Fast algorithms from low-rank updates

Daniel Kressner

Institute of Mathematics / EPFL

daniel.kressner@epfl.ch
`http://anchp.epfl.ch`

Based on joint work with
Bernhard Beckermann, Alice Cortinovis, Michel Crouzeix,
Stefano Massei, Marcel Schweitzer.

June 22, 2021

Matrix functions

Given $A \in \mathbb{C}^{n\times n}$, $f : \Omega \to \mathbb{C}$ analytic on a domain Ω containing eigenvalues of A, matrix function $f(A) \in \mathbb{C}^{n\times n}$ defined as

$$f(A) = \frac{1}{2\pi i}\oint_\Gamma f(z)(zI - A)^{-1}\,\mathrm{d}z,$$

for a contour Γ encircling eigenvalues of A.

Special cases:

- Polynomial $f = \alpha_0 + \alpha_1 z + \cdots + \alpha_d z^d$:

$$f(A) = \alpha_0 I + \alpha_1 A + \cdots + \alpha_d A^d.$$

- A diagonalizable, $A = X \cdot \operatorname{diag}(\lambda_1, \dots, \lambda_n) \cdot X^{-1}$:

$$f(A) = X \begin{bmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_n) \end{bmatrix} X^{-1}$$

Examples

- $f(z) = 1/z$: Matrix inverse.
- $f(z) = \exp(z)$: Matrix exponential.

  Exponential integrators for ODEs / PDEs.
  Centrality measures in network science.
- $f(z) = \text{sign}(z)$: Matrix sign function.

  Eigenvalue problems.
  Riccati equations for optimal control.
  Lyapunov equations for model reduction.
- $f(z) = \sqrt{z}$: Matrix square root.

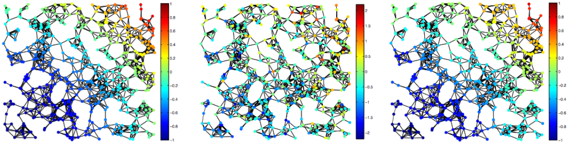
  Dirichlet-to-Neumann map.
  Affine stochastic processes. Fractional DEs.
- $f(z) = \log(z)$: Matrix logarithm.

  generators of continuous Markov chains.

  determinants: $\log\det(A) = \text{trace}(\log(A))$.
- $f(z) =$ rational. Matrix rational.

  Filter in graph signal processing.

Signal processing on graphs



Original Noisy Denoised

- Undirected weighted graph $G = (V, E, w)$ with n nodes (traffic network/social network/...).
- Graph Laplacian $L \in \mathbb{R}^{n \times n}$ with spectral decomp $L = U \Lambda U^T$.

Graph spectral filtering of signal $x \in \mathbb{R}^n$:

1. Graph Fourier transform: $\hat{x} \leftarrow U^T x$
2. Application of filter f in Fourier domain: $\hat{y}_i \leftarrow f(\lambda_i) \cdot y_i$
3. Graph inverse Fourier transform: $y \leftarrow U\hat{y}$

This is a matrix function: $y = f(L)x$ [Shuman et al.'2013].

`https://epfl-lts2.github.io/gspbox-html/`

Localized changes in nodes/edges = low-rank changes of L.

Fréchet derivatives

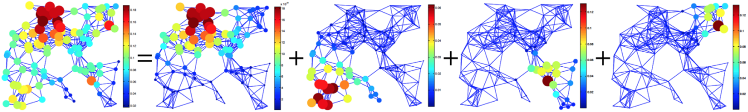
Fréchet derivative $Df\{A\} : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ of matrix function $f(A)$ defined via

$$f(A+E) = f(A) + Df\{A\}(E) + \mathcal{O}(\|E\|^2).$$

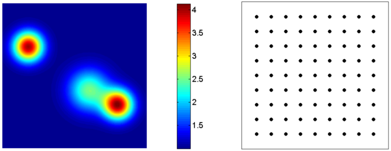
Evaluation required in optimization problems with matrix functions.

Examples:

- Learning heat diffusion graphs [Thanou et al.'2017].



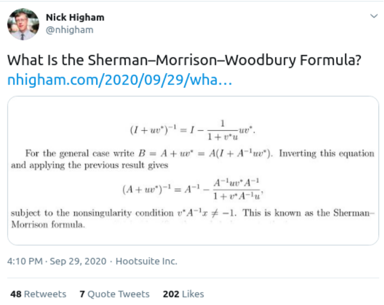
- Inverse parabolic problems



Sherman-Morrison for matrix functions?

Motivation: Sherman-Morrison formula

$$(A + bc^T)^{-1} = A^{-1} - \frac{1}{1 + c^T A^{-1} b} A^{-1} b c^T A^{-1}$$



Nick Higham
@nhigham

What Is the Sherman–Morrison–Woodbury Formula?
nhigham.com/2020/09/29/wha...

$$(I + uv^*)^{-1} = I - \frac{1}{1 + v^* u} uv^*.$$

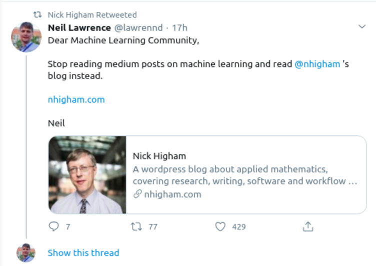
For the general case write $B = A + uv^* = A(I + A^{-1}uv^*)$. Inverting this equation and applying the previous result gives

$$(A + uv^*)^{-1} = A^{-1} - \frac{A^{-1}uv^*A^{-1}}{1 + v^*A^{-1}u},$$

subject to the nonsingularity condition $v^*A^{-1}x \neq -1$. This is known as the Sherman–Morrison formula.

4:10 PM · Sep 29, 2020 · Hootsuite Inc.

48 Retweets **7** Quote Tweets **202** Likes



Nick Higham Retweeted
Neil Lawrence @lawrennd · 17h
Dear Machine Learning Community,

Stop reading medium posts on machine learning and read @nhigham 's blog instead.

nhigham.com

Neil

Nick Higham
A wordpress blog about applied mathematics, covering research, writing, software and workflow ...
nhigham.com

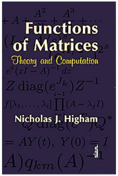
Sherman–Morrison formula for `expm`?

$$\exp(A + \text{rank-1}) = \exp(A) + \text{rank-1}$$

- would be nice
- important implications in time integration, (social) network analysis, signal processing on graphs, . . .

Sherman-Morrison for `expm`?

$A =$ identity matrix:



$$f(I + bc^T) = I + \frac{f(1 + c^T b) - f(1)}{c^T b} bc^T.$$

$f = \exp$ [Celledoni/Iserles'2000], general case [Higham'2008].

There is no Sherman-Morrison for general A but...

Quick Matlab test

```
>> A = -rand(200);
>> b = -rand(200,1); c = rand(200,1);
```

```
>> rank( expm(A+b*c') - expm(A) )
```

Quick Matlab test

```
>> A = -rand(200);
>> b = -rand(200,1); c = rand(200,1);

>> rank( expm(A+b⋆c’) - expm(A) )

ans = 11
```

Quick Matlab test

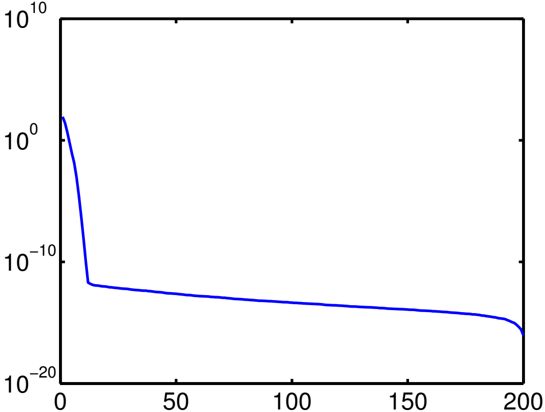
```
>> A = -rand(200);
>> b = -rand(200,1); c = rand(200,1);

>> rank( expm(A+b*c') - expm(A) )

ans = 16
```

Quick Matlab test

```
>> A = -rand(200);
>> b = -rand(200,1); c = rand(200,1);
>> svd( expm(A+b*c') - expm(A) )
```



Low-rank approximation

Schmidt-Mirsky/Eckart-Young:

$$\min\left\{\|A - CD^T\|_2 :\ C \in \mathbb{R}^{n\times r}, D \in \mathbb{R}^{m\times r}\right\} = \sigma_{r+1}.$$

with singular values $\sigma_1 \geq \cdots \geq \sigma_r \geq \sigma_{r+1} \geq \cdots$ of A.

- MATLAB's (numerical) rank:

$$\sigma_{r+1} \lesssim 10^{-16} \times \|A\|_2.$$

Quick singular value decay
$\rightsquigarrow$ A can be replaced by low-rank approximation

This talk: Construction, analysis, application of efficient methods for approximating $f(A + bc^T) - f(A)$.

Low-rank updates

Inspiration from Taylor expansion

$$\exp(A) = I + A + \frac{1}{2}A^2 + \frac{1}{3!}A^3 + \cdots$$

Truncating after third term ⇝

$$\begin{aligned} & \exp(A + bc^T) - \exp(A) \\ \approx \quad & I + A + bc^T + \frac{1}{2}(A + bc^T)^2 + \frac{1}{3!}(A + bc^T)^3 - I - A - \frac{1}{2}A^2 - \frac{1}{3!}A^3 \\ = \quad & \begin{bmatrix} b & Ab & A^2b \end{bmatrix} \times [\text{some } 3 \times 3 \text{ matrix}] \times \begin{bmatrix} c & A^Tc & (A^T)^2c \end{bmatrix}^T \end{aligned}$$

Difference contained in linear combinations of outer products of elements from $\operatorname{span}\{b, Ab, A^2b\}$ and $\operatorname{span}\{c, A^Tc, (A^T)^2c\}$.

Tensorized Krylov subspace

For

$$\begin{aligned}\mathcal{U}_k &= \mathcal{K}_k(A, c) = \mathrm{span}\{c, Ac, A^2c, \dots, A^{k-1}c\}\\ \mathcal{V}_\ell &= \mathcal{K}_\ell(B^T, d) = \mathrm{span}\{d, B^Td, (B^T)^2d, \dots, (B^T)^{\ell-1}d\}\end{aligned}$$

define

$$\mathcal{V}_\ell \otimes \mathcal{U}_k = \mathrm{span}\big\{v \otimes u : \ u \in \mathcal{U}_k, \ v \in \mathcal{V}_\ell\big\}.$$

In terms of matrices

$$\mathcal{V}_\ell \otimes \mathcal{U}_k = \mathrm{span}\big\{uv^T : \ u \in \mathcal{U}_k, \ v \in \mathcal{V}_\ell\big\}.$$

Plays prominent role in Krylov subspace methods for Lyapunov equations [Saad'1990, Jaimoukha/Kasenally'1994, Simoncini'2007, Benner/Truhar/Li'2009], tensor equations [K./Tobler'2010], functions of matrices with Kronecker structure [Benzi/Simoncini'2017], and divide&conquer for matrix equations [K./Massei/Robol'2018].

Tensorized Krylov subspace

Two useful characterizations:

1. Evaluate bivariate polynomial

$$p(x,y) = \sum_{i=0}^{k} \sum_{j=0}^{\ell} \alpha_{ij} x^i y^j \in \Pi_{k,\ell}$$

in commuting matrices $\mathcal{A} = I_n \otimes A$, $\mathcal{B} = B \otimes I_m$:

$$p(\mathcal{A},\mathcal{B}) = \sum_{i=0}^{k} \sum_{j=0}^{\ell} \alpha_{ij} \mathcal{A}^i \mathcal{B}^j .$$

$$\mathcal{V}_\ell \otimes \mathcal{U}_k = \big\{ p(\mathcal{A},\mathcal{B})(d \otimes c) : \; p \in \Pi_{k-1,\ell-1} \big\}$$

2. For two bases U_k, V_ℓ of $\mathcal{U}_k, \mathcal{V}_\ell$:

$$\mathcal{V}_\ell \otimes \mathcal{U}_k = \big\{ U_k X V_\ell^T \; : \; X \in \mathbb{R}^{k \times \ell} \big\}$$

Choice of $X_m(f)$

$$f(A + bc^T) - f(A) \approx U_m X_m(f) V_m^T$$

Natural to let $X_m(f)$ solve the compressed problem:

$$X_m(f) = f\big(U_m^T(A + bc^T)V_m\big) - f\big(U_m^T A V_m\big).$$

Choice of $X_m(f)$

$$f(A+bc^T)-f(A)\approx U_m X_m(f) V_m^T$$

Natural to let $X_m(f)$ solve the compressed problem:

$$X_m(f)=f\big(U_m^T(A+bc^T)V_m\big)-f\big(U_m^TAV_m\big).$$

Well, actually not!

- No polynomial exactness.
- No useful convergence (bounds).

A useful result

Lemma. Let

$$\mathcal{A} := \begin{bmatrix} A & bc^T \\ 0 & A + bc^T \end{bmatrix}.$$

Then

$$f(\mathcal{A}) = \begin{bmatrix} f(A) & f(A + bc^T) - f(A) \\ 0 & f(A + bc^T) \end{bmatrix}.$$

Sketch of proof. Use contour integration.

Choice of $X_m(f)$

Idea: Compress block matrix $\mathcal{A}$ instead of A and $A + bc^T$.

$$
\begin{aligned}
& f\left(\begin{bmatrix} U_m & 0 \\ 0 & V_m \end{bmatrix}^T \mathcal{A} \begin{bmatrix} U_m & 0 \\ 0 & V_m \end{bmatrix}\right) \\
= \; & f\left(\begin{bmatrix} U_m & 0 \\ 0 & V_m \end{bmatrix}^T \begin{bmatrix} A & bc^T \\ 0 & A + bc^T \end{bmatrix} \begin{bmatrix} U_m & 0 \\ 0 & V_m \end{bmatrix}\right) \\
= \; & f\left(\begin{bmatrix} G_m & \|b\|\|c\| e_1 e_1^T \\ 0 & H_m^T + \|c\| V_m^T b e_1^T \end{bmatrix}\right) =: \begin{bmatrix} f(G_m) & X_m(f) \\ 0 & f(H_m^* + \|c\| V_m^T b e_1^T) \end{bmatrix},
\end{aligned}
$$

with Hessenberg matrices $G_m = U_m^T A U_m$, $H_m = V_m^T A^T V_m$.

Theorem (Polynomial exactness)

$$p(A + bc^T) - p(A) = U_m X_m(p) V_m^T$$

for every polynomial p of degree at most m.

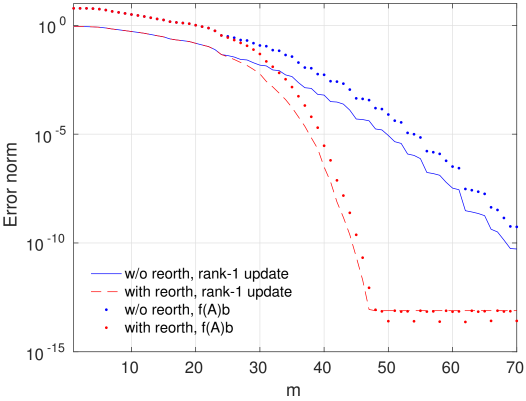
Algorithm

Krylov subspace approximation of $f(A+bc^*)-f(A)$:

```
Perform $m$ steps of the Arnoldi method to compute an orthonormal basis $U_m$ of $\mathcal{K}_m(A,b)$ and $G_m = U_m^* A U_m$.
Perform $m$ steps of the Arnoldi method to compute an orthonormal basis $V_m$ of $\mathcal{K}_m(A^*,c)$ and $H_m = V_m^* A^* V_m$.
Compute matrix function $F_m = f\left(\begin{bmatrix} G_m & \|b\|\|c\| e_1 e_1^* \\ 0 & H_m^* + \|c\| V_m^* b e_1^* \end{bmatrix}\right)$.
Set $X_m(f) = F_m(1:m, m+1:2m)$.
Return $U_m X_m(f) V_m^*$.
```

Example

f = expm, symmetric 100×100 matrix with logarithmically spaced eigenvalues in $[-10^3, 10^{-3}]$



Communicability measures of graphs

- Undirected graph $G = (V, E)$ with n nodes.
- Symmetric adjacency matrix $A \in \mathbb{R}^{n \times n}$.
- Subgraph centrality [Estrada/Rodríguez-Velázquez'2005] of ith node defined as

$$\frac{[\exp(A)]_{ii}}{\operatorname{trace}(\exp(A))}.$$

- Interesting to study impact of removing/adding edges on communicability [Arrigo/Benzi'2016].

Removing/adding edge $=$ symmetric rank-2 change of A.

Idea:

- Compute diagonal elements of $\exp(A)$ once in (expensive) offline computation, using Lanczos + quadrature [Golub/Meurant'2010]. Implementation in mmq toolbox by Meurant.
- Keep track of changes in $\operatorname{diag}(\exp(A))$ with tensorized Krylov subspace method.

Results

Compare:

update	Time needed for updating subgraph centrality of all nodes, when modifying 10 edges, using our method with accuracy 10^{-6}.
mmq	Time needed for recomputing subgraph centrality of all nodes from scratch *once*.

Network	# nodes	# edges	update	mmq
Gleich/Minnesota	2,642	6,606	0.15 s	1.80 s
Pajek/Erdos992	6,100	15,030	0.62 s	6.22 s
Pajek/USpowerGrid	4,941	13,188	0.48 s	4.96 s
SNAP/ca-HepTh	9,877	51,971	1.25 s	14.58 s
SNAP/email-Enron	36,692	367,662	1.96 s	147.47 s

Related work: S. Pozza and F. Tudisco. On the stability of network indices defined by means of matrix functions, SIMAX'2018.

Convergence analysis

Strategy

Aim: Get a priori bounds on approximation

$$f(A+bc^T)-f(A)\approx U_mX_m(f)V_m^T.$$

Strategy: Use polynomial exactness, replace f by $f-p$, and bound norms.

Important ingredient: Bounds of the form

$$\|f(A+D)-f(A)\|\lesssim\|D\|$$

Lipschitz constants for matrix functions?

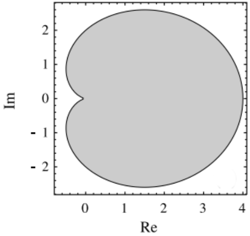
Via line integral (or other tricks), can be reduced to

$$\|Df\{A\}(E)\|\lesssim\|E\|$$

Crouzeix-Palencia for Fréchet derivative?

Norms of matrix functions

Numerical range
$\mathcal{W}(A) := \{x^T A x : \|x\|_2 = 1\}.$



Theorem [Crouzeix/Palencia'2017]. For function $f(x)$ analytic on domain Ω containing $\mathcal{W}(A)$:

$$\|f(A)\|_2 \leq (1+\sqrt{2}) \max_{x \in \mathcal{W}(A)} |f(x)|.$$

- 2021 SIAM Activity Group on Linear Algebra Best Paper Prize
- Crouzeix's Conjecture: Bound holds with 2 instead of $1+\sqrt{2}$

Untangling Noncommutativity with Operator Integrals

Anna Skripka

Bivariate matrix functions

Consider function $f(x, y)$ analytic on $\Omega_A \times \Omega_B \in \mathbb{C} \times \mathbb{C}$. Given $A \in \mathbb{C}^{m\times m}$, $B \in \mathbb{C}^{n\times n}$ with eigenvalues in Ω_A, Ω_B:

Bivariate matrix function $f\{A, B\}$ is linear operator on $\mathbb{C}^{m\times n}$ defined by

$$f\{A, B\}(C) = -\frac{1}{4\pi^2} \oint_{\Gamma_A} \oint_{\Gamma_B} f(x, y)(xI - A)^{-1} C (yI - B^T)^{-1} \, \mathrm{d}y \, \mathrm{d}x.$$

Given (univariate) matrix function $f(A)$ and finite difference quotient

$$f^{[1]}(x, y) := f[x, y] = \begin{cases} \frac{f(x)-f(y)}{x-y}, & \text{for } x \neq y, \\ f'(x), & \text{for } x = y, \end{cases}$$

Then $X = f^{[1]}\{A, A^T\}(C) = Df\{A\}(C)$ is Fréchet derivative of f at A in direction C.

Crouzeix/Palencia result for bivariate matrix functions

Constant $1+\sqrt{2}$ not achievable!

Example:

$$A = B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \qquad f(x,y) = xy.$$

$\rightsquigarrow \|f\{A,B\}\|_2 = 1$ but $|f(x,y)| \leq 1/4$ on $\mathcal{W}(A) \times \mathcal{W}(B)$.

Theorem [Crouzeix/K.'2020]. For function $f(x,y)$ analytic on $\Omega_A \times \Omega_B$ with $\mathcal{W}(A) \subset \Omega_A$, $\mathcal{W}(B) \subset \Omega_B$:

$$\|f\{A,B\}\|_2 \leq (1+\sqrt{2})^2 \sup_{x\in\mathcal{W}(A),y\in\mathcal{W}(B)} |f(x,y)|.$$

Corollaries [Beckermann/Cortinovis/K./Schweitzer'2020]

Bound on Fréchet derivative:

$$\|Df\{A\}\|_2 \leq (1+\sqrt{2})^2 \sup_{x \in \mathcal{W}(A)} |f'(x)|$$

Lipschitz constant for matrix functions:

$$\|f(A+D) - f(A)\|_F \leq (1+\sqrt{2})^2 \sup_{x \in \mathbb{E}} |f'(x)| \cdot \|D\|_F$$

where $\mathbb{E}$ is convex set containing numerical ranges of A and $A + D$.

Error bound for low-rank updates: Frobenius norm of $f(A + bc^T) - f(A) = U_m X_m(f) V_m^T$ bounded by

$$2(1+\sqrt{2})^2 \|b\|_2 \|c\|_2 \min_p \sup_{x \in \mathbb{E}} \cdot |(f' - p)(x)|.$$

Extension to rational Krylov subspaces/rational approximation possible.

Divide and Conquer Methods

Divide and Conquer for matrix functions

Divide: Decompose

$$A = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix} + \text{low rank}$$

Recurse: Compute/approximate recursively

$$f(\widetilde{A}) = \begin{bmatrix} f(A_{11}) & 0 \\ 0 & f(A_{22}) \end{bmatrix}$$

Conquer: Use low rank-updates to approximate

$$f(A) - f(\widetilde{A})$$

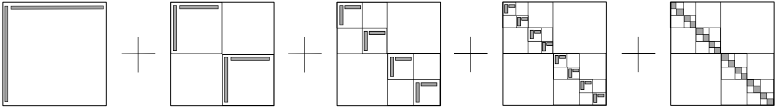
Complexity: Assuming bounded rank, n-independent convergence:

- $O(n \log n)$ for general matrices
- $O(n)$ for banded matrices (exploiting sparsity in low-rank updates)

Instead of $O(n^3)$ for general methods!

Divide and Conquer for matrix functions

Output:



Hierarchical low-rank structure (HODLR, HSS, ...)

Scope of divide and conquer:

- banded matrices
- sparse matrices with reasonably small edge separators (PDE discretizations, graph Laplacians, . . .)
- discretized integral equations
- Toeplitz/Hankel/... matrices (after Fourier transform [Martinsson/Rokhlin/Tygert'2005], [Xia/Xi/Gu'2012])

Main competitor: Matrix iterations in data-sparse arithmetic [Gavrilyuk/Hackbusch/Khoromskij'2002], [Grasedyck/Hackbusch/Khoromskij'2003], [K./Luce'2018], [Bini/Meini'2019], [Massei/Robol/K.'2020], . . .

Example: Fractional diffusion equation

$$\begin{cases} \frac{\partial u(x,t)}{\partial t} = \frac{\partial^\alpha u(x,t)}{\partial_- x^\alpha} + \frac{\partial^\alpha u(x,t)}{\partial_+ x^\alpha} & (x,t) \in (0,1) \times (0,T] \\ u(x,t) = 0 & (x,t) \in (\mathbb{R} \setminus [0,1]) \times [0,T] \\ u(x,0) = u_0(x) & x \in [0,1] \end{cases}$$

Spatial discretization with Grünwald-Letnikov $\rightsquigarrow$ exponential of (dense) matrix with hierarchical low-rank structure.

Execution time (in seconds)

n	D&C	expm(HSS)	dense
512	0.09	0.57	0.03
1024	0.18	1.01	0.15
2048	0.38	1.77	0.96
4096	1.09	3.99	8.94
8192	3.1	10.2	70.8
32768	22.2	37.2	–

Further examples

- Sampling from Gaussian Markov random field $\rightsquigarrow$
  matrix square root of $32\,768 \times 32\,768$ matrix. D&C: 13.6 seconds
- NtD map $\rightsquigarrow$
  inverse matrix square root of $32\,768 \times 32\,768$ matrix.
  D&C: 37.7 seconds
- subgraph centralities $\rightsquigarrow$
  diag(exp) of $45\,087 \times 45\,087$ matrix. D&C: 193 seconds

Conclusions

- Low-rank updates of $f(A)$ are cheap.
- Elegant convergence analysis via bivariate matrix functions.
- Basis of divide-and-conquer methods.

More details in

- B. Beckermann, D. Kressner, and M. Schweitzer. Low-rank updates of matrix functions. SIAM J. Matrix Anal. Appl., 2018.
- B. Beckermann, A. Cortinovis, D. Kressner, and M. Schweitzer. Low-rank updates of matrix functions II: Rational Krylov methods and the matrix sign function. SIAM J. Matrix Anal. Appl., 2021.
- A. Cortinovis, D. Kressner, and S. Massei. Divide and conquer methods for functions of matrices withbanded or hierarchical low-rank structure. In preparation, 2021.
- M. Crouzeix and D. Kressner. A bivariate extension of the Crouzeix-Palencia result with an application to Fréchet derivatives of matrix functions. arXiv'2020.
- L. Robol, D. Kressner, and S. Massei. Low-rank updates and divide-and-conquer for linear matrix equations. SIAM J. Sci. Comput., 2019.