Randomized Estimation of Spectral Densities

February 6, 2024

Fabio Matti

Outline

Introduction

Theoretical analysis

Algorithmic improvements

Numerical results

This document is provably reproducible.

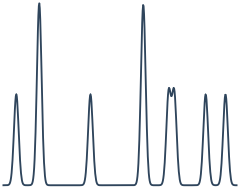
> hosted at https://github.com/FMatti/Rand-SD
> built on 2024-02-05 at 18:44:29 UTC from ca391a8

Introduction

Spectral density

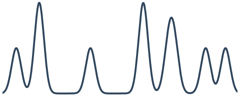
- Spectral density ϕ of symmetric $\boldsymbol{A} \in \mathbb{R}^{n \times n}$

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



- Smooth spectral density ϕ_σ

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



- Smoothing kernel g_σ

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

Chebyshev expansion

- Conversion to trace estimation

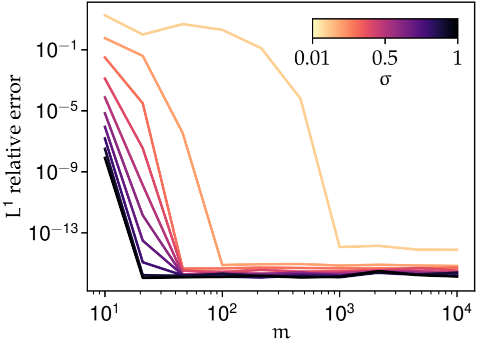
$$\phi_\sigma(t) = \mathrm{Tr}(g_\sigma(t\mathbf{I}_n - \boldsymbol{A})) \tag{4}$$

- Chebyshev expansion of matrix function

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

- Expanded spectral density

$$\phi_\sigma^{(m)}(t) = \mathrm{Tr}(g_\sigma^{(m)}(t\mathbf{I}_n - \boldsymbol{A})) \tag{6}$$



- Delta-Gauss-Chebyshev (DGC)

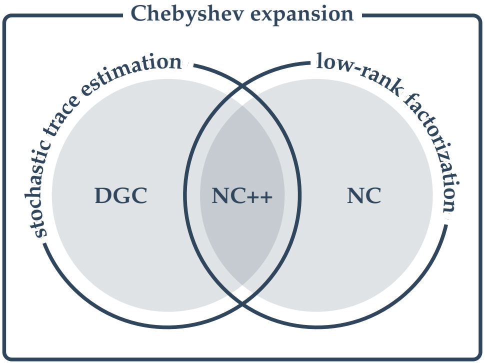
$$\widetilde{\phi}_{\sigma}^{(m)}(t) = \mathsf{H}_{n_{\Psi}}(g_{\sigma}^{(m)}(t\mathbf{I}_n - \boldsymbol{A})) \tag{7}$$

- Nyström-Chebyshev (NC)

$$\widehat{\phi}_{\sigma}^{(m)}(t) = \mathrm{Tr}(\widehat{g}_{\sigma}^{(m)}(t\mathbf{I}_n - \boldsymbol{A})) \tag{8}$$

- Nyström-Chebyshev++ (NC++)

$$\begin{aligned}\breve{\phi}_{\sigma}^{(m)}(t) &= \widehat{\phi}_{\sigma}^{(m)}(t) + \widetilde{\phi}_{\sigma}^{(m)}(t) \\ &\quad - \mathsf{H}_{n_{\Psi}}(\widehat{g}_{\sigma}^{(m)}(t\mathbf{I}_n - \boldsymbol{A}))\end{aligned} \tag{9}$$



Theoretical analysis

Theorem: Error of Delta-Gauss-Chebyshev method

$\widetilde{\phi}_\sigma^{(m)}(t)$ with DGC method on symmetric $\mathbf{A} \in \mathbb{R}^{n \times n}$ with spectrum in $[-1, 1]$ and Gaussian smoothing kernel g_σ. With high probability

$$\|\phi_\sigma - \widetilde{\phi}_\sigma^{(m)}\|_1 \leqslant \underbrace{\frac{\sqrt{2}}{\sigma^2}(1+\sigma)^{-m}\left(2 + c_\Psi \frac{1}{\sqrt{n n_\Psi}}\right)}_{\text{interpolation error and bias}} + \underbrace{c_\Psi \frac{1}{\sqrt{n_\Psi}}}_{\text{trace estimation}} \tag{10}$$

for some constant $c_\Psi \geqslant 0$.

- Even for "good" Chebyshev expansion, only $\mathcal{O}(\varepsilon^{-2})$ approximation

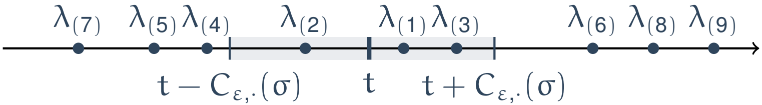
Numerical rank

- Numerical rank

$$r_{\varepsilon,\cdot}(g_\sigma(t\mathbf{I}_n - \boldsymbol{A})) \leqslant \#\{i : |t - \lambda_i| < C_{\varepsilon,\cdot}(\sigma)\} \tag{11}$$

- Constant (e.g. for nuclear norm $\|\cdot\|_*$)

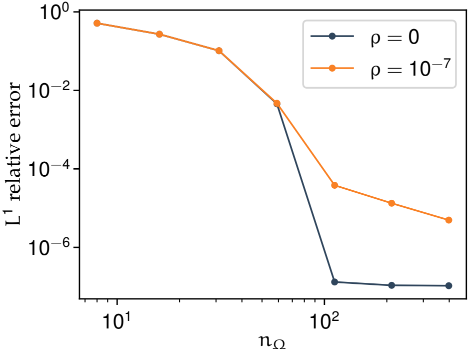
$$C_{\varepsilon,*}(\sigma) = \sigma\sqrt{-2\log(\sqrt{2\pi n}\sigma\varepsilon)} \tag{12}$$



- Most results for Nyström approximation only valid for PSD matrices
- Problem: Chebyshev expansion $g_\sigma^{(m)}$ may be negative
- Solution: Expand shifted kernel

$$\underline{g}_\sigma = g_\sigma + \rho \tag{13}$$

which is non-negative for large enough ρ



Theorem: Error of Nyström-Chebyshev method with shift

$\underline{\widehat{\phi}}_\sigma^{(m)}(t)$ with NC method on symmetric $\mathbf{A} \in \mathbb{R}^{n \times n}$ with spectrum in $[-1, 1]$ and shifted Gaussian smoothing kernel $\underline{g}_\sigma = g_\sigma + \rho$. If ρ large enough and $n_\Omega \gg r_{\varepsilon,*}(g_\sigma(t\mathbf{I}_n - \mathbf{A}))$ for all t, then with high probability

$$\|\underline{\phi}_\sigma - \underline{\widehat{\phi}}_\sigma^{(m)}\|_1 \lessapprox \underbrace{\frac{2\sqrt{2}}{\sigma^2}(1+\sigma)^{-m}}_{\text{interpolation error}} + \underbrace{4n(\varepsilon + 2\rho n)}_{\text{biased approximation error}} \tag{14}$$

- Significant approximation error because $\underline{g}_\sigma$ has a heavy tail

Theorem: Error of Nyström-Chebyshev++ method with shift

$\underline{\breve{\phi}}_{\sigma}^{(m)}(t)$ with NC++ method on symmetric $\mathbf{A} \in \mathbb{R}^{n \times n}$ with spectrum in $[-1, 1]$ and shifted Gaussian smoothing kernel $\underline{g}_{\sigma} = g_{\sigma} + \rho$. If ρ large enough and $n_{\Psi} = n_{\Omega} = \mathcal{O}(\varepsilon^{-1})$, then with high probability

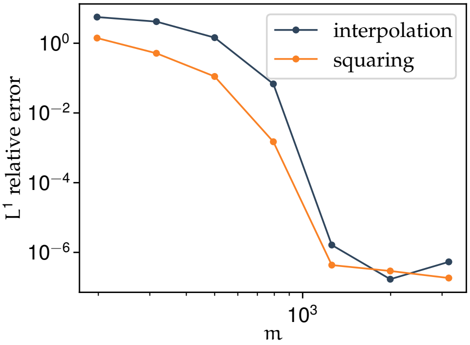
$$\|\underline{\phi}_{\sigma} - \underline{\breve{\phi}}_{\sigma}^{(m)}\|_1 \leqslant \underbrace{(1 + \varepsilon)\frac{2\sqrt{2}}{\sigma^2}(1 + \sigma)^{-m}}_{\text{interpolation error}} + \underbrace{\varepsilon(1 + 2n\rho)}_{\text{approximation error}} \tag{15}$$

Algorithmic improvements

- Expansion $f^{(m)}(s) = \sum_{l=0}^{m} \mu_l T_l(s)$ with discrete cosine transform (DCT)

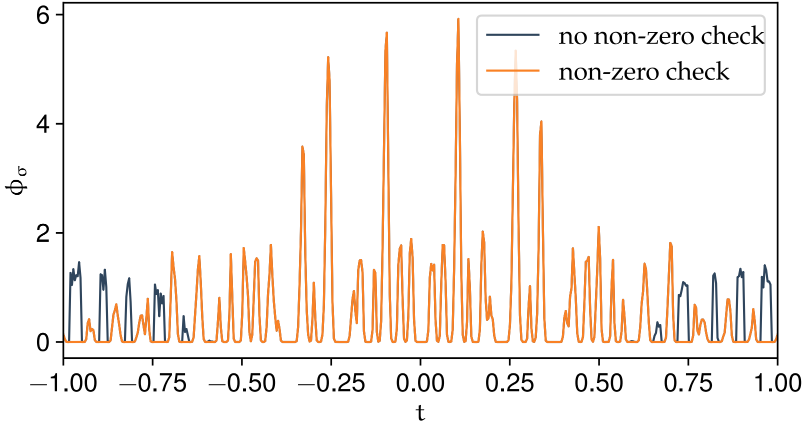
$$\{f(\cos(\pi i/m))\}_{i=0}^{m} \xleftrightarrow{\text{DCT}} \{\mu_l\}_{l=0}^{m} \tag{16}$$

- In the NC method we need to know $(g_\sigma(t\mathbf{I}_n - \mathbf{A}))^2$, e.g., with
  - $(g_\sigma^2)^{(m)}(t\mathbf{I}_n - \mathbf{A})$ (interpolation)
  - $(g_\sigma^{(m)}(t\mathbf{I}_n - \mathbf{A}))^2$ (squaring)



Non-zero check

- Problem: If $g_\sigma^{(m)}(t\mathbf{I}_n - \mathbf{A}) \approx \mathbf{0}$, then $(\boldsymbol{\Omega}^\top g_\sigma^{(m)}(t\mathbf{I}_n - \mathbf{A})\boldsymbol{\Omega})^\dagger$ bad idea
- Solution: If $\frac{1}{n_\Omega}\operatorname{Tr}(\boldsymbol{\Omega}^\top g_\sigma^{(m)}(t\mathbf{I}_n - \mathbf{A})\boldsymbol{\Omega}) < \kappa$ directly set $\widehat{\phi}_\sigma^{(m)}(t) = 0$



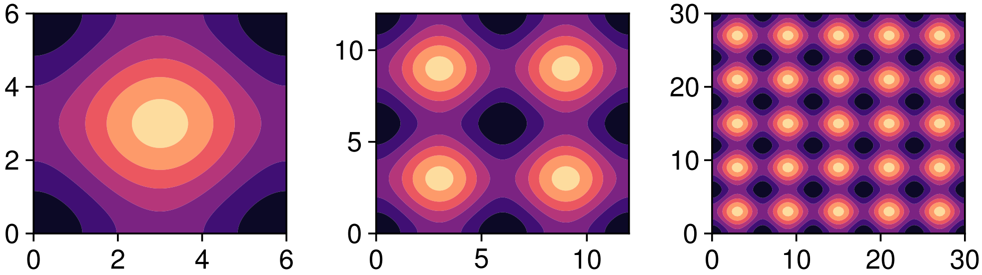
Numerical results

- Three-dimensional finite difference discretization [Lin, 2017]

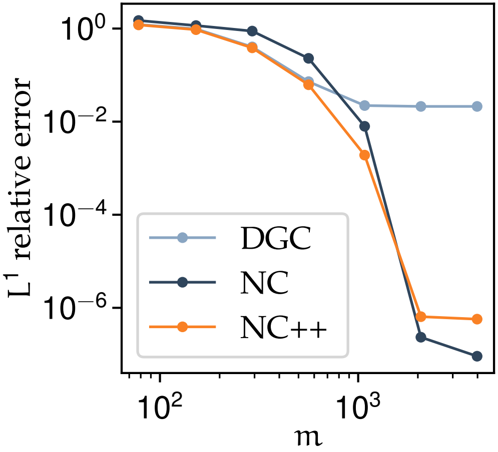
$$\mathcal{A}u(\mathbf{x}) = -\Delta u(\mathbf{x}) + V(\mathbf{x})u(\mathbf{x}) \tag{17}$$

- Periodic Gaussian wells

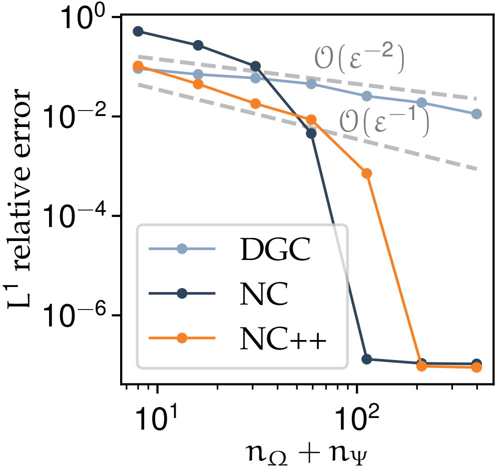
$$\alpha \exp\left(-\frac{\|\mathbf{x}\|_2^2}{2\beta^2}\right) \tag{18}$$



- Fix $n_\Omega + n_\Psi$, vary m

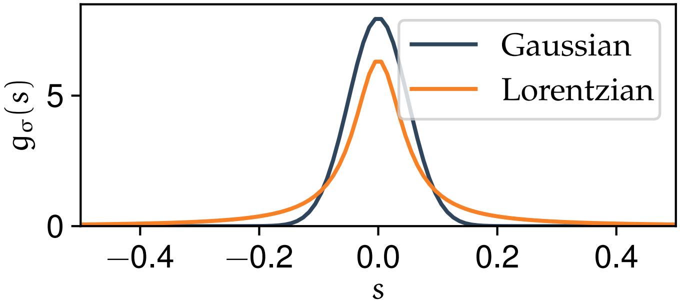


- Fix m, vary $n_\Omega + n_\Psi$

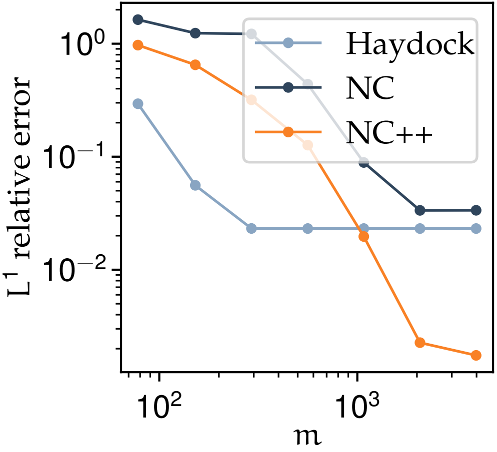


- Lorentzian smoothing kernel $\implies$ Haydock method [Lin et al., 2016]

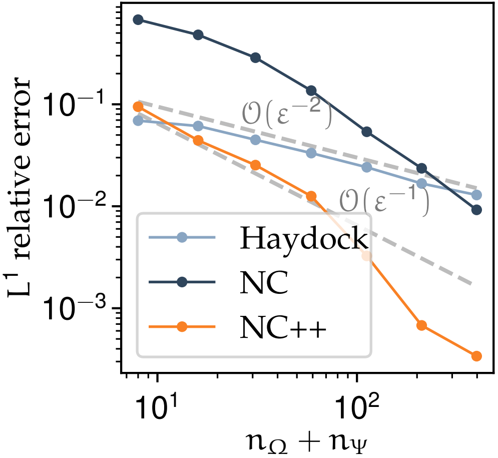
$$g_\sigma(s) = \frac{1}{\pi}\frac{\sigma}{s^2+\sigma^2} \tag{19}$$



- Fix $n_\Omega + n_\Psi$, vary m



- Fix m, vary $n_\Omega + n_\Psi$



Conclusion

Main contributions

- Developed a unified family of algorithms
- Introduced multiple improvements over [Lin, 2017]
- Derived error bounds for all methods

Outlook

- Make theoretical analysis without shift
- Find alternative to interpolation

- He, H., Kressner, D., Lam, H. L., and Matti, F. (2024).
  Parameter dependent Nystrom++ with application in spectral density function theory.
  In preparation.
- Lin, L. (2017).
  Randomized estimation of spectral densities of large matrices made accurate.
  Numerische Mathematik, 136:183–213.
- Lin, L., Saad, Y., and Yang, C. (2016).
  Approximating spectral densities of large matrices.
  SIAM Review, 58(1):34–65.

- Meyer, R. A., Musco, C., Musco, C., and Woodruff, D. P. (2021).
  Hutch++: Optimal stochastic trace estimation.
  arXiv.
- Trefethen, L. N. (2008).
  Is Gauss quadrature better than Clenshaw-Curtis?
  SIAM Review, 50(1):67–87.

Additional slides

Lemma: Chebyshev expansion (based on [Trefethen, 2008])

$\mathbf{A} \in \mathbb{R}^{n \times n}$ symmetric with spectrum in $[-1, 1]$. Then

$$\|\phi_\sigma - \phi_\sigma^{(m)}\|_1 \leqslant \frac{C_1}{\sigma^2}(1 + C_2\sigma)^{-m}. \tag{20}$$

Lemma: Parameter-dependent Girard-Hutchinson [He et al., 2024]

$\mathbf{B}(t) \in \mathbb{R}^{n \times n}$ symmetric and continuous in $t \in [a, b]$. With high probability

$$\int_a^b |\mathrm{Tr}(\mathbf{B}(t)) - H_{n_\Psi}(\mathbf{B}(t))| \, dt \leqslant c_\Psi \frac{1}{\sqrt{n_\Psi}} \int_a^b \|\mathbf{B}(t)\|_F dt. \tag{21}$$

Theorem: Error of Delta-Gauss-Chebyshev method

Let $\widetilde{\phi}_\sigma^{(m)}(t)$ be computed with the DGC method on a symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ with its spectrum contained in $[-1, 1]$ using a Gaussian smoothing kernel g_σ with smoothing parameter $\sigma > 0$, degree of expansion $m \in \mathbb{N}$, and number of Hutchinson's queries $n_\Psi \in \mathbb{N}$. For $\delta \in (0, e^{-1})$ it holds with probability $\geqslant 1 - \delta$, that

$$\|\phi_\sigma - \widetilde{\phi}_\sigma^{(m)}\|_1 \leqslant \frac{\sqrt{2}}{\sigma^2}(1+\sigma)^{-m}\left(2 + c_\Psi \frac{\log(1/\delta)}{\sqrt{n n_\Psi}}\right) + c_\Psi \frac{\log(1/\delta)}{\sqrt{n_\Psi}} \tag{22}$$

for $c_\Psi \geqslant 24e$.

- Parameter-dependent Nyström approximation

$$\widehat{f}(\mathbf{A}, t) = (f(\mathbf{A}, t)\mathbf{\Omega})(\mathbf{\Omega}^\top f(\mathbf{A}, t)\mathbf{\Omega})^\dagger (f(\mathbf{A}, t)\mathbf{\Omega})^\top \tag{23}$$

Theorem: Parameter-dependent Nyström [He et al., 2024]

$f(\mathbf{A}, t)$ function of symmetric $\mathbf{A} \in \mathbb{R}^{n \times n}$ which continuously depends on $t \in [a, b]$. Standard Gaussian $\mathbf{\Omega} \in \mathbb{R}^{n \times n_\Omega}$ with $n_\Omega > r + 3$. With high probability

$$\int_a^b |\mathrm{Tr}(f(\mathbf{A}, t)) - \mathrm{Tr}(\widehat{f}(\mathbf{A}, t))| dt < c(1 + r) \int_a^b \sum_{i=r+1}^{n} \sigma_i(t) dt \tag{24}$$

where $\sigma_i(t)$ are the (ordered) eigenvalues of $f(\mathbf{A}, t)$ at t.

Theorem: Error of Nyström-Chebyshev method with shift

Let $\widehat{\underline{\phi}}_{\sigma}^{(m)}$ be computed with the NC method on a symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ with its spectrum contained in $[-1, 1]$ using a shifted Gaussian smoothing kernel $\underline{g}_{\sigma} = g_{\sigma} + \rho$ with smoothing parameter $\sigma > 0$, degree of expansion $m \in \mathbb{N}$, and sketch size $n_{\Omega} = r + p$ for some numbers $r \geqslant 2$, $p \geqslant 4$. If shift $\rho \geqslant \frac{\sqrt{2}}{n\sigma^2}(1+\sigma)^{-m}$ and $r \geqslant r_{\varepsilon,*}(g_{\sigma}(t\mathbf{I}_n - \mathbf{A}))$ for all $t \in [-1, 1]$, then for all $\gamma \geqslant 1$, the inequality

$$\|\underline{\phi}_{\sigma} - \widehat{\underline{\phi}}_{\sigma}^{(m)}\|_1 \leqslant 2\gamma^2(1+r)(2\varepsilon + 4\rho(n-r)) + \frac{2\sqrt{2}}{\sigma^2}(1+\sigma)^{-m} \tag{25}$$

holds with probability $\geqslant 1 - \gamma^{-p}$.

- Goal: prove $\mathcal{O}(\varepsilon^{-1})$ result for Nyström-Chebyshev++ method

Lemma: Parameter-dependent Nyström 2.0 [He et al., 2024]

$\mathbf{B}(t) \in \mathbb{R}^{n \times n}$ symmetric positive semi-definite (PSD) and continuous in $t \in [a, b]$. With high probability

$$\int_a^b \|\mathbf{B}(t) - \widehat{\mathbf{B}}(t)\|_F \mathrm{d}t \leqslant c_\Omega \frac{1}{\sqrt{n_\Omega}} \int_a^b \mathrm{Tr}(\mathbf{B}(t)) \mathrm{d}t. \tag{26}$$

Nyström-Chebyshev++ (main theorem)

- Parameter-dependent result for [Meyer et al., 2021]

$$\mathrm{Tr}^{++}(\mathbf{B}(t)) = \mathrm{Tr}(\widehat{\mathbf{B}}(t)) + \mathrm{H}_{n_\Psi}(\boldsymbol{\Delta}(t)) \tag{27}$$

Theorem: Parmeter-dependent trace estimation

$\mathbf{B}(t) \in \mathbb{R}^{n \times n}$ is symmetric PSD and continuous in $t \in [a, b]$, and

$$\mathrm{Tr}(\mathbf{B}(t)) = \mathrm{Tr}(\widehat{\mathbf{B}}(t)) + \mathrm{Tr}(\boldsymbol{\Delta}(t)) \text{ and } \int_a^b \|\boldsymbol{\Delta}(t)\|_F \mathrm{d}t \leqslant c_\Omega \frac{1}{\sqrt{n_\Omega}} \int_a^b \mathrm{Tr}(\mathbf{B}(t)) \mathrm{d}t. \tag{28}$$

With high probability

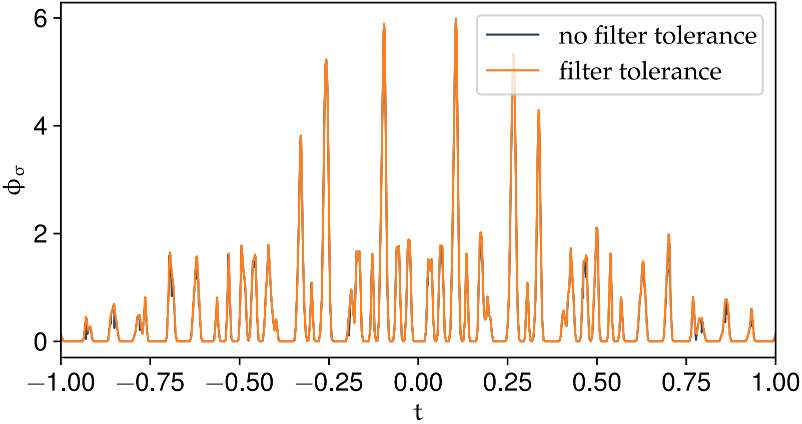
$$\int_a^b |\mathrm{Tr}^{++}(\mathbf{B}(t)) - \mathrm{Tr}(\mathbf{B}(t))| \mathrm{d}t \leqslant c \frac{1}{\sqrt{n_\Psi n_\Omega}} \int_a^b \mathrm{Tr}(\mathbf{B}(t)) \mathrm{d}t. \tag{29}$$

Theorem: Error of Nyström-Chebyshev++ method with shift

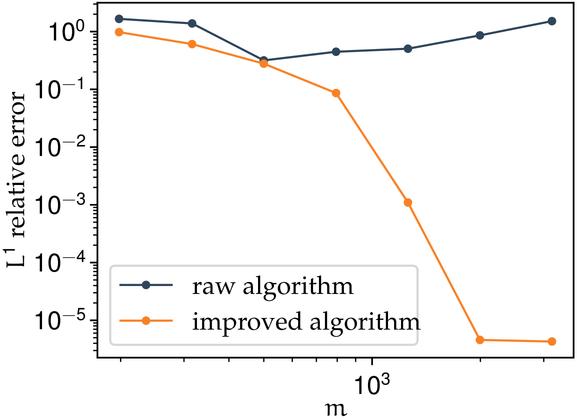
Let $\breve{\underline{\phi}}_{\sigma}^{(m)}$ be computed with the NC++ method on a symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ with its spectrum contained in $[-1, 1]$ using a shifted Gaussian smoothing kernel $\underline{g}_{\sigma} = g_{\sigma} + \rho$ and with the parameters smoothing parameter $\sigma > 0$, degree of expansion $m \in \mathbb{N}$, and number of Hutchinson's queries $n_{\Psi} + n_{\Omega} = \mathcal{O}(\log(1/\delta)/\varepsilon)$ with even $n_{\Omega} \geqslant 8 \log(1/\delta)$. If $\rho \geqslant \frac{\sqrt{2}}{n\sigma^2}(1 + \sigma)^{-m}$, then for $\delta \in (0, e^{-1})$ with probability $\geqslant 1 - \delta$

$$\|\underline{\phi}_{\sigma} - \breve{\underline{\phi}}_{\sigma}^{(m)}\|_1 \leqslant (1 + \varepsilon)\frac{2\sqrt{2}}{\sigma^2}(1 + \sigma)^{-m} + \varepsilon(1 + 2n\rho). \tag{30}$$

- Problem: Values just slightly outside range of g_σ are filtered out
- Solution: Add tolerance η to range $[0, (1+\eta)/n\sqrt{2\pi\sigma^2}]$



Combined algorithmic improvements



Allocation of matrix-vector products