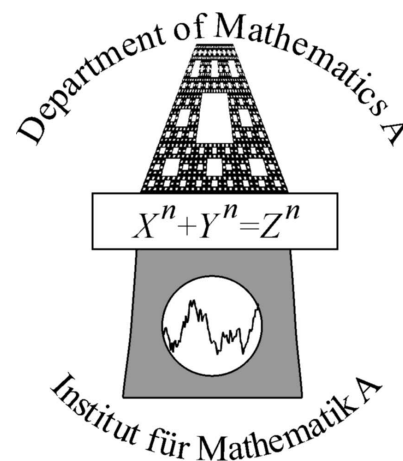


Transformation methods for the creation of non-uniformly distributed low-discrepancy sequences

Reinhold Kainhofer (reinhold@kainhofer.com)

Jürgen Hartinger (hartinger@finanz.math.tu-graz.ac.at)



Institut für Mathematik

Graz University of Technology

Steyrergasse 30, A-8010 Graz, Austria

Many problems from finance and engineering need sequences that follow a distribution other than the uniform distribution.

How can we generate them for QMC?

Methods known from Monte Carlo:

- **Inversion method**, using the quantile function of the distribution.
Preserves the discrepancy.
- **Acceptance-Rejection** method
- **Smoothing** method:
Adaption of acceptance-rejection for QMC (Wang)
Integrand with A-R is changed/modified to perform better, integral value stays the same

- Inversion:
Problem: Usually the quantile function is not explicitly known.
- Acceptance-Rejection: **unsuited for QMC**
 - ◆ introduces jumps in the integrand function \Rightarrow unbounded variation
 - ◆ lots of points are thrown away in high dimensions
- Smoothing:
Problem: Only applicable for integration problems, does not generate a sequence. For path simulation the h -distributed sequences are explicitly needed

- **Star-discrepancy** of a sequence:

$$D_N \left((\mathbf{x}_i)_{i \in \mathbb{N}} \right) = \sup_{J=[0, \mathbf{z}) \subset [0, 1)^s} \left| \frac{\#(\mathbf{x}_i, J)}{N} - \lambda_s(J) \right|$$

- **Low-discrepancy sequence:** $D_N((\mathbf{x}_i)) = \mathcal{O}\left(\frac{\log^s N}{N}\right)$

- **h -Discrepancy** of sequence $(\mathbf{x}_i)_{i \in \mathbb{N}}$:

$$D_{h,N} \left((\mathbf{x}_i)_{i \in \mathbb{N}} \right) = \sup_{J=[0, \mathbf{z}) \subset [0, 1)^s} \left| \frac{\#(\mathbf{x}_i, J)}{N} - H(J) \right|$$

- **L^p -Discrepancy** of sequence $(x_i)_{i \in \mathbb{N}}$:

$$D^{(p)} \left((\mathbf{x}_i)_{i \in \mathbb{N}} \right) = \left(\int_{[0, 1]^s} \left(\frac{\#(\mathbf{x}_i, [0, \mathbf{z}))}{N} - \lambda_s([0, \mathbf{z})) \right)^p d\mathbf{z} \right)^{\frac{1}{p}}$$

In 1972 Hlawka and Mück propose a **method to generate h -distributed (double) sequences.**

For $f(\mathbf{x}) = f_1(x_1) \cdot \dots \cdot f_s(x_s)$ simpler construction (Hlawka, '98):

Let $\omega = (x_i)_{i \in \mathbb{N}} \sim U(0, 1)$ with discrepancy $D_N(\omega)$, and $f(x_1, \dots, x_s)$ an s -dimensional density function. Define $M_f = \sup f(\mathbf{x})$. Then $\tilde{\omega}$ with

$$\tilde{y}_{k,j} = \frac{1}{N} \sum_{r=1}^N [1 + x_{k,j} - F_j(x_{r,j})] = \frac{1}{N} \sum_{r=1}^N \chi_{[0, x_{k,j}]}(F_j(x_{r,j}))$$

has an F -discrepancy of

$$D_{N,F}(\tilde{\omega}) \leq (1 + 3mM_f)D_N(\omega) .$$

The Hlawka-Mück method has two serious practical problems:

- **Effort:** $\mathcal{O}(N^2)$ for generation

Each y_k need the evaluation of $\sum_{r=1}^N \dots$

-) The order cannot be easily reduced

The evaluation of $F(x_i)$ can be done before.

- If you change N , the whole point set needs to be **regenerated**.

- Transformed sequence contains **identical points**

All points are transformed to the grid $\frac{k}{N}$, $0 \leq k \leq N$.

-) Can be solved, e.g. by our adaption, by using a continuous approximation to $F(x)$ instead of the step function.

Inversion, approximate distribution function F by continuous, piecewise linear function.

One-dimensional: Let $\omega = (x_i)_{i \in \mathbb{N}}$ a uniformly distributed sequence. Define $x_k^- = \max_{x_i} (F(x_i) < x_k)$ and $x_k^+ = \min_{x_i} (F(x_i) > x_k)$. Then for the the sequence generated by

$$y_k = \frac{F(x^+) - x_k}{F(x^+) - F(x^-)} x^- + \frac{x_k - F(x^-)}{F(x^+) - F(x^-)} x^+$$

it holds that $|y_k - F^{-1}(x_k)| < D_N(\omega)$ and $|F(y_k) - x_k| < M D_N(\omega)$. Consequently, $D_N(\tilde{\omega}) \leq (1 + M) D_N(\omega)$.

For s -dimensional distributions with $f = f_1 \dots f_s$ the bound is $D_N(\tilde{\omega}) \leq (1 + 2M)^s D_N(\omega)$.

Other possible modifications (approx., support pts)

- Why do we use the sequence to approximate $F(x)$?
 Quality of approximation needs to increase with N .
- Why don't we use $\frac{k}{N}$ as support points for the empirical distribution function?
 Problem: If we change N , all support points need to be recalculated. With the H-M method (and our adaption), only the sum needs to be redone, not the $F(x_k)$.
 Also, no significant improvement in order of discrepancy (just a constant).
- Why not use the same support points for all dimensions?
 Quality of approximation and of the sequence decoupled.
 Discrepancy bound includes both discrepancies, asymptotics unchanged.

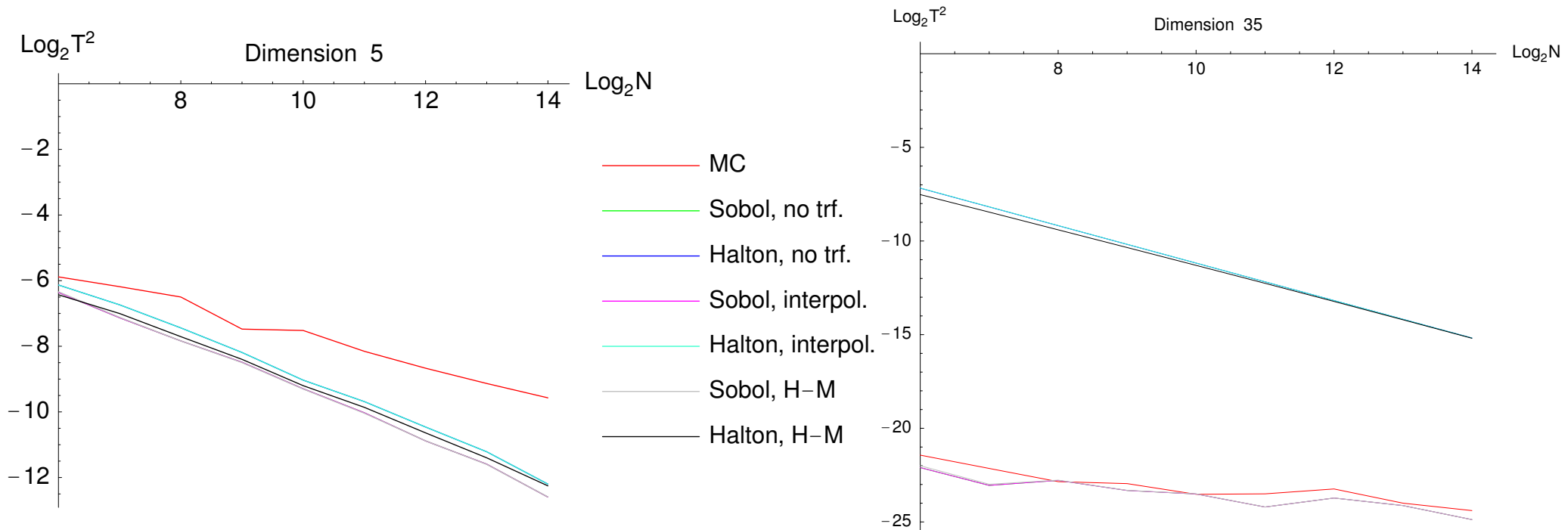
Integration error can also be **bounded by the $L^{(2)}$ discrepancy** (Sloan, Wozniakowski):

$$|I_d(f) - Q_{n,d}(f)| \leq \text{disc}^{(2)}((x_i)) \|f\|_d^2$$

with $\text{disc}^{(2)} = \left(\sum_{u \subset D} (T_u^{(2)})^2 \right)^{1/2}$ and $L^{(2)}$ -variation $\|f\|_d^2$.

$T^{(2)}$ can be **explicitly calculated**:

$$(T^{(2)})^2 = \frac{1}{N^2} \sum_{i,j=1}^N \prod_{k=1}^d (1 - \max(x_{ik}, x_{jk})) - \frac{2^{1-d}}{N} \sum_{i=1}^N \prod_{k=1}^d (1 - x_{ik}^2) - \frac{1}{3^d}$$



- Halton loses quality for high dimensions
- MC and Sobol have the same $L^{(2)}$ discrepancy in 50 dim
- H-M and our transformations do not affect discrepancy

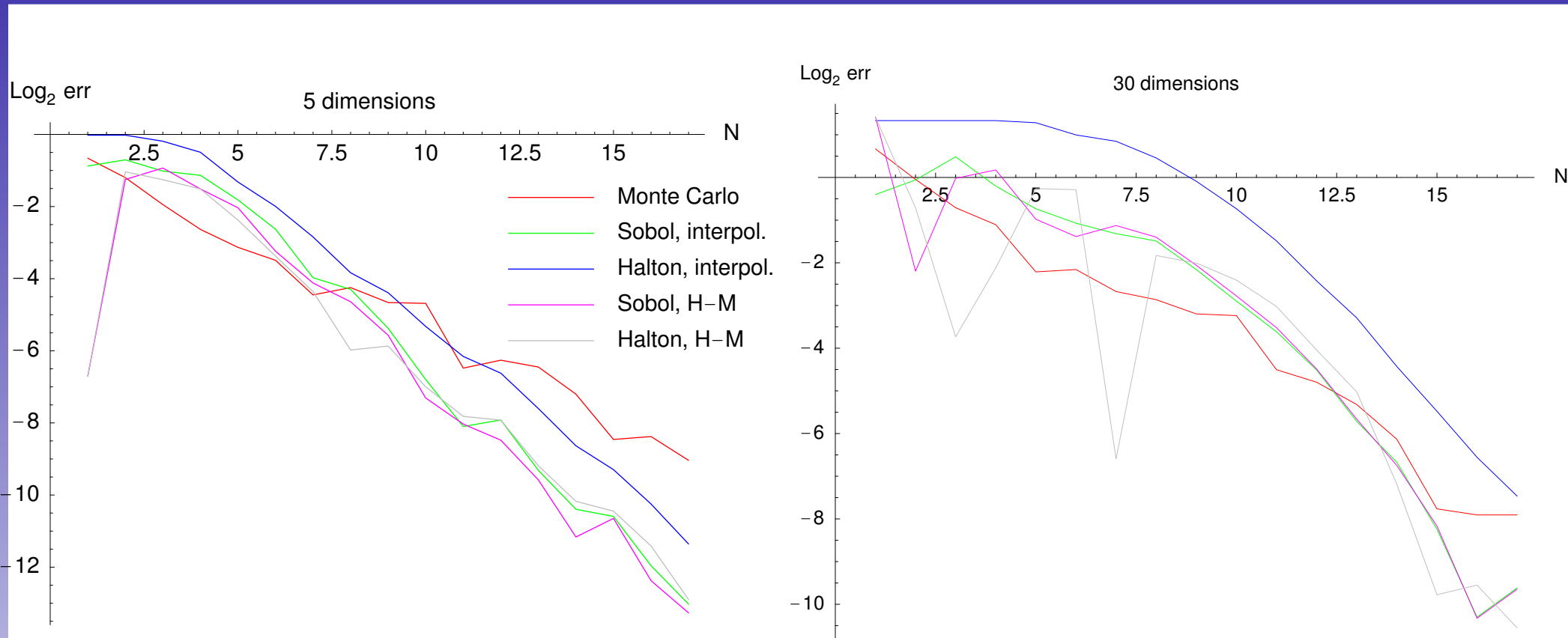
Asian option with payoff

$$P = \mathbb{E}_{NIG} \left(e^{-rT} \left(S_0 \sum_{i=1}^d \prod_{j=1}^i e^{x_j} - K \right)^+ \right)$$

When using inversion, integrand is unbounded ($V(f) = \infty$)!

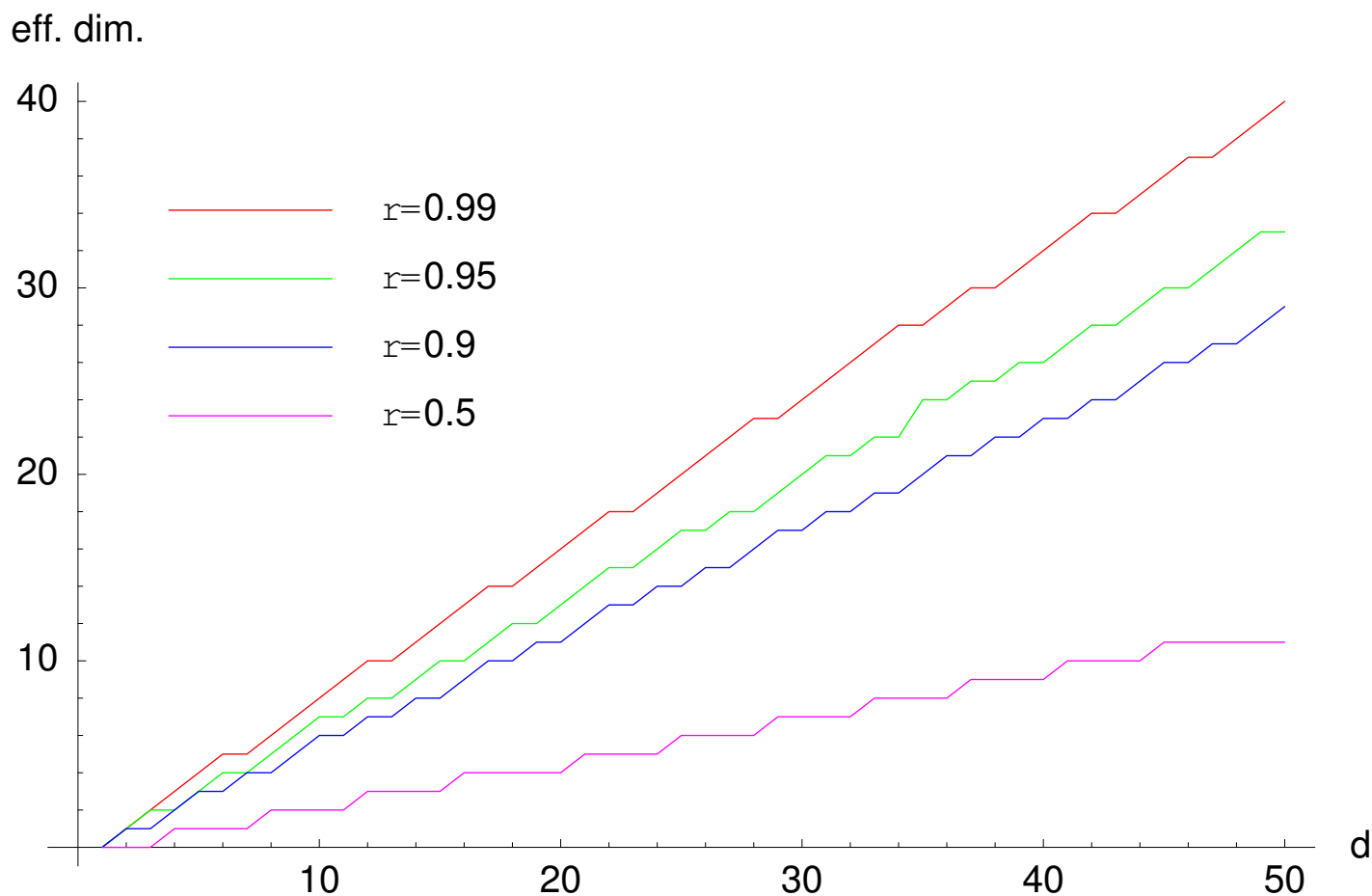
- Theorems by Sobol, and Hartinger/K./Tichy and Hartinger/K./Predota prove convergence.
- Only slightly worse convergence order than Koksma-Hlawka for bounded variation.

Asian options (50 dim): results



- QMC loose quality for high dimensions
- Halton with interpolation doesn't work
- Sobol works both with H-M and its interpolation adaption

Problem: high effective dimension



high effective dimension (e.g. $d=50$, eff. dim=41), no dimension reduction techniques (like BB, PCA) known for Asian options with NIG distribution