A Characterization of Continuous differentiability of Proximal Mappings of Composite Functions

Ebrahim Sarabi

Department of Mathematics
Miami University

Research is partially supported by the U.S. National Science Foundation under the grant DMS 2108546.

Based on the Joint works with N.T.V. Hang (Institute of Mathematics, Vietnam)

2022 Midwest Optimization Meeting (University of Waterloo)
Motivation

Proto-Differentiability

Strict Proto-Differentiability

Smoothness of Proximal Mappings
Motivation

Proto-Differentiability

Strict Proto-Differentiability

Smoothness of Proximal Mappings
Recall that for a convex function \( f : \mathbb{R}^n \rightarrow \mathbb{R} = [-\infty, \infty] \) and parameter value \( r > 0 \), the proximal mapping of \( f \), denoted by \( \text{prox}_{r} f \), is defined by

\[
\text{prox}_{r} f (x) = \arg\min_{w \in \mathbb{R}^n} \left\{ f(w) + \frac{1}{2r} \|w - x\|^2 \right\}, \quad x \in \mathbb{R}^n.
\]

When \( f = \delta_C \), namely the indicator function of a convex set \( C \subset \mathbb{R}^n \), this mapping reduces to the projection mapping of \( C \), defined by

\[
P_C(x) = \arg\min \left\{ \|w - x\|^2 \mid w \in C \right\}, \quad x \in \mathbb{R}^n.
\]

- **Question.** At what points is \( P_C \) is continuously differentiable \((C^1)\)?

---

\(^1\)“In spite of the elementary formulation of this question, a full answer is so far unknown.” J.-B. Hiriart-Urruty, At what points is the projection mapping differentiable? Amer. Math. Monthly 89(7), 456–458 (1982)
What we know so far:

- The projection mapping $P_C$ may fail to be differentiable in general\(^2\). For instance, assume that $C$ is the unit ball and $x$ is a vector that $\|x\| = 1$. Then $P_C$ fails to be continuously differentiable at $x$.

What we know so far:

• R. Holmes \(^3\) studied the smoothness of projection mapping onto a closed convex set in Hilbert spaces. His main result states that if \( C \subset \mathbb{R}^n \) is a closed convex set, \( x \in \mathbb{R}^n \), the boundary of \( C \) is a \( C^2 \) smooth manifold around \( y = P_C(x) \), then the projection mapping \( P_C \) is \( C^1 \) in a neighborhood of the open normal ray \( \{ y + t(x - y) | t > 0 \} \).

\[ x - y \]

\[ x \]

\[ y = P_C(x) \]

\[ C \]

• When the projection point \( y \) is a corner point, Holmes’s result fails because the boundary of \( C \) is not a \( C^2 \) smooth manifold around \( y \).

---

\(^3\)R.B. Holmes, Smoothness of certain metric projections on Hilbert space. Trans. Amer. Math. Soc. 183, 87–100 (1973)
What we know so far:

**Theorem.** (Facchinei-Pang\(^4\)) Assume that \( C \) is a polyhedral convex set. Then the projection mapping \( P_C \) is differentiable at \( x \) if and only if \( x - y \in \text{ri} \, N_C(y) \), where \( y = P_C(x) \).

---


\(^5\)\( N_C(\bar{x}) = \left\{ v \in \mathbb{R}^n : \langle v, x - \bar{x} \rangle \leq 0 \text{ for all } x \in C \right\} \)
What we know so far:

**Theorem.** (Facchini-Pang\(^4\)) Assume that \( C \) is a polyhedral convex set. Then the projection mapping \( P_C \) is differentiable at \( x \) if and only if \( x - y \in \text{ri } N_C(y) \), where \( y = P_C(x) \).

- \( P_C \) is not differentiable at \( y \) since \( 0 \notin \text{ri } N_C(y) \).

---


\(^5\) \( N_C(\bar{x}) = \left\{ v \in \mathbb{R}^n | \langle v, x - \bar{x} \rangle \leq 0 \text{ for all } x \in C \right\} \)
What we know so far:

- The projection mapping $P_C$ is always directionally differentiable if we assume a second-order regularity on $C$ such as parabolic regularity. Recall that a function $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is directionally differentiable at $\bar{x}$ if the following limit exists for any $w \in \mathbb{R}^n$:

$$\lim_{t \downarrow 0} \frac{g(\bar{x} + tw) - g(\bar{x})}{t}.$$

---

What we know so far:

- The projection mapping $P_C$ is always directionally differentiable if we assume a second-order regularity on $C$ such as parabolic regularity $^6$. Recall that a function $g : \mathbb{R}^n \to \mathbb{R}^m$ is directionally differentiable at $\bar{x}$ if the following limit exists for any $w \in \mathbb{R}^n$:

\[
\lim_{t \downarrow 0} \frac{g(\bar{x} + tw) - g(\bar{x})}{t}.
\]

- We likely need a second-order regularity to ensure continuous differentiability of the projection mapping onto a closed convex (prox-regular) set.

---

What we know so far:

Assume that $C \subset \mathbb{R}^n$ is a $C^2$ smooth manifold around a point $\bar{x} \in C$, meaning that there exists a neighborhood $O$ of $\bar{x}$ on which $C$ has the representation

$$C \cap O = \{ x \in O | \Phi(x) = 0 \},$$

where $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a $C^2$ function with $\nabla \Phi(\bar{x})$ having full rank.

---

What we know so far:

Assume that $C \subset \mathbb{R}^n$ is a $C^2$ smooth manifold around a point $\bar{x} \in C$, meaning that there exists a neighborhood $O$ of $\bar{x}$ on which $C$ has the representation

$$C \cap O = \{ x \in O | \Phi(x) = 0 \},$$

where $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a $C^2$ function with $\nabla \Phi(\bar{x})$ having full rank.

- It is well-known that the projection mapping $P_C$ is locally single-valued and Lipschitz continuous and directionally differentiable.
- Lewis and Malick\(^7\) showed that $P_C$ is $C^1$ around $\bar{x}$.

Motivation

Proto-Differentiability

Strict Proto-Differentiability

Smoothness of Proximal Mappings
• Given $C \subset \mathbb{R}^n$ and $\bar{x} \in C$, recall that the tangent cone and the adjacent cone to $C$ at $\bar{x}$ are defined, respectively, by

$$T_C(\bar{x}) = \limsup_{t \downarrow 0} \frac{C - \bar{x}}{t}$$
and

$$A_C(\bar{x}) = \liminf_{t \downarrow 0} \frac{C - \bar{x}}{t},$$

where both limits are understood in the sense of Painlevé-Kuratowski.

\[8\] $T_C(\bar{x}) = \{ w \in \mathbb{R}^n | \exists t_k \downarrow 0, \ w_k \to w \text{ as } k \to \infty \text{ with } \bar{x} + t_k w_k \in C \}$

\[9\] $A_C(\bar{x}) = \{ w \in \mathbb{R}^n | \forall t_k \downarrow 0 \ \exists w_k \to w \text{ as } k \to \infty \text{ with } \bar{x} + t_k w_k \in C \}$
Given $C \subset \mathbb{IR}^n$ and $\bar{x} \in C$, recall that the tangent cone and the adjacent cone to $C$ at $\bar{x}$ are defined, respectively, by

$$T_C(\bar{x}) = \limsup_{t \downarrow 0} \frac{C - \bar{x}}{t} \quad \text{and} \quad A_C(\bar{x}) = \liminf_{t \downarrow 0} \frac{C - \bar{x}}{t},$$

where both limits are understood in the sense of Painlevé-Kuratowski.

- Clearly we always have $A_C(\bar{x}) \subset T_C(\bar{x})$.

- **Definition.** Suppose that $f : \mathbb{IR}^n \to \mathbb{IR}$ is a convex function and $(\bar{x}, \bar{v}) \in \text{gph} \partial f$. We say $\partial f$ is proto-differentiable at $\bar{x}$ for $\bar{v}$ if

$$A_{\text{gph} \partial f}(\bar{x}, \bar{v}) = T_{\text{gph} \partial f}(\bar{x}, \bar{v}),$$

where

$$\text{gph} \partial f = \{(x, v) \in \mathbb{IR}^n \times \mathbb{IR}^n | v \in \partial f(x)\}.$$
**Theorem.** (Rockafellar\textsuperscript{10} (1990)). Suppose that $f : \mathbb{IR}^n \to \mathbb{IR}$ is a proper convex function and $(\bar{x}, \bar{v}) \in \text{gph} \partial f$. Then the following properties are equivalent:

- $\partial f$ is proto-differentiable at $\bar{x}$ for $\bar{v}$;
- $\text{prox}_f$ is directionally differentiable at $\bar{x} + \bar{v}$.

The proof is based on the identity

$$\text{prox}_f = (I + \partial f)^{-1},$$

which holds for any convex functions.

---

Theorem. (Rockafellar\textsuperscript{10} (1990)). Suppose that $f : \mathbb{R}^n \to \mathbb{IR}$ is a proper convex function and $(\bar{x}, \bar{v}) \in \text{gph} \partial f$. Then the following properties are equivalent:

- $\partial f$ is proto-differentiable at $\bar{x}$ for $\bar{v}$;
- $\text{prox}_f$ is directionally differentiable at $\bar{x} + \bar{v}$.

The proof is based on the identity

$$(w, q) \in T_{\text{gph} \partial f}(\bar{x}, \bar{v}) \iff (w + q, w) \in T_{\text{gph} \text{prox}_f}(\bar{x} + \bar{v}, \bar{y})$$

which holds for any convex functions.

---

Theorem. (Rockafellar\textsuperscript{10} (1990)). Suppose that $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a proper convex function and $(\bar{x}, \bar{v}) \in \text{gph} \partial f$. Then the following properties are equivalent:

- $\partial f$ is proto-differentiable at $\bar{x}$ for $\bar{v}$;
- $\text{prox}_f$ is directionally differentiable at $\bar{x} + \bar{v}$.

The proof is based on the identity

$$\text{prox}_f = (I + \partial f)^{-1},$$

which holds for any convex functions. Proto-differentiability holds for many important sets and functions including

- polyhedral convex sets, the second-order cone, the cone of positive semidefinite symmetric matrices;
- polyhedral functions; convex piecewise linear-quadratic functions, spectral functions.

Motivation

Proto-Differentiability

Strict Proto-Differentiability

Smoothness of Proximal Mappings
• Given $C \subset \mathbb{R}^n$ and $\bar{x} \in C$, recall that the paratingent cone and the regular (Clarke) tangent cone to $C$ at $\bar{x}$ are defined, respectively, by

$$\hat{T}_C(\bar{x}) = \limsup_{x \to \bar{x}, t \downarrow 0} \frac{C - \bar{x}}{t} \quad \text{and} \quad \tilde{T}_C(\bar{x}) = \liminf_{x \to \bar{x}, t \downarrow 0} \frac{C - x}{t},$$

where both limits are understood in the sense of Painlevé-Kuratowski.

\[\text{\cite{Poliquin}}\]
Given $C \subset \mathbb{R}^n$ and $\bar{x} \in C$, recall that the paratingent cone and the regular (Clarke) tangent cone to $C$ at $\bar{x}$ are defined, respectively, by

$$\hat{T}_C(\bar{x}) = \limsup_{x \to \bar{x}, t \downarrow 0} \frac{C - \bar{x}}{t} \quad \text{and} \quad \tilde{T}_C(\bar{x}) = \liminf_{x \to \bar{x}, t \downarrow 0} \frac{C - x}{t},$$

where both limits are understood in the sense of Painlevé-Kuratowski.

- Clearly we always have $\tilde{T}_C(\bar{x}) \subset \hat{T}_C(\bar{x})$.

- **Definition.** Suppose that $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a convex function and $(\bar{x}, \bar{v}) \in \text{gph } \partial f$. We say $\partial f$ is strictly proto-differentiable \(^{11}\) at $\bar{x}$ for $\bar{v}$ if

$$\hat{T}_{\text{gph } \partial f}(\bar{x}, \bar{v}) = \tilde{T}_{\text{gph } \partial f}(\bar{x}, \bar{v}).$$

Theorem. (Poliquin-Rockafellar\textsuperscript{12} (1996)). Suppose that $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a proper convex function and $(\bar{x}, \bar{v}) \in \text{gph} \partial f$. Then the following properties are equivalent:

- $\partial f$ is strictly proto-differentiable at $x$ for $v$ for any $(x, v) \in \text{gph} \partial f$ sufficiently close to $(\bar{x}, \bar{v})$;
- for any $r > 0$, $\text{prox}_{r f}$ is continuously differentiable in a neighborhood of $\bar{x} + r\bar{v}$.


Theorem. (Poliquin-Rockafellar\textsuperscript{12} (1996)). Suppose that $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ is a proper convex function and $(\bar{x}, \bar{v}) \in \text{gph} \, \partial f$. Then the following properties are equivalent:

- $\partial f$ is strictly proto-differentiable at $x$ for $v$ for any $(x, v) \in \text{gph} \, \partial f$ sufficiently close to $(\bar{x}, \bar{v})$;
- for any $r > 0$, $\text{prox}_{r f}$ is continuously differentiable in a neighborhood of $\bar{x} + r \bar{v}$.

Poliquin-Rockafellar showed that this result holds for prox-regular functions at $\bar{x}$ for $\bar{v} = 0$ provided that $\bar{x} \in \text{argmin} \, f$. It is, however, possible to show that the latter condition can be dropped using the stability properties of generalized equations.\textsuperscript{13}.


Theorem. (Poliquin-Rockafellar\textsuperscript{12} (1996)). Suppose that $f: \mathbb{R}^n \to \overline{\mathbb{R}}$ is a proper convex function and $(\bar{x}, \bar{v}) \in \text{gph} \partial f$. Then the following properties are equivalent:

- $\partial f$ is strictly proto-differentiable at $x$ for $v$ for any $(x, v) \in \text{gph} \partial f$ sufficiently close to $(\bar{x}, \bar{v})$;
- for any $r > 0$, $\text{prox}_{r f}$ is continuously differentiable in a neighborhood of $\bar{x} + r \bar{v}$.

Poliquin-Rockafellar showed that this result holds for prox-regular functions at $\bar{x}$ for $\bar{v} = 0$ provided that $\bar{x} \in \text{argmin} f$. It is, however, possible to show that the latter condition can be dropped using the stability properties of generalized equations.\textsuperscript{13}.

Question. When does strict proto-differentiability hold?


Recall that $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ is called polyhedral if $\text{epi } f$ is a polyhedral convex set. Important examples of polyhedral functions include

- the indicator function of a polyhedral convex set;
- $f(x) = \max\{\langle a_i, x \rangle + \alpha_i \mid i = 1, \ldots, m\}$ with $a_i \in \mathbb{R}^n$ and $\alpha_i \in \mathbb{R}$.

---

Recall that $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is called polyhedral if $\text{epi } f$ is a polyhedral convex set. Important examples of polyhedral functions include

- the indicator function of a polyhedral convex set;
- $f(x) = \max \{ \langle a_i, x \rangle + \alpha_i \mid i = 1, \ldots, m \}$ with $a_i \in \mathbb{R}^n$ and $\alpha_i \in \mathbb{R}$.

**Theorem.** (Hang-S\textsuperscript{14} (2022)). Suppose that $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a polyhedral function and $(\bar{x}, \bar{v}) \in \text{gph } \partial f$. Then the following properties are equivalent:

- $\partial f$ is strictly proto-differentiable at $x$ for $v$ for any $(x, v) \in \text{gph } \partial f$ sufficiently close to $(\bar{x}, \bar{v})$;
- $\bar{v} \in \text{ri } \partial f(\bar{x})$.

**Theorem.** (Hang-S\textsuperscript{15} (2022)). Suppose that $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a polyhedral function and $(\bar{x}, \bar{v}) \in \text{gph} \partial f$. Then the following properties are equivalent:

- for any $r > 0$, $\text{prox}_{r f}$ is continuously differentiable in a neighborhood of $\bar{x} + r \bar{v}$;
- $\bar{v} \in \text{ri} \partial f(\bar{x})$.

---

**Theorem.** (Hang-S\textsuperscript{15} (2022)). Suppose that $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a polyhedral function and $(\bar{x}, \bar{v}) \in \text{gph} \partial f$. Then the following properties are equivalent:

- for any $r > 0$, $\text{prox}_{r f}$ is continuously differentiable in a neighborhood of $\bar{x} + r \bar{v}$;
- $\bar{v} \in \text{ri} \partial f(\bar{x})$.

**Corollary.** (Hang-S (2022)). Assume that $C \subset \mathbb{R}^n$ is a polyhedral convex set and $x \in \mathbb{R}^n$. Then $P_C$ is continuously differentiable in a neighborhood of $x$ if and only if $x - y \in \text{ri} N_C(z)$, where $y = P_C(x)$.

For the polyhedral set $C$, $P_C$ is continuously differentiable at $\bar{x} + \bar{v}_1$ but is not continuously differentiable at $\bar{x} + \bar{v}_2$. 
• Similar results\textsuperscript{16} were established recently for the composite function
\[ f \circ \Phi, \]
where \( f \) is a polyhedral function and \( \Phi \) is a \( C^2 \) function, and the constraint qualification
\[ \text{par}\{\partial f(\Phi(\bar{x}))\}^{17} \cap \ker \nabla \Phi(\bar{x})^* = \{0\} \]
is satisfied at \( \bar{x} \in \mathbb{R}^n \) with \( \Phi(\bar{x}) \in \text{dom} \ f \).


\textsuperscript{17}the linear subspace parallel to the affine hull of \( \partial f(\Phi(\bar{x})) \).

• Similar results\textsuperscript{16} were established recently for the composite function

\[ f \circ \Phi, \]

where \( f \) is a polyhedral function and \( \Phi \) is a \( C^2 \) function, and the constraint qualification

\[ \text{par}\{\partial f(\Phi(\bar{x}))\}\cap \ker \nabla \Phi(\bar{x})^* = \{0\} \]

is satisfied at \( \bar{x} \in \mathbb{R}^n \) with \( \Phi(\bar{x}) \in \text{dom} \ f \).

• The condition above boils down to the classical linear independent constraint qualification when \( f = \delta_{\mathbb{R}^m \times \{0\}^{n-m}} \) with \( 0 \leq m \leq n \).

• This composite function is prox-regular and thus its proximal mapping is locally single-valued and Lipschitz continuous.\textsuperscript{18}


\textsuperscript{17}the linear subspace parallel to the affine hull of \( \partial f(\Phi(\bar{x})) \).

Motivation

Proto-Differentiability

Strict Proto-Differentiability

Smoothness of Proximal Mappings
Theorem. (Hang-S (2022)). Given the composite function $g = f \circ \Phi$ with $(\bar{x}, \bar{v}) \in \text{gph } \partial g$, the following properties are equivalent:

- $\partial g$ is strictly proto-differentiable at $x$ for $v$ for any $(x, v) \in \text{gph } \partial g$ sufficiently close to $(\bar{x}, \bar{v})$;
- $\bar{v} \in \text{ri } \partial g(\bar{x})$.

---

$^{19}$ the limiting subdifferential of $g$
Theorem.( Hang-S (2022)). Given the composite function $g = f \circ \Phi$ with $(\bar{x}, \bar{v}) \in \text{gph} \partial g$, the following properties are equivalent:

- $\partial g^{19}$ is strictly proto-differentiable at $x$ for $v$ for any $(x, v) \in \text{gph} \partial g$ sufficiently close to $(\bar{x}, \bar{v})$;
- $\bar{v} \in \text{ri} \partial g(\bar{x})$.

Theorem( Hang-S (2022)). For the composite function $g = f \circ \Phi$ with $(\bar{x}, \bar{v}) \in \text{gph} \partial g$, the following properties are equivalent:

- $\bar{v} \in \text{ri} \partial g(\bar{x})$;
- for any $r > 0$ sufficiently small, the proximal mapping $\text{prox}_{rg}$ is continuously differentiable in a neighborhood of $\bar{x} + r\bar{v}$.

\textsuperscript{19} the limiting subdifferential of $g$
Assume that $C \subset \mathbb{R}^n$ is fully amenable around a point $\bar{x} \in C$, meaning that there exists a neighborhood $O$ of $\bar{x}$ on which $C$ has the representation

$$C \cap O = \{ x \in O | \Phi(x) \in \Theta \},$$

where $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a $C^2$ function and $\Theta \subset \mathbb{R}^m$ is a polyhedral convex set, and the condition

$$\text{span}\{ N_C(\Phi(\bar{x})) \}^{20} \cap \text{ker} \nabla \Phi(\bar{x})^* = \{0\}$$

holds.

\[^{20}\text{the linear subspace } N_C(\Phi(\bar{x})).\]
Assume that $C \subset \mathbb{R}^n$ is fully amenable around a point $\bar{x} \in C$, meaning that there exists a neighborhood $O$ of $\bar{x}$ on which $C$ has the representation

$$C \cap O = \{ x \in O | \Phi(x) \in \Theta \},$$

where $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ is a $C^2$ function and $\Theta \subset \mathbb{R}^m$ is a polyhedral convex set, and the condition

$$\text{span}\{N_C(\Phi(\bar{x}))\}^{20} \cap \ker \nabla \Phi(\bar{x})^* = \{0\}$$

holds.

**Theorem** (Hang-S (2022)). For a fully amenable set $C$ with $(\bar{x}, \bar{v}) \in \text{gph } N_C$, the following properties are equivalent:

1. $\bar{v} \in \text{ri } N_C(\bar{x})$;
2. for any $r > 0$ sufficiently small, the projection mapping $P_C$ is continuously differentiable in a neighborhood of $\bar{x} + r\bar{v}$.

---

$^{20}$the linear subspace $N_C(\Phi(\bar{x}))$. 

---

22
**Example.** Assume that $C$ is the unit ball in $\mathbb{R}^n$. Then $C$ is full amenable at every point $x \in C$ since

$$C = \{ x \in \mathbb{R}^n | \Phi(x) \leq 0 \} \text{ with } \Phi(x) = \|x\|^2 - 1.$$ 

If $\|x\| = 1$, then we have $0 \notin \text{ri } N_C(x)$ and thus $P_C$ can’t be continuously differentiable around $x$. 

![Diagram of the unit ball and normal cone](attachment:image.png)
References:


Thank you for your attention!