Computing the Matrix Square Root in a Matrix Group

Nick Higham
Department of Mathematics
The Victoria University of Manchester

higham@ma.man.ac.uk
http://www.ma.man.ac.uk/~higham/

Joint work with Niloufer Mackey, D. Steven Mackey, and Françoise Tisseur.

THE UNIVERSITY of MANCHESTER

Matrix Square Root

- X is a square root of $A \in \mathbb{C}^{n\times n} \iff X^2 = A$.
- Number of square roots may be zero, finite or infinite.

Principal Square Root

- For A with no eigenvalues on $\mathbb{R}^- = \{x \in \mathbb{R} : x \le 0\}$ is the unique square root with spectrum in the open right half-plane.
- Denoted by $A^{1/2}$.

Newton's Method

$A = (X + E)^2$: drop second order term and solve for E.

$$\left.\begin{aligned} X_0 \text{ given,}& \\ \text{Solve } X_k E_k + E_k X_k &= A - X_k^2 \\ X_{k+1} &= X_k + E_k \end{aligned}\right\} k = 0, 1, 2, \ldots$$

Prohibitively expensive.

Newton's Method

$A = (X + E)^2$: drop second order term and solve for E.

$$\left.\begin{aligned} X_0 \text{ given}, \\ \text{Solve } X_k E_k + E_k X_k &= A - X_k^2 \\ X_{k+1} &= X_k + E_k \end{aligned}\right\} k = 0, 1, 2, \ldots$$

Prohibitively expensive.

But observe that

$$X_0 A = A X_0 \quad \Rightarrow \quad X_k A = A X_k \text{ for all } k.$$

Newton Iteration

$$X_{k+1} = \tfrac{1}{2}(X_k + X_k^{-1}A), \qquad X_0 = A.$$

If $A \in \mathbb{C}^{n\times n}$ has no eigenvalues on $\mathbb{R}^-$ then

- Iterates X_k nonsingular, converge quadratically to $A^{1/2}$.
- Related to Newton sign function iterates

$$S_{k+1} = \frac{1}{2}(S_k + S_k^{-1}), \qquad S_0 = A^{1/2}$$

by $X_k \equiv A^{1/2} S_k$.

Newton Iteration

$$X_{k+1} = \tfrac{1}{2}(X_k + X_k^{-1}A), \qquad X_0 = A.$$

If $A \in \mathbb{C}^{n\times n}$ has no eigenvalues on $\mathbb{R}^-$ then

- Iterates X_k nonsingular, converge quadratically to $A^{1/2}$.
- Related to Newton sign function iterates

$$S_{k+1} = \frac{1}{2}(S_k + S_k^{-1}), \qquad S_0 = A^{1/2}$$

by $X_k \equiv A^{1/2} S_k$.

Problem: *numerical instability* (Laasonen, 1958; H, 1996).

Newton Variants

DB iteration:

[Denman–Beavers, 1976]

$$X_{k+1} = \frac{1}{2}\left(X_k + Y_k^{-1}\right), \qquad X_0 = A,$$
$$Y_{k+1} = \frac{1}{2}\left(Y_k + X_k^{-1}\right), \qquad Y_0 = I.$$

Product form DB:

[Cheng-Higham-Kenney-Laub, 2001]

$$M_{k+1} = \frac{1}{2}\left(I + \frac{M_k + M_k^{-1}}{2}\right), \qquad M_0 = A,$$
$$X_{k+1} = \frac{1}{2}X_k(I + M_k^{-1}), \qquad X_0 = A,$$
$$Y_{k+1} = \frac{1}{2}Y_k(I + M_k^{-1}), \qquad Y_0 = I.$$

CR iteration:

[Meini, 2004]

$$Y_{k+1} = -Y_k Z_k^{-1} Y_k, \qquad Y_0 = I - A,$$
$$Z_{k+1} = Z_k + 2Y_{k+1}, \qquad Z_0 = 2(I + A).$$

Content

- ★ Characterizations of functions f that preserve automorphism group structure.
- ★ New, numerically stable square root iterations.
- ★ Unified stability analysis of square root iterations based on Fréchet derivatives.

Group Background

Given nonsingular M and $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$,

$$\langle x, y \rangle_{\mathrm{M}} = \begin{cases} x^T M y, & \text{real or complex bilinear forms,} \\ x^* M y, & \text{sesquilinear forms.} \end{cases}$$

Define automorphism group

$$\mathbb{G} = \{ A \in \mathbb{K}^{n \times n} : \langle Ax, Ay \rangle_{\mathrm{M}} = \langle x, y \rangle_{\mathrm{M}}, \ \forall x, y \in \mathbb{K}^n \}.$$

Adjoint $A^\star$ of $A \in \mathbb{K}^{n \times n}$ wrt $\langle \cdot, \cdot \rangle_{\mathrm{M}}$ defined by

$$\langle Ax, y \rangle_{\mathrm{M}} = \langle x, A^\star y \rangle_{\mathrm{M}} \quad \forall x, y \in \mathbb{K}^n.$$

Can show: $$A^\star = \begin{cases} M^{-1} A^T M, & \text{for bilinear forms,} \\ M^{-1} A^* M, & \text{for sesquilinear forms.} \end{cases}$$

$$\boxed{\mathbb{G} = \{ A \in \mathbb{K}^{n \times n} : A^\star = A^{-1} \}}.$$

Some Automorphism Groups

Space	M	$A^\star$	Automorphism group, $\mathbb{G}$
Groups corresponding to a bilinear form			
$\mathbb{R}^n$	I	A^T	Real orthogonals
$\mathbb{C}^n$	I	A^T	Complex orthogonals
$\mathbb{R}^n$	$\Sigma_{p,q}$	$\Sigma_{p,q}A^T\Sigma_{p,q}$	Pseudo-orthogonals
$\mathbb{R}^n$	R	RA^TR	Real perplectics
$\mathbb{R}^{2n}$	J	$-JA^TJ$	Real symplectics
$\mathbb{C}^{2n}$	J	$-JA^TJ$	Complex symplectics
Groups corresponding to a sesquilinear form			
$\mathbb{C}^n$	I	A^*	Unitaries
$\mathbb{C}^n$	$\Sigma_{p,q}$	$\Sigma_{p,q}A^*\Sigma_{p,q}$	Pseudo-unitaries
$\mathbb{C}^{2n}$	J	$-JA^*J$	Conjugate symplectics

$$R = \begin{bmatrix} & & 1 \\ & \cdot^{\cdot^{\cdot}} & \\ 1 & & \end{bmatrix}, \qquad J = \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix}, \qquad \Sigma_{p,q} = \begin{bmatrix} I_p & 0 \\ 0 & -I_q \end{bmatrix}$$

Bilinear Forms

Theorem 1
(a) *For any* f *and* $A \in \mathbb{K}^{n\times n}$, $f(A^\star) = f(A)^\star$.
(b) *For* $A \in \mathbb{G}$, $f(A) \in \mathbb{G}$ *iff* $f(A^{-1}) = f(A)^{-1}$.

Proof. (a) We have

$$f(A^\star) = f(M^{-1}A^TM) = M^{-1}f(A^T)M = M^{-1}f(A)^TM = f(A)^\star.$$

(b) For $A \in \mathbb{G}$, consider

$$\begin{array}{ccc} f(A)^\star & = & f(A^\star) \\ & & \| \\ & & f(A^{-1}) \end{array}$$

Bilinear Forms

Theorem 1
(a) *For any* f *and* $A \in \mathbb{K}^{n\times n}$, $f(A^\star) = f(A)^\star$.
(b) *For* $A \in \mathbb{G}$, $f(A) \in \mathbb{G}$ *iff* $f(A^{-1}) = f(A)^{-1}$.

Proof. (a) We have

$$f(A^\star) = f(M^{-1}A^TM) = M^{-1}f(A^T)M = M^{-1}f(A)^TM = f(A)^\star.$$

(b) For $A \in \mathbb{G}$, consider

$$\begin{array}{ccc} f(A)^\star & = & f(A^\star) \\ \| & & \| \\ f(A)^{-1} & = & f(A^{-1}) \end{array}$$

Implications

For bilinear forms, f preserves group structure of A when $f(A^{-1}) = f(A)^{-1}$.

This condition holds *for all A* for

- **Matrix sign function**, $\mathrm{sign}(A) = A(A^2)^{-1/2}$.
- Any **matrix power** A^α, subject to suitable choice of branches. In particular, the
  - **principal matrix pth root** $A^{1/p}$ ($p \in \mathbb{Z}^+$, $\Lambda(A) \cap \mathbb{R}^- = \emptyset$): unique X such that
    1. $X^p = A$.
    2. $-\pi/p < \arg(\lambda(X)) < \pi/p$.

Group Newton Iteration

Theorem 2 *Let* $A \in \mathbb{G}$ *(any group),* $\Lambda(A) \cap \mathbb{R}^- = \emptyset$*, and*

$$\begin{aligned} Y_{k+1} &= \frac{1}{2}(Y_k + Y_k^{-\star}) \\ &= \frac{1}{2}(Y_k + M^{-1}Y_k^{-T}M), \qquad Y_1 = \frac{1}{2}(I + A). \end{aligned}$$

Then $Y_k \to A^{1/2}$ *quadratically and* $Y_k \equiv X_k$ $(k \geq 1)$*, where* X_k *are the Newton iterates:* $X_{k+1} = \frac{1}{2}(X_k + X_k^{-1}A)$*,* $X_0 = A$*.*

Group Newton Iteration

Theorem 2 *Let* $A \in \mathbb{G}$ *(any group),* $\Lambda(A) \cap \mathbb{R}^- = \emptyset$*, and*

$$\begin{aligned} Y_{k+1} &= \frac{1}{2}(Y_k + Y_k^{-\star}) \\ &= \frac{1}{2}(Y_k + M^{-1} Y_k^{-T} M), \qquad Y_1 = \frac{1}{2}(I + A). \end{aligned}$$

Then $Y_k \to A^{1/2}$ *quadratically and* $Y_k \equiv X_k$ $(k \geq 1)$*, where* X_k *are the Newton iterates:* $X_{k+1} = \frac{1}{2}(X_k + X_k^{-1} A)$*,* $X_0 = A$*.*

Cf.

- Cardoso, Kenney & Silva Leite (2003, App. Num. Math.): bilinear forms with $M^T = \pm M$, $M^T M = I$.
- H (2003, SIREV): $M = \Sigma_{p,q}$.

Generalized Polar Decomposition

Theorem 2 *For any group* $\mathbb{G}$, $A \in \mathbb{K}^{n\times n}$ *has a unique generalized polar decomposition* $A = WS$ *where*

$$W \in \mathbb{G} \quad (\textit{i.e.}, W^\star = W^{-1}), \qquad S^\star = S,$$

and $\Lambda(S) \in$ *open right half-plane* (*i.e.*, $\mathrm{sign}(S) = I$) *iff* $(A^\star)^\star = A$ *and* $\Lambda(A^\star A) \cap \mathbb{R}^- = \emptyset$.

Note

- $(A^\star)^\star = A$ holds for all $\mathbb{G}$ in the earlier table.
- Other gpd's exist with different conditions on $\Lambda(S)$ (Rodman & co-authors).

GPD Iteration & Square Root

Theorem 3 *Suppose the iteration* $X_{k+1} = X_k h(X_k^2)$, $X_0 = A$ *converges to* $\operatorname{sign}(A)$ *with order* m. *If* A *has the generalized polar decomposition* $A = WS$ *w.r.t. a scalar product then*

$$Y_{k+1} = Y_k h(Y_k^\star Y_k), \qquad Y_0 = A$$

converges to W *with order of convergence* m.

GPD Iteration & Square Root

Theorem 3 *Suppose the iteration* $X_{k+1} = X_k h(X_k^2)$, $X_0 = A$ *converges to* $\mathrm{sign}(A)$ *with order* m. *If* A *has the generalized polar decomposition* $A = WS$ *w.r.t. a scalar product then*

$$Y_{k+1} = Y_k h(Y_k^\star Y_k), \qquad Y_0 = A$$

converges to W *with order of convergence* m.

Theorem 4 *Let* $\mathbb{G}$ *be any automorphism group and* $A \in \mathbb{G}$. *If* $\Lambda(A) \cap \mathbb{R}^- = \emptyset$ *then* $I + A = WS$ *is a generalized polar decomposition with* $W = A^{1/2}$ *and* $S = A^{-1/2} + A^{1/2}$.

Application

Newton

Sign: $X_{k+1} = X_k \cdot \frac{1}{2}(I + (X_k^2)^{-1}) \equiv X_k h(X_k^2), \quad X_0 = A$.

Group sqrt:

$$Y_{k+1} = \frac{1}{2} Y_k (I + (Y_k^\star Y_k)^{-1}) = \frac{1}{2}(Y_k + Y_k^{-\star}), \quad Y_0 = I + A.$$

Schulz

Sign: $X_{k+1} = X_k \cdot \frac{1}{2}(3I - X_k^2) \equiv X_k h(X_k^2), \quad X_0 = A$.

Group Schulz:

$$Y_{k+1} = \frac{1}{2} Y_k (3I - Y_k^\star Y_k), \quad Y_0 = I + A.$$

Class of Square Root Iterations

Theorem 5 *Suppose the iteration* $X_{k+1} = X_k h(X_k^2)$, $X_0 = A$ *converges to* $\operatorname{sign}(A)$ *with order* m. *If* $\Lambda(A) \cap \mathbb{R}^- = \emptyset$ *and*

$$\begin{aligned} Y_{k+1} &= Y_k h(Z_k Y_k), & Y_0 &= A, \\ Z_{k+1} &= h(Z_k Y_k) Z_k, & Z_0 &= I, \end{aligned}$$

then $Y_k \to A^{1/2}$ *and* $Z_k \to A^{-1/2}$ *as* $k \to \infty$, *both with order* m, *and* $Y_k = A Z_k$ *for all* k. *Moreover, if* $X \in \mathbb{G}$ *implies* $X h(X^2) \in \mathbb{G}$ *then* $A \in \mathbb{G}$ implies $Y_k \in \mathbb{G}$ and $Z_k \in \mathbb{G}$ *for all* k.

- Proof makes use of $\operatorname{sign}\left(\begin{bmatrix} 0 & A \\ I & 0 \end{bmatrix}\right) = \begin{bmatrix} 0 & A^{1/2} \\ A^{-1/2} & 0 \end{bmatrix}.$
- Newton sign leads to DB iteration.

Padé Square Root Iterations

Example: structure-preserving cubic:

$$Y_{k+1} = Y_k(3I + Z_kY_k)(I + 3Z_kY_k)^{-1}, \qquad Y_0 = A,$$
$$Z_{k+1} = (3I + Z_kY_k)(I + 3Z_kY_k)^{-1}Z_k, \qquad Z_0 = I.$$

If $\Lambda(A) \cap \mathbb{R}^- = \emptyset$ then

- $Y_k \to A^{1/2}$ and $Z_k \to A^{-1/2}$ cubically,
- $A \in \mathbb{G} \;\Rightarrow\; X_k \in \mathbb{G},\; Y_k \in \mathbb{G}$ **for all** k.

Stability

Define $X_{k+1} = g(X_k)$ to be **stable** in nbhd of fixed point $X = g(X)$ if for $X_0 := X + H_0$, with arbitrary error H_0, the $H_k := X_k - X$ satisfy

$$H_{k+1} = L_X(H_k) + O(\|H_k\|^2),$$

where L_X is a linear operator with **bounded powers**.

Theorem 6 *Consider the mapping*

$$G(Y, Z) = \begin{bmatrix} Yh(ZY) \\ h(ZY)Z \end{bmatrix},$$

where $X_{k+1} = X_k h(X_k^2)$ is any superlinear matrix sign iteration. Any $P = (B, B^{-1})$ is a fixed point for G, and

$$dG_P(E, F) = \frac{1}{2} \begin{bmatrix} E - BFB \\ F - B^{-1}EB^{-1} \end{bmatrix}.$$

*dG_P is **idempotent**, that is, $dG_P \circ dG_P = dG_P$.*

Experiment

Random **pseudo-orthogonal** $A \in \mathbb{R}^{10\times 10}$,

$$M = \mathrm{diag}(I_6, -I_4) \qquad (A^T M A = M),$$
$$\|A\|_2 = 10^5 = \|A^{-1}\|_2,$$

generated using alg of H (2003) and chosen to be symmetric positive definite.

$$\mathrm{err}(X) = \frac{\|X - A^{1/2}\|_2}{\|A^{1/2}\|_2},$$

$$\mu_{\mathbb{G}}(X) = \frac{\|X^\star X - I\|_2}{\|X\|_2^2}.$$

Results

k	Newton	Group Newton		Cubic, struc. pres.	
	$\mathrm{err}(X_k)$	$\mathrm{err}(Y_k)$	$\mu_{\mathbb{G}}(Y_k)$	$\mathrm{err}(Y_k)$	$\mu_{\mathbb{G}}(Y_k)$
0	3.2e2			3.2e2	1.4e-15
1	1.6e2	1.6e2	1.0e-5	1.0e2	7.2e-15
2	7.8e1	7.8e1	1.0e-5	3.4e1	6.1e-14
3	3.9e1	3.9e1	1.0e-5	1.1e1	5.1e-13
4	1.9e1	1.9e1	1.0e-5	3.0e0	2.9e-12
5	8.9e0	8.9e0	9.9e-6	5.5e-1	4.4e-12
6	4.0e0	4.0e0	9.6e-6	2.0e-2	4.3e-12
7	3.2e1	1.6e0	8.5e-6	2.0e-6	4.5e-12
8	2.3e5	4.9e-1	5.5e-6	2.1e-11	4.8e-12
9	4.6e9	8.2e-2	1.5e-6		
10	2.3e9	3.1e-3	6.1e-8		
11	1.1e9	4.7e-6	9.5e-11		
12	5.6e8	2.1e-11	2.4e-16		

Conclusions

- ★ Characterizations of f that preserve group structure (e.g., if $f(A^{-1}) = f(A)^{-1}$ for bilinear forms).
- ★ Using gen polar decomp, derived **numerically stable** form of Newton for $A^{1/2}$ when $A \in \mathbb{G}$.
- ★ Derived new family of coupled iterations for $A^{1/2}$ that is **structure preserving** for matrix groups.
- ★ **Stability analysis** using Fréchet derivative.

- *Functions Preserving Matrix Groups and Iterations for the Matrix Square Root*, NA Report 446, March 2004; to appear in SIMAX.
- Functions of a Matrix; book in preparation.