# Fast Gaussian Process Regression for Smooth Functions using Lattice and Digital Sequences with Matching Kernels

## Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing

Aleksei G Sorokin & Fred J Hickernell

Illinois Institute of Technology, Department of Applied Mathematics

August 19, 2024

# Motivation

$K_{\boldsymbol{\alpha}}^{\boldsymbol{\gamma}} : [0,1]^d \times [0,1]^d \to \mathbb{R}$ a symmetric positive semi-definite kernel of a RKHS with

$$K_{\boldsymbol{\alpha}}^{\boldsymbol{\gamma}}(\boldsymbol{x},\boldsymbol{y}) = \prod_{j=1}^{d} \left[1 + \gamma_j K_{\alpha_j}(x_j, y_j)\right]$$

- $\boldsymbol{\gamma}$ are (product) weights e.g. serendipitous weights [Kaarnioja et al., 2023]
- $K_{\alpha_j} : [0,1] \times [0,1] \to \mathbb{R}$ a one dimensional kernel with "smoothness" $\alpha_j$

Gaussian Process Regression (kriging, kernel interpolation) on nodes $\{\boldsymbol{x}_t\}_{t=1}^n \subset [0,1]^d$

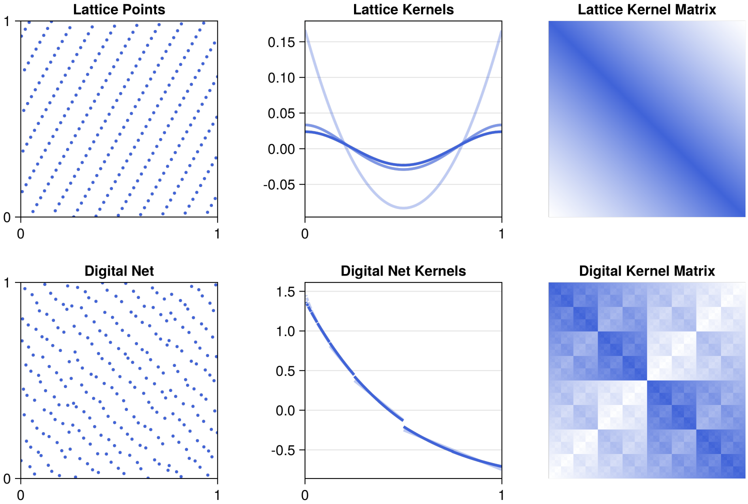
- Fitting requires solving a system in the Gram matrix $\mathsf{K} = (K_{\boldsymbol{\alpha}}^{\boldsymbol{\gamma}}(\boldsymbol{x}_t, \boldsymbol{x}_s))_{t,s=1}^n$
- Generally costs $\mathcal{O}(n^3)$ when $K_{\boldsymbol{\alpha}}^{\boldsymbol{\gamma}}$ and $\{\boldsymbol{x}_t\}_{t=1}^n$ unstructured
- Reduced cost of $\mathcal{O}(n \log n)$ when using special pairings of $K_{\boldsymbol{\alpha}}^{\boldsymbol{\gamma}}$ and $\{\boldsymbol{x}_t\}_{t=1}^n$
- GPR optimization of $\boldsymbol{\gamma}$ equivalent to optimizing RKHS for kernel interpolation

# Outline

Special pairings of $K_{\boldsymbol{\alpha}}^{\boldsymbol{\gamma}}$ and $\{\boldsymbol{x}_t\}_{t=1}^n$ enable GPR in $\mathcal{O}(n \log n)$

1. First special pairing: Lattice $\{\boldsymbol{x}_t\}_{t=1}^n$ and shift invariant $K_{\boldsymbol{\alpha}}^{\boldsymbol{\gamma}}$
   - Smooth shift-invariant $K_{\alpha_j}$ known for any $\alpha_j$
   - Connected to decay of Fourier coefficients of smooth functions
2. Second special pairing: Digital sequence $\{\boldsymbol{x}_t\}_{t=1}^n$ and digitally-shift-invariant $K_{\boldsymbol{\alpha}}^{\boldsymbol{\gamma}}$
   - NEW "Smooth" digitally-shift-invariant kernels $K_{\alpha_j}$ for $\alpha_j > 1$
   - Connected to bound on decay of Walsh coefficients of smooth functions
3. Connect special pairings to fast GPR and Quasi-Monte Carlo

# Fast Gaussian Process Regression



# Fourier Series

When $f:[0,1]\to\mathbb{R}$ has an absolutely convergent Fourier series

$$f(x)=\sum_{k\in\mathbb{Z}}\widehat{f}(k)e^{2\pi\mathrm{i}kx},\qquad \widehat{f}(k)=\int_0^1 f(x)e^{-2\pi\mathrm{i}kx}\mathrm{d}x.$$

When $f^{(\alpha)}$ has an absolutely convergent Fourier series for some $\alpha\in\mathbb{N}_0$ and $f^{(\beta)}$ periodic for all $\beta\in\{0,\dots,\alpha-1\}$

$$f^{(\alpha)}(x)=\sum_{k\in\mathbb{Z}}\widehat{f^{(\alpha)}}(k)e^{2\pi\mathrm{i}kx},\qquad \widehat{f^{(\alpha)}}(k)=(2\pi\mathrm{i}k)^{\alpha}\widehat{f}(k).$$

# Fourier Spaces

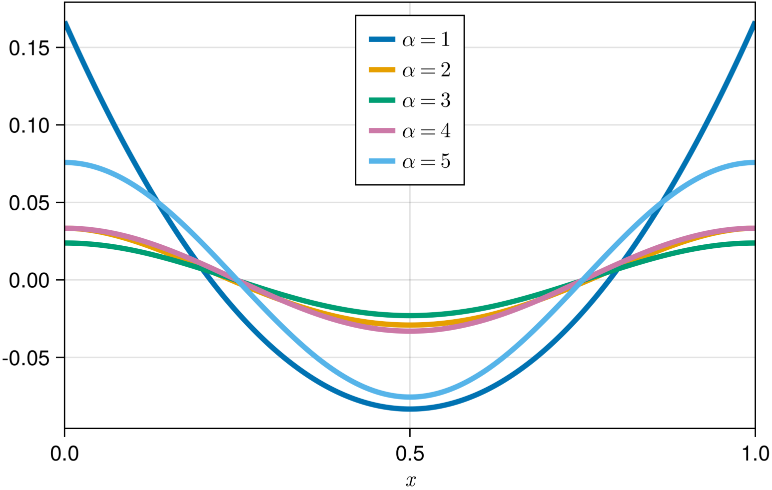
Let $\{x-y\} = (x-y) \mod 1$ and let $B_i$ denote the $i^{\text{th}}$ Bernoulli polynomial.

$$\mathring{K}_\alpha(x,y) = \sum_{k\in\mathbb{Z}\setminus\{0\}} \frac{e^{2\pi \mathrm{i} k(x-y)}}{k^{2\alpha}} = \frac{(-1)^{\alpha+1}(2\pi)^{2\alpha}}{(2\alpha)!} B_{2\alpha}(\{x-y\}) = \mathring{K}_\alpha(\{x-y\}),$$

is the kernel of Sobolev RKHS $\mathring{H}^\alpha$ with $\alpha \in \mathbb{N}$ and

$$\langle f, g\rangle^{\circ}_{\alpha} = (-1)^\alpha (2\pi)^{-2\alpha} \int_0^1 f^{(\alpha)}(x) g^{(\alpha)}(x) \mathrm{d}x.$$

$$\mathring{K}_\alpha(x) = \frac{(-1)^{\alpha+1}(2\pi)^{2\alpha}}{(2\alpha)!} B_{2\alpha}(x)$$



# Digitwise Operations

Prime base $b \geq 2$ expansion of $x \in [0,1)$ is

$$x = .\mathsf{x}_1\mathsf{x}_2\mathsf{x}_3\cdots_b = \sum_{\ell\in\mathbb{N}} \mathsf{x}_\ell b^{-\ell}, \qquad \text{e.g.} \qquad .375 = .011_2,$$

with digitwise addition (digitwise exclusive or in base $b=2$)

$$x \oplus y := \sum_{\ell\in\mathbb{N}} ((\mathsf{x}_\ell + \mathsf{y}_\ell) \mod b) b^{-\ell}, \qquad \text{e.g.} \qquad .375 \oplus .625 = .011_2 \oplus .101_2 = .110_2 = .75.$$

Similarly for $k \in \mathbb{N}_0$

$$k = \cdots\mathsf{k}_2\mathsf{k}_1\mathsf{k}_0.0_b = \sum_{\ell\in\mathbb{N}_0} \mathsf{k}_\ell b^{\ell}, \qquad \text{e.g.} \qquad 5 = 101_2,$$

$$k \oplus h := \sum_{\ell\in\mathbb{N}_0} ((\mathsf{k}_\ell + \mathsf{h}_\ell) \mod b) b^{\ell}, \qquad \text{e.g.} \qquad 5 \oplus 6 = 101_2 \oplus 110_2 = 011_2 = 3.$$

# Walsh Functions

Introduced for base $b = 2$ in [Walsh, 1923] with important results in [Fine, 1949]. Generalized to finite abelian group with a bijection in [Larcher et al., 1996].

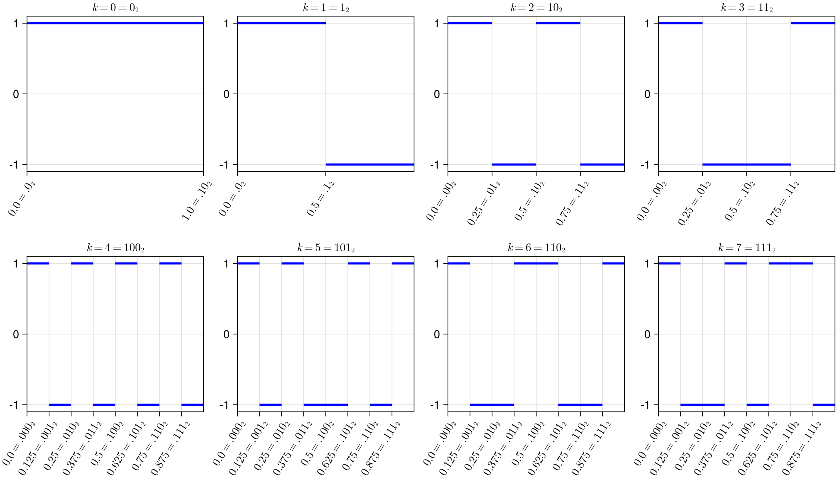
For $k \in \mathbb{N}_0$ with $\mathsf{k} = (\mathsf{k}_0, \mathsf{k}_1, \dots)$ and $x \in [0, 1)$ with $\mathsf{x} = (\mathsf{x}_1, \mathsf{x}_2, \dots)$,

$$\mathrm{wal}_k(x) = e^{2\pi \mathrm{i}/b \sum_{\ell=0}^{\infty} \mathsf{k}_\ell \mathsf{x}_{\ell+1}} = e^{2\pi \mathrm{i}/b \mathbf{k}.\mathbf{x}}$$

$$\text{e.g. for } b = 2, \qquad \mathrm{wal}_6(.75) = (-1)^{(0,1,1).(1,1,0)} = -1.$$

For any fixed $b$, $\{\mathrm{wal}_k : k \in \mathbb{N}_0\}$ is a complete orthonormal system in $\mathcal{L}_2([0, 1))$. Notice similarity to complex exponential basis $\left\{e^{2\pi \mathrm{i} k x} : k \in \mathbb{Z}\right\}$ for Fourier series.

# $b=2$ Walsh Functions $\mathrm{wal}_k(x) = (-1)^{\sum_{\ell \in \mathbb{N}_0} \mathsf{k}_\ell \mathsf{x}_{\ell+1}}$



# Walsh Function Properties

For any $x, y \in [0,1)$ and $k, h \in \mathbb{N}_0$ and $f \in \mathcal{L}_2([0,1))$

1. $\mathrm{wal}_k(x)\mathrm{wal}_h(x) = \mathrm{wal}_{k \oplus h}(x)$ and $\mathrm{wal}_k(x)\mathrm{wal}_k(y) = \mathrm{wal}_k(x \oplus y)$
2. $$\int_0^1 \mathrm{wal}_k(x)\mathrm{d}x = \begin{cases} 1, & k = 0 \\ 0, & k > 0 \end{cases}$$
3. $$\int_0^1 f(\sigma)\mathrm{d}\sigma = \int_0^1 f(x \oplus \sigma)\mathrm{d}\sigma$$
4. $$\sum_{k=0}^{b^a-1} \mathrm{wal}_k(x) = \begin{cases} b^a, & a < \beta(x) - 1 \\ 0, & \text{otherwise} \end{cases}$$

   where $\beta(x) = -\lfloor \log_b(x) \rfloor$ is the index first non-zero digit in the base $b$ expansion e.g. with $b = 2$ then $\beta(.375) = \beta(.011_2) = 2$.

# Weight Function

Write $k \in \mathbb{N}$ as

$$k = \sum_{\ell=1}^{\#k} \mathsf{k}_{a_\ell} b^{a_\ell}$$

with $a_1 > \cdots > a_{\#k} \geq 0$ and $k_{a_\ell} \in \{1,\ldots,b-1\}$.
Weight function for $\alpha \in \mathbb{N}_0$ has

$$\mu_\alpha(k) = \sum_{\ell=1}^{\min(\alpha,\#k)} (a_\ell + 1)$$

with $\mu_0(k) = \mu_\alpha(0) = 0$. $\mu$ sums indices of non-zero digits. For example, with $b = 2$

$$k = 13 = 1101_2 \qquad \text{has} \qquad (a_1, a_2, a_3) = (3, 2, 0)$$

$$\mu_1(k) = (3+1), \quad \mu_2(k) = (3+1)+(2+1), \quad \mu_3(k) = (3+1)+(2+1)+(0+1) = \mu_4(k) = \ldots.$$

# Walsh Series of Smooth Functions

For $\alpha \geq 2$ the Sobolev RKHS $H^\alpha$ with inner product

$$\langle f, g \rangle_\alpha = \sum_{\beta=1}^{\alpha-1} \int_0^1 f^{(\beta)}(x) \mathrm{d}x \int_0^1 g^{(\beta)}(x) \mathrm{d}x + \int_0^1 f^{(\alpha)}(x) g^{(\alpha)}(x) \mathrm{d}x$$

has kernel

$$K_\alpha(x,y) = \sum_{\beta=1}^{\alpha-1} \frac{B_\beta(x) B_\beta(y)}{(\beta!)^2} + \overbrace{(-1)^{\alpha+1} \frac{B_{2\alpha}(\{x-y\})}{(2\alpha)!}}^{\mathring{K}_\alpha(\{x-y\})/(2\pi)^{2\alpha}}.$$

[Dick, 2008, 2009] show that if $f \in H^\alpha$ then for $\widehat{f}(k) = \int_0^1 f(x) \overline{\mathrm{wal}_k(x)} \mathrm{d}x$ we have

$$\sup_{k \in \mathbb{N}_0} \left| \widehat{f}(k) \right| b^{\mu_\alpha(k)} < \infty \qquad \text{i.e.} \qquad \exists C_{f,\alpha} > 0 \quad \text{s.t.} \quad \left| \widehat{f}(k) \right| \leq \frac{C_{f,\alpha}}{b^{\mu_\alpha(k)}}.$$

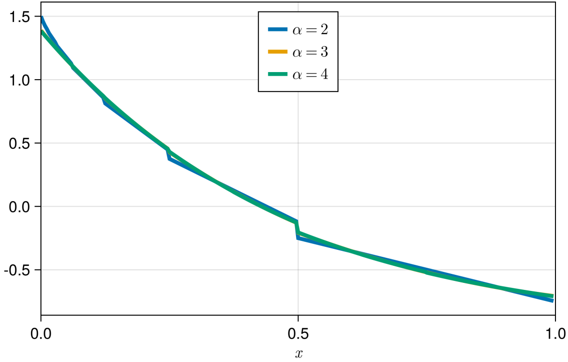
For the $\alpha = 1$ case see [Dick and Pillichshammer, 2005].

# Digitally Shift Invariant Walsh Kernel

$H^\alpha \subset \widetilde{H}^\alpha$ where $\widetilde{H}^\alpha$ is an RKHS with kernel

$$\widetilde{K}_\alpha(x,y) = \sum_{k \in \mathbb{N}} \frac{\mathrm{wal}_k(x \ominus y)}{b^{\mu_\alpha(k)}} = \widetilde{K}_\alpha(x \ominus y).$$

Below $\widetilde{K}_\alpha(x)$ with $b=2$ is shown. Discontinuities at $\{2^{-a} : a \in \mathbb{N}\}$ among others.



# Explicit forms of Low Order Kernels in $b = 2$

$$\beta(x) = -\lfloor \log_2(x) \rfloor, \qquad t_\nu(x) = 2^{-\nu\beta(x)}$$

$$\widetilde{K}_2(x) = -1 - \beta(x)x + \frac{5}{2}\left[1 - t_1(x)\right],$$

$$\widetilde{K}_3(x) = -1 + \beta(x)x^2 - 5\left[1 - t_1(x)\right]x + \frac{43}{18}\left[1 - t_2(x)\right],$$

new $$\widetilde{K}_4(x) = -1 - \frac{2}{3}\beta(x)x^3 + 5\left[1 - t_1(x)\right]x^2 - \frac{43}{9}\left[1 - t_2(x)\right]x + \frac{701}{294}\left[1 - t_3(x)\right] + \beta(x)\left[\frac{1}{48}\sum_{a_1 \geq 0} \frac{\mathrm{wal}_{2^{a_1}}(x)}{2^{3a_1}} - \frac{1}{42}\right].$$

# Research Connections for Multivariate Functions

**Kernel matrix** $\mathsf{K} = (K(\boldsymbol{x}_t, \boldsymbol{x}_s))_{t,s=1}^n$ **of pairwise evaluations at** $\{\boldsymbol{x}_t\}_{t=1}^n \subset [0,1)^d$ **is:**

- Generally dense and unstructured for $K_{\boldsymbol{\alpha}}^{\boldsymbol{\gamma}}(\boldsymbol{x},\boldsymbol{y}) := \prod_{j=1}^d \left[1 + \gamma_j K_{\alpha_j}(x_j, y_j)\right]$.
- circulant for $\mathring{K}_{\boldsymbol{\alpha}}^{\boldsymbol{\gamma}}(\boldsymbol{x},\boldsymbol{y}) := \prod_{j=1}^d \left[1 + \gamma_j \mathring{K}_{\alpha_j}(\{x_j - y_j\})\right]$, lattice $\{\boldsymbol{x}_t\}_{t=1}^n$.
- block Toeplitz for $\widetilde{K}_{\boldsymbol{\alpha}}^{\boldsymbol{\gamma}}(\boldsymbol{x},\boldsymbol{y}) := \prod_{j=1}^d \left[1 + \gamma_j \widetilde{K}_{\alpha_j}(x_j \ominus y_j)\right]$, digital net $\{\boldsymbol{x}_t\}_{t=1}^n$.

Kriging: Generally costs $\mathcal{O}(n^3)$. Costs $\mathcal{O}(n \log n)$ when K circulant or block Toeplitz.

**Quasi Monte Carlo Absolute Error**

$$\left| \int_{[0,1)^d} f(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} - \frac{1}{n} \sum_{t=1}^n f(\boldsymbol{x}_t) \right|$$

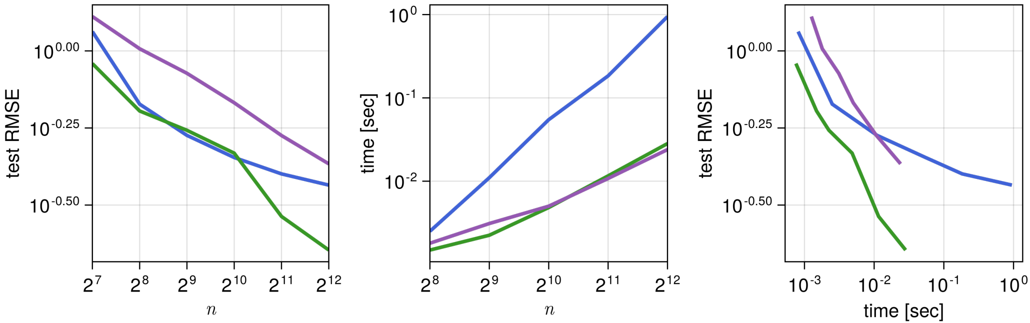
- is $\mathcal{O}(n^{-\alpha+\delta})$ for $f$ in RKHS of $\mathring{K}_{\boldsymbol{\alpha}}^{\boldsymbol{\gamma}}$ and certain lattice sequence $\{\boldsymbol{x}_t\}_{t=1}^n$.
- is $\mathcal{O}(n^{-\alpha+\delta})$ for $f$ in RKHS of $\widetilde{K}_{\boldsymbol{\alpha}}^{\boldsymbol{\gamma}}$ using certain digital sequences $\{\boldsymbol{x}_t\}_{t=1}^n$.

$\delta > 0$ arbitrary. Convergence rates independent of $d$ when $\boldsymbol{\gamma}$ chosen judiciously.

# Keister Example with $d = 3$

$$f(\boldsymbol{x}) = \pi^{d/2} \cos(\|\Phi^{-1}(\boldsymbol{x})\|/\sqrt{2}), \qquad \boldsymbol{x} \in [0,1]^d \qquad \text{Gaussian CDF } \Phi$$

IID + RBF  Lattice  Digital Net



# References I

Josef Dick. Walsh spaces containing smooth functions and quasi–monte carlo rules of arbitrary high order. *SIAM Journal on Numerical Analysis*, 46(3):1519–1553, 2008.

Josef Dick. The decay of the walsh coefficients of smooth functions. *Bulletin of the Australian Mathematical Society*, 80(3):430–453, 2009.

Josef Dick and Friedrich Pillichshammer. Multivariate integration in weighted hilbert spaces based on walsh functions and weighted sobolev spaces. *Journal of Complexity*, 21(2):149–195, 2005.

Nathan Jacob Fine. On the walsh functions. *Transactions of the American Mathematical Society*, 65(3):372–414, 1949.

Vesa Kaarnioja, Frances Y Kuo, and Ian H Sloan. Lattice-based kernel approximation and serendipitous weights for parametric pdes in very high dimensions. *arXiv preprint arXiv:2303.17755*, 2023.

# References II

Gerhard Larcher, Harald Niederreiter, and Wolfgang Ch Schmid. Digital nets and sequences constructed over finite rings and their application to quasi-monte carlo integration. *Monatshefte für Mathematik*, 121:231–253, 1996.

Joseph L Walsh. A closed set of normal orthogonal functions. *American Journal of Mathematics*, 45(1):5–24, 1923.