QMC methods for stochastic programs: ANOVA decomposition of integrands

W. Römisch

Humboldt-University Berlin

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

(H. Heitsch, I. H. Sloan)

MCQMC 2012, Sydney, February 12–17, 2012

Introduction

- Stochastic programs are optimization problems containing integrals in the objective function and/or constraints.
- Applied stochastic programming models in production, transportation, energy, finance etc. are typically large scale.
- Standard approach for solving such models are variants of Monte Carlo for generating scenarios (i.e., samples).
- A few recent approaches to scenario generation in stochastic programming besides MC:

  (a) Optimal quantization of probability distributions (Pflug-Pichler 2010).

  (b) Quasi-Monte Carlo (QMC) methods (Koivu-Pennanen 05, Homem-de-Mello 06).

  (c) Sparse grid quadrature rules (Chen-Mehrotra 08).

While the justification of MC and (a) may be based on available stability results for stochastic programs, there is almost no reasonable justification of applying (b) and (c).

Personal interest: Applying and justifying randomized QMC methods (randomly shifted and digitally shifted polynomial lattice rules) with application in energy models.

Two-stage linear stochastic programs

Two-stage stochastic programs arise as deterministic equivalents of improperly posed random linear programs

$$\min\{\langle c, x\rangle : x \in X,\, Tx = \xi\},$$

where X is a convex polyhedral subset of $\mathbb{R}^m$, T a matrix, ξ is a d-dimensional random vector.

A possible deviation $\xi - Tx$ is compensated by additional costs $\Phi(x, \xi)$ whose mean with respect to the probability distribution P of ξ is added to the objective. We assume that the additional costs represent the optimal value of a *second-stage program*, namely,

$$\Phi(x, \xi) = \inf\{\langle q, y\rangle : y \in \mathbb{R}^{\bar{m}},\, Wy = \xi - Tx, y \geq 0\},$$

where $q \in \mathbb{R}^{\bar{m}}$, W a $(d, \bar{m})$-matrix (having rank d) and t varies in the polyhedral cone $W(\mathbb{R}^{\bar{m}}_+)$.

The *deterministic equivalent* then is of the form

$$\min\left\{\langle c, x\rangle + \int_{\mathbb{R}^d} \Phi(x, \xi) P(d\xi) : x \in X\right\}.$$

We assume that the additional costs are of the form

$$\Phi(x,\xi) = \varphi(\xi - Tx)$$

with the second-stage optimal value function

$$\begin{aligned}\varphi(t) &= \inf\{\langle q, y\rangle : Wy = t, y \geq 0\}\\ &= \sup\{\langle t, z\rangle : W^\top z \leq q\} = \sup_{z\in\mathcal{D}} \langle t, z\rangle\,,\end{aligned}$$

There exist vertices v^j of the dual feasible set $\mathcal{D}$ and polyhedral cones $\mathcal{K}_j$, $j = 1, \ldots, \ell$, decomposing $\operatorname{dom}\varphi$ such that

$$\varphi(t) = \langle v^j, t\rangle, \; \forall t \in \mathcal{K}_j, \quad \text{and} \quad \varphi(t) = \max_{j=1,\ldots,\ell} \langle v^j, t\rangle.$$

Hence, the integrands are of the form

$$f(\xi) = \max_{j=1,\ldots,\ell} \langle v^j, \xi - Tx\rangle.$$

Problem: When transformed to $[0,1]^d$, f is not of bounded variation in the Hardy-Krause sense and does not belong to tensor product Sobolev spaces $\bigotimes_{i=1}^d W_2^1([0,1])$ in general.

Model extensions

- Two-stage models with *affine functions* $h(\xi)$ and/or $T(\xi)$, hence, the integrands f are of the form

$$f(\xi) = \max_{j=1,\ldots,\ell} \langle v^j, h(\xi) - T(\xi)x \rangle.$$

- Two-stage models with *random second-stage costs* $q(\xi)$

$$f(\xi) = \max_{j=1,\ldots,\ell} \langle v^j(\xi), h(\xi) - Tx \rangle = \max_{j=1,\ldots,\ell} \langle C_j q(\xi), h(\xi) - T(\xi)x \rangle.$$

- *Multi-period models*: Random vector $\xi = (\xi_1, \ldots, \xi_T)$

$$f(\xi) \;=\; \Psi_1(\xi, x),$$

  where Ψ_1 is given by the DP recursion

$$\Phi_t(\xi^t, u_{t-1}) := \sup \left\{ \langle u_{t-1}, z_t \rangle + \Psi_{t+1}(\xi^t, z_t) : W_t^\top z_t \le q_t(\xi_t) \right\}$$
$$\Psi_t(\xi^t, z_{t-1}) := \Phi_t(\xi^t, h_t(\xi_t) - T_t(\xi_t) z_{t-1}),\ t = T, \ldots, 1,$$

  where $z_0 = x$, $\xi^t = (\xi_t, \ldots, \xi_T)$ and $\Psi_{T+1}(\xi^{T+1}, z_T) \equiv 0$.

- *Multi-stage models*: The only difference to multi-period is

$$\Psi_t(\xi^t, z_{t-1}) := \mathbb{E}[\Phi_t(\xi^t, h_t(\xi_t) - T_t(\xi_t) z_{t-1}) | \xi_1, \ldots, \xi_t].$$

ANOVA decomposition of multivariate functions

Idea: Decompositions of f may be used, where most of them are smooth, but hopefully only some of them relevant.

Let $D = \{1, \ldots, d\}$ and $f \in L_{1,\rho_d}(\mathbb{R}^d)$ with $\rho_d(\xi) = \prod_{j=1}^{d} \rho_j(\xi_j)$. Let the projection P_k, $k \in D$, be defined by

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

Clearly, $P_k f$ is constant with respect to ξ_k. For $u \subseteq D$ we write

$$P_u f = \Big(\prod_{k \in u} P_k\Big)(f),$$

where the product means composition, and note that the ordering within the product is not important because of Fubini's theorem. The function $P_u f$ is constant with respect to all x_k, $k \in u$. Note that P_u satisfies the properties of a projection.

ANOVA-decomposition of f:

$$f = \sum_{u \subseteq D} f_u \,,$$

where $f_\emptyset = I_d(f) = P_D(f)$ and recursively

$$f_u = P_{-u}(f) - \sum_{v \subseteq u} f_v$$

or

$$f_u = \sum_{v \subseteq u} (-1)^{|u|-|v|} P_{-v} f = P_{-u}(f) + \sum_{v \subset u} (-1)^{|u|-|v|} P_{u-v}(P_{-u}(f)),$$

where P_{-u} and P_{u-v} mean integration with respect to ξ_j, $j \in D \setminus u$ and $j \in u \setminus v$, respectively. The second representation motivates that f_u is essentially as smooth as $P_{-u}(f)$.

If f belongs to $L_{2,\rho_d}(\mathbb{R}^d)$, the ANOVA functions $\{f_u\}_{u \subseteq D}$ are orthogonal in $L_{2,\rho_d}(\mathbb{R}^d)$.

We set $\sigma^2(f) = \|f - I_d(f)\|_{L_2}^2$ and have

$$\sigma^2(f) = \|f\|_{L_2}^2 - (I_d(f))^2 = \sum_{\emptyset \neq u \subseteq D} \|f_u\|_{L_2}^2 .$$

The truncation dimension d_t of f is the smallest $d_t \in \mathbb{N}$ such that

$$\sum_{u \subseteq \{1,\dots,d_t\}} \|f_u\|_{L_2}^2 \geq p\sigma^2(f) \quad \text{(where } p \in (0,1) \text{ is close to } 1\text{)}.$$

Then it holds

$$\Big\| f - \sum_{u \subseteq \{1,\dots,d_t\}} f_u \Big\|_{L_2} \leq (1-p)\sigma^2(f).$$

(Wang-Fang 03, Kuo-Sloan-Wasilkowski-Woźniakowski 10, Griebel-Holtz 10)

According to an observation of Griebel-Kuo-Sloan 10 the ANOVA terms f_u can be smoother than f under certain conditions.

ANOVA decomposition of two-stage integrands

Assumption:

(A1) $W(\mathbb{R}^{\bar{m}}_+) = \mathbb{R}^d$ (*complete recourse*).

(A2) $\mathcal{D} \neq \emptyset$ (*dual feasibility*).

(A3) $\int_{\mathbb{R}^d} \|\xi\| P(d\xi) < \infty$.

(A4) P has a density of the form $\rho_d(\xi) = \prod_{j=1}^d \rho_j(\xi_j)$ ($\xi \in \mathbb{R}^d$) with $\rho_j \in C(\mathbb{R})$, $j = 1, \ldots, d$.

(A1) and (A2) imply that $\mathrm{dom}\,\varphi = \mathbb{R}^d$ and $\mathcal{D}$ is bounded and, hence, it is the convex hull of its vertices. Furthermore, the cones $\mathcal{K}_j$ are the normal cones to $\mathcal{D}$ at the vertices v^j, i.e.,

$$\begin{aligned}\mathcal{K}_j &= \{t \in \mathbb{R}^d : \langle t, z - v^j \rangle \leq 0, \forall z \in \mathcal{D}\} \quad (j = 1, \ldots, \ell) \\ &= \{t \in \mathbb{R}^d : \langle t, v^i - v^j \rangle \leq 0, \forall i = 1, \ldots, \ell, \, i \neq j\}.\end{aligned}$$

It holds that $\cup_{j=1,\ldots,\ell} \mathcal{K}_j = \mathbb{R}^d$ and for $j \neq j'$ the intersection $\mathcal{K}_j \cap \mathcal{K}_{j'}$ is a common closed face of dimension $d-1$ iff the two cones are adjacent and is contained in

$$\{t \in \mathbb{R}^d : \langle t, v^{j'} - v^j \rangle = 0\}.$$

To compute projections $P_k(f)$ for $k \in D$. Let $\xi_i \in \mathbb{R}$, $i = 1, \ldots, d$, $i \neq k$, be given. We set $\xi^k = (\xi_1, \ldots, \xi_{k-1}, \xi_{k+1}, \ldots, \xi_d)$ and

$$\xi_s = (\xi_1, \ldots, \xi_{k-1}, s, \xi_{k+1}, \ldots, \xi_d) \in \mathbb{R}^d = \cup_{j=1,\ldots,\ell} \mathcal{K}_j.$$

Assuming (A1)–(A4) it is possible to derive an explicit representation of $P_k(f)$ that depends on ξ^k and on the finitely many points at which the one-dimensional affine subspace $\{\xi_s : s \in \mathbb{R}\}$ meets the common face of two adjacent cones. This leads to

Proposition:

Let $k \in D$. Assume (A1)–(A4) and that all adjacent vertices of $\mathcal{D}$ have different kth components.

The kth projection $P_k f$ is infinitely differentiable if the density ρ_k is in $C^\infty(\mathbb{R})$ and all its derivatives are bounded on $\mathbb{R}$.

Theorem:
Let $u \subset D$. Assume (A1)–(A4) and that all adjacent vertices of $\mathcal{D}$ have different kth components for some $k \in D \setminus u$.
Then the ANOVA term f_u belongs to $C^\infty(\mathbb{R}^{d-|u|})$ if $\rho_k \in C^\infty(\mathbb{R})$ and all its derivatives are bounded on $\mathbb{R}$.

Remark: The algebraic condition on the vertices of $\mathcal{D}$ is satisfied almost everywhere in the following sense:
Given $\mathcal{D}$ there are only finitely many orthogonal matrices Q performing rotations of $\mathbb{R}^d$ such that the condition is not satisfied for $Q\mathcal{D} = \{z \in \mathbb{R}^d : (QW)^\top z \leq q\}$. Note that then the optimal value $\phi(t)$ is equal to $\max\{\langle Qt, z\rangle : z \in Q\mathcal{D}\}$. Such an orthogonal transformation of $\mathcal{D}$ leads only to simple changes.

Example:

Let $\bar{m} = 3$, $d = 2$, P denote the two-dimensional standard normal distribution and let the following vector q and matrix W

$$W = \begin{pmatrix} -1 & 1 & 0 \\ 1 & 1 & -1 \end{pmatrix} \qquad q = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

be given. Then (A1) and (A2) are satisfied and the dual feasible set $\mathcal{D}$ is the triangle (in $\mathbb{R}^2$)

$$\mathcal{D} = \{z \in \mathbb{R}^2 : -z_1 + z_2 \leq 1, z_1 + z_2 \leq 1, -z_2 \leq 0\},$$

with the vertices

$$v^1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad v^2 = \begin{pmatrix} -1 \\ 0 \end{pmatrix} \quad v^3 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

The normal cones $\mathcal{K}_j$ to $\mathcal{D}$ at v^j, $j = 1, 2, 3$, are

$$\begin{aligned} \mathcal{K}_1 &= \{z \in \mathbb{R}^2 : z_1 \geq 0, z_2 \leq z_1\}, \\ \mathcal{K}_2 &= \{z \in \mathbb{R}^2 : z_1 \leq 0, z_2 \leq -z_1\}, \\ \mathcal{K}_3 &= \{z \in \mathbb{R}^2 : z_2 \geq z_1, z_2 \geq -z_1\}. \end{aligned}$$

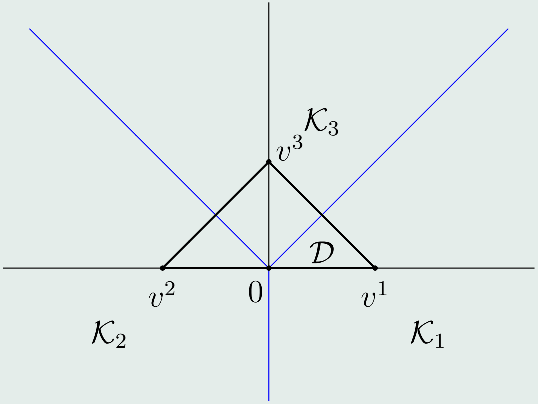

Figure 1: Illustration of $\mathcal{D}$, its vertices v^j and the normal cones $\mathcal{K}_j$ to its vertices

Hence, the second component of the two adjacent vertices v^1 and v^2 coincides. The function φ is of the form

$$\varphi(t) = \max_{i=1,2,3} \langle v^i, t\rangle = \max\{t_1, -t_1, t_2\} = \max\{|t_1|, t_2\}$$

and the integrand is

$$f(\xi) = \max\{|\xi_1 - [Tx]_1|, \xi_2 - [Tx]_2\}$$

The ANOVA projection $P_1 f$ is in C^∞, but $P_2 f$ is not differentiable.

Remark: Under the assumptions of the theorem the function

$$f_{d-1}(\xi) = \sum_{|u| \leq d-1} f_u$$

is in $C^\infty(\mathbb{R}^d)$ if $\rho_k \in C^\infty(\mathbb{R})$ and all its derivatives are bounded on $\mathbb{R}$ for every $k \in D$. On the other hand, it holds

$$f = f_{d-1} + f_D.$$

Hence, the question arises: For which two-stage linear stochastic programs is the L_2-norm of f_D small or, equivalently, is f_{d-1} a good approximation of f in L_{2,ρ_d}?

Open problem: Estimates of the truncation dimension of two-stage linear stochastic programs ?

Conclusions

- The results provide a first theoretical explanation of our computational results close to the optimal rate for randomly shifted lattice rules applied to two-stage stochastic programs.
- The results will be extended to more general two-stage situations.
- Numerical experiments with and without orthogonal transformations will hopefully lead to more computational insight into the geometric condition on adjacent vertices.
- **Challenge:** Multi-stage and integer stochastic programs.

References

M. Chen and S. Mehrotra: Epi-convergent scenario generation method for stochastic problems via sparse grid, *SPEPS* 7-2008.

J. Dick, F. Pillichshammer: *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration*, Cambridge University Press, 2010.

M. Griebel, F. Y. Kuo, I. H. Sloan: The smoothing effect of the ANOVA decomposition, *Journal of Complexity* 26 (2010), 523–551.

M. Griebel, F. Y. Kuo and I. H. Sloan: The smoothing effect of integration in $\mathbb{R}^d$ and the ANOVA decomposition, *Mathematics of Computation* (to appear).

T. Homem-de-Mello: On rates of convergence for stochastic optimization problems under non-i.i.d. sampling, *SIAM Journal on Optimization* 19 (2008), 524-551.

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.

F. Y. Kuo, I. H. Sloan, G. W. Wasilkowski, B. J. Waterhouse: Randomly shifted lattice rules with the optimal rate of convergence for unbounded integrands, *Journal of Complexity* 26 (2010), 135–160.

A. B. Owen: Multidimensional variation for Quasi-Monte Carlo, in J. Fan, G. Li (Eds.), International Conference on Statistics, World Scientific Publ., 2005, 49–74.

D. Nuyens and R. Cools: Fast algorithms for component-by-component constructions of rank-1 lattice rules in shift-invariant reproducing kernel Hilbert spaces, *Mathematics of Computation* 75 (2006), 903-922.

T. Pennanen, M. Koivu: Epi-convergent discretizations of stochastic programs via integration quadratures, *Numerische Mathematik* 100 (2005), 141–163.

G. Ch. Pflug, A. Pichler: Scenario generation for stochastic optimization problems, in: Stochastic Optimization Methods in Finance and Energy (M.I. Bertocchi, G. Consigli, M.A.H. Dempster eds.), Springer, 2011.

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

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

I. H. Sloan, F. Y. Kuo and S. Joe: On the step-by-step construction of Quasi-Monte Carlo integration rules that achieve strong tractability error bounds in weighted Sobolev spaces, *Mathematics of Computation* 71 (2002), 1609-1640.

X. Wang and K.-T. Fang: The effective dimension and Quasi-Monte Carlo integration, *Journal of Complexity* 19 (2003), 101–124.