



5.10 The Gram-Schmidt algorithm





Introduction

- In this topic, we will
 - Learn about the best approximation of a given vector of collection of m mutually orthogonal vectors
 - Define the Gram-Schmidt algorithm for taking a set of m vectors and returning a collection of at most m mutually orthogonal unit vectors
 - Look at examples as go through the steps





Best approximations

- Suppose we have two vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ that are orthogonal
 - Neither contains information about the other
 - Recall the best approximation of a given vector $\mathbf{v}$ of either of this is found by calculating the projection:

$$\text{proj}_{\mathbf{u}_1}(\mathbf{v})$$

$$\text{proj}_{\mathbf{u}_2}(\mathbf{v})$$





Best approximations

- Question: what is the best approximation of $\mathbf{v}$ as a linear combination of two orthogonal vectors $\mathbf{u}_1$ and $\mathbf{u}_2$?
 - As before, minimizing $\|\mathbf{v} - \alpha_1 \mathbf{u}_1 - \alpha_2 \mathbf{u}_2\|_2$ is equivalent to minimizing

$$\begin{aligned}\|\mathbf{v} - \alpha_1 \mathbf{u}_1 - \alpha_2 \mathbf{u}_2\|_2^2 &= \langle \mathbf{v} - \alpha_1 \mathbf{u}_1 - \alpha_2 \mathbf{u}_2, \mathbf{v} - \alpha_1 \mathbf{u}_1 - \alpha_2 \mathbf{u}_2 \rangle \\ &= \langle \mathbf{v}, \mathbf{v} \rangle - \alpha_1 \langle \mathbf{v}, \mathbf{u}_1 \rangle - \alpha_2 \langle \mathbf{v}, \mathbf{u}_2 \rangle \\ &\quad - \alpha_1 \langle \mathbf{u}_1, \mathbf{v} \rangle + \alpha_1^2 \langle \mathbf{u}_1, \mathbf{u}_1 \rangle + \cancel{\alpha_1 \alpha_2 \langle \mathbf{u}_1, \mathbf{u}_2 \rangle} \\ &\quad - \alpha_2 \langle \mathbf{u}_2, \mathbf{v} \rangle + \cancel{\alpha_1 \alpha_2 \langle \mathbf{u}_2, \mathbf{u}_1 \rangle} + \alpha_2^2 \langle \mathbf{u}_2, \mathbf{u}_2 \rangle \\ &= \langle \mathbf{v}, \mathbf{v} \rangle \\ &\quad - 2\alpha_1 \langle \mathbf{u}_1, \mathbf{v} \rangle + \alpha_1^2 \langle \mathbf{u}_1, \mathbf{u}_1 \rangle \\ &\quad - 2\alpha_2 \langle \mathbf{u}_2, \mathbf{v} \rangle + \alpha_2^2 \langle \mathbf{u}_2, \mathbf{u}_2 \rangle\end{aligned}$$





Best approximations

- Now, note that the minimum of

$$x^2 + 2y^2 + 3x + 5y + 6$$

can be found by minimizing both $x^2 + 3x$ and $2y^2 + 5y$ separately

- In this case, $x = -1.5$ and $y = -1.25$

- This is because if

- the minimum of $f(x)$ is at $x_{\min}$ and

- the minimum of $g(y)$ is at $y_{\min}$,

the minimum of $f(x) + g(y)$ is $f(x_{\min}) + g(y_{\min})$

- This can be shown by a simple proof by contradiction...

- The polynomial $x^2 + 2y^2 + 3x + 4xy + 5y + 6$ has a cross term,
so the minimum is at $x = -1$ and $y = -0.25$

- For this, we must solve $2x + 4y = -3$

$$4y + 4x = -5$$



Best approximations

- Thus, the best approximation of $\mathbf{v}$ as a linear combination of two orthogonal vectors is:

$$\text{proj}_{\mathbf{u}_1}(\mathbf{v}) + \text{proj}_{\mathbf{u}_2}(\mathbf{v})$$

- Thus, given m vectors $\mathbf{u}_1, \dots, \mathbf{u}_m$, all of which are *mutually orthogonal*, then the best approximation of $\mathbf{v}$ as a linear combination of all these vectors is

$$\text{proj}_{\mathbf{u}_1}(\mathbf{v}) + \text{proj}_{\mathbf{u}_2}(\mathbf{v}) + \dots + \text{proj}_{\mathbf{u}_m}(\mathbf{v})$$

- This can be shown using a proof by induction...





Best approximations

- Next, consider the work required to calculate this best approximation:

$$\frac{\langle \mathbf{u}_1, \mathbf{v} \rangle}{\langle \mathbf{u}_1, \mathbf{u}_1 \rangle} \mathbf{u}_1 + \frac{\langle \mathbf{u}_2, \mathbf{v} \rangle}{\langle \mathbf{u}_2, \mathbf{u}_2 \rangle} \mathbf{u}_2 + \cdots + \frac{\langle \mathbf{u}_m, \mathbf{v} \rangle}{\langle \mathbf{u}_m, \mathbf{u}_m \rangle} \mathbf{u}_m$$

- We could reduce the work by half if we normalized all of the vectors:

$$\langle \hat{\mathbf{u}}_1, \mathbf{v} \rangle \hat{\mathbf{u}}_1 + \langle \hat{\mathbf{u}}_2, \mathbf{v} \rangle \hat{\mathbf{u}}_2 + \cdots + \langle \hat{\mathbf{u}}_m, \mathbf{v} \rangle \hat{\mathbf{u}}_m$$





Gram-Schmidt algorithm

Problem:

Suppose we have m vectors $\mathbf{u}_1, \dots, \mathbf{u}_m$, none of which are necessarily normalized and no two of which are necessarily orthogonal

Can we convert these into an associated collection of at most m vectors, all of which are unit vectors, and all of which are mutually orthogonal?





Gram-Schmidt algorithm

Solution:

The Gram-Schmidt algorithm.

Step 1.

Normalize the first vector, so define $\hat{\mathbf{u}}_1 \leftarrow \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|_2}$





Gram-Schmidt algorithm

Step 2.

Subtract from $\mathbf{u}_2$ the projection of $\mathbf{u}_2$ onto $\hat{\mathbf{u}}_1$

$$\mathbf{u}_2 \leftarrow \mathbf{u}_2 - \langle \hat{\mathbf{u}}_1, \mathbf{u}_2 \rangle \hat{\mathbf{u}}_1$$

If $\mathbf{u}_2$ is now $\mathbf{0}$, it was a scalar multiple of $\hat{\mathbf{u}}_1$,
so it contains no additional information: discard it

Otherwise, normalize the second vector, so define

$$\hat{\mathbf{u}}_2 \leftarrow \frac{\mathbf{u}_2}{\|\mathbf{u}_2\|_2}$$





Example

- Suppose we have the two vectors $\mathbf{u}_1 = \begin{pmatrix} 1.5 \\ -2.0 \end{pmatrix}$ and $\mathbf{u}_2 = \begin{pmatrix} 3.5 \\ 4.5 \end{pmatrix}$
- Step 1: $\hat{\mathbf{u}}_1 \leftarrow \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|_2} = \frac{\begin{pmatrix} 1.5 \\ -2.0 \end{pmatrix}}{2.5} = \begin{pmatrix} 0.6 \\ -0.8 \end{pmatrix}$
- Step 2: $\mathbf{u}_2 \leftarrow \mathbf{u}_2 - \langle \hat{\mathbf{u}}_1, \mathbf{u}_2 \rangle \hat{\mathbf{u}}_1 = \begin{pmatrix} 3.5 \\ 4.5 \end{pmatrix} - \left\langle \begin{pmatrix} 0.6 \\ -0.8 \end{pmatrix}, \begin{pmatrix} 3.5 \\ 4.5 \end{pmatrix} \right\rangle \begin{pmatrix} 0.6 \\ -0.8 \end{pmatrix}$
 $= \begin{pmatrix} 3.5 \\ 4.5 \end{pmatrix} - (-1.5) \begin{pmatrix} 0.6 \\ -0.8 \end{pmatrix} = \begin{pmatrix} 4.4 \\ 3.3 \end{pmatrix}$
 $\hat{\mathbf{u}}_2 \leftarrow \frac{\mathbf{u}_2}{\|\mathbf{u}_2\|_2} = \frac{\begin{pmatrix} 4.4 \\ 3.3 \end{pmatrix}}{5.5} = \begin{pmatrix} 0.8 \\ 0.6 \end{pmatrix}$





Example

- Note that $\hat{\mathbf{u}}_1 = \begin{pmatrix} 0.6 \\ -0.8 \end{pmatrix}$ and $\hat{\mathbf{u}}_2 = \begin{pmatrix} 0.8 \\ 0.6 \end{pmatrix}$ are both unit vectors and are orthogonal to each other

- Note that we can easily reconstruct the original vectors as a linear combination of these two vectors, so no information was lost

- Also, given any vector in $\mathbf{R}^2$, say $\mathbf{v} = \begin{pmatrix} -7 \\ -20 \end{pmatrix}$, the best approximation is

$$\langle \hat{\mathbf{u}}_1, \mathbf{v} \rangle \hat{\mathbf{u}}_1 + \langle \hat{\mathbf{u}}_2, \mathbf{v} \rangle \hat{\mathbf{u}}_2 = 11.8 \begin{pmatrix} 0.6 \\ -0.8 \end{pmatrix} - 17.6 \begin{pmatrix} 0.8 \\ 0.6 \end{pmatrix} = \begin{pmatrix} -7 \\ -20 \end{pmatrix}$$

- In this case, the best approximation equals $\mathbf{v}$





Example

- Suppose we have the two vectors $\mathbf{u}_1 = \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}$ and $\mathbf{u}_2 = \begin{pmatrix} 6 \\ -2 \\ -1 \\ -3 \end{pmatrix}$
- Step 1: $\hat{\mathbf{u}}_1 \leftarrow \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|_2} = \frac{\mathbf{u}_1}{2} = \begin{pmatrix} 0.5 \\ -0.5 \\ 0.5 \\ -0.5 \end{pmatrix}$
- Step 2: $\mathbf{u}_2 \leftarrow \mathbf{u}_2 - \langle \hat{\mathbf{u}}_1, \mathbf{u}_2 \rangle \hat{\mathbf{u}}_1 = \begin{pmatrix} 6 \\ -2 \\ -1 \\ -3 \end{pmatrix} - 5 \begin{pmatrix} 0.5 \\ -0.5 \\ 0.5 \\ -0.5 \end{pmatrix} = \begin{pmatrix} 3.5 \\ 0.5 \\ -3.5 \\ -0.5 \end{pmatrix}$
 $\hat{\mathbf{u}}_2 \leftarrow \frac{\mathbf{u}_2}{\|\mathbf{u}_2\|_2} = \frac{\mathbf{u}_2}{5} = \begin{pmatrix} 0.7 \\ 0.1 \\ -0.7 \\ -0.1 \end{pmatrix}$





Example

- Thus, we have $\hat{\mathbf{u}}_1 = \begin{pmatrix} 0.5 \\ -0.5 \\ 0.5 \\ -0.5 \end{pmatrix}$ and $\hat{\mathbf{u}}_2 = \begin{pmatrix} 0.7 \\ 0.1 \\ -0.7 \\ -0.1 \end{pmatrix}$

- The best approximation of $\mathbf{v} = \begin{pmatrix} -5 \\ -4 \\ 11 \\ -3 \end{pmatrix}$ is found by calculating

$$\langle \hat{\mathbf{u}}_1, \mathbf{v} \rangle \hat{\mathbf{u}}_1 + \langle \hat{\mathbf{u}}_2, \mathbf{v} \rangle \hat{\mathbf{u}}_2 = 6.5 \begin{pmatrix} 0.5 \\ -0.5 \\ 0.5 \\ -0.5 \end{pmatrix} - 11.3 \begin{pmatrix} 0.7 \\ 0.1 \\ -0.7 \\ -0.1 \end{pmatrix} = \begin{pmatrix} -4.66 \\ -4.38 \\ 11.16 \\ -2.12 \end{pmatrix}$$





Gram-Schmidt algorithm

Step 3.

Subtract from $\mathbf{u}_3$ the projection of $\mathbf{u}_3$ onto $\hat{\mathbf{u}}_1$

$$\mathbf{u}_3 \leftarrow \mathbf{u}_3 - \langle \hat{\mathbf{u}}_1, \mathbf{u}_3 \rangle \hat{\mathbf{u}}_1$$

If $\mathbf{u}_3$ is now $\mathbf{0}$, it was a scalar multiple of $\hat{\mathbf{u}}_1$,

so it contains no additional information: discard it

Subtract from this new $\mathbf{u}_3$ the projection of $\mathbf{u}_3$ onto $\hat{\mathbf{u}}_2$

$$\mathbf{u}_3 \leftarrow \mathbf{u}_3 - \langle \hat{\mathbf{u}}_2, \mathbf{u}_3 \rangle \hat{\mathbf{u}}_2$$

If $\mathbf{u}_3$ is now $\mathbf{0}$, it was a linear combination of $\hat{\mathbf{u}}_1$ and $\hat{\mathbf{u}}_2$

so it contains no additional information: discard it

Otherwise, normalize the this new third vector, so define

$$\hat{\mathbf{u}}_3 \leftarrow \frac{\mathbf{u}_3}{\|\mathbf{u}_3\|_2}$$





Example

- Suppose we now have $\mathbf{u}_1 = \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}$, $\mathbf{u}_2 = \begin{pmatrix} 6 \\ -2 \\ -1 \\ -3 \end{pmatrix}$ and $\mathbf{u}_3 = \begin{pmatrix} 2 \\ 3 \\ -12 \\ 1 \end{pmatrix}$

- We have $\hat{\mathbf{u}}_1 = \begin{pmatrix} 0.5 \\ -0.5 \\ 0.5 \\ -0.5 \end{pmatrix}$ and $\hat{\mathbf{u}}_2 = \begin{pmatrix} 0.7 \\ 0.1 \\ -0.7 \\ -0.1 \end{pmatrix}$,

and thus we need only apply Gram-Schmidt to $\mathbf{u}_3$





Example

- Step 3:

$$\mathbf{u}_3 \leftarrow \mathbf{u}_3 - \langle \hat{\mathbf{u}}_1, \mathbf{u}_3 \rangle \hat{\mathbf{u}}_1 = \begin{pmatrix} 2 \\ 3 \\ -12 \\ 1 \end{pmatrix} - (-7) \begin{pmatrix} 0.5 \\ -0.5 \\ 0.5 \\ -0.5 \end{pmatrix} = \begin{pmatrix} 5.5 \\ -0.5 \\ -8.5 \\ -2.5 \end{pmatrix}$$

$$\mathbf{u}_3 \leftarrow \mathbf{u}_3 - \langle \hat{\mathbf{u}}_2, \mathbf{u}_3 \rangle \hat{\mathbf{u}}_2 = \begin{pmatrix} 5.5 \\ -0.5 \\ -8.5 \\ -2.5 \end{pmatrix} - 10 \begin{pmatrix} 0.7 \\ 0.1 \\ -0.7 \\ -0.1 \end{pmatrix} = \begin{pmatrix} -1.5 \\ -1.5 \\ -1.5 \\ -1.5 \end{pmatrix}$$

$$\hat{\mathbf{u}}_3 \leftarrow \frac{\mathbf{u}_3}{\|\mathbf{u}_3\|_2} = \frac{\mathbf{u}_3}{3} = \begin{pmatrix} -0.5 \\ -0.5 \\ -0.5 \\ -0.5 \end{pmatrix}$$





Example

- Thus, we have $\hat{\mathbf{u}}_1 = \begin{pmatrix} 0.5 \\ -0.5 \\ 0.5 \\ -0.5 \end{pmatrix}$, $\hat{\mathbf{u}}_2 = \begin{pmatrix} 0.7 \\ 0.1 \\ -0.7 \\ -0.1 \end{pmatrix}$ and $\hat{\mathbf{u}}_3 = \begin{pmatrix} -0.5 \\ -0.5 \\ -0.5 \\ -0.5 \end{pmatrix}$

- The best approximation of $\mathbf{v} = \begin{pmatrix} -5 \\ -4 \\ 11 \\ -3 \end{pmatrix}$ is found by calculating

$$\langle \hat{\mathbf{u}}_1, \mathbf{v} \rangle \hat{\mathbf{u}}_1 + \langle \hat{\mathbf{u}}_2, \mathbf{v} \rangle \hat{\mathbf{u}}_2 + \langle \hat{\mathbf{u}}_3, \mathbf{v} \rangle \hat{\mathbf{u}}_3$$

$$= 6.5 \begin{pmatrix} 0.5 \\ -0.5 \\ 0.5 \\ -0.5 \end{pmatrix} - 11.3 \begin{pmatrix} 0.7 \\ 0.1 \\ -0.7 \\ -0.1 \end{pmatrix} + 0.5 \begin{pmatrix} -0.5 \\ -0.5 \\ -0.5 \\ -0.5 \end{pmatrix} = \begin{pmatrix} -4.91 \\ -4.63 \\ 10.91 \\ -2.37 \end{pmatrix}$$





Gram-Schmidt algorithm

Step k .

Subtract from $\mathbf{u}_k$ the projection of $\mathbf{u}_k$ onto $\hat{\mathbf{u}}_\ell$ for $\ell = 1, 2, \dots, k-1$ where we did not discard that vector

$$\mathbf{u}_k \leftarrow \mathbf{u}_k - \langle \hat{\mathbf{u}}_1, \mathbf{u}_k \rangle \hat{\mathbf{u}}_1$$

$$\mathbf{u}_k \leftarrow \mathbf{u}_k - \langle \hat{\mathbf{u}}_2, \mathbf{u}_k \rangle \hat{\mathbf{u}}_2$$

$$\vdots$$

$$\mathbf{u}_k \leftarrow \mathbf{u}_k - \langle \hat{\mathbf{u}}_{k-1}, \mathbf{u}_k \rangle \hat{\mathbf{u}}_{k-1}$$

If at any point, $\mathbf{u}_k = \mathbf{0}$, we discard this vector.

Otherwise, normalize the this new vector, so define

$$\hat{\mathbf{u}}_k \leftarrow \frac{\mathbf{u}_k}{\|\mathbf{u}_k\|_2}$$





Gram-Schmidt algorithm

- To summarize:
 - Given m vectors $\mathbf{u}_1$ through $\mathbf{u}_m$, to convert these into a collection of mutually orthogonal unit vectors:
 - Starting with $\mathbf{u}_1$, normalize it and assign that to $\hat{\mathbf{u}}_1$
 - Then, letting $k = 2, 3, 4, \dots, m$, do the following:
 - Subtract from $\mathbf{u}_k$ the projection of it onto $\hat{\mathbf{u}}_1$
 - Repeat this, subtracting from $\mathbf{u}_k$ the projection of it onto each $\hat{\mathbf{u}}_\ell$ that was not discarded where $\ell = 2, \dots, k-1$
 - If at the end of this process, $\mathbf{u}_k = \mathbf{0}$, then discard this vector
 - Otherwise, normalize $\mathbf{u}_k$ and assign it to $\hat{\mathbf{u}}_k$





Gram-Schmidt algorithm

- Warning: There are two forms of this algorithm
 - The most common one taught is not numerically stable
 - That means, it works great for toy examples taught at the undergraduate level, but it fails when the vectors are numeric
 - The algorithm taught in this class is more stable numerically
 - You will cover numerical stability in your course on numerical analysis and methods





Summary

- Following this topic, you now
 - Know that you can easily calculate the best approximation of a given vector as a linear combination of mutually orthogonal vectors
 - Understand that the Gram-Schmidt algorithm can take a collection of arbitrary vectors and convert them into a collection of mutually orthogonal unit vectors





References

- [1] https://en.wikipedia.org/wiki/Gram%E2%80%93Schmidt_process
- [2] https://en.wikipedia.org/wiki/Inner_product_space





Acknowledgments

None so far.





Colophon

These slides were prepared using the Cambria typeface. Mathematical equations use Times New Roman, and source code is presented using Consolas. Mathematical equations are prepared in MathType by Design Science, Inc. Examples may be formulated and checked using Maple by Maplesoft, Inc.

The photographs of flowers and a monarch butter appearing on the title slide and accenting the top of each other slide were taken at the Royal Botanical Gardens in October of 2017 by Douglas Wilhelm Harder. Please see

<https://www.rbg.ca/>

for more information.





Disclaimer

These slides are provided for the NE 112 *Linear algebra for nanotechnology engineering* course taught at the University of Waterloo. The material in it reflects the authors' best judgment in light of the information available to them at the time of preparation. Any reliance on these course slides by any party for any other purpose are the responsibility of such parties. The authors accept no responsibility for damages, if any, suffered by any party as a result of decisions made or actions based on these course slides for any other purpose than that for which it was intended.

