Mathematics for Robotics (ROB-GY 6013 Section A)

- **Week 8:**
  - Gram-Schmidt Process
  - Projection Theorem

Norms and Inner Products

- Many norms (inspired by *length*)
- Many inner products (inspired by *dot product*)
- Connect these two concepts

Norms and Inner Products

- Many norms (inspired by *length*)
- Many inner products (inspired by *dot product*)
- Connect these two concepts
    - The square root of an inner product of a vector with itself ***induces*** a norm

$$\| x \| := \langle x, x \rangle^{1/2} = \sqrt{\langle x, x \rangle}$$

Norms and Inner Products

- Many norms (inspired by *length*)
- Many inner products (inspired by *dot product*)
- Connect these two concepts
  - The square root of an inner product of a vector with itself ***induces*** a norm

$$\| x \| := \langle x, x \rangle^{1/2} = \sqrt{\langle x, x \rangle}$$

  - Unless otherwise specified, we are using the above norm when we mention an **inner product space** $(\mathcal{X}, \mathcal{F}, \langle \cdot , \cdot \rangle)$

Gram-Schmidt: Building the Basis

- Review of n-dimensional vector spaces:
  - **Complete the basis:**
    - Given LI vectors $\{y^1, \ldots, y^k\}$
    - Find LI vectors $\{y^1, \ldots, y^k, y^{k+1}, \ldots, y^n\}$

Gram-Schmidt: Building the Basis

- Review of n-dimensional vector spaces:
  - **Complete the basis:**
    - Given LI vectors $\{y^1, \ldots, y^k\}$
    - Find LI vectors $\{y^1, \ldots, y^k, y^{k+1}, \ldots, y^n\}$
  - **Change the basis:**
    - Given basis $\{y^1, \ldots, y^n\}$
    - Find another basis $\{v^1, \ldots, v^n\}$

Gram-Schmidt: Building the Basis

- Review of n-dimensional vector spaces:
  - **Complete the basis:**
    - Given LI vectors $\{y^1, \dots, y^k\}$
    - Find LI vectors $\{y^1, \dots, y^k, y^{k+1}, \dots, y^n\}$
  - **Change the basis:**
    - Given basis $\{y^1, \dots, y^n\}$
    - Find another basis $\{v^1, \dots, v^n\}$
- **Build an orthonormal basis** (k not necessarily n)**:**
  - Given LI vectors $\{y^1, \dots, y^k\}$
  - Find LI vectors $\{v^1, \dots, v^k\} = \text{span}\{y^1, \dots, y^k\}$

Gram-Schmidt: Building the Basis

- Review of n-dimensional vector spaces:
  - **Complete the basis:**
    - Given LI vectors $\{y^1, \dots, y^k\}$
    - Find LI vectors $\{y^1, \dots, y^k, y^{k+1}, \dots, y^n\}$
  - **Change the basis:**
    - Given basis $\{y^1, \dots, y^n\}$ **Change of Basis Matrix**
    - Find another basis $\{v^1, \dots, v^n\}$
- **Build an orthonormal basis** (k not necessarily n)**:**
  - Given LI vectors $\{y^1, \dots, y^k\}$ **Gram-Schmidt Process**
  - Find LI vectors $\{v^1, \dots, v^k\} = \text{span}\{y^1, \dots, y^k\}$

Gram-Schmidt: Building the Basis

- Review of n-dimensional vector spaces:
  - **Complete the basis:**
    - Given LI vectors $\{y^1, \ldots, y^k\}$
    - Find LI vectors $\{y^1, \ldots, y^k, y^{k+1}, \ldots, y^n\}$
  - **Change the basis:**
    - Given basis $\{y^1, \ldots, y^n\}$
    - Find another basis $\{v^1, \ldots, v^n\}$
- **Build an orthonormal basis** (k not necessarily n):
  - Given LI vectors $\{y^1, \ldots, y^k\}$
  - Find LI vectors $\{v^1, \ldots, v^k\} = \text{span}\{y^1, \ldots, y^k\}$

Change of Basis Matrix

<u>Representation</u>

Gram-Schmidt Process

Gram-Schmidt: Building the Basis

- **Build an orthonormal basis** (k not necessarily n)**:**
  - Given LI vectors $\{y^1, \ldots, y^k\}$
  - Find LI vectors $\{v^1, \ldots, v^k\}$

Gram-Schmidt Process

- Why are we doing this?

Gram-Schmidt: Building the Basis

- **Build an orthonormal basis** (k not necessarily n)**:**
  - Given LI vectors $\{y^1, \ldots, y^k\}$
  - Find LI vectors $\{v^1, \ldots, v^k\}$

Gram-Schmidt Process

- Why are we doing this?
  - $(k < n)$ Is this useful?

Gram-Schmidt: Building the Basis

- **Build an orthonormal basis** (k not necessarily n):
  - Given LI vectors $\{y^1, \dots, y^k\}$
  - Find LI vectors $\{v^1, \dots, v^k\}$

Gram-Schmidt Process

- Why are we doing this?
  - $(k < n)$ Is this useful?
    - At a basic level, finding an orthonormal basis is convenient for any problem dealing with vector spaces (e.g., wave-functions in an infinite-dimensional space in quantum mechanics)

Gram-Schmidt: Building the Basis

- **Build an orthonormal basis** (k not necessarily n)**:**
  - Given LI vectors $\{y^1, \dots, y^k\}$
  - Find LI vectors $\{v^1, \dots, v^k\}$

Gram-Schmidt Process

- Why are we doing this?
  - $(k < n)$ Is this useful?
    - At a basic level, finding an orthonormal basis is convenient for any problem dealing with vector spaces (e.g., wave-functions in an infinite-dimensional space in quantum mechanics)
  - $(k = n)$ Can we just pick the natural basis?

Gram-Schmidt: Building the Basis

- **Build an orthonormal basis** (k not necessarily n)**:**
  - Given LI vectors $\{y^1, \dots, y^k\}$
  - Find LI vectors $\{v^1, \dots, v^k\}$

Gram-Schmidt Process

- Why are we doing this?
  - $(k < n)$ Is this useful?
    - At a basic level, finding an orthonormal basis is convenient for any problem dealing with vector spaces (e.g., wave-functions in an infinite-dimensional space in quantum mechanics)
  - $(k = n)$ Can we just pick the natural basis?
    - As a tool for the QR algorithm for finding eigenvalues and proving other theorems

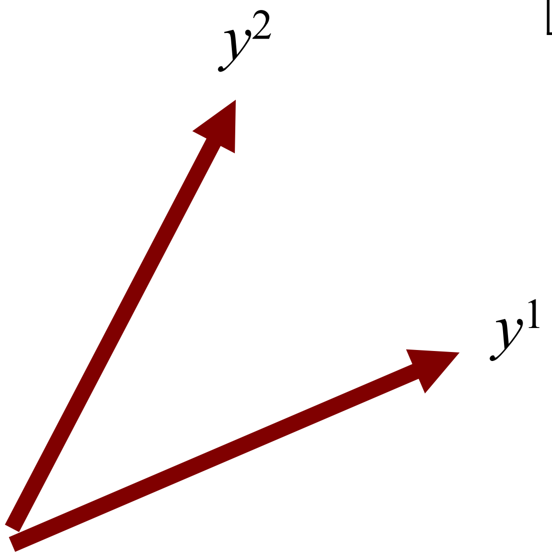
Orthonormal basis

- Why look for an **orthonormal basis**?
  - Because the orthonormal basis is the “best” basis
    - Basis vectors are orthogonal and unit length (norm of one) (the natural basis is an orthonormal basis)
  - Orthonormal bases are not unique.
  - Gram-Schmidt Process is not the only method to find them.
    - (Householder’s, etc.)

Inspiration

- Given two vectors that are not perpendicular to each other:

y^1	y^2

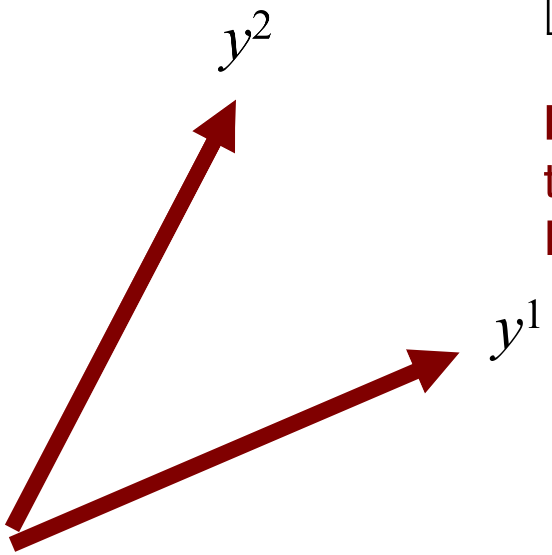


Inspiration

- Initialize

$v^1 = y^1$	y^2

Note: Think of the entries of the table as lines of code that update the vectors v^1 and v^2 rather than equations. In this step, here we are initializing v^1 to be y^1.



Inspiration

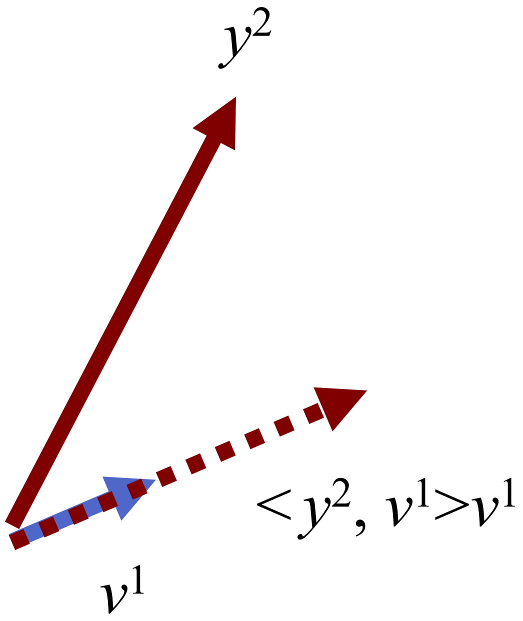
- Normalize

$v^1 = y^1$	y^2
$v^1 = v^1/\| v^1 \|$	

Inspiration

- Project

$v^1 = y^1$	y^2
$v^1 = v^1/\|\| v^1 \|\|$	

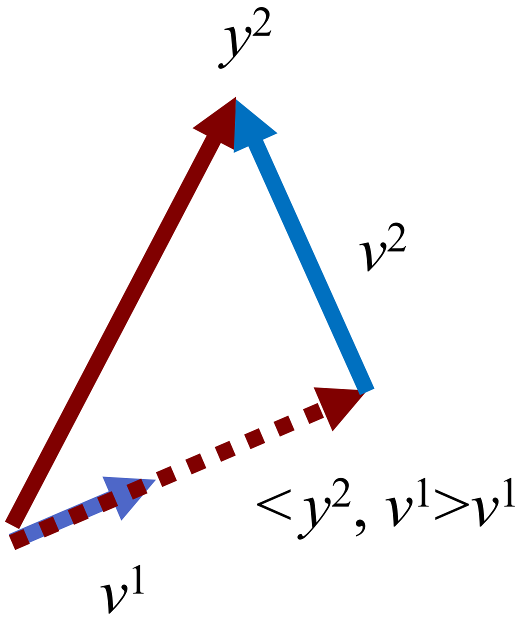


Projection of y^2 on v^1

Inspiration

- Orthogonalize (subtract out the projection)

$v^1 = y^1$	y^2
$v^1 = v^1/\|\| v^1 \|\|$	$v^2 = y^2 - \langle y^2, v^1\rangle v^1$

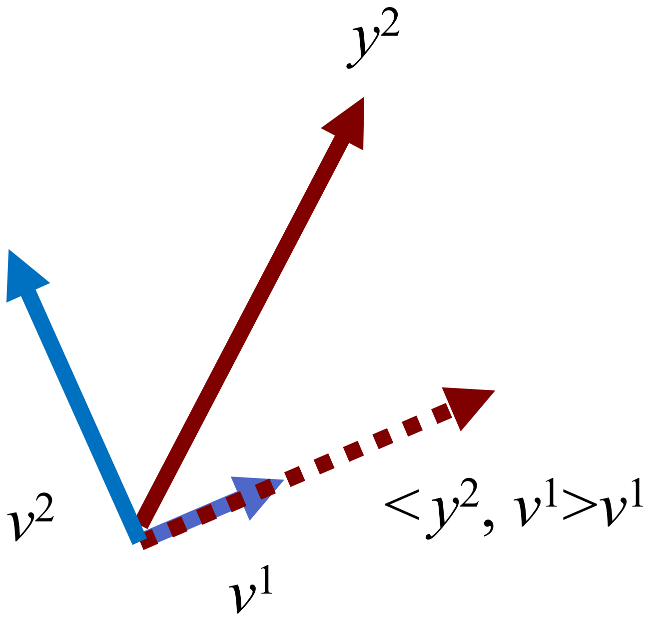


Projection of y^2 on v^1

Inspiration

- Orthogonalize (subtract out the projection)

$v^1 = y^1$	y^2
$v^1 = v^1/\|\| v^1 \|\|$	$v^2 = y^2 - <y^2, v^1>v^1$

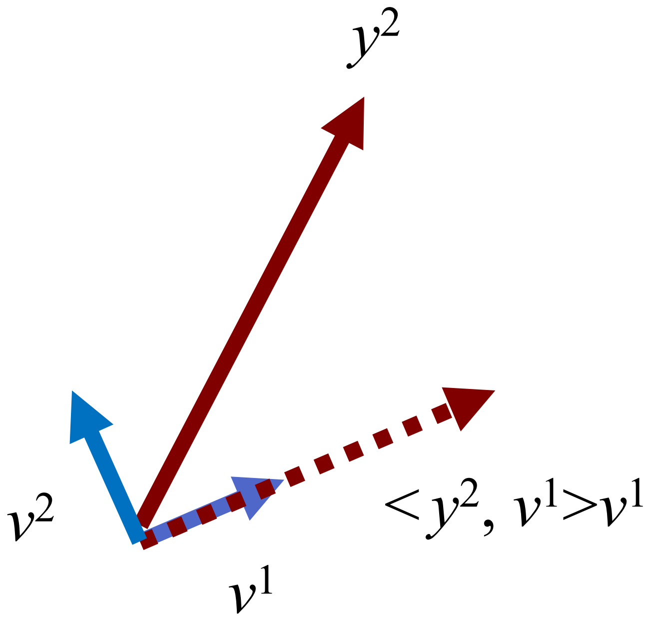


Projection of y^2 on v^1

Inspiration

- Normalize

$v^1 = y^1$	y^2
$v^1 = v^1/\| v^1 \|$	$v^2 = y^2 - \langle y^2, v^1 \rangle v^1$
	$v^2 = v^2/\| v^2 \|$

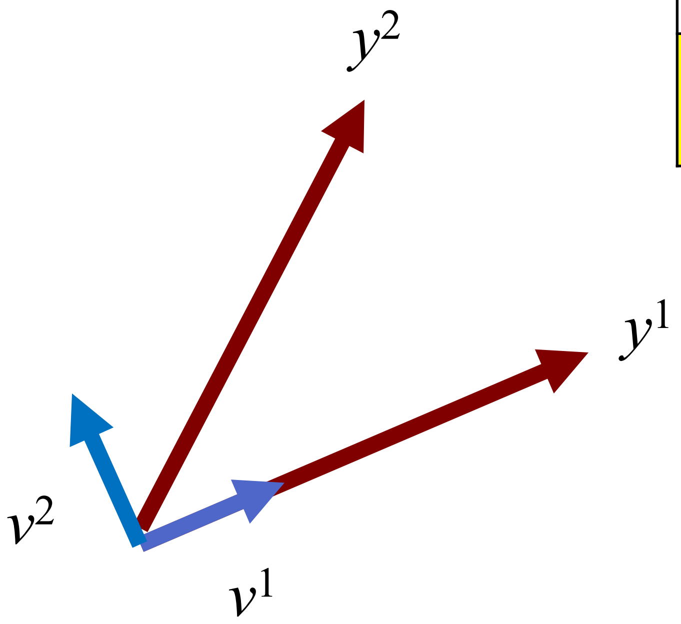


Projection of y^2 on v^1

Inspiration

- Compare old linearly independent basis y with new orthonormal basis v

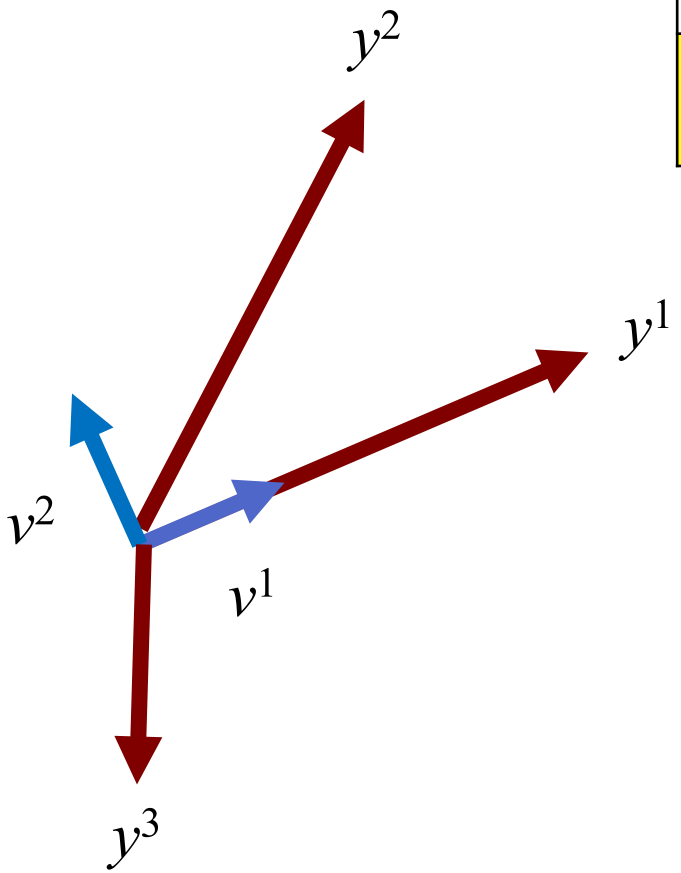
$v^1 = y^1$	y^2
$v^1 = v^1/\| v^1 \|$	$v^2 = y^2 - \langle y^2, v^1\rangle v^1$
	$v^2 = v^2/\| v^2 \|$



Classical Gram Schmidt Process

- Add a third vector

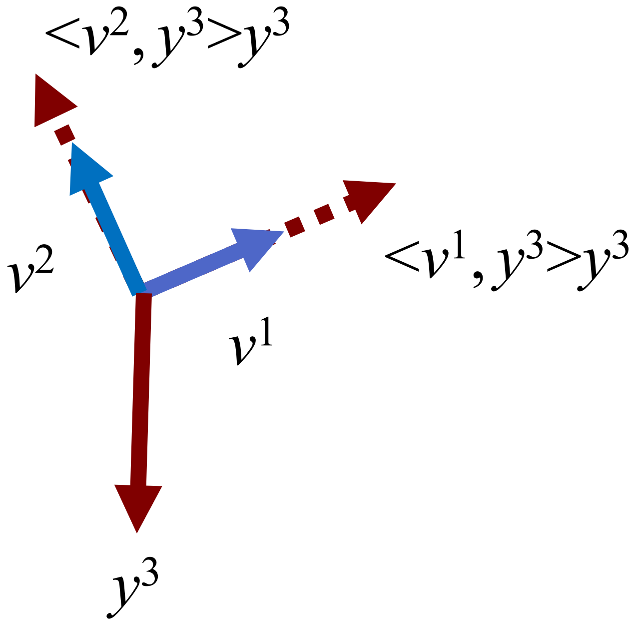
$v^1 = y^1$	y^2	y^3
$v^1 = v^1/\|\| v^1 \|\|$	$v^2 = y^2 - \langle y^2, v^1\rangle v^1$	
	$v^2 = v^2/\|\| v^2 \|\|$	



Classical Gram Schmidt Process

- Orthogonalize y^3 (subtract out ***two*** projections)

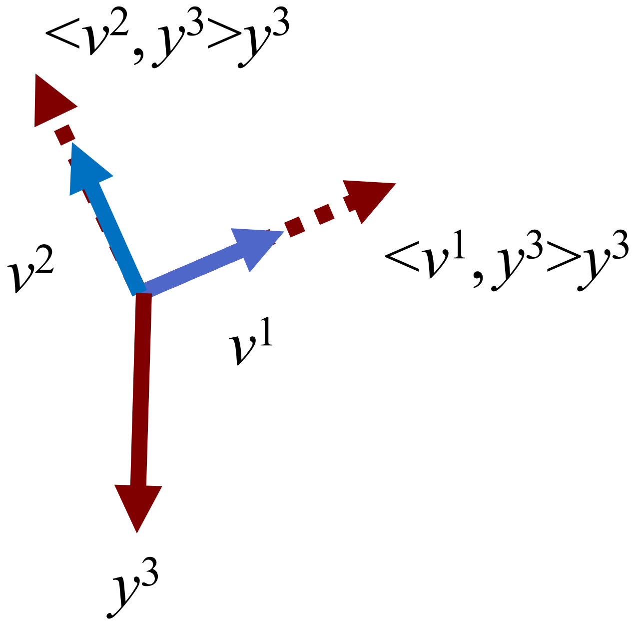
$v^1 = y^1$	y^2	y^3
$v^1 = v^1/\|\| v^1 \|\|$	$v^2 = y^2 - <y^2, v^1>v^1$	
	$v^2 = v^2/\|\| v^2 \|\|$	



Classical Gram Schmidt Process

- Subtract out projection onto v^1

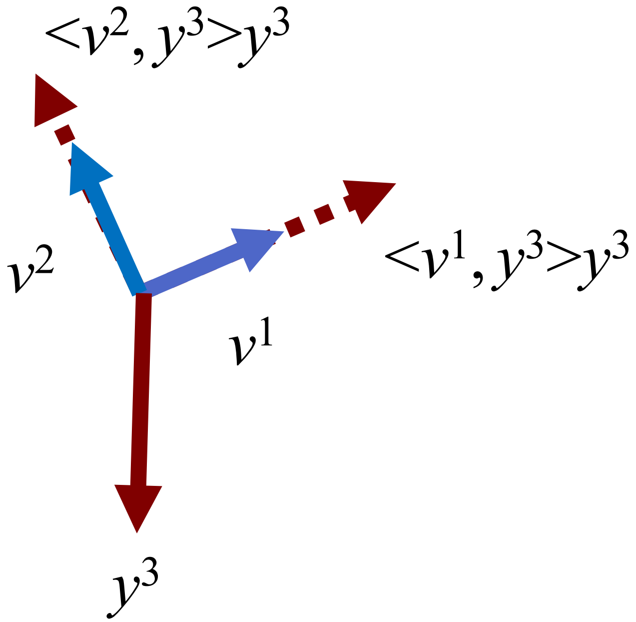
$v^1 = y^1$	y^2	y^3
$v^1 = v^1/\|\| v^1 \|\|$	$v^2 = y^2 - <y^2, v^1>v^1$	$v^3 = y^3 - <y^3, v^1>v^1$
	$v^2 = v^2/\|\| v^2 \|\|$	



Classical Gram Schmidt Process

- Subtract out projection onto v^2

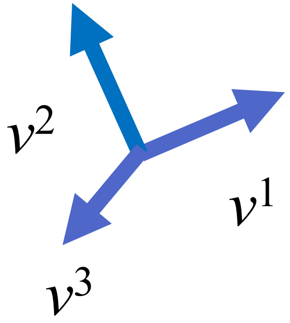
$v^1 = y^1$	y^2	y^3
$v^1 = v^1/\|\| v^1 \|\|$	$v^2 = y^2 - \langle y^2, v^1\rangle v^1$	$v^3 = y^3 - \langle y^3, v^1\rangle v^1$
	$v^2 = v^2/\|\| v^2 \|\|$	$v^3 = v^3 - \langle y^3, v^2\rangle v^2$



Classical Gram Schmidt Process

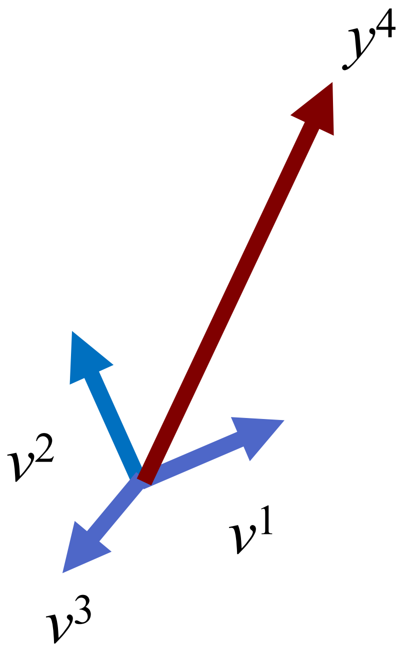
- Normalize

$v^1 = y^1$	y^2	y^3
$v^1 = v^1/\|\| v^1 \|\|$	$v^2 = y^2 - \langle y^2, v^1 \rangle v^1$	$v^3 = y^3 - \langle y^3, v^1 \rangle v^1$
	$v^2 = v^2/\|\| v^2 \|\|$	$v^3 = v^3 - \langle y^3, v^2 \rangle v^2$
		$v^3 = v^3/\|\| v^3 \|\|$



Classical Gram Schmidt Process

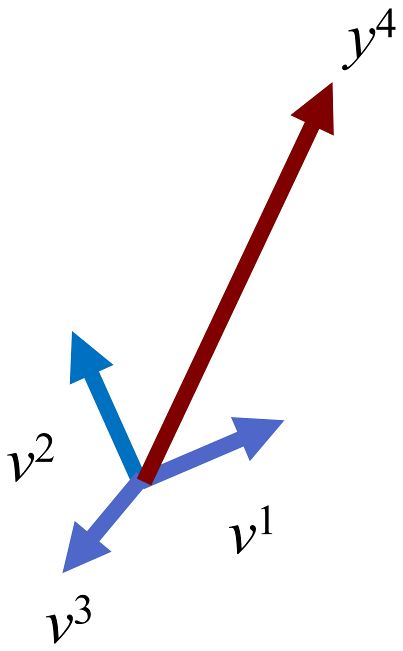
- Add a fourth vector: Try it yourself



$v^1 = y^1$	y^2	y^3	y^4
$v^1 = v^1/\|\| v^1 \|\|$	$v^2 = y^2 - <y^2, v^1>v^1$	$v^3 = y^3 - <y^3, v^1>v^1$	??
	$v^2 = v^2/\|\| v^2 \|\|$	$v^3 = v^3 - <y^3, v^2>v^2$	??
		$v^3 = v^3/\|\| v^3 \|\|$	??
			$v^4 = v^4/\|\| v^4 \|\|$

Classical Gram Schmidt Process

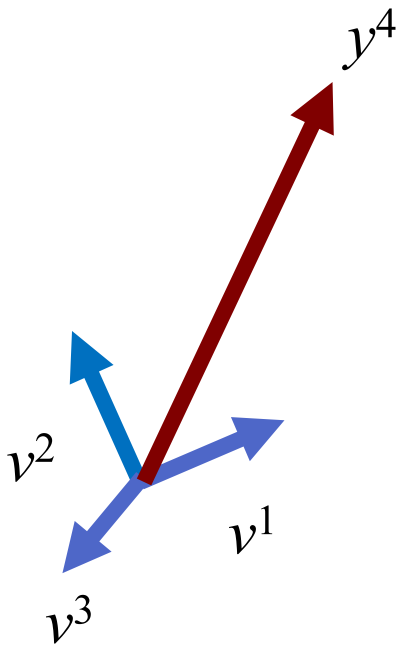
- What if one of the four vectors was linearly dependent?



$v^1 = y^1$	y^2	y^3	y^4
$v^1 = v^1/\| v^1 \|$	$v^2 = y^2 - \langle y^2, v^1\rangle v^1$	$v^3 = y^3 - \langle y^3, v^1\rangle v^1$	??
	$v^2 = v^2/\| v^2 \|$	$v^3 = v^3 - \langle y^3, v^2\rangle v^2$	??
		$v^3 = v^3/\| v^3 \|$	??
			$v^4 = v^4/\| v^4 \|$

Classical Gram Schmidt Process

- What if one of the four vectors was linearly dependent?



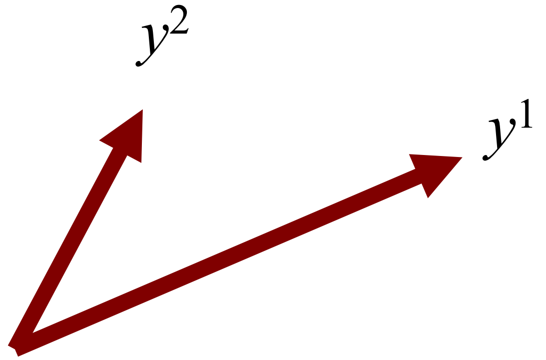
$v^1 = y^1$	y^2	y^3	y^4
$v^1 = v^1/\|\| v^1 \|\|$	$v^2 = y^2 - \langle y^2, v^1 \rangle v^1$	$v^3 = y^3 - \langle y^3, v^1 \rangle v^1$	??
	$v^2 = v^2/\|\| v^2 \|\|$	$v^3 = v^3 - \langle y^3, v^2 \rangle v^2$	??
		$v^3 = v^3/\|\| v^3 \|\|$	??
			$v^4 = v^4/\|\| v^4 \|\|$

If were y^3 were linearly independent, the orthogonalization process would result in $v^3 = 0$

Classical Gram Schmidt Process (Variation)

- We do not need to normalize at each step. Can do it all in one step at the end.

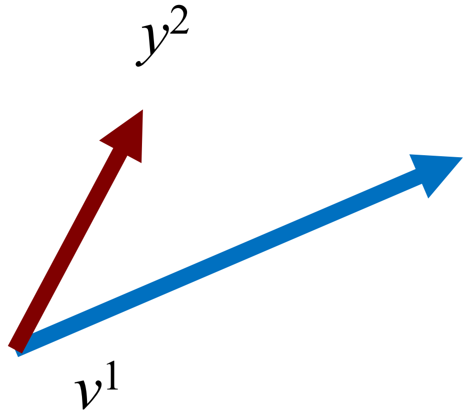
y^1	y^2



Classical Gram Schmidt Process (Variation)

- Initialize, but do not normalize yet.

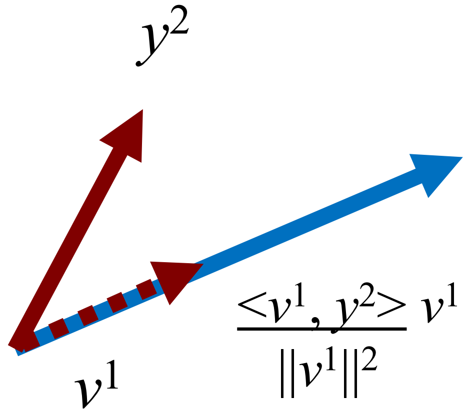
$v^1 = y^1$	y^2



Classical Gram Schmidt Process (Variation)

- In general, when projecting on v^i scale by $||v^i||^2$. (Skipped before when $||v^i||^2 = 1$)

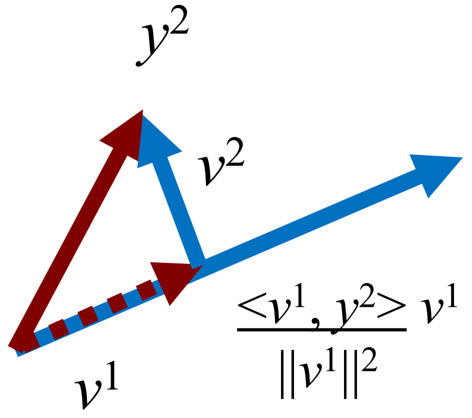
$v^1 = y^1$	y^2



Classical Gram Schmidt Process (Variation)

- In general, when projecting on v^i scale by $||v^i||^2$. (Skipped before when $||v^i||^2 = 1$)

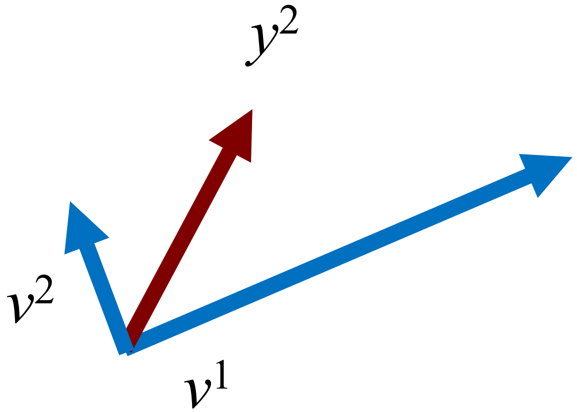
$v^1 = y^1$	y^2
	$v^2 = y^2 - \dfrac{<v^1, y^2>}{\|\|v^1\|\|^2} v^1$



Classical Gram Schmidt Process (Variation)

- In general, when projecting on v^i scale by $||v^i||^2$. (Skipped before when $||v^i||^2 = 1$)

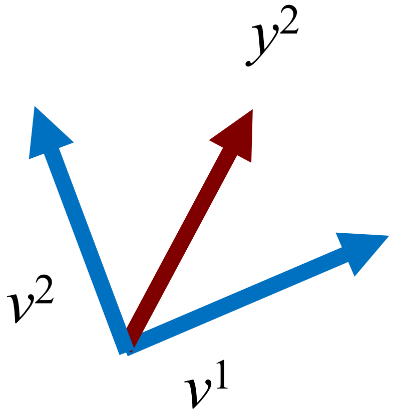
$v^1 = y^1$	y^2
	$v^2 = y^2 - \frac{<v^1, y^2>}{\|\|v^1\|\|^2} v^1$



Classical Gram Schmidt Process (Variation)

- Normalize at the end

$v^1 = y^1$	y^2
	$v^2 = y^2 - \dfrac{\langle v^1, y^2 \rangle}{\|v^1\|^2} v^1$
$v^1 = v^1 / \| v^1 \|$	$v^2 = v^2 / \| v^2 \|$



Classical Gram Schmidt Process (Variation)

- For 4 vectors:

$v^1 = y^1$	y^2	y^3	y^4
	$v^2 = y^2 - \frac{<v^1, y^2>}{\lVert v^1 \rVert^2} v^1$	$v^3 = y^3 - \frac{<v^1, y^3>}{\lVert v^1 \rVert^2} v^1$	$v^4 = y^4 - \frac{<v^1, y^4>}{\lVert v^1 \rVert^2} v^1$
		$v^3 = y^3 - \frac{<v^2, y^3>}{\lVert v^2 \rVert^2} v^2$	$v^4 = y^3 - \frac{<v^2, y^3>}{\lVert v^2 \rVert^2} v^2$
			$v^4 = y^4 - \frac{<v^3, y^3>}{\lVert v^3 \rVert^2} v^3$
$v^1 = v^1/\lVert v^1 \rVert$	$v^2 = v^2/\lVert v^2 \rVert$	$v^3 = v^3/\lVert v^3 \rVert$	$v^4 = v^4/\lVert v^4 \rVert$

Classical Gram Schmidt Process (Variation)

- Summarize table entries into compact formula for **orthogonalization**

$v^1 = y^1$	y^2	y^3	y^4

$$v_k = y_k - \sum_{i=1}^{k-1} \frac{\left\langle y^j, v^j \right\rangle}{\| v^j \|^2} \cdot v^j$$

$v^1 = v^1 / \| v^1 \|$	$v^2 = v^2 / \| v^2 \|$	$v^3 = v^3 / \| v^3 \|$	$v^4 = v^4 / \| v^4 \|$

Proposition 3.9: Recursion Gram Schmidt Process

- Let $(\mathcal{X}, \mathbb{R}, \langle \cdot , \cdot \rangle)$ be an **inner product space**, $\{y^1, \dots, y^k\}$ a **linearly independent set** and $\{v^1, \dots, v^{k-1}\}$ an **orthogonal set** satisfying

$$\operatorname{span}\{v^1, \dots, v^{k-1}\} = \operatorname{span}\{y^1, \dots, y^{k-1}\}$$

- Define

$$v_k = y_k - \sum_{i=1}^{k-1} \frac{\left\langle y^j, v^j \right\rangle}{\| v^j \|^2} \cdot v^j$$

- where $\|v^j\|^2 = \langle v^j, v^j \rangle$. Then $\{v^1, \dots, v^k\}$ is **orthogonal** and

$$\operatorname{span}\{v^1, \dots, v^k\} = \operatorname{span}\{y^1, \dots, y^k\}$$

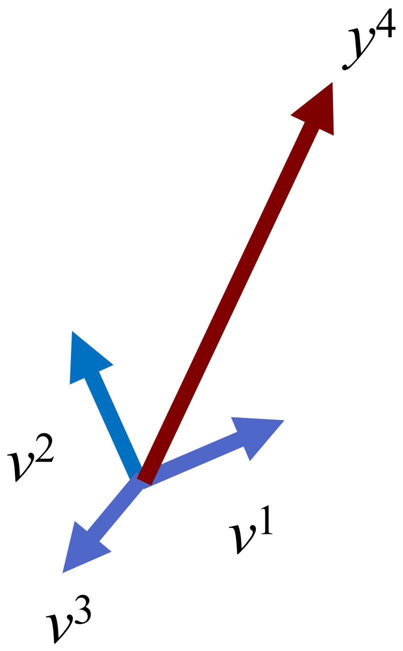
- The formula describes the steps for **orthogonalization**
- This is a theorem that tells you how to “grow” an orthogonal set out of a linearly independent set of vectors and that it is always possible to do so

Gram Schmidt Process

- Proof is straightforward, the orthogonality of $\{v^1, \ldots, v^k\}$ is by construction.
  - See main textbook
  - Also see pages 1-8 of "Gram_Schmidt Handout.pdf" from Content > Weekly Lectures

Revisit Classical Gram Schmidt Process

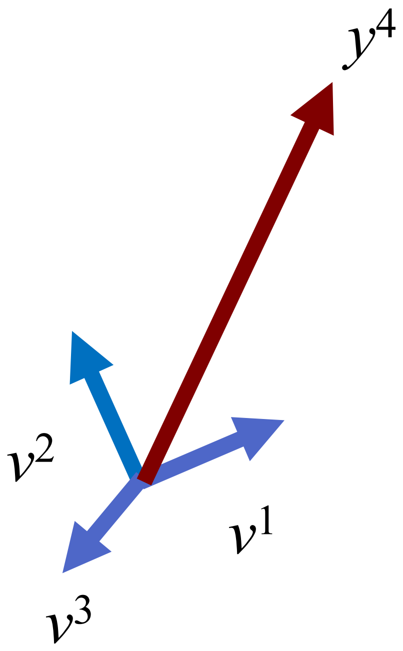
- Due to rounding, every layer of computation accumulates error

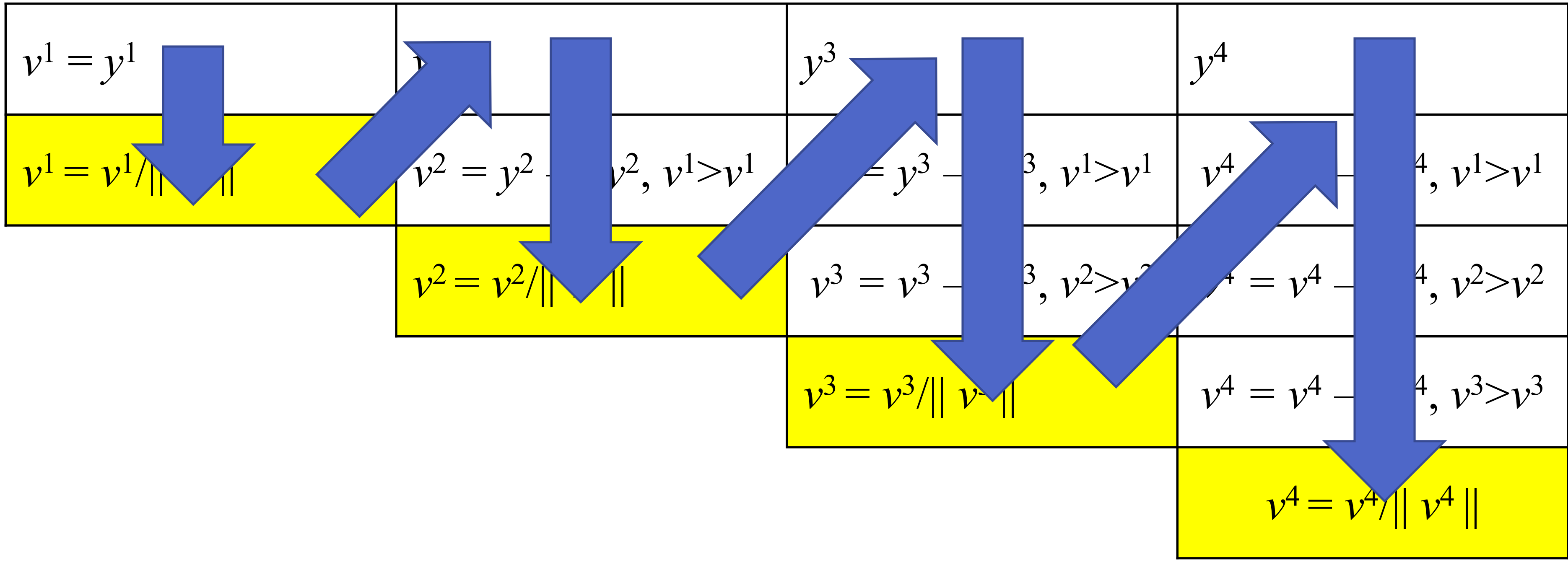


$v^1 = y^1$	y^2	y^3	y^4
$v^1 = v^1/\Vert v^1 \Vert$	$v^2 = y^2 - <y^2, v^1>v^1$	$v^3 = y^3 - <y^3, v^1>v^1$	$v^4 = y^4 - <y^4, v^1>v^1$
	$v^2 = v^2/\Vert v^2 \Vert$	$v^3 = v^3 - <y^3, v^2>v^2$	$v^4 = v^4 - <y^4, v^2>v^2$
		$v^3 = v^3/\Vert v^3 \Vert$	$v^4 = v^4 - <y^4, v^3>v^3$
			$v^4 = v^4/\Vert v^4 \Vert$

Revisit Classical Gram Schmidt Process

- Due to rounding, every layer of computation accumulates error



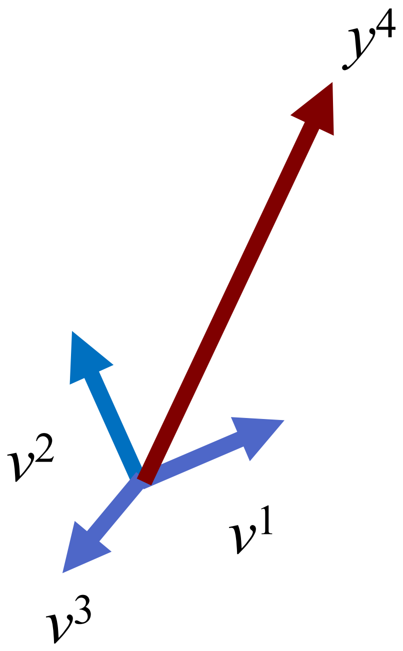


With many layers of computation here, the resulting vectors may longer be orthogonal due to accumulated error

Modified Gram-Schmidt

- Normalize

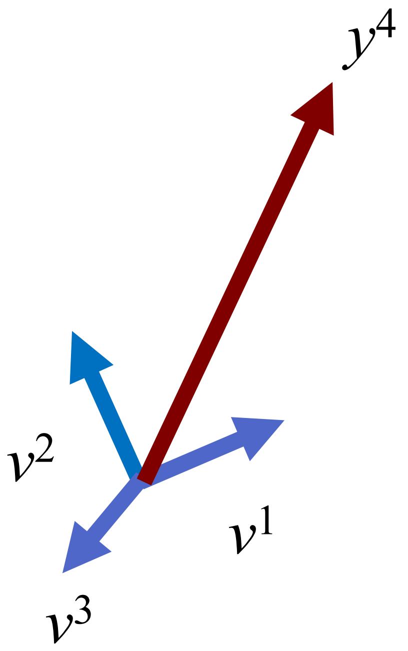
$v^1 = y^1$	y^2	y^3	y^4
$v^1 = v^1 / \| v^1 \|$			



Modified Gram-Schmidt

- Orthogonalize y^2, y^3, and y^4 in the same step (with respect to v^1)

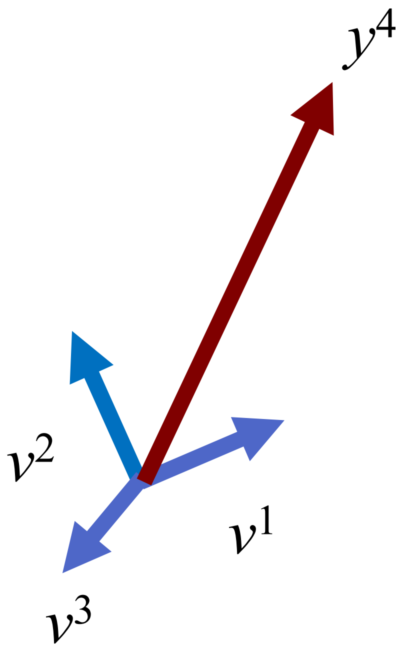
$v^1 = y^1$	y^2	y^3	y^4
$v^1 = v^1/\|\| v^1 \|\|$	$v^2 = y^2 - \langle y^2, v^1 \rangle v^1$	$v^3 = y^3 - \langle y^3, v^1 \rangle v^1$	$v^4 = y^4 - \langle y^4, v^1 \rangle v^1$



Modified Gram-Schmidt

- Normalize

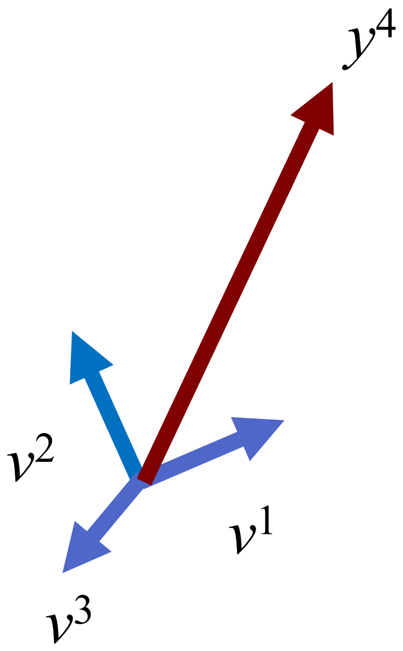
$v^1 = y^1$	y^2	y^3	y^4
$v^1 = v^1/\| v^1 \|$	$v^2 = y^2 - \langle y^2, v^1 \rangle v^1$	$v^3 = y^3 - \langle y^3, v^1 \rangle v^1$	$v^4 = y^4 - \langle y^4, v^1 \rangle v^1$
	$v^2 = v^2/\| v^2 \|$		



Modified Gram-Schmidt

- Orthogonalize v^3 and v^4 with respect to v^2. Note v here instead of y.

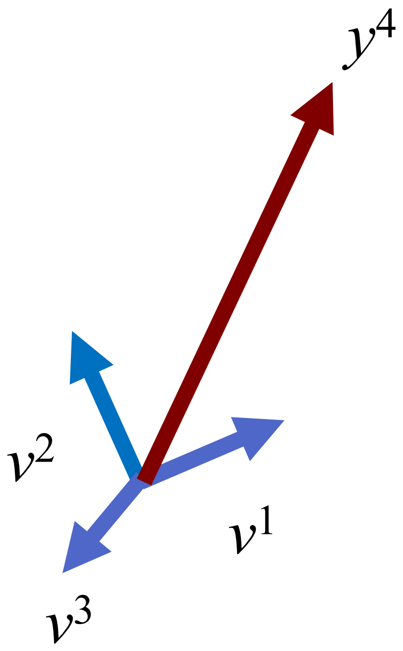
$v^1 = y^1$	y^2	y^3	y^4
$v^1 = v^1/\|\| v^1 \|\|$	$v^2 = y^2 - \langle y^2, v^1\rangle v^1$	$v^3 = y^3 - \langle y^3, v^1\rangle v^1$	$v^4 = y^4 - \langle y^4, v^1\rangle v^1$
	$v^2 = v^2/\|\| v^2 \|\|$	$v^3 = v^3 - \langle v^3, v^2\rangle v^2$	$v^4 = v^4 - \langle v^4, v^2\rangle v^2$



With Classical Gram Schmidt, errors accumulate without any mechanism to correct them. In Modified Gram-Schmidt, the method offers some self-correction. See textbook page 80.

Modified Gram-Schmidt

- Normalize

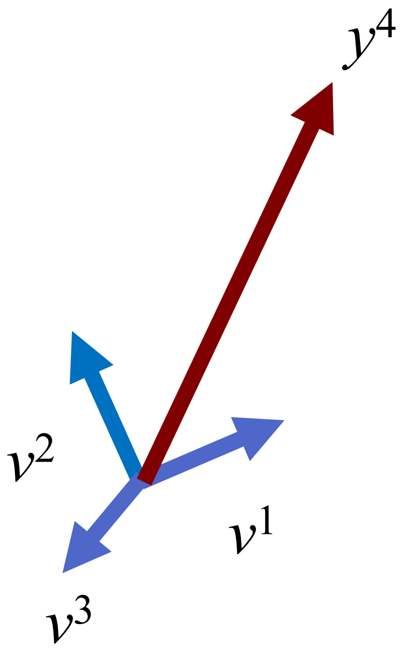


$v^1 = y^1$	y^2	y^3	y^4
$v^1 = v^1/\| v^1 \|$	$v^2 = y^2 - \langle y^2, v^1\rangle v^1$	$v^3 = y^3 - \langle y^3, v^1\rangle v^1$	$v^4 = y^4 - \langle y^4, v^1\rangle v^1$
	$v^2 = v^2/\| v^2 \|$	$v^3 = v^3 - \langle v^3, v^2\rangle v^2$	$v^4 = v^4 - \langle v^4, v^2\rangle v^2$
		$v^3 = v^3/\| v^3 \|$	

Modified Gram-Schmidt

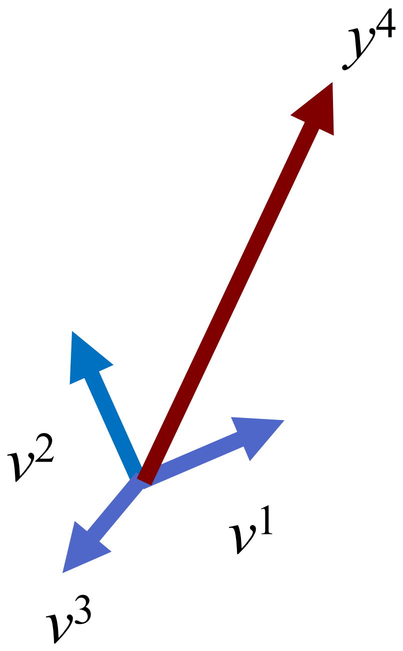
- Orthogonalize v^4 with respect to v^3:

$v^1 = y^1$	y^2	y^3	y^4
$v^1 = v^1/\| v^1 \|$	$v^2 = y^2 - \langle y^2, v^1 \rangle v^1$	$v^3 = y^3 - \langle y^3, v^1 \rangle v^1$	$v^4 = y^4 - \langle y^4, v^1 \rangle v^1$
	$v^2 = v^2/\| v^2 \|$	$v^3 = v^3 - \langle v^3, v^2 \rangle v^2$	$v^4 = v^4 - \langle v^4, v^2 \rangle v^2$
		$v^3 = v^3/\| v^3 \|$	$v^4 = v^4 - \langle v^4, v^3 \rangle v^3$



Modified Gram-Schmidt

- Normalize



$v^1 = y^1$	y^2	y^3	y^4
$v^1 = v^1 / \lVert v^1 \rVert$	$v^2 = y^2 - \langle y^2, v^1 \rangle v^1$	$v^3 = y^3 - \langle y^3, v^1 \rangle v^1$	$v^4 = y^4 - \langle y^4, v^1 \rangle v^1$
	$v^2 = v^2 / \lVert v^2 \rVert$	$v^3 = v^3 - \langle v^3, v^2 \rangle v^2$	$v^4 = v^4 - \langle v^4, v^2 \rangle v^2$
		$v^3 = v^3 / \lVert v^3 \rVert$	$v^4 = v^4 - \langle v^4, v^3 \rangle v^3$
			$v^4 = v^4 / \lVert v^4 \rVert$

Another view of Gram Schmidt

- **Change of basis:**
  - New orthonormal basis vectors v are linear combinations of the original linearly independent basis vectors y

Another view of Gram Schmidt

- **Change of basis:**
  - New orthonormal basis vectors v are linear combinations of the original linearly independent basis vectors y

$$v^1 = \begin{bmatrix} y^1 & y^2 & y^3 & y^4 \end{bmatrix} \begin{bmatrix} 1/\|y^1\| \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Another view of Gram Schmidt

- **Change of basis:**
  - New orthonormal basis vectors v are linear combinations of the original linearly independent basis vectors y

$$\begin{bmatrix} v^1 & v^2 & v^3 & v^4 \end{bmatrix} = \begin{bmatrix} y^1 & y^2 & y^3 & y^4 \end{bmatrix} \begin{bmatrix} 1/\|y^1\| & ? & ? & ? \\ 0 & ? & ? & ? \\ 0 & 0 & ? & ? \\ 0 & 0 & 0 & ? \end{bmatrix}$$

Another view of Gram Schmidt

- **Change of basis:**
  - New orthonormal basis vectors v are linear combinations of the original linearly independent basis vectors y

$$\begin{bmatrix} v^1 & v^2 & v^3 & v^4 \end{bmatrix} = \begin{bmatrix} y^1 & y^2 & y^3 & y^4 \end{bmatrix} \begin{bmatrix} & & & \\ & M & & \\ & & & \\ & & & \end{bmatrix}$$

Upper (or Right) Triangular Matrix

Another view of Gram Schmidt

- **Change of basis:**
  - New orthonormal basis vectors v are linear combinations of the original linearly independent basis vectors y

$$\begin{bmatrix} v^1 & v^2 & v^3 & v^4 \end{bmatrix} \begin{bmatrix} \\ \\ & & M^{-1} & & \\ \\ \\ \end{bmatrix} = \begin{bmatrix} y^1 & y^2 & y^3 & y^4 \end{bmatrix}$$

Upper (or Right) Triangular Matrix

Another view of Gram Schmidt

- **Change of basis:**
  - New orthonormal basis vectors v are linear combinations of the original linearly independent basis vectors y

$$\begin{bmatrix} & & \\ & Q & \\ & & \end{bmatrix}\begin{bmatrix} & & \\ & R & \\ & & \end{bmatrix} = \begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix}$$

Orthogonal Matrix **Upper (or Right) Triangular Matrix**

Applications: Gram Schmidt

- QR factorization (also known as QR decomposition)
  - “Gram Schmidt with Book-keeping”
  - Writing down Gram Schmidt naturally decomposes an invertible matrix (stack of linearly independent column vectors) into the product of an orthogonal matrix (stack of orthonormal vectors) and some upper triangular matrix.
- QR factorization plays a role in the QR algorithm to find eigenvalues (described briefly in the next few slides for your information)

QR Algorithm

- **Factor an invertible matrix A (LI columns):** $A = QR$
  - Matrix multiply $R \times Q$
  - Factor $(R \times Q)$ into new Q and R with QR factorization

QR Algorithm

- **Factor an invertible matrix A (LI columns):** $A = QR$
  - Matrix multiply $R \times Q$
  - Factor $(R \times Q)$ into new Q and R with QR factorization
- **Loop until $R \times Q$ converges to an upper triangular matrix**
- **Diagonal entries of an upper triangular matrix are the eigenvalues**

- Flipping the order of multiplication $(Q \times R)$ vs. $(R \times Q)$ changes the resulting matrix but not its eigenvalues
  - Proof: Show that QR and RQ are similar

$$(QR) = QR(QQ^{-1}) = Q(RQ)Q^{-1}$$

What else can we harvest from Gram Schmidt

- See Handout

Claim: Let $\{v_1, \cdots, v_n\}$ be an **orthonormal basis** for a inner product space X. Then the representation of $x \in X$ with respect to $\{v_1, \cdots, v_n\}$ is

$$\boxed{x = \sum_{i=1}^{n} < x, v_i > v_i}$$

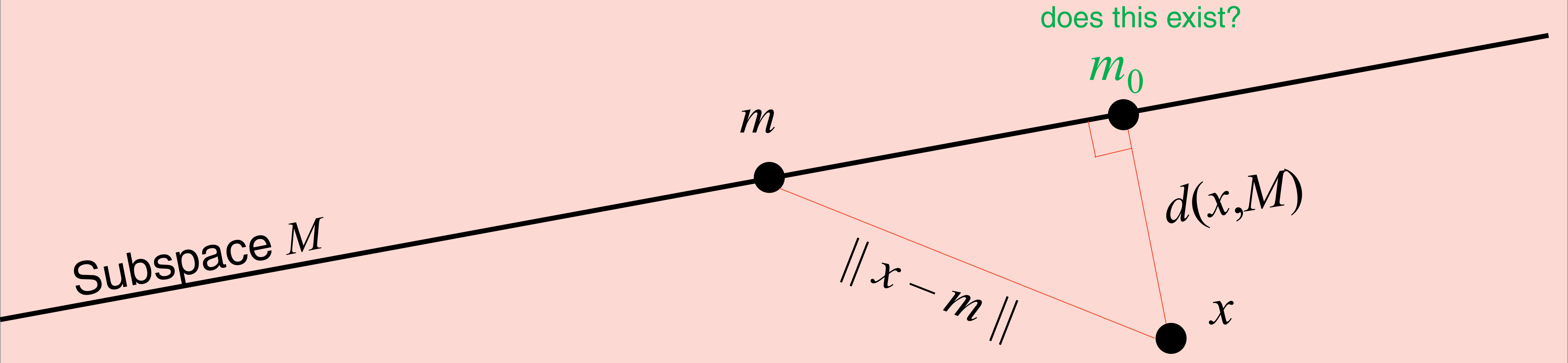
- $x \perp S$ if and only if $x \perp \text{span}\{S\}$

Warning

- The following section develops the journey to the Projection Theorem
- Keep the following pictures in your mind. The idea is to prove this picture is good.

Let $(\mathcal{X}, \mathcal{F}, \langle \cdot , \cdot \rangle)$ be an **inner product space**, M be a **subspace** of $\mathcal{X}$, and x be an arbitrary **point** in $\mathcal{X}$.

Inner product space $(\mathcal{X}, \mathcal{F}, \langle \cdot , \cdot \rangle)$



Pre-Projection Theorem

- Let $(\mathcal{X}, \mathcal{F}, \langle \cdot , \cdot \rangle)$ be an **inner product space**, M be a **subspace** of $\mathcal{X}$, and x be an arbitrary **point** in $\mathcal{X}$.
  a) If $\exists m_0 \in M$ such that $\|x - m_0\| \leq \|x - m\| \; \forall m \in M$, then m_0 is unique.
  b) A necessary and sufficient condition for m_0 to be a minimizing vector in M is that the vector $x - m_0$ is orthogonal to M.

Pre-Projection Theorem (Equivalent)

- Let $(\mathcal{X}, \mathcal{F}, \langle \cdot , \cdot \rangle)$ be an **inner product space**, M be a **subspace** of $\mathcal{X}$, and x be an arbitrary **point** in $\mathcal{X}$.
  - a') If $\exists m_0 \in M$ such that $\|x - m_0\| = d(x, M) = \inf_{m \in M} \|x - m\|$, then m_0 is unique.
  - b') $\|x - m_0\| = d(x, M) \Leftrightarrow x - m_0 \perp M$

Pre-Projection Theorem (Equivalent)

- Let $(\mathcal{X}, \mathcal{F}, \langle \cdot , \cdot \rangle)$ be an **inner product space**, M be a **subspace** of $\mathcal{X}$, and x be an arbitrary **point** in $\mathcal{X}$.

  a') If $\exists m_0 \in M$ such that $\|x - m_0\| = d(x, M) = \inf_{m \in M} \|x - m\|$, then m_0 is unique.

  b') $\|x - m_0\| = d(x, M) \Leftrightarrow x - m_0 \perp M$

- If there exists a minimizing vector, it is unique.
- The minimizing vector is completely characterized by the approximating error being orthogonal to the subspace M. (the set of approximates)
- Not a surprise, we expect the connection between orthogonality and minimum distance.

Proof: Pre-Projection Theorem

Lemma 3.29 *(called the Pre-Projection Theorem in Luenberger) Let $\mathcal{X}$ be a finite-dimensional (real) inner product space, M be a subspace of $\mathcal{X}$, and x be an arbitrary point in $\mathcal{X}$.*

(a) If $\exists m_0 \in M$ such that $\|x - m_0\| \leq \|x - m\| \ \forall m \in M$, then m_0 is unique.

(b) A necessary and sufficient condition for m_0 to be a minimizing vector in M is that the vector $x - m_0$ is orthogonal to M.

Remarks:

(a') If $\exists m_0 \in M$ such that $\|x - m_0\| = d(x, M) = \inf_{m \in M} \|x - m\|$, then m_0 is unique. [equivalent to (a)]

(b') $\|x - m_0\| = d(x, M) \iff x - m_0 \perp M$. [equivalent to (b)]

Proof: We break the proof up into a series of claims.

Claim 3.30 *If $m_0 \in M$ satisfies $\|x - m_0\| = d(x, M)$, then $x - m_0 \perp M$.*

Proof: (By contrapositive) Assume $x - m_0 \not\perp M$. We will produce $m_1 \in M$ such that $\|x - m_1\| < \|x - m_0\|$. Indeed, suppose $x - m_0 \not\perp M$. Then, $\exists m \in M$ such that $\langle x - m_0, m \rangle \neq 0$. We know $m \neq 0$, and hence we define

- $\tilde{m} = \frac{m}{\|m\|} \in M$;
- $\delta := \langle x - m_0, \tilde{m} \rangle \neq 0$; and
- $m_1 = m_0 + \delta \tilde{m} \implies m_1 \in M$.

Proof: Pre-Projection Theorem

The intuition behind the definition of m_1 is that $x - m_1$ is "closer" to being perpendicular to M than is $x - m_0$, and hence it should follow that $||x - m_1|| < ||x - m_0||$. To prove the latter point, we do a few computations:

$$\begin{aligned}
||x - m_1||^2 &= ||x - m_0 - \delta\tilde{m}||^2 \\
&= \langle x - m_0 - \delta\tilde{m}, x - m_0 - \delta\tilde{m}\rangle \\
&= \langle x - m_0, x - m_0\rangle - \delta \underbrace{\langle x - m_0, \tilde{m}\rangle}_{\delta} - \delta \underbrace{\langle \tilde{m}, x - m_0\rangle}_{\delta} + \delta^2 \underbrace{\langle \tilde{m}, \tilde{m}\rangle}_{=1} \\
&= ||x - m_0||^2 - \delta^2 \\
&< ||x - m_0||^2
\end{aligned}$$

because $\delta^2 > 0$. Hence, $||x - m_1||^2 < ||x - m_0||^2$ and therefore, $||x - m_0|| \neq \inf_{m \in M} ||x - m|| := d(x, M)$. $\square$

Proof: Pre-Projection Theorem

Claim 3.31 *If $x - m_0 \perp M$, then $\|x - m_0\| = d(x, M)$ and m_0 is unique.*

Proof: Recall the Pythagorean Theorem:

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2 \text{ when } x \perp y$$

Let $m \in M$ be arbitrary and suppose $x - m_0 \perp M$. Then $x - m_0 \perp m_0 - m$, and thus

$$\begin{aligned} \|x - m\|^2 &= \|x - m_0 + \underbrace{m_0 - m}_{\in M}\|^2 \\ &= \|x - m_0\|^2 + \|m_0 - m\|^2. \end{aligned}$$

It follows that

$$\inf_{m \in M} \|x - m\|^2 = \inf_{m \in M} \left(\|x - m_0\|^2 + \|m_0 - m\|^2\right) = \|x - m_0\|^2 + \inf_{m \in M} \|m_0 - m\|^2 = \|x - m_0\|^2.$$

The unique minimizer is m_0 because $\|m_0 - m\|^2 = 0$ only for $m = m_0$. $\square$

The two claims complete the proof.

$\blacksquare$

Proof Remarks: Pre-projection Theorem

- Applies to finite and infinite-dimensional space…
- …but does not imply existence of the minimizing vector in M
  - The Classical Projection Theorem does guarantee existence, but only on finite-dimensional inner product spaces

Definition: Orthogonal Complement

- Let $(\mathcal{X}, \mathcal{F}, \langle \cdot , \cdot \rangle)$ be an **inner product space**, and $S \subset \mathcal{X}$ a subset

$$S^{\perp} := \{x \in \mathcal{X} \mid x \perp S\} = \{x \in \mathcal{X} \mid \langle x, y \rangle = 0 \text{ for all } y \in S\}$$

  is the **orthogonal complement** of S.

- $S \subset \mathcal{X}$ is a subset but not necessarily a subspace of $\mathcal{X}$
  - $S^{\perp}$ ***is*** a subspace of $\mathcal{X}$
  - $S^{\perp} = (\mathrm{span}\{S\})^{\perp}$

Proposition

- Let $(\mathcal{X}, \mathcal{F}, \langle \cdot , \cdot \rangle)$ be a **finite**-dimensional **inner product space** and M a **subspace** of $\mathcal{X}$. Then,

$$\mathcal{X} = M \bigoplus M^{\perp}$$

- Note: $M \cap M^{\perp} = \{0\}$.
- Recall: $V + W := \{x \in \mathcal{X} \mid x = v + w$, for some $v \in V, w \in W\}$

Proof

Proposition 3.34 *Let $(\mathcal{X}, \mathcal{F}, \langle \cdot, \cdot \rangle)$ be a finite dimensional inner product space and M a subspace of $\mathcal{X}$. Then,*

$$\mathcal{X} = M \oplus M^\perp.$$

Remark 3.35 *Suppose that V and W are subspaces of $\mathcal{X}$. Then $V + W := \{x \in \mathcal{X} \mid x = v + w, \text{ for some } v \in V, w \in W\}$. Because V and W are subspaces, $0 \in V \cap W$ (the zero vector is in their intersection). If that is the only vector in the intersection, meaning $V \cap W = \{0\}$, the zero subspace, then we write $V \oplus W$, and it is called the* ***direct sum*** *of V and W. What does the direct sum get you that an ordinary sum would not? You can show that $(x \in V \oplus W) \iff (\exists$* **unique** *$v \in V, w \in W$ such that $x = v + w)$.*

Proof: If $x \in M \cap M^\perp$, then by the definition of $M^\perp$, $\langle x, x \rangle = 0$, which implies $x = 0$. Hence, $M \cap M^\perp = \{0\}$. Next, we need to show that $\mathcal{X} = M + M^\perp$, that is, every $X \in \mathcal{X}$ can be written as a sum of a vector in M and a vector in $M^\perp$.

Let $\{y^1, \ldots, y^k\}$ be a basis of M. By Corollary 2.35, it can be completed to a basis for $\mathcal{X}$, that is,

$$\mathcal{X} = \text{span}\{y^1, y^2, \ldots, y^k, y^{k+1}, \ldots, y^n\} \text{ and } \{y^1, y^2, \ldots, y^k, y^{k+1}, \ldots, y^n\} \text{ is linearly independent.}$$

- What next?

Proof

We can then apply Gram-Schmidt to produce orthonormal vectors $\{v^1, \dots, v^k, v^{k+1}, \dots, v^n\}$ such that

$$\text{span}\{v^1, \dots, v^k\} = \text{span}\{y^1, \dots, y^k\} = M \text{ and } \text{span}\{v^1, \dots, v^k, v^{k+1}, \dots, v^n\} = \mathcal{X}.$$

An easy calculation gives

$$M^\perp = \text{span}\{v^{k+1}, \dots, v^n\}.$$

Indeed, suppose $x = \alpha_1 v^1 + \cdots + \alpha_k v^k + \alpha_{k+1} v^{k+1} + \cdots + \alpha_n v^n$. Then $x \in M^\perp \iff x \perp M \iff \langle x, v^i \rangle = 0, 1 \le i \le k$. However,

$$\begin{aligned} \langle x, v^i \rangle &= \alpha_1 \langle v^1, v^i \rangle + \cdots + \alpha_i \langle v^i, v^i \rangle + \cdots + \alpha_n \langle v^n, v^i \rangle \\ &= \alpha_i \quad (\text{because } \langle v^j, v^i \rangle = 0,\ j \neq i, \text{ and } \langle v^i, v^i \rangle = 1) \end{aligned}$$

and therefore $x \perp M \iff \alpha_i = 0,\ 1 \le i \le k$. This yields $(x \in M^\perp) \iff (x = \alpha_{k+1} v^{k+1} + \cdots + \alpha_n v^n) \iff (x \in \text{span}\{v^{k+1}, \cdots, v^n\})$. Therefore,

$$M^\perp = \text{span}\{v^{k+1}, \dots, v^n\}.$$

■

- Gram Schmidt Process returns!

Classical Projection Theorem

- Let $(\mathcal{X}, \mathcal{F}, \langle \cdot , \cdot \rangle)$ be a finite-dimensional (real) **inner product space** and M a **subspace** of $\mathcal{X}$. Then, $\forall x \in \mathcal{X}$, $\exists$ **unique** $\hat{x} \in M$ such that

$$\|x - \hat{x}\| = d(x, M) := \inf_{m \in M} \|x - m\| = \min_{m \in M} \|x - m\|$$

where we can write **min**imum instead of **inf**imum, because the infimum is achieved. Moreover $\hat{x} \in M$ is characterized by $x - \hat{x} \perp M$

Classical Projection Theorem

- Let $(\mathcal{X}, \mathcal{F}, \langle \cdot , \cdot \rangle)$ be a finite-dimensional (real) **inner product space** and M a **subspace** of $\mathcal{X}$. Then, $\forall x \in \mathcal{X}$, $\exists$ **unique** $\hat{x} \in M$ such that

$$\|x - \hat{x}\| = d(x, M) := \inf_{m \in M} \|x - m\| = \min_{m \in M} \|x - m\|$$

  where we can write **min**imum instead of **inf**imum, because the infimum is achieved. Moreover $\hat{x} \in M$ is characterized by $x - \hat{x} \perp M$

- Use of $\hat{x}$ instead of m_0 because this is the standard notation for estimation problems.

Proof: Classical Projection Theorem

Theorem 3.36 ***(Classical Projection Theorem)*** *Let* $(\mathcal{X}, \mathbb{R})$ *be a finite dimensional real inner product space and* M *a subspace of* $\mathcal{X}$*. Then,* $\forall x \in \mathcal{X}, \exists$ unique $\widehat{x} \in M$ *such that*

$$\|x - \widehat{x}\| = d(x, M) := \inf_{m \in M} \|x - m\| = \min_{m \in M} \|x - m\|,$$

where we can write min *instead of* inf *because the infimum is achieved. Moreover,* $\widehat{x} \in M$ *is characterized by* $x - \widehat{x} \perp M$*.*

Proof: We only need to show that $\forall x \in \mathcal{X}$ there exists $\widehat{x} \in M$ such that $(x - \widehat{x}) \perp M$. This is because if such an $\widehat{x}$ exists, Lemma 3.29, the "Pre-projection Theorem," already shows that it is unique and $\|x - \widehat{x}\| = d(x, M)$. By Proposition 3.34, $\mathcal{X} = M \oplus M^{\perp}$. Therefore, there exist $\widehat{x} \in M$ and $m^{\perp} \in M^{\perp}$ such that

$$x = \widehat{x} + m^{\perp}.$$

Hence,

$$x - \widehat{x} = m^{\perp} \in M^{\perp} \implies (x - \widehat{x}) \perp M.$$

$\blacksquare$

Remark 3.37 *You may have observed that* $\mathcal{X} = M \oplus M^{\perp}$ *also shows that* $\widehat{x}$ *is unique. While this is true, it is based on Proposition 3.34, which is true when* $\mathcal{X}$ *is a "complete" inner product space and* M *is a "closed" subspace, properties that are automatically satisfied when* $\mathcal{X}$ *is finite dimensional. We will touch on these more subtle properties later when we do some basic Real Analysis.*

Next week

- Normal Equations
  - Solving actual, useful problems $\hat{x} = \arg\min_{x \in M} \|Ax - b\|$