Math 6610: Analysis of Numerical Methods, I
The singular value decomposition and consequences

Department of Mathematics, University of Utah

Fall 2025

Resources: Trefethen and Bau 1997, Lectures 4, 5
Salgado and Wise 2022, Chapter 2

Eigenevalue decompositions and "spectral" theorems

If $\boldsymbol{A} \in \mathbb{C}^{n \times n}$, then

- If $\boldsymbol{A}$ is not defective, it's diagonalizable
- $\boldsymbol{A}$ is Hermitian iff it's unitarily real diagonalizable

Unitary diagonalizability, i.e., a "spectral theorem", is conceptually attractive: such matrices are diagonal under an appropriate rigid rotation/reflection of coordinates.

Q: Hermitian + skew-Hermitian matrices are unitarily diagonalizable.
Any others?

In "our" proof strategies: (i) if $\underline{v}$ is an eigenvector of $\underline{A}$, then require span$\{\underline{v}\}^\perp$ to be an invariant subspace of $\underline{A}$.

(ii) Require if $\underline{v}$ is an eigenvector $\underline{A}$, then it's also an eigenvector of A^*.

(I.e. $\underline{\underline{A}}$ and $\underline{\underline{A}}^*$ should have the same eigenvectors)

(iii) Two matrices $\underline{\underline{A}}$ and $\underline{\underline{B}}$ have the same eigenvectors iff they commute ($AB = BA$)

"The" spectral theorem on finite-dimensional spaces is an equivalence between the geometric notion of unitary diagonalizability and an algebraic characterization of matrices.

Definition

A matrix $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ is a **normal matrix** if it commutes with its transpose, i.e., $\boldsymbol{A}\boldsymbol{A}^* = \boldsymbol{A}^*\boldsymbol{A}$.

E.g. $\underline{\underline{A}} = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}$

The spectral theorem

“The” spectral theorem on finite-dimensional spaces is an equivalence between the geometric notion of unitary diagonalizability and an algebraic characterization of matrices.

Definition

A matrix $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ is a **normal matrix** if it commutes with its transpose, i.e., $\boldsymbol{A}\boldsymbol{A}^* = \boldsymbol{A}^*\boldsymbol{A}$.

Now, here is the most general spectral theorem.

Theorem (The spectral theorem (for normal operators))

A square matrix is unitarily diagonalizable iff it is normal.

Normal matrices

The definition of eigenvalues, $\boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v}$, suggests that the eigenvalues of a matrix are sufficient to quantitatively characterize its action. This is in many contexts false for *general* $\boldsymbol{A}$.

Normal matrices are the class of matrices for which eigenvalues *precisely* characterize the action of a matrix.

The definition of eigenvalues, $\boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v}$, suggests that the eigenvalues of a matrix are sufficient to quantitatively characterize its action. This is in many contexts false for *general* $\boldsymbol{A}$.

Normal matrices are the class of matrices for which eigenvalues *precisely* characterize the action of a matrix.

If A is normal, ...

- then the 2-norm induced operator, spectral radius, and numerical radius $(\sup |W(\boldsymbol{A})|)$ coincide
- then $\|\boldsymbol{A}\boldsymbol{x}\|_2 = \|\boldsymbol{A}^*\boldsymbol{x}\|_2$ for any vector $\boldsymbol{x}$
- then $\boldsymbol{A} = \boldsymbol{A}^*\boldsymbol{U}$ for some unitary $\boldsymbol{U}$

Actually, any of the above also implies that $\boldsymbol{A}$ is normal.

The Schur decomposition

One relatively simple proof of the spectral theorem exercises a general matrix decomposition.

Theorem (Schur factorization/decomposition)

Let $\boldsymbol{A} \in \mathbb{C}^{n \times n}$. Then there is a unitary matrix $\boldsymbol{U} \in \mathbb{C}^{n \times n}$ and an upper triangular matrix $\boldsymbol{T} \in \mathbb{C}^{n \times n}$ such that $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{T}\boldsymbol{U}^*$.

(I.e.: all square matrices are unitarily triangularizable.)

The Schur decomposition

One relatively simple proof of the spectral theorem exercises a general matrix decomposition.

Theorem (Schur factorization/decomposition)

Let $\boldsymbol{A} \in \mathbb{C}^{n \times n}$. Then there is a unitary matrix $\boldsymbol{U} \in \mathbb{C}^{n \times n}$ and an upper triangular matrix $\boldsymbol{T} \in \mathbb{C}^{n \times n}$ such that $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{T}\boldsymbol{U}^*$.

(I.e.: all square matrices are unitarily triangularizable.)

For now, the Schur factorization is essentially "just" a crutch for proving the spectral theorem.

But here's a preview for why it's useful more broadly: By the Schur factorization, every matrix $\boldsymbol{A}$ is *unitarily* similar to a triangular matrix $\boldsymbol{T}$.

Recall:

- All non-defective square matrices are diagonalizable (eigenvalue decomposition)
- All square matrices are bidiagonalizable (Jordan normal form)
- All square matrices are unitarily triangularizable (Schur decomposition)
- All *normal* matrices are exactly the set of unitarily diagonalizable square matrices (spectral theorem)

A summary of similarity

Recall:

- All non-defective square matrices are diagonalizable (eigenvalue decomposition)
- All square matrices are bidiagonalizable (Jordan normal form)
- All square matrices are unitarily triangularizable (Schur decomposition)
- All *normal* matrices are exactly the set of unitarily diagonalizable square matrices (spectral theorem)

What about rectangular matrices?

Perhaps the most powerful matrix factorization is the following, which states that all matrices (even defective or rectangular ones) are *asymmetrically* unitarily diagonalizable, with non-negative diagonal.

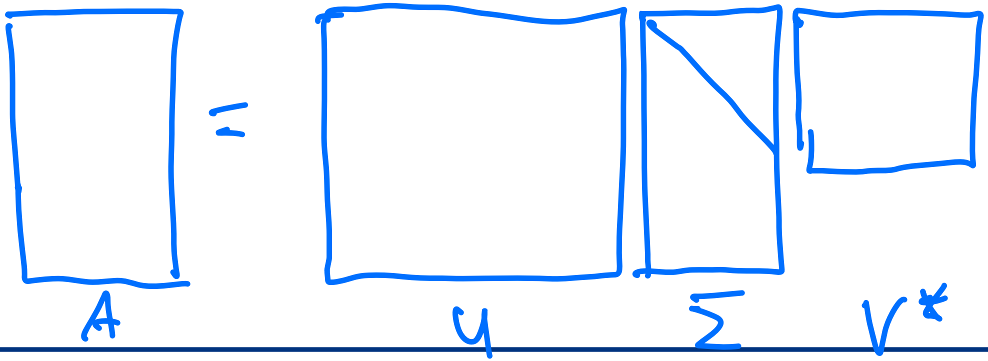
Theorem (SVD: Singular Value Decomposition)

Any matrix $\boldsymbol{A} \in \mathbb{C}^{m \times n}$ *can be written as the product,*

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^*,$$

where $\boldsymbol{U} \in \mathbb{C}^{m \times m}$ *and* $\boldsymbol{V} \in \mathbb{C}^{n \times n}$ *are unitary.*
The matrix $\boldsymbol{\Sigma} \in C^{m \times n}$ *is diagonal with non-negative entries.*
By convention the diagonal elements of $\boldsymbol{\Sigma}$ *are ordered in non-increasing order:* $\Sigma_{i,i} \geqslant \Sigma_{i+1,i+1}$.
This decomposition is unique up to unitary transformations in subspaces with equal singular values.

$m > n$:



The singular value decomposition

Perhaps the most powerful matrix factorization is the following, which states that all matrices (even defective or rectangular ones) are *asymmetrically* unitarily diagonalizable, with non-negative diagonal.

Theorem (SVD: Singular Value Decomposition)

Any matrix $\boldsymbol{A} \in \mathbb{C}^{m \times n}$ *can be written as the product,*

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^*,$$

where $\boldsymbol{U} \in \mathbb{C}^{m \times m}$ *and* $\boldsymbol{V} \in \mathbb{C}^{n \times n}$ *are unitary.*
The matrix $\boldsymbol{\Sigma} \in C^{m \times n}$ *is diagonal with non-negative entries.*
By convention the diagonal elements of $\boldsymbol{\Sigma}$ *are ordered in non-increasing order:* $\Sigma_{i,i} \geqslant \Sigma_{i+1,i+1}$.
This decomposition is unique up to unitary transformations in subspaces with equal singular values.

With $p = \min\{m, n\}$, notational convention:

- $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1, \ldots, \sigma_p)$
- $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_p \geqslant 0$, the "singular values"
- $\boldsymbol{U} = [\boldsymbol{u}_1, \ \boldsymbol{u}_2, \ \ldots, \ \boldsymbol{u}_m]$, the "left singular vectors"
- $\boldsymbol{V} = [\boldsymbol{v}_1, \ \boldsymbol{v}_2, \ \ldots, \ \boldsymbol{v}_n]$, the "right singular vectors"

$m > n$:
$$\underline{\underline{A}} = \begin{pmatrix} | & & | \\ u_1 & \cdots & u_m \\ | & & | \end{pmatrix} \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{pmatrix} \begin{pmatrix} - v_1^* - \\ \vdots \\ - v_n^* - \end{pmatrix}$$

Perhaps the most powerful matrix factorization is the following, which states that all matrices (even defective or rectangular ones) are *asymmetrically* unitarily diagonalizable, with non-negative diagonal.

Theorem (SVD: Singular Value Decomposition)

Any matrix $\boldsymbol{A} \in \mathbb{C}^{m \times n}$ *can be written as the product,*

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^*,$$

where $\boldsymbol{U} \in \mathbb{C}^{m \times m}$ *and* $\boldsymbol{V} \in \mathbb{C}^{n \times n}$ *are unitary.*
The matrix $\boldsymbol{\Sigma} \in C^{m \times n}$ *is diagonal with non-negative entries.*
By convention the diagonal elements of $\boldsymbol{\Sigma}$ *are ordered in non-increasing order:* $\Sigma_{i,i} \geqslant \Sigma_{i+1,i+1}$.
This decomposition is unique up to unitary transformations in subspaces with equal singular values.

Proof step 1: Assuming $m \geqslant n$, compute eigendecomposition of $\boldsymbol{A}^*\boldsymbol{A}$. $\in \mathbb{C}^{n \times n}$

NB: A^*A is Hermitian, and positive semi-definite.

$R_{A^*A}(x) = \langle A^*Ax, x\rangle / \|x\|^2 = \|Ax\|^2 / \|x\|^2 \geq 0 \Rightarrow W(A^*A) \subset [0,\infty)$

$$A^*A = V \Lambda V^* \quad , \quad \Lambda = \mathrm{diag}(\lambda_1 \dots \lambda_n),$$

unitary.

$$\lambda_1 \geq \lambda_2 \geq \dots \lambda_n \geq 0$$

$$V = \begin{pmatrix} | & & | \\ v_1 & \dots & v_n \\ | & & | \end{pmatrix}, \quad \langle v_j, v_k \rangle = \delta_{j,k} = \begin{cases} 1, & j = k \\ 0, & j \neq k \end{cases}$$

Perhaps the most powerful matrix factorization is the following, which states that all matrices (even defective or rectangular ones) are *asymmetrically* unitarily diagonalizable, with non-negative diagonal.

Theorem (SVD: Singular Value Decomposition)

Any matrix $\boldsymbol{A} \in \mathbb{C}^{m \times n}$ *can be written as the product,*

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^*,$$

where $\boldsymbol{U} \in \mathbb{C}^{m \times m}$ *and* $\boldsymbol{V} \in \mathbb{C}^{n \times n}$ *are unitary.*
The matrix $\boldsymbol{\Sigma} \in C^{m \times n}$ *is diagonal with non-negative entries.*
By convention the diagonal elements of $\boldsymbol{\Sigma}$ *are ordered in non-increasing order:* $\Sigma_{i,i} \geqslant \Sigma_{i+1,i+1}$.
This decomposition is unique up to unitary transformations in subspaces with equal singular values.

Proof step 2: Define σ_j, compute $\boldsymbol{u}_j = \frac{1}{\sigma_j}\boldsymbol{A}\boldsymbol{v}_j$, show they are orthonormal.

Define $\lambda_j = \sigma_j^2$, $(\lambda_j \geq 0)$
with $\sigma_j \geq 0$

· Suppose $\sigma_j > 0$ for $j \leq r$.

For $j \le r$: define $u_j := \frac{1}{\sigma_j} A v_j$

$$\langle u_j, u_j \rangle = \frac{1}{\sigma_j^2} \langle A v_j, A v_j \rangle = \frac{1}{\sigma_j^2} v_j^* A^* A v_j$$

$$= \frac{1}{\sigma_j^2} v_j^* \lambda_j v_j = 1$$

$j \neq k$, $j, k \le r$: $\langle u_j, u_k \rangle = \frac{1}{\sigma_j \sigma_k} v_k^* \lambda_j v_j = 0$

$\Longrightarrow \{u_j\}_{j \in [r]}$ is orthonormal.

For future reference: Let $\{u_j\}_{j=r+1}^m$ be any $\mathbb{C}^m$-orthonormal completion of $\{u_j\}_{j \in [r]}$.

Perhaps the most powerful matrix factorization is the following, which states that all matrices (even defective or rectangular ones) are *asymmetrically* unitarily diagonalizable, with non-negative diagonal.

Theorem (SVD: Singular Value Decomposition)

Any matrix $\boldsymbol{A} \in \mathbb{C}^{m \times n}$ *can be written as the product,*

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^*,$$

where $\boldsymbol{U} \in \mathbb{C}^{m \times m}$ *and* $\boldsymbol{V} \in \mathbb{C}^{n \times n}$ *are unitary.*
The matrix $\boldsymbol{\Sigma} \in C^{m \times n}$ *is diagonal with non-negative entries.*
By convention the diagonal elements of $\boldsymbol{\Sigma}$ *are ordered in non-increasing order:* $\Sigma_{i,i} \geqslant \Sigma_{i+1,i+1}$.
This decomposition is unique up to unitary transformations in subspaces with equal singular values.

Proof step 3: Concatenate vector equalities, complete bases for $\mathbb{C}^m$, $\mathbb{C}^n$, solve for $\boldsymbol{A}$.

$$u_j \sigma_j = A v_j \qquad j \in [r]$$

$$u_j \cdot 0 = A v_j \quad , \; j \geq r+1 \quad (A v_j = 0 v_j)$$

$$\begin{pmatrix} u_1\sigma_1 & \cdots & u_n\sigma_n \end{pmatrix} = A \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}$$

$$\begin{bmatrix} | & & | \\ u_1 & \cdots & u_n \\ | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} = AV$$

$$\begin{bmatrix} | & & | & | & & | \\ u_1 & \cdots & u_n & u_{n+1} & \cdots & u_m \\ | & & | & | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ & 0 & \end{bmatrix} = AV$$

$$U\Sigma = AV \Rightarrow U\Sigma V^* = A$$

The SVD is incredibly general and useful.

$$A = U\Sigma V^* = \sum_{j=1}^{p} \sigma_j \left(u_j v_j^*\right)$$

$p = \min\{m, n\}$

recall: for Hermitian matrices: $A = \sum_j \lambda_j (v_j v_j^*)$

The SVD

The SVD is incredibly general and useful.

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^* = \sum_{j=1}^{p} \sigma_j \left(\boldsymbol{u}_j \boldsymbol{v}_j^*\right)$$

If $r = \mathrm{rank}(\boldsymbol{A})$:

$$\mathrm{rank}(\boldsymbol{A}) = r \quad \Longleftrightarrow \quad \sigma_r > 0, \text{ and } \sigma_j = 0, \quad j > r$$

The SVD is incredibly general and useful.

"full" SVD $\rightarrow$ $$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^* = \sum_{j=1}^{p} \sigma_j \left(\boldsymbol{u}_j \boldsymbol{v}_j^*\right)$$

If $r = \mathrm{rank}(\boldsymbol{A})$:

$$\mathrm{rank}(\boldsymbol{A}) = r \quad \Longleftrightarrow \quad \sigma_r > 0, \text{ and } \sigma_j = 0, \quad j > r$$

$\boldsymbol{A}$ has a *reduced* SVD:

$$\boldsymbol{A} = \sum_{j=1}^{r} \sigma_j \left(\boldsymbol{u}_j \boldsymbol{v}_j^*\right) = \widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{V}}^*, \qquad \widetilde{\boldsymbol{U}} = \left[\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r\right], \qquad \widetilde{\boldsymbol{V}} = \left[\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\right]$$

$$A = \begin{bmatrix} | & & | \\ u_1 & \cdots & u_r \\ | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \begin{bmatrix} - & v_1^* & - \\ & \vdots & \\ - & v_r^* & - \end{bmatrix}$$

square, invertible

The SVD is incredibly general and useful.

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^* = \sum_{j=1}^{p} \sigma_j \left(\boldsymbol{u}_j \boldsymbol{v}_j^*\right)$$

If $r = \mathrm{rank}(\boldsymbol{A})$:

$$\mathrm{rank}(\boldsymbol{A}) = r \quad \Longleftrightarrow \quad \sigma_r > 0, \text{ and } \sigma_j = 0, \quad j > r$$

$\boldsymbol{A}$ has a *reduced* SVD:

$$\boldsymbol{A} = \sum_{j=1}^{r} \sigma_j \left(\boldsymbol{u}_j \boldsymbol{v}_j^*\right) = \widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{V}}^*, \qquad \widetilde{\boldsymbol{U}} = \left[\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r\right], \qquad \widetilde{\boldsymbol{V}} = \left[\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\right]$$

$$\|\boldsymbol{A}\|_2 = \sigma_1 = \sqrt{\lambda_{max}(A^*A)} = \sqrt{\lambda_{max}(AA^*)}$$

The SVD is, at last, a decomposition that actually reveals the "size" of a matrix.

Why? 1.) $\|\cdot\|_2$ is unitarily invariant.

2.) $\|\Sigma\|_2 = \sigma_1$ (direct computation)

The SVD

The SVD is incredibly general and useful.

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^* = \sum_{j=1}^{p} \sigma_j \left(\boldsymbol{u}_j \boldsymbol{v}_j^*\right)$$

If $r = \mathrm{rank}(\boldsymbol{A})$:

$$\mathrm{rank}(\boldsymbol{A}) = r \qquad \Longleftrightarrow \qquad \sigma_r > 0, \text{ and } \sigma_j = 0, \quad j > r$$

$\boldsymbol{A}$ has a *reduced* SVD:

$$\boldsymbol{A} = \sum_{j=1}^{r} \sigma_j \left(\boldsymbol{u}_j \boldsymbol{v}_j^*\right) = \widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{V}}^*, \qquad \widetilde{\boldsymbol{U}} = \left[\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r\right], \qquad \widetilde{\boldsymbol{V}} = \left[\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\right]$$

$$\|\boldsymbol{A}\|_2 = \sigma_1$$

The SVD is, at last, a decomposition that actually reveals the "size" of a matrix.

If A is square and invertible:

$$\boldsymbol{A}^{-1} = \boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^*$$

The four fundamental subspaces, redux

The SVD immediately reveals the 4 fundamental subspaces. Since,

$$\boldsymbol{A} = \widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{V}}^*, \text{ (reduced)}$$

$$= U \Sigma V^* \text{ (full)}$$

then:

- $\text{range}(\boldsymbol{A}) = \text{span}\{\boldsymbol{u}_1, \ldots \boldsymbol{u}_r\}$
- $\text{corange}(\boldsymbol{A}) = \ker(\boldsymbol{A}^*) = \text{span}\{\boldsymbol{u}_{r+1}, \ldots \boldsymbol{u}_m\}$
- $\ker(\boldsymbol{A}) = \text{span}\{\boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_n\}$
- $\text{coker}(\boldsymbol{A}) = \text{range}(\boldsymbol{A}^*) = \text{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\}$

I.e., the SVD (nearly) tells you *everything* about a matrix.

The four fundamental subspaces, redux

The SVD immediately reveals the 4 fundamental subspaces. Since,

$$\boldsymbol{A} = \widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{V}},$$

then:

- $\mathrm{range}(\boldsymbol{A}) = \mathrm{span}\{\boldsymbol{u}_1, \ldots \boldsymbol{u}_r\}$
- $\mathrm{corange}(\boldsymbol{A}) = \ker(\boldsymbol{A}^*) = \mathrm{span}\{\boldsymbol{u}_{r+1}, \ldots \boldsymbol{u}_m\}$
- $\ker(\boldsymbol{A}) = \mathrm{span}\{\boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_n\}$
- $\mathrm{coker}(\boldsymbol{A}) = \mathrm{range}(\boldsymbol{A}^*) = \mathrm{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r\}$

I.e., the SVD (nearly) tells you *everything* about a matrix.

Q: Let $\boldsymbol{A}$ be a square matrix. How are the singular values σ_i of $\boldsymbol{A}$ related to its spectrum?
(Or maybe to the modulus of its spectrum?)

Low rank approximation

There are numerous (quite potent) uses of the SVD. Here's one of the more popular ones:

Suppose $\boldsymbol{A} \in \mathbb{C}^{m \times n}$ is a given matrix, and $r < \min\{m, n\}$. We wish to solve the optimization problem,

$$\underset{\boldsymbol{B} \in \mathbb{C}^{m \times n} \;\; \operatorname{rank}(\boldsymbol{B}) \leqslant r}{\arg\min} \; \|\boldsymbol{A} - \boldsymbol{B}\|,$$

for some norm $\|\cdot\|$.

Low rank approximation

There are numerous (quite potent) uses of the SVD. Here's one of the more popular ones:

Suppose $\boldsymbol{A} \in \mathbb{C}^{m \times n}$ is a given matrix, and $r < \min\{m, n\}$. We wish to solve the optimization problem,

$$\underset{\boldsymbol{B} \in \mathbb{C}^{m \times n} \quad \operatorname{rank}(\boldsymbol{B}) \leqslant r}{\arg\min} \|\boldsymbol{A} - \boldsymbol{B}\|,$$

for some norm $\|\cdot\|$.

Why?

- data compression
- simplification or denoising
- interpretable modeling

There are numerous (quite potent) uses of the SVD. Here's one of the more popular ones:

Suppose $\boldsymbol{A} \in \mathbb{C}^{m \times n}$ is a given matrix, and $r < \min\{m, n\}$. We wish to solve the optimization problem,

$$\underset{\boldsymbol{B} \in \mathbb{C}^{m \times n} \quad \operatorname{rank}(\boldsymbol{B}) \leqslant r}{\arg\min} \|\boldsymbol{A} - \boldsymbol{B}\|,$$

for some norm $\|\cdot\|$.

Why?

- data compression
- simplification or denoising
- interpretable modeling

Unsurpisingly, the choice of norm matters. There's a particular choice that is convenient to consider.

Definition

A norm $\|\cdot\|$ on $m \times n$ matrices is *unitarily invariant* if $\|\boldsymbol{A}\| = \|\boldsymbol{U}\boldsymbol{A}\boldsymbol{V}\|$ for any $\boldsymbol{A}$ and arbitrary unitary matrices $\boldsymbol{U}, \boldsymbol{V}$ of appropriates sizes. This implies the norm is a function of only the singular values. I.e., with $p = \min\{m, n\}$, $\|\cdot\|$ is a function $f : [0, \infty)^p \to [0, \infty)$ such that,

$$\|\boldsymbol{A}\| = f(\sigma_1, \ldots, \sigma_p).$$

The main result is that the explicit solution to a(ny) nuclear norm low-rank approximation problem is a truncated SVD.

Theorem (Schmidt-Eckart-Young-Mirsky)

Let $\boldsymbol{A} \in \mathbb{C}^{m \times n}$, and $k < \operatorname{rank}(\boldsymbol{A}) \leqslant \min\{m, n\}$. Suppose $\|\cdot\|$ is a unitarily invariant norm on $m \times n$ matrices. Then a solution $\boldsymbol{B}_k$ to,

$$\boldsymbol{B}_k \in \underset{\boldsymbol{B} \in \mathbb{C}^{m \times n} \quad \operatorname{rank}(\boldsymbol{B}) \leqslant k}{\arg\min} \|\boldsymbol{A} - \boldsymbol{B}\|,$$

is

$$\boldsymbol{B}_k = \sum_{j \in [k]} \cancel{\sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^*} \quad \sigma_j u_j v_j^*$$

The minimizer $\boldsymbol{B}_k$ is unique iff $\sigma_k > \sigma_{k+1}$.

(NB: $k \geqslant \operatorname{rank}(\boldsymbol{A})$ is fine, but the problem is then vacuous since one can choose $\boldsymbol{B} = \boldsymbol{A}$.)

Computing the SVD

It's kind of unclear how the SVD should be computed. However, the proof we've presented is actually constructive:

1. Since $\boldsymbol{A}\boldsymbol{A}^* \in \mathbb{C}^{m \times m}$ and $\boldsymbol{A}^*\boldsymbol{A} \in \mathbb{C}^{n \times n}$ are both spd, we order their eigenvalues in decreasing order. Then $\sigma_i^2(\boldsymbol{A}) = \lambda_i(\boldsymbol{A}\boldsymbol{A}^*) = \lambda_i(\boldsymbol{A}^*\boldsymbol{A})$.

Computing the SVD

It's kind of unclear how the SVD should be computed. However, the proof we've presented is actually constructive:

1. Since $\boldsymbol{A}\boldsymbol{A}^* \in \mathbb{C}^{m\times m}$ and $\boldsymbol{A}^*\boldsymbol{A} \in \mathbb{C}^{n\times n}$ are both spd, we order their eigenvalues in decreasing order. Then $\sigma_i^2(\boldsymbol{A}) = \lambda_i(\boldsymbol{A}\boldsymbol{A}^*) = \lambda_i(\boldsymbol{A}^*\boldsymbol{A})$.

2. Suppose we compute $\lambda_i(\boldsymbol{A}^*\boldsymbol{A})$ for $i \in [n]$, so we have $\sigma_i(\boldsymbol{A})$.
   By the spectral theorem, we can also compute (orthonormal!) eigenvectors $\boldsymbol{v}_i$ such that $\boldsymbol{A}^*\boldsymbol{A}\boldsymbol{v}_i = \sigma_i^2\boldsymbol{v}_i$.

Computing the SVD

It's kind of unclear how the SVD should be computed. However, the proof we've presented is actually constructive:

1. Since $\boldsymbol{A}\boldsymbol{A}^* \in \mathbb{C}^{m\times m}$ and $\boldsymbol{A}^*\boldsymbol{A} \in \mathbb{C}^{n\times n}$ are both spd, we order their eigenvalues in decreasing order. Then $\sigma_i^2(\boldsymbol{A}) = \lambda_i(\boldsymbol{A}\boldsymbol{A}^*) = \lambda_i(\boldsymbol{A}^*\boldsymbol{A})$.

2. Suppose we compute $\lambda_i(\boldsymbol{A}^*\boldsymbol{A})$ for $i \in [n]$, so we have $\sigma_i(\boldsymbol{A})$.
   By the spectral theorem, we can also compute (orthonormal!) eigenvectors $\boldsymbol{v}_i$ such that $\boldsymbol{A}^*\boldsymbol{A}\boldsymbol{v}_i = \sigma_i^2 \boldsymbol{v}_i$.

3. As in the proof, set $\boldsymbol{u}_i = \boldsymbol{A}\boldsymbol{v}_i/\sigma_i$. We now have $\sigma_i, \boldsymbol{u}_i, \boldsymbol{v}_i$ for $i \in [n]$.

Computing the SVD

It's kind of unclear how the SVD should be computed. However, the proof we've presented is actually constructive:

1. Since $\boldsymbol{A}\boldsymbol{A}^* \in \mathbb{C}^{m \times m}$ and $\boldsymbol{A}^*\boldsymbol{A} \in \mathbb{C}^{n \times n}$ are both spd, we order their eigenvalues in decreasing order. Then $\sigma_i^2(\boldsymbol{A}) = \lambda_i(\boldsymbol{A}\boldsymbol{A}^*) = \lambda_i(\boldsymbol{A}^*\boldsymbol{A})$.

2. Suppose we compute $\lambda_i(\boldsymbol{A}^*\boldsymbol{A})$ for $i \in [n]$, so we have $\sigma_i(\boldsymbol{A})$.
   By the spectral theorem, we can also compute (orthonormal!) eigenvectors $\boldsymbol{v}_i$ such that $\boldsymbol{A}^*\boldsymbol{A}\boldsymbol{v}_i = \sigma_i^2 \boldsymbol{v}_i$.

3. As in the proof, set $\boldsymbol{u}_i = \boldsymbol{A}\boldsymbol{v}_i/\sigma_i$. We now have $\sigma_i, \boldsymbol{u}_i, \boldsymbol{v}_i$ for $i \in [n]$.

4. To compute $\boldsymbol{u}_{n+1}, \ldots \boldsymbol{u}_m$, we can choose a(ny) $\mathbb{C}^m$-orthogonal completion of $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$.

There are some minor details to iron out before this is actually implementable, e.g., we really only compute $\boldsymbol{u}_i$ for $i \leqslant \operatorname{rank}(\boldsymbol{A})$, and the rest we identify through orthogonal completion.

Back to eigenvalues

Hence, we arrive at the conclusion that the SVD is quite directly computable through an eigenvalue decomposition of symmetric (positive semi-definite) matrices!

So if we can compute eigenpairs of symmetric matrices, the rest is "easy".

Pro tip: computing eigenpairs of matrices of size $\gtrsim 10$ using roots of characteristic polynomials is an unattractive idea.

Back to eigenvalues

Hence, we arrive at the conclusion that the SVD is quite directly computable through an eigenvalue decomposition of symmetric (positive semi-definite) matrices!

So if we can compute eigenpairs of symmetric matrices, the rest is "easy".

Pro tip: computing eigenpairs of matrices of size $\gtrsim 10$ using roots of characteristic polynomials is an unattractive idea.

In summary: eigenvalues are important to compute!
Most tasks we discussed rely on computing eigenvalues, at least of Hermitian matrices. How is this done?

For the next few weeks, we'll talk about how to compute eigenvalues. However, there's a lot of buildup we'll need before arriving there:

- What makes a good computational algorithm?
- How do we solve linear systems?
- How to we orthogonalize vectors?

References I

Salgado, Abner J. and Steven M. Wise (2022). *Classical Numerical Analysis: A Comprehensive Course*. Cambridge: Cambridge University Press. ISBN: 978-1-108-83770-5. DOI: 10.1017/9781108942607.

Trefethen, Lloyd N. and David Bau (1997). *Numerical Linear Algebra*. SIAM: Society for Industrial and Applied Mathematics. ISBN: 0-89871-361-7.