17a}$$

$$\text{s.t.} \quad \mathbf{q}^T\mathbf{q} \leq \rho_1 \tag{17b}$$

Sketch of Proof

- By Cauchy-Schwarz inequality

$$\max_{\mathbf{q}^T\mathbf{q}\leq\rho_1} [h_1(\mathbf{x})^T + 2\mu^T h_2(\mathbf{x})]\Sigma^{\frac{1}{2}}\mathbf{q} \leq \sqrt{\rho_1}||\Sigma^{\frac{1}{2}}[h_1(\mathbf{x}) + 2h_2(\mathbf{x})\mu]||$$

The equality is satisfied when $\mathbf{q}^* = \sqrt{\rho_1}\frac{\Sigma^{\frac{1}{2}}[h_1(\mathbf{x})+2h_2(\mathbf{x})\mu]}{||\Sigma^{\frac{1}{2}}[h_1(\mathbf{x})+2h_2(\mathbf{x})\mu]||}$.

Sketch of Proof

- By Cauchy-Schwarz inequality

$$\max_{\mathbf{q}^T\mathbf{q}\leq\rho_1} [h_1(\mathbf{x})^T + 2\mu^T h_2(\mathbf{x})]\Sigma^{\frac{1}{2}}\mathbf{q} \leq \sqrt{\rho_1}||\Sigma^{\frac{1}{2}}[h_1(\mathbf{x}) + 2h_2(\mathbf{x})\mu]||$$

The equality is satisfied when $\mathbf{q}^* = \sqrt{\rho_1}\frac{\Sigma^{\frac{1}{2}}[h_1(\mathbf{x})+2h_2(\mathbf{x})\mu]}{||\Sigma^{\frac{1}{2}}[h_1(\mathbf{x})+2h_2(\mathbf{x})\mu]||}$.

- The surrogate problem is equivalent to

$$\begin{aligned}\min_{\mathbf{x},h_0(\mathbf{x}),h_1(\mathbf{x}),h_2(\mathbf{x})} \quad & h_0(\mathbf{x}) + h_1(\mathbf{x})^T\mu + (\mu\mu^T + \rho_2\Sigma) \bullet h_2(\mathbf{x}) \\ & + \sqrt{\rho_1}||\Sigma^{\frac{1}{2}}[h_1(\mathbf{x}) + 2h_2(\mathbf{x})\mu]|| \\ \text{s.t.} \quad & Q(\mathbf{x},\xi) \leq h_0(\mathbf{x}) + h_1(\mathbf{x})^T\xi + \xi^T h_2(x)\xi, \quad \forall\xi \in \mathcal{S} \\ & h_2(x) \succeq 0, h_0(\mathbf{x}) \in \mathbb{R}, h_1(\mathbf{x}) \in \mathbb{R}^m, \mathbf{x} \in \mathbf{X}\end{aligned}$$

which is the same as problem (5) in [1].

DRO with higher-order moments

Suppose we know the higher-oder moment information, i.e., $\mathcal{M}_{\boldsymbol{a}} = \mathbb{E}[\xi^{\boldsymbol{a}}] = \mathbb{E}[\xi_1^{a_1}\xi_2^{a_2}...\xi_m^{a_m}]$, $\boldsymbol{a} \in \mathcal{A} \subset N^m$. The corresponding polynomial surrogate of problem (10) is

$$\min_{\mathbf{x}\in\mathbf{X},\boldsymbol{h}} \sum_{\boldsymbol{a}\in\mathcal{A}} h_{\boldsymbol{a}}\mathcal{M}_{\boldsymbol{a}} \tag{19a}$$

$$\text{s.t.} \quad Q(\mathbf{x},\xi) \leq \sum_{\boldsymbol{a}\in\mathcal{A}} h_{\boldsymbol{a}}\xi^{\boldsymbol{a}}, \quad \forall \xi \in \mathcal{S} \tag{19b}$$

which is also the dual problem of the DRO with the moment information ([**?**]).

Summary

Conclusions and Future work

- We present a surrogate-based approximation for the general distributionally robust optimization(DRO) problem.
- We show that the surrogate-based approximation provided an upper bound of the DRO minimization problem.
- We prove that when the surrogate is polynomial, the surrogate-based approximation becomes exact for the DRO problem with support, first and second moments information.
- We provide a different angel to approximate DRO problem not from the primal-and-dual perspective ([2, 4]).

Reference I

[1] E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations research*, 58(3):595–612, 2010.

[2] K. Isii. On sharpness of tchebycheff-type inequalities. *Annals of the Institute of Statistical Mathematics*, 14(1):185–197, 1962.

[3] J. B. Lasserre. *Moments, positive polynomials and their applications*, volume 1. World Scientific, 2009.

[4] A. Shapiro. On duality theory of conic linear problems. In *Semi-infinite programming*, pages 135–165. Springer, 2001.

Thank you for your attention!