Supplementary Material

Proofs

Proof of Proposition 1
The scalar-vector pair \((\tilde{\sigma}_i^2, \tilde{u}^{(i)})\) satisfies the equation \((AA^H - \tilde{\sigma}_i^2 I_{m+})\tilde{u}^{(i)} = 0\). If we partition the \(i\)th left singular vector as

\[
\tilde{u}^{(i)} = \begin{pmatrix} \tilde{f}^{(i)} \\ \tilde{y}^{(i)} \end{pmatrix},
\]

we can write

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

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

\[
0 = [EE^H - EB^H (BB^H - \tilde{\sigma}_i^2 I_m)^{-1} BE^H - \tilde{\sigma}_i^2 I_s] \tilde{y}^{(i)}
= [E(I_s - B^H (BB^H - \tilde{\sigma}_i^2 I_m)^{-1} B)E^H - \tilde{\sigma}_i^2 I_s] \tilde{y}^{(i)}
= [E(VV^H + V \Sigma^T (\tilde{\sigma}_i^2 I_m - \Sigma \Sigma^T)^{-1} \Sigma V^H)E^H - \tilde{\sigma}_i^2 I_s] \tilde{y}^{(i)}
= [EV(I_n + \Sigma^T (\tilde{\sigma}_i^2 I_m - \Sigma \Sigma^T)^{-1} \Sigma)V^H E^H - \tilde{\sigma}_i^2 I_s] \tilde{y}^{(i)}.
\]

The proof concludes by noticing that

\[
I_n + \Sigma^T (\tilde{\sigma}_i^2 I_m - \Sigma \Sigma^T)^{-1} \Sigma = \begin{pmatrix}
1 + \frac{\sigma_i^2}{\tilde{\sigma}_i^2 - \sigma_i^2} & \cdots \\
\vdots & \ddots \\
1 + \frac{\sigma_n^2}{\tilde{\sigma}_i^2 - \sigma_n^2} & \cdots \\
\end{pmatrix} \begin{pmatrix}
\frac{\tilde{\sigma}_i^2}{\tilde{\sigma}_i^2 - \sigma_i^2} \\
\cdots \\
\frac{\tilde{\sigma}_i^2}{\tilde{\sigma}_i^2 - \sigma_n^2}
\end{pmatrix},
\]

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

Proof of Proposition 2
Since the left singular vectors of \(B\) span \(\mathbb{R}^m\), we can write

\[
BE^H \tilde{y}^{(i)} = \sum_{j=1}^{m} \sigma_j u^{(j)} (Ev^{(j)})^H \tilde{y}^{(i)}.
\]

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

\[
\tilde{f}^{(i)} = -(BB^H - \tilde{\sigma}_i^2 I_m)^{-1} BE^H \tilde{y}^{(i)}
= -U(\Sigma \Sigma^T - \tilde{\sigma}_i^2 I_m)^{-1} \Sigma (EV)^H \tilde{y}^{(i)}
= -\sum_{j=1}^{\min(m,n)} u^{(j)} \frac{\sigma_j}{\sigma_j^2 - \tilde{\sigma}_i^2} (Ev^{(j)})^H \tilde{y}^{(i)}
= -\sum_{j=1}^{\min(m,n)} u^{(j)} \frac{\sigma_j}{\sigma_j^2 - \tilde{\sigma}_i^2} (Ev^{(j)})^H \tilde{y}^{(i)}
= \sum_{j=1}^{\min(m,n)} u^{(j)} \chi_{j,i},
\]
Proof of Proposition 3

We have

\[
\min_{z \in \text{range}(Z)} \| \hat{u}^{(i)} - z \| \leq \left\| \begin{pmatrix} u^{(k+1)}, \ldots, u^{(\min(m,n))} \end{pmatrix} \begin{pmatrix} \chi_{k+1,i} \\ \vdots \\ \chi_{\min(m,n),i} \end{pmatrix} \right\| \\
= \left\| \begin{pmatrix} 0_{k,k} \\ \frac{\sigma_{k+1}}{\sigma_{k+1}^2 - \bar{\sigma}_i^2} \\ \vdots \\ \frac{\sigma_{\min(m,n)}}{\sigma_{\min(m,n)}^2 - \bar{\sigma}_i^2} \end{pmatrix} V^H \hat{E}^H \hat{y}^{(i)} \right\| \\
\leq \max \left\{ \left| \frac{\sigma_j}{\sigma_j^2 - \bar{\sigma}_i^2} \right| : j = k+1, \ldots, \min(m,n) \right\} \left\| \hat{E}^H \hat{y}^{(i)} \right\|.
\]

The proof follows by noticing that due to Cauchy’s interlacing theorem we have \( \sigma_{k+1}^2 \leq \bar{\sigma}_i^2 \), \( i = 1, \ldots, k \), and thus

\[
\left| \frac{\sigma_{k+1}}{\sigma_{k+1}^2 - \bar{\sigma}_i^2} \right| \geq \cdots \geq \left| \frac{\sigma_{\min(m,n)}}{\sigma_{\min(m,n)}^2 - \bar{\sigma}_i^2} \right|.
\]

Proof of Lemma 1

We can write

\[
B(\lambda) = (I - U_k U_k^H) U \begin{pmatrix} \sigma_1^2 - \lambda \\ \vdots \\ \sigma_m^2 - \lambda \end{pmatrix}^{-1} U^H \\
= U \begin{pmatrix} 0_{k,k} \\ \frac{1}{\sigma_{k+1}^2 - \lambda} \\ \vdots \\ \frac{1}{\sigma_{m}^2 - \lambda} \end{pmatrix} U^H,
\]

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

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

Accounting for all powers \( \rho = 0, 1, 2, \ldots \) gives

\[
B(\lambda) \sum_{\rho=0}^\infty \left[ (\bar{\sigma}_i^2 - \lambda) B(\lambda) \right]^\rho = U \begin{pmatrix} 0_{k,k} \\ \frac{\sum_{\rho=0}^\infty \gamma_{k+1,i}^\rho}{\sigma_{k+1}^2 - \lambda} \\ \vdots \\ \frac{\sum_{\rho=0}^\infty \gamma_{m,i}^\rho}{\sigma_m^2 - \lambda} \end{pmatrix} U^H.
\]
Since $\lambda > \tilde{\sigma}_k^2 \geq \sigma_k^2$, it follows that for any $j > k$ we have $|\gamma_{j,i}| < 1$. Therefore, the geometric series converges and \[
abla_{\rho=0}^\infty \gamma_{j,i} = \frac{1}{1 - \gamma_{j,i}} = \frac{1}{\sigma_j^2 - \hat{\sigma}_i^2}. \] It follows that \[
abla_{\rho=0}^\infty \gamma_{j,i} = \frac{1}{\sigma_j^2 - \hat{\sigma}_i^2}. \]

We finally have \[
abla_{\rho=0}^\infty \left[ (\tilde{\sigma}_i^2 - \lambda) B(\lambda) \right] = U \begin{pmatrix}
0_{k,k} & 1 \\
\sigma_{k+1} - \hat{\sigma}_i^2 & \ddots \\
\sigma_{m} - \hat{\sigma}_i^2 \end{pmatrix} U^H \]
\[= (I - U_k U_k^H) B(\tilde{\sigma}_i^2). \]

This concludes the proof.

**Proof of Proposition 4**

First, notice that \[
(BB^H - \tilde{\sigma}_i^2 I_m)^{-1} = U_k U_k^H (BB^H - \tilde{\sigma}_i^2 I_m)^{-1} + (I_m - U_k U_k^H) (BB^H - \tilde{\sigma}_i^2 I_m)^{-1}. \]

Therefore, we can write \[
(BB^H - \tilde{\sigma}_i^2 I_m)^{-1} B E H \tilde{y}^{(i)} = U_k (\Sigma_k^2 - \tilde{\sigma}_i^2 I_k)^{-1} \Sigma_k (E V_k) H \tilde{y}^{(i)} + (I_m - U_k U_k^H) (BB^H - \tilde{\sigma}_i^2 I_m)^{-1} B E H \tilde{y}^{(i)}. \]

The left singular vector $\tilde{\sigma}^{(i)}$ can be then expressed as \[
\tilde{u}^{(i)} = \begin{pmatrix}
\Sigma_k^2 - \tilde{\sigma}_i^2 I_k \\
\Sigma_k (E V_k) H \tilde{y}^{(i)} \\
(\Sigma_k V_k^H E) H \tilde{y}^{(i)} \\
(\Sigma_k^2 - \tilde{\sigma}_i^2 I_k) - (I_m - U_k U_k^H) (BB^H - \tilde{\sigma}_i^2 I_m)^{-1} B E H \tilde{y}^{(i)} \\
\end{pmatrix}. \]

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

**Proof of Proposition 5**

The proof exploits the formula \[
(B(\tilde{\sigma}_i^2) - B(\lambda)) B E H = (I - U_k U_k^H) U \left[ (\Sigma \Sigma^T - \tilde{\sigma}_i^2 I_m)^{-1} - (\Sigma \Sigma^T - \lambda I_m)^{-1} \right] U^H \Sigma V^H E H. \]
The asymptotic complexity analysis of the method in (Zha & Simon, 1999) is as follows. We need \(O(ns^2 + nsk)\) FLOPs to form \((I_k - V_kV_k^H)E^H\) and compute its QR decomposition. The SVD of the matrix \(Z^HAW\) requires \(O((k+s)^3)\) FLOPs. Finally, the cost to form the approximation of matrices \(\hat{U}_k\) and \(\hat{V}_k\) is equal to \(O(k^2(m+n) + nsk)\) FLOPs.

The asymptotic complexity analysis for the “SV” variant of the method in (Vecharynski & Saad, 2014) is as follows. We need \(O((\text{nnz}(E) + nk)\delta_1 + (n + s)\delta_2^2)\) FLOPs to approximate the \(r\) leading singular triplets of \((I_k - V_kV_k^H)E^H\), 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 \((k+s)(k+r)^2 + \text{nnz}(E)k + rs\) 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 \(\hat{U}_k\) and \(\hat{V}_k\) is equal to \(O(k^2m + (\text{nnz}(A) + nk)k)\) FLOPs. For the case where \(Z\) is set as in Proposition 5, termed as “Alg. 1 (b)”, we need

\[
\chi = O \left( \text{nnz}(A)\delta_3 + mn_3^2 \right)
\]

FLOPs to build \(X_{\delta_3, r}\), 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 \(k\).

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Building (Z)</th>
<th>Building (W)</th>
<th>Solving the projected problem</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Zha &amp; Simon, 1999)</td>
<td>-</td>
<td>(ns^2 + nsk)</td>
<td>((k + s)^3)</td>
<td>(k^2(m+n) + nsk)</td>
</tr>
<tr>
<td>(Vecharynski &amp; Saad, 2014)</td>
<td>-</td>
<td>((\text{nnz}(E) + nk)k + (n + s)k^2)</td>
<td>((k + s)(k+r)^2 + \text{nnz}(E)k + rs)</td>
<td>(k^2(m+n) + nsk)</td>
</tr>
<tr>
<td>Alg. 1 (a)</td>
<td>-</td>
<td>-</td>
<td>((\text{nnz}(E) + nk)k + (k+s)k^2)</td>
<td>(k^2m + (\text{nnz}(A) + nk)k)</td>
</tr>
<tr>
<td>Alg. 1 (b)</td>
<td>(\chi)</td>
<td>-</td>
<td>((\text{nnz}(E) + (n + r)k)k + (k + r+s)k^2)</td>
<td>(k^2m + (\text{nnz}(A) + nk)k)</td>
</tr>
</tbody>
</table>

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 \(k\).
Projection techniques to update the truncated SVD of evolving matrices with applications

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 $\hat{A} \in \mathbb{R}^{m \times n}$ denote the matrix where each column denotes a vectorized image of size $\sqrt{m} \times \sqrt{m}$ pixels. Moreover, let $A = \hat{A} - ze_n^T$, 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 $\frac{\lambda_1 + \ldots + \lambda_k}{\lambda_1 + \ldots + \lambda_n}$ is above a chosen threshold $\epsilon \in \mathbb{R}$. Let $X = [x^{(1)}, \ldots, x^{(k)}]$.
4. Compute the projection of the training dataset $F = \hat{A}X$.
5. For any new test image $b \in \mathbb{R}^m$, compute its projection $\hat{b} = X^T (b - z)$.
6. Classify the test image $b$ by $\rho$-Nearest Neighbor classification between $\hat{b}$ 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.