Quantitative stability and Monte Carlo approximations of PDE constrained optimization problems under uncertainty

M. Hoffhues[+], W. Römisch, T. M. Surowiec[+]

Humboldt-University Berlin
Institute of Mathematics
www.math.hu-berlin.de/~romisch

[+]Philipps-University Marburg
FB Mathematik und Informatik

Workshop *PDE Constrained Optimization Under Uncertainty and Mean Field Games*
WIAS Berlin, January 28–30, 2020

Introduction

(Quantitative) stability analysis for stochastic optimization problems is developed for finite-dimensional spaces so far. It may serve as theoretical justification for approximation schemes.

The latter require a combination of discretization and sampling techniques and specific solution methods.

Nowadays, infinite-dimensional optimization problems under uncertainty motivated by economic and engineering applications attracted more interest.

Partial differential equations (PDEs) with random coefficients are within the reach of efficient computational methods.

PDE constrained optimization under uncertainty

Let $D \subset \mathbb{R}^m$ be an open bounded domain with Lipschitz boundary, $V = H_0^1(D)$ the classical Sobolev space with inner product $(\cdot,\cdot)_V$, $V^\star = H^{-1}(D)$ its dual with norm $\|\cdot\|_\star$ und dual pairing $\langle\cdot,\cdot\rangle$ and $H = L^2(D)$ with inner product $(\cdot,\cdot)_H$. Let Ξ be a metric space and $\mathbb{P}$ be a Borel probability measure on Ξ.

We consider the bilinear form $a(\cdot,\cdot;\xi) : V \times V \to \mathbb{R}$ defined by

$$a(u,v;\xi) = \int_D \sum_{i,j=1}^n b_{ij}(x,\xi) \frac{\partial u(x)}{\partial x_i} \frac{\partial v(x)}{\partial x_j} \mathrm{d}x \quad (\xi \in \Xi).$$

We impose the condition that the functions $b_{ij} : D \times \Xi \to \mathbb{R}$ are measurable on $D \times \Xi$ and there exist $L > \gamma > 0$ such that

$$\gamma \sum_{i=1}^n y_i^2 \leq \sum_{i,j=1}^n b_{ij}(x,\xi) y_i y_j \leq L \sum_{i=1}^n y_i^2 \quad (\forall y \in \mathbb{R}^n)$$

for a.e. $x \in D$ and $\mathbb{P}$-a.e. $\xi \in \Xi$. This implies that each b_{ij} is essentially bounded on $D \times \Omega$ from both sides with respect to the associated product measure.

We consider the optimization problem: Minimize the functional

$$\begin{aligned}\mathcal{J}(u,z) &:= \frac{1}{2}\int_{\Xi}\int_{D} |u(x,\xi)-\widetilde{u}_d(x)|^2 \,\mathrm{d}x\mathrm{d}\mathbb{P}(\xi) + \frac{\alpha}{2}\int_{D} |z(x)|^2 \,\mathrm{d}x \\ &= \frac{1}{2}\mathbb{E}_{\mathbb{P}}[\|u-\widetilde{u}_d\|_H^2] + \frac{\alpha}{2}\|z\|_H^2\end{aligned}$$

subject to $z \in Z_{\mathrm{ad}}$ with $Z_{\mathrm{ad}} \subset H$ denoting a closed convex bounded set and u solving the random elliptic PDE

$$a(u,v;\xi) = \int_D (z(x)+g(x,\xi))v(x)dx \quad \text{for } \mathbb{P}\text{-a.e. } \xi \in \Xi$$

and all test functions $v \in C_0^\infty(D)$, where $\alpha > 0$, $\widetilde{u}_d \in H$ and $g : D \times \Xi \to \mathbb{R}$ is measurable on $D \times \Xi$ and at least square integrable on D.

For each $\xi \in \Xi$ we define the mapping $A(\xi) : V \to V^\star$ by means of the Riesz representation theorem

$$\langle A(\xi)u, v\rangle = a(u,v;\xi) \quad (u,v \in V).$$

Then $A(\xi)$ is linear, uniformly positive definite (with $\gamma > 0$) and uniformly bounded (with $L > 0$) and the random PDE may be written in the form

$$A(\xi)u = z + g(\xi) \quad (\mathbb{P}\text{-a.e. } \xi \in \Xi).$$

Let J the duality mapping $J : V \to V^\star$ given by

$$\langle Ju, v \rangle = (u, v)_V \quad (u, v \in V).$$

For any $b \in V^\star$ and $t > 0$ we consider the mapping

$$K_t(\xi)u = u - tJ^{-1}(A(\xi)u - b) \quad (v \in V).$$

Then $K_t(\xi)$ is a contraction of V with constant

$$0 < \kappa(t) = \sqrt{1 - 2\gamma t + L^2 t^2} < 1 \quad \text{iff} \quad t \in \Big(0, \frac{2\gamma}{L^2}\Big).$$

The unique fixed point of $K_t(\xi)$ is the unique solution of $A(\xi)u = b$ and belongs to the closed ball $\mathbb{B}\big(0, \frac{t}{1-\kappa(t)}\|b\|_\star\big)$ in V.

Hence, the inverse mapping $A(\xi)^{-1} : V^\star \to V$ exists and is linear, uniformly positive definite (with $\frac{1}{L}$) and uniformly bounded (with $\frac{1}{\gamma}$).

Existence and quadratic growth

Abstract optimization problem: We consider the integrand

$$\begin{aligned} f(z,\xi) &= \frac{1}{2}\left\|A(\xi)^{-1}(z+g(\xi)) - \widetilde{u}_d\right\|_H^2 + \frac{\alpha}{2}\|z\|_H^2 \\ &= \frac{1}{2}\left\|A(\xi)^{-1}z - (\widetilde{u}_d - A(\xi)^{-1}g(\xi))\right\|_H^2 + \frac{\alpha}{2}\|z\|_H^2 \end{aligned}$$

for any $z \in Z_{\mathrm{ad}}$ and $\xi \in \Xi$, and the infinite-dimensional stochastic optimization problem

$$\min\left\{F(z) = \int_\Xi f(z,\xi)\mathrm{d}\mathbb{P}(\xi) : z \in Z_{\mathrm{ad}}\right\}, \tag{1}$$

where $g \in L_2(\Xi,\mathbb{P};H)$ and $A(\xi)^{-1}$ as defined earlier.

Proposition 1:

The functional F is finite, strongly convex and lower semicontinuous, hence, weakly lower semicontinuous on the weakly compact set Z_{ad}. Hence, there exists a unique minimizer $z(\mathbb{P}) \in Z_{\mathrm{ad}}$ of (1) and it holds

$$\|z - z(\mathbb{P})\|^2 \leq \frac{8}{\alpha}(F(z) - F(z(\mathbb{P}))) \quad (\forall z \in Z_{\mathrm{ad}}).$$

Quantitative stability

Weak convergence in $\mathcal{P}(\Xi)$: $(\mathbb{P}_N)$ converges weakly to $\mathbb{P}$ iff

$$\lim_{N\to\infty} \int_\Xi f(\xi) \mathrm{d}\mathbb{P}_N(\xi) = \int_\Xi f(\xi) \mathrm{d}\mathbb{P}(\xi) \qquad (\forall f \in C_b(\Xi, \mathbb{R})).$$

The topology of weak convergence is metrizable if Ξ is separable.

Distance on $\mathcal{P}(\Xi)$: (Zolotarev 83)

$$d_{\mathfrak{F}}(\mathbb{P}, \mathbb{Q}) = \sup_{f \in \mathfrak{F}} \left| \int_\Xi f(\xi) \mathrm{d}\mathbb{P}(\xi) - \int_\Xi f(\xi) \mathrm{d}\mathbb{Q}(\xi) \right|,$$

where $\mathfrak{F}$ is a family of real-valued Borel measurable functions on Ξ.

A number of important probability metrics are of the form $d_{\mathfrak{F}}$, for example, the bounded Lipschitz metric (metrizing the topology of weak convergence) and the Fortet-Mourier type metrics.

Whether convergence with respect to $d_{\mathfrak{F}}$ implies or is implied by weak convergence depends on the richness and on analytical properties of $\mathfrak{F}$.

Lemma: (Topsøe 67)
Let $\mathfrak{F}$ be uniformly bounded and $\mathbb{P}(\{\xi \in \Xi : \mathfrak{F}$ is not equicontinuous at $\xi\}) = 0$ holds. Then $\mathfrak{F}$ is a $\mathbb{P}$-uniformity class, i.e., weak convergence of $(\mathbb{P}_N)$ to $\mathbb{P}$ implies

$$\lim_{N\to\infty} d_{\mathfrak{F}}(\mathbb{P}_N, \mathbb{P}) = 0.$$

Compared with classical probability metrics we consider a much smaller family $\mathfrak{F}$ of functions, namely,

$$\mathfrak{F} = \{f(z, \cdot) : z \in Z_{\mathrm{ad}}\}.$$

In this case we arrive at a semi-metric which we call problem-based or minimal information (m.i.) distance and $\mathfrak{F}$ the m.i. family, respectively.

Theorem 1:
Under the standing assumptions and with $\mathfrak{F} = \{f(z, \cdot) : z \in Z_{\mathrm{ad}}\}$ we obtain the following estimates for the optimal values $v(\mathbb{P})$ and solutions $z(\mathbb{P})$ of (1):

$$\begin{aligned} |v(\mathbb{Q}) - v(\mathbb{P})| &\le d_{\mathfrak{F}}(\mathbb{Q}, \mathbb{P}) \\ \|z(\mathbb{Q}) - z(\mathbb{P})\|_H &\le 2\sqrt{\frac{2}{\alpha}} d_{\mathfrak{F}}(\mathbb{Q}, \mathbb{P})^{\frac{1}{2}} \end{aligned}$$

for any $\mathbb{Q} \in \mathcal{P}(\Xi)$.

Properties of the integrands

Theorem 2:
Assume that all functions $b_{ij}(x,\cdot)$, $i,j=1,\dots,n$, and $g(x,\cdot)$ are Lipschitz continuous on Ξ uniformly with respect to $x\in D$. Furthermore, assume that $g\in L_\infty(\Xi,\mathbb{P};V^\star)$.
Then the m.i. family $\mathfrak{F}=\{f(z,\cdot):z\in Z_{\text{ad}}\}$ is uniformly bounded and equi-Lipschitz continuous on Ξ. In particular, $\mathfrak{F}$ is a $\mathbb{P}$-uniformity class.
Moreover, the family $\{f(\cdot,\xi):\xi\in\Xi\}$ is Lipschitz continuous on each bounded subset of H (with a constant not depending on ξ).

Remark:
Here, we consider more general PDE models under weaker assumptions than, for example, in Cohen-Devore-Schwab 11 and subsequent work on computational random PDEs in which regularity properties of solutions with respect to parameters play a central role.

Monte Carlo approximations

Let $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ be independent identically distributed Ξ-valued random variables on some probability space $(\Omega, \mathcal{F}, P)$ having the common distribution $\mathbb{P}$, i.e., $\mathbb{P} = P\xi_1^{-1}$. We consider the empirical measures

$$\mathbb{P}_n(\cdot) = \frac{1}{n}\sum_{i=1}^{n} \delta_{\xi_i(\cdot)} \qquad (n \in \mathbb{N})$$

and the empirical or Monte Carlo approximation of the stochastic program (1) with sample size n, i.e.,

$$\min\left\{\frac{1}{n}\sum_{i=1}^{n} f(z, \xi_i(\cdot)) : z \in Z_{\mathrm{ad}}\right\}. \tag{2}$$

The optimal value $v(\mathbb{P}_n(\cdot))$ of (2) is a real random variable and the solution $z(\mathbb{P}_n(\cdot))$ an H-valued random element.

Qualitative and quantitative results on the asymptotic behavior of optimal values and solutions to (2) are known in finite-dimensional settings so far (see Dupačová-Wets 88, and the surveys by Shapiro 03 and Pflug 03).

It is known that $(\mathbb{P}_n(\cdot))$ converges weakly to $\mathbb{P}$ P-almost surely.

Corollary:
The sequences $(v(\mathbb{P}_n(\cdot)))$ and $(z(\mathbb{P}_n(\cdot)))$ of empirical optimal values and solutions converge P-almost surely to the true optimal values and solutions $v(\mathbb{P})$ and $z(\mathbb{P})$, respectively.

Quantitative information on the asymptotic behavior of $v(P_n(\cdot))$ and $z(P_n(\cdot))$ is closely related to uniform convergence properties of the empirical process

$$\left\{\sqrt{n}(\mathbb{P}_n(\cdot) - \mathbb{P})f = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (f(\xi_i(\cdot)) - \mathbb{P}f)\right\}_{f \in \mathfrak{F}}$$

indexed by $\mathfrak{F} = \{f(z, \cdot) : z \in Z\}$ and, hence, to quantitative estimates of

$$\sqrt{n}\, d_{\mathfrak{F}}(\mathbb{P}_n(\cdot), \mathbb{P}) = \sqrt{n} \sup_{f \in \mathfrak{F}} |\mathbb{P}_n(\cdot)f - \mathbb{P}f|. \tag{3}$$

Here, we set $\mathbb{P}f = \int_\Xi f(\xi) \mathrm{d}\mathbb{P}(\xi)$ for any probability distribution $\mathbb{P}$ and any $f \in \mathfrak{F}$. Since the supremum in (3) is non-measurable in general, the outer probability $P^\star(A) = \inf\{P(B) : A \subseteq B, B \in \mathcal{F}\}$ is used in the following.

The empirical process is called uniformly bounded in outer probability with tail $C_{\mathfrak{F}}(\cdot)$ if the function $C_{\mathfrak{F}}(\cdot)$ is defined on $(0,\infty)$, decreasing to 0, and the estimate

$$P^{\star}(\{\omega\in\Omega:\sqrt{n}\,d_{\mathfrak{F}}(\mathbb{P}_n(\omega),\mathbb{P})\geq\varepsilon\})\leq C_{\mathfrak{F}}(\varepsilon)$$

holds for all $\varepsilon>0$ and $n\in\mathbb{N}$.

Whether such a property is satisfied depends on the size of the class $\mathfrak{F}$ measured in terms of so-called bracketing numbers. To introduce the latter concept, let $\mathfrak{F}$ be a subset of the linear normed space $L_p(\Xi,\mathbb{P})$ (for some $p\geq 1$) equipped with the usual norm

$$\|f\|_{\mathbb{P},p}=(\mathbb{P}|f|^p)^{\frac{1}{p}}=\Big(\int_{\Xi}|f(\xi)|^p\mathrm{d}\mathbb{P}(\xi)\Big)^{\frac{1}{p}}.$$

The bracketing number $N_{[\,]}(\varepsilon,\mathfrak{F},\|\cdot\|_{\mathbb{P},p})$ is the minimal number of brackets $[l,u]=\{f\in L_p(\Xi,\mathbb{P}):l\leq f\leq u\}$ with $l,u\in L_p(\Xi,\mathbb{P})$ and $\|l-u\|_{\mathbb{P},p}<\varepsilon$ needed to cover $\mathfrak{F}$.

The following results provide criteria for the uniform boundedness of the empirical process.

Lemma: (Talagrand 94)
Let $\mathfrak{F}$ be a class of real-valued measurable functions on Ξ. If $\mathfrak{F}$ is uniformly bounded and there exist constants $r \geq 1$ and $R \geq 1$ such that

$$N_{[\,]}(\varepsilon, \mathfrak{F}, \|\cdot\|_{\mathbb{P},2}) \leq \left(\frac{R}{\varepsilon}\right)^r$$

holds for every $\varepsilon > 0$, then the empirical process indexed by $\mathfrak{F}$ is uniformly bounded in outer probability with exponential tail, i.e.,

$$P^\star(\{\omega \in \Omega : \sqrt{n}\, d_{\mathfrak{F}}(\mathbb{P}_n(\omega), \mathbb{P}) \geq \varepsilon\}) \leq (K(R) r^{-\frac{1}{2}}\varepsilon)^r \exp(-2\varepsilon^2)$$

with some constant $K(R)$ depending only on R.

Lemma: (van der Vaart–Wellner 96)
Let Z denote a subset of a linear normed space with norm $\|\cdot\|$ and $\mathfrak{F} = \{f(z,\cdot) : z \in Z\}$ be a subset of $L_p(\Omega, \mathbb{P})$ having the property

$$|f(z,\xi) - f(z',\xi)| \leq \|z - z'\|\Phi(\xi) \quad (\forall z, z' \in Z; \xi \in \Xi),$$

where Φ belongs to $L_p(\Omega, \mathbb{P})$. Then it holds

$$N_{[\,]}(2\varepsilon\|\Phi\|_{\mathbb{P},p}, \mathfrak{F}, \|\cdot\|_{\mathbb{P},p}) \leq N(\varepsilon, Z, \|\cdot\|),$$

where the covering number $N(\varepsilon, Z, \|\cdot\|)$ denotes the minimal number of balls with respect to the norm $\|\cdot\|$ and radius ε needed to cover Z.

If Z is a bounded subset of a k-dimensional space, there exists $K > 0$ such that

$$N(\varepsilon, Z, \|\cdot\|) \leq K\varepsilon^{-k}.$$

Since Z_{ad} belongs to the infinite-dimensional space H, an intermediate step is needed to apply the second lemma.

Let Z_k, $k \in \mathbb{N}$, denote a sequence of piecewise constant subspaces of $H = L_2(D)$ such that $Z_{\mathrm{ad}}^{(k)} = Z_k \cap Z_{\mathrm{ad}}$ has the property

$$d(z, Z_{\mathrm{ad}}^{(k)}) = O(h_k) \quad \text{for any } z \in Z_{\mathrm{ad}},$$

where $h_k \to 0$ is the diameter of the cells of D.

Proposition:

Under the standing assumptions, there exists a constant $C > 0$ such that

$$|v(\mathbb{P}) - v(\mathbb{P}_n(\cdot))| \leq C(d(z(\mathbb{P}), Z_{\mathrm{ad}}^{(k)}) + d_{\mathfrak{F}_k}(\mathbb{P}_n(\cdot), \mathbb{P})),$$

where $\mathfrak{F}_k = \{f(z,\cdot) : z \in Z_{\mathrm{ad}}^{(k)}\}$. Let $k(n)$ be a sequence such that $\sqrt{n}h_{k(n)} \to 0$. Then for all $\varepsilon > 0$ and $n \in \mathbb{N}$

$$P^\star(\{\omega : \sqrt{n}|v(\mathbb{P}) - v(\mathbb{P}_n(\omega)| \geq \varepsilon\} \leq (\hat{C}\varepsilon)^{k(n)} \exp(-2\varepsilon^2).$$

Conclusions

- Quantitative stability results can be used to justify scenario reduction heuristics.
- Monte Carlo methods have a very slow convergence rate and require a very large sample size and, thus, a huge number of PDE solves.
- Randomized Quasi-Monte Carlo methods could be a viable alternative due to better convergence rates and the possibility of effective dimension reduction. However, their justification requires a completely different methodology.

References

H. Antil and D. Leykekhman: A brief introduction to PDE-constrained optimization, *Frontiers in PDE-Constrained Optimization* (H. Antil et al. eds.), IMA Volumes in Mathematics and its Applications, Springer, 2018, 3–40.

A. Cohen, R. Devore and C. Schwab: Analytic regularity and polynomial approximation of parametric and stochastic PDEs, *Analysis and Applications* 9 (2011), 11–47.

J. Dupačová and R. J-B Wets: Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems, *The Annals of Statistics* 16 (1988), 1517–1549.

H. Gajewski, K. Gröger and K. Zacharias: *Nichtlineare Operatorgleichungen und Operatordifferentialgleichungen*, Akademie-Verlag, Berlin, 1974.

M. Hoffhues, W. Römisch and T. M. Surowiec: On quantitative stability in infinite-dimensional optimization under uncertainty, preprint, 2020.

S. T. Rachev and W. Römisch: Quantitative stability in stochastic programming: The method of probability metrics, *Mathematics of Operations Research* 27 (2002), 792–818.

A. Ruszczyński and A. Shapiro (eds.): *Stochastic Programming*, Handbooks in Operations Research and Management Science, Volume 10, Elsevier, Amsterdam 2003.

M. Talagrand: Sharper bounds for Gaussian and empirical processes, *Annals of Probability* 22 (1994), 28–76.

F. Topsøe: On the connection of P-continuity and P-uniformity in weak convergence, *Probability Theory and its Applications* 12 (1967), 281–290.

A. W. van der Vaart, J. A. Wellner: *Weak Convergence and Empirical Processes*, Springer, New York, 1996.

V. M. Zolotarev: Probability metrics, *Theory of Probability and its Applications* 28 (1967), 278–302.