Derivative of the first order SVD

Randolf Scholz

May 14, 2023

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

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

Any triplet $(\sigma,\, u,\, v)$ for a **unique** singular value satisfies

The Jacobian and Lagrangian The derivative of the objective function is

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

Consider the function

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

Thus, gradient descent schema is

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

And the newton step with diagonal approximation of the hessian:

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

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

from this we can derive

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

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:

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

1 The VJP

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

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

Now, we compute the terms individually:

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

And for the second term we get

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

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

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

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:

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

We can use the **push-through identity** to convert these into 4 linear systems:

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

We can use the **push-through identity** to convert these into 4 linear systems:

$$
\begin{aligned}
Px &= \phi & \qquad Py &= \tilde{A}\psi \\
Qz &= \tilde{A}^\top\phi & \qquad Qw &= \psi
\end{aligned}
$$

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:

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