



Soviet-era science, translated into English

I. M. GLAZMAN

1964

SovietRxiv

View the original and related papers at <https://sovietrxiv.org/items/ru-196401.94412>

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

Abstract

Full Text

I. M. GLAZMAN

ON GRADIENT RELAXATION FOR NON-QUADRATIC FUNCTIONALS

(Presented by Academician S. N. Bernstein, 15 X 1963)

1. The descent method, proposed by L. V. Kantorovich in ⁽⁵⁾, was investigated by him and by other authors for the case of a positive-definite quadratic functional in Euclidean or Hilbert space (see ^(6, 7) and the bibliography given there). Outside the class of such functionals the descent method, despite its wide use in computational practice, constitutes a problem. Some aspects of this problem are reflected in the survey of B. Tompkins ⁽³⁾ and in the original paper of I. M. Gelfand and M. L. Tsetlin ⁽⁴⁾.

The present note is devoted to one particular question of the descent problem. Its main task is the effective construction of a universal algorithm of gradient relaxation for the class K of all functionals $\Phi(\mathbf{x})$ of a point $\mathbf{x}(x_1, x_2, \dots, x_p)$ of the Euclidean space \mathcal{E}_p satisfying the following three conditions: 1°. $\Phi(\mathbf{x})$ is twice continuously differentiable. 2°. $\Phi(\infty) = \infty$. 3°. $\Phi(\mathbf{x})$ has a unique stationary point $\check{\mathbf{x}}$. Obviously, $\Phi(\check{\mathbf{x}})$ is the least value of the functional $\Phi(\mathbf{x})$ in \mathcal{E}_p , and the desired gradient-relaxation process must produce the corresponding minimizing sequence $\mathbf{x}_n \rightarrow \check{\mathbf{x}}$.

In essence, in what follows, instead of 2° and 3° only the following is used: the level surface $\Phi(\mathbf{x}) = \Phi(\mathbf{x}_0)$, passing through the zero approximation \mathbf{x}_0 , is homeomorphic to a sphere; outside it $\Phi(\mathbf{x}) > \Phi(\mathbf{x}_0)$, while inside it $\Phi(\mathbf{x})$ satisfies the inequality $\Phi(\mathbf{x}) < \Phi(\mathbf{x}_0)$ and has a unique stationary point $\check{\mathbf{x}}$.

2. The methods of gradient relaxation known for quadratic functionals do not carry over to our case. Thus, for example, the method of steepest descent would require, at each step, in order to determine the next multiplier of complete relaxation, solving a nonlinear equation, which in itself requires carrying out an infinite computational procedure. The realization of the stationary process of gradient relaxation would require, before the beginning of the process, estimating from above the maximum M of the operator norm of the Hessian matrix

$$H(\mathbf{x}) = \left(\frac{\partial^2 \Phi}{\partial x_j \partial x_k} \right)_{j,k=1}^p$$

in the domain $\Omega \subset \mathcal{E}_p$ of all \mathbf{x} satisfying the condition $\Phi(\mathbf{x}) < \Phi(\mathbf{x}_0)$. The effective construction of an auxiliary algorithm for such an estimate, suitable for

the entire class K , apparently again requires constructing a minimizing sequence (now already for the functional $-\|H(\mathbf{x})\|$, and moreover with an estimate of the rate of convergence), which leads to a logical circle.

3. The **multiplier of complete relaxation** $\alpha(\mathbf{x})$ for a given point $\mathbf{x} \in \Omega$ is defined as the smallest positive t -root of the equation $\Phi'_t(\mathbf{x} - t\nabla\Phi(\mathbf{x})) = 0$.

Lemma 1. *The multiplier of complete relaxation $\alpha(\mathbf{x})$, for $\mathbf{x} \in \Omega$, satisfies the inequality $\alpha(\mathbf{x}) \geq 1/M$.*

Proof. For the function $\varphi(t) = \Phi(\mathbf{x} - t\nabla\Phi(\mathbf{x}))$ we have

$$\varphi'(0) = - \int_0^{\alpha(\mathbf{x})} (H_t \nabla\Phi(\mathbf{x}), \nabla\Phi(\mathbf{x})) dt, \quad (1)$$

where $H_t = H(x - t\nabla\Phi(x))$, but for $0 \leq t \leq \alpha(x)$

$$|(H_t \nabla\Phi(x), \nabla\Phi(x))| \leq M \|\nabla\Phi(x)\|^2,$$

and $\varphi'(0) = -\|\nabla\Phi(x)\|^2$, so that from (1) it follows that $[M\alpha(x) - 1]\|\nabla\Phi(x)\|^2 \geq 0$, and the lemma is proved.

From Lemma 1 there follow the following two existence theorems.

Theorem 1. *For any $x_0 \in \mathcal{E}_p$, for all points $x \in \Omega$ there exists a relaxation multiplier γ independent of x .*

Proof. It is sufficient to take γ from the interval $0 < \gamma < 1/M$.

Theorem 2. *For any initial approximation $x_0 \in \mathcal{E}_p$, there exist stationary gradient relaxation processes*

$$x_{n+1} = x_n - \gamma \nabla\Phi(x_n) \quad (n = 0, 1, 2, \dots), \quad (2)$$

converging to \tilde{x} .

Proof. It is sufficient to take as γ any common relaxation multiplier (not necessarily the lower one) for all points $x \in \Omega$.

Let us note that not every stationary relaxation process (2) converges, since the multiplier γ , being a relaxation one for all x_n ($n = 0, 1, 2, \dots$), may fail to be relaxation for the whole domain Ω . However, the following obvious statement holds.

Lemma 2. *If the process (2) is a relaxation process and, moreover, the sequence x_n converges, then $\lim_{n \rightarrow \infty} x_n = \tilde{x}$.*

In the general case of a nonstationary relaxation process

$$x_{n+1} = x_n - \gamma_n \nabla \Phi(x_n), \quad (n = 0, 1, 2, \dots) \quad (3)$$

we shall call the sum of the series

$$\sum_{n=0}^{\infty} \|x_{n+1} - x_n\| \quad (4)$$

the **length of the relaxation path**.

Lemma 3. *If in (3) the relaxation multipliers γ_n tend to zero as $n \rightarrow \infty$, but $\lim_{n \rightarrow \infty} \Phi(x_n) > \Phi(\check{x})$, then the length of the relaxation path is finite.*

The proof is based on comparing the series (4) with the series

$$\sum_{n=0}^{\infty} \Delta s_n, \quad (5)$$

where Δs_n is the length of the segment of the orthogonal trajectory of the family $\Phi(x) = C$ with beginning at the point x_n and end on the surface $\Phi(x) = \Phi(x_{n+1})$. For the partial sums of the series (5) one easily obtains the estimate

$$\sum_{n=0}^m \Delta s_n < \frac{1}{\beta} [\Phi(x_0) - \mu],$$

where $\mu = \lim_{n \rightarrow \infty} \Phi(x_n)$ and $\beta = \min \|\nabla \Phi(x)\|$ for $\mu \leq \Phi(x) \leq \Phi(x_0)$.

4. We turn to the description of an algorithm for constructing a process (3) converging to \check{x} .

To control the construction of the process (3), let us assign arbitrarily an increasing sequence of positive numbers $q_m \rightarrow \infty$, and for feedback with it an integer-valued function $i(n)$ will be defined. At each step of the process, after a finite number of trials, the next relaxation multiplier γ_{n-1} , the next approximation x_n , the current length of the relaxation path

$$l_n = \sum_{k=0}^n \|x_{k+1} - x_k\|,$$

and also the value of the function $i(n)$, will be computed.

The first step begins with the choice of an arbitrary point $\mathbf{x}_0 \in \mathcal{E}_p$ and an arbitrary trial multiplier $\gamma'_0 > 0$, which is tested for relaxation by means of the inequality

$$\Phi(\mathbf{x}_0 - \gamma'_0 \nabla \Phi(\mathbf{x}_0)) < \Phi(\mathbf{x}_0). \quad (6)$$

If (6) is true, then we successively double γ'_0 until (6) is violated, and the penultimate value of the trial multiplier in this sequence is taken as the final value γ_0 . If, however, (6) is false, then we successively halve γ'_0 until (6) is satisfied, and the last value of the trial multiplier is taken as γ_0 .

Now, having determined \mathbf{x}_1 by formula (3), we compute l_1 and set $i(1) = 1$, which completes the construction of the first step of process (3).

Suppose the n -th step has been carried out. We begin the next step by checking the relation

$$l_n < l_1 + q_{i(n)}. \quad (7)$$

If (7) is true, then we determine the initial trial value γ'_n of the next relaxation multiplier by the formula $\gamma'_n = \gamma_{n-1}$, set $i(n+1) = i(n)$, and test the multiplier γ'_n for relaxation by means of the inequality

$$\Phi(\mathbf{x}_n - \gamma'_n \nabla \Phi(\mathbf{x}_n)) < \Phi(\mathbf{x}_n). \quad (8)$$

If, however, (7) is false, then we determine γ'_n by the formula $\gamma'_n = \frac{1}{2}\gamma_{n-1}$, set $i(n+1) = i(n) + 1$, and test γ'_n for relaxation by means of (8).

If (8) is satisfied for γ'_n , then we put $\gamma'_n = \gamma_n$. If, however, (8) is false for γ'_n , then we successively halve γ'_n until (8) is satisfied, and the last value of the trial multiplier obtained in this way is taken as γ_n .

Now we compute by formula (3) the next approximation \mathbf{x}_{n+1} and the current length of the relaxation path

$$l_{n+1} = l_n + \|\mathbf{x}_{n+1} - \mathbf{x}_n\|,$$

while the value $i(n+1)$ has already been computed above.

Thus the $(n+1)$ -st step has been carried out and the description of the algorithm is complete.

5. The algorithm described above gives a solution of the problem posed, for the following holds.

Theorem 3. *The process (3), constructed with the aid of the algorithm of item 4, gives a minimizing sequence for the problem of determining the minimum of the functional $\Phi(\mathbf{x})$.*

Proof. In carrying out the algorithm of item 4, the sequence $i(n)$ either remains bounded or tends to infinity.

In the first case, from the boundedness of $i(n)$ it follows that the length of the relaxation path (4) is finite, so that the series

$$\sum_{k=0}^{\infty} (\mathbf{x}_{k+1} - \mathbf{x}_k)$$

converges, i.e., there exists $\lim_{n \rightarrow \infty} \mathbf{x}_n = \mathbf{z}$. Further, since, by Lemma 1, any multiplier smaller than $1/M$ is a relaxation multiplier, it also follows from the boundedness of $i(n)$ that process (3) is stationary. But then, by Lemma 2, $\mathbf{z} = \dot{\mathbf{x}}$, so that

$$\lim_{n \rightarrow \infty} \mathbf{x}_n = \dot{\mathbf{x}}. \quad (9)$$

In the second case, from the unboundedness of $i(n)$ it follows that $\gamma_n \rightarrow 0$. On the other hand, from the very relaxation property of process (3) there follows the existence of

$$\lim_{n \rightarrow \infty} \Phi(\mathbf{x}_n) = \mu.$$

Here it must be that $\mu = \Phi(\dot{\mathbf{x}})$, for otherwise, by Lemma 3, the length of the relaxation path would be finite, whereas the unboundedness of $i(n)$ indicates that the relaxation path is infinite. Thus $\mu = \Phi(\dot{\mathbf{x}})$, so that (9) again holds.

Obviously, for given $\Phi \in K$, x_0 , γ'_0 , and a prescribed error bound, the number of steps of process (3) required to attain the required accuracy depends on the choice of the control sequence $\{q_m\}$.

6. The algorithm described in Sec. 4 can be used to construct minimizing sequences in variational problems.

For illustration we shall confine ourselves to the simplest example of minimizing a "quasi-quadratic" functional

$$\Phi(x) = \int_a^b \left\{ \left(\frac{dx}{d\xi} \right)^2 + f(\xi, x) \right\} d\xi$$

on the class \mathfrak{U} of all absolutely continuous functions $x = x(\xi)$ satisfying the boundary conditions $x(a) = a_1$, $x(b) = b_1$. We shall assume that in the strip $a \leq \xi \leq b$, $-\infty < x < \infty$ the function $f(\xi, x)$ is continuous in ξ , twice continuously differentiable in x , and satisfies the conditions $f(\xi, x) \geq \beta > -\infty$, $f''_{xx}(\xi, x) \geq k > 0$, where β and k are some constants.

From a theorem of S. N. Bernstein ⁽²⁾ (see also ⁽¹⁾) it follows, in particular, that the problem under consideration has a solution $\bar{x}(\xi)$, and that it is unique. Further, using the compactness in $C[a, b]$ of the set Ω of all $x \in \mathfrak{U}$ satisfying, for some $x_0 \in \mathfrak{U}$, the inequality $\Phi(x) \leq \Phi(x_0)$, and other properties of the set Ω (see ⁽¹⁾), it is not difficult to establish convergence in $C[a, b]$ of the sequence of functions constructed by means of the algorithm of Sec. 4 to $\bar{x}(\xi)$. Note

that in the case when $f(\xi, x)$ is a polynomial in ξ and x , implementation of the algorithm of Sec. 4 requires only rational operations and leads to a polynomial minimizing sequence.

Physical-Technical Institute of Low Temperatures
Academy of Sciences of the Ukrainian SSR

Received
25 IX 1963

REFERENCES

- ¹ N. I. Akhiezer, *Lectures on the Calculus of Variations*, 1955.
- ² S. N. Bernstein, *Ann. sci. école norm. sup.*, **29**, 431 (1912).
- ³ *Modern Mathematics for Engineers*, ed. by E. F. Beckenbach, IL, 1958.
- ⁴ I. M. Gelfand, M. L. Tsetlin, *UMN*, **17**, no. 1, 3 (1962).
- ⁵ L. V. Kantorovich, *DAN*, **48**, no. 7 (1945).
- ⁶ L. V. Kantorovich, G. P. Akilov, *Functional Analysis in Normed Spaces*, Moscow, 1959.
- ⁷ D. K. Faddeev, V. N. Faddeeva, *Computational Methods of Linear Algebra*, Moscow, 1960.

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

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