Generalized Matrix Functions: Properties, Algorithms, and Applications

Michele Benzi

Department of Mathematics and Computer Science

Emory University

Atlanta, Georgia, USA

Numerical Linear Algebra with Applications (NLA2)
October 22-28, 2016
Luminy, France

Outline

Acknowledgements

- Joint work with Francesca Arrigo (Strathclyde) and Caterina Fenu (Aachen)
- Support: NSF (Grant DMS-1418889)

See F. Arrigo, M. Benzi, and C. Fenu, *Computation of generalized matrix functions*, SIMAX 37(3), pp. 836–860, 2016.

Motivation

Generalized matrix functions were originally introduced by J. B. Hawkins and A. Ben-Israel in 1973 in an attempt to extend the notion of a matrix function to rectangular matrices. The idea was to parallel the construction of the (Moore-Penrose) generalized inverse, using the SVD.

J. B. Hawkins and A. Ben-Israel, *On generalized matrix functions*, Linear and Multilinear Algebra, 1 (1973), pp. 163–171.

This paper is purely theoretical and does not mention any applications.

Motivation (cont.)

Note that the use of the term “generalized" is somewhat misleading, since this notion of matrix function does not reduce to the usual one when A is square, except in special situations.

The name was dropped in the treatment given later in

A. Ben-Israel and T. N. E. Greville, *Generalized Inverses: Theory and Applications*, Second Ed., Springer, New York, 2003.

Motivation (cont.)

This notion has gone largely unnoticed and has been dismissed by some as useless. Nevertheless, generalized matrix functions, usually unrecognized as such, have apppeared repeatedly in different contexts in the literature and do have important applications, for instance to matrix optimization and low-rank approximation problems arising in

- computer vision (photometric stereo and optical flow)
- regularization of discrete ill-posed problems
- financial mathematics
- MRI
- control theory
- complex frequency estimation

The notion of generalized matrix function also arises in the analysis of directed networks and as a subtask when computing (standard) functions of skew-symmetric matrices (Del Buono, Lopez, and Peluso, SISC 2005).

Definition of generalized matrix function

Definition

Let $A \in \mathbb{C}^{m \times n}$ be a rank r matrix and let $A = U_r \Sigma_r V_r^*$ be its compact SVD. Let $f : \mathbb{R} \to \mathbb{R}$ be a scalar function such that $f(\sigma_i)$ is defined for all $i = 1, 2, \ldots, r$. The generalized matrix function $f^\diamond : \mathbb{C}^{m \times n} \to \mathbb{C}^{m \times n}$ induced by f is defined as

$$f^\diamond(A) := U_r f(\Sigma_r) V_r^*,$$

where $f(\Sigma_r)$ is defined for the $r \times r$ matrix Σ_r in the standard way:

$$f(\Sigma_r) = \mathrm{diag}(f(\sigma_1), f(\sigma_2), \ldots, f(\sigma_r)).$$

Basic properties

For $A = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^* = \sum_{i=1}^{r} \sigma_i E_i$, let

$$E := \sum_{i=1}^{r} E_i = U_r V_r^*.$$

Note that $EE^* = P_{R(A)}$ and $E^*E = P_{R(A^*)}$.

Proposition

Let $f, g, h : \mathbb{R} \to \mathbb{R}$ be scalar functions and let $f^\diamond, g^\diamond, h^\diamond : \mathbb{C}^{m\times n} \to \mathbb{C}^{m\times n}$ be the corresponding generalized matrix functions. Then:

(i) *if $f(z) = k$, then $f^\diamond(A) = kE$;*

(ii) *if $f(z) = z$, then $f^\diamond(A) = A$;*

(iii) *if $f(z) = z^{-1}$, then $f^\diamond(A) = (A^\dagger)^*$;*

(iv) *if $f(z) = g(z) + h(z)$, then $f^\diamond(A) = g^\diamond(A) + h^\diamond(A)$;*

(v) *if $f(z) = g(z)h(z)$, then $f^\diamond(A) = g^\diamond(A)E^*h^\diamond(A)$.*

Basic properties (cont.)

Proposition

Let $A \in \mathbb{C}^{m \times n}$ *be a matrix of rank* r. *Let* $f : \mathbb{R} \to \mathbb{R}$ *be a scalar function and let* $f^\diamond : \mathbb{C}^{m \times n} \to \mathbb{C}^{m \times n}$ *be the induced generalized matrix function, assumed to be defined at* A. *Then the following properties hold true.*

(i) $[f^\diamond(A)]^* = f^\diamond(A^*)$;

(ii) *let* $X \in \mathbb{C}^{m \times m}$ *and* $Y \in \mathbb{C}^{n \times n}$ *be two unitary matrices, then* $f^\diamond(XAY) = X[f^\diamond(A)]Y$;

(iii) *if* $A = \mathrm{diag}(A_{11}, A_{22}, \ldots, A_{kk})$, *then*

$$f^\diamond(A) = \mathrm{diag}(f^\diamond(A_{11}), f^\diamond(A_{22}), \ldots, f^\diamond(A_{kk}));$$

(iv) $f^\diamond(I_k \otimes A) = I_k \otimes f^\diamond(A)$;

(v) $f^\diamond(A \otimes I_k) = f^\diamond(A) \otimes I_k$.

Basic properties (cont.)

The next two propositions describe the relation between generalized and standard matrix functions.

Proposition

Let $A \in \mathbb{C}^{m\times n}$ be a rank r matrix and let $f : \mathbb{R} \to \mathbb{R}$ be a scalar function. Let $f^\diamond : \mathbb{C}^{m\times n} \to \mathbb{C}^{m\times n}$ be the induced generalized matrix function. Then

$$f^\diamond(A) = \left(\sum_{i=1}^{r} \frac{f(\sigma_i)}{\sigma_i}\mathbf{u}_i\mathbf{u}_i^*\right) A = A\left(\sum_{i=1}^{r} \frac{f(\sigma_i)}{\sigma_i}\mathbf{v}_i\mathbf{v}_i^*\right), \tag{1a}$$

or, equivalently,

$$f^\diamond(A) = f(\sqrt{AA^*})(\sqrt{AA^*})^\dagger A = A(\sqrt{A^*A})^\dagger f(\sqrt{A^*A}). \tag{1b}$$

Basic properties (cont.)

Proposition

Let $A \in \mathbb{C}^{m\times n}$ have the polar decomposition

$$A = PH$$

with $P \in \mathbb{C}^{m\times n}$ having orthonormal columns and $H \in \mathbb{C}^{n\times n}$ Hermitian positive semidefinite. If $f^{\diamond}(A)$ is defined, then

$$f^{\diamond}(A) = Pf(H)\,, \tag{2}$$

where $f(H)$ is a standard matrix function of H.

Note: recall that $H = \sqrt{A^*A}$.

Basic properties (cont.)

When f is analytic, it is also possible to express $f^{\diamond}(A)$ in terms of *generalized power series* and also in terms of contour integrals. We will not need these representations here.

The notion of generalized matrix function also extends to compact operators on infinite-dimensional separable Hilbert spaces (since the SVD does).

F. Andersson, M. Carlsson, and K.-M. Perfekt, *Operator-Lipschitz estimates for the singular value functional calculus*, Proceedings of the AMS, 144(5), pp. 1867–1875, 2016.

Functions of block matrices

In network science and in the numerical solution of systems of ODEs it is required io compute (standard) matrix functions of matrices of the form

$$\mathcal{A} = \begin{pmatrix} 0 & A \\ A^* & 0 \end{pmatrix} \quad \text{or} \quad \mathcal{B} = \begin{pmatrix} 0 & -A \\ A^* & 0 \end{pmatrix}.$$

In particular, exponentials of matrices of the form $\mathcal{A}$ arise in the analysis of bipartite and directed networks, and exponentials of matrices of the form $\mathcal{B}$ arise, for example, in the numerical integration of the Korteveg-de Vries equation and other Hamiltonian systems.

M. Benzi, E. Estrada, and C. Klymko, *Ranking hubs and authorities using matrix functions*, LAA, 438 (2013), pp. 2447–2474.

N. Del Buono, L. Lopez, and R. Peluso, *Computation of the exponential of large sparse skew-symmetric matrices*, SISC, 27 (2005), pp. 278–293.

Functions of block matrices (cont.)

It is easy to check that the exponentials of $\mathcal{A}$ and $\mathcal{B}$ are given by

$$\exp(\mathcal{A}) = \begin{pmatrix} \cosh(\sqrt{AA^*}) & \sinh^\diamond(A) \\ \sinh^\diamond(A^*) & \cosh(\sqrt{A^*A}) \end{pmatrix}$$

and

$$\exp(\mathcal{B}) = \begin{pmatrix} \cos(\sqrt{AA^*}) & -\sin^\diamond(A) \\ \sin^\diamond(A^*) & \cos(\sqrt{A^*A}) \end{pmatrix},$$

respectively. Similar expressions hold for other functions of $\mathcal{A}$ and $\mathcal{B}$, leading to even functions of $\sqrt{AA^*}$ and $\sqrt{A^*A}$ on the diagonal and to generalized odd functions of A and A^* in the off-diagonal positions.

Hence, generalized matrix functions occur as submatrices of (standard) functions of certain block 2×2 matrices.

Another example

The “hybrid" matrix function

$$\Phi(A) = I - A + \frac{AA^T}{2!} - \frac{AA^TA}{3!} + \frac{AA^TAA^T}{4!} - \cdots ,$$

where A is the adjacency matrix of a directed network, was studied in

J. J. Crofts, E. Estrada, D. J. Higham, and A. Taylor, *Mapping directed networks*, ETNA, 37 (2010), pp. 337–350.

Note that if $A = U\Sigma V^T$, then

$$\Phi(A) = U\cosh(\Sigma)\; U^T - U\sinh(\Sigma)\; V^T = \cosh\left(\sqrt{AA^T}\right) - \sinh^\diamond(A).$$

Gaussian quadrature

Most problems involving generalized matrix functions lead to the computation of bilinear forms of the type

$$\mathbf{z}^* f^{\diamond}(A)\mathbf{w}. \tag{3}$$

Owing to the identities

$$\mathbf{z}^* f^{\diamond}(A)\mathbf{w} = \mathbf{z}^* \left(\sum_{i=1}^{r} \frac{f(\sigma_i)}{\sigma_i} \mathbf{u}_i \mathbf{u}_i^* \right) \widetilde{\mathbf{w}} = \widetilde{\mathbf{z}}^* \left(\sum_{i=1}^{r} \frac{f(\sigma_i)}{\sigma_i} \mathbf{v}_i \mathbf{v}_i^* \right) \mathbf{w},$$

where $\widetilde{\mathbf{w}} = A\mathbf{w}$, and $\widetilde{\mathbf{z}} = A^*\mathbf{z}$, we can rewrite (3) as

$$\mathbf{z}^* f^{\diamond}(A)\mathbf{w} = \widetilde{\mathbf{z}}^* \left(\sum_{i=1}^{r} \frac{f(\sigma_i)}{\sigma_i} \mathbf{v}_i \mathbf{v}_i^* \right) \mathbf{w} = \widetilde{\mathbf{z}}^* g(A^*A)\mathbf{w}, \tag{4a}$$

$$\mathbf{z}^* f^{\diamond}(A)\mathbf{w} = \mathbf{z}^* \left(\sum_{i=1}^{r} \frac{f(\sigma_i)}{\sigma_i} \mathbf{u}_i \mathbf{u}_i^* \right) \widetilde{\mathbf{w}} = \mathbf{z}^* g(AA^*)\widetilde{\mathbf{w}}, \tag{4b}$$

where in both cases $g(t) = (\sqrt{t})^{-1} f(\sqrt{t})$.

Gaussian quadrature (cont.)

Note that if $\mathbf{z}, \mathbf{w}$ are vectors such that $\widetilde{\mathbf{z}} \neq \mathbf{w}$, then we can use the polarization identity:

$$\widetilde{\mathbf{z}}^* g(A^*A)\mathbf{w} = \frac{1}{4}\left[(\widetilde{\mathbf{z}} + \mathbf{w})^* g(A^*A)(\widetilde{\mathbf{z}} + \mathbf{w}) - (\widetilde{\mathbf{z}} - \mathbf{w})^* g(A^*A)(\widetilde{\mathbf{z}} - \mathbf{w})\right]$$

to reduce the evaluation of the bilinear form of interest to the evaluation of two Hermitian forms.

Hence, we can assume that $\widetilde{\mathbf{z}} = \mathbf{w}$.

See G. H. Golub and G. Meurant, *Matrices, Moments and Quadrature with Applications*, Princeton University Press, Princeton, NJ, 2010.

Gaussian quadrature (cont.)

Let $\widetilde{\mathbf{z}} = \mathbf{w}$ be a unit vector (i.e., $\|\mathbf{w}\|_2 = 1$). We can rewrite the quantity (4a) as a Riemann–Stieltjes integral using the eigendecomposition of A^*A:

$$\mathbf{w}^* g(A^*A)\mathbf{w} = \mathbf{w}^* V_r g(\Sigma_r^2) V_r^* \mathbf{w} = \sum_{i=1}^{r} \frac{f(\sigma_i)}{\sigma_i} |\mathbf{v}_i^* \mathbf{w}|^2 = \int_{\sigma_r^2}^{\sigma_1^2} g(t)\, d\alpha(t),$$

where $\alpha(t)$ is a piecewise constant step function with jumps at the positive eigenvalues $\{\sigma_i^2\}_{i=1}^r$ of A^*A, defined as follows:

$$\alpha(t) = \begin{cases} 0, & \text{if } t < \sigma_r^2 \\ \sum_{i=j+1}^{r} |\mathbf{v}_i^* \mathbf{w}|^2, & \text{if } \sigma_{j+1}^2 \leq t < \sigma_j^2 \\ \sum_{i=1}^{r} |\mathbf{v}_i^* \mathbf{w}|^2, & \text{if } t \geq \sigma_r^2. \end{cases}$$

We use Gaussian quadrature to approximate the above Stieltjes integral. We now show that the nodes and weights can be obtained from the Golub–Kahan bidiagonalization of A, with starting vector $\mathbf{w}$.

Gaussian quadrature (cont.)

We can approximate the quadratic form (Stieltjes integral) with an ℓ point Gauss quadrature rule:

$$\mathcal{G}_\ell := \mathbf{e}_1^T g\left(B_\ell^* B_\ell\right) \mathbf{e}_1 = \mathbf{e}_1^T \left(\sqrt{B_\ell^* B_\ell}\right)^\dagger f\left(\sqrt{B_\ell^* B_\ell}\right) \mathbf{e}_1, \tag{5}$$

where B_ℓ is the bidiagonal matrix obtained after ℓ steps of Golub–Kahan bidiagonalization of A, with starting vector $\mathbf{w}$.

Proposition

Let $(\theta_i, \upsilon_i, \nu_i)$ for $i = 1, 2, \ldots, \ell$ be the singular triplets of $B_\ell = \mathcal{U}_\ell \Theta_\ell \mathcal{V}_\ell^*$. Then the nodes of the ℓ-point Gauss quadrature rule $\mathcal{G}_\ell$ are the singular values $\{\theta_i\}_{i=1}^\ell$. Furthermore, if $\mathbf{z} = \widetilde{\mathbf{w}}$, the weights of $\mathcal{G}_\ell$ are $(\mathbf{e}_1^T \upsilon_i)^2 \theta_i^{-1}$ for $i = 1, 2, \ldots, \ell$.
Similarly, if $\widetilde{\mathbf{z}} = \mathbf{w}$, then the weights of the rule are given by $(\mathbf{e}_1^T \nu_i)^2 \theta_i^{-1}$.

One can also prescribe some of the nodes (Gauss–Radau/Gauss–Lobatto rules).

Gaussian quadrature (cont.)

The previous approach requires computing the standard matrix function $f(\sqrt{T_\ell})$, with $T_\ell = B_\ell^* B_\ell$ tridiagonal. In alternative, we can work directly with B_ℓ:

Proposition

The ℓ-point Gauss quadrature rule $\mathcal{G}_\ell$ is given by

$$\mathcal{G}_\ell = \mathbf{e}_1^T B_\ell^\dagger f^\diamond(B_\ell)\mathbf{e}_1, \qquad \text{if } \widetilde{\mathbf{z}} = \mathbf{w},$$

or

$$\mathcal{G}_\ell = \mathbf{e}_1^T f^\diamond(B_\ell) B_\ell^\dagger \mathbf{e}_1, \qquad \text{if } \mathbf{z} = \widetilde{\mathbf{w}}.$$

While mathematically equivalent, the two approach can behave rather differently in practice.

Gaussian quadrature (cont.)

A third approach uses the Golub–Kahan decomposition $A = P_r B_r Q_r^*$:

$$\mathbf{z}^* f^\diamond(P_r B_r Q_r^*)\mathbf{w} = \mathbf{z}^* f^\diamond(P_r \mathcal{U}_r \Sigma_r \mathcal{V}_r^* Q_r^*)\mathbf{w} = \mathbf{z}^*(P_r \mathcal{U}_r) f(\Sigma_r)(Q_r \mathcal{V}_r)^* \mathbf{w},$$

hence $\mathbf{z}^* f^\diamond(A)\mathbf{w} = \widehat{\mathbf{z}}^* f^\diamond(B_r)\mathbf{e}_1$, with $\widehat{\mathbf{z}} = P_r^* \mathbf{z}$ and $Q_r^* \mathbf{w} = \mathbf{e}_1$.

Assume now that $\ell < r$. We can truncate the bidiagonalization process and approximate $f^\diamond(A)\mathbf{w}$ as

$$f^\diamond(A)\mathbf{w} \approx P_\ell f^\diamond(B_\ell)\mathbf{e}_1$$

and then obtain the approximation to the bilinear form of interest as

$$\mathbf{z}^T f^\diamond(A)\mathbf{w} \approx \mathbf{z}^T P_\ell f^\diamond(B_\ell)\mathbf{e}_1.$$

The quality of the approximation will depend in general on the distribution of the singular values of A and on the particular choice of f.

Test problems

We present some total communicability computations on two directed networks: ITwiki and SLASHDOT. Here $m = n$ and the computed quantities are row sums:

$$C(i) = [f^\diamond(A)\mathbf{1}]_i = \mathbf{e}_i^T f^\diamond(A)\mathbf{1}\,.$$

ITwiki is the Italian Wikipedia. Its adjacency matrix A is $49,728 \times 49,728$ and has $941,425$ nonzeros, and there is a link from node i to node j in the graph if page i refers to page j.

SLASHDOT is a social news website on science and technology (aka "news for nerds"). There is a connection from node i to node j if user i indicated user j as a friend or a foe. Its adjacency matrix A is $82,168 \times 82,168$ matrix with $948,464$ nonzeros.

Test problems (cont.)

We approximate $C(i)$ for ten different choices of i using Gauss quadrature with ℓ nodes using the stopping criterion

$$\mathcal{R}_\ell = \frac{\left|x^{(\ell)} - x^{(\ell-1)}\right|}{\left|x^{(\ell)}\right|} \leq \text{tol}$$

and $x^{(\ell)}$ is the approximation to $C(i)$ obtained with ℓ steps of the method being tested.

We also check the relative error

$$\mathcal{E}_\ell = \frac{|x^{(\ell)} - C(i)|}{|C(i)|},$$

where $C(i)$ is the “exact" quantity, computed using $\gg \ell$ terms.

All the computations were carried out with MATLAB Version 7.10.0.499 (R2010a) 64-bit for Linux, in double precision arithmetic, on an Intel Core i5 computer with 4 GB RAM.

Results for $f(z) = \sinh(z)$

Table: Network: ITwiki, $f(z) = \sinh(z)$ ($\mathrm{tol} = 10^{-6}$).

	First approach		Second approach		Third approach	
	ℓ	$\mathcal{E}_\ell$	ℓ	$\mathcal{E}_\ell$	ℓ	$\mathcal{E}_\ell$
1	5	3.88e-08	5	2.90e-08	6	8.02e-09
2	10	4.72e-05	9	4.68e-05	7	1.27e-08
3	5	3.20e-08	5	3.17e-08	6	7.01e-09
4	7	2.31e-05	9	2.33e-05	8	4.31e-09
5	8	4.20e-05	20	5.77e-05	8	5.91e-09
6	9	2.19e-04	24	2.13e-04	8	2.70e-08
7	6	4.26e-07	6	5.85e-07	7	3.15e-09
8	14	1.91e-04	29	2.24e-04	8	3.38e-09
9	5	8.57e-08	5	9.31e-08	6	5.07e-09
10	9	9.36e-06	8	1.12e-05	8	3.22e-10

Results for $f(z) = \sinh(z)$ (cont.).

Table: Network: SLASHDOT, $f(z) = \sinh(z)$ (tol $= 10^{-6}$).

	First approach		Second approach		Third approach	
	ℓ	$\mathcal{E}_\ell$	ℓ	$\mathcal{E}_\ell$	ℓ	$\mathcal{E}_\ell$
1	6	4.31e-07	6	5.61e-07	9	2.45e-08
2	9	3.24e-05	15	2.26e-06	9	1.56e-08
3	7	1.24e-06	8	1.75e-06	9	1.04e-07
4	14	2.21e-04	8	2.12e-04	10	1.74e-08
5	7	2.24e-05	7	2.35e-05	10	5.16e-09
6	10	4.84e-04	19	3.72e-04	10	1.99e-08
7	7	1.20e-06	7	1.20e-06	9	6.47e-08
8	7	7.11e-07	7	7.66e-07	9	7.68e-09
9	7	5.53e-06	7	5.98e-06	9	1.32e-09
10	6	6.98e-07	6	4.92e-07	8	8.68e-09

Block methods

It is also possible to use block variants of the Golub–Kahan and Lanczos algorithms to compute or approximate quantities of the form

$$Z^* f^\diamond(A) W$$

where now Z and W are $m \times k$ and $n \times k$ matrices; this problem reduces to the previous one when $k = 1$.

For instance, when Z and W are formed with selected columns of the corresponding identity marices I_m and I_n, these algorithms allow one to estimate multiple selected entries of $f^\diamond(A)$ simultaneously.

As is often the case, these block methods are usually more efficient than computing the corresponding entries of $f^\diamond(A)$ one by one by a "scalar" algorithm.

See our paper for details.

A numerical example

We compute blocks of hub-to-authority communicabilities in the Twitter network, which has $3,656$ nodes and $188,712$ edges. These are $k \times k$ submatrices of $\sinh^{\diamond}(A)$, where A is the adjacency matrix.

Define the relative error $\mathcal{E}_\ell$ and the relative distance $\mathcal{R}_\ell$ as

$$\mathcal{E}_\ell = \frac{\|F_\ell - Z^T f^{\diamond}(A)W\|_2}{\|Z^T f^{\diamond}(A)W\|_2} \qquad \text{and} \qquad \mathcal{R}_\ell = \frac{\|\mathcal{G}_\ell - \mathcal{H}_{\ell+1}\|_{\max}}{\|\mathcal{G}_\ell + \mathcal{H}_{\ell+1}\|_{\max}},$$

respectively, where

$$\|M\|_{\max} = \max_{\substack{1\le i\le m\\ 1\le j\le n}} \{M_{ij}\}, \text{ with } M \in \mathbb{C}^{m\times n},$$

and F_ℓ is the arithmetic mean

$$F_\ell = \frac{1}{2}\left(\mathcal{G}_\ell + \mathcal{H}_{\ell+1}\right)$$

between Gauss and anti-Gauss quadrature rules, which can be used as an approximation of the matrix-valued expression $Z^T f^{\diamond}(A)W$.

A numerical example (cont.)

Table: Execution time (in seconds), relative error and relative distance for the computation of the total communicabilities between k nodes of the Twitter network with $\ell = 5$ and $\ell = 10$ steps.

k	$\ell = 5$			$\ell = 10$		
	Time	$\mathcal{E}_5$	$\mathcal{R}_5$	Time	$\mathcal{E}_{10}$	$\mathcal{R}_{10}$
5	2.14e-01	4.62e-04	5.07e-09	3.50e-01	4.62e-04	9.74e-10
10	2.70e-01	1.04e-02	2.21e-09	5.62e-01	1.04e-02	9.96e-10
20	4.21e-01	3.78e-02	5.39e-10	1.10e+00	3.78e-02	8.12e-09
30	6.63e-01	2.24e-02	1.78e-11	2.12e+00	2.24e-02	3.14e-10
50	1.24e+00	4.59e-02	6.83e-12	5.57e+00	4.59e-02	1.63e-11
100	3.86e+00	5.65e-02	3.43e-11	2.72e+01	5.65e-02	1.60e-11

Conclusions and future work

The upshot:

- Generalized matrix functions arise in several important applications and are interesting mathematical objects to study.
- For large sparse, matrices, generalized matrix functions can be approximated via Gauss quadrature rules and Golub–Kahan bidiagonalization.
- Convergence will be fast if f is large on the large singular values of A, and small on the rest.

Topics for further research:

- A better theoretical understanding of generalized matrix functions is needed (see work of V. Noferini in this direction).
- The convergence properties of the various algorithms need to be better understood.
- Algorithms for more “difficult" choices of f.