On Recursive Multiscale Trust-Region Algorithms for Unconstrained Minimization

Gratton, Serge; Sartenaer, Annick; Toint, Philippe

Published in:
Oberwolfach Reports: Optimization and Applications

Publication date:
2005

Document Version
Peer reviewed version

On Recursive Multiscale Trust-Region Algorithms for Unconstrained Minimization

ANNICK SARTENAER, PHILIPPE L. TOINT
(joint work with Serge Gratton)

A large class of large-scale finite-dimensional minimization programs arise from the discretization of infinite-dimensional problems, such as optimal-control problems defined in terms of either ordinary or partial differential equations. We report here on a potentially efficient new class of algorithms using this structure and briefly discuss a first set of numerical experiments.

A simple first approach is to use coarser grids in order to compute approximate solutions which can then be used as starting points for the optimization problem on a finer grid (see [10], [3], [5] or [4], for instance). However, potentially more efficient techniques are inspired from the multigrid paradigm in the solution of partial differential equations and associated systems of linear algebraic equations (see, for example, [6] or [7]). The work presented here was in particular motivated by the "generalized truncated Newton algorithm" presented in Fisher [9], a talk by Moré [13], the contributions by Nash and Lewis [11, 12] and the computational success of the low/high-fidelity model management techniques of Alexandrov, Lewis and co-authors [1, 2].

We consider the solution of the unconstrained optimization problem

$$\min_{x \in \Re^n} f(x), \tag{1}$$

where f is a twice-continuously differentiable objective function which maps $\Re^n$ into $\Re$ and is bounded below. The trust-region methods which we study produce, given an initial point x_0, a sequence $\{x_k\}$ of iterates (hopefully) converging to a local first-order ciritical point for the problem. At each iterate x_k, these methods build a (typically quadratic) model $m_k(x_k + s)$ of $f(x_k + s)$. This model is then assumed to be adequate in a "trust region", defined as a sphere of radius $\Delta_k > 0$ centered at x_k, and a step s_k is then computed that sufficiently reduces this model in the region. The objective function is computed at the trial point $x_k + s_k$ and this trial point is accepted as the next iterate if and only if the achieved reduction in f is sufficiently large compared to the predicted reduction in m_k. The value of the radius is finally updated to ensure that it is decreased when the trial point cannot be accepted as the next iterate, and is increased otherwise. Obtaining sufficient decrease on this model then amounts to (approximately) solving the problem $\min_{\|s\| \leq \Delta_k} m_k(x_k + s)$.

We investigate what can be done to reduce the cost of solving (1) if one attempts to exploit the knowledge of simplified expressions of the objective function, when available. More specifically, we assume that we know a collection of functions $\{f_i\}_{i=0}^r$ such that each f_i is a twice-continuously differentiable function from $\Re^{n_i}$ to $\Re$ (with $n_i \geq n_{i-1}$), the connection with our original problem being that $n_r = n$ and $f_r(x) = f(x)$ for all $x \in \Re^n$. We will also assume that, for each $i = 1, \ldots, r$, f_i is "more costly" to minimize than f_{i-1}. This may be because f_i has more

variables than f_{i-1} (as would typically be the case if the f_i represent increasingly finer discretizations of the same infinite-dimensional objective), or because the structure (in terms of partial separability, sparsity or eigenstructure) of f_i is more complex than that of f_{i-1}, or for any other reason. To fix terminology, we will refer to a particular i as a *level*. Of course, for f_{i-1} to be useful at all in minimizing f_i, there should be some relation between the variables of these two functions. We henceforth assume that, for each $i = 1, \ldots, r$, there exist a full-rank linear operator R_i from $\Re^{n_i}$ into $\Re^{n_{i-1}}$ and another full-rank operator P_i from $\Re^{n_{i-1}}$ into $\Re^{n_i}$ such that $P_i = R_i^T$, where P_i and R_i are interpreted as restriction and prolongation between a fine and a coarse grid. The idea is then to use f_{r-1} to construct an alternative model h_{r-1} for $f_r = f$ in the neighbourhood of the current iterate, that is cheaper than m_k, and to use this alternative model to define the step in the trust-region algorithm whenever possible. If more than two levels are available ($r > 1$), this can be done recursively, the approximation process stopping at level 0, where the usual quadratic model is always used. The resulting algorithm can then be specified as a variant of the Basic Trust-Region Algorithm of [8].

We briefly describe the global convergence theory associated with this algorithm and show convergence from arbitrary starting points to first-order critical points under classical assumptions. We also discuss an associated dimension-independent worst-case complexity result. We finally present a first numerical application for one of the possible implementations. This implementation specifies the nature of the non-recursive iterations, which fall into two classes: smoothing iterations, aimed at decreasing high-frequency components of the gradient, and damping iterations, which decrease their low-frequency components (an important question is to modify these iterations as to ensure "sufficient decrease" in the sense of the Cauchy condition). Other implementation questions are concerned with the form of the recursive iterations, ranging from free form (where the optimization at lower levels is governed purely by accuracy requirements) to fixed cycles (such as the V and W cycles inspired by multigrid techniques). Demonstration of the efficiency of the method is done on a minimum surface problem with highly oscillatory boundary conditions. Problems of this type involving up to 1.1 million variables were solved by the new algorithm in MATLAB on an oldish laptop PC (Pentium 4 Mobile, 1.6 GHz). Perspectives are finally described, which are both numerous and interesting.

References

[1] N. M. Alexandrov and R. L. Lewis. An overview of first-order model management for engineering optimization. *Optimization and Engineering*, 2:413–430, 2001.

[2] N. M. Alexandrov, R. L. Lewis, C. R. Gumbert, L. L. Green, and P. A. Newman. Approximation and model management in aerodynamic optimization with variable fidelity models. *Journal of Aircraft*, 38(6):1093–1101, 2001.

[3] R. E. Bank, P. E. Gill, and R. F. Marcia. Interior point methods for a class of elliptic variational inequalities. In T. Biegler, O. Ghattas, M. Heinkenschloss, and B. Van Bloemen Waanders, editors, *High performance algorithms and software for nonlinear optimization*, pages 218–235, Heidelberg, Berlin, New York, 2003. Springer Verlag.

[4] S. J. Benson, L. C. McInnes, J. Moré, and J. Sarich. Scalable algorithms in optimization: Computational experiments. Preprint ANL/MCS-P1175-0604, Mathematics and Computer Science, Argonne National Laboratory, Argonne, Illinois, USA, 2004. To appear in the Proceedings of the 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization (MA&O) Conference, August 30 - September 1, 2004.

[5] J. T. Betts and S. O. Erb. Optimal low thurst trajectory to the moon. *SIAM Journal on Applied Dynamical Systems*, 2(2):144–170, 2003.

[6] A. Brandt. Multi-level adaptative solutions to boundary value problems. *Mathematics of Computation*, 31(138):333–390, 1977.

[7] W. L. Briggs, V. E. Henson, and S. F. McCormick. *A Multigrid Tutorial*. SIAM, Philadelphia, USA, 2nd edition, 2000.

[8] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. *Trust-Region Methods*. Number 01 in MPS-SIAM Series on Optimization. SIAM, Philadelphia, USA, 2000.

[9] M. Fisher. Minimization algorithms for variational data assimilation. In *Recent Developments in Numerical Methods for Atmospheric Modelling*, pages 364–385. ECMWF, 1998.

[10] A. Griewank and Ph. L. Toint. Local convergence analysis for partitioned quasi-Newton updates. *Numerische Mathematik*, 39:429–448, 1982.

[11] M. Lewis and S. G. Nash. Practical aspects of multiscale optimization methods for VLSICAD. In Jason Cong and Joseph R. Shinnerl, editors, *Multiscale Optimization and VLSI/CAD*, pages 265–291, Dordrecht, The Netherlands, 2002. Kluwer Academic Publishers.

[12] M. Lewis and S. G. Nash. Model problems for the multigrid optimization of systems governed by differential equations. *SIAM Journal on Scientific Computing*, (to appear), 2005.

[13] J. J. Moré. Terascale optimal PDE solvers. Talk at the ICIAM 2003 Conference in Sydney, 2003.