Disturbance Grassmann Kernels for Subspace-Based Learning

Junyuan Hong, Huanhuan Chen and Feng Lin
University of Science and Technology of China
Hefei, Anhui
Email: jyhong@mail.ustc.edu.cn, hchen@ustc.edu.cn and lf1995@mail.ustc.edu.cn

I. Proof of Lemma 5.1-5.2

The equivalence relation between two Grassmann elements, $\mathbf{U}$ and $\mathbf{U}'$, are defined such that there exists a $m \times m$ orthogonal matrix $\mathbf{Q}_m$ such that

$$\mathbf{U}' = \mathbf{U}\mathbf{Q}_m.$$

As a result, the basis matrix is rotation invariant to any $\mathbf{Q}_m$. Hence, let $\tilde{\mathbf{Q}}_m$ and $\mathbf{Q}_m$ be two orthogonal matrixes. Then $\mathbf{U}\mathbf{Q}_m$ and $\tilde{\mathbf{U}}\tilde{\mathbf{Q}}$ are the equivalent representations of $\mathbf{U}$ and $\tilde{\mathbf{U}}$, respectively. Due to the exponential mapping,

$$\tilde{\mathbf{U}} = \pi(\mathbf{H}) = \begin{pmatrix} \mathbf{U}\mathbf{V}_H & \mathbf{U}_H \end{pmatrix} \begin{pmatrix} \cos \mathbf{\Sigma}_H \\ \sin \mathbf{\Sigma}_H \end{pmatrix} \mathbf{V}_H^T, \quad (1)$$

$$\text{s.t. } \mathbf{H} = \mathbf{U}_H \mathbf{\Sigma}_H \mathbf{V}_H^T, \quad (2)$$

we have the relation:

$$\tilde{\mathbf{U}}\tilde{\mathbf{Q}}_m = \begin{pmatrix} \mathbf{U}\mathbf{Q}_m\mathbf{V}_H & \mathbf{U}_H \end{pmatrix} \begin{pmatrix} \cos \mathbf{\Sigma}_H \\ \sin \mathbf{\Sigma}_H \end{pmatrix} \mathbf{V}_H^T \tilde{\mathbf{Q}}_m. \quad (3)$$

Because $\mathbf{V}_H$ is a unitary matrix, let $\mathbf{Q}_m = \mathbf{V}_H^T$ and $\tilde{\mathbf{Q}}_m = \mathbf{V}_H$. Substituting them into Eq. (3), we obtain

$$\tilde{\mathbf{U}}\tilde{\mathbf{Q}}_m = \mathbf{U}\cos \mathbf{\Sigma}_H + \mathbf{U}_H \sin \mathbf{\Sigma}_H \quad (4)$$

for which we can construct a new tangent vector as

$$\mathbf{H}^* = \mathbf{U}_H \mathbf{\Sigma}_H. \quad (5)$$

And because

$$\mathbf{H} = \mathbf{U}_\perp \mathbf{Z} \quad (6)$$

we can rewrite the tangent vector as

$$\mathbf{H}^* = \mathbf{U}_\perp \mathbf{Z} = \hat{\mathbf{H}}\mathbf{\Theta} \quad (7)$$

where $\mathbf{\Theta}$ is a diagonal matrix with entries as vector length $\theta_i = \|\mathbf{U}_\perp \mathbf{z}_i\| = \|\mathbf{z}_i\|$, and $\hat{\mathbf{H}} = \mathbf{\Theta}^{-1}\mathbf{U}_\perp\mathbf{Z}$. The Eq. (5) together with Eq. (7) imply the equivalence, $\mathbf{U}_H \equiv \hat{\mathbf{H}}$ and $\mathbf{\Sigma}_H \equiv \mathbf{\Theta}$, which manifests the columns of $\hat{\mathbf{H}}$ are orthonormal. Substitution to Eq. (4) gives the result in Lemma 5.1.

We can further deduce that the $\mathbf{H}^* = \mathbf{H}\mathbf{V}$, thus any non-orthonormal $\mathbf{H}$ is mapped to the equivalent subspace. Therefore, the second lemma is proved.

II. Proof of Lemma 6.1

To reveal the form of approximated Gaussian noise, we start from the Gaussian perturbation in data. Suppose $\mathbf{X}$ is a $D \times N$ data matrix, and $\tilde{\mathbf{X}} = \mathbf{X} + \epsilon_X \mathbf{W}$ is its noisy version controlled by a small constant, where the entries of $\mathbf{W}$ are independently and identically distributed. For a given $m < D$, $\mathbf{X}$ and $\tilde{\mathbf{X}}$ have the conformally partitions SVDs

$$\mathbf{X} = \begin{pmatrix} \mathbf{U} & \mathbf{U}_L \end{pmatrix} \begin{pmatrix} \mathbf{\Sigma} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{V} & \mathbf{V}_L \end{pmatrix}^T \quad (8)$$

$$\tilde{\mathbf{X}} = \begin{pmatrix} \tilde{\mathbf{U}} & \tilde{\mathbf{U}}_L \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{\Sigma}} & 0 \\ 0 & \tilde{\mathbf{\Sigma}}_L \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{V}} & \tilde{\mathbf{V}}_L \end{pmatrix}^T \quad (9)$$

where $\mathbf{\Sigma}$ and $\tilde{\mathbf{\Sigma}}$ are $m \times m$ matrices, and all singular values denoted by λ_i, $i = 1, \ldots, m$ are in descending order.

According to the theorem of singular-vector perturbation [1], with an exponential probability, there exists a unitary $m \times m$ matrix $\mathbf{Q}$ such that

$$\left\| \tilde{\mathbf{U}}_H \mathbf{M} - (\mathbf{U}_H + \epsilon_X \mathbf{W}_U) \right\| \le \delta$$

where δ is a small value governed by noise level, singular vectors and singular values, and the Gaussian noise matrix $\mathbf{W}_U$ is given by $\mathbf{W}_U = \mathbf{U}_L \mathbf{U}_L^T \mathbf{W}\mathbf{V}\mathbf{\Sigma}^{-1}$. When $\mathbf{X}$ does not have two non-zero singular values that are very close to each other, the rotation by $\mathbf{Q}$ can be omitted.

It is obvious that the matrix $\mathbf{U}_L^T\mathbf{W}\mathbf{V}$ has a distribution identical to $\mathbf{W}$ since $\mathbf{W}$ is spherically Gaussian distributed and coordination transforming by $\mathbf{U}_L$ and $\mathbf{V}_H$ has no effect on such a spherical distribution. Let $\mathbf{W}_0 = \mathbf{U}_L^T\mathbf{W}\mathbf{V}$ which is an $m \times m$ matrix, then the entries of $\mathbf{W}_0$ are subject to $\mathcal{N}(0, 1/D)$. Thus we reach the conclusion for that the $\mathbf{U}_L$ presents a null space of $\mathbf{U}$.

References

[1] R. Wang, "Singular vector perturbation under gaussian noise," *SIAM Journal on Matrix Analysis and Applications*, vol. 36, no. 1, pp. 158–177, 2015.