J Glob Optim (2010) 46:155–161
DOI 10.1007/s10898-009-9457-4

Characterizing zero-derivative points

Sanjo Zlobec

Received: 27 January 2009 / Accepted: 16 June 2009 / Published online: 2 July 2009

Abstract We study smooth functions in several variables with a Lipschitz derivative. It is shown that these functions have the "envelope property": Around zero-derivative points, and only around such points, the functions are envelopes of a quadratic parabolloid. The property is used to reformulate Fermat's extreme value theorem and the theorem of Lagrange under slightly more restrictive assumptions but without the derivatives.

Keywords Zero-derivative point · Fermat's extreme value theorem · Theorem of Lagrange

Mathematics Subject Classification (2000) 26B05 · 90C30

1 Introduction

The points where functions have zero derivatives are important in many areas of mathematics. In particular, a classic result in single-variable calculus says that local extrema of differentiable functions can be achieved only at zero-derivative points. This result is often referred to as Fermat's extreme value theorem, e.g., [4] and it extends to functions of several variables and beyond, e.g., [3]. This note shows that around the zero-derivative points, and only around these points, smooth functions with a Lipschitz derivative are envelopes of quadratic parabolloids. In principle, one can use this "envelope property" to determine whether the derivative is zero without calculating the derivative and checking its roots. The property also shows that convergent methods for calculating zero-derivative points of smooth, generally non-convex, functions with Lipschitz derivatives have a quadratic rate of weak convergence.

The proof of the envelope property uses a result from [5] which says that every smooth function with a Lipschitz derivative, when considered on a compact convex set, is the difference of a convex function and a convex quadratic function. In Sect. 2 it is shown that an

S. Zlobec (✉)
Department of Mathematics and Statistics, McGill University, Burnside Hall, 805 Sherbrooke Street West, H3A 2K6 Montreal, QC, Canada
e-mail: zlobec@math.mcgill.ca

analogous result holds for concave functions. The behaviour of functions around zero-derivative points is studied in Sect. 3. Optimality conditions for a local extremum, that are based on the envelope property, are given in Sect. 4, where they are compared with Fermat's theorem and the theorem of Lagrange. A possible modification of the αBB algorithm [1] in global optimization, using the convexifier from Sect. 2, is suggested in Sect. 5.

2 Decomposition of smooth functions

We use the terminology and notation from [5]. Consider a smooth, i.e., continuously (Fréchet) differentiable, function in several variables $f\colon R^n \to R$. Suppose that the derivative of f at x, represented by the row gradient $\nabla f(x)$, has the Lipschitz property on a compact convex set K in R^n. This means that there exists a constant L, called a Lipschitz constant of the derivative, such that

$$\|\nabla^{\mathrm{T}} f(x) - \nabla^{\mathrm{T}} f(y)\| \leq L\|x - y\|$$

for every x and y in K. Here $\nabla^{\mathrm{T}} f(x)$ denotes the transpose of $\nabla f(x)$ and the vector norm is chosen to be Euclidean. Note that L depends on K.

Definition 1 Consider a continuous function $f\colon R^n \to R$ on a compact convex set K of its open domain. If there exists a number α such that $f(x) = C(\alpha, x) - q(\alpha, x)$, where $C(\alpha, x)$ is a convex function and $q(\alpha, x) = -1/2\alpha x^{\mathrm{T}} x$ is a convex quadratic function on K, then we say that f is convexifiable, that α is a convexifier, and that C(α, x) is a convexification of f on K. If there exists a number β such that $f(x) = \tilde{C}(\beta, x) - \tilde{q}(\beta, x)$, where $\tilde{C}(\beta, x)$ is a concave function and $\tilde{q}(\beta, x) = -1/2\beta x^{\mathrm{T}} x$ is a concave quadratic function on K then f is concavifiable, β is a concavifier, and $\tilde{C}(\beta, x)$ is a concavification of f on K.

A continuous function can be convexifiable but not concavifiable and vice versa.

Example 1 Scalar function $f(x) = -|x|$ is concavifiable (being a concave function) but it is not convexifiable on the interval $I = [-1, 1]$.

The derivatives of smooth functions may or may not have the Lipschitz property. If f is smooth and if its derivative does have the Lipschitz property, then f is both convexifiable and concavifiable.

Theorem 1 (Decomposition of Smooth Functions with a Lipschitz Derivative) *Consider a smooth function f: $R^n \to R$ on a compact convex set K in its open domain. If the derivative of f has the Lipschitz property on K then f is both convexifiable and concavifiable on K. Moreover, if L is a Lipschitz constant of the derivative of f on K, then any $\alpha \leq -L$ is a convexifier and any $\beta \geq L$ is a concavifier.*

Proof The convexification part was proved in [5]. In order to prove concavification, take $x, y \in K, x \neq y$. Using the inner product notation and the Cauchy-Schwartz inequality we have

$$\begin{aligned} |(\nabla^{\mathrm{T}} f(x) - \nabla^{\mathrm{T}} f(y), x - y)| &\leq \|\nabla^{\mathrm{T}} f(x) - \nabla^{\mathrm{T}} f(y)\| \cdot \|x - y\| \\ &\leq L\|x - y\|^2 \end{aligned}$$

by the Lipschitz property. Hence

$$|(\nabla^{\mathrm{T}} f(x) - \nabla^{\mathrm{T}} f(y), x - y)/\|x - y\|^2| \leq L.$$

This implies

$$-L \le (\nabla^{\mathrm{T}} f(x) - \nabla^{\mathrm{T}} f(y), x - y)/\|x - y\|^2 \le L.$$

Hence, for every $\beta \ge L$

$$\beta \ge (\nabla^{\mathrm{T}} f(x) - \nabla^{\mathrm{T}} f(y), x - y)/\|x - y\|^2$$

and then

$$\beta\|x - y\|^2 \ge (\nabla^{\mathrm{T}} f(x) - \nabla^{\mathrm{T}} f(y), x - y).$$

This is the same as

$$(\nabla^{\mathrm{T}} f(x) - \nabla^{\mathrm{T}} f(y) - \beta (x - y), x - y) \le 0.$$

The derivative of $\tilde{C}(\beta, x) = f(x) - 1/2\, \beta x^{\mathrm{T}} x$ is $\nabla\tilde{C}(\beta, x) = \nabla f(x) - \beta x^{\mathrm{T}}$. Hence

$$(\nabla^{\mathrm{T}}\tilde{C}(\beta, x) - \nabla^{\mathrm{T}}\tilde{C}(\beta, y), x - y) \le 0.$$

The inequality holds also if $x = y \in K$, so it holds for every x and y in K. This means that $\tilde{C}$ is concave, i.e., $\tilde{C}$ is a concavification of f on K. □

3 Characterization of zero-derivative points

Definition 2 Consider a smooth function *f*: $R^{\mathrm{n}} \to R$ on a compact convex set K with interior points. We say that *f* has the envelope property at an interior point x^* of K if there is a constant $\Lambda \ge 0$ such that

$$\left|f\left(x^*\right) - f(x)\right| \le \Lambda\|x^* - x\|^2 \quad \text{for every } x \text{ in } K.$$

Example 2 Function $f(x) = x^3$, considered on the interval $I = [-2, 2]$, has the envelope property at $x^* = 0$. Indeed, for any constant $\Lambda \ge 2$, $|f(x^*) - f(x)| = \left|x^3\right| \le \Lambda x^2$ on I. The function does not have the envelope property at $y^* = 1$ because there is no $\Lambda \ge 0$ such that $\left|1 - y^3\right| \le \Lambda (1 - y)^2$ for every y on I.

The main result of the paper claims that for smooth functions with a Lipschitz derivative, the envelope property is equivalent to the zero-derivative property:

Theorem 2 (Characterization of Zero-Derivative Points) *Consider a smooth function f: $R^n \to R$ on a compact convex set K with interior points. Assume that the derivative of f has a Lipschitz derivative with a constant L on K. Then f has the envelope property at an arbitrary interior point x^* of K if, and only if,* $\nabla f(x^*) = 0$. *In particular, one can specify* $\Lambda = 1/2L$.

Proof First we will show that $\nabla f(x^*) = 0$ implies the envelope property. Since f is convexifiable, it can be decomposed: $f(x) = 1/2\alpha\|x\|^2 + C(\alpha, x)$, where $C(\alpha, \mathrm{x})$ is a convex function in x and α is a convexifier of *f* on K. By convexity of C, it follows that

$$C\left(\alpha, \lambda x + (1 - \lambda) x^*\right) \le \lambda\, C(\alpha, x) + (1 - \lambda)\, C\left(\alpha, x^*\right)$$

for every x in K and $0 \le \lambda \le 1$. Hence, after substitution

$$\begin{aligned} f\left(\lambda x + (1 - \lambda) x^*\right) \le\ & 1/2\alpha\|\lambda x + (1 - \lambda) x^*\|^2 + \lambda f(x) - 1/2\alpha\lambda\|x\|^2 \\ & + (1 - \lambda) f\left(x^*\right) - 1/2\alpha (1 - \lambda) \|x^*\|^2 \end{aligned}$$

and using properties of the norm

$$\left[f\left(x^*+\lambda\left(x-x^*\right)\right)-f\left(x^*\right)\right]/\lambda \le f(x)-f\left(x^*\right)-1/2\alpha\,(1-\lambda)\,\|x-x^*\|^2.$$

Sending $\lambda \to 0$ yields

$$\left[\nabla f\left(x^*\right)\right]\left(x-x^*\right) \le f(x)-f\left(x^*\right)-1/2\alpha\|x-x^*\|^2.$$

When $\nabla f(x^*) = 0$ then $f(x^*) - f(x) \le -1/2\alpha\|x - x^*\|^2$. Similarly, since f is smooth with a Lipschitz derivative on K, we know that $\tilde{C}\,(\beta, x) = f(x) - 1/2\beta\|x\|^2$, where $\tilde{C}\,(\beta, x)$ is a concave function for every concavifier β. This implies $f(x^*) - f(x) \ge -1/2\beta\|x - x^*\|^2$. One can specify $\alpha = -L$ and $\beta = L$, where L is a Lipschitz constant of the derivative, by Theorem 1. Therefore

$$\left|f\left(x^*\right)-f(x)\right| \le 1/2L\|x-x^*\|^2.$$

After the identification $\Lambda = 1/2L$ we conclude that f has the envelope property at x^*. Let us show that the envelope property at an interior point x^* implies $\nabla f(x^*) = 0$. First, for every $x \in K, x \ne x^*$, after division by $\|x - x^*\|$, $|f(x^*) - f(x)|/\|x - x^*\| \le \Lambda\|x - x^*\|$. Using the zero row 0 and the substitution $h = x - x^*$, this yields

$$(1/\|h\|)\left|f\left(x^*\right)-f\left(x^*+h\right)-0h\right| \le \Lambda\|h\|.$$

Hence, in the limit $h \to 0$, this implies $\nabla f(x^*) = 0$, by continuity of the norm. □

The envelope property has the geometric interpretation: Function $F = |f(x^*) - f(x)|$ describes a distance between $f(x^*)$ and $f(x)$. The graph of $P = \|x^* - x\|^2$ is a quadratic parabolloid with the vertex at x^*. Theorem 2 says that around the zero-derivative points x^*, and only around these points, the function $E = F/\Lambda$ is an envelope of P for every $\Lambda > 0$ sufficiently large. One can specify $\Lambda \ge 1/2L$, where L is a Lipschitz constant of the derivative of f on K. The essence of the property is depicted in Fig. 1 using MATLAB. The figure depicts the situation described in Example 3.

Example 3 Consider $f(x) = x^3$ around $x^* = 0$ on $I = [-2, 2]$. Here $F = |f(x^*) - f(x)| = |x^3|$ and $P = x^2$. After division by, e.g., $\Lambda = 12$, F "moves downwards" to $E = F/12$ which is an envelope of P. Equivalently, F is an envelope of $12P$.

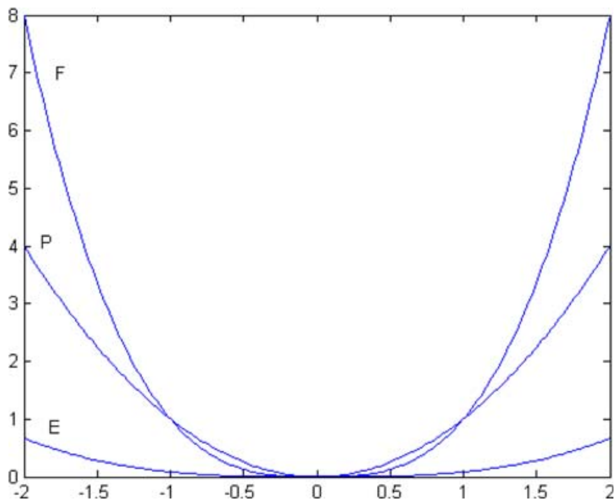

Fig. 1 The envelope property

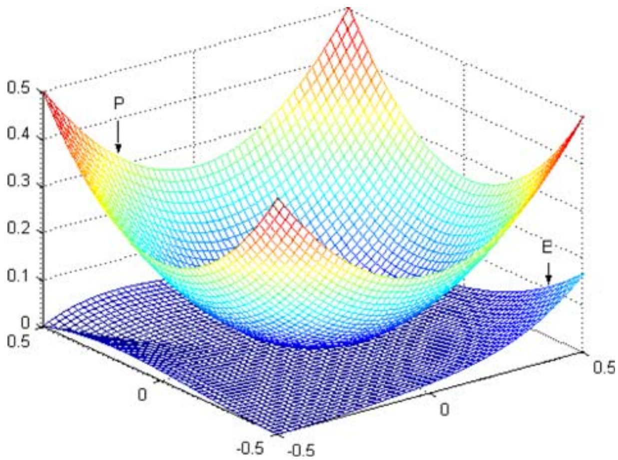

Fig. 2 The envelope criterion at a zero-derivative point

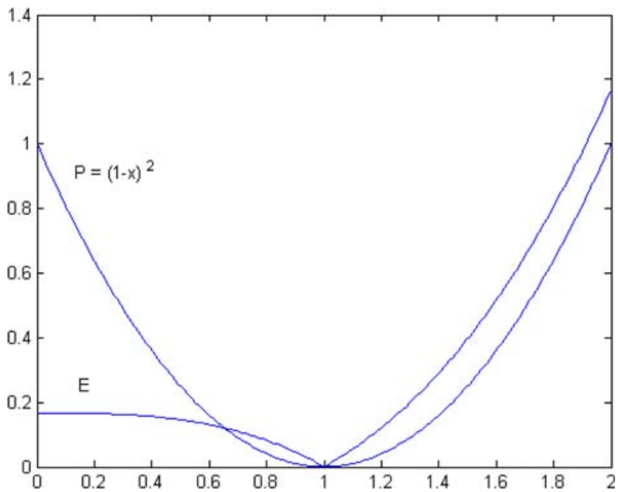

Fig. 3 Violation of the envelope criterion at a non-zero-derivative point

We refer to the "envelope property" as the "envelope criterion" when it is used for identification of zero-derivative points and local extrema.

Example 4 Consider $f(x) = \left(x_2 - {x_1}^2\right)\left(x_2 - 2{x_1}^2\right)$. Is $\nabla f(x^*) = 0$ at $x^* = 0$? Take, e.g., $-1 \le x_1, x_2 \le 1$ and $\Lambda = 6$. We have $E = \left|\left(x_2 - {x_1}^2\right)\left(x_2 - 2{x_1}^2\right)\right|/6$ and $P = {x_1}^2 + {x_2}^2$. The graphs of E and P are depicted in Fig. 2. The envelope criterion is satisfied and we can conclude that the origin is a zero-derivative point. (The origin is not a local extremum.)

Example 5 For $f(x) = x^3$ the envelope criterion is satisfied at $x^* = 0$ but not at $x^* = 1$. Hence $x^* = 0$ is a zero-derivative point but not $x^* = 1$. The violation is depicted in Fig. 3.

Suppose that one wants to calculate a zero-derivative point x^* of a smooth function $f : R^n \rightarrow R$ by a method that produces a convergent sequence of approximations

$x^k \to x^*, k = 0, 1, \ldots$ Then, from some $\tilde{k}$, i.e., for every $k \geq \tilde{k}$, we have $x^k \in K$, where K is some compact convex set containing x^* in its interior. If the derivative of f has a Lipschitz property on K, then Theorem 2 says that the rate of weak convergence on K is quadratic.

Example 6 Consider $f(x) = \cos x$ around its zero-derivative point $x^* = 0$. As $x^{\mathrm{k}} \to x^*$, $\cos x^{\mathrm{k}} \to 1$ with a quadratic rate of convergence. In fact $|1 - \cos x| \leq 1/2x^2$ for every x.

4 Envelope criteria for local extrema

Fermat's extremal value theorem says that $\nabla f(\mathrm{x}^*) = 0$ at a local extremum x^* of a differentiable function. We recall it for scalar functions.

Theorem 3 (Fermat's Theorem, [4, p.177].) *If f has a local extremum* (*that is, maximum or minimum*) *at c, and if $f'(c)$ exists, then $f'(c) = 0$.*

One can reformulate this result using Theorem 2 without the derivative but more assumptions on f around the local extremum are required:

Theorem 4 (Envelope Criterion for Local Extrema) *Consider a smooth function f on $I = [a, b]$ where it is assumed that the derivative of f has the Lipschitz property. If f has a local extremum at $c \in (a, b)$, then $|f(c) - f(x)| \leq \Lambda (c - x)^2$ for some $\Lambda \geq 0$ and every x in I.*

Example 7 Consider $f(x) = x^2 \cos x$. At $\mathrm{c} = 0$ the envelope criterion is $|\cos x| \leq \Lambda$, satisfied for, e.g., $\Lambda = 1$ and every x. At $c = \pi/2$, $\left|x^2 \cos x\right| / (\pi/2 - x)^2 \to \infty$ as $x \to \pi/2$. Hence $\mathrm{c} = 0$ can, but $c = \pi/2$ can not, be an extremal point.

The envelope criterion is applicable, in particular, to twice differentiable functions. Typically it is not required that one specifies the actual numerical value of Λ. The theorem holds for any Λ sufficiently large, e.g., $\Lambda \geq 1/2L$, where L is a Lipschitz constant of the derivative of f on K. However, the assumption that the derivative of f be Lipschitz can not be omitted as the following example shows.

Example 8 Function $f(x) = |x|^{3/2}$ is smooth but its derivative does not have the Lipschitz property on $I = [-1, 1]$. At $c = 0$ we have $|f(x)| / x^2 = 1/\sqrt{|x|} \to \infty$ as $x \to 0$ although $f'(0) = 0$.

The envelope property can be used to reformulate other results that use zero-derivative points, such as Rolle's theorem and the mean-value theorem. One can extend the above results to problems with constraints. In particular, consider the problem of optimizing a function on the set of points that are implicitly determined by a system of m equations in n variables:

$$\text{Opt } f^0(x), \text{ subject to } f^{\mathrm{i}}(x) = 0, \quad i \in P = \{1, \ldots, m\}. \tag{1}$$

The envelope criterion can be applied to the Lagrangian $L(x, \lambda) = \lambda_0 f^0(x) + \sum_{\mathrm{i} \in P} \lambda_{\mathrm{i}} f^{\mathrm{i}}(x)$ with some non-zero $(m+1)$-tuple $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_{\mathrm{m}})$. (If a "regularization condition" is satisfied, then one can specify $\lambda_0 = 1$.) At a local constrained extremum x^*, $\nabla L(x^*, \lambda) = 0$ by the theorem of Lagrange. Alternatively, Theorem 2 provides the following condition:

Theorem 5 (Envelope Criterion for Constrained Local Extrema) *Suppose that x^* is a constrained local extremum of* (1)*, where x^* is also an interior point of some arbitrary compact* *convex set K on which* $f^0, f^i, i \in P$ *are smooth functions with Lipschitz derivatives. Then there exists a non-zero* $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_m)$ *such that*

$$|\lambda_0 \left[f^0\left(x^*\right) - f^0\left(x\right)\right] - \sum_{i \in P} \lambda_i f^i\left(x\right)| \le \Lambda \|x^* - x\|^2$$

for some $\Lambda \ge 0$ *and every x in K.*

If the equality constraints "=0" are replaced by "≤0", then a non-zero and componentwise non-negative λ can be introduced in Theorem 5. This yields an analogue of the Fritz John condition. If the functions are convex, a slight modification of Theorem 5, pertaining to the complementarity condition, yields an analogue of the Karush–Kuhn–Tucker conditions. This is a possible direction for further research in the study of optimality in mathematical programming.

5 Note on the αBB algorithm

The αBB algorithm is well known for solving nonconvex programs with twice continuously differentiable functions, e.g., [1]. In the construction of convex underestimators the algorithm uses the convexifier

$$\lambda_{\min} = \min{}_{x \in K} \min{}_{i=1,\ldots,n} (\lambda_{\mathrm{i}} \left(H\left(x\right)\right)$$

where $\lambda_i \ H(x)$ is an eigenvalue of the Hessian matrix $H(x)$ of f at a given $x \in K$. This convexifier is regarded as a measure of nonconvexity of f and it is calculated over particular "intervals" K determined by a branch and bound approach. A method for calculating $\lambda_{\min}$ is given in [2] using "interval" Hessian matrices.

In our study of the envelope property we have used a different convexifier $\alpha = -L$ and a concavifier $\beta = L$, where L is a Lipschitz constant of the derivative of f on a compact convex set. The connection $\alpha + \beta = 0$ made it possible to formulate the property using the absolute value function. Unlike $\lambda_{\min}$, the convexifier α uses only the first-order information. The two are related by the relation $\alpha \le \lambda_{\min}$. It would be interesting to know whether $\lambda_{\min}$ can be "replaced" by α in the αBB algorithm. This may possibly lead to simpler calculations and broader applicability. Instead of programs with twice continuously differentiable functions, one could apply a modified algorithm to programs with once continuously differentiable functions with Lipschitz derivatives. Whatever the choice of convexifiers, the convergent methods retain their weakly quadratic rate of convergence.

Acknowledgments The author is indebted to the referees for their constructive comments. The research is partly supported by National Science and Engineering Research Council of Canada.

References

1. Androulakis, I.P., Maranas, C.D., Floudas, C.A.: αBB: A global optimization method for general constrained nonconvex problems. J. Glob. Optim. **7**, 337–363 (1995)
2. Adjiman, C.S., Floudas, C.A.: Rigorous convex underestimators for general twice-differentiable problems. J. Glob. Optim. **9**, 23–40 (1996)
3. Apostol, T. M.: Calculus, vol. II: Calculus of Several Variables with Applications to Probability and Vector Analysis. Blaisdell, New York (1965)
4. Stewart, J.: Single Variable Calculus. Brooks/Cole, Pacific Grove (1987)
5. Zlobec, S.: The fundamental theorem of calculus for Lipschitz functions. Math. Commun. **13**, 215–232 (2008)