Matrix Multiplication and Inverse

Pulkit Tandon

ENGR108
Stanford University

slides adopted from Prof. Stephen Boyd

July 21, 2022
Matrix multiplication
Matrix multiplication

Composition of linear functions

Matrix powers

QR factorization
can multiply $m \times p$ matrix $A$ and $p \times n$ matrix $B$ to get $C = AB$:

$$C_{ij} = \sum_{k=1}^{p} A_{ik}B_{kj} = A_{i1}B_{1j} + \cdots + A_{ip}B_{pj}$$

for $i = 1, \ldots, m$, $j = 1, \ldots, n$

- to get $C_{ij}$: move along $i$th row of $A$, $j$th column of $B$

- example:
Special cases of matrix multiplication

- scalar-vector product (with scalar on right!) $x\alpha$
- inner product $a^T b$
- matrix-vector multiplication $Ax$
- outer product of $m$-vector $a$ and $n$-vector $b$

$$ab^T = \begin{bmatrix}
a_1b_1 & a_1b_2 & \cdots & a_1b_n \\
a_2b_1 & a_2b_2 & \cdots & a_2b_n \\
& \vdots & \ddots & \vdots \\
a_mb_1 & a_mb_2 & \cdots & a_mb_n
\end{bmatrix}$$
Properties

- \((AB)C = A(BC)\), so both can be written \(ABC\)
- \(A(B + C) = AB + AC\)
- \((AB)^T = B^T A^T\)
- \(AI = A\) and \(IA = A\)
- \(AB = BA\) does not hold in general
block matrices can be multiplied using the same formula, e.g.,

\[
\begin{bmatrix}
A & B \\
C & D
\end{bmatrix}
\begin{bmatrix}
E & F \\
G & H
\end{bmatrix}
= \begin{bmatrix}
AE + BG & AF + BH \\
CE + DG & CF + DH
\end{bmatrix}
\]

(provided the products all make sense)
Column interpretation

- denote columns of $B$ by $b_i$:

$$B = \begin{bmatrix} b_1 & b_2 & \cdots & b_n \end{bmatrix}$$

- then we have

$$AB = A \begin{bmatrix} b_1 & b_2 & \cdots & b_n \end{bmatrix} = \begin{bmatrix} Ab_1 & Ab_2 & \cdots & Ab_n \end{bmatrix}$$

- so $AB$ is ‘batch’ multiply of $A$ times columns of $B$
Multiple sets of linear equations

- given $k$ systems of linear equations, with same $m \times n$ coefficient matrix
  \[ Ax_i = b_i, \quad i = 1, \ldots, k \]
- write in compact matrix form as $AX = B$
- $X = [x_1 \cdots x_k]$, $B = [b_1 \cdots b_k]$
Inner product interpretation

- with \( a_i^T \) the rows of \( A \), \( b_j \) the columns of \( B \), we have

\[
AB = \begin{bmatrix}
    a_1^T b_1 & a_1^T b_2 & \cdots & a_1^T b_n \\
    a_2^T b_1 & a_2^T b_2 & \cdots & a_2^T b_n \\
    \vdots & \vdots & & \vdots \\
    a_m^T b_1 & a_m^T b_2 & \cdots & a_m^T b_n
\end{bmatrix}
\]

- so matrix product is all inner products of rows of \( A \) and columns of \( B \), arranged in a matrix
Gram matrix

- let $A$ be an $m \times n$ matrix with columns $a_1, \ldots, a_n$
- the Gram matrix of $A$ is
  \[
  G = A^T A = \begin{bmatrix}
  a_1^T a_1 & a_1^T a_2 & \cdots & a_1^T a_n \\
  a_2^T a_1 & a_2^T a_2 & \cdots & a_2^T a_n \\
  \vdots & \vdots & \ddots & \vdots \\
  a_n^T a_1 & a_n^T a_2 & \cdots & a_n^T a_n
  \end{bmatrix}
  \]
- Gram matrix gives all inner products of columns of $A$
- example: $G = A^T A = I$ means columns of $A$ are orthonormal
 Complexity

- to compute $C_{ij} = (AB)_{ij}$ is inner product of $p$-vectors
- so total required flops is $(mn)(2p) = 2mnp$ flops
- multiplying two $1000 \times 1000$ matrices requires 2 billion flops
- ... and can be done in well under a second on current computers
Outline

Matrix multiplication

Composition of linear functions

Matrix powers

QR factorization
Composition of linear functions

- $A$ is an $m \times p$ matrix, $B$ is $p \times n$
- define $f : \mathbb{R}^p \to \mathbb{R}^m$ and $g : \mathbb{R}^n \to \mathbb{R}^p$ as
  \[ f(u) = Au, \quad g(v) = Bv \]
- $f$ and $g$ are linear functions
- composition of $f$ and $g$ is $h : \mathbb{R}^n \to \mathbb{R}^m$ with $h(x) = f(g(x))$
- we have
  \[ h(x) = f(g(x)) = A(Bx) = (AB)x \]
- composition of linear functions is linear
- associated matrix is product of matrices of the functions
Second difference matrix

- $D_n$ is $(n - 1) \times n$ difference matrix:
  \[ D_n x = (x_2 - x_1, \ldots, x_n - x_{n-1}) \]

- $D_{n-1}$ is $(n - 2) \times (n - 1)$ difference matrix:
  \[ D_{n-1} y = (y_2 - y_1, \ldots, y_{n-1} - y_{n-2}) \]

- $\Delta = D_{n-1} D_n$ is $(n - 2) \times n$ second difference matrix:
  \[ \Delta x = (x_1 - 2x_2 + x_3, x_2 - 2x_3 + x_4, \ldots, x_{n-2} - 2x_{n-1} + x_n) \]

- for $n = 5$, $\Delta = D_{n-1} D_n$ is
  \[
  \begin{bmatrix}
  1 & -2 & 1 & 0 & 0 \\
  0 & 1 & -2 & 1 & 0 \\
  0 & 0 & 1 & -2 & 1
  \end{bmatrix}
  =
  \begin{bmatrix}
  -1 & 1 & 0 & 0 \\
  0 & -1 & 1 & 0 \\
  0 & 0 & -1 & 1
  \end{bmatrix}
  \begin{bmatrix}
  -1 & 1 & 0 & 0 & 0 \\
  0 & -1 & 1 & 0 & 0 \\
  0 & 0 & -1 & 1 & 0 \\
  0 & 0 & 0 & -1 & 1
  \end{bmatrix}
  \]
Outline

Matrix multiplication

Composition of linear functions

Matrix powers

QR factorization
Matrix powers

- for $A$ square, $A^2$ means $AA$, and same for higher powers
- with convention $A^0 = I$ we have $A^kA^l = A^{k+l}$
- negative powers later; fractional powers in other courses
Directed graph

- $n \times n$ matrix $A$ is adjacency matrix of directed graph:

$$A_{ij} = \begin{cases} 1 & \text{there is an edge from vertex } j \text{ to vertex } i \\ 0 & \text{otherwise} \end{cases}$$

- example:

![Directed graph diagram]

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$
Paths in directed graph

▶ square of adjacency matrix:

\[(A^2)_{ij} = \sum_{k=1}^{n} A_{ik}A_{kj}\]

▶ \((A^2)_{ij}\) is number of paths of length 2 from \(j\) to \(i\)

▶ for the example,

\[
A^2 = \begin{bmatrix}
1 & 0 & 1 & 1 & 0 \\
0 & 1 & 1 & 1 & 2 \\
1 & 0 & 1 & 2 & 1 \\
0 & 1 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0
\end{bmatrix}
\]

e.g., there are two paths from 4 to 3 (via 3 and 5)

▶ more generally, \((A^\ell)_{ij}\) = number of paths of length \(\ell\) from \(j\) to \(i\)
Outline

Matrix multiplication

Composition of linear functions

Matrix powers

QR factorization
Gram–Schmidt in matrix notation

- run Gram–Schmidt on columns $a_1, \ldots, a_k$ of $n \times k$ matrix $A$
- if columns are linearly independent, get orthonormal $q_1, \ldots, q_k$
- define $n \times k$ matrix $Q$ with columns $q_1, \ldots, q_k$
- $Q^T Q = I$
- from Gram–Schmidt algorithm
  
  $$a_i = (q_1^T a_i)q_1 + \cdots + (q_{i-1}^T a_i)q_{i-1} + \|\tilde{q}_i\|q_i$$
  $$= R_{1i}q_1 + \cdots + R_{ii}q_i$$

  with $R_{ij} = q_i^T a_j$ for $i < j$ and $R_{ii} = \|\tilde{q}_i\|$ 

- defining $R_{ij} = 0$ for $i > j$ we have $A = QR$
- $R$ is upper triangular, with positive diagonal entries

QR factorization
QR factorization

- $A = QR$ is called *QR factorization* of $A$
- factors satisfy $Q^T Q = I$, $R$ upper triangular with positive diagonal entries
- can be computed using Gram–Schmidt algorithm (or some variations)
- has a *huge* number of uses, which we’ll see soon
Matrix inverses
Outline

Left and right inverses

Inverse

Solving linear equations

Examples

Pseudo-inverse
Left inverses

- A number $x$ that satisfies $xa = 1$ is called the inverse of $a$.
- The inverse ($i.e.$, $1/a$) exists if and only if $a \neq 0$, and is unique.
- A matrix $X$ that satisfies $XA = I$ is called a left inverse of $A$.
- If a left inverse exists we say that $A$ is left-invertible.
- Example: the matrix

$$A = \begin{bmatrix}
-3 & -4 \\
4 & 6 \\
1 & 1
\end{bmatrix}$$

has two different left inverses:

$$B = \frac{1}{9} \begin{bmatrix}
-11 & -10 & 16 \\
7 & 8 & -11
\end{bmatrix}, \quad C = \frac{1}{2} \begin{bmatrix}
0 & -1 & 6 \\
0 & 1 & -4
\end{bmatrix}$$
Left inverse and column independence

- if $A$ has a left inverse $C$ then the columns of $A$ are linearly independent

- to see this: if $Ax = 0$ and $CA = I$ then

$$0 = C0 = C(Ax) = (CA)x = Ix = x$$

- we’ll see later the converse is also true, so

  a matrix is left-invertible if and only if its columns are linearly independent

- matrix generalization of

  a number is invertible if and only if it is nonzero

- so left-invertible matrices are tall or square
Solving linear equations with a left inverse

▶ suppose $Ax = b$, and $A$ has a left inverse $C$
▶ then $Cb = C(Ax) = (CA)x =Ix = x$
▶ so multiplying the right-hand side by a left inverse yields the solution
Example

\[
A = \begin{bmatrix}
-3 & -4 \\
4 & 6 \\
1 & 1
\end{bmatrix}, \quad b = \begin{bmatrix}
1 \\
-2 \\
0
\end{bmatrix}
\]

▶ over-determined equations \( Ax = b \) have (unique) solution \( x = (1, -1) \)

▶ \( A \) has two different left inverses, 

\[
B = \frac{1}{9} \begin{bmatrix}
-11 & -10 & 16 \\
7 & 8 & -11
\end{bmatrix}, \quad C = \frac{1}{2} \begin{bmatrix}
0 & -1 & 6 \\
0 & 1 & -4
\end{bmatrix}
\]

▶ multiplying the right-hand side with the left inverse \( B \) we get 

\[
Bb = \begin{bmatrix}
1 \\
-1
\end{bmatrix}
\]

▶ and also 

\[
Cb = \begin{bmatrix}
1 \\
-1
\end{bmatrix}
\]

Left and right inverses
Right inverses

- a matrix \( X \) that satisfies \( AX = I \) is a right inverse of \( A \)
- if a right inverse exists we say that \( A \) is right-invertible
- \( A \) is right-invertible if and only if \( A^T \) is left-invertible:
  \[
  AX = I \iff (AX)^T = I \iff X^T A^T = I
  \]
- so we conclude
  \[A \text{ is right-invertible if and only if its rows are linearly independent}\]
- right-invertible matrices are wide or square
Solving linear equations with a right inverse

• suppose $A$ has a right inverse $B$

• consider the (square or underdetermined) equations $Ax = b$

• $x = Bb$ is a solution:

\[
Ax = A(Bb) = (AB)b = Ib = b
\]

• so $Ax = b$ has a solution for any $b$
Example

- same $A$, $B$, $C$ in example above
- $C^T$ and $B^T$ are both right inverses of $A^T$
- under-determined equations $A^T x = (1, 2)$ has (different) solutions

$$B^T(1, 2) = (1/3, 2/3, -2/3), \quad C^T(1, 2) = (0, 1/2, -1)$$

(there are many other solutions as well)
Outline

Left and right inverses

Inverse

Solving linear equations

Examples

Pseudo-inverse
if \( A \) has a left and a right inverse, they are unique and equal (and we say that \( A \) is invertible)

so \( A \) must be square

to see this: if \( AX = I, YA = I \)

\[
X = IX = (YA)X = Y(AX) = YI = Y
\]

we denote them by \( A^{-1} \):

\[
A^{-1}A = AA^{-1} = I
\]

inverse of inverse: \( (A^{-1})^{-1} = A \)
Solving square systems of linear equations

- Suppose $A$ is invertible
- For any $b$, $Ax = b$ has the unique solution
  \[ x = A^{-1}b \]
- Matrix generalization of simple scalar equation $ax = b$ having solution
  \[ x = (1/a)b \text{ (for } a \neq 0) \]
- Simple-looking formula $x = A^{-1}b$ is basis for many applications
Invertible matrices

the following are equivalent for a square matrix $A$:

▶ $A$ is invertible
▶ columns of $A$ are linearly independent
▶ rows of $A$ are linearly independent
▶ $A$ has a left inverse
▶ $A$ has a right inverse
Examples

- $I^{-1} = I$
- if $Q$ is orthogonal, i.e., square with $Q^T Q = I$, then $Q^{-1} = Q^T$
- $2 \times 2$ matrix $A$ is invertible if and only if $A_{11}A_{22} \neq A_{12}A_{21}$

$$A^{-1} = \frac{1}{A_{11}A_{22} - A_{12}A_{21}} \begin{bmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{bmatrix}$$

- you need to know this formula
- there are similar but much more complicated formulas for larger matrices (and no, you do not need to know them)
Non-obvious example

\[
A = \begin{bmatrix}
1 & -2 & 3 \\
0 & 2 & 2 \\
-3 & -4 & -4
\end{bmatrix}
\]

- \(A\) is invertible, with inverse

\[
A^{-1} = \frac{1}{30} \begin{bmatrix}
0 & -20 & -10 \\
-6 & 5 & -2 \\
6 & 10 & 2
\end{bmatrix}.
\]

- verified by checking \(AA^{-1} = I\) (or \(A^{-1}A = I\))

- we’ll soon see how to compute the inverse
Properties

- \((AB)^{-1} = B^{-1}A^{-1}\) (provided inverses exist)
- \((A^T)^{-1} = (A^{-1})^T\) (sometimes denoted \(A^{-T}\))
- Negative matrix powers: \((A^{-1})^k\) is denoted \(A^{-k}\)
- With \(A^0 = I\), identity \(A^kA^l = A^{k+l}\) holds for any integers \(k, l\)
Triangular matrices

- lower triangular $L$ with nonzero diagonal entries is invertible

- so see this, write $Lx = 0$ as

\[
\begin{align*}
L_{11}x_1 & = 0 \\
L_{21}x_1 + L_{22}x_2 & = 0 \\
& \quad \vdots \\
L_{n1}x_1 + L_{n2}x_2 + \cdots + L_{n,n-1}x_{n-1} + L_{nn}x_n & = 0
\end{align*}
\]

- from first equation, $x_1 = 0$ (since $L_{11} \neq 0$)
- second equation reduces to $L_{22}x_2 = 0$, so $x_2 = 0$ (since $L_{22} \neq 0$)
- and so on

this shows columns of $L$ are linearly independent, so $L$ is invertible

- upper triangular $R$ with nonzero diagonal entries is invertible
Inverse via QR factorization

- Suppose $A$ is square and invertible.
- So its columns are linearly independent.
- So Gram–Schmidt gives QR factorization:
  - $A = QR$
  - $Q$ is orthogonal: $Q^T Q = I$
  - $R$ is upper triangular with positive diagonal entries, hence invertible.

- So we have

$$A^{-1} = (QR)^{-1} = R^{-1} Q^{-1} = R^{-1} Q^T$$
Outline

Left and right inverses

Inverse

Solving linear equations

Examples

Pseudo-inverse
Back substitution

- Suppose $R$ is upper triangular with nonzero diagonal entries.
- Write out $Rx = b$ as

\[
R_{11}x_1 + R_{12}x_2 + \cdots + R_{1,n-1}x_{n-1} + R_{1n}x_n = b_1 \\
\vdots \\
R_{n-1,n-1}x_{n-1} + R_{n-1,n}x_n = b_{n-1} \\
R_{nn}x_n = b_n
\]

- From last equation we get $x_n = b_n/R_{nn}$.
- From 2nd to last equation we get

\[
x_{n-1} = (b_{n-1} - R_{n-1,n}x_n)/R_{n-1,n-1}
\]

- Continue to get $x_{n-2}, x_{n-3}, \ldots, x_1$.
Back substitution

- called *back substitution* since we find the variables in reverse order, substituting the already known values of $x_i$
- computes $x = R^{-1}b$
- complexity:
  - first step requires 1 flop (division)
  - 2nd step needs 3 flops
  - $i$th step needs $2i - 1$ flops

  total is $1 + 3 + \cdots + (2n - 1) = n^2$ flops
Solving linear equations via QR factorization

- assuming $A$ is invertible, let’s solve $Ax = b$, i.e., compute $x = A^{-1}b$
- with $QR$ factorization $A = QR$, we have
  \[ A^{-1} = (QR)^{-1} = R^{-1}Q^T \]
- compute $x = R^{-1}(Q^Tb)$ by back substitution
Solving linear equations via QR factorization

given an $n \times n$ invertible matrix $A$ and an $n$-vector $b$

1. **QR factorization**: compute the QR factorization $A = QR$
2. compute $Q^T b$.
3. **Back substitution**: Solve the triangular equation $Rx = Q^T b$ using back substitution

- complexity $2n^3$ (step 1), $2n^2$ (step 2), $n^2$ (step 3)
- total is $2n^3 + 3n^2 \approx 2n^3$
Multiple right-hand sides

- let’s solve $Ax_i = b_i$, $i = 1, \ldots, k$, with $A$ invertible
- carry out QR factorization once (2$n^3$ flops)
- for $i = 1, \ldots, k$, solve $Rx_i = Q^T b_i$ via back substitution (3$kn^2$ flops)
- total is $2n^3 + 3kn^2$ flops
- if $k$ is small compared to $n$, same cost as solving one set of equations
Outline

Left and right inverses

Inverse

Solving linear equations

Examples

Pseudo-inverse
Polynomial interpolation

▶ let’s find coefficients of a cubic polynomial

\[ p(x) = c_1 + c_2 x + c_3 x^2 + c_4 x^3 \]

that satisfies

\[ p(-1.1) = b_1, \quad p(-0.4) = b_2, \quad p(0.1) = b_3, \quad p(0.8) = b_4 \]

▶ write as \( Ac = b \), with

\[
A = \begin{bmatrix}
1 & -1.1 & (-1.1)^2 & (-1.1)^3 \\
1 & -0.4 & (-0.4)^2 & (-0.4)^3 \\
1 & 0.1 & (0.1)^2 & (0.1)^3 \\
1 & 0.8 & (0.8)^2 & (0.8)^3 \\
\end{bmatrix}
\]
Polynomial interpolation

- (unique) coefficients given by $c = A^{-1}b$, with

$$A^{-1} = \begin{bmatrix}
-0.0370 & 0.3492 & 0.7521 & -0.0643 \\
0.1388 & -1.8651 & 1.6239 & 0.1023 \\
0.3470 & 0.1984 & -1.4957 & 0.9503 \\
-0.5784 & 1.9841 & -2.1368 & 0.7310 \\
\end{bmatrix}$$

- so, e.g., $c_1$ is not very sensitive to $b_1$ or $b_4$

- first column gives coefficients of polynomial that satisfies

$$p(-1.1) = 1, \quad p(-0.4) = 0, \quad p(0.1) = 0, \quad p(0.8) = 0$$

called (first) *Lagrange polynomial*
Example
Lagrange polynomials

Lagrange polynomials associated with points $-1.1, -0.4, 0.2, 0.8$
Outline

Left and right inverses

Inverse

Solving linear equations

Examples

Pseudo-inverse
Invertibility of Gram matrix

- $A$ has linearly independent columns if and only if $A^T A$ is invertible
- to see this, we’ll show that $Ax = 0 \iff A^T Ax = 0$
- $\Rightarrow$: if $Ax = 0$ then $(A^T A)x = A^T(Ax) = A^T 0 = 0$
- $\Leftarrow$: if $(A^T A)x = 0$ then
  
  $0 = x^T (A^T A)x = (Ax)^T (Ax) = \|Ax\|^2 = 0$

so $Ax = 0$
the pseudo-inverse of $A$ with independent columns is

$$A^\dagger = (A^T A)^{-1} A^T$$

it is a left inverse of $A$:

$$A^\dagger A = (A^T A)^{-1} A^T A = (A^T A)^{-1} (A^T A) = I$$

(we’ll soon see that it’s a very important left inverse of $A$)

reduces to $A^{-1}$ when $A$ is square:

$$A^\dagger = (A^T A)^{-1} A^T = A^{-1} A^{-T} A^T = A^{-1} I = A^{-1}$$
Pseudo-inverse of wide matrix

- if $A$ is wide, with linearly independent rows, $AA^T$ is invertible
- pseudo-inverse is defined as

\[ A^\dagger = A^T (AA^T)^{-1} \]

- $A^\dagger$ is a right inverse of $A$:

\[ AA^\dagger = AA^T (AA^T)^{-1} = I \]

(we’ll see later it is an important right inverse)

- reduces to $A^{-1}$ when $A$ is square:

\[ A^T (AA^T)^{-1} = A^T A^{-T} A^{-1} = A^{-1} \]
Pseudo-inverse via QR factorization

- Suppose $A$ has linearly independent columns, $A = QR$
- Then $A^T A = (QR)^T (QR) = R^T Q^T QR = R^T R$
- So

$$A^\dagger = (A^T A)^{-1} A^T = (R^T R)^{-1} (QR)^T = R^{-1} R^{-T} R^T Q^T = R^{-1} Q^T$$

- Can compute $A^\dagger$ using back substitution on columns of $Q^T$
- For $A$ with linearly independent rows, $A^\dagger = QR^{-T}$