Appendix

Anonymous Author(s)
Affiliation
Address
email

1 Proof of Theorem 1

Given $\boldsymbol{Z} = \mathcal{D}(\boldsymbol{R}) = \boldsymbol{U}\boldsymbol{\Sigma}_\lambda \boldsymbol{V}^T$, we have

$$(vec(\boldsymbol{Z}^T))^T = [\boldsymbol{u}_1^T \boldsymbol{\Sigma}_\lambda \boldsymbol{V}^T, \cdots, \boldsymbol{u}_i^T \boldsymbol{\Sigma}_\lambda \boldsymbol{V}^T, \cdots, \boldsymbol{u}_m^T \boldsymbol{\Sigma}_\lambda \boldsymbol{V}^T] = (vec(\boldsymbol{U}^T))^T (\boldsymbol{I} \otimes \boldsymbol{\Sigma}_\lambda \boldsymbol{V}^T), \quad (1)$$

and

$$vec(\boldsymbol{Z}) = [(\boldsymbol{U}\boldsymbol{\Sigma}_\lambda \boldsymbol{v}_1)^T, \cdots, (\boldsymbol{U}\boldsymbol{\Sigma}_\lambda \boldsymbol{v}_i)^T, \cdots, (\boldsymbol{U}\boldsymbol{\Sigma}_\lambda \boldsymbol{v}_n)^T]^T = (\boldsymbol{I} \otimes \boldsymbol{U}\boldsymbol{\Sigma}_\lambda) vec(\boldsymbol{V}^T), \quad (2)$$

where $\boldsymbol{u}_i^T$ and $\boldsymbol{v}_i^T$ are respectively the ith row of $\boldsymbol{U}$ and $\boldsymbol{V}$. Then, the partial derivatives in (i) are straightfoward. The partial derivative of $\boldsymbol{Z}$ with respect to the r-th diagonal element of the matrix $\boldsymbol{\Sigma}$ is

$$\frac{\partial \boldsymbol{Z}}{\partial \sigma_{r,r}} = vec(\boldsymbol{G}),$$

where

$$G_{ij} = U_{i,r} V_{j,r},$$

and $\sigma_{r,r}$ is the r-th diagonal element of the singular value matrix $\boldsymbol{\Sigma}$.

In order to calculate $\frac{\partial \boldsymbol{U}}{\partial \boldsymbol{R}}, \frac{\partial \boldsymbol{\Sigma}}{\partial \boldsymbol{R}}, \frac{\partial \boldsymbol{V}^T}{\partial \boldsymbol{R}}$, we do a perturbation analysis on the equation $\boldsymbol{R} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$. Let $\boldsymbol{\Delta_R}, \boldsymbol{\Delta_U}, \boldsymbol{\Delta_\Sigma}$ and $\boldsymbol{\Delta_V}$ be infinitesimally small perturbations to $\boldsymbol{R}, \boldsymbol{U}, \boldsymbol{\Sigma}$ and $\boldsymbol{V}$, respectively. Then

$$\begin{aligned} \boldsymbol{R} + \boldsymbol{\Delta_R} &= (\boldsymbol{U} + \boldsymbol{\Delta_U})(\boldsymbol{\Sigma} + \boldsymbol{\Delta_\Sigma})(\boldsymbol{V} + \boldsymbol{\Delta_V})^T \\ &= \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T + \boldsymbol{\Delta_U}\boldsymbol{\Sigma}\boldsymbol{V}^T + \boldsymbol{U}\boldsymbol{\Delta_\Sigma}\boldsymbol{V}^T + \boldsymbol{U}\boldsymbol{\Delta_\Sigma}\boldsymbol{\Delta_V}^T \\ &\quad + \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{\Delta_V}^T + \boldsymbol{\Delta_U}\boldsymbol{\Sigma}\boldsymbol{\Delta_V}^T + \boldsymbol{\Delta_U}\boldsymbol{\Delta_\Sigma}\boldsymbol{\Delta_V}^T + \boldsymbol{\Delta_U}\boldsymbol{\Delta_\Sigma}\boldsymbol{V}^T. \end{aligned} \quad (3)$$

Ignoring the higher-order infinitesimals, we obtain

$$\boldsymbol{\Delta_R} = \boldsymbol{\Delta_U}\boldsymbol{\Sigma}\boldsymbol{V}^T + \boldsymbol{U}\boldsymbol{\Delta_\Sigma}\boldsymbol{V}^T + \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{\Delta_V}^T \quad (4)$$

or equivalently,

$$\boldsymbol{U}^T\boldsymbol{\Delta_R}\boldsymbol{V} = \boldsymbol{U}^T\boldsymbol{\Delta_U}\boldsymbol{\Sigma} + \boldsymbol{\Delta_\Sigma} + \boldsymbol{\Sigma}\boldsymbol{\Delta_V}^T\boldsymbol{V}. \quad (5)$$

$\boldsymbol{U}$ and $\boldsymbol{V}$ are both orthogonal matrices, and so are their perturbed versions, i.e.

$$\begin{aligned} (\boldsymbol{U} + \boldsymbol{\Delta_U})^T(\boldsymbol{U} + \boldsymbol{\Delta_U}) &= \boldsymbol{I} \\ (\boldsymbol{V} + \boldsymbol{\Delta_V})^T(\boldsymbol{V} + \boldsymbol{\Delta_V}) &= \boldsymbol{I}. \end{aligned} \quad (6)$$

Then

$$\boldsymbol{U}^T\boldsymbol{\Delta_U} + \boldsymbol{\Delta_U}^T\boldsymbol{U} = \boldsymbol{0} \quad (7a)$$

$$\boldsymbol{V}^T\boldsymbol{\Delta_V} + \boldsymbol{\Delta_V}^T\boldsymbol{V} = \boldsymbol{0}. \quad (7b)$$

Together with (5), we obtain

$$\boldsymbol{U}^T\boldsymbol{\Delta_R}\boldsymbol{V} = -\boldsymbol{\Delta_U}^T\boldsymbol{U}\boldsymbol{\Sigma} + \boldsymbol{\Delta_\Sigma} - \boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{\Delta_V}. \quad (8)$$

Submitted to 33rd Conference on Neural Information Processing Systems (NeurIPS 2019). Do not distribute.

Let $\Delta_{R_{ij}}$ be the perturbation on the element R_{ij} of matirx $\boldsymbol{R}$. Let $\boldsymbol{u}_i^T$ be the i-th row of $\boldsymbol{U}$, and $\boldsymbol{v}_j^T$ be the j-th row of $\boldsymbol{V}$. Setting the elements of $\boldsymbol{\Delta_R}$ to zero except $\Delta_{R_{ij}}$, we obtain from (8) that

$$\Delta_{R_{ij}} \boldsymbol{u}_i {\boldsymbol{v}_j}^T = -\boldsymbol{\Delta}_{\boldsymbol{U}}^T \boldsymbol{U\Sigma} + \boldsymbol{\Delta_\Sigma} - \boldsymbol{\Sigma V}^T \boldsymbol{\Delta_V}, \tag{9}$$

or equivalently

$$\boldsymbol{u}_i {\boldsymbol{v}_j}^T = -\frac{\boldsymbol{\Delta}_{\boldsymbol{U}}^T}{\Delta_{R_{ij}}} \boldsymbol{U\Sigma} + \frac{\boldsymbol{\Delta_\Sigma}}{\Delta_{R_{ij}}} - \boldsymbol{\Sigma V}^T \frac{\boldsymbol{\Delta_V}}{\Delta_{R_{ij}}}. \tag{10}$$

From (7), we obtain

$$\boldsymbol{U}^T \frac{\boldsymbol{\Delta_U}}{\Delta_{R_{ij}}} + \frac{\boldsymbol{\Delta}_{\boldsymbol{U}}^T}{\Delta_{R_{ij}}} \boldsymbol{U} = \boldsymbol{0} \tag{11a}$$

$$\boldsymbol{V}^T \frac{\boldsymbol{\Delta_V}}{\Delta_{R_{ij}}} + \frac{\boldsymbol{\Delta}_{\boldsymbol{V}}^T}{\Delta_{R_{ij}}} \boldsymbol{V} = \boldsymbol{0}. \tag{11b}$$

Define $\boldsymbol{F} = \lim_{\Delta_{R_{ij}} \to 0} \boldsymbol{U}^T \frac{\boldsymbol{\Delta_U}}{\Delta_{R_{ij}}}$, $\boldsymbol{E} = \lim_{\Delta_{R_{ij}} \to 0} \boldsymbol{V}^T \frac{\boldsymbol{\Delta_V}}{\Delta_{R_{ij}}}$, and $\boldsymbol{C} = \lim_{\Delta_{R_{ij}} \to 0} (\boldsymbol{u}_i {\boldsymbol{v}_j}^T - \frac{\boldsymbol{\Delta_\Sigma}}{\Delta_{R_{ij}}})$. Then (9) is rewritten as

$$\boldsymbol{C} = -\boldsymbol{F}^T \boldsymbol{\Sigma} - \boldsymbol{\Sigma E}. \tag{12}$$

From (11), $\boldsymbol{F}$ and $\boldsymbol{E}$ are both antisymmetric matrices, i.e.

$$TriU(\boldsymbol{F}) = -TriL(\boldsymbol{F})^T \text{and } TriU(\boldsymbol{E}) = -TriL(\boldsymbol{E})^T, \tag{13}$$

where the diagonal elements of $\boldsymbol{F}$ and $\boldsymbol{E}$ are all zero. Then from (12), we obtain

$$\frac{\partial \boldsymbol{\Sigma}}{\partial R_{ij}} = vec\left(\lim_{\Delta_{R_{ij}} \to 0} \frac{\boldsymbol{\Delta_\Sigma}}{\Delta_{R_{ij}}} \right) = vec(diag(\boldsymbol{u}_i {\boldsymbol{v}_j}^T)), \tag{14}$$

$$\boldsymbol{C} = TriL(\boldsymbol{u}_i {\boldsymbol{v}_j}^T) + TriU(\boldsymbol{u}_i {\boldsymbol{v}_j}^T), \tag{15}$$

$$TriU(\boldsymbol{C})\boldsymbol{\Sigma}^{-1} = -TriU(\boldsymbol{F}^T) - \boldsymbol{\Sigma} TriU(\boldsymbol{E})\boldsymbol{\Sigma}^{-1}, \tag{16}$$

and

$$TriL(\boldsymbol{C})\boldsymbol{\Sigma}^{-1} = -TriL(\boldsymbol{F}^T) - \boldsymbol{\Sigma} TriL(\boldsymbol{E})\boldsymbol{\Sigma}^{-1}. \tag{17}$$

Taking tranpose on both sides of (17), together with $TriU(\boldsymbol{F}^T) = TriL(\boldsymbol{F})^T = -TriU(\boldsymbol{F}) = -TriL(\boldsymbol{F}^T)^T$, we obtain

$$\boldsymbol{\Sigma}^{-1}(TriL(\boldsymbol{C}))^T = TriU(\boldsymbol{F}^T) + \boldsymbol{\Sigma}^{-1} TriU(\boldsymbol{E})\boldsymbol{\Sigma}. \tag{18}$$

Adding (16) to (18), we obtain

$$TriU(\boldsymbol{C})\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}(TriL(\boldsymbol{C}))^T = \boldsymbol{\Sigma}^{-1} TriU(\boldsymbol{E})\boldsymbol{\Sigma} - \boldsymbol{\Sigma} TriU(\boldsymbol{E})\boldsymbol{\Sigma}^{-1}. \tag{19}$$

Let

$$\boldsymbol{P} = TriU(\boldsymbol{C})\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}(TriL(\boldsymbol{C}))^T. \tag{20}$$

Then, solving (19), we obtain the (i,j)th element of $\boldsymbol{E}$ as

$$E_{ij} = \begin{cases} \frac{P_{ij}}{({\sigma_{i,i}}^{-1}\sigma_{j,j} - \sigma_{i,i}\sigma_{j,j}^{-1})} & i < j \\ 0 & i = j \\ \frac{-P_{ji}}{({\sigma_{j,j}}^{-1}\sigma_{i,i} - \sigma_{j,j}\sigma_{i,i}^{-1})} & i > j \end{cases}. \tag{21}$$

From (12), we obtain

$$\boldsymbol{F} = -((\boldsymbol{C} + \boldsymbol{\Sigma E})\boldsymbol{\Sigma}^{-1})^T = -((TriL(\boldsymbol{u}_i {\boldsymbol{v}_j}^T) + TriU(\boldsymbol{u}_i {\boldsymbol{v}_j}^T) + \boldsymbol{\Sigma E})\boldsymbol{\Sigma}^{-1})^T. \tag{22}$$

Finally, from the definition of $\boldsymbol{F}$ and $\boldsymbol{E}$, we obtain

$$\frac{\partial \boldsymbol{U}}{\partial R_{ij}} = vec\left(\lim_{\Delta_{R_{ij}} \to 0} \frac{\boldsymbol{\Delta_U}}{\Delta_{R_{ij}}} \right) = vec\left(\boldsymbol{UF}\right), \tag{23}$$

$$\frac{\partial \boldsymbol{V}}{\partial R_{ij}} = vec\left(\lim_{\Delta_{R_{ij}} \to 0} \frac{\boldsymbol{\Delta_V}}{\Delta_{R_{ij}}} \right) = vec\left(\boldsymbol{VE}\right). \tag{24}$$