Journal of Computational Mathematics, Vol.22, No.4, 2004, 489–500.

A DIRECT SEARCH FRAME-BASED CONJUGATE GRADIENTS METHOD *

I.D. Coope C.J. Price
(*Department of Mathematics and Statistics, University of Canterbury, Private Bag 4800, Christchurch, New Zealand*)

Abstract

A derivative-free frame-based conjugate gradients algorithm is presented. Convergence is shown for C^1 functions, and this is verified in numerical trials. The algorithm is tested on a variety of low dimensional problems, some of which are ill-conditioned, and is also tested on problems of high dimension. Numerical results show that the algorithm is effective on both classes of problems. The results are compared with those from a discrete quasi-Newton method, showing that the conjugate gradients algorithm is competitive. The algorithm exhibits the conjugate gradients speed-up on problems for which the Hessian at the solution has repeated or clustered eigenvalues. The algorithm is easily parallelizable.

Mathematics subject classification: 90C56, 90C30, 65K05.
Key words: Conjugate gradients, Derivative-free, Frame-based methods, Numerical results.

1. Introduction

The linear conjugate gradient was developed by Hestenes and Stiefel [10], and extended to the minimization of general functions by Fletcher and Reeves [6]. A description of conjugate gradients methods for minimising general functions can be found in [5, 6, 7, 8, 9, 12] and elsewhere. In this paper we consider the application of conjugate gradients techniques to unconstrained minimization of C^1 functions in a derivative-free context. A method is described which conforms to the frame-based template in [2], thereby guaranteeing convergence under standard conditions. The problem may be formally stated as

$$\min_{x \in R^n} f(x),$$

where a local, but not necessarily a global minimizer is sought. Here we restrict attention to objective functions f which are continuously differentiable, but do not assume that gradient information is available. Second order optimality conditions are not useable as second derivatives may not exist. Consequently stationary points are accepted as solutions. The algorithm forms an estimate of the gradient at each iterate, but does not rely on the accuracy of these estimates to guarantee convergence. These gradient estimates are used to form conjugate gradients search directions, and line searches are conducted along these directions. Hence the algorithm mimics a conjugate gradients method when it can (that is to say, when its gradient estimates happen to be accurate), which makes it more effective in practice. Convergence is guaranteed by the frame-based nature of the algorithm, not the fact that it mimics a conjugate gradients method. The theoretical convergence properties are unaffected by the accuracy of gradient estimates, although inaccurate estimates will, in general, degrade the numerical performance of the algorithm.

* Received August 26, 2002; final revised June 4, 2003.

The conjugate gradients method used is that of Polak and Ribière [12], and Polyak [13] (hereafter PRP). Limited numerical comparisons between frame based PRP and Fletcher–Reeves methods indicated that the former was more promising. This preference seems to be in accord with the case when exact gradients are available [7]. The automatic reset property [14] of the PRP method is also very desirable for higher dimensional problems.

The algorithm generates a sequence of iterates $\{x^{(k)}\}$, where k is the iteration number. At each iteration the function values at a set of points $\Phi^{(k)}$ called a frame are calculated. Frames are defined precisely in the following section, and a template for frame-based algorithms is given in Section 3. Loosely speaking, the points in the frame surround $x^{(k)}$. The gradient at $x^{(k)}$ is estimated using points from the frame $\Phi^{(k)}$. This gradient estimate allows a derivative-free conjugate search direction to be formed, as described in Section 4. A line search is conducted along this search direction, yielding the next iterate. The process is repeated until an adequate approximation to a stationary point is obtained. The choice of frames yields a second order gradient estimate at each iterate. The line search uses parabolic interpolation to locate an approximation to each line local minimum. On a quadratic these gradient estimates and approximations to line local minima are exact, and so the algorithm exactly minimizes convex quadratics in a finite time. A description of the line search is given in Section 5. Numerical results and concluding remarks are presented in Sections 6 and 7.

2. Frames

A frame Φ is defined by a frame centre x, a frame size $h > 0$, and a positive basis $\mathcal{V}_+$. A positive basis [4] (see also [18]) is a set of vectors $\mathcal{V}_+$ with the following two properties:

(i) every vector in R^n is a linear combination of the members of $\mathcal{V}_+$, where all coefficients of the linear combination are non-negative; and

(ii) no proper subset of $\mathcal{V}_+$ satisfies (i).

A frame $\Phi(x, h, \mathcal{V}_+)$ around x is a set of points of the form $\Phi(x, h, \mathcal{V}_+) = \{x + hv : v \in \mathcal{V}_+\}$. It is shown in [4] that positive bases (and hence also frames) have at least $n + 1$ and at most $2n$ members. Positive bases and frames containing $2n$ members are called maximal. Maximal positive bases [4] are of the form $\mathcal{V}_+ = \{v_1, v_2, \ldots, v_n, -v_1, -v_2, \ldots, -v_n\}$ where $v_1, \ldots, v_n$ are a basis for R^n. Herein we use $v_i = e_i$, where e_i is the i^{th} unit vector. This yields a frame around a frame centre $x^{(k)}$ of the form:

$$\Phi^{(k)} = \{x^{(k)} + h^{(k)} e_i : \forall i = 1, \ldots, n\} \cup \{x^{(k)} - h^{(k)} e_i : \forall i = 1, \ldots, n\}. \tag{1}$$

Each such maximal frame contains enough information to form second order estimates of the gradient at $x^{(k)}$, and also the diagonal entries of the Hessian at $x^{(k)}$. These second derivative estimates are used to scale the decision variables at each reset.

Frames which are called quasi-minimal are of particular interest. These frames have the property that no frame point is more than ϵ lower than the frame centre, where ϵ is a preselected non-negative constant. The convergence theory [2] shows that any method conforming to it will generate an infinite subsequence of quasi-minimal frames. The convergence theory also shows that, under mild conditions, the cluster points of this subsequence of quasi-minimal frame centres are stationary points of f.

At each iteration a positive constant $\epsilon^{(k)}$ is chosen. The frame $\Phi^{(k)}$ is called *quasi-minimal* if and only if

$$f\left(x^{(k)}\right) \le f(x) + \epsilon^{(k)} \quad \forall x \in \Phi^{(k)}. \tag{2}$$