Conditioning and Stability

Prof Franck K. Mutombo

University of Lubumbashi

franckm@aims.ac.za | *mutombof@unilu.ac.cd*

October 21, 2025

Conditioning and Stability

content...

Conditioning and Stability

A computing problem is well-posed if

1. a solution exists (e.g., we want to rule out situations that lead to division by zero),
2. the computed solution is unique,
3. the solution depends continuously on the data, i.e., a small change in the data should result in a small change in the answer. This phenomenon is referred to as stability of the problem.

Conditioning and Stability

Example Consider the following three different recursion algorithms to compute $x_n = \left(\frac{1}{3}\right)^n$:

1. $x_0 = 1, x_n = \frac{1}{3}x_{n-1}$ for $n \geq 1$,
2. $y_0 = 1, y_1 = \frac{1}{3}, y_{n+1} = \frac{4}{3}y_n - \frac{1}{3}y_{n-1}$ for $n \geq 1$,
3. $z_0 = 1, z_1 = \frac{1}{3}, z_{n+1} = \frac{10}{3}z_n - z_{n-1}$ for $n \geq 1$.

- The validity of the latter two approaches can be proved by induction.
- Use of slightly perturbed initial values shows will that the first algorithm yields stable errors throughout.
- The second algorithm has stable errors, but unstable relative errors.
- And the third algorithm is unstable in either sense.

Consider solution of the linear system $A\boldsymbol{x} = \boldsymbol{b}$, with exact answer $\boldsymbol{x}$ and computed answer $\tilde{\boldsymbol{x}}$. Thus, we expect an error

$$e = \boldsymbol{x} - \tilde{\boldsymbol{x}}$$

Since $\boldsymbol{x}$ is not known to us in general we often judge the accuracy of the solution by looking at the residual

$$\boldsymbol{r} = \boldsymbol{b} - A\tilde{\boldsymbol{x}} = A\boldsymbol{x} - A\tilde{\boldsymbol{x}} = A\boldsymbol{e}$$

and hope that a small residual guarantees a small error.

The Condition Number of Matrix

Example

We consider $A\boldsymbol{x} = \boldsymbol{b}$ with

$$A = \begin{bmatrix} 1.01 & 0.99 \\ 0.99 & 1.01 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

and exact solution $\boldsymbol{x} = [1, 1]^T$.

The Condition Number of Matrix

1. (a) Let's assume we computed a solution of $\tilde{\boldsymbol{x}} = [1.01, 1.01]^T$. Then the error

$$\boldsymbol{e} = \boldsymbol{x} - \tilde{\boldsymbol{x}} = \begin{bmatrix} -0.01 \\ -0.01 \end{bmatrix}$$

is small, and the residual

$$\boldsymbol{r} = \boldsymbol{b} - A\boldsymbol{x} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} - \begin{bmatrix} 2.02 \\ 2.02 \end{bmatrix} = \begin{bmatrix} -0.02 \\ -0.02 \end{bmatrix}$$

is also small. Everything looks good.

The Condition Number of Matrix

2. (b) Now, let's assume that we computed a solution of $\tilde{\boldsymbol{x}} = [2, 0]^T$. This "solutions" is obviously not a good one. Its error is

$$e = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

which is quite large. However, the residual is

$$\boldsymbol{r} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} - \begin{bmatrix} 2.02 \\ 1.98 \end{bmatrix} = \begin{bmatrix} -0.02 \\ 0.02 \end{bmatrix},$$

which is still small. This is not good. This shows that the residual is not a reliable indicator of the accuracy of the solution.

3. (c) If we change the right-hand side of the problem to $\boldsymbol{b} = [2, -2]^T$ so that the exact solution becomes $\boldsymbol{x} = [100, -100]^T$, then things behave "wrong" again. Let's assume we computed a solution $\tilde{\boldsymbol{x}} = [101, -99]^T$ with a relatively small error of $\boldsymbol{e} = [-1, -1]^T$. However, the residual now is

$$\boldsymbol{r} = \begin{bmatrix} 2 \\ -2 \end{bmatrix} - \begin{bmatrix} 4 \\ 0 \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \end{bmatrix}$$

which is relatively large. So again, the residual is not an accurate indicator of the error.

What is the explanation for the phenomenon we're observing? The answer is, the matrix A is ill-conditioned.

The Condition Number of Matrix

Let's try to get a better understanding of how the error and the residual are related for the problem $A\boldsymbol{x} = \boldsymbol{b}$. We will use the notation

$$\boldsymbol{e} = \boldsymbol{x} - \tilde{\boldsymbol{x}}, \quad \boldsymbol{r} = \boldsymbol{b} - A\tilde{\boldsymbol{x}} = \boldsymbol{b} - \tilde{\boldsymbol{b}}.$$

Thus,

$$\begin{aligned}\|\boldsymbol{e}\| = \|\boldsymbol{x} - \tilde{\boldsymbol{x}}\| &= \left\|A^{-1}\boldsymbol{b} - A^{-1}\tilde{\boldsymbol{b}}\right\| = \left\|A^{-1}(\boldsymbol{b} - \tilde{\boldsymbol{b}})\right\| \\ &\leq \left\|A^{-1}\right\| \|\boldsymbol{b} - \tilde{\boldsymbol{b}}\| = \left\|A^{-1}\right\| \|\boldsymbol{r}\|\end{aligned}$$

Therefore, the absolute error satisfies

$$\|\boldsymbol{e}\| \leq \left\|A^{-1}\right\| \|\boldsymbol{r}\|.$$

The Condition Number of Matrix

Often, however, it is better to consider the relative error, i.e., $\frac{\|\boldsymbol{e}\|}{\|\boldsymbol{x}\|}$ (and $\left.\frac{\|\boldsymbol{r}\|}{\|\boldsymbol{b}\|}\right)$:

$$
\begin{aligned}
\|\boldsymbol{e}\| &\leq \left\|A^{-1}\right\| \|\boldsymbol{r}\| \underbrace{\frac{\|A\boldsymbol{x}\|}{\|\boldsymbol{b}\|}}_{=1} \\
&\leq \left\|A^{-1}\right\| \|A\| \|\boldsymbol{x}\| \frac{\|\boldsymbol{r}\|}{\|\boldsymbol{b}\|}.
\end{aligned}
$$

This yields the bound

$$
\frac{\|\boldsymbol{e}\|}{\|\boldsymbol{x}\|} \leq \left\|A^{-1}\right\| \|A\| \frac{\|\boldsymbol{r}\|}{\|\boldsymbol{b}\|} = \kappa(A) \frac{\|\boldsymbol{r}\|}{\|\boldsymbol{b}\|},
$$

where $\kappa(A) = \left\|A^{-1}\right\| \|A\| \qquad (1)$ is called the condition number of A.

The Condition Number of Matrix

Remark

1. The condition number depends on the type of norm used.
2. For the 2-norm of a nonsingular $m \times m$ matrix A we know $\|A\|_2 = \sigma_1$ (the largest singular values of A), and $\left\|A^{-1}\right\|_2 = \frac{1}{\sigma_m}$.
3. If A is singular then $\kappa(A) = \infty$.

Also note that $\kappa(A) = \frac{\sigma_1}{\sigma_m} \geq 1$. In fact, this holds for any norm.

How should we interpret the bound (1)? If $\kappa(A)$ is large (i.e., the matrix is illconditioned), then relatively small perturbations of the right-hand side $\boldsymbol{b}$ (and therefore the residual) may lead to large errors; an instability.

The Condition Number of Matrix

For well-conditioned problems (i.e., $\kappa(A) \approx 1$) we can also get a useful bound telling us what sort of relative error $\frac{\|\boldsymbol{x}-\tilde{\boldsymbol{x}}\|}{\|\boldsymbol{x}\|}$ we should at least expect. Consider

$$\begin{aligned}
\|\boldsymbol{r}\|\|\boldsymbol{x}\| &= \|\boldsymbol{b}-\tilde{\boldsymbol{b}}\|\|\boldsymbol{x}\| \\
&= \|A\boldsymbol{x}-A\tilde{\boldsymbol{x}}\|\|\boldsymbol{x}\| = \|A(\boldsymbol{x}-\tilde{\boldsymbol{x}})\|\|\boldsymbol{x}\| \\
&= \|A\boldsymbol{e}\|\|\boldsymbol{x}\| \\
&= \|A\boldsymbol{e}\|\left\|A^{-1}\boldsymbol{b}\right\| \leq \|A\|\|\boldsymbol{e}\|\left\|A^{-1}\right\|\|\boldsymbol{b}\|,
\end{aligned}$$

so that

$$\frac{1}{\kappa(A)}\frac{\|\boldsymbol{r}\|}{\|\boldsymbol{b}\|} \leq \frac{\|\boldsymbol{e}\|}{\|\boldsymbol{x}\|} \qquad (2).$$

Of course, we can combine (1) and (2) to obtain

$$\frac{1}{\kappa(A)}\frac{\|\boldsymbol{r}\|}{\|\boldsymbol{b}\|} \leq \frac{\|\boldsymbol{x}-\tilde{\boldsymbol{x}}\|}{\|\boldsymbol{x}\|} \leq \kappa(A)\frac{\|\boldsymbol{r}\|}{\|\boldsymbol{b}\|} \qquad (3).$$

so that

$$\frac{1}{\kappa(A)}\frac{\|\boldsymbol{r}\|}{\|\boldsymbol{b}\|} \leq \frac{\|\boldsymbol{e}\|}{\|\boldsymbol{x}\|} \qquad (2).$$

Of course, we can combine (1) and (2) to obtain

$$\frac{1}{\kappa(A)}\frac{\|\boldsymbol{r}\|}{\|\boldsymbol{b}\|} \leq \frac{\|\boldsymbol{x}-\tilde{\boldsymbol{x}}\|}{\|\boldsymbol{x}\|} \leq \kappa(A)\frac{\|\boldsymbol{r}\|}{\|\boldsymbol{b}\|} \qquad (3).$$

These bounds are true for any A, but show that the residual is a good indicator of the error only if A is well-conditioned.

Exercises

1. Find the condition number of the matrix

$$A = \begin{bmatrix} 1000 & 999 \\ 999 & 998 \end{bmatrix}$$

2. Let H_n be the $n \times n$ Hilbert matrix, defined by $h_{ij} = 1/(i+j-1)$. Use Julia the find the condition number of H_4. Discuss.

Example: The SVD of the matrix A reveals

$$A = \begin{bmatrix} 1.01 & 0.99 \\ 0.99 & 1.01 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 0.02 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix},$$

which implies

$$\kappa(A) = \frac{\sigma_1}{\sigma_2} = \frac{2}{0.02} = 100.$$

For a 2×2 matrix this is an indication that A is fairly ill-conditioned. We see that the bounds (25) allow for large variations:

$$\frac{1}{100}\frac{\|\boldsymbol{r}\|}{\|\boldsymbol{b}\|} \leq \frac{\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|}{\|\boldsymbol{x}\|} \leq 100\frac{\|\boldsymbol{r}\|}{\|\boldsymbol{b}\|}.$$

Thus the relative residual is not a good error indicator (as we saw in our initial calculations).

The Effect of Changes in A on the Relative Error

We again consider the linear system $A\boldsymbol{x} = \boldsymbol{b}$. But now A may be perturbed to $\widetilde{A} = A + \delta A$. We denote by $\boldsymbol{x}$ the exact solution of $A\boldsymbol{x} = \boldsymbol{b}$, and by $\tilde{\boldsymbol{x}}$ the exact solution of $\widetilde{A}\tilde{\boldsymbol{x}} = \boldsymbol{b}$, i.e., $\tilde{\boldsymbol{x}} = \boldsymbol{x} + \boldsymbol{\delta x}$.
This implies

$$\begin{aligned} \tilde{A}\tilde{\boldsymbol{x}} = \boldsymbol{b} \quad &\iff \quad (A + \boldsymbol{\delta} A)(\boldsymbol{x} + \boldsymbol{\delta x}) = \boldsymbol{b} \\ \iff \underbrace{A\boldsymbol{x} - \boldsymbol{b}}_{=0} &+ (\boldsymbol{\delta} A)\boldsymbol{x} + A(\boldsymbol{\delta x}) + (\boldsymbol{\delta} A)(\boldsymbol{\delta x}) = 0. \end{aligned}$$

If we neglect the term with the product of the deltas then we get

$$(\boldsymbol{\delta} A)\boldsymbol{x} + A(\boldsymbol{\delta x}) = 0 \quad \text{ or } (\boldsymbol{\delta x}) = -A^{-1}(\boldsymbol{\delta} A)\boldsymbol{x}.$$

The Effect of Changes in A on the Relative Error

Taking norms this yields

$$\|\delta \boldsymbol{x}\| \leq \left\|A^{-1}\right\| \|\delta A\| \|\boldsymbol{x}\| \quad \Longleftrightarrow \quad \|\delta \boldsymbol{x}\| \leq \left\|A^{-1}\right\| \|A\| \frac{\|\delta A\|}{\|A\|} \|\boldsymbol{x}\|$$

or

$$\frac{\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|}{\|\boldsymbol{x}\|} \leq \kappa(A) \frac{\|A - \widetilde{A}\|}{\|A\|} \qquad (4).$$

We can interpret (4) as follows: For ill-conditioned matrices a small perturbation of the entries can lead to large changes in the solution of the linear system. This is also evidence of an instability.

The Effect of Changes in A on the Relative Error

Example: We consider

$$A = \begin{bmatrix} 1.01 & 0.99 \\ 0.99 & 1.01 \end{bmatrix} \quad \text{with} \quad \boldsymbol{\delta} A = \begin{bmatrix} -0.01 & 0.01 \\ 0.01 & -0.01 \end{bmatrix}.$$

Now

$$\widetilde{A} = A + \boldsymbol{\delta} A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

which is even singular, so that $\widetilde{A}\tilde{\boldsymbol{x}} = \boldsymbol{b}$ with $\boldsymbol{b} = [2, -2]^T$ has no solution at all.

Remark: For matrices with condition number $\kappa(A)$ one can expect to lose $\log_{10} \kappa(A)$ digits when solving $A\boldsymbol{x} = \boldsymbol{b}$.

Backward Stability

In light of the estimate (4) we say that an algorithm for solving $A\boldsymbol{x} = \boldsymbol{b}$ is backward stable if

$$\frac{\|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|}{\|\boldsymbol{x}\|} = \mathcal{O}\left(\kappa(A)\varepsilon_{\text{machine}}\right),$$

i.e., if the significance of the error produced by the algorithm is due only to the conditioning of the matrix.

Remark

We can view a backward stable algorithm as one which delivers the "right answer to a perturbed problem", namely $\widetilde{A}\tilde{\boldsymbol{x}} = \boldsymbol{b}$, with perturbation of the order $\frac{\|A - \widetilde{A}\|}{\|A\|} = \mathcal{O}\left(\varepsilon_{\text{machine}}\right)$