Note Title 2025-04-11

Newton method for the matrix square root:

$\boxed{X_0 = \alpha I \quad \text{or} \quad X_0 = \alpha M}$

$F(X) = X^2 - M$

$X_{k+1} = X_k - E$, where E solves $EX_k + X_k E = X_k^2 - M$

Modified Newton:

$$X_{k+1} = \frac{1}{2}\left(X_k + X_k^{-1} M\right)$$

Using the fact that $X_0 M = M X_0$, one can prove that $X_k M = M X_k$ for each k, and that the two iterations produce the same sequence X_k in exact arithmetic.

<u>Lemma</u>: suppose M has no negative eigenvalues, then $X_k \to M^{\frac{1}{2}}$, the principal square root of M

Multiply the modified Newton method equation by $M^{-\frac{1}{2}}$ to get

$$M^{-\frac{1}{2}} X_{k+1} = \frac{1}{2}\left(M^{-\frac{1}{2}} X_k + \underline{M^{-\frac{1}{2}} X_k^{-1} M}\right)$$

$$= \frac{1}{2}\left(M^{-\frac{1}{2}} X_k + X_k^{-1} M^{\frac{1}{2}}\right)$$ set $Y_k := M^{-\frac{1}{2}} X_k$ to get

$$Y_{k+1} = \frac{1}{2}\left(Y_k + Y_k^{-1}\right) \leftarrow \text{matrix sign iteration!}$$

So $\lim_{k\to\infty} Y_k = \operatorname{sign}(Y_0) = \operatorname{sign}\left(M^{-\frac{1}{2}} X_0\right)$

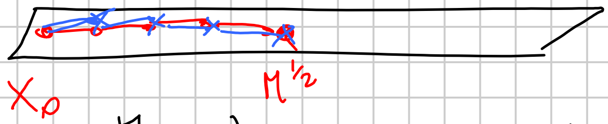
If $X_0 = \alpha M$, with $\alpha > 0$, then $M^{-\frac{1}{2}} X_0 = \alpha M^{\frac{1}{2}}$

Since $M^{\frac{1}{2}}$ has all eigenvalues in the RHP, $\operatorname{sign}(M^{-\frac{1}{2}} X_0) = I$

So $I = \lim_{k\to\infty} Y_k = \lim_{k\to\infty} M^{-\frac{1}{2}} X_k$

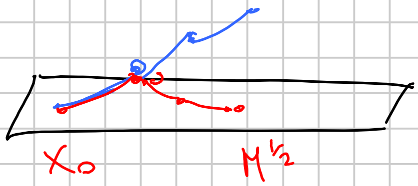
And $\lim X_k = M^{\frac{1}{2}}$

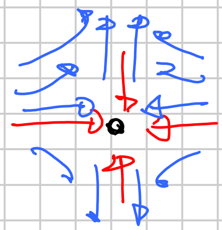
If $X_0 = \alpha I$, $Y_0 = \alpha M^{-\frac{1}{2}}$ and $\operatorname{sign}(Y_0) = I$ anyway.



← Matrices that commute with M

The iterations coincide on this manifold





But they have very different behaviour <u>outside</u> of the manifold: TN is attracting, MN is repulsive, so the iterates diverge even when started very close to $M^{1/2}$.

TN is quadratically convergent thanks to the general theory of Newton methods.

<u>Discrete-time dynamical system</u> associated to the map

$$F : \mathbb{C}^N \to \mathbb{C}^N$$

$$x_{k+1} = F(x_k) \qquad k = 0, 1, 2, 3, \dots$$

Suppose we start close to a fixed point $x_* = F(x_*)$

$x_0 = x_* + e$, with $e \in \mathbb{C}^N$ small.

$$x_1 = F(x_0) = F(x_* + e) = F(x_*) + \underbrace{J_{F,x_*}}_{N \times N \text{ matrix}} \cdot e + O(\|e\|^2)$$

$$= x_* + J_{F,x_*} e + O(\|e\|^2)$$

$$x_k = x_* + (J_{F,x_*})^k e + O(\|e\|^2)$$

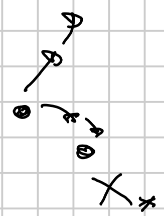
If $\rho(J_{F,X_*}) < 1$, the iterates get closer to X_*

$\Rightarrow X_*$ is an attractive fixed point

If $\rho(J_{F,X_*}) > 1$, then for most starting e the powers $(J_{F,X_*})^k e$ get larger and larger and the iterates do not approach X_*

$\Rightarrow X_*$ is a <u>repulsive</u> fixed point.

(cfr. analysis of 1-variable fixed-point methods).



For modified Newton, we have

$$X_{k+1} = \frac{1}{2}\left(X_k + X_k^{-1} M\right)$$

The Fréchet derivative of $F(X) = \frac{1}{2}(X + X^{-1}M)$

is obtained by

$$
\begin{aligned}
F(X+E) &= \frac{1}{2}\left(X + E + (X+E)^{-1} M\right) \\
&= \frac{1}{2}\left(X + E + \left(X^{-1} - X^{-1}EX^{-1} + X^{-1}EX^{-1}EX^{-1} + \dots\right) M\right) \\
&= \frac{1}{2}\left(X + X^{-1}M\right) + \frac{1}{2}\left(E - X^{-1}EX^{-1}M\right) + o(\|E\|)
\end{aligned}
$$

$$L_{F,X}[E] = \frac{1}{2}\left(E - X^{-1}EX^{-1}M\right)$$

$$L_{F,M^{1/2}}[E] = \frac{1}{2}\left(E - M^{-\frac{1}{2}} E M^{\frac{1}{2}}\right)$$

Using Kronecker products, we can see it as an $n^2 \times n^2$ matrix:

$$K = \frac{1}{2}\left(I_{n^2} - \left(M^{\frac{1}{2}}\right)^T \otimes M^{-\frac{1}{2}}\right) \qquad \operatorname{vec}(AXB) = \left(B^T \otimes A\right)\operatorname{vec} X$$

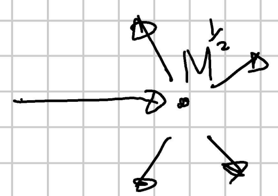
I can change basis so that this matrix becomes upper triangular:
take Schur forms $\left(M^{\frac{1}{2}}\right)^T = Q_1 U_1 Q_1^*$, $M^{-\frac{1}{2}} = Q_2 U_2 Q_2^*$

$$K = \frac{1}{2}\left(Q_1 \otimes Q_2\right)\left(I - U_1 \otimes U_2\right)\left(Q_1 \otimes Q_2\right)^*$$

On the diagonal, I can read off

$$\Lambda(K) = \left\{ \frac{1}{2}\left(1 - \lambda_i^{\frac{1}{2}} \cdot \lambda_j^{-\frac{1}{2}}\right) \; : \; i,j = 1, \dots n \right\} \text{ where } \lambda_1, \dots \lambda_n = \Lambda(M).$$

If M has two eigenvalues such that $\frac{\lambda_i}{\lambda_j}$ is large,
then $\rho(K) > 1 \Rightarrow M^{\frac{1}{2}}$ is a repulsive fixed point.



$$X_{k+1} = \frac{1}{2}\left(X_k + X_k^{-1} M\right)$$

Denman-Beavers iteration: set $X_k^{-1} M = Y_k^{-1}$ $\qquad Y_k = M^{-1} X_k$
to get

$$X_{k+1} = \frac{1}{2}\left(X_k + Y_k^{-1}\right)$$

$$Y_{k+1} = M^{-1} X_{k+1} = M^{-1} \cdot \frac{1}{2}\left(X_k + X_k^{-1} M\right)$$

$$= \frac{1}{2}\left(Y_k + X_k^{-1}\right)$$

$$\begin{cases} X_{k+1} = \frac{1}{2}\left(X_k + Y_k^{-1}\right) \\ Y_{k+1} = \frac{1}{2}\left(Y_k + X_k^{-1}\right) \end{cases} \qquad F: \mathbb{C}^{2n^2} \to \mathbb{C}^{2n^2}$$

The Jacobian of this iteration is idempotent: $J^2 = J$.

Functions of sparse matrices

$A \in \mathbb{C}^{n \times n}$ large and sparse

In general, $f(A)$ is going to be dense, but we can hope to compute $f(A)b$ for a vector b

e.g. to solve $\begin{cases} \dot{x} = Ax \\ x(0) = b \end{cases}$ we need to compute $x(t) = \exp(tA)b$

Problem: given f, A large and sparse, b, compute $f(A)b$.

1. Schur-based methods won't work
2. methods based on $f(A) \approx p(A)$ or $f(A) \approx r(A)$ (polynomial or rational approximation) work, because we can evaluate $p(A)b$, $q(A)^{-1}p(A)b$ even for large and sparse A.

$$A^3 b = A(A(Ab))$$

3. Cauchy formula:

$$f(A)b = \frac{1}{2\pi i}\int f(z)(zI - A)^{-1} b \, dz$$

$$\approx \sum_{k=1}^{N} w_k \underbrace{(z_k I - A)^{-1} b}_{\text{sparse linear system}}$$ for certain nodes z_k and weights w_k

This also reduces to a rational approximation:

$$r(z) = \sum_{k=1}^{N} w_k \frac{1}{z_k - z}$$

Problem in both cases: find $p(z)$ or $r(z)$ that approximates $f(z)$

in $\Lambda(A)$ (which itself is difficult to compute).

New idea: use Arnoldi.

Def: Given $A \in \mathbb{C}^{m \times m}$, $b \in \mathbb{C}^m$ $\quad m > n$

$$K_n(A,b) = \operatorname{span}\left(b, Ab, A^2b, \dots A^{n-1}b\right)$$
$$= \left\{ p(A)b : p \in \text{polynomials of degree } d < n \right\}$$

In many problems, we can get good approximations of the solution by projecting it onto a Krylov subspace.
Let $V_n \in \mathbb{C}^{m \times n}$ be a matrix whose columns are an orthonormal basis of $K_n(A,b)$, then to solve $Ax=b$, we project it to solve instead $\left(V_n^* A V_n\right) y = V_n^* b$
and then recover $x = V_n y$

Also, the eigenvalues of A are usually well approximated by those of $A_n = V_n^* A V_n \in \mathbb{C}^{n \times n}$ (<u>Ritz values</u>)
Usually $\Lambda(A_n)$ is a good approximation of the <u>outer</u> eigenvalues of A (those with largest modulus).

Arnoldi produces an orthonormal basis V_n for $K_n(A,b)$

$$V_n^* b = \begin{bmatrix} \beta \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = e_1 \beta \qquad \beta = \|b\|.$$

Lemma: For all polynomials of degree $d < n$,

$$\underbrace{p(A)b}_{} = V_n \, p(\underbrace{V_n^* A V_n}_{A_n}) \underbrace{V_n^* b}_{e_1\beta}.$$

$m \times m$ matrix $\quad$ $n \times n$ smaller matrix

Proof: it is enough to prove that $A^j b = V_n A_n^j V_n^* b$ for all $j = 0, 1, 2, \dots n-1$.

$\boxed{j=0}$: $b = V_n V_n^* b$? $\quad$ $V_n V_n^*$ is the orthogonal projection onto $K_n(A,b)$

Since $b \in K_n(A,b)$, the projection does nothing, so $b = V_n V_n^* b$

$\boxed{j=1}$ $\quad Ab \stackrel{?}{=} V_n V_n A V_n V_n^* b$?

$V_n V_n^* b = b$, as above

$V_n V_n^* A b = Ab$ because $Ab \in K_n(A,b)$

Similarly,

$$\underbrace{V_n V_n^* A V_n V_n^* A \cdots V_n V_n^*}_{j \text{ times } V_n^* A V_n} b = A^j b$$

as long as $j < n$, because we can cancel out these projections one by one: $V_n V_n^* A^j b = A^j b$ if $A^j b \in K_n(A,b)$. $\square$

$$\underbrace{p(A)}_{m \times m} b = V_n \underbrace{p(A_n)}_{n \times n} e_1 \beta \qquad A_n = V_n^* A V_n \in \mathbb{C}^{n \times n}$$

For a generic matrix function, we can hope that

$$f(A) b \approx V_n f(A_n) e_1 \beta =: c$$

is a good approximation.

Note that this is also a polynomial approximation:

$f(A_n) = p_n(A_n)$, where p_n is the interpolating polynomial for f on $\Lambda(A_n)$ of degree $d < n$.

This is <u>not</u> the interpolating polynomial $p(z)$ of degree m on $\Lambda(A)$ for which $f(A) = p(A)$, but one with smaller degree.

$$c = V_n f(A_n) e_1 \beta = V_n p_n(A_n) e_1 \beta = p_n(A) b$$

$$f(A) b = p(A) b$$

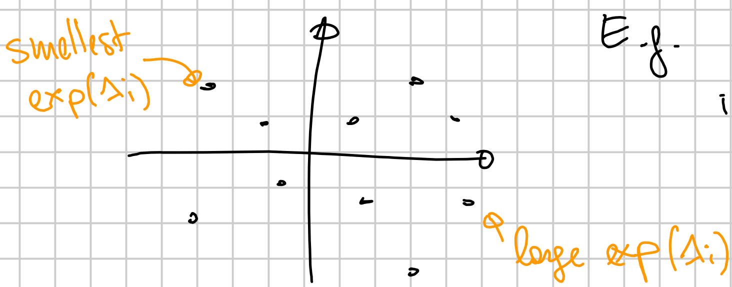
$$p_n(\mu) = f(\mu) \quad \text{for Ritz values} \quad \mu \in \Lambda(A_n)$$

The Ritz values approximate well <u>outer</u> eigenvalues, so this gives us a good approximation on those

$$f(A) b = V \begin{bmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_m) \end{bmatrix} V^* b$$

$$p_n(A) b = V \begin{bmatrix} p_n(\lambda_1) & & \\ & \ddots & \\ & & p_n(\lambda_m) \end{bmatrix} V^* b$$

This will be a good approximation if the eigenvalues for which $f(\lambda_i)$ is larger are <u>outer</u> eigenvalues



E.g. $f(\lambda_i) = \exp(\lambda_i)$ is larger when λ_i is <u>rightmost</u> in the complex plane

(Just an intuition for now, proofs in the next lecture)