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 5

Matrix Perturbations

5.1 Introduction

The following question arises frequently in matrix theory: What is the smallest possible perturbation of a matrix that causes it to lose rank? We discuss two cases next, with perturbations measured in the 2-norm, and then discuss the measurement of perturbations in the Frobenius norm. This provides us with a new formulation to the least squares estimation problem in which uncertainty is present in the matrix A as well as the vector y. This is known as *total least squares*.

5.2 Additive Perturbation

Theorem 5.1 Suppose $A \in \mathbb{C}^{m \times n}$ has full column rank $(= n)$. Then

$$\min_{\Delta \in \mathbb{C}^{m \times n}} \{\|\Delta\|_2 \mid A + \Delta \text{ has rank } < n\} = \sigma_n(A) \ . \tag{5.1}$$

Proof: Suppose $A + \Delta$ has rank $< n$. Then there exists $x \neq 0$ such that $\|x\|_2 = 1$ and

$$(A + \Delta)\, x = 0 \ .$$

Since $\Delta x = -Ax$,

$$
\begin{aligned}
\|\Delta x\|_2 &= \|Ax\|_2 \\
&\geq \sigma_n(A) \ .
\end{aligned}
\tag{5.2}
$$

From the properties of induced norms (see Section 3.1), we also know that

$$\|\Delta\|_2 \|x\|_2 \geq \|\Delta x\|_2.$$

Using Equation (24.3) and the fact that $\|x\|_2 = 1$, we arrive at the following:

$$\begin{aligned}
\|\Delta\|_2 &\geq \|\Delta x\|_2 \\
&\geq \sigma_n(A)
\end{aligned} \tag{5.3}$$

To complete the proof, we must show that the lower bound from Equation (5.3) can be achieved. Thus, we must construct a Δ so that $A + \Delta$ has rank $< n$ and $\|\Delta\|_2 = \sigma_n(A)$; such a Δ will be a minimizing solution. For this, choose

$$\Delta = -\sigma_n u_n v_n'$$

where u_n, v_n are the left and right singular vectors associated with the smallest singular value σ_n of A. Notice that $\|\Delta\|_2 = \sigma_n(A)$. This choice yields

$$\begin{aligned}
(A + \Delta)\, v_n &= \sigma_n u_n - \sigma_n u_n v_n^* v_n \\
&= \sigma_n u_n - \sigma_n u_n \\
&= 0 \ .
\end{aligned}$$

That is, $A + \Delta$ has rank $< n$. This completes the proof.

5.3 Multiplicative Perturbation

Theorem 5.2 (Small Gain) Given $A \in \mathbb{C}^{m \times n}$,

$$\min_{\Delta \in \mathbb{C}^{n \times m}} \{\|\Delta\|_2 \mid I - A\Delta \text{ is singular } \} = \frac{1}{\sigma_1(A)} \ . \tag{5.4}$$

Proof: Suppose $I - A\Delta$ is singular. Then there exists $x \neq 0$ such that

$$(I - A\Delta)\, x = 0$$

so

$$\|A\Delta x\|_2 = \|x\|_2 \ . \tag{5.5}$$

From the properties of induced norms (see Lecture 4 notes),

$$\begin{aligned}
\|A\Delta x\|_2 &\leq \|A\|_2 \|\Delta x\|_2 \\
&= \sigma_1(A) \|\Delta x\|_2 \ .
\end{aligned}$$

Upon substituting the result in Equation (5.5) for $\|A\Delta x\|_2$, we find

$$\|x\|_2 \leq \sigma_1(A) \|\Delta x\|_2 \ .$$

Dividing through by $\sigma_1(A)\|x\|_2$ yields

$$\frac{\|\Delta x\|_2}{\|x\|_2} \geq \frac{1}{\sigma_1(A)} \quad ,$$

which implies

$$\|\Delta\|_2 \geq \frac{1}{\sigma_1(A)} \quad . \tag{5.6}$$

To conclude the proof, we must show that this lower bound can be achieved. Thus, we construct a Δ which satisfies Equation (5.6) with equality and also causes $(I - A\Delta)$ to be singular. For this, choose

$$\Delta = \frac{1}{\sigma_1(A)} v_1 u_1' \quad .$$

Notice that the lower bound (Equation (5.6)) is satisfied with equality, *i.e.*, $\|\Delta\|_2 = 1/\sigma_1(A)$. Now choose $x = u_1$. Then:

$$
\begin{aligned}
(I - A\Delta)\, x &= (I - A\Delta)\, u_1 \\
&= \left(I - \frac{A v_1 u_1'}{\sigma_1}\right) u_1 \\
&= u_1 - \underbrace{\frac{A v_1}{\sigma_1}}_{u_1} \\
&= u_1 - u_1 \quad (\text{since } A v_1 = \sigma_1 u_1) \\
&= 0 \quad .
\end{aligned}
$$

This completes the proof.

The theorem just proved is called the **small gain theorem**. The reason for this is that it guarantees $(I - A\Delta)$ is nonsingular provided

$$\|\Delta\|_2 < \frac{1}{\|A\|_2} \quad .$$

This condition is most often written as

$$\|\Delta\|_2 \|A\|_2 < 1 \quad , \tag{5.7}$$

i.e., the product of the gains is less than one.

Remark: We can actually obtain the additive perturbation result from multiplicative perturbation methods. Assume A is invertible, and Δ is a matrix which makes its sum with A singular. Since

$$A + \Delta = A\left(I + A^{-1}\Delta\right) \quad ,$$

and A is nonsingular, then $(I + A^{-1}\Delta)$ must be singular. By our work with multiplicative perturbations, we know that the Δ associated with the smallest $\|\Delta\|_2$ that makes this quantity singular satisfies

$$\|\Delta\|_2 = \frac{1}{\sigma_1(A^{-1})} = \sigma_n(A) \ .$$

5.4   Perturbations Measured in the Frobenius Norm

We will now demonstrate that, for the multiplicative and additive perturbation cases where we minimized the induced 2-norm, we also minimized the Frobenius norm.

Let $A \in \mathbb{C}^{m \times n}$, and let $\mathrm{rank}(A) = r$.

$$
\begin{aligned}
\|A\|_F \ &\triangleq\ \left(\sum_{j=1}^{n} \sum_{i=1}^{m} |a_{ij}|^2 \right)^{\frac{1}{2}} && \text{(5.8)}\\
&=\ \left(\mathrm{trace}(A'A) \right)^{\frac{1}{2}} && \text{(5.9)}\\
&=\ \left(\sum_{i=1}^{r} \sigma_i^2 \right)^{\frac{1}{2}} \quad \text{(the trace of a matrix is the sum of its eigenvalues)} && \text{(5.10)}\\
&\geq\ \sigma_1(A) \ . && \text{(5.11)}
\end{aligned}
$$

Therefore,

$$\|A\|_F \geq \|A\|_2 \ , \tag{5.12}$$

which is a useful inequality.

In both the perturbation problems that we considered earlier, we found a rank-one solution, or dyad, for Δ:

$$\Delta = \alpha uv' \ , \tag{5.13}$$

where $\alpha \in \mathbb{C}$, $u \in \mathbb{C}^m$, $v \in \mathbb{C}^n$ such that $\|u\|_2 = \|v\|_2 = 1$. It is easy to show that the Frobenius norm and induced 2-norm are *equal* for rank one matrices of the form in Equation (5.13). It follows from this that the Δ which minimizes the induced 2-norm also minimizes the Frobenius norm, for the additive and multiplicative perturbation cases we have examined. In general, however, minimizing the induced 2-norm of a matrix does not imply the Frobenius norm is minimized (or vice versa.)

Example 5.1   This example is intended to illustrate the use of the singular value decomposition and Frobenius norms in the solution of a minimum distance problem. Suppose we have a matrix $A \in \mathbb{C}^{n \times n}$, and we are interested in finding the closest matrix to A of the form cW where c is a complex number and W is a

unitary matrix. The distance is to be measured by the Frobenius norm. This problem can be formulated as

$$\min_{c \in \mathbb{C}, W \in \mathbb{C}^{n \times n}} \|A - cW\|_F$$

where $W'W = I$. We can write

$$
\begin{aligned}
\|A - cW\|_F^2 &= \text{Tr}\left((A - cW)'(A - cW)\right) \\
&= \text{Tr}(A'A) - c'\text{Tr}(W'A) - c\text{Tr}(A'W) + |c|^2\text{Tr}(W'W).
\end{aligned}
$$

Note that $\text{Tr}(W'W) = \text{Tr}(I) = n$. Therefore, we have

$$\|A - cW\|_F^2 = \|A\|_F^2 - 2Re\left(c'\text{Tr}(W'A)\right) + n|c|^2, \tag{5.14}$$

and by taking

$$c = \frac{1}{n}\text{Tr}(W'A)$$

the right hand side of Equation (5.14) will be minimized. Therefore we have that

$$\|A - cW\|_F^2 \geq \|A\|_F^2 - \frac{1}{n}|\text{Tr}(W'A)|^2.$$

Now we must minimize the right hand side with respect to W, which is equivalent to maximizing $|\text{Tr}(W'A)|$. In order to achieve this we employ the singular value decomposition of A as $U\Sigma V'$, which gives

$$
\begin{aligned}
|\text{Tr}(W'A)|^2 &= |\text{Tr}(W'U\Sigma V')|^2 \\
&= |\text{Tr}(V'W'U\Sigma)|^2.
\end{aligned}
$$

The matrix $Z = V'W'U$ satisfies

$$
\begin{aligned}
ZZ' &= V'W'UU'WV \\
&= I.
\end{aligned}
$$

Therefore,

$$|\text{Tr}(Z\Sigma)|^2 = |\sum_{i=1}^{n} \sigma_i z_{ii}|^2 \leq \left(\sum_{i=1}^{n} \sigma_i\right)^2,$$

implies that

$$\min_{c,W} \|A - cW\|_F^2 \geq \|A\|_F^2 - \frac{1}{n}\left(\sum_{i=1}^{n} \sigma_i\right)^2. \tag{5.15}$$

In order to complete this example we show that the lower bound in Equation (5.15) can actually be achieved with a specific choice of W. Observe that

$$\text{Tr}(W'U\Sigma V') = \text{Tr}(W'UV'\Sigma),$$

and by letting $W' = VU'$ we obtain

$$\text{Tr}(W'A) = \text{Tr}(\Sigma) = \sum_{i=1}^{n} \sigma_i$$

and

$$c = \frac{1}{n} \sum_{i=1}^{n} \sigma_i.$$

Putting all the pieces together, we get that

$$\min_{c,W} \|A - cW\|_F^2 = \sum_{i=1}^{n} \sigma_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} \sigma_i^2 \right)^2,$$

and the minimizing unitary matrix is given by

$$cW = \frac{1}{n} \left(\sum_{i=1}^{n} \sigma_i \right) UV'.$$

It is clear also that, in order for a matrix to be exactly represented as a complex multiple of a unitary matrix, all of its singular values must be equal.

5.5   Total Least Squares

We have previously examined solving least squares problems of the form $y = Ax + e$. An interpretation of the problem we solved there is that we perturbed y as little as possible — in the least squares sense — to make the resulting equation $y - e = Ax$ consistent. It is natural to ask what happens if we allow A to be perturbed as well, in addition to perturbing y. This makes sense in situations where the uncertainty in our model and the noise in our measurements cannot or should not be attributed entirely to y, but also to A. The simplest least squares problem of this type is one that allows a perturbed model of the form

$$y = (A + \Delta) x + e \ . \tag{5.16}$$

The so-called *total least squares* estimation problem can now be stated as

$$\min_{\Delta,e} \left(\sum_{i,j} |\Delta_{ij}|^2 + \sum_i |e_i|^2 \right)^{\frac{1}{2}} = \min_{\Delta,e} \|\Delta \vdots e\|_F \tag{5.17}$$

$$= \min_{\Delta,e} \|\hat{\Delta}\|_F \ , \tag{5.18}$$

where

$$\hat{\Delta} = \begin{bmatrix} \Delta \vdots e \end{bmatrix} \ . \tag{5.19}$$

Weighted versions of this problem can also be posed, but we omit these generalizations.

Note that no constraints have been imposed on Δ in the above problem statement, and this can often limit the direct usefulness of the total least squares formulation in practical problems. In practice, the expected or allowed perturbations of A are often quite structured; however, the solution of the total least squares problem under such structural constraints is much harder than that of the unconstrained problem that we present the solution of next. Nevertheless, the total least squares formulation can provide a useful benchmark. (The same sorts of comments can of course be made about the conventional least squares formulation: it is often not the criterion that we would want to use, but its tractability compared to other criteria makes it a useful point of departure.)

If we make the definitions

$$\hat{A} = \left[\begin{array}{ccc} A & \vdots & -y \end{array} \right] , \quad \hat{x} = \left[\begin{array}{c} x \\ 1 \end{array} \right] \tag{5.20}$$

then the perturbed model in Equation (5.16) can be rewritten as

$$\left(\hat{A} + \hat{\Delta} \right) \hat{x} = 0 \ . \tag{5.21}$$

This equation makes evident that what we seek is the $\hat{\Delta}$ with minimal Frobenius norm that satisfies Equation (5.21)—the smallest $\hat{\Delta}$ that makes $\hat{A} + \hat{\Delta}$ singular.

Let us suppose that A has full column rank (n), and that it has more rows than columns (which is normally the case, since in least squares estimation we typically have many more measurements than parameters to estimate). In addition, let us assume that $\hat{A}$ has rank $(n + 1)$, which is also generally true. From what we've learned about additive perturbations, we now see that a minimal (in a Frobenius sense) $\hat{\Delta}$ that satisfies Equation (5.21) is

$$\hat{\Delta} = -\sigma_{n+1} u_{n+1} v'_{n+1} \ , \tag{5.22}$$

where the σ_{n+1}, u_{n+1} and v_{n+1} are derived from the SVD of $\hat{A}$ (i.e. σ_{n+1} is the smallest singular value of $\hat{A}$, etc.). Given that we now know $\hat{A}$ and $\hat{\Delta}$, choosing $\hat{x} = v_{n+1}$, and rescaling $\hat{x}$, we have

$$\left(\hat{A} + \hat{\Delta} \right) \left[\begin{array}{c} x \\ 1 \end{array} \right] = 0 \ ,$$

which gives us x, the total least squares solution. This solution is due to Golub and Van Loan (see their classic text on *Matrix Computations*, Second Edition, Johns Hopkins University Press, 1989).

5.6   Conditioning of Matrix Inversion

We are now in a position to address some of the issues that came up in Example 1 of Lecture 4, regarding the sensitivity of the inverse A^{-1} and of the solution $x = A^{-1}b$ to perturbations

in A (and/or b, for that matter). We first consider the case where A is invertible, and examine the sensitivity of A^{-1}. Taking differentials in the defining equation $A^{-1}A = I$, we find

$$d(A^{-1})\, A + A^{-1}\, dA = 0 \;, \tag{5.23}$$

where the order of the terms in each half of the sum is important, of course. (Rather than working with differentials, we could equivalently work with perturbations of the form $A + \epsilon P$, etc., where ϵ is vanishingly small, but this really amounts to the same thing.) Rearranging the preceding expression, we find

$$d(A^{-1}) = -A^{-1}\, dA\, A^{-1} \tag{5.24}$$

Taking norms, the result is

$$\|d(A^{-1})\| \le \|A^{-1}\|^2 \|dA\| \tag{5.25}$$

or equivalently

$$\frac{\|d(A^{-1})\|}{\|A^{-1}\|} \le \|A\| \|A^{-1}\| \frac{\|dA\|}{\|A\|} \tag{5.26}$$

This derivation holds for any submultiplicative norm. The product $\|A\|\|A^{-1}\|$ is termed the *condition number* of A with respect to inversion (or simply the condition number of A) and denoted by $K(A)$:

$$K(A) = \|A\|\|A^{-1}\| \tag{5.27}$$

When we wish to specify which norm is being used, a subscript is attached to $K(A)$. Our earlier results on the SVD show, for example, that

$$K_2(A) = \sigma_{max}/\sigma_{min} \tag{5.28}$$

The condition number in this 2-norm tells us how slender the ellipsoid Ax for $\|x\|_2 = 1$ is — see Figure 5.1. In what follows, we shall focus on the 2-norm condition number (but will omit the subscript unless essential).

Some properties of the 2-norm condition number (all of which are easy to show, and some of which extend to the condition number in other norms) are

- $K(A) \ge 1$;

- $K(A) = K(A^{-1})$;

- $K(AB) \le K(A)K(B)$;

- Given $U'U = I$, $K(UA) = K(A)$.

The importance of (5.26) is that the bound can actually be attained for some choice of the perturbation dA and of the matrix norm, so the situation can get as bad as the bound allows: *the fractional change in the inverse can be $K(A)$ times as large as the fractional change in the original.* In the case of 2-norms, a particular perturbation that attains the bound
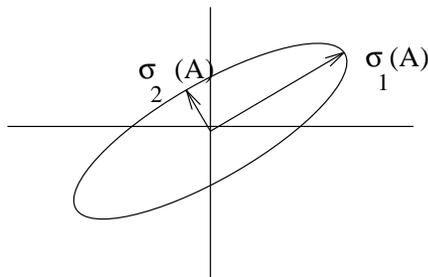
Figure 5.1: Depiction of how A (a real 2×2 matrix) maps the unit circle. The major axis of the ellipse corresponds to the largest singular value, the minor axis to the smallest.

can be derived from the Δ of Theorem 5.1, by simply replacing $-\sigma_n$ in Δ by a differential perturbation:

$$dA = -d\sigma \, u_n v'_n \tag{5.29}$$

We have established that a large condition number corresponds to a matrix whose inverse is very sensitive to relatively small perturbations in the matrix. Such a matrix is termed *ill conditioned* or poorly conditioned with respect to inversion. A perfectly conditioned matrix is one whose condition number takes the minimum possible value, namely 1.

A high condition number also indicates that a matrix is close to losing rank, in the following sense: There is a perturbation Δ of small norm $(= \sigma_{min})$ relative to $\|A\|$ $(= \sigma_{max})$ such that $A + \Delta$ has lower rank than A. This follows from our additive perturbation result in Theorem 5.1. This interpretation extends to non-square matrices as well. We shall term the ratio in (5.28) the condition number of A even when A is non-square, and think of it as a measure of *nearness to a rank loss*.

Turning now to the sensitivity of the solution $x = A^{-1}b$ of a linear system of equations in the form $Ax = b$, we can proceed similarly. Taking differentials, we find that

$$dx = -A^{-1} \, dA \, A^{-1}b + A^{-1} \, db = -A^{-1} \, dA \, x + A^{-1}b \tag{5.30}$$

Taking norms then yields

$$\|dx\| \leq \|A^{-1}\|\|dA\|\|x\| + \|A^{-1}\|\|db\| \tag{5.31}$$

Dividing both sides of this by $\|x\|$, and using the fact that $\|x\| \geq (\|b\|/\|A\|)$, we get

$$\frac{\|dx\|}{\|x\|} \leq K(A)\Big(\frac{\|dA\|}{\|A\|} + \frac{\|db\|}{\|b\|}\Big) \tag{5.32}$$

We can come close to attaining this bound if, for example, b happens to be nearly collinear with the column of U in the SVD of A that is associated with σ_{min}, and if appropriate perturbations occur. Once again, therefore, the fractional change in the answer can be close to $K(A)$ times as large as the fractional changes in the given matrices.

Example 5.2   For the matrix A given in Example 1 of Lecture 4, the SVD is

$$A = \begin{pmatrix} 100 & 100 \\ 100.2 & 100 \end{pmatrix} = \begin{pmatrix} .7068 & .7075 \\ .7075 & .-7068 \end{pmatrix} \begin{pmatrix} 200.1 & 0 \\ 0 & 0.1 \end{pmatrix} \begin{pmatrix} .7075 & .7068 \\ -.7068 & .7075 \end{pmatrix}$$

(5.33)

The condition number of A is seen to be 2001, which accounts for the 1000-fold magnification of error in the inverse for the perturbation we used in that example. The perturbation Δ of smallest 2-norm that causes $A + \Delta$ to become singular is

$$\Delta = \begin{pmatrix} .7068 & .7075 \\ .7075 & .-7068 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & -0.1 \end{pmatrix} \begin{pmatrix} .7075 & .7068 \\ -.7068 & .7075 \end{pmatrix}$$

whose norm is 0.1. Carrying out the multiplication gives

$$\Delta \approx \begin{pmatrix} .05 & -.05 \\ -.05 & .05 \end{pmatrix}$$

With $b = [1 \ -1]^T$, we saw large sensitivity of the solution x to perturbations in A. Note that this b is indeed nearly collinear with the second column of U. If, on the other hand, we had $b = [1 \ 1]$, which is more closely aligned with the first column of U, then the solution would have been hardly affected by the perturbation in A — a claim that we leave you to verify.

Thus $K(A)$ serves as a bound on the magnification factor that relates fractional changes in A or b to fractional changes in our solution x.

Conditioning of Least Squares Estimation

Our objective in the least-square-error estimation problem was to find the value $\widehat{x}$ of x that minimizes $\|y - Ax\|_2^2$, under the assumption that A has full column rank. A detailed analysis of the conditioning of this case is beyond our scope (see *Matrix Computations* by Golub and Van Loan, cited above, for a detailed treatment). We shall make do here with a statement of the main result in the case that the fractional residual is much less than 1, i.e.

$$\frac{\|y - A\widehat{x}\|_2}{\|y\|_2} \ll 1$$

(5.34)

This low-residual case is certainly of interest in practice, assuming that one is fitting a reasonably good model to the data. In this case, it can be shown that the fractional change $\|d\widehat{x}\|_2/\|\widehat{x}\|_2$ in the solution $\widehat{x}$ can approach $K(A)$ times the sum of the fractional changes in A and y, where $K(A) = \sigma_{max}(A)/\sigma_{min}(A)$. In the light of our earlier results for the case of invertible A, this result is perhaps not surprising.

Given this result, it is easy to explain why solving the normal equations

$$(A'A)\widehat{x} = A'y$$

to determine $\widehat{x}$ is numerically unattractive (in the low-residual case). The numerical inversion of $A'A$ is governed by the condition number of $A'A$, and this is the *square* of the condition number of A:

$$K(A'A) = K^2(A)$$

You should confirm this using the SVD of A. The process of directly solving the normal equations will thus introduce errors that are not intrinsic to the least-square-error problem, because this problem is governed by the condition number $K(A)$, according to the result quoted above. Fortunately, there are other algorithms for computing $\widehat{x}$ that are governed by the condition number $K(A)$ rather than the square of this (and Matlab uses one such algorithm to compute $\widehat{x}$ when you invoke its least squares solution command).

Exercises

Exercise 5.1 Suppose the complex $m \times n$ matrix A is perturbed to the matrix $A + E$.

(a) Show that
$$|\sigma_{max}(A + E) - \sigma_{max}(A)| \leq \sigma_{max}(E)$$

Also find an E that results in the inequality being achieved with equality.

(Hint: To show the inequality, write $(A + E) = A + E$ and $A = (A + E) - E$, take the 2-norm on both sides of each equation, and use the triangle inequality.)

It turns out that the result in (a) actually applies to *all* the singular values of A and $A + E$, not just the largest one. Part (b) below is one version of the result for the smallest singular value.

(b) Suppose A has *less than* full column rank, i.e. has rank$< n$, but $A + E$ has full column rank. Show (following a procedure similar to part (a) — but looking at $\min \|(A + E)x\|_2$ rather than the norm of $A + E$, etc.) that
$$\sigma_{min}(A + E) \leq \sigma_{max}(E)$$

Again find an E that results in the inequality being achieved with equality.

[The result in (b), and some extensions of it, give rise to the following sound (and widely used) procedure for estimating the rank of some underlying matrix A, given only the matrix $A + E$ and knowledge of $\|E\|_2$: Compute the SVD of $A + E$, then declare the "numerical rank" of A to be the number of singular values of $A + E$ that are larger than the threshold $\|E\|_2$. The given information is consistent with having an A of this rank.]

(c) Verify the above results using your own examples in MATLAB. You might also find it interesting to verify numerically that for large m, n, the norm of the matrix $E = s * randn(m, n)$ — which is a matrix whose entries are independent, zero-mean, Gaussian, with standard deviation s — is close to $s * (\sqrt{m} + \sqrt{n})$. So if A is perturbed by such a matrix, then a reasonable value to use as a threshold when determining the numerical rank of A is this number.

Exercise 5.2 Let A and E be $m \times n$ matrices. Show that
$$\min_{\text{rank } E \leq r} \|A - E\|_2 = \sigma_{r+1}(A).$$

To prove this, notice that the rank constraint on E can be interpreted as follows: If $v_1, \ldots, v_{r+1}$ are linearly independent vectors, then there exists a nonzero vector z, expressed as a linear combination of such vectors, that belongs to the nullspace of E. Proceed as follows:

1. Select the v_i's from the SVD of A.

2. Select a candidate element z with $\|z\|_2 = 1$.

3. Show that $\|(A - E)z\|_2 \geq \sigma_{r+1}$. This implies that $\|A - E\|_2 \geq \sigma_{r+1}$.

4. Construct an E that achieves the above bound.

Exercise 5.3 Consider the real, square system of equations $Ax = (U\Sigma V^T)x = y$, where U and V are orthogonal matrices, with

$$\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 10^{-6} \end{pmatrix} , \quad y = U \begin{pmatrix} 1 \\ 10^{-6} \end{pmatrix}$$

All norms in this problem are taken to be 2-norms.

(a) What is the norm of the exact solution x?

(b) Suppose y is perturbed to $y + \delta y$, and that correspondingly the solution changes from x in (a) to $x + \delta x$. Find a perturbation δy, with $\|\delta y\| = 10^{-6}$, such that

$$\frac{\|\delta x\|}{\|x\|} \approx \kappa(A) \frac{\|\delta y\|}{\|y\|}$$

where $\kappa(A)$ is the condition number of A.

(c) Suppose instead of perturbing y we perturb A, changing it to $A + \delta A$, with the solution correspondingly changing from x to $x + \delta x$ (for some δx that is different than in part (b)). Find a perturbation δA, with $\|\delta A\| = 10^{-7}$, such that

$$\frac{\|\delta x\|}{\|x\|} \approx \kappa(A) \frac{\|\delta A\|}{\|A\|}$$

Exercise 5.4 Positive Definite Matrices

A matrix A is positive semi-definite if $x'Ax \geq 0$ for all $x \neq 0$. We say Y is the square root of a Hermitian positive semi-definite matrix if $Y'Y = A$. Show that Y always exists and can be constructed from the SVD of A.

Exercise 5.5 Let A and B have compatible dimensions. Show that if

$$\|Ax\|_2 \leq \|Bx\|_2 \qquad \text{for all } x,$$

then there exists a matrix Y with $\|Y\|_2 \leq 1$ such that

$$A = YB.$$

Assume B has full rank to simplicity.

Exercise 5.6 (a) Suppose

$$\left\| \begin{pmatrix} X \\ A \end{pmatrix} \right\| \leq \gamma.$$

Show that there exists a matrix Y with $\|Y\|_2 \leq 1$ such that

$$X = Y(\gamma^2 I - A'A)^{\frac{1}{2}}$$

(b) Suppose

$$\|(\, X \quad A \,)\| \le \gamma.$$

Show that there exists a matrix Z with $\|Z\| \le 1$ such that $X = (\gamma^2 I - AA^*)^{\frac{1}{2}} Z$.

Exercise 5.7 Matrix Dilation

The problems above can help us prove the following important result:

$$\gamma_0 := \min_{X} \left\| \begin{pmatrix} X & B \\ C & A \end{pmatrix} \right\| = \max \left\{ \|(\, C \quad A \,)\|, \, \left\| \begin{pmatrix} B \\ A \end{pmatrix} \right\| \right\}.$$

This is known as the matrix dilation theorem. Notice that the left hand side is always greater than or equal to the right hand side irrespective of the choice of X. Below, we outline the steps necessary to prove that this lower bound is tight. Matrix dilations play an important role in systems theory particularly in model reduction problems.

1. Let γ_1 be defined as

$$\gamma_1 = \max \left\{ \|(\, C \quad A \,)\|, \, \left\| \begin{pmatrix} B \\ A \end{pmatrix} \right\| \right\}.$$

   Show that:

$$\gamma_0 \ge \gamma_1.$$

2. Use the previous exercise to show that there exists two matrices Y and Z with norms less than or equal to one such that

$$B = Y(\gamma_1^2 I - A^* A)^{\frac{1}{2}}, \quad C = (\gamma_1^2 I - AA^*)^{\frac{1}{2}} Z.$$

3. Define a candidate solution to be $\tilde{X} = -YA^* Z$. Show by direct substitution that

$$
\left\| \begin{pmatrix} \tilde{X} & B \\ C & A \end{pmatrix} \right\| = \left\| \begin{pmatrix} -YA^* Z & Y(\gamma_1^2 I - A^* A)^{\frac{1}{2}} \\ C = (\gamma_1^2 I - AA^*)^{\frac{1}{2}} Z & A \end{pmatrix} \right\|
$$
$$
= \left\| \begin{pmatrix} Y & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} -A^* & (\gamma_1^2 I - A^* A)^{\frac{1}{2}} \\ C = (\gamma_1^2 I - AA^*)^{\frac{1}{2}} & A \end{pmatrix} \begin{pmatrix} Z & 0 \\ 0 & I \end{pmatrix} \right\|
$$

4. Show that

$$\left\| \begin{pmatrix} \tilde{X} & B \\ C & A \end{pmatrix} \right\| \le \gamma_1.$$

   This implies that $\gamma_0 \le \gamma_1$ which proves the assertion.

Exercise 5.8 Prove or disprove (through a counter example) the following singular values inequalities.

1. $\sigma_{min}(A + B) \le \sigma_{min}(A) + \sigma_{min}(B)$ for any A and B.
2. $\sigma_{min}(A + E) \le \sigma_{max}(E)$ whenever A does not have column rank, and E is any matrix.

3. If $\sigma_{max}(A) < 1$, then
$$\sigma_{max}(I - A)^{-1} \leq \frac{1}{1 - \sigma_{max}(A)}$$

4. $\sigma_i(I + A) \leq \sigma_i(A) + 1$.

6.241J / 16.338J Dynamic Systems and Control

Spring 2011