Math 4336
Notes #2

Condition Number

Suppose $f: \mathbb{R}^n \to \mathbb{R}^m$ is a mathematical function (no computers involved).

The condition number depends on a fixed but arbitrary x and looks at the relative change in the function depending on the relative change in x. It's going to depend on how you chose to measure distance and what you mean by relative.

Example: Look at the function

$f: \mathbb{R}^2 \to \mathbb{R}; \ (x,y) \longmapsto xy,$

We take the The box norm on $\mathbb{R}^2$ and the usual distance on $\mathbb{R}$ To measure relative closeness to (x,y)

and a close by point $(\tilde{x}, \tilde{y})$ we use the componentwise relative change

$$\max\left(\frac{|x-\tilde{x}|}{|x|}, \frac{|y-\tilde{y}|}{|y|} \right) = R(\tilde{x},\tilde{y}).$$

The relative change on $\mathbb{R}$ between a fixed z and $\tilde{z}$ would be $\frac{|z-\tilde{z}|}{|z|}$.

Of course we're assuming to denominators are not zero

If $|R(\tilde{x},\tilde{y})| < \varepsilon$, where $\varepsilon > 0$ is some small number, we have $\tilde{x} = x(1+\delta_1)$, $\tilde{y} = y(1+\delta_2)$ for some δ_1, δ_2 with $|\delta_1|, |\delta_2| \le \varepsilon$.

Hence $\frac{|f(\tilde{x},\tilde{y}) - f(x,y)|}{|f(x,y)|} = |Q|$

is given by

$$Q = \frac{\tilde{x}\tilde{y} - xy}{xy} = \frac{x(1+\delta_1)\,y(1+\delta_2) - xy}{xy}$$

$$= \frac{(x + \delta_1 x)(y + \delta_2 y) - xy}{xy}$$

$$= \frac{xy + \delta_2 xy + \delta_1 xy + \delta_1 \delta_2 xy - xy}{xy}$$

$$= \delta_2 + \delta_1 + \delta_1 \delta_2 .$$

So $|Q| = |\delta_2 + \delta_1 + \delta_1\delta_2|$

$$\leq |\delta_2| + |\delta_1| + |\delta_1||\delta_2|$$

$$\leq \varepsilon + \varepsilon + \varepsilon^2 = 2\varepsilon + \varepsilon^2$$

Thus the relative error

$$|Q| \leq 2\varepsilon + \varepsilon^2$$

Since ε^2 is very small this is approximately 2ε.

Since $\left| \dfrac{f(\tilde{x},\tilde{y}) - f(x,y)}{f(x,y)} \right| \le 2\varepsilon + o(\varepsilon)$

We will define the condition number as 2. (which happens not to depend on $(x,y) \neq 0$

Example Consider an invertible $n \times n$ matrix A, and let

$f: \mathbb{R}^n \to \mathbb{R}^n$: $x \longmapsto Ax$

be multiplication by A.

Let $\|x\|$, $x \in \mathbb{R}^n$ denote the box norm

and let $\|A\|$ be the operator norm on matrices that results from our choice on $\mathbb{R}^n$. Recall

$$\|A\| = \max\left\{ \sum_{j=1}^{n} |a_{ij}| \;\middle|\; i = 1, \dots, n \right\}$$

We'll measure the relative distance for x to some $\tilde{x}$ as the componentwise relative error $\max\left\{ \frac{|\tilde{x}_i - x_i|}{|x_i|} \middle| i = 1, \dots, n \right\}$

If this is $\le \varepsilon$, then for each i, $|\tilde{x}_i - x_i| \le \varepsilon |x_i| \le \varepsilon \|x\|$. Taking the max over i, we have $\|\tilde{x} - x\| \le \varepsilon \|x\|$. Thus the componentwise relative distance yields an estimate $\frac{\|\tilde{x} - x\|}{\|x\|} \le \varepsilon$, measuring

things by $\frac{\|\tilde{x} - x\|}{\|x\|}$ is a cruder measure of relative error.

but lets use that for the values,

So $\frac{\|A\tilde{x} - Ax\|}{\|Ax\|}$ is our measure.

Then $\frac{\|A\tilde{x} - Ax\|}{\|Ax\|} = \frac{\|A(\tilde{x}-x)\|}{\|Ax\|}$

$\leq \frac{\|A\| \|\tilde{x}-x\|}{\|Ax\|} \leq \frac{\varepsilon \|A\| \|x\|}{\|Ax\|}$ $(*)$

But $\|x\| = \|A^{-1}Ax\| \leq \|A^{-1}\| \|Ax\|$

Thus $(*) \leq \varepsilon \frac{\|A\| \|A^{-1}\| \|Ax\|}{\|Ax\|}$

So $\frac{\|A\tilde{x} - Ax\|}{\|Ax\|} \leq \|A\| \|A^{-1}\| \varepsilon$

So we say the condition number of $f(x) = Ax$ is $\|A\| \|A^{-1}\|$.

This analysis is useful in solving linear equations numerically.

Here is one definition of forward error, with a particular choice of measuring methods. Different authors may give slightly different definitions

<u>Def'n</u> Let $f: \mathbb{R}^n \to \mathbb{R}^m$ and let x be a fixed but arbitrary point in $\mathbb{R}^n$. The <u>condition number</u> of f (at x) is the smallest number K such that

$$\max \frac{|\tilde{x}_i - x_i|}{|x_i|} \leq \varepsilon \Rightarrow$$

$$\frac{\| f(\tilde{x}) - f(x) \|}{\|x\|} \leq K\varepsilon + o(\varepsilon).$$

on the right are evaluated at the output of the previous stages. (this follows from the chain rule).

An algorithm A is called <u>forward stable</u> if

$$\mathcal{K}_A(x) \le C' \mathcal{K}_F(x)$$

where C' is "not too big".

To show this, it is sufficient to show

$$\mathcal{K}_{f_n} \mathcal{K}_{f_{n-1}} \cdots \mathcal{K}_{f_1} \le C' \mathcal{K}_F(x),$$

which is probably the way you would do the analysis. If n is changing you'd like C' to depend on n like $O(n)$ (or slower).

Such an analysis is not easy. Another approach is to look at the backward error.

Wilkinson's idea: the result of a numerical computation on the computer is the result of an exact computation with slightly perturbed initial data

Def'n An algorithm A for computing a mathematical function F is backward stable (at x) if the computed aproximation $\hat{y} = A(x)$ to $y = F(x)$ can be interpreted as the exact result at some point $\tilde{x}$, ie, $\hat{y} = F(\tilde{x})$, where $\tilde{x}$ is close to x in the error measure you're using, say

Sometimes Wilkonson is a litte off and the computed value $\hat{Y} = A(x)$ is not actually a value of F.

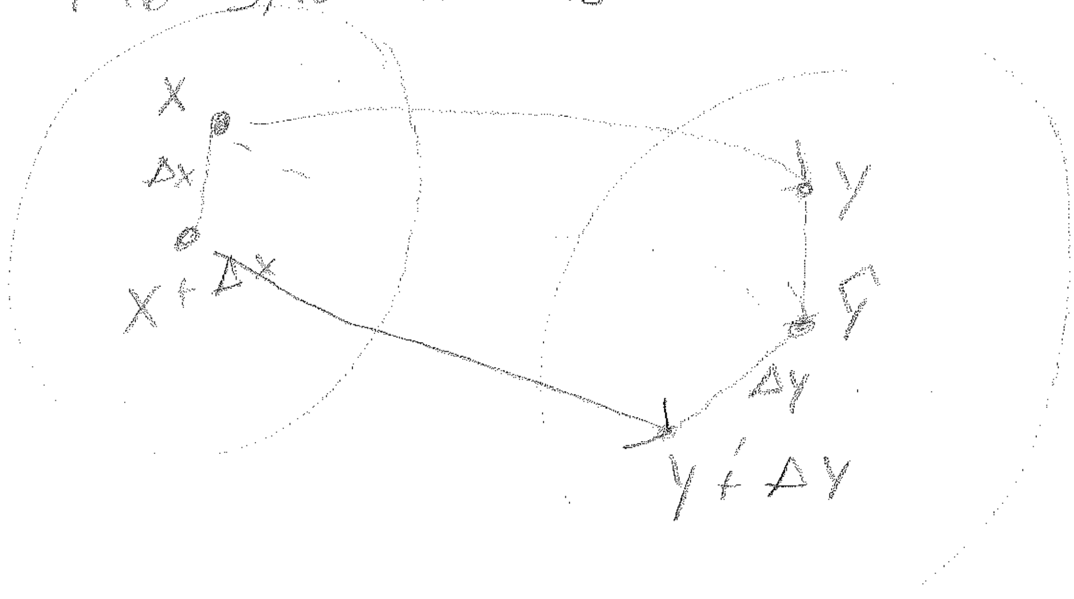
This might happen with a function $F(x) = \cos(x)$ if x is near zero, the computed value will be near one. It might turn out that $\hat{y}$ is a tiny bit bigger than 1, so not in the image of cos.

In a situation like this we expect at $A(x) = \hat{Y}$ will be close to a value $\hat{y} + \Delta Y$ of F, so we'll have

$$\hat{y} + \Delta \hat{y} = F(x + \Delta x)$$

for some small Δx

The Situation is



We say the algorithm is forward-backward stable if we have

$$\hat{y} + \Delta y = F(x + \Delta x)$$

where $\|\Delta y\| < \varepsilon \|y\|$

and $\|\Delta x\| < \eta \|x\|$

for some "small" numbers ε, η, somewhere near machine precision. In other words the computed answer $\hat{y}$ is little different from the exact value $\hat{y} + \Delta y$ that would be produce for input $x + \Delta x$ $(\hat{y} + \Delta y = F(x+\Delta x))$ which differs little from x.

In other words $\hat{y}$ is nearly the right anwer to nearly the right problem.

This notion what we called forward-backward stable. ("A backward stable algorithm to a well-conditioned problem is forward-backward stable.") A forward-backward stable algorithm is the best kind of computation we can hope for when solving real world problems and controlling round-off error.