Progress in high-dimensional numerical integration and its application to stochastic optimization

W. Römisch

Humboldt-University Berlin
Department of Mathematics

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

DFG Research Center MATHEON
Mathematics for key technologies

Short course, Rutgers University, May 21–23, 2013

Part I: Introduction to Monte Carlo, Quasi-Monte Carlo and sparse grid techniques

Contents:

Introduction

- Numerical integration belongs to the standard problems of numerical analysis.
- It is often needed as subproblem algorithm for solving more involved problems (e.g. PDEs).
- A number of problems in physics (e.g. quantum physics) require the computation of high-dimensional integrals.
- Any expectation in stochastic models means computing integrals.
- Computing the risk of decisions, e.g., the numerical evaluation of risk measures, requires numerical integration (often in high dimensions).
- Many applied stochastic optimization models (with mean-risk objective and/or risk constraints) in engineering, production, energy or finance contain a medium- or long-term time horizon and are highly complex. Their solution process requires repeatedly (very) high-dimensional numerical integration.

Introduction to numerical integration

Classical theory of numerical integration in dimension $d=1$:

Quadrature rule

$$\int_0^1 f(\xi)d\xi \approx \sum_{i=1}^{n} w_i f(\xi^i),$$

where $\xi^i \in [0,1]$ are the quadrature points (knots) and $w_i \in \mathbb{R}$, $i=1,\dots,n$, the quadrature weights satisfying $\sum_{i=1}^n w_i = 1$.

Quadrature error

$$e_n(f) = \int_0^1 f(\xi)d\xi - \sum_{i=1}^{n} w_i f(\xi^i).$$

Examples:

(Left) rectangle rule with equally-spaced points $\xi^i = \frac{i-1}{n}$ and weights $w_i = \frac{1}{n}$, $i=1,\dots,n$, with error $e_n(f) \leq \frac{\|f'\|_\infty}{2n}$ if $f \in C^1$;

the trapezoidal rule with $e_n(f) = O(\frac{1}{n^2})$ if $f \in C^2$;

the Simpson rule with $e_n(f) = O(\frac{1}{n^4})$ if $f \in C^4$;

the Gaussian quadrature rules with quadrature points being roots of certain polynomials are exact for all polynomials of degree $2n-1$.

Theorem: It holds

$$\lim_{n\to\infty} e_n(f) = 0 \text{ for all } f \in C([0,1]) \quad \text{iff} \quad \sup_{n\in\mathbb{N}} \sum_{i=1}^{n} |w_i| < \infty.$$

The result carries over to $[0,1]^d$, $d > 1$, and to more general domains.

How to extend the ideas to higher dimension $d > 1$?

An **obvious way** is the **product rule** in $[0,1]^d$:
Take d one-dimensional quadrature rules with weights $u_{i_j} \in \mathbb{R}$ and points $\xi^{i_j} \in [0,1]$, $j = 1, \ldots, d$ and consider

$$\int_0^1 \cdots \int_0^1 f(\xi_1, \ldots, \xi_d) d\xi_1 \cdots d\xi_d \approx \sum_{i_1=1}^{m_1} \cdots \sum_{i_d=1}^{m_d} \prod_{j=1}^{d} u_{i_j} f(\xi^{i_1}, \ldots, \xi^{i_d})$$

Total number of quadrature points is $n = \prod_{j=1}^{d} m_j$. For $m_j = m$, $j = 1, \ldots, d$, the total number is $n = m^d$, hence, it grows exponentially.
For example, the product rectangular rule has order $O(m^{-1}) = O(n^{-\frac{1}{d}})$.

("curse of dimensionality")

Alternative approaches:

(1) Use independent identically distributed random samples ξ^i, $i \in \mathbb{N}$, with common uniform probability distribution on $[0,1]^d$ (defined on some probability space) – **Monte Carlo method**.

(2) Determine a deterministic sequence $\xi^i \in [0,1]^d$, $i \in \mathbb{N}$, such that the sequence

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{\xi^i}, \quad n \in \mathbb{N},$$

of discrete (probability) measures converges to the uniform probability distribution on $[0,1]^d$, i.e., to the Lebesgue measure λ^d on $[0,1]^d$ in a suitable sense (e.g. uniform convergence of distribution functions) – **Quasi-Monte Carlo method**.

(3) Remove a suitably large number of equally spaced product quadrature points such that the convergence rate is close to that of one-dimensional quadrature rules (except for some logarithmic terms) – **sparse grid method**.

Original mathematical background:

(1) - asymptotic statistics, (2) - number theory, (3) - complexity theory.

Transformation of integrals for general probability distributions P

For some function $f : \mathbb{R}^d \to \mathbb{R}$ we consider the integral

$$\int_{\mathbb{R}^d} f(x)P(dx)\,.$$

First step: Transform a multivariate probability distribution P on $\mathbb{R}^d$ to a distribution on $\mathbb{R}^d$ with d independent one-dimensional marginal distributions by using methods related to the class $\mathcal{P}$ of distributions with $P \in \mathcal{P}$.

Example: If P is normal with zero mean and nonsingular covariance matrix Σ. If A is any matrix satisfying $\Sigma = A\,A^\top$, then the distribution $P \circ A$ has independent marginals.

Second step: Let $F_k : \mathbb{R} \to [0,1]$ denote the marginal distribution functions and ρ_k, $k = 1,\ldots,d$, the marginal densities of a probability distribution P with independent marginals. Then by $\xi_k = F_k(x_k)$, $d\xi_k = \rho_k(x_k)dx_k$, $k = 1,\ldots,d$,

$$\int_{\mathbb{R}^d} f(x)P(dx) = \int_{\mathbb{R}^d} f(x)\prod_{k=1}^{d}\rho_k(x_k)dx = \int_{[0,1]^d} f(F_1^{-1}(\xi_1),\ldots,F_d^{-1}(\xi_d))d\xi_1\cdots d\xi_d$$

Monte Carlo sampling

Monte Carlo methods are based on drawing independent identically distributed (iid) Ξ-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 such that

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

i.e., $Q_{n,d}(\cdot)$ is a random functional, and it holds by the law of large numbers

$$\lim_{n\to\infty} Q_{n,d}(\omega)(f) = \int_{[0,1]^d} f(\xi)d\xi = I_d(f) = \mathbb{E}[f] \quad \mathbb{P}\text{-almost surely}$$

for every real continuous and bounded function f on Ξ.

If P has a finite moment of order $r \geq 1$, the error estimate

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

is valid.

Hence, the mean square convergence rate is

$$\|Q_{n,d}(\cdot)(f) - I_d(f)\|_{L_2} \leq \sigma(f) n^{-\frac{1}{2}},$$

where $\sigma^2(f) = \mathbb{E}\left((f - \mathbb{E}(f))^2\right)$ is the variance of f. Note that even equality holds without any assumption on f except $\sigma(f) < \infty$.

Moreover, it holds

$$\begin{aligned}
\mathbb{E}[Q_{n,d}(\omega)(f)] &= \tfrac{1}{n}\sum_{i=1}^{n} \mathbb{E}[f(\xi^i(\omega))] = I_d(f) \\
\mathrm{Var}[Q_{n,d}(\omega)(f)] &= \frac{\sigma^2(f)}{n} \\
\lim_{n\to\infty} \mathbb{P}\Big(|I_d(f) - Q_{n,d}(\cdot)(f)| \leq c\frac{\sigma(f)}{\sqrt{n}}\Big) &= \frac{1}{\sqrt{2\pi}}\int_{-c}^{c} \exp\big(-\tfrac{t^2}{2}\big)dt
\end{aligned}$$

and an unbiased estimator for $\mathrm{Var}[Q_{n,d}(\omega)(f)]$ is given by

$$\frac{1}{n(n-1)}\Big(\sum_{i=1}^{n} f^2(\xi^i) - n[Q_{n,d}(\cdot)(f)]^2\Big).$$

Advantages:

(i) MC sampling works for (almost) all integrands and is unbiased.

(ii) The machinery of probability theory is available.

(iii) The convergence rate $O(n^{-\frac{1}{2}})$ does not depend on the dimension 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.

(iv) MC methods are in practice (distressingly) slow.

Practically, iid samples are approximately obtained by so-called pseudo random number generators as uniform samples in $[0, 1]^d$.

Survey: P. L'Ecuyer: Uniform random number generation, *AOR* 53 (1994).

Classical linear congruential generators:

Its parameters are a large $M \in \mathbb{N}$ (modulus), a multiplier $a \in \mathbb{N}$ with $1 \leq a < M$ and $gcd(a, M) = 1$, and $c \in Z_M = \{0, 1, \ldots, M - 1\}$.

Starting with $y_0 \in Z_M$ a sequence is generated by

$$y_n \equiv a y_{n-1} + c \mod M \qquad (n \in \mathbb{N})$$

and the linear congruential pseudo random numbers are

$$\xi^n = \frac{y_n}{M} \in [0, 1).$$

The period $M - 1$ is chosen as a large prime number, e.g., $M = 2^{32}$.
Linear congruential pseudo random numbers fall mainly into planes (Marsaglia 68)!

Use only pseudo random number generators having passed a series of statistical tests, e.g., uniformity test, serial correlation test, monkey tests etc.

CAT-test: There are $26^3 = 17576$ possible 3-letter words. With $a = 69069$, $c = 0$, $M = 2^{32}$ one gets CAT after n=13561, then after 18263, and the third after 14872 calls. Other generators are even unable to produce CAT after 10^6 calls (Marsaglia-Zaman 93).

Warning:
For linear congruential generators never use more than $\frac{M}{4}$ or even $\frac{M}{10}$ calls.

Excellent pseudo random number generator: Mersenne Twister.
It has the astronomical period $2^{19937} - 1$ and provides 623-dimensional equidistribution up to 32-bit accuracy (Matsumoto-Nishimura 98).

Quasi-Monte Carlo methods

The basic idea of Quasi-Monte Carlo (QMC) methods is to replace random samples in Monte Carlo methods by deterministic points that are (in some way) uniformly distributed in $[0,1]^d$. So, we consider the approximate computation of

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

by a QMC algorithm

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

with (non-random) points ξ^i, $i = 1, \ldots, n$, from $[0,1]^d$.

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

$$D_n^*(\xi^1, \ldots, \xi^n) := \sup_{\xi \in [0,1]^d} |\mathrm{disc}(\xi)|, \quad \mathrm{disc}(\xi) := \prod_{i=1}^{d} \xi_i - \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{[0,\xi)}(\xi^i) \, .$$

(uniform distance of the uniform distribution function and the sample distribution function)

A sequence $(\xi^i)_{i\in\mathbb{N}}$ is called uniformly distributed in $[0,1]^d$ if

$$D_n^*(\xi^1,\ldots,\xi^n) \to 0 \quad \text{for} \quad n\to\infty$$

How fast can $D_n^*(\xi^1,\ldots,\xi^n)$ **converge to zero** as $n\to\infty$?

A **classical result** due to Roth 54 provides the lower bound

$$D_n^*(\xi^1,\ldots,\xi^n) \geq B_d \frac{(\log n)^{\frac{d-1}{2}}}{n}$$

for some constant B_d and all sequences (ξ^i) in $[0,1]^d$.

Later it becomes clear that there exist sequences (ξ^i) in $[0,1]^d$ such that

$$D_n^*(\xi^1,\ldots,\xi^n) = O(n^{-1}(\log n)^{d-1}).$$

Are there standard constructions for determining such sequences?

Classical convergence results:

Theorem: (Proinov 88)
If the real function f is continuous on $[0,1]^d$, then there exists $C > 0$ such that

$$|Q_{n,d}(f) - I_d(f)| \leq C\omega_f\Big(D_n^*(\xi^1,\ldots,\xi^n)^{\frac{1}{d}}\Big),$$

where $\omega_f(\delta) = \sup\{|f(\xi) - f(\tilde{\xi})| : \|\xi - \tilde{\xi})\| \leq \delta,\ \xi, \tilde{\xi} \in [0,1]^d\}$ is the modulus of continuity of f.

Theorem: (Koksma-Hlawka 61)
If f is of bounded variation $V_{\mathrm{HK}}(f)$ in the sense of Hardy and Krause, it holds

$$|I_d(f) - Q_{n,d}(f)| \leq V_{\mathrm{HK}}(f) D_n^*(\xi^1,\ldots,\xi^n)\,.$$

for any $n \in \mathbb{N}$ and any $\xi^1,\ldots,\xi^n \in [0,1]^d$.

Note that $V_{\mathrm{HK}}(f) < \infty$ is more restrictive than one might think at first moment. For example, one needs the existence of the mixed derivative $\frac{\partial^d f}{\partial\xi_1\cdots\partial\xi_d} \in L_2$.

First general QMC construction: Digital nets (Sobol 69, Niederreiter 87)

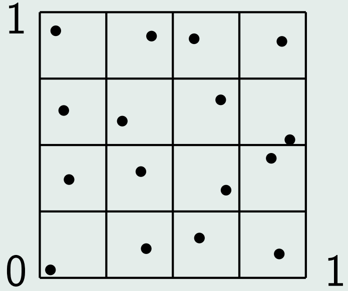
Elementary subintervals E in base b:

$$E = \prod_{j=1}^{d} \left[\frac{a_j}{b^{d_j}}, \frac{a_j + 1}{b^{d_j}} \right),$$

where $a_i, d_i \in \mathbb{Z}_+, 0 \leq a_i < b^{d_i}, i = 1, \ldots, d$.

Let $m, t \in \mathbb{Z}_+$, $m > t$. A set of b^m points in $[0,1)^d$ is a (t, m, d)-net in base b if every elementary subinterval E in base b with $\lambda^d(E) = b^{t-m}$ contains b^t points.

Illustration of a $(0, 4, 2)$-net with $b = 2$



A sequence (ξ^i) in $[0,1)^d$ is a (t, d)-sequence in base b if, for all integers $k \in \mathbb{Z}_+$ and $m > t$, the set

$$\{\xi^i : kb^m \leq i < (k+1)b^m\}$$

is a (t, m, d)-net in base b.

Theorem: (Niederreiter 92)
For fixed $d > 4$ and $b \in \mathbb{N}$, $b \geq 2$, there exists a constant $A(b,d)$ such that the star-discrepancy of a (t,m,d)-net $\{\xi^1, \ldots, \xi^n\}$ in base b with $m > 0$ satisfies

$$D_n^*(\xi^1, \ldots, \xi^n) \leq A(b,d)b^t \frac{(\log n)^{d-1}}{n} + O\Big(\frac{b^t (\log n)^{d-2}}{n}\Big).$$

Special cases: Sobol', Faure, Niederreiter and Niederreiter-Xing sequences.

Second general QMC construction: Lattices (Korobov 59, Sloan-Joe 94)

(Rank-1) lattice rules: Let $g \in \mathbb{Z}^d$ and consider the lattice points

$$\{\xi^i = \{\tfrac{i}{n} g\} : i = 1, \ldots, n\},$$

where $\{z\}$ is defined as *componentwise fractional part* of $z \in \mathbb{R}_+$, i.e., $\{z\} = z - \lfloor z \rfloor \in [0,1)$.

The generator $g \in \mathbb{Z}^d$ is chosen such that the (rank-1) lattice rule has good convergence properties. Such lattice rules may achieve better convergence rates $O(n^{-k+\delta})$, $k \in \mathbb{N}$, for integrands in C^k.

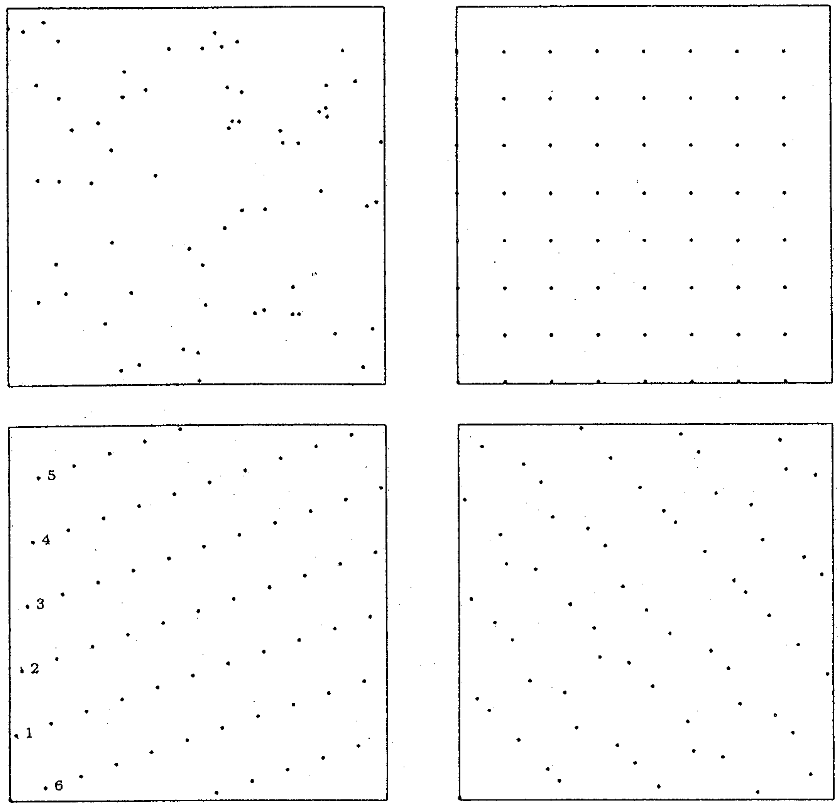


Fig. 5.3 Four different point sets with $n = 64$: random (top left), rectangular grid (top right), Korobov lattice (bottom left), and Sobol' (bottom right).

Recent development: Randomized lattice rules.

Randomly shifted lattice points:
If $\triangle$ is a sample from the uniform distribution in $[0,1]^d$, put

$$Q_{n,d}(f) = \frac{1}{n}\sum_{i=1}^{n} f\big(\{\tfrac{i-1}{n} g + \triangle\}\big).$$

Theorem:
Let n be prime, $f \in \mathbb{F}_d = \mathcal{W}^{(1,\ldots,1)}_{2,\gamma,\mathrm{mix}}([0,1]^d)$ (with $\gamma_j > 0$ denoting the weight of component j in the norm of $\mathbb{F}_d$).
Then $g \in \mathbb{Z}^d$ can be constructed componentwise such that for any $\delta \in (0, \frac{1}{2}]$ there exists a constant $C(\delta) > 0$ such that the mean worst-case quadrature error attains the optimal convergence rate

$$\hat{e}(Q_{n,d}) \le C(\delta) n^{-1+\delta},$$

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

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

(Sloan/Wožniakowski 98, Sloan/Kuo/Joe 02, Kuo 03)

Quadrature rules with sparse grids

Again we consider the unit cube $[0,1]^d$ in $\mathbb{R}^d$. Let a sequence of nested grids in $[0,1]$ be given, i.e.,

$$\Xi^i = \{\xi_1^i, \ldots, \xi_{m_i}^i\} \subset \Xi^{i+1} \subset [0,1] \quad (i \in \mathbb{N}),$$

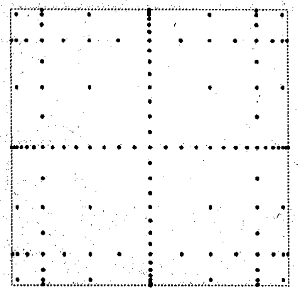
for example, the dyadic grid

$$\Xi^i = \left\{ \frac{j}{2^i} : j = 0, 1, \ldots, 2^i \right\} \quad (i \in \mathbb{N}).$$

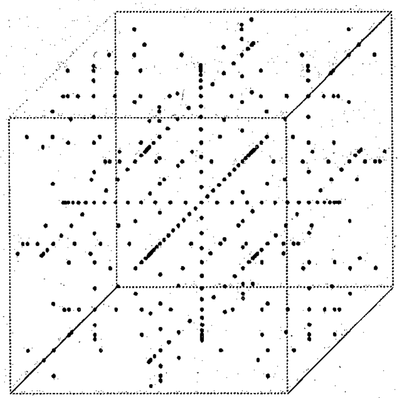
Then the point set in $[0,1]^d$ suggested by Smolyak (Smolyak 63) is

$$H(q,d) := \bigcup_{\sum_{j=1}^d i_j = q} \Xi^{i_1} \times \cdots \times \Xi^{i_d} \qquad (q \in \mathbb{N})$$

and called a sparse grid in $[0,1]^d$. Let $n = n(q,d)$ denote the number of points in $[0,1]^d$. In case of dyadic grids in $[0,1]$ the set $H(q,d)$ consists of all d-dimensional dyadic grids with product of mesh sizes given by $\frac{1}{2^q}$.



(a) $d = 2$



(b) $d = 3$

The corresponding tensor product quadrature rule for $n \geq d$ on $[0,1]^d$ (with the Lebesgue measure λ^d) is of the form

$$Q_{n(q,d),d}(f) = \sum_{q-d+1 \leq |\mathbf{i}| \leq q} (-1)^{q-|\mathbf{i}|} \binom{d-1}{q-|\mathbf{i}|} \sum_{j_1=1}^{m_{i_1}} \cdots \sum_{j_d=1}^{m_{i_d}} f(\xi_{j_1}^{i_1}, \ldots, \xi_{j_d}^{i_d}) \prod_{l=1}^{d} a_{j_l}^{i_l},$$

where $|\mathbf{i}| = \sum_{l=1}^{d} i_l$, $n(q,d)$ is the number of quadrature knots and the coefficients $a_j^{i_l}$ ($j = 1, \ldots, m_l$, $l = 1, \ldots, d$) are weights of d one-dimensional quadrature rules

$$\int_0^1 f(\xi) d\xi \approx Q^l(f) = \sum_{j=1}^{m_l} a_j^{i_l} f(\xi_j^{i_l}) \quad (l = 1, \ldots, d).$$

The weights are denoted by w_k, $k = 1, \ldots, n(q,d)$, and with a bijective mapping

$$\{\xi^k : k = 1, \ldots, n(q,d)\} \leftrightarrow \left\{(\xi_{j_1}^{i_1}, \ldots, \xi_{j_d}^{i_d}) : j_l = 1, \ldots, m_{i_l}, q-d+1 \leq |\mathbf{i}| \leq q\right\}$$

the tensor product quadrature rule $Q_{n(q,d),d}(f)$ may be rewritten as

$$Q_{n(q,d),d}(f) = \sum_{k=1}^{n(q,d)} w_k f(\xi^k).$$

Even if the one-dimensional weights are positive, some of the weights w_k may become negative. Hence, an interpretation as discrete probability measure is no longer possible.

Example: Consider the classical Clenshaw-Curtis rule Q^i with $m_1 = 1$, $m_i = 2^{i-1} + 1$, $i = 2, \ldots, d$, $\xi_1^1 = 0$ and

$$\xi_j^i = \tfrac{1}{2}\big(1 - \cos \tfrac{\pi(j-1)}{m_i - 1}\big) \quad (j = 1, \ldots, m_i,\ i = 2, \ldots, d)$$

and the weights a_j^i, $j = 1, \ldots, m_i$, be defined such that Q^i is exact for all univariate polynomials of degree at most m_i, $i = 1, \ldots, d$ (Novak-Ritter 96).

Proposition: $\|Q_{n(q,d),d}\|_\infty \leq c_d (\log n(q,d))^{d-1}$ for some $c_d > 0$ and fixed d.

Theorem: (Bungartz-Griebel 04)
If f belongs to $\mathbb{F}_d = \mathcal{W}_{2,\mathrm{mix}}^{(r,\ldots,r)}([0,1]^d)$, it holds

$$\left| \int_{[0,1]^d} f(\xi) d\xi - \sum_{k=1}^{n} w_k f(\xi^k) \right| \leq C_{r,d} \|f\|_d \frac{(\log n)^{(d-1)(r+1)}}{n^r}.$$

Conclusions

- High-dimensional numerical integration is a task met in a number of practical applications.
- Classical numerical integration fails for higher dimensions due to the curse of dimensionality.
- Alternatives are Monte Carlo, Quasi-Monte Carlo and sparse grid techniques.
- Monte Carlo methods are general, its convergence rate does not depend upon the dimension, but the convergence is slow.
- Classical Quasi-Monte Carlo methods converge faster than Monte Carlo schemes, but the convergence rate becomes effective only for $n \geq e^d$.
- Recently developed randomized lattice rules **lift the curse of dimensionality** and converge significantly faster than Monte Carlo.
- Sparse grid methods converge fast for (very) smooth functions.

Part II: Quasi-Monte Carlo methods and their recent developments

Wednesday, May 22, 2 pm.

Part III: QMC algorithms for solving stochastic optimization problems: Challenges and solutions

Thursday, May 23, 2 pm.

References

Monte Carlo:

P. L'Ecuyer: Uniform random number generation, *Annals of Operations Research* 53 (1994), 77–120.

P. L'Ecuyer: Uniform random number generation, in: *Simulation* (S. G. Henderson, B. L. Nelson eds.), Handbooks in Operations Research and Management Science, Elsevier, Amsterdam, 2006, 55–81.

L. Devroye: Non-uniform random variate generation, in: *Simulation* (S. G. Henderson, B. L. Nelson eds.), Handbooks in Operations Research and Management Science, Elsevier, Amsterdam, 2006.

L. Devroye: *Non-uniform random variate generation*, Springer-Verlag, New York, 1986 (2008).

P. Glasserman: *Monte-Carlo Methods in Financial Engineering*, Springer, New York, 2003.

M. Matsumoto, T. Nishimura: Mersenne Twister: A 623-dimensionally equidistributed uniform pseudo-random number generator, *ACM Transactions on Modeling and Computer Simulation* 8 (1998), 3–30.

G. Marsaglia: Random numbers fall mainly in the planes, *Proc. Nat. Acad. Sci.* 61 (1968), 25–28.

G. Marsaglia, A. Zaman: Monkey tests for random number generators, *Comp. Math. Applic.* 26, 9 (1993), 1–10.

G. Marsaglia, J. C.W. Marsaglia, A. Zaman: Rapid evaluation of the inverse of the normal distribution function, *Statistics & Probability Letters* 19 (1994), 259–266.

A. Genz, F. Bretz: *Computation of Multivariate Normal and t Probabilities*, Lecture Notes in Statistics 195, Springer-Verlag, New York, 2009.

A. Shapiro: Monte Carlo sampling methods, in: *Stochastic Programming* (A. Ruszczyński, A. Shapiro eds.), Handbooks in Operations Research and Management Science, Volume 10, Elsevier, Amsterdam 2003, 353–425.

Quasi-Monte Carlo:

I. M. Sobol': *Multidimensional Quadrature Formulas and Haar Functions*, Nauka, Moscow, 1969 (in Russian).

H. Niederreiter: *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM, Philadelphia, 1992.

I. H. Sloan, S. Joe: *Lattice Methods for Multiple Integration*, Clarendon Press, Oxford 1994.

F. Y. Kuo, I. H. Sloan: Lifting the curse of dimensionality, *Notices of the AMS* 52 (2005), 1320–1328.

C. Lemieux: *Monte Carlo and Quasi-Monte Carlo Sampling*, Springer, 2009.

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

F. Y. Kuo, Ch. Schwab, I. H. Sloan: Quasi-Monte Carlo methods for high-dimensional integration: the standard (weighted Hilbert space) setting and beyond, ANZIAM Journal 53 (2011), 1–37.

J. Dick, F. Y. Kuo, I. H. Sloan: High-dimensional integration – the Quasi-Monte Carlo way, *Acta Numerica* (2014), 1–153.

I. H. Sloan, 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.

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.

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.

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

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

M. Griebel, F. Y. Kuo, I. H. Sloan: The smoothing effect of integration in $\mathbb{R}^d$ and the ANOVA decomposition, *Mathematics of Computation* 82 (2013), 383–400.

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

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.

S. S. Drew, T. Homem-de-Mello: Quasi-Monte Carlo strategies for stochastic optimization, *Proceedings of the 2006 Winter Simulation Conference*, IEEE, 2006, 774–782.

Sparse Grids:

H.-J. Bungartz, M. Griebel: Sparse Grids, *Acta Numerica* (2004), 1–123.

T. Gerstner, M. Griebel: Numerical integration using sparse grids, *Numerical Algorithms* 18 (1998), 209–232.
E. Novak, K. Ritter: High dimensional integration of smooth functions over cubes, *Numerische Mathematik* 75 (1996), 79–97.

S. A. Smolyak: Quadrature and interpolation formulas for tensor products of certain classes of functions, *Soviet Math. Dokl.* 4 (1963), 240–243.

G. W. Wasilkowski, H. Woźniakowski: Explicit cost bounds of algorithms for multivariate tensor product problems, *Journal of Complexity* 11 (1995), 1–56.