Fast and Efficient MMD-based Fair PCA via Optimization over Stiefel Manifold

Junghyun Lee[1], Gwangsu Kim[2], Matt Olfat[3,4], Mark Hasegawa-Johnson[5], Chang Dong Yoo[2]

[1]GSAI, KAIST [2]School of EE, KAIST [3]IEOR, UC Berkeley [4]Citadel [5]Dept. of ECE, UIUC

See [4] for the full version of our paper. Also, checkout our Github: https://github.com/nick-jhlee/fair-manifold-pca

AAAI-22

Motivation

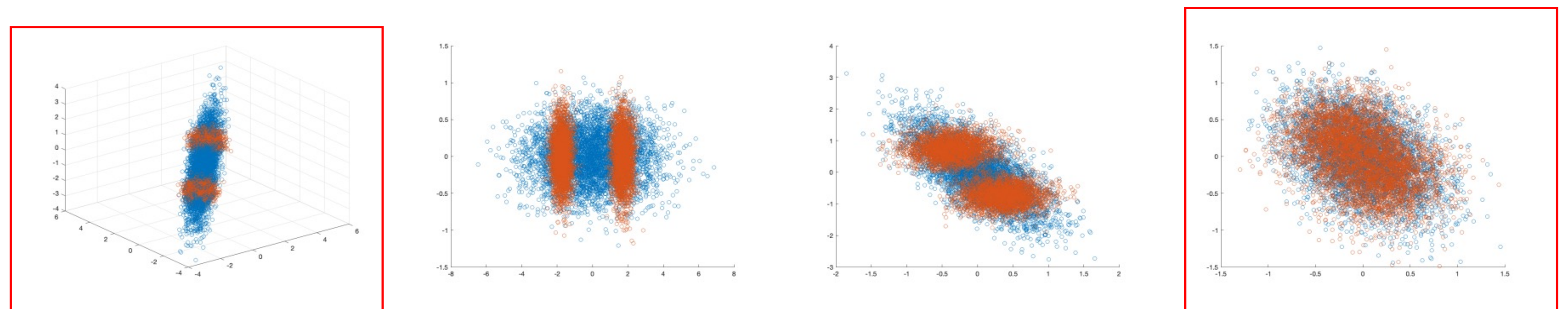
Motivated by the importance of dimensionality-reduction algorithm for dealing with high dim. datas and fairness, we propose a new definition/formulation of fair PCA.

Problem Setting: Fair PCA

- Data matrix: $X \in \mathbb{R}^{n \times p}$, its covariance matrix: $\Sigma \in \mathbb{R}^{p \times p}$,
- Projection(loading) matrix: $V \in \mathbb{R}^{p \times d}$ s.t. $d < p$, $V^T V = I_d$
- Projected data distributions of protected classes: Q_0, Q_1 (Q_0, Q_1 are *dependent on* V)

Two goals:

1. ***Reconstruction:*** Maximize the explained variance (exp. var.) after projection(%): $100 * \frac{tr(V^T \Sigma V)}{tr(\Sigma)}$
2. ***Fairness:*** Minimize the discrepancy between the projected data distributions of two protected classes



(a) Original data (b) PCA (c) FPCA (Olfat and Aswani 2019) (d) MbF-PCA (ours)

Problems with prev. approach

Previous approach [1] proposed an *adversarial* definition of fair PCA and an SDP algorithm. This has several problems:

1. *The estimator for fairness metric is **neither tractable nor asymptotically consistent**.*
2. *Exact distribution similarity (fairness) cannot be achieved as **only the first two moments are considered**.*

---

3. *The SDP is w.r.t. a $p \times p$ matrix variable P i.e. it is **inscalable to high dimensions**.*
4. *Relaxation was required, resulting in the outputted loading matrix to be **always suboptimal w.r.t. exp. var.***

New Definition of Fair PCA

Use **MMD** as our measure of discrepancy between the projected data distributions of protected classes!

1. MMD estimator is both ***tractable and consistent*** [2].
2. MMD is a ***metric***[a] for the space of probability measures.

[a]This is when the kernel used for MMD is *characteristic* (Fukumizu et al., 2008). RBF kernel is one example.

New Formulation of Fair PCA

MbF-PCA: Fair PCA as a ***constrained Riemannian optimization over the Stiefel manifold $St(p, d)$.***

- ***$St(p, d)$***: set of $V \in \mathbb{R}^{p \times d}$ with $V^T V = I_d$.
- **Constraint:** MMD of projected distributions $= 0$

$$\underset{V \in St(p,d)}{\text{minimize}} \quad f(V) := -\langle \Sigma, VV^\intercal \rangle$$

$$\text{subject to} \quad h(V) := \text{MMD}^2(Q_0, Q_1) = 0.$$

3. This optimization is w.r.t. $p \times d$ matrix variable V, and the manifold is utilized i.e. it becomes ***scalable to high dimensions***.
4. The resulting ***optimization is exact*** i.e. no relaxation.

Solving MbF-PCA

To solve MbF-PCA, we use REPMS[3], a *Riemannian counterpart for exact penalty method*. (See [4] for the pseudocode)

Under reasonable assumptions, we prove the following ***novel*** *optimality guarantees* for REPMS:

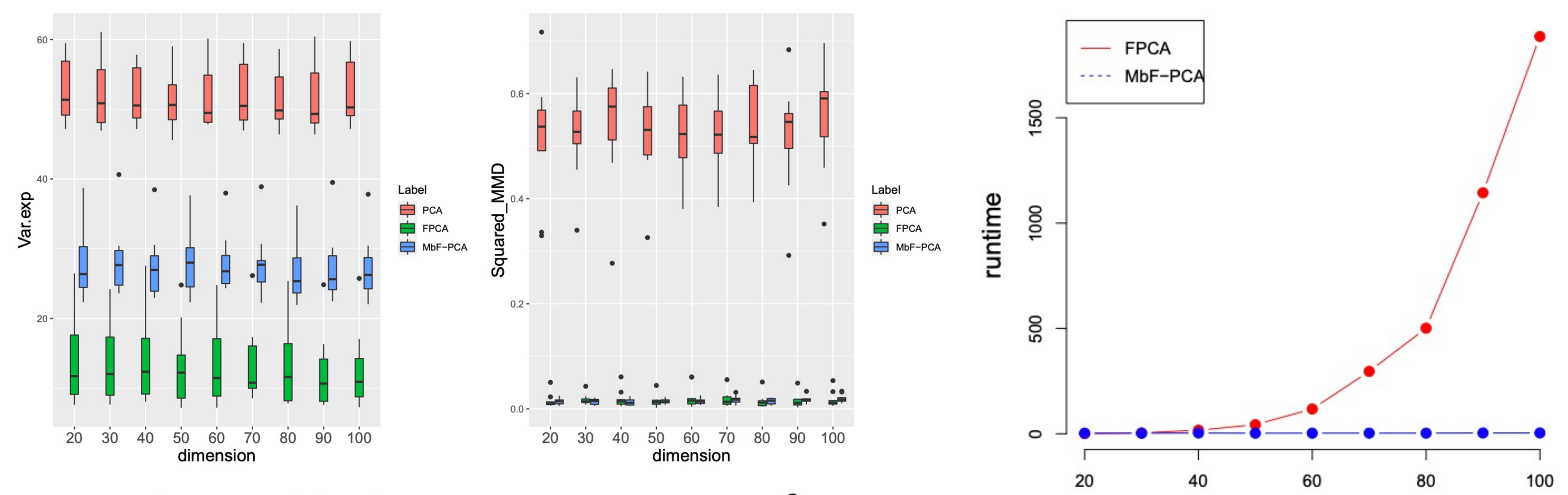
1. Under *ideal hyperparam. setting*, any fair limit point is a **local minimum**.
2. Moreover, *practical hyperparm. setting* **approximates** 1.

Compared to the analysis provided in [3], above is

- *more realistic* analysis (by considering hyperparameters)
- *stronger* optimality results (since we have local minimality)

Experiments

1. Synthetic dataset



(a) Variance explained (%) (b) MMD^2

- MbF-PCA outperforms in terms of exp. var., under similar level of fairness
- MbF-PCA is much more scalable to high dimensions

2. UCI dataset - Adult Income

- **%Var:** var. exp., **%Acc:** downstream task accuracy, Δ_{DP}: downstream task fairness[a] (demographic parity; DP)
- Reducing the dimension from 97 to 10

Algorithm (hyperparameter[b])	%Var	%Acc	MMD^2	Δ_{DP}
PCA	21.77	93.64	0.195	0.16
FPCA (0.1, 0.01)	15.75	91.94	0.006	0.13
FPCA (0, 0.01)	15.52	91.66	0.004	0.13
MbF-PCA (10^{-3})	18.71	92.81	0.005	0.14
MbF-PCA (10^{-6})	15.49	86.36	0.003	0.07

- MbF-PCA outperforms in all criteria
- MbF-PCA displays better trade-off

[a]The framework can be extended to other fairness such as equalized odds
[b]Lower hyperparameter implies tighter fairness constraint for the algorithm

References

[1] Olfat, M.; and Aswani, A. 2019. Convex Formulations for Fair Principal Component Analysis. In *AAAI*

[2] Gretton, A.; Borgwardt, K. M.; Rasch, M. J.; Scholkopf, B.; and Smola, A. 2012. A Kernel Two-Sample Test. In *JMLR*

[3] Liu, C.; and Boumal, N. Simple Algorithms for Optimization on Riemannian Manifolds with Constraints. 2019. In *Appl Math Optim*

[4] https://arxiv.org/pdf/2109.11196.pdf