Math 417
Mathematical Programming

Homework VIII

Frédéric Boileau

Prof. Tim Hoheisel

10 novembre 2016

Contents

1 Examples of locally Lipschitz functions

Find a function $f : \mathbb{R} \to \mathbb{R}$ which

(a) is continuous but not locally Lipschitz continous.

$$\text{Let } f(x) := x^{1/3}$$

Suppose $x \geq 0$. Then we have that $|f(x) - f(0)| = f(x)$. To be locally Lipschitz we should be able to find $L > 0$ and neighborhood $B_\epsilon(0)$ such that if $x \in B_\epsilon(0)$ then we have that$f(x) \leq L \times x$. Let's look at what happens for x smaller than 1. If we choose any $x < \frac{1}{\sqrt{L}^3}$ for some $\delta > 0$ we immediately have $f(x) > Lx$ which contradicts local Lipschitz continuity.

(b) is locally Lipschitz but not differentiable

$$\text{Let } f(x) := |x|$$

It is clear that $f(x)$ is not differentiable. It is also quite clearly locally Lipschitz everywhere except at the origin since it consists of two linear parts. It remains to show that it is Lipschitz at 0. Well but around the origin every $L > 1$ will work since $|f(x) - f(0)| = f(x) = |x|$. Moreover $|x - 0| = |x|$. So clearly :

$$|f(x) - f(0)| \leq L|x - 0| \quad \forall L \geq 1$$

2 Solvability of central path conditions

$$\min\ x_1 + x_2 \quad \text{s.t.} \quad x_1 + x_2 \geq 1,\ x_1, x_2 \geq 0 \tag{1}$$

(a) Determine all solutions of 1. It is easy to see that since $c = A$ all the points where the constraint is active are solutions of the LP. The two BFS are when either of the two variables are 0.

(b) Transform 1 into standard form We only need to add a slack variable to bring the problem into standard form. Let $x_3 \geq 0$. Then 1 becomes

$$\min\ x_1 + x_2 \quad \text{s.t.} \quad x_1 + x_2 - x_3 = 1,\ x_1, x_2, x_3 \geq 0 \tag{2}$$

So the matrix A in this case is only a row vector.

(c) For $\tau > 0$ determine the solution (x_τ, y_τ, z_τ) of the central path conditions with the reformulation from (b). Let's look at the equations we have:

$$1 = x_1 + x_2 - x_3 \tag{3}$$
$$\tau = x_i s_i \tag{4}$$
$$0 < x, s \tag{5}$$
$$1 = y + s_1 \tag{6}$$
$$1 = y + s_2 \tag{7}$$
$$0 = y + s_3 \tag{8}$$

From 6 and 7 we have that $s_1 = s_2$ and from 8 we have that $s_3 = -y$. Putting these two together we have $s_3 = s_1 - 1$. This gives us an additional constraint : $s_1 > 1$. We can now reformulate 3 as follows:

$$\frac{2\tau}{s_1} - \frac{\tau}{s_1 - 1} = 1 \quad s_1 > 1 \tag{9}$$

This equation in turn gives us a quadratic and we now express the possible values of s_1 in terms of τ

$$s_1 = \frac{1 + \tau \pm \sqrt{\tau^2 - 6\tau + 1}}{2} \tag{10}$$

The discriminant has two roots: $\tau = 3 \pm 2\sqrt{2}$ and is negative between them. Moreover we have a constraint on s_1 which is that is has to be strictly greater than one, hence :

$$\tau \pm \sqrt{\tau^2 - 6\tau + 1} > 1 \tag{11}$$

Which gives us the constraint $\tau > 3 + 2\sqrt{2}$ hence $s_1 > 2 + \sqrt{2}$.

Since the system is uniquely determined we have all the information we need for the solution of the central path conditions. From s_1 we first immediately get s_2 and s_3 which in turn give us x_1, x_2 and x_3. Finally the y's are uniquely determined since the matrix has only one row, it can only have full row rank. Actually y is a scalar. We do not explicitly write (x_τ, y_τ, s_τ) as this seems quite unnecessary and not very neat.

(d) As $\tau \to 0$ we cannot use the mapping found above as it is not defined for values of τ smaller than the constraint we have imposed on it. So we stop tripping and let s_1 take imaginary values as τ goes to 0. Then $s_1 \to 1$ hence so does s_2 and $s_3 \to 0$. Thus $x_1, x_2 = 0$ and $x_3 = -1$ and finally $y = -1$. Which is a bad solution since our s_i are non zero, which is not good, I think.

3 Strictly convex functions

A convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is called *strictly convex* on the convex set $C \subset \text{dom} f$ if

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y) \quad (x, y \in C, x \neq y, \lambda \in (0,1)) \tag{12}$$

(a) If f is strictly convex on $\text{dom} f \neq \emptyset$ then argmin f is either empty or a singleton. We will derive a proof by contradiction. Assume f is a strictly convex function on a set $C \subset \mathbb{R}^n$ and that there exists two distinct minimizers x_1 and x_2. WLOG assume that $f(x_1) \leq f(x_2)$.

$$\lambda f(x_1) \leq \lambda f(x_2) \quad \forall\ \lambda \in (0,1) \tag{13}$$

$$\lambda f(x_1) + (1-\lambda) f(x_2) \leq \lambda f(x_2) + (1-\lambda) f(x_2) \tag{14}$$

$$\lambda f(x_1) + (1-\lambda) f(x_2) \leq f(x_2) \tag{15}$$

Now we apply strict convexity to 15 to get

$$f(\lambda x_1 + (1-\lambda) x_2) < f(x_2) \quad \forall\ \lambda \in (0,1) \tag{16}$$

This is the desired contradiction. The strict inequality above means that every point between x_1 and x_2 has an objective value strictly less than $f(x_1)$. But this means:

$$\nexists\ \epsilon > 0 : f(x) \geq f(x_1) \quad \forall\ x \in B_\epsilon(x_1) \tag{17}$$

This clearly contradicts the assumption that x_1 was a minimizer in the first place.

$\blacksquare$

(b) Suppose C is open and f is continuously differentiable on C. Then f is strictly convex on C iff

$$f(x) > f(\bar{x}) + \nabla f(\bar{x})^\mathsf{T}(x - \bar{x}) \quad (x, \bar{x} \in C, x \neq \bar{x}) \tag{18}$$

By the mean value theorem we know there exists a point $\tau_\lambda \in \mathbb{R}^n$ on the line between $\bar{x}$ and $\bar{x} + \lambda(x - \bar{x})$ such that

$$\begin{aligned} \lambda \nabla f(\tau_\lambda)^\mathsf{T}(x - \bar{x}) &= f(\bar{x} + \lambda(x - \bar{x})) - f(\bar{x}) && (19) \\ &< \lambda f(x) + (1-\lambda) f(\bar{x}) - f(\bar{x}) && (20) \\ &= \lambda(f(x) - f(\bar{x})) && (21) \end{aligned}$$

By construction we have $\tau_\lambda \to \bar{x}$ as $\lambda \to 0$. Now since f is continuously differentiable we have that $\lim_{\tau_\lambda \to \bar{x}} \nabla f(\tau_\lambda) = \nabla f(\bar{x})$. This with the inequality derived previously directly yields:

$$f(x) - f(\bar{x}) > \nabla f(\bar{x})^\mathsf{T}(x - \bar{x}) \tag{22}$$

This gives us "$\Rightarrow$"

For the other direction, i.e. if 18 holds then f is convex, we take two arbitrary points in C, say x and y. Moreover let $\lambda \in (0,1)$ and define $z := \lambda x+(1-\lambda)y$. Then we have

$$f(x) - f(z) > \nabla f(z)^{\mathsf{T}}(x-z) \tag{23}$$
$$f(y) - f(z) > \nabla f(z)^{\mathsf{T}}(y-z) \tag{24}$$

We multiply 23 and 24 by λ and $1-\lambda$ respectively and add them. The RHS cancels out and we are left with

$$\lambda f(x) + (1-\lambda)f(y) - f(z) \tag{25}$$
$$= \lambda f(x) + (1-\lambda)f(y) - f(\lambda x + (1-\lambda)y) \tag{26}$$
$$> 0 \tag{27}$$

Which is the definition of strict convexity.
■

4 Log-barrier function

Let $lb : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be the log barrier function:

$$\text{lb}(x) := \begin{cases} -\sum_{i=1}^n \log(x_i) & \text{if} \quad x > 0 \\ +\infty & \text{else} \end{cases}$$

(a) lb is continuously differentiable on $\text{dom} f$ with

$$\nabla \text{lb}(x) = -\begin{pmatrix} x_1^{-1} \\ \vdots \\ x_n^{-1} \end{pmatrix} \quad (x > 0) \tag{28}$$

The proof of 28 is quite direct from the definition of the gradient. Let f denote the log-barrier function as defined previously.

$$\begin{aligned} \nabla \text{lb}(x) &= \sum_{i=1}^n \frac{\partial f}{\partial x_i} e_i \\ &= \sum_{i=1}^n e_i \left(-\frac{\partial}{\partial x_i} \sum_{i=1}^n \log(x_i) \right) \\ &= -\sum_{i=1}^n e_i \times \frac{1}{x_i} \end{aligned}$$

The last line is just another way to write the RHS of 28. Moreover it is clearly continuous for $x \in \mathbb{R}^n > 0$ where the last inequality is element wise. ∎

(b) We use the fact that a function is strictly convex if and only if its Hessian is definite positive. The later fact is the first we prove since otherwise this approach wouldn't be very useful now would it.

Claim:

$$\nabla^2 \mathrm{lb}(x) > 0 \quad \forall\, x \in \mathbb{R}^n \tag{29}$$

Proof:
This is pretty direct from the gradient we have already derived. Since the partial of the log-barrier w.r.t. x_i is $1/x_i$, clearly we have that :

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = 0 \quad \text{whenever} \quad i \neq j \tag{30}$$

Moreover:

$$\frac{\partial^2 f}{\partial x_i^2} = \frac{1}{x_i^2} \tag{31}$$

This gives us all the entries of the Hessian which can be expressed as follows:

$$\nabla^2 \mathrm{lb}(x) = \mathrm{diag}(x)^{-2} \tag{32}$$

Since the matrix is diagonal, we already have the eigenvalues given, namely $1/x_i^2$. Since these are clearly strictly positive for all $x \in \mathbb{R}$ we have that the Hessian is positive definite.
■

Now only need to show that a function $f : \mathbb{R}^n \to \mathbb{R} \cup \infty$ is convex if and only if its Hessian is definite positive for all $x \in \mathbb{R}^n$. We need not worry about the domain since as proven in the last assignment if the function is convex on a domain and defined to take ∞ outside it, the extended function will still be convex. Now the domain of the log-barrier function is the positive orthant so we will only consider values of x in this orthant.

<u>Claim</u> : Let f be a scalar field and have a positive definite hessian, then for all x, y in the domain and $\lambda \in (0, 1)$ we have that

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y) \tag{33}$$

We reduce it to a 2 dimensional problem by parameterizing a line from x to y, which is just all convex combinations of those two points:

$$g(\lambda) := f(\lambda x + (1-\lambda)y)$$

With a little bit of matrix calculus we arrive at the expression of the second derivative of $g(t)$

$$g''(\lambda) = (x-y)^{\mathsf{T}} \nabla^2 f(\lambda x + (1-\lambda)y)(x-y) \tag{34}$$

Since the Hessian is definite positive we clearly have that $g(\lambda) > 0$ for all λ. Now let's take the Taylor expansion of g around an arbitrary λ:

$$g(\mu) = g(\lambda) + (\mu - \lambda)g'(\lambda) + \mathcal{O}(g''(\lambda))$$

The fact that $g(\lambda)$ is positive definite allows us to write the following inequalities:

$$g(0) > g(\lambda) + g'(\lambda)(0 - \lambda) \tag{35}$$
$$g(1) > g(\lambda) + g'(\lambda)(1 - \lambda) \tag{36}$$

Now we multiply 35 by λ and 36 by $1 - \lambda$ and them. This gives us:

$$g(\lambda)` < \lambda g(1) + (1-\lambda)g(0) \tag{37}$$

Which is clearly just another way to write strict convexity.
■

(c) Let x_k be a sequence that is the in the domain for all k but goes to $\bar{x}$ which is on the boundary. Clearly the boundary of the domain is the surfaces that delimit the positive orthant. There are two cases, x_k goes to the origin or x_k has one of its components that goes to zero. If one of the x_i goes to zero than $-\log x_i \to \infty$ and so lb $\to \infty$. Also if x_k goes to zero for all its components (origin) then the log-barrier function blows up even faster. In all cases log-barrier goes to infinity. Moreover log-barrier is defined to be ∞ outside its domain so clearly $\lim_{x_k \to \bar{x}} \mathrm{lb}(x) = \infty = f(\bar{x})$ when $\bar{x} \in bd$. So it follows that the function is continuous on its domain. ■