Politecnico di Milano
School of Industrial and Information Engineering
Academic Year 2024-2025

Numerical Analysis for Machine Learning

Professor: Edie Miglio

Last updated: September 30, 2024

This document is intended for educational purposes only.
These are unreviewed notes and may contain errors.

Made by Roberto Benatuil Valera

Contents

Chapter 1

Numerical Linear Algebra tools

1.1 Introduction: Recap of Linear Algebra

In this section we will review some basic concepts of Linear Algebra that will be useful for the rest of the course.

1.1.1 Matrix-vector multiplication

Given a matrix $A \in \mathbb{R}^{m\times n}$ and a vector $x \in \mathbb{R}^n$, the matrix-vector multiplication $y = Ax$ is defined as:

$$y_i = \sum_{j=1}^{n} A_{ij} x_j \tag{1.1}$$

A matrix-vector multiplication can be considered as a linear combination of the columns of the matrix A. Lets see an example:

$$\begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} x_1 + \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} x_2 \tag{1.2}$$

1.1.2 Column space of a matrix

The column space of a matrix $A \in \mathbb{R}^{m\times n}$ is the subspace of $\mathbb{R}^m$ spanned by the columns of A. In other words, it is the set of all possible linear combinations of the columns of A. The column space of a matrix is denoted as $C(A)$.

If the columns of A are linearly independent, then the column space of A is the entire $\mathbb{R}^m$. If the columns of A are linearly dependent, then the column space of A is a subspace of $\mathbb{R}^m$ with dimension equal to the rank of A.

The rank of a matrix A is is the size of the largest set of linearly independent columns of A. It is denoted as $rank(A)$. Note that $rank(A) = rank(A^T)$.

1.1.3 System of linear equations

A system of linear equations is a set of m equations with n unknowns of the form:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m \end{aligned} \tag{1.3}$$

This system can be written in matrix form as $Ax = b$, where $A \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$.

The system $Ax = b$ has a solution if and only if $b \in C(A)$. If $b \in C(A)$, then the system has a unique solution if and only if $rank(A) = n$. If $rank(A) < n$, then the system has infinitely many solutions.

1.1.4 CR factorization

The CR factorization of a matrix $A \in \mathbb{R}^{m \times n}$, with $m \geq n$, is a factorization of A as $A = CR$, where $C \in \mathbb{R}^{m \times r}$ is a matrix with the linearly independent columns of A and $R \in \mathbb{R}^{r \times n}$ is obtained by determining the coefficients of the linear combination of the columns of C that give the columns of A. In this factorization, $r = rank(A)$.

Lets see an example:

$$A = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{bmatrix} = CR \tag{1.4}$$

The matrix C is also called the Row Reduced Echelon Form of A.

1.1.5 Matrix-matrix multiplication

Given two matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$, the matrix-matrix multiplication $C = AB$ is defined as:

$$C_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj} \tag{1.5}$$

A matrix-matrix multiplication can be considered as the outer product of the columns of A and the rows of B. Lets see an example:

$$\begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} + \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} \begin{bmatrix} 3 & 4 \end{bmatrix} \tag{1.6}$$

Note that each outer product generates a matrix of the same size as the result matrix, but always with rank 1. So the matrix-matrix multiplication can be considered as a sum of rank 1 matrices, obtained by the outer products of the columns of A and the rows of B.

1.1.6 Null space of a matrix

The null space of a matrix $A \in \mathbb{R}^{m \times n}$ is the set of all vectors $x \in \mathbb{R}^n$ such that $Ax = 0$. The null space of a matrix is denoted as $N(A)$. It is also called the kernel of A, denoted as $ker(A)$.

Formally, we have that:

$$N(A) = \{x \in \mathbb{R}^n : Ax = 0\} \tag{1.7}$$

The null space of a matrix is a subspace of $\mathbb{R}^n$. The dimension of the null space of a matrix is called the nullity of the matrix.

1.1.7 Fundamental subspaces of a matrix

Given a matrix $A \in \mathbb{R}^{m \times n}$, we can define four fundamental subspaces:

- The column space of A, denoted as $C(A)$
- The row space of A, denoted as $C(A^T)$
- The null space of A, denoted as $N(A)$
- The left null space of A, denoted as $N(A^T)$

These subspaces are related by the following properties:

$$\begin{aligned} C(A) &\perp N(A^T) \\ C(A^T) &\perp N(A) \end{aligned} \tag{1.8}$$

They also satisfy the following dimensions properties:

$$\begin{aligned} dim(C(A)) + dim(N(A)) &= n \\ dim(C(A^T)) + dim(N(A^T)) &= m \end{aligned} \tag{1.9}$$

This is known as the Rank-Nullity Theorem.

1.1.8 Orthogonal matrices

An orthogonal matrix is a square matrix $Q \in \mathbb{R}^{n \times n}$ such that $Q^T Q = I$, where I is the identity matrix. This implies that $Q^T = Q^{-1}$.

Now, consider that Q is an orthogonal matrix, and set $w = Q^T x$. Then we have that:

$$\begin{aligned} \|w\|^2 = w^T w &= x^T Q Q^T x \\ &= x^T x = \|x\|^2 \end{aligned} \tag{1.10}$$

This means that the norm of a vector is preserved under an orthogonal transformation. This is called an isometry. It is a useful property for numerical algorithms, as it helps to avoid numerical instability.

There are two main types of orthogonal transformations that we are interested:

Rotation matrices

A rotation matrix is an orthogonal matrix that represents a rotation in $\mathbb{R}^2$ or $\mathbb{R}^3$. In $\mathbb{R}^2$, a rotation matrix is of the form:

$$Q(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \tag{1.11}$$

Reflection matrices

A reflection matrix is an orthogonal matrix that represents a reflection with respect to a hyperplane. If n denotes the unit normal vector to the hyperplane, then the reflection matrix is of the form:

$$Q = I - 2nn^T \tag{1.12}$$

Note that the inverse of this matrix is itself, as $Q^T = Q^{-1}$ and in this case, Q is symmetric ($Q = Q^T$).

1.1.9 QR factorization

The QR factorization of a matrix $A \in \mathbb{R}^{m \times n}$, with $m \geq n$, is a factorization of A as $A = QR$, where $Q \in \mathbb{R}^{m \times n}$ is an orthogonal matrix and $R \in \mathbb{R}^{n \times n}$ is an upper triangular matrix.

Gram-Schmidt process

The Gram-Schmidt process is a method to compute the QR factorization of a matrix. Given a matrix $A \in \mathbb{R}^{m \times n}$, the Gram-Schmidt process computes an orthonormal basis for the column space of A, as follows:

$$\begin{aligned} q_1 &= \frac{a_1}{\|a_1\|} \\ q_i &= a_i - \sum_{j=1}^{i-1} (q_j^T a_i) q_j \quad \forall i = 2, \ldots, n \end{aligned} \tag{1.13}$$

where a_i denotes the i-th column of A. The matrix Q is obtained by stacking the vectors q_i as columns. The matrix R is obtained by computing the coefficients of the linear combination of the columns of Q that give the columns of A.

1.1.10 Eigenvalues and eigenvectors

Given a square matrix $A \in \mathbb{R}^{n \times n}$, a scalar λ is called an eigenvalue of A if there exists a vector $v \in \mathbb{R}^n$ such that:

$$Av = \lambda v \tag{1.14}$$

The vector v is called an eigenvector of A associated with the eigenvalue λ.

Let P be the matrix whose columns are the eigenvectors of A, and Λ be the diagonal matrix whose diagonal elements are the eigenvalues of A. Then we have that:

$$A = P\Lambda P^{-1} \tag{1.15}$$

This is called the eigendecomposition of A.

The eigenvalues of a matrix are the roots of the characteristic polynomial of A, which is defined as:

$$det(A - \lambda I) = 0 \tag{1.16}$$

1.1.11 Similar matrices

Two square matrices A and B are called similar if there exists a non-singular matrix M such that:

$$B = M^{-1}AM \tag{1.17}$$

Similar matrices have the same eigenvalues, but not necessarily the same eigenvectors. Let (λ, y) be an eigenpair of B, then we have:

$$By = \lambda y \Rightarrow M^{-1}AMy = \lambda y \Rightarrow A(My) = \lambda(My) \tag{1.18}$$

This means that My is an eigenvector of A associated with the eigenvalue λ. So, to obtain the eigenvectors of A from the eigenvectors of B, we need to multiply the eigenvectors of B by M.

1.2 Power method

The power method is an iterative algorithm to compute the largest eigenvalue of a matrix. The algorithm is as follows:

Algorithm 1 Power method

```
Choose a random vector x^(0), s.t. ||x^(0)|| = 1
for k = 1, 2, ... do
    y^(k) = A x^(k-1)
    x^(k) = y^(k) / ||y^(k)||
    λ^(k) = x^(k)T A x^(k)
end for
```

The power method converges to the eigenvector associated with the largest eigenvalue of A. The convergence rate is determined by the ratio of the largest eigenvalue to the second largest eigenvalue.