Vv417 Lecture 25

Jing Liu

UM-SJTU Joint Institute

December 2, 2019

- We have discussed that every Hermitian matrix $\mathbf{A}$ can be factored as

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{H}}$$

where $\mathbf{U}$ is an $n \times n$ unitary matrix of eigenvectors of $\mathbf{A}$, and $\mathbf{D}$ is the diagonal matrix whose entries are the corresponding eigenvalues.

- This factorization is often called the eigenvalue decomposition (EVD) of $\mathbf{A}$.
- We have also discussed that non-Hermitian matrices can also have eigenvalue decomposition of the form above as long as they are normal.
- We have also considered when the matrix is not normal. A similar expansion can be obtained using Schur decomposition as long as the matrix is square,

$$\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^{\mathrm{H}}$$

where $\mathbf{U}$ is an $n \times n$ unitary matrix and $\mathbf{T}$ is an upper triangular matrix.

- The eigenvalue and Schur decompositions are important in practice.
- Especially, in numerical algorithms when $\mathbf{A}$ is large, this is not only because the matrices $\mathbf{D}$, and $\mathbf{T}$ have simpler forms than $\mathbf{A}$, but also because of the unitary matrix $\mathbf{U}$ in these factorizations do not magnify roundoff error.

- To see why this is so, let $\hat{\mathbf{x}}$ be a vector whose entries are known exactly

$$\mathbf{x} = \hat{\mathbf{x}} + \mathbf{e}$$

is the vector that results when roundoff error $\mathbf{e}$ is present.

- If $\mathbf{U}$ is a unitary matrix, then it is length-preserving

$$\begin{aligned} \|\mathbf{U}\mathbf{x} - \mathbf{U}\hat{\mathbf{x}}\| &= \|\mathbf{U}(\mathbf{x} - \hat{\mathbf{x}})\| \\ &= (\mathbf{x} - \hat{\mathbf{x}})^{\mathrm{H}} \mathbf{U}^{\mathrm{H}} \mathbf{U} (\mathbf{x} - \hat{\mathbf{x}}) \\ &= \|\mathbf{x} - \hat{\mathbf{x}}\| \\ &= \|\mathbf{e}\| \end{aligned}$$

which shows that the error in computing

$$\mathbf{U}\hat{\mathbf{x}} \quad \text{by computing} \quad \mathbf{U}\mathbf{x}$$

has the same magnitude as the error in approximating $\hat{\mathbf{x}}$ by $\mathbf{x}$.

- There are two main paths that one might follow in looking for other kinds of decompositions of a general square matrix $\mathbf{A}$:
  1. One might look for decompositions of the form
  $$\mathbf{A} = \mathbf{P}\mathbf{J}\mathbf{P}^{-1}$$
  in which $\mathbf{P}$ is invertible but not necessarily having orthogonal columns.
  2. Alternatively, one might look for decompositions of the form
  $$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{H}}$$
  in which $\mathbf{U}$ and $\mathbf{V}$ are orthogonal but not necessarily the same.
- The first path leads to decompositions in which $\mathbf{J}$ is either

  diagonal or a certain kind of block diagonal matrix

  called a Jordan canonical form.
- Jordan canonical forms are important theoretically and in certain applications, but they are of lesser importance numerically because of the roundoff problems that result from the lack of orthogonality in $\mathbf{P}$.

Theorem

If $\mathbf{A}$ is an $m \times n$ matrix, then

1. $\mathbf{A}$ and $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ have the same null space.
2. $\mathbf{A}$ and $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ have the same row space.
3. $\mathbf{A}^{\mathrm{T}}$ and $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ have the same column space.
4. $\mathbf{A}$ and $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ have the same rank.

Proof

- Let us consider statement one, we need to show $\mathbf{x}_0$ is a solution of

$$\mathbf{A}\mathbf{x} = \mathbf{0}$$

if and only if $\mathbf{x}_0$ is a solution of

$$\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{x} = \mathbf{0}$$

Proof

- If $\mathbf{x}_0$ is a solution of $\mathbf{A}\mathbf{x} = \mathbf{0}$, then

$$\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{x}_0 = \mathbf{A}^{\mathrm{T}}\mathbf{0} = \mathbf{0}$$

- Conversely, if $\mathbf{x}_0$ is any solution of $\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{x} = \mathbf{0}$, then

$$\mathbf{x}_0 \in \mathrm{null}(\mathbf{A}^{\mathrm{T}}\mathbf{A})$$

- Hence $\mathbf{x}_0$ is orthogonal to all vectors in the row space of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$, thus

$$\mathbf{x}_0 \in \mathrm{null}(\mathbf{A}^{\mathrm{T}}\mathbf{A}) = \mathrm{col}\left(\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{\mathrm{T}}\right)^{\perp} = \mathrm{col}\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{\perp}$$

- So $\mathbf{x}_0$ is also orthogonal to every vector in the column space of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$.
- In particular, it is orthogonal to $\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)\mathbf{x}_0$, thus

$$\mathbf{x}_0^{\mathrm{T}}\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)\mathbf{x}_0 = 0 \implies (\mathbf{A}\mathbf{x}_0)^{\mathrm{T}}(\mathbf{A}\mathbf{x}_0) = 0 \implies \mathbf{A}\mathbf{x}_0 = \mathbf{0}$$

Theorem

If $\mathbf{A}$ is an $m \times n$ matrix, then

1. $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ is orthogonally diagonalizable
2. The eigenvalues of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ are nonnegative

Proof

- The matrix $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ is symmetric, so can be orthogonally diagonalized, thus there is an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$,

$$\mathcal{B} = \{\mathbf{v}_1,\ \mathbf{v}_2, \cdots,\ \mathbf{v}_n\}$$

corresponding to eigenvectors λ_1, λ_2, …, λ_n.

$$\begin{aligned}\|\mathbf{A}\mathbf{v}_i\|^2 = (\mathbf{A}\mathbf{v}_i)^{\mathrm{T}}(\mathbf{A}\mathbf{v}_i) = \mathbf{v}_i^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{v}_i = \mathbf{v}_i^{\mathrm{T}}\lambda_i\mathbf{v}_i \\ = \lambda_i\|\mathbf{v}_i\|^2 = \lambda_i \geq 0 \quad \text{for all } i.\end{aligned}$$

Definition

If $\mathbf{A}$ is an $m \times n$ matrix, and if λ_1, λ_2, ..., λ_n are the eigenvalues of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$, then

$$\sigma_1 = \sqrt{\lambda_1}, \quad \sigma_2 = \sqrt{\lambda_2}, \ldots, \quad \sigma_n = \sqrt{\lambda_n}$$

are called the singular values of $\mathbf{A}$.

- Let's extend the notion of a "main diagonal" to matrices that are not square

$$\begin{bmatrix} \boxed{*} & * & * & * & * & * & * \\ * & \boxed{*} & * & * & * & * & * \\ * & * & \boxed{*} & * & * & * & * \\ * & * & * & \boxed{*} & * & * & * \end{bmatrix} \qquad \begin{bmatrix} \boxed{*} & * & * & * \\ * & \boxed{*} & * & * \\ * & * & \boxed{*} & * \\ * & * & * & \boxed{*} \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{bmatrix}$$

- We define the main diagonal of an $m \times n$ matrix to be the line of entries that start at the upper left corner and extends diagonally as far as it can go, and we will refer to the entries on the main diagonal as the diagonal entries.

Singular Value Decomposition SVD

If $\mathbf{A}$ is an $m \times n$ of rank k, then $\mathbf{A}$ can be factored as

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{T}}$$

$$= \left[\begin{array}{ccc|ccc} \mathbf{u}_1 & \cdots & \mathbf{u}_k & \mathbf{u}_{k+1} & \cdots & \mathbf{u}_m \end{array}\right] \left[\begin{array}{ccc|c} \sigma_1 & \cdots & 0 & \\ \vdots & \ddots & \vdots & \mathbf{0}_{k\times(n-k)} \\ 0 & \cdots & \sigma_k & \\ \hline & \mathbf{0}_{(m-k)\times k} & & \mathbf{0}_{(m-k)\times(n-k)} \end{array}\right] \left[\begin{array}{c} \mathbf{v}_1^{\mathrm{T}} \\ \vdots \\ \mathbf{v}_k^{\mathrm{T}} \\ \hline \mathbf{v}_{k+1}^{\mathrm{T}} \\ \vdots \\ \mathbf{v}_n^{\mathrm{T}} \end{array}\right]$$

in which $\mathbf{U}$, $\boldsymbol{\Sigma}$ and $\mathbf{V}$ are matrices of size $m \times m$, $m \times n$, and $n \times n$, respectively.

- The matrix $\mathbf{V}$ unitarily diagonalize $\mathbf{A}^{\mathrm{T}}\mathbf{A}$, the columns of $\mathbf{V}$ are ordered so that the corresponding eigenvalues are in order of decreasing size.

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$$

Singular Value Decomposition SVD

- The nonzero diagonal entries of $\mathbf{\Sigma}$ are nonzero singular values of $\mathbf{A}$,

$$\sigma_1 = \sqrt{\lambda_1}, \quad \sigma_2 = \sqrt{\lambda_2}, \ldots, \quad \sigma_k = \sqrt{\lambda_k}$$

where λ_i are the nonzero eigenvalues of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ in order of decreasing size, so

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k > 0$$

- Vector $\mathbf{u}_j$ is defined as the normalized image of $\mathbf{v}_j$ under $\mathbf{A}$

$$\mathbf{u}_j = \frac{\mathbf{A}\mathbf{v}_j}{\|\mathbf{A}\mathbf{v}_j\|} = \frac{1}{\sigma_j}\mathbf{A}\mathbf{v}_j \qquad \text{for } j = 1, 2, \ldots, k$$

- The set $\{\mathbf{u}_{k+1}, \ldots, \mathbf{u}_m\}$ is an extension set of $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ so that

$$\{\mathbf{u}_1, \ldots \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_m\}$$

forms an orthonormal basis for $\mathbb{R}^m$.

Proof

- For notational simplicity we will prove this in the case where $\mathbf{A}$ is a square matrix of $n \times n$. To modify the argument for an $m \times n$ matrix you need only make the notational adjustment required to account for the possibility that

$$m > n \qquad \text{or} \qquad m < n$$

- The matrix $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ is symmetric, so it has eigenvalue decomposition

$$\mathbf{A}^{\mathrm{T}}\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{V}^{\mathrm{T}}$$

- Since $\mathbf{A}$ is assume to have rank k, then $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ also has rank k, it follows that $\mathbf{D}$ as well has rank k since $\mathbf{D}$ and $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ are similar.

$$\mathbf{D} = \begin{bmatrix} \lambda_1 & \cdots & \cdots & 0 \\ \vdots & \lambda_2 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix} \qquad \text{where} \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k > 0$$

Proof

- Take the set of image vectors, $\{\mathbf{A}\mathbf{v}_1, \mathbf{A}\mathbf{v}_2, \ldots \mathbf{A}\mathbf{v}_n\}$, and consider,

$$(\mathbf{A}\mathbf{v}_i)^{\mathrm{T}}\mathbf{A}\mathbf{v}_j = \mathbf{v}_i^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{v}_j = \mathbf{v}_i^{\mathrm{T}}\lambda_j\mathbf{v}_j = \lambda_j\mathbf{v}_i^{\mathrm{T}}\mathbf{v}_j$$

- The orthogonality of $\mathbf{v}_i$ and $\mathbf{v}_j$ implies

$$(\mathbf{A}\mathbf{v}_i)^{\mathrm{T}}\mathbf{A}\mathbf{v}_j = \lambda_j\mathbf{v}_i^{\mathrm{T}}\mathbf{v}_j = \lambda_j \cdot 0 = 0 \qquad \text{for } i \neq j.$$

- The first k image vectors $\mathbf{A}\mathbf{v}_i$ are nonzero for we have shown $\|\mathbf{A}\mathbf{v}_i\|^2 = \lambda_i$, and the first k eigenvalues are nonzero. Therefore

$$\mathcal{S} = \{\mathbf{A}\mathbf{v}_1, \mathbf{A}\mathbf{v}_2, \cdots, \mathbf{A}\mathbf{v}_k\}$$

  is an orthogonal set of nonzero vectors in $\mathrm{col}(\mathbf{A})$ since $\mathbf{A}\mathbf{v}_i \in \mathrm{col}(\mathbf{A})$ for $\forall i$.

- The column space of $\mathbf{A}$ has dimension k, hence $\mathcal{S}$ being orthogonal and thus linearly independent set of k vectors must be an orthogonal basis for $\mathrm{col}\,(\mathbf{A})$.

Proof

- If we now normalize the vectors in $\mathcal{S}$,

$$\mathbf{u}_j = \frac{\mathbf{A}\mathbf{v}_j}{\|\mathbf{A}\mathbf{v}_j\|} = \frac{1}{\sqrt{\lambda_j}}\mathbf{A}\mathbf{v}_j = \frac{1}{\sigma_j}\mathbf{A}\mathbf{v}_j \implies \sigma_j \mathbf{u}_j = \mathbf{A}\mathbf{v}_j \quad \text{for } j = 1, 2, \ldots, k$$

- And if we extend $\mathcal{S}$ to an orthonormal basis, say by using Gram-Schmidt

$$\{\underbrace{\mathbf{u}_1, \ldots, \mathbf{u}_k}_{\mathcal{S}}, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\} \quad \text{for } \mathbb{R}^n$$

- Now let $\mathbf{U} = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_k & \mathbf{u}_{k+1} & \cdots & \mathbf{u}_n \end{bmatrix}$ and $\mathbf{\Sigma} = \begin{bmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, then

$$\begin{aligned} \mathbf{U}\mathbf{\Sigma} &= \begin{bmatrix} \sigma_1 \mathbf{u}_1 & \cdots & \sigma_k \mathbf{u}_k & 0\mathbf{u}_{k+1} & \cdots & 0\mathbf{u}_n \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{A}\mathbf{v}_1 & \cdots & \mathbf{A}\mathbf{v}_k & \mathbf{A}\mathbf{v}_{k+1} & \cdots & \mathbf{A}\mathbf{v}_n \end{bmatrix} = \mathbf{A}\mathbf{V} \implies \mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}} \end{aligned}$$

□

since $\mathbf{V}$ is orthogonal.

Exercise

Compute the singular values and the singular value decomposition of

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

Solution

- Since $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ has eigenvalues $\lambda_1 = 4$ and $\lambda_2 = 0$, the singular values of $\mathbf{A}$ are

$$\sigma_1 = 2 \quad \text{and} \quad \sigma_2 = 0$$

- The corresponding eigenvectors of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ are $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$, so

$$\mathbf{V} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

- The rank of $\mathbf{A}$ is clearly 1, so $\mathbf{u}_1 = \frac{1}{\sigma_1}\mathbf{A}\mathbf{v}_1 = \frac{1}{\sqrt{2}}(\mathbf{e}_1 + \mathbf{e}_2)$

Solution

- The remaining column vectors of $\mathbf{U}$ must form an orthonormal basis for $\text{null}(\mathbf{A}^{\text{T}})$, we can compute a basis $\{\mathbf{x}_2, \mathbf{x}_3\}$ for $\text{null}(\mathbf{A}^{\text{T}})$ in the usual way

$$\mathbf{x}_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

- These vectors are already orthogonal, so we can skip Gram-Schmidt,

$$\mathbf{u}_2 = \frac{1}{\|\mathbf{x}_2\|}\mathbf{x}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{u}_3 = \mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

- Therefore $\mathbf{A} = \mathbf{U\Sigma V}^{\text{T}} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$

Q: What does the rank of a matrix $\mathbf{A}$ tell us?

- In terms of a linear equation

$$\mathbf{A}\mathbf{x} = \mathbf{b},$$

it gives the number of conditions that are imposed on the variables.

- In terms of transformation,

$$T\colon \mathcal{U} \to \mathcal{V}, \qquad \text{where } \mathbf{A} = [T] \text{ is the coordinate matrix of } T,$$

it gives the dimension of the range of T.

Q: What does the rank of a matrix

$$\mathbf{A}$$

tell us if the matrix is simply a way of storing information?

- SVD gives another interpretation of the rank of a matrix.

- In the case that $\mathbf{A}$ has rank $k < n$, it is often sufficient to consider

$$\begin{aligned}\mathbf{A} &= \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}} \\ &= \sigma_1\mathbf{u}_1\mathbf{v}_1^{\mathrm{T}} + \sigma_2\mathbf{u}_2\mathbf{v}_2^{\mathrm{T}} + \cdots + \sigma_k\mathbf{u}_k\mathbf{v}_k^{\mathrm{T}} + \underbrace{\sigma_{k+1}\mathbf{u}_{k+1}\mathbf{v}_{k+1}^{\mathrm{T}} + \cdots + \sigma_n\mathbf{u}_n\mathbf{v}_n^{\mathrm{T}}}_{\mathbf{0}_{(n-k)\times(n-k)}} \\ &= \mathbf{U}_k\mathbf{\Sigma}_k\mathbf{V}_k^{\mathrm{T}} \quad \text{where} \quad \mathbf{U}_k = \begin{bmatrix}\mathbf{u}_1 & \cdots & \mathbf{u}_k\end{bmatrix}, \mathbf{V}_k = \begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_k\end{bmatrix}\end{aligned}$$

  and $\mathbf{\Sigma}_k$ is a diagonal matrix containing all the nonzero singular values of $\mathbf{A}$.
- The above is called the compact form of the singular value decomposition, which is one type of reduced singular value decompositions RSVD.
- RSVD can be used to "compress" digital information.
- For example, a black and white photograph might be stored as a matrix

$$\mathbf{A}$$

  where each entry stores information about a pixel of the picture, a numerical value between 0 and 255 in accordance with the pixel's grey level.

- If the matrix $\mathbf{A}$ has size $m \times n$, then one might store every $m \times n$ elements
- Alternatively, we can compute the reduced singular value decomposition,

$$\mathbf{A} = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^{\mathrm{T}} + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^{\mathrm{T}} + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^{\mathrm{T}} = \mathbf{U}_r \boldsymbol{\Sigma}_r \mathbf{V}_r^{\mathrm{T}}$$

  in which $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq \sigma_{r+1} \geq \cdots \geq \sigma_n$,
- Instead storing the matrix $\mathbf{A}$, store the σ_j's, and the $\mathbf{u}_j$'s and the $\mathbf{v}_j$'s.
- When needed, the matrix $\mathbf{A}$ can be reconstructed. Since each $\mathbf{u}_j$ has m entries and each $\mathbf{v}_j$ has n entries, this method requires storage space for

$$rm + rn + r = r(m + n + 1)$$

- We call the above the rank r approximation of $\mathbf{A}$.
- The size of the singular value σ_j of a given matrix $\mathbf{A}$ gives the weight of

$$\mathbf{u}_j \mathbf{v}_j^{\mathrm{T}}$$

- Therefore the RSVD can also be a useful tool when we need to sort through noisy data and lift out relevant information.

- Recall diagonalizing a matrix or a linear system simplifies the problem, e.g.

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

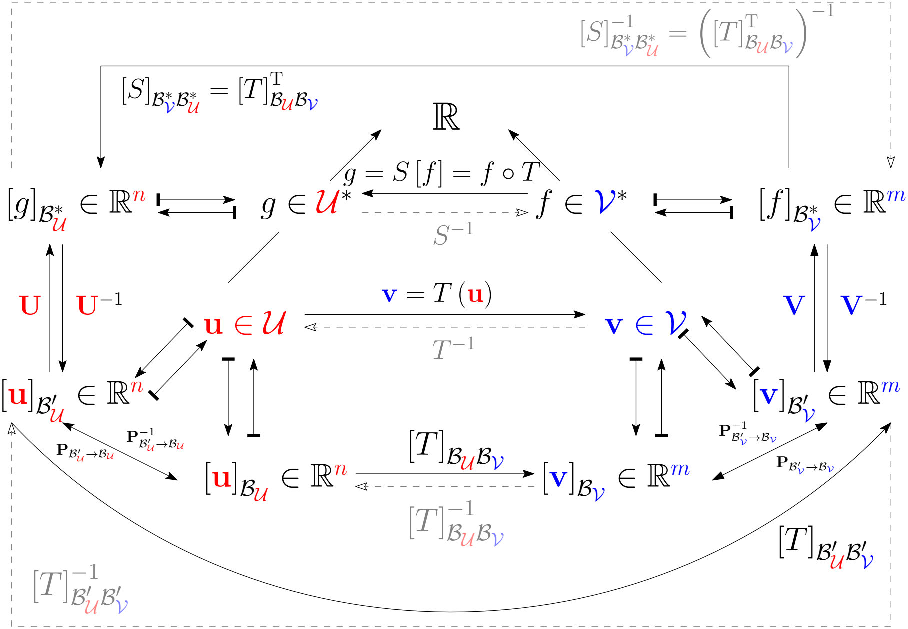
- If $\mathbf{A}$ is square and diagonalizable, then

$$\mathbf{P}\mathbf{D}\mathbf{P}^{-1}\mathbf{x} = \mathbf{b} \implies \mathbf{D}\left(\mathbf{P}^{-1}\mathbf{x}\right) = \left(\mathbf{P}^{-1}\mathbf{b}\right)$$

- With singular value decomposition, we can diagonalize a rectangular system,

$$\begin{aligned} \mathbf{A}\mathbf{x} &= \mathbf{b} \qquad\qquad \text{where } \mathbf{A} \in \mathbb{R}^{m\times n}. \\ \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{T}}\mathbf{x} &= \mathbf{b} \\ \boldsymbol{\Sigma}\left(\mathbf{V}^{\mathrm{T}}\mathbf{x}\right) &= \left(\mathbf{U}^{\mathrm{T}}\mathbf{b}\right) \end{aligned}$$

- Using the orthonormal bases $\mathcal{B}_{\mathcal{U}} = \{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ and $\mathcal{B}_{\mathcal{V}} = \{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ for the range and the domain of $T_{\mathbf{A}}$, respectively, converts $\mathbf{A}[\mathbf{x}]_{\mathcal{S}} = [\mathbf{b}]_{\mathcal{S}}$ to a block system $\boldsymbol{\Sigma}[\mathbf{x}]_{\mathcal{B}_{\mathcal{V}}} = \boldsymbol{\Sigma}[\mathbf{b}]_{\mathcal{B}_{\mathcal{U}}}$ with a diagonal block and zero blocks.


$[S]^{-1}_{\mathcal{B}^*_{\mathcal{V}}\mathcal{B}^*_{\mathcal{U}}} = \left([T]^{\mathrm{T}}_{\mathcal{B}_{\mathcal{U}}\mathcal{B}_{\mathcal{V}}}\right)^{-1}$
$[S]_{\mathcal{B}^*_{\mathcal{V}}\mathcal{B}^*_{\mathcal{U}}} = [T]^{\mathrm{T}}_{\mathcal{B}_{\mathcal{U}}\mathcal{B}_{\mathcal{V}}}$
$\mathbb{R}$
$g = S[f] = f \circ T$
$[g]_{\mathcal{B}^*_{\mathcal{U}}} \in \mathbb{R}^n$
$g \in \mathcal{U}^*$
$f \in \mathcal{V}^*$
$[f]_{\mathcal{B}^*_{\mathcal{V}}} \in \mathbb{R}^m$
S^{-1}
$\mathbf{U}$
$\mathbf{U}^{-1}$
$\mathbf{v} = T(\mathbf{u})$
$\mathbf{V}$
$\mathbf{V}^{-1}$
$\mathbf{u} \in \mathcal{U}$
$\mathbf{v} \in \mathcal{V}$
T^{-1}
$[\mathbf{u}]_{\mathcal{B}'_{\mathcal{U}}} \in \mathbb{R}^n$
$[\mathbf{v}]_{\mathcal{B}'_{\mathcal{V}}} \in \mathbb{R}^m$
$\mathbf{P}^{-1}_{\mathcal{B}'_{\mathcal{U}} \to \mathcal{B}_{\mathcal{U}}}$
$\mathbf{P}_{\mathcal{B}'_{\mathcal{U}} \to \mathcal{B}_{\mathcal{U}}}$
$[T]_{\mathcal{B}_{\mathcal{U}}\mathcal{B}_{\mathcal{V}}}$
$\mathbf{P}^{-1}_{\mathcal{B}'_{\mathcal{V}} \to \mathcal{B}_{\mathcal{V}}}$
$\mathbf{P}_{\mathcal{B}'_{\mathcal{V}} \to \mathcal{B}_{\mathcal{V}}}$
$[\mathbf{u}]_{\mathcal{B}_{\mathcal{U}}} \in \mathbb{R}^n$
$[\mathbf{v}]_{\mathcal{B}_{\mathcal{V}}} \in \mathbb{R}^m$
$[T]^{-1}_{\mathcal{B}_{\mathcal{U}}\mathcal{B}_{\mathcal{V}}}$
$[T]_{\mathcal{B}'_{\mathcal{U}}\mathcal{B}'_{\mathcal{V}}}$
$[T]^{-1}_{\mathcal{B}'_{\mathcal{U}}\mathcal{B}'_{\mathcal{V}}}$

- If $\mathbf{A}$ is an invertible $n \times n$ matrix with singular value decomposition

$$\mathbf{A} = \mathbf{U\Sigma V}^{\mathrm{T}}$$

then all three matrices on the left are invertible $n \times n$ matrices, and

$$\mathbf{A}^{-1} = \mathbf{V\Sigma}^{-1}\mathbf{U}^{\mathrm{T}}$$

- If $\mathbf{A}$ is $n \times n$ but singular with a rank $k < n$, then its corresponding

$$\mathbf{\Sigma}$$

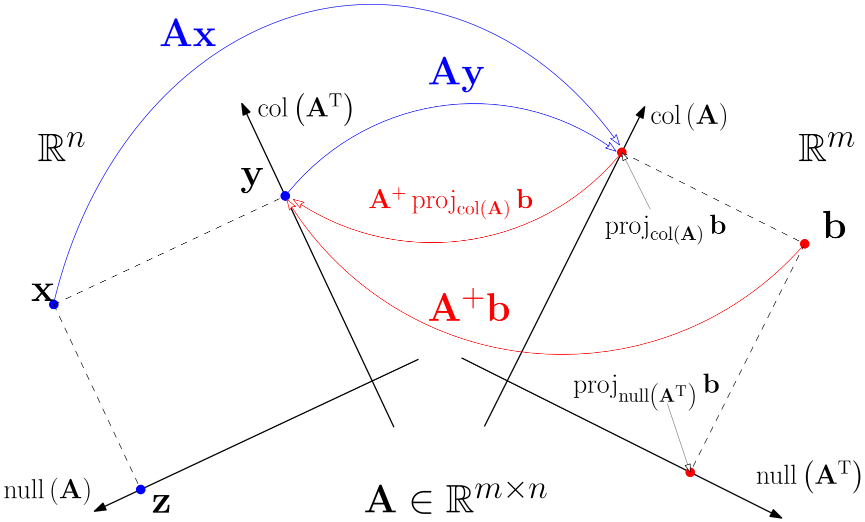
of a singular value decomposition of $\mathbf{A} = \mathbf{U\Sigma V}^{\mathrm{T}}$, is singular, but notice

$$\mathbf{\Sigma}_k$$

in the compact singular value decomposition $\mathbf{A} = \mathrm{U}_k\mathbf{\Sigma}_k\mathbf{V}_k^{\mathrm{T}}$ is invertible.

Q: Why are we not surprised $\mathbf{\Sigma}_k$ is always invertible in terms of transformation?

- Recall we have briefly mentioned the concept of generalised inverse, $\mathbf{A}^+$



Definition

Let $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}}$ be an SVD for an $m \times n$ matrix $\mathbf{A}$ of rank k, where

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

and $\mathbf{\Sigma}_k$ is a $k \times k$ diagonal matrix containing the nonzero singular values of $\mathbf{A}$,

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k > 0$$

The generalised inverse, aka, pseudoinverse of $\mathbf{A}$ is $n \times m$ matrix $\mathbf{A}^+$ defined by

$$\mathbf{A}^+ = \mathbf{V}_k \mathbf{\Sigma}_k^{-1} \mathbf{U}_k^{\mathrm{T}} = \mathbf{V}\mathbf{\Sigma}^+\mathbf{U}^{\mathrm{T}}$$

where $\mathbf{\Sigma}^+$ is the $n \times m$ matrix

$$\mathbf{\Sigma}^+ = \begin{bmatrix} \mathbf{\Sigma}_k^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

- It is generalised also in the sense that $\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}$ and $\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+$.

- In additional to $\mathbf{A}\mathbf{A}^{+}\mathbf{A}=\mathbf{A}$ and $\mathbf{A}^{+}\mathbf{A}\mathbf{A}^{+}=\mathbf{A}^{+}$, the followings are true

$$\left(\mathbf{A}\mathbf{A}^{+}\right)^{\mathrm{T}}=\mathbf{A}\mathbf{A}^{+}$$
$$\left(\mathbf{A}^{+}\mathbf{A}\right)^{\mathrm{T}}=\mathbf{A}^{+}\mathbf{A}$$
$$\left(\mathbf{A}^{\mathrm{T}}\right)^{+}=\left(\mathbf{A}^{+}\right)^{\mathrm{T}}$$
$$\mathbf{A}^{++}=\mathbf{A}$$

Theorem

If $\mathbf{A}$ is an $m\times n$ matrix with rank n, then

$$\mathbf{A}^{\mathrm{T}}\mathbf{A}$$

is invertible, and

$$\mathbf{A}^{+}=\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}$$

thus the least squares solution of $\mathbf{A}\mathbf{x}=\mathbf{b}$ is given by

$$\mathbf{x}=\mathbf{A}^{+}\mathbf{b}$$