Classical Numerical Analysis, Chapter 02

Abner J. Salgado and Steven M. Wise

asalgad1@utk.edu swise1@.utk.edu
University of Tennessee

Chapter 02
Singular Value Decomposition

Why is the Singular Value Decomposition (SVD) of Interest?

If $\mathsf{A} \in \mathbb{C}^{n\times n}$ is Hermitian, then it is unitarily diagonalizable, i.e., there is a diagonal matrix $\mathsf{D} \in \mathbb{R}^{n\times n}$ and a unitary matrix $\mathsf{U} \in \mathbb{C}^{n\times n}$ such that

$$\mathsf{A} = \mathsf{U}\mathsf{D}\mathsf{U}^{\mathsf{H}}.$$

This gives us, at least for this class of matrices, a nice geometric interpretation of the action of a matrix on a vector.

1. Since U^{H} is unitary and $\|\mathsf{U}^{\mathsf{H}}\mathbf{x}\|_2 = \|\mathbf{x}\|_2$, the action of U^{H} is *essentially* that of a rotation/reflection (it does not change the magnitude of the vector).
2. The matrix D is diagonal; its action is a (signed) dilation in each coordinate direction.
3. Finally, $\mathsf{U} = \left(\mathsf{U}^{\mathsf{H}}\right)^{-1}$ reverses the rotation/reflection implemented by U^{H}.

Unfortunately, this decomposition only holds for Hermitian matrices. Can we somehow generalize unitary diagonalization for general matrices, even non-square matrices?

Reduced and Full SVDs

Definition of the (Full) Singular Value Decomposition

Definition (SVD)

Let $A \in \mathbb{C}^{m\times n}$. A **singular value decomposition** (SVD) of the matrix A is a factorization of the form

$$A = U\Sigma V^{H},$$

where $U \in \mathbb{C}^{m\times m}$ and $V \in \mathbb{C}^{n\times n}$ are unitary, $\Sigma \in \mathbb{R}^{m\times n}$ is diagonal — meaning that $[\Sigma]_{i,j} = 0$, for $i \neq j$ — and the diagonal entries $[\Sigma]_{i,i} = \sigma_i$ are nonnegative and in non–increasing order: $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$ with $p = \min(m, n)$. The elements of the Σ are called the **singular values of** A. The columns of U and V are called the **left** and **right singular vectors**, respectively.

Full SVD to Reduced SVD

Let us, for the sake of definiteness, assume that $m \geq n$. If $\mathsf{A} \in \mathbb{C}^{m\times n}$ has an SVD, then we can write

$$\mathsf{A}\mathbf{v}_j = \sigma_j \mathbf{u}_j, \quad j = 1, \ldots, n,$$

or, equivalently,

$$\mathsf{A}\begin{bmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_n \\ | & & | \end{bmatrix}\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix}.$$

In other words, we have obtained the representation

$$\mathsf{A}\mathsf{V} = \hat{\mathsf{U}}\hat{\Sigma},$$

where (i) $\hat{\Sigma} \in \mathbb{R}^{n\times n}$ is square and diagonal, with non–negative diagonal entries; (ii) $\hat{\mathsf{U}} \in \mathbb{C}^{m\times n}$ has orthonormal columns; and (iii) $\mathsf{V} \in \mathbb{C}^{n\times n}$ is unitary. Thus, we have the **reduced SVD**,

$$\mathsf{A} = \hat{\mathsf{U}}\hat{\Sigma}\mathsf{V}^{\mathsf{H}}.$$

Reduced SVD to Full SVD

Now, assume that $m > n$. Suppose that we know (somehow) that $\mathsf{A} \in \mathbb{C}^{m\times n}$ has the reduced SVD

$$\mathsf{A} = \hat{\mathsf{U}}\hat{\Sigma}\mathsf{V}^{\mathsf{H}},$$

where (i) $\hat{\Sigma} \in \mathbb{R}^{n\times n}$ is square and diagonal, with non–negative diagonal entries; (ii) $\hat{\mathsf{U}} \in \mathbb{C}^{m\times n}$ has orthonormal columns; and (iii) $\mathsf{V} \in \mathbb{C}^{n\times n}$ is unitary. Use basis completion and the Gram-Schmidt process to add orthonormal vectors onto the columns of $\hat{\mathsf{U}}$. Define

$$\mathsf{U} = \begin{bmatrix} | & & | & | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_n & \mathbf{u}_{n+1} & \cdots & \mathbf{u}_m \\ | & & | & | & & | \end{bmatrix},$$

where $\mathbf{u}_{n+1}, \cdots, \mathbf{u}_m$ are the added orthonormal columns. The matrix $\mathsf{U} \in \mathbb{C}^{m\times m}$ is unitary.

Reduced SVD to Full SVD, Cont.

Next, define

$$\Sigma = \begin{bmatrix} \hat{\Sigma} \\ \mathrm{O} \end{bmatrix} \in \mathbb{R}^{m \times n},$$

where $\mathrm{O} \in \mathbb{R}^{(m-n) \times n}$ is a matrix all of whose entries are zero.

Now, it follows that

$$\mathrm{A} = \mathrm{U}\Sigma\mathrm{V}^{\mathrm{H}},$$

because the added columns of U are nullified by the added zeros in Σ. Essentially, no new information has been added.

We have show that, if A has a full SVD, it has a reduced SVD, and *vice versa*.

Theorem (existence of SVD)

Every matrix $\mathsf{A} \in \mathbb{C}^{m \times n}$ *has a singular value decomposition. The singular values are unique and*

$$\{\sigma_i^2\}_{i=1}^p = \begin{cases} \sigma(\mathsf{A}^{\mathsf{H}}\mathsf{A}) & \text{if} \quad m \geq n, \\ \sigma(\mathsf{A}\mathsf{A}^{\mathsf{H}}) & \text{if} \quad m \leq n, \end{cases}$$

where $p = \min(m, n)$. *Recall, the symbol* $\sigma(\mathsf{B})$ *stands for the spectrum of the square matrix* B.

Proof.

Assume, WLOG, that $m > n$. Clearly, $\mathsf{A}^{\mathsf{H}}\mathsf{A}$ and $\mathsf{A}\mathsf{A}^{\mathsf{H}}$ are Hermitian. Let $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$ be arbitrary. Then

$$\mathbf{x}^{\mathsf{H}}\mathsf{A}^{\mathsf{H}}\mathsf{A}\mathbf{x} = \|\mathsf{A}\mathbf{x}\|_2^2 \geq 0 \quad \text{and} \quad \mathbf{y}^{\mathsf{H}}\mathsf{A}\mathsf{A}^{\mathsf{H}}\mathbf{y} = \left\|\mathsf{A}^{\mathsf{H}}\mathbf{y}\right\|_2^2 \geq 0.$$

This proves that eigenvalues of $\mathsf{A}^{\mathsf{H}}\mathsf{A}$ and $\mathsf{A}\mathsf{A}^{\mathsf{H}}$ are non-negative. (Exercise.) Furthermore, $\mathsf{A}^{\mathsf{H}}\mathsf{A}$ and $\mathsf{A}\mathsf{A}^{\mathsf{H}}$ share the same non-zero eigenvalues. (Exercise.)

Proof, Cont.

By the spectral decomposition theorem, there are unitary matrices $\mathsf{U} \in \mathbb{C}^{m\times m}$ and $\mathsf{V} \in \mathbb{C}^{n\times n}$ such that

$$\mathsf{A}\mathsf{A}^{\mathsf{H}} = \mathsf{U}\Lambda_1\mathsf{U}^{\mathsf{H}} \quad \text{and} \quad \mathsf{A}^{\mathsf{H}}\mathsf{A} = \mathsf{V}\Lambda_2\mathsf{V}^{\mathsf{H}},$$

where

$$\Lambda_1 = \mathbf{diag}(\lambda_1, \ldots, \lambda_n, 0, \ldots, 0) \in \mathbb{R}^{m\times m} \quad \text{and} \quad \Lambda_2 = \mathbf{diag}(\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^{n\times n}.$$

Without loss of generality, we may assume that the eigenvalues are ordered so that

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0.$$

Now, we define

$$\sigma_i := \sqrt{\lambda_i}, \quad i = 1, \ldots, n,$$

and

$$\Sigma := \mathbf{diag}(\sigma_1, \ldots, \sigma_n) \in \mathbb{C}^{m\times n}.$$

Then,

$$\Sigma\Sigma^{\mathsf{H}} = \Lambda_1 \in \mathbb{R}^{m\times m} \quad \text{and} \quad \Sigma^{\mathsf{H}}\Sigma = \Lambda_2 \in \mathbb{R}^{n\times n}.$$

Proof, Cont.

Now, let us examine the decomposition

$$\mathsf{A}^{\mathsf{H}}\mathsf{A} = \mathsf{V}\boldsymbol{\Sigma}^{\top}\boldsymbol{\Sigma}\mathsf{V}^{\mathsf{H}},$$

as given above. It follows that

$$\mathsf{A}^{\mathsf{H}}\mathsf{A}\mathbf{v}_i = \sigma_i^2\mathbf{v}_i, \quad i = 1, \ldots, n,$$

where $\mathbf{v}_k$ is the k^{th} column of V. Let $r \in \{1, \ldots, n\}$ be the unique index satisfying

$$\sigma_r > 0 \quad \text{and} \quad \sigma_{r+1} = 0.$$

Except for the trivial case, there must be such an r. Define

$$\mathbf{u}_k = \frac{\mathsf{A}\mathbf{v}_k}{\sigma_k}, \quad k = 1, \ldots, r.$$

Proof, Cont.

Then

$$\mathsf{A}\mathsf{A}^{\mathsf{H}}\mathbf{u}_k = \mathsf{A}\left(\frac{\mathsf{A}^{\mathsf{H}}\mathsf{A}\mathbf{v}_k}{\sigma_k}\right) = \mathsf{A}\left(\sigma_k \mathbf{v}_k\right) = \sigma_k^2 \mathbf{u}_k, \quad k = 1, \ldots, r.$$

Thus the constructed vectors $\mathbf{u}_k$ can be identified as eigenvectors of $\mathsf{A}\mathsf{A}^{\mathsf{H}}$. They are also orthonormal: for $i, j \in \{1, \ldots, r\}$,

$$\mathbf{u}_i^{\mathsf{H}}\mathbf{u}_j = \frac{\mathbf{v}_i^{\mathsf{H}}\mathsf{A}^{\mathsf{H}}\mathsf{A}\mathbf{v}_j}{\sigma_i \sigma_j} = \frac{\sigma_j}{\sigma_i}\mathbf{v}_i^{\mathsf{H}}\mathbf{v}_j = \delta_{i,j}.$$

Therefore, the constructed vectors $\mathbf{u}_k$ can be identified as the first r columns of the matrix U that we found earlier. The remaining vectors $\mathbf{u}_{r+1}, \ldots \mathbf{u}_m$ can be obtained, via basis completion, so that

$$\mathsf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_m] \in \mathbb{C}^{m \times m}$$

is unitary, as desired. We have, finally,

$$\mathsf{A} = \mathsf{U}\boldsymbol{\Sigma}\mathsf{V}^{\mathsf{H}}.$$

$\square$

Uniqueness of Singular Vectors

Theorem

Suppose that $\mathsf{A} \in \mathbb{C}^{m\times n}$ *with* $m \geq n$. *If*

$$\mathsf{A} = \mathsf{U}_1\mathbf{\Sigma}_1\mathsf{V}_1^{\mathsf{H}} = \mathsf{U}_2\mathbf{\Sigma}_2\mathsf{V}_2^{\mathsf{H}}$$

are two SVDs for A, *then* $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$, *the columns of* V_1 *and* V_2 *form an orthonormal basis of eigenvectors of* $\mathsf{A}^{\mathsf{H}}\mathsf{A}$, *and, if* $\mathsf{A}^{\mathsf{H}}\mathsf{A}$ *has* n *distinct eigenvalues, then*

$$\mathsf{V}_1 = \mathsf{V}_2\mathsf{D},$$

for some $\mathsf{D} = \mathbf{diag}[e^{\mathrm{i}\theta_1}, \ldots, e^{\mathrm{i}\theta_n}]$, *with angles* $\theta_i \in \mathbb{R}$, $i = 1, \ldots, n$. *Finally, if* $\mathbf{rank}(\mathsf{A}) = n$ *(*A *has full rank), and*

$$\mathsf{A} = \mathsf{U}_1\mathbf{\Sigma}\mathsf{V}^{\mathsf{H}} = \mathsf{U}_2\mathbf{\Sigma}\mathsf{V}^{\mathsf{H}}$$

are two SVDs for A, *then the first* n *columns of* U_1 *and* U_2 *are equal.*

Further Properties of the SVD

SVD and Rank

Theorem

Let $\mathsf{A} \in \mathbb{C}^{m \times n}$*, then* $\mathbf{rank}(\mathsf{A})$ *coincides with the number of nonzero singular values.*

Proof.

We write the SVD: $\mathsf{A} = \mathsf{U}\boldsymbol{\Sigma}\mathsf{V}^{\mathsf{H}}$. Since U and V are unitary they are full rank. By a theorem, the rank is unchanged by multiplying by invertible (square, full rank) matrices. Thus, $\mathbf{rank}(\mathsf{A}) = \mathbf{rank}(\boldsymbol{\Sigma})$ and since $\boldsymbol{\Sigma}$ is diagonal the assertion follows. □

Range and Kernel Through SVD

Theorem

Let $\mathsf{A} \in \mathbb{C}^{m\times n}$ *with* $\mathsf{rank}(\mathsf{A}) = r$. *Suppose an SVD for* A *is given by* $\mathsf{A} = \mathsf{U}\boldsymbol{\Sigma}\mathsf{V}^{\mathsf{H}}$, *where* $\mathbf{u}_1, \ldots, \mathbf{u}_m$ *denote the columns of* U *and* $\mathbf{v}_1, \ldots, \mathbf{v}_n$ *denote the columns of* V. *Then*

$$\langle \mathbf{u}_1, \ldots, \mathbf{u}_r \rangle = \mathsf{im}(\mathsf{A}) \quad \textit{and} \quad \langle \mathbf{v}_{r+1}, \ldots, \mathbf{v}_n \rangle = \mathsf{ker}(\mathsf{A}).$$

Proof.

$(\subseteq)$: From the SVD one can easily write $\mathsf{A}\mathbf{v}_i = \sigma_i \mathbf{u}_i$, for $i = 1, \ldots, r$. This proves immediately that $\mathbf{u}_i \in \mathsf{im}(\mathsf{A})$, for $i = 1, \ldots, r$. Since $\mathsf{im}(\mathsf{A})$ is a subspace of $\mathbb{C}^m$, any linear combination of $\mathbf{u}_1, \ldots, \mathbf{u}_r$ is in $\mathsf{im}(\mathsf{A})$. Hence $\langle \mathbf{u}_1, \ldots, \mathbf{u}_r \rangle \subseteq \mathsf{im}(\mathsf{A})$.

Proof, Cont.

($\supseteq$): Let $\mathbf{y} \in \mathrm{im}(\mathsf{A})$. Then $\exists\ \mathbf{x} \in \mathbb{C}^n$ such that $\mathsf{A}\mathbf{x} = \mathbf{y}$. This implies that $\mathsf{U}\Sigma\mathsf{V}^{\mathsf{H}}\mathbf{x} = \mathbf{y}$ for some $\mathbf{x}$. Let $\mathbf{x}' = \mathsf{V}^{\mathsf{H}}\mathbf{x}$. Then for some $\mathbf{x}' \in \mathbb{C}^n$, $\mathsf{U}\boldsymbol{\Sigma}\mathbf{x}' = \mathbf{y}$. Set $\mathbf{x}'' = \boldsymbol{\Sigma}\mathbf{x}'$. Note that $\mathbf{x}'' \in \mathbb{C}^m$ and $x''_{r+1} = \cdots = x''_m = 0$. Hence for some $\mathbf{x}'' \in \mathbb{C}^m$, $\mathsf{U}\mathbf{x}'' = \mathbf{y}$. Now we write

$$\mathbf{y} = \mathsf{U}\mathbf{x}'' = \sum_{j=1}^{m} x''_j \mathbf{u}_j = \sum_{j=1}^{r} x''_j \mathbf{u}_j \in \langle \mathbf{u}_1, \ldots, \mathbf{u}_r \rangle \ .$$

This proves that $\langle \mathbf{u}_1, \ldots, \mathbf{u}_r \rangle \supseteq \mathrm{im}(\mathsf{A})$, and we are done.

($\subseteq$): From the SVD one can easily write $\mathsf{A}\mathbf{v}_i = \mathbf{0}$, for $i = r+1, \ldots, n$. This proves immediately that $\mathbf{v}_i \in \ker(\mathsf{A})$, for $i = r+1, \ldots, n$. Since $\ker(\mathsf{A})$ is subspace of $\mathbb{C}^n$, any linear combination of $\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n$ is in $\ker(\mathsf{A})$. Hence $\langle \mathbf{v}_{r+1}, \ldots, \mathbf{v}_n \rangle \subseteq \ker(\mathsf{A})$.

Proof, Cont.

($\supseteq$): Let $\mathbf{x} \in \ker(\mathsf{A})$. Then $\mathsf{A}\mathbf{x} = \mathbf{0}$. This implies that $\mathsf{U}\mathbf{\Sigma}\mathsf{V}^{\mathsf{H}}\mathbf{x} = \mathbf{0}$. Let $\mathbf{x}' = \mathsf{V}^{\mathsf{H}}\mathbf{x}$. This implies that $\mathbf{x} = \mathsf{V}\mathbf{x}'$ and $\mathsf{U}\mathbf{\Sigma}\mathbf{x}' = \mathbf{0}$. Since U is invertible, this implies that $\mathbf{\Sigma}\mathbf{x}' = \mathbf{0}$. This homogeneous system always has a solution of the form

$$\mathbf{x}' = \begin{bmatrix} x_1' = 0 \\ \vdots \\ x_r' = 0 \\ x_{r+1}' = \alpha_{r+1} \\ \vdots \\ x_n' = \alpha_n \end{bmatrix}$$

where $\alpha_{r+1}, \ldots, \alpha_n$ are arbitrary. But this shows that

$$\mathbf{x} = \mathsf{V}\mathbf{x}' = \sum_{j=1}^{n} x_j' \mathbf{v}_j = \sum_{j=r+1}^{n} \alpha_j \mathbf{v}_j \in \langle \mathbf{v}_{r+1}, \ldots, \mathbf{v}_n \rangle .$$

This proves that $\langle \mathbf{v}_{r+1}, \ldots, \mathbf{v}_n \rangle \supseteq \ker(\mathsf{A})$, and we are done. $\square$

Rank–Plus–Nullity Theorem

Corollary

Suppose that $\mathsf{A} \in \mathbb{C}^{m\times n}$. *Then,*

$$\operatorname{rank}(\mathsf{A}) + \operatorname{nullity}(\mathsf{A}) = \dim(\operatorname{im}(\mathsf{A})) + \dim(\ker(\mathsf{A})) = n.$$

Proof.

Set $r = \operatorname{rank}(\mathsf{A})$. The last theorem guarantees that

$$\langle \mathbf{u}_1, \ldots, \mathbf{u}_r \rangle = \operatorname{im}(\mathsf{A}) \quad \text{and} \quad \langle \mathbf{v}_{r+1}, \ldots, \mathbf{v}_n \rangle = \ker(\mathsf{A}).$$

□

When is $A^H A$ is Nonsingular?

Theorem

Suppose $A \in \mathbb{C}^{m \times n}$ *with* $m \geq n$. $A^H A$ *is nonsingular iff* $\operatorname{rank}(A) = n$.

Proof.

Let $A = U\Sigma V^H$ be an SVD for A. Then $A^H A = V\Sigma^\intercal \Sigma V^H$ yields a unitary diagonalization of $A^H A$. Note that

$$\Sigma^\intercal \Sigma = \operatorname{diag}\left[\sigma_1^2, \ldots, \sigma_n^2\right] \;,$$

where $\sigma_1, \ldots, \sigma_n$ are the singular values of A, some of which may be zero. It is clear that $\operatorname{rank}(A) = r$, where r is the number of nonzero singular values. Of course, it must be that $r \leq n$. Likewise, $\operatorname{rank}(A^H A)$ is the number of nonzero elements on the diagonal of $\Sigma^H \Sigma$. This number must also be r. In other words,

$$\operatorname{rank}(A) = r = \operatorname{rank}(A^H A),$$

which proves the result. □

Some Further Properties

Theorem (SVD and norms)

Let $\mathsf{A} \in \mathbb{C}^{m \times n}$, *then* $\|\mathsf{A}\|_2 = \sigma_1$ *and* $\|\mathsf{A}\|_F^2 = \sum_{i=1}^r \sigma_i^2$.

Theorem (SVD and self–adjoint matrices)

If A *is Hermitian, that is,* $\mathsf{A}^{\mathsf{H}} = \mathsf{A}$, *then the singular values of* A *are the absolute values of its eigenvalues. If* A *is Hermitian with non–negative eigenvalues, then the eigenvalues and the singular values coincide.*

Theorem (SVD and determinants)

Let $\mathsf{A} \in \mathbb{C}^{n \times n}$, *then*

$$|\det(\mathsf{A})| = \prod_{i=1}^n \sigma_i.$$

Low Rank Approximations

Rank–One Matrices

Definition

Given $\mathbf{u} \in \mathbb{C}^m_\star$, $\mathbf{v} \in \mathbb{C}^n_\star$ *and* $\sigma \in \mathbb{C}_\star$, *the matrix* $\mathsf{A} \in \mathbb{C}^{m\times n}$ *defined via*

$$\mathsf{A} = \sigma \mathbf{u}\mathbf{v}^{\mathsf{H}} = \sigma \mathbf{u} \otimes \mathbf{v},$$

is called a **rank–one matrix**.

For every $\mathbf{x} \in \mathbb{C}^n$, the rank–one matrix A, defined above, acts as

$$\mathsf{A}\mathbf{x} = \sigma \left(\mathbf{v}^{\mathsf{H}}\mathbf{x}\right) \mathbf{u} \in \operatorname{span}\{\mathbf{u}\}.$$

The rank of such a matrix is exactly equal to one. This justifies the name. The question we want to address now is whether *every* matrix can be represented (or at least approximated) by linear combinations of rank–one matrices.

Rank–One Decomposition Theorem

Theorem

Let $\mathsf{A} \in \mathbb{C}^{m \times n}$ *be such that* $r = \mathbf{rank}(\mathsf{A})$ *and* $\mathsf{A} = \mathsf{U}\mathsf{\Sigma}\mathsf{V}^{\mathsf{H}}$ *be a SVD. Then* A *is a linear combination of* r *rank–one matrices:*

$$\mathsf{A} = \sum_{j=1}^{r} \sigma_j \mathbf{u}_j \otimes \mathbf{v}_j .$$

Proof.

It suffices to write $\mathsf{\Sigma}$ as the sum of matrices of the form

$$\mathsf{\Sigma}_j = \mathbf{diag}[0, \ldots, 0, \sigma_j, 0, \ldots, 0],$$

where the element σ_j is in the j–th entry. The rest of the details are left as an exercise. □

Eckart–Young Theorem

Theorem

Let $\mathsf{A} \in \mathbb{C}^{m \times n}$ *be such that* $r = \operatorname{rank}(\mathsf{A})$. *Let* $\mathsf{A} = \mathsf{U}\boldsymbol{\Sigma}\mathsf{V}^{\mathsf{H}}$ *be an SVD of* A. *For* $k < r$ *define* $\mathsf{A}_k = \sum_{j=1}^{k} \sigma_j \mathbf{u}_j \otimes \mathbf{v}_j$. *Let us denote by* $\mathcal{C}_k$ *the collection of all matrices* $\mathsf{B} \in \mathbb{C}^{m \times n}$ *such that* $\operatorname{rank}(\mathsf{B}) \leq k$. *Then,* A_k *is of rank* k*, and*

$$\|\mathsf{A} - \mathsf{A}_k\|_2 = \sigma_{k+1} = \inf_{\mathsf{B} \in \mathcal{C}_k} \|\mathsf{A} - \mathsf{B}\|_2.$$

Furthermore,

$$\|\mathsf{A} - \mathsf{A}_k\|_F = \sqrt{\sum_{j=k+1}^{r} \sigma_j^2} = \inf_{\mathsf{B} \in \mathcal{C}_k} \|\mathsf{A} - \mathsf{B}\|_F.$$

Example

Consider the 7×7 matrix

$$A = \begin{bmatrix} 2.9057 & 6.5457 & 3.8587 & 4.6737 & 2.4171 & 3.8703 & 0.9922 \\ 2.6460 & 0.8889 & 3.2542 & 1.7574 & 6.8099 & 1.4357 & 6.2114 \\ 2.1268 & 3.4270 & 2.6255 & 2.6030 & 3.0228 & 2.1093 & 2.2459 \\ 2.1686 & 2.6778 & 3.8653 & 3.6416 & 3.6917 & 3.3553 & 2.2689 \\ 3.8927 & 6.6301 & 6.4863 & 6.8192 & 5.0281 & 6.0910 & 2.5521 \\ 4.0461 & 6.2798 & 3.1713 & 2.7880 & 6.1632 & 1.6456 & 5.8120 \\ 2.9612 & 1.9820 & 6.1018 & 5.1754 & 6.4646 & 5.0316 & 4.2031 \end{bmatrix}.$$

The singular values are, to 13 decimal digits of precision,

$$\begin{aligned} \sigma_1 &= 27.7754505112764, \\ \sigma_2 &= 08.0248423105149, \\ \sigma_3 &= 05.2245562622115, \\ \sigma_4 &= 00.0001965858656, \\ \sigma_5 &= 00.0000856660061, \\ \sigma_6 &= 00.0000628919629, \\ \sigma_7 &= 00.0000071697992. \end{aligned}$$

Example (Example, Cont.)

This matrix is very nearly singular and is well approximated by the rank–3 (compressed) matrix

$$\mathsf{A}_3 = \sum_{i=1}^{3} \sigma_i \mathbf{u}_i \mathbf{v}_i^{\intercal},$$

where $\mathbf{u}_i, \mathbf{v} \in \mathbb{R}^7$ are the singular vectors, which are suppressed for brevity. In other words, to a good approximation, there are really only 3 important components of A — namely, $\sigma_i \mathbf{u}_i \mathbf{v}_i^{\intercal}$, $i = 1, 2, 3$ — that express its action. In particular, according to the Eckart–Young Theorem 3.3, the relative error in the compressed matrix is relatively small,

$$\frac{\|\mathsf{A} - \mathsf{A}_3\|_2}{\|\mathsf{A}\|_2} = \frac{\sigma_4}{\sigma_1} = 7.07768 \times 10^{-06}.$$