# Fast learning in reproducing kernel Kreĭn spaces via signed measures

Fanghui Liu[1], Xiaolin Huang[2,3], Yingyi Chen[1], Johan A.K. Suykens[1]
[1]Department of Electrical Engineering (ESAT-STADIUS), KU Leuven, Belgium
[2]Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, China
[3]Institute of Medical Robotics, Shanghai Jiao Tong University, China

## Topic: Indefinite kernels

extensively used in practice:

- *intrinsic*: hyperbolic tangent kernel, log kernel, truncated $\ell_1$ distance kernel
- *extrinsic*: positive definite kernels degenerate to indefinite ones, e.g.,
  1) polynomial kernel (by $\ell_2$ normalization) on the unit sphere
  2) Gaussian kernel with geodesic distance

**Open question: positive decomposition**
Can a given non-PD kernel be decomposed into the difference of two PD kernels?

## Topic: Scalability

Scalibility due to its n-by-n indefinite kernel matrix by eigenvalue decomposition
kernel ridge regression (KRR)

- space complexity $\mathcal{O}(n^2)$
- time complexity $\mathcal{O}(n^3)$

Random features are infeasible: **indefiniteness**
Current results are **biased** to omit the negative eigenvalue part

**Key question in scalability**
How to obtain unbiased estimator by random features for indefinite kernels?

## Our Algorithm

**Algorithm 1**: Random features for various indefinite kernels via signed measures.

**Input**: An indefinite kernel function $k(\boldsymbol{x},\boldsymbol{x}')=k(z)$ with $z:=\|\boldsymbol{x}-\boldsymbol{x}'\|_2$ and the number of random features $s$.

**Output**: Unbised random feature map $\Phi(\cdot):\mathbb{R}^d\rightarrow\mathbb{R}^{4s}$ such that $k(\boldsymbol{x},\boldsymbol{x}')\approx \frac{1}{s}\sum_{i=1}^{s}\langle\varphi_i(\boldsymbol{x}),\varphi_i(\boldsymbol{x}')\rangle$.

1. Obtain the measure $\mu(\cdot)$ of the kernel $k$ via (generalized) Fourier transform
2. Given $\mu$, let $\mu:=\mu_+-\mu_-$ be the Jordan decomposition with two nonnegative measures $\mu_\pm$ and compute the total mass $\|\mu\|=\|\mu_+\|+\|\mu_-\|$
3. Sample $\{\boldsymbol{\omega}_i\}_{i=1}^s\sim\mu_+/\|\mu_+\|$ and $\{\boldsymbol{\nu}_i\}_{i=1}^s\sim\mu_-/\|\mu_-\|$
4. Output the explicit feature mapping $\Phi(\boldsymbol{x})$ with $\varphi_i(\boldsymbol{x})$ given in Eq. (1).

## Refs. & Acknowledgement

[1] **RFF:** A. Rahimi and B. Recht. Random features for large-scale kernel machines, NeurIPS2007.
[2] **SRF:** J. Pennington et al. Spherical random features for polynomial kernels, NeurIPS2015.
[3] J. Bognár. Indefinite inner product spaces. Springer, 1974.
[4] D. Oglic and T. Gärtner. Scalable learning in reproducing kernel kreĭn spaces, ICML2019.

NEW: ERC Advanced Grant E-DUALITY
Exploring duality for future data-driven modelling

## Preliminaries: RKKS, random features, signed measures

**Reproducing Kernel Kreĭn Spaces (RKKS):** There exists two Hilbert spaces $\mathcal{H}_\pm$ such that
i) $\forall f\in\mathcal{H}_\mathcal{K}$, it can be decomposed into $f=f_+\oplus f_-$, where $f_+\in\mathcal{H}_+$ and $f_-\in\mathcal{H}_-$, respectively.
ii) $\forall f,g\in\mathcal{H}_\mathcal{K}$, $\langle f,g\rangle_{\mathcal{H}_\mathcal{K}}=\langle f_+,g_+\rangle_{\mathcal{H}_+}-\langle f_-,g_-\rangle_{\mathcal{H}_-}$.
**Positive decomposition:** $k=k_+-k_-$ for $k$ asscoiated with RKKS and $k_\pm$ associated with $\mathcal{H}_\pm$.
**Random features:** A shift-invariant $k(\boldsymbol{x},\boldsymbol{x}')=k(\boldsymbol{x}-\boldsymbol{x}')$ and positive definite kernel (by nomarlizing $k(\boldsymbol{0})=1$) admit the spectral distribution representation

$$k(\boldsymbol{x},\boldsymbol{x}')=\int_{\mathbb{R}^d}p(\boldsymbol{\omega})\exp\left(\mathrm{i}\boldsymbol{\omega}^\top(\boldsymbol{x}-\boldsymbol{x}')\right)\mathrm{d}\boldsymbol{\omega}\approx\frac{1}{s}\sum_{j=1}^{s}\exp(\mathrm{i}\boldsymbol{\omega}_j^\top\boldsymbol{x})\exp(\mathrm{i}\boldsymbol{\omega}_j^\top\boldsymbol{x}')^*\ \text{with}\ \boldsymbol{\omega}\sim p(\boldsymbol{\omega})$$

**Signed Measures:** Let $\mu:\mathcal{A}\rightarrow(-\infty+\infty)$ be a signed measure on a set $\Omega$ satisfying $\mu(\emptyset)=0$ and $\sigma$-additivity. It covers Borel measure, finite measure, probability measure, etc.

## Our Model in Theory

**Theorem 1.** *Denote the (generalized) Fourier transform of a stationary indefinite kernel $k$ as the measure $\mu$, then we have the following results:*
*(i)* **Existence**: *$k$ admits the positive decomposition, i.e., $k=k_+-k_-$, if and only if the total mass of the measure $\mu$ is finite, i.e., $\|\mu\|<\infty$.*
*(ii)* **Representation**: *If $\|\mu\|<\infty$, the associated RKHSs $\mathcal{H}_\pm$ are given by*

$$\mathcal{H}_\pm=\left\{f:\|f\|^2_{\mathcal{H}_\pm}=\int_{\mathbb{R}^d}\frac{|F(\boldsymbol{\omega})|^2}{\mu_\pm(\boldsymbol{\omega})}\mathrm{d}\boldsymbol{\omega}<\infty\right\},$$

*where $F(\boldsymbol{\omega})$ is the Fourier transform of $f$.*

**Remark:** 1) much easier to be found than operator theory in harmonic analysis
2) computationally implementable in practice

## Our Model in Application: randomized fratures map

Our algorithm admits

$$k(\boldsymbol{x}-\boldsymbol{x}')\approx\frac{1}{s}\sum_{i=1}^{s}\langle\mathrm{Re}[\varphi_i(\boldsymbol{x})],\mathrm{Re}[\varphi_i(\boldsymbol{x}')]\rangle-\frac{1}{s}\sum_{i=1}^{s}\langle\mathrm{Im}[\varphi_i(\boldsymbol{x})],\mathrm{Im}[\varphi_i(\boldsymbol{x}')]\rangle\,.$$

where $\varphi_i(\boldsymbol{x})$ is

$$\varphi_i(\boldsymbol{x})=\left[\sqrt{c_1\|\mu_+\|}\cos(\boldsymbol{\omega}_i^\top\boldsymbol{x}),\sqrt{c_1\|\mu_+\|}\sin(\boldsymbol{\omega}_i^\top\boldsymbol{x}),\mathrm{i}\sqrt{c_2\|\mu_-\|}\cos(\boldsymbol{\nu}_i^\top\boldsymbol{x}),\mathrm{i}\sqrt{c_2\|\mu_-\|}\sin(\boldsymbol{\nu}_i^\top\boldsymbol{x})\right].\quad(1)$$

Hold for linear combination of PD kernels, polynomial kernels on the sphere.

**Theorem 2.** *For any $\boldsymbol{x},\boldsymbol{x}'\in\mathbb{R}^d$, by $\ell_2$ normalization, the NTK kernel of a two layer ReLU network of the form $f(\boldsymbol{x};\boldsymbol{\theta})=\sqrt{2}s\sum_{j=1}^{s}\sum_{j=1}^{s}a_j\max\{\boldsymbol{\omega}_j^\top\boldsymbol{x},0\}$ is stationary*

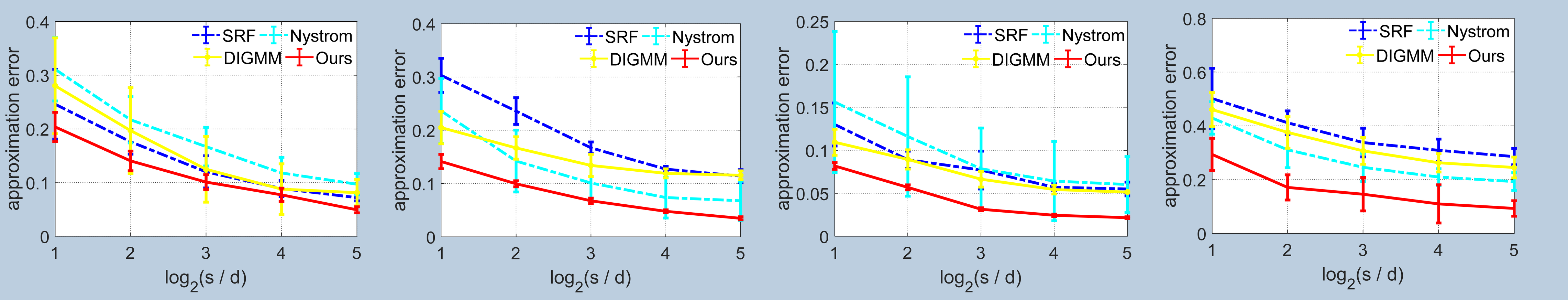
$$k(\boldsymbol{x},\boldsymbol{x}')=\frac{2-z^2}{\pi}\arccos\left(\frac{1}{2}z^2-1\right)+\frac{z}{2\pi}\sqrt{4-z^2}\,,$$

*where $z:=\|\boldsymbol{x}-\boldsymbol{x}'\|_2\in[0,2]$. However, the function $k(z),z\in[0,2]$ is not positive definite.*

## Experimental Results

Kernel approximation on *letter*, *ijcnn1*, *covtype*, and *cod-RNA*

delta-Gaussian kernel: $k(\boldsymbol{x},\boldsymbol{x}')=\exp(-\|\boldsymbol{x}-\boldsymbol{x}'\|^2/2a^2)-\exp(-\|\boldsymbol{x}-\boldsymbol{x}'\|^2/2b^2)$ with $a=1,b=10$




polynomial kernel by $\ell_2$ normalization: $k(\boldsymbol{x},\boldsymbol{x}')=(1-\|\boldsymbol{x}-\boldsymbol{x}'\|_2^2/a^2)^p$ with $a=p=2$