Journal of Uncertain Systems
Vol.7, No.3, pp.164-175, 2013
Online at: www.jus.org.uk

How to Explain Usefulness of Different Results when Teaching Calculus: Example of the Mean Value Theorem

Olga Kosheleva*
Department of Teacher Education, University of Texas at El Paso, El Paso, Texas 79968, USA

Received 1 April 2012; Revised 16 April 2012

Abstract

Students learn better when they understand why they need to learn the specific material, why this material is useful in their own discipline. In this paper, we show how to explain usefulness of simple calculus results to engineering and science students – on the example of such a seemingly theoretical result as the Mean Value Theorem.

Keywords: teaching calculus, Mean Value Theorem, interval computations

1 Introduction

Explaining usefulness is important. Many students come to college to become engineers, scientists, etc. When a student wants to be a civil engineer, this desire motivates the student to study civil engineering courses; when a student wants to be a physicist, this desire motivates the student to study physics courses.

However, to get a degree in Civil Engineering, it is not enough to only study Civil Engineering courses: a student also needs to study several courses in mathematics and physics. There is a reason why these courses are needed: the material studied in these courses provides a background needed for the Civil Engineering material. The problem is that when the students take the corresponding classes – such as calculus – they often do not have a clear understanding of why this material is useful for their profession. This lack of understanding inhibits their learning, and makes it harder to teach calculus to engineers and scientists.

To make this teaching easier – and to make students more willing to study – it is therefore desirable to explain to them (as clearly as possible) why the corresponding calculus material is important for their chosen profession.

The more details, the better. As we have just mentioned, it is desirable that the students understand why calculus is important for their profession. Once the students get this understanding, they become more eager to learn. However, often, this eagerness only applies to those parts of the material that the students perceive as useful, while about other parts of the material – for which the students do not have a good understanding of how useful these parts are – the students remain unmotivated.

It is therefore desirable to provide not only the explanation of why the course as a whole is useful, it is also desirable to provide students with an explanation of why each piece of the material taught in this course is useful.

What we do in this paper: general idea. In this paper, we describe the usefulness of a seemingly purely theoretical result: the Mean Value Theorem (its exact formulation is given below).

We will illustrate its usefulness on the example of practical problems related to so-called *interval computations* – a techniques for processing uncertainty in data processing in frequent situations when we do not know the probabilities of different approximation errors, we only know the upper bounds on the approximation error. In the description of the usefulness of interval computations, we will follow [2, 3, 5].

What we do in this paper: plan. The usual proof of the Mean Value Theorem is based on the Rolle's Theorem, which is, in turn, based on the basic result that when a differentiable function attains a local

*Corresponding author. Email: olgak@utep.edu (O. Kosheleva).

minimum or a local maximum, then its derivative at this minimum or maximum point is 0. Because of this dependence, we will start with this basic result; then, we turn to the Rolle's Theorem and after that, to the Mean Value Theorem.

2 Basic Result: the Value of the Derivative at Minima and Maxima

Why it is important to study maxima and minima: main reason. In practice, we are interested in making the best possible decisions. For example, when we invest money, we would like to get the largest possible return on the investment. When we design a bridge, we would like to find a design that – within the given specifications on its stability, durability, etc. – will cost the smallest amount of money. When we design a computer chip, we often want to make sure that its computing rate (as estimated, e.g., by the number of basic arithmetic operations per second) is the largest possible. When a GPS system makes a route for the emergency vehicle to reach the site of an accident, we want to find the path that would take the shortest possible time.

In principle, we can make many different decisions. Let $x = (x_1, \ldots, x_n)$ be a list of all the parameters that are needed to describe a decision. For example, in the financial investment case, x_i may be the proportion of the original investment amount that will be invested in the i-th financial instrument. We assume that, once we know the decision x, we can determine the outcome of this decision; let us denote the resulting outcome by $f(x) = f(x_1, \ldots, x_n)$. In these terms, our goal is to find the values of the parameters $x_1, \ldots, x_n$ that lead to the largest possible (or, if appropriate, smallest possible) value of this function $f(x_1, \ldots, x_n)$.

Why it is important to study maxima and minima: need to take uncertainty into account. The above main reason why optimization is useful is well known. However, there is another – less well known – reason why studying minima and maxima is very important for engineering and scientific applications. This reason is related to the fact that in the above motivations, we assumed that we can set the exact values of the corresponding parameters and, when these values are set, we can determine the exact outcome of the corresponding decision.

In practice, there is uncertainty. While it is possible to allocated a fixed amount of money into a certain type of investment, it is not that easy to set an exact size of a bridge width: the actual size may be somewhat different, we need to also set up a tolerance describing how accurately we need to maintain the original sizes: $\pm$ 10 cm, $\pm$ 1 cm? Similarly, the outcome may depend not only on the parameters describing our decision, but also on the parameters describing the outside world – and these parameters are also frequently only known with uncertainty.

For example, suppose that we want to design the trajectory of a spaceship that is intended to fly an automatic mission to the Moon. In principle, we know the equations that describe the motion of the spaceship. So, once we know the original orientation and velocity of the spaceship and we know the values of the parameters (e.g., the atmospheric density at different heights), we can predict where the spaceship will be and thus, check whether it will hit the Moon. In reality, it is difficult to maintain the exact values of the initial orientation and velocity, and we only know the atmospheric parameters with some uncertainty.

Interval uncertainty: a practically important situation. Uncertainty means that instead of the exact values of the corresponding quantities x_i, we only know the approximate values $\widetilde{x}_i$. In engineering and science courses, when students deal with uncertainty, it is usually assumed that we know the probability of different possible values of the corresponding approximation errors $\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i$. In practice, however, we often only know the upper bounds on these errors, i.e., the values Δ_i for which $|\Delta x_i| \le \Delta_i$.

For example, if we require that the bridge width is 10 m, with tolerance $\pm$ 10 cm (= 0.1 m), this means that the actual width of the corresponding bridge part can be anywhere between $10 - 0.1 = 9.9$ m and $10 + 0.1 = 10.1$ m. In general, the actual (unknown) value of the corresponding quantity x_i can take any value from the interval $\mathbf{x_i} \stackrel{\text{def}}{=} [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$.

Taking into account interval uncertainty necessitates computing minima and maxima. Under interval uncertainty, we want to make sure that for all possible values $x_i \in [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$, the corresponding value $f(x_1, \ldots, x_n)$ is within the given bounds $\underline{f}$ and $\overline{f}$ (e.g., that the spaceship hits the Moon and not flies past it). In other words, we need to check that for all possible values $x_i \in [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$, we have

$$\underline{f} \le f(x_1, \ldots, x_n) \le \overline{f}.$$

To check that *all* the values $f(x_1,\ldots,x_n)$ are smaller than or equal to $\overline{f}$, it is sufficient to check that the *largest* of these values does not exceed $\overline{f}$, i.e., that

$$\max_{x_i\in\mathbf{x}_i} f(x_1,\ldots,x_n)\le\overline{f}.$$

Similarly, to check that *all* the values $f(x_1,\ldots,x_n)$ are larger than or equal to $\underline{f}$, it is sufficient to check that the *smallest* of these values does not exceed $\overline{f}$, i.e., that

$$\min_{x_i\in\mathbf{x}_i} f(x_1,\ldots,x_n)\ge\underline{f}.$$

In other words, to check the desired condition, it is sufficient to find the minimum and the maximum of the function $f(x_1,\ldots,x_n)$ on the box $\mathbf{x}_1\times\cdots\times\mathbf{x}_n$, and then check whether the range

$$\mathbf{y}=f(\mathbf{x}_1,\ldots,\mathbf{x}_n)\stackrel{\rm def}{=}\{f(x_1,\ldots,x_n):x_1\in\mathbf{x}_1,\ldots,x_n\in\mathbf{x}_n\}=\left[\min_{x_i\in\mathbf{x}_i} f(x_1,\ldots,x_n),\max_{x_i\in\mathbf{x}_i} f(x_1,\ldots,x_n)\right]$$

is indeed contained in the desired interval $\left[\underline{f},\overline{f}\right]$.

Comment. From this viewpoint, if we cannot compute the exact range $\mathbf{y}$, it is desirable to at least compute the *enclosure* for this range, i.e., an interval $\mathbf{Y}\supseteq\mathbf{y}$. Indeed, if this enclosing (wider) interval $\mathbf{Y}$ is contained in $\left[\underline{f},\overline{f}\right]$, then we can be sure that the actual range $\mathbf{y}$ is also contained there.

3 How to Find Minima and Maxima

Now that we understand why finding minima and maxima is important, let us recall how calculus helps in this task. It is well known that for functions of one variable, this can be reduced to a problem of finding zeroes of the derivative. Let us recall the corresponding result and its proof – since this result is used in proving the Mean Value Theorem, the main topic of this paper.

Definition 1. *By the* derivative $f'(x)$ *of a function* $f(x)$ *at a point* x *we mean the limit*

$$\lim_{\Delta x\to 0}\frac{f(x+\Delta x)-f(x)}{\Delta x}.$$

Definition 2. *A function* $f(x)$ *is called* differentiable *at a point* x *if has a derivative at this point.*

Definition 3. *We say that a function* $f(x)$ *attains a* local maximum *at the point* x *if there exists a real number* $\varepsilon>0$ *such that* $f(x)\ge f(x')$ *for all* $x'\in(x-\varepsilon,x+\varepsilon)$.

Definition 4. *We say that a function* $f(x)$ *attains a* local minimum *at the point* x *if there exists a real number* $\varepsilon>0$ *such that* $f(x)\le f(x')$ *for all* $x'\in(x-\varepsilon,x+\varepsilon)$.

Basic Result. (case of one variable) *If a function* $f(x)$ *has a local minimum or a local maximum at a point* x*, and it is differentiable at this point* x*, then* $f'(x)=0$.

Proof. Let us proof this result for the case when the function $f(x)$ attains a local maximum; the local minimum case can be handled similarly.

For the local maximum case, by definition of the local maximum, for sufficiently small Δx, we have $f(x)\ge f(x+\Delta x)$, and hence, $f(x+\Delta x)-f(x)\le 0$.

For $\Delta x>0$, the division by Δx does not change the sign of the inequality. So, the ratio

$$\frac{f(x+\Delta x)-f(x)}{\Delta x}$$

is also non-positive. Thus, when $\Delta x\to 0$, the limit $f'(x)$ is non-positive: $f'(x)\le 0$.

For $\Delta x<0$, the division by Δx changes the sign of the inequality. So, the ratio

$$\frac{f(x+\Delta x)-f(x)}{\Delta x}$$

is non-negative. Thus, when $\Delta x \to 0$, the limit $f'(x)$ is non-negative: $f'(x) \ge 0$.

From $f'(x) \le 0$ and $f'(x) \ge 0$, we can now conclude that $f'(x) = 0$. The result is proven.

How this result can be used in optimization: case of a single variable. If we have no restrictions on the value x and we are interested in finding the optimum of a function $f(x)$, then it is sufficient to find the value x for which $f'(x) = 0$. This fact often simplifies the optimization problem.

This result can also be used to find the optimum of a function $f(x)$ on a given interval $[\underline{x}, \overline{x}]$. Indeed, the desired optimum is attained either at one of the endpoints $\underline{x}$ or $\overline{x}$ – or at an interior point. If the optimum is attained at an interior point, this means that it is either a local minimum or a local maximum of the function $f(x)$; thus, we must have $f'(x) = 0$.

So, to find the minimum (or maximum) of a given differentiable function $f(x)$ on a given interval $[\underline{x}, \overline{x}]$, it is sufficient to compute the values of the function at the two endpoints and at all the points x at which $f'(x) = 0$. The largest of these values is the maximum, the smallest of these values is the minimum.

Example. Let us use this idea to find the minimum and the maximum of a function $f(x) = x^2$ on the interval $[-2, 1]$. For the endpoints $x = 1$ and $x = -2$, we get values $f(1) = 1^2 = 1$ and $f(-2) = (-2)^2 = 4$.

The equation $f'(x) = 2x = 0$ has only one solution: $x = 0$. At this point $x = 0$, we have $f(0) = 0^2 = 0$. Now, we have computed three values of $f(x)$: 0, 1, and 4. The smallest of these three values is $\min(0, 1, 4) = 0$, the largest of these three values is $\max(0, 1, 4) = 4$. Thus, we can conclude that the smallest value of the given function $f(x) = x^2$ on the interval $[-1, 2]$ is equal to 0, while its largest value is equal to 4.

How this result can be used in optimization: case of several variables. If a function $f(x_1, \ldots, x_n)$ attains local maximum at some point, this means that changing all the values slightly can only decrease the value of this function (or keep the same). In particular, this is true if we only change the value of one variable. This change corresponds to the notion of a partial derivative:

Definition 5. *By the i-th* partial derivative $\dfrac{\partial f}{\partial x_i}$ *of a function $f(x_1, \ldots, x_n)$ at a point $x = (x_1, \ldots, x_n)$ we mean the limit*

$$\lim_{\Delta x_i \to 0} \frac{f(x_1, \ldots, x_{i-1}, x_i + \Delta x_i, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n)}{\Delta x_i}.$$

Basic Result. (case of several variables) *If a function $f(x_1, \ldots, x_n)$ has a local minimum or a local maximum at a point x, and it is differentiable at this point x, then all its partial derivatives at this point are equal to 0: $\dfrac{\partial f}{\partial x_i} = 0$ for all i.*

Because of this, if we are interested in finding the maximum or minimum of a function $f(x_1, \ldots, x_n)$ on a given box $[\underline{x}_1, \overline{x}_1] \times \cdots \times [\underline{x}_n, \overline{x}_n]$, then at the point x where this optimum is attained, for each i, one of the following three conditions is satisfied:

- $x_i = \underline{x}_i$;
- $x_i = \overline{x}_i$;
- $\dfrac{\partial f}{\partial x_i} = 0$.

Thus, to find the desired optimum, it is sufficient to try all possible combinations of such conditions. For each variable, we need to check at least two conditions, so overall, we need at least to check 2^n combinations $(x_1, \ldots, x_n)$:

- for $n = 1$, we have two possible values $\underline{x}_1$ and $\overline{x}_1$;
- when we add the second variable, to each of these combinations we add $\underline{x}_2$ or $\overline{x}_2$, so we end up with $2 \times 2 = 2^2 = 4$ possible combinations $(\underline{x}_1, \underline{x}_2)$, $(\underline{x}_1, \overline{x}_2)$, $(\overline{x}_1, \underline{x}_2)$, and $(\overline{x}_1, \overline{x}_2)$;
- when we add the third variable, to each of these combinations we add $\underline{x}_3$ or $\overline{x}_3$, so we end up with $2^2 \times 2 = 2^3 = 8$ possible combinations, etc.

Case of several variables: challenge. For small n, it is possible to check all 2^n combinations. However, in many practical problems, the number n of parameters affecting the result may be in the hundreds. In this case, we need to check 2^{100} combinations. To estimate this number, we can use the known approximate equality $2^{10} = 1024 \approx 10^3$ (this is why Kilobyte is called this way although it has not 1000 bytes but 1024 ones). Due to this approximate equality, we have $2^{100} = \left(2^{10}\right)^{10} \approx \left(10^3\right)^{10} = 10^{30}$. With 10^{12} operations per second, we need $10^{30}/10^{12} = 10^{18}$ seconds. At $\approx 3 \cdot 10^7$ seconds per year, this means that we need $10^{18}/(3 \cdot 10^7) \approx 3 \cdot 10^{10}$ years, 30 billion years, which is longer that the lifetime of the Universe.

This shows that it is not realistically possible to use the above calculus approach to find the exact range. Moreover, the problem of computing the exact range is NP-hard, meaning that most probably it is not possible to reduce this exponentially growing computation times; for exact definitions and proof; see, e.g., [4].

Computing the enclosure: main idea. As we have mentioned earlier, when we cannot easily compute the range exactly, it is reasonable to compute an enclosure for this range. One way to produce such an enclosure is called *straightforward interval arithmetic.*

This method is based on the fact that for the case when $f(x_1, x_2)$ is an arithmetic operation, we can use monotonicity to compute the exact range

$$f(\mathbf{x}_1, \mathbf{x}_2) = \{f(x_1, x_2) : x_1 \in \mathbf{x}_1 \text{ and } x_2 \in \mathbf{x}_2\}.$$

For example, addition $f(x_1, x_2) = x_1 + x_2$ is an increasing function of both variables, and thus:

- its largest value is attained when both x_1 and x_2 attain their largest values $x_1 = \overline{x}_1$ and $x_2 = \overline{x}_2$, and
- its smallest value is attained when both x_1 and x_2 attain their smallest values $x_1 = \underline{x}_1$ and $x_2 = \underline{x}_2$.

Thus, the range has the form

$$[\underline{x}_1, \overline{x}_1] + [\underline{x}_2, \overline{x}_2] = [\underline{x}_1 + \underline{x}_2, \overline{x}_1 + \overline{x}_2].$$

For subtraction $f(x_1, x_2) = x_1 - x_2$, the situation is similar, with the only difference that the difference increases with x_1 and decreases with x_2. Thus, e.g., to ind the largest possible value of $x_1 - x_2$, we need to take the largest possible value of x_1 and the smallest possible value of x_2. As a result, we get the following range:

$$[\underline{x}_1, \overline{x}_1] - [\underline{x}_2, \overline{x}_2] = [\underline{x}_1 - \overline{x}_2, \overline{x}_1 - \underline{x}_2].$$

For multiplication $f(x_1, x_2) = x_1 \cdot x_2$, the situation is more complex, since the product is an increasing function of x_1 when $x_2 \geq 0$ and a decreasing function of x_1 when $x_2 \leq 0$. However, in both cases, the dependence on x_1 is linear, and a linear function always attains its maximum and minimum at the endpoints. Thus, to find the maximum and the minimum of the product on the box $[\underline{x}_1, \overline{x}_1] \times [\underline{x}_1, \overline{x}_2]$, it is sufficient to consider all $2 \cdot 2 = 4$ combinations of endpoints. Thus, we get

$$[\underline{x}_1, \overline{x}_1] \cdot [\underline{x}_2, \overline{x}_2] = [\min(\underline{x}_1 \cdot \underline{x}_2, \underline{x}_1 \cdot \overline{x}_2, \overline{x}_1 \cdot \underline{x}_2, \overline{x}_1 \cdot \overline{x}_2), \max(\underline{x}_1 \cdot \underline{x}_2, \underline{x}_1 \cdot \overline{x}_2, \overline{x}_1 \cdot \underline{x}_2, \overline{x}_1 \cdot \overline{x}_2)].$$

Division x_1/x_2 is similarly monotonic in both variables, so unless $0 \in [\underline{x}_2, \overline{x}_2]$ (in which case the ratio can take infinite value) the range can be obtained similarly, as

$$[\underline{x}_1, \overline{x}_1]/[\underline{x}_2, \overline{x}_2] = [\min(\underline{x}_1/\underline{x}_2, \underline{x}_1/\overline{x}_2, \overline{x}_1/\underline{x}_2, \overline{x}_1/\overline{x}_2), \max(\underline{x}_1/\underline{x}_2, \underline{x}_1/\overline{x}_2, \overline{x}_1/\underline{x}_2, \overline{x}_1/\overline{x}_2)].$$

For a general function $f(x_1, \ldots, x_n)$, this method uses the fact that in a computer, each computation consists of a sequence of arithmetic operations. So, to compute the enclosure, we repeat the computations forming the program f step-by-step, replacing each operation with real numbers by the corresponding operation of interval arithmetic (given above). It can be proven, by simple induction, that, as a result, we get an enclosure $\mathbf{Y} \supseteq \mathbf{y}$ for the desired range.

In some cases, this enclosure is exact. In more complex cases (see example below), the enclosure has excess width.

Example. Let us illustrate the above idea on the example of estimating the range of the function $f(x_1) = x_1 - x_1^2$ on the interval $x_1 \in [0, 0.8]$.

We start with parsing the expression for the function, i.e., describing how a computer will compute this expression; it will implement the following sequence of elementary operations:

$$r_1 := x_1 \cdot x_1;$$

$$r_2 := x_1 - r_1.$$

According to straightforward interval computations, we perform the same operations, but with *intervals* instead of *numbers*:

$$\mathbf{r}_1 := [0, 0.8] \cdot [0, 0.8] = [\min(0 \cdot 0, 0 \cdot 0.8, 0.8 \cdot 0, 0.8 \cdot 0.8), \max(0 \cdot 0, 0 \cdot 0.8, 0.8 \cdot 0, 0.8 \cdot 0.8)]$$

$$= [\min(0, 0, 0, 0.64), \max(0, 0, 0, 0.64)] = [0, 0.64];$$

$$\mathbf{r}_2 := [0, 0.8] - [0, 0.64] = [0 - 0.64, 0.8 - 0] = [-0.64, 0.8].$$

For this function, the actual range is $f(\mathbf{x}_1) = [0, 0.25]$; see Fig. 1.

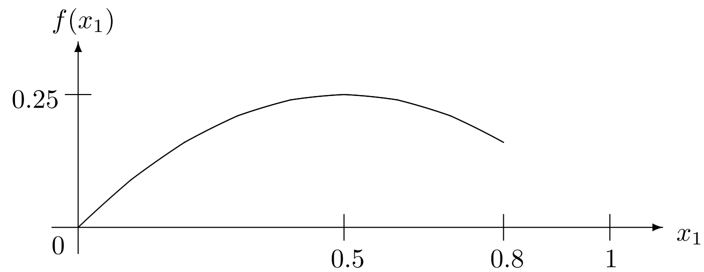

Figure 1: Range of the function $f(x_1) = x_1 - x_1^2$ on the interval $[0, 0.8]$

How can we improve this estimate? Interestingly, for that, we can use the Mean Value Theorem. Let us describe this theorem, and then let us describe how it can be used to improve the above estimates.

4 Mean Value Theorem: Reminder

The formulation of the Mean Value Theorem is as follows:

Mean Value Theorem. *Let $f(x)$ be a continuous function on the interval $[a, b]$ which is differentiable everywhere on the open interval (a, b). Then, there exists a value $\eta \in (a, b)$ for which*

$$\frac{f(b) - f(a)}{b - a} = f'(\eta).$$

Comment. Since the ratio $(f(b) - f(a))/(b - a)$ does not change if we swap a and b, the same result holds if we consider $a > b$. In other words, we can reformulate the Mean Value Theorem as follows:

Mean Value Theorem. *Let a and b be real numbers, and let a function $f(x)$ be continuous for all values x between a and b and differentiable for all values x which are strictly between a and b. Then, there exists a value η which is strictly between a and b and for which*

$$\frac{f(b) - f(a)}{b - a} = f'(\eta).$$

The Mean Value Theorem can also be equivalently reformulated as

$$f(b) = f(a) + f'(\eta) \cdot (b - a).$$

It is worth mentioning that this formula is similar to the Taylor series formula

$$f(b) = f(a) + f'(a) \cdot (b - a) + \frac{1}{2} \cdot f''(a) \cdot (b - a)^2 + \cdots ;$$

the main difference is that we use the derivative not at the point a but at some point $\eta \in (a, b)$, which helps us to make the Mean Value Theorem formula exact with the first two terms only.

From the geometric viewpoint, the ratio $(f(b)-f(a))/(b-a)$ is the slope of the secant, i.e., of a straight line connecting the points $(a, f(a))$ and $(b, f(b))$ from the function's graph, and the value $f'(\eta)$ is the slope of the tangent

$$y = f(\eta) + f'(\eta)\cdot(x-\eta)$$

to the graph. In these terms, the Mean Value Theorem claims that there exists a point η at which the tangent has the same slope as the secant, i.e., at which the tangent is parallel to the secant.

The proof of the Mean Value Theorem is usually based on the following Rolle's Theorem:

Rolle's Theorem. *For every continuous function $f(x)$ on the interval $[a,b]$ which is differentiable everywhere on the open interval (a,b) and for which $f(a)=f(b)$, there exists a point $\eta\in(a,b)$ for which $f'(\eta)=0$.*

Similarly to the Mean Value Theorem, Rolle's Theorem can be reformulated in the following form:

Rolle's Theorem. *Let a and b be real numbers, and let a function f for which $f(a)=f(b)$ be continuous for all values x between a and b and differentiable for all values x which are strictly between a and b. Then, there exists a point η which is strictly between a and b and for which $f'(\eta)=0$.*

In the following text, we will use the following multi-dimensional generalization of the Mean Value Theorem:

Mean Value Theorem. (multi-D version) *Let $a_1,\ldots,a_n$ and $b_1,\ldots,b_n$ be real numbers, and let a function $f(x_1,\ldots,x_n)$ be continuous for all values x_i which are between a_i and b_i and differentiable for all values x_i which are strictly between a_i and b_i. Then, there exists values η_i which are strictly between a_i and b_i and for which*

$$f(b_1,\ldots,b_n) = f(a_1,\ldots,a_n) + \sum_{i=1}^n \frac{\partial f}{\partial x_i}(\eta_1,\ldots,\eta_n)\cdot(b_i-a_i).$$

Proof of the Rolle's Theorem. To prove this theorem, we consider two possible cases: (1) when $f(x)$ is a constant, and (2) when $f(x)$ is not a constant.

1°. If the function $f(x)$ is a constant, i.e., if $f(x)=f(x')$ for all $x,x'\in[a,b]$, then, by definition, the derivative is always 0, so we can take any $\eta\in(a,b)$ as the desired point.

2°. If the function $f(x)$ is not a constant, this means that there exists a value $x_0\in[a,b]$ for which $f(x_0)\neq f(a)$ – because otherwise, the function $f(x)$ would be a constant. Since $f(x_0)\neq f(a)$, we have either $f(x_0)<f(a)$ or $f(x_0)>f(a)$. Let us analyze these two cases one by one.

2.1°. If $f(x_0)>f(a)$, this means that the maximum of $f(x)$ on the interval $[a,b]$ is larger than $f(a)=f(b)$ and thus, that this maximum is attained at a point $\eta\in(a,b)$. According to the basic result, at this maximum point, we have $f'(\eta)=0$.

2.2°. If $f(x_0)<f(a)$, this means that the minimum of $f(x)$ on the interval $[a,b]$ is smaller than $f(a)=f(b)$ and thus, that this minimum is attained as a point $\eta\in(a,b)$. According to the basic result, at this minimum point, we have $f'(\eta)=0$.

In all possible cases, we have found a value $\eta\in(a,b)$ for which $f'(\eta)=0$. The statement is proven.

Proof of the 1-D Mean Value Theorem. To prove this result, let us consider an auxiliary function

$$g(x) \stackrel{\text{def}}{=} f(x) - \frac{f(b)-f(a)}{b-a}\cdot(x-a).$$

One can easily check that this function is differentiable, with

$$g'(x) = f'(x) - \frac{f(b)-f(a)}{b-a}.$$

At the endpoints a and b of the interval $[a,b]$, the new function $g(x)$ takes the values

$$g(a) = f(a) - \frac{f(b)-f(a)}{b-a}\cdot(a-a) = f(a)$$

and

$$g(b) = f(b) - \frac{f(b) - f(a)}{b - a} \cdot (b - a) = f(b) - (f(b) - f(a)) = f(a).$$

So, we have $g(a) = g(b)$. Hence, due to the Rolle's Theorem, there exists a point $\eta \in (a, b)$ at which $g'(\eta) = 0$. Substituting the above expression for $g'(x)$ into this equality, we conclude that

$$f'(\eta) - \frac{f(b) - f(a)}{b - a} = 0,$$

i.e., that

$$f'(\eta) = \frac{f(b) - f(a)}{b - a}.$$

The statement is proven.

Proof of multi-D Mean Value Theorem. To prove this result, let us consider an auxiliary function

$$g(\lambda) = f(a_1 + \lambda \cdot (b_1 - a_1), \ldots, a_n + \lambda \cdot (b_n - a_n)).$$

For each $\lambda \in [0, 1]$, the corresponding values $a_i + \lambda \cdot (b_i - a_i)$ are between a_i and b_i. For $\lambda = 0$, we get $g(0) = f(a_1, \ldots, a_n)$, and for $\lambda = 1$, we get

$$g(1) = f(a_1 + (b_1 - a_1), \ldots, a_n + (b_n - a_n)) = f(b_1, \ldots, b_n).$$

The derivative of this function is equal to

$$g'(\lambda) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a_1 + \lambda \cdot (b_1 - a_1), \ldots, a_n + \lambda \cdot (b_n - a_n)) \cdot (b_i - a_i).$$

Due to the 1-D version of the Mean Value Theorem, we conclude that there exists a value $\lambda \in (0, 1)$ for which

$$g(1) = g(0) + g'(\lambda) \cdot (1 - 0) = g(0) + g'(\lambda).$$

Substituting the above expressions for $g(0)$, $g(1)$, and $g'(x)$ into this formula, we conclude that

$$f(b_1, \ldots, b_n) = f(a_1, \ldots, a_n) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a_1 + \lambda \cdot (b_1 - a_1), \ldots, a_n + \lambda \cdot (b_n - a_n)) \cdot (b_i - a_i),$$

i.e., the desired formula for $\eta_i = a_i + \eta \cdot (b_i - a_i)$. The statement is proven.

5 Usefulness of the Mean Value Theorem

At first glance. At first glance, it may seem that the Mean Value Theorem is a purely theoretical statement: indeed, all we conclude is that there exists some values η_i, without any idea on how to find these values η_i. However, as we will see, this result is actually very useful in interval computations.

The use of the Mean Value Theorem in interval computations. We start by representing each interval $\mathbf{x}_i = [\underline{x}_i, \overline{x}_i]$ in the form $[\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$, where $\widetilde{x}_i = (\underline{x}_i + \overline{x}_i)/2$ is the midpoint of the interval $\mathbf{x}_i$ and $\Delta_i = (\overline{x}_i - \underline{x}_i)/2$ is the half-width of this interval.

The Mean Value Theorem implies that

$$f(x_1, \ldots, x_n) = f(\widetilde{x}_1, \ldots, \widetilde{x}_n) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\eta_1, \ldots, \eta_n) \cdot (x_i - \widetilde{x}_i),$$

where each η_i is some value from the interval $\mathbf{x}_i$.

Since $\eta_i \in \mathbf{x}_i$, the value of the i-th derivative belongs to the interval range of this derivative on these intervals. We also know that $x_i - \widetilde{x}_i \in [-\Delta_i, \Delta_i]$. Thus, we can conclude that

$$f(\mathbf{x}_1, \ldots, \mathbf{x}_n) \subseteq f(\widetilde{x}_1, \ldots, \widetilde{x}_n) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{x}_1, \ldots, \mathbf{x}_n) \cdot [-\Delta_i, \Delta_i].$$

To compute the ranges of the partial derivatives, we can use, e.g., straightforward interval computations.

Example. Let us illustrate this method on the above example of estimating the range of the function $f(x_1) = x_1 - x_1^2$ over the interval $[0, 0.8]$. For this interval, the midpoint is $\widetilde{x}_1 = 0.4$; at this midpoint, $f(\widetilde{x}_1) = 0.24$. The half-width is $\Delta_1 = 0.4$. The only partial derivative here is $\dfrac{\partial f}{\partial x_1} = 1 - 2x_1$, its range on $[0, 0.8]$ is equal to

$$\begin{aligned} & 1 - 2 \cdot [0, 0.8] = [1, 1] - [2, 2] \cdot [0, 0.8] \\ = \ & [1, 1] - [\min(2 \cdot 0, 2 \cdot 0.8, 2 \cdot 0, 2 \cdot 0.8), \max(2 \cdot 0, 2 \cdot 0.8, 2 \cdot 0, 2 \cdot 0.8)] \\ = \ & [1, 1] - [\min(0, 1.6, 0.1.6), \max(0.1.6, 0, 1.6)] = [1, 1] - [0, 1.6] = [1 - 1.6, 1 - 0] = [-0.6, 1]. \end{aligned}$$

Thus, we get the following enclosure $\mathbf{Y} \supseteq \mathbf{y}$ for the desired range $\mathbf{y}$ of the function $f(x_1)$ on the interval $[0, 08]$:

$$\mathbf{Y} = 0.24 + [-0.6, 1] \cdot [-0.4, 0.4] = 0.24 + [-0.4, 0.4] = [-0.16, 0.64].$$

This enclosure is narrower than the straightforward enclosure $[-0.64, 0.8]$, but it still contains excess width.

How can we get better estimates? In this approach, we, in effect, ignored quadratic and higher order terms, i.e., terms of the type $\dfrac{\partial^2 f}{\partial x_i \partial x_j} \cdot \Delta x_i \cdot \Delta x_j$. When the estimate is not accurate enough, it means that this ignored term is too large. There are two ways to reduce the size of the ignored term:

- we can try to decrease this quadratic term, or
- we can try to explicitly include higher order terms in the Taylor expansion formula, so that the remainder term will be proportional to say Δx_i^3 and thus, be much smaller.

Let us describe these two ideas in detail.

First idea: bisection. Let us first describe the situation in which we try to minimize the second-order remainder term. In the above expression for this term, we cannot change the second derivative. The only thing we can decrease is the difference $\Delta x_i = x_i - \widetilde{x}_i$ between the actual value and the midpoint. This value is bounded by the half-width Δ_i of the box. So, to decrease this value, we can subdivide the original box into several narrower subboxes. Usually, we divide into two subboxes, so this subdivision is called *bisection*.

The range over the whole box is equal to the union of the ranges over all the subboxes. The widths of each subbox are smaller, so we get smaller Δx_i and hopefully, more accurate estimates for ranges over each of this subbox. Then, we take the union of the ranges over subboxes.

Example. Let us illustrate this idea on the above $x_1 - x_1^2$ example. In this example, we divide the original interval $[0, 0.8]$ into two subintervals $[0, 0.4]$ and $[0.4, 0.8]$. For both intervals, $\Delta_1 = 0.2$.

In the first subinterval, the midpoint is $\widetilde{x}_1 = 0.2$, so $f(\widetilde{x}_1) = 0.2 - 0.04 = 0.16$. The range of the derivative is equal to

$$1 - 2 \cdot [0, 0.4] = 1 - [0, 0.8] = [0.2, 1],$$

hence we get an enclosure

$$0.16 + [0.2, 1] \cdot [-0.2, 0.2] = [-0.04, 0.36].$$

For the second interval, $\widetilde{x}_1 = 0.6$, $f(0.6) = 0.24$, the range of the derivative is

$$1 - 2 \cdot [0.4, 0.8] = [-0.6, 0.2],$$

hence we get an enclosure

$$0.24 + [-0.6, 0.2] \cdot [-0.2, 0.2] = [0.12, 0.36].$$

The union of these two enclosures is the interval $[-0.04, 0.36]$. This enclosure is much more accurate than before.

Further bisection leads to even more accurate estimates – the smaller the subintervals, the more accurate the enclosure.

Bisection: general comment. The more subboxes we consider, the smaller Δx_i and thus, the more accurate the corresponding enclosures. However, once we have more boxes, we need to spend more time processing these boxes. Thus, we have a trade-off between computation time and accuracy: the more computation time we allow, the more accurate estimates we will be able to compute.

Additional idea: monotonicity checking. If the function $f(x_1, \ldots, x_n)$ is monotonic over the original box $\mathbf{x}_1 \times \cdots \times \mathbf{x}_n$, then we can easily compute its exact range. Since we used the Mean Value Theorem for the original box, this probably means that on that box, the function is not monotonic: for example, with respect to x_1, it may be increasing at some points in this box, and decreasing at other points.

However, as we divide the original box into smaller subboxes, it is quite possible that at least some of these subboxes will be outside the areas where the derivatives are 0 and thus, the function $f(x_1, \ldots, x_n)$ will be monotonic. So, after we subdivide the box into subboxes, we should first check monotonicity on each of these subboxes – and if the function is monotonic, we can easily compute its range.

In calculus terms, if a function is increasing with respect to x_i, then its partial derivative $d_i \stackrel{\text{def}}{=} \dfrac{\partial f}{\partial x_i}$ is a limit of non-negative numbers and is, thus, non-negative everywhere on this subbox (we used a similar argument in the proof of the Basic Result). Vice versa, if the derivative $f'(x)$ is everywhere non-negative $f'(x) \geq 0$, then we can use the 1-D Mean Value Theorem $f(b) = f(a) + f'(\eta) \cdot (b - a)$ to prove that if $a < b$, then $f(a) \leq f(b)$, i.e., that the function is increasing.

Similarly, a differentiable function is decreasing with respect to x_i if and only if the corresponding partial derivative is everywhere non-positive.

Thus, to check monotonicity, we should find the range $\left[\underline{d}_i, \overline{d}_i\right]$ of this derivative (we need to do it anyway to use the Mean Value Theorem to estimate the expression):

- if $\underline{d}_i \geq 0$, this means that the derivative is everywhere non-negative and thus, the function f is increasing in x_i;
- if $\overline{d}_i \leq 0$, this means that the derivative is everywhere non-positive and thus, the function f is decreasing in x_i.

If $\underline{d}_i < 0 < \overline{d}_i$, then we have to use the Mean Value Theorem.

If the function is monotonic (e.g., increasing) only with respect to some of the variables x_i, then

- to compute $\overline{y}$, it is sufficient to consider only the value $x_i = \overline{x}_i$, and
- to compute $\underline{y}$, it is sufficient to consider only the value $x_i = \underline{x}_i$.

For such subboxes, we reduce the original problem to two problems with fewer variables, problems which are thus easier to solve.

Example. For the example $f(x_1) = x_1 - x_1^2$, the partial derivative is equal to $1 - 2 \cdot x_1$.

On the first subbox $[0, 0.4]$, the range of this derivative is $1 - 2 \cdot [0, 0.4] = [0.2, 1]$. Thus, the derivative is always non-negative, the function is increasing on this subbox, and its range on this subbox is equal to $[f(0), f(0.4)] = [0, 0.16]$.

On the second subbox $[0.4, 0.8]$, the range of the derivative is $1 - 2 \cdot [0.4, 0.8] = [-0.6, 0.2]$. Here, we do not have guaranteed monotonicity, so we can use the Mean Value Theorem to get the enclosure $[0.12, 0.36]$ for the range.

The union of these two enclosures is the interval $[0, 0.36]$, which is slightly more accurate than before. Further bisection leads to even narrower enclosures.

Towards General Taylor techniques. As we have mentioned, another way to get narrower enclosures is to use so-called *Taylor techniques*, i.e., to explicitly consider second-order and higher-order terms in the Taylor expansion; see, e.g., [1, 6], and references therein. Let us illustrate the main ideas of Taylor analysis on the case when we allow second order terms.

Formulas involving quadratic and higher-order terms can also be obtained based on the Mean Value Theorem. Specifically, we need a generalization of the Mean Value Theorem originally proved by Cauchy, one of the 19 century founders of modern mathematical analysis.

Cauchy Mean Value Theorem. *Let $F(x)$ and $G(x)$ be continuous on the interval $[a, b]$ and differentiable everywhere on the open interval (a, b), and let $G'(x) \neq 0$ for all $x \in (a, b)$. Then, there exists a value $\eta \in (a, b)$ for which*

$$\frac{F(b) - F(a)}{G(b) - G(a)} = \frac{F'(\eta)}{G'(\eta)}.$$

Similarly to the original Mean Value Theorem, this result can be equivalently reformulated as follows:

Cauchy Mean Value Theorem. *Let a and b be real numbers, and let functions $F(x)$ and $G(x)$ be continuous for all values x between a and b and differentiable for all values x strictly between a_i and b_i, and let $G'(x) \neq 0$ for all such x. Then, there exists a value η which is strictly between a and b and for which*

$$\frac{F(b) - F(a)}{G(b) - G(a)} = \frac{F'(\eta)}{G'(\eta)}.$$

Proof. Similarly to the proof of the original Mean Value Theorem, the proof of this result follows when we apply the Rolle's Theorem to the auxiliary function $g(x) = f(x) - h \cdot (g(x) - g(a))$, where the parameter h is chosen in such a way that $g(b) = g(a)$.

Taylor's Theorem. (1-D quadratic case) *Let $f(x)$ be continuous on the interval $[a, b]$ and twice differentiable everywhere on the open interval (a, b). Then, there exists a point $\eta \in (a, b)$ for which*

$$f(b) = f(a) + f'(a) \cdot (b - a) + \frac{1}{2} \cdot f''(\eta) \cdot (b - a)^2.$$

Proof. To prove this result, we apply Cauchy Mean Value Theorem to two auxiliary functions: $F(x) = f(x) + f'(x) \cdot (b - x)$ and $G(x) = (x - b)^2$. Here,

$$F(b) = f(b) + f'(b) \cdot (b - x) = f(b),$$

$$F(a) = f(a) + f'(a) \cdot (b - a),$$

so

$$F(b) - F(a) = f(b) - f(a) - f'(a) \cdot (b - a).$$

Similarly,

$$G(b) - G(a) = (b - b)^2 - (b - a)^2 = -(b - a)^2.$$

For the derivatives, we get

$$F'(x) = f'(x) - f'(x) + f''(x) \cdot (b - x) = f''(x) \cdot (b - x)$$

and

$$G'(x) = 2 \cdot (x - b),$$

so

$$\frac{F'(x)}{G'(x)} = \frac{f''(x) \cdot (b - x)}{2 \cdot (x - b)} = -\frac{1}{2} \cdot f''(x).$$

Thus, the Cauchy Mean Value Theorem implies there exists a value η for which

$$\frac{f(b) - f(a) - f'(a) \cdot (b - a)}{-(b - a)^2} = -\frac{1}{2} \cdot f''(\eta).$$

Multiplying both sides by the denominator of the left-hand side, we get the desired equality. The statement is proven.

By applying this result to the same auxiliary function

$$g(\lambda) = f(a_1 + \lambda \cdot (b_1 - a_1), \ldots, a_n + \lambda \cdot (b_n - a_n))$$

that we used to prove the multi-D version of the Mean Value Theorem, we conclude that

$$\begin{aligned} f(x_1, \ldots, x_n) = \quad & f(\widetilde{x}_1, \ldots, \widetilde{x}_n) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\widetilde{x}_1, \ldots, \widetilde{x}_n) \cdot (x_i - \widetilde{x}_i) \\ & + \frac{1}{2} \cdot \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{\partial^2 f}{\partial x_i \partial x_j}(\eta_1, \ldots, \eta_n) \cdot (x_i - \widetilde{x}_i) \cdot (x_j - \widetilde{x}_j). \end{aligned}$$

Thus, we get the enclosure

$$f(\mathbf{x}_1,\ldots,\mathbf{x}_n)\subseteq \mathbf{Y}\stackrel{\text{def}}{=} f(\widetilde{x}_1,\ldots,\widetilde{x}_n)+\sum_{i=1}^n \frac{\partial f}{\partial x_i}(\widetilde{x}_1,\ldots,\widetilde{x}_n)\cdot[-\Delta_i,\Delta_i]$$

$$+\frac{1}{2}\cdot\sum_{i=1}^n\sum_{j=1}^m \frac{\partial^2 f}{\partial x_i\partial x_j}(\mathbf{x}_1,\ldots,\mathbf{x}_n)\cdot[-\Delta_i,\Delta_i]\cdot[-\Delta_j,\Delta_j].$$

Example. Let us illustrate this idea on the above example of $f(x_1)=x_1-x_1^2$. Here, $\Delta_1=0.4$, $\widetilde{x}_1=0.4$, so $f(\widetilde{x}_1)=0.24$ and $\dfrac{\partial f}{\partial x_1}(\widetilde{x}_1)=1-2\cdot 0.4=0.2$. The second derivative is equal to -2, so the Taylor estimate takes the form

$$\mathbf{Y}=0.24+0.2\cdot[-0.4,0.4]-[-0.4,0.4]^2.$$

Strictly speaking, if we interpret Δx_1^2 as $\Delta x_1\cdot\Delta x_1$ and use the formulas of interval multiplication, we get the interval

$$[-0.4,0.4]^2=[-0.4,0.4]\cdot[-0.4,0.4]=[-0.16,0.16]$$

and thus, the enclosure

$$\mathbf{Y}=0.24+[-0.08,0.08]-[-0.16,0.16]=[0.16,0.32]-[-0.16,0.16]=[0,0.48]$$

for the desired range. However, we can view x^2 as a special function, for which the range over $[-0.4,0.4]$ is known to be $[0,0.16]$. In this case, the above enclosure takes the form

$$\mathbf{Y}=0.24+[-0.08,0.08]-[0,0.16]=[0.16,0.32]-[0,0.16]=[0,0.32]$$

which is much closer to the actual range $[0,0.25]$.

Taylor methods: general comment. The more terms we consider in the Taylor expansion, the smaller the remainder term and thus, the more accurate the corresponding enclosures. However, once we have more terms, we need to spend more time computing these terms. Thus, for Taylor methods, we also have a trade-off between computation time and accuracy: the more computation time we allow, the more accurate estimates we will be able to compute.

References

[1] Berz, M., and G. Hoffstätter, Computation and application of Taylor polynomials with interval remainder bounds, *Reliable Computing*, vol.4, no.1, pp.83–97, 1998.

[2] Jaulin, L., M. Kieffer, O. Didrit, and E. Walter, *Applied Interval Analysis, with Examples in Parameter and State Estimation, Robust Control and Robotics*, Springer-Verlag, London, 2001.

[3] Kreinovich, V., Interval computations and interval-related statistical techniques: tools for estimating uncertainty of the results of data processing and indirect measurements, *Data Modeling for Metrology and Testing in Measurement Science*, pp.117–145, 2009.

[4] Kreinovich, V., A. Lakeyev, J. Rohn, and P. Kahl, *Computational Complexity and Feasibility of Data Processing and Interval Computations*, Kluwer, Dordrecht, 1997.

[5] Moore, R.E., R.B. Kearfott, and M.J. Cloud, *Introduction to Interval Analysis*, SIAM Press, 2009.

[6] Neumaier, A., Taylor forms, *Reliable Computing*, vol.9, pp.43–79, 2002.