RESULTS AND DISCUSSION

Algorithm and analysis

In this section, we present a polynomial-time algorithm for solving the fair PCA problem. Our algorithm outputs a matrix of rank at most *d*+1 and guarantees that it achieves the fair PCA objective value equal to the optimal *d*-dimensional fair PCA value. The algorithm has two steps: first, relax fair PCA to a semidefinite optimization problem and solve the SDP; second, solve an LP designed to reduce the rank of said solution. We argue using properties of extreme point solutions that the solution must satisfy a number of constraints of the LP with equality, and argue directly that this implies the solution must lie in *d* + 1 or fewer dimensions. We refer the reader to Lau et al. [2011] for basics and applications of this technique in approximation algorithms.

Theorem 5.1. *There is a polynomial-time algorithm that outputs an approximation matrix of the data such that it is either of rank d and is an optimal solution to the fair PCA problem OR it is of rank d* + 1*, has equal losses for the two populations and achieves the optimal fair PCA objective value for dimension d.*

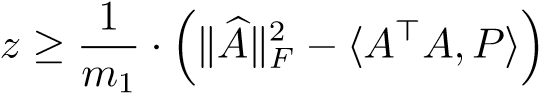
Proof of Theorem 5.1: The algorithm to prove Theorem 5.1 is presented in Algorithm 1. Using Lemma 4.7, we can write the semi-definite relaxation of the fair PCA objective (Def. 4.4) as SDP (4).

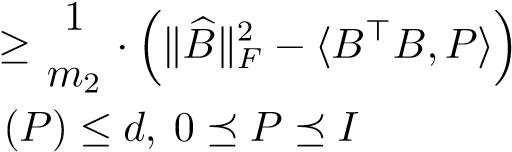
This semi-definite program can be solved in polynomial time. The system of constraints (5)-(9) is a

Algorithm 1: Fair PCA

InputOutput::*AU* 22RR*mm*1⇥⇥*nn,,B*rank2(*U*R)*m*2⇥*dn*,+ 1*d < n,m* = *m*1 + *m*2

1. Find optimal rank-*d* approximations of *A,B* as *A,B* (e.g. by Singular Value Decomposition).
2. Let (*P,*ˆ *z*ˆ) be a solution to the SDP: b b

min*P*2R*n*⇥*n*, *z*2R *z* (4) s.t.

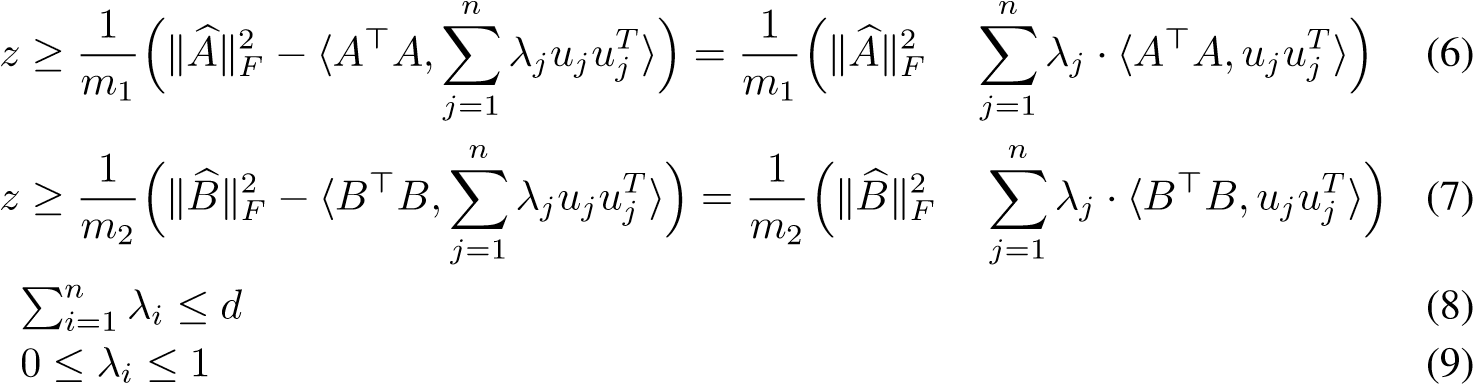
 *z*

Tr

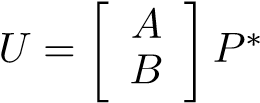
1. Apply Singular Value Decomposition to.
2. Find an extreme solution (¯*,z*⇤) of the LP:

min *z* (5)

2R*n*, *z*2R

s.t.

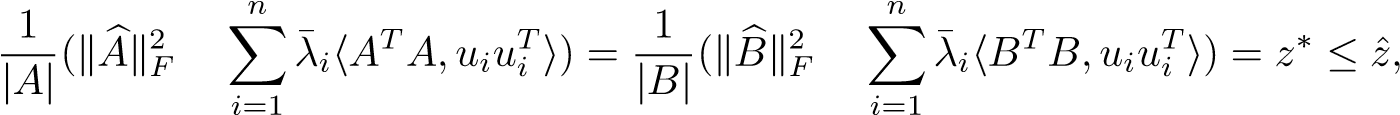
 where ⇤*j* = 1 q1 ¯*j*.

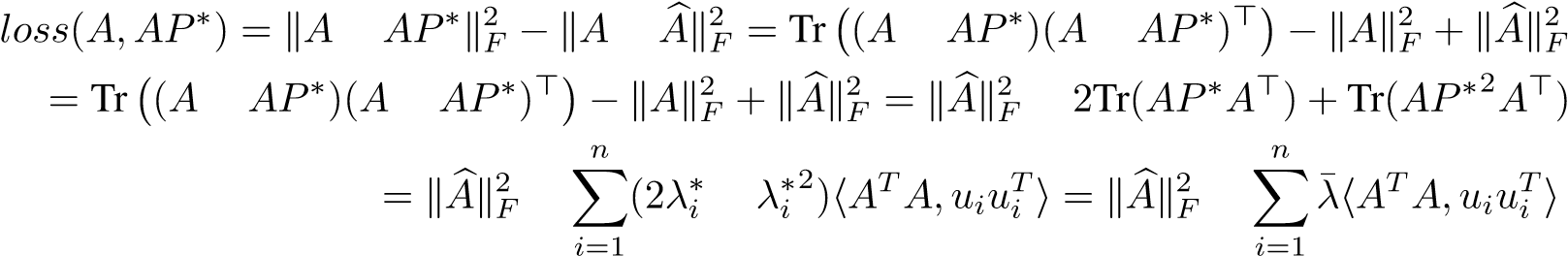
6 return

linear program in the variables *i* (with the *ui*’s fixed). Therefore, an extreme point solution (¯*,z*⇤) is defined by *n* + 1 equalities, at most three of which can be constraints in (6)-(8) and the rest (at leaston the sum of the*n* 2 of them) must be from the¯*i* ¯*i* = 0 or ¯*i* = 1 for *i* 2 [*n*]. Given the upper bound of1, i.e., at most two are*d* ’s, this implies that at least *d* 1 of them are equal to

fractional and add up to 1.

Case 1. All the eigenvalues are integral. Therefore, there are *d* eigenvalues equal to 1. This results in orthogonal projection to *d*-dimension.

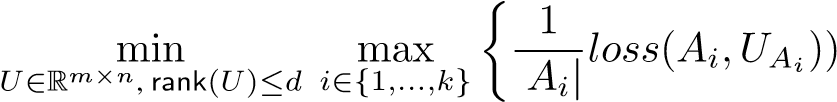
Case 2. *n* 2 of eigenvalues are in {0*,*1} and two eigenvalues 0 *<* ¯*d,* ¯*d*+1 *<* 1. Since we have *n* + 1 tight constraints, this means that both of the first two constraints are tight. Therefore where the inequality is by observing that (ˆ*,z*ˆ) is a feasible solution. Note that the loss of group *A* given by an affine projection is

*,* where the last inequality is by the choice of ⇤*j* = 1 q1 ¯*j*. The same equality holds true for group*x* ! (*x* · *uB*1*,...,x*. Therefore,· *ud* 1*,P*p⇤ gives the equal loss of⇤*d x* · *ud,*p ⇤*d*+1 *z*⇤  *z*ˆ for two groups. The embedding *x* · *ud*+1) corresponds to the affine projection of any

point (row) of *A,B* defined by the solution *P*⇤.

In both cases, the objective value is at most that of the original fairness objective. ⇤ The result of Theorem 5.1 in two groups generalizes to more than two groups as follows. Given *m* data points inprojected space, we generalize Definition 4.4 of fair PCA problem as optimizingR*n* with *k* subgroups *A*1*,A*2*,...,Ak*, and *d*  *n* the desired number of dimensions of

*,* (10)

|

where *UAi* are matrices with rows corresponding to rows of *U* for groups *Ai*.

Theorem 5.2. *There is a polynomial-time algorithm to find a projection such that it is of dimension at most d* + *k* 1 *and achieves the optimal fairness objective value for dimension d.*

In contrast to the case of two groups, when there are more than two groups in the data, it is possible that all optimal solutions to fair PCA will not assign the same loss to all groups. However, with *k* 1 extra dimensions, we can ensure that the loss of each group remains at most the optimal fairness objective in *d* dimension. The result of Theorem 5.2 follows by extending algorithm in Theorem 5.1 by adding linear constraints to SDP and LP for each extra group. An extreme solution (¯*,z*⇤) of the resulting LP contains at most *k* of *i*’s that are strictly in between 0 and 1. Therefore, the final projection matrix *P*⇤ has rank at most *d* + *k* 1.

Runtime We now analyze the runtime of Algorithm 1, which consists of solving SDP (4) and finding an extreme solution to an LP (5)-(9). The SDP and LP can be solved up to additive error of *✏ >* 0 in the objective value in *O*(*n*6*.*5 log(1*/✏*)) [Ben-Tal and Nemirovski, 2001] and *O*(*n*3*.*5 log(1*/✏*)) [Schrijver, 1998] time, respectively. The running time of SDP dominates the algorithm both in theory and practice, and is too slow for practical uses for moderate size of *n*.

We propose another algorithm of solving SDP using the multiplicative weight (MW) update method. In theory, our MW takes iterations of solving standard PCA, giving a total of runtime, which may or may not be faster than *O*(*n*6*.*5 log(1*/✏*)) depending on *n,✏*. In practice, however, we observe that after appropriately tuning one parameter in MW, the MW algorithm achieves accuracy *✏ <* 10 5 within tens of iterations, and therefore is used to obtain experimental results in this paper. Our MW can handle data of dimension up to a thousand with running time in less than a minute. The details of implementation and analysis of MW method are in Appendix A.

Experiments

We use two common human-centric data sets for our experiments. The first one is labeled faces in the wild (LFW) [Huang et al., 2007], the second is the Default Credit data set [Yeh and Lien, 2009]. We preprocess all data to have its mean at the origin. For the LFW data, we normalized each pixel value by . The gender information for LFW was taken from Afifi and Abdelhamed [2017], who manually verified the correctness of these labels. For the credit data, since different attributes are measurements of incomparable units, we normalized the variance of each attribute to be equal to 1.

Results We focus on projections into relatively few dimensions, as those are used ubiquitously in early phases of data exploration. As we already saw in Figure 1 left, at lower dimensions, there is a noticeable gap between PCA’s average reconstruction error for men and women on the LFW data set. This gap is at the scale of up to 10% of the total reconstruction error when we project to 20 dimensions. This still holds when we subsample male and female faces with equal probability from the data set, and so men and women have equal magnitude in the objective function of PCA (Figure 1 right).

Figure 3 shows the average reconstruction error of each population (Male/Female, Higher/Lower education) as the result of running vanilla PCA and Fair PCA on LFW and Credit data. As we expect,

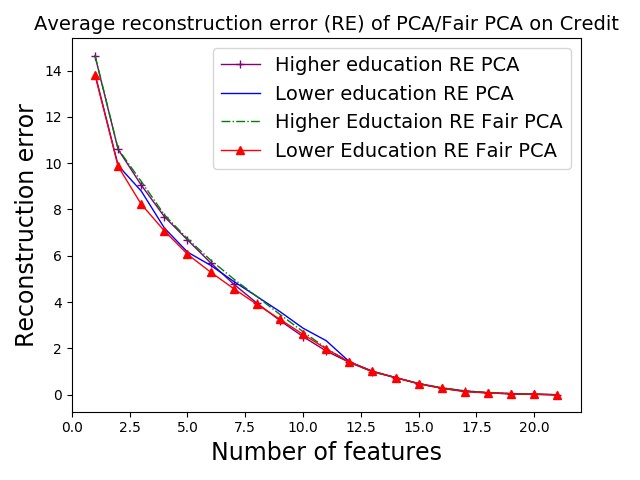
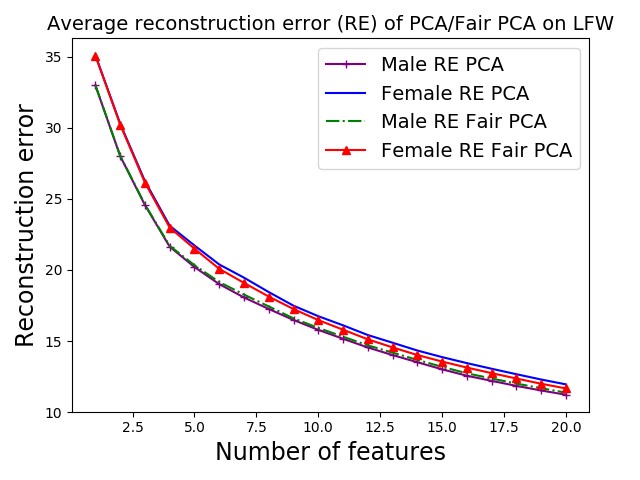


Figure 3: Reconstruction error of PCA/Fair PCA on LFW and the Default Credit data set.

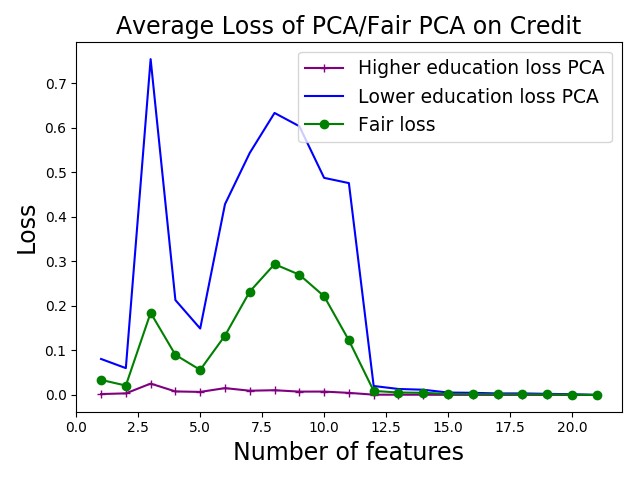
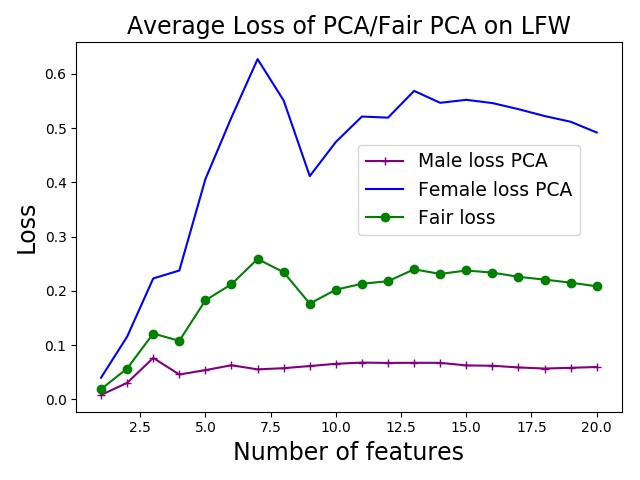


Figure 4: Loss of PCA/Fair PCA on LFW and the Default Credit data set.

as the number of dimensions increase, the average reconstruction error of every population decreases. For LFW, the original data is in 1764 dimensions (42⇥42 images), therefore, at 20 dimensions we still see a considerable reconstruction error. For the Credit data, we see that at 21 dimensions, the average reconstruction error of both populations reach 0, as this data originally lies in 21 dimensions. In order to see how fair are each of these methods, we need to zoom in further and look at the average loss of populations.

Figure 4 shows the average loss of each population as the result of applying vanilla PCA and Fair PCA on both data sets. Note that at the optimal solution of Fair PCA, the average loss of two populations are the same, therefore we have one line for “Fair loss”. We observe that PCA suffers much higher average loss for female faces than male faces. After running fair PCA, we observe that the average loss for fair PCA is relatively in the middle of the average loss for male and female. So, there is improvement in terms of the female average loss which comes with a cost in terms of male average loss. Similar observation holds for the Credit data set. In this context, it appears there is some cost to optimizing for the less well represented population in terms of the better-represented population.

Future work

This work is far from a complete study of when and how dimensionality reduction might help or hurt the fair treatment of different populations. Several concrete theoretical questions remain using our framework. What is the complexity of optimizing the fairness objective? Is it NP-hard, even for *d* = 1? Our work naturally extends to *k* predefined subgroups rather than just 2, where the number of additional dimensions our algorithm uses is *k* 1. Are these additional dimensions necessary for computational efficiency?

In a broader sense, this work aims to point out another way in which standard ML techniques might introduce unfair treatment of some subpopulation. Further work in this vein will likely prove very enlightening.