Lecture 23: Composition of linear maps and the chain rule

December 7, 2010

Announcements

- Problem set 11 due Monday, December 13, by 5PM.
- Upcoming office hours:
  - Wednesday, December 8, 2–3 PM.
  - Monday, December 13, 2–3 PM.
- Course evaluations
  - Available online now through December 15.
  - Important for us, to help improve teaching.
  - If not filled out now, grades blocked until January 7.
- Example: A non-differentiable function
Overview

Goal
Understand the chain rule better

Mantra
To find total derivative of a composition $G \circ F$, compose their total derivatives $DG \circ DF$.
Recall that for a function $F : \mathbb{R}^n \to \mathbb{R}^m$, the total derivative $DF(a)$ at $a \in \mathbb{R}^n$ is the best linear approximation.
So first we will understand composition of linear functions.
Lecture 23: Composition of linear maps and the chain rule

▶ Composition of linear maps and matrix multiplication

Properties of matrix multiplication

Chain rule
Composition of linear maps

Given two linear maps $L : \mathbb{R}^p \to \mathbb{R}^n$ and $M : \mathbb{R}^n \to \mathbb{R}^m$, the composition $M \circ L$ is also linear.

How do matrices look? How does $[M \circ L]$ relate to $[M]$ and $[L]$?

Definition

Given $m \times n$ and $n \times p$ matrices

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,p} \\ b_{2,1} & b_{2,2} & \cdots & b_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n,1} & b_{n,2} & \cdots & b_{n,p} \end{pmatrix},$$

the matrix product $AB$ of $A$ and $B$ is the $m \times p$ matrix whose $(i, j)$ entry is

$$c_{i,j} = a_{i,1}b_{1,j} + a_{i,2}b_{2,j} + \cdots + a_{i,n}b_{n,j}.$$

Example

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} = ? \quad \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} = ?$$
Matrix multiplication notes

- A matrix with $m$ rows and $n$ columns represents a map from $\mathbb{R}^n$ to $\mathbb{R}^m$. To be able to compose functions, the spaces must match. So in matrix product $AB$, must have
  
  $(\text{columns of } A) = (\text{rows of } B)$.
- Special case I: If $B$ has one column, looks like application of linear map.
- Special case II: If $A$ has one row and $B$ has one column, looks like dot product. (Put one vector vertically.)
- What happens if $A$ and $B$ are $1 \times 1$? What happens if $A$ is $1 \times 1$?

Examples

- How can we multiply these matrices so result makes sense?
  
  $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$

- Linear map $L(x, y) = (x + y, x - y, y)$ applied to vector $(2, 3)$ is
  
  $\begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 5 \\ -1 \\ 3 \end{pmatrix}$

- Dot product $\langle 1, 2 \rangle \cdot \langle 3, 4 \rangle$ is
  
  $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = (9)$
Matrix multiplication notes

- A matrix with \( m \) rows and \( n \) columns represents a map from \( \mathbb{R}^n \) to \( \mathbb{R}^m \). To be able to compose functions, the spaces must match. So in matrix product \( AB \), must have 
  \[ \#(\text{columns of } A) = \#(\text{rows of } B). \]
- Special case I: If \( B \) has one column, looks like application of linear map.
- Special case II: If \( A \) has one row and \( B \) has one column, looks like dot product. (Put one vector vertically.)
- What happens if \( A \) and \( B \) are \( 1 \times 1 \)? What happens if \( A \) is \( 1 \times 1 \)?

Examples

- How can we multiply these matrices so result makes sense?
  \[
  \begin{pmatrix}
  1 & 2 & 3 \\
  4 & 5 & 6 \\
  \end{pmatrix}
  \begin{pmatrix}
  1 & 1 \\
  0 & 1 \\
  \end{pmatrix}
  =
  \begin{pmatrix}
  1 & 1 & 1 \\
  0 & 1 & 1 \\
  0 & 0 & 1 \\
  \end{pmatrix}
  \]
- Linear map \( L(x, y) = (x + y, x - y, y) \) applied to vector \((2, 3)\) is
  \[
  \begin{pmatrix}
  1 & 1 \\
  1 & -1 \\
  0 & 1 \\
  \end{pmatrix}
  \begin{pmatrix}
  2 \\
  3 \\
  \end{pmatrix}
  =
  \begin{pmatrix}
  5 \\
  -1 \\
  3 \\
  \end{pmatrix}
  \]
- Dot product \( \langle 1, 2 \rangle \cdot \langle 3, 4 \rangle \) is
  \[
  (1 2)
  \begin{pmatrix}
  3 \\
  4 \\
  \end{pmatrix}
  = (9)
  \]
Matrix multiplication notes

- A matrix with $m$ rows and $n$ columns represents a map from $\mathbb{R}^n$ to $\mathbb{R}^m$. To be able to compose functions, the spaces must match. So in matrix product $AB$, must have
  $\#$(columns of $A) = \#$(rows of $B$).
- Special case I: If $B$ has one column, looks like application of linear map.
- Special case II: If $A$ has one row and $B$ has one column, looks like dot product. (Put one vector vertically.)
- What happens if $A$ and $B$ are $1 \times 1$?
  What happens if $A$ is $1 \times 1$?

Examples

- How can we multiply these matrices so result makes sense?
  \[
  \begin{pmatrix}
  1 & 2 & 3 \\
  4 & 5 & 6
  \end{pmatrix}
  \begin{pmatrix}
  1 & 1 \\
  0 & 1
  \end{pmatrix}
  \begin{pmatrix}
  1 & 1 & 1 \\
  0 & 1 & 1 \\
  0 & 0 & 1
  \end{pmatrix}
  \]
- Linear map $L(x, y) = (x + y, x - y, y)$ applied to vector $(2, 3)$ is
  \[
  \begin{pmatrix}
  1 & 1 \\
  1 & -1 \\
  0 & 1
  \end{pmatrix}
  \begin{pmatrix}
  2 \\
  3
  \end{pmatrix}
  =
  \begin{pmatrix}
  5 \\
  -1 \\
  3
  \end{pmatrix}
  \]
- Dot product $\langle 1, 2 \rangle \cdot \langle 3, 4 \rangle$ is
  \[
  \begin{pmatrix}
  1 & 2
  \end{pmatrix}
  \begin{pmatrix}
  3 \\
  4
  \end{pmatrix}
  =
  (9)
Matrix multiplication notes

- A matrix with \( m \) rows and \( n \) columns represents a map from \( \mathbb{R}^n \) to \( \mathbb{R}^m \). To be able to compose functions, the spaces must match. So in matrix product \( AB \), must have 
  \#(columns of A) = \#(rows of B).
- Special case I: If \( B \) has one column, looks like application of linear map.
- Special case II: If \( A \) has one row and \( B \) has one column, looks like dot product. (Put one vector vertically.)
- What happens if \( A \) and \( B \) are \( 1 \times 1 \)?
  What happens if \( A \) is \( 1 \times 1 \)?

Examples

- How can we multiply these matrices so result makes sense?
  \[
  \begin{pmatrix}
  1 & 2 & 3 \\
  4 & 5 & 6
  \end{pmatrix}
  \begin{pmatrix}
  1 & 1 \\
  0 & 1
  \end{pmatrix}
  \begin{pmatrix}
  1 & 1 & 1 \\
  0 & 1 & 1 \\
  0 & 0 & 1
  \end{pmatrix}
  \]
- Linear map \( L(x, y) = (x + y, x - y, y) \) applied to vector \((2, 3)\) is
  \[
  \begin{pmatrix}
  1 & 1 \\
  1 & -1 \\
  0 & 1
  \end{pmatrix}
  \begin{pmatrix}
  2 \\
  3
  \end{pmatrix}
  =
  \begin{pmatrix}
  5 \\
  -1 \\
  3
  \end{pmatrix}
  \]
- Dot product \( \langle 1, 2 \rangle \cdot \langle 3, 4 \rangle \) is
  \[
  \begin{pmatrix}
  1 & 2
  \end{pmatrix}
  \begin{pmatrix}
  3 \\
  4
  \end{pmatrix}
  = (9)
  \]
Matrix multiplication proof

Theorem
If $L : \mathbb{R}^p \to \mathbb{R}^n$ and $M : \mathbb{R}^n \to \mathbb{R}^m$ are linear maps, then
\[
[M \circ L] = [M][L].
\]

Proof.
We’ll do the $2 \times 2$ case to try to avoid notation overflow.

$L(x, y) = (ax + by, cx + dy)$ \hspace{1cm} $M(x, y) = (ex + fy, gx + hy)$.

On one hand:
\[
(M \circ L)(x, y) = (e(ax + by) + f(cx + dy), g(ax + by) + h(cx + dy))
\]
\[
= ((ea + fc)x + (eb + fd)y, (ga + hc)x + (gb + hd)y)
\]
\[
[M \circ L] = \begin{pmatrix}
ea + fc & eb + fd \\
ga + hc & gb + hd
\end{pmatrix}.
\]

On other hand:
\[
[M] = \begin{pmatrix}
e & f \\
g & h
\end{pmatrix} \hspace{1cm} [L] = \begin{pmatrix}
a & b \\
c & d
\end{pmatrix} \hspace{1cm} [M][L] = \begin{pmatrix}
ea + fc & eb + fd \\
ga + hc & gb + hd
\end{pmatrix}.
\]

\qed
Matrix multiplication proof

Theorem

If $L : \mathbb{R}^p \to \mathbb{R}^n$ and $M : \mathbb{R}^n \to \mathbb{R}^m$ are linear maps, then

$[M \circ L] = [M][L]$.  

Proof.

We’ll do the $2 \times 2$ case to try to avoid notation overflow.

$L(x, y) = (ax + by, cx + dy)$  
$M(x, y) = (ex + fy, gx + hy)$.  

On one hand:

$$(M \circ L)(x, y) = (e(ax + by) + f(cx + dy), g(ax + by) + h(cx + dy))$$

$= ((ea + fc)x + (eb + fd)y, (ga + hc)x + (gb + hd)y)$$

$[M \circ L] = \begin{pmatrix} ea + fc & eb + fd \\ ga + hc & gb + hd \end{pmatrix}$.  

On other hand:

$[M] = \begin{pmatrix} e & f \\ g & h \end{pmatrix}$  
$[L] = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$  
$[M][L] = \begin{pmatrix} ea + fc & eb + fd \\ ga + hc & gb + hd \end{pmatrix}$.
Matrix multiplication proof

**Theorem**

If $L : \mathbb{R}^p \to \mathbb{R}^n$ and $M : \mathbb{R}^n \to \mathbb{R}^m$ are linear maps, then

$$[M \circ L] = [M][L].$$

**Proof.**

We’ll do the $2 \times 2$ case to try to avoid notation overflow.

$L(x, y) = (ax + by, cx + dy)$  \quad  $M(x, y) = (ex + fy, gx + hy)$.

On one hand:

$$(M \circ L)(x, y) = (e(ax + by) + f(cx + dy), g(ax + by) + h(cx + dy))$$

$$= ((ea + fc)x + (eb + fd)y, (ga + hc)x + (gb + hd)y)$$

$$[M \circ L] = \begin{pmatrix} ea + fc & eb + fd \\ ga + hc & gb + hd \end{pmatrix}.$$  

On other hand:

$$[M] = \begin{pmatrix} e & f \\ g & h \end{pmatrix} \quad [L] = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad [M][L] = \begin{pmatrix} ea + fc & eb + fd \\ ga + hc & gb + hd \end{pmatrix}.$$  

$\square$
Matrix multiplication proof

**Theorem**

If \( L : \mathbb{R}^p \to \mathbb{R}^n \) and \( M : \mathbb{R}^n \to \mathbb{R}^m \) are linear maps, then

\[
[M \circ L] = [M][L].
\]

**Proof.**

We’ll do the \(2 \times 2\) case to try to avoid notation overflow.

\( L(x, y) = (ax + by, cx + dy) \quad M(x, y) = (ex + fy, gx + hy). \)

On one hand:

\[
(M \circ L)(x, y) = (e(ax + by) + f(cx + dy), g(ax + by) + h(cx + dy))
\]

\[
= ((ea + fc)x + (eb + fd)y, (ga + hc)x + (gb + hd)y)
\]

\[
[M \circ L] = \begin{pmatrix}
    ea + fc & eb + fd \\
    ga + hc & gb + hd
\end{pmatrix}.
\]

On other hand:

\[
[M] = \begin{pmatrix}
    e & f \\
    g & h
\end{pmatrix} \quad [L] = \begin{pmatrix}
    a & b \\
    c & d
\end{pmatrix} \quad [M][L] = \begin{pmatrix}
    ea + fc & eb + fd \\
    ga + hc & gb + hd
\end{pmatrix}.
\]

\( \square \)
Matrix multiplication proof

**Theorem**

If \( L : \mathbb{R}^p \to \mathbb{R}^n \) and \( M : \mathbb{R}^n \to \mathbb{R}^m \) are linear maps, then

\[
[M \circ L] = [M][L].
\]

**Proof.**

We’ll do the \( 2 \times 2 \) case to try to avoid notation overflow.

\[
L(x, y) = (ax + by, cx + dy) \quad M(x, y) = (ex + fy, gx + hy).
\]

On one hand:

\[
(M \circ L)(x, y) = (e(ax + by) + f(cx + dy), g(ax + by) + h(cx + dy))
\]

\[
= ((ea + fc)x + (eb + fd)y, (ga + hc)x + (gb + hd)y).
\]

\[
[M \circ L] = \begin{pmatrix}
ea + fc & eb + fd \\
(\ \ \ ga + hc & gb + hd
\end{pmatrix}.
\]

On other hand:

\[
[M] = \begin{pmatrix}
e & f \\
g & h
\end{pmatrix}, \quad [L] = \begin{pmatrix}
a & b \\
c & d
\end{pmatrix}, \quad [M][L] = \begin{pmatrix}
ea + fc & eb + fd \\
(\ \ \ ga + hc & gb + hd
\end{pmatrix}.
\]

\( \square \)
Matrix multiplication proof

Theorem
If \( L : \mathbb{R}^p \to \mathbb{R}^n \) and \( M : \mathbb{R}^n \to \mathbb{R}^m \) are linear maps, then
\[
[M \circ L] = [M][L].
\]

Proof.
We'll do the 2 \( \times \) 2 case to try to avoid notation overflow.

\[
L(x, y) = (ax + by, cx + dy) \quad \quad M(x, y) = (ex + fy, gx + hy).
\]

On one hand:
\[
(M \circ L)(x, y) = (e(ax + by) + f(cx + dy), g(ax + by) + h(cx + dy))
= ((ea + fc)x + (eb + fd)y, (ga + hc)x + (gb + hd)y)
\]
\[
[M \circ L] = \begin{pmatrix}
ea + fc & eb + fd \\
ga + hc & gb + hd
\end{pmatrix}.
\]

On other hand:
\[
[M] = \begin{pmatrix}
e & f \\
g & h
\end{pmatrix} \quad \quad [L] = \begin{pmatrix}
a & b \\
c & d
\end{pmatrix} \quad \quad [M][L] = \begin{pmatrix}
ea + fc & eb + fd \\
ga + hc & gb + hd
\end{pmatrix}.
\]

\( \square \)
Matrix multiplication proof

**Theorem**

If \( L : \mathbb{R}^p \to \mathbb{R}^n \) and \( M : \mathbb{R}^n \to \mathbb{R}^m \) are linear maps, then

\[
[M \circ L] = [M][L].
\]

**Proof.**

We’ll do the \( 2 \times 2 \) case to try to avoid notation overflow.

\( L(x, y) = (ax + by, cx + dy) \quad M(x, y) = (ex + fy, gx + hy). \)

On one hand:

\[
(M \circ L)(x, y) = (e(ax + by) + f(cx + dy), g(ax + by) + h(cx + dy))
= ((ea + fc)x + (eb + fd)y, (ga + hc)x + (gb + hd)y)
\]

\[
[M \circ L] = \begin{pmatrix}
  ea + fc & eb + fd \\
  ga + hc & gb + hd
\end{pmatrix}.
\]

On other hand:

\[
[M] = \begin{pmatrix}
  e & f \\
  g & h
\end{pmatrix} \quad [L] = \begin{pmatrix}
  a & b \\
  c & d
\end{pmatrix} \quad [M][L] = \begin{pmatrix}
  ea + fc & eb + fd \\
  ga + hc & gb + hd
\end{pmatrix}.
\]

\( \square \)
Matrix multiplication proof

Theorem

If $L : \mathbb{R}^p \to \mathbb{R}^n$ and $M : \mathbb{R}^n \to \mathbb{R}^m$ are linear maps, then

$$[M \circ L] = [M][L].$$

Proof.

We’ll do the $2 \times 2$ case to try to avoid notation overflow.

$L(x, y) = (ax + by, cx + dy)$  \hspace{1cm}  $M(x, y) = (ex + fy, gx + hy)$.

On one hand:

$$(M \circ L)(x, y) = (e(ax + by) + f(cx + dy), g(ax + by) + h(cx + dy))$$

$$= \left((ea + fc)x + (eb + fd)y, (ga + hc)x + (gb + hd)y\right)$$

$$[M \circ L] = \begin{pmatrix} ea + fc & eb + fd \\ ga + hc & gb + hd \end{pmatrix}.$$ 

On other hand:

$$[M] = \begin{pmatrix} e & f \\ g & h \end{pmatrix} \quad [L] = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad [M][L] = \begin{pmatrix} ea + fc & eb + fd \\ ga + hc & gb + hd \end{pmatrix}.$$  \hspace{1cm} $\square$
Lecture 23: Composition of linear maps and the chain rule

Composition of linear maps and matrix multiplication

Properties of matrix multiplication

Chain rule
Associativity

**Theorem**

Matrix multiplication is associative: For matrices $A$, $B$, $C$ (with matching dimensions),

$$A(BC) = (AB)C.$$  

**Proof.**

Suppose $A$, $B$, and $C$ represent linear maps $L$, $M$, and $N$, respectively.

Then $L \circ (M \circ N) = (L \circ M) \circ N$ (it means, do $N$, then $M$, then $L$, in that order), so

$$A(BC) = [L \circ (M \circ N)] = [(L \circ M) \circ N] = (AB)C. \quad \Box$$

(Alternatively, could write out the sums.)

**Warning**

Matrix product is rarely commutative: $AB \neq BA$ in general. (May not be same size, or even both defined.)

**Example**

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

**Theorem**

*Matrix multiplication is associative:* For matrices $A$, $B$, $C$ (with matching dimensions),

$$A(BC) = (AB)C.$$  

**Proof.**

Suppose $A$, $B$, and $C$ represent linear maps $L$, $M$, and $N$, respectively.

Then $L \circ (M \circ N) = (L \circ M) \circ N$ (it means, do $N$, then $M$, then $L$, in that order), so

$$A(BC) = [L \circ (M \circ N)] = [(L \circ M) \circ N] = (AB)C. \quad \square$$

(Alternatively, could write out the sums.)

**Warning**

Matrix product is rarely commutative: $AB \neq BA$ in general. (May not be same size, or even both defined.)

**Example**

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & (2) \\ 1 & 1 & (3) \end{pmatrix} \begin{pmatrix} 1 & 2 \\ -1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 2 \\ -1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$
Special matrices

For more properties, introduce two special matrices:

- The $n \times n$ identity matrix is the matrix with 1's on its diagonal:

$$I_{n \times n} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}.$$  

The matrix $I_{n \times n}$ represents the identity map $\mathbb{I}: \mathbb{R}^n \to \mathbb{R}^n$ given by $\mathbb{I}(\vec{x}) = \vec{x}$.

- The $m \times n$ zero matrix $0_{m \times n}$ has all entries 0:

$$0_{m \times n} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}.$$
 Arithmetic with matrices

Can also add matrices: if $A$ and $B$ are both $n \times m$ matrices then $A + B$ is obtained by adding corresponding entries of $A$ and $B$.

**Theorem**

Suppose $A$, $B$, $C$ and $D$ are $n \times m$, $n \times m$, $m \times l$ and $p \times n$ matrices, respectively. Then:

1. *(Addition is commutative:)* $A + B = B + A$.
2. *(Additive identity:)* $0_{n \times m} + A = A$.
3. *(Multiplicative identity:)* $I_{n \times n} A = A = AI_{m \times m}$.
4. *(Multiplicative zero:)* $A 0_{m \times k} = 0_{n \times k}$.

5. *(Multiplication is associative:)* $D(AC) = (DA)C$.
6. *(Distributivity:)* $(A + B)C = AC + BC$ and $D(A + B) = DA + DB$.

**Examples**

$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} + \begin{pmatrix} 2 & 2 & 3 \\ 3 & 4 & 4 \end{pmatrix} = \begin{pmatrix} 3 & 4 & 6 \\ 7 & 9 & 10 \end{pmatrix}$

$\begin{pmatrix} 2 & 2 & 3 \\ 3 & 4 & 4 \end{pmatrix} + \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} 3 & 4 & 6 \\ 7 & 9 & 10 \end{pmatrix}$

$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$

etc.
More examples

Exercise

- Let $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. What is $A^2$? $A^3$? $A^4$? Do you see a pattern?

- Let $B = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$. What is $B^2$? $B^3$? $B^4$? Do you see a pattern?
Lecture 23: Composition of linear maps and the chain rule

Composition of linear maps and matrix multiplication

Properties of matrix multiplication

▶ Chain rule
Chain rule statement

On to the point!
Recall that, for $F: \mathbb{R}^n \to \mathbb{R}^m$ any function, the total derivative $DF(\vec{a})$ is the best linear approximation to $F$ near $\vec{a} \in \mathbb{R}^n$:

$$F(\vec{a} + \vec{h}) \approx F(\vec{a}) + DF(\vec{a})(\vec{h}).$$

Theorem

For $F: \mathbb{R}^n \to \mathbb{R}^m$ and $G: \mathbb{R}^m \to \mathbb{R}^p$ differentiable functions,

$$D(G \circ F)(\vec{a}) = DG(F(\vec{a})) \circ DF(\vec{a})$$

or, in terms of the Jacobian matrices,

$$[D(G \circ F)(\vec{a})] = [DG(F(\vec{a}))][DF(\vec{a})].$$

Example

For $F(x, y) = (x^3 - y^2, x^2y)$, have

$$[DF(x, y)] = \begin{pmatrix} 3x^2 & -2y \\ 2xy & x^2 \end{pmatrix}$$

$F(-1, 1) = (-2, 1)$

$$[DF(-1, 1)] = \begin{pmatrix} 3 & -2 \\ -2 & 1 \end{pmatrix}.$$

For $G(x, y) = (y^3, x^2)$,

$$[DG(-2, 1)] = \begin{pmatrix} 0 & 3 \\ 4 & 0 \end{pmatrix}$$

$$[D(G \circ F)(-1, 1)] = \begin{pmatrix} 0 & 3 \\ 4 & 0 \end{pmatrix} \begin{pmatrix} 3 & -2 \\ -2 & 1 \end{pmatrix} = \begin{pmatrix} -6 & 3 \\ 12 & -8 \end{pmatrix}.$$
Chain rule statement

On to the point!
Recall that, for $F : \mathbb{R}^n \to \mathbb{R}^m$ any function, the total derivative $DF(\vec{a})$ is the best linear approximation to $F$ near $\vec{a} \in \mathbb{R}^n$:

$$F(\vec{a} + \vec{h}) \approx F(\vec{a}) + DF(\vec{a})(\vec{h}).$$

**Theorem**

For $F : \mathbb{R}^n \to \mathbb{R}^m$ and $G : \mathbb{R}^m \to \mathbb{R}^p$ differentiable functions,

$$D(G \circ F)(\vec{a}) = DG(F(\vec{a})) \circ DF(\vec{a})$$

or, in terms of the Jacobian matrices,

$$[D(G \circ F)(\vec{a})] = [DG(F(\vec{a}))][DF(\vec{a})].$$

**Example**

For $F(x, y) = (x^3 - y^2, x^2y)$, have

$$[DF(x, y)] = \begin{pmatrix} 3x^2 & -2y \\ 2xy & x^2 \end{pmatrix}$$

$$F(-1, 1) = (-2, 1)$$

$$[DF(-1, 1)] = \begin{pmatrix} 3 & -2 \\ -2 & 1 \end{pmatrix}.$$
Chain rule example

Let’s do that example the other way around. Had

\[ G(x, y) = (y^3, x^2) \quad F(x, y) = (x^3 - y^2, x^2y) \]

\[ [DG(F(-1, 1))] = \begin{pmatrix} 0 & 3 \\ -4 & 0 \end{pmatrix} \quad [DF(-1, 1)] = \begin{pmatrix} 3 & -2 \\ -2 & 1 \end{pmatrix} \]

\[ [DG(F(-1, 1))] [DF(-1, 1)] = \begin{pmatrix} -6 & 3 \\ -12 & 8 \end{pmatrix} \]

On the other hand,

\[ (G \circ F)(x, y) = G(x^3 - y^2, x^2y) \]
\[ = ((x^2y)^3, (x^3 - y^2)^2) \]
\[ = (x^6y^3, x^6 - 2x^3y^2 + y^4) \]

\[ [D(G \circ F)(x, y)] = \begin{pmatrix} 6x^5y^3 & 3x^6y^3 \\ 6x^5 - 6x^2y^2 & -4x^3y + 4y^3 \end{pmatrix} \]

\[ [D(G \circ F)(-1, 1)] = ? \]
Chain rule example

Let’s do that example the other way around. Had

\[ G(x, y) = (y^3, x^2) \quad F(x, y) = (x^3 - y^2, x^2y) \]

\[
\begin{bmatrix}
    DG(F(-1, 1))
\end{bmatrix} = \begin{pmatrix}
    0 & 3 \\
    -4 & 0
\end{pmatrix}
\quad
\begin{bmatrix}
    DF(-1, 1)
\end{bmatrix} = \begin{pmatrix}
    3 & -2 \\
    -2 & 1
\end{pmatrix}
\]

\[
\begin{bmatrix}
    DG(F(-1, 1))\end{bmatrix}\begin{bmatrix}
    DF(-1, 1)
\end{bmatrix} = \begin{pmatrix}
    -6 & 3 \\
    -12 & 8
\end{pmatrix}
\]

On the other hand,

\[(G \circ F)(x, y) = G(x^3 - y^2, x^2y)\]

\[
= ((x^2y)^3, (x^3 - y^2)^2) \\
= (x^6y^3, x^6 - 2x^3y^2 + y^4)
\]

\[
\begin{bmatrix}
    D(G \circ F)(x, y)
\end{bmatrix} = \begin{pmatrix}
    6x^5y^3 & 3x^6y^3 \\
    6x^5 - 6x^2y^2 & -4x^3y + 4y^3
\end{pmatrix}
\]

\[
\begin{bmatrix}
    D(G \circ F)(-1, 1)
\end{bmatrix} = ?
\]
Chain rule example

Let’s do that example the other way around. Had

\[
G(x, y) = (y^3, x^2) \quad F(x, y) = (x^3 - y^2, x^2y)
\]

\[
[DG(F(-1, 1))] = \begin{pmatrix} 0 & 3 \\ -4 & 0 \end{pmatrix} \quad [DF(-1, 1)] = \begin{pmatrix} 3 & -2 \\ -2 & 1 \end{pmatrix}
\]

\[
[DG(F(-1, 1))] [DF(-1, 1)] = \begin{pmatrix} -6 & 3 \\ -12 & 8 \end{pmatrix}
\]

On the other hand,

\[
(G \circ F)(x, y) = G(x^3 - y^2, x^2y)
\]

\[
= ((x^2y)^3, (x^3 - y^2)^2)
\]

\[
= (x^6y^3, x^6 - 2x^3y^2 + y^4)
\]

\[
[D(G \circ F)(x, y)] = \begin{pmatrix} 6x^5y^3 & 3x^6y^3 \\ 6x^5 - 6x^2y^2 & -4x^3y + 4y^3 \end{pmatrix}
\]

\[
[D(G \circ F)(-1, 1)] = ?
\]
Chain rule example

Let’s do that example the other way around. Had

\[ G(x, y) = (y^3, x^2) \quad F(x, y) = (x^3 - y^2, x^2 y) \]

\[
[DG(F(-1, 1))] = \begin{pmatrix} 0 & 3 \\ -4 & 0 \end{pmatrix} \quad [DF(-1, 1)] = \begin{pmatrix} 3 & -2 \\ -2 & 1 \end{pmatrix}
\]

\[
[DG(F(-1, 1))] [DF(-1, 1)] = \begin{pmatrix} -6 & 3 \\ -12 & 8 \end{pmatrix}
\]

On the other hand,

\[
(G \circ F)(x, y) = G(x^3 - y^2, x^2 y)
\]

\[
= ((x^2 y)^3, (x^3 - y^2)^2)
\]

\[
= (x^6 y^3, x^6 - 2x^3 y^2 + y^4)
\]

\[
[D(G \circ F)(x, y)] = \begin{pmatrix} 6x^5 y^3 & 3x^6 y^3 \\ 6x^5 - 6x^2 y^2 & -4x^3 y + 4y^3 \end{pmatrix}
\]

\[
[D(G \circ F)(-1, 1)] = ?
\]
Chain rule example

Let’s do that example the other way around. Had

\[ G(x, y) = (y^3, x^2) \quad F(x, y) = (x^3 - y^2, x^2y) \]

\[
\begin{bmatrix}
DG(F(-1, 1))
\end{bmatrix} =
\begin{pmatrix}
0 & 3 \\
-4 & 0
\end{pmatrix}
\quad
[DF(-1, 1)] =
\begin{pmatrix}
3 & -2 \\
-2 & 1
\end{pmatrix}
\]

\[
[DG(F(-1, 1))][DF(-1, 1)] =
\begin{pmatrix}
-6 & 3 \\
-12 & 8
\end{pmatrix}
\]

On the other hand,

\[ (G \circ F)(x, y) = G(x^3 - y^2, x^2y) \]

\[ = (((x^2y)^3, (x^3 - y^2)^2) \]

\[ = (x^6y^3, x^6 - 2x^3y^2 + y^4) \]

\[
[D(G \circ F)(x, y)] =
\begin{pmatrix}
6x^5y^3 & 3x^6y^3 \\
6x^5 - 6x^2y^2 & -4x^3y + 4y^3
\end{pmatrix}
\]

\[ [D(G \circ F)(-1, 1)] = ? \]
**Chain rule example**

Let’s do that example the other way around. Had

\[ G(x, y) = (y^3, x^2) \quad F(x, y) = (x^3 - y^2, x^2y) \]

\[ [DG(F(-1, 1))] = \begin{pmatrix} 0 & 3 \\ -4 & 0 \end{pmatrix} \quad [DF(-1, 1)] = \begin{pmatrix} 3 & -2 \\ -2 & 1 \end{pmatrix} \]

\[ [DG(F(-1, 1))] [DF(-1, 1)] = \begin{pmatrix} -6 & 3 \\ -12 & 8 \end{pmatrix} \]

On the other hand,

\[(G \circ F)(x, y) = G(x^3 - y^2, x^2y)\]
\[ = ((x^2y)^3, (x^3 - y^2)^2) \]
\[ = (x^6y^3, x^6 - 2x^3y^2 + y^4) \]

\[ [D(G \circ F)(x, y)] = \begin{pmatrix} 6x^5y^3 & 3x^6y^3 \\ 6x^5 - 6x^2y^2 & -4x^3y + 4y^3 \end{pmatrix} \]

\[ [D(G \circ F)(-1, 1)] = ? \]
Chain rule example

Let’s do that example the other way around. Had

\[ G(x, y) = (y^3, x^2) \quad F(x, y) = (x^3 - y^2, x^2y) \]

\[ [DG(F(-1, 1))] = \begin{pmatrix} 0 & 3 \\ -4 & 0 \end{pmatrix} \quad [DF(-1, 1)] = \begin{pmatrix} 3 & -2 \\ -2 & 1 \end{pmatrix} \]

\[ [DG(F(-1, 1))] [DF(-1, 1)] = \begin{pmatrix} -6 & 3 \\ -12 & 8 \end{pmatrix} \]

On the other hand,

\[ (G \circ F)(x, y) = G(x^3 - y^2, x^2y) \]
\[ = ((x^2y)^3, (x^3 - y^2)^2) \]
\[ = (x^6y^3, x^6 - 2x^3y^2 + y^4) \]

\[ [D(G \circ F)(x, y)] = \begin{pmatrix} 6x^5y^3 & 3x^6y^3 \\ 6x^5 - 6x^2y^2 & -4x^3y + 4y^3 \end{pmatrix} \]

\[ [D(G \circ F)(-1, 1)] = ? \]
Some special cases

For $f : \mathbb{R} \rightarrow \mathbb{R}$ and $g : \mathbb{R} \rightarrow \mathbb{R}$, let $s = f(x)$.

$$(g \circ f)'(x) = g'(f(x))f'(x)$$

$$
\left( \frac{dg}{dx} \right) = \left( \frac{dg}{df} \right) \left( \frac{df}{dx} \right).
$$

For $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ and $g : \mathbb{R} \rightarrow \mathbb{R}$,

$$
\frac{\partial g}{\partial x} = \frac{dg}{df} \frac{\partial f}{\partial x} \quad \frac{\partial g}{\partial y} = \frac{dg}{df} \frac{\partial f}{\partial y}
$$

$$
\left( \frac{\partial g}{\partial x} \frac{\partial g}{\partial y} \right) = \left( \frac{dg}{df} \right) \left( \frac{\partial f}{\partial x} \frac{\partial f}{\partial y} \right).
$$

For $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ and $g : \mathbb{R}^2 \rightarrow \mathbb{R}$, write

$$f(r, s) = (x(r, s), y(r, s)).$$

$$
\frac{\partial g}{\partial r} = \frac{\partial g}{\partial x} \frac{\partial x}{\partial r} + \frac{\partial g}{\partial y} \frac{\partial y}{\partial r} \quad \frac{\partial g}{\partial s} = \frac{\partial g}{\partial x} \frac{\partial x}{\partial s} + \frac{\partial g}{\partial y} \frac{\partial y}{\partial s}
$$

$$
\left( \frac{\partial g}{\partial r} \frac{\partial g}{\partial s} \right) = \left( \frac{\partial g}{\partial x} \frac{\partial g}{\partial y} \right) \left( \begin{array}{c} \frac{\partial x}{\partial r} \\ \frac{\partial x}{\partial s} \\ \frac{\partial y}{\partial r} \\ \frac{\partial y}{\partial s} \end{array} \right).
$$

Examples

What is the derivative of $\sin(x^2)$?

If the height of a mountain is $h(x, y) = 5 - x^2 - 4y^2$ and the temperature varies with altitude as $T(h) = h^2$, what is $\partial T/\partial x$ at $(1, 1)$?

If the height of a mountain is $h(x, y) = 5 - x^2 - 4y^2$, what is the change in the height if you are at $(1, 1)$ and move directly away from the peak at $(0, 0)$?
Some special cases

▶ For $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$, let $s = f(x)$.

$$(g \circ f)'(x) = g'(f(x))f'(x)$$

$$\left( \frac{dg}{dx} \right) = \left( \frac{dg}{df} \right) \left( \frac{df}{dx} \right).$$

▶ For $f : \mathbb{R}^2 \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$,

$$\frac{\partial g}{\partial x} = \frac{dg}{df} \frac{\partial f}{\partial x}, \quad \frac{\partial g}{\partial y} = \frac{dg}{df} \frac{\partial f}{\partial y},$$

$$\left( \frac{\partial g}{\partial x} \frac{\partial g}{\partial y} \right) = \left( \frac{dg}{df} \right) \left( \frac{\partial f}{\partial x} \frac{\partial f}{\partial y} \right).$$

▶ For $f : \mathbb{R}^2 \to \mathbb{R}^2$ and $g : \mathbb{R}^2 \to \mathbb{R}$, write

$$f(r, s) = (x(r, s), y(r, s)).$$

$$\frac{\partial g}{\partial r} = \frac{\partial g}{\partial x} \frac{\partial x}{\partial r} + \frac{\partial g}{\partial y} \frac{\partial y}{\partial r}, \quad \frac{\partial g}{\partial s} = \frac{\partial g}{\partial x} \frac{\partial x}{\partial s} + \frac{\partial g}{\partial y} \frac{\partial y}{\partial s},$$

$$\left( \frac{\partial g}{\partial r} \frac{\partial g}{\partial s} \right) = \left( \frac{\partial g}{\partial x} \frac{\partial g}{\partial y} \right) \left( \frac{\partial x}{\partial r} \frac{\partial x}{\partial s} \frac{\partial y}{\partial r} \frac{\partial y}{\partial s} \right).$$

Examples

▶ What is the derivative of $\sin(x^2)$?

▶ If the height of a mountain is $h(x, y) = 5 - x^2 - 4y^2$ and the temperature varies with altitude as $T(h) = h^2$, what is $\partial T/\partial x$ at $(1, 1)$?

▶ If the height of a mountain is $h(x, y) = 5 - x^2 - 4y^2$, what is the change in the height if you are at $(1, 1)$ and move directly away from the peak at $(0, 0)$?
Some special cases

For $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$, let $s = f(x)$.

$$(g \circ f)'(x) = g'(f(x))f'(x)$$

$$
\left( \frac{dg}{dx} \right) = \left( \frac{dg}{df} \right) \left( \frac{df}{dx} \right).
$$

For $f : \mathbb{R}^2 \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$,

$$
\frac{\partial g}{\partial x} = \frac{dg}{df} \frac{\partial f}{\partial x} \quad \frac{\partial g}{\partial y} = \frac{dg}{df} \frac{\partial f}{\partial y}
$$

$$
\left( \frac{\partial g}{\partial x} \frac{\partial g}{\partial y} \right) = \left( \frac{dg}{df} \right) \left( \frac{\partial f}{\partial x} \frac{\partial f}{\partial y} \right).
$$

For $f : \mathbb{R}^2 \to \mathbb{R}^2$ and $g : \mathbb{R}^2 \to \mathbb{R}$, write

$$
f(r, s) = (x(r, s), y(r, s)).
$$

$$
\frac{\partial g}{\partial r} = \frac{\partial g}{\partial x} \frac{\partial x}{\partial r} + \frac{\partial g}{\partial y} \frac{\partial y}{\partial r} \quad \frac{\partial g}{\partial s} = \frac{\partial g}{\partial x} \frac{\partial x}{\partial s} + \frac{\partial g}{\partial y} \frac{\partial y}{\partial s}
$$

$$
\left( \frac{\partial g}{\partial x} \frac{\partial g}{\partial y} \right) \left( \frac{\partial x}{\partial r} \frac{\partial x}{\partial s} \frac{\partial y}{\partial r} \frac{\partial y}{\partial s} \right)
$$

Examples

What is the derivative of $\sin(x^2)$?

If the height of a mountain is $h(x, y) = 5 - x^2 - 4y^2$ and the temperature varies with altitude as $T(h) = h^2$, what is $\partial T / \partial x$ at $(1, 1)$?

If the height of a mountain is $h(x, y) = 5 - x^2 - 4y^2$, what is the change in the height if you are at $(1, 1)$ and move directly away from the peak at $(0, 0)$?
On the chain rule proof

Theorem
For $F : \mathbb{R}^n \to \mathbb{R}^m$ and $G : \mathbb{R}^m \to \mathbb{R}^p$ differentiable functions,
\[
D(G \circ F)(\vec{a}) = DG(F(\vec{a})) \circ DF(\vec{a})
\]
or, in terms of the Jacobian matrices,
\[
[D(G \circ F)(\vec{a})] = [DG(F(\vec{a}))][DF(\vec{a})].
\]

Proof idea.
This boils down to using approximations. If $\vec{a} \in \mathbb{R}^n$ and $\vec{b} = F(\vec{a})$ and have
\[
F(\vec{a} + \vec{h}) \approx F(\vec{a}) + L(\vec{h}) \quad G(\vec{b} + \vec{k}) \approx G(\vec{b}) + M(\vec{k})
\]
for linear $L, M$, then
\[
G(F(\vec{a} + \vec{h})) \approx G(F(\vec{a}) + L(\vec{h}))
\]
\[
\approx G(F(\vec{a})) + M(L(\vec{h}))
\]
\[
= (G \circ F)(\vec{a}) + (M \circ L)(\vec{h}).
\]
(by setting $\vec{k} = L(\vec{h})$).