Orthogonality

Orthonormal Bases, Orthogonal Matrices

An Orthonormal Basis

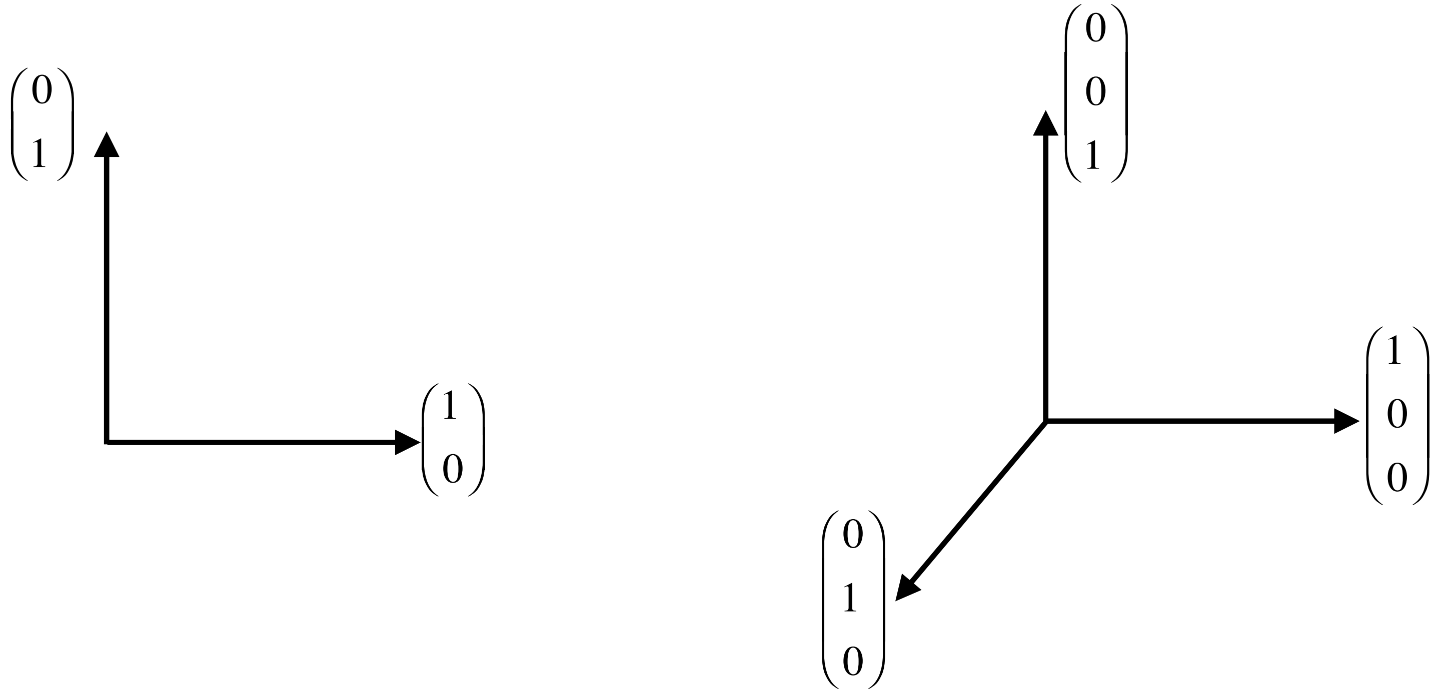
- Implicit in our previous discussion was the idea of an **orthonormal basis**.

- A collection of vectors is **orthonormal** if they are mutually **orthogonal** (perpendicular) and every vector in the collection is a **unit vector** (has *length* 1. $||\mathbf{x}||=1$)

An Orthonormal Basis

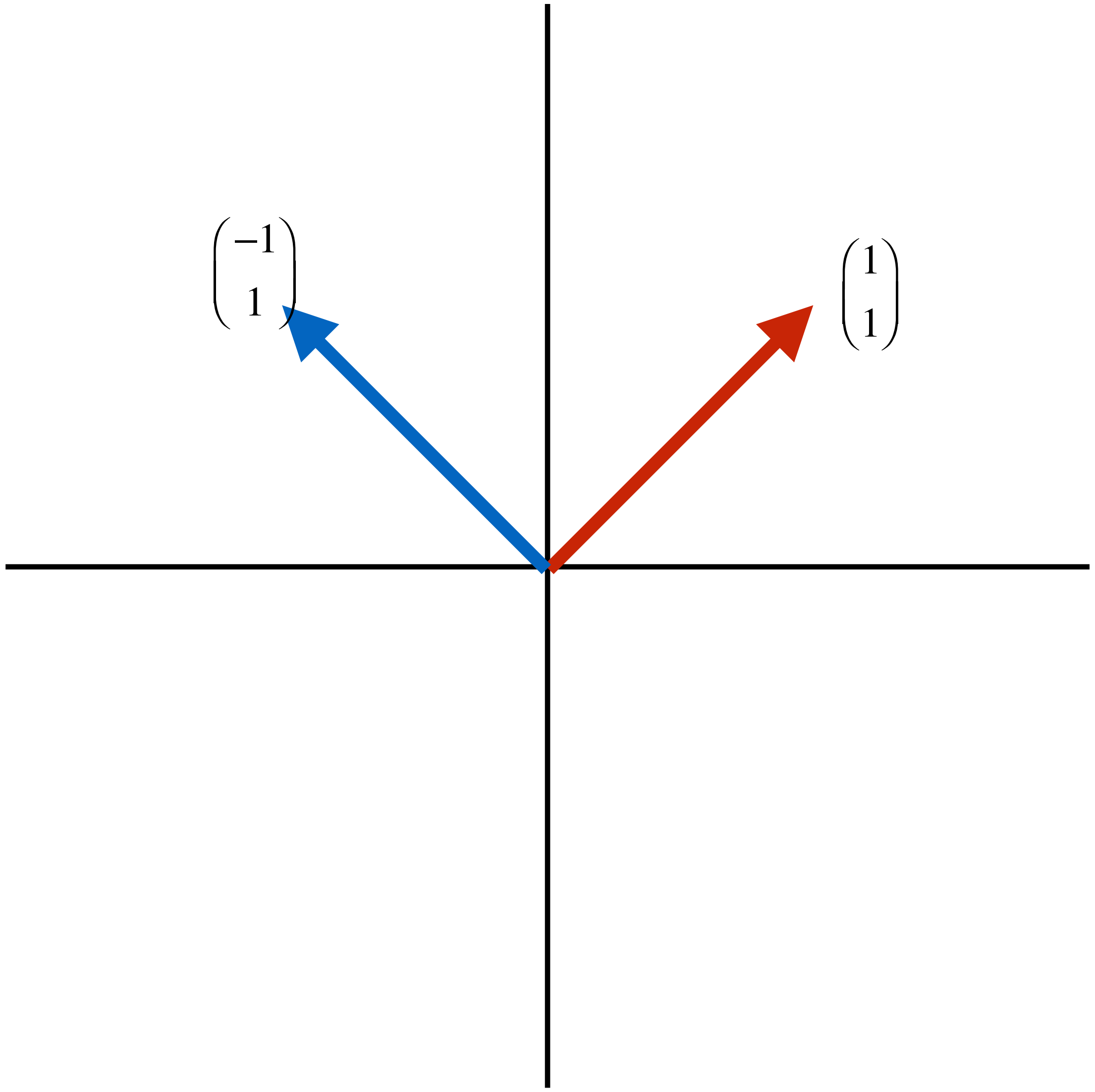
Easiest example of an orthonormal basis?

The elementary basis vectors!



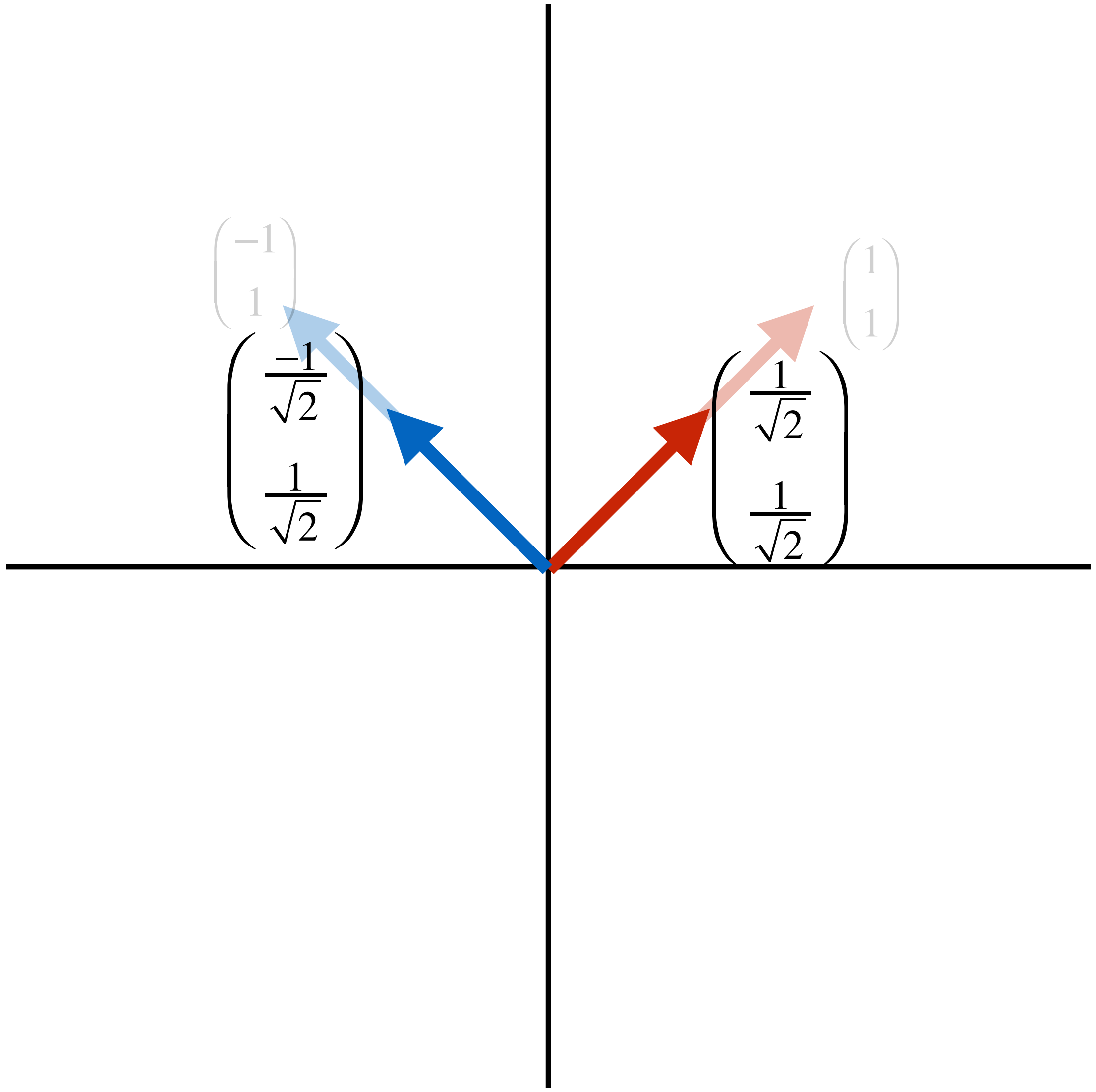
An Orthonormal Basis

What if I wanted to change the basis to the red and blue ***directions*** shown? (I still want it to be orthonormal)



An Orthonormal Basis

Orthonormal basis with the same directions, but now *normalized* so basis vectors have unit length.

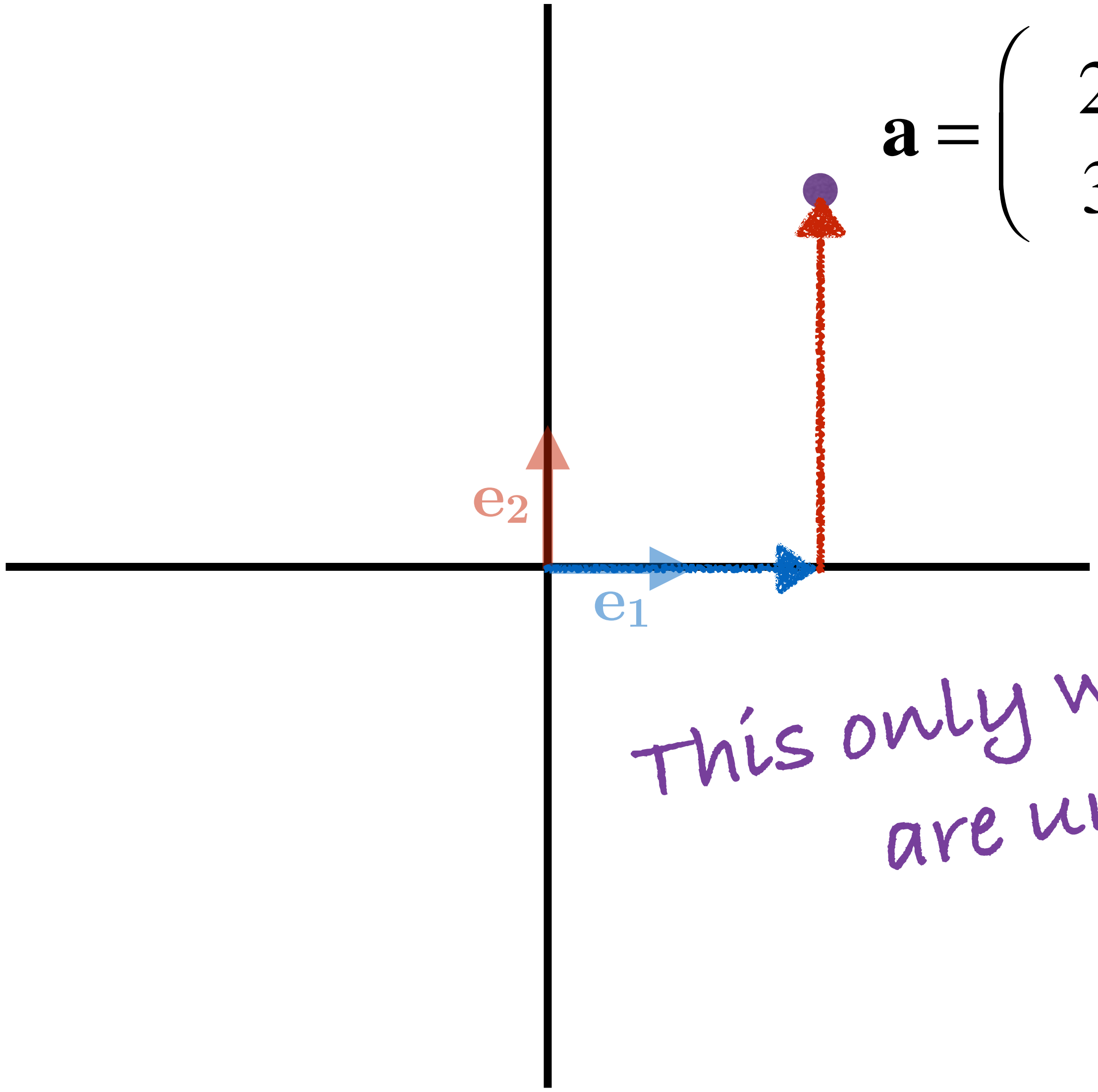


Why make a point about this?

- There are **infinitely many basis vectors to specify an axis**!
- The computer is going to provide a **unit vector**.
- **Want the coordinates to tell us "how far to go in each direction."** This only works if the basis vectors have length 1!

Bases and Coordinates

- Coordinate pairs are represented in a basis. Each **coordinate** tells you how far to move along each basis direction.



$$\mathbf{a} = \begin{pmatrix} 2 \\ 3 \end{pmatrix} = 2\mathbf{e}_1 + 3\mathbf{e}_2$$

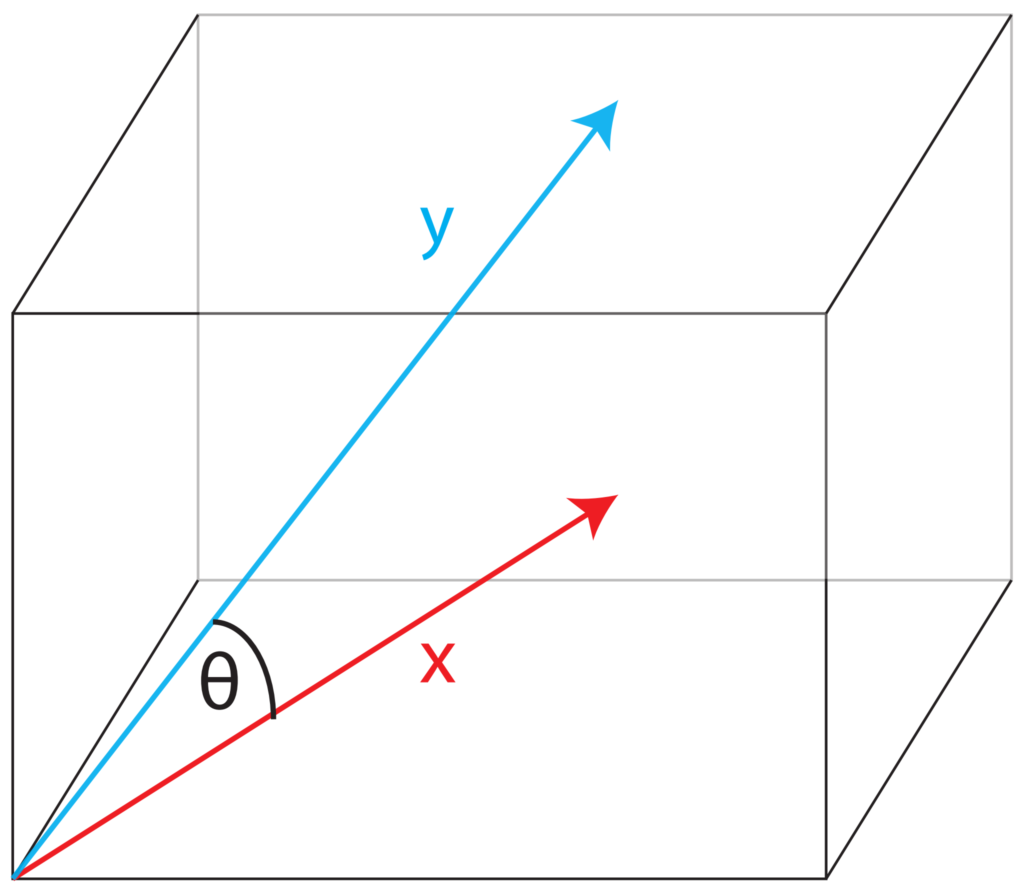
This only works if e_1 and e_2 are unit vectors!!

Determining Orthogonality

(When the angle between vectors is 90 degrees)

Angle between vectors

- Cosine of the angle between two vectors, $\mathbf{x}$ and $\mathbf{y}$, is the inner product of their unit vectors:



$$\cos(\theta) = \frac{\mathbf{x}^T \mathbf{y}}{\|\mathbf{x}\| \ \|\mathbf{y}\|}$$

$$-1 \leq \cos(\theta) \leq 1$$

Vectors are linearly dependent when $|\cos(\theta)| = 1$

Common measure of similarity for high dimensional data like text.

Angle between vectors

$$\cos(\theta) = \frac{\mathbf{x}^T \mathbf{y}}{\|\mathbf{x}\| \ \|\mathbf{y}\|}$$

- Vectors are orthogonal when:

  $$\theta = 90^\circ \ \rightarrow \ \cos(90^\circ) =$$

- Two vectors, x and y, are **orthogonal** when their inner product is zero: i.e. when $\mathbf{x}^{\mathrm{T}}\mathbf{y}=0$.

Practice

1

What's the cosine of the angle between x=(1,-1) and y=(1,0)?

2

Are the vectors $v_1=(1,-1,1)$ and $v_2=(0,1,1)$ orthogonal?

3

What are the two conditions necessary for a collection of vectors to be orthonormal?

Orthonormal Basis

If a set of basis vectors forms an orthonormal basis, it must be that:

1. $\mathbf{v}_i^T\mathbf{v}_j = 0 \quad \text{when } i \neq j$ (i.e. mutually orthogonal)

2. $\mathbf{v}_i^T\mathbf{v}_i = 1 \quad \text{for all } i$ (i.e. each vector is unit vector)

Orthonormal Columns

Suppose the columns of a matrix are orthonormal:

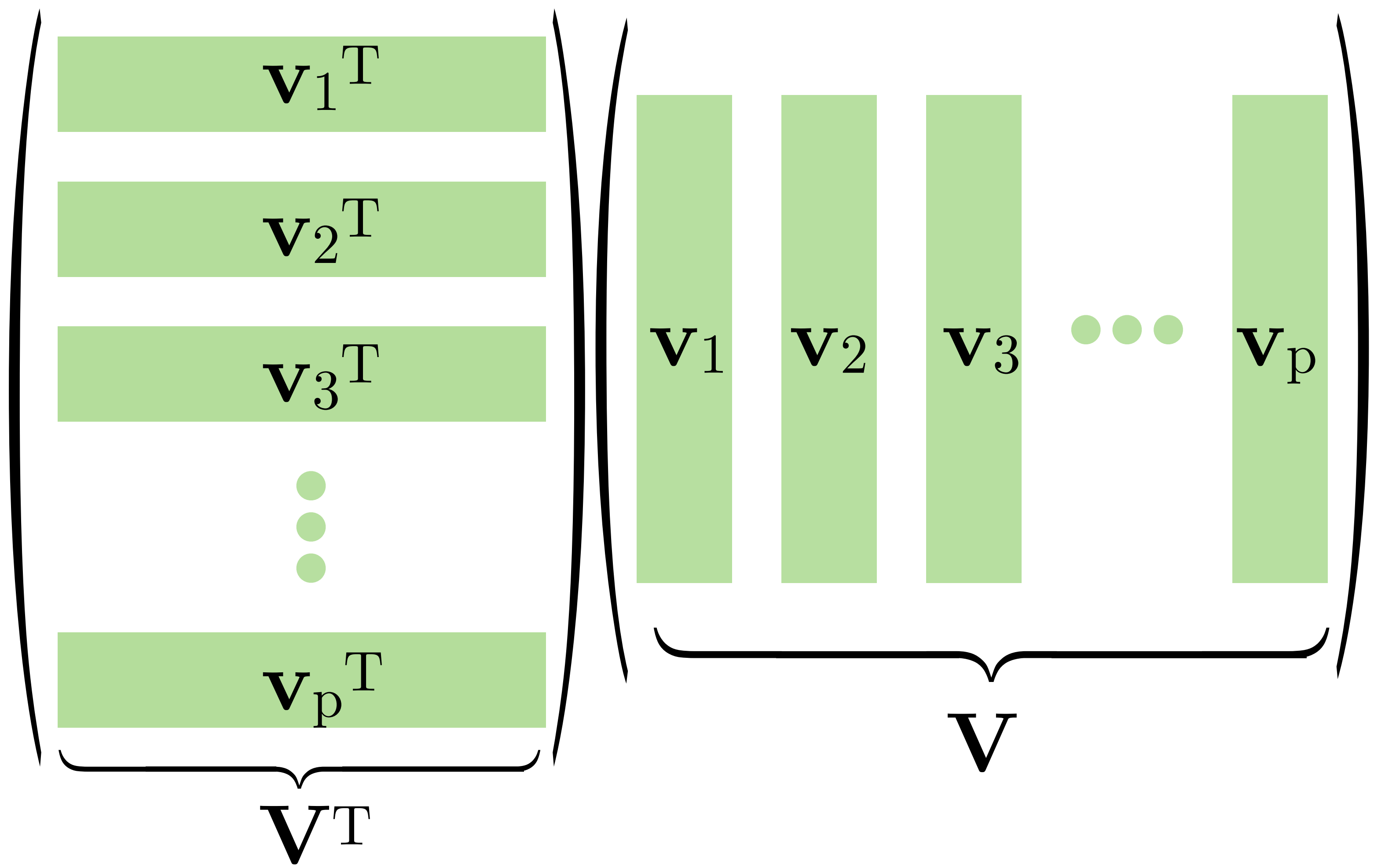
$$\mathbf{V} = \left[\mathbf{v}_1 \mid \mathbf{v}_2 \mid \mathbf{v}_3 \mid \ldots \mid \mathbf{v}_p\right]$$

Consider the matrix product $\mathbf{V}^T\mathbf{V}$

Orthonormal Columns

Suppose the columns of a matrix are orthonormal:

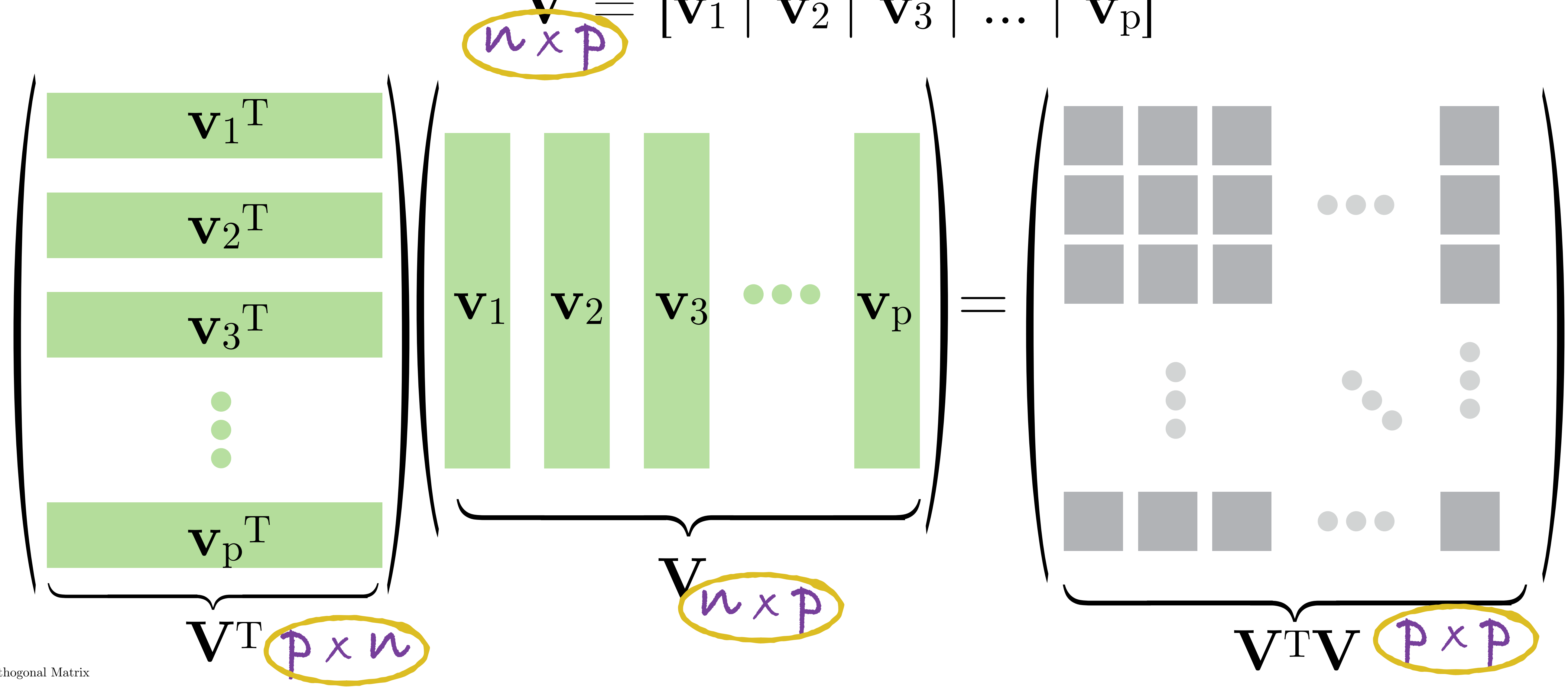
$$\mathbf{V} = [\mathbf{v}_1 \mid \mathbf{v}_2 \mid \mathbf{v}_3 \mid \dots \mid \mathbf{v}_p]$$



Orthonormal Columns

Suppose the columns of a matrix are orthonormal:

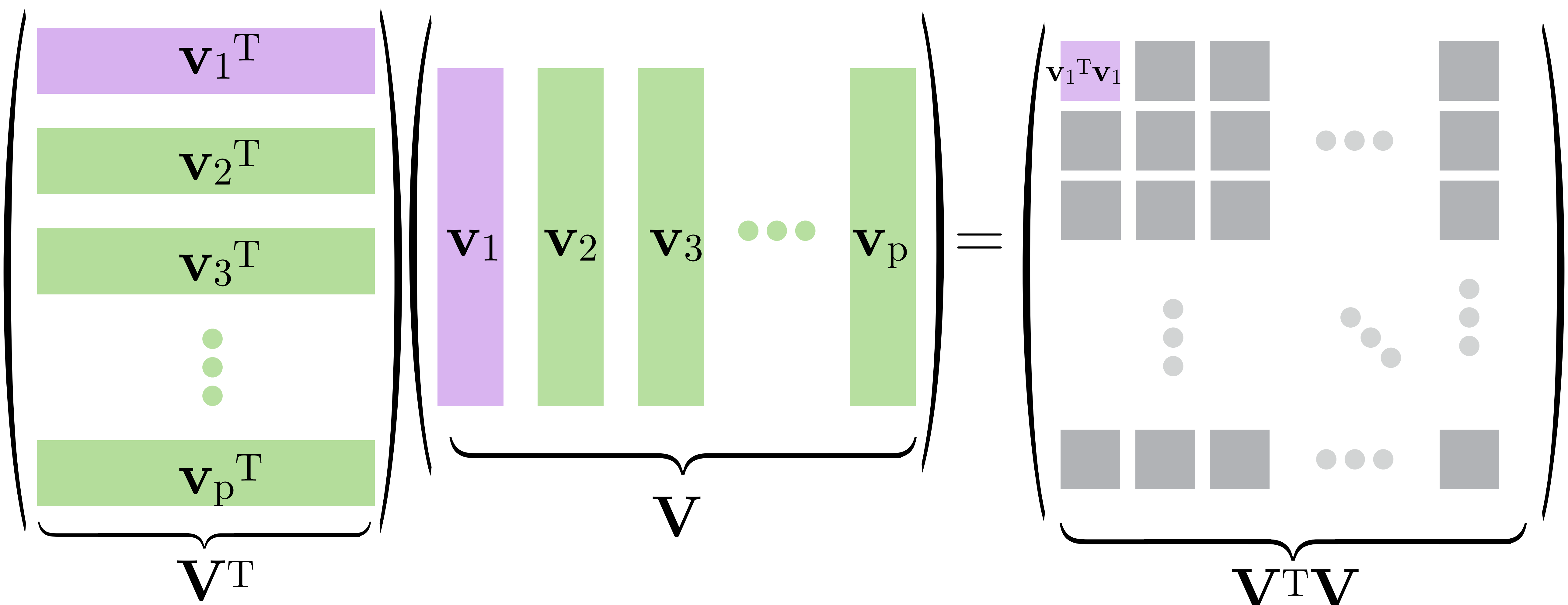
$$\mathbf{V} = [\mathbf{v}_1 \mid \mathbf{v}_2 \mid \mathbf{v}_3 \mid \dots \mid \mathbf{v}_\mathrm{p}]$$



Orthonormal Columns

Suppose the columns of a matrix are orthonormal:

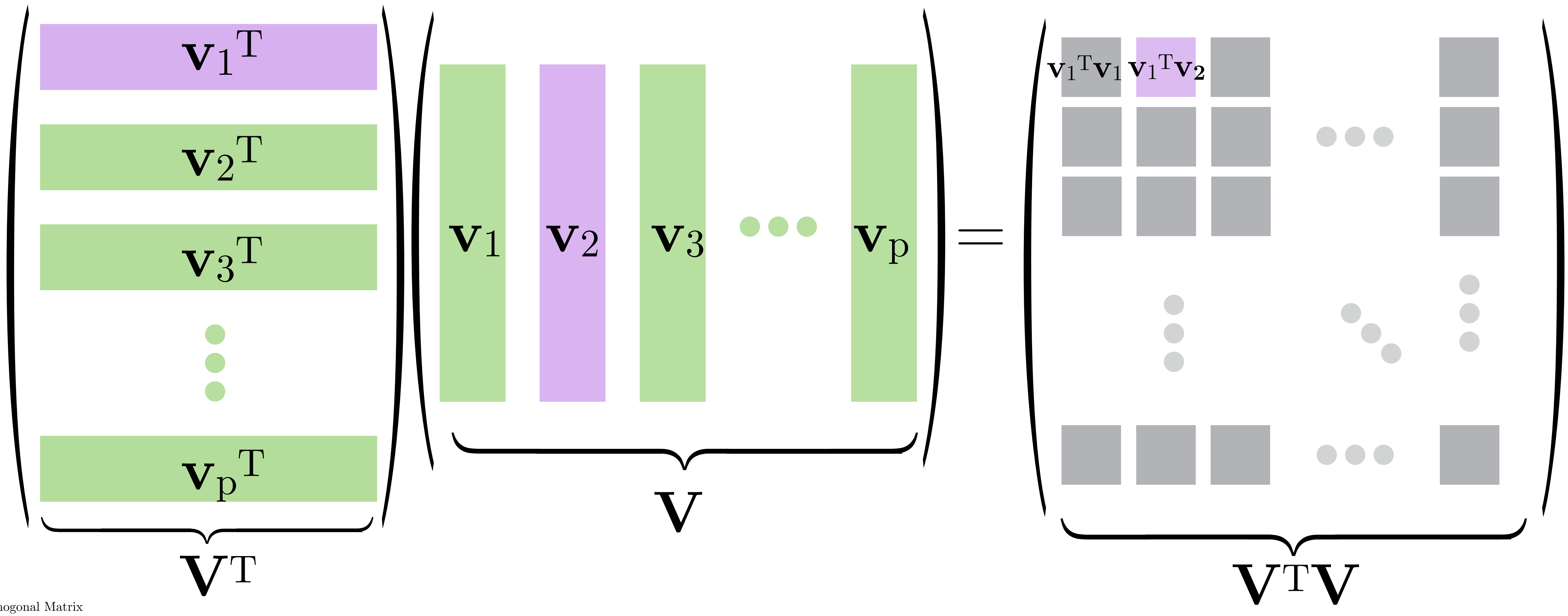
$$\mathbf{V} = [\mathbf{v}_1 \mid \mathbf{v}_2 \mid \mathbf{v}_3 \mid \ldots \mid \mathbf{v}_p]$$



Orthonormal Columns

Suppose the columns of a matrix are orthonormal:

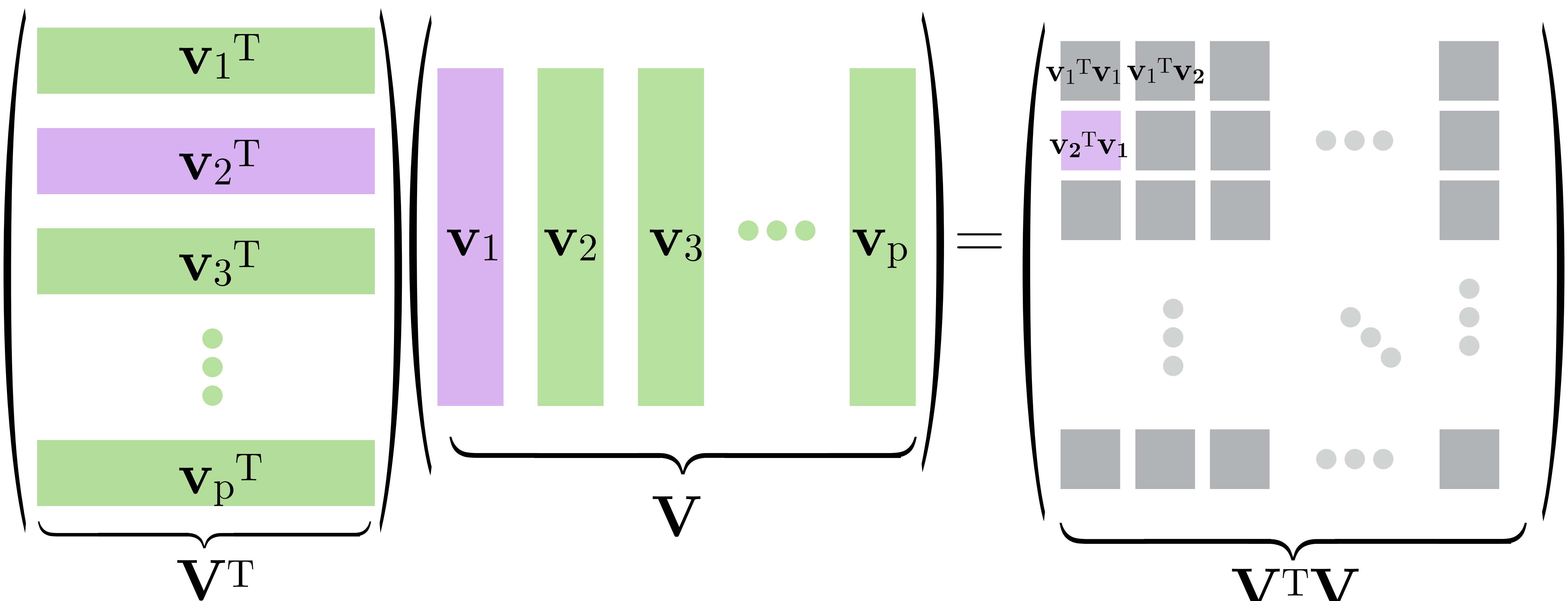
$$\mathbf{V} = [\mathbf{v}_1 \mid \mathbf{v}_2 \mid \mathbf{v}_3 \mid \ldots \mid \mathbf{v}_p]$$



Orthonormal Columns

Suppose the columns of a matrix are orthonormal:

$$\mathbf{V} = [\mathbf{v}_1 \mid \mathbf{v}_2 \mid \mathbf{v}_3 \mid \ldots \mid \mathbf{v}_p]$$

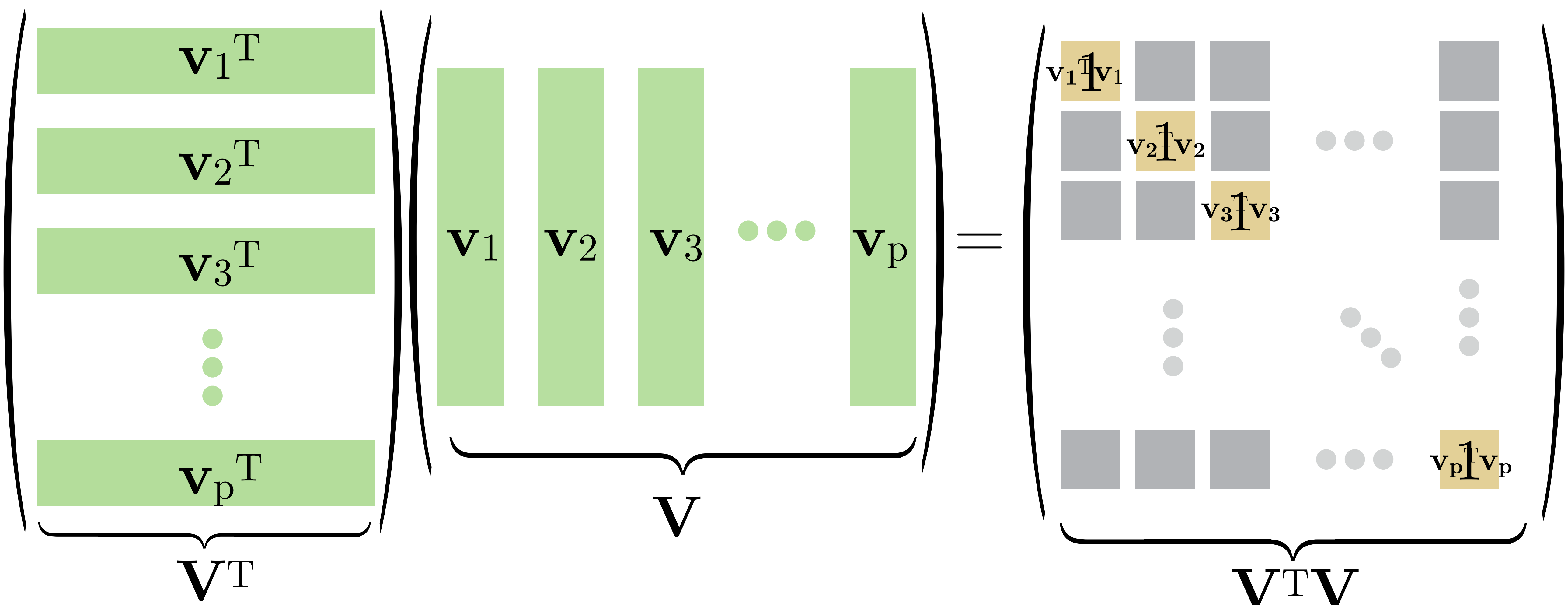


Orthonormal Columns

Suppose the columns of a matrix are orthonormal:

$$\mathbf{v}_i^T \mathbf{v}_i = 1$$

$$\mathbf{V} = [\mathbf{v}_1 \mid \mathbf{v}_2 \mid \mathbf{v}_3 \mid \dots \mid \mathbf{v}_p]$$



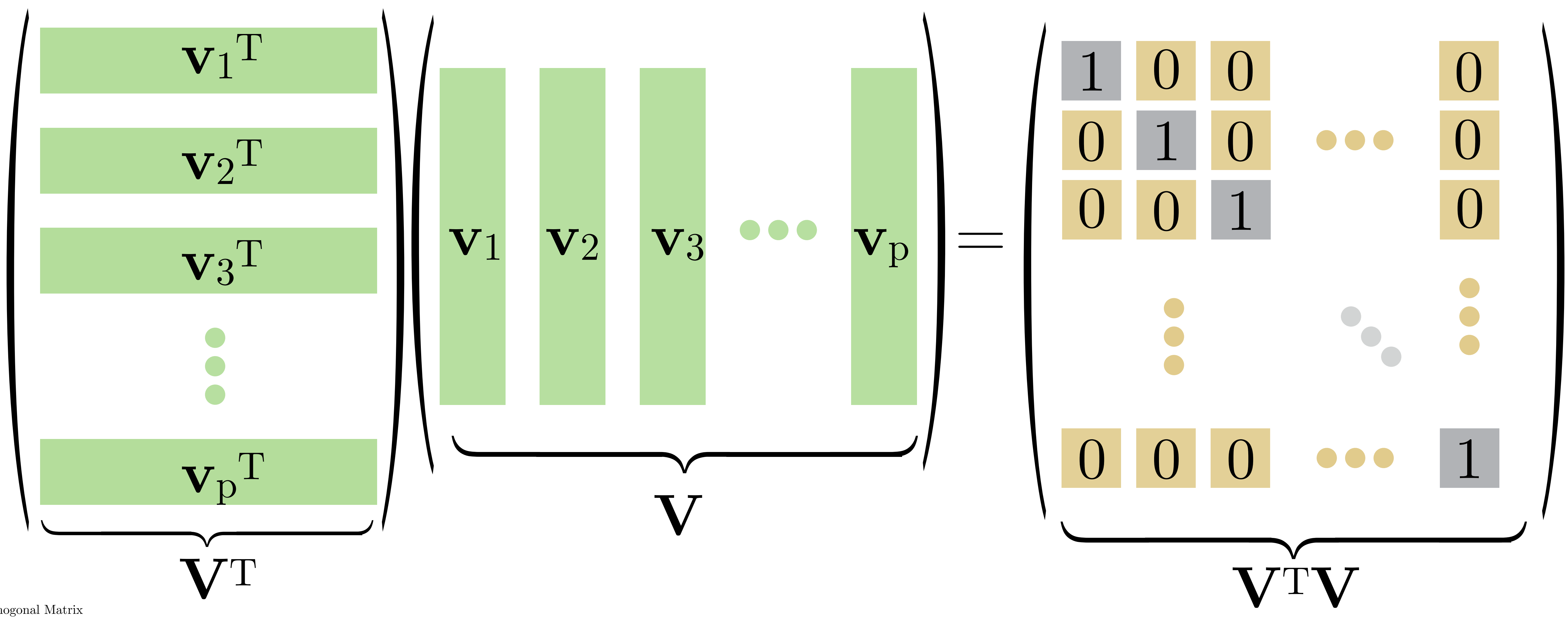
Orthonormal Columns

Suppose the columns of a matrix are orthonormal:

$$\mathbf{v}_i^T \mathbf{v}_j = 0$$

$$i \neq j$$

$$\mathbf{V} = [\mathbf{v}_1 \mid \mathbf{v}_2 \mid \mathbf{v}_3 \mid \dots \mid \mathbf{v}_p]$$



Orthonormal Columns

When a matrix, V, has orthonormal columns

$$\mathbf{V}^{\mathrm{T}}\mathbf{V} = \mathbf{I}$$

However, we can't say anything about $\mathbf{V}\mathbf{V}^{\mathrm{T}}$ unless the matrix is square.

Orthogonal Matrix

When a square matrix has orthonormal columns, it also has orthonormal rows. Such a matrix is called an **orthogonal matrix** and its inverse is equal to its transpose:

$$\mathbf{V}^{\mathrm{T}}\mathbf{V} = \mathbf{V}\mathbf{V}^{\mathrm{T}} = \mathbf{I}$$

$$\mathbf{V}^{-1} = \mathbf{V}^{\mathrm{T}}$$

Orthogonal Matrix

- An orthogonal matrix is easy to maneuver inside matrix equations, since $\mathbf{V}^{-1} = \mathbf{V}^{T}$

- For example if U and V are orthogonal, the following equations are equivalent:

$$\mathbf{XV} = \mathbf{UD}$$

$$\mathbf{X} = \mathbf{UDV}^{T}$$

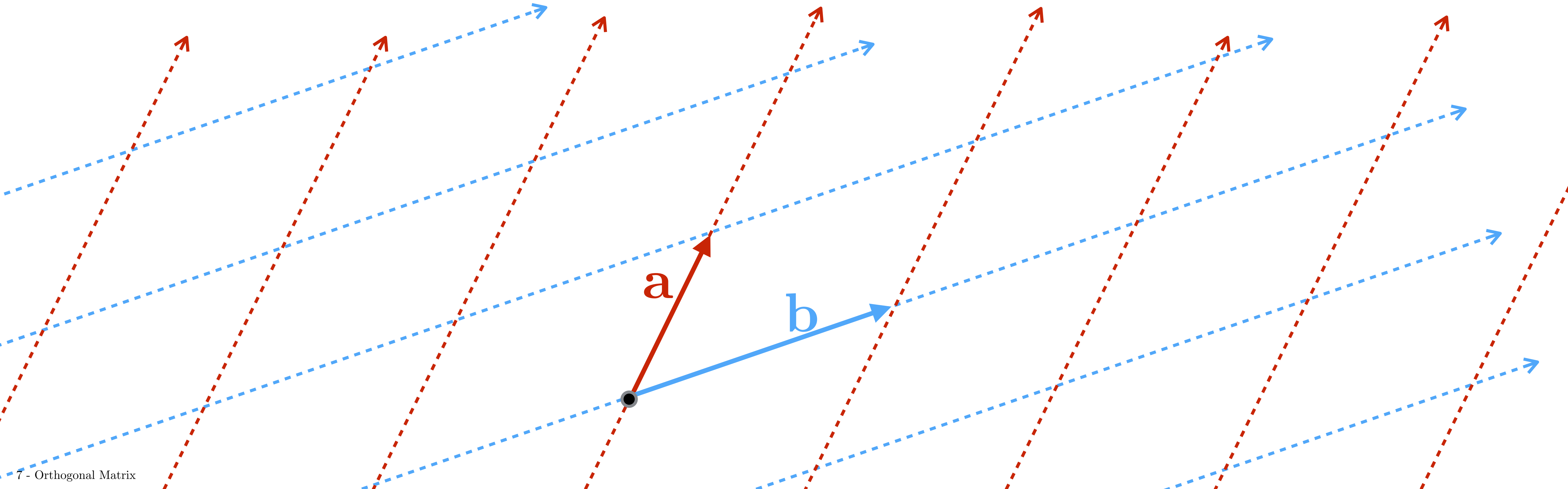
$$\mathbf{U}^{T}\mathbf{X} = \mathbf{DV}^{T}$$

$$\mathbf{U}^{T}\mathbf{XV} = \mathbf{D}$$

Why an Orthonormal Basis?

- Two Conditions → Two Reasons

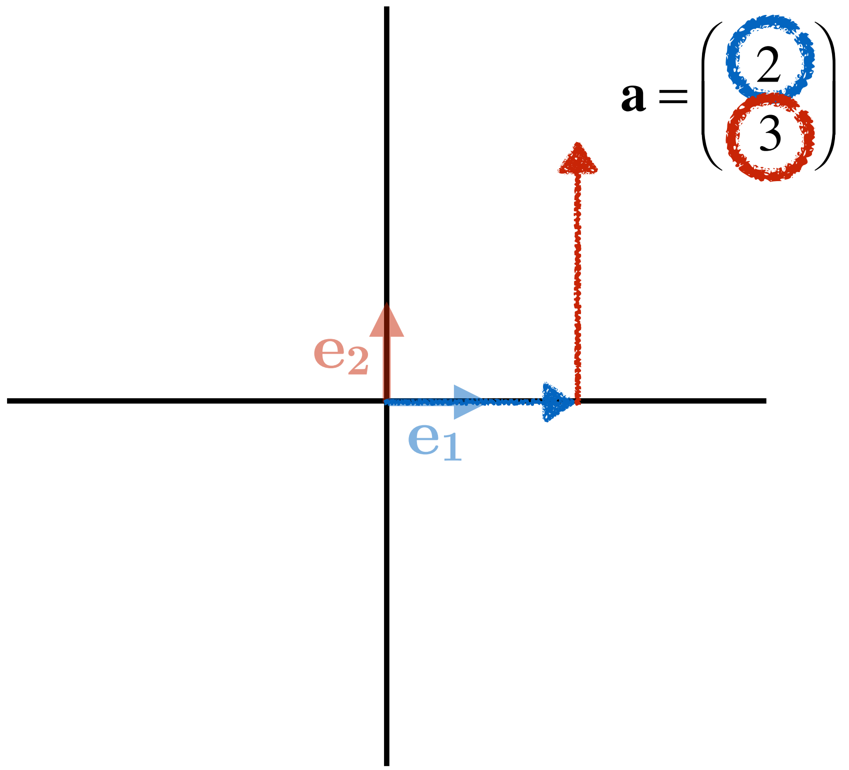
1. Basis vectors mutually perpendicular. They are just rotations of the elementary basis vectors. Can still plot/consider data coordinates in a familiar way. Anything else would just be weird (Non-Euclidean/Affine)!



Why an Orthonormal Basis?

- Two Conditions → Two Reasons

  2. The basis vectors have length 1. Want the coordinates to tell us how many *units* to go in each basis direction. In this way, we can focus on the coordinates alone and almost ignore the existence of basis vectors!



To investigate,
you might consider
writing the point **a**
in the orthogonal *but*
not orthonormal basis
$v_1=(2,0)$ $v_2=(0,1)$

Summary: Orthonormal Bases

- A basis that is NOT orthonormal will distort the data.
- A basis that IS orthonormal will merely rotate the data
- Most dimension reduction methods create a new orthonormal basis for the data.
  - ★ Principal Components Analysis
  - ★ Singular Value Decomposition
  - ★ Factor Analysis
  - ★ Correspondence Analysis

Practice

Let $\mathbf{U} = \frac{1}{3}\begin{pmatrix} -1 & 2 & 0 & -2 \\ 2 & 2 & 0 & 1 \\ 0 & 0 & 3 & 0 \\ -2 & 1 & 0 & 2 \end{pmatrix}$

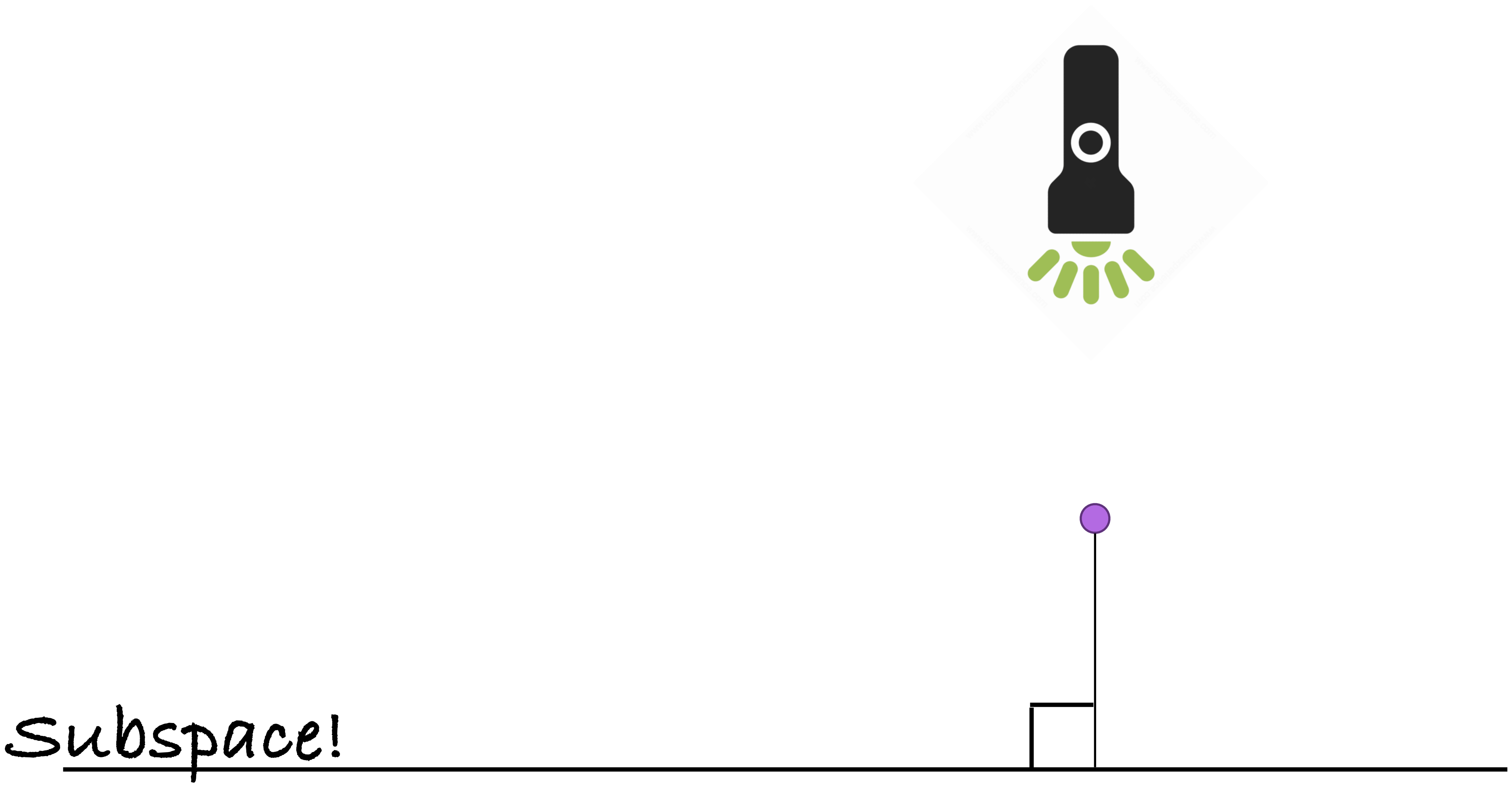
1 Show that $\mathbf{U}$ is an orthogonal matrix

2 Let $\mathbf{b}$=(1,1,1,1). Solve the equation $\mathbf{Ux}=\mathbf{b}$

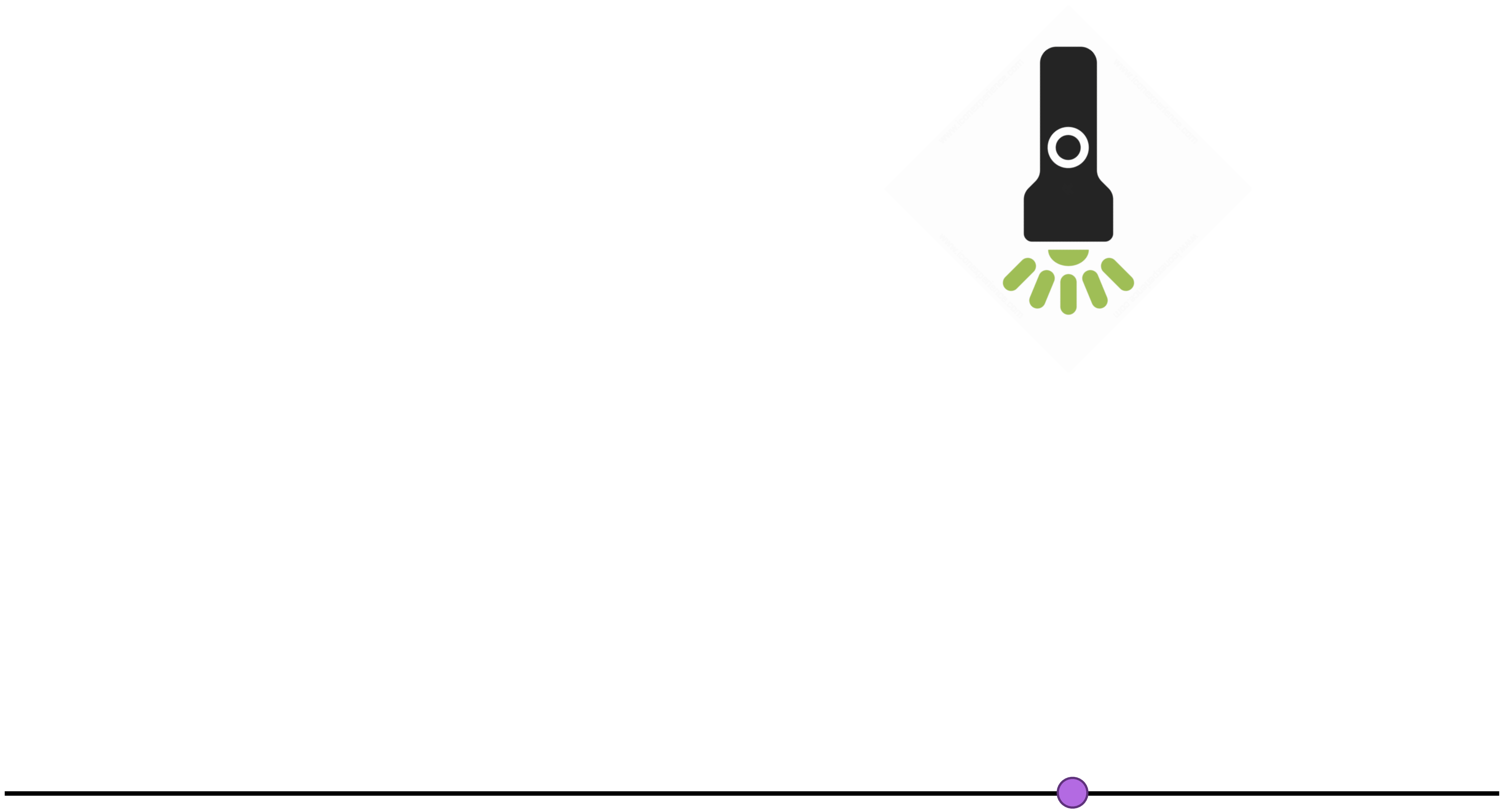
3 Find two vectors which are orthogonal to $\mathbf{x}$=(1,1,1)

Orthogonal Projections

Orthogonal Projection

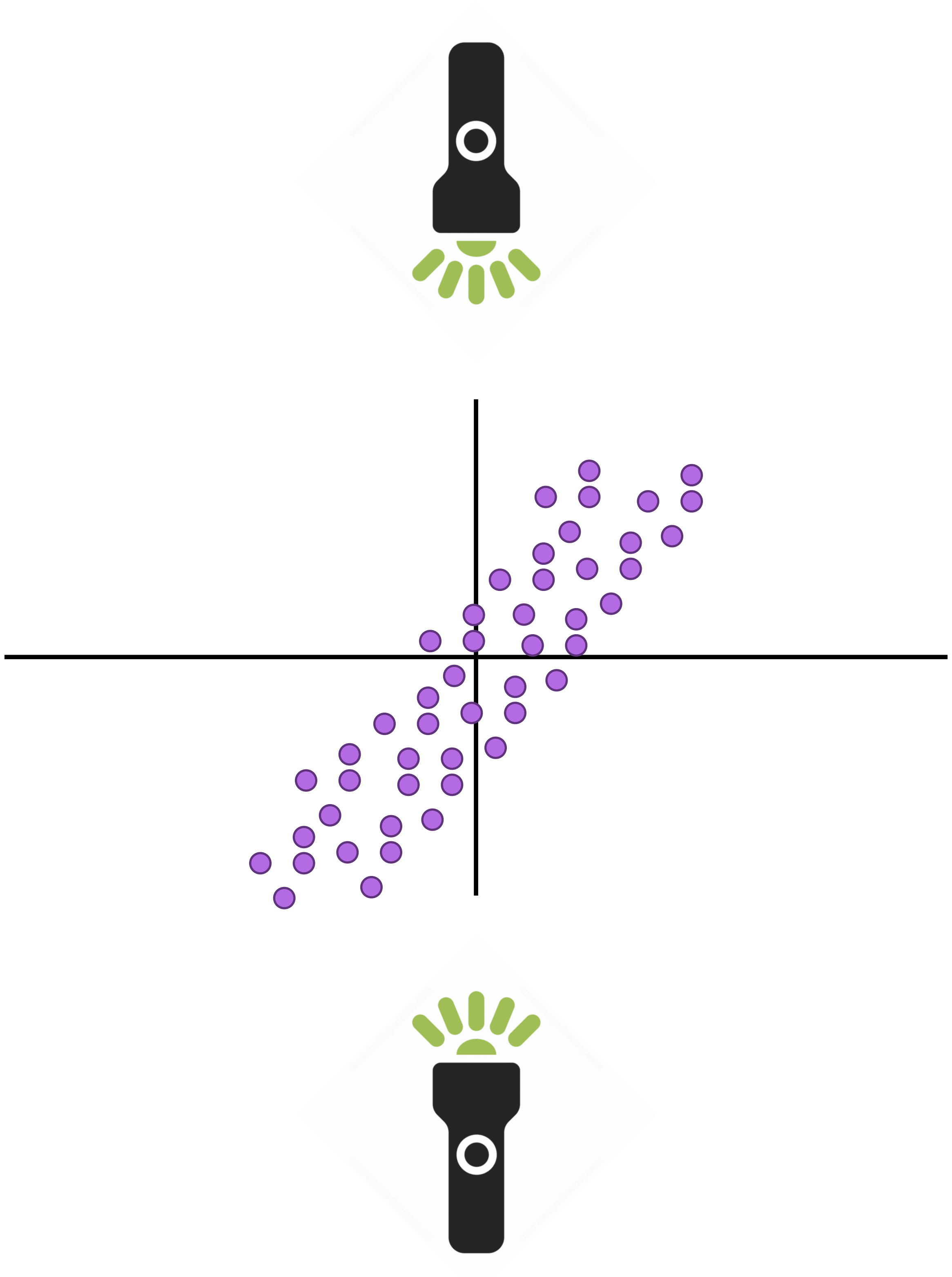


Orthogonal Projection

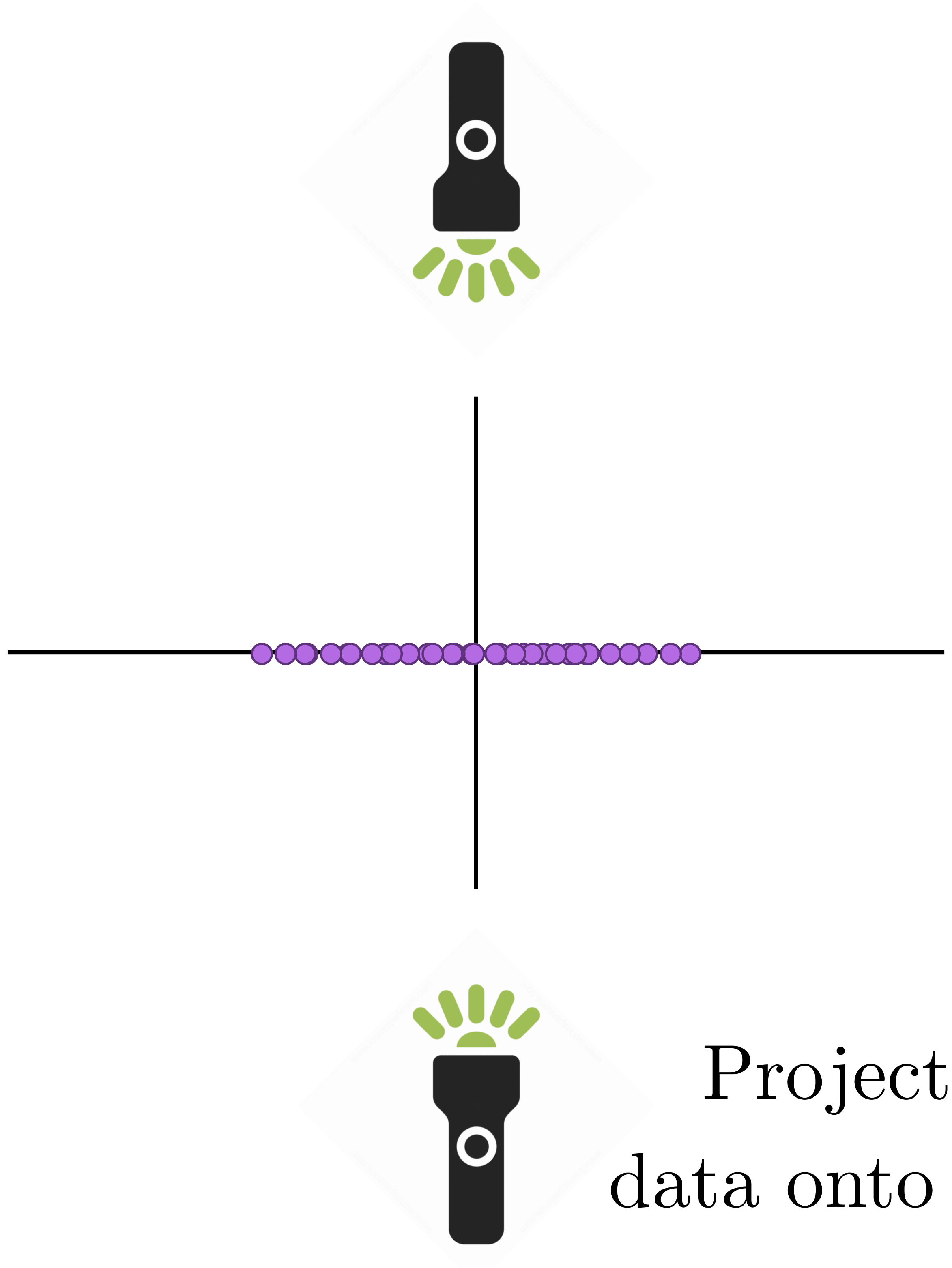


The projection of the
point onto the subspace.

Orthogonal Projection

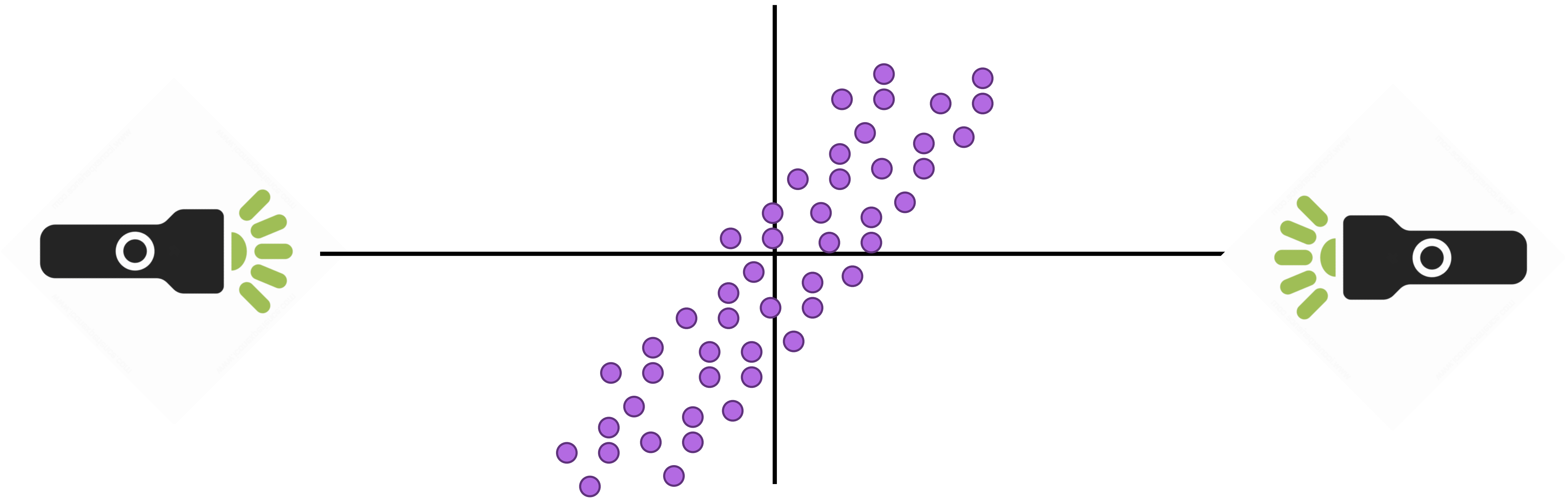


Orthogonal Projection

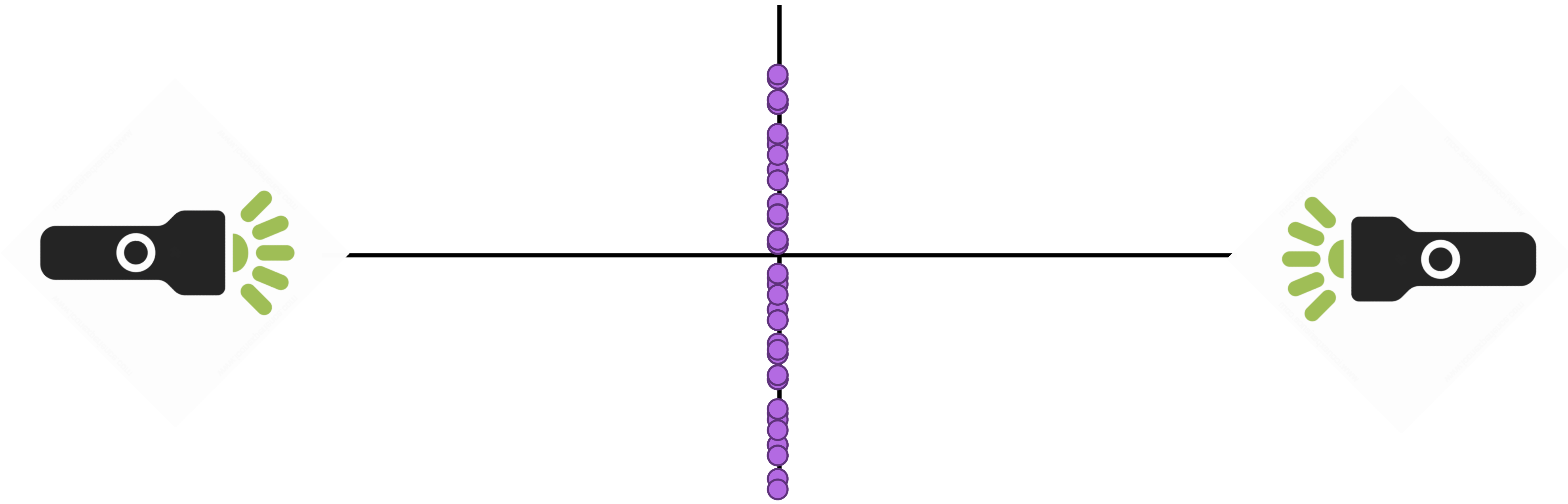


Projection of the data onto the span(e_1)

Orthogonal Projection



Orthogonal Projection



Projection of the
data onto the span(e_2)

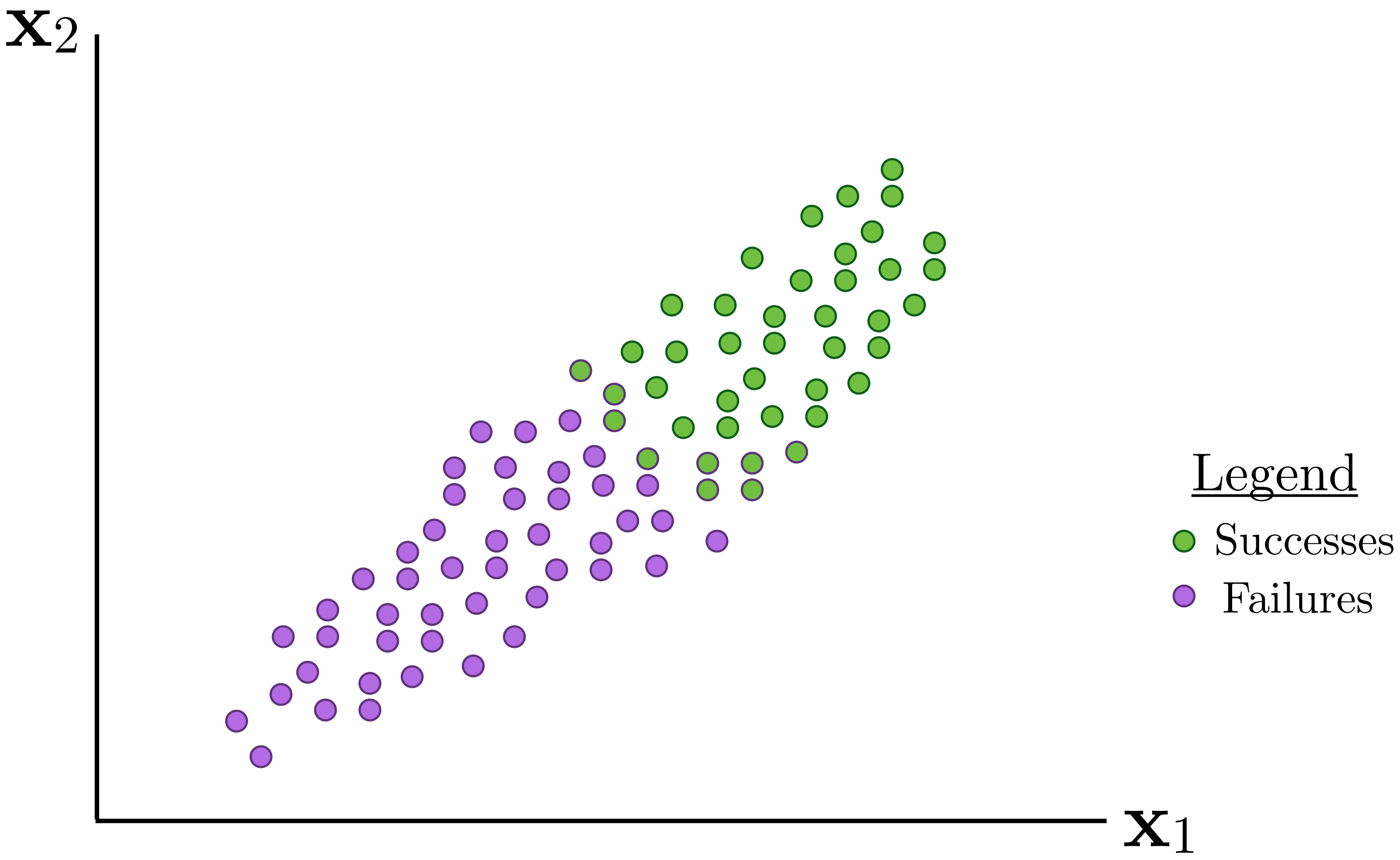
Orthogonal Projection

When we "drop" a variable, this is essentially what is happening! We've projected the data onto the span of the other basis vectors.

A Glimpse of Application

Combining a change of basis with an orthogonal projection

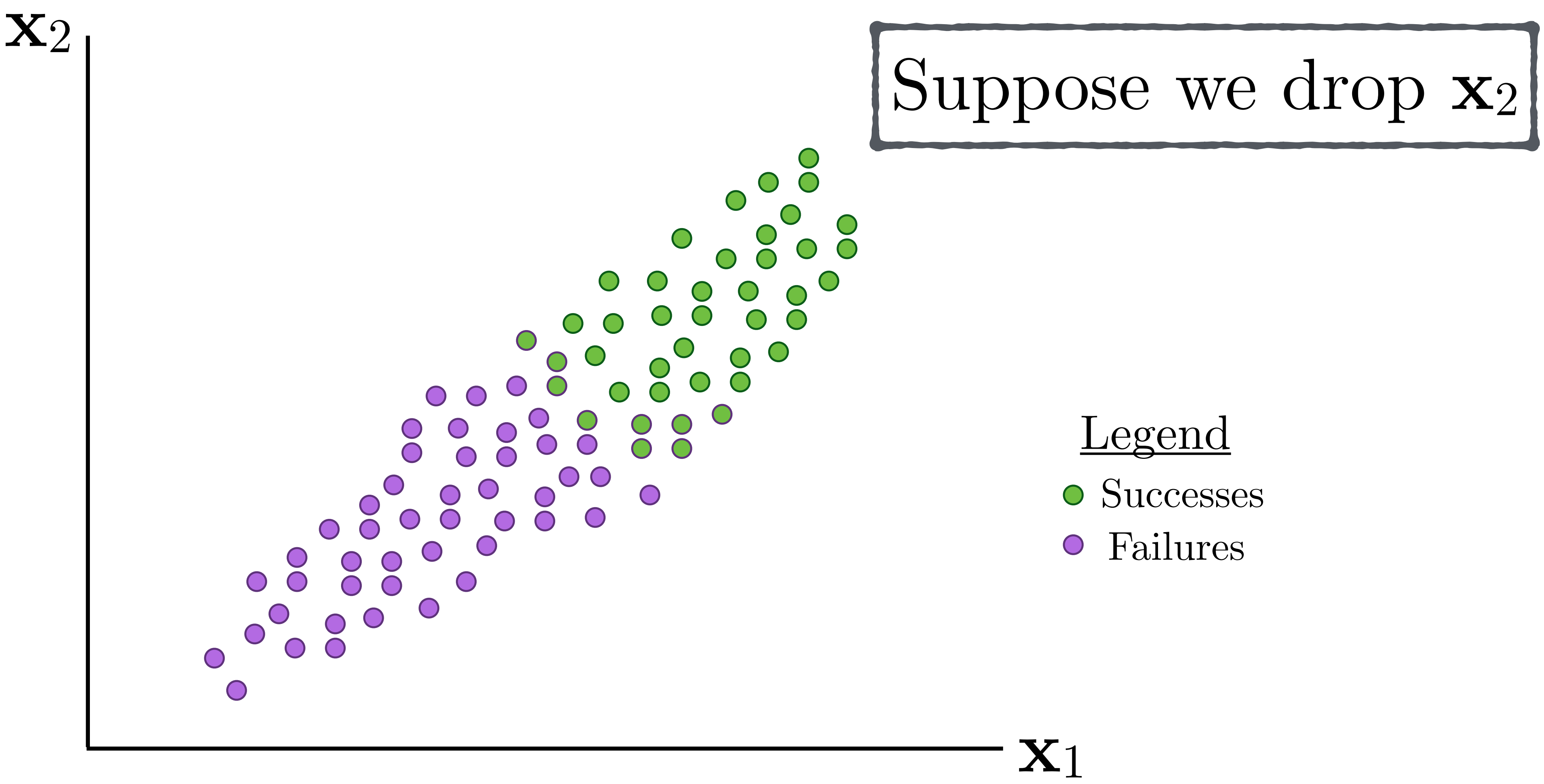
Orthogonal Projection



Suppose 2 variables is just too many.

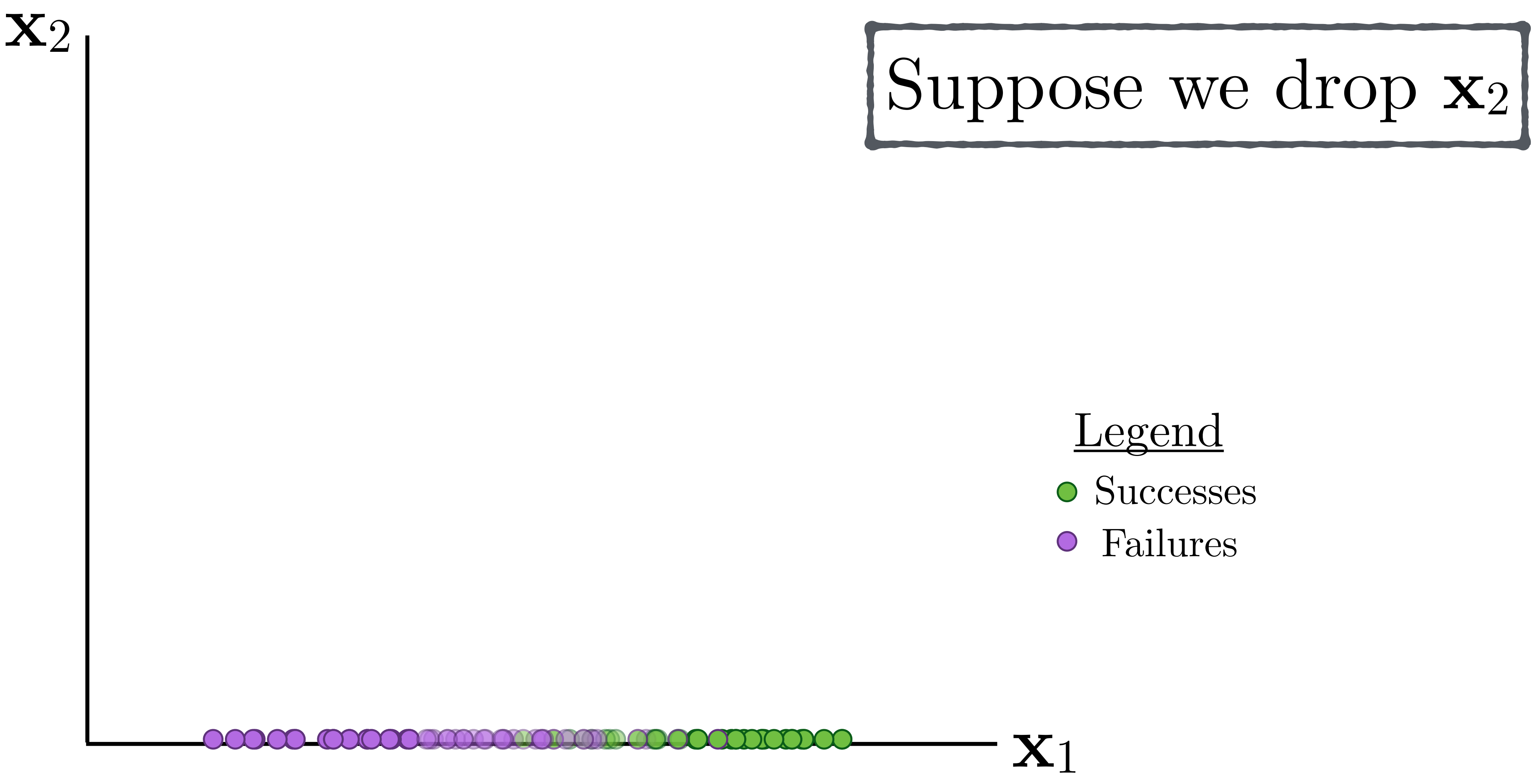
Need to reduce the dimensions.

Orthogonal Projection

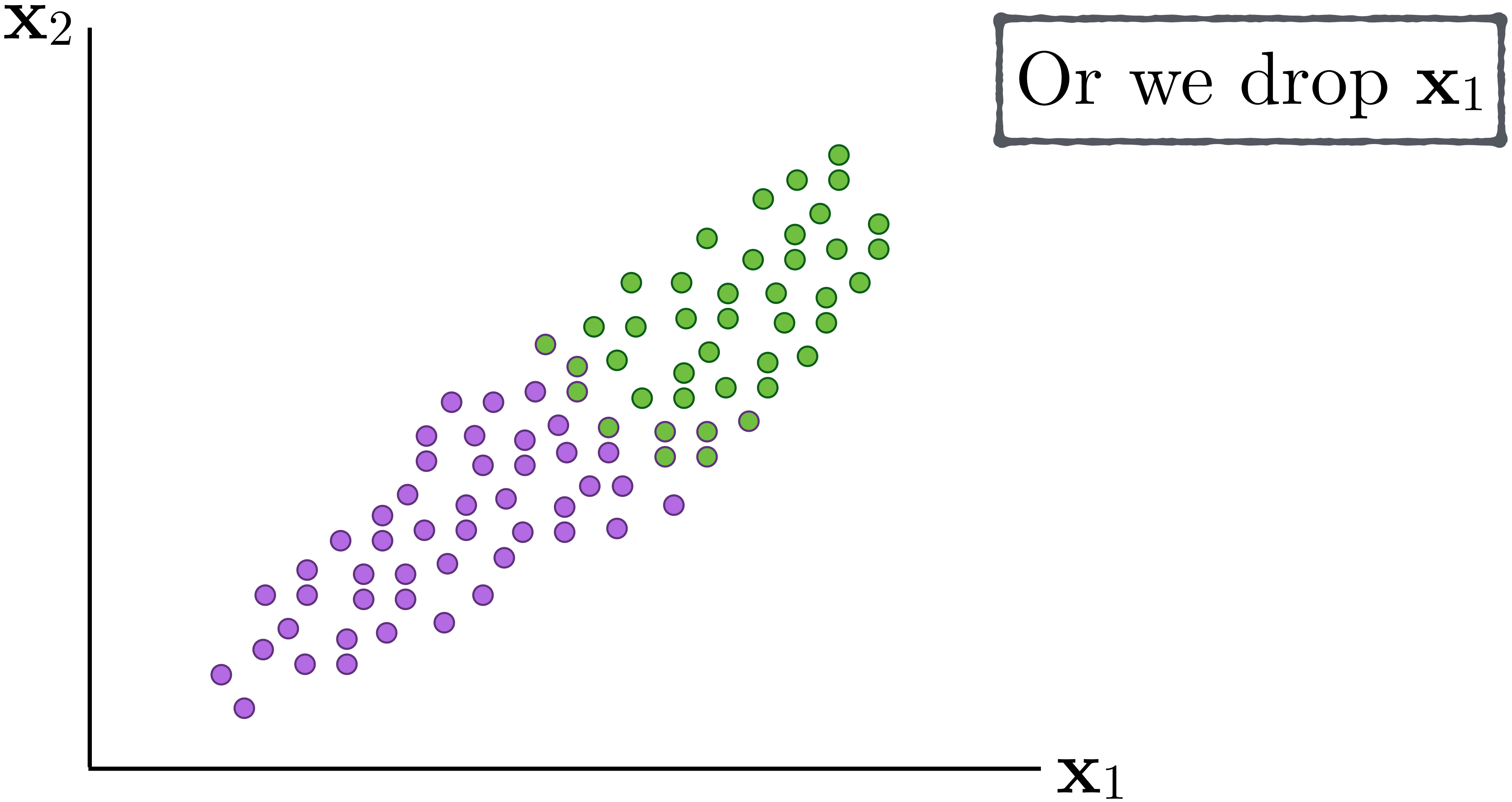


One option is to simply drop one of the variables.

Orthogonal Projection



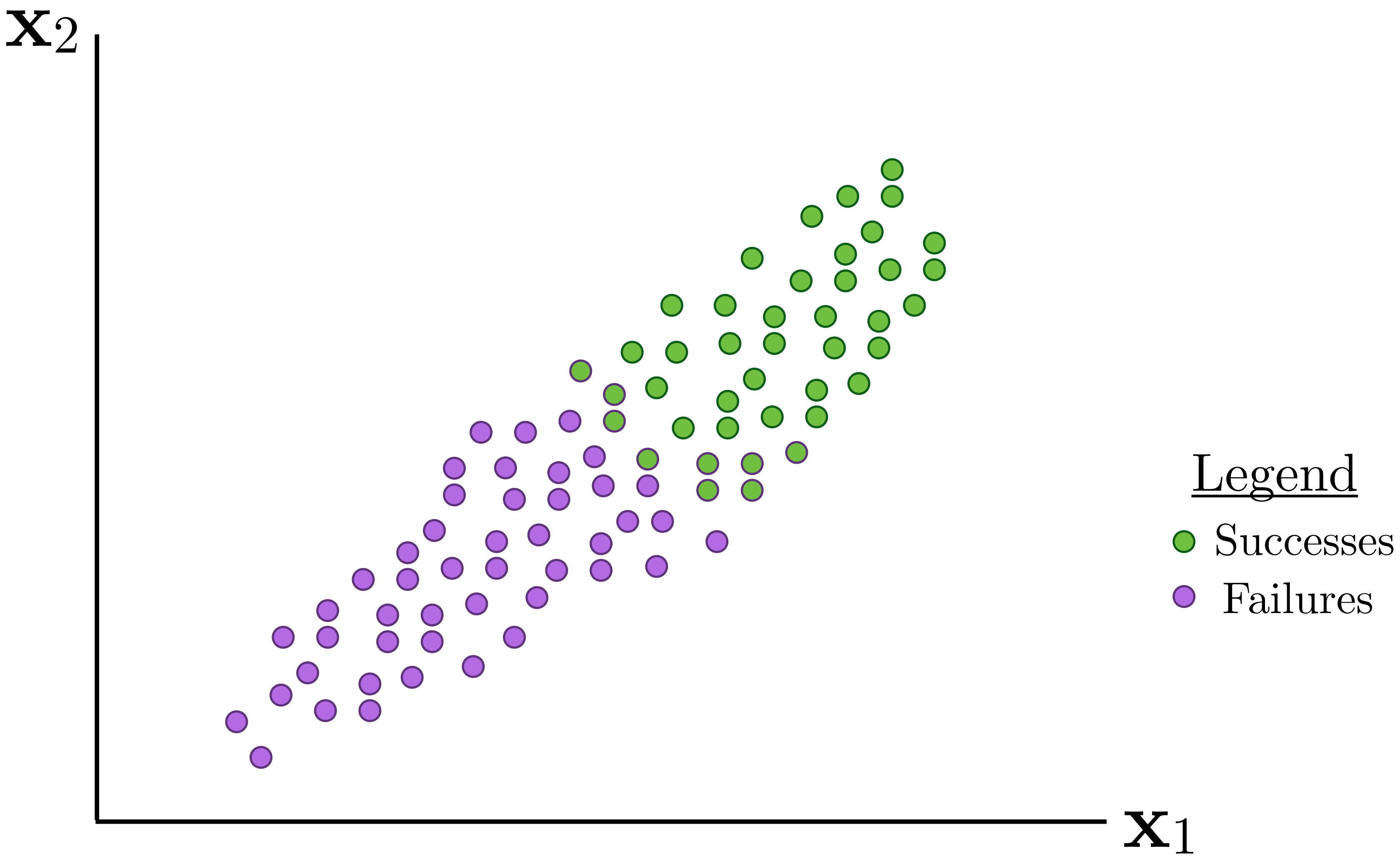
Orthogonal Projection



Orthogonal Projection

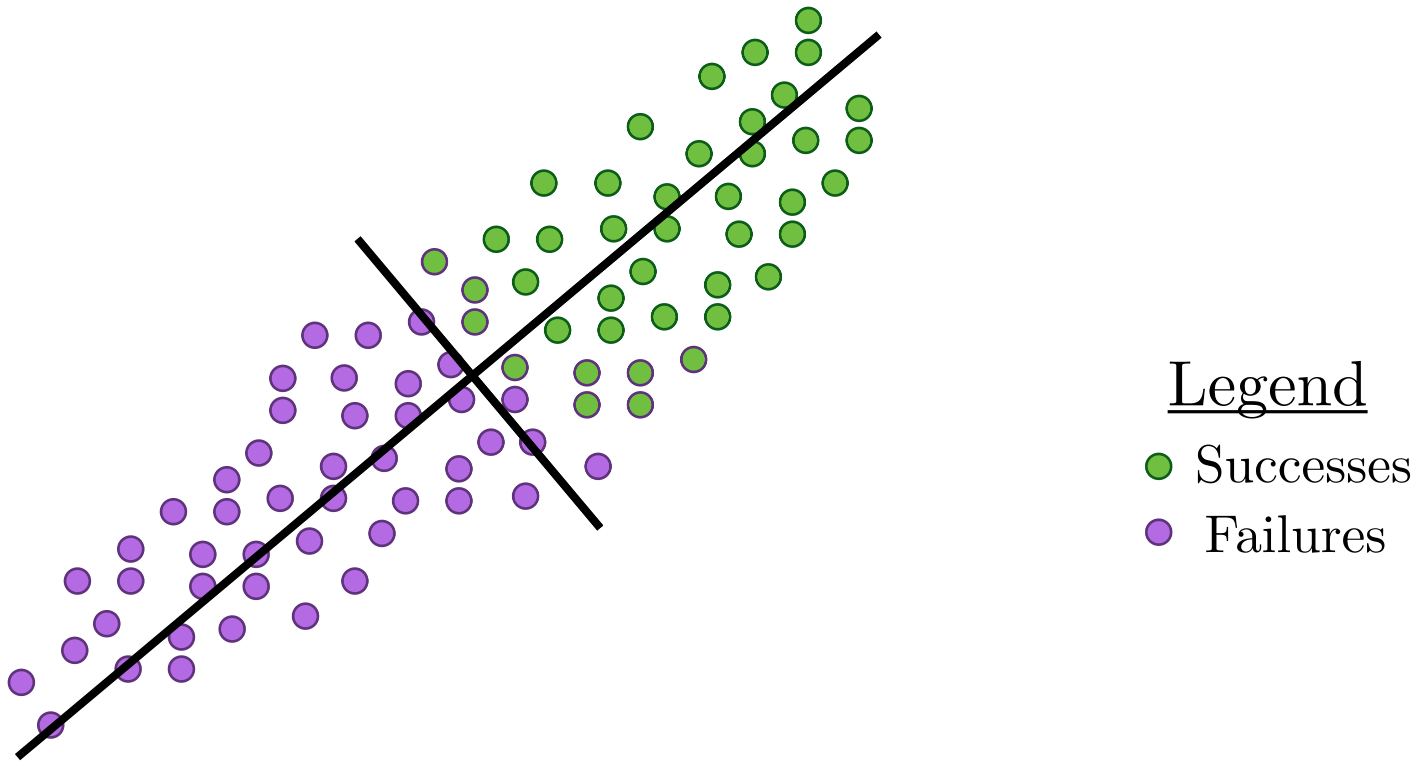
Or we drop $\mathbf{x}_1$

Orthogonal Projection



What if we took a different approach and changed the basis?

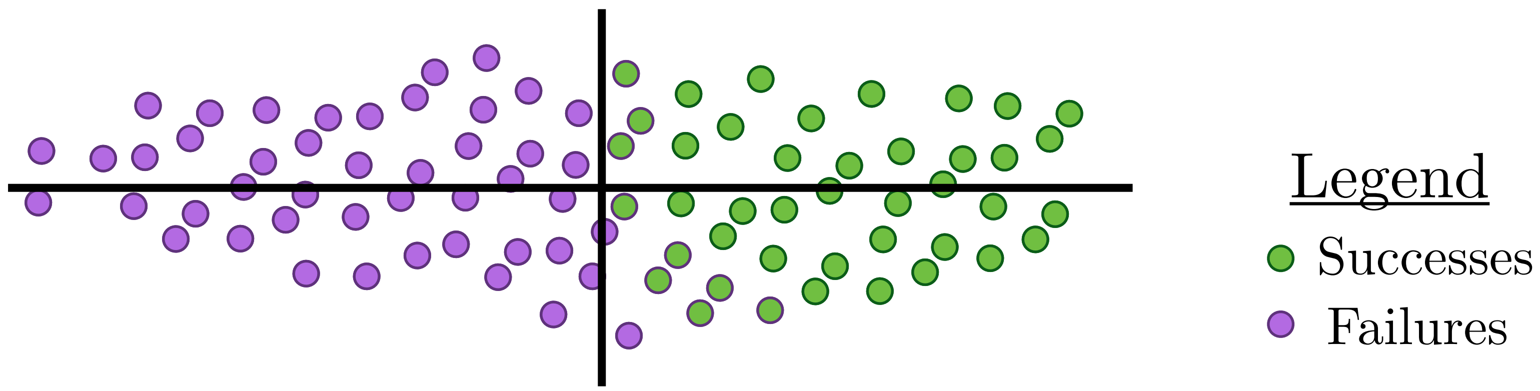
Orthogonal Projection



What if we took a different approach and changed the basis?

Orthogonal Projection

Suppose we drop $\mathbf{v}_2$

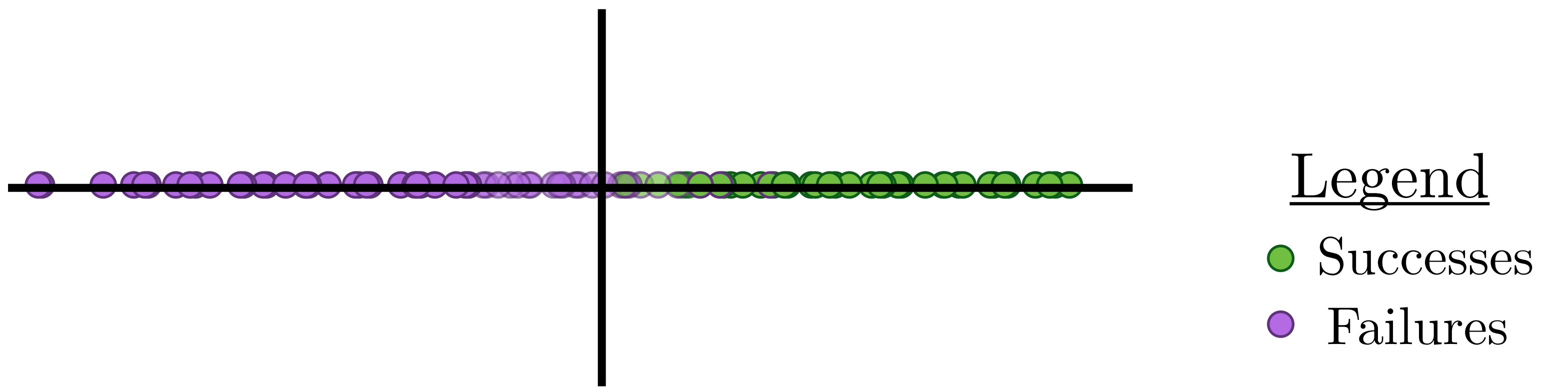


Now that we have these new variables, $\mathbf{v}_1$ and $\mathbf{v}_2$, what happens when we drop one?

Orthogonal Projection

Suppose we drop $\mathbf{v}_2$

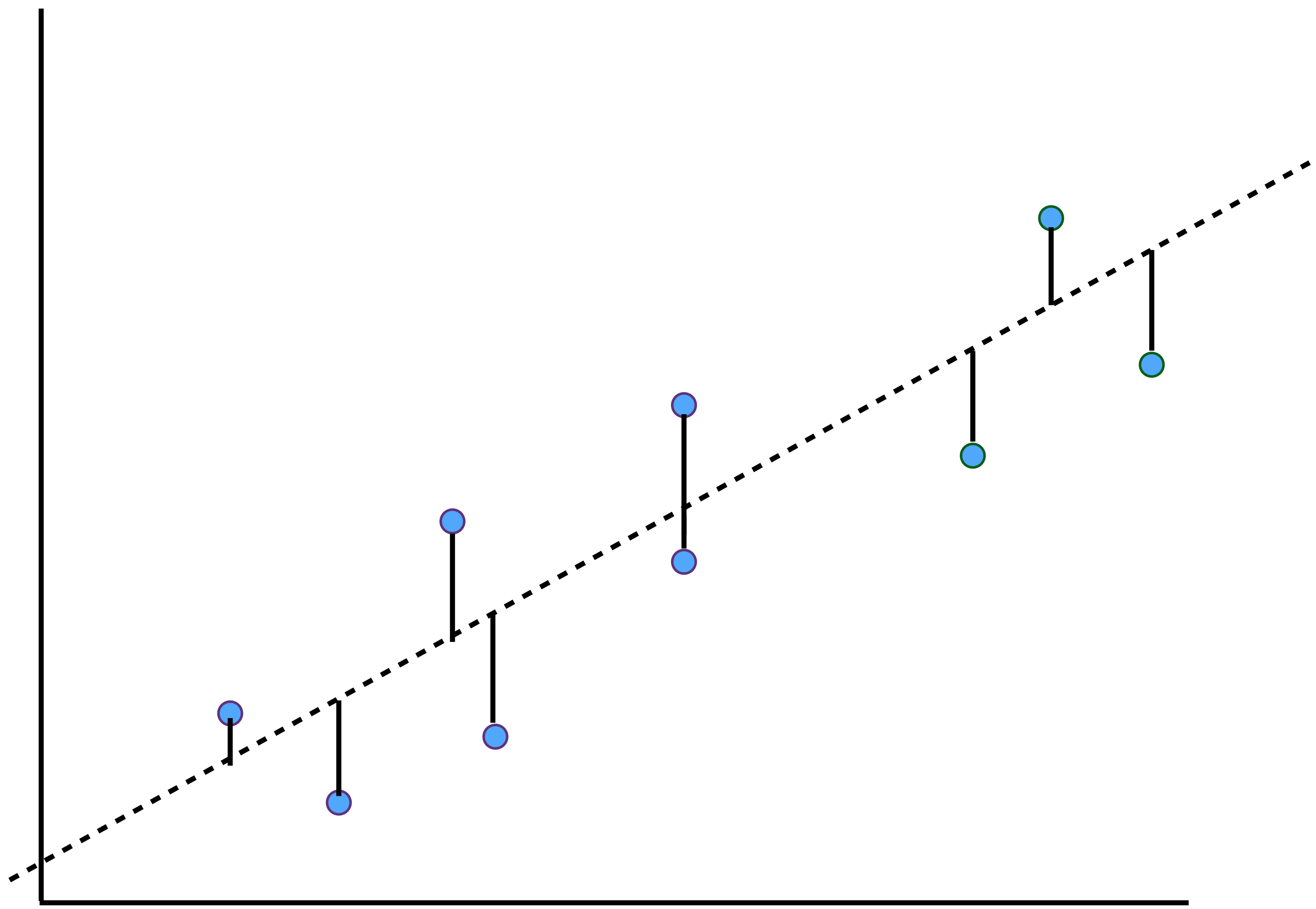
data nearly perfectly separable in 1 dimension!



Summary: Orthogonal Projections

- Most dimension reduction methods do what we just saw:
  - Draw new axes, create a new set of coordinates for the data along the associated basis vectors
  - Project the data orthogonally onto the preferred axes.
    - Preference is given to the preservation of patterns and information (i.e. variance).

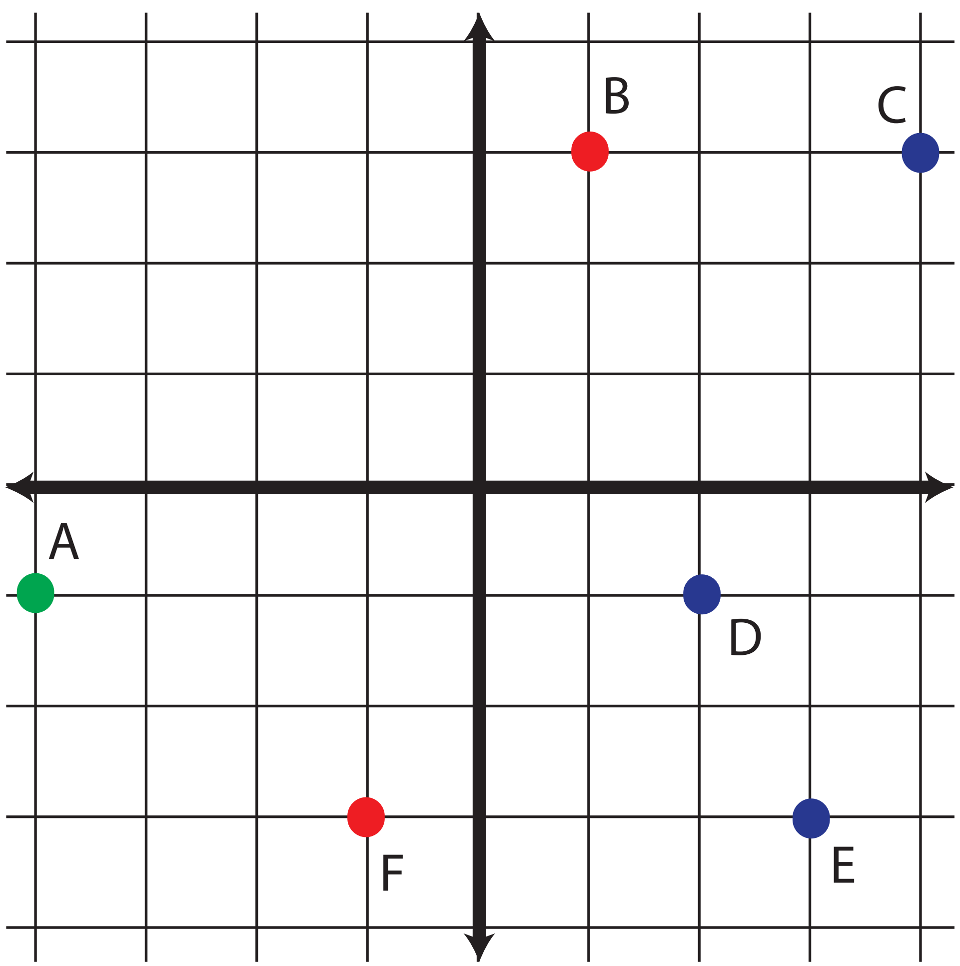
Are predicted values in regression orthogonal projections?



No!

Practice

Draw the orthogonal projections of the 6 points labeled A-F onto the following subspaces:



1. The span($\mathbf{e}_1$)

2. The span($\mathbf{e}_2$)

3. The span((-1,-1))

Linear Regression as a projection

Just for fun.

Linear Regression as a System of Equations

$$\beta_0 + \beta_1 \mathbf{x}_1 + \cdots + \beta_p \mathbf{x}_p = \mathbf{y}$$

$$\begin{array}{c} \\ obs_1 \\ obs_2 \\ \vdots \\ obs_n \end{array}
\underbrace{\begin{array}{c} \begin{array}{ccccc} & \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_p \end{array} \\ \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} \end{array}}_{\mathbf{X}}
\underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}}_{\boldsymbol{\beta}}
=
\underbrace{\begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{pmatrix}}_{\mathbf{y}}$$

$$\mathbf{X}\boldsymbol{\beta} = \mathbf{y}$$

Question: Is the vector $\mathbf{y}$ in the ***span*** of the columns of $\mathbf{X}$?

$$\beta_0 + \beta_1 \mathbf{x}_1 + \cdots + \beta_p \mathbf{x}_p = \mathbf{y}$$

Translation: Is there an *exact* solution for $\boldsymbol{\beta}$?

Answer: No. (That's why we use "least squares")

$$\boxed{\mathbf{X}\boldsymbol{\beta} = \mathbf{y}}$$

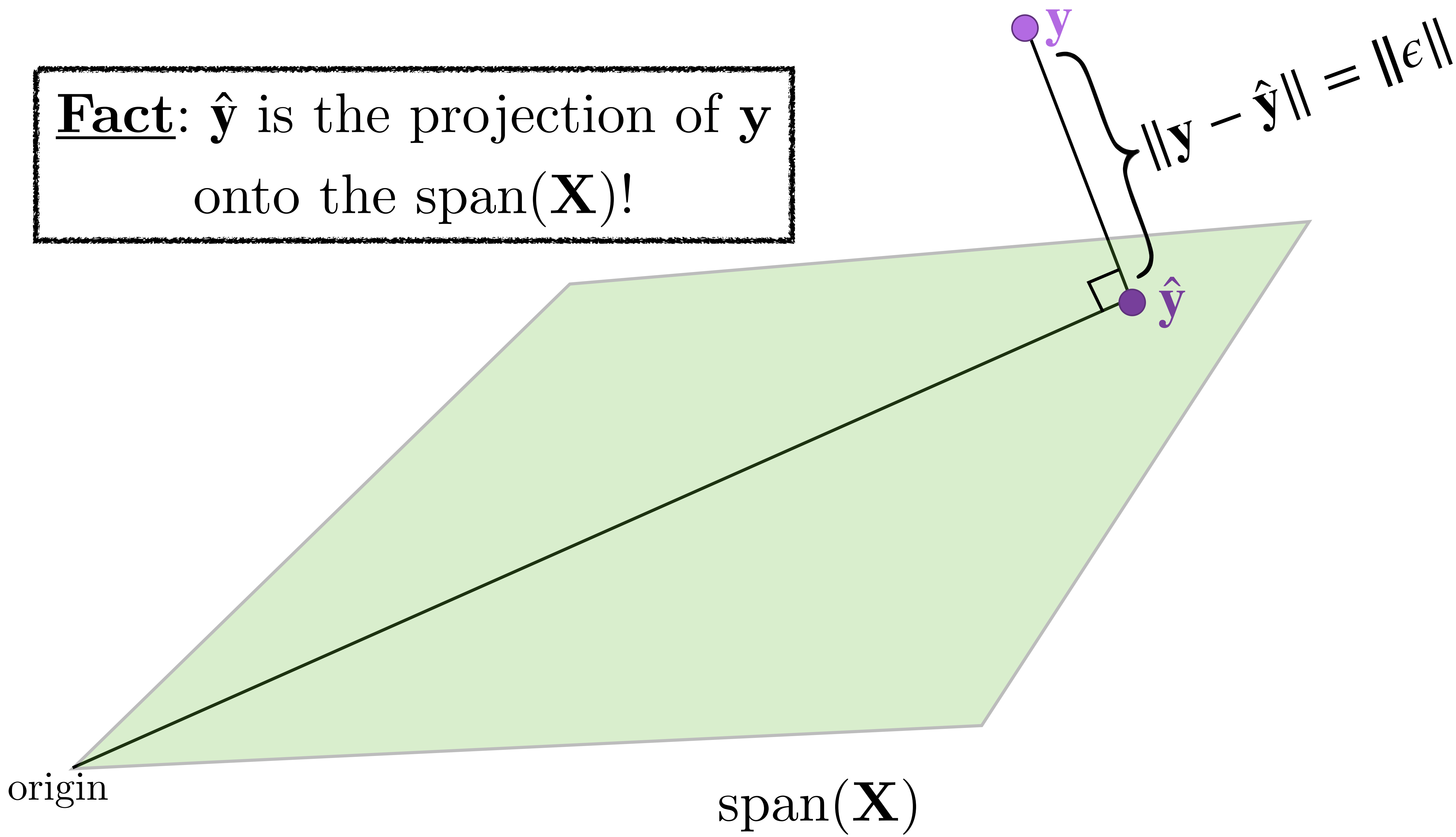
Objective: Find the point, $\hat{\mathbf{y}}$, contained in the span($\mathbf{X}$) that minimizes squared error ($\|\mathbf{y} - \hat{\mathbf{y}}\|$)

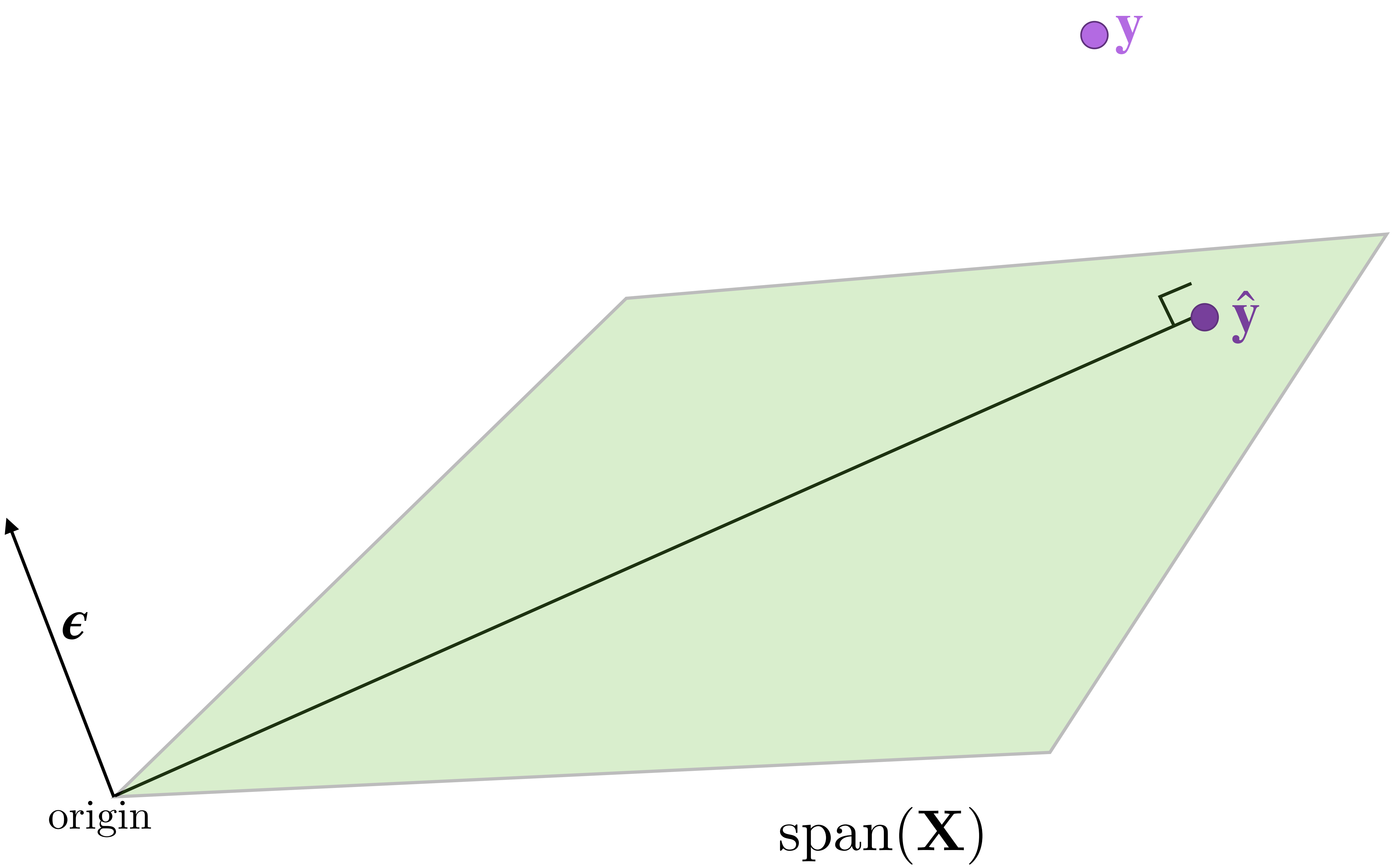
$$\widehat{\beta}_0 + \widehat{\beta}_1 \mathbf{x}_1 + \cdots + \widehat{\beta}_p \mathbf{x}_p = \widehat{\mathbf{y}}$$

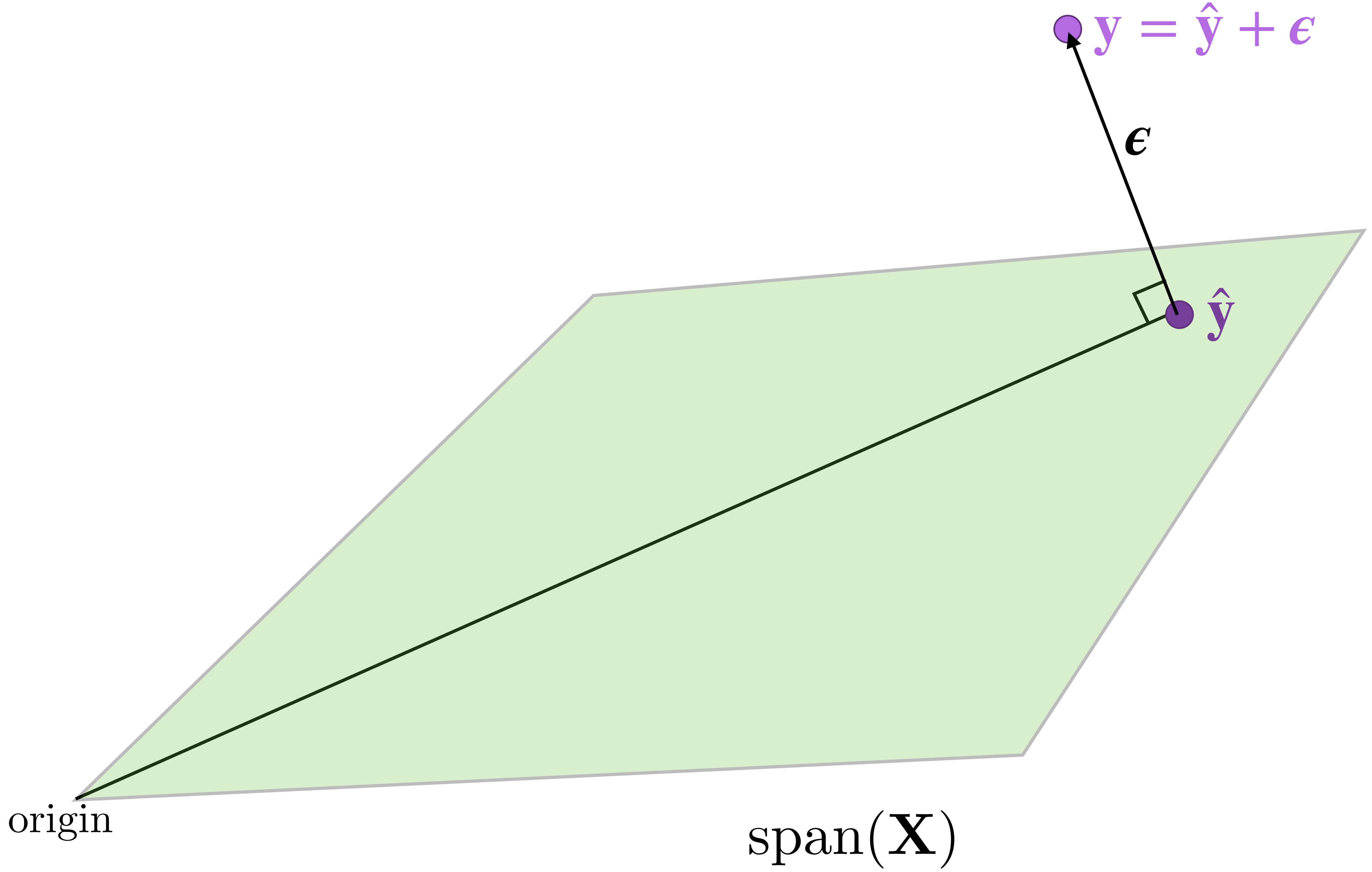
Translation: Find the closest point $\hat{\mathbf{y}}$ (closest in the *Euclidean sense*), contained in the span($\mathbf{X}$)

Translation: Find the closest point $\hat{\mathbf{y}}$ (closest in the *Euclidean sense*), contained in the span($\mathbf{X}$)

Fact: $\hat{\mathbf{y}}$ is the projection of $\mathbf{y}$ onto the span($\mathbf{X}$)!




$\mathbf{y}$
$\hat{\mathbf{y}}$
$\boldsymbol{\epsilon}$
origin
$\text{span}(\mathbf{X})$



Fine print: The span($\mathbf{X}$) is a (p+1)-dimensional subspace of $\mathbb{R}^n$. $\mathbb{R}^n$ here is what I would call the "sample (vector) space" which has an axis for each observation.

Projection Matrix

To obtain the projection of a point onto the span of the columns of any matrix $\mathbf{X}$, we multiply by the **projection matrix $\mathbf{P}_{\mathbf{X}}$,** defined as

$$\mathbf{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}$$

This is sometimes referred to as the "hat" matrix in statistics, because it puts the "hat" on $\mathbf{y}$:

$$\mathbf{X}\underbrace{(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{y}}_{\hat{\boldsymbol{\beta}}} = \hat{\mathbf{y}}$$

Major Ideas from Section

- Cosine/Angle between vectors
- Orthonormality
- Orthonormal Basis
- Orthogonal Matrix
- Orthogonal Projections