

Chapter 8

Sensitivity and Condition of the $A\vec{X} = \vec{b}$ Problem

Fortunately, for many practical problems that lead to an equation of the form $A\vec{X} = \vec{b}$, the solution can be calculated easily and accurately. However, there are important physical problems that are described by linear algebraic equations that are difficult to solve accurately. For these physical problems, small errors in the A matrix or \vec{b} input vector can produce large errors in the solution \vec{X} . That is, the problem is very *sensitive* to errors in its description. This chapter discusses this problem, and bounds are obtained on the relative error in the solution.

Sensitivity of Linear Algebraic Systems

In this section we consider the *sensitivity* of the system $A\vec{X} = \vec{b}$. We study how small changes (*i.e.*, errors) in matrix A and/or input data \vec{b} influence the solution (output) \vec{X} . The terms *sensitivity* and *condition* are associated with studies of this nature. In our analysis, we require that A be an $n \times n$ non-singular matrix and \vec{b} an $n \times 1$ vector.

We can utilize SVD to gain some insight into the sensitivity problem. Let's consider the case where A is "almost rank deficient" so that σ_n is "small" (recall that σ_1 and σ_n are the largest and smallest, respectively, singular values of $n \times n$ non-singular A). Then, small changes in \vec{b} and/or A *can* cause large changes in the solution \vec{X} . To see this, we use the SVD expansion of A and expand the solution $\vec{X} = A^{-1}\vec{b}$ in the series

$$\begin{aligned} \vec{X} &= A^{-1} \vec{b} = (\mathcal{U}\Sigma\mathcal{V}^*)^{-1}\vec{b} = \mathcal{V}\Sigma^{-1}\mathcal{U}^* \vec{b} \\ &= \sum_{i=1}^n \left(\frac{\vec{u}_i^* \vec{b}}{\sigma_i} \right) \vec{v}_i \end{aligned} \quad , \quad (8-1)$$

where \vec{u}_k and \vec{v}_k , $1 \leq k \leq n$, are the n columns of $n \times n$ unitary \mathcal{U} and \mathcal{V} , respectively (see (7-7)). A casual inspection of (8-1) reveals that, if σ_n is "small" so that A is almost rank deficient, "small" changes in A and/or \vec{b} can produce large changes in solution \vec{X} (since "small" changes in A could

lead to large changes in $1/\sigma_n$).

To proceed further, we need to establish a bit of notation. Suppose, for example, that the system matrix and input result from imprecise measurements. Let A and \vec{b} denote the correct values, and A_Δ and \vec{b}_Δ denote the errors. That is, $A + A_\Delta$ and $\vec{b} + \vec{b}_\Delta$ are measured quantities that contain errors A_Δ and \vec{b}_Δ (also, A_Δ and \vec{b}_Δ can be thought of as *perturbations* of A and \vec{b} , respectively). The *relative errors* are quantified by

$$\rho_A = \frac{\|A_\Delta\|_2}{\|A\|_2} \quad (8-2)$$

$$\rho_b = \frac{\|\vec{b}_\Delta\|_2}{\|\vec{b}\|_2}. \quad (8-3)$$

In terms of the above notation, the general system becomes

$$(A + A_\Delta)(\vec{X} + \vec{X}_\Delta) = \vec{b} + \vec{b}_\Delta. \quad (8-4)$$

That is, the errors A_Δ and \vec{b}_Δ produce an error \vec{X}_Δ in the output. A measure of the relative output error is

$$\rho_x = \frac{\|\vec{X}_\Delta\|_2}{\|\vec{X}\|_2}, \quad (8-5)$$

where \vec{X} satisfies $A\vec{X} = \vec{b}$ (*i.e.*, \vec{X} is the solution of the unperturbed system).

In what follows, we develop a bound on ρ_x , this bound depends on ρ_A and ρ_b . We would like to be able to claim that $\rho_x \leq \kappa(A)(\rho_A + \rho_b)$, where $\kappa(A)$ is the A -matrix dependent *condition number* defined below. This is our "dream bound"; the bound that we would like to have but

cannot have - **because it is not true!** But, as is shown below, we can develop a bound that comes "close" to what we want!

A simple example illustrates the need to consider sensitivity issues. Consider the system $A\vec{X} = \vec{b}$ with

$$A = \begin{bmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{bmatrix}, \vec{b} = \begin{bmatrix} 32 \\ 23 \\ 33 \\ 31 \end{bmatrix}. \quad (8-6)$$

The unique solution is $\vec{X} = [1 \ 1 \ 1 \ 1]^T$. Now, assume that $A_\Delta = 0$ (and $\rho_A = 0$: there is no error in the measurement of matrix A) and $\vec{b}_\Delta = [.1 \ .9 \ .1 \ .9]^T$ so that $\rho_b = .0213$. The solution to

$$A(\vec{X} + \vec{X}_\Delta) = \vec{b} + \vec{b}_\Delta \quad (8-7)$$

yields an output error of $\vec{X}_\Delta = [8.2 \ -13.6 \ 3.5 \ -2.1]^T$ and $\rho_x = 8.1985$. The system "amplified" the input error by $\rho_x/\rho_b = 384.9$; obviously, this system is very sensitive to input errors.

Let's continue the previous example described by (8-6). This time, assume that we measure \vec{b} exactly so that $\vec{b}_\Delta = \vec{0}$. However, we make errors in measuring A; suppose that the error in A is

$$A_\Delta = \begin{bmatrix} 0 & 0 & .1 & .2 \\ .08 & .04 & 0 & 0 \\ 0 & -.02 & -.11 & 0 \\ -.01 & -.01 & 0 & -.02 \end{bmatrix}. \quad (8-8)$$

$$\rho_A = \frac{\|A_\Delta\|_2}{\|A\|_2} = .0076209$$

The solution of the imprecise system

$$(A + A_\Delta)(\vec{X} + \vec{X}_\Delta) = \vec{b} \quad (8-9)$$

yields a relative output error of $\vec{X}_\Delta = [-82 \ 136 \ -35 \ 21]^T$, so that $\rho_X = 81.9848$. This time, our error A_Δ was “amplified” by $\rho_X/\rho_A = 10,758$; obviously, this system is very sensitive to small errors in the A matrix.

Upper Bound on Relative Error Due to a Perturbation b_Δ Acting Alone

We want to develop an upper bound on ρ_X ; this upper bound should depend on ρ_b and ρ_A . But first, we develop a simple relationship between ρ_X and ρ_b . Suppose there are only errors in \vec{b} ; that is, suppose $A_\Delta = 0$, $\vec{b}_\Delta \neq \vec{0}$ so that (8-7) holds. Also, assume that A^{-1} exists. From (8-6), we have $A\vec{X}_\Delta = \vec{b}_\Delta$ so that

$$\vec{X}_\Delta = A^{-1}\vec{b}_\Delta. \quad (8-10)$$

Now, take the norm of both sides to obtain

$$\|\vec{X}_\Delta\|_2 = \|A^{-1}\vec{b}_\Delta\|_2 \leq \|A^{-1}\|_2 \|\vec{b}_\Delta\|_2. \quad (8-11)$$

Now, use (8-10) with $\|\vec{b}\|_2 = \|A\vec{X}\|_2 \leq \|A\|_2 \|\vec{X}\|_2$ to obtain

$$\rho_X = \frac{\|\vec{X}_\Delta\|_2}{\|\vec{X}\|_2} \leq \frac{\|A^{-1}\|_2 \|\vec{b}_\Delta\|_2}{\left(\|\vec{b}\|_2 / \|A\|_2\right)} = \|A\|_2 \|A^{-1}\|_2 \frac{\|\vec{b}_\Delta\|_2}{\|\vec{b}\|_2} = \kappa(A)\rho_b, \quad (8-12)$$

where

$$\kappa(A) \equiv \|A\|_2 \|A^{-1}\|_2 \quad (8-13)$$

is called the *condition number* of matrix A.

As shown by (8-12), the condition number $\kappa(A)$ is an upper bound for the “error amplification” ρ_x/ρ_b . Equation (8-12) is a “decent” upper bound. Naturally, there are some perturbations \vec{b}_Δ for which $\kappa(A)$ is much larger than ρ_x/ρ_b (so that the bound (8-12) appears to be “loose”). However, there are other perturbations \vec{b}_Δ for which the bound is “right on” and $\kappa(A)$ is only slightly larger than ρ_x/ρ_b .

As discussed in the paragraph following (4-24), the 2-norm of matrix A is equal to the largest singular value of A; that is, $\|A\|_2 = \sigma_1$. Also, Equation (7-23) states that the 2-norm of A^{-1} is the reciprocal of the smallest singular value; in other words, $\|A^{-1}\|_2 = 1/\sigma_n$. By combining these results with (8-13), we can write the condition number as

$$\kappa(A) \equiv \|A\|_2 \|A^{-1}\|_2 = \frac{\sigma_1}{\sigma_n}, \quad (8-14)$$

the ratio of the largest to smallest singular values.

When $\kappa(A)$ is large the system is said to be *ill-conditioned*. Generally speaking, large values for $\kappa(A)$ imply that it may be difficult to compute an accurate solution to $A\vec{X} = \vec{b}$ (and, as it turns out, compute accurately the eigenvalues of A).

Notice that the condition number κ is not directly affected by the “size” (*i.e.*, norm) of the matrix. For example, the condition numbers of $A = I$ and $A' = I/10$ are both unity. By comparison, *the determinant of $n \times n$ matrix A is an awful measure of ill-conditioning*. $\text{Det}(A)$ depends on the scaling of A (*i.e.*, $\det(\alpha A) = \alpha^n \det(A)$). Also, $\det(A)$ depends on the order n ; if $A = I/10$, then $\det(A) = 10^{-n}$. In fact, the “nearly singular” matrix $A = I/10$ is very well conditioned!

MatLab's Condition Number Function

MatLab has a built-in function for computing the condition number $\kappa(A)$ of a matrix. The

syntax for this command is

$$k = \text{cond}(A), \quad (8-15)$$

where A is a non-singular, $n \times n$ matrix. If $\kappa(A)$ is large, then matrix A is said to be *ill-conditioned*, and it may be difficult to compute an accurate solution for $A\vec{X} = \vec{b}$.

Upper Bound on Relative Error Due to a Perturbation A_Δ Acting Alone

Now, let us consider the second problem, where the $n \times n$ nonsingular matrix A is perturbed, but \vec{b} does not ($A_\Delta \neq 0$, $\vec{b}_\Delta = \vec{0}$). For this problem, Equation (8-9) applies, where \vec{X} satisfies $A\vec{X} = \vec{b}$. From (8-9) we have

$$A\vec{X}_\Delta = \vec{b} - (A + A_\Delta)\vec{X} - A_\Delta\vec{X}_\Delta = -A_\Delta(\vec{X} + \vec{X}_\Delta), \quad (8-16)$$

so that

$$\vec{X}_\Delta = -A^{-1}A_\Delta(\vec{X} + \vec{X}_\Delta). \quad (8-17)$$

Now, take the norm of (8-17) to obtain

$$\|\vec{X}_\Delta\|_2 = \|A^{-1}A_\Delta(\vec{X} + \vec{X}_\Delta)\|_2 \leq \|A^{-1}\|_2 \|A_\Delta\|_2 \|\vec{X} + \vec{X}_\Delta\|_2, \quad (8-18)$$

and this leads to

$$\frac{\|\vec{X}_\Delta\|_2}{\|\vec{X} + \vec{X}_\Delta\|_2} \leq \kappa(A) \rho_A, \quad (8-19)$$

where $\kappa(A)$ is the condition number given by (8-13) and (8-14). While not exactly the bound we would like, this result is useful (for this case, we would like a bound like (8-12), but it is not to be!!).

Sensitivity Results in Terms of a Power Series Expansion

In this section, we consider the effects of both A_Δ and \vec{b}_Δ on the solution. In a manner similar to that established in Chapter 4 (see Theorem 4.3), we use power series to obtain our results.

Again, we consider (8-4), repeated here as $(A+A_\Delta)(\vec{X}+\vec{X}_\Delta) = \vec{b} + \vec{b}_\Delta$. As before, we have $A\vec{X} = \vec{b}$ so that $\vec{X} = A^{-1}\vec{b}$ (i.e., A^{-1} is assumed to exist in what follows). Now, for the moment, assume that $(A + A_\Delta)^{-1}$ exists so that

$$\begin{aligned}\vec{X} + \vec{X}_\Delta &= (A + A_\Delta)^{-1}(\vec{b} + \vec{b}_\Delta) = \left[A(I + A^{-1}A_\Delta) \right]^{-1}(\vec{b} + \vec{b}_\Delta) \\ &= (I + A^{-1}A_\Delta)^{-1}A^{-1}(\vec{b} + \vec{b}_\Delta).\end{aligned}\tag{8-20}$$

As shown by Theorem 4.3, if we require $\|A^{-1}A_\Delta\|_2 < 1$, then $(A + A_\Delta)^{-1} = (I + A^{-1}A_\Delta)^{-1}A^{-1}$ exists, and we have

$$\begin{aligned}\vec{X} + \vec{X}_\Delta &= (I + A^{-1}A_\Delta)^{-1}A^{-1}(\vec{b} + \vec{b}_\Delta) \\ &= \left[I - A^{-1}A_\Delta + (A^{-1}A_\Delta)^2 - \dots \right] A^{-1}(\vec{b} + \vec{b}_\Delta) \quad , \\ &= \underbrace{A^{-1}\vec{b}}_{0^{\text{th}}\text{-order}} + \underbrace{A^{-1}[\vec{b}_\Delta - A_\Delta A^{-1}\vec{b}]}_{1^{\text{st}}\text{-order in } \vec{b}_\Delta, A_\Delta} + O(\vec{b}_\Delta, A_\Delta)\end{aligned}\tag{8-21}$$

where $O(\vec{b}_\Delta, A_\Delta)$ means 2nd and higher-order terms in \vec{b}_Δ, A_Δ (2nd-order terms are A_Δ^2 and $A_\Delta\vec{b}_\Delta$; 3rd-order terms are A_Δ^3 and $A_\Delta^2\vec{b}_\Delta$; and so forth). Use (8-21), and the fact that $\vec{X} = A^{-1}\vec{b}$, to write

$$\bar{X}_\Delta = A^{-1}[\bar{b}_\Delta - A_\Delta A^{-1}\bar{b}] + O(\bar{b}_\Delta, A_\Delta). \quad (8-22)$$

Finally, this last result can be used to obtain

$$\rho_x = \frac{\|\bar{X}_\Delta\|_2}{\|\bar{X}\|_2} \leq \|A^{-1}\|_2 \left(\frac{\|\bar{b}_\Delta\|_2}{\|\bar{X}\|_2} + \|A_\Delta\|_2 \right) + O(\bar{b}_\Delta, A_\Delta), \quad (8-23)$$

However, note that $\|\bar{b}\|_2 \leq \|A\|_2 \|\bar{X}\|_2$ or $\|\bar{b}\|_2 / \|A\|_2 \leq \|\bar{X}\|_2$; this result can be used in (8-23) to obtain

$$\begin{aligned} \rho_x &\leq \|A^{-1}\|_2 \left(\|A\|_2 \frac{\|\bar{b}_\Delta\|_2}{\|\bar{b}\|_2} + \|A_\Delta\|_2 \right) + O(\bar{b}_\Delta, A_\Delta) \\ &= \kappa(A) [\rho_b + \rho_A] + O(\bar{b}_\Delta, A_\Delta) \end{aligned} \quad (8-24)$$

a useful bound (depending on ρ_b and ρ_A) on the relative error ρ_x . Thus, to first order, the solution's relative error ρ_x is bounded by $\kappa(A)$ times the sum of the relative errors ρ_A and ρ_b in the matrix A and input data \bar{b} , respectively (when using (8-24), one would drop the $O(\bar{b}_\Delta, A_\Delta)$ terms). In this sense, the condition number $\kappa(A)$ quantifies the sensitivity of the $A\bar{X} = \bar{b}$ problem.

An Absolute Bound on the Relative Error

While useful, Equation (8-24) provides information only for "sufficiently small" $\|\bar{b}_\Delta\|_2$ and $\|A_\Delta\|_2$ (so that 2nd and higher-order terms can be ignored). In some applications, questions may be raised about those pesky second-and-higher-order terms that must be ignored to use (8-24). Hence, in this section, we do away with this "truncated series bound" and develop an *absolute* bound on the relative error.

Theorem 8-1

Consider the $A\bar{X} = \bar{b}$ problem where $n \times n$ matrix A is nonsingular and $n \times 1$ input vector $\bar{b} \neq$

$\bar{0}$. Now, let

$$(A + A_\Delta)(\bar{X} + \bar{X}_\Delta) = \bar{b} + \bar{b}_\Delta \quad (8-25)$$

be the perturbed equation. Let ε be an upper bound on the relative errors ρ_A and ρ_b ; that is, let the small positive number $\varepsilon > 0$ be such that

$$\rho_A = \frac{\|A_\Delta\|_2}{\|A\|_2} \leq \varepsilon \quad \text{and} \quad \rho_b = \frac{\|\bar{b}_\Delta\|_2}{\|\bar{b}\|_2} \leq \varepsilon. \quad (8-26)$$

If $\varepsilon\kappa(A) < 1$, then $A + A_\Delta$ is nonsingular, and

$$\frac{\|\bar{X} + \bar{X}_\Delta\|_2}{\|\bar{X}\|_2} \leq \frac{1 + \varepsilon\kappa(A)}{1 - \varepsilon\kappa(A)}. \quad (8-27)$$

Proof: From (8-25) we have

$$(I + A^{-1}A_\Delta)(\bar{X} + \bar{X}_\Delta) = \bar{X} + A^{-1}\bar{b}. \quad (8-28)$$

From Theorem 4-3 and

$$\|A^{-1}A_\Delta\|_2 \leq \|A^{-1}\|_2 \|A_\Delta\|_2 \leq \varepsilon \|A^{-1}\|_2 \|A\|_2 = \varepsilon\kappa(A) < 1, \quad (8-29)$$

it follows that $(I + A^{-1}A_\Delta)$ is nonsingular. Hence, we can use (8-28) write

$$\begin{aligned} \|\bar{\mathbf{X}} + \bar{\mathbf{X}}_{\Delta}\|_2 &= \left\| (\mathbf{I} + \mathbf{A}^{-1}\mathbf{A}_{\Delta})^{-1} (\bar{\mathbf{X}} + \mathbf{A}^{-1}\bar{\mathbf{b}}_{\Delta}) \right\|_2 \leq \left\| (\mathbf{I} + \mathbf{A}^{-1}\mathbf{A}_{\Delta})^{-1} \right\|_2 \left(\|\bar{\mathbf{X}}\|_2 + \|\mathbf{A}^{-1}\bar{\mathbf{b}}_{\Delta}\|_2 \right) \\ &\leq \left\| (\mathbf{I} + \mathbf{A}^{-1}\mathbf{A}_{\Delta})^{-1} \right\|_2 \left(\|\bar{\mathbf{X}}\|_2 + \epsilon \|\mathbf{A}^{-1}\|_2 \|\bar{\mathbf{b}}\|_2 \right) \end{aligned} \quad (8-30)$$

However, from Theorem 4-3 and (8-29) we have

$$\left\| (\mathbf{I} + \mathbf{A}^{-1}\mathbf{A}_{\Delta})^{-1} \right\|_2 \leq \frac{1}{1 - \|\mathbf{A}^{-1}\mathbf{A}_{\Delta}\|_2} \leq \frac{1}{1 - \epsilon\kappa(\mathbf{A})}, \quad (8-31)$$

so that (8-30) becomes

$$\begin{aligned} \|\bar{\mathbf{X}} + \bar{\mathbf{X}}_{\Delta}\|_2 &\leq \frac{1}{1 - \epsilon\kappa(\mathbf{A})} \left(\|\bar{\mathbf{X}}\|_2 + \epsilon \|\mathbf{A}^{-1}\|_2 \|\bar{\mathbf{b}}\|_2 \right) \leq \frac{1}{1 - \epsilon\kappa(\mathbf{A})} \left(\|\bar{\mathbf{X}}\|_2 + \frac{\|\bar{\mathbf{b}}\|_2}{\|\mathbf{A}\|_2} \epsilon \|\mathbf{A}^{-1}\|_2 \|\mathbf{A}\|_2 \right) \\ &\leq \frac{1}{1 - \epsilon\kappa(\mathbf{A})} \left(\|\bar{\mathbf{X}}\|_2 + \epsilon\kappa(\mathbf{A}) \frac{\|\bar{\mathbf{b}}\|_2}{\|\mathbf{A}\|_2} \right) \end{aligned} \quad (8-32)$$

Now, note that $\|\bar{\mathbf{b}}\|_2 = \|\mathbf{A}\bar{\mathbf{X}}\|_2 \leq \|\mathbf{A}\|_2 \|\bar{\mathbf{X}}\|_2$, so that (8-32) can be written as

$$\|\bar{\mathbf{X}} + \bar{\mathbf{X}}_{\Delta}\|_2 \leq \frac{1}{1 - \epsilon\kappa(\mathbf{A})} \left(\|\bar{\mathbf{X}}\|_2 + \epsilon\kappa(\mathbf{A}) \|\bar{\mathbf{X}}\|_2 \right) \quad (8-33)$$

which leads to (8-27) as claimed.♥ This result is used in the following bound on $\rho_{\mathbf{X}}$.

Theorem 8-2

Consider the same $\mathbf{A}\bar{\mathbf{X}} = \bar{\mathbf{b}}$ problem with (8-25), (8-26) and $\epsilon\kappa(\mathbf{A}) < 1$, where ϵ is a small positive number. An absolute bound on $\rho_{\mathbf{X}}$ is given by

$$\rho_{\mathbf{X}} = \frac{\|\bar{\mathbf{X}}_{\Delta}\|_2}{\|\bar{\mathbf{X}}\|_2} \leq \kappa(\mathbf{A}) \left(\rho_{\mathbf{b}} + \frac{1 + \epsilon\kappa(\mathbf{A})}{1 - \epsilon\kappa(\mathbf{A})} \rho_{\mathbf{A}} \right). \quad (8-34)$$

Proof: From $(A + A_\Delta)(\vec{X} + \vec{X}_\Delta) = \vec{b} + \vec{b}_\Delta$ we have $A\vec{X}_\Delta = -A_\Delta(\vec{X} + \vec{X}_\Delta) + \vec{b}_\Delta$. This leads to

$$\vec{X}_\Delta = -A^{-1}A_\Delta(\vec{X} + \vec{X}_\Delta) + A^{-1}\vec{b}_\Delta \quad (8-35)$$

so that

$$\|\vec{X}_\Delta\|_2 \leq \|A^{-1}\vec{b}_\Delta\|_2 + \|A^{-1}A_\Delta(\vec{X} + \vec{X}_\Delta)\|_2 \leq \frac{\kappa(A)}{\|A\|_2} \|\vec{b}_\Delta\|_2 + \kappa(A)\rho_A \|\vec{X} + \vec{X}_\Delta\|_2 \quad (8-36)$$

$$\rho_X = \frac{\|\vec{X}_\Delta\|_2}{\|\vec{X}\|_2} \leq \frac{\kappa(A)}{\|A\|_2 \|\vec{X}\|_2} \|\vec{b}_\Delta\|_2 + \kappa(A)\rho_A \frac{\|\vec{X} + \vec{X}_\Delta\|_2}{\|\vec{X}\|_2}. \quad (8-37)$$

Now, since $\|\vec{b}\|_2 \leq \|A\|_2 \|\vec{X}\|_2$ we have

$$\rho_X \leq \kappa(A)\rho_b + \kappa(A)\rho_A \frac{\|\vec{X} + \vec{X}_\Delta\|_2}{\|\vec{X}\|_2}. \quad (8-38)$$

Now, use Theorem 8.1 with (8-38) to obtain the desired result

$$\rho_X \leq \kappa(A) \left(\rho_b + \frac{1 + \varepsilon\kappa(A)}{1 - \varepsilon\kappa(A)} \rho_A \right) \quad (8-39)$$

as claimed.♥ Equation (8-39) is a handy absolute upper bound (in terms of ρ_A and ρ_b) on ρ_X .

Example

Consider the problem

$$\begin{bmatrix} 1 & 0 \\ 0 & 10^{-6} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 10^{-6} \end{bmatrix} \quad (8-40)$$

with solution $\vec{X} = [1 \ 1]^T$. First, consider the perturbation $A_\Delta = 0$, $\vec{b}_\Delta = [10^{-6} \ 0]^T$. The system $A(\vec{X} + \vec{X}_\Delta) = \vec{b} + \vec{b}_\Delta$ has solution $\vec{X} + \vec{X}_\Delta = [1+10^{-6} \ 1]^T$, so

$$\rho_x = \frac{\|\vec{X}_\Delta\|_2}{\|\vec{X}\|_2} = \frac{10^{-6}}{\sqrt{2}}. \quad (8-41)$$

However, $\kappa(A) = \|A\|_2 \|A^{-1}\|_2 \approx 10^6$, and $\rho_b \approx 10^{-6}$. Hence, the bound (8-39) yields

$$\rho_x \leq \kappa(A)\rho_b \approx 1. \quad (8-42)$$

A comparison of (8-41) and (8-42) leads to the conclusion that bound (8-39) produces a *gross* overestimate of the error. One might be led to believe that the bound is “very loose”. But this is not *always* the case. For consider a second example with the same (8-40). However, this time, set $A_\Delta = 0$, $\vec{b}_\Delta = [0 \ 10^{-6}]^T$. The solution to $A(\vec{X} + \vec{X}_\Delta) = \vec{b} + \vec{b}_\Delta$ is $\vec{X} + \vec{X}_\Delta = [1 \ 2]^T$ and

$$\rho_x = \frac{\|\vec{X}_\Delta\|_2}{\|\vec{X}\|_2} = \frac{1}{\sqrt{2}} \leq \kappa(A)\rho_b \approx 1. \quad (8-43)$$

This time, the true relative error ρ_x is close to the upper bound (8-39) on ρ_x . So, as seen by these two rather extreme examples, bound (8-39) may be “conservative”, or it might be “right-on”, depending on the particular problem and perturbations.