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

the essence of knowledge
Boston — Delft

Foundations and Trends® 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
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® in Optimization, vol. 6, no. 1, pp. 1–62, 2024.

ISBN: 978-1-63828-349-2

Foundations and Trends® in Optimization
Volume 6, Issue 1, 2024

Editorial Board

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

A Tutorial on Hadamard Semidifferentials

Kenneth Lange

Departments of Computational Medicine, Human Genetics, and Statistics, University of California, Los Angeles, USA; 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: 10.1561/2400000041.

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 $f\colon \mathbb{R}^p \to \mathbb{R}^q$. The geometric definition depends on tangent curves $v\colon \mathbb{R} \to \mathbb{R}^n$ through $\boldsymbol{x}$ and defines $df(\boldsymbol{x})$ as the $q \times p$ Jacobi matrix satisfying

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

Here $v(0) = \boldsymbol{x}$, 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

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

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 $\boldsymbol{e}_i$ 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 $\boldsymbol{x}$ is denoted by $\|\boldsymbol{x}\|$ and the spectral and

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

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

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

References

[1] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008.

[2] J.-P. Aubin and H. Frankowska, *Set-Valued Analysis*. Springer, 2009.

[3] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2nd ed. Springer, 2017.

[4] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, "Automatic differentiation in machine learning: A survey," *Journal of Machine Learning Research*, vol. 18, 2018, pp. 1–43.

[5] A. Beck, *First-Order Methods in Optimization*. SIAM, 2017.

[6] D. P. Bertsekas, *Nonlinear Programming*. Belmont MA: Athena Scientific, 1999.

[7] E. Beutner and H. Zähle, "Deriving the asymptotic distribution of U-and V-statistics of dependent data using weighted empirical processes," *Bernoulli*, vol. 18, 2012, pp. 803–822.

[8] J. M. Borwein and A. S. Lewis, *Convex Analysis and Nonlinear Optimization: Theory and Examples*, 2nd ed. Springer, 2010.

[9] N. Boumal, *An Introduction to Optimization on Smooth Manifolds*. Cambridge University Press, 2023.

[10] F. Clarke, *Functional Analysis, Calculus of Variations and Optimal Control*. Springer, 2013.

[11] J. M. Danskin, "The theory of max-min, with applications," *SIAM Journal of Applied Mathematics*, vol. 14, 1966, pp. 641–664.

[12] M. C. Delfour, *Introduction to Optimization and Hadamard Semidifferential Calculus*, 2nd ed. Philadelphia: SIAM, 2019.

[13] M. C. Delfour, "Hadamard semidifferential of functions on an unstructured subset of a TVS," *Journal of Pure Applied Functional Analysis*, vol. 5, 2020, pp. 1039–1072.

[14] M. C. Delfour, "Hadamard semidifferential, oriented distance function, and some applications," *Communications on Pure and Applied Analysis*, vol. 21, 2022, pp. 1917–1951.

[15] M. C. Delfour and J.-P. Zolésio, *Shapes and Geometries: Metrics, Analysis, Differential Calculus, and Optimization*. Philadelphia: SIAM, 2011.

[16] A. L. Dontchev and R. T. Rockafellar, *Implicit Functions and Solution Mappings: A View from Variational Analysis*. Springer, 2009.

[17] A. Edelman, T. A. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM Journal on Matrix Analysis and Applications*, vol. 20, 1998, pp. 303–353.

[18] K. Fan, I. Glicksberg, and A. J. Hoffman, "Systems of inequalities involving convex functions," *Proceedings of the American Mathematical Society*, vol. 8, 1957, pp. 617–622.

[19] M. Fréchet, "Sur la notion de différentielle," *Comptes Rendus de l'Académie des Sciences*, vol. 152, 1911, pp. 845–847.

[20] M. Fréchet, "La notion de différentielle dans l'analyse générale," *Annales Scientifiques de l'École Normale Supérieure*, vol. XLII, 1925, pp. 293–323.

[21] M. Fréchet, "Sur la notion de différentielle," *Journal de Mathématiques Pures et Appliquées*, vol. 16, 1937, pp. 233–250.

[22] A. Garon and M. C. Delfour, *Transfinite Interpolations and Eulerian/Lagrangian Dynamics*. Philadelphia: SIAM, 2022.

[23] J. Hadamard, "La notion de différentielle dans l'enseignement," *The Mathematical Gazette*, vol. 19, 1935, pp. 341–342.

[24] Z. Lai, L.-H. Lim, and Y. Ke, "Simpler Grassmannian optimization," *arXiv preprint* arXiv: 2009.13502, 2020.

[25] K. Lange, *MM Optimization Algorithms*. Philadelphia: SIAM, 2016.

[26] D. G. Luenberger and Y. Ye, *Linear and Nonlinear Programming*, 3rd ed. Springer, 2008.

[27] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 3rd ed. Wiley, 2019.

[28] R. D. Neidinger, “Introduction to automatic differentiation and matlab object-oriented programming,” *SIAM Review*, vol. 52, 2010, pp. 545–563.

[29] Y. Nesterov, “Smooth minimization of non-smooth functions,” *Mathematical Programming*, vol. 103, 2005, pp. 127–152.

[30] J. P. Penot, “Calcul sous-différentiel et optimisation,” *Journal of Functional Analysis*, vol. 27, 1978, pp. 248–276.

[31] J. Pierpont, *The Theory of Functions of Real Variables*, vol. I. Boston: Ginn and Company, 1905.

[32] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*. Springer, 2009.

[33] S. Särkkä and L. Svensson, *Bayesian Filtering and Smoothing*. Cambridge University Press, 2023.

[34] O. Stolz, *Grundzüge der Differential und Integralrechnung, I*. Leipzig: BG Teubner, 1893.

[35] J. Tanner and K. Wei, “Normalized iterative hard thresholding for matrix completion,” *SIAM Journal on Scientific Computing*, vol. 35, 2013, S104–S125.

[36] A. W. Van der Vaart, *Asymptotic Statistics*. Cambridge University Press, 2000.

[37] B. Vandereycken and S. Vandewalle, “A riemannian optimization approach for computing low-rank solutions of lyapunov equations,” *SIAM Journal on Matrix Analysis and Applications*, vol. 31, 2010, pp. 2553–2579.

[38] W. H. Young, *The Fundamental Theorems of Differential Calculus*. Cambridge: Cambridge University Press, 1910.