STRUCTURE IN MATRICES

David F. Gleich

August 27, 2019

Lecture 3 – Learning objectives.
How to understand *structure* in matrices and where it comes from.
What a Hankel matrix is.
What a sparse matrix is.

One of the goals of matrix computations is to exploit structure inside the problem. That is, there are algorithms that will work for essentially any matrix that is a valid input for the problem. But you are unlikely to have any matrix! You have a matrix that arises in your application. The idea is that we should be able to take advantage of that structure in order to write better algorithms!

1 A FIRST EXAMPLE OF STRUCTURE

Consider those least-squares problems we had from the last lecture. The matrix

$$\boldsymbol{A} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & & \\ 1 & x_N & x_N^2 \end{bmatrix}.$$

One way to solve a least-squares problem (which we will derive eventually, see a future lecture!) is to convert it into a set of linear equations called the normal equations.

Recap of the normal equations approach for least-squares [1] To find $\mathbf{x}$ such that $\|\mathbf{b} - \boldsymbol{A}\mathbf{x}\|_2^2$ is minimized, we can solve the normal equations

$$\boldsymbol{A}^T\boldsymbol{A}\mathbf{x} = \boldsymbol{A}^T\mathbf{b}.$$

[1] See the future lectures on why there are usually better ways to solve these problems.

Let's look at the matrix $\boldsymbol{A}^T\boldsymbol{A}$ defined in the least squares problem with above:

$$\boldsymbol{A}^T\boldsymbol{A} = \begin{bmatrix} N & \sum_i x_i & \sum_i x_i^2 \\ \sum_i x_i & \sum_i x_i^2 & \sum_i x_i^3 \\ \sum_i x_i^2 & \sum_i x_i^3 & \sum_i x_i^4. \end{bmatrix}.$$

Again, this is a very specific type of matrix. Notice that the values are constant in certain regions:

$$\boldsymbol{A}^T\boldsymbol{A} = \begin{bmatrix} y_1 & y_2 & y_3 \\ y_2 & y_3 & y_4 \\ y_3 & y_4 & y_5 \end{bmatrix}.$$

So if you are interested in solving these types of equations, you do *not* need a general purpose means of solving a linear system of equations. What you need is just an algorithm to solve these specific types of inputs.

This type of input arises frequently, and so it has acquired a name. The general form is called a *Hankel* matrix.[2]

[2] Who was Hankel? Is he really the first to look at these? Short answer: I don't know. Wikipedia has a little bit of information, https://en.wikipedia.org/wiki/Hermann_Hankel, so does the Encyclopedia of Mathematics, https://www.encyclopediaofmath.org/index.php/Hankel_matrix, but neither gets at the question of who was the first to specialize on these types of matrices.

DEFINITION 1 *A matrix is a* Hankel matrix *if* $A_{i,j} = A_{i-1,j+1}$ *whenever* $i-1, j+1$ *is a valid index into the matrix.*

The following are straightforward results of this definition.

THEOREM 2

- *A square Hankel matrix is symmetric.*
- *An* $m \times n$ *Hankel matrix is defined by* $m+n-1$ *numbers.*

Thus, as an example, we could work out how to solve linear equations with Hankel matrices and show how to use that to fit 1 dimensional curves. This would *hopefully* make it better for this single purpose.

2 SPARSE STRUCTURE IN MATRICES

The next type of structure we'll discuss is *sparse structure* or simply put *sparsity*. Simply put, a sparse matrix consists of mostly zeros. Conversely, a dense matrix is a matrix that is not sparse.

Matrices with mostly zeros arise often when we are studying real-world systems. There are a large number of examples of this at the SuiteSparse Matrix Collection (`https://sparse.tamu.edu/`) which has a database of tens of thousands of examples of these real-world system and the matrices that result.

We'll see a few examples as we go along through our lectures. Let's get started with a simple one that comes up for a problem that has all the of the actual interesting detail removed.

EXAMPLE 3 *Consider the following problem. Suppose you are sitting on the number line at 0, and you move left and right with equal probability (1/2) at each step. What is the expected length of time until you first hit the integer* $+6$ *or* -4. [3]

[3] More generally, this is an instance of first-hitting time in Markov chains, but we don't need to get into the formalities of that characterization and generalization of the problem.

We can solve this by letting $x_{\{i\}}$ *be the expected length of time before you first hit either integer given that you start at state* $\{i\}$. *Note that* i *can range from* -4 *to* $+6$. *Also note that* $x_{\{-4\}} = x_{\{+6\}} = 0$. *Let's work out the others. Suppose we were to start at state* $\{-3\}$. *Then we have at least one move that results in two cases: we move to* $\{-4\}$ *with probability 1/2 and stop, or we move to* $\{-2\}$ *with probability 1/2 and continue. Consequently:*

$$x_{\{-3\}} = 1 + \tfrac{1}{2}x_{\{-4\}} + \tfrac{1}{2}x_{\{-2\}}.$$

Likewise, we can apply the same analysis to get:

$$\begin{aligned} x_{\{-2\}} &= 1 + \tfrac{1}{2}x_{\{-3\}} + \tfrac{1}{2}x_{\{-1\}} \\ x_{\{-1\}} &= 1 + \tfrac{1}{2}x_{\{-2\}} + \tfrac{1}{2}x_{\{0\}} \\ &\vdots \\ x_{\{+5\}} &= 1 + \tfrac{1}{2}x_{\{+4\}} + \tfrac{1}{2}x_{\{+6\}}. \end{aligned}$$

This gives us an overall linear system of equations:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{2} & 1 & -\frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{1}{2} & 1 & -\frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{\{-4\}} \\ x_{\{-3\}} \\ x_{\{-2\}} \\ x_{\{-1\}} \\ x_{\{0\}} \\ x_{\{+1\}} \\ x_{\{+2\}} \\ x_{\{+3\}} \\ x_{\{+4\}} \\ x_{\{+5\}} \\ x_{\{+6\}} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}.$$

Note that most of the elements are zero.

EXAMPLE 4 *Consider the following revised setting. Suppose you are sitting on the number line at 0, and you move right with probability* p *and stay put with probability* q. *With probability* r, *you restart at* 0 *at any value between* 0 *and your current position-1, with the restart chosen uniformly at random. What is the expected time until you hit* $+10$? *We also have* $p+q+r = 1$. *If* $p = 1/2, q = 1/4, r = 1/4$, *how long do we expect to take to reach* $+10$?

Again, this can be solved by a similar linear system. First, we have $x_{\{0\}} = 1 + (1 - p)x_{\{0\}} + px_{\{+1\}}$. *The other equations all proceed in a similar fashion.*

$$\begin{bmatrix} p & -p & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -r & (1-q) & -p & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{2}r & -\frac{1}{2}r & (1-q) & -p & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{3}r & -\frac{1}{3}r & -\frac{1}{3}r & (1-q) & -p & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{4}r & -\frac{1}{4}r & -\frac{1}{4}r & -\frac{1}{4}r & (1-q) & -p & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{5}r & -\frac{1}{5}r & -\frac{1}{5}r & -\frac{1}{5}r & -\frac{1}{5}r & (1-q) & -p & 0 & 0 & 0 & 0 \\ -\frac{1}{6}r & -\frac{1}{6}r & -\frac{1}{6}r & -\frac{1}{6}r & -\frac{1}{6}r & -\frac{1}{6}r & (1-q) & -p & 0 & 0 & 0 \\ -\frac{1}{7}r & -\frac{1}{7}r & -\frac{1}{7}r & -\frac{1}{7}r & -\frac{1}{7}r & -\frac{1}{7}r & -\frac{1}{7}r & (1-q) & -p & 0 & 0 \\ -\frac{1}{8}r & -\frac{1}{8}r & -\frac{1}{8}r & -\frac{1}{8}r & -\frac{1}{8}r & -\frac{1}{8}r & -\frac{1}{8}r & -\frac{1}{8}r & (1-q) & -p & 0 \\ -\frac{1}{9}r & -\frac{1}{9}r & -\frac{1}{9}r & -\frac{1}{9}r & -\frac{1}{9}r & -\frac{1}{9}r & -\frac{1}{9}r & -\frac{1}{9}r & -\frac{1}{9}r & (1-q) & -p \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{\{0\}} \\ x_{\{+1\}} \\ x_{\{+2\}} \\ x_{\{+3\}} \\ x_{\{+4\}} \\ x_{\{+5\}} \\ x_{\{+6\}} \\ x_{\{+7\}} \\ x_{\{+8\}} \\ x_{\{+9\}} \\ x_{\{+10\}} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}.$$

A system of equations closely related to this arose when were were studying the ability of algorithms to make progress *when there are failures. In this case,* p *is the probability of progress on a given step and* q *and* r *model two different failure scenarios that take us back to previous levels of progress.*

— TODO – Aside on Candyland Markov chain.
— TODO – Aside on PageRank linear system.
— TODO – Forward reference to creating a problem on a grid.

3 SYMMETRIC POSITIVE DEFINITE MATRICES

When we solved the least squares problem via the normal equations, the matrix $\boldsymbol{A}^T\boldsymbol{A}$ came up. It turns out that matrices with this form are extremely important and so they are called *symmetric positive semi-definite matrices*. More explicitly, a symmetric positive definite matrix is one that can be written:

$$\boldsymbol{B} = \boldsymbol{F}^T\boldsymbol{F}$$

for some matrix $\boldsymbol{F}$. Note that $\boldsymbol{F}$ need not be unique. Another way of characterizing symmeric positive semi-definite matrices is:

$$\mathbf{x}^T\boldsymbol{A}\mathbf{x} \geq 0 \text{ for all } \mathbf{x}.$$

DEFINITION 5 *A matrix* $\boldsymbol{A}$ *is symmetric positive definite if*

$$\mathbf{x}^T\boldsymbol{A}\mathbf{x} > 0 \textit{ for all } \mathbf{x} \neq 0.$$

4 FAST OPERATORS AND DATA-SPARSE MATRICES

— TODO – Fill in this section or move until later.

5 M-MATRICES

An M-matrix is a square matrix where the inverse is non-negative. [4]

[4] There are a few variations on these classes, so this may not be quite right, so I want to double-check.

DEFINITION 6 *Let* $\boldsymbol{A}$ *be a square matrix, then* $\boldsymbol{A}$ *is an* M*-matrix if* $\boldsymbol{A}^{-1}$ *is elementwise non-negative.*

The class of M-matrices arises frequently when dealing with Markov chains (although that is not why it is called an M-matrix).
— TODO – More on where M-matrices come from.

6 OTHER CLASSES OF MATRICES

TOEPLITZ A Toeplitz matrix is one where $A_{i,j} = A_{i+1,j+1}$ whenever the latter index is valid.

DIAGONAL A diagonal matrix is one where only the main diagonal entries are set to non-zero values. An example is the identity matrix, where all the diagonal entries are 1.

TRIDIAGONAL A tridiagonal matrix is one where only three diagonals are set.

The most important classes of matrices:
Symmetric positive definite. M-matrices. Orthogonal matrices. Sparse.