RSCS: A parallel simplex algorithm for the Nimrod/O optimization toolset

Author
Lewis, Andrew, Abramson, David, Peachey, Tom

Published
2006

Journal Title
Scientific Programming

Copyright Statement
Copyright 2006 IOS Press. Reproduced here in accordance with publisher policy. Please refer to the journal for the definitive published version.

Downloaded from
http://hdl.handle.net/10072/11561

Link to published version
http://www.iospress.nl/loadtop/load.php?isbn=10589244
RSCS: A Parallel Simplex Algorithm for the Nimrod/O Optimization Toolset

Andrew Lewis  
Division of Information Services  
Griffith University  
Brisbane, Qld, Australia  
Email: a.lewis@griffith.edu.au

David Abramson  
Dept. of Computer Science and Software Engineering  
Monash University  
Melbourne, Vic., Australia  
Email: davida@csse.monash.edu.au

Tom Peachey  
Dept. of Computer Science and Software Engineering  
Monash University  
Melbourne, Vic., Australia  
Email: tcp@csse.monash.edu.au

Abstract—This paper describes a method of parallelisation of the popular Nelder-Mead simplex optimization algorithms that can lead to enhanced performance on parallel and distributed computing resources. A reducing set of simplex vertices are used to derive search directions generally closely aligned with the local gradient. When tested on a range of problems drawn from real-world applications in science and engineering, this reducing set concurrent simplex (RSCS) variant of the Nelder-Mead algorithm compared favourably with the original algorithm, and also with the inherently parallel multidirectional search algorithm (MDS). All algorithms were implemented and tested in a general-purpose, grid-enabled optimization toolset.

Index terms—Parallel programming, optimization, Nelder-Mead algorithm

I. INTRODUCTION

In scientific and engineering research and design increasingly sophisticated, rigorous and realistic numerical simulations of physical systems are used to understand these systems, and aid in the design process. Engineers can use computational models instead of building physical prototypes to examine the behaviour of components or systems. Such an approach is usually faster and cheaper and hence allows the user to explore various design scenarios. In particular the user may search through combinations of design parameters in order to achieve a design that is optimal in some sense. Similarly, scientific research is increasingly using computational models, and there is often a need to determine those model parameters that produce the best fit to real-world data.

While use of computational models is becoming routine across a wider range of applications, they are often used in an informal manner: an engineer might use a simulation to test a handful of different cases and pick the best of them. More rigorous use to comprehensively explore the design parameter space has the potential to deliver better outcomes. This drives a demand for the capability to perform automatic optimization, minimising or maximising some derived quantity, a measure of “fitness” of the design, to reach a desired objective.

The computation of these objective function values is generally an extremely computationally intensive process when models, which may each take hours to compute, must be run tens, or maybe hundreds, of times to effectively refine possible designs. An ability to deploy the parallel and distributed computing resources that form the basis of contemporary high performance computing architectures would be a distinct advantage in making automatic optimization a practical tool in the engineering design process.

The design engineer who uses the optimization program, while an expert in the application domain, cannot always also be expected to be an expert in computer science. Ideally, to be useful a general purpose optimization tool should be easily applicable to a wide range of problems without assuming specialised knowledge in methods of optimization from the user. The more it can be treated as a “black box”, the wider its potential adoption and the greater its end benefit. To meet these needs, the algorithms described in this thesis have been implemented as components of a fully integrated optimization toolset, Nimrod/O [1]–[3].

Many different optimization algorithms have been developed, from traditional gradient descent methods [4] [5] to more recent innovations inspired by systems in nature, evolutionary and genetic algorithms [6]. Of enduring popularity, particularly for problems with “noisy” objective functions, or where gradient information is unreliable, unavailable or difficult to obtain, are direct search methods. Foremost among these is the simplex algorithm of Nelder and Mead [7].

In the Nelder-Mead simplex algorithm, the \( n + 1 \) vertices of a simplex of approximations to an optimal point in \( n \)-dimensional parameter space are sampled, ordered by objective function value, and an attempt made to replace the worst vertex by reflection through the convex hull of the remaining vertices, using limited sampling along the search direction so defined. Use of the Nelder-Mead simplex algorithm remains current, largely because, on a range of practical engineering problems, it is capable of returning a very good result [8]. It is also robust to small perturbations or inaccuracies in objective function values [9].

Since the original algorithm treats a single vertex at a time, the overall optimization process can be very time-consuming. With the wide availability of parallel and distributed computing resources, an obvious approach to attempt to reduce the total optimization time is to simultaneously consider and relocate several vertices. This paper proposes a method for concurrent execution of a simplex optimization algorithm, and...
Fig. 1. The Nimrod/O Architecture

presents results from numerical tests of the revised algorithm on a number of case studies derived from real-world problems in scientific and engineering applications.

II. THE NIMROD/O TOOLSET

Nimrod/O is a development of the Nimrod research project [10]–[12], incorporating automatic optimization into the framework of what was originally a parameter sweep toolset. Nimrod allows a scientist or engineer to succinctly describe their numerical simulation, define parameter ranges and perform automatic explorations of parameter space using enumeration of the cross-product of the defined parameters on parallel or distributed computers.

While an extremely useful tool, Nimrod suffers from the shortcoming of combinatorial explosion of required model evaluations as problem dimensionality and desired solution resolution increases. Nimrod/O avoids this problem by extending the toolset to include automatic optimization in the same, easily usable framework. Nimrod/O is equipped to run a number of very different optimization algorithms, those described in this paper being just one set. Importantly, different problems are better suited to different algorithms, and Nimrod/O allows the user to try different algorithms (in parallel if the resources are available) and to choose the one that performs best. In this paper we are describing some enhancements to one of these algorithms.

The structure of Nimrod/O and outline of its operation is shown in Figure 1. It is designed to be modular, the incorporation of additional optimization algorithms being a relatively simple operation. Since the C source code for Nimrod/O is freely available [13], the developer of a new algorithm can rapidly obtain the service of distributed evaluations Nimrod/O provides. Nimrod/O will also provide an extensive constraint parser, caching of completed jobs, granularity control for the parameters and statistical summaries of the evaluations.

The simplest way to incorporate an algorithm is to use the “hooks” provided. For example, if the algorithm is coded in the function userSuppliedOptimization1 (for which a skeleton is provided) then that algorithm may be invoked by the line

method special 1
in a schedule file.

The user-supplied function will initially be passed a starting point in search space. What it does with this point is defined by the optimization method. For example, in simplicial methods the starting simplex is constructed around the starting point. However, it should be borne in mind that differing starting points are the only differentiation between the multiple, simultaneous optimizations available through Nimrod/O’s multi-start feature.

The dimensionality of parameter space is available through a global variable, and its bounds through passed variable attribute structures. The code needs to use an array of structures of type Point. Each structure will hold the coordinates of a point in the search space; simultaneous evaluation of the objective function at a number of points constitutes a “batch of jobs”. Execution of a batch is performed by passing the array to the function evalBatchOfJobs. On return each structure gives the status of the corresponding job, the result of any imposed constraints, the location of the “optimum” found and the value of the objective function at that point. More detail is supplied in Chapter 10 of the Nimrod/O User’s Guide [14].

A limitation of this approach is that only the tolerance setting, a value that can be used to decide convergence, is available to the algorithm. Other static, operational parameters of the optimization method may be read from a file, but if the developer wants settings to be passed from the schedule file to the algorithm, it is necessary to modify other parts of the Nimrod/O code. In this case it is advisable to fully incorporate the new method using an existing method as a guide.

Several changes are required. The files involved are shown in parentheses.

- Additions to the parser lexicon to recognise new statements in the schedule (parsesched.c and definitions.h).
- Additions to the structure Opt_Settings to include the information required (definitions.h).
- Insertion of parsing options in ParseInnerLine (parsesched.c).
- Additions of the new method as a case in Optimize
The plan file starts by defining the parameters. A parameter is named, its type defined, how it will be expressed and, where appropriate, the bounds on its value. Supported parameter types are float, integer and text. Float and integer types must be expressed as a range of values, text as a list of values. Where text parameters are used, the listed cases will be evaluated by simple enumeration, i.e. multiple optimizations will be performed, one for each value of the text parameter.

Then, in the section labelled as “task main”, a brief description is given of how to run the numerical simulation. Input and output of the simulation are assumed to be via named files. This has been a common method of interaction with large-scale simulations, and is an easy method independent of the model implementation. For each objective function evaluation task input files are copied to the “node” on which execution is to be scheduled. All other files necessary for execution of the simulation should also be copied, including all necessary scripts and executables.

Then a user-provided script is run on the node. It is assumed that the script will provide the commands to run the simulation and derive the single, floating-point objective function value, possibly by post-processing of the model output. It may be noted that the parameters are provided to the script via named environment variables. These will be substituted by the particular parameter values supplied by the optimization algorithm at run-time. The results, written to a file, are copied back to the scheduling node, and placed in a file with the standard name “output” and an extension identifying the particular job.

Following the description of the model execution is the section specifying the algorithm to be used. Multiple algorithms can be specified, and will be executed simultaneously. Parameters for the algorithms are kept to a minimum, are generally intuitive, and provided with sensible default values. In the example shown, eight simultaneous runs will be performed using a parallel implementation of the Nelder-Mead Simplex algorithm, from random starting points. The algorithm description also specifies a desired solution tolerance. For the Simplex algorithm, and the new variant described in this paper, this specifies the convergence criterion by defining the magnitude of the fractional gradient of the final simplex.

This is sufficient to specify the entire optimization experiment to Nimrod/O. Importantly, the user is not concerned with the computational platform, since support for this is provided by Nimrod or EnFuzion. Issues relating to the Grid are managed through the Nimrod Portal.

---

### III. Reducing Set Concurrent Simplex (RSCS)

The application of supplementary search directions to the Nelder-Mead algorithm, drawing on the methods of the Multidimensional Search (MDS) algorithm of Dennis and Torczon [16]–[18] has been suggested previously by Hamma [19]. This still implemented additional searches sequentially. As illustrated in Figures 2 and 3, a straightforward, concurrent implementation of all the possible Nelder-Mead search directions is potentially inefficient.
When the search directions of the Nelder-Mead simplex algorithm are applied sequentially, step one producing the new vertex 3a and step 2 producing the new vertex 2b, the search directions are generally downhill, relative to the local gradient. However, if they are applied concurrently, it may quite often be the case that one or more may be to some degree uphill (for example, the trial vertex 2a in Figure 3). Note: the MDS search directions have been omitted for clarity: they would be from vertices 2 and 3, through vertex 1.

In this paper a different approach to generating supplementary search directions for concurrent search is proposed. The search direction from the worst vertex is through the centroid of the remaining $n$ vertices, as in the normal Nelder-Mead algorithm. But the search direction from the next worst vertex is through the $n-1$ remaining vertices that are better than it, and so on, until the search direction from the second best vertex is reduced to the MDS search direction though the best vertex. All searches are performed concurrently, and all vertices are independently relocatable.

The method can be considered as deriving from a hybrid of the Nelder-Mead and MDS algorithms. In this work it will be referred to as the Reducing Set Concurrent Simplex (RSCS) algorithm. The set of search directions generated is illustrated in Figure 4. Comparing the search directions illustrated in Figure 4 with the concurrent search directions of Figure 3, it may be noted that the search directions generated by RSCS are more likely to be downhill.

Viewed in two dimensions RSCS can appear to be a simple hybrid of Nelder-Mead and MDS. However, if the 3-dimensional case is considered it becomes apparent there are differences. Figures 5, 6 and 7 show the three search directions constructed by RSCS, and the edges and vertices used in their construction. The search direction shown in Figure 6 will not appear in either the Nelder-Mead or MDS algorithms.
IV. NUMERICAL EXPERIMENTS

We have assembled a number of case studies drawn from interesting and challenging scientific and engineering applications. These were used to test and assess the performance of the individual algorithms. The following problems were used for these investigations:

A. Laser 1 and 2

A two-dimensional test surface was derived from the computation of a quantum electrodynamical simulation of a laser-atom interaction experiment [10]. The base case, Laser 1, is quite a smooth surface, the dataset containing only 4 minima, of which the global minimum is quite dominant, as can be seen in Figure 8(a). Additive fractal noise was overlaid on this dataset to develop a "noisier", more challenging surface to test the algorithms. This dataset, Laser 2, contained 1157 local minima of varying severity, and is illustrated in Figure 8(b).

B. Crack 1 and 2

Finite element analysis of a thin plate under cyclic loading, with a cutout specified by parameters, was used to generate the Crack datasets [20]. Common practice in damage tolerant design has been to minimise the maximum stress under load. Isosurfaces of these stress values are shown in Figure 8(c). This dataset, Crack 1, was reasonably smooth, with only 26 local minima. A new approach in modeling stressed components is to attempt to maximize durability. The Crack 2 model seeks to maximise the life of the part as determined by the minimum time taken for fatigue crack growth to a defined length from any of a number of starting crack locations. Isosurfaces at a number of values are shown in Figure 8(d). In contrast to Crack 1, this dataset was "noisy", with 540 local maxima, and discontinuous isosurfaces.

C. Aerofoil

This test case models the aerodynamic properties of a two dimensional aerofoil. The objective function to be minimised is the lift-drag ratio [2], and this is computed by executing a Computational Fluid Dynamic model of the object. Figure 8(e) shows a number of isosurfaces in the parameter space investigated. The dataset was generally smooth, with only 12 local minima and a dominant global minimum.

D. Bead

The application from which this case study was drawn used a ceramic bead to minimise distortion of the radiation pattern of a mobile telecommunications handset during testing [21]. The objective function value, derived from an FDTD full-wave analysis of the handset and signal feed cable structure, was a measure of transmission strength through the bead at 1 GHz. The dataset for the Bead case study, of which isosurfaces for a particular value are shown in Figure 8(f), is quite complex and contains 298 local minima.

E. Rosenbrock’s function

In order to provide a point of comparison, the well-known Rosenbrock’s function in two dimensions was included. The objective function values for this test case were directly computed from $f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$ for $x_i \in [-2, 2]$ which has one local minimum at $f(1, 1) = 0$.

F. Case study assembly

It is generally not practical to directly use complex, real-world problems as test cases. The “black box”, when queried, can take a considerable amount of time and computational resource to provide a response. So parameter sweeps were made of the numerical models that form the basis of these test cases, and the output data stored. These pre-computed data are interrogated, and linear interpolation employed to provide realistic responses from what are, in effect, “sandboxes” in which optimization programs can readily be tested. These “sandboxes” themselves represent a large investment of time and computational resource – for example, the data acquisition necessary to build the “Bead” test case required over 2 months of continuous computation on a multi-processor supercomputer.

Generally, the case studies fell into 2 sets:
- Smooth, with a dominant global minimum (Laser 1, Crack 1, Aerofoil, Rosenbrock’s function)
- Multiple/many local minima, non-convex ( Laser 2, Crack 2, Bead)
In formulating ideas about the most appropriate algorithm for use with a particular problem, a great deal may depend on whether the problem encountered is "noisy" or "smooth". Particular attention has been addressed to this issue when assessing algorithm performance.

V. RESULTS OF EXPERIMENTS

To evaluate the new RSCS algorithm, it was compared with the original, Nelder-Mead algorithm and an implementation of the MDS algorithm. The Nelder-Mead algorithm performs evaluation of four points when considering relocation of each vertex, corresponding to the various alternatives of reflection, extension and contraction of the simplex. All these points are independent of each other, completely defined by the existing simplex geometry, and could be evaluated concurrently. The algorithm was slightly modified to exploit this concurrency, and all results reported for the "original" algorithm in this section refer to this parallel implementation.

Each of the algorithms was run on each of the test cases from 10 randomly distributed start points. For the purposes of comparison, in a given test case the same set of start points were used for each algorithm. The starting simplices were right simplices aligned with the coordinate axes. By default they were scaled to 10% of the parameter range for each coordinate, as use of reasonably large simplices has been shown to enhance performance [22]. Convergence criterion for most cases was a fractional step-wise gradient of $10^{-3}$.

Function evaluations are performed concurrently in batches in Nimrod/O. The batch count can be interpreted as equivalent to Effective Serial Function Evaluations (ESFE), a measure of the wall-clock time taken for completion, providing the machine has enough processors to concurrently evaluate all points, a reasonable assumption given the ready access to cheap clusters.

Analysis of the returned objective function values using the Shapiro-Wilk W test statistic determined that the results are not normally distributed. For this reason, median values and non-parametric, descriptive statistical methods are used for comparison of algorithms.

Tables I, II and III show, for each algorithm on each test case over the 10 runs performed, respectively:

- The median objective function value obtained,
- The median number of function evaluations performed (FE) and
- The median Effective Serial Function Evaluations (ESFE).

For each test case, the best median objective function value is highlighted in bold type. Also highlighted is the fastest time, in terms of ESFE, for an algorithm to achieve a result within 10% of the best median objective function, as a percentage of the range of median values obtained.

Table IV shows the best objective function value obtained in 10 runs for each algorithm on each test case. Table V shows the actual ESFEs required to obtain that result. For each test case, the best objective function value obtained by any algorithm, and the fastest ESFE to obtain that value, are highlighted in bold type.
From Table I, it can be seen that RSCS appears to deliver slightly better results than Nelder-Mead on the majority of test cases. From Table IV it can be seen that RSCS was also capable of equalling the best objective function value returned by both the Nelder-Mead and MDS algorithms on almost all test cases. In the one case it fell slightly behind, it was by less than 0.2%. Analysis of objective function values returned, using the Kruskal-Wallis H test statistic and pair-wise comparisons using Mann-Whitney U test statistics indicated there was no statistically significant difference in the quality of results returned by RSCS, Nelder-Mead simplex and MDS algorithms.

The only significant remaining difference between RSCS and the other algorithms is thus its speed. RSCS gave median ESFE that were consistently better than the Nelder-Mead algorithm. On average, RSCS was 78% faster.

From Table III, it can be seen that RSCS is approximately 20% faster than MDS on average. This excludes the time taken by MDS on Rosenbrock’s function, which was considered a pathological example of the tendency of simplicial methods toward premature convergence [23]. If the median time taken on Rosenbrock’s function is included, RSCS is on average faster than MDS by a factor of 4.5. The decision to include these poorer results may be justified by reference to Table V in which it can be seen that on two different test cases MDS terminated by exceeding the maximum permissible iterations (1000), rather than satisfying convergence criteria. It can be conjectured that the insistence on congruency of consecutive simplices in the MDS algorithm forces premature contraction of the simplex diameter.

To further investigate the behaviour of the algorithms, the median objective function values achieved after each iteration of each algorithm across all 10 runs on each of the real-world test cases were examined. As noted above, the case studies used can be classified into two classes, “smooth” and “noisy”. In Figures 9 and 10 the two case studies derived from the
quantum electrodynamical models, Laser 1 and Laser 2, are shown. The first of these exemplifies a “smooth” test case, the latter a “noisy” one.

VI. CONCLUSIONS

A method of concurrent execution of a simplex optimization algorithm has been proposed, and its performance on a range of real-world problems compared with the popular Nelder-Mead and MDS algorithms from which it is derived. The supplementary search directions used are constructed from reducing sets of vertices in a manner which increases the probability they will be aligned with the local gradient.

In general, the Nelder-Mead and MDS algorithms perform well on the “smooth” test cases, but poorly on the class of cases with noise or many local minima. The new algorithm, RSCS, performs reasonably (or very) well on both types of problems. It generally provides equivalent or slightly better results, and delivers them considerably faster.

A general-purpose optimization toolset, Nimrod/O, in which the prototype and test algorithms have been implemented has also been described. The toolset makes the use of parallel and distributed computing for automatic optimization in the engineering design process readily and easily applicable. Nimrod/O can run a number of very different optimisation algorithms, in parallel if the resources are available. It allows a user to try different algorithms and select those which perform best on a particular problem. The toolset is configured for use on parallel computers, collections of workstations on networks, or grid-based computing resources.

By increasing the speed of the optimization algorithm using parallel computing resources, the method described delivers automatic optimization as a feasible tool for use by the practising engineer tackling real-world problems. With the growing availability of inexpensive computing clusters, rigorous exploration of design alternatives can be made, rather than haphazard, ad hoc evaluation of limited numbers of prototype designs.

REFERENCES


