2. Matrices

- definition and notation
- matrix-vector product
- matrix-matrix product
- orthogonal matrices
- matrix norm

Matrices

$m \times n$-matrix A:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

- a_{ij} are the *elements* or *coefficients*
- set of $m \times n$-matrices is denoted $\mathbf{R}^{m \times n}$
- m, n are the *dimensions*

special matrices (dimensions follow from context)

- $A = 0$ (zero matrix): $a_{ij} = 0$ for $i = 1, \ldots, m$, $j = 1, \ldots, n$
- $A = I$ (identity matrix): $m = n$, $a_{ii} = 1$, $a_{ij} = 0$ for $i \neq j$

Block matrix notation

example: 2×2-block matrix A

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

for example, if B, C, D, E are defined as

$$B = \begin{bmatrix} 2 & 2 \\ 1 & 3 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 2 & 3 \\ 5 & 4 & 7 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad E = \begin{bmatrix} -1 & 6 & 0 \end{bmatrix}$$

then A is the matrix

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

note: dimensions of the blocks must be compatible!

Matrix transpose

$$A^T = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}$$

- A^T is $n \times m$ if A is $m \times n$
- $(A^T)^T = A$
- a square matrix A is *symmetric* if $A = A^T$, *i.e.*, $a_{ij} = a_{ji}$

Scalar multiplication and addition

scalar multiplication of an $m \times n$-matrix A with a scalar β

$$\beta A = \begin{bmatrix} \beta\, a_{11} & \beta\, a_{12} & \cdots & \beta\, a_{1n} \\ \beta\, a_{21} & \beta\, a_{22} & \cdots & \beta\, a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \beta\, a_{m1} & \beta\, a_{m2} & \cdots & \beta\, a_{mn} \end{bmatrix}$$

addition of two $m \times n$-matrices A and B

$$A + B = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2n} + b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \cdots & a_{mn} + b_{mn} \end{bmatrix}$$

Matrix-vector product

product of $m \times n$-matrix A with n-vector x

$$Ax = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix}$$

- dimensions must be compatible: $\#$ columns in $A = \#$ elements in x
- explains notation $a^T x$ for inner product of vectors a and x:

$$a^T x = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

matrix-vector product of $1 \times n$-matrix a^T with x

Linear functions

definition: a function $f : \mathbf{R}^n \to \mathbf{R}^m$ is linear if superposition holds

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$$

for all n-vectors x, y and all scalars α, β

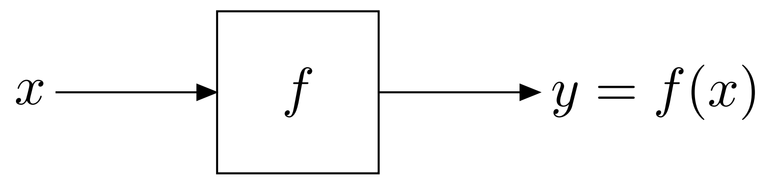
property: f is linear if and only if it can be expressed as

$$f(x) = Ax$$

where A is $m \times n$ and constant (independent of x)

Operator interpretation

think of a function $f : \mathbf{R}^n \to \mathbf{R}^m$ in terms of its effect on x



- signal processing interpretation: a system with n inputs x_i, m outputs y_i
- programming interpretation: a subroutine with n input arguments x_i and m output arguments y_i

f is linear if we can represent its action on x as a a product $f(x) = Ax$

Examples ($f : \mathbf{R}^3 \to \mathbf{R}^3$)

- reverse the order of the components of x: $y_1 = x_3$, $y_2 = x_2$, $y_3 = x_1$

  a linear function: $f(x) = Ax$ with

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

- sort the components of x in decreasing order: not linear
- scale x_1 by a given number d_1, x_2 by d_2, x_3 by d_3

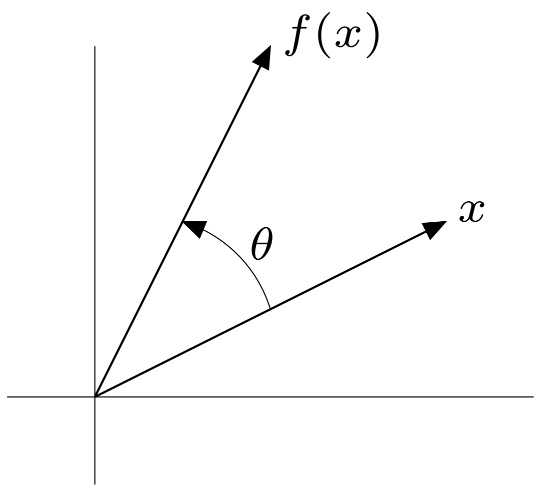
  a linear function: $f(x) = Ax$ with

$$A = \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix}$$

- replace each x_i by its absolute value $|x_i|$: not linear

Geometrical examples ($f : \mathbf{R}^2 \to \mathbf{R}^2$)

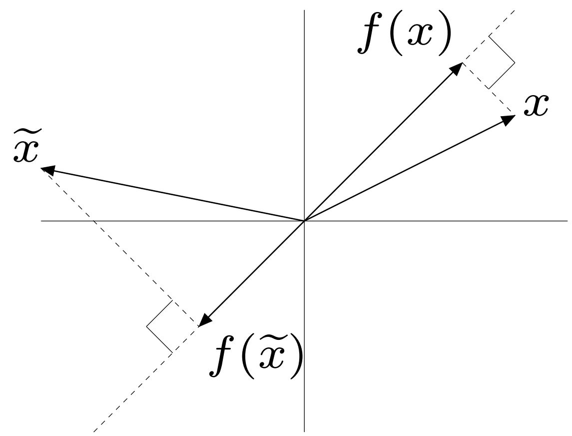
- rotate x counterclockwise over an angle θ



$f(x) = Ax$ with

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

- project x on the line through 0 and $(1,1)$



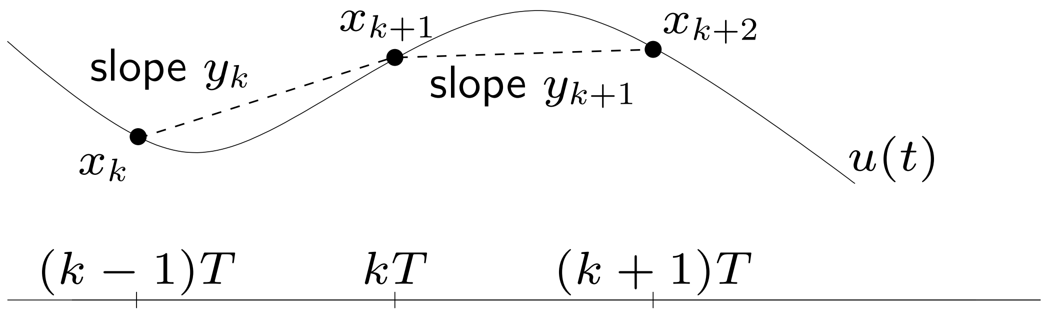
$f(x) = Ax$ with

$$A = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

Example: finite differences of sampled signal

for x, u as defined on page 1-5, define $(n-1)$-vector y as

$$y_k = \frac{1}{T}(u(kT) - u((k-1)T)) = \frac{1}{T}(x_{k+1} - x_k), \quad k = 1, \ldots, n-1$$



a linear function of x: $y = Ax$ with

$$A = \frac{1}{T}\begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}$$

Linearization

derivative or Jacobian matrix of $f : \mathbf{R}^n \to \mathbf{R}^m$

$$Df(\hat{x}) = \begin{bmatrix} \partial f_1(\hat{x})/\partial x_1 & \partial f_1(\hat{x})/\partial x_2 & \cdots & \partial f_1(\hat{x})/\partial x_n \\ \partial f_2(\hat{x})/\partial x_1 & \partial f_2(\hat{x})/\partial x_2 & \cdots & \partial f_2(\hat{x})/\partial x_n \\ \vdots & \vdots & & \vdots \\ \partial f_m(\hat{x})/\partial x_1 & \partial f_m(\hat{x})/\partial x_2 & \cdots & \partial f_m(\hat{x})/\partial x_n \end{bmatrix}$$

first-order Taylor approximation of f around $\hat{x}$

$$f_{\text{aff}}(x) = f(\hat{x}) + Df(\hat{x})(x - \hat{x})$$

- an affine function of x
- a good approximation of $f(x)$ for x around $\hat{x}$

Matrix-matrix product

product of $m \times n$-matrix A with $n \times p$-matrix B:

$$(AB)_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}$$

dimensions must be compatible: $\#$ of columns in $A = \#$ of rows in B

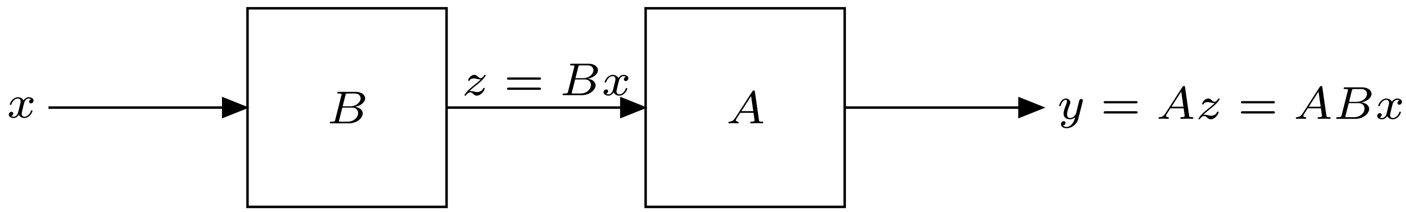
- $AB \neq BA$ in general! (even if the dimensions make sense)

$$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \neq \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

- there are exceptions, $e.g.$, $A\,I = I\,A$ for all square A
- $(AB)^T = B^T A^T$

Operator interpretation

$$y = (AB)x = A(Bx)$$



explains why in general $AB \neq BA$:

- $f(x) = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

  swaps the two elements of x; then changes the sign of the first element

- $f(x) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

  changes the sign of first element of x; then swaps the two elements

Exercise

undirected graph

- n vertices (nodes)
- undirected edges (links) between certain pairs of vertices i, j

adjacency matrix: matrix of order n defined as

$$\begin{array}{ll} a_{ij} = 1 & \text{if there is an edge between nodes } i \text{ and } j \\ a_{ij} = 0 & \text{otherwise} \end{array}$$

question: give a graph interpretation of powers A^k, $k = 2, 3 \ldots$

Orthogonal matrices

a matrix Q is orthogonal if

$$Q^TQ = I$$

examples

$$\left[\begin{array}{cc} 1/\sqrt{3} & 1/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{6} \\ -1/\sqrt{3} & 2/\sqrt{6} \end{array}\right], \qquad \left[\begin{array}{ccc} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{array}\right], \qquad \left[\begin{array}{cc} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{array}\right]$$

$$Q = I - 2uu^T \quad \text{(if } u \text{ is a vector with } \|u\| = 1\text{)}$$

Column properties

denote the columns of Q by q_k:

$$Q = \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix}$$

then

$$Q^TQ = \begin{bmatrix} q_1^Tq_1 & q_1^Tq_2 & \cdots & q_1^Tq_n \\ q_2^Tq_1 & q_2^Tq_2 & \cdots & q_2^Tq_n \\ \vdots & \vdots & \ddots & \vdots \\ q_n^Tq_1 & q_n^Tq_2 & \cdots & q_n^Tq_n \end{bmatrix} = I$$

- columns q_i have unit norm: $q_i^Tq_i = 1$ for $i = 1, \ldots, n$
- columns are mutually orthogonal: $q_i^Tq_j = 0$ for $i \neq j$

Properties of orthogonal matrices

suppose Q is $m \times n$ and orthogonal

- multiplication with Q preserves norms:

$$\|Qx\| = (x^T Q^T Q x)^{1/2} = (x^T x)^{1/2} = \|x\|$$

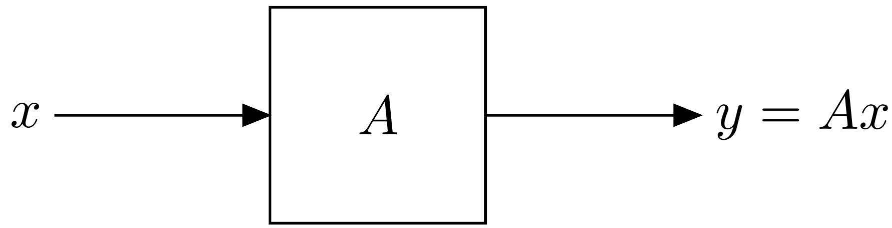
- multiplication with Q preserves inner products:

$$(Qx)^T (Qy) = x^T Q^T Q y = x^T y$$

- multiplication with Q preserves angles between vectors

Norm of a matrix

$m \times n$-matrix A defines a linear function $f(x) = Ax$



- $\|Ax\|/\|x\|$ gives the *amplification factor* or *gain* in the direction x

- we define the *norm* of A as the maximum gain (over all directions):

$$\|A\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \max_{\|x\|=1} \|Ax\|$$

(also called *spectral norm* or *2-norm*)

Computing the norm of a matrix

simple examples where it is easy to maximize $\|Ax\|/\|x\|$:

- zero matrix: $\|0\| = 0$
- identity matrix: $\|I\| = 1$
- diagonal matrix: if

$$A = \begin{bmatrix} \alpha_1 & 0 & \cdots & 0 \\ 0 & \alpha_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha_n \end{bmatrix}$$

  then $\|A\| = \max_{i=1,\ldots,n} |\alpha_i|$

- orthogonal matrix: $\|A\| = 1$

for **general** A, we have to use a numerical algorithm (MATLAB: `norm(A)`)

Other matrix norms

many other definitions of matrix norms exist, $e.g.$, the *Frobenius* norm

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2}$$

- has a simple explicit expression
- in MATLAB: use `norm(A,'fro')`

in this course: $\|A\|$ stands for the 'operator' norm (page 2-19)

- no simple explicit expression (except for special A)
- readily computed numerically

Properties of the matrix norm

- $\|\alpha A\| = |\alpha|\|A\|$
- $\|A\| \geq 0$ for all A; $\|A\| = 0$ iff $A = 0$
- $\|A + B\| \leq \|A\| + \|B\|$
- $\|Ax\| \leq \|A\|\|x\|$ if the product Ax exists
- $\|AB\| \leq \|A\|\|B\|$ if the product AB exists
- if A is nonsingular: $\|A\|\|A^{-1}\| \geq 1$
- if A is nonsingular: $1/\|A^{-1}\| = \min_{x \neq 0} (\|Ax\|/\|x\|)$
- $\|A^T\| = \|A\|$