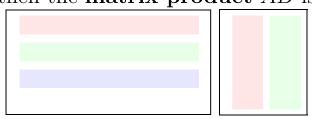
Math 19b: Linear Algebra with Probability **Oliver Knill, Spring 2011**

Lecture 9: Matrix algebra

If A is a $n \times m$ matrix and A is a $m \times p$ matrix, then the **matrix product** AB is defined as the $n \times p$ matrix with entries $(BA)_{ij} = \sum_{k=1}^{m} B_{ik} A_{kj}$.



1 If B is a 3×4 matrix, and A is a 4×2 matrix then BA is a 3×2 matrix. For example:

$$B = \begin{bmatrix} 1 & 3 & 5 & 7 \\ 3 & 1 & 8 & 1 \\ 1 & 0 & 9 & 2 \end{bmatrix}, \; A = \begin{bmatrix} 1 & 3 \\ 3 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \; BA = \begin{bmatrix} 1 & 3 & 5 & 7 \\ 3 & 1 & 8 & 1 \\ 1 & 0 & 9 & 2 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 3 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 15 & 13 \\ 14 & 11 \\ 10 & 5 \end{bmatrix}.$$

If A is a $n \times n$ matrix and $T : \vec{x} \mapsto Ax$ has an inverse S, then S is linear. The matrix A^{-1} belonging to $S = T^{-1}$ is called the **inverse matrix** of A.

Matrix multiplication generalizes the common multiplication of numbers. We can write the dot product between two vectors as a matrix product when writing the first vector as a $1 \times n$ matrix (= row vector) and the second as a $n \times 1$ matrix (=column vector) like in $[1\ 2\ 3] \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} = 20$. Note that $AB \neq BA$ in general and for $n \times n$ matrices, the inverse A^{-1} does not always exist, otherwise, for $n \times n$ matrices the same rules apply as for numbers: $A(BC) = (AB)C$, $AA^{-1} = A^{-1}A = 1_n$, $(AB)^{-1} = B^{-1}A^{-1}$, $A(B+C) = AB + AC$, $(B+C)A = BA + CA$ etc.

2 The entries of matrices can themselves be matrices. If B is a $n \times p$ matrix and A is a $p \times m$ matrix, and assume the entries are $k \times k$ matrices, then BA is a $n \times m$ matrix, where each entry $(BA)_{ij} = \sum_{l=1}^{p} B_{il} A_{lj}$ is a $k \times k$ matrix. Partitioning matrices can be useful to improve the speed of matrix multiplication. If $A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$, where A_{ij} are $k \times k$ matrices with the property that A_{11} and A_{22} are invertible, then one can write the inverse as $B = \begin{bmatrix} A_{11}^{-1} & -A_{11}^{-1} A_{12} A_{22}^{-1} \\ 0 & A_{22}^{-1} \end{bmatrix}$ is the inverse of A.

3 Let us associate to a small bloging network a matrix $\begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$ and look at the spread of some news. Assume the source of the news about some politician is the first entry (maybe the gossip news "gawker") $\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$. The vector Ax has a 1 at the places, where the news could be in the next hour. The vector $(AA)(x)$ tells, in how many ways the news can go in 2 steps. In our case, it can go in three different ways back to the page itself.

Matrices help to solve combinatorial problems. One appears in the movie "Good will hunting". For example, what does $[A^{100}]$ tell about the news distribution on a large network. What does it mean if A^{100} has no zero entries?

If A is a $n \times n$ matrix and the system of linear equations $Ax = y$ has a unique solution for all y, we write $x = A^{-1}y$. The inverse matrix can be computed using Gauss-Jordan elimination. Lets see how this works.

Let 1_n be the $n \times n$ identity matrix. Start with $[A|1_n]$ and perform Gauss-Jordan elimination. Then

$$\text{rref}([A|1_n]) = [1_n|A^{-1}]$$

Proof. The elimination process solves $A\vec{x} = \vec{e}_i$ simultaneously. This leads to solutions $\vec{v}_i$ which are the columns of the inverse matrix A^{-1} because $A^{-1}\vec{e}_i = \vec{v}_i$.

$$\left[\begin{array}{cc|cc} 2 & 6 & 1 & 0 \\ 1 & 4 & 0 & 1 \end{array}\right] \qquad [\ A \ |\ 1_2\ ]$$

$$\left[\begin{array}{cc|cc} 1 & 3 & 1/2 & 0 \\ 1 & 4 & 0 & 1 \end{array}\right] \qquad [\ .... \ |\ ...\ ]$$

$$\left[\begin{array}{cc|cc} 1 & 3 & 1/2 & 0 \\ 0 & 1 & -1/2 & 1 \end{array}\right] \qquad [\ .... \ |\ ...\ ]$$

$$\left[\begin{array}{cc|cc} 1 & 0 & 2 & -3 \\ 0 & 1 & -1/2 & 1 \end{array}\right] \qquad [\ 1_2 \ |\ A^{-1}\ ]$$

The inverse is $A^{-1} = \begin{bmatrix} 2 & -3 \\ -1/2 & 1 \end{bmatrix}$.

If $ad - bc \neq 0$, the inverse of a linear transformation $\vec{x} \mapsto Ax$ with $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is given by the matrix $A^{-1} = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} / (ad - bc)$.

Shear:

$$A = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \qquad A^{-1} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

Diagonal:

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \qquad A^{-1} = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/3 \end{bmatrix}$$

Reflection:

$$A = \begin{bmatrix} \cos(2\alpha) & \sin(2\alpha) \\ \sin(2\alpha) & -\cos(2\alpha) \end{bmatrix} \qquad A^{-1} = A = \begin{bmatrix} \cos(2\alpha) & \sin(2\alpha) \\ \sin(2\alpha) & -\cos(2\alpha) \end{bmatrix}$$

Rotation:

$$A = \begin{bmatrix} \cos(\alpha) & \sin(\alpha) \\ -\sin(\alpha) & \cos(-\alpha) \end{bmatrix} \qquad A^{-1} = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix}$$

Rotation=Dilation:

$A = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ $\qquad A^{-1} = \begin{bmatrix} a/r^2 & b/r^2 \\ -b/r^2 & a/r^2 \end{bmatrix}, r^2 = a^2 + b^2$

Homework due February 16, 2011

1 Find the inverse of the following matrix

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} .$$

2 The probability density of a multivariate normal distribution centered at the origin is a multiple of

$$f(x) = \exp(-x \cdot A^{-1}x)$$

We will see the covariance matrix later. It encodes how the different coordinates of a random vector are correlated. Assume

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

Find $f(\langle[1, 2, -3])$.

3 This is a system we will analyze more later. For now we are only interested in the algebra. **Tom** the cat moves each minute randomly from on spots 1,5,4 jumping to neighboring sites only. At the same time **Jerry**, the mouse, moves on spots 1,2,3, also jumping to neighboring sites. The possible position combinations $(2, 5), (3, 4), (3, 1), (1, 4), (1, 1)$ and transitions are encoded in a matrix

$$A = \begin{array}{ccccc} 0 & 1 & 1 & 1 & 0 \\ 1/4 & 0 & 0 & 0 & 0 \\ 1/4 & 0 & 0 & 0 & 0 \\ 1/4 & 0 & 0 & 0 & 0 \\ 1/4 & 0 & 0 & 0 & 1 \end{array}$$

This means for example that we can go from state $(2, 5)$ with equal probability to all the other states. In state $(3, 4)$ or $(3, 1)$ the pair (Tom,Jerry) moves back to $(2, 5)$. If state $(1, 1)$ is reached, then Jerry's life ends and Tom goes to sleep there. We can read off that the probability that Jerry gets eaten in one step as $1/4$. Compute A^4. The first column of this matrix gives the probability distribution after four steps. The last entry of the first column gives the probability that Jerry got swallowed after 4 steps.