Lifting the Curse of Dimensionality

Numerical Integration in High Dimensions

Frances Kuo

f.kuo@unsw.edu.au

University of New South Wales, Sydney, Australia

Curse of dimensionality

$\sim$ Richard Bellman (1957) $\sim$

... describes the extraordinarily rapid growth in the difficulty of problems as the number of variables (the dimension, d) increase.

e.g. The *cost* of an algorithm (the number of function evaluations, N) grows exponentially with d.

How large is d in practical applications?

- Collateralized Mortgage Obligations (CMO)

  30 years $\times$ 12 monthly repayment calculations $=$ 360 dimensions

- Daily counts of asthma patients seeking hospital treatments

  5 years $\times$ 365 days $=$ 1825 dimensions

- Macquarie Bank ALPS series (a.k.a. CEO)

  5 years $\times$ 250 trading days $\times$ 80 stocks $=$ one million dimensions

- Porous flow with permeability modeled as a random field

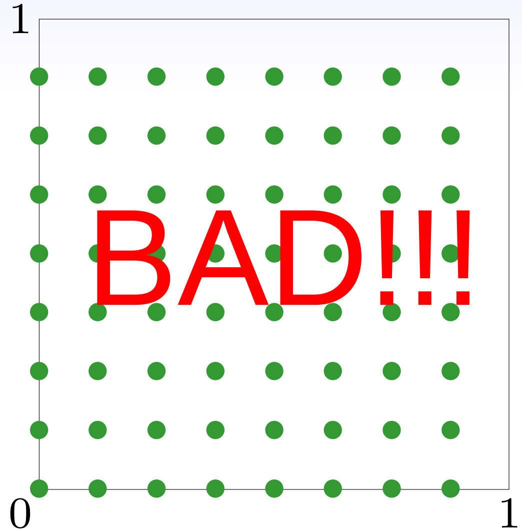
  501 by 501 mesh with circulant embedding $=$ one million dimensions

High dimensional numerical integration

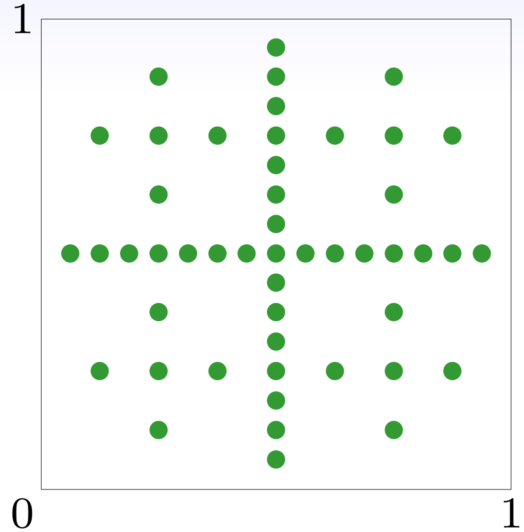
$$\int_{[0,1]^d} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

$$\approx \sum_{i=1}^{N} w_i f(\boldsymbol{t}_i)$$

$N = n^d$

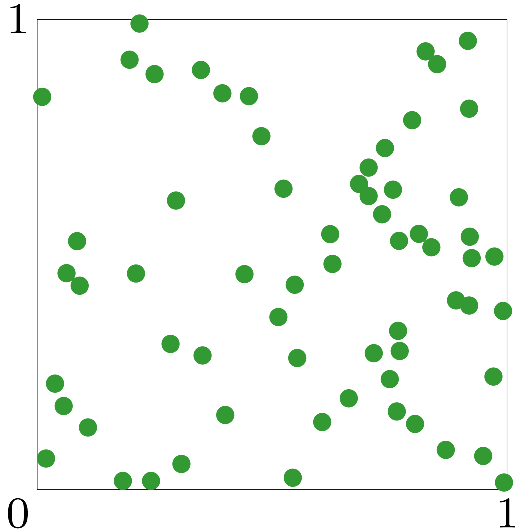


A product rule with 64 points

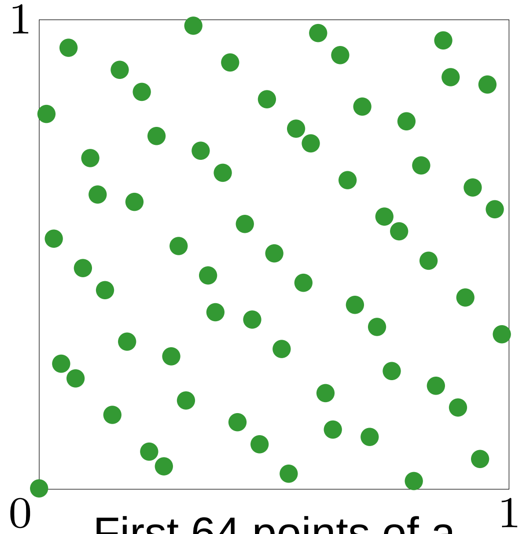


A sparse grid with 49 points

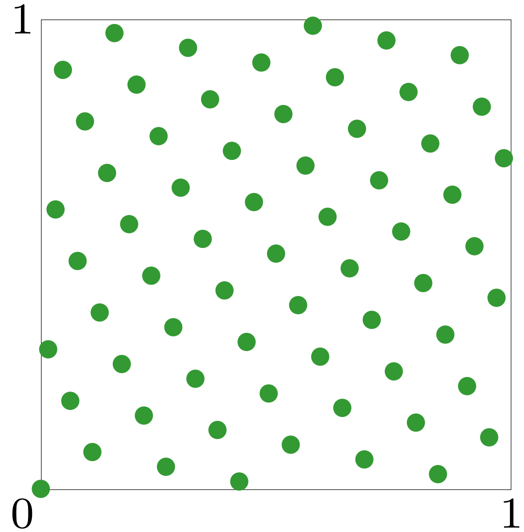
If $f(\boldsymbol{x}) = x_1$ then the error of product rule is $\mathcal{O}(n^{-1}) = \mathcal{O}(N^{-1/d})$



64 random points



First 64 points of a 2D Sobol′ sequence



A lattice rule with 64 points

MC v.s. QMC

$$\int_{[0,1]^d} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \approx \frac{1}{N}\sum_{i=1}^{N} f(\boldsymbol{t}_i)$$

Monte Carlo method

$\boldsymbol{t}_i$ random uniform

$N^{-1/2}$ convergence

order of variables irrelevant

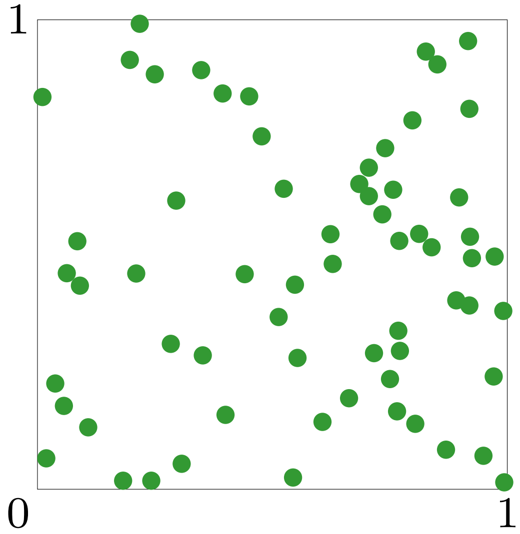
Quasi-Monte Carlo methods

$\boldsymbol{t}_i$ deterministic

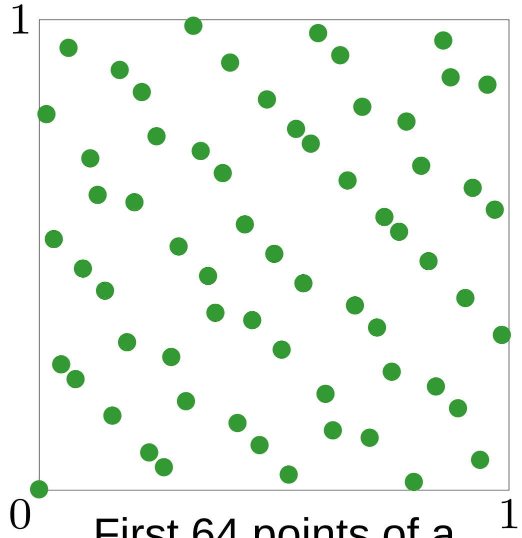
close to N^{-1} convergence (or better)

more effective for earlier variables and lower-order projections

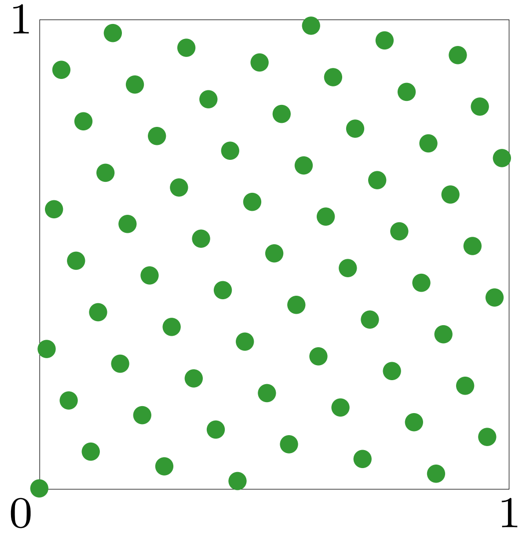
order of variables very important



64 random points



First 64 points of a 2D Sobol′ sequence



A lattice rule with 64 points

use *randomized* QMC methods for error estimation

QMC

Two types of QMC methods:

- open: infinite sequence which is independent of N (and/or d)
- closed: finite point set which depends on N (and/or d)

Two main families of QMC methods:

- **(t,m,s)-nets** (closed) and **(t,s)-sequences** (open)
- **lattice rules** (traditionally closed; now also open)

a group under addition modulo $\mathbb{Z}$ and includes the integer points

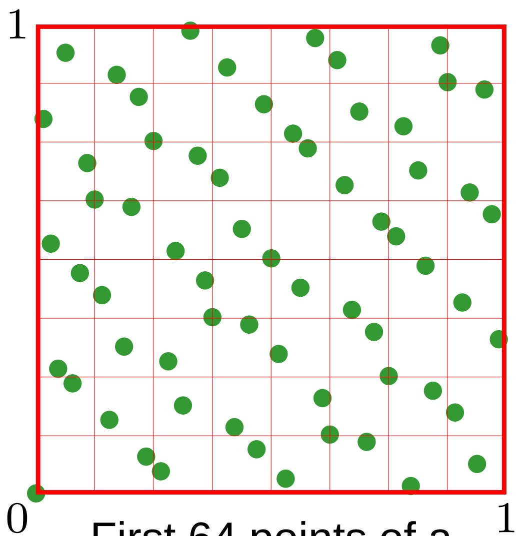
Niederreiter book (1992)

Sloan and Joe book (1994)

Kuo and Sloan AMS Notices (2005)

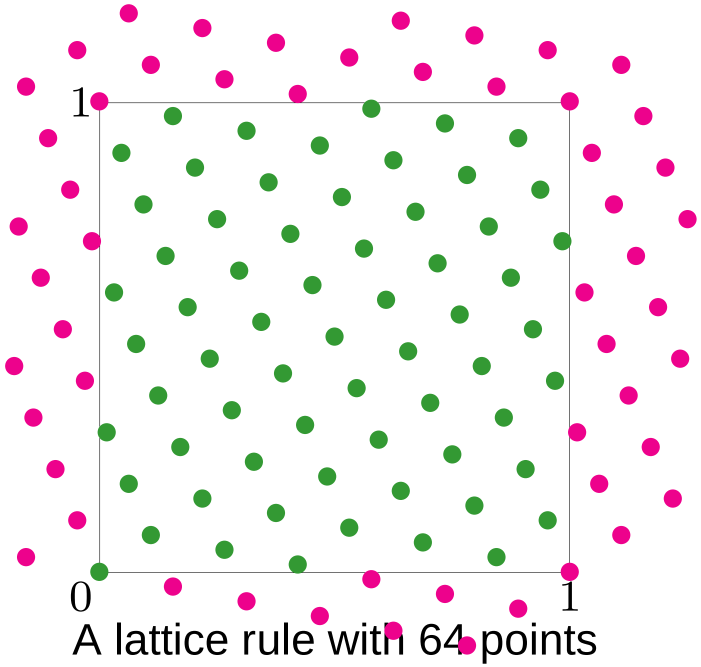
Dick and Pillichshammer book (2010)

having the right number of points in various sub-cubes



First 64 points of a 2D Sobol′ sequence

(0,6,2)-net



A lattice rule with 64 points

Overview of the “lattice” strategy

Given a complicated integral over $\mathbb{R}^d$, where d is hundreds or thousands or even more, what do we do?

$$\boxed{\int_{\mathbb{R}^d} F(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} \;=\; \cdots \;=\; \int_{[0,1]^d} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \;\approx\; \frac{1}{N}\sum_{i=1}^{N} f(\boldsymbol{t}_i)}$$

1. Transform the problem into an integral over the **unit cube** $[0,1]^d$.
2. Identify a *weighted* function space to which the transformed integrand belongs.
3. Find a lattice rule which gives a *small* worst case error.
4. Compute the lattice points $\boldsymbol{t}_i$ and approximate the integral.
5. Use a number of random shifts for error estimation.

Transformation plays a crucial role

... because it determines the features of the transformed integrand.

$$\int_{\mathbb{R}^d} F(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} \xrightarrow[\text{change of variable}]{\text{transformation}} = \int_{[0,1]^d} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \approx \frac{1}{N}\sum_{i=1}^{N} f(\boldsymbol{t}_i)$$

MC - $\boldsymbol{t}_i$ random uniform

QMC - $\boldsymbol{t}_i$ deterministic

importance sampling

$$= \int_{\mathbb{R}^d} g(\boldsymbol{u})\, p(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u}$$

$$\approx \frac{1}{N}\sum_{i=1}^{N} g(\boldsymbol{\tau}_i)$$

MC - $\boldsymbol{\tau}_i$ random samples drawn from the distribution p

DIRECT CONNECTION:

If $p(\boldsymbol{u}) = \prod_{j=1}^{d} \phi(u_j)$ $\quad \begin{cases} \phi & \text{pdf} \\ \Phi & \text{cdf} \\ \Phi^{-1} & \text{icdf} \end{cases}$

substitute $\boldsymbol{u} = \Phi^{-1}(\boldsymbol{x})$

then $f = g(\Phi^{-1}(\cdot))$, $\boldsymbol{\tau}_i \overset{\text{MC}}{\sim} \Phi^{-1}(\boldsymbol{t}_i)$

If p is *not* of a product form...

- rotation and re-scaling are required

e.g. BB/PCA/LT

- also re-centering may be necessary

Effective dimension

Caflisch, Morokoff, Owen (1997)

- **ANOVA** (ANalysis Of VAriance) **decomposition**

$$\boxed{f(\boldsymbol{x}) = \sum_{\mathfrak{u}\subseteq\{1,\ldots,d\}} f_{\mathfrak{u}}(\boldsymbol{x}_{\mathfrak{u}})}$$

  - $f_{\mathfrak{u}}$ depends only on $\boldsymbol{x}_{\mathfrak{u}} = (x_j)_{j\in\mathfrak{u}}$
  - $f_{\emptyset} = \int_{[0,1]^d} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$

e.g. $f(x_1,x_2,x_3) = f_{\emptyset} + f_{\{1\}}(x_1) + f_{\{2\}}(x_2) + f_{\{3\}}(x_3)$
$+ f_{\{1,2\}}(x_1,x_2) + f_{\{1,3\}}(x_1,x_3) + f_{\{2,3\}}(x_2,x_3)$
$+ f_{\{1,2,3\}}(x_1,x_2,x_3)$

  - unique, under the condition $\int_0^1 f_{\mathfrak{u}}(\boldsymbol{x}_{\mathfrak{u}})\,\mathrm{d}x_j = 0$ for all $j \in \mathfrak{u}$
  - orthogonal in L_2 (and in "unanchored" Sobolev space)
  - decomposition of variance $\boxed{\sum_{\mathfrak{u}\subseteq\{1,\ldots,d\}} \sigma^2(f_{\mathfrak{u}}) = \sigma^2(f)}$

- **Truncation dimension** d_T: $\sum_{\mathfrak{u}\subseteq\{1,\ldots,d_T\}} \sigma^2(f_{\mathfrak{u}}) \geq 0.99\,\sigma^2(f)$
- **Superposition dimension** d_S: $\sum_{|\mathfrak{u}|\leq d_S} \sigma^2(f_{\mathfrak{u}}) \geq 0.99\,\sigma^2(f)$

---

e.g. $f(x_1,x_2,x_3,x_4) = x_1 + \cos(x_2 x_3)$ "truncation" dimension is 3
"superposition" dimension is 2

Weighted function spaces

Sloan and Woźniakowski (1998);...

- Associate a **weight** $\gamma_{\mathfrak{u}}$ with each group of variables $\boldsymbol{x}_{\mathfrak{u}}$:

  small $\gamma_{\mathfrak{u}}$ $\implies$ f depends weakly on $\boldsymbol{x}_{\mathfrak{u}}$

  Then choose $\gamma_{\mathfrak{u}}$ to model the dimension structure of the integrand...

- Assume that f belong to a ***weighted*** ("unanchored") **Sobolev space** H.
  - It consists of functions with square-integrable mixed first derivatives.
  - The norm is weighted: ~ *the norm can also be written solely in terms of* f ~

$$\|f\|^2 = \sum_{\mathfrak{u}\subseteq\{1,\ldots,d\}} \|f_{\mathfrak{u}}\|^2, \qquad \|f_{\mathfrak{u}}\|^2 = \frac{1}{\gamma_{\mathfrak{u}}}\left\|\left(\prod_{j\in\mathfrak{u}} \frac{\partial}{\partial x_j}\right) f_{\mathfrak{u}}\right\|_{L_2}^2$$

  - It has a reproducing kernel.

$$K(\boldsymbol{x},\cdot) \in H, \quad \langle f, K(\boldsymbol{x},\cdot)\rangle = f(\boldsymbol{x}) \text{ for all } f \in H \text{ and } \boldsymbol{x} \in [0,1]^d.$$

- Analyze the **worst case error** of an integration rule in H:

$$e^{\mathrm{wor}}(\boldsymbol{t}_1,\ldots,\boldsymbol{t}_N) := \sup_{\|f\|\le 1} \left| \int_{[0,1]^d} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \frac{1}{N}\sum_{i=1}^N f(\boldsymbol{t}_i)\right|$$

~ *there is an explicit formula in terms of the reproducing kernel* ~

Then $\text{error} \le e^{\mathrm{wor}}(\boldsymbol{t}_1,\ldots,\boldsymbol{t}_N)\,\|f\|$.

Lattice rules

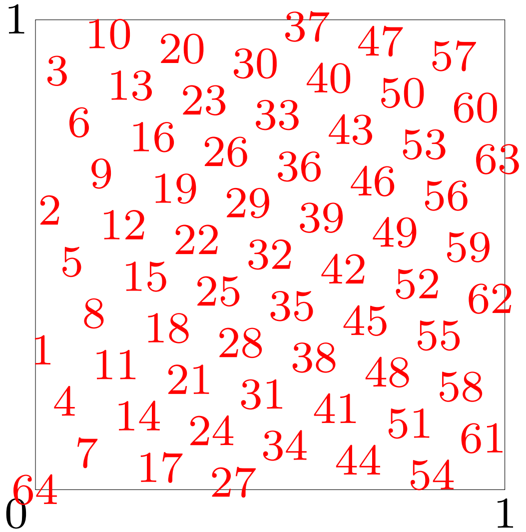
Rank-1 lattice rules have points

$$\boxed{\boldsymbol{t}_i = \mathrm{frac}\left(\frac{i}{N}\,\boldsymbol{z}\right), \qquad i = 1, 2, \ldots, N}$$

$\boldsymbol{z} \in \mathbb{Z}^d$ – the generating vector, with all components *coprime* to N

$\mathrm{frac}(\cdot)$ – means to take the fractional part of all components

$\sim$ quality determined by the choice of $\boldsymbol{z}$ $\sim$



A lattice rule with 64 points

$$N = 64 \quad \boldsymbol{z} = (1, 19) \quad \boldsymbol{t}_i = \mathrm{frac}\left(\frac{i}{64}(1, 19)\right)$$

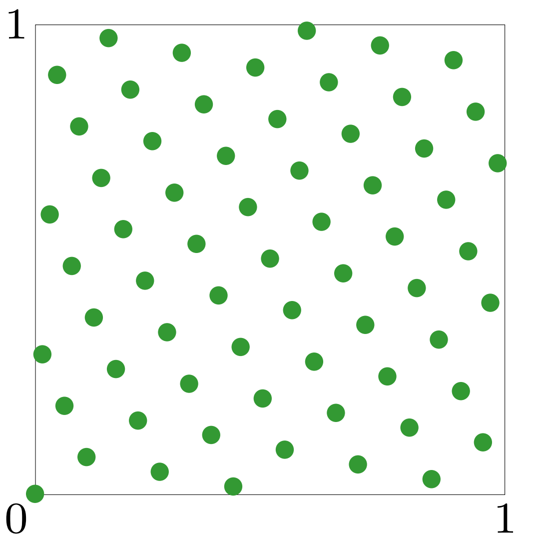
Randomly shifted lattice rules

Shifted rank-1 lattice rules have points

$$\boxed{\boldsymbol{t}_i = \mathrm{frac}\left(\frac{i}{N}\,\boldsymbol{z} + \boldsymbol{\Delta}\right), \qquad i = 1, 2, \ldots, N}$$

$\boldsymbol{\Delta} \in [0,1)^d$ – the shift

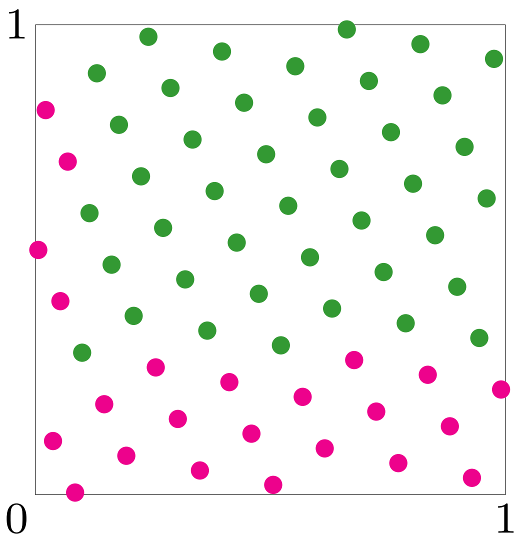
$\sim$ use a number of random shifts for error estimation $\sim$



A lattice rule with 64 points

shifted by

$\boldsymbol{\Delta} = (0.1, 0.3)$



A shifted lattice rule with 64 points

Component-by-component construction

- Want to find $\boldsymbol{z}$ with (shifted-averaged) *worst case error* as small possible.

  $\sim$ Exhaustive search is practically impossible - too many choices! $\sim$

- **CBC algorithm** [Sloan, Kuo, Joe (2002);...]
  1. Set $z_1 = 1$.
  2. With z_1 fixed, choose z_2 to minimize the worst case error in 2D.
  3. With z_1, z_2 fixed, choose z_3 to minimize the worst case error in 3D.
  4. etc.

  - Cost of algorithm is only $\mathcal{O}(N \log N\, d)$ using FFTs. [Nuyens, Cools (2005)]
  - Optimal rate of convergence $\mathcal{O}(N^{-1+\delta})$ in weighted Sobolev space, with the implied constant independent of d under an appropriate condition on the weights. [Kuo (2003); Dick (2004)]

    $\sim$ Averaging argument: there is always one choice as good as average! $\sim$

  - Extensible/embedded variants. [Cools, Kuo, Nuyens (2006); Dick, Pillichshammer, Waterhouse (2007)]

MC v.s. Randomized QMC

Monte Carlo method

$$\boxed{\bar{M} = \frac{1}{N}\sum_{i=1}^{N} f(\underset{\text{random}}{\boldsymbol{t}_i})}$$

$$\boxed{N = qn}$$

r.m.s. error $\stackrel{\text{Theory}}{=} \dfrac{\sqrt{\text{variance of } f}}{\sqrt{N}}$

r.m.s. error $\stackrel{\text{Practice}}{\approx}$

$$\sqrt{\frac{1}{N(N-1)}\sum_{i=1}^{N}\left(f(\boldsymbol{t}_i) - \bar{M}\right)^2}$$

Randomly shifted lattice rule

$$Q_k = \frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{t}_i^{(k)}), \quad k = 1, 2, \ldots, q$$

$$\boldsymbol{t}_i^{(k)} = \mathrm{frac}\left(\frac{i}{n}\underset{\text{deterministic}}{\boldsymbol{z}} + \underset{\text{random}}{\boldsymbol{\Delta}_k}\right)$$

$$\boxed{\bar{Q} = \frac{1}{q}\sum_{k=1}^{q} Q_k}$$

r.m.s. error $\stackrel{\text{Theory}}{\leq} \dfrac{C_\delta \, \|f\|}{\sqrt{q}\, n^{1-\delta}}$

r.m.s. error $\stackrel{\text{Practice}}{\approx}$

$$\sqrt{\frac{1}{q(q-1)}\sum_{k=1}^{q}\left(Q_k - \bar{Q}\right)^2}$$

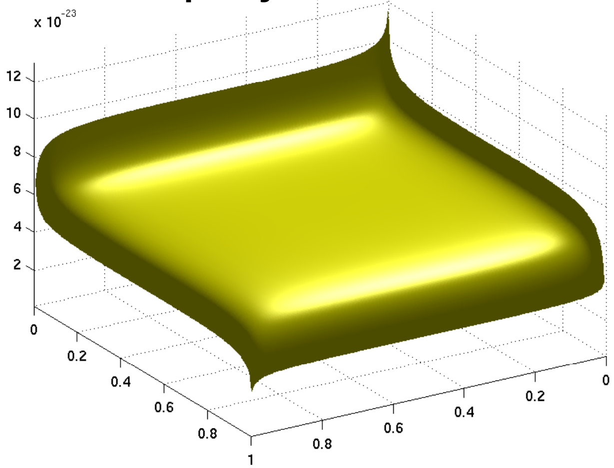
$\sim$ Randomized QMC methods combine the best of two worlds $\sim$

faster rate of convergence + unbiased + simple error estimation

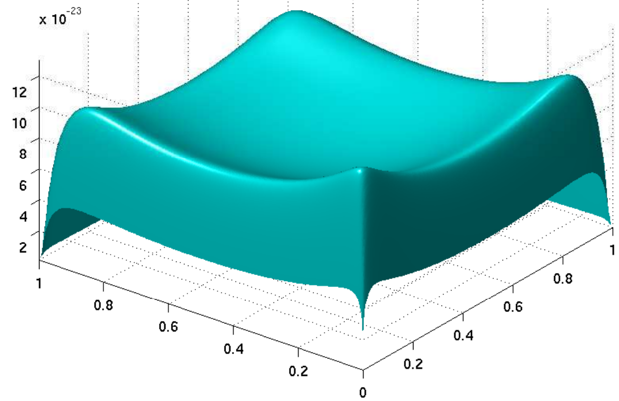
Applications from statistics

[Kuo, Dunsmuir, Sloan, Wand, Womersley (2008)]

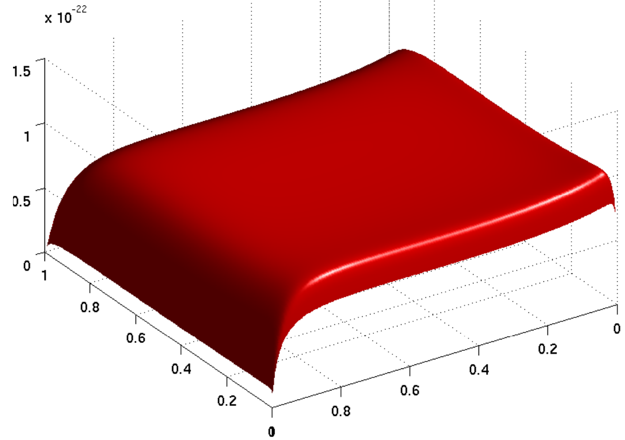
2D projections of transformed integrands from maximum likelihood problems:



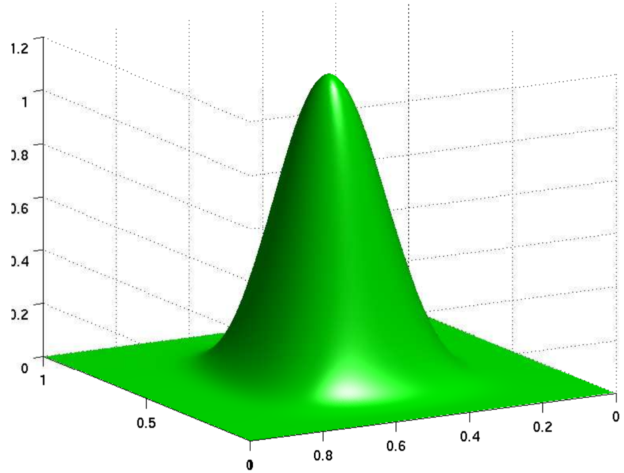
Normal (good)



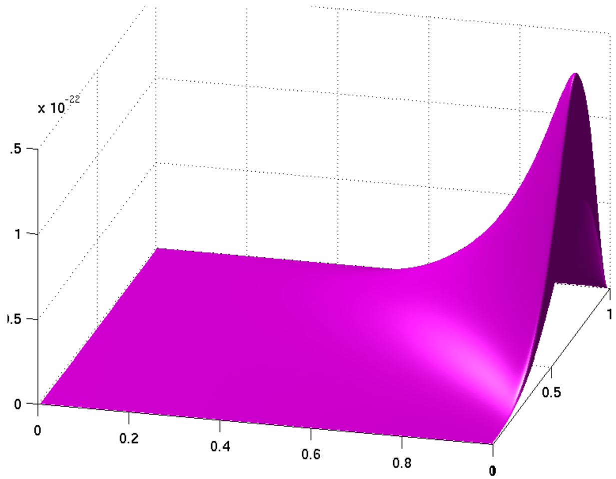
Logistic (better)



Student-t (best)



no re-scaling (bad)

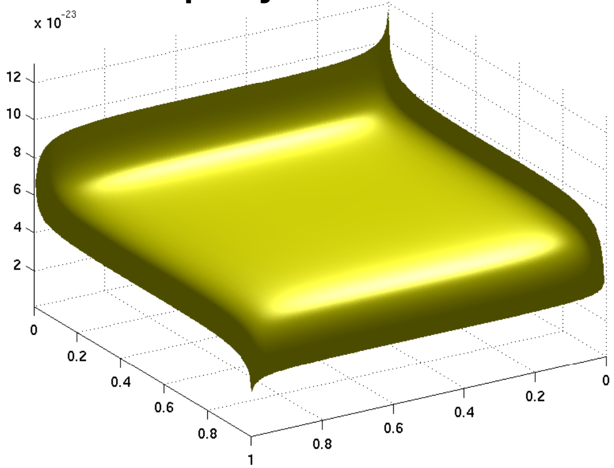


no centering (worse)

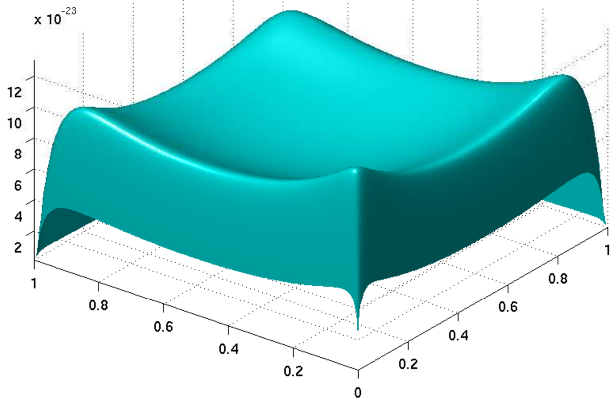
Applications from statistics

[Kuo, Dunsmuir, Sloan, Wand, Womersley (2008)]

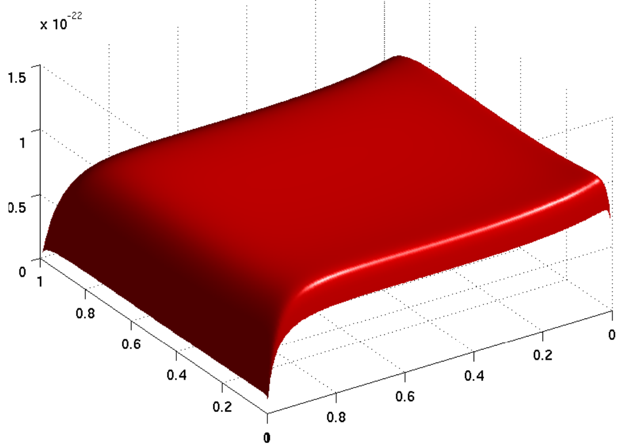
2D projections of transformed integrands from maximum likelihood problems:



Normal (good)



Logistic (better)



Student-t (best)

These integrands fail to lie in our weighted Sobolev space, because

- they are unbounded near the boundary of the unit cube, or
- they have huge derivatives near the boundary of the unit cube.

∼ our nice QMC theory cannot be applied ∼

NEW theoretical analysis [Kuo, Sloan, Wasilkowski, Waterhouse (2010)]

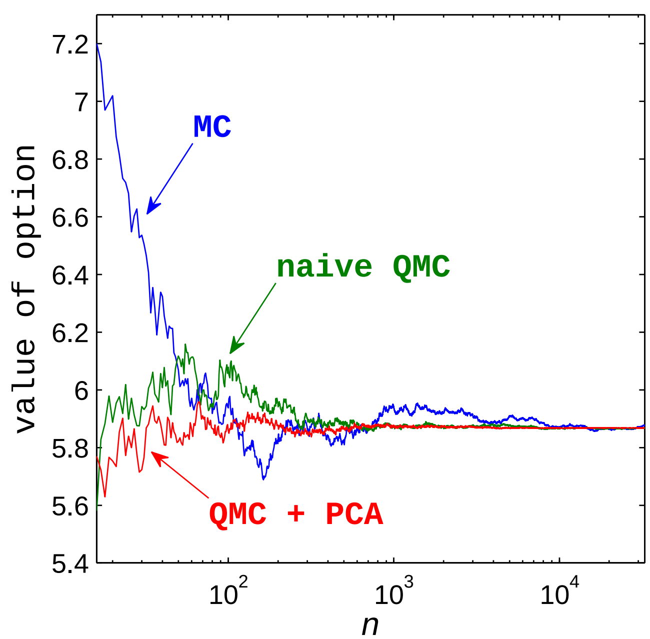
- a different reproducing kernel Hilbert space (also weighted) which includes these integrands (and more) by introducing a weight function in the norm
- optimal rate of convergence achieved by randomly-shifted lattice rules

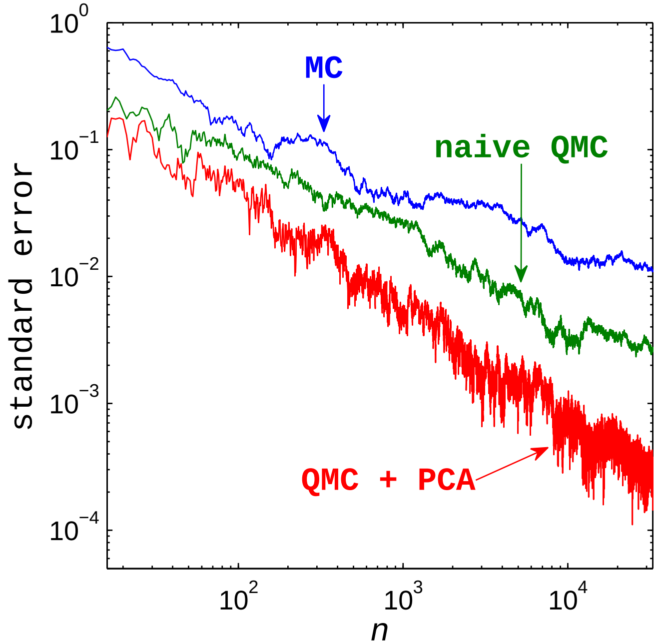
Applications from finance

[Giles, Kuo, Sloan, Waterhouse (2008)]

Arithmetic-average Asian call option with 5 stocks and 256 time steps

$$\text{payoff} = \max\left(\frac{1}{5 \times 256} \sum_{s=1}^{5} \sum_{j=1}^{256} (\text{price of stock } s \text{ at time } t_j)\ ,\ 0\right)$$





- Black-Scholes model: stock price follows a geometric Brownian motion
- Path construction (affects the relative importance of integration variables):

  RW - random walk, BB - Brownian bridge, PCA - principal components analysis

Ordering the variables is crucial for the success of QMC!

Applications from finance

[Giles, Kuo, Sloan, Waterhouse (2008)]

Arithmetic-average Asian call option with 5 stocks and 256 time steps

$$\text{payoff} = \max\left(\frac{1}{5 \times 256} \sum_{s=1}^{5} \sum_{j=1}^{256} (\text{price of stock } s \text{ at time } t_j)\ ,\ 0\right)$$

The associated integrands fail to lie in our weighted Sobolev space, because

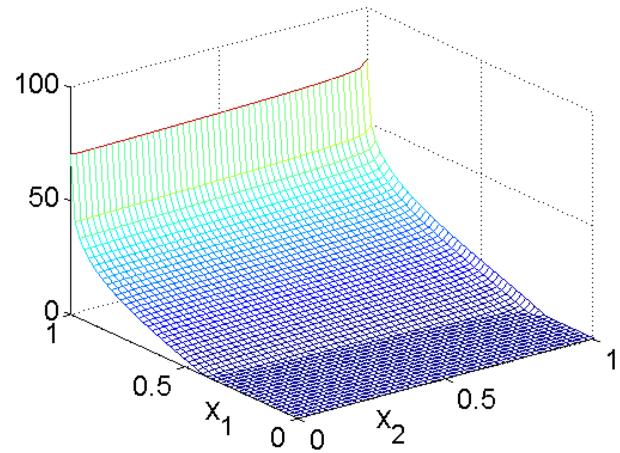
- they are unbounded near the boundary of the unit cube, and
- they have kinks, i.e., no square-integrable mixed first derivatives.

$\sim$ *our nice QMC theory cannot be applied* $\sim$

Why does QMC still work?

... low effective dimension (under BB/PCA)

recall ANOVA decomposition $f = \sum_{\mathfrak{u} \subseteq \{1,\ldots,d\}} f_{\mathfrak{u}}$



NEW theoretical analysis [Griebel, Kuo, Sloan (2010)]

- RW/BB: all $f_{\mathfrak{u}}$ with $|\mathfrak{u}| \le \frac{d+1}{2}$ belong to our Sobolev space
- PCA: similar

Applications from physics

[Graham, Kuo, Nuyens, Scheichl, Sloan (2010)]

Flow through random porous media

PDEs: Darcy's law $\vec{q} + k\,\vec{\nabla} p = \vec{0}$

mass conservation law $\vec{\nabla} \cdot \vec{q} = 0$

on the unit square in 2D

Input: permeability is a lognormal random field $k(\vec{x},\omega) = \exp(Z(\vec{x},\omega))$

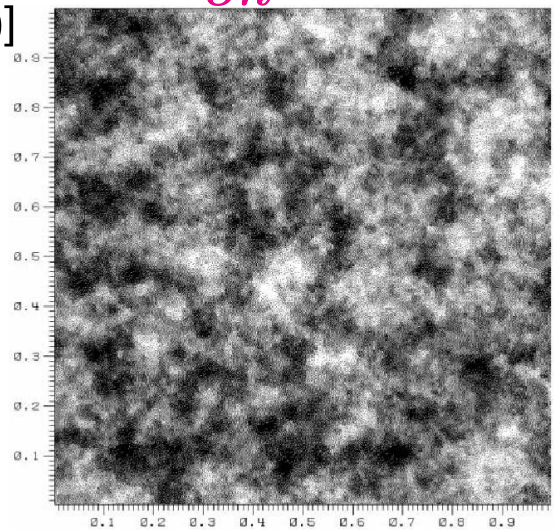
Unknowns to be determined: velocity $\vec{q}$ and pressure p

Boundary conditions:

$\dfrac{\partial p}{\partial n} = 0$

[Cliffe, Graham, Scheichl, Stals (2000)]

$p = 1$



$p = 0$

$\dfrac{\partial p}{\partial n} = 0$

Quantities of interest:

- pressure head at a point
- effective permeability
- breakthrough time

Applications from physics

[Graham, Kuo, Nuyens, Scheichl, Sloan (2010)]

Flow through random porous media

- Require the permeability field only at a discrete set of M points
  - ✗ truncate the Karhunen-Loève expansion of the covariance function
  - ✓ factorize the covariance matrix
- Use QMC to obtain N realizations of the permeability field
  - circulant embedding and FFT – fast factorization of covariance matrix
  - very high dimensional integral $d = \mathcal{O}(M)$

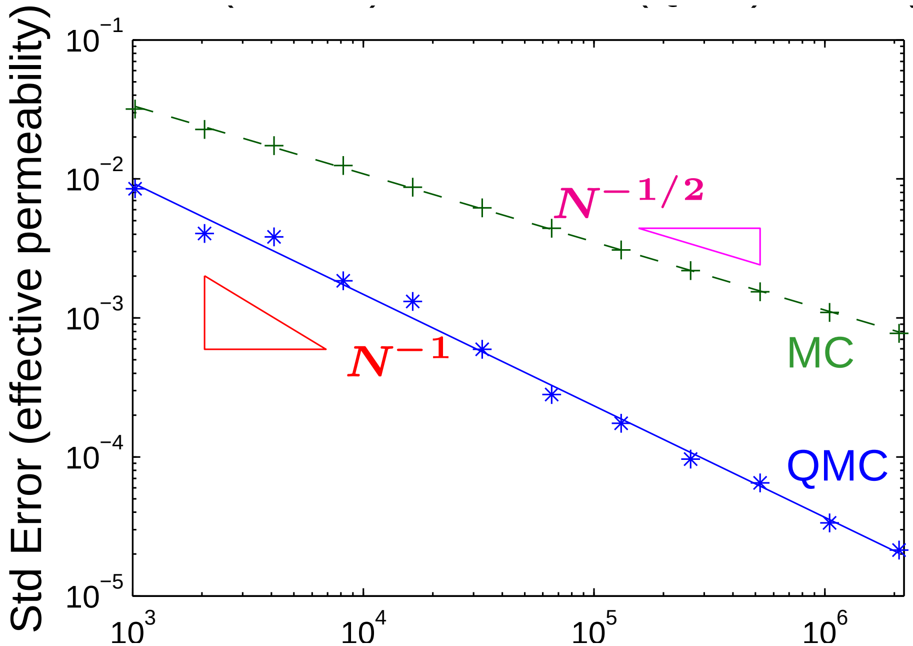
    e.g. d is **one million** for a 501×501 grid
- For each realization, solve the PDE by mixed finite element method
  - Raviart-Thomas elements
  - divergence-free reduction – solve an auxiliary problem using standard FEM
  - algebraic multigrid `amg1r5.f`
- Total cost is $\mathcal{O}(N\, M \log M)$ and is highly parallelizable
- Balance the discretization error with quadrature error (no truncation error)
- Multilevel technique should work well

Applications from physics

[Graham, Kuo, Nuyens, Scheichl, Sloan (2010)]

Flow through random porous media

Example: effective permeability



$$r(\vec{x},\vec{y}) = \sigma^2 \exp\left(-\frac{\|\vec{x}-\vec{y}\|_1}{\lambda}\right)$$

$\sigma^2 = 1$, $\lambda = 1$

$M = 129 \times 129$, $d = 16384$

To have discretization error $< 10^{-3}$ and quadrature error $< 10^{-3}$...

σ^2	λ	M	N_{QMC}		N_{MC}	
1	0.3	129×129	16700	(28 min)	982000	(28 h)
1	0.1	513×513	4900	(3 h)	142000	(3 d)
3	0.1	1025×1025	28700	(35 h)	525000	(67 d)

Summary

- Product rules are bad
- Sparse grids have pros and cons...
- MC method converges slowly
- QMC methods are equal-weight quadrature rules over the unit cube
  - Transformation to the unit cube plays a crucial role (also for MC and SG)
  - Better convergence rates than MC
  - Good for earlier variables and lower-order projections
  - Ordering the variables is very important
  - Randomized QMC:
    unbiased, simple error estimation, good convergence rate
- Challenges:
  - Integrands from practical problems do not fit into existing QMC theory
  - How to choose the weights?
  - In need of ultra-high dimensional QMC for $d > N$