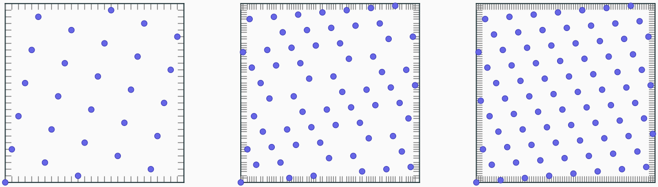
Lattice point sets and applications (part II)

Dirk Nuyens — NUMA, KU Leuven, Belgium



Workshop and Summer School on Applied Analysis 2023
TU Chemnitz
Chemnitz, Germany
September 2023

The plan for today

The plan for today

- Weighted function spaces and norms.
- Results for numerical integration.
- Function approximation using truncated Fourier series.
- Maybe: Integration on $\mathbb{R}^d$.
- Again some Julia code to demonstrate things...

Small recap

Lattice rule = equal weight quadrature using lattice points

For $f \in \mathcal{H}_{d,\alpha,\boldsymbol{\gamma}}$ approximate the d-dimensional integral

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

by an n-point lattice rule with generating vector $\boldsymbol{z} \in \mathbb{Z}_n^d$

$$Q_{n,\boldsymbol{z}}(f) := \frac{1}{n} \sum_{k \in \mathbb{Z}_n} f\left(\frac{\boldsymbol{z} k \bmod n}{n}\right).$$

Worst-case error for $f \in \mathcal{H}_{d,\alpha,\boldsymbol{\gamma}}$ for a given algorithm Q_n (e.g. $Q_{n,\boldsymbol{z}}$):

$$e(Q_n, \mathcal{H}_{d,\alpha,\boldsymbol{\gamma}}) := \sup_{\substack{f \in \mathcal{H}_{d,\alpha,\boldsymbol{\gamma}} \\ \|f\|_{d,\alpha,\boldsymbol{\gamma}} \le 1}} |I(f) - Q_n(f)|.$$

$\rightsquigarrow$ For good lattice rule $Q_{n,\boldsymbol{z}}$ converges like $n^{-\alpha} \, \|f\|_{d,\alpha,\boldsymbol{\gamma}}$.
Optimal. Bakhvalov. Matching upper and lower bounds (mod logs).

Function space

Korobov space* of dominating mixed smoothness $\alpha > 0$ ($\alpha > 1/2$):

$$\mathcal{H}_{d,\alpha,\boldsymbol{\gamma}} := \left\{ f \in L_2([0,1]^d) : \|f\|^2_{d,\alpha,\boldsymbol{\gamma}} < \infty \right\},$$

with

$$\|f\|^2_{d,\alpha,\boldsymbol{\gamma}} := \sum_{\boldsymbol{h} \in \mathbb{Z}^d} r^2_{d,\alpha,\boldsymbol{\gamma}}(\boldsymbol{h}) \, |\hat{f}(\boldsymbol{h})|^2$$

and

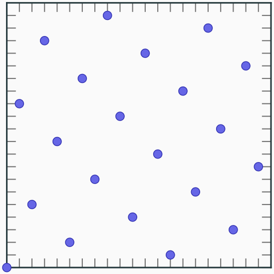
$$r_{d,\alpha,\boldsymbol{\gamma}}(\boldsymbol{h}) := \gamma^{-1}_{\mathrm{supp}(\boldsymbol{h})} \prod_{j \in \mathrm{supp}(\boldsymbol{h})} |h_j|^{\alpha}.$$

Weighted spaces: Sloan & Woźniakowski (2001), Novak & Woźniakowski (2008, 2010, 2012), ...

*Korobov used ℓ_∞ norm.

Example of a good lattice rule

Eg: $n = 21$ and $\boldsymbol{z} = (1, 13)$: Fibonacci rule: $n = F_k$, $\boldsymbol{z} = (1, F_{k-1})$.



Only $d = 2$, $d \geq 2$: Constructive methods for deterministic error:
Fast component-by-component (Nuyens & Cools 2006, . . .)
$\rightarrow$ Fixed vector $\boldsymbol{z}$ for a given n.
(Or sequence of $n = p^m$, Cools, Kuo & Nuyens 2006).

Julia – Simple lattice rule example

Given n and $\boldsymbol{z} \in \mathbb{Z}_n^d$:

$$\boldsymbol{x}_k := \frac{k\boldsymbol{z} \bmod n}{n}, \qquad Q_{n,\boldsymbol{z}}(f) := \frac{1}{n} \sum_{k \in \mathbb{Z}_n} f(\boldsymbol{x}_k).$$

```
lattice_points(z, n) = ( ( (k * z) .% n ) ./ n for k in 0:n-1 )

f = x -> prod(1 .+ (x .- 1/2))

using Statistics: mean

mean(f, lattice_points([1, 8], 13))
```

Julia – Lattice sequence in base 2 (as a plain rule sequence)

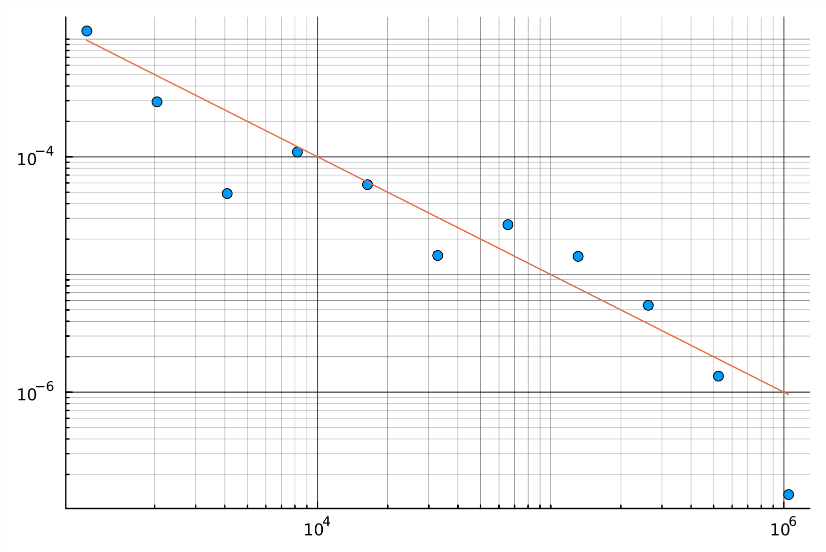
```
# exew_base2_m20_a3_HKKN.txt from Magic Point Shop:
z = [1, 364981, 245389, 97823, 488939, 62609, 400749, 385317,
     21281, 223487]       # 10 dimension with max 2^20 points

d = 2; m1 = 10; m2 = 20;
seq = ( lattice_points(z[1:d], 2^m) for m in m1:m2 )

# Such nice vectorisation...
Es = abs.(mean.(f, seq) .- 1)       # true integral is 1

using Plots
ns = 2 .^ (m1:m2)
scatter(ns, Es, xscale=:log10, yscale=:log10)
plot!(ns, ns .^ -1, xscale=:log10, yscale=:log10)
```

Absolute error versus n for $d = 2 \rightarrow$ order 1 convergence



Open problem

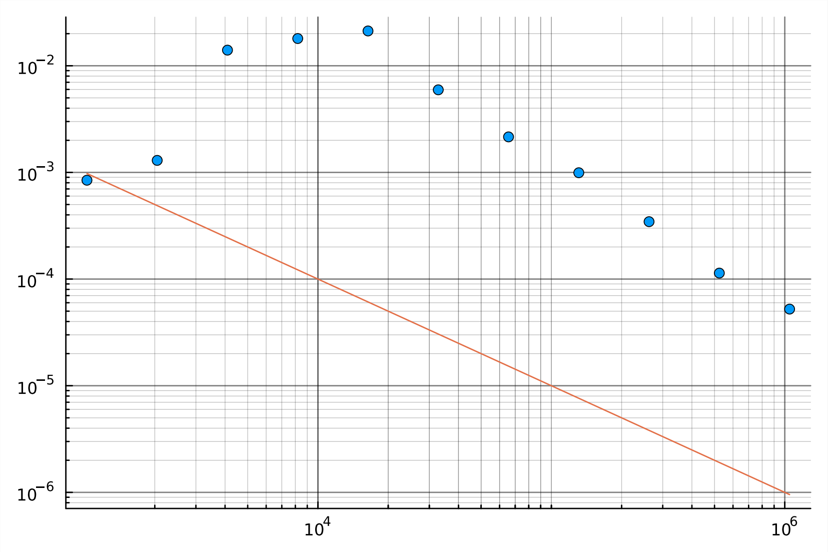
The sequence in the previous plot is using a base-2 radical inverse function (van der Corput), e.g.

$$(1011)_2 \mapsto (0.1101)_2.$$

But yesterday I also showed the Korobov sequence trick...

- The Korobov sequence trick:
  Given a good generating vector $\boldsymbol{z}^* = (z_0, z_1, \ldots, z_d) \in \mathbb{Z}_n^{d+1}$ with $z_0 = 1$, use $\boldsymbol{z} = (z_1, \ldots, z_d) \in \mathbb{Z}_n^d$ as a sequence, i.e., point by point, and get error n^{-1}.
  Can we show $n^{-\alpha}$, when $n = p^m$, $m_1 \leq m \leq m_2$, for a lattice sequence using this same trick?

Absolute error versus n for $d = 10 \to$ order 1 after bump



What do we see?

- The curse of dimensionality...
- Why does this happen?
- When does this happen?

Weighted function spaces

How to measure deterministic algorithms? (Intro to IBC)

- Worst-case error for approximating $I(f)$ by $Q_n(f)$ for $f \in \mathcal{F}_d$:

$$e(Q_n, \mathcal{H}_{d,\alpha,\boldsymbol{\gamma}}) := \sup_{\substack{f \in \mathcal{H}_{d,\alpha,\boldsymbol{\gamma}} \\ \|f\|_{d,\alpha,\boldsymbol{\gamma}} \leq 1}} |I(f) - Q_n(f)| \leq \text{ upper bound for } Q_n.$$

- Best possible error using n function values (benchmark):

$$e(n, \mathcal{H}_{d,\alpha,\boldsymbol{\gamma}}) := \inf_{Q_n : \{(w_k, \boldsymbol{x}_k)\}_{k=1}^n} e(Q_n; \mathcal{H}_{d,\alpha,\boldsymbol{\gamma}}) \geq \text{ lower bound for any}$$

$$= \textit{error of best algorithm using n function evaluations.}$$

- Information complexity: the minimal number of function values needed to reach error at most ϵ:

$$n(\epsilon, \mathcal{H}_{d,\alpha,\boldsymbol{\gamma}}) := \min\{n : \exists Q_n \text{ for which } e(Q_n, \mathcal{H}_{d,\alpha,\boldsymbol{\gamma}}) \leq \epsilon\}$$

$$= \textit{number of function evaluations of best algorithm.}$$

See a multitude of references, e.g., Novak (2016) or the Novak–Woźniakowski trilogy (2008,2010,2012), . . .

The curse of dimensionality & types of tractability

Tractability started by Woźniakowski (1994) and since then vastly expanded...

- The curse of dimensionality is defined as needing an exponential number of function values in d to reach an error $\epsilon \leq \epsilon_0$:

$$n(\epsilon, \mathcal{H}_{d,\alpha,\boldsymbol{\gamma}}) \geq c\,(1+\gamma)^d, \qquad \text{for some } c, \gamma, \epsilon_0 > 0.$$

- A problem is called (weakly) tractable if

$$\lim_{\epsilon^{-1}+d\to\infty} \frac{\ln n(\epsilon, d)}{\epsilon^{-1} + d} = 0,$$

  and intractable otherwise.
- Different types, e.g., polynomial tractability

$$n(\epsilon, \mathcal{H}_{d,\alpha,\boldsymbol{\gamma}}) \leq c\,\epsilon^{-p}\,d^q, \qquad \text{for some } c, p, q \geq 0.$$

See a multitude of references, in particular the Novak–Woźniakowski trilogy (2008,2010,2012), ...

The curse might always be there...

Define $\mathcal{F}_d$ with $f \in \mathcal{F}_d$ when

$$\|f\|_{\mathcal{F}_d} := \max_{\boldsymbol{x},\boldsymbol{y}\in[0,1]^d} \frac{|f(\boldsymbol{x}) - f(\boldsymbol{y})|}{\|\boldsymbol{x} - \boldsymbol{y}\|_\infty} \quad < \quad \infty,$$

then (Maung Zho Newn and Sharygin, 1971)

$$e(n, \mathcal{F}_d) = \frac{d}{2d+2}\, n^{-1/d}.$$

This is for any (linear) algorithm!

See also Novak (2016).

The aim is to not just avoid the "curse by construction" (product rule $n = m^d$), but also

- rate independent of $d \Rightarrow$ "mixed dominating smoothness".
- constant $C_{d,\alpha,\gamma}$ independent of $d \Rightarrow$ "weighted spaces".

Tools / assumptions

- **Mixed dominating smoothness spaces:**
  Move from typical Sobolev norm with $\|D^{\boldsymbol{\tau}} f\|_{L_2}$ bounded for $\tau_1 + \cdots + \tau_d \le \alpha$, which gives $O(n^{-\alpha/d})$ to $\tau_1, \ldots, \tau_d \le \alpha$ which gives $\sim O(n^{-\alpha})$. I.e., define $\|f\|_{d,\alpha}^2$ by

$$\sum_{\substack{\boldsymbol{\tau} \in \{0,\ldots,\alpha\}^d \\ \|\boldsymbol{\tau}\|_\infty \le \alpha}} \|D^{\boldsymbol{\tau}} f\|_{L_2}^2 \quad \text{versus} \quad \sum_{\substack{\boldsymbol{\tau} \in \{0,\ldots,\alpha\}^d \\ \|\boldsymbol{\tau}\|_1 \le \alpha}} \|D^{\boldsymbol{\tau}} f\|_{L_2}^2.$$

- **Dimension-independent error bounds:**
  Switch to weighted spaces: not all combinations of variables are as important. Denote the importance of the variables in $\mathfrak{u} \subseteq \{1, \ldots, d\}$ by $\gamma_{\mathfrak{u}}$. I.e., define $\|f\|_{d,\alpha,\boldsymbol{\gamma}}^2$ by

$$\sum_{\substack{\boldsymbol{\tau} \in \{0,\ldots,\alpha\}^d \\ \|\boldsymbol{\tau}\|_\infty \le \alpha}} \gamma_{\mathfrak{u}}^{-1} \|D^{\boldsymbol{\tau}} f\|_{L_2}^2.$$

Mixed spaces: Novak, Sickel, Temlyakov, Kühn, Ullrich, Ullrich, Potts, . . .
Weights: Hickernell (1998), Sloan & Woźniakowski (1998), Novak–Woźniakowski. . .

Again our favourite function space

Korobov space of dominating mixed smoothness $\alpha > 1/2$:

$$\mathcal{H}_{\boldsymbol{d},\alpha,\boldsymbol{\gamma}} := \left\{ f \in L_2([0,1]^d) : \|f\|^2_{d,\alpha,\boldsymbol{\gamma}} < \infty \right\},$$

with

$$\|f\|^2_{d,\alpha,\boldsymbol{\gamma}} := \sum_{\boldsymbol{h} \in \mathbb{Z}^s} r^2_{d,\alpha,\boldsymbol{\gamma}}(\boldsymbol{h}) \, |\hat{f}(\boldsymbol{h})|^2$$

and

$$r^2_{d,\alpha,\boldsymbol{\gamma}}(\boldsymbol{h}) := \gamma^{-1}_{\mathrm{supp}(\boldsymbol{h})} \prod_{j \in \mathrm{supp}(\boldsymbol{h})} |h_j|^{2\alpha}.$$

For integer smoothness

When $\alpha \in \mathbb{N}$ then this norm can be written as the norm of a more usual unanchored periodic Sobolev space of dominating mixed smoothness α:

$$\|f\|_{d,\alpha,\boldsymbol{\gamma}}^2 := \sum_{\boldsymbol{h} \in \mathbb{Z}^d} r_{d,\alpha,\boldsymbol{\gamma}}^2(\boldsymbol{h}) \, |\hat{f}(\boldsymbol{h})|^2 = \sum_{\boldsymbol{h} \in \mathbb{Z}^d} \gamma_{\mathrm{supp}(\boldsymbol{h})}^{-1} \, |\hat{f}(\boldsymbol{h})|^2 \prod_{j \in \mathrm{supp}(\boldsymbol{h})} |h_j|^{2\alpha}$$

$$= \sum_{\substack{\boldsymbol{\nu} \in \{0,\alpha\}^d \\ \mathfrak{u} := \mathrm{supp}(\boldsymbol{\nu})}} \frac{\gamma_{\mathfrak{u}}^{-1}}{\prod_{j \in \mathfrak{u}} (2\pi)^{2\nu_j}} \int_{[0,1]^{|\mathfrak{u}|} } \Bigg| \underbrace{\int_{[0,1]^{d-|\mathfrak{u}|}} f^{(\boldsymbol{\nu})}(\boldsymbol{y}_{-\mathfrak{u}}, \boldsymbol{y}_{\mathfrak{u}}) \, \mathrm{d}\boldsymbol{y}_{-\mathfrak{u}}}_{\text{``unanchored''}} \Bigg|^2 \mathrm{d}\boldsymbol{y}_{\mathfrak{u}}$$

$$= \sum_{\substack{\boldsymbol{\nu} \in \{0,\alpha\}^d \\ \mathfrak{u} := \mathrm{supp}(\boldsymbol{\nu})}} \gamma_{\mathfrak{u}}^{-1} \, \|P_{\mathfrak{u}} \, f^{(\boldsymbol{\nu})}\|_{L_2}^2 .$$

Usual error bounds

Example theorem.

For $f \in \mathcal{H}_{d,\alpha,\gamma}$ with $\alpha > 1/2$ and $n \in \mathbb{N}$ we can construct a generating vector $z \in \mathbb{Z}_n^d$ such that

$$|I(f) - Q_{n,z}(f)| \leq \frac{C_{d,\alpha,\gamma,\lambda}}{n^\lambda} \|f\|_{d,\alpha,\gamma} \qquad \text{for all } \lambda \in [1/2, \alpha)$$

with

$$C_{d,\alpha,\gamma,\lambda} = \dots$$

With the right summability conditions on the weights this becomes a dimension-independent convergence bound for some $C'_{\alpha,\gamma,\lambda}$ with $C_{d,\alpha,\gamma,\lambda} < C'_{\alpha,\gamma,\lambda} < \infty$.

See a lot of CBC and fast CBC papers: Kuo, Sloan, Dick, N., Kritzer, Ebert, Wilkes, Schwab, . . .

Function approximation

Function approximation in the worst-case setting

- Consider the embedding of $f \in \mathcal{H}_{d,\alpha,\boldsymbol{\gamma}}$ into L_2:

$$\mathrm{APP}_d : \mathcal{H}_{d,\alpha,\boldsymbol{\gamma}} \to L_2([0,1]^d)$$

where $\mathrm{APP}_d\, f = f$ for all $f \in H_{d,\alpha,\boldsymbol{\gamma}}$
and $\mathcal{H}_{d,\alpha,\boldsymbol{\gamma}}$ continuously embedded in L_2.

- Approximate APP_d by a deterministic linear algorithm $A_{d,n}$ which uses n function values (i.e., standard information Λ^{std}):

$$A_{d,n}(f; \{\boldsymbol{t}_k, a_k\}_{k=1}^n)(\boldsymbol{x}) = \sum_{k=1}^{n} f(\boldsymbol{t}_k)\, a_k(\boldsymbol{x})$$

where the $\{\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n\}$ are deterministic points (to be chosen), and the a_k are a set of functions (to be chosen).

- Use the worst-case error as quality measurement:

$$e^{\mathrm{APP}}(A_{d,n}, \mathcal{H}_{d,\alpha,\boldsymbol{\gamma}}, L_2) := \sup_{\substack{f \in \mathcal{H}_{d,\alpha,\boldsymbol{\gamma}} \\ \|f\|_{d,\alpha,\boldsymbol{\gamma}} \leq 1}} \|f - A_{d,n}(f)\|_{L_2}.$$

Best L_2 approximation

Consider the compact operator $W_d = \mathrm{APP}_d^* \mathrm{APP}_d : H_d \to H_d$ with eigenpairs $(\lambda_{d,j}, \eta_{d,j})$, ordered by $\lambda_{d,1} \geq \lambda_{d,2} \geq \cdots$. The best L_2 approximation for Λ^{all} e.g., Novak & Woźniakwoski (2010)

$$A_{d,n}^*(f)(\boldsymbol{x}) := \sum_{j=1}^{n} \langle f, \eta_{d,j} \rangle_{d,\alpha,\boldsymbol{\gamma}} \, \eta_{d,j}(\boldsymbol{x}),$$

with

$$e_{d,n}^{\mathrm{APP}}(A_{d,n}^*) = \sqrt{\lambda_{d,n+1}}.$$

Our space $\mathcal{H}_{d,\alpha,\boldsymbol{\gamma}}$ is a reproducing kernel Hilbert space with kernel

$$K_{d,\alpha,\boldsymbol{\gamma}}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{\boldsymbol{h} \in \mathbb{Z}^d} \frac{\exp(2\pi \mathrm{i}\, \boldsymbol{h} \cdot \boldsymbol{x})}{r_{d,\alpha,\boldsymbol{\gamma}}(\boldsymbol{h})} \frac{\overline{\exp(2\pi \mathrm{i}\, \boldsymbol{h} \cdot \boldsymbol{y})}}{r_{d,\alpha,\boldsymbol{\gamma}}(\boldsymbol{h})}.$$

Hence

$$\eta_{d,j}(\boldsymbol{x}) = \frac{\exp(2\pi \mathrm{i}\, \boldsymbol{h}_{d,j} \cdot \boldsymbol{x})}{r_{d,\alpha,\boldsymbol{\gamma}}(\boldsymbol{h}_{d,j})}, \qquad \lambda_{d,j} = r_{d,\alpha,\boldsymbol{\gamma}}^{-2}(\boldsymbol{h}_{d,j}) = \|\eta_{d,j}\|_{L_2}^2.$$

Approximate the best L_2 approximation

General idea:

- Enumerate Fourier indices in order of importance: for $M \geq 0$:

$$\mathcal{A}_d(M) := \{\boldsymbol{h} \in \mathbb{Z}^d : r_{d,\alpha,\gamma}(\boldsymbol{h}) \leq M\}.$$

- Approximate $\hat{f}_{\boldsymbol{h}}$ by $\hat{f}^a_{\boldsymbol{h}}$ for all $\boldsymbol{h} \in \mathcal{A}_d(M)$ using cubature.
- Approximate f by

$$A_{d,M}(f)(\boldsymbol{x}) = \sum_{\boldsymbol{h} \in \mathcal{A}_d(M)} \hat{f}^a_{\boldsymbol{h}} \, \mathrm{e}^{2\pi \mathrm{i} \, \boldsymbol{h} \cdot \boldsymbol{x}}.$$

- With error

$$(f - A_{d,M}(f))(\boldsymbol{x}) = \sum_{\boldsymbol{h} \notin \mathcal{A}_d(M)} \hat{f}_{\boldsymbol{h}} \, \mathrm{e}^{2\pi \mathrm{i} \, \boldsymbol{h} \cdot \boldsymbol{x}} + \sum_{\boldsymbol{h} \in \mathcal{A}_d(M)} (\hat{f}_{\boldsymbol{h}} - \hat{f}^a_{\boldsymbol{h}}) \, \mathrm{e}^{2\pi \mathrm{i} \, \boldsymbol{h} \cdot \boldsymbol{x}}.$$

A lot of refs, e.g., Li & Hickernell (2003), Kuo, Sloan & Woźniakowski (2006 & 2008), Byrenheid, Kämmerer, Ullrich & Volkmer (2017), . . .

L_2 error of lattice algorithm

$$\|f - A_{d,n}(f;\boldsymbol{z})\|_{L_2}^2 = \sum_{\boldsymbol{h}\notin\mathcal{A}_d(M)} |\hat{f}_{\boldsymbol{h}}|^2 + \sum_{\boldsymbol{h}\in\mathcal{A}_d(M)} |\hat{f}_{\boldsymbol{h}} - \hat{f}_{\boldsymbol{h}}^a|^2$$

$$\leq \|f\|_{d,\alpha,\gamma}^2 \left(\frac{1}{M} + \sum_{\boldsymbol{h}\in\mathcal{A}_d(M)} \sum_{\substack{0\neq\boldsymbol{\ell}\in\mathbb{Z}^d \\ \boldsymbol{\ell}\cdot\boldsymbol{z}\equiv 0 \pmod n}} \frac{1}{r_{\alpha,\gamma}(\boldsymbol{h}+\boldsymbol{\ell})} \right)$$

$$\leq \|f\|_{d,\alpha,\gamma}^2 \left(\frac{1}{M} + \sum_{\boldsymbol{h}\in\mathbb{Z}^d} \frac{M}{r_{\alpha,\gamma}(\boldsymbol{h})} \sum_{\substack{0\neq\boldsymbol{\ell}\in\mathbb{Z}^d \\ \boldsymbol{\ell}\cdot\boldsymbol{z}\equiv 0 \pmod n}} \frac{1}{r_{\alpha,\gamma}(\boldsymbol{h}+\boldsymbol{\ell})} \right)$$

$\Rightarrow$ Three methods to find good generating vectors.

Three methods for good generating vectors for APP_d

1. Satisfy the reconstruction property $\mathcal{A}_d(M)$:

$$\hat{f}^a_{\boldsymbol{h}} = \hat{f}_{\boldsymbol{h}} \quad \forall \boldsymbol{h} \in \mathcal{A}_d(M) \quad \text{for all } f \text{ with finite support } \mathcal{A}_d(M)$$

$$\Leftrightarrow \qquad \text{all } \boldsymbol{h} \cdot \boldsymbol{z} \bmod n \text{ for } \boldsymbol{h} \in \mathcal{A}_d(M) \text{ unique.}$$

Kämmerer (2013,2014), Kämmerer, Potts, Volkmer (2015), Kuo, Migliorati, Nobile, N. (2021), ...

2. Minimize

$$E_d(\boldsymbol{z}) := \sum_{\boldsymbol{h} \in \mathcal{A}_d(M)} \sum_{\substack{0 \neq \boldsymbol{\ell} \in \mathbb{Z}^d \\ \boldsymbol{\ell} \cdot \boldsymbol{z} \equiv 0 \pmod n}} \frac{1}{r_{d,\alpha,\gamma}(\boldsymbol{h} + \boldsymbol{\ell})}.$$

Kuo, Sloan, Woźniakowski (2006,2008), Cools, Kuo, N., Suryanarayana (2016), ...

3. Minimize $\qquad \rightarrow$ no dependence on $\mathcal{A}_d(M)$

$$S_d(\boldsymbol{z}) := \sum_{\boldsymbol{h} \in \mathbb{Z}^d} \frac{1}{r_{d,\alpha,\gamma}(\boldsymbol{h})} \sum_{\substack{0 \neq \boldsymbol{\ell} \in \mathbb{Z}^d \\ \boldsymbol{\ell} \cdot \boldsymbol{z} \equiv 0 \pmod n}} \frac{1}{r_{d,\alpha,\gamma}(\boldsymbol{h} + \boldsymbol{\ell})}.$$

Cools, Kuo, N., Sloan (2020,2021); product weights: Dick, Kritzer, Kuo, Sloan (2007)
Composite n and embedded point sets: Kuo, Mo, Nuyens (2023)

Final Julia intermezzo

The final Julia intermezzo

First some things I didn't say yet:

- We have fast CBC construction algorithms to obtain good generating vectors for approximation (also sequences!).
- This only gives us half of the optimal rate.
- To improve this: Kämmerer, Potts, Bartel, Volkmer, Ullrich, ...
- Rank-1 lattice points in d dimensions gives you 1D FFT.
- Kernel interpolation completely avoids the index set!

Julia:

- Show some index sets.
- How they grow...
- Inner products on the index sets.

The end

The end

- Thanks for listening...