Randomized Methods for Joint Eigenvalue Problems

Daniel Kressner

Institute of Mathematics, EPFL

Based on joint work with:

Haoze He (EPFL), Artemis Pados (MIT), and Bor Plestenjak (Ljubljana)

Linear Algebra 101

Symmetric $A_1, \ldots, A_d \in \mathbb{R}^{n\times n}$ **commute pairwise** if and only if there is orthogonal $Q \in \mathbb{R}^{n\times n}$ that **jointly diagonalizes** them:

$$Q^T A_1 Q = \begin{bmatrix} \diagdown \end{bmatrix}, \; \ldots, \; Q^T A_d Q = \begin{bmatrix} \diagdown \end{bmatrix}.$$

Questions one might ask:

- How does one actually compute Q?
- Do commuting matrices ever show up in practice / applications?
- Why is randomization important?

Linear Algebra 101

Symmetric $A_1, \ldots, A_d \in \mathbb{R}^{n \times n}$ **commute pairwise** if and only if there is orthogonal $Q \in \mathbb{R}^{n \times n}$ that **jointly diagonalizes** them:

$$Q^T A_1 Q = \begin{bmatrix} \diagdown \end{bmatrix}, \ \ldots, \ Q^T A_d Q = \begin{bmatrix} \diagdown \end{bmatrix}.$$

Structure of talk:

1. Motivating example:
   Roots of polynomials
2. Detailed picture:
   Randomized joint diagonalization of symmetric commuting matrices
3. Further extensions and applications
4. Back to the roots

Roots of polynomials

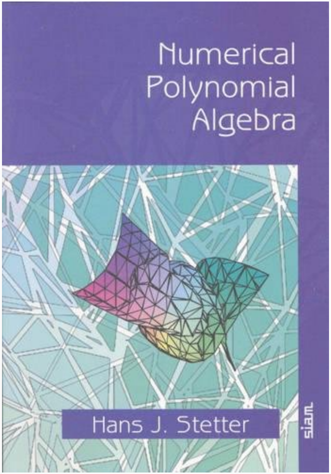
Polynomial roots as eigenvalue problems

1 polynomial in 1 variable:

1. Build linearization, e.g., companion matrix
2. Solve matrix eigenvalue problem

m polynomials in m variables:

1. Build m multiplication matrices
2. Solve **joint** matrix eigenvalue problem



Construction of multiplication matrices

Example:

$$p_1(x, y) = x^2 - 1 = 0, \qquad p_2(x, y) = y^2 - 1 = 0.$$

Step 1: Build basis for *quotient algebra*

$$A = \mathbb{C}[x, y]/I$$

with ideal $I = \langle x^2 - 1,\ y^2 - 1 \rangle$. Because $x^2 \equiv 1,\ y^2 \equiv 1$, can choose

$$B = \{1,\ x,\ y,\ xy\}.$$

General construction involves Gröbner bases or Sylvester-like / Macaulay-like resultant matrices.

Construction of multiplication matrices

Step 2: Consider multiplication by x and y:

$$L_x : A \to A, \quad L_x : p(x,y) \mapsto x \cdot p(x,y),$$

$$L_y : A \to A, \quad L_y : p(x,y) \mapsto y \cdot p(x,y).$$

Find matrix representation with respect to B:

$$M_x = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad M_y = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

Key property: $M_x M_y = M_y M_x$
(reflects that multiplication by x and y commute).

Eigenvalues of multiplication matrices

Slightly nontrivial: A common root (λ_x, λ_y) of p_1, p_2 is a joint eigenvalue

$$M_x v = \lambda_x v, \quad M_y v = \lambda_y v,$$

with joint eigenvector v (obtained from evaluating B in $x = \lambda_x$, $y = \lambda_y$).

Problem:

$$M_x = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad M_y = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

have double eigenvalues $\rightsquigarrow$

- v not well-defined by individual matrices
- difficult to pair eigenvalues of M_x and M_y

Eigenvalues of multiplication matrices

Simple trick: Choose $\mu_x, \mu_y \sim \mathcal{N}(0,1)$ and form

$$M = \mu_x M_x + \mu_y M_y.$$

Eigenvalues of M are simple a.s.!

Rayleigh quotients with each of the four eigenvectors v of M recover common roots:

$$\lambda_x = v^T M_x v, \quad \lambda_y = v^T M_y v.$$

Idea has been around for long time in root finding [Corless et al.'1997].

Difficulties encountered in general case:

- Multiplication matrices are generally nonsymmetric
- Multiplication matrices quickly become huge
- Joint eigenvector method only works well in generic situations (simple common roots / joint eigenvalues)
- Numerical stability issues [Graf/Townsend'2026]

Joint diagonalization of commuting symmetric matrices

Nearly commuting matrices

Linear Algebra 101 task

Given commuting symmetric matrices $A_1, \ldots, A_d$, find orthogonal matrix Q that **jointly diagonalizes** family:

$$Q^T A_1 Q = \begin{bmatrix} \diagdown \end{bmatrix}, \ \ldots, \ Q^T A_d Q = \begin{bmatrix} \diagdown \end{bmatrix}.$$

Problem: Not well-posed in finite-precision arithmetic. Roundoff destroys commutativity. Moreover, in many applications, commutativity is just an idealistic model assumption.

A, B are *nearly commuting* if $\exists \ \|\triangle A\|_2 \approx 0$, $\|\triangle B\|_2 \approx 0$ such that $A + \triangle A, B + \triangle B$ commute.

Nearly commuting matrices

Robustified Linear Algebra 101 task

Given *nearly* commuting symmetric matrices $A_1, \dots, A_d$ find orthogonal matrix Q that *nearly* diagonalizes $A_1, \dots, A_d$.

Applications beyond root finding:

- Blind Source Separation: Separation of source signal by joint diagonalization of cumulant tensor slices [Cardoso/Soloumiac'1993], covariance matrices [Pham/Cardoso'2001], autocorrelation matrices [Belouchrani et al.'1997].
- Parameter estimation in latent variable models through orthogonal decomposition of moment tensors = joint diagonalization of slices [Anandkumar et al.'2014]
- Manifold learning [Eynard et al.'2015]
- Parameter identification [Ehler et al.'2019]
- Computer graphics [Kovantsky et al.'2013]

Existing optimization-based approaches

Most numerical algorithms for joint diagonalization based on optimization:

$$\min_{Q^T Q = I} \mathcal{L}(Q),$$

where loss function $\mathcal{L}$ measures distance from being diagonal.

Popular choices:

- General symmetric $A_1, \ldots, A_d$:

$$\mathcal{L}(Q) := \sum_{k=1}^{d} \left\| \mathrm{offdiag}(Q^T A_k Q) \right\|_F^2,$$

  where offdiag extracts off-diagonal part of a matrix.
- Symmetric positive definite $A_1, \ldots, A_d$:

$$\mathcal{L}(Q) := \frac{1}{2n} \sum_{k=1}^{d} [\log \det \mathrm{diag}(Q^T A_k Q) - \log \det(Q^T A_k Q)],$$

  KL divergence [Pham'2001]

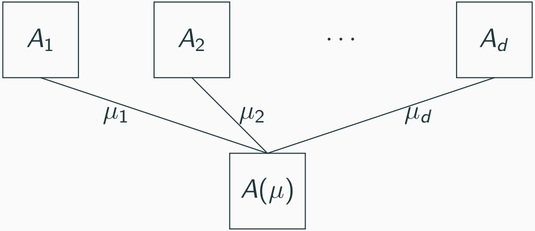
Existing optimization-based approaches

Optimization algorithms:

- Jacobi-like algorithms = coordinate descent [Bunse-Gerstner et al.'1993]
- Block coordinate descent [Pham'2001]
- Quasi-Newton methods [Ziehe et al.'2003], [Ablin et al.'2018]
- Riemannian gradient descent [Afsari/Krishnaprasad'2004]
- Riemannian trust region [Absil/Gallivan'2006]
- Riemannian Newton [Alyani et al.'2017]
- ...

Randomized approach

Sketched family: $A(\mu) = \mu_1 A_1 + \mu_1 A_2 + \cdots + \mu_d A_d$ for $\mu \sim \mathcal{N}(0, I_d)$.



Simple idea: Obtain Q from diagonalizing symmetric matrix $A(\mu)$.

Proposed and applied several times in the literature:

- Parameter identification problems [Ehler et al.'2019]
- Joint eigenvector method for root finding [Corless et al.'1997]
- Orthogonal decomposition of tensors [Anandkumar et al.'2012/2014]
- CP and related decompositions for tensors [Evert et al.'2022], [Telen/Vannieuwenhoven'2021], . . .
- . . .

How does randomness help?

Assume exact commutativity and $n = 2$.

Two cases:

- Every matrix $A_1,\ A_2,\ ,\dots,\ A_d$ has double eigenvalues:

$$A_1 = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_1 \end{pmatrix},\ A_2 = \begin{pmatrix} \lambda_2 & 0 \\ 0 & \lambda_2 \end{pmatrix},\ \dots\ , A_d = \begin{pmatrix} \lambda_d & 0 \\ 0 & \lambda_d \end{pmatrix}.$$

  Any orthogonal matrix Q will do.
- At least one matrix A_k has two simple eigenvalues $\rightsquigarrow$

$$A(\mu) = \mu_1 A_1 + \mu_2 A_2 + \cdots + \mu_d A_d$$

  has two simple eigenvalues a.s. and, in turn, Q is essentially uniquely defined and can be obtained from diagonalizing $A(\mu)$.

New contribution [He/DK'2024]: Analysis when $A_1, \dots, A_d$ *nearly* commute.

Note: Previous analyses either assume exact commutativity [Anandkumar et al.'2014] or sufficiently large eigenvalue gaps [Ehler et al.'2019].

Idea of analysis for nearly commuting family

$A_1, \dots, A_d$ *nearly* commuting $\rightsquigarrow$ $\exists$ commuting $D_1, \dots, D_d$ s.t.

$$A_k = D_k + E_k, \quad \|E_1\|_F^2 + \cdots + \|E_d\|_F^2 \leq \varepsilon^2.$$

Let x be joint eigenvector for simple (joint) eigenvalue $(\lambda_1, \dots, \lambda_d)$ of $D_1, \dots, D_k$. Then

$$\tilde{x} = x + (\lambda(\mu)I - D(\mu))^\dagger E(\mu)x + O(\varepsilon^2).$$

is corresp. eigenvector of $A(\mu)$ [Stewart/Sun'1990, Greenbaum/Li/Overton'2020].

Residual measures whether $\tilde{x}$ is a good eigenvector for A_k:

$$r_k = \lambda_k \tilde{x} - A_k \tilde{x}.$$

Residuals determine off-diagonal part in $Q^T A_k Q$.

$$\|r_k\|_2 \lesssim \| \underbrace{(\lambda_k I - D_k)(\lambda(\mu)I - D(\mu))^\dagger}_{\text{wlog diagonal}} \|_2 \|E(\mu)\|_2$$

Idea of analysis for nearly commuting family

Norm of diagonal matrix

$$(\lambda_1 I - D_1)(\lambda(\mu) I - D(\mu))^\dagger$$

measures (asymptotically) magnification of input error. Given another joint eigenvalue $(\xi_1, \ldots, \xi_d)$, diagonal entry takes the form

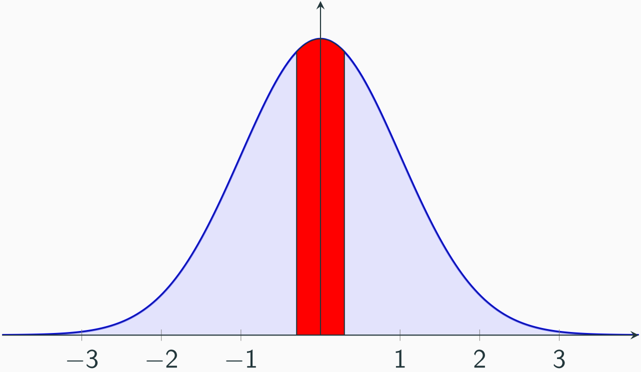
$$\begin{aligned} & \frac{\lambda_1 - \xi_1}{\mu_1(\lambda_1 - \xi_1) + \mu_2(\lambda_1 - \xi_2) + \cdots + \mu_d(\lambda_1 - \xi_d)} \\ = \quad & \frac{1}{\mu_1 + \mu_2 \frac{\lambda_1 - \xi_2}{\lambda_1 - \xi_1} + \cdots + \mu_d \frac{\lambda_1 - \xi_d}{\lambda_1 - \xi_1}} = \langle \mu, v \rangle^{-1} \end{aligned}$$

Important observations:

- Cancellation of (small/problematic) gap $\lambda_1 - \xi_1$
- $\langle \mu, v \rangle \sim \mathcal{N}(0, \|v\|_2^2)$ and $\|v\|_2 \geq 1$

Anticoncentration

pdf of $\langle \mu, v \rangle$ for Gaussian random μ



$$\mathbb{P}\Big\{ -R \leq \langle \mu, v \rangle^{-1} \leq R \Big\} \sim R^{-1}$$

Small probability that $|\langle \mu, v \rangle|$ is large!

Idea of analysis for nearly commuting family

Summary:

- Residual has norm $\leq O(R\varepsilon)$ with probability $\geq 1 - R^{-1}$.
- Union bound: Off-diagonal parts of $Q^T A_1 Q, \ldots, Q^T A_d Q$ have norm $\leq O(R\varepsilon)$ with probability $\geq 1 - R^{-1}$

BUT:

- Analysis assumes simple eigenvalues.
- Neglected higher-order terms in asymptotic perturbation result critically depend on eigenvalue gaps.

Solution:

- Cluster eigenvalues and consider invariant subspaces belonging to each cluster.
- Use non-asymptotic perturbation results for invariant subspaces [Stewart'1971, Karow/DK'2014]

Robust recovery

Robust recovery of Q **[He/DK'2024]**

Suppose that $A_1, \ldots, A_d$ is ε-near to a commuting family and Q diagonalizes $A(\mu) = \mu_1 + \cdots + \mu_d$ for $\mu \sim N(0, I_n)$. Then:

Q diagonalizes each matrix $A_1, \ldots, A_d$ up to error $O(R\varepsilon)$

with probability at least $1 - R^{-1}$.

+ Universal result. No assumptions on eigenvalue gaps.
− Constant in $O(R\varepsilon)$ proportional to $n^{3.5}$.

DRJD = Increased accuracy by combining several samples through successive deflation.

Experiments for synthetic data

10 commuting randomly generated 100×100 symm pos def matrices
only perturbed by roundoff error

	Time (in msec)	Offdiag error
QNDIAG	400	9.6×10^{-10}
FFDIAG	3600	8.3×10^{-11}
JADE	2200	2.6×10^{-7}
RJD	20	2.8×10^{-11}
DRJD	25	9.9×10^{-12}

QNDIAG Quasi-Newton method for Pham's cost function [Ablin et al.'2018].

FFDIAG Newton-like method for standard cost function [Ziehe et al.'2003].

JADE Jacobi method [Bunse-Gerstner et al.'1993], [Cardoso/Souloumiac'1996].

RJD/DRJD Randomized algorithms (3 trials for RJD).

Experiments for synthetic data

10 commuting randomly generated 100×100 symm pos def matrices perturbed by $\approx 10^{-5}$

	Time (in msec)	Offdiag error
QNDIAG	290	1.1×10^{-5}
FFDIAG	3400	9.5×10^{-6}
JADE	2070	9.5×10^{-6}
RJD	20	8.1×10^{-4}
DRJD	310	1.3×10^{-5}

QNDIAG Quasi-Newton method for Pham's cost function [Ablin et al.'2018].

FFDIAG Newton-like method for standard cost function [Ziehe et al.'2003].

JADE Jacobi method [Bunse-Gerstner et al.'1993], [Cardoso/Souloumiac'1996].

RJD/DRJD Randomized algorithms (3 trials for RJD).

Experiments for synthetic data

10 commuting randomly generated 100×100 symm pos def matrices perturbed by $\approx 10^{-1}$

	Time (in msec)	Offdiag error
QNDIAG	350	1.1×10^{-1}
FFDIAG	3830	9.5×10^{-2}
JADE	2530	9.5×10^{-2}
RJD	20	1.94
DRJD	318	1.4×10^{-1}

QNDIAG Quasi-Newton method for Pham's cost function [Ablin et al.'2018].

FFDIAG Newton-like method for standard cost function [Ziehe et al.'2003].

JADE Jacobi method [Bunse-Gerstner et al.'1993], [Cardoso/Souloumiac'1996].

RJD/DRJD Randomized algorithms (3 trials for RJD).

Fun application: Diagonalizing normal matrices

Linear Algebra 101

Square matrix A is normal iff its Hermitian part $(A + A^*)/2$ and skew-Hermitian part $(A - A^*)/2$ commute.

Structure-exploiting approach for diagonalizing normal matrix:

```
[U,~] = eig(randn*(A+A')+1i*randn*(A-A'));
D = diag(diag(U'*A*U));
```

Requires 0.55 seconds and attains residual 3×10^{-11} for 1000×1000 unitary matrix.

Versus unstructured approach:

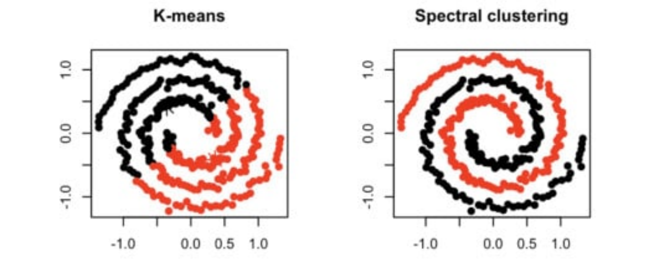
```
[U,S] = schur(A); D = diag(diag(S));
```

Requires 3.4 seconds and attains residual 3×10^{-14}.

Partial joint diagonalization

Spectral clustering

Spectral clustering [von Luxburg'2007] uses eigenvectors belonging to smallest k nonzero eigenvalues of graph Laplacian L to cluster data.



Eigenvectors $X \in \mathbb{R}^{n \times k}$ belonging to smallest k eigenvalues minimizes

$$\mathrm{trace}(X^T L X)$$

among all matrices satisfying $X^T X = I$, $X^T \mathbf{1} = \mathbf{0}$.

Multi-modal spectral clustering

Graph Laplacians $L_1, \ldots, L_d$ capturing different modalities of the same data. Taking different modalities into account can improve robustness of clustering.

Existing multi-modal spectral clustering methods [Dodero et al.'2013, Eynard et al.'2015, Khachatrian et al.'2021] aim at computing shared approximate basis of eigenvectors across modalities. Computed through optimization-based joint diagonalization of $L_1, \ldots, L_d$.

Measure for quality of joint embedding:

$$s(X) = \left\| \left[\text{trace}(X^T L_1 X), \ldots, \text{trace}(X^T L_d X) \right] \right\|_\infty.$$

Multi-modal spectral clustering

Theorem [He/Pados/DK'2025]

Letting Δ^{d-1} denote standard d-dimensional simplex, it holds that

$$\min_{\substack{X^T\mathbf{1}=\mathbf{0}\\ X^TX=I_k}} s(X) = \max_{\mu\in\Delta^{d-1}} \sum_{i=1}^{k} \lambda_i(L(\mu)).$$

Proof via duality.

Extends result for $k=1$ by [Coifman/Marshall/Steinerberger'2023].

$\rightsquigarrow$ Simple randomized method:

- Sample several random $\mu \in \Delta^{d-1}$.
- Choose $L(\mu) = \mu_1 L_1 + \cdots + \mu_d L_d$ that has largest sum of k smallest eigenvalues.
- Return X = eigenvectors belonging to smallest k eigenvalues of $L(\mu)$.

Results on benchmark examples

Clustering performance in terms of normalized mutual information (NMI); higher is better:

Method	**Weighted SBM**	**Caltech-7**	**Digits**
Single Laplacians	0.640 (1) 0.512 (2) 0.624 (3) 0.659 (4)	0.158 (Gabor) 0.322 (Wavelet) 0.355 (Centrist) 0.421 (HOG) 0.341 (GIST) 0.507 (LBP)	0.665 (DCT) 0.607 (Patch)
RJD Average	0.711 ($\pm$0.012)	0.491 ($\pm$0.002)	0.650 ($\pm$0.001)
RJD-BASE	**0.803**	**0.531**	**0.665**
QN-Diag	0.743	0.285	0.627
QN-Diag (RJD-BASE init.)	0.743	0.274	0.627
JADE	0.773	0.415	0.650
JADE (RJD-BASE init.)	0.601	0.024	0.075
MVSC	0.737	0.476	0.661
CoReg-MVSC	0.688	0.431	0.679
MV-KMeans	–	–	0.489
MV-SphKMeans	–	–	0.528
Single-Dir. Smoothness	0.774	0.499	0.682
BASE Smoothness	**0.780**	**0.530**	**0.682**

RJD-BASE = random. method. (200 samples)

BASE Smoothness = projected GD applied to $s(X)$.

Simultaneous diagonalization by congruence

Simultaneous diagonalization by congruence (SDC)

SDC

Family of symmetric matrices $A_1, \dots, A_d \in \mathbb{R}^{n\times n}$ is SDC if there exists invertible X such that

$$X^T A_1 X = \begin{bmatrix} \diagdown \end{bmatrix}, \ \dots, \ X^T A_d X = \begin{bmatrix} \diagdown \end{bmatrix}.$$

[He/Nguyen'2022]: Family is SDC if and if there is invertible P such that $P^T A_1 P, \dots P^T A_d P$ commute pairwise.

Important special case: If $\exists \theta$ such that

$$A(\theta) = \theta_1 A_1 + \cdots + \theta_d A_d \quad \text{is positive definite,}$$

can choose $P = L^{-1}$ for Cholesky factorization $A(\theta) = LL^T$.

Simultaneous diagonalization by congruence (SDC)

Applications:

- SDC equivalent to tensor rank-n CP decomposition of third-order tensor with partial symmetries [Domanov/de Lathauwer'2015]
- Blind Source Separation [Chabriel et al.'2014]
- Remote sensing [Khachatrian et al.'2021]
- Quadratic Programming [Jiang/Li'2016]

All previous algorithms optimization-based:

- General case: Riemannian optimization to minimize off-diagonal norm over $\mathcal{OB}(n, n)$.
  Riemannian trust region [Absil/Gallivan'2006], Riemannian BFGS [Bouchard et al.'2020], . . .
- Positive definite case: Minimization of KL divergence.
  Jacobi-like [Pham'2001], Quasi-Newton [Ablin/Cardoso/Gramfort'2018], . . .

Randomized SDC (RSDC)

Randomized SDC (RSDC)

Choose *two* independent $\theta, \mu \sim \mathcal{N}(0, I_d)$ and form

$$A(\mu) = \mu_1 A_1 + \cdots + \mu_d A_d, \quad A(\theta) = \theta_1 A_1 + \cdots + \theta_d A_d.$$

Diagonalize matrix pencil

$$X^T A(\mu) X - \lambda X^T A(\theta) X = \begin{bmatrix} \diagdown \end{bmatrix} - \lambda \begin{bmatrix} \diagdown \end{bmatrix},$$

using, e.g., QZ algorithm (this may not be possible).

Exact recovery: If $A_1, \ldots, A_d$ is exactly SDC then X diagonalizes family by congruence a.s.

RSDC: Robust recovery

Robust recovery **[He/DK'2024]**

Suppose that $A_1, \ldots, A_d$ is ε-near to SDC and X diagonalizes $A(\mu) - \lambda A(\theta)$ for $[\mu, \theta] \sim N(0, I_{2n})$. Then:

X diagonalizes each matrix $A_1, \ldots, A_d$ up to error $O(R^2 \|X\|_2^2 \varepsilon)$

with probability at least $1 - R^{-1}$.

Simplifications when $A_1, \ldots, A_d$ are positive (semi-)definite:

- Choose $\theta_1 = \cdots = \theta_d = 1/d$.
- Diagonalization of $A(\mu) - \lambda A(\theta)$ via Cholesky of $A(\theta)$.
- Robust recovery guarantee improves to $O(R\|X\|_2^2 \varepsilon)$.

Application to BSS

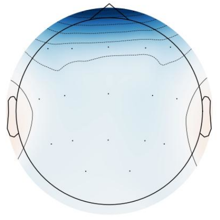
Blind Source Separation for EEG recordings [Congedo et al.'2014]:

Joint diagonalization of sample covariance matrices by *congruence*:

Randomized SDC [He/DK'2024]

+ quasi-Newton opt [Ziehe et al.'2004]

requires 3s vs. 280s for SOTA (Pham's Jacobi-like method) applied to eye-blinking benchmark, at comparable accuracy.



Back to the roots

Eigenvalues of multiplication matrices

Recall that computing roots of d-variate polynomials requires computing joint eigenvalues of (commuting) multiplication matrices:

$$M_{x_1},\ M_{x_2},\ \dots\ , M_{x_d}.$$

Randomized approaches based on linear combination

$$M(\mu) = \mu_1 M_{x_1} + \mu_2 M_{x_2} + \cdots + \mu_d M_{x_d}, \quad \mu \sim \mathrm{Unif}\big(\mathbb{S}_{\mathbb{C}}^{d-1}\big).$$

RQ1 Compute (right) eigenvectors v of $M(\mu)$ and obtain common roots from (one-sided) Rayleigh quotients:

$$\lambda_1 = \frac{v^T M_{x_1} v}{v^T v}, \lambda_2 = \frac{v^T M_{x_2} v}{v^T v}, \dots, \lambda_d = \frac{v^T M_{x_1} v}{v^T v}.$$

RQ2 Compute right *and* left eigenvectors v, w of $M(\mu)$ and obtain common roots from (two-sided) Rayleigh quotients:

$$\lambda_1 = \frac{w^T M_{x_1} v}{w^T v}, \lambda_2 = \frac{w^T M_{x_2} v}{w^T v}, \dots, \lambda_d = \frac{w^T M_{x_1} v}{w^T v}.$$

Results for randomized approaches

Robust computation of simple roots **[He/DK/Plestenjak'2025]**

Let $\lambda = (\lambda_1, \ldots, \lambda_d)$ be simple common root of d-variate polynomial system. Suppose that $\tilde{\lambda}^{\mathrm{RQ1}}$ is computed from applying RQ1 to multiplication matrices perturbed by error of norm $\leq \varepsilon$. Then

$$\left\|\lambda - \lambda^{\mathrm{RQ1}}\right\| \lesssim R \cdot \mathrm{cond}(\lambda) \cdot \varepsilon + O(\varepsilon^2)$$

holds with probability $1 - R^{-2}$.

- Improved tail behavior R^{-2} thanks to using complex random numbers: $\mu \sim \mathrm{Unif}\left(\mathbb{S}_{\mathbb{C}}^{d-1}\right)$.
- $\mathrm{cond}(\lambda)$: condition number of λ as a joint eigenvalue of multiplication matrices.
- Result extends to semi-simple roots.
- Analysis of RQ2 more complicated / more pessimistic, although RQ2 performs better in practice.

Results for randomized approaches

	random	rose	katsura7	redeco8
Rschur	1.31e-11	1.98e-08	1.74e-10	3.09e-12
schur	5.58e-12	1.12e-08	8.19e-10	2.75e-12
RQ2	9.72e-14	9.26e-09	1.61e-11	3.52e-12

Accuracy of different methods for computing common roots of four standard benchmark examples.

Rschur Common roots via Schur decomposition of randomized linear combination of multiplication matrices [Telen/Van Barel'2018].

schur Common roots via Schur decomposition of first multiplication matrix [Vermeersch/de Moor'2023]

RQ2 Our method.

Timing nearly identical as compt of multiplication matrices expensive.

Conclusions

Conclusions

- Randomization turns complicated joint eigenvalue problems into standard eigenvalue problems that can be solved with off-the-shelve software.
- Randomization yields highly competitive performance (time, accuracy) in a wide range of applications.
  For joint diagonalization tasks, randomization replaces optimization-based approaches or delivers excellent initial iterates.
- Analysis based on (lots of) perturbation analysis + anticoncentation.

H. He and DK. *Randomized joint diagonalization of symmetric matrices*. SIAM J. Matrix Anal. Appl., 45(1):661–684, 2024.

H. He and DK. *A simple, randomized algorithm for diagonalizing normal matrices*. Calcolo 62:30, 2025.

H. He and DK. *A randomized algorithm for simultaneously diagonalizing symmetric matrices by congruence*. arXiv:2402.16557, 2024.

H. He, DK, B. Plestenjak. *Randomized methods for computing joint eigenvalues, with applications to multiparameter eigenvalue problems and root finding*, Numer. Algor. 100, 861–892, 2025.

H. He, A. Pados, DK. *RJD-BASE: Multi-modal spectral clustering via randomized joint diagonalization*. arXiv:2509.11981, 2025.