A generalization of the randomized singular value decomposition

Nicolas Boullé (University of Oxford) and Alex Townsend (Cornell University)

Highlights

- We provide **new theoretical bounds** for the randomized SVD when using random input vectors generated from **any multivariate Gaussian distribution**.
- We generalize the randomized SVD to **Hilbert-Schmidt operators** and provide numerical examples to learn integral kernels.
- **Covariance matrices with prior knowledge** outperform the standard identity matrix used in the literature and lead to near-optimal approximation errors.

Background: randomized SVD

Computing the **singular value decomposition** (SVD) is a fundamental linear algebra task in machine learning, statistics, and signal processing but this can be **computationally infeasible** for large matrices.

The randomized SVD uses matrix-vector products **with standard Gaussian random vectors** and is one of the most popular algorithms for constructing a low-rank approximation to a matrix A.

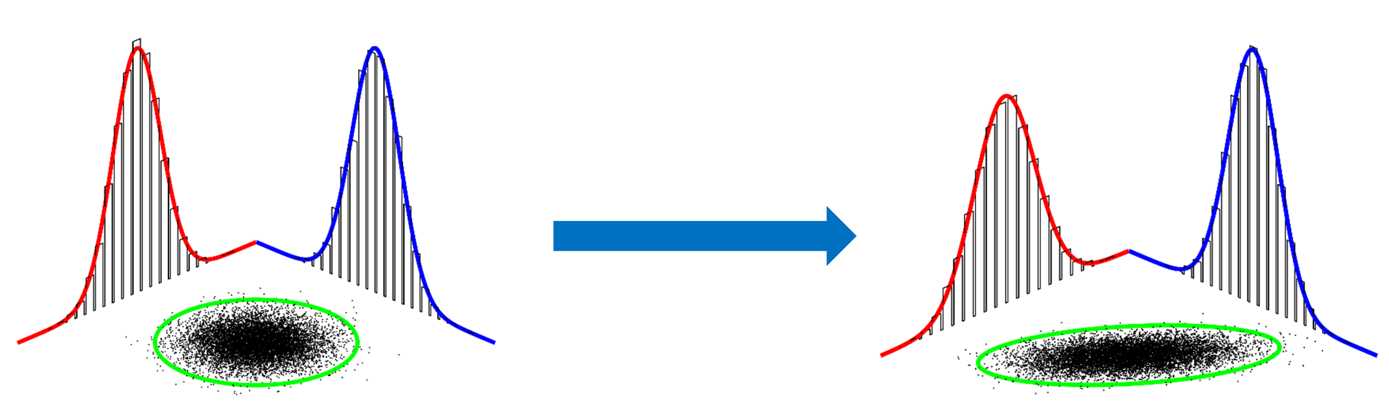
Theorem (Halko et al. (2011)). *Let $\mathbf{A}$ be an $m\times n$ matrix, $k\geq 1$ an integer, and choose an oversampling parameter $p\geq 4$. If $\boldsymbol{\Omega}\in\mathbb{R}^{n\times(k+p)}$ is a standard Gaussian random matrix and $\mathbf{QR}=\mathbf{A}\boldsymbol{\Omega}$ is the economized QR decomposition of $\mathbf{A}\boldsymbol{\Omega}$, then for all $u,t\geq 1$,*

$$\|\mathbf{A}-\mathbf{QQ}^*\mathbf{A}\|_{\mathrm{F}}\leq\left(1+t\sqrt{\frac{3k}{p+1}}\right)\sqrt{\sum_{j=k+1}^{n}\sigma_j^2(\mathbf{A})}+ut\frac{\sqrt{k+p}}{p+1}\sigma_{k+1}(\mathbf{A}),$$

with failure probability at most $2t^{-p}+e^{-u^2}$.

The squared tail of the singular values of A gives the best rank k approximation error to A in the Frobenius norm.

This result shows that the randomized SVD can compute a **near-best low-rank approximation** to A with high probability.



Standard Gaussian vectors — Correlated Gaussian vectors

We generalize the randomized SVD to allow for correlated Gaussian input vectors.

Generalized randomized SVD

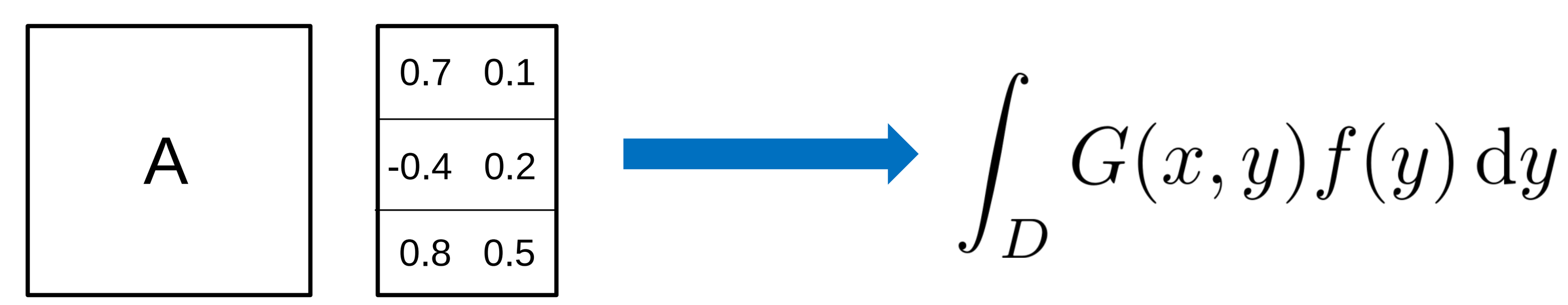
New theoretical bounds are obtained for the randomized SVD with nonstandard covariance matrices.

Theorem. *Let $\mathbf{A}$ be an $m\times n$ matrix, $k\geq 1$ an integer, and choose an oversampling parameter $p\geq 4$. If $\boldsymbol{\Omega}\in\mathbb{R}^{n\times(k+p)}$ is a Gaussian random matrix, where each column is sampled from a multivariate Gaussian distribution with covariance matrix $\mathbf{K}\in\mathbb{R}^{n\times n}$, and $\mathbf{QR}=\mathbf{A}\boldsymbol{\Omega}$ is the economized QR decomposition of $\mathbf{A}\boldsymbol{\Omega}$, then for all $u,t\geq 1$,*

$$\|\mathbf{A}-\mathbf{QQ}^*\mathbf{A}\|_{\mathrm{F}}\leq\left(1+ut\sqrt{(k+p)\frac{3k}{p+1}\frac{\beta_k}{\gamma_k}}\right)\sqrt{\sum_{j=k+1}^{n}\sigma_j^2(\mathbf{A})},$$

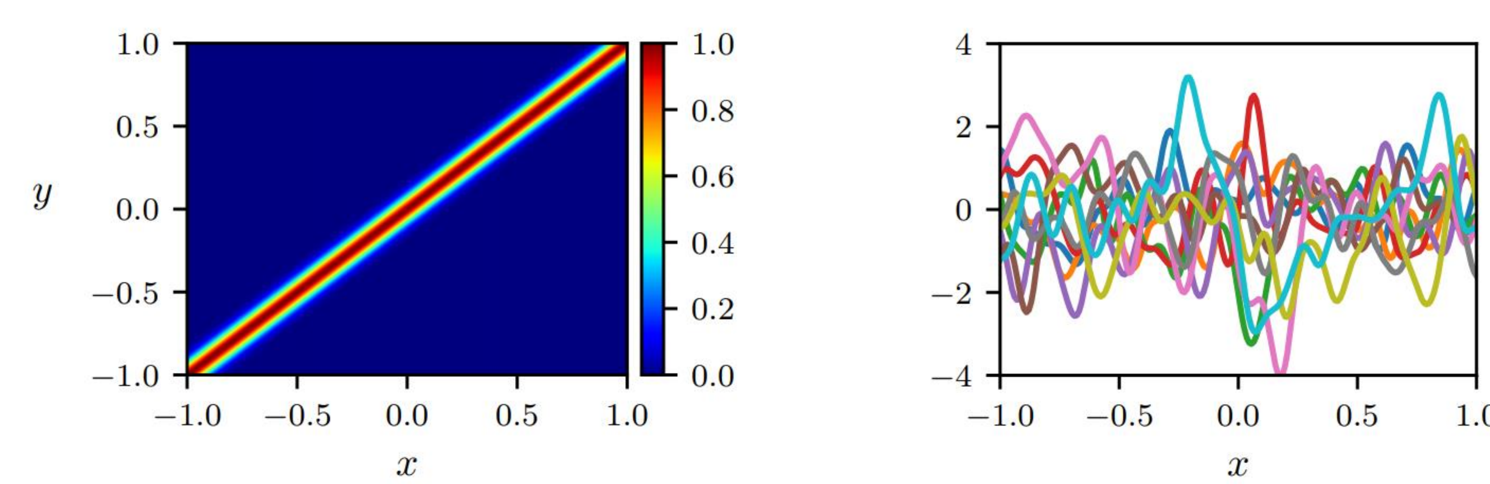
with failure probability at most $t^{-p}+[ue^{-(u^2-1)/2}]^{k+p}$. Here, γ_k and β_k denote the covariance quality factors.

The factors β_k and γ_k depend on how much prior information of the matrix A is encoded in the covariance matrix.



Matrix-vector product — Integral against a random function

The randomized SVD is generalized to learn **Hilbert-Schmidt operators** using random input functions, sampled from a Gaussian process (GP).



Covariance kernel — Random functions

The **choice of the covariance kernel** in the GP is crucial and impacts both the theoretical bounds and numerical results of the randomized SVD.

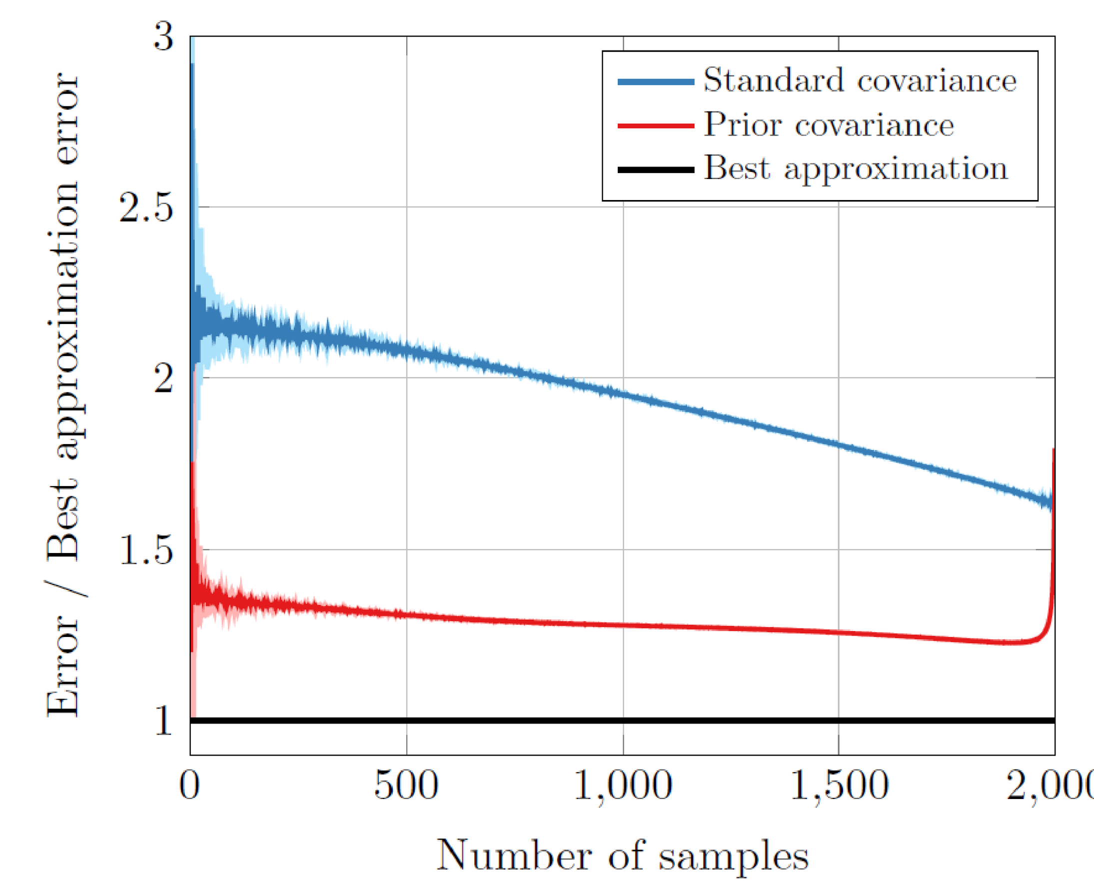
This leads us to introduce a **new covariance kernel** based on weighted Jacobi polynomials for learning kernels of operators.

The **smoothness of the functions** sampled from a GP with the Jacobi kernel can be **controlled** as it is related to the decay rate of the kernel's eigenvalues.

Paper: https://openreview.net/forum?id=hgKtwSb4S2

Incorporating prior knowledge to learn matrices

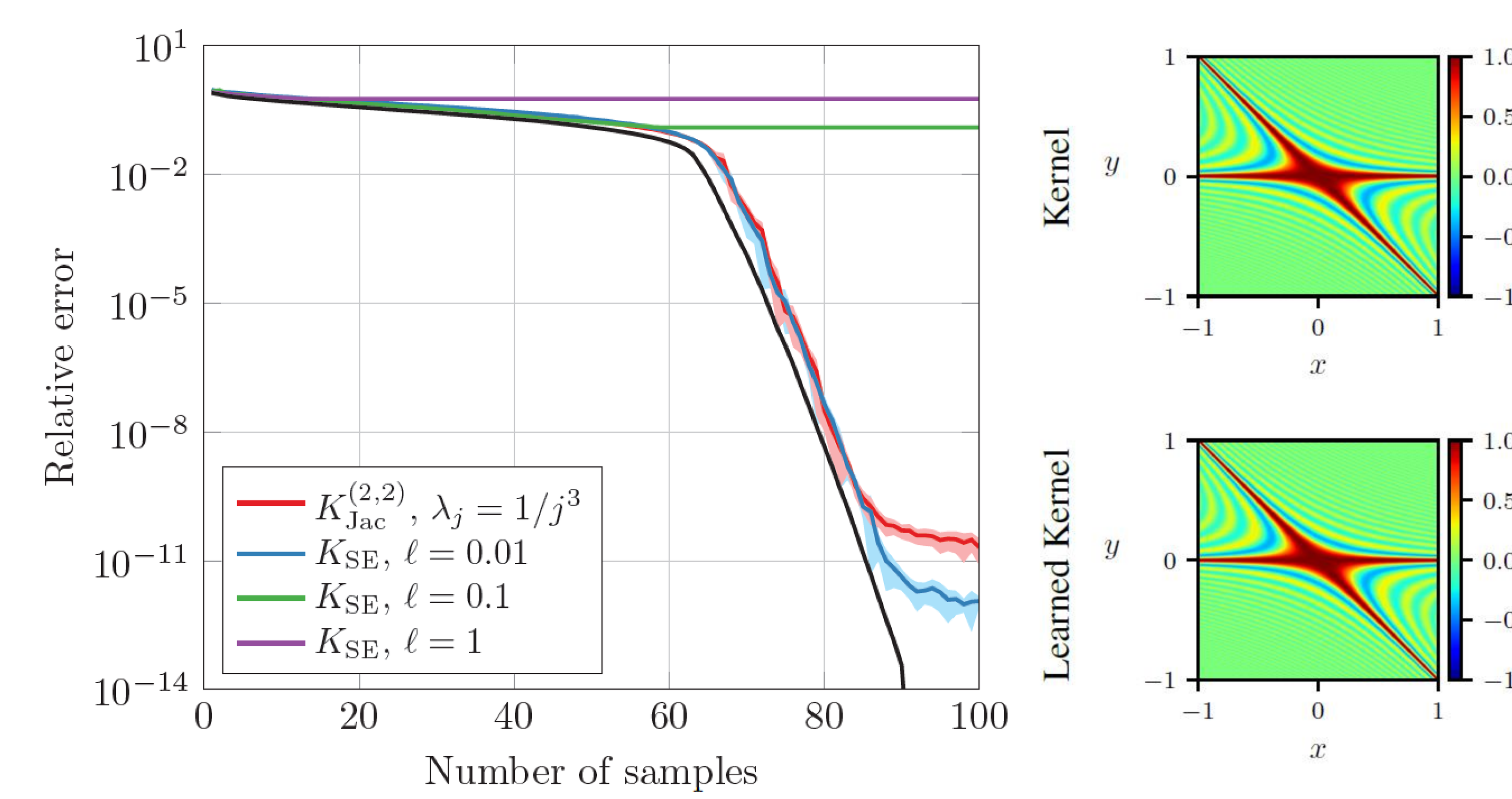
Prior knowledge of the matrix can be incorporated into the covariance matrix to **outperform the randomized SVD** with standard Gaussian inputs.



Randomized SVD error using standard and prior covariance matrices

Learning Hilbert-Schmidt operators

The randomized SVD for HS operators is applied to **learn kernels** of integral operators.



Randomized SVD error using standard and prior covariance matrices

One can obtain **near-optimal approximation error** with the randomized SVD depending on the choice of the covariance kernel used to sample the functions.