Stochastic Comparison of Ellipsoidal and Interval Error Estimation in Vector Operations

Alexander Ovseevich

1 Introduction

Consider the following elementary problem of numerical linear algebra. Suppose we are given a vector $x \in \mathbf{R}^n$, not known exactly but located within a known bounded domain Ω, and a matrix A which is known exactly. We would like to localize the vector Ax as good as possible. Certainly, Ax is contained in $A\Omega$, and that's the best one can say. In practice this answer may be not good enough, since it might be unfit for computer. In particular, the domains Ω of uncertainty should have a simple description, that would allow to check easily (for a computer) whether a given vector is contained in it.

There are at least two classes of suitable domains: boxes $\mathcal{B} = \{x \in \mathbf{R}^n : |x_i - a_i| \leq b_i\}$, and ellipsoids $\mathcal{E} = \{x \in \mathbf{R}^n : \langle Q^{-1}(x-a), x-a\rangle \leq 1\}$. Methods of computations with vectors, localized in boxes, are known as interval analysis, similar methods for vectors, localized in ellipsoids, are known as ellipsoidal analysis.

The present paper is inspired by [4], where some evidences are presented that in the problem of multiplication of a vector by matrix the ellipsoidal analysis is, in certain sense, better than the interval one. More precisely, suppose the vector is localized in a box $\mathcal{B}$, and $\mathcal{E}$ is the minimum volume ellipsoid containing $\mathcal{B}$. Certainly, $\mathcal{E}$ also localizes the vector, and, at this stage, the substitution of $\mathcal{E}$ for $\mathcal{B}$ results in a loss of accuracy. However, upon multiplication by A the domain $A\mathcal{B}$ is not necessarily a box, while the domain $A\mathcal{E}$ is still an ellipsoid. To stay within the interval framework one should substitute the minimal box $\mathrm{Box}(A\mathcal{B})$, containing $A\mathcal{B}$, for $A\mathcal{B}$. Finally we get two localization domains: $\mathrm{Box}(A\mathcal{B})$ and $A\mathcal{E}$. It is suggested in [4] to compare the quality of methods by means of volumes of the final localization domains.

1.1 Main inequality

The result of comparison does not depend on the initial box, but only on the matrix A, and is determined by the sign $\begin{matrix}\leq\\ \geq\end{matrix}$ in the inequality

$$\prod_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|\begin{matrix}\leq\\ \geq\end{matrix}\frac{(\pi n)^{\frac{n}{2}}\,|\det A|}{2^n\,\Gamma\left(\dfrac{n}{2}+1\right)}. \tag{1.1}$$

The $\leq$ sign specifies the set of matrices such that the ellipsoidal method turns to be worse than the interval one. The inequality (1.1) comes directly from exact formulas for volumes of ellipsoid and box, while the factor $\pi^{n/2}/\Gamma(\frac{n}{2}+1)$ arises as volume of the circumscribed ball for unit cube. Real problems of numerical linear algebra correspond to a large dimension n. That's why we will compare ellipsoids and boxes as $n\to\infty$.

2 Random matrices

The set Ω_n of $n\times n$-matrices A such that (1.1) holds with $\leq$ sign is quite complicated. In a rather vague way, one can say that Ω_n is relatively poor, i.e. most matrices do not belong to it. Still it is not clear in advance how to measure properly the size of Ω_n, and establish that it is small. We suggest a stochastic approach to this issue. Namely, we assume that the matrix A is random so that its elements are independent Gaussian random variables with zero mean and unit covariance. In particular, the distribution of any element a_{ij} of A takes the form

$$p(x)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right). \tag{2.1}$$

Then, a natural measure for the size of the set Ω_n is its probability $\mathbf{P}(\Omega_n)=(2\pi)^{-n^2/2}\int_{\Omega_n}e^{-\frac{1}{2}\operatorname{Tr}AA^*}dA$. Here, $\operatorname{Tr}AA^*=\sum a_{ij}^2$, and $dA=\prod da_{ij}$.

3 Main result

Theorem 1 *The probability of the event Ω_n that intervals are better than ellipsoids tends to zero as $n\to\infty$. In other words,*

$$\mathbf{P}(\Omega_n)=(2\pi)^{-n^2/2}\int_{\Omega_n}e^{-\frac{1}{2}\operatorname{Tr}AA^*}dA=o(1) \tag{3.1}$$

More precisely, $\mathbf{P}(\Omega_n)=O(1/(n^2\log n))$.

Denote $\log \sum\limits_{j=1}^{n} \left|\frac{a_{ij}}{\sqrt{n}}\right|$ by χ_i, and define

$$\psi_n = \frac{1}{n}\sum_{i=1}^{n} \chi_i,\ C_n = \log \frac{(\pi n)^{\frac{1}{2}}}{2\,\Gamma\left(\frac{n}{2}+1\right)^{1/n}},\ \Delta_n = \frac{1}{n}\log\left|\det\left(\frac{A}{\sqrt{n}}\right)\right|.$$

The set Ω_n is defined by inequality

$$\psi_n \le C_n + \Delta_n \tag{3.2}$$

and our main result says that this inequality holds with a very small probability. The reason is that each term in (3.2) has a definite, and even deterministic "limit in probability" as $n \to \infty$: $\psi_n = \frac{1}{2}\log\frac{2n}{\pi} + o(1)$, $C_n = \frac{1}{2}\log\frac{\pi e}{2} + o(1)$, $\Delta_n = -\frac{1}{2} + o(1)$, but the limit inequality $\frac{1}{2}\log\frac{2n}{\pi} \le \frac{1}{2}\log\frac{\pi e}{2} - \frac{1}{2}$ is totally false. In what follows we expound the above arguments.

3.1 A heuristic analysis of inequality (3.2)

The functions χ_i can be regarded as independent random variables on the Gaussian probability space of $n \times n$-matrices, and the left-hand side of (3.2) has a form of a mean value $\psi_n = \frac{1}{n}\sum\limits_{i=1}^{n}\chi_i$. Hence, when $n \to \infty$ one can apply the Law of Large Numbers (LLN) to analyze the left-hand side of (3.2) and conclude that

$$\psi_n = \frac{1}{n}\sum_{i=1}^{n}\chi_i \to \mathbf{E}\chi_1 \text{ in probability.} \tag{3.3}$$

By virtue of the Central Limit Theorem (CLT) the distribution of $f_i = \sum\limits_{j=1}^{n}\left|\frac{a_{ij}}{\sqrt{n}}\right|$ is approximately Gaussian with covariance $1 - \frac{2}{\pi}$ and mathematical expectation $\sqrt{\frac{2n}{\pi}}$. Therefore,

$$\mathbf{E}\chi_1 = \mathbf{E}\log f_1 = \log\sqrt{\frac{2n}{\pi}} + o(1) \tag{3.4}$$

and ψ_n is contained in $o(1)$-neighborhood of $\log\sqrt{\frac{2n}{\pi}}$ with probability $1+o(1)$. Hence, if n is large the inequality (3.2) with an overwhelming probability takes the form

$$\Delta_n \ge \log\sqrt{\frac{2n}{\pi}} + \frac{1}{2}\log\frac{\pi e}{2} + o(1) = \frac{1}{2}\log n + \frac{1}{2} + o(1), \tag{3.5}$$

where ψ_n is absent. Thanks to Siegel [1, 2] we have explicit expression

$$\mathbf{E}|\det A|^k = (2\pi)^{-n^2/2} \int |\det A|^k \, e^{-\frac{1}{2}\operatorname{Tr}(AA^*)} dA = 2^{\frac{kn}{2}} \prod_{i=1}^{n} \frac{\Gamma\left(\frac{k+i}{2}\right)}{\Gamma\left(\frac{i}{2}\right)} \tag{3.6}$$

for $\mathbf{E}|\det A|^k$ with any complex k. In particular, it follows from (3.6) that $\mathbf{E}e^{n\Delta_n} = o(1)$. Therefore, Δ_n can be large only with (exponentially) small probability. In particular, the probability of (3.5) decays as $n \to \infty$.

3.2 Rigorous analysis of the left-hand side of (3.2)

The above arguments tacitly assume that some limit processes commute. We will not justify exactly this, and use subgaussian random variables instead of CLT.

A real random variable is said to be subgaussian if

$$\mathbf{E}e^{\lambda\xi} \le e^{\frac{1}{2}\lambda^2}$$

for any real λ. The fact which is very important for us is this:

Theorem 2 *If x is a standard Gaussian random variable, then, the random variable $\xi = |x| - \mathbf{E}|x|$ is subgaussian.*

A proof is based on the so-called theory of logarithmic concavity [5]. This immediately implies the following corollary.

Corollary 1 *Each random variable $f_i - \sqrt{\frac{2n}{\pi}} = \sum_{j=1}^{n} \frac{|a_{ij}| - \mathbf{E}|a_{ij}|}{\sqrt{n}}$ is subgaussian.*

On the basis of this corollary one can show that

$$\mathbf{E}\log f_i = \log\sqrt{\frac{2n}{\pi}} + o(1) \tag{3.7}$$

$$\mathbf{E}\log^2 f_i = \log^2\sqrt{\frac{2n}{\pi}} + o(1) \tag{3.8}$$

$$\mathbf{E}\left|\log f_i - \mathbf{E}\log f_i\right|^2 = O(\frac{\log n}{n}) \tag{3.9}$$

In particular, the asymptotic equality (3.4) holds, and LLN can be applied in order to justify (3.3). Thus, $\psi_n - \log\sqrt{\frac{2n}{\pi}} \to 0$ in a reasonable sense.

Finally, from (3.9), (3.4) and the Chebyshev inequality we obtain:

$$\mathbf{P}\left(\psi_n \le \frac{1}{4}\log n + C\right) = o(1), \tag{3.10}$$

where C is an arbitrary constant, while $o(1)$ is, in fact, $O(\frac{1}{n^2 \log n})$. Therefore, ψ_n is large with a large probability.

3.3 Analysis of the right-hand side of (3.2)

As to the random variable $\Delta_n = \frac{1}{n}\log\left|\det\left(\frac{A}{\sqrt{n}}\right)\right|$ in the right-hand side of (3.2), it is not large with an overwhelming probability. In fact, one can show that $\Delta_n \to -\frac{1}{2}$ in probability so that the (absolute value of) determinant of a random matrix becomes more and more deterministic as $n \to \infty$.

It follows from the Siegel formula (3.6) that

$$\mathbf{E}e^{n\Delta_n} = \left(\frac{2}{n}\right)^{\frac{n}{2}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(1/2\right)}. \tag{3.11}$$

The logarithm of the right-hand side of (3.11) is

$$-\frac{n}{2} - \frac{1}{2} + o(1)$$

in view of the Stirling formula. Now, by means of the Chebyshev inequality we can estimate the probability of large values of Δ_n as follows:

$$\mathbf{P}(\Delta_n \geq x) \leq e^{-nx}\mathbf{E}e^{n\Delta_n} \sim e^{-nx-\frac{n}{2}-\frac{1}{2}}.$$

In particular,

$$\mathbf{P}\left(\Delta_n \geq \frac{1}{4}\log n + C\right) \leq (1+o(1))e^{-\frac{1}{4}n\log n}, \tag{3.12}$$

where C is an arbitrary constant.

3.4 Summing up

Now we get back to inequality (3.2), where $C_n \to C = \frac{1}{2}\log\frac{\pi e}{2}$ by the Stirling formula. If the inequality (3.2) holds for a large n, then either $\psi_n \leq \frac{1}{4}\log n + C + 1$, or $\Delta_n \geq \frac{1}{4}\log n - C - 1$. But, in view of (3.10), (3.12) these events have small probabilities as $n \to \infty$. This proves the main Theorem 1 to the effect that probability of advantage of intervals over ellipsoids is small as $n \to \infty$. In fact, it is shown that this probability is $O(1/(n^2\log n))$. These considerations can be regarded as an evidence in favor of ellipsoids vs. boxes in linear algebraic computations with a guaranteed accuracy.

References

[1] *Siegel, C.L.* Über die analytische Theorie der quadratischen Formen, Ann. Math. Ser. 2, **36** (1935), 527–606

[2] *Godement, R.*, Fonctions holomorphes de carré sommable dans le demi-plan de Siegel, Seminaire H. Cartan, Ecole Normale Supérieure, Paris, 1958.

[3] *Helgason, Sigurdur*, Differential Geometry and Symmetric Spaces, Academic Press, New York London 1962

[4] Felix L. Chernousko, Alexander J. Ovseevich, and Yuri V. Tarabanko, Comparison of Interval and Ellipsoidal Bounds for the Errors of Vector Operations, Doklady math. 2005, v. 71, 1, 127-130

[5] Barry Simon, Functional Integration and Quantum Physics, Academic Press New York San Francisco London 1979