Revista de

Matemáticas Aplicadas

ex SIGMA

Volumen 9 Número 1 Julio 1987

Editor

Servet Martínez — Universidad de Chile

Comité Editorial

Robert Anderson — Université de Montreal
Octavio Betancourt — New York University
Enrique Cabaña — Universidad Simón Bolívar
John Dennis — Rice University
Eric Goles — Universidad de Chile
Jean B. Hiriart-Urruty — Université de Toulouse
Patricio A. Laura — Inst. de Mecánica Apl. Belgrano
Alan C. Lazer — University of Miami
Ludovic Lébart — Credoc-Paris
Juan P. Milasciewicz — Universidad de Buenos Aires
Robert Miller — University of Wisconsin-Madison
Andrew M. Olson — Universidad de Puerto Rico
Rolando Rebolledo — Universidad Católica de Chile
François Robert — Université de Grenoble
Marcel P. Schutzenberger — Université de Paris VII
Florencio I. Utreras — Universidad de Chile
Pierre - Jean Laurent — Université de Grenoble
Raúl Manasevich — Universidad de Chile

ISSN 0716-3096
Departamento de Matemáticas Aplicadas
Facultad de Ciencias Físicas y Matemáticas
Universidad de Chile

Rev. Mat. Apl. 9:1-7

Revista de
Matemáticas Aplicadas
©1987 Universidad de Chile
Departamento de Matemáticas Aplicadas

A Differential Equations Approach to Function Minimization

Jorge Amaya

Abstract

In this paper we establish a differential equation whose solution is a descent curve for a certain function. The existence, uniqueness and properties of this solution are also analyzed. Finally, a generalized gradient path algorithm is proposed

Key Words: Unconstrained optimization, descent curves, gradient path

AMS Classifications: 90C30, 65K05, 49D07.

1.- Introduction

In the last years, a good volume of research has been dedicated to curvilinear search algorithms for unconstrained optimization ([1], [4], [5], [8], [9]). The origin of many of these algorithms seems to be the differential equation

$$\begin{aligned} \dot{x}(t) &= -A(x(t))\nabla f(x(t)) \qquad t \geq 0 \\ x(0) &= x_0 \end{aligned} \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a C^2 - class function and $A : \mathbb{R}^n \to \mathbb{R}^{n\times n}$ is a symmetric C^1 - class matrix.

These case $A(x) = I, \quad \forall x \in \mathbb{R}^n$, has been already analyzed and, in particular, Botsaris [4] under appropriate hypothesis is that the solution of

$$\begin{aligned} \dot{x}(t) &= -\nabla f(x(t)) \qquad t \geq 0 \\ x(0) &= x_0 \end{aligned} \tag{1.2}$$

converges to a minimizer of f, as $t \to \infty$.

From the differential equation (1.2) a curvilinear search algorithm has been proposed based on the iteration formula ([5],[8])

$$x_{k+1} = x_k + (e^{-t_k H(x_k)} - I)H(x_k)^{-1}\nabla f(x_k) \tag{1.3}$$

where the matrix $H(x_k)$ denotes the hessian matrix of f evaluated at x_k and t_k is chosen according to some stepsize strategy. In fact, by approximating the right hand side of (1.2) to the first order, we obtain

$$\begin{aligned} \dot{x}(t) &= -H(x_0)x(t) + H(x_0)x_0 - \nabla f(x_0) \\ x(0) &= x_0 \end{aligned} \tag{1.4}$$

which is a linear differential equation whose solution is of the form (1.3). For more details, see the references ([4],[5],[7],[8]).

Our interest is to analyze existence, uniqueness, and some properties of the solution of (1.1).

2.- Existence and Uniqueness of the Solution

The first question arising is if a unique solution for the differential equation (1.1) exists. To prove that it does, we will assume that $x_0 \in \mathbb{R}^n$ is chosen such that the set $L_0 = \{x \in \mathbb{R}^n / f(x) \leq f(x_0)\}$ is compact.

The following result will be needed in our analysis.

Lemma 2.1. *Let $y :]a, b[\to \mathbb{R}^n, a < b < \infty$, be a C^1 -class function such that*

$$\|\dot{y}(t)\| \leq K \qquad \forall t \in]a, b[$$

for some $K > 0$. Then, $\lim_{t \to a^+} y(t)$ and $\lim_{t \to b^-} y(t)$ exist. ∎

Lef $F : \mathbb{R}^n \to \mathbb{R}^n$ be a C^1 - class function defined by

$$F(x) = -A(x)\nabla f(x) \tag{2.1}$$

and let U be a bounded open set containing L_0. The function F is Lipschitzian in U; in fact, ∇f is bounded in U. So, for every $(\bar{t}, \bar{x}) \in \mathbb{R} \times U$ there exists a neighbourhood $I(\bar{t}) \times V(\bar{x}) \subseteq \mathbb{R} \times U$ such that the equation

$$\dot{x}(t) = F(x(t)) \qquad t \geq 0 \tag{2.2}$$

posses a unique solution in $I(\bar{t})$ passing through $(\bar{t}, \bar{x})$.

Definition 2.1. *Let $I, J \subset \mathbb{R}$ be intervals and let $\psi : J \to \mathbb{R}^n$ be a solution of (2.2) on J. Let $\phi \to \mathbb{R}^n$. Then ϕ is a maximal solution of (2.2) and I is a maximal interval for (2.2) if $I \subseteq J$ and $\phi(t) = \psi(t)$ for all $t \in I$ implies $I = J$.* ∎

In the case $\bar{t} = 0$, $\bar{x} = x_0$, let $z : I(0) \to V(x_0)$ be the solution of (2.2). Let us assume that z is the maximal solution and $I(0) = [0, w_+[$, with

$w_+ < \infty$. Owing to the continuity of F and to the boundness of U, it is clear that there exists a constant $K > 0$ verifying $\|\dot{z}(t)\| \leq K$, $\forall t \in I(0)$. Then, by applying lemma 2.1 $z_+ = \lim_{t \to w_+^-} z(t)$ exists, which implies that the solution z can be continued to $[0, w_+]$.

Now, it is easy to see that there exists a local solution $\bar{z}$, for the equation

$$\begin{aligned} \dot{x}(t) &= F(x(t)) \\ x(w_+) &= z_+ \end{aligned} \tag{2.3}$$

in $]w_+ - \Delta, w_+ + \Delta[$ for some $\Delta > 0$. By uniqueness, we have $\dot{z}(t) = z(t)$, $\forall t \in]w_+ - \Delta, w_+]$, so z is not the maximal solution. In this manner, the maximal interval of the solution of (1.1) must be $[0, \infty[$. In the next section we will investigate some properties of that solution and prove that this curve converges to a stationary point of f in L_0, as $t \to \infty$.

3.- Asymptotic Behaviour of the Solution

We will assume the following hypothesis:

(H1) L_0 is convex.

(H2) There exists a pair of constants $m_1, M_1 > 0$ such that, for all $u \in L_0$

$$m_1 \|h\|^2 \leq < A(u)h, h > \leq M_1 \|h\|^2 \qquad \forall h \in \mathbb{R}^n \tag{3.1}$$

where $\| \ \|$ denotes the euclidean norm in $\mathbb{R}^n$.

(H3) There exists a constant $m_2 > 0$ such that, for all $u \in \mathbb{R}^n, y \in L_0$

$$m_2 \|h\|_u^2 \leq < H(y)h, h >_u \qquad \forall h \in \mathbb{R}^n \tag{3.2}$$

where $< v, w >_u = < A(u)v, w >$ and $\|h\|_u^2 = < h, h >_u$.

First, it is easy to see that the solution x of (1.1) satisfies $x(t) \in L_0$, $\forall t \geq 0$. In fact, x is a descent curve, i.e., the function $\rho : [0, \infty[\to \mathbb{R}$ defined by $\rho(t) = f(x(t))$, satisfies (from (H2))

$$\dot{\rho}(t) \leq -m_1 \|\nabla f(x(t))\|^2 \leq 0.$$

Thus, ρ is decreasing in $[0, \infty[$ which implies that $x(t) \in L_0$, $\forall t \geq 0$.

The next theorem establishes the relationship between the differential equation (1.1) and a stationary point of f in L_0.

Theorem 3.1. *Let x be the solution of (1.1), and assume that the hypothesis (H1), (H2) and (H3) are satisfied. Then, if x^* is the unique stationary point of f in L_0.*

$$\|x(t) - x^*\| \leq e^{-m_1 m_2 t} \|x_0 - x^*\| \tag{3.3}$$

Proof: Let σ, η_t be the functions defined by

$$\sigma(t) = < x(t) - x^*, x(t) - x^* > \qquad t \geq 0 \qquad (3.4)$$

$$\eta_t(s) = < A(x(t)) \nabla f(sx(t) + (1-s)x^*), x(t) - x^* > \qquad 0 \leq s \leq 1 \quad (3.5)$$

It is clear that

$$\dot{\sigma}(t) = -2 < A(x(t)) \nabla f(x(t)), x(t) - x^* > = -2\eta_t(1) \qquad (3.6)$$

and, by applying the mean value theorem to η_t in $[0,1]$, we have

$$\eta_t(1) = < A(x(t)) H(rx(t) + (1-r)x^*)(x(t) - x^*), x(t) - x^* > \qquad (3.7)$$

for some $r \in]0,1[$. By substituting (3.7) in (3.6)

$$\dot{\sigma}(t) = -2 < H(rx(t) + (1-r)x^*)(x(t) - x^*), x(t) - x^* >_{x(t)} \qquad (3.8)$$

for some $r \in]0,1[$.

From (H1) we have $rx(t) + (1-r)x^* \in L_0$ for all $r \in]0,1[$ and, by applying the hypothesis (H3) and (H2) to (3.8).

$$\begin{aligned} \dot{\sigma}(t) &\leq -2m_2 \|x(t) - x^*\|^2_{x(t)} \\ &\leq -2m_1 m_2 \|x(t) - x^*\|^2 = -2m_1 m_2 \sigma(t) \end{aligned}$$

which implies

$$\sigma(t) \leq e^{-2m_1 m_2 t} \sigma(0)$$

or

$$\|x(t) - x^*\| \leq e^{-m_1 m_2 t} \|x_0 - x^*\| \qquad \blacksquare$$

The above theorem shows that $x(t)$ converges to x^*, as $t \to \infty$. On the other hand, it can be proved that x^* is asymptotically stable,in the Lyapunov sense [7]. In fact, let V be an open set such that $x(t) \in V \subseteq L_0$ for all $t \geq 0$, and let $\ell : V \to I\!R$ be a function defined by $\ell(x) = f(x) - f(x^*)$. It is not hard to show that

(i) $\ell(x^*) = 0$ and $\ell(x) > 0 \qquad \forall x \in V, \qquad x \neq x^*$

(ii) $\frac{d}{dt}\ell(x(t)) < 0 \qquad \forall t \geq 0$

so, ℓ is a Lyapunov function for the equation (1.1) and x^* is an asymptotically stable point.

4.- A Generalized Gradient Path Algorithm

In the general case the solution of (1.1) cannot be found in analytic form, but an approximated scheme can be used. By expanding the right hand side of (1.1) to the first order around x_0, we obtain the linear differential equation

$$\begin{aligned} \dot{x}(t) &= -J(x_0)x(t) + J(x_0)x_0 - A(x_0)\nabla f(x_0) \\ x(0) &= x_0 \end{aligned} \tag{4.1}$$

where $J(x_0)$ denotes the Jacobian matrix of F evaluated at x_0. It is easy to see that

$$J(x_0) = \sum_{j=1}^{n} g_j(x_0)J_j(x_0) + A(x_0)H(x_0) \tag{4.2}$$

where g_j denotes the j^{th} coordinate of ∇f and J_j denotes the Jacobian matrix of the j^{th} column of A.

Under the assumption that $J(x_0)^{-1}$ exists, the solution of (4.1) is

$$x(t) = x_0 + (e^{-tJ(x_0)} - I)J(x_0)^{-1}A(x_0)\nabla f(x_0)$$

which suggest a curvilinear search algorithm based on the iteration formula

$$x_{k+1} = x_k + (e^{-t_k J_k} - I)J_k^{-1}A_k g_k \tag{4.3}$$

where $J_k = J(x_k), A_k = A(x_k), g_k = \nabla f(x_k)$ and t_k is the stepsize.

The algorithm (4.3) belongs to a family characterized by the formula

$$x^k(t) = x_k - S_k(t)g_k \tag{4.4}$$

which has been treated in [1]. In our case

$$S_k(t) = -(e^{-tJ_k} - I)J_k^{-1}A_k \tag{4.5}$$

We will assume that the stepsize t_k is calculated by using an Armijo type criterion [2]: let us consider $\delta, \beta \in]0,1[$ and the sequence $\{\alpha^k\}$ in $I\!R^n$; then

$$\begin{aligned} t_k = \max\{\lambda/\lambda = \alpha_k\beta^s, s = 0,1,2,..., \\ f(x^k(\lambda)) \leq f(x_k) - \lambda\delta < g_k, \dot{S}_k(0)g_k >\} \end{aligned} \tag{4.6}$$

The convergence of the algorithm defined by (4.4) with the choice (4.6) of t_k, has been established in [1], under the following hypothesis concerning $\{S_k\}$ and x_0:

(h1) x_0 is chosen such that L_0 is compact.

(h2) $S_k(0) = 0 \quad \forall k = 0,1,2,...$

(h3) $\dot{S}_k(0)$ is symmetric, for all $k = 0, 1, 2, \ldots$ and there exists $m, M > 0$ such that, for all $z \in \mathbb{R}^n$, $k = 0, 1, 2, \ldots$

$$m\|z\|^2 \leq z^T \dot{S}_k(0) z \leq M\|z\|^2 \tag{4.7}$$

(h4) The sequence $\{\dot{S}_k\}$ is equicontinuous at zero, i.e., $\forall \epsilon > 0, \exists \eta > 0$ such that

$$t \in]0, \eta[\Rightarrow \|\dot{S}_k(0) - \dot{S}_k(t)\|_\infty \leq \epsilon \quad \forall k = 0, 1, 2, \ldots \tag{4.8}$$

where $\| \ \|_\infty$ denotes de matrix norm $\|A\|_\infty = \sup\{\|Av\| / v \in \mathbb{R}^n, \|v\| = 1\}$.

The hypothesis (h2) is obviously satisfied by the choice (4.5) of $\{S_k\}$. Let us examine (h3) and (h4). It is clear that

$$\dot{S}_k(t) = e^{-tJ_k} A_k \tag{4.9}$$

and

$$\dot{S}_k(0) = A_k \tag{4.10}$$

so, (h3) is satisfied. In fact, the function A satisfies (H2).

On the other hand, by (4.9), (4.10), (H2) and the mean value theorem,

$$\begin{aligned} \|\dot{S}_k(0) - \dot{S}_k(t)\|_\infty &= \|A_k - e^{-tJ_k} A_k\|_\infty \\ &\leq \|e^{-tJ_k} - I\|_\infty \|A_k\|_\infty \\ &\leq M_1 \|e^{-tJ_k} - I\|_\infty \\ &\leq M_1 t \|e^{-\xi_k J_k}\|_\infty \|J_k\|_\infty \\ &\leq M_1 t e^{\xi_k \|J_k\|_\infty} \|J_k\|_\infty \end{aligned} \tag{4.11}$$

for some $\xi_k \in]0, t[$.

The choice (4.6) of t_k ensures that $x_k \in L_0, \forall k = 0, 1, 2, \ldots$ and owing to the continuity of J we have

$$\|J_k\|_\infty = \|J(x_k)\|_\infty \leq \gamma \tag{4.12}$$

for some $\gamma > 0$. From (4.11) and (4.12),

$$\|\dot{S}_k(0) - \dot{S}_k(t)\|_\infty \leq \gamma M_1 t e^{t\gamma} \tag{4.13}$$

which implies the equicontinuity of $\{\dot{S}_k\}$ at zero, i.e. (h4) is verified.

In this manner, the algorithm defined by (4.3) and (4.6) converges to stationary point of f.

References

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

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

[3] Botsaris, C.A., A Class of Methods for Unscontrained Optimization Based on Stable Numerical Integration Techniques, *Journal of Mathematical Analysis and Applications* **63**(**3**) (1978).

[4] Botsaris, C.A., Differential Gradient Methods, *Journal of Mathematical Analysis and Applications* **63** (1978), 177-198.

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

[6] Coddington, E.A.,Levinson, N., Theory of Ordinary Differential Equations, Mc Graw-Hill, New York, 1955.

[7] Sotomayor, J., Liçoes de Equaçoes Diferenciais Ordinárias, IMPA, Rio de Janeiro, 1979.

[8] Vial, J. Ph., Zang,I., Unconstrained Optimization by Approximation of the Gradient Path, *Mathematics of Operations Research* **2**(**3**) (1977).

[9] Zang, I., A New Arc Algorithm for Unconstrained Optimization, *Mathematical Programming*, **15** (1978), 36-52.

Partially supported by Departamento de Investigación y Bibliotecas, Universidad de Chile, under project E-1947-8522.

Jorge Amaya A.
Departamento de Matemáticas Aplicadas,
Facultad de Ciencias Físicas y Matemáticas,
Universidad de Chile,
Casilla 170
Correo 3,
Santiago,
Chile.

Fecha de recepción: Marzo de 1985.