# Numerical Matrix Analysis

## Notes #5 — The Singular Value Decomposition

Peter Blomgren
⟨blomgren@sdsu.edu⟩

Department of Mathematics and Statistics
Dynamical Systems Group
Computational Sciences Research Center
San Diego State University
San Diego, CA 92182-7720

**http://terminus.sdsu.edu/**

Spring 2024
(Revised: February 1, 2024)



## Outline







## Student Learning Targets, and Objectives

Target The Singular Value Decomposition
- Objective Existence and Uniqueness statements
- Objective Impact: "diagonalizability"

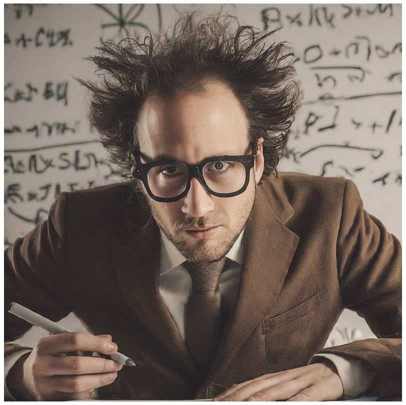
Target The SVD ↔ Matrix Properties
- Objective rank, range, null-space, norms
- Objective relation to eigenvalues, determinant
- Objective Linearly Optimal **Low Rank Approximations**





## Gratuitous "AI"

**Google Bard, 2004-02-01:** create an image of a nerdy mathematician preparing slides for a lecture on computational matrix algebra



*"'Sup, bruh?!? I invert giant matrices by hand. What's your superpower?!?"*

## Last Time

- Hölder and Cauchy-[Bunyakovsky]-Schwarz inequalities:

$$|\vec{v}^*\vec{w}| \overset{H}{\leq} \|\vec{v}\|_p \, \|\vec{w}\|_q, \quad \frac{1}{p} + \frac{1}{q} = 1, \qquad |\vec{v}^*\vec{w}| \overset{CBS}{\leq} \|\vec{v}\|_2 \, \|\vec{w}\|_2$$

- Bounds on the norms of matrix products

$$\|AB\| \leq \|A\| \, \|B\|$$

- General matrix norms: The Frobenius norm $\|A\|_F^2 = \sum_{ij} |a_{ij}|^2$.
- A geometrical introduction to the SVD.
- The reduced vs. the full SVD.





## Missing Proof

We ended last lecture with: "**If** *we can show that every matrix A has a SVD,* **then** *it follows that the image of the unit sphere under any linear map is a hyper-ellipse...*"

We now turn our attention to showing that this indeed is the case...

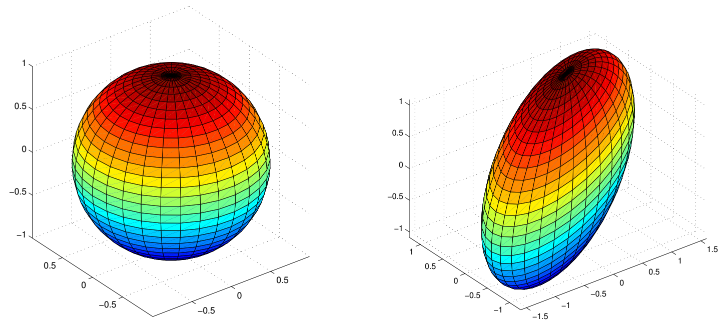


**Figure:** The unit-sphere $\mathbb{S}^2$, and the image $A\mathbb{S}^2$, where $A = \begin{bmatrix} 1.3127 & 0.6831 & 0.6124 \\ 0.0129 & 1.0928 & 0.6085 \\ 0.3840 & 0.0353 & 1.0158 \end{bmatrix}$.





## Theorem: $A = U\Sigma V^*$ — Existence and Uniqueness

**Theorem (Existence and Uniqueness of the SVD)**

*Every matrix $A \in \mathbb{C}^{m\times n}$ has a singular value decomposition $A = U\Sigma V^*$, where*

$$\begin{array}{lcll} U & \in & \mathbb{C}^{m\times m} & \text{is unitary} \\ V & \in & \mathbb{C}^{n\times n} & \text{is unitary} \\ \Sigma & \in & \mathbb{R}^{m\times n} & \text{is diagonal, non-negative.} \end{array}$$

*Furthermore, the singular values $\{\sigma_k\}$ are uniquely determined, and if A is square and the $\sigma_k$ are distinct, the left $\{\vec{u}_k\}$ and right $\{\vec{v}_k\}$ singular vectors are uniquely determined up to complex scalar factors $s \in \mathbb{C} : |s| = 1$.*

We present a proof that is very "matrix-y," a completely different approach is presented [MATH 524 (NOTES#7.1–7.2)]





## Theorem: $A = U\Sigma V^*$ — Proof, 1 of 5

THE PROOF IS BY INDUCTION. Let $\sigma_1 = \|A\|_2$. There must exist $\vec{u}_1 \in \mathbb{C}^m$, $\|\vec{u}_1\|_2 = 1$, and $\vec{v}_1 \in \mathbb{C}^n$, $\|\vec{v}_1\|_2 = 1$, such that $A\vec{v}_1 = \sigma_1\vec{u}_1$:

$$\sigma_1 = \frac{\|A\vec{x}^*\|_2}{\|\vec{x}^*\|_2}, \text{ for some } \vec{x}^*. \quad \text{Let } \vec{v}_1 = \frac{\vec{x}^*}{\|\vec{x}^*\|_2}.$$

Clearly, $A\vec{v}_1 = \vec{p}$, for some $\vec{p}$. Let $\vec{u}_1 = \dfrac{\vec{p}}{\|\vec{p}\|_2}$, and $\sigma_1 = \|\vec{p}\|_2$.

Consider any extension (∃ Movie, see also [MATH 524]) of $\vec{v}_1$ to an orthonormal basis $\{\vec{v}_k\}_{k=1,\dots,n}$ of $\mathbb{C}^n$ and of $\vec{u}_1$ to an orthonormal basis $\{\vec{u}_k\}_{k=1,\dots,n}$ of $\mathbb{C}^m$. Let $U_1$ and $V_1$ denote the matrices with columns $\vec{u}_k$ and $\vec{v}_k$, respectively.

Theorem: $A = U\Sigma V^*$ Proof, 2 of 5

We have (by construction)

$$U_1^* A V_1 = S = \begin{bmatrix} \sigma_1 & \vec{w}^* \\ \vec{0} & B \end{bmatrix},$$

where $\vec{0}$ is a column-vector of size $(m-1)$, and $\vec{w}^*$ is a row vector of size $(n-1)$, and the matrix $B \in \mathbb{C}^{(m-1)\times(n-1)}$.

Now,

$$\left\| \begin{bmatrix} \sigma_1 & \vec{w}^* \\ \vec{0} & B \end{bmatrix} \begin{bmatrix} \sigma_1 \\ \vec{w} \end{bmatrix} \right\|_2 \geq \sigma_1^2 + \vec{w}^*\vec{w} = \sqrt{\sigma_1^2 + \vec{w}^*\vec{w}} \left\| \begin{bmatrix} \sigma_1 \\ \vec{w} \end{bmatrix} \right\|_2,$$

Hence, $\|S\|_2 \geq \sqrt{\sigma_1^2 + \vec{w}^*\vec{w}}$.





Theorem: $A = U\Sigma V^*$ Proof, 3 of 5

We have $\|S\|_2 \geq \sqrt{\sigma_1^2 + \vec{w}^*\vec{w}}$, and $S = U_1^* A V_1$. Since $U_1$ and $V_1$ are unitary, we must have $\|S\|_2 = \|A\|_2 = \sigma_1$.

Therefore $\|\vec{w}\|_2^2 = \vec{w}^*\vec{w} = 0$, which means $\vec{w} = 0$, hence

$$U_1^* A V_1 = S = \begin{bmatrix} \sigma_1 & \vec{0}^* \\ \vec{0} & B \end{bmatrix}, \quad \Leftrightarrow \quad A = U_1 \begin{bmatrix} \sigma_1 & \vec{0}^* \\ \vec{0} & B \end{bmatrix} V_1^*$$

If $m = 1$, or $n = 1$, we are done. Otherwise, the sub-matrix B describes the action of $A$ on the subspace orthogonal to $\vec{v}_1$.

We can now recursively (inductively) apply the same process to $B$, and establish existence of the SVD of $A$:

$$A = U_1 \begin{bmatrix} 1 & \vec{0}^* \\ \vec{0} & U_2 \end{bmatrix} \begin{bmatrix} \sigma_1 & \vec{0}^* \\ \vec{0} & \Sigma_2 \end{bmatrix} \begin{bmatrix} 1 & \vec{0}^* \\ \vec{0} & V_2 \end{bmatrix}^* V_1^* = U\Sigma V^*.$$





Theorem: $A = U\Sigma V^*$ Proof, 4 of 5

THE UNIQUENESS PROOF REMAINS —

[Geometric Version] If the singular values $\sigma_k$ are distinct, then the lengths of the semi-axes of the hyper-ellipse $A\mathbb{S}^{(n-1)}$ must be distinct.

The semi-axes themselves are determined by the geometry, up to a complex sign. $\square_{\text{geometric}}$.

[Algebraic Version] $\sigma_1 = \|A\|_2$ is uniquely determined. Now, suppose that in addition to $\vec{v}_1$, there is another linearly independent vector $\vec{w}_1$ with $\|\vec{w}_1\| = 1$, and $\|A\vec{w}_1\| = \sigma_1$.

We define a unit vector $\vec{v}_2$, orthogonal to $\vec{v}_1$, as a linear combination of $\vec{v}_1$ and $\vec{w}_1$:

$$\vec{v}_2 = \frac{\vec{w}_1 - (\vec{v}_1^*\vec{w}_1)\vec{v}_1}{\|\vec{w}_1 - (\vec{v}_1^*\vec{w}_1)\vec{v}_1\|_2}. \quad (\vec{v}_2 = \vec{w}_1^{\perp \vec{v}_1})$$





Theorem: $A = U\Sigma V^*$ Proof, 5 of 5

Since $\|A\|_2 = \sigma_1$, $\|A\vec{v}_2\|_2 \leq \sigma_1$; but this must be an equality, otherwise since for some $\theta$

$$\vec{w}_1 = \cos(\theta)\vec{v}_1 + \sin(\theta)\vec{v}_2, \quad \vec{v}_1 \perp \vec{v}_2, \quad \cos^2(\theta) + \sin^2(\theta) = 1$$

we would have $\|A\vec{w}_1\|_2 < \sigma_1$.

This vector $\vec{v}_2$ is a **second** right singular vector corresponding to the singular value $\sigma_1$; it will lead to the appearance of a $\vec{y}$ (the last $(n-1)$ elements of $V_1^*\vec{v}_2$) with $\|\vec{y}\|_2 = 1$, and $\|B\vec{y}\|_2 = \sigma_1$.

Hence, if the singular vector $\vec{v}_1$ is not unique, then the corresponding singular value $\sigma_1$ is not simple ($\sigma_1 \not> \sigma_2$). Therefore there cannot exist a vector $\vec{w}_1$ as above.

Now, the uniqueness of the remaining singular vectors follows by induction. $\square_{\text{algebraic}}$

## The SVD: $A = U\Sigma V^*$

**Bold Statement**

SVD enables us to say that **every matrix is "diagonal"** — as long as we use the proper bases for the domain $\in \mathbb{C}^n$, and range (image) $\in \mathbb{C}^m$ spaces.

**Changing Bases — Rotating the Map!**

Any $\vec{b} \in \mathbb{C}^m$ can be expanded in the basis of the left singular vectors of $A$ (*i.e.* the columns of $U$), and any $\vec{x} \in \mathbb{C}^n$ in the basis of the right singular vectors of $A$ (*i.e.* the columns of $V$)...

The coordinates for these expansions are

$$\vec{b}' = U^*\vec{b}, \qquad \vec{x}' = V^*\vec{x}.$$

Now, the relation $\vec{b} = A\vec{x}$ can be written in terms of $\vec{b}'$ and $\vec{x}'$:

$$\vec{b} = A\vec{x} \quad \Leftrightarrow \quad U^*\vec{b} = U^*A\vec{x} = U^*\underbrace{U\Sigma V^*}_{A}\vec{x} \quad \Leftrightarrow \quad \tilde{\mathbf{b}}' = \mathbf{\Sigma}\tilde{\mathbf{x}}'$$





## Singular Value vs. Eigenvalue Decomposition



The idea of **diagonalizing** a matrix by a change of basis is the foundation for the study of eigenvalues.

A non-defective square matrix $A$ can be expressed as a diagonal matrix of eigenvalues $\Lambda$, if the range (image) and domain are expressed in a basis of the eigenvectors. The **eigenvalue decomposition** of $A \in \mathbb{C}^{m\times m}$ is

$$A = X\Lambda X^{-1}$$

where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_m)$, and the columns of $X \in \mathbb{C}^{m\times m}$ contain linearly independent eigenvectors of $A$.
We can change basis for the expression $\vec{b} = A\vec{x}$:

$$\vec{b}' = X^{-1}\vec{b}, \qquad \vec{x}' = X^{-1}\vec{x}.$$

and find that

$$\vec{b}' = \Lambda\vec{x}'$$





## Singular Value vs. Eigenvalue Decomposition



### The SVD and Eigenvalue Decomposition

| **The SVD**, $A = U\Sigma V^*$ | **Eigenvalue Decomp.**, $A = X\Lambda X^{-1}$ |
|---|---|
| **Properties** | |
| Uses two different bases — the set of right and left singular vectors. | Uses one basis — the eigenvectors. |
| Uses orthonormal bases | Uses a basis which is generally not orthogonal. |
| All matrices (even rectangular ones) have a singular value decomposition. | Not all matrices (even square ones) have an eigenvalue decomposition. |
| **(Typical) Application Relevance** | |
| Behavior of $A$ itself, or $A^{-1}$. Information in $A$. | Behavior of $A^k$, $e^{tA}$. |





## The SVD ⇝ Matrix Properties



The SVD has many connections with other fundamental topics in linear algebra...

In the following slides, assume that $A \in \mathbb{C}^{m\times n}$, let $p = \min(m, n)$, and let $r \le p$ denote the number of non-zero singular values of $A$; finally let $\mathrm{span}(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m)$ denote the space spanned by the vectors $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m$, *i.e.* all linear combinations of the vectors.

**Theorem (Rank of a Matrix)**

$\mathrm{rank}(A) = r$.

**Proof (Rank of a Matrix)**

The rank of a diagonal matrix is the number of non-zero entries. In the decomposition $A = U\Sigma V^*$, both $U$ and $V$ are full rank. Therefore $\mathrm{rank}(A) = \mathrm{rank}(\Sigma) = r$. □

## The SVD ⇝ Matrix Properties — The Range (Image) and Null-space

**Theorem (Range (Image) and Nullspace of a Matrix)**

$$\text{range}(A) = \text{span}(\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_r),$$
$$\text{null}(A) = \text{span}(\vec{v}_{r+1}, \vec{v}_{r+2}, \ldots, \vec{v}_n).$$

**Proof (Range (Image) and Nullspace of a Matrix)**

This follows directly from the change of bases induced by $A = U\Sigma V^*$ and the fact that

$$\begin{aligned} \text{range}(\Sigma) &= \text{span}(\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_r) &\subseteq \mathbb{C}^m, \\ \text{null}(\Sigma) &= \text{span}(\vec{e}_{r+1}, \vec{e}_{r+2}, \ldots, \vec{e}_n) &\subseteq \mathbb{C}^n. \end{aligned}$$



Recap
Existence and Uniqueness of the SVD
The SVD
"Every Matrix is Diagonal"
Singular Values and Eigenvalues
The SVD ⇝ Matrix Properties

## The SVD ⇝ Matrix Properties — Euclidean and Frobenius Norms

**Theorem (Euclidean and Frobenius Matrix Norms)**

$$\|A\|_2 = \sigma_1, \quad \textit{and} \quad \|A\|_F = \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_r^2}.$$

**Proof (Euclidean and Frobenius Matrix Norms)**

We already established that $\sigma_1 = \|A\|_2$ in the existence proof, since $A = U\Sigma V^*$ with unitary $U$ and $V$, $\|A\|_2 = \|\Sigma\|_2 = \max\{|\sigma_i|\} = \sigma_1$.

Now, since the Frobenius norm is invariant under unitary transformations, $\|A\|_F = \|\Sigma\|_F = \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_r^2}$.



Recap
Existence and Uniqueness of the SVD
The SVD
"Every Matrix is Diagonal"
Singular Values and Eigenvalues
The SVD ⇝ Matrix Properties

## The SVD ⇝ Matrix Properties — Singular Values / Eigenvalues

**Theorem**

*The non-zero singular values of $A$ are the square roots of the non-zero eigenvalues of $A^*A$ or $AA^*$ (these two matrices have the same non-zero eigenvalues).*

**Proof (Singular Values from $AA^*$ or $A^*A$)**

From

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

we see that $A^*A$ and $\Sigma^*\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2)$ have the same eigenvalues, $\lambda_i = \sigma_i^2$, $i = 1, 2, \ldots, p$.

If $n > p$, we have an additional $(n-p)$ zero eigenvalues.

The same argument works for $AA^*$ (just substitute $m$ for $n$)...



Recap
Existence and Uniqueness of the SVD
The SVD
"Every Matrix is Diagonal"
Singular Values and Eigenvalues
The SVD ⇝ Matrix Properties

## The SVD ⇝ Matrix Properties — Singular Values / Eigenvalues

**Theorem ($\sigma_k = |\lambda_k|$ for Hermitian Matrices)**

*If $A = A^*$, then the singular values of $A$ are the absolute values of the eigenvalues of $A$.* **Note:** *In the language of* [MATH 524] *$A$ is self-adjoint.*

**Proof (part 1)**

The eigenvalues of a Hermitian matrix are real since if $(\lambda, \vec{v})$ is an eigenvalue-eigenvector pair ($\lambda \neq 0$), then

$$\begin{aligned} \langle \vec{v}, A\vec{v} \rangle &= \vec{v}^* A\vec{v} = (A^*\vec{v})^*\vec{v} = \langle A^*\vec{v}, \vec{v} \rangle \\ \langle \vec{v}, A\vec{v} \rangle &= \langle \vec{v}, \lambda\vec{v} \rangle = \lambda\langle \vec{v}, \vec{v} \rangle \\ \langle \vec{v}, A\vec{v} \rangle &= \langle A^*\vec{v}, \vec{v} \rangle = \langle A\vec{v}, \vec{v} \rangle = \langle \lambda\vec{v}, \vec{v} \rangle = \lambda^*\langle \vec{v}, \vec{v} \rangle \end{aligned}$$

Hence, $\lambda = \lambda^* \Rightarrow \lambda \in \mathbb{R}$. Further, a Hermitian matrix has a complete set of orthogonal eigenvectors. This means that we can diagonalize $A$

$$A = Q\Lambda Q^* = Q(|\Lambda|\text{sign}(\Lambda))Q^*$$

for some unitary matrix $Q$ and $\Lambda$ a real diagonal matrix...

## The SVD ⇝ Matrix Properties — Singular Values / Eigenvalues

**Proof (part 2)**

Since $\mathrm{sign}(\Lambda)Q^*$ is unitary, we have

$$A = \underbrace{Q}_{U}\,\underbrace{|\Lambda|}_{\Sigma}\,\underbrace{(\mathrm{sign}(\Lambda)Q^*)}_{V^*}$$

a SVD of $A$, where $\sigma_i = |\lambda_i|$, $i = 1, 2, \ldots, p$. (An appropriate ordering of the columns of $U$ guarantees that the singular values are ordered in decreasing order.) □





## The SVD ⇝ Matrix Properties — The Determinant

**Theorem**

*For $A \in \mathbb{C}^{m\times m}$, $|\det(A)| = \prod_{i=1}^{m} \sigma_i$.*

**Proof (Magnitude of Determinant is Product of Singular Values)**

$$\begin{aligned} |\det(A)| &= |\det(U\Sigma V^*)| = |\det(U)| \cdot |\det(\Sigma)| \cdot |\det(V^*)| \\ &= 1 \cdot |\det(\Sigma)| \cdot 1 = |\det(\Sigma)| = \prod_{i=1}^{m} \sigma_i \end{aligned}$$

where we have used the fact that $\det(AB) = \det(A)\det(B)$ and that the magnitude of the determinant of a unitary matrix is one.





## The SVD ⇝ Matrix Properties — Low-Rank Approximations, 1 of 5

This discussion is a significant part of WHY this course exists!

Given the SVD of $A$, $A = U\Sigma V^*$, we can represent $A$ as a sum of $r$ rank-one matrices

$$A = \sum_{k=1}^{r} \sigma_k \vec{u}_k \vec{v}_k^*$$

This is certainly not the only way to write $A$ as a sum of rank-one matrices: it could be written as a sum of its $m$ rows, $n$ columns, or even its $mn$ entries...

The decomposition above has the special property that if we truncate the sum at $\nu < r$, then that partial sum captures as much "energy" of $A$ as possible for a rank-$\nu$ sub-matrix of $A$.

We formalize this in a theorem...





## The SVD ⇝ Matrix Properties — Low-Rank Approximations, 2 of 5

**Theorem (Optimal Low-Rank Approximation)**

*For any $\nu$ with $0 \le \nu < r$, define*

$$A_\nu = \sum_{k=1}^{\nu} \sigma_k \vec{u}_k \vec{v}_k^*$$

*if $\nu = p = \min(m, n)$, define $\sigma_{\nu+1} = 0$. Then*

$$\|A - A_\nu\|_2 = \inf_{\substack{B \in \mathbb{C}^{m\times n} \\ \mathrm{rank}(B) \le \nu}} \|A - B\|_2 = \sigma_{\nu+1}$$

## The SVD $\rightsquigarrow$ Matrix Properties — Low-Rank Approximations, 2.5 of 5

**Low Rank Approximations in DS/Machine Learning/Generative AI**

Low-Rank Adaptation (LoRA) is a family of methods for fine-tuning large-scale AI/Machine Learning models in an efficient manner.

"Base-Models" (e.g. LLMs like ChatGPT; or image-generative models like the Stable Diffusion SD1.5 or SDXL models) are trained on extremely large data sets — this training uses significant resources, *i.e.* they are "expensive."

**Very Simplified:** fine-tuning is "retraining" (parts of) the model using a smaller specific data set; *e.g.* published peer-reviewed mathematics research papers, or images created in a particular "style."

The Model parameters use usually collected in a large matrix $A \in \mathbb{R}^{M\times N}$; and the fine-tuning computes "a few" — collected in much smaller matrices $B \in \mathbb{R}^{M\times p}$, and $C \in \mathbb{R}^{p\times N}$, so that the effective fine-tuned model can be represented as

$$A + BC$$

$M$ and $N$ are usually "quite large" ($> 1,000$), and $p$ "small" ($< 10$).





## The SVD $\rightsquigarrow$ Matrix Properties — Low-Rank Approximations, 3 of 5

Proof (Optimal Low-Rank Approximation)

Suppose that there is some $B$ with $\mathrm{rank}(B) \le \nu$ such that $\|A - B\|_2 < \|A - A_\nu\|_2 = \sigma_{\nu+1}$.

Then there is an $(n-\nu)$-dimensional subspace $\mathrm{null}(B) = \mathbb{W} \subseteq \mathbb{C}^n$ such that $\vec{w} \in \mathbb{W} \Rightarrow B\vec{w} = 0$. Thus $\forall \vec{w} \in \mathbb{W}$:

$$\|A\vec{w}\|_2 = \|(A-B)\vec{w}\|_2 \le \|A-B\|_2\|\vec{w}\|_2 < \sigma_{\nu+1}\|\vec{w}\|_2.$$

Now, $\mathbb{W}$ is an $(n-\nu)$-dimensional subspace where $\|A\vec{w}\|_2 < \sigma_{\nu+1}\|\vec{w}\|_2$. But there is a $(\nu+1)$-dimensional subspace where $\|A\vec{w}\|_2 \ge \sigma_{\nu+1}\|\vec{w}\|_2$ — $\mathbb{V} = \mathrm{span}(u_1, \dots, u_{\nu+1})$ the space spanned by the first $(\nu+1)$ right singular vectors of $A$.

Since the sum of the dimensions of the two subspaces $(\nu+1) + (n-\nu) = (n+1)$ exceeds $n$, there must be a non-zero vector lying in both. This is a contradiction.





## The SVD $\rightsquigarrow$ Matrix Properties — Low-Rank Approximations, 4 of 5

The preceding theorem has a nice geometrical interpretation.

Ponder the issue of finding the best approximation of an $n$-dimensional hyper-ellipsoid.

$\Rightarrow$ The best approximation by a 2-dimensional ellipse must be the ellipse spanned by the largest and second largest axis.

$\Rightarrow$ We get the best 3-dimensional approximation by adding the span of the 3rd largest axis, etc...

This is useful in many applications, *e.g.* signal compression (images, audio, etc.), analysis of large data sets, etc.





## The SVD $\rightsquigarrow$ Matrix Properties — Low-Rank Approximations, 5 of 5

We state the following theorem, and leave the proof as an "exercise."

Theorem

*For the matrix $A_\nu$ as defined in the previous theorem*

$$\|A - A_\nu\|_F = \inf_{\substack{B \in \mathbb{C}^{m\times n} \\ \mathrm{rank}(B) \le \nu}} \|A - B\|_F = \sqrt{\sigma_{\nu+1}^2 + \sigma_{\nu+2}^2 + \cdots + \sigma_r^2}$$

We will get back to **how to compute** the SVD later. For now, we note that it is a powerful tool which can be used to

- find the numerical rank of a matrix;
- find the orthonormal basis for the range (image) and null-space;
- computing $\|A\|_2$;
- computing low-rank approximations.

The SVD shows up in least squares fitting, regularization, intersection of subspaces (video games?), and many, many other problems.