Data 345

Applied Linear Algebra for Statistical Learning

Class 10 (Sep. 29, 2025)

Linear Transformations & Matrices

- Last time: we introduced a linear transformation $f_\theta: \mathbb{R}^2 \to \mathbb{R}^2$ which took vectors in $\mathbb{R}^2$ and rotated them θ radians counterclockwise.
- We noted that determining a formula for f_θ boiled down to the effect of f_θ on basis elements of $\mathbb{R}^2$.
- Let $T: V \to W$ be a linear transformation between vector spaces V and W, with $\dim(V) = n$ and $\dim(W) = m$. Let $B = (\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n)$ and $C = (\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_m)$ be two <u>ordered</u> bases for V and W.
- Since T sends vectors in V to vectors in W, then for any $j \in \{1, 2, \dots, n\}$, the vector $T(\mathbf{b}_j)$ may be expressed as a linear combination of basis elements from C, i.e.:

$$T(\mathbf{b}_j) = \alpha_{j1}\mathbf{c}_1 + \alpha_{j2}\mathbf{c}_2 + \cdots + \alpha_{jm}\mathbf{c}_m$$

- These coefficients can be arranged into a matrix A_T, where $A_T[i, j] = \alpha_{ij}$.

Linear Transformations & Matrices

- The matrix A_T is called the **transformation matrix** of T with respect to B and C.
- Since linear transformations are entirely determined by their effect on basis elements, a transformation essentially just summarizes the effect of T on each individual basis element from B in terms of the basis C.
- Example: Define a linear mapping $T: \mathbb{R}^3 \to \mathbb{R}^4$ as:

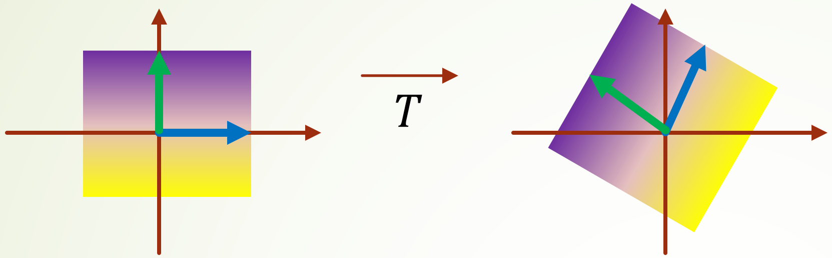
$$T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} 3x_1 + 2x_2 + x_3 \\ x_1 + x_2 + x_3 \\ x_1 - 3x_2 \\ 2x_1 + 3x_2 + x_3 \end{bmatrix}$$

$$T\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 3 \\ 1 \\ 1 \\ 2 \end{bmatrix} \qquad T\left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 1 \\ -3 \\ 3 \end{bmatrix} \qquad T\left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} \qquad A_T = \begin{bmatrix} 3 & 2 & 1 \\ 1 & 1 & 1 \\ 1 & -3 & 0 \\ 2 & 3 & 1 \end{bmatrix}$$

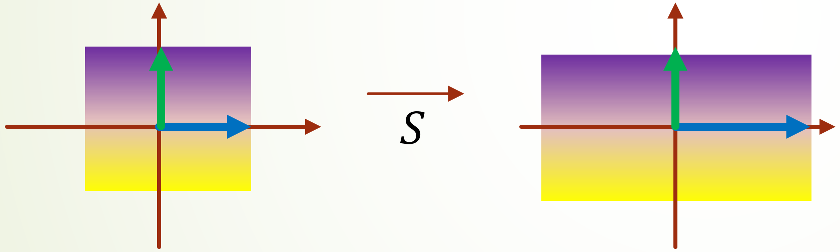
Visualizing Linear Transformations

- Since a linear transformation's behavior is entirely determined by its effect on basis vectors, one strategy for visualizing linear transformations (particularly from $\mathbb{R}^2$ to $\mathbb{R}^2$) is as follows:
- Draw two separate coordinate axes.
- Draw two arrows for the standard basis vectors $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ in the first pair of coordinate axes.
- Draw a dense collection of points around the unit square in the first pair of coordinate axes.
- Draw $T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right)$ and $T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)$ as vectors in the second pair of axes.
- Draw the images of the individual points from the unit square in the second pair of coordinate axes.

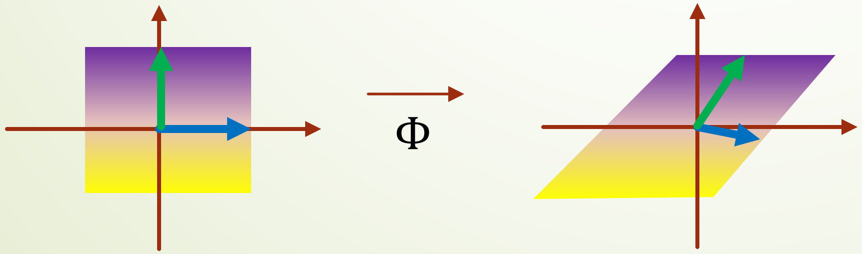
Visualizing Linear Transformations



For $T: \mathbb{R}^2 \to \mathbb{R}^2$ where: $A_T = \begin{bmatrix} \cos\frac{\pi}{3} & -\sin\frac{\pi}{3} \\ \sin\frac{\pi}{3} & \cos\frac{\pi}{3} \end{bmatrix}$



For $S: \mathbb{R}^2 \to \mathbb{R}^2$ where $A_S = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$



For a general linear transformation
$\Phi: \mathbb{R}^2 \to \mathbb{R}^2$

Linear Transformations & Basis

- **Caution:** Matrices for linear transformations are always made with respect to <u>an ordered basis</u> for both the domain and the codomain. Choosing a different basis for V or W (or even a different ordering) leads to a different matrix.
- Example: Consider the transformation $S: \mathbb{R}^2 \to \mathbb{R}^2$ that simply stretches in the horizontal direction by a factor of 2 and leaves the vertical direction unchanged. That is, $S\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 2x_1 \\ x_2 \end{bmatrix}$.
- Letting $B = C = \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)$, we get:

$$\begin{aligned} S\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) &= \begin{bmatrix} 2 \\ 0 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 0\begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ S\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) &= \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 1\begin{bmatrix} 0 \\ 1 \end{bmatrix} \end{aligned} \longrightarrow A_S = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

- Letting $C = \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)$, however, gives:

$$\begin{aligned} S\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) &= \begin{bmatrix} 2 \\ 0 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 0\begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ S\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) &= \begin{bmatrix} 0 \\ 1 \end{bmatrix} = -1\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 1\begin{bmatrix} 1 \\ 1 \end{bmatrix} \end{aligned} \longrightarrow \widetilde{A_S} = \begin{bmatrix} 2 & -1 \\ 0 & 1 \end{bmatrix}$$

Linear Transformation & Basis

- All this is to say that <u>different matrices may represent the same linear transformation</u>.
- This would be genuinely awful news under most circumstances because it means that the assignment of transformation matrices is not well-defined.
- Let $T: V \to W$ be a linear transformation between two vector spaces. Suppose that $B, \tilde{B}$ and $C, \tilde{C}$ are two ordered bases for V and W, respectively.

$$\begin{tikzcd} V \arrow[r, "T"'] & W \end{tikzcd}$$

Create a linear transformation $V \to V$ which simply "converts" the basis elements of $\tilde{B}$ to basis elements of B. Its transformation matrix is P.

$$\begin{tikzcd} B \arrow[r, "A_T"'] & C \\ \tilde{B} \arrow[u, "P"] \arrow[r, "\widetilde{A_T}"'] \arrow[ur, phantom, "\widetilde{A_T} = Q^{-1} A_T P"] & \tilde{C} \arrow[u, "Q"'] \end{tikzcd}$$

Create a linear transformation $W \to W$ which converts the basis elements of $\tilde{C}$ to basis elements of C. Its transformation matrix is Q.

Equivalent Linear Transformations

- Let $A, B \in \mathbb{R}^{m \times n}$. Then A and B are said to be **equivalent** matrices if there exists invertible $Q \in \mathbb{R}^{m \times m}$ and invertible $P \in \mathbb{R}^{n \times n}$ such that $B = Q^{-1}AP$.
- Two matrices are equivalent if, and only if, they *represent the same linear transformation*.
- Under the equation $B = Q^{-1}AP$, the matrix A represents a linear transformation from V to W under one basis, while B represents the same linear transformation using a different basis. The matrices P and Q are called **change-of-basis matrices**.
- Now: **why should we care** about this? After all, each vector space has its own standard basis – why not just represent every transformation with respect to its standard bases?

Equivalent Linear Transformations

- Insisting on using the same bases every time makes some degree of sense for defining matrices, but it's incredibly restrictive and limits real solving power.
- Here's an extremely simple example of how change-of-basis speeds up computation.
- Consider a matrix $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ corresponding to a linear transformation of the standard basis vectors of $\mathbb{R}^2$.
- Instead of the basis $B = \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)$, let's consider a different basis $\tilde{B} = \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix}\right)$.

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \end{bmatrix} = 3\begin{bmatrix} 1 \\ 1 \end{bmatrix} + 0\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 0\begin{bmatrix} 1 \\ 1 \end{bmatrix} + 1\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$\tilde{A} = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$$

Equivalent Linear Transformations

- The big difference between A and $\tilde{A}$ is that $\tilde{A}$ is a **diagonal** matrix: its only nonzero entries lie on the main diagonal.
- If we wanted to repeatedly apply the transformation, we would chain together the transformation (i.e., multiply the matrices by themselves multiple times).

$$A^2 = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} (2 \cdot 2) + (1 \cdot 1) & (2 \cdot 1) + (1 \cdot 2) \\ (1 \cdot 2) + (2 \cdot 1) & (1 \cdot 1) + (2 \cdot 2) \end{bmatrix} = \begin{bmatrix} 4 + 1 & 2 + 2 \\ 2 + 2 & 1 + 4 \end{bmatrix} = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$$

$$A^3 = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix} \cdot \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} (5 \cdot 2) + (4 \cdot 1) & (5 \cdot 1) + (4 \cdot 2) \\ (4 \cdot 2) + (5 \cdot 1) & (4 \cdot 1) + (5 \cdot 2) \end{bmatrix} = \begin{bmatrix} 10 + 4 & 5 + 8 \\ 8 + 5 & 4 + 10 \end{bmatrix} = \begin{bmatrix} 14 & 13 \\ 13 & 14 \end{bmatrix}$$

$$A^4 = \begin{bmatrix} 14 & 13 \\ 13 & 14 \end{bmatrix} \cdot \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} (14 \cdot 2) + (13 \cdot 1) & (14 \cdot 1) + (13 \cdot 2) \\ (13 \cdot 2) + (14 \cdot 1) & (13 \cdot 1) + (14 \cdot 2) \end{bmatrix} = \begin{bmatrix} 28 + 13 & 14 + 26 \\ 26 + 14 & 13 + 28 \end{bmatrix} = \begin{bmatrix} 41 & 40 \\ 40 & 41 \end{bmatrix}$$

$$\tilde{A}^2 = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} (3 \cdot 3) + (0 \cdot 0) & (3 \cdot 0) + (0 \cdot 1) \\ (0 \cdot 3) + (1 \cdot 0) & (0 \cdot 0) + (1 \cdot 1) \end{bmatrix} = \begin{bmatrix} 9 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\tilde{A}^3 = \begin{bmatrix} 9 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} (9 \cdot 3) + (0 \cdot 0) & (9 \cdot 0) + (0 \cdot 1) \\ (0 \cdot 3) + (1 \cdot 0) & (0 \cdot 0) + (1 \cdot 1) \end{bmatrix} = \begin{bmatrix} 27 & 0 \\ 0 & 1 \end{bmatrix}$$

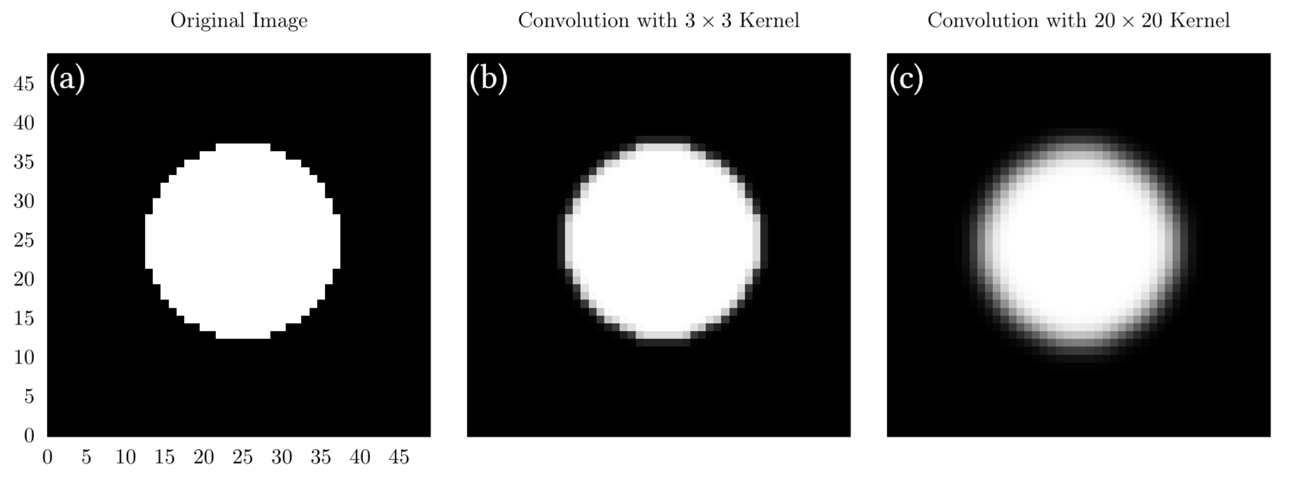
Diagonal vs. Non-Diagonal

- If you look carefully at the entries in $\tilde{A}^k$, you can see that:

$$\tilde{A}^k = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}^k = \begin{bmatrix} 3^k & 0 \\ 0 & 1^k \end{bmatrix}$$

- In other words, to compute a high power of a diagonal matrix all you need to do is compute the exponentiated diagonals.
- If $A, B \in \mathbb{R}^{n\times n}$, then matrix multiplication – computing AB – is $\mathcal{O}(n^3)$. So to compute A^k (knowing that A is not a diagonal matrix) would have a time complexity of $\mathcal{O}(k \cdot n^3)$. For large powers of k, it would be extremely slow.
- On the other hand, for a diagonal matrix, one only needs to compute n separate exponentiated real numbers. So, if $D \in \mathbb{R}^{n\times n}$ is diagonal, then computing D^k is $\mathcal{O}(n \cdot \log k)$ instead.
- This is such a huge difference in time that even though it is slow and expensive to find a "diagonal" equivalent to a matrix it is **often worth the trade-off** by an order of magnitude.

Application – Image Processing



Applying a “blur” filter to an image requires a type of transformation called a **convolution**, which distributes pixel intensity values based on a sort of weighted average of pixel values of neighboring pixel values.

For an $N \times M$ image, “direct” convolution (i.e. using the standard basis elements) would take $O(NMnm)$ time. Taking the time to diagonalize first changes the operation to $O(NM \log NM)$ time instead.

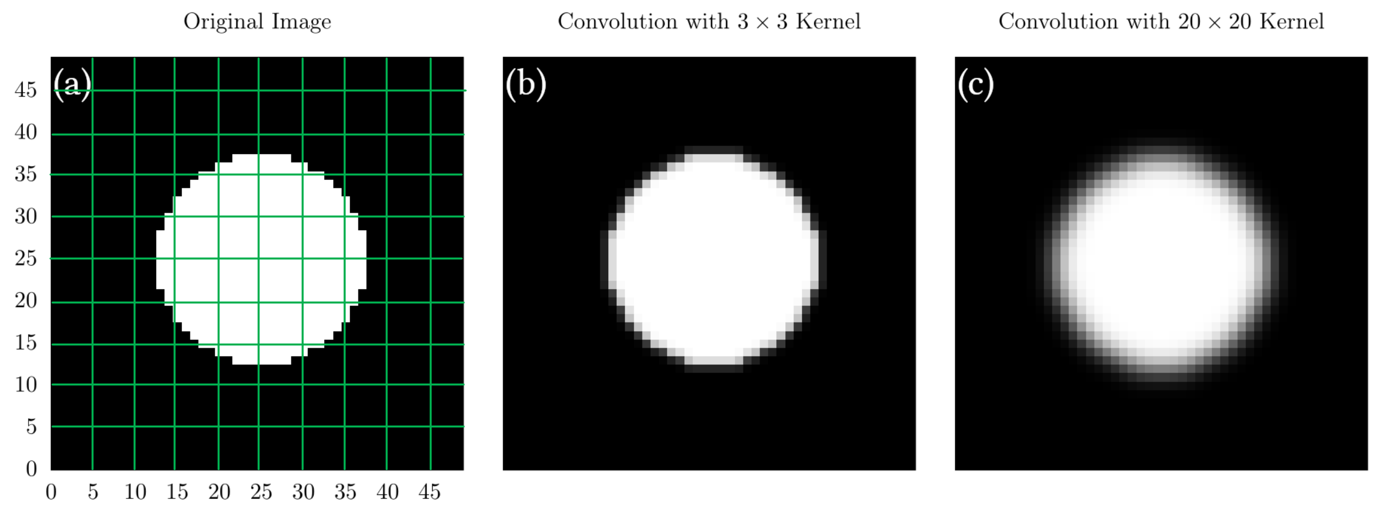
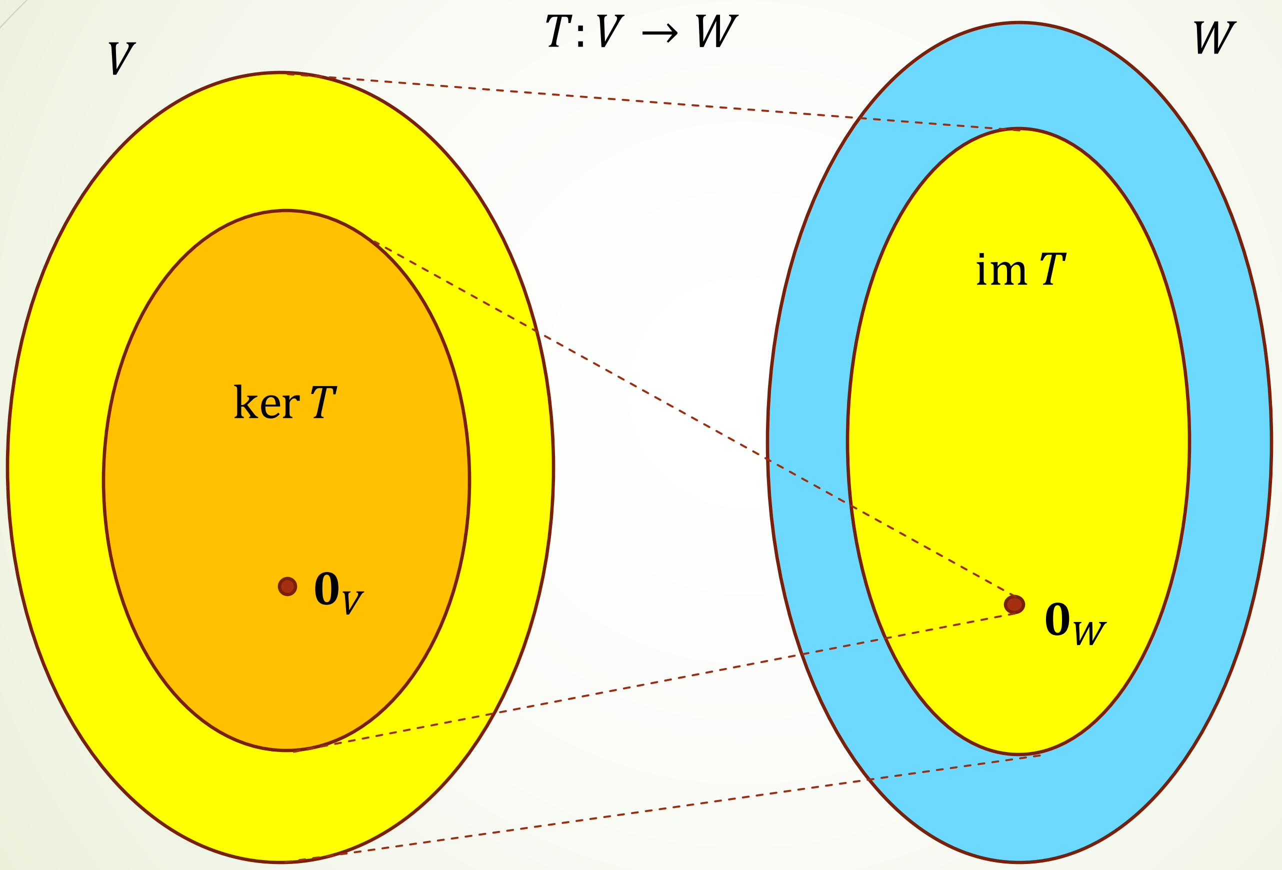


Image & Kernel

- Let $T: V \to W$ be a linear transformation. The **image** or **range** of T is simply $T(V)$. The **kernel** or **null space** of T is the set of all vectors $\mathbf{v}$ in V such that $T(\mathbf{v}) = \mathbf{0}_W$.
- In other words, the kernel of T is the set of vectors in V that get annihilated (sent to the zero vector) by the transformation T. The image of T is often written $\operatorname{im} T$, while the kernel is often written $\ker T$.
- For any linear transformation $T: V \to W$, we have that $T(\mathbf{0}_V) = \mathbf{0}_W$ so the null space / kernel is never empty.
- Because T is linear, the image of T is always a subspace of W, while the kernel of T is always a subspace of V.

Image & Kernel



The null space $\ker T$ is the part of the space V that is squashed into the zero vector.

The image $\operatorname{im} T$ is the version of V that has been successfully "transferred" to W.