Exercise 4: Iterative PCA (10 P)

When performing principal component analysis, computing the full eigendecomposition of the scatter matrix $\boldsymbol{S}$ is typically slow, and we are often only interested in the first principal components. An efficient procedure to find the first principal component is *power iteration*. It starts with a random unit vector $\boldsymbol{w}^{(0)} \in \mathbb{R}^d$, and iteratively applies the parameter update

$$\boldsymbol{w}^{(t+1)} = \boldsymbol{S}\boldsymbol{w}^{(t)} / \|\boldsymbol{S}\boldsymbol{w}^{(t)}\|$$

until some convergence criterion is met. Here, we would like to show the exponential convergence of power iteration. For this, we look at the error terms

$$\mathcal{E}_k(\boldsymbol{w}) = \left|\frac{\boldsymbol{w}^\top \boldsymbol{u}_k}{\boldsymbol{w}^\top \boldsymbol{u}_1}\right| \qquad \text{with} \quad k = 2, \dots, d,$$

and observe that they should all converge to zero as $\boldsymbol{w}$ approaches the eigenvector $\boldsymbol{u}_1$ and becomes orthogonal to other eigenvectors.

(a) Show that $\mathcal{E}_k(\boldsymbol{w}^{(T)}) = |\lambda_k/\lambda_1|^T \cdot \mathcal{E}_k(\boldsymbol{w}^{(0)})$, i.e. the convergence of the algorithm is exponential with the number of time steps T.

Proof by induction:

1) For $T=0$, it is clearly true

Let $T=1$, then $\mathcal{E}_k(w^{(1)}) = \left|\frac{w^{(1)T} u_k}{w^{(1)T} u_1}\right| = \left|\frac{(Sw^{(0)})^T u_k}{(Sw^{(0)})^T u_1}\right| = \left|\frac{w^{(0)T} \overbrace{S^T u_k}^{\lambda_k u_k}}{w^{(0)T} \underbrace{S^T u_1}_{\lambda_1 u_1}}\right| = \left|\frac{\lambda_k}{\lambda_1}\right| \cdot \left|\frac{w^{(0)T} u_k}{w^{(0)T} u_1}\right|$

$= \left|\frac{\lambda_k}{\lambda_1}\right| \mathcal{E}_k(w^{(0)})$

2) Assume that for some $T \geq 0$ we have $\mathcal{E}_k(w^{(T)}) = \left|\frac{\lambda_k}{\lambda_1}\right|^T \mathcal{E}_k(w^{(0)})$

3) $\mathcal{E}_k(w^{(T+1)}) = \left|\frac{w^{(T+1)T} u_k}{w^{(T+1)T} u_1}\right| = \left|\frac{(Sw^{(T)})^T u_k}{(Sw^{(T)})^T u_1}\right| = \left|\frac{w^{(T)T} S^T u_k}{w^{(T)T} S^T u_1}\right| = \left|\frac{\lambda_k}{\lambda_1}\right| \cdot \underbrace{\left|\frac{w^{(T)T} u_k}{w^{(T)T} u_1}\right|}_{\mathcal{E}_k(w^{(T)})} =$

hypothesis
$= \left|\frac{\lambda_k}{\lambda_1}\right|^{T+1} \cdot \mathcal{E}_k(w^{(0)})$

Using mathematical induction, we have proven that $\forall T \in \mathbb{N}$

$$\mathcal{E}_k(w^{(T)}) = \left|\frac{\lambda_k}{\lambda_1}\right|^T \mathcal{E}_k(w^{(0)}) \qquad \square$$