

Modelling with twice continuously differentiable functions*

Sanjo Zlobec^{1,†}

¹ *Department of Mathematics and Statistics, McGill University
Burnside Hall, 805 Sherbrooke Street West, Montreal, Quebec, Canada H3A 2K6
E-mail: {zlobec@math.mcgill.ca}*

Abstract. Many real life situations can be described using twice continuously differentiable functions over convex sets with interior points. Such functions have an interesting separation property: At every interior point of the set they separate particular classes of quadratic convex functions from classes of quadratic concave functions. Using this property we introduce new characterizations of the derivative and its zero points. The results are applied to the study of sensitivity of the Cobb-Douglas production function. They are also used to describe the least squares solutions in linear and nonlinear regression.

Key words: twice continuously differentiable function, zero derivative point, separation property of functions, Cobb-Douglas production function, least squares solution, Newton's second law

Invited paper, available online: March 4, 2014

1. Introduction

Operations research is the study of improvable real life situations using mathematical models [11]. One of its basic topics is the identification of extreme values of functions and their corresponding optimal solutions. The classic extreme value theorem of Fermat is still widely used in various frameworks in the study of extreme points. The theorem says that a locally optimal solution can occur only at a zero-derivative point, also called a “stationary point”, e.g., [6, 11, 12]. These points have been recently characterized for continuously differentiable functions with a Lipschitz derivative and, in particular, for twice continuously differentiable functions in several variables [17, 18, 19]. The characterizations are based on the quadratic envelope property of functions introduced in [17] and they do not require infinitesimal calculus. They provide an alternative approach to the study of optimality. Let us loosely illustrate these ideas for a twice continuously differentiable function $f(x)$ of the single variable x on an interval $I = [a, b]$ around an interior point x^* of I . Since $f(x)$ is assumed to be twice continuously differentiable we know that its second derivative assumes its extreme values on I . In particular, there exists a number

$$\rho = \max_{x \in I} |f''(x)|.$$

*This research is partly supported by NSERC of Canada.

†Corresponding author.

Following [17, 18, 19] (and Theorems 3 and 4 below) we know that an arbitrary number $g = g(x^*)$ is the derivative of f at x^* if, and only if, the ratio function

$$\frac{|f(x) - f(x^*) - g \cdot (x - x^*)|}{(x - x^*)^2} \quad (1)$$

is bounded on the set $I \setminus \{x^*\} = \{x \in I, x \neq x^*\}$ with an upper bound $\frac{1}{2}\rho$. For the purpose of studying sensitivity of the function we can state this result differently: An arbitrary number $g = g(x^*)$ is the derivative of f at x^* if, and only if, there exists $\Lambda \geq \rho$ (in fact, for every Λ “sufficiently large”) such that

$$|f(x) - f(x^*) - g \cdot (x - x^*)| \leq \frac{1}{2}\Lambda(x - x^*)^2 \quad (2)$$

for every $x \in I$.

Example 1 (Verification of Derivative). Consider $f(x) = x^5 - x^2$ on $I = [-1, 2]$ and $x^* = 0$. We would like to know whether $f'(x^*) = 0$. Since the ratio function (1), with $g = 0$, is

$$\frac{|x^5 - x^2|}{x^2} = |x^3|$$

and it is bounded on $I \setminus \{0\}$, we conclude that $g = f'(0) = 0$. How about, say, $x^* = 1$? Since $|x^5 - x^2|/(x - 1)^2 \rightarrow \infty$ as $x \rightarrow 1$ we conclude that at this point $f'(x^*) \neq 0$.

Example 2 (Sensitivity). Let us approximate $f(x) = x^5 - x^2$ on $I = [-1, 2]$ around $x^* = 0$. We know by (2) that, for a given tolerance $\varepsilon > 0$

$$\frac{1}{2}\Lambda(x - x^*)^2 \leq \varepsilon$$

implies

$$|f(x) - f(x^*) - g \cdot (x - x^*)| \leq \varepsilon$$

for $x \in I$. One can specify $\Lambda = \rho = 158$. Therefore, since we already know that $g = f'(0) = 0$, for every x satisfying $79x^2 \leq \varepsilon$ we have $|f(x)| \leq \varepsilon$. In particular, for the choice $\varepsilon = 1$, we know that $|f(x)| \leq 1$ for every x such that $79x^2 \leq 1$.

There are real life situations which possibly cannot be improved, such as situations described by the classical laws of physics. The above results may provide different mathematical descriptions of these situations.

Example 3. Consider an object of mass m moving in time along a twice continuously differentiable trajectory $f(x)$ governed by Newton’s second law. If $F(x)$ denotes some “force” acting on the object then on a time interval $I = [a, b]$ the law says that $F(x) = mf''(x)$. When the force is explicitly known, one can obtain the object’s trajectory by finding a solution of this differential equation. Suppose that we do not necessarily know the force. Nevertheless, using (1) we know that for the trajectory $f(x)$ on I and the instantaneous velocity $v(x^*)$, at every moment x^* , $a < x^* < b$ the ratio function

$$\frac{|f(x) - f(x^*) - v(x^*) \cdot (x - x^*)|}{(x - x^*)^2}$$

is bounded on the set $I \setminus \{x^*\}$. After specifying $\Lambda = \rho$ in (2), a bound is

$$\frac{1}{2}\Lambda = \frac{1}{2}\rho = \frac{1}{2} \max_{x \in I} |f''(x)| = \frac{1}{2m} \max_{x \in I} |F(x)|.$$

In particular, if the object is moving with a constant velocity, then $\rho = 0$ and we conclude that its position is $f(x) = f(x^*) + v(x^*) \cdot (x - x^*)$ at every $x \in I$. It is easy to verify these properties for the trajectory of an object during its free fall.

The preceding examples show that the new characterizations of the derivative may not always be suitable for calculating the derivative but they have three applications. They can be used to verify (rather than calculate) the derivative, determine a region around a given x^* where all values of the function fall within a specified tolerance, and one can possibly use them to learn more about the model.

The objective of this paper is to extend some of the recent results for functions of the single variable to functions in several variables. The extensions will be illustrated on the Cobb-Douglas production function in economics and also in regression. In particular, we obtain new characterizations of partial derivatives of the production function, which are important to study sensitivity of the model, and new characterizations of the best least squares solutions. These illustrations have different mathematical properties. The Cobb-Douglas function is studied around non-zero derivative points and therefore the related problems are well posed [13, 14]. In contrast, the least squares problems generally are ill posed and they are also known to be notoriously ill conditioned [2]. We will not study numerical topics. In optimization these are studied in, e.g., [4] and [7]. Mathematical background of the paper is set at the level of intermediate calculus and linear algebra. The most recent material related to this paper can be found in [19].

2. Separation of functions

Consider a twice continuously differentiable (abbreviated: C^2) function f in n variables on a closed and bounded (abbreviated: compact) convex set K with interior points. Denote by $H = \nabla^2 f(x)$ the Hessian matrix of f at x and by $\lambda_i = \lambda_i(x)$ its i -th eigenvalue, $i = 1, \dots, n$. Since the eigenvalues of H are real numbers we can introduce the “global” spectral radius of H on K , denoted by

$$\rho = \max_{x \in K} \max_{i=1, \dots, n} |\lambda_i|.$$

This is a non-negative number which depends on H and K . It makes the function $f(x) + \frac{1}{2}\rho\|x^2\|$ convex. Here $\|x\|$ denotes the Euclidean norm of a column n -tuple x . In fact, one has a more general claim.

Theorem 1. *Consider a C^2 function f in n variables on a compact convex set K in its open domain and the global spectral radius ρ of its Hessian matrix on K . Then for every number $\Lambda \geq \rho$, the function $f(x) + \frac{1}{2}\Lambda\|x^2\|$ is convex on K and the function $f(x) - \frac{1}{2}\Lambda\|x^2\|$ is concave on K .*

Proof. The function $f(x) + \frac{1}{2}\Lambda\|x^2\|$ is convex on K if, and only if, the matrix $\nabla^2 f(x) + \Lambda I$, where I is the $n \times n$ unit matrix, is positive semi-definite for every $x \in K$. Using the transpose u^T of u , and matrix multiplication, this is equivalent to

$$\frac{u^T \nabla^2 f(x) u}{\|u^2\|} + \Lambda \geq 0,$$

for every $u \neq 0$. Since $\Lambda \geq \rho$, we have $\lambda_i(x) + \rho \geq 0$, and it follows that $\lambda_i(x) + \Lambda \geq 0, i = 1, \dots, n$. Hence we conclude that $f(x) + \frac{1}{2}\Lambda\|x^2\|$ is convex. Similarly one can prove the concavity part. \square

Example 4. If f is a C^2 function of the single variable on an interval $I = [a, b]$, then

$$\rho = \max_{x \in I} |f''(x)|.$$

This number was introduced in Section 1. For $f(x) = \sin x$ on $I = [-\pi, \pi]$, it is $\rho = 1$. For the product function $f(x) = x_1 x_2$, the eigenvalues of the Hessian matrix on any compact convex set K are $\lambda_1 = -1$ and $\lambda_2 = 1$, hence $\rho = 1$.

We use ρ to describe a separation property of C^2 functions.

Theorem 2 (Separation property of C^2 functions). *Consider a C^2 function $f(x)$ in n variables defined on an open set of \mathbb{R}^n containing a compact convex set K . Denote by ρ the global spectral radius of the Hessian matrix of f on K . Take an arbitrary interior point x^* of K and denote by $G = \nabla f(x^*)$ the gradient of f at x^* . Then for every number Λ such that $\Lambda \geq \rho$, we have*

$$-\frac{1}{2}\Lambda\|x - x^*\|^2 + f(x^*) + G \cdot (x - x^*) \leq f(x) \leq f(x^*) + G \cdot (x - x^*) + \frac{1}{2}\Lambda\|x - x^*\|^2 \quad (3)$$

for every $x \in K$.

Proof. We know, for $\Lambda \geq \rho$, that $C(x, \Lambda) = f(x) + \frac{1}{2}\Lambda\|x\|^2$ is a convex function on K , by Theorem 1. Therefore, for x and x^* in K , we have

$$C(\lambda x + (1 - \lambda)x^*, \Lambda) \leq \lambda C(x, \Lambda) + (1 - \lambda)C(x^*, \Lambda)$$

for every $0 \leq \lambda \leq 1$. Hence

$$\begin{aligned} f(\lambda x + (1 - \lambda)x^*) &\leq \frac{1}{2}\Lambda\|\lambda x + (1 - \lambda)x^*\|^2 + \lambda f(x) - \frac{1}{2}\Lambda\lambda\|x\|^2 \\ &\quad + (1 - \lambda)f(x^*) + \frac{1}{2}\Lambda(1 - \lambda)\|x^*\|^2. \end{aligned}$$

Using properties of the norm, and after division by $\lambda > 0$, this yields

$$\frac{f(x^* + \lambda(x - x^*)) - f(x^*)}{\lambda} \leq f(x) - f(x^*) + \frac{1}{2}\Lambda(1 - \lambda)\|x - x^*\|^2.$$

On the left-hand side we have a quotient of functions in the single variable λ of the type $\frac{0}{0}$. Using L'Hopital's rule, in the limit $\lambda \rightarrow 0$ this becomes

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

Similarly, since $\tilde{C}(x, \Lambda) = f(x) - \frac{1}{2}\Lambda\|x\|^2$ is a concave function, it follows that

$$\nabla f(x^*) \cdot (x - x^*) \geq f(x) - f(x^*) - \frac{1}{2}\Lambda\|x - x^*\|^2.$$

□

Theorem 2 gives relationships between a C^2 function and its first and second derivatives. The second derivative enters the inequalities indirectly through ρ .

Example 5. Consider $f(x) = \sin x$ on a compact interval and its interior point x^* . The separation property, after specifying $\Lambda = \rho$ and $\rho = 1$, says that

$$-\frac{1}{2}(x - x^*)^2 + \sin x^* + \cos x^* \cdot (x - x^*) \leq \sin x \leq \sin x^* + \cos x^* \cdot (x - x^*) + \frac{1}{2}(x - x^*)^2$$

for every $x \in I$. At $x^* = 0$ this yields

$$x - \frac{1}{2}x^2 \leq \sin x \leq x + \frac{1}{2}x^2 \quad \text{for} \quad -\pi \leq x \leq \pi.$$

The choice of a smaller interval, e.g., $I = [-\frac{\pi}{4}, \frac{\pi}{4}]$ and $\rho = \frac{1}{2} \max_{x \in I} |f''(x)| = \frac{\sqrt{2}}{4}$ yields

$$x - \frac{\sqrt{2}}{4}x^2 \leq \sin x \leq x + \frac{\sqrt{2}}{4}x^2 \quad \text{for} \quad -\frac{\pi}{4} \leq x \leq \frac{\pi}{4}.$$

The separation is depicted in Figure 1 on the region $-\frac{\pi}{2} \leq x, x^* \leq \frac{\pi}{2}$ by an “extended graph” of $f(x)$ in the space with coordinates (x, x^*) ; function $\sin x$ is represented in that space as $\sin x + 0 \cdot x^*$.

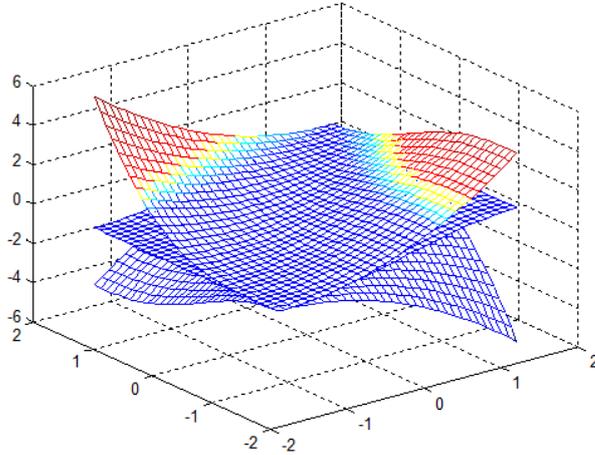


Figure 1: Separating property of sine function

The inequalities (3) can be written in a more concise form using the absolute value function. For $f(x)$ on K we introduce its “extension function” around an interior point x^* of K as

$$E(x, x^*) = \frac{|f(x) - f(x^*) - \nabla f(x^*) \cdot (x - x^*)|}{\|x - x^*\|^2}, \quad x \in K \setminus \{x^*\} = \{x \in K, x \neq x^*\}.$$

Theorem 2 says that $E(x, x^*)$ is bounded on the set $K \setminus \{x^*\}$ by $\frac{1}{2}\rho$.

Example 6. Consider $f(x) = x_1x_2$ on $K = \{(x_1, x_2)^T : -1 \leq x_1, x_2 \leq 1\}$. Its extension function around $x^* = 0$ is

$$E(x, x^*) = \frac{|x_1x_2|}{x_1^2 + x_2^2}.$$

Since $\rho = 1$, by Example 4, an upper bound of $E(x, x^*)$ on $K \setminus \{x^*\}$ is $\frac{1}{2}$. The graph of this function is depicted in a different context in [19].

Theorem 2 can also be used to find lower bounds of the global spectral radius.

Example 7. Consider the product function $f(x) = x_1x_2 \cdots x_n$ in $n \geq 2$ variables on a compact convex set K in \mathbb{R}^n containing $x^* = 0$ in its interior. Since $G = \nabla f(x^*) = 0$, (3) gives a lower bound of the global spectral radius ρ of the Hessian matrix of f on K

$$\frac{2x_1x_2 \cdots x_n}{x_1^2 + x_2^2 + \cdots + x_n^2} \leq \rho \quad \text{for every } x = (x_i) \neq 0.$$

In particular, for $n = 2$, the inequality is obvious. But it is not obvious for $n \geq 3$.

3. Characterizations of the gradient and its zero points

The well-known geometric property of the gradient of $f(x)$ at x^* is that it is an n -tuple in \mathbb{R}^n orthogonal to the level set $\{x : f(x) = f(x^*)\}$ at x^* and pointing in the direction of steepest ascent of f from x^* . This property is used in the classic steepest ascent method of Cauchy and in its many variations and numerical improvements. In mathematical modelling it is used to describe many interesting situations such as flight of insects toward a light source, trajectory of a heat-seeking missile and movements of sharks in water towards higher blood concentration [1]. In this section we will use Theorem 2 to give an essentially different property of the gradient. In fact, we will characterize the gradient.

Theorem 3 (Global characterization of the gradient). *Consider a C^2 function $f(x)$ in n variables defined on an open set of \mathbb{R}^n containing a compact convex set K with a nonempty interior. Also consider the global spectral radius ρ of the Hessian matrix of f on K and a number $\Lambda, \Lambda \geq \rho$. A row n -tuple $G = G(x^*)$ is the gradient of f at an interior point x^* of K if, and only if, the inequalities (3) hold for every $x \in K$.*

Proof. In view of Theorem 2, we only have to show that G in (3), is represented by an n -tuple of partial derivatives. This can be seen after dividing (3) by $\|x - x^*\| \neq 0$ where $x \neq x^*$ is chosen on the i -th coordinate axis. The i -th component of $G = (G_i)$,

in the limit $x \rightarrow x^*$, is the partial derivative $\frac{\partial f}{\partial x_i}(x^*)$, $i = 1, \dots, n$. Hence G is the gradient of f at x^* . \square

Theorem 3 can be formulated differently using the extension function of f .

Theorem 4 (Uniform bound characterization of the gradient). *Consider a C^2 function $f(x)$ in n variables defined on an open set of \mathbb{R}^n containing a compact convex set K with a nonempty interior. Also consider the global spectral radius ρ of the Hessian matrix of f on K and a number $\Lambda, \Lambda \geq \rho$. A row n -tuple $G = G(x^*)$ is the gradient of f at an interior point x^* of K if, and only if*

$$\frac{|f(x) - f(x^*) - G \cdot (x - x^*)|}{\|x - x^*\|^2} \leq \frac{1}{2}\Lambda \quad (4)$$

for every $x \in K \setminus \{x^*\}$.

Let us illustrate Theorem 4.

Example 8. Consider $f(x) = x^{0.4}$ on the interval $I = [2, 4]$. Take $x^* = 3$. Is $f'(3) = 0.2$? Since the left-hand side in (4), with $G = 0.2$, is not bounded on the set $I \setminus \{3\}$ because

$$\frac{|x^{0.4} - 3^{0.4} - 0.2(x - 3)|}{(x - 3)^2} \rightarrow \infty \quad \text{as } x \rightarrow 3$$

we conclude that $f'(3) \neq 0.2$. This situation is depicted in Figure 2.

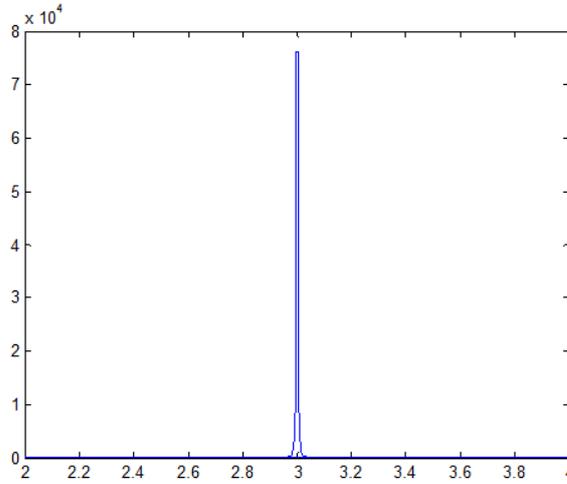


Figure 2: Incorrect derivative

Is $f'(3) = 0.4$? Since the ratio function

$$\frac{|x^{0.4} - 3^{0.4} - 0.4(x - 3)|}{(x - 3)^2}$$

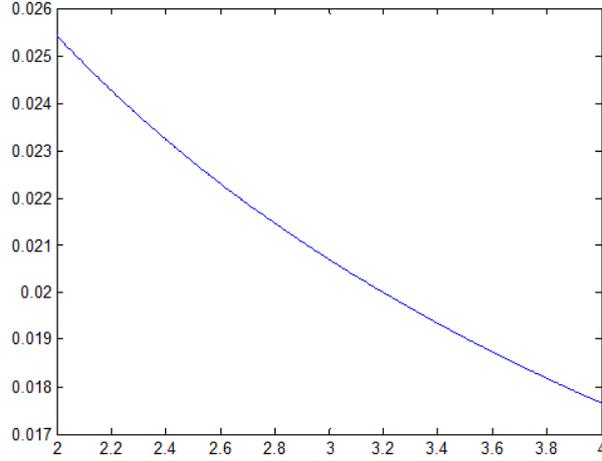


Figure 3: Correct derivative

is bounded on $I \setminus \{3\}$ we conclude that $f'(3) = 0.4$. This situation is depicted in Figure 3.

Warning. Theorems 3 and 4 do not generally hold for continuously differentiable functions. A counter example is $f(x) = |x|^{3/2}$ on $I = [-1, 1]$ considered around $x^* = 0$, [17].

Theorem 3 can also be used to estimate values of functions around a given point x^* falling within a prescribed tolerance $\varepsilon \geq 0$. Let us follow-up on Example 2 for functions in several variables.

Theorem 5. Consider a C^2 function $f(x)$ in n variables defined on an open set of \mathbb{R}^n containing a compact convex set K with nonempty interior. Let $\rho > 0$ be the global spectral radius of the Hessian matrix of f on K and take $\Lambda \geq \rho$. Consider an interior point x^* of K . Then for a given $\varepsilon \geq 0$

$$|f(x) - f(x^*) - \nabla f(x^*) \cdot (x - x^*)| \leq \varepsilon \quad \text{whenever} \quad \|x - x^*\|^2 \leq \frac{2\varepsilon}{\Lambda}, x \in K.$$

Example 9. Consider $f(x) = \cos x$ around $x^* = 0$. Here $\rho = 1$ and we can take $\Lambda = 1$. Therefore for any tolerance $\varepsilon \geq 0$, we know that $|1 - \cos x| \leq \varepsilon$, for every x which satisfies $x^2 \leq 2\varepsilon$. In general, the bigger choice of Λ , the smaller interval for estimation.

Theorems 3 and 4 yield, in particular, characterizations of zero-derivative points $\nabla f(x^*) = 0$. One only specifies $G = 0$ in the theorems. One can find illustrations of this special case in, e.g., [12, 17, 18, 19].

4. Cobb-Douglas production models

The Cobb-Douglas models are formulated as products of several functions each of the single variable x_i raised to some constant powers $a_i, i = 1, \dots, k$. In economics, they might represent the technological relationships between k different inputs such as capital and labour and the amount of output that is produced by these inputs. The results from the preceding sections are directly applicable to these models to verify derivatives and study sensitivity under input perturbations. Let us outline how this can be done on an example borrowed from [12, p. 291] with the original notation.

Example 10. Consider the Cobb-Douglas production model $Q(L, K) = 100L^{0.4}K^{0.6}$ studied around $L^* = 1024$ and $K^* = 32768$. Here Q denotes total production (the real value of all goods produced in, say, a year), L denotes labour input (the total number of person hours worked in a year), K denotes capital input (the real value of all machinery, equipment, and buildings), 0.4 and 0.6 are the output elasticities of capital and labour respectively. Constant $A = 100$ is total factor productivity. The partial derivatives are found to be $\frac{\partial Q(L^*, K^*)}{\partial L} \approx 320$ and $\frac{\partial Q(L^*, K^*)}{\partial K} \approx 15$. This means that, at K^* , if L^* is increased from 1024 to 1025, Q is roughly estimated to increase by 320. Similarly, at the fixed L^* , if K is increased from 32768 to 32769, Q is estimated to increase by 15.

The extended function of $Q(L, K)$ around L^* and K^* is the ratio function

$$E(L, K, L^*, K^*) = \frac{|Q(L, K) - Q(L^*, K^*) - (320, 15)(L - L^*, K - K^*)^T|}{(L - L^*)^2 + (K - K^*)^2}$$

where $(L, K) \neq (L^*, K^*)$ in a compact convex set containing (L^*, K^*) in its interior. Its graph is depicted by Figure 4.

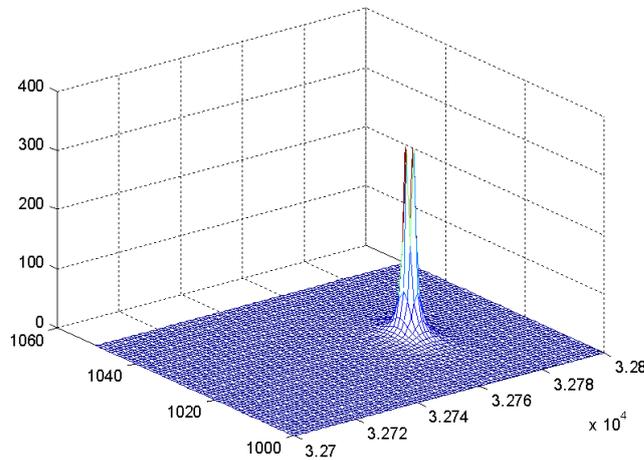


Figure 4: Extended Cobb-Douglas function around fixed inputs

For the results to be meaningful, the function $E(L, K, L^*, K^*)$ should be bounded on every compact convex set around (L^*, K^*) . From the graph (done in MATLAB) we see that the bound has the value about 200. This implies that the global spectral radius ρ of the Hessian matrix on the chosen region in (L, K) around (L^*, K^*) is about 400. One can use this information for a more refined sensitivity analysis using Theorem 5. Finer the grid in the (L, K) plane, smoother is the shape of the graph and more accurate are the estimates of ρ and perturbed values of the function.

Partial derivatives of a general Cobb-Douglas function $f(x_1, \dots, x_k)$, at a given input x_i^* are $G_i = \frac{\partial f(x_i^*)}{\partial x_i}$, $i = 1, \dots, k$. They can be verified directly using Theorems 3 and 4 if we study the k “coordinate-wise” functions of the single variable: $f(x_1, x_2^*, \dots, x_k^*), \dots, f(x_1^*, \dots, x_{k-1}^*, x_k)$. Since the model is stated in terms of C^2 functions and studied around non-zero inputs, the verifications of the derivatives and study of sensitivity are well posed problems according to the general results given in [13] and [14].

5. Regression

Theorems 3 and 4 can be used in linear regression to characterize least squares solutions of possibly inconsistent systems of linear algebraic equations $Ax = b$. For a given $m \times n$ matrix A and an m -tuple b , consider the function

$$F(x) = \|Ax - b\|^2.$$

The Hessian matrix of $F(x)$ is $\nabla^2 F(x) = 2A^T A$ which is positive semi definite. Therefore $F(x)$ is a convex function regardless of the choice of A and b . Its optimal solution x^* is called a least squares solution. It is a zero-derivative point of $F(x)$ so let us use, e.g., Theorem 4 for its characterization. Denote by σ the largest eigenvalue of $A^T A$.

Theorem 6 (Characterization of least squares solutions). *An n -tuple x^* is a least squares solution of $Ax = b$ if, and only if*

$$\frac{|\|Ax - b\|^2 - \|Ax^* - b\|^2|}{\|x - x^*\|^2} \leq \sigma \quad (5)$$

on the set $K \setminus \{x^*\}$, where K is an arbitrary compact convex set containing x^* in its interior.

Example 11 (Trivial example). *Consider $Ax = b$ where $A = 0$ is the $m \times n$ zero matrix. Any n -tuple x^* is its least squares solution.*

Remark 1. *Theorem 6 can be formulated using properties of generalized inverses of matrices, e.g., [2, 8, 15]. In particular, the term $Ax^* - b$ is the orthogonal projection of b on the null space of A^T .*

Example 12 (Estimating number of divorces in Canada in 1890). *The number of registered divorces in Canada from 1880 to 1889 is given by the following data according to the “Statistical Year-book of Canada 1902”:*

Year (t_i)	1880	1881	1882	1883	1884	1885	1886	1887	1888	1889
Divorces (d_i)	5	7	6	13	10	12	11	10	9	15

Using linear regression we wish to estimate the number of divorces in 1890. The ten years can be ordered from, say, 1 to 10 instead of 1880 to 1889. The linear regression line can be assumed to be of the form $d = x_1 t + x_2$. Passing this line through the ten points gives a system $Ax = b$ of 10 equations in 2 unknowns where

$$A^T = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix},$$

$$b^T = [5 \ 7 \ 6 \ 13 \ 10 \ 12 \ 11 \ 10 \ 9 \ 15].$$

The least squares solution $x^* = (x_1^*, x_2^*)$ is calculated to be $x_1^* = 0.73$, $x_2^* = 5.8$. Hence the regression line is $d = 0.73t + 5.8$ and the estimated number of divorces in 1890 is $d = 0.73 \cdot 11 + 5.8 = 13.83 \approx 14$. The result can be verified using Theorem 6. An upper bound of the ratio function in (5) on any compact convex set containing x^* in its interior is $\sigma = 392.90$.

In general, the “error function” $F(x)$ can be any C^2 function. If data (d_i, t_i) , $i = 1, \dots, N$ in the (d, t) plane are approximated by, say, the exponential function

$$d = x_1 \cdot e^{x_2 t}$$

describing, e.g., radioactive decay, then passing this function through the N points yields a generally inconsistent system of N equations in two variables. The error function could be of the form

$$F(x) = \sum_{i=1, \dots, N} (d_i - x_1 \cdot e^{x_2 t_i})^2.$$

The nonlinear least squares problem here is to find minimizing points $x^* = (x_1^*, x_2^*)$ of $F(x)$. At these points the derivative of $F(x)$ must be equal to zero. One can use Theorems 3 and 4 with $G = 0$ to characterize zero-derivative points.

Interesting approximations of data by exponential functions were described in [9] and [10]. The measurements were done *in vivo* on patients using positron emission tomography after radioactive tissue tracer was administered into the patients’ “vascular space”. Also the influence of scan intervals on improvement of “parameter estimates” (x^* in minimization of the error function) was studied.

In the context of input optimization [16] the problem of finding “best scanning intervals” can be identified as a problem of finding “optimal inputs”. Indeed, in input optimization some or all data are identified as a vector input θ . The error function is thus formulated in terms of x and θ , i.e., as some $F = F(x, \theta)$. The problem is to find a particular θ^* which minimizes the optimal value function $F^\circ(\theta) = F(x^*(\theta), \theta)$ subject to constraints imposed on x and θ . Here $x^*(\theta)$ is an optimal solution of $F(x, \theta)$ for a given θ . This $x^*(\theta)$ can be calculated by a method of “robust optimization” [3], which may take into consideration stochastic nature of data. Characterizations of “optimal inputs” θ^* over “regions of stability” for convex parametric models are formulated in [16] using suitable Lagrange functions. The study of optimal inputs is important in many applied areas including designs of experiments [5].

6. Conclusion

We have studied some of the basic problems in mathematical modeling using twice continuously differentiable functions. One of these problems is how to characterize the derivative and its zero points. We have found these characterizations using a separating property of functions. The new results are based on the global quadratic envelope property of functions and they do not explicitly require infinitesimal calculus. One can use them to verify solutions of mathematical models which use derivatives. They can also be used to study the model's sensitivity to data and, in some cases, to reformulate the model. The results are applied to the Cobb-Douglas production model and to linear and nonlinear regression. In these applications we have obtained equivalent, but geometrically different, characterizations of partial derivatives of the Cobb-Douglas function and we have characterized least squares solutions in linear regression.

Acknowledgement

The author is indebted to his colleagues for their constructive comments. In particular, Adi Ben-Israel made valuable remarks on Theorem 6 which clarified and improved its presentation. The author is indebted to Mirko Dikšić for many discussions of the research topics related to regression which were implemented at Brain Imaging Center at McGill University. He has also provided the author with related references. Biagio Ricceri has communicated to the author his latest research on well-posed problems. Also many thanks go to Ivan Soldo for his patience and technical support.

References

- [1] Barkey, D. D. (1984). *Calculus*. Saunders College Publishing. Philadelphia.
- [2] Ben-Israel, A and Greville T. N. (2013). *Generalized Inverses: Theory and Applications*. CMS Books in Mathematics. Second edition. Springer.
- [3] Ben-Tal, A. and Nemirovski A. (1998). Robust convex optimization. *Math. Oper. Res.*, 23, 769–805.
- [4] Bertsekas, D. P. (1979). Convexification procedures and decomposition methods for nonconvex optimization problems. *J. Optim. Theory Appl.*, 29, 169–197.
- [5] Box, G. E. P. and Lucas H. L. (1959). Design of experiments in nonlinear situations. *Biometrika*, 46, 77–99.
- [6] Crnjac, M., Jukić, D. and Scitovski, R. (1994). *Matematika*. Faculty of Economics. University of Osijek.
- [7] Floudas, C. A. and Gounaris, C. E. (2009). An overview of advances in global optimization during 2003-2008. A chapter in the book *Lectures on Global Optimization*. Pardalos, P. M. and Coleman, T. F. (Eds.). Fields Institute Communications. Vol. 55, 105–154.
- [8] Golan, J. S. (2011). *The Linear Algebra a Beginning Graduate Student Ought to Know*. Third edition. Springer.
- [9] Jovkar, S., Evans, A. C., Dikšić, M., Nakai, H., and Yamamoto, Y. L. (1989). Minimization of parameter estimation errors in dynamic PET: choice of scanning schedules. *Phys. Med. Biol.*, 34, 895–908.

- [10] Kato, A., Dikšić, M., Yamamoto, Y. L., Strother, S. C. and Feindel, W. (1984). An improved approach for measurement of regional cerebral rate constants in the deoxy glucose method with positron emission tomography. *J. Cerebral Blood Flow and Metabolism*, 4, 555–563.
- [11] Lukač, Z. and Neralić, L. (2012). *Operacijska Istraživanja*. Zagreb: Element.
- [12] Neralić, L. and Šego, B. (2013). *Matematika*. Second edition. Zagreb: Element.
- [13] Ricceri, B. (2007). The problem of minimizing locally a C^2 functional around non-critical points is well posed. *Proceedings of the American Mathematical Society*, 135, 2187–2191.
- [14] Ricceri, B. (2012). A strict minimax inequality criterion and some of its consequences. *Positivity*, 16, 455–470.
- [15] Zlobec, S. (1970). An explicit form of the Moore-Penrose inverse of an arbitrary complex matrix. *SIAM Review*, 12, 132–134.
- [16] Zlobec, S. (2001). *Stable Parametric Programming*. Boston: Kluwer.
- [17] Zlobec, S. (2010). Characterizing zero-derivative points. *J. Global Optim.*, 46, 155–161.
- [18] Zlobec, S. (2011). Alternative formulations of the gradient. *J. Global Optim.*, 46, 549–553.
- [19] Zlobec, S. (2014). Separating property of continuously differentiable functions. To appear in *Math. Commun.*