Generalized Kernel Thinning

Raaz Dwivedi
raaz@mit.edu
MIT HARVARD UNIVERSITY

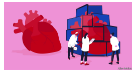
Lester Mackey
lmackey@microsoft.com
Microsoft Research

python™ **pip install goodpoints**

GitHub **microsoft/goodpoints**

arXiv.org
2105.05842 **Kernel Thinning**, COLT 2021
2110.01593 **Generalized Kernel Thinning**, ICLR 2022

Motivation: Computational cardiology



- Modeling **digital twin heart** to predict therapy response in a *non-invasive way* requires single-cell modeling. Commonly used strategy:

$$\mathbb{P}^\star f \triangleq \int f(x) d\mathbb{P}^\star(x) \approx \frac{1}{n}\sum_{i=1}^{n} f(x_i) \triangleq \mathbb{P}_n f$$

  - Estimate single cell model using Bayesian set-up: Use millions of Markov chain Monte Carlo (MCMC) points to approximate posterior $\mathbb{P}^\star$
  - Feed these samples to whole-heart simulator

x_i = MCMC sample for single cell model parameters
f = heart simulator

1 Million MCMC samples ~ 2 weeks
Single evaluation of f ~ 5 weeks

Can NOT use all million samples….!

Goal: Represent $\mathbb{P}^\star$ using a few high quality points $(x_i)_{i=1}^{n}$

- **I.I.D. sampling, and MCMC sampling** exhibit the **slow root-n Monte Carlo rate** $|\mathbb{P}^\star f - \mathbb{P}_n f| = \Theta(n^{-1/2})$, e.g., ~ 10^6 points for 0.1% error
- Computationally prohibitive for expensive f, and common fixes uniform thinning, or **standard thinning**–-choose every t–th point
- **Accuracy degrades with such thinning**–$\Theta(\sqrt{t/n})$ worst-case error–same as the slow rate with n/t points, e.g., $n^{-1/4}$ rate with $\sqrt{n}$ points

Kernel Thinning: Better than Monte Carlo rate

$x_1, x_2, \ldots, x_n \in \mathbb{R}^d$, kernel $\mathbf{k}$
$\mathbb{P}_{in} \triangleq \frac{1}{n}\sum_{i=1}^{n}\delta_{x_i}$ → Kernel Thinning (KT) → $y_1, \ldots, y_{\sqrt{n}}$, $\mathbb{P}_{KT} \triangleq \frac{1}{\sqrt{n}}\sum_{i=1}^{\sqrt{n}}\delta_{y_i}$

Non-uniform sub-sample of size $\sqrt{n}$

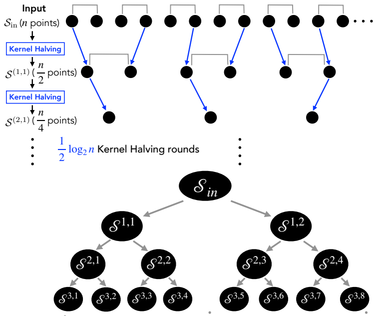
- With high probability, the worst-case error—MMD error—in the reproducing kernel Hilbert space (RKHS) satisfies

$$\sup_{\|g\|_{\mathbf{k}}\leq 1} |\mathbb{P}_{in} g - \mathbb{P}_{KT} g| \lesssim_d \begin{cases} n^{-1/2}\sqrt{\log n} & \text{(Compactly supported; e.g., B-spline } \mathbf{k}) \\ n^{-1/2}\sqrt{\log^{d/2+1} n \log\log n} & \text{(Sub-Gaussian tails; e.g., Gaussian } \mathbf{k}) \\ n^{-1/2}\sqrt{\log^{d+1} n \log\log n} & \text{(Sub-exponential tails; e.g., Matern } \mathbf{k}) \end{cases}$$

For output size s, the MMD error is $\widetilde{O}(1/s)$

KT: A two-stage algorithm

- **KT-Split**: Repeated rounds of **non-uniform halving** (prefer diverse points based on kernel alignment) the parent coreset in two equal-sized children coresets
- **KT-Swap**: Point-by-point refinement of the best child coreset by swapping with the best alternative in $\mathcal{S}_{in}$ if it improves the MMD error



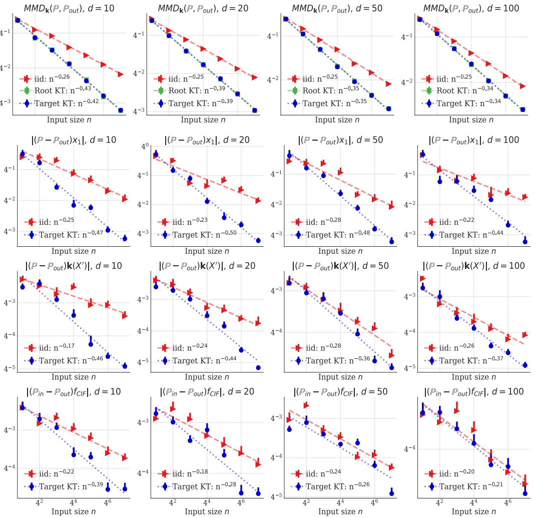
Generalized kernel thinning

	Root KT [Original Algorithm]	Target KT	KT+ [Best of both worlds]
KT-Split kernel	$\mathbf{k}_{rt}$	$\mathbf{k}$	$\mathbf{k} + \mathbf{k}_{\alpha-rt}$
Single-function error	Same as MMD error	$\sqrt{\frac{\log n}{n}}$ For arbitrary $\mathbf{k}$ on arbitrary domain	$\sqrt{\frac{\log n}{n}}$
MMD error	See left panel	$\sqrt{\frac{\log^{ad+b} n}{n}}$ Analytic $\mathbf{k}$ $\sqrt{\frac{n^{d/m}}{n}}$ m-times differentiable $\mathbf{k}$	Min(Target KT Error, α-Root KT)

$\mathbf{k}(x,y) = \int \mathbf{k}_{rt}(x,z)\mathbf{k}_{rt}(z,y)dz$ & $\mathbf{k}_{\alpha-rt} = (\widehat{\mathbf{k}})^{\alpha}$ where $\hat{}$ denotes Fourier transform

- **Significantly superior to $n^{-1/4}$ rates from Standard-$\sqrt{n}$ Thinning**
- **In fact, nearly minimax integration error in many settings**
- **Quasi Monte Carlo like guarantees, but KT guarantees apply to non-uniform targets with unbounded support**

Good performance inside and outside of RKHS



MMD error rates for popular kernels

Kernel k	Target KT	Root KT	KT+
Gauss(σ)	$\frac{(\log n)^{\frac{3d}{4}+1}}{\sqrt{n \cdot c_n^d}}$	$\frac{(\log n)^{\frac{d}{4}+\frac{1}{2}}\sqrt{c_n}}{\sqrt{n}}$	$\frac{(\log n)^{\frac{d}{4}+\frac{1}{2}}\sqrt{c_n}}{\sqrt{n}}$
Laplace(σ)	$n^{-\frac{1}{4}}$	N/A	$\left(\frac{c_n(\log n)^{1+2d(1-\alpha)}}{n}\right)^{\frac{1}{4\alpha}}$
Matérn(ν,γ) $\nu \in (\frac{d}{2}, d]$	$n^{-\frac{1}{4}}$	N/A	$\left(\frac{c_n(\log n)^{1+2d(1-\alpha)}}{n}\right)^{\frac{1}{4\alpha}}$
Matérn(ν,γ) $\nu > d$	$\min(n^{-\frac{1}{4}}, \frac{(\log n)^{\frac{d+1}{2}}}{n^{(\nu-d)/(2\nu-d)}})$	$\frac{(\log n)^{\frac{d+1}{2}}\sqrt{c_n}}{\sqrt{n}}$	$\frac{(\log n)^{\frac{d+1}{2}}\sqrt{c_n}}{\sqrt{n}}$
IMQ(ν,γ) $\nu < \frac{d}{2}$	$\frac{(\log n)^{d+1}}{\sqrt{n}}$	$\min(n^{-\frac{1}{4}}, \frac{\log n}{\sqrt{n^{1-d/(4\nu)}}})$	$\frac{(\log n)^{d+1}}{\sqrt{n}}$
IMQ(ν,γ) $\nu \geq \frac{d}{2}$	$\frac{(\log n)^{d+1}}{\sqrt{n}}$	$\frac{(\log n)^{\frac{d+1}{2}}\sqrt{c_n}}{\sqrt{n}}$	$\frac{(\log n)^{\frac{d+1}{2}}\sqrt{c_n}}{\sqrt{n}}$
sinc(θ)	$\frac{(\log n)^2}{\sqrt{n}}$	$n^{-\frac{1}{4}}$	$\frac{(\log n)^2}{\sqrt{n}}$
B-spline($2\beta+1,\gamma$) $\beta \in 2\mathbb{N}$	$\sqrt{\frac{\log n}{n^{2\beta/(2\beta+1)}}}$	N/A	$\sqrt{\frac{\log n}{n}}$
B-spline($2\beta+1,\gamma$) $\beta \in 2\mathbb{N}_0+1$	$\sqrt{\frac{\log n}{n^{2\beta/(2\beta+1)}}}$	$\sqrt{\frac{\log n}{n}}$	$\sqrt{\frac{\log n}{n}}$