SpringerBriefs in Optimization

SpringerBriefs in Optimization showcases algorithmic and theoretical techniques, case studies, and applications within the broad-based field of optimization. Manuscripts related to the ever-growing applications of optimization in applied mathematics, engineering, medicine, economics, and other applied sciences are encouraged.

More information about this series at http://www.springer.com/series/8918

Adam B. Levy

Attraction in Numerical Minimization

Iteration Mappings, Attractors, and Basins of Attraction

Springer

Adam B. Levy
Bowdoin College
Brunswick
ME, USA

ISSN 2190-8354 ISSN 2191-575X (electronic)
SpringerBriefs in Optimization
ISBN 978-3-030-04048-2 ISBN 978-3-030-04049-9 (eBook)
https://doi.org/10.1007/978-3-030-04049-9

Library of Congress Control Number: 2018963116

Mathematics Subject Classification: 49J53, 90C, 49M, 65K

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To Sarah

Preface

Equilibria are fundamental objects of study in dynamical systems theory because they can be used to classify and understand the behavior of the system. We introduce a parallel concept of "attractors" in numerical minimization and use various related objects from dynamical systems theory as inspiration for analogues in numerical minimization. We then employ our new tools to carefully analyze a variety of particular examples in numerical minimization in order to demonstrate how these tools can enrich our understanding. Our definition of attractor rests on the platform of "iteration mappings" which are a special type of "multiset-mapping" that we also introduce and develop here.

Set-valued mappings associate sets to individual inputs, generalizing the notion of a single-valued mapping. Set-valued mappings arise naturally in many areas, for instance, in optimization where solutions may not be unique and where generalized (and set-valued) notions of continuity and derivatives allow the variational analysis of problems with constraints and/or non-smooth objective functions. The origins of variational analysis can be traced to at least 1925 in the work of Vasilesco [24], and this research area remains very active today. Some important books covering various stages of this subject to date include [4, 6, 12, 18, 21].

Numerical variational analysis is a related area that uses generalized notions of continuity and differentiability to analyze the convergence and stability of numerical optimization methods, and the focus of this analysis has traditionally been on individual "iterate" vectors generated by a method at each step. It has long been recognized (e.g., [25]) that we may conceive of these iterates as the input and/or output of a set-valued iteration mapping because the next iterate may need to be chosen from several promising trial vectors generated by the current iterate. This formulation is important since so many numerical optimization methods use multiple trial vectors to determine an iterate vector at each step (e.g., pattern-search methods, dating back at least to [11], or any method employing a backtracking line search). However, this formulation is limited to methods that use a single iterate vector to generate the next trial vectors, and some numerical optimization methods actually generate and use multiple iterate vectors at each step, including the Nelder-Mead method [19] and trust-region methods using interpolating model functions (e.g., [8]). Therefore, to more thoroughly represent the evolution of optimization methods from one iteration to the next via an iteration mapping, we will extend the notion of set-valued mapping to accommodate set inputs as well as set outputs. Moreover, the vectors generated at each iteration of an optimization method sometimes appear with copies (especially in the limit), so the input and output of an iteration mapping actually might be "multisets." Because of this, we will define and develop multiset-mappings that take multisets to multisets; and we call the special multisets generated by iteration mappings "iterate-multisets."

In order to study the convergence properties of the sequences of iterate-multisets generated by numerical minimization methods, we introduce a generalized continuity property for multiset-mappings called "calmness." A localized version of calmness for set-valued mappings was introduced in [20], and the version we introduce here for multiset-mappings uses pre-distance functions on the domain and range spaces. We characterize calmness for multiset-mappings via a generalized derivative for multiset-mappings that we also introduce here, and we use calmness to provide two versions of fixed-point theorems for iteration mappings that generalize the Banach fixed-point theorem [5, Theorem 6]. A crucial concept for iteration mappings is the "viability" of initial iterate-multisets, which identifies the iterate-multisets from which the repeated application of the iteration mapping produces a sequence of non-empty iterate-multisets. We precisely define iteration mappings for three different well-known minimization methods: coordinate-search, steepest-descent (with three different line-searches), and Nelder-Mead; and we discuss viability in each case.

We define appropriate notions of stability and asymptotic stability for attractors in numerical optimization, where the second notion involves the additional presence of a positive "radius of attraction" (signaling that the attractor attracts every viable initial iterate-multiset whose elements are close to the attractor). We apply our generalized Banach fixed-point theorems to deduce conditions on the iteration mapping that ensure a positive radius of attraction. We also give conditions under which a positive radius of attraction implies that the attractor is a local minimizer, and we provide a companion result involving a weaker notion of "radius of restricted attraction." In addition, we prove that the reverse implication of the companion result holds when the attractor is stable. These results rely on important properties of "local dense viability" (viable initial iterate-multisets are sufficiently abundant near an attractor) and "minvalue-monotonicity" (the minimum value of the objective function does not increase with iteration). All of these notions depend on both the objective function being minimized and the method of numerical minimization. So, for instance, an attractor for one method might not be an attractor for another, even if the objective function remains the same. This is fundamentally different from the objects from dynamical systems we use for inspiration, since those are fixed by the system and do not vary with the method one might use to explore the system.

Another object we adapt from dynamical systems is the "basin of attraction" associated with an attractor in numerical optimization; and we pair this with companion notions of "basin size" and "basin entropy." Both of these notions are computed by simulation, and the latter quantifies the complexity of basin boundaries. We illustrate simulated basins of attraction for four example objective functions using the same four different numerical methods for each (coordinate-search and steepest-descent with three different line-searches). For all 16 combinations, we compute and classify basin size, as well as computing and illustrating basin entropy. We apply the Nelder-Mead method to the four example functions in a separate section because its non-singleton iterate-multisets require different illustrations than the other methods.

Using the tools we have developed here, we investigate the practical significance of two well-known counterexamples to good convergence behavior in numerical minimization: the canoe function with coordinate-search and McKinnon's function [17] with Nelder-Mead. We use our notions of basin size and basin entropy to quantify the extent to which initial data are likely to lead to undesirable consequences. This investigation was stimulated by the curious fact that the Nelder-Mead method is widely used despite its possibility of failure for high-dimensional problems [23], as well as the impossibility of traditional convergence theory to support it (as demonstrated by McKinnon [17]). The authors of [15] suggest some possible reasons for the continued popularity of the Nelder-Mead method, and one minor consequence of our work here is to add to their list by demonstrating that the theory-stifling result of [17] can be viewed as practically insignificant.

Harpswell, ME, USA Adam B. Levy
December 2017

Contents

1 Multisets and Multiset Mappings 1
1.1 Pre-distance Functions 2
1.2 Calmness 4
1.2.1 Calmness for Set-Valued Mappings 6
1.3 Derivative Characterization of Calmness 8

2 Iteration Mappings 11
2.1 Coordinate-Search 12
2.1.1 Iteration Mapping $I_{\mathrm{CS},f}$ 12
2.2 Steepest-Descent 13
2.2.1 Iteration Mapping $I_{\mathrm{SD}_i,f}$ 13
2.3 Nelder–Mead 16
2.3.1 Iteration Mapping $I_{\mathrm{NM},f}$ 18
2.4 Practical Considerations 20
2.5 Convergence Theorems for Iteration Mappings 21

3 Equilibria in Dynamical Systems 23
3.1 Basins of Attraction 24
3.1.1 Basin Size 25
3.1.2 Basin Entropy 27
3.2 Stability 29
3.3 Equilibria as Minimizers 30

4 Attractors 33
4.1 Local Dense Viability 34
4.2 Basins of Attraction 34
4.3 Stability 35
4.4 Minvalue-Monotonicity 37
4.5 Attractors as Minimizers 38

5 Basin Analysis via Simulation 43
5.1 Example 1: Two Global Minimizers 45
5.1.1 Simulated Basins 46
5.1.2 Basin Sizes 48
5.1.3 Basin Entropy 50
5.2 Example 2: One Global Minimizer and a Saddle Point 51
5.2.1 Simulated Basins 52
5.2.2 Basin Sizes 53
5.2.3 Basin Entropy 56
5.3 Example 3: One Global Minimizer and One Local Minimizer 57
5.3.1 Simulated Basins 58
5.3.2 Basin Sizes 58
5.3.3 Basin Entropy 61
5.4 Example 4: One Global Minimizer and a Ring of Local Minimizers 62
5.4.1 Simulated Basins 63
5.4.2 Basin Sizes 64
5.4.3 Basin Entropy 64
5.5 Nelder–Mead Method 66
5.5.1 Simulated Basins 66
5.5.2 Basin Sizes 68
5.5.3 Basin Entropy 69
5.6 Practical Significance of Counterexamples 71
5.6.1 Coordinate-Search and the Canoe Function 71
5.6.2 The Nelder–Mead Method and McKinnon's Function 74

References 77