Are Quasi-Monte Carlo methods efficient for two-stage stochastic programs ?

W. Römisch

Humboldt-University Berlin

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

(H. Heitsch, I. H. Sloan)

Anniversary Workshop in honour of Georg Pflug, Vienna, September 9, 2011

Personal reminiscences

- First meeting with Georg in Köszeg, during August 1981,
- Similar mathematical interests during the last 20 years (empirical approximations in SP, use of probability metrics, scenario (tree) generation, risk measures, energy etc.),
- Cooperation during the last 10 years, joint project, visits, joint book in 2007.

Congratulations and many thanks Georg !

Introduction

The following are recent approaches to scenario generation in stochastic programming besides Monte Carlo (MC):

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

(b) Quasi-Monte Carlo (QMC) methods (Koivu-Pennanen 05).

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

While the justification of (a) may be based on stability analysis, there is almost no reasonable justification of applying (b) and (c) to two- and multi-stage models. The basic theoretical background for (b) and (c) is similar.

There is encouraging progress of the underlying theory and of the available computational experience of both methodologies during the last 10 years, in particular, in finance.

Known convergence rates: MC $O(n^{-\frac{1}{2}})$, (a) $O(n^{-\frac{1}{d}})$
(d dimension of random vector, n number of scenarios).

Quasi-Monte Carlo methods

We consider the numerical integration of (Riemann) integrable functions f over the unit cube $[0,1]^d$ in $\mathbb{R}^d$. The approximate computation of

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

by a Quasi-Monte Carlo (QMC) algorithm $Q_{n,d}$ means

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

where the points ξ^i, $i = 1, \ldots, n$, belong to $[0,1]^d$.

We assume that f belongs to a linear normed space $\mathbb{F}_d$ with norm $\| \cdot \|_d$ and unit ball $\mathbb{B}_d$. The worst-case error of $Q_{n,d}$ over $\mathbb{B}_d$ is

$$e(Q_{n,d}) = \sup_{f \in \mathbb{B}_d} |I_d(f) - Q_{n,d}(f)|$$

and for $n = 0$ we formally set $Q_{0,d} = 0$.

Usally the smallest $n = n_{\min}(\varepsilon, d, \{Q_{n,d}\}) \in \mathbb{N}$ is considered such that

$$e(Q_{n,d}) \leq \varepsilon\, e(Q_{0,d}) = \varepsilon\, \|I_d\|,$$

holds for every $\varepsilon \in (0,1)$.

A family $\{Q_{n,d}\}$ of QMC algorithms is called tractable if there exist nonnegative constants C, q and p such that

$$n_{\min}(\varepsilon, d, \{Q_{n,d}\}) \leq C\, d^q \varepsilon^{-p}$$

holds for every $\varepsilon \in (0,1)$. Of course, $q = 0$ is desirable.

Example of $\mathbb{F}_d$: Tensor product Sobolev space

$$\mathbb{F}_d = W^{(s,\dots,s)}_{r,\text{mix}}([0,1]^d) = \bigotimes_{i=1}^{d} W^s_r([0,1]) \quad (s \geq 1,\ 1 \leq r \leq \infty).$$

contains all functions for which weak partial derivatives of order s exist with respect to each variable. For $s = 1$ the partial derivative

$$\frac{\partial^d f(\xi)}{\partial \xi_1 \cdots \partial \xi_d}$$

has to exist (in the sense of Sobolev).

Classical QMC results

For each $n \in \mathbb{N}$ and $\xi_i \in [0,1]^d$, $i = 1, \ldots, n$, the star-discrepancy is considered

$$D^*_{n,d}(\xi^1, \ldots, \xi^n) = \sup_{\xi \in [0,1]^d} \left| \lambda^d([0,\xi)) - \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{[0,\xi)}(\xi^i) \right| .$$

Theorem: There exist sequences $\{\xi_i\}_{i \in \mathbb{N}}$ such that

$$D^*_{n,d}(\xi^1, \ldots, \xi^n) = O(n^{-1}(\log n)^{d-1}) = O(n^{-1+\delta}) \quad (\forall \delta \in (0, 1/2]).$$

However, the leading coefficients depend on d and increase with d even for the best known sequences by Sobol, Faure and Niederreiter.

Theorem: (Koksma-Hlawka 1961)
If f is of bounded variation in the sense of Hardy and Krause, it holds for any d and n belonging to $\mathbb{N}$

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

Integrands in linear two-stage stochastic programming

Two-stage linear stochastic programs with random right-hand sides:

$$\min\left\{\langle c,x\rangle+\int_{\Xi}\Phi(\xi-Tx)P(d\xi):x\in X\right\}$$

where $c\in\mathbb{R}^m$, X is a polyhedral subset of $\mathbb{R}^m$, Ξ a closed subset of $\mathbb{R}^d$, T a (r,m)-matrix, P a Borel probability measure on Ξ and

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

where $q\in\mathbb{R}^{\bar{m}}$, W a $(r,\bar{m})$-matrix (having rank r) and t varies in the polyhedral cone $W(\mathbb{R}^{\bar{m}})$. There exist vertices v^j of $\mathcal{D}$ and polyhedral cones $\mathcal{K}_j$, $j=1,\ldots,\ell$, decomposing $\mathrm{dom}\,\Phi$ such that $\Phi(t)=\langle v^j,t\rangle$, $\forall t\in\mathcal{K}_j$, and $\Phi(t)=\max_{j=1,\ldots,\ell}\langle v^j,t\rangle$. Hence, the integrands are of the form

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

Problem: f is not of bounded variation in the HK sense.

Multivariate integration by randomly shifted lattice rules

Let $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_d > 0$ and set $\gamma_u = \prod_{j\in u} \gamma_j$.

Theorem: (weighted Koksma-Hlawka inequality)
If all partial derivatives of f exist and are continuous, it holds

$$|I_d(f) - Q_{n,d}(f)| \leq D_{n,d,\gamma}(\xi^1, \ldots, \xi^n)\|f\|_{d,\gamma},$$

where $\mathrm{disc}(\xi) = |\lambda^d([0,\xi)) - \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{[0,\xi)}(\xi^i)|$,

$$D_{n,d,\gamma}(\xi^1, \ldots, \xi^d) = \left(\sum_{\emptyset \neq u \subset D} \gamma_u \int_{[0,1]^{|u|}} \mathrm{disc}^2(\xi^u, 1^{-u}) d\xi^u\right)^{\frac{1}{2}}$$

is the weighted L_2-discrepancy and $\|f\|_{d,\gamma}^2 = \langle f, f\rangle_{d,\gamma}$ with

$$\langle f, g\rangle_{d,\gamma} = \sum_{u\subset D} \gamma_u^{-1} \int_{[0,1]^{|u|}} \frac{\partial^{|u|}}{\partial \xi^u} f(\xi^u, 1^{-u}) \frac{\partial^{|u|}}{\partial \xi^u} g(\xi^u, 1^{-u}) d\xi^u.$$

Weighted tensor product Sobolev space:

$$F_{d,\gamma} = \left\{ f \in W_{2,\mathrm{mix}}^{(1,\ldots,1)}([0,1]^d) : \|f\|_{d,\gamma} < \infty \right\}$$

Theorem: (Sloan-Wožniakowski 98)
There exist QMC algorithms satisfying $n_{\min}(\varepsilon, d, \{Q_{n,d}\}) \leq C\,\varepsilon^{-p}$ with $p \in [1, 2]$ on $F_{d,\gamma}$ iff $\sum_{j=1}^{\infty} \gamma_j < \infty$.

Randomly shifted rank-1 lattice rules:

$$Q_{n,d}(f) = \frac{1}{n} \sum_{k=0}^{n-1} f\left(\left\{\frac{kz}{n} + \triangle\right\}\right),$$

where $z \in \mathbb{Z}^d$, $\{x\}$ means componentwise the fractional part of x and $\triangle$ is a uniformly distributed random variable in $[0, 1]^d$.

Theorem: (Sloan-Kuo-Joe 02, Kuo 03)
Let n be prime. Then $z \in \mathbb{Z}^d$ can be constructed component-by-component such that for every $0 < \delta \leq \frac{1}{2}$

$$e(Q_{n,d}) \leq C_d(\delta) n^{-1+\delta} \|I_d\|$$

holds on $F_{d,\gamma}$ for some $C_d(\delta)$.
The constant $C_d(\delta)$ may be chosen to be independent on d if

$$\sum_{j=1}^{\infty} \gamma_j^{\frac{1}{2(1-\delta)}} < \infty.$$

The construction and convergence results are extended from $[0,1]^d$ to $\mathbb{R}^d$ and probability densities on $\mathbb{R}^d$ of the form

$$\rho_d(\xi) = \prod_{j=1}^{d} \rho(\xi_j) \quad (\xi \in \mathbb{R}^d)$$

by the transformation $\Phi^{-1}(\xi^i)$, $i = 1, \ldots, n$, where Φ is a mapping from $\mathbb{R}^d$ to $[0,1]^d$ given by

$$\Phi(u) = (\phi(u_1, \ldots, \phi(u_d)),$$

and ϕ is the distribution function of the density ρ.

(Kuo-Sloan-Wasilkowski-Waterhouse 10)

ANOVA decomposition of multivariate functions

Idea: If f isn't of bounded variation or smooth, decompositions of f may be used, where only some of the terms are relevant and, hopefully, are of bounded variation or smooth.

Let $D = \{1, \ldots, d\}$ and $f \in L_{1,\rho_d}(\mathbb{R}^d)$. The projection P_k, $k \in D$, is defined by

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

Clearly, the function $P_k f$ 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, namely, P_u is linear and it holds $P_u^2 = P_u$.

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 $P_{-u}(f)$.

Proposition:
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

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

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 f_u can be smoother than f under certain conditions.

ANOVA decomposition of integrands in two-stage models

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 $\rho_d(\xi) = \prod_{j=1}^d \rho(\xi_j)$ ($\xi \in \mathbb{R}^d$).

(A1) and (A2) imply that $\mathcal{D}$ 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.,

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

It holds that $\cup_{j=1,\ldots,\ell} \mathcal{K}_j = \mathrm{dom}\,\Phi$ 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. In the latter case, the intersection 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 $\xi_i \in \mathbb{R}$, $i = 1, \dots, d$, $i \neq k$, be given. We set $\xi^k = (\xi_1, \dots, \xi_{k-1}, \xi_{k+1}, \dots, \xi_d)$ and

$$\xi_s = (\xi_1, \dots, \xi_{k-1}, s, \xi_{k+1}, \dots, \xi_d) \in \operatorname{dom} \Phi = \cup_{j=1,\dots,\ell} \mathcal{K}_j.$$

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 continuously differentiable if the one-dimensional density ρ is continuous. $P_k f$ is in C^∞ if $\rho \in C^\infty(\mathbb{R})$.

Theorem:
Let $u \subset D$. Assume (A1)–(A4) and that all adjacent vertices of $\mathcal{D}$ have different components.
Then the ANOVA term f_u is infinitely differentiable if $\rho \in C^\infty(\mathbb{R})$.

Example:

Let $\bar{m} = 3$, $d = 2$, $\Xi = \mathbb{R}^2$, 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 $\mathbb{R}^2$)

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

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

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

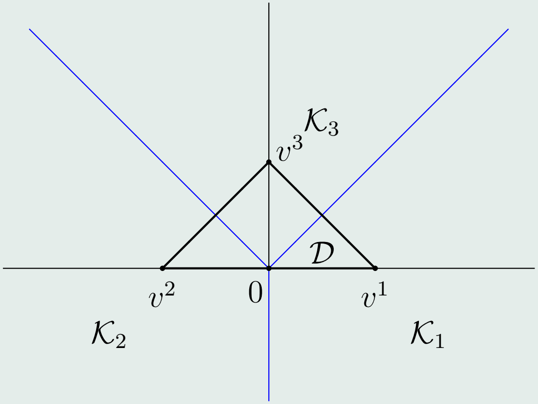

Figure 1: Illustration of $\mathcal{D}$, its vertices v^j and the normal cones $\mathcal{K}_j$ to its vertices

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

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

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.

Open problem: Truncation dimension of linear two-stage stochastic programs ?

Appendix: Functions of bounded variation

Let $D = \{1, \ldots, d\}$ and we consider subsets u of D with cardinality $|u|$. By $-u$ we mean $-u = D \setminus u$.
The expression ξ^u denotes the $|u|$-tuple of the components ξ_j, $j \in u$, of $\xi \in \mathbb{R}^d$. For example, we write

$$f(\xi) = f(\xi^u, \xi^{-u}).$$

We set the d-fold alternating sum of f over the d-dimensional interval $[a, b]$ as

$$\triangle(f; a, b) = \sum_{u \subseteq D} (-1)^{|u|} f(a^u, b^{-u}).$$

Furthermore, we set for any $v \subseteq u$

$$\triangle_u(f; a, b) = \sum_{v \subseteq u} (-1)^{|v|} f(a^v, b^{-v}).$$

Let G_j denote finite grids in $[a_j, b_j)$, $a_j < b_j$, $j = 1, \ldots, d$, and $G = \times_{i=1}^{d} G_j$ a grid in $[a, b) = \times_{i=1}^{d}[a_j, b_j)$. For $g \in G$ let $g^+ = (g_1^+, \ldots, g_d^+)$, where g_j^+ is the successor of g_j in $G_j \cup \{b_j\}$. Then the variation of f over G is

$$V_G(f) = \sum_{g \in G} \left|\triangle(f; g, g^+)\right| .$$

If $\mathcal{G}$ denotes the set of all finite grids in $[a, b)$, the variation of f on $[a, b]$ in the sense of Vitali is

$$V_{[a,b]}(f) = \sup_{G \in \mathcal{G}} V_G(f) .$$

The variation of f on $[a, b]$ in the sense of Hardy and Krause is

$$V_{\mathrm{HK}}(f; a, b) = \sum_{u \subset D} V_{[a^{-u}, b^{-u}]}(f(\xi^{-u}, b^u)) .$$

Bounded variation on $[a, b]$ in the sense of Hardy and Krause then means $V_{\mathrm{HK}}(f; a, b) < \infty$.

Proposition: (Owen 05)

Let $d \geq 3$, $b_i \in \mathbb{R}$, $i = 0, 1, \ldots, d$, and we consider for $\xi \in [0,1]^d$

$$f(\xi) = \max\{\langle b, \xi \rangle - b_0, 0\}.$$

If $\{\xi \in [0,1]^d : \langle b, \xi \rangle = b_0\}$ has positive $(d-1)$-dimensional volume and none of $b_1, \ldots, b_d$ is zero, it holds $V_{\mathrm{HK}}(f) = \infty$.

References

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

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

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.

V. Bally, G. Pagès, J. Printems: First order schemes in the numerical quantization method, *Math. Finance* 13 (2003), 1–16.

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 (to appear).

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.

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

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

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 and 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 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, manuscript, University of New South Wales, December 2010.

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.

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.

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.

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.