AN ALGORITHM USING QUADRATIC INTERPOLATION FOR UNCONSTRAINED DERIVATIVE FREE OPTIMIZATION

Conn Andrew R.[1] and Toint Philippe L.[2]

[1] IBM T.J. Watson Research Center,
P.O.Box 218, Yorktown Heights, NY 10598, USA.
[2] Department of Mathematics,
Facultés Universitaires ND de la Paix,
61, rue de Bruxelles, B-5000 Namur, Belgium, EU.

Abstract. This paper explores the use of multivariate interpolation techniques in the context of methods for unconstrained optimization that do not require derivative of the objective function. A new algorithm is proposed that uses quadratic models in a trust region framework. The algorithm is constructed to require few evaluations of the objective function and is designed to be relatively insensitive to noise in the objective function values. Its performance is analyzed on a set of 20 examples, both with and without noise.

Key words. Nonlinear optimization, trust regions, derivative free algorithms.

1. INTRODUCTION

We are concerned, in this paper, with the problem of minimizing an objective function whose value is determined by measuring some quantity in the real world. This measure may be of a physical nature (for instance, the depth of a certain layer in geophysical exploration) or be related to other contexts. We will focus on the case where there are no constraints on the problem variables. The generalization to simple bounds, in particular, is quite straightforward, as indicated at the end of the paper. Moreover, since the proposed method is derivative free, one might want to handle constrained problems using an exact penalty function.

Three important features characterize these types of problems. Firstly, the cost of obtaining a function value, that is of performing the measure for particular values of the problem variables, is typically very high. This calls for optimization techniques that make optimal use of all such evaluations, possibly at the expense of more extensive linear algebra calculations within the algorithm itself. The second important feature is that the nature of the function evaluation or some other reasons, prevents the computation of any associated derivatives (gradient or Hessian), a serious drawback for many optimization methods. Finally, the considered measure is usually subject to error itself, introducing some "noise" on the objective evaluation, which puts additional requirements on the minimization's robustness.

G. Di Pillo et al. (eds.), *Nonlinear Optimization and Applications*

Note that these problem features may make the calculation of derivatives by finite differences unattractive. Indeed, the additional function evaluations required in the differencing, by this technique, may be very costly and, most importantly, finite differencing can be unreliable in the presence of noise if no specific action is taken to adapt the differentiation step size to the noise level. Since automatic differentiation (see Griewank [22], for example) is not applicable to a "physical" measurement procedure, we thus may be forced to consider algorithms that do not approximate objective function derivatives for a given value of the problem variables.

By extension, we will also consider in this paper unconstrained optimization problems whose objective function is the result of a complex and costly numerical procedure (such as, for example, in the analysis of the in flight vibration of a helicopter rotor), possibly involving some considerable noise (due, for instance, to truncation or approximation in the calculation defining the objective). At variance with the framework described above, automatic differentiation may often be applied to such cases, but the computed derivatives then include differentiation of the noise itself, making the calculated gradients of questionable value to measure the local slope. Furthermore, automatic differentiation is not applicable when the source code for evaluating the objective function is unavailable. Finally, it may not always be straightforward to use, as is for example the case in fluid dynamics calculations where, according to Burns [7], it may generate unwanted dependence on discretization parameters or on the introduction of artificial viscosity.

Derivative free optimization methods have a long history and we refer the reader to Dixon [16], Himmelblau [24] or Polyak [34] for extensive discussion and references. These methods come in essentially five different classes. The first class contains the algorithms which use finite-difference approximations of the objective function's derivatives in the context of a gradient based method, such as nonlinear conjugate gradients or quasi-Newton methods (see, for instance, Stewart [46], Dennis and Schnabel [14], Gill *et al.* [19] and Gill *et al.* [20]). The methods in the second class are often referred to as "pattern search" methods, because they are based on the exploration of the variables' space using a well specified geometric pattern, typically a simplex. They were investigated by Spendley *et al.* [44], Hooke and Jeeves [25] and Nelder and Mead [31], the algorithm proposed by the latter still being one of the most popular minimization technique in use today. More recent developments of pattern search methods include proposals by Torczon [50], Dennis and Torczon [15], Buckley [6] and Elster and Neumaier [18]. The approaches of the third type are based instead on random sampling and were developed by Box [3], Brooks [5] and Kelly and Wheeling [26], to cite a few. The methods of the fourth class are based, as for many methods using derivatives, on the use of successive one-dimensional minimizations (line searches) along selected directions. These directions may be chosen amongst the set of coordinate basis vectors, as in Elkin [17], Ortega and Rheinboldt [32] or Lucidi and Sciandrone [27], with possible reorientation of the basis as described in Rosenbrock [40] and Swann [47], or on sets of mutually conjugate directions, as proposed by Powell [35] and later developed by Brent [4]. Finally, the algorithms of the fifth class are based on the progressive building and updating of a model of the objective function, as proposed by Powell in [38] for linear models and in [39] for quadratic ones. There is also a related class of "global modelling" methods, that uses Design of Experiments (DOE) interpolation models. For instance, in a problem with ten variables one may determine 50 suitably chosen function values (perhaps by using optimal designs, see for example Owen [33]) for determining an initial model that satisfies a maximum likelihood estimator (MLE)