

# Quasi-Monte Carlo Methoden für Optimierungsmodelle der Energiewirtschaft

H. Leövey und W. Römisch

Humboldt-Universität Berlin

Institut für Mathematik

<http://www.math.hu-berlin.de/~romisch>



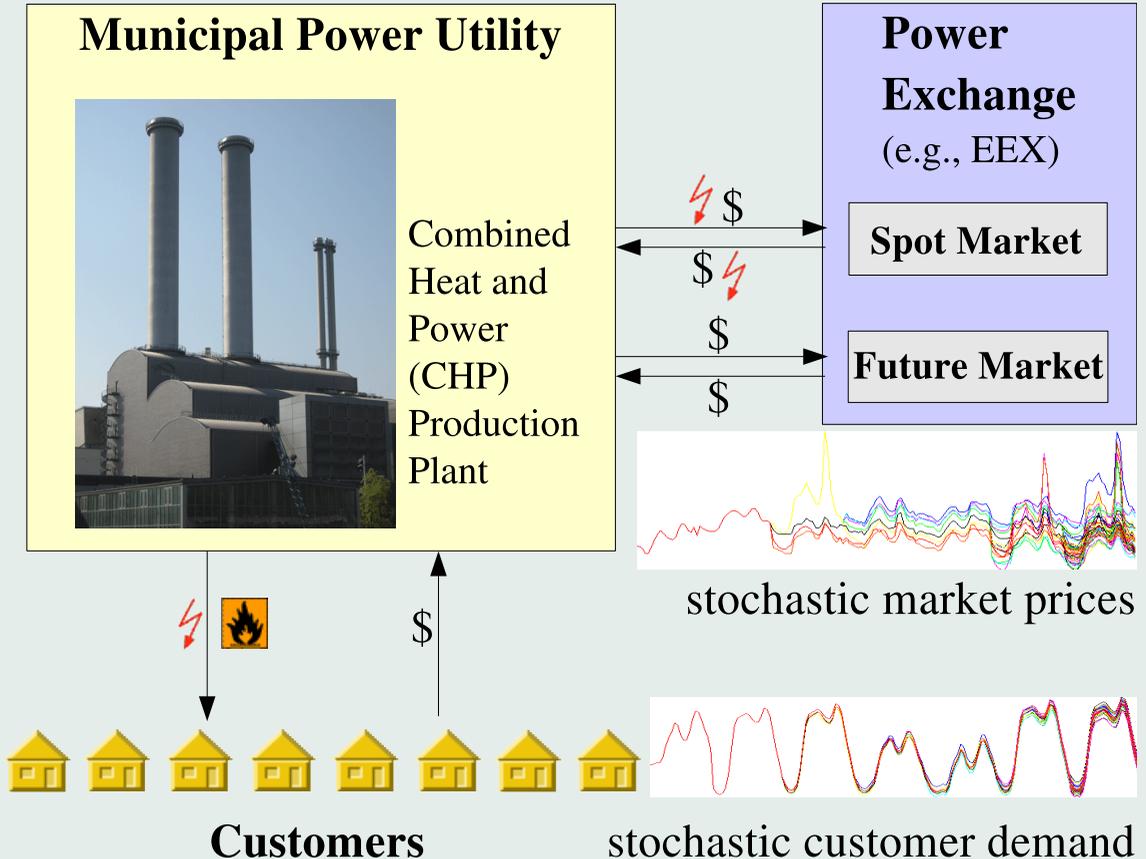
VDI-Tagung "Optimierung in der Energiewirtschaft"  
Düsseldorf, 25./26.11.15

# Einleitung

- Energiewirtschaftliche Modelle enthalten oft **unsichere Parameter (Marktpreise, Zuflüsse, Wind, Last)**, für die (statistische) Daten und Modelle existieren.
- Wir betrachten hier Modelle für **kleinere Marktteilnehmer**, da deren Entscheidungen die Preise am Strommarkt nicht bzw. kaum beeinflussen.
- Die unsicheren Parameter werden durch eine **endliche Anzahl von Szenarien mit zugehörigen Wahrscheinlichkeiten** (evtl. in Form eines Szenariobaumes) dargestellt.
- Damit kann eine **gleichzeitige Maximierung des erwarteten Gewinns und evtl. Minimierung des Risikos** erfolgen.
- In den letzten 15 Jahren gab es einen **Durchbruch** bei der numerischen Berechnung hochdimensionaler Integrale durch randomisierte Quasi-Monte Carlo Methoden.

**Ziel: Modellierung der unsicheren Parameter mittels randomisierter Quasi-Monte Carlo Methoden**

# Strom-Portfolio Optimierung



Wir betrachten das **Strom Portfolio Management** eines **Stadtwerkes**, dessen Portfolio sich aus folgenden Bestandteilen zusammensetzt:

- **Stromproduktion** (mit Hilfe von eigenen thermischen Anlagen),
- **bilaterale Verträge**,
- **(day-ahead) Spotmarkt-Handel**.
- **(Future Handel für Risiko-Management)**.

Der **Zeithorizont** wird in **stündliche Intervalle** unterteilt. Der stochastische Prozess besteht aus den Komponenten **Strom-Last** und **Spotmarkt-Preise** und ist **multivariat**, d.h., es existieren vielfältige statistische Abhängigkeiten. Der stochastische Prozess wird approximativ durch eine endliche Anzahl von Szenarien dargestellt. Das Ziel besteht in der **Maximierung des erwarteten Gewinns** (und der eventuellen gleichzeitigen Minimierung des Risikos). Das Strom Portfolio Management Modell stellt ein **sehr großes gemischt-ganzzahliges lineares Optimierungsproblem** dar.

Das Modell enthält die **elektrische Last**  $\xi_\delta$  und **Strommarktpreise**  $\xi_c$  als stochastische Parameter. Beide sind Komponenten des Zufallsvektors

$$\xi = (\xi_{\delta,1}, \xi_{c,1}, \dots, \xi_{\delta,T}, \xi_{c,T})^\top.$$

Der Zeithorizont ist in  $T$  Zeitintervalle unterteilt. In jedem Zeitpunkt  $t \in \{1, \dots, T\}$  muss die Firma die Last decken, was ihr aber bei Spitzenlast nur durch Stromhandel auf der Basis bilateraler Verträge zu festen Preisen bzw. durch Zukäufe am Strommarkt zu stochastischen Preisen gelingt. Bei Niedriglast ist ein Abschalten eigener Einheiten erforderlich, die später wieder zugeschaltet werden müssen.

Das **zweistufige Produktionsmodell** besitzt die Gestalt

$$\min \left\{ \sum_{t=1}^T \left[ c_t^\top x_t + \int_{\mathbb{R}^T} q_t(\xi)^\top y_t P(d\xi) \right] : W y + V x \geq h(\xi), y \in Y, x \in X \right\}$$

wobei der Vektor  $x_t$  die Leistungen der Grundlast-Blöcke und  $c_t$  deren Kosten zum Zeitpunkt  $t$  beschreiben. Die Menge  $X$  enthält Leistungsgrenzen und mögliche Schranken für Leistungsänderungen der thermischen Einheiten in jedem Zeitpunkt.

Der Vektor  $y_t$  der Zweitstufen-Entscheidung enthält die 0-1-Schaltentscheidungen und Leistungen der zyklischen Blöcke und des bilateralen Vertrages sowie den Zukauf vom Strommarkt. Die Restriktionen  $Wy + Vx \geq h(\xi)$  beschreiben die Lastdeckung in jedem Zeitpunkt  $t$  und Mindestlauf- bzw. Mindeststillstandszeiten der zyklischen Blöcke. Die Restriktion  $y \in Y$  enthält Leistungsgrenzen und Ganzzahligkeitsbedingungen.  $P$  ist die Wahrscheinlichkeitsverteilung von  $\xi$  auf  $\mathbb{R}^{2T}$ .

Wir nehmen an, dass der **stochastische Last- und Preis-Prozess**  $\{\xi_{\delta,t}, \xi_{c,t}\}$  **als lineare Zeitreihe** (Eichhorn-Römisch-Wegner 05) modelliert werden kann, der Einfachheit halber als ARMA(1,1) Prozess

$$\begin{pmatrix} \xi_{\delta,t} \\ \xi_{c,t} \end{pmatrix} = \begin{pmatrix} \bar{\xi}_{\delta,t} \\ \bar{\xi}_{c,t} \end{pmatrix} + \begin{pmatrix} E_{1,t} \\ E_{2,t} \end{pmatrix} \quad (t = 1, \dots, T), \quad \begin{pmatrix} \bar{\xi}_{\delta,1} \\ \bar{\xi}_{c,1} \end{pmatrix} = B_1 \begin{pmatrix} \gamma_{1,1} \\ \gamma_{2,1} \end{pmatrix},$$

$$\begin{pmatrix} \bar{\xi}_{\delta,t} \\ \bar{\xi}_{c,t} \end{pmatrix} = A \begin{pmatrix} \bar{\xi}_{\delta,t-1} \\ \bar{\xi}_{c,t-1} \end{pmatrix} + B_1 \begin{pmatrix} \gamma_{1,t} \\ \gamma_{2,t} \end{pmatrix} + B_2 \begin{pmatrix} \gamma_{1,t-1} \\ \gamma_{2,t-1} \end{pmatrix} \quad (t = 2, \dots, T),$$

wobei  $E_{1,t} = \mathbb{E}[\xi_{\delta,t}]$ ,  $E_{2,t} = \mathbb{E}[\xi_{c,t}]$ ,  $A$ ,  $B_1$ ,  $B_2$  (2, 2)-Matrizen und  $\gamma_{1,t}$ ,  $\gamma_{2,t}$  unabhängige und standard normalverteilte Zufallsvariablen sind. Die Zeitreihe erlaubt die **Berücksichtigung von Korrelationen**.

# Modellierung der unsicheren Parameter

1. Schritt: Anpassung eines statistischen Modells der Form

$$\xi_t = G_t(\xi_1, \dots, \xi_{t-1}; \eta_1, \dots, \eta_t), \quad t = 1, 2, \dots, T,$$

wobei die  $\eta_t$ ,  $t = 1, \dots, T$ , Zufallsvektoren mit unabhängigen Komponenten und identischer Verteilung (sog. **Innovationen**) sind, (Beispiel: Lineare Zeitreihen

$$\xi_t + \sum_{i=1}^k a_i \xi_{t-i} = \sum_{i=0}^k b_i \eta_{t-i}, \quad t = k, k+1, \dots, T.)$$

oder **Schätzung einer multivariaten Wahrscheinlichkeitsverteilung** jeweils auf der Basis vorhandener statistischer Daten.

2. Schritt: **Sampling** vom statistischen Modell bzw. von der multivariaten Wahrscheinlichkeitsverteilung **mit Hilfe von Monte Carlo oder Quasi-Monte Carlo Methoden**.

## Monte Carlo Methoden: Vorteile und Nachteile

Monte Carlo Methoden beruhen darauf, **unabhängige Samples**  $\xi^n(\cdot)$ ,  $n \in \mathbb{N}$ , mit **Werten in  $\mathbb{R}^d$**  und **gemeinsamer Verteilung  $P$**  bzw. Dichte  $\rho$  zu bestimmen, so dass

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

das Integral  $\int_{\mathbb{R}^d} f(\xi)P(d\xi)$  bzw.  $\int_{\mathbb{R}^d} f(\xi)\rho(\xi)d\xi$  approximiert.

- Monte Carlo (MC) Sampling funktioniert für alle  $L_2$ -Integranden.
- Die MC Konvergenzrate  $O(\frac{1}{\sqrt{n}})$  hängt nicht von  $d$  ab.
- MC Methoden sind in der Praxis aber enttäuschend langsam.
- Mögliche Glattheit des Integranden verbessert die Rate nicht.
- Die Erzeugung unabhängiger zufälliger Samples ist schwierig. Praktisch werden unabhängige und identisch verteilte Samples approximativ durch sog. **Pseudo-Zufallszahlengeneratoren** zunächst als gleichverteilte Samples in  $[0, 1]^d$  erzeugt.

# Quasi-Monte Carlo Methoden

Wir betrachten die approximative Berechnung von

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

durch eine Quasi-Monte Carlo (QMC) Methode

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

mit (determinierten) Punkten  $\xi^i$ ,  $i = 1, \dots, n$ , aus  $[0, 1]^d$ .

Es existieren zwei Haupt-Gruppen von QMC Methoden:

(Dick-Pillichshammer 10, Dick-Kuo-Sloan 13, Leobacher-Pillichshammer 14)

- (1) Digital Netze und Folgen,
- (2) Gitter-Regeln (lattice rules).

## Spezielle digitale Folgen:

Sobol' Folgen (Sobol' 67);

Faure Folgen (Faure 82);

Niederreiter Folgen (Niederreiter 87);

verallgemeinerte Niederreiter Folgen (Niederreiter 05)

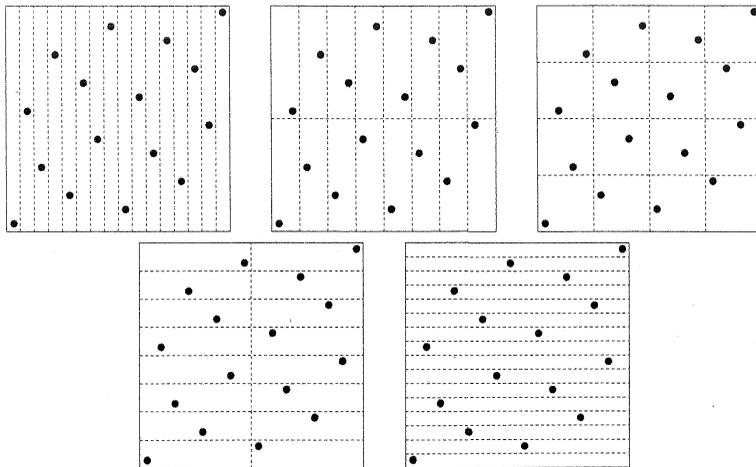
## Spezielle Rang-1 Gitter:

$$\left\{ \frac{(i-1)}{n} g \right\}, \quad i = 1, \dots, n,$$

wobei  $g \in \mathbb{Z}_+^d$  der Erzeuger des Gitters ist und die Klammern  $\{\cdot\}$  bedeuten, dass komponentenweise nur die Ziffern nach dem Dezimalpunkt verwendet werden.

Klassische Konvergenzrate:  $|Q_n(f) - I_d(f)| = O(n^{-1}(\log n)^d)$

Quasi-Monte Methoden besitzen besonders gute Konvergenzeigenschaften, falls die Integranden eine niedrige effektive Dimension besitzen.



**Fig. 5.2** A  $(0, 4, 2)$ -net in base 2; every 2-dimensional elementary interval in base 2 of area  $2^{-4}$  contains exactly one point

# Randomisierte QMC Methoden

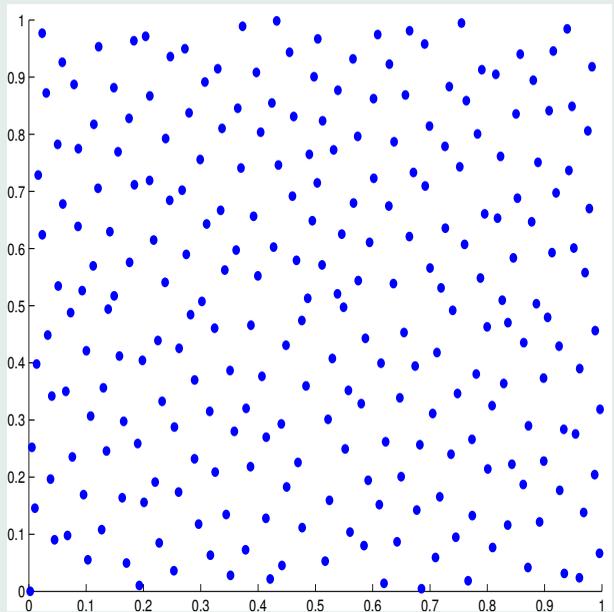
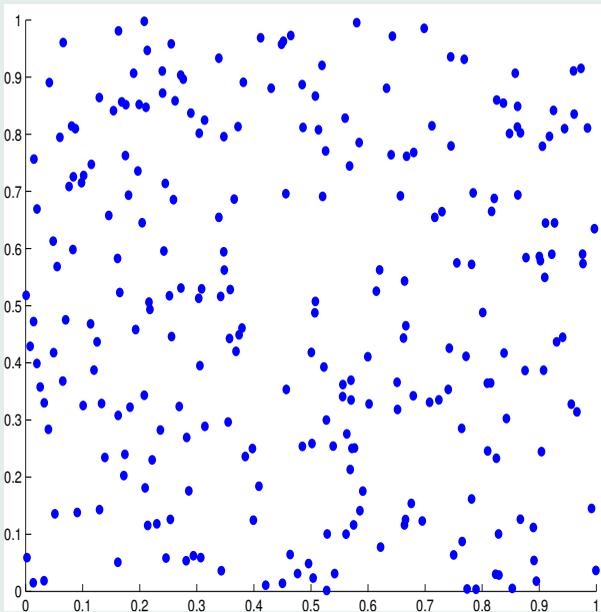
Randomisierte Versionen von QMC Punktmengen haben die Eigenschaft dass (Owen 95, L'Ecuyer-Lemieux 02, Dick-Pillichshammer 10)

- (i) jeder Punkt der randomisierten Punktmenge eine Gleichverteilung über  $[0, 1)^d$  besitzt, und
- (ii) die QMC Eigenschaften unter der Randomisierung mit Wahrscheinlichkeit 1 erhalten bleiben.

Beispiele für Randomisierungstechniken sind

- (a) zufällige Verschiebungen (shifts) von Gitter-Regeln,
- (b) scrambling, d.h., zufällige Permutationen der Zahlen in  $\mathbb{Z}_b = \{0, 1, \dots, b-1\}$  mit Anwendung auf die Ziffern von Zahldarstellungen zur Basis  $b$ ,
- (c) affines scrambling erzeugt zufällige Ziffern durch zufällige lineare Transformationen der Original-Ziffern, wobei die Elemente aller Matrizen und Vektoren zufällig, unabhängig und gleichverteilt in  $\mathbb{Z}_b$  gewählt sind.

Eigenschaften (i) und (ii) ermöglichen Fehlerabschätzungen und führen zu verbesserten Konvergenzeigenschaften verglichen mit der Original QMC Methode.



Vergleich von  $n = 2^7$  Monte Carlo Mersenne Twister Punkten und randomly binary shifted Sobol' Punkten der Dimension  $d = 500$ , Projektion auf die 8. und 9. Komponente

**Randomly scrambled Sobol' Folgen** besitzen die folgende **Konvergenzrate** des Quadratmittel-Fehlers für  $f \in \mathcal{W}_{2,\gamma,\text{mix}}^{(1,\dots,1)}([0, 1]^d)$

$$\sup_{f \in \mathbb{B}_d} \sqrt{\mathbb{E}[Q_n(\omega)(f) - I_d(f)]^2} \leq C_d n^{-\frac{3}{2}} (\log n)^{\frac{d-1}{2}} .$$

(Dick-Pillichshammer 10)

## Randomly shifted lattice rules

Ist  $\Delta$  ein Zufallsvektor mit Gleichverteilung in  $[0, 1]^d$ , so betrachten wir

$$Q_n(\omega)(f) = \frac{1}{n} \sum_{i=1}^n f\left(\left\{\frac{(i-1)}{n}g + \Delta(\omega)\right\}\right).$$

Es sei  $n$  eine Primzahl und  $f \in \mathcal{W}_{2,\gamma,\text{mix}}^{(1,\dots,1)}([0, 1]^d)$ .

Dann kann  $g \in \mathbb{Z}_+^d$  komponentenweise so konstruiert werden, dass für jedes  $\delta \in (0, \frac{1}{2}]$  eine Konstante  $C(\delta) > 0$  existiert, so dass der **worst-case Quadratmittel Integrationsfehler** die optimale Konvergenzrate

$$\sup_{f \in \mathbb{B}_d} \sqrt{\mathbb{E}[Q_n(\omega)(f) - I_d(f)]^2} \leq C(\delta) n^{-1+\delta}$$

erreicht, wobei die **Konstante**  $C(\delta)$  wächst wenn  $\delta$  kleiner wird, aber **nicht von der Dimension**  $d$  abhängt, wenn die Folge  $(\gamma_j)$  die Bedingung erfüllt

$$\sum_{j=1}^{\infty} \gamma_j^{\frac{1}{2(1-\delta)}} < \infty \quad (\text{z.B. } \gamma_j = \frac{1}{j^3}).$$

(Sloan-Kuo-Joe 02, Kuo 03, Nuyens-Cools 06)

# Transformation von Integralen für allgemeine Dichten $\rho$

Wir betrachten eine Funktion  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  und das Integral

$$\int_{\mathbb{R}^d} f(\xi) \rho(\xi) d\xi.$$

**1. Schritt:** Transformiere die multivariate Dichtefunktion  $\rho$  auf  $\mathbb{R}^d$  in eine Produktdichte auf  $\mathbb{R}^d$  mit  $d$  unabhängigen eindimensionalen Marginal-Dichten mit Hilfe von Methoden, die für die Klasse von Verteilungen, zu der  $P$  gehört, geeignet sind.

**Example:** Falls  $P$  normal mit Mittelwert 0 und regulärer Kovarianzmatrix  $\Sigma$  ist, so gilt für jede Matrix  $A$  mit  $\Sigma = AA^\top$ , dass die Dichte von  $P \circ A$  Produktform besitzt.

**2. Schritt:** Seien  $\rho_k$ ,  $k = 1, \dots, d$ , die Marginal-Dichten einer Verteilung  $P$  mit Produktdichte  $\rho(\xi) = \prod_{k=1}^d \rho_k(\xi_k)$ . Dann entsteht durch die Transformationen  $\varphi_k(\xi_k) = \int_{-\infty}^{\xi_k} \rho_k(t) dt$ ,  $x_k = \varphi_k(\xi_k)$ ,  $k = 1, \dots, d$ :

$$\int_{\mathbb{R}^d} f(\xi) \prod_{k=1}^d \rho_k(\xi_k) d\xi = \int_{[0,1]^d} f(\varphi_1^{-1}(x_1), \dots, \varphi_d^{-1}(x_d)) dx_1 \cdots dx_d$$

# ANOVA Darstellung und effektive Dimension

Wir betrachten eine nichtlineare Funktion  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  und möchten den Erwartungswert von  $f(\xi)$  berechnen, d.h.

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

wobei  $\xi$  ein  $d$ -dimensionaler Zufallsvektor mit unabhängigen Komponenten und der Dichtfunktion

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

Da die  $d$ -dimensionale Funktion  $f$  aus additiven Teilen bestehen kann, die von viel weniger Variablen abhängen und damit zu niederdimensionalen Integralen führt, wie z.B. für

$$f(\xi) = f_1(\xi_1) + \cdots + f_d(\xi_d) = \sum_{i=1}^d f_i(\xi_i)$$

$$f(\xi) = \sum_{i=1}^d \sum_{j=i}^d f_{ij}(\xi_i, \xi_j),$$

interessieren wir uns für eine **Darstellung von  $f$**  in der Form

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

die aus  $2^d$  Termen besteht.

Die obige Darstellung kann kompakter wie folgt geschrieben werden

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

wobei  $D = \{1, \dots, d\}$  und  $\xi^u$  zu  $\mathbb{R}^{|u|}$  gehört und nur die Komponenten  $\xi_j$  mit  $j \in u$  enthält. Hierbei bezeichnet  $|u|$  die Anzahl der Elemente der Menge  $u$ .

Im folgenden verwenden wir die Menge  $L_{2,\rho}(\mathbb{R}^d)$  aller quadratisch integrierbarer Funktionen mit der Norm

$$\|f\|_{2,\rho} = \left( \int_{\mathbb{R}^d} |f(\xi)|^2 \rho(\xi) d\xi \right)^{\frac{1}{2}},$$

und dem Skalarprodukt

$$\langle f, g \rangle_\rho = \int_{\mathbb{R}^d} f(\xi) g(\xi) \rho(\xi) d\xi.$$

Eine Darstellung der Form (\*) für eine Funktion  $f \in L_{2,\rho}(\mathbb{R}^d)$  nennt man eine **ANOVA Darstellung von  $f$**  falls

$$\int_{\mathbb{R}} f_u(\xi^u) \rho_k(\xi_k) d\xi_k = 0 \quad (\text{für alle } k \in u \text{ und } u \subseteq D).$$

Falls (\*) eine ANOVA Darstellung von  $f \in L_{2,\rho}(\mathbb{R}^d)$  ist, sind die Funktionen  $f_u$ ,  $\emptyset \neq u \subseteq D$ , orthogonal in  $L_{2,\rho}(\mathbb{R}^d)$ , d.h. es gilt

$$\langle f_u, f_v \rangle_\rho = \int_{\mathbb{R}^d} f_u(\xi) f_v(\xi) \rho(\xi) d\xi = 0 \quad \text{genau dann, wenn } u \neq v,$$

Die Funktionen  $f_u$  einer ANOVA Darstellung erlauben eine Darstellung durch sogenannte **Projektionen**, d.h. Integrale der Art

$$(P_k f)(\xi) = \int_{-\infty}^{\infty} f(\xi_1, \dots, \xi_{k-1}, s, \xi_{k+1}, \dots, \xi_d) \rho_k(s) ds \quad (\xi \in \mathbb{R}^d; k \in D)$$

Die Idee der Projektionen ist erweiterbar

$$P_u f = \left( \prod_{k \in u} P_k \right) (f) \quad (u \subseteq D).$$

Dann ist die nachfolgende Darstellung von  $f_u$  gültig

(Kuo-Sloan-Wasilkowski-Woźniakowski 10):

$$f_u = \left( \prod_{j \in u} (I - P_j) \right) P_{-u}(f) = P_{-u}(f) + \sum_{v \subsetneq u} (-1)^{|u|-|v|} P_{-v}(f),$$

d.h.  $f_u$  ist nur mittels Projektionen von  $f$  charakterisierbar

( $-u$  bezeichnet das Komplement von  $u$  bzgl.  $D$ ).

Wir betrachten die Varianzen von  $f$  und  $f_u$

$$\sigma^2(f) = \|f - I_{d,\rho}(f)\|_{2,\rho}^2 \quad \text{und} \quad \sigma_u^2(f) = \|f_u\|_{2,\rho}^2$$

und erhalten, dass sich die Varianzen der Variablengruppen zur Gesamt-Varianz aufsummieren, d.h.

$$\sigma^2(f) = \|f\|_{L_2}^2 - (I_{d,\rho}(f))^2 = \sum_{\emptyset \neq u \subseteq D} \sigma_u^2(f).$$

Die Quotienten

$$\frac{\sigma_u^2(f)}{\sigma^2(f)} \quad (u \subseteq D)$$

heissen **globale Sensitivitäts-Indizes** für die Wichtigkeit der Gruppe von Variablen  $\xi_j$ ,  $j \in u$ , innerhalb von  $f$ .

Für kleines  $\varepsilon \in (0, 1)$  nennt man

$$d_S(\varepsilon) = \min \left\{ s \in D : \sum_{|u| \leq s} \frac{\sigma_u^2(f)}{\sigma^2(f)} \geq 1 - \varepsilon \right\}$$

effektive (Superpositions-) Dimension der Funktion  $f$ .

Es gilt die Abschätzung

$$\left\| f - \sum_{|u| \leq d_S(\varepsilon)} f_u \right\|_{2,\rho} \leq \sqrt{\varepsilon} \sigma(f),$$

d.h. die Funktion  $f$  wird bereits gut durch alle ANOVA Terme  $f_u$  approximiert, falls die Anzahl  $|u|$  der Elemente von  $u$  nicht größer als die effektive Dimension  $d_S(\varepsilon)$  ist.

Sind die ANOVA Terme  $f_u$ ,  $|u| \leq d_S(\varepsilon)$ , glatter als  $f$ , so bedeutet die Abschätzung de facto eine **Glättung von  $f$** .

**Beispiel 1:** Wir betrachten die additive Function  $f(\xi) = \sum_{j=1}^d f_j(\xi_j)$  und bezeichnen mit  $\mu_j$  den Mittelwert und mit  $\gamma_j$  die Varianz von  $f_j$  bzgl. der Dichte  $\rho_j$ , d.h.

$$\mu_j = \int_{\mathbb{R}} f_j(\xi_j) \rho_j(\xi_j) d\xi_j \quad \text{und} \quad \gamma_j^2 = \int_{\mathbb{R}} (f_j(\xi_j) - \mu_j)^2 \rho_j(\xi_j) d\xi_j$$

für  $j = 1, \dots, d$ . Die Projektionen  $P_u$  of  $f$  für  $u \subseteq D$  sind

$$P_u f(\xi) = \sum_{j \in -u} f_j(\xi_j) + \sum_{j \in u} \mu_j.$$

Insbesondere gilt

$$f_{\{j\}}(\xi) = f_j(\xi_j) - \mu_j \quad (j = 1, \dots, d), \quad f_u(\xi) = 0 \text{ falls } |u| > 1,$$

und die effektive Dimension von  $f$  ist  $d_S(\varepsilon) = 1$  ( $\forall \varepsilon \in (0, 1)$ ).

**Beispiel 2:**  $f(\xi) = \sum_{i=1}^d \sum_{j=i}^d f_{ij}(\xi_i, \xi_j)$

Die effektive Dimension von  $f$  ist  $d_S(\varepsilon) = 2$  ( $\forall \varepsilon \in (0, 1)$ ).

**Beispiel 3: Multiplikative Funktionen**  $f(\xi) = \prod_{j=1}^d f_j(\xi_j)$ .

Mit  $\mu_j$  und  $\gamma_j$ ,  $j = 1, \dots, d$ , wie in Beispiel 1, erhält man

$$f_u(\xi) = \prod_{j \in u} (f_j(\xi_j) - \mu_j) \prod_{j \in -u} \mu_j$$

$$\sigma_u^2(f) = \prod_{j \in u} \gamma_j^2 \prod_{j \in -u} \mu_j^2 \quad \text{und} \quad \sigma^2(f) = \prod_{j=1}^d (\mu_j^2 + \gamma_j^2) - \prod_{j=1}^d \mu_j^2.$$

Beispiel: (Hellekalek 88)

Betrachtet man die Dichte  $\rho$  der Gleichverteilung auf  $[0, 1]^d$  und

$$f(\xi) = \prod_{j=1}^d \left( \xi_j - \frac{1}{2} \right) \quad (\xi \in [0, 1]^d),$$

so gilt  $\mu_j = 0$ ,  $\gamma_j^2 = \frac{1}{12}$ ,  $j = 1, \dots, d$ ,  $\sigma_u^2(f) = 0$ ,  $u \subsetneq D$ , sowie  $I_{d,\rho}(f) = 0$ ,  $f_D = f$  und  $d_S(\varepsilon) = d$ .

Leider ist die effektive Dimension schwer berechenbar, aber eine obere Schranke kann berechnet werden, indem das kleinste  $s \in D$ , so dass

$$\frac{1}{\sigma^2(f)} \sum_{v \subseteq \{1, \dots, s\}} \sigma_v^2(f) \geq (1 - \varepsilon)$$

erfüllt ist, berechnet wird. Dies beruht auf einer speziellen Integraldarstellung von

$$\sum_{v \subseteq u} \sigma_v^2(f) \quad (u \subseteq D),$$

wobei die auftretenden Integrale mit Hilfe von Monte Carlo bzw. Quasi-Monte Carlo Methoden basierend auf großen Samples näherungsweise berechnet werden.

Im Ernstfall kann auch der Quotient

$$\frac{\sigma_v^2(f)}{\sigma^2(f)}$$

mit beträchtlichem Aufwand für eine große Zahl von Mengen  $v \subseteq D$  mit Hilfe von Monte Carlo bzw. Quasi-Monte Carlo Methoden basierend auf großen Samples approximativ berechnet werden.

## Numerische Resultate

Als Ergebnis einer linearen Zeitreihe mit normalverteilten Innovationen ist der Zufallsvektor  $\xi$  normalverteilt mit Mittelwert  $m = \mathbb{E}[\xi]$  und Kovarianzmatrix  $\Sigma$ . Die übliche Vorgehensweise zur Realisierung von MC bzw. QMC Methoden, besteht nun darin, die Kovarianzmatrix  $\Sigma$  mit Hilfe einer geeigneten Matrix  $A$  in der Form  $\Sigma = A A^\top$  zu faktorisieren. Der Zufallsvektor  $z = (z_1, \dots, z_{2T})$  mit der Eigenschaft

$$\xi = A z + m$$

ist normalverteilt mit Mittelwert 0 und seine Komponenten besitzen Varianz 1. Bezeichnet  $\phi$  die Verteilungsfunktion der Standard-Normalverteilung, so ist der Vektor  $\eta = (\eta_1, \dots, \eta_{2T})^\top$  mit  $z_i = \phi^{-1}(\eta_i)$ ,  $i = 1, \dots, 2T$ , gleichverteilt in  $[0, 1]^{2T}$ . Die inverse Funktion  $\phi^{-1}$  kann effizient und genau berechnet werden.

Für die numerischen Tests wurde  $T = 100$ , also  $d = 2T = 200$ , gewählt. Die Matrizen  $A$ ,  $B_1$ ,  $B_2$  und die anderen Parameter des Modells wurden wie bei den Testrechnungen in (Leövey-Römisch 15) für zweistufige stochastische Planungsmodelle ohne ganzzahlige Variable in der zweiten Stufe gewählt.

Als Matrix  $A$  wurden in den numerischen Tests die Dreiecks-Matrix  $A = L_{\text{Ch}}$  der Cholesky Faktorisierung (CH) sowie die Matrix  $A = U_{\text{PCA}}$  der **Hauptkomponentenanalyse (PCA)**

$$A = U_{\text{PCA}} = (u_1 \sqrt{\lambda_1} \cdots u_d \sqrt{\lambda_d})$$

mit den **Eigenwerten**  $\lambda_1 \geq \cdots \geq \lambda_d > 0$  und **Eigenvektoren**  $u_1, \dots, u_d$  von  $\Sigma$  verwendet.

Durch Berechnung der oberen Schranken für die effektive Dimension mit  $2^{15}$  randomly scrambled Sobol' points ergaben sich  **$d_S(0.01) \leq 2$  bei Verwendung von PCA** und  **$d_S(0.01) \leq 200$  bei CH**. Detailliertere Untersuchungen ergaben bei PCA  $d_S(0.01) = 2$  und bei CH  $d_S(0.01) > 2$ . Die **Anwendung der Hauptkomponentenanalyse führt also zu einer sehr starken Reduktion der effektiven Dimension**.

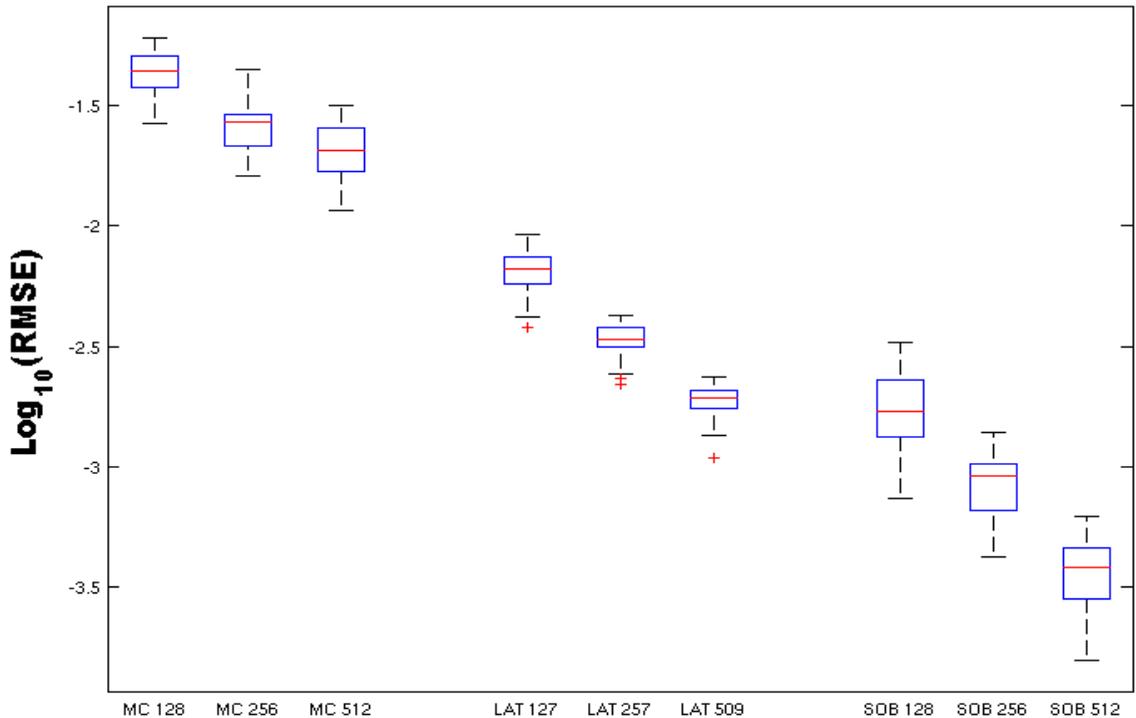
Für die numerischen Tests werden jetzt  **$n$  MC und randomisierte QMC Punkte in  $[0, 1]^d$**  für drei verschiedene Sample Anzahlen  $n$  erzeugt, die nach Transformation mit  $\phi^{-1}$  und Einsetzen in die Zeitreihe zu Samples  $\xi^j \in \mathbb{R}^d$ ,  $j = 1, \dots, n$ , führen.

Mit diesen Samples entsteht das **approximative zweistufige Optimierungsproblem**

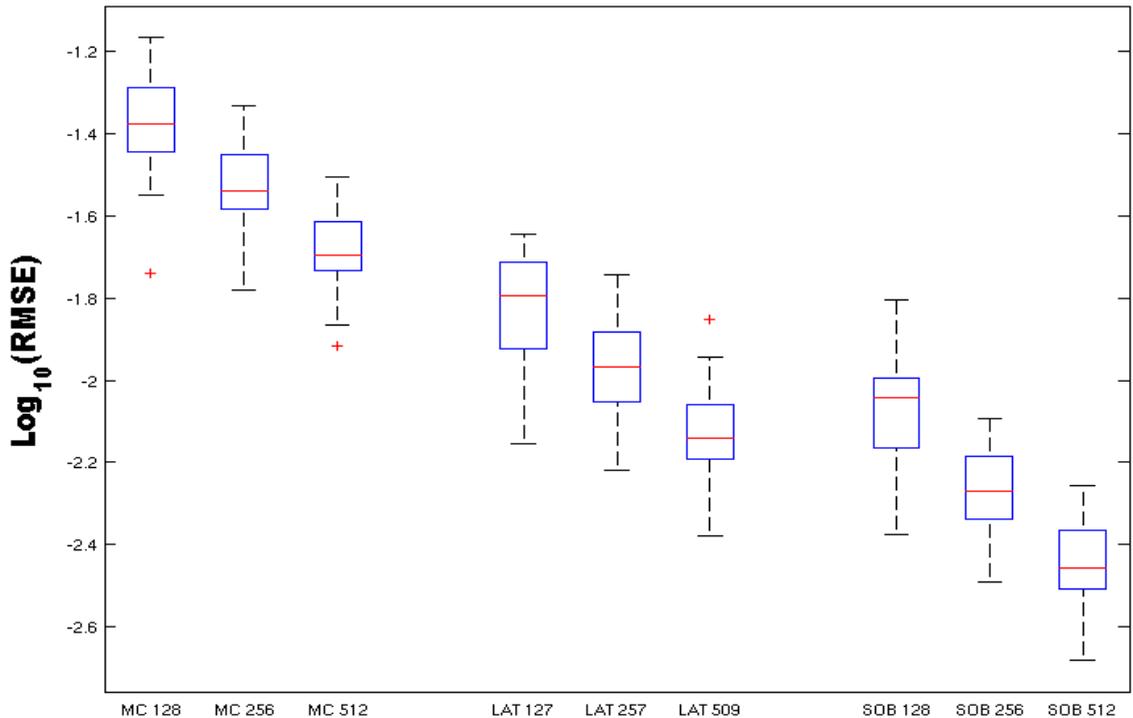
$$\min \left\{ \sum_{t=1}^T c_t^\top x_t + \frac{1}{n} \sum_{j=1}^n \Phi(q(\xi^j), h(\xi^j) - Vx) : x \in X \right\},$$

in dem  $\Phi$  die **Optimalwertfunktion des gemischt-ganzzahligen linearen Optimierungsproblems der zweiten Stufe bezeichnet**. Sie ist stetig auf gewissen polyedrischen Mengen und hat eventuell Sprungstellen an den Rändern dieser Polyeder wegen des Vorhandenseins der 0-1-Entscheidungen.

Für MC points in  $[0, 1]^d$  wird der **Mersenne Twister** als Generator für unabhängige Pseudo-Zufallszahlen und  $n = 128, 256, 512$  verwendet. Als randomisierte QMC Methoden verwenden wir **randomly scrambled Sobol' points** für  $n = 128, 256, 512$  und **randomly shifted lattice rules** mit  $n = 127, 257, 509$  (da Primzahlen  $n$  für diese Methoden günstiger sind). Dabei wird das scrambling mit Hilfe des affinen scrambling realisiert. Hierbei und beim Generieren der random shifts wird ebenfalls der Mersenne Twister eingesetzt.



Gezeigt wird  $\text{Log}_{10}$  des relativen Quadratmittelfehlers für die optimalen Kosten des zweistufigen Modells unter Verwendung der PCA Faktorisierung der Kovarianzmatrix. Resultate für Mersenne Twister MC und randomly scrambled Sobol' (SOB) mit  $n = 128, 256$  und  $512$  Punkten und randomly shifted lattice rules (LAT) mit  $n = 127, 257$  und  $509$  Punkten.



Gezeigt wird  $\text{Log}_{10}$  des relativen Quadratmittelfehlers der optimalen Kosten des zweistufigen Modells unter Verwendung der Cholesky (CH) Faktorisierung der Kovarianzmatrix. Resultate für Merseenne Twister MC und randomly scrambled Sobol' (SOB) mit  $n = 128, 256$  und  $512$  Punkten und randomly shifted lattice rules (LAT) mit  $n = 127, 257$  und  $509$  Punkten.

Geschätzt wird der relative Fehler der Wurzel aus dem Quadratmittel (RMSE) der optimalen Kosten des zweistufigen Modells durch Realisieren von 10 Versuchen für jedes Experiment und von 30 Wiederholungen. Die Box-plots in beiden Abbildungen entstehen dadurch, dass die untere bzw. obere Seite der Boxen durch das erste und dritte Quartil der Schätzungen gebildet werden. Der Median entspricht der Linie zwischen unterer und oberer Seite. Ausreißer, die nicht in den Boxen liegen, werden mit + markiert.

Für alle drei Methoden wurden die **Konvergenzraten** geschätzt. Im Mittel ergab sich für MC die theoretisch erwartete Rate von  $-0.5$ , bei Verwendung von PCA zur Faktorisierung etwa  $-0.9$  für randomly shifted lattice rules und  $-1.0$  für randomly scrambled Sobol' points.

Bei Verwendung von CH ergaben sich für alle drei Methoden als mittlere Konvergenzrate etwa  $-0.5$ . Die Konvergenzraten aller Methoden sind optisch in der logarithmischen Skala beider Abbildungen zu erkennen. Deutlich wird aber bei CH ebenfalls, dass **die beiden randomisierten QMC Methoden zu kleineren Fehlern als MC** führen.

Eine **Erklärung für das deutlich bessere Verhalten** beider randomisierter QMC Methoden unter **Verwendung von PCA** ist die durch die **niedrige effektive Dimension** erreichte Glättung des Integranden.

## Zusammenfassung

- Moderne randomisierte Quasi-Monte Carlo Methoden besitzen deutliche Vorteile im Vergleich zu Monte Carlo Methoden insbesondere bei niedriger effektiver Dimension der Integranden.
- Die Vorteile bestehen in der besseren Genauigkeit für gegebene Anzahlen von Samples bzw. Szenarios oder in der Möglichkeit kleinere Anzahlen von Samples für dieselbe Genauigkeit verwenden zu können und damit kürzere Rechenzeiten zu erreichen.  
Statt  $10^4$  Samples bei MC nur gut  $10^2$  Samples bei QMC.
- Dies ist auch theoretisch abgesichert für stochastische Optimierungsprobleme ohne Ganzzahligkeitsbedingungen (Leövey-Römisch 15).
- Die vorgestellten numerischen Ergebnisse zeigen diese Vorteile auch für gemischt-ganzzahlige stochastische Optimierungsmodelle der Energiewirtschaft und lassen eine Erweiterung der Theorie erhoffen.
- Die gleichen Vorteile sind auch bei Anwendung der Methoden im integrierten Risikomanagement zu erwarten.

## Literatur:

J. Dick, F. Y. Kuo and I. H. Sloan: High-dimensional integration – the Quasi-Monte Carlo way, *Acta Numerica* 22 (2013), 133–288.

J. Dick and F. Pillichshammer: *Digital Nets and Sequences*, Cambridge University Press, Cambridge 2010.

A. Eichhorn, W. Römisch and I. Wegner: Mean-risk optimization of electricity portfolios using multiperiod polyhedral risk measures. IEEE St. Petersburg PowerTech Proceedings (2005).

A. Eichhorn, H. Heitsch and W. Römisch: Stochastic optimization of electricity portfolios: Scenario tree modeling and risk management, in *Handbook of Power Systems* (S. Rebennack, P.M. Pardalos, M. V. F. Pereira, N. A. Iliadis Eds.), Springer, 2010, 405–432.

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 and H. Woźniakowski: On decomposition of multivariate functions, *Mathematics of Computation* 79 (2010), 953–966.

P. L'Ecuyer and Ch. Lemieux: Recent advances in randomized quasi-Monte Carlo methods, in *Modeling Uncertainty* (M. Dror, P. L'Ecuyer, F. Szidarovski eds.), Kluwer, Boston, 2002, 419–474.

G. Leobacher and F. Pillichshammer: *Introduction to Quasi-Monte Carlo Integration and Applications*, Birkhäuser, Heidelberg, 2014.

H. Leövey and W. Römisch: Quasi-Monte Carlo methods for linear two-stage stochastic programming problems, *Mathematical Programming* 151 (2015), 315–345.

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.

- 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.
- A. B. Owen: Randomly permuted  $(t, m, s)$ -nets and  $(t, s)$ -sequences, in: *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing* (H. Niederreiter and P. J.-S. Shiue eds.), Lecture Notes in Statistics, Vol. 106, Springer, New York, 1995, 299–317.
- A. B. Owen: The dimension distribution and quadrature test functions, *Statistica Sinica* 13 (2003), 1–17.
- G. Ch. Pflug and W. Römisch: *Modeling, Measuring and Managing Risk*, World Scientific, Singapore, 2007.
- 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.
- I. M. Sobol': The distribution of points in a cube and the approximate evaluation of integrals, *U.S.S.R. Comput. Math. and Math. Phys.* 7 (1967), 86–112.
- X. Wang and K.-T. Fang: The effective dimension and Quasi-Monte Carlo integration, *Journal of Complexity* 19 (2003), 101–124.
- X. Wang and I. H. Sloan: Quasi-Monte Carlo methods in financial engineering: An equivalence principle and dimension reduction, *Operations Research* 59 (2011), 80–95.