Fun with Flags: Robust Principal Directions via Flag Manifolds

Nathan Mankovich
University of Valencia

Gustau Camps-Valls
University of Valencia

Tolga Birdal
Imperial College London

A. Theoretical Justifications & Discussions

On the unifying aspects of our framework. In our framework, the link between RPCA & Dual-PCA, established also in the discussed earlier works, emerges as a by-product of our unifying formulation. To elucidate, our flag-based framework allows for: (i) extending DPCP to manifold-valued data (f$\mathcal{T}$DPCP), (ii) interpolating between L_1/L_2–DPCP via the use of non-trivial flag types, and (iii) an efficient algorithms for computing flag–(tangent) DPCP for any flag type. To the best of our knowledge, Alg. 1 (main paper) is the only method for finding non-trivial flags of robust directions and when used for both fRPCA & fWPCA.

A.1. Proof of Prop. 3

Let us recall the proposition before delving into the proof.

Proposition 1 (Stiefel optimization of (weighted) fPCA**).** *Suppose we have weights $\{w_{ij}\}_{i=1,j=1}^{i=k,j=p}$ for a dataset $\{\mathbf{x}_j\}_{j=1}^p \subset \mathbb{R}^n$ along with a flag type $(n_1, n_2, \dots, n_k; n)$. We store the weights in the diagonal weight matrices $\{\mathbf{W}_i\}_{i=1}^k$ with diagonals $(\mathbf{W}_i)_{jj} = w_{ij}$. If*

$$\mathbf{U}^* = \underset{\mathbf{U} \in St(n_k,n)}{\arg\max} \sum_{i=1}^k \mathrm{tr}\left(\mathbf{U}^T \mathbf{X} \mathbf{W}_i \mathbf{X}^T \mathbf{U} \mathbf{I}_i\right) \tag{1}$$

where $\mathbf{I}_i$ is determined as a function of the flag signature. For example, for $\mathcal{FL}(n+1)$:

$$(\mathbf{I}_i)_{l,s} = \begin{cases} 1, & l = s \in \{n_{i-1}+1, n_{i-1}+2, \dots, n_i\} \\ 0, & \text{otherwise} \end{cases}$$

Then $[\![\mathbf{U}^*]\!] = [\![\mathbf{U}]\!]^*$ is the weighted fPCA of the data with the given weights (e.g., solves **??***) as long as we restrict ourselves to a region on $\mathcal{FL}(n+1)$ and $St(n_k, n)$ where weighted fPCA is convex.*

Proof. First we will show that the flag and Stiefel objective functions are equivalent. Take

$$[\![\mathbf{U}]\!] \in \mathcal{FL}(n+1) = \mathcal{FL}(n_1, n_2, \dots, n_k; n). \tag{2}$$

We decompose $\mathbf{U} = [\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_k]$ where $\mathbf{U}_i \in \mathbb{R}^{n \times m_i}$ and $\sum_{l=1}^i m_l = n_i$. Using $\mathbf{I}_i$ (defined above) we have $\mathbf{U}\mathbf{I}_i\mathbf{U}^T = \mathbf{U}_i$.

Recall the objective function for both fRPCA and fDPCP is

$$\mathbb{E}_j\left[\sum_{i=1}^k w_{ij} \|\pi_{\mathbf{U}_i}(\mathbf{x}_j)\|_2^2\right] = \sum_{j=1}^p \sum_{i=1}^k w_{ij} \|\pi_{\mathbf{U}_i}(\mathbf{x}_j)\|_2^2, \tag{3}$$

$$= \sum_{j=1}^p \sum_{i=1}^k w_{ij} \|\mathbf{U}_i \mathbf{U}_i^T \mathbf{x}_j\|_2^2 \tag{4}$$

Using the definition of norms and $\mathbf{U}_i^T\mathbf{U}_i = \mathbf{I}$, Eq. (3) is equivalent to

$$\sum_{j=1}^p \sum_{i=1}^k w_{ij} \mathrm{tr}\left(\mathbf{x}_j^T \mathbf{U}_i \mathbf{U}_i^T \mathbf{x}_j\right) \tag{5}$$

Now, using properties of trace, matrix multiplication, and our handy $\{\mathbf{I}_i\}_{i=1}^k$ we reach our desired result

$$\sum_{j=1}^p \sum_{i=1}^k w_{ij} \mathrm{tr}\left(\mathbf{U}_i^T \mathbf{x}_j \mathbf{x}_j^T \mathbf{U}_i\right), \tag{6}$$

$$= \sum_{i=1}^k \mathrm{tr}\left(\mathbf{U}_i^T \left(\sum_{j=1}^p w_{ij} \mathbf{x}_j \mathbf{x}_j^T\right) \mathbf{U}_i\right), \tag{7}$$

$$= \sum_{i=1}^k \mathrm{tr}\left(\mathbf{U}_i^T \left(\mathbf{X}\mathbf{W}_i\mathbf{X}^T\right) \mathbf{U}_i\right), \tag{8}$$

$$= \sum_{i=1}^k \mathrm{tr}\left(\mathbf{U}_i \mathbf{U}_i^T \mathbf{X}\mathbf{W}_i\mathbf{X}^T\right), \tag{9}$$

$$= \sum_{i=1}^k \mathrm{tr}\left(\mathbf{U}\mathbf{I}_i \mathbf{U}^T \mathbf{X}\mathbf{W}_i\mathbf{X}^T\right), \tag{10}$$

$$= \sum_{i=1}^k \mathrm{tr}\left(\mathbf{U}^T \mathbf{X}\mathbf{W}_i\mathbf{X}^T \mathbf{U}\mathbf{I}_i\right). \tag{11}$$

So we have shown that the flag and Stiefel objective functions are equivalent.

Finally, we show $[\![\mathbf{U}^*]\!] = [\![\mathbf{U}]\!]^*$. Notice that the objective function for weighted flag PCA is invariant to different flag manifold representatives. First, let f denote the objective function in Eq. (11). Suppose $\mathbf{U}^*$ solves

$\arg\max_{\mathbf{Y}\in St(n_k,n)} f(\mathbf{Y})$. Then take some other representative for $[\![\mathbf{U}^*]\!]$, namely $\mathbf{U}^*\mathbf{M}$ where

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_2 & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{M}_k \end{bmatrix} \text{ and } \mathbf{M}_1 \in O(m_i). \quad (12)$$

Then $f(\mathbf{U}^*\mathbf{M}) = f(\mathbf{U}^*)$ because

$$f(\mathbf{U}^*\mathbf{M}) = \sum_{i=1}^{k} \mathrm{tr}((\mathbf{U}_i^*\mathbf{M})^T \mathbf{X}\mathbf{W}_i\mathbf{X}^T(\mathbf{U}_i^*\mathbf{M})), \quad (13)$$

$$= \sum_{i=1}^{k} \mathrm{tr}(\mathbf{U}_i^*\mathbf{M}\mathbf{M}^T{\mathbf{U}_i^*}^T\mathbf{X}\mathbf{W}_i\mathbf{X}^T), \quad (14)$$

$$= \sum_{i=1}^{k} \mathrm{tr}(\mathbf{U}_i^*{\mathbf{U}_i^*}^T\mathbf{X}\mathbf{W}_i\mathbf{X}^T), \quad (15)$$

$$= \sum_{i=1}^{k} \mathrm{tr}({\mathbf{U}_i^*}^T\mathbf{X}\mathbf{W}_i\mathbf{X}^T\mathbf{U}_i^*), \quad (16)$$

$$= f(\mathbf{U}^*). \quad (17)$$

So $f(\cdot)$ has the same value for any representative for $[\![\mathbf{U}^*]\!]$. Since $f(\mathbf{U}^*) \geq f(\mathbf{Y})$ for all $\mathbf{Y} \in St(n_k, n)$, then

$$f(\mathbf{U}^*\mathbf{M}) = f(\mathbf{U}^*) \geq f(\mathbf{Y}) = f(\mathbf{U}^*\boldsymbol{O}) \quad (18)$$

for all $[\![\mathbf{Y}]\!] \in \mathcal{FL}(n+1)$ where $\boldsymbol{O}$ is of the same block structure as $\mathbf{M}$.

Recall $[\![\mathbf{U}]\!]^* \in \mathcal{FL}(n+1)$ maximizes f, so $f(\mathbf{U}) \geq f(\mathbf{Y})$ for all $[\![\mathbf{Y}]\!] \in \mathcal{FL}(n+1)$ and since $f(\cdot)$ has the same value for any representative of $[\![\mathbf{Y}]\!]$, we have $f(\mathbf{U}) \geq f(\mathbf{Y})$ for all $\mathbf{Y} \in St(n_k, n)$.

Recall, that $f(\mathbf{U}^*) \geq f(\mathbf{Y})$ for all $\mathbf{Y} \in St(n_k, n)$. So $f(\mathbf{U}^*) = f(\mathbf{U})$. Since f has a unique maximizer over $\mathcal{FL}(n+1)$, we have $[\![\mathbf{U}^*]\!] = [\![\mathbf{U}]\!]^* = \arg\max_{[\![\mathbf{Y}]\!]\in\mathcal{FL}(n+1)} f(\mathbf{Y})$. □

A.2. Proof of Prop. 4

Let us recall the proposition before delving into the proof.

Proposition 2 (Stiefel optimization for flagified Robust (Dual-)PCAs)**.** *We can formulate fRPCA, fWPCA, fDPCP, and fWDPCP as optimization problems over the Stiefel manifold using $[\![\mathbf{U}]\!]^* = [\![\mathbf{U}^*]\!]$ and the following:*

$$\mathbf{U}^\star = \quad (19)$$

$$\begin{cases} \underset{\mathbf{U}\in St(n,n_k)}{\arg\max} \sum_{i=1}^k \mathrm{tr}\left(\mathbf{U}^T\mathbf{P}_i^+\mathbf{U}\mathbf{I}_i\right), & \text{(fRPCA)} \\ \underset{\mathbf{U}\in St(n,n_k)}{\arg\min} \sum_{i=1}^k \mathrm{tr}\left(\mathbf{P}_i^- - \mathbf{U}^T\mathbf{P}_i^-\mathbf{U}\mathbf{I}_i\right), & \text{(fWPCA)} \end{cases}$$

$$\mathbf{U}^\star = \quad (20)$$

$$\begin{cases} \underset{\mathbf{U}\in St(n,n_k)}{\arg\min} \sum_{i=1}^k \mathrm{tr}\left(\mathbf{U}^T\mathbf{P}_i^+\mathbf{U}\mathbf{I}_i\right), & \text{(fDPCP)} \\ \underset{\mathbf{U}\in St(n,n_k)}{\arg\max} \sum_{i=1}^k \mathrm{tr}\left(\mathbf{P}_i^- - \mathbf{U}^T\mathbf{P}_i^-\mathbf{U}\mathbf{I}_i\right) & \text{(fWDPCP)} \end{cases}$$

where $\mathbf{P}^- = \mathbf{X}\mathbf{W}_i^-([\![\mathbf{U}]\!])\mathbf{X}^T$, $\mathbf{P}^+ = \mathbf{X}\mathbf{W}_i^+([\![\mathbf{U}]\!])\mathbf{X}^T$ and $\mathbf{W}_i^-([\![\mathbf{U}]\!])$, $\mathbf{W}_i^+([\![\mathbf{U}]\!])$ are defined in **??** *as long as we restrict ourselves to a region on $\mathcal{FL}(n+1)$ and $St(n_k, n)$ where flag robust and dual PCAs are convex.*

Proof. First, we write the objective functions for fRPCA and fDPCP over $St(n_k, n)$ using **??** to define each $\mathbf{W}_i^+$ as

$$f^+(\mathbf{U}) = \mathbb{E}\left[\sum_{i=1}^k \|\pi_{\mathbf{U}_i}(\mathbf{x}_j)\|_2\right], \quad (21)$$

$$= \sum_{j=1}^p\sum_{i=1}^k \|\pi_{\mathbf{U}_i}(\mathbf{x}_j)\|_2, \quad (22)$$

$$= \sum_{j=1}^p\sum_{i=1}^k \sqrt{\mathrm{tr}\left(\mathbf{x}_j^T\mathbf{U}_i\mathbf{U}_i^T\mathbf{x}_j\right)}, \quad (23)$$

$$= \sum_{j=1}^p\sum_{i=1}^k \sqrt{\mathrm{tr}\left(\mathbf{U}_i^T\mathbf{x}_j\mathbf{x}_j^T\mathbf{U}_i\right)} \quad (24)$$

$$= \sum_{j=1}^p\sum_{i=1}^k \sqrt{\mathrm{tr}\left(\mathbf{U}^T\mathbf{x}_j\mathbf{x}_j^T\mathbf{U}\mathbf{I}_i\right)}, \quad (25)$$

$$= \sum_{j=1}^p\sum_{i=1}^k \frac{\mathrm{tr}\left(\mathbf{U}^T\mathbf{x}_j\mathbf{x}_j^T\mathbf{U}\mathbf{I}_i\right)}{\sqrt{\mathrm{tr}\left(\mathbf{U}^T\mathbf{x}_j\mathbf{x}_j^T\mathbf{U}\mathbf{I}_i\right)}}, \quad (26)$$

$$= \sum_{i=1}^k \mathrm{tr}\left(\mathbf{U}^T\sum_{j=1}^p \frac{\mathbf{x}_j\mathbf{x}_j^T}{\|\mathbf{U}\mathbf{I}_i\mathbf{U}^T\mathbf{x}_j\|_2}\mathbf{U}\mathbf{I}_i\right), \quad (27)$$

$$= \sum_{i=1}^k \mathrm{tr}\left(\mathbf{U}^T\mathbf{X}\mathbf{W}_i^+\mathbf{X}^T\mathbf{U}\mathbf{I}_i\right), \quad (28)$$

$$= \sum_{i=1}^k \mathrm{tr}\left(\mathbf{U}^T\mathbf{P}_i^+\mathbf{U}\mathbf{I}_i\right). \quad (29)$$

Now we write the objective functions for fWPCA and fWDPCP over $St(n_k, n)$ using **??** to define each $\mathbf{W}_i^-$ as

$$f^-(\mathbf{U}) = \mathbb{E}\left[\sum_{i=1}^{k} \|\mathbf{x}_j - \pi_{\mathbf{U}_i}(\mathbf{x}_j)\|_2\right], \tag{30}$$

$$= \sum_{j=1}^{p}\sum_{i=1}^{k} \|\mathbf{x}_j - \pi_{\mathbf{U}_i}(\mathbf{x}_j)\|_2, \tag{31}$$

$$= \sum_{j=1}^{p}\sum_{i=1}^{k} \sqrt{\mathrm{tr}\left(\mathbf{x}_j^T\mathbf{x}_j - \mathbf{x}_j^T\mathbf{U}_i\mathbf{U}_i^T\mathbf{x}_j\right)}, \tag{32}$$

$$= \sum_{j=1}^{p}\sum_{i=1}^{k} \sqrt{\mathbf{x}_j^T\mathbf{x}_j - \mathrm{tr}\left(\mathbf{U}_i^T\mathbf{x}_j\mathbf{x}_j^T\mathbf{U}_i\right)}, \tag{33}$$

$$= \sum_{j=1}^{p}\sum_{i=1}^{k} \sqrt{\mathbf{x}_j^T\mathbf{x}_j - \mathrm{tr}\left(\mathbf{U}^T\mathbf{x}_j\mathbf{x}_j^T\mathbf{U}\mathbf{I}_i\right)}, \tag{34}$$

$$= \sum_{j=1}^{p}\sum_{i=1}^{k} \frac{\mathbf{x}_j^T\mathbf{x}_j - \mathrm{tr}\left(\mathbf{U}^T\mathbf{x}_j\mathbf{x}_j^T\mathbf{U}\mathbf{I}_i\right)}{\sqrt{\mathbf{x}_j^T\mathbf{x}_j - \mathrm{tr}\left(\mathbf{U}^T\mathbf{x}_j\mathbf{x}_j^T\mathbf{U}\mathbf{I}_i\right)}}, \tag{35}$$

$$= \sum_{j=1}^{p} \frac{\mathbf{x}_j\mathbf{x}_j^T}{\|\mathbf{x}_j - \mathbf{U}\mathbf{I}_i\mathbf{U}^T\mathbf{x}_j\|_2} \tag{36}$$

$$- \sum_{i=1}^{k} \mathrm{tr}\left(\mathbf{U}^T \sum_{j=1}^{p} \frac{\mathbf{x}_j\mathbf{x}_j^T}{\|\mathbf{x}_j - \mathbf{U}\mathbf{I}_i\mathbf{U}^T\mathbf{x}_j\|_2}\mathbf{U}\mathbf{I}_i\right) \tag{37}$$

$$= \sum_{i=1}^{k} \mathrm{tr}\left(\mathbf{X}\mathbf{W}_i^-\mathbf{X}^T - \mathbf{U}^T\mathbf{X}\mathbf{W}_i\mathbf{X}^T\mathbf{U}\mathbf{I}_i\right), \tag{38}$$

$$= \sum_{i=1}^{k} \mathrm{tr}\left(\mathbf{P}^- - \mathbf{U}^T\mathbf{P}^-\mathbf{U}\mathbf{I}_i\right). \tag{39}$$

Now, we can write the Lagrangians for these problems with the symmetric matrix of Lagrange multipliers $\mathbf{\Lambda}$ as

$$\mathcal{L}^+(\mathbf{U}) = f^+(\mathbf{U}) + \mathrm{tr}(\mathbf{\Lambda}^+(\mathbf{I} - \mathbf{U}^T\mathbf{U})),$$
$$\mathcal{L}^-(\mathbf{U}) = f^-(\mathbf{U}) + \mathrm{tr}(\mathbf{\Lambda}^-(\mathbf{I} - \mathbf{U}^T\mathbf{U})).$$

Then, we collect our gradients in the following equations

$$\nabla_{\mathbf{U}}\mathcal{L}^+ = \sum_{j=1}^{p}\sum_{i=1}^{k} \frac{\mathbf{x}_j\mathbf{x}_j^T\mathbf{U}\mathbf{I}_i}{\|\mathbf{U}\mathbf{I}_i\mathbf{U}^T\mathbf{x}_j\|_2} - 2\mathbf{U}\mathbf{\Lambda}^+ \tag{40}$$

$$\nabla_{\mathbf{U}}\mathcal{L}^- = -\sum_{j=1}^{p}\sum_{i=1}^{k} \frac{\mathbf{x}_j\mathbf{x}_j^T\mathbf{U}\mathbf{I}_i}{\|\mathbf{x}_j\mathbf{x}_j^T - \mathbf{U}\mathbf{I}_i\mathbf{U}^T\mathbf{x}_j\|_2} - 2\mathbf{U}\mathbf{\Lambda}^- \tag{41}$$

$$\nabla_{\mathbf{\Lambda}^+}\mathcal{L}^+ = \nabla_{\mathbf{\Lambda}^-}\mathcal{L}^- = \mathbf{I} - \mathbf{U}^T\mathbf{U}. \tag{42}$$

(43)

Then setting $\nabla_{\mathbf{U}}\mathcal{L}_1 = \mathbf{0}$, $\nabla_{\mathbf{\Lambda}_1}\mathcal{L}_1 = \mathbf{0}$, left multiplying by $\mathbf{U}^T$, and playing with properties of trace results in

$$\sum_{i=1}^{k} \mathrm{tr}\left(\mathbf{U}^T\mathbf{X}\mathbf{W}_i\mathbf{X}^T\mathbf{U}\mathbf{I}_i\right) = 2\mathrm{tr}(\mathbf{\Lambda}^+), \tag{44}$$

$$\sum_{i=1}^{k} \mathrm{tr}\left(\mathbf{U}^T\mathbf{X}\mathbf{W}_i\mathbf{X}^T\mathbf{U}\mathbf{I}_i\right) = -2\mathrm{tr}(\mathbf{\Lambda}^-). \tag{45}$$

Then we have the following cases: we choose

- (fRPCA) $\mathbf{U}^*$ to maximize $\mathrm{tr}(\mathbf{\Lambda}^+)$ so that we maximize f^+,
- (fDPCP) $\mathbf{U}^*$ to minimize $\mathrm{tr}(\mathbf{\Lambda}^+)$ so that we minimize f^+,
- (fWPCA) $\mathbf{U}^*$ to minimize $-\mathrm{tr}(\mathbf{\Lambda}^-)$ so that we minimize f^-,
- (fWDPCP) $\mathbf{U}^*$ to maximize $-\mathrm{tr}(\mathbf{\Lambda}^-)$ so that we maximize f^-.

$$\sum_{j=1}^{p}\sum_{i=1}^{k} \frac{\mathbf{x}_j\mathbf{x}_j^T\mathbf{U}\mathbf{I}_i}{\|\mathbf{U}_i^T\mathbf{x}_j\|_2} = \mathbf{\Lambda}_1\mathbf{U}, \tag{46}$$

$$\sum_{j=1}^{p}\sum_{i=1}^{k} \frac{\mathbf{U}^T\mathbf{x}_j\mathbf{x}_j^T\mathbf{U}\mathbf{I}_i}{\|\mathbf{U}_i^T\mathbf{x}_j\|_2} = \mathbf{\Lambda}_1, \tag{47}$$

$$\sum_{j=1}^{p}\sum_{i=1}^{k} (\mathbf{W}_i)_{jj}\mathbf{U}^T\mathbf{x}_j\mathbf{x}_j^T\mathbf{U}\mathbf{I}_i = \mathbf{\Lambda}_1, \tag{48}$$

$$\mathrm{tr}\left(\sum_{i=1}^{k} \mathbf{U}^T\left(\sum_{j=1}^{p}(\mathbf{W}_i)_{jj}\mathbf{x}_j\mathbf{x}_j^T\right)\mathbf{U}\mathbf{I}_i\right) = \mathrm{tr}(\mathbf{\Lambda}_1), \tag{49}$$

$$\sum_{i=1}^{k} \mathrm{tr}\left(\mathbf{U}^T\left(\sum_{j=1}^{p}(\mathbf{W}_i)_{jj}\mathbf{x}_j\mathbf{x}_j^T\right)\mathbf{U}\mathbf{I}_i\right) = \mathrm{tr}(\mathbf{\Lambda}_1), \tag{50}$$

$$h_{[\![U]\!]}(\mathbf{U}) = \mathrm{tr}(\mathbf{\Lambda}_1). \tag{51}$$

Similarly, setting $\nabla_{\mathbf{U}}\mathcal{L}_2 = \mathbf{0}$, $\nabla_{\mathbf{\Lambda}_2}\mathcal{L}_2 = \mathbf{0}$ and leveraging **??** to define $\{\mathbf{W}_i\}_i$ results in

$$\sum_{j=1}^{p}\sum_{i=1}^{k} \frac{\mathbf{x}_j\mathbf{x}_j^T\mathbf{U}\mathbf{I}_i}{\|\mathbf{x}_j - \mathbf{U}_i\mathbf{U}_i^T\mathbf{x}_j\|_2} = \mathbf{\Lambda}_2\mathbf{U}, \tag{52}$$

$$-\sum_{j=1}^{p}\sum_{i=1}^{k} \frac{\mathbf{U}^T\mathbf{x}_j\mathbf{x}_j^T\mathbf{U}\mathbf{I}_i}{\|\mathbf{x}_j - \mathbf{U}_i\mathbf{U}_i^T\mathbf{x}_j\|_2} = \mathbf{\Lambda}_2, \tag{53}$$

$$-\sum_{j=1}^{p}\sum_{i=1}^{k} (\mathbf{W}_i)_{jj}\mathbf{U}^T\mathbf{x}_j\mathbf{x}_j^T\mathbf{U}\mathbf{I}_i = \mathbf{\Lambda}_2, \tag{54}$$

$$-\sum_{i=1}^{k} \mathrm{tr}(\mathbf{U}^T\left(\mathbf{X}\mathbf{W}_i\mathbf{X}^T\right)\mathbf{U}\mathbf{I}_i) = \mathrm{tr}(\mathbf{\Lambda}_2), \tag{55}$$

$$-h_{[\![U]\!]}(\mathbf{U}) = \mathrm{tr}(\mathbf{\Lambda}_2). \tag{56}$$

Finally, using a similar argument to that for the proof of the Stiefel optimization of fPCA leveraging assumed convexity, we have that $[\![\mathbf{U}^*]\!] = [\![\mathbf{U}]\!]^*$. □

A.3. Proof of Prop. 5

We now prove the convergence of our algorithm. Let us recall the proposition from the main paper before delving into the proof.

Proposition 3 (Convergence of **??** for fDPCP)**.** **??** *for fDPCP converges as long as* $\|\mathbf{U}\mathbf{I}_i\mathbf{U}^T\mathbf{x}_j\|_2 \geq \epsilon \; \forall i, j$ *and we restrict ourselves to a region on* $\mathcal{FL}(n+1)$ *and* $St(n_k, n)$ *where fDPCPis convex.*

Proof. This proof follows closely to what was done in [1]. First let $f^+ : \mathcal{FL}(n+1) \times \mathcal{FL}(n+1) \to \mathbb{R}$ denote the fDPCP objective function and $T : \mathcal{FL}(n+1) \to \mathcal{FL}(n+1)$ denote an iteration of **??**. Then, assuming that $\|\mathbf{U}\mathbf{I}_i\mathbf{U}^T\mathbf{x}_j\|_2 \geq \epsilon$ for $i = 1, 2, \dots, k$ and $j = 1, 2, \dots, p$, we define the function $h : \mathcal{FL}(d+1) \times \mathcal{FL}(d+1) \to \mathbb{R}$ as

$$h([\![\mathbf{Z}]\!], [\![\mathbf{U}]\!]) = \sum_{i=1}^{p} \mathrm{tr}(\mathbf{Z}^T\mathbf{X}\mathbf{W}_i^+([\![\mathbf{U}]\!])\mathbf{X}^T\mathbf{Z}\mathbf{I}_i), \tag{57}$$

using the definition in **??** for $\mathbf{W}_i^+([\![\mathbf{U}]\!])$.

Some algebra reduces $h([\![\mathbf{Z}]\!], [\![\mathbf{U}]\!])$ to

$$h([\![\mathbf{Z}]\!], [\![\mathbf{U}]\!]) = \sum_{i=1}^{p}\sum_{j=1}^{k} \frac{\|\mathbf{Z}\mathbf{I}_i\mathbf{Z}^T\mathbf{x}_j\|_2^2}{\|\mathbf{U}\mathbf{I}_i\mathbf{U}^T\mathbf{x}_j\|_2}. \tag{58}$$

From Eq. (57), we see that $h([\![\mathbf{Z}]\!], [\![\mathbf{U}]\!])$ is the weighted flag PCA objective function of $\{\mathbf{x}_j\}_{j=1}^p$ with weights on the diagonals of $\mathbf{W}_i([\![\mathbf{U}]\!])$. The weighted flagified orthogonal PCA (f⊥PCA) optimization problem with weights in the diagonals $\mathbf{W}_i^+([\![\mathbf{U}]\!])$ can be solved using a similar algorithm to Alg. 1 by just minimizing instead of maximizing (see Alg. 2). Thus minimizing $h([\![\mathbf{Z}]\!], [\![\mathbf{U}]\!])$ over $[\![\mathbf{Z}]\!]$ is an iteration of Alg. 2 for fDPCP which means

$$T([\![\mathbf{U}]\!]) = \operatorname*{arg\,min}_{[\![\mathbf{Z}]\!] \in \mathcal{FL}(d+1)} h([\![\mathbf{Z}]\!], [\![\mathbf{U}]\!]). \tag{59}$$

Using this, we have

$$h(T([\![\mathbf{U}]\!]), [\![\mathbf{U}]\!]) \leq h([\![\mathbf{U}]\!], [\![\mathbf{U}]\!]). \tag{60}$$

By the definition of h

$$h([\![\mathbf{U}]\!], [\![\mathbf{U}]\!]) = \sum_{i=1}^{p}\sum_{j=1}^{k} \frac{\|\mathbf{U}\mathbf{I}_i\mathbf{U}^T\mathbf{x}_j\|_2^2}{\|\mathbf{U}\mathbf{I}_i\mathbf{U}^T\mathbf{x}_j\|_2}, \tag{61}$$

$$= \sum_{i=1}^{p}\sum_{j=1}^{k} \|\mathbf{U}\mathbf{I}_i\mathbf{U}^T\mathbf{x}_j\|_2, \tag{62}$$

$$= f([\![\mathbf{U}]\!]). \tag{63}$$

This means, we have

$$h(T([\![\mathbf{U}]\!]), [\![\mathbf{U}]\!]) \leq f([\![\mathbf{U}]\!]). \tag{64}$$

Now we use the identity from algebra: $\frac{a^2}{b} \geq 2a - b$ for any $a, b \in \mathbb{R}$ and $b > 0$. Let

$$a = \|\mathbf{Z}\mathbf{I}_i\mathbf{Z}^T\mathbf{x}_j\|_2 \text{ and } b = \|\mathbf{U}\mathbf{I}_i\mathbf{U}^T\mathbf{x}_j\|_2. \tag{65}$$

Then

$$h([\![\mathbf{Z}]\!], [\![\mathbf{U}]\!]) \geq 2\sum_{j=1}^{p}\sum_{i=1}^{k} \|\mathbf{Z}\mathbf{I}_i\mathbf{Z}^T\mathbf{x}_j\|_2 \tag{66}$$

$$- \sum_{j=1}^{p}\sum_{i=1}^{k} \|\mathbf{U}\mathbf{I}_i\mathbf{U}^T\mathbf{x}_j\|_2, \tag{67}$$

$$= 2f([\![\mathbf{Z}]\!]) - f([\![\mathbf{U}]\!]). \tag{68}$$

Now, take $[\![\mathbf{Z}]\!] = T([\![\mathbf{U}]\!])$. This gives us

$$h(T([\![\mathbf{U}]\!]), [\![\mathbf{U}]\!]) \geq 2f(T([\![\mathbf{U}]\!])) - f([\![\mathbf{U}]\!]). \tag{69}$$

Then, combining Eq. 69 with Eq. 64, we have

$$2f(T([\![\mathbf{U}]\!])) - f([\![\mathbf{U}]\!]) \leq f([\![\mathbf{U}]\!]), \tag{70}$$

$$f(T([\![\mathbf{U}]\!])) \leq f([\![\mathbf{U}]\!]). \tag{71}$$

Finally, notice that the real sequence with terms $f^+(T([\![\mathbf{U}^{(m-1)}]\!])) = f^+([\![\mathbf{U}^{(m)}]\!]) \in \mathbb{R}$ is bounded below by 0 and is decreasing. So it converges as $m \to \infty$.

□

B. Further Notes on Flagified PCA

We now generalize PCA and its variants using flags by grouping eigenvectors using the flag type. The PCA optimization problem is naturally an optimization problem on the Stiefel manifold, $St(k, n) := \{\mathbf{U} \in \mathbb{R}^{k \times n} : \mathbf{U}^T\mathbf{U} = \mathbf{I}\}$. Suppose $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k] \in St(k, n)$ are the $k < n$ principal components of a data matrix $\mathbf{X}$. These are naturally ordered according to their decreasing associated objective function values[1]. This results in the nested subspace structure

$$[\![\mathbf{U}]\!] = [\mathbf{u}_1] \subset [\mathbf{u}_1, \mathbf{u}_2] \subset \cdots \subset [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k] \subset \mathbb{R}^n. \tag{72}$$

So one can think of $[\![\mathbf{U}]\!] \in \mathcal{FL}(1, 2, \dots, k; n)$, and consequently, reformulate PCA as an optimization problem over $\mathcal{FL}(1, 2, \dots, k; n)$. Thinking of $\mathbf{U}$ as $[\![\mathbf{U}]\!]$ emphasizes the nested subspace structure of the principal components according to their associated objective function values.

What if we have multiple principal components with the same objective function value? In other words, suppose we have at least one eigenvalue of $\mathbf{X}\mathbf{X}^T$ with a geometric multiplicity greater than 1? For example, assume

[1] The objective function values are also referred to as explained variances, eigenvalues and squared singular values

our dataset has a large variance on some 2-plane, and all other directions orthogonal to that plane have smaller, unequal variance. Then, the first two principal components, $\mathbf{u}_1$ and $\mathbf{u}_2$, will have the same objective function value in **??**. Additionally, any rotation of the two vectors within the plane $\text{span}(\mathbf{u}_1, \mathbf{u}_2)$ will still produce the same objective function values. So, **??** is no longer a convex optimization problem over $St(k, n)$ because the first two principal components are not unique. However, if we remove $[\mathbf{u}_1] \subset [\mathbf{u}_1, \mathbf{u}_2]$ from the nested subspace structure and consider $[\![\mathbf{U}]\!] \in \mathcal{FL}(2, 3, \ldots, k; n)$ as

$$[\![\mathbf{U}]\!] = [\mathbf{u}_1, \mathbf{u}_2] \subset [\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3] \subset \cdots \subset [\mathbf{u}_1, \ldots, \mathbf{u}_k] \subset \mathbb{R}^n. \tag{73}$$

Then we have a unique solution to **??** over $\mathcal{FL}(2, 3, \ldots, k; n)$ in place of $St(k, n)$. In practice, it is unlikely that we will have two eigenvectors with the same eigenvalue. However, we can consider two eigenvalues the same as long as $|\lambda_i - \lambda_j| < \epsilon$ for some $\epsilon > 0$.

Motivated by this example, we state a generalization of PCA, which optimizes over flags of a given type.

Definition 1 (Flagified PCA (fPCA) [4])**.** *A flag of principal components is the solution to:*

$$\underset{[\![\mathbf{U}]\!] \in \mathcal{FL}(n+1)}{\arg\max} \; \mathbb{E}\left[\sum_{i=1}^{k} \|\pi_{\mathbf{U}_i}(\mathbf{x}_j)\|_2^2\right] \tag{74}$$

Ye *et al*. find a solution Eq. (74) using Newton's method on the flag manifold [6] and Nguyen offers a method for solving such a problem using RTR on flag manifolds [3]. These algorithms produce the same basis vectors for flags regardless of flag type. These basis vectors are different than those found using standard PCA. But, for $[\![\mathbf{U}]\!] \in \mathcal{FL}(n_1, n_2, \ldots, n_k, n)$ that solves Eq. (74) using either Newton's method or RTR, the column space of $\mathbf{U}_{:,:n_k}$ is the same as the span of the first n_k principal components. This is because the objective function in Eq. (74) is invariant to ordering the columns of $\mathbf{U}$.

Variants on flagified PCA that maximize $\text{tr}(\mathbf{U}^T \mathbf{X}\mathbf{X}^T \mathbf{U})^q$ over $\mathcal{FL}(n+1)$ are coined "nonlinear eigenflags" and are difficult to solve for $q = 2$ [6]. Yet, methods from Mankovich *et al*. can be adapted to solve such problems, especially for $q = 1/2$. Another variant of fPCA is weighted fPCA where we assume a weight for each subspace dimension in the flag i and each data point j as $w_{ij} \in \mathbb{R}$. We propose this formulation in the manuscript.

C. DPCP-IRLS and the Grassmannian

This concept was first unearthed in [2]. Expanding the matrix norm we have

$$\|\mathbf{X}^T\mathbf{B}\|_{1,2} = \sum_{j=1}^{p} \|\mathbf{B}^T\mathbf{x}_j\|_2, \tag{75}$$

$$= \sum_{j=1}^{p} \sqrt{\sum_{i=1}^{k} |\mathbf{b}_i^T\mathbf{x}_j|^2}, \tag{76}$$

$$= \sum_{j=1}^{p} \sqrt{\mathbf{x}_j^T\mathbf{B}\mathbf{B}^T\mathbf{x}_j}, \tag{77}$$

$$= \sum_{j=1}^{p} \sqrt{\text{tr}\left(\mathbf{B}^T\mathbf{x}_j\mathbf{x}_j^T\mathbf{B}\right)}. \tag{78}$$

This can be phrased using principal angles as

$$\underset{\mathbf{B}^T\mathbf{B}=\mathbf{I}}{\arg\min} \sum_{j=1}^{p} \cos\theta([\mathbf{x}_j], [\mathbf{B}]). \tag{79}$$

Suppose $\{[\mathbf{X}_j]\}_{j=1}^p \subset \text{Gr}(k, n)$. Namely, $\mathbf{X}_j \in \mathbb{R}^{n \times k}$ where $\mathbf{X}_j^T\mathbf{X}_j = \mathbf{I}$ for each j. A natural generalization of DPCP-IRLS is the optimization problem on the Grassmannian,

$$\underset{[\mathbf{B}] \in \text{Gr}(k,n)}{\arg\min} \sum_{j=1}^{p} \|\cos\theta([\mathbf{X}_j], [\mathbf{B}])\|_2. \tag{80}$$

This can also be solved by an IRLS scheme.

The "flagified" version of Eq. (80) is

$$\underset{[\![\mathbf{B}]\!] \in \mathcal{FL}(n+1)}{\arg\min} \sum_{j=1}^{p} \|\cos\theta([\mathbf{X}_j], [\mathbf{B}])\|_2. \tag{81}$$

D. Novel Flagified Robust and Dual PCA and TPCA Variants

We present the intuition behind the geometry of Robust and Dual PCA versus $\mathcal{T}$PCA in Fig. 1. Then we provide a visual comparison between Euclidean and manifold variants of RPCA and DPCP in Fig. 2.

Tab. 1 summarizes our novel flagified robust and dual PCA variants and emphasizes that flag types other than $(1, 2 \ldots, k; n)$ and $(k; n)$ produce novel principal directions that are "in between' L_1 and L_2 formulations.

Finally Tab. 2 summarizes the naming schemes of all of the algorithms intriduced in this paper

E. Rest of the Proposed Algorithms

In the paper, we proposed three new algorithms. We now present these algorithms as well as the objective functions

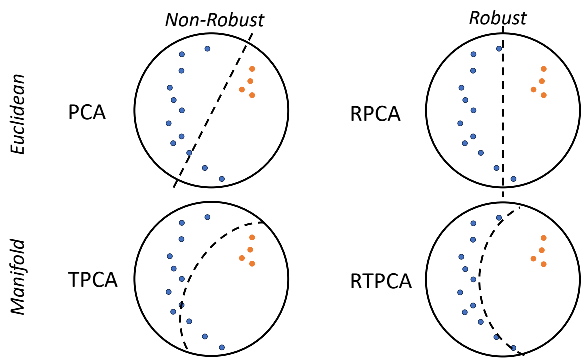

Figure 1. Inliers (blue) and outliers (orange) on the 2-sphere. The first row are Euclidean algorithms and the second row are manifold (tangent space) algorithms. The dashed lines are the first principal subspace (first row) and geodesic (second row) spanned by the first principal direction. Note: first principal subspaces pass through the center of the sphere and first principal geodesics are great circles on the sphere.

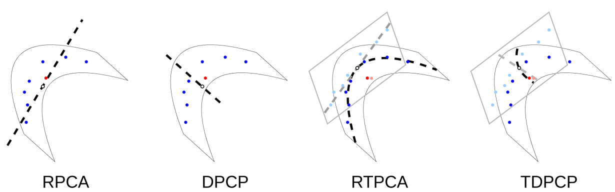

Figure 2. Given manifold valued data with inliers (blue) and outliers (red). The dashed black lines are the 1st principal component for RPCA and DPCP, for R$\mathcal{T}$PCA and $\mathcal{T}$DPCP this is the 1st principal geodesic. For RPCA and R$\mathcal{T}$PCA this line / geodesic should contain the inliers. Due to the reversal in the objective, for DPCP and $\mathcal{T}$DPCP this geodesic should contain the outliers.

they minimize. First, Alg. 1 finds a solution to weighted flagified PCA

$$[\![\mathbf{U}]\!]^\star = \underset{[\![\mathbf{U}]\!] \in \mathcal{FL}(n+1)}{\arg\max} \ \mathbb{E}_j \left[\sum_{i=1}^{k} w_{ij} \|\pi_{\mathbf{U}_i}(\mathbf{x}_j)\|_2^2 \right]. \quad (82)$$

Second, Alg. 2 finds a solution to weighted flagified orthogonal PCA (f⊥PCA)

$$[\![\mathbf{U}]\!]^\star = \underset{[\![\mathbf{U}]\!] \in \mathcal{FL}(n+1)}{\arg\min} \ \mathbb{E}_j \left[\sum_{i=1}^{k} w_{ij} \|\pi_{\mathbf{U}_i}(\mathbf{x}_j)\|_2^2 \right]. \quad (83)$$

Flagified (Dual-)PCA	Robust PCA Variant
fRPCA$(1, ..., k)$	L_1-RPCA
fRPCA$(\cdot)$	–
fRPCA(k)	L_2-RPCA
fWPCA$(1, ..., k)$	L_1-WPCA
fWPCA$(\cdot)$	–
fWPCA(k)	L_2-WPCA
fDPCP$(1, ..., k)$	L_1-DPCP
fDPCP$(\cdot)$	–
fDPCP(k)	L_2-DPCP
fR$\mathcal{T}$PCA$(1, ..., k)$	L_1–R$\mathcal{T}$PCA
fR$\mathcal{T}$PCA$(\cdot)$	–
fR$\mathcal{T}$PCA(k)	L_2–R$\mathcal{T}$PCA
fW$\mathcal{T}$PCA$(1, ..., k)$	L_1–W$\mathcal{T}$PCA
fW$\mathcal{T}$PCA$(\cdot)$	–
fW$\mathcal{T}$PCA(k)	L_2–W$\mathcal{T}$PCA
f$\mathcal{T}$DPCP$(1, ..., k)$	L_1–$\mathcal{T}$DPCP
f$\mathcal{T}$DPCP$(\cdot)$	–
f$\mathcal{T}$DPCP(k)	L_2–$\mathcal{T}$DPCP

Table 1. Flag types for Euclidean optimization (first half) and manifold optimization (second half). Flag optimization in these algorithms provides a new objective functions which live in between L_1 and L_2 robust PCA formulations. Note: we remove the number of the ambient dimension in the flag signature for less redundant notation and we assume we are computing the first k principal components.

Lastly, Alg. 3 approximates solutions to

$$[\![\mathbf{U}]\!]^\star \approx \quad (84)$$
$$\begin{cases} \underset{[\![\mathbf{U}]\!] \in \mathcal{FL}(n+1)}{\arg\max} \ \mathbb{E}_j \left[\sum_{i=1}^{k} w_{ij} d(\boldsymbol{\mu}, \pi_{\mathbf{U}_i}(\mathbf{x}_j))^2 \right], & (\text{f}\mathcal{T}\text{PCA}) \\ \underset{[\![\mathbf{U}]\!] \in \mathcal{FL}(n+1)}{\arg\min} \ \mathbb{E}_j \left[\sum_{i=1}^{k} w_{ij} d(\boldsymbol{\mu}, \pi_{\mathbf{U}_i}(\mathbf{x}_j))^2 \right], & (\text{f}\perp\mathcal{T}\text{PCA}) \\ \underset{[\![\mathbf{U}]\!] \in \mathcal{FL}(n+1)}{\arg\max} \ \mathbb{E}_j \left[\sum_{i=1}^{k} d(\boldsymbol{\mu}, \pi_{\mathbf{U}_i}(\mathbf{x}_j)) \right], & (\text{fR}\mathcal{T}\text{PCA}) \\ \underset{[\![\mathbf{U}]\!] \in \mathcal{FL}(n+1)}{\arg\min} \ \mathbb{E}_j \left[\sum_{i=1}^{k} d(\mathbf{x}_j, \pi_{\mathbf{U}_i}(\mathbf{x}_j)) \right], & (\text{fW}\mathcal{T}\text{PCA}) \\ \underset{[\![\mathbf{U}]\!] \in \mathcal{FL}(n+1)}{\arg\min} \ \mathbb{E}_j \left[\sum_{i=1}^{k} d(\boldsymbol{\mu}, \pi_{\mathbf{U}_i}(\mathbf{x}_j)) \right], & (\text{f}\mathcal{T}\text{DPCP}) \end{cases}$$

F. Extra Experiments

Impact of flag-type on cluster detection. To assess the impact of flag-type, we generate a dataset $\{\mathbf{x}_j\}_{j=1}^{300} \subset \mathbb{R}^{10}$ with 3 clusters (C_1, C_2, C_3) in which we curate the flag type corresponding to the data structure: $\mathcal{FL}(2, 5, 7; 10)$. To do this we sample $\{\mathbf{x}_j\}_{j=1}^{300} \subset \mathbb{R}^{10}$ with 3 clusters. The lth

Abbreviation	Name
PCA	Principal Component Analysis
RPCA	Robust PCA
WPCA	Weiszfeld PCA
DPCP	Dual Principal Component Pursuit
WDPCP	Weiszfeld DPCP
fPCA	Flagified PCA
fRPCA	Flagified RPCA
fWPCA	Flagified WPCA
fDPCP	Flagified DPCP
fWDPCP	Flagified WDPCP
$\mathcal{T}$PCA	Tangent PCA
R$\mathcal{T}$PCA	Robust $\mathcal{T}$PCA
W$\mathcal{T}$PCA	Weiszfeld $\mathcal{T}$PCA
$\mathcal{T}$DPCP	Tangent DPCP
W$\mathcal{T}$DPCP	Tangent WDPCP
f$\mathcal{T}$PCA	Flagified $\mathcal{T}$PCA
fR$\mathcal{T}$PCA	Flagified R$\mathcal{T}$PCA
fW$\mathcal{T}$PCA	Flagified W$\mathcal{T}$PCA
f$\mathcal{T}$DPCP	Flagified $\mathcal{T}$DPCP
fW$\mathcal{T}$DPCP	Flagified W$\mathcal{T}$DPCP

Table 2. The names of the major algorithms covered in this work.

Algorithm 1: Weighted fPCA

Inputs: Dataset $\{\mathbf{x}_j \in \mathbb{R}^n\}_{j=1}^p$,
weights $\{w_{ij}\}_{i,j=1}^{i=k,j=p} \subset \mathbb{R}$,
flag type $(n+1)$
Output: Weighted flagified principal directions $[\![\mathbf{U}]\!]^* \in \mathcal{FL}(n+1)$
for $i = 1, 2, \dots, k$ **do**
  $(\mathbf{W}_i)_{jl} \leftarrow \begin{cases} w_{ij}, & j = l \\ 0, & \text{elsewhere} \end{cases}$
$\mathbf{U}^* \leftarrow$ Solve **??** with $\{\mathbf{W}_i\}_{i=1}^k$ via Stiefel-CGD.
$[\![\mathbf{U}]\!]^* \leftarrow [\![\mathbf{U}^*]\!]$

entry of $\mathbf{x}$, $(\mathbf{x})_l \in \mathbb{R}$, is sampled from

$$C(1) : (\mathbf{x})_l \sim \begin{cases} \mathcal{U}[0,1), & l \leq 2 \\ \mathcal{U}[0,0.1), & l \geq 3 \end{cases}, \tag{85}$$

$$C(2) : (\mathbf{x})_l \sim \begin{cases} \mathcal{U}[0,1), & 3 \leq i \leq 5 \\ \mathcal{U}[0,0.1), & i \leq 2 \text{ or } i \geq 6 \end{cases}, \tag{86}$$

$$C(3) : (\mathbf{x})_l \sim \begin{cases} \mathcal{U}[0,1), & i = 6,7 \\ \mathcal{U}[0,0.1), & i \leq 5 \text{ or } i \geq 8 \end{cases}. \tag{87}$$

We then compute 2 sets of $k = 7$ principal directions by running fWPCA with flag type $(2, 5, 7; 10)$

Algorithm 2: Weighted flag $\perp$PCA (f$\perp$PCA)

Inputs: Dataset $\{\mathbf{x}_j \in \mathbb{R}^n\}_{j=1}^p$,
weights $\{w_{ij}\}_{i,j=1}^{i=k,j=p} \subset \mathbb{R}$,
flag type $(n+1)$
Output: Weighted flagified principal directions $[\![\mathbf{U}]\!]^* \in \mathcal{FL}(n+1)$
for $i = 1, 2, \dots, k$ **do**
  $(\mathbf{W}_i)_{jl} \leftarrow \begin{cases} w_{ij}, & j = l \\ 0, & \text{elsewhere} \end{cases}$
$\mathbf{U}^* \leftarrow$ Minimize the objective in **??** with $\{\mathbf{W}_i\}_{i=1}^k$ via Stiefel-CGD.
$[\![\mathbf{U}]\!]^* \leftarrow [\![\mathbf{U}^*]\!]$

Algorithm 3: f$\mathcal{T}$PCA/fR$\mathcal{T}$PCA/fW$\mathcal{T}$PCA/f$\mathcal{T}$DPCP

Input: Dataset: $\{\mathbf{x}_j\}_{j=1}^p \subset \mathcal{M}$, flag type $(n+1)$,
fPCA Variant: $\Phi : \mathcal{W} \rightarrow \mathcal{FL}(n+1)$
Output: Flagified principal tangent directions $[\![\mathbf{U}]\!]^*$
if *robust* **then**
  $\boldsymbol{\mu} \leftarrow \text{KarcherMedian}\left(\{\mathbf{x}_j\}_{j=1}^p\right)$
else
  $\boldsymbol{\mu} \leftarrow \text{KarcherMean}\left(\{\mathbf{x}_j\}_{j=1}^p\right)$
$\{\mathbf{v}_j\}_j \leftarrow \{\text{Exp}_{\boldsymbol{\mu}}(\mathbf{x}_j)\}_j$
$\mathcal{W} \leftarrow \{\text{vec}(\mathbf{v}_j)\}_j$
$[\![\mathbf{U}]\!]^* \leftarrow \Phi(\mathcal{W}, n+1)$

	Cluster 1		Cluster 2		Cluster 3	
fWPCA$(\cdot)$	(7)	$(2,5,7)$	(7)	$(2,5,7)$	(7)	$(2,5,7)$
AUC ↑	0.72	**0.73**	0.48	**1.00**	0.43	**0.49**

Table 3. AUC for cluster classification using fWPCA. We see higher AUCs when we match the flag type for fWPCA with the cluster dimensions (e.g., $(2, 5, 7)$).

(fWPCA$(2, 5, 7)$) and fWPCA(k) using **??** with 200 max. iters. Both of these methods result in a flag representative $\mathbf{U} = [\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3] \in \mathbb{R}^{10\times 7}$ where $\mathbf{U}_1 \in \mathbb{R}^{10\times 2}$, $\mathbf{U}_2 \in \mathbb{R}^{10\times 3}$, and $\mathbf{U}_3 \in \mathbb{R}^{10\times 2}$. We compute the reconstruction error for point j against each $\mathbf{U}_i$ as $\sum_{j=1}^p \|\mathbf{x}_j - \mathbf{U}_i\mathbf{U}_i^T\mathbf{x}_j\|_2$. These errors are used for 3 classification tasks, predicting C_i using $\mathbf{U}_i$ for $i = 1, 2, 3$. The corresponding AUC values are in Tab. 3. fWPCA$(2, 5, 7)$ produces higher AUCs because it is optimized over a more optimal flag type, respecting the subspace structure of the data.

Data generation for "Convergence on 4-sphere". We first sample a random center $\mathbf{x} \in \mathbb{S}^4$, and then sample 100 inlier tangent vectors from $\mathcal{U}[0, .01)$. Another 20 outlier tangent vectors $\mathbf{v}$, have entries $v_1, v_2 \sim \mathcal{U}[0, .01)$ and $v_3, v_4, v_5 \sim \mathcal{U}[0, .1)$. We wrap these vectors to have our

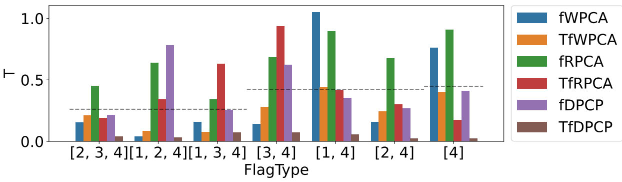

Figure 3. Smaller T corresponds to principal directions which are more similar to those computed with flag type $(1, 2, 3, 4; 5)$. The mean T values for each class of flag type are the horizontal dashed lines. Notice that, these mean values increase as we increase the distance between flag types. We truncate flag types by removing the ambient dimension (5).

dataset, $\{\text{Exp}_{\mathbf{x}}(\mathbf{v})\}$.

Impact of flag type on principal directions. We run flagified robust PCA and $\mathcal{T}$PCA variants using **??** (with 200 max. iters.) with different flag types on data on $Gr(2, 4)$ with 100 inliers and 20 outliers sampled as described in the "Outlier detection on $Gr(2, 4)$" section of the manuscript. We call $(1, 2, 3, 4; 5)$ the "base" flag type. We use T to measure the different between principal directions from the base flag type $\{\mathbf{u}_1 \dots, \mathbf{u}_4\}$ and other principal directions $\{\mathbf{v}_1 \dots, \mathbf{v}_4\}$ as

$$T = \frac{1}{4}\sum_{i=1}^{4} \theta(\mathbf{u}_i, \mathbf{v}_i)^2. \tag{88}$$

We plot T values for different flagified robust PCA and $\mathcal{T}$PCA variants in Fig. 3. We separate flag types into classes based on the number of nested subspaces. Flag types with the same number of nested subspaces are considered "closer" flags. We find that closer flag types have smaller T values. This experiment verifies that running flagified robust PCA variants with different flag types recover different principal directions and these differences are directly proportional to the "distance" between flag types. This also emphasizes that flag types other than $(1, \dots, k; n)$ and $(k; n)$ indeed recover novel principal directions. The direct utility of these gap-filling methods to real-world datasets is future work.

Outlier detection on $Gr(2, 4)$. We present the result of using PCA, fWPCA$(1, ..., k)$, fWPCA(k), fRPCA$(1, ..., k)$, fRPCA(k), fDPCP$(1, ..., k)$, and fDPCP(k) on $Gr(2, 4)$ data for outlier detection in Fig. 4. This is the same data as the data used for **??**; but, in this case, we run our algorithms on the vectorized matrix representatives for points on $Gr(2, 4)$ and do outlier detection using Euclidean distance and variances.

Hand reconstructions. We use the 2D Hands dataset and add "hairball" outliers by sampling from a normal distribution with mean 0 and standard deviation 10 ($\mathcal{N}(0, 10)$), then we divide by the Frobenius norm and mean center to obtain

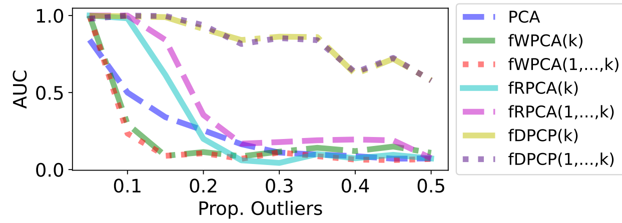

Figure 4. AUC of different algorithms for outlier detection using the first $k = 2$ principal directions of outlier-contaminated data on $Gr(2, 4)$. All algorithms other than PCA are optimized with **??** with 100 max. iters.

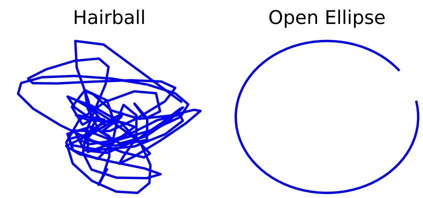

Figure 5. Examples of outliers used for contamination of the hands dataset. Hairballs are used in hand reconstruction and open ellipses are used in outlier detection.

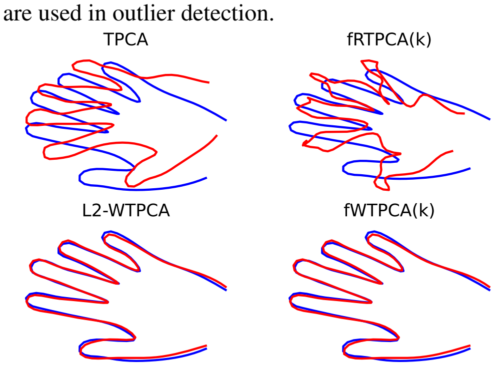

Figure 6. Reconstruction of hand 6 using the first principal direction computed on a dataset with 40 hands and 5 outliers. The cumulative reconstruction errors for the 40 inlier hands from L to R, Top to Bottom, are: 8.19, 6.20, 5.35, and 5.35.

a point on Σ_2^{56}. A figure with an example of an outlier ellipse and a hairball outlier is in Fig. 5.

We run fW$\mathcal{T}$PCA$(1, ..., k)$, L_1–W$\mathcal{T}$PCA using Alg. 1 from [5] run on the tangent space, fR$\mathcal{T}$PCA$(1, ..., k)$, and $\mathcal{T}$PCA to find different versions of the first $k = 1$ principal direction on a dataset with all 40 hands and 5 outliers. We compute reconstruction error for each method using the framework described in the $Gr(2, 4)$ experiments. Our cumulative reconstruction errors for the 40 inlier hands and a visualization of a hand reconstruction is in Fig. 6. L_1–W$\mathcal{T}$PCA and fW$\mathcal{T}$PCA$(1, ..., k)$ produce the lowest reconstruction errors on the hands and have the most sensible reconstructions. Additionally, Alg. 3 preforms just as well as Alg. 1 from [5] run on the tangent space.

We move on to computing cumulative inlier reconstruction errors as we gradually add outliers and report results in Fig. 7. fW$\mathcal{T}$PCA have the most stable reconstruction errors followed by fR$\mathcal{T}$PCA, then $\mathcal{T}$PCA.

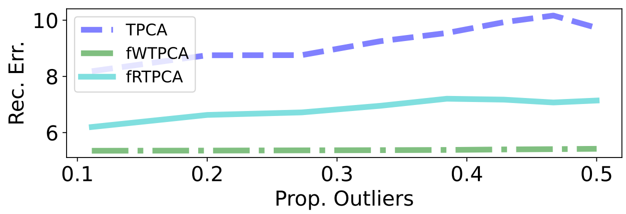

Figure 7. The cumulative reconstruction error of the 40 inlier hands using the first $k = 1$ principal direction where we gradually add hairball outliers.

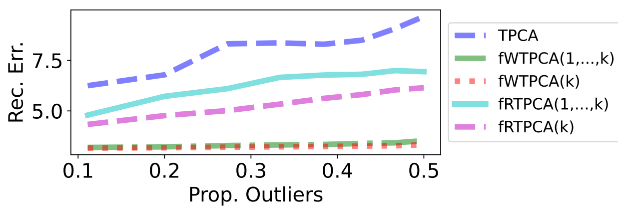

Figure 8. The cumulative reconstruction error of the 40 inlier hands using the first $k = 2$ principal directions where we gradually add hairball outliers.

References

[1] Nathan Mankovich and Tolga Birdal. Chordal averaging on flag manifolds and its applications. In *ICCV*, pages 3881–3890, 2023. 4

[2] Nathan J Mankovich. *Subspace and Network Averaging for Computer Vision and Bioinformatics*. PhD thesis, Colorado State University, 2023. 5

[3] Du Nguyen. Closed-form geodesics and optimization for Riemannian logarithms of Stiefel and flag manifolds. *Journal of Optimization Theory and Applications*, 194(1), 2022. 5

[4] Xavier Pennec. Barycentric subspace analysis on manifolds. *Annals of Statistics*, 46(6A), 2018. 5

[5] Qianqian Wang, Quanxue Gao, Xinbo Gao, and Feiping Nie. $\ell_{2,p}$-norm based PCA for image recognition. *IEEE Transactions on Image Processing*, 27(3):1336–1346, 2017. 8

[6] Ke Ye, Ken Sze-Wai Wong, and Lek-Heng Lim. Optimization on flag manifolds. *Mathematical Programming*, 194(1):621–660, 2022. 5