Project Polar

Kaustubh Roy

December 11, 2023

1 Part 1

I took some more time to read the papers as well after I mailed you the first time. I hope it has improved my answers.

2 Part 2

Let the SVD of A be $P\Sigma Q^*$, where P, Q are Unitary Matrices and Σ is a diagonal matrix containing the singular values of A: $\Sigma_{ii} = \sigma_i$.

Now, the polar decomposition of A: $A = UH$ can be constructed by using the property of the Unitary Matrix Q:

$$QQ^* = I$$

$$A = PQ^*Q\Sigma Q^*$$

$$A = (PQ^*)(Q\Sigma Q^*)$$

Now, since P and Q are both $Unitary$, PQ^* is also $Unitary$, and it can be trivially shown that $Q\Sigma Q^*$ is $Hermitian$. So, we have constructed the Polar Decomposition of A. Looking at the terms, $U = PQ^*$ and $H = Q\Sigma Q^*$.

Let λ, x be an eignenpair of H such that $Hx = \lambda x$. Then

$$Q\Sigma Q^*x = \lambda x$$

$$Q^*Q\Sigma Q^*x = Q^*\lambda x$$

$$\Sigma(Q^*x) = \lambda(Q^*x)$$

So, λ is an eigenvalue of Σ. Now, since Σ is a diagonal matrix, it's diagonal elements must be its eigenvalues. And, by construction of SVD, the diagonal elements of Σ are the singular values of A. Thus, the singular values of A are the eigenvalues of H.

3 Part 3

First, let us start with $A = UH$. Then $A^* = (UH)^* = H^*U^*$

$$A^*A = H^*U^*UH$$

Since U is $Unitary$, $U^*U = I$.

$$A^*A = H^*(U^*U)H$$
$$A^*A = H^*H$$
$$A^*A = H^2$$

Now, we need to examine AA^*:

$$AA^* = UHH^*U^*$$
$$AA^* = UH^2U^*$$

if $[U, H] \neq 0$: we are at an impasse... But, if U and H commute, we have:

$$UH = HU$$
$$AA^* = UH^2U^*$$
$$AA^* = (UH)HU^*$$
$$AA^* = H(UH)U^*$$
$$AA^* = HH(UU^*)$$
$$AA^* = H^2$$

So $AA^* = A^*A$ if $[U, H] = 0$ Alternatively: if $AA^* = A^*A$:

$$UH^2U^* = H^2$$
$$UH(U^*U)HU^* = H^2$$
$$(UHU^*)^2 = H^2$$

Taking principal square root:

$$UHU^* = H$$
$$UHU^*U = HU$$
$$UH = HU$$

4 Part 4

From construction, we know that the polar decomposition of A: $A = UH$ is related to the SVD $A = P\Sigma Q^*$ as:

$$U = PQ^*$$

$$H = Q\Sigma Q^*$$

In the formula:

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

We can substitute the SVD of A to get:

$$U = \frac{2}{\pi} P\Sigma Q^* \int_0^\infty (t^2 I + (P\Sigma Q^*)^*(P\Sigma Q^*))^{-1}\, dt$$

$$U = \frac{2}{\pi} P\Sigma Q^* \int_0^\infty (t^2 I + ((Q^*)^*\Sigma^* P^*)(P\Sigma Q^*))^{-1}\, dt$$

$$U = \frac{2}{\pi} P\Sigma Q^* \int_0^\infty (t^2 I + Q\Sigma^*(P^* P)\Sigma Q^*)^{-1}\, dt$$

$$U = \frac{2}{\pi} P\Sigma Q^* \int_0^\infty (t^2 (QQ^*) + Q\Sigma^*\Sigma Q^*)^{-1}\, dt$$

$$U = \frac{2}{\pi} P\Sigma Q^* \int_0^\infty (Q(t^2 I + \Sigma^*\Sigma)Q^*)^{-1}\, dt$$

Since $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$, let $A = Q$, $B = (t^2 I - \Sigma^*\Sigma)$ and $C = Q^*$. We get:

$$U = \frac{2}{\pi} P\Sigma Q^* \int_0^\infty (Q^*)^{-1}(t^2 I + \Sigma^*\Sigma))^{-1} Q^{-1}\, dt$$

$$U = \frac{2}{\pi} P\Sigma Q^* (Q^*)^{-1} (\int_0^\infty (t^2 I + \Sigma^*\Sigma))^{-1}\, dt) Q^{-1}$$

Substituting X as the integral in the middle:

$$U = \frac{2}{\pi} P\Sigma X Q^{-1}$$

Now, we need to compute X:

$$X = \int_0^\infty (t^2 I + \Sigma^*\Sigma))^{-1}\, dt$$

Now, Since Σ is a diagonal matrix of singular values of A, $\Sigma^* = \Sigma$. So:

$$X = \int_0^\infty (t^2 I + \Sigma^2))^{-1}\, dt$$

Σ^2 is a diagonal matrix with elements σ_i^2 where σ_i are the singular values of A. So, $t^2 I - \Sigma^2$ is a diagonal matrix with elements $t^2 - \sigma_i^2$. The inverse of a diagonal matrix is a diagonal matrix with its elements being equal to the reciprocal of the original matrix. So, if $\Lambda = t^2 I - \Sigma^2$, elements of Λ are $\lambda_i = (t^2 - \sigma_i^2)^{-1}$

$$X = \int_0^\infty \Lambda\, dt$$

Now, this is a diagonal integral and can be treated as a separate integral of each term. Each term individually integrates to:

$$X_{ii} = \int_0^\infty \frac{1}{t^2 - \sigma_i^2}\, dt$$

$$X_{ii} = \frac{1}{\sigma_i} \tanh^{-1} \frac{t}{\sigma_i} \Big|_0^\infty$$

$$X_{ii} = \frac{1}{\sigma_i} \tanh^{-1} \frac{\infty}{\sigma_i} - \frac{1}{\sigma_i} \tanh^{-1} \frac{0}{\sigma_i}$$

$$X_{ii} = \frac{-i\pi}{2\sigma_i}$$

$$X_{ii} = \frac{-i\pi}{2} \frac{1}{\sigma_i}$$

Now, The matrix with diagonal elements $\frac{1}{\sigma_i}$ is Σ^{-1}. So, the integral can be written as:

$$X = \frac{-i\pi}{2} \Sigma^{-1}$$

Plugging the result back into the original equation, we get:

$$U = \frac{2}{\pi} P\Sigma(\frac{-i\pi}{2}\Sigma^{-1})Q^{-1}$$

$$U = -iPQ^{-1}$$

Since $Q^{-1} = Q^*$

$$U = -iPQ^*$$

5 Part 5

I was not sure about the differences in the questions, since both required me to approach the problems using the SVD. I have clubbed them together.

$$A = UH = PQ^*Q\Sigma Q^* = P\Sigma Q^*$$

Taking $X_0 = A$: Newton Iteration Proceeds as:

$$X_1 = 0.5(X_0 + X_0^{-*})$$

Replacing X_0 with $P\Sigma Q^*$ we get:

$$X_1 = 0.5(P\Sigma Q^* + (P\Sigma Q^*)^{-*})$$

$$X_1 = 0.5(P\Sigma Q^* + ((P\Sigma Q^*)^{-1})^*)$$

$$X_1 = 0.5(P\Sigma Q^* + ((Q^*)^{-1}\Sigma^{-1}P^{-1})^*)$$

Now, since P and Q are orthogonal, $P^*P = Q^*Q = I$

$$X_1 = 0.5(P\Sigma Q^* + (Q\Sigma^{-1}P^{-1})^*)$$

$$X_1 = 0.5(P\Sigma Q^* + ((P^{-1})^*\Sigma^{-*}Q^*)$$

Now, since Σ is a diagonal matrix with positive real values, $\Sigma^* = \Sigma$

$$X_1 = 0.5(P\Sigma Q^* + P\Sigma^{-1}Q^*)$$

$$X_1 = 0.5P(\Sigma + \Sigma^{-1})Q^*$$

$$X_1 = P\frac{1}{2}(\Sigma + \Sigma^{-1})Q^*$$

Let $\Lambda_1 = \frac{1}{2}(\Sigma + \Sigma^{-1})$ it can be shown that:

$$X_2 = P\Lambda_2 Q^*$$

where $\Lambda_2 = \frac{1}{2}(\Lambda_1 + \Lambda_1^{-1})$ and the same iteration follows. All the iterations of Λ_k are diagonal matrices with $\Lambda_0 = \Sigma$. So, we can compute Λ_k as a set of uncoupled sequences of diagonal elements:

$$(\lambda_k)_i = \frac{1}{2}((\lambda_{k-1})_i + \frac{1}{(\lambda_{k-1})_i})$$

This is the Newton iteration method for finding the square root of a. Where the iterations followed:

$$X_{k+1} = \frac{1}{2}(X_k + \frac{a}{X_k})$$

With a = 1. So, each of these iterations should converge to $\sqrt{1} = 1$.

6 Part 6

To prove convergence of the iterations, let us denote the iterative map for the singular values as:

$$(\lambda_k)_i = \frac{1}{2}((\lambda_{k-1})_i + \frac{1}{(\lambda_{k-1})_i})$$

$$(\lambda_k)_i = \frac{1}{2(\lambda_{k-1})_i}((\lambda_{k-1})_i^2 + 1)$$

$$(\lambda_k)_i - 1 = \frac{1}{2(\lambda_{k-1})_i}((\lambda_{k-1})_i^2 + 1 - 2(\lambda_{k-1})_i)$$

Now, $((\lambda_{k-1})_i^2 + 1 - 2(\lambda_{k-1})_i)$ can be written as $((\lambda_{k-1})_i - 1)^2$

$$(\lambda_k)_i - 1 = \frac{1}{2(\lambda_{k-1})_i}((\lambda_{k-1})_i - 1)^2$$

Similarly:

$$(\lambda_k)_i + 1 = \frac{1}{2(\lambda_{k-1})_i}((\lambda_{k-1})_i + 1)^2$$

Dividing the two expressions, we are left with:

$$\frac{(\lambda_k)_i - 1}{(\lambda_k)_i + 1} = \frac{((\lambda_{k-1})_i - 1)^2}{((\lambda_{k-1})_i + 1)^2}$$

We can then repeat this iteration backward $k-2$ more times and reach let's call this (eq 6.1):

$$\frac{(\lambda_k)_i - 1}{(\lambda_k)_i + 1} = \left(\frac{(\lambda_0)_i - 1}{(\lambda_0)_i + 1}\right)^{2k-1}$$

Now, since $(\lambda_0)_i$ are the singular values of A, and A is non-singular, all of them are positive definite. So, the fraction:

$$\left(\frac{(\lambda_0)_i - 1}{(\lambda_0)_i + 1}\right) \leq 1$$

for all i. Therefore, as $k \to \infty$:

$$\frac{(\lambda_k)_i - 1}{(\lambda_k)_i + 1} \to 0$$

equivalently $(\lambda_k)_i \to 1$ for all i.

Thus the diagonal matrix $\Lambda_k \to I$ and $X_k \to U$. At each iteration, X_k can be written as $X_k = P\Lambda_k Q^*$ So, the diagonal elements of Λ_k will be the singular values of X_k. Thus, we can analyze the convergence by studying the convergence of $\Lambda_k \to I$. From eq 6.1 (if we write it in matrix form):

$$||(\Lambda_k + I)^{-1}(\Lambda_k - I)||_2 = \max_{1 \leq i \leq n} \left(\frac{(\lambda_0)_i - 1}{(\lambda_0)_i + 1}\right)^{2k-1}$$

If we call $\frac{(\lambda_0)_i - 1}{(\lambda_0)_i + 1} = \delta_i$:

$$||(\Lambda_k + I)^{-1}(\Lambda_k - I)||_2 = \max_{1 \leq i \leq n} \delta_i^{2k-1}$$

Since all $\delta_i \leq 1$, for each iteration, the matrix Λ approaches Identity quadratically.

7 Part 7

In order to solve this, we can approach the problem using the same approach we used in the previous section. Taking the Iteration:

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

We begin by taking $X_0 = A$ and use the SVD of A: $A = P\Sigma Q^*$ The iteration yields:

$$X_1 = \frac{1}{2} A(3I - A^* A)$$

$$X_1 = \frac{1}{2} P\Sigma Q^* (3I - (P\Sigma Q^*)^* (P\Sigma Q^*))$$

Simplifying this expression, we get:

$$X_1 = \frac{1}{2} P\Sigma Q^* (3I - (Q\Sigma P^*)(P\Sigma Q^*))$$

$$X_1 = \frac{1}{2} P\Sigma Q^* (3I - (Q\Sigma^2 Q^*))$$

Using the identiy $QQ^* = I$:

$$X_1 = \frac{1}{2} P\Sigma Q^* (3QQ^* - (Q\Sigma^2 Q^*))$$

$$X_1 = \frac{1}{2} P\Sigma Q^* (Q3IQ^* - (Q\Sigma^2 Q^*))$$

$$X_1 = \frac{1}{2} P\Sigma Q^* (Q(3I - \Sigma^2)Q^*)$$

$$X_1 = \frac{1}{2} P\Sigma Q^* Q(3I - \Sigma^2)Q^*$$

$$X_1 = \frac{1}{2} P\Sigma (3I - \Sigma^2)Q^*$$

$$X_1 = P(\frac{1}{2}\Sigma(3I - \Sigma^2))Q^*$$

Now, let us name $(\frac{1}{2}\Sigma(3I - \Sigma^2))$ as Λ_0. And Λ_0 is a diagonal real matrix, similar to Σ. So, We can write $X_1 = P\Lambda_0 Q^*$ Now, using the same method and substitutions, we can show that:

$$X_2 = P(\frac{1}{2}\Lambda_0(3I - \Lambda_0^2))Q^*$$

giving us another expression of the similar form $X_2 = P\Lambda_1 Q^*$

Thus, we have a similar sequence of individual diagonal terms which must converge to unity, for our iterations to converge to Identity.

The iteration in this case is of the form:

$$\lambda_{k+1}^i = \frac{1}{2}\lambda_k^i (3 - (\lambda_k^i)^2)$$

where λ_0^i is the $i-th$ singular value of A. Each singular value must converge to 1 for the iteration to converge to U. Let us perform the analysis of the iterative map:

$$x_{k+1} = \frac{1}{2} x_k (3 - x_k^2)$$

to find the suitable conditions for x_0. Firstly, this map has three fixed points, $0, +1$, and -1. We must find the condition under which the iterations converge to $+1$.

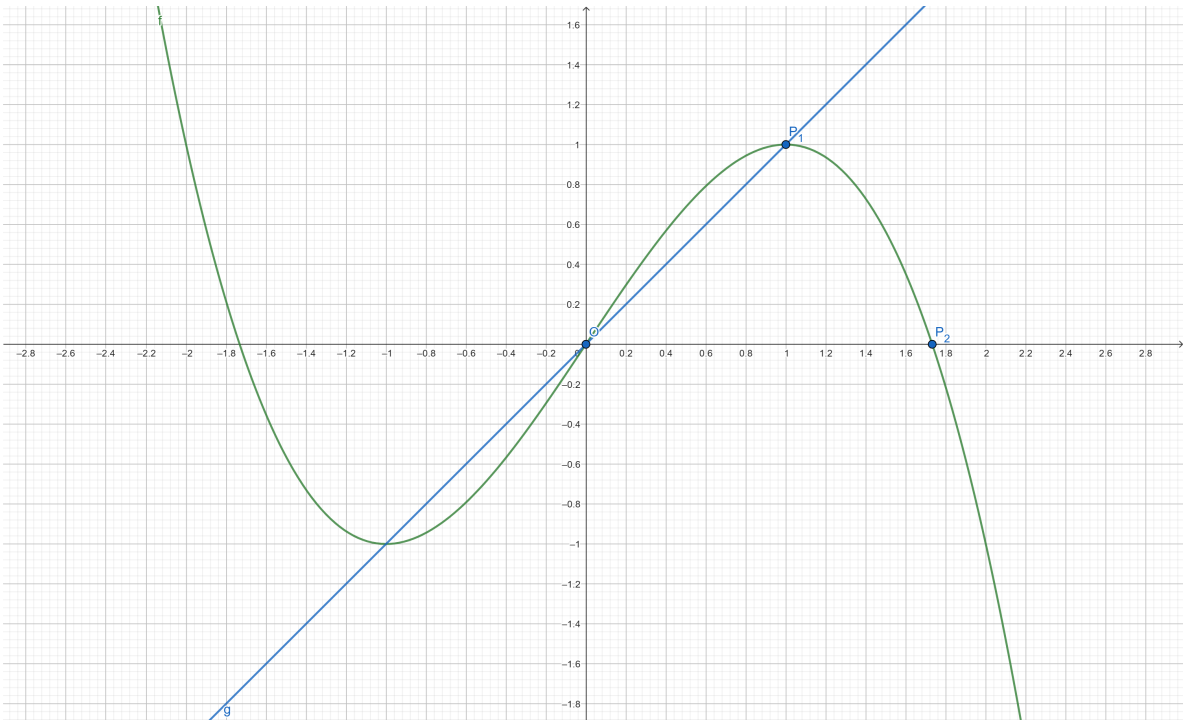

Figure 1: The functions $f : y = \frac{1}{2}x(3 - x^2)$ and $g : y = x$ are plotted. Points O, P_1 and P_2 are $(0, 0)$, $(1, 1)$ and $(\sqrt{3}, 0)$ respectively.

As is evident from the figure, the map has 3 fixed points, of which $x^* = 1$, and $x^* = -1$ are attractive fixed points, and $x^* = 0$ is a repelling fixed point. To show this, We can show that $f'(x)$ at $x = 1$ and $x = -1$ is less than 1. It is 0. for $x = 0$, however, it is $\frac{3}{2}$.

Between the points O and P_1, $f > g$, so, any iteration that enters the region between 0 and 1 will end up at P_1. Similarly, any iteration that enters the region between 0 and -1 will end up attracted to -1. As is evident, initial values up to P_2 are guaranteed to enter the region $(0, 1)$ and will be attracted to $x^* = 1$. The points immediately beyond P_2 will enter a region where they will be attracted to $x* = -1$. and then followed by a region from which the points are attracted to $x^* = 1$ again, till they reach a point where they diverge completely. We can obtain the exact points that mark the complicated basins of attraction of the two fixed points, but it will suffice for this problem to only regard the first region where the points are guaranteed to converge to $x^* = 1$. Thus, the largest singular value of A must be less than $\sqrt{3}$.

8 Part 8

Assuming Gauss-Jordan Elimination to compute the inverse of the square matrix of size n, and ignoring the addition operations, we have an operation count of $\frac{2n^3}{3}$. For Matrix Multiplication, we have an operation count of $2n^3$. But, for the first multiplication of Newton-Schulz, the term inside the bracket $X_k^* X_k$, due to symmetry, would only need half the number of operations. And, the second multiplication, $X_k * (3I - X_k^* X_k)$, would be a regular matrix multiplication with $2n^3$ operations. So, the total operation count for one step of Newton-Schulz is $3n^3$.

So, one step of Newton-Schulz takes 4.5 times the number of operations of a step of Newton method.

Ignoring operation counts, however, in a step of Newton iteration, we have one matrix inversion, while a step of Newton-Schulz has two matrix multiplication steps. So, if matrix multiplication is more than twice as fast as matrix inversion, Newton-Schulz would be a faster operation.

9 Part 9

The condition for convergence of the Newton-Schulz iteration is that the largest singular value of the matrix X_0, from which we start the iteration, be smaller than $\sqrt{3}$. This means that the 2-norm of the matrix must be smaller than $\sqrt{3}$. But, it is expensive to calculate the 2-norm of a Matrix, and doing it at every step of the Newton iterations would be very inefficient. One can estimate the 2-norm of a Matrix by using the identity $|A|_2 \leq \sqrt{|A|_1 . |A|_\infty}$ since calculating $|A|_1$ and $|A|_\infty$ are cheaper operations.

The remarks for running the poldec function, which is written in poldec.m on the given matrices are in the following subsections.

9.1 rand(10)

The method took 9 iterations and did not switch to N-S. All of the iterations were Newton-iterations.

9.2 eye(8)

The method did not iterate at all since the convergence condition was satisfied with $X_0 = I_8$

9.3 hilb(6)

The program took 29 iterations to complete. It switched to N-S iterations from the 23rd iteration.

9.4 magic(6)

The program ran 56 iterations but did not converge to a polar decomposition. The reason is that the magic matrices with even size are singular (except for size 2). In the case of magic(6), the rank is 5. Warning: Matrix is close to singular or badly scaled. Results may be inaccurate. RCOND = 2.000402e-18.

9.5 hadamard(8)

The program converged in 6 iterations. Did not switch to N-S iterations.

10 Part 10

For a symmetric positive definite matrix A, the Cholesky decomposition breaks the matrix into factors: $A = LL^*$, where L is a Lower Triangular matrix (For Symmetric matrices, L would be real, so $L^* = L'$). Now:

$$A = LL^*$$

Let L^* have the polar decomposition $L^* = UH$. Then, $LL^* = (UH)^*.(UH)$

$$LL^* = H^*(U^*U)H$$

$$LL^* = H^*H$$

Since $H^* = H$

$$LL^* = H^2$$

So, $A = H^2$ and H is the square root of the matrix A.

Code is written in sqrt.m. Note: the function chol in Matlab returns the upper triangular matrix L^* instead of L.