A medley of fun techniques for solving interesting problems in RandNLA

Fabio Matti[†]

[†]ANCHP, École polytechnique fédérale de Lausanne (EPFL), Lausanne, Switzerland

Sometimes problems can be solved in surprisingly fun or simple ways. Below is a collection of examples we have encountered recently.

PROBLEM I

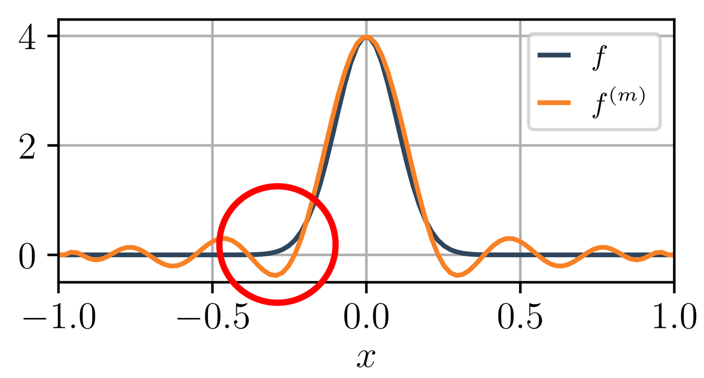
Task: Non-negative polynomial approx.

Approximate non-negative function $f : \mathbb{R} \to \mathbb{R}_{\geq 0}$ with non-negative polynomial $p_m \in \mathbb{P}_m$.

Chebyshev interpolation: Scale Chebyshev polynomials T_ℓ with $\mu_\ell \in \mathbb{R}$ to approximate

$$f(x) \approx f^{(m)}(x) = \sum_{\ell=0}^{m} \mu_\ell T_\ell(x)$$

Observation: Chebyshev interpolant of a non-negative function is not always non-negative.

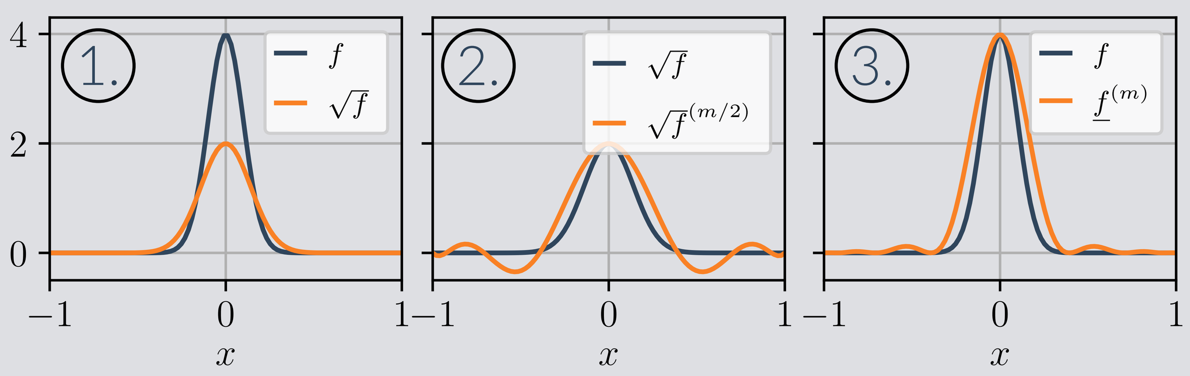


Issue: When used as matrix function approximation $f(\boldsymbol{A}) \approx f^{(m)}(\boldsymbol{A})$, might change definiteness.

Positivity-preserving interpolation

To get a non-negative polynomial approximation of f, we propose the following simple procedure:

1. Take the square root $\sqrt{f} : \mathbb{R} \to \mathbb{R}^{+}$ of f
2. Compute degree $m/2$ interpolant $\sqrt{f}^{(m/2)}$ of $\sqrt{f}$
3. Square interpolant $\underline{f}^{(m)} := (\sqrt{f}^{(m/2)})^2$



Clearly, $\underline{f}^{(m)}$ interpolates f and, being a square, is non-negative.

Computation and convergence

Fast discrete cosine transform (FDCT) enables cheap conversion between function evaluations at Chebyshev nodes $s_j = \cos(\pi j/m)$, $j = 0, \dots, m$ and coefficients μ_ℓ, $\ell = 0, \dots, m$ of $f^{(m)}$:

$$\{\mu_\ell\}_{\ell=0}^{m} \xleftrightarrow[\mathcal{O}(m\log(m))]{\text{FDCT}} \{f(s_j)\}_{j=0}^{m}$$

Non-negative approximation $\underline{f}^{(m)}$ keeps convergence properties of Chebyshev interpolant:

Theorem: Maintained convergence [1]

Let $E^{(m)}$ denote the error of a degree m Chebyshev interpolant $f^{(m)}$ of f. Then

$$\|f - \underline{f}^{(m)}\|_\infty \leq (C + E^{(m/2)}(\sqrt{f}))E^{(m/2)}(\sqrt{f})$$

PROBLEM II

Task: Higher-order moment bound

Given $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ and Gaussian random vector $\boldsymbol{\omega} \in \mathbb{R}^n$, bound the higher-order moments

$$\mathbb{E}^p\left[\|\boldsymbol{A\omega}\|_2^2\right] := \left(\mathbb{E}\left[\|\boldsymbol{A\omega}\|_2^{2p}\right]\right)^{1/p}.$$

Basic bounds: Simple arguments show that the singular values matrix $\boldsymbol{\Sigma} = \operatorname{diag}(\sigma_1, \dots, \sigma_n)$, $\sigma_1 \geq \dots \geq \sigma_n$ of $\boldsymbol{A}$ satisfies $\mathbb{E}^p\left[\|\boldsymbol{A\omega}\|_2^2\right] = \mathbb{E}^p\left[\|\boldsymbol{\Sigma\omega}\|_2^2\right]$.

- With $\sigma_1 \geq \sigma_i$, $i = 1, \dots, n$:
$$\begin{aligned}\mathbb{E}^p\left[\|\boldsymbol{A\omega}\|_2^2\right] &= \mathbb{E}^p\left[\sigma_1^2\omega_1^2 + \dots + \sigma_n^2\omega_n^2\right] \\ &\leq \sigma_1^2\mathbb{E}^p\left[\|\boldsymbol{\omega}\|_2^2\right] \leq \|\boldsymbol{A}\|_2^2(p+n),\end{aligned}$$
- With triangle inequality:
$$\begin{aligned}\mathbb{E}^p\left[\|\boldsymbol{A\omega}\|_2^2\right] &= \mathbb{E}^p\left[\sigma_1^2\omega_1^2 + \dots + \sigma_n^2\omega_n^2\right] \\ &\leq \sigma_1^2\mathbb{E}^p\left[\omega_1^2\right] + \dots + \sigma_n^2\mathbb{E}^p\left[\omega_n^2\right] \leq \|\boldsymbol{A}\|_F^2 p.\end{aligned}$$

Issue: Not tight enough for some purposes, as they might scale as $\mathcal{O}(n)$ and $\mathcal{O}(np)$, respectively.

Grouping singular values

To get a tighter bound, we split singular values into $\ell = \lceil n/k \rceil$ groups of k:

$$\underbrace{\sigma_1, \dots, \sigma_k}_{\text{group 1}}, \underbrace{\sigma_{k+1}, \dots, \sigma_{2k}}_{\text{group 2}}, \dots, \underbrace{\sigma_{(\ell-1)k+1}, \dots, \sigma_{\ell k}}_{\text{group } \ell}.$$

With $\sigma_{ik+1}^2 \leq (\sigma_{(i-1)k+1}^2 + \dots + \sigma_{ik-1}^2 + \sigma_{ik}^2)/k$:

$$\begin{aligned}\sum_{i=0}^{\ell-1} \sigma_{ik+1}^2 &= \sigma_1^2 + \sum_{i=1}^{\ell-1} \sigma_{ik+1}^2 \leq \|\boldsymbol{A}\|_2^2 + \frac{1}{k}\sum_{j=1}^{(\ell-1)k} \sigma_j^2 \\ &\leq \|\boldsymbol{A}\|_2^2 + \frac{1}{k}\sum_{j=1}^{n} \sigma_j^2 \leq \|\boldsymbol{A}\|_2^2 + \frac{\|\boldsymbol{A}\|_F^2}{k}.\end{aligned}$$

Applying triangle inequality to each group:

$$\mathbb{E}^p\left[\|\boldsymbol{A\omega}\|_2^2\right] = \mathbb{E}^p\left[\sum_{i=1}^{n}\sigma_i^2\omega_i^2\right] \leq \sum_{i=0}^{\ell-1}\sigma_{ik+1}^2\,\mathbb{E}^p\left[\sum_{j=1}^{k}\omega_{ik+j}^2\right]$$

$$\leq (k+p-1)\sum_{i=0}^{\ell-1}\sigma_{ik+1}^2 \leq (k+p-1)\left(\|\boldsymbol{A}\|_2^2 + \frac{\|\boldsymbol{A}\|_F^2}{k}\right)$$

Proving some interesting results

Generalizes Lem. B.1 in [2] from $p = 2, 4$ to $p \geq 2$:

Theorem: Moment bound [1]

Let $\boldsymbol{\Omega} \in \mathbb{R}^{n\times k}$ with $k \geq 2$. Then for any $p \geq 2$,

$$\mathbb{E}^p\left[\|\boldsymbol{A\Omega}\|_2\right] \leq \sqrt{k + p/2 - 1}\left(\sqrt{2}\|\boldsymbol{A}\|_2 + \frac{\|\boldsymbol{A}\|_F}{\sqrt{k}}\right).$$

Helped analyze parameter-dependent Nyström approx. $\widehat{\boldsymbol{B}}_{\boldsymbol{\Omega}}(t) := (\boldsymbol{B}(t)\boldsymbol{\Omega})(\boldsymbol{\Omega}^\top\boldsymbol{B}(t)\boldsymbol{\Omega})^\dagger(\boldsymbol{B}(t)\boldsymbol{\Omega})^\top$:

Theorem: Parameterized Nyström [1]

With high probability

$$\int_a^b \|\boldsymbol{B}(t) - \widehat{\boldsymbol{B}}_{\boldsymbol{\Omega}}(t)\|_F \,\mathrm{d}t \sim \frac{1}{\sqrt{n_{\boldsymbol{\Omega}}}}\int_a^b \operatorname{Tr}(\boldsymbol{B}(t))\,\mathrm{d}t$$

PROBLEM III

Task: Approximate inverse quadratic form

Given the symmetric positive definite matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ and a vector $\boldsymbol{b} \in \mathbb{R}^n$, approximate the quadratic form

$$\boldsymbol{b}^\top\boldsymbol{A}^{-1}\boldsymbol{b}.$$

Standard approximation: Such quadratic forms are often approximated by building an orthonormal basis $\boldsymbol{Q}_k \in \mathbb{R}^{n\times k}$ of the Krylov subspace

(1) $\quad \mathcal{K}_k(\boldsymbol{A}, \boldsymbol{b}) = \operatorname{span}(\boldsymbol{b}, \boldsymbol{Ab}, \dots, \boldsymbol{A}^{k-1}\boldsymbol{b}),$

along with the tridiagonal matrix $\boldsymbol{T}_k = \boldsymbol{Q}_k^\top\boldsymbol{AQ}_k$, for example with the Lanczos method.

Then, an approximation is of the form

(2) $\quad \boldsymbol{b}^\top\boldsymbol{A}^{-1}\boldsymbol{b} \approx \boldsymbol{b}^\top\boldsymbol{Q}_k\boldsymbol{T}_k^{-1}\boldsymbol{Q}_k^\top\boldsymbol{b} = \|\boldsymbol{b}\|_2^2(\boldsymbol{T}_k^{-1})_{11}.$

Issue: Common approximation results are not versatile enough when, e.g., non-standard choices for the Krylov subspace (1) are used.

Connecting it to conjugate gradient

Consider the linear system $\boldsymbol{Ax} = \boldsymbol{b}$ with corresponding conjugate gradient iterates $\boldsymbol{x}_k$.

We have:

$$\begin{aligned}\|\boldsymbol{x} - \boldsymbol{x}_k\|_{\boldsymbol{A}}^2 &= (\boldsymbol{x} - \boldsymbol{x}_k)^\top\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{x}_k) \\ &= \boldsymbol{x}^\top\boldsymbol{Ax} - 2\boldsymbol{x}^\top\boldsymbol{Ax}_k + \boldsymbol{x}_k^\top\boldsymbol{Ax}_k \\ &= \boldsymbol{b}^\top\boldsymbol{A}^{-1}\boldsymbol{b} - 2\boldsymbol{b}^\top\boldsymbol{x}_k + \boldsymbol{x}_k^\top\boldsymbol{Ax}_k.\end{aligned}$$

With $\boldsymbol{x}_k = \|\boldsymbol{b}\|_2\boldsymbol{Q}_k\boldsymbol{T}_k^{-1}\boldsymbol{e}_1$:

$$\begin{aligned}\|\boldsymbol{x} - \boldsymbol{x}_k\|_{\boldsymbol{A}}^2 &= \boldsymbol{b}^\top\boldsymbol{A}^{-1}\boldsymbol{b} - 2\|\boldsymbol{b}\|_2\underbrace{\boldsymbol{b}^\top\boldsymbol{Q}_k}_{=\|\boldsymbol{b}\|_2\boldsymbol{e}_1^\top}\boldsymbol{T}_k^{-1}\boldsymbol{e}_1 \\ &\quad + \|\boldsymbol{b}\|_2^2\boldsymbol{e}_1^\top\boldsymbol{T}_k^{-1}\underbrace{\boldsymbol{Q}_k^\top\boldsymbol{AQ}_k}_{=\boldsymbol{T}_k}\boldsymbol{T}_k^{-1}\boldsymbol{e}_1 \\ &= \boldsymbol{b}^\top\boldsymbol{A}^{-1}\boldsymbol{b} - \|\boldsymbol{b}\|_2^2(\boldsymbol{T}_k^{-1})_{11},\end{aligned}$$

which is exactly the error of the quadratic form approximation (2).

Some helpful consequences of this

Helps us establish the following bound:

Theorem: Inverse quadratic form [3]

For any $k \geq 1$

$$\frac{\boldsymbol{b}^\top\boldsymbol{A}^{-1}\boldsymbol{b} - \|\boldsymbol{b}\|_2^2(\boldsymbol{T}_k^{-1})_{11}}{\boldsymbol{b}^\top\boldsymbol{A}^{-1}\boldsymbol{b}} \leq 4\left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}\right)^{2k}$$

where κ is condition number of $\boldsymbol{A}$.

The optimality property of CG,

$$\boldsymbol{b}^\top\boldsymbol{A}^{-1}\boldsymbol{b} - \|\boldsymbol{b}\|_2^2(\boldsymbol{T}_k^{-1})_{11} = \min_{\boldsymbol{u}\in\mathcal{K}_k(\boldsymbol{A},\boldsymbol{b})}\|\boldsymbol{x} - \boldsymbol{u}\|_{\boldsymbol{A}}^2,$$

allows us to extend bounds for (2) to augmented Krylov subspaces $\mathcal{K}_k(\boldsymbol{A}, \boldsymbol{b}) + \mathcal{U}$ [3].

[1] FM, H. Y. Lam, H. He, and D. Kressner, "Stochastic trace estimation for parameter-dependent matrices applied to spectral density approximation," *SIAM J. Sci. Comput.*, 2026.

[2] J. A. Tropp and R. J. Webber, "Randomized algorithms for low-rank matrix approximation: design, analysis, and applications," *arXiv:2306.12418*, 2023.

[3] FM, M. S. Andersen, T. Chen, and D. Kressner, "Kernel-based linear system identification using augmented krylov subspaces," *arXiv:2605.08362*, 2026.