Available online at www.sciencedirect.com

ScienceDirect

Journal of Computational and Applied Mathematics 223 (2009) 62–74

JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS

www.elsevier.com/locate/cam

Unconstrained derivative-free optimization by successive approximation[☆]

Árpád Bűrmen*, Tadej Tuma

University of Ljubljana, Faculty of Electrical Engineering, Tržaška 25, SI-1000 Ljubljana, Slovenia

Received 21 April 2006; received in revised form 20 December 2007

Abstract

We present an algorithmic framework for unconstrained derivative-free optimization based on dividing the search space in regions (partitions). Every partition is assigned a representative point. The representative points form a grid. A piecewise-constant approximation to the function subject to optimization is constructed using a partitioning and its corresponding grid. The convergence of the framework to a stationary point of a continuously differentiable function is guaranteed under mild assumptions. The proposed framework is appropriate for upgrading heuristics that lack mathematical analysis into algorithms that guarantee convergence to a local minimizer. A convergent variant of the Nelder–Mead algorithm that conforms to the given framework is constructed. The algorithm is compared to two previously published convergent variants of the NM algorithm. The comparison is conducted on the Moré–Garbow–Hillstrom set of test problems and on four variably-dimensional functions with dimension up to 100. The results of the comparison show that the proposed algorithm outperforms both previously published algorithms.

MSC: 65K05; 90C56

Keywords: Unconstrained minimization; Direct search; Successive approximation; Grid; Simplex

1. Introduction

Solving unconstrained optimization problems of the form

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ has received a lot of attention lately, in particular methods that search for local minima of f. Several different methods for solving such problems without using derivative information (direct search) were proposed in the past. These so-called direct search methods were despised by the optimization community at first [22] because most of them lacked mathematical analysis. In the past 20 years the advancements in computational

☆ The research was co-funded by the Ministry of Education, Science, and Sport (Ministrstvo za Šolstvo, Znanost in Šport) of the Republic of Slovenia through the programme P2-0246 Algorithms and optimization methods in telecommunications.

* Corresponding author.

E-mail address: arpadb@fides.fe.uni-lj.si (Á. Bűrmen).

0377-0427/$ - see front matter
doi:10.1016/j.cam.2007.12.017

capabilities and simulation techniques lead to many optimization problems where no derivatives of f are available. Consequently direct search methods became interesting for optimization practitioners.

The situation began to change with the advent of the multi-directional search by Torczon [20]. Its convergence theory was based on the fact that all visited points lie on successively finer grids. The convergence theory that followed [21] established the class of pattern search methods. Several well-known algorithms belong to that class, among others also the Hooke–Jeeves algorithm [11]. The developments continued by allowing larger flexibility in choosing the grid [8] and introducing a sufficient descent condition [7] which removes the requirement that the points must lie on a grid.

On the other side developments occurred on generalizing the convergence theory in the direction of nonsmooth functions (functions that are not continuously differentiable). The activities in this field started with the introduction of the generalized pattern search (GPS) [1] and the nonsmooth approach of Coope and Price [9]. GPS evolved into mesh-adaptive direct search (MADS) [2] where an asymptotically dense set of search directions is used. A very good overview which encompasses mostly the analysis for continuously differentiable functions is given in [12]. A somewhat older review of direct search methods can be found in [15]. Some of the above-mentioned convergence analyses are developed for constrained optimization algorithms [1,9,2,12].

This paper presents a framework for ensuring convergence to a local minimizer of continuously differentiable functions. The framework is based on the idea of grid restrainment from [4] where it was used with a very special form of a grid. The generalization presented here allows non-uniform grids provided that some simple requirements are satisfied. These requirements are equivalent to those imposed on the admissible sets in [9] (i.e. the intersection of a bounded set and the grid must always be finite). The division of $\mathbb{R}^n$ into regions (partitions) which define the behavior of the grid-restrainment operator can also be chosen in a very flexible manner.

Our framework is a byproduct of the search for simple convergent variants of the Nelder–Mead (NM) algorithm [18] which gave rise to the notion of grid restrainment. The effect of grid restrainment to successively finer grids can also be viewed from a different perspective. Instead of grid-restrained points we are working with increasingly finer piecewise-constant approximations to f. This interpretation leads to the successive approximation NM (SANM) algorithm. SANM requires less linear algebra operations than its predecessor, the grid-restrained NM (GRNM) algorithm [4], and is also faster.

The paper is divided as follows. First the background for analyzing our framework is developed. The framework is presented and its convergence is established under mild assumptions. Next a variant of the Nelder–Mead algorithm conforming to the presented framework is described. The algorithm is tested on the Moreé–Garbow–Hillstrom [17] test suite and on some multi-dimensional test problems with dimension ranging up 100. The results are compared to those obtained with the convergent simplex variants proposed in [19,5] and [4]. The variant [4] is shown to conform to the proposed framework. Finally the conclusions are given.

Notation. Vectors are denoted by lowercase letters and are assumed to be column vectors so that $\mathbf{x}^T\mathbf{y}$ denotes the scalar product of $\mathbf{x}$ and $\mathbf{y}$. Matrices are denoted by uppercase letters e.g. $\mathbf{A}$. A_{ij} denotes jth element in the ith row of matrix $\mathbf{A}$. The corresponding lowercase letter with a superscript is reserved for matrix columns (e.g. $\mathbf{a}^i$). Set members are also denoted with a superscript. Members of a sequence $\{\mathbf{x}_k\}_{k=1}^{\infty}$ are denoted by a subscript (e.g. $\mathbf{x}_k$). Calligraphic uppercase letters are reserved for maps and sets. $\mathbb{R}$ and $\mathbb{Z}$ denote the set of real and integer numbers, respectively. Function $o(x)$ is such that $\lim_{x\downarrow 0} o(x)/x = 0$. $\mathcal{W}_r$ denotes an open ball with an arbitrary center and radius r. An open ball with radius r centered at $\mathbf{x}$ is denoted by $\mathcal{W}_r(\mathbf{x})$. The remaining notation is introduced in the text as needed.

2. Background

In the case of n-dimensional unconstrained optimization the search is conducted in $\mathbb{R}^n$.

Definition 1. Partitioning $\mathcal{P}(\mathbb{R}^n)$ divides $\mathbb{R}^n$ into a set of partitions P^i such that $\bigcup_i P^i = \mathbb{R}^n$ and $P^i \cap P^j \neq \emptyset$ iff $i = j$.

Now suppose that every partition P^i is assigned a representative point $\mathbf{p}^i \in P^i$. Let $\operatorname{diam}(P^i) = \max_{\mathbf{x},\mathbf{y}\in P^i} \|\mathbf{x}-\mathbf{y}\|$ denote the diameter of a partition.

Definition 2. A grid $\mathcal{G}(\mathbb{R}^n, \mathcal{P})$ is a one-to-one map between the set of representative points and the partitioning $\mathcal{P}(\mathbb{R}^n)$.