NE 155, Class 12 and 13, S16

Vectors and Matrices: Feb 17-19, 2016

Vector Review

A real n-dimensional vector $\vec{x}$ is an ordered set of n real numbers that expresses magnitude and direction:

$$\vec{x} = (x_1, x_2, \dots, x_n)$$

Properties:

1. sum: two vectors of the same size give a new vector of that size: $\vec{x} + \vec{y} = (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n)$
2. scalar multiple: $c\vec{x} = (cx_1, cx_2, \dots, cx_n)$
3. dot product: takes two equal length vectors and results in a scalar.

   Algebraically, it is the sum of the products of the corresponding entries of the two sequences of numbers: $\vec{x} \cdot \vec{y} = \sum_{i=1}^{n} a_i b_i = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$

   Geometrically, it is the product of the magnitudes of the two vectors and the cosine of the angle between them. $\vec{x} \cdot \vec{y} = ||\vec{x}||||\vec{y}||cos\theta$.
4. $||\vec{x}||^2 = \vec{x} \cdot \vec{x}$
5. distance from $\vec{x}$ to $\vec{y}$: $||\vec{x} - \vec{y}|| = ((x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2)^{1/2}$
6. commutative property: $\vec{x} + \vec{y} = \vec{y} + \vec{x}$
7. associative property: $(\vec{x} + \vec{y}) + \vec{z} = \vec{x} + (\vec{y} + \vec{z})$
8. distributive property: $c(\vec{x} + \vec{y}) = c\vec{x} + c\vec{y}$

Matrix Review

$$\mathbf{A} = [a_{ij}]_{m\times n} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mj} & \cdots & a_{mn} \end{pmatrix}$$

where $i = 1, \ldots, m$ is the row index and $j = 1, \ldots, n$ is the column index.

$\mathbf{A} \in \mathbb{R}^{m\times n}$ is an $m \times n$ real matrix
$\mathbf{A} \in \mathbb{C}^{m\times n}$ is an $m \times n$ complex matrix

Properties:

1. sum: $\mathbf{A} + \mathbf{B} = [a_{ij} + b_{ij}]_{m\times n}$
2. scalar multiple: $c\mathbf{A} = [ca_{ij}]_{m\times n}$
3. multiplication: $\mathbf{C} = \mathbf{AB}$;

   $\mathbf{A} \in \mathbb{C}^{m\times n}$, and $\mathbf{B} \in \mathbb{C}^{n\times p}$, and $\mathbf{C} \in \mathbb{C}^{m\times p}$, then $c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$

   $\mathbf{AB} \neq \mathbf{BA}$
4. commutative property: $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$
5. associative property: $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$
6. distributive property: $c(\mathbf{A} + \mathbf{B}) = c\mathbf{A} + c\mathbf{B}$

Definitions

Given $\mathbf{A} \in \mathbb{C}^{m\times n}$, $\mathbf{A}$ is

1. Transpose: $\mathbf{A} \in \mathbb{C}^{m\times n}$, and $\mathbf{B} \in \mathbb{C}^{n\times m} = \mathbf{A}^T$ from

$b_{ij} = a_{ji}$ for $i = 1, \ldots, n$ and $j = 1, \ldots, m$

$$\mathbf{A} = \begin{pmatrix} 1-i & 2 \\ 3+2i & 4 \\ 5 & 6+0.4i \end{pmatrix}, \qquad \mathbf{A}^T = \begin{pmatrix} 1-i & 3+2i & 5 \\ 2 & 4 & 6+0.4i \end{pmatrix}$$

2. Conjugate Transpose / adjoint, $\mathbf{A}^H = \overline{\mathbf{A}^T}$ is the complex conjugate of the transpose.

   Recall that complex conjugates are a pair of complex numbers, both the same except with imaginary parts of opposite signs. For example, The conjugate of the complex number $z = a + ib$, where a and b are real numbers, is $\overline{z} = a - ib$.

$$\mathbf{A}^H = \begin{pmatrix} 1+i & 3-2i & 5 \\ 2 & 4 & 6-i \end{pmatrix}$$

3. Inverse: $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$, where $\mathbf{I}$ is a diagonal matrix containing ones on the diagonal. If this exists, $\mathbf{A}$ is non-singular / invertible.

$$\mathbf{A} = \begin{pmatrix} 4 & 3 \\ 3 & 2 \end{pmatrix}$$

$$\mathbf{A}^{-1} = \begin{pmatrix} -2 & 3 \\ 3 & -4 \end{pmatrix}$$

4. Regular if $\mathbf{A}^{-1}$ exists

5. Hermitian / self-adjoint if $\mathbf{A} = \mathbf{A}^H$

$$\mathbf{A} = \mathbf{A}^H = \begin{pmatrix} 2 & 2+i & 4 \\ 2-i & 3 & i \\ 4 & -i & 1 \end{pmatrix}$$

6. Symmetric if $\mathbf{A} = \mathbf{A}^T$

$$\mathbf{A} = \mathbf{A}^T = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$$

7. Antisymmetric / skew-symmetric if $\mathbf{A}^T = -\mathbf{A}$ (or, skew-Hermitian if $\mathbf{A}^H = -\mathbf{A}$)

$$\mathbf{A} = \begin{pmatrix} 0 & -2 \\ 2 & 0 \end{pmatrix}, \qquad \mathbf{A}^T = \begin{pmatrix} 0 & 2 \\ -2 & 0 \end{pmatrix}$$

8. Unitary if $\mathbf{A}\mathbf{A}^H = \mathbf{I}$

$$\mathbf{A} = \frac{1}{2}\begin{pmatrix} 1+i & 1-i \\ 1-i & 1+i \end{pmatrix} ,$$

$$\mathbf{A}\mathbf{A}^H = \frac{1}{2}\begin{pmatrix} 1-i & 1+i \\ 1+i & 1-i \end{pmatrix} \times \frac{1}{2}\begin{pmatrix} 1+i & 1-i \\ 1-i & 1+i \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix} = \mathbf{I}$$

9. Normal if $\mathbf{A}\mathbf{A}^H = \mathbf{A}^H\mathbf{A}$

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} , \qquad \mathbf{A}^H = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} ,$$

$$\mathbf{A}\mathbf{A}^H = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \times \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} = \mathbf{A}^H\mathbf{A}$$

10. Orthogonal if A is real and $\mathbf{A}\mathbf{A}^T = \mathbf{A}^T\mathbf{A}$, which means $\mathbf{A}^{-1} = \mathbf{A}^T$.

$$\mathbf{A} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} , \qquad \mathbf{A}^T = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} , \qquad \mathbf{A}^T\mathbf{A} = \frac{1}{2}\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = \mathbf{I}$$

Among complex matrices, all unitary ($\mathbf{A}\mathbf{A}^H = \mathbf{I}$), Hermitian/self-adjoint ($\mathbf{A} = \mathbf{A}^H$), and skew-Hermitian ($\mathbf{A}^H = -\mathbf{A}$) matrices are normal (($\mathbf{A}\mathbf{A}^H = \mathbf{A}^H\mathbf{A}$)).

Likewise, among real matrices, all orthogonal ($\mathbf{A}\mathbf{A}^T = \mathbf{A}^T\mathbf{A}$), symmetric ($\mathbf{A} = \mathbf{A}^T$), and skew-symmetric ($\mathbf{A}^T = -\mathbf{A}$) matrices are normal.

However, it is *not* the case that all normal matrices are either unitary or (skew-)Hermitian (see example above).

Equations and Special Matrices

All of the information above is context to help us solve actual problems.

We often write systems of equations as $\mathbf{A}\vec{x} = \vec{b}$ from

$$\begin{aligned} &a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ &\vdots \\ &a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = b_n \end{aligned}$$

To find $\vec{x}$, we need to find a way to affect $\mathbf{A}^{-1}\vec{b}$. This can be done many many ways, and the first thing we'll talk about are methods for inverting the matrix.

- Tridiagonal matrix has entries on only the main, upper, and lower diagonal
- Lower triangular has entries on the diagonal and below
- Upper triangular has entries on the diagonal and above
- Block Tridagonal has blocks of elements (like sub-matrices) on the diagonal. The blocks may be full or only partially full. The blocks look like $\mathbf{D}_k = [\mathbf{D_{ij}}]_k$

The inverse of a diagonal is simply: $d_{ii}^{-1} = 1/d_{ii}$, another diagonal matrix.

Theorem: The following are equivalent (see Math 54 or a textbook for proof):

1. $\mathbf{A}$ is regular ($\mathbf{A}^{-1}$ exists)
2. Rank($\mathbf{A}$) = n
3. $\mathbf{A}\vec{x} = 0$ iff $x = 0$
4. $\mathbf{A}\vec{x} = \vec{b}$ is uniquely solveable $\forall\vec{b}$
5. det($\mathbf{A}$) $\neq 0$

Minors, Cofactors, Determinants

If the determinant is zero, then the matrix is singular, meaning we cannot invert it and numerical solutions are pretty much impossible.

Properties of Determinants

- det($\alpha\mathbf{A}$) = α^n det($\mathbf{A}$)
- det($\mathbf{A}^T$) = det($\mathbf{A}$)
- det($\mathbf{A}^{-1}$) = 1/det($\mathbf{A}$)
- det($\mathbf{AB}$) = det($\mathbf{A}$)det($\mathbf{B}$)
- det($\mathbf{A}^k$) = [det($\mathbf{A}$)]^k
- in general, det($\mathbf{A} + \mathbf{B}$) $\neq$ det($\mathbf{A}$) + det($\mathbf{B}$)

Good illustration: http://www.mathsisfun.com/algebra/matrix-inverse-minors-cofactors-adjugate.html

For a square matrix, the **first order minor**, M_{ij}, just deletes the i^{th} row and j^{th} column and takes the determinant. E.g.

$$\mathbf{A} = \begin{pmatrix} 1 & 4 & 7 \\ 3 & 0 & 5 \\ -1 & 9 & 11 \end{pmatrix} \qquad M_{23} = \text{det} \begin{pmatrix} 1 & 4 \\ -1 & 9 \end{pmatrix} = ((1 \times 9) - (4 \times -1)) = 13$$

The corresponding i, j **cofactor** of $\mathbf{A}$ is

$$C_{ij} = (-1)^{i+j} M_{ij}$$

You can compute minors and cofactors for all of $\mathbf{A}$. The $n \times n$ matrix containing all of the cofactors is denoted $\mathbf{C}$ in this context.

Using these terms, the <u>determinant</u> (which we use for lots of stuff) can be defined in terms of the Laplace expansion

$$\begin{aligned} \text{det}(\mathbf{A}) &= \sum_{j=1}^{n} a_{ij} C_{ij} \text{ for any i} \in \{1, ..., n\} \\ &= \sum_{i=1}^{n} a_{ij} C_{ij} \text{ for any j} \in \{1, ..., n\} \end{aligned}$$

For the above matrix, lets look at the Laplace expansion along the second column ($j = 2$; sum runs over i)

$$\begin{aligned}
\det(\mathbf{A}) &= (-1)^{1+2}a_{12}M_{12} + (-1)^{2+2}a_{22}M_{22} + (-1)^{3+2}a_{32}M_{32} \\
&= (-1)^{1+2} \cdot 4 \cdot \det \begin{pmatrix} 3 & 5 \\ -1 & 11 \end{pmatrix} + (-1)^{2+2} \cdot 0 \cdot \det \begin{pmatrix} 1 & 7 \\ -1 & 11 \end{pmatrix} + (-1)^{3+2} \cdot 9 \cdot \det \begin{pmatrix} 1 & 7 \\ 3 & 5 \end{pmatrix} \\
&= -4 \cdot ((3 \cdot 11) - (5 \cdot -1)) + 0 - 9 \cdot ((1 \cdot 5) - (7 \cdot 3)) = -8
\end{aligned}$$

The inverse of $\mathbf{A}$ (which we often need) can be obtained from the determinant and cofactor matrix:

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})}\mathbf{C}^T$$

The benefit is that if we've used the cofactor method to get the determinant then we get the inverse for free.

Norms and Convergence

We're going to look at direct and iterative methods to solve problems; we'll need these concepts to understand how the solution methods behave.

Vector Norms

Given $\vec{x}, \vec{y} \in \mathbb{R}^n$, a vector norm, denoted by $||\cdot||$, has the following properties:

1. $||\vec{x}|| > 0$; $||\vec{x}|| = 0$ iff $\vec{x} = 0$ (positive definite)
2. $||\vec{x} + \vec{y}|| \leq ||\vec{x}|| + ||\vec{y}||$ (triangle inequality)
3. $||\alpha\vec{x}|| = |\alpha|||\vec{x}||$ (homogeneous)

The p-norm:

$$||\vec{x}||_p \equiv (|x_1|^p + |x_2|^p + \cdots + |x_n|^p)^{1/p} \qquad p \geq 1$$

- $||\vec{x}||_1 \equiv |x_1| + |x_2| + \cdots + |x_n|$

- Euclidean norm (length) $||\vec{x}||_2 \equiv (|x_1|^2 + |x_2|^2 + \cdots + |x_n|^2)^{1/2}$
- $||\vec{x}||_\infty \equiv \max_{1 \leq i \leq n} |x_i|$

Inner Products

Inner products are symmetric and bilinear form in $\mathbb{R}^n$:

- bi-linear (depends on two arguments) $< \vec{x}, \vec{y} >$
- symmetric $< \vec{x}, \vec{y} >=< \vec{y}, \vec{x} >$
- linear $< \vec{x} + \alpha\vec{z}, \vec{y} >=< \vec{x}, \vec{y} > +\alpha < \vec{y}, \vec{z} >$
- positive $< \vec{x}, \vec{x} > > 0$ for $\vec{x} \neq 0$

The dot product is also known as the Euclidean inner product. One can define many inner products, but *in this class assume we're using the Euclidean inner product* (dot product) unless otherwise specified.

Convergence

We can use the concepts of norms and inner products to develop some relationships and talk about convergence.

The *Cauchy-Schwartz* inequality states

$$< \vec{x}, \vec{y} >\leq ||\vec{x}||_2 ||\vec{y}||_2$$

Two norms, denoted $|| \cdot ||_a$ and $|| \cdot ||_b$ here, are defined as being equivalent if there exists C_1 and C_2 such that $\forall \vec{x} \in \mathbb{R}^n$

$$C_1 ||\vec{x}||_a \leq ||\vec{x}||_b \leq C_2 ||\vec{x}||_a \qquad \text{where } C_1, C_2 = f(n)$$

This leads to the theorem that in $\mathbb{R}^n$ all norms are equivalent (offered without proof). E.g.:

$$||\vec{x}||_\infty \leq ||\vec{x}||_2 \leq \sqrt{n}||\vec{x}||_\infty$$
$$||\vec{x}||_\infty \leq ||\vec{x}||_1 \leq n||\vec{x}||_\infty$$
$$||\vec{x}||_2 \leq ||\vec{x}||_1 \leq \sqrt{n}||\vec{x}||_2$$

The inequalities above may not be 'sharp'. A sharp inequality is when there could be no better inequality when making the comparison.

E.g., when comparing two real numbers/expressions, an inequality is sharp because we could not increase the left or decrease the right by a positive factor and still have it be true ($2 \leq 2$).

Norm equivalence is very important because if we can prove some property (usually convergence or an error bound) in **some** norm, we have effectively proven it in **all** norms.

Error

For $\vec{x}$ (the real solution) and $\hat{x}$ (representing our numerical solution) $\in \mathbb{R}^n$ and some norm p we define

- Absolute error: $||\hat{x} - \vec{x}||_p$
- Relative error: $||\hat{x} - \vec{x}||_p / ||\vec{x}||_p$, where $\vec{x} \neq 0$

Convergence

Given a sequence $\{\hat{x}^{(k)}\}_{k=1,2,\ldots,\infty}$ and some norm p, we say that $\{\hat{x}^{(k)}\}$ converges to $\vec{x}$ if

$$\lim_{k\to\infty} ||\hat{x}^{(k)} - \vec{x}||_p = 0$$

Matrix Norms

Given $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times n}$, a matrix norm, denoted by $||\cdot||$, has the following properties:

1. $||\mathbf{A}|| > 0$ and $||\mathbf{A}|| = 0$ iff $\mathbf{A} = 0$ (positive definite)

2. $||\mathbf{A}+\mathbf{B}|| \leq ||\mathbf{A}|| + ||\mathbf{B}||$ (triangle inequality)

3. $||\alpha\mathbf{A}|| = |\alpha|||\mathbf{A}||$ (homogeneous)

E.g., the Frobenius norm

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

Definitions:

- submultiplicative if $||\mathbf{AB}|| \leq ||\mathbf{A}||||\mathbf{B}||$
- Not all norms are submultiplicative, we will only deal with those that are.
- subordinate matrix norm for $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\vec{x} \in \mathbb{R}^n$:

  $||\mathbf{A}|| \equiv \sup_{\vec{x}\neq\vec{0}} ||\mathbf{A}\vec{x}||/||\vec{x}||$
- **supremum**, or least upper bound, of a set S of real numbers is denoted by sup$S$ and is defined to be the smallest real number that is greater than or equal to every number in S.

  Consequently, the supremum is also referred to as the least upper bound (or LUB). If the supremum exists, it is unique, meaning that there will be only one supremum.

examples:

$$||\mathbf{A}||_1 = \sup_{\vec{x}\neq\vec{0}} \frac{||\mathbf{A}\vec{x}||_1}{||\vec{x}||_1} = \max_{1\leq j\leq n}\sum_{i=1}^{m}|a_{ij}| \qquad \text{max absolute col sum}$$

$$||\mathbf{A}||_\infty = \sup_{\vec{x}\neq\vec{0}} \frac{||\mathbf{A}\vec{x}||_\infty}{||\vec{x}||_\infty} = \max_{1\leq i\leq m}\sum_{j=1}^{n}|a_{ij}| \qquad \text{max absolute row sum}$$

$$||\mathbf{A}||_2 = \sup_{\vec{x}\neq\vec{0}} \frac{||\mathbf{A}\vec{x}||_2}{||\vec{x}||_2} = \text{ the sqrt of the max eigenvalue of } \mathbf{A}^H\mathbf{A}$$

(largest singular value for any matrix) or the spectral radius of a square matrix

For vectors, the infinity norm is the largest value. For matrices, it's the maximum absolute row sum - which is really the row that will most strongly change a vector, so that's analogous. For the 1 matrix norm, the maximum absolute column sum corresponds to the vector component that will, overall, be changed the most. We'll talk about the last after we talk about spectral radius.

The equivalence of norms we talked about with vectors also holds here. Some of the relationships are

$$
\begin{aligned}
||\mathbf{A}||_2 \leq ||\mathbf{A}||_F \leq \sqrt{n}||\mathbf{A}||_2 \\
\frac{1}{\sqrt{n}}||\mathbf{A}||_\infty \leq ||\mathbf{A}||_2 \leq \sqrt{m}||\mathbf{A}||_\infty \\
\frac{1}{\sqrt{m}}||\mathbf{A}||_1 \leq ||\mathbf{A}||_2 \leq \sqrt{n}||\mathbf{A}||_2
\end{aligned}
$$

Again, as with vector norms, showing a property in some matrix norms implies that property in other matrix norms (which may be harder to compute). The inequalities above may not be 'sharp'.

All matrix norms are bounded

For $\lambda \in \sigma(\mathbf{A})$

$$|\lambda| \leq ||\mathbf{A}||$$

$$\rho(\mathbf{A}) \leq ||\mathbf{A}||$$

proof:

1. The eigenvalues of $\mathbf{A}$ satisfy $\mathbf{A}\vec{x} = \lambda x$, $\vec{x} \neq \vec{0}$.
2. Let the eigenvectors of $\mathbf{A}$ be $\vec{v} = [\vec{v_1}, \vec{v_2}, \ldots, \vec{v_n}]$
3. then

$$
\begin{aligned}
|\lambda|||\vec{v}|| &= ||\lambda\vec{v}|| && \text{property of norms} \\
&= ||\mathbf{A}\vec{v}|| && \text{substitution} \\
&\leq ||\mathbf{A}||||\vec{v}|| && \text{triangle inequality} \\
\therefore |\lambda| &\leq ||\mathbf{A}|| && \text{cancellation} \\
|\lambda| \leq \rho(\mathbf{A}) &\leq ||\mathbf{A}|| && \text{definition of } \rho \text{ as max } |\lambda|
\end{aligned}
$$

Note, we can use any submultiplicative matrix norm to bound from above the spectral radius (this will be important).