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 2016

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: 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: H"older and Cauchy-Bunyakovsky-Schwarz

Last time we noted that the 1-norm and \( \infty \)-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 expressions like the ones above.
However, we can usually find useful bounds on vector- and matrix-norms, using the Hölder inequality:

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

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

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

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

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

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

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

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

Hence,

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

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

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

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

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

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

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

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

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):

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

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

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

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

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

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)
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
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 \( \sigma_1, \sigma_2, \ldots, \sigma_m \) in some orthogonal directions \( \bar{u}_1, \bar{u}_2, \ldots, \bar{u}_m \).

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

---

Example #1: SVD of a 2 \times 2 Matrix

For \( A \in \mathbb{R}^{m \times n} \), if \( \text{rank}(A) = r \), then exactly \( r \) of the lengths \( \sigma_i \) 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 \times 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 #2: SVD of a 2 \times 2 Matrix

For \( A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \), the SVD is given by

\[
\text{SVD} \left( \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \right) = \left[ \begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right] \left[ \begin{array}{cc} 2 & 0 \\ 0 & 1 \end{array} \right] \left[ \begin{array}{cc} 1 & 0 \\ \sigma_1 & \sigma_2 \end{array} \right]^{*}
\]

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

Here, the principal semi-axes of the ellipse are

\[
\sigma_1 \bar{u}_1 = 1.8011 \begin{bmatrix} 0.2955 \\ 0.9553 \end{bmatrix}, \quad \sigma_2 \bar{u}_2 = 0.5011 \begin{bmatrix} -0.9553 \\ 0.2955 \end{bmatrix}
\]
The Singular Value Decomposition

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

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

Let $A \in \mathbb{R}^{m \times n} (m \geq n)$ be of full rank, i.e. $\text{rank}(A) = n$, and let $A S^{n-1}$ 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 $A S^{n-1}$ (some lengths may be zero), written as $\sigma_1, \sigma_2, \ldots, \sigma_n$. By convention, they are ordered in descending order, so that

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

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

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

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

$$A \bar{v}_j = \sigma_j \bar{u}_j.$$
The Reduced and Full SVDs

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

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 \(\mathbb{C}^m\).

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 \(\Sigma\), by adding \(n - m\) rows of zeros at the bottom of \(\hat{\Sigma}\).

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 \(\hat{u}_j\) 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 \(\Sigma\) will have \(r\) positive diagonal entries, with the remaining \(n - r\) equal to zero.
The SVD: 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

- $U \in \mathbb{C}^{m \times m}$ is unitary
- $V \in \mathbb{C}^{n \times n}$ is unitary
- $\Sigma \in \mathbb{R}^{m \times n}$ is diagonal

The diagonal entries of $\Sigma$ are non-negative, and ordered in decreasing order, i.e. $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p \geq 0$, 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 $\Sigma$ 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 11:00am, Friday February 19, 2016

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

Use your software to solve tb-4.1 and tb-4.3 (pp.30–31).

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

The web is full of matlab tutorials and help; you may want to check out Joe Mahaffy’s Lab Manual (MATLAB and Maple).