Revista de

Matemáticas Aplicadas

Volume 10 Number 2 July 1989

CONTENTS

ISSN 0716-5803
Universidad de Chile
Facultad de Ciencias Físicas y Matemáticas
Departamento de Ingeniería Matemática

Rev. Mat. Apl. **10**:71-79

Revista de Matemáticas Aplicadas

Departamento de Ingeniería Matemática

CONVERGENCE OF CURVILINEAR SEARCH ALGORITHMS TO SECOND ORDER POINTS

J. AMAYA[1-2]

[1]CORE, Université Catholique de Louvain,
Voie du Roman Pays 34
1348 Louvain-la-Neuve, BELGIUM
[2]Departamento de Ingeniería Matemática
Facultad de Ciencias Físicas y Matemáticas, U. de Chile
Casilla 170 Correo 3
Santiago - Chile

Abstract

The purpose of this paper is to establish, for curvilinear algorithms, general hypotheses under which convergence occurs to points satisfying second order necessary conditions for minimality. A gradient path algorithm and a negative curvature strategy are both examined in this context.

AMS Classifications: 65K05, 49D37, 90C30.

Key Words: unconstrained optimization, second order conditions, curvilinear algorithms, negative curvature.

1.- Introduction

Most unconstrained mathematical programming algorithms are such that all a user can be assured of is that they will converge to stationary points. Besides these, however, there exist algorithms aimed at approximating stationary points satisfying second order necessary minimality conditions as well.

Attempts to devise algorithms of the latter type can be found in several papers, e.g., McCormick [6], Moré and Sorensen [7] and Goldfarb [5], who have all used the notion of negative curvature. Apart from their results, gradient path algorithms ([4], [8]) have also been adapted to approximate points of such nature [3].

In this work, hypotheses are established under which the stationary points to which general curvilinear search algorithms will converge are guaranted to also satisfy second order necessary conditions to be local minima. We assume that the trajectory, where the new iterate will be chosen at each iteration, is sufficiently smooth.

2.- The algorithm.

Let $f : R^n \to R$ be a class-C^2 function which we want to minimize, and $d : R^+ \times R^n \to R^n$ a function such that, for all $x \in R^n, d(0,x) = x$. We suppose that, for every $x \in R^n$, the trajectory $d(t,x)$ is C^2 in $t \geq 0$. This is the case, for example, of linear combinations of descent and/or negative curvature directions, with twice continuously differentiable coefficients, as shown in section 5.

Given x in R^n, the function $d(t,x)$ describes a trajectory in R^n originating at x. The minimization algorithm gives rise to a sequence $\{x_k\}$ in R^n through

$$x_{k+1} = \begin{cases} x_k & \text{if } x_k \in M \\ d(t_k, x_k) & \text{if } x_k \notin M \end{cases} \tag{2.1}$$

where $M = \{x \in R^n / \nabla f(x) = 0 \text{ and } < H(x)h, h > \geq 0, \quad h \in R^n\}$, ∇f and H denoting the gradient vector and the hessian matrix of f, respectively.

The step size is determined in accordance with a strategy of the Armijo type ([1], [2], [6]) as follows: for $x \in R^n$, we define the class-C^2 function $\varphi_x : R^+ \to R$ by

$$\varphi_x(t) = f(d(t,x)) \qquad t \in R^+ \tag{2.2}$$

This function is shown to satisfy

$$\varphi'_{x_k}(0) = < g_k, \dot{d}(0, x_k) > \qquad \text{and}$$

$$\varphi''_{x_k}(0) = < H_k \dot{d}(0,x_k), \dot{d}(0,x_k) > + < g_k, \ddot{d}(0,x_k) >,$$

where g_k and H_k respectively denote the gradient and the hessian matrix of f, both evaluated at x. $\dot{d}$ and $\ddot{d}$ denotes respectively the first and second derivatives of d with respect to t.

Let $\delta, \beta \in]0,1[$ and $\alpha_k > 0$ be given, and be Δ_k denotes the $\min\{0, \varphi''_{x_k}(0)\}$. The stepsize t_k is then defined by the following rule:

$$\text{if} \quad \varphi_{x_k}(\alpha_k) > \varphi_{x_k}(0) + \delta \alpha_k \varphi'_{x_k}(0) + \delta \frac{\alpha_k^2}{2} \Delta_k, \text{ then}$$

$$t_k = \max\{\lambda = \alpha_k \beta^s / s = 0, 1, 2, ...; \varphi_{x_k}(\lambda) \leq \varphi_{x_k}(0) + \delta\lambda\varphi'_{x_k}(0) + \delta\frac{\lambda^2}{2}\Delta_k\}. \tag{2.3}$$

Otherwise

$$t_k = \beta \min\{\lambda = \frac{\alpha_k}{\beta^s} / s = 0, 1, 2, ...; \varphi_{x_k}(\lambda) > \varphi_{x_k}(0) + \delta\lambda\varphi'_{x_k}(0) + \delta\frac{\lambda^2}{2}\Delta_k\}. \tag{2.4}$$

This procedure guarantees that the interval $[t_k, t_k/\beta]$ contains at least one value $\bar{t}_k$ satisfying

$$\varphi_{x_k}(\bar{t}_k) = \varphi_{x_k}(0) + \delta\bar{t}_k\varphi'_{x_k}(0) + \delta\frac{\bar{t}_k^2}{2}\Delta_k.$$

We shall make use of this fact in the next section to prove the convergence of algorithm (2.1) without resorting to additional hypotheses on the sequence $\{\alpha_k\}$.

3.- Convergence of the algorithm.

It is our objective to have algorithm (2.1) approximate points of M, in view of which we set down the following assumptions:

$$L = \{x \in R^n / f(x) \leq f(x_0)\} \text{ is compact}, \tag{3.1}$$

$$\varphi'_x(0) \leq 0 \quad \text{for all } x \notin M, \tag{3.2}$$

$$\text{if } x \notin M \text{ and } \varphi'_x(0) = 0, \text{ then } \varphi''_x(0) < 0. \tag{3.3}$$

The first assumption is standard. The second one says that the trajectory $d(t, x)$ emerges from x in a nonincreasing direction. The third assumption guarants that, if the first derivative of φ_x is zero, the trayectory emerges from x in a negative curvature direction.

The next lemma shows that the stepsize t_k is well defined.

Lemma 3.1. Let $x_k \notin M$ and δ, β, α_k and Δ_k as defined before. Under assumptions (3.1), (3.2) and (3.3), the stepsize t_k defined by (2.3) and (2.4) exists.

Proof: Let us suppose that t_k does not exists. There are two cases:

i) For all $\quad \lambda = \alpha_k \beta^s, s = 0, 1, 2, ...$

$$\varphi_{x_k}(\lambda) > \varphi_{x_k}(0) + \delta\lambda\varphi'_{x_k}(0) + \delta\frac{\lambda^2}{2}\Delta_k. \tag{3.4}$$

If $\varphi'_{x_k}(0) = 0$ then from (3.4) we have

$$\frac{\varphi_{x_k}(\lambda) - \varphi_{x_k}(0)}{\lambda^2} > \frac{\delta}{2}\Delta_k$$

and letting $s \to \infty$ we find

$$\frac{1}{2}\varphi''_{x_k}(0) \geq \frac{\delta}{2}\Delta_k.$$

But, in view of (3.3), we have $\Delta_k = \varphi''_{x_k}(0) < 0$, which contradicts the choice of δ.

If $\varphi'_{x_k}(0) < 0$ then from (3.4) we have

$$\frac{\varphi_{x_k}(\lambda) - \varphi_{x_k}(0)}{\lambda} > \delta\varphi'_{x_k}(0) + \delta\frac{\lambda}{2}\Delta_k$$

and letting $s \to \infty$, we find

$$\varphi'_{x_k}(0) \geq \delta\varphi'_{x_k}(0)$$

which contradicts the choice of δ.

(ii) For all $\lambda = \frac{\alpha_k}{\beta^s}$, $s = 0, 1, 2, \dots$

$$\varphi_{x_k}(\lambda) \leq \varphi_{x_k}(0) + \delta\lambda\varphi'_{x_k}(0) + \delta\frac{\lambda^2}{2}\Delta_k. \tag{3.5}$$

The right hand side of (3.5) is obviously decreasing and unbounded as $\lambda \to \infty$. Then, we have $\varphi_{x_k}(\lambda) \to -\infty$ as $s \to \infty$, which contradicts the boundedness of f, which arises from assumption (3.1). ■

The next theorem establishes the convergence of the algorithm.

Theorem 3.1. *Under assumptions (3.1), (3.2) and (3.3), every adherent point (and there exists one) of the sequence $\{x_k\}$ generated by algorithm (2.1) belongs to M.*

Proof: *Let us suppose that $x_k \notin M$ for $k = 0, 1, 2, \dots$ From the choice of t_k we have $\varphi_{x_k}(t_k) \leq \varphi_{x_k}(0)$, i.e., the sequence $\{f(x_k)\}$ is decreasing, so $\{x_k\} \subset L$. Due to (3.1), the sequence $\{x_k\}$ has an adherent point $\bar{x}$. Let $\{x_{k_j}\}$ be a subsequence converging to $\bar{x}$ and let us suppose that $\bar{x} \notin M$.*

From (2.3) and (2.4) we find that

$$\varphi_{x_{k_j}}(t_{k_j}/\beta) > \varphi_{x_{k_j}}(0) + \delta\frac{t_{k_j}}{\beta}\varphi'_{x_{k_j}}(0) + \delta\frac{t_{k_j}^2}{2\beta^2}\Delta_{k_j} \tag{3.6}$$

Besides, from Taylor's theorem we know that there exists a value $\theta_{k_j} \in [0, t_{k_j}/\beta]$ such that

$$\varphi_{x_{k_j}}(t_{k_j}/\beta = \varphi_{x_{k_j}}(0) + \frac{t_{k_j}}{\beta}\varphi'_{x_{k_j}}(0) + \frac{t_{k_j}^2}{2\beta^2}\varphi''_{x_{k_j}}(\theta_{k_j}),$$

which takes us to

$$(1-\delta)\varphi'_{x_{k_j}}(0) + \frac{t_{k_j}}{2\beta}(\varphi''_{x_{k_j}}(\theta_{k_j}) - \delta\Delta_{k_j}) > 0 \tag{3.7}$$

when substituted in (3.6).

We now analize the cases that may occur.

A) If $\{t_{k_j}\}$ is not bounded away from zero, without loss of generality we can assume that $t_{k_j} \to 0$ as $j \to \infty$. Then $\theta_{k_j} \to 0$ and by continuity and other familiar arguments we trivially see that:

$$\varphi''_{x_{k_j}}(\theta_{k_j}) \to \varphi''_{\bar{x}}(0),$$

$$\varphi''_{x_{k_j}}(0) \to \varphi''_{\bar{x}}(0),$$

$$\varphi'_{x_{k_j}}(0) \to \varphi'_{\bar{x}}(0) \qquad \text{and}$$

$$\Delta_{k_j} \to \bar{\Delta} = \min\{0, \varphi''_{\bar{x}}(0)\}.$$

Thus, from (3.7) we have

$$\varphi'_{\bar{x}}(0) \geq 0. \tag{3.8}$$

Given that $\bar{x} \notin M$, by (3.2) and (3.3) the expression (3.8) implies

$$\varphi'_{\bar{x}}(0) = 0 \tag{3.9}$$

and

$$\varphi''_{\bar{x}}(0) < 0. \tag{3.10}$$

Let us return to (3.7). Given that $x_{k_j} \notin M$, by (3.2) we have $\varphi''_{k_j}(\theta_{k_j}) - \delta\Delta_{k_j} > 0$, which entails

$$\varphi''_{\bar{x}}(0) - \delta\bar{\Delta} > 0. \tag{3.11}$$

On the other hand, from (3.10) and the definition of $\bar{\Delta}$ we have $\bar{\Delta} = \varphi''_{\bar{x}}(0)$. Hence from (3.11) we get $(1-\delta)\varphi''_{\bar{x}}(0) \geq 0$, which implies $\varphi''_{\bar{x}}(0) \geq 0$. This is a contradiction with (3.10).

B) If $\{t_{k_j}\}$ is bounded away from zero, without loss of generality we assume that there exists $\tau > 0$ such that

$$t_{k_j} \geq \tau \quad \text{for all} \quad j = 0, 1, 2, \ldots \tag{3.12}$$

From (2.3) and (2.4) we have, for all $k = 0, 1, 2, \ldots$

$$\varphi_{x_k}(t_k) \leq \varphi_{x_k}(0) + \delta t_k \varphi'_{x_k}(0) + \delta \frac{t_k^2}{2} \Delta_k.$$

That is, in particular, by considering $f(x_{k_{j+1}}) \leq f(x_{k_j+1})$, we have

$$f(x_{k_{j+1}}) - f(x_{k_j}) \leq \delta t_{k_j} \varphi'_{x_{k_j}}(0) + \delta \frac{t_{k_j}^2}{2} \Delta_{k_j}. \tag{3.13}$$

From (3.2) and the definition of Δ_{k_j}, the right hand side of (3.13) is nonpositive. So, by summing from $j = 0$ to $j = N$ in (3.13) and combining with (3.12) we get

$$f(x_{k_{N+1}}) - f(x_{k_0}) \leq \delta \tau \sum_{j=0}^{N} \varphi'_{x_{k_j}}(0) + \delta \frac{\tau^2}{2} \sum_{j=0}^{N} \Delta_{k_j} \tag{3.14}$$

The series $\sum \varphi'_{x_{k_j}}(0)$ and $\sum \Delta_{k_j}$ are convergent, because if either diverges to $-\infty$, then $f(x_{k_{N+1}})$ diverges to $-\infty$, as $N \to \infty$, which contradicts the boudedness of f in L.

Hence we necessarily have $\varphi'_{x_{k_j}}(0) \to 0$ and $\Delta_{k_j} \to 0$ as $j \to \infty$. Thus $\varphi'_{\bar{x}}(0) = 0$ and $\bar{\Delta} = 0$. By the definition of $\bar{\Delta}$, we finally have $\varphi''_{\bar{x}}(0) \geq 0$, which contradicts (3.3).

From (A) and (B) we conclude that $\bar{x} \in M$. ■

4.- A gradient path type algorithm.

Theorem 3.1 is used here to show that a modification proposed by Auslender [3] of the gradient path local approximation ([4], [8]) converges to a point in M. His strategy consist of adding, to the trajectory, a non-linear term emerging in a direction of negative curvature. Such term depends on the eigenvectors associated to the negative eigenvalues of the hessian matrix. The resulting trajectory is given by

$$d(t, x_k) = x_k + \sum_{\lambda_i \neq 0} \frac{e^{-t\lambda_i} - 1}{\lambda_i} < u_i, g_k > u_i - t \sum_{\lambda_i = 0} < u_i, g_k > u_i +$$

$$+ \sum_{\lambda_i < 0} \frac{e^{-t\lambda_i} - 1}{\lambda_i} \sigma_i u_i \tag{4.1}$$

where the hessian matrix H_k is decomposed in the spectral form

$$H_k = U_k D_k U_k^T = [u_1, ..., u_n] \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \begin{bmatrix} u_1^T \\ \vdots \\ u_n^T \end{bmatrix} \tag{4.2}$$

and

$$\sigma_i = \begin{cases} 0 & \text{if} \quad \lambda_i \geq 0 \\ 1 & \text{if} \quad \lambda_i < 0 \text{ and } < u_i, g_k >> 0 \\ -1 & \text{if} \quad \lambda_i < 0 \text{ and } < u_i, g_k > \leq 0 \end{cases} \tag{4.3}$$

It is clear that $d(., x_k)$ is of class-C^2 and

$$\begin{aligned} \dot{d}(0, x_k) &= -g_k - U_k \sigma, \qquad \text{and} \\ \ddot{d}(0, x_k) &= H_k g_k + D_k U_k \sigma \end{aligned} \tag{4.4}$$

where $\sigma = (\sigma_1, ..., \sigma_n)^T$.

From (4.4) we have

$$\varphi'_{x_k}(0) = -\|g_k\|^2 - \sigma^T U_k^T g_k, \qquad \text{and}$$

$$\varphi''_{x_k}(0) = 2g_k^T H_k g_k + 3\sigma^T D_k U_k^T g_k + \sigma^T D_k \sigma \tag{4.5}$$

From (4.3) it is clear that $\sigma^T U_k^T g_k \geq 0$ and so if $x_k \notin M$, we have $\varphi'_{x_k}(0) \leq 0$, which shows that (3.2) holds.

On the other hand, if $x_k \notin M$ and $\varphi'_{x_k}(0) = 0$, then from (4.5), $g_k = 0$, and therefore

$$\varphi''_{x_k}(0) = \sigma^T D_k \sigma = \sum_{\lambda_i < 0} \lambda_i < 0$$

Hence (3.3) also holds.

Thus the algorithm defined by trajectory (4.1) is shown to converge in accordance with theorem 3.1.

5.- A negative curvature strategy.

We can use theorem 3.1 to prove convergence in the case of a trajectory proposed by Moré and Sorensen [7], which is given by

$$d(t, x_k) = x_k + t^2 s_k + t r_k \tag{5.1}$$

where (s_k, r_k) is a descent pair.

A descent pair satisfies

$$< g_k, s_k >< 0 \tag{5.2}$$

$$< g_k, r_k > \leq 0 \tag{5.3}$$

and

$$r_k^T H_k r_k = 0 \tag{5.4}$$

if H_k is positive semidefinite, and

$$< g_k, s_k > \leq 0 \tag{5.5}$$

$$< g_k, r_k > \leq 0 \tag{5.6}$$

and

$$r_k^T H_k r_k < 0 \tag{5.7}$$

if H_k is not positive semidefinite.

The only case where a descent pair does not exists occurs when $x_k \in M$.

For trajectory (5.1) we have

$$\varphi'_{x_k}(0) = < g_k, r_k > \text{ and} \tag{5.8}$$

$$\varphi''_{x_k}(0) = r_k^T H_k r_k + 2 < g_k, s_k > \tag{5.9}$$

From (5.3) and (5.6), (3.2) is trivially verified.

Let us examine assumption (3.3). Assuming $\varphi'_{x_k}(0) = 0$, we have two cases:

(a) If H_k is positive semidefinite, by applying (5.2) and (5.4), to the expression (5.9), we obtain $\varphi''_{x_k}(0) < 0$.

(b) If H_k is not positive semidefinite, by applying (5.5) and (5.7), to the expression (5.9), we obtain $\varphi''_{x_k}(0) < 0$.

The algorithm defined by curve (5.1) hence converges to a point belonging to M.

References

[1] Amaya, J., On the Convergence of Curvilinear Search Algorithms in Unconstrained Optimization, *Operation Research Letters*, **4(1)** (1985).

[2] Armijo, L., Minimization of Functions Having Lipschitz Continuous Partial Derivatives, *Pacific Journal of Mathematics*, **16** (1966), 1-3.

[3] Auslender, A., Algorithms for Computing Points that Satisfie Second Order Necessary Conditions, *Proceedings IIASA Meeting*, Ed. by Wierzbicki, Springer Verlag, (1979).

[4] Botsaris, C. A. and D. H. Jacobson, A Newton-Type Curvilinear Search Method for Optimization, *Journal of Mathematical Analysis and Applications*, **54** (1976), 217-229.

[5] Goldfarb, D., Curvilinear Path Steplenght Algorithms for Minimization which use Directions of Negative Curvature, *Mathematical Programming*, **18** (1980), 31-40.

[6] McCormick, G., A Modification of Armijo's Step-Size Rule for Negative Curvature, *Mathematical Programming*, **13** (1977), 111-115.

[7] Moré, J. and D.C. Sorensen, On the use of Directions of Negative Curvature in a Modified Newton Method, *Mathematical Programming*, **16** (1979), 1-20.

[8] Vial, J. Ph. and I. Zang, Unconstrained Optimization by Approximation of the Gradient Path, *Mathematics of Operations Research*, **2** (1977), 253-265.

Supported by FONDECYT under Project 0316-87, TWAS under Project 87-73, DTI (U. de Chile) under Project E-2618-87 and A.G.C.D.

[Paper received January 1989]