Section 2.4/2.5: Matrix Partitions and Factorizations

February 22, 2008

Abstract

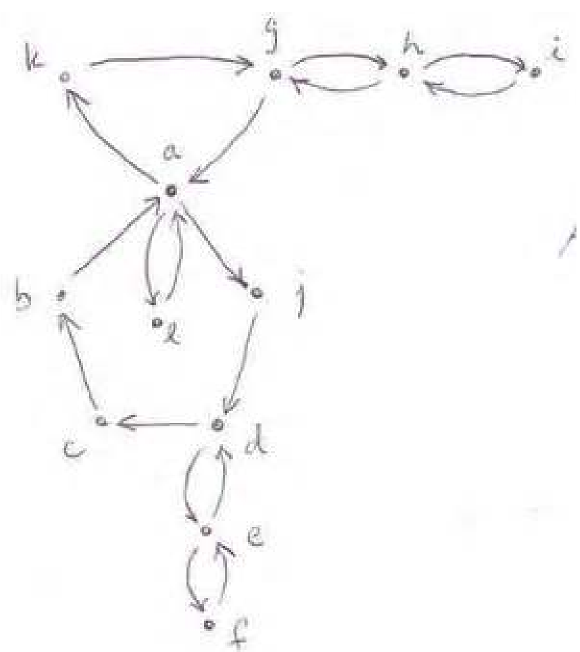
The basic idea of matrix partitioning is to create and study matrices whose elements are matrices – that might seem to be compounding pain with pain, but is actually quite useful.

Matrix factorization of matrix A means to break a matrix into a product. It is carried out for (at least) two different reasons:

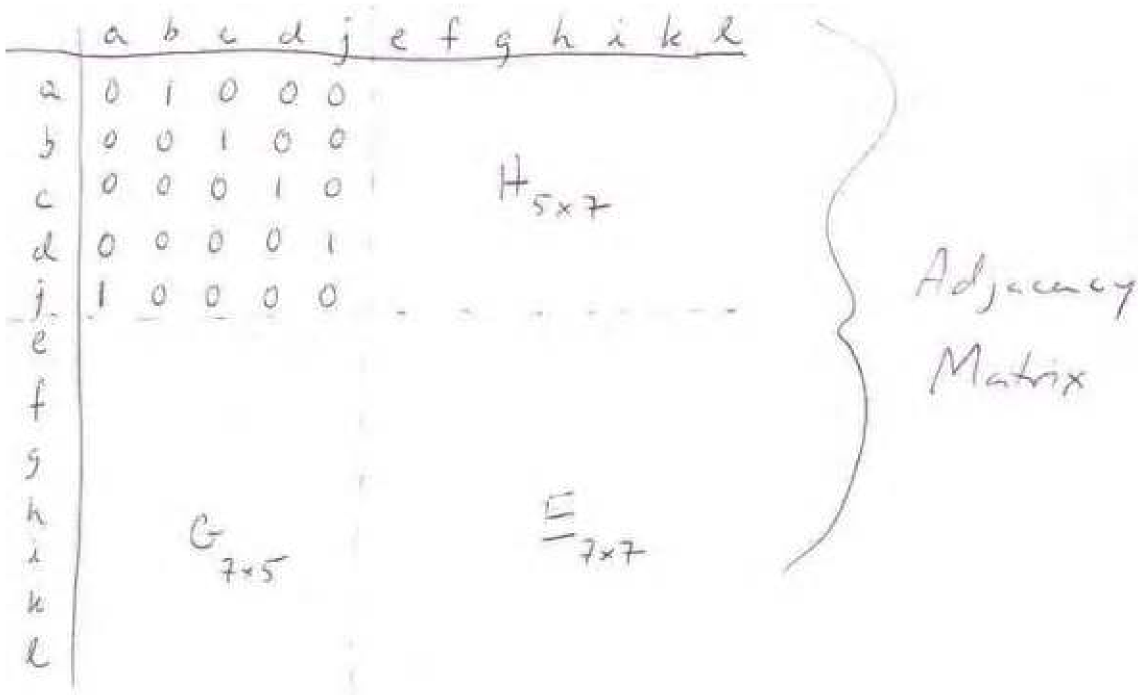
- because it's advantageous to think of A as a series of successive linear transformations, or
- to bring out some structure in the matrix A.

1 Section 2.4

For example, in the "proof graph" of Theorem 8 of section 2.3, we



might isolate the nodes making up the pentagonal cycle (a, b, c, d, and j) and form a partitioned matrix



The matrix in the upper left-hand corner is a “permutation matrix”, because it simply permutes the elements of the set of nodes a, b, c, d, and j. By contrast to the whole matrix, this matrix can be multiplied by itself as long as you want, and you will never get a “full” matrix (a matrix with few zeros): you will always get a matrix with exactly five non-zero elements.

Matlab code of this permutation matrix:

```
M=[0,1,0,0,0
   0,0,1,0,0
   0,0,0,1,0
   0,0,0,0,1
   1,0,0,0,0]
M^2
M^3
M^4
M^5
M+M^2+M^3+M^4
```

What happens when M is multiplied by itself five times?

Note that the remaining matrices also have well-defined meanings pertaining to different “activities” among the nodes:

- matrix E stands for “etcetera”: connections among the nodes other than a, b, c, d, and j;
- matrix G represents connections from the nodes a, b, c, d, and j to the others; and
- matrix H represents connections from the others to the nodes a, b, c, d, and j.

If either of G or H are zero matrices, then the nodes (representing different statements) will not have been shown to be equivalent, because, though we may be able to get from one group of nodes to the other, we won't be able to get back.

Example: #2 and 3, p. 139

#2 $\begin{bmatrix} E & 0 \\ 0 & F \end{bmatrix}\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} EA & EB \\ FC & FD \end{bmatrix}$

#3 $\begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix}\begin{bmatrix} W & X \\ Y & Z \end{bmatrix} = \begin{bmatrix} Y & Z \\ W & X \end{bmatrix}$

The other really neat thing that this section presents is the idea of matrix multiplication as a sum of outer-products. An outer-product of two vectors $\mathbf{u}$ and $\mathbf{v}$ is the matrix product

$$\mathbf{u}_{m \text{ x } 1}\mathbf{v}^T_{n \text{ x } 1}$$

$$\begin{bmatrix} \underline{u} \end{bmatrix}_{m \times 1}\begin{bmatrix} \underline{v}^T \end{bmatrix}_{1\times n} = \begin{bmatrix} u_1v_1 & \cdots & u_1v_n \\ \vdots & & \vdots \\ u_mv_1 & \cdots & u_mv_n \end{bmatrix}_{m\times n}$$

so the result is a matrix $A_{m \text{ x } n}$, whose entries are $a_{ij} = u_i v_j$. We can also form the outer-product $\mathbf{vu}^T$ from these two vectors, of course.

So a matrix product AB can be thought of as

$$AB = [\mathbf{a}_{\cdot 1}\mathbf{a}_{\cdot 2}\cdots\mathbf{a}_{\cdot n}]\begin{bmatrix} \mathbf{b}^T_{1\cdot} \\ \mathbf{b}^T_{2\cdot} \\ \cdots \\ \mathbf{b}^T_{n\cdot} \end{bmatrix} = \mathbf{a}_{\cdot 1}\mathbf{b}^T_{1\cdot} + \mathbf{a}_{\cdot 2}\mathbf{b}^T_{2\cdot} + \cdots \mathbf{a}_{\cdot n}\mathbf{b}^T_{n\cdot}$$

(where the "$\cdot$" in the indices indicates which of rows or columns is being chosen – if the dot occurs first, it's a column; second, it's a row).

Example: #17, p. 140

$X_k = [\underline{x}_1 \cdots \underline{x}_k]$

$G_k = X_k X_k^T$

$= \begin{bmatrix} | & | & \cdots \end{bmatrix}\begin{bmatrix} — \\ \vdots \\ — \end{bmatrix}$

$= \underline{x}_1\underline{x}_1^T + \underline{x}_2\underline{x}_2^T + \cdots + \underline{x}_k\underline{x}_k^T$

$\underline{x}_{k+1}$ arrives; we want G_{k+1}

$G_{k+1} = G_k + \underline{x}_{k+1}\cdot\underline{x}_{k+1}^T$

$AB = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} | & | \end{bmatrix}\begin{bmatrix} — \\ — \end{bmatrix}$

$= \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix}$

$= \begin{bmatrix} a_{11}b_{11} & a_{11}b_{12} \\ a_{21}b_{11} & a_{21}b_{12} \end{bmatrix} + \begin{bmatrix} a_{12}b_{21} & a_{12}b_{22} \\ a_{22}b_{21} & a_{22}b_{22} \end{bmatrix}$

$= \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix}\begin{bmatrix} b_{11} & b_{12} \end{bmatrix} + \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix}\begin{bmatrix} b_{21} & b_{22} \end{bmatrix}$

2 Section 2.5

An example that we've already encountered is thinking of the matrix inverse of square matrix A, A^{-1}, as a product of elementary matrices:

$$E_p \cdots E_2 E_1 = A^{-1}$$

$$E_p \cdots E_2 E_1 \begin{bmatrix} A & I \end{bmatrix} \sim \begin{bmatrix} I & A^{-1} \end{bmatrix}$$

We're focusing on the LU decomposition, which is one strategy for solving linear equations $A\mathbf{x} = \mathbf{b}$. Rather than compute A^{-1} (when possible) and multiply $\mathbf{b}$ by it, it's more advantageous to simply factor $A = LU$, where L is lower-triangular and U is upper-triangular.

This process is better conditioned numerically, and may have other advantages: for example, if A is sparse (has lots of zeros), the LU decomposition may also have many zeros, but the inverse A^{-1} tends to be "full" (that is, of non-zero elements).

In order to solve $A\mathbf{x} = \mathbf{b}$, we proceed as follows:

$$A\mathbf{x} = L\underbrace{U\mathbf{x}}_{\mathbf{y}} = \mathbf{b},$$

so we can solve this in two steps: first, solve for $\mathbf{y}$ in

$$L\mathbf{y} = \mathbf{b}, \quad \text{(by back substitution)}$$

and then $\boxed{\text{solve for } \mathbf{x}}$ in

$$U\mathbf{x} = \mathbf{y}. \quad \text{(again by back substitution)}$$

It doesn't seem that we've made much progress, since we've replaced one equation by two, until we notice that it's easy to solve both of the new equations since they're triangular (just use back-substitution). If we need to solve many equations of the form $A\mathbf{x} = \mathbf{b}$, with fixed A, then it often makes sense to first factor into LU.

Example: #2, p. 149

$$A = \begin{bmatrix} 4 & 3 & -5 \\ -4 & -5 & 7 \\ 8 & 6 & -8 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 4 & 3 & -5 \\ 0 & -2 & 2 \\ 0 & 0 & 2 \end{bmatrix} \qquad \underline{b} = \begin{bmatrix} 2 \\ -4 \\ 6 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 2 \\ -4 \\ 6 \end{bmatrix} \Rightarrow \begin{aligned} y_1 &= 2 \\ y_2 &= -4 + y_1 = -4 + 2 = -2 \\ y_3 &= 6 - 2y_1 = 6 - 2 \cdot 2 = 2 \end{aligned}$$

Now backsolve $U\underline{x} = \underline{y}$

$$\begin{bmatrix} 4 & 3 & -5 \\ 0 & -2 & 2 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ -2 \\ 2 \end{bmatrix}$$

$$\begin{aligned} 4x_1 &= 2 - 3x_2 + 5x_3 = 2 - 6 + 5 = 1 \\ -2x_2 &= -2 - 2x_3 = -4 \qquad x_2 = 2 \\ x_3 &= 1 \end{aligned}$$

$$\underline{x} = \begin{bmatrix} 1/4 \\ 2 \\ 1 \end{bmatrix}$$

The algorithm for finding the LU factorization is simple:

(a) Reduce A to an echelon form U by a sequence of row replacement operations, if possible.

(b) Place entries in L such that the same sequence of row operations reduces L to I (and such that the "diagonal entries" (those with equal indices, a_{ii}) of L are one – remember, L is not necessarily square).

Example: #10, p. 150 (forget pivoting)

$$A = \begin{bmatrix} -5 & 3 & 4 \\ 10 & -8 & -9 \\ 15 & 1 & 2 \end{bmatrix} \sim \begin{bmatrix} -5 & 3 & 4 \\ 0 & -2 & -1 \\ 0 & 10 & 14 \end{bmatrix} \sim \begin{bmatrix} -5 & 3 & 4 \\ 0 & -2 & -1 \\ 0 & 0 & 9 \end{bmatrix} = U$$

$$L = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -3 & -5 & 1 \end{bmatrix}$$

$$\frac{1}{-5}\begin{bmatrix} -5 \\ 10 \\ 15 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ -3 \end{bmatrix} \qquad \frac{1}{-2}\begin{bmatrix} -2 \\ 10 \end{bmatrix} = \begin{bmatrix} 1 \\ -5 \end{bmatrix}$$

Example: #16, p. ~~150~~ 149

$$A = \begin{bmatrix} 2 & -6 & 6 \\ -4 & 5 & -7 \\ 3 & 5 & -1 \\ -6 & 4 & -8 \\ 8 & -3 & 9 \end{bmatrix}_{5\times 3} \sim \begin{bmatrix} 2 & -6 & 6 \\ 0 & -7 & 5 \\ 0 & 14 & -10 \\ 0 & -14 & 10 \\ 0 & 21 & -15 \end{bmatrix} \sim \begin{bmatrix} 2 & -6 & 6 \\ 0 & -7 & 5 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}_{5\times 3} = U$$

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 & 0 \\ 3/2 & -2 & 1 & 0 & 0 \\ -3 & 2 & 0 & 1 & 0 \\ 4 & -3 & 0 & 0 & 1 \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & 0 \\ -2 & 1 \\ 3/2 & -2 \\ -3 & 2 \\ 4 & -3 \end{bmatrix} \begin{bmatrix} 2 & -6 & 6 \\ 0 & -7 & 5 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ -2 \\ 3/2 \\ -3 \\ 4 \end{bmatrix} \begin{bmatrix} 2 & -6 & 6 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ -2 \\ 2 \\ -3 \end{bmatrix} \begin{bmatrix} 0 & -7 & 5 \end{bmatrix}$$

Example: #25, p. 150

#18, p140 $\quad X_{m\times n}$, $\quad X^TX$ invertible

$$W = [X \quad \underline{x}_0]$$

$$W^T = \begin{bmatrix} X^T \\ \underline{x}_0^T \end{bmatrix}$$

$$W^TW = \begin{bmatrix} X^TX & X^T\underline{x}_0 \\ \underline{x}_0^TX & \underline{x}_0^T\underline{x}_0 \end{bmatrix}$$

Identify X^TX with A_{11} from #15

Schur complement of X^TX is

$$S = \underline{x}_0^T\underline{x}_0 - \underline{x}_0^TX(X^TX)^{-1}X^T\underline{x}_0$$

Let $M = I_m - X(X^TX)^{-1}X^T$

$$S = \underline{x}_0^T I_m \underline{x}_0 - \underline{x}_0^TX(X^TX)^{-1}X^T\underline{x}_0$$

$$= \underline{x}_0^T\left[I_m\underline{x}_0 - X(X^TX)^{-1}X^T\underline{x}_0\right]$$

$$= \underline{x}_0^T\left[I_m - X(X^TX)^{-1}X^T\right]\underline{x}_0$$

$$= \underline{x}_0^T M \underline{x}_0$$

#15 p139

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} A_{11} & 0 \\ XA_{11} & S \end{bmatrix}\begin{bmatrix} I & Y \\ 0 & I \end{bmatrix}$$

A_{11} invertible

$$= \begin{bmatrix} A_{11} & A_{11}Y \\ XA_{11} & XA_{11}Y + S \end{bmatrix}$$

Find X & Y.

$$A_{11}Y = A_{12}$$
$$XA_{11} = A_{21}$$

$$A_{11}^{-1}A_{11}Y = A_{11}^{-1}A_{12}$$
$$\boxed{Y = A_{11}^{-1}A_{12}}$$

$$XA_{11}A_{11}^{-1} = A_{21}A_{11}^{-1}$$
$$\boxed{X = A_{21}A_{11}^{-1}}$$

$$A_{22} = X A_{11} Y + S$$

$$A_{22} = (A_{21} A_{11}^{-1}) A_{11} (A_{11}^{-1} A_{12}) + S$$

$$= A_{21} A_{11}^{-1} A_{12} + S$$

$$S = A_{22} - A_{21} A_{11}^{-1} A_{12}$$