W20 CSE 203B: Linear Algebra Review

Xinyuan Wang
01/10/2020
Contents

• Linear System and Matrix Form
• Matrix Subspaces
• Matrix Operations and Properties
• Eigenvalues and Eigenvectors
• Functions and Properties
Contents

• Linear System and Matrix Form
  • Solving linear equations
Matrix and Vector

• A vector $x \in \mathbb{R}^n$

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

• A matrix $A \in \mathbb{R}^{m \times n}$

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

• How to solve a linear system $Ax = b$ \implies

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$
Linear Systems

• Solve $Ax = b$, equivalent to find the $n$ unknown variables $x_1, \ldots, x_n$ satisfying equations

\[ a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\
\vdots \\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \]

• The LHS defines a linear combination of $A$’s column vectors

\[ x_1c_1(A) + x_2c_2(A) + \cdots + x_nc_n(A) = b \]

where $c_i(A)$ is the $i^{th}$ column of $A$.

• When does the system have a solution?
Linear Systems

• A necessary and sufficient condition that $Ax = b$ has a solution is that $b \in S(c_1(A), \ldots, c_n(A))$, which is a space in $\mathbb{R}^m$ spanned by the columns of $A$.

• Gaussian elimination with row operations
  
  • The solution won’t be changed

\[
\begin{bmatrix}
a_{11} & \cdots & a_{1n} \\
a_{m1} & \cdots & a_{mn}
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
\vdots \\
x_n
\end{bmatrix}
= b_1
\]

\[
\begin{bmatrix}
x_1 \\
x_2 \\
\vdots \\
x_n
\end{bmatrix}
= b_m
\]
Contents

• Matrix Subspaces
  • Linear independence and rank
  • Range and nullspace of a matrix
Range of a matrix

• The range of matrix $A \in \mathbb{R}^{m \times n}$ is the set of vectors that can be expressed as a linear combination of the column vectors of $A$
  $$\mathcal{R}(A) = \{y : y = Ax \mid x \in \mathbb{R}^n\} \in \mathbb{R}^m$$

• The columns of $A$ are linearly independent if no column is in the range of the remaining columns
  $$a_i \notin \mathcal{R}(A_{-i}), \forall i = 1, \ldots, n$$

• Row subspace $\text{row}(A) = \{A^T y \mid y \in \mathbb{R}^m\} = \mathcal{R}(A^T)$
Nullspace of a matrix

• The **nullspace** of \( A \in \mathbb{R}^{m \times n} \) is the set of all vector \( x \) s.t. \( Ax = 0 \)

\[
\mathcal{N}(A) = \{ x \mid Ax = 0 \} \in \mathbb{R}^n
\]

• **Properties:**
  • Disjoint, orthogonal complement \( \mathcal{N}(A) \perp \mathcal{R}(A^T) \)
  • Span the entire space \( \mathbb{R}^n \)
    \[
    \dim(\mathcal{N}(A)) + \dim(\text{row}(A)) = n
    \]

**Excises:** what is the range of the matrix below? Nullspace?

1) \( A = \begin{pmatrix}
1 & 2 & 0 \\
0 & 1 & 0 \\
0 & 0 & -1
\end{pmatrix} \)  
2) \( A = \begin{pmatrix}
2 & 1 & -1 \\
1 & -1 & 0 \\
-1 & 1 & 0
\end{pmatrix} \)
Orthogonality

• Two vectors $x, y \in \mathbb{R}^n$ are orthogonal if
  $$ x^T y = 0 $$

• They are orthonormal if
  $$ \|x\|_2 = \|y\|_2 = 1 $$

• A matrix $U \in \mathbb{R}^{n \times n}$ is orthogonal if all its columns are orthonormal
  $$ U^T U = I = UU^T $$

• Columns of an orthogonal matrix are linearly independent
Rank

• Definition: the rank of any matrix $A \in \mathbb{R}^{m \times n}$, denoted by $r(A)$, is the number of independent columns.

• Properties
  • $r(A) \leq \min(m, n)$
  • $r(A) = r(A^T)$
  • $r(AB) \leq \min(r(A), r(B))$
  • $r(A + B) \leq r(A) + r(B)$
  • $A$ has full rank if $r(A) = \min(m, n)$
  • If $r(A) < \min(m, n)$, then the rows/columns of $A$ are not linearly independent
Matrix Inverse

- A square matrix $A \in \mathbb{R}^{n \times n}$ is invertible if $r(A) = n$.
  - There exists $B \in \mathbb{R}^{n \times n}$ so that $AB = I$ and $B = A^{-1}$ is unique.
  - $A$ if full rank and $\dim(\mathcal{N}(A)) = 0$.

- Solve a linear system $Ax = b$
  - Has a solution for any $b \in \mathbb{R}^n$ (consistent).
  - The solution $x = A^{-1}b$ is unique.
  - If $b = 0$, then $x = 0$.

Excises: does the linear system have unique solution?

a) \[
\begin{pmatrix}
2 & -1 \\
1 & 0
\end{pmatrix} x = \begin{pmatrix} 1 \\ 0 \end{pmatrix}
\]
Yes, $x = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$

b) \[
\begin{pmatrix}
1 & -1 \\
-1 & 1
\end{pmatrix} x = \begin{pmatrix} 1 \\ -1 \end{pmatrix}
\]
No, $x = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \mathcal{N} \left( \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \right)$ any nonzero solution
Contents

• Matrix Operations and Properties
  • Trace of a matrix
  • Norm
  • Determinant of a matrix
Trace

• The trace of a square matrix $A \in \mathbb{R}^{n \times n}$, denoted $\text{tr}A$, is the sum of the diagonal elements in the matrix

$$\text{tr}A = \sum_{i=1}^{n} A_{ii}$$

• Properties
  • For $A \in \mathbb{R}^{n \times n}$, $\text{tr}A = \text{tr}A^T$.
  • For $A, B \in \mathbb{R}^{n \times n}$, $\text{tr}(A + B) = \text{tr}A + \text{tr}B$.
  • For $A \in \mathbb{R}^{n \times n}, t \in \mathbb{R}$, $\text{tr}(tA) = t \text{tr}A$.
  • For $A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{n \times m}$, $\text{tr}AB = \text{tr}BA$. 
Norms

• Given a vector \( x \in \mathbb{R}^n \), its norm is given by
  • Euclidean norm (\( L_2 \)-norm): \( \|x\|_2 = \langle x, x \rangle = (x_1^2 + \cdots + x_n^2)^{1/2} \)
  • Sum-absolute-value (\( L_1 \)-norm): \( \|x\|_1 = |x_1| + \cdots + |x_n| \)
  • Chebyshev (\( L_\infty \)-norm): \( \|x\|_\infty = \max_i |x_i| \) for \( i = 1, \ldots, n \)

• The vector norm satisfies
  • \( \|cx\| = |c| \|x\| \)
  • \( \|x\| = 0 \iff x = 0 \)
  • \( \|x + y\| \leq \|x\| + \|y\| \) (triangle inequality)

\[ x = (1,1) \quad x + y = (3,1) \]
\[ y = (2,0) \]
Determinants

- The determinant of a square matrix $A \in \mathbb{R}^{n \times n}$ is a function $\text{det}: \mathbb{R}^{n \times n} \to \mathbb{R}$, denoted $|A|$ or $\det A$. Consider the set of all linear combinations of the row vectors of $A$ with coefficient in $[0,1]$
  $$S = \{v \in \mathbb{R}^n : v = \sum_{i=1}^{n} \alpha_i a_i \text{ where } 0 \leq \alpha_i \leq 1, i = 1, \ldots, n\}.$$ 

- $|\det A|$ is the area of the n-dimensional parallelootope.
Compute Determinants

• Let \( A \in \mathbb{R}^{n \times n} \), for any row \( i \) and column \( j \) of \( A \) define the \( ij - minor \)

\[
M_{ij} = \text{det} A \left|_{i^{th} \text{ row}, j^{th} \text{ column removed}} \right.
\]

1) For any fixed row \( k \)

\[
\text{det} A = \sum_{j=1}^{n} a_{kj} (-1)^{k+j} M_{kj}
\]

2) For any fixed column \( m \)

\[
\text{det} A = \sum_{j=1}^{n} a_{jm} (-1)^{m+j} M_{jm}
\]

Excises: compute the determinant of

\[
\begin{pmatrix}
3 & 2 & -1 \\
0 & 1 & 3 \\
1 & 2 & -1
\end{pmatrix}
\]
Determinants

• Let $A \in \mathbb{R}^{n \times n}$, properties of determinant
  • If $r(A) < n$, then $\det A = 0$
  • If $A$ is triangular, then $\det A = \prod_{i=1}^{n} a_{ii}$
  • If $r(A) = n$, then $\det A \neq 0$

• If $A, B \in \mathbb{R}^{n \times n}$, then $\det AB = \det A \det B$.

Excises: calculate the determinants of $A, B, AB$

$$A = \begin{pmatrix} 2 & 4 \\ 2 & 1 \end{pmatrix}, B = \begin{pmatrix} 3 & -1 \\ 1 & -2 \end{pmatrix}$$

• Define the adjugate (adjoint) $\hat{A}_{ij} = (-1)^{i+j}M_{ji}$, if $A$ is invertible then $A^{-1} = \frac{1}{\det A} \hat{A}$

Excises: find the adjugate and inverse of

$$\begin{pmatrix} 2 & 4 \\ 2 & 1 \end{pmatrix}, \quad \hat{A} = \begin{pmatrix} 1 & -4 \\ -2 & 2 \end{pmatrix}, \det A = -6$$
Contents

• Eigenvalues and Eigenvectors
  • Properties
  • Decompositions
Eigenvalues

• Let $A \in \mathbb{R}^{n \times n}$, $Ax = \lambda x$ leads to $\det(A - \lambda I) = 0$.

• If $A$ is symmetric, then it could be decomposed into

$$A = Q \Lambda Q^T = \sum_{i=1}^{n} \lambda_i q_i q_i^T,$$

where $Q \in \mathbb{R}^{n \times n}$ is orthogonal which satisfies $Q^T Q = I$, and $\Lambda = \text{diag}(\lambda_1, ..., \lambda_n)$. Each vector $q_i$ is eigenvector corresponds to eigenvalue $\lambda_i$.

• If $A$ is not symmetric but diagonalizable, then

$$A = Q \Lambda Q^{-1}$$

where $Q$ is not necessarily orthogonal.

• For general $A \in \mathbb{R}^{m \times n}$, singular value decomposition $A = U \Lambda V^T$
Properties of Eigenvalues

• Let $A \in \mathbb{R}^{n \times n}$, the properties of its eigenvalues
  \[ \det A = \prod_{i=1}^{n} \lambda_i, \quad \text{tr} \ A = \sum_{i=1}^{n} \lambda_i. \]

• The rank of $A$ is equal to the number of non-zero eigenvalues of $A$.

• If $A$ is non-singular then $1/\lambda_i$ is an eigenvalue of $A^{-1}$ with associated vector $x_i$.

• The largest and smallest eigenvalues satisfy
  \[ \lambda_{\text{max}}(A) = \sup_{x \neq 0} \frac{x^T Ax}{x^T x}, \quad \lambda_{\text{min}}(A) = \inf_{x \neq 0} \frac{x^T Ax}{x^T x}. \]
Singular Value Decomposition

• Let $A \in \mathbb{R}^{m \times n}$, $Ax = \lambda x$ leads to $\det(A - \lambda I) = 0$. Then it could be factorized as

$$A = U\Sigma V^*$$

where $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are unitary matrices which are orthogonal if $A$ is real. $\Sigma$ is a diagonal matrix with non-negative real diagonal, called singular values.

• The columns of $V$ are eigenvectors of $A^*A$

$$A^*A = V\Sigma(U^*U)\Sigma V^* = V\Sigma^2V^*$$

• The columns of $U$ are eigenvectors of $AA^*$

$$AA^* = U\Sigma(V^*V)\Sigma U^* = U\Sigma^2U^*$$

• The non-zero elements of $\Sigma$ are the square roots of the non-zero eigenvalues of $A^*A$ and $AA^*$

Related to eigenvalue decomposition, can we figure out the range and null space of a matrix from its SVD results?
Contents

• Functions and Properties
Functions and Related Properties

• For \( f : A \to B \) we say that \( f \) is a function on the set \( \text{dom} \ f \subseteq A \) into the set \( B \).
  • \( f : R^n \to R^m \) means that the function maps some n-vectors into m-vectors; the \( \text{dom} \ f \) is a subset of \( R^n \) but not necessarily covers the whole space.

• Continuity
  • A function \( f : R^n \to R^m \) is continuous at \( x \in \text{dom} \ f \) if for all \( \epsilon > 0 \) there exists a \( \delta \) such that
    \[
    y \in \text{dom} \ f, \quad ||y - x||_2 \leq \delta \Rightarrow ||f(y) - f(x)||_2 \leq \epsilon
    \]

• Closed function
  • A function \( f : R^n \to R \) is closed if for each \( \alpha \in R \), the sublevel set
    \[
    \{x \in \text{dom} \ f \mid f(x) \leq \alpha\}
    \]
    is closed.
  • Open and closed set: a set \( C \) is open if \( C = \text{int} C \); a set \( C \in R^n \) is closed if its complement \( R^n \setminus C \) is open.
Derivative and Gradient

- Suppose \( f: \mathbb{R}^n \rightarrow \mathbb{R}^m \) and \( x \in \text{int} \, \text{dom} \, f \). The function \( f \) is differentiable at \( x \) if there exists a matrix \( Df(x) \in \mathbb{R}^{m \times n} \) that satisfies

\[
\lim_{z \in \text{dom} \, f, z \neq x, z \rightarrow x} \frac{||f(z) - f(x) - Df(x)(z - x)||}{||z - x||_2} = 0.
\]

- First-order approximation

We call \( Df(x) \) the derivative (Jacobian) of \( f \) at \( x \) where

\[
Df(x)_{ij} = \frac{\partial f_i(x)}{\partial x_j}, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n.
\]

- The function \( f \) is differentiable if \( \text{dom} \, f \) is open, and differentiable at \( \forall x \in \text{dom} \, f \).

- Gradient: transpose of derivative (row -> column)

\[
\nabla f(x) = Df(x)^T
\]

first-order approximation \( \hat{f}(x) = f(x) + \nabla f(x)^T(z - x) \)
Derivative and Gradient

- Suppose $f: \mathbb{R}^n \rightarrow \mathbb{R}$ and second-order differentiable

$$
\text{Gradient: } \nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}, \quad \text{Hessian } \nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}
$$

second-order approximation

$$
\hat{f}(x) = f(x) + \nabla f(x)^T (z - x) + \frac{1}{2} (z - x)^T \nabla^2 f(x) (z - x)
$$

Example: calculate the gradient and Hessian of a quadratic function $f(x) = \frac{1}{2} x^T Px + q^T x + r$, where $P \in S^n, q \in \mathbb{R}^n, r \in \mathbb{R}$.

- derivative $Df(x) = x^T P + q^T$, gradient $\nabla f(x) = Px + q$
- Hessian $\nabla^2 f(x) = D\nabla f(x) = P$
Chain Rule

- Suppose \( f: \mathbb{R}^n \rightarrow \mathbb{R}^m \) is differentiable at \( x \in \text{int dom } f \) and \( g: \mathbb{R}^m \rightarrow \mathbb{R}^p \) is differentiable at \( f(x) \in \text{int dom } g \). Define the composition \( h: \mathbb{R}^n \rightarrow \mathbb{R}^p \) \( h(x) = g(f(x)) \). Then \( h \) is differentiable at \( x \)

\[
Dh(x) = Dg(f(x))Df(x)
\]

Example: prove the convexity of \( g(x) = f(Ax + b) \) with \( \text{dom } g = \{x|Ax + b \in \text{dom } f\} \). It is a composition with affine function discussed in class.

By the chain rule: \( Dg(x) = Df(Ax + b)A \)

the gradient \( \nabla g(x) = A^T \nabla f(Ax + b) \)

the Hessian \( \nabla^2 g(x) = D\nabla g(x) = A^T \nabla^2 f(Ax + b)A \)
• Read “Convex Optimization” Appendix A by Stephen Boyd for more relevant mathematical background.