Quasi-Monte Carlo for Bayesian optimal experimental design

Vesa Kaarnioja (FU Berlin)

With special thanks to collaborators: Claudia Schillings (FU Berlin), Tapio Helin (LUT University), Philipp Wacker (University of Canterbury)

Table of contents

Part I: Quasi-Monte Carlo methods

High-dimensional numerical integration

$$\int_{[0,1]^s} f(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} \approx \sum_{i=1}^{n} w_i f(\boldsymbol{t}_i)$$

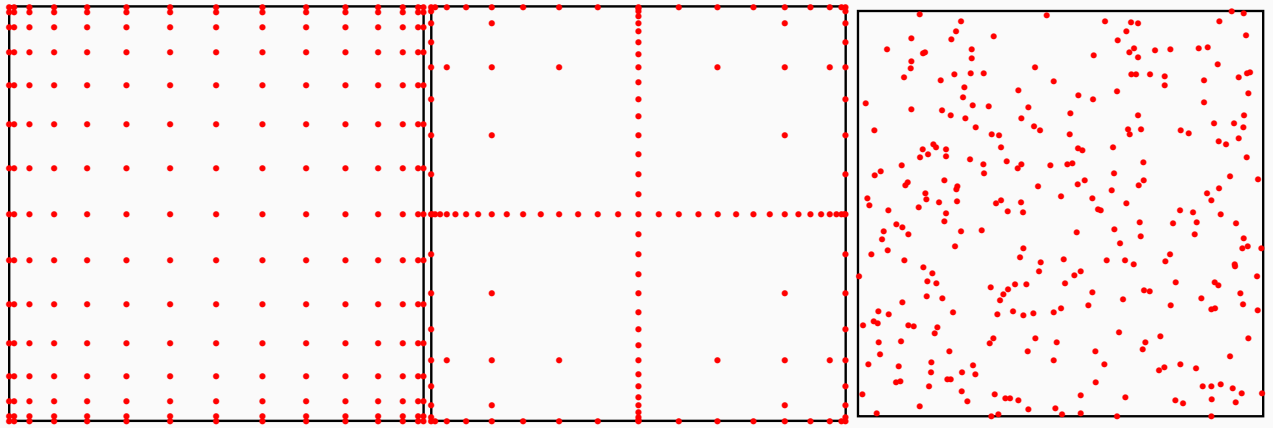


Figure 1: Tensor product grid, sparse grid, Monte Carlo nodes (not QMC rules)

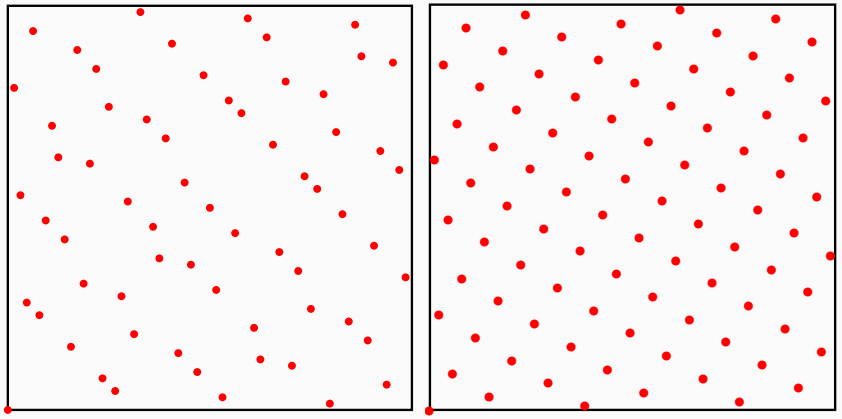


Figure 2: Sobol′ points, lattice rule (examples of QMC rules)

Quasi-Monte Carlo (QMC) methods are a class of *equal weight* cubature rules

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

where $(\boldsymbol{t}_i)_{i=1}^n$ is an ensemble of *deterministic* nodes in $[0,1]^s$.

The nodes $(\boldsymbol{t}_i)_{i=1}^n$ are NOT random!! Instead, they are *deterministically chosen*.

QMC methods exploit the smoothness and anisotropy of an integrand in order to achieve better-than-Monte Carlo rates.

Lattice rules

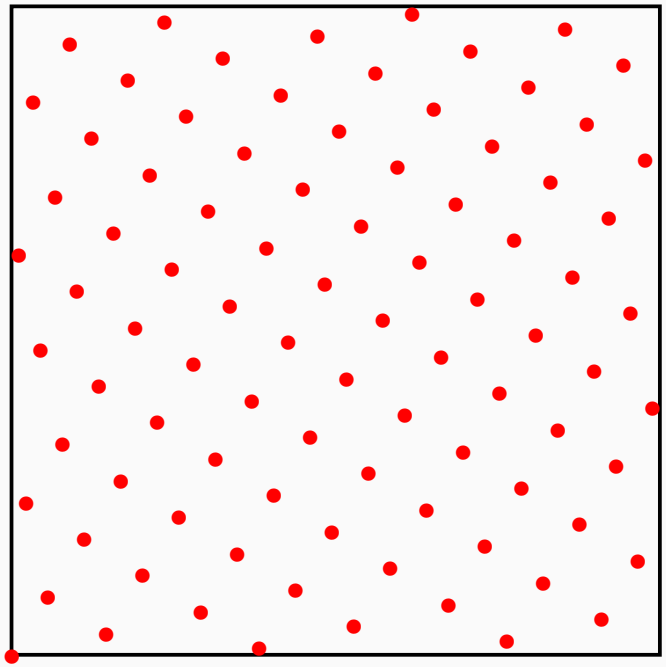
Rank-1 lattice rules

$$Q_{s,n}(f) = \frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{t}_i)$$

have the points

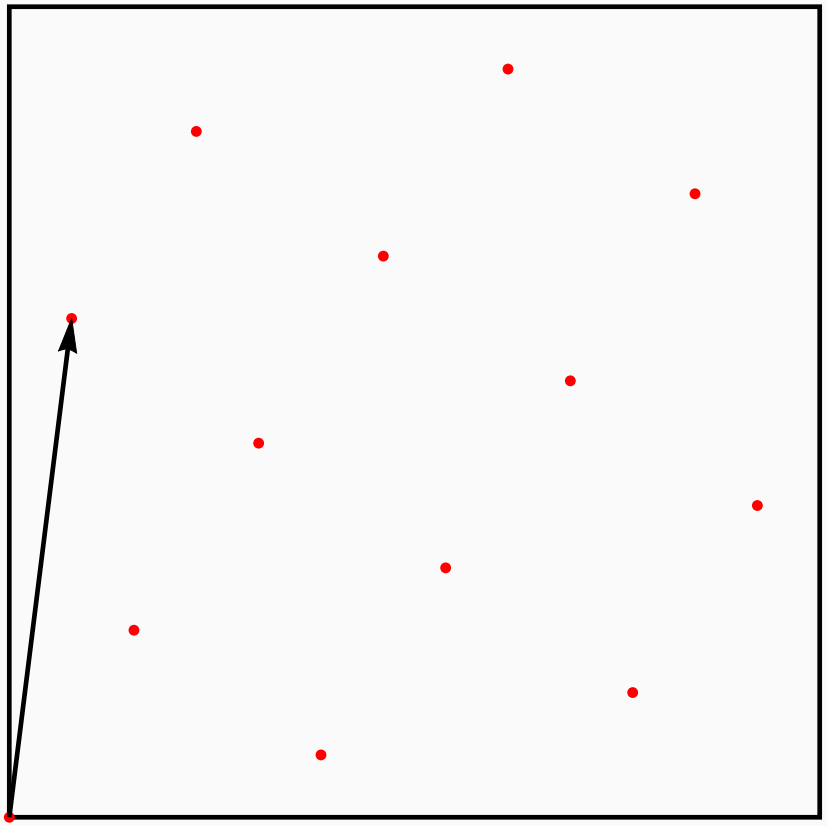
$$\boldsymbol{t}_i = \mathrm{mod}\left(\frac{i\boldsymbol{z}}{n}, 1\right), \quad i \in \{1, \ldots, n\},$$

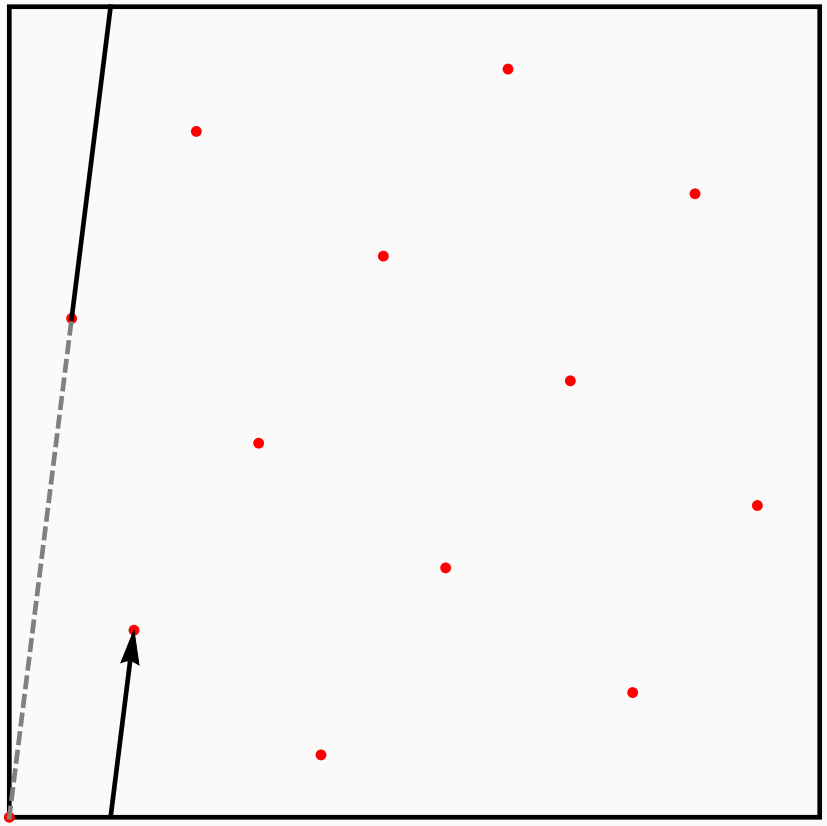
where the entire point set is determined by the *generating vector* $\boldsymbol{z} \in \mathbb{N}^s$, with all components *coprime* to n.

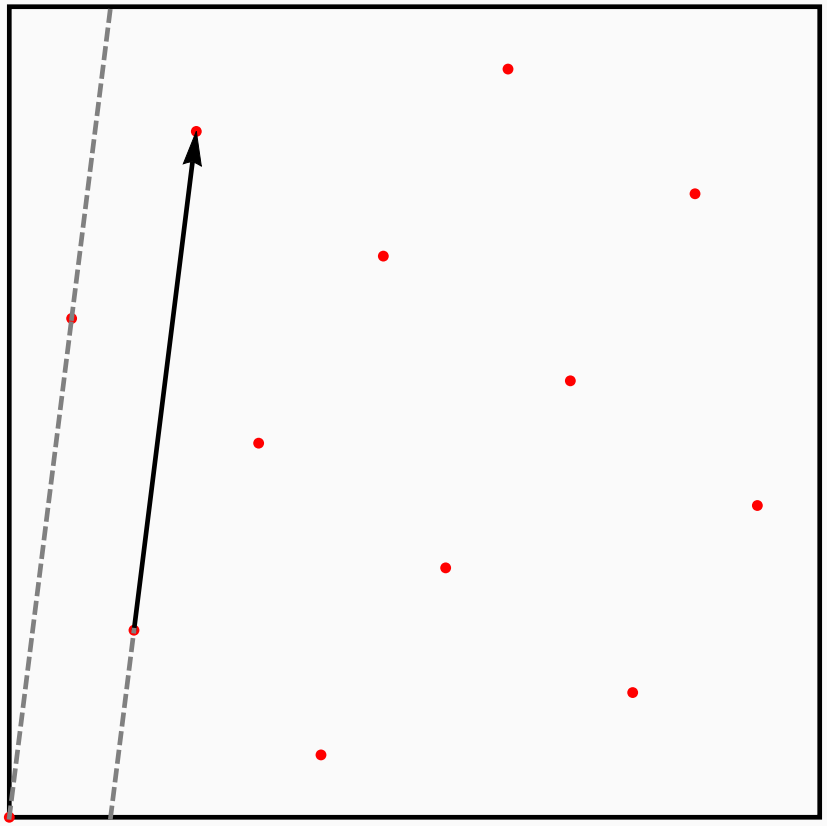


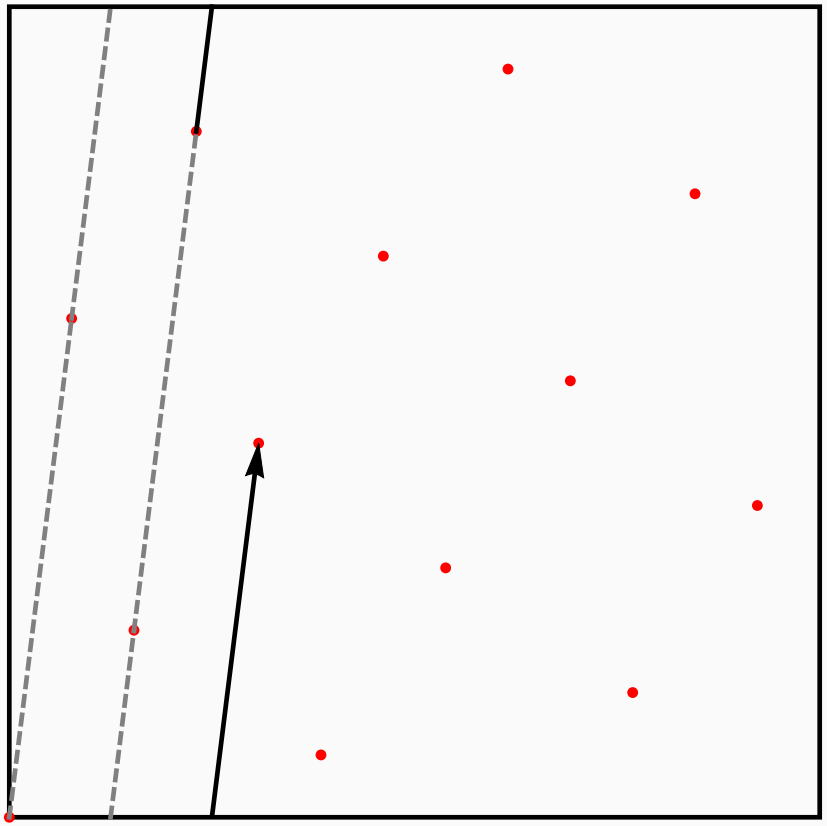
Lattice rule with $\boldsymbol{z} = (1, 55)$ and $n = 89$ nodes in $[0, 1]^2$

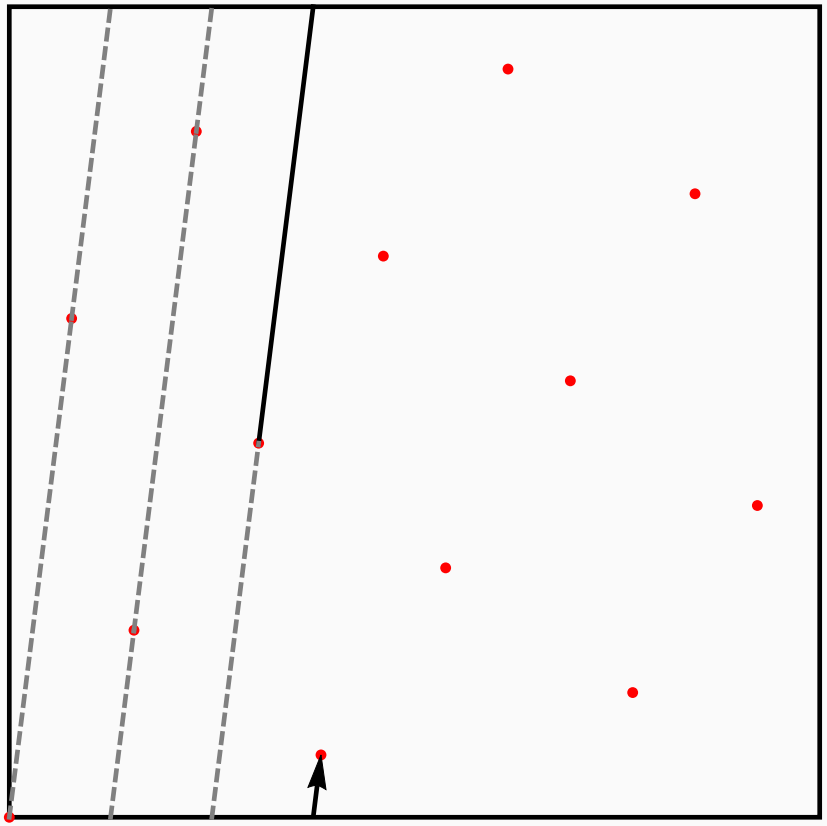
The quality of the lattice rule is determined by the choice of $\boldsymbol{z}$.

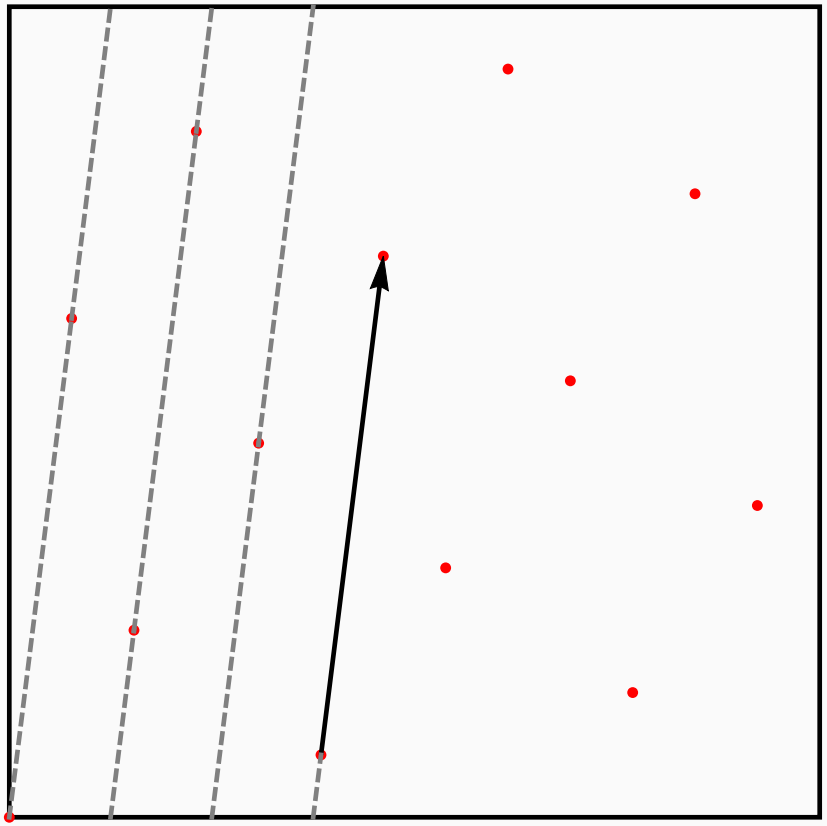


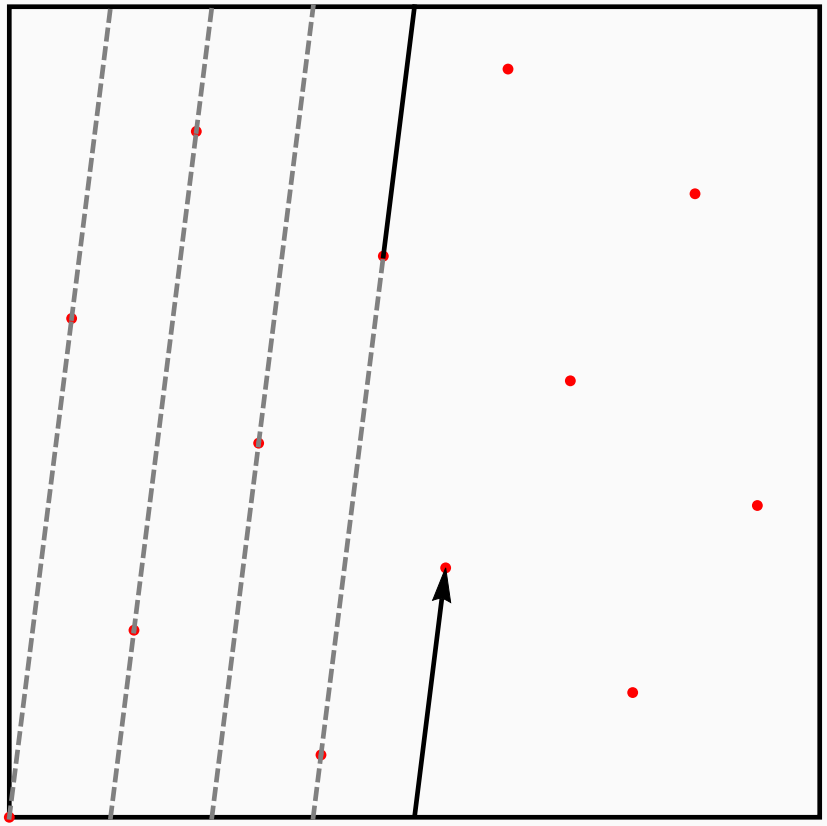


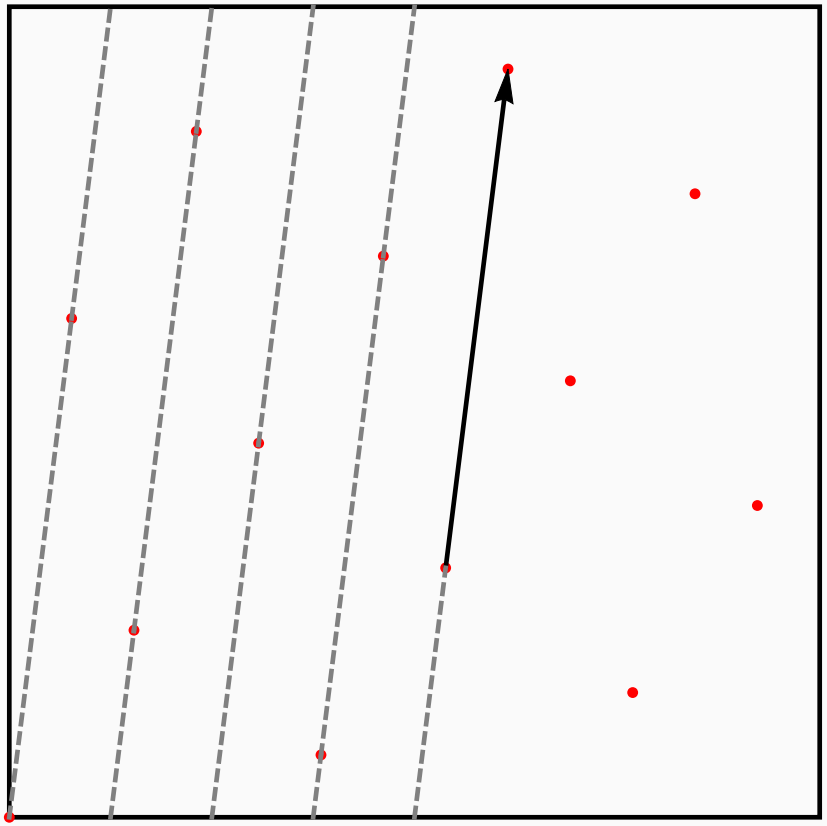


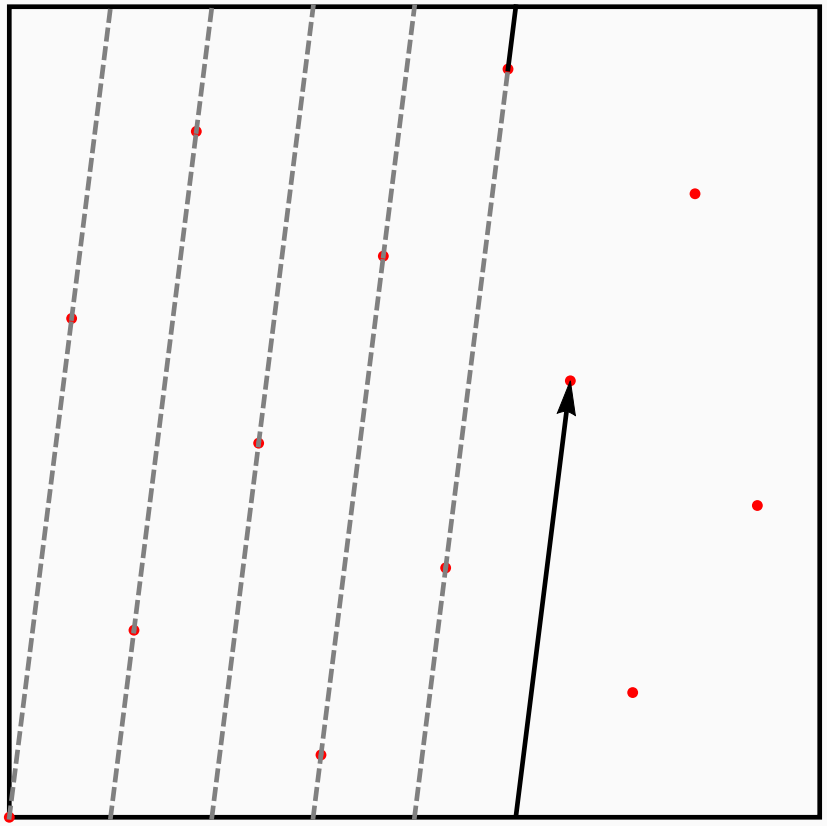


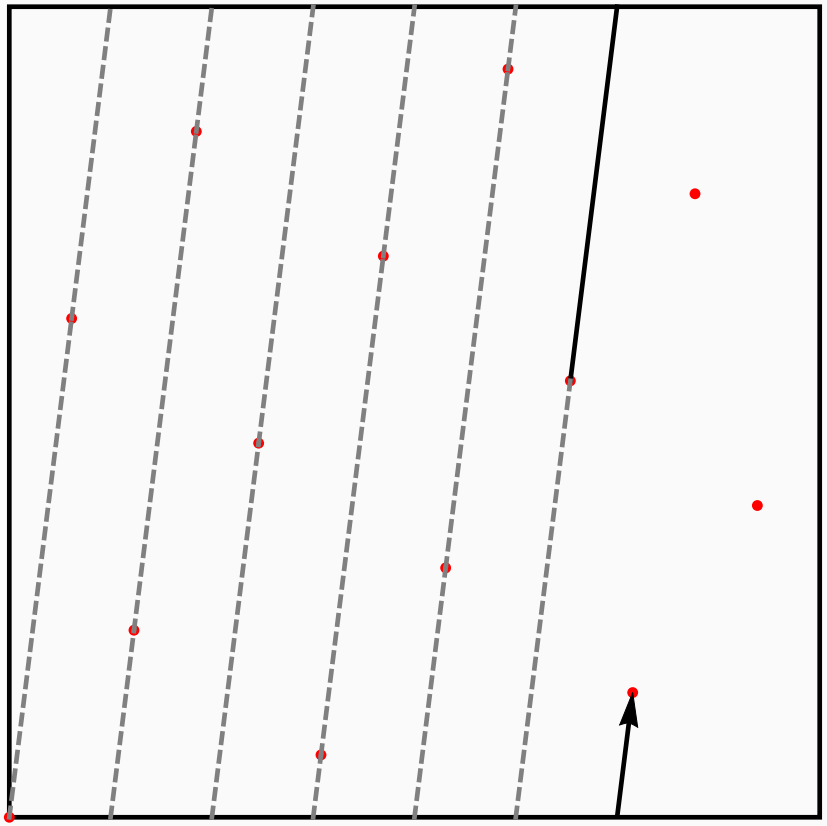


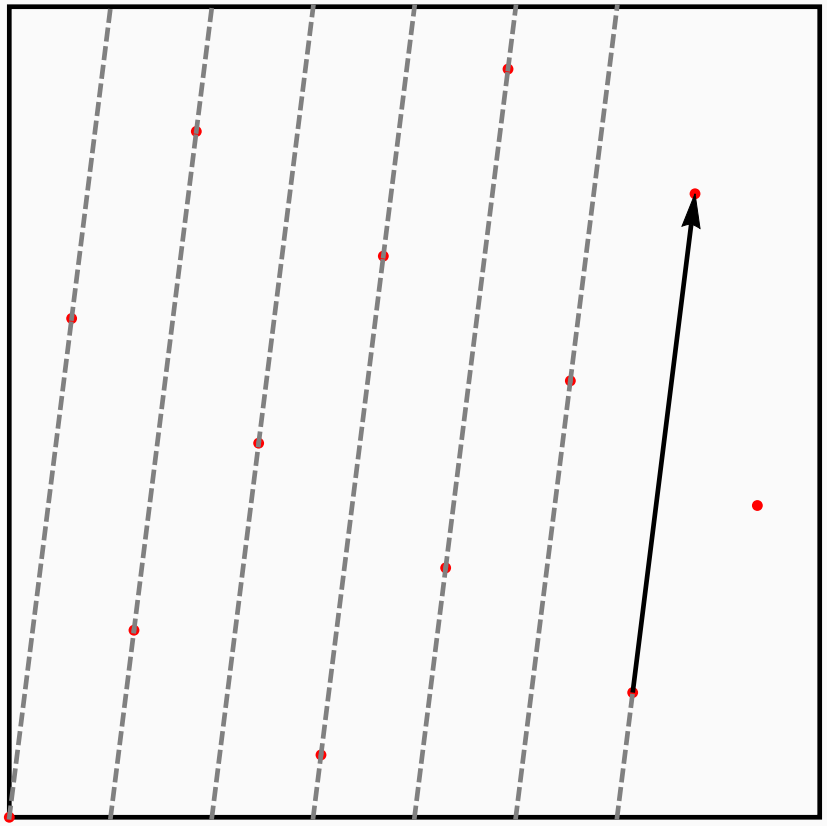


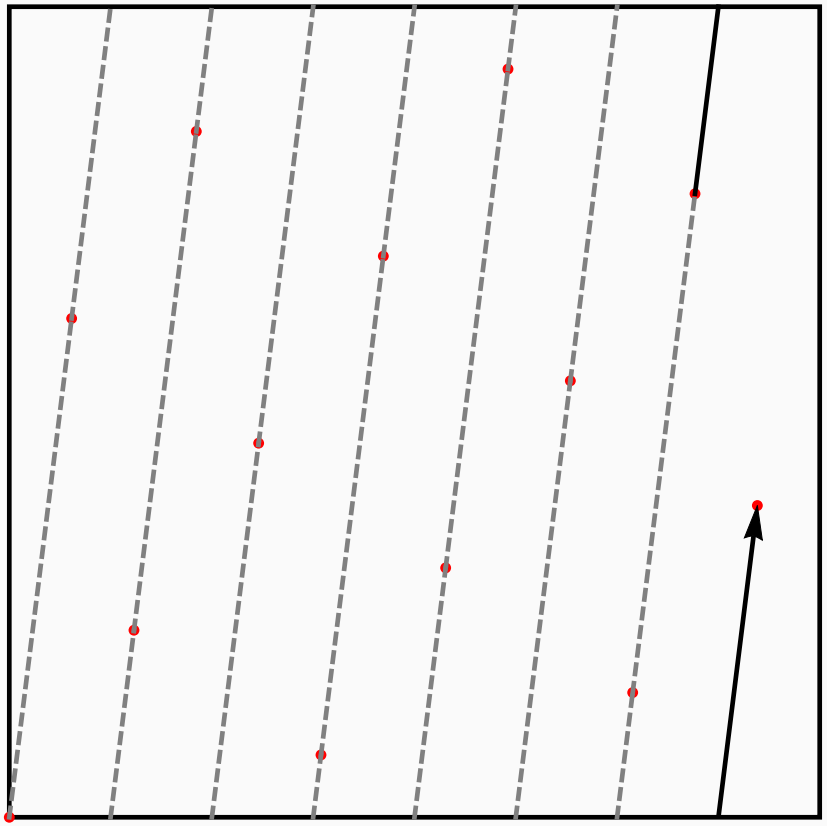












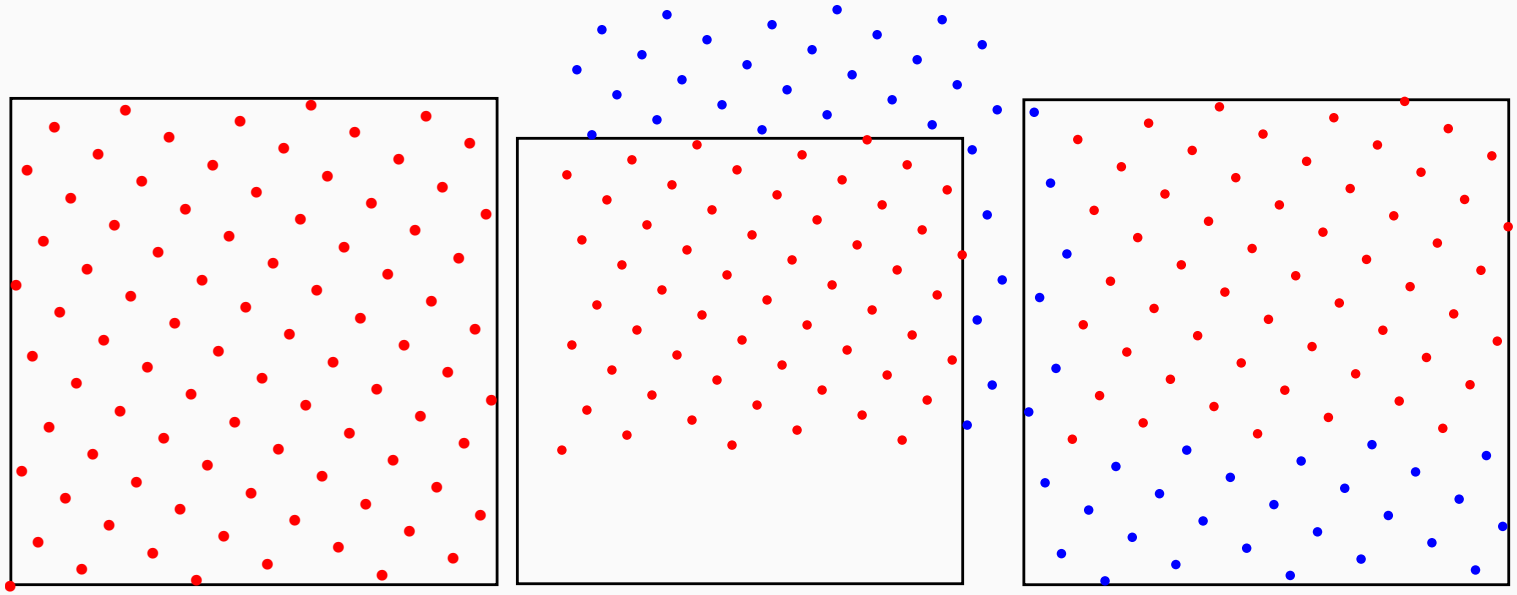
Randomly shifted lattice rules

Shifted rank-1 lattice rules have points

$$\boldsymbol{t}_i = \mathrm{mod}\left(\frac{i\boldsymbol{z}}{n} + \boldsymbol{\Delta}, 1\right), \quad i \in \{1, \ldots, n\}.$$

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

Use a number of random shifts for error estimation.



Lattice rule shifted by $\boldsymbol{\Delta} = (0.1, 0.3)$.

Let $\boldsymbol{\Delta}^{(r)}$, $r = 1, \ldots, R$, be independent random shifts drawn from $U([0,1]^s)$ and define

$$Q_{s,n}^{(r)}(f) := \frac{1}{n} \sum_{i=1}^{n} f(\mathrm{mod}(\boldsymbol{t}_i + \boldsymbol{\Delta}^{(r)}, 1)). \quad \text{(QMC rule with 1 random shift)}$$

Then

$$\overline{Q}_{s,n}(f) = \frac{1}{R} \sum_{r=1}^{R} Q_{s,n}^{(r)} f \qquad \text{(QMC rule with } R \text{ random shifts)}$$

is an unbiased estimator of $I_s(f)$.

Let $f : [0,1]^s \to \mathbb{R}$ be sufficiently smooth.

Error bound (one random shift):

$$|I_s(f) - Q^{\boldsymbol{\Delta}}_{s,n}(f)| \le e^{\boldsymbol{\Delta}}_{s,n,\boldsymbol{\gamma}}(\boldsymbol{z})\|f\|_{\boldsymbol{\gamma}}.$$

R.M.S. error bound (shift-averaged):

$$\sqrt{\mathbb{E}_{\boldsymbol{\Delta}}[|I_s(f) - \overline{Q}_{s,n}(f)|^2]} \le e^{\mathrm{sh}}_{s,n,\boldsymbol{\gamma}}(\boldsymbol{z})\|f\|_{\boldsymbol{\gamma}}.$$

We consider weighted Sobolev spaces with dominating mixed smoothness, equipped with norm

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

and (squared) worst case error

$$P(\boldsymbol{z}) := e^{\mathrm{sh}}_{s,n,\boldsymbol{\gamma}}(\boldsymbol{z})^2 = \frac{1}{n}\sum_{k=0}^{n-1} \sum_{\varnothing\neq\mathfrak{u}\subseteq\{1:s\}} \gamma_{\mathfrak{u}} \prod_{j\in\mathfrak{u}} \omega\left(\left\{\frac{kz_j}{n}\right\}\right)$$

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

CBC algorithm (Sloan, Kuo, Joe 2002)

The idea of the *component-by-component* (CBC) algorithm is to find a good generating vector $\boldsymbol{z} = (z_1, \ldots, z_s)$ by proceeding as follows:

1. Set $z_1 = 1$ (this is a freebie since $P(1) = P(z)$ for all $z \in \mathbb{N}$);
2. With z_1 fixed, choose z_2 to minimize error criterion $P(z_1, z_2)$;
3. With z_1 and z_2 fixed, choose z_3 to minimize error criterion $P(z_1, z_2, z_3)$

   $\vdots$

- The CBC algorithm is a *greedy algorithm*: in general, it will not find the generating vector $\boldsymbol{z}$ that minimizes $P(\boldsymbol{z})$. However, it can be shown that the generating vector obtained by the CBC algorithm satisfies an error bound (more on this later).
- For generic $\boldsymbol{\gamma} = (\gamma_{\mathfrak{u}})_{\mathfrak{u} \subseteq \{1:s\}}$, evaluating $P(\boldsymbol{z}) = P(\boldsymbol{\gamma}, \boldsymbol{z})$ takes $\mathcal{O}(2^s)$ operations. For an efficient implementation, it is desirable that the weights $\boldsymbol{\gamma}$ can be characterized by an expression that does not contain too many degrees of freedom.
- Efficient implementation using FFT! (QMC4UQ, QMCPy, etc.)

Theorem (CBC error bound)
The generating vector $\boldsymbol{z} \in \mathbb{U}_n^s$ constructed by the CBC algorithm, minimizing the squared shift-averaged worst-case error $[e^{\mathrm{sh}}_{s,n,\boldsymbol{\gamma}}(\boldsymbol{z})]^2$ for the weighted unanchored Sobolev space in each step, satisfies

$$[e^{\mathrm{sh}}_{s,n,\boldsymbol{\gamma}}(\boldsymbol{z})]^2 \le \left(\frac{1}{\varphi(n)} \sum_{\varnothing \ne \mathfrak{u} \subseteq \{1:s\}} \gamma_{\mathfrak{u}}^{\lambda} \left(\frac{2\zeta(2\lambda)}{(2\pi^2)^{\lambda}}\right)^{|\mathfrak{u}|}\right)^{1/\lambda} \quad \text{for all } \lambda \in (1/2, 1],$$

where $\zeta(x) := \sum_{k=1}^{\infty} k^{-x}$ denotes the Riemann zeta function for $x > 1$.

Remarks:

- Optimal rate of convergence $\mathcal{O}(n^{-1+\delta})$ in weighted Sobolev spaces, independently of s under an appropriate condition on the weights.
- Cost of algorithm for POD weights is $\mathcal{O}(s\, n \log n + s^2\, n)$ using FFT.
- Fast CBC works for any (composite) number $n \ge 2$, but the implementation is more involved when n is not prime.

Significance: Suppose that $f \in H_{s,\boldsymbol{\gamma}}$ for all $\boldsymbol{\gamma} = (\gamma_{\mathfrak{u}})_{\mathfrak{u} \subseteq \{1:s\}}$. Then for any given sequence of weights $\boldsymbol{\gamma}$, we can use the CBC algorithm to obtain a generating vector satisfying the error bound

$$\sqrt{\mathbb{E}_{\boldsymbol{\Delta}} |I_s f - Q_{n,s}^{\boldsymbol{\Delta}} f|^2} \le \left(\frac{1}{\varphi(n)} \sum_{\varnothing \ne \mathfrak{u} \subseteq \{1:s\}} \gamma_{\mathfrak{u}}^{\lambda} \left(\frac{2\zeta(2\lambda)}{(2\pi^2)^{\lambda}} \right)^{|\mathfrak{u}|} \right)^{1/(2\lambda)} \|f\|_{s,\boldsymbol{\gamma}} \tag{1}$$

for all $\lambda \in (1/2, 1]$. We can use the following strategy:

- For a given integrand f, estimate the norm $\|f\|_{s,\boldsymbol{\gamma}}$.
- Find weights $\boldsymbol{\gamma}$ which *minimize* the error bound (1).
- Using the optimized weights $\boldsymbol{\gamma}$ as input, use the CBC algorithm to find a generating vector which *satisfies* the error bound (1).

Remarks:

- If n is prime, then $\frac{1}{\varphi(n)} = \frac{1}{n-1}$. If $n = 2^k$, then $\frac{1}{\varphi(n)} = \frac{2}{n}$. For general (composite) $n \ge 3$, $\frac{1}{\varphi(n)} \le \frac{\mathrm{e}^{\gamma} \log\log n + \frac{3}{\log\log n}}{n}$, where $\gamma = 0.57721566\ldots$ (Euler–Mascheroni constant).
- The optimal convergence rate close to $\mathcal{O}(n^{-1})$ is obtained with $\lambda \to 1/2$, but $\lambda = 1/2$ is not permitted since $\zeta(2\lambda) \overset{\lambda \to 1/2+}{\to} \infty$.

Ongoing lecture course about QMC for UQ!

http://www.mi.fu-berlin.de/math/groups/naspde/teaching/UQQMC.html

Uncertainty Quantification and Quasi-Monte Carlo (Wintersemester 2022/23)

Lecture notes

Lecture notes will be published here after each week's lecture.

Week 1: Hilbert spaces, Hilbert projection theorem, orthogonal decomposition
Week 2: Dual space, Riesz representation theorem, adjoint operator, Lax–Milgram lemma
Week 3: Fourier transform, Sobolev spaces
Week 4: Lipschitz domain, Trace theorem, elliptic PDE
Week 5: Galerkin method, Céa's lemma, finite element programming (files: pde_ex.m, FEMdata.m, UpdateStiffness.m)
Week 6: Brief overview of probability theory, Karhunen–Loève expansion, elliptic PDEs with random coefficients (files: lognormal_demo.m)
Week 7: Quasi-Monte Carlo methods (finally!), reproducing kernel Hilbert space (RKHS), worst-case error (files: lognormal_demo2.m / note about the implementation)
Week 8: Randomly shifted rank-1 lattice rules, shift-averaged worst-case error, component-by-component (CBC) construction
Week 9: CBC error bound

Part II: Application to Bayesian optimal experimental design

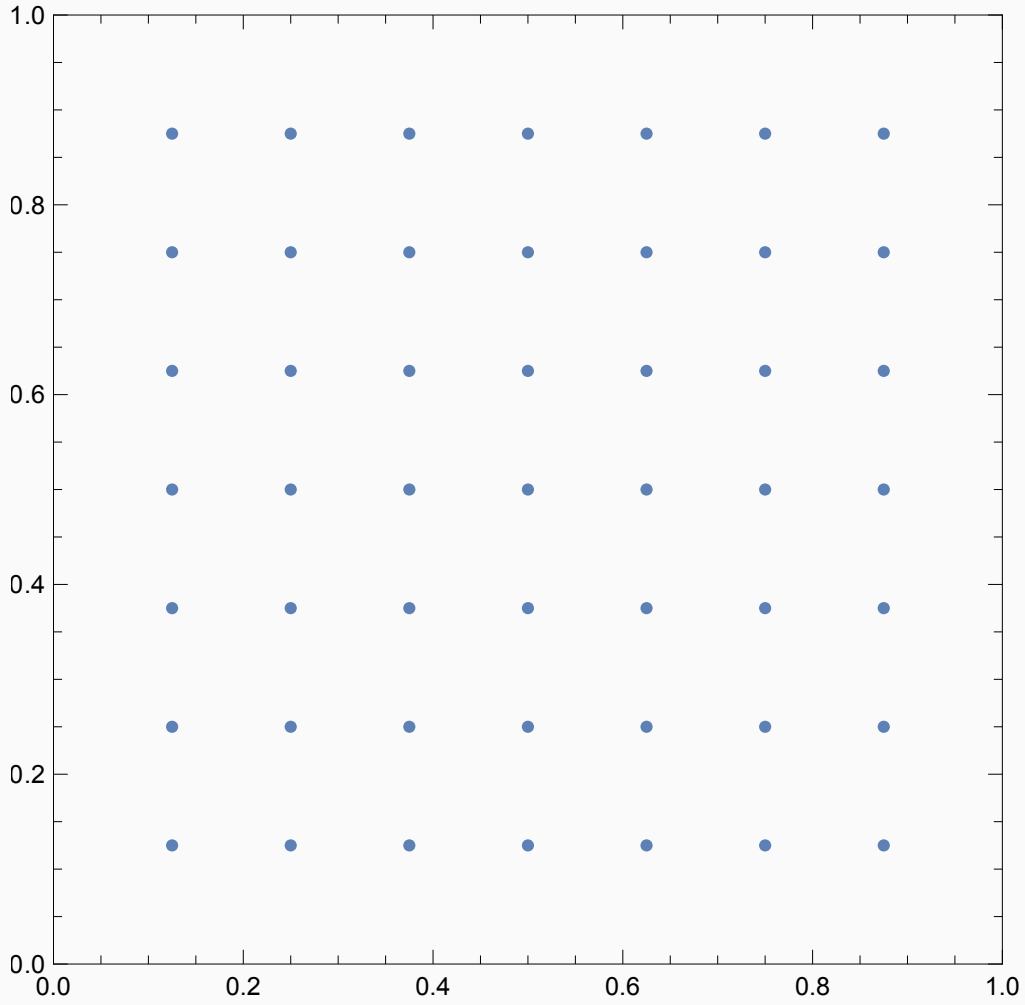
Let $G : \Theta \times \Xi \to \mathbb{R}^k$ be a forward mapping depending on a true parameter $\boldsymbol{\theta} \in \Theta$ and a design parameter $\boldsymbol{\xi} \in \Xi$.

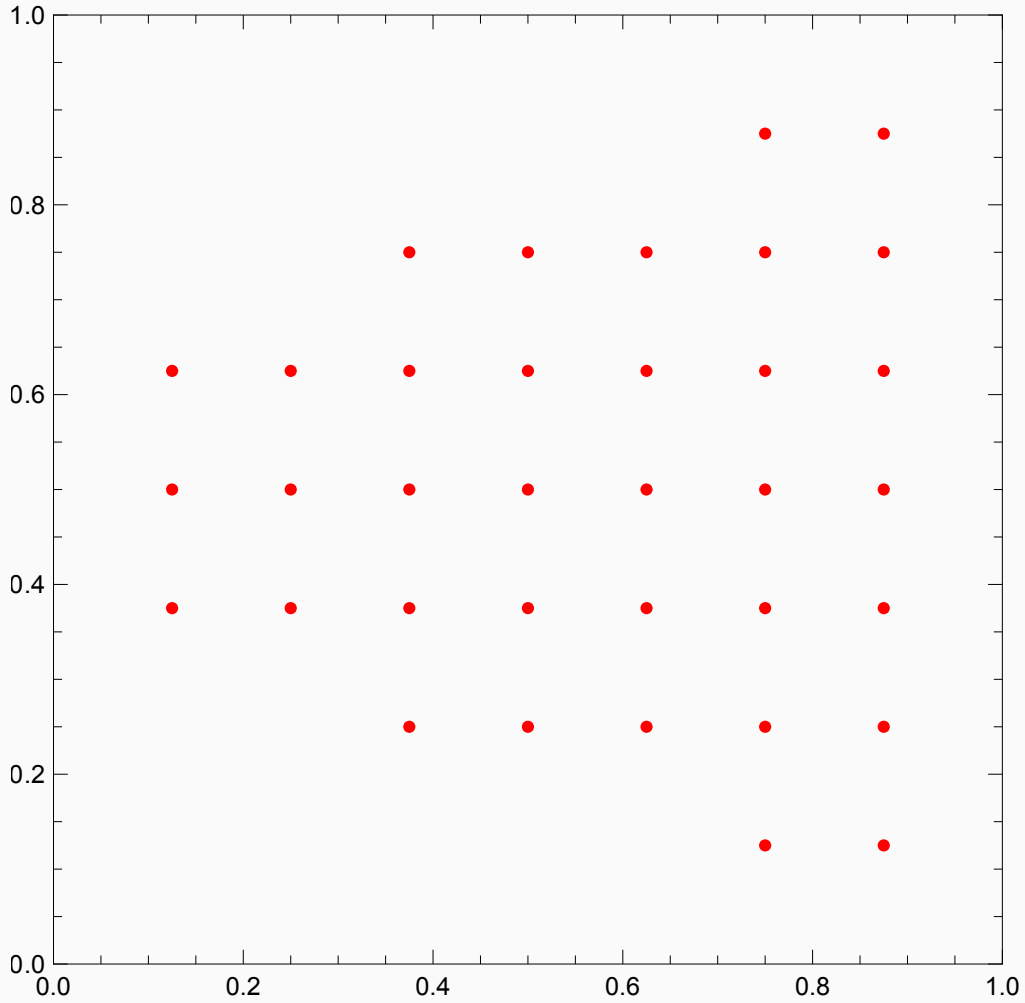
Measurement model:

$$\boldsymbol{y} = G(\boldsymbol{\theta}, \boldsymbol{\xi}) + \boldsymbol{\eta},$$

where $\boldsymbol{y} \in \mathbb{R}^k$ is the measurement data and $\boldsymbol{\eta} \in \mathbb{R}^k$ is Gaussian noise such that $\boldsymbol{\eta} \sim \mathcal{N}(0, \Gamma)$ with positive definite covariance matrix $\Gamma \in \mathbb{R}^{k \times k}$.

Goal in Bayesian optimal experimental design: Recover the design parameter $\boldsymbol{\xi}$ for the Bayesian inference of $\boldsymbol{\theta}$, which we model as a random variable endowed with prior distribution $\pi(\boldsymbol{\theta})$.


1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0


1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

A measure of the information gain for a given design $\boldsymbol{\xi}$ and data $\boldsymbol{y}$ is given by the Kullback–Leibler divergence

$$D_{\mathrm{KL}}(\pi(\cdot|\boldsymbol{y},\boldsymbol{\xi})\|\pi(\cdot)) := \int_{\Theta} \log\left(\frac{\pi(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{\xi})}{\pi(\boldsymbol{\theta})}\right)\pi(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\theta}. \tag{2}$$

A Bayesian optimal design $\boldsymbol{\xi}^*$ maximizing the expected utility (2) over the design space Ξ with respect to the data $\boldsymbol{y}$ and model parameters $\boldsymbol{\theta}$ is given by

$$\boldsymbol{\xi}^* = \arg\max_{\boldsymbol{\xi}\in\Xi} \underbrace{\int_{Y}\int_{\Theta} \log\left(\frac{\pi(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{\xi})}{\pi(\boldsymbol{\theta})}\right)\pi(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{\xi})\pi(\boldsymbol{y}|\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\theta}\,\mathrm{d}\boldsymbol{y}}_{=:\mathrm{EIG}},$$

where $\pi(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{\xi})$ corresponds to the posterior distribution of the parameter $\boldsymbol{\theta}$ and $\pi(\boldsymbol{y}|\boldsymbol{\xi}) = \int \pi(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{\xi})\pi(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}$ is the marginal distribution of the data $\boldsymbol{y}$.

The posterior is given by Bayes' theorem

$$\pi(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{\xi}) = \frac{\pi(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{\xi})\pi(\boldsymbol{\theta})}{\pi(\boldsymbol{y}|\boldsymbol{\xi})},$$

which means that the expected utility can be written as

$$\begin{aligned}\mathrm{EIG} &= \int_Y \int_\Theta \log\left(\frac{\pi(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{\xi})}{\pi(\boldsymbol{\theta})}\right)\pi(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\theta}\,\pi(\boldsymbol{y}|\boldsymbol{\xi})\mathrm{d}\boldsymbol{y} \\ &= \int_\Theta \left[\int_Y \log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{\xi})}{\pi(\boldsymbol{y}|\boldsymbol{\xi})}\right)\pi(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{\xi})\mathrm{d}\boldsymbol{y}\right]\,\pi(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}.\end{aligned}$$

Approaches taken in the literature:

- Double-loop Monte Carlo (Beck, Mansour, Espath, Long, Tempone)
- MCLA (Beck, Mansour, Espath, Long, Tempone)
- DLMCIS (Beck, Mansour, Espath, Long, Tempone)

Why QMC might work

Assume, e.g., the following:

- The forward model $\boldsymbol{y} = G(\boldsymbol{\theta}, \boldsymbol{\xi}) + \boldsymbol{\eta}$ satisfies

$$|\partial_{\boldsymbol{\theta}}^{\boldsymbol{\nu}} G_j(\boldsymbol{\theta}, \boldsymbol{\xi})| \leq C|\boldsymbol{\nu}|!\boldsymbol{b}^{\boldsymbol{\nu}}, \quad j \in \{1, \ldots, k\},$$

  where $\boldsymbol{b} := (b_j)_{j \geq 1} \in \ell^p$ for some $p \in (0, 1)$ and $C > 0$ is independent of $\boldsymbol{\xi}$.
- $\boldsymbol{\Theta} = [-\frac{1}{2}, \frac{1}{2}]^s$ and $\pi(\boldsymbol{\theta}) = 1$ for $\boldsymbol{\theta} \in \Theta$ and 0 otherwise.
- We have the likelihood

$$\pi(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{\xi}) = C\mathrm{e}^{-\frac{1}{2}\|\boldsymbol{y} - G(\boldsymbol{\theta}, \boldsymbol{\xi})\|_{\Gamma^{-1}}^2}, \quad C = \frac{1}{(2\pi)^{k/2}\sqrt{\det \Gamma}}.$$

Under these conditions, it is easy to see that

$$\mathrm{EIG} = \int_Y \int_\Theta \log\left(\frac{\pi(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{\xi})}{\pi(\boldsymbol{y}|\boldsymbol{\xi})}\right) \pi(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{\xi}) \pi(\boldsymbol{\theta}) \,\mathrm{d}\boldsymbol{\theta} \,\mathrm{d}\boldsymbol{y}$$

$$= \log C - 1 - \int_Y \log\left(\int_\Theta C\mathrm{e}^{-\frac{1}{2}\|\boldsymbol{y} - G(\boldsymbol{\theta}, \boldsymbol{\xi})\|_{\Gamma^{-1}}^2} \,\mathrm{d}\boldsymbol{\theta}\right) \int_\Theta C\mathrm{e}^{-\frac{1}{2}\|\boldsymbol{y} - G(\boldsymbol{\theta}, \boldsymbol{\xi})\|_{\Gamma^{-1}}^2} \,\mathrm{d}\boldsymbol{\theta} \,\mathrm{d}\boldsymbol{y}.$$

Consider

$$\int_Y \log\left(\int_\Theta C\mathrm{e}^{-\frac{1}{2}\|\boldsymbol{y}-G(\boldsymbol{\theta},\boldsymbol{\xi})\|^2_{\Gamma^{-1}}}\,\mathrm{d}\boldsymbol{\theta}\right)\int_\Theta C\mathrm{e}^{-\frac{1}{2}\|\boldsymbol{y}-G(\boldsymbol{\theta},\boldsymbol{\xi})\|^2_{\Gamma^{-1}}}\,\mathrm{d}\boldsymbol{\theta}\,\mathrm{d}\boldsymbol{y} \tag{3}$$

Observations:

- Parametric regularity of the integrand

  $$\int_\Theta C\mathrm{e}^{-\frac{1}{2}\|\boldsymbol{y}-G(\boldsymbol{\theta},\boldsymbol{\xi})\|^2_{\Gamma^{-1}}}\,\mathrm{d}\boldsymbol{\theta}$$

  well-understood (as long as the parametric regularity of G can be quantified), e.g., Herrmann–Keller–Schwab (2021); slight generalization of this result by K–Schillings–Schwab (in progress).
- If we replace $\int_\Theta C\mathrm{e}^{-\frac{1}{2}\|\boldsymbol{y}-G(\boldsymbol{\theta},\boldsymbol{\xi})\|^2_{\Gamma^{-1}}}\,\mathrm{d}\boldsymbol{\theta}$ by a QMC approximation

  $$\int_\Theta C\mathrm{e}^{-\frac{1}{2}\|\boldsymbol{y}-G(\boldsymbol{\theta},\boldsymbol{\xi})\|^2_{\Gamma^{-1}}}\,\mathrm{d}\boldsymbol{\theta} \approx \frac{C}{n}\sum_{i=1}^n \mathrm{e}^{-\frac{1}{2}\|\boldsymbol{y}-G(\boldsymbol{\theta}_i,\boldsymbol{\xi})\|^2_{\Gamma^{-1}}}$$

  then (3) can be approximated by

  $$\int_Y \log\left(\frac{C}{n}\sum_{i=1}^n \mathrm{e}^{-\frac{1}{2}\|\boldsymbol{y}-G(\boldsymbol{\theta}_i,\boldsymbol{\xi})\|^2_{\Gamma^{-1}}}\right)\frac{C}{n}\sum_{i=1}^n \mathrm{e}^{-\frac{1}{2}\|\boldsymbol{y}-G(\boldsymbol{\theta}_i,\boldsymbol{\xi})\|^2_{\Gamma^{-1}}}\,\mathrm{d}\boldsymbol{y}.$$

Rewrite

$$\int_Y \log\left(\frac{C}{n}\sum_{i=1}^{n} \mathrm{e}^{-\frac{1}{2}\|\boldsymbol{y}-G(\boldsymbol{\theta}_i,\boldsymbol{\xi})\|^2_{\Gamma^{-1}}}\right)\frac{C}{n}\sum_{i=1}^{n} \mathrm{e}^{-\frac{1}{2}\|\boldsymbol{y}-G(\boldsymbol{\theta}_i,\boldsymbol{\xi})\|^2_{\Gamma^{-1}}}\,\mathrm{d}\boldsymbol{y}$$

$$= \frac{C}{n}\sum_{i=1}^{n}\int_Y \log\left(\frac{C}{n}\sum_{i=1}^{n} \mathrm{e}^{-\frac{1}{2}\|\boldsymbol{y}-G(\boldsymbol{\theta}_i,\boldsymbol{\xi})\|^2_{\Gamma^{-1}}}\right)\underbrace{\mathrm{e}^{-\frac{1}{2}\|\boldsymbol{y}-G(\boldsymbol{\theta}_i,\boldsymbol{\xi})\|^2_{\Gamma^{-1}}}\,\mathrm{d}\boldsymbol{y}}_{=\text{Gaussian weight}}.$$

QMC for integrals subject to Gaussian probability densities well-understood. Essentially, we only need to worry about the parametric regularity of

$$\log\left(\frac{C}{n}\sum_{i=1}^{n} \mathrm{e}^{-\frac{1}{2}\|\boldsymbol{y}-G(\boldsymbol{\theta}_i,\boldsymbol{\xi})\|^2_{\Gamma^{-1}}}\right)$$

as long as we carry out the QMC analysis in an appropriate weighted Sobolev space.

Lemma
Under the previous conditions, we have the a priori *bound for the higher-order derivatives*

$$\partial_{\boldsymbol{y}}^{\boldsymbol{\nu}} \log\left(C\frac{1}{n}\sum_{j=1}^{n} \mathrm{e}^{-\frac{1}{2}\|\boldsymbol{y}-G(\boldsymbol{\theta}_j,\boldsymbol{\xi})\|_{\Gamma^{-1}}^2} \right)$$

$$\leq 1.1^{k|\boldsymbol{\nu}|} k^{|\boldsymbol{\nu}|} \mu_{\min}^{-|\boldsymbol{\nu}|/2} \sum_{\lambda=1}^{|\boldsymbol{\nu}|} (\lambda-1)! \left(\sum_{j=1}^{n} \mathrm{e}^{-\frac{1}{2}\|\boldsymbol{y}-G(\boldsymbol{\theta}_j,\boldsymbol{\xi})\|_{\Gamma^{-1}}^2} \right)^{-\lambda} S(|\boldsymbol{\nu}|,\lambda)$$

for all $\boldsymbol{y} \in \mathbb{R}^k =: Y$ *and* $\boldsymbol{\nu} \in \mathbb{N}_0^k$. *Here* $\mu_{\min}$ *is a lower bound for the largest eigenvalue of* Γ *and* $S(n,k)$ *denotes the* Stirling number of the second kind.

$\Rightarrow$ At least when Γ is diagonal, one expects to obtain linear convergence by applying a QMC rule to the transformed (outer) integral(s)

$$\int_{\mathbb{R}^k} \log\left(\frac{C}{n}\sum_{i=1}^{n} \mathrm{e}^{-\frac{1}{2}\|\boldsymbol{y}-G(\boldsymbol{\theta}_i,\boldsymbol{\xi})\|_{\Gamma^{-1}}^2} \right) \mathrm{e}^{-\frac{1}{2}\|\boldsymbol{y}-G(\boldsymbol{\theta}_i,\boldsymbol{\xi})\|_{\Gamma^{-1}}^2} \,\mathrm{d}\boldsymbol{y}$$

$$= \int_{(0,1)^k} \log\left(\frac{C}{n}\sum_{i=1}^{n} \mathrm{e}^{-\frac{1}{2}\|\Phi^{-1}(\boldsymbol{t})-G(\boldsymbol{\theta}_i,\boldsymbol{\xi})\|_{\Gamma^{-1}}^2} \right) \mathrm{d}\boldsymbol{t}.$$

Numerical example

Consider the parametric PDE

$$\begin{cases} -\nabla \cdot (a(\boldsymbol{x}, \boldsymbol{\theta}) \nabla u(\boldsymbol{x}, \boldsymbol{\theta})) = x_1, & \boldsymbol{x} \in (0,1)^2, \ \boldsymbol{\theta} \in \Theta, \\ u((x_1, 0), \boldsymbol{\theta}) = 0, \ u((x_1, 1), \boldsymbol{\theta}) = 1, & 0 \leq x_1 \leq 1, \ \boldsymbol{\theta} \in \Theta \\ \partial_{\mathrm{n}} u((0, x_2), \boldsymbol{\theta}) = \partial_{\mathrm{n}} u((1, x_2), \boldsymbol{\theta}) = 0, & 0 \leq x_2 \leq 1, \ \boldsymbol{\theta} \in \Theta \end{cases}$$

equipped with an uncertain diffusion coefficient

$$a(\boldsymbol{x}, \boldsymbol{\theta}) = 1 + \sum_{j=1}^{s} j^{-2} \theta_j \sin(2\pi x_1) \sin(2\pi x_2).$$

Experiment 1: QMC convergence of the inner integral

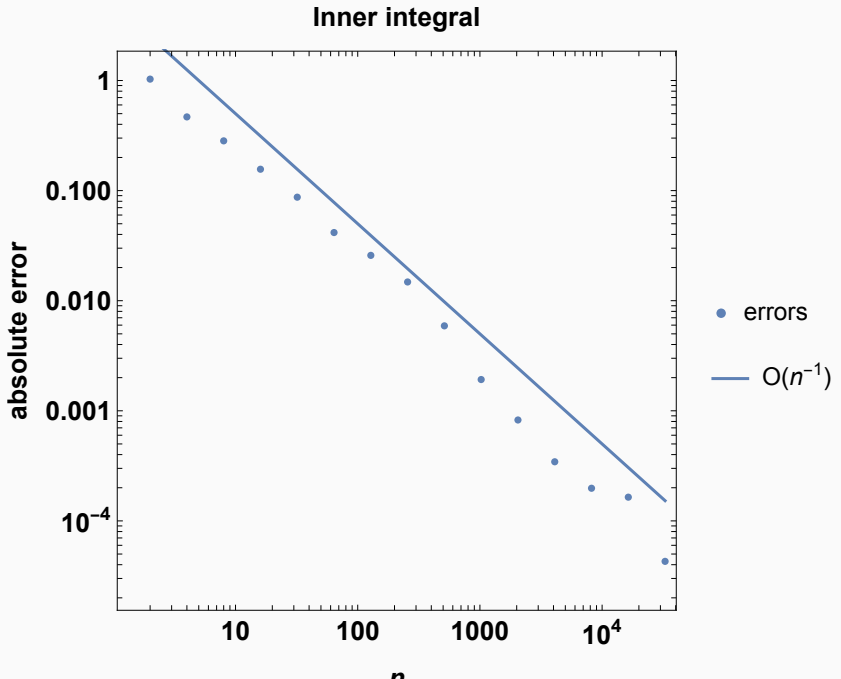
First we consider the QMC convergence of the "inner integral"

$$\int_{\Theta} C \mathrm{e}^{-\frac{1}{2}\|\boldsymbol{y}-G(\boldsymbol{\theta},\boldsymbol{\xi})\|^2_{\Gamma^{-1}}} \, \mathrm{d}\boldsymbol{\theta}.$$

Set $\boldsymbol{\xi} := (\frac{7}{8}, \frac{1}{2})$, $\Gamma := 0.1^2 I$, and with $k = 1$ consider

$$G(\boldsymbol{\xi}, \boldsymbol{\theta}) = u(\boldsymbol{\xi}, \boldsymbol{\theta}).$$

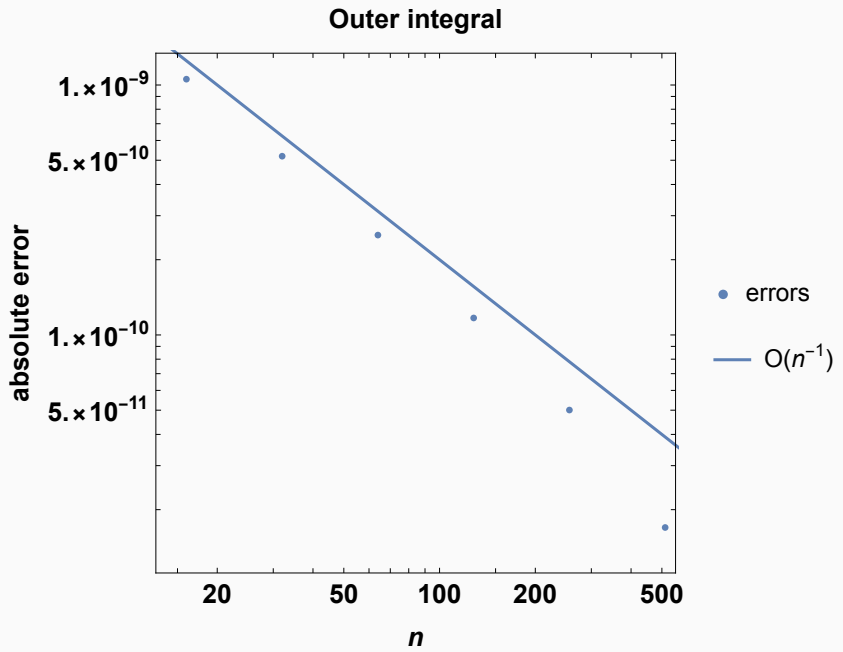
Finally, we use $\boldsymbol{y} = y = 0.396586$, and $s = 100$. We use an off-the-shelf cubature rule to compute QMC approximations of this integral.



Experiment 2: QMC convergence of the outer integral

$$\int_Y \log\left(\frac{C}{n}\sum_{i=1}^{n} \mathrm{e}^{-\frac{1}{2}\|\boldsymbol{y}-G(\boldsymbol{\theta}_i,\boldsymbol{\xi})\|^2_{\Gamma^{-1}}}\right)\frac{C}{n}\sum_{i=1}^{n} \mathrm{e}^{-\frac{1}{2}\|\boldsymbol{y}-G(\boldsymbol{\theta}_i,\boldsymbol{\xi})\|^2_{\Gamma^{-1}}}\,\mathrm{d}\boldsymbol{y}.$$

Fix QMC discretization of the inner integral ($n = 32$). Also, take $Y = [-1/2, 1/2]^5$ and fix design $\boldsymbol{\xi}$ s.t. the observation operator corresponds to $k = 5$ point evaluations within the computational domain.



- Use Laplace approximation (or "Laplace-preconditioned QMC" à la Schillings, Sprungk, Wacker 2020) to relax the dependence on the dimension k?
- In addition to discrete optimal design problems, continuous design problems? Consider the input current as a (continuous) design variable for EIT?