236330 - Introduction to Optimization: Homework #1
BONUS QUESTION

April 16, 2020

Amit Rotner
123456789
Or Steiner
123456789

Problem:

Let:

$$A = USV^T$$

Be singular-value decomposition of A.

We want to find the gradient of:

$$f_i(A) = \sigma_i$$

Solution:

Let $1 \le i \le k$.

The diagonal entries of S are the singular values of A, this implies that $S_{ii} = \sigma_i$. Hence:

$$f_i(A) = \sigma_i = e_i^T S e_i$$

Therefore:

$$df_i(A) = d\sigma_i = de_i^T S e_i = e_i^T dS e_i$$

Since U and V are orthogonal matrices, it follows:

$$U^T U = V^T V = I$$

Therefore:

$$dU^T U + U^T dU = 0$$

$$U^T dU = -dU^T U = -\left(U^T dU\right)^T$$

Thus, $U^T dU$ is skew-symmetric. And:

$$dV^T V + V^T dV = 0$$

$$dV^T V = -V^T dV = -\left(dV^T V\right)^T$$

Thus $dV^T V$ is skew-symmetric, as well.

Using the product rule of differentials on $A = USV^T$,we get:

$$dA = dUSV^T + UdSV^T + USdV^T$$

Left multiplying by U^T and right multiplying by V the above, yields:

$$U^T dAV = U^T dUS + dS + SdV^T V$$

Since $U^T dU$ and $dV^T V$ are skew-symmetric matrices, their diagonal is zero.

S is diagonal, hence dS is diagonal and therefore,

$$dS = I_k \odot \left[U^T dAV\right]$$

where $A \odot B$ is the Hadamard product of A and B.

Combining all together gives us:

$$\begin{aligned}
df_i(A) &= e_i^T dSe_i \\
&\underset{(1)}{=} Tr\left(e_i^T dSe_i\right) \\
&= Tr\left(e_i^T I^k \odot \left[U^T dAV\right] e_i\right) \\
&\underset{(2)}{=} Tr\left(e_i e_i^T I^k \odot \left[U^T dAV\right]\right) \\
&\underset{(3)}{=} Tr\left(I_{ii}^k \odot \left[U^T dAV\right]\right) \\
&= \left[U^T dAV\right]_{ii} \\
&\underset{(1)}{=} Tr\left(\left[U^T dAV\right]_{ii}\right) \\
&= Tr\left(I_{ii}^k U^T dAV\right) \\
&\underset{(2)}{=} Tr\left(V I_{ii}^k U^T dA\right)
\end{aligned}$$

(1) - $\forall a \in \mathbb{R}$: $Tr\left(a\right) = a$

(2) - Trace is invariant under cyclic permutation.

(3) - I_{ii}^k is a $k \times k$ matrix with all entries equal to 0, except for the ii element which is equal to 1.

Using the external definition of the gradient, we get:

$$\boxed{\nabla f_i\left(A\right) = U I_{ii}^k V^T}$$