

6. Matrices

Outline

Matrices

Matrix-vector multiplication

Examples

Matrices

- ▶ a *matrix* is a rectangular array of numbers, *e.g.*,

$$\begin{bmatrix} 0 & 1 & -2.3 & 0.1 \\ 1.3 & 4 & -0.1 & 0 \\ 4.1 & -1 & 0 & 1.7 \end{bmatrix}$$

- ▶ its *size* is given by (row dimension) \times (column dimension)
e.g., matrix above is 3×4
- ▶ *elements* also called *entries* or *coefficients*
- ▶ B_{ij} is i, j element of matrix B
- ▶ i is the *row index*, j is the *column index*; indexes start at 1
- ▶ two matrices are *equal* (denoted with $=$) if they are the same size and corresponding entries are equal

Matrix shapes

an $m \times n$ matrix A is

- ▶ *tall* if $m > n$
- ▶ *wide* if $m < n$
- ▶ *square* if $m = n$

Column and row vectors

- ▶ we consider an $n \times 1$ matrix to be an n -vector
- ▶ we consider a 1×1 matrix to be a number
- ▶ a $1 \times n$ matrix is called a *row vector*, *e.g.*,

$$\begin{bmatrix} 1.2 & -0.3 & 1.4 & 2.6 \end{bmatrix}$$

which is *not* the same as the (column) vector

$$\begin{bmatrix} 1.2 \\ -0.3 \\ 1.4 \\ 2.6 \end{bmatrix}$$

Columns and rows of a matrix

- ▶ suppose A is an $m \times n$ matrix with entries A_{ij} for $i = 1, \dots, m, j = 1, \dots, n$
- ▶ its j th *column* is (the m -vector)

$$\begin{bmatrix} A_{1j} \\ \vdots \\ A_{mj} \end{bmatrix}$$

- ▶ its i th *row* is (the n -row-vector)

$$\begin{bmatrix} A_{i1} & \cdots & A_{in} \end{bmatrix}$$

- ▶ *slice* of matrix: $A_{p:q,r:s}$ is the $(q - p + 1) \times (s - r + 1)$ matrix

$$A_{p:q,r:s} = \begin{bmatrix} A_{pr} & A_{p,r+1} & \cdots & A_{ps} \\ A_{p+1,r} & A_{p+1,r+1} & \cdots & A_{p+1,s} \\ \vdots & \vdots & & \vdots \\ A_{qr} & A_{q,r+1} & \cdots & A_{qs} \end{bmatrix}$$

Block matrices

- ▶ we can form *block matrices*, whose entries are matrices, such as

$$A = \begin{bmatrix} B & C \\ D & E \end{bmatrix}$$

where B , C , D , and E are matrices (called *submatrices* or *blocks* of A)

- ▶ matrices in each block row must have same height (row dimension)
- ▶ matrices in each block column must have same width (column dimension)
- ▶ example: if

$$B = \begin{bmatrix} 0 & 2 & 3 \end{bmatrix}, \quad C = \begin{bmatrix} -1 \end{bmatrix}, \quad D = \begin{bmatrix} 2 & 2 & 1 \\ 1 & 3 & 5 \end{bmatrix}, \quad E = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$$

then

$$\begin{bmatrix} B & C \\ D & E \end{bmatrix} = \begin{bmatrix} 0 & 2 & 3 & -1 \\ 2 & 2 & 1 & 4 \\ 1 & 3 & 5 & 4 \end{bmatrix}$$

Column and row representation of matrix

- ▶ A is an $m \times n$ matrix
- ▶ can express as block matrix with its (m -vector) columns a_1, \dots, a_n

$$A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}$$

- ▶ or as block matrix with its (n -row-vector) rows b_1, \dots, b_m

$$A = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

Examples

- ▶ *image*: X_{ij} is i, j pixel value in a monochrome image
- ▶ *rainfall data*: A_{ij} is rainfall at location i on day j
- ▶ *multiple asset returns*: R_{ij} is return of asset j in period i
- ▶ *contingency table*: A_{ij} is number of objects with first attribute i and second attribute j
- ▶ *feature matrix*: X_{ij} is value of feature i for entity j

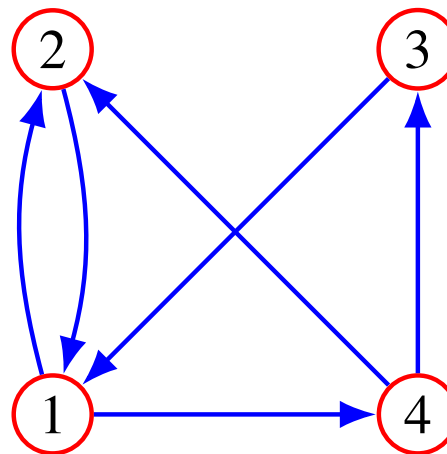
in each of these, what do the rows and columns mean?

Graph or relation

- ▶ a *relation* is a set of pairs of *objects*, labeled $1, \dots, n$, such as

$$\mathcal{R} = \{(1,2), (1,3), (2,1), (2,4), (3,4), (4,1)\}$$

- ▶ same as *directed graph*



- ▶ can be represented as $n \times n$ matrix with $A_{ij} = 1$ if $(i,j) \in \mathcal{R}$

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

Special matrices

- ▶ $m \times n$ zero matrix has all entries zero, written as $0_{m \times n}$ or just 0
- ▶ identity matrix is square matrix with $I_{ii} = 1$ and $I_{ij} = 0$ for $i \neq j$, e.g.,

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

- ▶ *sparse matrix*: most entries are zero
 - examples: 0 and I
 - can be stored and manipulated efficiently
 - **nnz**(A) is number of nonzero entries

Diagonal and triangular matrices

- ▶ *diagonal matrix*: square matrix with $A_{ij} = 0$ when $i \neq j$
- ▶ **diag**(a_1, \dots, a_n) denotes the diagonal matrix with $A_{ii} = a_i$ for $i = 1, \dots, n$
- ▶ example:

$$\mathbf{diag}(0.2, -3, 1.2) = \begin{bmatrix} 0.2 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 1.2 \end{bmatrix}$$

- ▶ *lower triangular matrix*: $A_{ij} = 0$ for $i < j$
- ▶ *upper triangular matrix*: $A_{ij} = 0$ for $i > j$
- ▶ examples:

$$\begin{bmatrix} 1 & -1 & 0.7 \\ 0 & 1.2 & -1.1 \\ 0 & 0 & 3.2 \end{bmatrix} \text{ (upper triangular),} \quad \begin{bmatrix} -0.6 & 0 \\ -0.3 & 3.5 \end{bmatrix} \text{ (lower triangular)}$$

Transpose

- ▶ the *transpose* of an $m \times n$ matrix A is denoted A^T , and defined by

$$(A^T)_{ij} = A_{ji}, \quad i = 1, \dots, n, \quad j = 1, \dots, m$$

- ▶ for example,

$$\begin{bmatrix} 0 & 4 \\ 7 & 0 \\ 3 & 1 \end{bmatrix}^T = \begin{bmatrix} 0 & 7 & 3 \\ 4 & 0 & 1 \end{bmatrix}$$

- ▶ transpose converts column to row vectors (and vice versa)
- ▶ $(A^T)^T = A$

Addition, subtraction, and scalar multiplication

- ▶ (just like vectors) we can add or subtract matrices of the same size:

$$(A + B)_{ij} = A_{ij} + B_{ij}, \quad i = 1, \dots, m, \quad j = 1, \dots, n$$

(subtraction is similar)

- ▶ scalar multiplication:

$$(\alpha A)_{ij} = \alpha A_{ij}, \quad i = 1, \dots, m, \quad j = 1, \dots, n$$

- ▶ many obvious properties, e.g.,

$$A + B = B + A, \quad \alpha(A + B) = \alpha A + \alpha B, \quad (A + B)^T = A^T + B^T$$

Matrix norm

- ▶ for $m \times n$ matrix A , we define

$$\|A\| = \left(\sum_{i=1}^m \sum_{j=1}^n A_{ij}^2 \right)^{1/2}$$

- ▶ agrees with vector norm when $n = 1$
- ▶ satisfies norm properties:

$$\|\alpha A\| = |\alpha| \|A\|$$

$$\|A + B\| \leq \|A\| + \|B\|$$

$$\|A\| \geq 0$$

$$\|A\| = 0 \text{ only if } A = 0$$

- ▶ distance between two matrices: $\|A - B\|$
- ▶ (there are other matrix norms, which we won't use)

Outline

Matrices

Matrix-vector multiplication

Examples

Matrix-vector product

- ▶ *matrix-vector product* of $m \times n$ matrix A , n -vector x , denoted $y = Ax$, with

$$y_i = A_{i1}x_1 + \cdots + A_{in}x_n, \quad i = 1, \dots, m$$

- ▶ for example,

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

Row interpretation

- ▶ $y = Ax$ can be expressed as

$$y_i = b_i^T x, \quad i = 1, \dots, m$$

where b_1^T, \dots, b_m^T are rows of A

- ▶ so $y = Ax$ is a ‘batch’ inner product of all rows of A with x
- ▶ example: $A\mathbf{1}$ is vector of row sums of matrix A

Column interpretation

- ▶ $y = Ax$ can be expressed as

$$y = x_1 a_1 + x_2 a_2 + \cdots + x_n a_n$$

where a_1, \dots, a_n are columns of A

- ▶ so $y = Ax$ is linear combination of columns of A , with coefficients x_1, \dots, x_n
- ▶ important example: $Ae_j = a_j$
- ▶ columns of A are linearly independent if $Ax = 0$ implies $x = 0$

Outline

Matrices

Matrix-vector multiplication

Examples

General examples

- ▶ $0x = 0$, *i.e.*, multiplying by zero matrix gives zero
- ▶ $Ix = x$, *i.e.*, multiplying by identity matrix does nothing
- ▶ inner product $a^T b$ is matrix-vector product of $1 \times n$ matrix a^T and n -vector b
- ▶ $\tilde{x} = Ax$ is de-meaned version of x , with

$$A = \begin{bmatrix} 1 - 1/n & -1/n & \cdots & -1/n \\ -1/n & 1 - 1/n & \cdots & -1/n \\ \vdots & & \ddots & \vdots \\ -1/n & -1/n & \cdots & 1 - 1/n \end{bmatrix}$$

Difference matrix

- ▶ $(n - 1) \times n$ difference matrix is

$$D = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 & 0 \\ & & \ddots & \ddots & & & \\ & & & \ddots & \ddots & & \\ 0 & 0 & 0 & \cdots & -1 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -1 & 1 \end{bmatrix}$$

$y = Dx$ is $(n - 1)$ -vector of differences of consecutive entries of x :

$$Dx = \begin{bmatrix} x_2 - x_1 \\ x_3 - x_2 \\ \vdots \\ x_n - x_{n-1} \end{bmatrix}$$

- ▶ *Dirichlet energy*: $\|Dx\|^2$ is measure of wiggleness for x a time series

Return matrix – portfolio vector

- ▶ R is $T \times n$ matrix of asset returns
- ▶ R_{ij} is return of asset j in period i (say, in percentage)
- ▶ n -vector w gives portfolio (investments in the assets)
- ▶ T -vector R_w is time series of the portfolio return
- ▶ **avg**(R_w) is the portfolio (mean) return, **std**(R_w) is its risk

Feature matrix – weight vector

- ▶ $X = [x_1 \cdots x_N]$ is $n \times N$ *feature matrix*
- ▶ column x_j is feature n -vector for object or example j
- ▶ X_{ij} is value of feature i for example j
- ▶ n -vector w is weight vector
- ▶ $s = X^T w$ is vector of scores for each example; $s_j = x_j^T w$

Input – output matrix

- ▶ A is $m \times n$ matrix
- ▶ $y = Ax$
- ▶ n -vector x is *input* or *action*
- ▶ m -vector y is *output* or *result*
- ▶ A_{ij} is the factor by which y_i depends on x_j
- ▶ A_{ij} is the *gain* from input j to output i
- ▶ e.g., if A is lower triangular, then y_i only depends on x_1, \dots, x_i

Complexity

- ▶ $m \times n$ matrix stored A as $m \times n$ array of numbers
(for sparse A , store only $\mathbf{nnz}(A)$ nonzero values)
- ▶ matrix addition, scalar-matrix multiplication cost mn flops
- ▶ matrix-vector multiplication costs $m(2n - 1) \approx 2mn$ flops
(for sparse A , around $2\mathbf{nnz}(A)$ flops)