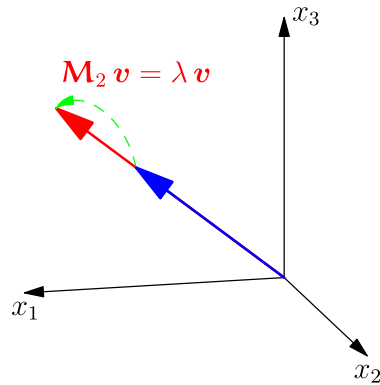
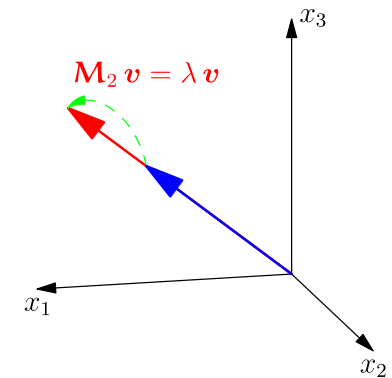


Eigensystems



Eigenvectors, Orthogonal Matrices, Eigenvector Decomposition, Rank

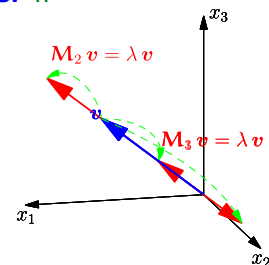
1. **Eigenvectors**
2. Orthogonal Matrices
3. Eigen Decomposition
4. Low Rank Approximation



Eigenvector equation

- Eigen-systems help us to understand mappings
- A vector v is said to be an **eigenvector** if

$$Mv = \lambda v$$



- M is square (i.e. $n \times n$)
- Where the number λ is the **eigenvalue**
- Eigenvalues play a fundamental role in understanding operators

Symmetric Matrices

- If M is an $n \times n$ **symmetric** matrix then it has n real orthogonal eigenvectors with real eigenvalues
- We denote the i^{th} eigenvector by v_i and the corresponding eigenvalue by λ_i so that

$$Mv_i = \lambda_i v_i$$

- Orthogonal means that if $i \neq j$ then

$$v_i^T v_j = 0$$

- (We can always normalise eigenvectors if we want)

Proof of Orthogonality

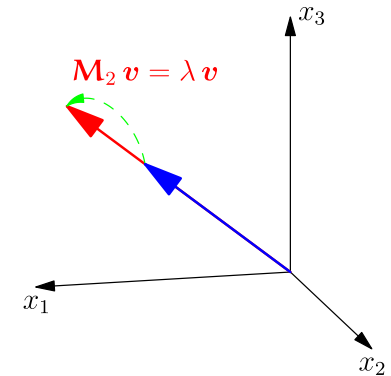
- $(\mathbf{M}v_i = \lambda_i v_i)^\top$ implies $v_i^\top \mathbf{M}^\top = \lambda_i v_i^\top$
- When \mathbf{M} is symmetric then $\mathbf{M}v_i = \lambda_i v_i \Rightarrow v_i^\top \mathbf{M} = \lambda_i v_i^\top$
- Consider two eigenvectors v_i and v_j of \mathbf{M}

$$\begin{aligned} v_i^\top \mathbf{M} v_j &= (v_i^\top \mathbf{M}) v_j = \lambda_i v_i^\top v_j \\ &= v_i^\top (\mathbf{M} v_j) = \lambda_j v_i^\top v_j \end{aligned}$$

- So either $\lambda_i = \lambda_j$ or $v_i^\top v_j = 0$
- If $\lambda_i = \lambda_j$ then any linear combination of v_i and v_j is an eigenvector $(\mathbf{M}(av_i + bv_j) = \lambda_i(av_i + bv_j))$. So I can choose two eigenvectors that are orthogonal to each other.

Outline

1. Eigenvectors
2. Orthogonal Matrices
3. Eigen Decomposition
4. Low Rank Approximation



Orthogonal Matrices

- We can construct an **orthogonal** matrix \mathbf{V} from the eigenvectors

$$\mathbf{V} = (v_1, v_2, \dots, v_n)$$

- Matrix \mathbf{V} is an $n \times n$ matrix
- Because of the orthogonality of the vectors v_i

$$\mathbf{V}^\top \mathbf{V} = \begin{pmatrix} v_1^\top v_1 & v_1^\top v_2 & \dots & v_1^\top v_n \\ v_2^\top v_1 & v_2^\top v_2 & \dots & v_2^\top v_n \\ \vdots & \vdots & \ddots & \vdots \\ v_n^\top v_1 & v_n^\top v_2 & \dots & v_n^\top v_n \end{pmatrix} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} = \mathbf{I}$$

The Other Way Around

- We have shown that $\mathbf{V}^\top \mathbf{V} = \mathbf{I}$
- Thus multiply both sides on the left by \mathbf{V}

$$\mathbf{V} \mathbf{V}^\top \mathbf{V} = \mathbf{V} \mathbf{I}$$

- \mathbf{V} will have an inverse, \mathbf{V}^{-1} , such that $\mathbf{V} \mathbf{V}^{-1} = \mathbf{I}$
- Multiplying the equation on the right by \mathbf{V}^{-1}

$$\begin{aligned} (\mathbf{V} \mathbf{V}^\top) \mathbf{V} \mathbf{V}^{-1} &= \mathbf{V} \mathbf{V}^{-1} \mathbf{I} \\ \mathbf{V} \mathbf{V}^\top &= \mathbf{I} \end{aligned}$$

- Note that, $\mathbf{V}^{-1} = \mathbf{V}^\top$ (definition of orthogonal matrix)

Invertible Matrices

- A matrix, \mathbf{M} , will be singular (uninvertible) if there exists a vector $\mathbf{x} (\neq \mathbf{0})$ such that

$$\mathbf{M}\mathbf{x} = \mathbf{0}$$

- Now if there exists such a vector such that $\mathbf{V}\mathbf{x} = \mathbf{0}$ then multiply by \mathbf{V}^T we get

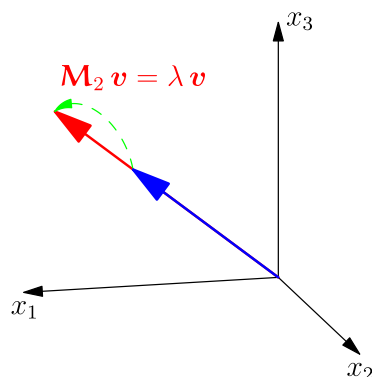
$$\begin{aligned}\mathbf{V}^T\mathbf{V}\mathbf{x} &= \mathbf{V}^T\mathbf{0} \\ \mathbf{x} &= \mathbf{0}\end{aligned}$$

since $\mathbf{V}^T\mathbf{V} = \mathbf{I}$

- Thus \mathbf{V} is invertible

Outline

1. Eigenvectors
2. Orthogonal Matrices
3. **Eigen Decomposition**
4. Low Rank Approximation



Rotations

- Orthogonal matrices satisfy $\mathbf{V}^T\mathbf{V} = \mathbf{V}\mathbf{V}^T = \mathbf{I}$
- As a consequent they define rotations (and possibly a reflection)
- Consider a vector \mathbf{x} and $\mathbf{x}' = \mathbf{V}\mathbf{x}$, now

$$\|\mathbf{x}'\|_2^2 = \mathbf{x}'^T\mathbf{x}' = (\mathbf{V}\mathbf{x})^T(\mathbf{V}\mathbf{x}) = \mathbf{x}^T\mathbf{V}^T\mathbf{V}\mathbf{x} = \mathbf{x}^T\mathbf{x} = \|\mathbf{x}\|_2^2$$

- Similarly if additionally $\mathbf{y}' = \mathbf{V}\mathbf{y}$ then

$$\langle \mathbf{x}', \mathbf{y}' \rangle = (\mathbf{V}\mathbf{x})^T(\mathbf{V}\mathbf{y}) = \mathbf{x}^T\mathbf{V}^T\mathbf{V}\mathbf{y} = \mathbf{x}^T\mathbf{y} = \langle \mathbf{x}, \mathbf{y} \rangle = \|\mathbf{x}\|_2\|\mathbf{y}\|_2\cos(\theta)$$

- Rotations and reflections preserve lengths and angles

Matrix Decomposition

- Taking the matrix of eigenvectors, \mathbf{V} , then

$$\mathbf{M}\mathbf{V} = \mathbf{M}(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) = (\lambda_1\mathbf{v}_1, \lambda_2\mathbf{v}_2, \dots, \lambda_n\mathbf{v}_n) = \mathbf{V}\mathbf{\Lambda}$$

- where $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}$

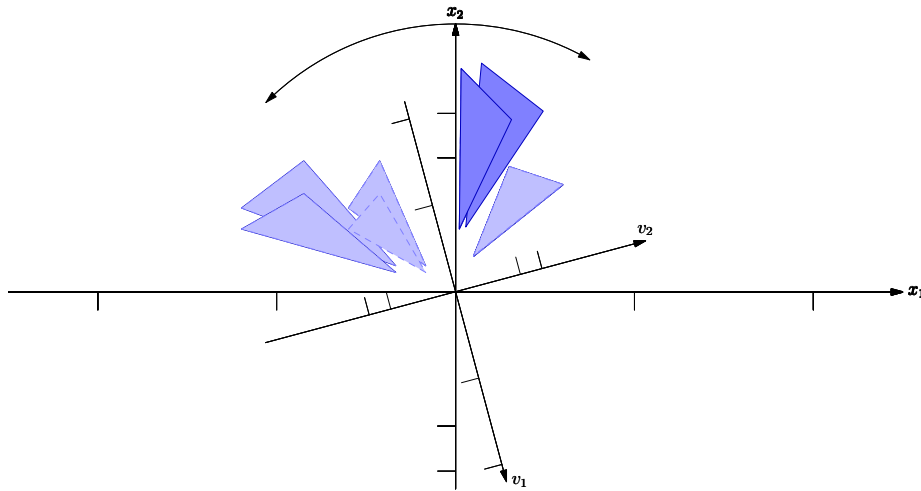
- Now

$$\mathbf{M} = \mathbf{M}\mathbf{V}\mathbf{V}^T = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$$

- Very important *similarity transform*

Mappings by Symmetric Matrices

$$\mathbf{M} = \begin{pmatrix} 0.83 & -0.31 \\ -0.31 & 1.9 \end{pmatrix} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T = \begin{pmatrix} \cos(-75) & \sin(-75) \\ -\sin(-75) & \cos(-75) \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 0.75 \end{pmatrix} \begin{pmatrix} \cos(75) & \sin(75) \\ -\sin(75) & \cos(75) \end{pmatrix}$$



Inverses

- For any square matrix

$$\mathbf{M} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$$

$$\mathbf{M}^{-1} = \mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{V}^T$$

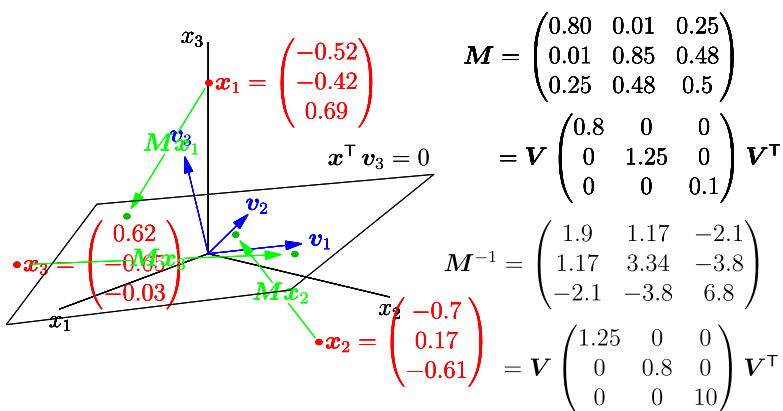
- Where $\mathbf{\Lambda}^{-1} = \text{diag}(\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \dots, \frac{1}{\lambda_n}) = \begin{pmatrix} \frac{1}{\lambda_1} & 0 & \dots & 0 \\ 0 & \frac{1}{\lambda_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\lambda_n} \end{pmatrix}$

- Since

$$\begin{aligned} \mathbf{M}\mathbf{M}^{-1} &= (\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T)(\mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{V}^T) = \mathbf{V}\mathbf{\Lambda}(\mathbf{V}^T\mathbf{V})\mathbf{\Lambda}^{-1}\mathbf{V}^T \\ &= \mathbf{V}\mathbf{\Lambda}\mathbf{\Lambda}^{-1}\mathbf{V}^T = \mathbf{V}\mathbf{V}^T = \mathbf{I} \end{aligned}$$

- I.e, Small eigenvalues become large eigenvalues and visa versa

Ill-Conditioning Again



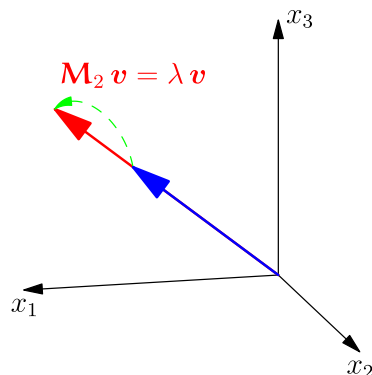
Condition Number

- Taking matrix inverses can be inherently unstable
- Any small error can be amplified by taking the inverse
- The stability of the inverse depends on the ratio of smallest eigenvalue to the largest eigenvalue (i.e. the biggest possible amplification compared to the smallest)
- Note that the Hilbert-norm of a matrix is the absolute value of the largest eigenvalue
- The condition number is given by

$$\|\mathbf{M}\|_H \times \|\mathbf{M}^{-1}\|_H = \frac{|\lambda_{\max}|}{|\lambda_{\min}|}$$

- Large condition number implies very ill-conditioned

1. Eigenvectors
2. Orthogonal Matrices
3. Eigen Decomposition
4. **Low Rank Approximation**



- The rank of a matrix, \mathbf{M} , is the number of non-zero eigenvalues
- The space spanned by the eigenvectors v_a , v_b , etc. with zero eigenvalue forms a **null space**
- Any vector in the null space will get projected to the zero vector

$$\mathbf{M}(av_a + bv_b + \dots) = \mathbf{0}$$

- A square matrix is said to be **rank deficient** if it has any eigenvectors with eigenvalue equal to 0
- This happens when the columns of the matrix are not linearly independent

“Inverting” Rank Deficient Matrices

- Rank deficient matrices are non-invertible (i.e. we don't know the vector x such that $\mathbf{M}x = b$) as we don't know the component of the x in the null space
- Although we don't know x we can find a vector, x , that satisfies $\mathbf{M}x = b$
- Given a symmetric $n \times n$ matrix with k non-zero eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_k$ we can construct a “pseudo inverse” \mathbf{M}^+ as $\mathbf{V}\mathbf{\Lambda}^+\mathbf{V}^T$ where $\mathbf{\Lambda}^+ = \text{diag}(\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \dots, \frac{1}{\lambda_k}, 0, \dots, 0)$
- This finds the vector x with no component in the null space (it is the solution with the smallest norm)
- This is a different to the pseudo inverse for non-square matrices

Low Rank Approximation

- Recall that matrices with large and small eigenvectors are ill-conditions so the inverse has the potential to greatly amplify any measurement error
- One work around is to set all small eigenvalues to zero and use the pseudo inverse
- Setting small eigenvalues to zero reduces the rank of the matrix and is an example of a low rank approximation
- Low rank approximations are much used to obtain approximate models for arrays of data (we will revisit this when we look at SVD)

Summary

- Linear mappings are commonly used in machine learning algorithms such as regression■
- We can understand symmetric operators by looking at their eigenvectors■
- Any symmetric matrix can be decomposed as $\mathbf{M} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$
 - ★ where \mathbf{V} are orthogonal matrices whose rows are the eigenvector
 - ★ and $\mathbf{\Lambda}$ is a diagonal matrix of the eigenvalues■
- This decomposition allows us to understand inverse mappings■