

REVIEWS AND DESCRIPTIONS OF TABLES AND BOOKS

The numbers in brackets are assigned according to the American Mathematical Society classification scheme accessible from www.ams.org/msc/.

5[65K05, 65K10]—*Introduction to derivative-free optimization*, by Andrew R. Conn, Katya Scheinberg, and Luis N. Vicente, SIAM, Philadelphia, PA, 2009, xii+277 pp., softcover, US\$73.00, ISBN 978-0-898716-68-9

This monograph is a state-of-the-art presentation of algorithms for solving a class of optimization problems that are benign in the sense that their objective functions are reasonably smooth, unconstrained, and defined over a relatively small number of variables (say up to a hundred). However, *the problems are extremely challenging from an algorithmic standpoint*, because gradient vectors and higher derivatives of an objective function cannot be supplied and evaluation of the latter is often expensive and/or corrupted by noise (making accurate finite-difference derivative estimation prohibitive).

Optimization problems of the foregoing type are encountered in many areas of science, engineering, and business. The standard workhorse for solving them used to be the direct-search algorithm of Nelder and Mead [4], which has been popular with practitioners from its inception in 1965. A little over a decade ago, during the mid-nineties, the algorithm and its variants attracted the attention of theoreticians, who shed considerable light on its convergence properties, or lack thereof. This research, along with the increasing need to solve practical applications, helped to spur renewed research into optimization algorithms that do not require derivative information. For obvious reasons, they are called “derivative-free”, and they can be characterized, in broad-brush fashion, as techniques based on “sampling the objective function at several points, sufficiently spaced apart” (p. 57). The algorithms remain *gradient-related*, because their convergence to a local minimizing point, which is considered in detail in this monograph, relies on certain smoothness assumptions. In this way, derivative-free algorithms can also be distinguished from “non-differentiable” optimization algorithms, which compute subgradient (or subdifferential) information and seek a local minimizing point where the objective function may not possess a first derivative.

As stated in its preface, the monograph’s “main aims include giving an interested reader a good idea of the state of the art of derivative free optimization, with a *detailed description of the basic theory* to the extent that the reader can well understand what is needed to ensure convergence, how it affects algorithm design, and what kind of success one can expect and where” (p. xi–xii, italics ours). The authors motivate their subject in a brief introductory chapter, by outlining a few examples of problems that arise in practice and by illustrating the typical behaviour of key algorithms (formulated subsequently) on a pair of standard test problems, namely, Rosenbrock’s function and a perturbed quadratic. The main body of the monograph is organized into two parts titled “Sampling and modeling” and “Frameworks and algorithms”, respectively, each approximately 100 pages in

length. These two parts are followed by a brief third part on a “Review of other topics.” We will consider each part in turn.

Part I brings together the mathematical machinery used to formulate derivative-free algorithms in Part II. The material is nicely organized and well presented. Much of it is of interest in its own right and also a nice application of interpolation techniques that students often encounter for the first time in an introductory course on numerical analysis. The reader is introduced to the theory of positive linear independence, in particular, positive spanning sets and positive bases upon which directional direct-search methods rely to ensure descent; the key notion of “poisedness” of a set of sample points for linear interpolation/regression; and the connection between linear interpolation and simplex gradient estimation. Subsequent chapters consider nonlinear polynomial interpolation (used to form quadratic interpolation models within derivative-free algorithms). In particular, they describe polynomial bases (Lagrange, Newton) and give a comprehensive discussion of the quality of an interpolation set (well-poisedness, Λ -poisedness, condition number as a measure of poisedness, derivation of error bounds, etc.). These topics are considered in three consecutive chapters for different types of interpolation sets: fully determined, overdetermined (polynomial regression), and underdetermined. The concluding chapter of Part I addresses the key topic of modifying an interpolation set to improve its quality, i.e., ensure well-poisedness, and it presents particular algorithms for this purpose.

Part II utilizes the “mathematical toolkit” of Part I to formulate and analyze specific derivative-free optimization algorithms. Two main classes of algorithms are discussed, namely, direct-search methods and trust-region methods, and a brief chapter on implicit-filtering, or line search methods based on simplex gradients, is sandwiched in between. Direct search is considered in two consecutive chapters. The first describes algorithms based on classical coordinate-search and a directional direct-search framework—a typical iteration involves a scatter/mesh search step, a poll step using a positive basis, and an internal parameter update—and considers their global convergence for both continuously differentiable and nonsmooth cases. The second chapter of the pair addresses simplicial direct-search and gives a comprehensive theoretical discussion of the Nelder-Mead algorithm and recent modifications that are designed to guarantee convergence. The discussion of the other main class of methods—trust-region algorithms based on derivative-free linear and quadratic models—is the centerpiece of the monograph. Two key algorithmic frameworks are formulated along with a proof of their global convergence to first-order and second-order critical points. They provide the basis for several practical algorithms—the “DFO” approach, Powell’s methods, and wedge methods—that are presented in the concluding chapter.

Finally, Part III is a brief review of other topics, including the use of surrogate models and extensions of algorithms discussed in Part II when constraints are present. A useful list of available software (pertaining to the chapters of Part II) is given in an appendix.

The book suffers from certain weaknesses as follows. The exercises at the end of each chapter are mostly elaborations of theory developed within its body. (It might have been better to embed them directly within the text.) This shortcoming diminishes the book’s usefulness as a textbook for an introductory course. The handful of practical examples given in the introductory chapter are “atypical of applications of derivative-free optimization but are easy to understand” (p. 3). The

book provides few numerical illustrations—compare, for example, the discussion of derivative-free algorithms in Kelley [2]—and no implementational details, i.e., its focus is not on the “algorithmic engineering” side of the subject. The elegant theory of *unidimensional* derivative-free algorithms is also not covered in this monograph. It can be found in the classic work of Brent [1] and is relevant, for example, when designing a line search for use within the simplex gradient-based approach. Finally, note that derivative-free algorithms can be employed as *heuristic techniques* for solving non-differentiable optimization problems, and recall from Rademacher’s theorem that a Lipschitz continuous objective function has a Fréchet derivative almost everywhere. The important intersection between derivative-free optimization and non-differentiable optimization is not adequately addressed.

In conclusion, and on a more personal note, we view the monograph under review as *emblematic* of the emerging discipline of algorithmic science & engineering within scientific computing—more specifically, its *theoretical algorithmic science* mode—as delineated in Nazareth [3]. In meeting their main goals (quoted in the third paragraph above), the authors have succeeded admirably. Their book is a gracefully-written, well-organized, and timely contribution to an important area at the current forefront of differentiable optimization research. It provides useful guidance to practitioners on the merits and demerits of a range of choices of algorithms when derivative information cannot be supplied, and its state-of-the-art survey will motivate researchers on new directions to pursue. It also provides useful theoretical advice for developers of mathematical software, for example, the numerical recipes collection of Press et al. [5] and its subsequent editions.

REFERENCES

- [1] R.P. Brent, *Algorithms for Minimization without Derivatives*, Prentice-Hall, Englewood Cliffs, New Jersey, 1973 (reissued by Dover Publications, Mineola, New York, 2002). MR0339493 (49:4251)
- [2] C.T. Kelley, *Iterative Methods for Optimization*, SIAM, Philadelphia, 1999. MR1678201 (2000h:90003)
- [3] J.L. Nazareth, *On algorithmic science and engineering: an emerging core discipline of computing*, SIAM News, 39 (2006), No. 2, pp. 2-4.
- [4] J.A. Nelder and R. Mead, *A simplex method for function minimization*, Computer J., 7 (1965), pp. 308-313.
- [5] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, Cambridge University Press, Cambridge, England, 1992. MR1196230 (93i:65001a)

J. L. NAZARETH

UNIVERSITY OF WASHINGTON, SEATTLE, WA

E-mail address: nazareth@amath.washington.edu