Pure and Applied Functional Analysis
Volume 3, Number 4, 2018, 743–751

Yokohama Publishers
ISSN 2189-3764 ONLINE JOURNAL

GEOMETRY OF LOCAL OPTIMA

SANJO ZLOBEC

Abstract. We use the quadratic envelope characterization of zero-derivative points to introduce primal and dual necessary conditions for local optimality. This is done for unconstrained problems and for problems with equality constraints. In the last section we give primal and dual characterizations of roots of functions of the single variable.

1. Introduction

A characterization of zero-derivative points for functions in several variables was introduced at the conference [13] and then published in several papers, e.g., [8, 9, 10, 11]. For a twice continuously differentiable function in two variables this result was proved in a mathematics text for economics students at University of Zagreb [6]. The topic of characterizing zero-derivative points, without explicitly using differentiation, has been discussed with several colleagues on Research Gate (RG), particularly on the author's Q & A page. It has also been studied in the research project [12] with selected applications to fixed points and roots of equations. In this paper we recall some of these ideas. There are at least two different proofs of the characterization of x^* where $\nabla f(x^*) = 0$. One of these uses the fact, loosely speaking, that C1 functions in n variables with Lipschitz derivative (i.e., the derivatives are Lipschitz functions) can be represented as differences of convex and quadratic convex functions. This approach was used in [8] and we will outline it in Theorem 2.1. Another approach uses, in part, the Mean Value Theorem of Calculus and it was recently given in [5]. The latter was essentially suggested, but not implemented, by one of the anonymous referees of [11]. The assumption that functions have Lipschitz derivative is justified by the fact that, otherwise, the functions may describe unrealistic situations with unbounded energy.

Example 1.1 (C1 Function with Non-Lipschitz Derivative)**.** Function $f(x) = |x|^{\frac{3}{2}}$ is C1 but its derivative does not have Lipschitz derivative at $x^* = 0$. (See a counter-example in [8].) In our proof of Theorem 2.1 we will use, possibly a non-intuitive result, that every C1 function in several variables with Lipschitz derivative on a convex set K is the difference of two convex functions. This decomposition holds

2010 *Mathematics Subject Classification.* 30C46.
Key words and phrases. zero derivative point, local optimum, nonlinear program, Lipschitz derivative, roots of functions.

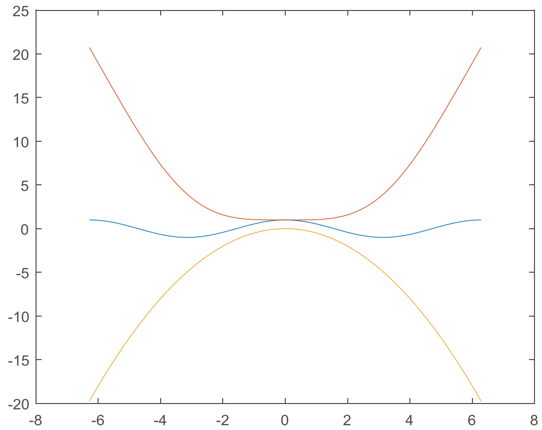

FIGURE 1. Decomposition of cos x into difference of two convex functions.

because

$$f(x) = \left[f(x) + \frac{1}{2}\alpha \cdot \|x\|^2\right] - \left[\frac{1}{2}\alpha \cdot \|x\|^2\right] \quad \text{for} \quad x \in K. \tag{1.1}$$

For $\alpha > 0$ sufficiently large, each term in the parentheses is a convex function [8].

Example 1.2 (Convex decomposition of periodic functions)**.** Consider f(x) = cos x on interval $K = [-2\pi, 2\pi]$. A decomposition is $\cos x = (\cos x + \frac{1}{2} \cdot x^2) - (\frac{1}{2} \cdot x^2)$. See Fig 1.

2. Zero-Derivative point

For a function in several variables denote by $\nabla f(x^*)$ the gradient of f at x^*. If f denotes a function of the single variable the notation is $f'(x)$. The norm $\|x\|$ used below is Euclidean. The key theorem of the paper follows.

Theorem 2.1 (Quadratic Envelope Characterization of Zero-Derivative Points, [8])**.** *Consider a C1 function $f(x)$ with Lipschitz derivative on a convex set K in its domain. Assume that K has interior points and consider an arbitrary interior point x^*. Then $\nabla f(x^*) = 0$ if, and only if, there exists a number $\Lambda \geq 0$ such that*

$$|f(x) - f(x^*)| \leq \Lambda \cdot \|x - x^*\|^2 \tag{2.1}$$

for every $x \in K$.

Proof. (Outline, for details see [8]) We know that $f(x) = \frac{1}{2}\alpha \cdot \|x\|^2 + C(\alpha, x)$ where C is a convex function in x and α is some convexifier [8, Theorem 2]. By convexity of C it follows that $C(\alpha, \lambda x + (1-\lambda)x^*) \leq \lambda C(\alpha, x) + (1-\lambda)C(\alpha, x^*)$, for every $x \in K, 0 \leq \lambda \leq 1$. After substitution for C, using Cauchy inequality and properties of the norm we find

$$f[(x^* + \lambda(x - x^*)] - f(x^*)] \leq f(x) - f(x^*) - \frac{1}{2}\alpha \cdot (1-\lambda) \cdot \|x - x^*\|^2.$$

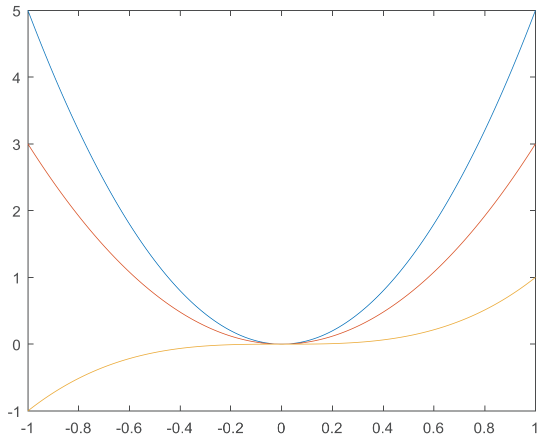

FIGURE 2. Primal necessary conditions for $f(x^*) = 0$.

Dividing by $\lambda > 0$ and sending $\lambda \to 0$ yields $[\nabla f(x^*)](x - x^*) \leq f(x) - f(x^*) - \frac{1}{2}\alpha \cdot \|x - x^*\|^2$. Suppose that $\nabla f(x^*) = 0$. Then

$$f(x^*) - f(x) \leq -\frac{1}{2}\alpha \cdot \|x - x^*\|^2.$$

Similarly $f(x^*) - f(x) \geq -\frac{1}{2}\beta \cdot \|x - x^*\|^2$ with a concavifier β. Now properties of α and β complete the necessity part of the proof. Sufficiency is easier to prove: Divide (2.1) by $\|x - x^*\| \neq 0$ and go to the limit $x \to 0$. □

Remarks. For the sake of simplicity, the inequality (2.1) is also called "the formula". If (2.1) holds for some $\Lambda \geq 0$ then it also holds for every bigger Λ. So we can talk about a class of $\Lambda's$. If f is a single-variable C2 function, one can specify $\Lambda = \frac{1}{2} \max t \in K|f'(t)|$. Note that the point$x^*$ in (2.1) is an apex (lowest point) of the paraboloid $\|x - x^*\|^2$ on K. One can split Theorem 2.1 into its "primal" and "dual" formulations. The former is also referred to as the apex property.

Corollary 2.2 (Primal Characterization of Zero-Derivative Points; Apex Property)**.** *Consider a C1 function $f(x)$ with Lipschitz derivative on a convex set K in its domain. Assume that K has interior points and consider its arbitrary interior point x^*. Then $\nabla f(x^*) = 0$ if, and only if, x^* is an apex of a class of paraboloids $\Lambda \cdot \|x - x^*\|^2$, $\Lambda \geq 0$ over-estimating $|f(x) - f(x^*)|$ on K.*

Example 2.3. Consider $f(x) = x^3$ on $K = [-1, 1]$ and $x^* = 0$. This point is an apex of the class of parabolas $\Lambda \cdot x^2$ over-estimating $|f(x)|$. Fig.2 depicts $f(x), |f(x]$, and the over-estimation with $\Lambda = 3$ and with $\Lambda = 5$. We can say, by Corollary 2.2, that an over-estimation of $|f(x) - f(x^*)|$ by $\Lambda \cdot \|x - x^*\|^2, \Lambda \geq 0$ is a necessary conditions for zero-derivative point $\nabla f(x^*) = 0$.

After dividing (2.1) by $\Lambda \neq 0$, Theorem 2.1 assumption a "dual form" which is given in terms of the ratio function

$$R(x) = |f(x) - f(x^*)|/\|x - x^*\|^2$$

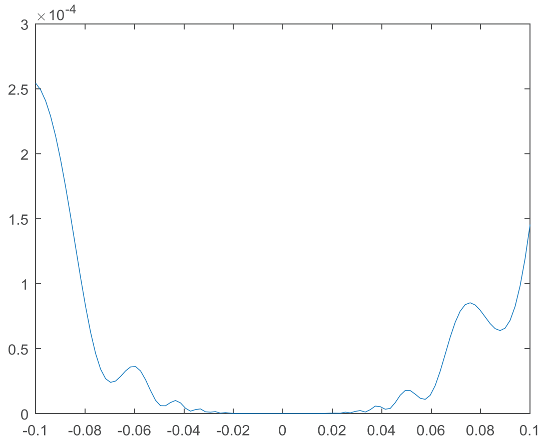

FIGURE 3. Function with infinitely many isolated local optima on a compact interval.

on K. $R(x)$ is not defined at the point $x^* \in K$ and it is uniformly bounded (by some Lipschitz derivative constant Λ) on $K \backslash \{x^*\}$. Functions with this property were studied in [9] in connection with L'Hopital's rule. Since R is not defined at $x = x^*$ we call such points the "vacant points". Characterization of zero-derivative points using $R(x^*)$ is called a vacant point property.

Corollary 2.4 (Dual Characterization of Zero-Derivative Points; Vacant Point Property.)**.** *Consider a C1 function $f(x)$ with Lipschitz derivative on a convex set K in its domain. Assume that K has interior points and consider its arbitrary interior point x^*. Then $\nabla f(x^*) = 0$ if, and only if, the ratio function $R(x)$ is not defined at x^* and it is uniformly bounded on the set $x \in K \backslash \{x^*\}$.*

Example 2.5 (Trivial)**.** Consider $f(x) = x^3$ on $K = [-1, 1]$. Take $x^* = 0$. Since $R(x) = |x|$ and $x^* \in K$, the conditions of Corollary 2.4 are trivially satisfied. This confirms that x^* is a zero-derivative point.

Example 2.6 (Non-trivial; non-monotonic function in every neighborhood of a local optimum)**.** Consider $f(x) = x^4 \cdot [2 + \sin\left(\frac{1}{x}\right)]$ for $x \neq x^* = 0, f(0) = 0$.

The graph of f is depicted in Fig. 2. Gelbaum and Olmsted studied this function in their book [4] in a different context. They observed that this is a differentiable function with an "extreme value at a point where the derivative does not make a simple change in sign. Also that "in no interval of the form $(a, 0)$ or $(0, b)$ is f monotonic. We note that the function is C1 with Lipschitz derivative on compact intervals around x^*. Also that f is highly nonconvex being non-monotonic in every neighborhood of $x^* = 0$. The function has infinitely many isolated local optima, including x^* which is a global minimum. Our objective hereby is to verify that both the apex and vacant- point properties hold at $x^* = 0$. This double checks $f'(x^*) = 0$.

In Fig. 2 the primal and dual conditions at zero-derivative point of $f(x)$ are compared.

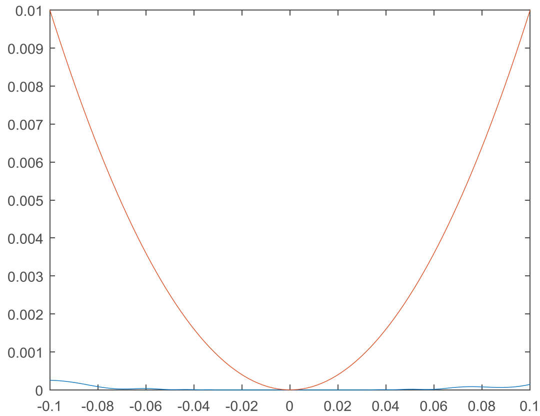

FIGURE 4. Apex property at $x^* = 0$ with $\Lambda = 1$.

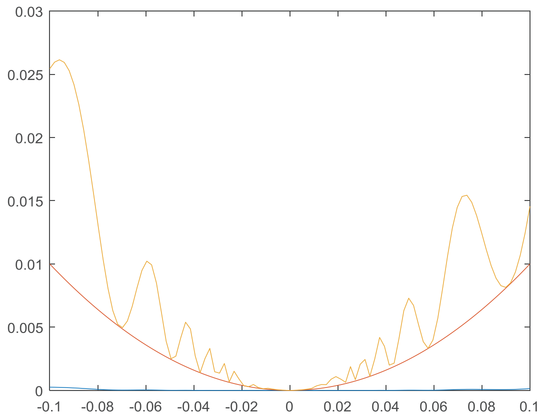

FIGURE 5. Comparison of two conditions for zero derivative at $x^* = 0$.

The above characterizations of zero-derivative points, in their necessity parts yield, possibly new, necessary conditions for local optimality.

3. GEOMETRY OF LOCAL OPTIMA

In this section we extend previous results to the classic problem of optimizing a function $f^0(x)$ in n variables subject to m equality constraints

$$\text{Opt} f^0(x), \text{ s.t. } f^i(x) = 0, i \in P = 1, \ldots, m. \tag{3.1}$$

Take a feasible point $x^*$of (3.1) which is also in the interior of some convex set K, not necessarily compact [5]. (Point x^* may not be an interior point of the feasible set of (3.1).) If x^* is a local optimum of $f^0(x)$ (unconstrained version of (3.1))then we know that x^* has two properties: the apex property and the vacant point property. These are necessary conditions for unconstrained local optima. Let us extend these to (3.1). For comparison we use the "complete" Lagrange function

$$L(x, \lambda) = \lambda_0 \cdot f^0 + \sum_{i \in P} \lambda_i \cdot fi(x), \text{ where } \lambda = (\lambda_i) \in R^{m+1}.$$

The classic theorem of Lagrange says that, at a local optimum x^* of (3.1), $\nabla L(x^*, \lambda*) = 0$ for some $(m+1) - tuple\lambda^* = (\lambda_i) \neq 0$. This implies that x^* is a zero-derivative point of $L(x, \lambda*)$ for some $\lambda^* \neq 0$. Hence his theorem can be reformulated using the " formula". We note that the reformulations will not require any constraint qualification (CQ), [2, 3].

Theorem 3.1 (Apex Property: Primal Formulation of the Theorem of Lagrange)**.** *Consider* (3.1) *where all functions are assumed to be C1 with Lipschitz derivatives on some convex set K which contains a feasible point x^*. If x^* is an interior point of K, which is also a local optimum of* (3.1)*, then x^* is an apex of a class of paraboloids $\Lambda \cdot \|x - x^*\|^2$, $\Lambda \geq 0$ over-estimating*

$$|L(x, \lambda*) - L(x^*, \lambda^*)| \text{ on } K \text{ for some } \lambda^* \neq 0.$$

Example 3.2 (Example 11.2 in [2] which illustrates the need for CQ)**.** Consider

$$\min f^0(x) = x_1, \text{ s.t. } f^1(x) = x_1^2 + (x_2 - 1)^2 - 1 = 0, f^2(x) = x_1^2 + (x_2 + 1)^2 - 1 = 0.$$

Clearly, the point $x_1^* = x_2^* = 0$ is optimal. Since $\nabla L(x^*, \lambda^*) = 0$ only if $\lambda_0^* = 0$, we know that no CQ holds. In contrast, the choice of any sphere centered at $x^* = 0 \in R^2$ for K, and any choice of $\lambda \geq 2$, confirms that $x^* = 0$ is an apex of a class of parabolas over-estimating

$$|L(x, \lambda^*) - L(x^*, \lambda^*)| = 2 \cdot (x_1^2 + x_2^2) \text{ on } K \text{ for } \lambda^* = (0, 1, 1)'.$$

In the context of optimality this means that $x^* = 0$ could be a local optimum. Let us formulate an equivalent necessary condition for optimality.

Theorem 3.3 (Vacant Point Property: Dual Formulation of the Theorem of Lagrange)**.** *Consider* (3.1) *where all functions are assumed to be C1 with Lipschitz derivatives on some convex set K. Suppose that x^* is an interior point of K, containing a feasible point x^*, which is also a local optimum of* (3.1)*. Then x^* is a vacant point of*

$$R(x) = |L(x, \lambda^*) - L(x^*, \lambda^*)| / \|x - x^*\|^2$$

on K for some $\lambda^* \neq 0$, i.e., $R(x)$ is not defined at x^* and $R(x)$ is uniformly bounded on $K \backslash x^*$, for some $\lambda^* \neq 0$.

Comments. Problems with mixed inequalities, studied in nonlinear programming [3], can be reduced to the form (3.1) if non-active constrains at a feasible point x^* are omitted. All above results essentially follow from the quadratic envelope characterization of zero-derivative points, i.e., from the "formula. It is easy to see that the assumption of compactness of K in the original proof of the formula can be omitted, see [5] for an alternative proof in parts.

4. "Formula" in integral calculus: characterizing roots

In this section we use the "formula in integral calculus to characterize roots of a single variable function $f(x) = 0$.

Theorem 4.1 ("Formula in Integral Calculus)**.** *Consider a continuous Lipschitz function f of the single variable x on $I = [a, b]$ in its domain and a point x^* such that $a < x^* < b$. Denote the integral of f(t) from x^* to x in I by*

$$Y(x) = \inf[x^*, x] f(t) dt.$$

Then x^* is a root of f, i.e., $f(x^*) = 0$, if and only if

$$|Y(x)| \leq \Lambda \cdot (x - x^*)^2, \quad x \in I$$

for some class of $\Lambda \geq 0$.

Proof. (Outline) Suppose that $f(x^*) = 0$. We know that $Y'(x) = f(x)$ by the fundamental theorem of calculus. Hence $Y'(x^*) = 0$ and then

$$|Y(x) - Y(x^*)| \leq \Lambda \cdot (x - x^*)^2 \text{ by the "formula".}$$

But $Y(x^*) = 0$, completing the necessity part of the proof. The implications also hold in the reverse order. Indeed, if $Y(x^*) = 0$ then also $|Y(x)| \leq \Lambda \cdot (x - x^*)^2$, $x \in I$ for some $\Lambda \geq 0$. This follows by the "formula because $Y'(x^*) = 0$. Hence $f(x^*) = 0$. □

The above yields primal ("apex") and dual ("vacant point") characterizations of roots:

Theorem 4.2 (Primal Characterization of Roots)**.** *Using the assumptions of Theorem 4.1, an arbitrary interior point x^* of I is a root of $f(x)$ if, and only if, x^* is an apex of a class of parabolas $r \cdot x^2, r \geq 0$ over-estimating the absolute value of the function $Y(x)$ on I.*

Example 4.3. Consider $f(x) = \sin x$ on $I = [-\pi, \pi]$. Property of the root $x^* = 0$ is depicted by the Primal characterization $Y(x)| \leq x^2$,with the choice $r = 1$, Fig. 4.

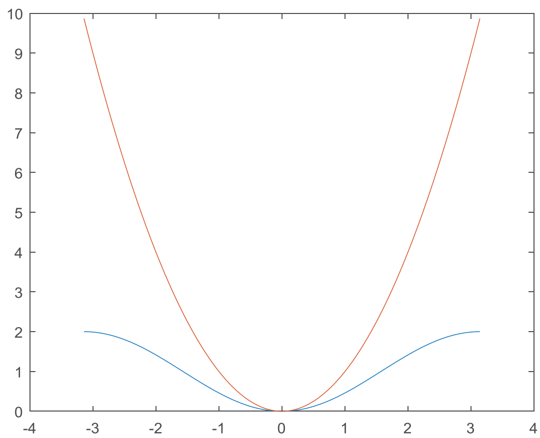

FIGURE 6. Primal characterization of a root.

We also have a "dual characterization of roots:

Theorem 4.4 (Dual Characterization of Roots)**.** *Under the assumptions of Theorem 4.1, an arbitrary interior point x^* of I is a root of $f(x)$ if, and only if, $|Y(x)|/(x - x^*)^2$ is uniformly bounded on $I \backslash x^*$.*

Example 4.5. Consider $f(x) = \sin x$ and $x^* = 0$. Take $I = [-\pi, \pi]$. Is $f(0) = 0$? True, by the dual result, since $Y(x) = 1 - \cos x$ and $R(x) = Y(x)|/x^2, x \neq 0$ is uniformly bounded on $I \backslash 0$. Fig.4 depicts a difference between primal and dual characterizations. We can say that the primal is an "apex characterization" and the dual is a "vacant point" characterization of the root x^*.

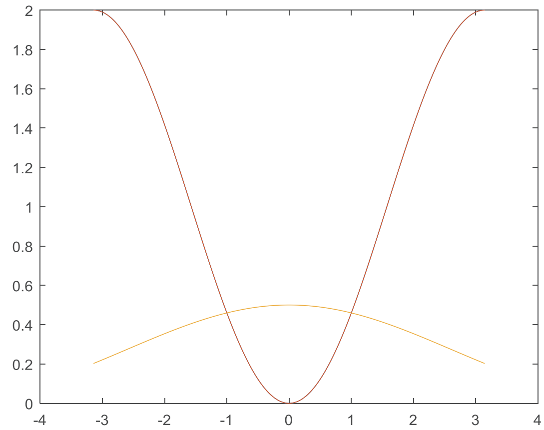

Figure 7. Primal and dual characterizations of a root compared.

Other applications of the "formula include a study of fixed points [11], sensitivity of the Cobb-Douglas production function in economics, and selected results from linear and nonlinear regression [1, 9].

5. Conclusion

Leonhard Euler proclaimed that "nothing at all takes place in the universe in which some rule of maximum or minimum does not appear [3, 7]. Fermat et al. formalized this claim by proving that, for differentiable functions $f(x)$, the optimal points can occur only at zero-derivative points $\nabla f(x^*) = 0$. For C1 functions with Lipschitz derivatives, the points where $\nabla f(x^*) = 0$ were characterized at the conference [13] and in its follow-up papers, e.g., [8, 9, 10]. Combining these results with the quadratic envelope characterization of zero- derivative points, a.k.a. the "formula [11, 12, 13], we have introduced hereby two equivalent, but geometrically different, necessary conditions for local optima. In the final section we used both the formula and the fundamental theorem of calculus to characterize roots of Lipschitz functions of the single variable. Hence from now on we can talk about "primal and 電ual necessary conditions for local optimality of functions in several variables and 菟rimal and 電ual characterizations of roots of functions of the single variable.

Acknowledgment

The author is indebted to his project followers on RG and all colleagues who made comments and recommendations, in particular to Adi Ben-Israel (Rutgers), Dragan Jukić (Osijek), Luka Neralić (Zagreb), Dejenie Lakew (Stratford), Arpad Szaz (Debrecen), Octav Olteanu (Bucharest), Jyotiranjan Nayak (Hyderabad), Firdaus

Udwadia (Southern California), Joachim Domsta (Elblag), Milan Vlach (Prague), Qingying Xue (Bejing), Miodrag Mateljevic and Miroslav Pavlovi (Belgrade), Mikhail Lvovich Goldman (Moscow) and Semen Kutateladze (Novosibirsk). However, the author is responsible for all possible errors and omissions in the text.

References

[1] A. Ben-Israel and T. N. Greville, *Generalized Inverses: Theory and Applications*, CMS Books in Mathematics, Second Edition, Springer, 2013.

[2] A. Ben-Israel, A. Ben-Tal and Sanjo Zlobec, *Optimality in Nonlinear Programming: A Feasible Directions Approach*, John Wiley & Sons, 1981.

[3] R. W. Cottle and M. N. Thapa, *Linear and Nonlinear Optimization*, Int. Ser. Operat. Res. & Manag., Springer, 2017.

[4] B. R. Gelbaum and J. M. H. Olmstead, *Counterexamples in Analysis*, Holden-Day, 1964.

[5] D. Jukić, *An elementary proof of the quadratic envelope characterization of zero-derivative points*, Optim Lett, July 2017. https://doi.org/10.1007/s11590-017-1174-1

[6] L. Neralić and B. Šego, *Matematika*, Fourth edition, Zagreb, 2014. Element.

[7] W. A. Oldfather, C. A. Ellis and D. M. Brown, *Leonhard Euler 痴 elastic curves*, ISIS **20** (1933), 72–160 (English translation of L. Euler, Methodus Inviendi Lineas Curvas Maximi Minimive Proprietate Gaudentes, Lausanne & Geneva (1744), p.245.

[8] S. Zlobec, *Characterizing zero-derivative points*, J. Global Optim. **46** (2010), 155-161.

[9] S. Zlobec, *L 迦 opital 痴 rule without derivatives*, Math. Comm. **17** (2012), 665–672.

[10] S. Zlobec, *Modelling with twice continuously differentiable functions*, Croatian Operational Research Review (CRORR) **5** (2014), 1–13.

[11] S. Zlobec, *Characterizing fixed points*, Croatian Op. Res. Review (CRORR) **8** (2017), 353–358.

[12] S. Zlobec, *Studying the essence of fixed points*, RG project (2017).

[13] S. Zlobec, *On the derivative of smooth meaningful functions*, in: 13th Int. Conference Op. Research, Faculty of Economics, Split, (2010) Extended abstract Croatian Op. Res. Review (CRORR) v.2 (2011) p.3.

Manuscript received October 5 2017
revised November 24 2017

Sanjo Zlobec
McGill University, Department of Mathematics and Statistics, Burnside Hall, 805 Sherbrook West, Montreal, Quebec, Canada H3A 2K6
E-mail address: zlobec@math.mcgill.ca