DOI 10.1007/s10958-020-04778-1
*Journal of Mathematical Sciences, Vol. 246, No. 6, May, 2020*

# PROBABILISTIC ESTIMATION OF MATRIX CONDITION NUMBER

**V. S. Antyufeev**
Institute Numerical Mathematics and Mathematical Geophysics SB RAS
6, Akad. Lavrent'eva pr., Novosibirsk 630090, Russia
Novosibirsk State University
1, Pirogova St., Novosibirsk 630090, Russia
ant@osmf.sscc.ru

UDC 519.6

*We consider ill-conditioned matrices of systems of linear algebraic equations with random error of vector right-hand side. We study the condition number $\nu$ of the matrix of the system. We show that, under certain natural assumptions, the number $\nu$ can be considerably diminished. Bibliography*: 3 *titles. Illustrations*: 3 *figures.*

**0.1.** We consider the system of linear algebraic equations $A\boldsymbol{x} = \boldsymbol{b}$. Let $\Delta\boldsymbol{b}$ and $\Delta\boldsymbol{x}$ are errors of the right-hand side $\boldsymbol{b}$ and solution $\boldsymbol{x}$ respectively. The relative errors of the solution and right-hand side are connected by the inequality

$$\frac{\|\Delta\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \leqslant \nu \cdot \frac{\|\Delta\boldsymbol{b}\|}{\|\boldsymbol{b}\|}, \tag{1}$$

where the condition number $\nu := \|A^{-1}\| \cdot \|A\|$ determines the quality of the matrix $A$. The value of the condition number can be used to determine the coefficient of the regularization system. The number $\nu$ in the inequality (1) cannot be replaced with a less number. However, in a sense, the estimate (1) is rough. We will specify this assertion below and show how to diminish $\nu$ considerably in the estimate.

**0.2.** It is natural to assume that the error vector $\Delta\boldsymbol{b}$ is random. If the components of this vector are independent, then it is natural to assume that the random radius vector $\boldsymbol{s} := \Delta\boldsymbol{b}/|\Delta\boldsymbol{b}|$ is isotropic, i.e., a random point $\boldsymbol{s}$ is uniformly distributed on the unit sphere $S^{n-1} \subset R^n$. Assume that the vectors $\boldsymbol{x}$ and $\boldsymbol{b}$ are fixed (nonrandom).

We propose a new method for estimating the condition number in the norm $\|\cdot\|_2$ taking into account the randomness of the error vector. We describe ideas providing to such an estimate.

**0.3.** The inequality (1) is equivalent to the inequality

$$\frac{|\Delta\boldsymbol{x}|}{|\Delta\boldsymbol{b}|} \cdot \frac{|\boldsymbol{b}|}{|\boldsymbol{x}|} \leqslant \|A^{-1}\| \cdot \|A\|$$

Translated from *Sibirskii Zhurnal Chistoi i Prikladnoi Matematiki* **18**, No. 1, 2018, pp. 28-34.

which is the product of the inequalities

$$\frac{|\Delta \boldsymbol{x}|}{|\Delta \boldsymbol{b}|} \leqslant \|A^{-1}\| \tag{2}$$

and

$$\frac{|\boldsymbol{b}|}{|\boldsymbol{x}|} \leqslant \|A\|.$$

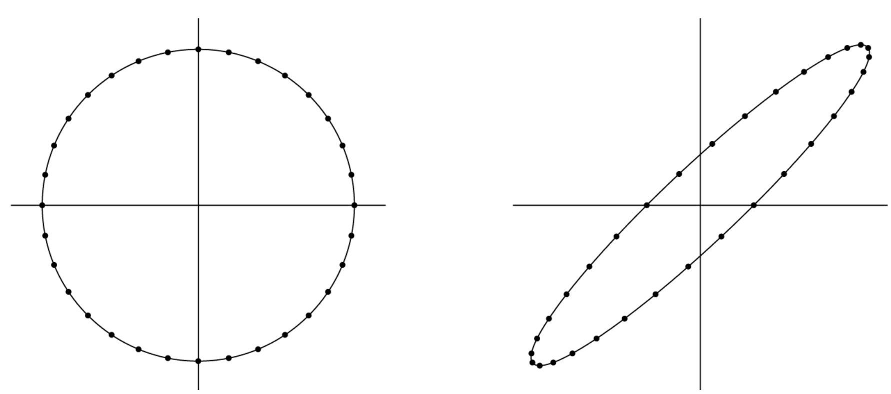
The inequality (2) is related to the random vectors $\Delta \boldsymbol{x}$ and $\Delta \boldsymbol{b}$. We write the inequality (2) in the equivalent form

$$|A^{-1}\boldsymbol{s}| \leqslant \|A^{-1}\|. \tag{3}$$

**0.4.** The length of the radius vector $\boldsymbol{e} = A^{-1}\boldsymbol{s}$ varies in $[\lambda_n, \lambda_1]$, where $\lambda_n \ll \lambda_1$, if the matrix $A$ is ill-conditioned.

To illustrate the situation, we consider the simplified two-dimensional case. In Figure 1, the points $\boldsymbol{s}_k$ are uniformly distributed on the unit circle $S^1$, whereas the corresponding points $\boldsymbol{e}_k = A^{-1}\boldsymbol{s}_k$ are located on the ellipse $E$. It is seen that the length of the larger half-axis of the ellipse $E$ is considerably greater than the mean value of the random length of the radius vector $\boldsymbol{e} = A^{-1}\boldsymbol{s}$:

$$\mathbf{E}(|A^{-1}\boldsymbol{s}|) \ll \max_{s \in S^1} |A^{-1}\boldsymbol{s}| \equiv \|A^{-1}\|. \tag{4}$$



(a) (b)

FIGURE 1. Location of points (a) $\boldsymbol{s}_k \in S^1$ and (b) $\boldsymbol{e}_k \in E$.

In the case of a real ill-conditioned matrix, the condition number can be equal to several hundred or even thousands so that the inequality (4) becomes stronger and, respectively, the inequality (1) presents a rather rough estimate. The above raises a question about distribution of the random variable $|A^{-1}\boldsymbol{s}|$.

## 1 The Random Variable $L$

**1.1.** We write the expression for the length $\mathbf{L}$ of the radius vector of the point $\boldsymbol{e} = A^{-1}\boldsymbol{s} \in E$

$$\mathbf{L} := |A^{-1}\boldsymbol{s}| = (A^{-1}\boldsymbol{s} \cdot A^{-1}\boldsymbol{s})^{1/2} = (\boldsymbol{s} \cdot (A^{-1})^{*}A^{-1}\boldsymbol{s})^{1/2} = ((A^{*}A)^{-1}\boldsymbol{s} \cdot \boldsymbol{s})^{1/2}.$$

Here, $\boldsymbol{s}$ is a random point and, consequently, $\mathbf{L}$ is a random variable.

Let $\lambda_1^2 > \ldots > \lambda_n^2 > 0$ be eigenvalues of the matrix $(A^*A)^{-1}$. We have [1]

$$\frac{1}{\|A\|} = \lambda_n \leqslant \mathbf{L} = \left(\sum_{k=1}^{n} \lambda_k^2 s_k^2\right)^{1/2} \leqslant \lambda_1 = \|A^{-1}\|. \tag{5}$$

**1.2.** Since $\boldsymbol{s}$ is a random point, the inequality (3) is an estimate for the random variable $|A^{-1}\boldsymbol{s}|$ and is satisfied with probability 1. We write this inequality in the equivalent probabilistic form

$$P(\mathbf{L} < \lambda_1) = 1. \tag{6}$$

In turn, (6) can be generalized as follows. Let $F$ be a distribution [2] of a random variable $\mathbf{L}$. Then $P(\mathbf{L} < t) = F(t)$ is a generalization of (6).

**1.3.** We consider calculations of the values of the function $F$. Let $\mu$ be the Lebesgue measure on the sphere $S^{n-1}$. Since the point $\boldsymbol{s}$ is uniformly distributed on $S^{n-1}$ (cf. Subsection 0.2), for any (measurable) set $M \subset S^{n-1}$

$$P(\boldsymbol{s} \in M) = \mu(M)/\mu(S^{n-1}). \tag{7}$$

Thus,

$$F(t) = \begin{cases} 0, & t \leqslant \lambda_n, \\ \dfrac{\mu(\mathbf{L} < t)}{\mu(S^{n-1})}, & \lambda_n < t < \lambda_1, \\ 1, & t \geqslant \lambda_1. \end{cases} \tag{8}$$

We cannot analytically determine $\mu(\mathbf{L} < t)$ in (8) (together with the function $F$). Therefore, we calculate a sample distribution function for $\mathbf{L}$ by the Monte–Carlo method. For this purpose we simulate a sequence of random points $\boldsymbol{s}_k$, $k = 1, 2, \ldots$, uniformly on $S^{n-1}$ with the help of known formulas. For each point $\boldsymbol{s}_k$ we calculate $t_k = |A^{-1}\boldsymbol{s}_k|$. The numbers $t_k$, $k = 1, 2, \ldots$, form a sequence of sample values of the random variable $\mathbf{L}$. If the number $N_s$ of points $\boldsymbol{s}_k$ is sufficiently large, then we construct the approximate distribution function.

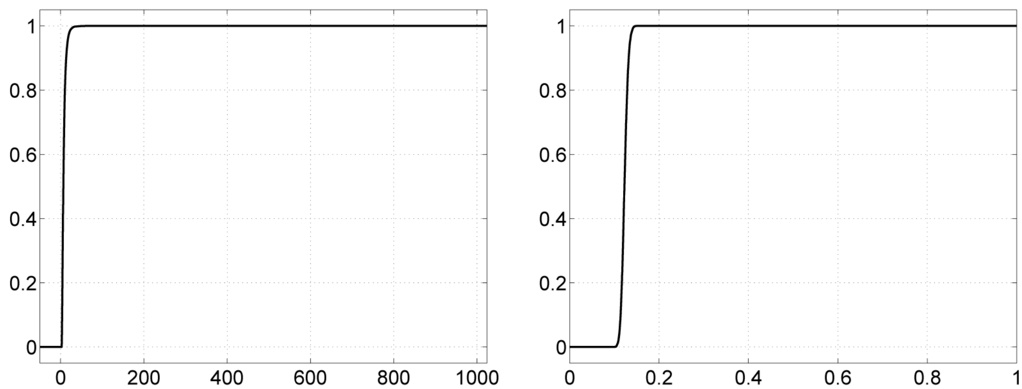
Note that $t_k$ for every $k = 1, 2, \ldots, N_s$ satisfies the inequalities $\lambda_N \leqslant t_k \leqslant \lambda_1$. For every $t_i$ we calculate the number $N_i$ of $t_k$ such that $t_k < t_i$. We define the function $F_{N_s}$ at the nodes $t_k$, $k = 1, 2, \ldots, N_s$, by the formula

$$F_{N_s}(t_k) = \frac{N_i}{N_s}.$$

The piecewise constant approximation of the function $F_{N_s}$ between neighboring nodes yields the sample function of distribution. Below, we will consider piecewise linear approximation in order to obtain convenient graphs.

The random variable $\mathbf{L}$ and its distribution function $F$ are determined by the choice of a finite sequence $\Lambda(1{:}n) = \{\lambda_k, k = 1, \ldots, n\}$. Therefore, we use the notation $\mathbf{L} \equiv \mathbf{L}_{\Lambda(1:n)}$ and $F \equiv F_{\Lambda(1:n)}$.

Further, we will consider only number sequences $\Lambda(1{:}n)$, where $n$ is large and $\lambda_n/\lambda_1$ is small, without mention matrices from which they are obtained.




(a) (b)

FIGURE 2. (a) $\Lambda^1 = \{\lambda_k = 2^{10}, 2^9, \ldots, 1\}$ and (b) $\Lambda^2 = \{\lambda_k = 1.01^0, 1.01^{-1}, \ldots, 1.01^{-300}\}$.

For two such sequences $\Lambda^1$ and $\Lambda^2$ we present the graphs of the corresponding functions $F_{\Lambda^1}$ and $F_{\Lambda^2}$. They look like the graph of the Heaviside function $\chi$. This fact will be explained in the theorem proved in the following section.

## 2 Convergence Theorem

We show that for finite number sequences $\Lambda(1{:}n)$ (cf. Section 1) $F_{\Lambda(1:n)} \simeq \chi$. To prove the corresponding limit relation, we form the auxiliary normed sequence $\overline{\Lambda}(1{:}n) = \{\overline{\lambda}_k = \lambda_k/\lambda_1, k = 1, \ldots, n\}$. Here, $\overline{\lambda}_1 = 1$ and $\overline{\lambda}_n \simeq 0$. The finite normed sequence $\overline{\Lambda}(1{:}n)$ can be regarded as a part of the infinite sequence $\{\overline{\lambda}_k,\ k = 1, 2, \ldots\}$, where $\overline{\lambda}_k \to 0$. For such sequences the following assertion holds.

**Theorem.** *Assume that $\overline{\lambda}_k \to 0$ and $\boldsymbol{s}^n = (s_1^n, \ldots, s_n^n)$ is a random point uniformly distributed on the unit sphere $S^{n-1}$; $\mathbf{L}_{\overline{\Lambda}(1:n)} = \left(\overline{\lambda}_1^2 (s_1^n)^2 + \ldots + \overline{\lambda}_n^2 (s_n^n)^2\right)^{1/2}$. Then for all $t \in R$*

$$\lim_{n\to\infty} F_{\overline{\Lambda}(1:n)}(t) = \chi(t), \tag{9}$$

*Moreover, the convergence $F_{\overline{\Lambda}(1:n)} \to \chi$ for any $\varepsilon > 0$ is uniform on the set $R \setminus (0, \varepsilon)$.*

**Proof.** Below, all formulas with the Γ-function are taken from [3]. We first transform the relation (9).

**2.1.** By (5), we have

$$0 < \overline{\lambda}_n \leqslant \mathbf{L}_{\overline{\Lambda}(1:n)} \leqslant \overline{\lambda}_1 = 1.$$

Consequently,

$$F_{\overline{\Lambda}(1:n)}(t) \equiv \begin{cases} 0, & t \in (-\infty, \overline{\lambda}_n], \\ 1, & t \in [1, +\infty), \end{cases}$$

i.e., $F_{\overline{\Lambda}(1:n)}(t) \equiv \chi(t)$ for $t \in (-\infty, 0] \cup [1, +\infty)$. Now, instead of (9), it suffices to prove that for $t \in (0, 1)$

$$\lim_{n\to\infty} F_{\overline{\Lambda}(1:n)} = 1; \tag{10}$$

moreover, the convergence is uniform on the set $[\varepsilon, 1)$ for any $\varepsilon \in (0, 1)$.

**2.2.** Since

$$F_{\overline{\Lambda}(1:n)}(t) = P(\mathbf{L}_{\overline{\Lambda}(1:n)} < t) = 1 - P(\mathbf{L}_{\overline{\Lambda}(1:n)} \geqslant t),$$

(10) is equivalent to the relation

$$P(\mathbf{L}_{\overline{\Lambda}(1:n)} \geqslant t) \overset{(7)}{=} \frac{\mu(\mathbf{L}_{\overline{\Lambda}(1:n)} \geqslant t)}{\mu(S^{n-1})} = \frac{\mu\Big(\sum\limits_{k=1}^{n} \overline{\lambda}_k^2 (s_k^n)^2 \geqslant t^2\Big)}{\mu(S^{n-1})} \to 0.$$

Thus, instead of (9), we have the equivalent relation

$$\frac{\mu\Big(\sum\limits_{k=1}^{n} \overline{\lambda}_k^2 (s_k^n)^2 \geqslant t^2\Big)}{\mu(S^{n-1})} \to 0. \tag{11}$$

**2.3.** We prove an auxiliary inequality. Let a function $f \geqslant 0$ be defined on the set $S$, and let $h > 0$. Then

$$\int\limits_S f = \int\limits_{f(\boldsymbol{x})<h} + \int\limits_{f(\boldsymbol{x})\geqslant h} \geqslant h \cdot \mu\{f(\boldsymbol{x}) \geqslant h\}.$$

Hence

$$\mu\{f(\boldsymbol{x}) \geqslant h\} \leqslant \frac{1}{h} \cdot \int\limits_S f. \tag{12}$$

**2.4.** Assume that

$$f(\boldsymbol{s}) = \sum \overline{\lambda}_k^2 (s_k^n)^2, \quad h = t^2,$$

in (12). We estimate the numerator of the fraction in (11):

$$\mu\left(\sum_{k=1}^{n} \overline{\lambda}_k^2 (s_k^n)^2 \geqslant t^2\right) \leqslant \frac{1}{t^2} \int\limits_{S^{n-1}} \sum_{k=1}^{n} \overline{\lambda}_k^2 (s_k^n)^2 \, d\boldsymbol{s}^n = \left(\frac{1}{t^2} \sum_{k=1}^{n} \overline{\lambda}_k^2\right) \cdot \int\limits_{S^{n-1}} (s_k^n)^2 d\boldsymbol{s}^n. \tag{13}$$

Since the integral on the right-hand side of (13) is independent of $k$, we can interchange summation and integration if $1 \leqslant k \leqslant n$ is arbitrary.

**2.5.** We calculate the integral on the right-hand side of (13):

$$\int\limits_{S^{n-1}} (s_k^n)^2 d\boldsymbol{s}^n = 2\frac{\pi^{(n-1)/2}}{\Gamma((n-1)/2)} \cdot \int\limits_0^\pi \cos^2\varphi \sin^{n-2}\varphi d\varphi = \frac{\pi^{(n-1)/2}}{\Gamma((n-1)/2)} \int\limits_0^\pi (\sin^{n-2}\varphi - \sin^n\varphi) d\varphi.$$

Here,

$$\int\limits_0^\pi \sin^n\varphi \, d\varphi = \sqrt{\pi} \cdot \frac{\Gamma\big(\frac{n+1}{2}\big)}{\Gamma\big(\frac{n+2}{2}\big)}.$$

We use the following two simple assertions [3].

**2.6.1.** If $c_n \to 0$, then $\frac{1}{n} \sum\limits_{k=1}^{n} c_k \to 0$.

**2.6.2.** If $\lim\limits_{k\to\infty} a_k = 0$ and $a_k \geqslant b_k \geqslant 0$ for all $k$, then $\lim\limits_{k\to\infty} b_k = 0$.

**2.7.** We return to the proof of (11):

$$0 \leqslant \frac{\mu\Big(\sum\limits_{k=1}^{n} \overline{\lambda}_k^2 (s_k^n)^2 \geqslant t^2\Big)}{\mu(S^{n-1})} \leqslant \ldots .$$

Taking into account Subsections 2.4 and 2.5 and the formula

$$\mu(S^{n-1}) = 2\frac{\pi^{n/2}}{\Gamma(n/2)},$$

we have

$$\ldots \leqslant \left(\frac{1}{t^2}\sum_{k=1}^{n}\overline{\lambda}_k^2\right) \cdot \frac{2\frac{\pi^{(n-1)/2}}{\Gamma\left(\frac{n-1}{2}\right)}}{2\frac{\pi^{n/2}}{\Gamma\left(\frac{n}{2}\right)}} \cdot \sqrt{\pi} \cdot \left(\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} - \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n+2}{2}\right)}\right)$$

$$= \left(\frac{1}{t^2}\sum_{k=1}^{n}\overline{\lambda}_k^2\right) \cdot \left(1 - \frac{\Gamma\left(\frac{n}{2}\right)\cdot\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)\cdot\Gamma\left(\frac{n+2}{2}\right)}\right) = \ldots .$$

Using the formula

$$\Gamma(x+1) = x\cdot\Gamma(x), \quad x = (n-1)/2,\ x = n/2,$$

and taking into account that $t \geqslant \varepsilon$, we continue as follows:

$$\ldots = \left(\frac{1}{t^2}\sum_{k=1}^{n}\overline{\lambda}_k^2\right) \cdot \left(1 - \frac{\Gamma\left(\frac{n}{2}\right)\cdot\Gamma\left(\frac{n-1}{2}\right)\cdot\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)\cdot\Gamma\left(\frac{n}{2}\right)\cdot\left(\frac{n}{2}\right)}\right) = \frac{1}{t^2}\cdot\frac{1}{n}\sum_{k=1}^{n}\overline{\lambda}_k^2 \leqslant \frac{1}{\varepsilon^2}\left[\frac{1}{n}\sum_{k=1}^{n}\overline{\lambda}_k\right].$$

Returning to the beginning of Subsection 2.7, we see that

$$0 \leqslant \frac{\mu\Big(\sum\limits_{k=1}^{n} \overline{\lambda}_k^2 (s_k^n)^2 \geqslant t^2\Big)}{\mu(S^{n-1})} \leqslant \frac{1}{\varepsilon^2}\Big[\frac{1}{n}\sum_{k=1}^{n}\overline{\lambda}_k\Big]; \tag{14}$$

moreover,

$$\frac{1}{\varepsilon^2}\Big[\frac{1}{n}\sum_{k=1}^{n}\overline{\lambda}_k\Big] \to 0$$

by Subsection 2.6.1 and

$$\frac{\mu\Big(\sum\limits_{k=1}^{n} \overline{\lambda}_k^2 (s_k^n)^2 \geqslant t^2\Big)}{\mu(S^{n-1})} \to 0$$

by Subsection 2.6.2. The convergence is uniform since the majorant in (14) is independent of $t$. The theorem is proved. □

We return to the finite sequence $\Lambda(1{:}n) = \{\lambda_k > 0,\ k = 1,\ldots,n\}$, where $n$ is large and $\lambda_n/\lambda_1$ is small. With this sequence we associate the auxiliary sequence $\overline{\Lambda}(1{:}n)$. Since $\mathbf{L}_{\Lambda(1:n)}$ and $\mathbf{L}_{\overline{\Lambda}(1:n)}$ are connected by

$$\mathbf{L}_{\Lambda(1:n)} = (\lambda_1^2(s_1^n)^2 + \ldots + \lambda_n^2(s_n^n)^2)^{1/2} = \lambda_1\cdot(\overline{\lambda}_1^2(s_1^n)^2 + \ldots + \overline{\lambda}_n^2(s_n^n)^2)^{1/2} = \lambda_1\mathbf{L}_{\overline{\Lambda}(1:n)},$$

the functions $F_{\Lambda(1:n)}$ and $F_{\overline{\Lambda}(1:n)}$ are connected by the equality

$$F_{\Lambda(1:n)}(t) = F_{\overline{\Lambda}(1:n)}(t/\lambda_1).$$

By the above theorem, $F_{\overline{\Lambda}(1:n)} \simeq \chi$ for large $n$ outside the interval $(0, \varepsilon)$. Consequently, $F_{\Lambda(1:n)} \simeq \chi$ outside the interval $(0, \lambda_1\varepsilon)$.

In the following section, we show how to use the above-obtained results.

## 3 Replacement of $\nu$ with $\nu_\varepsilon$

By definition and notation, the condition coefficient $\nu(A)$ is the least number satisfying (with probability 1) the inequality $\mathbf{L} \cdot \|A\| \leqslant \nu$.

We choose a sufficiently small $\varepsilon > 0$ such that it is possible to ignore some random event in calculations provided that the probability of this event is less than $\varepsilon$. We find a numerical solution to Equation $F_{\overline{\Lambda}(1:n)}(L_\varepsilon) = 1 - \varepsilon$ (cf. Figure 3; a fragment of Figure 2).

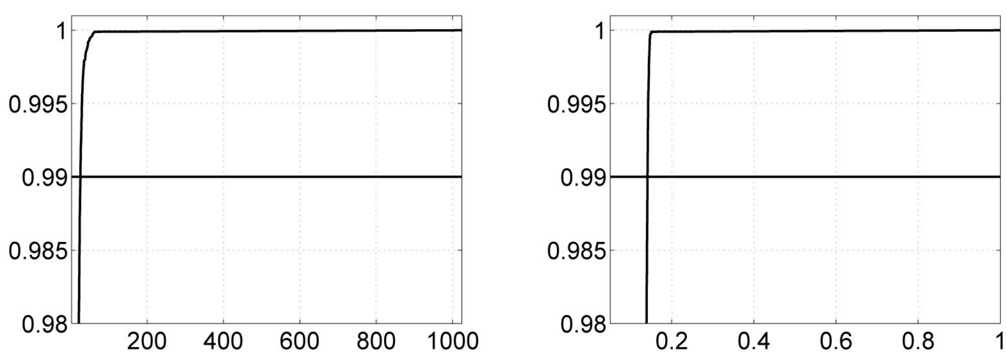



FIGURE 3. The graphical solution of the equations $F_{\Lambda^k}(L_\varepsilon) = 1 - \varepsilon$ $(k = 1, 2)$.

Since $P(\mathbf{L} < L_\varepsilon) \equiv F_{\overline{\Lambda}(1:n)}(L_\varepsilon) = 1 - \varepsilon$, we have $P(\mathbf{L} \geqslant L_\varepsilon) = 1 - P(\mathbf{L} < L_\varepsilon) = \varepsilon$. In other words, the event $\{\mathbf{L} \geqslant L_\varepsilon\}$ is so improbable that it can be ignored in calculations.

Let $\nu_\varepsilon = L_\varepsilon \cdot \|A\|$. Then the following relation holds with probability $1 - \varepsilon$:

$$\frac{|\Delta \boldsymbol{x}|}{|\Delta \boldsymbol{b}|} \cdot \|A\| \equiv \mathbf{L} \cdot \|A\| \leqslant \nu_\varepsilon.$$

Thus, the coefficient $\nu$ can be replaced with $\nu_\varepsilon$ (although, cf. below $\nu_\varepsilon/\nu \ll 1$).

The ratio of $\nu_\varepsilon$ and $\nu$ is expressed as follows:

$$\frac{\nu_\varepsilon}{\nu} = \frac{L_\varepsilon\|A\|}{\lambda_1\|A\|} = \frac{L_\varepsilon}{\lambda_1}.$$

We compute $\nu_\varepsilon/\nu$ for the sequences $\Lambda^1$ and $\Lambda^2$ (cf. Figure 2) with $\varepsilon = 0,01$:

$$\Lambda^1: \quad L_\varepsilon = 0,2481, \quad \lambda = 1024, \quad \frac{\nu_\varepsilon}{\nu} = 0,00024,$$

$$\Lambda^2: \quad L_\varepsilon = 0,1411, \quad \lambda = 1, \quad \frac{\nu_\varepsilon}{\nu} = 0,14.$$

The values of $\frac{\nu_\varepsilon}{\nu}$ in the first and second cases are essentially different, which could be caused by the fact that the sequence $1,01^{-k}$ $(\Lambda^2)$ converges to zero much slower than $2^{-k}$ $(\Lambda^1)$.

## References


1. R. Bellman, *Introduction to Matrix Analysis*, McGrow-Hill, New York etc. (1960).

2. W. Feller, *An Introduction to Probability Theory and Its Applications*, John Wiley and Sons, New York etc. (1970).

3. G. M. Fikhtengolts, *Course of Differential and Integral Calculus* [in Russian], Nauka, Moscow (1977).


Submitted on January 18, 2017