Applications of SVD/PCA

Rafikul Alam
Department of Mathematics
Indian Institute of Technology Guwahati
Guwahati - 781039, INDIA

Outline

- SVD/PCA based algorithm for face recognition
- SVD and Polar decomposition

Face recognition

Figure: Library of twelve Presidents of the United States.

We store these images as vectors or matrices and apply PCA (Principal Component Analysis) to process the images.

Face recognition

Assume that each image has the same pixel resolution with m rows and n columns. Let $P_1, \ldots, P_N$ be $m \times n$ matrices representing N images. Define

$$P := \begin{bmatrix} \mathrm{Vec}(P_1) & \cdots & \mathrm{Vec}(P_N) \end{bmatrix} \in \mathbb{R}^{mn \times N},$$

Face recognition

Assume that each image has the same pixel resolution with m rows and n columns. Let $P_1, \ldots, P_N$ be $m \times n$ matrices representing N images. Define

$$P := \begin{bmatrix} \mathrm{Vec}(P_1) & \cdots & \mathrm{Vec}(P_N) \end{bmatrix} \in \mathbb{R}^{mn \times N},$$

where $\mathrm{Vec} : \mathbb{R}^{m \times n} \longrightarrow \mathbb{R}^{mn}$ converts a matrix to a column vector by stacking the columns.

Face recognition

Assume that each image has the same pixel resolution with m rows and n columns. Let $P_1, \ldots, P_N$ be $m \times n$ matrices representing N images. Define

$$P := \begin{bmatrix} \mathrm{Vec}(P_1) & \cdots & \mathrm{Vec}(P_N) \end{bmatrix} \in \mathbb{R}^{mn \times N},$$

where $\mathrm{Vec} : \mathbb{R}^{m \times n} \longrightarrow \mathbb{R}^{mn}$ converts a matrix to a column vector by stacking the columns. Consider the average image

$$\mathbf{p} := \frac{1}{N}(Pe_1 + \cdots + Pe_N) = \frac{1}{N}P\mathbf{e}, \quad \text{where } \mathbf{e} := [1, 1, \ldots, 1]^\top \in \mathbb{R}^N.$$

Face recognition

Assume that each image has the same pixel resolution with m rows and n columns. Let $P_1, \ldots, P_N$ be $m \times n$ matrices representing N images. Define

$$P := \begin{bmatrix} \mathrm{Vec}(P_1) & \cdots & \mathrm{Vec}(P_N) \end{bmatrix} \in \mathbb{R}^{mn \times N},$$

where $\mathrm{Vec} : \mathbb{R}^{m \times n} \longrightarrow \mathbb{R}^{mn}$ converts a matrix to a column vector by stacking the columns. Consider the average image

$$\mathbf{p} := \frac{1}{N}(Pe_1 + \cdots + Pe_N) = \frac{1}{N}P\mathbf{e}, \quad \text{where } \mathbf{e} := [1, 1, \ldots, 1]^\top \in \mathbb{R}^N.$$

Subtracting $\mathbf{p}$ from each columns of P, we have mean-subtracted images $A := P - \mathbf{p}\mathbf{e}^\top$.

Face recognition

Assume that each image has the same pixel resolution with m rows and n columns. Let $P_1, \ldots, P_N$ be $m \times n$ matrices representing N images. Define

$$P := \begin{bmatrix} \mathrm{Vec}(P_1) & \cdots & \mathrm{Vec}(P_N) \end{bmatrix} \in \mathbb{R}^{mn \times N},$$

where $\mathrm{Vec} : \mathbb{R}^{m \times n} \longrightarrow \mathbb{R}^{mn}$ converts a matrix to a column vector by stacking the columns. Consider the average image

$$\mathbf{p} := \frac{1}{N}(Pe_1 + \cdots + Pe_N) = \frac{1}{N}P\mathbf{e}, \quad \text{where } \mathbf{e} := [1, 1, \ldots, 1]^\top \in \mathbb{R}^N.$$

Subtracting $\mathbf{p}$ from each columns of P, we have mean-subtracted images $A := P - \mathbf{p}\mathbf{e}^\top$. Note that each row of A has zero mean. Hence $S := \frac{1}{N-1}AA^\top$ is the covariance matrix.

Face recognition

Assume that each image has the same pixel resolution with m rows and n columns. Let $P_1, \ldots, P_N$ be $m \times n$ matrices representing N images. Define

$$P := \begin{bmatrix} \mathrm{Vec}(P_1) & \cdots & \mathrm{Vec}(P_N) \end{bmatrix} \in \mathbb{R}^{mn \times N},$$

where $\mathrm{Vec} : \mathbb{R}^{m \times n} \longrightarrow \mathbb{R}^{mn}$ converts a matrix to a column vector by stacking the columns. Consider the average image

$$\mathbf{p} := \frac{1}{N}(Pe_1 + \cdots + Pe_N) = \frac{1}{N}P\mathbf{e}, \quad \text{where } \mathbf{e} := [1, 1, \ldots, 1]^\top \in \mathbb{R}^N.$$

Subtracting $\mathbf{p}$ from each columns of P, we have mean-subtracted images $A := P - \mathbf{p}\mathbf{e}^\top$. Note that each row of A has zero mean. Hence $S := \frac{1}{N-1}AA^\top$ is the covariance matrix.

Eigenvectors of S are called eigenfaces. Note that eigenfaces are left singular vectors of A and span the columns space of A. PCA uses eigenfaces to recapture an approximation to the original faces.

Face recognition

Assume that each image has the same pixel resolution with m rows and n columns. Let $P_1, \ldots, P_N$ be $m \times n$ matrices representing N images. Define

$$P := \begin{bmatrix} \mathrm{Vec}(P_1) & \cdots & \mathrm{Vec}(P_N) \end{bmatrix} \in \mathbb{R}^{mn \times N},$$

where $\mathrm{Vec} : \mathbb{R}^{m \times n} \longrightarrow \mathbb{R}^{mn}$ converts a matrix to a column vector by stacking the columns. Consider the average image

$$\mathbf{p} := \frac{1}{N}(Pe_1 + \cdots + Pe_N) = \frac{1}{N}P\mathbf{e}, \quad \text{where } \mathbf{e} := [1, 1, \ldots, 1]^\top \in \mathbb{R}^N.$$

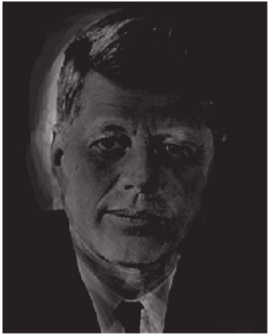
Subtracting $\mathbf{p}$ from each columns of P, we have mean-subtracted images $A := P - \mathbf{p}\mathbf{e}^\top$. Note that each row of A has zero mean. Hence $S := \frac{1}{N-1}AA^\top$ is the covariance matrix.

Eigenvectors of S are called eigenfaces. Note that eigenfaces are left singular vectors of A and span the columns space of A. PCA uses eigenfaces to recapture an approximation to the original faces.

Usually 50 or 100 eigenfaces are required to create strong approximations to a set of $10,000$ faces.

Face recognition

(a) label 1



(b) label 2

Figure: (a) The average image of twelve U.S. Presidents and (b) the average image subtracted from the image of President Kennedy .

Face recognition

For the presidential library images, S is a 50000×50000 matrix. Finding eigenvectors would be prohibitively time-consuming even for a supercomputer.

Face recognition

For the presidential library images, S is a 50000×50000 matrix. Finding eigenvectors would be prohibitively time-consuming even for a supercomputer.

However, when there are N images in the library, there will be at most N linearly independent eigenvectors of S.

Face recognition

For the presidential library images, S is a 50000×50000 matrix. Finding eigenvectors would be prohibitively time-consuming even for a supercomputer.

However, when there are N images in the library, there will be at most N linearly independent eigenvectors of S.

Since $A^\top A$ and $AA^\top$ have the same nonzero eigenvalues, we solve $A^\top Av = \lambda v$. If $\lambda \neq 0$ then $AA^\top(Av) = \lambda(Av)$ and $Av \neq 0$. Hence $u := Av$ is an eigenvector of S corresponding to the eigenvalue $\lambda/(N-1)$.

Face recognition

For the presidential library images, S is a 50000×50000 matrix. Finding eigenvectors would be prohibitively time-consuming even for a supercomputer.

However, when there are N images in the library, there will be at most N linearly independent eigenvectors of S.

Since $A^\top A$ and $AA^\top$ have the same nonzero eigenvalues, we solve $A^\top Av = \lambda v$. If $\lambda \neq 0$ then $AA^\top(Av) = \lambda(Av)$ and $Av \neq 0$. Hence $u := Av$ is an eigenvector of S corresponding to the eigenvalue $\lambda/(N-1)$.

For the presidential library images, the eigenfaces are obtained from eigenvectors of the 12×12 matrix $A^\top A$.

Figure: A half dozen eigenfaces of the library of twelve U.S. Presidents.

Face recognition

Nonzero singular values of A are square roots of the nonzero eigenvalues of $A^T A$. The singular values have tendency to drop off in value quickly.

Face recognition

Nonzero singular values of A are square roots of the nonzero eigenvalues of $A^T A$. The singular values have tendency to drop off in value quickly. This allows us to approximate a large collection of images, like $N = 5,000$, with only a small subset of eigenfaces.

Face recognition

Nonzero singular values of A are square roots of the nonzero eigenvalues of $A^T A$. The singular values have tendency to drop off in value quickly. This allows us to approximate a large collection of images, like $N = 5,000$, with only a small subset of eigenfaces.

Eigenfaces can be used to recognize a new face as follows. Subtract the average image $\mathbf{p}$ from the new image $\mathbf{x}$.

Face recognition

Nonzero singular values of A are square roots of the nonzero eigenvalues of $A^T A$. The singular values have tendency to drop off in value quickly. This allows us to approximate a large collection of images, like $N = 5,000$, with only a small subset of eigenfaces.

Eigenfaces can be used to recognize a new face as follows. Subtract the average image $\mathbf{p}$ from the new image $\mathbf{x}$. Project the mean-subtracted image $\mathbf{v} := \mathbf{x} - \mathbf{p}$ on the space spanned by the eigenfaces $\mathbf{u}_1, \ldots, \mathbf{u}_r$:

Face recognition

Nonzero singular values of A are square roots of the nonzero eigenvalues of $A^T A$. The singular values have tendency to drop off in value quickly. This allows us to approximate a large collection of images, like $N = 5,000$, with only a small subset of eigenfaces.

Eigenfaces can be used to recognize a new face as follows. Subtract the average image $\mathbf{p}$ from the new image $\mathbf{x}$. Project the mean-subtracted image $\mathbf{v} := \mathbf{x} - \mathbf{p}$ on the space spanned by the eigenfaces $\mathbf{u}_1, \ldots, \mathbf{u}_r$:

$$\mathbf{v}_{proj} := \sum_{j=1}^{r} (\mathbf{u}_j^\top \mathbf{v})\mathbf{u}_j = [\mathbf{u}_1 \cdots \mathbf{u}_r][\mathbf{u}_1 \cdots \mathbf{u}_r]^\top \mathbf{v}.$$

Then adding the average presidential image we obtain a recognizable face (or an approximation) $\widehat{\mathbf{v}} := \mathbf{v}_{proj} + \mathbf{p}$.

Face recognition

Nonzero singular values of A are square roots of the nonzero eigenvalues of $A^T A$. The singular values have tendency to drop off in value quickly. This allows us to approximate a large collection of images, like $N = 5,000$, with only a small subset of eigenfaces.

Eigenfaces can be used to recognize a new face as follows. Subtract the average image $\mathbf{p}$ from the new image $\mathbf{x}$. Project the mean-subtracted image $\mathbf{v} := \mathbf{x} - \mathbf{p}$ on the space spanned by the eigenfaces $\mathbf{u}_1, \ldots, \mathbf{u}_r$:

$$\mathbf{v}_{proj} := \sum_{j=1}^{r} (\mathbf{u}_j^\top \mathbf{v})\mathbf{u}_j = [\mathbf{u}_1 \cdots \mathbf{u}_r][\mathbf{u}_1 \cdots \mathbf{u}_r]^\top \mathbf{v}.$$

Then adding the average presidential image we obtain a recognizable face (or an approximation) $\widehat{\mathbf{v}} := \mathbf{v}_{proj} + \mathbf{p}$. If $\|\mathrm{Vec}(P_j) - \widehat{\mathbf{v}}\|_2$ is smallest then $\widehat{\mathbf{v}} \approx \mathrm{Vec}(P_j)$.

Face recognition

Nonzero singular values of A are square roots of the nonzero eigenvalues of $A^T A$. The singular values have tendency to drop off in value quickly. This allows us to approximate a large collection of images, like $N = 5,000$, with only a small subset of eigenfaces.

Eigenfaces can be used to recognize a new face as follows. Subtract the average image $\mathbf{p}$ from the new image $\mathbf{x}$. Project the mean-subtracted image $\mathbf{v} := \mathbf{x} - \mathbf{p}$ on the space spanned by the eigenfaces $\mathbf{u}_1, \ldots, \mathbf{u}_r$:

$$\mathbf{v}_{proj} := \sum_{j=1}^{r} (\mathbf{u}_j^\top \mathbf{v})\mathbf{u}_j = [\mathbf{u}_1 \cdots \mathbf{u}_r][\mathbf{u}_1 \cdots \mathbf{u}_r]^\top \mathbf{v}.$$

Then adding the average presidential image we obtain a recognizable face (or an approximation) $\widehat{\mathbf{v}} := \mathbf{v}_{proj} + \mathbf{p}$. If $\|\mathrm{Vec}(P_j) - \widehat{\mathbf{v}}\|_2$ is smallest then $\widehat{\mathbf{v}} \approx \mathrm{Vec}(P_j)$.

This is part of the power of PCA in facial recognition. One can use a large library of images, reduce it to a much smaller set of eigenfaces, and then use it to recognize a face or create an approximation, even with some disguising.

Face recognition

(a) label 1

(b) label 2

Figure: (a) An altered image of President Kennedy and (b) the image recreated using the six eigenfaces.

Polar decomposition of a matrix

If $z \in \mathbb{C}$ then z has a polar representation $z = rw$, where $r = |z| = \sqrt{\bar{z}z}$ and $\bar{w}w = 1$. A similar decomposition holds for a matrix.

Polar decomposition of a matrix

If $z \in \mathbb{C}$ then z has a polar representation $z = rw$, where $r = |z| = \sqrt{\bar{z}z}$ and $\bar{w}w = 1$. A similar decomposition holds for a matrix.

Definition: Let $A \in \mathbb{C}^{n\times n}$. Then a decomposition of the form $A = WR$, where $W \in \mathbb{C}^{n\times n}$ is unitary and $R \succeq \mathbf{0}$ (positive semidefinite), is called a polar decomposition of A. The matrices W and R are called polar factors.

Polar decomposition of a matrix

If $z \in \mathbb{C}$ then z has a polar representation $z = rw$, where $r = |z| = \sqrt{\bar{z}z}$ and $\bar{w}w = 1$. A similar decomposition holds for a matrix.

Definition: Let $A \in \mathbb{C}^{n\times n}$. Then a decomposition of the form $A = WR$, where $W \in \mathbb{C}^{n\times n}$ is unitary and $R \succeq \mathbf{0}$ (positive semidefinite), is called a polar decomposition of A. The matrices W and R are called polar factors.

Note that if polar decomposition $A = WR$ exists then $R^2 = A^*A \Longrightarrow R$ is the positive square root of A^*A. Often R is also denoted by $|A|$, that is, $|A| = \sqrt{A^*A} \succeq \mathbf{0}$.

Polar decomposition of a matrix

If $z \in \mathbb{C}$ then z has a polar representation $z = rw$, where $r = |z| = \sqrt{\bar{z}z}$ and $\bar{w}w = 1$. A similar decomposition holds for a matrix.

Definition: Let $A \in \mathbb{C}^{n\times n}$. Then a decomposition of the form $A = WR$, where $W \in \mathbb{C}^{n\times n}$ is unitary and $R \succeq \mathbf{0}$ (positive semidefinite), is called a polar decomposition of A. The matrices W and R are called polar factors.

Note that if polar decomposition $A = WR$ exists then $R^2 = A^*A \Longrightarrow R$ is the positive square root of A^*A. Often R is also denoted by $|A|$, that is, $|A| = \sqrt{A^*A} \succeq \mathbf{0}$.

If A is nonsingular then $R = \sqrt{A^*A} \succ \mathbf{0}$ (positive definite). Define $W := AR^{-1}$.

Polar decomposition of a matrix

If $z \in \mathbb{C}$ then z has a polar representation $z = rw$, where $r = |z| = \sqrt{\bar{z}z}$ and $\bar{w}w = 1$. A similar decomposition holds for a matrix.

Definition: Let $A \in \mathbb{C}^{n\times n}$. Then a decomposition of the form $A = WR$, where $W \in \mathbb{C}^{n\times n}$ is unitary and $R \succeq \mathbf{0}$ (positive semidefinite), is called a polar decomposition of A. The matrices W and R are called polar factors.

Note that if polar decomposition $A = WR$ exists then $R^2 = A^*A \Longrightarrow R$ is the positive square root of A^*A. Often R is also denoted by $|A|$, that is, $|A| = \sqrt{A^*A} \succeq \mathbf{0}$.

If A is nonsingular then $R = \sqrt{A^*A} \succ \mathbf{0}$ (positive definite). Define $W := AR^{-1}$. Then we have $A = WR$ and $W^*W = R^{-1}A^*AR^{-1} = I_n$. This shows that $A = WR$ is a polar decomposition.

Polar decomposition of a matrix

If $z \in \mathbb{C}$ then z has a polar representation $z = rw$, where $r = |z| = \sqrt{\bar{z}z}$ and $\bar{w}w = 1$. A similar decomposition holds for a matrix.

Definition: Let $A \in \mathbb{C}^{n\times n}$. Then a decomposition of the form $A = WR$, where $W \in \mathbb{C}^{n\times n}$ is unitary and $R \succeq \mathbf{0}$ (positive semidefinite), is called a polar decomposition of A. The matrices W and R are called polar factors.

Note that if polar decomposition $A = WR$ exists then $R^2 = A^*A \Longrightarrow R$ is the positive square root of A^*A. Often R is also denoted by $|A|$, that is, $|A| = \sqrt{A^*A} \succeq \mathbf{0}$.

If A is nonsingular then $R = \sqrt{A^*A} \succ \mathbf{0}$ (positive definite). Define $W := AR^{-1}$. Then we have $A = WR$ and $W^*W = R^{-1}A^*AR^{-1} = I_n$. This shows that $A = WR$ is a polar decomposition.

Remark: Note that polar decomposition is unique when A is nonsingular.

Polar decomposition of a matrix

If $z \in \mathbb{C}$ then z has a polar representation $z = rw$, where $r = |z| = \sqrt{\bar{z}z}$ and $\bar{w}w = 1$. A similar decomposition holds for a matrix.

Definition: Let $A \in \mathbb{C}^{n\times n}$. Then a decomposition of the form $A = WR$, where $W \in \mathbb{C}^{n\times n}$ is unitary and $R \succeq \mathbf{0}$ (positive semidefinite), is called a polar decomposition of A. The matrices W and R are called polar factors.

Note that if polar decomposition $A = WR$ exists then $R^2 = A^*A \Longrightarrow R$ is the positive square root of A^*A. Often R is also denoted by $|A|$, that is, $|A| = \sqrt{A^*A} \succeq \mathbf{0}$.

If A is nonsingular then $R = \sqrt{A^*A} \succ \mathbf{0}$ (positive definite). Define $W := AR^{-1}$. Then we have $A = WR$ and $W^*W = R^{-1}A^*AR^{-1} = I_n$. This shows that $A = WR$ is a polar decomposition.

Remark: Note that polar decomposition is unique when A is nonsingular.

Theorem: Let $A \in \mathbb{C}^{m\times n}$ with $m \geq n$. Then there exists an isometry $W \in \mathbb{C}^{m\times n}$ and a positive semidefinite matrix $R \in \mathbb{C}^{n\times n}$ such that $A = WR$.

Polar decomposition

Proof: Consider the trimmed SVD $A = U_n \Sigma_n V_n^*$, where $U_n \in \mathbb{C}^{m \times n}$ is an isometry, $V_n \in \mathbb{C}^{n \times n}$ is unitary, and $\Sigma_n = \mathrm{diag}(\sigma_1, \cdots, \sigma_n) \in \mathbb{R}^{n \times n}$.

Polar decomposition

Proof: Consider the trimmed SVD $A = U_n \Sigma_n V_n^*$, where $U_n \in \mathbb{C}^{m \times n}$ is an isometry, $V_n \in \mathbb{C}^{n \times n}$ is unitary, and $\Sigma_n = \mathrm{diag}(\sigma_1, \cdots, \sigma_n) \in \mathbb{R}^{n \times n}$. Then

$$A = U_n \Sigma_n V_n^* = U_n V_n^* (V_n \Sigma_n V_n^*) = WR,$$

where $W := U_n V_n^* \in \mathbb{C}^{m \times n}$ is an isometry and $R := V_n \Sigma_n V_n^* \succeq \mathbf{0}$. ■

Polar decomposition

Proof: Consider the trimmed SVD $A = U_n \Sigma_n V_n^*$, where $U_n \in \mathbb{C}^{m\times n}$ is an isometry, $V_n \in \mathbb{C}^{n\times n}$ is unitary, and $\Sigma_n = \mathrm{diag}(\sigma_1, \cdots, \sigma_n) \in \mathbb{R}^{n\times n}$. Then

$$A = U_n \Sigma_n V_n^* = U_n V_n^* (V_n \Sigma_n V_n^*) = WR,$$

where $W := U_n V_n^* \in \mathbb{C}^{m\times n}$ is an isometry and $R := V_n \Sigma_n V_n^* \succeq \mathbf{0}$. $\blacksquare$

SVD from polar decomposition: Let $A \in \mathbb{C}^{m\times n}$ and $A = WR$ be a polar decomposition.

Polar decomposition

Proof: Consider the trimmed SVD $A = U_n \Sigma_n V_n^*$, where $U_n \in \mathbb{C}^{m\times n}$ is an isometry, $V_n \in \mathbb{C}^{n\times n}$ is unitary, and $\Sigma_n = \mathrm{diag}(\sigma_1, \cdots, \sigma_n) \in \mathbb{R}^{n\times n}$. Then

$$A = U_n \Sigma_n V_n^* = U_n V_n^* (V_n \Sigma_n V_n^*) = WR,$$

where $W := U_n V_n^* \in \mathbb{C}^{m\times n}$ is an isometry and $R := V_n \Sigma_n V_n^* \succeq \mathbf{0}$. ■

SVD from polar decomposition: Let $A \in \mathbb{C}^{m\times n}$ and $A = WR$ be a polar decomposition. Since $R \succeq 0$, by spectral theorem, we have $R = V\Sigma V^*$, where V is unitary, $\Sigma := \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$ and $\sigma_1 \geq \sigma_2 \cdots \geq \sigma_n \geq 0$ are eigenvalues of R.

Polar decomposition

Proof: Consider the trimmed SVD $A = U_n \Sigma_n V_n^*$, where $U_n \in \mathbb{C}^{m\times n}$ is an isometry, $V_n \in \mathbb{C}^{n\times n}$ is unitary, and $\Sigma_n = \mathrm{diag}(\sigma_1, \cdots, \sigma_n) \in \mathbb{R}^{n\times n}$. Then

$$A = U_n \Sigma_n V_n^* = U_n V_n^* (V_n \Sigma_n V_n^*) = WR,$$

where $W := U_n V_n^* \in \mathbb{C}^{m\times n}$ is an isometry and $R := V_n \Sigma_n V_n^* \succeq \mathbf{0}$. ■

SVD from polar decomposition: Let $A \in \mathbb{C}^{m\times n}$ and $A = WR$ be a polar decomposition. Since $R \succeq 0$, by spectral theorem, we have $R = V\Sigma V^*$, where V is unitary, $\Sigma := \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$ and $\sigma_1 \geq \sigma_2 \cdots \geq \sigma_n \geq 0$ are eigenvalues of R.

This shows that $A = WR = WV\Sigma V^* = U\Sigma V^*$ is a trimmed SVD of A, where $U := WV$ is an isometry.

Polar decomposition

Proof: Consider the trimmed SVD $A = U_n \Sigma_n V_n^*$, where $U_n \in \mathbb{C}^{m\times n}$ is an isometry, $V_n \in \mathbb{C}^{n\times n}$ is unitary, and $\Sigma_n = \mathrm{diag}(\sigma_1, \cdots, \sigma_n) \in \mathbb{R}^{n\times n}$. Then

$$A = U_n \Sigma_n V_n^* = U_n V_n^* (V_n \Sigma_n V_n^*) = WR,$$

where $W := U_n V_n^* \in \mathbb{C}^{m\times n}$ is an isometry and $R := V_n \Sigma_n V_n^* \succeq \mathbf{0}$. ■

SVD from polar decomposition: Let $A \in \mathbb{C}^{m\times n}$ and $A = WR$ be a polar decomposition. Since $R \succeq 0$, by spectral theorem, we have $R = V\Sigma V^*$, where V is unitary, $\Sigma := \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$ and $\sigma_1 \geq \sigma_2 \cdots \geq \sigma_n \geq 0$ are eigenvalues of R.

This shows that $A = WR = WV\Sigma V^* = U\Sigma V^*$ is a trimmed SVD of A, where $U := WV$ is an isometry.

Example: Consider $A := \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{bmatrix}$. We now compute polar decomposition $A = WR$ from SVD of A.

Example

The trimmed SVD of A is given by

```
>> [U, S, V] = svd(A, 0)
```

$$U = \begin{bmatrix} -0.1409 & 0.8247 & 0.5477 \\ -0.3439 & 0.4263 & -0.7276 \\ -0.5470 & 0.0278 & -0.1880 \\ -0.7501 & -0.3706 & 0.3679 \end{bmatrix}, \quad S = \begin{bmatrix} 25.4624 & 0 & 0 \\ 0 & 1.2907 & 0 \\ 0 & 0 & 0.0000 \end{bmatrix},$$

$$V = \begin{bmatrix} -0.5045 & -0.7608 & -0.4082 \\ -0.5745 & -0.0571 & 0.8165 \\ -0.6445 & 0.6465 & -0.4082 \end{bmatrix}$$

Example

The desired polar factors are given by

```
>> W = U*V', R = V*S*V'
```

$$W = \begin{bmatrix} -0.7799 & 0.4810 & 0.4004 \\ 0.1463 & -0.4208 & 0.7943 \\ 0.3316 & 0.1592 & 0.4473 \\ 0.5102 & 0.7525 & 0.0936 \end{bmatrix}, \quad R = \begin{bmatrix} 7.2286 & 7.4367 & 7.6448 \\ 7.4367 & 8.4085 & 9.3804 \\ 7.6448 & 9.3804 & 11.1160 \end{bmatrix}.$$

***