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

\[ \min \{ f(x) ; x \in \mathbb{R}^n \} \]

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 \( x_0 \in \mathbb{R}^n \) 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 \( \alpha_i \) 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 \( \alpha_i \) 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 $\beta_{j+1} = \frac{-\beta_j^2}{2\gamma_j} d_i^T g_i$.

**Step 3.5.** If $\frac{\beta_j}{\beta_{j+1}} < 2$, 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

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

if the rule stopped in Step 3.3 and

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

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

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

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

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

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

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