Projection onto a line

Reading: Strang 4.2

Learning objective: Be able to find the projection of a point onto a line.

Review

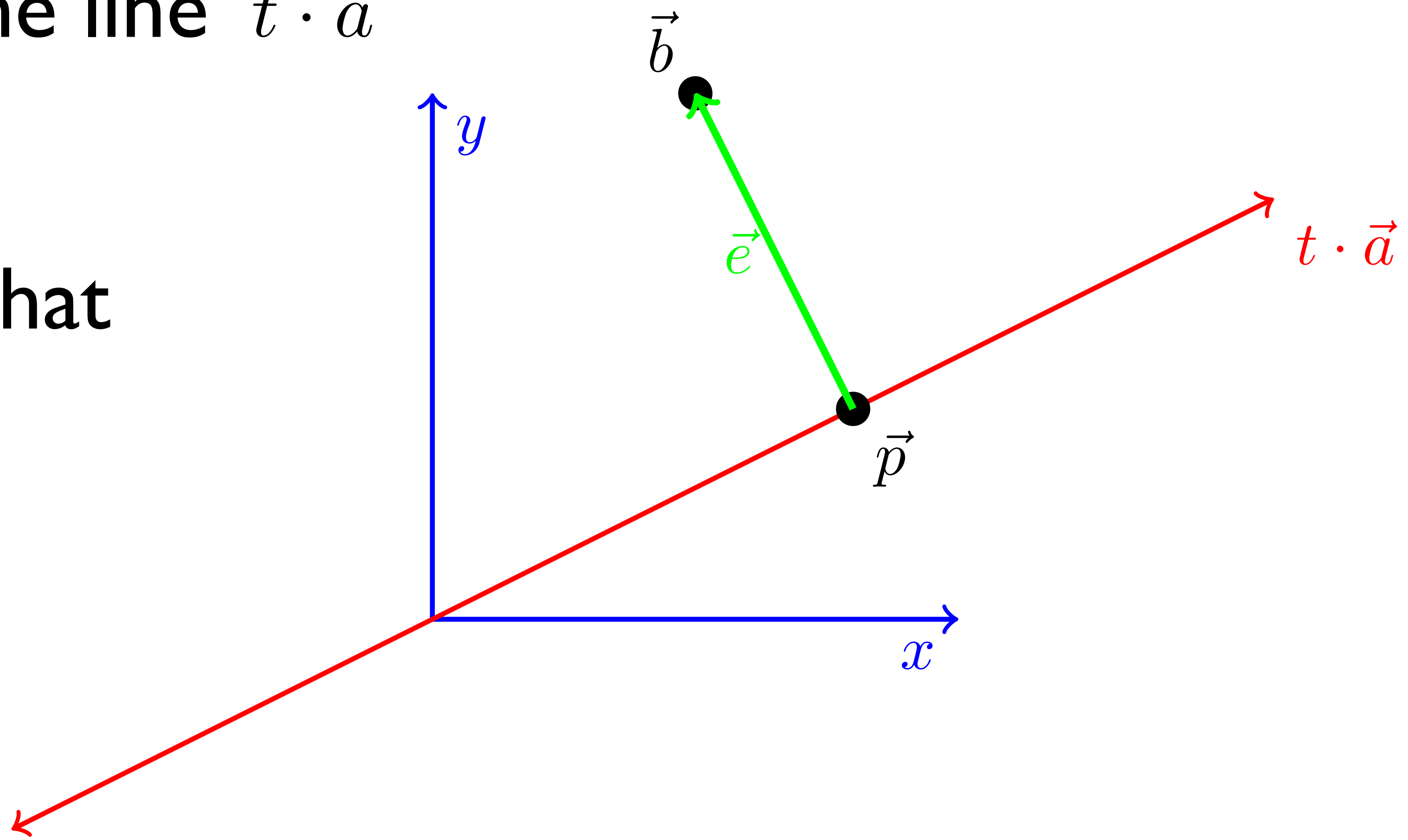
Let $\vec{a} = (2, 1)$ and $\vec{b} = (\frac{1}{2}, 1)$.

What point $\vec{p}$ on the line $t \cdot \vec{a}$ is closest to $\vec{b}$?

Key fact: $\vec{p}$ is such that

$$\vec{e} = \vec{b} - \vec{p}$$

is orthogonal to $\vec{a}$.

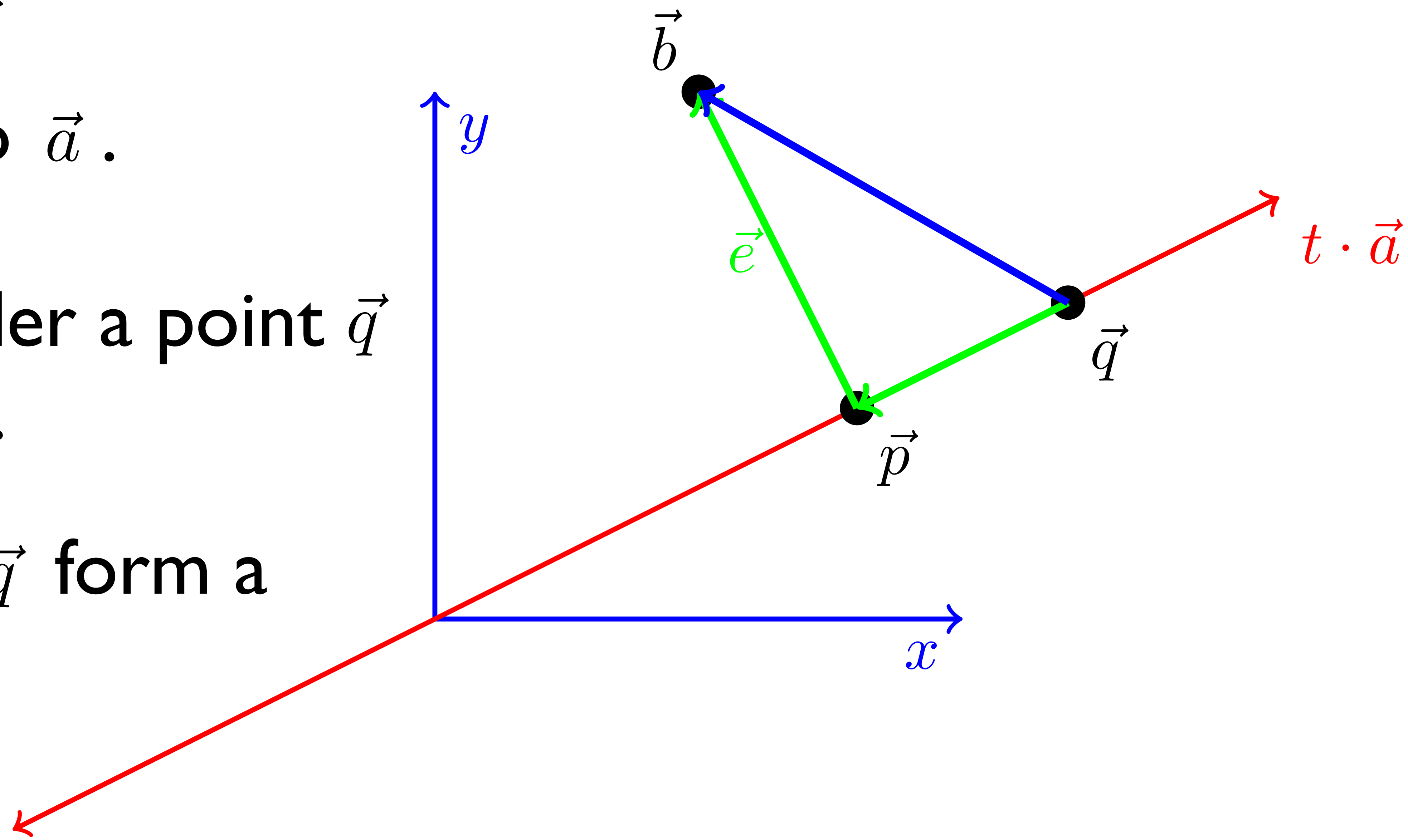


What point $\vec{p}$ on the line $t \cdot \vec{a}$ is closest to $\vec{b}$?

Claim: $\vec{p}$ is such that

$$\vec{e} = \vec{b} - \vec{p}$$

is orthogonal to $\vec{a}$.



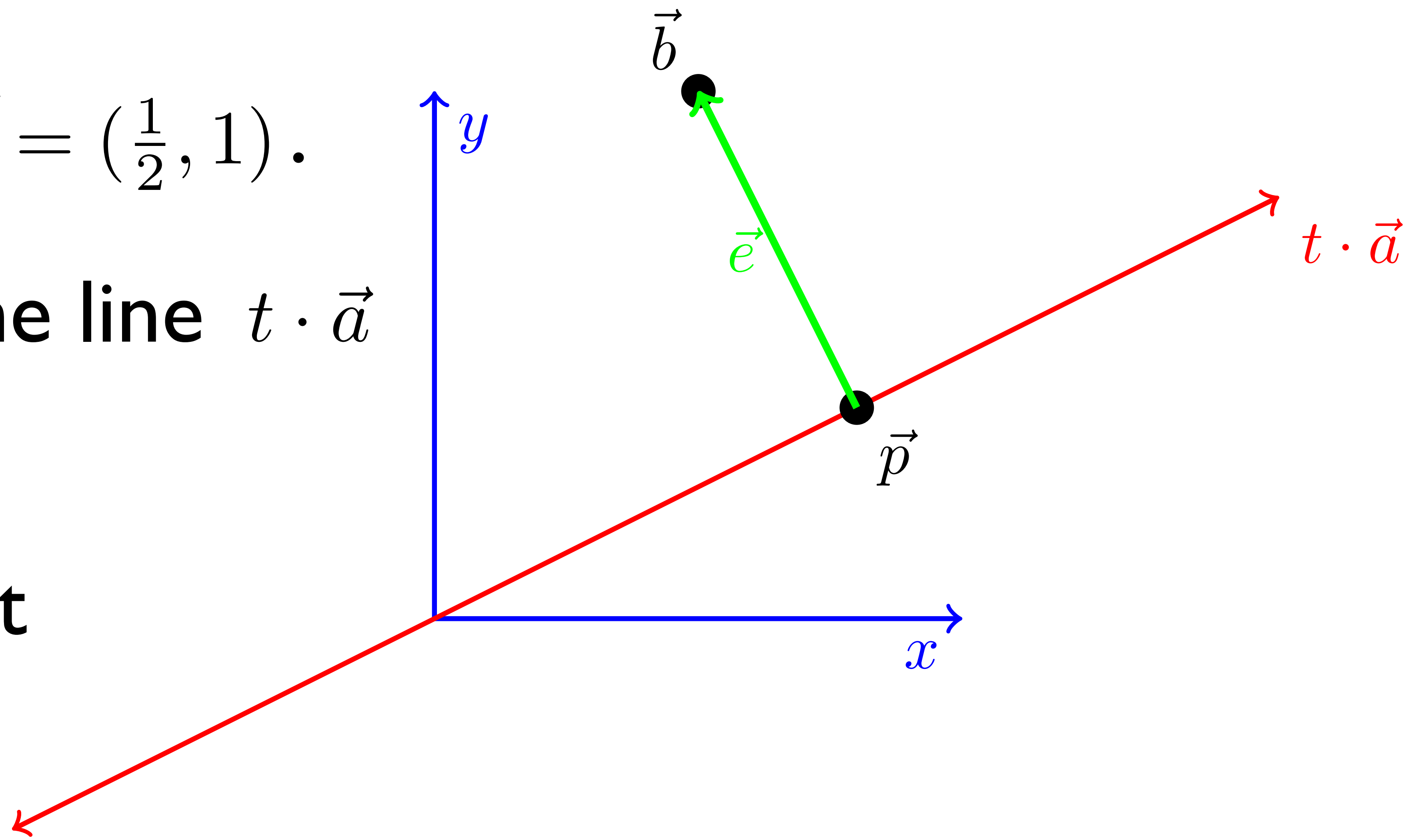
Reason: Consider a point $\vec{q}$ on the line $t \cdot \vec{a}$.

The points $\vec{b}, \vec{p}, \vec{q}$ form a right triangle.

$$\|\vec{b} - \vec{q}\|^2 = \|\vec{e}\|^2 + \|\vec{p} - \vec{q}\|^2$$

Closest point on a line

Let $\vec{a} = (2,1)$ and $\vec{b} = (\frac{1}{2}, 1)$.



What point $\vec{p}$ on the line $t \cdot \vec{a}$ is closest to $\vec{b}$?

Claim: $\vec{p}$ is such that

$$\vec{e} = \vec{b} - \vec{p}$$

is orthogonal to $\vec{a}$.

Now we can determine $\hat{x}$ and $\vec{p} = \hat{x} \cdot \vec{a}$.

$$\langle \vec{a}, \vec{b} - \hat{x} \cdot \vec{a} \rangle = 0 \implies \hat{x} = \frac{\langle \vec{a}, \vec{b} \rangle}{\| \vec{a} \|^2}$$

Closest point on a line

Let $\vec{a} = (2, 1)$ and $\vec{b} = (\frac{1}{2}, 1)$.

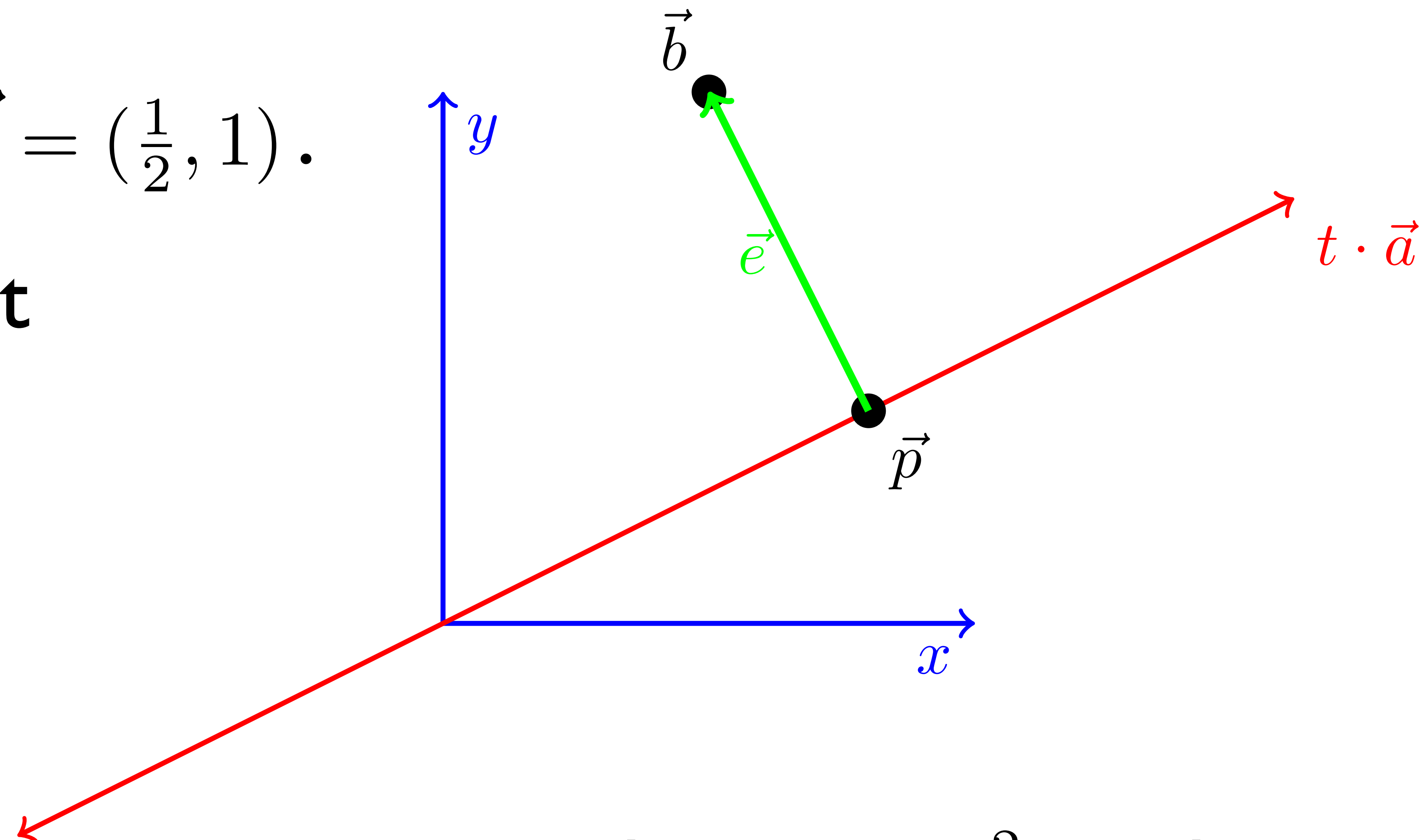
Claim: $\vec{p}$ is such that

$$\vec{e} = \vec{b} - \vec{p}$$

is orthogonal to $\vec{a}$.

$\vec{p} = \hat{x} \cdot \vec{a}$ where

$$\hat{x} = \frac{\langle \vec{a}, \vec{b} \rangle}{\|\vec{a}\|^2}$$

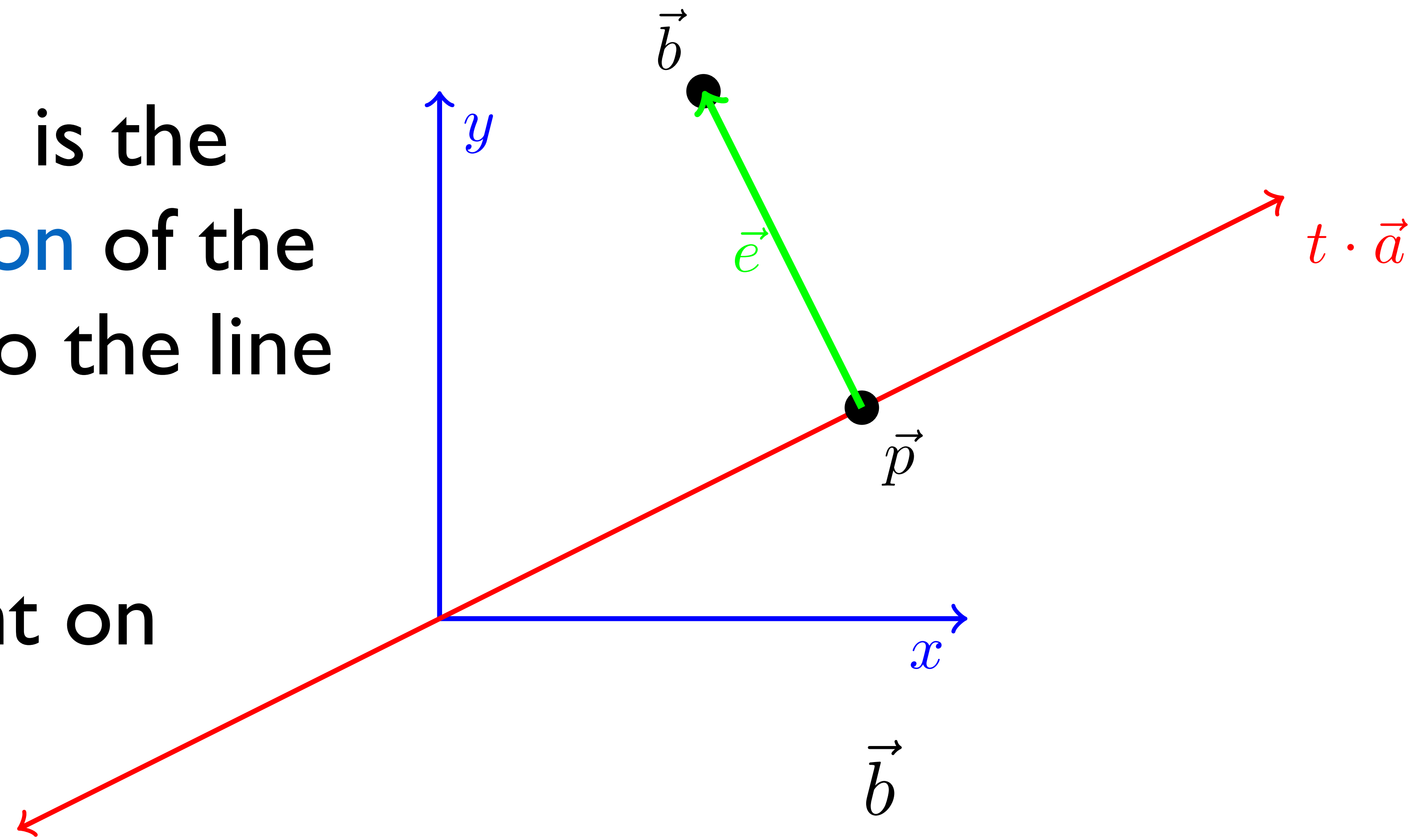


In our problem $\hat{x} = \frac{2}{5}$ and

$$\vec{p} = (\tfrac{4}{5}, \tfrac{2}{5})$$

Orthogonal Projection

The point $\vec{p} = (\frac{4}{5}, \frac{2}{5})$ is the orthogonal projection of the point $\vec{b} = (\frac{1}{2}, 1)$ onto the line $t \cdot (2, 1)$.



It is the closest point on the line to $\vec{b}$.

In this course we will only talk about orthogonal projections. We simply call $\vec{p}$ the projection of $\vec{b}$ onto the line $t \cdot \vec{a}$.

Projection onto a line

Now let's find the general formula for the projection of a vector $\vec{b} \in \mathbb{R}^n$ onto a line $t \cdot \vec{a}$.

The principle is the same: the projection is the point $\vec{p}$ such that the difference $\vec{b} - \vec{p}$ is orthogonal to $\vec{a}$.

Letting $\vec{p} = \hat{x} \cdot \vec{a}$ this means

$$\langle \vec{a}, \vec{b} - \hat{x} \cdot \vec{a} \rangle = 0 \implies \hat{x} = \frac{\langle \vec{a}, \vec{b} \rangle}{\|\vec{a}\|^2}$$

$$\vec{p} = \vec{a} \cdot \frac{\langle \vec{a}, \vec{b} \rangle}{\|\vec{a}\|^2} = \vec{a} \cdot \frac{\vec{a}^T \vec{b}}{\vec{a}^T \vec{a}}$$

Projection onto a line

Now let's find the general formula for the projection of a point $\vec{b} \in \mathbb{R}^n$ onto a line $t \cdot \vec{a}$.

$$\vec{p} = \vec{a} \cdot \frac{\langle \vec{a}, \vec{b} \rangle}{\| \vec{a} \|^2} = \vec{a} \cdot \frac{\vec{a}^T \vec{b}}{\vec{a}^T \vec{a}}$$

We can find a matrix P such that $\vec{p} = P\vec{b}$. This is called the projection matrix.

$$P = \frac{\vec{a}\,\vec{a}^T}{\vec{a}^T \vec{a}}$$

The denominator is just a number, $\|\vec{a}\|^2$.

$\vec{a}\,\vec{a}^T$ is a matrix of rank one. All column are multiples of $\vec{a}$.

Example

$$P = \frac{\vec{a}\,\vec{a}^T}{\vec{a}^T\,\vec{a}}$$

Let's go back to our example of the line $t \cdot (2, 1)$.

The projection matrix to project onto this line is given by

$$P = \frac{1}{5}\begin{bmatrix} 2 \\ 1 \end{bmatrix}\begin{bmatrix} 2 & 1 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}$$

The projection matrix allows us to project any point $\vec{b}$ onto the line $t \cdot (2, 1)$.

Example

$$P = \frac{\vec{a}\,\vec{a}^T}{\vec{a}^T\vec{a}}$$

What if we project onto the line $t \cdot (2c, c)$?

This is the same line! The projection matrix does not change.

$$P = \frac{1}{5c^2}\begin{bmatrix} 2c \\ 1c \end{bmatrix}\begin{bmatrix} 2c & 1c \end{bmatrix} = \frac{1}{5c^2}\begin{bmatrix} 4c^2 & 2c^2 \\ 2c^2 & c^2 \end{bmatrix}$$

Projecting Again

What happens if we project twice?

$$P^2 = \left(\frac{\vec{a}a^{\vec{T}}}{\vec{a}^T\vec{a}}\right)\left(\frac{\vec{a}a^{\vec{T}}}{\vec{a}^T\vec{a}}\right)$$

$$= \frac{\vec{a}}{\vec{a}^T\vec{a}}\vec{a}^T\vec{a}\frac{a^{\vec{T}}}{\vec{a}^T\vec{a}}$$

$$= \frac{\vec{a}\vec{a}^T}{\vec{a}^T\vec{a}}$$

$$= P$$

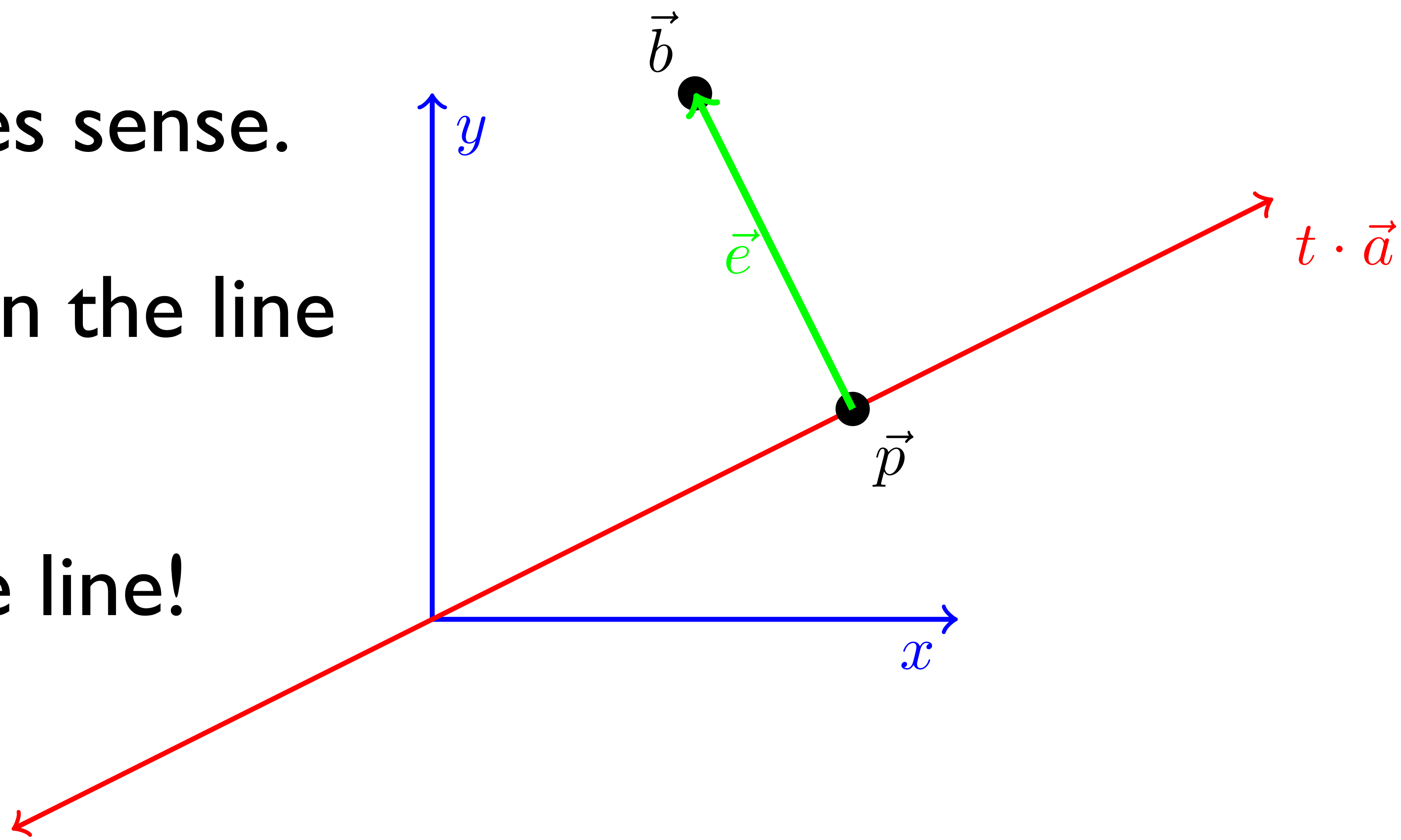
Projecting Again

What happens if we project twice? $P^2 = P$

Intuitively this makes sense.

The closest point on the line to $P\vec{b}$ is $P\vec{b}$.

$P\vec{b}$ is already on the line!



This means $PP\vec{b} = P\vec{b}$ for any vector $\vec{b}$.

Symmetry

$$P = \frac{\vec{a}\,\vec{a}^T}{\vec{a}^T\vec{a}}$$

The projection onto a line matrix is also symmetric.

What happens to the 4 subspaces with a symmetric matrix?

The column space is equal to the row space.

The nullspace is equal to the left nullspace.

The nullspace is the orthogonal complement of the column space.

Projection Onto a Subspace

Reading: Strang 4.2

Learning objective: Be able to find the projection of a point onto a subspace.

Projection Onto a Subspace

So far we have just projected onto a line through the origin.

We can project onto any subspace $S \subseteq \mathbb{R}^n$.

The projection of $\vec{b}$ onto S is the closest point to $\vec{b}$ in S.

It is the solution to the problem $\underset{\vec{p} \in S}{\text{minimize}} \; \|\vec{b} - \vec{p}\|$

Example

Example: Let $S = \text{span}(\{(1,0,0),(0,1,0)\})$ be the x-y plane.

What is the projection of $\vec{b} = (3,4,5)$ onto S?

Answer: A point in S is of the form $(a_1, a_2, 0)$.

The distance from $\vec{b}$ is

$$\|(3,4,5) - (a_1, a_2, 0)\| = \sqrt{(3 - a_1)^2 + (4 - a_2)^2 + 5^2}$$

How should we choose a_1, a_2 to minimize this?

Example

Example: Let $S = \text{span}(\{(1,0,0),(0,1,0)\})$ be the x-y plane.

What is the projection of $\vec{b} = (3,4,5)$ onto S?

Answer: A point in S is of the form $(a_1, a_2, 0)$.

$$\|(3,4,5) - (a_1, a_2, 0)\| = \sqrt{(3-a_1)^2 + (4-a_2)^2 + 5^2}$$

How should we choose a_1, a_2 to minimize this?

$$a_1 = 3, a_2 = 4$$

The projection of $\vec{b}$ onto S is $(3,4,0)$.

Example

Example: Let $S = \mathrm{span}(\{(1,0,0),(0,1,0)\})$ be the x-y plane.

What is the projection of $\vec{b} = (3,4,5)$ onto S?

The projection of $\vec{b}$ onto S is $\vec{p} = (3,4,0)$.

Note that the difference $\vec{b} - \vec{p} = (0,0,5)$ is orthogonal to S.

This is a general principle!

Review: Orthogonal Complements

Let $S \subseteq \mathbb{R}^n$ be a subspace.

A vector $\vec{u} \in \mathbb{R}^n$ is orthogonal to S iff

$$\langle \vec{u}, \vec{v} \rangle = 0 \text{ for every } \vec{v} \in S$$

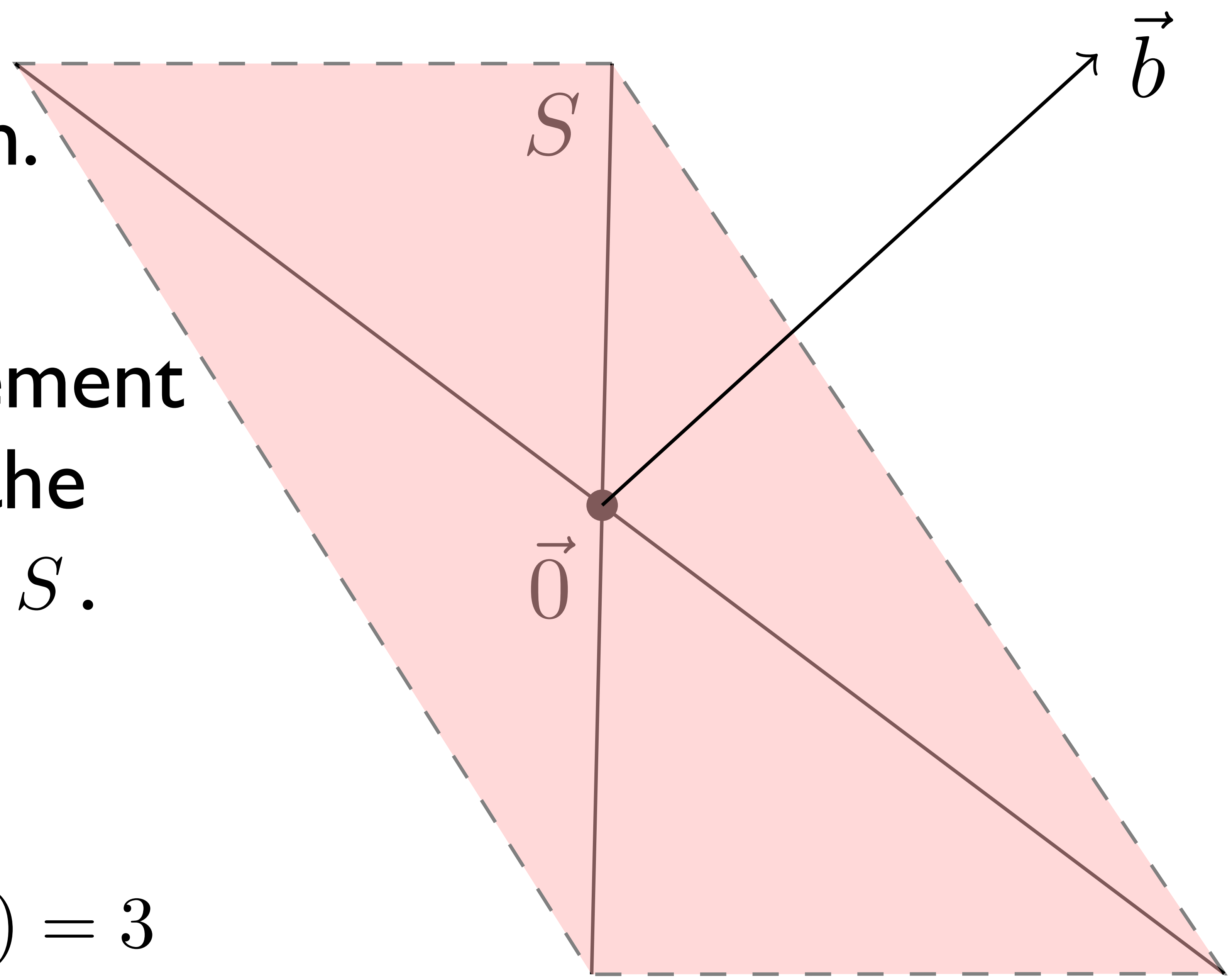
Definition: The orthogonal complement of a subspace S, denoted $S^\perp$, is the set of all vectors orthogonal to S.

$$S^\perp = \{\vec{u} : \langle \vec{u}, \vec{v} \rangle = 0 \text{ for all } \vec{v} \in S\}$$

$S^\perp$ is the largest subspace orthogonal to S.

Example

Let $S \subseteq \mathbb{R}^3$ be a plane through the origin.



The orthogonal complement of S is a line through the origin perpendicular to S.

The dimensions add up:

$$\dim(S) + \dim(S^{\perp}) = 3$$

Every vector in $\vec{u} \in \mathbb{R}^3$ can be written as

$$\vec{u} = \vec{u}_s + \vec{u}_{s^{\perp}} \text{ with } \vec{u}_s \in S, \vec{u}_{s^{\perp}} \in S^{\perp}$$

Key Fact

Let $S \subseteq \mathbb{R}^n$ be a subspace and $\vec{b} \in \mathbb{R}^n$.

> The projection $\vec{p}$ of $\vec{b}$ onto S is such that $\vec{b} - \vec{p}$ is orthogonal to S.

Reason: Write $\vec{b} = \vec{b}_s + \vec{b}_{s^\perp}$ where $\vec{b}_s \in S, \vec{b}_{s^\perp} \in S^\perp$.

$$\|\vec{b} - \vec{p}\|^2 = \|\vec{b}_{s^\perp} + \vec{b}_s - \vec{p}\|^2$$

Now $\vec{b}_{s^\perp}$ and $\vec{b}_s - \vec{p} \in S$ are orthogonal for $\vec{p} \in S$.

The projection $\vec{p}$ of $\vec{b}$ onto S is such that $\vec{b} - \vec{p}$ is orthogonal to S.

Reason: Write $\vec{b} = \vec{b}_s + \vec{b}_{s^\perp}$ where $\vec{b}_s \in S, \vec{b}_{s^\perp} \in S^\perp$.

$$\|\vec{b} - \vec{p}\|^2 = \|\vec{b}_{s^\perp} + \vec{b}_s - \vec{p}\|^2$$

Now $\vec{b}_{s^\perp}$ and $\vec{b}_s - \vec{p} \in S$ are orthogonal for $\vec{p} \in S$.

For orthogonal vectors $\vec{v}, \vec{w}$

$$\|\vec{v} + \vec{w}\|^2 = \|\vec{v}\|^2 + \|\vec{w}\|^2$$

This implies:

$$\|\vec{b}_{s^\perp} + \vec{b}_s - \vec{p}\|^2 = \|\vec{b}_{s^\perp}\|^2 + \|\vec{b}_s - \vec{p}\|^2$$

which is minimized by taking $\vec{p} = \vec{b}_s$.

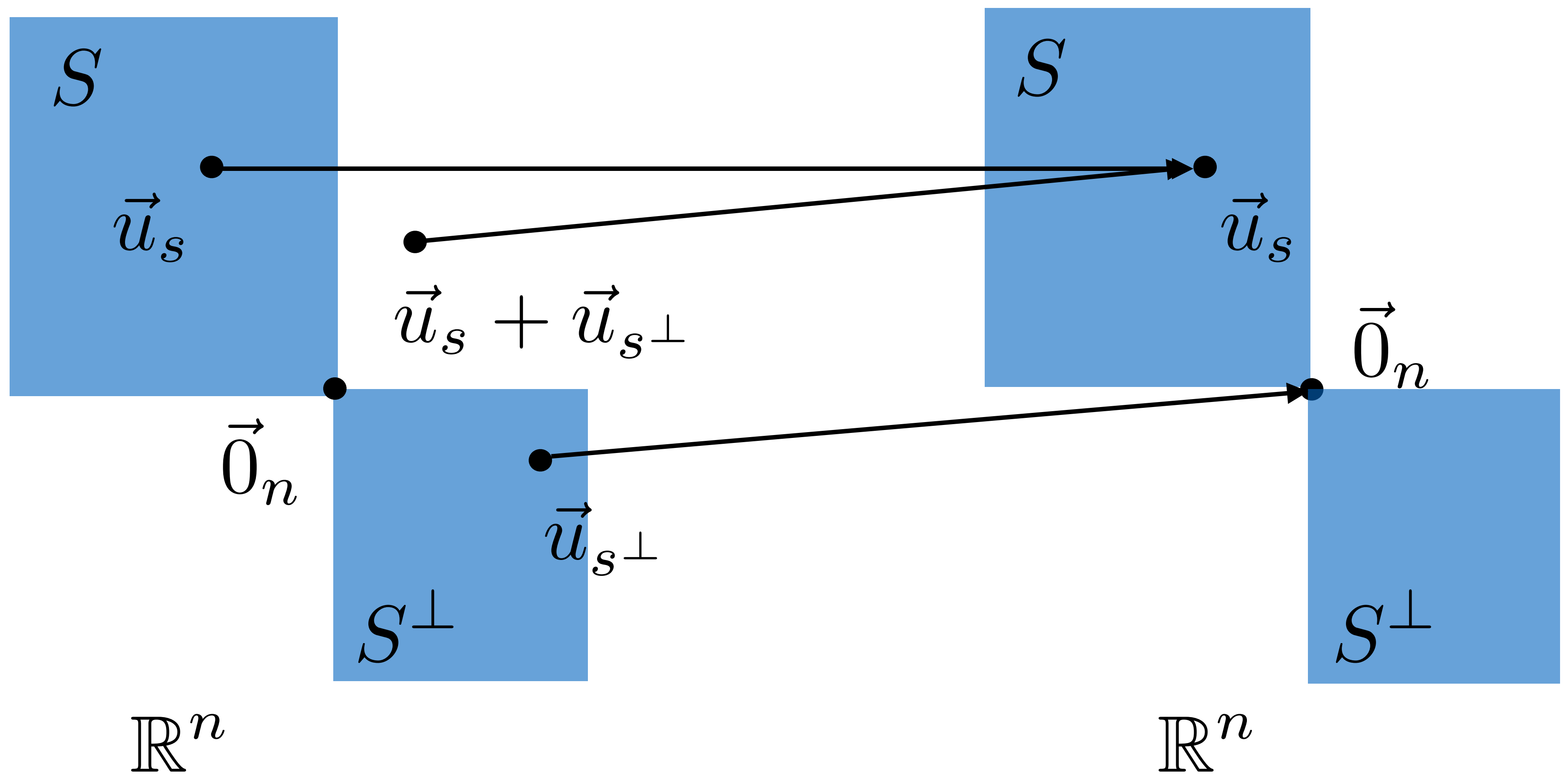
The projection $\vec{p}$ of $\vec{b}$ onto S is such that $\vec{b} - \vec{p}$ is orthogonal to S.

Conclusion: If we write $\vec{b} = \vec{b}_s + \vec{b}_{s^\perp}$ for $\vec{b}_s \in S, \vec{b}_{s^\perp} \in S^\perp$ the projection of $\vec{b}$ onto S is $\vec{p} = \vec{b}_s$.

$\vec{b} - \vec{p} = \vec{b}_{s^\perp}$ is orthogonal to S. It is in $S^\perp$.

The projection matrix P has the action $P\vec{b} = \vec{u}_s$.

Projection Onto a Subspace



Pictorial representation of the action of a projection onto a subspace $S \subseteq \mathbb{R}^n$.

Least Squares

Reading: Strang 4.3

Learning objective: Be able to find the least squares solution to a system of linear equations and know when it is appropriate to do so.

Least Squares

Sometimes the equation $A\vec{x} = \vec{b}$ does not have a solution.

Usually this is because A is a tall skinny matrix—there are more equations than unknowns.

What can we do in this situation?

We can try to find a solution that gets "close" to $\vec{b}$.

The vector of errors is given by $\vec{e} = \vec{b} - A\vec{x}$.

Least Squares

Sometimes the equation $A\vec{x} = \vec{b}$ does not have a solution.

We can try to find a solution that gets "close" to $\vec{b}$.

The vector of errors is given by $\vec{e} = \vec{b} - A\vec{x}$.

When $A\vec{x} = \vec{b}$ has a solution, we can make $\vec{e} = \vec{0}$.

When this is not possible we can try to make the length of $\vec{e}$ as small as possible.

Least Squares

Sometimes the equation $A\vec{x} = \vec{b}$ does not have a solution.

To make the length of the error vector as small as possible we want to find the $\hat{\mathrm{x}}$ that minimizes

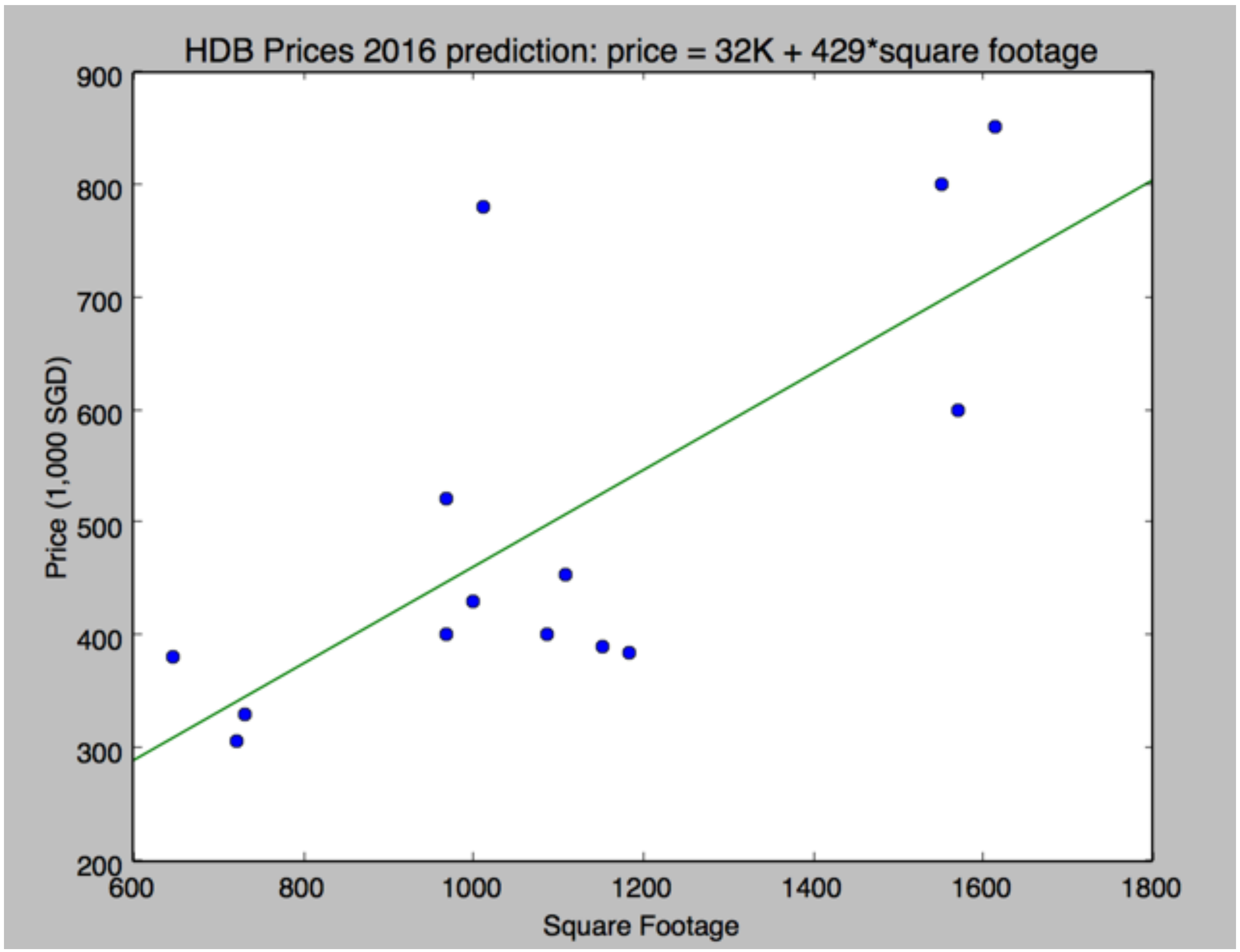
$$\|\vec{b} - A\hat{\mathbf{x}}\|$$

This $\hat{\mathrm{x}}$ is called the least squares solution to $A\vec{x} = \vec{b}$.

It minimizes the sum of the squares of the components of the error vector.

Housing Prices

Least squares solutions are enormously useful!



This line was found using least squares.

Housing Prices

A line exactly fitting the data is a solution to this equation.

$$\text{price} = a + b \cdot \text{square footage}$$

$$\begin{bmatrix} 1 & 1614 \\ 1 & 968 \\ 1 & 1184 \\ 1 & 968 \\ 1 & 1000 \\ 1 & 1152 \\ 1 & 1087 \\ 1 & 1108 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 850 \\ 400 \\ 385 \\ 520 \\ 430 \\ 390 \\ 400 \\ 453 \end{bmatrix}$$

constant term

square footage

housing price (1000 SGD)

But this equation has no solution!

Least Squares

If $A\vec{x} = \vec{b}$ has no solution then $\vec{b}$ is not in the column space of A.

If $\hat{\mathrm{x}}$ minimizes $\|\vec{b} - A\hat{\mathrm{x}}\|$ then $A\hat{\mathrm{x}}$ is the closest point in the column space of A to $\vec{b}$.

In other words, $A\hat{\mathrm{x}}$ is the projection of $\vec{b}$ onto the column space of A.

By the key fact, this means the error vector $\vec{e} = \vec{b} - A\hat{\mathrm{x}}$ is orthogonal to the column space of A.

The projection $\vec{p}$ of $\vec{b}$ onto S is such that $\vec{b} - \vec{p}$ is orthogonal to S.

The projection of $\vec{b}$ onto the column space of A is the vector $A\hat{\mathbf{x}}$ such that $\vec{b} - A\hat{\mathbf{x}}$ is orthogonal to $C(A)$.

The orthogonal complement of the column space is the left nullspace.

This means $\vec{b} - A\hat{\mathbf{x}}$ is in the left nullspace of A.

$$A^T(\vec{b} - A\hat{\mathbf{x}}) = \vec{0} \implies A^T A\hat{\mathbf{x}} = A^T\vec{b}$$

Normal Equation

A least squares solution $\hat{\mathbf{x}}$ to $A\vec{x} = \vec{b}$ satisfies the equation

$$A^T A\hat{\mathbf{x}} = A^T\vec{b}$$

This is known as the normal equation.

If $A^T A$ is invertible (almost always the case in practice) the least squares solution is unique, and given by

$$\hat{x} = (A^T A)^{-1} A^T \vec{b}$$

Projection onto column space

Assume that $A^T A$ is invertible. The minimizer of

$$\|\vec{b} - A\vec{x}\|$$

is given by $\hat{x} = (A^T A)^{-1} A^T \vec{b}$.

The projection of $\vec{b}$ onto the column space of A is

$$A\hat{\mathbf{x}} = A(A^T A)^{-1} A^T \vec{b}$$

The projection matrix onto the column space of A is

$$A(A^T A)^{-1} A^T$$