

Matrices

Last modified: 2026-10-27 20:53:54 (PDT)

Matrices

Definition 0.1 (Matrix). A **matrix** [1] of dimensions $m \times n$ is a rectangular array of $m \cdot n$ numbers, arranged in m rows and n columns:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

i Remark (Matrix entries, and vectors as matrices)

Remark 0.1 (Matrix entries, and vectors as matrices). The entry in row i and column j of $\mathbf{A}$ is written a_{ij} or $(\mathbf{A})_{ij}$. For example, for the 2×3 matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix},$$

$a_{12} = 2$ and $a_{23} = 6$.

A column vector of length p is a $p \times 1$ matrix, and a row vector of length p is a $1 \times p$ matrix. For example, the column vector $(7, 8)$ is a 2×1 matrix, and the row vector $[7, 8]$ is a 1×2 matrix.

Definition 0.2 (Tensor). A **tensor** [2] of order $k \geq 1$ is an array of numbers with k indices,

$$\mathcal{A} = (a_{i_1 i_2 \dots i_k}), \quad i_1 = 1, \dots, n_1, \dots, i_k = 1, \dots, n_k,$$

and its dimensions are $n_1 \times n_2 \times \dots \times n_k$.

Example 0.1 (Vectors, matrices and images as tensors). A vector of length p is a tensor of order 1, and an $m \times n$ **matrix** is a tensor of order 2. A 28×28 grayscale image is a matrix: the entry a_{ij} is the brightness of the pixel in row i and column j . A 28×28 color image stores three numbers per pixel (red, green, blue), so it is an order-3 tensor with dimensions $28 \times 28 \times 3$, and a_{ijc} is the value of color channel c at the pixel in row i and column j . It has $28 \cdot 28 \cdot 3 = 2352$ entries.

Matrix transpose

Definition 0.3 (Matrix transpose). The **transpose** [3] of an $m \times n$ matrix $\mathbf{A}$ is the $n \times m$ matrix $\mathbf{A}^\top$ obtained by swapping the rows and columns of $\mathbf{A}$:

$$(\mathbf{A}^\top)_{ij} = a_{ji}$$

Example 0.2 (Transposing a 2×3 matrix). For $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$,

$$\mathbf{A}^\top = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix},$$

a 3×2 matrix. For instance, entry $(1, 2)$ of the transpose is $a_{21} = 4$.

Matrix addition

Definition 0.4 (Zero matrix). The $m \times n$ **zero matrix** [4] $\mathbf{0}_{m \times n}$ (or $\mathbf{0}$ when dimensions are clear from context) has all entries equal to zero:

$$\mathbf{0}_{m \times n} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

Example 0.3 (Zero matrices of two shapes).

$$\mathbf{0}_{2 \times 3} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{0}_{3 \times 2} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Both have all entries 0, but one is 2×3 and the other 3×2 , so they are different matrices.

Definition 0.5 (Matrix addition). Two matrices $\mathbf{A}$ and $\mathbf{B}$ of the same dimensions $m \times n$ can be added element-wise; their **matrix sum** [5] is:

$$(\mathbf{A} + \mathbf{B})_{ij} = a_{ij} + b_{ij}$$

Example 0.4 (Adding matrices, and a sum that is not defined).

$$\begin{aligned} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} &= \begin{bmatrix} 1+0 & 2+1 \\ 3+1 & 4+0 \end{bmatrix} \quad (\text{add entry by entry}) \\ &= \begin{bmatrix} 1 & 3 \\ 4 & 4 \end{bmatrix}. \quad (\text{add}) \end{aligned}$$

A 2×2 matrix and a 2×3 matrix cannot be added: they do not have the same dimensions, so entry $(1, 3)$ of the sum would have no first term.

Theorem 0.1 (Matrix addition is commutative). For $m \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$, matrix addition is *commutative*:

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$$

Theorem 0.2 (Matrix addition is associative). For $m \times n$ matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$, matrix addition is *associative*:

$$(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$$

Theorem 0.3 (Zero matrix is the additive identity). For an $m \times n$ matrix $\mathbf{A}$:

$$\mathbf{A} + \mathbf{0}_{m \times n} = \mathbf{A}$$

Theorem 0.4 (Additive inverse). For any $m \times n$ matrix $\mathbf{A}$, the $m \times n$ matrix $-\mathbf{A}$ (defined by $(-\mathbf{A})_{ij} = -a_{ij}$) satisfies:

$$\mathbf{A} + (-\mathbf{A}) = \mathbf{0}_{m \times n}$$

Scalar multiplication

Definition 0.6 (Scalar multiplication). The **scalar multiple** [6] of a matrix $\mathbf{A}$ by a scalar c is:

$$(c\mathbf{A})_{ij} = c \cdot a_{ij}$$

Example 0.5 (A scalar multiple).

$$\begin{aligned} 3 \begin{bmatrix} 1 & -2 \\ 0 & 4 \end{bmatrix} &= \begin{bmatrix} 3 \cdot 1 & 3 \cdot (-2) \\ 3 \cdot 0 & 3 \cdot 4 \end{bmatrix} \\ &= \begin{bmatrix} 3 & -6 \\ 0 & 12 \end{bmatrix}. \end{aligned}$$

Matrix multiplication

Definition 0.7 (Matrix multiplication). The **product** [7] of an $m \times k$ matrix $\mathbf{A}$ and a $k \times n$ matrix $\mathbf{B}$ is the $m \times n$ matrix $\mathbf{C} = \mathbf{AB}$ with entries:

$$c_{ij} = \sum_{s=1}^k a_{is} b_{sj}$$

i Remark (Matrix multiplication is not commutative)

Remark 0.2 (Matrix multiplication is not commutative). When $\mathbf{AB}$ and $\mathbf{BA}$ are both defined, they are usually different. For example, with

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix},$$

$$\mathbf{AB} = \begin{bmatrix} 3 & 2 \\ 1 & 1 \end{bmatrix}$$

$$\mathbf{BA} = \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix},$$

so $\mathbf{AB} \neq \mathbf{BA}$.

Example 0.6 (Dot product as matrix multiplication). The dot product of two column vectors $\tilde{x}$ and $\tilde{\beta}$ can be written as a matrix product of the row vector $\tilde{x}^\top$ with the column vector $\tilde{\beta}$:

$$\begin{aligned}\tilde{x} \cdot \tilde{\beta} &= \tilde{x}^\top \tilde{\beta} \\ &= [x_1, x_2, \dots, x_p] \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} \\ &= x_1\beta_1 + x_2\beta_2 + \dots + x_p\beta_p\end{aligned}$$

Theorem 0.5 (Matrix multiplication is associative). For an $m \times k$ matrix $\mathbf{A}$, a $k \times l$ matrix $\mathbf{B}$, and an $l \times n$ matrix $\mathbf{C}$:

$$\underbrace{(\mathbf{AB})\mathbf{C}}_{m \times n} = \underbrace{\mathbf{A}(\mathbf{BC})}_{m \times n}$$

i Proof

Proof. Entry (i, j) of each side, from Definition 0.7:

$$\begin{aligned}[(\mathbf{AB})\mathbf{C}]_{ij} &= \sum_{t=1}^l (\mathbf{AB})_{it} c_{tj} && \text{(definition of } (\mathbf{AB})\mathbf{C} \text{)} \\ &= \sum_{t=1}^l \left(\sum_{s=1}^k a_{is} b_{st} \right) c_{tj} && \text{(definition of } \mathbf{AB} \text{)} \\ &= \sum_{s=1}^k a_{is} \left(\sum_{t=1}^l b_{st} c_{tj} \right) && \text{(distribute and swap the finite sums)} \\ &= \sum_{s=1}^k a_{is} (\mathbf{BC})_{sj} && \text{(definition of } \mathbf{BC} \text{)} \\ &= [\mathbf{A}(\mathbf{BC})]_{ij} && \text{(definition of } \mathbf{A}(\mathbf{BC}) \text{)}\end{aligned}$$

□

Theorem 0.6 (Matrix multiplication is distributive over addition). Matrix multiplication *distributes* over addition from the left and from the right. For an $m \times k$ matrix $\mathbf{A}$ and $k \times n$ matrices $\mathbf{B}$ and $\mathbf{C}$:

$$\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$$

For $m \times k$ matrices $\mathbf{A}$ and $\mathbf{B}$ and a $k \times n$ matrix $\mathbf{C}$:

$$(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$$

Theorem 0.7 (Product of two sums of matrices). For $m \times k$ matrices $\mathbf{A}$ and $\mathbf{B}$ and $k \times n$ matrices $\mathbf{C}$ and $\mathbf{D}$,

$$(\mathbf{A} + \mathbf{B})(\mathbf{C} + \mathbf{D}) = \mathbf{AC} + \mathbf{AD} + \mathbf{BC} + \mathbf{BD}.$$

Each product keeps its factors in the order they appear on the left, because matrix multiplication is not commutative. So $\mathbf{AD}$ and $\mathbf{BC}$ are not collected into $2\mathbf{AD}$.

Proof

Proof.

$$\begin{aligned} (\mathbf{A} + \mathbf{B})(\mathbf{C} + \mathbf{D}) &= (\mathbf{A} + \mathbf{B})\mathbf{C} + (\mathbf{A} + \mathbf{B})\mathbf{D} && \text{(Theorem 0.6, from the left)} \\ &= \mathbf{AC} + \mathbf{BC} + \mathbf{AD} + \mathbf{BD} && \text{(Theorem 0.6, from the right, twice)} \\ &= \mathbf{AC} + \mathbf{AD} + \mathbf{BC} + \mathbf{BD} && \text{(matrix addition is commutative)} \end{aligned}$$

□

Example 0.7 (Why the middle terms do not combine). Take $\mathbf{A} = \mathbf{C} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and

$\mathbf{B} = \mathbf{D} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$. Then $\mathbf{A} + \mathbf{B} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$, so

$$\begin{aligned} (\mathbf{A} + \mathbf{B})^2 &= \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}, \\ \mathbf{A}^2 + \mathbf{AB} + \mathbf{BA} + \mathbf{B}^2 &= \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}. \end{aligned}$$

Here $\mathbf{AB} \neq \mathbf{BA}$, so the middle terms do not combine into $2\mathbf{AB}$. The squared norm of a sum of vectors, $\|x + y\|^2$, is a case where they do combine ([Theorem 2 in Inner Products and Orthogonality](#)), because $\tilde{x} \cdot \tilde{y} = \tilde{y} \cdot \tilde{x}$. Matrices that commute, such as $\mathbf{A}$ and $\mathbf{A}^2$, combine the same way.

Exercise 0.1 (Which of these products are defined?). Let $\mathbf{A}$ be 4×3 , let $\mathbf{B}$ be 3×2 , and let $\tilde{x} \in \mathbb{R}^3$. For each expression, say whether it is defined, and if so give the shape of the result:

1. $\mathbf{AB}$
2. $\mathbf{BA}$
3. $\mathbf{Ax}$
4. $\tilde{x}^\top \tilde{x}$
5. $\tilde{x} \tilde{x}^\top$

Solution

Solution 0.1. A vector in $\mathbb{R}^3$ is a 3×1 matrix, so the same rule settles every case.

1. $\mathbf{AB}$: $(4 \times 3)(3 \times 2)$. The inner pair is 3 and 3, so it is defined, and the result is 4×2 .
2. $\mathbf{BA}$: $(3 \times 2)(4 \times 3)$. The inner pair is 2 and 4, which do not match, so it is **not defined**. Matrix multiplication is not commutative, and this is the blunt form of that: swapping the order can leave an expression that means nothing.
3. $\mathbf{Ax}$: $(4 \times 3)(3 \times 1)$, defined, result 4×1 — a vector in $\mathbb{R}^4$.
4. $\tilde{x}^\top \tilde{x}$: $(1 \times 3)(3 \times 1)$, defined, result 1×1 — a single number. The product is the inner product of [Equation 1 in Vectors](#).
5. $\tilde{x} \tilde{x}^\top$: $(3 \times 1)(1 \times 3)$, defined, result 3×3 — a matrix.

The last two use the same two vectors and differ only in order, and they return objects of different kinds. Reading a transpose as decoration rather than as a shape change is the commonest way to lose track of an expression.

Square and identity matrices

Definition 0.8 (Square matrix). A matrix is **square** [8] if it has the same number of rows as columns.

Example 0.8 (Square and not square). $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ is square (2 rows, 2 columns); $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ is not (2 rows, 3 columns).

Definition 0.9 (Main diagonal). The **main diagonal** [9] of an $m \times n$ matrix $\mathbf{A}$ is the list of its entries $a_{11}, a_{22}, \dots, a_{kk}$, where k is the smaller of m and n .

Definition 0.10 (Diagonal entry). The entries on the main diagonal (Definition 0.9) of a matrix $\mathbf{A}$ are the **diagonal entries** of $\mathbf{A}$.

Definition 0.11 (Off-diagonal entry). In an $m \times n$ matrix $\mathbf{A}$ with main diagonal (Definition 0.9), every entry a_{ij} with $i \neq j$ is an **off-diagonal entry**.

Example 0.9 (Diagonal entries of a square and a non-square matrix).

- For $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, the diagonal entries are $a_{11} = 1$ and $a_{22} = 4$, and the off-diagonal entries are $a_{12} = 2$ and $a_{21} = 3$.
- For $\mathbf{B} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$, $m = 2$ and $n = 3$, so $k = 2$, and the diagonal entries are $b_{11} = 1$ and $b_{22} = 5$; the other four entries, 2, 3, 4 and 6, are off-diagonal.

Definition 0.12 (Identity matrix). The $p \times p$ **identity matrix** [10] $\mathbf{I}_p$ (or $\mathbf{I}$ when the size is clear from context) has ones on the [main diagonal](#) and zeros elsewhere:

$$(\mathbf{I}_p)_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \quad \mathbf{I}_p = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

Example 0.10 (The 3×3 identity).

$$\mathbf{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

so, for instance, $(\mathbf{I}_3)_{22} = 1$ and $(\mathbf{I}_3)_{23} = 0$.

i Remark (The identity matrix and the Kronecker delta)

Remark 0.3 (The identity matrix and the Kronecker delta). The entries of the identity matrix are given by the [Kronecker delta](#): $(\mathbf{I}_p)_{ij} = \delta_{ij}$.

Theorem 0.8 (Identity matrix is a multiplicative identity). *For any $m \times p$ matrix $\mathbf{A}$:*

$$\mathbf{A} \mathbf{I}_p = \mathbf{A}$$

$$\mathbf{I}_m \mathbf{A} = \mathbf{A}$$

Matrix-vector multiplication

Definition 0.13 (Matrix-vector multiplication). The **matrix-vector product** of an $m \times p$ matrix $\mathbf{A}$ and a $p \times 1$ column vector $\tilde{x}$ is the $m \times 1$ column vector $\mathbf{A}\tilde{x}$ with entries:

$$(\mathbf{A}\tilde{x})_i = \sum_{j=1}^p a_{ij} x_j$$

i Remark (Each entry is a dot product)

Remark 0.4 (Each entry is a dot product). Matrix-vector multiplication extends the dot product ([Definition 8 in Vectors](#)): entry i of $\mathbf{A}\tilde{x}$ is the dot product of row i of $\mathbf{A}$ with $\tilde{x}$. For example,

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \cdot 1 + 2 \cdot (-1) \\ 3 \cdot 1 + 4 \cdot (-1) \end{bmatrix} \\ = \begin{bmatrix} -1 \\ -1 \end{bmatrix}.$$

Definition 0.14 (Linear system, system of linear equations). A **linear system** [11] (or **system of linear equations**) with m equations in n unknowns is an equation $\mathbf{A}\tilde{x} = \tilde{b}$ in which the $m \times n$ matrix $\mathbf{A}$ and the vector $\tilde{b} \in \mathbb{R}^m$ are given. A **solution** of the system is a vector $\tilde{x} \in \mathbb{R}^n$ for which $\mathbf{A}\tilde{x} = \tilde{b}$ holds.

Example 0.11 (Two equations in two unknowns). By [Definition 0.13](#), row i of $\mathbf{A}\tilde{x} = \tilde{b}$ is the equation $a_{i1}x_1 + \cdots + a_{in}x_n = b_i$. So the two equations $x_1 + 2x_2 = 5$ and $3x_1 + 4x_2 = 6$ form the linear system $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \tilde{x} = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$. The vector $\tilde{x} = (-4, 4.5)$ is a solution:

$$\begin{aligned}
\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} -4 \\ 4.5 \end{bmatrix} &= \begin{bmatrix} 1 \cdot (-4) + 2 \cdot 4.5 \\ 3 \cdot (-4) + 4 \cdot 4.5 \end{bmatrix} && \text{(Definition 0.13)} \\
&= \begin{bmatrix} -4 + 9 \\ -12 + 18 \end{bmatrix} && \text{(multiply)} \\
&= \begin{bmatrix} 5 \\ 6 \end{bmatrix} && \text{(add)}
\end{aligned}$$

The vector $(1, 2)$ is not a solution: its first equation holds, since $1 + 2 \cdot 2 = 5$, but its second gives $3 \cdot 1 + 4 \cdot 2 = 11 \neq 6$.

Linear and affine maps

Definition 0.15 (Linear map, linear transformation, linear function). A [function](#) $f : \mathbb{R}^p \rightarrow \mathbb{R}^m$ is a **linear map** [12] (also called a **linear transformation** or **linear function**) if it preserves vector addition and scalar multiplication: for all $\tilde{x}, \tilde{y} \in \mathbb{R}^p$ and $c \in \mathbb{R}$,

1. $f(\tilde{x} + \tilde{y}) = f(\tilde{x}) + f(\tilde{y})$ (*additivity*)
2. $f(c\tilde{x}) = cf(\tilde{x})$ (*homogeneity*)

Example 0.12 (A linear map, and two functions that are not).

- $f(\tilde{x}) = (2x_1, x_1 + x_2)$ on $\mathbb{R}^2$ is linear. For $\tilde{x}, \tilde{y} \in \mathbb{R}^2$ and $c \in \mathbb{R}$:

$$\begin{aligned}
f(\tilde{x} + \tilde{y}) &= (2(x_1 + y_1), (x_1 + y_1) + (x_2 + y_2)) \\
&= f(\tilde{x}) + f(\tilde{y})
\end{aligned}$$

and

$$\begin{aligned}
f(c\tilde{x}) &= (2cx_1, cx_1 + cx_2) \\
&= cf(\tilde{x}).
\end{aligned}$$

- $g(x) = x + 1$ on $\mathbb{R}$ is not linear: additivity fails, since $g(0+0) = 1$ but $g(0) + g(0) = 2$.
- $h(x) = x^2$ on $\mathbb{R}$ is not linear: $h(2 \cdot 1) = 4$ but $2h(1) = 2$, so $h(c\tilde{x}) \neq ch(\tilde{x})$ for $c = 2$, $x = 1$.

Definition 0.16 (Linear operator). A **linear operator** on $\mathbb{R}^p$ is a linear map (Definition 0.15) from $\mathbb{R}^p$ to itself, that is, a linear map $f : \mathbb{R}^p \rightarrow \mathbb{R}^p$ whose input and output have the same length. Some authors use “operator” more loosely, for any linear map.

Example 0.13 (A linear operator, and a linear map that is not one). The map $f(\tilde{x}) = (2x_1, x_1 + x_2)$ from Example 0.12 sends $\mathbb{R}^2$ to $\mathbb{R}^2$, so it is a linear operator on $\mathbb{R}^2$. The linear map $g(\tilde{x}) = x_1 + x_2$ sends $\mathbb{R}^2$ to $\mathbb{R}$, so it is linear but not an operator: its outputs have length 1, not 2.

Theorem 0.9 (A linear map sends zero to zero). *If $f : \mathbb{R}^p \rightarrow \mathbb{R}^m$ is a linear map (Definition 0.15), then $f(\tilde{0}) = \tilde{0}$.*

i Proof

Proof. On the left, $\tilde{0}$ is the zero vector of length p ; on the right, it is the zero vector of length m .

$$\begin{aligned} f(\tilde{0}) &= f(\tilde{0} \cdot \tilde{0}) \quad (\text{every entry of } \tilde{0} \cdot \tilde{0} \text{ is } 0 \cdot 0 = 0; \text{ Definition 0.6}) \\ &= \tilde{0} \cdot f(\tilde{0}) \quad (\text{homogeneity, with } c = 0 \text{ and } \tilde{x} = \tilde{0}) \\ &= \tilde{0} \quad (\text{every entry of } \tilde{0} \cdot f(\tilde{0}) \text{ is } 0; \text{ Definition 0.6}) \end{aligned}$$

□

Example 0.14 (Ruling out linearity at zero). In Example 0.12, $g(x) = x + 1$ has $g(\tilde{0}) = 1 \neq \tilde{0}$, so by Theorem 0.9 it is not linear, without checking additivity or homogeneity. Sending $\tilde{0}$ to $\tilde{0}$ is necessary but not sufficient: $h(x) = x^2$ in the same example has $h(\tilde{0}) = \tilde{0}$ and is still not linear.

Theorem 0.10 (Multiplying by a matrix is a linear map). *For any $m \times p$ matrix $\mathbf{A}$, the function $f : \mathbb{R}^p \rightarrow \mathbb{R}^m$ defined by $f(\tilde{x}) = \mathbf{A}\tilde{x}$ is a linear map (Definition 0.15).*

i Proof

Proof. Both conditions are checked one entry at a time. For each $i \in \{1, \dots, m\}$, additivity:

$$\begin{aligned}
(\mathbf{A}(\tilde{x} + \tilde{y}))_i &= \sum_{j=1}^p a_{ij}(\tilde{x} + \tilde{y})_j && \text{(Definition 0.13)} \\
&= \sum_{j=1}^p a_{ij}(x_j + y_j) && \text{(Definition 7 in Vectors)} \\
&= \sum_{j=1}^p (a_{ij}x_j + a_{ij}y_j) && \text{(multiplication distributes over addition)} \\
&= \sum_{j=1}^p a_{ij}x_j + \sum_{j=1}^p a_{ij}y_j && \text{(a sum of sums splits)} \\
&= (\mathbf{A}\tilde{x})_i + (\mathbf{A}\tilde{y})_i && \text{(Definition 0.13)} \\
&= (\mathbf{A}\tilde{x} + \mathbf{A}\tilde{y})_i && \text{(Definition 7 in Vectors)}
\end{aligned}$$

and homogeneity:

$$\begin{aligned}
(\mathbf{A}(c\tilde{x}))_i &= \sum_{j=1}^p a_{ij}(c\tilde{x})_j && \text{(Definition 0.13)} \\
&= \sum_{j=1}^p a_{ij} c x_j && \text{(Definition 0.6)} \\
&= \sum_{j=1}^p c a_{ij} x_j && \text{(products are symmetric)} \\
&= c \sum_{j=1}^p a_{ij} x_j && \text{(a constant factor comes out of a sum)} \\
&= c (\mathbf{A}\tilde{x})_i && \text{(Definition 0.13)} \\
&= (c \mathbf{A}\tilde{x})_i && \text{(Definition 0.6)}
\end{aligned}$$

The four algebra rules cited here are on the [algebra page](#): [distributivity](#), [a sum of sums](#), [symmetry of products](#), and [a constant factor](#). $\square$

Theorem 0.11 (A linear map preserves linear combinations). *If $f : \mathbb{R}^p \rightarrow \mathbb{R}^m$ is a linear map (Definition 0.15), then for any vectors $\tilde{v}_1, \dots, \tilde{v}_k \in \mathbb{R}^p$ and any coefficients $c_1, \dots, c_k \in \mathbb{R}$:*

$$f\left(\sum_{j=1}^k c_j \tilde{v}_j\right) = \sum_{j=1}^k c_j f(\tilde{v}_j)$$

i Proof

Proof. By [induction](#) on k . For $k = 1$, the claim $f(c_1 \underline{v}_1) = c_1 f(\underline{v}_1)$ is homogeneity. Suppose the claim holds for k vectors. Then for $k + 1$:

$$\begin{aligned}
 f\left(\sum_{j=1}^{k+1} c_j \underline{v}_j\right) &= f\left(\sum_{j=1}^k c_j \underline{v}_j + c_{k+1} \underline{v}_{k+1}\right) && \text{(split off the last term)} \\
 &= f\left(\sum_{j=1}^k c_j \underline{v}_j\right) + f(c_{k+1} \underline{v}_{k+1}) && \text{(additivity)} \\
 &= \sum_{j=1}^k c_j f(\underline{v}_j) + f(c_{k+1} \underline{v}_{k+1}) && \text{(induction hypothesis)} \\
 &= \sum_{j=1}^k c_j f(\underline{v}_j) + c_{k+1} f(\underline{v}_{k+1}) && \text{(homogeneity)} \\
 &= \sum_{j=1}^{k+1} c_j f(\underline{v}_j) && \text{(join the last term back on)}
 \end{aligned}$$

□

Theorem 0.12 (Every linear map is multiplication by a matrix). *If $f : \mathbb{R}^p \rightarrow \mathbb{R}^m$ is a linear map (Definition 0.15), then there is exactly one $m \times p$ matrix $\mathbf{A}$ such that*

$$f(\underline{x}) = \mathbf{A} \underline{x} \quad \text{for all } \underline{x} \in \mathbb{R}^p$$

Column j of $\mathbf{A}$ is $f(\underline{e}_j)$, the value of f at the j th indicator vector (Definition 18 in Vectors); that is, $a_{ij} = (f(\underline{e}_j))_i$.

i Proof

Proof. Existence. Let $\mathbf{A}$ be the $m \times p$ matrix with $a_{ij} = (f(\underline{e}_j))_i$. For any $\underline{x} \in \mathbb{R}^p$,

$$\begin{aligned}
 f(\underline{x}) &= f\left(\sum_{j=1}^p x_j \underline{e}_j\right) && \text{(Theorem 4 in Vectors)} \\
 &= \sum_{j=1}^p x_j f(\underline{e}_j) && \text{(Theorem 0.11)}
 \end{aligned}$$

and entry i of that sum is

$$\begin{aligned}
\left(\sum_{j=1}^p x_j f(\underline{e}_j)\right)_i &= \sum_{j=1}^p x_j (f(\underline{e}_j))_i \quad (\text{Definition 9 in Vectors, entry by entry}) \\
&= \sum_{j=1}^p x_j a_{ij} \quad (\text{definition of } \mathbf{A}) \\
&= \sum_{j=1}^p a_{ij} x_j \quad (\text{products are symmetric}) \\
&= (\mathbf{A}\underline{x})_i \quad (\text{Definition 0.13})
\end{aligned}$$

Uniqueness. Suppose $\mathbf{B}$ is any $m \times p$ matrix with $f(\underline{x}) = \mathbf{B}\underline{x}$ for all $\underline{x}$. Taking $\underline{x} = \underline{e}_j$, for each i :

$$\begin{aligned}
(f(\underline{e}_j))_i &= (\mathbf{B}\underline{e}_j)_i \quad (\text{assumption on } \mathbf{B}) \\
&= \sum_{k=1}^p b_{ik} (\underline{e}_j)_k \quad (\text{Definition 0.13}) \\
&= b_{ij} \cdot 1 + \sum_{k \neq j} b_{ik} \cdot 0 \quad (\text{Definition 18 in Vectors, splitting off the } k = j \text{ term}) \\
&= b_{ij} + 0 \quad (\text{times 1 changes nothing; times 0 gives 0}) \\
&= b_{ij} \quad (\text{adding 0 changes nothing})
\end{aligned}$$

So

$$\begin{aligned}
b_{ij} &= (f(\underline{e}_j))_i \\
&= a_{ij}
\end{aligned}$$

for every i and j , and $\mathbf{B} = \mathbf{A}$. □

Definition 0.17 (Matrix of a linear map). The **matrix of** a linear map $f : \mathbb{R}^p \rightarrow \mathbb{R}^m$ is the unique $m \times p$ matrix $\mathbf{A}$ with $f(\underline{x}) = \mathbf{A}\underline{x}$ for all $\underline{x}$ (Theorem 0.12). Its column j is $f(\underline{e}_j)$.

Example 0.15 (Finding the matrix of a linear map). For $f(\underline{x}) = (2x_1, x_1 + x_2)$ from Example 0.12, the values of f at the indicator vectors are

$$f(\underline{e}_1) = f\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 2 \cdot 1 \\ 1 + 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad f(\underline{e}_2) = f\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 2 \cdot 0 \\ 0 + 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

and these are the columns of the matrix of f (Definition 0.17): $\mathbf{A} = \begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix}$.

Corollary 0.1 (Linear functions from vectors to numbers). *A function $f : \mathbb{R}^p \rightarrow \mathbb{R}$ is linear (Definition 0.15) if and only if there is a vector $\underline{w} \in \mathbb{R}^p$ such that*

$$f(\underline{x}) = \underline{w} \cdot \underline{x} \quad \text{for all } \underline{x} \in \mathbb{R}^p$$

and then $w_j = f(\underline{e}_j)$ for each j .

Proof

Proof. If f is linear. Theorem 0.12 with $m = 1$ gives a $1 \times p$ matrix $\mathbf{A}$ with $f(\underline{x}) = \mathbf{A}\underline{x}$ and $a_{1j} = f(\underline{e}_j)$. Let $\underline{w}$ be the vector with $w_j = a_{1j}$, so that $\mathbf{A} = \underline{w}^\top$. Then

$$\begin{aligned} f(\underline{x}) &= \mathbf{A}\underline{x} && \text{(Theorem 0.12)} \\ &= \underline{w} \cdot \underline{x} && (\mathbf{A} = \underline{w}^\top) \end{aligned}$$

and

$$\begin{aligned} w_j &= a_{1j} \\ &= f(\underline{e}_j). \end{aligned}$$

If $f(\underline{x}) = \underline{w}^\top \underline{x}$. Here $\underline{w}^\top$ is a $1 \times p$ matrix, so f is linear by Theorem 0.10. □

Example 0.16 (A linear function from $\mathbb{R}^2$ to $\mathbb{R}$). $f(\underline{x}) = 3x_1 + 5x_2$ is $\underline{w} \cdot \underline{x}$ with $\underline{w} = (3, 5)^\top$, so it is linear by Corollary 0.1. The entries of $\underline{w}$ are the values at the indicator vectors:

$$\begin{aligned} f(\underline{e}_1) &= 3 \cdot 1 + 5 \cdot 0 \\ &= 3 \end{aligned}$$

and

$$\begin{aligned} f(\underline{e}_2) &= 3 \cdot 0 + 5 \cdot 1 \\ &= 5. \end{aligned}$$

$g(\underline{x}) = 3x_1 + 5x_2 + 1$ is not linear, because $g(\underline{0}) = 1 \neq 0$ (Theorem 0.9).

Videos

Two chapters of the [Essence of linear algebra](#) series by 3Blue1Brown show the results above geometrically:

- [Linear transformations and matrices](#) (Chapter 3) shows why a linear map is fixed by where it sends the indicator vectors, which become the columns of its matrix (Definition 0.17).

- [Dot products and duality](#) (Chapter 9) shows why every linear map from $\mathbb{R}^p$ to $\mathbb{R}$ is a dot product with a fixed vector (Corollary 0.1).

Definition 0.18 (Translation). For a fixed vector $\underset{\sim}{b} \in \mathbb{R}^m$, the **translation** [13] by $\underset{\sim}{b}$ is the function $t_{\underset{\sim}{b}} : \mathbb{R}^m \rightarrow \mathbb{R}^m$ defined by

$$t_{\underset{\sim}{b}}(\underset{\sim}{y}) = \underset{\sim}{y} + \underset{\sim}{b}$$

Example 0.17 (Translating two points). The translation by $\underset{\sim}{b} = (1, -2)^\top$ sends

$$t_{\underset{\sim}{b}}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 0+1 \\ 0+(-2) \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}, \quad t_{\underset{\sim}{b}}\left(\begin{bmatrix} 3 \\ 4 \end{bmatrix}\right) = \begin{bmatrix} 3+1 \\ 4+(-2) \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

Both points move one unit right and two units down.

i Remark (A translation moves every point by the same vector)

Remark 0.5 (A translation moves every point by the same vector). A translation by $\underset{\sim}{b}$ moves every point by the same vector $\underset{\sim}{b}$.

Definition 0.19 (Affine map, affine transformation, affine function). A [function](#) $f : \mathbb{R}^p \rightarrow \mathbb{R}^m$ is an **affine map** (also called an **affine transformation** or **affine function**) if it is a matrix product followed by a translation (Definition 0.18) (a [composition](#), with the matrix product as the inner function): there are an $m \times p$ matrix $\mathbf{A}$ and a vector $\underset{\sim}{b} \in \mathbb{R}^m$ such that

$$\begin{aligned} f(\underset{\sim}{x}) &= t_{\underset{\sim}{b}}(\mathbf{A}\underset{\sim}{x}) \\ &= \mathbf{A}\underset{\sim}{x} + \underset{\sim}{b} \quad \text{for all } \underset{\sim}{x} \in \mathbb{R}^p \end{aligned}$$

Definition 0.20 (Offset of an affine map, intercept, bias). In Definition 0.19, the vector $\underset{\sim}{b}$ is called the **offset** of f (statistics also calls it the **intercept**, and machine learning the **bias**).

i Remark (Real-valued affine maps)

Remark 0.6 (Real-valued affine maps). For a real-valued function ([Definition 4 in Vectors](#), $m = 1$), $\mathbf{A}$ is a single row, written $\underset{\sim}{w}^\top$, and $\underset{\sim}{b}$ is a single number b , so an affine function is a linear function $\underset{\sim}{w} \cdot \underset{\sim}{x}$ (Corollary 0.1) plus an offset:

$$f(\underset{\sim}{x}) = \underset{\sim}{w} \cdot \underset{\sim}{x} + b, \quad \underset{\sim}{w} \in \mathbb{R}^p, \quad b \in \mathbb{R} \quad (1)$$

Definition 0.21 (Augmented vector). The **augmented vector** of $\underset{\sim}{x} \in \mathbb{R}^p$ is the vector $\underset{\sim}{x}^{\text{aug}} \in \mathbb{R}^{p+1}$ formed by placing a 1 above the entries of $\underset{\sim}{x}$:

$$\begin{aligned} \underset{\sim}{x}^{\text{aug}} &= \begin{bmatrix} 1 \\ \underset{\sim}{x} \end{bmatrix} \\ &= (1, x_1, \dots, x_p)^\top \end{aligned}$$

Example 0.18 (An affine function as a dot product with an augmented vector). For $\underset{\sim}{x} = (2, 5)^\top$, the augmented vector is $\underset{\sim}{x}^{\text{aug}} = (1, 2, 5)^\top$. With $\underset{\sim}{w} = (3, -1)^\top$ and $b = 4$, the affine function $\underset{\sim}{w} \cdot \underset{\sim}{x} + b$ (Equation 1) equals the dot product of $(b, w_1, w_2)^\top = (4, 3, -1)^\top$ with $\underset{\sim}{x}^{\text{aug}}$:

$$\begin{aligned} \underset{\sim}{w} \cdot \underset{\sim}{x} + b &= 3 \cdot 2 + (-1) \cdot 5 + 4 \\ &= 5, \\ (4, 3, -1)^\top \cdot \underset{\sim}{x}^{\text{aug}} &= 4 \cdot 1 + 3 \cdot 2 + (-1) \cdot 5 \\ &= 5 \end{aligned}$$

Theorem 0.13 (An affine map sends zero to its offset). *An affine map $f(\underset{\sim}{x}) = \mathbf{A}\underset{\sim}{x} + \underset{\sim}{b}$ (Definition 0.19) satisfies $f(\underset{\sim}{0}) = \underset{\sim}{b}$.*

Proof

Proof. First, $\mathbf{A}\underset{\sim}{0} = \underset{\sim}{0}$: for each i ,

$$\begin{aligned} (\mathbf{A}\underset{\sim}{0})_i &= \sum_{j=1}^p a_{ij} \cdot 0 \quad (\text{Definition 0.13}) \\ &= \sum_{j=1}^p 0 \quad (\text{any number times 0 is 0}) \\ &= 0 \quad (\text{a sum of zeros is 0}) \end{aligned}$$

Then

$$\begin{aligned} f(\underset{\sim}{0}) &= \mathbf{A}\underset{\sim}{0} + \underset{\sim}{b} \quad (\text{Definition 0.19}) \\ &= \underset{\sim}{0} + \underset{\sim}{b} \quad (\mathbf{A}\underset{\sim}{0} = \underset{\sim}{0}) \\ &= \underset{\sim}{b} \quad (\text{adding } \underset{\sim}{0} \text{ changes no entry; Definition 7 in Vectors}) \end{aligned}$$

□

Theorem 0.14 (An affine map is linear exactly when its offset is zero). *An affine map $f(\tilde{x}) = \mathbf{A}\tilde{x} + \tilde{b}$ (Definition 0.19) is linear (Definition 0.15) if and only if $\tilde{b} = \tilde{0}$.*

i Proof

Proof. If $\tilde{b} = \tilde{0}$, then

$$\begin{aligned} f(\tilde{x}) &= \mathbf{A}\tilde{x} + \tilde{0} \\ &= \mathbf{A}\tilde{x}, \end{aligned}$$

which is linear by Theorem 0.10. Conversely, if f is linear, then $f(\tilde{0}) = \tilde{0}$ by Theorem 0.9, and $f(\tilde{0}) = \tilde{b}$ by Theorem 0.13, so $\tilde{b} = \tilde{0}$. □

Example 0.19 (Two affine maps with the same matrix). Let $\mathbf{A} = \begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix}$, the matrix of the linear map in Example 0.12. $f(\tilde{x}) = \mathbf{A}\tilde{x}$ has offset $\tilde{0}$, so it is linear. $g(\tilde{x}) = \mathbf{A}\tilde{x} + \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ has offset $\begin{bmatrix} 1 \\ 0 \end{bmatrix} \neq \tilde{0}$, so by Theorem 0.14 it is not linear: indeed $g(\tilde{0}) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, not $\tilde{0}$.

Corollary 0.2 (Every linear map is affine). *Every linear map $f : \mathbb{R}^p \rightarrow \mathbb{R}^m$ (Definition 0.15) is an affine map (Definition 0.19) with offset $\tilde{0}$.*

i Proof

Proof. If f is linear, Theorem 0.12 gives a matrix $\mathbf{A}$ with

$$\begin{aligned} f(\tilde{x}) &= \mathbf{A}\tilde{x} \\ &= \mathbf{A}\tilde{x} + \tilde{0}, \end{aligned}$$

which is Definition 0.19 with $\tilde{b} = \tilde{0}$. □

Theorem 0.15 (An affine map determines its matrix and offset). *If $f(\tilde{x}) = \mathbf{A}\tilde{x} + \tilde{b}$ and $f(\tilde{x}) = \mathbf{A}'\tilde{x} + \tilde{b}'$ for all $\tilde{x} \in \mathbb{R}^p$ (Definition 0.19), then $\mathbf{A}' = \mathbf{A}$ and $\tilde{b}' = \tilde{b}$.*

Proof

Proof. By Theorem 0.13, applied to each formula,

$$\begin{aligned}\tilde{b} &= f(\tilde{0}) \\ &= \tilde{b}'.\end{aligned}$$

Then

$$\begin{aligned}\mathbf{A}\tilde{x} &= f(\tilde{x}) - \tilde{b} \\ &= \mathbf{A}'\tilde{x}\end{aligned}$$

for all $\tilde{x}$, so $\mathbf{A}$ and $\mathbf{A}'$ are both matrices (Definition 0.17) of the linear map $\tilde{x} \mapsto f(\tilde{x}) - \tilde{b}$ (linear by Theorem 0.10), and the uniqueness in Theorem 0.12 gives $\mathbf{A}' = \mathbf{A}$. $\square$

Remark (“Linear” in this site, and in machine learning and statistics)

Remark 0.7 (“Linear” in this site, and in machine learning and statistics). This site uses “linear” only in the sense of Definition 0.15, so by Theorem 0.14 a function with a nonzero offset is affine but not linear. For functions of one variable, this is the same split as [affine](#) $ax + b$ versus [linear](#) ax on the algebra page. (The algebra page writes these as $mx + b$ and mx ; here m already names the number of outputs.)

Machine learning and statistics often call any function of the form Equation 1 “linear”, as in a *linear model* in statistics or a *linear layer* in a neural network, and call the offset b the bias (machine learning) or the intercept (statistics), as Definition 0.19 notes. A model with an intercept is affine in $\tilde{x}$ but linear in its coefficients $(b, \tilde{w})$, and that is the sense of “linear” in “linear model”: it stays linear in the coefficients even when the inputs are transformed: $w_1x + w_2x^2 + b$ is not affine in x , but it is linear in (b, w_1, w_2) . Separately, the affine function $\tilde{w} \cdot \tilde{x} + b$ is a linear function of the augmented vector $\tilde{x}^{\text{aug}}$ (Definition 0.21), with b as the coefficient on its first entry.

Further reading

- Zhou [14], sections 6 (“Linear functions and operators”; see Definition 0.16) and 7 (“Affine functions”), proves that every linear function is a matrix-vector product by the same route as Theorem 0.12, and that every affine function has the form $\mathbf{A}\tilde{x} + \tilde{b}$. It defines “affine” differently, by $f(\alpha\tilde{x} + \beta\tilde{y}) = \alpha f(\tilde{x}) + \beta f(\tilde{y})$ whenever $\alpha + \beta = 1$, and shows that this condition holds exactly for the maps $\mathbf{A}\tilde{x} + \tilde{b}$, so the two definitions agree.
- Boyd and Vandenberghe [15], chapter 2 (“Linear functions”), covers the vector-to-number case: linear functions as inner products, affine functions, the [first-order Taylor approximation](#) as an affine function, and the regression model. Section 8.1

(“Linear and affine functions”) covers the vector-to-vector case. A free PDF is on the book’s website.

Definition 0.22 (Hyperplane). A **hyperplane** [16] in $\mathbb{R}^p$ is the set of points $\tilde{x} \in \mathbb{R}^p$ satisfying

$$\tilde{w} \cdot \tilde{x} + b = 0$$

for some vector $\tilde{w} \in \mathbb{R}^p \setminus \{0\}$ and number $b \in \mathbb{R}$. The number b is the offset of the hyperplane.

Definition 0.23 (Normal vector of a hyperplane). In Definition 0.22, the vector $\tilde{w}$ is a **normal vector** [17] of the hyperplane.

Example 0.20 (A line in $\mathbb{R}^2$, and why the normal vector must be nonzero).

- With $w = (1, 1)$ and $b = -1$, the hyperplane in $\mathbb{R}^2$ is the line $x_1 + x_2 - 1 = 0$. The point $(1, 0)$ is on it, since $1 + 0 - 1 = 0$; the point $(1, 1)$ is not, since $1 + 1 - 1 = 1 > 0$, so it lies in the half-space where $\tilde{w} \cdot \tilde{x} + b > 0$.
- With $\tilde{w} = 0$, the equation reads $b = 0$: if $b \neq 0$ no point satisfies it, and if $b = 0$ every point does, so the set is empty or all of $\mathbb{R}^2$, not a line. That degenerate case is why the definition requires $\tilde{w} \neq 0$.

The course website’s version of this page has an interactive plot here.

Exercise 0.2 (Write a rule in matrix form). A clinic scores a patient’s risk from three measurements.

symbol	meaning	units
x_1	age	years
x_2	systolic blood pressure	mmHg
x_3	body mass index	kg/m ²

The rule in use is

$$\text{score} = 0.4 x_1 + 0.2 x_2 + 1.5 x_3 - 30$$

Write this rule in the form of Equation 1, naming $\tilde{w}$, b and p explicitly.

Solution

Solution 0.2. Read the [coefficients](#) straight off the rule. There are three inputs, so $p = 3$:

$$\underset{\sim}{w} = \begin{bmatrix} 0.4 \\ 0.2 \\ 1.5 \end{bmatrix}, \quad b = -30$$

and the rule is $f(\underset{\sim}{x}) = \underset{\sim}{w}^\top \underset{\sim}{x} + b$.

Exercise 0.3 (Score a patient). Give the score from the rule in Exercise [0.2](#) for a patient aged 50 with blood pressure 130 and BMI 28.

Solution

Solution 0.3. For the patient,

$$\underset{\sim}{x} = \begin{bmatrix} 50 \\ 130 \\ 28 \end{bmatrix}$$

$$\begin{aligned} \underset{\sim}{w}^\top \underset{\sim}{x} &= (0.4)(50) + (0.2)(130) + (1.5)(28) \\ &= 20 + 26 + 42 \\ &= 88 \end{aligned}$$

$$\begin{aligned} f(\underset{\sim}{x}) &= 88 - 30 \\ &= 58 \end{aligned}$$

Two things are worth noticing. The [coefficients](#) are not comparable as they stand. The [weight](#) on BMI, 1.5, is 7.5 times the weight on blood pressure, 0.2, but the contribution of BMI, 42, is less than twice the contribution of blood pressure, 26, because the BMI [value](#), 28, is much smaller than the blood pressure value, 130. So the weight on a feature says nothing on its own about how much that feature contributes to the score. And $b = -30$ changes the score: without it, the score would be 88 instead of 58.

Exercise 0.4 (Which of these are affine in $\underset{\sim}{w}$?). Each of the five expressions is a function of the weights $\underset{\sim}{w} \in \mathbb{R}^p$, with the data $\underset{\sim}{x}$ and y held fixed. Say which ones are affine (Definition [0.19](#)), that is, have the form Equation [1](#) — an inner product plus an offset — and which of those are also linear (Definition [0.15](#)).

1. $\underset{\sim}{w}^\top \underset{\sim}{x}$
2. $\underset{\sim}{w}^\top \underset{\sim}{x} - y$
3. $(\underset{\sim}{w}^\top \underset{\sim}{x} - y)^2$

4. $\|\tilde{w}\|_2^2$
5. $w_1x_1 + 3$

Solution

Solution 0.4.

1. **Affine, and linear.** It is Equation 1 with $b = 0$, which is a linear map by Theorem 0.10: the matrix $\tilde{x}^\top$ times $\tilde{w}$.
2. **Affine.** Still an inner product plus an offset; here the offset is $b = -y$, which is a constant because y is held fixed. It is linear only when $y = 0$: otherwise $\tilde{w} = \tilde{0}$ gives $-y \neq 0$, and a linear map sends 0 to 0 (Theorem 0.9).
3. **Not affine.** Squaring is not an inner product plus an offset. This expression is the squared error of one observation, and its curvature in $\tilde{w}$ is exactly why fitting a model is an optimization problem rather than a lookup.
4. **Not affine.** By Equation 2 in Vectors this is $\tilde{w}^\top \tilde{w}$, which has $\tilde{w}$ in both slots at once. Hold either slot fixed and it is linear in the other; as a function of the single variable $\tilde{w}$ it is quadratic.
5. **Affine, not linear.** It is Equation 1 with coefficient vector $\tilde{a} = (x_1, 0, \dots, 0)^\top$ and $b = 3$; at $\tilde{w} = \tilde{0}$ it equals $3 \neq 0$, so it is not linear (Theorem 0.9). The data $\tilde{x}$ is untouched — $\tilde{a}$ is a new vector built from its first entry.

Items 1, 2 and 5 are affine and items 3 and 4 are not, which is the split that matters: a *model* can be affine in its weights while the *quantity we minimize* is not. Squared error and the norm penalty are both curved in $\tilde{w}$, and that curvature is what makes fitting an optimization problem at all: there is always a downhill direction to follow. Curvature alone does not promise a *single* answer. The squared error of one observation is completely flat along a whole line of weight vectors, every one of which fits that observation exactly, and pinning down one of them is part of what the norm penalty is added for in week 3.

Transposes of sums and products

Theorem 0.16 (Transpose of a sum). *For $m \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$:*

$$(\mathbf{A} + \mathbf{B})^\top = \mathbf{A}^\top + \mathbf{B}^\top$$

In particular, for column vectors $\tilde{x}$ and $\tilde{y}$ of the same length:

$$(\tilde{x} + \tilde{y})^\top = \tilde{x}^\top + \tilde{y}^\top$$

i Proof

Proof. Entry (i, j) of each side:

$$\begin{aligned}
 [(\mathbf{A} + \mathbf{B})^\top]_{ij} &= (\mathbf{A} + \mathbf{B})_{ji} && \text{(definition of the transpose)} \\
 &= a_{ji} + b_{ji} && \text{(definition of matrix addition)} \\
 &= (\mathbf{A}^\top)_{ij} + (\mathbf{B}^\top)_{ij} && \text{(definition of the transpose)} \\
 &= [\mathbf{A}^\top + \mathbf{B}^\top]_{ij} && \text{(definition of matrix addition)}
 \end{aligned}$$

□

Theorem 0.17 (Transpose of a product). *For an $m \times k$ matrix $\mathbf{A}$ and a $k \times n$ matrix $\mathbf{B}$:*

$$\underbrace{(\mathbf{AB})^\top}_{n \times m} = \underbrace{\mathbf{B}^\top}_{n \times k} \underbrace{\mathbf{A}^\top}_{k \times m}$$

i Proof

Proof. Entry (i, j) of each side:

$$\begin{aligned}
 [(\mathbf{AB})^\top]_{ij} &= (\mathbf{AB})_{ji} && \text{(definition of the transpose)} \\
 &= \sum_{s=1}^k a_{js} b_{si} && \text{(definition of matrix multiplication)} \\
 &= \sum_{s=1}^k (\mathbf{B}^\top)_{is} (\mathbf{A}^\top)_{sj} && \text{(definition of the transpose; reorder each product)} \\
 &= [\mathbf{B}^\top \mathbf{A}^\top]_{ij} && \text{(definition of matrix multiplication)}
 \end{aligned}$$

□

i Remark (The order of the factors reverses)

Remark 0.8 (The order of the factors reverses). Transposing a product reverses the order of the factors. Keeping the original order usually gives an undefined product or a different matrix. For example, with $\mathbf{A} = [1, 2]$ (1×2) and $\mathbf{B} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$ (2×1),

$$\begin{aligned}
 \mathbf{AB} &= 1 \cdot 3 + 2 \cdot 4 \\
 &= 11,
 \end{aligned}$$

so $(\mathbf{AB})^\top = 11$, and

$$\mathbf{B}^\top \mathbf{A}^\top = [3, 4] \begin{bmatrix} 1 \\ 2 \end{bmatrix} = 11, \quad \mathbf{A}^\top \mathbf{B}^\top = \begin{bmatrix} 1 \\ 2 \end{bmatrix} [3, 4] = \begin{bmatrix} 3 & 4 \\ 6 & 8 \end{bmatrix}.$$

If instead $\mathbf{A}$ is 2×3 and $\mathbf{B}$ is 3×1 , then $\mathbf{B}^\top \mathbf{A}^\top$ is a $(1 \times 3)(3 \times 2)$ product, but $\mathbf{A}^\top \mathbf{B}^\top$ is a $(3 \times 2)(1 \times 3)$ product, which is undefined.

Theorem 0.18 (A 1×1 product equals its own transpose). *If $\underline{u}$ and $\underline{v}$ are vectors of length p and $\mathbf{M}$ is a $p \times p$ matrix, then $\underline{u}^\top \mathbf{M} \underline{v}$ is a number, and*

$$\underline{u}^\top \mathbf{M} \underline{v} = \underline{v}^\top \mathbf{M}^\top \underline{u}.$$

In particular, $\underline{u}^\top \underline{v} = \underline{v}^\top \underline{u}$.

i Proof

Proof. The product $\underline{u}^\top \mathbf{M} \underline{v}$ is a $(1 \times p)(p \times p)(p \times 1)$ product, so it is a 1×1 matrix. The transpose of a 1×1 matrix is itself, so

$$\begin{aligned} \underline{u}^\top \mathbf{M} \underline{v} &= (\underline{u}^\top \mathbf{M} \underline{v})^\top && \text{(a } 1 \times 1 \text{ matrix equals its transpose)} \\ &= \underline{v}^\top \mathbf{M}^\top (\underline{u}^\top)^\top && \text{(transpose of a product, twice)} \\ &= \underline{v}^\top \mathbf{M}^\top \underline{u} && \text{(transposing twice changes nothing)} \end{aligned}$$

The transpose step is Theorem 0.17. Taking $\mathbf{M} = \mathbf{I}$ gives the last statement. $\square$

Example 0.21 (Two ways to write the same number). Take $\underline{u} = (1, 2)$, $\underline{v} = (3, -1)$ and $\mathbf{M} = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}$. Then $\mathbf{M} \underline{v} = (1, -3)$, so

$$\underline{u}^\top \mathbf{M} \underline{v} = 1 \cdot 1 + 2 \cdot (-3) = -5.$$

Also $\mathbf{M}^\top = \begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix}$ and $\mathbf{M}^\top \underline{u} = (1, 8)$, so

$$\underline{v}^\top \mathbf{M}^\top \underline{u} = 3 \cdot 1 + (-1) \cdot 8 = -5.$$

The two products are equal, as the theorem says. In least squares, this is why the two cross terms of $(\mathbf{X} \underline{\beta} - \underline{y})^\top (\mathbf{X} \underline{\beta} - \underline{y})$ are equal and can be combined.

- [1] Wikipedia contributors, “Matrix (mathematics) — Wikipedia, the free encyclopedia.” 2026. Available: [https://en.wikipedia.org/wiki/Matrix_\(mathematics\)](https://en.wikipedia.org/wiki/Matrix_(mathematics))
- [2] Wikipedia contributors, “Tensor — Wikipedia, the free encyclopedia.” 2026. Available: <https://en.wikipedia.org/wiki/Tensor>

- [3] Wikipedia contributors, “Transpose — Wikipedia, the free encyclopedia.” 2026. Available: <https://en.wikipedia.org/wiki/Transpose>
- [4] Wikipedia contributors, “Zero matrix — Wikipedia, the free encyclopedia.” 2026. Available: https://en.wikipedia.org/wiki/Zero_matrix
- [5] Wikipedia contributors, “Matrix addition — Wikipedia, the free encyclopedia.” 2026. Available: https://en.wikipedia.org/wiki/Matrix_addition
- [6] Wikipedia contributors, “Scalar multiplication — Wikipedia, the free encyclopedia.” 2026. Available: https://en.wikipedia.org/wiki/Scalar_multiplication
- [7] Wikipedia contributors, “Matrix multiplication — Wikipedia, the free encyclopedia.” 2026. Available: https://en.wikipedia.org/wiki/Matrix_multiplication
- [8] Wikipedia contributors, “Square matrix — Wikipedia, the free encyclopedia.” 2026. Available: https://en.wikipedia.org/wiki/Square_matrix
- [9] Wikipedia contributors, “Main diagonal — Wikipedia, the free encyclopedia.” 2026. Available: https://en.wikipedia.org/wiki/Main_diagonal
- [10] Wikipedia contributors, “Identity matrix — Wikipedia, the free encyclopedia.” 2026. Available: https://en.wikipedia.org/wiki/Identity_matrix
- [11] Wikipedia contributors, “System of linear equations — Wikipedia, the free encyclopedia.” 2026. Available: https://en.wikipedia.org/wiki/System_of_linear_equations
- [12] Wikipedia contributors, “Linear map — Wikipedia, the free encyclopedia.” 2026. Available: https://en.wikipedia.org/wiki/Linear_map
- [13] Wikipedia contributors, “Translation (geometry) — Wikipedia, the free encyclopedia.” 2026. Available: [https://en.wikipedia.org/wiki/Translation_\(geometry\)](https://en.wikipedia.org/wiki/Translation_(geometry))
- [14] H. Zhou, “Matrices.” Lecture notes for Biostat 216, Mathematical Methods for Biostatistics, University of California, Los Angeles, 2024. Accessed: Oct. 05, 2026. [Online]. Available: <https://ucla-biostat-216.github.io/2024fall/slides/03-matrix/03-matrix.html>
- [15] S. Boyd and L. Vandenberghe, *Introduction to applied linear algebra: Vectors, matrices, and least squares*. Cambridge University Press, 2018. doi: [10.1017/9781108583664](https://doi.org/10.1017/9781108583664).
- [16] Wikipedia contributors, “Hyperplane — Wikipedia, the free encyclopedia.” 2026. Available: <https://en.wikipedia.org/wiki/Hyperplane>
- [17] Wikipedia contributors, “Normal (geometry) — Wikipedia, the free encyclopedia.” 2026. Available: [https://en.wikipedia.org/wiki/Normal_\(geometry\)](https://en.wikipedia.org/wiki/Normal_(geometry))