Randomized Estimation of Spectral Densities

Fabio Matti Haoze He Daniel Kressner

ANCHP, École polytechnique fédérale de Lausanne (EPFL), Lausanne, Switzerland

In many problems in physics, engineering, and computer science, the eigenvalues of certain matrices help understand the nature of a system. However, computing the eigenvalues of a matrix can be prohibitively expensive. Furthermore, it is often not crucial to know the exact individual eigenvalues, but more so their approximate locations with respect to each other. The goal of spectral density theory is to find the approximate distribution of the eigenvalues of large matrices.

Spectral density

The spectral density ϕ of symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ with eigenvalues $\lambda_1, \dots, \lambda_n \in \mathbb{R}$, which we assume to be contained in $[-1, 1]$, is defined as

$$\phi(t) = \frac{1}{n}\sum_{i=1}^{n} \delta(t - \lambda_i). \tag{1}$$

Since the Dirac delta distribution δ cannot be approximated using smooth functions, and because we do not care about the exact locations of the eigenvalues anyway, we work with the smooth spectral density ϕ_σ defined as

$$\phi_\sigma(t) = \sum_{i=1}^{n} g_\sigma(t - \lambda_i) \tag{2}$$

with a smoothing kernel g_σ, which is parametrized by a smoothing parameter $\sigma > 0$. Usually, g_σ is a Gaussian of width σ

$$g_\sigma(s) = \frac{1}{n\sqrt{2\pi\sigma^2}} e^{-\frac{s^2}{2\sigma^2}}. \tag{3}$$

We may then convert the problem of computing ϕ_σ to the trace estimation problem

$$\phi_\sigma(t) = \operatorname{Tr}(g_\sigma(t\boldsymbol{I}_n - \boldsymbol{A})). \tag{4}$$

Chebyshev expansion

It is often prohibitively expensive to directly evaluate the matrix function $g_\sigma(t\boldsymbol{I}_n - \boldsymbol{A})$. For this and other reasons, a Chebyshev expansion with degree of expansion m

$$g_\sigma^{(m)}(t\boldsymbol{I}_n - \boldsymbol{A}) = \sum_{l=0}^{m} \mu_l(t) T_l(\boldsymbol{A}) \tag{5}$$

is first computed instead. For each t, the coefficients $\{\mu_l\}_{l=0}^m$ can efficiently be computed through a discrete cosine transform (DCT). When defining the vector of coefficients $\boldsymbol{\mu} \in \mathbb{R}^{m+1}$ and the vector $\boldsymbol{g} \in \mathbb{R}^{m+1}$ with components $g_l = g_\sigma(t - \cos(\pi l/m))$, $l = 0, \dots, m$, we can switch between the two in $\mathcal{O}(m\log(m))$ complexity using

$$\boldsymbol{\mu} = \mathrm{DCT}^{-1}(\boldsymbol{g}) \iff \boldsymbol{g} = \mathrm{DCT}(\boldsymbol{\mu}). \tag{6}$$

This way of computing the Chebyshev expansion of a function can also be used to efficiently square Chebyshev polynomials, which will be crucial in our implementation

$$\Big(\sum_{l=0}^{m} \mu_l T_l(\boldsymbol{A})\Big)^2 = \sum_{l=0}^{m} \nu_l T_l(\boldsymbol{A}) \implies \boldsymbol{\nu} = \mathrm{DCT}^{-1}\left(\mathrm{DCT}(\boldsymbol{\mu})^2\right). \tag{7}$$

Finally, the starting point for all our methods will be the expanded spectral density

$$\phi_\sigma^{(m)}(t) = \operatorname{Tr}(g_\sigma^{(m)}(t\boldsymbol{I}_n - \boldsymbol{A})). \tag{8}$$

Delta-Gauss-Chebyshev

For a symmetric matrix $\boldsymbol{B} \in \mathbb{R}^{n\times n}$ and standard Gaussian random vector $\boldsymbol{\psi} \in \mathbb{R}^n$ the estimate $\boldsymbol{\psi}^\top \boldsymbol{B}\boldsymbol{\psi}$ satisfies

$$\operatorname{Tr}(\boldsymbol{B}) = \mathbb{E}\left[\boldsymbol{\psi}^\top \boldsymbol{B}\boldsymbol{\psi}\right]. \tag{9}$$

The Girard-Hutchinson trace estimator averages over $n_\Psi \in \mathbb{N}$ independent realizations of this estimate to get

$$\mathrm{H}_{n_\Psi}(\boldsymbol{B}) = \frac{1}{n_\Psi}\sum_{i=1}^{n_\Psi} \boldsymbol{\psi}_i^\top \boldsymbol{B}\boldsymbol{\psi}_i. \tag{10}$$

The Delta-Gauss-Chebyshev (DGC) method [1] applies this estimator to (8) to obtain

$$\boxed{\widetilde{\phi}_\sigma^{(m)}(t) = \mathrm{H}_{n_\Psi}(g_\sigma^{(m)}(t\boldsymbol{I}_n - \boldsymbol{A}))}. \tag{11}$$

We can derive the following result for the DGC method:

Theorem 1: Delta-Gauss-Chebyshev method

If $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is symmetric, then with high probability

$$\|\phi_\sigma - \widetilde{\phi}_\sigma^{(m)}\|_1 \le \frac{\sqrt{2}}{\sigma^2}(1+\sigma)^{-m}\left(2 + \frac{c_\Psi}{\sqrt{nn_\Psi}}\right) + \frac{c_\Psi}{\sqrt{n_\Psi}}$$

for a constant $c_\Psi \ge 0$.

Nyström-Chebyshev

The Nyström approximation offers a way to compress a symmetric positive semi-definite (PSD) matrix of low rank by multiplying it with a small sketching matrix $\boldsymbol{\Omega} \in \mathbb{R}^{n\times n_\Omega}$, where $n_\Omega \ll n$, and forming

$$\widehat{\boldsymbol{B}} = (\boldsymbol{B}\boldsymbol{\Omega})(\boldsymbol{\Omega}^\top \boldsymbol{B}\boldsymbol{\Omega})^\dagger (\boldsymbol{B}\boldsymbol{\Omega})^\top. \tag{12}$$

Since $g_\sigma(t\boldsymbol{I}_n - \boldsymbol{A})$ exhibits a low-rank structrue, particularly for small σ, the Nyström-Chebyshev (NC) method uses the Nyström approximation $\widehat{g}_\sigma^{(m)}$ of (5) to compute

$$\boxed{\widehat{\phi}_\sigma^{(m)}(t) = \operatorname{Tr}(\widehat{g}_\sigma^{(m)}(t\boldsymbol{I}_n - \boldsymbol{A}))}. \tag{13}$$

An efficient implementation can be achieved due to the cyclic property of the trace, which allows to write

$$\widehat{\phi}_\sigma^{(m)}(t) = \operatorname{Tr}(\boldsymbol{K}_1(t)^\dagger \boldsymbol{K}_2(t)) \tag{14}$$

with the two matrices

$$\boldsymbol{K}_1(t) = \boldsymbol{\Omega}^\top g_\sigma^{(m)}(t\boldsymbol{I}_n - \boldsymbol{A})\boldsymbol{\Omega} \in \mathbb{R}^{n_\Omega \times n_\Omega} \tag{15}$$

$$\boldsymbol{K}_2(t) = \boldsymbol{\Omega}^\top (g_\sigma^{(m)}(t\boldsymbol{I}_n - \boldsymbol{A}))^2 \boldsymbol{\Omega} \in \mathbb{R}^{n_\Omega \times n_\Omega} \tag{16}$$

which can be computed efficiently with (6) and (7).

Nyström-Chebyshev++

The Nyström++ estimator is a variance-reduced version of the Girard-Hutchinson estimator. It is defined as

$$\mathrm{Tr}^{++}(\boldsymbol{B}) = \operatorname{Tr}(\widehat{\boldsymbol{B}}) + \mathrm{H}_{n_\Psi}(\boldsymbol{\Delta}) \tag{17}$$

with $\boldsymbol{\Delta} = \boldsymbol{B} - \widehat{\boldsymbol{B}}$. Generalizing it to the parameter-dependent case yields the Nyström-Chebyshev++ (NC++) method, which combines NC and DGC

$$\boxed{\breve{\phi}_\sigma^{(m)}(t) = \widehat{\phi}_\sigma^{(m)}(t) + \widetilde{\phi}_\sigma^{(m)}(t) - \mathrm{H}_{n_\Psi}(\widehat{g}_\sigma^{(m)}(t\boldsymbol{I}_n - \boldsymbol{A}))}. \tag{18}$$

The lack of theory for Nyström approximations of indefinite matrices only allows results for the shifted density

$$\underline{\phi}_\sigma(t) = \operatorname{Tr}(\underline{g}_\sigma(t\boldsymbol{I}_n - \boldsymbol{A})) = \phi_\sigma(t) + n\rho \tag{19}$$

with $\underline{g}_\sigma = g_\sigma + \rho$. With [2] we get the following result:

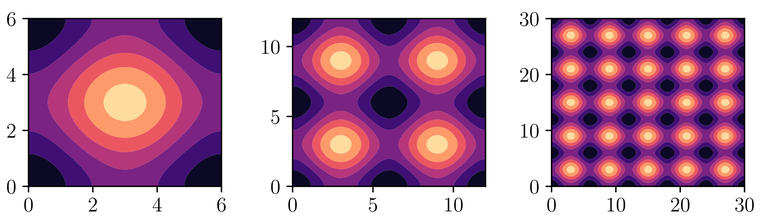
Theorem 2: Nyström-Chebyshev++ with shift

Let $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ be symmetric. If $n_\Omega = n_\Psi = \mathcal{O}(\varepsilon^{-1})$ and $\rho \ge \frac{\sqrt{2}}{n\sigma^2}(1+\sigma)^{-m}$, then with high probability

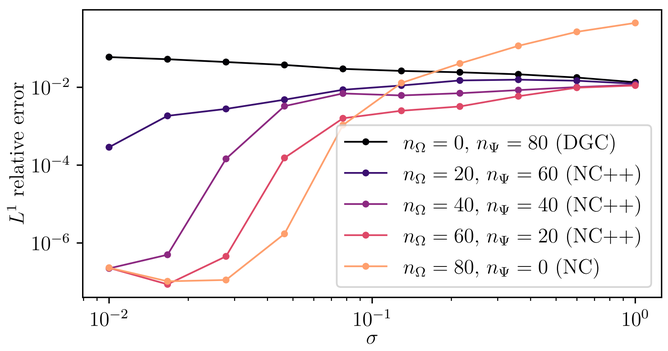
$$\|\underline{\phi}_\sigma - \breve{\underline{\phi}}_\sigma^{(m)}\|_1 \le (1+\varepsilon)\frac{2\sqrt{2}}{\sigma^2}(1+\sigma)^{-m} + \varepsilon(1+2n\rho).$$

Model problem

We consider the three-dimensional finite difference discretization matrix $\boldsymbol{A} \in \mathbb{R}^{1000\times 1000}$ of the Laplace operator in a potential of periodic Gaussian wells [1].

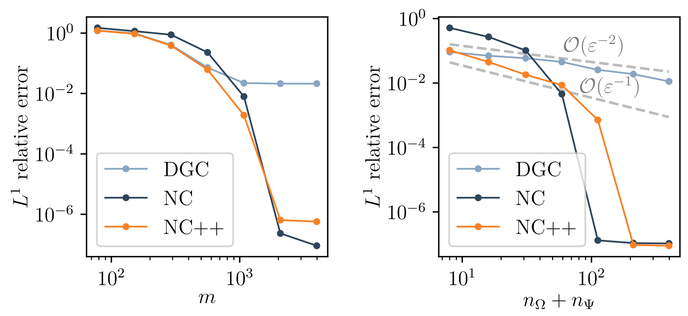


NC++ is a hybrid of DGC and NC. For small σ, the matrix in (8) has low rank, so the NC method performs well. For large σ, the Nyström approximation alone is not sufficient.

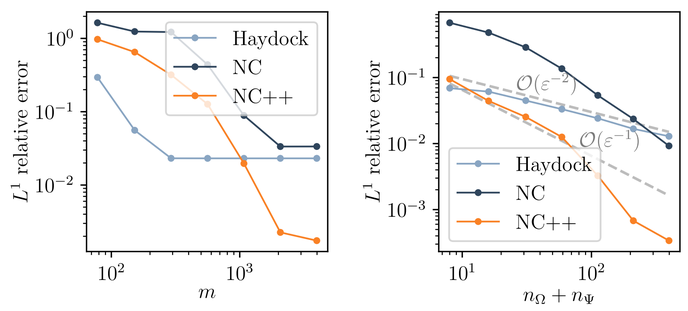


Convergence study

For fixed m but varying $n_\Omega + n_\Psi$, and vice versa, we plot the L^1 approximation error of all three methods on the model problem.



If instead of a Gaussian g_σ we use a Lorentzian $g_\sigma(s) \propto \sigma/(s^2+\sigma^2)$, there exists a specialized method called the Haydock method. We repeat the same plots as before.



Conclusion

We were able to significantly improve many algorithmic aspects of the methods. The development of an alternative expansion framework based on the DCT allowed us to vastly simplify the Chebyshev expansion, while obtaining provable accuracy, all this in addition to making this stage orders of magnitude faster in many cases. Furthermore, we derived theoretical error bounds for all the methods.

References

[1] L. Lin, "Randomized estimation of spectral densities of large matrices made accurate," *Numerische Mathematik*, vol. 136, pp. 183–213, 2017.

[2] H. He, D. Kressner, H. L. Lam, and F. Matti, "Parameter dependent Nystrom++ with application in spectral density function theory," *In preparation*, 2024.

This document is provably reproducible.
> hosted at https://github.com/FMatti/Rand-SD
> built on 2024-02-05 at 18:44:29 UTC from ca391a8

Master's thesis
Lausanne, 02/02/2024