

Math417
Mathematical Programming

Homework VII

Frédéric Boileau

Prof. Tim Hoheisel

3 novembre 2016

Contents

1	Nonsmooth Fermat's rule	2
2	Convergence	2
3	Operator Norms	4
4	Operator 1-norm	5
5	Local Lipschitz continuity of smooth functions	6

1 Nonsmooth Fermat's rule

Let $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ and $\bar{x} \in \text{dom} f$.

Claim: \bar{x} is a minimizer of f iff $0 \in \partial f(\bar{x})$, where the last expression denotes the subdifferential of f .

$$\begin{aligned}\partial f(\bar{x}) &:= \{v \in \mathbb{R}^n \mid f(x) \geq f(\bar{x}) + \langle v, x - \bar{x} \rangle\} \\ 0 \in \partial f(\bar{x}) &\iff f(x) \geq f(\bar{x}) + \langle 0, x - \bar{x} \rangle = f(\bar{x}) \quad \forall \quad x \in \text{dom} f\end{aligned}$$

The last line completes the proof as this is the definition of a minimizer.

2 Convergence

We will first prove a theorem which is quite powerful in its own right. First let $\rho(A)$ denote the spectral radius of the matrix A which means the following:

$$\rho(A) := \max\{|\lambda_1|, \dots, |\lambda_n|\} \quad (1)$$

Where the lambdas designate the eigenvalues of the matrix A .

Claim: Let $A \in \mathbb{R}^n$ with spectral radio $\rho(A)$, then $\rho(A) < 1$ is a necessary and sufficient condition for $A^k \rightarrow 0$.

Proof: Let (v, λ) be a non trivial eigen pair.

$$\begin{aligned}\left(\lim_{k \rightarrow \infty} A^k\right)v &= 0 \\ \lim_{k \rightarrow \infty} (A^k v) &= 0 \\ \lim_{k \rightarrow \infty} (\lambda^k v) &= 0 \\ v \lim_{k \rightarrow \infty} (\lambda^k) &= 0 \\ \therefore \lim_{k \rightarrow \infty} \lambda^k &= 0\end{aligned}$$

Because we assumed v was non zero. Moreover the last line implies $|\lambda| < 1$ for all eigen values, which is what we set out to prove. (Notice we can read the reasoning in the other direction which means we also proved sufficiency).

We now prove another statement and those two together complete the proof.
Claim: Let $A \in \mathbb{R}^n$ equipped with a consistent matrix norm $\rho(A)$, then

$$\rho(A) \leq \|A^k\|^{1/k} \quad \forall k \in \mathbb{N} \quad (2)$$

Proof: Once again let (v, λ) be a non trivial eigen pair for a matrix A . We know the norm is submultiplicative:

$$|\lambda|^k \|v\| = \|\lambda^k v\| \leq \|A^k\| \cdot \|v\| \quad (3)$$

And since we chose a non trivial pair ($v \neq 0$) and the norm of a vector is by definition 0 iff the vector is 0, we have

$$|\lambda|^k \leq \|A^k\| \quad \forall k \in \mathbb{N} \quad (4)$$

and for all eigen values. Now 4 easily yields the desired inequality 2.

We now prove the original claim using the sub-multiplicativity of the norm:

$$\lim_{k \rightarrow \infty} \|A\|^k = 0 \quad \Rightarrow \quad \lim_{k \rightarrow \infty} \|A^k\| = 0 \quad \Rightarrow \quad \lim_{k \rightarrow \infty} \rho(A)^k = 0 \quad (5)$$

The last equality implies that the spectral radius is less than 1. We can now finally invoke the first claim we made to conclude that necessarily $A^k \rightarrow 0$ which concludes the proof.

■

3 Operator Norms

Let $\|\cdot\|_X$ and $\|\cdot\|_Y$ be vector norms on $X := \mathbb{R}^n$ and $Y := \mathbb{R}^m$ respectively. We wish to show that the operator norm as defined in class is indeed a norm. This means we wish to show that the mapping is absolutely homogeneous, that it respects the triangle inequality and maps 0 to and only to 0. Let the operator norm be defined as follows:

$$A \in R^{m \times n} \mapsto \sup \frac{\|Ax\|_Y}{\|x\|_X} \quad (6)$$

We assume the norm of x is not zero when we are supping throughtout this section for latex speed reasons. For the absolute homogeneous property it is a pretty direct consequence of vector norm and sup basic properties.

$$\sup \frac{\|\alpha Ax\|_Y}{\|x\|_X} = \sup \frac{|\alpha| \|Ax\|_Y}{\|x\|_X} = |\alpha| \sup \frac{\|Ax\|_Y}{\|x\|_X} \quad \checkmark \quad (7)$$

We now verify the triangle inequality which is again a basic result of norm and sup basic properties:

$$\|A+B\| = \sup \frac{\|(A+B)x\|_Y}{\|x\|_X} \leq \sup \frac{\|Ax\|_Y}{\|x\|_X} + \sup \frac{\|Bx\|_Y}{\|x\|_X} = \|A\| + \|B\| \quad \checkmark \quad (8)$$

Finally it is clear to see that any matrix maps 0 to 0 and that supping over 0 gives us 0 so we have now all the basic properties of a well defined norm.

We now wish to show that the norm could be defined as supping the norm of the matrix product over the unit ball and and also restrict it over its closure. We start by proving one of the claims and adding an obvious inequality:

$$\sup \frac{\|Ax\|_Y}{\|x\|_X} = \sup \|A \frac{x}{\|x\|_X}\| = \sup_{\|x\|_X=1} \|Ax\|_Y \leq \sup_{\|x\|_X \leq 1} \|Ax\|_Y$$

Moreover:

$$\sup_{\|x\|_X \leq 1} \|Ax\|_Y \leq \sup_{\|x\|_X \leq 1} \frac{\|Ax\|}{\|x\|_X} = \sup \|A \frac{x}{\|x\|_X}\| = \sup_{\|\xi\|=1} \|A\xi\| \quad (9)$$

This inequality with the one derived previously shows that equality holds.

■

4 Operator 1-norm

Let $A \in \mathbb{R}^{m \times n}$, we wish to show that

$$\|A\|_1 = \max_{j=1, \dots, n} \sum_{i=1}^m |a_{ij}| \quad (10)$$

We will derive the norm, also using the result from the previous exercise:

$$\|Ax\|_1 = \sup_{\|x\|_1=1} \sum_{i=1}^m \left| \sum_{j=1}^n a_{ij} x_j \right| \quad (11)$$

$$\sum_{i=1}^m \left| \sum_{j=1}^n a_{ij} x_j \right| \leq \sum_{i=1}^m \sum_{j=1}^n |a_{ij} x_j| \quad (12)$$

$$= \sum_{j=1}^n |x_j| \sum_{i=1}^m |a_{ij}| \quad (13)$$

Now let $A_j = \sum_{i=1}^m |a_{ij}|$ and take its argmax over the indices to be J . In other words, the columns with the highest 1-norm. We then have the obvious inequality:

$$\|Ax\|_1 \leq A_J \sum_{j=1}^n |x_j| = A_J \|x\|_1 \quad (14)$$

Where the last equality is simply by the definition of the 1-norm of a vector. This last line gives us the following inequality:

$$\sup_{\|x\|_1=1} \|Ax\|_1 \leq \sup_{\|x\|_1=1} A_J \|x\|_1 = A_J \quad (15)$$

We finally get the desired inequality

$$\|A\|_1 \leq A_J \quad (16)$$

But A_J exists and is exactly the RHS of 10, which completes the proof.

■

5 Local Lipschitz continuity of smooth functions

Let $G : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be continuously differentiable. Show that G is locally Lipschitz continuous at every point $\bar{x} \in \mathbb{R}^n$. We first observe that while global Lipschitz continuity is a strong condition, local Lipschitz continuity merely means that the rate of change is bounded everywhere, which means that the second derivatives exist and are bounded which is equivalent to saying the function is continuously differentiable. To do the actual formal proof we first prove the following:

Claim: Let G be defined as above and continuously differentiable, then there exists a neighborhood about every point in \mathbb{R}^n such that the Jacobian is bounded, i.e. its norm is bounded. We will just prove the statement for the spectral norm.

Proof: Let $J(x)$, $\rho(A)$ denote the jacobian of G and the spectral radius of A respectively. Since $G(x)$ is continuously differentiable $D_j f_i(x)$ is differentiable for all $i = 1, \dots, m$ and $j = 1, \dots, n$. Moreover we know from basic real analysis that linear combinations of differentiable functions are differentiable. Let $A = (J^T J)(x)$. By the above statement all the entries of A are differentiable. Let $B_\epsilon(\bar{x})$ be an open ball centered at \bar{x} and take L^2 to be the maximum value that $\rho(A)$ takes over that neighborhood. Formally let

$$L^2 = \max\{\rho(A(x)) \mid x \in B_\epsilon(\bar{x})\} \quad (17)$$

Clearly $L^2 \neq \pm\infty$ since $A(x)|_{\hat{x}}$ is a matrix with real entries for all $\hat{x} \in B_\epsilon(\bar{x})$. Now since $\|A\|_2$ can be shown as in class to be $\sqrt{\rho(A)}$ we have proven the claim which we state formally:

$$\forall \bar{x} \in \mathbb{R}^n \quad \exists \quad \epsilon > 0 \quad \text{and} \quad L \in \mathbb{R} \quad \text{s.t.} \quad \|A\|_2 \leq L \quad \forall x \in B_\epsilon(\bar{x}) \quad (18)$$

Moreover L is defined by 17 and is clearly the least upper bound. Now let $a, b \in B_\epsilon(\bar{x})$. An open ball is a convex set so it contains all its inner convex combinations. Let us define the following parametrized curve:

$$\phi(t) := G(a + \lambda(b - a)) \quad \lambda \in [0, 1] \quad (19)$$

By convexity the curve lies entirely in $B_\epsilon(\bar{x})$.

$$G(b) - G(a) = \phi(1) - \phi(0) = \int_0^1 \phi'(t) dt = \int_0^1 dG(a + \lambda(b - a)) \cdot (b - a) dt$$

Where the last equality is due to the chain rule.

Now using the triangle inequality for integrals, the bound from 18 and the fact that the curve is entirely in $B_\epsilon(\bar{x})$ we have:

$$|G(b) - G(a)| = \left| \int_0^1 dG(a + \lambda(b - a)) \cdot (b - a) dt \right| \quad (20)$$

$$\leq \int_0^1 |dG(a + \lambda(b - a)) \cdot (b - a)| dt \quad (21)$$

$$\leq \int_0^1 L|b - a| \quad (22)$$

$$= L|b - a| \quad (23)$$

Which proves local Lipschitz continuity.

■

Now I'd like to thank my manager, my family, my friends, all those people that have made this possible to happen, the crowd at stackexchange, let's not forget them, this is incredible, I never thought I'd come this far, just goes to remind you everything is possible! life is but a dream, but

*-In the end,
the water was too cold for us*

-Robert Lowell