Fast Summation of Radial Kernels via QMC Slicing

Johannes Hertrich[1], Tim Jahn[2], Michael Quellmalz[2]
[1] Université Paris Dauphine-PSL and UCL, [2] Technische Universität Berlin

Dauphine | PSL UNIVERSITÉ PARIS — TECHNISCHE UNIVERSITÄT BERLIN

The Problem

Given: Points $x_n, y_m \in \mathbb{R}^d$ and coefficients $w_n \in \mathbb{R}$ for $n = 1, \dots, N$, $m = 1, \dots, M$ and a **radial kernel** $K\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$,

$$K(x, y) = F(\|x - y\|),$$

with **basis function** $F\colon [0, \infty) \to \mathbb{R}$.

Task: Compute the kernel sums

$$s_m = \sum_{n=1}^N w_n K(x_n, y_m) = \sum_{n=1}^N w_n F(\|x_n - y_m\|) \quad \forall\, m = 1, \dots, M. \qquad (1)$$

Applications: KDE, classification via SVMs, dimensionality reduction with kernelized PCA, computation of MMDs or energy distance, MMD gradient flows, Stein variational GD for Bayesian inference

Existing approaches

The direct/exact computation of (1) is unfeasible. $\rightsquigarrow$ Complexity $\mathcal{O}(NM)$

- **Fast Fourier summation** $\rightsquigarrow$ Complexity $\mathcal{O}(N + M + N_{\mathrm{ft}}^d \log N_{\mathrm{ft}})$
  - ☺ Efficient for $d \in \{1, 2, 3\}$
  - ☹ Becomes unfeasible for larger d
- **Random Fourier features** (RFF) $\rightsquigarrow$ Complexity $\mathcal{O}(P(N + M))$
  - ☺ Error bound $\mathbb{E}[\|F - \tilde{F}\|_\infty] \in \mathcal{O}(1/\sqrt{P})$
  - ☹ Only applicable for positive definite functions (not for Riesz $F(\|x\|) = -\|x\|$)

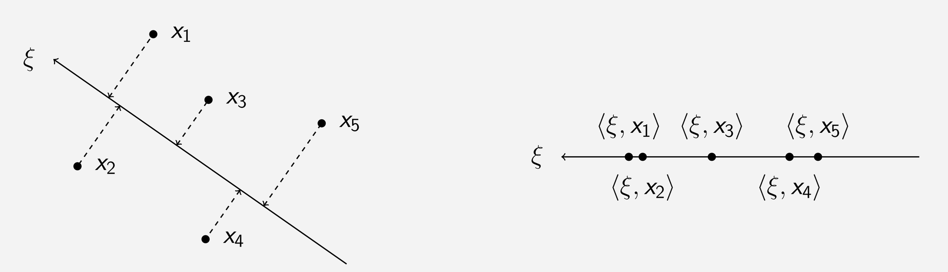
Slicing Summation

Idea: Construct $f\colon \mathbb{R} \to \mathbb{R}$ such that

$$F(\|x\|) = \mathbb{E}_{\xi \sim \mathcal{U}_{\mathbb{S}^{d-1}}}[f(\langle x, \xi \rangle)]. \qquad (2)$$

Insert (2) into the sum (1) and sample $\xi_p \sim \mathcal{U}_{\mathbb{S}^{d-1}}$ (uniformly from $\mathbb{S}^{d-1}$)

$$s_m = \sum_{n=1}^N w_n \mathbb{E}_{\xi \sim \mathcal{U}_{\mathbb{S}^{d-1}}} f(\langle x_n - y_m, \xi \rangle) \approx \frac{1}{P} \sum_{p=1}^P \underbrace{\sum_{n=1}^N w_n f(\langle x_n - y_m, \xi_p \rangle)}_{\text{1D kernel sum}}.$$



- Reduce the problem to P one-dimensional sums
- ☺ Complexity $\mathcal{O}(P(N + M + N_{\mathrm{ft}} \log N_{\mathrm{ft}}))$

Known Kernel Pairs (F, f)

Kernel	$F(x)$	$f(x)$
Gauss	$\exp(-\frac{x^2}{2\sigma^2})$	${}_1F_1(\frac{d}{2}; \frac{1}{2}; \frac{-x^2}{2\sigma^2})$
Laplace	$\exp(-\alpha x)$	${}_1F_2(\frac{d}{2}; \frac{1}{2}, \frac{1}{2}; \frac{\alpha^2 x^2}{4}) - \frac{\sqrt{\pi}\alpha x \Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})} {}_1F_2(\frac{d+1}{2}; 1, \frac{3}{2}; \frac{\alpha^2 x^2}{4})$
Matérn	$\frac{2^{1-\nu}}{\Gamma(\nu)}(\frac{\sqrt{2\nu}}{\beta}x)^\nu K_\nu(\frac{\sqrt{2\nu}}{\beta}x)$	[Hertrich 2025, Appx C]
Riesz for $r > -1$	$-x^r$	$-\frac{\sqrt{\pi}\Gamma(\frac{d+r}{2})}{\Gamma(\frac{d}{2})\Gamma(\frac{r+1}{2})} x^r$
Thin Plate Spline	$x^2 \log(x)$	$dx^2 \log(x) + \frac{d}{2}(H_{d/2} - 2 + \log(4)) x^2$

Approximation Error for Random Directions

Theorem: The mean squared error is

$$\mathbb{E}_{\xi_1, \dots, \xi_P \sim \mathcal{U}_{\mathbb{S}^{d-1}}}\Big[\Big(\frac{1}{P}\sum_{p=1}^P f(\langle x, \xi_p \rangle) - F(\|x\|)\Big)^2\Big] = \frac{\mathbb{V}_d[f](x)}{P}$$

with

Kernel	$\mathbb{V}_d[f](x)$
Any positive definite	$\mathbb{V}_d[f](x) \le F(0)^2 - F(\|x\|)^2$
Riesz $F(\|x\|) = -\|x\|^r$, $r > 0$	$\mathbb{V}_d[f](x) = \mathrm{const}(d, r)\, F(\|x\|)^2$.

- ☺ This gives the exact rate of the expected error
- ☹ Error rate $1/\sqrt{P}$ for $P \to \infty$

Quasi Monte Carlo (QMC) on the Sphere

Idea: Improve the error rate $\mathcal{O}(1/\sqrt{P})$ via replacing random points ξ_p by a quadrature rule on $\mathbb{S}^{d-1}$.

Definition [Brauchart et al. 2014]: A **QMC Design** for the Sobolev space $H^s(\mathbb{S}^{d-1})$ is a sequence $\boldsymbol{\xi}^P = (\xi_1^P, \dots, \xi_P^P) \in (\mathbb{S}^{d-1})^P$ s.t.

$$\sup_{\substack{g \in H^s(\mathbb{S}^{d-1}) \\ \|g\|_{H^s(\mathbb{S}^{d-1})} \le 1}} \left| \frac{1}{|\mathbb{S}^{d-1}|} \int_{\mathbb{S}^{d-1}} g(\xi)\, \mathrm{d}\xi - \frac{1}{P} \sum_{p=1}^P g(\xi_p^P) \right| \le \frac{c(s, d)}{P^{s/(d-1)}} \in \mathcal{O}(P^{\frac{-s}{d-1}}).$$

$\Rightarrow$ How smooth is our integrand $g_x\colon \mathbb{S}^{d-1} \to \mathbb{R}$, $g_x(\xi) = f(\langle \xi, x \rangle)$?

Theorem: We have $g_x \in H^s(\mathbb{S}^{d-1})$ for

Kernel	s
Gauss	for all $s \in \mathbb{R}_{\ge 0}$
Matérn (smoothness ν)	for $s < 2\nu + \frac{1}{2}$
Laplace	for $s < \frac{3}{2}$
neg. distance (energy / Riesz)	for $s < \frac{3}{2}$.

(Computable) QMC Designs

Spherical t-designs: Integrate polynomials of degree $\le t$ exactly

- ☺ Always exists and is a QMC design for $s = t$ whenever $s > \frac{d-1}{2}$
- ☹ Hard to compute (numerically done only for $d \le 4$)

Distance points: Maximize

$$\mathcal{E}(\boldsymbol{\xi}^P) = \sum_{p,q=1}^P \|\xi_p^P - \xi_q^P\|$$

- ☺ Is a QMC design for $s = \frac{d+1}{2}$
- ☺ Resulting worst case error rate $\mathcal{O}(1/P^{\frac{d}{2(d-1)}})$
  For $d = 3$: error rate $\mathcal{O}(1/P^{\frac{3}{4}})$

Sobol/Halton points:

- ☺ Sobol/Halton QMC sequences on $[0, 1]^d$ can be transformed to $\mathbb{S}^{d-1}$
- ☹ Unclear whether the transformed sequences are QMC designs on $\mathbb{S}^{d-1}$

Numerics

Slicing points ξ_p:

- Random directions
- Orthogonal directions
- Distance QMC points

Compared methods:

- Random Fourier features (RFF)
- Orthogonal Random Features (ORF)
- QMC(Sobol)-RFF for Gauss kernel

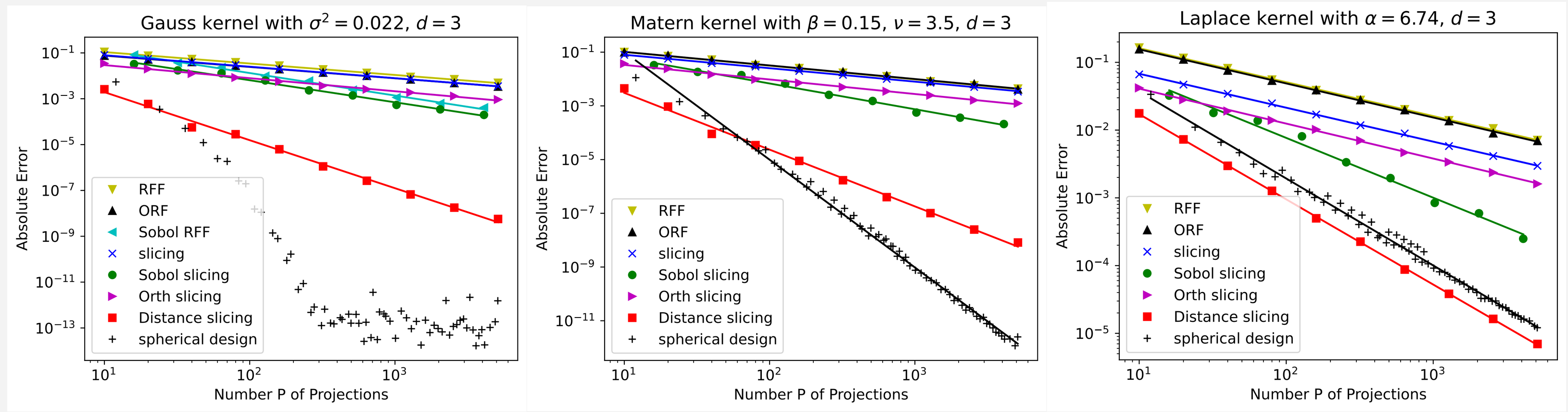


Figure: Mean approximation error $\left|F(\|x\|) - \frac{1}{P}\sum_{p=1}^P f(\langle x, \xi_p \rangle)\right|$ for dimension $d = 3$

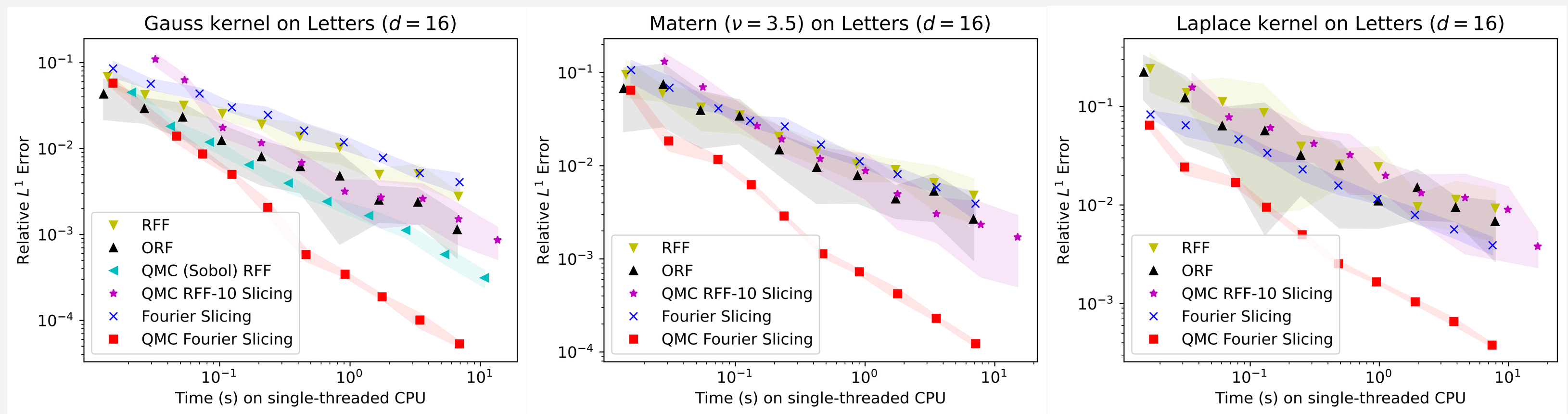


Figure: Slicing Summation on CPU (Letters dataset with $d = 16$ and $M = N = 20000$).

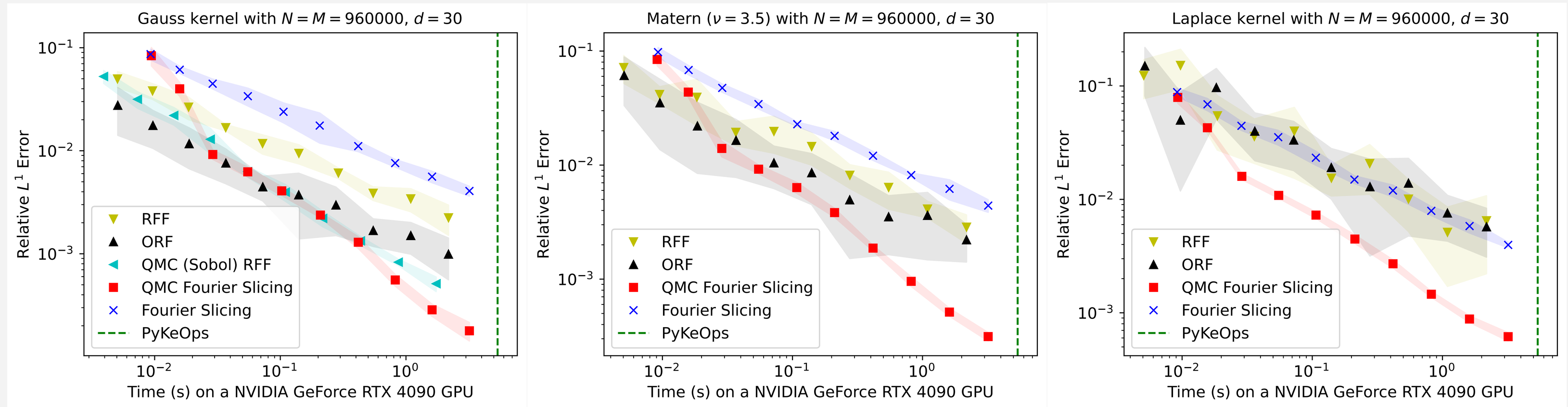
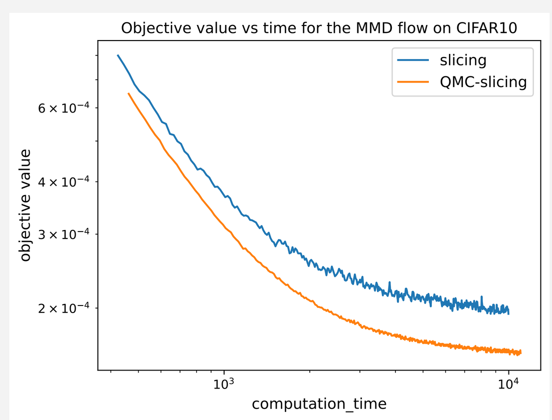


Figure: GPU times. The dashed green line is the direct computation with PyKeOps.

Application: MMD as Loss Function

- Minimize $\mathrm{Loss}(\mathbf{x}) = \mathrm{MMD}(\mathbf{x}, \mathbf{y})$ via gradient descent
- use Riesz kernel $F(x) = -\|x\|$
  $\Rightarrow$ RFFs are not applicable
- Direct computation is extremely slow
- CIFAR10 dataset ($d = 3072$)



Paper: **Code:** **Library:**