Comparison of least squares algorithms

	Normal eqns	QR	SVD
$m \approx n$	$\frac{4}{3}n^3$	$\frac{4}{3}n^3$	$13n^3$
$m \gg n$	mn^2	$2mn^2$	$2mn^2$

TL;DR: Normal equation faster than QR faster than SVD.

Is that all there is to say?

```
>> A = [1 1 2; 1 2 3; 3 1 4; 1 2 3+1e-8];
>> b = A*[3;4;5];
>> [Q1, R1] = qr(A, 0); R1 \ (Q1'*b)
ans =
   2.999999449262360
   3.999999449262359
   5.000000550737639
>> [U, S, V] = svd(A, 0); V*pinv(S)*U'*b
ans =
   2.999999450230202
   3.999999718451101
   5.000000549769798
>> (A'*A) \ (A'*b)
Warning: Matrix is close to singular or badly scaled. Resul
2.619349e-18.
ans =
 -33.566262533223330
 -32.566262676876534
  41.566262572401477
```

Sensitivity issues

```
>> A = [1 1 2; 1 2 3; 3 1 4; 1 2 3+1e-8];
```

A is at distance 10^{-8} from a non-full-rank matrix:

```
>> svd(A)
ans =
   7.553509024056715
   1.715954977117343
   0.000000004225771
```

This will be a common trend: problems close to unsolvable are numerically troublesome.

Big question

Why are normal equations much less accurate than QR/SVD?

To understand more, we need to study sensitivity and stability.

Sensitivity of a problem

Basic question: how does the output of a problem change when we change its input.

Example: compute $f(x, y) = x + 2y$. If I change x to $\tilde{x} = x + \delta$, the output becomes $f(\tilde{x}, y) = x + \delta + 2y$.

Change in input: $|\tilde{x} - x| = |\delta|$.
Change in output: $|x + \delta + y - (x + y)| = |\delta|$.

Example: compute $f(x) = x^2$. If I change x to $\tilde{x} = x + \delta$, the output becomes $(x + \delta)^2 = x^2 + 2\delta x + \delta^2$.

Change in input: $|\tilde{x} - x| = |\delta|$
Change in output: $|x^2 + 2\delta x + \delta^2| = |2\delta x + \delta^2|$.

Definition The (absolute) condition number of a function f is the maximum possible output change / input change ratio for infinitely small change of the inputs.

$$\kappa_{abs}(f, x) = \lim_{\delta \to 0} \sup_{|\tilde{x} - x| \leq \delta} \frac{|f(\tilde{x}) - f(x)|}{|\tilde{x} - x|}.$$

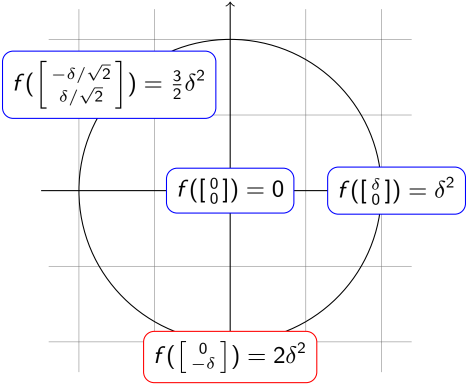
In this case,

$$\lim_{|\delta| \to 0} \frac{|2\delta x + \delta^2|}{|\delta|} = 2|x|.$$

Close relative of the derivative — in this case it's $\left|\frac{\partial f}{\partial x}\right|$.

Multivariate functions

For multivariate functions, we may need vector and matrix norms to measure input / output.
That 'sup' means that we take the worst case over all directions:
Example $f(\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]) = x^2 + 2y^2$ around $(x, y) = (0, 0)$.



Maximum change in output (for a change in input of norm δ, over all directions): $2\delta^2$.

Example

In that case, $f([\begin{smallmatrix} x \\ y \end{smallmatrix}]) = \begin{bmatrix} x \\ y \end{bmatrix}^T \begin{bmatrix} 1 & \\ & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$, and for all vectors z of norm $\|z\| \leq \delta$ we have

$$\|z^T \begin{bmatrix} 1 & \\ & 2 \end{bmatrix} z\| \leq \|z^T\| \left\| \begin{bmatrix} 1 & \\ & 2 \end{bmatrix} \right\| \|z\| \leq 2\delta^2$$

(equality is attained for instance for $z = [\begin{smallmatrix} 0 \\ -\delta \end{smallmatrix}]$).
So the absolute condition number of $f(z) = z^T [\begin{smallmatrix} 1 & \\ & 2 \end{smallmatrix}] z$ at the point $[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}]$ is

$$\kappa_{abs}(f, 0) = \lim_{\delta \to 0} \sup_{\|z\| \leq \delta} \frac{|f(0 + z) - f(0)|}{\|z\|} \leq \lim_{\delta \to 0} \frac{2\delta^2}{\delta} = 0.$$

Again, the absolute condition number is related to the derivative: for the norm-2, $\kappa_{abs}(f, x) = \|\nabla f_x\|$.

Relative condition number

Example Sensitivity of $f(x) = x^2$ around $x = 1000$: perturbing the input to $\tilde{x} = 1000.01$ changes the output from $f(x) = 1\,000\,000$ to $f(\tilde{x}) = 1\,000\,020.000\,1$.

Change in input: 0.01.
Change in output: 20.000 1.

Does not fit our intuition that $f(x)$ and $f(\tilde{x})$ are close. Better to measure input/output changes as relative changes:

Relative change in input: $\frac{\|\tilde{x}-x\|}{\|x\|} = 10^{-5}$.
Relative change in output: $\frac{\|f(\tilde{x})-f(x)\|}{\|f(x)\|} \approx 2 \times 10^{-5}$.

Definition The relative condition number of a function f is

$$\kappa_{rel}(f,x) = \lim_{\delta \to 0} \sup_{\frac{\|\tilde{x}-x\|}{\|x\|} \le \delta} \frac{\frac{\|f(\tilde{x})-f(x)\|}{\|f(x)\|}}{\frac{\|\tilde{x}-x\|}{\|x\|}} = \kappa_{abs}(f,x) \frac{\|x\|}{\|f(x)\|}.$$

Condition number of solving linear equations

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

$$f(b) = (\text{the solution of } Ax = b) = A^{-1}b$$

with respect to its input b?
For a certain $\tilde{b}$, let $\tilde{x} = A^{-1}\tilde{b}$ be the solution of $A\tilde{x} = \tilde{b}$. Then,

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

Combining the two inequalities, one gets

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

Valid for all $\tilde{b}$ — hence also in the limit $\|\tilde{b} - b\| \to 0$.

Condition number of a matrix

Theorem

The relative condition number of solving linear equations (with A fixed and b 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 b fixed? (Condition number with respect to input A)

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

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

$$Ax = b, \quad (A + \Delta A)(x + \Delta x) = b$$

Ignoring the second-order term $\Delta A \Delta x$, we get

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

i.e., rearranging,

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

Example — well-conditioned matrix

```
>> A = [2 1; 1 1];
>> b = [1;1];
>> cond(A)
ans =
   6.8541e+00
>> btilde = b + [0;1e-6];
>> x = A \ b;
>> xtilde = A \ btilde;
>> norm(x-xtilde) / norm(x)
ans =
   2.2361e-06
>> norm(b-btilde)/norm(b)
ans =
   7.0711e-07
>> norm(b-btilde)/norm(b) * 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 \ b; xtilde = A \ btilde;
>> norm(x-xtilde)/norm(x)
ans =
   1.4142e-01
>> norm(b-btilde)/norm(b)
ans =
   7.0711e-07
```

'Ill-conditioned' = large condition number (whatever that means in your context).

Condition number and SVD

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

For any matrix $A \in \mathbb{R}^{n\times n}$ (with singular values $\sigma_1 \geq \cdots \geq \sigma_n$)

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

Indeed,

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

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

Condition number and distance to singularity

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

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

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

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

Remarks

Condition number is a theoretical property — close to a derivative.

It does not depend on floating point computations, or on the choice of algorithms...

TL;DR: some computations will eat up your significant digits.

Again, take it as a warning sign: if your model calls for computing an ill-conditioned quantity, it's probably going to be useless.

Exercises

1. What is the absolute condition number of $f(x, y) = x + 2y$ with respect to its input $y \in \mathbb{R}$?
2. Show that the absolute condition number of solving a linear system (w.r.t input b) is $\|A^{-1}\|$.
3. What is the relative condition number of $f(x, y) = x - y$ (with $x, y \in \mathbb{R}$)?