Math 346 Lecture #11
7.5 Conditioning

7.5.1 Condition Number of a Function

We quantity the "condition number" that measures how sensitive the output of a function is on its input:

$$\textit{change in output} \;=\; \textit{condition number} \;\times\; \textit{change in input}.$$

Definition 7.5.2. For normed linear spaces $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$, nonempty open $U \subset X$, and $f : U \to Y$, the absolute condition number of f at $\mathrm{x} \in U$ is

$$\hat{\kappa}(\mathrm{x}) = \lim_{\delta \to 0^+} \sup_{0<\|\mathrm{h}\|_X<\delta} \frac{\|f(\mathrm{x}+\mathrm{h}) - f(\mathrm{x})\|_Y}{\|\mathrm{h}\|_X}.$$

The absolute condition number "exists" (it might be infinite) because the function

$$g(\delta) = \sup_{0<\|\mathrm{h}\|_X<\delta} \frac{\|f(\mathrm{x}+\mathrm{h}) - f(\mathrm{x})\|_Y}{\|\mathrm{h}\|_X}$$

is bounded below by 0 and is monotone nonincreasing for decreasing δ. We show its value is finite and determine it when f is differentiable at x.

Proposition 7.5.3. For Banach spaces X and Y, nonempty open $U \subset X$, if $f : U \to Y$ is differentiable at $\mathrm{x} \in U$, then

$$\hat{\kappa}(\mathrm{x}) = \|Df(\mathrm{x})\|_{X,Y}.$$

Proof. By the differentiability of f at x, for $\epsilon > 0$ there exists $\nu > 0$ such that for all $\mathrm{h} \in X$ satisfying $\|\mathrm{h}\|_X < \nu$ there holds

$$\frac{\|f(\mathrm{x}+\mathrm{h}) - f(\mathrm{x}) - Df(\mathrm{x})\mathrm{h}\|_Y}{\|\mathrm{h}\|_X} < \epsilon.$$

By the "reverse" triangle inequality we have

$$\big|\|f(\mathrm{x}+\mathrm{h}) - f(\mathrm{x})\|_Y - \|Df(\mathrm{x})\mathrm{h}\|_Y\big| \leq \|f(\mathrm{x}+\mathrm{h}) - f(\mathrm{x}) - Df(\mathrm{x})\mathrm{h}\|_Y.$$

Dividing by $\|\mathrm{h}\|_X$ then gives

$$\left| \frac{\|f(\mathrm{x}+\mathrm{h}) - f(\mathrm{x})\|_Y}{\|\mathrm{h}\|_X} - \frac{\|Df(\mathrm{x})\mathrm{h}\|_Y}{\|\mathrm{h}\|_X} \right| < \epsilon.$$

[At this point in the book, the proof appears to goes horribly wrong in that it looks like it has canceled the common $\|\mathrm{h}\|_X$ in the second term inside the absolute value. Here is the fix.]

This implies that

$$-\epsilon + \frac{\|Df(\mathrm{x})\mathrm{h}\|_Y}{\|\mathrm{h}\|_X} < \frac{\|f(\mathrm{x}+\mathrm{h}) - f(\mathrm{x})\|_Y}{\|\mathrm{h}\|_X} < \epsilon + \frac{\|Df(\mathrm{x})\mathrm{h}\|_Y}{\|\mathrm{h}\|_X}.$$

Applying the supremum over $0 < \|\mathrm{h}\|_X < \delta$, for any $0 < \delta \leq \nu$, to the terms in these inequalities gives

$$-\epsilon + \|Df(\mathrm{x})\|_{X,Y} \leq \sup_{0<\|\mathrm{h}\|_X<\delta} \frac{\|f(\mathrm{x}+\mathrm{h}) - f(\mathrm{x})\|_Y}{\|\mathrm{h}\|_X} \leq \epsilon + \|Df(\mathrm{x})\|_{X,Y}$$

because

$$\sup_{0<\|\mathrm{h}\|_X<\delta} \frac{\|Df(\mathrm{x})\mathrm{h}\|_Y}{\|\mathrm{h}\|_X} = \sup\left\{\frac{\|Df(\mathrm{x})\mathrm{h}\|_Y}{\|\mathrm{h}\|_X} : \mathrm{h} \in X, \mathrm{h} \neq 0\right\} = \|Df(x)\|_{X,Y}$$

by the scaling property of the norms, i.e., $\|c\mathrm{y}\| = |c|\,\|\mathrm{y}\|$ for any norm $\|\cdot\|$.

Thus we have

$$-\epsilon \leq \sup_{0<\|\mathrm{h}\|_X<\delta} \frac{\|f(\mathrm{x}+\mathrm{h}) - f(\mathrm{x})\|_Y}{\|\mathrm{h}\|_X} - \|Df(\mathrm{x})\|_{X,Y} \leq \epsilon,$$

from which we obtain: for all $\epsilon > 0$ there exists $\nu > 0$ such that for all $0 < \delta \leq \nu$ there holds

$$\left|\sup_{0<\|\mathrm{h}\|_X<\delta} \frac{\|f(\mathrm{x}+\mathrm{h}) - f(\mathrm{x})\|_Y}{\|\mathrm{h}\|_X} - \|Df(\mathrm{x})\|_{X,Y}\right| \leq \epsilon.$$

This says that the function

$$g(\delta) = \sup_{0<\|\mathrm{h}\|_X<\delta} \frac{\|f(\mathrm{x}+\mathrm{h}) - f(\mathrm{x})\|_Y}{\|\mathrm{h}\|_X}$$

has limit $\|Df(\mathrm{x})\|$ as $\delta \to 0^+$. □

A better measurement of how sensitive the output of a function is on the input is the relative error. Quantifying this gives the relative condition number.

Definition 7.5.4. For normed linear spaces $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$, nonempty open U in X, and $f : U \to Y$, the relative condition number of f at $\mathrm{x} \in U$ is

$$\kappa(\mathrm{x}) = \lim_{\delta\to 0^+} \sup_{0<\|\mathrm{h}\|_X<\delta} \left(\frac{\|f(\mathrm{x}+\mathrm{h}) - f(\mathrm{x})\|_Y}{\|f(\mathrm{x})\|_Y} \Big/ \frac{\|\mathrm{h}\|_X}{\|\mathrm{x}\|_X}\right) = \frac{\hat{\kappa}(\mathrm{x})}{\|f(\mathrm{x})\|_Y/\|\mathrm{x}\|_X}$$

provided $f(\mathrm{x}) \neq 0$, or the limit of $\kappa(\mathrm{y})$ exists as $\mathrm{y} \to \mathrm{x}$ where $f(\mathrm{x}) = 0$.

Remark 7.5.5. We say a problem is well conditioned at x if $\kappa(\mathrm{x})$ exists (i.e., $f(\mathrm{x}) \neq 0$) and is small, where small depends on the problem. We say a problem is ill conditioned at x if $\kappa(\mathrm{x})$ exists and is large, where large depends on the problem.

Nota Bene 7.5.6. Roughly speaking, the relation condition number give

relative change in ouput = *relative condition number* × *relative change in input.*

The rule of thumb associated with the relative condition number is that without any error in the algorithm itself, we expect to lose k digits of accuracy when the relative condition number is 10^k.

Differentiability of f at x, along with $f(\mathrm{x}) \neq 0$, is sufficient to give an exact value for the relative condition number.

Corollary 7.5.7. For Banach spaces $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$, nonempty U in X, and $f : U \to Y$, if f is differentiable at $\mathrm{x} \in U$, then

$$\kappa(\mathrm{x}) = \frac{\|Df(\mathrm{x})\|_{X,Y}}{\|f(\mathrm{x})\|_Y / \|\mathrm{x}\|_X}$$

provided $f(\mathrm{x}) \neq 0$ or the limit of $\kappa(\mathrm{y})$ exists as $\mathrm{y} \to \mathrm{x}$ where $f(\mathrm{x}) = 0$.

Example (in lieu of 7.5.8). (i) For the function

$$f(x) = \frac{x^2}{x-2}$$

we have

$$Df(x) = \frac{2x(x-2) - x^2}{(x-2)^2} = \frac{x(2(x-2) - x)}{(x-2)^2}$$

so that

$$\kappa(x) = \frac{\dfrac{|x|\,|(2(x-2)-x)|}{(x-2)^2}}{\dfrac{x^2}{|x-2|}\Big/|x|} = \frac{|x-4|}{|x-2|}.$$

The relative condition is defined at $x = 0$ where $f(0) = 0$.

The problem is well conditioned when x is not near $x = 2$, and is especially well conditioned when x is near 4 (which is a critical point of f).

The problem is ill conditioned near $x = 2$.

(ii) Consider the function $y = f(x)$ implicitly defined by

$$0 = F(x, y) = x^5 - y^2 - 4.$$

Since

$$D_2F(x, y) = -2y$$

is invertible when $y \neq 0$, there is by the Implicit Function Theorem a differentiable function $y = f(x)$ such that

$$0 = F(x, f(x)).$$

Differentiation of this with respect to x gives $0 = D_1F(x, f(x)) + D_2F(x, f(x))Df(x)$ so that

$$Df(x) = -\frac{D_1F(x, f(x))}{D_2F(x, f(x))} = -\frac{5x^4}{-2y} = \frac{5x^4}{2y}.$$

Computing the relative condition number we have

$$\kappa = \frac{|5x^4|/|2y|}{|y|/|x|} = \frac{5|x|^5}{2|y^2|} = \frac{5|x|^5}{2|x^5 - 4|}$$

where we have used $0 = x^5 - y^2 - 4$ to get $y^2 = x^5 - 4$.

The problem is ill conditioned when x is close to $4^{1/5}$ and well condition elsewhere, especially when x is close to 0 (which is a critical point of f).

7.5.2 Condition of Finding a Simple Root of a Polynomial

The roots of a quadratic polynomial have an explicit formula a function of the coefficients of the quadratic polynomial: for $a, b, c \in \mathbb{F}$, the roots of $az^2 + bz + c = 0$ are

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Each of these roots, as long as $b^2 - 4ac \neq 0$, is a simple root and is a C^∞ function of a, b, c. Explicit formulas for the roots of cubic and quartic polynomials as functions of the coefficients are known. These formulas show that the simple roots are C^∞ functions of the coefficients of the polynomials. It is a mathematical result due to Galois, that there are no explicit formulas for the roots of a quintic or higher degree polynomial in terms of the coefficients. Are the simple roots of these polynomials C^∞ functions of the coefficients?

We use the Implicit Function Theorem to show that a simple root (i.e., of multiplicity one) of a polynomial of fixed degree varies as a C^∞ function of the coefficients of the polynomial, and we compute the relative condition number of the root as a function of the i^{th} coefficient of the polynomial.

Proposition 7.5.9. For $P : \mathbb{F}^{n+1} \times \mathbb{F} \to \mathbb{F}$ defined by

$$P(\mathrm{a}, x) = \sum_{i=0}^{n} a_i x^i$$

where $\mathrm{a} = \begin{bmatrix} a_0 & a_1 & \cdots & a_n \end{bmatrix}^{\mathrm{T}} \in \mathbb{F}^{n+1}$, if z is a simple root of $p(x) = P(\mathrm{b}, x)$ for some $\mathrm{b} \in \mathbb{F}^{n+1}$, then there is an open neighbourhood U of b in $\mathbb{F}^{n+1}$ and a C^∞ function $r : U \to \mathbb{F}$ that satisfies $r(\mathrm{b}) = z$ and $P(\mathrm{a}, r(\mathrm{a})) = 0$ for all $\mathrm{a} \in U$. Moreover, the relative condition number of r as a function of the i^{th} coefficient a_i at the point (b, z) is

$$\kappa = \left| \frac{z^{i-1} b_i}{p'(z)} \right|.$$

Proof. A root z of p is simple if and only if $p'(z) \neq 0$.

Differentiating P with respect to x and evaluating at (b, z) gives

$$D_x P(\mathrm{b}, z) = \sum_{i=1}^{n} i b_i z^{i-1} = p'(z) \neq 0.$$

By the Implicit Function Theorem there exists an open neighbourhood U of b in $\mathbb{F}^{n+1}$ and a unique C^∞ function $r : U \to \mathbb{F}$ such that $r(\mathrm{b}) = z$ and $P(\mathrm{a}, r(\mathrm{a})) = 0$ for all $\mathrm{a} \in U$.

Differentiating of $P(\mathrm{a}, r(\mathrm{a})) = 0$ with respect to a and evaluating at (b, z) gives

$$D_r(\mathrm{b}) = -D_x P(\mathrm{b}, z)^{-1} D_{\mathrm{a}} P(\mathrm{b}, z) = -\frac{1}{p'(z)} \begin{bmatrix} 1 & z & \cdots & z^n \end{bmatrix}$$

because

$$\frac{\partial P}{\partial a_i} = \frac{\partial}{\partial a_i}\left(\sum_{i=0}^{n} a_i x^i\right) = x^i.$$

Fixing all but the i^{th} coefficient a_i in $r(\mathrm{a})$, we obtain a function whose relative condition number at (b, z) is

$$\kappa = \frac{|-z^i/p'(z)|}{|r(\mathrm{b})|/|b_i|} = \left|\frac{z^i b_i}{zp'(z)}\right| = \left|\frac{z^{i-1}b_i}{p'(z)}\right|.$$

This gives the result. □

7.5.3 Condition Number of a Matrix

We compute the relative condition number of three types of functions that involve invertible square matrices. We find a common upper bound for all three relative condition numbers and describe conditions for which the common upper bound is realized for each of the three relative conditions numbers. These conditions involve right singular vectors x for an invertible $n \times n$ matrix A, i.e., $A = U\Sigma V^{\mathrm{H}}$ where $\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_n)$ are the n singular values of A in decreasing magnitude, and the i^{th} column v_i of V is a right singular vector of A corresponding to the singular value σ_i, i.e., for $\mathrm{u}_i = U\mathrm{e}_i$, we have

$$A\mathrm{v}_i = U\Sigma V^{\mathrm{H}}\mathrm{v}_i = U\Sigma\mathrm{e}_i = \sigma_i U\mathrm{e}_i = \sigma_i\mathrm{u}_i.$$

Theorem 7.5.11. Let $A \in M_n(\mathbb{F})$ be invertible and $\mathrm{x} \in \mathbb{F}^n$.

(i) The relative condition number of $f(\mathrm{x}) = A\mathrm{x}$ is

$$\kappa = \|A\|\frac{\|\mathrm{x}\|}{\|A\mathrm{x}\|} \leq \|A\|\,\|A^{-1}\|.$$

Equality holds when the norm is $\|\cdot\|_2$ and x is a right singular vector corresponding to the minimal singular value of A.

(ii) The relative condition number of $g(A) = A\mathrm{x}$ is

$$\kappa = \|\mathrm{x}\|\frac{\|A\|}{\|A\mathrm{x}\|} \leq \|A\|\,\|A^{-1}\|.$$

Equality holds when the norm is $\|\cdot\|_2$ and x is a right singular vector corresponding to the minimal singular value of A.

(iii) The relative condition number of $h(\mathrm{x}) = A^{-1}\mathrm{x}$ is

$$\kappa = \|A^{-1}\|\frac{\|\mathrm{x}\|}{\|A^{-1}\mathrm{x}\|} \leq \|A\|\,\|A^{-1}\|.$$

Equality holds when the norm is $\|\cdot\|_2$ and x is a left singular vector corresponding to the maximum singular value of A.

Proof. (i) Since $Df(\mathrm{x}) = A$ we have by Corollary 7.5.7 that

$$\kappa = \frac{\|Df(\mathrm{x})\|}{\|f(\mathrm{x})\|/\|\mathrm{x}\|} = \frac{\|A\|}{\|A\mathrm{x}\|/\|\mathrm{x}\|} = \|\mathrm{x}\|\frac{\|A\|}{\|A\mathrm{x}\|}.$$

To get the upper bound of $\|A\|\,\|A^{-1}\|$ on κ we use the invertibility of A to get

$$\begin{aligned}\|A^{-1}\| &= \sup\left\{\frac{\|A^{-1}\mathrm{y}\|}{\|\mathrm{y}\|} : \mathrm{y} \in \mathbb{F}^n, \mathrm{y} \neq 0\right\}\\ &= \sup\left\{\frac{\|A^{-1}A\mathrm{x}\|}{\|A\mathrm{x}\|} : \mathrm{x} \in \mathbb{F}^n, \mathrm{x} \neq 0\right\}\\ &= \sup\left\{\frac{\|\mathrm{x}\|}{\|A\mathrm{x}\|} : \mathrm{x} \in \mathbb{F}^n, \mathrm{x} \neq 0\right\},\end{aligned}$$

which implies for all nonzero $\mathrm{x} \in \mathbb{F}^n$ that

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

whence

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

The rest of the proof is HW (Exercise 7.2.7; make use of Exercise 4.31).

(ii) The function $g(A) = A\mathrm{x}$ is differentiable at each invertible A because

$$g(A+H)\mathrm{x} - g(A) - H\mathrm{x} = A\mathrm{x} - H\mathrm{x} - A\mathrm{x} - H\mathrm{x} = 0$$

implies that

$$\lim_{H\to 0}\frac{\|g(A+H) - g(A) - H\mathrm{x}\|}{\|H\|} = 0,$$

whence $Dg(A)H = H\mathrm{x}$. With

$$\begin{aligned}\|Dg(A)\| &= \sup\left\{\frac{\|Dg(A)H\|}{\|H\|} : H \in M_n(\mathbb{F}), H \neq 0\right\}\\ &= \sup\left\{\frac{\|H\mathrm{x}\|}{\|H\|} : H \in M_n(\mathbb{F}), H \neq 0\right\}\\ &\leq \sup\left\{\frac{\|H\|\,\|\mathrm{x}\|}{\|H\|} : H \in M_n(\mathbb{F}), H \neq 0\right\}\\ &= \|\mathrm{x}\|,\end{aligned}$$

we have

$$\kappa = \frac{\|Dg(A)\|}{\|g(A)\|/\|A\|} \leq \frac{\|\mathrm{x}\|\|A\|}{\|A\mathrm{x}\|} \leq \|A\|\,\|A^{-1}\|,$$

where the last inequality was proved in part (i).

When the norm is $\|\cdot\|_2$, then $\|Dg(A)\| = \|\mathrm{x}\|$ by Exercise 3.29, and Exercise 7.2.7 gives equality of κ with $\|A\|\,\|A^{-1}\|$ for a right singular vector of A corresponding to the minimal singular value of A.

(iii) Replacing A with A^{-1} in part (i) gives

$$\kappa = \|\mathrm{x}\|\frac{\|A^{-1}\|}{\|A^{-1}\mathrm{x}\|} \leq \|A^{-1}\|\,\|A\|.$$

Equality holds in the norm $\|\cdot\|_2$ when x is a left singular vector of A corresponding to the largest singular value of A: from $A = U\Sigma V^{\mathrm{H}}$ we have $A^{-1} = V\Sigma^{-1}U^{\mathrm{H}}$, so that with the left singular vector $\mathrm{u}_1 = U\mathrm{e}_1$ corresponding to the maximum singular value σ_1 we have $\|\mathrm{u}_1\|_2 = 1$ and $A^{-1}\mathrm{u}_1 = V\Sigma^{-1}U^{\mathrm{H}}\mathrm{u}_1 = V\Sigma^{-1}\mathrm{e}_1 = (1/\sigma_1)V\mathrm{e}_1 = (1/\sigma_1)\mathrm{v}_1$, hence $\|A^{-1}\mathrm{u}_1\|_2 = \|(1/\sigma_1)\mathrm{v}_1\|_2 = 1/\sigma_1$, so that $\|\mathrm{u}_1\|_2/\|A^{-1}\mathrm{u}_1\|_2 = \sigma_1 = \|A\|_2$. □

Definition 7.5.12. *The* condition number of an invertible $A \in M_n(\mathbb{F})$ is defined to be

$$\kappa = \|A\|\,\|A^{-1}\|$$

(as this is the common upper bound for all three functions involving an invertible matrix).

Example. (a) We compute the relative condition number for the invertible

$$A = \begin{bmatrix} 9 & 8 \\ 1 & 1 \end{bmatrix} \in M_2(\mathbb{R})$$

with inverse

$$A^{-1} = \begin{bmatrix} 1 & -8 \\ -1 & 9 \end{bmatrix}.$$

In the induced 1-norm, the relative condition number of A is

$$\kappa = \|A\|_1\|A^{-1}\|_1 = 10 \cdot 17 = 170,$$

and in the induced ∞-norm the relative condition number of A is also

$$\kappa = \|A\|_\infty\|A^{-1}\|_\infty = 17 \cdot 10 = 170.$$

(b) The relative condition number for a matrix is norm-dependent as we show for the invertible

$$A = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 2 \\ -3 & -5 & 7 \end{bmatrix} \in M_3(\mathbb{R})$$

with inverse

$$A^{-1} = \begin{bmatrix} 10 & -7 & 2 \\ 1 & 0 & 0 \\ 5 & -3 & 1 \end{bmatrix}.$$

In the induced 1-norm, the relative condition number of A is

$$\kappa = \|A\|_1\|A^{-1}\|_1 = 9 \cdot 16 = 144,$$

but in the induced ∞-norm, the relative condition number of A is

$$\kappa = \|A\|_\infty\|A^{-1}\|_\infty = 15 \cdot 19 = 285.$$