ELSEVIER
Available online at www.sciencedirect.com
ScienceDirect
Journal of Functional Analysis 259 (2010) 2999–3024
JOURNAL OF Functional Analysis
www.elsevier.com/locate/jfa

The chain rule as a functional equation

Shiri Artstein-Avidan [a,1], Hermann König [b,*], Vitali Milman [a,2]

[a] *School of Mathematical Sciences, Tel Aviv University, Ramat Aviv, Tel Aviv 69978, Israel*
[b] *Mathematisches Seminar, Universität Kiel, 24098 Kiel, Germany*

Received 18 June 2010; accepted 7 July 2010

Available online 17 July 2010

Communicated by J. Bourgain

Abstract

We consider operators T from $C^1(\mathbb{R})$ to $C(\mathbb{R})$ satisfying the "chain rule"

$$T(f \circ g) = (Tf) \circ g \cdot Tg, \quad f, g \in C^1(\mathbb{R}),$$

and study under which conditions this functional equation admits only the derivative or its powers as solutions. We also consider T operating on other domains like $C^k(\mathbb{R})$ for $k \in \mathbb{N}_0$ or $k = \infty$ and study the more general equation $T(f \circ g) = (Tf) \circ g \cdot Ag$, $f, g \in C^1(\mathbb{R})$ where both T and A map $C^1(\mathbb{R})$ to $C(\mathbb{R})$.

Keywords: Chain rule; Functional equation

1. Introduction and results

The derivative $D : C^1(\mathbb{R}) \to C(\mathbb{R})$ satisfies the chain rule

$$D(f \circ g) = \big((Df) \circ g\big) \cdot Dg, \quad f, g \in C^1(\mathbb{R}).$$

* Corresponding author.
E-mail address: hkoenig@math.uni-kiel.de (H. König).
1 The author is supported by ISF grant 865-07 and by BSF grant 2006079.
2 The author is supported by the Alexander von Humboldt Foundation, by ISF grant 387/09 and by BSF grant 2006079.

0022-1236/$ – see front matter
doi:10.1016/j.jfa.2010.07.002

We show in this paper that the chain rule functional equation

$$T(f \circ g) = \big((Tf) \circ g\big) \cdot Tg, \quad f, g, f \circ g \in \mathcal{D}(T) \tag{1}$$

characterizes the derivative, or a power of the derivative, under very mild conditions of non-degeneracy. Here $\mathcal{D}(T)$ denotes the domain of the operation T and $\mathrm{Im}(T)$ will denote its image. Various assumptions can be made on both and the answer depends on the choice of $\mathcal{D}(T)$ and $\mathrm{Im}(T)$. We mainly consider the natural case $\mathcal{D}(T) = C^1(\mathbb{R})$ and $\mathrm{Im}(T) \subset C(\mathbb{R})$.

Let us emphasize that no continuity assumptions on T are made; however, the continuity of T follows from (1). Also the linearity of T in the derivative is a consequence of (1) and some initial condition; it is not assumed a priori. In contrast, the Leibniz product rule (for a derivation) does not characterize the derivative; it is known that without conditions of continuity or linearity, a derivation does not necessarily give the derivative.

The results of the paper may be seen in the light of some recent results on classical operations or transforms in Geometry and Analysis which were shown to be characterized by some very elementary and natural conditions like e.g. anti-monotonicity or multiplicativity. The Fourier transform was studied in [1,2], geometry duality and the Legendre transform in [3,4], mixed volumes in [7]. This paper is concerned with characterizations of the derivative by the chain rule.

Let us remark that there are no non-trivial examples with (1) where $\mathcal{D}(T)$ and $\mathrm{Im}(T)$ are both *equal* to $C(\mathbb{R})$; this will be shown at the end. In this paper, we mainly consider the "natural" setting $\mathcal{D}(T) = C^1(\mathbb{R})$, $\mathrm{Im}(T) \subset C(\mathbb{R})$. In this case, the usual derivative D is surjective onto $C(\mathbb{R})$, of course. In chapter 5 we give modifications for the setting $\mathcal{D}(T) = C^k(\mathbb{R})$ and $C^\infty(\mathbb{R})$.

Without further restrictions on T, there are examples very different from D which satisfy the functional equation (1).

Consider a continuous positive function $H : \mathbb{R} \to \mathbb{R}_{>0}$ and define

$$Tf := \frac{H \circ f}{H}, \quad f \in C(\mathbb{R}). \tag{2}$$

This defines a map of $C^1(\mathbb{R})$ or $C(\mathbb{R})$ *into* $C(\mathbb{R})$ satisfying (1). We are grateful to L. Polterovich for pointing out this example to us. Since H needs to be never zero to make (2) a valid example, this map $T : C(\mathbb{R}) \to C(\mathbb{R})$ is *not onto* since no functions in its image have zeros.

Another example of a map $T : C^1(\mathbb{R}) \to C(\mathbb{R})$ satisfying (1) is given by

$$(Tf) := \left\{ \begin{matrix} f', & f \in C^1(\mathbb{R}) \text{ bijective}, \\ 0, & f \in C^1(\mathbb{R}) \text{ not bijective} \end{matrix} \right\}. \tag{3}$$

Checking (1), note that for $f, g \in C^1(\mathbb{R})$, $f \circ g : \mathbb{R} \to \mathbb{R}$ is bijective if and only if f *and* $g : \mathbb{R} \to \mathbb{R}$ are bijective ($f \circ g$ being bijective implies first that $f : \mathrm{Im}(g) \to \mathbb{R}$ is bijective and thus strictly monotone; then $f \in C^1(\mathbb{R})$ yields $\mathrm{Im}(g) = \mathbb{R}$).

Let $C_b^1(\mathbb{R}) = \{f \in C^1(\mathbb{R}) \mid f \text{ bounded from above or from below (or both)}\}$ be the class of half-bounded (or bounded) functions. We call

$$T : \mathcal{D}(T) = C^1(\mathbb{R}) \to \mathrm{Im}(T) \subset C(\mathbb{R})$$

non-degenerate if $T|_{C_b^1(\mathbb{R})} \neq 0$. This means explicitly

$$\exists x_1 \in \mathbb{R},\ h_1 \in C_b^1(\mathbb{R}), \quad (Th_1)(x_1) \neq 0. \tag{4}$$

We also consider two additional *non-degeneracy conditions* for T:

$$\exists x_0 \in \mathbb{R},\ h_0 \in C^1(\mathbb{R}), \quad (Th_0)(x_0) = 0 \tag{4.1}$$

and

$$\exists x_- \in \mathbb{R},\ h_- \in C^1(\mathbb{R}), \quad (Th_-)(x_-) < 0. \tag{4.2}$$

Note that (4) is not satisfied in example (3) and that (4.1) does not hold in example (2). The map $T : C^1(\mathbb{R}) \to C(\mathbb{R})$, $Tf := |f'|$ also fulfills (1) but does not satisfy (4.2). Our first main result states

Theorem 1. *Let $T : \mathcal{D}(T) := C^1(\mathbb{R}) \to \operatorname{Im}(T) \subset C(\mathbb{R})$ be an operation satisfying the chain rule*

$$T(f \circ g) = (Tf) \circ g \cdot Tg, \quad f, g \in \mathcal{D}(T). \tag{1}$$

Assume that T is non-degenerate in the sense of (4). *Then there exists some $p \geqslant 0$ and a positive continuous function $H \in C(\mathbb{R})$ such that either*

$$\left.\begin{aligned} &Tf = \frac{H \circ f}{H} |f'|^p \\ \text{or in the case of } p > 0, \quad &Tf = \frac{H \circ f}{H} |f'|^p \operatorname{sgn}(f'). \end{aligned}\right\} \tag{5}$$

If in addition to (4) *the non-degeneracy condition* (4.1) *is satisfied, we have $p > 0$, i.e. solutions of the type $Tf = H \circ f/H$ are excluded. If in addition to* (4)*, $T(2\,\mathrm{Id}) = c$ is a constant function, H is constant and therefore $Tf = |f'|^p$ or $Tf = |f'|^p \operatorname{sgn}(f')$, where $p = \log_2(c)$. Assuming also* (4.2) *and $c = 2$, the only solution to the chain rule equation is $Tf = f'$. If (4.2) holds and only $T(2\,\mathrm{Id})(0) = 2$ is satisfied, T is of the form $Tf = \frac{H \circ f}{H} \cdot f'$. Any map with* (5) *satisfies* (1)*, of course.*

Remarks.

(i) Take $p > 0$ and let G be the anti-derivative of $H^{1/p} > 0$, $G' = H^{1/p}$. Then G is a continuously differentiable strictly monotone function, and we get the following alternative formulation of the result

$$Tf = \left|\frac{(G \circ f)'}{G'}\right|^p = \left|\frac{d(G \circ f)}{dG}\right|^p$$

or

$$Tf = \left|\frac{(G \circ f)'}{G'}\right|^p \operatorname{sgn}(f') = \left|\frac{d(G \circ f)}{dG}\right|^p \operatorname{sgn}\left(\frac{d(G \circ f)}{dG}\right).$$

In this sense, all solutions of the chain rule for non-degenerate maps T are p-th powers of some derivative, up to signs. Under surjectivity conditions and the normalization $T(2\,\mathrm{Id}) = 2$, the usual derivative is the only solution. We would like to emphasize that no linearity or continuity conditions were imposed on T, but that they are a consequence of the answer.

(ii) Note that if T_1 and T_2 are operations satisfying (1), also $T_1 \cdot T_2$ is a solution of (1), and actually for any $p > 0$ also

$$f \mapsto |T_1 f|^p, \qquad f \mapsto |T_1 f|^p \operatorname{sgn}(T_1 f).$$

The first of these solutions, also in (5), is not (pointwise) surjective.

(iii) Obviously, the non-degeneracy condition is necessary for (5) to hold.

(iv) The form of an operator $T : C^1(\mathbb{R}) \to C(\mathbb{R})$ satisfying the chain rule (1) and the non-degeneracy conditions (4), (4.1) and (4.2) is actually *determined completely* by the function $T(2\,\mathrm{Id})$: as a restatement of Theorem 1 and its proof below let $p := \log_2 T(2\,\mathrm{Id})(0)$ and $\varphi(x) := T(2\,\mathrm{Id})(x)/T(2\,\mathrm{Id})(0)$. For this $\varphi \in C(\mathbb{R})$, the product $H(x) := \prod_{n=1}^{\infty} \varphi(\frac{x}{2^n})$ converges uniformly on compact subsets to a continuous function H with $H(0) = 1$ such that T is given by the formula

$$Tf = H \circ f / H \cdot \left|f'\right|^p \operatorname{sgn}\left(f'\right)$$

not only for $f = 2\,\mathrm{Id}$, but *for all* $f \in C^1(\mathbb{R})$.

An analogue of Theorem 1 for $C^k(\mathbb{R})$ or $C^\infty(\mathbb{R})$ functions will be given in the last chapter.

L. Polterovich mentioned to us that there is a cohomological interpretation of Theorem 1. We give this as a remark after the proof of Theorem 1.

As a matter of fact, the standard chain rule is to a certain extent enforced by a much weaker version: One may ask to what extent does the functional equation

$$T(f \circ g) = (Tf) \circ g \cdot Ag, \quad f, f \circ g \in \mathcal{D}(T),\ g \in \mathcal{D}(A) \tag{6}$$

characterize the derivative if $T : \mathcal{D}(T) = C^1(\mathbb{R}) \to \mathrm{Im}(T) \subset C(\mathbb{R})$ and $A : \mathcal{D}(A) = C^1(\mathbb{R}) \to \mathrm{Im}(A) \subset C(\mathbb{R})$ are operations connected by (6).

If T is non-degenerate, (6) implies that $A(\mathrm{Id}) = 1$. The setting $A := \mathbb{1}$ would allow solutions not associated with derivatives like $Tf := H \circ f$ for some $H \in C(\mathbb{R})$.

Let us introduce the following *strong non-degeneracy* condition for T,

$$\left.\begin{array}{lll} \text{(i)} & \forall x_1 \in \mathbb{R},\ \exists h_1 \in C_b^1(\mathbb{R}),\ y \in \mathbb{R}, & \big(h_1(y) = x_1\big) \wedge \big((Th_1)(y) \neq 0\big), \\ \text{(ii)} & \exists x_0 \in \mathbb{R},\ h_0 \in C^1(\mathbb{R}), & (Th_0)(x_0) = 0. \end{array}\right\} \tag{7}$$

If T satisfies (7) and the operators T and A fulfill (6), we will show that A has the form $Ag = Tg \cdot A_1 \circ g$, $g \in \mathcal{D}(A)$ where A_1 is a suitable function on $\mathbb{R}$.

If $G \in C^1(\mathbb{R})$ is strictly monotone and $H \in C(\mathbb{R})$ is positive, for any $p > 0$,

$$Tf := \left|(G \circ f)'\right|^p / H$$

and

$$Tf := \left|(G \circ f)'\right|^p / H \cdot \operatorname{sgn}\left(f'\right)$$

provide examples of operations T satisfying (6) if

$$Ag = Tg \cdot A_1 \circ g \quad \text{where } A_1 = H / \left|G'\right|^p.$$

In the case that $H = |G'|^p$, we come back to the solutions

$$Tf = \left|\frac{d(G \circ f)}{dG}\right|^p, \qquad Tf = \left|\frac{d(G \circ f)}{dG}\right|^p \operatorname{sgn}\left(\frac{dG \circ f}{dG}\right)$$

of (1) mentioned previously.

Differently from equation (1), the functional equation (6) is not stable under taking products of maps T_1, T_2 satisfying (6) for fixed A.

Our second main result is:

Theorem 2. *Let* $T : \mathcal{D}(T) = C^1(\mathbb{R}) \to \operatorname{Im}(T) \subset C(\mathbb{R})$ *and*

$$A : \mathcal{D}(A) = C^1(\mathbb{R}) \to \operatorname{Im}(A) \subset C(\mathbb{R})$$

be operations satisfying the functional equation

$$T(f \circ g) = (Tf) \circ g \cdot Ag, \quad f, f \circ g \in \mathcal{D}(T),\ g \in \mathcal{D}(A). \tag{6}$$

Assume that T *is strongly non-degenerate in the sense of* (7). *Then there exists* $p > 0$ *and there are positive functions* $G_1, G_2 \in C(\mathbb{R})$ *such that*

$$Tf = G_1 \circ f \cdot \frac{G_2}{G_1} \cdot \left|f'\right|^p \quad \text{or} \quad Tf = G_1 \circ f \cdot \frac{G_2}{G_1} \cdot \left|f'\right|^p \operatorname{sgn}\left(f'\right),$$

$$Af = Tf / G_2 \circ f = \frac{H \circ f}{H} \cdot K\left(f'\right),$$

where $H := \frac{G_1}{G_2}$ *and* $K(u) = |u|^p$ *or* $K(u) = |u|^p \operatorname{sgn}(u)$. *Conversely, any such operations* T *and* A *satisfy* (6).

Therefore the weak form (6) of the chain rule is actually only a small modification of the chain rule in the sense that

$$T(f \circ g) = (Tf) \circ g \cdot Tg / G_2 \circ g, \quad f, g \in C^1(\mathbb{R}).$$

Let us note that condition (7) of strong non-degeneracy of T is necessary for T and A to be of the above form. The example $Tf := H \circ f$, $H \in C(\mathbb{R})$ and $A := \mathbb{1}$ (in particular $T = \mathrm{Id}$, $A = \mathbb{1}$) satisfies (6) but is not of the form given: this T does not fulfill (7). Namely (7)(i) would require that H has no zeros while (7)(ii) only holds if H has at least one zero.

The condition (7) of strong non-degeneracy of T implies that A is non-degenerate in the sense of (4) if T and A are intertwined by satisfying (6): Applying (6) to $h_1 = \mathrm{Id} \circ h_1$ for the h_1

in (7)(i), one finds $(Ah_1)(y) \neq 0$ for some $y \in \mathbb{R}$ and $T(\mathrm{Id})(x) \neq 0$ for all $x \in \mathbb{R}$. Applying (6) then to $h_0 = \mathrm{Id} \circ h_0$ for the h_0 in (7)(ii) yields $(Ah_0)(x_0) = 0$ since $T(\mathrm{Id})(h(x_0)) \neq 0$ holds.

The following result was already indicated in the beginning of the introduction, concerning operators T on *all continuous* functions.

Proposition 3. *Assume that $T : C(\mathbb{R}) \to C(\mathbb{R})$ satisfies the functional equation $T(f \circ g) = (Tf) \circ g \cdot Tg$ for all $f, g \in C(\mathbb{R})$ and that there exist $g_0 \in C(\mathbb{R})$ and $x_0 \in \mathbb{R}$ with $(Tg_0)(x_0) = 0$. Then T is zero on the class of half-bounded continuous functions, $T|_{C_b(\mathbb{R})} = 0$.*

Remarks. Clearly, for any continuous function $H : \mathbb{R} \to \mathbb{R}_{>0}$, $T(f) = \frac{H \circ f}{H}$ defines a non-trivial operation $T : C(\mathbb{R}) \to C(\mathbb{R})$ with $T(f \circ g) = (Tf) \circ g \cdot Tg$. However, $Tf(x) \neq 0$ for all $f \in C(\mathbb{R})$ and all $x \in \mathbb{R}$. Also, the example of $T : C(\mathbb{R}) \to C(\mathbb{R})$ analogous to (3)

$$Tf := \left\{ \begin{array}{ll} \mathbb{1}, & f \in C(\mathbb{R}),\ f \text{ bijective}, \\ 0, & f \in C(\mathbb{R}),\ f \text{ not bijective} \end{array} \right\}$$

shows that there are operations T with $T(f \circ g) = (Tf) \circ g \cdot Tg$ and $T|_{C(\mathbb{R})} \neq 0$, $T|_{C_b(\mathbb{R})} = 0$.

2. Localization results

The derivative is a local operator. In order to prove our main results, we need to show a similar property for T. These localization results are the subject of this section. We assume throughout that the operator T satisfies the chain rule (1) and the non-degeneracy condition (4) and that

$$\mathcal{D}(T) = C^1(\mathbb{R}), \qquad \mathrm{Im}(T) \subset C(\mathbb{R}).$$

We start with a strengthening of (4).

Lemma 4.

(i) *For any $x_1 \in \mathbb{R}$ there exists $h_1 \in C_b^1(\mathbb{R})$ with $(Th_1)(x_1) \neq 0$.*
(ii) *If additionally* (4.1) *holds, for any $x_0 \in \mathbb{R}$ there exists $h_0 \in C^1(\mathbb{R})$ with $(Th_0)(x_0) = 0$.*

Proof. (i) Let $x_1 \in \mathbb{R}$. By (4) there exists $x_2 \in \mathbb{R}$ and $h_2 \in C_b^1(\mathbb{R})$ with $(Th_2)(x_2) \neq 0$. Define $g, h_1 \in C_b^1(\mathbb{R})$ by

$$g(s) = s + x_1 - x_2, \qquad h_1(s) = h_2(s + x_2 - x_1).$$

Then $g(x_2) = x_1$, $h_2 = h_1 \circ g$ and by (1),

$$0 \neq (Th_2)(x_2) = (Th_1)(x_1)(Tg)(x_2),$$

and hence $(Th_1)(x_1) \neq 0$. If h_2 is actually bounded, also h_1 is.

(ii) Let $x_0 \in \mathbb{R}$. By (4.1) there exists $x_2 \in \mathbb{R}$ and $h_2 \in C^1(\mathbb{R})$ with $(Th_2)(x_2) = 0$. Define $g, h_0 \in C^1(\mathbb{R})$ by

$$g(s) = s + x_2 - x_0, \qquad h_0(s) = h_2(s + x_2 - x_0).$$

Then $g(x_0) = x_2$, $h_0 = h_2 \circ g$ and by (1),

$$(Th_0)(x_0) = (Th_2)(x_2)(Tg)(x_0) = 0. \qquad \square$$

The following lemmas and proofs contain statements about different sets of open intervals in $\mathbb{R}$ depending on whether $Th = 0$ for all bounded functions h or not. If – case (a) – there is a *bounded* function $h \in C^1(\mathbb{R})$ and $x_1 \in \mathbb{R}$ with $(Th)(x_1) \neq 0$, we allow all *open* intervals. In this case, let

(a) $\mathcal{I} := \{J \subset \mathbb{R} \mid J \text{ open interval}\}$.

If – case (b) – there is only a *half-bounded* function, $h \in C^1(\mathbb{R})$, but no bounded function, and $x_1 \in \mathbb{R}$ with $(Th)(x_1) \neq 0$, the statements only hold for the smaller class of half-infinite *open* intervals

(b) $\mathcal{I} := \{J \subset \mathbb{R} \mid J = (c, \infty) \text{ or } (-\infty, c) \text{ for some } c \in \mathbb{R}\}$.

We use this class $\mathcal{I}$ as just defined throughout the rest of this section.

Lemma 5. *For any interval $J \in \mathcal{I}$, any $y \in J$ and any $x \in \mathbb{R}$ there exists $g \in \mathcal{D}(T)$ such that $g(x) = y$, $\operatorname{Im}(g) \subset J$ and $(Tg)(x) \neq 0$.*

Proof. Let $J \in \mathcal{I}$, $y \in J$ and $x \in \mathbb{R}$. By (4) and Lemma 4, there exists a bounded [case (a)] or half-bounded but not bounded [case (b)] function $h \in C^1(\mathbb{R})$ with $(Th)(x) \neq 0$. Pick some open interval $I \in \mathcal{I}$ for which $\operatorname{Im}(h) \subset I$ (bounded in case (a), half-infinite in case (b)). Choose any bijective C^1-map $f : J \to I$ with $f(y) = h(x)$ [note $h(x) \in I$]. Clearly, this may be done in such way that f is extendable to a C^1-map $\tilde{f}$ on $\mathbb{R}$, $\tilde{f}|_J = f$. Let

$$g := f^{-1} \circ h : \mathbb{R} \to J \subset \mathbb{R}.$$

Then $g \in \mathcal{D}(T)$, $g(x) = y$ and $\operatorname{Im}(g) \subset J$. Since $h = f \circ g = \tilde{f} \circ g$, we find using (1),

$$0 \neq (Th)(x) = T(\tilde{f} \circ g)(x) = (T\tilde{f})(y) \cdot (Tg)(x).$$

Hence $(Tg)(x) \neq 0$, $g(x) = y$ and $\operatorname{Im}(g) \subset J$ are satisfied. $\square$

The next lemma is a localized version of the claim that under our conditions, the identity function is mapped to the constant function 1, and that the constant functions are mapped onto the function 0.

Lemma 6. *For any interval $J \in \mathcal{I}$ and $f \in \mathcal{D}(T)$, $c \in \mathbb{R}$ the following holds*:

(i) *If* (4.1) *is satisfied and $f|_J = c$, we have $(Tf)|_{\bar{J}} = 0$.*
(ii) *If $f|_J = \mathrm{Id}_J$, we have $(Tf)|_{\bar{J}} = 1$.*

Proof. (i) Take $c \in \mathbb{R}$. We claim that $Tc = 0$ for the constant function given by c. Indeed, if there were $x \in \mathbb{R}$ with $(Tc)(x) \neq 0$, then $c \circ g = c$ for any $g \in \mathcal{D}(T)$ and by (1),

$$0 \neq (Tc)(x) = T(c \circ g)(x) = (Tc)\big(g(x)\big)(Tg)(x).$$

Thus $(Tg)(x) \neq 0$ for all $g \in \mathcal{D}(T)$ which contradicts (4.1). Hence $Tc = 0$.

For the local version of this, let $J \in \mathcal{I}$, $y \in J$ and $x \in \mathbb{R}$. By Lemma 5 there is a $g \in \mathcal{D}(T)$ with $\mathrm{Im}(g) \subset J$, $g(x) = y$ and $(Tg)(x) \neq 0$. From $f|_J = c$ we get that $f \circ g = c$, so we have by the preceding and (1) that

$$0 = (Tc)(x) = T(f \circ g)(x) = (Tf)(y)(Tg)(x).$$

Since $(Tg)(x) \neq 0$, we must have $(Tf)(y) = 0$. Since y was arbitrary in J, we get $(Tf)|_J = 0$. Since Tf is continuous, this also holds on the closure of J, $(Tf)|_{\bar{J}} = 0$.

(ii) Assume now that $f|_J = \mathrm{Id}|_J$. Take any $y \in J$, $x \in \mathbb{R}$. By Lemma 5 choose $g \in \mathcal{D}(T)$ with $\mathrm{Im}(g) \subset J$, $g(x) = y$ and $(Tg)(x) \neq 0$. Then $f \circ g = g$ so that by (1),

$$0 \neq (Tg)(x) = (Tf)(y) \cdot (Tg)(x).$$

We conclude that $(Tf)(y) = 1$. Hence $(Tf)|_J = 1$, and by the continuity of Tf, $(Tf)|_{\bar{J}} = 1$. □

Lemma 7. *In case* (a) *let $J \in \mathcal{I}$ be a bounded open interval and assume that $f \in \mathcal{D}(T)$ is such that $f|_J$ is strictly monotone. Then $(Tf)(x) \neq 0$ for all $x \in J$.*

Proof. For $I := f(J)$, $f|_J : J \to I$ is bijective. Restricting to subintervals $J \in \mathcal{I}$ if necessary, we may assume that both I, J are bounded. Then the inverse map $g := (f|_J)^{-1} : I \to J$ may be extended to some $\tilde{g} \in \mathcal{D}(T)$. Then $h := g \circ f$ satisfies $h|_J = \mathrm{Id}|_J$, and thus by Lemma 6,

$$1 = (Th)(x) = (T\tilde{g})\big(f(x)\big)(Tf)(x)$$

for any $x \in J$. Therefore $(Tf)(x) \neq 0$. □

The following is the localization of the operation T on an interval $J \in \mathcal{I}$.

Lemma 8. *Let $J \in \mathcal{I}$ and assume that $f_1, f_2 \in \mathcal{D}(T)$ satisfy $f_1|_J = f_2|_J$. Then $(Tf_1)|_{\bar{J}} = (Tf_2)|_{\bar{J}}$.*

Proof. Let $x \in J$ be arbitrary. Choose a smaller interval $J_1 \subset J$, $J_1 \in \mathcal{I}$ and a function $g \in C^1(\mathbb{R})$ such that $x \in J_1$, $\mathrm{Im}\, g \subset J$ and $g|_{J_1} = \mathrm{Id}|_{J_1}$. We then have $f_1 \circ g = f_2 \circ g$, and by Lemma 6, $(Tg)|_{J_1} = 1$. In particular, $g(x) = x$ and $(Tg)(x) = 1$. Hence, using (1),

$$\begin{aligned}(Tf_1)(x) &= (Tf_1)\big(g(x)\big) = (Tf_1)\big(g(x)\big) \cdot (Tg)(x) = T(f_1 \circ g)(x) \\ &= T(f_2 \circ g)(x) = (Tf_2)\big(g(x)\big) \cdot (Tg)(x) = (Tf_2)(x).\end{aligned}$$

This shows that $(Tf_1)|_J = (Tf_2)|_J$ holds. By the continuity of Tf_1 and Tf_2, also $(Tf_1)|_{\bar{J}} = (Tf_2)|_{\bar{J}}$ is true. □

The following lemma gives a converse result to case (i) of Lemma 6.

Lemma 9. *Let $J \in \mathcal{I}$ and assume that* (4.1) *holds and that $g \in \mathcal{D}(T)$ satisfies $(Tg)|_J = 0$. Then g is constant on J: there exists $c \in \mathbb{R}$ such that $g|_J = c$.*

We omit the proof here since it follows directly from Lemma 11 below.
We now prove a "pure" localization result.

Proposition 10. *Assume that $T : C^1(\mathbb{R}) \to C(\mathbb{R})$ satisfies the chain rule* (1) *and is non-degenerate in the sense of* (4). *Then there is a function $F : \mathbb{R}^3 \to \mathbb{R}$ such that*

$$Tf(x) = F\big(x, f(x), f'(x)\big) \tag{8}$$

holds for any $f \in C^1(\mathbb{R})$ and any $x \in \mathbb{R}$.

Proof. Fix $x_0 \in \mathbb{R}$ and consider $f \in C^1(\mathbb{R})$ with given values $f(x_0)$ and $f'(x_0)$. Let $J_1 := (x_0, \infty)$ and $J_2 := (-\infty, x_0)$. Then $J_1, J_2 \in \mathcal{I}$. Consider the tangent of f at x_0, $g(x) := f(x_0) + (x - x_0)f'(x_0)$, $x \in \mathbb{R}$. It suffices to prove that $(Tf)(x_0) = (Tg)(x_0)$. Define $h \in C^1(\mathbb{R})$ by

$$h(x) := \left\{\begin{matrix} g(x), & x \in J_1, \\ f(x), & x \in \bar{J}_2 \end{matrix}\right\}.$$

Then $h|_{J_1} = g|_{J_1}$ and $h|_{J_2} = f|_{J_2}$. Hence by Lemma 8,

$$(Tg)|_{\bar{J}_1} = (Th)|_{\bar{J}_1} \quad \text{and} \quad (Th)|_{\bar{J}_2} = (Tf)|_{\bar{J}_2}.$$

Since $x_0 \in \bar{J}_1 \cap \bar{J}_2$, we conclude that $(Tg)(x_0) = (Th)(x_0) = (Tf)(x_0)$. Therefore the value $(Tf)(x_0)$ depends only on the two parameters $f(x_0)$ and $f'(x_0)$, for any fixed $x_0 \in \mathbb{R}$. We encode this information by letting $(Tf)(x_0) = F_{x_0}(f(x_0), f'(x_0))$, where $F_{x_0} : \mathbb{R}^2 \to \mathbb{R}$ is a fixed function for any $x_0 \in \mathbb{R}$. Finally denoting $F(x, y, z) := F_x(y, z)$, we have that for any $x \in \mathbb{R}$ and $f \in \mathcal{D}(T)$,

$$Tf(x) = F\big(x, f(x), f'(x)\big). \qquad \square$$

3. The strong form of the chain rule

In this section we give the proof of Theorem 1. We generally assume that $\mathcal{D}(T) = C^1(\mathbb{R})$, $\mathrm{Im}(T) \subset C(\mathbb{R})$ and that T satisfies the chain rule and the condition (4) of non-degeneracy.

Lemma 11. *If in addition to* (4) *also* (4.1) *holds, for any $g \in \mathcal{D}(T)$ and $x \in \mathbb{R}$ we have: $(Tg)(x) = 0$ holds if and only if $g'(x) = 0$.*

Remark. Lemma 9 is a direct corollary of this result.

Proof. Let $g \in \mathcal{D}(T)$ and $x_0 \in \mathbb{R}$. By (8) of Proposition 10,

$$(Tf)(x) = F\big(x, f(x), f'(x)\big), \quad x \in \mathbb{R},\ f \in \mathcal{D}(T).$$

Inserting a constant function $f = c$ and using Lemma 6, we see that $0 = (Tf)(x) = F(x, c, 0)$ for any $x, c \in \mathbb{R}$. This shows that $g'(x_0) = 0$ implies $(Tg)(x_0) = 0$.

Assume now that $(Tg)(x_0) = 0$. We then have by the chain rule (1) for any $f \in \mathcal{D}(T)$,

$$F\big(x_0, f\big(g(x_0)\big), f'\big(g(x_0)\big)g'(x_0)\big) = T(f \circ g)(x_0) = (Tf)\big(g(x_0)\big)(Tg)(x_0) = 0.$$

This implies that for any two numbers $b, c \in \mathbb{R}$,

$$F\big(x_0, b, cg'(x_0)\big) = 0.$$

If $g'(x_0) \neq 0$ would hold, $F(x_0, \cdot, \cdot) : \mathbb{R}^2 \to \mathbb{R}$ would be the zero function, and hence $(Tf)(x_0) = 0$ would follow for any $f \in \mathcal{D}(T)$, contradicting assumption (4.1). Hence $(Tf)(x_0) = 0$ if and only if $f'(x_0) = 0$ is true. □

Remark. If for some x_0, $(Tg)(x_0) = 0$ holds for all *bounded* functions $g \in \mathcal{D}(T)$, this means that $F(x_0, \cdot, \cdot) : \mathbb{R}^2 \to \mathbb{R}$ is zero and hence $(Tg)(x_0) = 0$ for all functions $g \in \mathcal{D}(T)$, in particular for all *half-bounded* functions $g \in \mathcal{D}(T)$. Thus case (b) in the definition of the set $\mathcal{I}$ of intervals (after Lemma 4) is impossible and hence Lemmas 5 to 9 hold for *all* open intervals.

Proposition 12. *The function F with* (8) *representing T has the form*

$$F(x, y, z) = H(y)/H(x) \cdot K(z), \quad x, y, z \in \mathbb{R}$$

where $H : \mathbb{R} \to \mathbb{R}_{\neq 0}$ and $K : \mathbb{R} \to \mathbb{R}$ are suitable functions and where K is multiplicative and zero exactly in zero,

$$K(uv) = K(u)K(v), \quad u, v \in \mathbb{R}.$$

Moreover, either $K \equiv \mathbb{1}$ or ($K(u) = 0$ if and only if $u = 0$). In the case of (4.1) *the second case applies.*

Proof. (i) Consider $x_0, y_0 \in \mathbb{R}$ and $f, g \in \mathcal{D}(T)$ such that $g(x_0) = y_0$, $f(y_0) = x_0$. We have by (1),

$$\begin{aligned} T(f \circ g)(x_0) &= (Tf)(y_0)(Tg)(x_0) \\ &= (Tg)(x_0)(Tf)(y_0) \\ &= T(g \circ f)(y_0). \end{aligned} \tag{9}$$

Since $(f \circ g)(x_0) = x_0$, $(g \circ f)(y_0) = y_0$, this means, using (8) and the chain rule,

$$F\big(x_0, x_0, f'(y_0)g'(x_0)\big) = F\big(y_0, y_0, g'(x_0)f'(y_0)\big).$$

Since there are such functions f, g for which the derivatives $f'(y_0)$ and $g'(x_0)$ attain arbitrary values, we conclude that for all $x_0, y_0, u \in \mathbb{R}$, $F(x_0, x_0, u) = F(y_0, y_0, u)$. We may thus define $K : \mathbb{R} \to \mathbb{R}$ by $K(u) := F(x_0, x_0, u)$ independently of $x_0 \in \mathbb{R}$. Expressing (9) in terms of F, we find

$$K\big(f'(y_0)g'(x_0)\big) = F\big(x_0, y_0, g'(x_0)\big)F\big(y_0, x_0, f'(x_0)\big),$$

and hence for any $u, v \in \mathbb{R}$ and $x_0, y_0 \in \mathbb{R}$

$$K(uv) = F(x_0, y_0, u)F(y_0, x_0, v). \tag{10}$$

Moreover, $K(1) = F(x_0, x_0, 1) = T(\mathrm{Id})(x_0) = 1$ by Lemma 6. Therefore

$$F(x_0, y_0, u)F\big(y_0, x_0, u^{-1}\big) = 1 \tag{11}$$

for any $u \neq 0$; in particular $F(x_0, y_0, u) \neq 0$. Hence

$$F(x_0, y_0, u) = \frac{K(uv)}{F(y_0, x_0, v)}, \quad v \neq 0$$

where the right side is independent of $v \neq 0$. Taking $v = 1$, we define $G : \mathbb{R}^2 \to \mathbb{R}_{\neq 0}$ by

$$G(x_0, y_0) := 1/F(y_0, x_0, 1),$$

so that $G(x_0, x_0) = 1$ and

$$F(x_0, y_0, u) = G(x_0, y_0)K(u), \quad x_0, y_0, u \in \mathbb{R}$$

and by (11), $G(x_0, y_0)G(y_0, x_0) = 1$. Putting $x_0 = y_0$ in (10), we also have that $K(uv) = K(u)K(v)$. Since for $u \neq 0$, $F(x_0, x_0, u) \neq 0$, $K(u) \neq 0$ holds for $u \neq 0$. For the constant function $f = x_0$, $K(0) = F(x_0, x_0, 0) = (Tf)(x_0)$. Assuming (4.1), this is zero by Lemma 6. If (4.1) does not hold, $K(0) \neq 0$ is possible. But then $K(0)^2 = K(0 \cdot 0) = K(0)$ yields $K(0) = 1$ and $K(0) = K(0 \cdot v) = K(0) \cdot K(v)$ implies that $K(v) = 1$ for all $v \in \mathbb{R}$, i.e. $K = \mathbb{1}$.

(ii) To finish the proof, we will show that there exists a function $H : \mathbb{R} \to \mathbb{R}_{\neq 0}$ so that $G(x, y) = H(y)/H(x)$ for all $x, y \in \mathbb{R}$.

Using the functional equation (1) for $f, g \in C^1(\mathbb{R})$,

$$T(f \circ g)(x) = (Tf)\big(g(x)\big)(Tg)(x),$$

this is expressed in terms of G and K by

$$\begin{aligned} &G\big(x, f\big(g(x)\big)\big) \cdot K\big(f'\big(g(x)\big)g'(x)\big) \\ &\quad = G\big(g(x), f\big(g(x)\big)\big) \cdot K\big(f'\big(g(x)\big)\big) \cdot G\big(x, g(x)\big) \cdot K\big(g'(x)\big), \quad x \in \mathbb{R}. \end{aligned}$$

Together with $K(uv) = K(u)K(v)$, we find choosing f, g with $f'(g(x)) \neq 0$, $g'(x) \neq 0$,

$$G\big(x, f\big(g(x)\big)\big) = G\big(x, g(x)\big)G\big(g(x), f\big(g(x)\big)\big).$$

Since the function values $g(x)$, $f(g(x))$ can be chosen arbitrarily, we conclude that

$$G(x, y) = G(x, w)G(w, y)$$

holds for any $x, y, w \in \mathbb{R}$. Further, $G(w, x) = 1/G(x, w)$. We find, defining $H : \mathbb{R} \to \mathbb{R}_{\neq 0}$ by $H(y) := G(0, y)$ that

$$G(x, y) = G(x, 0)G(0, y) = G(0, y)/G(0, x) = H(y)/H(x). \qquad \square$$

Remarks.

1. $H(x_0)$ is just the value of the image of the shift by x_0 in 0,

$$H(y_0) = G(0, y_0) = F(0, y_0, 1) = T(\mathrm{Id} + y_0)(0).$$

2. The conclusion of Proposition 10 and 12 is that

$$(Tf)(x) = \frac{(H \circ f)(x)}{H(x)} \cdot K\big(f'(x)\big), \quad x \in \mathbb{R},\ f \in C^1(\mathbb{R}) \tag{12}$$

with H being never zero, K being multiplicative and

$$H\big(\cdot, f(\cdot)\big)K\big(f'(\cdot)\big)$$

being continuous for any $f \in C^1(\mathbb{R})$.

Lemma 13. *Assume that $K : \mathbb{R} \to \mathbb{R}$ is measurable, not identically zero and satisfies for all $u, v \in \mathbb{R}$ that $K(uv) = K(u)K(v)$. Then there exists some $p \geqslant 0$ such that*

$$K(u) = |u|^p \quad \textit{or} \quad K(u) = |u|^p \operatorname{sgn}(u).$$

Proof. For $u \neq 0$, $K(u) \neq 0$ since otherwise K is zero identically. Therefore we can define $F : \mathbb{R} \to \mathbb{R}$, $F(t) := \log|K(e^t)|$. Then for $s, t \in \mathbb{R}$,

$$F(t + s) = \log\big|K\big(e^{t+s}\big)\big| = \log\big|K\big(e^t\big)\big| + \log\big|K\big(e^s\big)\big| = F(t) + F(s).$$

Since F is measurable, by a result of Sierpinski [8] and Banach [5], F is linear, i.e. there is $p \in \mathbb{R}$ such that $F(t) = pt$, $t \in \mathbb{R}$. Thus $|K(e^t)| = e^{pt}$.

Since $K(1) = K(1)^2 > 0$, it follows that $K(u) = u^p$ for any $u > 0$.

Since $K(-1)^2 = K(1) = 1$, $K(-1) \in \{\pm 1\}$. If $K(-1) = 1$, $K(u) = |u|^p$ for any $u \in \mathbb{R}$, and if $K(-1) = -1$, $K(u) = |u|^p \operatorname{sgn}(u)$ for any $u \in \mathbb{R}$. Since $K(0) = K(0)^2 = 0$ or 1, p has to be non-negative. $\square$

Proof of Theorem 1. Under the assumptions of Theorem 1, (12) gives the general form of T.

(a) We will show that the function K in (12) is measurable and apply Lemma 13. Applying (12) first to $f(x) = 2x$, we get that $(Tf)(x) = \frac{H(2x)}{H(x)}K(2)$. Since $K(2) \neq 0$ and $Tf \in C(\mathbb{R})$, $\varphi : \mathbb{R} \to \mathbb{R}$, $\varphi(x) := \frac{H(2x)}{H(x)}$ is continuous on $\mathbb{R}$.

Then $H(2x)/(\varphi(x)H(x)) = 1$. Take any $a, b \in \mathbb{R}$. Then

$$\frac{H(b)}{\varphi(\frac{b}{2})H(\frac{b}{2})} = 1 = \frac{H(a)}{\varphi(\frac{a}{2})H(\frac{a}{2})},$$

$$\frac{H(b)}{H(a)} = \frac{H(\frac{b}{2})}{H(\frac{a}{2})} \cdot \frac{\varphi(\frac{b}{2})}{\varphi(\frac{a}{2})}.$$

Using this iteratively with $(b/2^i, a/2^i)$ instead of (b, a), we find for any $k \in \mathbb{N}$,

$$\frac{H(b)}{H(a)} = \frac{H(\frac{b}{2^k})}{H(\frac{a}{2^k})} \prod_{i=1}^{k} \left(\frac{\varphi(\frac{b}{2^i})}{\varphi(\frac{a}{2^i})} \right). \tag{13}$$

Choose $a = 1$ and apply (12) to $g(x) = bx$ to get

$$(Tg)(x) = \frac{H(bx)}{H(x)} K(b)$$

Since Tg is continuous in 0, we find that

$$K(b) = (Tg)(0) = \lim_{x \to 0} (Tg)(x) = \lim_{k \to \infty} (Tg)\left(\frac{1}{2^k}\right) = \left(\lim_{k \to \infty} \frac{H(\frac{b}{2^k})}{H(\frac{1}{2^k})}\right) K(b)$$

exists for any b; using this and (13) yields

$$1 = \lim_{k \to \infty} \frac{H(\frac{b}{2^k})}{H(\frac{1}{2^k})} = \lim_{k \to \infty} \left(\prod_{i=1}^{k} \frac{\varphi(\frac{1}{2^i})}{\varphi(\frac{b}{2^i})} \right) \cdot \frac{H(b)}{H(1)},$$

$$H(b) = H(1) \cdot \lim_{k \to \infty} \left(\prod_{i=1}^{k} \frac{\varphi(\frac{b}{2^i})}{\varphi(\frac{1}{2^i})} \right).$$

As a function of b, the right side is a pointwise limit of continuous functions in b since φ is continuous. Hence H is measurable as a function of b.

(b) Next, choose $h(x) = x^2/2$ in (12) to conclude that

$$(Th)(x) = \frac{H(x^2/2)}{H(x)} K(x),$$

so that

$$K(x) = \frac{H(x)}{H(x^2/2)} (Th)(x).$$

Since H is measurable and (Th) is continuous, the right side and therefore K is measurable as a function of x. Moreover, K is multiplicative and by Lemma 13,

$$K(u) = |u|^p \quad \text{or} \quad K(u) = |u|^p \operatorname{sgn}(u)$$

for a suitable $p \geqslant 0$. Hence

$$(Tf)(x) = \frac{H(f(x))}{H(x)} \left|f'(x)\right|^p$$

or

$$(Tf)(x) = \frac{H(f(x))}{H(x)} \left|f'(x)\right|^p \operatorname{sgn}\bigl(f'(x)\bigr).$$

We conclude that for any $f \in C^1(\mathbb{R})$, $H(f(x))/H(x)$ is continuous in x (if $f'(x) \neq 0$). To conclude from this that H is continuous, it suffices to prove that H is continuous in one point, say x_0: the continuity in any other point x_1 follows from considering $f(x) = x + x_1 - x_0$ and the continuity of $H \circ f/H$.

(c) For any $c \in \mathbb{R}$, let $b(c) := \overline{\lim}_{y\to c} H(y)$ and $a(c) := \underline{\lim}_{x\to c} H(x)$. We claim that $b(c)/H(c)$ and $a(c)/H(c)$ are constant functions of c; in the case that for some c_0, $b(c_0)$ or $a(c_0)$ are infinity or zero, this should mean that all other values $b(c)$ or $a(c)$ are infinity or zero, too.

Assume to the contrary that there are c_0 and c_1 such that $b(c_1)/H(c_1) < b(c_0)/H(c_0)$.

Choose any maximizing sequence y_n, $\lim_n y_n = c_0$ with $\lim_n H(y_n) = b(c_0)$. Since for $f(t) = t + c_1 - c_0$, $H \circ f/H$ is continuous, $\lim_n H(y_n + c_1 - c_0)/H(y_n) = H(c_1)/H(c_0)$ exists and $\lim_n(y_n + c_1 - c_0) = c_1$ implies that $\overline{\lim}_n H(y_n + c_1 - c_0) \leqslant b(c_1)$, we arrive at the contradiction

$$\begin{aligned}
\frac{b(c_0)}{H(c_0)} &= \frac{H(c_1)}{H(c_0)} \frac{b(c_0)}{H(c_1)} \\
&= \lim_n \frac{H(y_n + c_1 - c_0)}{H(y_n)} \frac{H(y_n)}{H(c_1)} \\
&\leqslant \frac{\overline{\lim}_n H(y_n + c_1 - c_0)}{H(c_1)} \\
&= \frac{b(c_1)}{H(c_1)} < \frac{b(c_0)}{H(c_0)}.
\end{aligned}$$

This argument is also valid assuming $b(c_1) < b(c_0) = \infty$. The proof for $a(c)$ is similar. We now use a similar reasoning as in [2].

If H would be discontinuous at some point, it would be discontinuous anywhere and could not be extended to be continuous. Assume that this is the case and take any sequence $(c_n)_{n\in\mathbb{N}}$ of pairwise disjoint numbers ($c_n \neq c_m$ for $n \neq m$) with $\lim_n c_n = 0$. Let $\delta_n := \frac{1}{4}\min\{|c_n - c_m| \mid m \neq n\}$ and choose $0 < \varepsilon_n < \delta_n$ such that $\sum_{n\in\mathbb{N}} \varepsilon_n/\delta_n < \infty$. Since H is discontinuous in any point c_n, $b(c_n)/a(c_n) > 1$ holds. By the above argument, this value is independent of n,

$$1 < \frac{b}{a} := \frac{b(c_n)}{a(c_n)} \quad \text{for all } n \in \mathbb{N}.$$

Hence by definition of $b(c_n)$ and $a(c_n)$ we may find $y_n, x_n \in \mathbb{R}$ with $|y_n - c_n| < \varepsilon_n$, $|x_n - c_n| < \varepsilon_n$ and $H(y_n)/H(x_n) > \frac{b+a}{2a} > 1$ (if $b/a = \infty$, choose them with $H(y_n)/H(x_n) > 2$). Let ψ be a smooth cut-off function like

$$\psi(x) = \begin{Bmatrix} \exp(1 - \frac{1}{1-x^2}), & |x| < 1, \\ 0, & |x| \geqslant 1 \end{Bmatrix}$$

and put $g_n(x) := (y_n - x_n)\psi(\frac{x-x_n}{\delta_n})$. The functions g_n are supported disjointly since $|x_n - x_m| \geqslant |c_n - c_m| - 2\varepsilon_n \geqslant 4\delta_n - 2\varepsilon_n \geqslant 2\delta_n$ for any $m \neq n$. Hence $g_n(x_m) = (y_n - x_n)\delta_{nm}$. Since

$$\sum_n \|g_n'\|_\infty \leqslant \sum_n |y_n - x_n|/\delta_n \|\psi'\|_\infty \leqslant 2\left(\sum_n \varepsilon_n/\delta_n\right)\|\psi'\|_\infty < \infty$$

holds,

$$f(x) := x + \sum_n g_n(x), \quad x \in \mathbb{R}$$

defines a differentiable function $f \in C^1(\mathbb{R})$ with $f(x_n) = x_n + (y_n - x_n) = y_n$, $f(0) = 0$ and $f'(0) = 1 \neq 0$. Since $x_n \to 0$, $y_n = f(x_n) \to 0$, the continuity of $H \circ f/H$ yields the contradiction

$$1 = \frac{H(0)}{H(0)} = \lim_n \frac{H(y_n)}{H(x_n)} > \frac{b+a}{2a} > 1.$$

This proves that H is continuous.

(d) In the case that $T(2\,\mathrm{Id}) = c$ is constant, the function φ in part (a) is constant and consequently H is constant, giving that $Tf = |f'|^p$ or $Tf = |f'|^p \operatorname{sgn}(f')$. Clearly $T(2\,\mathrm{Id}) = 2$ yields $p = 1$, and condition (4.2) excludes the first possibility, implying $Tf = f'$. If only $T(2\,\mathrm{Id})(0) = 2$, H may be non-constant, but $p = 1$ follows directly from the form of T.

Choosing $a = 0$ in part (a) gives $H(b) = H(0)\prod_{i\in\mathbb{N}} \varphi(\frac{b}{2^i})$. This, the form of K and the continuity of H justify Remark (iv) after the statement of Theorem 1.

A direct calculation shows that any operator T given by (5) satisfies the chain rule functional equation (1). □

Remark. We now mention a cohomological interpretation of Theorem 1. The semigroup $G = (C^1(\mathbb{R}), 0)$ with the operation of composition acts on the abelian semigroup $M = (C(\mathbb{R}), \cdot)$ with the operation of pointwise multiplication by composition from the right, $G \times M \to M$, $fH := H \circ f$. Thus M is a module over G. We denote the functions from G^n to M by $F^n(G, M)$ and define the coboundary operators

$$d^n : F^n(G, M) \to F^{n+1}(G, M), \quad n \in \mathbb{N}_0$$

using *additive* notation $+$ for the operation $\cdot$ on M by

$$
\begin{aligned}
d^n\varphi(g_1,\dots,g_{n+1}) = {} & g_1\varphi(g_2,\dots,g_{n+1}) \\
& + \sum_{i=1}^{n}(-1)^i\varphi(g_1,\dots,g_{i-1},g_ig_{i+1},g_{i+2},\dots,g_{n+1}) \\
& + (-1)^{n+1}\varphi(g_1,\dots,g_n)
\end{aligned}
$$

where φ maps G^n into M and $g_1,\dots,g_n \in G$. Theorem 1 characterizes the cocycles in $\operatorname{Ker}(d^1)$ for $n=1$; then $\varphi = T : G = C^1(\mathbb{R}) \to M = C(\mathbb{R})$ has coboundaries

$$d^1T(g_1,g_2) = g_1T(g_2) - T(g_1g_2) + T(g_1), \quad g_1,g_2 \in C^1(\mathbb{R}).$$

As for cocycles T, $d^1T = 0$ means in multiplicative notation that

$$T(g_2 \circ g_1) = T(g_2) \circ g_1 \cdot Tg_1.$$

The cocycles are just the solutions of the chain rule equation, $Tg = H \circ g/H \cdot |g'|^p[\operatorname{sgn}(g')]$. For $n=0$, $\varphi \in F^0(G,M)$ can be identified with $\varphi = H \in M = C(\mathbb{R})$ and thus the coboundaries are

$$d^0H(g) = gH - H, \quad g \in C^1(\mathbb{R})$$

and, in multiplicative notation, for functions H which are never zero,

$$d^0H(g) = H \circ g/H.$$

The cohomology group $H^1(G,M) = \operatorname{Ker}(d^1)/\operatorname{Im}(d^0)$ is thus represented by the maps $g \mapsto |g'|^p$ and $g \mapsto |g'|^p \operatorname{sgn}(g')$ from G to M.

4. The weak form of the chain rule

We now turn to the proof of Theorem 2 concerning the equation

$$T(f \circ g) = T(f) \circ g \cdot Ag, \quad f,g \in C^1(\mathbb{R}). \tag{6}$$

We assume throughout this section that $\mathcal{D}(T) = \mathcal{D}(\mathcal{A}) = C^1(\mathbb{R})$, $\operatorname{Im}(T), \operatorname{Im}(A) \subset C(\mathbb{R})$ and that T is strongly non-degenerate as defined by (7), i.e. that the assumptions of Theorem 2 hold.

By (7)(i), for any $x \in \mathbb{R}$ there is $g \in C_b^1(\mathbb{R})$ and $y \in \mathbb{R}$ such that $g(y) = x$ and $(Tg)(y) \neq 0$. Since $\operatorname{Id} \circ g = g$, by (6)

$$0 \neq (Tg)(y) = T(\operatorname{Id})\big(g(y)\big)(Ag)(y) = T(\operatorname{Id})(x)(Ag)(y).$$

Hence for any $x \in \mathbb{R}$, $T(\operatorname{Id})(x) \neq 0$ holds.

We will reduce the proof of Theorem 2 to the proof of Theorem 1. For this, we need two localization results.

Lemma 14. *Let $c \in \mathbb{R}$ be arbitrary and $J := (c,\infty)$ or $J = (-\infty,c)$. Then for any $x \in J$ there exists $y \in \mathbb{R}$ and $g \in C_b^1(\mathbb{R})$ with $g(y) = x$, $(Ag)(y) \neq 0$ and $\operatorname{Im}(g) \subset J$.*

This is a stronger condition than (7)(i), but now for A, since additionally $\mathrm{Im}(g) \subset J$ holds.

Proof. Let $c \in \mathbb{R}$ be arbitrary. We consider the case $J = (c, \infty)$. Take any $x \in J$, thus $c < x$. By (7) there exists $h_1 \in C_b^1(\mathbb{R})$ and $y \in \mathbb{R}$ with $h_1(y) = x$ and $(Th_1)(y) \neq 0$.

(i) We may assume that h_1 is bounded from below: If h_1 were bounded from above, consider $h_2 := 2x - h_1$. Then h_2 is bounded from below, $h_2(y) = x = h_1(y)$. Let $f(s) := 2x - s$. Then $h_1 = f \circ h_2$ and $h_2 = f \circ h_1$. By (6),

$$0 \neq (Th_1)(y) = (Tf)(x)(Ah_2)(y),$$

implying $(Tf)(x) \neq 0$. Using $h_1 = \mathrm{Id} \circ h_1$ and (6), we get

$$0 \neq (Th_1)(y) = T(\mathrm{Id})(x)(Ah_1)(y),$$

and hence $(Ah_1)(y) \neq 0$. Finally, (6) and $h_2 = f \circ h_1$ yield

$$(Th_2)(y) = (Tf)(x)(Ah_1)(y) \neq 0.$$

(ii) Assume therefore that $h_1(y) = x$, $(Th_1)(y) \neq 0$ and $h_1 \in C^1(\mathbb{R})$ is bounded from below. Then $\mathrm{Im}(h_1) \subset (c_1, \infty)$ for some $c_1 \in \mathbb{R}$. Since $x \in \mathrm{Im}(h_1)$, $c_1 < x$.

If c_1 can be chosen such that $c_1 \geqslant c$, $\mathrm{Im}(h_1) \subset (c_1, \infty) \subset (c, \infty) = J_1$. Using (6) for $h_1 = \mathrm{Id} \circ h_1$ yields $(Ah_1)(y) \neq 0$ as noted already. Taking $g = h_1$, the proof is finished in this case.

If c_1 is such that $c_1 < c$, we may find $\lambda \in (0, 1)$ such that $(1 - \lambda)x + \lambda c_1 > c$ since $x > c$ (with a possibly small $\lambda > 0$).

Let $g(s) := (1 - \lambda)x + \lambda h_1(s)$. Then $g(y) = x$, $\mathrm{Im}(g) \subset (c, \infty) = J$ since for any $s \in \mathbb{R}$, $(1 - \lambda)x + \lambda h_1(s) > (1 - \lambda)x + \lambda c_1 > c$. Note that $h_1(s) = \gamma g(s) - (\gamma - 1)x$ for $\gamma = 1/\lambda$. Let $f(t) := \gamma t - (\gamma - 1)x$. Then $h_1 = f \circ g$ and by (6),

$$0 \neq (Th_1)(y) = T(f \circ g)(y) = (Tf)(x)(Ag)(y).$$

This means that $g(y) = x$, $(Ag)(y) \neq 0$ and $\mathrm{Im}(g) \subset J$ holds as required. □

Lemma 15. *Let $c \in \mathbb{R}$ be arbitrary and $J = (c, \infty)$ or $J = (-\infty, c)$. Assume that $f_1, f_2 \in C^1(\mathbb{R})$ satisfy $f_1|_J = f_2|_J$. Then $(Tf_1)|_{\bar{J}} = (Tf_2)|_{\bar{J}}$.*

Proof. Let $x \in J$ be arbitrary. By Lemma 14, we can find $y \in \mathbb{R}$ and $g \in C_b^1(\mathbb{R})$ with $g(y) = x$, $(Ag)(y) \neq 0$ and $\mathrm{Im}(g) \subset J$. If $f_1, f_2 \in C^1(\mathbb{R})$ satisfy $f_1|_J = f_2|_J$, $f_1 \circ g = f_2 \circ g$ and (6) implies

$$(Tf_1)(x)(Ag)(y) = T(f_1 \circ g)(y) = T(f_2 \circ g)(y) = (Tf_2)(x)(Ag)(y).$$

Since $(Ag)(y) \neq 0$, we find $(Tf_1)(x) = (Tf_2)(x)$. This shows that $(Tf_1)|_J = (Tf_2)|_J$, and by continuity of Tf_1, Tf_2, $(Tf_1)|_{\bar{J}} = (Tf_2)|_{\bar{J}}$ holds. □

The "pure" localization result, the analogue of Proposition 10, states:

Proposition 16. *Assume* (6) *holds for T and A and that T is strongly non-degenerate. Then there is a function $F:\mathbb{R}^3\to\mathbb{R}$ such that*

$$(Tf)(x)=F\big(x,f(x),f'(x)\big)$$

holds for any $f\in C^1(\mathbb{R})$ and $x\in\mathbb{R}$.

Proof. The proof is the same as the one of Proposition 10, only replacing Lemma 8 by Lemma 15. Note that Lemma 8 was only applied to open intervals $J_1=(x_0,\infty)$ and $J_2=(-\infty,x_0)$ which are just those intervals allowed in Lemma 15. □

Proposition 17. *Assume that T and A satisfy* (6)*,*

$$T(f\circ g)=(Tf)\circ g\cdot Ag,\quad f,g\in C^1(\mathbb{R})$$

and that T is strongly non-degenerate (7)*. Then there is a function $G:\mathbb{R}\to\mathbb{R}_{\neq 0}$ such that $Af=Tf\cdot G\circ f$; $f\in C^1(\mathbb{R})$. Moreover, there are functions $G_1,G_2:\mathbb{R}\to\mathbb{R}_{\neq 0}$ and $K:\mathbb{R}\to\mathbb{R}$ such that the function F in Proposition* 16 *has the form*

$$F(x,y,z)=\frac{G_1(y)}{G_1(x)}G_2(x)K(z),\quad x,y,z\in\mathbb{R}$$

where G_1 is continuous and $K(u)=|u|^p$ or $K(u)=|u|^p\operatorname{sgn}(u)$ for a suitable $p>0$.

Proof. Choose any $g\in C^1(\mathbb{R})$ and $x\in\mathbb{R}$. The chain rule (6) yields for any $f\in C^1(\mathbb{R})$,

$$T(f\circ g)(x)=(Tf)\big(g(x)\big)(Ag)(x),$$

which by Proposition 16 means

$$F\big(x,f\big(g(x)\big),f'\big(g(x)\big)g'(x)\big)=F\big(g(x),f\big(g(x)\big),f'\big(g(x)\big)\big)(Ag)(x).$$

Now choose f such that $f(g(x))=g(x)$ and $f'(g(x))=1$. Then

$$F\big(x,g(x),g'(x)\big)=F\big(g(x),g(x),1\big)(Ag)(x).$$

We noted in the introduction of this section that $T(\mathrm{Id})(u)\neq 0$ for all $u\in\mathbb{R}$. Hence

$$F\big(g(x),g(x),1\big)=T(\mathrm{Id})\big(g(x)\big)\neq 0,$$

and hence we have

$$(Ag)(x)=\frac{F(x,g(x),g'(x))}{F(g(x),g(x),1)},\quad x\in\mathbb{R},\ g\in C^1(\mathbb{R}).$$

This means that A is also a local operator, and in fact

$$(Ag)(x) = (Tg)(x)/G_2\big(g(x)\big),$$

if we define $G_2 : \mathbb{R} \to \mathbb{R}_{\neq 0}$ by $G_2(u) := F(u, u, 1) = T(\mathrm{Id})(u) \neq 0$. Insert this into (6) to find

$$T(f \circ g)(x) = (Tf)\big(g(x)\big)(Tg)(x)/G_2\big(g(x)\big).$$

We introduce the operator $T' : C^1(\mathbb{R}) \to C(\mathbb{R})$ by

$$\big(T'f\big)(x) := (Tf)(x)/G_2(x).$$

Then for $f, g \in C^1(\mathbb{R})$,

$$\begin{aligned} T'(f \circ g)(x) &= T(f \circ g)(x)/G_2(x) \\ &= (Tf)\big(g(x)\big) \cdot (Tg)(x)/\big(G_2\big(g(x)\big) \cdot G_2(x)\big) \\ &= \big(T'f\big)\big(g(x)\big) \cdot \big(T'g\big)(x). \end{aligned}$$

Hence $T' : C^1(\mathbb{R}) \to C(\mathbb{R})$ satisfies the strong chain rule (1), and the (strong) non-degeneracy assumption on T implies that also T' is non-degenerate in the sense of (4).

By Theorem 1, applied to T', there are continuous functions $G_1 : \mathbb{R} \to \mathbb{R}_{\neq 0}$ and $K : \mathbb{R} \to \mathbb{R}$, $K(u) = |u|^p$ or $K(u) = |u|^p \operatorname{sgn}(u)$ for some $p > 0$ such that

$$\begin{aligned} F\big(x, f(x), f'(x)\big)/G_2(x) &= (Tf)(x)/G_2(x) \\ &= \big(T'f\big)(x) \\ &= \frac{(G_1 \circ f)(x)}{G_1(x)} K\big(f'(x)\big), \quad x \in \mathbb{R},\ f \in C^1(\mathbb{R}). \end{aligned}$$

Since the values of $f(x)$ and $f'(x)$ can be chosen arbitrarily, we find

$$F(x, y, z) = \frac{G_1(y)}{G_1(x)} G_2(x) K(z)$$

for all $x, y, z \in \mathbb{R}$. □

Proof of Theorem 2. The conclusion of Propositions 16 and 17 is that T and A have the form

$$\begin{aligned} (Tf)(x) &= \frac{G_1(f(x))}{G_1(x)} G_2(x) K\big(f'(x)\big), \\ (Af)(x) &= (Tf)(x)/G_2\big(f(x)\big), \quad x \in \mathbb{R},\ f \in C^1(\mathbb{R}) \end{aligned}$$

where G_1 is continuous and $K(u) = |u|^p$ or $K(u) = |u|^p \operatorname{sgn}(u)$.

Since $T(f)$ is continuous, also G_2 has to be continuous. □

5. The chain rule in $C^k(\mathbb{R})$ for $k \neq 1$

In this chapter we consider the chain rule (1) for functions in $\mathcal{D}(T) = C^k(\mathbb{R})$ for $k \neq 1$. We first prove Proposition 3 showing that there are no non-trivial examples of solutions to (1) if $k = 0$ and then consider the chain rule operation as maps

$$T : C^k \to C, \qquad T : C^\infty \to C, \qquad T : C^k \to C^{k-1}, \qquad T : C^\infty \to C^\infty$$

for $k \geqslant 2$. We start with the

Proof of Proposition 3. Assume that $T|_{C_b(\mathbb{R})} = 0$ is false. Then there are $g_1 \in C_b(\mathbb{R})$ and $x_1 \in \mathbb{R}$ with $(Tg_1)(x_1) \neq 0$. By assumption, there are also $g_0 \in C(\mathbb{R})$ and $x_0 \in \mathbb{R}$ with $(Tg)(x_0) = 0$. Using this, the analogues of Lemmas 4, 5, 6 and 8 may be proved in the same way for continuous functions as for differentiable functions, using $T(f \circ g) = (Tf) \circ g \cdot Tg$ for *all continuous* functions. In particular, for all open half-infinite intervals $J \subset \mathbb{R}$ and $f_1, f_2 \in C(\mathbb{R})$ with $f_1|_J = f_2|_J$, we have that $(Tf_1)|_{\bar{J}} = (Tf_2)|_{\bar{J}}$. Similarly as in the proof of Proposition 10, this can be used to prove that there is a function $F : \mathbb{R}^2 \to \mathbb{R}$ such that

$$Tf(x) = F\bigl(x, f(x)\bigr), \quad x \in \mathbb{R},\ f \in C(\mathbb{R}).$$

Namely, fix $x_0 \in \mathbb{R}$ and $f \in C(\mathbb{R})$ with a given value $f(x_0)$. Take any other function $g \in C(\mathbb{R})$ with $g(x_0) = f(x_0)$ and define $h \in C(\mathbb{R})$ by

$$h(x) := \begin{Bmatrix} g(x), & x \in J_1, \\ f(x), & x \in \bar{J}_2 \end{Bmatrix},$$

where $J_1 = (x_0, \infty)$, $J_2 = (-\infty, x_0)$. Then $h|_{J_1} = g|_{J_1}$ and $h|_{J_2} = f|_{J_2}$. Hence $(Tg)|_{\bar{J}_1} = (Th)|_{\bar{J}_1}$ and $(Th)|_{\bar{J}_2} = (Tf)|_{\bar{J}_2}$. Since $x_0 \in \bar{J}_1 \cap \bar{J}_2$, we conclude that $(Tg)(x_0) = (Th)(x_0) = (Tf)(x_0)$. Therefore the value $(Tf)(x_0)$ depends only on x_0 and $f(x_0)$, i.e. $Tf(x_0) = F(x_0, f(x_0))$ for a suitable function $F : \mathbb{R}^2 \to \mathbb{R}$ and for all $x_0 \in \mathbb{R}$, $f \in C(\mathbb{R})$.

By assumption, we may choose $x_0 \in \mathbb{R}$ and $g_0 \in C(\mathbb{R})$ with $(Tg_0)(x_0) = 0$. Take any $x \in \mathbb{R}$ and consider $h \in C(\mathbb{R})$, $h(s) = s + x_0 - x$. Then $h(x) = x_0$ and $g := g_0 \circ h$ satisfies by the chain rule

$$(Tg)(x) = T(g_0 \circ h)(x) = (Tg_0)(x_0)(Th)(x) = 0.$$

Hence for any x there is $g \in C(\mathbb{R})$ with $(Tg)(x) = 0$, i.e. $F(x, g(x)) = 0$. For any $f \in C(\mathbb{R})$, again by the chain rule

$$T(f \circ g)(x) = (Tf)\bigl(g(x)\bigr)(Tg)(x),$$

i.e. $F(x, f(g(x))) = F(g(x), f(g(x)))F(x, g(x)) = 0$. Since f may attain arbitrary values on $g(x)$, we conclude that $F(x, y) = 0$ for all $x, y \in \mathbb{R}$. Hence $Tf = 0$ for all $f \in C(\mathbb{R})$, contradicting the assumption $T|_{C_b(\mathbb{R})} \neq 0$. $\square$

To formulate an analogue of Theorem 1 for $C^k(\mathbb{R})$ and $C^\infty(\mathbb{R})$ functions, we need corresponding non-degeneracy conditions for T satisfying the chain rule (1). For $k \geqslant 2$, let $C_b^k(\mathbb{R}) :=$

$C_b^1(\mathbb{R}) \cap C^k(\mathbb{R})$ and $C_b^\infty(\mathbb{R}) := C_b^1(\mathbb{R}) \cap C^\infty(\mathbb{R})$. T is *non-degenerate* on $\mathcal{D}(T) = C^k(\mathbb{R})$ provided that

$$\exists x_1 \in \mathbb{R},\ h_1 \in C_b^k(\mathbb{R}) \quad (Th_1)(x_1) \neq 0. \tag{4_k}$$

Similarly (4_∞) denotes this if $h_1 \in C_b^\infty(\mathbb{R})$. The analogous conditions to (4.1) and (4.2) will be denoted by (4.1_k), (4.2_k). We then have:

Proposition 18. *Assume that $k \in \mathbb{N}_{\geqslant 2}$ and that $T : \mathcal{D}(T) = C^k(\mathbb{R}) \to \mathrm{Im}(T) \subset C(\mathbb{R})$ satisfies the chain rule*

$$T(f \circ g) = (Tf) \circ g \cdot Tg, \quad f, g \in \mathcal{D}(T). \tag{1}$$

If T is non-degenerate in the sense of (4_k), there exists $p \geqslant 0$ and a positive continuous function $H \in C(\mathbb{R})$ such that

$$Tf = \frac{H \circ f}{H} \cdot |f'|^p$$

or in the case of $p > 0$,

$$Tf = \frac{H \circ f}{H} \cdot |f'|^p \operatorname{sgn}(f').$$

This also holds for $k = \infty$. If (4.1_k) holds, we have $p > 0$. If (4.2_k) is valid and $T(2\,\mathrm{Id})(0) = 2$, $p = 1$ and T has the form $Tf = \frac{H\circ f}{H} \cdot f'$; the stronger assumption $T(2\,\mathrm{Id}) = 2$ implies even $Tf = f'$.

If the image of T consists of smooth functions, i.e. if $T : C^k(\mathbb{R}) \to C^{k-1}(\mathbb{R})$ or $T : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ satisfies (1) and (4_k) or (4_∞), the function H is in $C^{k-1}(\mathbb{R})$ or $C^\infty(\mathbb{R})$ and p satisfies

$$\big(p \in \{0, \dots, k-1\} \text{ or } p > k-1 \text{ for } k \in \mathbb{N}\big) \quad \text{or} \quad (p \in \mathbb{N}_0 \text{ for } k = \infty).$$

We need the following lemma for the Proof of Proposition 18.

Lemma 19. *Let $0 < a \leqslant 1$ and $L \in C(\mathbb{R})$ be a continuous function such that for any fixed $x_0 \in \mathbb{R}$,*

$$\psi(x) := L(x_0 + x) - aL\left(x_0 + \frac{x}{2}\right), \quad x \in \mathbb{R}$$

defines a C^1-function $\psi \in C^1(\mathbb{R})$. Then L is a C^1-function, $L \in C^1(\mathbb{R})$.

Proof. (a) We first prove the differentiability of L in x_0 where we may assume that $x_0 = 0$. Then $\psi(x) = L(x) - aL(\frac{x}{2})$. Since $\psi \in C^1(\mathbb{R})$ and $\psi(0) = (1-a)L(0)$, given $\varepsilon > 0$, there is $\delta > 0$ such that for all $|x| \leqslant \delta$, we have that

$$\left|\frac{\psi(x) - \psi(0)}{x} - \psi'(0)\right| \leqslant \varepsilon.$$

We use the following iteration procedure: for any $n \in \mathbb{N}$,

$$\sum_{j=0}^{n-1} a^j \psi\left(\frac{x}{2^j}\right) = L(x) - a^n L\left(\frac{x}{2^n}\right),$$

and replacing x successively by $x/2^j$, we find

$$\left|\frac{(L(x) - a^n L(\frac{x}{2^n})) - (\sum_{j=0}^{n-1} a^j)(1-a)L(0)}{x} - \left(\sum_{j=0}^{n-1} \frac{a^j}{2^j}\right)\psi'(0)\right| \leqslant \frac{\varepsilon}{1-a/2}$$

using $\sum_{j=0}^{\infty} \frac{a^j}{2^j} = \frac{1}{1-a/2}$. Since L is continuous in $x_0 = 0$, for $a < 1$, $a^n L(\frac{x}{2^n}) \to 0$. We may then take the limit for $n \to \infty$,

$$\left|\frac{L(x) - L(0)}{x} - \frac{\psi'(0)}{1-a/2}\right| \leqslant \frac{\varepsilon}{1-a/2} \leqslant 2\varepsilon.$$

This also holds if $a = 1$, then $L(\frac{x}{2^n}) \to L(0)$ and the term with $(1-a)$ disappears, yielding the same estimate. This shows that $L'(0) = \psi'(0)/(1-a/2)$ exists. Similarly, L is differentiable in any point $x_0 \in \mathbb{R}$, using the function ψ defined above in terms of x_0.

(b) We next prove the continuity of L' in $x_0 = 0$. By assumption, $\psi'(x) = L'(x) - \frac{a}{2}L'(\frac{x}{2})$ is continuous in $x_0 = 0$. Hence for $\varepsilon > 0$, there is $\delta > 0$ such that for $|x| \leqslant \delta$,

$$\left|L'(x) - \frac{a}{2}L'\left(\frac{x}{2}\right) - \left(1 - \frac{a}{2}\right)L'(0)\right| \leqslant \varepsilon.$$

Using a similar iteration technique as before,

$$\sum_{j=0}^{n-1} \left(\frac{a}{2}\right)^j \psi'\left(\frac{x}{2^j}\right) = L'(x) - \left(\frac{a}{2}\right)^n L'\left(\frac{x}{2^n}\right),$$

we find

$$\left|L'(x) - \left(\frac{a}{2}\right)^n L'\left(\frac{x}{2^n}\right) - \sum_{j=0}^{n-1} \left(\frac{a}{2}\right)^n \left(1 - \frac{a}{2}\right)L'(0)\right| \leqslant \frac{\varepsilon}{1-a/2} \leqslant 2\varepsilon.$$

The argument in (a) yields that L' remains bounded in a small neighborhood of 0 since $\psi \in C^1(\mathbb{R})$, and hence $(\frac{a}{2})^n L'(\frac{x}{2^n}) \to 0$ as $n \to \infty$. In the limit, we get for sufficiently small $|x| \leqslant \delta_2 \leqslant \delta$,

$$\left|L'(x) - L'(0)\right| \leqslant 2\varepsilon$$

and hence L is continuously differentiable in $x_0 = 0$ and any other $x_0 \in \mathbb{R}$ as well. □

Proof of Proposition 18. The proof is basically similar to the one of Theorem 1. We only stress and outline the modifications which are needed.

(a) Using the non-degeneracy condition (4_k) for $C^k(\mathbb{R})$ functions, the localization on intervals in Lemmas 4, 5, 6 and 8 works the same way; for the pointwise localization in the proof of Proposition 10 one has to use a function $h \in C^k(\mathbb{R})$ defined by f on one side of x_0 and by a Taylor approximation of the k-th order $g \in C^k(\mathbb{R})$ on the other side. This way, we may conclude that

$$Tf(x) = F\big(x, f(x), f'(x), f''(x), \ldots, f^{(k)}(x)\big), \quad f \in C^k(\mathbb{R}),\ x \in \mathbb{R} \tag{14}$$

holds for a suitable function $F : \mathbb{R}^{k+2} \to \mathbb{R}$. In the case of $C^\infty(\mathbb{R})$, $k = \infty$, $Tf(x)$ depends on x, $f(x)$ and all derivatives $f^{(j)}(x)$, $j \in \mathbb{N}$.

We claim that $Tf(x)$ actually does not depend on the values of the derivatives of order $\geqslant 2$, i.e. that $Tf(x) = \tilde{F}(x, f(x), f'(x))$. The further proof in the case of $T : C^k(\mathbb{R}) \to C(\mathbb{R})$ then proceeds exactly as the one for Theorem 1.

For $f, g \in C^k(\mathbb{R})$, the derivatives of $f \circ g$ have the form

$$(f \circ g)^{(k)} = f^{(k)} \circ g \cdot g'^k + \varphi_k\big(f' \circ g, \ldots, f^{(k-1)} \circ g, g', \ldots, g^{(k-1)}\big) + f' \circ g \cdot g^{(k)} \tag{15}$$

where φ_k depends only on the lower order derivatives of f and g up to the order $(k-1)$; e.g. $\varphi_2 = 0$, $\varphi_3(f' \circ g, f'' \circ g, g', g'') = 3f'' \circ g \cdot g' \cdot g''$.

Let us also remark that for any $x_0, y_0 \in \mathbb{R}$ and any sequence $(t_n)_{n\in\mathbb{N}}$ of real numbers there is $g \in C^\infty(\mathbb{R})$ with $g(x_0) = y_0$ and $g^{(n)}(x_0) = t_n$ for any $n \in \mathbb{N}$. This may be shown by adding infinitely many small bump functions, see [6], p. 16.

We now show successively that the function F with (14) does not depend on the variables with values $f''(x), \ldots, f^{(k)}(x)$. Starting with $f''(x)$, to simplify notation, we will not always write the further variables $f'''(x)$ etc. in detail, but just put $\ldots$ for them. We explain below how to work with $f'''(x)$ etc.

If $T : C^k(\mathbb{R}) \to C(\mathbb{R})$ satisfies (1), (14) means for functions $f, g \in C^k(\mathbb{R})$ with $g(x_0) = y_0$, $f(y_0) = z_0$, where $x_0, y_0, z_0 \in \mathbb{R}$ are arbitrary,

$$\begin{aligned} T(f \circ g)(x_0) &= F\big(x_0, z_0, f'(y_0)g'(x_0), f''(y_0)g'(x_0)^2 + g''(x_0)f'(y_0), \ldots\big) \\ &= (Tf)(y_0)(Tg)(x_0) \\ &= F\big(y_0, z_0, f'(y_0), f''(y_0), \ldots\big)F\big(x_0, y_0, g'(x_0), g''(x_0), \ldots\big), \end{aligned} \tag{16}$$

also for $k = \infty$. If $x_0 = z_0$, also $(g \circ f)(y_0)$ is defined and

$$T(f \circ g)(x_0) = (Tf)(y_0)(Tg)(x_0) = (Tg)(x_0)(Tf)(y_0) = T(g \circ f)(y_0),$$

i.e.

$$\begin{aligned} &F\big(x_0, x_0, f'(y_0)g'(x_0), f''(y_0)g'(x_0)^2 + g''(x_0)f'(y_0), \ldots\big) \\ &\quad = F\big(y_0, y_0, g'(x_0)f'(y_0), f''(y_0)g'(x_0) + g''(x_0)f'(y_0)^2, \ldots\big). \end{aligned}$$

Choosing arbitrary real sequences $(t_n)_{n\in\mathbb{N}}$, $(s_n)_{n\in\mathbb{N}}$, we find functions $f, g \in C^\infty(\mathbb{R})$ with $g(x_0) = y_0$, $f(y_0) = x_0$, $g^{(n)}(x_0) = t_n$, $f^{(n)}(y_0) = s_n$ for all $n \in \mathbb{N}$. Therefore

$$F\big(x_0, x_0, s_1t_1, s_2t_1^2 + s_1t_2, \ldots\big) = F\big(y_0, y_0, s_1t_1, s_2t_1 + s_1^2t_2, \ldots\big). \tag{17}$$

Since the (t_n) and (s_n) were chosen independently of x_0 and y_0, the left and right side of (17) does not depend on x_0 and y_0.

Let $K(u_1, u_2, \ldots) := F(x_0, x_0, u_1, u_2, \ldots)$. Then (17) means that

$$K\big(s_1t_1, t_1^2s_2 + s_1t_2, \ldots\big) = K\big(s_1t_1, t_1s_2 + s_1^2t_2, \ldots\big).$$

For $k \geqslant 3$ or $k = \infty$, the structure of the k-th derivative of $f \circ g$ given by (15) means that in more precise notation

$$\begin{aligned} &K\big(s_1t_1, t_1^2s_2 + s_1t_2, \ldots, t_1^ks_k + s_1t_k + \varphi_{k1}, \ldots\big) \\ &\quad = K\big(s_1t_1, t_1s_2 + s_1^2t_2, \ldots, t_1s_k + s_1^kt_k + \varphi_{k2}, \ldots\big), \end{aligned} \tag{18}$$

where $\varphi_{k1}, \varphi_{k2} \in \mathbb{R}$ depend only on the values of $s_1, \ldots, s_{k-1}$ and $t_1, \ldots, t_{k-1}$. The notation here is for the $C^\infty(\mathbb{R})$ case; for $C^k(\mathbb{R})$ the last dots in (18) and also below should be dropped. Assume that $s_1t_1 \neq +1, -1$ and $s_1t_1 \neq 0$. We claim that, given arbitrary values of $a_2, \ldots, a_k, \ldots$ and $b_2, \ldots, b_k, \ldots$, we may choose $s_2, t_2, \ldots, s_k, t_k, \ldots$ such that (18) yields

$$K(s_1t_1, a_2, \ldots, a_k, \ldots) = K(s_1t_1, b_2, \ldots, b_k, \ldots), \tag{19}$$

i.e. that K only depends on the first variable $u = s_1t_1$ (if $u \neq 0, 1, -1$). Note that $\det\begin{pmatrix} t_1^k & s_1 \\ t_1 & s_1^k \end{pmatrix} = (s_1t_1)((s_1t_1)^{k-1} - 1) \neq 0$. Thus successively, we may solve uniquely the linear equations for $s_2, t_2, \ldots, s_k, t_k$, starting with $j = 2$ and continuing up to $j = k$ (and further in the C^∞-case), and using the obtained values of $s_2, t_2, \ldots$ in the later equations to determine the values of $\varphi_{k1}, \varphi_{k2}$,

$$\begin{aligned} &t_1^2s_2 + s_1t_2 = a_2, \qquad t_1s_2 + s_1^2t_2 = b_2, \qquad \ldots, \\ &t_1^ks_k + s_1t_k = a_k - \varphi_{k1}, \qquad t_1s_k + s_1^kt_k = b_k - \varphi_{k2}. \end{aligned}$$

This means that (18) implies (19) and for $u_1 \neq 0, 1, -1$, $K(u_1, u_2, \ldots)$ is independent of all variables (u_j), $j \geqslant 2$, and we write it as $\tilde{K}(u_1)$.

In the case of $s_1t_1 = 1$, putting $x_0 = y_0 = z_0$ in (16), we find that for any $s_2, \ldots, s_k, t_2, \ldots, t_k$ (and choosing $t_1 = 2$, $s_1 = 1/2$),

$$K\big(1, 4s_2 + t_2/2, \ldots, 2^ks_k + t_k/2^k + \varphi_k, \ldots\big) = \tilde{K}(2)\tilde{K}(1/2). \tag{20}$$

Again for arbitrary values of $a_2, \ldots, a_k$, we find successively $s_2, t_2, \ldots, s_k, t_k$ such that the left side of (20) equals $K(1, a_2, \ldots, a_k, \ldots)$ and therefore $\tilde{K}(1) = K(1, \ldots)$ is also independent of the variables (u_j), $j \geqslant 2$. Similarly for $K(-1, \ldots)$.

To show that also $K(0, \ldots)$ is independent of the further variables, choose $t_1 = a$, $s_1 = 0$ (thus $s_1t_1 = 0$) in (18) to conclude

$$K\big(0, a^2 s_2, \ldots, a^k s_k + \varphi_{k1}, \ldots\big) = K(0, as_2, \ldots, as_k + \varphi_{k2}, \ldots)$$

which again implies independence of the further variables. We now write again K instead of $\tilde{K}$: we know that $K(u_1) = F(x_0, x_0, u_1, u_2, \ldots)$ is independent of x_0 and $(u_2, \ldots)$. For values $y_0 \neq x_0$, we have by (16),

$$F(x_0, y_0, t_1, t_2, \ldots) = \frac{K(s_1 t_1)}{F(y_0, x_0, s_1, s_2, \ldots)}.$$

Since the left side is independent of $s_1, s_2, \ldots$ and the right side is independent of $t_2, t_3, \ldots$, the equation is of the form

$$F(x_0, y_0, t) = K(t)/F(y_0, x_0, 1).$$

From here, the same arguments as in the proof of Theorem 1 yield that

$$F(x_0, y_0, t) = H(y_0)/H(x_0) t^p \big[\operatorname{sgn}(t)\big]$$

with $p \geqslant 0$ and $H \in C(\mathbb{R})$, i.e. that T has the form given in Proposition 18.

(b) If $H \in C^{k-1}(\mathbb{R})$ and $p \in \{0, \ldots, k-1\}$ or $p > k-1$, the chain rule (1) defines a map $T : C^k(\mathbb{R}) \to C^{k-1}(\mathbb{R})$ into smooth functions; the condition on p is obviously needed to have continuous derivatives in zeros of f.

Let us conversely assume that T maps $C^k(\mathbb{R})$ into $C^{k-1}(\mathbb{R})$ and verifies (1) where $k \in \mathbb{N}$, $k \geqslant 2$. We know that T has the form

$$Tf = H \circ f/H \big|f'\big|^p \big[\operatorname{sgn}\big(f'\big)\big]$$

where $H \in C(\mathbb{R})$, $H > 0$. We claim that H is smooth, namely $H \in C^{k-1}(\mathbb{R})$. Let $L := -\log H$. Obviously $L \in C^{k-1}(\mathbb{R})$ if and only if $H \in C^{k-1}(\mathbb{R})$. Take $f(x) = \frac{x}{2}$. By assumption $Tf \in C^{k-1}(\mathbb{R})$ and hence $\varphi(x) := L(x) - L(\frac{x}{2})$ defines a function $\varphi \in C^{k-1}(\mathbb{R})$. We prove by induction on $k \geqslant 2$ that $\varphi \in C^{k-1}(\mathbb{R})$ and $L \in C^{k-2}(\mathbb{R})$ implies that $L \in C^{k-1}(\mathbb{R})$.

For $k = 2$, $T : C^2(\mathbb{R}) \to C^1(\mathbb{R})$, $\varphi \in C^1(\mathbb{R})$. Since $H \in C(\mathbb{R})$, also $L \in C(\mathbb{R})$. By Lemma 19, with $\psi = \varphi$ and $a = 1$, $L \in C^1(\mathbb{R})$.

To prove the induction step, assume that $k \geqslant 3$ and $\varphi \in C^{k-1}(\mathbb{R})$ as well as $L \in C^{k-2}(\mathbb{R})$ holds. We have to prove that $L \in C^{k-1}(\mathbb{R})$.

Define $\psi(x) := \varphi^{(k-2)}(x) = L^{(k-2)}(x) - \frac{1}{2^{k-2}} L^{(k-2)}(\frac{x}{2})$. Then $\psi \in C^1(\mathbb{R})$ and $L^{(k-2)} \in C(\mathbb{R})$. By Lemma 19, with $a = 1/2^{k-2}$, $L^{(k-2)} \in C^1(\mathbb{R})$, i.e. $L \in C^{k-1}(\mathbb{R})$. Hence $H \in C^{k-1}(\mathbb{R})$ is true if $T : C^k(\mathbb{R}) \to C^{k-1}(\mathbb{R})$ satisfies (1). □

References

[1] S. Alesker, S. Artstein-Avidan, V. Milman, A characterization of the Fourier transform and related topics, in: A. Alexandrov, et al. (Eds.), Linear and Complex Analysis, in: Adv. Math. Sci., vol. 63, 2009, pp. 11–26, dedicated to V.P. Havin, Amer. Math. Soc. Transl., vol. 226.

[2] S. Alesker, S. Artstein-Avidan, D. Faifman, V. Milman, A characterization of product preserving maps with applications to a characterization of the Fourier transform, in press.

[3] S. Artstein-Avidan, V. Milman, The concept of duality for measure projections of convex bodies, J. Funct. Anal. 254 (2008) 2648–2666.

[4] S. Artstein-Avidan, V. Milman, The concept of duality in convex analysis, and the characterization of the Legendre transform, Ann. of Math. 169 (2009) 661–674.
[5] S. Banach, Sur l'équation fonctionelle $f(x+y)=f(x)+f(y)$, Fund. Math. 1 (1920) 123–124.
[6] L. Hörmander, The Analysis of Linear Partial Differential Operators I, Springer, 1983.
[7] V. Milman, R. Schneider, Characterizing the mixed volume, submitted for publication.
[8] W. Sierpinski, Sur l'équation fonctionelle $f(x+y)=f(x)+f(y)$, Fund. Math. 1 (1920) 116–122.