

# ON THE CONVERGENCE OF THE STEEPEST DESCENT PROCESS

MATHEMATICS

1968

SovietRxiv

---

View the original and related papers at <https://sovietrxiv.org/items/ru-196801.39044>

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.

## Abstract

## Full Text

UDC 518 : 517.948

MATHEMATICS

Yu. I. LYUBICH

# ON THE CONVERGENCE OF THE STEEPEST DESCENT PROCESS

(Presented by Academician L. V. Kantorovich on 3 VI 1967)

The idea of the steepest descent process was put forward already by Cauchy, but only after the appearance of the works <sup>(1-4)</sup> did this process become widely known. At present it and its modifications are a customary means for the numerical solution of extremal problems. However, the theory of the steepest descent process is still insufficiently developed. In the present communication some gaps in this theory will be filled.

Let  $G \subset R^n$  be a domain;  $\varphi(x)$  ( $x \in G$ ) a real-valued function of class  $C^1$ ;  $S = \{x \mid \nabla\varphi(x) = 0\}$ . Put  $\varphi_x(\tau) = \varphi(x - \tau\nabla\varphi(x))$  ( $\tau \geq 0$ ). For  $x \in S$  put

$$\gamma(x) = \sup\{t \mid t > 0, \dot{\varphi}_x(\tau) < 0 \ (0 \leq \tau \leq t)\}.$$

If  $\gamma(x) < \infty$  and is attained, then at the point  $x$  there is defined the **descent operator**

$$\Gamma x = x - \gamma(x)\nabla\varphi(x).$$

In addition,  $\Gamma x = x$  ( $x \in S$ ). Put

$$I = \bigcap_{k=1}^{\infty} \mathfrak{D}(\Gamma^k)$$

( $\mathfrak{D}(\cdot)$  is the domain of definition). The steepest descent process (in what follows we shall say briefly—the process) is defined by the iterative formula

$$x_{k+1} = \Gamma x_k \quad (k = 0, 1, 2, \dots)$$

for  $x_0 \in I$ .

**Lemma 1.** If the sequence  $\{y_k\}_0^\infty \subset \mathfrak{D}(\Gamma)$  converges to a point  $y \in G$  and

$$\lim_{k \rightarrow \infty} \varphi(\Gamma y_k) = \lim_{k \rightarrow \infty} \varphi(y_k),$$

then  $y$  is a stationary point, i.e.  $y \in S$ .

From this it follows immediately:

**Theorem 1.** For any  $x_0 \in I$ , all limit points of the process lying in  $G$  are stationary.

Numerous consequences follow from Theorem 1.

Consider the Lebesgue set  $\{x \mid x \in G, \varphi(x) \leq \varphi(x_0)\}$ . Denote by  $\Omega_{x_0}$  the component of linear connectedness containing  $x_0$ . Obviously,  $\Gamma$  does not take points out of  $\Omega_{x_0}$ .

**Corollary 1.** If  $x_0 \in I$  and  $\Omega_{x_0}$  is closed, then all limit points of the process are stationary.

If  $\Omega_{x_0}$  is compact,\* then  $x_0 \in I$ . For compactness of  $\Omega_{x_0}$  it is sufficient that

$$\varphi(x_0) < \lim_{x \rightarrow \partial G} \varphi(x).$$

The analogous non-strict inequality is necessary.

**Corollary 2.** If  $\Omega_{x_0}$  is compact, then the limit points of the process exist and are stationary.

**Corollary 3.** If  $\Omega_{x_0}$  is compact and the points of the set  $\Omega_{x_0} \cap S$  lie on pairwise distinct  $\varphi$ -levels, then the process converges to some stationary point.

Let us generalize Corollary 3. We shall call points  $u, v \in S$  **connected** if: 1) the rectilinear segment  $[u, v]$  lies entirely in the closure  $\bar{G}$ ; 2)  $\varphi(x) = \text{const}$  on  $[u, v] \cap G$ .

**Lemma 2.** Under the conditions of Lemma 1, all limit points of the sequence  $\{\Gamma y_k\}_0^\infty$  lying in  $G$  are connected with  $y$ .

\* Obviously, in this case  $\Omega_{x_0}$  contains a point at which the absolute minimum of the function  $\varphi$  is attained.

The connectedness relation defines a graph  $\mathfrak{G}$  with vertex set  $S$ . Denote by  $\mathfrak{G}_{x_0}$  the subgraph of the graph  $\mathfrak{G}$  with vertex set  $S \cap \Omega_{x_0}$ .

**Theorem 2.** If  $\Omega_{x_0}$  is compact and on each  $\varphi$ -level in  $\Omega_{x_0}$  there lie no more than a finite number of stationary points, then all limit points of the process belong to one connected component of the graph  $\mathfrak{G}_{x_0}$ .

**Corollary 4.** If, under the hypotheses of Theorem 2, no two stationary points in  $\Omega_{x_0}$  are connected, then the process converges to some stationary point.

With the help of this corollary one easily obtains

**Theorem 3.** For the square of the modulus of a complex polynomial\* the process, for any initial approximation, converges to a root of the polynomial or of its derivative.

Convergence to a root of the derivative\*\* is an exceptional phenomenon. This fact follows from the general scheme developed below.

Let  $a \in S$  and let  $\varphi$  be twice differentiable at the point  $a$ . We shall call the point  $a$  **hyperbolic** if its Morse index is equal to  $n - 1$ , i.e. the Hessian  $H(a) = (\nabla \nabla \varphi)(a)$ , being nonsingular, has only one positive eigenvalue.

The notion of a hyperbolic stationary point can be extended to some degenerate cases. A homogeneous polynomial  $\rho(x) \neq \text{const}$  is called hyperbolic (in the sense of Petrovskii <sup>(5)</sup>) if: 1)  $(\exists y)(\rho(y) > 0)$ ; 2) for any linearly independent  $x, y$  and  $\rho(y) > 0$ , the equation  $\rho(x + \lambda y) = 0$  in  $\lambda$  has only real simple roots. Now let  $a \in S$  and let  $\varphi$  be twice differentiable in a neighborhood of  $a$ . If, as  $x \rightarrow a$ , the Hessian  $H(x)$  has the form

$$H(x) = \nabla \nabla \rho(x - a) + o(\|x - a\|^{r-2}),$$

where  $\rho$  is a hyperbolic polynomial of degree  $r$ , then the point  $a$  is called **hyperbolic**.

**Theorem 4.** Let  $a$  be a hyperbolic stationary point. There exists a neighborhood  $U_a$  such that, if  $x_k \in U_a$ , starting with some index  $m - 1$ , then  $x_m = a$ .

Thus a process converging to a hyperbolic saddle breaks off. Hyperbolicity is necessary if one requires that the break-off phenomenon be preserved under small perturbations of the function  $\varphi$ .

**Lemma 3.** All roots of the derivative of a complex polynomial<sup>\*\*\*</sup>, but not of the polynomial itself, are hyperbolic stationary points for the square of the modulus of the polynomial.

Therefore, for the square of the modulus of a complex polynomial, convergence to a root of the derivative can occur only in a finite number of steps. Relying on this fact, one can establish the following proposition.

**Theorem 5.** For the square of the modulus of a complex polynomial, the set of initial approximations for which the process converges to a root of the derivative, but not of the polynomial itself, is an algebraic curve in  $\mathbb{R}^2$ .

In  $\mathbb{R}^2$ , for any  $\varphi$ , nondegenerate saddle points are hyperbolic, and therefore convergence to them can occur only in a finite number of steps.

**Theorem 6.** Let, in a domain  $G \subset \mathbb{R}^2$ , the function  $\varphi$  be twice continuously differentiable and all its stationary points be nondegenerate and pairwise unconnected. If the function  $\varphi$  in no subdomain of the domain  $G$  satisfies a differential equation of the form

$$|\nabla \varphi|^2 = F(\varphi)$$

---

\* Here and below, distinct from a constant.

\*\* Which, unlike the root of the polynomial itself, is a saddle.

\*\*\* Independently of multiplicity.

and tends to infinity as  $x \rightarrow \partial G$ , then the set of those initial approximations for which the process converges to the saddle point is contained in the union of no more than a countable set of smooth curves.

Thus, if the function  $\varphi$  and the initial approximation  $x_0$  are taken in “general position,” then the process converges to a point of local minimum. The justification of an analogous conclusion in  $\mathbf{R}^n$  ( $n > 2$ ) encounters considerable difficulties.

Kharkov State University  
named after A. M. Gorky

Received  
27 V 1967

## REFERENCES

1. G. Temple, Proc. Roy. Soc., Ser. A, **169**, 476 (1939).
2. H. B. Curry, Quart. Appl. Math., **2**, No. 3 (1944).
3. L. V. Kantorovich, DAN, **48**, No. 7 (1945).
4. L. V. Kantorovich, DAN, **56**, No. 3 (1947).
5. I. G. Petrovsky, Matem. sborn., **2** (44), No. 3 (1937).

*Note: Figure translations are in progress. See original paper for figures.*

*Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.*