

Stochastic trace estimation for parameter-dependent matrices applied to spectral density approximation

May 19, 2026

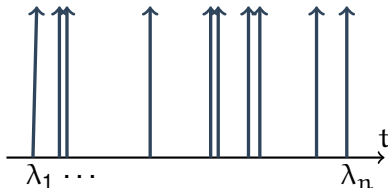
Fabio Matti (EPFL)

Joint work at EPFL with Haoze He,
Daniel Kressner, and Hei Yin Lam
(to appear in SIAM J. Sci. Comput.)

Spectral density

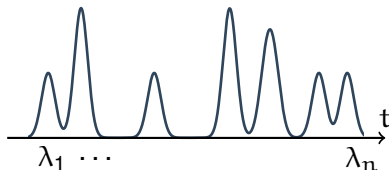
$\mathbf{A} \in \mathbb{R}^{n \times n}$ symmetric with eigenvalues $\lambda_1, \dots, \lambda_n$, $t \in \mathbb{R}$

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

*Smoothed* spectral density

$\mathbf{A} \in \mathbb{R}^{n \times n}$ symmetric with eigenvalues $\lambda_1, \dots, \lambda_n$,
 $g_\sigma : \mathbb{R} \rightarrow \mathbb{R}^+$ smooth (e.g. Gaussian of width σ), $t \in \mathbb{R}$

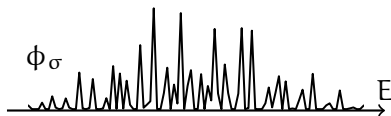
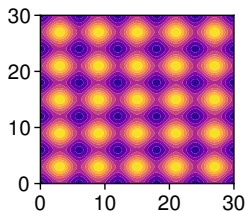
$$\phi_\sigma(t) = \sum_{i=1}^n g_\sigma(t - \lambda_i)$$



Electronic structure Hamiltonian

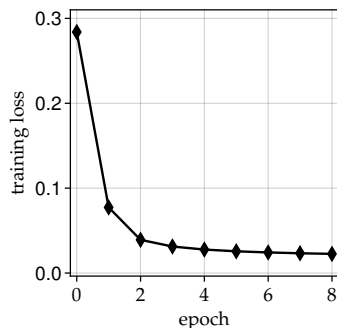
$$\mathcal{H}u(\mathbf{x}) = -\Delta u(\mathbf{x}) + V(\mathbf{x})u(\mathbf{x})$$

$V(\mathbf{x})$ periodic Gaussian wells potential [Lin, 2017]

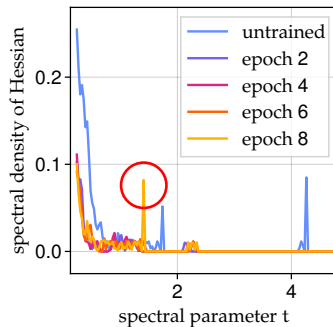
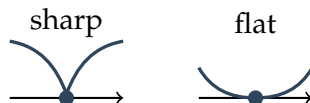


Neural network optimization

$$\text{loss} := \sum_i \|\mathbf{y}_i - \text{NN}(\mathbf{x}_i)\|$$



Type of minimum



- ▶ Often large/implicit $\mathbf{A}$
 - $\implies$ difficult to compute $\lambda_1, \dots, \lambda_n$
 - $\implies$ can't form $\phi_\sigma(t) = \sum_{i=1}^n g_\sigma(t - \lambda_i)$

- ▶ Since $\sum_{i=1}^n f(\lambda_i) = \text{Tr}(f(\mathbf{A}))$:

$$\phi_\sigma(t) = \text{Tr}(\mathbf{B}(t)), \quad \mathbf{B}(t) = g_\sigma(t\mathbf{I} - \mathbf{A})$$

$\implies \phi_\sigma$ is trace of parameter-dependent SPD matrix $\mathbf{B}(t)$

- ▶ **Common approach:** Do trace estimation [Lin et al., 2016]

Trace estimation for parameter-dependent matrices

Sample $\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_{n_\Psi} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$ and form

$$\text{Tr}(\mathbf{B}) \approx \text{Tr}_\Psi(\mathbf{B}) := \frac{1}{n_\Psi} \sum_{j=1}^{n_\Psi} \boldsymbol{\psi}_j^\top \mathbf{B} \boldsymbol{\psi}_j$$

[Hutchinson, 1990, Girard, 1989]

Constant case: Girard-Hutchinson [Meyer et al., 2021]

Provided $n_\Psi = \mathcal{O}(\varepsilon^{-2} \log(\delta^{-1}))$, then with probability $\geq 1 - \delta$

$$|\text{Tr}(\mathbf{B}) - \text{Tr}_\Psi(\mathbf{B})| \leq \varepsilon \|\mathbf{B}\|_F$$

Sample $\psi_1, \dots, \psi_{n_\Psi} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$ (same for all t) and form

$$\mathrm{Tr}(\mathbf{B}(t)) \approx \mathrm{Tr}_\Psi(\mathbf{B}(t)) := \frac{1}{n_\Psi} \sum_{j=1}^{n_\Psi} \psi_j^\top \mathbf{B}(t) \psi_j$$

Lemma I: Girard-Hutchinson [M. et al., 2026]

Provided $n_\Psi = \mathcal{O}(\varepsilon^{-2} \log(\delta^{-1}))$, then with probability $\geq 1 - \delta$

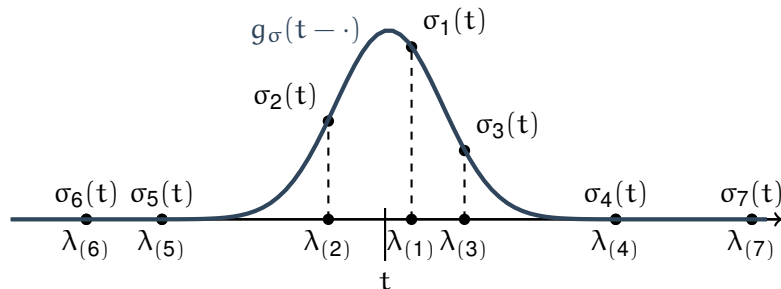
$$\int_a^b |\mathrm{Tr}(\mathbf{B}(t)) - \mathrm{Tr}_\Psi(\mathbf{B}(t))| \, dt \leq \varepsilon \int_a^b \|\mathbf{B}(t)\|_F \, dt$$

Singular values of $\mathbf{B}(t) = g_\sigma(t\mathbf{I} - \mathbf{A})$:

$$\sigma_i(t) = g_\sigma(t - \lambda_{(i)}), \quad i = 1, \dots, n$$

with eigenvalues $\lambda_{(1)}, \dots, \lambda_{(n)}$ of $\mathbf{A}$ sorted as

$$|t - \lambda_{(1)}| \leq \dots \leq |t - \lambda_{(n)}|$$



Sample Gaussian $\mathbf{\Omega} \in \mathbb{R}^{n \times n_{\Omega}}$ (same for all t) and form

$$\mathbf{B}_{\Omega}(t) = (\mathbf{B}(t)\mathbf{\Omega})(\mathbf{\Omega}^{\top}\mathbf{B}(t)\mathbf{\Omega})^{\dagger}(\mathbf{B}(t)\mathbf{\Omega})^{\top}$$

to approximate

$$\text{Tr}(\mathbf{B}(t)) \approx \text{Tr}_{\Omega}(\mathbf{B}(t)) := \text{Tr}(\mathbf{B}_{\Omega}(t)),$$

Theorem: Nyström estimator [M. et al., 2026]

If $n_{\Omega} \geq r + 4 \geq 6$, $\gamma \geq 1$, then with probability $\geq 1 - \gamma^{r-n_{\Omega}}$

$$\int_a^b |\text{Tr}(\mathbf{B}(t)) - \text{Tr}_{\Omega}(\mathbf{B}(t))| \, dt \leq \gamma^2(1+r) \int_a^b \sum_{i=r+1}^n \sigma_i(t) \, dt$$

- ▶ Girard-Hutchinson converges slowly (Monte-Carlo)
- ▶ Nyström not effective if eigenvalues of $\mathbf{A}$ are clustered

Use residual trace estimation [Lin, 2017, Meyer et al., 2021, Persson et al., 2022]

$$\mathrm{Tr}(\mathbf{B}(t)) \approx \mathrm{Tr}_{\Psi, \Omega}(\mathbf{B}(t)) := \mathrm{Tr}(\mathbf{B}_{\Omega}(t)) + \underbrace{\mathrm{Tr}_{\Psi}(\mathbf{B}(t) - \mathbf{B}_{\Omega}(t))}_{\text{Nyström residual}}$$

Lemma II: Nyström estimator [M. et al., 2026]

Provided $n_{\Omega} = \mathcal{O}(\varepsilon^{-2} + \log(\delta^{-1}))$, then with probability $\geq 1 - \delta$

$$\int_a^b \|\mathbf{B}(t) - \mathbf{B}_{\Omega}(t)\|_F dt \leq \varepsilon \int_a^b \mathrm{Tr}(\mathbf{B}(t)) dt$$

Theorem: Nyström++ estimator [M. et al., 2026]

Provided $n_{\Psi} = n_{\Omega} = \mathcal{O}(\varepsilon^{-1} \log(\delta^{-1}))$, with probability $\geq 1 - \delta$

$$\int_a^b |\mathrm{Tr}(\mathbf{B}(t)) - \mathrm{Tr}_{\Psi, \Omega}(\mathbf{B}(t))| \, dt \leq \varepsilon \int_a^b \mathrm{Tr}(\mathbf{B}(t)) \, dt$$

Proof:

$$\begin{aligned} & \int_a^b |\mathrm{Tr}(\mathbf{B}(t)) - \mathrm{Tr}_{\Psi, \Omega}(\mathbf{B}(t))| \, dt \\ &= \int_a^b |\mathrm{Tr}(\mathbf{B}(t) - \mathbf{B}_{\Omega}(t)) - \mathrm{Tr}_{\Psi}(\mathbf{B}(t) - \mathbf{B}_{\Omega}(t))| \, dt \\ &\stackrel{\text{Lem I}}{\leq} \tilde{\varepsilon} \int_a^b \|\mathbf{B}(t) - \mathbf{B}_{\Omega}(t)\|_F \, dt \\ &\stackrel{\text{Lem II}}{\leq} \tilde{\varepsilon}^2 \int_a^b \mathrm{Tr}(\mathbf{B}(t)) \, dt \end{aligned}$$

Nystrom++ estimator

$$\text{Tr}_{\Psi, \Omega}(\mathbf{B}(t)) = \text{Tr}(\mathbf{B}_{\Omega}(t)) + \text{Tr}_{\Psi}(\mathbf{B}(t) - \mathbf{B}_{\Omega}(t))$$

- If $n_{\Omega} = 0$:

$$\text{Tr}_{\Psi, \mathbf{0}}(\mathbf{B}(t)) = \text{Tr}(\mathbf{B}_{\mathbf{0}}(t)) + \text{Tr}_{\Psi}(\mathbf{B}(t) - \mathbf{B}_{\mathbf{0}}(t))$$

$\implies$ Girard-Hutchinson estimator

- If $n_{\Psi} = 0$:

$$\text{Tr}_{\mathbf{0}, \Omega}(\mathbf{B}(t)) := \text{Tr}(\mathbf{B}_{\Omega}(t)) + \text{Tr}_{\mathbf{0}}(\mathbf{B}(t) - \mathbf{B}_{\Omega}(t))$$

$\implies$ Nystrom estimator

Algorithm for spectral density estimation

We consider the following spectral density approximation

$$\phi_{\sigma}(t) \approx \hat{\phi}_{\sigma}(t) := \text{Tr}_{\Psi, \Omega} (g_{\sigma}(t\mathbf{I} - \mathbf{A}))$$

Problem: Difficult to evaluate $g_{\sigma}(t\mathbf{I} - \mathbf{A})$:

- ▶ $\mathbf{A}$ typically large/implicit
- ▶ g_{σ} usually Gaussian

Solution: Use Chebyshev interpolant of g_{σ} ¹ [Lin, 2017]

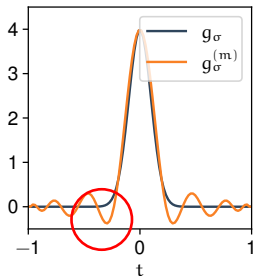
¹assume WLOG spectrum of $\mathbf{A}$ in $[-1, 1]$

Approximate

$$g_{\sigma}(t - \cdot) \approx g_{\sigma}^{(m)}(t - \cdot) := \sum_{\ell=0}^m \mu_{\ell}(t) T_{\ell}(\cdot)$$

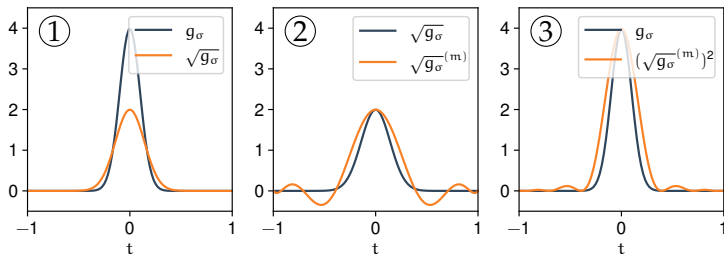
with $\{\mu_{\ell}(t)\}_{\ell=0}^m = \text{DCT}^{-1}\{g_{\sigma}(t - s_k)\}_{k=0}^m$, $s_k = \cos(\pi k/m)^2$

Problem: $g_{\sigma}^{(m)}$ can be negative $\implies g_{\sigma}^{(m)}(t\mathbf{I} - \mathbf{A})$ indefinite



Solution: Non-negative Chebyshev interpolation:

- 1 Take square root
- 2 Interpolate square root
- 3 Square interpolant



- $\mathcal{O}(m \log(m))$ squaring with DCT [Baszenski, Tasche, 1997]
- Exponential convergence maintained [M. et al., 2026]

The final spectral density estimator is

$$\phi_\sigma(t) \approx \hat{\phi}_\sigma^{(m)}(t) := \overbrace{\text{Tr}_{\Psi, \Omega}(\underbrace{g_\sigma^{(m)}(t\mathbf{I} - \mathbf{A}))}_{\text{Chebyshev interp.}}}_{\text{Nyström++}}$$

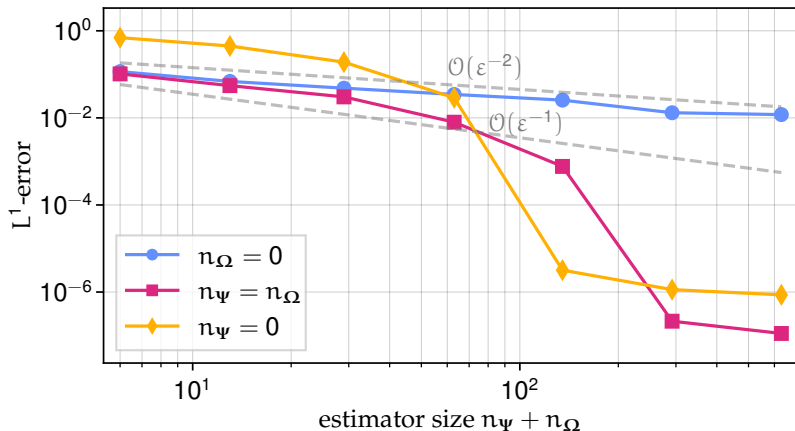
Theorem: Chebyshev-Nyström++ method [M. et al., 2026]

Provided $m = \mathcal{O}((\log(\varepsilon^{-1}) + \log(\sigma^{-1}))/\log(1 + \sigma))$ and $n_\Omega = n_\Psi = \mathcal{O}(\varepsilon^{-1} \log(\delta^{-1}))$, then with probability $\geq 1 - \delta$

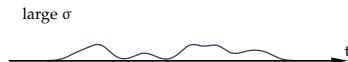
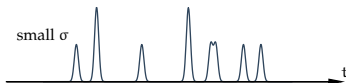
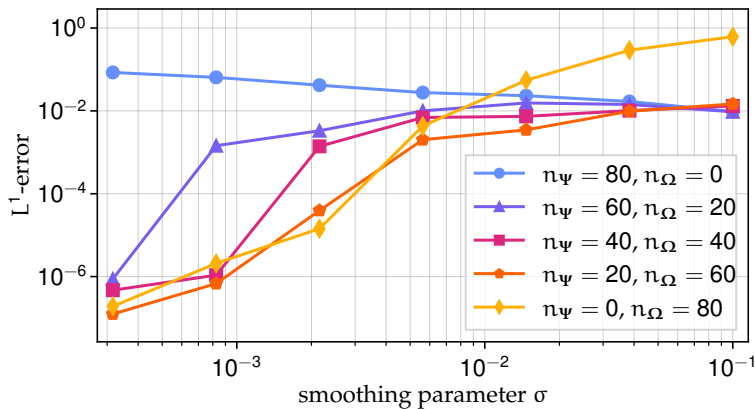
$$\int_{-1}^1 \left| \phi_\sigma(t) - \hat{\phi}_\sigma^{(m)}(t) \right| dt \leq \varepsilon$$

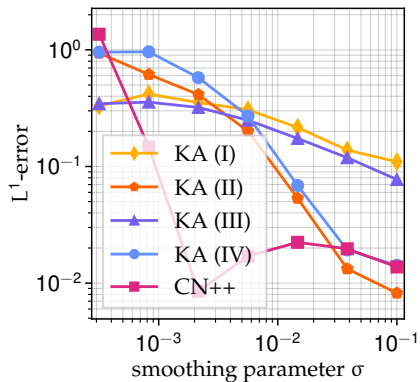
Numerical experiments

Girard-Hutchinson queries + Nyström sketch size = fixed



Girard-Hutchinson queries + Nyström sketch size = fixed





	n_{Ω}	n_{Ψ}	q	r	time (s)
KA (I)	10	60	20	10	3.42
KA (II)	20	30	20	10	2.12
KA (III)	10	30	40	10	3.92
KA (IV)	10	30	20	20	2.20

	n_{Ω}	n_{Ψ}	m	time (s)
CN++	20	20	2000	3.77

- Baszenski Günter, Tasche Manfred.* Fast polynomial multiplication and convolutions related to the discrete cosine transform // Linear Algebra Appl. 1997. 252, 1. 1–25.
- Girard Didier A.* A fast ‘Monte-Carlo cross-validation’ procedure for large least squares problems with noisy data // Numer. Math. 1989. 56, 1. 1–23.
- Hutchinson Michael F.* A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines // Comm. Statist. Simulation Comput. 1990. 19, 2. 433–450.
- Lin Lin.* Randomized estimation of spectral densities of large matrices made accurate // Numer. Math. 2017. 136, 1. 183–213.
- Lin Lin, Saad Yousef, Yang Chao.* Approximating spectral densities of large matrices // SIAM Rev. 2016. 58, 1. 34–65.

- M. F., He Haoze, Kressner Daniel, Lam Hei Yin.* Stochastic trace estimation for parameter-dependent matrices applied to spectral density approximation // SIAM J. Sci. Comput. 2026. to appear.
- Meyer Raphael A., Musco Cameron, Musco Christopher, Woodruff David P.* Hutch++: Optimal stochastic trace estimation // Proc. 2021 Symp. Simpl. Algorithms. Philadelphia, PA: SIAM, 2021. 142–155.
- Persson David, Cortinovis Alice, Kressner Daniel.* Improved variants of the Hutch++ algorithm for trace estimation // SIAM J. Matrix Anal. Appl. 2022. 43, 3. 1162–1185.