Homework 10

This homework is due on Saturday, April 6, 2024, at 11:59PM.

1. Gram-Schmidt Basics

(a) **Use Gram-Schmidt to find a matrix U whose columns form an orthonormal basis for the column space of V.**

$$V = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \tag{1}$$

(b) Show that you get the same resulting vector when you project $\vec{w} = \begin{bmatrix} 1 \\ -1 \\ 0 \\ -1 \\ 0 \end{bmatrix}$ onto the columns of V as you do when you project onto the columns of U, i.e. **show that**

$$V(V^\top V)^{-1}V^\top \vec{w} = U(U^\top U)^{-1}U^\top \vec{w}. \tag{2}$$

Feel free to use `NumPy` for the projection onto the columns of V, but compute the projection onto the columns of U by hand. Comment on whether projecting upon the V or U basis is computationally more efficient. *(HINT: Which of these matrices allow us to circumvent the matrix inversion in the projection formula?)*

2. Correctness of the Gram-Schmidt Algorithm

Suppose we take a list of vectors $\{\vec{a}_1, \vec{a}_2, \dots, \vec{a}_n\}$ and run the following Gram-Schmidt algorithm on it to perform orthonormalization. It produces the vectors $\{\vec{q}_1, \vec{q}_2, \dots, \vec{q}_n\}$.

```
for $i = 1$ up to $n$ do                                          ▷ Iterate through the vectors
    $\vec{r}_i = \vec{a}_i - \sum_{j<i} \vec{q}_j \left(\vec{q}_j^\top \vec{a}_i\right)$       ▷ Find the amount of $\vec{a}_i$ that remains after we project
    if $\vec{r}_i = \vec{0}$ then
        $\vec{q}_i = \vec{0}$
    else
        $\vec{q}_i = \frac{\vec{r}_i}{\|\vec{r}_i\|}$                                        ▷ Normalize the vector.
    end if
end for
```

In this problem, we prove the correctness of the Gram-Schmidt algorithm by showing that the following three properties hold on the vectors output by the algorithm.

1. If $\vec{q}_i \neq \vec{0}$, then $\vec{q}_i^\top \vec{q}_i = \|\vec{q}_i\|^2 = 1$ (i.e. the $\vec{q}_i$ have unit norm whenever they are nonzero).
2. For all $1 \leq \ell \leq n$, $\text{Span}(\{\vec{a}_1, \dots, \vec{a}_\ell\}) = \text{Span}(\{\vec{q}_1, \dots, \vec{q}_\ell\})$.
3. For all $i \neq j$, $\vec{q}_i^\top \vec{q}_j = 0$ (i.e. $\vec{q}_i$ and $\vec{q}_j$ are orthogonal).

(a) First, we show that the first property holds by construction from the if/then/else statement in the algorithm. It holds when $\vec{q}_i = \vec{0}$, since the first property has no restrictions on $\vec{q}_i$ if it is the zero vector. **Show that $\|\vec{q}_i\| = 1$ if $\vec{q}_i \neq \vec{0}$.**

(b) Next, we show the second property by considering each ℓ from 1 to n, and showing the statement that $\text{Span}(\{\vec{a}_1, \dots, \vec{a}_\ell\}) = \text{Span}(\{\vec{q}_1, \dots, \vec{q}_\ell\})$. This statement is true when $\ell = 1$ since the algorithm produces $\vec{q}_1$ as a scaled version of $\vec{a}_1$. Now assume that this statement is true for $\ell = k - 1$. Under this assumption, **show that the spans are the same for $\ell = k$.**

This implies that because $\text{Span}(\{\vec{a}_1\}) = \text{Span}(\{\vec{q}_1\})$, then so too is $\text{Span}(\{\vec{a}_1, \vec{a}_2\}) = \text{Span}(\{\vec{q}_1, \vec{q}_2\})$, and so forth, until we get that $\text{Span}(\{\vec{a}_1, \dots, \vec{a}_n\}) = \text{Span}(\{\vec{q}_1, \dots, \vec{q}_n\})$.

(HINT: What you need to show is: if there exists $\vec{\alpha} = \begin{bmatrix} \alpha_1 & \cdots & \alpha_k \end{bmatrix} \neq \vec{0}_k$ so that $\vec{y} = \sum_{j=1}^{k} \alpha_j \vec{a}_j$, then there exists $\vec{\beta} = \begin{bmatrix} \beta_1 & \cdots & \beta_k \end{bmatrix} \neq \vec{0}_k$ such that $\vec{y} = \sum_{j=1}^{k} \beta_j \vec{q}_j$ (this is the forward direction). And vice versa from $\vec{\beta}$ to $\vec{\alpha}$ (this is the reverse direction).)

(HINT: To show the forward direction, write $\vec{a}_k$ in terms of $\vec{q}_k$ and earlier $\vec{q}_j$. Use the condition for $\ell = k - 1$ to show the condition for $\ell = k$. Don't forget the case that $\vec{q}_k = \vec{0}$. The reverse direction may be approached similarly.)

(c) Lastly, we establish orthogonality between every pair of vectors in $\{\vec{q}_1, \vec{q}_2, \dots, \vec{q}_n\}$. Consider each ℓ from 1 to n. We want to show the statement that for all $j < \ell$, $\vec{q}_j^\top \vec{q}_\ell = 0$. The statement holds for $\ell = 1$ since there are no $j < 1$. Assume that this statement holds for ℓ up to and including $k - 1$. That is, we assume that for all $i \leq k - 1$, $\vec{q}_j^\top \vec{q}_i = 0$ for all $j < i$.

Under this assumption, **show that for all $i \leq k$, that $\vec{q}_j^\top \vec{q}_i = 0$ for all $j < i$.** This shows that every pair of distinct vectors up to $1, 2, \dots, \ell$ are orthogonal for each ℓ from 1 to n.

(HINT: The cases $i \leq k - 1$ are already covered by the assumption. So you can focus on $i = k$. Next, notice that the case $\vec{q}_k = \vec{0}$ is also true, since the inner product of any vector with $\vec{q}_k = \vec{0}$ is $\vec{0}$. So, focus on the case $\vec{q}_k \neq \vec{0}$ and expand what you know about $\vec{q}_k$.)

3. Orthonormality, and Gram-Schmidt Potpourri

(a) Suppose $Q \in \mathbb{R}^{m \times n}$ is tall (i.e. $m \geq n$) matrix and has orthonormal columns, which of the following is true:

○	$Q^\top Q = I_n$
○	$QQ^\top = I_m$
○	neither
○	both

(b) Suppose $Q \in \mathbb{R}^{n \times n}$ is a square, orthonormal matrix, which of the following is true:

○	$Q^\top Q = I_n$
○	$QQ^\top = I_n$
○	neither
○	both

(c) Suppose $Q \in \mathbb{R}^{m \times n}$ is a wide (i.e. $m \leq n$) matrix and has orthonormal rows, which of the following is true:

○	$Q^\top Q = I_n$
○	$QQ^\top = I_m$
○	neither
○	both

(d) **Using Gram-Schmidt, find an orthonormal basis for $\mathbb{R}^2$ starting with the vectors $\vec{q}_1 = \begin{bmatrix} 3 \\ \sqrt{3} \end{bmatrix}$ and $\vec{q}_2 = \begin{bmatrix} -\sqrt{3} \\ 3 \end{bmatrix}$. Let $\vec{v}_1$ be the corresponding orthonormal basis vector for $\vec{q}_1$ and $\vec{v}_2$ be the corresponding orthonormal basis vector for $\vec{q}_2$.**

© UCB EECS 16B, Spring 2024. All Rights Reserved. This may not be publicly shared without explicit permission.

4. Spectral Theorem for Real Symmetric Matrices

We want to show that every real symmetric matrix can be diagonalized by a matrix of its orthonormal eigenvectors. In other words, a symmetric matrix $S \in \mathbb{R}^{n \times n}$, i.e., a matrix S such that $S = S^\top$, can be written as $S = V \Lambda V^\top$, where $V \in \mathbb{R}^{n \times n}$ is an orthonormal matrix of eigenvectors of S and $\Lambda \in \mathbb{R}^{n \times n}$ is a diagonal matrix of corresponding real eigenvalues of S. This is called the Spectral Theorem for real symmetric matrices.

To prove this, we will use block matrix manipulation and the induction proof technique.

(a) One part of the spectral theorem can be proved without any further delay. **Prove that the eigenvalues λ of a real, symmetric matrix S are real.**

(HINT: Let λ be an eigenvalue of S with corresponding nonzero eigenvector $\vec{v}$. Evaluate $\overline{\vec{v}}^\top S \vec{v}$ in two different ways: $\overline{\vec{v}}^\top (S\vec{v})$ and $(\overline{\vec{v}}^\top S)\vec{v}$. What does this show about λ?)

(b) For the main proof that every real symmetric matrix is diagonalized by a matrix of its orthonormal real eigenvectors, we will proceed by induction.

Show the base case: every 1×1 symmetric matrix S can be written as $S = V \Lambda V^\top$, where V is a real and orthonormal matrix of eigenvectors of S, and Λ is a real and diagonal matrix of corresponding eigenvalues of S.

(HINT: Every 1×1 matrix is symmetric, and also diagonal, by definition; the only real orthonormal 1×1 matrices are $\begin{bmatrix} 1 \end{bmatrix}$ and $\begin{bmatrix} -1 \end{bmatrix}$.)

(c) With the base case done, we are now in the inductive step. Let S be an arbitrary $n \times n$ symmetric matrix; ultimately, we want to show that $S = V \Lambda V^\top$, where V is a real and orthonormal matrix of eigenvectors of S, and Λ is a real and diagonal matrix of corresponding eigenvalues of S.

To start, let λ be an eigenvalue of S, and let $\vec{q}$ be any normalized eigenvector of S corresponding to eigenvalue λ. Let $\widetilde{Q} \in \mathbb{R}^{n \times (n-1)}$ be a set of orthonormal vectors chosen so that $Q := \begin{bmatrix} \vec{q} & \widetilde{Q} \end{bmatrix} \in \mathbb{R}^{n \times n}$ is an orthonormal matrix.[1] **Show the following equality:**

$$Q^\top S Q = \begin{bmatrix} \lambda & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & S_0 \end{bmatrix} \qquad \textbf{where} \qquad S_0 := \widetilde{Q}^\top S \widetilde{Q}. \tag{3}$$

(HINT: Expand Q as a block matrix $\begin{bmatrix} \vec{q} & \widetilde{Q} \end{bmatrix}$ and multiply $Q^\top S Q = \begin{bmatrix} \vec{q} & \widetilde{Q} \end{bmatrix}^\top S \begin{bmatrix} \vec{q} & \widetilde{Q} \end{bmatrix}$.)

(HINT: Since Q is orthonormal, we have $Q^\top Q = I_n$. What does this mean for the values of $\vec{q}^\top \vec{q}$ and $\widetilde{Q}^\top \vec{q}$? Use block matrix multiplication on $Q^\top Q = \begin{bmatrix} \vec{q} & \widetilde{Q} \end{bmatrix}^\top \begin{bmatrix} \vec{q} & \widetilde{Q} \end{bmatrix}$ again.)

(d) **Show that the matrix S_0 is a real symmetric matrix.**

---

[1] This matrix $\widetilde{Q}$ can be generated via Gram-Schmidt, for example.

© UCB EECS 16B, Spring 2024. All Rights Reserved. This may not be publicly shared without explicit permission.

(e) Since S_0 is a real symmetric $(n-1) \times (n-1)$ matrix, by our inductive assumption, S_0 can be orthonormally diagonalized as $S_0 = V_0 \Lambda_0 V_0^\top$, where Λ_0 is a real diagonal matrix of eigenvalues of S_0 and $V_0 \in \mathbb{R}^{(n-1)\times(n-1)}$ is a real orthonormal matrix of corresponding eigenvectors of S_0.

Define

$$V := Q \begin{bmatrix} 1 & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & V_0 \end{bmatrix} \qquad \textbf{and} \qquad \Lambda := V^\top S V. \tag{4}$$

i. **Show that V is orthonormal.**

ii. **Show that Λ is diagonal.**

iii. **Show that $S = V \Lambda V^\top$.**

(HINT: Use block matrix multiplication again.)

We have found a real orthonormal V and real diagonal Λ such that $S = V\Lambda V^\top = V \Lambda V^{-1}$. We have seen in a previous homework that if $A = V\Lambda V^{-1}$, then Λ are the eigenvalues of A, and V are the corresponding eigenvectors. Thus, we've proved the Spectral Theorem for real symmetric matrices by induction!

Contributors:
- Regina Eckert.
- Anant Sahai.
- Nathan Lambert.
- Sally Hui.
- Siddharth Iyer.
- Yu-Yun Dai.
- Sanjit Batra.
- Sidney Buchbinder.
- Gaoyue Zhou.
- Druv Pai.

© UCB EECS 16B, Spring 2024. All Rights Reserved. This may not be publicly shared without explicit permission.