

EN530.678 Nonlinear Control and Planning in Robotics

Lecture 1: Matrix Algebra Basics

January 24, 2022

Lecturer: Marin Kobilarov

1 Mathematical Preliminaries I: Matrix Algebra

- vectors $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ and matrices $A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \cdot & \ddots & \cdot \\ \cdot & \ddots & \cdot \\ a_{n1} & \dots & a_{nn} \end{bmatrix} \in \mathbb{R}^{n \times n}$

- scalar t denotes time, we write $x(t)$ and $A(t)$ when they are function of time

- Inner products

$$x^T y \equiv x' y \equiv x \cdot y \equiv \langle x, y \rangle \equiv \sum_{i=1}^n x_i y_i$$

- Matrix *determinant* $\det(A)$ or $|A|$ is

$$\det(A) = \sum_{i=1}^n a_{ij} C_{ij},$$

for a single fixed $j \in \{1, \dots, n\}$, where C_{ij} is called the ij -th cofactor, which is the determinant of the reduced matrix obtained by crossing out the i -th row and j -th column multiplied by $(-1)^{i+j}$.

- The determinant is also the *signed volume* of the parallelepiped whose sides corresponds to the columns of the matrix

- Matrix Inverse

$$(A^{-1})_{ij} = \frac{1}{\det(A)} C_{ji}, \quad \text{for } \det(A) \neq 0$$

- Linear Independence: a set of vectors $a_1 \in \mathbb{R}^n, \dots, a_n \in \mathbb{R}^n$ are linearly independent if it is not possible to express one a linear combination of the others, i.e.

$$x_1 a_1 + \dots + x_n a_n = 0$$

implies that all scalars x_1, \dots, x_n are zero. The *rank* of a matrix is the maximum number of linearly independent columns or rows. A square n -by- n matrix with rank less than n is called *singular*.

- The solutions λ_i to the equation

$$\det(A - \lambda I) = 0,$$

where I is the identity matrix, are called the *eigenvalues* of A . If $Ax = y$ then $\lambda x = y$ and the vectors x^i corresponding to λ_i are called the *eigenvectors* of A . Combining all solutions we have

$$A [x^1 \mid \cdots \mid x^n] = [x^1 \mid \cdots \mid x^n] \text{diag}([\lambda_1, \dots, \lambda_n]) \Leftrightarrow AS = S\Lambda,$$

or

$$S^{-1}AS = \Lambda,$$

which is called *similarity transformation*, i.e. A is similar to the diagonal matrix Λ . Two similar matrices A and B satisfy $\lambda_i(A) = \lambda_i(B)$. We have the relationship

$$\text{trace}(A) = \sum_1^n a_{ii} = \sum_1^n \lambda_i(A)$$

If A is symmetric then $S^{-1} = S^T$, i.e. S is an orthogonal transformation.

- Consider the equation $Ax = y$, where $A \in \mathbb{R}^{n \times n}$. The following are equivalent:

1. $\det(A) \neq 0$
2. A^{-1} exists
3. $Ax = y$ has a unique solution for $y \neq 0$
4. A is full rank;
5. we have $\lambda_i(A) \neq 0, i = 1, \dots, n$ where $\lambda_i(A)$ is the i -th eigenvalue

- The *norm* of a vector is $\|x\|^2 = x^T x$. For $y = Ax$ for non-singular matrix A we have

$$\|y\|^2 = x^T A^T A x = \|x\|_{A^T A}^2,$$

where $\|x\|_B^2$ is called a generalized norm, i.e. a norm in new coordinates defined by B . The matrix B is *positive definite* if $\|x\|_B^2 > 0$ for all $x \neq 0$, which is written as $B > 0$. If $\|x\|_B^2 \geq 0$ for all $x \neq 0$ then B is *positive semidefinite*, i.e. $B \geq 0$.

- The *norm* of a matrix

$$\|A\| = \max_{\|x\|=1} \|Ax\|$$

- *Symmetric* matrices have real eigenvalues and mutually orthogonal, real, non-zero eigenvectors x_1, \dots, x_n . Assuming normalized $\|x_i\| = 1$ we have

$$A = \sum_{i=1}^n \lambda_i x_i x_i^T$$

Let $\lambda_1 \leq \dots \leq \lambda_n$ be the eigenvalues of symmetric matrix A , then we have

$$\|A\| = \max\{|\lambda_1|, |\lambda_n|\}, \quad \lambda_1 \|y\|^2 \leq y^T A y \leq \lambda_n \|y\|^2, \text{ for all } y \in \mathbb{R}^n$$

- Geometric Notions:

- The scalar equation $(a^i)^T x - b_i = 0$ for a given scalar b_i and vector a^i defines a *hyperplane* in \mathbb{R}^n with normal vector a^i . The intersection of n such hyperplanes is a point determined by $Ax = b$, where $(a^i)^T$ are the rows of A and $b = (b_1, \dots, b_n)$
- the equation $x^T \Sigma^{-1} x - c = 0$ determines a quadratic surface. If $\Sigma > 0$ and $\Sigma = \Sigma^T$ then this is an hyperellipsoid in \mathbb{R}^n with principal axes equal to $\sqrt{\lambda_i/c}$ where λ_i are the eigenvalues of Σ . Furthermore, since $\Sigma = S^T \Lambda S$ the axis of the ellipsoid are rotated by S . Clearly, if $\lambda_i = 0$ for some i then the hyperellipsoid is flat along that dimension and its volume (i.e. determinant) is zero. Another way to think about it is as the relationship $y^T y - c = 0$, where $y = \Lambda^{-\frac{1}{2}} S x$, which means that to construct the set of points x we can start with a set of points y on a sphere of radius c^2 , then scale that sphere by $\sqrt{\lambda_i}$ along its i -th axis for all $i = 1, \dots, n$ (to obtain an axis-aligned ellipsoid) and then rotate the axis-aligned ellipsoid by S .
- more generally, a scalar function $f(x) = 0$ defines a hypersurface in \mathbb{R}^n . Taylor expansion gives:

$$f(x) \approx f(x_0) + \left. \frac{\partial f}{\partial x} \right|_{x=x_0} (x - x_0) = 0,$$

so that the *normal* to the surface is simply the gradient. A closer approximation results from second-order expansion

$$f(x) \approx f(x_0) + \left. \frac{\partial f}{\partial x} \right|_{x=x_0} (x - x_0) + \frac{1}{2} (x - x_0)^T \left. \frac{\partial^2 f}{\partial x^2} \right|_{x=x_0} (x - x_0) = 0,$$

where $\left. \frac{\partial^2 f}{\partial x^2} \right|_{x=x_0} \equiv B$ is the n -by- n *Hessian* matrix. If $B \geq 0$ (> 0) we call the function *locally convex* (strictly locally convex) near x_0 . If it is true for all x_0 then f is convex (strictly convex).

- Derivative Notation: Let f be a function of two variables $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$. The following equivalent notations will be used

$$\frac{\partial f}{\partial x}(x, y) \equiv \partial_x f(x, y) \equiv f_x(x, y) \equiv D_1 f(x, y)$$

$$\frac{\partial f}{\partial y}(x, y) \equiv \partial_y f(x, y) \equiv f_y(x, y) \equiv D_2 f(x, y)$$

Similar notation is used for higher derivatives, e.g.

$$\frac{\partial^2 f}{\partial x^2}(x, y) \equiv \partial_x^2 f(x, y) \equiv f_{xx}(x, y) \equiv D_1^2 f(x, y).$$

We regard $\partial_x f$ as a *row vector*, i.e.

$$\partial_x f = \left[\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right]$$

The *gradient* of f denoted by $\nabla_x f$ is the column vector

$$\nabla_x f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = \partial_x f^T.$$

The notation extends when $f(x)$ is a column vector of functions, in which case $\partial_x f$ is a matrix called the Jacobian.

The differential df of a function $f(x, y)$ is

$$df = f_x \cdot dx + f_y \cdot dy,$$

where dx and dy are regarded as infinitesimal changes in x and y . In other words, df defines how f changes subject to infinitesimal changes in its parameters.