Available online at www.sciencedirect.com
SCIENCE DIRECT®
Nonlinear Analysis 63 (2005) e2621 – e2628
Nonlinear Analysis
www.elsevier.com/locate/na

Conjugate gradient methods in Banach spaces

I. Stein, Jr.

Department of Mathematics, The University of Toledo, Toledo, OH 43606-3390, USA

Abstract

One purpose of this paper is to extend the conjugate gradient method to minimize functions on Banach spaces that are norm reflexive, strictly convex and non-Hilbert. The algorithms in this paper are based upon the notion of the metric gradient defined by Golomb and Tapia (Numer. Math. 20 (1972) 115–124). Generalizations are made for the algorithms of Daniel (The Approximate Minimization of Functionals, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1971; The conjugate gradient method for linear and nonlinear operator equations, Ph.D. Thesis, Stanford University, Palo Alto, CA, 1965), Polak (Computational Methods in Optimization—a Unified Approach, Academic Press, New York and London, 1971), Polak–Ribière (Rev. Française Informat. Recherche Opérationnelle 3 (16) (1969) 35–43), and Fletcher–Reeves (Comput. J. 7 (1964) 149–154). A local convergence theorem is given for a class of descent conjugate gradient methods for functionals defined on Banach spaces. Applications include problems in differential equations and the calculus of variations where the Banach spaces are Sobolev spaces and the second Frechet differential has finite signature and nullity as defined by Hestenes (Pacific J. Math. 1 (1951) 525–581).

Keywords: Conjugate gradient methods; Metric gradient; Numerical optimization; Direct methods in the calculus of variations; Numerical solutions of differential equations

1. Introduction

In 1952, Hestenes and Stiefel [9] developed the conjugate gradient method primarily as a practical method for solving a large linear system of equations that arose in finding a numerical solution to an elliptic partial differential equation. Their idea in solving a large linear system was based upon the equivalent problem of minimizing a related quadratic function on a finite dimensional space. In 1952, Hayes [6,7] extended their method from

E-mail address: Ivie.Stein@utoledo.edu.

doi:10.1016/j.na.2005.02.111

finite dimensional spaces to Hilbert spaces. In 1964, Fletcher and Reeves [4] extended the conjugate gradient method to minimize nonquadratic functions on finite dimensional spaces. In 1965, Daniel [2,3] extended the conjugate gradient method to minimize nonquadratic functions on Hilbert spaces. In 1972, Golomb and Tapia [5] introduced the idea of a gradient in Banach spaces. It is that idea upon which this paper is based. In 1975, Byrd and Tapia [1] extended steepest descent to normed linear spaces, and in 2002, Penot [10] used the gradient of Golomb and Tapia [5] to obtain convergence of descent algorithms in normed linear spaces. The procedures provided in this paper relative to the norm used for weak extremums in variational theory provide computationally direct methods in the calculus of variations. They can work for Banach spaces which are not norm reflexive and strictly convex.

Assume that a nonquadratic function F has a second Frechet differential that is positive definite and bounded in a closed convex set C with F bounded below on C. As for notation let $\mathrm{d}F(x, h)$ denote the Frechet differential of F at x in the direction h as defined in [13, p. 40], and let $\nabla F(x)$ denote the metric gradient of F at x as defined in [5, p. 116], and let $\mathrm{d}^2F(x)(h_1, h_2)$ denote the second Frechet differential of F at x as defined in [13, p. 55].

Consider the second Frechet differential as a quadratic form $\mathrm{d}^2F(x)(h) = \mathrm{d}^2F(x)(h, h)$ in h, and assume it has finite signature and nullity as defined by Hestenes [8].

2. Conjugate gradient algorithms in Banach spaces

In this section we present a class of downhill or descent algorithms for minimizing functions defined on a Banach space.

1. Polak–Ribière algorithm (refer to Polak [11, pp. 53–54] and to Polak–Ribière [12]).

   (a) Select an x_0 in the Banach space, and compute $\nabla F(x_0)$.
   (b) If $\nabla F(x_0) = 0$, then terminate the algorithm.
   (c) If $\nabla F(x_0) \neq 0$, then set $p_0 = -\nabla F(x_0)$.
   (d) For $i = 0, 1, 2, \ldots,$ compute x_{i+1} by

   $$x_{i+1} = x_i + a_i p_i,$$

   where a_i is the first positive real number a, which minimizes $F(x_i + ap_i)$ locally.
   (e) Compute $\nabla F(x_{i+1})$.
   (f) If $\nabla F(x_{i+1}) = 0$, then terminate the algorithm.
   (g) If $\nabla F(x_{i+1}) \neq 0$, then compute p_{i+1} by

   $$p_{i+1} = -\nabla F(x_{i+1}) + b_i p_i, \quad \text{where } b_i = \frac{(\mathrm{d}F(x_{i+1}) - \mathrm{d}F(x_i))(\nabla F(x_{i+1}))}{\mathrm{d}F(x_i)(\nabla F(x_i))}$$

   (h) Set $i = i + 1$, and go to step (d).
2. Daniel algorithm (refer to Daniel [11, p. 125]). Replace b_i in step (g) of the Polak–Ribière algorithm by

$$b_i = \frac{\mathrm{d}^2F(x_{i+1}, p_i)(\nabla F(x_{i+1}))}{\mathrm{d}^2F(x_{i+1}, p_i)(p_i)}.$$

3. Fletcher–Reeves algorithm (refer to Polak [11, p. 52]). Replace b_i in step (g) of the Polak–Ribière algorithm by

$$b_i = \frac{\mathrm{d}F(x_{i+1})(\nabla F(x_{i+1}))}{\mathrm{d}F(x_i)(\nabla F(x_i))}.$$

3. Assumptions

Let X be a norm reflexive strictly convex Banach space. Using the notations of Sections 1 and 2, assume that the algorithms we consider generate a sequence $\{x_n\}$, $n = 1, 2, \ldots,$ from a starting point x_0 which is in a convex set C having the properties that F is bounded below on C, F is continuous on C, $\mathrm{d}F(x, h)$ is continuous as a function of h for all x in C, and that the following condition holds: there are positive constants m and M for which

$$m\|h\|^2 \leqslant \mathrm{d}^2 F(x, h) \leqslant M\|h\|^2$$

for all x in C and for all h in X.

For the algorithms and functions we consider, we assume that for all n, $\mathrm{d}F(xn) \neq 0$, the minimums a_n exist, $p_n \neq 0$, and $x_{n+1} \neq x_n$. In addition, we observe later that some of the algorithms we consider satisfy the condition that

$$\|p_n\| \leqslant B\|\nabla F(x_n)\|$$

for some $B > 1$. This is analogous to the approach by Daniel [2, p. 128], where $\|b_{n-1} p_{n-1}\| \leqslant D\|r_{n-1}\|$ for some positive D which leads to $\|p_n\|^2 \leqslant (1 + D^2)\|r_n\|^2$ for his algorithm. Our condition with B generalizes to a class of algorithms in Banach space what Daniel did for his algorithm in Hilbert space. This generalization includes Daniel's algorithm and the algorithm of Polak–Ribière in the Banach space setting. For Daniel's algorithm in the Hilbert space setting our B is Daniel's $(1 + D^2)^{1/2}$.

4. Preliminaries

Throughout this section we work under the assumptions and notation of Sections 1, 2, and 3. Assume x is in C described in Section 3. From Theorem 4.6 below we observe that the class of algorithms we consider are descent or downhill methods.

Lemma 4.1. *If* $\mathrm{d}F(x_i) \neq 0$ *and* x_i *is in* C, *then* $\mathrm{d}F(x_{i+1})(p_i) = 0$.

Theorem 4.1. *Let* x *be in* C. *Assume that* $\mathrm{d}F(x_i) \neq 0$. *Then*

(1) $\|\mathrm{d}F(x)\| = \|\nabla F(x)\|$,
(2) $\mathrm{d}F(x)(\nabla F(x)) = \|\mathrm{d}F(x)\|^2$,
(3) $\mathrm{d}F(x_i)(p_i) = -\mathrm{d}F(x_i)(\nabla F(x_i))$, *and*
(4) $\mathrm{d}F(x_i)(p_i) = -\|\nabla F(x_i)\|^2$.

Remark. Theorem 4.1 tells us that $\mathrm{d}F(x) = 0$ iff $\nabla F(x) = 0$ for x in C.

One of our objectives is to establish bounds on a_i such as those described in Theorem 4.5. We do this through the sequence of lemmas and theorems given below describing the relationship between norms of $\nabla F(x_i)$ and p_i.

Lemma 4.2. *If* $\mathrm{d}F(x_i) \neq 0$ *then* $\|\nabla F(x_i)\| \leqslant \|p_i\|$.

Theorem 4.2. *If* $\mathrm{d}F(x_i) \neq 0$ *and* $p_i \neq 0$ *for all* i, *then for all* i

(1) $a_i = \dfrac{\|\nabla F(x_i)\|^2}{\int_0^1 \mathrm{d}^2 F(x_i + t a_i p_i, p_i)\,\mathrm{d}t}$,

(2) $\dfrac{\|\nabla F(x_i)\|^2}{M\|p_i\|^2} \leqslant a_i \leqslant \dfrac{\|\nabla F(x_i)\|^2}{m\|p_i\|^2}$,

(3) *If* $\|p_i\| \leqslant B\|\nabla F(x_i)\|$ *for all* i *with* $B > 1$, *then* $1/MB^2 \leqslant a_i \leqslant 1/m$.

In Part (3) of Theorem 4.2 above, we assume the condition $\|p_i\| \leqslant B\|\nabla F(x_i)\|$ for some constant $B > 1$. Below we shall outline the details of finding the constants B explicitly in terms of m and M for different algorithms.

Lemma 4.3. *If* $\mathrm{d}F(x_i) \neq 0$, *then* $\|\nabla F(x_{i+1})\| \leqslant (1 + M/m)\|\nabla F(x_i)\|$ *is true for all algorithms we consider.*

Corollary 4.1. *If* $\mathrm{d}F(x_i) \neq 0$, *then* $\|\nabla F(x_{i+1})\| \leqslant (1 + M/m)\|p_i\|$ *is true for all algorithms we consider.*

Lemma 4.4. *If* $p_i \neq 0$ *and if the assumptions of Section* 3 *of this paper hold*, *then*

$$\|p_i\| \leqslant (1 + M/m)\|\nabla F(x_i)\|$$

holds for the Polak–Ribière algorithm. Hence, *the quantity* B *defined in Section* 3 *of this paper is given by*

$$B = \left(1 + \frac{M}{m}\right)$$

for the Polak–Ribière algorithm.

Corollary 4.2. $\mathrm{d}F(x_i) \neq 0$ *and* $p_i \neq 0$ *then*

$$\|p_{i+1}\| \leqslant \left(1 + \frac{M}{m}\right)^2 \|p_i\|$$

for the Polak–Ribière algorithm.

Lemma 4.5. *If* $p_i \neq 0$, *then* $\|p_{i+1}\| \leqslant (1 + M/m)\|\nabla F(x_{i+1})\|$ *for Daniel's algorithm. Hence*, B *defined in Section* 3 *of this paper is given by* $B = M/m$ *for Daniel's algorithm.*

Corollary 4.3. *If* $\mathrm{d}F(x_i) \neq 0$ *and* $p_i \neq 0$, *then for Daniel's algorithm*

$$\|p_{i+1}\| \leqslant (M/m)\left(1 + \frac{M}{m}\right)^2 \|p_i\|.$$

Lemma 4.6. *If* $\mathrm{d}F(x_i) \neq 0$, *then for all algorithms we consider*

$$\|p_{i+1}\| \leqslant \left(1 + \frac{M}{m} + |b_i|\right) \|p_i\|.$$

Corollary 4.4. *If* $\mathrm{d}F(x_i) \neq 0$, *then for the Fletcher–Reeves algorithm*

$$b_i \leqslant \left(1 + \frac{M}{m}\right)^2.$$

Corollary 4.5. *If* $\mathrm{d}F(x_i) \neq 0$, $p_i \neq 0$, *then for the Fletcher–Reeves algorithm*

$$\|p_{i+1}\| \leqslant (1 + M/m)(2 + M/m)\|p_i\|.$$

Theorem 4.3. *Assume that* $p_i \neq 0$. *Then*

$$\|b_i\| \leqslant (1 + B)(1 + M/m)$$

is true for all algorithms satisfying $\|p_n\| \leqslant B\|\nabla F(x_n)\|$, *where* $B > 1$.

Theorem 4.4. *If* $\mathrm{d}F(x_i) \neq 0$, $p_i \neq 0$, *then we have*

(1) $\|b_i\| \leqslant M(m + M)/m^2$ *for the Polak–Ribière algorithm*,
(2) $\|b_i\| \leqslant (M/m)(1 + M/m)$ *for Daniel's algorithm*,
(3) $\|b_i\| \leqslant (1 + M/m)^2$ *for the Fletcher–Reeves algorithm.*

Theorem 4.5. *If* $\mathrm{d}F(x_i) \neq 0$, $p_i \neq 0$, *then*

(1) $m^2/[M(m + M)^2] \leqslant a_i \leqslant 1/m$ *for the Polak–Ribière algorithm*,
(2) $m^2/M^3 \leqslant a_i \leqslant 1/m$ *for Daniel's algorithm, and*
(3) $a_i \leqslant 1/m$ *for the Fletcher–Reeves algorithm.*

Theorem 4.6. *If* $\mathrm{d}F(x_i) \neq 0$ *and* x_i *is in* C, *then* $F(x_{i+1}) < F(x_i)$.

5. Convergence of conjugate gradient algorithms

Throughout this section we want to emphasize that for all n, $\mathrm{d}F(x_n) \neq 0$, $p_n \neq 0$, and $x_{n+1} \neq x_n$ hold in addition to the other assumptions of Section 3.

Lemma 5.1. *Let* $\{x_n\}$ *be a sequence generated by an algorithm from the class we consider in Section* 2 *under the assumptions given in Section* 3. *Suppose that* $\|p_n\| \leqslant B\|\nabla F(x_n)\|$

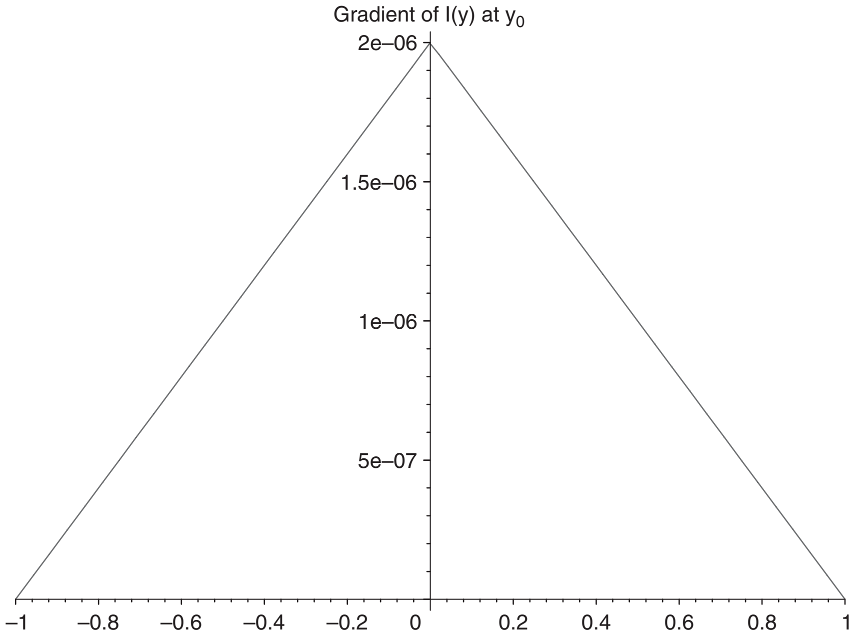

Fig. 1.

for some constant $B > 1$. Then there is a positive constant K such that

$$F(x_n) - F(x_{n+1}) \geqslant K\|\nabla F(x_n)\|^2$$

for each n. In fact, K can be chosen by

$$K = a\left(1 - \frac{aMB}{2}\right), \quad \textit{where } 0 < a < \frac{2}{MB}.$$

Lemma 5.2. *Under the assumptions of Lemma* 5.1 *of this paper, $\{F(x_n)\}$ converges, $\{x \in C : F(x) < F(x_0)\}$ is a bounded set, $\|x_{n+k} - x_n\|$ is bounded for all n and k, and we have*

$$(\mathrm{d}F(x_{n+k}) - \mathrm{d}F(x_n))(x_{n+k} - x_n) \geqslant m\|x_{n+k} - x_n\|^2.$$

Hence, $\{x_n\}$ is a Cauchy sequence in C.

Theorem 5.1. *Let $\{x_n\}$* and *$\{p_n\}$ be generated by a conjugate gradient algorithm from the class of algorithms for which*

$$\|p_n\| \leqslant B\|\nabla F(x_n)\|$$

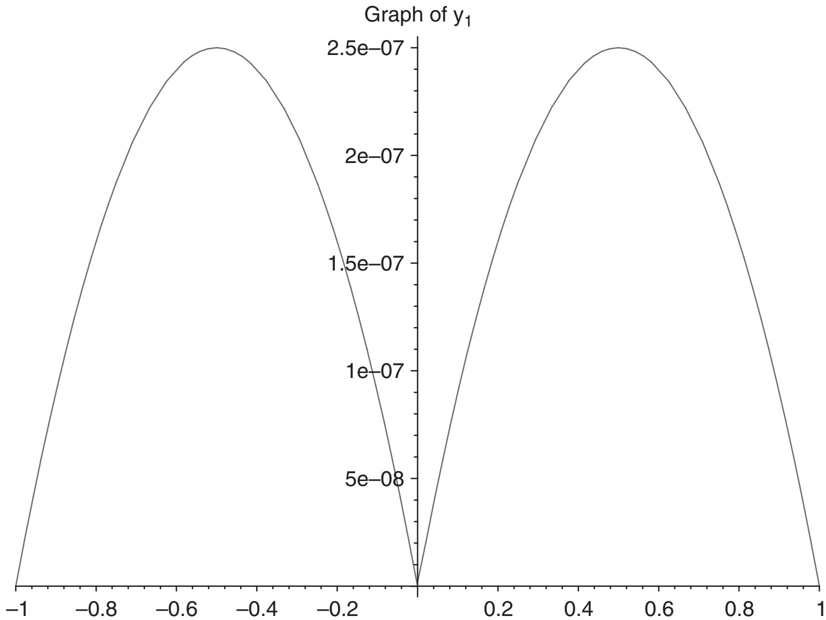

Fig. 2.

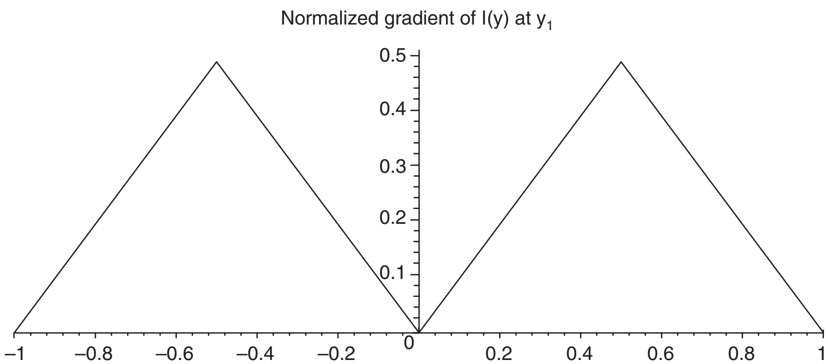

Fig. 3.

for some $B > 1$. *With the notation and assumptions of Sections* 1, 2, *and* 3, *it follows that* (i) $\mathrm{d}F(x_n) \to 0$, (ii) $p_n \to 0$, (iii) *there is a vector* x^* *in* C *such that* $x_n \to x^*$, (iv) $F(x_n) \to F(x^*)$, (v) $\mathrm{d}F(x_n) \to \mathrm{d}F(x^*)$, (vi) $\mathrm{d}F(x^*) = 0$. *Furthermore, if* x^* *is in the interior of* C, *then it is the unique local minimum for* F *on the interior of* C.

6. Numerical computations

The arc length functional provides an interesting example. Consider the problem of minimizing $I(y) = \int_{-1}^{1} \sqrt{(1 + (y'(t))^2)}\, dt$, $y(-1) = 0$, $y(1) = 0$, with y continuous on $[-1, 1]$ and y' piecewise continuous on $[-1, 1]$. Although the norm suggested by Golomb and Tapia [5, pp. 123–124] may not be strictly convex and norm reflexive, we use it to illustrate the procedure. With starting value $y_0(t) = -10^{-6}(t-1)(t+1)$ we obtain Figs. 1–3.

References

[1] R.H. Byrd, R.A. Tapia, An extension of Curry's theorem to steepest descent in normed linear spaces, Math. Programming 9 (2) (1975) 247–254.

[2] J.W. Daniel, The Approximate Minimization of Functionals, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1971.

[3] J.W. Daniel, The conjugate gradient method for linear and nonlinear operator equations, Ph.D. Thesis, Stanford University, Palo Alto, CA, 1965.

[4] R.M. Fletcher, C.M. Reeves, Function minimization by conjugate gradients, Comput. J. 7 (1964) 149–154.

[5] M. Golomb, R.A. Tapia, The metric gradient in normed linear spaces, Numer. Math. 20 (1972) 115–124.

[6] R.M. Hayes, Iterative methods for solving linear problems in Hilbert space, Ph.D. Thesis, University of California at Los Angeles, 1952.

[7] R.M. Hayes, Iterative methods for solving linear problems in Hilbert space, in: O. Tausky (Ed.), Contribution to the Solutions of Linear Systems and the Determination of Eigenvalues, Applied Mathematics Series, vol. 39, National Bureau of Standards, Washington, DC, 1954, pp. 71–103.

[8] M.R. Hestenes, Applications to the theory of quadratic forms in Hilbert space to the calculus of variations, Pacific J. Math. 1 (1951) 525–581.

[9] M.R. Hestenes, E. Stiefel, Methods of conjugate gradients for solving linear systems, J. Res. Nat. Bur. Standards 49 (6) (1952) 409–439.

[10] J.P. Penot, On the convergence of descent algorithms, Comput. Optim. Appl. 23 (3) (2002) 279–284.

[11] E. Polak, Computational Methods in Optimization—a Unified Approach, Academic Press, New York and London, 1971.

[12] E. Polak, G. Ribière, Note sur la convergence de méthodes de directions conjuguées, Rev. Française Informat. Recherche Opérationnelle 3 (16) (1969) 35–43.

[13] M.M. Vainberg, Variational Methods for the Study of Nonlinear Operators, Holden Day, San Francisco, 1964.