



Soviet-era science, translated into English

CONDITIONING IN GENERAL COMPUTATIONAL PROBLEMS

MATHEMATICS

1966

SovietRxiv

View the original and related papers at <https://sovietrxiv.org/items/ru-196601.89529>

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

Abstract

Full Text

UDC 518:517.948

MATHEMATICS

Yu. I. LYUBICH

CONDITIONING IN GENERAL COMPUTATIONAL PROBLEMS

(Presented by Academician S. N. Bernstein on 11 II 1966)

Consider the problem of computing the values of a continuous operator F , acting from one linear normed space (l.n.s.) into another and defined on an open set D_F . Suppose that $0 \in D_F$, $0 \in FD_F$.

Let a point $x \in D_F$ be given with absolute error not exceeding a certain $\varepsilon > 0$. This means that a point $a \in D_F$ is given such that $\|x - a\| \leq \varepsilon$. We shall take ε so small that

$$K(a, \varepsilon) \equiv \{x \mid \|x - a\| \leq \varepsilon\} \subset D_F.$$

We denote the problem of computing Fx in the situation described by $\text{Comp}(F, a, \varepsilon)$.

Since the position of the point x in the ball $K(a, \varepsilon)$ is not fixed by any additional conditions, each of the values Fx ($x \in K(a, \varepsilon)$) may turn out to be "true." In reality, however, only the approximation Fa is computed. But then it is very important to be able to estimate the degree of reliability of the result of the computation.* In this connection we propose the following

Definition 1. The **condition measure** of the problem $\text{Comp}(F, a, \varepsilon)$ is the quantity

$$c(F, a, \varepsilon) = \sup_{x \in K(a, \varepsilon), x \neq a} \left\{ \frac{\|Fx - Fa\|}{\|Fa\|} : \frac{\|x - a\|}{\|a\|} \right\}.$$

The expression in braces is the quotient of the relative errors of the approximate equalities $Fx \approx Fa$, $x \approx a$.

Definition 1, in the case of the problem of solving a linear equation, is equivalent to the definition given by N. P. Zhidkov ⁽¹⁾. The term "condition measure" goes back to the well-known work of A. Turing ⁽²⁾, in which the "condition measure of a matrix A ," $c(A) = \|A\| \|A^{-1}\|$, was introduced** in connection with problems of solving linear equations and inverting matrices. The relation of the functional $c(A)$ to the concepts proposed by us will be established below.

Let some number system with base $p > 1$ be fixed.

Definition 2. The **uncertainty measure** of the problem $\text{Comp}(F, a, \varepsilon)$ is the quantity

$$I(F, a, \varepsilon) = \log_p c(F, a, \varepsilon).$$

The uncertainty measure $I(F, a, \varepsilon)$ is the maximum possible loss^{***} of information occurring when x is transformed into Fx by the operator F in the ε -neighborhood of the point a . Here the information about x contained in a is compared with the information about Fx contained in Fa . Information is measured by the number of correct digits.

* We assume that Fa is computed absolutely exactly. Thus the question is not of rounding errors, but of the degree of stability of the result with respect to the initial data.

** Turing used the matrix norm $N(A) = \sqrt{\text{sp}(A^*A)}$, as well as the norm $\max |a_{ij}|$ (see also (3)). In (1) arbitrary subordinate matrix norms are used. Subordinate norms are operator norms in the sense of (4).

*** The inequality $I(F, a, \varepsilon) \geq 0$ is not necessary, but this is the typical and most important case.

Knowing the measure of uncertainty of the problem (or its estimate), one can estimate the number of reserve digits in the initial data sufficient for reliable computation.

Let now $\varepsilon \rightarrow 0$. Consider the “limiting” problem $\text{Comp}(F, a)$. Suppose that the operator F has at the point a a Fréchet derivative $F'(a)$:

$$F(x) = F(a) + F'(a)(x - a) + o(\|x - a\|) \quad (x \rightarrow a).$$

As usual, $F'(a)$ is a bounded linear operator.

Definition 3. The **measure of conditioning** of the problem $\text{Comp}(F, a)$ is the quantity

$$c(F, a) = \frac{\|a\| \|F'(a)\|}{\|Fa\|}.$$

Definition 4. The **measure of uncertainty** of the problem $\text{Comp}(F, a)$ is the quantity $I(F, a) = \log_p c(F, a)$.

It is easily established that

$$c(F, a) = \lim_{\varepsilon \rightarrow 0} c(F, a, \varepsilon). \quad (1)$$

But $c(F, a, \varepsilon)$ does not increase as ε decreases. Therefore

Theorem 1. *The inequality holds*

$$c(F, a, \varepsilon) \geq c(F, a). \quad (2)$$

Consequently,

$$I(F, a) = \lim_{\varepsilon \rightarrow 0} I(F, a, \varepsilon); \quad I(F, a, \varepsilon) \geq I(F, a), \quad (3)$$

i.e. the measure of uncertainty of the problem $\text{Comp}(F, a)$ represents the minimal irreducible loss of information when using the approximation a .

We now show that Turing's measure of conditioning in the operator norm coincides with the measure of conditioning of the limiting problem of inverting a linear operator.

Theorem 2. *Let F be the inversion operator in the space of bounded linear operators in some normed linear space. Then*

$$c(F, a) = \|a\| \|a^{-1}\|. \quad (4)$$

Proof. Since $F'(a) = -a^{-1}(\cdot)a^{-1}$, we have $\|F'(a)\| \leq \|a^{-1}\|^2$. On the other hand, let $\theta \in (0, 1)$; let u be such a vector that $\|u\| = 1$, $\|a^{-1}u\| \geq \theta \|a^{-1}\|$, and let f be such a linear functional that $\|f\| = 1$, $f(a^{-1}u) = \|a^{-1}u\|$. Then for the operator $z = f \otimes u$ we have $\|z\| = 1$, $\|F'(a)z\| \geq \theta \|a^{-1}\|^2$. Consequently, $\|F'(a)\| = \|a^{-1}\|^2$, whence (4) follows.

Corollary.

$$\sup_{\|x-a\| \leq \varepsilon} \left\{ \frac{\|x^{-1} - a^{-1}\|}{\|a^{-1}\|} : \frac{\|x-a\|}{\|a\|} \right\} \geq \|a\| \|a^{-1}\|.$$

Let us consider several more important examples.

1. F is a **linear operator**. Then $F' = F$ and

$$c(F, a) = \frac{\|a\| \|F\|}{\|Fa\|} \geq 1 \quad (a \notin \text{Ker } F). \quad (5)$$

In particular, in the problem of solving the linear equation $Ay = x$ we have $F = A^{-1}$, and formula (5) is applicable. If the operator A is bounded together with its inverse, then

$$\sup_a c(A^{-1}, a) = \|A\| \|A^{-1}\|. \quad (6)$$

Formula (6) was noted in ⁽¹⁾.

2. F -a **bilinear operator** on a pair of l.n.s. E_1, E_2 . The Cartesian product $E_1 \times E_2$ is normed in some way: $\|(x, y)\| = \nu(\|x\|, \|y\|)$, where $x \in E_1$, $y \in E_2$, and $\nu(\xi, \eta)$ is some norm in the two-dimensional cone $\xi \geq 0$, $\eta \geq 0$, nondecreasing in each argument. With the aid of the linear operators associated with F , one obtains the inequality

$$c(F, (a, b)) \leq \|F\| \frac{\nu(\|a\|, \|b\|) \nu^*(\|b\|, \|a\|)}{\|F(a, b)\|}, \quad (7)$$

where ν^* is the norm conjugate* to ν .

An example of a bilinear operator is the operator of multiplication of elements in a normed ring, $F(x, y) = xy$. Here usually $\|F\| = 1$. If one sets $\nu(\xi, \eta) = \sqrt{\xi^2 + \eta^2}$, then $\nu^* = \nu$, and (7) takes the form

$$c(F, (a, b)) \leq \frac{\|a\|^2 + \|b\|^2}{\|ab\|}. \quad (8)$$

3. F -a **scalar function of one scalar variable**. Then the Fréchet derivative coincides with the ordinary derivative, and

$$c(F, a) = \left| a \frac{F'(a)}{F(a)} \right| = \left| a \left(\frac{d}{dx} \ln F(x) \right)_{x=a} \right|. \quad (9)$$

The role of the right-hand side of formula (9) in the elementary theory of errors is well known (see, for example, (5)).

If, in particular, $F(x) = x^\alpha$ ($x > 0$), then $c(F, a) = |\alpha|$. Hence it is evident that when $|\alpha| < 1$ no loss of information occurs. When $|\alpha| > 1$, there is a loss of not less than $\log_p |\alpha|$ correct digits. All these circumstances are also well known.

In conclusion we note one important formal property of the condition measure. Consider the product $F_2 F_1$ of two admissible operators. This is again an admissible operator. Introduce the modulus of continuity $\omega(F, a, \varepsilon) = \sup_{x \in K(a, \varepsilon)} \|Fx - Fa\|$. Then, as is easy to see,

$$c(F_2 F_1, a, \varepsilon) \leq c(F_2, F_1 a, \omega(F_1, a, \varepsilon)) c(F_1, a, \varepsilon), \quad (10)$$

whence, as $\varepsilon \rightarrow 0$,

$$c(F_2 F_1, a) \leq c(F_2, F_1 a) c(F_1, a). \quad (11)$$

By virtue of (10) and (11), measures of uncertainty are submultiplicative.

Kharkov State University
named after A. M. Gorky

Received
9 II 1966

CITED LITERATURE

- ¹ N. P. Zhidkov, *Zhurn. vychisl. matem. i matem. fiz.*, **3**, No. 5, 803 (1963).
- ² A. M. Turing, *Quart. J. Mech. and Appl. Math.*, **1**, 287 (1948).
- ³ D. K. Faddeev, V. N. Faddeeva, *Computational Methods of Linear Algebra*, Moscow–Leningrad, 1963.
- ⁴ Yu. I. Lyubich, *UMN*, **18**, No. 4, 161 (1963).
- ⁵ I. S. Berezin, N. P. Zhidkov, *Methods of Computation*, Moscow, 1959.

$$* \text{ I.e. } \nu^*(\xi, \eta) = \sup_{\alpha, \beta} \frac{\xi\alpha + \eta\beta}{\nu(\alpha, \beta)}.$$

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

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