

Derivative of the first order SVD

Randolf Scholz

May 11, 2023

Consider computing the first order SVD expansion. By the Eckart–Young–Mirsky theorem, this is equivalent to solving

$$\underset{\sigma, u, v}{\text{minimize}} \frac{1}{2} \|A - \sigma uv^\top\|_F^2 \quad \text{s.t.} \quad \|u\| = 1 \quad \text{and} \quad \|v\| = 1$$

Any triplet (σ, u, v) for a **unique** singular value satisfies

The jacobian and lagrangian The derivative of the objective function is

$$\mathbf{J}_f(A, \begin{bmatrix} \sigma \\ u \\ v \end{bmatrix}) = \begin{bmatrix} A - \sigma uv^\top & \sigma - u^\top Av \\ \sigma^2 u - \sigma Av & \\ \sigma^2 v - \sigma A^\top u & \end{bmatrix} \implies \mathbf{H}_f(\begin{bmatrix} \sigma \\ u \\ v \end{bmatrix}) = \begin{bmatrix} 1 & 2\sigma u - Av & 2\sigma v - A^\top u \\ -Av & \sigma^2 \mathbb{I}_m & -\sigma A \\ -A^\top u & -\sigma A^\top & \sigma^2 \mathbb{I}_n \end{bmatrix}$$

Consider the function

$$f(A, \begin{bmatrix} \sigma \\ u \\ v \end{bmatrix}) = \begin{pmatrix} \sigma - u^\top Av \\ \sigma^2 u - \sigma Av \\ \sigma^2 v - \sigma A^\top u \end{pmatrix} \equiv \mathbf{0} \implies \mathbf{J}_f(A, \begin{bmatrix} \sigma \\ u \\ v \end{bmatrix}) = \begin{bmatrix} -\xi v u^\top & 1 & 2\sigma u - Av & 2\sigma v - A^\top u \\ -\sigma v \phi^\top & -Av & \sigma^2 \mathbb{I}_m & -\sigma A \\ -\sigma u \psi^\top & -A^\top u & -\sigma A^\top & \sigma^2 \mathbb{I}_n \end{bmatrix}$$

Thus, gradient descent schema is

$$\begin{aligned} \sigma' &= \sigma - \eta_\sigma (\sigma - u^\top Av) \\ u' &= u - \eta_u (\sigma^2 u - \sigma Av) \\ v' &= v - \eta_v (\sigma^2 v - \sigma A^\top u) \end{aligned}$$

And the newton step with diagonal approximation of the hessian:

$$\begin{aligned} \sigma' &= \sigma - 1(\sigma - u^\top Av) &= u^\top Av \\ u' &= u - \frac{1}{\sigma^2} (\sigma^2 u - \sigma Av) &= \frac{1}{\sigma} Av \\ v' &= v - \frac{1}{\sigma^2} (\sigma^2 v - \sigma A^\top u) &= \frac{1}{\sigma} A^\top u \end{aligned}$$

$$\begin{aligned}
\sigma &= u^\top Av \\
Av &= \sigma u \\
A^\top u &= \sigma v
\end{aligned}$$

from this we can derive

$$\begin{aligned}
\Delta\sigma &= \Delta u^\top Av + u^\top \Delta Av + u^\top A \Delta v \\
&= \Delta u^\top u + u^\top \Delta Av + v^\top \Delta v \\
&= u^\top \Delta Av
\end{aligned}$$

Where in the last step we used $\Delta u \perp u$ and $\Delta v \perp v$, which follows from the side condition. Further we have:

$$\begin{aligned}
\Delta Av + A \Delta v &= \Delta\sigma u + \sigma \Delta u \\
\Delta A^\top u + A^\top \Delta u &= \Delta\sigma v + \sigma \Delta v
\end{aligned}
\iff \underbrace{\begin{bmatrix} \sigma \mathbb{I}_m & -A \\ -A^\top & \sigma \mathbb{I}_n \end{bmatrix}}_{=:K} \cdot \begin{bmatrix} \Delta u \\ \Delta v \end{bmatrix} = \begin{bmatrix} \Delta Av - \Delta\sigma u \\ \Delta A^\top u - \Delta\sigma v \end{bmatrix}$$

1 The VJP

The last equation allows us to compute the VJP at ease:

$$\begin{aligned}
\left\langle \begin{bmatrix} \phi \\ \psi \end{bmatrix} \middle| \begin{bmatrix} \Delta u \\ \Delta v \end{bmatrix} \right\rangle &= \left\langle \begin{bmatrix} \phi \\ \psi \end{bmatrix} \middle| K^{-1} \begin{bmatrix} \Delta Av - \Delta\sigma u \\ \Delta A^\top u - \Delta\sigma v \end{bmatrix} \right\rangle \\
&= \left\langle K^{-\top} \begin{bmatrix} \phi \\ \psi \end{bmatrix} \middle| \begin{bmatrix} \Delta Av - \Delta\sigma u \\ \Delta A^\top u - \Delta\sigma v \end{bmatrix} \right\rangle \\
&= \left\langle \begin{bmatrix} \tilde{\phi} \\ \tilde{\psi} \end{bmatrix} \middle| \begin{bmatrix} \Delta Av - \Delta\sigma u \\ \Delta A^\top u - \Delta\sigma v \end{bmatrix} \right\rangle
\end{aligned}$$

Now, we compute the terms individually:

$$\begin{aligned}
\langle \tilde{\phi} \mid \Delta Av - \Delta\sigma u \rangle &= \langle \tilde{\phi} v^\top \mid \Delta A \rangle - \langle u^\top \tilde{\phi} \mid \Delta\sigma \rangle \\
&= \langle \tilde{\phi} v^\top \mid \Delta A \rangle - \langle u^\top \tilde{\phi} \mid u^\top \Delta Av \rangle \\
&= \langle (\mathbb{I}_m - uu^\top) \tilde{\phi} v^\top \mid \Delta A \rangle
\end{aligned}$$

And for the second term we get

$$\begin{aligned}
\langle \tilde{\psi} \mid \Delta A^\top u - \Delta\sigma v \rangle &= \langle \tilde{\psi} u^\top \mid \Delta A^\top \rangle - \langle v^\top \tilde{\psi} \mid \Delta\sigma \rangle \\
&= \langle u \tilde{\psi}^\top \mid \Delta A \rangle - \langle \tilde{\psi}^\top v \mid u^\top \Delta Av \rangle \\
&= \langle u \tilde{\psi} (\mathbb{I}_n - vv^\top) \mid \Delta A \rangle
\end{aligned}$$

Using the formula for inverting a block-matrix, we can give an explicit solution to $K^{-\top}$:

$$\begin{aligned} \begin{bmatrix} \sigma \mathbb{I}_m & -A \\ -A^\top & \sigma \mathbb{I}_n \end{bmatrix}^{-1} &= \begin{bmatrix} (\sigma \mathbb{I}_m - \frac{1}{\sigma} A A^\top)^{-1} & \mathbf{0}_{m \times n} \\ \mathbf{0}_{n \times m} & (\sigma \mathbb{I}_n - \frac{1}{\sigma} A^\top A)^{-1} \end{bmatrix} \cdot \begin{bmatrix} \mathbb{I}_m & \frac{1}{\sigma} A \\ \frac{1}{\sigma} A^\top & \mathbb{I}_n \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{\sigma} (\mathbb{I}_m - \frac{1}{\sigma^2} A A^\top)^{-1} & \frac{1}{\sigma^2} (\mathbb{I}_m - \frac{1}{\sigma^2} A A^\top)^{-1} A \\ \frac{1}{\sigma^2} (\mathbb{I}_n - \frac{1}{\sigma^2} A^\top A)^{-1} A^\top & \frac{1}{\sigma} (\mathbb{I}_n - \frac{1}{\sigma^2} A^\top A)^{-1} \end{bmatrix} \\ &= \frac{1}{\sigma} \begin{bmatrix} (\mathbb{I}_m - \tilde{A} \tilde{A}^\top)^{-1} & (\mathbb{I}_m - \tilde{A} \tilde{A}^\top)^{-1} \tilde{A} \\ (\mathbb{I}_n - \tilde{A}^\top \tilde{A})^{-1} \tilde{A}^\top & (\mathbb{I}_n - \tilde{A}^\top \tilde{A})^{-1} \end{bmatrix} \end{aligned}$$

And we see it's basically projection operators with respect to the image/kernel of $\tilde{A} = \frac{1}{\sigma} A$. In summary, we obtain the following formula for the VJP:

$$\begin{aligned} \begin{bmatrix} \sigma \mathbb{I}_m & -A \\ -A^\top & \sigma \mathbb{I}_n \end{bmatrix}^{-\top} \begin{bmatrix} \tilde{\phi} \\ \tilde{\psi} \end{bmatrix} &= \begin{bmatrix} \phi \\ \psi \end{bmatrix} \iff \begin{bmatrix} \tilde{\phi} \\ \tilde{\psi} \end{bmatrix} = \frac{1}{\sigma} \begin{bmatrix} (\mathbb{I}_m - \tilde{A} \tilde{A}^\top)^{-1} & (\mathbb{I}_m - \tilde{A} \tilde{A}^\top)^{-1} \tilde{A} \\ (\mathbb{I}_n - \tilde{A}^\top \tilde{A})^{-1} \tilde{A}^\top & (\mathbb{I}_n - \tilde{A}^\top \tilde{A})^{-1} \end{bmatrix}^\top \begin{bmatrix} \phi \\ \psi \end{bmatrix} \\ &\iff \begin{bmatrix} \tilde{\phi} \\ \tilde{\psi} \end{bmatrix} = \frac{1}{\sigma} \begin{bmatrix} (\mathbb{I}_m - \tilde{A} \tilde{A}^\top)^{-1} & \tilde{A} (\mathbb{I}_n - \tilde{A}^\top \tilde{A})^{-1} \\ \tilde{A}^\top (\mathbb{I}_m - \tilde{A} \tilde{A}^\top)^{-1} & (\mathbb{I}_n - \tilde{A}^\top \tilde{A})^{-1} \end{bmatrix} \begin{bmatrix} \phi \\ \psi \end{bmatrix} \end{aligned}$$

Thus, we need to solve 4 linear systems:

$$\begin{aligned} (\mathbb{I}_m - \tilde{A} \tilde{A}^\top) x &= \phi & (\mathbb{I}_n - \tilde{A}^\top \tilde{A}) y &= \tilde{A} \psi \\ (\mathbb{I}_m - \tilde{A} \tilde{A}^\top) z &= \tilde{A}^\top \phi & (\mathbb{I}_n - \tilde{A}^\top \tilde{A}) w &= \psi \end{aligned}$$

Then $\tilde{\phi} = \frac{1}{\sigma}(x + y)$ and $\tilde{\psi} = \frac{1}{\sigma}(z + w)$, and the VJP are given by the previous equations:

$$\begin{aligned} \xi^\top \frac{\partial \sigma}{\partial A} &= \xi u v^\top \\ \phi^\top \frac{\partial u}{\partial A} &= (\mathbb{I}_m - u u^\top) \tilde{\phi} v^\top \\ \psi^\top \frac{\partial v}{\partial A} &= u \tilde{\psi} (\mathbb{I}_n - v v^\top) \end{aligned}$$