Are (continuous) two-stage stochastic programs solvable?

W. Römisch

Humboldt-University Berlin
Institute of Mathematics

www.math.hu-berlin.de/~romisch

ICSP 2016, Buzios (Brazil), Round-table, June 29, 2016

Introduction

- It was recently proved that even the approximate solution of linear two-stage stochastic programs with fixed recourse for a sufficiently high accuracy is NP-hard (Hanasusanto-Kuhn-Wiesemann 15).
- Computational methods for solving stochastic programs require a discretization of the underlying probability distribution induced by a numerical integration scheme for computing expectations.
- Standard approach: Variants of Monte Carlo (MC) methods. However, MC methods are extremely slow and may require enormous sample sizes.
- On the other hand, it is known that numerical integration is strongly polynomially tractable for integrands belonging to weighted tensor product mixed Sobolev spaces if the weights satisfy certain condition (Sloan-Woźniakowski 98).
- Moreover, the optimal order of convergence of numerical integration in such spaces can essentially be achieved by certain randomized Quasi-Monte Carlo methods (Sloan-Kuo-Joe 02, Kuo 03).
- Typical integrands in two-stage stochastic programming can be approximated by functions from mixed Sobolev spaces if their effective dimension is low.

Complexity of two-stage stochastic programs

The two papers Dyer-Stougie 06, Hanasusanto-Kuhn-Wiesemann 15 consider the following second-stage optimal value function

$$Q(\xi; \alpha, \beta) = \max \left\{ \xi^\top y - \beta z : y \leq \alpha z,\ y \in \mathbb{R}^d_+,\ z \in [0,1] \right\} = \max\{\xi^\top \alpha - \beta, 0\},$$

where $\alpha \in \mathbb{R}^d_+$ and $\beta \in \mathbb{R}_+$ are parameters and the random vector ξ is uniformly distributed in $[0,1]^d$. Then the expected recourse function is of the form

$$\mathcal{Q}(\alpha, \beta) = \mathbb{E}[Q(\xi; \alpha, \beta)] = \frac{1}{2}\alpha^\top e - \beta + \int_0^\beta \mathrm{Vol}\, P(\alpha, t)\, dt,$$

where $P(\alpha, \beta) = \{z \in [0,1]^d : \alpha^\top z \leq \beta\}$ is the knapsack polytope and $e = (1, \ldots, 1)^\top \in \mathbb{R}^d$.

Theorem: (Hanasusanto-Kuhn-Wiesemann 15)
For any pair $(\alpha, \beta) \in \mathbb{R}^{d+1}_+$ there exists $\varepsilon(d, \alpha)$ such that the computation of $\mathcal{Q}(\alpha, \beta)$ within an absolute accuracy of $\varepsilon < \varepsilon(d, \alpha)$ is NP-hard.

Note that for any $\alpha \in \mathbb{R}^d \setminus \{0\}$ the constant $\varepsilon(d, \alpha)$ tends to 0 exponentially with respect to the dimension d.

Note also that the function

$$f(\xi) = \max\{\xi^\top \alpha - \beta, 0\} \quad (\xi \in [0,1]^d)$$

is not of bounded variation in the sense of Hardy and Krause if $d > 2$ (Owen 05) and does not belong to mixed Sobolev spaces on $[0,1]^d$.

But, both properties are particularly relevant for the application of Quasi-Monte Carlo methods for numerical integration.

For general linear two-stage stochastic programs, the second-stage optimal value function is of the form $\Phi(q(\xi), h(\xi) - T(\xi)x)$, where

$$\Phi(u,t) = \inf\{u^\top y : Wy = t,\, y \in Y\} \quad ((u,t) \in \mathbb{R}^{\overline{m}} \times \mathbb{R}^s),$$

with first-stage decision $x \in X \subset \mathbb{R}^m$, $s \times \overline{m}$ recourse matrix W, polyhedral cone Y, $s \times m$-matrix $T(\cdot)$, $q(\cdot) \in \mathbb{R}^{\overline{m}}$, $h(\cdot) \in \mathbb{R}^s$ being affine functions and d-dimensional random vector ξ with support Ξ. If the condition **(A1)** relatively complete recourse and dual feasibility is satisfied, the second-stage optimal value function is continuous and piecewise linear-quadratic on Ξ and it holds

$$\Phi(q(\xi), h(\xi) - T(\xi)x) = \max_{j=1,\dots,\ell} (C_j q(\xi))^\top (h(\xi) - T(\xi)x) \quad ((x,\xi) \in X \times \Xi).$$

Complexity of numerical integration

We consider the approximate computation of

$$I_d(f) = \int_{[0,1]^d} f(\xi) d\xi$$

by a linear numerical integration or quadrature method of the form

$$Q_n(f) = \sum_{i=1}^{n} w_i f(\xi^i)$$

with points $\xi^i \in [0,1]^d$ and weights $w_i \in \mathbb{R}$, $i = 1, \ldots, n$.
We assume that f belongs to a linear normed space $\mathbb{F}_d$ of functions on $[0,1]^d$ with norm $\|\cdot\|_d$ and unit ball $\mathbb{B}_d = \{f \in \mathbb{F}_d : \|f\|_d \leq 1\}$ such that I_d and Q_n are linear bounded functionals on $\mathbb{F}_d$.

Worst-case error of Q_n over $\mathbb{B}_d$ and optimal error are given by:

$$\begin{aligned} e(Q_n) &= \sup_{f \in \mathbb{B}_d} |I_d(f) - Q_n(f)| \\ e(n, \mathbb{B}_d) &= \inf_{Q_n} e(Q_n) \,. \end{aligned}$$

(Novak 14)

It is known that due to the convexity and symmetry of $\mathbb{B}_d$ linear algorithms are optimal among nonlinear and adaptive ones (Bakhvalov 71, Novak 88).

The information complexity $n(\varepsilon, \mathbb{B}_d)$ is the minimal number of function values which is needed that the worst-case error is at most ε, i.e.,

$$n(\varepsilon, \mathbb{B}_d) = \min\{n : \exists Q_n \text{ such that } e(Q_n) \leq \varepsilon\}$$

Of course, the behavior of $n(\varepsilon, \mathbb{B}_d)$ as function of (ε, d) depends heavily on $\mathbb{F}_d$.

Numerical integration is said to
be polynomially tractable if there exist constants $C > 0$ $q \geq 0$, $p > 0$ such that

$$n(\varepsilon, \mathbb{B}_d) \leq C d^q \varepsilon^{-p},$$

be strongly polynomially tractable if there exist constants $C > 0$, $p > 0$ such that

$$n(\varepsilon, \mathbb{B}_d) \leq C \varepsilon^{-p},$$

have the curse of dimension if there exist $c > 0$, $\varepsilon_0 > 0$ and $\gamma > 0$ such that

$$n(\varepsilon, \mathbb{B}_d) \geq c(1+\gamma)^d \text{ for all } \varepsilon \leq \varepsilon_0 \text{ and for infinitely many } d \in \mathbb{N}.$$

Randomized algorithms:

A randomized quadrature algorithm is denoted by $(Q(\omega))_{\omega\in\Omega}$ and considered on a probabability space $(\Omega, \mathcal{F}, \mathbb{P})$.. We assume that $Q(\omega)$ is a quadrature algorithm for each ω and that it depends on ω in a measurable way. Let $n(f,\omega)$ denote the number of evaluations of $f \in \mathbb{F}_d$ needed to perform $Q(\omega)f$. The number

$$n(Q) = \sup_{f\in\mathbb{B}_d} \int_\Omega n(f,\omega)\mathbb{P}(d\omega)$$

is called the cardinality of the randomized algorithm Q and

$$e^{\mathrm{ran}}(Q) = \sup_{f\in\mathbb{B}_d} \left(\int_\Omega \|I_d f - Q(\omega)f\|^2 \, \mathbb{P}(d\omega)\right)^{\frac{1}{2}}$$

the error of Q. The minimal error of randomized algorithms is

$$e^{\mathrm{ran}}(n, \mathbb{B}_d) = \inf\{e^{\mathrm{ran}}(Q) : n(Q) \le n\}\,.$$

By construction it is clear that $e^{\mathrm{ran}}(n, \mathbb{B}_d) \le e(n, \mathbb{B}_d)$ holds.

Standard Monte Carlo (MC) method Q based on n i.i.d. samples: (Mathé 95)

$$e^{\mathrm{ran}}(Q) = (1+\sqrt{n})^{-1} \le n^{-\frac{1}{2}}$$

if $\mathbb{B}_d$ is the unit ball of $\mathbb{F}_d = L_p([0,1]^d)$ for $2 \le p < \infty$.

Example:

Consider the Banach space $\mathbb{F}_d = C^r([0,1]^d)$ ($r \in \mathbb{N}$) with the norm

$$\|f\|_{r,d} = \max_{|\alpha| \le r} \|D^\alpha f\|_\infty,$$

where $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}_0^d$ and $D^\alpha f$ denotes the mixed partial derivative of order $|\alpha| = \sum_{i=1}^d \alpha_i$, i.e.,

$$D^\alpha f(\xi) = \frac{\partial^{|\alpha|} f}{\partial \xi_1^{\alpha_1} \cdots \partial \xi_d^{\alpha_d}}(\xi)\,.$$

It is long known (Bakhvalov 59) that there exist constants $C_{r,d}$, $c_{r,d} > 0$ such that

$$c_{r,d}\, n^{-\frac{r}{d}} \le e(n, \mathbb{B}_d) \le C_{r,d}\, n^{-\frac{r}{d}}.$$

But, surprisingly it was shown only recently that the numerical integration on $C^r([0,1]^d)$ suffers from the curse of dimension (Hinrichs-Novak-Ullrich-Woźniakowski 14).

For the tensor product mixed Sobolev space

$$W_{2,\mathrm{mix}}^{(r,\ldots,r)}([0,1]^d) = \{f : [0,1]^d \to \mathbb{R} : D^\alpha f \in L_2([0,1]^d) \text{ if } \|\alpha\|_\infty \le r\}$$

it is known that $e(n, \mathbb{B}_d) = O(n^{-r} (\log n)^{\frac{(d-1)}{2}})$ (Frolov 76, Bykovskii 85).

We consider the linear space $W^1_{2,\gamma}([0,1])$ of all absolutely continuous functions on $[0,1]$ with derivatives belonging to $L_2([0,1])$ and the weighted inner product

$$\langle f, g\rangle_\gamma = \int_0^1 f(x)dx \int_0^1 g(x)dx + \frac{1}{\gamma}\int_0^1 f'(x)g'(x)dx.$$

Then the weighted tensor product mixed Sobolev space

$$W^{(1,\ldots,1)}_{2,\gamma,\mathrm{mix}}([0,1]^d) = \bigotimes_{j=1}^{d} W^1_{2,\gamma_j}([0,1])$$

is equipped with the inner product

$$\langle g, \tilde{g}\rangle_\gamma = \sum_{u\subseteq\mathfrak{D}} \gamma_u^{-1} \int_{[0,1]^{|u|}} \Big(\int_{[0,1]^{d-|u|}} \frac{\partial^{|u|}}{\partial t^u} g(t)dt^{-u}\Big)\Big(\int_{[0,1]^{d-|u|}} \frac{\partial^{|u|}}{\partial t^u}\tilde{g}(t)dt^{-u}\Big)dt^u\,,$$

where $\mathfrak{D} = \{1,\ldots,d\}$, the weights γ_i are positive and γ_u is given in product form $\gamma_u = \prod_{i\in u}\gamma_i$ for $u \subseteq \mathfrak{D}$, where $\gamma_\emptyset = 1$. For $u \subseteq \mathfrak{D}$ we use the notation $|u|$ for its cardinality, $-u$ for $\mathfrak{D}\setminus u$ and t^u for the $|u|$-dimensional vector with components t_j for $j \in u$.

Theorem: (Sloan-Woźniakowski 98, Sloan-Wang-Woźniakowski 04)
Numerical integration is strongly polynomially tractable on $W^{(1,\ldots,1)}_{2,\gamma,\mathrm{mix}}([0,1]^d)$ if

$$\sum_{j=1}^{\infty} \gamma_j < \infty\,.$$

Randomly shifted lattice rules

We consider the randomized Quasi-Monte Carlo method

$$Q_n(\omega)(f) = \frac{1}{n}\sum_{i=1}^{n} f\Big(\Big\{\frac{(i-1)}{n}g + \triangle(\omega)\Big\}\Big),$$

where $\triangle$ is a random vector with uniform distribution on $[0,1]^d$.

Theorem:
Let n be prime, $\mathbb{B}_d$ be the unit ball in $\mathcal{W}^{(1,\ldots,1)}_{2,\gamma,\mathrm{mix}}([0,1]^d)$. Then $g \in \mathbb{Z}^d$ can be constructed component-by-component such that for any $\delta \in (0,\frac{1}{2}]$ there exists a constant $C(\delta) > 0$ and the randomized minimal error allows the estimate

$$e^{\mathrm{ran}}(Q_n, \mathbb{B}_d) \le C(\delta)\, n^{-1+\delta},$$

where the constant $C(\delta)$ increases when δ decreases, but does not depend on the dimension d if the sequence (γ_j) satisfies the condition

$$\sum_{j=1}^{\infty} \gamma_j^{\frac{1}{2(1-\delta)}} < \infty \qquad (\text{e.g. } \gamma_j = \tfrac{1}{j^3}).$$

(Sloan-Kuo-Joe 02, Kuo 03, Nuyens-Cools 06)

ANOVA decomposition and effective dimension

We consider a multivariate function $f : \mathbb{R}^d \to \mathbb{R}$ and intend to compute the mean of $f(\xi)$, i.e.

$$\mathbb{E}[f(\xi)] = I_{d,\rho}(f) = \int_{\mathbb{R}^d} f(\xi_1, \ldots, \xi_d)\rho(\xi_1, \ldots, \xi_d) d\xi_1 \cdots d\xi_d \,,$$

where ξ is a d-dimensional random vector with density

$$\rho(\xi) = \prod_{k=1}^{d} \rho_k(\xi_k) \quad (\xi \in \mathbb{R}^d).$$

We are interested in a representation of f consisting of 2^d terms

$$f(\xi) = f_0 + \sum_{i=1}^{d} f_i(\xi_i) + \sum_{\substack{i,j=1 \\ i<j}}^{d} f_{ij}(\xi_i, \xi_j) + \cdots + f_{12\cdots d}(\xi_1, \ldots, \xi_d).$$

The previous representation can be more compactly written as

$$(*) \qquad f(\xi) = \sum_{u \subseteq \mathfrak{D}} f_u(\xi^u) \,,$$

where $\mathfrak{D} = \{1, \ldots, d\}$ and ξ^u contains only the components ξ_j with $j \in u$ and belongs to $\mathbb{R}^{|u|}$. Here, $|u|$ denotes the cardinality of u.

Next we make use of the space $L_{2,\rho}(\mathbb{R}^d)$ of all square integrable functions with inner product

$$\langle f, \tilde{f} \rangle_\rho = \int_{\mathbb{R}^d} f(\xi)\tilde{f}(\xi)\rho(\xi)d\xi \,.$$

A representation of the form $(*)$ of $f \in L_{2,\rho}(\mathbb{R}^d)$ is called ANOVA decomposition of f if

$$\int_{\mathbb{R}} f_u(\xi^u)\rho_k(\xi_k)d\xi_k = 0 \quad (\text{for all } k \in u \text{ and } u \subseteq \mathfrak{D}).$$

The ANOVA terms f_u, $\emptyset \neq u \subseteq \mathfrak{D}$, are orthogonal in $L_{2,\rho}(\mathbb{R}^d)$, i.e.

$$\langle f_u, f_v \rangle_\rho = \int_{\mathbb{R}^d} f_u(\xi) f_v(\xi)\rho(\xi)d\xi = 0 \quad \text{if and only if} \quad u \neq v,$$

The ANOVA terms f_u allow a representation in terms of (so-called) (ANOVA) projections, i.e.

$$(P_k f)(\xi) = \int_{-\infty}^{\infty} f(\xi_1, \ldots, \xi_{k-1}, s, \xi_{k+1}, \ldots, \xi_d)\rho_k(s)ds \; (\xi \in \mathbb{R}^d; k \in \mathfrak{D}).$$

and

$$P_u f = \Big(\prod_{k \in u} P_k\Big)(f) \quad (u \subseteq \mathfrak{D}).$$

Then it holds (Kuo-Sloan-Wasilkowski-Woźniakowski 10):

$$f_u = \Big(\prod_{j\in u}(I - P_j)\Big)P_{-u}(f) = P_{-u}(f) + \sum_{v\subsetneq u}(-1)^{|u|-|v|}P_{-v}(f)\,,$$

Consider the variances of f and f_u

$$\sigma^2(f) = \|f - I_{d,\rho}(f)\|_{2,\rho}^2 \quad \text{und} \quad \sigma_u^2(f) = \|f_u\|_{2,\rho}^2$$

and obtain

$$\sigma^2(f) = \|f\|_{L_2}^2 - (I_{d,\rho}(f))^2 = \sum_{\emptyset\neq u\subseteq\mathfrak{D}} \sigma_u^2(f)\,.$$

For small $\varepsilon \in (0,1)$ (e.g. $\varepsilon = 0.01$)

$$d_S(\varepsilon) = \min\Big\{s \in \mathfrak{D} : \sum_{|u|\leq s}\frac{\sigma_u^2(f)}{\sigma^2(f)} \geq 1-\varepsilon\Big\}$$

is called effective (superposition) dimension of f and it holds

$$(+) \qquad \Big\|f - \sum_{|u|\leq d_S(\varepsilon)} f_u\Big\|_{2,\rho} \leq \sqrt{\varepsilon}\sigma(f)\,,$$

i.e., the function f is approximated by a truncated ANOVA decomposition which contains all ANOVA terms f_u such that $|u| \leq d_S(\varepsilon)$. If f is nonsmooth and the ANOVA terms f_u, $|u| \leq d_S(\varepsilon)$, are smoother than f, the estimate $(+)$ means an approximate smoothing of f.

ANOVA decomposition of two-stage integrands

Assumptions: (A1) and

(A2) P has fourth order absolute moments.

(A3) P has a density of the form $\rho(\xi) = \prod_{i=1}^{d} \rho_i(\xi_i)$ ($\xi \in \mathbb{R}^d$) with continuous marginal densities ρ_i, $i \in \mathfrak{D}$.

(A4) For each $x \in X$ all common faces of the adjacent convex polyhedral sets

$$\Xi_j(x) = \{\xi \in \Xi : (q(\xi), h(\xi) - T(\xi)x) \in \mathcal{K}_j\} \quad (j = 1, \ldots, \ell)$$

do not parallel any coordinate axis, where the polyhedral cones $\mathcal{K}_j$, $j = 1, \ldots, \ell$, represent a partition of $\operatorname{dom} \Phi$ (geometric condition).

Theorem: Let $x \in X$, assume (A1)–(A4) and $f = f(x, \cdot)$ be the two-stage integrand. Then the second order truncated ANOVA decomposition of f

$$f^{(2)} := \sum_{|u| \leq 2} f_u \qquad \text{where} \qquad f = f^{(2)} + \sum_{|u|=3}^{d} f_u$$

belongs to $W_{2,\rho,\text{mix}}^{(1,\ldots,1)}(\mathbb{R}^d)$ if all marginal densities ρ_k, $k \in \mathfrak{D}$, belong to $C^1(\mathbb{R})$.

Remark: The second order truncated ANOVA decomposition $f^{(2)}$ is a good approximation of f if the effective superposition dimension $d_S(\varepsilon)$ is at most 2.

Conclusions

- The approximate solution of linear two-stage stochastic programs with fixed recourse for a sufficiently high accuracy is NP-hard.
- The numerical integration on weighted tensor product mixed Sobolev spaces on $[0,1]^d$ is strongly polynomially tractable if the weights satisfy a suitable condition.
- Randomly shifted lattice rules attain the optimal order of convergence on such spaces if the weights satisfy a slightly stronger condition. Hence, such methods are superior to Monte Carlo methods and reduce the sample sizes from n to almost $\sqrt{n}$.
- The second order ANOVA decomposition of two-stage integrands belongs to a mixed Sobolev space on $\mathbb{R}^d$ if the marginal densities are in C^1 and represent a good $L_{2,\rho}(\mathbb{R}^d)$ approximation if the effective superposition dimension $d_S(\varepsilon)$ of the integrands is at most two. It is conjectured that this result extends to higher effective dimensions.
- Unfortunately, mixed Sobolev spaces on $\mathbb{R}^d$ are in general not of tensor product type. Sufficient conditions have to be studied!

References:

M. Dyer and L. Stougie: Computational complexity of stochastic programming problems, *Mathematical Programming* 106 (2006), 423–432.

J. Dick, F. Y. Kuo and I. H. Sloan: High-dimensional integration – the Quasi-Monte Carlo way, *Acta Numerica* 22 (2013), 133–288.

J. Dick, F. Pillichshammer: *Digital Nets and Sequences*, Cambridge University Press, Cambridge 2010.

G. A. Hanasusanto, D. Kuhn and W. Wiesemann: A comment on "computational complexity of stochastic programming problems", *Mathematical Programming* (to appear).

A. Hinrichs, E. Novak, M. Ullrich and H. Woźniakowski: The curse of dimensionality for numerical integration of smooth functions, *Mathematics of Computation* 83 (2014), 2853–2863.

F. Y. Kuo: Component-by-component constructions achieve the optimal rate of convergence in weighted Korobov and Sobolev spaces, *Journal of Complexity* 19 (2003), 301–320.

F. Y. Kuo, I. H. Sloan, G. W. Wasilkowski, H. Woźniakowski: On decomposition of multivariate functions, *Mathematics of Computation* 79 (2010), 953–966.

H. Leövey and W. Römisch: Quasi-Monte Carlo methods for linear two-stage stochastic programming problems, *Mathematical Programming* 151 (2015), 315–345.

E. Novak: Some results on the complexity of numerical integration, arXiv:1409.6714v1 [math.NA], 2014.

A. B. Owen: The dimension distribution and quadrature test functions, *Statistica Sinica* 13 (2003), 1–17.

I. H. Sloan and H. Woźniakowski: When are Quasi Monte Carlo algorithms efficient for high-dimensional integration, *Journal of Complexity* 14 (1998), 1–33.

I. H. Sloan, F. Y. Kuo and S. Joe: Constructing randomly shifted lattice rules in weighted Sobolev spaces, *SIAM Journal Numerical Analysis* 40 (2002), 1650–1665.