Lifting the Curse of Dimensionality — Numerical Integration in Very High Dimensions

Ian Sloan

i.sloan@unsw.edu.au

University of New South Wales, Sydney, Australia
Richard Bellman coined the phrase ‘curse of dimensionality’ to describe the extraordinarily rapid increase in the difficulty of problems as the number of variables increases.
Consider
\[ \int_0^1 \cdots \int_0^1 f(x^1, \ldots, x^{360}) \, dx^1 \cdots dx^{360} \]

In 1995 Traub and Paskov computed values of ‘mortgage-backed obligations’ - in the US mortgages last for 30 years, and may be repaid each month, making \(30 \times 12 = 360\) repayment possibilities. The computed quantity is a 360-dimensional expected value.

They used a ‘quasi-Monte Carlo’ method.
The Problem after Transformation

Problem after transformation to unit cube:

\[ I_s(f) := \int_{[0,1]^s} f(x) \, dx. \]

Message from the finance experiments: ‘Quasi-Monte Carlo’ methods can sometimes work better than classical Monte Carlo methods.
Monte Carlo

\[ M_{n,s}(f) = \frac{1}{n} \sum_{k=1}^{n} f(t_k), \]

with \( t_1, \ldots, t_n \) chosen randomly from a uniform distribution on \([0, 1]^s\).

**Error:** For \( f \in L^2([0, 1]^s) \),

\[
\text{error} = \frac{\sigma(f)}{\sqrt{n}},
\]

where

\[
\sigma^2(f) = I_s((f - I_s(f))^2) = I_s(f^2) - (I_s(f))^2.
\]
Quasi-Monte Carlo

$Q_{n,s}(f) = \frac{1}{n} \sum_{k=1}^{n} f(t_k)$,

with $t_1, \ldots, t_n$ deterministic (and cleverly chosen).

How to choose $t_1, \ldots, t_n$?
Quasi-Monte Carlo

\[ Q_{n,s}(f) = \frac{1}{n} \sum_{k=1}^{n} f(t_k), \]

with \( t_1, \ldots, t_n \) deterministic (and cleverly chosen).

How to choose \( t_1, \ldots, t_n \)?

- Low discrepancy points
  
  (Sobol, Faure, Niederreiter, . . . )
Quasi-Monte Carlo

\[ Q_{n,s}(f) = \frac{1}{n} \sum_{k=1}^{n} f(t_k), \]

with \( t_1, \ldots, t_n \) deterministic (and cleverly chosen).

How to choose \( t_1, \ldots, t_n \)?

- Low discrepancy points
  (Sobol, Faure, Niederreiter, \ldots)
- Lattice rules
Product Grid

NOTE:

A ‘product’ grid is very bad for large $s$.

For example, if there are

$$n = N^s$$

points, and if

$$f(x^1, \ldots, x^s) = x^1,$$

then the 1-$d$ rectangle rule gives

$$\text{ERROR} = O\left(\frac{1}{N}\right)$$
NOTE:

A ‘product’ grid is very bad for large $s$.

For example, if there are

$$n = N^s$$

points, and if

$$f(x^1, \ldots, x^s) = x^1,$$

then the 1-$d$ rectangle rule gives

$$\text{ERROR} = O\left(\frac{1}{N}\right) = O\left(\frac{1}{n^{1/s}}\right)!$$

This is the ‘Curse of Dimensionality’.
Koksma-Hlawka inequality is

\[ |I_s(f) - Q_{n,s}(f)| \leq D^*(t_1, \ldots, t_n)V(f) \]

(This implies that we should make \( D^* \), the ‘star discrepancy’, small!)
Star Discrepancy $D^*(t_1, \ldots, t_n)$

The **star discrepancy** is the supremum, over all red boxes, of

$$\left| \frac{\text{number of QMC points in the box}}{\text{total number of points}} - \text{volume of the box} \right|.$$
Star Discrepancy $D^*(t_1, \ldots, t_n)$

The star discrepancy is the supremum, over all red boxes, of

$$\frac{\text{number of QMC points in the box}}{\text{total number of points}} - \text{volume of the box}.$$

For the point $t_1 = (0.6, 0.5)$:

$$\left| \frac{5}{13} - 0.6 \times 0.5 \right| = 0.0846$$

For the point $t_1 = (0.8, 0.9)$:

$$\left| \frac{10}{13} - 0.8 \times 0.9 \right| = 0.0492$$
A LOW-DISCREPANCY sequence $t_1, t_2, \ldots$ of points in $[0, 1]^s$ is one for which the first $n$ members satisfy

$$D^*(t_1, \ldots, t_n) \leq C_s \frac{(\log n)^s}{n}.$$ 

For fixed $s$ the resulting error bound is “better than MC” if $n$ is large enough.
A LOW-DISCREPANCY sequence $t_1, t_2, \ldots$ of points in $[0, 1]^s$ is one for which the first $n$ members satisfy

$$D^*(t_1, \ldots, t_n) \leq C_s \frac{(\log n)^s}{n}.$$ 

For fixed $s$ the resulting error bound is “better than MC” if $n$ is large enough.

But the bound on the right grows with $n$ until $n \approx e^s$. 
Random vs Sobol points

A comparison of 64 (pseudo)random and 64 Sobol points ...
Other low discrepancy sequences are due to

Faure

Niederreiter

Xing
Lattice Rules

Lattice rules for periodic functions
Korobov, 1959 and Hlawka, 1961
L K Hua and Y Wang
Frolov, 1977
S. & Kachoyan, 1984
Lyness
Sloan & Joe (book), 1994

Lattice rules for non-periodic functions
L’Ecuyer, .... , in \( L_2 \) context
S. & Woźniakowski, Hickernell, Joe, Kuo, Dick, Wang, Waterhouse, ..... , from 2001
**Lattice Rule Definition**

**Lattice Rule** (of rank 1)

\[
Q_{n,s}f = \frac{1}{n} \sum_{k=1}^{n} f \left( \left\{ k \frac{z}{n} \right\} \right),
\]

\[z \in \{1, \ldots, n - 1\}^s\]

**Shifted lattice rule**

\[
Q_{n,s}f = \frac{1}{n} \sum_{k=1}^{n} f \left( \left\{ k \frac{z}{n} + \Delta \right\} \right),
\]

\[\Delta \text{ (the "shift")} \in [0, 1]^s\]

Example of Lattice & Shifted Lattice Rules

\[ n = 34, \ z = (1, 21) \]

\[ n = 34, \ z = (1, 21), \ \Delta = (0.8, 0.1) \]
Questions we might ask

WHAT IS THE CRITERION FOR A ‘GOOD’ LATTICE?

(We need to define a function space setting.)
Questions we might ask

WHAT IS THE CRITERION FOR A ‘GOOD’ LATTICE?
(We need to define a function space setting.)

DO ‘GOOD’ LATTICES EXIST WHEN $s$ IS LARGE?
(Yes, we can prove existence.)
Questions we might ask

WHAT IS THE CRITERION FOR A ‘GOOD’ LATTICE?
(We need to define a function space setting.)

DO ‘GOOD’ LATTICES EXIST WHEN $s$ IS LARGE?
(Yes, we can prove existence.)

AND IF THEY EXIST, CAN WE CONSTRUCT GOOD ONES?
The Function Space Setting

Weighted spaces

(I.H. Sloan & H. Woźniakowski, J. of Complexity ’98)
‘Weighted’ spaces are function spaces in which the successive coordinate directions $x^1, x^2, \ldots$ are weighted by parameters $\gamma_1, \gamma_2, \ldots$, with

$$\gamma_1 \geq \gamma_2 \geq \cdots > 0,$$

BECAUSE IN SOME PROBLEMS THE VARIABILITY OF $f$ IS:

$\bullet$ greatest in $x^1$ direction
The Function Space Setting

Weighted spaces

(I.H. Sloan & H. Woźniakowski, J. of Complexity ’98)
‘Weighted’ spaces are function spaces in which the successive coordinate directions $x^1, x^2, \ldots$ are weighted by parameters $\gamma_1, \gamma_2, \ldots$, with

$$\gamma_1 \geq \gamma_2 \geq \cdots > 0,$$

BECAUSE IN SOME PROBLEMS THE VARIABILITY OF $f$ IS:

- greatest in $x^1$ direction
- less in $x^2$ direction
The Function Space Setting

Weighted spaces

(I.H. Sloan & H. Woźniakowski, J. of Complexity ‘98)
‘Weighted’ spaces are function spaces in which the successive coordinate directions $x^1, x^2, \ldots$ are weighted by parameters $\gamma_1, \gamma_2, \ldots$, with

$$\gamma_1 \geq \gamma_2 \geq \cdots > 0,$$

BECAUSE IN SOME PROBLEMS THE VARIABILITY OF $f$ IS:

- greatest in $x^1$ direction
- less in $x^2$ direction
- still less in $x^3$ direction
The Function Space Setting

Weighted spaces

(I.H. Sloan & H. Woźniakowski, J. of Complexity ’98)

‘Weighted’ spaces are function spaces in which the successive coordinate directions $x^1, x^2, \ldots$ are weighted by parameters $\gamma_1, \gamma_2, \ldots$, with

$$\gamma_1 \geq \gamma_2 \geq \cdots > 0,$$

Because in some problems the variability of $f$ is:

- greatest in $x^1$ direction
- less in $x^2$ direction
- still less in $x^3$ direction
- \ldots
The Spaces

The spaces $H_{s,\gamma}$ are Hilbert spaces, of non-periodic functions, with norms $\|f\|_{s,\gamma}$:

- $s = 1$:
  $$\|f\|_{1,\gamma}^2 = |f(1)|^2 + \frac{1}{\gamma_1} \int_0^1 \left| \frac{df}{dx} \right|^2 dx$$
The Spaces

The spaces $H_{s,\gamma}$ are Hilbert spaces, of non-periodic functions, with norms $\|f\|_{s,\gamma}$:

- **$s = 1$**: $\|f\|_{1,\gamma}^2 = |f(1)|^2 + \frac{1}{\gamma_1} \int_0^1 |\frac{df}{dx}|^2 dx$
- **$s = 2$**: $\|f\|_{2,\gamma}^2 = |f(1, 1)|^2$

\[
+ \frac{1}{\gamma_1} \int_0^1 \left| \frac{\partial f}{\partial x^1}(x^1, 1) \right|^2 dx^1 + \frac{1}{\gamma_2} \int_0^1 \left| \frac{\partial f}{\partial x^2}(1, x^2) \right|^2 dx^2
+ \frac{1}{\gamma_1 \gamma_2} \int_0^1 \int_0^1 \left| \frac{\partial^2 f}{\partial x^1 \partial x^2} \right|^2 dx^1 dx^2
\]
The spaces $H_{s,\gamma}$ are Hilbert spaces, of non-periodic functions, with norms $\|f\|_{s,\gamma}$:

- $s = 1:\quad \|f\|_{1,\gamma}^2 = |f(1)|^2 + \frac{1}{\gamma_1} \int_0^1 \left| \frac{df}{dx} \right|^2 dx$

- $s = 2:\quad \|f\|_{2,\gamma}^2 = |f(1,1)|^2$

$$+ \frac{1}{\gamma_1} \int_0^1 \left| \frac{\partial f}{\partial x_1} (x_1,1) \right|^2 dx_1 + \frac{1}{\gamma_2} \int_0^1 \left| \frac{\partial f}{\partial x_2} (1,x_2) \right|^2 dx_2$$

$$+ \frac{1}{\gamma_1 \gamma_2} \int_0^1 \int_0^1 \left| \frac{\partial^2 f}{\partial x_1 \partial x_2} \right|^2 dx_1 dx_2$$

- $s = 3:\quad \cdots$
Reproducing Kernel Property

The special feature is that $H_{s,\gamma}$ is a tensor product, reproducing-kernel Hilbert space with a simple kernel.

Reproducing kernel property:

$$(f, K_{s,\gamma}(x, \cdot))_{s,\gamma} = f(x) \quad \forall x \in [0, 1]^s.$$ 

Here the kernel is

$$K_{s,\gamma}(x, y) = \prod_{j=1}^{s} (1 + \gamma_j \min(1 - x^j, 1 - y^j)).$$
Existence of a Good QMC Rule

This is expressed in terms of the ‘worst-case-error’ in $H_{s,\gamma}$:

$$e_{n,s,\gamma}(t_1, \ldots, t_n) :=$$

$$\sup \left\{ \left| I_s(f) - \frac{1}{n} \sum_{k=1}^{n} f(t_k) \right| : \| f \|_{s,\gamma} \leq 1 \right\}$$
Existence of a Good QMC Rule

This is expressed in terms of the ‘worst-case-error’ in $H_{s,\gamma}$:

$$e_{n,s,\gamma}(t_1, \ldots, t_n) := \sup \left\{ \left| I_s(f) - \frac{1}{n} \sum_{k=1}^{n} f(t_k) \right| : \|f\|_{s,\gamma} \leq 1 \right\}$$

Main result (Existence)

**THEOREM** Sloan & Woźniakowski, ’98

If $\sum_{j=1}^{\infty} \gamma_j < \infty$ then there exist points $t_1, \ldots, t_n \in [0, 1]^s$ such that

$$e_{n,s,\gamma}(t_1, \ldots, t_n) \leq \frac{D \gamma}{\sqrt{n}}$$
Remarks

1. Result holds e.g. for $\gamma_j = 1/j^2$, not for $\gamma_j = 1$. 
Remarks

1. Result holds e.g. for $\gamma_j = 1/j^2$, not for $\gamma_j = 1$.

2. The condition $\sum_{j=1}^{\infty} \gamma_j < \infty$ is necessary as well as sufficient.
Remarks

1. Result holds e.g. for $\gamma_j = 1/j^2$, not for $\gamma_j = 1$.

2. The condition $\sum_{j=1}^{\infty} \gamma_j < \infty$ is necessary as well as sufficient.

3. $D_\gamma$ is known explicitly: e.g.

$$D_\gamma = \exp \left( \frac{1}{4} \sum_{j=1}^{\infty} \gamma_j \right)$$
Remarks

1. Result holds e.g. for $\gamma_j = 1/j^2$, not for $\gamma_j = 1$.

2. The condition $\sum_{j=1}^{\infty} \gamma_j < \infty$ is necessary as well as sufficient.

3. $D_\gamma$ is known explicitly: e.g.

$$D_\gamma = \exp \left( \frac{1}{4} \sum_{j=1}^{\infty} \gamma_j \right)$$

4. Proof is not constructive.
Proof idea

Let $M_{n,s,\gamma}$ be the RMS average of $e_{n,s,\gamma}$ over all possible choices of $t_1, \ldots, t_n$. It can be shown that

$$M_{n,s,\gamma}^2 = \frac{1}{n} \left( \prod_{j=1}^{s} \left( 1 + \frac{\gamma_j}{2} \right) - \prod_{j=1}^{s} \left( 1 + \frac{\gamma_j}{3} \right) \right)$$
Proof idea

Let $M_{n,s,\gamma}$ be the RMS average of $e_{n,s,\gamma}$ over all possible choices of $t_1, \ldots, t_n$. It can be shown that

$$M_{n,s,\gamma}^2 = \frac{1}{n} \left( \prod_{j=1}^{s} \left( 1 + \frac{\gamma_j}{2} \right) - \prod_{j=1}^{s} \left( 1 + \frac{\gamma_j}{3} \right) \right)$$

$$\leq \frac{1}{n} \exp \left( \frac{1}{2} \sum_{j=1}^{s} \gamma_j \right).$$
Proof idea

Let \( M_{n,s,\gamma} \) be the RMS average of \( e_{n,s,\gamma} \) over all possible choices of \( t_1, \ldots, t_n \). It can be shown that

\[
M^2_{n,s,\gamma} = \frac{1}{n} \left( \prod_{j=1}^{s} \left( 1 + \frac{\gamma_j}{2} \right) - \prod_{j=1}^{s} \left( 1 + \frac{\gamma_j}{3} \right) \right)
\]

\[
\leq \frac{1}{n} \exp \left( \frac{1}{2} \sum_{j=1}^{s} \gamma_j \right).
\]

Finally, there exists at least one choice \( t_1, \ldots, t_n \) that is as good as average!
Worst-Case Error in a RKHS

The key fact behind that last result is the following expression for the worst case error in any reproducing kernel space $H(K)$:

$$e_{n,s,\gamma}(t_1, \ldots, t_n)^2 = \int\int_{[0,1]^{2s}} K(x, y) dx \, dy$$

$$-2 \frac{1}{n} \sum_{k=1}^{n} \int_{[0,1]^s} K(x, t_k) dx$$

$$+ \frac{1}{n^2} \sum_{k=1}^{n} \sum_{k' = 1}^{n} K(t_k, t_{k'})$$
Worst-Case Error in a RKHS

The key fact behind that last result is the following expression for the worst case error in any reproducing kernel space $H(K)$:

$$e_{n,s,\gamma}(t_1, \ldots, t_n)^2 = \int \int [0,1]^{2s} K(x,y) dx \, dy$$

$$-2 \frac{1}{n} \sum_{k=1}^{n} \int [0,1]^s K(x, t_k) dx$$

$$+ \frac{1}{n^2} \sum_{k=1}^{n} \sum_{k'=1}^{n} K(t_k, t_{k'})$$

Thus

$$M_{n,s,\gamma}^2 = \int [0,1]^s \cdots \int [0,1]^s e_{n,s,\gamma}(t_1, \ldots, t_s)^2 dt_1 \cdots dt_s = \cdots$$
Faster Convergence than the MC Rate

**THEOREM** Hickernell & Woźniakowski (’99):

If

\[
\sum_{j=1}^{\infty} \gamma_j^{1/2} < \infty \quad \left( \text{e.g. } \gamma_j = \frac{1}{j^3} \right)
\]

then \( \exists t_1, \ldots, t_n \in [0, 1]^s \) such that \( \forall \delta > 0 \)

\[
e_{n, s, \gamma}(t_1, \ldots, t_n) \leq \frac{C_{\gamma, \delta}}{n^{1-\delta}}.
\]

Proof is again not constructive.
THEOREM Sloan and Woźniakowski ('01):

\[
\sum_{j=1}^{\infty} \gamma_j^{1/2} < \infty,
\]

and if \( n \) is prime, then \( \exists \) a SHIFTED LATTICE RULE \( L_{n, s} \) such that

\[
e_{n, s, \gamma}(L_{n, s}) \leq \frac{C_{\gamma, \delta}}{n^{1-\delta}} \quad \forall \delta > 0.
\]

A shifted lattice rule (with \( n \) prime) is a QMC rule of special form

\[
L_{n, s}(f) = \frac{1}{n} \sum_{k=1}^{n} f \left( \left\{ \frac{kz}{n} + \Delta \right\} \right),
\]

\( z \in \{1, 2, \ldots, n - 1\}^s, \Delta \in [0, 1)^s \)
Towards construction

**THEOREM** Sloan and Woźniakowski ('01):

If

\[ \sum_{j=1}^{\infty} \gamma_j^{1/2} < \infty, \]

and if \( n \) is prime, then \( \exists \) a SHIFTED LATTICE RULE \( L_{n,s} \) such that

\[ e_{n,s,\gamma}(L_{n,s}) \leq \frac{C_{\gamma,\delta}}{n^{1-\delta}} \quad \forall \delta > 0. \]

A shifted lattice rule (with \( n \) prime) is a QMC rule of special form

\[ L_{n,s}(f) = \frac{1}{n} \sum_{k=1}^{n} f \left( \left\{ \frac{kz}{n} + \Delta \right\} \right), \]

\[ z \in \{1, 2, \ldots, n-1\}^s, \Delta \in [0, 1)^s \]

Proof is by averaging in a different way over \( \Delta \) and \( z \), then using

Jensen's inequality.
CONSTRUCTION!
The original idea of component-by component (or CBC) construction was to construct $z$ and $\Delta$ one component at a time:

If $z = (z^1, \ldots, z^s), \quad \Delta = (\Delta^1, \ldots, \Delta^s),$

we construct successively

$$z^1, \Delta^1, z^2, \Delta^2, z^3, \Delta^3, \ldots.$$
Original CBC Algorithm - sketch

For $s = 2, \ldots, s_{\text{max}}$, $z^1 = 1$, $\Delta^1 = \ldots$.

a) Choose $z^s$ from $\{1, 2, \ldots, n - 1\}$ to minimise

\[
\text{average of } e_{n,s,\gamma}^2(z, \Delta) \text{ over } \Delta^s
\]

b) Choose $\Delta^s$ from $\left\{ \frac{1}{2n}, \frac{3}{2n}, \ldots, \frac{2n-1}{2n} \right\}$ to minimise

\[
e_{n,s,\gamma}^2(z, \Delta)
\]

[The cost of computing $e_{n,s,\gamma}^2(z, \Delta)$ is $O(n^2 \delta)$. The cost of computing its average over $\Delta^s$ is the same.]
THEOREM: If $\sum_{j=1}^{\infty} \gamma_j < \infty$, then with $z = (z^1, \ldots, z^s)$ and $\Delta = (\Delta^1, \ldots, \Delta^s)$ calculated by the algorithm, we have

$$e_{n,s,\gamma}(z, \Delta) \leq \frac{D\gamma}{\sqrt{n}}.$$
THEOREM: If $\sum_{j=1}^{\infty} \gamma_j < \infty$, then with $z = (z^1, \ldots, z^s)$ and $\Delta = (\Delta^1, \ldots, \Delta^s)$ calculated by the algorithm, we have

$$e_{n,s,\gamma}(z, \Delta) \leq \frac{D_{\gamma}}{\sqrt{n}}.$$ 

Proof: By induction on $s$. At each step the result holds for the RMS average over both $\Delta^s$ and $z^s$. 
THEOREM: If \( \sum_{j=1}^{\infty} \gamma_j < \infty \), then with \( z = (z^1, \ldots, z^s) \) and \( \Delta = (\Delta^1, \ldots, \Delta^s) \) calculated by the algorithm, we have

\[
    e_{n,s,\gamma}(z, \Delta) \leq \frac{D\gamma}{\sqrt{n}}.
\]

Proof: By induction on \( s \). At each step the result holds for the RMS average over both \( \Delta^s \) and \( z^s \).

Cost of evaluating all components

\[
    z^1, \Delta^1, z^2, \Delta^2, \ldots, z^s, \Delta^s
\]

is

\[
    O(n^3 s^2).
\]
THERE ARE STILL SOME THINGS TO CRITICISE:

- Though the $O(n^3 s^2)$ cost is only polynomial in $n$, it is still too high. (We stopped at $n \approx 4,000$, $s = 40$.)
THERE ARE STILL SOME THINGS TO CRITICISE:

- Though the $O(n^3 s^2)$ cost is only polynomial in $n$, it is still too high. (We stopped at $n \approx 4,000, s = 40$.)

- The known error bound for this algorithm is still no better than Monte Carlo.
THERE ARE STILL SOME THINGS TO CRITICISE:

- Though the $O(n^3s^2)$ cost is only polynomial in $n$, it is still too high. (We stopped at $n \approx 4,000$, $s = 40$.)

- The known error bound for this algorithm is still no better than Monte Carlo.

- This method (LIKE OTHER QMC METHODS, BUT UNLIKE THE MC METHOD) gives no idea of the actual integration error.
There are advantages if we work instead with

RANDOMLY SHIFTED LATTICE RULES
Randomly Shifted Lattice Rules


Define

\[ L_{n,s}(z, \Delta; f) := \frac{1}{n} \sum_{k=0}^{n-1} f \left( \left\{ k \frac{z}{n} + \Delta \right\} \right) \]

where \( z \) is fixed BUT NOW \( \Delta \) IS CHOSEN FROM A UNIFORM DISTRIBUTION ON \([0, 1)^s\).

This family is an unbiased estimator of \( I_s(f) \)!
Randomly Shifted Lattice Rules


Define

\[ L_{n,s}(z, \Delta; f) := \frac{1}{n} \sum_{k=0}^{n-1} f \left( \left\{ \frac{kz}{n} + \Delta \right\} \right) \]

where \( z \) is fixed BUT NOW \( \Delta \) IS CHOSEN FROM A UNIFORM DISTRIBUTION ON \([0, 1)^s\).

This family is an unbiased estimator of \( I_s(f) \)!

And now we do not need to compute the shift \( \Delta \)
In practice we compute

\[ \bar{L}_{n,s}(z; f) = \frac{1}{q} \sum_{i=1}^{q} L_{n,s}(z, \Delta_i; f) \]

where \( \Delta_1, \ldots, \Delta_q \) are chosen randomly from a uniform distribution on \( [0, 1)^s \), and \( q \) is, say, 10.

Now the approximation has Monte Carlo features - the error can be estimated: The error estimate is

\[ \frac{1}{\sqrt{q}} \left[ \frac{1}{q - 1} \sum_{i=1}^{q} \left( L_{n,s}(z, \Delta_i; f) - \bar{L}_{n,s}(z; f) \right)^2 \right]^{1/2}. \]
CBC Algorithm for Random Shifts

The new algorithm constructs $z^1, z^2, \ldots$ from $\{1, 2, \ldots, n - 1\}^s$, one component at a time, for use in randomly shifted lattice rules.
CBC Algorithm for Random Shifts

The new algorithm constructs $z^1, z^2, \ldots$ from $\{1, 2, \ldots, n - 1\}^s$, one component at a time, for use in randomly shifted lattice rules.

At each stage we minimise the worst case error averaged over shifts.
CBC Algorithm for Random Shifts

The new algorithm constructs $z^1, z^2, \ldots$ from $\{1, 2, \ldots, n - 1\}^s$, one component at a time, for use in randomly shifted lattice rules.

At each stage we minimise the worst case error averaged over shifts.

This is fast, because (as we shall see) the average over shifts,

$$ e_{n,s,\gamma}^{\text{rms}}(z) := \left( \int_{[0,1]^s} e_{n,s,\gamma}^2(z, \Delta) d\Delta \right)^{\frac{1}{2}}, $$

can be computed with cost only $O(ns)$. 
The CBC Algorithm for Random Shifts

- $z^1 = 1$

- For $s = 2, 3, \ldots, s_{\text{max}}$, choose $z^s$ from
  \[ \{1, 2, \ldots, n - 1\} \]
  so as to minimise
  \[
e_{n,s,\gamma}^{\text{rms}}(z^1, z^2, \ldots, z^s)^2 = -\prod_{j=1}^{s} (1 + \gamma_j / 3) \]
  \[+ \frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{s} \left[ 1 + \gamma_j \left( B_2(\{iz_j/n\}) + \frac{1}{3} \right) \right] \]

Here $B_2(x) = x^2 - x + \frac{1}{6}$.

THE TOTAL COST of a naive implementation IS ONLY $O(n^2 s_{\text{max}}^2)$. 
Error of CBC with Random Shifts

**THEOREM** Sloan, Kuo, Joe, 2002

Let $n$ be prime, and assume $\sum_{j=1}^{\infty} \gamma_j < \infty$.

If $z^1, \ldots, z^s$ are chosen by the new algorithm then

$$ e_{n,s,\gamma}(z) \leq \frac{D_\gamma}{n^{1/2}} $$
Error of CBC with Random Shifts

**THEOREM** Sloan, Kuo, Joe, 2002

Let \( n \) be prime, and assume \( \sum_{j=1}^{\infty} \gamma_j < \infty \).

If \( z^1, \ldots, z^s \) are chosen by the new algorithm then

\[
e_{n,s,\gamma}^{\text{rms}}(z) \leq \frac{D_\gamma}{n^{1/2}}
\]

As before,

\[
D_\gamma = \left( \prod_{j=1}^{\infty} \left( 1 + \frac{\gamma_j}{2} \right) - \prod_{j=1}^{\infty} \left( 1 + \frac{\gamma_j}{3} \right) \right)^{1/2} \leq \exp \left( \frac{1}{4} \sum_{j=1}^{\infty} \gamma_j \right)
\]
THEOREM Sloan, Kuo, Joe, 2002

Let \( n \) be prime, and assume \( \sum_{j=1}^{\infty} \gamma_j < \infty \).

If \( z^1, \ldots, z^s \) are chosen by the new algorithm then

\[
e_{\text{rms}}^{n,s,\gamma}(z) \leq \frac{D_\gamma}{n^{1/2}}
\]

As before,

\[
D_\gamma = \left( \prod_{j=1}^{\infty} \left( 1 + \frac{\gamma_j}{2} \right) - \prod_{j=1}^{\infty} \left( 1 + \frac{\gamma_j}{3} \right) \right)^{1/2} \leq \exp \left( \frac{1}{4} \sum_{j=1}^{\infty} \gamma_j \right)
\]

Proof: By induction, using an averaging argument at each step.
EXAMPLE: \( \gamma_j = \frac{1}{j^2}, \ j = 1, 2, \ldots \)

<table>
<thead>
<tr>
<th>( s )</th>
<th>( z^j )</th>
<th>( e_{n,s,\gamma}^{\text{rms}}(z) )</th>
<th>( z^j )</th>
<th>( e_{n,s,\gamma}^{\text{rms}}(z) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2.04(-4)</td>
<td>1</td>
<td>1.28(-5)</td>
</tr>
<tr>
<td>2</td>
<td>765</td>
<td>3.02(-4)</td>
<td>9376</td>
<td>2.03(-5)</td>
</tr>
<tr>
<td>3</td>
<td>605</td>
<td>3.72(-4)</td>
<td>11835</td>
<td>2.58(-5)</td>
</tr>
<tr>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
</tr>
<tr>
<td>31</td>
<td>450</td>
<td>6.51(-4)</td>
<td>13604</td>
<td>5.51(-5)</td>
</tr>
<tr>
<td>32</td>
<td>241</td>
<td>6.53(-4)</td>
<td>3393</td>
<td>5.53(-5)</td>
</tr>
<tr>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
</tr>
<tr>
<td>99</td>
<td>973</td>
<td>7.01(-4)</td>
<td>15017</td>
<td>6.07(-5)</td>
</tr>
<tr>
<td>100</td>
<td>304</td>
<td>7.02(-4)</td>
<td>10489</td>
<td>6.08(-5)</td>
</tr>
</tbody>
</table>
**Numerical Example**

**EXAMPLE:** \( \gamma_j = \frac{1}{j^2}, \; j = 1, 2, \ldots \)

<table>
<thead>
<tr>
<th>s</th>
<th>( z_j )</th>
<th>( e_{n,s,\gamma}^{\text{rms}}(z) )</th>
<th>( n = 2,003 )</th>
<th>( e_{n,s,\gamma}^{\text{rms}}(z) )</th>
<th>( n = 32,003 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2.04(-4)</td>
<td></td>
<td>1</td>
<td>1.28(-5)</td>
</tr>
<tr>
<td>2</td>
<td>765</td>
<td>3.02(-4)</td>
<td>9376</td>
<td>2.03(-5)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>605</td>
<td>3.72(-4)</td>
<td>11835</td>
<td>2.58(-5)</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>31</td>
<td>450</td>
<td>6.51(-4)</td>
<td>13604</td>
<td>5.51(-5)</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>241</td>
<td>6.53(-4)</td>
<td>3393</td>
<td>5.53(-5)</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>99</td>
<td>973</td>
<td>7.01(-4)</td>
<td>15017</td>
<td>6.07(-5)</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>304</td>
<td>7.02(-4)</td>
<td>10489</td>
<td>6.08(-5)</td>
<td></td>
</tr>
</tbody>
</table>

The apparent rate of convergence is better than \( O \left( \frac{1}{\sqrt{n}} \right) \).
NEW THEOREM


Let \( n \) be prime, and let \( z^1, z^2, \ldots, z^s \) be chosen by the new algorithm.

Assume

\[
\sum_{j=1}^{\infty} \gamma_j^{1/2} < \infty.
\]

Then \( \forall \delta > 0 \)

\[
e_{n,s,\gamma}^{\text{rms}}(z) \leq \frac{C_{\gamma,\delta}}{n^{1-\delta}}.
\]

– the optimal rate is achieved by the algorithm!
Faster construction

Nuyens & Cools (MC$^2$ QMC 2004)

The cost of the CBC construction for the randomly shifted case and $n$ prime can be reduced to $O(n \log n s_{\text{max}})!$.

The Nuyens and Cools implementation allows the CBC algorithm to be run with $s$ in thousands, $n$ in millions.
Nuyens & Cools (MC² QMC 2004)

The cost of the CBC construction for the randomly shifted case and \( n \) prime can be reduced to \( \mathcal{O}(n \log n \ s_{\text{max}})! \).

The Nuyens and Cools implementation allows the CBC algorithm to be run with \( s \) in thousands, \( n \) in millions.

How does it work? Recall

\[
\varepsilon_{n,s,\gamma}^{\text{rms}}(z^1, z^2, \ldots, z^s)^2 = -\prod_{j=1}^{s} \left(1 + \frac{\gamma_j}{3}\right) + \\
\frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{s-1} \left[1 + \gamma_j \left(B_2(\{iz_j/n\}) + \frac{1}{3}\right)\right] \left[1 + \gamma_s \left(B_2(\{iz_s/n\}) + \frac{1}{3}\right)\right]
\]
Before and after, for $n = 53$

Because $n$ is prime, each row of the matrix $B_2\left(\{iz_s/n\}\right)$ contains the same numbers $B_2(1/n), B_2(2/n), \ldots, B_2((n-1)/n)$ in some order.

Graphic by Dirk Nuyens
The big question for applications

For a particular application, how should the weights $\gamma_1, \gamma_2, \ldots$ be chosen?
THE BIG QUESTION FOR APPLICATIONS

For a particular application, how should the weights $\gamma_1, \gamma_2, \ldots$ be chosen?

That depends on the problem.
The basic finance problem: formulation

A typical one-asset problem with payoff $G$ often has the form

\[ E(G) = \int_{\mathbb{R}^s} G(x) \frac{1}{\sqrt{(2\pi)^s \det C}} \exp\left(-\frac{1}{2} x^T C^{-1} x\right) \, dx \]

where in the case of geometric Brownian motion

\[ C_{i,j} = \min(t_i, t_j) \quad i, j = 1, \ldots, s. \]
The basic finance problem: formulation

A typical one-asset problem with payoff $G$ often has the form

$$
E(G) = \int_{\mathbb{R}^s} G(x) \frac{1}{\sqrt{(2\pi)^s \det C}} \exp^{-\frac{1}{2} x^T C^{-1} x} \, dx
$$

where in the case of geometric Brownian motion

$$
C_{i,j} = \min(t_i, t_j) \quad i, j = 1, \cdots, s.
$$

The first step is to factorise $C$:

$$
C = AA^T.
$$
The basic finance problem: continued

Then

\[ E(G) = \int_{\mathbb{R}^s} G(Az) \frac{1}{\sqrt{(2\pi)^s}} \exp\left(-\frac{1}{2} z^T z\right) \, dz, \quad x = Az \]

\[ = \int_{[0,1]^s} G(A\Phi^{-1}(t)) \, dt, \quad z = \Phi^{-1}(t) \]

\[ \approx \frac{1}{n} \sum_{k=1}^{n} G(A\Phi^{-1}(t_k)) \]
Then

\[ E(G) = \int_{\mathbb{R}^s} G(Az) \frac{1}{\sqrt{(2\pi)^s}} \exp^{-\frac{1}{2}z^Tz} \, dz, \quad x = Az \]

\[ = \int_{[0,1]^s} G(A\Phi^{-1}(t)) \, dt, \quad z = \Phi^{-1}(t) \]

\[ \approx \frac{1}{n} \sum_{k=1}^{n} G(A\Phi^{-1}(t_k)) \]

Sometimes it makes sense to choose the matrix \( A \) so as to put as much variance into the first few variables as possible. This choice is known as principal components analysis, or PCA. We know that this is a good choice for the case of an ‘Asian option’.
The basic finance problem: continued

Then

\[
E(G) = \int_{\mathbb{R}^s} G(Az) \frac{1}{\sqrt{(2\pi)^s}} \exp\left(-\frac{1}{2}z^Tz\right) \, dz, \quad x = Az
\]

\[
= \int_{[0,1]^s} G(A\Phi^{-1}(t)) \, dt, \quad z = \Phi^{-1}(t)
\]

\[
\approx \frac{1}{n} \sum_{k=1}^{n} G(A\Phi^{-1}(t_k))
\]

Sometimes it makes sense to choose the matrix \(A\) so as to put as much variance into the first few variables as possible. This choice is known as \textit{principal components analysis}, or PCA. We know that this is a good choice for the case of an ‘Asian option’.

(Alternatively, the \textit{standard construction} corresponds to a Cholesky factorisation of \(C\).)
For an Asian option with PCA the importance of the new variables (ordered by the eigenvalues of $C$) declines rapidly — so we may use decaying weights, for example

$$\gamma_j = \text{const} \times \frac{1}{j^2}.$$ 

For such a case the numerical results as $n$ increases are very good.
Performance of QMC

Comparison of MC and QMC: digital Asian option

<table>
<thead>
<tr>
<th>Value of option</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.005</td>
</tr>
</tbody>
</table>

$n = 2^{12}, 2^{13}, 2^{14}$
Performance of QMC

Comparison of MC and QMC: digital Asian option

<table>
<thead>
<tr>
<th>n</th>
<th>Value of option</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.50</td>
<td>0.005</td>
</tr>
<tr>
<td>1</td>
<td>0.51</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.52</td>
<td></td>
</tr>
</tbody>
</table>

Graphs showing the value of the option and the standard error for different values of n, comparing MC and QMC.
Performance of QMC

Comparison of MC and QMC: digital Asian option

<table>
<thead>
<tr>
<th>Value of option</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC</td>
<td>QMC</td>
</tr>
<tr>
<td>QMC + PCA</td>
<td></td>
</tr>
</tbody>
</table>

The diagrams show the evolution of the value of the option and the standard error as a function of the number of simulations (n).
In a current application for an Australian bank (related to the valuation of a certain financial instrument) Kuo and Waterhouse are currently using rules for dimension 9,125 and up to approximately 1 million points!
In a current application for an Australian bank (related to the valuation of a certain financial instrument) Kuo and Waterhouse are currently using rules for dimension 9,125 and up to approximately 1 million points!

This calculation uses an embedded sequence of rules, with \( n \) from \( 2^{10} \) to \( 2^{20} \), which is extensible in dimension by a variant of the fast CBC algorithm. (Cools, Kuo, Nuyens, submitted)
Open Questions

- For which applications is the weighted Sobolev space model appropriate? (It is NOT strictly appropriate for option pricing problems.) Modifications?

- For a particular application (e.g. option pricing, or likelihood integrals in statistics), how should the weights be chosen?

- For a particular application, how best to transform from $\mathbb{R}^s$ to $[0, 1]^s$?
For a general reference, see:


or

http://www.maths.unsw.edu.au, then select successively "applied", "reports", "2005"
1. **Natural ordering of the indices**

2. **Grouping on divisors**

3. **Generator ordering of the indices**

4. **Symmetric reduction after application of $B_2$ kernel function**