Supplementary Material for the Paper: An Interface between Grassmann manifolds and vector spaces

Lincon S. Souza[1] Naoya Sogi[1] Bernardo B. Gatto[2]
Takumi Kobayashi[3] Kazuhiro Fukui[1,2]
[1]Graduate School of Science and Technology, University of Tsukuba
[2]Center for Artificial Intelligence Research (C-AIR), University of Tsukuba
[3] National Institute of Advanced Industrial Science and Technology (AIST)
{lincons, sogi, bernardo}@cvlab.cs.tsukuba.ac.jp takumi.kobayashi@aist.go.jp
kfukui@cs.tsukuba.ac.jp

1. Gradient computation of the Grassmann Log Map

In this document, we present a derivation of the gradient of the log map for the Grassmann manifold $\mathcal{G}(d, m)$. For that, we utilise various conventional techniques to operate differential forms [1].

Given the data and the loss gradient $\nabla_{\boldsymbol{H}} L = \dot{\boldsymbol{H}}$, we compute the gradients with respect to the tangency point and data.

The Grassmann log consists of the three equations below:

$$\boldsymbol{B} = (\boldsymbol{K}^\top \boldsymbol{X})^{-1}(\boldsymbol{K}^\top - \boldsymbol{K}^\top \boldsymbol{X}\boldsymbol{X}^\top), \quad (1)$$

$$\boldsymbol{W}\boldsymbol{\Theta}\boldsymbol{Z}^\top = \boldsymbol{B}^\top, \quad (2)$$

$$\mathrm{Log}_{\boldsymbol{K}} \boldsymbol{X} = \boldsymbol{H} = \boldsymbol{W}^* \arctan(\boldsymbol{\Theta}^*) \boldsymbol{Z}^{*\top}, \quad (3)$$

where $\boldsymbol{W}^*, \boldsymbol{\Theta}^*, \boldsymbol{Z}^*$ represent the matrices with the first m columns of $\boldsymbol{W}, \boldsymbol{\Theta}$ and $\boldsymbol{Z}^*$ respectively. It should be noted that equation 2 is the transposed SVD of $\boldsymbol{B}$. The differential of the expression 3 may be written as:

$$\boldsymbol{H} = \boldsymbol{W}^* \arctan(\boldsymbol{\Theta}^*) \boldsymbol{Z}^{*\top} = \boldsymbol{W}^* \boldsymbol{S} \boldsymbol{Z}^{*\top}, \quad (4)$$

$$d\boldsymbol{H} = d\boldsymbol{W}^* \boldsymbol{S} \boldsymbol{Z}^{*\top} + \boldsymbol{W}^* d\boldsymbol{S} \boldsymbol{Z}^{*\top} + \boldsymbol{W}^* \boldsymbol{S} d\boldsymbol{Z}^{*\top}. \quad (5)$$

Above, we performed the change of variables $\boldsymbol{S} = \arctan(\boldsymbol{\Theta}^*)$. This contains an element-wise differential, simple to compute:

$$d\boldsymbol{S} = \boldsymbol{\Omega} d\boldsymbol{\Theta}, \quad (6)$$

where $\boldsymbol{\Omega}_i = 1/(1 + \boldsymbol{\Theta}_i^2)$ is a diagonal matrix and $i = 1, \ldots, d$ iterates over the diagonal. Since $\boldsymbol{W}^*$ is an orthogonal matrix, $\boldsymbol{W}^{*\top} d\boldsymbol{W}^*$ is skew-symmetric. This constraint leads to Townsend's solution [2] for equation 5. We use this result and reverse the change of variables to obtain the update equations for each variable:

$$\begin{aligned}\dot{\boldsymbol{W}}^* = \boldsymbol{W}^*(\boldsymbol{F} \circ [&\boldsymbol{W}^{*\top} \dot{\boldsymbol{H}} \boldsymbol{Z}^* \arctan(\boldsymbol{\Theta}^*) \\ &+ \arctan(\boldsymbol{\Theta}^*) \boldsymbol{Z}^{*\top} \dot{\boldsymbol{H}}^\top \boldsymbol{W}^*] \\ &+ (\boldsymbol{I} - \boldsymbol{W}^* \boldsymbol{W}^{*\top}) \dot{\boldsymbol{H}} \boldsymbol{Z}^* \arctan(\boldsymbol{\Theta}^*)^{-1}).\end{aligned} \quad (7)$$

$$\dot{\boldsymbol{\Theta}}^* = \boldsymbol{I} \circ [\boldsymbol{W}^{*\top} \dot{\boldsymbol{H}} \boldsymbol{Z}^*] \boldsymbol{\Omega}^{-1} \quad (8)$$

$$\begin{aligned}\dot{\boldsymbol{Z}}^* = \boldsymbol{Z}^*(\boldsymbol{F} \circ [&\arctan(\boldsymbol{\Theta}^*) \boldsymbol{W}^{*\top} \dot{\boldsymbol{H}} \boldsymbol{Z}^* \\ &+ \boldsymbol{Z}^{*\top} \dot{\boldsymbol{H}}^\top \boldsymbol{W}^* \arctan(\boldsymbol{\Theta}^*)]) \\ &+ (\boldsymbol{I} - \boldsymbol{Z}^* \boldsymbol{Z}^{*\top}) \dot{\boldsymbol{H}}^\top \boldsymbol{W}^* \arctan(\boldsymbol{\Theta}^*)^{-1}.\end{aligned} \quad (9)$$

$\circ$ represents the Hadamard product, and $\boldsymbol{I}$ represents the identity matrix. $\boldsymbol{F}$ is a matrix of the form:

$$\boldsymbol{F}_{ij} = \begin{cases} 1/(\arctan^2(\boldsymbol{\Theta}_j) - \arctan^2(\boldsymbol{\Theta}_i)), & i \neq j \\ 0, & i = j. \end{cases} \quad (10)$$

The results $\dot{\boldsymbol{W}}^*$, $\dot{\boldsymbol{\Theta}}^*$ and $\dot{\boldsymbol{Z}}^*$ are m-leftmost matrices, so to continue back to the full matrix gradients we can fill in columns of zeros until the matrices become square, where then we write then as $\dot{\boldsymbol{W}}$, $\dot{\boldsymbol{\Theta}}$ and $\dot{\boldsymbol{Z}}$. Then, the next step is to consider the equation $\boldsymbol{W}\boldsymbol{\Theta}\boldsymbol{Z}^\top = \boldsymbol{B}^\top$. Since it is a reconstruction rather than a decomposition, its update equation can be obtained as:

$$\begin{aligned}\dot{\boldsymbol{B}}^\top = [&\boldsymbol{W}(\boldsymbol{F} \circ [\boldsymbol{W}^\top \dot{\boldsymbol{W}} - \dot{\boldsymbol{W}}^\top \boldsymbol{W}]) \boldsymbol{\Theta} \\ &+ (\boldsymbol{I} - \boldsymbol{W}\boldsymbol{W}^\top) \dot{\boldsymbol{W}} \boldsymbol{\Theta}^{-1}] \boldsymbol{Z}^\top \\ &+ \boldsymbol{W}(\boldsymbol{I} \circ \dot{\boldsymbol{\Theta}}) \boldsymbol{Z}^\top + \boldsymbol{W}[\boldsymbol{\Theta}(\boldsymbol{F} \circ [\boldsymbol{Z}^\top \dot{\boldsymbol{Z}} - \dot{\boldsymbol{Z}}^\top \boldsymbol{Z}]) \boldsymbol{Z}^\top \\ &+ \boldsymbol{\Theta}^{-1} \dot{\boldsymbol{Z}}^\top (\boldsymbol{I} - \boldsymbol{Z}\boldsymbol{Z}^\top)],\end{aligned} \quad (11)$$

where $\boldsymbol{F}$ follows equation 10, but the non-diagonals are defined as $1/(\boldsymbol{\Theta}_j^2 - \boldsymbol{\Theta}_i^2)$.

Finally we consider the equation $\boldsymbol{B}^\top = (\boldsymbol{X}^\top\boldsymbol{K})^{-1}(\boldsymbol{X}^\top - \boldsymbol{X}^\top\boldsymbol{K}\boldsymbol{K}^\top)$. We transpose it and call $\boldsymbol{A} = (\boldsymbol{X}^\top - \boldsymbol{X}^\top\boldsymbol{K}\boldsymbol{K}^\top)$ and $\boldsymbol{C} = (\boldsymbol{X}^\top\boldsymbol{K})^{-1}$, then derivate it by the product rule, massaging the equation to obtain a general form:

$$d\boldsymbol{B} = d\boldsymbol{A}\boldsymbol{C} + \boldsymbol{A}d\boldsymbol{C} \tag{12}$$

$$d\boldsymbol{A} = d\boldsymbol{X} - (d\boldsymbol{K}\boldsymbol{K}^\top\boldsymbol{X} + \boldsymbol{K}d\boldsymbol{K}^\top\boldsymbol{X} + \boldsymbol{K}\boldsymbol{K}^\top d\boldsymbol{X}) \tag{13}$$

$$d\boldsymbol{C} = -(\boldsymbol{K}^\top\boldsymbol{X})^{-1}(d\boldsymbol{K}^\top\boldsymbol{X} + \boldsymbol{K}^\top d\boldsymbol{X})(\boldsymbol{K}^\top\boldsymbol{X})^{-1} \tag{14}$$

We obtain two update rules, one to from $d\boldsymbol{B}$ with respect to $\boldsymbol{X}$ in case a gradient-based pre-processing needs it, and one with respect to $\boldsymbol{K}$ to update it as a parameter. The derivative with respect to $\boldsymbol{X}$ is calculated as follows. First we consider $d\boldsymbol{K} = 0$:

$$\begin{aligned} d\boldsymbol{B} = (\boldsymbol{X} - (d\boldsymbol{K}\boldsymbol{K}^\top\boldsymbol{X} + \boldsymbol{K}d\boldsymbol{K}^\top\boldsymbol{X} \\ + \boldsymbol{K}\boldsymbol{K}^\top d\boldsymbol{X}))\boldsymbol{C} + \boldsymbol{A}(-(\boldsymbol{K}^\top\boldsymbol{X})^{-1}(d\boldsymbol{K}^\top\boldsymbol{X} \\ + \boldsymbol{K}^\top d\boldsymbol{X} + \boldsymbol{K}^\top d\boldsymbol{X})(\boldsymbol{K}^\top\boldsymbol{X})^{-1}) \end{aligned} \tag{15}$$

$$\begin{aligned} d\boldsymbol{B} = (\boldsymbol{I} - \boldsymbol{K}\boldsymbol{K}^\top)d\boldsymbol{X}(\boldsymbol{K}^\top\boldsymbol{X})^{-1} \\ - \boldsymbol{A}(\boldsymbol{K}^\top\boldsymbol{X})^{-1}\boldsymbol{K}^\top d\boldsymbol{X}(\boldsymbol{K}^\top\boldsymbol{X})^{-1} \end{aligned} \tag{16}$$

$$d\boldsymbol{B} = (\boldsymbol{I} - \boldsymbol{K}\boldsymbol{K}^\top - \boldsymbol{A}(\boldsymbol{K}^\top\boldsymbol{X})^{-1}\boldsymbol{K}^\top)d\boldsymbol{X}(\boldsymbol{K}^\top\boldsymbol{X})^{-1} \tag{17}$$

Then, since our loss function outputs a single real value, we can massage the equations to a canonical form $dL = \operatorname{tr}\dot{\boldsymbol{B}}^\top d\boldsymbol{B}$:

$$\begin{aligned} dL = \operatorname{tr}\dot{\boldsymbol{B}}^\top((\boldsymbol{I} - \boldsymbol{K}\boldsymbol{K}^\top \\ - \boldsymbol{A}(\boldsymbol{K}^\top\boldsymbol{X})^{-1}\boldsymbol{K}^\top)d\boldsymbol{X}(\boldsymbol{K}^\top\boldsymbol{X})^{-1}), \end{aligned} \tag{18}$$

which leads to the final update equation for the gradient of $\boldsymbol{X}$:

$$\dot{\boldsymbol{X}} = (\boldsymbol{K}^\top\boldsymbol{X})^{-1}\dot{\boldsymbol{B}}^\top[(\boldsymbol{I} - \boldsymbol{K}\boldsymbol{K}^\top)(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{K}^\top\boldsymbol{X})^{-1}\boldsymbol{K}^\top]. \tag{19}$$

Repeating the same technique, the update equation for $\boldsymbol{K}$ can be derived as:

$$\begin{aligned} \dot{\boldsymbol{K}} = -(\boldsymbol{K}^\top\boldsymbol{X})^{-1}\dot{\boldsymbol{B}}^\top[\boldsymbol{K}^\top\boldsymbol{X} \\ + \boldsymbol{K}^\top\boldsymbol{X}^\top + (\boldsymbol{X}^\top\boldsymbol{K})^{-1}\boldsymbol{X}^\top(\boldsymbol{I} - \boldsymbol{K}\boldsymbol{K}^\top)]. \end{aligned} \tag{20}$$

References

[1] Thomas P Minka. Old and new matrix algebra useful for statistics. *See www. stat. cmu. edu/minka/papers/matrix. html*, 2000. 1

[2] James Townsend. Differentiating the singular value decomposition, 2016. 1