

1. Understanding Newton-Schulz

Let us consider a parameter matrix $W \in \mathbb{R}^{d_{\text{out}} \times d_{\text{in}}}$. Define the degree-3 odd polynomial p as:

$$p(W) = \frac{1}{2} \left(3I_{d_{\text{out}}} - WW^T \right) W.$$

In this problem, we will study how the iteration $W_{k+1} = p(W_k)$ affects the singular values of W_k .

- (a) **Show that the iteration acts only on the singular values of W .** i.e. if $W = U\Sigma V^T$ is the SVD, then show that

$$p(W) = Up(\Sigma)V^T.$$

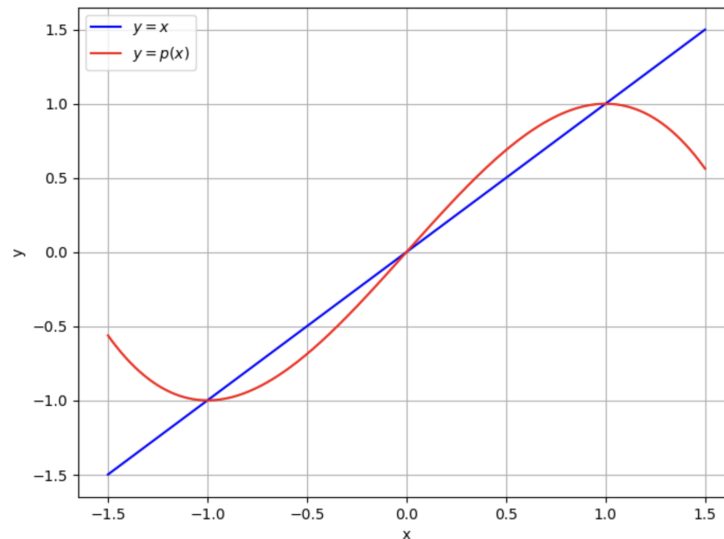
Hint: First show that $WW^T = U(\Sigma\Sigma^T)U^T$.

- (b) Write down the fixed point equation for $p(x) = \frac{3}{2}x - \frac{1}{2}x^3$. **Solve for all fixed points**, i.e. x^* such that $x^* = p(x^*)$.

- (c) We define a fixed point x^* of $p(x)$ as *locally stable* if $|\frac{d}{dx}p(x^*)| < 1$. First, convince yourself that a stable fixed point means that the distance towards the fixed point decreases with more iterations. **Determine which fixed points of $p(x)$ are stable and which are unstable.**

- (d) Below are plots of $y = p(x)$ and $y = x$. Pick different starting points for x and **show graphically how iteration $x = p(x)$ eventually converges to a stable fixed point.**

Use “cobweb diagram” to show how x evolves over time.



- (e) Suppose the singular value starts as $+\sigma$. **For which values of σ does it converge to $+1$? What does it do for other values?**
- (f) **Explain why this iteration can be viewed as an approximate way to make W closer to an orthogonal matrix** (with singular values near ± 1) and what we must ensure before using the iterations.

Contributors:

- Joey Hong.
- Anant Sahai.
- Gireeja Ranade.