Linear Algebra Online and Offline: The 2015 Lin Lectures

by Gilbert Strang*

Lin Chia-Chiao was a remarkable man. He was very kind to this young assistant professor at MIT. All the faculty knew that he had come from China's famous university Tsinghua (and we were told that Tsinghua was the "MIT of China"). But I never expected to have this chance for lectures in his memory right here.

Everyone was permitted to call him C. C. His specialties were fluid mechanics and astrophysics. He was promoted to the rank of Institute Professor, which is an enormous honor at MIT. (The one and only Institute Professor in our department right now is Isadore Singer, winner of the Abel Prize jointly with Michael Atiyah.) After retirement C. C. and his wife eventually moved to a home provided for them here on the Tsinghua campus. On my previous visit I had the good fortune to talk to him—we enjoyed remembering the MIT Math Department of Wiener and Levinson and Lin. He passed away at the age of 96.

I was always grateful to him, and now it is a special honor to visit Tsinghua. I chose to speak partly about my experience of preparing video lectures on linear algebra. I want to encourage others to think (and to act!) in this direction too—it can change your life. I will also describe a sequence of ideas that could help to complete large matrices when data is missing, and to compute with those matrices when the size of the data is overwhelming.

OpenCourseWare and Online Teaching

My principal courses at MIT are 18.06 Linear Algebra and 18.085 Computational Science and Engineering. One is for undergraduates, with 250 students in Spring 2015. The other is for graduate students, with 125 students in Fall 2014. At some time in the 20th century I was asked to reorganize these subjects. At that time linear algebra was studied by a small group of math majors (which once included me—I liked it but we almost never saw a matrix). And the equivalent to 18.085 never saw a computer. This had to change—without losing the mathematical content.

The new courses needed new textbooks: *Linear Algebra and Its Applications* (1976 … 2006) was followed by *Introduction to Linear Algebra* (Wellesley-Cambridge Press, 1993 … 2009). A Chinese translation of that second book is just completed. For the graduate course, *Introduction to Applied Mathematics* (1986) was followed by *Computational Science and Engineering* (2007). And now there are video lectures; those are my subject today.

MIT's OpenCourseWare was a new answer to an old question: How to help students outside the classroom? It was the right answer—to make the lectures free for everyone. There are 2000 MIT courses on `ocw.mit.edu`, many with videos. They show MIT as it is.

The new problem in 2015 is to go further. OCW led to MITx (on campus) and edX (worldwide) with homeworks and exams and grades and certificates. The 18.06 homeworks are now graded by computer, with instant response that students like. We have not yet offered an edX course on linear algebra (a MOOC = Massive Open Online Course).

The linear algebra course was filmed in my normal class with normal mistakes; you could look at Lecture 10. Probably 4,000,000 viewers have watched, and quite a large subset has sent email! I especially wanted to provide a version with Chinese subtitles.

* Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Mass., U.S.A.
E-mail: gs@math.mit.edu

I searched for the right person to help, then I was happy to learn that the subtitles already existed (possibly now on `163.com`).

Perhaps my main message here at Tsinghua is to say that the technical part, the recording itself, is not difficult. I think we will soon have full-scale successful efforts starting with high school algebra—and why not all the way back to ordinary multiplication? I just completed a series on Differential Equations using my own camera (planned for my YouTube channel as well as `ocw.mit.edu` and `mathworks.com`). The Khan Academy shows that one person can make *thousands* of videos. Those are not highly developed but Khan has a relaxed style that students appreciate.

Matrix Completion with Maximum Determinant and Banded Inverse

These pages will outline a series of steps in the "linear algebra of banded matrices". These matrices have zero entries outside a band of width w on each side of the main diagonal: $a_{ij} = 0$ for $|i-j| > w$. A tridiagonal matrix (bandwidth $w = 1$) is the simplest and most important example.

The nonzero entries in A might correspond to edges in a graph or a network, when the nodes lie along a line—and connections can only reach w nodes to the left and right.

We will outline without proofs a series of ideas in the literature that potentially lead to fast completions and computations with large banded matrices. Ordinary elimination solves a linear system $Ax = b$ in $O(Nw^2)$ steps, but that is not our problem. Here are the main ideas:

1. The Nullity Theorem
2. Inverses of banded matrices
3. Positive definite completion with maximum determinant
4. Fast inverse of the completed (full) matrix

We see possible future applications for genetic data.

1. The Nullity Theorem applies to any invertible matrix A and its inverse, partitioned into four (possibly rectangular) blocks:

$$A^{-1} = \left[\begin{array}{c|c} P & Q \\ \hline R & S \end{array}\right]^{-1} = \left[\begin{array}{c|c} T & U \\ \hline V & W \end{array}\right].$$

The theorem says that each block T, U, V, W in A^{-1} has the *same nullity* (dimension of nullspace) as the complementary block S, Q, R, P in A. For a 2 by 2 matrix with 1 by 1 blocks, the entries t, u, v, w are exactly $s, -q, -r, p$ divided by $\det A \neq 0$.

The proof reproduced in [10] starts from the block equations that are given by $AA^{-1} = I$ and $A^{-1}A = I$.

2. The inverses of banded matrices have a special structure: *Low rank above and below the diagonal.* If A is a tridiagonal matrix, all submatrices of A^{-1} that do not cross the main diagonal have rank ≤ 1. If A has bandwidth w, then all submatrices above the w^{th} subdiagonal of A^{-1} and all submatrices below the w^{th} superdiagonal of A^{-1} have rank $\leq w$ [2, 9] Then A^{-1} is "semiseparable of rank w."

Here is a typical step in the proof. When A is tridiagonal, the submatrix Q in rows 1, 2 and columns 4 to n is zero. Its nullity is $n-3$ (the number of columns). The complementary submatrix U includes rows $1, 2, 3$ and columns 3 to n of A^{-1}. Its nullity is $n-3$ by the Nullity Theorem. Since it has $n-2$ columns, its rank is 1.

The rank of any matrix equals the size of its largest invertible square submatrix. Thus if A is tridiagonal, every 2 by 2 submatrix of A^{-1} (rows $i, i+1$ and columns $j, j+1$) is singular—zero determinant—except if $i = j$.

$$A^{-1} = \begin{bmatrix} a & b & 0 \\ c & d & e \\ 0 & f & g \end{bmatrix}^{-1} = \begin{bmatrix} . & x & \mathbf{xy^{-1}z} \\ . & y & z \\ . & . & . \end{bmatrix}$$

In this case the zero entries in two corners of A match the zero cofactors in two corners of A^{-1}.

Important. That example shows how the entries of A^{-1} can be systematically constructed—building outward from its tridiagonal part B that contains x, y, z. The construction multiplies the column $[x\ y]^T$ times the row $[1\ y^{-1}z]$ to produce that 2 by 2 matrix of rank 1.

This step also applies when B is *block* tridiagonal, including blocks x, y, z. The completed block in A^{-1} is still $[x\ y]^T[I\ y^{-1}x]$. Our whole analysis extends to "block-banded" matrices A and to low rank off-diagonal blocks in A^{-1}. These are dual properties.

3. To explain matrix completion, we reverse direction. Begin with the tridiagonal part B of an incomplete matrix C. All other entries of C are to be determined. In applications C might represent a covariance matrix in which we only know the variances $b_{ii} = c_{ii} = \sigma_i^2$ and the covariances $\sigma_{i\ i+1}$ between neighboring pairs i and $i+1$. We do not know the other covariances—only that C should be a positive definite matrix. Starting with that tridiagonal part B, is there a better way to complete the matrix C than simply padding it with zeros?

Yes, there is. This is a widely discussed problem, to fill in missing data in an optimal way. One popular principle is "maximum entropy". The probability distribution $p(x) = e^{-x^T C^{-1} x/2}/(2\pi)^{n/2}\sqrt{\det C}$ is multivariate Gaussian. Its entropy $\int p(x) \log p(x)\ dx$ turns out to depend directly on the determinant of C.

Maximizing the entropy reduces to *maximizing the determinant of* C, when the tridiagonal part is known. See Dempster [3] for an early and convincing analysis of the statistical reasoning.

Now comes the happy fact about maximum determinant. It is achiev-ed by that same step by step construction from B to C, maintaining low rank off the diagonal. In other words, *the inverse of the optimally completed matrix* C *is a banded matrix* A [5]:

> Every entry that is completed in C corresponds to a zero entry in the inverse matrix A.

We can connect this fact to the cofactor formula for the determinant of C. Each new entry is multiplied by its cofactor, so calculus (zero derivative at the maximum point) asks for cofactor = zero. This result extends to every incomplete matrix B (including all block-banded matrices) provided its entries produce a *chordal graph*: Every cycle of length ≥ 4 has a chord that cuts across to produce a shorter cycle.

B will be outside this completion theory if its tridiagonal part *and also* b_{1n} *and* b_{n1} are specified. And a square grid in the plane is not chordal. Its minimal cycles have length 4.

Another way to recognize the optimal completion $c_{13} = c_{31}$ of a 3 by 3 symmetric tridiagonal matrix B is the determinant formula

$$b_{22}\ \det C = \begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix} \begin{vmatrix} b_{22} & b_{23} \\ b_{32} & b_{33} \end{vmatrix} - \begin{vmatrix} b_{12} & c_{13} \\ b_{22} & b_{23} \end{vmatrix}^2$$

$\det C$ is a maximum when that last determinant is zero.

4. Here is a remarkable **local inverse formula** for the optimal $C^{-1} = A$. The formula uses only the specified part B of C (this is possible because C is constructed from B). We know that the unspecifed part of C (the completed part) corresponds to zeros in $C^{-1} = A$. In our example, the local inverse formula gives the *tridiagonal part of* C^{-1} *directly from the tridiagonal part* B *of* C.

Statistical computations usually involve C^{-1} more than the actual covariance matrix C. So this formula is potentially useful. It computes C^{-1} by adding the inverses of the 2 by 2 submatrices of B and subtracting the inverses of their 1 by 1 overlaps:

$$C^{-1} = \begin{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}^{-1} & & \\ & + \begin{bmatrix} b_{22} & b_{23} \\ b_{32} & b_{33} \end{bmatrix}^{-1} & - [b_{22}]^{-1} \end{bmatrix}$$

The extension to wider bands remains correct: Add the inverses of principal submatrices of size $w+1$ and subtract the inverses of size w.

I believe that this remarkable formula was first found by statisticians; it is implicit in [8, page 145]. Johnson and Lundquist [6] establish the formula for block-banded matrices and all chordal matrices. And the signal processing literature [1, 7] offers a fast inverse that begins with the Cholesky factorization $B = LL^T$. This comes from ordinary Gaussian elimination.

Altogether a fascinating series of ideas about banded matrices. I want to thank many friends (above all Shev MacNamara) for joining me in this continuing adventure.

References

[1] Amir Asif and José Moura, *Block matrices with L-block-banded inverse: Inversion algorithms*, IEEE Trans. Signal Proc. **53** (2005) 630–642.

[2] Edgar Asplund, *Inverses of matrices which satisfy* $a_{ij} = 0$ *for* $j > i + p$, Mathematica Scandinavica **7** (1959) 57–60.

[3] Arthur Dempster, *Covariance selection*, Biometrica **28** (1972) 157–175.

[4] Harry Dym and Israel Gohberg, *Extensions of band matrices with band inverses*, Linear Algebra Appl. **36** (1981) 1–24.

[5] Charles Johnson, *Matrix completion problems: A survey* in *Matrix Theory and Applications*, Proc. Symp. Appl. Math., American Math. Society **40** (1990) 171–198.

[6] Charles Johnson and Michael Lundquist, *Local inverse of matrices with sparse inverses*, Linear Algebra Appl. **277** (1998) 33–39.

[7] Aleksandar Kavcic and José Moura, *Matrices with banded inverses: Inversion algorithms*, IEEE Trans. Inf. Th. **46** (2000) 1495–1509.

[8] Steffen Lauritzen, *Graphical Models*, Clarendon Press, 2002.

[9] Gilbert Strang and Tri-Dong Nguyen, *The interplay of ranks of submatrices*, SIAM Review **46** (2004) 637–646.

[10] Raf Vandebril, Marc van Borel and Nicola Mastronardi, *Matrix Computations and Semiseparable Matrices*, Johns Hopkins, 2008.