

Numerical Matrix Analysis

Lecture Notes #4

— Fundamentals —

Matrix Norms, the Singular Value Decomposition

Peter Blomgren,

`<blomgren.peter@gmail.com>`

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 2020

Outline

- 1 Introduction: Matrix Norms
 - Recap
 - Inequalities
 - General Matrix Norms
- 2 The Singular Value Decomposition
 - Many Names... One Powerful Tool!
 - Examples for 2×2 Matrices
 - More Details, and Examples Revisited
- 3 The SVD of a Matrix: Formal Definition
 - Spheres and Hyper-ellipses
 - Homework

Last Time

Orthogonal Vectors, Matrices and Norms:

- The Adjoint / Hermitian Conjugate of a Matrix, A^*
- The Inner Product of Two Vectors, $\langle \vec{x}, \vec{y} \rangle = \vec{x}^* \vec{y}$
- Orthogonal, $\langle \vec{x}, \vec{y} \rangle = 0$, and Orthonormal, $\|\vec{x}\| = 1$, Vectors
- Orthogonal and Orthonormal Sets — Linear Independence; Basis for \mathbb{C}^m
- Unitary Matrices $Q^* Q = I$
- Vector Norms, $\|\cdot\|_p$ (p -norms), weighted p -norms
- Induced Matrix Norms

Inequalities: Hölder and Cauchy-Bunyakovsky-Schwarz

1 of 2

Last time we noted that the 1-norm and ∞ -norm of a matrix simplify to the maximal column- and row-sum, respectively, *i.e.* for $A \in \mathbb{C}^{m \times n}$

$$\|A\|_1 = \max_{1 \leq j \leq n} \|\vec{a}_j\|_1$$

$$\|A\|_\infty = \max_{1 \leq i \leq m} \|\vec{a}_i^*\|_1$$

For other p -norms, $1 \leq p \leq \infty$, the matrix-norms do not reduce to simple direct-computable expressions like the ones above.

Inequalities: Hölder and Cauchy-Bunyakovsky-Schwarz

2 of 2

However, we can usually find useful bounds on vector- and matrix-norms, using the **Hölder inequality**:

$$|\vec{x}^* \vec{y}| \leq \|\vec{x}\|_p \|\vec{y}\|_q, \quad \frac{1}{p} + \frac{1}{q} = 1.$$

In the special case $p = q = 2$, the inequality is known as the **Cauchy-Schwarz**, or **Cauchy-Bunyakovsky-Schwarz** inequality

$$|\vec{x}^* \vec{y}| \leq \|\vec{x}\|_2 \|\vec{y}\|_2.$$

Example: 2-norm of a Rank-1 Matrix $A = \vec{u}\vec{v}^*$

Rank-1 matrices formed by an outer product $\vec{u}\vec{v}^*$ show up in many numerical schemes:

$$\vec{u}\vec{v}^* = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{bmatrix} \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{bmatrix} \vec{v}^* = \begin{bmatrix} u_1 \vec{v}^* \\ u_2 \vec{v}^* \\ \vdots \\ u_m \vec{v}^* \end{bmatrix}$$

Now, for any $\vec{x} \in \mathbb{C}^n$, we get

$$\|A\vec{x}\|_2 = \|\vec{u}\vec{v}^* \vec{x}\|_2 = \|\vec{u}\|_2 |\vec{v}^* \vec{x}| \leq \|\vec{u}\|_2 \|\vec{v}\|_2 \|\vec{x}\|_2$$

Hence,

$$\|A\|_2 = \sup_{\vec{x} \in \mathbb{C}^n - \{\vec{0}\}} \frac{\|A\vec{x}\|_2}{\|\vec{x}\|_2} \leq \|\vec{u}\|_2 \|\vec{v}\|_2$$

Example: 2-norm of a Rank-1 Matrix $A = \vec{u}\vec{v}^*$

2 of 2

Since $\vec{v} \in \mathbb{C}^n$, the inequality

$$\|A\|_2 = \sup_{\vec{x} \in \mathbb{C}^n - \{\vec{0}\}} \frac{\|A\vec{x}\|_2}{\|\vec{x}\|_2} \leq \|\vec{u}\|_2 \|\vec{v}\|_2$$

is actually an equality. Let $\vec{x} = \vec{v}$:

$$\|A\vec{v}\|_2 = \|\vec{u}\vec{v}^*\vec{v}\|_2 = \|\vec{u}\|_2 |\vec{v}^*\vec{v}| = \|\vec{u}\|_2 \|\vec{v}\|_2^2$$

Bounds on the Norms of Matrix Products, $\|AB\|$

Let $A \in \mathbb{C}^{\ell \times m}$, $B \in \mathbb{C}^{m \times n}$, and $\vec{x} \in \mathbb{C}^n$: and let $\|\cdot\|$ denote compatible p -norms, then

$$\|AB\vec{x}\| \leq \|A\| \|B\vec{x}\| \leq \|A\| \|B\| \|\vec{x}\|.$$

Therefore, we have

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

where, in general $\|AB\| \neq \|A\| \|B\|$.

General (Non-Induced) Matrix Norms

Matrix norms induced by vector norms are quite common, but as long as the following norm-conditions are satisfied:

- (1) $\|A\| \geq 0$, and $\|A\| = 0$ only if $A = 0$
- (2) $\|A + B\| \leq \|A\| + \|B\|$
- (3) $\|\alpha A\| = |\alpha| \|A\|$

for $A \in \mathbb{C}^{m \times n}$, $\|\cdot\|$ is a valid matrix-norm.

The most commonly used non-induced matrix norm is the **Frobenius norm** (sometimes referred to as the **Hilbert-Schmidt norm**):

$$\|A\|_F = \left[\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 \right]^{1/2}.$$

The Frobenius Norm

We can view the Frobenius Norm in terms of column- or row-sums:

$$\|A\|_F = \left[\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 \right]^{1/2} = \left[\sum_{j=1}^n \|\vec{a}_j\|_2^2 \right]^{1/2} = \left[\sum_{i=1}^m \|\vec{a}_i^*\|_2^2 \right]^{1/2}$$

...or in terms of the trace (sum of diagonal entries)

$$\|A\|_F = \sqrt{\text{trace}(A^*A)} = \sqrt{\text{trace}(AA^*)}$$

Invariance under Unitary Multiplication

Both the 2-norm and the Frobenius norm are invariant under multiplication by unitary matrices, *i.e.*

Theorem

For any $A \in \mathbb{C}^{m \times n}$ and unitary $Q \in \mathbb{C}^{m \times m}$, we have

$$\|QA\|_2 = \|A\|_2, \quad \|QA\|_F = \|A\|_F$$

... an indication of the importance (and usefulness) of unitary matrices!

Invariance under Unitary Multiplication

Both the 2-norm and the Frobenius norm are invariant under multiplication by unitary matrices, *i.e.*

Theorem

For any $A \in \mathbb{C}^{m \times n}$ and unitary $Q \in \mathbb{C}^{m \times m}$, we have

$$\|QA\|_2 = \|A\|_2, \quad \|QA\|_F = \|A\|_F$$

... an indication of the importance (and usefulness) of unitary matrices!

This ends our quick introduction to basic linear algebra concepts

Next: A first look at the Singular Value Decomposition (SVD)



Linear Algebra References

Updated Spring 2020

Introduction to Linear Algebra — “Optional” for Math 254

Gilbert Strang, Wellesley-Cambridge Press, 5th edition (2016),
ISBN-0980232775

Linear Algebra Done Right — “Required” for Math 524

Sheldon Axler, Springer-Verlag, 3rd edition (November 5,
2014), ISBN-3319110799

Linear Algebra and Learning from Data

Gilbert Strang, Wellesley-Cambridge Press, 1st edition (January 2, 2019), ISBN-0692196382

The Singular Value Decomposition

The SVD [*mathematics*] is known by many names:

- the Proper Orthogonal Decomposition (POD)
- the Karhunen-Loève (KL-) Decomposition [*signal analysis*]
- Principal Component Analysis (PCA) [*statistics*]
- Empirical Orthogonal Functions, etc...

“[The SVD is] absolutely a high point of linear algebra.”
Prof. Gilbert Strang, MIT

Hits on scholar.google.com

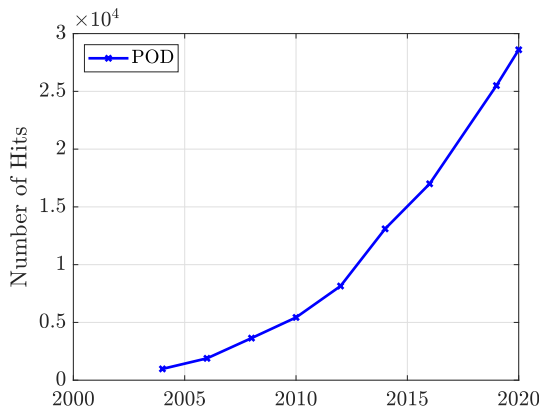


Figure: The many names, faces, and close relatives of the Singular Value Decomposition... Number of hits for "Proper Orthogonal Decomposition"

Hits on scholar.google.com

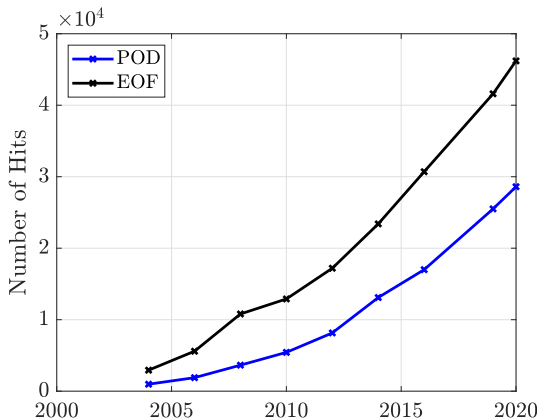


Figure: The many names, faces, and close relatives of the Singular Value Decomposition... Number of hits for “Proper.Orthogonal.Decomposition”, “Empirical.Orthogonal.(Function|Functions)”

Hits on scholar.google.com

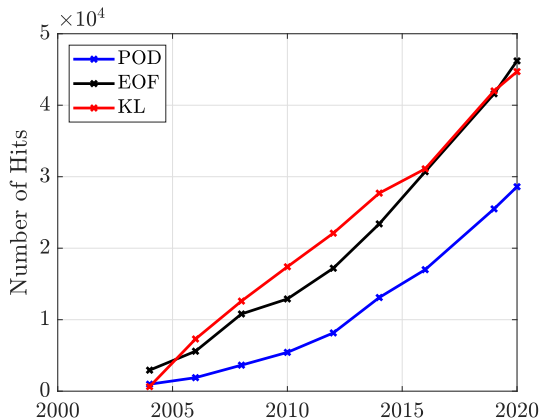


Figure: The many names, faces, and close relatives of the Singular Value Decomposition... Number of hits for “Proper.Orthogonal.Decomposition”, “Empirical.Orthogonal.(Function|Functions)”, “Karhunen.Loeve”

Hits on scholar.google.com

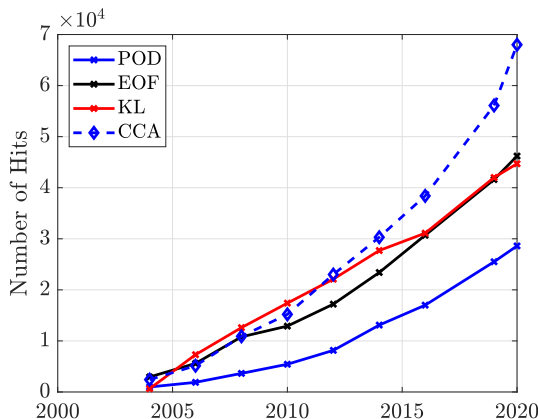


Figure: The many names, faces, and close relatives of the Singular Value Decomposition... Number of hits for “Proper.Orthogonal.Decomposition”, “Empirical.Orthogonal.(Function|Functions)”, “Karhunen.Loeve”, “Canonical.Correlation.Analysis”

Hits on scholar.google.com

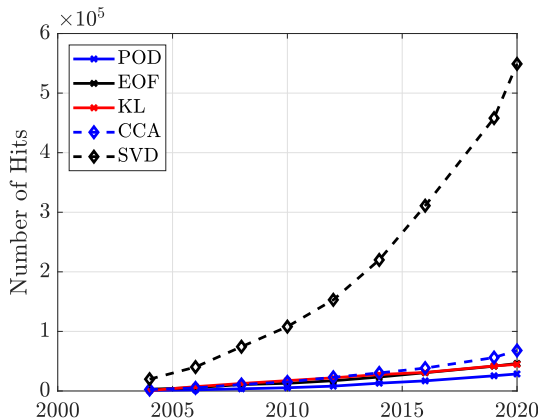


Figure: The many names, faces, and close relatives of the Singular Value Decomposition... Number of hits for “Proper.Orthogonal.Decomposition”, “Empirical.Orthogonal.(Function|Functions)”, “Karhunen.Loeve”, “Canonical.Correlation.Analysis”, “Singular.Value.Decomposition”

Hits on scholar.google.com

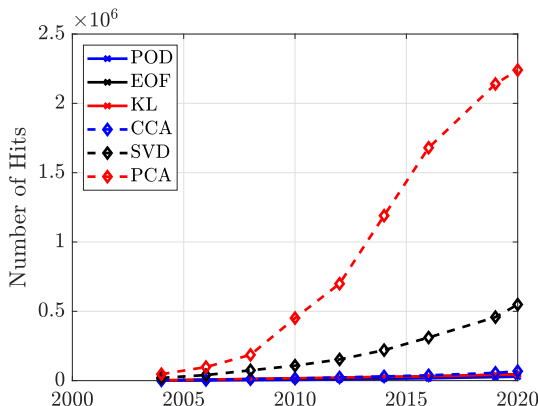


Figure: The many names, faces, and close relatives of the Singular Value Decomposition... Number of hits for “Proper.Orthogonal.Decomposition”, “Empirical.Orthogonal.(Function|Functions)”, “Karhunen.Loeve”, “Canonical.Correlation.Analysis”, “Singular.Value.Decomposition”, “Principal.Component.Analysis”

Hits on scholar.google.com

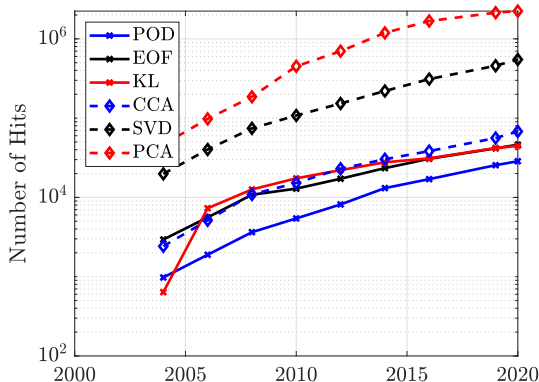


Figure: The many names, faces, and close relatives of the Singular Value Decomposition... Number of hits for “Proper.Orthogonal.Decomposition”, “Empirical.Orthogonal.(Function|Functions)”, “Karhunen.Loeve”, “Canonical.Correlation.Analysis”, “Singular.Value.Decomposition”, “Principal.Component.Analysis”

The Singular Value Decomposition

In our first look at the SVD, we will **not** consider **how to** compute the SVD, but will focus on the meaning of the SVD; — especially its geometric interpretation.

The motivating geometric fact:

The image of the unit sphere under any $m \times n$ matrix, A , is a hyper-ellipse.

The hyper-ellipse in \mathbb{R}^m is the surface we get when stretching the unit sphere by some factors $\sigma_1, \sigma_2, \dots, \sigma_m$ in some orthogonal directions $\vec{u}_1, \vec{u}_2, \dots, \vec{u}_m$.

We take \vec{u}_i to be unit vectors, *i.e.* $\|\vec{u}_i\|_2 = 1$, thus the vectors $\{\sigma_i \vec{u}_i\}$ are the **principal semi-axes** of the hyper-ellipse.

The Singular Value Decomposition

For $A \in \mathbb{R}^{m \times n}$, if $\text{rank}(A) = r$, then exactly r of the lengths σ_i will be non-zero. In particular, if $m \geq n$, at most n of them will be non-zero.

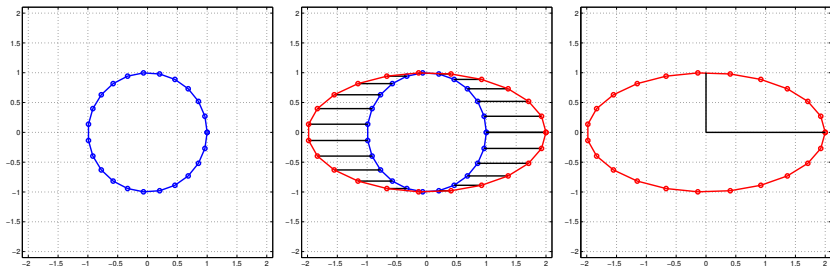
Before we take this discussion further, let's look at some examples of the SVD of some 2×2 matrices.

Keep in mind that computing the SVD of a matrix A answers the question:

“What are the principal semi-axes of the hyper-ellipse generated when A operates on the unit sphere?”

In some sense, this constitutes to most complete information you can extract from a matrix.

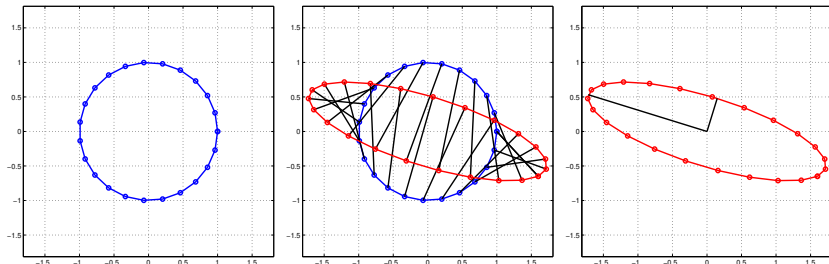
Example#1: SVD of a 2×2 Matrix



$$\text{SVD} \left(\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \right) = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{[\vec{u}_1 \quad \vec{u}_2]} \underbrace{\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}}_{\text{diag}(\sigma_1, \sigma_2)} \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^*}_{\begin{bmatrix} \vec{v}_1 & \vec{v}_2 \end{bmatrix}^*}$$

For now, let's sweep the matrix V^* under the carpet, and note that the SVD has identified the directions of stretching (\vec{u}_1, \vec{u}_2) and the amount of stretching $(\sigma_1, \sigma_2) = (2, 1)$.

Example#2: SVD of a 2×2 Matrix

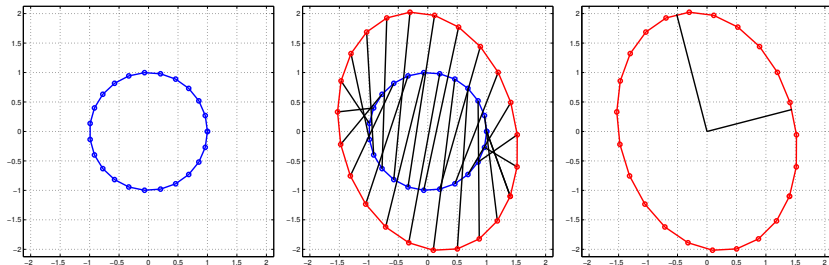


$$\text{SVD} \left(\begin{bmatrix} 1.6 & -0.65 \\ -0.65 & -0.3 \end{bmatrix} \right) = \begin{bmatrix} -0.9553 & 0.2955 \\ 0.2955 & 0.9553 \end{bmatrix} \begin{bmatrix} 1.8011 & 0 \\ 0 & 0.5011 \end{bmatrix} \begin{bmatrix} -0.9553 & -0.2955 \\ 0.2955 & -0.9553 \end{bmatrix}^*$$

Here, the principal semi-axes of the ellipse are

$$\sigma_1 \vec{u}_1 = 1.8011 \begin{bmatrix} -0.9553 \\ 0.2955 \end{bmatrix}, \quad \sigma_2 \vec{u}_2 = 0.5011 \begin{bmatrix} 0.2955 \\ 0.9553 \end{bmatrix}$$

Example#3: SVD of a 2×2 Matrix



$$\text{SVD} \left(\begin{bmatrix} 1.4 & -0.62 \\ -1.1 & -1.7 \end{bmatrix} \right) = \begin{bmatrix} -0.2501 & 0.9682 \\ 0.9682 & 0.2501 \end{bmatrix} \begin{bmatrix} 2.0556 & 0 \\ 0 & 1.4896 \end{bmatrix} \begin{bmatrix} -0.6885 & 0.7253 \\ -0.7253 & -0.6885 \end{bmatrix}^*$$

Here, the principal semi-axes of the ellipse are

$$\sigma_1 \vec{u}_1 = 2.0556 \begin{bmatrix} -0.2501 \\ 0.9682 \end{bmatrix}, \quad \sigma_2 \vec{u}_2 = 1.4896 \begin{bmatrix} 0.9682 \\ 0.2501 \end{bmatrix}$$

The Singular Value Decomposition

More Details, 1 of 2

Let \mathbb{S}^{n-1} be the unit sphere in \mathbb{R}^n , i.e.

$$\mathbb{S}^{n-1} = \{\vec{x} \in \mathbb{R}^n : \|\vec{x}\|_2 = 1\}$$

Let $A \in \mathbb{R}^{m \times n}$ ($m \geq n$) be of full rank, i.e. $\text{rank}(A) = n$, and let $A\mathbb{S}^{n-1}$ denote the image of the unit sphere (our hyper-ellipse).

The **n singular values of A** are the lengths of the n principal semi-axes of $A\mathbb{S}^{n-1}$ (some lengths may be zero), written as $\sigma_1, \sigma_2, \dots, \sigma_n$. By convention, they are ordered in descending order, so that

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n > 0$$

The **n left singular vectors of A** are the unit vectors $\{\vec{u}_1, \vec{u}_2, \dots, \vec{u}_n\}$ oriented in the directions of the principal semi-axes of $A\mathbb{S}^{n-1}$.



The Singular Value Decomposition

More Details, 2 of 2

Note that the vector $\sigma_i \vec{u}_i$ is the i th largest principal semi-axis of $A\mathbb{S}^{n-1}$.

The n **right singular vectors of A** are the unit vectors $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\} \in \mathbb{S}^{n-1}$ that are **pre-images** of the principal semi-axes of $A\mathbb{S}^{n-1}$, i.e.

$$A\vec{v}_j = \sigma_j \vec{u}_j.$$

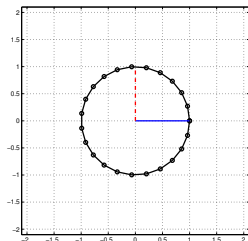
Note how this is similar to *and* different from an eigen-vector – eigen-value pair:

$$A\vec{\xi}_j = \lambda_j \vec{\xi}_j.$$

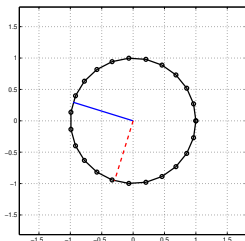
With that knowledge we can re-visit the three examples...

Revisited: Our 3 Examples

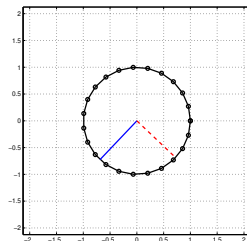
3D Movies \exists



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



$$A = \begin{bmatrix} 1.6 & -0.65 \\ -0.65 & -0.3 \end{bmatrix}$$



$$A = \begin{bmatrix} 1.4 & -0.62 \\ -1.1 & -1.7 \end{bmatrix}$$

The Reduced SVD

1 of 2

What we have described so far is known as the reduced (or thin) SVD, if we collect the relations between the right and left singular vectors,

$$A\vec{v}_j = \sigma_j\vec{u}_j, \quad j = 1, \dots, n$$

in full-blown matrix notation we get

$$\begin{bmatrix} A \end{bmatrix} \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \end{bmatrix} = \begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_n \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{bmatrix}$$

Usually written in the compact form

$$AV = \hat{U}\hat{\Sigma}$$

The Reduced SVD

2 of 2

In looking at the reduced SVD in this form

$$AV = \hat{U}\hat{\Sigma}$$

we note that $A \in \mathbb{C}^{m \times n}$ (if $\text{rank}(A) = n$), $V \in \mathbb{C}^{n \times n}$ (unitary), $\hat{U} \in \mathbb{C}^{m \times n}$ (unitary), and $\hat{\Sigma} \in \mathbb{R}^{n \times n}$ (diagonal, real).

If we multiply by V^* from the right, and use the fact that $VV^* = I$, we get the reduced SVD in its standard form:

$$A = \hat{U}\hat{\Sigma}V^*$$

From the Reduced to the Full SVD

In most applications the SVD is used as we have described (*i.e.* the reduced SVD “version”).

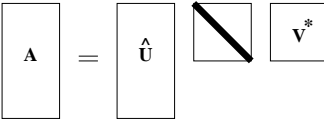
However, the SVD can be extended as follows: The columns of \hat{U} are n orthonormal vectors in \mathbb{C}^m ($m \geq n$). If $m < n$, then they do not form a basis for \mathbb{C}^m .

[LINEARLY INDEPENDENT LIST EXTENDS TO A BASIS (MATH 524, NOTES#2)]

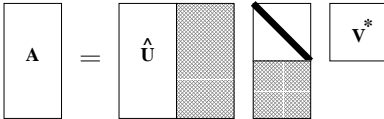
By adding an additional $(n - m)$ orthonormal columns to \hat{U} , we get a new unitary matrix $U \in \mathbb{C}^{n \times n}$.

Further, we form the matrix Σ , by adding $(n - m)$ rows of zeros at the bottom of $\hat{\Sigma}$.

The Reduced and Full SVDs



$$A = \hat{U} \hat{\Sigma} \hat{V}^*$$



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

We can now drop the simplifying assumption that $\text{rank}(A) = n$.

If A is rank-deficient, *i.e.* $\text{rank}(A) = r < n$, the full SVD is still appropriate; however, we only get r left singular vectors \vec{u}_j from the geometry of the hyper-ellipse.

In order to construct U , we add $(n - r)$ additional arbitrary orthonormal columns. In addition V will need $(n - r)$ additional arbitrary orthonormal columns. The matrix Σ will have r positive diagonal entries, with the remaining $(n - r)$ equal to zero.

The SVD of a Matrix: Formal Definition

Definition (Singular Value Decomposition)

Let m and n be arbitrary integers. Given $A \in \mathbb{C}^{m \times n}$, a **Singular Value Decomposition** of A is a factorization

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

where

$$\begin{array}{lll} U & \in \mathbb{C}^{m \times m} & \text{is unitary} \\ V & \in \mathbb{C}^{n \times n} & \text{is unitary} \\ \Sigma & \in \mathbb{R}^{m \times n} & \text{is diagonal} \end{array}$$

The diagonal entries of Σ are non-negative, and ordered in decreasing order, *i.e.* $\sigma_1 \geq \sigma_2 \geq \dots \sigma_p \geq 0$, where $p = \min(m, n)$.

Note: We do not require $m \geq n$. $\text{rank}(A) = r \leq \min(m, n)$.

Spheres and Hyper-ellipses

Clearly[@] if A has a SVD, *i.e.* $A = U\Sigma V^*$, then A must map the unit sphere into a hyper-ellipse:

- V^* preserves the sphere, since multiplication by a unitary matrix preserves the 2-norm. (Multiplication by V^* is a rotation + possibly a reflection).
- Multiplication by Σ stretches the sphere into a hyper-ellipse aligned with the basis.
- Multiplication by the unitary U preserves all 2-norms, and angles between vectors; hence the shape of the hyper-ellipse is preserved (albeit rotated and reflected).

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; something we stated boldly on slide 15.

[@] clearly = “Are you lost yet?” 😊

Next Time: More on the SVD

We save the proof that indeed every matrix A has a SVD for next lecture.

We also discuss the connection between the SVD and (the more familiar?) eigenvalue decomposition.

Further we make connections between the SVD and the rank, range, and null-space of A ... etc...

It takes some time to digest the SVD...

We will return to the computation of the SVD later, when we have developed a toolbox of numerical algorithms.

Homework #2 — Due at 12:00pm, Friday February 14, 2020

Figure out how to get your favorite piece of mathematical software (e.g. Matlab) to compute the SVD.

Use your software (NOT “hand calculation”) to solve tb-4.1 and tb-4.3 (pp.30–31).

Hint: To get started in matlab, try **help svd**, and **help plot**.