11. Matrix Inverse

- column by column

Ex.
$$A = \begin{bmatrix} 3 & -0.1 & -0.2 \\ 0.1 & 7 & -0.3 \\ 0.3 & -0.2 & 10 \end{bmatrix}$$

Step 1: LU Factorization $(A = L \cdot U)$

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 0.0333 & 1 & 0 \\ 0.1 & -0.0271 & 1 \end{bmatrix} \qquad U = \begin{bmatrix} 3 & -0.1 & -0.2 \\ 0 & 7.0033 & -0.2933 \\ 0 & 0 & 10.012 \end{bmatrix}$$

Step 2:

1st column

$$\begin{bmatrix} 1 & 0 & 0 \\ 0.0333 & 1 & 0 \\ 0.1 & -0.0271 & 1 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \Rightarrow d = \begin{bmatrix} 1 \\ -0.0333 \\ -0.1009 \end{bmatrix}$$

$$\begin{bmatrix} 3 & -0.1 & -0.2 \\ 0 & 7.0033 & -0.2933 \\ 0 & 0 & 10.012 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ -0.0333 \\ -0.1009 \end{bmatrix} \Rightarrow X = \begin{bmatrix} 0.3325 \\ -0.0052 \\ -0.0101 \end{bmatrix}$$ ✓

2nd column

$$\begin{bmatrix} 1 & 0 & 0 \\ 0.0333 & 1 & 0 \\ 0.1 & -0.0271 & 1 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \Rightarrow d = \begin{bmatrix} 0 \\ 1 \\ 0.0271 \end{bmatrix}$$

$$\begin{bmatrix} 3 & -0.1 & -0.2 \\ 0 & 7.0033 & -0.2933 \\ 0 & 0 & 10.012 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0.271 \end{bmatrix} \Rightarrow X = \begin{bmatrix} 0.0049 \\ 0.1429 \\ 0.0027 \end{bmatrix}$$ ✓

3rd Column

$$\begin{bmatrix} 1 & 0 & 0 \\ 0.0333 & 1 & 0 \\ 0.1 & -0.0271 & 1 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \Rightarrow d = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 3 & -0.1 & -0.2 \\ 0 & 7.0033 & -0.2933 \\ 0 & 0 & 10.012 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \Rightarrow X = \begin{bmatrix} 0.0068 \\ 0.0042 \\ 0.0999 \end{bmatrix}$$ ✓

Check $A^{-1}A = I$

$$\underbrace{\begin{bmatrix} 0.3325 & 0.0049 & 0.0068 \\ -0.0052 & 0.1429 & 0.0042 \\ -0.0101 & 0.0027 & 0.0999 \end{bmatrix}}_{A^{-1}} \cdot \underbrace{\begin{bmatrix} 3 & -0.1 & -0.2 \\ 0.1 & 7 & -0.3 \\ 0.3 & -0.2 & 10 \end{bmatrix}}_{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

A^{-1}:

Solve $Ax = I$ or

$$\underbrace{A}_{LU} x_j = e_j \quad , \; j = 1, 2, \cdots n \qquad \text{- } n \text{ linear systems of equations}$$

No. of operations: $\underbrace{\frac{n^3}{3}}_{\text{Gaussian Elimination}} + \underbrace{n^3}_{\begin{cases} Ux = y : \frac{1}{2}n^2 \\ Ly = b : \frac{1}{2}n^2 \end{cases} * n \text{ equations}} = \frac{4}{3}n^3 \longrightarrow n^3$ ($\because$ zeros in rhs)

L^{-1}: $\quad Y = L^{-1}$

then $Ly_j = e_j \; , \; j = 1, 2, \cdots n$ $\qquad e_j = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \leftarrow j\text{th}$

forward substitution. (Note first $(j-1)$ elements in y_j are zero)
(i.e. L^{-1} is also lower triangular)

$$\therefore y_{ij} = \frac{\delta_{ij} - \sum_{k=j}^{i-1} l_{ik} y_{kj}}{l_{ii}} \; , \quad i = j, j+1, \cdots n$$

No of operations: $\frac{n^3}{6}$

If L and U are known, $(A = L \cdot U)$

$$A^{-1} = (L \cdot U)^{-1} = U^{-1} \cdot L^{-1}$$

No of operations: $\frac{n^3}{6} + \frac{n^3}{6} + \frac{n^3}{3} = \frac{2}{3}n^3$

Vector and Matrix Norms — necessary for error analysis

- **Vector norm** — scala measures its magnitude (concept of Length)

  Lp-norms (or Lp-norms): $\|X\|_p = (|x_1|^p + |x_2|^p + \cdots + |x_n|^p)^{1/p}, \quad 1 \le p < \infty$

  Special case

  $p=2$: Euclidean Norm: $\|X\|_2 = (|x_1|^2 + |x_2|^2 + \cdots + |x_n|^2)^{\frac{1}{2}}$

  $p=1$: Manhattan Distance $\|X\|_1 = (|x_1| + |x_2| + \cdots + |x_n|)$

  $p \to \infty$: Maximum Norm: $\|X\|_\infty = \max_{1 \le i \le n} |x_i|$

Properties:

1(a): $\|x\| > 0$ if $x \neq 0$, $\|x\| = 0$ if $x = 0$

2(a): $\|\alpha x\| = |\alpha| \cdot \|x\|$, α a scalar

3(a): $\|x+y\| \le \|x\| + \|y\|$

- **Matrix Norm's property**

1:(b): $\|A\| > 0$ if $A \neq 0$, $\|A\| = 0$ if $A = 0$

2(b): $\|\alpha A\| = |\alpha| \cdot \|A\|$, α a scalar

3(b) $\|A+B\| \le \|A\| + \|B\|$.

4(b) $\|A \cdot B\| \le \|A\| \cdot \|B\|$

If a matrix norm and a vector norm are related in such a way that

5(b) $\|Ax\| \le \|A\| \cdot \|x\|$,

is satisfied for any A and x, then two norms are said to be consistent

For any vector norm, there exists a consistent matrix norm (matrix-bound norm) subordinate to the vector norm .

$$\|A\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|}$$

$\|A\|_f = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^2}$ // Frobenius norm //

$\|A\|_1 = \max_{1 \le j \le n} \sum_{i=1}^{n} |a_{ij}|$ // column-sum norm //

$\|A\|_\infty = \max_{1 \le i \le n} \sum_{j=1}^{n} |a_{ij}|$ // row-sum norm //

Example.

$$A = \begin{pmatrix} 1.2969 & 0.8648 \\ 0.2161 & 0.1441 \end{pmatrix} \qquad b = \begin{pmatrix} 0.8642 \\ 0.1440 \end{pmatrix}$$

$\|A\|_\infty = \max (1.2969 + 0.8648, 0.2161 + 0.1441) = 2.1617$

$\|b\|_\infty = 0.8642$

Perturbation Analysis

$$Ax = b$$

Let's find the effect of perturbations in b and A. If we let

$$A(x + \delta x) = b + \delta b$$

then $\delta x = A^{-1}\,\delta b$

$$\therefore \|\delta x\| \le \|A\| \cdot \|\delta b\| \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

If we let $(A + \delta A)(x + \delta x) = b$

then $A\,\delta x + \delta A\,(x + \delta x) = 0$

$$\therefore \delta x = -A^{-1}\,\delta A\,(x + \delta x)$$

$$\|\delta x\| \le \|A^{-1}\| \cdot \|\delta A\| \cdot \|x + \delta x\|$$

$$\frac{\|\delta x\|}{\|x + \delta x\|} \le k(A)\,\frac{\|\delta A\|}{\|A\|}$$

where $k(A) = \|A\| \cdot \|A^{-1}\|$:condition number

From (1) and inequality $\|b\| = \|Ax\| \le \|A\|\,\|x\|$, it follows that

$$\frac{\|\delta x\|}{\|x\|} \le K(A)\,\frac{\|\delta b\|}{\|b\|}$$

If k(A) is large, then small perturbation in A and b will produce large relative perturbations in x - ill-conditioned problem.

Example. $A = \begin{bmatrix} 1.2969 & 0.8648 \\ 0.2161 & 0.1441 \end{bmatrix}$ $b = \begin{bmatrix} 0.8642 \\ 0.1440 \end{bmatrix}$ $\|A\|_\infty = 2.1617$

$$A^{-1} = 10^8 \begin{bmatrix} 0.1441 & -0.8648 \\ -0.2161 & 1.2969 \end{bmatrix} \qquad \|A^{-1}\|_\infty = 1.5130 \times 10^8$$

$$k(A) = \|A\| \cdot \|A^{-1}\| = 2.1617 \times 1.5130 \times 10^8 \approx \underline{\underline{3.3 \times 10^8}}$$ ill-conditioned

Ill-conditioned Problems

(1) Hilbert Matrix

$$a_{ij} = \frac{1}{i+j-1} \qquad 1 \le i, j \le n$$

Ex. $n=4$, $x_i = (1, -1, 1, -1)^t$, 5-decimals

$$\left[\begin{array}{cccc|c} 1.00000 & 0.50000 & 0.33333 & 0.25000 & 0.58333 \\ 0.50000 & 0.33333 & 0.25000 & 0.20000 & 0.21667 \\ 0.33333 & 0.25000 & 0.20000 & 0.16667 & 0.11666 \\ 0.25000 & 0.20000 & 0.16667 & 0.14286 & 0.07381 \end{array}\right]$$

$$\left[\begin{array}{cccc|c} 1.00000 & 0.50000 & 0.33333 & 0.25000 & 0.58333 \\ 0 & 0.08333 & 0.08333 & 0.07500 & -0.07500 \\ 0 & 0.08333 & 0.08889 & 0.08333 & -0.07778 \\ 0 & 0.07500 & 0.08333 & 0.08036 & -0.07202 \end{array}\right]$$

$$\left[\begin{array}{cccc|c} 1.00000 & 0.50000 & 0.33333 & 0.25000 & 0.58333 \\ 0 & 0.08333 & 0.08333 & 0.07500 & -0.07500 \\ 0 & 0 & 0.00556 & 0.00833 & -0.00278 \\ 0 & 0 & 0.00833 & 0.01286 & -0.00452 \end{array}\right]$$

$$\left[\begin{array}{cccc|c} 1.00000 & 0.50000 & 0.33333 & 0.25000 & 0.58333 \\ 0 & 0.08333 & 0.08333 & 0.07500 & -0.07500 \\ 0 & 0 & 0.00556 & 0.00833 & -0.00278 \\ 0 & 0 & 0 & -0.00038 & 0.00036 \end{array}\right]$$

$x_1 = 0.99424$

$x_2 = -0.96101$

$x_3 = 0.91933$

$x_4 = -0.94737$

	Gaussian	True	Terror
x_1	0.99424	1	0.00576
x_2	-0.96101	-1	0.03899
x_3	0.91933	1	0.08067
x_4	-0.94737	-1	0.05263

[2] Vandermonde Matrix

$$a_{ij} = (1+i)^{j-1} \qquad b_i = \frac{(1+i)^n - 1}{i}$$

Ex. $n = 4$

$$\begin{bmatrix} 1 & 2 & 4 & 8 \\ 1 & 3 & 9 & 27 \\ 1 & 4 & 16 & 64 \\ 1 & 5 & 25 & 125 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 15 \\ 40 \\ 85 \\ 156 \end{bmatrix}$$

* For large n, round-off error propagates and magnifies throughout back substitution.

$n = 9$

1	2	4	8	16	32	64	128	256	511
1	3	9	27	81	243	729	2187	6561	9841
1	4	16	64	256	1024	4096	16384	65536	87381
1	5	25	125	625	3125	15625	78125	390625	488281
1	6	36	216	1296	7776	46656	279936	1679616	2015539
1	7	49	343	2401	16807	117649	823543	5746801	6725601
1	8	64	512	4096	32768	262144	2097152	16777216	19173960
1	9	81	729	6561	59049	531441	4782969	43046720	48427560
1	10	100	1000	10000	100000	1000000	10000000	100000000	111111106 ?

1	2	4	8	16	32	64	128	256	511
	1	5	19	65	211	665	2059	6305	9330
		2	18	110	570	2702	12138	52670	68210
			6	84	750	5460	35406	213494	255150
				24	480	5880	57120	484344	547848
					120	3240	52080	650160	705600
						720	25200	514080	539998
							5040	221758	226812
								40336	40294

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
-683	1240	-931	387	-95	15.7	-0.37	1.07	0.998

5.5.6. Iterative Improvement of a Solution

We have seen that when A is ill-conditioned, the computed solution $\bar{x}$ may be inaccurate without any indication in the form of a large solution vector. This will happen for particular right-hand sides b, such that $A^{-1}b$ is small even though the norm of A^{-1} is large. If we compute the inverse matrix, we cannot be deceived in this way and ill-conditioning will show up in the form of large elements in the computed inverse $\bar{A}^{-1}$. If the computed condition number $\|A\|_\infty \|\bar{A}^{-1}\|_\infty$ is small, then $\bar{A}^{-1}$ certainly is close to the true inverse.

Unfortunately, the extra work involved in computing the inverse is often prohibitively large. We shall here describe an alternative approach, which requires very little extra work when n is large, and which also gives a correction to $\bar{x}$ and not just an estimate of the error. If $r = b - A\bar{x}$ is the residual vector to a computed solution $\bar{x}$, then

$$A(x - \bar{x}) = r.$$

Now assume that Gaussian elimination has given the approximate triangular factors $\bar{L}$ and $\bar{U}$. From Theorem 5.5.1 we know that $\bar{L}\bar{U} = A + E$, where E is small. We can therefore approximate the correction $x - \bar{x}$ with the solution to

$$\bar{L}(\bar{U}\delta x) = r,$$

which splits into the two triangular systems $\bar{L}y = r$ and $\bar{U}\delta x = y$. The computation of r and δx, therefore, takes only $n^2 + 2 \cdot \frac{1}{2}n^2 = 2n^2$ operations, which is an order of magnitude less than the $n^3/3$ operations required for computing $\bar{x}$.

New rounding errors are introduced in the computation of δx, and $\bar{x} + \delta x$ may not be a more accurate solution than $\bar{x}$. A more detailed analysis shows that, because of the cancellation which will take place in computing $r = b - A\bar{x}$, it is essential that it be computed with sufficient accuracy. It is often advisable to proceed as follows. The components in r are

$$r_i = b_i - \sum_{k=1}^{n} a_{ik}\bar{x}_k, \quad i = 1, 2, \ldots, n.$$

If a_{ik} and $\bar{x}_k$ are given with t digits, then the products $a_{ik}\bar{x}_k$ contain at most $2t$ digits. We compute these products exactly and accumulate the sum using $2t$ digits. Finally, $b_i - (A\bar{x})_i$ is computed and rounded to t digits. This can be done very conveniently on most computers, and will insure that the error from this part of the calculation is small.

The improved solution $\bar{x} + \delta x$ can, of course, be corrected in the same way, etc., and we can carry out the following iterative process:

Iterative improvement. Put $x^{(1)} = \bar{x}$ and compute $x^{(s)}$, $s = 2, 3, \ldots$, from

$$r^{(s)} = b - Ax^{(s)}, \qquad \bar{L}(\bar{U}\delta x^{(s)}) = r^{(s)}, \quad x^{(s+1)} = x^{(s)} + \delta x^{(s)}, \tag{5.5.17}$$

where only the computation of $r^{(s)}$ requires double precision. If A is not too ill-conditioned—say,

$$nu\kappa(A) \leq 0.1,$$

—then $x^{(s)}$ will converge rapidly to the correct solution rounded to single precision. We can also get a good estimate of $\kappa(A)$ from

$$\kappa(A) \leq \frac{1}{nu} \frac{\|\delta x^{(1)}\|_\infty}{\|x^{(2)}\|_\infty}. \tag{5.5.18}$$

If convergence is not obtained in practice, then we must assume that A is so ill-conditioned that higher-precision arithmetic throughout is unavoidable.

Example 5.5.5

We illustrate the method on the equations

$$\begin{pmatrix} 0.20000 & 0.16667 & 0.14286 \\ 0.16667 & 0.14286 & 0.12500 \\ 0.14286 & 0.12500 & 0.11111 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0.50953 \\ 0.43453 \\ 0.37897 \end{pmatrix},$$

which have the exact solution $x_1 = x_2 = x_3 = 1$. If floating-point arithmetic with $t = 5$ digits is used, Gaussian elimination will give the computed triangular factors

$$\bar{L} = \begin{pmatrix} 1 & 0 & 0 \\ 0.83335 & 1 & 0 \\ 0.71430 & 1.49874 & 1 \end{pmatrix}, \qquad \bar{U} = \begin{pmatrix} 0.20000 & 0.16667 & 0.14286 \\ 0 & 0.00397 & 0.00595 \\ 0 & 0 & 0.00015 \end{pmatrix},$$

and the computed solution

$$\bar{x} = (1.03845, 0.89673, 1.06667)^T.$$

We compute first $A\bar{x}$ using $2t$ digits, then $r^{(1)} = b - A\bar{x}$,

$$A\bar{x} = \begin{pmatrix} 0.5095324653 \\ 0.4345190593 \\ 0.3789619207 \end{pmatrix}, \qquad r^{(1)} = 10^{-5} \begin{pmatrix} -0.24653 \\ 1.09407 \\ 0.80793 \end{pmatrix},$$

and then solve for $\delta x^{(1)}$. We get

$$\delta x^{(1)} = \begin{pmatrix} -0.03709 \\ 0.09955 \\ -0.06424 \end{pmatrix}, \qquad x^{(2)} = \bar{x} + \delta x^{(1)} = \begin{pmatrix} 1.00136 \\ 0.99628 \\ 1.00243 \end{pmatrix}.$$

The errors in the corrected solution are about 30 times smaller than those in $\bar{x}$. The rapid convergence obtained if we continue the iterations is clearly illustrated in the tables below.

s	$x^{(s)}$		
1	1.03845	0.89673	1.06667
2	1.00136	0.99628	1.00243
3	1.00005	0.99986	1.00009
4	1.00000	1.00000	1.00000
s	$10^5 \cdot r^{(s)}$		
1	−0.24653	1.09407	0.80793
2	0.08626	0.10180	0.07131
3	0.04764	0.04169	0.03571

It is interesting to note that the residuals do not decrease at the same rate as the errors in successive $x^{(s)}$. Using (5.5.18), we can derive the estimate

$$\kappa(A) \approx \frac{1}{3 \cdot \frac{1}{2} \cdot 10^{-5}} \frac{0.1}{1} = 0.7 \cdot 10^4,$$

which agrees well with the known value.

residual vector $\tilde{r} = A\tilde{x} - b$

error vector $\tilde{e} = \tilde{x} - x$

Question. Does smaller residual vector always means good $\tilde{x}$?

Counterexample. (Kahan)

$$A = \begin{pmatrix} 1.2969 & 0.8648 \\ 0.2161 & 0.1441 \end{pmatrix} \qquad b = \begin{pmatrix} 0.8642 \\ 0.1440 \end{pmatrix}$$

Suppose $\tilde{x} = \begin{pmatrix} 0.9911 \\ -0.4870 \end{pmatrix}$ then $r = \begin{pmatrix} 10^{-8} \\ -10^{-8} \end{pmatrix}$

but $x = \begin{pmatrix} 2 \\ -2 \end{pmatrix}$

Reason.

After elimination,

$$a_{22}^{(2)} = 0.1441 - \frac{0.2161}{1.2969} \cdot 0.8648 = 0.1441 - 0.144099923 \approx 10^{-8}$$

∴ small change in 0.1441 → large change in $a_{22}^{(2)}$ → large change in x_2