Linear Algebra 4: Orthogonality & Symmetric Matrices and the SVD

MODULE 13: QUADRATIC FORMS AND CONSTRAINED OPTIMIZATION

TOPIC 1: Diagonalization of Symmetric Matrices

Symmetric Matrices

DEFINITION
↳ If matrix $A = A^T$, then A is symmetric.

Example: Which of the following matrices are symmetric?

$A = (2)$ ✓

$B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ ✓
$= B^T$

$C = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \neq C^T = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$. ✗

$D = \begin{pmatrix} 4 & 2 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$ ✗

A very common example: $A^T A$.

For any rectangular matrix A with columns $a_1, \dots, a_n$,

$$A^T A = \begin{pmatrix} - & a_1^T & - \\ - & a_2^T & - \\ \vdots & \vdots & \vdots \\ - & a_n^T & - \end{pmatrix} \begin{pmatrix} | & | & \cdots & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & \cdots & | \end{pmatrix} = \underbrace{\begin{pmatrix} a_1^T a_1 & a_1^T a_2 & \cdots & a_1^T a_n \\ a_2^T a_1 & a_2^T a_2 & \cdots & a_2^T a_n \\ \vdots & \vdots & \ddots & \vdots \\ a_n^T a_1 & a_n^T a_2 & \cdots & a_n^T a_n \end{pmatrix}}_{\text{entries are dot products of columns of } A.}$$

And because $a_i^T a_j = a_i \cdot a_j = a_j \cdot a_i$, $A^T A$ is symmetric.

Example: Suppose A and C are $n \times n$ matrices, $\vec{x} \in \mathbb{R}^n$, and C is symmetric. Which of the following products are equal to a symmetric matrix?

1. AA^T $\quad (AA^T)^T = (A^T)^T A^T = AA^T$ ⇒ symmetric.
2. $\vec{x}\vec{x}^T$ $\quad (\vec{x}\vec{x}^T)^T = (\vec{x}^T)^T (\vec{x})^T = \vec{x}\vec{x}^T$ ⇒ symmetric
3. C^2 $\quad = (C\,C)^T = C^T C^T = CC = C^2$ ⇒ symmetric

Additional Notes on Symmetric Matrices:
* If a matrix is symmetric, the matrix must be square.
* If a matrix is square and diagonal, the matrix must be symmetric.
* Symmetric matrices have other useful properties that we will introduce and take advantage of.

Exercise: Suppose that A is a symmetric $n \times n$ matrix, B is a symmetric $n \times n$ matrix, and $\vec{x}$ is a vector in $\mathbb{R}^n$. Which of the following expressions are symmetric?

1. AB ✗ $\quad (AB)^T = B^T A^T = BA$ (not necessarily equal to AB).
2. $\vec{x}^T A \vec{x}$ ✓ $\quad$ This product is a number. $\forall k \in \mathbb{R}, k^T = k$. ∴ Symmetric.
3. $A + B$ ✓ $\quad (A+B)^T = A^T + B^T = A + B$. ∴ Symmetric.

(True or False): If A is an invertible symmetric matrix, then A^{-1} is also symmetric.

True. Because we know A is invertible, we can say that $A^{-1}A = I$.
But if we take the transpose of both sides of this equation and we obtain

$$(A^{-1}A)^T = I^T = I.$$

And: $I = (A^{-1}A)^T = A^T(A^{-1})^T = A(A^{-1})^T$
So if $A(A^{-1})^T = I$, then the inverse of A is also $(A^{-1})^T$.
However, if A is invertible, then A has a unique inverse. So $(A^{-1})^T = A^{-1}$.

Orthogonal Diagonalization

Motivation → Many algorithms rely on symmetric matrices.
↳ The normal equations use A^TA, which is symmetric.
↳ The SVD (a popular data analysis tool) also uses A^TA.

↳ What properties do symmetric matrices have that we can use to develop and understand algorithms and the results they produce?

Theorem: Symmetric Matrices and their Eigenspaces
If A is a symmetric matrix, with eigenvectors $\vec{v}_1$ and $\vec{v}_2$ corresponding to two distinct eigenvalues, then $\vec{v}_1$ and $\vec{v}_2$ are orthogonal.
More generally, eigenspaces associated to distinct eigenvalues are orthogonal subspaces.

Proof: We can show that eigenvectors $\vec{v}_1$ and $\vec{v}_2$ must be orthogonal, or $\vec{v}_1 \cdot \vec{v}_2 = 0$, for $\lambda_1 \neq \lambda_2$ and symmetric A.

$$\begin{aligned}
\lambda_1 \vec{v}_1 \cdot \vec{v}_2 &= A\vec{v}_1 \cdot \vec{v}_2 && \text{using } A\vec{v}_i = \lambda_i \vec{v}_i \\
&= (A\vec{v}_1)^T \vec{v}_2 && \text{using the definition of the dot product} \\
&= \vec{v}_1^{\,T} A^T \vec{v}_2 && \text{property of transpose of product} \\
&= \vec{v}_1^{\,T} A \vec{v}_2 && \text{given that } A = A^T \\
&= \vec{v}_1 \cdot A\vec{v}_2 \\
&= \vec{v}_1 \cdot \lambda_2 \vec{v}_2 && \text{using } A\vec{v}_i = \lambda_i \vec{v}_i \\
&= \lambda_2 \vec{v}_1 \cdot \vec{v}_2
\end{aligned}$$

But $\lambda_1 \neq \lambda_2$. $\therefore \vec{v}_1 \cdot \vec{v}_2 = 0$.

Orthogonal Diagonalization Example

Diagonalize A using an orthogonal matrix, P. The eigenvalues of A are given.

$$A = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \lambda = -1, 1$$

$$\lambda = -1, \quad A - (-1)I = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \Rightarrow \vec{v}_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$$

$\lambda = 1,\ A - I = \begin{pmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & -1 \end{pmatrix}$. x_2, x_3 is free.

Let $x_2 = 1, x_3 = 0. \Rightarrow \vec{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$

Let $x_2 = 0, x_3 = 1 \Rightarrow \vec{v}_3 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$

$A = PDP^T$

$P = \begin{pmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 0 \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \end{pmatrix},\ D = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.\ P^T = \begin{pmatrix} 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 0 & 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{pmatrix}$

* Recall that if P is an orthogonal $n \times n$ matrix, then $P^{-1} = P^T$
* Symmetric matrices can be diagonalized as $A = PDP^T$. Gram-Schmidt may be needed when eigenvalues are repeated to construct a full set of orthonormal eigenvectors that span $\mathbb{R}^n$.
* What about the converse: if $A = PDP^T$, then is A symmetric?

$$A^T = (PDP^T)^T = P^{TT} D^T P^T = PDP^T = A \Rightarrow A \text{ is symmetric}$$

* Thus, if we can write $A = PDP^T$, then A must be both diagonalizable and symmetric.

<u>The Spectral Theorem</u>

The set of eigenvalues of a matrix are sometimes referred to as the <u>spectrum of A</u>.

The Spectral Theorem
An $n \times n$ symmetric matrix A has the following properties.
↳ All eigenvalues of A are real.
↳ The eigenspaces are mutually orthogonal.
↳ A can be diagonalized as $A = PDP^T$, where D is diagonal and P is orthogonal.

(True or False) Suppose A is a symmetric matrix, and that $\vec{v}_1$ & $\vec{v}_2$ are eigenvectors of A. Then $\vec{v}_1 \cdot \vec{v}_2 = 0$.

<u>False</u>. We are told whether the eigenvectors correspond to distinct eigenvalues. If the eigenvectors did correspond to distinct eigenvalues, then the eigenvectors would be orthogonal because A is symmetric. We would have that $\vec{v}_1 \cdot \vec{v}_2 = 0$.

Exercise:
Suppose A is a 2×2 real matrix that satisfies $A = A^T$. An eigenvector of A corresponding to eigenvalue λ_1 is $\vec{v}_1 = \begin{pmatrix} 5 \\ -4 \end{pmatrix}$.

If the eigenvector for eigenvalue λ_2 is $\vec{v}_2 = \begin{pmatrix} k \\ 1 \end{pmatrix}$ and $\lambda_1 \neq \lambda_2$, what does k need to be equal to?

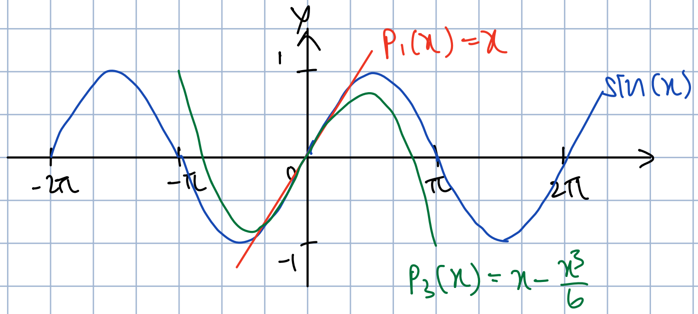
The eigenvectors correspond to distinct eigenvalues of a symmetric matrix, so $\vec{v}_1 \cdot \vec{v}_2 = 0$.

$$5k - 4(1) = 0 \Rightarrow k = \underline{4/5}.$$

The Spectral Decomposition of a Matrix

Motivation: Approximation

Recall from calculus: Taylor expansions and Taylor polynomials, can be used to approximate functions near a point.



Students are not expected to be familiar with Taylor expansions in this course, but, can we use expansions to approximate matrices?

2 cases:
1. A is a symmetric matrix (using an orthogonal diagonalization)
2. A is any real, $m \times n$ matrix (using the SVD)

Spectral Decomposition:

Suppose A can be orthogonally diagonalized as

$$A = PDP^T = \begin{pmatrix} \vec{u}_1 & \cdots & \vec{u}_n \end{pmatrix} \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix} \begin{pmatrix} \vec{u}_1^T \\ \vdots \\ \vec{u}_n^T \end{pmatrix}$$

Then A has the decomposition

$$A = \lambda_1 \vec{u}_1 \vec{u}_1^T + \dots + \lambda_n \vec{u}_n \vec{u}_n^T = \sum_{i=1}^{n} \lambda_i \vec{u}_i \vec{u}_i^T$$

Outline of the Spectral Decomposition Proof

A short explanation on why A has the decomposition

$$A = \lambda_1 \vec{u}_1 \vec{u}_1^T + \dots + \lambda_n \vec{u}_n \vec{u}_n^T = \sum_{i=1}^{n} \lambda_i \vec{u}_i \vec{u}_i^T$$

We assume that we can write $A = PDP^T$. If the columns of D are $d_1, d_2, \dots, d_n$, then, using the definition of matrix multiplication,

$$PD = \begin{pmatrix} Pd_1 & Pd_2 & \cdots & Pd_n \end{pmatrix}$$

Then what is Pd_i equal to?

Recall that a matrix times a vector is a linear combination of the columns of the matrix weighted by the entries of the vector. Column i of PD is

$$Pd_i = P\begin{pmatrix}0\\ \vdots\\ 0\\ \lambda_i\\ 0\\ \vdots\end{pmatrix} = 0 + \cdots + 0 + \lambda_i\vec{u}_i + 0 + \cdots = \lambda_i\vec{u}_i$$

Therefore, the columns of PD are $\lambda_i\vec{u}_i$. We can now simplify our expression for $A = PDP^T$ to a product of two $n\times n$ matrices.

Thus, A can be expressed as follows:

$$A = PDP^T = \begin{pmatrix}\lambda_1\vec{u}_1 & \lambda_2\vec{u}_2 & \cdots & \lambda_n\vec{u}_n\end{pmatrix}\begin{pmatrix}\vec{u}_1^T\\ \vec{u}_2^T\\ \vdots\\ \vec{u}_n^T\end{pmatrix}$$

Using the column-row expansion for the product of two matrices, this becomes

$$A = \lambda_1\vec{u}_1\vec{u}_1^T + \cdots + \lambda_n\vec{u}_n\vec{u}_n^T = \sum_{i=1}^{n}\lambda_i\vec{u}_i\vec{u}_i^T$$

* The row-column expansion for the product of two matrices is a way of defining matrix multiplication.

Example: Construct a spectral decomposition for A.

$$A = \begin{pmatrix}3 & 1\\ 1 & 3\end{pmatrix} = \begin{pmatrix}1/\sqrt{2} & -1/\sqrt{2}\\ 1/\sqrt{2} & 1/\sqrt{2}\end{pmatrix}\begin{pmatrix}4 & 0\\ 0 & 2\end{pmatrix}\begin{pmatrix}1/\sqrt{2} & 1/\sqrt{2}\\ -1/\sqrt{2} & 1/\sqrt{2}\end{pmatrix}$$

$$A = \sum_{i=1}^{2}\lambda_i\vec{u}_i\vec{u}_i^T = 4\begin{pmatrix}1/\sqrt{2}\\ 1/\sqrt{2}\end{pmatrix}\begin{pmatrix}1/\sqrt{2} & 1/\sqrt{2}\end{pmatrix} + 2\begin{pmatrix}-1/\sqrt{2}\\ 1/\sqrt{2}\end{pmatrix}\begin{pmatrix}-1/\sqrt{2} & 1/\sqrt{2}\end{pmatrix}$$

$$= 4\begin{pmatrix}1/2 & 1/2\\ 1/2 & 1/2\end{pmatrix} + 2\begin{pmatrix}1/2 & -1/2\\ -1/2 & 1/2\end{pmatrix}$$ *

Notes:

* Each term in the sum $\lambda_i\vec{u}_i\vec{u}_i^T$ will be an $n\times n$ matrix with rank 1 (each column will be a multiple of $\vec{u}_i$)

* Ordering the eigenvalues from largest to smallest (in absolute value),

$$|\lambda_i| \geq |\lambda_{i+1}|$$

we may be able to truncate the sum

$$A = \sum_{i=1}^{n}\lambda_i\vec{u}_i\vec{u}_i^T$$

to exclude smaller terms. This gives us a way to approximate A.

Exercise:

Suppose A is the symmetric matrix given below:

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

A has the eigenvalues and the unit eigenvectors:

$$\lambda_1 = 1, \quad \vec{v}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \lambda_2 = -1, \quad \vec{v}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

The spectral decomposition of A has the form:

$$A = A_1 + A_2 = \lambda_1 \vec{v}_1 \vec{v}_1^T + \lambda_2 \vec{v}_2 \vec{v}_2^T$$

Suppose:

$$A_1 = \lambda_1 \vec{v}_1 \vec{v}_1^T = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}, \quad A_2 = \lambda_2 \vec{v}_2 \vec{v}_2^T = \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix}$$

What are the values of p_{11} and q_{11}?

$$A_1 = \lambda_1 \vec{v}_1 \vec{v}_1^T = 1\left(\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right)\left(\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \end{pmatrix}\right) = \frac{1}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

$$A_2 = \lambda_2 \vec{v}_2 \vec{v}_2^T = -\left(\frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 1 \end{pmatrix}\right)\left(\frac{1}{\sqrt{2}}\begin{pmatrix} -1 & 1 \end{pmatrix}\right) = -\frac{1}{2}\begin{pmatrix} -1 \\ 1 \end{pmatrix}\begin{pmatrix} -1 & 1 \end{pmatrix} = -\frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$$

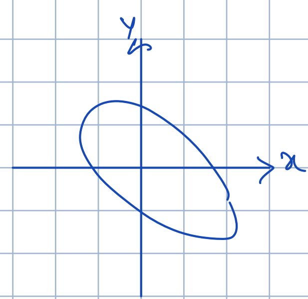
Note that we must have that $p_{11} + q_{11} = 0$ because the entry of A that is in the first row, first column, is zero and

$$\lambda_2 \vec{v}_2 \vec{v}_2^T = \frac{1}{2}\begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}.$$

TOPIC 2: Quadratic Forms

Motivation:

The equation $5x^2 + 4xy + 8y^2 = 4$ represents an ellipse.



An ellipse is an example of a quadratic form. If we can represent quadratic forms using a symmetric matrix, we can take advantage of the properties that symmetric matrices have to them.

DEFINITION

↳ A quadratic form is a function $Q: \mathbb{R}^n \to \mathbb{R}$, given by

$$Q(\vec{x}) = \vec{x}^T A \vec{x} = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix}\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

Matrix A is $n \times n$ and symmetric.

In the above, $\vec{x}$ is a vector of variables.

Quadratic Forms and Symmetric Matrices

Compute the quadratic form $Q = \vec{x}^T A \vec{x}$, using $\vec{x} = \begin{pmatrix} x \\ y \end{pmatrix}$.

(a) $A = \begin{pmatrix} 4 & 0 \\ 0 & 3 \end{pmatrix}$ $\quad Q = (x \ \ y)\begin{pmatrix} 4 & 0 \\ 0 & 3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = (x \ \ y)\begin{pmatrix} 4x \\ 3y \end{pmatrix} = 4x^2 + 3y^2$

(b) $A = \begin{pmatrix} 4 & 1 \\ 1 & -3 \end{pmatrix}$ $\quad Q = (x \ \ y)\begin{pmatrix} 4 & 1 \\ 1 & -3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = (x \ \ y)\begin{pmatrix} 4x+y \\ x-3y \end{pmatrix} = x(4x+y) + y(x-3y)$

$$= 4x^2 + xy + xy - 3y^2$$
$$= 4x^2 - 3y^2 + \underline{2xy} \ *$$

The $2xy$ term is a <u>cross-product</u> term because it contains both variables.

A Quadratic Form

Express $Q = x^2 - 6xy + 9y^2$ in the form of $Q = \vec{x}^T A \vec{x}$, where $\vec{x} \in \mathbb{R}^2$ and $A = A^T$.

Placing coefficients of x^2 and y^2 on main diagonal, and dividing coefficient of xy by 2, we obtain

$$x^2 - 6xy + 9y^2 = (x \ \ y)\begin{pmatrix} 1 & -3 \\ -3 & 9 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$$

We can verify this result by multiplying $\vec{x}^T A \vec{x}$.

Example: Quadratic Form in Three Variables

Write Q in the form $\vec{x}^T A \vec{x}$ for $\vec{x} \in \mathbb{R}^3$.

$$Q(\vec{x}) = 5x_1^2 - x_2^2 + 3x_3^2 + 6x_1x_3 - 12x_2x_3$$

Note that we can write Q as

$$Q = 5x_1^2 - x_2^2 + 3x_3^2 + 0x_1x_2 + 6x_1x_3 - 12x_2x_3$$

Taking a similar approach to the previous exercise, we obtain

$$Q = (x_1 \ \ x_2 \ \ x_3)\begin{pmatrix} 5 & 0 & 3 \\ 0 & -1 & -6 \\ 3 & -6 & 3 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

Again, we can verify this result by multiplying $\vec{x}^T A \vec{x}$.

Notes on Quadratic Forms:

* One of the reasons we are interested in quadratic forms relates to their use in describing linear transforms.
* Consider the transform $\vec{x} \Rightarrow A\vec{x}$. The squared length of the vector $A\vec{x}$ is a quadratic form:

$$\|A\vec{x}\|^2 = (A\vec{x}) \cdot (A\vec{x}) = \vec{x}^T A^T A \vec{x}$$

* Note that $A^T A$ is symmetric.
* In other words, we can use symmetric matrices and their properties to, for example, characterize linear transforms.

Exercise:
Consider the quadratic form: $Q = 4x_1^2 + 6x_1x_2 + 8x_2^2$.

We can represent Q using a symmetric matrix, A:

$$Q = \vec{x}^T A \vec{x}, \quad \vec{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

What values should be used for a_{11}, a_{12} and a_{22}?

$$Q = (x_1 \;\; x_2) \begin{pmatrix} 4 & 3 \\ 3 & 8 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

Change of Variable

Motivation: Does this inequality hold for all x, y?

$$x^2 - 6xy + 9y^2 \geq 0$$

To answer this question, consider the following:
* The polynomial $Q = x^2 - 6xy + 9y^2$ is an example of a quadratic form.
* The $-6xy$ term is a cross-product term because it contains both variables.
* The cross-product term makes problems like this one more complicated.

Given $Q = \vec{x}^T A \vec{x}$, where $\vec{x} \in \mathbb{R}^n$ is a variable vector and A is a real $n \times n$ symmetric matrix. Then,

$$A = PDP^T$$

where P is an $n \times n$ orthogonal matrix. A change of variable can be represented as

$$\vec{x} = P\vec{y} \quad \text{or} \quad \vec{y} = P^{-1}\vec{x}.$$

With this change of variable, the quadratic form $\vec{x}^T A \vec{x}$ becomes:

$$\begin{aligned} Q = \vec{x}^T A \vec{x} &= (P\vec{y})^T A (P\vec{y}) \\ &= \vec{y}^T P^T A P \vec{y} \\ &= \vec{y}^T D \vec{y}, \text{ using } A = PDP^T. \end{aligned}$$

Thus, Q is expressed without cross-product terms.

proof up there!

Principle Axes Theorem
If A is a symmetric matrix then there exists an orthogonal change of variable $\vec{x} = P\vec{y}$ that transforms $\vec{x}^T A \vec{x}$ to $\vec{y}^T D \vec{y}$ with no cross-product terms.

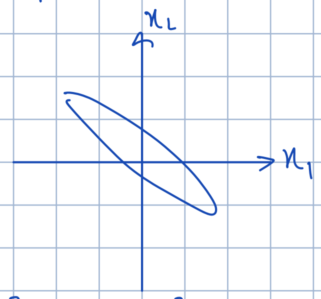
Example:
Compute the quadratic form $Q = \vec{x}^T A \vec{x}$ for $A = \begin{pmatrix} 5 & 2 \\ 2 & 8 \end{pmatrix}$, and identify a change of variable that removes the cross-product term. The eigenvalues and eigenvectors of A are given below:

$$\lambda_1 = 9, \quad \lambda_2 = 4, \quad \vec{v}_1 = \begin{pmatrix} 2 \\ -1 \end{pmatrix}, \quad \vec{v}_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

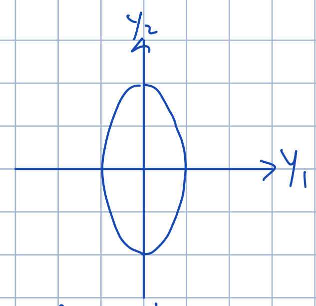
Our change of variable is $\vec{x} = P\vec{y}$. $P = \frac{1}{\sqrt{5}}\begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}$

Using this change of variable, $Q = \vec{x}^T A \vec{x} = \vec{y}^T D \vec{y} = 9y_1^2 + 4y_2^2$

If, for example, we set $Q = 1$, we obtain 2 curves:



$Q = 5x^2 + 4xy + 8y^2 = 1$



$Q = 9y_1^2 + 4y_2^2 = 1.$

Using the change of variable, we can more easily
→ identify points on the ellipse that are closest\furthest from the origin
→ determine whether Q can take on negative\positive values.

Return to motivating question: does this inequality hold for all x, y?

$$x^2 - 6xy + 9y^2 \geq 0$$

Let $Q = x^2 - 6xy + 9y^2$.

$$Q = x^2 - 6xy + 9y^2 = \vec{x}^T \begin{pmatrix} 1 & -3 \\ -3 & 9 \end{pmatrix} \vec{x}, \quad A = \begin{pmatrix} 1 & -3 \\ -3 & 9 \end{pmatrix}$$

By inspection, A is singular. $\Rightarrow \lambda_1 = 0$.
The sum of the eigenvalues is the trace $\Rightarrow \lambda_2 = 10$

$$\Rightarrow Q = \vec{y}^T D \vec{y} = 0y_1^2 + 10y_2^2.$$

Q can be zero but is never negative. ∴ The inequality holds.

Exercise:
In this exercise, our goal is to create a change of variable that transforms the quadratic form below into another quadratic form that has no cross-product terms.

$$Q = x_1^2 - 12x_1x_2 - 4x_2^2$$

First we need to express Q in the form $Q = \vec{x}^T A \vec{x}$ for symmetric A.

$A = \begin{pmatrix} 1 & -6 \\ -6 & -4 \end{pmatrix}$. The coefficient of the cross-product term divided by 2 is used for the off-diagonal entries of A.

Next we need to orthogonally diagonalize A. One of the eigenvalues of A is $\lambda_1 = -8$. What is the other eigenvalue of A equal to?

We can obtain the eigenvalues using the characteristic polynomial. But it is less work to use the idea that the sum of the entries along the main diagonal is equal to the sum of the eigenvalues.

$$\lambda_1 + \lambda_2 = 1 + (-4) = -3.$$

Since $\lambda_1 = -8$, $-8 + \lambda_2 = -3 \Rightarrow \lambda_2 = 5$ *

An eigenvector associated with $\lambda_1 = -8$ is $\vec{v}_1 = \begin{pmatrix} c \\ 3 \end{pmatrix}$. What is c equal to?

$\lambda_1 = -8$, $A - (-8)I = \begin{pmatrix} 9 & -6 \\ -6 & 4 \end{pmatrix}$. We do not need the second row when looking for the eigenvector of a 2x2 matrix.

$\therefore c = 2$. $\quad \vec{v}_1 = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$.

Next, we use the eigenvectors of A to construct the change of variable:

$$\vec{x} = P\vec{y}, \quad P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}$$

When constructing the change of variable, don't forget that we need unit vectors: the column of P should be orthonormal.

$\lambda_2 = 5$, $A - 5I = \begin{pmatrix} -4 & -6 \\ -6 & -9 \end{pmatrix}$. $\quad \therefore \vec{v}_2 = \begin{pmatrix} 3 \\ -2 \end{pmatrix}$

$$\vec{x} = P\vec{y}, \quad P = \frac{1}{\sqrt{13}} \begin{pmatrix} 2 & 3 \\ 3 & -2 \end{pmatrix}$$

Finally, we can write our quadratic form below into another quadratic form that has no cross-product terms:

$$Q = -8y_1^2 + c_2 y_2^2$$

What number should we use for c_2? $\underline{\underline{5}}$ (the other eigenvalue of A).

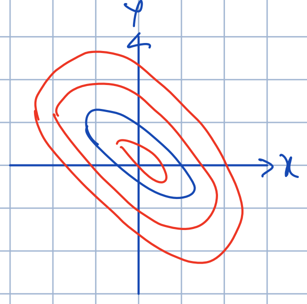
Quadratic Surfaces

Motivation: We will want to solve optimization problems of the form

$$\text{minimize } Q = \|A\vec{x}\|^2, \text{ subject to constraints on } \vec{x}$$

* To help understand this problem it is helpful to have a geometric interpretation of $Q = \|A\vec{x}\|^2$.
* Recall that $Q = \|A\vec{x}\|^2 = \vec{x}^T A^T A \vec{x}$ is a quadratic function, and A^TA is symmetric.
* Terminology that describes the shape of Q will also help us discuss the optimization problem.

Example of a curve in $\mathbb{R}^2$: $\quad Q = \vec{x}^T \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \vec{x} = 2x^2 + 2y^2 + 2xy = 1.$



As we vary the value of c, the size of our ellipse will change.
↳ When we let c vary continuously, we generate a surface in $\mathbb{R}^3$

Surfaces

Suppose $z = Q(\vec{x}) = \vec{x}^T A \vec{x}$, where $A \in \mathbb{R}^{2\times 2}$ is symmetric. Then the set of points that satisfy

$$z = \vec{x}^T A \vec{x}$$

defines a surface in $\mathbb{R}^3$.

Classifying Quadratic Forms

DEFINITION

↳ A quadratic form Q is → positive definite if $Q > 0$ $\forall \vec{x} \neq \vec{0}$
↳ negative definite if $Q < 0$ $\forall \vec{x} \neq \vec{0}$
↳ positive semidefinite if $Q \geq 0$ $\forall \vec{x}$.
↳ negative semidefinite if $Q \leq 0$ $\forall \vec{x}$
↳ indefinite if Q takes on positive and negative values for $\vec{x} \neq \vec{0}$.

Quadratic Forms and Eigenvalues

Theorem:

If A is a symmetric matrix with eigenvalues λ_i, then $Q = \vec{x}^T A \vec{x}$ is
↳ positive definite when all eigenvalues are positive
↳ negative definite when all eigenvalues are negative
↳ indefinite when at least one eigenvalue is negative and at least one eigenvalue is positive

Proof: Assuming A symmetric, we can write $A = PDP^T$ and set $\vec{x} = P\vec{y}$, so

$$\begin{aligned} Q = \vec{x}^T A \vec{x} &= (P\vec{y})^T A (P\vec{y}) \\ &= \vec{y}^T P^T A P \vec{y} \\ &= \vec{y}^T D \vec{y}, \text{ using } A = PDP^T \\ &= \textstyle\sum \lambda_i y_i^2, \text{ because } D \text{ is diagonal} \end{aligned}$$

Because y_i^2 is always non-negative, for $\vec{y} \neq 0$,
* Q is positive when, $\forall i$, $\lambda_i > 0$ ⇒ Q is positive definite
* Q is negative when, $\forall i$, $\lambda_i < 0$ ⇒ Q is negative definite
* indefinite when λ_i has both negative and positive values

Quadratic Forms and Eigenvalues Example

$$Q = 4x^2 + 2xy - 2y^2 = \vec{x}^T A \vec{x} \Rightarrow A = \begin{pmatrix} 4 & 1 \\ 1 & -2 \end{pmatrix}, \quad \lambda = 1 \pm \sqrt{10}$$

eigenvalues are both positive and negative ⇒ Q is indefinite

(True or False) The quadratic form $Q = -x^2 - 2xy - y^2$ is negative definite.

$$Q = \vec{x}^T A \vec{x} = \vec{x}^T \begin{pmatrix} -1 & -1 \\ -1 & -1 \end{pmatrix} \vec{x}, \quad \vec{x} = \begin{pmatrix} x \\ y \end{pmatrix}$$

The matrix A has zero eigenvalue, so there are many locations where Q can be zero. Any vector in the null space of A corresponds to a point where $Q = 0$.

For example, the vector $\begin{pmatrix}1\\-1\end{pmatrix}$ is in the null space of A. At the point $(x,y)=(1,-1)$, we have that

$$Q = -(1)^2 - 2(1)(-1) - (-1)^2 = 0.$$

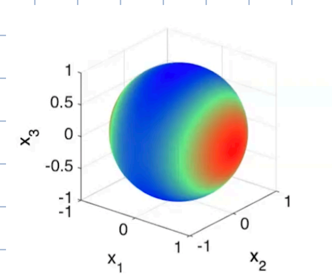
The quadratic form is zero away from the origin, so the form is negative semi-definite, but not negative-definite. $\therefore$ FALSE

TOPIC 3: Constrained Optimization

Temperature on a Sphere

The surface of a unit sphere in $\mathbb{R}^3$ is given by

$$1 = x_1^2 + x_2^2 + x_3^2 = \|x\|^2$$



Q is a quantity (eg. temperature) that we want to optimize

$$Q(\vec{x}) = 9x_1^2 + 4x_2^2 + 3x_3^2$$

Identify the largest and smallest values of Q on the surface of the sphere, and where they are located.

Solution: Largest value of Q on a Sphere

We will identify the largest value of Q on the sphere first.

$$Q = \vec{x}^T \begin{pmatrix} 9 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 3 \end{pmatrix} \vec{x} = 9x_1^2 + 4x_2^2 + 3x_3^2 \leq 9x_1^2 + 9x_2^2 + 9x_3^2$$
$$= 9(x_1^2 + x_2^2 + x_3^2)$$
$$= 9\|\vec{x}\|^2 = 9.$$

We are only considering points on the surface of the sphere: $\|\vec{x}\|^2 = 1$.
Thus,

$$\max\{Q(x) : \|\vec{x}\| = 1\} = 9, \text{ and max occurs at } \vec{x} = \begin{pmatrix} \pm 1 \\ 0 \\ 0 \end{pmatrix}$$

A similar analysis yields

$$\min\{Q(x) : \|\vec{x}\| = 1\} = 3, \text{ and min occurs at } \vec{x} = \begin{pmatrix} 0 \\ 0 \\ \pm 1 \end{pmatrix}.$$

Notice that the minimum and maximum values of Q were the eigenvalues of A, and the corresponding eigenvectors gave their locations.

A Constrained Optimization Problem

Suppose we wish to find the maximum or minimum values of

$$Q(\vec{x}) = \vec{x}^T A \vec{x}, \quad \vec{x} \in \mathbb{R}^n, \quad A \in \mathbb{R}^{n \times n}$$

subject to $\|\vec{x}\| = 1$. That is, we want to find

$$m = \min\{Q(\vec{x}) : \|\vec{x}\| = 1\}$$
$$M = \max\{Q(\vec{x}) : \|\vec{x}\| = 1\}$$

This is an example of a constrained optimization problem. Note that we may also want to know where these extreme values are obtained.

Constrained Optimization and Eigenvalues

Theorem:
If $Q = \vec{x}^T A \vec{x}$, A is a real $n \times n$ symmetric matrix, with eigenvalues $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_n$ and associated normalized eigenvectors $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n$.

Then, subject to the constraint $\|\vec{x}\| = 1$,
* the maximum value of $Q(\vec{x}) = \lambda_1$, attained at $\vec{x} = \pm\vec{u}_1$.
* the minimum value of $Q(\vec{x}) = \lambda_n$, attained at $\vec{x} = \pm\vec{u}_n$.

Proof for maximum value of Q:

Assume λ_1 is the largest eigenvalue with corresponding unit eigenvector $\vec{u}_1$

$$
\begin{aligned}
Q = \vec{x}^T A \vec{x} &= \vec{y}^T D \vec{y}, && \text{using } A = PDP^T,\ \vec{x} = P\vec{y} \\
&= \sum \lambda_i y_i^2, && \text{because } D \text{ is diagonal} \\
&\leq \sum \lambda_1 y_i^2, && \text{because } \lambda_1 \text{ is the largest eigenvalue} \\
&= \lambda_1 \sum y_i^2 \\
&= \lambda_1 \|\vec{y}\|^2 \\
&= \lambda_1, && \text{because } \|\vec{y}\|^2 = 1.
\end{aligned}
$$

So the maximum value of Q is at most λ_1. And $Q = \lambda_1$ at $\pm\vec{u}_1$ because

$$Q(\pm\vec{u}_1) = \vec{u}_1^T A \vec{u}_1 = \vec{u}_1^T(\lambda_1 \vec{u}_1) = \lambda_1.$$

Exercise:
A unit vector that gives the location of the maximum value of $Q(\vec{x}) = x_1^2 - 2x_2^2$ subject to $\vec{x}^T\vec{x} = 1$, $\vec{x} \in \mathbb{R}^2$, is

$$\vec{x} = \begin{pmatrix} 1 \\ k \end{pmatrix}$$

What is k equal to?

The matrix of the quadratic form is $A = \begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix}$. Eigenvalues of matrix $A = 1, -2$.

$\Rightarrow$ maximum value of Q subject to $\vec{x}^T\vec{x} = 1$ is 1.

Applying the usual approach for identifying an eigenvector:

$$A - \lambda I = A - I = \begin{pmatrix} 0 & 0 \\ 0 & -3 \end{pmatrix}.$$

A unit eigenvector in the null space of $A - \lambda I$ is $\vec{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. $\therefore k = 0$.

Constrained Optimization with a Repeated Eigenvalue

Example:
Calculate the maximum and minimum values of $Q(\vec{x}) = \vec{x}^T A \vec{x}$, $\vec{x} \in \mathbb{R}^3$, subject to $\|\vec{x}\| = 1$, and identify points where these values are obtained.

$$Q(\vec{x}) = x_1^2 + 2x_2x_3$$

For $Q(\vec{x}) = x_1^2 + 2x_2x_3$, we have $Q(\vec{x}) = \vec{x}^T A \vec{x}$, $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$

By inspection, A has eigenvalues ± 1 (don't forget that an eigenvalue, λ, is a number that makes $A - \lambda I$ singular). For $\lambda = 1$,

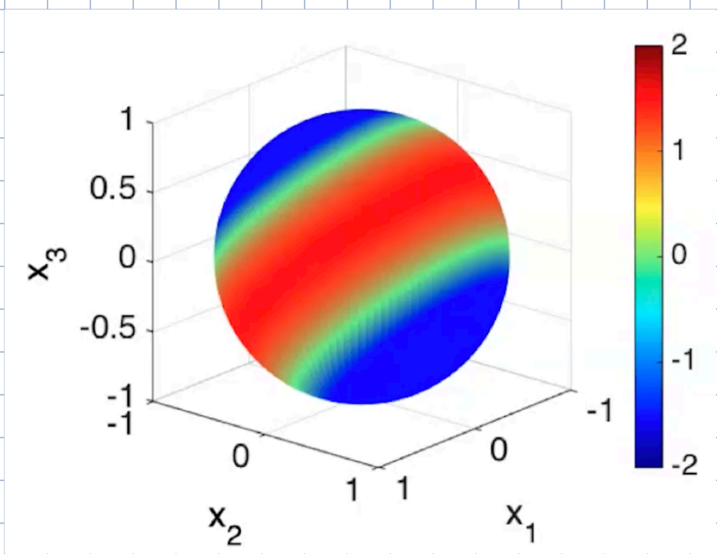
$$A - I = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 1 & -1 \end{pmatrix} \Rightarrow \vec{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \vec{v}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

Because A is symmetric, the eigenvector for eigenvalue $\lambda = -1$ must be orthogonal to $\vec{v}_1$ and $\vec{v}_2$. So by inspection,

$$\vec{v}_3 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$$

Therefore, the minimum value of Q is -1, and is obtained at $\pm\vec{v}_3$. The maximum value of Q is $+1$, which is obtained at any unit vector in the span of $\vec{v}_1$ and $\vec{v}_2$.

The image shown below is the unit sphere whose surface is colored according to the quadratic from the previous example:



Notice the agreement between our solution and the image.

Exercise:
Suppose Q is the quadratic form below.

$$Q = \vec{x}^T A \vec{x}, \quad \vec{x} = \begin{pmatrix} 5 & 2 & 1 \\ 2 & 8 & 2 \\ 1 & 2 & 5 \end{pmatrix}$$

The minimum value of Q subject to $\|\vec{x}\| = 1$ is $Q = 4$.
An eigenvector of A associated with eigenvalue $\lambda = 4$ is

$$\vec{v} = \begin{pmatrix} 0 \\ 1 \\ c \end{pmatrix}$$

What is c equal to?

$$\lambda = 4, \quad A - \lambda I = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

The entries of the eigenvector associated with $\lambda = 4$ must therefore satisfy the equation $x_1 + 2x_2 + x_3 = 0$. That is, if the eigenvector is

$$\vec{v} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix},$$

then we must have that $x_1 + 2x_2 + x_3 = 0$.

Given $\vec{v} = \begin{pmatrix} 0 \\ 1 \\ c \end{pmatrix}$, $x_1 = 0$, $x_2 = 1$. $0 + 2(1) + x_3 = 0 \Rightarrow \underline{\underline{x_3 = -2}}$.

The minimum value of Q, subject to $\|\vec{x}\| = 1$, is obtained at $\vec{x}_0$, where:

$$\vec{x}_0 = \begin{pmatrix} 0 \\ k_0 \\ k_1 \end{pmatrix}$$

If $\vec{v}$ is parallel to $\vec{x}_0$ and $k_0 > 0$, what must k_1 be equal to? Don't forget that we need $\|\vec{x}\| = 1$.

To obtain the location of the minimum, we can take our eigenvector and divide by its length.

$$\vec{x}_0 = \frac{1}{\sqrt{5}} \begin{pmatrix} 0 \\ 1 \\ -2 \end{pmatrix}. \quad \therefore k_1 = -2/\sqrt{5}$$

Orthogonality Constraints

Theorem:
Suppose $Q = \vec{x}^T A \vec{x}$, where $A \in \mathbb{R}^{n \times n}$ is symmetric and has eigenvalues $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_n$ and associated eigenvectors $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n$.

Subject to the constraints $\|\vec{x}\| = 1$ and $\vec{x} \cdot \vec{u}_1 = 0$,
* the maximum value of $Q(\vec{x}) = \lambda_2$, attained at $\vec{x} = \vec{u}_2$
* the minimum value of $Q(\vec{x}) = \lambda_n$, attained at $\vec{x} = \vec{u}_n$

Proof: The proof of this theorem uses a similar approach to the theorem that gives the maximum of Q subject to $\|\vec{x}\| = 1$.

* We would use a change of variable, and express Q using a diagonal matrix and an orthonormal basis for $\mathbb{R}^n$.
* In fact, we could use this approach to identify maximum values with additional orthogonality constraints. This would go beyond the scope of what we need.

Example:

Calculate the maximum value of $Q(\vec{x}) = \vec{x}^T A \vec{x}$, $\vec{x} \in \mathbb{R}^3$, subject to $\|\vec{x}\| = 1$ and to $\vec{x} \cdot \vec{u}_1 = 0$, and identify a point where this maximum is obtained.

$$Q(\vec{x}) = x_1^2 + 2x_2x_3, \quad \vec{u}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$$Q = \vec{x}^T A \vec{x} = \vec{x}^T \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \vec{x}, \quad \lambda = \pm 1$$

$$\text{If } \lambda = 1, \quad A - I = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 1 & -1 \end{pmatrix}, \quad \vec{u}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \vec{u}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

$\Rightarrow$ maximum value of Q is $+1$, obtained at $\pm\vec{u}_2$.

Exercise:

Suppose Q is the quadratic form below.

$$Q = \vec{x}^T A \vec{x}, \quad A = \begin{pmatrix} -2 & -2 & -1 \\ -2 & -1 & -2 \\ -1 & -2 & -2 \end{pmatrix}, \quad \vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

The eigenvalues of A are -5, -1 and 1. Thus, the maximum value of Q subject to $\|\vec{x}\| = 1$ is $Q_1 = \lambda_1 = 1$. And this maximum is obtained at the corresponding unit eigenvector, $\vec{v}_1$.

Suppose the maximum value of Q, subject to $\|\vec{x}\| = 1$ and $\vec{x} \cdot \vec{v}_1 = 0$ is Q_2. What is Q_2 equal to?

The maximum value of Q subject to the two constraint is the second largest eigenvalue, -1.
$\therefore Q_2 = -1$.

Suppose $Q = Q_2$ is obtained at the point $(k_1, 0, k_2)$, and $k_2 > 0$. What must k_1 be equal to?

There are 2 locations where $Q = Q_2$, and they are given by the eigenvector for the eigenvalue $\lambda = -1$.

$$\lambda = -1, \quad A - (-1)I = \begin{pmatrix} -1 & -2 & -1 \\ -2 & 0 & -2 \\ -1 & -2 & -1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

The eigenvector $\vec{v}_2$ must therefore satisfy $x_2 = 0$ and equation $x_1 + x_3 = 0$.

An eigenvector corresponding to $\lambda = -1$ is $\vec{v}_2 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$.

The eigenvector must also have unit length. The unit eigenvector that we need is:

$$\vec{v}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}. \qquad \therefore k_1 = -1/\sqrt{2}.$$