More About VLAD: A Leap from Euclidean to Riemannian Manifolds Supplementary Material

Masoud Faraki Mehrtash T. Harandi Fatih Porikli
College of Engineering and Computer Science, Australian National University, Australia
NICTA, Canberra Research Laboratory, Australia
{masoud.faraki,mehrtash.harandi,fatih.porikli}@nicta.com.au

In this supplementary material, we provide the proofs of Theorem 3 and Theorem 4 in the paper. The Theorem 3 states that the Fréchet mean of a set of SPD matrices based on the Jeffrey divergence, δ_J, admits a closed form solution.

Theorem 3. *The Fréchet mean of a set of SPD matrices* $\{\boldsymbol{X}_i\}_{i=1}^m \in \mathcal{S}_{++}^d$ *with* δ_J *is*

$$\boldsymbol{\mu} = \boldsymbol{P}^{-1/2}(\boldsymbol{P}^{1/2}\boldsymbol{Q}\boldsymbol{P}^{1/2})^{1/2}\boldsymbol{P}^{-1/2}\,, \tag{1}$$

where $\boldsymbol{P} = \sum_i \boldsymbol{X}_i^{-1}$ *and* $\boldsymbol{Q} = \sum_i \boldsymbol{X}_i$.

Proof. The solution is obtained by zeroing out the derivative of $\sum_i^m \delta_J^2(\boldsymbol{X}_i, \boldsymbol{\mu})$ with respect to $\boldsymbol{\mu}$. At $\boldsymbol{\mu}$, $\frac{\partial \delta_J^2(\boldsymbol{X}_i,\boldsymbol{\mu})}{\partial \boldsymbol{\mu}} = \frac{1}{2}(\boldsymbol{X}_i^{-1} - \boldsymbol{\mu}^{-1}\boldsymbol{X}_i\boldsymbol{\mu}^{-1})$, we get

$$\begin{aligned}\frac{\partial \sum_i^m \delta_J^2(\boldsymbol{X}_i, \boldsymbol{\mu})}{\partial \boldsymbol{\mu}} &= \sum_{i=1}^m \boldsymbol{X}_i^{-1} - \sum_{i=1}^m \boldsymbol{\mu}^{-1}\boldsymbol{X}_i\boldsymbol{\mu}^{-1} = 0 \\ &\Rightarrow \boldsymbol{\mu}\sum_{i=1}^m \boldsymbol{X}_i^{-1}\boldsymbol{\mu} = \sum_{i=1}^m \boldsymbol{X}_i. \end{aligned} \tag{2}$$

The quadratic equation $\boldsymbol{ABA} = \boldsymbol{C}$ is called a *Riccati* equation [1] and has the following unique and closed form solution for $\boldsymbol{B} \succ 0$ and $\boldsymbol{C} \succeq 0$

$$\boldsymbol{A} = \boldsymbol{B}^{-1/2}\big(\boldsymbol{B}^{1/2}\boldsymbol{C}\boldsymbol{B}^{1/2}\big)^{1/2}\boldsymbol{B}^{-1/2}$$

Comparing the form of (2) with the Riccati equation concludes the proof. We note that a different proof is also provided in [3]. □

Similarly the Theorem 4 states that, for a set of linear subspaces under the projection metric, δ_P, we have the luxury of obtaining the Fréchet mean analytically.

Theorem 4. *The Fréchet mean for a set of points* $\{\boldsymbol{X}_i\}_{i=1}^m$, $\boldsymbol{X}_i \in \mathcal{G}_d^p$ *based on* δ_P *admits a closed-form solution.*

Proof. We need to solve

$$\begin{aligned}\boldsymbol{\mu}^* &\triangleq \arg\min_{\boldsymbol{\mu}} \sum_{i=1}^m \left\| \boldsymbol{\mu}\boldsymbol{\mu}^T - \boldsymbol{X}_i\boldsymbol{X}_i^T \right\|_F^2, \\ &\text{s.t. } \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu} = \mathbf{I}_{\mathrm{p}}.\end{aligned} \tag{3}$$

We note that with the orthogonality constraint on points, *i.e.*, $\boldsymbol{\mu}^T\boldsymbol{\mu} = \boldsymbol{X}_i^T\boldsymbol{X}_i = \mathbf{I}_p$

$$\begin{aligned}\sum_{i=1}^m \left\| \boldsymbol{\mu}\boldsymbol{\mu}^T - \boldsymbol{X}_i\boldsymbol{X}_i^T \right\|_F^2 &= 2mp - 2\sum_{i=1}^m \mathrm{Tr}\{\boldsymbol{\mu}^T\boldsymbol{X}_i\boldsymbol{X}_i^T\boldsymbol{\mu}\} \\ &= 2mp - 2\,\mathrm{Tr}\{\boldsymbol{\mu}^T\Big(\sum_{i=1}^m \boldsymbol{X}_i\boldsymbol{X}_i^T\Big)\boldsymbol{\mu}\}.\end{aligned}$$

Therefore to minimize (3), one should maximize $\mathrm{Tr}\{\boldsymbol{\mu}^T\Big(\sum_{i=1}^m \boldsymbol{X}_i\boldsymbol{X}_i^T\Big)\boldsymbol{\mu}\}$ by taking into account the constraint $\boldsymbol{\mu}^T\boldsymbol{\mu} = \mathbf{I}_p$, *i.e.*,

$$\begin{aligned}\boldsymbol{\mu}^* &\triangleq \arg\max_{\boldsymbol{\mu}}\ \mathrm{Tr}\{\boldsymbol{\mu}^T\Big(\sum_{i=1}^m \boldsymbol{X}_i\boldsymbol{X}_i^T\Big)\boldsymbol{\mu}\}, \\ &\text{s.t. } \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu} = \mathbf{I}_{\mathrm{p}}.\end{aligned} \tag{4}$$

The solution of (4) is obtained by computing the p largest eigenvectors of $\sum_{i=1}^m \boldsymbol{X}_i\boldsymbol{X}_i^T$ according to the Rayleigh-Ritz theorem [2], which concludes the proof. □

References

[1] R. Bhatia. *Positive Definite Matrices*. Princeton University Press, 2007. 1

[2] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge University Press, 2012. 1

[3] Z. Wang and B. C. Vemuri. An affine invariant tensor dissimilarity measure and its applications to tensor-valued image segmentation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 223–228. IEEE, 2004. 1