A Novel Approach for Fair Principal Component Analysis Based on Eigendecomposition

Guilherme Dean Pelegrina and Leonardo Tomazeli Duarte, Senior Member, IEEE

Abstract—Principal component analysis (PCA), a ubiquitous dimensionality reduction technique in signal processing, searches for a projection matrix that minimizes the mean squared error between the reduced dataset and the original one. As the classical PCA is not tailored to address concerns related to fairness, its application to actual problems may lead to disparity in the reconstruction errors of different groups (e.g., men and women, whites and blacks, etc.), with potentially harmful consequences. For instance, in terms of quality of representation in the projected space, one may retain more information from a specific group (e.g., men) instead of another one (e.g., women), which may introduce bias towards sensitive groups. Although several fair versions of PCA have been proposed recently, there still remains a fundamental gap in the search for algorithms that are simple enough to be deployed in real systems. Moreover, the considered fairness measure does not minimize, necessarily, the reconstruction errors of different groups. To address this, we propose a novel PCA algorithm which tackles fairness issues by means of a simple strategy comprising a 1-D search which exploits the closed-form solution of PCA. As attested by numerical experiments, the proposal can significantly improve fairness, by reducing disparities in reconstruction errors, with a very small loss in the overall reconstruction error and without resorting to complex optimization schemes. Moreover, our findings are consistent in several real situations as well as in scenarios with both unbalanced and balanced datasets.

Impact Statement—Works in PCA-based fair dimensionality reduction generally propose complex algorithms which do not necessarily reduce disparities between reconstruction errors of different groups. We provide a simple one-dimensional search approach that exploits the closed-form solution of PCA and enhance fairness in dimensionality reduction. We show that our proposal minimizes the disparities between reconstruction errors. Moreover, we highlight that the development of fair methods, such as our proposal, can be useful in applications where the data have some type of bias. We also provide all codes supporting this paper that can be used to address fairness in any dimensionality reduction problem.

Index Terms—Ethical implications of artificial intelligence, machine learning, unsupervised learning.

I. INTRODUCTION

Dimensionality reduction techniques have been used in signal processing and machine learning problems in order to deal with high dimensional datasets and, therefore, to enhance data visualization and reduce the complexity of learning algorithms [11], [2], [3]. Among such techniques, one of the most used is the principal component analysis (PCA) [4], which has been applied in several problems [5], [6], [7], [8], [9]. In summary, PCA aims at reducing the dimensionality of a dataset while preserving as much information as possible from this dataset. Besides maximizing the retained information, there is also an interest in the development of methods that avoid bias when providing dimensionality reduction. Indeed, the classical PCA formulation does not take into account different sensitive groups when projecting the data. As a consequence, the reduced dataset may contain distinct representation errors for those different groups, which may introduce a bias in practical applications. Fig. 1 illustrates this scenario when projecting images into a lower dimension. If one does not take into account the presence of different groups within the dataset, one may retain more information from men instead of women. Therefore, unless strategies to mitigate bias are deployed, the application of PCA in machine learning-based systems may suffer from fairness issues.

Several recent studies have addressed bias and fairness issues in real world problems [10], [11], [12], [13]. As an example of an unfair scenario, in recidivism risk prediction [14], blacks are more likely to be classified as possible recidivists in comparison with whites. Moreover, in face recognition [15], one empirically observes better classification rates for males than females. In credit concession [16], disparities towards race have also been observed. A topic of interest when dealing with disparities is the tradeoff between model accuracy and fairness [17], [18], [19], [20], [21], [22]. In PCA-based dimensionality reduction, tradeoffs between the quality of representation and fairness have also been studied in the literature [23], [24]. A central question is how much one is willing to lose in the overall representation error in order to decrease the disparity between the sensitive groups. Aiming at reducing such a disparity, several works in the literature proposed fair PCA-based dimensionality reduction techniques [23], [24], [25], [26], [27], [28], [29], [30].

One may note two main lines of reasoning in these approaches. In the FairPCA algorithm [25] and other related approaches [24], [27], [28], fairness is measured by means of the loss suffered by each sensitive group with respect to their individual optimal projection. In the optimal scenario, one finds a projection matrix that leads to the same average loss for each sensitive group. On the other hand, in the multiojective fair principal component analysis (MOFPCA) algorithm [23], one simply
measures fairness by taking the squared difference between the averaged reconstruction errors of both sensitive groups. In this case, the fairest scenario is the one that minimizes the disparity between the groups. Moreover, the formulation in MOFPCA does not require to find $r$ projection vectors that improve fairness. It exploits the eigenvectors provided by the classical PCA and, in a different combination (i.e., not necessarily the eigenvectors associated with the $r$ highest eigenvalues), evaluates if fairness was improved without a large loss in the overall reconstruction error. The price paid by MOFPCA is that the projection vectors are restricted to the ones provided by the classical PCA.

In this article, we propose a novel approach to exploit fairness in PCA-based dimensionality reduction. Similarly as in the MOFPCA, we also consider the disparity between the reconstruction errors as a fairness measure. However, we exploit any projection matrix that improve fairness in the reduced space. For this purpose, we formulate an optimization problem whose cost function includes both overall reconstruction error and the adopted fairness measure. In order to cast this cost function as a monoobjective optimization problem, the objectives are weighted by a scalar factor. The interesting aspect is that, given a predefined weighting factor, the solution has a closed-form based on the eigenvector/eigenvalue decomposition. Therefore, a first contribution of this article consists in formulating an optimization problem that provides a compromise solution between the overall reconstruction error and the fairness measure, which depends on the adopted weighting factor and can be solved by an eigendecomposition.

Moreover, as a central concern is to reduce disparities between sensitive groups, one may investigate which weighting factor leads to the fairest scenario. Therefore, our second contribution is to propose an efficient fair PCA-based dimensionality reduction algorithm that minimizes the disparity between reconstruction errors by means of a 1-D search in closed-form solutions. By minimizing this disparity, we achieve a projected data in which both groups have similar amount of information retained. For this purpose, we formulate an optimization problem whose cost function includes both overall reconstruction error and the fairness measure, which depends on the adopted weighting factor and can be solved by an eigendecomposition.

Moreover, as a central concern is to reduce disparities between sensitive groups, one may investigate which weighting factor leads to the fairest scenario. Therefore, our second contribution is to propose an efficient fair PCA-based dimensionality reduction algorithm that minimizes the disparity between reconstruction errors by means of a 1-D search in closed-form solutions. By minimizing this disparity, we achieve a projected data in which both groups have similar amount of information retained from the original data. Note that the existing fair PCA methods either require a complex monoobjective scheme or a multiobjective scheme which is limited to the principal components which stem from the classical ones frequently used in classification tasks, such as statistical parity [31] or equalized odds [16]. However, we follow the same reasoning as in classification tasks. We consider that the model is fair when we equalize the adopted performance measure (i.e., the reconstruction error). Although fairness in dimensionality reduction can be addressed regardless a downstream task, we also investigate its effect in a credit default classification problem.

The rest of this article is organized as follows. Section II describes the classical PCA formulation and the possible disparities in dimensionality reduction tasks. In Section III, we present the proposed approach for fair dimensionality reduction. The numerical experiments as well as the obtained results are discussed in Section IV. Finally, Section V concludes the article.

II. DISPARITIES IN PRINCIPAL COMPONENT ANALYSIS

Let $X \in \mathbb{R}^{n \times d}$ denote a dataset with $n$ samples and $d$ attributes. For convenience, assume that $X$ has zero mean (otherwise, one should center the data by extracting its mean). Aiming at reducing the dimensionality of $X$ from $d$ to $r$-dimensional samples, the goal in PCA is to find a projection matrix $U \in \mathbb{R}^{d \times r}$ such that the projected data $XU$ is uncorrelated and retain as much information from $X$ as possible. PCA can be formulated by the following optimization problem:

$$\min_U \left\| X - XU^T \right\|^2_F $$

$$\text{s.t. } U^TU = I$$

where $\mathcal{R}_X(U) = \|X - XU^T\|^2_F$ represents the overall reconstruction error, $\|\cdot\|^2_F$ is the Frobenius norm [32], and $U^TU = I$, where $I$ is the identity matrix, and ensures that $U$ is orthogonal. It is easy to show that minimizing $\|X - XU^T\|^2_F$ is equivalent to maximize $\text{tr}(XX^T) - \text{tr}(U^TX^TXU)$ (or maximize $\text{tr}(U^T C_X U)$), where $C_X = \frac{1}{d} XX^T$ is the covariance matrix of $X$, as $\text{tr}(XX^T)$ is a constant) [33], [34], [35]. Moreover, because $\text{tr}(U^T X^T X U) = \|XU\|^2_F = \sum_{j=1}^r \|u_j^T X\|^2$, minimizing the reconstruction error is also equivalent to maximize the total variance of the projected data. It is also known from the literature [4] that the solution of PCA is achieved by setting the columns of $U$ as the eigenvectors of $C_X$ associated with its highest eigenvalues (see Appendix A).

![Fig. 1. Illustrative example of disparities provided by PCA.](image)
Although one expects to optimize the overall model performance (i.e., to achieve the minimum reconstruction error in PCA), there is a growing interest in the literature in the development of automatic decision systems that are also fair. We here consider that a fair algorithm should not lead to disparities between sensitive groups (e.g., males and females, whites and blacks, etc.). Therefore, a concern in the classical PCA is that it does not take into account possible disparities between different groups when projecting the dataset. For instance, assume that $X = [X_A; X_B]$, where $X_A \in \mathbb{R}^{n_A \times d}$ and $X_B \in \mathbb{R}^{n_B \times d}$ represent different sensitive groups with $n_A$ and $n_B$ samples, respectively. Without loss of generality, let us also assume that group $A$ is privileged in comparison with group $B$. This means that the application of the classical PCA in $X$ leads to a reduced dataset in which the average reconstruction error of group $A$ is lower in comparison with group $B$, i.e., $\bar{R}_{X_B}(U) = \frac{1}{n_B}R_{X_B}(U) < \frac{1}{n_A}R_{X_A}(U) = \bar{R}_{X_A}(U)$. Therefore, this disparity may harm the representation of group $B$ in comparison with group $A$ in the projected data. In order to illustrate this scenario, let us consider the example presented in Fig. 2(a), (b), and (c). One may see that, by projecting the dataset into the first principal component, the variance of group $A$ is greater than the variance of group $B$. Therefore, group $B$ has a worse representation in comparison with group $A$.

In order to avoid disparities in the reconstruction errors, it is convenient to adopt a disparity measure when performing dimensionality reduction. Here, we propose the following one:

$$D_{X_B, X_A}(U) = \frac{\|X_B - X_BUU^T\|_F^2}{n_B} - \frac{\|X_A - X_AUU^T\|_F^2}{n_A}$$

which calculates the disparity between the average reconstruction errors of the harmed group ($\bar{R}_{X_B}$) and the privileged one ($\bar{R}_{X_A}$). The fairest scenario is achieved when $D_{X_B, X_A}(U) = 0$. If one considers the example illustrated in Fig. 2(d), (e), and (f), by assuming another projection vector, the retained information for groups $A$ and $B$ are similar, which indicates that the disparity between their representations was mitigated.

III. PROPOSED FAIR PCA-BASED DIMENSIONALITY REDUCTION APPROACH

Our formulation for fair PCA consists in minimizing both overall reconstruction error and $D_{X_B, X_A}(U)$. Therefore, we deal with the following optimization problem:

$$\min_{U} \bar{R}_{X}(U) + (1 - \alpha) D_{X_B, X_A}(U)$$

s.t. $U^T U = I$
where $\mathcal{R}_X(U) = \frac{1}{n} \mathcal{R}_X(U)$ and $\alpha \in [0, 1]$ is the weighting factor that controls the importance given to each objective. Note that, if $\alpha = 1$, (3) is equivalent to the classical PCA formulation. However, the lower is the value of $\alpha$, the greater is the importance assigned to the disparity measure. Therefore, different values of $\alpha$ lead to tradeoff solutions between the fairness metric and the overall reconstruction error.

Similarly as in the classical PCA, the interesting aspect of our proposal is that the optimal solution can also be achieved by means of an eigendecomposition. This finding is stated in the following theorem:

**Theorem 1:** Assume a predefined weighting factor $\alpha$. The optimal solution of the fair principal component analysis problem expressed in (3) is given by the eigenvectors of the weighted covariance matrices $\mathcal{C} = \alpha \frac{X^T X}{n} + (1 - \alpha)\left(\frac{X^{T}_A X_A}{n_A} - \frac{X^{T}_B X_B}{n_B}\right)$ associated with its highest eigenvalues.

**Proof:** Seeing that $\alpha \mathcal{R}_X(U) + (1 - \alpha)D_{X_B, X_A}(U) = \alpha \mathcal{R}_X(U) + (1 - \alpha)(\mathcal{R}_B(U) - \mathcal{R}_A(U))$, where $\mathcal{R}_X(U) = \frac{X^T X}{n}$ and $\mathcal{R}_{X_B, X_A}(U)$, it is easy to show that

$$\begin{align*}
\alpha \mathcal{R}_X(U) + (1 - \alpha)D_{X_B, X_A}(U) &= \alpha \frac{tr(U^T X^T X U)}{n} \\
&+ (1 - \alpha)\left(\frac{tr(U^T X^T A U)}{n_A} - \frac{tr(U^T X^T B U)}{n_B}\right) \\
&= \kappa - tr\left(U^T \frac{XX^T}{n} U\right) \\
&+ tr\left(U^T \frac{(1 - \alpha) X^T A X_A}{n_A} - \frac{(1 - \alpha) X^T B X_B}{n_B}\right) \\
&= \kappa - \frac{tr(U^T \hat{C} U)}{n}
\end{align*}$$

where $\kappa = \kappa = \frac{tr(XX^T)}{n} + (1 - \alpha)\left(\frac{tr(X_B X^T B)}{n_B} - \frac{tr(X_A X^T A)}{n_A}\right)$ and $\hat{C} = \alpha \frac{X^T X}{n} + (1 - \alpha)\left(\frac{X^T A X_A}{n_A} - \frac{X^T B X_B}{n_B}\right)$. As $\kappa$ is a constant, the minimization of $\alpha \mathcal{R}_X(U) + (1 - \alpha)D_{X_B, X_A}(U)$ leads to the maximization of $tr(U^T \hat{C} U)$. Similarly as in the classical PCA, our proposal also turns to the maximization of a trace operator. Therefore, one may follow an iterative procedure to find the columns of $U$. By starting with the first principal component, we deal with the following optimization problem:

$$\begin{align*}
\max_{u_1} u_1^T \hat{C} u_1 \\
\text{s.t. } u_1^T u_1 = 1..
\end{align*}$$

We know from the classical PCA solution (see Appendix A) that $u_1$ is the eigenvector of $\hat{C}$ associated with its highest eigenvalue $\lambda_1$. For the next $r - 1$ projection vectors, we must include a constraint that ensures that $u_j^T u_{j-1} = 0$ and solve the following optimization problem for $j = 2, \ldots, r$:

$$\begin{align*}
\max_{u_j} u_j^T \hat{C} u_j \\
\text{s.t. } u_j^T u_{j-1} = 1, \quad \forall j = 2, \ldots, r.
\end{align*}$$

However, for all $j = 2, \ldots, r$, instead of assuming $u_j^T \mathcal{C} U u_{j-1} = 0$, we here assume that $u_j^T \mathcal{C} U u_{j-1} = 0$. As a consequence, we do not ensure that the projected data are uncorrelated. By following the derivation of the classical PCA, we conclude that the columns of $U$ are composed, in its columns, by the eigenvectors of $\mathcal{C}$ associated with its highest eigenvalues.

In summary, given a weighting factor $\alpha$, we attempt to reduce the disparities by solving an optimization problem very similar to the classical PCA. However, in our approach, the eigendecomposition of matrix $\mathcal{C}$ takes into account the disparity between sensitive groups. Clearly, for $\alpha = 1$, one only minimizes the overall reconstruction error and does not reduce possible disparities (i.e., one may have $\mathcal{R}_{X_A}(U) < \mathcal{R}_{X_B}(U)$). On the other hand, $\alpha = 0$ may minimize the disparity measure more than the necessary, which may invert the privileged group. In other words, one may achieve a projected data in which group $B$ has a better representation in comparison with group $A$ (i.e., $\mathcal{R}_{X_B}(U) < \mathcal{R}_{X_A}(U)$). Therefore, it is possible that there is a value of $\alpha$ that leads to $\mathcal{R}_{X_B}(U) \approx \mathcal{R}_{X_B}(U)$, i.e., that minimizes the considered fairness measure

$$F_{X_B, X_A}(U) = (\mathcal{R}_{X_B}(U) - \mathcal{R}_{X_A}(U))^2.$$

Note that (6) is the square of $D_{X_B, X_A}(U)$. Therefore, when minimizing $F_{X_B, X_A}(U)$, one reduces the disparity and avoids the inversion of the privileged group. In the sequel, we present the proposed algorithms for fair dimensionality reduction.

**A. Unconstrained Fair Dimensionality Reduction**

In view of the fact that by fixing $\alpha$ the optimization problem has a closed solution, the (unconstrained) fair PCA-based dimensionality reduction algorithm proposed in this article (called here $\nu$-FPCA) consists in a 1-D search (in $\alpha$) of eigenvectors/eigenvalues solutions aiming at minimizing the fairness measure (6). Mathematically, this formulation can be expressed as follows:

$$\min_{\alpha} \left(\frac{\|X_B - X_B U U^T\|^2}{n_B} - \frac{\|X_A - X_A U U^T\|^2}{n_A}\right)^2$$

where the columns of $U \in \mathbb{R}^{d \times r}$ are composed by the eigenvectors of $\hat{C}$ associated with its highest eigenvalues. Therefore, it is a simple and efficient approach that can be solved, e.g., by applying a golden section search algorithm [36]. Based on the dataset $X$ divided into the sensitive groups $G_1$ and $G_2$, the steps of $\nu$-FPCA are presented in Algorithm 1 (in this algorithm, eig($P$, q) represents the operator that returns the $q$ eigenvectors of $P$ associated with its $q$ highest eigenvalues). Recall that, without loss of generality, we assume that group $A$ is the privileged one when applying the classical PCA.

**B. Constrained Fair Dimensionality Reduction**

Without further considerations, the optimization problem (7) will only minimize the fairness measure. In this case, there are
Algorithm 1: (u-FPCA).

Input: Dataset \(X\), sensitive data \(X_{G_1}\) and \(X_{G_2}\), data sizes \(n, n_{G_1}\) and \(n_{G_2}\), reduced dimension \(r\) and tolerance \(tol\) for the golden section search.

Output: Fair projection matrix \(U\).

1: Compute the covariance matrix of \(X\): \(C_X \leftarrow X^T X / n\)
2: Compute PCA: \(U \leftarrow eig(C_X, r)\)
3: Compute the averaged reconstruction errors:
\[
\overline{R}_{PCA}^{A}(U) \leftarrow \frac{\|X_{A} - X_{A}UU^T\|^2_F}{n_A} \quad \text{and}
\overline{R}_{PCA}^{B}(U) \leftarrow \frac{\|X_{B} - X_{B}UU^T\|^2_F}{n_B}
\]
4: Define the privileged and harmed groups:
if \(\overline{R}_{PCA}^{A}(U) \leq \overline{R}_{PCA}^{B}(U)\) then
\[A \leftarrow G_1, n_A = n_{G_1}, B \leftarrow G_2 \text{ and } n_B = n_{G_2}\]
else
\[A \leftarrow G_2, n_A = n_{G_2}, B \leftarrow G_1 \text{ and } n_B = n_{G_1}\]
end if
5: Apply the golden section search:
Define: \(\alpha_0 \leftarrow 0, \alpha_1 \leftarrow 1 \text{ and } g_{ratio} \leftarrow (\sqrt{5} + 1)/2\)
while \(\alpha_1 - \alpha_0 \leq tol\)
Compute the candidates for \(\alpha\):
\[
\alpha_{1a} \leftarrow \alpha_0 + (\alpha_1 - \alpha_0) / g_{ratio} \quad \text{and}
\alpha_{0a} = \alpha_1 - (\alpha_1 - \alpha_0) / g_{ratio}
\]
Compute the weighted covariance matrices:
\[
\hat{C}_1 \leftarrow \alpha_{1a} C_X + (1 - \alpha_{1a})(X_{A}X_{A} - X_{B}X_{B}) \quad \text{and}
\hat{C}_0 \leftarrow \alpha_{0a} C_X + (1 - \alpha_{0a})(X_{A}X_{A} - X_{B}X_{B})
\]
Compute the candidates for \(U\): \(U_1 \leftarrow eig(\hat{C}_1, r)\) and \(U_0 \leftarrow eig(\hat{C}_0, r)\)
Compute the averaged reconstruction errors:
\[
\overline{R}_{X_A}(U_1) \leftarrow \frac{\|X_{A} - X_AUU^T\|^2_F}{n_A} \quad \text{and}
\overline{R}_{X_B}(U_0) \leftarrow \frac{\|X_{B} - X_BUU^T\|^2_F}{n_B}
\]
6: Compute the fair projection matrix: \(U \leftarrow eig(\hat{C}, r)\)

IV. Experiments

The experiments benchmark our proposals against the classical PCA and the algorithms MOFPCA [23] and FairPCA [25]. We used the following datasets and sensitive attributes (the first two were also used in [23], [25]).

1) Taiwanese Credit Default (TCRED) \[^2\] [37]: Credit history information used to predict default payments in Taiwan. It consists of 30,000 samples and 20 noncategorical attributes. We considered as sensitive attribute the education level (higher or lower, with 26,615 and 3,385 samples, respectively).

2) Labeled Faces in the Wild (LFW) \[^3\] [38]: Public benchmark of photographs frequently used for face recognition. It consists of 13,232 samples and 1,764 attributes (36 x 49 pixels). We considered as sensitive attribute the gender (female or male, with 2,962 and 10,270 samples, respectively).

3) The Law School Admissions Council’s (LSAC) \[^4\] [39]: A dataset collected from law school students that investigates ethical concerns in the bar passage rates. There are 23,726 samples and 10 attributes. We considered as sensitive attribute the race (black or white, with 1,790 and 21,936 samples, respectively).

For each one, we collected the predictive attributes and defined the sensitive one, which will be only used to split the dataset into groups \(A\) and \(B\). Therefore, we do not consider the sensitive attribute in the dimensionality reduction task. We evaluate the considered methods in scenarios in which there are unbalanced and balanced datasets with respect to the sensitive attributes. The obtained results are presented in the sequel.

\[^1\] It is worth mentioning that all codes supporting this article are available in the public repository https://github.com/GuilhermePelegrina/FPCA. The preprocessed data are available at https://github.com/GuilhermePelegrina/Datasets/tree/main/FPCA.

\[^2\] https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients.

\[^3\] http://vis-www.cs.umass.edu/lfw/.

\[^4\] http://www.seaphe.org/databases.php.
Algorithm 2: (c-FPCA).

**Input:** Dataset $X$, sensitive data $X_{G_1}$ and $X_{G_2}$, data sizes $n, n_{G_1}$ and $n_{G_2}$, reduced dimension $r$ and tolerance $tol$ for the golden section search.

**Output:** Fair projection matrix $U$.

1. Compute the covariance matrix of $X$: $C_X \leftarrow \frac{1}{n} X^T X$
2. Compute PCA: $U \leftarrow eig(C_X, r)$
3. Compute the averaged reconstruction errors:
   - $\bar{R}^PCA_{X_{G_1}}(U) \leftarrow \frac{\|X_{G_1} - X_{G_1}UU^T\|^2}{n_{G_1}}$
   - $\bar{R}^PCA_{X_{G_2}}(U) \leftarrow \frac{\|X_{G_2} - X_{G_2}UU^T\|^2}{n_{G_2}}$
4. Define the privileged and harmed groups:
   - if $\bar{R}^PCA_{X_{G_1}}(U) \leq \bar{R}^PCA_{X_{G_2}}(U)$ then
     - $A \leftarrow G_1, n_A = n_{G_1}, B \leftarrow G_2$ and $n_B = n_{G_2}$
   - else
     - $A \leftarrow G_2, n_A = n_{G_2}, B \leftarrow G_1$ and $n_B = n_{G_1}$
5. Apply the golden section search:
   - Define: $\alpha_0 \leftarrow 0$, $\alpha_1 \leftarrow 1$ and $g_{ratio} \leftarrow (\sqrt{5} + 1)/2$
   - while $\alpha_0 - \alpha_1 \leq tol$ do
     - Compute the candidates for $\alpha$:
       - $\alpha_{1a} = \alpha_0 + (\alpha_1 - \alpha_0)/g_{ratio}$ and $\alpha_{0a} = \alpha_1 - (\alpha_1 - \alpha_0)/g_{ratio}$
     - Compute the weighted covariance matrices:
       - $\bar{C}_{1} \leftarrow \alpha_{1a} C_X + (1 - \alpha_{1a}) \frac{X_{A}X_{A}}{n_A} - \frac{X_{B}X_{B}}{n_B}$ and $\bar{C}_{0} \leftarrow \alpha_{0a} C_X + (1 - \alpha_{0a}) \frac{X_{A}X_{A}}{n_A} - \frac{X_{B}X_{B}}{n_B}$
     - Compute the candidates for $U$: $U_1 \leftarrow eig(\bar{C}_1, r)$ and $U_0 \leftarrow eig(\bar{C}_0, r)$
     - Compute the averaged reconstruction errors:
       - $\bar{R}^PCA_{X_{A}}(U_1) \leftarrow \frac{\|X_{A} - X_{A}U_1U_1^T\|^2}{n_A}$
       - $\bar{R}^PCA_{X_{A}}(U_0) \leftarrow \frac{\|X_{A} - X_{A}U_0U_0^T\|^2}{n_A}$
       - $\bar{R}^PCA_{X_{B}}(U_1) \leftarrow \frac{\|X_{B} - X_{B}U_1U_1^T\|^2}{n_B}$ and $\bar{R}^PCA_{X_{B}}(U_0) \leftarrow \frac{\|X_{B} - X_{B}U_0U_0^T\|^2}{n_B}$
     - Compute the fairness measures:
       - $F_{X_{B}, X_{A}}(U_1) \leftarrow (\bar{R}^PCA_{X_{A}}(U_1) - \bar{R}^PCA_{X_{B}}(U_1))^2$ and $F_{X_{B}, X_{A}}(U_0) \leftarrow (\bar{R}^PCA_{X_{A}}(U_0) - \bar{R}^PCA_{X_{B}}(U_0))^2$
       - if $F_{X_{B}, X_{A}}(U_0) \leq F_{X_{B}, X_{A}}(U_1)$ then
         - if $\bar{R}^PCA_{X_{A}}(U_0) \leq \bar{R}^PCA_{X_{B}}(U)$ then
           - $\alpha_1 \leftarrow \alpha_{1a}$
           - $\alpha_0 \leftarrow \alpha_{0a}$
           - end if
         - else
           - $\alpha_0 \leftarrow (\alpha_1 + \alpha_0)/2$
         - end if
   - end while
6. Compute the weighted covariance matrix:
   - $C \leftarrow \alpha C_X + (1 - \alpha) \frac{X_{A}X_{A}}{n_A} - \frac{X_{B}X_{B}}{n_B}$
7. Compute the fair projection matrix: $U \leftarrow eig(C, r)$

### Table I: Computational Time in Covariance Matrices Calculations

| Dataset | Time (in seconds) |
|---------|------------------|
| TCRED   |                  |
|         |                  |
|         |                  |
|         |                  |
| LFW     |                  |
|         |                  |
|         |                  |
|         |                  |
| LSAC    |                  |
|         |                  |
|         |                  |
|         |                  |

#### A. Fair Dimensionality Reduction: Unbalanced Datasets

We first addressed scenarios with unbalanced datasets, i.e., when $n_A \neq n_B$. Fig. 3 presents the reconstruction errors and the fairness measures for different reduced dimensions. In all cases, the u-FPCA approach led to the lower values of fairness measure. However, as this approach allows an increase of the reconstruction error of both groups in order to achieve fairness, in some scenarios we obtained higher overall reconstruction error in comparison with MOFP and FairPCA [see Fig. 3(a)]. If we consider the c-FPCA approach, we achieve good values of fairness with a small loss in the overall reconstruction error in comparison with the u-FPCA approach. In other words, if we compare the reconstruction errors of groups $A$ and $B$ in different reduced dimensions, the c-FPCA approach could reduce the disparities between these values without damaging the overall reconstruction error. Therefore, by comparing with the classical PCA, we achieved reconstruction errors such that $\bar{R}^A(U), \bar{R}^B(U) \leq \bar{R}^{PCA}_B(U)$, where group $B$ is the unprivileged group. Note, in Fig. 4, that this is not always true for the u-FPCA approach.

Although the c-FPCA approach led to interesting results in fair dimension reduction problems, in the TCRED dataset the MOFP achieved similar results in terms of fairness. However, as can be seen in Fig. 4(a), it paid the same price as the u-FPCA: The reconstruction error of both groups increased. As we have already mentioned, this can be a problem in practical applications, and a projection with a better compromise between the objectives, such as the one provided by the c-FPCA proposed approach, could be a good solution for the dimensionality reduction problem. With respect to the FairPCA algorithm, most results are close (or even worst) to the classical PCA (see Fig. 3).

However, it is important to recall that the fairness measure in FairPCA is different from the one adopted in this article.

#### B. Fair Dimensionality Reduction: Balanced Datasets

In the previous experiment, we considered unbalanced datasets and showed the interesting results provided by the proposed c-FPCA approach. However, because some bias in machine learning come from unbalanced datasets, there may be a concern if our proposals only works on such scenarios. Motivated by this question, in this experiment, we verify the consistency of our proposals in balanced scenarios for each dataset. We considered the same datasets and selected a subset...
of samples\footnote{We selected $n_{\text{min}}$ samples of each group, where $n_{\text{min}}$ is the minimum between $n_A$ and $n_B$.} such that $n_A = n_B$. Therefore, there will not be an interference of unbalanced data in the dimension reduction problem.

Figs. 5 and 6 present the obtained results. Similarly as in the previous experiment, the u-FPCA approach achieved very good results in terms of fairness. However, there were some considerable loss in the overall reconstruction errors. Although it reduced the disparity, the reconstruction errors of both groups increased (see the u-FPCA results in Fig. 6). Conversely, for all datasets, the c-FPCA approach could improve fairness with a very small loss in the overall reconstruction errors. Therefore, it was consistent with the previous experiment.

If one analyzes the results provided by the MOFPCA, there was a slightly improvement in some reduced dimensions [see Fig. 5(b)]. However, the c-FPCA could further improve fairness with a lower loss in the overall reconstruction error. With respect to the FairPCA, the achievements were consistent with the previous experiment, with results very similar to the classical PCA, or even worst in some scenarios (see Fig. 5). We note such a similarity in the reconstruction errors of both sensitive groups, which are practically the same as in the application of the classical PCA (see the FairPCA results in Fig. 6).

C. On the Computational Time of the Proposed Approaches

Recall that our proposals consist in a 1-D search on closed-form solutions. As these solutions comprise the eigendecomposition of the weighted covariance matrix $\hat{C}$, they are similar to the classical PCA solution, which is given by the eigendecomposition of $C_X$. Therefore, we can safely say that, in terms of the eigendecomposition, our proposals and the classical PCA have the same computational cost as the (weighted) covariance matrices are $d \times d$. However, the difference lies in the covariance matrices calculation. While in the classical PCA one only calculates $X^T X$, in our proposals we compute $X^T \hat{X}$, $X^T_X X_A$ and $X_B^T X$. Table II presents the average time when calculating

\begin{table}[h]
\centering
\begin{tabular}{|l|c|c|}
\hline
Dataset & Time (in seconds) & \\
\hline
& u-FPCA & c-FPCA \\
\hline
TCRED & 0.9608 & 0.7830 \\
& 0.3557 & 0.3520 \\
Unbalanced & & \\
Balanced & & \\
\hline
LFW & 168.3090 & 171.4607 \\
& 96.5968 & 96.4768 \\
Unbalanced & & \\
Balanced & & \\
\hline
LSAC & 0.3044 & 0.3105 \\
& 0.0402 & 0.0365 \\
Unbalanced & & \\
Balanced & & \\
\hline
\end{tabular}
\caption{Computational Time in the Search Procedure Until Convergence.}
\end{table}
Fig. 4. Reconstruction errors of each sensitive group in unbalanced datasets. (a) TCRED dataset. (b) LFW dataset. (c) LSAC dataset.

Both \( \hat{C} \) and \( C_X \). Although the cost of \( \hat{C} \) is around twice the cost of \( C_X \), the computational time is very low even in the LFW dataset in which the number of attributes is \( d = 1764 \). Moreover, note that the higher the number of attributes (or the number of samples), the higher is the computational cost.

Another aspect with influence in the computational time of our proposals is the search procedure. While in the classical PCA one has a single eigendecomposition, in our approaches, we conduct several eigendecompositions until the convergence. Clearly, this depends on the adopted search strategy. In our experiments based on the golden section search, the computational times until the convergence for all datasets and proposed approaches\(^6\) are presented in Table I. As in the covariance matrices calculations, the higher the number of attributes (or the number of samples), the higher is the computational cost. However, even for the LFW dataset whose number of attributes are very high, our proposals could be applied in a feasible time. Therefore, we can assume that both u-FPCA and c-FPCA can be applied in high-dimensional data.

\(^6\)Computational time was practically the same regardless the reduced dimension \( r \).
D. Effect of Fair Dimensionality Reduction Into Classification Tasks

Frequently, the machine learning community is interested in classification problems. Although fairness in dimensionality reduction can be useful regardless of a downstream task, in the last experiment of this paper, we investigate the effect of achieving fair dimensionality reduction into a classification task. We consider the TCRED dataset, whose goal is to predict default payments. As classification-based fairness measures, we considered the statistical parity \( \text{SP} \), equalized odds \( \text{EO} \), and (overall) accuracy equality \( \text{AE} \). They are defined as follows (see [40] for further details):

1) **Statistical parity:** A classifier satisfies statistical parity \( \text{SP} \) if both groups have the same probability of being classified as the positive class. Mathematically, one minimizes \( f_{\text{SP}} = \frac{TP_A + FP_B}{TP_A + FP_B} - \frac{TP_B + FP_A}{TP_B + FP_A} \), where \( TP_k \) and \( FP_k \), \( k \in \{A, B\} \), are the number of true positives and false positives, respectively.

2) **Equalized odds:** A classifier satisfies equalized odds \( \text{EO} \) if both groups have equal true positive and false positive rates. Mathematically, one minimizes \( f_{\text{EO}} = \frac{TP_A}{TP_A + FN_A} - \frac{TP_B}{TP_B + FN_B} + \frac{FP_A}{FP_A + TN_A} - \frac{FP_B}{FP_B + TN_B} \), where \( TN_k \) and \( FN_k \), \( k \in \{A, B\} \), are the number of true negatives and false negatives, respectively.

3) **Accuracy equality:** A classifier satisfies accuracy equality \( \text{AE} \) if both groups achieve correct classifications for both positive and negative classes. Mathematically, one minimizes \( f_{\text{AE}} = \frac{TP_A + TN_A}{n_A} - \frac{TP_B + TN_B}{n_B} \).

We evaluate these measures based on Neural Networks\(^7\) trained from reduced datasets provided by PCA, u-FPCA, and c-FPCA. As an illustrative scenario, we considered the projection into \( r = 4 \) dimensions. Note that, in this scenario, there were disparities between the reconstruction errors [see Figs. 4(a) and 6(a)]. The results (averaged over 100 simulations) for unbalanced and balanced datasets are presented in Fig. 7(a) and (b), respectively. In both cases, we note that the proposed u-FPCA and c-FPCA achieved better results in comparison with the classical PCA. If we compare the proposed approaches, the u-FPCA leads to better results. This result is in accordance with the previous results. As the u-FPCA minimizes the disparity even with a high loss in the overall reconstruction error, it could lead to lower classification-based fairness measures in comparison with c-FPCA.

It is worth remarking that, although we consider the Neural Networks in our experiments, the proposed approach can be applied regardless the adopted regression or classification strategy. For instance, in order to reduce the dimension of the dataset before applying a deep learning model, our proposal

\[^7\]We borrowed the MLP classifier from Scikit-learn library [41] in Python and adopted the max_iter = 10\(^6\) as the maximum number on iterations. Moreover, we split the data into 80% for training and 20% for testing. We also assumed 0.5 as the classification threshold.
can be used as a preprocessing step which projects the data by retaining similar amount of information from different group of individuals.

V. CONCLUSION

In this article, we addressed the problem of fair dimensionality reduction based on principal component analysis. As the classical PCA only minimizes the overall reconstruction error of a dataset, it was not conceived in order to avoid possible disparities between sensitive groups. As a consequence, the application of such a technique may lead to a reduced dataset in which a specific group is underrepresented with respect to another one. This may create (or even increase) social bias.

In order to maximize the information retained in the dimensionality reduction while mitigating disparities between sensitive groups, we formulate an optimization problem that exploits both overall reconstruction error and fairness measure when searching for the projection matrix. We also proved that the solution of such a formulation is as simple as the solution of the classical PCA, which consists of an eigendecomposition. Moreover, we proposed two 1-D algorithms, which exploit...
eigendecomposition solutions, to achieve a fair dimensionality reduction. Therefore, our proposal can be easily deployed in any already running systems.

The experimental results in three real datasets attested that our proposal can find a projection matrix that minimizes the disparity between the sensitive groups without a large loss in the overall reconstruction error. Moreover, in contrast with other existing methods, our results were consistent in scenarios with both unbalanced and balanced data.

We also investigated the effect of our proposals into classification-based fairness measures. The obtained results suggested that addressing fairness in dimensionality reduction can help in mitigating fairness in downstream tasks. In order to further investigate this effect, future works can extend our proposal for supervised principal component analysis. In this context, the fair dimensionality reduction technique can find a projection matrix that directly improves the classification accuracy while reducing the disparity between the true/false positive/negative rates of sensitive groups. However, it is important to highlight that there are several other sources of bias that must be investigated in order to address fairness in classification tasks.

**APPENDIX**

**EIGENVECTORS AS A SOLUTION FOR PCA**

In this section, we demonstrate that the solution of PCA consists of the eigenvectors associated with the highest eigenvalues of the covariance matrix of $X$. We follow an iterative approach [4], which starts by searching for an unitary projection vector $u_1$ that maximizes the variance of the projected data $x_1 = Xu_1$. This variance is given by $\text{Var}[\tilde{x}_1] = \frac{1}{n}\tilde{x}_1^T \tilde{x}_1 = \frac{1}{n} u_1^T X^T X u_1 = u_1^T C_X u_1$, where $C_X$ is the covariance of $X$. The optimization problem is as follows:

$$\begin{align*}
\max_{u_1} & \quad u_1^T C_X u_1 \\
\text{s.t.} & \quad u_1^T u_1 = 1.
\end{align*}$$

This optimization problem can be easily solved by using the Lagrange multipliers [42], which leads to

$$\max_{u_1, \lambda_1} u_1^T C_X u_1 - \lambda_1 (u_1^T u_1 - 1).$$

By taking the gradient of this cost function, one obtains that

$$2C_X u_1 - \lambda_1 u_1 = 0 \rightarrow C_X u_1 = \lambda_1 u_1$$

where one recognizes that $u_1$ is an eigenvector of $C_X$ and $\lambda_1$ is the associated eigenvalue. Therefore, the first projection vector is the eigenvector associated with the highest eigenvalue of the covariance matrix $C_X$. Moreover, as $C_X u_1 = \lambda_1 u_1$, $\text{Var}[\tilde{x}_1] = u_1^T C_X u_1 = u_1^T \lambda_1 u_1 = \lambda_1 u_1^T u_1 = \lambda_1$, i.e., the variance in $\tilde{x}_1$ is given by the highest eigenvalue of $C_X$.

Once one has found the first principal component, one may move to the second one. Other than the constraint that ensures a unitary vector, one also needs to guarantee that the second principal component is orthogonal to the first one, i.e., $u_2^T u_1 = 0$. This leads to the following optimization problem:

$$\begin{align*}
\max_{u_2} & \quad u_2^T C_X u_2 \\
\text{s.t.} & \quad u_2^T u_2 = 1, \\
& \quad u_2^T u_1 = 0.
\end{align*}$$

One may also deal with (12) by means of the Lagrange multipliers

$$\max_{u_2, \lambda_2} u_2^T C_X u_2 - \lambda_2 (u_2^T u_2 - 1) - \phi (u_2^T u_1).$$

By taking the gradient of this cost function, one obtains that

$$2C_X u_2 - \lambda_2 u_2 - \phi u_1 = 0. \quad (14)$$

If one multiplies the aforementioned equation on the left by $u_1^T$ and assuming that $u_2^T u_1 = u_1^T u_2 = 0$, one obtains that

$$2u_1^T C_X u_2 - \lambda_2 u_1^T u_2 - \phi u_1^T u_1 = 0 \rightarrow 2u_1^T C_X u_2 = \phi u_1^T u_1. \quad (15)$$

Since we want that the projections $\tilde{x}_1 = X u_1$ and $\tilde{x}_2 = X u_2$ are uncorrelated, i.e., $\frac{1}{n} u_1^T X^T X u_2 = u_1^T C_X u_2 = 0$, this implies that $\phi u_1^T u_1 = 0$ and, therefore, $\phi$ must be equal to zero. One may also verify this condition on $\phi$ by taking (14) and, since $u_1^T C_X u_2 = u_1^T u_2$, $\phi u_1^T u_1 = 0$, $\phi = 0$. This leads to the following expression:

$$2C_X u_2 - \lambda_2 u_2 = 0 \rightarrow C_X u_2 = \lambda_2 u_2 \quad (16)$$

where one also recognizes that $u_2$ is an eigenvector of $C_X$ and $\lambda_2$ is the associated eigenvalue. In other words, the second principal component coefficients are the eigenvector associated with the second highest eigenvalue of $C_X$. Moreover, the variance in the projected data $\tilde{x}_2$ is also equivalent to the second highest eigenvalue of $C_X$.  

Fig. 7. Effect of fair dimensionality reduction into classification-based fairness measures. (a) Unbalanced dataset. (b) Balanced dataset.
For the next \( r - 2 \) principal components, the aforementioned technique can be iteratively used to find the projection matrix \( U \) whose columns are composed by the eigenvectors \( u_1, u_2, \ldots, u_r \) of \( C_X \).

REFERENCES

[1] S. Kaski and J. Peltonen, “Dimensionality reduction for data visualization [Applications corner],” IEEE Signal Process. Mag., vol. 28, no. 2, pp. 100–104, Mar. 2011.

[2] X. Huang, L. Wu, and Y. Ye, “A review on dimensionality reduction techniques,” Int. J. Pattern Recognit. Artif. Intell., vol. 33, no. 10, 2019, Art. no. 1950017.

[3] G. T. Reddy et al., “Analysis of dimensionality reduction techniques on Big Data,” IEEE Access, vol. 8, pp. 54776–54788, 2020.

[4] I. T. Jolliffe, Principal Component Analysis, 2nd ed. New York, NY, USA: Springer, 2002.

[5] M.-S. Kang, J.-H. Bae, B.-S. Kang, and K.-T. Kim, “ISAR cross-range scaling using iterative processing via principal component analysis and bisecition algorithm,” IEEE Trans. Signal Process., vol. 64, no. 15, pp. 3909–3918, Aug. 2016.

[6] H. Zhao, J. Zheng, J. Xu, and W. Deng, “Fault diagnosis method based on principal component analysis and broad learning system,” IEEE Access, vol. 7, pp. 99263–99272, 2019.

[7] C.-M. Feng, Y. Xu, J.-X. Liu, Y.-L. Gao, and C.-H. Zheng, “Supervised discriminative sparse PCA for com-characteristic gene selection and tumor classification on multiview biological data,” IEEE Trans. Neural Netw. Learn. Syst., vol. 30, no. 10, pp. 2926–2937, Oct. 2019.

[8] I. Chakraborty, D. Roy, I. Garg, A. Ankit, and K. Roy, “Constructing energy-efficient mixed-precision neural networks through principal component analysis for edge intelligence,” Nature Mach. Intell., vol. 2, no. 1, pp. 43–55, 2020.

[9] D. Rajani and P. R. Kumar, “An optimized blind wavefronting scheme based on principal component analysis in redundant discrete wavelet domain,” Signal Process., vol. 172, 2020, Art. no. 107556.

[10] S. Barocas, M. Hardt, and A. Narayanan, “Fairness in machine learning,” fairmlbook.org, 2019. [Online]. Available: http://fairmlbook.org

[11] V. Mhasawade, Y. Zhao, and R. Chumara, “Machine learning and algorithmic fairness in public and population health,” Nature Mach. Intell., vol. 3, pp. 659–666, 2021.

[12] B. M. Booth, L. Hickman, S. K. Subburaj, L. Tay, S. E. Woo, and S. K. D’Mello, “Integrating psychometrics and computing perspectives on bias and fairness in affective computing: A case study of automated video interviews,” IEEE Signal Process. Mag., vol. 38, no. 6, pp. 84–95, Nov. 2021.

[13] J. Cheong, S. Kalkan, and H. Gunes, “The Hitchhiker’s guide to bias and fairness in facial affective signal processing: Overview and techniques,” IEEE Signal Process. Mag., vol. 38, no. 6, pp. 39–49, Nov. 2021.

[14] J. Angwin, J. Larson, S. Mattu, and L. Kirchner, “Machine bias—ProPublica,” 2016. [Online]. Available: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[15] M. Ngan and P. Grother, “Face recognition vendor test (FRVT) - Performance of automated gender classification algorithms,” US Department of Commerce, National Institute of Standards and Technology (NIST), Gaithersburg, MD, USA, Interagency/Internal Report 8052, 2015.

[16] M. Hardt, E. Price, and N. Srebro, “Equality of opportunity in supervised learning,” in Proc. Adv. Neural Inf. Process. Syst., Barcelona, Spain, 2016, pp. 3315–3323.

[17] J. Kleinberg, S. Mullainathan, and M. Raghavan, “Inherent trade-offs in the fair determination of risk scores,” in Proc. 8th Innov. Theor. Comput. Sci. Conf., 2017, pp. 43:1–43:23.

[18] I. Dresel and H. Fariad, “The accuracy, fairness, and limits of predicting recidivism,” Sci. Adv., vol. 4, 2018, Art. no. eaa05580.

[19] C. Haas, “The price of fairness - A framework to explore trade-offs in algorithmic fairness,” in Proc. 40th Int. Conf. Inf. Syst., Munich, Germany, 2019, pp. 1–17.

[20] T. Zhang, T. Zhu, J. Li, M. Han, W. Zhou, and P. S. Yu, “Fairness in semi-supervised learning: Unlabeled data help to reduce discrimination,” IEEE Trans. Knowl. Data Eng., vol. 34, no. 4, pp. 1763–1774, Apr. 2022.

[21] K. T. Rodolfá, H. Lambda, and R. Ghani, “Empirical observation of negligible fairness–accuracy trade-offs in machine learning for public policy,” Nature Mach. Intell., vol. 3, no. 10, pp. 896–904, 2021.

[22] T. Zhang, T. Zhu, K. Gao, W. Zhou, and P. S. Yu, “Balancing learning model privacy, fairness, and accuracy with early stopping criteria,” IEEE Trans. Neural Netw. Learn. Syst., to be published, doi: 10.1109/TNNLS.2021.3129592.

[23] G. D. Pelegrina, R. D. B. Brotto, L. T. Duarte, R. Attux, and J. M. T. Romano, “Analysis of trade-offs in fair principal component analysis based on multi-objective optimization,” in Proc. Int. Joint Conf. Neural Netw., 2022, pp. 1–8.

[24] M. M. Kamani, F. Haddadpour, R. Forsati, and M. Mahdavi, “Efficient fair principal component analysis,” Mach. Learn., vol. 111, pp. 3671–3702, 2022.

[25] S. Samadui, U. Tantipongpipat, J. Morgenstern, M. Singh, and S. Vempala, “The price of fair PCA: One extra dimension,” in Proc. Adv. Neural Inf. Process. Syst., 2018, pp. 10976–10987.

[26] M. Olait and A. Aswani, “Convex formulations for fair principal component analysis,” in Proc. AAAI Conf. Artif. Intell., 2019, pp. 663–670.

[27] U. Tantipongpipat, S. Samadui, M. Singh, J. Morgenstern, and S. Vempala, “Multi-criteria dimensionality reduction with applications to fairness,” in Proc. Adv. Neural Inf. Process. Syst., Vancouver, Canada, 2019, pp. 15161–15171.

[28] J. Morgenstern, S. Samadui, M. Singh, U. Tantipongpipat, and S. Vempala, “Fair dimensionality reduction and iterative rounding for SDPs,” vol. 2019, 2019. [Online]. Available: http://dml.mathdoc.fr/item/1902.11281

[29] G. Zalberg and A. Wiesel, “Fair principal component analysis and filter design,” IEEE Trans. Signal Process., vol. 69, pp. 4835–4842, 2021.

[30] J. Lee, G. Kim, M. Olait, M. Hasegawa-Johnson, and C. D. Yoo, “Fast and efficient MMD-based fair PCA via optimization over Stiefel manifold,” in Proc. 36th AAAI Conf. Artif. Intell., 2022, pp. 7363–7371.

[31] C. Dwork, M. Hardt, T. Pittasi, O. Reingold, and R. Zemel, “Fairness through awareness,” in Proc. 3rd Innov. Theor. Comput. Sci. Conf., Cambridge, MA, USA, 2012, pp. 214–226.

[32] G. H. Golub and C. F. Van Loan, Matrix Computations, 4th ed. Baltimore, MD, USA: Johns Hopkins Univ. Press, 2013.

[33] F. Bro and A. K. Smilde, “Principal component analysis,” Wiley Interdiscip. Rev. Comput. Stat. Vol., vol. 3, no. 10, pp. 896–904, 2011.

[34] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, “Labeled faces in the wild: A database for studying face recognition in unconstrained environments,” in Proc. Workshop Faces ‘Real-Life’ Images: Detection, Alignment, Recognition, Marseille, France, 2008, pp. 1–14.

[35] L. F. Wightman, “LSAC national longitudinal bar passage study,” Law School Admission Council, LSAC Research Report Series, 1996. [Online]. Available: https://archive.lawschooltransparency.com/reform/projects/investigations/2015/documents/NLBPS.pdf

[36] S. Verma and J. Rubin, “Fairness definitions explained,” in Proc. Adv. Neural Inf. Process. Syst., Vancouver, Canada, 2019, pp. 15161–15171.

[37] T. Zhang, T. Zhu, K. Gao, W. Zhou, and P. S. Yu, “Balancing learning model privacy, fairness, and accuracy with early stopping criteria,” IEEE Trans. Neural Netw. Learn. Syst., to be published, doi: 10.1109/TNNLS.2021.3129592.