3.VectorsMatrices

April 15, 2025

0.1 STAT 207: Vectors and matrices

Zhe Fei (zhe.fei@ucr.edu)

- NAS Chapter 6

Many scientific computational problems involve vectors and matrices.

The distinction between the set of real numbers, $\mathbb{R}$, and the set of floating-point numbers, $\mathbb{F}$, that we use in the computer has important implications for numerical computations.

We will consider the floating-point representation of elements of vectors and matrices and the computations in $\mathbb{F}$.

In multidimensional calculus, vector and matrix norms quantify notions of topology and convergence.

- Norms are devices for deriving explicit bounds, theoretical developments in numerical analysis rely heavily on norms.
- They are particularly useful in establishing convergence and in estimating rates of convergence of iterative methods for solving linear and nonlinear equations.
- Norms also arise in almost every other branch of theoretical numerical analysis. Functional analysis, which deals with infinite-dimensional vector spaces, uses norms on functions.

0.1.1 Elementary Properties of Vector Norms

Euclidean vector norm:

$$\|x\|_2 = \sqrt{\sum_{i=1}^{m} x_i^2}$$

A norm on R^m is formally defined by four properties:

(a) $\|x\| \geq 0$,

(b) $\|x\| = 0$ if and only if $x = \mathbf{0}$,

(c) $\|cx\| = |c| \cdot \|x\|$ for all real number c,

(d) $\|x + y\| \leq \|x\| + \|y\|$.

ℓ-1 norm:

$$\|x\|_1 = \sum_{i=1}^{m} |x_i|,$$

ℓ-∞ norm:

$$\|x\|_\infty = \max_i |x_i|.$$

For each of the norms $\|x\|_p, p = 1, 2$, and ∞, a sequence of vectors x_n converges to a vector y if and only if each component sequence x_{ni} converges to y_i. Thus, all three norms give the same topology on R^m.

NAS Proposition 6.2.1 Let $\|x\|$ be any norm on R^m. Then there exist positive constants k_l and k_u such that $k_l\|x\|_1 \leq \|x\| \leq k_u\|x\|_1$ holds for all $x \in R^m$.

- $\sup_{x \neq 0} \|x\|/\|x\|^*$ is finite for any pair of norms $\|x\|$ and $\|x\|^*$.
- For $p < q$ from $\{1, 2, \infty\}$,

$$\|x\|_q \leq \|x\|_p$$

$$\|x\|_p \leq m^{1/p-1/q}\|x\|_q$$

0.1.2 Elementary Properties of Matrix Norms

An $m \times m$ matrix $A = (a_{ij})$ is simply a vector in R^{m^2}.

It is preferred that a matrix norm is compatible with matrix multiplication.

In addition to (a) - (d), we require

(e) $\|AB\| \leq \|A\| \cdot \|B\|$

for any product of $m \times m$ matrices A and B.

- Frobenius norm: $\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^m a_{ij}^2} = \sqrt{tr(AA^T)} = \sqrt{tr(A^TA)}$
- Matrix norm induced from any vector norm $\|x\|$ on R^m:

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

- What is $\|I\|$ and $\|I\|_F$?

In infinite-dimensional settings, induced matrix norms are called operator norms.

NAS Proposition 6.2.1 If $A = (a_{ij})$ is an $m \times m$ matrix, then

(a) $\|A\|_1 = \max_j \sum_i |a_{ij}|$,

(b) $\|A\|_2 = \sqrt{\rho(A^TA)}$, which reduces to $\rho(A)$ if A is symmetric,

(c) $\|A\|_2 = \max_{\|u\|_2=1, \|v\|_2=1} u^TAv$,

(d) $\|A\|_\infty = \max_i \sum_j |a_{ij}|$.

0.2 Norm Preserving Linear Transformations

- Orthogonal matrix: $OO^T = I$;
- Orthonormal set of vectors S: every vector in S has magnitude 1 and the set of vectors are mutually orthogonal.

Following $(\det O)^2 = 1$, we can divide the orthogonal matrices into *rotations* with $\det O = 1$ and *reflections* with $\det O = -1$. Take 2×2 matrices for example,

- Rotation of angle θ can be presented as:

$$O = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

- Reflection of a point across the line at angle $\theta/2$ with the x axis:

$$O = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}$$

Remarks

- The set of orthogonal matrices forms a group under matrix multiplication.
- The identity matrix is the unit of the group.
- The rotations constitute a subgroup of the orthogonal group, but the reflections do not since the product of two reflections is a rotation.

Orthogonal transformations preserve inner products and Euclidean norms: $(Ou)^T Ov = u^T O^T Ov = u^T v$.

As a result, all eigenvalues of O lie on the boundary of the unit circle.

Norm invariance for vectors also leads to norm invariance for matrices.

- $\|OA\|_2^2 = \|A\|_2^2$
- $\|AO\|_2^2 = \|A\|_2^2$
- $\|OA\|_F^2 = \|A\|_F^2$
- $\|AO\|_F^2 = \|A\|_F^2$

Householder matrices Given a unit vector u, the Householder matrix $H = I - 2uu^T$ represents a reflection across the plane perpendicular to u.

H is orthogonal and symmetric,

$$HH^T = I - 4uu^T + 4u\|u\|_2^2 u^T = I.$$

- $Hv = v$ whenever v lies in the plane perpendicular to u.
- $Hu = -u$.
- H has one eigenvalue equal to -1 and all others equal to 1.

Example: Let

$$u = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Then

$$H = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - 2 \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

For any vector $v = \begin{pmatrix} x \\ y \end{pmatrix}$, the transformation is

$$Hv = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ -y \end{pmatrix}.$$

Applications:

- Matrix decomposition: QR, SVD,
- Orthogonalization:
- Stabilizing Numerical Algorithms:

0.3 Iterative Solution of Linear Equations

Many numerical problems involve iterative schemes of the form

$$x_n = Bx_{n-1} + w \tag{1}$$

for solving the vector-matrix equation $(I - B)x = w$. The map $f(x) = Bx + w$ satisfies

$$\|f(y) - f(x)\| = \|B(y - x)\| \leq \|B\| \cdot \|y - x\|$$

and is contractive for a vector norm $\|x\|$ if $\|B\| < 1$ holds for the induced matrix norm.

NAS Proposition 6.5.1 Let B be an arbitrary matrix with spectral radius $\rho(B)$ (largest absolute value of the eigenvalues). Then $\rho(B) < 1$ if and only if $\|B\| < 1$ for some induced matrix norm. The inequality $\|B\| < 1$ implies:

(a) $\lim_{n\to\infty} |B^n| = 0$,

(b) $(I - B)^{-1} = \sum_{n=0}^{\infty} B^n$,

(c) $\frac{1}{1+\|B\|} \leq \| (I - B)^{-1} \| \leq \frac{1}{1-\|B\|}$.

Linear iteration is especially useful in solving the equation $Ax = b$ for x when an approximation C to A^{-1} is known, let $B = I - CA$ and $w = Cb$ and iterate equation (1).

0.3.1 Jacobi's Iteration

- Jacobi's method offers a typical example of this strategy.
- Suppose $A = (a_{ij})$ is strictly diagonally dominant ($|a_{ii}| > \sum_{j\neq i} |a_{ij}|$), let $D = \text{diag}(a_{ii})$ be the diagonal matrix. Then $C = D^{-1}$ is an approximate inverse of A, and $B = I - CA$ satisfies $\|B\|_\infty < 1$.

[1]:
```python
import numpy as np

def jacobi(A, b, x0, tol=1e-6, max_iter=1000):
    """
    Solves the system of linear equations Ax = b using Jacobi's method.
    """
    n = len(A)
    C = np.diag(1/np.diag(A))
    B = np.identity(n) - C@A
    w = C@b

    x = B@x0 + w
    k = 1
    while max(abs(x-x0))> tol:
        x0 = x
        x = B@x0 + w
        k += 1

    if k<max_iter:
        print(f'Converged in {k} iterations.')
        return x
    else:
        print(f'Maximum iterations reached.')
        return x
```

[2]:
```python
# Example usage
A = np.array([[4, 1, 1], [2, 5, 1], [1, 2, 4]])
b = np.array([4, 1, 2])
x0 = np.array([0, 0, 0])

x = jacobi(A, b, x0)
print(f'Solution: {x}')

print(A@x)
```

```
Converged in 28 iterations.
Solution: [ 0.96874976 -0.26562527  0.39062468]
[3.99999845 0.99999787 1.99999795]
```

0.3.2 Landweber's Iteration Scheme

In practice, the approximate inverse C can be rather crude. $C = \epsilon A^T$ for ϵ small and positive.

- $A^T A$ is positive definite with eigenvalues $0 < \lambda_1 \leq ... \leq \lambda_m$.
- $I - \epsilon A^T A$ has eigenvalues $1 - \epsilon\lambda_1, ...1 - \epsilon\lambda_m$.
- Need $1 - \epsilon\lambda_m > -1$, so that $\|I - \epsilon A^T A\|_2 < 1$.

0.3.3 Equilibrium Distribution of a Markov Chain

Movement among the m states of a Markov chain is governed by its $m \times m$ transition matrix $P = (p_{ij})$, whose entries are nonnegative and satisfy $\sum_j p_{ij} = 1$ for all i.

A column vector x with nonnegative entries and norm $\|x\|_1 = \sum_i x_i = 1$ is said to be an equilibrium (stationary) distribution for P provided $x^T P = x^T$, or equivalently $Qx = x$ for $Q = P^T$.

Let $S = \{x : x_i \geq 0, i = 1, 2, ..m, \sum_i x_i = 1\}$. Assume for some power k, Q^k has all positive entries, then consider the matrix $R = Q^k - c\mathbf{1}\mathbf{1}^T$.

- The map $x \to Q^k x$ is contractive on S with unique fixed point x_∞.
- $Qx_\infty = x_\infty$.
- $\lim_{n\to\infty} Q^n x = x_\infty$ for all $x \in S$.

This power method is also used in PageRank:

- Eldén L (2007) *Matrix Methods in Data Mining and Pattern Recognition.* SIAM, Philadelphia
- Langville AN, Meyer CD (2006) *Google's PageRank and Beyond: The Science of Search Engine Rankings.* Princeton University Press, Princeton NJ

0.3.4 Condition Number of a Matrix

Consider the matrix

$$A = \begin{pmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{pmatrix}$$

that is symmetric and positive definite, and $Ax = b$.

- for $b = (32, 23, 33, 31)^T$, we find $x = (1, 1, 1, 1)^T$,
- for $b + \Delta b = (32.1, 22.9, 33.1, 30.9)^T$, the solution is violently perturbed $x + \Delta x = (9.2, -12.6, 4.5, -1.1)^T$,
- for $b = (4, 3, 3, 1)^T$, $x = (1, -1, 1, -1)^T$,
- if we perturb A to $A + 0.01I$, the solution is $x + \Delta x = (.59, -.32, .82, -.89)^T$.

How to explain these patterns? Define the condition number

$$\mathrm{cond}(A) = \|A\| \cdot \|A^{-1}\|$$

of matrix A relative to the given norm. We can have

- $$\frac{\|\Delta x\|}{\|x\|} \leq \mathrm{cond}(A)\frac{\|\Delta b\|}{\|b\|},$$
  if $Ax = b$ and $A(x + \Delta x) = b + \Delta b$.
- $$\frac{\|\Delta x\|}{\|x + \Delta x\|} \leq \mathrm{cond}(A)\frac{\|\Delta A\|}{\|A\|},$$
  if $Ax = b$ and $(A + \Delta A)(x + \Delta x) = b$.

The condition number $\text{cond}_2(A)$ relative to the matrix norm $\|A\|_2$ is the ratio of the largest and smallest eigenvalues of A, which are $\lambda_1 = 0.01015, \lambda_4 = 30.2887$. Therefore, $\text{cond}_2(A) = 2984$.

[ ]: