1111: Linear Algebra I

Dr. Vladimir Dotsenko (Vlad)

Michaelmas Term 2015

Elementary matrices

Let us define elementary matrices. By definition, an elementary matrix is an $n \times n$-matrix obtained from the identity matrix I_n by one elementary row operation.

Recall that there were elementary operations of three types: swapping rows, re-scaling rows, and combining rows. This leads to elementary matrices S_{ij}, obtained from I_n by swapping rows i and j, $R_i(c)$, obtained from I_n by multiplying the row i by c, and $E_{ij}(c)$, obtained from the identity matrix by adding to the row i the row j multiplied by c.

Exercise. Write these matrices explicitly.

Main property of elementary matrices

Our definition of elementary matrices may appear artificial, but we shall now see that it agrees wonderfully with the definition of the matrix product.

Theorem. Let E be an elementary matrix obtained from I_n by a certain elementary row operation $\mathcal{E}$, and let A be some $n \times k$-matrix. Then the result of the row operation $\mathcal{E}$ applied to A is equal to $E \cdot A$.

Proof. By inspection, or by noticing that elementary row operations combine rows, and the matrix product $I_n \cdot A = A$ computes dot products of rows with columns, so an operation on rows of the first factor results in the same operation on rows of the product.

Invertible matrices

An $m \times n$-matrix A is said to be invertible, if there exists an $n \times m$-matrix B such that $A \cdot B = I_m$ and $B \cdot A = I_n$.

Why are invertible matrices useful? If a matrix is invertible, it is very easy to solve $A \cdot \mathbf{x} = \mathbf{b}$! Indeed,

$$B \cdot \mathbf{b} = B \cdot A \cdot \mathbf{x} = I_n \cdot \mathbf{x} = \mathbf{x} \ .$$

Some important properties:

- The equalities $A \cdot B = I_m$ and $B \cdot A = I_n$ can hold for at most one matrix B; indeed, if it holds for two matrices B_1 and B_2, we have

$$B_1 = B_1 \cdot I_m = B_1 \cdot (A \cdot B_2) = (B_1 \cdot A) \cdot B_2 = I_n \cdot B_2 = B_2 \ .$$

  Thus the matrix B can be called *the inverse of* A and be denoted A^{-1}.
- If both matrices A_1 and A_2 are invertible, and their product is defined, then $A_1 A_2$ is invertible, and $(A_1 A_2)^{-1} = A_2^{-1} A_1^{-1}$; indeed, for example

$$(A_1 A_2) A_2^{-1} A_1^{-1} = A_1 (A_2 A_2^{-1}) A_1^{-1} = A_1 I_{m_2} A_1^{-1} = A_1 A_1^{-1} = I_{m_1} \ .$$

  (As they say, "you put your socks on before putting on your shoes, but take them off after taking off your shoes").

Invertible matrices

Theorem. 1. An elementary matrix is invertible.
2. If an $n \times m$-matrix A is invertible, then $m = n$.
3. An $n \times n$-matrix A is invertible if and only if it can be represented as a product of elementary matrices.

Proof. 1. If $A = E$ is an elementary matrix, then for B we can take the matrix corresponding to the inverse row operation. Then $AB = I_n = BA$ since we know that multiplying by an elementary matrix performs the actual row operation.
2. Suppose that $m \neq n$, and there exist matrices A and B such that $A \cdot B = I_m$ and $B \cdot A = I_n$. Without loss of generality, $m > n$ (otherwise swap A with B). Let us show that $AB = I_m$ leads to a contradiction. We have $E_1 \cdot E_2 \cdots E_p \cdot A = R$, where R is the reduced row echelon form of A, and E_i are appropriate elementary matrices. Therefore,

$$R \cdot B = E_1 \cdot E_2 \cdots E_p \cdot A \cdot B = E_1 \cdot E_2 \cdots E_p \ .$$

Invertible matrices

From $R \cdot B = E_1 \cdot E_2 \cdots E_p$, we immediately deduce

$$R \cdot B \cdot (E_p)^{-1} \cdots (E_2)^{-1} \cdot (E_1)^{-1} = I_m \ .$$

But if we assume $m > n$, the last row of R is inevitably zero (there is no room for m pivots), so the last row of $R \cdot B \cdot (E_p)^{-1} \cdots (E_2)^{-1} \cdot (E_1)^{-1}$ is zero too, a contradiction.

3. If A can be represented as a product of elementary matrices, it is invertible, since products of invertible matrices are invertible. If A is invertible, then the last row of its reduced row echelon form must be non-zero, or we get a contradiction like in the previous argument. Therefore, each row of the reduced row echelon form of A, and hence, by previous result, each column of the reduced row echelon form of A, has a pivot, so the reduced row echelon form of A is the identity matrix. We conclude that $E_1 \cdot E_2 \cdots E_p \cdot A = I_n$, so $A = (E_p)^{-1} \cdots (E_2)^{-1} \cdot (E_1)^{-1}$, which is a product of elementary matrices. □

One more property of inverses

There is another useful property that is proved completely analogously:

> *If for an $n \times n$-matrix A, there exists a "one-sided" inverse (that is, B for which only one of the two conditions $AB = I_n$ and $BA = I_n$ are satisfied), then $B = A^{-1}$.*

To prove it, it is enough to consider the case $AB = I_n$ (otherwise we can swap the roles of A and B). In this case, we proceed as before to conclude that the reduced row echelon form of A cannot have a row of zeros, hence that reduced row echelon form is the identity matrix, hence A is invertible. Finally, $A^{-1}(AB) = (A^{-1}A)B = I_nB = B$.

Warning: we know that for $m \neq n$ an $m \times n$-matrix cannot be invertible, but such a matrix can have a one-sided inverse. You will be asked to construct an example in the next homework.

Computing inverses

Our results lead to an elegant algorithm for computing the inverse of an $n \times n$-matrix A.

Form an $n \times (2n)$-matrix $(A \mid I_n)$. Apply the usual algorithm to compute its reduced row echelon form. If A is invertible, the output is a matrix of the form $(I_n \mid B)$, where $B = A^{-1}$.

Justification. If A is invertible, its reduced row echelon form is the identity matrix I_n. Therefore, the computation of the reduced row echelon form of $(A \mid I_n)$ will produce a matrix of the form $(I_n \mid B)$, since pivots emerge from the left to the right. This matrix is clearly in its reduced row echelon form. Let us take the elementary matrices corresponding to the appropriate row operations, so that $E_1 \cdot E_2 \cdots E_p \cdot A = I_n$. This means, as we just proved, that $A^{-1} = E_1 \cdot E_2 \cdots E_p$. It remains to remark that

$$E_1 \cdot E_2 \cdots E_p \cdot (A \mid I_n) = (E_1 \cdot E_2 \cdots E_p \cdot A \mid E_1 \cdot E_2 \cdots E_p),$$

so $(I_n \mid B) = (I_n \mid E_1 \cdot E_2 \cdots E_p) = (I_n \mid A^{-1})$. □