Turing's Paper on Rounding-Off Errors in Matrix Processes

Garry J Tee

Before 1940, solving a set of n linear equations with $n \geq 10$ was regarded as a very challenging problem. Several direct methods were used for solving linear equations, but only during the 1940s did mathematicians realise that most of those direct methods were closely related to Gaußian elimination, and also to the factorisation of a square matrix as a product of lower and upper triangular matrices. Some linear problems in physics and engineering had been solved by indirect methods of successive approximation, which later were systematised as various iterative methods.

As more powerful calculating machines were developed, some scientists began solving larger sets of linear equations. But serious concern was felt about the effects of round-off in the computation, which might produce large errors in the computed solution for $n > 10$ (von Neumann & Goldstine [5, footnote 11]). At a conference in 1941, Harold Hotelling argued (Hotelling [2]) that in solving a set of n equations the calculation should be done with $n \log_{10} 4$ extra figures — for each type of machine that implied severe limitations on the size of linear systems which could be solved.

In 1947, John von Neumann and Herman H Goldstine published their very detailed study of "Numerical inverting of matrices of high order", in which they concentrated mainly on inversion of symmetric positive-definite matrices. They expressed upper bounds for the error of the computed inverse in terms of "various quantities and properties which cannot be supposed to be known when the problem of inverting a matrix $\overline{A}$ or $\overline{A}_I$ comes up", including the maximum and minimum moduli of eigenvalues of the matrix (von Neumann & Goldstine [5, p. 1089]). They anticipated that, when electronic digital computers become available for inverting matrices, then $n \sim 100$ will become manageable (von Neumann & Goldstine [5, p. 1031]).

They remarked that "An approximate inverse of a matrix $\overline{P}$ might be defined as one which lies close to the exact inverse $\overline{P}^{-1}$. From the point of view of numerical procedure it seems more appropriate, however, to interpret it as the inverse P'^{-1} of a matrix P' that lies close to $\overline{P}$ that is, to permit an uncertainty of, say, ϵ in every element of $\overline{P}$". (von Neumann & Goldstine [5, p. 1092]). That seems to be the earliest published mention of backward error analysis.

In 1948, Alan Mathison Turing published his influential paper on "Rounding-off errors in matrix processes" (Turing [4]). At the National Physical Laboratory he had collaborated with Leslie Fox, Harry D Huskey and James H Wilkinson in extensive tests of the elimination method on many sets of linear equations (Fox, Huskey & Wilkinson [1]). He reported that "Fox found that no exponential build-up of errors such as that envisaged by Hotelling actually occurred. In the meantime another theoretical investigation was being carried out by J von Neumann, who reached conclusions similar to those of this paper for the case of positive definite matrices, and communicated them to the writer at Princeton in January 1947 before the proofs given here were complete" (Turing [4, p. 288]).

If a set of linear equations $\mathbf{Ax} = \mathbf{b}$ is to be solved for a single vector $\mathbf{b}$ then the computational cost is about one-third of that for inverting $\mathbf{A}$ and then evaluating $\mathbf{x}$ as $\mathbf{A}^{-1}\mathbf{b}$; but if solutions are required for several vectors $\mathbf{b}$ then $\mathbf{A}$ should be inverted. Turing considered that "It seems probable that with the advent of electronic computers it will become standard practice to find the inverse". However, good computing practice still uses solution (and not inversion) for the usual case of a single vector $\mathbf{b}$. Turing gave a constructive proof that if the principal minors of the matrix $\mathbf{A}$ are non-singular then $\mathbf{A}$ factorises as $\mathbf{A} = \mathbf{LDU}$ uniquely, where $\mathbf{L}$ is a unit lower triangle, $\mathbf{D}$ is a diagonal matrix and $\mathbf{U}$ is a unit upper triangle. (Likewise $\mathbf{A} = \mathbf{U'D'L'}$ uniquely, where $\mathbf{L'}$ is a unit

lower triangle, **D′** is a diagonal matrix and **U′** is a unit upper triangle.) (Turing, [4, p. 289]). He then used the **LDU** factorisation to construct the Gaußian elimination method, Jordan's method for inversion, Morris's escalator method and the orthogonalisation method for constructing an upper triangular matrix **M** such that **AM** is orthogonal.

The transpose of a real matrix **A** is denoted by $\mathbf{A}^*$. Choleski's method is recommended for symmetric **A**, constructing the lower triangular matrix $\mathbf{LD}^{1/2}$, which "may involve some purely imaginary numbers, but no strictly complex ones" (Turing [4, p. 295]). He did not mention the simpler real symmetric rational Choleski factorisation $\mathbf{A} = \mathbf{L}^*\mathbf{DL}$, in which some diagonal elements of **D** will be negative if **A** is not positive definite; and he did not discuss the numerical stability of the factorisation in that case.

Turing considered three ways in which the magnitude of a matrix may be measured by a real number (Turing [4, p. 297])

The norm $N(\mathbf{A}) \stackrel{\text{def}}{=} (\text{trace } \mathbf{A}^*\mathbf{A})^{1/2} = \sqrt{\sum_{i,j} a_{ij}^2}$.

The maximum expansion

$$B(\mathbf{A}) \stackrel{\text{def}}{=} \max_{\mathbf{x}} \frac{|\mathbf{Ax}|}{|\mathbf{x}|} = \max_{\mathbf{x}} \frac{\sqrt{(\mathbf{Ax},\, \mathbf{Ax})}}{\sqrt{(\mathbf{x},\, \mathbf{x})}}\,.$$

The maximum coefficient $M(\mathbf{A}) \stackrel{\text{def}}{=} \max_{i,j} |a_{ij}|$.

Of those three measures, $N(\mathbf{A})$ and $B(\mathbf{A})$ are probably of greatest theoretical significance. But $B(\mathbf{A})$ is difficult to compute, and so Turing dealt chiefly with the much simpler measure $M(\mathbf{A})$. He listed several inequalities relating those three measures.

Ill-conditioned matrices and equations are considered in [4, Sec. 8]. "The expression 'ill-conditioned' is sometimes used merely as a term of abuse applicable to matrices or equations, but it seems most often to carry a meaning somewhat similar to that defined below.

Consider the equations

$$\left.\begin{aligned} 1 \cdot 4x + 0 \cdot 9y &= 2 \cdot 7 \\ -0 \cdot 8x + 1 \cdot 7y &= -1 \cdot 2 \end{aligned}\right\} \qquad (8.1)$$

and form from them another set by adding one-hundredth of the first to the second, to give a new equation replacing the first

$$\left.\begin{aligned} -0 \cdot 786x + 1 \cdot 709y &= -1 \cdot 173 \\ -0 \cdot 800x + 1 \cdot 700y &= -1 \cdot 200 \end{aligned}\right\}. \qquad (8.2)$$

The set of Eq. (8.2) is fully equivalent to (8.1), but clearly if we attempt to solve (8.2) by numerical methods involving round-off errors, we are almost certain to get much less accuracy than if we worked with Eq. (8.1). We should describe the Eq. (8.2) as an *ill-conditioned* set, or, at any rate, as ill-conditioned compared with (8.1). It is characteristic of ill-conditioned sets of equations that small percentage errors in the coefficients given may lead to large percentage errors in the solution" ([4, pp. 297–298]).

He defined $N(\mathbf{A})\,N(\mathbf{A}^{-1})/n$ as *the N-condition number of* **A**, and he defined $nM(\mathbf{A})\,M(\mathbf{A}^{-1})$ as *the M-condition number of* **A**. Nowadays the standard condition number of **A** is $\kappa(\mathbf{A}) \stackrel{\text{def}}{=} \sqrt{\lambda_n/\lambda_1}$ where λ_1 and λ_n are the minimum and maximum eigenvalues of $\mathbf{A}^*\mathbf{A}$, and that can be represented as $\kappa(\mathbf{A}) = B(\mathbf{A})B(\mathbf{A}^{-1})$.

The problem of inverting a general (non-singular) matrix **A** can be reduced to the inversion of a positive definite matrix, since $\mathbf{A}^*\mathbf{A}$ is positive definite and $\mathbf{A}^{-1} = (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*$. (von Neumann & Goldstine [5, p. 1056]) and similarly for solving the equation $\mathbf{Ax} = \mathbf{b}$. But Turing pointed out that, as well of the cost of the extra matrix multiplication, such a normalisation makes the equations more ill-conditioned (Turing [4, p. 296]). Indeed, if that process is applied to a positive definite matrix **A** then the condition number of the normalised matrix $\mathbf{A}^*\mathbf{A}$ is the square of the condition number of **A**. Hence, if **A** is ill-conditioned for inversion, then $\mathbf{A}^*\mathbf{A}$ is much more ill-conditioned for inversion.

After **A** has been factorised and used (with round-off) to compute an approximate solution $\mathbf{x}_1$ to the equation $\mathbf{Ax} = \mathbf{b}$, that approximate solution can be refined by computing the "residual" vector $\mathbf{b}_1 = \mathbf{b} - \mathbf{Ax}_1$. Using the existing factorisation of **A** the equation $\mathbf{Ax}_2 = \mathbf{b}_1$ can be solved (with round-off) to give a refined estimate of the solution **x** as $\mathbf{x}_1 + \mathbf{x}_2$. And that refinement could be iterated, until an acceptably small residual vector is found (Turing [4, p. 300]). But Turing did not point out that the residual vector $\mathbf{x}_1$ should be computed with more significant figures than were used in computing $\mathbf{x}_1$.

Turing then analysed the effects of round-off errors in several methods for inverting matrices and solving equations, and he explained that "Our main purpose in this paper is to establish that the exponential build-up of errors need not occur, and this will be proved when we have found one method of inversion where it is absent" (Turing [4, p. 302]).

He gave a statistical analysis of the effect of round-off errors in Jordan's inversion method when "we are working to a fixed number of decimal places both in the reduction of the original matrix to unity and in the building up of the inverse. It is not easy to obtain corresponding results for the case where a definite number of *significant figures* are kept, but we may make some qualitative suggestions. ... In the case of positive definite, symmetric matrices it is possible to give more definite estimates for the case where calculation is limited to a specific number of significant figures. Results of this nature have been obtained by J v Neumann and H H Goldstine" (Turing [4, pp. 304–305]).

He then remarked that "It is instructive to compare the estimates of error given above with the errors liable to arise from the inaccuracy of the original matrix. If we desire the inverse of $\mathbf{A}$, but the figures given to us are not those of $\mathbf{A}$ but of $\mathbf{A} - \mathbf{S}$, then if we invert perfectly correctly we shall get $(\mathbf{A} - \mathbf{S})^{-1}$ instead of $\mathbf{A}^{-1}$, that is, we shall make an error of $(\mathbf{A} - \mathbf{S})^{-1} - \mathbf{A}^{-1}$, i.e. of $(1-\mathbf{A}^{-1}\mathbf{S})^{-1}\mathbf{A}^{-1}\mathbf{S}\mathbf{A}^{-1}$. If we ignore the second-order terms this is $\mathbf{A}^{-1}\mathbf{S}\mathbf{A}^{-1}$". (Turing [4, p. 306]).

Round-off error analysis had been performed by constructing bounds for the error of the rounded result of a single arithmetic operation upon a pair of variables, each of which had initial error bounds (which could be 0). Throughout a lengthy computation the successive error bounds generally increased very rapidly. That so-called forward error analysis became extremely complicated for large calculations (as in Hotelling [3] and von Neumann & Goldstine [5]), and the errors of the final results were usually very much smaller than the computed error bounds. In the previous paragraph, Turing gave one of the earliest examples of backward error analysis, which starts from the actual solution (computed with round-off and truncation errors etc.) of a given equation, and which constructs bounds for a perturbation of the original equation such that the actual computed output satisfies exactly the perturbed equation. That backward error analysis is usually simpler than forward error analysis, and if the original equation includes measured data (which is subject to uncertainty) then such backward error analysis is more physically meaningful than forward error analysis. James H Wilkinson greatly developed backward error analysis, and he demonstrated its effectiveness for many types of problem in his very influential treatise (Wilkinson [6]).

In 1947 Turing's colleagues concluded their joint study of round-off error in many sets of linear equations solved by the elimination method with the following statement: "Hotelling (7) has implied that the building-up error in the elimination methods is so large that these methods can be used only with extreme caution. Turing's analysis shows that this estimate can be reached only in the most exceptional circumstances, and in our practical experience on matrices of orders up to the twentieth, some of them very ill-conditioned, the errors were in fact quite small." (Fox, Huskey & Wilkinson [1, p. 173]).

In the *Collected Works of A M Turing* (1992), Frank D Burgoyne commented that "Turing's paper was one of the earliest attempts to examine the error analysis of the various methods of solving linear equations and inverting matrices. His analysis was basically sound. The main importance of the paper was that it was published at the dawn of the modern computing era, and it gave indications of which methods were 'safe' when solving such problems on a computer". (Turing [4, p. XII]).

References

[1] L. Fox, H. D. Huskey and J. H. Wilkinson, Notes on the solution of algebraic linear simultaneous equations, *Quarterly J. Mech. Appl. Math.* **1** (1948) 149–173.

[2] H. H. Goldstine and J. von Neumann, Numerical inverting of matrices of high order. II, *Proc. Am. Phil. Soc.* **2** (1951) 188–202.

[3] H. Hotelling, Some new methods in matrix calculation, *Ann. Math. Stat.* **14** (1943) 1–34.

[4] A. M. Turing, Rounding-off errors in matrix processes, *Quarterly J. Mech. Appl. Math.* **1** (1948) 287–308. Reprinted (with notes and corrections) in *Collected Works of A. M. Turing*, *Pure Mathematics*, ed. J. L. Britton (North-Holland, Amsterdam, 1992), pp. 41–62.

[5] J. von Neumann and H. H. Goldstine, Numerical inverting of matrices of high order, *Bull. Amer. Math. Soc.* **53** (1947) 1021–1099.

[6] J. H. Wilkinson, *Rounding Errors in Algebraic Processes* (Her Majesty's Stationery Office, London, 1963).

Garry J Tee

University of Auckland, New Zealand
g.tee@auckland.ac.nz

Garry Tee (HonD AUT) studied at Auckland University College, and graduated as MSc (University of New Zealand). In 1957 his first job was as a Computer, with a geophysical prospecting team in the Northwest Australian desert. He analysed geophysical data by cranking a Brunsviga calculating machine and flipping a slide rule. In 1958, he realised that electronic digital computers were going to become extremely important, and so he went to England where he became a mathematical consultant in industry. He started computing on a DEUCE computer, the reduced commercial version of Alan Turing's 1946 design for ACE. In 1964, he became a foundation member of the Department of Mathematics at the University of Lancaster, and in 1968, he joined the Department of Mathematics at the University of Auckland, where he works on numerical analysis and on the history of science. In 1980, he became also a foundation member of the Department of Computer Science.