Lectures on Dynamic Systems and Control

Mohammed Dahleh   Munther A. Dahleh   George Verghese
Department of Electrical Engineering and Computer Science
Massachuasetts Institute of Technology[1]

Chapter 4

Matrix Norms and Singular Value Decomposition

4.1 Introduction

In this lecture, we introduce the notion of a *norm* for matrices. The *singular value decomposition* or SVD of a matrix is then presented. The SVD exposes the 2-norm of a matrix, but its value to us goes much further: it enables the solution of a class of *matrix perturbation problems* that form the basis for the stability robustness concepts introduced later; it solves the so-called *total least squares* problem, which is a generalization of the least squares estimation problem considered earlier; and it allows us to clarify the notion of *conditioning*, in the context of matrix inversion. These applications of the SVD are presented at greater length in the next lecture.

Example 4.1 To provide some immediate motivation for the study and application of matrix norms, we begin with an example that clearly brings out the issue of matrix conditioning with respect to inversion. The question of interest is how sensitive the inverse of a matrix is to perturbations of the matrix.

Consider inverting the matrix

$$A = \begin{pmatrix} 100 & 100 \\ 100.2 & 100 \end{pmatrix} \tag{4.1}$$

A quick calculation shows that

$$A^{-1} = \begin{pmatrix} -5 & 5 \\ 5.01 & -5 \end{pmatrix} \tag{4.2}$$

Now suppose we invert the perturbed matrix

$$A + \Delta A = \begin{pmatrix} 100 & 100 \\ 100.1 & 100 \end{pmatrix} \tag{4.3}$$

The result now is

$$(A + \Delta A)^{-1} = A^{-1} + \Delta(A^{-1}) = \begin{pmatrix} -10 & 10 \\ 10.01 & -10 \end{pmatrix} \tag{4.4}$$

Here ΔA denotes the perturbation in A and $\Delta(A^{-1})$ denotes the resulting perturbation in A^{-1}. Evidently a 0.1% change in one entry of A has resulted in a 100% change in the entries of A^{-1}. If we want to solve the problem $Ax = b$ where $b = [1 \ \ -1]^T$, then $x = A^{-1}b = [-10 \ \ 10.01]^T$, while after perturbation of A we get $x + \Delta x = (A + \Delta A)^{-1}b = [-20 \ \ 20.01]^T$. Again, we see a 100% change in the entries of the solution with only a 0.1% change in the starting data.

The situation seen in the above example is much worse than what can ever arise in the scalar case. If a is a scalar, then $d(a^{-1})/(a^{-1}) = -da/a$, so the fractional change in the inverse of a has the same maginitude as the fractional change in a itself. What is seen in the above example, therefore, is a purely matrix phenomenon. It would seem to be related to the fact that A is nearly singular — in the sense that its columns are nearly dependent, its determinant is much smaller than its largest element, and so on. In what follows (see next lecture), we shall develop a sound way to measure nearness to singularity, and show how this measure relates to sensitivity under inversion.

Before understanding such sensitivity to perturbations in more detail, we need ways to measure the "magnitudes" of vectors and matrices. We have already introduced the notion of vector norms in Lecture 1, so we now turn to the definition of matrix norms.

4.2 Matrix Norms

An $m \times n$ complex matrix may be viewed as an operator on the (finite dimensional) normed vector space $\mathbb{C}^n$:

$$A^{m\times n} \ : \ (\, \mathbb{C}^n, \|\cdot\|_2 \,) \longrightarrow (\, \mathbb{C}^m, \|\cdot\|_2 \,) \tag{4.5}$$

where the norm here is taken to be the standard Euclidean norm. Define the **induced 2-norm** of A as follows:

$$\begin{aligned} \|A\|_2 &\triangleq \sup_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} &(4.6)\\ &= \max_{\|x\|_2 = 1} \|Ax\|_2 \ . &(4.7) \end{aligned}$$

The term "induced" refers to the fact that the definition of a norm for *vectors* such as Ax and x is what enables the above definition of a *matrix* norm. From this definition, it follows that the induced norm measures the amount of "amplification" the matrix A provides to vectors on the unit sphere in $\mathbb{C}^n$, i.e. it measures the "gain" of the matrix.

Rather than measuring the vectors x and Ax using the 2-norm, we could use any p-norm, the interesting cases being $p = 1, 2, \infty$. Our notation for this is

$$\|A\|_p = \max_{\|x\|_p = 1} \|Ax\|_p \ . \tag{4.8}$$

An important question to consider is whether or not the induced norm is actually a norm, in the sense defined for vectors in Lecture 1. Recall the three conditions that define a norm:

1. $\|x\| \geq 0$, and $\|x\| = 0 \Longleftrightarrow x = 0$;
2. $\|\alpha x\| = |\alpha|\ \|x\|$;
3. $\|x + y\| \leq \|x\| + \|y\|$.

Now let us verify that $\|A\|_p$ is a norm on $\mathbb{C}^{m \times n}$ — using the preceding definition:

1. $\|A\|_p \geq 0$ since $\|Ax\|_p \geq 0$ for any x. Futhermore, $\|A\|_p = 0 \Longleftrightarrow A = 0$, since $\|A\|_p$ is calculated from the *maximum* of $\|Ax\|_p$ evaluated on the unit sphere.
2. $\|\alpha A\|_p = |\alpha|\ \|A\|_p$ follows from $\|\alpha y\|_p = |\alpha|\ \|y\|_p$ (for any y).
3. The triangle inequality holds since:

$$\begin{aligned} \|A + B\|_p &= \max_{\|x\|_p = 1} \|(A + B)x\|_p \\ &\leq \max_{\|x\|_p = 1} \left(\|Ax\|_p + \|Bx\|_p \right) \\ &\leq \|A\|_p + \|B\|_p \ . \end{aligned}$$

Induced norms have two additional properties that are very important:

1. $\|Ax\|_p \leq \|A\|_p \|x\|_p$, which is a direct consequence of the definition of an induced norm;
2. For $A^{m \times n}$, $B^{n \times r}$,

$$\|AB\|_p \leq \|A\|_p \|B\|_p \tag{4.9}$$

which is called the *submultiplicative property.* This also follows directly from the definition:

$$\begin{aligned} \|ABx\|_p &\leq \|A\|_p \|Bx\|_p \\ &\leq \|A\|_p \|B\|_p \|x\|_p \ \text{ for any } x. \end{aligned}$$

Dividing by $\|x\|_p$:

$$\frac{\|ABx\|_p}{\|x\|_p} \leq \|A\|_p \|B\|_p \ ,$$

from which the result follows.

Before we turn to a more detailed study of ideas surrounding the induced 2-norm, which will be the focus of this lecture and the next, we make some remarks about the other induced norms of practical interest, namely the induced 1-norm and induced ∞-norm. We shall also say something about an important matrix norm that is *not* an induced norm, namely the *Frobenius* norm.

It is a fairly simple exercise to prove that

$$\|A\|_1 = \max_{1 \le j \le n} \sum_{i=1}^{m} |a_{ij}| \quad \text{(max of absolute column sums of } A) \quad , \tag{4.10}$$

and

$$\|A\|_\infty = \max_{1 \le i \le m} \sum_{j=1}^{n} |a_{ij}| \quad \text{(max of absolute row sums of } A) \quad . \tag{4.11}$$

(Note that these definitions reduce to the familiar ones for the 1-norm and ∞-norm of *column vectors* in the case $n = 1$.)

The proof for the induced ∞-norm involves two stages, namely:

1. Prove that the quantity in Equation (4.11) provides an upper bound γ:

$$\|Ax\|_\infty \le \gamma \|x\|_\infty \quad \forall x \quad ;$$

2. Show that this bound is achievable for some $x = \hat{x}$:

$$\|A\hat{x}\|_\infty = \gamma \|\hat{x}\|_\infty \quad \text{for some } \hat{x} \quad .$$

In order to show how these steps can be implemented, we give the details for the ∞-norm case. Let $x \in \mathbb{C}^n$ and consider

$$\begin{aligned}
\|Ax\|_\infty &= \max_{1 \le i \le m} |\sum_{j=1}^{n} a_{ij} x_j| \\
&\le \max_{1 \le i \le m} \sum_{j=1}^{n} |a_{ij}||x_j| \\
&\le \left(\max_{1 \le i \le m} \sum_{j=1}^{n} |a_{ij}| \right) \max_{1 \le j \le n} |x_j| \\
&= \left(\max_{1 \le i \le m} \sum_{j=1}^{n} |a_{ij}| \right) \|x\|_\infty
\end{aligned}$$

The above inequalities show that an upper bound γ is given by

$$\max_{\|x\|_\infty = 1} \|Ax\|_\infty \le \gamma = \max_{1 \le i \le m} \sum_{j=1}^{n} |a_{ij}|.$$

Now in order to show that this upper bound is achieved by some vector $\hat{x}$, let $\bar{i}$ be an index at which the expression of γ achieves a maximum, that is $\gamma = \sum_{j=1}^{n} |a_{\bar{i}j}|$. Define the vector $\hat{x}$ as

$$\hat{x} = \begin{bmatrix} sgn(a_{\bar{i}1}) \\ sgn(a_{\bar{i}2}) \\ \vdots \\ sgn(a_{\bar{i}n}) \end{bmatrix}.$$

Clearly $\|\hat{x}\|_\infty = 1$ and

$$\|A\hat{x}\|_\infty = \sum_{j=1}^{n} |a_{\bar{i}j}| = \gamma.$$

The proof for the 1-norm proceeds in exactly the same way, and is left to the reader.

There are matrix norms — i.e. functions that satisfy the three defining conditions stated earlier — that are *not* induced norms. The most important example of this for us is the **Frobenius norm**:

$$\|A\|_F \triangleq \left(\sum_{j=1}^{n} \sum_{i=1}^{m} |a_{ij}|^2 \right)^{\frac{1}{2}} \tag{4.12}$$

$$= \left(\mathrm{trace}(A'A)\right)^{\frac{1}{2}} \quad \text{(verify)} \tag{4.13}$$

In other words, the Frobenius norm is defined as the root sum of squares of the entries, i.e. the usual Euclidean 2-norm of the matrix when it is regarded simply as a vector in $\mathbb{C}^{mn}$. Although it can be shown that it is not an induced matrix norm, the Frobenius norm still has the submultiplicative property that was noted for induced norms. Yet other matrix norms may be defined (some of them without the submultiplicative property), but the ones above are the only ones of interest to us.

4.3 Singular Value Decomposition

Before we discuss the singular value decomposition of matrices, we begin with some matrix facts and definitions.

Some Matrix Facts:

- A matrix $U \in \mathbb{C}^{n\times n}$ is unitary if $U'U = UU' = I$. Here, as in Matlab, the superscript $'$ denotes the (entry-by-entry) *complex conjugate* of the *transpose*, which is also called the *Hermitian transpose* or *conjugate transpose*.

- A matrix $U \in \mathbb{R}^{n\times n}$ is orthogonal if $U^TU = UU^T = I$, where the superscript ^{T} denotes the transpose.

- Property: If U is unitary, then $\|Ux\|_2 = \|x\|_2$.

- If $S = S'$ (*i.e.* S equals its Hermitian transpose, in which case we say S is *Hermitian*), then there exists a unitary matrix such that $U'SU =$ [diagonal matrix].[1]
- For any matrix A, both $A'A$ and AA' are Hermitian, and thus can always be diagonalized by unitary matrices.
- For any matrix A, the eigenvalues of $A'A$ and AA' are always real and non-negative (proved easily by contradiction).

Theorem 4.1 (Singular Value Decomposition, or SVD) Given any matrix $A \in \mathbb{C}^{m \times n}$, A can be written as

$$A = \overset{m\times m}{U} \overset{m\times n}{\Sigma} \overset{n\times n}{V'} \quad , \tag{4.14}$$

where $U'U = I$, $V'V = I$,

$$\Sigma = \left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & 0 \\ & & \sigma_r & \\ \hline & 0 & & 0 \end{array} \right] \quad , \tag{4.15}$$

and $\sigma_i = \sqrt{i\text{th nonzero eigenvalue of } A'A}$. The σ_i are termed the **singular values** of A, and are arranged in order of descending magnitude, *i.e.*,

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0 \ .$$

Proof: We will prove this theorem for the case $\text{rank}(A) = m$; the general case involves very little more than what is required for this case. The matrix AA' is Hermitian, and it can therefore be diagonalized by a unitary matrix $U \in \mathbb{C}^{m \times m}$, so that

$$U \Lambda_1 U' = AA'.$$

Note that $\Lambda_1 = diag(\lambda_1, \lambda_2, \ldots, \lambda_m)$ has real positive diagonal entries λ_i due to the fact that AA' is positive definite. We can write $\Lambda_1 = \Sigma_1^2 = diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2)$. Define $V_1' \in \mathbb{C}^{m \times n}$ by $V_1' = \Sigma_1^{-1} U' A$. V_1' has orthonormal rows as can be seen from the following calculation: $V_1' V_1 = \Sigma_1^{-1} U' A A' U \Sigma_1^{-1} = I$. Choose the matrix V_2' in such a way that

$$V' = \begin{bmatrix} V_1' \\ V_2' \end{bmatrix}$$

is in $\mathbb{C}^{n \times n}$ and unitary. Define the $m \times n$ matrix $\Sigma = [\Sigma_1 | 0]$. This implies that

$$\Sigma V' = \Sigma_1 V_1' = U' A.$$

In other words we have $A = U \Sigma V'$.

[1]One cannot always diagonalize an arbitrary matrix—cf the Jordan form.

Example 4.2 For the matrix A given at the beginning of this lecture, the SVD — computed easily in Matlab by writing $[u, s, v] = svd(A)$ — is

$$A = \begin{pmatrix} .7068 & .7075 \\ .7075 & .-7068 \end{pmatrix} \begin{pmatrix} 200.1 & 0 \\ 0 & 0.1 \end{pmatrix} \begin{pmatrix} .7075 & .7068 \\ -.7068 & .7075 \end{pmatrix} \tag{4.16}$$

Observations:

i)

$$\begin{aligned} AA' &= U\Sigma V'V\Sigma^T U' \\ &= U\Sigma\Sigma^T U' \\ &= U \left[\begin{array}{ccc|c} \sigma_1^2 & & & \\ & \ddots & & 0 \\ & & \sigma_r^2 & \\ \hline & 0 & & 0 \end{array} \right] U' , \end{aligned} \tag{4.17}$$

which tells us U diagonalizes AA';

ii)

$$\begin{aligned} A'A &= V\Sigma^T U'U\Sigma V' \\ &= V\Sigma^T\Sigma V' \\ &= V \left[\begin{array}{ccc|c} \sigma_1^2 & & & \\ & \ddots & & 0 \\ & & \sigma_r^2 & \\ \hline & 0 & & 0 \end{array} \right] V' , \end{aligned} \tag{4.18}$$

which tells us V diagonalizes $A'A$;

iii) If U and V are expressed in terms of their columns, *i.e.*,

$$U = \begin{bmatrix} u_1 & u_2 & \cdots & u_m \end{bmatrix}$$

and

$$V = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} ,$$

then

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i' , \tag{4.19}$$

which is another way to write the SVD. The u_i are termed the **left singular vectors** of A, and the v_i are its **right singular vectors**. From this we see that we can alternately interpret Ax as

$$Ax = \sum_{i=1}^{r} \sigma_i \, u_i \, \underbrace{(v_i' x)}_{\text{projection}} \quad , \tag{4.20}$$

which is a weighted sum of the u_i, where the weights are the products of the singular values and the projections of x onto the v_i.

Observation (iii) tells us that $\mathcal{R}a(A) = \text{span}\,\{u_1, \; \ldots \; u_r\}$ (because $Ax = \sum_{i=1}^{r} c_i u_i$ — where the c_i are scalar weights). Since the columns of U are independent, dim $\mathcal{R}a(A) = r = \text{rank}\;(A)$, and $\{u_1, \; \ldots \; u_r\}$ constitute a *basis* for the range space of A. The null space of A is given by $\text{span}\{v_{r+1}, \ldots, v_n\}$. To see this:

$$\begin{aligned}
U\Sigma V' x = 0 \quad & \Longleftrightarrow \quad \Sigma V' x = 0 \\
& \Longleftrightarrow \quad \begin{bmatrix} \sigma_1 v_1' x \\ \vdots \\ \sigma_r v_r' x \end{bmatrix} = 0 \\
& \Longleftrightarrow \quad v_i' x = 0 \;\; , \;\; i = 1, \; \ldots \; , r \\
& \Longleftrightarrow \quad x \in \text{span}\{v_{r+1}, \ldots, v_n\}.
\end{aligned}$$

Example 4.3 One application of singular value decomposition is to the solution of a system of algebraic equations. Suppose A is an $m \times n$ complex matrix and b is a vector in $\mathbb{C}^m$. Assume that the rank of A is equal to k, with $k < m$. We are looking for a solution of the linear system $Ax = b$. By applying the singular value decomposition procedure to A, we get

$$\begin{aligned}
A &= U\Sigma V' \\
&= U \left[\begin{array}{c|c} \Sigma_1 & 0 \\ \hline 0 & 0 \end{array} \right] V'
\end{aligned}$$

where Σ_1 is a $k \times k$ non-singular diagonal matrix. We will express the unitary matrices U and V columnwise as

$$\begin{aligned}
U &= \begin{bmatrix} u_1 & u_2 & \ldots & u_m \end{bmatrix} \\
V &= \begin{bmatrix} v_1 & v_2 & \ldots & v_n \end{bmatrix}.
\end{aligned}$$

A necessary and sufficient condition for the solvability of this system of equations is that $u_i' b = 0$ for all i satisfying $k < i \leq m$. Otherwise, the system of equations is inconsistent. This condition means that the vector b must be orthogonal to the last $m-k$ columns of U. Therefore the system of linear equations can be written as

$$\left[\begin{array}{c|c} \Sigma_1 & 0 \\ \hline 0 & 0 \end{array}\right] V'x = U'b$$

$$\left[\begin{array}{c|c} \Sigma_1 & 0 \\ \hline 0 & 0 \end{array}\right] V'x = \begin{bmatrix} u_1'b \\ u_2'b \\ \vdots \\ \vdots \\ u_m'b \end{bmatrix} = \begin{bmatrix} u_1'b \\ \vdots \\ u_k'b \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Using the above equation and the invertibility of Σ_1, we can rewrite the system of equations as

$$\begin{bmatrix} v_1' \\ v_2' \\ \vdots \\ v_k' \end{bmatrix} x = \begin{bmatrix} \frac{1}{\sigma_1}u_1'b \\ \frac{1}{\sigma_2}u_2'b \\ \dots \\ \frac{1}{\sigma_k}u_k'b \end{bmatrix}$$

By using the fact that

$$\begin{bmatrix} v_1' \\ v_2' \\ \vdots \\ v_k' \end{bmatrix} \begin{bmatrix} v_1 & v_2 & \dots & v_k \end{bmatrix} = I,$$

左乘 （v1, v2, v3, .. vk）

we obtain a solution of the form

$$x = \sum_{i=1}^{k} \frac{1}{\sigma_i} u_i'b \, v_i.$$

From the observations that were made earlier, we know that the vectors $v_{k+1}, v_{k+2}, \dots, v_n$ span the kernel of A, and therefore a general solution of the system of linear equations is given by

$$x = \sum_{i=1}^{k} \frac{1}{\sigma_i}(u_i'b) \, v_i + \sum_{i=k+1}^{n} \beta_i v_i,$$

where the coefficients β_i, with i in the interval $k+1 \le i \le n$, are arbitrary complex numbers.

4.4 Relationship to Matrix Norms

The singular value decomposition can be used to compute the induced 2-norm of a matrix A.

Theorem 4.2

$$\begin{aligned} \|A\|_2 &\triangleq \sup_{x\neq 0} \frac{\|Ax\|_2}{\|x\|_2} \\ &= \sigma_1 \\ &= \sigma_{max}(A) \ , \end{aligned} \tag{4.21}$$

which tells us that the maximum amplification is given by the maximum singular value.

Proof:

$$\begin{aligned} \sup_{x\neq 0} \frac{\|Ax\|_2}{\|x\|_2} &= \sup_{x\neq 0} \frac{\|U\Sigma V'x\|_2}{\|x\|_2} \\ &= \sup_{x\neq 0} \frac{\|\Sigma V'x\|_2}{\|x\|_2} \\ &= \sup_{y\neq 0} \frac{\|\Sigma y\|_2}{\|Vy\|_2} \\ &= \sup_{y\neq 0} \frac{\left(\sum_{i=1}^{r} \sigma_i^2 |y_i|^2\right)^{\frac{1}{2}}}{\left(\sum_{i=1}^{r} |y_i|^2\right)^{\frac{1}{2}}} \\ &\leq \sigma_1 \ . \end{aligned}$$

For $y = [1\ 0\ \cdots\ 0]^T$, $\|\Sigma y\|_2 = \sigma_1$, and the supremum is attained. (Notice that this correponds to $x = v_1$. Hence, $Av_1 = \sigma_1 u_1$.)

Another application of the singular value decomposition is in computing the *minimal* amplification a full rank matrix exerts on elements with 2-norm equal to 1.

Theorem 4.3 Given $A \in \mathbb{C}^{m\times n}$, suppose $\mathrm{rank}(A) = n$. Then

$$\min_{\|x\|_2=1} \|Ax\|_2 = \sigma_n(A) \ . \tag{4.22}$$

Note that if $\mathrm{rank}(A) < n$, then there is an x such that the minimum is zero (rewrite A in terms of its SVD to see this).
Proof: For any $\|x\|_2 = 1$,

$$\begin{aligned} \|Ax\|_2 &= \|U\Sigma V'x\|_2 \\ &= \|\Sigma V'x\|_2 \ \text{(invariant under multiplication by unitary matrices)} \\ &= \|\Sigma y\|_2 \end{aligned}$$

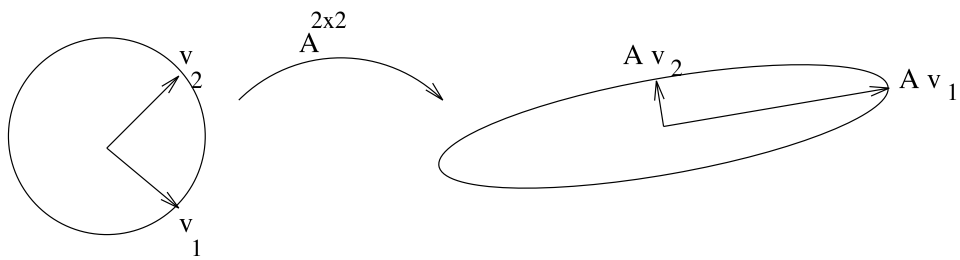

Figure 4.1: Graphical depiction of the mapping involving $A^{2\times 2}$. Note that $Av_1 = \sigma_1 u_1$ and that $Av_2 = \sigma_2 u_2$.

for $y = V'x$. Now

$$\begin{aligned} \|\Sigma y\|_2 &= \left(\sum_{i=1}^{n} |\sigma_i y_i|^2 \right)^{\frac{1}{2}} \\ &\geq \sigma_n \ . \end{aligned}$$

Note that the minimum is achieved for $y = [0\ 0\ \cdots\ 0\ 1]^T$; thus the proof is complete.

The Frobenius norm can also be expressed quite simply in terms of the singular values. We leave you to verify that

$$\begin{aligned} \|A\|_F &\triangleq \left(\sum_{j=1}^{n} \sum_{i=1}^{m} |a_{ij}|^2 \right)^{\frac{1}{2}} \\ &= (\text{trace}(A'A))^{\frac{1}{2}} \\ &= \left(\sum_{i=1}^{r} \sigma_i^2 \right)^{\frac{1}{2}} \end{aligned} \tag{4.23}$$

Example 4.4 Matrix Inequality

We say $A \leq B$, two square matrices, if

$$x'Ax \leq x'Bx \qquad \text{for all } x \neq 0.$$

It follows that for any matrix A, not necessarily square,

$$\|A\|_2 \leq \gamma \leftrightarrow A'A \leq \gamma^2 I.$$

Exercises

Exercise 4.1 Verify that for any A, an $m \times n$ matrix, the following holds:

$$\frac{1}{\sqrt{n}}||A||_1 \leq ||A||_2 \leq \sqrt{m}||A||_\infty.$$

Exercise 4.2 Suppose $A' = A$. Find the exact relation between the eigenvalues and singular values of A. Does this hold if A is not conjugate symmetric?

Exercise 4.3 Show that if $\text{rank}(A) = 1$, then, $||A||_F = ||A||_2$.

Exercise 4.4 This problem leads you through the argument for the existence of the SVD, using an iterative construction. Showing that $A = U\Sigma V'$, where U and V are unitary matrices is equivalent to showing that $U'AV = \Sigma$.

a) Argue from the definition of $||A||_2$ that there exist unit vectors (measured in the 2-norm) $x \in \mathbb{C}^n$ and $y \in \mathbb{C}^m$ such that $Ax = \sigma y$, where $\sigma = ||A||_2$.

b) We can extend both x and y above to orthonormal bases, i.e. we can find *unitary matrices* V_1 and U_1 whose first columns are x and y respectively:

$$V_1 = [x \ \tilde{V}_1] \ , \qquad U_1 = [y \ \tilde{U}_1]$$

Show that one way to do this is via Householder transformations, as follows:

$$V_1 = I - 2\frac{hh'}{h'h} \ , \qquad h = x - [1, 0, \ldots, 0]'$$

and likewise for U_1.

c) Now define $A_1 = U_1' A V_1$. Why is $||A_1||_2 = ||A||_2$?

d) Note that

$$A_1 = \begin{pmatrix} y'Ax & y'A\tilde{V}_1 \\ \tilde{U}_1'Ax & \tilde{U}_1'A\tilde{V}_1 \end{pmatrix} = \begin{pmatrix} \sigma & w' \\ 0 & B \end{pmatrix}$$

What is the justification for claiming that the lower left element in the above matrix is 0?

e) Now show that

$$||A_1 \begin{pmatrix} \sigma \\ w \end{pmatrix}||_2 \geq \sigma^2 + w'w$$

and combine this with the fact that $||A_1||_2 = ||A||_2 = \sigma$ to deduce that $w = 0$, so

$$A_1 = \begin{pmatrix} \sigma & 0 \\ 0 & B \end{pmatrix}$$

At the next iteration, we apply the above procedure to B, and so on. When the iterations terminate, we have the SVD.

[The reason that this is only an existence proof and not an algorithm is that it begins by invoking the existence of x and y, but does not show how to compute them. Very good algorithms do exist for computing the SVD — see Golub and Van Loan's classic, *Matrix Computations*, Johns Hopkins Press, 1989. The SVD is a cornerstone of numerical computations in a host of applications.]

Exercise 4.5 Suppose the $m \times n$ matrix A is decomposed in the form

$$A = U \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} V'$$

where U and V are unitary matrices, and Σ is an invertible $r \times r$ matrix (— the SVD could be used to produce such a decomposition). Then the "Moore-Penrose inverse", or *pseudo-inverse* of A, denoted by A^+, can be defined as the $n \times m$ matrix

$$A^+ = V \begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{pmatrix} U'$$

(You can invoke it in Matlab with `pinv(A)`.)

a) Show that A^+A and AA^+ are symmetric, and that $AA^+A = A$ and $A^+AA^+ = A^+$. (These four conditions actually constitute an alternative definition of the pseudo-inverse.)

b) Show that when A has full column rank then $A^+ = (A'A)^{-1}A'$, and that when A has full row rank then $A^+ = A'(AA')^{-1}$.

c) Show that, of all x that minimize $\|y - Ax\|_2$ (and there will be many, if A does not have full column rank), the one with smallest length $\|x\|_2$ is given by $\hat{x} = A^+y$.

Exercise 4.6 All the matrices in this problem are real. Suppose

$$A = Q \begin{pmatrix} R \\ 0 \end{pmatrix}$$

with Q being an $m \times m$ orthogonal matrix and R an $n \times n$ invertible matrix. (Recall that such a decomposition exists for any matrix A that has full column rank.) Also let Y be an $m \times p$ matrix of the form

$$Y = Q \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$$

where the partitioning in the expression for Y is conformable with the partitioning for A.

(a) What choice $\hat{X}$ of the $n \times p$ matrix X minimizes the Frobenius norm, or equivalently the squared Frobenius norm, of $Y - AX$? In other words, find

$$\hat{X} = \text{argmin}\ ||Y - AX||_F^2$$

Also determine the value of $||Y - A\hat{X}||_F^2$. (Your answers should be expressed in terms of the matrices Q, R, Y_1 and Y_2.)

(b) Can your $\hat{X}$ in (a) also be written as $(A'A)^{-1}A'Y$? Can it be written as $A^{+}Y$, where A^{+} denotes the (Moore-Penrose) pseudo-inverse of A ?

(c) Now obtain an expression for the choice $\overline{X}$ of X that minimizes

$$||Y - AX||_F^2 + ||Z - BX||_F^2$$

where Z and B are given matrices of appropriate dimensions. (Your answer can be expressed in terms of A, B, Y, and Z.)

Exercise 4.7 Structured Singular Values

Given a complex square matrix A, define the *structured singular value function* as follows.

$$\mu_{\underline{\Delta}}(A) = \frac{1}{\min_{\Delta \in \underline{\Delta}}\{\sigma_{max}(\Delta) \mid \det(I - \Delta A) = 0\}}$$

where $\underline{\Delta}$ is some set of matrices.

a) If $\underline{\Delta} = \{\alpha I\ :\ \alpha \in \mathbb{C}\}$, show that $\mu_{\underline{\Delta}}(A) = \rho(A)$, where ρ is the *spectral radius* of A, defined as: $\rho(A) = \max_i |\lambda_i|$ and the λ_i's are the eigenvalues of A.

b) If $\underline{\Delta} = \{\Delta \in \mathbb{C}^{n \times n}\}$, show that $\mu_{\underline{\Delta}}(A) = \sigma_{max}(A)$

c) If $\underline{\Delta} = \{diag(\alpha_1, \cdots, \alpha_n) \mid \alpha_i \in \mathbb{C}\}$, show that

$$\rho(A) \le \mu_{\underline{\Delta}}(A) = \mu_{\underline{\Delta}}(D^{-1}AD) \le \sigma_{max}(D^{-1}AD)$$

where

$$D \in \{\text{diag}(d_1, \cdots, d_n) \mid d_i > 0\}$$

Exercise 4.8 Consider again the *structured singular value function* of a complex square matrix A defined in the preceding problem. If A has more structure, it is sometimes possible to compute $\mu_{\underline{\Delta}}(A)$ exactly. In this problem, assume A is a rank-one matrix, so that we can write $A = uv'$ where u, v are complex vectors of dimension n. Compute $\mu_{\underline{\Delta}}(A)$ when

(a) $\underline{\Delta} = diag(\delta_1, \ldots, \delta_n), \quad \delta_i \in \mathbb{C}$.

(b) $\underline{\Delta} = diag(\delta_1, \ldots, \delta_n), \quad \delta_i \in \mathbb{R}$.

To simplify the computation, minimize the Frobenius norm of Δ in the defintion of $\mu_{\underline{\Delta}}(A)$.

6.241J / 16.338J Dynamic Systems and Control
Spring 2011