Matrix functions and network analysis

Paola Boito, Federico Poloni

Pisa–Hokkaido–Roma2 Summer School, August–September 2018
Motivation

Functions of matrices: allow to study in the same framework objects such as

- Matrix power series, e.g., \( \exp(A) = I + A + \frac{1}{2}A^2 + \frac{1}{3!}A^3 + \ldots; \)
- Maps on eigenvalues, e.g.,
  \[
  A = V \Lambda V^{-1} \mapsto V \begin{bmatrix} f(\lambda_1) & & \\
  & \ddots & \\
  & & f(\lambda_n) \end{bmatrix} V^{-1};
  \]
- Matrix rational iterations, e.g., \( X_{k+1} = \frac{1}{2}X_k + \frac{1}{2}X_k^{-1}; \)
- Solutions to matrix equations, e.g., \( X^2 = A. \)

Many objects that appear in applications can be naturally described as functions of matrices — you have probably already encountered \( \exp(A) \) and \( A^{1/2} \), for instance.
Motivation

On top of this, an overview of two interesting applications:

- Solving certain boundary-value problems / matrix equations appearing in control and queuing theory.
- Discovering ‘important’ vertices in a graph (centrality measures).

Reference books

Polynomials of matrices

Take a scalar polynomial, and evaluate it in a (square) matrix, e.g.,

\[ p(x) = 1 + 3x - 5x^2 \implies p(A) = I + 3A - 5A^2. \]

**Lemma**

If \( A = S \text{ blkdiag}(J_1, J_2, \ldots, J_s) S^{-1} \) is a Jordan form, then \( p(A) = S \text{ blkdiag}(p(J_1), p(J_2), \ldots, p(J_s)) S^{-1} \), and

\[
p(J_i) = \begin{bmatrix}
p(\lambda_i) & p'(\lambda_i) & \cdots & 1/k! p^{(k)}(\lambda_i) \\
p(\lambda_i) & \ddots & & \vdots \\
\vdots & \ddots & p'(\lambda_i) & \\
p(\lambda_i) & & p'(\lambda_i) & p(\lambda_i)
\end{bmatrix}.
\]

**Proof** Taylor expansion of \( p \) at \( \lambda_i \) and powers of shift matrix.
Functions of matrices [Higham book, ’08]

We can extend the same definition to arbitrary scalar functions:

**Definition**

If \( A = S \ \text{blkdiag}(J_1, J_2, \ldots, J_s)S^{-1} \) is a Jordan form, then
\[
f(A) = S \ \text{blkdiag}(f(J_1), f(J_2), \ldots, f(J_s))S^{-1},
\]
where
\[
f(J_i) = \begin{bmatrix}
f(\lambda_i) & f'(\lambda_i) & \cdots & \frac{1}{k!}f^{(k)}(\lambda_i) \\
f(\lambda_i) & \ddots & \ddots & \\
\vdots & \ddots & f'(\lambda_i) & \\
f(\lambda_i) & & & f(\lambda_i)
\end{bmatrix}.
\]

Given \( f: U \subseteq \mathbb{C} \rightarrow \mathbb{C} \), we say that \( f \) is defined on \( A \) if \( f \) is defined and differentiable at least \( n_i - 1 \) times on each eigenvalue \( \lambda_i \) of \( A \). (\( n_i = \max. \text{ size of a Jordan block with eigenvalue } \lambda_i. \))

**Reasonable doubt:** is it independent of the choice of \( S \)?
Alternate definition: via Hermite interpolation

**Definition**

\[ f(A) = p(A), \text{ where } p \text{ is a polynomial such that} \]
\[ f(\lambda_i) = p(\lambda_i), \ f'(\lambda_i) = p'(\lambda_i), \ldots, \ f^{n_i - 1}(\lambda_i) = p^{n_i - 1}(\lambda_i) \text{ for each } i. \]

We may use this as a definition of \( f(A) \):

- Does not depend on \( S \);
- Does not depend on \( p \).

Obvious from the definitions that it coincides with the previous one.

**Remark:** be careful when you say “all matrix functions are polynomials”, because \( p \) depends on \( A \).
Some properties

- If the eigenvalues of $A$ are $\lambda_1, \ldots, \lambda_s$, the eigenvalues of $f(A)$ are $f(\lambda_1), \ldots, f(\lambda_s)$. (geometric multiplicities may decrease)

- $f(A)g(A) = g(A)f(A) = (fg)(A)$ (since they are all polynomials in $A$).

- If $f_n \to f$ together with ‘enough derivatives’ (for instance because they are analytic and the convergence is uniform), then $f_n(A) \to f(A)$.

- **continuity** If $A_n \to A$, then $f(A_n) \to f(A)$.
  
  **Proof** let $p_n$ be the (Hermite) interpolating polynomial on the eigenvalues of $A_n$. Interpolating polynomials are continuous in the nodes, so $p_n \to p$ (coefficient by coefficient). Then $\|p_n(A_n) - p(A)\| \leq \|p_n(A_n) - p_n(A)\| + \|p_n(A) - p(A)\| \leq \ldots$. 
Example: square root

\[ A = \begin{bmatrix} 4 & 1 \\ 4 & 1 \\ 4 & 0 \end{bmatrix}, \quad f(x) = \sqrt{x} \]

We look for an interpolating polynomial with

\[ p(0) = 0, \ p(4) = 2, \ p'(4) = f'(4) = \frac{1}{4}, \ p''(4) = f''(4) = -\frac{1}{32}. \]

I.e.,

\[
\begin{bmatrix}
0 & 0 & 0 & 1 \\
4^3 & 4^2 & 4 & 1 \\
3 \cdot 4^2 & 2 \cdot 4 & 1 & 0 \\
6 \cdot 4 & 2 & 0 & 0 \\
\end{bmatrix}
\begin{bmatrix}
p_3 \\
p_2 \\
p_1 \\
p_0 \\
\end{bmatrix}
=
\begin{bmatrix}
0 \\
2 \\
\frac{1}{4} \\
\frac{1}{32} \\
\end{bmatrix},
\]

\[ p(x) = \frac{3}{256}x^3 - \frac{5}{32}x^2 + \frac{15}{16}x. \]
Example – continues

\[ p(A) = \frac{3}{256}A^3 - \frac{5}{32}A^2 + \frac{15}{16}A = \begin{bmatrix} 2 & \frac{1}{4} & -\frac{1}{64} \\ \frac{1}{4} & 2 & \frac{1}{4} \\ 2 & 2 & 0 \end{bmatrix}. \]

One can check that \( f(A)^2 = A \).
Example – square root

\[ A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad f(x) = \sqrt{x} \]

does not exist (because \( f'(0) \) is not defined).

(Indeed, there is no matrix such that \( X^2 = A \).)
Example – matrix exponential

\[ A = S \begin{bmatrix} -1 & 0 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} S^{-1}, \quad f(x) = \exp(x). \]

\[ \exp(A) = S \begin{bmatrix} e^{-1} & 1 \\ e & e \\ e & e \end{bmatrix} S^{-1} \]

Can also be obtained as \( I + A + \frac{1}{2} A^2 + \frac{1}{6} A^3 + \ldots \) (not so obvious, for Jordan blocks...)
Example – matrix sign

\[
A = S \begin{bmatrix}
-3 & -2 \\
1 & 1
\end{bmatrix} S^{-1}, \quad f(x) = \text{sign}(x) = \begin{cases}
1 & \text{Re } x > 0, \\
-1 & \text{Re } x < 0.
\end{cases}
\]

\[
f(A) = S \begin{bmatrix}
-1 & -1 \\
1 & 1
\end{bmatrix} S^{-1}.
\]

Not a multiple of \( I \), in general.

Instead, we can recover stable / unstable invariant subspaces of \( A \) as \( \ker(f(A) \pm I) \).

If we found a way to compute \( f(A) \) without diagonalizing, we could use it to compute eigenvalues via bisection...
Example – complex square root

\[ A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad f(x) = \sqrt{x} \]

We can choose branches arbitrarily: let us say \( f(i) = \frac{1}{\sqrt{2}}(1 + i) \), \( f(-i) = \frac{1}{\sqrt{2}}(1 - i) \).

Polynomial: \( p(x) = \frac{1}{\sqrt{2}}(1 + x) \).

\[ p(A) = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}. \]

(This is the so-called \textit{principal} square root: we have chosen the values of \( f(\pm i) \) in the right half-plane — other choices are possible).

(We get a non-real square root of \( A \) if we choose non-conjugate values for \( f(i) \) and \( f(-i) \))
Example – nonprimary square root

With our definition, if we have

\[ A = S \begin{bmatrix} 1 & 1 \\ & 2 \end{bmatrix} S^{-1}, \quad f(x) = \sqrt{x} \]

we cannot get

\[ f(A) = S \begin{bmatrix} 1 & -1 \\ \quad & \sqrt{2} \end{bmatrix} S^{-1} : \]

either \( f(1) = 1 \), or \( f(1) = -1 \ldots \)

This would also be a solution of \( X^2 = A \), though. This is called a nonprimary square root of \( A \). We get nonprimary roots/functions if we choose different branches for Jordan blocks with the same eigenvalue.

Not functions of matrices, with our definition. Also, they are not polynomials in \( A \).
Cauchy integrals

If $f$ is analytic on and inside a contour $\Gamma$ that encloses the eigenvalues of $A$, 

$$f(A) = \frac{1}{2\pi i} \int_{\Gamma} f(z)(zI - A)^{-1} dz.$$ 

Generalizes the analogous scalar formula.

**Proof** If $A = V\Lambda V^{-1} \in \mathbb{C}^{m\times m}$ is diagonalizable, the integral equals

$$V \begin{bmatrix} \frac{1}{2\pi i} \int_{\Gamma} \frac{f(z)}{z-\lambda_1} dz \\ \vdots \\ \frac{1}{2\pi i} \int_{\Gamma} \frac{f(z)}{z-\lambda_m} dz \end{bmatrix} V^{-1} = V \begin{bmatrix} f(\lambda_1) \\ \vdots \\ f(\lambda_m) \end{bmatrix} V^{-1}.$$ 

By continuity, the equality holds also for non-diagonalizable $A$. 
Methods

Matrix functions arise in several areas: solving ODEs (e.g. $\exp(A)$), matrix analysis (square roots), physics, ... 

Main methods to compute them:

- Factorizations (eigendecompositions, Schur...),
- Matrix versions of scalar iterations (e.g., Newton on $x^2 = a$),
- Interpolation / approximation,
- Complex integrals.

We will study them in this course. But, first, a detour.
Vectorization

Matrix functions are maps $\mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ (or $\mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$). We introduce some terminology / notation to study linear maps between these spaces.

**Definition**

For $A \in \mathbb{C}^{m\times n}$, $v = \text{vec}(A)$ is the vector $v \in \mathbb{C}^{mn}$ obtained by concatenating the columns of $A$.

$$\text{vec} \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{bmatrix}.$$
**Kronecker products**

### Definition

Given $M = (m_{ij}) \in \mathbb{C}^{m_1 \times m_2}$, $N \in \mathbb{C}^{n_1 \times n_2}$, the **Kronecker product** $M \otimes N \in \mathbb{C}^{m_1 n_1 \times m_2 n_2}$ is the matrix with blocks $m_{ij}N$.

\[
\begin{bmatrix}
1 & 2 \\
3 & 4
\end{bmatrix} \otimes 
\begin{bmatrix}
1 & 2 \\
0 & 1
\end{bmatrix} =
\begin{bmatrix}
1 & 2 & 2 & 4 \\
0 & 1 & 0 & 2 \\
3 & 6 & 4 & 8 \\
0 & 3 & 0 & 4
\end{bmatrix}.
\]

### Lemma

\[
\text{vec}(AXB) = (B^T \otimes A) \text{vec}(X),
\]

i.e., $B^T \otimes A$ is the matrix that represents the linear map $X \mapsto AXB$.

**Warning:** this is $B^T$, not $B^*$ (no conjugation).
Properties of Kronecker product

1. Linear in both factors:
   \[(\lambda L + \mu M) \otimes N = \lambda (L \otimes N) + \mu (M \otimes N).\]

2. \(M^* \otimes N^* = (M \otimes N)^*\).

3. \(LM \otimes NP = (L \otimes N)(M \otimes P)\), if the dimensions are compatible. Follows from \((AB)X(CD) = A(BXC)D\).

4. \((M \otimes N)^{-1} = M^{-1} \otimes N^{-1}\).

5. \(Q_1, Q_2\) unitary \(\implies\) \(Q_1 \otimes Q_2\) unitary.

6. If \(M = V_1 \Lambda_1 V_1^{-1}, N = V_2 \Lambda_2 V_2^{-1}\) are eigendecompositions, then \(M \otimes N = (V_1 \otimes V_2)(\Lambda_1 \otimes \Lambda_2)(V_1 \otimes V_2)^{-1}\) is an eigendecomposition.

7. Analogously for SVD, Schur factorization, …

8. The eigenvalues (singular values) of \(M \otimes N\) are the pairwise products of the eigenvalues (singular values) of \(M\) and \(N\).
Example: Sylvester equations

Given $A, B, C \in \mathbb{C}^{n \times n}$, find $X \in \mathbb{C}^{n \times n}$ that solves the matrix equation $AX - XB = C$.
When does it have a unique solution?

It is a linear system in $\mathbb{C}^{n^2}$.

$$AX - XB = C \iff (I \otimes A - B^T \otimes I) \text{vec}(X) = \text{vec}(C).$$

If $A = Q_A T_A Q_A^*$, $B^T = Q_B T_B Q_B^*$ are Schur decompositions, then

$$I \otimes A - B^T \otimes I = (Q_A \otimes Q_B)(I \otimes T_A - T_B \otimes I)(Q_A \otimes Q_B)^*$$

is a Schur decomposition.

Hence, $\Lambda(I \otimes A - B^T \otimes I) = (\alpha_i - \beta_j : i, j = 1, \ldots, n)$, where $\Lambda(A) = (\alpha_1, \ldots, \alpha_n)$, $\Lambda(B) = (\beta_1, \ldots, \beta_n)$. 
Solution of Sylvester equations

We have proved

Lemma

\[ AX - XB = C \] has a unique solution iff \( A \) and \( B \) have no common eigenvalues.

Corollary: \( AX - XB = C \) is ill-conditioned if \( A, B \) have two close eigenvalues. (It’s an iff when they are normal.)

Numerical solution: can we beat the naive \( O(n^6) \) algorithm “form \( I \otimes A - B^T \otimes I \) and treat it as a \( n^2 \times n^2 \) linear system”?

Yes! [Bartels-Stewart algorithm, 1972].

Idea: invert that Schur decomposition.

\[ (Q_A \otimes Q_B)^* \text{vec}(C) \text{ equals vec}(Q_B^*CQ_A) \]
\[ \sim \text{ product in } O(n^3). \]

\[ I \otimes T_A - T_B \otimes I \text{ has } O(n) \text{ nonzeros per row} \]
\[ \sim \text{ back-substitution in } O(n^3). \]
Extensions

- \( A \in \mathbb{C}^{m \times m}, \ X \in \mathbb{C}^{m \times n}, \ B \in \mathbb{C}^{n \times n} \): everything works without changes.

- **Stein’s equation** \( X - AXB = C \): works analogously. Solvable iff \( \alpha_i \beta_j \neq 1 \) for all \( i, j \).

- \( AXB - CXD = E \) (generalized Sylvester’s equation): works analogously, using generalized Schur factorization \( \text{schur}(A, C) \) and \( \text{schur}(D', B') \).
Lyapunov equations

\[ AX + XA^* = C. \] (*

They are simply Sylvester equations with \( B = -A^* \) (and \( C = C^* \)). They have a few notable properties.

Lemma

Suppose \( A \) has all its eigenvalues in the right half-plane \( RHP = \{ z \in \mathbb{C} : \text{Re}(z) > 0 \} \). Then,

1. (*) has a unique solution.
2. \( X = \int_0^\infty \exp(-tA)C \exp(-tA^*) \, dt \).
3. \( X \succ 0 \) if \( C \succ 0 \). (positive definite ordering)
Proof

With \( X = \int_0^\infty e^{-tA}Ce^{-tA^*} \, dt \), one has

\[
AX + XA^* = \int_0^\infty \left( Ae^{-tA}Ce^{-tA^*} + e^{-tA}Ce^{-tA^*}A^* \right) \, dt
\]

\[
= \left[ -e^{-tA}Ce^{-tA^*} \right]_0^\infty = 0 - (-C).
\]

The converse holds, too:

**Lemma**

If (*) holds with \( C \succ 0 \) and \( X \succ 0 \), then \( A \) has all its eigenvalues in the RHP.

**Proof** Let \( A^* \nu = \lambda \nu \); then,

\[
\nu^* C\nu = \nu^* (AX + XA^*)\nu = \bar{\lambda} \nu^* X\nu + \lambda \nu^* X\nu = 2 \text{Re}(\lambda) \nu^* X\nu.
\]
Lyapunov’s use of these equations

Proving that certain dynamical systems are stable!

Let \( y(t) : [0, \infty] \to \mathbb{C}^n \) be the solution of \( \frac{d}{dt} y(t) = -A y(t) \).

If I can find \( X \succ 0 \) and \( C \succ 0 \) such that \( A^T X + X A^T = C \), then

\[
\frac{d}{dt} y(t)^* X y(t) = y(t)^* (-A^* X - X A) y(t) = -y(t)^* C y(t) < 0.
\]

\( \implies \) The ‘energy’ \( y(t)^* X y(t) \) decreases.