# The Mean Value Inequality
# (without the Mean Value Theorem)

Michael Müger

September 13, 2017

**Abstract**

## 1 Introduction

Once one has defined the notion of differentiability of a function of one variable, it is easy to prove:

1.1 LEMMA *Assume $f : [a,b] \to \mathbb{R}$ is continuous, differentiable on $(a,b)$ and non-decreasing. Then $f'(x) \geq 0$ for all $x \in (a,b)$.*

(The function $f : \mathbb{R} \to \mathbb{R},\ x \mapsto x^3$ has $f'(0) = 0$ despite being strictly increasing. Thus the implication 'strictly increasing $\Rightarrow$ $f'(x) > 0$ for all $x \in (a,b)$' is not true.)

As to converses, the following is true, but harder to prove:

1.2 THEOREM *Let $f : [a,b] \to \mathbb{R}$ be continuous and differentiable on $(a,b)$.*

*(i) If $f'(x) \geq 0$ for all $x \in (a,b)$ then $f$ is non-decreasing (=weakly increasing).*

*(ii) If $f'(x) > 0$ for all $x \in (a,b)$ then $f$ is strictly increasing.*

*(iii) If $f'(x) = 0$ for all $x \in (a,b)$ then $f$ is constant.*

*(iv) If $m \leq f'(x) \leq M$ for all $x \in (a,b)$ then $m(x'-x) \leq f(x') - f(x) \leq M(x'-x)$ for all $x, x' \in [a,b]$.*

1.3 REMARK 1. (iv) has been called the Mean Value Inequality. Of course, (iii) is just the special case $m = M = 0$.

2. It is an understandable, yet serious, beginner's mistake to think that the above results are obvious. E.g., $f' \equiv 0$ only means that $\lim_{y \to x} \frac{f(y)-f(x)}{y-x} = 0$ for all $x$. A priori, this implies *nothing* about $f(y) - f(x)$ for finite non-zero $y - x$. Proving that such an implication actually does exist is quite non-trivial and one of the deeper results of classical analysis.

3. The standard textbook proof of the theorem uses the Mean Value Theorem (MVT): Under the given assumptions there is a $c \in (a,b)$ such that $f'(c) = \frac{f(b)-f(a)}{b-a}$. Combining this with the hypotheses on $f'$, the theorem follows immediately, despite the fact that nothing more can be said about $c$. However, dissecting this proof shows that it is quite involved: While the MVT is an easy corollary of Rolle's theorem (just as (iv) of the Theorem follows from (i)), proving Rolle's theorem is not straightforward. One combines the (easy) fact that $f'(x) = 0$ at a local extremum $x \in (a,b)$ with the (harder) result that every continuous function on a closed bounded interval is bounded and assumes its bounds. The latter in turn can be proven using abstract compactness arguments, but in an introductory course one usually invokes the Bolzano-Weierstrass theorem to the effect that bounded sequences have convergent subsequences. One way of proving this uses iterated halving of intervals. Such an argument will also be used to give one proof of part (i) of the Theorem.

4. The theorem also follows easily from the half of the Fundamental Theorem of Calculus according to which $f(b) - f(a) = \int_a^b f'(t)dt$ if $f : [a, b] \to \mathbb{R}$ is everywhere differentiable and $f'$ is Riemann integrable. (The second assumption is trivially true in case (iii), but not in the others.) However, the proof of this half of the Fundamental Theorem invariably uses the MVT, so that such an approach of proving Theorem 1.2 would be more complicated than the more direct invocation of MVT described above.

5. There has been some controversy about the rôle of the MVT in the teaching of classical analysis: Some authors, e.g. [1, 2, 3, 5, 9, 10], advocate the elimination of the MVT in favour of of the (supposedly) more natural and direct proofs of the theorem given below, while others disagree [12]. However this may be, such an approach is rarely spelt out in full detail, and doing so is the motivation for this note. We begin by reducing the proof to that of (i). □

*Proof of Theorem 1.2(ii)-(iv), assuming (i).* (ii) If $f'(x) > 0$ for all $x \in (a, b)$ then of course $f' \geq 0$, thus $f$ is non-decreasing by (i). Thus for $a \leq x < x' \leq b$ we have $f(x) \leq f(x')$. Thus if $f$ is not strictly increasing, there are $x < x'$ with $f(x) = f(x')$. But then $f$ is constant on $[x, x']$, thus $f' \equiv 0$ on $(x, x')$, contradicting the hypothesis. Thus $f$ is strictly increasing.

(iii) We have $f' \geq 0$ everywhere, so that $f$ is non-decreasing by (i). On the other hand, $(-f)' \geq 0$ everywhere, so that $-f$ is non-decreasing. Thus $f$ is non-increasing. Combining, we get that $f$ is constant.

(iv) Consider the auxiliary function $g(x) = f(x) - mx$. Then $g' = f' - m$, which is non-negative by the assumption. Thus $g$ is non-decreasing by (i), so that $x < x'$ implies $f(x) - mx \leq f(x') - mx'$. This is equivalent to $f(x') - f(x) \geq m(x' - x)$, which is one of the desired inequalities. The other one is proven analogously by considering the auxiliary function $h(x) = Mx - f(x)$, which satisfies $h' = M - f' \geq 0$ and therefore is non-decreasing. ■

## 2 First proof of (i)

*Proof.* Assume that $f$ is not non-decreasing. Then there are $x_1, x_2$ with $a \leq x_1 < x_2 \leq b$ such that $f(x_2) < f(x_1)$. If necessary, we can use the continuity of $f$ to change $x_1, x_2$ slightly so that $a < x_1 < x_2 < b$ and $f(x_2) < f(x_1)$. Thus $f$ is differentiable at $x_1, x_2$. Now $\lambda := \frac{f(x_2)-f(x_1)}{x_2-x_1} < 0$. The set

$$A = \left\{ x \in (x_1, x_2] \mid \frac{f(x) - f(x_1)}{x - x_1} \leq \lambda \right\}$$

is bounded below (by $x_1$) and non-empty (since $x_2 \in A$), thus $x^* = \inf(A)$ exists (order completeness of $\mathbb{R}$). In view of $\lambda < 0 \leq f'(x_1) = \lim_{x \searrow x_1} \frac{f(x)-f(x_1)}{x-x_1}$, we have $\frac{f(x)-f(x_1)}{x-x_1} > \lambda$ for all $x \in (x_1, x_2]$ close enough to $x_1$. Thus $x^* > x_1$. If $\{y_n\} \subseteq A$ is a sequence converging to $x^*$ then $f(y_n) - f(x_1) \leq \lambda(y_n - x_1)$ for all $n$, and taking $n \to \infty$ gives

$$f(x^*) - f(x_1) \leq \lambda(x^* - x_1). \tag{1}$$

(Thus $x^* \in A$.) If now $x_1 < x' < x^*$ (here we use $x_1 < x^*$) then $x' \notin A$, thus $f(x') - f(x_1) > \lambda(x' - x_1)$, which is equivalent to $f(x_1) - f(x') < \lambda(x_1 - x')$. Adding this to (1), we obtain

$$f(x^*) - f(x') < \lambda(x^* - x') \qquad \forall x' \in (x_1, x^*). \tag{2}$$

But in view of $\lambda < 0 \leq f'(x^*) = \lim_{x' \nearrow x^*} \frac{f(x^*)-f(x')}{x^*-x'}$, for $x' \in (x_1, x^*)$ close enough to $x^*$ we have $\frac{f(x^*)-f(x')}{x^*-x'} > \lambda$, which contradicts (2). ■

The above proof is similar to those in [9] and in [1] (where the hypotheses are $f' = 0$ and $f' > 0$, respectively).

## 3 Second proof of (i), by interval dissection

3.1 DEFINITION *If $f : [a, b] \to \mathbb{R}$ is a function and $a \le c < d \le b$, we call $\frac{f(d)-f(c)}{d-c}$ the* inclination *of $f$ on the interval $[c, d]$.*

*Proof.* Assume that (i) is false. Then there are $c_0, d_0$ such that $a \le c_0 < d_0 \le b$ such that $f(d_0) < f(c_0)$. This means that the inclination $\lambda_0$ of $f$ on $[c_0, d_0]$ is negative. Define $e_0 = (c_0 + d_0)/2$. Assume that the inclination of $f$ on both intervals $[c_0, e_0]$ and $[e_0, d_0]$ is larger than $\lambda_0$, to wit $\frac{f(e_0)-f(c_0)}{e_0-c_0} > \lambda_0$ and $\frac{f(d_0)-f(e_0)}{d_0-e_0} > \lambda_0$. In view of $c_0 < e_0 < d_0$, this is equivalent to $f(e_0) - f(c_0) > \lambda_0(e_0 - c_0)$ and $f(d_0) - f(e_0) > \lambda_0(d_0 - e_0)$. Adding these inequalities gives $f(d_0) - f(c_0) > \lambda_0(d_0 - c_0)$, which is equivalent to $\lambda_0 = \frac{f(d_0)-f(c_0)}{d_0-c_0} > \lambda_0$, which is absurd. Thus one of the two inclinations must be $\le \lambda_0$. [A more geometric argument goes like this: If $f(e_0) = \frac{f(c_0)+f(d_0)}{2}$ then the inclinations of $f$ on $[c_0, e_0]$ and $[e_0, d_0]$ both equal $\lambda_0$. If $f(e_0) > \frac{f(c_0)+f(d_0)}{2}$ then the inclinations of $f$ on $[c_0, d_0]$ and $[e_0, d_0]$ are $> \lambda_0$ and $< \lambda_0$, respectively. For $f(e_0) < \frac{f(c_0)+f(d_0)}{2}$, the converse holds.] If this is true for $[c_0, e_0]$, we put $c_1 = c_0, d_1 = e_0$, otherwise we put $c_1 = e_0, d_1 = d_0$. It is clear that this construction can be iterated, giving a sequence $\{[c_n, d_n]\}$ of intervals such that $d_n - c_n = 2^{-n}(d_0 - c_0)$ and such that the inclination of $f$ is $\le \lambda_0$ on all intervals $[c_n, d_n]$. The sequence $\{c_n\}$ is non-decreasing and bounded above (by $d_0$) and therefore converges to some $x \in [a, b]$. In view of $d_n - c_n = 2^{-n}(d_0 - c_0)$, we also have $d_n \to x$. (In fact, $\{x\} = \bigcap_n [c_n, d_n]$.) Since we have $c_n \le x \le d_n$ for all $n$ and $d_n - c_n \searrow 0$, Lemma 3.2 below gives that the inclination $\lambda_n$ of $f$ on $[c_n, d_n]$ converges to $f'(x)$. By construction, we have $\lambda_n \le \lambda_0 < 0$ for all $n$, implying $f'(x) \le \lambda_0 < 0$. This, however, contradicts the assumption $f' \ge 0$, proving that Theorem 1.2(i) is true. ■

3.2 LEMMA *Assume $f : [a, b] \to \mathbb{R}$ is differentiable at $x$ (one-sided differentiability if $x \in \{a, b\}$). Let $\{u_n\}, \{v_n\}$ be sequences satisfying $a \le u_n \le x \le v_n \le b$ and $v_n - u_n > 0$ for all $n$ and $v_n - u_n \to 0$. Then*

$$\lim_{n\to\infty} \frac{f(v_n) - f(u_n)}{v_n - u_n} = f'(x).$$

*Proof.* Since $f$ is differentiable at $x$, there is a function $g$ such that $g(z) = o(z)$ (i.e. $\lim_{z\to 0} \frac{g(z)}{z} = 0$) and $f(y) = f(x) + (y - x)f'(x) + g(y - x)$ for all $y \in [a, b]$. Then

$$f(v_n) - f(u_n) = (v_n - u_n)f'(x) + g(u_n - x) - g(v_n - x).$$

Since $v_n - u_n > 0$ for all $n$, this gives

$$\frac{f(v_n) - f(u_n)}{v_n - u_n} = f'(x) + \frac{g(u_n - x) - g(v_n - x)}{v_n - u_n}. \tag{3}$$

The assumption $u_n \le x \le v_n$ implies $|u_n - x| \le v_n - u_n$ and $|v_n - x| \le v_n - u_n$, thus in particular $u_n \to x$ and $v_n \to x$. Combining these facts with $g(z) = o(z)$ we have

$$\lim_{n\to\infty} \frac{|g(u_n - x)|}{v_n - u_n} \le \lim_{n\to\infty} \frac{|g(u_n - x)|}{|u_n - x|} = 0.$$

In the same way we prove $\lim_{n\to\infty} \frac{g(v_n-x)}{v_n-u_n} = 0$, and in view of (3) we have proven the claim. ■

3.3 REMARK 1. If one considers the lemma as obvious (as some authors seem to do, e.g. [13]), the second proof of Theorem 1.2(i) is more 'natural' and intuitive than the first. But if one takes the proof of the lemma into account, the first proof is shorter.

2. Note that neither the statement of the lemma nor the proof assumes that $u_n \ne x$ or $v_n \ne x$. This would be inconsistent with the other assumptions when $x \in \{a, b\}$. □

## 4 Third approach

In this approach, we first prove (ii) of the Theorem and then deduce (i).

*Proof.* We will first prove that $f'(x) > 0 \ \forall x \in (a,b)$ implies that $f$ is non-decreasing. (This statement is weaker than either of (i) or (ii).)

To this purpose, assume $f' > 0$ and assume that $f$ is not non-decreasing. Then there are $x_1, x_2$ with $a \le x_1 < x_2 \le b$ such that $f(x_2) < f(x_1)$. Choose any $t$ satisfying $f(x_2) < t < f(x_1)$. The set $S = \{x \in [x_1, x_2] \mid f(x) \ge t\}$ is bounded above and non-empty (since $x_1 \in S$), thus we can put $s = \sup(S)$. We have $f(x) < t$ for all $x > s$, and there are $x < s$ arbitrarily close to $s$ with $f(x) \ge t$, by definition of the supremum. Thus continuity of $f$ implies $f(s) = t$. [Note that we just have reproven the Intermediate Value Theorem.] Also by continuity, there is an $\varepsilon > 0$ such that $f(x) < t$ for all $x \in (x_2 - \varepsilon, x_2]$. Thus $s < x_2$, so that $(s, x_2] \ne \emptyset$. Similarly, $s > x_1 \ge a$, so that $f$ is differentiable at $s$. For $u \in (s, x_2]$ we have $f(u) < t = f(s)$, thus $\frac{f(u)-f(s)}{u-s} < 0$. Thus $f'(s) = \lim_{u \searrow s} \frac{f(u)-f(s)}{u-s} \le 0$, contradicting the assumption $f' > 0$.

If $f' > 0$, the statement just proven implies (ii) in exactly the same way as earlier. Now, assume $f' \ge 0$. For each $\varepsilon > 0$, the function $f(x) + \varepsilon x$ has derivative $\ge \varepsilon > 0$, and therefore is strictly increasing by (ii). Thus for $x' > x$ we have $f(x') + \varepsilon x' > f(x) + \varepsilon x$, which is equivalent to $f(x') - f(x) > -\varepsilon(x' - x)$. Taking the limit $\varepsilon \searrow 0$ gives $f(x') - f(x) \ge 0$, thus (i) holds. ∎

The above proof, which is from [14] with just a few details added, is the simplest known to the author. In the next two sections, we discuss alternative proofs that can be found in the literature.

## 5 Overview of the related literature

- As explained above, the standard way of proving Theorem 1.2 is via the MVT, deduced from Rolle's theorem, which in turn is proven using the Extreme Value Theorem (EVT): A continuous real-valued function on a closed bounded interval is bounded and assumes its bounds. While the standard proof of the EVT uses Bolzano-Weierstrass (or compactness), there are direct proofs using interval dissection. Cf. e.g. [7, p. 67-68].
- Theorem 1.2(i) has been called the Increasing Function Theorem (IFT), cf. [13]. Proofs like the one given in Section 3 can be found e.g. in [13], but not always with sufficient detail or rigor. In particular our Lemma 3.2 seems unavoidable.
- Theorem 1.2(iii) could be called the Zero Derivative Theorem. There are direct proofs of the latter using interval dissection, cf. e.g. [5] or [11, p. 206-207], or without dissection [9]. But since it is less immediate to deduce the IFT from the Zero Derivative Theorem, it seems more natural to consider the IFT as point of departure.
- Deducing Theorem 1.2 from the MVT is easy, but it is not clear that the converse is true. But one can prove the MVT by similar methods, cf. [8, Theorem 4].
- The discussion about the merits or drawbacks of the Mean Value Theorem in teaching calculus has been intense and often unnecessarily ideological, as is witnessed by the titles of the papers [1, 3, 5]. This ideological fervor is hard to understand since all this just is about alternative proofs, while outstanding mathematicians like M. Atiyah often emphasize that one only properly understands a theorem if one has seen many different proofs. (Furthermore, the title of [13] is misleading since the proofs given in this note are just as rigorous as the standard one using the MVT.)
- It is a fact that certain results, like the more general versions of l'Hôpital's theorem, are best proven using the MVT (or Cauchy's generalized MVT). Avoiding the MVT tends to lead to more restrictive assumptions, like the continuity of $f'$ imposed in [2].

- Another common application of the MVT is to proving remainder estimates for Taylor expansions. Here, however, rather satisfactory results can be proven using only the Mean Value Inequality, cf. [13, Theorem 2].
- For rather more sophisticated approaches to the results discussed in this note see [4, 6].

## References


[1] L. Bers: On avoiding the mean value theorem. Amer. Math. Monthly **74**, 583 (1967).

[2] R. P. Boas: L'Hospital's rule without mean value theorems. Amer. Math. Monthly **76**, 1051-1053 (1969).

[3] R. P. Boas: Who needs those Mean-Value-Theorems anyway? The 2-year College Math. J. **12**, 178-181 (1981).

[4] M. W. Botsko: A unified treatment of various theorems in elementary analysis. Amer. Math. Monthly **94**, 450-452 (1987).

[5] L. W. Cohen: On being mean to the Mean Value Theorem. Amer. Math. Monthly **74**, 581-582 (1967).

[6] J. J. Koliha: Mean, meaner, and the meanest mean value theorem. Amer. Math. Monthly **116**, 356-361 (2009).

[7] P. D. Lax; M. S. Terrell: *Calculus with applications*. 2nd. ed. Springer, 2014.

[8] J. Plante: A proof of Bonnet's version of the mean value theorem by methods of Cauchy. Amer. Math. Monthly **124**, 269-273 (2017).

[9] M. Powderly: A simple proof of a basic theorem of the calculus. Amer. Math. Monthly **70**, 544 (1963).

[10] D. E. Richmond: An elementary proof of a theorem in calculus. Amer. Math. Monthly **92**, 589-590 (1985).

[11] J. Stillwell: *The elements of mathematics. From Euclid to Gödel*. Princeton University Press, 2016.

[12] H. Swann: Commentary on Rethinking rigor in calculus: The role of the mean value. Amer. Math. Monthly **104**, 231-240 (1997).

[13] T. W. Tucker: Rethinking rigor in calculus: The role of the mean value theorem. Amer. Math. Monthly **104**, 241-245 (1997).

[14] A. C. M. van Rooij, W. H. Schikhof: *A second course on real functions*. Cambridge University Press, 1982.