# Supplementary Material

## Proofs

### Proof of Proposition 1

The scalar-vector pair $(\widehat{\sigma}_i^2, \widehat{u}^{(i)})$ satisfies the equation $(AA^H - \widehat{\sigma}_i^2 I_{m+s})\widehat{u}^{(i)} = 0$. If we partition the $i$'th left singular vector as

$$\widehat{u}^{(i)} = \begin{pmatrix} \widehat{f}^{(i)} \\ \widehat{y}^{(i)} \end{pmatrix},$$

we can write

$$\begin{pmatrix} BB^H - \widehat{\sigma}_i^2 I_m & BE^H \\ EB^H & EE^H - \widehat{\sigma}_i^2 I_s \end{pmatrix} \begin{pmatrix} \widehat{f}^{(i)} \\ \widehat{y}^{(i)} \end{pmatrix} = 0.$$

The leading $m$ rows satisfy $(BB^H - \widehat{\sigma}_i^2 I_m)\widehat{f}^{(i)} = -BE^H\widehat{y}^{(i)}$. Plugging the expression of $\widehat{f}^{(i)}$ in the second block of rows and considering the full SVD $B = U\Sigma V^H$ leads to

$$\begin{aligned}
0 &= \left[EE^H - EB^H(BB^H - \widehat{\sigma}_i^2 I_m)^{-1}BE^H - \widehat{\sigma}_i^2 I_s\right]\widehat{y}^{(i)} \\
&= \left[E(I_s - B^H(BB^H - \widehat{\sigma}_i^2 I_m)^{-1}B)E^H - \widehat{\sigma}_i^2 I_s\right]\widehat{y}^{(i)} \\
&= \left[E(VV^H + V\Sigma^T(\widehat{\sigma}_i^2 I_m - \Sigma\Sigma^T)^{-1}\Sigma V^H)E^H - \widehat{\sigma}_i^2 I_s\right]\widehat{y}^{(i)} \\
&= \left[EV(I_n + \Sigma^T\left(\widehat{\sigma}_i^2 I_m - \Sigma\Sigma^T\right)^{-1}\Sigma)V^H E^H - \widehat{\sigma}_i^2 I_s\right]\widehat{y}^{(i)}.
\end{aligned}$$

The proof concludes by noticing that

$$I_n + \Sigma^T\left(\widehat{\sigma}_i^2 I_m - \Sigma\Sigma^T\right)^{-1}\Sigma = \begin{pmatrix} 1 + \dfrac{\sigma_1^2}{\widehat{\sigma}_i^2 - \sigma_1^2} & & \\ & \ddots & \\ & & 1 + \dfrac{\sigma_n^2}{\widehat{\sigma}_i^2 - \sigma_n^2} \end{pmatrix} = \begin{pmatrix} \dfrac{\widehat{\sigma}_i^2}{\widehat{\sigma}_i^2 - \sigma_1^2} & & \\ & \ddots & \\ & & \dfrac{\widehat{\sigma}_i^2}{\widehat{\sigma}_i^2 - \sigma_n^2} \end{pmatrix},$$

where for the case $m < n$, we have $\sigma_j = 0$ for any $j = m+1, \ldots, n$. In case $\widehat{\sigma}_i = \sigma_j$, the Moore-Penrose pseudoinverse $(BB^H - \widehat{\sigma}_i^2 I_m)^\dagger$ is considered instead.

### Proof of Proposition 2

Since the left singular vectors of $B$ span $\mathbb{R}^m$, we can write

$$BE^H\widehat{y}^{(i)} = \sum_{j=1}^{m} \sigma_j u^{(j)} \left(Ev^{(j)}\right)^H \widehat{y}^{(i)}.$$

The proof concludes by noticing that the top $m \times 1$ part of $\widehat{u}^{(i)}$ can be written as

$$\begin{aligned}
\widehat{f}^{(i)} &= -(BB^H - \widehat{\sigma}_i^2 I_m)^{-1}BE^H\widehat{y}^{(i)} \\
&= -U(\Sigma\Sigma^T - \widehat{\sigma}_i^2 I_m)^{-1}\Sigma\,(EV)^H\,\widehat{y}^{(i)} \\
&= -\sum_{j=1}^{\mathtt{min}(m,n)} u^{(j)} \frac{\sigma_j}{\sigma_j^2 - \widehat{\sigma}_i^2}\left(Ev^{(j)}\right)^H \widehat{y}^{(i)} \\
&= -\sum_{j=1}^{\mathtt{min}(m,n)} u^{(j)} \frac{\sigma_j}{\sigma_j^2 - \widehat{\sigma}_i^2}\left(Ev^{(j)}\right)^H \widehat{y}^{(i)} \\
&= \sum_{j=1}^{\mathtt{min}(m,n)} u^{(j)}\chi_{j,i}.
\end{aligned}$$

**Proof of Proposition 3**

We have

$$
\begin{aligned}
\texttt{min}_{z\in\texttt{range}(Z)}\|\widehat{u}^{(i)}-z\| &\leq \left\| \begin{pmatrix} u^{(k+1)},\ldots,u^{(\texttt{min}(m,n))} \end{pmatrix} \begin{pmatrix} \chi_{k+1,i} \\ \vdots \\ \chi_{\texttt{min}(m,n),i} \end{pmatrix} \right\| \\
&= \left\| \begin{pmatrix} 0_{k,k} & & & \\ & \dfrac{\sigma_{k+1}}{\sigma_{k+1}^2-\widehat{\sigma}_i^2} & & \\ & & \ddots & \\ & & & \dfrac{\sigma_{\texttt{min}(m,n)}}{\sigma_{\texttt{min}(m,n)}^2-\widehat{\sigma}_i^2} \end{pmatrix} V^H E^H \hat{y}^{(i)} \right\| \\
&\leq \texttt{max}\left\{\left|\frac{\sigma_j}{\sigma_j^2-\widehat{\sigma}_i^2}\right|\right\}_{j=k+1,\ldots,\texttt{min}(m,n)} \left\| E^H \widehat{y}^{(i)} \right\|.
\end{aligned}
$$

The proof follows by noticing that due to Cauchy's interlacing theorem we have $\sigma_{k+1}^2 \leq \widehat{\sigma}_i^2,\ i=1,\ldots,k$, and thus $\left|\dfrac{\sigma_{k+1}}{\sigma_{k+1}^2-\widehat{\sigma}_i^2}\right| \geq \cdots \geq \left|\dfrac{\sigma_{\texttt{min}(m,n)}}{\sigma_{\texttt{min}(m,n)}^2-\widehat{\sigma}_i^2}\right|$.

**Proof of Lemma 1**

We can write

$$
\begin{aligned}
B(\lambda) &= \left(I-U_kU_k^H\right) U \begin{pmatrix} \sigma_1^2-\lambda & & \\ & \ddots & \\ & & \sigma_m^2-\lambda \end{pmatrix}^{-1} U^H \\
&= U \begin{pmatrix} 0_{k,k} & & & \\ & \dfrac{1}{\sigma_{k+1}^2-\lambda} & & \\ & & \ddots & \\ & & & \dfrac{1}{\sigma_m^2-\lambda} \end{pmatrix} U^H,
\end{aligned}
$$

where $\sigma_j=0$ for any $j>\texttt{min}(m,n)$. Let us now define the scalar $\gamma_{j,i}=\dfrac{\widehat{\sigma}_i^2-\lambda}{\sigma_j^2-\lambda}$. Then,

$$
B(\lambda)\left[(\widehat{\sigma}_i^2-\lambda)B(\lambda)\right]^\rho = U \begin{pmatrix} 0_{k,k} & & & \\ & \dfrac{\gamma_{k+1,i}^\rho}{\sigma_{k+1}^2-\lambda} & & \\ & & \ddots & \\ & & & \dfrac{\gamma_{m,i}^\rho}{\sigma_m^2-\lambda} \end{pmatrix} U^H.
$$

Accounting for all powers $p=0,1,2,\ldots$, gives

$$
B(\lambda)\sum_{\rho=0}^{\infty}\left[(\widehat{\sigma}_i^2-\lambda)B(\lambda)\right]^\rho = U \begin{pmatrix} 0_{k,k} & & & \\ & \dfrac{\sum_{\rho=0}^\infty \gamma_{k+1,i}^\rho}{\sigma_{k+1}^2-\lambda} & & \\ & & \ddots & \\ & & & \dfrac{\sum_{\rho=0}^\infty \gamma_{m,i}^\rho}{\sigma_m^2-\lambda} \end{pmatrix} U^H.
$$

Since $\lambda > \widehat{\sigma}_k^2 \geq \sigma_k^2$, it follows that for any $j > k$ we have $|\gamma_{j,i}| < 1$. Therefore, the geometric series converges and $\sum_{\rho=0}^{\infty} \gamma_{j,i}^{\rho} = \dfrac{1}{1-\gamma_{j,i}} = \dfrac{\sigma_j^2 - \lambda}{\sigma_j^2 - \widehat{\sigma}_i^2}$. It follows that $\dfrac{1}{\sigma_j^2 - \lambda} \sum_{\rho=0}^{\infty} \gamma_j^{\rho} = \dfrac{1}{\sigma_j^2 - \widehat{\sigma}_i^2}$.

We finally have

$$
\begin{aligned}
B(\lambda) \sum_{\rho=0}^{\infty} \left[(\widehat{\sigma}_i^2 - \lambda) B(\lambda)\right]^{\rho} &= U \begin{pmatrix} 0_{k,k} & & & \\ & \dfrac{1}{\sigma_{k+1}^2 - \widehat{\sigma}_i^2} & & \\ & & \ddots & \\ & & & \dfrac{1}{\sigma_m^2 - \widehat{\sigma}_i^2} \end{pmatrix} U^H \\
&= \left(I - U_k U_k^H\right) B(\widehat{\sigma}_i^2).
\end{aligned}
$$

This concludes the proof.

**Proof of Proposition 4**

First, notice that

$$
(BB^H - \widehat{\sigma}_i^2 I_m)^{-1} = U_k U_k^H (BB^H - \widehat{\sigma}_i^2 I_m)^{-1} + (I_m - U_k U_k^H)(BB^H - \widehat{\sigma}_i^2 I_m)^{-1}.
$$

Therefore, we can write

$$
(BB^H - \widehat{\sigma}_i^2 I_m)^{-1} BE^H \widehat{y}^{(i)} = U_k(\Sigma_k^2 - \widehat{\sigma}_i^2 I_k)^{-1} \Sigma_k (EV_k)^H \widehat{y}^{(i)} + (I_m - U_k U_k^H)(BB^H - \widehat{\sigma}_i^2 I_m)^{-1} BE^H \widehat{y}^{(i)}.
$$

The left singular vector $\widehat{u}^{(i)}$ can be then expressed as

$$
\begin{aligned}
\widehat{u}^{(i)} &= \begin{pmatrix} -(BB^H - \widehat{\sigma}_i^2 I_m)^{-1} BE^H \\ I_s \end{pmatrix} \widehat{y}^{(i)} \\
&= \begin{pmatrix} u^{(1)}, \ldots, u^{(k)} & \\ & I_s \end{pmatrix} \begin{pmatrix} \chi_{1,i} \\ \vdots \\ \chi_{k,i} \\ \widehat{y}^{(i)} \end{pmatrix} - \begin{pmatrix} B(\widehat{\sigma}_i^2) BE^H \widehat{y}^{(i)} \\ \end{pmatrix}.
\end{aligned}
$$

The proof concludes by noticing that by Lemma 1 we have $B(\widehat{\sigma}_i^2) = B(\lambda) \sum\limits_{\rho=0}^{\infty} \left[(\widehat{\sigma}_i^2 - \lambda) B(\lambda)\right]^{\rho}$.

**Proof of Proposition 5**

The proof exploits the formula

$$
(B(\widehat{\sigma}_i^2) - B(\lambda)) BE^H = (I - U_k U_k^H) U \left[(\Sigma\Sigma^T - \widehat{\sigma}_i^2 I_m)^{-1} - (\Sigma\Sigma^T - \lambda I_m)^{-1}\right] U^H U \Sigma V^H E^H.
$$

It follows

$$
\begin{aligned}
\texttt{min}_{z\in\texttt{range}(Z)}\|\widehat{u}^{(i)}-z\| &\leq \left\|\begin{pmatrix}\left[B(\widehat{\sigma}_i^2)-B(\lambda)\right]BE^H\widehat{y}^{(i)}\\ \end{pmatrix}\right\| \\
&\leq \left\|\begin{pmatrix}0_{k,k} & & & \\ & \frac{\sigma_{k+1}(\widehat{\sigma}_i^2-\lambda)}{(\sigma_{k+1}^2-\widehat{\sigma}_i^2)\left(\sigma_{k+1}^2-\lambda\right)} & & \\ & & \ddots & \\ & & & \frac{\sigma_{\texttt{min}(m,n)}(\widehat{\sigma}_i^2-\lambda)}{(\sigma_{\texttt{min}(m,n)}^2-\widehat{\sigma}_i^2)\left(\sigma_{\texttt{min}(m,n)}^2-\lambda\right)}\end{pmatrix}\right\|\left\|E^H\widehat{y}^{(i)}\right\| \\
&= \texttt{max}\left\{\left|\frac{\sigma_j(\widehat{\sigma}_i^2-\lambda)}{(\sigma_j^2-\widehat{\sigma}_i^2)\left(\sigma_j^2-\lambda\right)}\right|\right\}_{j=k+1,\ldots,\texttt{min}(m,n)}\left\|E^H\widehat{y}^{(i)}\right\|.
\end{aligned}
$$

# Asymptotic complexity

The asymptotic complexity analysis of the method in (Zha & Simon, 1999) is as follows. We need $O\left(ns^2+nsk\right)$ FLOPs to form $(I_s-V_kV_k^H)E^H$ and compute its QR decomposition. The SVD of the matrix $Z^HAW$ requires $O\left((k+s)^3\right)$ FLOPs. Finally, the cost to form the approximation of matrices $\widehat{U}_k$ and $\widehat{V}_k$ is equal to $O\left(k^2(m+n)+nsk\right)$ FLOPs.

The asymptotic complexity analysis for the "SV" variant of the method in (Vecharynski & Saad, 2014) is as follows. We need $O\left((\texttt{nnz}(E)+nk)\delta_1+(n+s)\delta_1^2\right)$ FLOPs to approximate the $r$ leading singular triplets of $(I_s-V_kV_k^H)E^H$, where $\delta_1\in\mathbb{Z}^*$ is greater than or equal to $r$ (i.e., $\delta_1$ is the number of Lanczos bidiagonalization steps). The cost to form and compute the SVD of the matrix $Z^HAW$ is equal to $(k+s)(k+r)^2+\texttt{nnz}(E)k+rs$ where the first term stands for the actual SVD and the rest of the terms stand for the formation of the matrix $Z^HAW$. Finally, the cost to form the approximation of matrices $\widehat{U}_k$ and $\widehat{V}_k$ is equal to $O\left(k^2(m+n)+nrk\right)$ FLOPs.

The asymptotic complexity analysis of Algorithm 1 is as follows. First, notice that Algorithm 1 requires no effort to build $W$. For the case where $Z$ is set as in Proposition 3, termed as "Alg. 1 (a)", we also need no FLOPs to build $Z$. The cost to solve the projected problem by unrestarted Lanczos is then equal to $O\left((\texttt{nnz}(E)+nk)\delta_2+(k+s)\delta_2^2\right)$ FLOPs, where $\delta_2\in\mathbb{Z}^*$ is greater than or equal to $k$ (i.e., $\delta_2$ is the number of steps in unrestarted Lanczos). Finally, the cost to form the approximation of matrices $\widehat{U}_k$ and $\widehat{V}_k$ is equal to $O\left(k^2m+(\texttt{nnz}(A)+n)k\right)$ FLOPs. For the case where $Z$ is set as in Proposition 5, termed as "Alg. 1 (b)", we need

$$\chi=O\left(\texttt{nnz}(A)\delta_3+m\delta_3^2\right)$$

FLOPs to build $X_{\lambda,r}$, where $\delta_3\in\mathbb{Z}^*$ is greater than or equal to $k$ (i.e., $\delta_3$ is either the number of Lanczos bidiagonalization steps or the number of columns of matrix $R$ in randomized SVD).

*Table 6. Detailed asymptotic complexity of Algorithm 1 and the schemes in (Zha & Simon, 1999) and (Vecharynski & Saad, 2014). All $\delta$ variables are replaced by $k$.*

| Scheme | Building $Z$ | Building $W$ | Solving the projected problem | Other |
|---|---|---|---|---|
| (Zha & Simon, 1999) | - | $ns^2+nsk$ | $(k+s)^3$ | $k^2(m+n)+nsk$ |
| (Vecharynski & Saad, 2014) | - | $(\texttt{nnz}(E)+nk)k+(n+s)k^2$ | $(k+s)(k+r)^2+\texttt{nnz}(E)k+rs$ | $k^2(m+n)+nrk$ |
| Alg. 1 (a) | - | - | $(\texttt{nnz}(E)+nk)k+(k+s)k^2$ | $k^2m+(\texttt{nnz}(A)+n)k$ |
| Alg. 1 (b) | $\chi$ | - | $(\texttt{nnz}(E)+(n+r)k)k+(k+r+s)k^2$ | $k^2m+(\texttt{nnz}(A)+n)k$ |

The above discussion is summarized in Table 6 where we list the asymptotic complexity of Algorithm 1 and the schemes in (Zha & Simon, 1999) and (Vecharynski & Saad, 2014). The complexities of the latter two schemes were also verified by adjusting the complexity analysis from (Vecharynski & Saad, 2014). To allow for a practical comparison, we replaced all $\delta$ variables with $k$ since in practice these variables are equal to at most a small integer multiple of $k$.

Consider now a comparison between Algorithm 1 (a) and the method in (Zha & Simon, 1999). For all practical purposes, these two schemes return identical approximations to $A_k$. Nonetheless, Algorithm 1 (a) requires no effort to build $W$. Moreover, the cost to solve the projected problem is linear with respect to $s$ and cubic with respect to $k$, instead of cubic with respect to the sum $s+k$ in (Zha & Simon, 1999). The only scenario where Algorithm 1 can be potentially more expensive than (Zha & Simon, 1999) is when matrix $A$ is exceptionally dense, and both $k$ and $s$ are very small. Similar observations can be made for the relation between Algorithm 1 (b) and the methods in (Vecharynski & Saad, 2014), although the comparison is more involved.

# Eigenfaces

A brief description of the eigenfaces technique is as follows.

1. Load the training dataset consisting of $n$ images, where each image is of size $\sqrt{m} \times \sqrt{m}$ pixels.
2. Let $\widehat{A} \in \mathbb{R}^{m \times n}$ denote the matrix where each column denotes a vectorized image of size $\sqrt{m} \times \sqrt{m}$ pixels. Moreover, let $A = \widehat{A} - ze_n^T$, where $z \in \mathbb{R}^m$ denotes the column mean, and $e_n \in \mathbb{R}^n$ denotes the vector of all ones.
3. Form the covariance matrix $M = A^T A/(n-1)$, and compute its $k$ leading eigenpairs $(\lambda_i, x^{(i)}),\ i = 1, \ldots, k$. The value of $k$ is set as the smallest integer such that the explained variance $\dfrac{\lambda_1 + \ldots + \lambda_k}{\lambda_1 + \ldots + \lambda_n}$ is above a chosen threshold $\epsilon \in \mathbb{R}$. Let $X = [x^{(1)}, \ldots, x^{(k)}]$.
4. Compute the projection of the training dataset $F = \widehat{A}X$.
5. For any new test image $b \in \mathbb{R}^m$, compute its projection $\widehat{b} = X^T(b - z)$.
6. Classify the test image $b$ by $\rho$-Nearest Neighbor classification between $\widehat{b}$ and the rows of matrix $F$.

Our implementation of the eigenfaces technique replaces Step 3 as follows. Instead of computing the covariance matrix $M$, we set $k$ a-priori and compute $X$ by instead computing the $k$ leading singular triplets of $A^T$. Note that the left singular vectors of $A^T$ and the eigenvectors of $A^T A$ are the same up to sign. Instead of using a standard SVD solver, we compute the rank-$k$ truncated SVD of $A^T$ using our updating scheme. This can be especially useful for very large data collections.