Parabola Volume 30, Issue 3 (1994)

AN INTRODUCTION TO NONSMOOTH CALCULUS

V. Jeyakumar[1] and B. M. Glover[2]

In this short article we will introduce some important notions concerned with the class of *convex* functions. These functions are not necessarily differentiable in the usual sense of calculus. The motivation for studying such functions is that they frequently arise in many applications of mathematics to finance, economics and engineering. One branch of mathematics in which convex functions are often studied is *Mathematical Programming.* This is the study of *optimization problems*; that is problems in which we wish to optimize (either maximize or minimize) a function (usually of several variables) often subject to a collection of restrictions on these variables. The restrictions are known as *constraints* and the function to be optimized is the *objective* function. A classical example of an optimization problem is the problem of maximizing profit subject to limitations on available resources, manpower etc. For the sake of simplicity and to focus on the key ideas we will be interested only in functions of a single variable and simple optimization problems. Not perhaps surprisingly many of the simple concepts and techniques we will touch upon have extensions to the most general multi- dimensional case involving functions depending on many variables.

An important aspect of the study of optimization problems is to determine conditions for optimality of a function at a given point. For example from elementary calculus you should be familiar with the fact that for a differentiable function $y = f(x)$ a *necessary* condition for $f(x)$ to achieve a local optimum (either maximum or minimum) at a point a is that

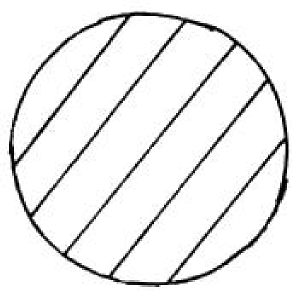
$$f'(a) = 0. \tag{1}$$

This says that at a local optimum the graph of the function has a horizontal tangent. We say that (1) is a *first order* necessary optimality condition (since it involves first order derivatives). A condition such as (1) can be used in a computational method for locating local optima; that is if we can locate a point a satisfying (1) then this point is a candidate for local optimum of the function. Of course we should note that (1) need *not* indicate either a local maximum or minimum. Consider the

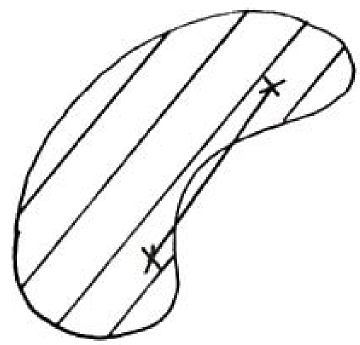
[1]Applied Mathematician at the School of Mathematics, The University of New South Wales.
[2]Applied Mathematician at the School of Mathematics and Computing, University of Ballarat, Victoria.

simple example $f(x) = x^3$, here $x = 0$ is a stationary point (i.e satisfying (1)) but this point corresponds to a point of inflexion rather than a local optimum. Thus, in general, (1) is a necessary but not *sufficient* condition for optimality. However, for example, there is a broad class of functions for which every stationary point is a *global* minimum. These functions are called *convex* functions. To consider the properties of these functions we first define a *convex* set in the Euclidean plane $\mathbb{R}^2$.

Definition 1 A set C is *convex* if it contains all points on the line segments joining any two points in the set.



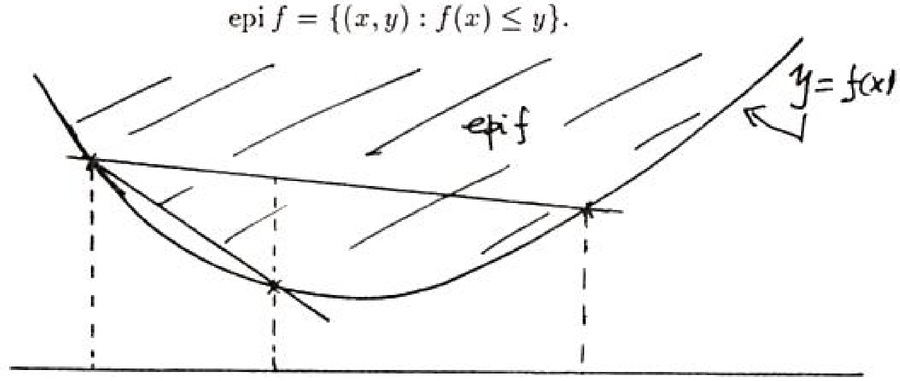
A convex set



A nonconvex set

Definition 2 A function $f(x)$ is said to be *convex* if, the chord connecting any two points on the graph of the function always lies above the graph. In other words, a function f is convex means that its *epigraph*, epi f, is a convex set. Here

$$\text{epi}\, f = \{(x, y) : f(x) \leq y\}.$$



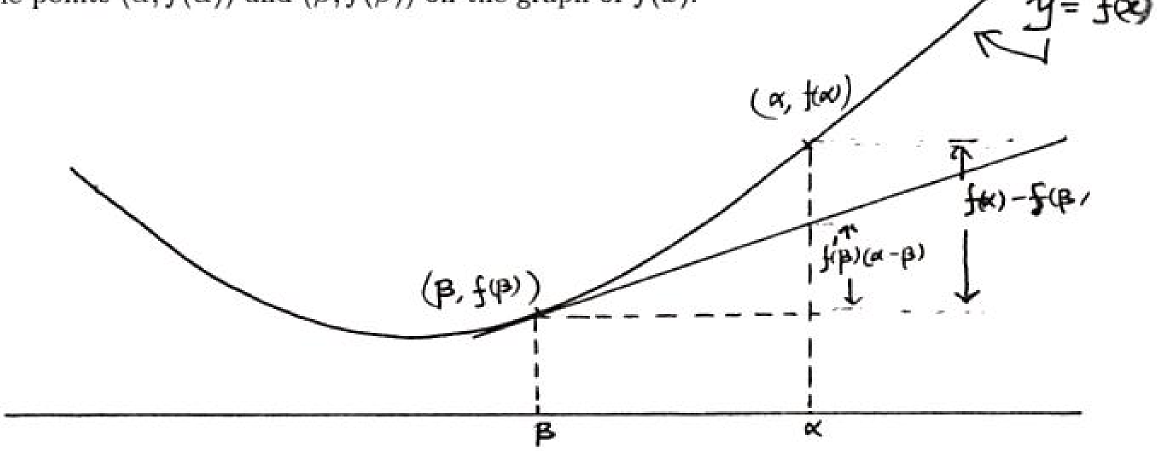
The epigraph of a convex function.

Exercise 1 By considering their graphs show that $f(x) = x^2$ is a convex function whereas $g(x) = x^3$ is not.

Another important property of a convex function f is that, for any α, β, we have

$$f(\alpha) - f(\beta) \geq f'(\beta)(\alpha - \beta). \tag{2}$$

The inequality tells us that the secants that emanate from the point $(\beta, f(\beta))$ always lie above the tangent line at this point. Here the secant is the line segment joining the points $(\alpha, f(\alpha))$ and $(\beta, f(\beta))$ on the graph of $f(x)$.



It is straightforward to show, for example, that (2) is satisfied by the function $f(x) = x^2$:

$$\begin{aligned} f(\alpha) - f(\beta) - f'(\beta)(\alpha - \beta) &= \alpha^2 - \beta^2 - 2\beta(\alpha - \beta) \\ &= \alpha^2 - 2\beta\alpha + \beta^2 \\ &= (\alpha - \beta)^2 \\ &\geq 0. \end{aligned}$$

Hence if we know that f is a convex function then any point satisfying (1) is a global minimum. This follows easily from (2) above. Consequently (1) is both necessary and sufficient for global optimality for a differentiable convex function.

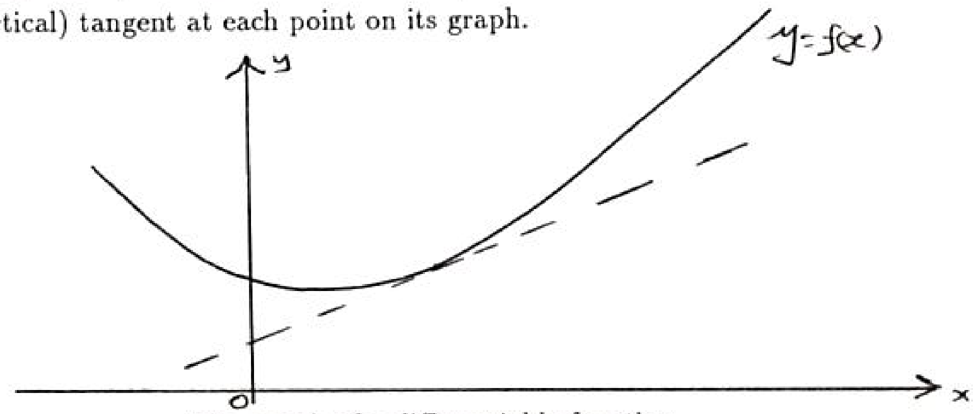
Exercise 2 By considering simple examples (such as x^2, x^4, e^x, $-\ln x$ etc) see if you can determine second order properties of convex functions (that is what can we say about $f''(x)$ if f is convex).

We should note that if (1) holds and $f''(a) > 0$ then a is a local minimum. This provides us with a second order optimality condition.

The discussion to date has focussed on differentiable functions. Recall that a function $y = f(x)$ is differentiable at $x = a$ if we can find the limit of

$$\frac{f(a+h) - f(a)}{h} \tag{3}$$

as h approaches zero. We say that f is differentiable if it is differentiable every where. We often refer to differentiable functions as *smooth* functions. This follows from the geometry of calculus. That is, if f is differentiable then it has a unique (non-vertical) tangent at each point on its graph.



The graph of a differentiable function

Thus the graph of a differentiable function appears *smooth*, that is it has no 'sharp' points. However in many applications we encounter functions which are not smooth. In the simplest cases their graphs contain one or more sharp points indicating point(s) of nondifferentiability. For example consider the absolute value function

$$f(x) = |x| = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0 \end{cases}$$

This function has no derivative at $x = 0$. It is, however, continuous (i.e no jumps or missing points in its graph), convex and has a global minimum at $x = 0$. The question we are now interested in is whether there is an analogue of (1) for functions such as the absolute value function. In particular is there an analogue for non-smooth convex functions? The answer is yes. Before we discuss this further we need some more basic information about convex functions.

Theorem 1 If $y = f(x)$ is a convex function then the following are valid:

(i) f is continuous,

(ii) f possesses *one-sided* derivatives at each point; that is, for any a we can calculate the following limits:

$$f'_+(a) = \lim_{h \to 0^+} \frac{f(a+h) - f(a)}{h},$$

and

$$f'_-(a) = \lim_{h \to 0^-} \frac{f(a+h) - f(a)}{h}.$$

Here $h \to 0^+$ means that we are only concerned with the limit as h approaches zero from the positive side, whereas $h \to 0^-$ refers to the limit as h approaches zero from the negative side.

Exercise 3 Show that for the absolute value function $f(x) = |x|$, $f'_+(0) = 1$ and $f'_-(0) = -1$. Discuss what this means geometrically in terms of tangents to the graph of f near zero.

We should note that if a function f actually has a derivative at a point a then the two one-sided derivatives coincide at that point. However for a convex function at a point of nondifferentiability these one-sided derivatives exist and are distinct. We now define a set, known as the *subdifferential*, which will replace the non-existent derivative at points of nondifferentiability for a convex function. The notation for the subdifferential of a convex function f at a point a is $\partial f(a)$ and it is defined as the interval

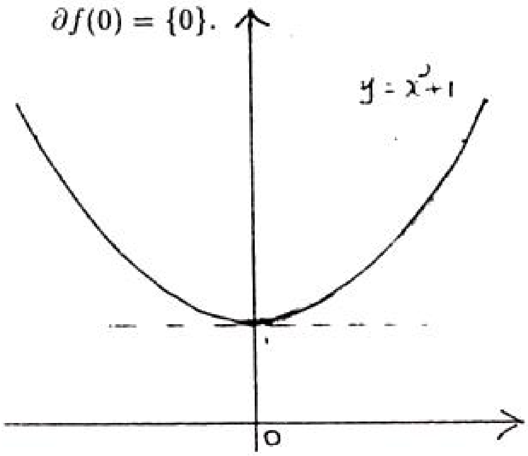
$$\partial f(a) = [f'_-(a), f'_+(a)].$$

Note here that the interval $[c, d]$ is defined as the collection of numbers between c and d, where $c \leq d$. The subdifferential has also the following equivalent form:

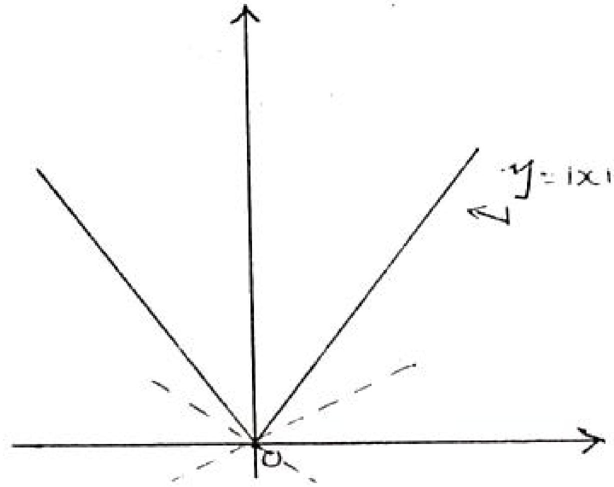
$$\partial f(a) = \{\alpha : f(x) - f(a) \geq \alpha(x - a), \text{ for all } x \}.$$

This is a fairly complicated definition and we need to have some geometric interpretation of it. So, for the absolute value function $f(x) = |x|$, $\partial f(0) = [-1, 1]$. This, intuitively, coincides with the gradients of the 'tangents' which can be drawn to the graph of f at the point $(0, 0)$. Thus at any point at which a convex function f actually has derivative the subdifferential reduces to a set with one element - the derivative at that point. At other points the subdifferential is an interval whose end points are the one-sided derivatives at that point. For instance, $f(x) = x^2 + 1$ has

$\partial f(0) = \{0\}$.



Unique tangent



Non-unique tangents

Exercise 4 Calculate the subdifferential at the origin for each of the following convex functions:

(i) $f(x) = \begin{cases} 2x & \text{if } x \geq 0 \\ x & \text{if } x < 0 \end{cases}$

(ii) $f(x) = \begin{cases} (x+1)^2 & \text{if } x \geq 0 \\ -x+1 & \text{if } x < 0 \end{cases}$

Now we return to consider optimality conditions. Recall that for the absolute value function $f(x) = |x|$ we have $\partial f(0) = [-1, 1]$ and $x = 0$ is a global minimum. From this very simple example we note that $x = 0$ is a minimum and 0 is inside $\partial f(0)$. It can be shown that the analogue of (1) for convex functions is that if a is a minimum of a convex function f then 0 is inside the $\partial f(a)$. In fact this condition is both necessary and sufficient for a to be a global optimum for convex functions. Test the functions in exercise 4 above to show that this condition is indeed satisfied at any minimum point.

Example Find the line of 'best' fit passing through the points $(1,1)$, $(2,0)$, $(3,2)$. Before we can begin to consider a solution to this problem we need to decide what exactly is meant by 'best fit'. This problem of finding a straight line which in some sense provides a best possible fit to a set of data points is extremely important in many areas of mathematical modelling and statistics. The favoured method of defining the line of 'best' fit is to construct the least-squares line of best fit. In this case we let our line have equation $y = \alpha + \beta x$, where α and β are to be determined so that the sum of the distances from the data points to the line is a minimum. This means we wish to find α and β which minimize the following function:

$$f_1(\alpha, \beta) = (\alpha + \beta - 1)^2 + (\alpha + 2\beta)^2 + (\alpha + 3\beta - 2)^2.$$

This is a function of two variables which is differentiable everywhere (one of the reasons for using this approach) and it can easily be shown that f is convex and attains its minimum at the point $(0, 0.5)$. Thus the least-squares line of 'best' fit is $y = 0.5x$. Plot the points and the line to verify this seems a reasonable solution.

However there are many ways to define 'best' for our straight line. Another possible definition is to choose α and β so that the sum of the vertical deviations from the data points to the line are a minimum. In this case we need to find α and β which minimize:

$$f_2(\alpha, \beta) = |\alpha + \beta - 1| + |\alpha + 2\beta| + |\alpha + 3\beta - 2|.$$

This is also a function of two variables and it is also convex but it does not have derivatives everywhere. If you have access to computer graphing software (such as Maple or Derive) it is helpful to plot the graph of this function (a 3-dimensional plot of course). It can be shown that f_2 reaches a minimum at $(0.5, 0.5)$ and that the subdifferential of f_2 is given by:

$$\partial f_2(0.5, 0.5) = \{(u + v + 1, u + 3v + 2) : u, v \in [-1, 1]\}.$$

It is not difficult to see (by solving a pair of simultaneous equations) that $(0, 0) \in \partial f_2(0.5, 0.5)$. Thus our necessary and sufficient condition for optimality is satisfied so that $(0.5, 0.5)$ is indeed the minimum. But you should notice that $y = 0.5 + 0.5x$ is a different line of 'best' fit to that obtained by the least-squares method. For certain practical problems of data analysis the second method discussed above provides a more reasonable line of best fit than the first. The difficulty with the second method is that the function to be minimized is not differentiable and so it is inherently more difficult to locate the optimum.

We have briefly discussed an important class of functions which arises frequently in optimization. In particular we have endeavoured to illustrate that a lack of differentiability is not always an impediment to developing verifiable optimality conditions. Essentially we use a set of gradients to replace the non-existent derivative at points of nondifferentiability. Such a set always exists for convex functions and standard first order necessary conditions such as (1) have a natural analogue for these functions. This discussion just touches upon some of the concepts that form the basis for a branch of mathematics known as ***nonsmooth calculus***, which has received considerable attention over the past 20 years.

On October 25, 1994, the following announcement was circulated on the Internet math. announce newsgroup, by Karl Rubin of Harvard University.

As of this morning, two manuscripts have been released

Modular elliptic curves and Fermat's Last Theorem, (by Andrew Wiles)

Ring theoretic properties of certain Hecke algebras, (by Richard Taylor and Andrew Wiles)

The first one (long) announces a proof of, among other things, Fermat's Last Theorem, relying on the second one (short) for one crucial step.

As most of you know, the argument described by Wiles in his Cambridge lectures turned out to have a serious gap, namely the construction of an Euler system. After trying unsuccessfully to repair that construction, Wiles went back to a different approach, which he had tried earlier but abandoned in favor of the Euler system idea. He was able to complete his proof, under the hypothesis that certain Hecke algebras are local complete intersections. This and the rest of the ideas described in Wiles' Cambridge lectures are written up in the first manuscript. Jointly, Taylor and Wiles establish the necessary property of the Hecke algebras in the second paper.

The overall outline of the argument is similar to the one Wiles described in Cambridge. The new approach turns out to be significantly simpler and shorter than the original one, because of the removal of the Euler system. (In fact, after seeing these manuscripts Faltings has apparently come up with a further significant simplification of that part of the argument.)

Versions of these manuscripts have been in the hands of a small number of people for (in some cases) a few weeks. While it is wise to be cautious for a little while longer, there is certainly reason for optimism.