Proceedings of Machine Learning Research vol 134:1–1, 2021 34th Annual Conference on Learning Theory

Kernel Thinning

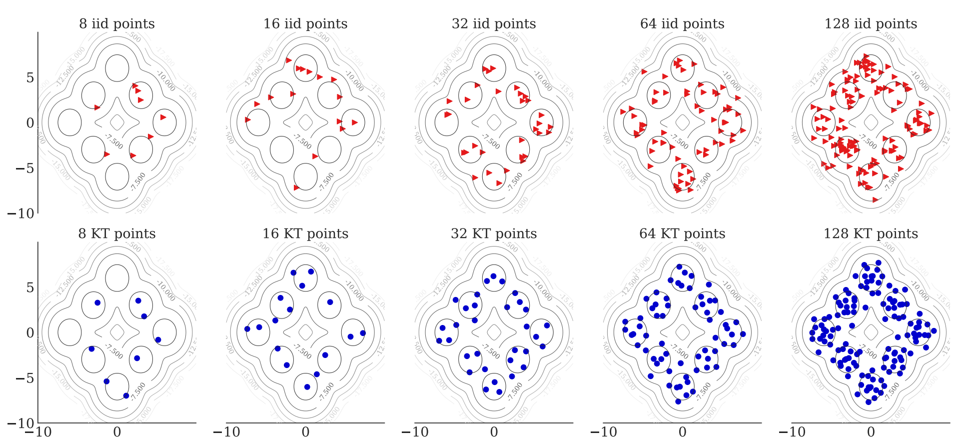
Raaz Dwivedi RAAZ.RSK@BERKELEY.EDU
EECS, UC Berkeley
Lester Mackey LMACKEY@MICROSOFT.COM
Microsoft Research New England

Editors: Mikhail Belkin and Samory Kpotufe

Abstract

[1]We introduce kernel thinning, a new procedure for compressing a distribution $\mathbb{P}$ more effectively than i.i.d. sampling or standard thinning. Given a suitable reproducing kernel $\mathbf{k}$ and $\mathcal{O}(n^2)$ time, kernel thinning compresses an n-point approximation to $\mathbb{P}$ into a $\sqrt{n}$-point approximation with comparable worst-case integration error across the associated reproducing kernel Hilbert space. With high probability, the maximum discrepancy in integration error is $\mathcal{O}_d(n^{-\frac{1}{2}}\sqrt{\log n})$ for compactly supported $\mathbb{P}$ and $\mathcal{O}_d(n^{-\frac{1}{2}}\sqrt{(\log n)^{d+1}\log\log n})$ for sub-exponential $\mathbb{P}$ on $\mathbb{R}^d$. In contrast, an equal-sized i.i.d. sample from $\mathbb{P}$ suffers $\Omega(n^{-\frac{1}{4}})$ integration error. Our sub-exponential guarantees resemble the classical quasi-Monte Carlo error rates for uniform $\mathbb{P}$ on $[0,1]^d$ but apply to general distributions on $\mathbb{R}^d$ and a wide range of common kernels. We use our results to derive explicit non-asymptotic maximum mean discrepancy bounds for Gaussian, Matérn, and B-spline kernels and present two vignettes illustrating the practical benefits of kernel thinning over i.i.d. sampling and standard Markov chain Monte Carlo thinning.

The Python package for kernel thinning can be found at `https://github.com/rzrsk/kernel_thinning`. In the figure below, we provide a simple illustration of the benefits of kernel thinning over i.i.d. sampling from an 8-mixture of Gaussian target $\mathbb{P}$. We provide a scatter plot of i.i.d. points from $\mathbb{P}$, and points output by kernel thinning using a Gaussian kernel $\mathbf{k}(x,y) = \exp(-\frac{1}{2}\|x-y\|_2^2)$ and i.i.d. points as the input. The equidensity contours of the target distribution are underlaid. Even for small sample sizes, the kernel thinning (KT) points exhibit a better spatial distribution with less clumping and fewer gaps, suggestive of a better approximation to $\mathbb{P}$.



1. Extended abstract. Full version appears as [`https://arxiv.org/abs/2105.05842`, v3]