# Stochastic optimization, multivariate numerical integration and Quasi-Monte Carlo methods


W. Römisch

Humboldt-University Berlin
Institute of Mathematics

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

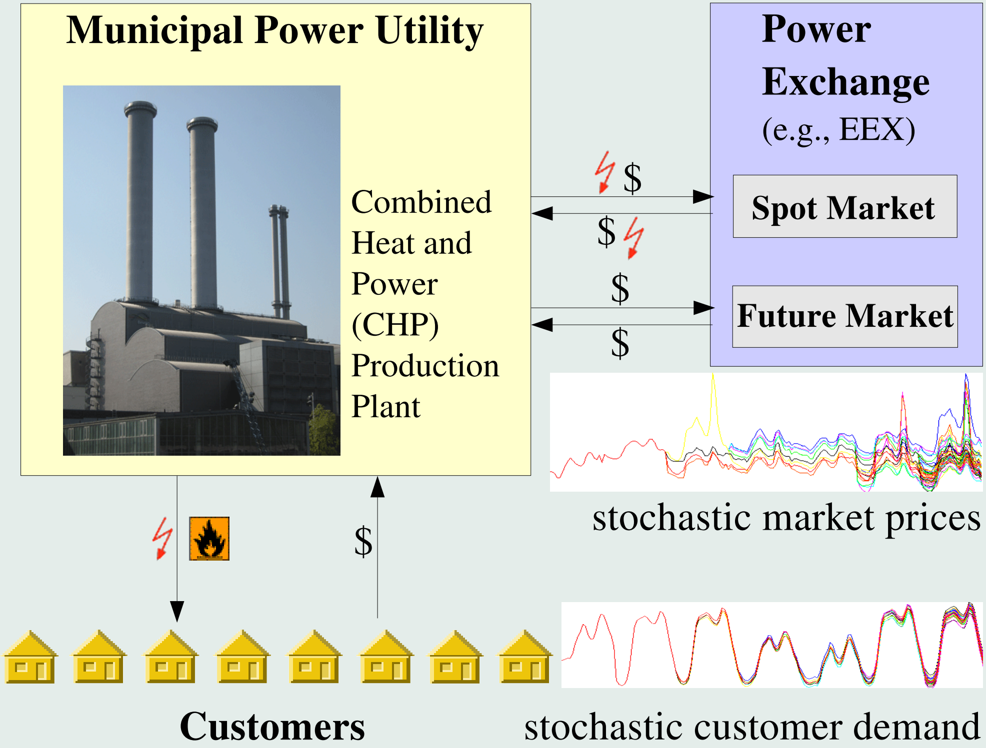


Chemnitzer Mathematisches Colloquium, TU Chemnitz, 29.6.2017

## Introduction and overview

- Stochastic optimization: Mathematics of decision making under uncertainty.
- Two-stage stochastic optimization is a standard problem. But, the evaluation of the objective of such models is #P-hard (Hanasusanto-Kuhn-Wiesemann 16).
- Computational methods for solving stochastic optimization problems 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.

## Application: Mean-Risk Electricity Portfolio Management




(Eichhorn-Römisch-Wegner 05, Eichhorn-Heitsch-Römisch 10)

## Linear two-stage stochastic programming models

Consider a linear program with stochastic parameters of the form

$$\min\{\langle c, x\rangle : x \in X,\, T(\xi)x = h(\xi)\},$$

where $\xi : \Omega \to \Xi$ is a random vector defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, $c \in \mathbb{R}^m$, $\Xi$ and $X$ are polyhedral subsets of $\mathbb{R}^d$ and $\mathbb{R}^m$, respectively, and the $r \times m$-matrix $T(\cdot)$ and vector $h(\cdot) \in \mathbb{R}^r$ are affine functions of $\xi$.

**Idea:** Introduce a recourse variable $y \in \mathbb{R}^{\overline{m}}$, recourse costs $q(\cdot) \in \mathbb{R}^{\overline{m}}$ as affine function of $\xi$, a fixed recourse $r \times \overline{m}$-matrix $W$, a polyhedral cone $Y \subseteq \mathbb{R}^{\overline{m}}$, and solve the second-stage or recourse program

$$\min\{\langle q(\xi), y\rangle : y \in Y, Wy = h(\xi) - T(\xi)x\}.$$

Define the optimal recourse costs

$$Q(x, \xi) := \Phi(q(\xi), h(\xi) - T(\xi)x) = \inf\{\langle q(\xi), y\rangle : y \in Y, Wy = h(\xi) - T(\xi)x\}$$

and add the expected recourse costs $\mathbb{E}[Q(x, \xi)]$ (depending on the first-stage decision $x$) to the original objective and consider the two-stage program

$$\min\left\{\langle c, x\rangle + \mathbb{E}[Q(x, \xi)] : x \in X\right\}.$$

## Structural properties of two-stage models

We consider the infimum function $\Phi(\cdot,\cdot)$ of the parametrized linear (second-stage) program, namely,

$$\begin{aligned}\Phi(u,t) &= \inf\left\{\langle u,y\rangle : Wy=t, y\in Y\right\} \quad ((u,t)\in\mathbb{R}^{\overline{m}}\times\mathbb{R}^{r})\\ &= \sup\left\{\langle t,z\rangle : W^{\top}z-u\in Y^{*}\right\} \quad ((u,t)\in\mathcal{D}\times W(Y))\\ \mathcal{D} &= \left\{u\in\mathbb{R}^{\overline{m}} : \{z\in\mathbb{R}^{r} : W^{\top}z-u\in Y^{*}\}\neq\emptyset\right\}\end{aligned}$$

where $W^{\top}$ denotes the transposed of the recourse matrix $W$ and $Y^{\star}$ the polar cone of $Y$ and we used linear programming duality.

**Theorem:** (Walkup-Wets 69)
The function $\Phi(\cdot,\cdot)$ is finite and continuous on the polyhedral cone $\mathcal{D}\times W(Y)$.
Furthermore, the function $\Phi(u,\cdot)$ is piecewise linear convex on $W(Y)$ for fixed $u\in\mathcal{D}$, and $\Phi(\cdot,t)$ is piecewise linear concave on $\mathcal{D}$ for fixed $t\in W(Y)$.
There exists a decompostion of $\mathcal{D}\times W(Y)$ into polyhedral cones $\mathcal{K}_j$, $j=1,\ldots,\ell$, and $\overline{m}\times r$ matrices $C_j$ such that

$$\Phi(u,t)=\max_{j=1,\ldots,\ell}\langle C_j u,t\rangle.$$

**Assumptions:**

**(A1)** *relatively complete recourse*: for any $(\xi, x) \in \Xi \times X$, $h(\xi) - T(\xi)x \in W(Y)$;

**(A2)** *dual feasibility*: $q(\xi) \in \mathcal{D}$ holds for all $\xi \in \Xi$.

**(A3)** *finite second order moment*: $\int_\Xi \|\xi\|^2 P(d\xi) < \infty$.

Note that (A1) is satisfied if $W(Y) = \mathbb{R}^d$ (complete recourse). In general, (A1) and (A2) impose a condition on the support of $P$.

**Proposition:** (Wets 74)

Assume (A1) and (A2). Then the deterministic equivalent of the two-stage model represents a convex program (with linear constraints) if the integrals $\int_\Xi \Phi(q(\xi), h(\xi) - T(\xi)x)P(d\xi)$ are finite for all $x \in X$. For the latter it suffices to assume (A3).

An element $x \in X$ minimizes the convex two-stage program if and only if

$$0 \in \int_\Xi \partial Q(x, \xi) P(d\xi) + N_X(x)\,,$$

$$\partial Q(x, \xi) = c - T(\xi)^\top \arg \max_{z \in \mathbb{R}^r, W^\top z - q(\xi) \in Y^\star} \langle z, h(\xi) - T(\xi)x \rangle.$$

## Complexity of two-stage stochastic programs

The two papers Dyer-Stougie 06, Hanasusanto-Kuhn-Wiesemann 16 consider the following second-stage optimal value function

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

where $\alpha \in \mathbb{R}^d_+$ and $\beta \in \mathbb{R}_+$ are parameters and the random vector $\xi$ is uniformly distributed in $[0,1]^d$. Starting with the identity

$$\max\{\gamma - \beta, 0\} = \gamma - \int_0^\beta \mathbb{1}_{\{\gamma \ge t\}} dt$$

we find for the expected recourse function

$$\begin{aligned} \mathbb{E}[Q(\alpha,\beta,\xi)] &= \mathbb{E}[\alpha^\top \xi] - \int_0^\beta \mathbb{E}\left[\mathbb{1}_{\{\alpha^\top \xi \ge t\}}\right] dt \\ &= \frac{1}{2}\alpha^\top e - \beta + \int_0^\beta \operatorname{Vol} P(\alpha,t) dt \\ \frac{\partial \mathbb{E}[Q(\alpha,\beta,\xi)]}{\partial \beta} &= \operatorname{Vol} P(\alpha,\beta) - 1\,, \end{aligned}$$

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

**Theorem:** (Hanasusanto-Kuhn-Wiesemann 16)
For any pair $(\alpha, \beta) \in \mathbb{R}_+^{d+1}$ there exists $\varepsilon(d, \alpha) > 0$ such that the computation of $\mathbb{E}[Q(\alpha, \beta, \xi)]$ within an absolute accuracy of $\varepsilon < \varepsilon(d, \alpha)$ is at least as hard as computing the volume $\mathrm{Vol}\, P(\alpha, \beta)$ of the knapsack polytope.
The computation of the latter is #P-hard (Dyer-Frieze 88).

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 $d$.

The complexity class #P contains the counting problems associated with the decision problems in the complexity class NP. A counting problem is in #P if the items to be counted can be validated as such in polynomial time. A #P problem is at least as difficult as the corresponding NP problem.
It is therefore commonly believed that #P-hard problems do not admit polynomial-time solution methods.

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 have mixed Sobolev derivatives 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 $\Phi(q(\cdot), h(\cdot) - T(\cdot)x)$ is continuous and piecewise linear-quadratic on $\Xi$ if (A1) and (A2) is satisfied. It holds

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

for some $\ell \in \mathbb{N}$ and $r \times \overline{m}$-matrices $C_j$, $j = 1, \ldots, \ell$. The latter correspond to some decomposition of $\mathcal{D} \times W(Y)$ into $\ell$ polyhedral cones.

## Discrete approximations of two-stage models

Replace the (original) probability measure $P$ by measures $P_n$ having (finite) discrete support $\{\xi_1, \ldots, \xi_n\}$ $(n \in \mathbb{N})$, i.e.,

$$P_n = \sum_{i=1}^{n} w_i \delta_{\xi_i},$$

and insert it into the infinite-dimensional stochastic program:

$$\min\{\langle c, x\rangle + \sum_{i=1}^{n} w_i \langle q(\xi_i), y_i\rangle : x \in X, y_i \in Y, \, i = 1, \ldots, n,$$

$$\begin{array}{ccccccc} Wy_1 & & & & +T(\xi_1)x & = & h(\xi_1) \\ & Wy_2 & & & +T(\xi_2)x & = & h(\xi_2) \\ & & \ddots & & \vdots & = & \vdots \\ & & & Wy_n & +T(\xi_n)x & = & h(\xi_n)\} \end{array}$$

Hence, we arrive at an (often) large scale block-structured linear program which is solvable in polynomial time and allows for specific decomposition methods.

(Ruszczyński-Shapiro 2003, Kall-Mayer 2005 (2010))

## Complexity of numerical integration

Each absolutely continuous probability distribution on $\mathbb{R}^d$ can be transformed into the uniform distribution on $[0,1]^d$ (Rosenblatt 52).

Hence, we may 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 minimal error are given by:

$$e(Q_n) = \sup_{f \in \mathbb{B}_d} |I_d(f) - Q_n(f)| \quad \text{and} \quad e(n, \mathbb{B}_d) = \inf_{Q_n} e(Q_n) \, .$$

(Novak 16)

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 $\varepsilon$, 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 $(\varepsilon, d)$ 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 probability space $(\Omega, \mathcal{F}, \mathbb{P})$.. We assume that $Q(\omega)$ is a quadrature algorithm for each $\omega$ and that it depends on $\omega$ 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

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

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

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

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 $\gamma_i$ are positive and $\gamma_u$ 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

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

## Monte Carlo sampling

Monte Carlo methods are based on drawing independent identically distributed (i.i.d.) $\Xi$-valued random samples $\xi^1(\cdot), \ldots, \xi^n(\cdot), \ldots$ (defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$) from an underlying probability distribution $P$ (on $\Xi$) such that

$$Q_n(\omega)(f) = \frac{1}{n}\sum_{i=1}^{n} f(\xi^i(\omega)),$$

i.e., $Q_n(\cdot)$ is a random functional, and it holds

$$\lim_{n\to\infty} Q_n(\omega)(f) = \int_{\Xi} f(\xi)P(d\xi) = \mathbb{E}(f) \quad \mathbb{P}\text{-almost surely}$$

for every real continuous and bounded function $f$ on $\Xi$.
If $P$ has a finite moment of order $r \geq 1$, the error estimate

$$\mathbb{E}\left[\left|\frac{1}{n}\sum_{i=1}^{n} f(\xi^i(\omega)) - \mathbb{E}(f)\right|^r\right] \leq \frac{\mathbb{E}\left[(f - \mathbb{E}(f))^r\right]}{n^{r-1}}$$

is valid.

(Shapiro-Dentcheva-Ruszczyński 2009 (2014))

Hence, the mean square convergence rate is

$$\|Q_n(\omega)(f) - \mathbb{E}(f)\|_{L_2} = \sigma(f)n^{-\frac{1}{2}},$$

where $\sigma^2(f) = \mathbb{E}\left[(f - \mathbb{E}(f))^2\right]$ is assumed to be finite.

Advantages:

(i) MC sampling works *for (almost) all integrands*.

(ii) The machinery of probability theory is available.

(iii) The convergence *rate does not depend on* $d$.

Deficiencies: (Niederreiter 92)

(i) There exist 'only' *probabilistic error bounds*.

(ii) Possible regularity of the integrand *does not improve* the rate.

(iii) Generating (independent) random samples is *difficult*.

Practically, iid samples are approximately obtained by pseudo random number generators as uniform samples in $[0,1]^d$ and later transformed to more general sets $\Xi$ and distributions $P$.

Good pseudo random number generator: Mersenne Twister (Matsumoto-Nishimura 98)

## Quasi-Monte Carlo methods

The basic idea of Quasi-Monte Carlo (QMC) methods is to use deterministic points that are (in some way) uniformly distributed in $[0,1]^d$ and to consider the approximate computation of

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

by a QMC algorithm with (non-random) points $\xi^i$, $i = 1, \ldots, n$, from $[0,1]^d$:

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

The uniform distribution property of point sets may be defined in terms of the so-called $L_p$-discrepancy of $\xi^1, \ldots, \xi^n$ for $1 \leq p \leq \infty$

$$d_{p,n}(\xi^1, \ldots, \xi^n) = \Big( \int_{[0,1]^d} |\mathrm{disc}(\xi)|^p d\xi \Big)^{\frac{1}{p}}, \quad \mathrm{disc}(\xi) := \prod_{j=1}^{d} \xi_j - \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{[0,\xi)}(\xi^i) \,.$$

There exist sequences $(\xi^i)$ in $[0,1]^d$ such that for all $\delta \in (0, \frac{1}{2}]$

$$d_{\infty,n}(\xi^1, \ldots, \xi^n) = O(n^{-1}(\log n)^d) \quad \text{or} \quad d_{\infty,n}(\xi^1, \ldots, \xi^n) \leq C(d,\delta) n^{-1+\delta} \,.$$

## 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 $\delta$ decreases, but does not depend on the dimension $d$ if the sequence $(\gamma_j)$ satisfies the condition

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

(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.

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

where $\xi$ 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

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

The previous representation can be more compactly written as

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

where $\mathfrak{D} = \{1, \dots, d\}$ and $\xi^u$ contains only the components $\xi_j$ 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)**

**(A3)** $P$ has fourth order absolute moments.

**(A4)** $P$ 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 $\rho_i$, $i \in \mathfrak{D}$.

**(A5)** For each $x \in X$ all common faces of 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 $\mathcal{K}_j$, $j = 1, \ldots, \ell$, decompose $\operatorname{dom} \Phi = \mathcal{D} \times W(Y)$ (geometric condition).

**Theorem:** Let $x \in X$, assume (A1)–(A5) 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 $\rho_k$, $k \in \mathfrak{D}$, 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 computation of the objective of linear two-stage stochastic programs with fixed recourse with a sufficiently high accuracy is $\#$P-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.

**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.

G. A. Hanasusanto, D. Kuhn and W. Wiesemann: A comment on "computational complexity of stochastic programming problems", *Mathematical Programming* 159 (2016), 557–569.

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, in *Monte Carlo and Quasi-Monte Carlo Methods* (R. Cools and D. Nuyens eds.), Springer, 2016.

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

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

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.