Computing 49, 373–376 (1993)

Computing

Short Communication/Kurze Mitteilung

A Step Size Rule for Unconstrained Optimization

J. Rohn, Prague

Received March 9, 1992; revised version July 13, 1992

Abstract — Zusammenfassung

A Step Size Rule for Unconstrained Optimization. We describe a step size rule for unconstrained optimization. The rule is proved to be finite and to perform the exact line search in one iteration in case of a strictly convex quadratic function.

AMS(MOS) Subject Classifications: 49M07, 90C30

Key words: Unconstrained optimization, step size rule, convergence.

Ein Schrittweitenalgorithmus für unrestringierte Optimierung. Wir beschreiben einen Schrittweitenalgorithmus für Lösung unrestringierter Optimierungsprobleme, der im Falle einer streng konvexen quadratischen Funktion die exakte Schrittweite in einer Iteration liefert.

Gradient methods for solving an unconstrained optimization problem

$$\min\{f(x); x \in R^n\}$$

as the steepest descent method or the methods by Fletcher-Reeves, Polak-Ribiére, Davidon-Fletcher-Powell or Broyden-Fletcher-Goldfarb-Shanno described e.g. in Fletcher [1], Luenberger [2] or Polak [3], construct a sequence of iterations $\{x_i\}$ according to the following general scheme (which we call the "main algorithm" to distinguish it from its specifications; we denote $g_i = \nabla f(x_i)$, the gradient of f at x_i):

Main algorithm.
Step 0. Select an $x_0 \in R^n$ and set $i := 0$.
Step 1. If $g_i = 0$, terminate: x_i is a stationary point of f.
Step 2. Otherwise find a search direction d_i such that $d_i^T g_i < 0$.
Step 3. Find a nonnegative real number α_i satisfying

$$f(x_i + \alpha_i d_i) = \min\{f(x_i + \alpha d_i); \alpha \geq 0\}.$$

Step 4. Set $x_{i+1} := x_i + \alpha_i d_i$, $i := i + 1$ and go to Step 1.

The methods listed above differ from each other only in the choice of the search direction d_i in Step 2. The computation of the step size α_i required in Step 3 cannot be performed exactly in a finite number of steps in general case and therefore must be replaced by some inexact line search procedure in practice; several such standard procedures are described in [1], [2], [3]. The purpose of this note is to propose another step size rule which runs as follows:

Step Size Rule (to replace Step 3 of the main algorithm).
Step 3.1. Set $\beta_0 := 1$ and $j := 0$.
Step 3.2. Compute $\gamma_j = f(x_i + \beta_j d_i) - f(x_i) - \beta_j d_i^T g_i$.
Step 3.3. If $\gamma_j \leq 0$, set $\alpha_i := \beta_j$ and go to Step 4.
Step 3.4. Otherwise compute $\beta_{j+1} = -\dfrac{\beta_j^2}{2\gamma_j} d_i^T g_i$.
Step 3.5. If $\dfrac{\beta_j}{\beta_{j+1}} < 2$, set $\alpha_i := \beta_j$ and go to Step 4.
Step 3.6. Otherwise set $j := j + 1$ and go to Step 3.2.

For each i, let us denote by j_i the index j for which $\alpha_i := \beta_j$ is set in Step 3.3 or Step 3.5. The basic properties of the rule are summed up in the following theorem:

Theorem 1. *Let $f \in C^1$. Then the step size rule is finite and the main algorithm using this rule generates a sequence of points satisfying*

$$f(x_{i+1}) - f(x_i) \leq \alpha_i d_i^T g_i \tag{1}$$

if the rule stopped in Step 3.3 and

$$\alpha_i d_i^T g_i < f(x_{i+1}) - f(x_i) = \left(\alpha_i - \frac{\alpha_i^2}{2\beta_{j_i+1}}\right) d_i^T g_i \tag{2}$$

if it stopped in Step 3.5. In particular, the sequence $\{f(x_i)\}$ is strictly decreasing. Moreover, if f is a strictly convex quadratic function, then the line search 3.1–3.6 is exact and $\alpha_i = \beta_1$ for each i.

Proof. First assume to the contrary that the rule does not terminate for some i, so that it constructs an infinite sequence $\{\beta_j\}_{j=0}^{\infty}$. Then from Steps 3.4 and 3.6 we obtain that

$$0 < \beta_{j+1} \leq \tfrac{1}{2}\beta_j$$

holds for each j, implying $\beta_j \to 0$. From Steps 3.2 and 3.4 we have

$$\gamma_j = f(x_i + \beta_j d_i) - f(x_i) - \beta_j d_i^T g_i = -\frac{\beta_j^2}{2\beta_{j+1}} d_i^T g_i,$$

hence

$$f(x_i + \beta_j d_i) - f(x_i) = \left(\beta_j - \frac{\beta_j^2}{2\beta_{j+1}}\right) d_i^T g_i \tag{3}$$

which gives

$$\frac{f(x_i + \beta_j d_i) - f(x_i)}{\beta_j} = \left(1 - \frac{\beta_j}{2\beta_{j+1}}\right) d_i^T g_i$$

for each j. Since the left-hand side tends to $d_i^T g_i$ as j approaches infinity, we obtain

$$\lim_{j \to \infty} \frac{\beta_j}{\beta_{j+1}} = 0,$$

but this is a contradiction since $\frac{\beta_j}{\beta_{j+1}} \geq 2$ for each j in view of Step 3.6. Hence the rule is finite and $\alpha_i := \beta_{j_i}$ is set for some j_i. If it stopped in Step 3.3, then we have (1), and if it terminated in Step 3.5, then (2) holds because of (3) and the fact that $\gamma_{j_i} > 0$. Since $d_i^T g_i < 0$, $\alpha_i > 0$ and $\alpha_i - \frac{\alpha_i^2}{2\beta_{j_i+1}} > 0$ in Step 3.5, both (1) and (2) imply $f(x_{i+1}) < f(x_i)$, hence the sequence $\{f(x_i)\}$ is strictly decreasing. Finally, let f be of the form

$$f(x) = \tfrac{1}{2}x^T Cx + b^T x + a$$

where C is a symmetric positive definite matrix. Then for each real β we have

$$f(x_i + \beta d_i) = f(x_i) + \beta d_i^T g_i + \tfrac{1}{2}\beta^2 d_i^T C d_i,$$

hence

$$\gamma_j = \tfrac{1}{2}\beta_j^2 d_i^T C d_i > 0$$

for each j and consequently in Step 3.4 we obtain

$$\beta_{j+1} = -\frac{d_i^T g_i}{d_i^T C d_i},$$

which is independent of j and obviously equal to the exact minimizer of $f(x_i + \beta d_i)$ over the nonnegative half-ray. Hence $\beta_2 = \beta_1$, so that $\alpha_i := \beta_1$ is set in Step 3.5. ■

Next we have this convergence theorem:

Theorem 2. *Let $f \in C^1$ and let the sequence generated by the main algorithm using the step size rule have the property $x_{i+1} - x_i \to 0$. Then each accumulation point (x_*, d_*) of the sequence $\{(x_i, d_i)\}$ satisfies*

$$d_*^T \nabla f(x_*) = 0. \tag{4}$$

Proof. Assume to the contrary that $d_*^T \nabla f(x_*) \neq 0$. Let $x_i \xrightarrow{K} x_*$, $d_i \xrightarrow{K} d_*$ along some subsequence $K \subset \{0, 1, 2, \ldots\}$ which may be chosen so that $\{\alpha_i\}_{i \in K}$ converges to some α. If $\alpha > 0$, then from $\alpha_i d_i = x_{i+1} - x_i \to 0$ we obtain $d_* = 0$, hence (4) holds. Thus assume that $\alpha = 0$. Then $j_i \geq 1$ from some i on and from Step 3.4 we have

$$\alpha_i = \beta_{j_i} = -\frac{\beta_{j_i-1}^2}{2\gamma_{j_i-1}} d_i^T g_i.$$

Here $\alpha_i \xrightarrow{K} 0$, $\{\gamma_{j_i-1}\}_K$ is bounded and $\{d_i^T g_i\}_K$ has a nonzero limit, hence $\beta_{j_i-1} \xrightarrow{K} 0$. Now, Theorem 1 gives

$$f(x_i + \alpha_i d_i) - f(x_i) < 0,$$

but from Step 3.6 we have

$$f(x_i + \beta_{j_i-1} d_i) - f(x_i) = \left(\beta_{j_i-1} - \frac{\beta_{j_i-1}^2}{2\beta_{j_i}}\right) d_i^T g_i \geq 0,$$

which together gives

$$f(x_i + \alpha_i' d_i) - f(x_i) = 0$$

for some $\alpha_i' \in [\alpha_i, \beta_{j_i-1}]$ and by the mean-value theorem we have

$$d_i^T \nabla f(x_i + \xi_i d_i) = 0$$

for some $\xi_i \in [0, \alpha_i']$. Since $\xi_i \xrightarrow{K} 0$ because of $\beta_{j_i-1} \xrightarrow{K} 0$, taking the limit we obtain $d_*^T \nabla f(x_*) = 0$ contrary to the assumption. This concludes the proof.

For the steepest descent method, where $d_i = -g_i$ is set for each i, we immediately obtain from (4) that $d_*^T \nabla f(x_*) = -\|\nabla f(x_*)\|_2^2 = 0$, hence x_* is a stationary point. Convergence properties of some other methods endowed with this step size rule are given in [4]. A limited computational experience shows that the rule performs best when implemented into the DFP or BFGS method.

References

[1] Fletcher, R.: Practical methods of optimization (Volume 1: Unconstrained optimization). New York: Wiley and Sons 1980.

[2] Luenberger, D. A.: Introduction to linear and nonlinear programming. Reading, Massachusetts: Addison-Wesley 1973.

[3] Polak, E.: Computational methods in optimization: a unified approach. New York: Academic Press 1971.

[4] Rohn, J.: Step size rule for unconstrained optimization. Report NI-92-04, Institute for Numerical Analysis, The Technical University of Denmark, Lyngby 1992.

Jiri Rohn
Faculty of Mathematics and Physics
Charles University
Malostranské nám. 25
11800 Prague
Czechoslovakia

Offenlegung gem. § 25 Abs. 1 bis 3 Mediengesetz:
Unternehmensgegenstand: Verlag von wissenschaftlichen Büchern und Zeitschriften

An den Springer-Verlag KG ist beteiligt: Dr. Konrad F. Springer, Sachsenplatz 4–6, A-1201 Wien, als Kommanditist zu 52,38%. Geschäftsführer: Dr. Konrad F. Springer, Prof. Dr. Dietrich Götze, Ing. Wolfram F. Joos, Dipl.-Kfm. Claus Michaletz und Rudolf Siegle, alle Sachsenplatz 4–6, A-1201 Wien.

Verleger: Springer-Verlag KG, Sachsenplatz 4–6, A-1201 Wien. – Herausgeber: Prof. Dr. Hans J. Stetter, Institut für Angewandte und Numerische Mathematik der Technischen Universität Wien, Wiedner Hauptstraße 6–10, A-1040 Wien. – Redaktion: Wiedner Hauptstraße 6–10, A-1040 Wien. – Satz und Umbruch: Asco Trade Typesetting Limited, Hong Kong; Reproduktion und Offsetdruck: Novographic, Ing. W. Schmid, Maurer-Lange-Gasse 64, A-1238 Wien. – Verlagsort: Wien. – Herstellungsort: Wien. – Printed in Austria.