1. Vectors

- definition and notation
- inner products and norms
- linear functions
- vector norm
- angle between vectors

Vectors

n-vector x:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

- also written as $x = (x_1, x_2, \ldots, x_n)$
- set of n-vectors is denoted $\mathbf{R}^n$
- x_i: ith *element* or *component* or *entry* of x

special vectors (size follows from context)

- $x = 0$ (zero vector): $x_i = 0$, $i = 1, \ldots, n$
- $x = e_i$ (ith *basis vector* or ith *unit vector*): $x_i = 1$, $x_k = 0$ for $k \neq i$

Vector operations

scalar multiplication of a vector x with a scalar α

$$\alpha\, x = \begin{bmatrix} \alpha\, x_1 \\ \alpha\, x_2 \\ \vdots \\ \alpha\, x_n \end{bmatrix}$$

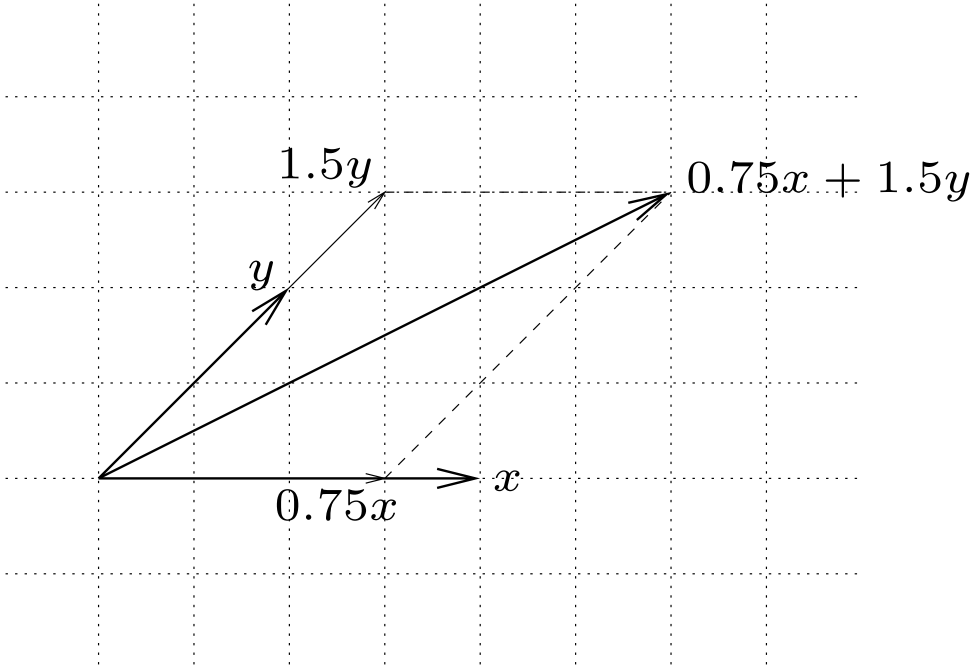
addition and subtraction of two n-vectors x, y

$$x + y = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}, \qquad x - y = \begin{bmatrix} x_1 - y_1 \\ x_2 - y_2 \\ \vdots \\ x_n - y_n \end{bmatrix}$$

Geometrical interpretation

for $n \leq 3$: x is a point with coordinates x_i

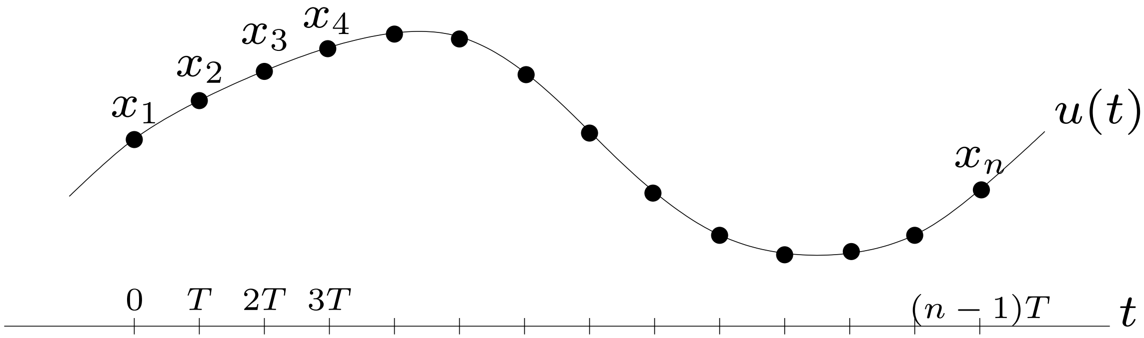
example: $x = (4, 0)$, $y = (2, 2)$



$$0.75\,x = \left[\begin{array}{c} 3 \\ 0 \end{array}\right], \qquad 1.5\,y = \left[\begin{array}{c} 3 \\ 3 \end{array}\right], \qquad 0.75\,x + 1.5\,y = \left[\begin{array}{c} 6 \\ 3 \end{array}\right]$$

Example: sampled signal

a function $u(t)$ of time, sampled with period T for $n-1$ periods

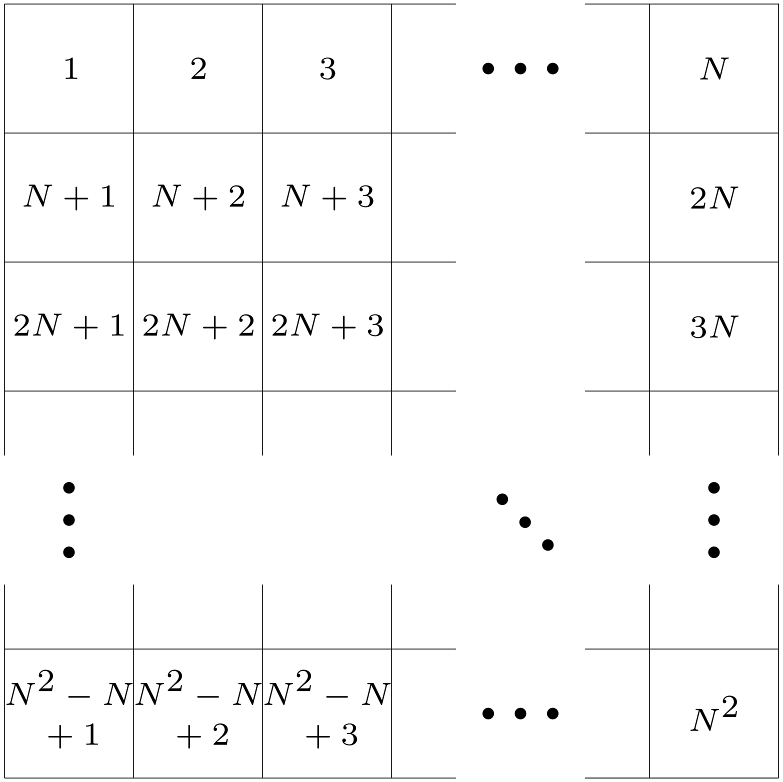


can be represented as an n-vector with

$$x_1 = u(0), \qquad x_2 = u(T), \qquad \ldots, \qquad x_n = u((n-1)T)$$

applications in signal processing, communications, control, finance, . . .

Example: image



1	2	3		$\cdots$		N
$N+1$	$N+2$	$N+3$				$2N$
$2N+1$	$2N+2$	$2N+3$				$3N$
$\vdots$				$\ddots$		$\vdots$
N^2-N $+1$	N^2-N $+2$	N^2-N $+3$		$\cdots$		N^2

- grayscale image of $N \times N$ pixels: can be represented by a vector x of size $n = N^2$; x_i is the grayscale level of pixel i
- color image: color of each pixel is given by three values (red, green, blue intensity); image can be represented by three vectors x, y, z of size N^2

Inner products

definition: the inner product of two n-vectors x, y is

$$x_1y_1 + x_2y_2 + \cdots + x_ny_n$$

notation: x^Ty

properties

- $(\alpha x)^Ty = \alpha(x^Ty)$ for scalar α
- $(x+y)^Tz = x^Tz + y^Tz$
- $x^Ty = y^Tx$

Linear functions

definition: a function $f : \mathbf{R}^n \to \mathbf{R}$ is *linear* if superposition holds, $i.e.$,

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y) \tag{1}$$

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

property: f is linear if and only if there exists a constant a such that

$$f(x) = a^T x = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n \quad \text{for all } x$$

affine function

- f satisfies (1) for all x, y, α, β with $\alpha + \beta = 1$
- f it can be written as $f(x) = a^T x + b$

Examples in $\mathbf{R}^3$

- $f(x) = (x_1 + x_2 + x_3)/3$ is linear: $f(x) = a^T x$ with

$$a = (1/3, 1/3, 1/3)$$

- $f(x) = -x_1$ is linear: $f(x) = a^T x$ with

$$a = (-1, 0, 0)$$

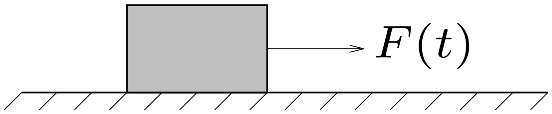
- $f(x) = \max\{x_1, x_2, x_3\}$ is not linear: superposition does not hold for

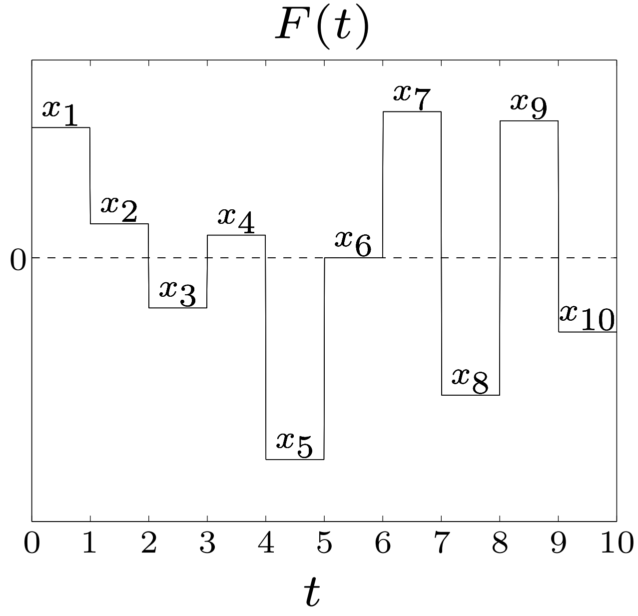
$$x = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad y = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \qquad \alpha = -1, \qquad \beta = 1$$

  we have $f(x) = 1$, $f(y) = 0$, $f(\alpha x + \beta y) = 0$, $\alpha f(x) + \beta f(y) = -1$

Mechanics example

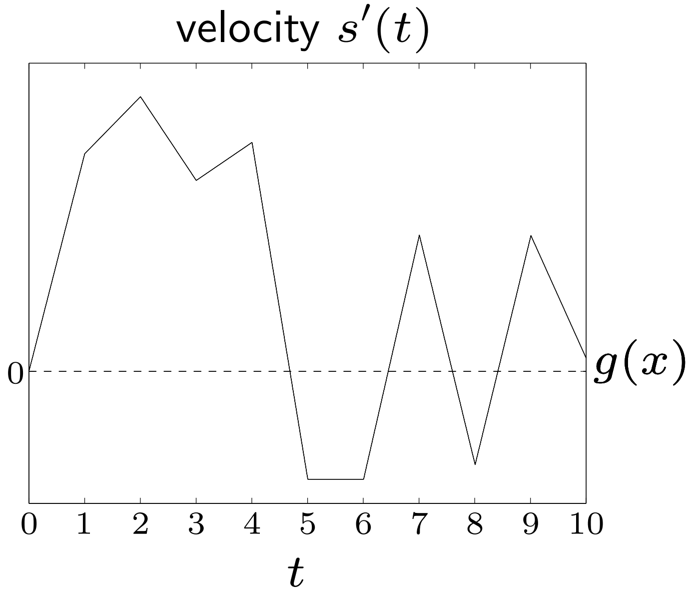
unit mass, with zero position/velocity at $t = 0$, subject to force $F(t)$

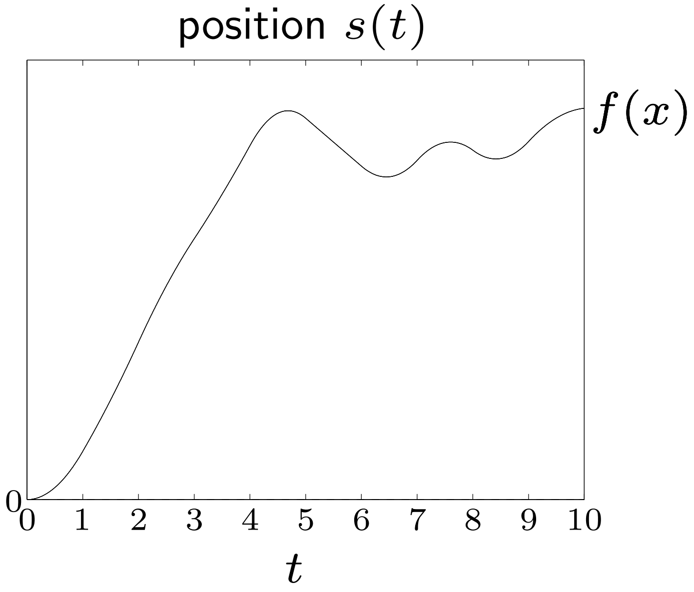




- $F(t) = x_j$ for $j - 1 \leq t < j$, $j = 1, \ldots, 10$ (x is sequence of forces)
- $f(x)$ is position of mass at $t = 10$
- $g(x)$ is velocity of mass at $t = 10$

are f and g linear functions of x?





let $s(t)$ be the position at time t; from Newton's law: $s''(t) = F(t)$

$$
\begin{aligned}
g(x) &= s'(10) = \int_0^{10} F(t)\,dt = \sum_{i=1}^{10} x_i \\
f(x) &= s(10) = \int_0^{10} s'(t)\,dt = \sum_{i=1}^{10} (\frac{1}{2} + 10 - i)\, x_i
\end{aligned}
$$

$i.e.$, the two functions are linear

Linearization

gradient of function $f : \mathbf{R}^n \to \mathbf{R}$ is

$$\nabla f(\hat{x}) = \left[\begin{array}{c} \partial f(\hat{x})/\partial x_1 \\ \partial f(\hat{x})/\partial x_2 \\ \vdots \\ \partial f(\hat{x})/\partial x_n \end{array} \right]$$

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

$$f_{\mathrm{aff}}(x) = f(\hat{x}) + \nabla f(\hat{x})^T (x - \hat{x})$$

- an affine function of x
- good approximation of $f(x)$ for x around $\hat{x}$
- generalizes Taylor approximation of function of one variable

$$f_{\mathrm{aff}}(x) = f(\hat{x}) + f'(\hat{x})(x - \hat{x})$$

Example

$$f(x_1, x_2) = e^{x_1+x_2-1} + e^{x_1-x_2-1} + e^{-x_1-1}$$

gradient

$$\nabla f(\hat{x}) = \left[\begin{array}{c} e^{\hat{x}_1+\hat{x}_2-1} + e^{\hat{x}_1-\hat{x}_2-1} - e^{-\hat{x}_1-1} \\ e^{\hat{x}_1+\hat{x}_2-1} - e^{\hat{x}_1-\hat{x}_2-1} \end{array} \right]$$

linearization around $\hat{x} = 0$

$$\begin{aligned} f_{\mathrm{aff}}(x) &= f(0) + \nabla f(0)^T (x - 0) \\ &= 3/e + x_1/e \end{aligned}$$

Euclidean norm

$$\|x\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \sqrt{x^T x}$$

properties

- also written $\|x\|_2$ to distinguish from other norms
- $\|\alpha x\| = |\alpha| \|x\|$ for scalar α
- $\|x + y\| \leq \|x\| + \|y\|$ (triangle inequality)
- $\|x\| \geq 0$ and $\|x\| = 0$ only if $x = 0$

interpretation

- $\|x\|$ measures the *magnitude* or length of x
- $\|x - y\|$ measures the *distance* between x and y

Cauchy-Schwarz inequality

$$-\|x\|\|y\| \leq x^T y \leq \|x\|\|y\|$$

- holds for all vectors x, y of the same size
- $x^T y = \|x\|\|y\|$ iff x and y are aligned (nonnegative multiples)
- $x^T y = -\|x\|\|y\|$ iff x and y are opposed (nonpositive multiples)

Exercises

1. use the Cauchy-Schwarz inequality to prove that

$$-\sqrt{n}\|x\| \leq \sum_{i=1}^{n} x_i \leq \sqrt{n}\|x\|$$

for all n-vectors x; when does equality hold?

2. use the Cauchy-Schwarz inequality to prove that

$$\frac{1}{n}\sum_{k=1}^{n} x_k \geq \left(\frac{1}{n}\sum_{k=1}^{n}\frac{1}{x_k}\right)^{-1}$$

for all n-vectors x with positive elements x_k

(l.h.s. is *arithmetic mean*; r.h.s. is *harmonic mean* of $x_1, \ldots, x_n$)

Angle between vectors

correlation coefficient of nonzero vectors x, y

$$\rho = \frac{x^T y}{\|x\| \|y\|}$$

$-1 \leq \rho \leq 1$ for all x, y

angle between x and y

$$\theta = \arccos \rho = \arccos \left(\frac{x^T y}{\|x\| \|y\|} \right)$$

we normalize θ so that $0 \leq \theta \leq \pi$

Terminology

- orthogonal vectors

$$x^T y = 0, \qquad \theta = \pi/2, \qquad \rho = 0$$

- aligned vectors

$$x^T y = \|x\|\|y\|, \qquad \theta = 0, \qquad \rho = 1$$

- anti-aligned (opposed) vectors

$$x^T y = -\|x\|\|y\|, \qquad \theta = \pi, \qquad \rho = -1$$

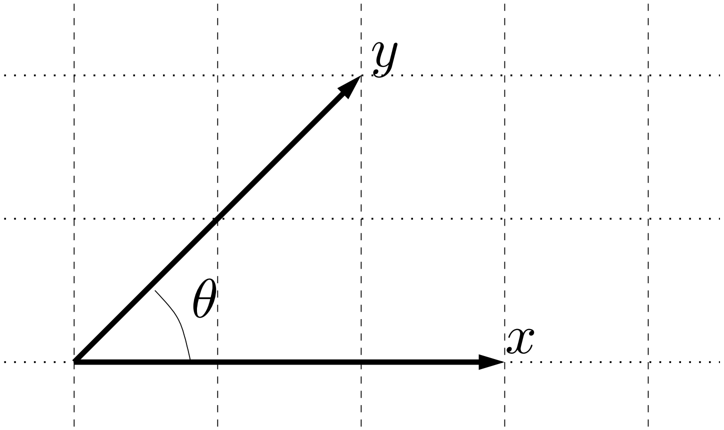
- acute angle

$$x^T y > 0, \qquad \theta < \pi/2, \qquad \rho > 0$$

- obtuse angle

$$x^T y < 0, \qquad \theta > \pi/2, \qquad \rho < 0$$

Example



$$x = (3,0), \qquad y = (2,2)$$

$$\theta = \arccos\left(\frac{x^T y}{\|x\|\|y\|}\right) = \arccos\left(\frac{6}{3\sqrt{8}}\right) = \pi/4 \text{ radians}$$

Exercise

given two n-vectors $x \neq 0$, y

$$\text{minimize (over } t\text{)} \quad \|tx - y\|$$

geometrically, tx is the projection of a vector y on the line through 0 and x