Polar Decomposition

Thomas Seleiro

November 29, 2020

2. Prove that the singular values of A are the eigenvalues of H.

We know that any matrix $A \in \mathbb{C}^{m\times n}$, $m \geq n$ has a thin singular value decomposition $A = P\Sigma Q^*$ where $P \in \mathbb{C}^{m\times n}$ has orthogonal columns, $Q \in \mathbb{C}^{n\times n}$ is unitary, and $\Sigma \in \mathbb{C}^{n\times n}$ is diagonal with $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$ where $\mathrm{rank}(A) = r$ and $\sigma_1 \geq \ldots \geq \sigma_r \geq 0$, the singular values of A. Thus we can write

$$A = (PQ^*)(Q\Sigma Q^*) =: UH \tag{1}$$

where U and H satisfy the properties of a polar decomposition.

In particular we have $H = Q\Sigma Q^*$ where Σ is diagonal and Q is orthogonal. Thus the diagonal values of Σ are the eigenvalues of H, which are the singular values of A.

3. Prove that A is normal ($A^*A = AA^*$) iff U and H commute.

We first suppose that U and H commute. Note that for the product HU to be well defined, we must have $m = n$ which implies $U \in \mathbb{C}$ is unitary. Since $A = UH = HU$ we get

$$A^*A = (UH)^*(UH) = H^*(U^*U)H = H^2 \tag{2}$$

$$AA^* = (HU)(HU)^* = H(UU^*)H^* = H^2 \tag{3}$$

so A is normal.

Now suppose A is normal. Since $A^*A \in \mathbb{C}^{n\times n}$ and $AA^* \in \mathbb{C}^{m\times m}$, A normal requires $m = n$. Using the singular value decomposition of A, we have

$$AA^* = (P\Sigma Q^*)(Q\Sigma P^*) = P\Sigma^2 P^* \tag{4}$$

where $\Sigma^2 = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_r^2)$. Equating (2) and (4), we get $H^2 = P\Sigma^2 P^*$. From [1, p.405], we know that there is a unique Hermitian positive semi-definite matrix $(AA^*)^{1/2}$ such that $(AA^*)^{1/2}(AA^*)^{1/2} = AA^* = H^2$. It is obvious by its construction that H is said matrix, but we also note that $(P\Sigma P^*)(P\Sigma P^*) = P\Sigma^2 P^* = AA^*$. Therefore $H = P\Sigma P^*$ and

$$HU = (P\Sigma P^*)(PQ^*) = P\Sigma Q^* = A = UH \tag{5}$$

by the properties of the SVD of A. Therefore U and H commute.

4. Verify the formula

$$U = \frac{2}{\pi} A \int_0^\infty (t^2 I - A^* A)^{-1} dt \qquad (*)$$

for full rank A by using the singular value decomposition (SVD) of A to diagonalize the formula.

Since $A^*A = (Q\Sigma P^*)(P\Sigma Q^*) = Q\Sigma^2 Q^*$, we have

$$t^2 I + A^* A = Q(t^2 I)Q^* + Q\Sigma^2 Q^* = QDQ^* \qquad (6)$$

where $D := \operatorname{diag}(t^2 + \sigma_1, \ldots, t^2 + \sigma_n)$. Inverting (6) gives

$$(t^2 I + A^* A)^{-1} = QD^{-1}Q^*, \qquad D^{-1} = \operatorname{diag}\left(\frac{1}{\sigma_1^2 + t^2}, \ldots, \frac{1}{\sigma_n^2 + t^2}\right) \qquad (7)$$

Since Q and Q^* do not depend on t, they can be taken outside the integral, leaving the right hand side of (*) in the form

$$\frac{2}{\pi} A\, Q \int_0^\infty D^{-1} dt\, Q^* \qquad (8)$$

The integral is a diagonal matrix where the ith diagonal component is

$$\int_0^\infty \frac{1}{\sigma_i^2 + t^2}\, dt = \left[\frac{1}{\sigma_i} \arctan\left(\frac{t}{\sigma_i}\right)\right]_0^\infty = \frac{\pi}{2\sigma_i} \qquad (9)$$

using [2, 4.2.4.4]. So the right-hand side of (*) is

$$\begin{aligned} A\, Q \operatorname{diag}\left(\sigma_1^{-1}, \ldots, \sigma_n^{-1}\right) Q^* &= P\Sigma Q^*\, Q\Sigma^{-1} Q^* \\ &= PQ^* = U \end{aligned} \qquad (10)$$

5. Derive Newton's method for computing U by considering equations $(X + E) * (X + E) = I$, where E is a "small perturbation". (Newton's method is $X_{k+1} = (X_k + X_k^{-*})/2, X_0 = A$)

We know that U is the closest unitary matrix to A, and since $U^*U = I$, we try to find a solution to the equation

$$F(X) = 0, \qquad F(X) := X^* X - I \qquad (11)$$

using a Newton method starting at A. The general form of the Newton method [(3), p.] is

$$F(X_{k+1}) + DF_{X_k}[X_{k+1} - X_k] = 0 \qquad (12)$$

where DF_{X_k} is the Fréchet derivative and, is the first order E term in

$$F(X + E) - F(X) = X^* E + E^* X + E^* E. \qquad (13)$$

So $DF_{X_k}[E] = X^*E + E^*X$. Substituting in (12),

$$X_k^* X_k - I + X_k^* \left(X_{k+1} - X_k\right) + \left(X_{k+1}^* - X_k^*\right) X_k = 0 \tag{14}$$

$$X_k^* X_k - I + X_k^* X_{k+1} - X_k^* X_k + X_{k+1}^* X_k - X_k^* X_k = 0 \tag{15}$$

$$X_k^* X_{k+1} + X_{k+1}^* X_k = X_k^* X_k + I \tag{16}$$

We know that for any matrix we can write $B = 1/2(B + B^*) + 1/2(B - B^*)$, where the terms on the right-hand side are the Hermitian and skew Hermitian components respectively [1, p.170]. Setting the skew Hermitian part to zero, and taking $B = X_k^* X_{k+1}$ gives

$$X_k^* X_{k+1} = \frac{1}{2}\left(X_k^* X_k + I\right) \tag{17}$$

$$X_{k+1} = \frac{1}{2}\left(X_k + X_k^{-*}\right) \tag{18}$$

6. Prove that Newton's method converges, and at a quadratic rate, by using the SVD of A.

7. Use the SVD to analyze the convergence of the Newton-Schulz iteration for computing U:

$$X_{k+1} = \frac{1}{2} X_k (3I - X_k^* X_k), \qquad X_0 = A$$

8. Evaluate the operation count for one step of Newton's method and one step of the Newton-Schulz iteration (taking account of symmetry). Ignoring operation counts, how much faster does matrix multiplication have to be than matrix inversion for Newton-Schulz to be faster than Newton (assuming both take the same number of iterations)?

References

[1] Roger A Horn and Charles R Johnson. *Matrix analysis.* Cambridge University Press, Cambridge, 1985.

[2] Alan. Jeffrey and Hui-Hui Dai. *Handbook of mathematical formulas and integrals.* Academic Press/Elsevier, Burlington, MA, 4th ed. / edition, 2008.

[3] C T Kelley. *Solving nonlinear equations with Newton's method.* Fundamentals of Algorithms. Society for Industrial and Applied Mathematics, Philadelphia, 2003.