# QMC methods in quantitative finance, tradition and perspectives

Gunther Leobacher

Johannes Kepler University Linz (JKU)

WU Research Seminar

# What is the content of the talk

- Valuation of financial derivatives
- in stochastic market models
- using (QMC-)simulation
- and why it might be a good idea

FWF SFB "Quasi-Monte Carlo methods: Theory and Applications"

http://www.finanz.jku.at

# Derivative pricing

# Derivative pricing
## BS and SDE models

Black-Scholes model:

- Share: $S_t = S_0 \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t\right)$, $t \in [0, T]$,
  - $\mu$ is the (log)-drift
  - $\sigma$ is the (log)-volatility
- Bond: $B_t = B_0 \exp(rt)$, $t \in [0, T]$,
  - $r > 0$ is the interest rate

# Derivative pricing
## BS and SDE models

SDE-model ($m$-dimensional):

$$
\begin{aligned}
dS_t &= b(t, S_t)dt + a(t, S_t)dW_t\,,\ \ t \in [0, T],\\
S_0 &= s_0
\end{aligned}
$$

Black-Scholes model is special case of SDE models,
$dS_t = \mu S_t dt + \sigma S_t dW_t$
Other popular SDE-model:

$$
\begin{aligned}
dB_t &= rB_t dt\\
dS_t &= \mu S_t dt + \sqrt{V_t} S_t(\rho dW_t^1 + \sqrt{1-\rho^2} dW_t^2)\\
dV_t &= \kappa(\theta - V_t)dt + \xi \sqrt{V_t}\, dW_t^1\\
(B_0, S_0, V_0) &= (b_0, s_0, v_0)\,.
\end{aligned}
$$

for $t \in [0, T]$. “Heston model”

# Derivative pricing
## BS and SDE models

- A contingent claim is a contract that pays its owner an amount of money that depends on the evolution of the price processes
- more technically: a function that is $\mathscr{F}_T^S$-measurable (Information generated by price processes)

# Derivative pricing
## BS and SDE models

Examples:

- European Call option on $S^1$ with strike $K$ and maturity $T$ pays $\max(S^1_T - K, 0)$ at time $T$;
- Asian Call option on $S^1$ pays $\max\left(\frac{1}{T-T_0}\int_{T_0}^{T} S^1_t\,dt - K, 0\right)$ at time $T$;
- an example of a Basket option pays $\max\left(\frac{1}{d}\sum_{k=1}^{d} S^k_T - K, 0\right)$ at time $T$;
- much more complicated payoffs exist in practice.

# Derivative pricing
## Prices as expectations

Under technical "no arbitrage condition" and existence of a "riskless" asset $B = S^j$ that we may use as a numeraire we have a (not necessarily unique) price, the price at time $0$ of the claim $C$ with payoff $\phi$ at time $T$ can be written in the form

$$\pi_0(C) = \mathrm{E}_Q\left(B_0 B_T^{-1} \phi\right)$$

where $Q$ is a pricing measure.

Only in rare cases can this expected value be computed explicitely.

# Derivative pricing
## Prices as expectations

Assume a Black-Scholes modell and suppose we want to price an Arithmetic average option

$$\phi = \max\left(\frac{1}{d}\sum_{k=1}^{d} S_{\frac{k}{d}T} - K, 0\right)$$

that is, the derivative's payoff depends on $S_{\frac{T}{d}}, \ldots, S_T$.
Let us compute its value, $\pi_0(\phi) = \mathrm{E}_Q(\exp(-rT)\phi)$.

# Derivative pricing
## Prices as integrals

Share price at time $\frac{k}{n}T$ under pricing measure

$$S_{\frac{k}{d}T} = s_0 \exp\left(\left(r - \frac{\sigma^2}{2}\right)\frac{k}{d}T + \sigma W_{\frac{k}{d}T}\right)$$

$$\stackrel{d}{=} s_0 \exp\left(\left(r - \frac{\sigma^2}{2}\right)\frac{k}{d}T + \sigma\sqrt{\frac{1}{d}T}\sum_{j=1}^{k} Z_j\right),$$

where $Z_1, \ldots, Z_d$ are independent standard normals.

$$\begin{aligned}\pi_0(\phi) &= \mathrm{E}_Q\left(\psi(Z_1, \ldots, Z_d)\right) \\ &= \int_{\mathbb{R}^d} \psi(z_1, \ldots, z_d) \exp\left(-\frac{1}{2}(z_1^2 + \cdots + z_d^2)\right)(2\pi)^{-\frac{d}{2}} dz_1 \ldots dz_d\end{aligned}$$

where $\psi$ is some (moderately complicated) function in $d$ variables.

# Derivative pricing
## Prices as integrals

That is, the price of the claim can be calculated as a $d$-dimensional integral over $\mathbb{R}^d$.

# Derivative pricing

## Prices as integrals

Same argument can be made for SDE models and much simpler payoff.
Solve SDE using, for example, Euler-Maruyama method with $d$ steps:

$$\hat{S}_0 = s_0$$

$$\hat{S}_{(k+1)\frac{T}{d}} = \hat{S}_{k\frac{T}{d}} + \mu(\hat{S}_{k\frac{T}{d}}, \frac{k}{d}T)\frac{T}{d} + \sigma(\hat{S}_{k\frac{T}{d}}, \frac{k}{d}T)\sqrt{\frac{T}{d}}Z_{k+1}$$

$k = 1, \ldots, d$.
Means that $\hat{S}_T$ is a function of $Z_1, \ldots, Z_d$. Expectation over payoff is again an integral over $\mathbb{R}^d$.

# Derivative pricing

## Prices as integrals

### Remark

*Let $\Phi$ be the cumulative distribution function of the standard normal distribution, and let $\Phi^{-1}$ denote its inverse. Then*

$$\int_{\mathbb{R}^d} \psi(z_1, \ldots, z_d) \exp\left( -\frac{1}{2}(z_1^2 + \cdots + z_d^2)\right)(2\pi)^{-\frac{d}{2}} dz_1 \ldots dz_d$$
$$= \int_{(0,1)^d} \psi\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\right) du_1 \ldots du_d$$

# MC and QMC methods

# MC and QMC methods

## High-dimensional integration

Suppose $f:[0,1)^d \longrightarrow \mathbb{R}$ is integrable and we want to know

$$I = \int_{[0,1)^d} f(\mathbf{x}) d\mathbf{x} \,.$$

For small $d$ we may use product rules with $n$ nodes per coordinate. For example, set $x_k = \frac{k}{n}$ and consider the one-dimensional rule

$$\int_0^1 g(x)dx \approx \frac{1}{n}\sum_{k=0}^{n-1} g(x_k)$$

By Fubini's theorem

$$I = \int_{[0,1)^d} f(\mathbf{x}) d\mathbf{x} = \int_0^1 \dots \int_0^1 f(x_1, \dots, x_d) dx_1 \dots dx_d$$

and thus

$$I \approx \frac{1}{n} \sum_{k_1=0}^{n-1} \dots \frac{1}{n} \sum_{k_d=0}^{n-1} f(x_{k_1}, \dots, x_{k_d}) \, .$$

- doubling $n$ in the one-dimensional integration rule multplies the number of function evaluations in the product rule by $2^d$.
- calculation cost increases exponentially in required accuracy
- this is known as "Curse of dimension"

# MC and QMC methods

## Monte Carlo

Idea: $I = \mathrm{E}(f(U_1, \ldots, U_d))$, where $U_1, \ldots, U_d$ are independent uniform random variables.

Consider an independent sequence $(\mathbf{U}_k)_{k \geq 0}$ of uniform random vectors. Then

$$\mathrm{P}\left(\lim_{N \to \infty} \frac{1}{N} \sum_{k=0}^{N-1} f(\mathbf{U}_k) = I\right) = 1\,,$$

by the strong law of large numbers.

# MC and QMC methods
## Monte Carlo

If, in addition, $\sigma^2 := \mathrm{E}(f(U_1,\ldots,U_d)^2) - I^2 < \infty$, we have by Tchebychev's inequality for every $\varepsilon > 0$

$$\mathrm{P}\left(\left|\frac{1}{N}\sum_{k=0}^{N-1} f(\mathbf{U}_k) - I\right| > \varepsilon\right) \leq \frac{\sigma^2}{N\varepsilon^2}\,.$$

Suppose we want an error less than $\varepsilon$ with probability $1-\alpha$, or, equivalently, an error greater than $\varepsilon$ with probability $\alpha$.

$$\frac{\sigma^2}{N\varepsilon^2} \leq \alpha \Leftrightarrow N \geq \frac{\sigma^2}{\varepsilon^2\alpha}$$

The number of integration nodes grows (only) quadratically in $\frac{1}{\varepsilon}$.

# MC and QMC methods

## Quasi-Monte Carlo

For $\mathbf{a} = (a_1, \ldots, a_d) \in [0,1]^d$
let $[0,\mathbf{a}) := [0,a_1) \times \ldots \times [0,a_d)$.

**Definition (Discrepancy function)**

Let $\mathscr{P}_N = \{\mathbf{x}_0, \ldots, \mathbf{x}_{N-1}\} \in [0,1)^d$. Then the discrepancy function $\Delta_{\mathscr{P}_N} : [0,1]^d \longrightarrow \mathbb{R}$ is defined by

$$\Delta_{\mathscr{P}_N}(\mathbf{a}) := \frac{\#\{0 \le k < N : \mathbf{x}_k \in [0,\mathbf{a})\}}{N} - \lambda^d([0,\mathbf{a}))\,,\ (\mathbf{a} \in [0,1]^d)$$

$\lambda^d$ denotes $d$-dimensional Lebesgue measure

# MC and QMC methods

## Quasi-Monte Carlo

### Definition (Star discrepancy)

$\mathscr{P}_N = \{\mathbf{x}_0, \ldots, \mathbf{x}_{N-1}\} \in [0,1)^d$. Then the star discrepancy $D^*(\mathscr{P}_N)$ is defined by

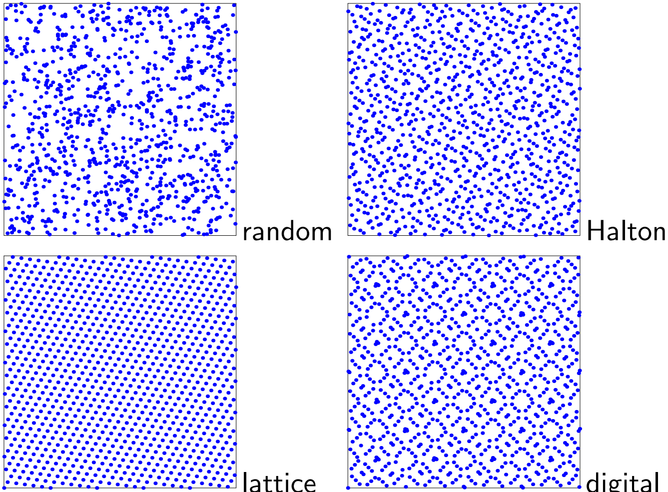
$$D^*(\mathscr{P}_N) := \sup_{\mathbf{a} \in [0,1]^d} |\Delta_{\mathscr{P}_N}(\mathbf{a})| = \|\Delta_{\mathscr{P}_N}\|_\infty$$

# MC and QMC methods
## Quasi-Monte Carlo

If $U_0, U_1, \ldots$ is a sequence of random vectors uniform in $[0,1)^d$, and $\mathscr{P}_N = \{U_0, \ldots, U_{N-1}\}$, then

$$\lim_{N\to\infty} D^*(\mathscr{P}_N) = 0 \qquad \text{a.s.}$$

"Low discrepancy sequences" are designed to have this convergence as fast as possible


random
Halton
lattice
digital

# MC and QMC methods
## Quasi-Monte Carlo

- For every dimension $d \geq 1$ there exist a sequence $(\mathbf{x}_n)_{n\geq 0}$ in $[0,1)^d$, such that $D^*(\mathscr{P}_N) = O\left(\frac{(\log N)^d}{N}\right)$, where $\mathscr{P}_N = \{\mathbf{x}_0, \ldots, \mathbf{x}_{N-1}\}$

  We call such a sequence with this property a low discrepancy sequence
- There exists a constant $c > 0$ such that for any sequence $\mathbf{x}_0, \mathbf{x}_2, \ldots \in [0,1)$ we have $\liminf_N D^*(\mathscr{P}_N) \geq c \frac{\log(N)^{\frac{d}{2}}}{N}$, where $\mathscr{P}_N = \{\mathbf{x}_0, \ldots, \mathbf{x}_{N-1}\}$

# MC and QMC methods

## Koksma-Hlawka inequality

Idea: let $\mathbf{a} \in [0,1)^d$. Let $(\mathbf{x}_k)_{k\geq 0}$ be a low-discrepancy sequence.

We have seen that for some $C$

$$\left|\frac{\#\{0 \leq k < N : \mathbf{x}_k \in [0,\mathbf{a})\}}{N} - \lambda^s([0,\mathbf{a}))\right| \leq C\frac{\log(N)^d}{N}$$

i.e.

$$\left|\frac{1}{N}\sum_{k=0}^{N-1} 1_{[0,\mathbf{a})}(\mathbf{x}_k) - \int_{[0,1)^d} 1_{[0,\mathbf{a})}(\mathbf{x})d\mathbf{x}\right| \leq C\frac{\log(N)^d}{N}$$

# MC and QMC methods
## Koksma-Hlawka inequality

We may have the hope that, for suitably behaved integrands, and a low discrepancy sequence $(\mathbf{x}_k)_{k\geq 0}$,

$$\left|\frac{1}{N}\sum_{k=0}^{N-1} f(\mathbf{x}_k) - \int_{[0,1)^d} f(\mathbf{x})d\mathbf{x}\right| \leq C\frac{\log(N)^d}{N}$$

(For large $N$ this convergence would be much faster than $N^{-\frac{1}{2}}$.)
The Koksma-Hlawka states that this is true indeed.

# MC and QMC methods

## Koksma-Hlawka inequality

**Theorem (Koksma-Hlawka inequality)**

*Let* $f : [0,1)^d \longrightarrow \mathbb{R}$ *and* $\mathscr{P} = \{\mathbf{x}_0, \ldots, \mathbf{x}_{N-1}\} \subseteq [0,1)^d$. *Then*

$$\left| \frac{1}{N} \sum_{k=0}^{N-1} f(\mathbf{x}_k) - \int_{[0,1)^d} f(\mathbf{x}) d\mathbf{x} \right| \leq V(f) D^*(\mathscr{P}),$$

*where* $V(f)$ *denotes the total variation of* $f$ *in the sense of Hardy and Krause.*

# MC and QMC methods

## Koksma-Hlawka inequality

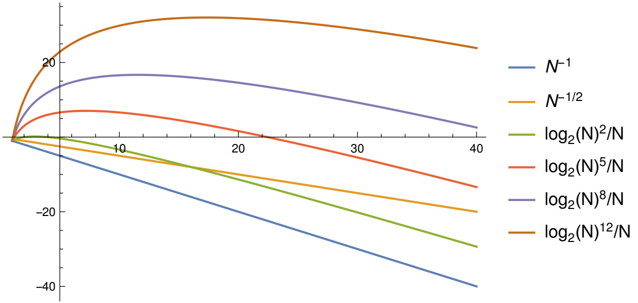
We have

$$V(f) = \sum_{\mathbf{u} \subseteq \{1,\ldots,d\}} \int_{[0,1]^{|\mathbf{u}|}} \left| \frac{\partial^{|\mathbf{u}|} f}{\partial x_{\mathbf{u}}}(x_{\mathbf{u}}, 1) \right| dx_{\mathbf{u}}$$

if the mixed derivatives of $f$ exist and are integrable.
Here, $(x_{\mathbf{u}}, 1)$ denotes the vector ones obtains by replacing coordinates with index not in $\mathbf{u}$ by 1
and $\frac{\partial^{|\mathbf{u}|} f}{\partial x_{\mathbf{u}}}$ means derivative by every variable with index in $\mathbf{u}$

# MC and QMC methods
## Koksma-Hlawka inequality

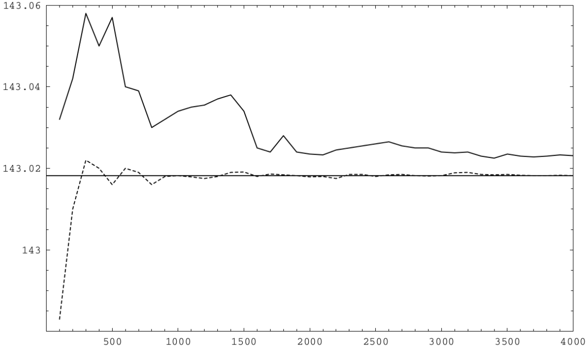
Double logarithmic plot:



# MC and QMC methods
## Koksma-Hlawka inequality

Suggests to use QMC for small to moderate dimensions only. However, in the late 20th century, starting with work by Paskov and Traub, "practitioners" started to observe the following phenomenon

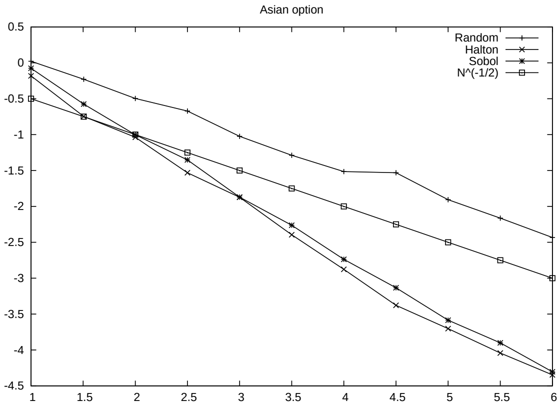
# MC and QMC methods
## Koksma-Hlawka inequality




Comparison between Monte Carlo (continuous) and Quasi-Monte Carlo (dotted) convergence in valuing a mortgage backed security

# MC and QMC methods

## Koksma-Hlawka inequality



# MC and QMC methods
## Koksma-Hlawka inequality

This phenomenon frequently occured in applications from mathematical finance, or, more concretely, in derivative pricing.

Where does this apparent superiority come from?

# Generation of Brownian paths

# Generation of Brownian paths
## Classical constructions

- Brownian motion is a process is continuous time
- For numerical computation one usually only needs the Brownian path at finitely many nodes $t_1, \dots, t_d$
- define a discrete Brownian path on nodes $0 < t_1 < \dots < t_d$ as Gaussian vector $(B_{t_1}, \dots, B_{t_d})$ with mean zero and covariance matrix

$$\left(\min(t_j, t_k)\right)_{j,k=1}^{d} = \begin{pmatrix} t_1 & t_1 & t_1 & \dots & t_1 \\ t_1 & t_2 & t_2 & \dots & t_2 \\ t_1 & t_2 & t_3 & \dots & t_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_1 & t_2 & t_3 & \dots & t_d \end{pmatrix}$$

# Generation of Brownian paths

## Classical constructions

For simplicity we only consider the evenly spaced case, i.e., $t_k = \frac{k}{d}T$, $k = 1, \ldots, d$. And we specialize to $T = 1$. Then the covariance matrix takes on the form

$$\left(\min(t_j, t_k)\right)_{j,k=1}^{d} = \frac{1}{d}\begin{pmatrix} 1 & 1 & 1 & \ldots & 1 \\ 1 & 2 & 2 & \ldots & 2 \\ 1 & 2 & 3 & \ldots & 3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 2 & 3 & \ldots & d \end{pmatrix}$$

# Generation of Brownian paths
## Classical constructions

Three classical constructions of discrete Brownian paths:

- the forward method, a.k.a. step-by-step method or piecewise method
- the Brownian bridge construction or Lévy-Ciesielski construction
- the principal component analysis construction (PCA construction)

# Generation of Brownian paths

## Classical constructions

Forward method:

- given a standard normal vector $X = (X_1, \dots, X_d)$
- compute discrete Brownian path inductively by

$$B_{\frac{1}{d}} = \sqrt{\frac{1}{d}} X_1 \,, \quad B_{\frac{k+1}{d}} = B_{\frac{k}{d}} + \sqrt{\frac{1}{d}} X_{k+1}$$

- Using that $\mathrm{E}(X_j X_k) = \delta_{jk}$, it is easy to see that $(B_{\frac{1}{d}}, \dots, B_1)$ has the required correlation matrix
- simple and efficient: generation of the normal vector plus $d$ multiplications and $d-1$ additions.

# Generation of Brownian paths

## Classical constructions

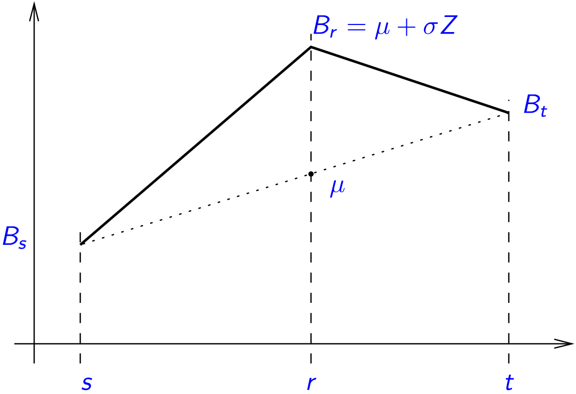
Brownian bridge construction: allows the values $B_{\frac{1}{d}}, \ldots, B_{\frac{d}{d}}$ to be computed in any given order

**Lemma**

*Let $B$ be a Brownian motion and let $r < s < t$.*
*Then the conditional distribution of $B_s$ given $B_r, B_t$ is $N(\mu, \sigma^2)$ with*

$$\mu = \frac{t-s}{t-r} B_s + \frac{s-r}{t-r} B_t \text{ and } \sigma^2 = \frac{(t-s)(s-r)}{t-r}.$$

# Generation of Brownian paths
## Classical constructions



# Generation of Brownian paths
## Classical constructions

- Discrete path can be computed in time proportional to $d$, given that factors are precomputed
- typical order of construction $B_1, B_{\frac{1}{2}}, B_{\frac{1}{4}}, B_{\frac{3}{4}}, B_{\frac{1}{8}}, \ldots$ (for $d = 2^m$)

# Generation of Brownian paths
## Classical constructions

PCA construction:

- exploits the fact that correlation matrix $\Sigma$ of $(B_{\frac{1}{d}}, \ldots, B_{\frac{d}{d}})$ is positive definite
- can be written $\Sigma = VDV^{-1}$ for a diagonal matrix $D$ with positive entries and an orthogonal matrix $V$
- $D$ can be written as $D = D^{\frac{1}{2}} D^{\frac{1}{2}}$, where $D^{\frac{1}{2}}$ is the element-wise positive square root of $D$
- now compute
$$(B_{\frac{1}{d}}, \ldots, B_{\frac{d}{d}})^\top = VD^{\frac{1}{2}} X \,.$$
  $X$ a standard normal random vector
- matrix-vector multiplication can be done in time proportional to $d \log(d)$ (Scheicher 2007)

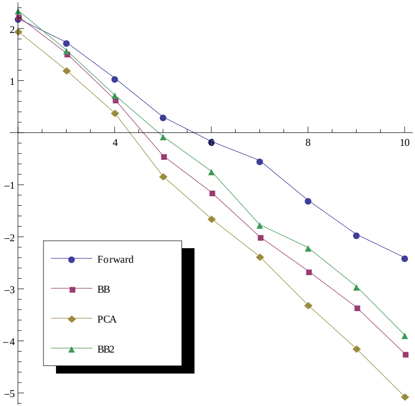
# Generation of Brownian paths
## Examples

Why do we need more than one construction?

- Consider the problem of valuating an average value option in the Heston model.
- Use Euler-Maruyama method to solve SDE
- Test the different approaches numerically:
  - model parameters: $s_0 = 100$, $v_0 = 0.3$, $r = 0.03$, $\rho = 0.2$, $\kappa = 2$, $\theta = 0.3$, $\xi = 0.5$,
  - option parameters: $K = 100$, $T = 1$.
  - $d = 64$

# Generation of Brownian paths

## Examples



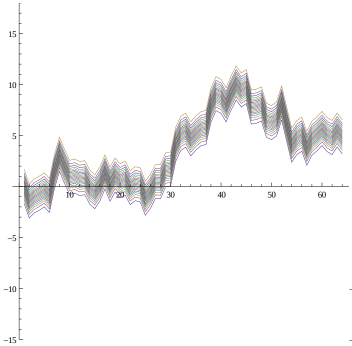
# Generation of Brownian paths
## Examples

Can we explain this behavior?

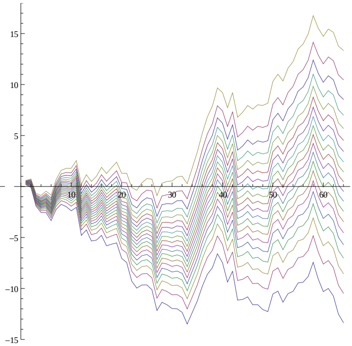
- QMC seems to perform better if some of the variables are more important than the others
- alternative construction often help to put more weight on earlier variables

# Generation of Brownian paths
## Examples

All variables but the first left constant:



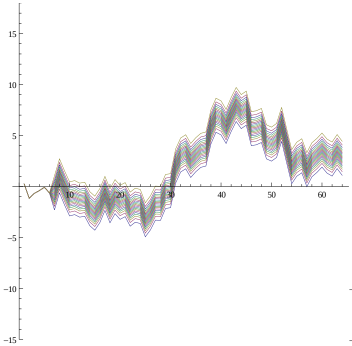

forward



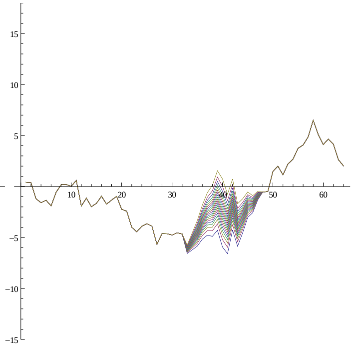

brownian bridge

# Generation of Brownian paths
## Examples

All variables but the seventh left constant:




forward




brownian bridge

# Generation of Brownian paths

## Examples

- notion of effective dimension
  - tries to explain why problem behaves low-dimensional w.r.t. QMC
  - uses concept of ANOVA decomposition of a function into lower-dimensional components
- alternative concept: weighted Korobov- or Sobolev spaces
  - give Koksma-Hlawka type inequalities with weighted norm/discrepancy
  - sequence need not be as well-distributed in coordinates that are less important
- both concepts have some connections

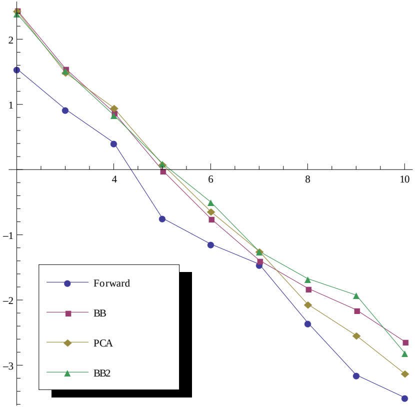
# Generation of Brownian paths
## Examples

But caution is in order

"Ratchet" Option: (Papageorgiou 2002) Same example model, but different payoff:

$$f(S_{\frac{T}{d}}, S_{\frac{2T}{d}}, \ldots, S_T) = \frac{1}{d} \sum_{j=1}^{d} 1_{[0,\infty)}\left(S_{\frac{jT}{d}} - S_{\frac{(j-1)T}{d}}\right) S_{\frac{jT}{d}} \, .$$

# Generation of Brownian paths
## Examples



# Generation of Brownian paths
## Orthogonal transforms

- Whether a path construction is "good" or not depends on the payoff as well
- before we continue with asking why QMC is good when combined with some pairs of payoffs/constructions
- we want a general framework for the constructions

# Generation of Brownian paths
## Orthogonal transforms

Cholesky decomposition of $\Sigma^{(d)}$: $\Sigma^{(d)} = SS^\top$, where

$$S = S^{(d)} := \frac{1}{\sqrt{d}} \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \dots & 1 \end{pmatrix} .$$

Note that $Sy$ is the cumulative sum over $y$ divided by $\sqrt{d}$,

$$Sy = \frac{1}{\sqrt{d}}(y_1, y_1 + y_2, \dots, y_1 + \dots + y_d)^\top$$

# Generation of Brownian paths

## Orthogonal transforms

### Lemma (Papageorgiou 2002)

*Let $A$ be any $d \times d$ matrix with $AA^\top = \Sigma$ and let $X$ be a standard normal vector. Then $B = AX$ is a discrete Brownian path with discretization $\frac{1}{d}, \frac{2}{d}, \ldots, \frac{d-1}{d}, 1$.*

### Lemma (Papageorgiou 2002)

*Let $A$ be any $d \times d$ matrix with $AA^\top = \Sigma$. Then there is an orthogonal $d \times d$ matrix $V$ with $A = SV$. Conversely, $SV(SV)^\top = \Sigma$ for every orthogonal $d \times d$ matrix $V$.*

# Generation of Brownian paths
## Orthogonal transforms

- orthogonal transform corresponding to forward method is $\mathrm{id}_{\mathbb{R}^d}$
- Brownian bridge construction for $d = 2^k$, with order $B_1, B_{\frac{1}{2}}, B_{\frac{1}{4}}, B_{\frac{3}{4}}, B_{\frac{1}{8}}, B_{\frac{3}{8}}, B_{\frac{5}{8}}, \ldots$, is given by the inverse Haar transform
- for the PCA, the orthogonal transform has been given explicitly in terms of the fast sine transform
- many orthogonal transforms can be computed using $O(d \log(d))$ operations (L. 2012)
- Examples include: Walsh, discrete sine/cosine, Hilbert, Hartley, wavelet and others
- orthogonal transforms have no influence on the probabilistic structure of the problem

# Weighted norms

# Weighted norms

Consider functions on $[0,1]^d$ and the following norm:

$$\|f\|^2 = \sum_{\mathbf{u}\subseteq\{1,\ldots,d\}} \int_{[0,1]^{|\mathbf{u}|}} \left| \frac{\partial^{|\mathbf{u}|} f}{\partial x_{\mathbf{u}}}(x_{\mathbf{u}}, 1) \right|^2 dx_{\mathbf{u}}$$

Here, $(x_{\mathbf{u}}, 1)$ denotes the vector one obtains by replacing coordinates with index not in $\mathbf{u}$ by 1
and $\frac{\partial^{|\mathbf{u}|} f}{\partial x_{\mathbf{u}}}$ means derivative by every variable with index in $\mathbf{u}$
(the corresponding 1-norm, if defined and finite, equals the variation in the sense of Hardy and Krause)

# Weighted norms

Sloan & Wożniakowski (1998) indroduced a sequence of weights $\gamma_1 \geq \gamma_2 \geq \ldots > 0$ and defined a weighted norm instead

$$\|f\|^2 = \sum_{\mathbf{u} \subseteq \{1,\ldots,d\}} \gamma_{\mathbf{u}}^{-1} \int_{[0,1]^{|\mathbf{u}|}} \left| \frac{\partial f}{\partial x_{\mathbf{u}}}(x_{\mathbf{u}}, 1) \right|^2 dx_{\mathbf{u}}$$

where $\boldsymbol{\gamma}_{\mathbf{u}} = \prod_{k \in \mathbf{u}} \gamma_k$

For example, if $\sum_{k=1}^{\infty} \gamma_k < \infty$, this makes contributions of larger indices bigger

# Weighted norms

- Rephrased: higher dimensions need to be less relevant to make norm still small
- Sloan & Wożniakowski (1998) present corresponding weighted discrepancy and weighted Koksma-Hlawka inequality
- Requirements on discrepancy more relaxed $\Longrightarrow$ have better dependence on dimension
- Sloan & Wożniakowski (1998) only show existence of good integration nodes
- for example, if $\sum_{k=1}^{\infty} \gamma_k < \infty$, then we can make the integration error small, independently of dimension!!

# Weighted norms

- Since then deviation from "One-size-fits-all approach" for construction of QMC point sets and sequences
- (Fast) Component-by-component constructions of point sets for given weights Dick & L. & Pillichshammer, Cools & Nuyens, Kritzer & L. & Pillichshammer, Dick & Kritzer & L. & Pillichshammer
- Many different norms/spaces and equi-distribution measures
- Main tool: reproducing kernel Hilbert space,
  that is,
  a Hilbert space of functions for which function evaluation is continuous

# Weighted norms

Thus, the problem was solved

The end

Or is it ?

# Weighted norms

Not quite!

- Transformation of financial problems to unit cube usually leads to infinite (weighted) norm
- there is no guarantee that finite norm with respect to one path construction gives finite norm in another
  directions of coordinate axes are special
- no or very few means to find “optimal” construction

# Weighted norms

Idea by Irrgeher & L.(2015): find a class of reproducing kernel Hilbert spaces

- of functions on the $\mathbb{R}^d$
- with a weighted norm
- that is continuous w.r.t. orthogonal transforms of the $\mathbb{R}^d$
- and allows for tractability/complexity discussions

# Hermite spaces

# Hermite spaces
## One-dimensional Hermite space

- $\phi(x) := \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$, $x \in \mathbb{R}$
- $L^2(\mathbb{R}, \phi) = \{f : \text{measurable and } \int_{\mathbb{R}} |f|^2 \phi < \infty\}$
- $(\bar{H}_k)_k$ ... sequence of normalized Hermite polynomials (i.e. $\bar{H}_0, \bar{H}_1, \bar{H}_2, \ldots$ is the Gram-Schmidt orthogonalization of $1, x, x^2, \ldots$ in $L^2(\mathbb{R}, \phi)$
- $(\bar{H}_k)_k$ ... forms Hilbert space basis of $L^2(\mathbb{R}, \phi)$, i.e.

$$f = \sum_{k \geq 0} \hat{f}(k) \bar{H}_k \qquad \text{in } L^2(\mathbb{R}, \phi)$$

# Hermite spaces

## One-dimensional Hermite space

**Theorem (Irrgeher & L. (?))**

*Let* $(r_k)_{k\geq 0}$ *be a sequence with*

- $r_k > 0$
- $\sum_{k\geq 0} r_k < \infty$

*If* $f : \mathbb{R} \to \mathbb{R}$ *is continuous,* $\int_{\mathbb{R}} f(x)^2 \phi(x) dx < \infty$*, and* $\sum_{k\geq 0} r_k^{-1} |\hat{f}(k)|^2 < \infty$ *then*

$$f(x) = \sum_{k\geq 0} \hat{f}(k) \bar{H}_k(x) \qquad \textit{for all } x \in \mathbb{R}$$

# Hermite spaces

## One-dimensional Hermite space

- Fix some positive summable sequence $r = (r_k)_{k \geq 0}$
- Introduce new inner product:

$$\langle f, g \rangle_{\mathrm{her}} := \sum_{k=0}^{\infty} r_k^{-1} \hat{f}(k) \hat{g}(k)$$

- and corresponding norm $\|.\|_{\mathrm{her}} := \langle ., . \rangle^{1/2}$,

$$\|f\|_{\mathrm{her}}^2 := \sum_{k=0}^{\infty} r_k^{-1} \hat{f}(k)^2$$

# Hermite spaces

## One-dimensional Hermite space

**Theorem (Irrgeher & L.(2015))**

*The Hilbert space*

$$\mathscr{H}_{\mathrm{her}}(\mathbb{R}) := \{f \in L^2(\mathbb{R}, \phi) \cap C(\mathbb{R}) : \|f\|_{\mathrm{her}} < \infty\}$$

*is a reproducing kernel Hilbert space with reproducing kernel*

$$K_{\mathrm{her}}(x, y) = \sum_{k \in \mathbb{N}_0} r(k) \bar{H}_k(x) \bar{H}_k(y)$$

(Can compute function evaluation by inner product
$f(x) = \langle f(.), K(x, .)\rangle_{\mathrm{her}}, \ \forall x \in \mathbb{R}$)

# Hermite spaces
## One-dimensional Hermite space

There are indeed some interesting functions in $\mathscr{H}_{\mathrm{her}}(\mathbb{R})$:

**Theorem (Irrgeher & L.(2015))**

*Let $r_k = k^{-\alpha}$, let $\beta > 2$ be an integer, and let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be a $\beta$ times differentiable function such that*

(i) $\int_{\mathbb{R}} |D_x^j f(x)| \phi(x)^{1/2} dx < \infty$ *for each* $j \in \{0, \ldots, \beta\}$ *and*

(ii) $D_x^j f(x) = O(e^{x^2/(2c)})$ *as* $|x| \to \infty$ *for each* $j \in \{0, \ldots, \beta - 1\}$ *and some* $c > 1$.

*Then $f \in \mathscr{H}_{\mathrm{her}}(\mathbb{R})$ for all $\alpha$ with $1 < \alpha < \beta - 1$.*

(derivatives up to order $\beta > \alpha + 1$ exist, satisfy an integrability and growth condition)

# Hermite spaces

## $d$-dimensional Hermite space

- For a $d$-multi-index $\mathbf{k} = (k_1, \ldots, k_d)$ define

$$\bar{H}_{\mathbf{k}}(x_1, \ldots, x_d) := \prod_{j=1}^{d} \bar{H}_{k_j}(x_j)$$

- defines Hilbert space basis of $L^2(\mathbb{R}^d, \phi)$
- write $f_{\mathbf{k}} := \langle f, \bar{H}_{\mathbf{k}} \rangle = \int_{\mathbb{R}^d} f(\mathbf{x})\bar{H}_{\mathbf{k}}(\mathbf{x})\phi(\mathbf{x})d\mathbf{x}$

We consider

$$\mathscr{H}_{\mathrm{her},\boldsymbol{\gamma}}(\mathbb{R}^d) := \mathscr{H}_{\mathrm{her}}(\mathbb{R}) \otimes \ldots \otimes \mathscr{H}_{\mathrm{her}}(\mathbb{R}).$$

with the inner product

$$\langle f, g \rangle_{\mathrm{her},\boldsymbol{\gamma}} = \sum_{\mathbf{k} \in \mathbb{N}_0^d} \mathbf{r}(\boldsymbol{\gamma}, \mathbf{k})^{-1} \hat{f}(\mathbf{k}) \hat{g}(\mathbf{k})$$

where the function $\mathbf{r}(\gamma, .) : \mathbb{N}_0^d \longrightarrow \mathbb{R}$ is given by

$$\mathbf{r}(\gamma, \mathbf{k}) = \prod_{j=1}^{d} (\delta_0(k_j) + (1 - \delta_0(k_j))\gamma_j^{-1} r_{k_j})$$

i.e. $\mathbf{r}(\gamma, \mathbf{k}) = \prod_{j=1}^{d} \tilde{r}(\gamma_j, k_j)$ where

$$\tilde{r}(\gamma, k) := \begin{cases} 1 & k = 0 \\ \gamma^{-1} r_k & k \geq 1 \end{cases}$$

# Hermite spaces
## $d$-dimensional Hermite space

"Canonical" RK:

$$K_{\mathrm{her},\boldsymbol{\gamma}}(\mathbf{x},\mathbf{y}) := \sum_{\mathbf{k}\in\mathbb{N}_0^d} \mathbf{r}(\gamma,\mathbf{k})\bar{H}_{\mathbf{k}}(\mathbf{x})\bar{H}_{\mathbf{k}}(\mathbf{y})$$

With this $\mathscr{H}_{\mathrm{her},\boldsymbol{\gamma}}$ is weighted RKHS of functions on the $\mathbb{R}^d$

# Hermite spaces
## $d$-dimensional Hermite space

Integration:

$$I(f) = \int_{\mathbb{R}^d} f(\mathbf{x})\phi(\mathbf{x})d\mathbf{x}$$

**Theorem (Irrgeher & L.(2015))**

*Integration in the RKHS $\mathscr{H}_{\mathrm{her},\gamma}(\mathbb{R}^d)$ is*

- *strongly tractable if $\sum_{j=1}^{\infty} \gamma_j < \infty$,*
- *tractable if $\limsup_d \frac{1}{\log d} \sum_{j=1}^{d} \gamma_j < \infty$.*

Irrgeher, Kritzer, L., Pillichshammer (2015) study Hermite spaces of analytic functions and find lower bounds on complexity

# Hermite spaces

## $d$-dimensional Hermite space

Why are we interested in this kind of space?

- Let $f \in \mathscr{H}_{\mathrm{her},\boldsymbol{\gamma}}$ and let $U : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ some orthogonal transform, $U^\top U = 1_{\mathbb{R}^d}$
- then $f \circ U \in \mathscr{H}_{\mathrm{her},\boldsymbol{\gamma}}$
- also $\int_{\mathbb{R}^d} f \circ U(\mathbf{x})\phi(\mathbf{x})d\mathbf{x} = \int_{\mathbb{R}^d} f(\mathbf{x})\phi(\mathbf{x})d\mathbf{x}$
- but in general $\|f \circ U\|_{\mathrm{her},\boldsymbol{\gamma}} \neq \|f\|_{\mathrm{her},\boldsymbol{\gamma}}$

# Hermite spaces
## Example

Example from Irrgeher & L.(2015): compute $\mathrm{E}(\exp(W_1))$, where $W$ is a Brownian path
Corresponds to integrating function a $f : \mathbb{R}^d \to \mathbb{R}$, if $W_1$ is computed using the foirward construction

- $f \in \mathscr{H}_{\mathrm{her},\gamma}$ for a sensible choice of $\gamma$
- $\|f\|_{\mathrm{her},\gamma} \geq ce^d$ for some $c$ and all $d \in \mathbb{N}$
- $\|f \circ U\|_{\mathrm{her},\gamma} \leq C$ for some $C$, where $U$ is the inverse Haar transform

# Hermite spaces
## Example

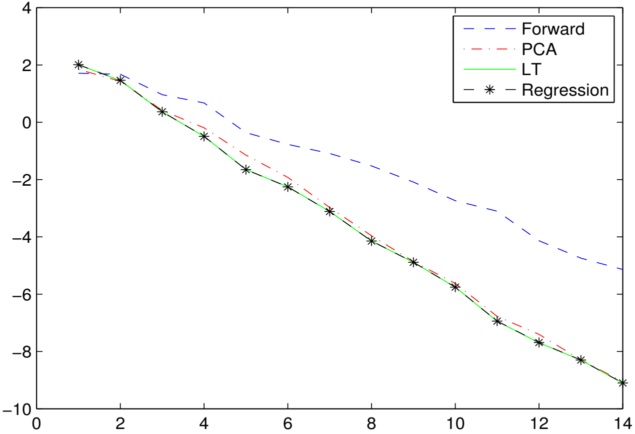
- norm of $\|f \circ U\|$ depends on $U$ in a continuous fashion.
- We can – in principle – use optimization techniques to find best transform

# Hermite spaces

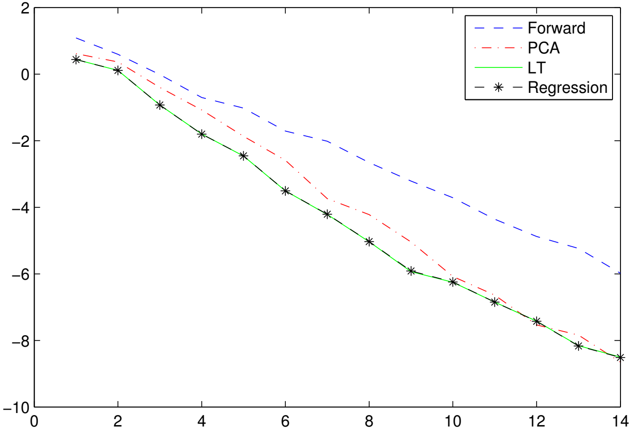
## Example

- An earlier result/method by Irrgeher & L. is better understood in the context of Hermite spaces
- instead of minimizing the weighted norm of $\|f \circ U\|$, minimize a seminorm which does not take into account all Hermite coefficients
- for example, only consider order one coefficients
- method is termed linear regression method and generates paths in linear time

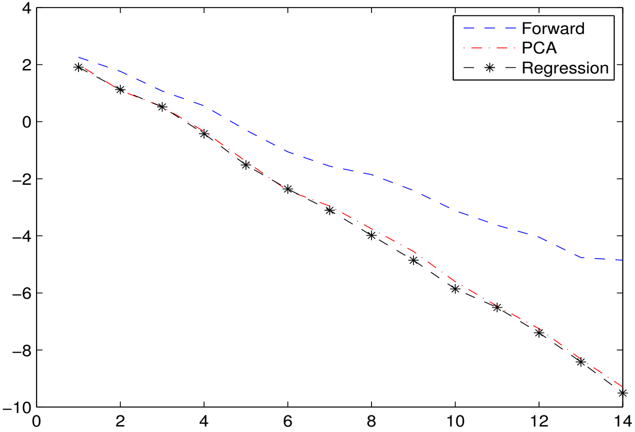
## Average value option



# Average value basket option



## Average value barrier option



# Hermite spaces

## Conclusion

- We have provided a potential approach to explaining to effectiveness of QMC for high-dimensional financial applications
- the approach enabled us to find a method that is practically the best available at the moment
- different lines of research:
  - construct point sets/sequences for those spaces
  - generalize regression method to higher oder approximations
  - make regression method more "automatic"
  - deal with "kinks"

Thank you !

- C. Irrgeher, P. Kritzer, G. Leobacher, F. Pillichshammer Integration in Hermite spaces of analytic functions. J. Complexity (31), pp. 380-404. 2015
- J. Dick, P. Kritzer, G. Leobacher, F. Pillichshammer A reduced fast component-by-component construction of lattice points for integration in weighted spaces with fast decreasing weights. J. Comput. Appl. Math. (276), pp. 1-15. 2015
- C. Irrgeher, G. Leobacher High-dimensional integration on $\mathbb{R}^d$, weighted Hermite spaces, and orthogonal transforms. J. Complexity (31), pp. 174-205. 2015
- G. Leobacher, F. Pillichshammer Introduction to quasi-Monte Carlo Integration and Applications. Compact Textbooks in Mathematics, Birkhäuser/Springer
- G. Leobacher Fast orthogonal transforms and generation of Brownian paths. Journal of Complexity (28), pp. 278-302. 2012
- I. Sloan, H. Woňiakowski: When Are Quasi-Monte Carlo Algorithms Efficient for High Dimensional Integrals? Journal of Complexity (14), pp. 1–33, 1998.
- A. Papageorgiou: The Brownian Bridge Does Not Offer a Consistent Advantage in Quasi-Monte Carlo Integration. Journal of Complexity (18), pp. 171–186, 2002.