# 1 Lecture 7: Matrix norms and matrix functions

This lecture we start to introduce an application of Cauchy's integral formula and trapezium rule: computing matrix functions, e.g., the matrix exponential $e^A$. There are three possible definitions for matrix functions:

1. Taylor series
2. Diagonalisation/Jordan canonical form
3. Cauchy's integral formula

## 1.1 Matrix norms

Before discussing matrix functions we first review the notion of a matrix norm induced by a vector norm:

$$\|A\| := \sup_{v:\|v\|=1} \|Av\| = \sup_v \frac{\|Av\|}{\|v\|}$$

some examples of matrix norms are the 1-norm

$$\|A\|_1 = \max_j \|A\mathbf{e}_j\|_1$$

that is the maximum column sum, the $\infty$-norm

$$\|A\|_\infty = \max_k \|A^\top \mathbf{e}_k\|_1$$

that is the maximum row sum and the 2-norm

$$\|A\|_2$$

which equals the largest singular value, or the square-root of the largest eigenvalue of $A^\top A$.

We denote the set of eigenvalues of a square matrix $A$, also known as the spectrum of $A$, by $\sigma(A)$. The *spectral radius* of $A$ is defined as

$$\rho(A) = \max\{|\lambda| \;:\; \lambda \in \sigma(A)\}$$

Hence,

$$\|A\|_2 = \sqrt{\rho(A^\top A)}.$$

Matrix norms have the usual norm properties, here $\alpha$ is constant and $B$ has same dimensions as $A$:

1. $$\|\alpha A\| = |\alpha|\|A\|$$

2. $$\|A + B\| \leq \|A\| + \|B\|$$

They have the extra feature that

$$\|AB\| \leq \|A\|\|B\|$$

which implies that

$$\|A^k\| \leq \|A\|^k.$$

### 1.1.1 Taylor series definition

One way to construct matrix exponentials is by Taylor series.

**Definition (Taylor series matrix function)** Suppose

$$f(z) = \sum_{k=0}^{\infty} f_k z^k$$

has radius of convergence $R$. If $\|A\| < R$ holds true for any matrix norm, then define

$$f(A) := \sum_{k=0}^{\infty} f_k A^k.$$

The well-posedness of this construction follows from the fact that the partial sums form a Cauchy sequence. In particular, let $\epsilon > 0$. Then for $r = \|A\| < R$ we can choose $N$ such that

$$\sum_{k=n}^{m} |f_k| r^k \leq \epsilon$$

for all $n, m > N$. Therefore

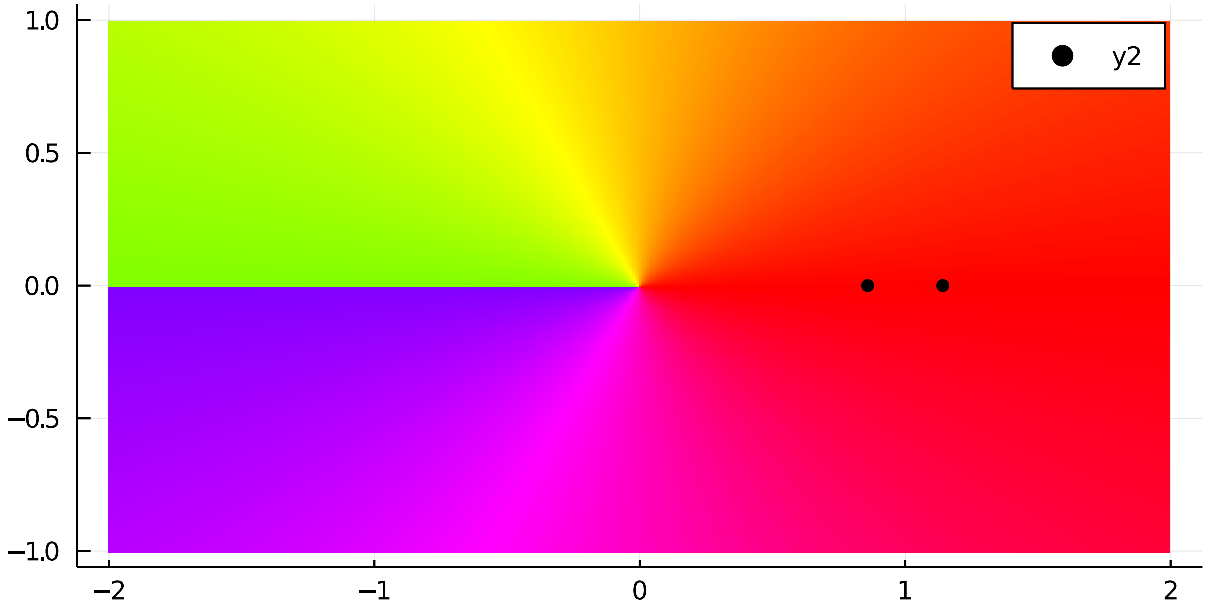
$$\|\sum_{k=n}^{m} f_k A^k\| \leq \sum_{k=n}^{m} |f_k| \|A\|^k \leq \epsilon.$$

The issue with Taylor series is that it requires analyticity to converge. We can see this clearly for the case $\sqrt{A}$ with Taylor series around 1:

$$\sqrt{z+1} = 1 + z/2 - z^2/8 + z^3/16 - \cdots = 1 + z/2 + \sum_{k=3}^{\infty} \frac{3 \cdots (2k-3)}{2^k k!} (-z)^k$$

Consider

```
using Plots, ComplexPhasePortrait, IntervalSets, LinearAlgebra
A = [1 0.1; 0.2 1]
λ = eigvals(A)
phaseplot(-2..2, -1..1, z -> sqrt(z))
scatter!(real(λ),imag(λ); color=:black)
```



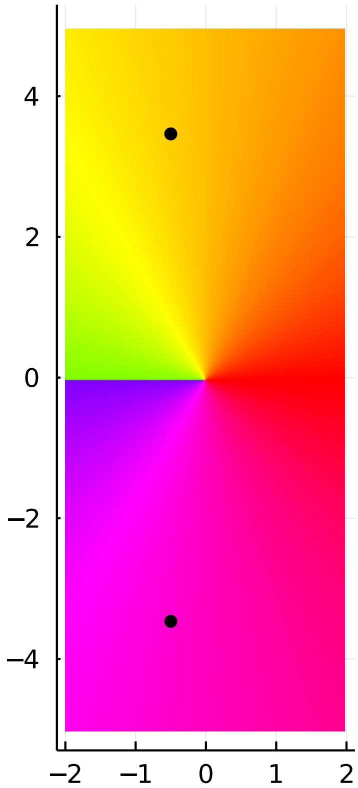

Then, using a slightly modified `sqrt_n` from Lecture 6 to compute the Taylor series and comparing with `sqrt(A)`, which is Julia's implementation of matrix square root, we see convergence, as the spectrum of `A` lies inside the disc of convergence of the Taylor series of $\sqrt{z}$ around 1:

```
function sqrt_n(n,z,z_0)
    ret = sqrt(z_0) * one(z)
    c = 0.5*inv(ret)*(z-z_0*I)
    for k=1:n
        ret += c
        c *= -(2k-1)/(2*(k+1)*z_0)*(z-z_0*I)
    end
    ret
end
norm(sqrt(A) - sqrt_n(100,A,1))
```

```
7.738033469234878e-16
```

But if the eigenvalues become much larger this won't work. Here we see an example of a matrix whose eigenvalues are not within the disc of convergence of the Taylor series:

```
A = [-0.5 -4; 3 -0.5]
λ = eigvals(A)
phaseplot(-2..2, -5..5, z -> sqrt(z))
scatter!(real(λ),imag(λ); color=:black, legend=false)
```




Using the partial sum of the Taylor series completely fails, and catastrophically so:

```
norm(sqrt(A) - sqrt_n(100,A,1))
```

```
2.1239250131186014e54
```

## 1.2 Diagonalisation/Jordan canonical form

**Definition (Diagonalisation of a matrix function)** Suppose $A = V\Lambda V^{-1}$ where

$$\Lambda = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_d \end{pmatrix}$$

Then define

$$f(A) := Vf(\Lambda)V^{-1} = V \begin{pmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_d) \end{pmatrix} V^{-1}$$

This works very well as a definition but is slow for large matrices and does not take advantage of sparsity.

We can extend it to Jordan canonical form:

**Definition (Jordan canonical form)** Suppose

$$A = V \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_{\tilde{d}} \end{pmatrix} V^{-1}$$

where $J_k$ are Jordan blocks, that is,

$$J_k = \begin{pmatrix} \lambda_k & 1 & & \\ & \ddots & \ddots & \\ & & \lambda_k & 1 \\ & & & \lambda_k \end{pmatrix}$$

Then define

$$f(A) := V \begin{pmatrix} f(J_1) & & \\ & \ddots & \\ & & f(J_{\tilde{d}}) \end{pmatrix} V^{-1}$$

using

$$f\left(\begin{pmatrix} \lambda & 1 & & \\ & \ddots & \ddots & \\ & & \lambda & 1 \\ & & & \lambda \end{pmatrix}\right) = \begin{pmatrix} f(\lambda) & f'(\lambda) & \cdots & f^{(m-1)}(\lambda)/(m-1)! \\ & \ddots & \ddots & \vdots \\ & & f(\lambda) & f'(\lambda) \\ & & & f(\lambda) \end{pmatrix}.$$

This definition can be verified to be consistent with Taylor series when both are valid by direct inspection. For example, we have

$$\begin{pmatrix} \lambda & 1 \\ & \lambda \end{pmatrix}^k = \begin{pmatrix} \lambda^k & k\lambda^{k-1} \\ & \lambda^k \end{pmatrix}.$$

## 1.3 Cauchy's integral formula

The last approach we detail in the next lecture is to use the Cauchy integral formula:

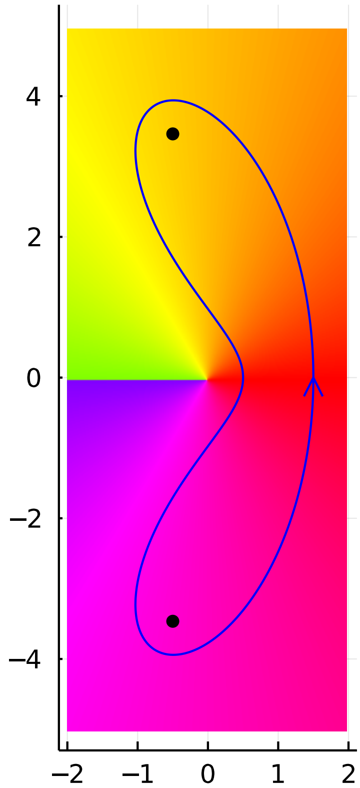
**Definition (Cauchy's integral matrix function)** Suppose $\gamma$ is a simple, closed contour that surrounds the spectrum of $A$ and $f$ is analytic in the interior. Then define

$$f(A) := \frac{1}{2\pi \mathrm{i}} \oint_\gamma f(\zeta)(\zeta I - A)^{-1} \mathrm{d}\zeta$$

Here the integrand is a matrix-valued function, hence consider the integral as defined entry-wise.

Before we explain where this comes from, we can test on a simple example.

```
A = [-0.5 -4; 3 -0.5]
λ= eigvals(A)
θ = range(0,2π,length=1000)
γ = θ -> (z = exp(im*θ);  2z + z^2 - 1.5/z) # a curve containing λ
phaseplot(-2..2, -5..5, z -> sqrt(z))
scatter!(real(λ),imag(λ); color=:black, legend=false)
plot!(real(γ.(θ)), imag(γ.(θ)); color=:blue, arrow=true)
```




We use the periodic Trapezium rule to integrate over the contour, showing it is valid:

```
periodic_rule(N) = 2π/N*(0:(N-1)), 2π/N*ones(N)
γp = θ -> (z = exp(im*θ); im*z*(2 + 2z + 1.5/z^2))
N = 1000
θ,w = periodic_rule(N)

norm(sum(w .* γp.(θ).*sqrt.(γ.(θ)) .* [inv(γ(θ)*I-A) for θ in θ])/(2π*im) - sqrt(A))
```

```
1.6279683670392478e-15
```

We will discuss more next lecture.