Curves in $\mathbb{R}^2$: Graphs vs Level Sets

Graphs ($y = f(x)$): The **graph** of $f\colon \mathbb{R} \to \mathbb{R}$ is

$$\{(x, y) \in \mathbb{R}^2 \mid y = f(x)\}.$$

Example: When we say "the curve $y = x^2$," we really mean: "The graph of the function $f(x) = x^2$." That is, we mean the set $\{(x, y) \in \mathbb{R}^2 \mid y = x^2\}$.

Level Sets ($F(x, y) = c$): The **level set** of $F\colon \mathbb{R}^2 \to \mathbb{R}$ at height c is

$$\{(x, y) \in \mathbb{R}^2 \mid F(x, y) = c\}.$$

Example: When we say "the curve $x^2 + y^2 = 1$," we really mean: "The level set of the function $F(x, y) = x^2 + y^2$ at height 1." That is, we mean the set $\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$.

Note: Every graph is a level set (why?). But not every level set is a graph. Graphs must pass the vertical line test. (Level sets may or may not.)

Surfaces in $\mathbb{R}^3$: Graphs vs Level Sets

Graphs ($z = f(x, y)$): The **graph** of $f\colon \mathbb{R}^2 \to \mathbb{R}$ is

$$\{(x, y, z) \in \mathbb{R}^3 \mid z = f(x, y)\}.$$

Example: When we say "the surface $z = x^2 + y^2$," we really mean: "The graph of the function $f(x, y) = x^2 + y^2$." That is, we mean the set $\{(x, y, z) \in \mathbb{R}^3 \mid z = x^2 + y^2\}$.

Level Sets ($F(x, y, z) = c$): The **level set** of $F\colon \mathbb{R}^3 \to \mathbb{R}$ at height c is

$$\{(x, y, z) \in \mathbb{R}^3 \mid F(x, y, z) = c\}.$$

Example: When we say "the surface $x^2 + y^2 + z^2 = 1$," we really mean: "The level set of the function $F(x, y, z) = x^2 + y^2 + z^2$ at height 1." That is, $\{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1\}$.

Again: Every graph is a level set (why?). But not every level set is a graph. Graphs must pass the vertical line test. (Level sets may or may not.)

Q: Do you see the patterns here? For example, suppose $G\colon \mathbb{R}^7 \to \mathbb{R}$.

◦ What does the graph of G look like? (A: 7-dimensional object in $\mathbb{R}^8$. That is, $\{(x_1, \ldots, x_7, x_8) \in \mathbb{R}^8 \mid x_8 = G(x_1, \ldots, x_7)\}$.)

◦ What do the level sets of G look like? (A: They are (generically) 6-dim objects in $\mathbb{R}^7$. That is, $\{(x_1, \ldots, x_7) \in \mathbb{R}^7 \mid G(x_1, \ldots, x_7) = c\}$.)

Curves in $\mathbb{R}^2$: Examples of Graphs

$\{(x, y) \in \mathbb{R}^2 \mid y = ax + b\}$: Line (not vertical)
$\{(x, y) \in \mathbb{R}^2 \mid y = ax^2 + bx + c\}$: Parabola

Curves in $\mathbb{R}^2$: Examples of Level Sets

$\{(x, y) \in \mathbb{R}^2 \mid ax + by = c\}$: Line
$\left\{(x, y) \in \mathbb{R}^2 \,\middle|\, \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1\right\}$: Ellipse
$\left\{(x, y) \in \mathbb{R}^2 \,\middle|\, \frac{x^2}{a^2} - \frac{y^2}{b^2} = 1\right\}$: Hyperbola
$\left\{(x, y) \in \mathbb{R}^2 \,\middle|\, -\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1\right\}$: Hyperbola

Surfaces in $\mathbb{R}^3$: Examples of Graphs

$\left\{(x, y, z) \in \mathbb{R}^3 \,\middle|\, z = ax + by\right\}$: Plane (not containing a vertical line)
$\left\{(x, y, z) \in \mathbb{R}^3 \,\middle|\, z = \frac{x^2}{a^2} + \frac{y^2}{b^2}\right\}$: Elliptic Paraboloid
$\left\{(x, y, z) \in \mathbb{R}^3 \,\middle|\, z = -\frac{x^2}{a^2} - \frac{y^2}{b^2}\right\}$: Elliptic Paraboloid
$\left\{(x, y, z) \in \mathbb{R}^3 \,\middle|\, z = \frac{x^2}{a^2} - \frac{y^2}{b^2}\right\}$: Hyperbolic Paraboloid ("saddle")
$\left\{(x, y, z) \in \mathbb{R}^3 \,\middle|\, z = -\frac{x^2}{a^2} + \frac{y^2}{b^2}\right\}$: Hyperbolic Paraboloid ("saddle")

Surfaces in $\mathbb{R}^3$: Examples of Level Sets

$\left\{(x, y, z) \in \mathbb{R}^3 \,\middle|\, ax + by + cz = d\right\}$: Plane
$\left\{(x, y, z) \in \mathbb{R}^3 \,\middle|\, \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1\right\}$: Ellipsoid
$\left\{(x, y, z) \in \mathbb{R}^3 \,\middle|\, \frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1\right\}$: Hyperboloid of 1 Sheet
$\left\{(x, y, z) \in \mathbb{R}^3 \,\middle|\, \frac{x^2}{a^2} - \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1\right\}$: Hyperboloid of 2 Sheets

Two Model Examples

Example 1A (Elliptic Paraboloid): Consider $f\colon \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x,y) = x^2 + y^2.$$

The level sets of f are curves in $\mathbb{R}^2$. Level sets are $\{(x,y) \in \mathbb{R}^2 \mid x^2+y^2 = c\}$.
The graph of f is a surface in $\mathbb{R}^3$. Graph is $\{(x,y,z) \in \mathbb{R}^3 \mid z = x^2+y^2\}$.

Notice that $(0,0,0)$ is a <u>local minimum</u> of f.
Note that $\frac{\partial f}{\partial x}(0,0) = \frac{\partial f}{\partial y}(0,0) = 0$. Also, $\frac{\partial^2 f}{\partial x^2}(0,0) > 0$ and $\frac{\partial^2 f}{\partial y^2}(0,0) > 0$.
Sketch the level sets of f and the graph of f:

Example 1B (Elliptic Paraboloid): Consider $f\colon \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x,y) = -x^2 - y^2.$$

The level sets of f are curves in $\mathbb{R}^2$. Level sets are $\{(x,y) \in \mathbb{R}^2 \mid -x^2-y^2 = c\}$.
The graph of f is a surface in $\mathbb{R}^3$. Graph is $\{(x,y,z) \in \mathbb{R}^3 \mid z = -x^2-y^2\}$.

Notice that $(0,0,0)$ is a <u>local maximum</u> of f.
Note that $\frac{\partial f}{\partial x}(0,0) = \frac{\partial f}{\partial y}(0,0) = 0$. Also, $\frac{\partial^2 f}{\partial x^2}(0,0) < 0$ and $\frac{\partial^2 f}{\partial y^2}(0,0) < 0$.
Sketch the level sets of f and the graph of f:

Example 2 (Hyperbolic Paraboloid): Consider $f\colon \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x,y) = x^2 - y^2.$$

The level sets of f are curves in $\mathbb{R}^2$. Level sets are $\{(x,y) \in \mathbb{R}^2 \mid x^2-y^2 = c\}$.
The graph of f is a surface in $\mathbb{R}^3$. Graph is $\{(x,y,z) \in \mathbb{R}^3 \mid z = x^2-y^2\}$.

Notice that $(0,0,0)$ is a <u>saddle point</u> of the graph of f.
Note that $\frac{\partial f}{\partial x}(0,0) = \frac{\partial f}{\partial y}(0,0) = 0$. Also, $\frac{\partial^2 f}{\partial x^2}(0,0) > 0$ while $\frac{\partial^2 f}{\partial y^2}(0,0) < 0$.
Sketch the level sets of f and the graph of f:

Introduction: Derivatives of Functions $\mathbb{R}^n \to \mathbb{R}^m$

Def: Let $F\colon \mathbb{R}^n \to \mathbb{R}^m$ be a function, say

$$F(x_1, \ldots, x_n) = (F_1(x_1, \ldots, x_n), \ldots, F_m(x_1, \ldots, x_n)).$$

Its **derivative** at the point $(x_1, \ldots, x_n)$ is the linear transformation $DF(x_1, \ldots, x_n)\colon \mathbb{R}^n \to \mathbb{R}^m$ whose $(m \times n)$ matrix is

$$DF(x_1, \ldots, x_n) = \begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial F_m}{\partial x_1} & \cdots & \frac{\partial F_m}{\partial x_n} \end{bmatrix}.$$

Note: The columns are the partial derivatives with respect to x_1, then to x_2, etc. The rows are the gradients of the component functions F^1, F^2, etc.

$$DF(x_1, \ldots, x_n) = \begin{bmatrix} | & & | \\ \frac{\partial F}{\partial x_1} & \cdots & \frac{\partial F}{\partial x_n} \\ | & & | \end{bmatrix} = \begin{bmatrix} \text{---} & \nabla F_1 & \text{---} \\ & \vdots & \\ \text{---} & \nabla F_m & \text{---} \end{bmatrix}.$$

Example: Let $F\colon \mathbb{R}^2 \to \mathbb{R}^3$ be the function

$$F(x, y) = (x + 2y,\ \sin(x),\ e^y) = (F_1(x, y), F_2(x, y), F_3(x, y)).$$

Its **derivative** at (x, y) is a linear transformation $DF(x, y)\colon \mathbb{R}^2 \to \mathbb{R}^3$. The matrix of the linear transformation $DF(x, y)$ is:

$$DF(x, y) = \begin{bmatrix} \frac{\partial F_1}{\partial x} & \frac{\partial F_1}{\partial y} \\ \frac{\partial F_2}{\partial x} & \frac{\partial F_2}{\partial y} \\ \frac{\partial F_3}{\partial x} & \frac{\partial F_3}{\partial y} \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ \cos(x) & 0 \\ 0 & e^y \end{bmatrix}.$$

Notice that (for instance) $DF(1, 1)$ is a linear transformation, as is $DF(2, 3)$, etc. That is, each $DF(x, y)$ is a linear transformation $\mathbb{R}^2 \to \mathbb{R}^3$.

Goals: We will:

- ∘ Interpret the derivative of F as the "best linear approximation to F."
- ∘ State a Chain Rule for multivariable derivatives.

Introduction: Gradient of Functions $\mathbb{R}^n \to \mathbb{R}$

Def: Let $f \colon \mathbb{R}^n \to \mathbb{R}$ be a function.

Recall: The derivative of $f \colon \mathbb{R}^n \to \mathbb{R}$ at the point $\mathbf{x} = (x_1, \ldots, x_n)$ is the $1 \times n$ matrix

$$Df(\mathbf{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}.$$

The **gradient** of $f \colon \mathbb{R}^n \to \mathbb{R}$ at the point $\mathbf{x} = (x_1, \ldots, x_n)$ is the vector

$$\nabla f(\mathbf{x}) = \left[\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right].$$

The **directional derivative** of $f \colon \mathbb{R}^n \to \mathbb{R}$ at the point $\mathbf{x} = (x_1, \ldots, x_n)$ in the direction $\mathbf{v} \in \mathbb{R}^n$ is the dot product of the vectors $\nabla f(\mathbf{x})$ and $\mathbf{v}$:

$$D_{\mathbf{v}} f(\mathbf{x}) = \nabla f(\mathbf{x}) \cdot \mathbf{v}.$$

We will give geometric interpretations of these concepts later in the course.

Introduction: Hessian of Functions $\mathbb{R}^n \to \mathbb{R}$

Theorem: Let $f \colon \mathbb{R}^n \to \mathbb{R}$ be a function whose second partial derivatives are all continuous. Then:

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}.$$

In brief: "Second partials commute."

Def: Let $f \colon \mathbb{R}^n \to \mathbb{R}$ be a function.

The **Hessian** of $f \colon \mathbb{R}^n \to \mathbb{R}$ at the point $\mathbf{x} = (x_1, \ldots, x_n)$ is the $n \times n$ matrix

$$Hf(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

The **directional second derivative** of $f \colon \mathbb{R}^n \to \mathbb{R}$ at the point $\mathbf{x} = (x_1, \ldots, x_n)$ in the direction $\mathbf{v} \in \mathbb{R}^n$ is

$$\mathbf{v}^T \, Hf(\mathbf{x}) \, \mathbf{v}.$$

Again, we will give geometric interpretations of these concepts later on.

Linear Approximation: Single-Variable Calculus

Review: In single-variable calc, we look at functions $f\colon \mathbb{R} \to \mathbb{R}$. We write $y = f(x)$, and at a point $(a, f(a))$ write:

$$\Delta y \approx dy.$$

Here, $\Delta y = f(x) - f(a)$, while $dy = f'(a)\Delta x = f'(a)(x-a)$. So:

$$f(x) - f(a) \approx f'(a)(x-a).$$

Therefore:

$$f(x) \approx f(a) + f'(a)(x-a).$$

The right-hand side $f(a) + f'(a)(x-a)$ can be interpreted as follows:

- It is the **best linear approximation** to $f(x)$ at $x = a$.
- It is the **1st Taylor polynomial** to $f(x)$ at $x = a$.
- The line $y = f(a) + f'(a)(x-a)$ is the **tangent line** at $(a, f(a))$.

Linear Approximation: Multivariable Calculus

Now consider functions $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$. At a point $(\mathbf{a}, \mathbf{f}(\mathbf{a}))$, we have exactly the same thing:

$$\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a}) \approx D\mathbf{f}(\mathbf{a})(\mathbf{x} - \mathbf{a}).$$

That is:

$$\mathbf{f}(\mathbf{x}) \approx \mathbf{f}(\mathbf{a}) + D\mathbf{f}(\mathbf{a})(\mathbf{x} - \mathbf{a}). \tag{$*$}$$

Note: The object $D\mathbf{f}(\mathbf{a})$ is a *matrix*, while $(\mathbf{x} - \mathbf{a})$ is a *vector*. That is, $D\mathbf{f}(\mathbf{a})(\mathbf{x} - \mathbf{a})$ is matrix-vector multiplication.

Example: Let $f\colon \mathbb{R}^2 \to \mathbb{R}$. Let's write $\mathbf{x} = (x_1, x_2)$ and $\mathbf{a} = (a_1, a_2)$. Then $(*)$ reads:

$$\begin{aligned} f(x_1, x_2) &\approx f(a_1, a_2) + \begin{bmatrix} \frac{\partial f}{\partial x_1}(a_1, a_2) & \frac{\partial f}{\partial x_2}(a_1, a_2) \end{bmatrix} \begin{bmatrix} x_1 - a_1 \\ x_2 - a_2 \end{bmatrix} \\ &= f(a_1, a_2) + \frac{\partial f}{\partial x_1}(a_1, a_2)(x_1 - a_1) + \frac{\partial f}{\partial x_2}(a_1, a_2)(x_2 - a_2). \end{aligned}$$

Review: Taylor Polynomials in Single-Variable Calculus

Review: In single-variable calculus, we look at functions $f\colon \mathbb{R} \to \mathbb{R}$.

At a point $a \in \mathbb{R}$, the linear approximation (1st-deg Taylor polynomial) to f is:

$$f(x) \approx f(a) + f'(a)(x-a).$$

More accurate is the quadratic approximation (2nd-deg Taylor polynomial)

$$f(x) \approx f(a) + f'(a)(x-a) + \frac{1}{2!}f''(a)(x-a)^2.$$

We would like to have similar ideas for multivariable functions.

Linear Approximation: 1st-Deg Taylor Polynomials

Let $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$. The **linear approximation** of $\mathbf{f}$ at the point $\mathbf{a} \in \mathbb{R}^n$ is:

$$\mathbf{f}(\mathbf{x}) \approx \mathbf{f}(\mathbf{a}) + D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a}). \tag{1}$$

Note that $D\mathbf{f}(\mathbf{a})$ is a *matrix*, while $(\mathbf{x}-\mathbf{a})$ is a *vector*. That is, $D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a})$ is matrix-vector multiplication.

Note that (1) is the best linear approximation to $\mathbf{f}(\mathbf{x})$ for points $\mathbf{x}$ near $\mathbf{a}$. It is the 1st-degree Taylor polynomial to $\mathbf{f}(\mathbf{x})$ at $\mathbf{a}$.

Quadratic Approximation: 2nd-Deg Taylor Polynomials

Let $f\colon \mathbb{R}^n \to \mathbb{R}$. The **quadratic approximation** of $\mathbf{f}$ at the point $\mathbf{a} \in \mathbb{R}^n$ is:

$$f(\mathbf{x}) \approx f(\mathbf{a}) + Df(\mathbf{a})(\mathbf{x}-\mathbf{a}) + \frac{1}{2!}(\mathbf{x}-\mathbf{a})^T Hf(\mathbf{a})(\mathbf{x}-\mathbf{a}). \tag{2}$$

Note that $(\mathbf{x}-\mathbf{a})^T$ is a row vector, while $(\mathbf{x}-\mathbf{a})$ is a column vector, while $Hf(\mathbf{a})$ is a matrix. So, $\frac{1}{2!}(\mathbf{x}-\mathbf{a})^T Hf(\mathbf{a})(\mathbf{x}-\mathbf{a})$ is of the form $\mathbf{v}^T A\mathbf{v}$.

Note that (2) is the best quadratic approximation to $f(\mathbf{x})$ for points $\mathbf{x}$ near $\mathbf{a}$. It is the 2nd-degree Taylor polynomial to $f(\mathbf{x})$ at $\mathbf{a}$.

Example: Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be $f(x,y) = x^3\sin(y)$. For (x,y) near $\mathbf{a} = (2, \frac{\pi}{2})$:

$$\begin{aligned}
f(x,y) &\approx f(2,\tfrac{\pi}{2}) + Df(2,\tfrac{\pi}{2})\begin{bmatrix} x-2 \\ y-\frac{\pi}{2}\end{bmatrix} + \frac{1}{2!}\begin{bmatrix} x-2 & y-\frac{\pi}{2}\end{bmatrix} Hf(2,\tfrac{\pi}{2})\begin{bmatrix} x-2 \\ y-\frac{\pi}{2}\end{bmatrix} \\
&= 8 + \begin{bmatrix} 12 & 0\end{bmatrix}\begin{bmatrix} x-2 \\ y-\frac{\pi}{2}\end{bmatrix} + \frac{1}{2}\begin{bmatrix} x-2 & y-\frac{\pi}{2}\end{bmatrix}\begin{bmatrix} 12 & 0 \\ 0 & -8\end{bmatrix}\begin{bmatrix} x-2 \\ y-\frac{\pi}{2}\end{bmatrix} \\
&= 8 + 12(x-2) + 6(x-2)^2 - 4(y-\tfrac{\pi}{2})^2.
\end{aligned}$$

Tangent Lines/Planes to Graphs

Fact: Suppose a curve in $\mathbb{R}^2$ is given as a graph $y = f(x)$. The equation of the tangent line at $(a, f(a))$ is:

$$y = f(a) + f'(a)(x - a).$$

Okay, you knew this from single-variable calculus. How does the multivariable case work? Well:

Fact: Suppose a surface in $\mathbb{R}^3$ is given as a graph $z = f(x, y)$. The equation of the tangent plane at $(a, b, f(a, b))$ is:

$$z = f(a, b) + \frac{\partial f}{\partial x}(a, b)(x - a) + \frac{\partial f}{\partial y}(a, b)(y - b).$$

Notice the similarity between this and the linear approximation to f at (a, b).

Tangent Lines/Planes to Level Sets

Def: For a function $F\colon \mathbb{R}^n \to \mathbb{R}$, its **gradient** is the vector in $\mathbb{R}^n$ given by:

$$\nabla F = \left[\frac{\partial F}{\partial x_1}, \frac{\partial F}{\partial x_2}, \ldots, \frac{\partial F}{\partial x_n}\right].$$

Theorem: Consider a level set $F(x_1, \ldots, x_n) = c$ of a function $F\colon \mathbb{R}^n \to \mathbb{R}$. If $(a_1, \ldots, a_n)$ is a point on the level set, then $\nabla F(a_1, \ldots, a_n)$ is normal to the level set.

Corollary 1: Suppose a curve in $\mathbb{R}^2$ is given as a level curve $F(x, y) = c$. The equation of the tangent line at a point (x_0, y_0) on the level curve is:

$$\frac{\partial F}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial F}{\partial y}(x_0, y_0)(y - y_0) = 0.$$

Corollary 2: Suppose a surface in $\mathbb{R}^3$ is given as a level surface $F(x, y, z) = c$. The equation of the tangent plane at a point (x_0, y_0, z_0) on the level surface is:

$$\frac{\partial F}{\partial x}(x_0, y_0, z_0)(x - x_0) + \frac{\partial F}{\partial y}(x_0, y_0, z_0)(y - y_0) + \frac{\partial F}{\partial z}(x_0, y_0, z_0)(z - z_0) = 0.$$

Q: Do you see why Cor 1 and Cor 2 follow from the Theorem?

Composition and Matrix Multiplication

Recall: Let $f\colon X \to Y$ and $g\colon Y \to Z$ be functions. Their **composition** is the function $g \circ f\colon X \to Z$ defined by

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

Observations:

(1) For this to make sense, we must have: co-domain(f) = domain(g).

(2) Composition is **not** generally commutative: that is, $f \circ g$ and $g \circ f$ are usually different.

(3) Composition is always associative: $(h \circ g) \circ f = h \circ (g \circ f)$.

Fact: If $T\colon \mathbb{R}^k \to \mathbb{R}^n$ and $S\colon \mathbb{R}^n \to \mathbb{R}^m$ are both linear transformations, then $S \circ T$ is also a linear transformation.

Question: How can we describe the matrix of the linear transformation $S \circ T$ in terms of the matrices of S and T?

Fact: Let $T\colon \mathbb{R}^n \to \mathbb{R}^n$ and $S\colon \mathbb{R}^n \to \mathbb{R}^m$ be linear transformations with matrices B and A, respectively. Then the matrix of $S \circ T$ is the product AB.

We can multiply an $m \times n$ matrix A by an $n \times k$ matrix B. The result, AB, will be an $m \times k$ matrix:

$$(m \times n)(n \times k) \to (m \times k).$$

Notice that n appears twice here to "cancel out." That is, we need the number of rows of A to equal the number of columns of B – otherwise, the product AB makes no sense.

Example 1: Let A be a (3×2)-matrix, and let B be a (2×4)-matrix. The product AB is then a (3×4)-matrix.

Example 2: Let A be a (2×3)-matrix, and let B be a (4×2)-matrix. Then AB is not defined. (But the product BA is defined: it is a (4×3)-matrix.)

Chain Rule

Chain Rule (Matrix Form): Let $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$ and $\mathbf{g}\colon \mathbb{R}^m \to \mathbb{R}^p$ be any differentiable functions. Then

$$D(\mathbf{g} \circ \mathbf{f})(\mathbf{x}) = D\mathbf{g}(\mathbf{f}(\mathbf{x})) \cdot D\mathbf{f}(\mathbf{x}).$$

Here, the product on the right-hand side is a product of matrices.

Many texts describe the chain rule in the following more classical form. While there is a "classical" form in the general case of functions $\mathbf{g}\colon \mathbb{R}^m \to \mathbb{R}^p$, we will keep things simple and only state the case of functions $g\colon \mathbb{R}^m \to \mathbb{R}$ with codomain $\mathbb{R}$.

Chain Rule (Classical Form): Let $g = g(x_1, \ldots, x_m)$ and suppose each $x_1, \ldots, x_m$ is a function of the variables $t_1, \ldots, t_n$. Then:

$$\begin{aligned}
\frac{\partial g}{\partial t_1} &= \frac{\partial g}{\partial x_1}\frac{\partial x_1}{\partial t_1} + \frac{\partial g}{\partial x_2}\frac{\partial x_2}{\partial t_1} + \cdots + \frac{\partial g}{\partial x_m}\frac{\partial x_m}{\partial t_1}, \\
&\vdots \\
\frac{\partial g}{\partial t_n} &= \frac{\partial g}{\partial x_1}\frac{\partial x_1}{\partial t_n} + \frac{\partial g}{\partial x_2}\frac{\partial x_2}{\partial t_1} + \cdots + \frac{\partial g}{\partial x_m}\frac{\partial x_m}{\partial t_n}.
\end{aligned}$$

Example 1: Let $z = g(u, v)$, where $u = h(x, y)$ and $v = k(x, y)$. Then the chain rule reads:

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial z}{\partial v}\frac{\partial v}{\partial x}$$

and

$$\frac{\partial z}{\partial y} = \frac{\partial z}{\partial u}\frac{\partial u}{\partial y} + \frac{\partial z}{\partial v}\frac{\partial v}{\partial y}.$$

Example 2: Let $z = g(u, v, w)$, where $u = h(t)$, $v = k(t)$, $w = \ell(t)$. Then the chain rule reads:

$$\frac{\partial z}{\partial t} = \frac{\partial z}{\partial u}\frac{\partial u}{\partial t} + \frac{\partial z}{\partial v}\frac{\partial v}{\partial t} + \frac{\partial z}{\partial w}\frac{\partial w}{\partial t}.$$

Since u, v, w are functions of just a single variable t, we can also write this formula as:

$$\frac{\partial z}{\partial t} = \frac{\partial z}{\partial u}\frac{du}{dt} + \frac{\partial z}{\partial v}\frac{dv}{dt} + \frac{\partial z}{\partial w}\frac{dw}{dt}.$$

For Clarification: The Two Forms of the Chain Rule

Q: How exactly are the two forms of the chain rule the same?
A: If we completely expand the matrix form, writing out everything in components, we end up with the classical form. The following examples may clarify.

Example 1: Suppose $z = g(u, v)$, where $u = h(x, y)$ and $v = k(x, y)$.
This setup means we essentially have two functions:

$$g\colon \mathbb{R}^2 \to \mathbb{R} \qquad \text{and} \qquad f\colon \mathbb{R}^2 \to \mathbb{R}^2$$
$$g(u, v) = z \qquad\qquad f(x, y) = (h(x, y), k(x, y)) = (u, v).$$

The matrix form of the Chain Rule reads:

$$D(g \circ f)(x, y) = Dg(f(x, y)) \cdot Df(x, y)$$
$$\begin{bmatrix} \frac{\partial(g\circ f)}{\partial x}, & \frac{\partial(g\circ f)}{\partial y} \end{bmatrix} = \begin{bmatrix} \frac{\partial g}{\partial u}, & \frac{\partial g}{\partial v} \end{bmatrix} \begin{bmatrix} \frac{\partial h}{\partial x} & \frac{\partial h}{\partial y} \\ \frac{\partial k}{\partial x} & \frac{\partial k}{\partial y} \end{bmatrix}$$
$$\begin{bmatrix} \frac{\partial z}{\partial x}, & \frac{\partial z}{\partial y} \end{bmatrix} = \begin{bmatrix} \frac{\partial g}{\partial u}\frac{\partial h}{\partial x} + \frac{\partial g}{\partial v}\frac{\partial k}{\partial x}, & \frac{\partial g}{\partial u}\frac{\partial h}{\partial y} + \frac{\partial g}{\partial v}\frac{\partial k}{\partial y} \end{bmatrix}$$

Setting components equal to each other, we conclude that

$$\frac{\partial z}{\partial x} = \frac{\partial g}{\partial u}\frac{\partial h}{\partial x} + \frac{\partial g}{\partial v}\frac{\partial k}{\partial x}$$

and

$$\frac{\partial z}{\partial y} = \frac{\partial g}{\partial u}\frac{\partial h}{\partial y} + \frac{\partial g}{\partial v}\frac{\partial k}{\partial y}.$$

This is exactly the classical form of the Chain Rule. $\square$

Example 2: Suppose $z = g(u, v, w)$, where $u = h(t)$, $v = k(t)$, $w = \ell(t)$.
This setup means we essentially have two functions:

$$g\colon \mathbb{R}^3 \to \mathbb{R} \qquad \text{and} \qquad f\colon \mathbb{R} \to \mathbb{R}^3$$
$$g(u, v, w) = z \qquad\qquad f(t) = (h(t), k(t), \ell(t)) = (u, v, w).$$

The matrix form of the Chain Rule reads:

$$D(g \circ f)(t) = Dg(f(t)) \cdot Df(t)$$
$$\begin{bmatrix} \frac{\partial(g\circ f)}{\partial t} \end{bmatrix} = \begin{bmatrix} \frac{\partial g}{\partial u}, & \frac{\partial g}{\partial v}, & \frac{\partial g}{\partial w} \end{bmatrix} \begin{bmatrix} \frac{\partial h}{\partial t} \\ \frac{\partial k}{\partial t} \\ \frac{\partial \ell}{\partial t} \end{bmatrix}$$
$$\frac{\partial z}{\partial t} = \frac{\partial g}{\partial u}\frac{\partial h}{\partial t} + \frac{\partial g}{\partial v}\frac{\partial k}{\partial t} + \frac{\partial g}{\partial w}\frac{\partial \ell}{\partial t}.$$

Again, we recovered the classical form of the Chain Rule. $\square$

Inverses: Abstract Theory

Def: A function $f\colon X \to Y$ is **invertible** if there is a function $f^{-1}\colon Y \to X$ satisfying:

$$\begin{aligned} f^{-1}(f(x)) &= x, \quad \text{for all } x \in X, \text{ and} \\ f(f^{-1}(y)) &= y, \quad \text{for all } y \in Y. \end{aligned}$$

In such a case, f^{-1} is called an **inverse function** for f.

In other words, the function f^{-1} "undoes" the function f. For example, an inverse function of $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = x^3$ is $f^{-1}\colon \mathbb{R} \to \mathbb{R}$, $f^{-1}(x) = \sqrt[3]{x}$. An inverse of $g\colon \mathbb{R} \to (0,\infty)$, $g(x) = 2^x$ is $g^{-1}\colon (0,\infty) \to \mathbb{R}$, $g^{-1}(x) = \log_2(x)$.

Whenever a new concept is defined, a mathematician asks two questions:

(1) **Uniqueness:** Are inverses unique? That is, must a function f have at most one inverse f^{-1}, or is it possible for f to have several different inverses?

Answer: Yes.

Prop 16.1: If $f\colon X \to Y$ is invertible (that is, f has an inverse), then the inverse function f^{-1} is unique (that is, there is only one inverse function).

(2) **Existence:** Do inverses always exist? That is, does every function f have an inverse function f^{-1}?

Answer: No. Some functions have inverses, but others don't.

New question: Which functions have inverses?

Prop 16.3: A function $f\colon X \to Y$ is invertible if and only if f is both "one-to-one" and "onto."

Despite their fundamental importance, there's no time to talk about "one-to-one" and "onto," so you don't have to learn these terms. This is sad :-(

Question: If inverse functions "undo" our original functions, can they help us solve equations? Yes! That's the entire point:

Prop 16.2: A function $f\colon X \to Y$ is invertible if and only if for every $b \in Y$, the equation $f(x) = b$ has exactly one solution $x \in X$.

In this case, the solution to the equation $f(x) = b$ is given by $x = f^{-1}(b)$.

Inverses of Linear Transformations

Question: Which linear transformations $T\colon \mathbb{R}^n \to \mathbb{R}^m$ are invertible? (Equiv: Which $m \times n$ matrices A are invertible?)

Fact: If $T\colon \mathbb{R}^n \to \mathbb{R}^m$ is invertible, then $m = n$.
So: If an $m \times n$ matrix A is invertible, then $m = n$.

In other words, non-square matrices are never invertible. But square matrices may or may not be invertible. Which ones are invertible? Well:

Theorem: Let A be an $n \times n$ matrix. The following are equivalent:
(i) A is invertible
(ii) $N(A) = \{\mathbf{0}\}$
(iii) $C(A) = \mathbb{R}^n$
(iv) $\text{rref}(A) = I_n$
(v) $\det(A) \neq 0$.

To Repeat: An $n \times n$ matrix A is invertible if and only if for every $\mathbf{b} \in \mathbb{R}^n$, the equation $A\mathbf{x} = \mathbf{b}$ has exactly one solution $\mathbf{x} \in \mathbb{R}^n$.
In this case, the solution to the equation $A\mathbf{x} = \mathbf{b}$ is given by $\mathbf{x} = A^{-1}\mathbf{b}$.

Q: How can we find inverse matrices? This is accomplished via:

Prop 16.7: If A is an invertible matrix, then $\text{rref}[A \mid I_n] = [I_n \mid A^{-1}]$.

Useful Formula: Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be a 2×2 matrix. If A is invertible ($\det(A) = ad - bc \neq 0$), then:

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Prop 16.8: Let $f\colon X \to Y$ and $g\colon Y \to Z$ be invertible functions. Then:
(a) f^{-1} is invertible and $(f^{-1})^{-1} = f$.
(b) $g \circ f$ is invertible and $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

Corollary: Let A, B be invertible $n \times n$ matrices. Then:
(a) A^{-1} is invertible and $(A^{-1})^{-1} = A$.
(b) AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$.

The Gradient: Two Interpretations

Recall: For a function $F\colon \mathbb{R}^n \to \mathbb{R}$, its **gradient** is the vector in $\mathbb{R}^n$ given by:

$$\nabla F = \left[\frac{\partial F}{\partial x_1}, \frac{\partial F}{\partial x_2}, \ldots, \frac{\partial F}{\partial x_n}\right].$$

There are two ways to think about the gradient. They are interrelated.

Gradient: Normal to Level Sets

Theorem: Consider a level set $F(x_1, \ldots, x_n) = c$ of a function $F\colon \mathbb{R}^n \to \mathbb{R}$. If $(a_1, \ldots, a_n)$ is a point on the level set, then $\nabla F(a_1, \ldots, a_n)$ is normal to the level set.

Example: If we have a level curve $F(x, y) = c$ in $\mathbb{R}^2$, the gradient vector $\nabla F(x_0, y_0)$ is a normal vector to the level curve at the point (x_0, y_0).

Example: If we have a level surface $F(x, y, z) = c$ in $\mathbb{R}^3$, the gradient vector $\nabla F(x_0, y_0, z_0)$ is a normal vector to the level surface at the point (x_0, y_0, z_0).

Normal vectors help us find tangent planes to level sets (see the handout "Tangent Lines/Planes...") But there's another reason we like normal vectors.

Gradient: Direction of Steepest Ascent for Graphs

Observation: A normal vector to a level set $F(x_1, \ldots, x_n) = c$ in $\mathbb{R}^n$ is the direction of steepest ascent for the graph $z = F(x_1, \ldots, x_n)$ in $\mathbb{R}^{n+1}$.

Example (Elliptic Paraboloid): Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be $f(x, y) = 2x^2 + 3y^2$.
The level sets of f are the ellipses $2x^2 + 3y^2 = c$ in $\mathbb{R}^2$.
The graph of f is the elliptic paraboloid $z = 2x^2 + 3y^2$ in $\mathbb{R}^3$.

At the point $(1, 1) \in \mathbb{R}^2$, the gradient vector $\nabla f(1, 1) = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$ is normal to the level curve $2x^2 + 3y^2 = 5$. So, if we were hiking on the surface $z = 2x^2 + 3y^2$ in $\mathbb{R}^3$ and were at the point $(1, 1, f(1, 1)) = (1, 1, 5)$, to ascend the surface the fastest, we would hike in the direction of $\begin{bmatrix} 4 \\ 6 \end{bmatrix}$. $\square$

Warning: Note that ∇f is normal to the level sets of f. It is <u>*not*</u> a normal vector to the graph of f.

Directional Derivatives

Def: For a function $f\colon \mathbb{R}^n \to \mathbb{R}$, its **directional derivative** at the point $\mathbf{x} \in \mathbb{R}^n$ in the direction $\mathbf{v} \in \mathbb{R}^n$ is:

$$\boxed{D_{\mathbf{v}} f(\mathbf{x}) = \nabla f(\mathbf{x}) \cdot \mathbf{v}.}$$

Here, $\cdot$ is the dot product of vectors. Therefore,

$$D_{\mathbf{v}} f(\mathbf{x}) = \|\nabla f(\mathbf{x})\| \|\mathbf{v}\| \cos\theta, \qquad \text{where } \theta = \measuredangle(\nabla f(\mathbf{x}), \mathbf{v}).$$

Usually, we assume that $\mathbf{v}$ is a unit vector, meaning $\|\mathbf{v}\| = 1$.

Example: Let $f\colon \mathbb{R}^2 \to \mathbb{R}$. Let $\mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix}$. Then:

$$D_{\mathbf{v}} f(x,y) = \nabla f(x,y) \cdot \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} = a\frac{\partial f}{\partial x} + b\frac{\partial f}{\partial y}.$$

In particular, we have two important special cases:

$$D_{\mathbf{e}_1} f(x,y) = \nabla f(x,y) \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{\partial f}{\partial x}$$

$$D_{\mathbf{e}_2} f(x,y) = \nabla f(x,y) \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \frac{\partial f}{\partial y}.$$

Point: Partial derivatives are themselves examples of directional derivatives!

Namely, $\frac{\partial f}{\partial x}$ is the directional derivative of f in the $\mathbf{e}_1$-direction, while $\frac{\partial f}{\partial y}$ is the directional derivative in the $\mathbf{e}_2$-direction.

Question: At a point $\mathbf{a}$, in which direction $\mathbf{v}$ will the function f grow the most? i.e.: At a given point $\mathbf{a}$, for which unit vector $\mathbf{v}$ is $D_{\mathbf{v}} f(\mathbf{a})$ maximized?

Theorem 6.3: Fix a point $\mathbf{a} \in \mathbb{R}^n$.

(a) The directional derivative $D_{\mathbf{v}} f(\mathbf{a})$ is <u>maximized</u> when $\mathbf{v}$ points in the same direction as $\nabla f(\mathbf{a})$.

(b) The directional derivative $D_{\mathbf{v}} f(\mathbf{a})$ is <u>minimized</u> when $\mathbf{v}$ points in the opposite direction as $\nabla f(\mathbf{a})$.

In fact: The maximum and minimum values of $D_{\mathbf{v}} f(\mathbf{a})$ at the point $\mathbf{a} \in \mathbb{R}^n$ are $\|\nabla f(\mathbf{a})\|$ and $-\|\nabla f(\mathbf{a})\|$. (Assuming we only care about unit vectors $\mathbf{v}$.)

Determinants

There are two reasons why determinants are important:

(1) Algebra: Determinants tell us whether a matrix is invertible or not.

(2) Geometry: Determinants are related to area and volume.

Determinants: Algebra

Prop 17.3: An $n \times n$ matrix A is invertible $\iff$ $\det(A) \neq 0$.

Moreover: if A is invertible, then

$$\det(A^{-1}) = \frac{1}{\det(A)}.$$

Properties of Determinants (17.2, 17.4):

(1) (Multiplicativity) $\det(AB) = \det(A)\det(B)$.

(2) (Alternation) Exchanging two rows of a matrix reverses the sign of the determinant.

(3) (Multilinearity): First:

$$\det\begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix} + \det\begin{bmatrix} b_1 & b_2 & \cdots & b_n \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix} = \det\begin{bmatrix} a_1+b_1 & a_2+b_2 & \cdots & a_n+b_n \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix}$$

and similarly for the other rows; Second:

$$\det\begin{bmatrix} ka_{11} & ka_{12} & \cdots & ka_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = k\det\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

and similarly for the other rows. Here, $k \in \mathbb{R}$ is any scalar.

Warning! Multilinearity does <u>*not*</u> say that $\det(A+B) = \det(A) + \det(B)$. It also does <u>*not*</u> say $\det(kA) = k\det(A)$. But: $\det(kA) = k^n\det(A)$ is true.

Determinants: Geometry

Prop 17.5: Let A be any 2×2 matrix. Then the area of the parallelogram generated by the columns of A is $|\det(A)|$.

Prop 17.6: Let $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation with matrix A. Let R be a region in $\mathbb{R}^2$. Then:

$$\text{Area}(T(R)) = |\det(A)| \cdot \text{Area}(R).$$

Coordinate Systems

Def: Let V be a k-dim subspace of $\mathbb{R}^n$. Each basis $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ determines a **coordinate system** on V.

That is: Every vector $\mathbf{v} \in V$ can be written uniquely as a linear combination of the basis vectors:

$$\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k.$$

We then call $c_1, \ldots, c_k$ the **coordinates** of $\mathbf{v}$ with respect to the basis $\mathcal{B}$. We then write

$$[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix}.$$

Note that $[\mathbf{v}]_{\mathcal{B}}$ has k components, even though $\mathbf{v} \in \mathbb{R}^n$.

Note: Levandosky (L21: p 145-149) explains all this very clearly, in much more depth than this review sheet provides. The examples are also quite good: make sure you understand all of them.

Def: Let $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a basis for a k-dim subspace V of $\mathbb{R}^n$. The **change-of-basis matrix** for the basis $\mathcal{B}$ is:

$$C = \begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_k \\ | & | & & | \end{bmatrix}.$$

Every vector $\mathbf{v} \in V$ in the subspace V can be written

$$\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k.$$

In other words:

$$\mathbf{v} = C[\mathbf{v}]_{\mathcal{B}}.$$

This formula tells us how to go between the standard coordinates for $\mathbf{v}$ and the $\mathcal{B}$-coordinates of $\mathbf{v}$.

Special Case: If $V = \mathbb{R}^n$ and $\mathcal{B}$ is a basis of $\mathbb{R}^n$, then the matrix C will be invertible, and therefore:

$$[\mathbf{v}]_{\mathcal{B}} = C^{-1}\mathbf{v}.$$