Condition number of solving linear equations

Let A be a fixed square invertible matrix. What is the variation in the output of

$$f(A, \mathbf{y}) = (\text{the solution of } A\mathbf{x} = \mathbf{y}) = A^{-1}\mathbf{y}$$

with respect to its input $\mathbf{y}$?

Consider two systems $A\mathbf{x} = \mathbf{y}$ and $A\tilde{\mathbf{x}} = \tilde{y}$ with $\tilde{\mathbf{y}} \neq \mathbf{y}$; let $\mathbf{x}$ and $\tilde{\mathbf{x}}$ be their solutions. Then,

- $\|\tilde{\mathbf{x}} - \mathbf{x}\| = \|A^{-1}\tilde{\mathbf{y}} - A^{-1}\mathbf{y}\| = \|A^{-1}(\tilde{\mathbf{y}} - \mathbf{y})\| \leq \|A^{-1}\|\|\tilde{\mathbf{y}} - \mathbf{y}\|$,
- $\|\mathbf{y}\| = \|A\mathbf{x}\| \leq \|A\|\|\mathbf{x}\|$.

Combining the two inequalities, one gets

$$\frac{\|\tilde{\mathbf{x}} - \mathbf{x}\|}{\|\mathbf{x}\|} \leq \frac{\|A^{-1}\|\|\tilde{\mathbf{y}} - \mathbf{y}\|}{\frac{\|\mathbf{y}\|}{\|A\|}} = \|A\|\|A\|^{-1}\frac{\|\tilde{\mathbf{y}} - \mathbf{y}\|}{\|\mathbf{y}\|}.$$

This bound holds for all $\tilde{\mathbf{y}}$, hence also in the limit $\|\tilde{\mathbf{y}} - \mathbf{y}\| \to 0$.

Condition number of a matrix

Theorem

The relative condition number of solving linear equations (with A fixed and y as input) is

$$\kappa(A) = \|A\| \|A^{-1}\|.$$

This quantity appears often; it is called 'the condition number of the matrix A'.

(Slight abuse of terminology, since we should speak of 'condition number of a problem', not 'of a matrix'.)

Condition number with respect to A

What if one changes A and keeps $\mathbf{y}$ fixed?

The relative condition number of the problem $A\mathbf{x} = \mathbf{y}$ with respect to its input A is, again, $\kappa(A) = \|A\|\|A^{-1}\|$.

Slightly different notation: A perturbed to $A + \Delta A$, $\mathbf{x}$ to $\mathbf{x} + \Delta\mathbf{x}$.

$$A\mathbf{x} = \mathbf{y}, \quad (A + \Delta A)(\mathbf{x} + \Delta\mathbf{x}) = \mathbf{y}$$

We can ignore the second-order term $\Delta A\, \Delta\mathbf{x}$, getting

$$\mathbf{y} + \Delta A\, \mathbf{x} + A\, \Delta\mathbf{x} + O(\|\Delta x\|) = \mathbf{y},$$

Rearranging,

$$\Delta\mathbf{x} = -A^{-1}\, \Delta A\, \mathbf{x}, \quad \frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} \leq \|A^{-1}\|\|A\| \frac{\|\Delta A\|}{\|A\|}.$$

Example — well-conditioned matrix

```
>> A = [2 1; 1 1];
>> y = [1;1];
>> cond(A)
ans =
   6.8541e+00
>> ytilde = y + [0;1e-6];
>> x = A \ y;
>> xtilde = A \ ytilde;
>> norm(x - xtilde) / norm(x)
ans =
   2.2361e-06
>> norm(y - ytilde) / norm(y)
ans =
   7.0711e-07
>> norm(y - ytilde) / norm(y) * cond(A)
ans =
   4.8466e-06
```

Example 2 — ill-conditioned matrix

```
>> A = [1 1; 1 1+1e-5];
>> cond(A)
ans =
   4.0000e+05
>> x = A \ y; xtilde = A \ ytilde;
>> norm(x - xtilde) / norm(x)
ans =
   1.4142e-01
>> norm(y - ytilde) / norm(y)
ans =
   7.0711e-07
```

'Ill-conditioned' $=$ large condition number (where 'large' is subjective; for instance, $\kappa(A) \approx 10^6$ usually is considered large).

Condition number and SVD

Recall: $\|A\| = \sigma_1$ (largest singular value) (with norm-2).

For a matrix $A \in \mathbb{R}^{n \times n}$, with singular values $\sigma_1 \geq \cdots \geq \sigma_n$, we have

$$\kappa(A) = \frac{\sigma_1}{\sigma_n}.$$

Indeed,

$$\|A\| = \|U \Sigma V^T\| = \|\Sigma\| = \sigma_1.$$

Moreover $A^{-1} = V \Sigma^{-1} U^T$, and $\|\Sigma^{-1}\| = \max_i \frac{1}{\sigma_i} = \frac{1}{\sigma_n}$.

Another property tells us that matrices with high condition number are those that are almost singular.

Condition number and distance to singularity

$$\frac{1}{\kappa(A)} = \min_{\tilde{A} \text{ singular}} \frac{\|A - \tilde{A}\|}{\|A\|} \quad (\text{"relative distance to singularity"})$$

Recall: the best rank-k approximation is truncated SVD.
The closest singular matrix to $A = U\Sigma V^T$ is

$$\tilde{A} = U \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_{n-1} & \\ & & & 0 \end{bmatrix} V^T.$$

$$\|\tilde{A} - A\| = \left\| U \begin{bmatrix} 0 & & & \\ & \ddots & & \\ & & 0 & \\ & & & \sigma_n \end{bmatrix} V^T \right\| = \left\| \begin{bmatrix} 0 & & & \\ & \ddots & & \\ & & 0 & \\ & & & \sigma_n \end{bmatrix} \right\| = \sigma_n.$$

Conditioning of least squares problems

Conditioning of linear least squares is a more complicated problem than the one for linear systems.
We will not give a full proof:

Theorem (Trefethen, Bau, Theorem 18.1)

Consider the linear least squares problem $\min \|A\mathbf{x} - \mathbf{y}\|$, with $A \in \mathbb{R}^{m \times n}$ with full column rank. Its relative condition number with respect to the input $\mathbf{y}$ is bounded by

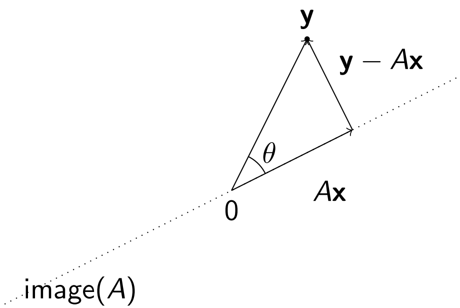
$$\kappa_{rel,\mathbf{y}\to\mathbf{x}} \leq \frac{\kappa(A)}{\cos\theta},$$

and with respect to A it is bounded by

$$\kappa_{rel,A\to\mathbf{x}} \leq \kappa(A) + \kappa(A)^2 \tan\theta,$$

where θ is the angle such that $\cos\theta = \frac{\|A\mathbf{x}\|}{\|\mathbf{y}\|}$.

The geometric picture



y 'split' into two orthogonal components: $A\mathbf{x}$ and $\mathbf{y} - A\mathbf{x}$.
QR and SVD reveal their norms: if $A = QR, Q = \begin{bmatrix} Q_0 & Q_c \end{bmatrix}$ or $A = U\Sigma V^T$, $U = \begin{bmatrix} U_0 & U_c \end{bmatrix}$ (as in their thin versions) then

$$\|A\mathbf{x}\| = \|Q_0^T\mathbf{y}\| = \|U_0^T\mathbf{y}\| = \|\mathbf{y}\|\cos\theta,$$
$$\|\mathbf{y} - A\mathbf{x}\| = \|Q_c^T\mathbf{y}\| = \|U_c^T\mathbf{y}\| = \|\mathbf{y}\|\sin\theta.$$

Some intuition

- $\theta \approx 90^\circ$: $\mathbf{y}$ almost orthogonal to $\operatorname{Im} A$: a small (relative) change in $\mathbf{y}$ causes a large (relative) change in the solution.
- $\kappa(A)$ tells us 'how well we can extract $\operatorname{Im} A$ from A': for instance,

$$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \text{ and } A_2 = \begin{bmatrix} 30\,000 & 30\,000 \\ 30\,000 & 30\,001 \\ 30\,000 & 30\,000 \end{bmatrix}$$

  have the same image, but a small (relative) perturbation to A_2 alters $\operatorname{Im} A_2$ more.
- Actually, $\kappa_2(A)$ is the relative distance to the nearest matrix $\tilde{A}$ without full column rank, generalizing the square case.
- $\theta \approx 0^\circ$ gives more well-behaved problems: the condition number is $\approx \kappa(A)$ instead of $\approx \kappa(A)^2$).

Book references: Trefethen-Bau, Lecture 18 (with more detail).