Notes on Singular Value and Eigenvalue Decompositions

Paul F. Roysdon, Ph.D.

I. Introduction

Matrix factorization, a.k.a. decomposition, is a method to "factor" a matrix into a product of matrices, e.g. the QR Decomposition $\mathbf{A} = \mathbf{QR}$ (see Notes on Matrix Decomposition, Section VI), and different factorizations reveal different properties of a matrix. Factorization facilitates solving a system of linear equations $\mathbf{Ax} = \mathbf{b}$ by forward and/or backward substitution, resulting in fewer additions and multiplications. In practice, matrix decomposition is most commonly used to solve the matrix inverse $\mathbf{A}^{-1}$. Herein we focus on Singular Value and Eigenvalue Decompositions.

II. Singular Value Decomposition

A. Definition

The singular value decomposition (SVD) [1], [2], [3], [4] factorizes a linear operator $\mathbf{A} : \mathbb{R}^n \mapsto \mathbb{R}^m$ into three simplified linear operators:

1) Projection $\mathbf{z} = \mathbf{V}^\intercal \mathbf{x}$ into an r-dimensional space, where r is the rank of $\mathbf{A}$.
2) Element-wise multiplication with r singular values $\boldsymbol{\sigma}_i$, i.e., $\mathbf{z}' = \boldsymbol{\Sigma}\mathbf{z}$.
3) Transformation $\mathbf{y} = \mathbf{U}\mathbf{z}'$ to the m-dimensional output space.

Combining these statements, $\mathbf{A}$ can be re-written as

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\intercal \tag{1}$$

with $\mathbf{U}$ an $m \times r$ orthonormal matrix spanning $\mathbf{A}$'s column space $im(\mathbf{A})$ (the **left singular vectors** of $\mathbf{A}$), $\boldsymbol{\Sigma}$ an $r \times r$ diagonal matrix of **singular values** of $\mathbf{A}$, and $\mathbf{V}$ an $n \times r$ orthonormal matrix spanning $\mathbf{A}$'s row space $im(\mathbf{A}^\intercal)$ (the **right singular vectors** of $\mathbf{A}$).

B. Useful SVD facts

- *Any* $m \times n$ matrix $\mathbf{A}$ has a SVD.
- The singular values of $\mathbf{A}$ are uniquely determined.
- If $\mathbf{A}$ is square and the singular values are distinct, then the left and right singular vectors are uniquely determined up to complex signs.
- Suppose $\mathbf{A}$ has r non-zero singular values. Then
  - $\text{rank}(\mathbf{A}) = r$
  - $\text{null}(\mathbf{A}) = \text{span}\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$
  - $\text{range}(\mathbf{A}) = \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$
  - $\text{null}(\mathbf{A}^\intercal) = \text{span}\{\mathbf{u}_{r+1}, \dots, \mathbf{u}_m\}$
  - $\text{range}(\mathbf{A}^\intercal) = \text{span}\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$
- If $\mathbf{A}$ is an $n \times n$ square matrix, then $|\det(\mathbf{A})| = \prod_{i=1}^{n} \boldsymbol{\sigma}_i$.
- If $\mathbf{A} \in \mathbb{C}^{m\times n}$, $\mathbf{U}$ and $\mathbf{V}$ are unitary, but the singular values are always in $\mathbb{R}$.
- **Thin SVD**: If the SVD of an $m \times n$ matrix $\mathbf{A}$ is given by $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\intercal$ then the **thin SVD** of $\mathbf{A}$ where $m \geq n$ expresses $\mathbf{A} = \mathbf{U}_1\boldsymbol{\Sigma}_1\mathbf{V}^\intercal$, where $\mathbf{U}_1$ is $m \times n$ and contains the first n columns of $\mathbf{U}$, and $\boldsymbol{\Sigma}_1$ is an $n \times n$ diagonal matrix containing the n singular values of $\mathbf{A}$.
- If the SVD of $\mathbf{A}$ is given by $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\intercal$ then
  - The singular values of $\mathbf{A}$ are the square roots of the non-zero eigenvalues of both $\mathbf{A}\mathbf{A}^\intercal$ and $\mathbf{A}^\intercal\mathbf{A}$.
  - The columns of $\mathbf{U}$ are the orthonormal eigenvectors of $\mathbf{A}^\intercal\mathbf{A}$.
  - The columns of $\mathbf{V}$ are the orthonormal eigenvectors of $\mathbf{A}\mathbf{A}^\intercal$.
- If $\mathbf{A}$ is symmetric, then the singular values of $\mathbf{A}$ are absolute values of the eigenvectors of $\mathbf{A}$.
- The SVD can be written as a sum of rank-1 matrices:

  $$\mathbf{A} = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\intercal,$$

  where $r = \text{rank}(\mathbf{A})$, σ_i is the i^{th} singular value, and $\mathbf{u}_i$, $\mathbf{v}_i$ are the i^{th} left and right singular vectors, respectively.
- If $k < r = \text{rank}(\mathbf{A})$ and $\mathbf{A}_k = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^\intercal$, then

  $$\min_{\text{rank}(\mathbf{B})=k} \|\mathbf{A} - \mathbf{B}\|_2 = \|\mathbf{A} - \mathbf{A}_k\|_2 = \boldsymbol{\sigma}_{k+1}$$

- The least squares solution $\mathbf{x}$ to the problem $\min \mathbf{x} \in \mathbb{R}^n \|\mathbf{b} - \mathbf{Ax}\|_2$ is given by the SVD:

  $$\mathbf{x} = \sum_{i=1}^{r} \frac{\mathbf{u}_i^\intercal \mathbf{b}}{\sigma_i} \mathbf{v}_i$$

- The pseudo inverse (or **Moore-Penrose pseudo-inverse**) of a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ is the matrix $\mathbf{A}^+$ that fulfills

  $$\begin{array}{rl} I & \mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A} \\ II & \mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+ \\ III & \mathbf{A}\mathbf{A}^+ \text{ symmetric} \\ IV & \mathbf{A}^+\mathbf{A} \text{ symmetric} \end{array}$$

  The matrix $\mathbf{A}^+$ is unique and does always exist. Note that in case of complex matrices, the symmetric condition is substituted by a condition of being Hermitian. The pseudo-inverse of $\mathbf{A}^+ \in \mathbb{R}^{n\times m}$ of $\mathbf{A} \in \mathbb{R}^{m\times n}$ is computed by the SVD of $\mathbf{A}$. Specifically, if $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\intercal$ is the SVD of $\mathbf{A}$, then $\mathbf{A}^+ = \mathbf{V}\boldsymbol{\Sigma}^+\mathbf{U}^\intercal$, where $\boldsymbol{\Sigma}^+ \in \mathbb{R}^{n\times m}$ has diagonal elements $\frac{1}{\sigma_i}$ for $i = 1, \dots, r$

followed by $n-r$ zeros ($r = \text{rank}(\mathbf{A})$). The pseudo-inverse of $\mathbf{A}^+ \in \mathbb{R}^{n\times m}$ of $\mathbf{A} \in \mathbb{R}^{m\times n}$ is the unique Frobenius norm solution to the problem

$$\min \mathbf{X} \in \mathbb{R}^{n\times m} \|\mathbf{AX} - \mathbf{I}_m\|_F.$$

- If $\mathbf{A}$ is $m \times n$ and has singular values $\sigma_1, \ldots, \sigma_r$, then
  - $\|\mathbf{A}\|_2 = \sigma_1$.
  - $\|\mathbf{A}\|_F = \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_r^2}$.
  - The
    $$\min_{\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{Ax}\|_2}{\|\mathbf{x}\|_2} = \sigma_n$$
    if $m \geq n$.
  - The solution to
    $$\max_{\mathbf{x}\in\mathbb{R}^m, \mathbf{y}\in\mathbb{R}^n} \frac{|\mathbf{x}^\intercal \mathbf{A} \mathbf{y}|}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2}$$
    is σ_1 with $\mathbf{x} = \mathbf{u}_1$, $\mathbf{y} = \mathbf{v}_1$, i.e. maximization of bilinear forms.
  - If $\mathbf{A} \in \mathbb{R}^{m\times n}$, then for $k = 1, \ldots, \min(m, n)$
    $$\boldsymbol{\sigma}_k = \max_{\substack{\dim(\mathbf{S})=k \\ \dim(\mathbf{T})=k}} \min_{\substack{\mathbf{x}\in\mathbf{S} \\ \mathbf{y}\in\mathbf{T}}} \frac{\mathbf{y}^\intercal \mathbf{A} \mathbf{x}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2}$$
    i.e. the minimax theorem.

C. Common SVD applications

- Principle Component Analysis (PCA) for dimension reduction.
- Latent Semantic Indexing.
- EigenFaces: PCA combined with Expectation-Maximization (EM).
- Collaborative Filtering.
- Data mining.
- Cross-Modal Factor Analysis (CFA).
- Canonical Correlation Analysis (CCA).
- Geometric Applications.
- Compression.
- Other applications.

TODO: finish this section.

III. Eigenvalue Decomposition

A. Definition

The eigenvalue decomposition (EVD) [1], [2], [3], [4] applies to mappings from $\mathbb{R}^n$ to itself, i.e., a linear operator $\mathbf{A} : \mathbb{R}^n \mapsto \mathbb{R}^n$ described by a square matrix. An **eigenvector** $\mathbf{e}$ of $\mathbf{A}$ is a vector that is mapped to a scaled version of itself, i.e., $\mathbf{Ae} = \boldsymbol{\lambda}\mathbf{e}$, where $\boldsymbol{\lambda}$ is the corresponding **eigenvalue**. For a matrix $\mathbf{A}$ of rank r, we can group the r non-zero eigenvalues in an $r \times r$ diagonal matrix $\boldsymbol{\Lambda}$ and their eigenvectors in an $n \times r$ matrix $\mathbf{E}$, and we have

$$\mathbf{AE} = \mathbf{E\Lambda}$$

Furthermore, if $\mathbf{A}$ is full rank ($r = n$) then $\mathbf{A}$ can be factorized as

$$\mathbf{A} = \mathbf{E\Lambda E}^{-1}$$

which is a diagonalization similar to the SVD in Eq. 1. In fact, if and only if $\mathbf{A}$ is symmetric and positive definite (abbreviated SPD), we have that the SVD and the eigen-decomposition coincide

$$\mathbf{A} = \mathbf{USV}^\intercal = \mathbf{E\Lambda E}^{-1}$$

with $\mathbf{U} = \mathbf{E}$ and $\mathbf{S} = \boldsymbol{\Lambda}$.

Given a non-square matrix $\mathbf{A} = \mathbf{USV}^\intercal$, two matrices and their factorization are of special interest:

$$\mathbf{A}^\intercal\mathbf{A} = \mathbf{VS}^2\mathbf{V}^\intercal \mathbf{AA}^\intercal = \mathbf{US}^2\mathbf{U}^\intercal$$

Thus, for these matrices the SVD on the original matrix $\mathbf{A}$ can be used to compute their SVD. And since these matrices are by definition SPD, this is also their eigen-decomposition, with eigenvalues $\boldsymbol{\Lambda} = \mathbf{S}^2$.

B. Useful EVD facts

- A square matrix $\mathbf{A}$ has n (not necessarily distinct) eigenvalues.
- $\mathbf{A}$ is invertible *iff* all its eigenvalues are nonzero.
- $\mathbf{A}$ is diagonalizable *iff* $\mathbf{A}$ has n linearly independent eigenvectors.
- $\prod_{i=1}^{n} \boldsymbol{\lambda}_i = \det(\mathbf{A})$
- $\sum_{i=1}^{n} \boldsymbol{\lambda}_i = \text{Tr}(\mathbf{A})$
- If $\mathbf{A}$ is symmetric then $\boldsymbol{\lambda}_i \in \mathbb{R}$ for $i = 1, \ldots, n$.
- If $\mathbf{A}$ is symmetric then there exists an orthogonal $\mathbf{Q} \in \mathbb{R}^{n\times n}$ such that its eigendecomposition can be written as $\mathbf{A} = \mathbf{QDQ}^\intercal$.
- $\boldsymbol{\lambda}$ is an eigenvalue *iff* $\det(\mathbf{A} - \boldsymbol{\lambda}\mathbf{I}) = \mathbf{0}$. The characteristic polynomial of $\mathbf{A}$ is $p_{\mathbf{A}}(\boldsymbol{\lambda}) = \det(\mathbf{A} - \boldsymbol{\lambda}\mathbf{I})$.

C. Common EVD applications

- Principle Component Analysis for dimension reduction.
- Computing powers of matrices: $\mathbf{A}^N$ where N is large.
- Computing operations with matrices, e.g. $e^{\mathbf{A}}$, $\ln \mathbf{A}$, etc.
- Google's PageRank.
- EigenFaces: PCA combined with Expectation-Maximization (EM).

IV. Examples

A. Simple Numerical Example

Consider the symmetric matrix $\mathbf{A} \in \mathbb{R}^3$,

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}.$$

The singular value decomposition produces

$$\mathbf{U} = \begin{bmatrix} -0.5 & -0.70711 & 0.5 \\ 0.70711 & -3.3215e-16 & 0.70711 \\ -0.5 & 0.70711 & 0.5 \end{bmatrix}$$

$$\mathbf{S} = \begin{bmatrix} 3.4142 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0.58579 \end{bmatrix}$$

$$\mathbf{V}^\intercal = \begin{bmatrix} -0.5 & -0.70711 & 0.5 \\ 0.70711 & 1.9725e-16 & 0.70711 \\ -0.5 & 0.70711 & 0.5 \end{bmatrix},$$

using the Matlab command

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

Note that the singular values in $\mathbf{S}$ are in ascending order!

The eigenvalue decomposition produces

$$\mathbf{E} = \begin{bmatrix} 0.5 & -0.70711 & -0.5 \\ 0.70711 & 6.0825e-17 & 0.70711 \\ 0.5 & 0.70711 & -0.5 \end{bmatrix}$$

$$\mathbf{V} = \begin{bmatrix} 0.58579 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3.4142 \end{bmatrix},$$

using the Matlab command

```
[E,V] = eig(A);
```

Note that the eigenvalues in $\mathbf{V}$ are in descending order!

B. Photo Compression Example

At the advent of the United States space-exploration program, the scientists at NASA required a method to compress images sent from distant spacecraft. Using technology available at the time (1960's), a full-resolution image of Mars would take several hours to transmit due to (primarily) the distance from Mars to Earth.

- The speed of light (in a vacuum): $299,792,458\ m/s$
- The average distance between Earth and Mars (depending on their elliptical orbits): $225,000,000,000\ m$

Therefore the average transit time is 751 seconds, or 12.5 minutes. The minimum (theoretical) time is 3.03 minutes, while the maximum (observed) time is 22.4 minutes. Applying image compression, an initial "discernible" image could be sent to verify the spacecraft location at Mars. Later, a higher-resolution could be sent.

Given the original full-color photo in Fig. 1, we can use the SVD to compute the singular values of the photo. First, convert the photo to grey-scale (see Fig. 2) thereby reducing the number of values (for each pixel) required to represent the image. The result is a reduction from $6,003,000$ to $2,001,000$ total values! Then, using the submatrices of $\mathbf{U}_n$, $\mathbf{\Sigma}_n$, and $\mathbf{V}_n^\intercal$ corresponding to the first n-singular values, perform an image reconstruction. As we use fewer singular values (Fig. 3-8), the dominant elements (the mountains) in the photo are less recognizable. Yet, if given the task to find the *minimum* number of singular values to reconstruct a "discernible" image, we might chose something between $10-50$ singular values resulting in a 99.6% file-size reduction!

This example is easily implemented in Matlab with 13 lines of code.

```
% open the original photo and map to matrix A
A_color = im2double(imread('arizona_photo.jpg'));
% convert the image to a grey-scale photo
A_grey = rgb2gray(A_color);
% calculate the SVD
[U,S,V] = svd(A_grey);
% select only the first n-singular values
n = 50;
U_n = U(:,1:n);
V_n = V(:,1:n);
S_n = S(1:n,1:n);
% re-compute the "compressed" photo
A_compressed = U_n * S_n * V_n';
```

References

[1] D. S. Bernstein, *Matrix Mathematics: Theory, Facts and Formulas*. Princeton University Press, 2009.

[2] G. Strang, *Linear Algebra and its Applications*. Thompson Brooks/-Cole, 2006.

[3] I. Bronshtein, K. Semendyayev, G. Musiol, and H. Muehlig, *Mandbook of Mathematics*. Springer, 2007.

[4] G. Golub and C. Van Loan, *Matrix Computations*. Johns Hopkins, 1996.

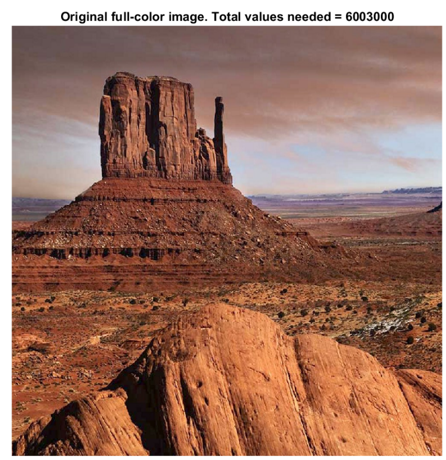

Fig. 1. Original photo, comprised of $6,003,000$ total values.

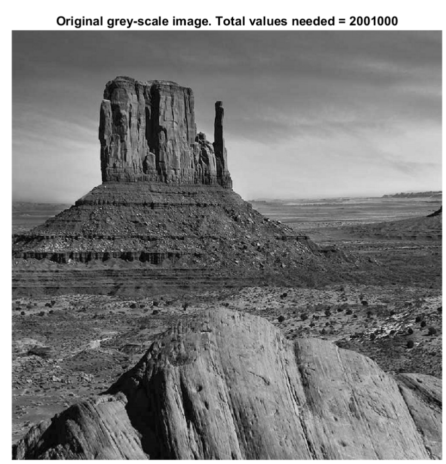

Fig. 2. Grey-scale photo, comprised of $2,001,000$ total values.

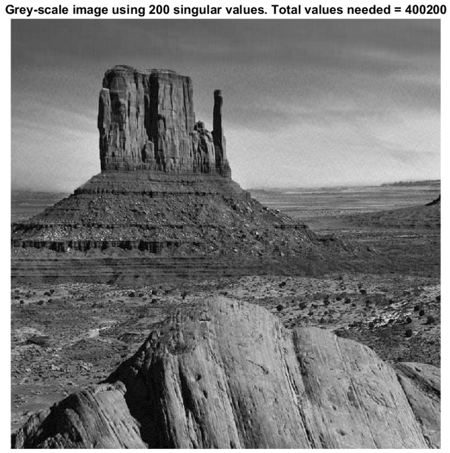

Fig. 3. Photo reconstruction using 200 singular values, $400,200$ total values.

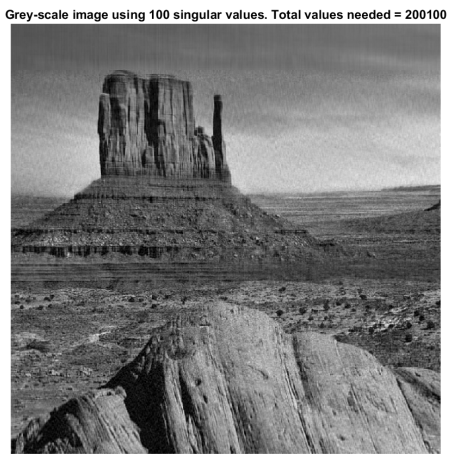

Fig. 4. Photo reconstruction using 100 singular values, $200,100$ total values.

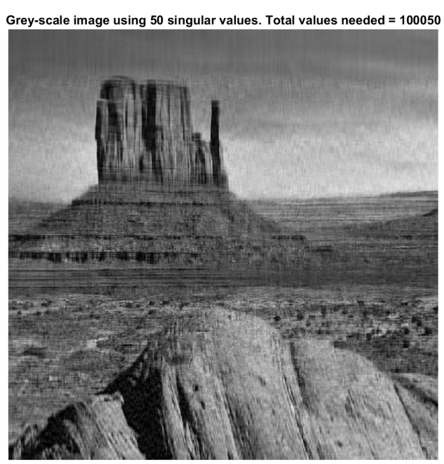

Fig. 5. Photo reconstruction using 50 singular values, $100,050$ total values.

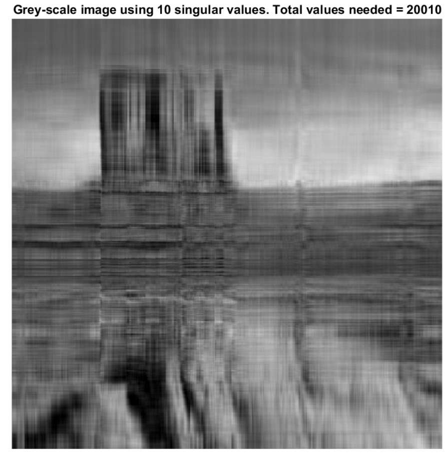

Fig. 6. Photo reconstruction using 10 singular values, $20,010$ total values.

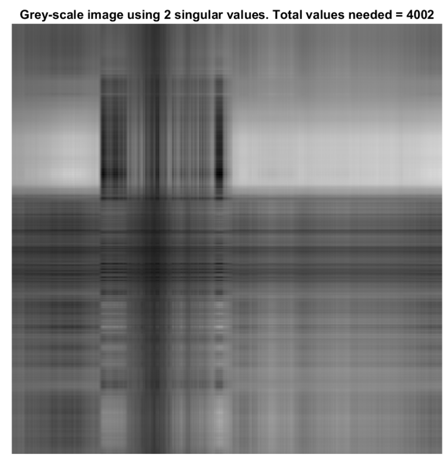

Fig. 7. Photo reconstruction using 2 singular values, $4,002$ total values.

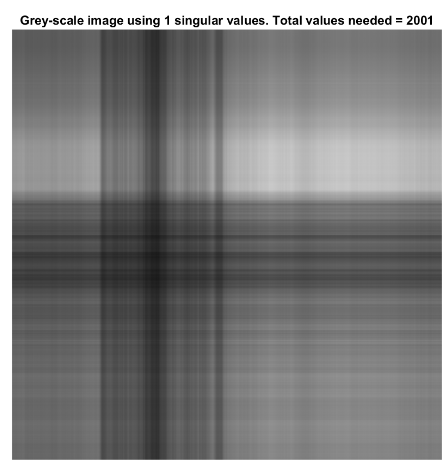

Fig. 8. Photo reconstruction using 1 singular value, $2,001$ total values.