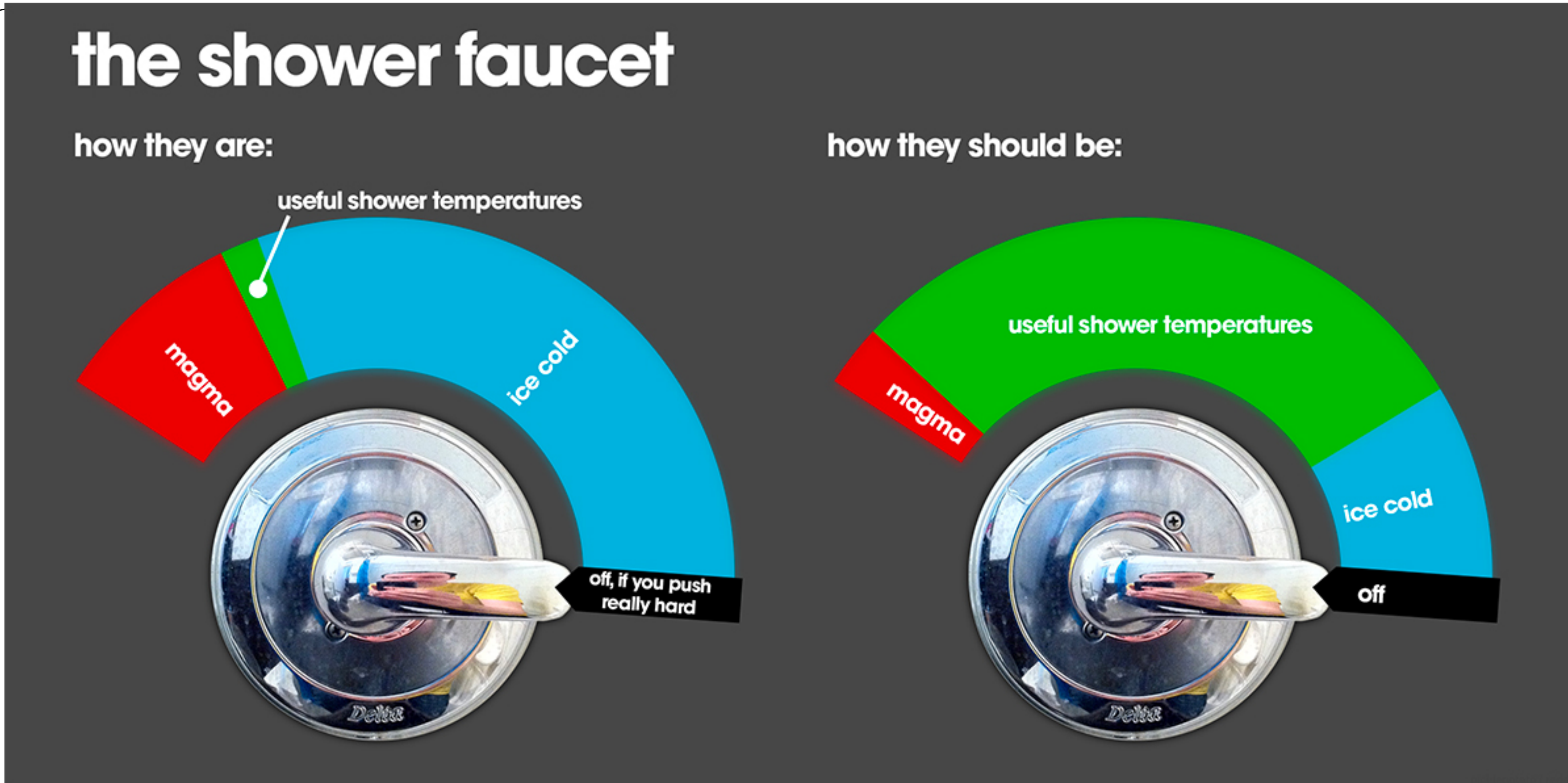
Linear System of Equations - Conditioning


the shower faucet
how they are:
useful shower temperatures
magma
ice cold
off, if you push really hard
how they should be:
useful shower temperatures
magma
ice cold
off
Delta
Delta


WHAT IT LOOKS LIKE
WHAT IT FEELS LIKE

Numerical experiments

$Ax = b$

$x_{true} = [1, 1, \dots, 1]$

$b = A x_{true}$

$x = \mathrm{solve}(A, b)$

Input has uncertainties:

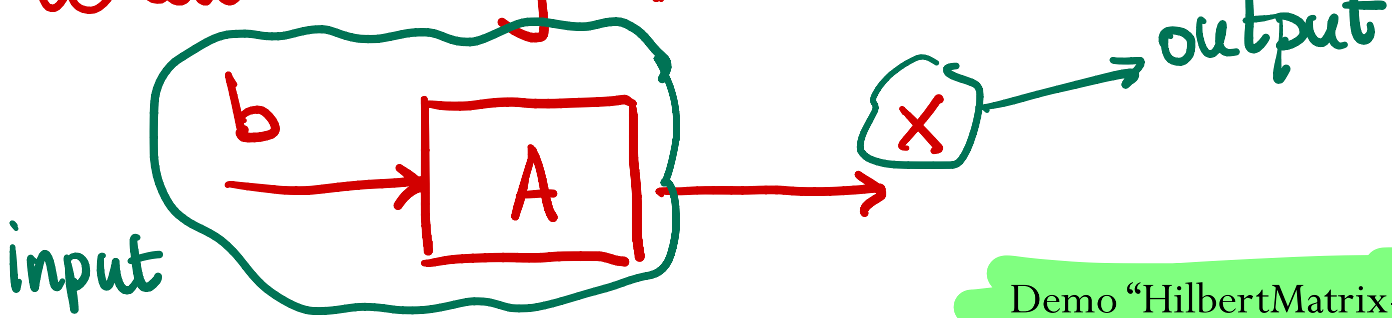
- Errors due to representation with finite precision
- Error in the sampling

$x \stackrel{?}{=} x_{true}$

Once you select your numerical method , how much error should you expect to see in your **output?**

Is your method sensitive to errors (perturbation) in the input?

When solving $Ax = b$



Demo “HilbertMatrix-ConditionNumber”

Sensitivity of Solutions of Linear Systems

Suppose we start with a non-singular system of linear equations $\boldsymbol{A}\,\boldsymbol{x} = \boldsymbol{b}$.

We change the right-hand side vector $\boldsymbol{b}$ (input) by a small amount $\Delta\boldsymbol{b}$.

How much the solution $\boldsymbol{x}$ (output) changes, i.e., how large is $\Delta\boldsymbol{x}$?

$$\frac{\text{Output Relative error}}{\text{Input Relative error}} = \frac{\|\Delta\boldsymbol{x}\|/\|\boldsymbol{x}\|}{\|\Delta\boldsymbol{b}\|/\|\boldsymbol{b}\|} = \frac{\|\Delta\boldsymbol{x}\|\,\|\boldsymbol{b}\|}{\|\Delta\boldsymbol{b}\|\,\|\boldsymbol{x}\|}$$

$Ax = b \rightarrow$ exact

$A\hat{x} = \hat{b} \rightarrow$ perturbed

$\hat{x} = x + \Delta x \qquad \hat{b} = b + \Delta b$

$A(x + \Delta x) = b + \Delta b \Rightarrow A\,\Delta x = \Delta b$

Sensitivity of Solutions of Linear Systems

$* \; Ax = b$

$\|Ax\| = \|b\|$

$$\frac{\text{Output Relative error}}{\text{Input Relative error}} = \frac{\|\Delta \boldsymbol{x}\|/\|\boldsymbol{x}\|}{\|\Delta \boldsymbol{b}\|/\|\boldsymbol{b}\|} = \frac{\|\Delta \boldsymbol{x}\|\,\|\boldsymbol{b}\|}{\|\Delta \boldsymbol{b}\|\,\|\boldsymbol{x}\|}$$

$A\,\Delta x = \Delta b$

$\Delta x = A^{-1}\Delta b$

$$\frac{\text{Output Relative error}}{\text{Input Relative error}} = \frac{\|A^{-1}\Delta b\|\,\|b\|}{\|\Delta b\|\,\|x\|} \leq \frac{\|A^{-1}\|\,\cancel{\|\Delta b\|}\,\|b\|}{\cancel{\|\Delta b\|}\,\|x\|}$$

Δ ineq. $\|A^{-1}\Delta b\| \leq \|A^{-1}\|\,\|\Delta b\|$

$$\leq \|A^{-1}\|\frac{\|b\|}{\|x\|} \overset{(*)}{=} \frac{\|A^{-1}\|\,\|Ax\|}{\|x\|}$$

$\to \Delta$ ineq.

$\|Ax\| \leq \|A\|\,\|x\|$

$$\leq \frac{\|A^{-1}\|\,\|A\|\,\|x\|}{\|x\|}$$

$$\frac{\|\Delta x\|/\|x\|}{\|\Delta b\|/\|b\|} \leq \|A^{-1}\|\,\|A\|$$

Sensitivity of Solutions of Linear Systems

We can also add a perturbation to the matrix $\boldsymbol{A}$ (input) by a small amount $\boldsymbol{E}$, such that

$$(\boldsymbol{A}+\boldsymbol{E})\,\widehat{\boldsymbol{x}}=\boldsymbol{b}$$

and in a similar way obtain:

$$\frac{\|\Delta\boldsymbol{x}\|}{\|\boldsymbol{x}\|}\leq\|\boldsymbol{A}^{-1}\|\,\|\boldsymbol{A}\|\,\frac{\|\boldsymbol{E}\|}{\|\boldsymbol{A}\|}$$

Condition number

The condition number is a measure of sensitivity of solving a linear system of equations to variations in the input.

The condition number of a matrix $\boldsymbol{A}$:

$$cond(\boldsymbol{A}) = \|\boldsymbol{A}^{-1}\| \, \|\boldsymbol{A}\|$$

if we don't include the index, it means we are using p=2 norm.

Recall that the induced matrix norm is given by

$$\|\boldsymbol{A}\| = \max_{\|\boldsymbol{x}\|=1} \|\boldsymbol{A}\boldsymbol{x}\|$$

And since the condition number is relative to a given norm, we should be precise and for example write:

$$cond_2(\boldsymbol{A}) \text{ or } cond_\infty(\boldsymbol{A})$$

Demo "HilbertMatrix-ConditionNumber"

Condition Number.

How "small" can it be?

$$\text{cond}(A) = \|A\| \|A^{-1}\|$$

1) $\|X\| > 0$ if $X \neq$ zero matrix $\longrightarrow \text{cond}(A) > 0$

2) $\|XY\| \leq \|X\| \|Y\| \longrightarrow \|AA^{-1}\| \leq \|A\| \|A^{-1}\|$

$\Rightarrow \|A\| \|A^{-1}\| \geq \|I\|$

$$\|I\| = \max_{\|x\|=1} \|Ix\| = 1$$

$$\boxed{\|A\| \|A^{-1}\| \geq 1}$$

Condition number

$$\frac{\|\Delta \boldsymbol{x}\|}{\|\boldsymbol{x}\|} \leq cond(\boldsymbol{A}) \frac{\|\Delta \boldsymbol{b}\|}{\|\boldsymbol{b}\|}$$

Small condition numbers mean not a lot of error amplification. Small condition numbers are good!

Recall that

$$\|\boldsymbol{I}\| = \max_{\|\boldsymbol{x}\|=1} \|\boldsymbol{I}\,\boldsymbol{x}\| = 1$$

Which provides with a lower bound for the condition number:

$$cond(\boldsymbol{A}) = \|\boldsymbol{A}^{-1}\|\,\|\boldsymbol{A}\|$$

If $\boldsymbol{A}^{-1}$ does not exist, then $cond(\boldsymbol{A}) = \infty$ (by convention)

Demo "HilbertMatrix-ConditionNumber"

Recall Induced Matrix Norms

$$\|\boldsymbol{A}\|_1 = \max_j \sum_{i=1}^{n} |A_{ij}|$$

Maximum absolute column sum of the matrix $\boldsymbol{A}$

$$\|\boldsymbol{A}\|_\infty = \max_i \sum_{j=1}^{n} |A_{ij}|$$

Maximum absolute row sum of the matrix $\boldsymbol{A}$

$$\|\boldsymbol{A}\|_2 = \max_k \sigma_k$$

σ_k are the singular value of the matrix $\boldsymbol{A}$

Condition Number of a Diagonal Matrix

What is the 2-norm-based condition number of the diagonal matrix

$$A = \begin{bmatrix} 100 & 0 & 0 \\ 0 & 13 & 0 \\ 0 & 0 & 0.5 \end{bmatrix}?$$

$$A^{-1} = \begin{bmatrix} 1/100 & & \\ & 1/13 & \\ & & 1/0.5 \end{bmatrix}$$

$$\operatorname{cond}(A) = \|A\| \, \|A^{-1}\| \qquad \|\cdot\|_2 = \max_i \sigma_i$$

$$\|A\| = 100$$

$$\|A^{-1}\| = 2$$

$$\operatorname{cond}_2(A) = 100 \times 2 = 200$$

Condition Number of Orthogonal Matrices

What is the 2-norm condition number of an orthogonal matrix A?

$$cond(\boldsymbol{A}) = \|\boldsymbol{A}^{-1}\|_2 \; \|\boldsymbol{A}\|_2 = \left\|\boldsymbol{A}^{\boldsymbol{T}}\right\|_2 \|\boldsymbol{A}\|_2 = 1$$

That means orthogonal matrices have optimal conditioning.

They are very well-behaved in computation.

About condition numbers

1. For any matrix $\boldsymbol{A}$, $cond(\boldsymbol{A}) \geq 1$

2. For the identity matrix $\boldsymbol{I}$, $cond(\boldsymbol{I}) = 1$

3. For any matrix $\boldsymbol{A}$ and a nonzero scalar γ, $cond(\gamma\boldsymbol{A}) = cond(\boldsymbol{A})$

4. For any diagonal matrix $\boldsymbol{D}$, $cond(\boldsymbol{D}) = \frac{\boldsymbol{max}|d_{\boldsymbol{i}}|}{\boldsymbol{min}|d_{\boldsymbol{i}}|}$

5. The condition number is a measure of how close a matrix is to being singular: a matrix with large condition number is nearly singular, whereas a matrix with a condition number close to 1 is far from being singular

6. The determinant of a matrix is NOT a good indicator is a matrix is near singularity

Residual versus error

Our goal is to find the solution $\boldsymbol{x}$ to the linear system of equations $\boldsymbol{A}\,\boldsymbol{x} = \boldsymbol{b}$

Let us recall the solution of the perturbed problem

$$\widehat{\boldsymbol{x}} = (\boldsymbol{x} + \Delta\boldsymbol{x})$$

which could be the solution of

$$\boldsymbol{A}\,\widehat{\boldsymbol{x}} = (\boldsymbol{b} + \Delta\boldsymbol{b}), \qquad (\boldsymbol{A} + \boldsymbol{E})\widehat{\boldsymbol{x}} = \boldsymbol{b}, \qquad (\boldsymbol{A} + \boldsymbol{E})\,\widehat{\boldsymbol{x}} = (\boldsymbol{b} + \Delta\boldsymbol{b})$$

And the **error vector** as

$$\boldsymbol{e} = \Delta\boldsymbol{x} = \widehat{\boldsymbol{x}} - \boldsymbol{x}$$

We can write the **residual vector** as

$$\boldsymbol{r} = \boldsymbol{b} - \boldsymbol{A}\,\widehat{\boldsymbol{x}}$$

demo

Relative residual: $\frac{\|\boldsymbol{r}\|}{\|\boldsymbol{A}\|\|\boldsymbol{x}\|}$

(How well the solution satisfies the problem)

Relative error: $\frac{\|\Delta\boldsymbol{x}\ \|}{\|\boldsymbol{x}\|}$

(How close the approximated solution is from the exact one)

Residual versus error

It is possible to show that the residual satisfy the following inequality:

$$\frac{||\boldsymbol{r}||}{||\boldsymbol{A}||\,||\widehat{\boldsymbol{x}}||} \leq c\ \epsilon_m$$

Where c is “large” constant when LU/Gaussian elimination is performed without pivoting and “small” with partial pivoting.

Therefore, Gaussian elimination with partial pivoting yields **small relative residual regardless of conditioning of the system**.

When solving a system of linear equations via LU with partial pivoting, the relative residual is guaranteed to be small!

Residual versus error

Let us first obtain the norm of the error:

$$\|\Delta x\| = \|\hat{x} - x\| = \|A^{-1}A\hat{x} - A^{-1}b\| = \|A^{-1}(A\hat{x} - b)\|$$

$$= \|A^{-1}r\|$$

$$\frac{\|\Delta x\|}{\|x\|} = \frac{\|A^{-1}r\|}{\|x\|} \leq \frac{\|A^{-1}\|\,\|r\|}{\|x\|} = \frac{\|A^{-1}\|\|A\|\,\|r\|}{\|A\|\,\|x\|}$$

$$\boxed{\frac{\|\Delta x\|}{\|x\|} \leq \mathrm{cond}(A)\,\frac{\|r\|}{\|A\|\,\|x\|}}$$

well conditioned matrices, if $\|r\|$ is small, then $\|\Delta x\|$ is also small!

Rule of thumb for conditioning

Suppose we want to find the solution $\boldsymbol{x}$ to the linear system of equations $\boldsymbol{A}\,\boldsymbol{x} = \boldsymbol{b}$ using LU factorization with partial pivoting and backward/forward substitutions.

$\hookrightarrow$ guarantee residual is small

Suppose we compute the solution $\hat{\boldsymbol{x}}$.

$e_r \leqslant 10^{-S}$

If the entries in $\boldsymbol{A}$ and $\boldsymbol{b}$ are accurate to S decimal digits,

$e_r \leqslant 10^{W} 10^{-S} = 10^{-(S-W)}$

and $cond(\boldsymbol{A}) = \mathbf{10}^{W}$,

then the elements of the solution vector $\hat{\boldsymbol{x}}$ will be accurate to about

$S - W$

decimal digits

$b \longrightarrow \boxed{A} \longrightarrow \hat{x}$

(solution $\hat{x}$ will lose W decimal digits of accuracy from input)