Real Analysis Exchange
Vol. 29(1), 2003/2004, pp. 465–471

J. J. Koliha, Department of Mathematics and Statistics, University of Melbourne, VIC 3010, Australia. e-mail: `j.koliha@ms.unimelb.edu.au`

APPROXIMATION OF CONVEX FUNCTIONS

Abstract

In this note we give an elementary proof that an arbitrary convex function can be uniformly approximated by a convex C^∞-function on any closed bounded subinterval of the domain. An interesting byproduct of our proof is a global equation for a polygonal (piecewise affine) function.

1 Convex functions

A real valued, not necessarily differentiable, function f defined on an open interval I is *convex* if for any two points u, v in the domain of f the portion of the graph between u and v lies below the corresponding secant. Formally, for any $u, v \in I$,

$$f((1-t)u+tv) \le (1-t)f(u)+tf(v) \quad \text{for all} \quad t \in [0,1]. \tag{1.1}$$

Geometrically (1.1) means that the slopes of secants of f increase in the following sense: If $x_1 < x_2 < x_3$, then according to [1] or [4],

$$\frac{f(x_2)-f(x_1)}{x_2-x_1} \le \frac{f(x_3)-f(x_1)}{x_3-x_1} \le \frac{f(x_3)-f(x_2)}{x_3-x_2}.$$

The function f is *strictly convex* if the inequality (1.1) is strict for all $t \in (0,1)$, and *affine* if the inequality (1.1) is always an equality. (It is well known that the general form of an affine function g is $g(x) = \alpha x + \beta$ for some constants $\alpha, \beta \in \mathbb{R}$.) A convex function f is continuous on I. Hence on each compact subinterval $[a, b]$ of I, f can be uniformly approximated by a C^∞-function. An extra effort must go into ensuring that we can choose a *convex* approximating function on $[a, b]$.

Key Words: convex function, convex C^∞-function, approximation
Mathematical Reviews subject classification: 26A51, 41A30
Received by the editors December 30, 2002
Communicated by: Peter Bullen

It is well known that such an approximation is possible (see Ghomi [2]), based on a *convolution* of f with an infinitely smooth nonnegative function θ. This approach requires the knowledge of Lebesgue integral and convolutions, and some sophisticated mathematical reasoning.

The aim of the present paper is to give an elementary proof of the approximation of convex functions by convex C^∞-functions, requiring only first year calculus and linear algebra. Details of elementary properties of convex functions needed in this note can be found in [1] or in the monograph [4] by Roberts and Varberg.

First we look at one of the simplest convex functions, $f(x) = |x|$, defined on $\mathbb{R}$. To approximate f by a differentiable convex function g, we may try to round off the cusp of $|x|$ by a suitable circular arc touching $|x|$; this leads to a once differentiable convex function. To improve on differentiability, we may round off the cusp with an arc of a sinusoid touching $|x|$ whose tangent points are also inflection points, this time obtaining twice differentiable function.

Our ultimate objective is attained when we use a branch of a hyperbola which has $|x|$ for asymptotes.

Lemma 1. *Let $\varepsilon > 0$ and let Φ_ε be defined by*

$$\Phi_\varepsilon(x) = \sqrt{x^2+\varepsilon^2}, \quad x \in \mathbb{R}. \tag{1.2}$$

Then Φ_ε is a strictly convex C^∞-function satisfying

$$\big| |x| - \Phi(x) \big| \le \varepsilon \quad \text{for all } x \in \mathbb{R}. \tag{1.3}$$

Proof. For any $x \in \mathbb{R}$ and any $\varepsilon > 0$,

$$0 \le \sqrt{x^2+\varepsilon^2} - |x| = \frac{\varepsilon^2}{\sqrt{x^2+\varepsilon^2}+|x|} \le \frac{\varepsilon^2}{\varepsilon} = \varepsilon.$$

Figure 1

The strict convexity of Φ_ε follows from the positivity of the second derivative,

$$\Phi_\varepsilon''(x) = \frac{\varepsilon^2}{(x^2+\varepsilon^2)^{3/2}} > 0.$$

By induction we find that, for any $n \in \mathbb{N}$,

$$\Phi_\varepsilon^{(n)}(x) = \frac{q_n(x)}{(x^2+\varepsilon^2)^{n-1/2}},$$

where $q_n(x)$ is a polynomial in x of degree $n-2$; hence Φ_ε is a C^∞-function. $\square$

2 Polygonal functions

Since we know how to approximate $|x|$ and, more generally $|x-a|$, by C^∞-functions, we also know how to approximate linear combinations of such functions. It turns out that every polygonal function can be expressed as a linear combination of functions of the form $|x-a_k|$.

A function P defined on a compact interval $[a,b]$ is *polygonal* (or *piecewise affine*) if it is continuous and if there is a partition $\mathcal{Q}$ given by

$$a = a_0 < a_1 < \cdots < a_{n-1} < a_n = b \tag{2.1}$$

such that $P(x)$ is affine on each subinterval $[a_{k-1}, a_k]$, $k = 1, \ldots, n$. The points $(a_k, P(a_k)) \in \mathbb{R}^2$ are the *vertices* of P. We define the norm of $\mathcal{Q}$ by $\|\mathcal{Q}\| = \max_{i=1,\ldots,n}(a_i - a_{i-1})$. A polygonal function P is uniquely identified by its vertices (a_0, b_0), (a_1, b_1), …, (a_n, b_n). We now derive a global representation for polygonal functions.

Theorem 1. *The polygonal function P with the vertices (a_0, b_0), …, (a_n, b_n) based on a partition* (2.1) *has a unique representation of the form*

$$P(x) = \sum_{k=0}^{n} p_k |x - a_k| = (p_0 - p_n)x + (p_n a_n - p_0 a_0) + \sum_{k=1}^{n-1} p_k |x - a_k| \tag{2.2}$$

for all $x \in [a,b]$. The coefficients p_k are given by

$$p_k = \tfrac{1}{2}(s_{k+1} - s_k), \quad k = 0, 1, \ldots, n, \tag{2.3}$$

where

$$s_{n+1} = \frac{b_n + b_0}{a_n - a_0} = -s_0, \quad s_k = \frac{b_k - b_{k-1}}{a_k - a_{k-1}},\ k = 1, \ldots, n. \tag{2.4}$$

Proof. Let us define $R(x) = \sum_{k=0}^{n} p_k |x - a_k|$ with some real coefficients p_k. Then R is continuous and affine on each subinterval $[a_{k-1}, a_k]$ $(k = 1, \ldots, n)$; hence R is a polygonal function with the vertices $(a_k, R(a_k))$, $k = 0, \ldots, n$. We determine the coefficients p_k to satisfy

$$R(a_k) = b_k, \quad k = 0, \ldots, n.$$

This is a linear system which can be written in matrix form as

$$\boldsymbol{A}\boldsymbol{p} = \boldsymbol{b},$$

where

$$\boldsymbol{A}=\begin{bmatrix} 0 & a_1-a_0 & a_2-a_0 & \dots & a_n-a_0\\ a_1-a_0 & 0 & a_2-a_1 & \dots & a_n-a_1\\ a_2-a_0 & a_2-a_1 & 0 & \dots & a_n-a_2\\ \dots & \dots & \dots & \dots & \dots\\ a_n-a_0 & a_n-a_1 & a_n-a_2 & \dots & 0\end{bmatrix},\ \boldsymbol{b}=\begin{bmatrix} b_0\\ b_1\\ b_2\\ \dots\\ b_n\end{bmatrix},\ \boldsymbol{p}=\begin{bmatrix} p_0\\ p_1\\ p_2\\ \dots\\ p_n\end{bmatrix}.$$

It is not difficult to find the inverse of $\boldsymbol{A}$. Each row of the symmetric matrix $\boldsymbol{A}^{-1}$ has only three nonzero entries,

$$\boldsymbol{A}^{-1}=\begin{bmatrix} c_{00} & c_{01} & 0 & 0 & \dots & 0 & 0 & c_{0n}\\ c_{01} & c_{11} & c_{12} & 0 & \cdots & 0 & 0 & 0\\ 0 & c_{12} & c_{22} & c_{23} & \dots & 0 & 0 & 0\\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots\\ 0 & 0 & 0 & \dots & 0 & c_{n-2,n-1} & c_{n-1,n-1} & c_{n-1,n}\\ c_{0n} & 0 & 0 & \dots & 0 & 0 & c_{n-1,n} & c_{nn}\end{bmatrix},$$

where

$$c_{00}=\frac{1}{2}\left(\frac{1}{a_n-a_0}-\frac{1}{a_1-a_0}\right),\quad c_{01}=\frac{1}{2}\frac{1}{a_1-a_0},\quad c_{0n}=\frac{1}{2}\frac{1}{a_n-a_0},$$
$$c_{k,k-1}=\frac{1}{2}\frac{1}{a_k-a_{k-1}},\quad c_{kk}=-\frac{1}{2}\left(\frac{1}{a_{k+1}-a_k}+\frac{1}{a_k-a_{k-1}}\right),$$
$$c_{k,k+1}=\frac{1}{2}\frac{1}{a_{k+1}-a_k},\qquad k=1,\dots,n-1,$$
$$c_{n-1,n}=\frac{1}{2}\frac{1}{a_n-a_{n-1}},\quad c_{nn}=\frac{1}{2}\left(\frac{1}{a_n-a_0}-\frac{1}{a_n-a_{n-1}}\right).$$

This can be checked by matrix multiplication. Then $\boldsymbol{p}=\boldsymbol{A}^{-1}\boldsymbol{b}$. A straightforward calculation shows that the coordinates of $\boldsymbol{p}$ are given by (2.3) and (2.4). □

Remark. In numerical mathematics, polygonal functions are regarded as *linear splines* (see, for instance, [3, Chapter 6]). Our construction amounts to showing that the vector space of all linear splines based on the partition $\mathcal{Q}$ defined by (2.1) has a basis consisting of the functions $\alpha+\beta x$, $|x-a_1|,\dots,$ $|x-a_{n-1}|$, while giving an explicit expression for the coordinates relative to this basis. An alternative basis, given in [3, p. 231], consists of the functions $\alpha+\beta x$, $(x-a_1)^+,\dots,$ $(x-a_{n-1})^+$. We can recover the results in [3] for linear splines using the relation $u^+=\frac{1}{2}(|u|+u)$.

3 Approximation of convex functions

We can now present our main theorem.

Theorem 2. *Let f be a function convex on an open interval I. For any subinterval $[a, b]$ of I and any $\varepsilon > 0$ there exists a convex C^∞-function h on $[a, b]$ such that $|f(x) - h(x)| \leq \varepsilon$ for all $x \in [a, b]$.*

Proof. Let $[a, b] \subset I$ and let $\varepsilon > 0$ be given. Then f is uniformly continuous on $[a, b]$, and there exists $\delta > 0$ such that for any $x_1, x_2 \in [a, b]$ satisfying $|x_1 - x_2| < \delta$ we have $|f(x_1) - f(x_2)| \leq \frac{1}{2}\varepsilon$. Let $\mathcal{Q}$ be a partition of $[a, b]$ defined by (2.1), with $\|\mathcal{Q}\| < \frac{1}{2}\delta$. Let P be the polygonal function on $[a, b]$ with the vertices $(a_0, b_0), \ldots, (a_n, b_n)$, where $b_k = f(a_k)$ $(k = 0, \ldots, n)$. We observe that

$$|f(x) - P(x)| \leq \tfrac{1}{2}\varepsilon \quad \text{for all } x \in [a, b]. \tag{3.1}$$

Since f is convex, the slopes $s_k = (b_k - b_{k-1})/(a_k - a_{k-1})$ are increasing with k $(k = 1, \ldots, n-1)$. Hence P is of the form (2.2) with $p_k = \frac{1}{2}(s_{k+1} - s_k) \geq 0$ for $k = 1, \ldots, n-1$. Let $\rho = \sum_{k=1}^{n-1} p_k$; we may assume that $\rho > 0$, otherwise P is affine. Let $\Phi = \Phi_{\varepsilon/2\rho}$ and define

$$h(x) = (p_0 - p_n)x + (p_n a_n - p_0 a_0) + \sum_{k=1}^{n-1} p_k \Phi(x - a_k).$$

Then h is a C^∞-function whose second derivative is nonnegative in view of Lemma 1:

$$h''(x) = \sum_{k=1}^{n-1} p_k \Phi''(x - a_k) \geq 0 \quad \text{for all } x \in [a, b];$$

hence, h is convex. Further, by Lemma 1,

$$|P(x) - h(x)| \leq \sum_{k=1}^{n-1} p_k \big| |x - a_k| - \Phi(x - a_k) \big| \leq \rho \frac{\varepsilon}{2\rho} = \frac{1}{2}\varepsilon. \tag{3.2}$$

Combining (3.1) and (3.2), we obtain the result. □

4 Example

We give an example of the approximation process described in our main theorem. Given the interval $[-3, 3]$, we choose the partition $\mathcal{Q}$ as

$$-3 < -\tfrac{3}{2} < -\tfrac{1}{2} < 0 < 1 < \tfrac{5}{2} < 3.$$

The polygonal function $P(x)$ based on this partition with vertices on the graph of the convex function $f(x) = x^2$ is given by

$$P(x) = \tfrac{1}{2}x - 6 + \tfrac{5}{4}\big|x + \tfrac{3}{2}\big| + \tfrac{3}{4}\big|x + \tfrac{1}{2}\big| + \tfrac{3}{4}\big|x\big| + \tfrac{5}{4}\big|x - 1\big| + \big|x - \tfrac{5}{2}\big|$$

(Figure 2 on the left.) The smoothing function h with $\varepsilon^2 = 0.05$ (Figure 2 on the right) is given by

$$h(x) = \tfrac{1}{2}x - 6 + \tfrac{5}{4}\Phi\big(x + \tfrac{3}{2}\big) + \tfrac{3}{4}\Phi\big(x + \tfrac{1}{2}\big) + \tfrac{3}{4}\Phi(x) + \tfrac{5}{4}\Phi(x - 1) + \Phi\big(x - \tfrac{5}{2}\big),$$

where $\Phi(x) = \sqrt{x^2 + 0.05}$. Even with this rather coarse partition we get a reasonable resemblance of the function $f(x) = x^2$.

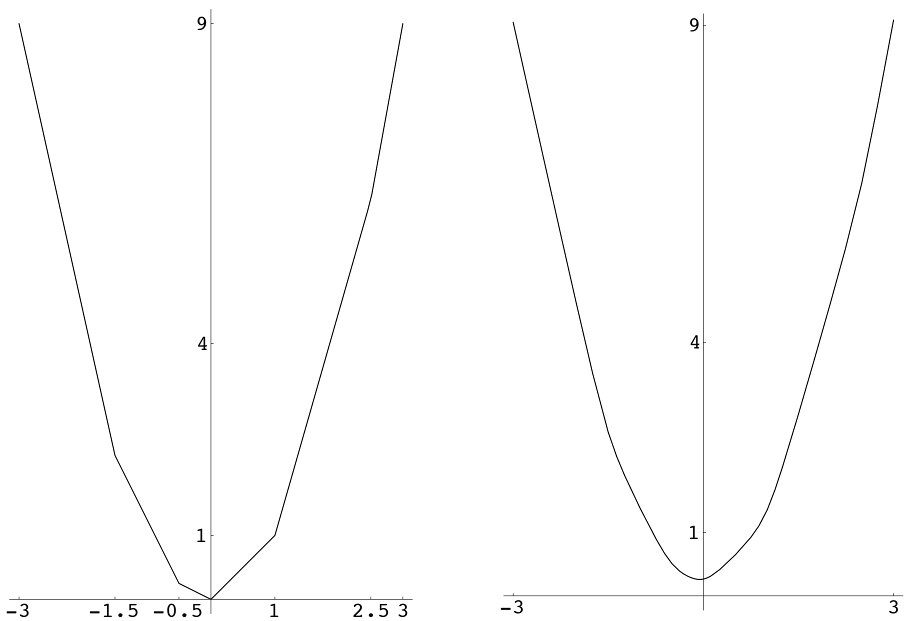

Figure 2

Acknowledgement. The author is indebted to Don Handley for pointing to asymptotes in connection with the approximation of $|x|$, and to an anonymous referee for drawing his attention to Ghomi's paper.

References

[1] N. Bourbaki, *Éléments de mathématique. I: Les structures fondamentales de l'analyse. Livre IV: Fonctions d'une variable réelle* (*théorie élémentaire*), Hermann, Paris 1958.

[2] Mohammad Ghomi, The problem of optimal smoothing for convex functions, *Proc. Amer. Math. Soc.* **130** (2002), 2255-2259.

[3] G. Hämmerlin and K.-H. Hoffmann, *Numerical Mathematics*, Springer, New York, 1991.

[4] A. W. Roberts and D. E. Varberg, *Convex functions*, Academic Press, New York-London, 1973.