Spectral nearness problems via Riemannian optimization

Miryam Gnazzo[1], Vanni Noferini[2], Lauri Nyman[2], Federico Poloni[1,3]

1: University of Pisa 2: Aalto University
3: supported by the University of Bologna

INdAM Workshop: Low-rank Structures and Numerical Methods in Matrix and Tensor Computations
Cortona, September 2025

Spectral nearness problems

Given a matrix $A \in \mathbb{C}^{n\times n}$, what is the closest matrix X to it with a certain spectral property? E.g.,

- X singular
- X Hurwitz (all eigenvalues in left half-plane)
- X non-Hurwitz (at least one eigenvalue in right half-plane)
- ...

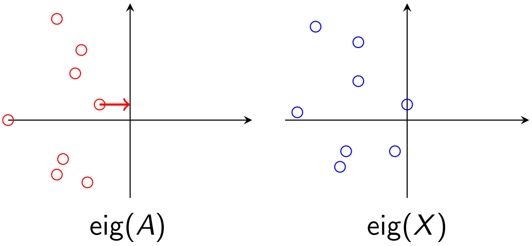
$$f(A) = \min_{X\in\mathcal{S}} \|A - X\|_F^2.$$

Optionally, add structure: X and A must share a certain (linear) structure: sparsity pattern, Toeplitz, ...

Idea: the Oracle

These problems become much easier if an oracle reveals a certain property of the solution.

Example from [Van Loan '85, Byers '88]: nearest (non-structured) non-Hurwitz matrix.



- Given A Hurwitz, the nearest non-Hurwitz perturbation must have an imaginary eigenvalue $\lambda = i\omega \in \mathtt{eig}(X)$.
- If the Oracle reveals ω, the problems becomes easy: $X - i\omega I$ must be the singular matrix nearest to $A - i\omega I$. So $f(A, \omega) = \sigma_{\min}(A - i\omega I)^2$.
- Now minimize on ω: $f(A) = \min_{\omega \in \mathbb{R}} f(A, \omega)$.

A better oracle

Problem: with this approach, it is hard to add structure!
Solution: a better oracle: reveal the eigenvector $\mathbf{v}$.

We start from a simpler problem: nearest singular structured matrix

$$f(A) = \min_{\Delta \text{ structured}} \|\Delta\|_F^2 \text{ s.t. } A + \Delta \text{ singular}$$

If we prescribe $\mathbf{v}$ in the kernel, the resulting problem is an underdetermined least-squares problem:

$$f(A, \mathbf{v}) = \min \|\Delta\|_F^2 \text{ s.t. } (A + \Delta)\mathbf{v} = 0 \iff \min \|\boldsymbol{\delta}\|^2 \text{ s.t. } M\boldsymbol{\delta} = \mathbf{r}.$$

for suitable $M, \mathbf{r}$ which depend on A and $\mathbf{v}$.
This holds even if we impose a linear structure on Δ: sparsity, Toeplitz, etc. In this case, the vector $\boldsymbol{\delta}$ contains the parameters that define Δ; e.g., $\Delta = \texttt{hankel}(\boldsymbol{\delta})$).

Again, the inner minimization problem is easy: $f(A, \mathbf{v}) = \|\boldsymbol{\delta}\|^2 = \|M^+\mathbf{r}\|^2$,

$$f(A) = \min_{\mathbf{v} \in \mathbb{R}^n \setminus \{\mathbf{0}\}} f(A, \mathbf{v}).$$

Some history

- For the two-level optimization idea: Variable projection for least-squares fitting problems that are linear only in some variables [N.E. Dahl '65, Lawton–Sylvestre '71, Golub–Pereira '73, Kaufman '75, Bolstad '77 `varpro.f`, ..., O'Leary–Rust '13 `varpro.m`].
- For matrix structured distance to singularity: Structured low-rank approximation [Manton–Mahony–Hua '03, Markovsky–Willems–Van Huffel–De Moor '05, Markovsky–Usevich '13, '14, '17 `slra.m`]
- Nearest stable/unstable matrix: [Noferini–Poloni '21] and this work

An instructive example

$$f(A) = \min_{\Delta \text{ diagonal}} \|\Delta\|_F^2 \text{ s.t. } A + \Delta \text{ singular}, \quad A = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}.$$

Fix $\mathbf{v} = \left[\begin{smallmatrix} v_1 \\ v_2 \end{smallmatrix}\right]$ s.t. $(A + \Delta)\mathbf{v} = \mathbf{0}$, set $\Delta = \left[\begin{smallmatrix} \delta_1 & 0 \\ 0 & \delta_2 \end{smallmatrix}\right]$. The problem becomes

$$f(A, \mathbf{v}) = \min \delta_1^2 + \delta_2^2 \text{ s.t. } \underbrace{\begin{cases} (1 + \delta_1)v_1 + v_2 = 0, \\ (2 + \delta_2)v_2 = 0. \end{cases}}_{M\boldsymbol{\delta} = \mathbf{r}}$$

When $v_1, v_2 \neq 0$, we get $f(A, \mathbf{v}) = \|M^+\mathbf{v}\|^2 = \left(1 + \frac{v_2}{v_1}\right)^2 + 2^2$. We obtain $f(A)$ by minimizing over $\mathbf{v}$.

However, we cannot ignore the cases $v_1, v_2 = 0$ (i.e., when $\operatorname{rk} M$ drops). When $v_2 = 0$, the second equation no longer forces $\delta_2 = -2$, and there we have the global minimum $f_{[1,0]^T}(A) = 1$.

An instructive example

$$f(A) = \min_{\Delta \text{ diagonal}} \|\Delta\|_F^2 \text{ s.t. } A + \Delta \text{ singular}, \quad A = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}.$$

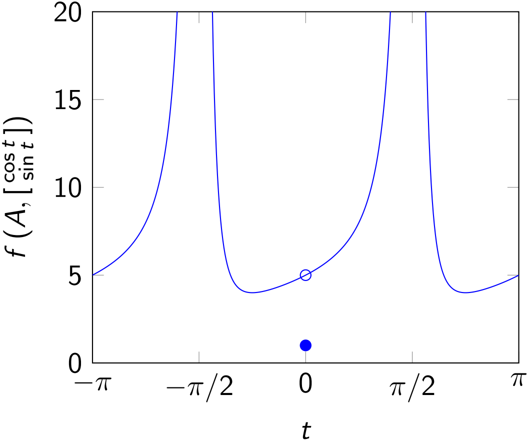
Fix $\mathbf{v} = \left[\begin{smallmatrix} v_1 \\ v_2 \end{smallmatrix}\right]$ s.t. $(A + \Delta)\mathbf{v} = \mathbf{0}$, set $\Delta = \left[\begin{smallmatrix} \delta_1 & 0 \\ 0 & \delta_2 \end{smallmatrix}\right]$. The problem becomes

$$f(A, \mathbf{v}) = \min \delta_1^2 + \delta_2^2 \text{ s.t. } \underbrace{\begin{cases} (1+\delta_1)v_1 + v_2 = 0, \\ (2+\delta_2)v_2 = 0. \end{cases}}_{M\boldsymbol{\delta} = \mathbf{r}}$$

When $v_1, v_2 \neq 0$, we get $f(A, \mathbf{v}) = \|M^+\mathbf{v}\|^2 = \left(1 + \frac{v_2}{v_1}\right)^2 + 2^2$. We obtain $f(A)$ by minimizing over $\mathbf{v}$.

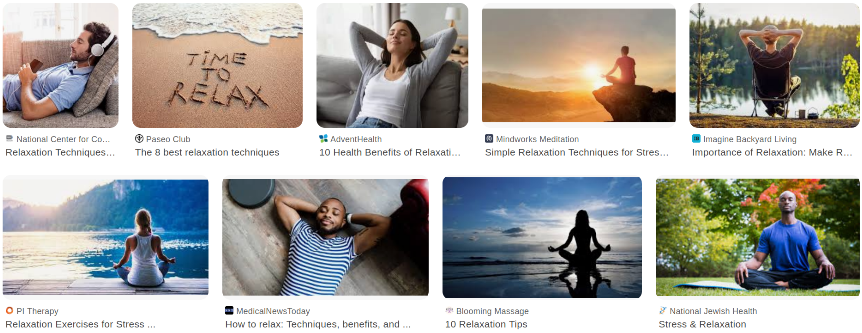
However, we cannot ignore the cases $v_1, v_2 = 0$ (i.e., when $\operatorname{rk} M$ drops). When $v_2 = 0$, the second equation no longer forces $\delta_2 = -2$, and there we have the global minimum $f_{[1,0]^T}(A) = 1$.

A problematic objective function



How can we optimize a discontinuous function? By regularizing it, i.e., relaxing the constraint.

Relaxation

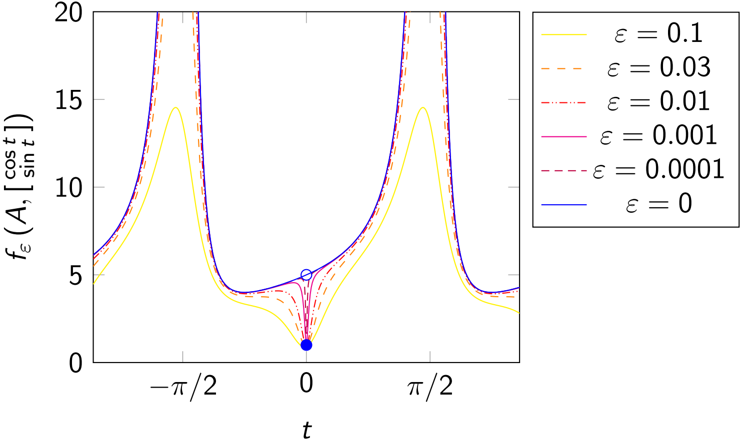


$$f(A) = \min_{\Delta \text{ structured}} \|\Delta\|_F^2 \text{ s.t. } (A+\Delta)\mathbf{v} = \mathbf{0}$$

$$\Downarrow$$

$$f_\varepsilon(A) = \min_{\Delta \text{ structured}} \|\Delta\|_F^2 + \varepsilon^{-1} \|(A+\Delta)\mathbf{v}\|^2$$

Constraint relaxation



We replace $f(A, \mathbf{v})$ with $f_\varepsilon(A, \mathbf{v}) = \min_\Delta \|\Delta\|_F^2 + \varepsilon^{-1}\|(A + \Delta)\mathbf{v}\|^2$.

When $\varepsilon \to 0$, this continuous function converges to $f(A, \mathbf{v})$.

Riemannian penalty method

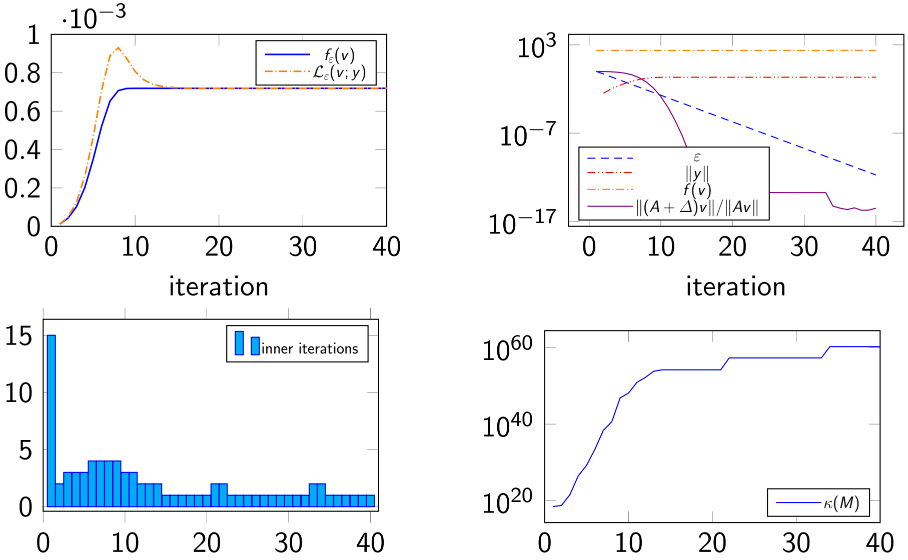
Solution strategy: penalty method.

1. Set ε to an initial value, e.g., $\varepsilon = 1$;
2. Find $\mathbf{v}_* = \arg\min f_\varepsilon(A, \mathbf{v})$;
3. Reduce ε, e.g., $\varepsilon \leftarrow \frac{1}{2}\varepsilon$;
4. Find $\mathbf{v}_* = \arg\min f_\varepsilon(A, \mathbf{v})$, using the previous $\mathbf{v}_*$ as initial value.
5. Repeat until ε is sufficiently small.

Remarks:

- A better variant: augmented Lagrangian method (cfr. ADMM)
- At each step, optimize on the sphere $\|\mathbf{v}\| = 1$ [Manopt].
- In practice, 0–2 iterations for most values of ε.
- Nice formula: $f_\varepsilon(A, \mathbf{v}) = \mathbf{r}^*(MM^* + \varepsilon I)^{-1}\mathbf{r}$.
- If the structure is sparsity, MM^* is diagonal.

Orani678: sparse structure, $n = 2529$, nnz $= 90158$.

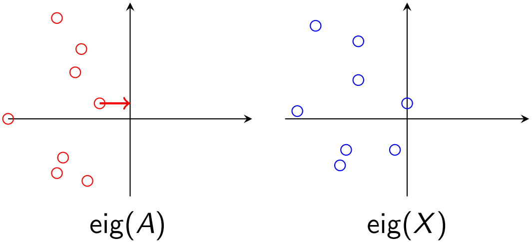


From nearest-singular to nearest-unstable

A more general problem: given a region $\Omega \subset \mathbb{C}$, find the nearest (structured) matrix to A with at least one eigenvalue in Ω.

$$f(A, \Omega) = \min_{\Delta \text{ structured}} \|\Delta\|_F^2 \text{ s.t. } A + \Delta \text{ has an eigenvalue in } \Omega.$$

E.g., set $\Omega =$ right-half plane for the nearest non-Hurwitz matrix problem.



If the oracle gives us A and $\lambda \in \Omega$, we reduce to the previous problem:

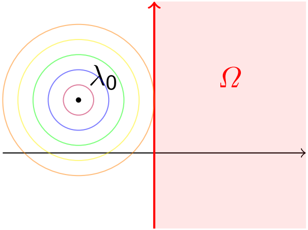
$$f(A, \Omega, \mathbf{v}, \lambda) = f(A - \lambda I, \mathbf{v}).$$

$$f(A, \Omega) = \min_{\mathbf{v}} \min_{\lambda \in \Omega} f(A, \Omega, \mathbf{v}, \lambda).$$

Double variable projection

However, the problem in λ is also easy to solve:

$$f(A - \lambda I, \mathbf{v}) = \alpha|\lambda - \lambda_0|^2 + \gamma \quad \text{for certain } \alpha, \gamma > 0, \lambda_0 \in \mathbb{C}.$$



So $\lambda_* = \text{projection}(\lambda_0, \Omega)$.

After eliminating Δ and λ, we are left again with a nonlinear problem in $\mathbf{v}$.

$$f(A, \Omega) = \min_{\mathbf{v} \in \mathbb{R}^n \setminus \{\mathbf{0}\}} f(A, \Omega, \mathbf{v}).$$

Example

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}, \quad \Delta = \begin{bmatrix} \delta_1 & 0 \\ 0 & \delta_2 \end{bmatrix}.$$

We assume $\lambda, \mathbf{v}$ real to ease notation.

$$\begin{aligned} f_\varepsilon(A - \lambda I, \mathbf{v}) &= \min_{\Delta} \|\Delta\|_F^2 + \varepsilon^{-1} \|(A + \Delta - \lambda I) v\|^2 \\ &= \min_{\delta_1, \delta_2} \delta_1^2 + \delta_2^2 + \varepsilon^{-1} \underbrace{\left(((1 + \delta_1 - \lambda) v_1 + v_2)^2 + ((2 + \delta_2) v_2)^2 \right)}_{= \|M\boldsymbol{\delta} - \mathbf{r}\|^2} \\ &= \underbrace{\begin{bmatrix} \lambda v_1 - (v_1 + v_2) \\ \lambda v_2 - 2 v_2 \end{bmatrix}^* \begin{bmatrix} v_1^2 + \varepsilon & \\ & v_2^2 + \varepsilon \end{bmatrix}^{-1} \begin{bmatrix} \lambda v_1 - (v_1 + v_2) \\ \lambda v_2 - 2 v_2 \end{bmatrix}}_{= \mathbf{r}^* (MM^* + \varepsilon I)^{-1} \mathbf{r}} \\ &= \alpha |\lambda - \lambda_0|^2 + \gamma, \quad \lambda_0 = \left(\frac{v_1^2}{v_1^2 + \varepsilon} + \frac{v_2^2}{v_2^2 + \varepsilon} \right)^{-1} \left(\frac{v_1 (v_1 + v_2)}{v_1^2 + \varepsilon} + \frac{2 v_2^2}{v_2^2 + \varepsilon} \right) \end{aligned}$$

Application: `Orani678`, Ω = small disc

Comparing with [Guglielmi, Lubich, Sicilia '23], who solve this problem as a substep to compute distance from singularity.

$\varepsilon = \|\Delta\|_F$	$\phi(\varepsilon) = \|\lambda_{\min}(A+\Delta)\|^2$
0.0000000	$1.1019564 \cdot 10^{-2}$
0.0176409	$9.5284061 \cdot 10^{-4}$
0.0219541	$2.5263758 \cdot 10^{-4}$
0.0243116	$6.5050153 \cdot 10^{-5}$
0.0261739	$4.1561289 \cdot 10^{-6}$
0.0264097	$1.6503282 \cdot 10^{-6}$
0.0264923	$1.0428313 \cdot 10^{-6}$
0.0266524	$2.6118300 \cdot 10^{-7}$
0.0267327	$6.5355110 \cdot 10^{-8}$
0.0267822	$9.6346192 \cdot 10^{-9}$
0.0268089	$1.7293467 \cdot 10^{-10}$
0.0268131	$6.1232190 \cdot 10^{-32}$

Application: approximate polynomial GCD

Nearest singular matrix with the same structure as the Sylvester matrix

$$S = \begin{bmatrix} a_0 & & & & b_0 & & & \\ \vdots & a_0 & & & \vdots & b_0 & & \\ a_d & \vdots & \ddots & & b_d & \vdots & \ddots & \\ & a_d & & a_0 & & b_d & & b_0 \\ & & \ddots & \vdots & & & \ddots & \vdots \\ & & & a_d & & & & b_d \end{bmatrix}.$$

Slightly worse than UVGCD, FastGCD on [Boito thesis, 8.1.1]:

d	$\|(a - gu, b - gw)\|$
9	$3.996 \cdot 10^{-3}$
8	$1.729 \cdot 10^{-4}$
7	$7.089 \cdot 10^{-6}$
6	$1.829 \cdot 10^{-7}$
5	$4.4\underline{91} \cdot 10^{-9}$
4	$\underline{*}$

Conclusions

- Variable projection / two-level optimization / "oracles": a technique that is useful also for spectral distance problems.
- Regularization is necessary to avoid missing global minima.
- More applications: polynomial GCD, nearest-singular matrix polynomial, etc.

Thank you for your attention!

Conclusions

- Variable projection / two-level optimization / "oracles": a technique that is useful also for spectral distance problems.
- Regularization is necessary to avoid missing global minima.
- More applications: polynomial GCD, nearest-singular matrix polynomial, etc.

Thank you for your attention!