

# **A Tutorial on Hadamard Semidifferentials**

## Other titles in Foundations and Trends® in Optimization

*Massively Parallel Computation: Algorithms and Applications*

Sungjin Im, Ravi Kumar, Silvio Lattanzi, Benjamin Moseley  
and Sergei Vassilvitskii

ISBN: 978-1-63828-216-7

*Acceleration Methods*

Alexandre d'Aspremont, Damien Scieur and Adrien Taylor

ISBN: 978-1-68083-928-9

*Atomic Decomposition via Polar Alignment: The Geometry of  
Structured Optimization*

Zhenan Fan, Halyun Jeong, Yifan Sun and Michael P. Friedlander

ISBN: 978-1-68083-742-1

*Optimization Methods for Financial Index Tracking: From Theory  
to Practice*

Konstantinos Benidis, Yiyong Feng and Daniel P. Palomar

ISBN: 978-1-68083-464-2

*The Many Faces of Degeneracy in Conic Optimization*

Dmitriy Drusvyatskiy and Henry Wolkowicz

ISBN: 978-1-68083-390-4

# A Tutorial on Hadamard Semidifferentials

---

Kenneth Lange  
UCLA  
klange@ucla.edu

**now**

the essence of knowledge

Boston — Delft

**AVA** Documents  
Document Sharing Hub

# Foundations and Trends<sup>®</sup> in Optimization

*Published, sold and distributed by:*

now Publishers Inc.  
PO Box 1024  
Hanover, MA 02339  
United States  
Tel. +1-781-985-4510  
[www.nowpublishers.com](http://www.nowpublishers.com)  
[sales@nowpublishers.com](mailto:sales@nowpublishers.com)

*Outside North America:*

now Publishers Inc.  
PO Box 179  
2600 AD Delft  
The Netherlands  
Tel. +31-6-51115274

The preferred citation for this publication is

K. Lange. *A Tutorial on Hadamard Semidifferentials*. Foundations and Trends<sup>®</sup> in Optimization, vol. 6, no. 1, pp. 1–62, 2024.

ISBN: 978-1-63828-349-2

© 2024 K. Lange

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, mechanical, photocopying, recording or otherwise, without prior written permission of the publishers.

Photocopying. In the USA: This journal is registered at the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923. Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by now Publishers Inc for users registered with the Copyright Clearance Center (CCC). The 'services' for users can be found on the internet at: [www.copyright.com](http://www.copyright.com)

For those organizations that have been granted a photocopy license, a separate system of payment has been arranged. Authorization does not extend to other kinds of copying, such as that for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale. In the rest of the world: Permission to photocopy must be obtained from the copyright owner. Please apply to now Publishers Inc., PO Box 1024, Hanover, MA 02339, USA; Tel. +1 781 871 0245; [www.nowpublishers.com](http://www.nowpublishers.com); [sales@nowpublishers.com](mailto:sales@nowpublishers.com)

now Publishers Inc. has an exclusive license to publish this material worldwide. Permission to use this content must be obtained from the copyright license holder. Please apply to now Publishers, PO Box 179, 2600 AD Delft, The Netherlands, [www.nowpublishers.com](http://www.nowpublishers.com); e-mail: [sales@nowpublishers.com](mailto:sales@nowpublishers.com)

# Foundations and Trends<sup>®</sup> in Optimization

## Volume 6, Issue 1, 2024

### Editorial Board

#### Editors-in-Chief

**Garud Iyengar**  
Columbia University

#### Editors

Dimitris Bertsimas  
*Massachusetts Institute of Technology*

John R. Birge  
*The University of Chicago*

Robert E. Bixby  
*Rice University*

Emmanuel Candes  
*Stanford University*

David Donoho  
*Stanford University*

Laurent El Ghaoui  
*University of California, Berkeley*

Donald Goldfarb  
*Columbia University*

Michael I. Jordan  
*University of California, Berkeley*

Zhi-Quan (Tom) Luo  
*University of Minnesota, Twin Cities*

George L. Nemhauser  
*Georgia Institute of Technology*

Arkadi Nemirovski  
*Georgia Institute of Technology*

Yurii Nesterov  
*HSE University*

Jorge Nocedal  
*Northwestern University*

Pablo A. Parrilo  
*Massachusetts Institute of Technology*

Boris T. Polyak  
*Institute for Control Science, Moscow*

Tamás Terlaky  
*Lehigh University*

Michael J. Todd  
*Cornell University*

Kim-Chuan Toh  
*National University of Singapore*

John N. Tsitsiklis  
*Massachusetts Institute of Technology*

Lieven Vandenbergh  
*University of California, Los Angeles*

Robert J. Vanderbei  
*Princeton University*

Stephen J. Wright  
*University of Wisconsin*

## Editorial Scope

Foundations and Trends® in Optimization publishes survey and tutorial articles in the following topics:

- algorithm design, analysis, and implementation (especially, on modern computing platforms)
- models and modeling systems, new optimization formulations for practical problems
- applications of optimization in machine learning, statistics, and data analysis, signal and image processing, computational economics and finance, engineering design, scheduling and resource allocation, and other areas

### Information for Librarians

Foundations and Trends® in Optimization, 2024, Volume 6, 4 issues. ISSN paper version 2167-3888. ISSN online version 2167-3918. Also available as a combined paper and online subscription.

# Contents

---

|           |                                                   |           |
|-----------|---------------------------------------------------|-----------|
| <b>1</b>  | <b>Introduction</b>                               | <b>2</b>  |
| <b>2</b>  | <b>Background on Convexity</b>                    | <b>6</b>  |
| <b>3</b>  | <b>Basic Properties of Semidifferentials</b>      | <b>11</b> |
| <b>4</b>  | <b>Tangent Vectors and Adjacent Cones</b>         | <b>25</b> |
| <b>5</b>  | <b>First-Order Optimality Conditions</b>          | <b>35</b> |
| <b>6</b>  | <b>A KKT Rule for Semidifferentiable Programs</b> | <b>40</b> |
| <b>7</b>  | <b>Second-Order Optimality Conditions</b>         | <b>43</b> |
| <b>8</b>  | <b>Differentials in Optimization Practice</b>     | <b>48</b> |
| <b>9</b>  | <b>Optimization on Embedded Submanifolds</b>      | <b>53</b> |
| <b>10</b> | <b>Discussion</b>                                 | <b>57</b> |
|           | <b>Acknowledgements</b>                           | <b>59</b> |
|           | <b>References</b>                                 | <b>60</b> |

# A Tutorial on Hadamard Semidifferentials

Kenneth Lange

*Departments of Computational Medicine, Human Genetics, and Statistics, University of California, Los Angeles, USA; [klange@ucla.edu](mailto:klange@ucla.edu)*

---

## ABSTRACT

The Hadamard semidifferential is more general than the Fréchet differential now dominant in undergraduate mathematics education. By slightly changing the definition of the forward directional derivative, the Hadamard semidifferential rescues the chain rule, enforces continuity, and permits differentiation across maxima and minima. It also plays well with convex analysis and naturally extends differentiation to smooth embedded submanifolds, topological vector spaces, and metric spaces of shapes and geometries. The current elementary exposition focuses on the more familiar territory of analysis in Euclidean spaces and applies the semidifferential to some representative problems in optimization and statistics. These include algorithms for proximal gradient descent, steepest descent in matrix completion, and variance components models.

---

**MSC Codes:** Primary 28A15, 65C60.

---

This research is supported in part by USPHS grants GM53275 and HG006139.

---

Kenneth Lange (2024), “A Tutorial on Hadamard Semidifferentials”, Foundations and Trends® in Optimization: Vol. 6, No. 1, pp 1–62. DOI: 10.1561/24000000041.

©2024 K. Lange

# 1

---

## Introduction

---

In line with earlier authors [19], [31], [34], [38], Hadamard in 1923 [23] gave a geometric definition of the differential of a function  $f: \mathbb{R}^p \rightarrow \mathbb{R}^q$ . The geometric definition depends on tangent curves  $v: \mathbb{R} \rightarrow \mathbb{R}^n$  through  $\mathbf{x}$  and defines  $df(\mathbf{x})$  as the  $q \times p$  Jacobi matrix satisfying

$$df(\mathbf{x})v'(0) = (f \circ v)'(0).$$

Here  $v(0) = \mathbf{x}$ , and the tangent vector  $v'(0)$  must exist. In 1937, Fréchet [21] extended this geometric definition from Euclidian spaces to function spaces and emphasized that the geometric definition is more general than his earlier 1925 analytic definition [20] of the differential. In normed vector spaces the analytic definition reads

$$\lim_{v \rightarrow 0} \frac{f(\mathbf{x} + \mathbf{v}) - f(\mathbf{x}) - df(\mathbf{x})\mathbf{v}}{\|\mathbf{v}\|} = \mathbf{0}.$$

This is the standard textbook-definition widely taught today. Hadamard always insisted that the directional derivative be linear with respect to the direction. But in his 1937 paper, Fréchet relaxed the linearity requirement and gave an example of a class of non-differentiable functions for which all the rules of the differential calculus hold, including

the chain rule. Yet, he was unable to handle the Euclidean norm at the origin and continuous but non-differentiable convex functions. This further step involves replacing tangent curves by half-tangent curves and ultimately half-tangent curves by tangent sequences. These successive generalizations emphasize forward directional derivatives rather than two-sided directional derivatives.

The Hadamard semidifferentiable functions as defined in Section 3 constitute the largest known class of nondifferentiable functions that retain all of the features of classical differential calculus, including the chain rule and automatic continuity. Norms, continuous convex functions, and semiconvex functions are all Hadamard semidifferentiable. The Hadamard semidifferential readily extends to functions defined on embedded submanifolds, topological vector spaces, and even metric spaces of shapes and geometries [13], [16], [30].

Given these advantages and the simplicity of the underlying theory, it seems that the time is ripe for the adoption of semidifferentials in advanced undergraduate and beginning graduate courses in the mathematical sciences. The current tutorial provides a brief survey of semidifferentials in the familiar context of finite-dimensional Euclidean space. This restriction exposes the most critical ideas, important connections to convexity and optimization, and a few novel applications. The text [12] delves more deeply into these topics and is highly recommended for a systematic course and self study.

Our table of contents provides a roadmap to the remainder of this tutorial. Section 2 briefly reviews relevant topics from convexity, Euclidean distance functions, and projection operators. After our presentation of semidifferentials in Section 3, Section 4 takes up the intertwined concepts of tangent vectors and tangent cones. We stress adjacent cones rather than contingent tangent cones and Clarke tangent cones. The notion of tangency underlying adjacent cones has a direct connection to semidifferentiability through Euclidean distance functions. Furthermore, adjacent cones enjoy a wide range of useful properties that are relatively easy to prove. Fortunately, adjacent and contingent tangent cones coincide for points in convex sets and embedded submanifolds.

Sections 5–7 demonstrate the strong ties between local optimality and semidifferentiability. The usual necessary conditions become

sufficient in the presence of convexity. Local optimality then becomes global as well. Local optimality is encoded algebraically by the KKT multiplier conditions, which extend to a subclass of semidifferentiable functions [25].

Section 8 showcases connections to convex optimization and computational statistics. Our statistical examples illustrate the value of directional derivatives in extracting first and second differentials. Optimization on manifolds has risen to prominence in the past few decades. Section 9 takes up this subtle subject through the lense of classical mathematical analysis. Fortunately, the most important manifolds in practice are embedded submanifolds situated firmly in Euclidean spaces. We contend that at least some of the vocabulary and apparatus of differential geometry can be ignored if optimization on manifolds is approached through classical analysis. Whether this is a good thing or not depends on one's background. Differential geometry requires intense effort to master. The route through classical analysis potentially allows more individuals to enter the field of constrained optimization at the risk of losing the important geometric insights and analytic tools afforded by differential geometry.

Section 10 concludes our treatment of Hadamard semidifferentials. We briefly mention there numerical methods and applications to infinite dimensional spaces. This tutorial is a snapshot in time of an incredibly rich subject. The efforts to understand and extend the range of the differential calculus have gone on for centuries and will doubtless continue for decades, if not centuries, more. For workers in the mathematical sciences, it is crucial to reach the right balance between generality and ease of use. In our view, Hadamard semidifferentials achieve that balance.

For the record, here are some notation conventions used in the sequel. All vectors and matrices appear in boldface. The entries of the vector  $\mathbf{0}$  consist of 0's, and the vector  $\mathbf{e}_i$  has all entries 0 except a 1 in entry  $i$ . The  $^\top$  superscript indicates a vector or matrix transpose. The Euclidean norm of a vector  $\mathbf{x}$  is denoted by  $\|\mathbf{x}\|$  and the spectral and

Frobenius norms of a matrix  $\mathbf{M} = (m_{ij})$  by

$$\|\mathbf{M}\| = \sup_{\mathbf{x} \neq \mathbf{0}} \frac{\|\mathbf{M}\mathbf{x}\|}{\|\mathbf{x}\|} \quad \text{and} \quad \|\mathbf{M}\|_F = \sqrt{\sum_i \sum_j m_{ij}^2},$$

respectively. All positive semidefinite matrices are symmetric by definition. For a smooth real-valued function  $f(\mathbf{x})$ , we write its gradient (column vector of partial derivatives) as  $\nabla f(\mathbf{x})$ , its first differential (row vector of partial derivatives) as  $df(\mathbf{x}) = \nabla f(\mathbf{x})^\top$ , and its second differential (Hessian matrix) as  $d^2f(\mathbf{x})$ . If  $g(\mathbf{x})$  is vector-valued with  $i$ th component  $g_i(\mathbf{x})$ , then the differential (Jacobi matrix)  $dg(\mathbf{x})$  has  $i$ th row  $dg_i(\mathbf{x})$ ; for a scalar-valued function,  $d^2f(\mathbf{x}) = d\nabla f(\mathbf{x})$ . The transpose  $dg(\mathbf{x})^\top$  is termed the gradient of  $g(\mathbf{x})$ .

# 2

---

## Background on Convexity

---

We assume that the reader is familiar with the basic notions of convexity. Here the emphasis is on a few results pertinent to our subsequent theory. Projection onto a closed convex set  $S \subset \mathbb{R}^d$  is key. The Euclidean distance function  $\text{dist}(\mathbf{y}, S) = \inf_{\mathbf{x} \in S} \|\mathbf{y} - \mathbf{x}\|$  achieve its infimum at a unique point  $P_S(\mathbf{y})$  called the projection of  $\mathbf{y}$  onto  $S$ . Our first proposition collects some relevant facts about distance functions and their corresponding projections.

**Proposition 2.1.** If  $S$  is a nonempty closed set, then the following properties hold:

- (a) For each  $\mathbf{y}$  the minimum distance  $\text{dist}(\mathbf{y}, S)$  is attained by one or more points  $\mathbf{x} \in S$ .
- (b) The Lipschitz bound

$$|\text{dist}(\mathbf{y}, S) - \text{dist}(\mathbf{z}, S)| \leq \|\mathbf{y} - \mathbf{z}\|$$

holds for all  $\mathbf{y}$  and  $\mathbf{z}$  in  $\mathbb{R}^n$ ; consequently,  $\text{dist}(\mathbf{y}, S)$  is continuous.

- (c) The distance function  $\text{dist}(\mathbf{y}, S)$  is convex if and only if the set  $S$  is convex.

- (d) The projection set  $P_S(\mathbf{y}) = \{\mathbf{x} \in S: \|\mathbf{y} - \mathbf{x}\| = \text{dist}(\mathbf{y}, S)\}$  is compact.
- (e) The projection set  $P_S(\mathbf{y})$  reduces to a single point for all  $\mathbf{y}$  if and only if  $S$  is convex.
- (f) The projection  $P_S(\mathbf{y})$  onto a convex set  $S$  is uniquely characterized by the obtuse angle criterion

$$(\mathbf{y} - \mathbf{x})^\top (\mathbf{z} - \mathbf{x}) \leq 0$$

for  $\mathbf{x} = P_S(\mathbf{y})$  and all  $\mathbf{z} \in S$ .

- (g) If  $\mathbf{a} + S$  is an affine subspace, then  $[\mathbf{y} - P_S(\mathbf{y})]^\top \mathbf{v} = 0$  for all  $\mathbf{v} \in S$ .

**Proof:** Choose any point  $\mathbf{z} \in S$ . Because we can confine  $\mathbf{x}$  to the compact sublevel set

$$S \cap \{\mathbf{x}: \|\mathbf{x} - \mathbf{y}\| \leq \|\mathbf{z} - \mathbf{y}\|\},$$

part (a) follows from the fact that a continuous function achieves its minimum on a compact set. For part (b) take the infimum over  $\mathbf{x} \in S$  of both sides of the triangle inequality

$$\|\mathbf{y} - \mathbf{x}\| \leq \|\mathbf{y} - \mathbf{z}\| + \|\mathbf{z} - \mathbf{x}\|.$$

This demonstrates that  $\text{dist}(\mathbf{y}, S) \leq \text{dist}(\mathbf{z}, S) + \|\mathbf{y} - \mathbf{z}\|$ . Reversing the roles of  $\mathbf{y}$  and  $\mathbf{z}$  then leads to the Lipschitz inequality

$$|\text{dist}(\mathbf{y}, S) - \text{dist}(\mathbf{z}, S)| \leq \|\mathbf{y} - \mathbf{z}\|.$$

For part (c) suppose  $S$  is convex and take  $\mathbf{y}$  and  $\mathbf{z}$  in  $C$ . Then for  $\alpha \in [0, 1]$

$$\begin{aligned} & \alpha \text{dist}(\mathbf{y}, S) + (1 - \alpha) \text{dist}(\mathbf{z}, S) \\ &= \alpha \|\mathbf{y} - P_S(\mathbf{y})\| + (1 - \alpha) \|\mathbf{z} - P_S(\mathbf{z})\| \\ &\geq \|\alpha \mathbf{y} + (1 - \alpha) \mathbf{z} - \alpha P_S(\mathbf{y}) - (1 - \alpha) P_S(\mathbf{z})\| \\ &\geq \text{dist}[\alpha \mathbf{y} + (1 - \alpha) \mathbf{z}, S]. \end{aligned}$$

Conversely, if the function  $\text{dist}(\mathbf{y}, S)$  is convex, then the sublevel set

$$S = \{\mathbf{y}: \text{dist}(\mathbf{y}, S) \leq 0\}$$

is convex. For part (d) it suffices to prove that  $P_S(\mathbf{x})$  is closed and bounded. Suppose  $\mathbf{x}_m \in P_S(\mathbf{y})$  converges to  $\mathbf{x}$ . Since  $S$  is closed,  $\mathbf{x} \in S$ . Taking limits in the equality  $\|\mathbf{y} - \mathbf{x}_m\| = \text{dist}(\mathbf{y}, S)$  demonstrates the required equality  $\|\mathbf{y} - \mathbf{x}\| = \text{dist}(\mathbf{y}, S)$ . The boundedness of  $P_S(\mathbf{y})$  is obvious from the coerciveness condition  $\lim_{\|\mathbf{x}\| \rightarrow \infty} \|\mathbf{y} - \mathbf{x}\| = \infty$ . The first claim of part (e) succumbs to Proposition 5.2. The converse of this claim is more subtle, and its proof is omitted. Part (f) is taken up in Example 5.5. Finally for part (g),  $P_S(\mathbf{y}) + \mathbf{v} \in \mathbf{a} + S$  for every  $\mathbf{v} \in S$ . It follows that

$$[\mathbf{y} - P_S(\mathbf{y})]^\top [P_S(\mathbf{y}) + \mathbf{v} - P_S(\mathbf{y})] = [\mathbf{y} - P_S(\mathbf{y})]^\top \mathbf{v} \leq 0.$$

Replacing  $\mathbf{v}$  by  $-\mathbf{v}$  now entails  $[\mathbf{y} - P_S(\mathbf{y})]^\top \mathbf{v} = 0$ .  $\square$

The separation properties of convex sets are crucial in advancing convex analysis. The most basic result is summarized in the next proposition.

**Proposition 2.2.** Consider a nonempty closed convex set  $C$  of  $\mathbb{R}^d$  and a point  $\mathbf{y}$  outside  $C$ . There exist a unit vector  $\mathbf{v}$  and real number  $c$  such that

$$\mathbf{v}^\top \mathbf{y} > c \geq \mathbf{v}^\top \mathbf{x} \tag{2.1}$$

for all  $\mathbf{x} \in C$ . If  $\mathbf{y}$  is a boundary point of a nonempty convex set  $C$ , then there exists a unit vector  $\mathbf{v}$  such that  $\mathbf{v}^\top \mathbf{y} \geq \mathbf{v}^\top \mathbf{x}$  for all  $\mathbf{x} \in C$ .

**Proof:** In view of the obtuse angle criterion, take  $\mathbf{v} = \mathbf{y} - P_C(\mathbf{y})$ . It then follows that  $c = \mathbf{v}^\top P_C(\mathbf{y}) \geq \mathbf{v}^\top \mathbf{x}$  for all  $\mathbf{x} \in C$ . Furthermore,  $\mathbf{v}^\top \mathbf{y} > \mathbf{v}^\top P_C(\mathbf{y}) = c$  because  $\mathbf{v}^\top \mathbf{v} = \|\mathbf{v}\|^2 > 0$ . One can replace  $\mathbf{v}$  by the unit vector  $\|\mathbf{v}\|^{-1} \mathbf{v}$  without invalidating the separation inequalities (2.1).

If  $\mathbf{y}$  is a boundary point of  $C$ , then there exists a sequence of points  $\mathbf{y}_i$  outside  $C$  that converge to  $\mathbf{x}$ . Let  $\mathbf{v}_i$  be the unit vector defining the hyperplane separating  $\mathbf{y}_i$  from  $C$ . Since the unit sphere is compact, we can assume that the  $\mathbf{v}_i$  converge to a unit vector  $\mathbf{v}$ . Taking limits

in the strict inequality  $\mathbf{v}_i^\top \mathbf{y}_i > \mathbf{v}_i^\top \mathbf{x}$  for  $\mathbf{x} \in C$  then yields the result  $\mathbf{v}^\top \mathbf{y} \geq \mathbf{v}^\top \mathbf{x}$ .  $\square$

The next technical proposition [18] is instrumental in proving a general Lagrange multiplier rule for semidifferentiable functions.

**Proposition 2.3.** For  $n$  finite convex functions  $f_i(\mathbf{x})$ , suppose

$$\max_{1 \leq i \leq n} f_i(\mathbf{x}) \geq 0$$

for all  $\mathbf{x}$  in a convex set  $C$ . Then some convex combination  $\sum_{i=1}^n \lambda_i f_i(\mathbf{x})$  is nonnegative throughout  $C$ .

**Proof:** Define the nonempty set

$$S_{\mathbf{x}} = \{\mathbf{y}: y_i > f_i(\mathbf{x}), 1 \leq i \leq n\}$$

for each  $\mathbf{x} \in C$  [18]. It is straightforward to check that  $S = \cup_{\mathbf{x} \in C} S_{\mathbf{x}}$  is convex. By hypothesis  $S$  excludes the point  $\mathbf{y} = \mathbf{0}$ . According to Proposition 2.2, there exists a unit vector  $\boldsymbol{\lambda}$  such that  $\boldsymbol{\lambda}^\top \mathbf{y} \geq \boldsymbol{\lambda}^\top \mathbf{0} = 0$  for all  $\mathbf{y} \in S$ . If any component  $\lambda_i < 0$ , then choose  $\mathbf{y} \in S$ , fix all  $y_j$  with  $j \neq i$ , and send  $y_i$  to  $\infty$ . These actions contradict the inequality  $\boldsymbol{\lambda}^\top \mathbf{y} \geq 0$ . The unit vector  $\boldsymbol{\lambda}$  can clearly be normalized so that its components sum to 1.  $\square$

In the subsequent sections we will discuss the applications of projection operators. It turns out that projections play a crucial role in constrained optimization over well-behaved sets. We now state a few projection operators without proof. For instance, let  $\mathbf{A}$  be a matrix with full row rank, and let  $S = \{\mathbf{x} \in \mathbb{R}^d: \mathbf{A}\mathbf{x} = \mathbf{b}\}$  be an affine subspace. The closest point to  $\mathbf{y}$  in  $S$  is

$$P_S(\mathbf{y}) = \mathbf{y} - \mathbf{A}^\top (\mathbf{A}\mathbf{A}^\top)^{-1} (\mathbf{A}\mathbf{y} - \mathbf{b}).$$

The matrix  $\mathbf{P} = \mathbf{I} - \mathbf{A}^\top (\mathbf{A}\mathbf{A}^\top)^{-1} \mathbf{A}$  is an orthogonal projection satisfying  $\mathbf{P}^2 = \mathbf{P}$ ,  $\mathbf{P}^\top = \mathbf{P}$ , and  $\mathbf{P}\mathbf{x} = \mathbf{x}$  for  $\mathbf{x} \in S$ . Projection onto a rectangle  $[\mathbf{a}, \mathbf{b}]$  simply clamps each coordinate  $y_i$  of  $\mathbf{y}$  to  $[a_i, b_i]$ . Projection onto the sphere  $\{\mathbf{x}: \|\mathbf{x}\| \leq c\}$  maps  $\mathbf{y}$  into itself when  $\|\mathbf{y}\| \leq c$  and into  $\frac{c}{\|\mathbf{y}\|} \mathbf{y}$  when  $\|\mathbf{y}\| > c$ . Other convex sets and their projections are explored in the references [3], [5], [25].

For matrix sets under the Frobenius inner product and norm, the singular value decomposition (svd) is key. If  $\mathbf{A}$  is a square matrix with

full svd  $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ , then the nearest orthogonal matrix to  $\mathbf{A}$  is  $\mathbf{U}\mathbf{V}^\top$ . This is also the closest rotation matrix to  $\mathbf{A}$  if  $\det(\mathbf{U}\mathbf{V}^\top) = 1$ . If  $\det(\mathbf{U}\mathbf{V}^\top) = -1$ , the closest rotation matrix can be constructed by replacing the last column  $\mathbf{u}$  of  $\mathbf{U}$  by  $-\mathbf{u}$ . For  $p < d$  the Stiefel manifold  $\text{St}(p)$  consists of  $d \times p$  matrices  $\mathbf{X}$  satisfying  $\mathbf{X}^\top \mathbf{X} = \mathbf{I}_p$ . The closest matrix  $\mathbf{X}$  in the Stiefel manifold to a given  $d \times p$  matrix  $\mathbf{A}$  with thin svd  $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$  is  $\mathbf{U}\mathbf{V}^\top$ . The Eckart–Young theorem says that the best rank  $r$  approximation of the matrix  $\mathbf{A}$  with full svd  $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$  is  $\sum_{j=1}^r \sigma_j \mathbf{u}_j \mathbf{v}_j^\top$ . Here the  $\sigma_j$  are singular values, and the  $\mathbf{u}_j$  and  $\mathbf{v}_j$  are singular vectors. The Grassmann manifold  $\text{G}(d, r)$  can be viewed as the set of  $d \times d$  orthogonal projection matrices of rank  $r$ . If the symmetric matrix  $\frac{1}{2}(\mathbf{A} + \mathbf{A}^\top)$  has ordered spectral decomposition  $\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$ , then the closest matrix to  $\mathbf{A}$  in  $\text{G}(d, r)$  is  $\sum_{j=1}^r \mathbf{u}_j \mathbf{u}_j^\top$ .

# 3

---

## Basic Properties of Semidifferentials

---

A function  $f(\mathbf{y})$  mapping an open set  $U$  of  $\mathbb{R}^p$  into  $\mathbb{R}^q$  is said to be Hadamard semidifferentiable at  $\mathbf{x} \in U$  in the direction  $\mathbf{v}$  if the limit

$$\lim_{t \downarrow 0, \mathbf{w} \rightarrow \mathbf{v}} \frac{f(\mathbf{x} + t\mathbf{w}) - f(\mathbf{x})}{t} = d_{\mathbf{v}}f(\mathbf{x}) \quad (3.1)$$

exists [12], [32]. If this limit exists for all directions  $\mathbf{v}$ , then  $f(\mathbf{x})$  is said to be Hadamard semidifferentiable at  $\mathbf{x}$ . Taking  $\mathbf{w}$  identically equal to  $\mathbf{v}$  shows that  $d_{\mathbf{v}}f(\mathbf{x})$  coincides with the forward directional derivative of  $f(\mathbf{y})$  at  $\mathbf{x}$  in the direction  $\mathbf{v}$ . Gateaux semidifferentiability at  $\mathbf{x}$  postulates the existence of all possible forward directional derivatives. Hadamard's definition is more restrictive and yields a richer theory. For the sake of brevity, we will omit the prefix Hadamard in discussing semidifferentials. It is also convenient to restate the definition in terms of sequences. Thus, semidifferentiability requires the limit

$$\lim_{n \rightarrow \infty} \frac{f(\mathbf{x} + t_n \mathbf{w}_n) - f(\mathbf{x})}{t_n} = d_{\mathbf{v}}f(\mathbf{x})$$

to exist and to be independent of the sequences  $t_n \downarrow 0$  and  $\mathbf{w}_n \rightarrow \mathbf{v}$ .

Implicit in the definition (3.1) is the fact that  $f(\mathbf{y})$  is semidifferentiable at  $\mathbf{x}$  if and only if each of its components is semidifferentiable at  $\mathbf{x}$ . The left and right  $i$ th partial semiderivative of  $f(\mathbf{x})$  is recovered

by taking  $\mathbf{v} = \pm \mathbf{e}_i$ . If  $f(\mathbf{x}) = \mathbf{M}\mathbf{x}$ , then  $d_{\mathbf{v}}f(\mathbf{x}) = \mathbf{M}\mathbf{v}$ . Finally, the positive homogeneity rule  $d_{c\mathbf{v}}f(\mathbf{x}) = c d_{\mathbf{v}}f(\mathbf{x})$  for  $c > 0$  follows from the representation

$$\frac{f(\mathbf{x} + t\mathbf{c}\mathbf{w}) - f(\mathbf{x})}{t} = c \frac{f(\mathbf{x} + t\mathbf{c}\mathbf{w}) - f(\mathbf{x})}{ct}.$$

More importantly, virtually all of the standard rules of differentiation remain intact in a broader setting than the class of differentiable functions. In addition, new rules for differentiating minima and maxima make their appearance, and many connections with convexity can be profitably pursued.

The function

$$f(\mathbf{x}) = \begin{cases} \frac{x_1^6}{(x_2 - x_1^2)^2 + x_1^8} & \mathbf{x} \neq \mathbf{0} \\ 0 & \mathbf{x} = \mathbf{0} \end{cases}$$

illustrates the difference between Hadamard semidifferentiability at a point and the mere existence of all forward directional derivatives at the point. One can prove that the forward directional derivative  $d_{\mathbf{v}}f(\mathbf{0}) = 0$  for all  $\mathbf{v}$ , but  $f(\mathbf{x})$  is discontinuous at  $\mathbf{0}$ , and the Hadamard semidifferential  $d_{(1,0)}f(\mathbf{0})$  does not exist. The last assertion follows from contrasting the convergence of the relevant difference quotient for the scalar sequence  $t_n = \frac{1}{n} \rightarrow 0$  and the two vector sequences  $\mathbf{w}_n = (1, 0)^\top$  and  $\mathbf{w}_n = (1, \frac{1}{n})^\top \rightarrow (1, 0)^\top$ .

The relation between differentiability and semidifferentiability is spelled out in our next proposition.

**Proposition 3.1.** A function  $f(\mathbf{y})$  differentiable at  $\mathbf{x}$  is also semidifferentiable at  $\mathbf{x}$ , and the identity  $d_{\mathbf{v}}f(\mathbf{x}) = df(\mathbf{x})\mathbf{v}$  holds, where  $df(\mathbf{x})$  is the Jacobi matrix of partial derivatives. If  $f(\mathbf{y})$  is semidifferentiable at  $\mathbf{x}$ , and the map  $\mathbf{v} \mapsto d_{\mathbf{v}}f(\mathbf{x})$  is linear, then  $f(\mathbf{y})$  is differentiable at  $\mathbf{x}$ .

**Proof:** If  $f(\mathbf{y})$  is differentiable at  $\mathbf{x}$ , then

$$f(\mathbf{y}) - f(\mathbf{x}) = df(\mathbf{x})(\mathbf{y} - \mathbf{x}) + o(\|\mathbf{y} - \mathbf{x}\|)$$

as  $\mathbf{y}$  approaches  $\mathbf{x}$ . Choosing  $\mathbf{y} = \mathbf{x} + t\mathbf{w}$  for  $t > 0$  and  $\|\mathbf{w} - \mathbf{v}\|$  small shows that the difference quotient approaches  $d_{\mathbf{v}}f(\mathbf{x}) = df(\mathbf{x})\mathbf{v}$ .

To prove the partial converse, define

$$g(\mathbf{u}) = \frac{f(\mathbf{x} + \mathbf{u}) - f(\mathbf{x}) - d_{\mathbf{u}}f(\mathbf{x})}{\|\mathbf{u}\|}.$$

It suffices to prove that  $c = \limsup_{\|\mathbf{u}\| \rightarrow 0} \|g(\mathbf{u})\| = 0$ . Choose a sequence  $\mathbf{u}_n \neq \mathbf{0}$  such that  $t_n = \|\mathbf{u}_n\|$  converges to 0 and  $\|g(\mathbf{u}_n)\|$  converges to  $c$ . Because the unit sphere is compact, some subsequence of the sequence  $\mathbf{w}_n = \|\mathbf{u}_n\|^{-1}\mathbf{u}_n$  converges to a unit vector  $\mathbf{v}$ . Without loss of generality, take the subsequence to be the original sequence. Linearity then implies that

$$\begin{aligned} \lim_{n \rightarrow \infty} g(\mathbf{u}_n) &= \lim_{n \rightarrow \infty} \frac{f(\mathbf{x} + t_n \mathbf{w}_n) - f(\mathbf{x}) - d_{t_n \mathbf{w}_n}f(\mathbf{x})}{t_n} \\ &= \lim_{n \rightarrow \infty} \left[ \frac{f(\mathbf{x} + t_n \mathbf{w}_n) - f(\mathbf{x})}{t_n} - df(\mathbf{x})\mathbf{w}_n \right] \\ &= d_{\mathbf{v}}f(\mathbf{x}) - d_{\mathbf{v}}f(\mathbf{x}) \end{aligned}$$

and that  $c = 0$ . □

Our next proposition highlights an important implication of Hadamard's definition of semidifferentiability.

**Proposition 3.2.** A function  $f(\mathbf{y})$  semidifferentiable at a point  $\mathbf{x}$  is continuous at  $\mathbf{x}$ . Furthermore, the map  $\mathbf{v} \mapsto d_{\mathbf{v}}f(\mathbf{x})$  is continuous.

**Proof:** Suppose  $\mathbf{x}_n$  tends to  $\mathbf{x}$ , but  $f(\mathbf{x}_n)$  does not tend to  $f(\mathbf{x})$ . Then there exists an  $\epsilon > 0$  such that  $\|f(\mathbf{x}_n) - f(\mathbf{x})\| \geq \epsilon$  for infinitely many  $n$ . Without loss of generality, we may assume that the entire sequence possesses this property. Now write

$$\mathbf{x}_n = \mathbf{x} + \mathbf{x}_n - \mathbf{x} = \mathbf{x} + t_n \frac{\mathbf{x}_n - \mathbf{x}}{\|\mathbf{x}_n - \mathbf{x}\|}$$

by taking  $t_n = \|\mathbf{x}_n - \mathbf{x}\|$ . Again some subsequence of the sequence

$$\mathbf{w}_n = \frac{\mathbf{x}_n - \mathbf{x}}{\|\mathbf{x}_n - \mathbf{x}\|}$$

of unit vectors converges to a unit vector  $\mathbf{v}$ . Passing to the subsequence where this occurs if necessary, we have

$$f(\mathbf{x} + t_n \mathbf{w}_n) - f(\mathbf{x}) = t_n d_{\mathbf{v}}f(\mathbf{x}) + o(t_n), \quad (3.2)$$

contrary to the assumption that  $\|f(\mathbf{x}_n) - f(\mathbf{x})\| \geq \epsilon$  for all large  $n$ .

To prove the second assertion, suppose that  $\mathbf{v}_n$  tends to  $\mathbf{v}$ . For each  $n$  choose  $t_n$  and  $\mathbf{w}_n$  so that  $0 < t_n \leq \frac{1}{n}$ ,  $\|\mathbf{w}_n - \mathbf{v}_n\| \leq \frac{1}{n}$ , and

$$\left\| \frac{f(\mathbf{x} + t_n \mathbf{w}_n) - f(\mathbf{x})}{t_n} - d_{\mathbf{v}_n} f(\mathbf{x}) \right\| \leq \frac{1}{n}.$$

Taking limits in the identity

$$\begin{aligned} & \frac{f(\mathbf{x} + t_n \mathbf{w}_n) - f(\mathbf{x})}{t_n} - d_{\mathbf{v}} f(\mathbf{x}) \\ &= \frac{f(\mathbf{x} + t_n \mathbf{w}_n) - f(\mathbf{x})}{t_n} - d_{\mathbf{v}_n} f(\mathbf{x}) + d_{\mathbf{v}_n} f(\mathbf{x}) - d_{\mathbf{v}} f(\mathbf{x}) \end{aligned}$$

now implies that  $d_{\mathbf{v}_n} f(\mathbf{x})$  tends to  $d_{\mathbf{v}} f(\mathbf{x})$ .  $\square$

The approximate equality (3.2) has an important consequence in minimization of real-valued functions. Suppose we can find a direction  $\mathbf{v}$  with  $d_{\mathbf{v}} f(\mathbf{x}) < 0$ . Then  $\mathbf{v}$  is a descent direction from  $\mathbf{x}$  in the sense that  $f(\mathbf{x} + t\mathbf{v}) < f(\mathbf{x})$  for all sufficiently small  $t > 0$ . Thus, back-tracking is bound to produce a decrease in  $f(\mathbf{y})$ .

Here is a simple test for establishing semidifferentiability.

**Proposition 3.3.** Suppose  $f(\mathbf{y})$  is locally Lipschitz around  $\mathbf{x}$  and possesses all possible forward directional derivatives there. Then  $f(\mathbf{y})$  is semidifferentiable at  $\mathbf{x}$ .

**Proof:** Consider the expansion

$$\frac{f(\mathbf{x} + t\mathbf{w}) - f(\mathbf{x})}{t} = \frac{f(\mathbf{x} + t\mathbf{w}) - f(\mathbf{x} + t\mathbf{v})}{t} + \frac{f(\mathbf{x} + t\mathbf{v}) - f(\mathbf{x})}{t}.$$

As  $t \downarrow 0$ , the second fraction on the right of this equation tends to the forward directional derivative of  $f(\mathbf{y})$  at  $\mathbf{x}$  in the direction  $\mathbf{v}$ . If  $c$  is a Lipschitz constant for  $f(\mathbf{y})$  in a neighborhood of  $\mathbf{x}$ , then the first fraction on the right of the equation is locally bounded in norm by  $c\|\mathbf{w} - \mathbf{v}\|$ , which can be made arbitrarily small by taking  $\mathbf{w}$  close to  $\mathbf{v}$ .  $\square$

**Example 3.1.** Euclidean norm excluding the origin.

Suppose  $\mathbf{x} \neq \mathbf{0}$ . Taking limits in the identity

$$\frac{\|\mathbf{x} + t\mathbf{w}\| - \|\mathbf{x}\|}{t} = \frac{\|\mathbf{x} + t\mathbf{w}\|^2 - \|\mathbf{x}\|^2}{t(\|\mathbf{x} + t\mathbf{w}\| + \|\mathbf{x}\|)} = \frac{2\mathbf{x}^\top \mathbf{w} + t\|\mathbf{w}\|^2}{\|\mathbf{x} + t\mathbf{w}\| + \|\mathbf{x}\|}$$