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

Abstract

Principal component analysis (PCA), along with its extensions to manifolds and outlier contaminated data, have been indispensable in computer vision and machine learning. In this work, we present a unifying formalism for PCA and its variants, and introduce a framework based on the flags of linear subspaces, i.e. *a hierarchy of nested linear subspaces of increasing dimension, which not only allows for a common implementation but also yields novel variants, not explored previously. We begin by generalizing traditional PCA methods that either maximize variance or minimize reconstruction error. We expand these interpretations to develop a wide array of new dimensionality reduction algorithms by accounting for outliers and the data manifold. To devise a common computational approach, we recast robust and dual forms of PCA as optimization problems on flag manifolds. We then integrate tangent space approximations of principal geodesic analysis (tangent-PCA) into this flag-based framework, creating novel robust and dual geodesic PCA variations. The remarkable flexibility offered by the 'flagification' introduced here enables even more algorithmic variants identified by specific flag types. Last but not least, we propose an effective convergent solver for these flag-formulations employing the Stiefel manifold. Our empirical results on both real-world and synthetic scenarios, demonstrate the superiority of our novel algorithms, especially in terms of robustness to outliers on manifolds.*

1. Introduction

Dimensionality reduction is at the heart of machine learning, statistics, and computer vision. Principal Component Analysis (PCA) [27, 52] is a well-known technique for reducing the dimensionality of a dataset by linearly transforming the data into a new coordinate system where (most of) the variation in the data can be described with fewer dimensions than the initial data. Thanks to its simplicity, effectiveness and versatility, PCA has quickly been extended to nonlinear transforms [6, 46, 61], Riemannian manifolds [21, 37, 55] or to unknown number of subspaces [71]. These have proven indispensable for extracting meaningful information from complex datasets.

In this paper, we present a unifying framework marrying a large family of PCA variants such as robust PCA [36, 44, 57], dual PCA [69], PGA [63] or tangent PCA [21] specified by the norms and powers in a common objective function. Being able to accommodate all these different versions into the same framework allows us to innovate novel ones, for example, *tangent dual PCA*, which poses a strong method for outlier filtering on manifolds. We further enrich the repertoire of available techniques by representing the space of eigenvectors as 'flags' [4], a hierarchy of nested linear subspaces, in a vein similar to [54]. This 'flagification' paves the way to a common computational basis, and we show how all these formulations can be efficiently implemented via a single algorithm that performs Riemannian optimization on Stiefel manifolds [8]. This algorithm additionally contributes to the landscape of optimization techniques for dimensionality reduction, and we prove its convergence for the particular case of dual PCA.

In summary, **our contributions are**:

- Generalization of PCA, PGA, and their robust versions leading to new novel variants of these principal directions
- A unifying flag manifold-based framework for computing principal directions of (non-)Euclidean data yielding novel (tangent) PCA formulations between L_1 and L_2 robust and dual principal directions controlled by flag types
- Novel weighting schemes, not only weighting the directions but also the subspaces composed of these directions.
- A practical way to optimize objectives on flags by mapping the problems into Stiefel-optimization, which removes the need to convey optimization on flag manifolds, an issue remaining open to date [66, 77]

Our theoretical exposition translates to excellent and remarkable findings, validating the usefulness of our novel toolkit in several applications, from outlier prediction to shape analysis. The implementation can be found here.

2. Related Work

Our work will heavily combine PCA with flag manifolds.

PCA and its variants. Although there are many variants of PCA [5, 11, 27–29, 61, 71, 78], we focus on certain forms of Robust PCA (RPCA) [35, 44, 46] and Dual PCA (DPCA) [69]. As opposed to RPCA, DPCA finds directions orthogonal to RPCA and is designed to work on datasets with outliers by minimizing a 0-norm problem [69].

We also consider the generalization of PCA to Riemannian manifolds, Principal Geodesic Analysis (PGA), which finds geodesic submanifolds that best represent the data [21] and has been applied to the manifold of SPD matrices on "real-world" datasets [25, 62]. Though, exact PGA is hard to compute [63, 64]. Principal Curves [26] finds curves passing through the mean of a dataset which maximize variance and have seen their own enhancements [37]. Linearized (tangent) versions of PGA perform PCA on the tangent space to the mean of the data [1, 6]. Geodesic PCA (GPCA) removes the mean requirement [30, 31]. Barycentric Subspace Analysis (BSA) realizes PCA on more classes of manifolds than just Riemannian manifolds by generalizing geodesic subspaces using weighted means of reference points [54]. Recent years have witnessed further generalizations [9, 55, 58] as we will mention later.

Flag manifolds. Flag manifolds are useful mathematical objects [4, 15, 34, 74, 77]. Nishimori *et al*. use Riemannian optimization and flag manifolds to formulate variations of independent component analysis [48–50, 50, 51]. Others represent an average subspace as a flag [16, 42, 43] and find average flags [41]. Flags even arise as nested principal directions and in manifold variants of PCA [54, 55, 77].

3. Preliminaries

Let us start by briefly introducing Riemannian geometry, flag manifolds, and methods for finding principal directions. The flag and flag manifold definitions follow [41].

Definition 1 (Riemannian manifold [38])**.** *A* Riemannian manifold $\mathcal{M}$ *is a smooth manifold with a positive definite inner product* $\langle\cdot,\cdot\rangle : \mathcal{T}_{\mathbf{x}}\mathcal{M} \times \mathcal{T}_{\mathbf{x}}\mathcal{M} \to \mathbb{R}$ *defined on the tangent space* $\mathcal{T}_{\mathbf{x}}\mathcal{M}$ *at* $\mathbf{x} \in \mathcal{M}$.

Definition 2 (Geodesics & Exp/Log-maps [38])**.** *A* geodesic $\gamma : \mathbb{R} \to \mathcal{M}$ *parameterizes a path from* $\gamma(0) = \mathbf{x}$ *to* $\gamma(1) = \mathbf{y}$. *The* exponential map $\mathrm{Exp}_{\mathbf{x}}(\mathbf{v}) : \mathcal{T}_{\mathbf{x}}\mathcal{M} \to \mathcal{M}$ *maps a vector* $\mathbf{v} \in \mathcal{T}_{\mathbf{x}}\mathcal{M}$ *to the manifold in a length preserving fashion such that* $\dot{\gamma}_{\mathbf{v}}(0) = \mathbf{v}$ *and* $\mathrm{Exp}_{\mathbf{x}}(\mathbf{v}) = \mathbf{y}$. *Its inverse, the* logarithmic map *is* $\mathrm{Log}_{\mathbf{x}}(\mathbf{y}) : \mathcal{M} \to \mathcal{T}_{\mathbf{x}}\mathcal{M}$ *for* $\mathbf{x}, \mathbf{y} \in \mathcal{M}$ *and computes the tangent direction from* $\mathbf{x}$ *to* $\mathbf{y}$. *Hence,* $\gamma(t) = \mathrm{Exp}_{\mathbf{x}}(\tau \mathrm{Log}_{\mathbf{x}}(\mathbf{y}))$ *for* $\tau \in \mathbb{R}$ *is the geodesic curve.* $\mathcal{H} \in \mathcal{M}$ *is said to be a geodesic submanifold at* $\mathbf{x} \in \mathcal{M}$ *if all geodesics through* $\mathbf{x}$ *in* $\mathcal{H}$ *are geodesics in* $\mathcal{M}$.

Definition 3 (Flag)**.** *A* flag *is a nested sequence of subspaces of a finite-dimensional vector space* $\mathcal{V}$ *of* increasing *dimension, is the filtration* $\{\emptyset\} = \mathcal{V}_0 \subset \mathcal{V}_1 \subset \ldots \subset \mathcal{V}_k \subset \mathcal{V}$ *with* $0 = n_0 < n_1 < \ldots < n_k < n$ *where* $\dim\mathcal{V}_i = n_i$ *and* $\dim\mathcal{V} = n$. *The* type *or* signature *of this flag is* $(n_1, \ldots, n_k; n)$ *or* $(n_1, \ldots, n_k)$.

Notation*:* *We denote a flag using a point* $\mathbf{X}$ *on the Stiefel manifold [18]* $St(n_k, n) = \{\mathbf{X} \in \mathbb{R}^{n\times n_k} : \mathbf{X}^T\mathbf{X} = \mathbf{I}\}$. *Given* $\mathbf{X} \in St(n_k, n)$ *and* $\mathbf{x}_i$, i^{th} *column of* $\mathbf{X}$, *we define*

$$\mathbf{X}_{i+1} = \begin{bmatrix}\mathbf{x}_{n_i+1} & \mathbf{x}_{n_i+2} & \cdots & \mathbf{x}_{n_{i+1}}\end{bmatrix} \in \mathbb{R}^{n\times m_{i+1}}. \quad (1)$$

where $m_i = n_i - n_{i-1}$ *for* $i = 1, 2, \ldots, k$. $[\mathbf{X}_1, \ldots, \mathbf{X}_i]$ *denotes the span of the columns of* $\{\mathbf{X}_1, \ldots, \mathbf{X}_i\}$. *Then*

$$[\mathbf{X}_1] \subset [\mathbf{X}_1, \mathbf{X}_2] \subset \cdots \subset [\mathbf{X}_1, \ldots, \mathbf{X}_k] = [\mathbf{X}] \subset \mathbb{R}^n.$$

is a flag of type $(n_1, \ldots, n_k; n)$ *and is denoted* $[\![\mathbf{X}]\!]$.

Definition 4 (Flag manifold)**.** *The set of all flags of type* $(n_1, \ldots, n_k; n)$ *is called the flag manifold due to its manifold structure. We refer to this* flag manifold *as* $\mathcal{FL}(n_1, \ldots, n_k; n)$ *or* $\mathcal{FL}(n+1)$. *Flags generalize Grassmann and Stiefel manifolds [18] because* $\mathcal{FL}(n_k; n) = Gr(n_k, n)$ *and* $\mathcal{FL}(1, \ldots, n_k; n) = St(n_k, n)$. *We denote flags as* $\mathcal{FL}(n+1)$ *using the fact from [77]:*

$$\mathcal{FL}(n+1) = St(n_k, n)/O(m_1)\times O(m_2)\times\cdots\times O(m_{k+1}).$$

4. Generalizing PCA and Its Robust Variants

Principal directions are the directions where the data varies. We now review and go beyond the celebrated principal component analysis (PCA) [27] algorithm and its variants. In what follows, we present PCA in a generalizing framework, which further yields novel variants. We consider a set of p centered samples (points with a sample mean of 0) with n random variables (features) $\mathcal{X} = \{\mathbf{x}_j \subset \mathbb{R}^n\}_{j=1}^p$ and collect these data in the matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p]$.

Definition 5 (PCA [27])**.** *PCA aims to linearly transform the data into a new coordinate system, specified by a set of* $k < n$ *orthonormal vectors* $\{\mathbf{u}_i \in \mathbb{R}^n\}_{j=1}^k$, *where (most of) the variation in the data can be described with fewer dimensions than the initial data. The* i^{th} *principal direction is obtained either by maximizing variance Eq. (2) or minimizing reconstruction error (Eq. (3)):*

$$\mathbf{u}_i = \underset{\mathbf{u}^T\mathbf{u}=1,\, \mathbf{u}\in S_i^\perp}{\arg\max} \; \mathbb{E}_j\left[\|\pi_{S_\mathbf{u}}(\mathbf{x}_j)\|_2^2\right] \quad (2)$$

$$\mathbf{u}_i = \underset{\mathbf{u}^T\mathbf{u}=1,\, \mathbf{u}\in S_{i-1}^\perp}{\arg\min} \; \mathbb{E}_j\left[\|\mathbf{x}_j - \pi_{S_\mathbf{u}}(\mathbf{x}_j)\|_2^2\right], \quad (3)$$

where $\pi_{S_\mathbf{u}}(\mathbf{x}) := \mathbf{u}\mathbf{u}^T\mathbf{x}$, $S_i = \mathrm{span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_i\}$, $S_i^\perp$ *denotes its orthogonal complement, and* $\mathbb{E}_j$ *denotes expectation over* j. *Although these objectives can be achieved jointly [73], in this work, we focus on the more common practice specified above. Notice that Eq. (2) is equivalent (up to rotation) to solving the following optimization:*

$$\underset{\mathbf{U}^T\mathbf{U}=\mathbf{I}}{\arg\max}\, \mathrm{tr}(\mathbf{U}^T\mathbf{X}\mathbf{X}^T\mathbf{U}) \quad (4)$$

Directions of maximum variance captured by the naive PCA are known to be susceptible to outliers in data. This motivated a body of work devising more robust versions.

Definition 6 (Generalized PCA). *The formulations in Dfn. 5 can be generalized by using arbitrary L_p-norms, q-powers, and weighted with real weights $\{w_1, w_2, \ldots, w_p\}$:*

$$\mathbf{U}^\star = \underset{\mathbf{U}^T\mathbf{U}=\mathbf{I}}{\arg\max}\ \mathbb{E}_j \left[w_j \|\pi_{S_\mathbf{U}}(\mathbf{x}_j)\|_p^q \right], \tag{5}$$

$$\mathbf{U}^\star = \underset{\mathbf{U}^T\mathbf{U}=\mathbf{I}}{\arg\min}\ \mathbb{E}_j \left[w_j \|\mathbf{x_j} - \pi_{S_\mathbf{U}}(\mathbf{x}_j)\|_p^q \right], \tag{6}$$

where $\pi_{S_\mathbf{U}}(\mathbf{x}) := \mathbf{U}\mathbf{U}^T\mathbf{x}$. Both problems recover PCA up to a rotation when $p = q = 2$. When $p \leq 2$ and $q < 2$, a more outlier robust *version of PCA is achieved, and the two formulations become different. The variance-maximizing (Eq. (5)) $q = 1$ case is known as L_p-RPCA [36]. Specifically, when $q = p = 1$, it recovers L_1-RPCA [44] and when $q = 1$, $p = 2$, it recovers L_2-RPCA [57]. On the other hand, minimizing the reconstruction error (Eq. (6)) leads to L_1-Weiszfeld PCA (L_1-WPCA) [46] for $q = p = 1$ and L_2-WPCA [14] for $q = 1$, $p = 2$.*

Generalizing Dual-PCA. Momentarily assume that the data matrix can be decomposed into inliers $\mathbf{X}_I$ and outliers $\mathbf{X}_O$: $\mathbf{X} = [\mathbf{X}_I, \mathbf{X}_O]\mathbf{P}$, where $\mathbf{P}$ is a permutation matrix.

Definition 7 (Dual-PCA [69]). *Assuming the samples live on the unit sphere, $\{\mathbf{x}_i\}_{i=1}^p \in \mathbb{S}^{n-1}$, DPCA seeks a subspace $S_* = \operatorname{span}\{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_k\}$ so that $S_*^\perp$ contains the most inliers (e.g., columns of $\mathbf{X}_I$). In other words, we seek vectors $\{\mathbf{b}_i\}_{i=1}^k$ that are* as orthogonal as possible *to the span of the inliers. This is formalized in the iterative optimization problem*

$$\mathbf{b}_i = \underset{\|\mathbf{b}\|=1,\ \mathbf{b}\in S_{i-1}^\perp}{\arg\min} \|\mathbf{X}^T\mathbf{b}\|_0. \tag{7}$$

When we re-write the maximum variance formulation of the PCA optimization in Eq. (5) as the minimization (e.g., $\mathbf{b}_i = \arg\min_{\|\mathbf{b}\|=1,\, \mathbf{u}\in S_{i-1}^\perp} \|\mathbf{X}^T\mathbf{b}\|_2^2$) we see that Dual-PCA minimizes a similar objective that is robustified by considering a 0-norm.

Definition 8 (Dual Principal Component Pursuit (DPCP)). *Relaxing the L_0-norm sparse problem into an L_1-norm one turns the DPCA problem into L_1-DPCP [69]. When solved via an L_2-relaxed scheme, we recover L_2-DPCP (referred to as DPCP-IRLS by [69]).*

Definition 9 (Dual PCA Generalizations). *In general, we can think of Dfn. 7 and 8 as variance-minimizing (Eq. (8)) and reconstruction-error-maximizing (Eq. (9)), respectively:*

$$\mathbf{B}^\star = \underset{\mathbf{B}^T\mathbf{B}=\mathbf{I}}{\arg\min}\, \mathbb{E}_j \left[\|\pi_{S_\mathbf{B}}(\mathbf{x}_j)\|_p^q \right] \tag{8}$$

$$\mathbf{B}^\star = \underset{\mathbf{B}^T\mathbf{B}=\mathbf{I}}{\arg\max}\, \mathbb{E}_j \left[\|\mathbf{x_j} - \pi_{S_\mathbf{B}}(\mathbf{x}_j)\|_p^q \right]. \tag{9}$$

Using Eq. (8), we recover L_1-DPCP with $p = q = 1$ and L_2-DPCP with $p = 2$ and $q = 1$.

Remark 1 (New DPCP Variants). *Two other possibilities optimize Eq. (9) when $p = q = 1$ and $p = 2, q = 1$. We call these methods L_p-Weiszfeld DPCPs (WDPCPs), specifically, L_1-WDPCP and L_2-WDPCP for different p values.*

Extension to Riemannian manifolds. Principal geodesic analysis (PGA) [21] generalizes PCA for describing the variability of data $\{\mathbf{x}_i \in \mathcal{M}\}_{i=1}^p$ on a Riemannian manifold $\mathcal{M}$, induced by the *geodesic distance* $d(\cdot,\cdot) : \mathcal{M} \times \mathcal{M} \to \mathbb{R}_+$. To this end, PGA first requires a manifold-mean:

$$\boldsymbol{\mu} = \underset{\mathbf{y}\in\mathcal{M}}{\arg\min}\, \mathbb{E}_j \left[d(\mathbf{x}_j, \mathbf{y})^2 \right], \tag{10}$$

where the local minimizer is called the *Karcher mean* and if there is a global minimizer, it is called the *Fréchet mean*. A more robust version of the Karcher mean is the *Karcher median* which minimizes $\mathbb{E}_j [d(\mathbf{x}_j, \mathbf{y})]$ and can be estimated by running a Weiszfeld-type algorithm [3]. Next, PGA uses geodesics rather than lines, locally the shortest path between two points, as one-dimensional subspaces.

Definition 10 (PGA [20, 63]). *The i^{th} principal geodesic for (exact) PGA is defined as $\gamma_i(t) = \operatorname{Exp}_{\boldsymbol{\mu}}(\mathbf{u}_i t)$ constructed either by maximizing variance (Eq. (11)) or by minimizing reconstruction error (or unexplained variance) (Eq. (12)):*

$$\mathbf{u}_i = \underset{\|\mathbf{u}\|=1,\ \mathbf{u}\in S_{i-1}^\perp}{\arg\max}\ \mathbb{E}_j \left[d(\boldsymbol{\mu}, \pi_{\mathcal{H}(S_\mathbf{u})}(\mathbf{x}_j))^2 \right] \tag{11}$$

$$\mathbf{u}_i = \underset{\|\mathbf{u}\|=1,\ \mathbf{u}\in S_{i-1}^\perp}{\arg\min}\ \mathbb{E}_j \left[d(\mathbf{x}_j, \pi_{\mathcal{H}(S_\mathbf{u})}(\mathbf{x}_j))^2 \right], \tag{12}$$

where $S_\mathbf{u} = \operatorname{span}\{\mathbf{u}\}$, the i^{th} subspace of $\mathcal{T}_{\boldsymbol{\mu}}\mathcal{M}$, and

$$S_i = \operatorname{span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_i\}. \tag{13}$$

The projection operator onto $\mathcal{H}(S) \subset \mathcal{M}$, the geodesic submanifold of $S \subset \mathcal{T}_{\boldsymbol{\mu}}\mathcal{M}$, is [1]

$$\pi_{\mathcal{H}}(\mathbf{x}) = \underset{\mathbf{z}\in\mathcal{H}}{\arg\min}\, d(\mathbf{z}, \mathbf{x}). \tag{14}$$

where $\mathcal{H}(S) = \{\operatorname{Exp}_{\boldsymbol{\mu}}(\mathbf{v}) : \mathbf{v} \in S\}$.

Remark 2. *In contrast to PCA, PGA needs to explicitly define $\mathcal{H}(S)$ because the tangent space and manifold are distinct. While in Euclidean space, maximizing variances is equivalent to minimizing residuals, Eq. (11) and Eq. (12) are not equivalent on Riemannian manifolds [63]. PGA results in a flag of subspaces of the tangent space of type $(1, 2, \ldots, k; \dim(T_{\boldsymbol{\mu}}(\mathcal{M}))$ in Eq. (15) along with an increasing sequence of geodesic submanifolds in Eq. (16)*

$$S_1 \subset S_2 \subset \cdots \subset S_k \subset T_{\boldsymbol{\mu}}(\mathcal{M}), \tag{15}$$

$$\mathcal{H}(S_1) \subset \mathcal{H}(S_2) \subset \cdots \subset \mathcal{H}(S_k) \subset \mathcal{M}. \tag{16}$$

[1] π_H can use any monotonically increasing function of distance on $\mathcal{M}$.

Additionally, a set of principal directions, $\{\mathbf{u}_i\}_{i=1}^k$, form a (totally) geodesic submanifold $\mathcal{H}(S_k) \subseteq \mathcal{M}$ as long as geodesics in $\mathcal{H}(S_k)$ are carried to geodesics in $\mathcal{M}$ [67].

Similar to Dfn. 6, we now generalize PGA.

Definition 11 (PGA Generalizations). *Let $\{w_j\}_{j=1}^p \subset \mathbb{R}$ denote a set of weights. The weighted principle geodesic is $\gamma_i(t) = \mathrm{Exp}_{\boldsymbol{\mu}}(\mathbf{u}_i t)$ where $\mathbf{u}_i$ maximizes / minimizes:*

$$\mathbf{u}_i = \underset{\|\mathbf{u}\|=1,\, \mathbf{u}\in S_{i-1}^{\perp}}{\arg\max} \; \mathbb{E}_j \left[w_j d(\boldsymbol{\mu}, \pi_{\mathcal{H}(S_{\mathbf{u}})}(\mathbf{x}_j))^q \right] \tag{17}$$

$$\mathbf{u}_i = \underset{\|\mathbf{u}\|=1,\, \mathbf{u}\in S_{i-1}^{\perp}}{\arg\min} \; \mathbb{E}_j \left[w_j d(\mathbf{x}_j, \pi_{\mathcal{H}(S_{\mathbf{u}})}(\mathbf{x}_j))^q \right], \tag{18}$$

This recovers PGA when $q = 2$ and $w_j = 1$ for all j.

Remark 3 (Tangent-PCA ($\mathcal{T}$PCA) [21]). *PGA is known to be computationally expensive to compute except on a few simple manifolds [59, 67]. As a remedy, Fletcher* et al. *[21] leverage the Euclidean-ness of the tangent space to define principal geodesics as $\gamma(t) = \mathrm{Exp}_{\boldsymbol{\mu}}(t\mathbf{u}_i)$ where $\{\mathbf{u}_i\}_{i=1}^k$ are the principal components of $\{\mathrm{Log}_{\boldsymbol{\mu}}\mathbf{x}_j\}_{j=1}^p$. This approximation, known as* Tangent-PCA*, and we will later approximately invert it to reconstruct data on $\mathcal{M}$ by (i) using principal directions to reconstruct the data on $T_{\boldsymbol{\mu}}(\mathcal{M})$, then (ii) mapping the reconstruction to $\mathcal{M}$ using $\mathrm{Exp}_{\boldsymbol{\mu}}(\cdot)$.*

Proposition 1 (Robust PGAs (RPGA & WPGA)). *Setting $1 \leq q < 2$, gives us novel, robust formulations of the PGA problem (RPGA and WPGA) defined in Dfn. 11, which we will solve in the unifying flag framework we provide. While general robust manifold-optimizers such as* robust median-of-means *[40] can be used to implement RPGA and WPGA, to be consistent with $\mathcal{T}$PCA, we will approximate these problems by performing RPCA and WPCA in the tangent space of the robust Karcher median (removing the square in Eq. (10)). We will refer to these tangent space versions as tangent RPCA (R$\mathcal{T}$PCA) and tangent WPCA (W$\mathcal{T}$PCA).*

We are now ready to formulate novel, dual versions of PGA.

Proposition 2 (Dual PGA (DPGA)). *Given a dataset on a Riemannian manifold, we define dual robust principal directions, analogous to DPCA (Dfn. 9):*

$$\mathbf{b}_i = \underset{\|\mathbf{b}\|=1\,,\, \mathbf{b}\in S_{i-1}^{\perp}}{\arg\min} \; \mathbb{E}_j \left[w_j d(\boldsymbol{\mu}, \pi_{H(S_{\mathbf{b}})}(\mathbf{x}_j))^q \right] \tag{19}$$

$$\mathbf{b}_i = \underset{\|\mathbf{b}\|=1\,,\, \mathbf{b}\in S_{i-1}^{\perp}}{\arg\max} \; \mathbb{E}_j \left[w_j d(\mathbf{x}_j, \pi_{H(S_{\mathbf{b}})}(\mathbf{x}_j))^q \right]. \tag{20}$$

We refer to these novel principal directions as DPGP (Eq. (19)) and WDPGP (Eq. (20)). Again, we can approximate these problems by performing DPCP and WDPCP in the tangent space, resulting in the tractable algorithms of tangent DPCP ($\mathcal{T}$DPCP) and tangent WDPCP (W$\mathcal{T}$DPCP).

		(p,q)	Variance	Rec. Err.	$\mathcal{FL}(\cdot;n)$
Euclidean	Primal	$(2,2)$	PCA [27]	PCA [27]	$(1,2,...,k)$
		$(2,1)$	L_2-RPCA [57]	L_2-WPCA [14]	(k)
		$(1,1)$	L_1-RPCA [44]	L_1-WPCA [46]	$(1,2,...,k)$
	Dual	$(2,2)$	⊥PCA [69]	⊥PCA [69]	$(1,2,...,k)$
		$(2,1)$	L_2-DPCP [69]	L_2-WDPCP	(k)
		$(1,1)$	L_1-DPCP [69]	L_1-WDPCP	$(1,2,...,k)$
Manifold	Primal	$(2,2)$	$\mathcal{T}$PCA [21]	$\mathcal{T}$PCA [21]	$(1,2,...,k)$
		$(2,1)$	L_2–R$\mathcal{T}$PCA	L_2–W$\mathcal{T}$PCA	(k)
		$(1,1)$	L_1–R$\mathcal{T}$PCA	L_1–W$\mathcal{T}$PCA	$(1,2,...,k)$
	Dual	$(2,2)$	⊥ $\mathcal{T}$PCA	⊥ $\mathcal{T}$PCA	$(1,2,...,k)$
		$(2,1)$	L_2–$\mathcal{T}$DPCP	L_2–W$\mathcal{T}$DPCP	(k)
		$(1,1)$	L_1–$\mathcal{T}$DPCP	L_1–W$\mathcal{T}$DPCP	$(1,2,...,k)$

Table 1. A summary of variants of PCA, robust PCA and tangent PCA. The new PCA variants introduced in this paper are highlighted in blue. For robust variants of PCA: optimizing over $\mathcal{FL}(1,2,\ldots,k;n) = St(k,n)$ recovers L_1 and optimizing over $\mathcal{FL}(k;n) = Gr(k,n)$ recovers L_2 formulations. Optimizing for any other flag type will provide a collection of novel algorithms between L_1 and L_2 versions.

Remark 4 (Normalization). *Classical DPCA works with datasets normalized to the unit sphere. Our tangent formulations of DPCP variations do not perform this preprocessing on the tangent space.*

We summarize all the PCA methods as well as our extensions in Tab. 1.

5. Flagifying PCA and Its Robust Variants

We now re-interpret PCA in Euclidean spaces as an optimization on flags of linear subspaces. This flagification will later enable us to introduce more variants and algorithms.

Definition 12 (Flagified (weighted-)PCA (fPCA) [54]). *A (weighted-)flag of principal components is the solution to:*

$$[\![\mathbf{U}]\!]^{\star} = \underset{[\![\mathbf{U}]\!]\in\mathcal{FL}(n+1)}{\arg\max} \; \mathbb{E}_j \left[\sum_{i=1}^{k} w_{ij} \|\pi_{\mathbf{U}_i}(\mathbf{x}_j)\|_2^2 \right]. \tag{21}$$

where w_{ij} denote the weights. We refer to a weighted flag PCA algorithm optimized over $\mathcal{FL}(n_1, n_2, \ldots, n_k; n)$ as weighted − fPCA$(n_1, n_2, \ldots, n_k; n)$. *When $w_{ij} = 1 \,\forall i, j$, we recover* fPCA.

Remark 5. *The solution to Eq. (4) is only unique up to rotation, and PCA is unique (up to column signs) because we order by eigenvalues. This ordering imposes a flag structure. Interpreting this optimization problem over a flag emphasizes the nested structure of principal subspaces [13] and provides a slight loosening of the strict eigenvalue ordering scheme from PCA. Also note that the joint optimization* *over the whole flag of subspaces (instead of optimizing each subspace independently) poses a computational challenge, preventing [54] from a practical implementation. This gap is filled via a manifold optimization in [77] by characterizing the Riemannian geometry of $\mathcal{FL}(\cdot)$. We provide further details in the supplementary.*

Building off of these, we now flagify the robust variants of PCA before introducing new dimensionality reduction algorithms and moving onto the principal geodesic.

Flagified Robust (Dual-)PCA variants. To respect the nested structure of flags, we must embed the flag structure into the optimization problem. Generalized versions of robust PCA in Eq. (22) and Dual PCA in Eq. (23) change the objective function value and the space over which we optimize. We state these flagified formulations below.

Definition 13 (Flagified (Dual-)PCA). *In the sequel, we define flagified (f) RPCA / WPCA / DPCP / WDPCP:*

$$[\![\mathbf{U}]\!]^\star = \begin{cases} \underset{[\![\mathbf{U}]\!]\in\mathcal{FL}(n+1)}{\arg\max} \ \mathbb{E}_j\left[\sum_{i=1}^k \|\pi_{\mathbf{U}_i}(\mathbf{x}_j)\|_2\right], & \text{(fRPCA)}\\ \underset{[\![\mathbf{U}]\!]\in\mathcal{FL}(n+1)}{\arg\min} \ \mathbb{E}_j\left[\sum_{i=1}^k \|\mathbf{x}_j - \pi_{\mathbf{U}_i}(\mathbf{x}_j)\|_2\right], & \text{(fWPCA)}\end{cases} \tag{22}$$

$$[\![\mathbf{B}]\!]^\star = \begin{cases} \underset{[\![\mathbf{B}]\!]\in\mathcal{FL}(n+1)}{\arg\min} \ \mathbb{E}_j\left[\sum_{i=1}^k \|\pi_{\mathbf{B}_i}(\mathbf{x}_j)\|_2\right], & \text{(fDPCP)}\\ \underset{[\![\mathbf{B}]\!]\in\mathcal{FL}(n+1)}{\arg\max} \ \mathbb{E}_j\left[\sum_{i=1}^k \|\mathbf{x}_j - \pi_{\mathbf{B}_i}(\mathbf{x}_j)\|_2\right], & \text{(fWDPCP)}\end{cases} \tag{23}$$

where $\mathbf{U}_i$ and $\mathbf{B}_i$ as $\mathbf{X}_i$ is defined using Eq. (1).

Remark 6. *Formulating these flagified robust PCAs over $\mathcal{FL}(1,2,\dots,k;n)$ recovers L_1 formulations and over $\mathcal{FL}(k;n)$ recovers L_2 of robust PCA and DPCP formulations. This fact is enforced in Tab. 1.*

Inspired by Mankovich and Birdal [41], we now show how to implement these robust variants by showing equivalent optimization problems on the Stiefel manifold [18]. We start by viewing weighted fPCA in Eq. (21) as a Stiefel optimization problem in Prop. 3. For the rest of this section, we will slightly abuse notation and use $[\![\mathbf{U}]\!]$ for flags of both primal and dual principal directions, discarding $\mathbf{B}$. We will provide the necessary proofs in our supplementary material.

Proposition 3 (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*

$$\mathbf{U}^* = \underset{\mathbf{U}\in St(n_k,n)}{\arg\max} \sum_{i=1}^k \operatorname{tr}\left(\mathbf{U}^T\mathbf{X}\mathbf{W}_i\mathbf{X}^T\mathbf{U}\mathbf{I}_i\right) \tag{24}$$

PCA Variant	fRPCA / fDPCP	fWPCA / fWDPCP
Weight	$\mathbf{W}_i^+$ from Eq. (25)	$\mathbf{W}_i^-$ from Eq. (26)

Table 2. Weight matrix assignment according to the flagified robust PCA formulation.

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 Eq. (21)) as long as we restrict ourselves to a region on $\mathcal{FL}(n+1)$ and $St(n_k,n)$ where (weighted) fPCA is convex.

Sketch of the proof. Our proof, whose details are in the supp. material, closely follows [41]. □

We propose an algorithm for finding (weighted) fPCA using Stiefel Conjugate Gradient Descent (Stiefel-CGD) [22, 60] in the supplementary.

Next, we translate the flagified robust PCA optimizations in Eqs. (22) and (23) to problems over the Stiefel manifold with diagonal weight matrices

$$(\mathbf{W}_i^+([\![\mathbf{U}]\!]))_{jj} = \max\left\{\|\mathbf{U}\mathbf{I}_i\mathbf{U}^T\mathbf{x}_j\|_2,\epsilon\right\}^{-1}, \tag{25}$$

$$(\mathbf{W}_i^-([\![\mathbf{U}]\!]))_{jj} = \max\left\{\|\mathbf{x}_j - \mathbf{U}\mathbf{I}_i\mathbf{U}^T\mathbf{x}_j\|_2,\epsilon\right\}^{-1}, \tag{26}$$

chosen according to the robust fPCA optimization of concern, as outlined in Tab. 2.

Proposition 4 (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:*

$$\mathbf{U}^\star = \begin{cases} \underset{\mathbf{U}\in St(n,n_k)}{\arg\max} \ \sum_{i=1}^k \operatorname{tr}\left(\mathbf{U}^T\mathbf{P}_i^+\mathbf{U}\mathbf{I}_i\right), & \text{(fRPCA)}\\ \underset{\mathbf{U}\in St(n,n_k)}{\arg\min} \ \sum_{i=1}^k \operatorname{tr}\left(\mathbf{P}_i^- - \mathbf{U}^T\mathbf{P}_i^-\mathbf{U}\mathbf{I}_i\right), & \text{(fWPCA)}\end{cases} \tag{27}$$

$$\mathbf{U}^\star = \begin{cases} \underset{\mathbf{U}\in St(n,n_k)}{\arg\min} \ \sum_{i=1}^k \operatorname{tr}\left(\mathbf{U}^T\mathbf{P}_i^+\mathbf{U}\mathbf{I}_i\right), & \text{(fDPCP)}\\ \underset{\mathbf{U}\in St(n,n_k)}{\arg\max} \ \sum_{i=1}^k \operatorname{tr}\left(\mathbf{P}_i^- - \mathbf{U}^T\mathbf{P}_i^-\mathbf{U}\mathbf{I}_i\right) & \text{(fWDPCP)}\end{cases} \tag{28}$$

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 Tab. 2 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.

Algorithm 1: fRPCA, fWPCA, fDPCP

Input: Data $\{\mathbf{x}_j \in \mathbb{R}^n\}_{j=1}^p$, flag type $(n+1)$, $\epsilon > 0$
Output: Flagified principal directions $[\![\mathbf{U}]\!]^*$

Initialize $[\![\mathbf{U}]\!]$
while *(not converged)* **do**
  Assign weights:
  case *fRPCA or fDPCP* **do**
    Assign $\{\mathbf{W}_i^+([\![\mathbf{U}]\!])\}_{i=1}^k$ using Eq. (25)
  case *fWPCA* **do**
    Assign $\{\mathbf{W}_i^-([\![\mathbf{U}]\!])\}_{i=1}^k$ using Eq. (26)
  Update estimate:
  case *fRPCA* **do**
    $\mathbf{A}^+ \leftarrow \sum_{i=1}^k \mathbf{I}_i \mathbf{U}^T \mathbf{X} \mathbf{W}_i^+ \mathbf{X}^T$
    $\mathbf{U} \leftarrow \arg\max_{\mathbf{Z} \in St(k,n)} \mathbf{A}^+([\![\mathbf{U}]\!])\mathbf{Z}$
  case *fWPCA* **do**
    $\mathbf{A}^- \leftarrow \sum_{i=1}^k \mathbf{I}_i \mathbf{U}^T \mathbf{X} \mathbf{W}_i^- \mathbf{X}^T$
    $\mathbf{U} \leftarrow \arg\min_{\mathbf{Z} \in St(k,n)} \mathbf{A}^-([\![\mathbf{U}]\!])\mathbf{Z}$
  case *fDPCP* **do**
    Assign $\mathbf{U}$ using Eq. (28) with $\{\mathbf{W}_i^+([\![\mathbf{U}]\!])\}$
$[\![\mathbf{U}]\!]^* \leftarrow [\![\mathbf{U}]\!]$

Eqs. (27) and (28) offer natural iterative re-weighted optimization schemes on the Stiefel manifold for obtaining flagified robust PCA variants, where we calculate a weighted flagified PCA at each iteration with weights defined in Tab. 2. This is similar to [10]. We summarize these algorithms in Alg. 1. We further establish the convergence guarantee for the case of fDPCP Prop. 5 and leave other convergence results to future work. The assumption in our convergence guarantee is realistic because in the presence of real-world, noisy data, we cannot expect to recover dual principal directions that are perfectly orthogonal to the inlier data points. We leave dropping this assumption, leveraging optimizing our algorithm, and more advanced proof techniques similar to those in [2, 53] for a future study.

Proposition 5 (Convergence of Alg. 1 for fDPCP)**.** *Alg. 1 for fDPCP converges as long as* $\|\mathbf{U}\mathbf{I}_i\mathbf{U}^T\mathbf{x}_j\|_2 \geq \epsilon\ \forall i,j$ *as long as we restrict ourselves to a region on* $\mathcal{FL}(n+1)$ *and* $St(n_k, n)$ *where fDPCP is convex.*

Sketch of the proof. Similar to [7, 41, 43], we first show that an iteration of Alg. 1 decreases the fDPCP objective value and then convergence follows easily. □

Remark 7 (Flagifying PGA or Tangent-PCA)**.** *True flagification of exact PGA is a difficult task. While Eq. (15) resembles a flag structure, this is not the case for the nonlinear submanifolds in Eq. (16) [12, 23, 24]. Instead, we will focus on its tangent approximations, where we map the data to the tangent space of the mean and perform (weighted) fPCA along with fRPCA, fWPCA, and fDPCP in the tangent space, just like in* $\mathcal{T}$*PCA (remark 3).*

Remark 8 (Computational Complexity (CC))**.** *Alg. 1 for fRPCA, fWPCA, and fDPCP has a CC of* $O(N_o M)$ *and Alg. 1 in* $\mathcal{T}_\mathbf{x}\mathcal{M}$ *for fR*$\mathcal{T}$*PCA, fW*$\mathcal{T}$*PCA, and f*$\mathcal{T}$*DPCP has a CC of* $O(N_\mu p n n_k^2) + O(N_o M)$*, where* N_o *is the number of iterations of the outer loop,* N_μ *is the number of iterations of the Karcher median,* M *is the CC of Stiefel CGD,* p *is the number of points, and the flag is of type* $(n_1, n_2, \ldots, n_k; n)$*.*

Remark 9 (Flagifying Tangent (Dual)-PCA)**.** *All these flagified PCA formulations can be run in the tangent space of a manifold centroid, producing a corresponding* tangent *version. Following the same convention, we dub these f*$\mathcal{T}$*PCA, fR*$\mathcal{T}$*PCA, fW*$\mathcal{T}$*PCA, and f*$\mathcal{T}$*DPCP and propose an algorithm for their computation in the supplementary.*

Remark 10 (Even further fPCA variants)**.** *As summarized in* **??***, optimizing over* $\mathcal{FL}(1,2,\ldots,k;n)$ *recovers* L_1 *versions of robust PCA while* $\mathcal{FL}(k;n)$ *recovers* L_2 *versions, in particular RPCA, WPCA, and DPCP. Naturally, one can employ other flag types recovering robust PCAs "in between"* L_1 *and* L_2 *that differ from* L_p*,* $1 < p < 2$*. Moreover, these flagified PCA formulations can be run in the tangent space of a manifold centroid to recover tangent robust* L_1 *and* L_2 *principal directions, and even ones in between. These generalizations immediately produce a plethora of novel dimensionality reduction algorithms. While we glimpse their potential advantages in Sec. 6.1, we leave their thorough investigation for a future study.*

6. Results

Baselines. Our algorithm results in a family of novel PCA/$\mathcal{T}$PCA algorithms (*cf*. Tab. 1 in blue). We compare these to their known versions using state-of-the-art implementations. In particular, we use the bit-flipping algorithm of [45] for L_1-RPCA, the alternating scheme of [72] for L_2-WPCA, and the iteratively reweighted algorithm of [69] (DPCP-IRLS) for L_2-DPCP. Finally, we use the Pymanopt [68] implementations for Stiefel CGD and Riemannian Trust Region (RTR) methods on flag manifolds [47] to directly optimize the objectives in Prop. 4.

Implementation details. We always initialize Alg. 1 randomly and determine convergence either if we reach a maximum number of iterations (max. iters. of 50) or meet at least one of $\left|f([\![\mathbf{U}^{(m)}]\!]) - f([\![\mathbf{U}^{(m+1)}]\!])\right| < 10^{-9}$ or $d_c([\![\mathbf{U}^{(m)}]\!], [\![\mathbf{U}^{(m+1)}]\!]) < 10^{-9}$ where $d_c(\cdot,\cdot)$ is the chordal distance on $\mathcal{FL}(n+1)$ [56]. Karcher's mean/median convergence parameter is 10^{-8}, and step size is 0.05. All algorithms are run on a 2020 M1 MacBook Pro.

Outlier detection. Euclidean formulations of PCA yield the residuals of $\|\mathbf{x}_j - \mathbf{U}\mathbf{U}^T\mathbf{x}_j\|_2$ (for WPCA, RPCA, and

	L_1-RPCA		L_2-WPCA		L_2-DPCP	
	Obj.↑	Time	Obj.↓	Time	Obj.↓	Time
Baseline	**54.69**	70.66	**42.89**	**0.24**	34.83	**0.26**
Flag (Alg. 1)	54.66	**0.19**	42.92	0.45	**34.66**	0.38

Table 3. Objective function values and run times comparing fRPCA$(1, ..., k)$/ fWPCA(k)/ fDPCP(k) found with Alg. 1 to baselines L_1-RPCA/ L_2-WPCA/ L_2-DPCP respectively.

PCA) and $\|\mathbf{B}\mathbf{x}_j\|_2$ (for DPCP). To predict labels for outliers, we normalize these residuals between $[0, 1]$ and decide on a threshold during AUC computation. Non-Euclidean versions, (fW$\mathcal{T}$PCA, fR$\mathcal{T}$PCA, and $\mathcal{T}$PCA), given k flattened principal directions $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_k]$ at a base point $\mathbf{x} \in \mathcal{M}$, we compute $\pi_{\mathbf{U}}(x_j) = \mathbf{U}\mathbf{U}^T\mathbf{x}_j$ and reshape it into $\mathbf{v}_j \in \mathcal{T}_{\mathbf{x}}\mathcal{M}$. The predicted label is then obtained from the reconstruction error for $\mathbf{x}_j$ by thresholding the manifold distance $d(\mathbf{x}_j, \hat{\mathbf{x}}_j)$, where $\hat{\mathbf{x}}_j = \mathrm{Exp}_{\mathbf{x}}(\mathbf{v}_j)$. The predictions for f$\mathcal{T}$DPCP follows a slightly different scheme which uses flattened estimations for the dual principal directions in $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_k]$ and the data in the tangent space $\{\mathbf{v}_j\}$. The predicted label for point j is obtained by thresholding $\|\mathbf{B}\mathbf{v}_j\|_2$.

6.1. Evaluating Euclidean Principal Directions

Can flagified PCAs recover specified algorithms? To ensure that our robust algorithms in Alg. 1 can recover traditional, specific PCA variants, we compare fRPCA$(1, ..., k)$ to L_1-RPCA, fWPCA(k) to L_2-WPCA, and fDPCP(k) to L_2-DPCP in Tab. 3 (with 200 max. iters.) by computing the first $k = 2$ principal directions of $\{\mathbf{x}_i\}_{i=1}^{100} \in \mathbb{R}^5$ where $\mathbf{x}_i \sim \mathcal{U}[0, 1)$ is sampled uniformly. As seen, our algorithms converge to similar objective values as the baselines while L_2-WPCA and L_2-DPCP run faster than the flag versions. Yet, the novel fRPCA$(1, ..., k)$ is much faster than L_1-RPCA and fDPCP(k) converges to a more optimal objective albeit being initialized randomly, as opposed to SVD-initialization of L_2-DPCP.

Is our algorithm advantageous to direct optimization on manifolds? We compare Alg. 1 to direct optimization with Stiefel CGD and Flag RTR on data $\{\mathbf{x}_i\}_{i=1}^{30} \in \mathbb{R}^4$ where $\mathbf{x}_i \sim \mathcal{U}[0, 1)$. Fig. 1 presents run times and objective values attained when computing the first $k = 2$ principal directions via fRPCA$(1, ..., k)$, fWPCA$(1, ..., k)$, and fDPCP$(1, ..., k)$ with 20 random initializations. Our algorithms converge faster and to more optimal objective values than naive Stiefel-CGD and Flag-RTR.

Outlier detection on remote sensing data. We use the UCMercedLandUseDataset [76] with 100 inlier 'runway' and introduce outlier 'mobilehomepark' images. We use the benchmark RPCA by Candès *et al.* [11]. Results in Fig. 2 (top) indicate a slight yet consistent increase in performance using novel robust fDPCP$(1, 40)$.

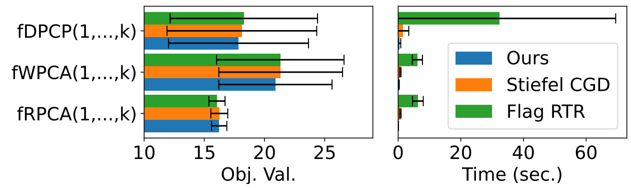

Figure 1. Alg. 1 converges faster to more optimal cost values compared to Stiefel CGD or Flag RTR.

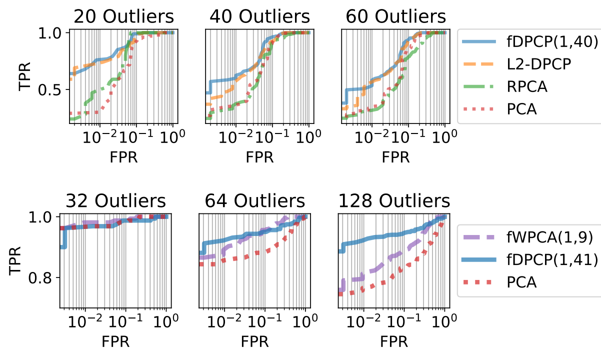

Figure 2. Average ROC curves over five trials of outlier samples for UCMercedLandUse (**top**) and YaleFaceDB-B (**bottom**). All data is reshaped and projected to $\mathbb{R}^{50}$ before outlier detection.

Outlier detection on Cropped YaleFaceDB-B. Similar to DPCP [70], we use the 64 illuminations of one face from YaleFaceDB-B [75] as inliers and introduce outliers as random images from Caltech101 [39]. Results in Fig. 2 (bottom) indicate that our robust flag methods are advantageous and the dual variant dominates as the outlier contamination increases.

6.2. Evaluating Non-Euclidean Extensions

We now evaluate flagified tangent-PCA and its robust variants starting with a synthetic evaluation of the sphere and Grassmannian before moving to real datasets. See the supplementary for additional experiments.

Convergence on 4**-sphere.** To sample a dataset of inliers and outliers on the 4-sphere $\mathbb{S}^4 = \left\{\mathbf{x} \in \mathbb{R}^5 : \|\mathbf{x}\|_2 = 1\right\}$ (see supplementary for details). Then we compute the first $k = 2$ principal directions of fR$\mathcal{T}$PCA$(1, ..., k)$, fW$\mathcal{T}$PCA$(1, ..., k)$, and f$\mathcal{T}$DPCP$(1, ..., k)$ and plot objective values as Euclidean optimizations in the tangent space of the Karcher median at each iteration of Alg. 1 in Fig. 3. All methods converge quickly, while the spread of objective function values due to initializations decreases.

Outlier detection on $Gr(2, 4)$. To compare between different flag type realizations of flagified robust PCAs, we now synthesize data with inliers and outliers on $Gr(2, 4)$, the set of all 2-planes in $\mathbb{R}^4$ represented as $Gr(2, 4) = \left\{[\mathbf{X}] : \mathbf{X} \in \mathbb{R}^{4\times 2} \text{ and } \mathbf{X}^T\mathbf{X} = \mathbf{I}\right\}$ [18]. To do so, consider two random points $[\mathbf{X}], [\mathbf{Y}] \in Gr(2, 4)$ acting as centroids for inliers and outliers, respectively. Inliers are sampled as $\mathrm{Exp}_{[\mathbf{X}]}(a\mathbf{V}_i)$ where $a \sim \mathcal{U}[0, 1)$, $\mathbf{V}_1, \mathbf{V}_2 \in$

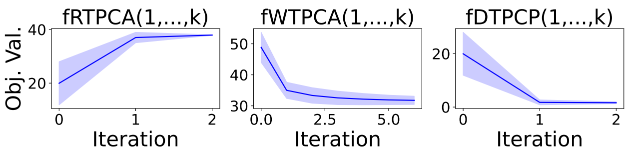

Figure 3. 50 random initializations of f$\mathcal{T}$PCA variations. The blue line is the mean and the shaded region is the standard deviation. The x-axis of this plot is the number of iterations of Alg. 1 performed in the tangent space.

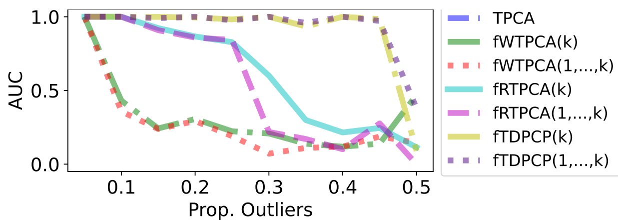

Figure 4. AUC of different algorithms for outlier detection using the first $k = 2$ principal directions of outlier-contaminated data on $Gr(2, 4)$. All iterative variants are optimized with 100 max. iters.

$\mathcal{T}_{[\mathbf{X}]}(Gr(2, 4))$ are two random tangent vectors and $\text{Exp}_{[\mathbf{X}]}$ is the exp-map of $Gr(2, 4)$. We randomly choose $i \in \{1, 2\}$. Outliers are sampled similarly as $\text{Exp}_{[\mathbf{Y}]}(b\mathbf{V})$ where $b \sim \mathcal{U}[0, 0.1)$ and gradually added to the dataset. Fig. 4 plots the AUCs for outlier detection using the first $k = 2$ principal directions. f$\mathcal{T}$DPCP$(1, ..., k)$ produces the highest AUC and is more stable to the presence of outliers. We also found that Euclidean PCA variants with the same data produce lower AUC (see supplementary).

Outlier detection and reconstruction on Kendall pre-shape space. We use an outlier-contaminated version of the 2D Hands [65] to probe the performance on a real dataset. We represent the 44 total inlier Procrustes-aligned hands and added outliers in the Kendall pre-shape space [33]: $\Sigma_2^{56} := \left\{\mathbf{X} \in \mathbb{R}^{56\times 2} : \|\mathbf{X}\|_F = 1 \text{ and } \sum_{i=1}^{56} \mathbf{x}_i = 0\right\}$. We sample outliers as open ellipses with axes sampled from $\mathcal{N}(.4, .5)$, centers from $\mathcal{N}(0, .1)$, and a hole that is $\approx 6.8\%$ of the entire ellipse. We project these outliers onto Σ_2^{56} by normalization and mean-centering. Fig. 5 reports the AUC on outlier detection as we gradually add outliers. f$\mathcal{T}$DPCP has the best outlier detections for both flag variants followed by fR$\mathcal{T}$PCA, fW$\mathcal{T}$PCA, and $\mathcal{T}$PCA with flag type $(1, 2, \dots, k)$ producing different AUC than $\mathcal{FL}(k)$ (*cf*. Tab. 1). All algorithms in these experiments are initialized with the SVD. We further consider a dataset with 30 outliers to isolate the hands (inliers). We run $\mathcal{T}$PCA with $k = 4$ principal directions to reconstruct the first hand in Fig. 6. In a slight abuse of notation, we reconstruct a hand $\mathbf{x} \in \Sigma_2^{56}$ using $k = 4$ principal tangent directions $\{\mathbf{u}_1, \dots, \mathbf{u}_4\} \in \mathcal{T}_{\boldsymbol{\mu}}\left(\Sigma_2^{56}\right)$ as $\hat{\mathbf{x}} = \text{Log}_{\mu}\left(\mathbf{U}\mathbf{U}^T\text{Exp}_{\mu}(\mathbf{x})\right)$ where $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_4]$. Since f$\mathcal{T}$DPCP$(1, ..., k)$ and f$\mathcal{T}$DPCP(k) almost perfectly detect all the outliers, they

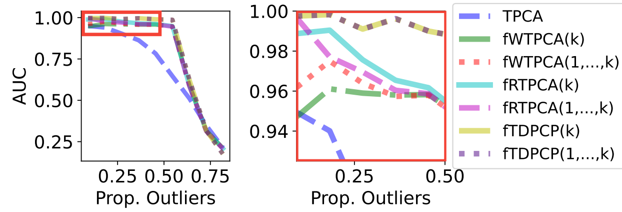

Figure 5. Mean AUC for outlier predictions using the first $k = 4$ principal directions where we gradually add outlier ellipses to the 2D Hands dataset. The mean is over 20 trials of adding outliers.

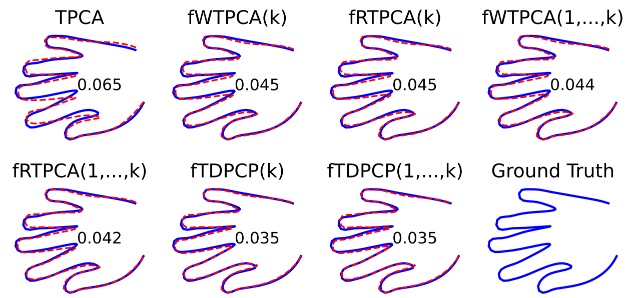

Figure 6. Reconstructions using a PCA of the inliers detected by variants of fW$\mathcal{T}$PCA, fR$\mathcal{T}$PCA, f$\mathcal{T}$DPCP and $\mathcal{T}$PCA on a the 2D Hands dataset with 28 hands and 16 ellipses. The reconstruction error is reported inside each hand.

produce the best reconstructions.

7. Conclusion

Having fun with flags, we have presented a unifying flag-manifold-based framework for computing robust principal directions of Euclidean and non-Euclidean data. Covering PCA, Dual-PCA, and their tangent versions in the same framework has given us a generalization power to develop novel, manifold-aware outlier detection and dimensionality reduction algorithms, either by modifying flag-type or by altering norms. We further devised practical algorithms on Stiefel manifolds to efficiently compute these robust directions without requiring direct optimization on the flag manifold. Our experimental evaluations revealed that new variants of robust and dual PCA/tangent PCA discovered in our framework can be useful in a variety of applications.

Limitations & future Work. We cannot handle *non-linear flags* [24] and hence cannot cover nested spheres/hyperbolic spaces [17, 19, 32]. We have also not included Barycentric subspaces *et al*. [54]. We leave these for a future study.

Acknowledgements. N. Mankovich and G. Camps-Valls acknowledge support from the project "Artificial Intelligence for complex systems: Brain, Earth, Climate, Society," funded by the Department of Innovation, Universities, Science, and Digital Society, code: CIPROM/2021/56. T. Birdal acknowledges support from the Engineering and Physical Sciences Research Council [grant EP/X011364/1].

References

[1] Michel Abboud, Abdesslam Benzinou, and Kamal Nasreddine. A robust tangent PCA via shape restoration for shape variability analysis. *Pattern Analysis and Applications*, 23: 653–671, 2020. 2

[2] Khurrum Aftab and Richard Hartley. Convergence of iteratively re-weighted least squares to robust m-estimators. In *IEEE Winter Conference on Applications of Computer Vision*, pages 480–487. IEEE, 2015. 6

[3] Khurrum Aftab, Richard Hartley, and Jochen Trumpf. Generalized Weiszfeld algorithms for lq optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(4):728–745, 2014. 3

[4] DV Alekseevsky. Flag manifolds. *Sbornik Radova*, 11, 1997. 1, 2

[5] Jerónimo Arenas-García, Kaare Brandt Petersen, Gustavo Camps-Valls, and Lars Kai Hansen. Kernel multivariate analysis framework for supervised subspace learning: A tutorial on linear and kernel multivariate methods. *IEEE Signal Processing Magazine*, 30(4):16–29, 2013. 1

[6] Suyash P Awate, Yen-Yun Yu, and Ross T Whitaker. Kernel principal geodesic analysis. In *ECML*, pages 82–98. Springer, 2014. 1, 2

[7] Amir Beck and Shoham Sabach. Weiszfeld's method: Old and new results. *Journal of Optimization Theory and Applications*, 164:1–40, 2015. 6

[8] Nicolas Boumal. *An introduction to optimization on smooth manifolds*. Cambridge University Press, 2023. 1

[9] Blanche Buet and Xavier Pennec. Flagfolds. *arXiv preprint arXiv:2305.10583*, 2023. 2

[10] Benjamin Busam, Tolga Birdal, and Nassir Navab. Camera pose filtering with local regression geodesics on the Riemannian manifold of dual quaternions. In *CVPR*, pages 2436–2445, 2017. 6

[11] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM*, 58(3):1–37, 2011. 1, 7

[12] Ioana Ciuclea, Alice Barbora Tumpach, and Cornelia Vizman. Shape spaces of nonlinear flags. In *International Conference on Geometric Science of Information*, pages 41–50. Springer, 2023. 6

[13] James Damon and JS Marron. Backwards principal component analysis and principal nested relations. *Journal of Mathematical Imaging and Vision*, 50(1-2), 2014. 4

[14] Chris Ding, Ding Zhou, Xiaofeng He, and Hongyuan Zha. R1-PCA: Rotational invariant L1-norm principal component analysis for robust subspace factorization. In *International Conference on Machine learning*, pages 281–288, 2006. 3, 4

[15] Ron Donagi and Eric Sharpe. Glsms for partial flag manifolds. *Journal of Geometry and Physics*, 58(12), 2008. 2

[16] Bruce Draper, Michael Kirby, Justin Marks, Tim Marrinan, and Chris Peterson. A flag representation for finite collections of subspaces of mixed dimensions. *Linear Algebra and its Applications*, 451:15–32, 2014. 2

[17] Ian L Dryden, Kwang-Rae Kim, Charles A Laughton, and Huiling Le. Principal nested shape space analysis of molecular dynamics data. *The Annals of Applied Statistics*, 13(4): 2213–2234, 2019. 8

[18] Alan Edelman, Tomás A Arias, and Steven T Smith. The geometry of algorithms with orthogonality constraints. *SIAM Journal on Matrix Analysis and Applications*, 20(2):303–353, 1998. 2, 5, 7

[19] Xiran Fan, Chun-Hao Yang, and Baba C Vemuri. Nested hyperbolic spaces for dimensionality reduction and hyperbolic nn design. In *CVPR*, 2022. 8

[20] P Thomas Fletcher, Conglin Lu, and Sarang Joshi. Statistics of shape via principal geodesic analysis on Lie groups. In *CVPR*. IEEE, 2003. 3

[21] P Thomas Fletcher, Conglin Lu, Stephen M Pizer, and Sarang Joshi. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions on Medical Imaging*, 23(8):995–1005, 2004. 1, 2, 3, 4

[22] William W Hager and Hongchao Zhang. A survey of nonlinear conjugate gradient methods. *Pacific Journal of Optimization*, 2(1):35–58, 2006. 5

[23] Stefan Haller and Cornelia Vizman. Nonlinear flag manifolds as coadjoint orbits. *Annals of global analysis and geometry*, 58(4):385–413, 2020. 6

[24] Stefan Haller and Cornelia Vizman. Weighted nonlinear flag manifolds as coadjoint orbits. *arXiv preprint arXiv:2301.00428*, 2023. 6, 8

[25] Mehrtash T Harandi, Mathieu Salzmann, and Richard Hartley. From manifold to manifold: Geometry-aware dimensionality reduction for SPD matrices. In *ECCV*, pages 17–32. Springer, 2014. 2

[26] Trevor Hastie and Werner Stuetzle. Principal curves. *Journal of the American Statistical Association*, 84(406), 1989. 2

[27] Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417, 1933. 1, 2, 4

[28] Long-Kai Huang and Sinno Pan. Communication-efficient distributed PCA by Riemannian optimization. In *International Conference on Machine Learning*, pages 4465–4474. PMLR, 2020.

[29] Wen Huang and Ke Wei. An extension of fast iterative shrinkage-thresholding algorithm to Riemannian optimization for sparse principal component analysis. *Numerical Linear Algebra with Applications*, 29(1):e2409, 2022. 1

[30] Stephan Huckemann and Herbert Ziezold. Principal component analysis for Riemannian manifolds, with an application to triangular shape spaces. *Advances in Applied Probability*, 38(2):299–319, 2006. 2

[31] Stephan Huckemann, Thomas Hotz, and Axel Munk. Intrinsic shape analysis: Geodesic PCA for Riemannian manifolds modulo isometric Lie group actions. *Statistica Sinica*, pages 1–58, 2010. 2

[32] Sungkyu Jung, Ian L Dryden, and James Stephen Marron. Analysis of principal nested spheres. *Biometrika*, 99(3):551–568, 2012. 8

[33] David G Kendall. Shape manifolds, procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society*, 16(2):81–121, 1984. 8

[34] Michael Kirby. *Geometric data analysis: An empirical approach to dimensionality reduction and the study of patterns*. Wiley New York, 2001. 2

[35] Nojun Kwak. Principal component analysis based on l1-norm maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(9):1672–1680, 2008. 1

[36] Nojun Kwak. Principal component analysis by $l\_\{p\}$-norm maximization. *IEEE Transactions on Cybernetics*, 44(5): 594–609, 2013. 1, 3

[37] Valero Laparra, S Jiménez, Gustavo Camps-Valls, and Jesús Malo. Nonlinearities and adaptation of color vision from sequential principal curves analysis. *Neural Computation*, 24(10):2751–2788, 2012. 1, 2

[38] John M Lee. *Riemannian manifolds: An introduction to curvature*. Springer Science & Business Media, 2006. 2

[39] Fei-Fei Li, Marco Andreetto, and Marc 'Aurelio Ranzato. Caltech101 image dataset. 2003. 7

[40] Lizhen Lin, Drew Lazar, Bayan Sarpabayeva, and David B Dunson. Robust optimization and inference on manifolds. *arXiv preprint arXiv:2006.06843*, 2020. 4

[41] Nathan Mankovich and Tolga Birdal. Chordal averaging on flag manifolds and its applications. In *ICCV*, pages 3881–3890, 2023. 2, 5, 6

[42] Nathan Mankovich, Emily J King, Chris Peterson, and Michael Kirby. The flag median and FlagIRLS. In *CVPR*, pages 10339–10347, 2022. 2

[43] Nathan J Mankovich. *Subspace and Network Averaging for Computer Vision and Bioinformatics*. PhD thesis, Colorado State University, 2023. 2, 6

[44] Panos P Markopoulos, George N Karystinos, and Dimitris A Pados. Optimal algorithms for L_1-subspace signal processing. *IEEE Transactions on Signal Processing*, 62(19):5046–5058, 2014. 1, 3, 4

[45] Panos P Markopoulos, Sandipan Kundu, Shubham Chamadia, and Dimitris A Pados. Efficient L1-norm principal-component analysis via bit flipping. *IEEE Transactions on Signal Processing*, 65(16):4252–4264, 2017. 6

[46] Sebastian Neumayer, Max Nimmer, Simon Setzer, and Gabriele Steidl. On the robust PCA and Weiszfeld's algorithm. *Applied Mathematics & Optimization*, 82(3):1017–1048, 2020. 1, 3, 4

[47] Du 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. 6

[48] Yasunori Nishimori, Shotaro Akaho, and Mark D Plumbley. Riemannian optimization method on generalized flag manifolds for complex and subspace ICA. In *AIP Conference*, 2006. 2

[49] Yasunori Nishimori, Shotaro Akaho, and Mark D Plumbley. Riemannian optimization method on the flag manifold for independent subspace analysis. In *International conference on independent component analysis and signal separation*, pages 295–302. Springer, 2006.

[50] Yasunori Nishimori, Shotaro Akaho, Samer Abdallah, and Mark D Plumbley. Flag manifolds for subspace ICA problems. In *ICASSP*, pages IV–1417. IEEE, 2007. 2

[51] Yasunori Nishimori, Shotaro Akaho, and Mark D Plumbley. Natural conjugate gradient on complex flag manifolds for complex independent subspace analysis. In *International Conference on Artificial Neural Networks*. Springer, 2008. 2

[52] Karl Pearson. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 1901. 1

[53] Liangzu Peng, Christian Kümmerle, and René Vidal. On the Convergence of IRLS and its Variants in Outlier-Robust Estimation. In *CVPR*, pages 17808–17818, 2023. 6

[54] Xavier Pennec. Barycentric subspace analysis on manifolds. *Annals of Statistics*, 46(6A), 2018. 1, 2, 4, 5, 8

[55] Xavier Pennec. Advances in geometric statistics for manifold dimension reduction. *Handbook of Variational Methods for Nonlinear Geometric Data*, pages 339–359, 2020. 1, 2

[56] Renaud-Alexandre Pitaval and Olav Tirkkonen. Flag orbit codes and their expansion to Stiefel codes. In *IEEE Information Theory Workshop*, pages 1–5. IEEE, 2013. 6

[57] Boris T Polyak and Mikhail V Khlebnikov. Robust principal component analysis: An IRLS approach. *IFAC-PapersOnLine*, 50(1):2762–2767, 2017. 1, 3, 4

[58] Dimbihery Rabenoro and Xavier Pennec. A geometric framework for asymptotic inference of principal subspaces in PCA. *arXiv preprint arXiv:2209.02025*, 2022. 2

[59] Salem Said, Nicolas Courty, Nicolas Le Bihan, and Stephen J Sangwine. Exact principal geodesic analysis for data on so (3). In *European Signal Processing Conference*, pages 1701–1705. IEEE, 2007. 4

[60] Hiroyuki Sato. Riemannian conjugate gradient methods: General framework and specific algorithms with convergence analyses. *SIAM Journal on Optimization*, 32(4):2690–2717, 2022. 5

[61] Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller. Kernel principal component analysis. In *International Conference on Artificial Neural Networks*, pages 583–588. Springer, 1997. 1

[62] Alexander Smith, Benjamin Laubach, Ivan Castillo, and Victor M Zavala. Data analysis using Riemannian geometry and applications to chemical engineering. *Computers & Chemical Engineering*, 168:108023, 2022. 2

[63] Stefan Sommer, François Lauze, Søren Hauberg, and Mads Nielsen. Manifold valued statistics, exact principal geodesic analysis and the effect of linear approximations. In *ECCV*, pages 43–56. Springer, 2010. 1, 2, 3

[64] Stefan Sommer, François Lauze, and Mads Nielsen. Optimization over geodesics for exact principal geodesic analysis. *Advances in Computational Mathematics*, 40, 2014. 2

[65] M. B. Stegmann and D. D. Gomez. A brief introduction to statistical shape analysis. page 15, 2002. 8

[66] Tom Szwagier and Xavier Pennec. Rethinking the Riemannian logarithm on flag manifolds as an orthogonal alignment problem. pages 375–383, 2023. 1

[67] Puoya Tabaghi, Michael Khanzadeh, Yusu Wang, and Sivash Mirarab. Principal component analysis in space forms. *arXiv preprint arXiv:2301.02750*, 2023. 4

[68] James Townsend, Niklas Koep, and Sebastian Weichwald. Pymanopt: A Python toolbox for optimization on man-

ifolds using automatic differentiation. *arXiv preprint arXiv:1603.03236*, 2016. 6

[69] Manolis Tsakiris and René Vidal. Dual principal component pursuit. *Journal of Machine Learning Research*, pages 1—-50, 2018. 1, 2, 3, 4, 6

[70] Manolis C. Tsakiris and Rene Vidal. Dual principal component pursuit. In *ICCV*, 2015. 7

[71] René Vidal, Yi Ma, and Shankar Sastry. Generalized principal component analysis (GPCA). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(12), 2005. 1

[72] 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. 6

[73] Sisi Wang, Feiping Nie, Zheng Wang, Rong Wang, and Xuelong Li. Max–min robust principal component analysis. *Neurocomputing*, 521:89–98, 2023. 2

[74] Mark Wiggerman. The fundamental group of a real flag manifold. *Indagationes Mathematicae*, 9(1):141–153, 1998. 2

[75] Yale. The extended yale face database b (cropped). 2001. 7

[76] Yi Yang and Shawn Newsam. Bag-of-visual-words and spatial extensions for land-use classification. In *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*, pages 270–279, 2010. 7

[77] Ke Ye, Ken Sze-Wai Wong, and Lek-Heng Lim. Optimization on flag manifolds. *Mathematical Programming*, 194(1):621–660, 2022. 1, 2, 5

[78] Teng Zhang and Yi Yang. Robust PCA by manifold optimization. *The Journal of Machine Learning Research*, 19 (1):3101–3139, 2018. 1