

That the Condition Number Squares Is an Identity, Not an Approximation

— Half the Significant Digits Are Gone the Moment $A^T A$ Is Formed — For the 8×8 Hilbert Matrix the Digits

Remaining Are **−4.4124**, Which Is to Say None — [Paper 330]

Yuuki YAMAGISHI (山岸 勇輝)

Yamagishi Complexity Torus Lab, Japan

September 2026

— *From thinking to syncing.* —

Abstract

A least-squares problem is often rewritten as $A^T A x = A^T b$ before being solved. This paper asks what that rewriting costs—the answer is exactly half the significant digits, and they are lost as an identity rather than as an approximation. No new mathematical theorem and no new law is claimed. First, $\kappa_2(A^T A)/\kappa_2(A)^2$ is **1.0000000000** for every matrix tried (Section 2). Second, this is the core of the paper. This is not “usually squares”; it follows from the definition of singular values (Section 3). Third, it holds for matrices that are neither square nor symmetric. A 30×11 Vandermonde gives 1.0000000000 (Section 2). Fourth, double precision carries only **15.9546** digits. For the 8×8 Hilbert matrix $\kappa(A^T A) = 10^{20.3670}$ and the digits remaining are **−4.4124** (Section 4). Fifth, there are matrices that do not degrade. An orthogonal matrix keeps $\kappa = 1$ (Section 4). Sixth, the separator is whether $A^T A$ was formed or not (Section 5).

Scope of this paper (scope note): No new mathematical theorem and no new law is claimed——condition numbers, the singular value decomposition, the normal equations, and the Hilbert matrix are all standard. We do not build numerical linear algebra——all we use is one equality and a table confirming it in extended precision. We do not adjudicate between solvers——it is not written here that QR or the SVD “should” be used. Only what is lost is put down as a number. We do no backward error analysis——the general stability of particular algorithms is not entered. It is mentioned and no more. We do not treat regularisation——lowering the condition number artificially, as ridge regression does, is left out. We treat no norm but the 2-norm——the condition number is defined in the 2-norm; in other norms the squaring is no longer an identity. Relation to earlier papers: Paper 326 showed the square-cube “law” to be an identity and not a law——this paper performs the same move in numerical computation. When what is called a “law” follows from a definition, hunting for counterexamples is beside the point. Paper 221 showed that changing the last bit moves an eigenvalue by 0.16——there a perturbed input moves the answer; here a rewriting squares the sensitivity itself. Paper 329 showed divergence and convergence coming out of the same points——there the construction of the approximation fails; here the digits run out. There are at least three separate reasons numbers cannot be trusted. Paper 95 separated convention from fact——the squaring is not a fact but a consequence of how κ is defined. Paper 300 showed whether two things share a root is decidable——by that test, this squaring and the perturbation of Paper 221 have distinct roots. What is added is giving the ratio to ten digits in extended precision, killing the explanation “because it is symmetric” by exhibiting rectangular non-symmetric cases, tabulating digits lost against the digits double precision has, negative values included, and putting the separator on whether $A^\top A$ was formed.

1. Notation and the Question

Setting: Given A of size $m \times n$ with $m \geq n$, and b , find the x minimising $\|Ax - b\|_2$. Rewritten as the normal equations,

$$A^\top A x = A^\top b,$$

one square symmetric system is all that has to be solved——the formula gets shorter and the implementation easier.

Observation 1.1

The condition number $\kappa_2(A) = \sigma_{\max}/\sigma_{\min}$ measures by what factor a relative error in the input reaches the answer. If κ is 10^k , about k digits are lost——this is the standard rule of thumb.

Observation 1.2

The way of measuring is fixed first. Condition numbers in the 2-norm, arithmetic in extended precision (160 digits)——because measured in double precision, the quantity being measured falls outside double precision.

Question: How much worse does the condition number get when A is replaced by $A^\top A$? Is it “usually”, or is it “always”?

2. The Ratio Was 1.0000000000 in Every Case Tried

Calculation 2.1: For the Hilbert matrix H_n with $H_{ij} = 1/(i + j - 1)$, compute $\kappa_2(A)$ and $\kappa_2(A^\top A)$ in extended precision (computed in this paper).

n	$\kappa_2(A)$	$\kappa_2(A^\top A)$	Ratio $\kappa_2(A^\top A)/\kappa_2(A)^2$
4	1.551374×10^4	2.406761×10^8	1.0000000000
6	1.495106×10^7	2.235342×10^{14}	1.0000000000
8	1.525758×10^{10}	2.327936×10^{20}	1.0000000000
12	1.713229×10^{16}	2.935153×10^{32}	1.0000000000

Observation 2.2

A check on the tool—— $\kappa_2(H_5) = 4.766073 \times 10^5$, agreeing with the textbook value of about 4.77×10^5 . The instrument is tested against an independently known case before the table is read.

Observation 2.3

The Hilbert matrix is symmetric. For a symmetric matrix $A^\top A = A^2$, so a ratio of 1 looks like nothing at all——from this table alone one cannot tell whether the squaring comes from symmetry or from something more general.

Calculation 2.4: So check a matrix that is neither symmetric nor square. Take the Vandermonde matrix built from 30 equispaced points of $[0, 1]$, of size 30 by degree plus one——this is precisely the design matrix that appears in least-squares polynomial fitting (computed in this paper).

Degree	$\kappa_2(A)$	$\kappa_2(A^\top A)$	Ratio
4	6.163814×10^2	3.799260×10^5	1.0000000000
6	1.887237×10^4	3.561662×10^8	1.0000000000
8	6.080104×10^5	3.696767×10^{11}	1.0000000000
10	2.067173×10^7	4.273202×10^{14}	1.0000000000

Observation 2.5

Remove symmetry and the ratio stays at 1. On a tall 30×11 matrix that is neither symmetric nor square——the explanation “it squares because it is symmetric” is gone.

3. The Core: This Is an Identity, Not a Rule of Thumb

Setting: Why it is always 1 takes one line of the singular value decomposition to see.

Observation 3.1

Write $A = U\Sigma V^\top$; then $A^\top A = V\Sigma^2 V^\top$. That is, the singular values of $A^\top A$ are the singular values of A each squared——and therefore

$$\kappa_2(A^\top A) = \frac{\sigma_{\max}^2}{\sigma_{\min}^2} = \left(\frac{\sigma_{\max}}{\sigma_{\min}}\right)^2 = \kappa_2(A)^2.$$

Observation 3.2

This is the core of the paper. The two sides are one thing written two ways——not a property of A , not a property of the problem, not a property of how the matrix is filled. It comes from the definition of κ_2 and the definition of the singular values of $A^\top A$.

Observation 3.3

Here is where it matters in practice.If it were a rule of thumb, hunting for a matrix that does not square would be worth doing—if it is an identity, there is none to find. The only way out is not to form $A^T A$.

Observation 3.4

The same move as Paper 326.Paper 326 wrote that the square-cube “law” is a geometric identity and not a law—there too, hunting for counterexamples was beside the point. When what is called a “law” follows from a definition, there is nothing there to break.

Observation 3.5

To be stated honestly—it is an identity in the 2-norm.Define the condition number in the 1- or ∞ -norm and $\kappa(A^T A)$ does not coincide exactly with $\kappa(A)^2$ —the claim of this paper is exact only inside the 2-norm.

4. How Many Digits Are Lost

Setting: The mantissa of a double is 53 bits.In decimal that is $53 \log_{10} 2 = 15.9546$ digits, and that is all the digits there are.

Calculation 4.1: Subtract the decade count of $\kappa_2(A^T A)$ to get what remains (computed in this paper).

n (Hilbert)	$\log_{10} \kappa_2(A^T A)$	Digits remaining
4	8.3814	7.5732
6	14.3493	1.6052
8	20.3670	−4.4124
12	32.4676	−16.5130

Observation 4.2

At $n = 8$ what remains goes negative. A negative count of digits means nothing in itself—what means something is “not one digit is left”: the normal-equations solution at $n = 8$ carries no significant figure at all in double precision.

Observation 4.3

Kept as A , the same $n = 8$ has 10.1835 decades and 5.7711 digits remain. Same problem, same precision, same machine—a change of writing decides whether the answer has significant figures or none.

Observation 4.4

The boundary lies between $n = 6$ and $n = 8$. At $n = 6$, 1.6052 digits remain and a barely meaningful answer comes out—two more in n and 6.0176 digits are gone. The condition number of the Hilbert matrix grows exponentially.

Observation 4.5

There are matrices that do not degrade. For an orthogonal matrix every σ is 1, so $\kappa_2 = 1$, and $A^\top A = I$ has condition number 1 as well—squaring 1 leaves 1. “Forming $A^\top A$ always makes things worse” is, said exactly, too strong.

Observation 4.6

Observation 4.5 says where the way out is. Nothing degrades only when κ is near 1, and QR and the SVD split A into orthogonal factors before solving precisely in order to work in that regime—this paper does not write that one should. It only notes that the numbers show the loss to be avoidable.

5. The Separator: Was $A^\top A$ Formed, or Not?

Observation 5.1

The separator——whether at some point in the solution $A^\top A$ was explicitly formed, or whether A was carried through as it stood.

	Matrix handled	Condition number	Digits left at $n = 8$
Normal equations	$A^\top A$	$\kappa_2(A)^2$	−4.4124
Solve with A itself	A	$\kappa_2(A)$	5.7711

Observation 5.2

The separator is not on the side of the matrix. The same A , the same b , the same answer sought——what makes the difference is what was assembled along the way. The sensitivity is set by the procedure, not by the problem.

Observation 5.3

The loss does not happen because numerical error “enters”. Form $A^\top A$ exactly and the condition number is squared exactly——with not one rounding anywhere, that matrix already carries κ^2 . What is lost is lost not during the arithmetic but at the moment the problem was rewritten.

Observation 5.4

To be stated honestly——if κ is small, nothing whatever goes wrong with the normal equations. At $\kappa_2(A) = 10^3$ the square is 10^6 and 9.9546 of the 15.9546 digits remain——this paper does not say “do not use the normal equations”. It says the range in which they may be used is set by the square root of κ .

Observation 5.5

Where this stands relative to Paper 221. Paper 221 wrote that changing the last bit of the input moves an eigenvalue by 0.16—there the cause is a perturbed input. Here the input never moves and only the way the equation is written squares the sensitivity—under the criterion of Paper 300 the two have distinct roots.

Observation 5.6

One last thing—the normal equations were invented to make the formula shorter. A rectangle becomes a square, the square is symmetric positive definite, and Cholesky applies—the price of that neatness is exactly half the significant digits, and the price is not written anywhere in the formula.

6. The Census and the Close

1. The ratio is **1.0000000000** (Calculation 2.1): Hilbert $n = 4, 6, 8, 12$.
2. $\kappa_2(H_5) = 4.766073 \times 10^5$ (Observation 2.2): agrees with the textbook.
3. Still 1 for a 30×11 rectangle (Calculation 2.4): symmetry plays no part.
4. The singular values of $A^\top A$ are σ^2 (Observation 3.1): the ratio follows from definitions.
5. Being an identity, it has no counterexample (Observation 3.3): none to find.
6. Exact only in the 2-norm (Observation 3.5): other norms do not coincide.
7. Double precision has **15.9546** digits (Section 4): a **53-bit** mantissa.
8. **−4.4124** digits remain at $n = 8$ (Calculation 4.1): not one is left.
9. **5.7711** remain if A is kept (Observation 4.3): same problem, same machine.
10. An orthogonal matrix does not degrade (Observation 4.5): 1 squared is 1.

Closing: Rewrite a least-squares problem as $A^\top A x = A^\top b$ and the condition number is squared exactly—for the Hilbert matrix, and equally for a 30×11 Vandermonde that is neither symmetric nor square, $\kappa_2(A^\top A)/\kappa_2(A)^2$ is **1.0000000000** in extended precision. And this is not a rule of thumb—if $A = U\Sigma V^\top$ then $A^\top A = V\Sigma^2 V^\top$, so the singular values of $A^\top A$ are those of A squared. It is not a property of A nor of the problem, but a consequence of how κ_2 is defined. So hunting for “a matrix that does not square” is pointless—as with the square-cube law in Paper 326, what follows from a definition has nothing there to break. The identity holds inside the 2-norm. The price can be counted in digits—double precision carries a 53-bit mantissa, that is **15.9546** decimal digits. For the 8×8

Hilbert matrix $\kappa_2(A^\top A) = 10^{20.3670}$ and the digits remaining are **—4.4124**: not one is left. Handle the same problem as A and 5.7711 digits survive. Yet it does not always degrade—an orthogonal matrix has $\kappa_2 = 1$ and squaring leaves 1. While κ is small, nothing whatever goes wrong. One thing separates them—whether at some point $A^\top A$ was explicitly formed, or whether A was carried through as it stood. The same A , the same b , the same answer sought; what differs is what was assembled along the way. With not one rounding anywhere, that matrix already carries κ^2 —what is lost is lost not during the arithmetic but at the moment the problem was rewritten. One last thing—the normal equations were invented to make the formula shorter. A rectangle becomes a square, the square is symmetric positive definite, and the handling gets easier. The price of that neatness is exactly half the significant digits, and the price is not written anywhere in the formula.

References

1. Gauss, C. F. (1809). *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium*. Hamburg: Perthes und Besser.
 2. Hilbert, D. (1894). Ein Beitrag zur Theorie des Legendre'schen Polynoms. *Acta Mathematica*, 18.
 3. Turing, A. M. (1948). Rounding-Off Errors in Matrix Processes. *Quarterly Journal of Mechanics and Applied Mathematics*, 1(1).
 4. Golub, G. H., Van Loan, C. F. (2013). *Matrix Computations*, 4th edition. Baltimore: Johns Hopkins University Press.
 5. Higham, N. J. (2002). *Accuracy and Stability of Numerical Algorithms*, 2nd edition. Philadelphia: SIAM.
 6. Yamagishi, Y. (2026). *Convention and Fact* [Paper 95]. Zenodo. 10.5281/zenodo.21758621
 7. Yamagishi, Y. (2026). *Change the Last Bit and an Eigenvalue Moves by 0.16* [Paper 221]. Zenodo. 10.5281/zenodo.22120012
 8. Yamagishi, Y. (2026). *The Three Hundredth Census: Whether Two Things Share a Root Is Decidable* [Paper 300]. Zenodo. 10.5281/zenodo.22171914
 9. Yamagishi, Y. (2026). *The Square-Cube “Law” Is Not a Law but an Identity* [Paper 326]. Zenodo. 10.5281/zenodo.22209992
 10. Yamagishi, Y. (2026). *From the Same Points, One Method Diverges by a Thousand and the Other Converges to a Thousandth* [Paper 329]. Zenodo. 10.5281/zenodo.22681343
-

— *From thinking to syncing.* —

Yuuki YAMAGISHI (山岸 勇輝)

© 2026 Yamagishi Complexity Torus Lab, Japan · Licensed under Creative Commons Attribution 4.0 International (CC BY 4.0)