Orthogonality of the four subspaces

Reading: Strang 4.1

Learning objective: See the role of the row space, nullspace, and column space in matrix multiplication.

Review

Let A be an m-by-n matrix with r pivots.

Row space: $C(A^T) \subseteq \mathbb{R}^n$ and $\dim(C(A^T)) = r$.

Nullspace: $N(A) \subseteq \mathbb{R}^n$ and $\dim(N(A)) = n - r$.

The row space and the nullspace are orthogonal subspaces of $\mathbb{R}^n$.

This means that $\langle \vec{v}, \vec{w} \rangle = 0$ for every $\vec{v} \in C(A^T), \vec{w} \in N(A)$.

The only vector in both the row space and the nullspace is $\vec{0}_n$.

Sum of Subspaces

Let $S, T \subseteq \mathbb{R}^n$ be subspaces.

The sum of S and T is

$$S + T = \{\vec{u} + \vec{v} : \vec{u} \in S, \vec{v} \in T\}$$

On this week's problem set you show

1) $S + T$ is a subspace.

2) If $S \cap T = \{\vec{0}_n\}$ then $\dim(S + T) = \dim(S) + \dim(T)$.

Let $S, T \subseteq \mathbb{R}^n$ be subspaces.

The sum of S and T is

$$S + T = \{\vec{u} + \vec{v} : \vec{u} \in S, \vec{v} \in T\}$$

1) $S + T$ is a subspace.

2) If $S \cap T = \{\vec{0}_n\}$ then $\dim(S + T) = \dim(S) + \dim(T)$.

Let's apply this to $N(A)$ and $C(A^T)$.

$$N(A) \cap C(A^T) = \{\vec{0}_n\}$$

$$\dim(N(A)) + \dim(C(A^T)) = n$$

This means that $C(A^T) + N(A) \subseteq \mathbb{R}^n$ is a subspace of dimension n .

Row space-nullspace decomposition

$C(A^T) + N(A) \subseteq \mathbb{R}^n$ is a subspace of dimension n .

Therefore $C(A^T) + N(A) = \mathbb{R}^n$.

Every vector in $\mathbb{R}^n$ can be written as a sum of a vector in the row space of A plus a vector in the nullspace of A.

For any $\vec{u} \in \mathbb{R}^n$ we can find $\vec{u}_r \in C(A^T)$ and $\vec{u}_n \in N(A)$ such that $\vec{u} = \vec{u}_r + \vec{u}_n$.

Row space-nullspace decomposition

For any $\vec{u} \in \mathbb{R}^n$ we can find $\vec{u}_r \in C(A^T)$ and $\vec{u}_n \in N(A)$ such that $\vec{u} = \vec{u}_r + \vec{u}_n$.

Moreover, this decomposition is unique. Suppose

$$\vec{u}_r + \vec{u}_n = \vec{u}_{r'} + \vec{u}_{n'}$$

with $\vec{u}_r, \vec{u}_{r'} \in C(A^T), \vec{u}_n, \vec{u}_{n'} \in N(A)$.

Then $\vec{u}_r - \vec{u}_{r'} = \vec{u}_{n'} - \vec{u}_n$ is a vector both in $C(A^T)$ and $N(A)$.

Thus $\vec{u}_r - \vec{u}_{r'} = \vec{u}_{n'} - \vec{u}_n = \vec{0}_n$.

Multiplication by A

This gives us a new view of the action of multiplication by A. Say that $\vec{u} = \vec{u}_r + \vec{u}_n$.

$$\begin{aligned} A\vec{u} = A(\vec{u}_r + \vec{u}_n) &= A\vec{u}_r + A\vec{u}_n \\ &= A\vec{u}_r + \vec{0}_m \\ &= A\vec{u}_r \end{aligned}$$

The part of $\vec{u}$ in the row space goes into the column space of A. The part of $\vec{u}$ in the nullspace goes to $\vec{0}_m$.

For any vector $\vec{b}$ in the column space of A, there is a vector $\vec{u}_r$ in the row space with $A\vec{u}_r = \vec{b}$.

Multiplication by A

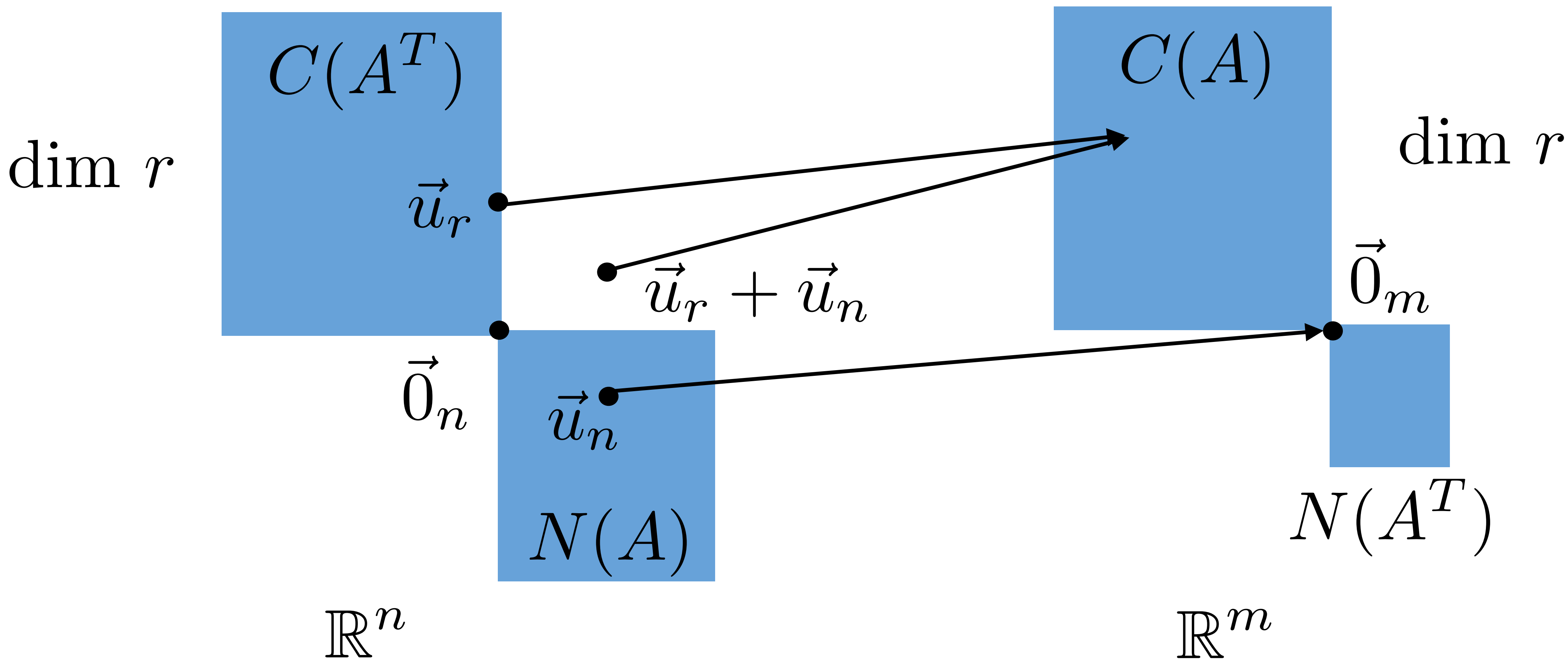
Moreover, for each vector $\vec{b}$ in the column space of A there is a unique vector $\vec{u}_r \in C(A^T)$ such that $A\vec{u}_r = \vec{b}$.

Suppose that $\vec{b} = A\vec{u}_r = A\vec{u}_{r'}$ with $\vec{u}_r, \vec{u}_{r'} \in C(A^T)$.

Then
$$\begin{aligned} \vec{0}_m &= A\vec{u}_r - A\vec{u}_{r'} \\ &= A(\vec{u}_r - \vec{u}_{r'}) \end{aligned}$$

Thus

$$\vec{u}_r - \vec{u}_{r'} \in C(A^T) \cap N(A) \implies \vec{u}_r - \vec{u}_{r'} = \vec{0}_n$$



Pictorial representation of multiplication by an m-by-n matrix A with r pivots.

Multiplication by A is one-to-one and onto as a function from $C(A^T)$ to $C(A)$.

Orthogonal Complements

Reading: Strang 4.1

Learning objective: Understand the properties of orthogonal complements.

Orthogonal Complement

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

We say a vector $\vec{u} \in \mathbb{R}^n$ is orthogonal to S if

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

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

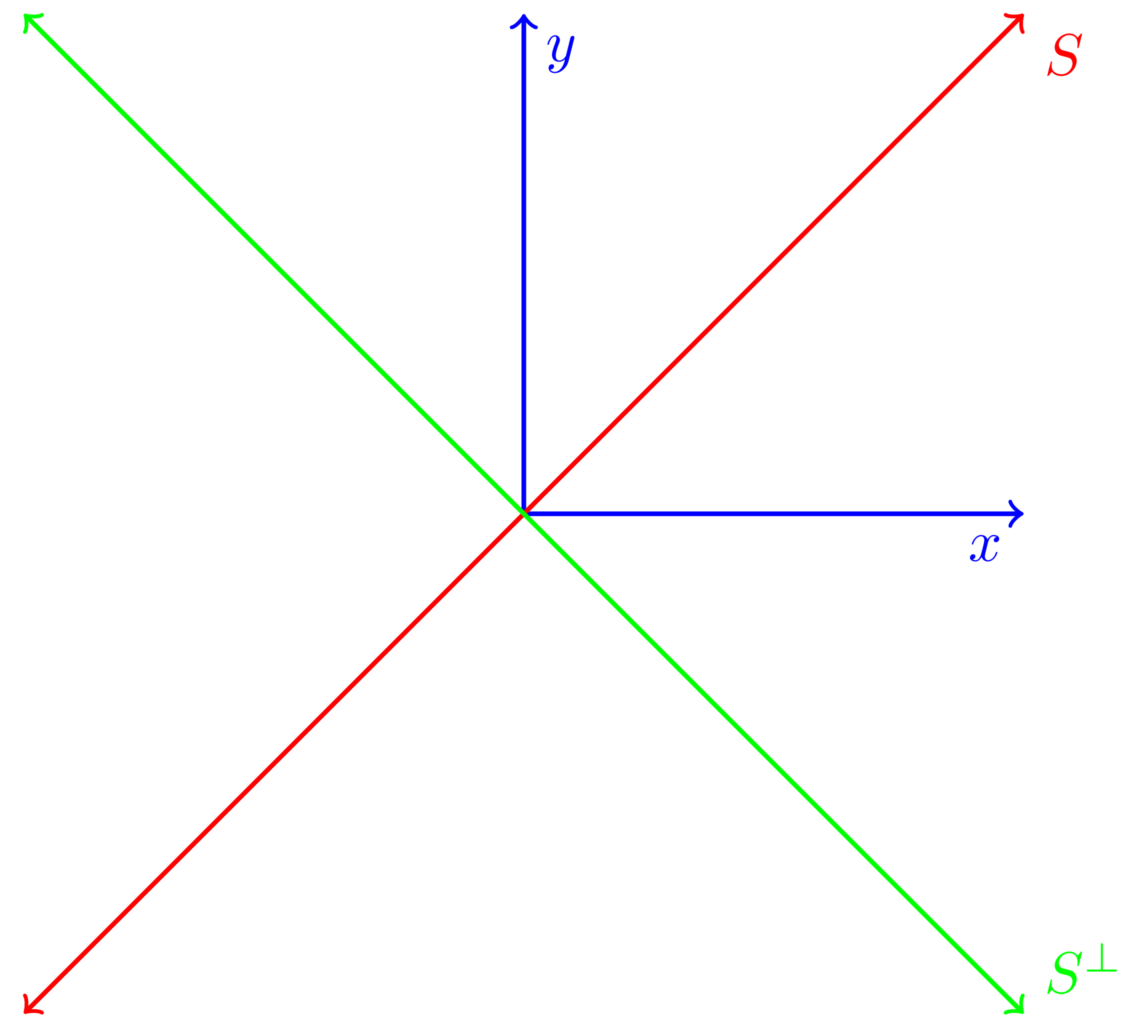
$$S^\perp = \{\vec{u} : \langle \vec{u}, \vec{v} \rangle = 0 \text{ for all } \vec{v} \in S\}$$

$S^\perp$ is read as "S perp".

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

Example

Let $S = \{t \cdot (1,1) : t \in \mathbb{R}\}$.



Let's compute $S^{\perp}$.

If $\vec{v} = (v_1, v_2)$ is orthogonal to S then

$$\begin{bmatrix} v_1 & v_2 \end{bmatrix} \begin{bmatrix} t \\ t \end{bmatrix} = 0 \text{ for all } t \in \mathbb{R}$$

$$\implies t(v_1 + v_2) = 0 \text{ for all } t \in \mathbb{R}$$

$$\implies (v_1 + v_2) = 0$$

$$S^{\perp} = \{s \cdot (1, -1) : s \in \mathbb{R}\}$$

Orthogonal Complement is a Subspace

Let $S \subseteq \mathbb{R}^n$ be a subspace. The orthogonal complement $S^\perp$ is also a subspace.

1) $\vec{0}_n$ is orthogonal to everything in S thus $\vec{0}_n \in S^\perp$.

2) If $\vec{v}_1, \vec{v}_2 \in S^\perp$ then for any $\vec{w} \in S$

$$\begin{aligned}\langle \vec{v}_1 + \vec{v}_2, \vec{w} \rangle &= \langle \vec{v}_1, \vec{w} \rangle + \langle \vec{v}_2, \vec{w} \rangle \\ &= 0\end{aligned}$$

3) If $\vec{v} \in S^\perp$ then for any $\vec{w} \in S$

$$\langle c \cdot \vec{v}, \vec{w} \rangle = c \cdot \langle \vec{v}, \vec{w} \rangle = 0$$

Example

Let A be an m-by-n matrix with r pivots.

The nullspace of A is the orthogonal complement of the row space $C(A^T)$.

$$C(A^T)^\perp \subseteq N(A)$$

If $\vec{v}$ is orthogonal to $C(A^T)$ then in particular it is orthogonal to every row of A. Thus $A\vec{v} = \vec{0}_m$.

$$N(A) \subseteq C(A^T)^\perp$$

If $\vec{v}$ is in $N(A)$ then $\vec{v}$ is orthogonal to all rows of A as $A\vec{v} = \vec{0}_m$. Then $\vec{v}$ is also orthogonal to all linear combinations of rows of A.

Example

Let A be an m-by-n matrix with r pivots.

The row space of A is the orthogonal complement of the nullspace.

$$C(A^T) \subseteq N(A)^{\perp}$$

The row space is orthogonal to the nullspace.

Let A be an m-by-n matrix with r pivots.

The row space of A is the orthogonal complement of the nullspace.

$$N(A)^{\perp} \subseteq C(A^T)$$

If there was a vector $\vec{v}$ orthogonal to the nullspace of A and not in the row space, then we could make a new matrix $B = [A; \vec{v}^T]$.

Then $N(A) = N(B)$ but

$$\dim(C(B^T)) = \dim(C(A^T)) + 1$$

a contradiction.

General Properties

The properties we have seen for the row space and the null space hold in general for orthogonal complements.

Let $S \subseteq \mathbb{R}^n$ be a subspace of dimension k. Then

§ $S^\perp$ is a subspace of dimension $n-k$.

§ $S \cap S^\perp = \{\vec{0}_n\}$.

§ $S + S^\perp = \mathbb{R}^n$.

§ Each $\vec{u} \in \mathbb{R}^n$ can be written as $\vec{u} = \vec{u}_s + \vec{u}_{s^\perp}$ with $\vec{u}_s \in S, \vec{u}_{s^\perp} \in S^\perp$ in a unique way.

§ $(S^\perp)^\perp = S$.

General Properties

The properties we have seen for the row space and the null space hold in general for orthogonal complements.

Let $S \subseteq \mathbb{R}^n$ be a subspace of dimension k. Then

§ $S^\perp$ is a subspace of dimension $n-k$.

To prove any of these, let $\vec{v}_1, \ldots, \vec{v}_k$ be a basis for S.

Create a matrix A whose rows are $\vec{v}_1, \ldots, \vec{v}_k$.

Then $C(A^T) = S$ and $N(A) = S^\perp$.

We can apply everything we have already learned about the row space and nullspace!

Example

Determine $S^{\perp}$ when $S = \mathrm{span}\{(1,2,5,4),(3,7,3,12)\}$.

Let's create a matrix A whose row space is S.

$$A = \begin{bmatrix} 1 & 2 & 5 & 4 \\ 3 & 7 & 3 & 12 \end{bmatrix}$$

Now the question boils down to finding the nullspace.

Determine $S^{\perp}$ when $S = \text{span}\{(1,2,5,4),(3,7,3,12)\}$.

Let's create a matrix A whose row space is S.

$$A = \begin{bmatrix} 1 & 2 & 5 & 4 \\ 3 & 7 & 3 & 12 \end{bmatrix}$$

Now the question boils down to finding the nullspace.

$$\begin{bmatrix} 1 & 2 & 5 & 4 \\ 3 & 7 & 3 & 12 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 2 & 5 & 4 \\ 0 & 1 & -12 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 & 29 & 4 \\ 0 & 1 & -12 & 0 \end{bmatrix}$$

The special solution for x_3 is $(-29, 12, 1, 0)$.

The special solution for x_4 is $(-4, 0, 0, 1)$.

Determine $S^\perp$ when $S = \text{span}\{(1,2,5,4),(3,7,3,12)\}$.

Let's create a matrix A whose row space is S.

$$A = \begin{bmatrix} 1 & 2 & 5 & 4 \\ 3 & 7 & 3 & 12 \end{bmatrix}$$

Now the question boils down to finding the nullspace.

The special solution for x_3 is $(-29, 12, 1, 0)$.

The special solution for x_4 is $(-4, 0, 0, 1)$.

$$S^\perp = \text{span}\{(-29,12,1,0),(-4,0,0,1)\}$$

Column space and Left Nullspace

Let A be an m-by-n matrix with r pivots.

Column space: $C(A) \subseteq \mathbb{R}^m$ and $\dim(C(A)) = r$.

Left nullspace: $N(A^T) \subseteq \mathbb{R}^m$ and $\dim(N(A^T)) = m - r$.

The column space and left nullspace live in the same vector space so we can look at their relationship.

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

Example

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ -2 & -1 & 0 \end{bmatrix}$$

The vector $(-2, 1, 1)$ is in the left nullspace.

It is orthogonal to each column of the matrix.

It is orthogonal to any linear combination of the columns.

Column space and Left Nullspace

Let A be an m-by-n matrix with r pivots.

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

We can see this by considering the matrix A^T.

The column space of A is the row space of A^T.

The left nullspace of A is the nullspace of A^T.

The nullspace of A^T is the orthogonal complement of the row space of A^T.

Summary

Let A be an m-by-n matrix with r pivots.

Row space: $C(A^T) \subseteq \mathbb{R}^n$ and $\dim(C(A^T)) = r$.

Nullspace: $N(A) \subseteq \mathbb{R}^n$ and $\dim(N(A)) = n - r$.

The nullspace is the orthogonal complement of the row space.

Column space: $C(A) \subseteq \mathbb{R}^m$ and $\dim(C(A)) = r$.

Left nullspace: $N(A^T) \subseteq \mathbb{R}^m$ and $\dim(N(A^T)) = m - r$.

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

Projections

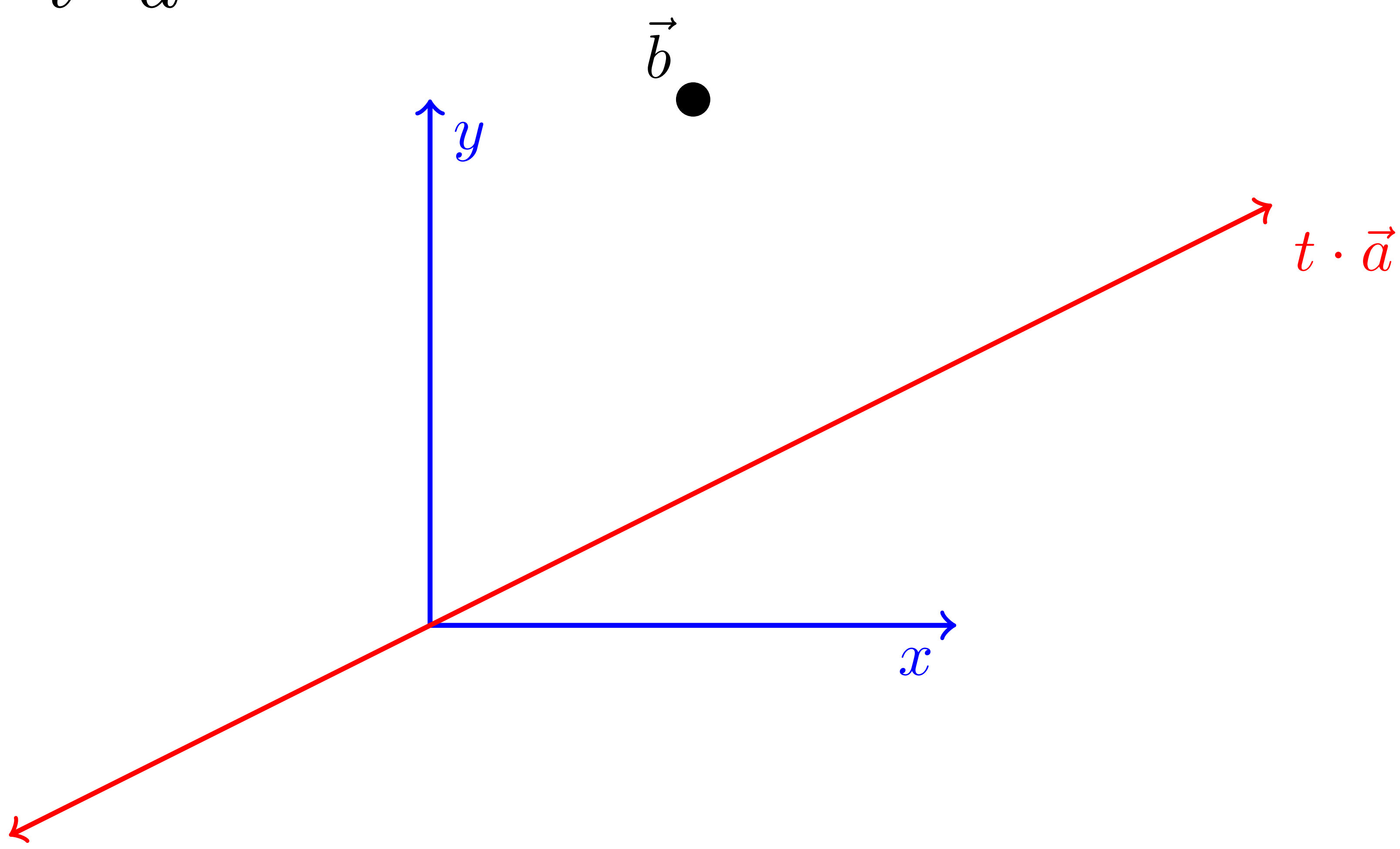
Reading: Strang 4.2

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

Closest point on a line

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

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



Let's call this point $\vec{p}$.

As $\vec{p}$ lies on the line,

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

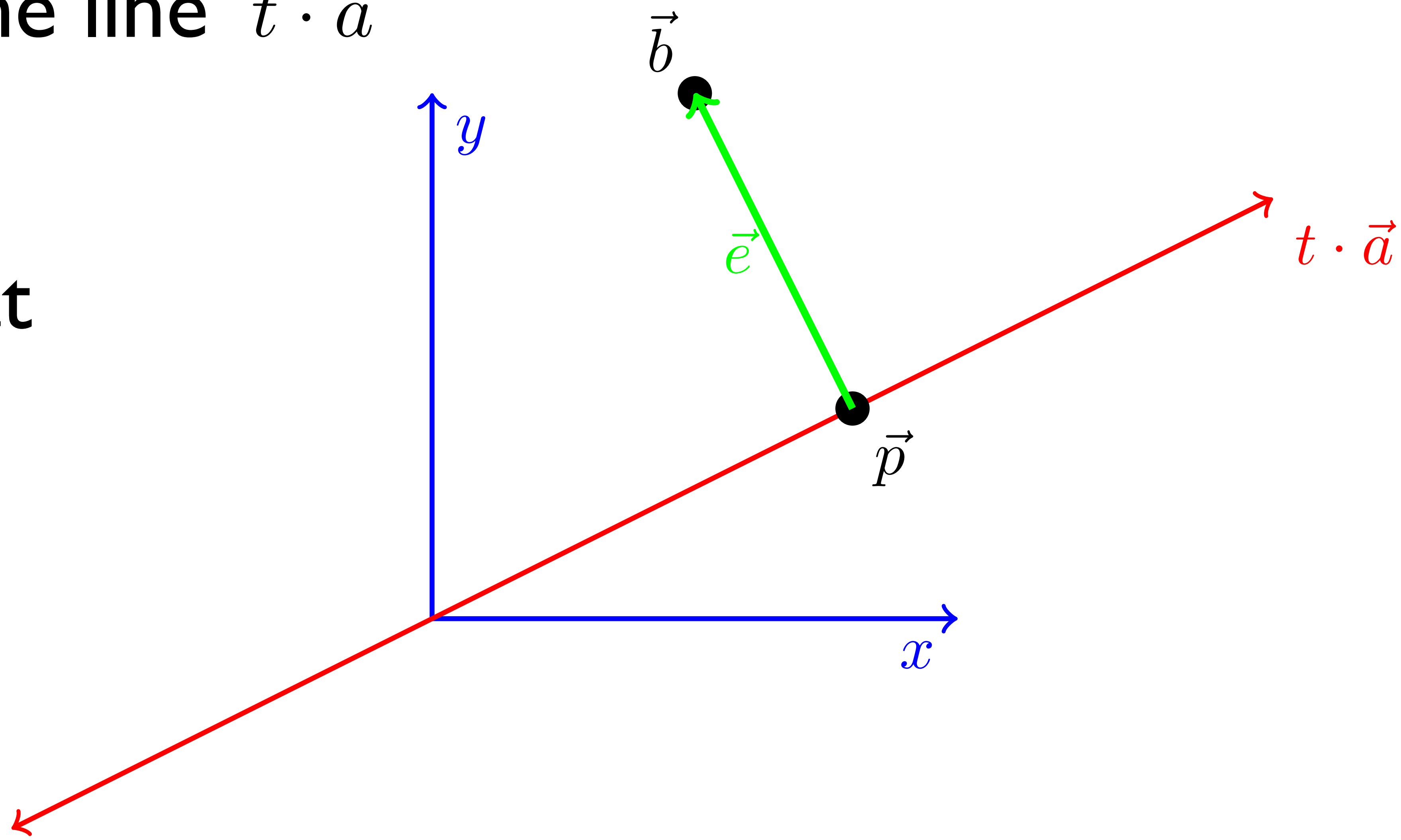
for some $\hat{x} \in \mathbb{R}$.

We want to determine the value of $\hat{x}$.

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 $\vec{a}$.

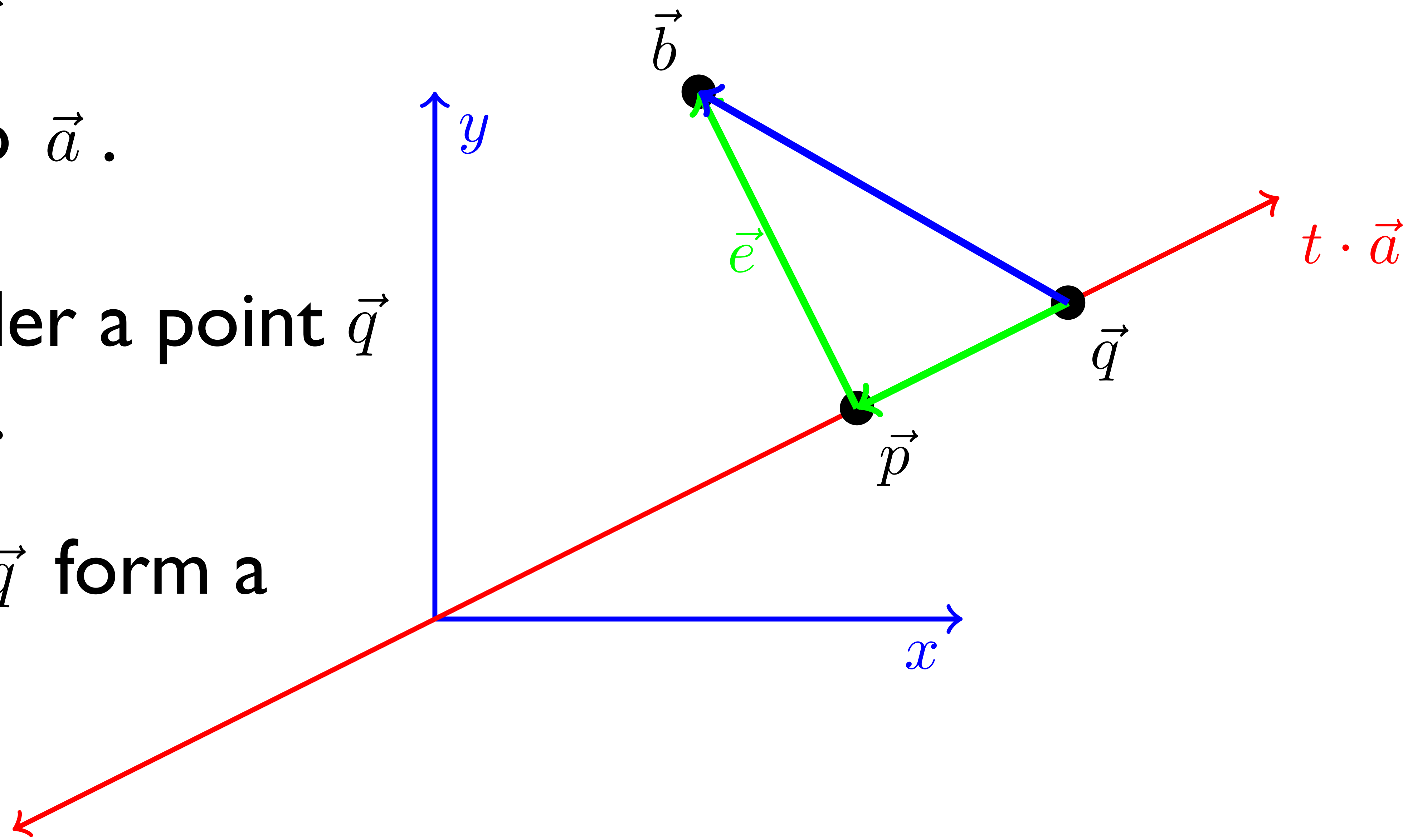
In this case the distance is $\|\vec{e}\|$.

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.

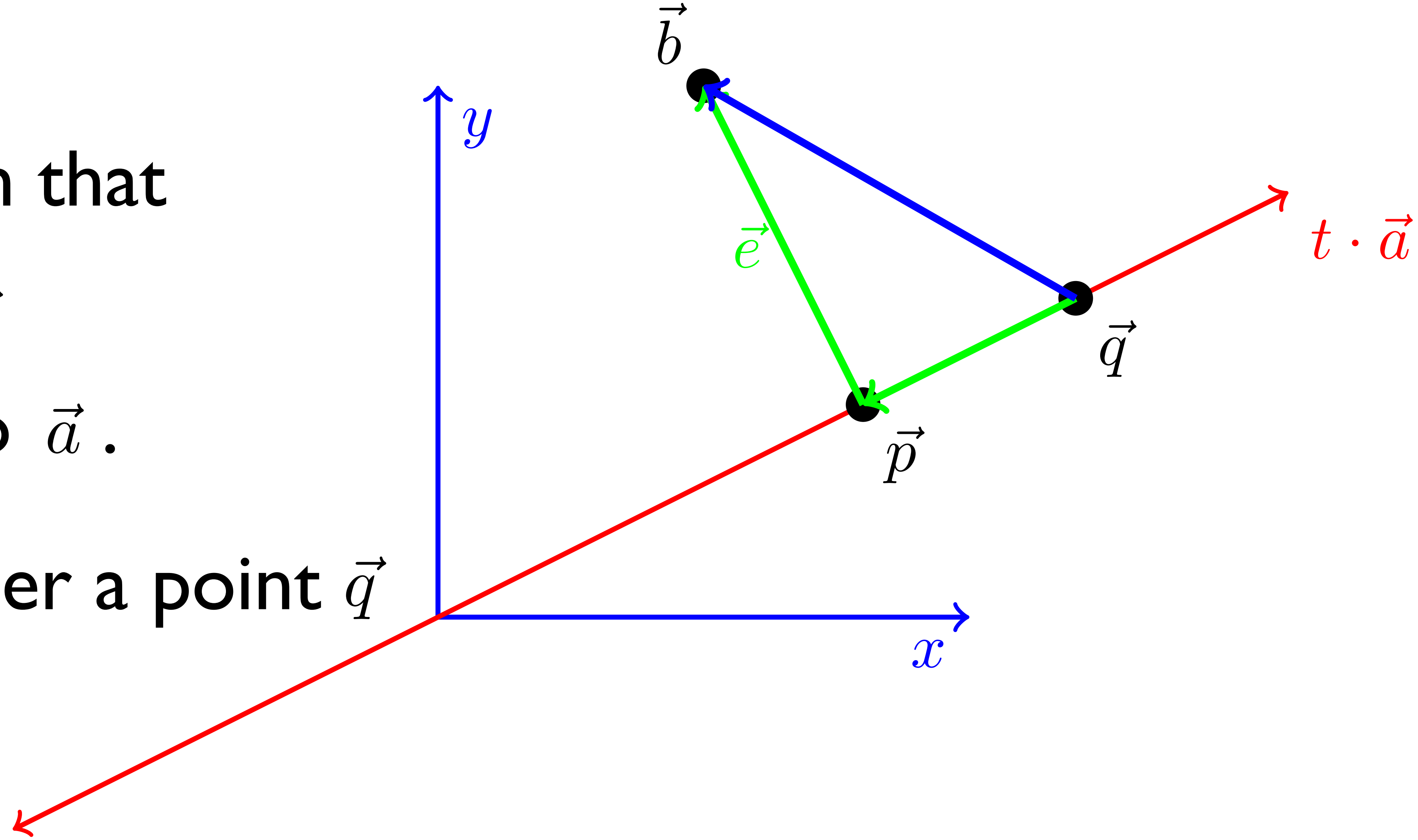
$$\|\vec{b} - \vec{q}\|^2 = \|\vec{e}\|^2 + \|\vec{p} - \vec{q}\|^2$$

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 $t \cdot \vec{a}$.



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

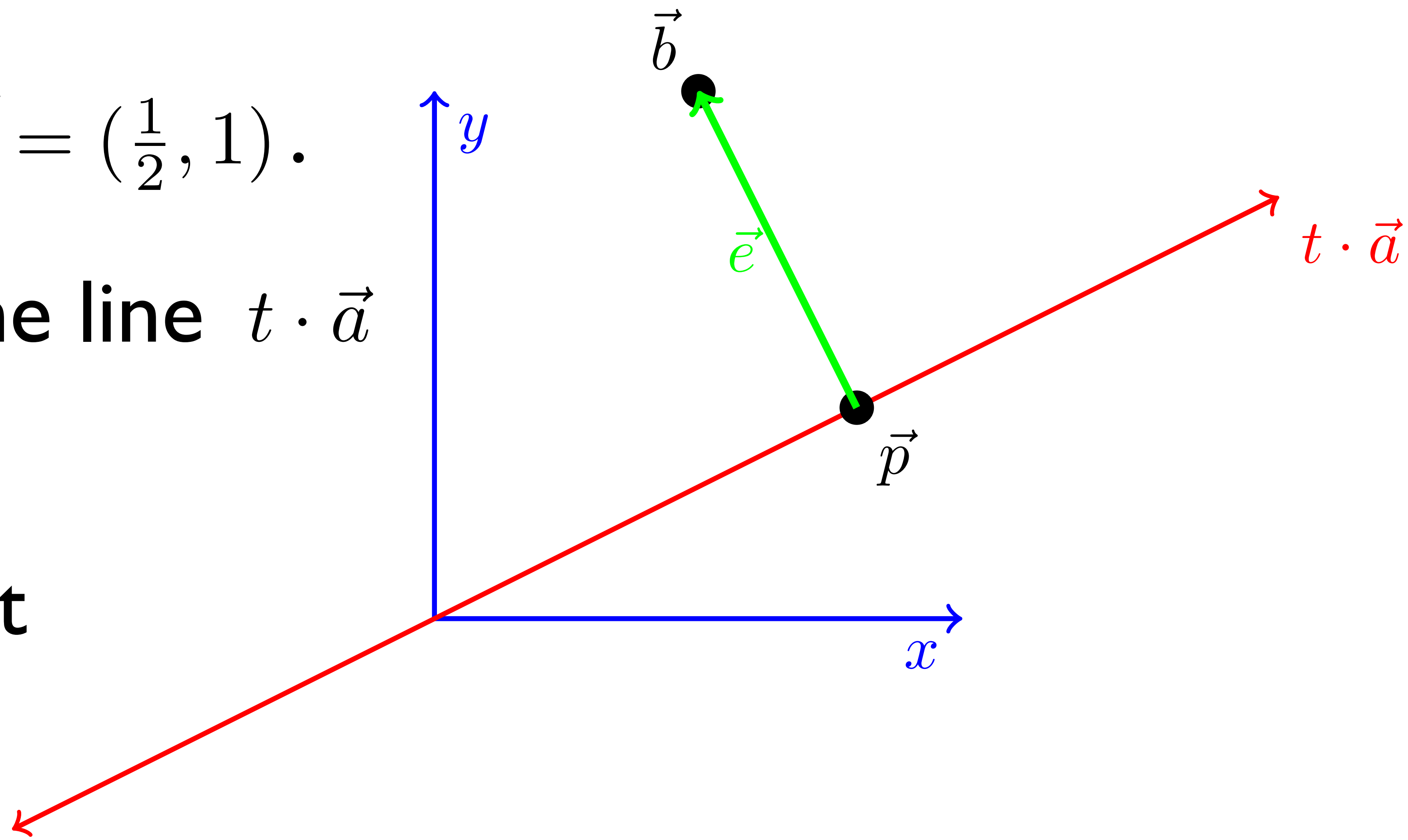
independent of $\vec{q}$

minimized when $\vec{p} = \vec{q}$

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

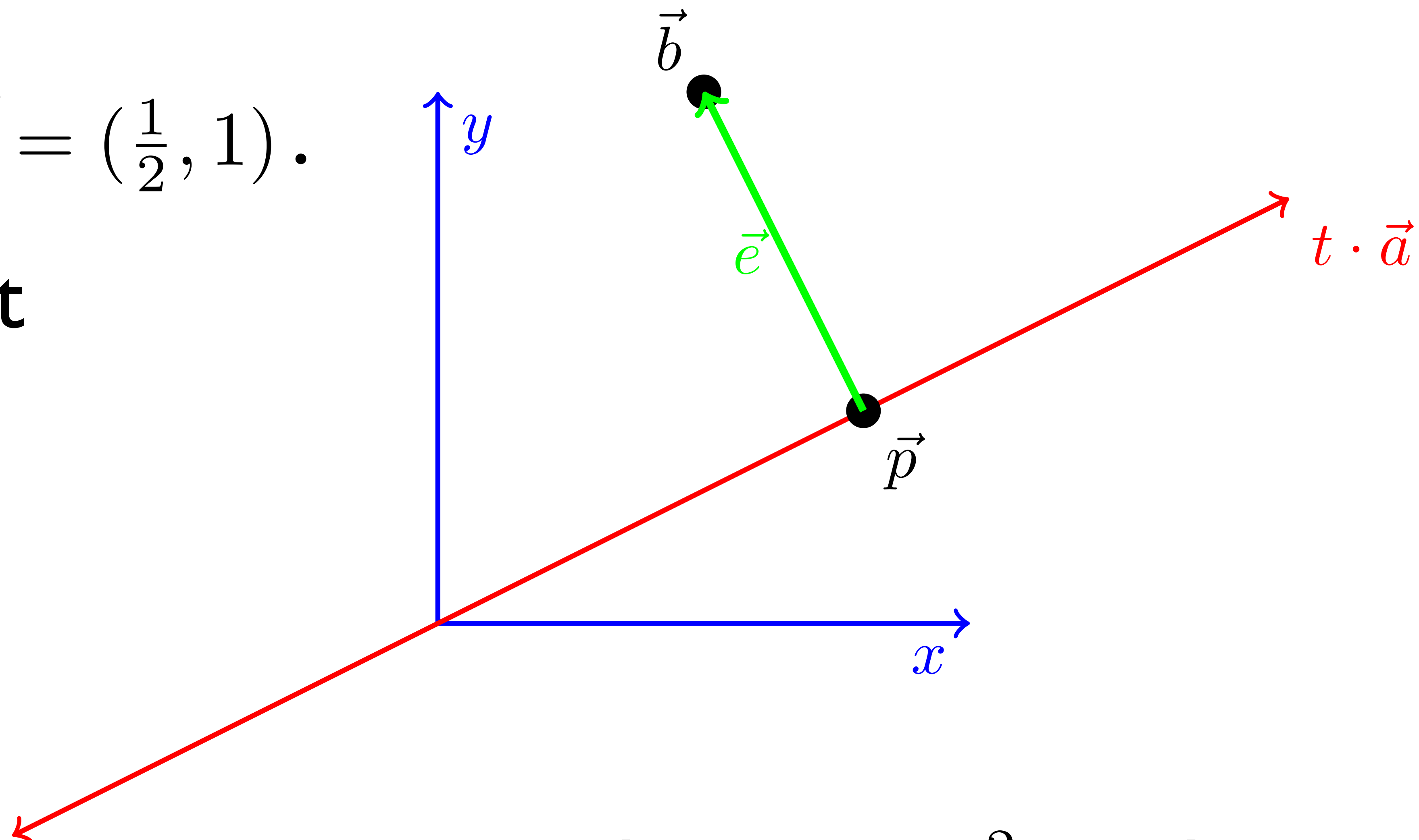
Claim: $\vec{p}$ is such that

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

is orthogonal to $\vec{a}$.

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

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



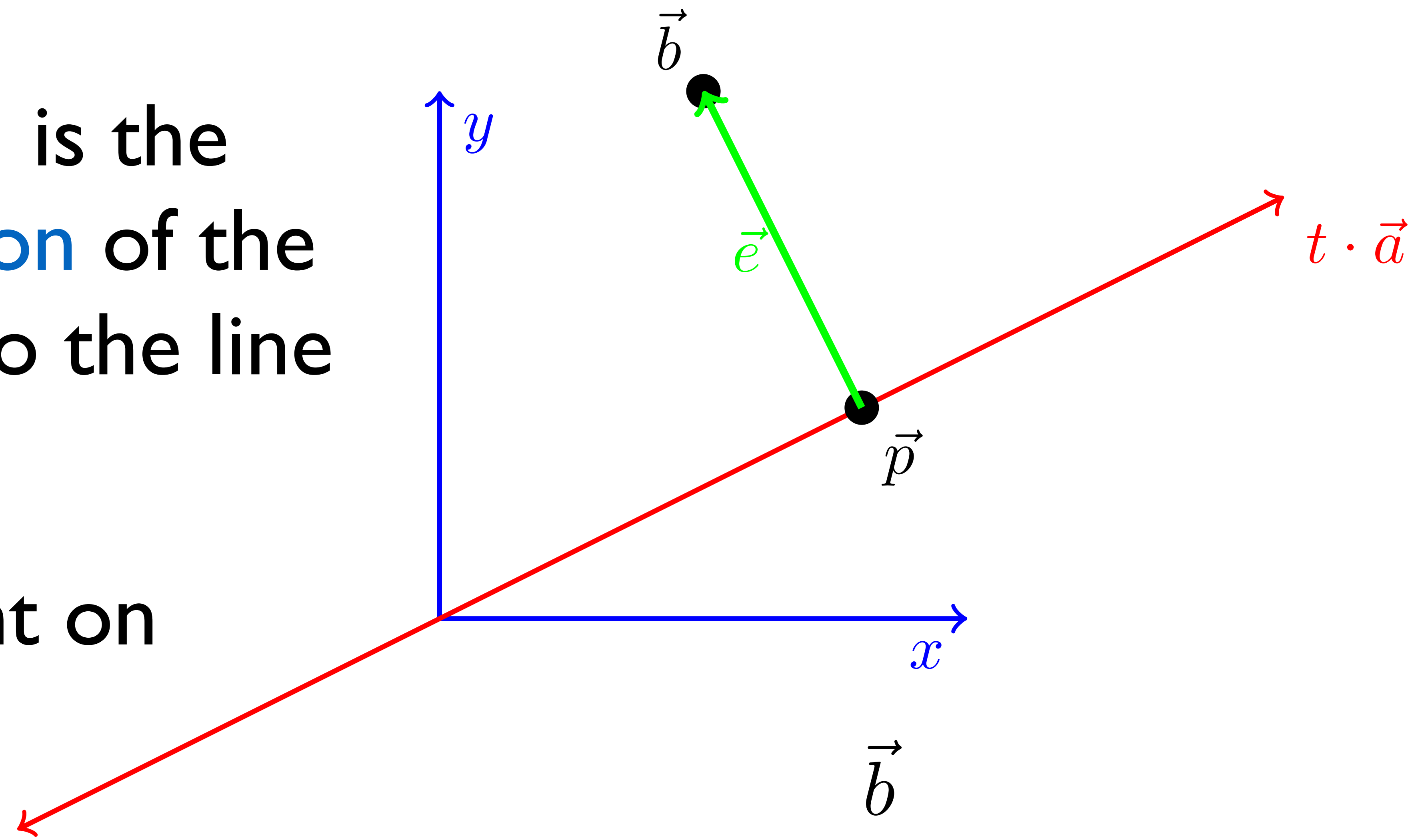
In our problem $\hat{x} = \frac{2}{5}$ and

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

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?

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

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

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

$$= P$$

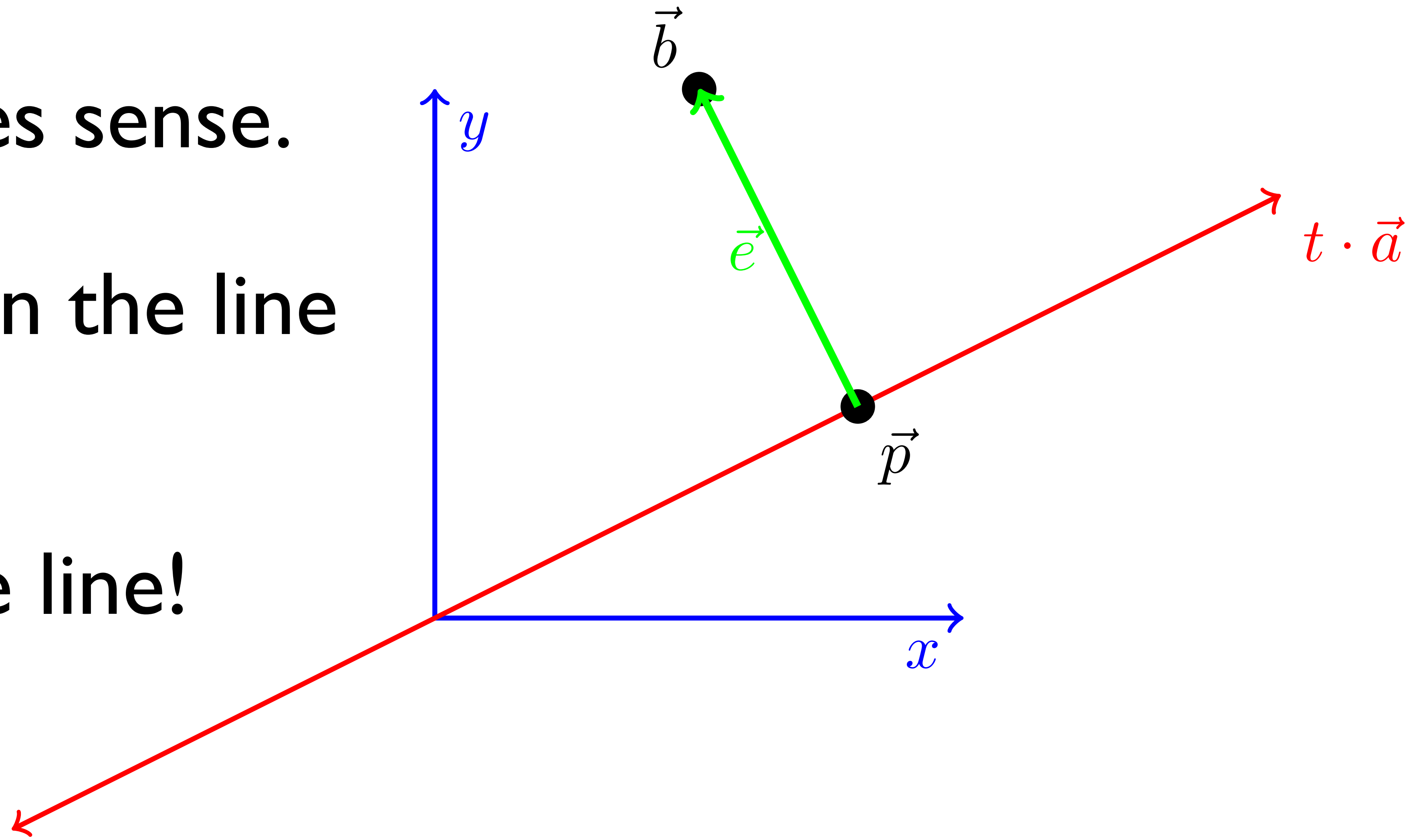
Projecting Again

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

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 $\displaystyle \underset{\vec{p} \in S}{\text{minimize}} \; \|\vec{b} - \vec{p}\|$

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?

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!

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 $\vec{p} = \vec{b}_s$.

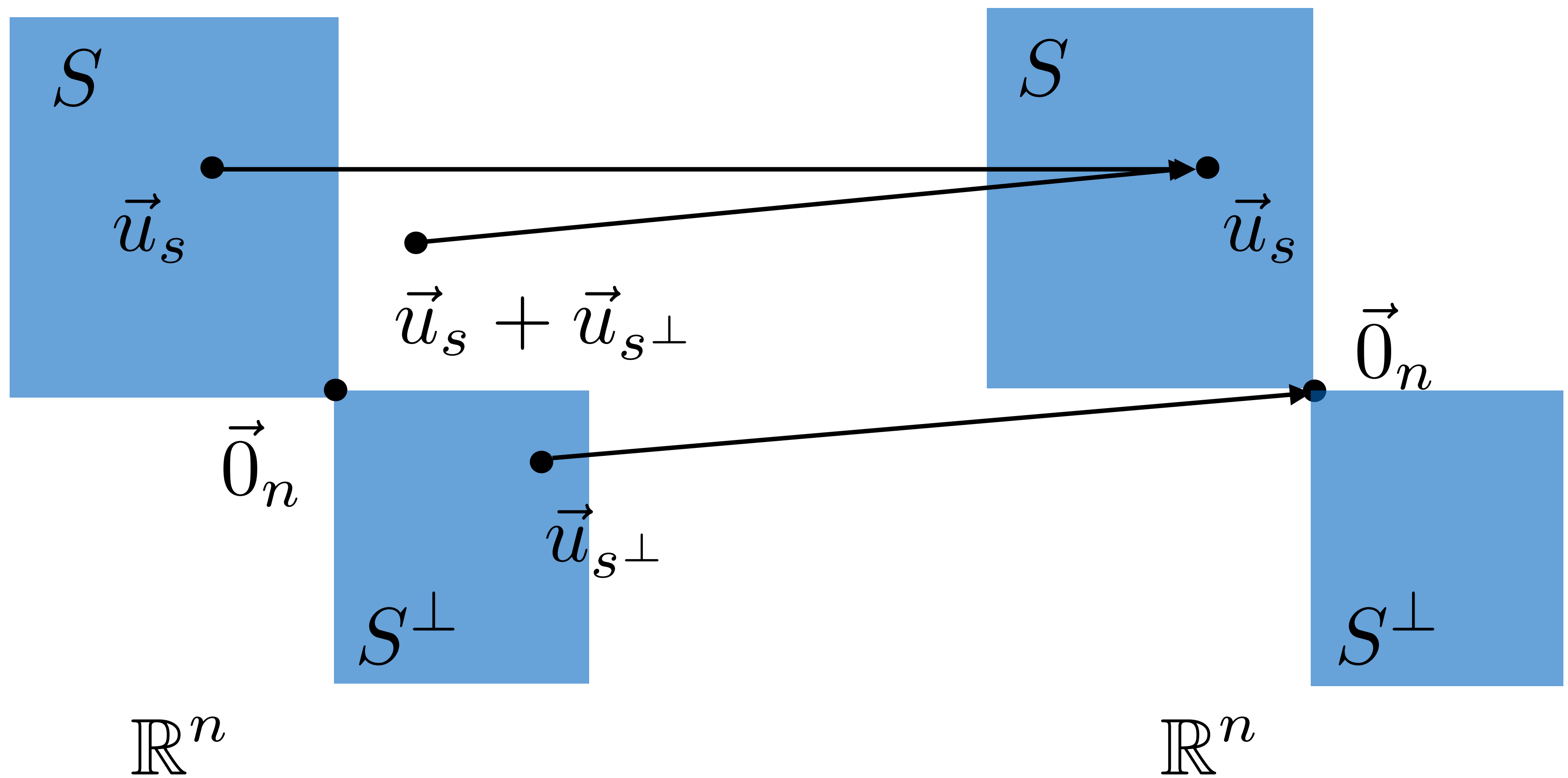
The projection $\vec{p}$ of $\vec{b}$ onto S is such that $\vec{b} - \vec{p}$ is orthogonal to S.

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

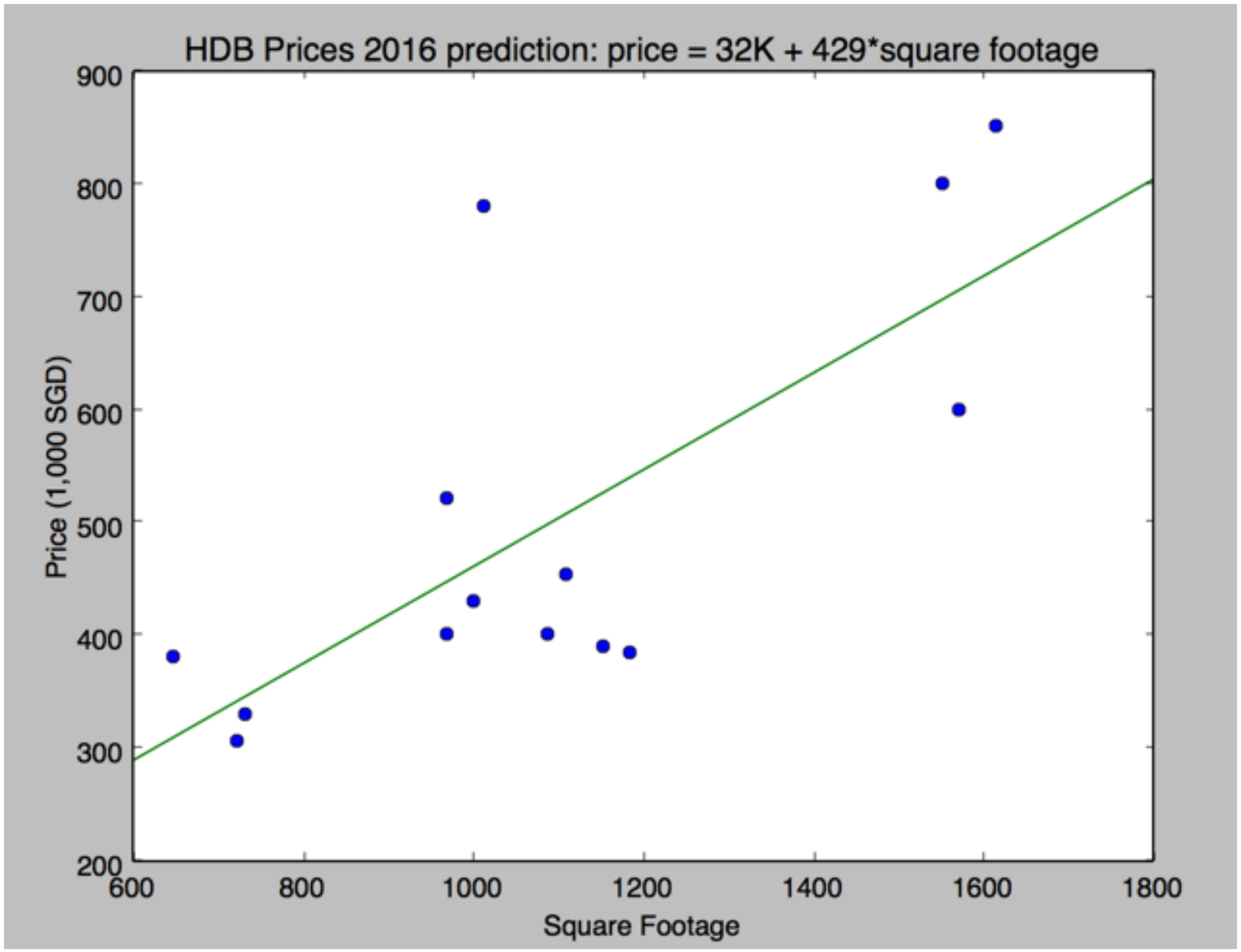
$$\|\vec{b} - A\hat{\mathbf{x}}\|$$

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.

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

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

But this equation has no solution!

constant term

square footage

housing price (1000 SGD)

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.

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.

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

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

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