Lec 1: Mathematics Review

15-369/669/769: Numerical Computing

Instructor: Minchen Li

Table of Content

- Vector Spaces
- Linearity
- Nonlinearity: Differential Calculus

Table of Content

- Vector Spaces
- Linearity
- Nonlinearity: Differential Calculus

Vector Spaces

Definition

A member $\vec{v} \in \mathcal{V}$ is known as a *vector*; arrows will be used to indicate vector variables.

Definition 1.1 (Vector space over $\mathbb{R}$)**.** A *vector space* over $\mathbb{R}$ is a set $\mathcal{V}$ closed under addition and scalar multiplication satisfying the following axioms:

- *Additive commutativity and associativity:* For all $\vec{u}, \vec{v}, \vec{w} \in \mathcal{V}$, $\vec{v} + \vec{w} = \vec{w} + \vec{v}$ and $(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$.
- *Distributivity:* For all $\vec{v}, \vec{w} \in \mathcal{V}$ and $a, b \in \mathbb{R}$, $a(\vec{v}+\vec{w}) = a\vec{v}+a\vec{w}$ and $(a+b)\vec{v} = a\vec{v}+b\vec{v}$.
- *Additive identity:* There exists $\vec{0} \in \mathcal{V}$ with $\vec{0} + \vec{v} = \vec{v}$ for all $\vec{v} \in \mathcal{V}$.
- *Additive inverse:* For all $\vec{v} \in \mathcal{V}$, there exists $\vec{w} \in \mathcal{V}$ with $\vec{v} + \vec{w} = \vec{0}$.
- *Multiplicative identity:* For all $\vec{v} \in \mathcal{V}$, $1 \cdot \vec{v} = \vec{v}$.
- *Multiplicative compatibility:* For all $\vec{v} \in \mathcal{V}$ and $a, b \in \mathbb{R}$, $(ab)\vec{v} = a(b\vec{v})$.

Vector Spaces

Examples

Example 1.1 ($\mathbb{R}^n$ as a vector space)**.** The most common example of a vector space is $\mathbb{R}^n$. Here, addition and scalar multiplication happen component-by-component:

$$\begin{aligned}(1,2)+(-3,4) &= (1-3,2+4) = (-2,6)\\ 10\cdot(-1,1) &= (10\cdot -1, 10\cdot 1) = (-10,10).\end{aligned}$$

Example 1.2 (Polynomials)**.** A second example of a vector space is the ring of polynomials with real-valued coefficients, denoted $\mathbb{R}[x]$. A polynomial $p \in \mathbb{R}[x]$ is a function $p : \mathbb{R} \to \mathbb{R}$ taking the form*

$$p(x) = \sum_k a_k x^k.$$

Addition and scalar multiplication are carried out in the usual way, e.g., if $p(x) = x^2+2x-1$ and $q(x) = x^3$, then $3p(x) + 5q(x) = 5x^3 + 3x^2 + 6x - 3$, which is another polynomial.

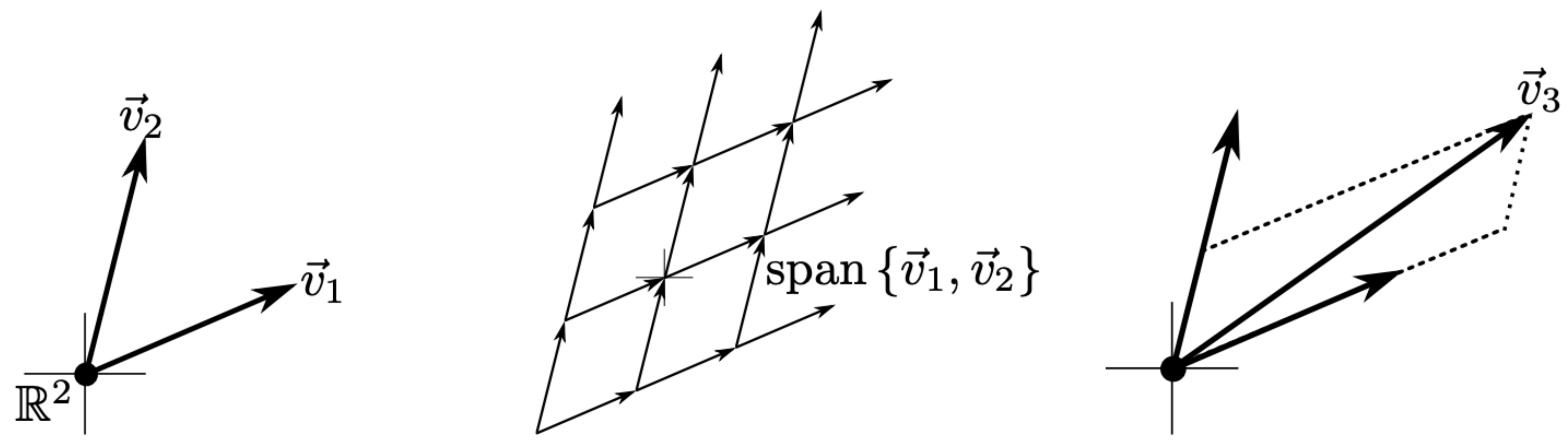
Vector Spaces

Span

Definition 1.2 (Span). The *span* of a set $S \subseteq \mathcal{V}$ of vectors is the set

$$\operatorname{span} S \equiv \{a_1\vec{v}_1 + \cdots + a_k\vec{v}_k : \vec{v}_i \in S \text{ and } a_i \in \mathbb{R} \text{ for all } i\}.$$

A weighted sum $\sum_i a_i\vec{v}_i$, where $a_i \in \mathbb{R}$ and $\vec{v}_i \in \mathcal{V}$, is known as a *linear combination* of the $\vec{v}_i$'s.



(a) $\vec{v}_1, \vec{v}_2 \in \mathbb{R}^2$ (b) span $\{\vec{v}_1, \vec{v}_2\}$ (c) span $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$

Figure 1.1 (a) Vectors $\vec{v}_1, \vec{v}_2 \in \mathbb{R}^2$; (b) their span is the plane $\mathbb{R}^2$; (c) span $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ = span $\{\vec{v}_1, \vec{v}_2\}$ because $\vec{v}_3$ is a linear combination of $\vec{v}_1$ and $\vec{v}_2$.

$\Leftrightarrow$ The set $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is linearly dependent

span S is a *subspace* of $\mathcal{V}$, a subset of $\mathcal{V}$ that is itself a vector space.

Vector Spaces

Linear Dependence

Definition 1.3 (Linear dependence)**.** We provide three equivalent definitions. A set $S \subseteq \mathcal{V}$ of vectors is *linearly dependent* if:

1. One of the elements of S can be written as a linear combination of the other elements, or S contains zero.

2. There exists a non-empty linear combination of elements $\vec{v}_k \in \mathcal{S}$ yielding $\sum_{k=1}^{m} c_k \vec{v}_k = 0$ where $c_k \neq 0$ for all k.

3. There exists $\vec{v} \in S$ such that $\operatorname{span} S = \operatorname{span} S \backslash \{\vec{v}\}$. That is, we can remove a vector from S without affecting its span.

If S is not linearly dependent, then we say it is *linearly independent*.

Vector Spaces

Dimensionality and Basis

Definition 1.4 (Dimension and basis)**.** The *dimension* of $\mathcal{V}$ is the maximal size $|S|$ of a linearly independent set $S \subset \mathcal{V}$ such that $\operatorname{span} S = \mathcal{V}$. Any set S satisfying this property is called a *basis* for $\mathcal{V}$.

Example 1.5 ($\mathbb{R}^n$)**.** The *standard basis* for $\mathbb{R}^n$ is the set of vectors of the form

$$\vec{e}_k \equiv (\underbrace{0, \ldots, 0,}_{k-1 \text{ elements}} \ 1, \underbrace{0, \ldots, 0}_{n-k \text{ elements}}).$$

That is, $\vec{e}_k$ has all zeros except for a single one in the k-th position. These vectors are linearly independent and form a basis for $\mathbb{R}^n$; for example in $\mathbb{R}^3$ any vector (a, b, c) can be written as $a\vec{e}_1 + b\vec{e}_2 + c\vec{e}_3$. Thus, the dimension of $\mathbb{R}^n$ is n, as expected.

Example 1.6 (Polynomials)**.** The set of monomials $\{1, x, x^2, x^3, \ldots\}$ is a linearly independent subset of $\mathbb{R}[x]$. It is infinitely large, and thus the dimension of $\mathbb{R}[x]$ is ∞.

Vector Spaces

$\mathbb{R}^n$, the n-Dimensional Euclidean Space

Definition 1.5 (Dot product)**.** The dot product of two vectors $\vec{a} = (a_1, \ldots, a_n)$ and $\vec{b} = (b_1, \ldots, b_n)$ in $\mathbb{R}^n$ is given by

$$\vec{a} \cdot \vec{b} \equiv \sum_{k=1}^{n} a_k b_k.$$

Example 1.7 ($\mathbb{R}^2$)**.** The dot product of $(1, 2)$ and $(-2, 6)$ is $1 \cdot -2 + 2 \cdot 6 = -2 + 12 = 10$.

the *norm* or *length* of a vector $\vec{a}$ $\qquad \|\vec{a}\|_2 \equiv \sqrt{a_1^2 + \cdots + a_n^2} = \sqrt{\vec{a} \cdot \vec{a}}.$

the distance between two points $\vec{a}, \vec{b} \in \mathbb{R}^n$ is $\|\vec{b} - \vec{a}\|_2$.

the angle θ between $\vec{a}$ and $\vec{b}$ $\qquad \theta \equiv \arccos \dfrac{\vec{a} \cdot \vec{b}}{\|\vec{a}\|_2 \|\vec{b}\|_2}$ $\qquad$ When $\vec{a} = c\vec{b}$ for some $c \in \mathbb{R}$, we have $\theta = \arccos 1 = 0$

Definition 1.6 (Orthogonality)**.** Two vectors $\vec{a}, \vec{b} \in \mathbb{R}^n$ are perpendicular, or *orthogonal*, when $\vec{a} \cdot \vec{b} = 0$.

Table of Content

- Vector Spaces
- Linearity
- Nonlinearity: Differential Calculus

Linearity

Definition

Definition 1.7 (Linearity)**.** Suppose $\mathcal{V}$ and $\mathcal{V}'$ are vector spaces. Then, $\mathcal{L} : \mathcal{V} \to \mathcal{V}'$ is *linear* if it satisfies the following two criteria for all $\vec{v}, \vec{v}_1, \vec{v}_2 \in \mathcal{V}$ and $c \in \mathbb{R}$:

- $\mathcal{L}$ *preserves sums:* $\mathcal{L}[\vec{v}_1 + \vec{v}_2] = \mathcal{L}[\vec{v}_1] + \mathcal{L}[\vec{v}_2]$
- $\mathcal{L}$ *preserves scalar products:* $\mathcal{L}[c\vec{v}] = c\mathcal{L}[\vec{v}]$

Linearity

Example

Example 1.8 (Linearity in $\mathbb{R}^n$)**.** The following map $f : \mathbb{R}^2 \to \mathbb{R}^3$ is linear:

$$f(x, y) = (3x, 2x + y, -y).$$

We can check linearity as follows:

- *Sum preservation:*

$$\begin{aligned} f(x_1 + x_2, y_1 + y_2) &= (3(x_1 + x_2), 2(x_1 + x_2) + (y_1 + y_2), -(y_1 + y_2)) \\ &= (3x_1, 2x_1 + y_1, -y_1) + (3x_2, 2x_2 + y_2, -y_2) \\ &= f(x_1, y_1) + f(x_2, y_2) \quad \checkmark \end{aligned}$$

- *Scalar product preservation:*

$$\begin{aligned} f(cx, cy) &= (3cx, 2cx + cy, -cy) \\ &= c(3x, 2x + y, -y) \\ &= cf(x, y) \quad \checkmark \end{aligned}$$

Contrastingly, $g(x, y) \equiv xy^2$ is not linear. For instance, $g(1, 1) = 1$, but $g(2, 2) = 8 \neq 2 \cdot g(1, 1)$, so g does not preserve scalar products.

Linearity

Example

Example 1.9 (Integration)**.** The following "functional" $\mathcal{L}$ from $\mathbb{R}[x]$ to $\mathbb{R}$ is linear:

$$\mathcal{L}[p(x)] \equiv \int_0^1 p(x)\,dx.$$

This more abstract example maps polynomials $p(x) \in \mathbb{R}[x]$ to real numbers $\mathcal{L}[p(x)] \in \mathbb{R}$. For example, we can write

$$\mathcal{L}[3x^2 + x - 1] = \int_0^1 (3x^2 + x - 1)\,dx = \frac{1}{2}.$$

Linearity of $\mathcal{L}$ is a result of the following well-known identities from calculus:

$$\int_0^1 c \cdot f(x)\,dx = c \int_0^1 f(x)\,dx$$
$$\int_0^1 [f(x) + g(x)]\,dx = \int_0^1 f(x)\,dx + \int_0^1 g(x)\,dx.$$

Linearity

Expanding a Linear Map

The vector $\vec{a} = (a_1, \ldots, a_n)$ is equal to the sum $\sum_k a_k \vec{e}_k$, where $\vec{e}_k$ is the k-th standard basis vector

if $\mathcal{L}$ is linear we can expand:

$$\begin{aligned}
\mathcal{L}[\vec{a}] &= \mathcal{L}\left[\sum_k a_k \vec{e}_k\right] \text{ for the standard basis } \vec{e}_k \\
&= \sum_k \mathcal{L}\left[a_k \vec{e}_k\right] \text{ by sum preservation} \\
&= \sum_k a_k \mathcal{L}\left[\vec{e}_k\right] \text{ by scalar product preservation.}
\end{aligned}$$

A linear operator $\mathcal{L}$ on $\mathbb{R}^n$ is completely determined by its action on the standard basis vectors $\vec{e}_k$.

That is, for any vector $\vec{a} \in \mathbb{R}^n$, we can use the sum above to determine $\mathcal{L}[\vec{a}]$ by linearly combining $\mathcal{L}[\vec{e}_1], \ldots, \mathcal{L}[\vec{e}_n]$.

Linearity

Matrices

The expansion of linear maps above suggests a context in which it is useful to store multiple vectors in the same structure. More generally, say we have n vectors $\vec{v}_1, \ldots, \vec{v}_n \in \mathbb{R}^m$. We can write each as a column vector:

$$\vec{v}_1 = \begin{pmatrix} v_{11} \\ v_{21} \\ \vdots \\ v_{m1} \end{pmatrix}, \vec{v}_2 = \begin{pmatrix} v_{12} \\ v_{22} \\ \vdots \\ v_{m2} \end{pmatrix}, \cdots, \vec{v}_n = \begin{pmatrix} v_{1n} \\ v_{2n} \\ \vdots \\ v_{mn} \end{pmatrix}.$$

Carrying these vectors around separately can be cumbersome notationally, so to simplify matters we combine them into a single $m \times n$ matrix:

$$\begin{pmatrix} | & | & & | \\ \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \\ | & | & & | \end{pmatrix} = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ v_{m1} & v_{m2} & \cdots & v_{mn} \end{pmatrix}.$$

We will call the space of such matrices $\mathbb{R}^{m \times n}$.

Linearity

Identity Matrix

Example 1.11 (Identity matrix)**.** We can store the standard basis for $\mathbb{R}^n$ in the $n \times n$ "identity matrix" $I_{n\times n}$ given by:

$$I_{n\times n} \equiv \begin{pmatrix} | & | & & | \\ \vec{e}_1 & \vec{e}_2 & \cdots & \vec{e}_n \\ | & | & & | \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}.$$

Linearity

Matrix-Vector Product

a matrix in $\mathbb{R}^{m\times n}$ can be multiplied by a column vector in $\mathbb{R}^n$

$$\begin{pmatrix} | & | & & | \\ \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \\ | & | & & | \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} \equiv c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n.$$

Expanding this sum yields the following explicit formula for matrix-vector products:

$$\begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ v_{m1} & v_{m2} & \cdots & v_{mn} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} = \begin{pmatrix} c_1v_{11} + c_2v_{12} + \cdots + c_nv_{1n} \\ c_1v_{21} + c_2v_{22} + \cdots + c_nv_{2n} \\ \vdots \\ c_1v_{m1} + c_2v_{m2} + \cdots + c_nv_{mn} \end{pmatrix}.$$

Example 1.12 (Identity matrix multiplication)**.** For any $\vec{x} \in \mathbb{R}^n$, we can write $\vec{x} = I_{n\times n}\vec{x}$, where $I_{n\times n}$ is the identity matrix from Example 1.11.

Linearity

Matrix-Matrix Product

We similarly define a product between a matrix $M \in \mathbb{R}^{m\times n}$ and another matrix in $\mathbb{R}^{n\times p}$ with columns $\vec{c}_i$ by concatenating individual matrix-vector products:

$$M\begin{pmatrix} | & | & & | \\ \vec{c}_1 & \vec{c}_2 & \cdots & \vec{c}_p \\ | & | & & | \end{pmatrix} \equiv \begin{pmatrix} | & | & & | \\ M\vec{c}_1 & M\vec{c}_2 & \cdots & M\vec{c}_p \\ | & | & & | \end{pmatrix}.$$

We will use capital letters to represent matrices, like $A \in \mathbb{R}^{m\times n}$. We will use the notation $A_{ij} \in \mathbb{R}$ to denote the element of A at row i and column j.

Linearity

Matrix Transpose

Definition 1.8 (Transpose)**.** The *transpose* of a matrix $A \in \mathbb{R}^{m\times n}$ is a matrix $A^\top \in \mathbb{R}^{n\times m}$ with elements $(A^\top)_{ij} = A_{ji}$.

Example 1.15 (Transposition)**.** The transpose of the matrix

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

is given by

$$A^\top = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}.$$

Geometrically, we can think of transposition as flipping a matrix over its diagonal.

$$\boxed{(A^\top)^\top = A, \qquad (A+B)^\top = A^\top + B^\top, \qquad \text{and} \qquad (AB)^\top = B^\top A^\top}$$

Linearity

Residual Norm

Example 1.16 (Residual norm)**.** Suppose we have a matrix A and two vectors $\vec{x}$ and $\vec{b}$. If we wish to know how well $A\vec{x}$ approximates $\vec{b}$, we might define a *residual* $\vec{r} \equiv \vec{b} - A\vec{x}$; this residual is zero exactly when $A\vec{x} = \vec{b}$. Otherwise, we can use the norm $\|\vec{r}\|_2$ as a proxy for the similarity of $A\vec{x}$ and $\vec{b}$. We can use the identities above to simplify:

$$\begin{aligned}
\|\vec{r}\|_2^2 &= \|\vec{b} - A\vec{x}\|_2^2 \\
&= (\vec{b} - A\vec{x}) \cdot (\vec{b} - A\vec{x}) \text{ as explained in §1.2.3} \\
&= (\vec{b} - A\vec{x})^\top (\vec{b} - A\vec{x}) \text{ by our expression for the dot product above} \\
&= (\vec{b}^\top - \vec{x}^\top A^\top)(\vec{b} - A\vec{x}) \text{ by properties of transposition} \\
&= \vec{b}^\top \vec{b} - \vec{b}^\top A\vec{x} - \vec{x}^\top A^\top \vec{b} + \vec{x}^\top A^\top A\vec{x} \text{ after multiplication}
\end{aligned}$$

All four terms on the right-hand side are scalars, or equivalently 1×1 matrices. Scalars thought of as matrices enjoy one additional nice property $c^\top = c$, since there is nothing to transpose! Thus,

$$\vec{x}^\top A^\top \vec{b} = (\vec{x}^\top A^\top \vec{b})^\top = \vec{b}^\top A\vec{x}.$$

This allows us to simplify even more:

$$\begin{aligned}
\|\vec{r}\|_2^2 &= \vec{b}^\top \vec{b} - 2\vec{b}^\top A\vec{x} + \vec{x}^\top A^\top A\vec{x} \\
&= \|A\vec{x}\|_2^2 - 2\vec{b}^\top A\vec{x} + \|\vec{b}\|_2^2.
\end{aligned}$$

Linearity

Linear System: $A\overrightarrow{x} = \overrightarrow{b}$

For example:

$$\begin{aligned} 3x + 2y + 5z &= 0 \\ -4x + 9y - 3z &= -7 \\ 2x - 3y - 3z &= 1. \end{aligned} \quad \Leftrightarrow \quad \underbrace{\begin{pmatrix} 3 & 2 & 5 \\ -4 & 9 & -3 \\ 2 & -3 & -3 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} x \\ y \\ z \end{pmatrix}}_{\overrightarrow{x}} = \underbrace{\begin{pmatrix} 0 \\ -7 \\ 1 \end{pmatrix}}_{\overrightarrow{b}}$$

Write $\vec{b}$ as a linear combination of the columns of A

Definition 1.9 (Column space and rank)**.** The *column space* of a matrix $A \in \mathbb{R}^{m \times n}$ is the span of the columns of A. It can be written as

$$\text{col } A \equiv \{A\vec{x} : \vec{x} \in \mathbb{R}^n\}.$$

The *rank* of A is the dimension of col A.

$A\vec{x} = \vec{b}$ is solvable exactly when $\vec{b} \in \text{col } A$.

Linearity

Inverse Matrix

Suppose A is square and $A\overrightarrow{x} = \overrightarrow{b}$ has a solution for all choices of $\overrightarrow{b}$

Then, we can substitute the standard basis $\overrightarrow{e}_1, \overrightarrow{e}_2, \ldots, \overrightarrow{e}_n$ to solve equations of the form $A\overrightarrow{x}_i = \overrightarrow{e}_i$:

$$\begin{aligned} A\begin{pmatrix} | & | & & | \\ \vec{x}_1 & \vec{x}_2 & \cdots & \vec{x}_n \\ | & | & & | \end{pmatrix} &= \begin{pmatrix} | & | & & | \\ A\vec{x}_1 & A\vec{x}_2 & \cdots & A\vec{x}_n \\ | & | & & | \end{pmatrix} \\ &= \begin{pmatrix} | & | & & | \\ \vec{e}_1 & \vec{e}_2 & \cdots & \vec{e}_n \\ | & | & & | \end{pmatrix} = I_{n\times n}, \end{aligned}$$

where $I_{n\times n}$ is the identity matrix from Example 1.11. We will call the matrix with columns $\vec{x}_k$ the *inverse* A^{-1}, which satisfies

$$AA^{-1} = A^{-1}A = I_{n\times n}.$$

By construction, $(A^{-1})^{-1} = A$. If we can find such an inverse, solving any linear system $A\vec{x} = \vec{b}$ reduces to matrix multiplication, since:

$$\vec{x} = I_{n\times n}\vec{x} = (A^{-1}A)\vec{x} = A^{-1}(A\vec{x}) = A^{-1}\vec{b}.$$

Table of Content

- Vector Spaces
- Linearity
- Nonlinearity: Differential Calculus

Nonlinearity: Differential Calculus

Differentiation in One Variable

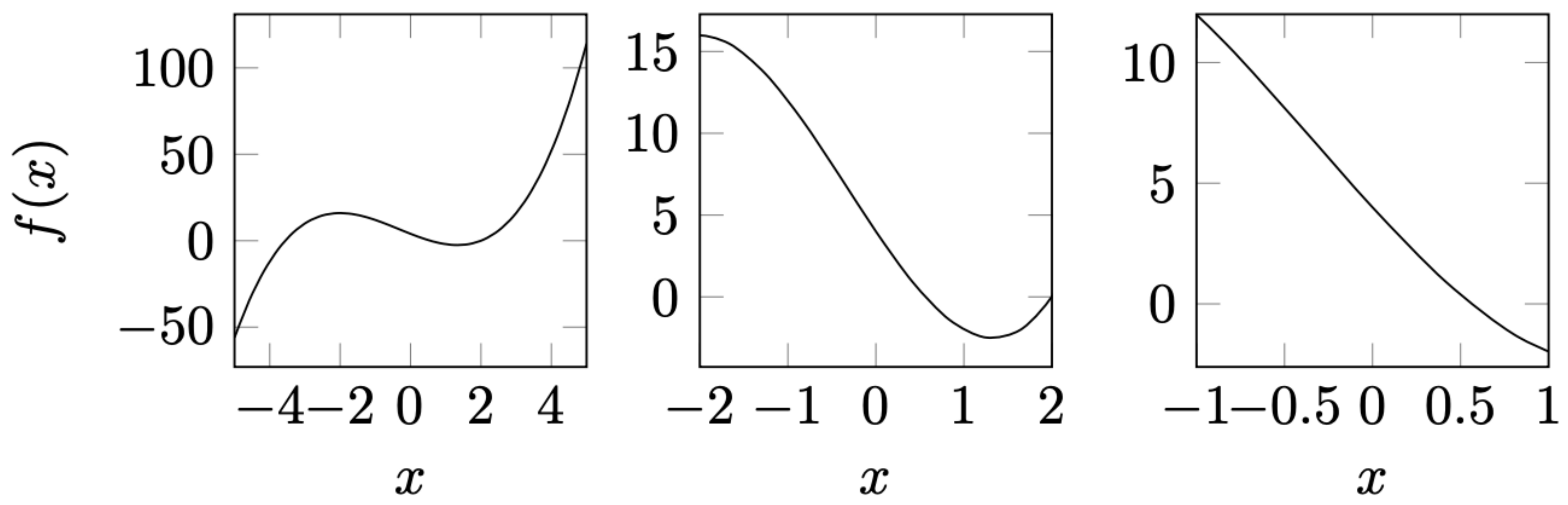


Figure 1.4 The closer we zoom into $f(x) = x^3 + x^2 - 8x + 4$, the more it looks like a line.

The derivative $f'(x)$ of a function $f(x) : \mathbb{R} \to \mathbb{R}$ is the slope of the approximating line, computed by finding the slope of lines through closer and closer points to x:

$$f'(x) = \lim_{y \to x} \frac{f(y) - f(x)}{y - x}.$$

Nonlinearity: Differential Calculus
Approximating Derivatives

Take $x, y \in \mathbb{R}$. Then, we can expand:

$$\begin{aligned} f(y) - f(x) &= \int_x^y f'(t)\,dt \text{ by the Fundamental Theorem of Calculus} \\ &= yf'(y) - xf'(x) - \int_x^y tf''(t)\,dt, \text{ after integrating by parts} \\ &= (y-x)f'(x) + y(f'(y) - f'(x)) - \int_x^y tf''(t)\,dt \\ &= (y-x)f'(x) + y\int_x^y f''(t)\,dt - \int_x^y tf''(t)\,dt \\ &\qquad \text{again by the Fundamental Theorem of Calculus} \\ &= (y-x)f'(x) + \int_x^y (y-t)f''(t)\,dt. \end{aligned}$$

Rearranging terms and defining $\Delta x \equiv y - x$ shows:

$$\begin{aligned} |f'(x)\Delta x - [f(y) - f(x)]| &= \left|\int_x^y (y-t)f''(t)\,dt\right| \text{ from the relationship above} \\ &\le |\Delta x| \int_x^y |f''(t)|\,dt, \text{ by the Cauchy-Schwarz inequality} \\ &\le D|\Delta x|^2, \text{ assuming } |f''(t)| < D \text{ for some } D > 0. \end{aligned}$$

Nonlinearity: Differential Calculus

Infinitesimal Big-O

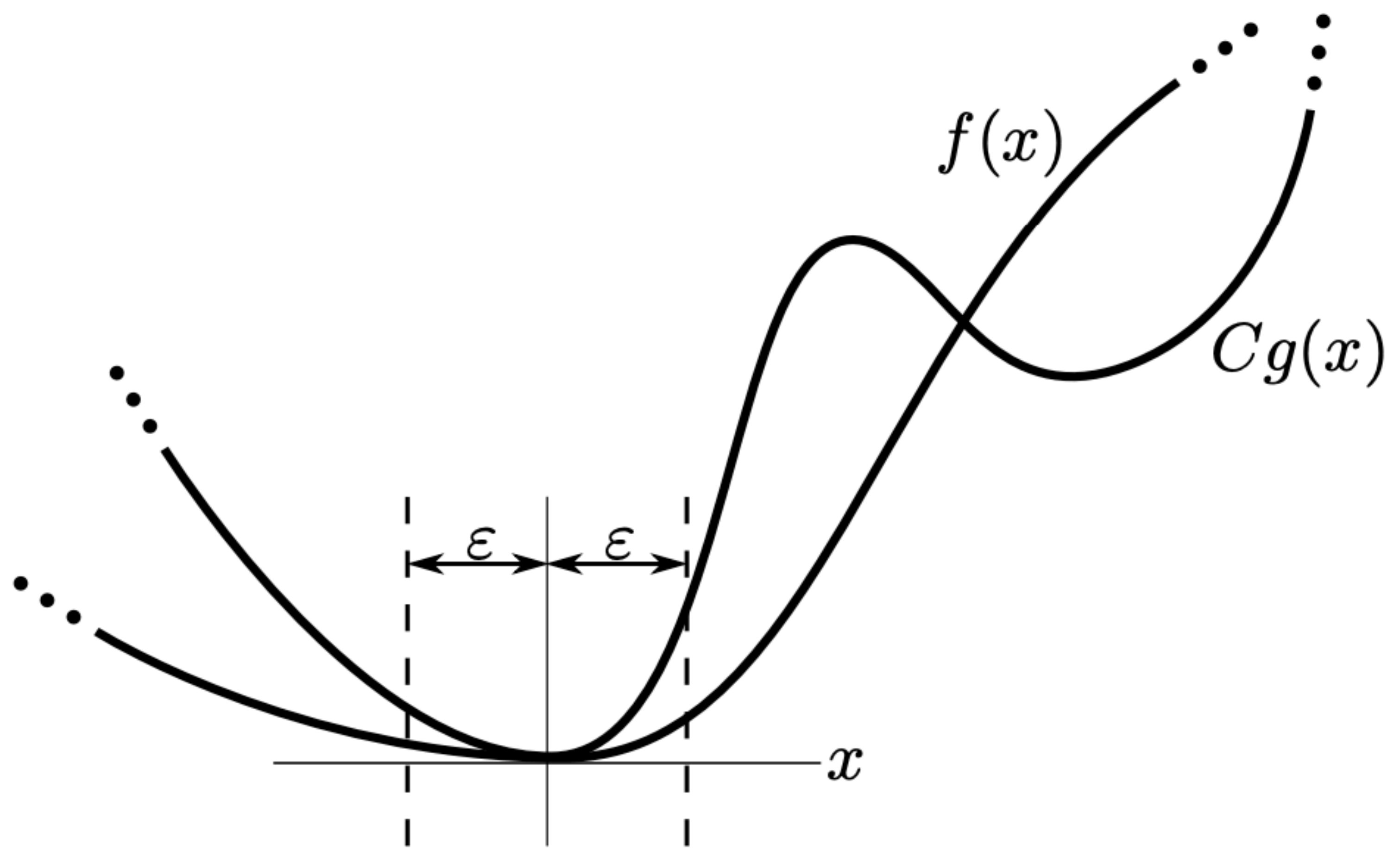


Figure 1.5 Big-O notation; in the ε neighborhood of the origin, $f(x)$ is dominated by $Cg(x)$; outside this neighborhood, $Cg(x)$ can dip back down.

Definition 1.10 (Infinitesimal big-O)**.** We will say $f(x) = O(g(x))$ if there exists a constant $C > 0$ and some $\varepsilon > 0$ such that $|f(x)| \leq C|g(x)|$ for all x with $|x| < \varepsilon$.

Nonlinearity: Differential Calculus

Taylor's Theorem

Our derivation above shows the following relationship for smooth functions $f : \mathbb{R} \to \mathbb{R}$:

$$f(x + \Delta x) = f(x) + f'(x)\Delta x + O(\Delta x^2).$$

This is an instance of Taylor's theorem, which we will apply copiously when developing strategies for integrating ordinary differential equations. More generally, this theorem shows how to approximate differentiable functions with polynomials:

$$f(x + \Delta x) = f(x) + f'(x)\Delta x + f''(x)\frac{\Delta x^2}{2!} + \cdots + f^{(k)}(x)\frac{\Delta x^k}{k!} + O(\Delta x^{k+1}).$$

Nonlinearity: Differential Calculus

Differentiation in Multiple Variables

If a function f takes multiple inputs, then it can be written $f(\vec{x}) : \mathbb{R}^n \to \mathbb{R}$ for $\vec{x} \in \mathbb{R}^n$. In other words, to each point $\vec{x} = (x_1, \ldots, x_n)$ in n-dimensional space, f assigns a single number $f(x_1, \ldots, x_n)$.

Definition 1.11 (Partial derivative)**.** The k-th *partial derivative* of f, notated $\frac{\partial f}{\partial x_k}$, is given by differentiating f in its k-th input variable:

$$\frac{\partial f}{\partial x_k}(x_1, \ldots, x_n) \equiv \frac{d}{dt} f(x_1, \ldots, x_{k-1}, t, x_{k+1}, \ldots, x_n)|_{t=x_k}.$$

Nonlinearity: Differential Calculus

Gradient

Using single-variable calculus, for a function $f : \mathbb{R}^n \to \mathbb{R}$,

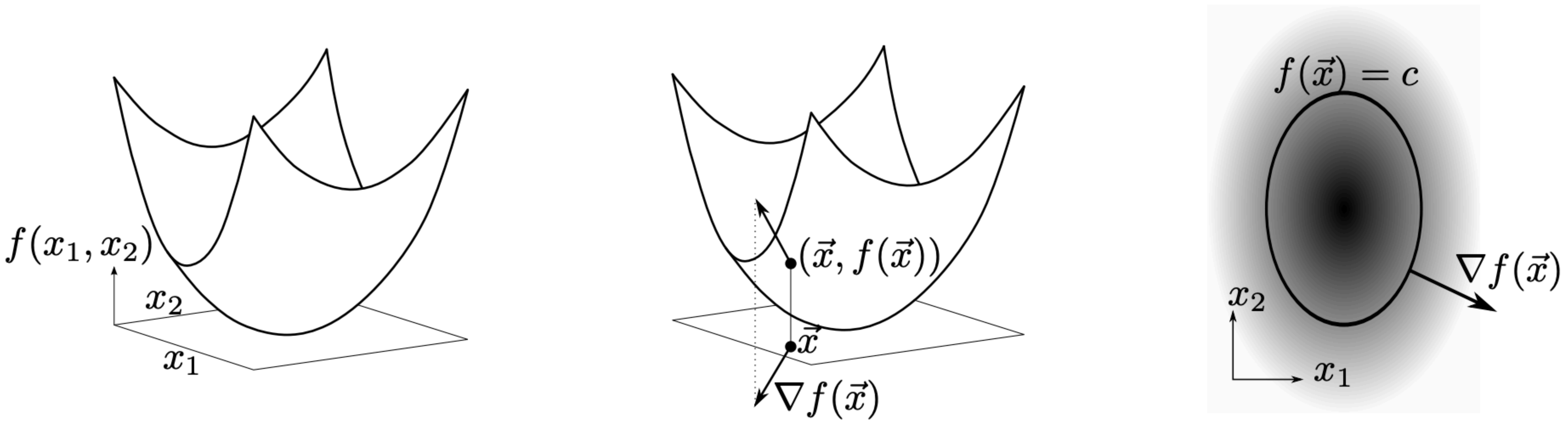
$$\begin{aligned}
f(\vec{x} + \Delta\vec{x}) &= f(x_1 + \Delta x_1, x_2 + \Delta x_2, \ldots, x_n + \Delta x_n) \\
&= f(x_1, x_2 + \Delta x_2, \ldots, x_n + \Delta x_n) + \frac{\partial f}{\partial x_1}\Delta x_1 + O(\Delta x_1^2) \\
&\qquad \text{by single-variable calculus in } x_1 \\
&= f(x_1, \ldots, x_n) + \sum_{k=1}^{n} \left[\frac{\partial f}{\partial x_k}\Delta x_k + O(\Delta x_k^2) \right] \\
&\qquad \text{by repeating this } n-1 \text{ times in } x_2, \ldots, x_n \\
&= f(\vec{x}) + \nabla f(\vec{x}) \cdot \Delta\vec{x} + O(\|\Delta\vec{x}\|_2^2),
\end{aligned}$$

where we define the *gradient* of f as

$$\nabla f(\vec{x}) \equiv \left(\frac{\partial f}{\partial x_1}(\vec{x}), \frac{\partial f}{\partial x_2}(\vec{x}), \cdots, \frac{\partial f}{\partial x_n}(\vec{x}) \right) \in \mathbb{R}^n.$$

Nonlinearity: Differential Calculus

Gradient



Graph of $f(\vec{x})$ Steepest ascent Level sets of $f(\vec{x})$

Figure 1.6 We can visualize a function $f(x_1, x_2)$ as a three-dimensional graph; then $\nabla f(\vec{x})$ is the direction on the (x_1, x_2) plane corresponding to the steepest ascent of f. Alternatively, we can think of $f(x_1, x_2)$ as the *brightness* at (x_1, x_2) (dark indicates a low value of f), in which case ∇f points perpendicular to level sets $f(\vec{x}) = c$ in the direction where f is increasing and the image gets lighter.

Nonlinearity: Differential Calculus

Directional Derivative

We can differentiate f in any direction $\vec{v}$ via the *directional derivative* $D_{\vec{v}}f$:

$$D_{\vec{v}}f(\vec{x}) \equiv \frac{d}{dt}f(\vec{x}+t\vec{v})|_{t=0} = \nabla f(\vec{x}) \cdot \vec{v}.$$

We allow $\vec{v}$ to have any length, with the property $D_{c\vec{v}}f(\vec{x}) = cD_{\vec{v}}f(\vec{x})$.

Example 1.18 ($\mathbb{R}^2$)**.** Take $f(x,y) = x^2y^3$. Then,

$$\frac{\partial f}{\partial x} = 2xy^3 \qquad \frac{\partial f}{\partial y} = 3x^2y^2.$$

Equivalently, $\nabla f(x,y) = (2xy^3, 3x^2y^2)$. So, the derivative of f at $(x,y) = (1,2)$ in the direction $(-1,4)$ is given by $(-1,4)\cdot\nabla f(1,2) = (-1,4)\cdot(16,12) = 32$.

Nonlinearity: Differential Calculus

Quadratic Forms

Example 1.20 (Quadratic forms)**.** Take any matrix $A \in \mathbb{R}^{n\times n}$, and define $f(\vec{x}) \equiv \vec{x}^\top A\vec{x}$. Writing this function element-by-element shows

$$f(\vec{x}) = \sum_{ij} A_{ij}x_i x_j.$$

Expanding f and checking this relationship explicitly is worthwhile. Take some $k \in \{1,\ldots,n\}$. Then, we can separate out all terms containing x_k:

$$f(\vec{x}) = A_{kk}x_k^2 + x_k\left(\sum_{i\neq k} A_{ik}x_i + \sum_{j\neq k} A_{kj}x_j\right) + \sum_{i,j\neq k} A_{ij}x_i x_j.$$

With this factorization,

$$\frac{\partial f}{\partial x_k} = 2A_{kk}x_k + \left(\sum_{i\neq k} A_{ik}x_i + \sum_{j\neq k} A_{kj}x_j\right) = \sum_{i=1}^{n}(A_{ik} + A_{ki})x_i.$$

This sum looks a lot like the definition of matrix-vector multiplication! Combining these partial derivatives into a single vector shows $\nabla f(\vec{x}) = (A + A^\top)\vec{x}$. In the special case when A is symmetric, that is, when $A^\top = A$, we have the well-known formula $\nabla f(\vec{x}) = 2A\vec{x}$.

Nonlinearity: Differential Calculus

Jacobian

we should consider $f : \mathbb{R}^n \to \mathbb{R}^m$. That is, f inputs n numbers and outputs m numbers

$$f(\vec{x}) = \begin{pmatrix} f_1(\vec{x}) \\ f_2(\vec{x}) \\ \vdots \\ f_m(\vec{x}) \end{pmatrix}$$

Definition 1.12 (Jacobian)**.** The *Jacobian* of $f : \mathbb{R}^n \to \mathbb{R}^m$ is the matrix $Df(\vec{x}) \in \mathbb{R}^{m\times n}$ with entries

$$(Df)_{ij} \equiv \frac{\partial f_i}{\partial x_j}.$$

Example 1.21 (Jacobian computation)**.** Suppose $f(x, y) = (3x, -xy^2, x + y)$. Then,

$$Df(x,y) = \begin{pmatrix} 3 & 0 \\ -y^2 & -2xy \\ 1 & 1 \end{pmatrix}.$$

Example 1.22 (Matrix multiplication)**.** Unsurprisingly, the Jacobian of $f(\vec{x}) = A\vec{x}$ for matrix A is given by $Df(\vec{x}) = A$.

Table of Content

- Vector Spaces
- Linearity
- Nonlinearity: Differential Calculus