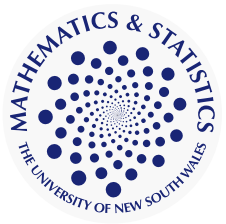
# A contemporary view of high-dimensional Quasi Monte Carlo


## Ian H. Sloan

i.sloan@unsw.edu.au

**The University of New South Wales – UNSW Australia**


**SIAM UQ 2018**

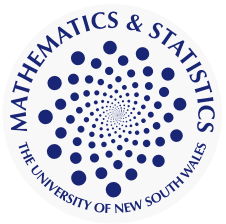


In UQ we often meet integrals with hundreds or thousands of variables.

Let's define the $s$-dimensional integral,

$$I_s(F) := \int_0^1 \cdots \int_0^1 F(y_1, \ldots, y_s) \mathrm{d}y_1 \cdots \mathrm{d}y_s,$$

where $s$ may be large.



In UQ we often meet integrals with hundreds or thousands of variables.

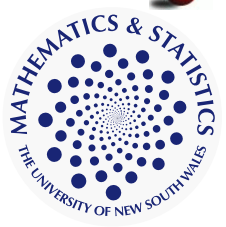
Let's define the $s$-dimensional integral,

$$I_s(F) := \int_0^1 \dots \int_0^1 F(y_1, \dots, y_s) \mathrm{d}y_1 \cdots \mathrm{d}y_s,$$

where $s$ may be large.

The main classes of methods for $\int_{[0,1]^s} F(\mathbf{y}) \mathbf{dy}$ are:

- Monte Carlo (MC)
- Sparse grids
- Quasi Monte Carlo (QMC)



In UQ we often meet integrals with hundreds or thousands of variables.

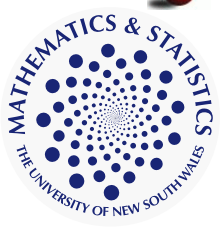
Let's define the $s$-dimensional integral,

$$I_s(F) := \int_0^1 \cdots \int_0^1 F(y_1, \ldots, y_s) \mathrm{d}y_1 \cdots \mathrm{d}y_s,$$

where $s$ may be large.

The main classes of methods for $\int_{[0,1]^s} F(\mathbf{y}) \mathbf{dy}$ are:

- Monte Carlo (MC)
- Sparse grids
- Quasi Monte Carlo (QMC)



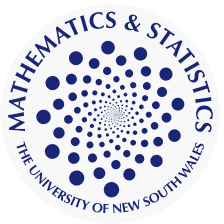

And when $s$ is **very** large, only MC and QMC remain.

# MC and QMC

Both MC and QMC approximate $I_s(F)$ by an equal-weight sum:

$$Q_{N,s}(F) := \frac{1}{N} \sum_{k=1}^{N} F(\mathrm{t}_k),$$

where $\mathrm{t}_1, \ldots, \mathrm{t}_N$ are points in $[0,1]^s$.



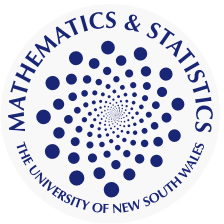
# MC and QMC

Both MC and QMC approximate $I_s(F)$ by an equal-weight sum:

$$Q_{N,s}(F) := \frac{1}{N} \sum_{k=1}^{N} F(\mathbf{t}_k),$$

where $\mathbf{t}_1, \ldots, \mathbf{t}_N$ are points in $[0,1]^s$.

MC: For MC $\mathbf{t}_1, \ldots, \mathbf{t}_N$ are chosen **randomly** and independently from a uniform distribution on $[0,1]^s$.



# MC and QMC

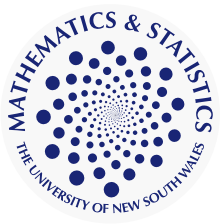
Both MC and QMC approximate $I_s(F)$ by an equal-weight sum:

$$Q_{N,s}(F) := \frac{1}{N} \sum_{k=1}^{N} F(\mathbf{t}_k),$$

where $\mathbf{t}_1, \ldots, \mathbf{t}_N$ are points in $[0,1]^s$.

MC: For MC $\mathbf{t}_1, \ldots, \mathbf{t}_N$ are chosen **randomly** and independently from a uniform distribution on $[0,1]^s$.

QMC: For QMC $\mathbf{t}_1, \ldots, \mathbf{t}_N$ are **deterministic** (and cleverly chosen).



# MC and QMC

Both MC and QMC approximate $I_s(F)$ by an equal-weight sum:

$$Q_{N,s}(F) := \frac{1}{N} \sum_{k=1}^{N} F(\mathbf{t}_k),$$

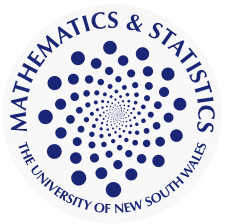
where $\mathbf{t}_1, \ldots, \mathbf{t}_N$ are points in $[0,1]^s$.

MC: For MC $\mathbf{t}_1, \ldots, \mathbf{t}_N$ are chosen **randomly** and independently from a uniform distribution on $[0,1]^s$.

QMC: For QMC $\mathbf{t}_1, \ldots, \mathbf{t}_N$ are **deterministic** (and cleverly chosen).

But how to choose good QMC points $\mathbf{t}_1, \ldots, \mathbf{t}_N$?

And how to prove that your choice is good?



# The traditional view e.g. Wikipedia

The famous **Koksma-Hlawka inequality** says,

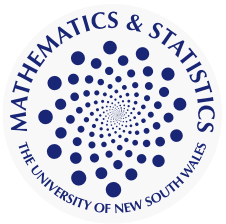
$$|I_s(F) - Q_{N,s}(F)| \leq D_N^*(\mathrm{t}_1, \ldots, \mathrm{t}_N)\ V(F),$$

– $V(F)$ is the **variation of $F$ in the sense of Hardy-Krause**

– $D_N^*$ is the **star discrepancy**,

$$D_N^*(\mathrm{t}_1, \ldots, \mathrm{t}_N) = \sup_B \Big|(\text{fraction of points in} B) - (\text{volume of} B)\Big|.$$

where the supremum is over all rectangular boxes $B \subseteq [0,1]^s$ with bottom left-hand corner at the origin.

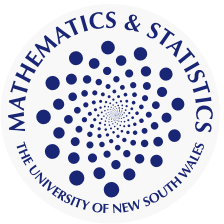


# The traditional view (continued)

"Good" point sets are the first $N$ members of a **low-discrepancy sequence**:

**Definition:** A **low-discrepancy sequence** $\mathbf{t_1, t_2, \ldots}$ (such as the **Sobol** sequence) satisfies

$$D_N^* \leq C_s \frac{(\log N)^s}{N}$$



# The traditional view (continued)

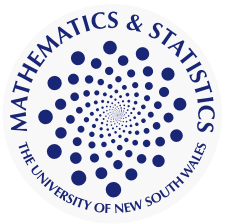
"Good" point sets are the first $N$ members of a **low-discrepancy sequence**:

**Definition:** A **low-discrepancy sequence** $t_1, t_2, \ldots$ (such as the **Sobol** sequence) satisfies

$$D_N^* \leq C_s \frac{(\log N)^s}{N}$$

$$\Longrightarrow \quad |I_s(F) - Q_{N,s}(F)| \leq C_s \frac{(\log N)^s}{N} V(F) \, .$$

Cf. $\sqrt{\mathbb{E}[(I_s(F) - Q_{N,s}(F)^{\mathrm{MC}})^2]} \leq \frac{1}{\sqrt{N}} \sqrt{I_s(F^2) - (I_s(F))^2}.$



# The traditional view (continued)

"Good" point sets are the first $N$ members of a **low-discrepancy sequence**:

**Definition:** A **low-discrepancy sequence** $t_1, t_2, \ldots$ (such as the **Sobol** sequence) satisfies

$$D_N^* \leq C_s \frac{(\log N)^s}{N}$$

$$\Longrightarrow \quad |I_s(F) - Q_{N,s}(F)| \leq C_s \frac{(\log N)^s}{N} V(F) \, .$$

Cf. $\sqrt{\mathbb{E}[(I_s(F) - Q_{N,s}(F)^{\mathrm{MC}})^2]} \leq \frac{1}{\sqrt{N}} \sqrt{I_s(F^2) - (I_s(F))^2}.$

This is fine if $s$ is not too large, but if $s$ is large ...

# The KH inequality isn't useful if $s$ is large!

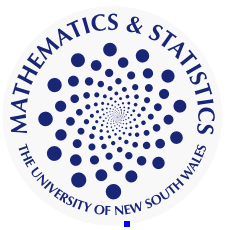
It isn't useful because for fixed $s$ **the bound is increasing with** $N$ – it continues to increase until $N \approx e^s$.

And by the time $N = e^s$ the bound can be truly astronomical:

$$\frac{(\log N)^s}{N} = \left(\frac{s}{e}\right)^s.$$

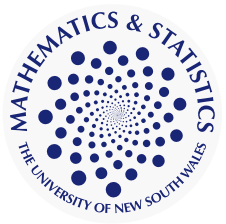
For example, if $s = 100$ then

$$\left(\frac{s}{e}\right)^s \approx 10^{150}.$$



# The contemporary point of view on QMC

is that the dependence on $s$ is as important as the dependence $N$!

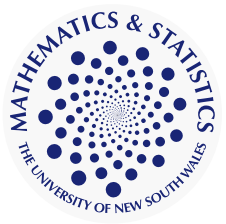


# The contemporary point of view on QMC

is that the dependence on $s$ is as important as the dependence $N$!

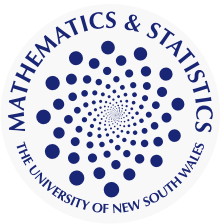
But how should we characterise the dependence on $s$?

How much harder does a problem become as the dimension increases?



# Applications are the key

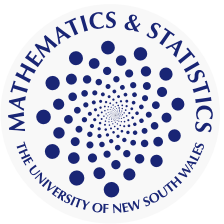
The dependence on $s$ should be driven by **applications**.



# Applications are the key

The dependence on $s$ should be driven by **applications**.

In applications (an example is coming!), we often have a **sequence** of integrands $F_s = F_s(y_1, \ldots, y_s)$, with $s \to \infty$. So we may want a sequence of QMC rules for ever increasing dimension $s$.



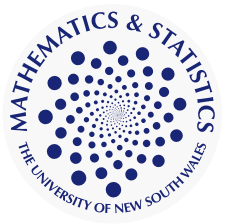
# Applications are the key

The dependence on $s$ should be driven by **applications**.

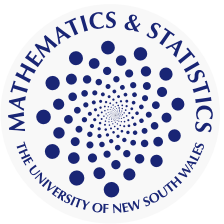
In applications (an example is coming!), we often have a **sequence** of integrands $F_s = F_s(y_1, \ldots, y_s)$, with $s \to \infty$. So we may want a sequence of QMC rules for ever increasing dimension $s$.

Given an application, how to decide what would be **good** QMC rules?

A convincing way is to **make some error bound small**.



# What sort of error bound would we wish for?



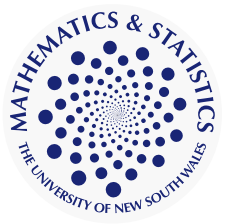
# What sort of error bound would we wish for?

What sort of error bound do we want? An ideal result might look like:

$$\boxed{|I_s(F_s) - Q_{N,s}(F_s)| \leq \frac{C}{N^\kappa}\|F_s\| \leq \frac{C'}{N^\kappa}},$$

where

- $\|\cdot\|$ is some norm that depends on the smoothness of $F_s$,
- $\kappa$ is as large as possible,
- and $C$ and $\|F_s\|$ (and their product $C'$) are independent of $s$.



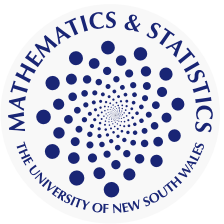
But if we want to have

$$|I_s(F_s) - Q_{N,s}(F_s)| \leq \frac{C}{N^\kappa}\|F_s\|_s \leq \frac{C'}{N^\kappa},$$

with $C, C'$ independent of $s$, then clearly

- we cannot use star discrepancy, because of its $(\log N)^s$ factor
- the norms $\|F_s\|$ must be bounded, so

  **the norm $\|\cdot\|$ MUST depend on the application**
- and we may need an essentially different QMC rule for each $s$.



But if we want to have

$$|I_s(F_s) - Q_{N,s}(F_s)| \leq \frac{C}{N^\kappa}\|F_s\|_s \leq \frac{C'}{N^\kappa},$$

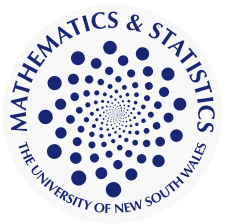
with $C, C'$ independent of $s$, then clearly

- we cannot use star discrepancy, because of its $(\log N)^s$ factor
- the norms $\|F_s\|$ must be bounded, so

  **the norm $\|\cdot\|$ MUST depend on the application**
- and we may need an essentially different QMC rule for each $s$.

For guidance, let's turn to an important class of problems:

**PDE with random coefficients**.



# PDE with random coefficients

A motivating example: flow through a porous medium

Darcy's law is $\vec{q}(\mathbf{x}) = -a(\mathbf{x})\nabla p(\mathbf{x}),$

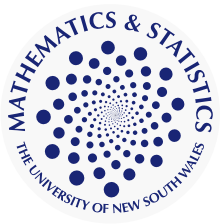
where

$p(\mathbf{x})$ is pressure of the fluid

$\vec{q}(\mathbf{x})$ is velocity of the fluid

$a(\mathbf{x})$ is "permeability" of the medium

Incompressibility: $\nabla \cdot \vec{q} = 0$



# PDE with random coefficients

A motivating example: flow through a porous medium

Darcy's law is $\vec{q}(\mathbf{x}) = -a(\mathbf{x})\nabla p(\mathbf{x}),$

where

$p(\mathbf{x})$ is pressure of the fluid

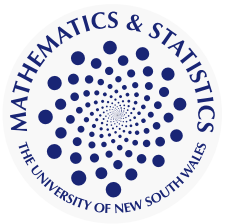
$\vec{q}(\mathbf{x})$ is velocity of the fluid

$a(\mathbf{x})$ is "permeability" of the medium

Incompressibility: $\nabla \cdot \vec{q} = 0$

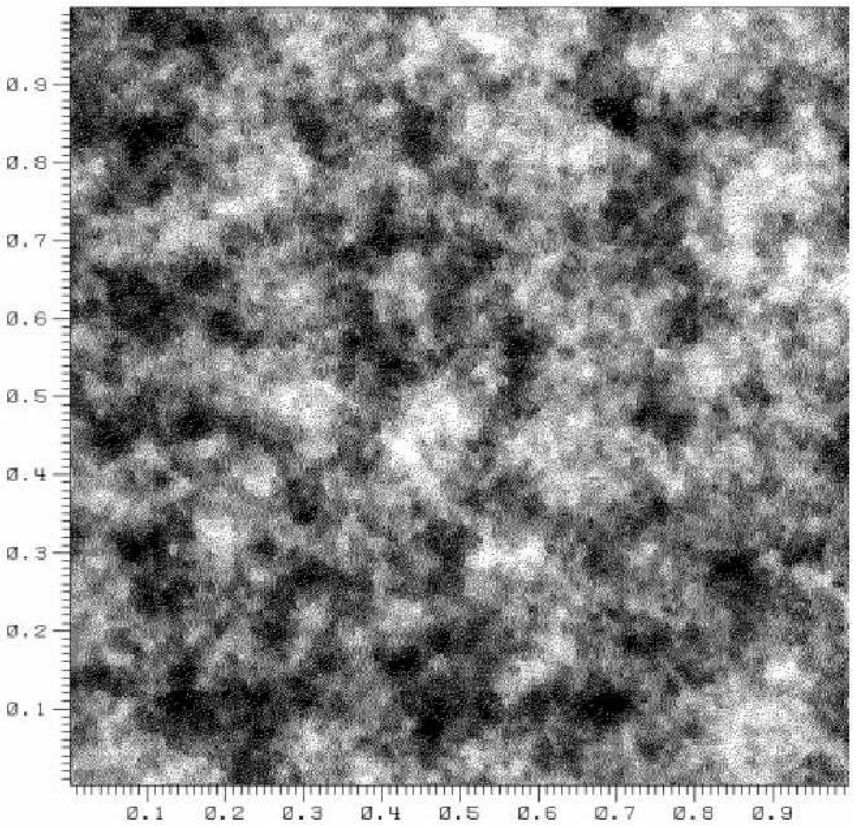
Together these give a second order elliptic PDE:

$$\nabla \cdot (a(\mathbf{x})\nabla p(\mathbf{x})) = 0$$



# Modeling the permeability

Describing all the microscopic pores and channels in a real material is commonly considered much too hard. So it is common engineering practice to model the permeability as a random field:




MATHEMATICS & STATISTICS
THE UNIVERSITY OF NEW SOUTH WALES

# A simple model problem – the "uniform" case

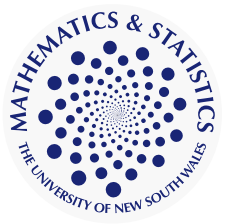
$$-\nabla \cdot (a(\mathrm{x},\mathrm{y})\,\nabla u(\mathrm{x},\mathrm{y})) \;=\; f(\mathrm{x}) \quad \text{for } \mathrm{x} \in D\,,$$

$$u(\mathrm{x},\mathrm{y}) \;=\; 0 \quad \text{on} \quad \partial D, \quad \mathrm{y} \in U := [0,1]^{\mathbb{N}},$$

with $D$ a bounded Lipschitz domain in $\mathbb{R}^d$, with $d = 1, 2,$ or $3$ and

$$a(\mathrm{x},\mathrm{y}) \;=\; \overline{a} + \sum_{j=1}^{\infty} (y_j - \tfrac{1}{2})\,\psi_j(\mathrm{x}), \qquad \mathrm{x} \in D, \quad \mathrm{y} \in U\,,$$

where $y_1, y_2, \ldots$ are parameters representing independent random variables uniformly distributed on $[0, 1]$;



# A simple model problem – the "uniform" case

$$-\nabla \cdot (a(\mathrm{x},\mathrm{y})\,\nabla u(\mathrm{x},\mathrm{y})) = f(\mathrm{x}) \quad \text{for } \mathrm{x} \in D\,,$$

$$u(\mathrm{x},\mathrm{y}) = 0 \quad \text{on} \quad \partial D, \quad \mathrm{y} \in U := [0,1]^{\mathbb{N}},$$

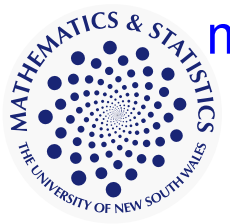
with $D$ a bounded Lipschitz domain in $\mathbb{R}^d$, with $d = 1, 2,$ or $3$ and

$$a(\mathrm{x},\mathrm{y}) = \overline{a} + \sum_{j=1}^{\infty} (y_j - \tfrac{1}{2})\,\psi_j(\mathrm{x}), \qquad \mathrm{x} \in D, \quad \mathrm{y} \in U\,,$$

where $y_1, y_2, \ldots$ are parameters representing independent random variables uniformly distributed on $[0,1]$; with $\overline{a}, \psi_j$ such that $\sum_j \|\psi_j\|_\infty < \infty$, and

$$a_{\max} \geq a(\mathrm{x},\mathrm{y}) \geq a_{\min} > 0,$$

making the PDE **strongly elliptic** for every $\mathrm{y}$.



# A simple model problem – the "uniform" case

In practice it is necessary to truncate the infinite sum to $s$ terms:

$$-\nabla \cdot (a(\mathrm{x},\mathrm{y})\,\nabla u(\mathrm{x},\mathrm{y})) \;=\; f(\mathrm{x}) \quad \text{in} \quad D\,,$$

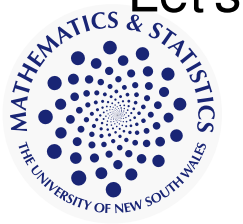
$$u(\mathrm{x},\mathrm{y}) \;=\; 0 \quad \text{on} \quad \partial D, \quad \mathrm{y} \in U := [0,1]^{\mathbb{N}},$$

with $D$ a bounded Lipschitz domain in $\mathbb{R}^d$, and

$$a(\mathrm{x},\mathrm{y}) \;=\; \overline{a} + \sum_{j=1}^{s} (y_j - \tfrac{1}{2})\,\psi_j(\mathrm{x}), \qquad \mathrm{x} \in D, \quad \mathrm{y} \in U\,,$$

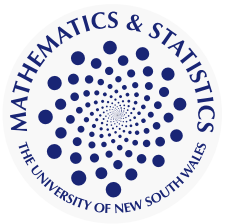
where $y_1, y_2, \ldots, y_s$ are independent random variables uniformly distributed on $[0,1]$.

Let's call the solution of the truncated system $u_s$.



# What might we want to compute?

- The mean pressure at a particular point or over a particular small region
- The effective permeability
- The mean “breakthrough time”
- . . .

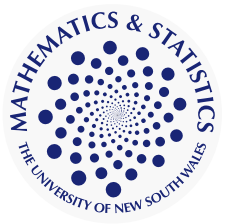


# What might we want to compute?

- The mean pressure at a particular point or over a particular small region
- The effective permeability
- The mean "breakthrough time"
- . . .

**All are expected values – and expected values are integrals.**

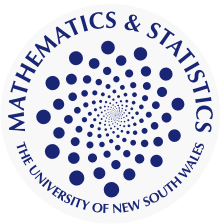
**And since there are $s$ uniform random variables, the expected value is an integral over $[0, 1]^s$.** And it is easy for $s$ to be large!



Suppose the problem is to compute the expected value of

$$F(\mathrm{y}) := G(u(\cdot;\mathrm{y})),$$

where $G$ is the **mean pressure $\overline{u}$ over some specified subregion** or some other linear functional.



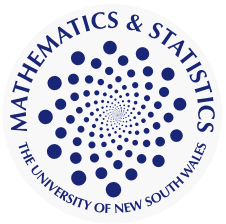
Suppose the problem is to compute the expected value of

$$F(\mathbf{y}) := G(u(\cdot;\mathbf{y})),$$

where $G$ is the **mean pressure $\overline{u}$ over some specified subregion** or some other linear functional.

The expected value is then an infinite-dimensional **integral**

$$I(F) := \int_{[0,1]^{\mathbb{N}}} F(\mathbf{y})\mathrm{d}\mathbf{y}.$$



Suppose the problem is to compute the expected value of

$$F(\mathrm{y}) := G(u(\cdot;\mathrm{y})),$$

where $G$ is the **mean pressure $\overline{u}$ over some specified subregion** or some other linear functional.

The expected value is then an infinite-dimensional **integral**

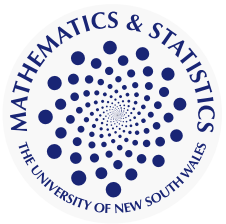
$$I(F) := \int_{[0,1]^{\mathbb{N}}} F(\mathrm{y})\mathrm{d}\mathrm{y}.$$

This is now approximated by

$$I_s(F_s) = \int_{[0,1]^s} F_s(y_1,\ldots,y_s)\mathrm{d}y_1\ldots\mathrm{d}y_s,$$

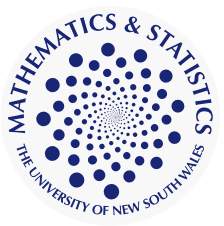
where

$$F_s(y_1,\ldots,y_s) := G(u_s(\cdot;y_1,\ldots,y_s,\tfrac{1}{2},\tfrac{1}{2},\cdots)),$$



# A bound of the desired shape

For this random PDE problem a recent paper [Kuo/Schwab/IHS, SINUM 2012] obtained a first error bound of the desired form:



# A bound of the desired shape

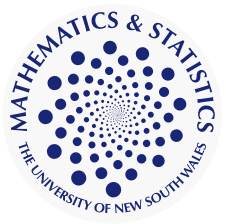
For this random PDE problem a recent paper [Kuo/Schwab/IHS, SINUM 2012] obtained a first error bound of the desired form:

**Theorem** [KSS12] For all $\delta > 0$,

$$\boxed{|I_s(F_s) - Q_{N,s}(F_s)| \leq \frac{C_\delta}{N^{1-\delta}} \|F_s\| \leq \frac{C'_\delta}{N^{1-\delta}},}$$

with $C_\delta$ independent of $s$; AND $\sup_s \|F_s\| < \infty$

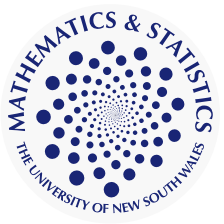
**if** $\sum_{j=1}^{\infty} \|\psi_j\|_\infty^{2/3} < \infty$ and other technical conditions.



# So what was the norm?

So what was the norm $\|\boldsymbol{F_s}\|$ in [KSS12]?

Later!

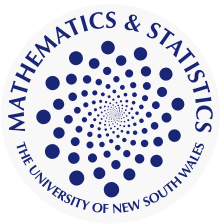


# So what was the norm?

So what was the norm $\|\boldsymbol{F}_{\boldsymbol{s}}\|$ in [KSS12]?

Later!

And what was the QMC rule in [KSS12]?



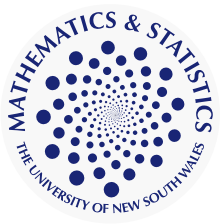
# So what was the norm?

So what was the norm $\|\boldsymbol{F}_{\boldsymbol{s}}\|$ in [KSS12]?

Later!

And what was the QMC rule in [KSS12]?

The QMC rule in that work was a **lattice rule** (more precisely, a **randomly shifted lattice rule**).

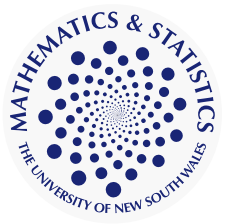


# Lattice rules

Lattice rules were introduced by number theorists in the late 1950s and 1960s, especially:

**Korobov, Hlawka, Zaremba, L. K. Hua**

They were designed for **periodic** functions.



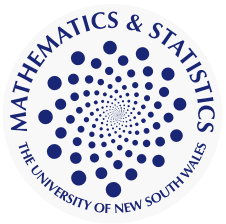
# Lattice rules

Lattice rules were introduced by number theorists in the late 1950s and 1960s, especially:

**Korobov, Hlawka, Zaremba, L. K. Hua**

They were designed for **periodic** functions.

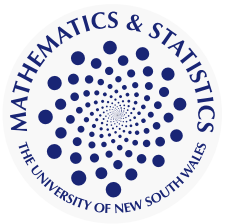
Most integrands are **not** periodic, but lattice rules can still be useful.



The simplest kind of lattice rule has the form

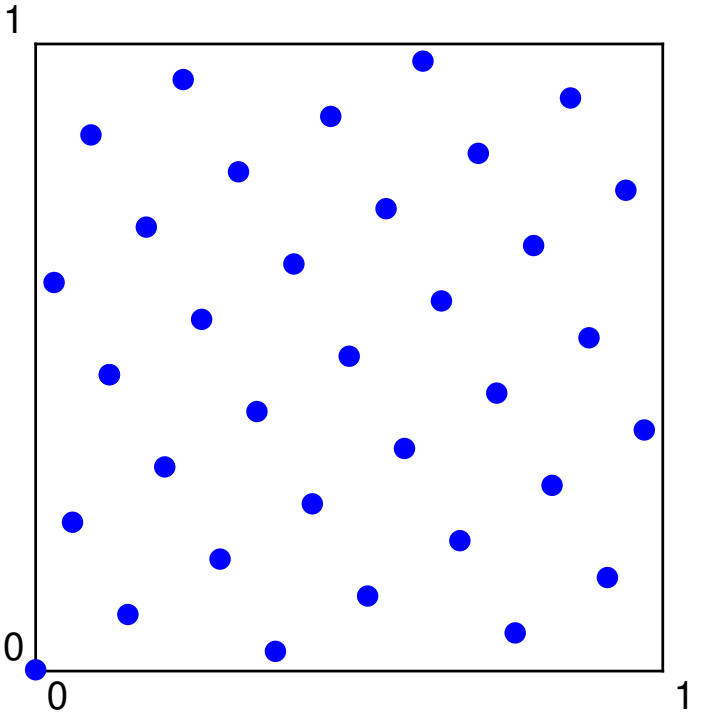
$$Q_{N,s}(\mathbf{z}; F) = \frac{1}{N} \sum_{k=0}^{N-1} F\left(\left\{k\frac{\mathbf{z}}{N}\right\}\right),$$

where $\mathbf{z} \in \{1, \ldots, N-1\}^s$, and the braces mean that each component of the $s$-vector in the braces is to be replaced by its fractional part.



# Example of a (good) 2-D lattice rule

$$s = 2, N = 34, \mathbf{z} = (1, 21)$$




MATHEMATICS & STATISTICS
THE UNIVERSITY OF NEW SOUTH WALES

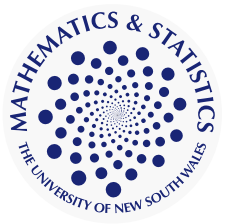
For the lattice rule

$$Q_{N,s}(\mathbf{z}; F) = \frac{1}{N} \sum_{k=0}^{N-1} F\left(\left\{k\frac{\mathbf{z}}{N}\right\}\right),$$

the corresponding **shifted lattice rule** is

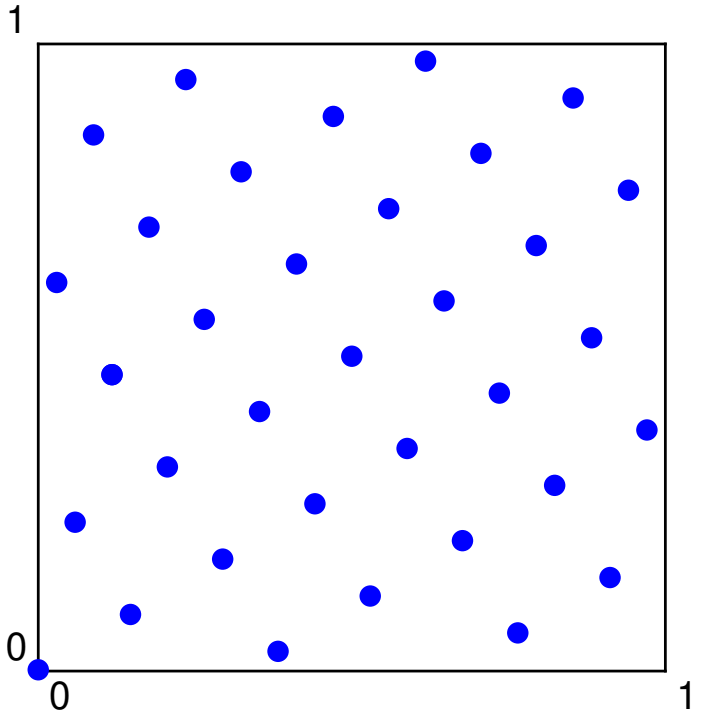
$$Q_{N,s}(\mathbf{z}, \boldsymbol{\Delta}; F) = \frac{1}{N} \sum_{k=0}^{N-1} F\left(\left\{k\frac{\mathbf{z}}{N} + \boldsymbol{\Delta}\right\}\right),$$

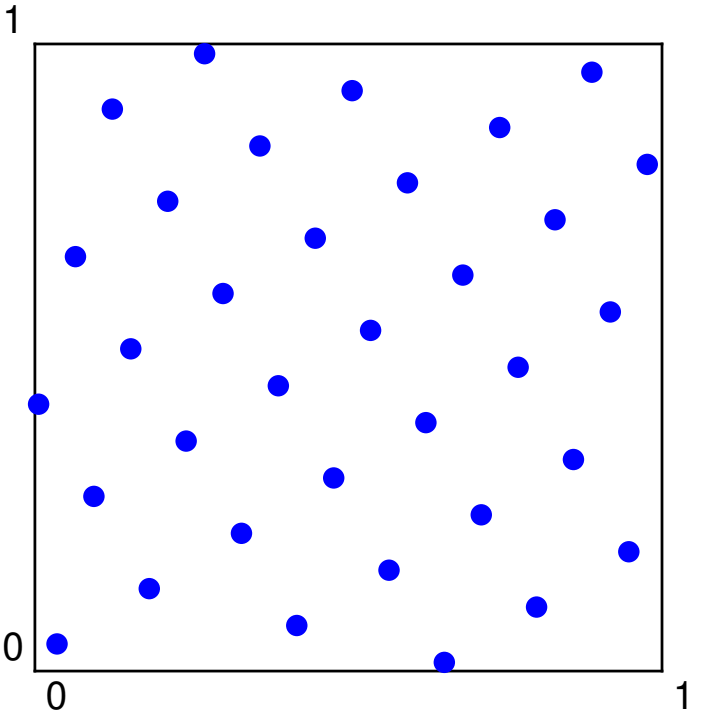
where $\boldsymbol{\Delta} \in [0,1)^s$ is the " **shift**".



# Lattice rule & shifted lattice rule

$N = 34, \mathbf{z} = (1, 21)$ $N = 34, \mathbf{z} = (1, 21), \boldsymbol{\Delta} = (0.8, 0.1)$

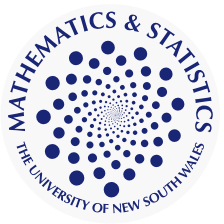



MATHEMATICS & STATISTICS
THE UNIVERSITY OF NEW SOUTH WALES

# Randomly shifted lattice rule

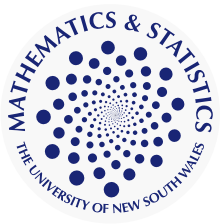
In a **randomly shifted** lattice rule $\boldsymbol{\Delta}$ is a **random** vector, uniformly distributed over $[0,1]^s$. In principle we now compute **expected values**.



# Randomly shifted lattice rule

In a **randomly shifted** lattice rule $\boldsymbol{\Delta}$ is a **random** vector, uniformly distributed over $[0,1]^s$. In principle we now compute **expected values**.

In practice we make 20 or 30 random choices of $\boldsymbol{\Delta} \in [0,1]^s$, do the calculation with each choice, and average the results.

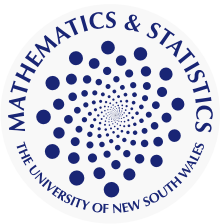


# Randomly shifted lattice rule

In a **randomly shifted** lattice rule $\boldsymbol{\Delta}$ is a **random** vector, uniformly distributed over $[0,1]^s$. In principle we now compute **expected values**.

In practice we make 20 or 30 random choices of $\boldsymbol{\Delta} \in [0,1]^s$, do the calculation with each choice, and average the results.

In a randomly shifted lattice rule the only thing to decide is how to choose $\mathbf{z}$!

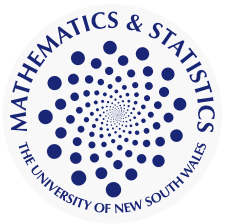


# The true error bound

The true error bound found in [KSS12] (no fake news here!) was for a randomly shifted lattice rule with a special choice of $\mathbf{z}$, and was a bound on the **root mean square expected error averaged over shifts**:

For all $\delta < \infty$,

$$\sqrt{\mathbb{E}_{\boldsymbol{\Delta}}[|I_s(F_s) - Q_{N,s}(\mathbf{z}, \boldsymbol{\Delta}; F)|^2]} \leq \frac{C_\delta}{N^{1-\delta}} \|F_s\| \leq \frac{C'_\delta}{N^{1-\delta}},$$

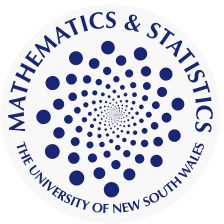


# How to choose $\mathbf{z}$?

Recall: the lattice rule for the integral over $[0,1]^s$ is

$$Q_{N,s}(\mathbf{z};F) = \frac{1}{N}\sum_{k=0}^{N-1} F\left(\left\{k\frac{\mathbf{z}}{N}\right\}\right).$$

So we need to choose $\mathbf{z}$. However, **there is no known formula for a "good" $\mathbf{z}$, beyond $s = 2$.**



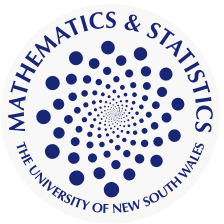
# How to choose $\mathbf{z}$?

Recall: the lattice rule for the integral over $[0,1]^s$ is

$$Q_{N,s}(\mathbf{z};F) = \frac{1}{N}\sum_{k=0}^{N-1} F\left(\left\{k\frac{\mathbf{z}}{N}\right\}\right).$$

So we need to choose $\mathbf{z}$. However, **there is no known formula for a "good" $\mathbf{z}$, beyond $s = 2$.**

But a good $\mathbf{z}$ can be **constructed**:



# How to choose $\mathbf{z}$?

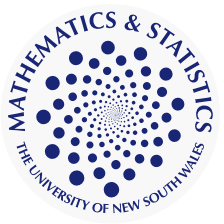
Recall: the lattice rule for the integral over $[0,1]^s$ is

$$Q_{N,s}(\mathbf{z};F) = \frac{1}{N}\sum_{k=0}^{N-1} F\left(\left\{k\frac{\mathbf{z}}{N}\right\}\right).$$

So we need to choose $\mathbf{z}$. However, **there is no known formula for a "good" $\mathbf{z}$, beyond $s=2$.**

But a good $\mathbf{z}$ can be **constructed**:

We choose a convenient Hilbert space $H$ to which $F$ belongs.



# How to choose $\mathbf{z}$?

Recall: the lattice rule for the integral over $[0,1]^s$ is

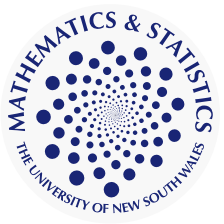
$$Q_{N,s}(\mathbf{z};F) = \frac{1}{N}\sum_{k=0}^{N-1} F\left(\left\{k\frac{\mathbf{z}}{N}\right\}\right).$$

So we need to choose $\mathbf{z}$. However, **there is no known formula for a "good" $\mathbf{z}$, beyond $s = 2$.**

But a good $\mathbf{z}$ can be **constructed**:

We choose a convenient Hilbert space $H$ to which $F$ belongs.

Then choose $\mathbf{z}$ to make the **worst-case error** for that space small.

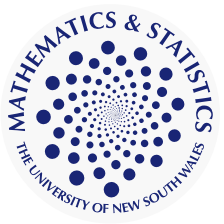


# Worst-case error

**Definition**: The **worst-case error** in the space $H$ of a given QMC rule $Q_{N,s}(\mathbf{z};\cdot)$ is

$$\boxed{e_{N,s,\mathbf{z}}(H) := \sup_{\|F\|_H \le 1} |I_s(F) - Q_{N,s}(\mathbf{z};F)|},$$

i.e. it is the largest error of $Q_{N,s}(\mathbf{z};F)$ for $F$ in the unit ball of $H$.



# Worst-case error

**Definition**: The **worst-case error** in the space $H$ of a given QMC rule $Q_{N,s}(\mathbf{z};\cdot)$ is

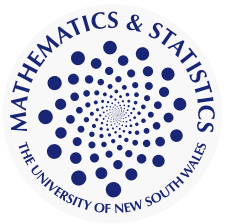
$$\boxed{e_{N,s,\mathbf{z}}(H) := \sup_{\|F\|_H \le 1} |I_s(F) - Q_{N,s}(\mathbf{z};F)|},$$

i.e. it is the largest error of $Q_{N,s}(\mathbf{z};F)$ for $F$ in the unit ball of $H$.

But in practice we are given $F$, not $H$. For a given $F$ we deduce

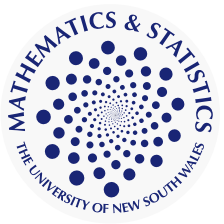
$$|I_s(F) - Q_{N,s}(\mathbf{z};F)| \le e_{N,s,\mathbf{z}}(H)\|F\|_H.$$

The **shift averaged worst-case error** has a similar error bound.



# So how to choose the function space $H$?

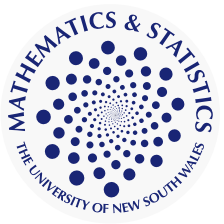
- We choose $H$ to be a Hilbert space of smooth enough functions, for which the worst-case error is easily computable.



# So how to choose the function space $H$?

- We choose $H$ to be a Hilbert space of smooth enough functions, for which the worst-case error is easily computable.

- The space $H$ also needs to have many free parameters, which we can adjust to make the final error bound small.

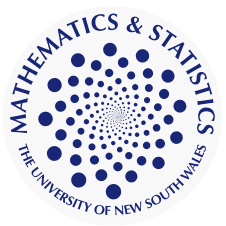


# So how to choose the function space $H$?

- We choose $H$ to be a Hilbert space of smooth enough functions, for which the worst-case error is easily computable.

- The space $H$ also needs to have many free parameters, which we can adjust to make the final error bound small.

Specifically, we take $H$ to be a space of functions with **square-integrable mixed first derivatives**, and the following norm:



# The space $H$ for the uniform case

The norm squared in $H$ is

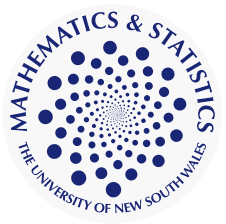
$$\|F\|_{s,\boldsymbol{\gamma}}^2 := \sum_{\mathfrak{u}\subseteq\{1,\ldots,s\}} \frac{1}{\gamma_{\mathfrak{u}}} \int_{[0,1]^{|\mathfrak{u}|}} \left| \frac{\partial^{|\mathfrak{u}|} F}{\partial \mathbf{y}_{\mathfrak{u}}}(\mathbf{y}_{\mathfrak{u}}; \tfrac{\mathbf{1}}{\mathbf{2}}) \right|^2 \mathrm{d}\mathbf{y}_{\mathfrak{u}},$$

where

$$(\mathbf{y}_{\mathfrak{u}}; \tfrac{\mathbf{1}}{\mathbf{2}})_j = \begin{cases} y_j \text{ if } j \in \mathfrak{u}, \\ \frac{1}{2} \text{ if } j \notin \mathfrak{u}. \end{cases}$$

For example, for $\mathfrak{u} = \{1, 3\}$ the corresponding term is

$$\frac{1}{\gamma_{1.3}} \int_0^1 \int_0^1 \left| \frac{\partial^2 F}{\partial y_1 \partial y_3}(y_1, \tfrac{1}{2}, y_3, \tfrac{1}{2}, \tfrac{1}{2} \ldots) \right|^2 \mathrm{d}y_1 \mathrm{d}y_3.$$



# The space $H$ for the uniform case

The norm squared in $H$ is

$$\|F\|^2_{s,\boldsymbol{\gamma}} := \sum_{\mathfrak{u}\subseteq\{1,\ldots,s\}} \frac{1}{\gamma_{\mathfrak{u}}} \int_{[0,1]^{|\mathfrak{u}|}} \left| \frac{\partial^{|\mathfrak{u}|} F}{\partial \mathbf{y}_{\mathfrak{u}}}(\mathbf{y}_{\mathfrak{u}}; \tfrac{\mathbf{1}}{\mathbf{2}}) \right|^2 \mathrm{d}\mathbf{y}_{\mathfrak{u}},$$

where

$$(\mathbf{y}_{\mathfrak{u}}; \tfrac{\mathbf{1}}{\mathbf{2}})_j = \begin{cases} y_j \text{ if } j \in \mathfrak{u}, \\ \frac{1}{2} \text{ if } j \notin \mathfrak{u}. \end{cases}$$

For example, for $\mathfrak{u} = \{1, 3\}$ the corresponding term is

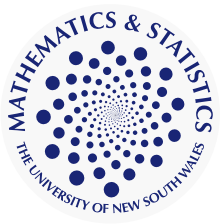
$$\frac{1}{\gamma_{1.3}} \int_0^1 \int_0^1 \left| \frac{\partial^2 F}{\partial y_1 \partial y_3}(y_1, \tfrac{1}{2}, y_3, \tfrac{1}{2}, \tfrac{1}{2} \ldots) \right|^2 \mathrm{d}y_1 \mathrm{d}y_3.$$

The $\gamma_{\mathfrak{u}}$ are **weights** - positive numbers chosen to fit the problem.

# Weights

Weights are in most applications **essential**. They play a big role in the contemporary QMC story.

Nowadays there are many kinds of weights ("order-dependent weights", "finite-order weights", "POD" weights, "SPOD weights", ...).

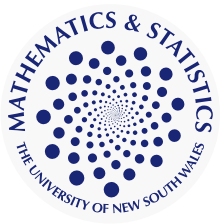


# Weights

Weights are in most applications **essential**. They play a big role in the contemporary QMC story.

Nowadays there are many kinds of weights ("order-dependent weights", "finite-order weights", "POD" weights, "SPOD weights", ...).

But the original weights, the simplest to handle, and still very important, are PRODUCT WEIGHTS.

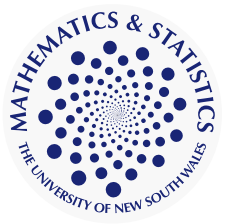


# Product weights

**Product weights** take the form

$$\gamma_{\boldsymbol{u}} = \prod_{j \in \boldsymbol{u}} \alpha_j, \quad \boldsymbol{u} \subseteq \{1, \ldots, s\},$$

where $\alpha_1 \geq \alpha_2 \geq \ldots > 0$ (also called weights!) are positive numbers chosen to quantify the decreasing importance of successive variables.

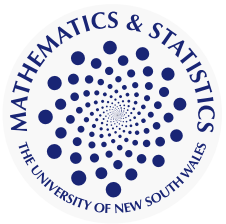


# Product weights are nice!

For product weights $\gamma_{\mathfrak{u}} = \prod_{j \in \mathfrak{u}} \alpha_j$, it has been known for $\mathbf{20}$ years that:

**Theorem** (IHS/Woźniakowski 98). There exist QMC points for which the worst case error is bounded independently of $s$ **iff**

$$\sum_{j=1}^{\infty} \alpha_j < \infty.$$



# Product weights are nice!

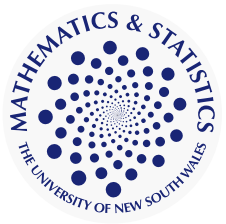
For product weights $\gamma_{\mathfrak{u}} = \prod_{j \in \mathfrak{u}} \alpha_j$, it has been known for $\mathbf{20}$ years that:

**Theorem** (IHS/Woźniakowski 98). There exist QMC points for which the worst case error is bounded independently of $s$ **iff**

$$\sum_{j=1}^{\infty} \alpha_j < \infty.$$

This is satisfied, for example, by $\alpha_j = 1/j^2$.

It is not satisfied by $\alpha_j = 1$.



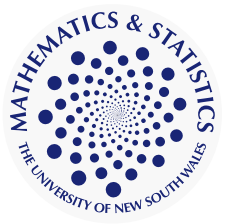
# The convergence rate can be $N^{-1+\delta}$

**Theorem** (IHS/Woźniakowski 01). If $N$ is prime, and if

$$\sum_{j=1}^{\infty} \alpha_j^{1/2} < \infty,$$

then for each $s$ there exists a **shifted lattice rule** $Q^{\text{lattice}}_{N,s,\mathbf{z},\Delta}$ such that

$$e_{N,s,\gamma}(Q^{\text{lattice}}_{N,s,\mathbf{z},\Delta}) \le \frac{C_{\gamma,\delta}}{N^{1-\delta}} \qquad \forall \delta > 0.$$



# The convergence rate can be $N^{-1+\delta}$

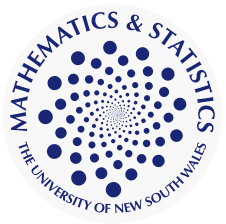
**Theorem** (IHS/Woźniakowski 01). If $N$ is prime, and if

$$\sum_{j=1}^{\infty} \alpha_j^{1/2} < \infty,$$

then for each $s$ there exists a **shifted lattice rule** $Q^{\text{lattice}}_{N,s,\mathbf{z},\Delta}$ such that

$$e_{N,s,\gamma}(Q^{\text{lattice}}_{N,s,\mathbf{z},\Delta}) \le \frac{C_{\gamma,\delta}}{N^{1-\delta}} \qquad \forall \delta > 0.$$

The original proofs of both theorems were non-constructive! They were existence proofs only. (But now ...)



# The convergence rate can be $N^{-1+\delta}$

**Theorem** (IHS/Woźniakowski 01). If $N$ is prime, and if

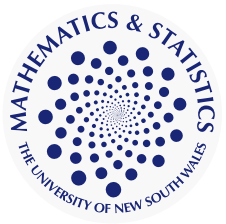
$$\sum_{j=1}^{\infty} \alpha_j^{1/2} < \infty,$$

then for each $s$ there exists a **shifted lattice rule** $Q^{\text{lattice}}_{N,s,\mathbf{z},\Delta}$ such that

$$e_{N,s,\gamma}(Q^{\text{lattice}}_{N,s,\mathbf{z},\Delta}) \le \frac{C_{\gamma,\delta}}{N^{1-\delta}} \qquad \forall \delta > 0.$$

The original proofs of both theorems were non-constructive! They were existence proofs only. (But now ...)

So now to choose $\mathbf{z}$:

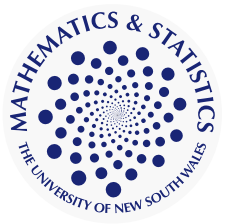


# Finding a good $z$ for product weights

For product weights the shift-averaged WCE of a lattice rule is easily computed:

$$e_{N,s,\mathbf{z}}(H)^2 = \frac{1}{N}\sum_{k=0}^{N-1}\prod_{j=1}^{s}\left(1+\alpha_j\left[B_2\left(\left\{\frac{kz_j}{N}\right\}\right)+\frac{1}{12}\right]\right)-\prod_{j=1}^{s}\left(1+\frac{\alpha_j}{12}\right),$$

where $B_2(x) := x^2 - x + 1/6$.



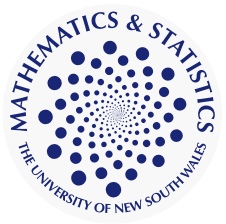
# Finding a good $z$ for product weights

For product weights the shift-averaged WCE of a lattice rule is easily computed:

$$e_{N,s,\mathbf{z}}(H)^2 = \frac{1}{N} \sum_{k=0}^{N-1} \prod_{j=1}^{s} \left(1 + \alpha_j \left[B_2\left(\left\{\frac{k z_j}{N}\right\}\right) + \frac{1}{12}\right]\right) - \prod_{j=1}^{s} \left(1 + \frac{\alpha_j}{12}\right),$$

where $B_2(x) := x^2 - x + 1/6$.

So we can compute the WCE for all possible choices of $\mathbf{z}$, and choose the best. Problem solved? No! For an exhaustive search would take $(N-1)^s$ evaluations of the worst-case error.



# Finding a good $z$ for product weights

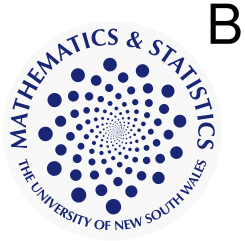
For product weights the shift-averaged WCE of a lattice rule is easily computed:

$$e_{N,s,\mathbf{z}}(H)^2 = \frac{1}{N}\sum_{k=0}^{N-1}\prod_{j=1}^{s}\left(1+\alpha_j\left[B_2\left(\left\{\frac{kz_j}{N}\right\}\right)+\frac{1}{12}\right]\right)-\prod_{j=1}^{s}\left(1+\frac{\alpha_j}{12}\right),$$

where $B_2(x) := x^2 - x + 1/6$.

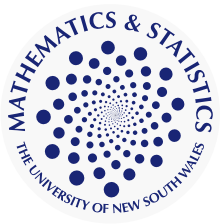
So we can compute the WCE for all possible choices of $\mathbf{z}$, and choose the best. Problem solved? No! For an exhaustive search would take $(N-1)^s$ evaluations of the worst-case error.

But there is a way: the **CBC construction**.



# CBC construction of a good $\mathbf{z}$

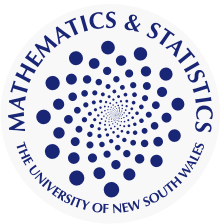
In the component-by-component (or CBC) algorithm for product weights (Korobov 1950s, IHS/Kuo/Joe 2002), a good generator $\mathbf{z} = (z_1, \ldots, z_s)$ is constructed one component at a time:



# CBC construction of a good $\mathbf{z}$

In the component-by-component (or CBC) algorithm for product weights (Korobov 1950s, IHS/Kuo/Joe 2002), a good generator $\mathbf{z} = (z_1, \ldots, z_s)$ is constructed one component at a time:

- choose $z_1 = 1$
- choose $z_2$ to minimise $e_{N,2}(z_1, z_2)$, then
- choose $z_3$ to minimise $e_{N,3}(z_1, z_2, z_3)$, then
- ...



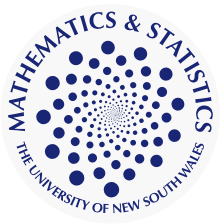
# CBC construction of a good $\mathbf{z}$

In the component-by-component (or CBC) algorithm for product weights (Korobov 1950s, IHS/Kuo/Joe 2002), a good generator $\mathbf{z} = (z_1, \ldots, z_s)$ is constructed one component at a time:

- choose $z_1 = 1$
- choose $z_2$ to minimise $e_{N,2}(z_1, z_2)$, then
- choose $z_3$ to minimise $e_{N,3}(z_1, z_2, z_3)$, then
- ...

so that at each step there are only (at most) $N - 1$ choices.

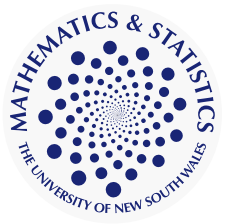
A naive implementation costs $O(s^2 N^2)$ operations.



# Fast CBC for product weights weights and shift-averaged WCEs

For product weights Cools and Nuyens (2006) developed **Fast CBC** – requires a time of order only $O(s\,N\log N)$.

The Nuyens and Cools implementation allows the CBC algorithm for product weights to be run with $s$ in thousands, $N$ in millions.



# The CBC algorithm has optimal convergence!

THEOREM Frances Kuo, J. Complexity, (2003)

*Let $N$ be prime, and let $z_1, z_2, \ldots, z_s$ be chosen by the CBC algorithm. Assume product weights, with*

$$\sum_{j=1}^{\infty} \alpha_j^{1/2} < \infty.$$

*Then* $\forall \delta > 0$

$$e_{N,s,\gamma}(\mathbf{z}) \leq \frac{C_{\gamma,\delta}}{N^{1-\delta}}.$$

Thus the optimal rate is achieved by the CBC algorithm!

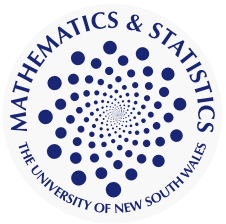
# All that remains is to choose the weights $\gamma_u$

Recall that we can bound the error in terms of the **worst case error**:

$$|I_s(F) - Q_{N,s}(\mathbf{z}; F)| \leq e_{N,s,\mathbf{z},\gamma} \|F\|_{s,\gamma},$$

or more properly,

$$\sqrt{\mathbb{E}_{\boldsymbol{\Delta}}[|I_s(F) - Q_{N,s}(\mathbf{z}, \boldsymbol{\Delta}; F)|^2]} \leq e_{N,s,\mathbf{z},\gamma} \|F\|_{s,\gamma},$$



# All that remains is to choose the weights $\gamma_u$

Recall that we can bound the error in terms of the **worst case error**:

$$|I_s(F) - Q_{N,s}(\mathbf{z}; F)| \leq e_{N,s,\mathbf{z},\gamma} \|F\|_{s,\gamma},$$

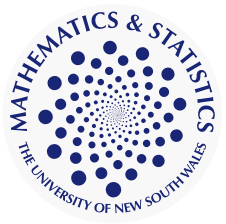
or more properly,

$$\sqrt{\mathbb{E}_{\boldsymbol{\Delta}}[|I_s(F) - Q_{N,s}(\mathbf{z}, \boldsymbol{\Delta}; F)|^2]} \leq e_{N,s,\mathbf{z},\gamma} \|F\|_{s,\gamma},$$

**Both factors on the right depend on the weights $\gamma_u$.**

**In [KSS12] we chose weights to minimise the product.**

Or rather, to minimise the product of known upper bounds on the two factors.



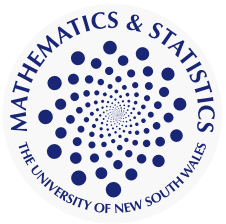
# The best weights weren't product weights!

The weights found in [KSS12] by the minimising strategy above turned out to be of the form

$$\gamma_{\mathfrak{u}} = (|\mathfrak{u}|!)^{(\text{known power bigger than 1})} \prod_{j=1}^{s} \alpha_j,$$

with known exponent, and known $\alpha_j$.

Such weights are called **POD weights** (for "product and order dependent").



# The best weights weren't product weights!

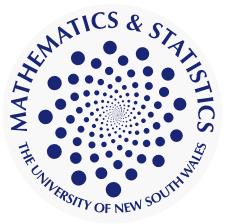
The weights found in [KSS12] by the minimising strategy above turned out to be of the form

$$\gamma_{\mathfrak{u}} = (|\mathfrak{u}|!)^{(\text{known power bigger than 1})} \prod_{j=1}^{s} \alpha_j,$$

with known exponent, and known $\alpha_j$.

Such weights are called **POD weights** (for "product and order dependent").

And fast CBC also exists for POD weights (Kuo/Schwab/IHS 11).

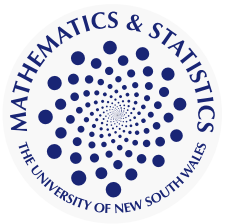


# There exist many extensions

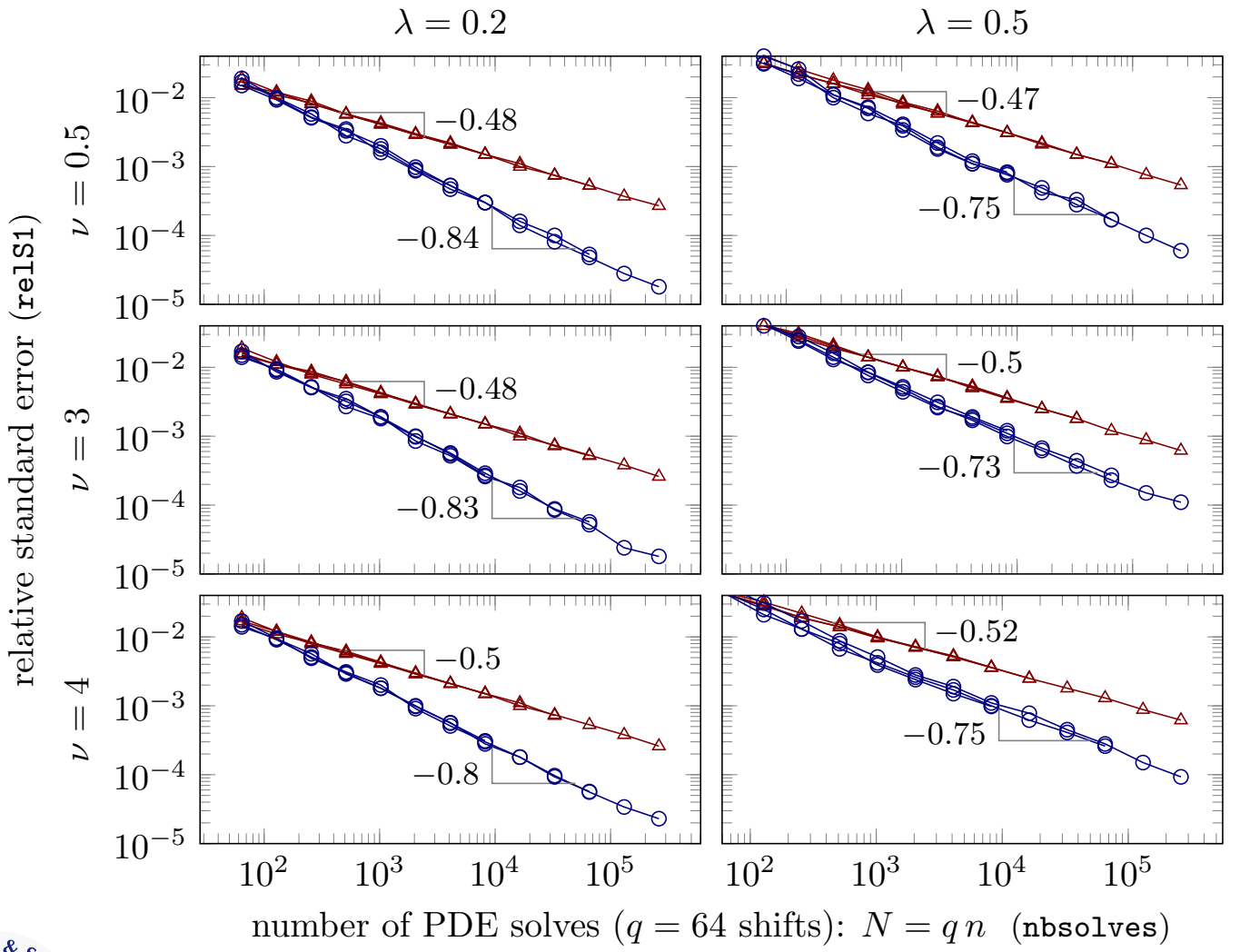
- Similar analysis of the **lognormal** case (i.e $\log a(x)$ is a Gaussian random field). (This is more difficult.)
- Analysis of **multilevel** QMC for both uniform and lognormal cases.
- **Higher order** QMC rules (Josef Dick): e.g. for the uniform case we now have

$$|I_s(F) - Q_{N,s}(F)| \le \frac{C}{N^2}\|F\|_s \le \frac{C'}{N^2}.$$

Lattice rules are now replaced by **interlaced polynomial lattice rules**, and POD weights by **SPOD** weights (standing for "smoothness-driven product and order-dependent weights"). And now the norm needs mixed **second** derivatives; and so on, to arbitrary order.



# QMC in practice

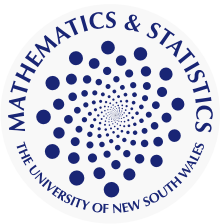



Graham/Kuo/Nuyens/Scheichl/IHS, Numer.Math. 2018

# Summarising the contemporary view of QMC:

- dependence on $s$ is as important as dependence on $N$
- applications should drive the study of dimension dependence
- that we should try to construct methods with provably good error bounds that are independent of dimension

  (or grow only slowly with dimension)



# Surveys and software

*Surveys*:

J. Dick, F. Kuo and I. Sloan, *High dimensional integration – the Quasi-Monte Carlo Way*, Acta Numerica **22** (2013).

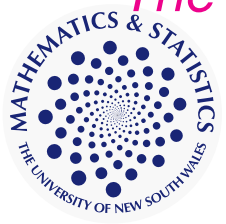
F. Kuo and D Nuyens, *Application of quasi-Monte Carlo methods to PDEs with random coefficients - a survey of analysis and implementation*, Found. Comput. Math. **16** (2016).

*The qmc4pde software from Leuven*:

http://people.cs.kuleuven.be/ dirk.nuyens/qmc4pde

*The gMLQMC software from ETH*:

http://www.sam.math.ethz.ch/HOQMC/gMLQMC



# Contributors to the contemporary view of QMC

Ronald Cools, Leuven
Josef Dick, UNSW
Takashi Goda, Tokyo
Ivan Graham, Bath
Fred Hickernell, IIT
Stephen Joe, Waikato
Frances Kuo, UNSW
Q Thong Le Gia, UNSW
James Nichols, UNSW
Dirk Nuyens, Leuven
Andrei Reztsov
Rob Scheichl, Bath
Christoph Schwab, ETH
Elisabeth Ullmann, Hamburg
Grzegorz Wasilkowski, Kentucky
Henryk Woźniakowski, Warsaw and Columbia

And apologies to those I missed!