Roman DMYTRYSHYN

Rzeszow University of Technology

Condition Number for a Matrix

Abstract. *A new more accurate formula to calculate condition number of a matrix is proposed. The effectiveness of the new formula was confirmed by the examples of the Hilbert matrix. In order to validate the new approach comparison with statistical Monte Carlo calculation was used.*

Streszczenie. *Opracowano nowy dokładniejszy wzór do obliczenia wskaźnika uwarunkowania macierzy. Skuteczność nowego wzoru potwierdzono na przykładach macierzy Hilberta. Dla sprawdzenia dokładności wzorów wykorzystano analizę Monte-Carlo.* (***Współczynnik uwarunkowania macierzy***).

Keywords: condition number of matrix, Hilbert's matrix, standard deviation, Fresenius's norm.
Słowa kluczowe: wskaźnik uwarunkowania macierzy, macierz Hilberta, standardowe odchylenie, norma Frobeniusa

Introduction

As per [1], condition number of the square nonsingular matrix $\mathbf{A}$ of $\mathbf{A}\cdot\mathbf{X}=\mathbf{B}$ equation (SLAE) shows the magnitude of the error increase of the calculation of $\mathbf{X}$ vector having the value of inaccuracy of $\mathbf{B}$ vector and inaccuracies of the entries of matrix $\mathbf{A}$. More precisely, condition number shows the maximum relation of the inaccuracies of the solution vector $\mathbf{X}$ relative to the value of the relative inaccuracy of the vector $\mathbf{B}$. Unlike the rounding error that is introduced by the numeric algorithm, condition number is an indicator of inaccuracies that are introduced by input data.

In the numeric methods of applied mathematics condition number is playing important role. Extended notation of the value of condition number is given in the form of the product of norm of original ($||\mathbf{A}||$) and inverse ($||\mathbf{A}^{-1}||$) matrices or relationship of eigenvalues of matrix $\mathbf{A}$.

$$\nu(\mathbf{A}) = ||\mathbf{A}|| \cdot ||\mathbf{A}^{-1}|| = |\boldsymbol{\lambda}_{\max}| / |\boldsymbol{\lambda}_{\min}| \tag{1}$$

The bigger value of $\nu(\mathbf{A})$, the bigger instabilities occur while solving SLAE. Let us note that eventually (1) is giving the possibility to estimate the top limitation of the inaccuracy.

The above belongs to the classic knowledge, which is described in different forms in the numeric methods textbooks. None less during the calculation of (1) different values are obtained due to the use of different norm. In table 1 the few examples of the calculation in MAPLE V5 R4 of condition number (cond) of Hilbert matrices of rank 5, 8, 12, 50 and 100 are shown. We have: $|\boldsymbol{\lambda}_{max}/\boldsymbol{\lambda}_{min}|$, condF (Frobenius's norm) and condI (infinity norm).

Table 1. Example one of condition number of Hilbert matrix

H	power	$\lambda_{max}/\lambda_{min}$	condF	condI
$H_{5\times5}$	$\times10^{4}$	47.661..	48.085..	94.366..
$H_{8\times8}$	$\times10^{8}$	152.58..	154.94..	338.73..
$H_{12\times12}$	$\times10^{16}$	1.7132..	1.7518..	4.1155..
$H_{50\times50}$	$\times10^{71}$	1422.9..	1500.9..	4330.3..
$H_{100\times100}$	$\times10^{148}$	* * *	405.37..	1267.2..

* * * - very long CPU time

As we see above especially differ norm condF and condI. As rank of Hilbert matrix is increasing this difference is growing. The difference for matrix $H_{100\times100}$ for condF and for condI is almost 3 times. The search for the choice of "the best" matrix norm has proven to be in vain…Obviously that it is preferable to choose the lowest value between different values of condition number. Therefore, the purpose of this work is to develop a more accurate formula for the condition number calculation.

The alternative solution

The source for the derivation of the alternative formula is a known dependency of the standard variation of the function of many variables from standard variations of it variables [3]. Let $y = f(x1, x2,\ldots,xn)$. $\sigma_{x1}, \sigma_{x2}, \ldots, \sigma_{xn}$ are given as well. Let us assume that variables and their derivations are independent. Then, after decomposition of function y into Taylor's series and ignoring the members of the higher rang, we are getting the formula of the standard derivation of the y function:

$$\sigma_y = \sqrt{\sum_{i=1}^{n}\left(\frac{\partial y}{\partial x_i}\cdot\sigma_{x_i}\right)^2} \tag{2}$$

Let us assign: $\sigma_{xi} = x_i\cdot\delta_{xi}$, where $\cdot\ \delta_{xi}$ is a standard relative deviation or the independent variable x_i. Then (2) will look:

$$\sigma_y = \sqrt{\sum_{i=1}^{n}\left(\frac{\partial y}{\partial x_i}\cdot x_i\cdot\delta_{x_i}\right)^2} \tag{3}$$

In practice the standard deviations of the relative errors of all variables are the same, therefore, let $\delta_{x1} = \delta_{x2}\ldots = \delta_x$. Then (3) will become:

$$\sigma_y = \delta_x\cdot\sqrt{\sum_{i=1}^{n}\left(\frac{\partial y}{\partial x_i}\cdot x_i\right)^2} \tag{4}$$

Let us show (4) accordingly to the function y in the form $\det\mathbf{A}$ as determinant of the nonsingular matrix $\mathbf{A} = [a_{ij}]_{nxn}$ $_{i,\,j\,=\,1..n.}$ where a_{ij} is a real or complex number

$$\sigma_{\det\mathbf{A}} = \delta_a\cdot\sqrt{\sum_{i=1}^{n}\left(\frac{\partial(\det\mathbf{A})}{\partial a_{ij}}\cdot a_{ij}\right)^2} \tag{5}$$

From the products inside parenthetic we will create square matrix $\mathbf{C}$ with elements $c_{ij} = \mathbf{M}_{ij}\cdot a_{ij}$, where $\mathbf{M}_{ij}$ is corresponding minor of the matrix $\mathbf{A}$. The sign of the minor is irrelevant here as later during the calculation of the norm the signs of the squares are always positive. Now let us rewrite (5), using matrix $\mathbf{C}$.

$$\sigma_{\det A} = \delta_a \cdot \sqrt{\sum_{i=1}^{n} c_{ij}{}^2} = \delta_a \cdot \|\mathbf{C}\|_F \tag{6}$$

where $||\mathbf{C}||_F$ is the Fresenius's norm of the matrix $\mathbf{C}$.

For the clarity sake of the explanation of $\nu(\mathbf{A})$ the formula for the estimation of the lost decimal digits of mantissa during the evaluation of the determinant of matrix $\mathbf{A}$ of arbitrary method [2] is used

$$L = \log_{10}(\nu(\mathbf{A})) \quad , \tag{7}$$

where L means the number of younger decimal digits of the mantissa of the determinant, that during its calculation become inaccurate. It is assumed that before the calculation of the determinant all *md* digits of mantissa of the matrix coefficients are exact $\delta_x = 10^{-md}$. Obviously, that after the calculation of the determinant the number of the accurate digits Q of mantissa is

$$Q = md - L = \log_{10} \frac{|\det \mathbf{A}|}{\sigma_{\det A}} = \log_{10} \frac{|\det \mathbf{A}|}{10^{-md} \cdot \|\mathbf{C}\|_F}$$

Having matrix $\mathbf{C}$, using the same way as in (7) we can determine the quantity of the lost digits during the calculation of the determinant of the arbitrary nonsingular square matrix $\mathbf{A}$.

$$L = md - \log_{10} \frac{\det \mathbf{A}}{10^{-md} \cdot \|\mathbf{C}\|_F} = \log_{10}(\nu(\mathbf{A})) \tag{8}$$

and from (7) we obtain

$$\nu(\mathbf{A}) = \text{condT} = \frac{\|\mathbf{C}\|_F}{|\det \mathbf{A}|} \tag{9}$$

Table 2 contains comparison of the results of the calculations of condition number of the Hilbert matrices using: $|\lambda_{max}| / |\lambda_{min}|$, identify norm (condI), (condT) from (9) and with Monte-Carlo calculation (condS). As we see, the values of condI and condT significantly differ.

Table 2. Second example of condition number of Hilbert matrix

Hilbert's matrix	power	$\lvert\lambda_{max}/\lambda_{min}\rvert$ (MAPLE)	condI (MAPLE)	condT (MAPLE)	condS (DELPHI)	Lost*
$H_{5\times5}$	$\times10^{4}$	47.661..	94.366..	4.6781..	4.6809..	4.67
$H_{8\times8}$	$\times10^{8}$	152.58..	338.73..	8.3703..	8.3726..	8.923..
$H_{12\times12}$	$\times10^{14}$	171.32..	411.55..	5.6787..	5.8304..	14.754
$H_{50\times50}$	$\times10^{71}$	1422.9..	4330.3..	9.7697..	* *	71.99
$H_{100\times100}$	$\times10^{148}$	* * *	1267.2..	1.2283..	* *	148.09

* - from *condT,*

** - short mantissa in Turbo DELPHI.

The "Lost" column shows the quantity of lost decimal digit numbers of mantissa during the determinant calculation as given by MAPLE software. As we see for matrix $\mathbf{H}_{50x50}$ at the minimum 80_{10} - digits arithmetics is required. $\mathbf{H}_{100x100}$ requires 160_{10} digits. The usage of *extended* precision in Turbo DELPHI allows precise calculation of the determinant Hilbert's matrix up to n =12.

Computer experiments

In order to verify adequate (closest to the exact) value of the condition number it is necessary to choose a referee, in other words method of calculating of the condition number, reliability and truthfulness of which is absolute. Such a referee can be found in the method of multivariate calculations based on statistical approach [3]. Obviously the "worse" the matrix is - the bigger is the standard derivation of determinant.

In order to implement Monte-Carlo method here it is necessary to choose the model of the discrete density of the random values distribution of matrix elements. Let us use the normal distribution of those values. The calculation of the mode and standard deviation will be done accordingly to the classic formulae [3]. Due to the unsatisfactory quality of the standard generator of the normal random numbers that was found in included library of the DELPHI software the author has developed generator with better quality. The main idea of the design of the better generator is based on optimization of the limited quantity ($750 \cdot 10^6$) of generated numbers. Those numbers obtained from a standard generator of uniform distributed numbers and discrete values of inverted cumulative distribution of normal numbers. The verification of the generator quality is based on the analysis of the random values and histogram of the density of the distribution (asymmetry, excess and χ^2).

As example we choose above mentioned Hilbert matrix $\mathbf{H}$ with elements $h_{ij} = 1/(i+j-1)$ $_{i,j = 1..n}$. The calculation of ill conditioned Hilbert matrix is an accepted test of the numerical methods as of today [2]. The experiments were done with the help of custom developed by author software using PASCAL language in Turbo-DELPHI environment, and also with analytical program MAPLE V5 R4. Calculations of determinant in PASCAL were made using the maximum precision *extended* (~22_{10} digits). Calculation of every determinant was made by the known formula of Gauss reduction [1]

$$a_{i,j}^{(k)} = a_{i,j}^{(k-1)} - \frac{a_{i,r}^{(r-1)} \cdot a_{r,j}^{(r-1)}}{a_{r-1,r-1}^{(r-1)}} \quad k = 1..n\text{-}1,\ i,j = k+1..n \tag{10}$$

which allows bringing the triangular matrix, determinant of which is the multiplication of its diagonal elements. Pivot-element was not selected.

After many tries the quantity of the variants of the calculations was chosen to be 1 million. In every calculation of the determinant matrix elements have got the discrete value accordingly to the normal distribution in $\pm3\sigma$ range. In the mantissa of the matrix elements the change was occurring from 15th position (md = 15), i.e. $\delta_{aij} = 10^{-md}$. Due to the above precautions the negative influence of rounding errors on the results of statistical experiments was avoided.

As the first step 10^6 variants were calculated using the Monte-Carlo method. This was done to get the range of the determinant values (from h_{min} till h_{max}), and also the closer to the exact value that was previously determined with MAPLE. As the second step the same 10^6 determinants were calculated to get histogram with 21 columns (Fig. 1).

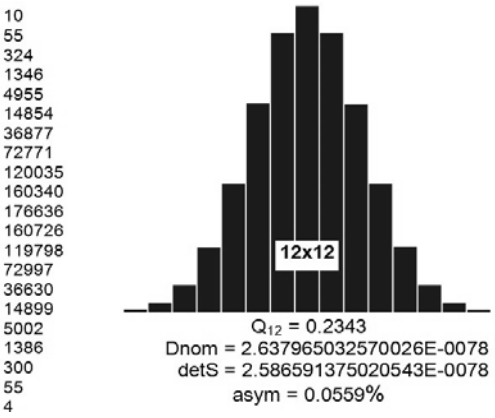

Fig. 1. An example of histogram for det $\mathbf{H}_{12x12}$

As the third step analysis of the histogram in order to calculate both the mode of determinant value (det **A)** and derivation of determinant (σ) was done. On Fig 1 whole numbers to the left of histograms show the quantity of the determinant values in the certain range (1 - 21) from h_{min} till h_{max}. Q_{12} mean the numbers of the reliable digits of determinant mantissa. D_{nom} represents exact value of the determinant (from MAPLE), DetS – mode of the determinant. Below we can see (asym) the value of the coefficient of the histogram asymmetry calculated as the ratio of the sum of heights of left columns to the sum of the right columns (in percents). Let us note that without using the improved generator of normal numbers histograms were noticeably asymmetrical. The column of condS in table 2 was obtained after appropriate histogram analysis.

From the Monte-Carlo calculations we obtain following formula for calculating of the condition number of matrix

$$\text{condS} = 10^{md-Q}$$

where $Q = \log_{10}(|\det\mathbf{A}| / 3\sigma)$.

After Monte-Carlo experiments we can see that condS from table 2 is almost the same as the condT. This is a reliable proof of the advantage of the proposed new formula (9) of the calculation of condition number of matrix. Calculation abilities of the numeric processor have limited multivariate research in PASCAL of the Hilbert matrix to 12th order.

The new formula (9) (condT) is universal and is suitable for the calculation of condition number of matrix for any square nonsingular matrix. Here are two matrices $\mathbf{A1}$ and $\mathbf{A2}$.

$$\mathbf{A1}=\begin{vmatrix} 1 & 2 & 3 \\ 7 & 5 & 4 \\ 9 & 8 & 6 \end{vmatrix} \qquad \mathbf{A2}=\begin{vmatrix} 3 & 2 & 1 \\ 7 & 5\cdot 10^{5} & 4 \\ 9 & 8 & 6 \end{vmatrix}$$

Table 3 shows the results of calculations of the condition numbers.

Table 3. The condition number for the matrix **A1** and **A2**

matrix	det	$\lvert\lambda_{max}\rvert/\lvert\lambda_{min}\rvert$ (MAPLE)	condF (MAPLE)	condT (MAPLE)	condS (DELPHI)
A1	19	14.326..	31.602..	11.325..	11.346..
A2	$4.5\cdot 10^{6}$	$4.363\cdot 10^{5}$	$6.261\cdot 10^{5}$	3.317..	3.279..

The observation of Table 3 reveals that the condition number of the matrix calculated according by the classical formula (1) differs significantly from the condition number obtained from the new formula of condT (9). The correct value of condT confirmed by Monte-Carlo calculation (condS).

Comparison of algorithms for the calculation of the determinant

By means of computer experiments the accuracy of the calculation of determinant using formula (10) and classic permutation by Leibnitz formula (11) without divisions were compared

$$\det\mathbf{A} = \sum(-1)^{\text{Inv}(k)}\cdot a_{1,k_1}\cdot\ldots\cdot a_{n,k_n}, \quad k_i \neq k_j;\ i, j, k_{i\neq j}, k_{i\neq j} \in \{1..n\}. \tag{11}$$

It was observed that popular statement about the advantage in accuracy of symbolic methods based on usage of the formulas without divisions (11) over formula with divisions like (10) is not always true. It is sufficient to look at experimental calculations of the determinant of matrices H_{5x5} - H_{7x7}, shown in Table 4. It turns out that «precise» formula (11) posse many of the same digits, highlighted with bold font. That is why in case of ill conditioned Hilbert's matrices using extended precision is not possible to satisfactory calculating of the determinant of the matrix for n > 6.

Table 4. Positive and negative components of determinant of Hilbert matrices according to (11)

+3.41251094831671646.. e-02
–3.41251094794178694.. e-02
detH_{5x5} = 3.7492951180783523.. e-12
+1.57016575114036218.. e-02
–1.57016575114036160.. e-02
detH_{6x6} = 5.3752710832757899.. e-18
+7.1215692969469809.. e-03
–7.1215692969469809.. e-03
detH_{7x7} = 0 !!!

Let us compare the results of calculations of the same determinants using Gauss formula (10) (Table 5).

Table 5. Comparison of the Hilbert's matrix determinant: exact values (MAPLE) and in DELPHI calculated

	detH_{5x5}
MAPLE	**3.749295132515087..**e-12
DELPHI	**3.749295132515086..**e-12
	detH_{6x6}
MAPLE	**5.367299887358**688..e-18
DELPHI	**5.367299887358**365..e-18
	detH_{7x7}
MAPLE	**4.835802623926117**..e-25
DELPHI	**4.835802623926110**..e-25

Table 5 contains exact values of the determinants, calculated by (10) in MAPLE and in DELPHI (PASCAL) using extended precision.

Conclusions

The new formula (9) for the calculation of condition number of matrix using the product of values of the every entry of the matrix $\mathbf{A}$ on its minor was developed. Obtained values of condition number are more accurate comparing to classic values. The accuracy of the new formula was proven with the Monte-Carlo method. The convenience of the usage as value of inaccuracy the number of lost (inaccurate) digits of mantissa (L) versa condition number matrix $\nu(\mathbf{A})$ was shown as added benefit.

REFERENCES

[1] Bakhvalov N. S., Numerical methods, Moscow, "Nauka", 1975, s.632. (in Russian)

[2] James Kesling J. The Condition Number for a Matrix, www.math.ufl.edu/~kees/ConditionNumber.pdf

[3] Dorozhovets M., Processing of the Measurement Result., Lviv Polytechnic National University, Ukraine, 2007, s. 624 (in Ukrainian)

***Author**: prof. PRz, dr hab. inż. Roman Dmytryshyn, Politechnika Rzeszowska, Wydział Elektrotechniki i Informatyki, 35-959 Rzeszów, ul. W. Pola 2, E-mail: rdmytr@prz.rzeszow.pl*