Fixed-rank PSD check

October 13, 2023

Given integers $n \in \mathbb{N}$ and $r \in [n]$, we examine the set of $n \times n$ rank-r positive semidefinite matrices, $\mathbb{S}_+^{n,r}$. (Note that $\mathbb{S}_+^{n,n} = \mathbb{S}_{++}^n$.)

1 Geometry

We will consider the map

$$\pi : \mathrm{St}(r,n) \times \mathbb{S}_{++}^r \to \mathbb{S}_+^{n,r}$$

given by

$$\pi(U,P) = UPU^\top, \quad \forall U \in \mathrm{St}(r,n), P \in \mathbb{S}_{++}^r.$$

We also consider a group action $\cdot : (\mathrm{St}(r,n) \times \mathbb{S}_{++}^r) \times O(r) \to \mathrm{St}(r,n) \times \mathbb{S}_{++}^r$ given by

$$(U,P) \cdot V \mapsto (UV, V^\top PV), \forall (U,P) \in \mathrm{St}(r,n) \times \mathbb{S}_{++}^r, V \in O(r).$$

Since $((U,P) \cdot V) \cdot W = (U,P) \cdot VW$ for any $V, W \in O(r)$, this is a *right* action. We can check that

$$\pi(U,P) = \pi((U,P) \cdot V), \quad \forall V \in O(r).$$

In fact, the following holds:

Lemma 1.1. *We have*

$$\pi(U,P) = \pi(U',P') \iff \exists V \in O(r) \text{ s.t. } (U',V') = (U,P) \cdot V.$$

Proof. TO DO. ☐

Also, for any $(U,P) \in \tilde{M}$, we have a linear map $d\pi_{U,P} : T_{U,P}\tilde{M} \to T_{\pi(U,P)}M$. We also know that π is a *submersion*, hence $d\pi_{U,P}$ is surjective.

For notational ease, define

$$\tilde{M} := \mathrm{St}(r,n) \times \mathbb{S}_{++}^r, \quad G := O(r), \quad M := \mathbb{S}_+^{n \times r}.$$

The tangent spaces on $\tilde{M}$ are given by

$$T_{U,P}\tilde{M} = T_U \,\mathrm{St}(r,n) \oplus T_P \mathbb{S}_{++}^r = \left\{ UA + U^\perp B : A \in \mathrm{Skew}(r), B \in \mathbb{R}^{(n-r)\times r} \right\} \oplus \mathbb{S}^r.$$

We will take our metric to be:

$$\left\langle (UA + U^\perp B, D), (UA' + U^\perp B', D') \right\rangle_{U,P} := \frac{1}{2} \mathrm{Tr}(A^\top A') + \mathrm{Tr}(B^\top B') + \mathrm{Tr}(P^{-1/2} D P^{-1} D' P^{-1/2}). \tag{1}$$

This is simply the sum of two known metrics on $\mathrm{St}(r,n)$ and $\mathbb{S}^r_{++}$, i.e., it is a product metric. Therefore the exponential map on $\tilde{M}$, denoted $\widetilde{\mathrm{Exp}}$, is given by

$$\begin{aligned}\widetilde{\mathrm{Exp}}_{U,P}\left(UA + U^{\perp}B, D\right) &:= \left(\mathrm{Exp}_U(UA + U^{\perp}B), \mathrm{Exp}_P(D)\right)\\ &= \left(\begin{bmatrix}U & U^{\perp}\end{bmatrix}\mathrm{Expm}\left(\begin{bmatrix}A & -B^{\top}\\ B & 0_{n-r,n-r}\end{bmatrix}\right)\begin{bmatrix}I_r\\ 0_{n-r,r}\end{bmatrix}, P^{1/2}\,\mathrm{Expm}(P^{-1/2}WP^{-1/2})P^{1/2}\right).\end{aligned}$$

Rationale. We are interested in optimization over M, and we use π to map points from $\tilde{M}$ to M. It is easy to describe a Riemannian structure over $\tilde{M}$, which is a product of the Stiefel manifold and the positive definite manifold, as Riemannian structures over each of these are well-studied. We have described one such structure for $\tilde{M}$ above. However, we are interested in M. By the *quotient manifold theorem*, M may inherit a Riemannian structure from $\tilde{M}$. It turns out that, in order to operate on M, we may instead consider operations on $\tilde{M}$, which are often much easier to work with. To describe this precisely, we need to introduce the notion of *vertical* and *horizontal* spaces.

2 Subspace decomposition and exponential map updates

Given $(U,P) \in \tilde{M}$, the *vertical space* is a subspace of $T_{U,P}M$ defined as

$$\tilde{V}_{U,P} := \ker(d\pi_{U,P}) := \left\{\left(UA + U^{\perp}B, D\right) \in T_{\tilde{M}}(U,P) : d\pi_{U,P}\left(UA + U^{\perp}B, D\right) = 0 \in T_{\pi(U,P)}M\right\}.$$

The *horizontal space* is

$$\tilde{H}_{U,P} := \tilde{V}_{U,P}^{\perp}$$

where the orthogonal complement is taken with respect to the Riemannian metric.

Lemma 2.1. *For any $(U,P) \in \tilde{M}$, we have*

$$\begin{aligned}\tilde{V}_{U,P} &= \left\{(UA, PA - AP) \in T_{\tilde{M}}(U,P) : A \in \mathrm{Skew}(r)\right\}\\ \tilde{H}_{U,P} &= \left\{\left(U(P^{-1}D - DP^{-1}) + U^{\perp}B, \frac{1}{2}D\right) : B \in \mathbb{R}^{(n-r)\times r}, D \in \mathbb{S}^r\right\}.\end{aligned}$$

Proof. TO DO. □

Suppose that P has the following spectral decomposition: $P = Q\Lambda Q^{\top}$ where $Q \in O(r)$ and $\Lambda = \mathrm{Diag}(\lambda_1, \ldots, \lambda_r) \succ 0$. For $i, j \in [r]$, denote

$$D_{ij}^P := Q\Lambda^{1/2}(e_ie_j^{\top} + e_je_i^{\top})\Lambda^{1/2}Q^{\top}.$$

If we denote the eigenvectors of P (equivalently, columns of Q) by $q_1, \ldots, q_r$, then we have

$$D_{ij}^P = (\lambda_i\lambda_j)^{1/2}(q_iq_j^{\top} + q_jq_i^{\top}).$$

We propose the following subspace decomposition for $\tilde{H}_{U,P}$:

$$\mathcal{S}_k^{U,P} := \left\{(vq_k^{\top}, 0) : v \in \ker(U^{\top})\right\}, \quad k \in [r] \tag{2a}$$

$$\mathcal{S}_{ij}^{U,P} := \left\{t\left(U(P^{-1}D_{ij}^P - D_{ij}^P P^{-1}), \frac{1}{2}D_{ij}^P\right) : t \in \mathbb{R}\right\}, \quad 1 \le i \le j \le r. \tag{2b}$$

Proposition 2.2. *The subspaces described in* (2) *are mutually orthogonal with respect to metric* (1).

Proof. It is clear that $\mathcal{S}_k^{U,P}$ and $\mathcal{S}_{ij}^{U,P}$ are orthogonal, and that $\{\mathcal{S}_k^{U,P} : k \in \mathbb{R}\}$ are mutually orthogonal. We will prove that $\mathcal{S}_{ij}^{U,P}$ and $\mathcal{S}_{kl}^{U,P}$ are orthogonal whenever $(i,j) \neq (k,l)$.

To see this, observe that

$$\begin{aligned}
&\left\langle \left(U(P^{-1}D_{ij}^P - D_{ij}^P P^{-1}), \frac{1}{2}D_{ij}^P \right), \left(U(P^{-1}D_{kl}^P - D_{kl}^P P^{-1}), \frac{1}{2}D_{kl}^P \right) \right\rangle_{U,P} \\
&= \frac{1}{2}\operatorname{Tr}\left(\left(P^{-1}D_{ij}^P - D_{ij}^P P^{-1}\right)^\top \left(P^{-1}D_{kl}^P - D_{kl}^P P^{-1}\right) \right) \\
&\quad + \frac{1}{4}\operatorname{Tr}\left(\left(P^{-1/2}D_{ij}^P P^{-1/2}\right)^\top \left(P^{-1/2}D_{kl}^P P^{-1/2}\right) \right).
\end{aligned}$$

Denote $S_{ij} = q_i q_j^\top + q_j q_i^\top$, and analogously for S_{kl}.

For the second term, notice that by definition, we have $P^{-1/2}D_{ij}^P P^{-1/2} = S_{ij}$, and analogously for $P^{-1/2}D_{kl}^P P^{-1/2}$. Therefore

$$\operatorname{Tr}\left(\left(P^{-1/2}D_{ij}^P P^{-1/2}\right)^\top \left(P^{-1/2}D_{kl}^P P^{-1/2}\right) \right) = \operatorname{Tr}(S_{ij}S_{kl}).$$

This is 0 unless $(i,j) = (k,l)$ or $(i,j) = (l,k)$.

For the first term, notice that $P^{-1}D_{ij}^P - D_{ij}^P P^{-1} = (\lambda_i \lambda_i)^{1/2}(P^{-1}S_{ij} - S_{ij}P^{-1}) = (\lambda_i\lambda_i)^{1/2}(\lambda_i^{-1}q_i q_j^\top + \lambda_j^{-1}q_j q_i^\top - \lambda_j^{-1}q_i q_j^\top - \lambda_i^{-1}q_j q_i^\top) = (\lambda_i\lambda_j)^{1/2}(\lambda_i^{-1} - \lambda_j^{-1})(q_i q_j^\top - q_j q_i^\top)$. Denote $H_{ij} = q_i q_j^\top - q_j q_i^\top$ and analogously for H_{kl}. Therefore

$$\begin{aligned}
&\left(P^{-1}D_{ij}^P - D_{ij}^P P^{-1}\right)^\top \left(P^{-1}D_{kl}^P - D_{kl}^P P^{-1}\right) \\
&= (\lambda_i\lambda_j)^{1/2}(\lambda_i^{-1} - \lambda_j^{-1})(\lambda_k\lambda_l)^{1/2}(\lambda_k^{-1} - \lambda_l^{-1})H_{ij}^\top H_{kl}.
\end{aligned}$$

Therefore

$$\operatorname{Tr}\left(\left(P^{-1}D_{ij}^P - D_{ij}^P P^{-1}\right)^\top \left(P^{-1}D_{kl}^P - D_{kl}^P P^{-1}\right) \right) = (\lambda_i\lambda_j)^{1/2}(\lambda_i^{-1} - \lambda_j^{-1})(\lambda_k\lambda_l)^{1/2}(\lambda_k^{-1} - \lambda_l^{-1})\operatorname{Tr}(H_{ij}^\top H_{kl}).$$

This is equal to 0 unless $(i,j) = (k,l)$ or $(i,j) = (l,k)$.

Therefore, combining both cases,

$$\left\langle \left(U(P^{-1}D_{ij}^P - D_{ij}^P P^{-1}), \frac{1}{2}D_{ij}^P \right), \left(U(P^{-1}D_{kl}^P - D_{kl}^P P^{-1}), \frac{1}{2}D_{kl}^P \right) \right\rangle_{U,P} = 0$$

unless $(i,j) = (k,l)$ or $(i,j) = (l,k)$. However, since $i \leq j$, $k \leq l$, this only occurs if $i = k$, $j = l$. Therefore the subspaces $\mathcal{S}_{ij}^{U,P}$ are mutually orthogonal. $\square$

We now compute the exponential map for vectors from our subspaces.

Proposition 2.3. *Fix $t \in \mathbb{R}$. Then:*

- *For any $k \in [r]$ and $v \in \ker(U^\top)$ we have*

$$\widetilde{\operatorname{Exp}}_{U,P}\left(\tau(vq_k^\top, 0)\right) = \left(U(I - q_k q_k^\top) + \left(\cos(\tau\|v\|_2)Uq_k + \sin(\tau\|v\|_2)\frac{Uv}{\|v\|_2} \right) q_k^\top, P \right).$$

- *For any $1 \le i \le j \le r$, we have*

$$\widetilde{\mathrm{Exp}}_{U,P}\left(\tau U(P^{-1}D_{ij}^P - D_{ij}^P P^{-1}), \frac{\tau}{2} D_{ij}^P\right) = (U_{+,ij}(\tau), P_{+,ij}(\tau))$$

where

$$\begin{aligned}
U_{+,ij}(\tau) &:= U(I - q_i q_i^\top - q_j q_j^\top) + (\cos(\alpha_{ij}(\tau))Uq_i - \sin(\alpha_{ij}(\tau))Uq_j)q_i^\top \\
&\quad + (\sin(\alpha_{ij}(\tau))Uq_i + \cos(\alpha_{ij}(\tau))Uq_j)q_j^\top \\
\alpha_{ij}(\tau) &:= \tau(\lambda_i\lambda_j)^{1/2}(\lambda_i^{-1} - \lambda_j^{-1}) \\
P_{+,ij}(\tau) &:= P - (\lambda_i q_i q_i^\top + \lambda_j q_j q_j^\top) + \cosh(\tau/2)(\lambda_i q_i q_i^\top + \lambda_j q_j q_j^\top) + \sinh(\tau/2)\sqrt{\lambda_i\lambda_j}(q_i q_j^\top + q_j q_i^\top).
\end{aligned}$$

Proof. The first result follows from [3, Lemma 5].

For the second result, note from the proof of Proposition 2.2, we have $P^{-1/2}D_{ij}^P P^{-1/2} = S_{ij}$ and $P^{-1}D_{ij}^P - D_{ij}^P P^{-1} = (\lambda_i\lambda_j)^{1/2}(\lambda_i^{-1} - \lambda_j^{-1})H_{ij}$, where $S_{ij} = q_i q_j^\top + q_j q_i^\top$ and $H_{ij} = q_i q_j^\top - q_j q_i^\top$. Then

$$\widetilde{\mathrm{Exp}}_{U,P}\left(\tau U(P^{-1}D_{ij}^P - D_{ij}^P P^{-1}), \frac{\tau}{2} D_{ij}^P\right) = \left(U \,\mathrm{Expm}\left(\tau(\lambda_i\lambda_j)^{1/2}(\lambda_i^{-1} - \lambda_j^{-1})H_{ij}\right), P^{1/2}\,\mathrm{Expm}((\tau/2)S_{ij})P^{1/2}\right).$$

It is straightforward to check that $\mathrm{Expm}\,(\alpha H_{ij})$ is a Givens rotation in the subspace spanned by q_i and q_j, i.e.,

$$\mathrm{Expm}(\alpha H_{ij}) = I_r - (q_i q_i^\top + q_j q_j^\top) + \cos(\alpha)(q_i q_i^\top + q_j q_j^\top) + \sin(\alpha)(q_i q_j^\top - q_j q_i^\top).$$

Therefore

$$\begin{aligned}
&U\,\mathrm{Expm}\left(\tau(\lambda_i\lambda_j)^{1/2}(\lambda_i^{-1} - \lambda_j^{-1})H_{ij}\right) \\
&= U(I_r - q_i q_i^\top - q_j q_j^\top) + (\cos(\alpha_{ij})Uq_i - \sin(\alpha_{ij})Uq_j)q_i^\top + (\sin(\alpha_{ij})Uq_i + \cos(\alpha_{ij})Uq_j)q_j^\top,
\end{aligned}$$

where $\alpha_{ij} = \tau(\lambda_i\lambda_j)^{1/2}(\lambda_i^{-1} - \lambda_j^{-1})$.

We also have

$$\mathrm{Expm}(\alpha S_{ij}) = I_r - (q_i q_i^\top + q_j q_j^\top) + \cosh(\alpha)(q_i q_i^\top + q_j q_j^\top) + \sinh(\alpha)(q_i q_j^\top + q_j q_i^\top),$$

hence, with $\alpha = t/2$, we have

$$\begin{aligned}
&P^{1/2}\,\mathrm{Expm}(P^{-1/2}(t/2)D_{ij}^P P^{-1/2})P^{1/2} \\
&= P^{1/2}\,\mathrm{Expm}(\alpha S_{ij})P^{1/2} \\
&= P - (\lambda_i q_i q_i^\top + \lambda_j q_j q_j^\top) + \cosh(\alpha)(\lambda_i q_i q_i^\top + \lambda_j q_j q_j^\top) + \sinh(\alpha)\sqrt{\lambda_i\lambda_j}(q_i q_j^\top + q_j q_i^\top).
\end{aligned}$$

□

Storage. Given $1 \le i \le j \le r$, to compute $U_{+,ij}(\tau), P_{+,ij}(\tau)$, we need to have access to the following quantities:

- Eigenvectors q_i, q_j and corresponding eigenvalues λ_i, λ_j for P.
- The products $Uq_i, Uq_j \in \mathbb{R}^n$.

We now explore how to efficiently update these quantities.

First, we analyse the spectral decomposition of $P_{+,ij}(\tau)$. Denote $\alpha = \tau/2$. Then

$$P_{+,ij}(\tau) = P - (\lambda_i q_i q_i^\top + \lambda_j q_j q_j^\top) + \cosh(\alpha)(\lambda_i q_i q_i^\top + \lambda_j q_j q_j^\top) + \sinh(\alpha)\sqrt{\lambda_i \lambda_j}(q_i q_j^\top + q_j q_i^\top).$$

Denote

$$\begin{aligned} P_{ij}(\alpha) &:= P_{+,ij}(\tau) - (P - (\lambda_i q_i q_i^\top + \lambda_j q_j q_j^\top)) \\ &= \cosh(\alpha)(\lambda_i q_i q_i^\top + \lambda_j q_j q_j^\top) + \sinh(\alpha)\sqrt{\lambda_i \lambda_j}(q_i q_j^\top + q_j q_i^\top) \\ &= Q_{ij} M(\lambda_i, \lambda_j, \alpha) Q_{ij}^\top \end{aligned}$$

where $Q_{ij} := \begin{bmatrix} q_i & q_j \end{bmatrix} \in \mathbb{R}^{r \times 2}$ and $M(\lambda_i, \lambda_j, \alpha)$ is the following 2×2 symmetric matrix:

$$M(\lambda_i, \lambda_j, \alpha) := \begin{bmatrix} \cosh(\alpha)\lambda_i & \sinh(\alpha)\sqrt{\lambda_i\lambda_j} \\ \sinh(\alpha)\sqrt{\lambda_i\lambda_j} & \cosh(\alpha)\lambda_j \end{bmatrix} = \cosh(\alpha)\lambda_j \begin{bmatrix} \lambda_i/\lambda_j & \tanh(\alpha)\sqrt{\lambda_i/\lambda_j} \\ \tanh(\alpha)\sqrt{\lambda_i/\lambda_j} & 1 \end{bmatrix}.$$

Finding the updated eigenvalues/vectors is equivalent to computing the spectral decomposition of $M(\lambda_i, \lambda_j, \alpha)$. To see this, write the spectral decomposition as

$$M(\lambda_i, \lambda_j, \alpha) = \begin{bmatrix} v_1(\alpha, \lambda_i, \lambda_j) & v_2(\alpha, \lambda_i, \lambda_j) \end{bmatrix} \begin{bmatrix} \gamma_1(\alpha, \lambda_i, \lambda_j) & 0 \\ 0 & \gamma_2(\alpha, \lambda_i, \lambda_j) \end{bmatrix} \begin{bmatrix} v_1^\top(\alpha, \lambda_i, \lambda_j) \\ v_2^\top(\alpha, \lambda_i, \lambda_j). \end{bmatrix}.$$

Then

$$P_{ij}(\alpha) = \begin{bmatrix} Q_{ij} v_1(\alpha, \lambda_i, \lambda_j) & Q_{ij} v_2(\alpha, \lambda_i, \lambda_j) \end{bmatrix} \begin{bmatrix} \gamma_1(\alpha, \lambda_i, \lambda_j) & 0 \\ 0 & \gamma_2(\alpha, \lambda_i, \lambda_j) \end{bmatrix} \begin{bmatrix} v_1(\alpha, \lambda_i, \lambda_j)^\top Q_{ij}^\top \\ v_2(\alpha, \lambda_i, \lambda_j)^\top Q_{ij}^\top \end{bmatrix}.$$

It is easy to check that $Q_{ij} v_1(\alpha, \lambda_i, \lambda_j)$ and $Q_{ij} v_2(\alpha, \lambda_i, \lambda_j)$ have norm 1, are orthonormal, and that they are also orthogonal to every other eigenvector of P. Therefore they are eigenvectors of $P^{1/2} \operatorname{Expm}(\alpha S_{ij}) P^{1/2}$ with corresponding eigenvalues $\gamma_1(\alpha, \lambda_i, \lambda_j), \gamma_2(\alpha, \lambda_i, \lambda_j)$. Explicit formulas for the spectral decomposition are as follows:

$$\begin{aligned} \gamma_1(\alpha, \lambda_i, \lambda_j) &= \frac{\cosh(\alpha)}{2}\left(\lambda_i + \lambda_j - \sqrt{(\lambda_i - \lambda_j)^2 + 4\tanh^2(\alpha)\lambda_i\lambda_j}\right) \\ \gamma_2(\alpha, \lambda_i, \lambda_j) &= \frac{\cosh(\alpha)}{2}\left(\lambda_i + \lambda_j + \sqrt{(\lambda_i - \lambda_j)^2 + 4\tanh^2(\alpha)\lambda_i\lambda_j}\right) \end{aligned}$$

and the corresponding (un-normalized) eigenvectors are

$$\begin{aligned} \bar{v}_1(\alpha, \lambda_i, \lambda_j) &= \begin{bmatrix} \lambda_i - \lambda_j - \sqrt{(\lambda_i - \lambda_j)^2 + 4\tanh^2(\alpha)\lambda_i\lambda_j} \\ 2\tanh(\alpha)\sqrt{\lambda_i\lambda_j} \end{bmatrix}, \quad v_1(\alpha, \lambda_i, \lambda_j) = \frac{\bar{v}_1(\alpha, \lambda_i, \lambda_j)}{\|\bar{v}_1(\alpha, \lambda_i, \lambda_j)\|_2} \\ \bar{v}_2(\alpha, \lambda_i, \lambda_j) &= \begin{bmatrix} \lambda_i - \lambda_j + \sqrt{(\lambda_i - \lambda_j)^2 + 4\tanh^2(\alpha)\lambda_i\lambda_j} \\ 2\tanh(\alpha)\sqrt{\lambda_i\lambda_j} \end{bmatrix}, \quad v_2(\alpha, \lambda_i, \lambda_j) = \frac{\bar{v}_2(\alpha, \lambda_i, \lambda_j)}{\|\bar{v}_2(\alpha, \lambda_i, \lambda_j)\|_2}. \end{aligned}$$

We also need to update the product of $U_{+,ij}(\tau)$ with the new eigenvectors $Q_{ij} v_1(\tau/2, \lambda_i, \lambda_j)$ and

$Q_{ij}v_2(\tau/2, \lambda_i, \lambda_j)$. This is computed as

$$\begin{aligned}
U_{+,ij}(\tau)Q_{ij}v_1(\tau/2, \lambda_i, \lambda_j) &= U(I - q_iq_i^\top - q_jq_j^\top)Q_{ij}v_1(\tau/2, \lambda_i, \lambda_j) \\
&\quad + (\cos(\alpha_{ij}(\tau))Uq_i - \sin(\alpha_{ij}(\tau))Uq_j)q_i^\top Q_{ij}v_1(\tau/2, \lambda_i, \lambda_j) \\
&\quad + (\sin(\alpha_{ij}(\tau))Uq_i + \cos(\alpha_{ij}(\tau))Uq_j)q_j^\top Q_{ij}v_1(\tau/2, \lambda_i, \lambda_j) \\
&= \begin{bmatrix} Uq_i & Uq_j \end{bmatrix} \begin{bmatrix} \cos(\alpha_{ij}(\tau)) & \sin(\alpha_{ij}(\tau)) \\ -\sin(\alpha_{ij}(\tau)) & \cos(\alpha_{ij}(\tau)) \end{bmatrix} v_1(\tau/2, \lambda_i, \lambda_j) \\
U_{+,ij}(\tau)Q_{ij}v_2(\tau/2, \lambda_i, \lambda_j) &= \begin{bmatrix} Uq_i & Uq_j \end{bmatrix} \begin{bmatrix} \cos(\alpha_{ij}(\tau)) & \sin(\alpha_{ij}(\tau)) \\ -\sin(\alpha_{ij}(\tau)) & \cos(\alpha_{ij}(\tau)) \end{bmatrix} v_2(\tau/2, \lambda_i, \lambda_j).
\end{aligned}$$

With these formulas, we can store and efficiently update eigenvectors $q_1, \ldots, q_r$ of P, corresponding eigenvalues $\lambda_1, \ldots, \lambda_r$ and products $Uq_1, \ldots, Uq_r$ (which fully determine U).

Note about updating indices: TO DO. Need to choose them to be consistent.

3 Gradients

Recall $\tilde{M} = \mathrm{St}(r, n) \times \mathbb{S}^r_{++}$ and $M = \mathbb{S}^{n,r}_{+}$. Now suppose we have a function $f : M \to \mathbb{R}$. We lift this function to $\tilde{f} = f \circ \pi : \tilde{M} \to \mathbb{R}$, recalling that $\pi : \tilde{M} \to M$ is the Riemannian submersion. We will show how to compute gradients for $\tilde{f}$.

Suppose now that $\tilde{f}$ is a *generic* function from $\tilde{M} \to \mathbb{R}$, which can be extended to a function $\tilde{\tilde{f}} : \mathbb{R}^{n\times r} \times \mathbb{R}^{r\times r} \to \mathbb{R}$ (or some open set containing $\tilde{M}$ in $\mathbb{R}^{n\times r} \times \mathbb{R}^{r\times r}$). Thus at a point $(U, P) \in \tilde{M}$, we may compute the *Euclidean* gradient $\nabla \tilde{\tilde{f}}(U, P) = (G_U, G_P)$. We will first provide a formula for the *Riemannian* gradient $\nabla \tilde{f}(U, P)$.

Lemma 3.1. *Given $(U, P) \in \tilde{M}$ and a Euclidean gradient $\nabla \tilde{\tilde{f}}(U, P) = (G_U, G_P)$, the Riemannian gradient of $\tilde{f}$ at (U, P) is*

$$\nabla \tilde{f}(U, P) = (U\tilde{A} + U^\perp \tilde{B}, \tilde{D}) \in T_{U,P}\tilde{M},$$

where

$$\begin{aligned}
\tilde{A} &= U^\top G_U - G_U^\top U \in \mathrm{Skew}(r) \\
\tilde{B} &= (U^\perp)^\top G_U \in \mathbb{R}^{(n-r)\times r} \\
\tilde{D} &= \frac{1}{2} P \left(G_P + G_P^\top \right) P \in \mathbb{S}^r.
\end{aligned}$$

Proof. We know that the Riemannian gradient satisfies

$$D\tilde{f}(U, P)[UA + U^\perp B, D] = \langle \nabla \tilde{f}(U, P), (UA + U^\perp B, D) \rangle_{U,P}$$

for any tangent vector $(UA + U^\perp B, D) \in T_{U,P}\tilde{M}$. Furthermore, Boumal [2, Eq. (3.36)] states that

$$D\tilde{f}(U, P)[UA + U^\perp B, D] = D\tilde{\tilde{f}}(U, P)[UA + U^\perp B, D] = \langle \nabla \tilde{\tilde{f}}(U, P), (UA + U^\perp B, D) \rangle$$

with the Euclidean inner product. Now denoting $\nabla \tilde{f}(U, P) = (U\tilde{A} + U^\perp \tilde{B}, \tilde{D})$, we observe that

$$\begin{aligned}
\langle \nabla \tilde{f}(U, P), (UA + U^\perp B, D) \rangle_{U,P} &= \frac{1}{2} \langle \tilde{A}, A \rangle + \langle \tilde{B}, B \rangle + \langle P^{-1} \tilde{D} P^{-1}, D \rangle \\
\langle \nabla \tilde{\tilde{f}}(U, P), (UA + U^\perp B, D) \rangle &= \langle U^\top G_U, A \rangle + \langle (U^\perp)^\top G_U, B \rangle + \langle G_P, D \rangle.
\end{aligned}$$

These two terms must be equal for all $A \in \mathrm{Skew}(r)$, $B \in \mathbb{R}^{(n-r)\times r}$ and $D \in \mathbb{S}^r$. Note that $\mathrm{Skew}(r)^\perp = \mathbb{S}^r$ and vice versa, and $(\mathbb{R}^{(n-r)\times r})^\perp = \{0\}$. Therefore there exists $S \in \mathbb{S}^r$ and $K \in \mathrm{Skew}(r)$ such that

$$\begin{aligned} \tilde{A} &= 2U^\top G_U + S \\ \tilde{B} &= (U^\perp)^\top G_U \\ P^{-1}\tilde{D}P^{-1} &= G_P + K. \end{aligned}$$

Since $\tilde{A} \in \mathrm{Skew}(r)$, we have

$$0 = \tilde{A} + \tilde{A}^\top = 2(U^\top G_U + G_U^\top U) + 2S \implies S = -(U^\top G_U + G_U^\top U) \implies \tilde{A} = U^\top G_U - G_U^\top U.$$

Since $\tilde{D} \in \mathbb{S}^r$, we have

$$G_P^\top + K^\top = P^{-1}\tilde{D}^\top P^{-1} = P^{-1}\tilde{D}P^{-1} = G_P + K \implies K = \frac{1}{2}(G_P^\top - G_P) \implies P^{-1}\tilde{D}P^{-1} = \frac{1}{2}(G_P + G_P^\top).$$

Multiplying on the left and right by P gives the result. $\square$

Now we know that when $\tilde{f} = f \circ \pi$, the gradients are horizontal vectors, i.e., $\nabla \tilde{f}(U,P) \in \tilde{H}_{U,P}$. We verify this holds when f can be extended to a function $\bar{f} : \mathbb{R}^{n\times n} \to \mathbb{R}$. Note that in this case $\tilde{f}(U,P) = \tilde{\bar{f}}(U,P) = \bar{f}(UPU^\top)$. Using the chain rule, we derive the following expressions for $\nabla \tilde{\bar{f}}(U,P)$:

$$\begin{aligned} \langle \nabla_U \tilde{\bar{f}}(U,P), V\rangle &= \frac{d}{dt}\ \bar{f}((U+tV)P(U^\top + tV^\top))\Big|_{t=0} \\ &= \frac{d}{dt}\ \bar{f}(UPU^\top + t(VPU^\top + UPV^\top) + t^2 VPV^\top)\Big|_{t=0} \\ &= \langle \nabla \bar{f}(UPU^\top), VPU^\top + UPV^\top\rangle \\ &= \langle V, (\nabla \bar{f}(UPU^\top) + \nabla \bar{f}(UPU^\top)^\top)UP\rangle \\ \implies \nabla_U \tilde{\bar{f}}(U,P) &= (\nabla \bar{f}(UPU^\top) + \nabla \bar{f}(UPU^\top)^\top)UP \\ \langle \nabla_P \tilde{\bar{f}}(U,P), Q\rangle &= \frac{d}{dt}\ \bar{f}(U(P+tQ)U^\top)\Big|_{t=0} = \frac{d}{dt}\ \bar{f}(UPU^\top + tUQU^\top)\Big|_{t=0} \\ &= \langle \nabla \bar{f}(UPU^\top), UQU^\top\rangle \\ \implies \nabla_P \tilde{\bar{f}}(U,P) &= U^\top \nabla \bar{f}(UPU^\top)U. \end{aligned}$$

Note that in the above, $\nabla \bar{f}(UPU^\top)$ is the *Euclidean* gradient of $\bar{f}$, and inner products are taken with respect to the Euclidean geometry. Denoting $G = \nabla \bar{f}(UPU^\top)$, we have

$$G_U = (G + G^\top)UP, \quad G_P = U^\top G U.$$

With these expressions, the Riemannian gradient $\nabla \tilde{f}(U,P)$ has the following expression:

$$\begin{aligned} \nabla \tilde{f}(U,P) &= (U\tilde{A} + U^\perp \tilde{B}, \tilde{D}) \\ \text{where } \tilde{A} &= U^\top (G+G^\top)UP - PU^\top(G+G^\top)U \\ \tilde{B} &= (U^\perp)^\top (G+G^\top)UP \\ \tilde{D} &= \frac{1}{2}PU^\top(G+G^\top)UP. \end{aligned}$$

This is horizontal because

$$\tilde{A} = P^{-1}PU^\top(G+G^\top)UP - PU^\top(G+G^\top)UPP^{-1} = 2(P^{-1}\tilde{D} - \tilde{D}P^{-1}).$$

4 Projections of gradients

When f can be extended to a function $\bar{f} : \mathbb{R}^{n\times n} \to \mathbb{R}$, we have the following formula for $\nabla \tilde{f}(U, P)$:

$$\begin{aligned}
\nabla \tilde{f}(U, P) &= (U\tilde{A} + U^{\perp}\tilde{B}, \tilde{D}) \\
\text{where } \tilde{A} &= U^{\top}(G + G^{\top})UP - PU^{\top}(G + G^{\top})U \\
\tilde{B} &= (U^{\perp})^{\top}(G + G^{\top})UP \\
\tilde{D} &= \frac{1}{2}PU^{\top}(G + G^{\top})UP \\
G &= \nabla \bar{f}(UPU^{\top}).
\end{aligned}$$

We know that this is horizontal. As before, let $P = Q\Lambda Q^{\top}$, with eigenpairs (λ_i, q_i) for $i \in [r]$.

Lemma 4.1. *Let V be an inner product space with subspace $S \subset V$. Let $u \in V$. Then $\bar{v}$ is the projection of u onto S if and only if $u - \bar{v} \in S^{\perp}$.*

Proof. This follows by considering the optimality condition of minimizing $f(v) = \frac{1}{2}\langle u - v, u - v\rangle$ over $v \in S$, which is $\langle u - \bar{v}, v - \bar{v}\rangle \geq 0$ for all $v \in S$. Necessarily, we need $u - \bar{v} \in \bar{S}^{\perp}$. $\square$

Proposition 4.2. *Let $\mathcal{S}_k^{U,P} = \{(vq_k^{\top}, 0) : v \in \mathrm{Ker}(U^{\top})\}$ and $\mathcal{S}_{ij}^{U,P} = \{\gamma\left(U(P^{-1}D_{ij}^P - D_{ij}^P P^{-1}), \frac{1}{2}D_{ij}^P\right) : \gamma \in \mathbb{R}\}$ where $D_{ij}^P = \sqrt{\lambda_i\lambda_j}(q_i q_j^{\top} + q_j q_i^{\top})$. Then*

$$\begin{aligned}
\mathrm{Proj}_{\mathcal{S}_k^{U,P}}(\nabla \tilde{f}(U, P)) &= \left(U^{\perp}\tilde{B}q_k q_k^{\top}, 0\right) = \left(\lambda_k(I_n - UU^{\top})(G + G^{\top})Uq_k q_k^{\top}, 0\right) \\
\mathrm{Proj}_{\mathcal{S}_{ij}^{U,P}}(\nabla \tilde{f}(U, P)) &= \frac{\langle \nabla \tilde{f}(U, P), v_{ij}^P\rangle_{U,P}}{\|v_{ij}^P\|_{U,P}^2} v_{ij}^P = \sqrt{\lambda_i\lambda_j}(Uq_j)^{\top}(G + G^{\top})(Uq_i)v_{ij}^P \\
\textit{where } v_{ij}^P &:= \left(U(P^{-1}D_{ij}^P - D_{ij}^P P^{-1}), \frac{1}{2}D_{ij}^P\right).
\end{aligned}$$

Proof. We have

$$\nabla \tilde{f}(U, P) - (U^{\perp}\tilde{B}q_k q_k^{\top}, 0) = \left(U\tilde{A} + U^{\perp}\tilde{B}(I_r - q_k q_k^{\top}), \tilde{D}\right).$$

Therefore, for any $(vq_k^{\top}, 0) \in \mathcal{S}_k^{U,P}$, write $v = U^{\perp}w$, thus we have

$$\begin{aligned}
\left\langle \nabla \tilde{f}(U, P) - (U^{\perp}\tilde{B}q_k q_k^{\top}, 0), (vq_k^{\top}, 0)\right\rangle_{U,P} &= \left\langle \tilde{B}(I_r - q_k q_k^{\top}), wq_k^{\top}\right\rangle \\
&= \mathrm{Tr}\left(q_k^{\top}(I_r - q_k q_k^{\top})\tilde{B}^{\top}w\right) \\
&= 0.
\end{aligned}$$

Lemma 4.1 then states that $\left(U^{\perp}\tilde{B}q_k q_k^{\top}, 0\right)$ is exactly the projection.

Since $\mathcal{S}_{ij}^{U,P}$ is a one-dimensional subspace spanned by v_{ij}^P, the formula $\mathrm{Proj}_{\mathcal{S}_{ij}^{U,P}}(\nabla \tilde{f}(U, P)) = \frac{\langle \nabla \tilde{f}(U,P), v_{ij}^P\rangle_{U,P}}{\|v_{ij}^P\|_{U,P}^2} v_{ij}^P$ is immediate. We now compute the constants. As in the proof of Proposition 2.3, notice that $P^{-1}D_{ij}^P - D_{ij}^P P^{-1} = \sqrt{\lambda_i\lambda_i}(1/\lambda_i - 1/\lambda_j)H_{ij}$, where $H_{ij} = q_i q_j^{\top} - q_j q_i^{\top}$. Then

$$\langle P^{-1}D_{ij}^P - D_{ij}^P P^{-1}, P^{-1}D_{ij}^P - D_{ij}^P P^{-1}\rangle = \lambda_i\lambda_j(1/\lambda_i - 1/\lambda_j)^2\|H_{ij}\|_{\mathrm{Fro}}^2 = \frac{2(\lambda_i - \lambda_j)^2}{\lambda_i\lambda_j}.$$

In addition, we note that $P^{-1/2}D^P_{ij}P^{-1/2} = S_{ij} := q_iq_j^\top + q_jq_i^\top$, and

$$\langle P^{-1/2}D^P_{ij}P^{-1/2}, P^{-1/2}D^P_{ij}P^{-1/2}\rangle = \|S_{ij}\|^2_{\mathrm{Fro}} = 2.$$

Then

$$\begin{aligned}\|v^P_{ij}\|^2_{U,P} &= \frac{1}{2}\langle P^{-1}D^P_{ij} - D^P_{ij}P^{-1}, P^{-1}D^P_{ij} - D^P_{ij}P^{-1}\rangle + \left\langle P^{-1/2}\left(\frac{1}{2}D^P_{ij}\right)P^{-1/2}, P^{-1/2}\left(\frac{1}{2}D^P_{ij}\right)P^{-1/2}\right\rangle \\ &= \frac{(\lambda_i-\lambda_j)^2}{\lambda_i\lambda_j} + \frac{1}{2}.\end{aligned}$$

We compute the inner product $\langle\nabla\tilde{f}(U,P), v^P_{ij}\rangle_{U,P}$ as follows:

$$\begin{aligned}&\langle\nabla\tilde{f}(U,P), v^P_{ij}\rangle_{U,P} \\ &= \frac{1}{2}\langle\tilde{A}, P^{-1}D^P_{ij} - D^P_{ij}P^{-1}\rangle + \left\langle P^{-1/2}\tilde{D}P^{-1/2}, P^{-1/2}\left(\frac{1}{2}D^P_{ij}\right)P^{-1/2}\right\rangle \\ &= \frac{1}{2}\langle\tilde{A}, P^{-1}D^P_{ij} - D^P_{ij}P^{-1}\rangle + \frac{1}{2}\left\langle P^{-1}\tilde{D}P^{-1}, D^P_{ij}\right\rangle \\ &= \frac{\sqrt{\lambda_i\lambda_j}}{2}\left(\langle U^\top(G+G^\top)UP - PU^\top(G+G^\top)U, P^{-1}S_{ij} - S_{ij}P^{-1}\rangle + \frac{1}{2}\left\langle U^\top(G+G^\top)U, S_{ij}\right\rangle\right) \\ &= \frac{\sqrt{\lambda_i\lambda_j}}{2}\left(\left\langle P^{-1}U^\top(G+G^\top)UP, S_{ij}\right\rangle - \left\langle U^\top(G+G^\top)U, S_{ij}\right\rangle\right. \\ &\quad\left. - \left\langle U^\top(G+G^\top)U, S_{ij}\right\rangle + \left\langle PU^\top(G+G^\top)UP^{-1}, S_{ij}\right\rangle + \frac{1}{2}\left\langle U^\top(G+G^\top)U, S_{ij}\right\rangle\right).\end{aligned}$$

Now recognise that $\left\langle U^\top(G+G^\top)U, S_{ij}\right\rangle = 2(Uq_j)^\top(G+G^\top)(Uq_i)$ and $\left\langle PU^\top(G+G^\top)UP^{-1}, S_{ij}\right\rangle = \left\langle PU^\top(G+G^\top)UP^{-1}, S_{ij}\right\rangle = \left(\frac{\lambda_i}{\lambda_j} + \frac{\lambda_j}{\lambda_i}\right)(Uq_j)^\top(G+G^\top)(Uq_i)$. Therefore

$$\langle\nabla\tilde{f}(U,P), v^P_{ij}\rangle_{U,P} = \frac{\sqrt{\lambda_i\lambda_j}}{2}\left(2\left(\frac{\lambda_i}{\lambda_j} + \frac{\lambda_j}{\lambda_i}\right) - 3\right)(Uq_j)^\top(G+G^\top)(Uq_i).$$

The result then follows by recognising that $2\left(\frac{\lambda_i}{\lambda_j} + \frac{\lambda_j}{\lambda_i}\right) - 3 = \frac{2(\lambda_i-\lambda_j)^2+\lambda_i\lambda_j}{\lambda_i\lambda_j} = 2\|v^P_{ij}\|^2_{U,P}$. $\square$

5 Application: trace regression

Recall $\tilde{M} = \mathrm{St}(r,n)\times\mathbb{S}^r_{++}$ and $M = \mathbb{S}^{n,r}_+$. We aim to solve

$$\min_{W\in M}\frac{1}{N}\sum_{\ell\in[N]}(y_p - x_p^\top Wx_p)^2.$$

We will lift this to the total space $\tilde{M}$ as follows:

$$\min_{(U,P)\in\tilde{M}}\frac{1}{N}\sum_{\ell\in[N]}(y_\ell - x_\ell^\top UPU^\top x_\ell)^2.$$

Here,

$$f(W) = \frac{1}{2N}\sum_{p\in[N]}(y_p - x_p^\top Wx_p)^2\tilde{f}(U,P) \qquad = f(UPU^\top) = \frac{1}{2N}\sum_{p\in[N]}(y_p - x_p^\top UPU^\top x_p)^2.$$

5.1 Algorithm (deterministic)

Given equal-sized matrices or vectors a and b, denote $a \odot b$ to be the element-wise product. The procedure is as follows:

- *Initialization.*
  - Denote $X = \begin{bmatrix} x_1 & \cdots & x_N \end{bmatrix} \in \mathbb{R}^{n \times N}$ to be the matrix of features, and $y = (y_1, \ldots, y_N) \in \mathbb{R}^N$ to be the vector of responses.
  - Pick initial point $U_1 = \begin{bmatrix} e_{1,r} & \cdots & e_{r,r} \end{bmatrix} \in \mathrm{St}(r,n)$, $P_1 = I_r \in \mathbb{S}^r_{++}$.
  - Initialize orthogonal quantities $Uq_k = e_{k,r}$, $\lambda_k = 1$ for $k \in [r]$. Let $UQ = \begin{bmatrix} Uq_1 & \cdots & Uq_r \end{bmatrix} \in \mathbb{R}^{n \times r}$ and $\lambda = (\lambda_1, \ldots, \lambda_r) \in \mathbb{R}^r$, $\Lambda = \mathrm{Diag}(\lambda) \in \mathbb{R}^{r \times r}$. (Here we think of $P = Q \Lambda Q^\top$.)
  - Initialize product quantities $w_{p,k} = x_p^\top U q_k$ for each $p \in [N]$ and $k \in [r]$. For each $k \in [r]$, define vectors $w_k = (w_{1,k}, \ldots, w_{N,k}) \in \mathbb{R}^N$. Let $W = \begin{bmatrix} w_1 & \cdots & w_r \end{bmatrix} = X^\top UQ \in \mathbb{R}^{N \times r}$.
  - Initialize trace quantities $\tau_p = x_p^\top U P U^\top x_p$; for $U = U_1, P = P_1$, this is just the sum of squares of the first r entries of x_p. Denote $\tau = (\tau_1, \ldots, \tau_N) \in \mathbb{R}^N$.

- *Outer iterations.* For each $t = 1, 2, \ldots$
  1. Set $U_{t,1} := U_t$, $P_{t,1} := P_t$.
  2. *Inner iterations.* For each $\ell = 1, \ldots, m$:
     (a) For convenience denote $(U_{t,\ell}, P_{t,\ell}) = (U, P)$
     (b) Pick a subspace $\mathcal{S}_k^{U,P}$ for some $k \in [r]$ or $\mathcal{S}_{ij}^{U,P}$ for some $(i,j) \in [r] \times [r]$ such that $i \leq j$.
       - *When the subspace is $\mathcal{S}_k^{U,P}$.*
         i. *Gradient projection.* We wish to project $\nabla \tilde{f}(U,P)$ onto $\mathcal{S}_k^{U,P}$. By Proposition 4.2 the formula is $\lambda_k (I_n - UU^\top)(G + G^\top) U q_k q_k^\top$, where $G = \nabla f(UPU^\top)$. We will denote $\hat{v}_k := \lambda_k (I_n - UU^\top)(G + G^\top) U q_k$, and provide a method to compute this. First notice that

$$G = \frac{1}{N} \sum_{p \in [N]} (y_p - \tau_p) x_p x_p^\top, \quad G + G^\top = 2G.$$

Then

$$(G + G^\top) U q_k = \frac{2}{N} \sum_{p \in [N]} (y_p - \tau_p) x_p w_{p,k} = \frac{2}{N} X((y - \tau) \odot w_k).$$

Since $UU^\top = (UQ)(UQ)^\top$, we have

$$UQ(UQ)^\top (G+G^\top) U q_k = \frac{2}{N} UQ(UQ)^\top X((y-\tau) \odot w_k) = \frac{2}{N} UQ \left(W^\top ((y - \tau) \odot w_k) \right).$$

We thus can compute

$$\hat{v}_k = \frac{2\lambda_k}{N} \left(X((y - \tau) \odot w_k) - UQ \left(W^\top ((y - \tau) \odot w_k) \right) \right)$$

using only matrix-vector products. The projection is then $(\tilde{v}_k q_k^\top, 0)$. We let

$$v_k := \frac{\hat{v}_k}{\|\hat{v}_k\|_2}.$$

ii. *Exponential map update formula.* We now need to update (U, P) in the direction of $(v_k q_k^\top, 0)$. According to Proposition 2.3, the update with step size γ will be of form

$$\widetilde{\operatorname{Exp}}_{U,P}(\gamma(v_k q_k^\top, 0)) = (U_{+,k}(\tau), P),$$

where

$$U_{+,k}(\gamma) = U + ((\cos(\gamma) - 1)Uq_k + \sin(\gamma)v_k)\, q_k^\top .$$

iii. *Line search.* We now find the step size τ that will minimize $\tilde{f}(U_{+,k}(\gamma), P)$. Note that $U_{+,k}(\gamma)$ is periodic over $\gamma \in [0, 2\pi]$, so we just need to do a line search over this interval. Then we have

$$\begin{aligned} x_p^\top U_{+,k}(\gamma) P U_{+,p}^\top(\gamma) x_p = \tau_p &+ 2\lambda_k \Big((\cos(\gamma) - 1)w_{p,k} + \sin(\gamma) v_k^\top x_p\Big) \\ &+ \lambda_k \Big((\cos(\gamma) - 1)w_{p,k} + \sin(\gamma) v_k^\top x_p\Big)^2 . \end{aligned}$$

Our ultimate goal is to do a line search over

$$h_k(\gamma) := \tilde{f}(U_{+,k}(\gamma), P) = \frac{1}{2N} \sum_{p \in [N]} \Big(y_p - x_p U_{+,k}(\gamma) P U_{+,p}^\top(\gamma) x_p\Big)^2 .$$

We need an efficient way to compute $h_k(\gamma)$ given γ. We can do this as follows. Let $z_k = X^\top v_k \in \mathbb{R}^N$ (where each entry is $x_p^\top v_k$). Denote the vector

$$s_k(\tau) := (\cos(\gamma) - 1)w_k + \sin(\gamma) z_k.$$

Then

$$h_k(\gamma) = \frac{1}{2N} \left\| y - \tau - 2\lambda_k s_k(\gamma) - \lambda_k s_k(\gamma) \odot s_k(\gamma) \right\|_2^2 .$$

In fact, the line search need not involve the first term $\|y - \tau\|_2^2$.
We can use any pre-existing line search routine in Python to find γ^* that minimizes $h_k(\gamma)$ over $\gamma \in [0, 2\pi]$.

iv. *Update tracking quantities.* We set $U_{t,\ell+1} := U_{+,k}(\gamma^*)$, $P_{t,\ell+1} = P$. But now we need to update the tracking quantities so we can use them later. This is done as follows:

$$\begin{aligned} w_k &\leftarrow X^\top(\cos(\gamma^*) U q_k + \sin(\gamma^*) v_k) = X^\top(s_k(\gamma^*) + U q_k) \\ \tau &\leftarrow \tau + 2\lambda_k s_k(\gamma^*) + \lambda_k s_k(\gamma^*) \odot s_k(\gamma^*) \\ U q_k &\leftarrow s_k(\gamma^*) + U q_k \end{aligned}$$

The first updates the product quantities $w_{p,k} = x_p^\top U q_k$ (using the new Uq_k) for each $p \in [N]$, which means we just need to update the kth column of W. The second updates trace quantities $\tau_p = x_p^\top U P U^\top x_p$. The third one updates orthogonal quantities Uq_k, i.e., just the kth column of UQ.

– *When the subspace is* $\mathcal{S}_{ij}^{U,P}$.

i. *Gradient projection.* We wish to project $\nabla \tilde{f}(U, P)$ onto $\mathcal{S}_{ij}^{U,P} = \mathrm{Span}(\{v_{ij}^P\})$ where $v_{ij}^P = \left(U(P^{-1}D_{ij}^P - D_{ij}^P P^{-1}), \frac{1}{2}D_{ij}^P\right)$, $D_{ij}^P = (\lambda_i\lambda_j)^{1/2}(q_i q_j^\top + q_j q_i^\top)$. By Proposition 4.2 the formula is $\sqrt{\lambda_i\lambda_j}(Uq_j)^\top(G + G^\top)(Uq_i)v_{ij}^P$, where

$$G = \nabla f(UPU^\top) = \frac{1}{N}\sum_{p\in[N]}(y_p - \tau_p)x_p x_p^\top, \quad G + G^\top = 2G.$$

Thus we have

$$\begin{aligned}\sqrt{\lambda_i\lambda_j}(Uq_j)^\top(G + G^\top)(Uq_i) &= \frac{2\sqrt{\lambda_i\lambda_j}}{N}\sum_{p\in[N]}(y_p - \tau_p)w_{p,i}w_{p,j}\\ &= \frac{2\sqrt{\lambda_i\lambda_j}}{N}(y - \tau)^\top(w_i \odot w_j).\end{aligned}$$

ii. *Exponential map update formula.* From Proposition 2.3, we have

$$\widetilde{\mathrm{Exp}}_{U,P}\left(\gamma v_{ij}^P\right) = (U_{+,ij}(\gamma), P_{+,ij}(\gamma))$$

where

$$\begin{aligned}U_{+,ij}(\gamma) &:= U(I - q_i q_i^\top - q_j q_j^\top) + (\cos(\alpha_{ij}\gamma)Uq_i - \sin(\alpha_{ij}\gamma)Uq_j)q_i^\top\\ &\quad + (\sin(\alpha_{ij}\gamma)Uq_i + \cos(\alpha_{ij}\gamma)Uq_j)q_j^\top\\ \alpha_{ij} &:= \sqrt{\lambda_i\lambda_j}(\lambda_i^{-1} - \lambda_j^{-1})\\ P_{+,ij}(\gamma) &:= P - (\lambda_i q_i q_i^\top + \lambda_j q_j q_j^\top) + \cosh(\gamma/2)(\lambda_i q_i q_i^\top + \lambda_j q_j q_j^\top)\\ &\quad + \sinh(\gamma/2)\sqrt{\lambda_i\lambda_j}(q_i q_j^\top + q_j q_i^\top).\end{aligned}$$

iii. *Line search.* We wish to optimize $h_{ij}(\gamma) := \tilde{f}(U_{+,ij}(\gamma), P_{+,ij}(\gamma))$. We need to first describe an efficient method for computing $h_{ij}(\gamma)$ given γ. Denote

$$\begin{aligned}S_{ij}(\gamma) &:= \begin{bmatrix}\cos(a_{ij}\gamma) & \sin(a_{ij}\gamma)\\ -\sin(a_{ij}\gamma) & \cos(a_{ij}\gamma)\end{bmatrix}\\ T_{ij}(\gamma) &:= \begin{bmatrix}\lambda_i\cosh(\gamma/2) & \sqrt{\lambda_i\lambda_j}\sinh(\gamma/2)\\ \sqrt{\lambda_i\lambda_j}\sinh(\gamma/2) & \lambda_j\cosh(\gamma/2)\end{bmatrix}\\ V_{ij}(\gamma) &:= S_{ij}(\gamma)T_{ij}(\gamma)S_{ij}^\top(\gamma)\\ Q_{ij} &:= \begin{bmatrix}q_i & q_j\end{bmatrix},\end{aligned}$$

where $S_{ij}(\gamma), T_{ij}(\gamma), V_{ij}(\gamma) \in \mathbb{R}^{2\times 2}$, $Q_{ij} \in \mathrm{St}(2, r) \subset \mathbb{R}^{r\times 2}$, so $Q_{ij}^\top Q_{ij} = I_2$. Also, $(I_r - q_i q_i^\top - q_j q_j^\top)Q_{ij} = (P - \lambda_i q_i q_i^\top - \lambda_j q_j q_j^\top)Q_{ij} = 0_{r\times 2}$. With these defined, the updates can be written as

$$\begin{aligned}U_{+,ij}(\gamma) &= U(I_r - q_i q_i^\top - q_j q_j^\top) + UQ_{ij}S_{ij}(\gamma)Q_{ij}^\top\\ P_{+,ij}(\gamma) &= P - \lambda_i q_i q_i^\top - \lambda_j q_j q_j^\top + Q_{ij}T_{ij}(\gamma)Q_{ij}^\top,\end{aligned}$$

and thus

$$U_{+,ij}(\gamma)P_{+,ij}(\gamma)U_{+,ij}^\top(\gamma) = U(P - \lambda_i q_i q_i^\top - \lambda_j q_j q_j^\top)U^\top + UQ_{ij}V_{ij}^\top(\gamma)(UQ_{ij})^\top.$$

Now notice that $x_p^\top U(P - \lambda_i q_i q_i^\top - \lambda_j q_j q_j^\top)U^\top = \tau_p - \lambda_i w_{p,i}^2 - \lambda_j w_{p,j}^2$. Furthermore, notice that $(UQ_{ij})^\top x_p = \begin{bmatrix} w_{p,i} \\ w_{p,j} \end{bmatrix}$; denote this by $w_{p,ij} \in \mathbb{R}^2$. Then we have

$$x_p^\top U_{+,ij}(\gamma)P_{+,ij}(\gamma)U_{+,ij}^\top(\gamma)x_p = \tau_p - \lambda_i w_{p,i}^2 - \lambda_j w_{p,j}^2 + w_{p,ij}^\top V_{ij}(\tau)w_{p,ij}.$$

In vectorized form over $p \in [N]$, we have

$$\begin{aligned} v_{ij}(\gamma) &:= \left\{x_p^\top U_{+,ij}(\gamma)P_{+,ij}(\gamma)U_{+,ij}^\top(\gamma)x_p\right\}_{p\in[N]} \\ &= \tau - \lambda_i w_i \odot w_i - \lambda_j w_j \odot w_j + \left(\left(\begin{bmatrix} w_i & w_j \end{bmatrix} V_{ij}(\gamma)\right) \odot \begin{bmatrix} w_i & w_j \end{bmatrix}\right)^\top \mathbf{1}_2 \in \mathbb{R}^N. \end{aligned}$$

Then $h_{ij}(\gamma) = \frac{1}{2N}\|y - v_{ij}(\gamma)\|_2^2$.
To implement the line search, we need to take into account the projection of $\nabla\tilde{f}(U,P)$. Recall that this is

$$\frac{2\sqrt{\lambda_i\lambda_j}}{N}(y-\tau)^\top(w_i \odot w_j)v_{ij}^P.$$

Since we want to look for descent directions, we move in the negative direction of this projection. Concretely, if $\frac{2\sqrt{\lambda_i\lambda_j}}{N}(y-\tau)^\top(w_i \odot w_j) > 0$, then we minimize $h_{ij}(\gamma)$ over $\gamma \le 0$, and if it's < 0, we minimize over $\gamma \ge 0$.

iv. *Update tracking quantities.* Suppose now that we have found the optimal γ^* to minimize $h_{ij}(\gamma)$. Then (U,P) is updated to $(U_{+,ij}(\gamma^*), P_{+,ij}(\gamma^*))$. We now need to update the tracking quantities. First, we update the trace quantities as

$$\tau \leftarrow \tau - \lambda_i w_i \odot w_i - \lambda_j w_j \odot w_j + \left(\left(\begin{bmatrix} w_i & w_j \end{bmatrix} V_{ij}(\gamma^*)\right) \odot \begin{bmatrix} w_i & w_j \end{bmatrix}\right)^\top \mathbf{1}_2$$

We now need to update the orthogonal and product quantites. Denote $T_{ij} = T_{ij}(\gamma^*)$. Since

$$P_{+,ij}(\gamma^*) = P - \lambda_i q_i q_i^\top - \lambda_j q_j q_j^\top + Q_{ij}T_{ij}Q_{ij}^\top,$$

any eigenvector of P that is orthogonal to both q_i and q_j is remains an eigenvector of $P_{+,ij}(\gamma^*)$, with the same eigenvalue. Therefore the new eigenvectors/values of $P_{+,ij}(\gamma^*)$ can be obtained by diagonalizing T_{ij}. Specifically, if we have

$$T_{ij} = \begin{bmatrix} v_{i,+} & v_{j,+} \end{bmatrix} \begin{bmatrix} \lambda_{i,+} & 0 \\ 0 & \lambda_{j,+} \end{bmatrix} \begin{bmatrix} v_{i,+}^\top \\ v_{j,+}^\top \end{bmatrix},$$

then $(Q_{ij}v_{i,+}, \lambda_{i,+})$ and $(Q_{ij}v_{j,+}, \lambda_{j,+})$ are the new eigenpairs, replacing (q_i, λ_i) and (q_j, λ_j) respectively. With these formulas, we update

$$\begin{aligned} w_i &\leftarrow \begin{bmatrix} w_i & w_j \end{bmatrix} v_{+,i} \\ w_j &\leftarrow \begin{bmatrix} w_i & w_j \end{bmatrix} v_{+,j} \\ Uq_i &\leftarrow \begin{bmatrix} Uq_i & Uq_j \end{bmatrix} v_{+,i} \\ Uq_j &\leftarrow \begin{bmatrix} Uq_i & Uq_j \end{bmatrix} v_{+,j} \end{aligned}$$

There are exact formulas for $v_{i,+}, v_{j,+}, \lambda_{i,+}, \lambda_{j,+}$, but we can also just do it numerically since $T_{ij} \in \mathbb{R}^{2\times 2}$.

3. Set $U_{t+1} := U_{t,m+1}$, $P_{t+1} := P_{t,m+1}$.

References

[1] S. Bonnabel and R. Sepulchre. Riemannian metric and geometric mean for positive semidefinite matrices of fixed rank. *SIAM Journal on Matrix Analysis and Applications*, 31(3):1055–1070, 2010.

[2] N. Boumal. *An introduction to optimization on smooth manifolds.* Cambridge University Press, 2023.

[3] D. H. Gutman and N. Ho-Nguyen. Coordinate descent without coordinates: Tangent subspace descent on riemannian manifolds. *Mathematics of Operations Research*, 2022.

[4] D. Nguyen. Operator-valued formulas for riemannian gradient and hessian and families of tractable metrics in riemannian optimization. *Journal of Optimization Theory and Applications*, pages 1–30, 2023.