Joint Dimensionality Reduction and Metric Learning: A Geometric Take Supplementary Material

Mehrtash Harandi [1 2] **Mathieu Salzmann** [3] **Richard Hartley** [2 1]

1. Quotient Manifolds: Theory and Derivations

In the following, we first provide some general theory about quotient manifolds, and then derive the quantities that are of interest for our particular purpose.

1.1. Theory of Quotient Manifolds

Let $\mathcal{M}_p$ be a Riemannian manifold and G be a group acting on $\mathcal{M}_p$, that is, G defines a mapping $f : G \times \mathcal{M}_p \to \mathcal{M}_p$. The orbit of a point $\boldsymbol{x} \in \mathcal{M}_p$ is the set of images of $\boldsymbol{x}$ under all elements of the group, *i.e.*,

$$\mathrm{Orb}(\boldsymbol{x}) = \{f(g, \boldsymbol{x}) : \forall g \in G\} \ .$$

The action of group G defines a relation denoted by $\sim$ in the form $\boldsymbol{x} \sim \boldsymbol{q} : \exists g \in G \ \text{s.t.} \ \boldsymbol{q} = f(g, \boldsymbol{x})$. Recall that a relation $\sim$ on a set $\mathcal{X}$ is said to be an equivalence relation iff $\forall a, b, c \in \mathcal{X}$, $a \sim a$ (reflexive), $a \sim b \Rightarrow b \sim a$ (symmetric) and if $a \sim b$ and $b \sim c$ then $a \sim c$(transitive). If the action of group G defines an equivalence relation, then under the conditions of Theorem 1 below, the set of orbits denoted by $\mathcal{M}_p \backslash G$ forms a smooth manifold.

Theorem 1 (Quotient Manifold Theorem)**.** *Suppose that a Lie group G acts* **smoothly**, **freely** *and* **properly** *on a smooth manifold $\mathcal{M}_p$. Then $\mathcal{M}_p \backslash G$ is a topological manifold of dimension $\dim(\mathcal{M}_p) - \dim(G)$, and has a unique smooth structure.*

Proof. See Theorem 9.16 in (Lee, 2003). □

Intuitively, a quotient manifold $\mathcal{M} \triangleq \mathcal{M}_p \backslash G$ is formed by gluing points of $\mathcal{M}_p$ together using the equivalence relation $\sim$. Assuming that $\mathcal{M}$ is a smooth manifold, it is possible to make use of the structure of $\mathcal{M}_p$ to define the Riemannian geometry of $\mathcal{M}$, as discussed below. To this end, let us first define the tangent space of a quotient manifold in general, and then focus on the specific quotient manifold of interest in this work.

1.1.1. Tangent spaces.

The equivalence relation $\sim$ splits the tangent space of $\mathcal{M}_t$ at Ω into two complementary parts, namely the horizontal space $\mathcal{H}_\Omega \mathcal{M}_p$ and the vertical space $\mathcal{V}_\Omega \mathcal{M}_p$ with respect to the metric $g_p(\cdot, \cdot)$ on the original manifold $\mathcal{M}_p$. That is, $T_\Omega \mathcal{M}_p = \mathcal{H}_\Omega \mathcal{M}_p \oplus \mathcal{V}_\Omega \mathcal{M}_p$, with

$$\mathcal{H}_\Omega \mathcal{M}_p = \{\boldsymbol{h}_\Omega \in T_\Omega \mathcal{M}_p : g_p(\boldsymbol{h}_\Omega, \boldsymbol{v}_\Omega) = 0, \ \forall \boldsymbol{v}_\Omega \in \mathcal{V}_\Omega \mathcal{M}_p\} \ . \tag{1}$$

The tangent space of the quotient manifold is identified with the horizontal space of the original manifold, *i.e.*, $T_{[\Omega]} \mathcal{M} \triangleq \mathcal{H}_\Omega \mathcal{M}_p$. Once the horizontal space $\mathcal{H}_\Omega \mathcal{M}_p$ is identified, a tangent vector on $\xi_{[\Omega]} \in T_{[\Omega]} \mathcal{M}$ can be uniquely represented

[1]Data61, CSIRO, Canberra, Australia [2]Australian National University, Canberra, Australia [3]CVLab, EPFL, Switzerland. Correspondence to: Mehrtash Harandi <mehrtash.harandi@anu.edu.au>.

Proceedings of the 34^{th} International Conference on Machine Learning, Sydney, Australia, PMLR 70, 2017.

by a horizontal tangent vector $\xi_{\Omega}^{\uparrow} \in \mathcal{H}_{\Omega}\mathcal{M}_p$. For two tangent vectors $\xi_{[\Omega]}, \vartheta_{[\Omega]} \in T_{[\Omega]}\mathcal{M}$, the metric in the original space induces a Riemannian metric on the quotient manifold given by

$$g(\xi_{[\Omega]}, \vartheta_{[\Omega]}) \triangleq g_p(\xi_{\Omega}^{\uparrow}, \vartheta_{\Omega}^{\uparrow}) \; . \tag{2}$$

1.2. Derivations for our Specific Quotient Manifold

For the product manifold $\mathcal{M}_p \triangleq \mathrm{St}(p,n) \times \mathcal{S}_{++}^{p}$, the tangent space is given by

$$T_{(\boldsymbol{W},\boldsymbol{M})} = T_{\boldsymbol{W}}\mathrm{St}(p,n) \times T_{\boldsymbol{M}}\mathcal{S}_{++}^{p} \; . \tag{3}$$

This lets us define the Riemannian metric of the product manifold as

$$g_{(\boldsymbol{W},\boldsymbol{M})}((\xi_{\boldsymbol{W}}, \xi_{\boldsymbol{M}}), (\varsigma_{\boldsymbol{W}}, \varsigma_{\boldsymbol{M}})) = 2\,\mathrm{Tr}(\xi_{\boldsymbol{W}}^T \varsigma_{\boldsymbol{W}}) + \mathrm{Tr}(\boldsymbol{M}^{-1}\xi_{\boldsymbol{M}}\boldsymbol{M}^{-1}\varsigma_{\boldsymbol{M}}) \; . \tag{4}$$

The Riemannian gradient of such a space directly follows from the geometry of the Stiefel and SPD manifolds. That is, for a smooth function $f_p(\boldsymbol{W}, \boldsymbol{M}) : \mathcal{M}_p \to \mathbb{R}$, we have

$$\mathrm{grad}\; f_p = \Big(\nabla_{\boldsymbol{W}}(f_p) - \boldsymbol{W}\mathrm{sym}\Big(\boldsymbol{W}^T\nabla_{\boldsymbol{W}}(f_p)\Big), \boldsymbol{M}\mathrm{sym}\Big(\nabla_{\boldsymbol{M}}(f_p)\Big)\boldsymbol{M}\Big). \tag{5}$$

Here $\nabla_{\boldsymbol{W}}(f)$ and $\nabla_{\boldsymbol{M}}(f)$ denote the $n \times p$ and $p \times p$ Euclidean gradients, *i.e.*, the matrix of partial derivatives such that

$$\nabla_{\boldsymbol{W}}(f_p)_{i,j} = \frac{\partial f_p(\boldsymbol{W}, \boldsymbol{M})}{\partial \boldsymbol{W}_{i,j}} \;, \qquad \nabla_{\boldsymbol{M}}(f_p)_{i,j} = \frac{\partial f_p(\boldsymbol{W}, \boldsymbol{M})}{\partial \boldsymbol{M}_{i,j}}. \tag{6}$$

Theorem 2. *The set* $\mathcal{M} \triangleq \big(\mathrm{St}(p,n) \times \mathcal{S}_{++}^{p}\big)\backslash\mathcal{O}(p)$ *with the equivalence relation* $[\big(\boldsymbol{W}, \boldsymbol{M}\big)] \sim \Big\{\big(\boldsymbol{W}\boldsymbol{R}, \boldsymbol{R}^T\boldsymbol{M}\boldsymbol{R}\big);\, \forall \boldsymbol{R} \in \mathcal{O}(p)\Big\}$. *and Riemannian metric* $g_{(\boldsymbol{W},\boldsymbol{M})}((\xi_{\boldsymbol{W}}, \xi_{\boldsymbol{M}}), (\varsigma_{\boldsymbol{W}}, \varsigma_{\boldsymbol{M}})) = 2\,\mathrm{Tr}(\xi_{\boldsymbol{W}}^T\varsigma_{\boldsymbol{W}}) + \mathrm{Tr}(\boldsymbol{M}^{-1}\xi_{\boldsymbol{M}}\boldsymbol{M}^{-1}\varsigma_{\boldsymbol{M}})$ *forms a Riemannian quotient manifold.*

Proof. $\mathcal{M}$ is a smooth manifold as a direct result of Theorem 9.16 in (Lee, 2003). In particular, the mapping $\pi : \big(\boldsymbol{W}, \boldsymbol{M}\big) \to \big(\boldsymbol{W}\boldsymbol{R}, \boldsymbol{R}^T\boldsymbol{M}\boldsymbol{R}\big)$ for $\boldsymbol{R} \in \mathcal{O}(p)$ is smooth, free, as $\big(\boldsymbol{W}, \boldsymbol{M}\big) = \big(\boldsymbol{W}\boldsymbol{R}, \boldsymbol{R}^T\boldsymbol{M}\boldsymbol{R}\big)$ iff $\boldsymbol{R} = \mathbf{I}_p$, and proper, since the actions of compact Lie groups ($\mathcal{O}(p)$ here) are always proper. This concludes the proof. □

Projecting onto the Horizontal Space

Let $\gamma : t \to (\boldsymbol{W}\boldsymbol{R}(t), \boldsymbol{R}(t)^T\boldsymbol{M}\boldsymbol{R}(t))$, with $\boldsymbol{R}(t) \in \mathcal{O}_p$, be a curve on $\mathcal{M}_p$ passing through $(\boldsymbol{W}, \boldsymbol{M})$ at $t = 0$. The vertical space at $(\boldsymbol{W}, \boldsymbol{M})$ is given by $\dot{\gamma}(t)$ evaluated at $t = 0$, which yields

$$\dot{\gamma}(t)|_{t=0} = (\boldsymbol{W}\Theta, \boldsymbol{M}\Omega - \Theta\boldsymbol{M}),$$

with $\Theta^T = -\Theta \in \mathbb{R}^{p\times p}$. As such,

$$\mathcal{V}_{(\boldsymbol{W},\boldsymbol{M})}\mathcal{M}_p = \{(\boldsymbol{W}\Theta, \boldsymbol{M}\Theta - \Theta\boldsymbol{M}) : \;\; \Theta^T = -\Theta \in \mathbb{R}^{p\times p}\}.$$

A horizontal vector $(\xi_{\boldsymbol{W}}, \xi_{\boldsymbol{M}})$ is then defined by the condition

$$g_p\big((\xi_{\boldsymbol{W}}, \xi_{\boldsymbol{M}}), (\boldsymbol{W}\Theta, \boldsymbol{M}\Theta - \Theta\boldsymbol{M})\big) = 0 \; .$$

Expanding the metric yields

$$\begin{aligned}
&g_p\big((\xi_{\boldsymbol{W}}, \xi_{\boldsymbol{M}}), (\boldsymbol{W}\Theta, \boldsymbol{M}\Theta - \Theta\boldsymbol{M})\big) = 0\\
&\Rightarrow 2\,\mathrm{Tr}(\xi_{\boldsymbol{W}}^T\boldsymbol{W}\Theta) + \mathrm{Tr}(\xi_{\boldsymbol{M}}\boldsymbol{M}^{-1}(\boldsymbol{M}\Theta - \Theta\boldsymbol{M})\boldsymbol{M}^{-1}) = 0\\
&\Rightarrow 2\,\mathrm{Tr}(\xi_{\boldsymbol{W}}^T\boldsymbol{W}\Theta) + \mathrm{Tr}(\xi_{\boldsymbol{M}}\Theta\boldsymbol{M}^{-1}) - \mathrm{Tr}(\xi_{\boldsymbol{M}}\boldsymbol{M}^{-1}\Theta) = 0\\
&\Rightarrow \mathrm{Tr}\,\Big((2\xi_{\boldsymbol{W}}^T\boldsymbol{W} + \boldsymbol{M}^{-1}\xi_{\boldsymbol{M}} - \xi_{\boldsymbol{M}}\boldsymbol{M}^{-1})\Theta\Big) = 0 \; .
\end{aligned}$$

Since Θ is a skew-symmetric matrix, the above equality is satisfied iff $2\xi_{\boldsymbol{W}}^T \boldsymbol{W} + \boldsymbol{M}^{-1}\xi_{\boldsymbol{M}} - \xi_{\boldsymbol{M}}\boldsymbol{M}^{-1}$ is symmetric. That is

$$2\xi_{\boldsymbol{W}}^T \boldsymbol{W} + \boldsymbol{M}^{-1}\xi_{\boldsymbol{M}} - \xi_{\boldsymbol{M}}\boldsymbol{M}^{-1} = 2\boldsymbol{W}^T\xi_{\boldsymbol{W}} + \xi_{\boldsymbol{M}}\boldsymbol{M}^{-1} - \boldsymbol{M}^{-1}\xi_{\boldsymbol{M}} \ . \tag{7}$$

Let $(\xi_{\boldsymbol{W}}, \xi_{\boldsymbol{M}}) \in T_{(\boldsymbol{W},\boldsymbol{M})}\mathcal{M}_p$ be a tangent vector of the product manifold. Removing the vertical component results in the horizontal vector $(\xi_{\boldsymbol{W}} - \boldsymbol{W}\Theta, \xi_{\boldsymbol{M}} - \boldsymbol{M}\Theta + \Theta\boldsymbol{M}) \in \mathcal{H}_{(\boldsymbol{W},\boldsymbol{M})}\mathcal{M}$. As derived above, this vector should satisfy the symmetry constraint of Eq. 7. To obtain Θ, we plug the horizontal vector $(\xi_{\boldsymbol{W}} - \boldsymbol{W}\Theta, \xi_{\boldsymbol{M}} - \boldsymbol{M}\Theta + \Theta\boldsymbol{M})$ into Eq. 7, which yields

$$\begin{aligned}
& 2(\xi_{\boldsymbol{W}} - \boldsymbol{W}\Theta)^T\boldsymbol{W} + \boldsymbol{M}^{-1}(\xi_{\boldsymbol{M}} - \boldsymbol{M}\Theta + \Theta\boldsymbol{M}) - (\xi_{\boldsymbol{M}} - \boldsymbol{M}\Theta + \Theta\boldsymbol{M})\boldsymbol{M}^{-1} \\
& = 2\boldsymbol{W}^T(\xi_{\boldsymbol{W}} - \boldsymbol{W}\Theta) + (\xi_{\boldsymbol{M}} - \boldsymbol{M}\Theta + \Theta\boldsymbol{M})\boldsymbol{M}^{-1} - \boldsymbol{M}^{-1}(\xi_{\boldsymbol{M}} - \boldsymbol{M}\Theta + \Theta\boldsymbol{M}) \\
\Rightarrow \quad & 2\xi_{\boldsymbol{W}}^T\boldsymbol{W} - 2\Theta^T\boldsymbol{W}^T\boldsymbol{W} + \boldsymbol{M}^{-1}\xi_{\boldsymbol{M}} - \Theta + \boldsymbol{M}^{-1}\Theta\boldsymbol{M} - \xi_{\boldsymbol{M}}\boldsymbol{M}^{-1} + \boldsymbol{M}\Theta\boldsymbol{M}^{-1} - \Theta \\
& = 2\boldsymbol{W}^T\xi_{\boldsymbol{W}} - 2\boldsymbol{W}^T\boldsymbol{W}\Theta + \xi_{\boldsymbol{M}}\boldsymbol{M}^{-1} - \boldsymbol{M}\Theta\boldsymbol{M}^{-1} + \Theta - \boldsymbol{M}^{-1}\xi_{\boldsymbol{M}} + \Theta - \boldsymbol{M}^{-1}\Theta\boldsymbol{M} \\
\Rightarrow \quad & 2\xi_{\boldsymbol{W}}^T\boldsymbol{W} - 2\Theta^T + \boldsymbol{M}^{-1}\xi_{\boldsymbol{M}}^T + \boldsymbol{M}^{-1}\Theta\boldsymbol{M} - \Theta - \xi_{\boldsymbol{M}}^T\boldsymbol{M}^{-1} - \Theta + \boldsymbol{M}\Theta\boldsymbol{M}^{-1} \\
& = 2\boldsymbol{W}^T\xi_{\boldsymbol{W}} - 2\Theta + \xi_{\boldsymbol{M}}\boldsymbol{M}^{-1} - \boldsymbol{M}\Theta\boldsymbol{M}^{-1} + \Theta - \boldsymbol{M}^{-1}\xi_{\boldsymbol{M}} + \Theta - \boldsymbol{M}^{-1}\Theta\boldsymbol{M} \\
\Rightarrow \quad & \boldsymbol{M}^{-1}\Theta\boldsymbol{M} + \boldsymbol{M}\Theta\boldsymbol{M}^{-1} = \xi_{\boldsymbol{W}}^T\boldsymbol{W} - \boldsymbol{W}^T\xi_{\boldsymbol{W}} + \boldsymbol{M}^{-1}\xi_{\boldsymbol{M}} - \xi_{\boldsymbol{M}}\boldsymbol{M}^{-1} \\
\Rightarrow \quad & \Theta\boldsymbol{M}^2 + \boldsymbol{M}^2\Theta = \boldsymbol{M}\Big(\xi_{\boldsymbol{W}}^T\boldsymbol{W} - \boldsymbol{W}^T\xi_{\boldsymbol{W}} + \boldsymbol{M}^{-1}\xi_{\boldsymbol{M}} - \xi_{\boldsymbol{M}}\boldsymbol{M}^{-1}\Big)\boldsymbol{M} \ .
\end{aligned}$$

In other words, Θ can be obtained as the solution to a Sylvester equation.

Remark 1. *To have more insights on the structure of $\mathcal{M}$, it is useful to study the form of the tangent vectors on $\mathcal{M}$. Since both the Stiefel and SPD manifolds are homogeneous spaces, their geometry and hence the geometry of their product space can be studied by analyzing their structure around any predefined point. Choosing $\boldsymbol{W} = [\mathbf{I}_p \ , \ \mathbf{0}_{p\times(n-p)}]^T$ and following the derivation above, it can be shown that the tangent vectors of $\mathcal{M}$ can be represented as*

$$\left\{ \left(\boldsymbol{U} \begin{bmatrix} \boldsymbol{V}\boldsymbol{M}^{-1} - \boldsymbol{M}^{-1}\boldsymbol{V} \\ \boldsymbol{B} \end{bmatrix}, \boldsymbol{V} \right) : \boldsymbol{V} \in \mathrm{Sym}(p), \ \boldsymbol{B} \in \mathbb{R}^{(n-p)\times p} \right\},$$

with $\boldsymbol{U} \in \mathcal{O}_n$. In the form considered above, a basis for the horizontal subspace can be formed from bases derived from $\boldsymbol{B} \in \boldsymbol{M}_{(n-p)\times p}$ and $\boldsymbol{V} \in \mathrm{Sym}(p)$. The basis elements for the $\boldsymbol{V}$ part are of the form

$$\big([\boldsymbol{V}_i\boldsymbol{M}^{-1} - \boldsymbol{M}^{-1}\boldsymbol{V}_i; \mathbf{0}], \boldsymbol{V}_i\big)$$

where the $\boldsymbol{V}_i$ are basis elements for $\mathrm{Sym}(p)$. We note that this form of horizontal projection cannot be used in first order optimization techniques as the basis elements are not orthogonal.

Optimization on $\mathcal{M}$

Given the definitions above, the only missing quantities to perform first order optimization on $\mathcal{M}$ are the Euclidean gradients $\nabla_{\boldsymbol{W}}$ and $\nabla_{\boldsymbol{M}}$ of the cost function in Eq. 10 of the main paper. Defining $\delta_i = \boldsymbol{x}_i - \tilde{\boldsymbol{x}}_i$, these gradients can be computed as

$$\nabla_{\boldsymbol{W}}(\mathcal{L}(\mathbb{X}, \boldsymbol{W}, \boldsymbol{M})) = 2\beta\Big(\sum_{i|y_i=1}\frac{p_i}{1+p_i}\delta_i\delta_i^T - \sum_{i|y_i=0}\frac{1}{1+p_i}\delta_i\delta_i^T\Big)\boldsymbol{W}\boldsymbol{M} \ , \tag{8}$$

$$\nabla_{\boldsymbol{M}}(\mathcal{L}(\mathbb{X}, \boldsymbol{W}, \boldsymbol{M})) = \beta\sum_{i|y_i=1}\frac{p_i}{1+p_i}\boldsymbol{W}^T\delta_i\delta_i^T\boldsymbol{W} - \beta\sum_{i|y_i=0}\frac{1}{1+p_i}\boldsymbol{W}^T\delta_i\delta_i^T\boldsymbol{W} + \nabla_{\boldsymbol{M}}(r(\boldsymbol{M}, \boldsymbol{M}_0)) \ . \tag{9}$$

For the Burg regularizer, $\nabla_{\boldsymbol{M}}(r(\boldsymbol{M}, \boldsymbol{M}_0))$ simply becomes $\boldsymbol{M}_0^{-1} - \boldsymbol{M}^{-1}$.

References

Lee, John M. *Smooth manifolds*. Springer, 2003.