Bonus Question

Let A be an $m \times n$ matrix of rank $k \leq min(m, n)$. Then we may decompose A as $A = USV^T$, where:

- U is $m \times k$ **orthonormal** matrix.
- S is $k \times k$ **diagonal** matrix, of which its diagonal entries are called the singular values of A. We denote them by $\sigma_i = S_{ii}$.
- V is $n \times k$ **orthonormal** matrix.

This decomposition is called the singular-value decomposition of A.

Jojo, Roy's African grey parrot, has found a function $f_i: \mathbb{R}^{m \times n} \to \mathbb{R}$, which is given by $f_i(A) = \sigma_i$.
We are interested to calculate the gradient of the function $f_i(A)$.

Hints:

1. Use the product rule of differentials to calculate dA, where A is considered as a function of U, S and V.
2. The entries of the diagonal of an anti-symmetric matrix are all zeros.
3. The Hadamard product of two matrices A, B of the same size, is denoted by $(A \odot B)_{ij} = A_{ij} \cdot B_{ij}$.
4. Use the cyclic property of the trace operator. That is:
$$Tr(ABC) = Tr(CAB) = Tr(BCA).$$
5. The trace of a scalar is a scalar. That is, given $a \in \mathbb{R}$:
$$Tr(a) = a$$