3^{rd} Summer School on Machine Learning
IIIT Hyderabad

Linear Algebra for Machine Learning

Lovish Chum
Research Fellow
Center for Visual Information Technology
lovish1234@gmail.com

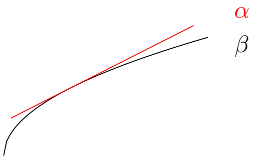
July 9, 2018

Contents

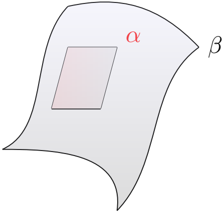
I personally believe that many more people need linear algebra than calculus

- Gilbert Strang

- Working with curves, first step is always to **linearize**
  - Approximate a curve via it's tangent
  - Approximate a curved surface locally via a plane
  - eg. Local Linear Embedding Link



(a) Curve app. by tangent



(b) Surface app. by plane

- Linear algebra provides a bedrock to Machine Learning. Much of the theoretical foundations of ML rely on solving an optimization objective.
  - While solving these problems, a data structure called "matrix" helps you to denote and operate on the data

$$\mathbf{F}_{HO} = \begin{bmatrix} f_{H_1O_1} & \dots & f_{H_1O_n} \\ \vdots & \vdots & \vdots \\ f_{H_nO_1} & \dots & f_{H_nO_n} \end{bmatrix} \qquad \mathbf{B}_{OH} = \begin{bmatrix} b_{O_1H_1} & \dots & b_{O_1H_n} \\ \vdots & \vdots & \vdots \\ b_{O_nH_1} & \dots & b_{O_nH_n} \end{bmatrix}$$

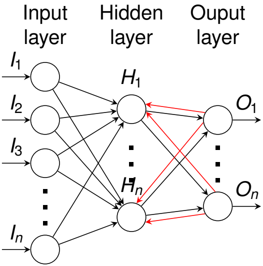


Figure: Neural Network

- As machine learning practitioners, it is important to understand the notations and operations in Linear Algebra as they form the basis of understanding **calculus**, **statistics** and **optimization**

ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION

Diederik P. Kingma* University of Amsterdam, OpenAI dpkingma@openai.com

Jimmy Lei Ba* University of Toronto jimmy@psi.utoronto.ca

ABSTRACT

We introduce *Adam*, an algorithm for first-order gradient-based optimization of stochastic objective functions, based on adaptive estimates of lower-order moments. The method is straightforward to implement, is computationally efficient, has little memory requirements, is invariant to diagonal rescaling of the gradients, and is well suited for problems that are large in terms of data and/or parameters. The method is also appropriate for non-stationary objectives and problems with very noisy and/or sparse gradients. The hyper-parameters have intuitive interpretations and typically require little tuning. Some connections to related algorithms, on which *Adam* was inspired, are discussed. We also analyze the theoretical convergence properties of the algorithm and provide a regret bound on the convergence rate that is comparable to the best known results under the online convex optimization framework. Empirical results demonstrate that Adam works well in practice and compares favorably to other stochastic optimization methods. Finally, we discuss *AdaMax*, a variant of *Adam* based on the infinity norm.

(a) Abstract

10 APPENDIX

10.1 CONVERGENCE PROOF

Definition 10.1. *A function $f : R^d \to R$ is convex if for all $x, y \in R^d$, for all $\lambda \in [0,1]$,*

$$\lambda f(x) + (1-\lambda) f(y) \geq f(\lambda x + (1-\lambda) y)$$

Also, notice that a convex function can be lower bounded by a hyperplane at its tangent.

Lemma 10.2. *If a function $f : R^d \to R$ is convex, then for all $x, y \in R^d$,*

$$f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

The above lemma can be used to upper bound the regret and our proof for the main theorem is constructed by substituting the hyperplane with the Adam update rules.

The following two lemmas are used to support our main theorem. We also use some definitions simplify our notation, where $g_t \triangleq \nabla f_t(\theta_t)$ and $g_{t,i}$ as the i^{th} element. We define $g_{1:t,i} \in \mathbb{R}^t$ as a vector that contains the i^{th} dimension of the gradients over all iterations till t, $g_{1:t,i} = [g_{1,i}, g_{2,i}, \cdots, g_{t,i}]$

Lemma 10.3. *Let $g_t = \nabla f_t(\theta_t)$ and $g_{1:t}$ be defined as above and bounded, $\|g_t\|_2 \leq G$, $\|g_t\|_\infty \leq G_\infty$. Then,*

$$\sum_{t=1}^{T} \sqrt{\frac{g_{t,i}^2}{t}} \leq 2G_\infty \|g_{1:T,i}\|_2$$

Proof. We will prove the inequality using induction over T.

The base case for $T = 1$, we have $\sqrt{g_{1,i}^2} \leq 2G_\infty \|g_{1,i}\|_2$.

For the inductive step,

$$\begin{aligned}
\sum_{t=1}^{T} \sqrt{\frac{g_{t,i}^2}{t}} &= \sum_{t=1}^{T-1} \sqrt{\frac{g_{t,i}^2}{t}} + \sqrt{\frac{g_{T,i}^2}{T}} \\
&\leq 2G_\infty \|g_{1:T-1,i}\|_2 + \sqrt{\frac{g_{T,i}^2}{T}} \\
&= 2G_\infty \sqrt{\|g_{1:T,i}\|_2^2 - g_T^2} + \sqrt{\frac{g_{T,i}^2}{T}}
\end{aligned}$$

(b) Appendix

- As programmers, it is important to understand underlying operations in libraries such as *numpy*, *torch*
  - Almost all of these libraries we use today act on matrices or tensors
- How can we perform matrix computations with **acceptable speed** and **accuracy** ?
  - Usually, we sit back and relax while libraries such as BLAS and LAPACK do these jobs
- How to know if the operation we are performing is **numerically stable** ?

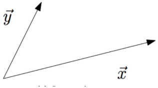
- Get an intuition behind linear transformations. Appreciate the duality between linear transformations and matrices
- Matrix Factorization
  - Eigendecomposition
  - Singular Value Decomposition
  - ~~LU Decomposition~~
  - ~~QR Decomposition~~
- Applications

- Linear Algebra by Gilbert Strang Link
    - Standard course on LA. Though it is 36 lectures long, it helps you start from the basics.
- Essence of Linear Algebra by Grand Sanderson Link
    - Amazing course which helps on gaining intuition of several notations and operations
- Computational Linear Algebra by Fast.ai Link
    - One of the most relevant course for ML practitioners. Flows from applications back to the concepts
- Immersive Linear Algebra by J. Storm et al Link
    - Quiet similar to the essence of linear algebra but a bit broader and interactive.

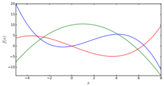
- scalars are written in lowecase and *italics* eg. s
- vectors are written in lowercase and **bold** eg. $\mathbf{v}$
- Matrices are written in Uppercase and **bold** eg. $\mathbf{M}$

Mathematics is the art of giving same name to different things - Henri Poincare

- A vector can be interpreted as a an arrow with a direction and scale, an abstract concept or a grid of numbers



(a) Geometric Notion

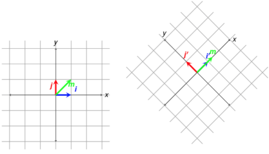


(b) Abstract Concepts

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

(c) List of Numbers

- Similarly a matrix can be thought of as a transformation of space (third lecture) or a grid of numbers



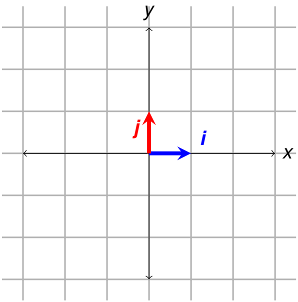
(a) Transformation (Geometrical Notion)

$$\begin{bmatrix} 1 & 1 & -1 \\ 1 & -1 & 2 \\ 2 & 0 & 3 \end{bmatrix}$$

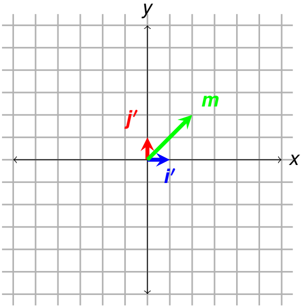
(b) Grid of Numbers (Data Structure)

No one can be told what the matrix is, you have to ***see*** *it for yourself !!*

- Morpheus in "The Matrix" !!



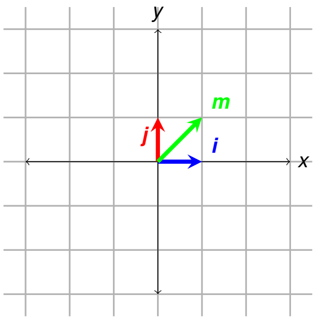
(a) Original Space



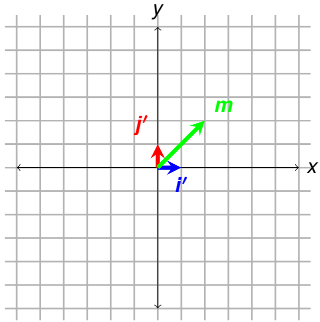
(b) Scaled Space

How to determine the coordinates of ***m*** in the original space ?

$$i' \rightarrow \frac{1}{2}i \rightarrow \frac{1}{2}\begin{bmatrix}1\\0\end{bmatrix} \rightarrow \begin{bmatrix}\frac{1}{2}\\0\end{bmatrix} \qquad j' \rightarrow \frac{1}{2}j \rightarrow \frac{1}{2}\begin{bmatrix}0\\1\end{bmatrix} \rightarrow \begin{bmatrix}0\\\frac{1}{2}\end{bmatrix}$$



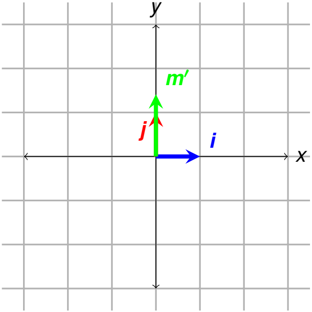
(a) Original Space



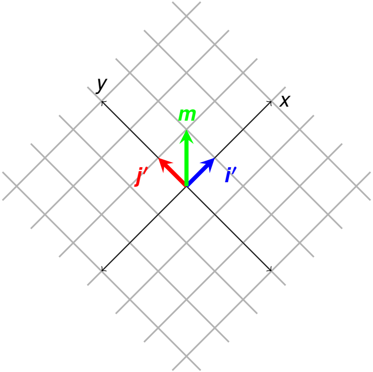
(b) Scaled Space (by factor of 2)

If we know where the i and j vectors (called the **basis** vectors) end up after a transformation, we can easily know where rest of the vectors will end up after the transformation. This information is contained in a matrix called the **transformation matrix**

$$\text{Transformation Matrix } \boldsymbol{A} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$$

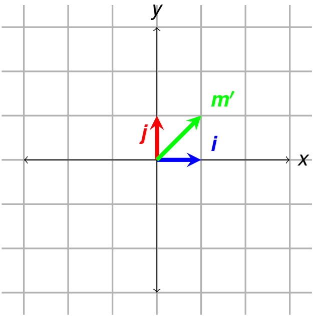


(a) Original Space

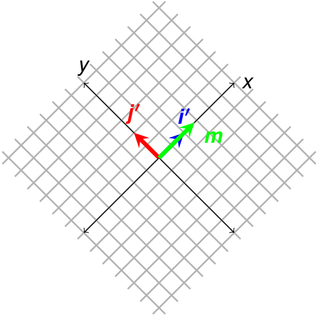


(b) Rotated Space
(45 degrees clockwise)

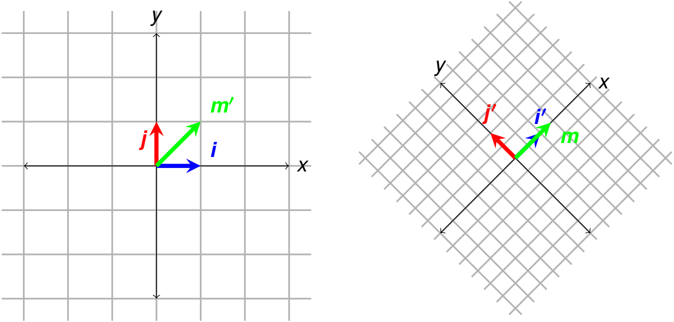
Transformation Matrix $\boldsymbol{A} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$ $\quad m' = \boldsymbol{A}m$



(a) Original Space



(b) Transformed Space
(45 degrees clockwise)
(scaled by a factor of 2)

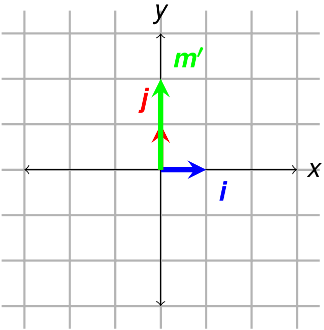


(a) Original Space

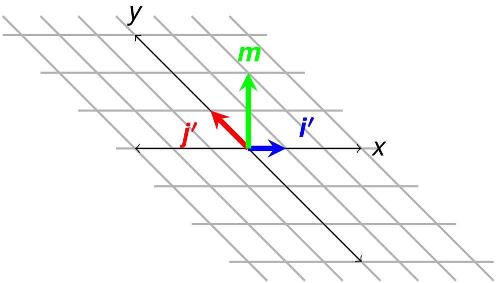
(b) Transformed Space
(45 degrees clockwise)
(scaled by a factor of 2)

Transformation Matrix

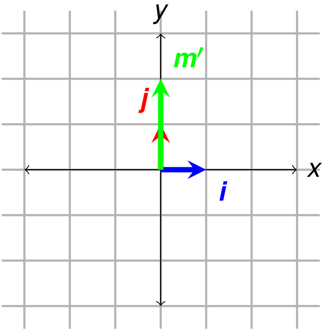
$$\boldsymbol{A} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{1}{2\sqrt{2}} & -\frac{1}{2\sqrt{2}} \\ \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \end{bmatrix}$$



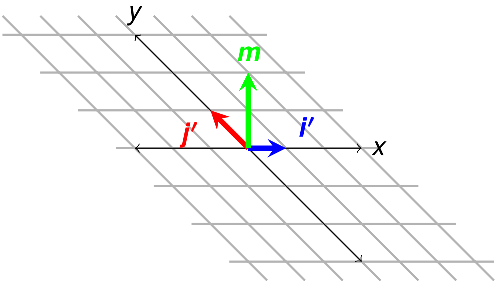
(a) Original Space



(b) Sheared Space



(a) Original Space

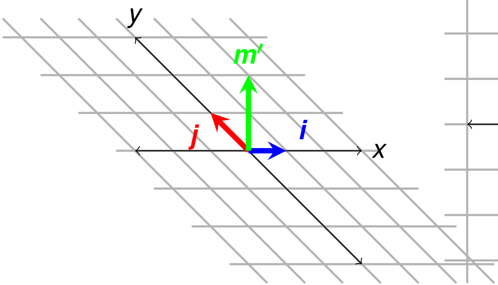


(b) Sheared Space

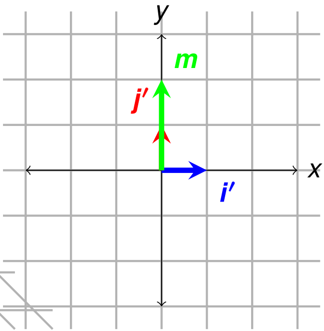
Transformation Matrix

$$\boldsymbol{A} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$$

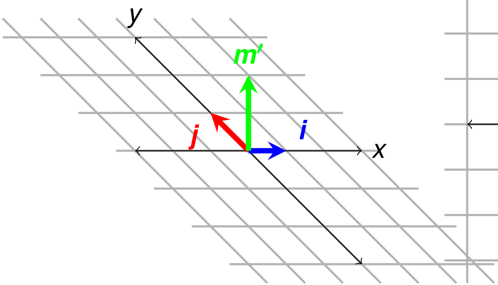
Linear Transformations
Inverse Transform



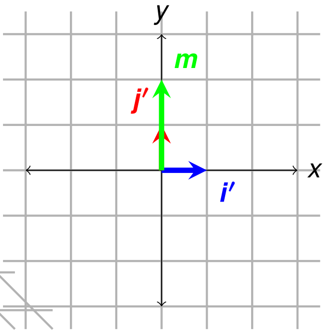
(a) Sheared Space



(b) Original Space



(a) Sheared Space



(b) Original Space

Inverse of Transformation Matrix

$$\boldsymbol{A}^{-1} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

- Scaling
- Rotation
- Shear
- Combination
- Inverse
- A transformation is said to be **linear** when
    - The origin does not move after the transformation
    - Any two straight lines in the original space which are parallel remain parallel after the transformation

Interest Bearing Bank Accounts

Suppose Account 1 yields 5% interest and Account 2 yields 3% interest. We represent the balance in account by 2-D vector $\mathbf{x}^t = \begin{bmatrix} x_1^t \\ x_2^t \end{bmatrix}$. We can write $\mathbf{x}^{t+1}$ in terms of $\mathbf{x}_t$ as

$$\mathbf{x}^{t+1} = \begin{bmatrix} 1.05 & 0 \\ 0 & 1.03 \end{bmatrix} \mathbf{x}^t$$

As it's a **diagonal** matrix, we can find $\mathbf{x}^{20}$ in a relatively easy manner

$$\mathbf{x}^{20} = \mathbf{A}\mathbf{x}^{19}$$

$$\mathbf{x}^{20} = \underbrace{\mathbf{A}.\mathbf{A} \cdots \mathbf{A}}_{\text{20 times}} \mathbf{x}^0$$

$$\mathbf{x}^{20} = \underbrace{\begin{bmatrix} 1.05 & 0 \\ 0 & 1.03 \end{bmatrix} \begin{bmatrix} 1.05 & 0 \\ 0 & 1.03 \end{bmatrix} \cdots \begin{bmatrix} 1.05 & 0 \\ 0 & 1.03 \end{bmatrix}}_{\text{20 times}} \mathbf{x}^0 = \begin{bmatrix} 1.05^{20} & 0 \\ 0 & 1.03^{20} \end{bmatrix} \mathbf{x}^0$$

Pokemon Reproduction

Let's assume for now that bearing a child is not contingent upon any other activity. Assume two fixed rules

- Each adult gives birth to a child every month
- Each child becomes an adult in a month
- No one dies

Pokemon Reproduction

Let's assume for now that bearing a child is not contingent upon any other activity. Assume two fixed rules

- Each adult gives birth to a child every month
- Each child becomes an adult in a month
- No one dies

The equation can be written as

$$\begin{bmatrix} A^{m+1} \\ C^{m+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} A^{m} \\ C^{m} \end{bmatrix}$$

Calculating the adult and child population after 20 months

$$\begin{bmatrix} A^{20} \\ C^{20} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \cdots \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}}_{\text{20 times}} \begin{bmatrix} A^{0} \\ C^{0} \end{bmatrix}$$

Pokemon Reproduction

Let's assume for now that bearing a child is not contingent upon any other activity. Assume two fixed rules

- Each adult gives birth to a child every month
- Each child becomes an adult in a month
- No one dies

The equation can be written as

$$\begin{bmatrix} A^{m+1} \\ C^{m+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} A^{m} \\ C^{m} \end{bmatrix}$$

Calculating the adult and child population after 20 months

$$\begin{bmatrix} A^{20} \\ C^{20} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \cdots \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}}_{\text{20 times}} \begin{bmatrix} A^{0} \\ C^{0} \end{bmatrix}$$

The matrix is **non-diagonal**, what can be done to get around the issue ?

- Diagonalize the matrix

Pokemon Reproduction - Surprising Fact

Let us take a matrix $\mathbf{S} = \begin{bmatrix} \frac{\sqrt{5}+1}{2} & \frac{1-\sqrt{5}}{2} \\ 1 & 1 \end{bmatrix}$. Observe that

$$\mathbf{\Lambda} = \mathbf{S}^{-1} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \mathbf{S} = \begin{bmatrix} \frac{\sqrt{5}+1}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{bmatrix}$$

is a diagonal matrix. How can this information help you ?

Pokemon Reproduction - Surprising Fact

Let us take a matrix $\mathbf{S} = \begin{bmatrix} \frac{\sqrt{5}+1}{2} & \frac{1-\sqrt{5}}{2} \\ 1 & 1 \end{bmatrix}$. Observe that

$$\mathbf{\Lambda} = \mathbf{S}^{-1} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \mathbf{S} = \begin{bmatrix} \frac{\sqrt{5}+1}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{bmatrix}$$

is a diagonal matrix. How can this information help you ?

$$\implies \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} = \mathbf{S\Lambda S}^{-1}$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^{20} = \underbrace{\mathbf{S\Lambda S}^{-1}\mathbf{S\Lambda S}^{-1} \cdots \mathbf{S\Lambda S}^{-1}}_{\text{20 times}} = \mathbf{S\Lambda}^{20}\mathbf{S}^{-1}$$

where $\mathbf{\Lambda}$ is a diagonal matrix.

If we are able to decompose the original matrix in such a way $\mathbf{S\Lambda S}^{-1}$, then we call the matrix **diagonalizable**, otherwise it's called **defective**.

Pokemon Reproduction - Surprising Fact

Let us take a matrix $\mathbf{S} = \begin{bmatrix} \frac{\sqrt{5}+1}{2} & \frac{1-\sqrt{5}}{2} \\ 1 & 1 \end{bmatrix}$. Observe that

$$\mathbf{\Lambda} = \mathbf{S}^{-1} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \mathbf{S} = \begin{bmatrix} \frac{\sqrt{5}+1}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{bmatrix}$$

is a diagonal matrix. How can this information help you ?

$$\implies \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} = \mathbf{S\Lambda S}^{-1}$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^{20} = \underbrace{\mathbf{S\Lambda S}^{-1}\mathbf{S\Lambda S}^{-1} \cdots \mathbf{S\Lambda S}^{-1}}_{\text{20 times}} = \mathbf{S\Lambda}^{20}\mathbf{S}^{-1}$$

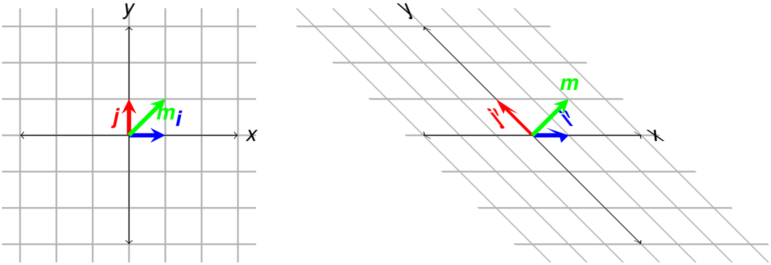
where $\mathbf{\Lambda}$ is a diagonal matrix.

If we are able to decompose the original matrix in such a way $\mathbf{S\Lambda S}^{-1}$, then we call the matrix **diagonalizable**, otherwise it's called **defective**.

Pokemon Reproduction - Surprising Fact 2

The columns of $\mathbf{S}$ contain eigenvectors and the diagonal values of $\mathbf{\Lambda}$ are eigenvalues corresponding to the eigenvectors.

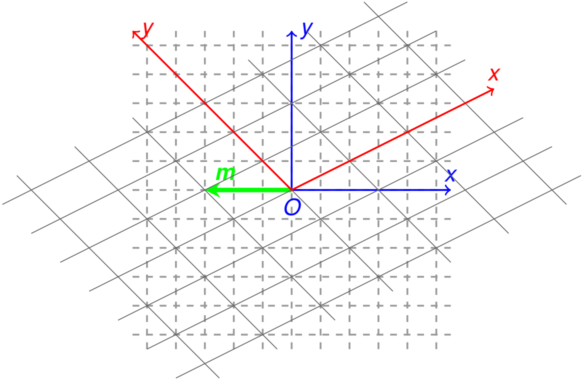
- In the Linear Transformation section, we saw that there are **many ways to represent a vector**. In fact, the representation depends on the basis vectors.
- For example, a vector which is represented as $\begin{bmatrix}1\\1\end{bmatrix}$ in the original space can be represented as $\begin{bmatrix}2\\1\end{bmatrix}$ in the transformed space



(a) Original Space

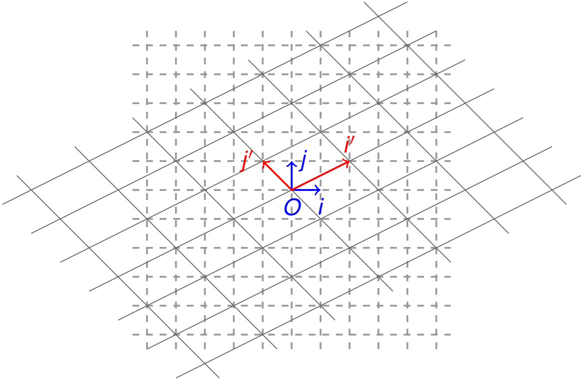
(b) Transformed Space

- Changing the basis is akin to speaking a different language. As two distinct words in different languages convey same entity, different coordinates in different **bases** can be used to describe the same vector.



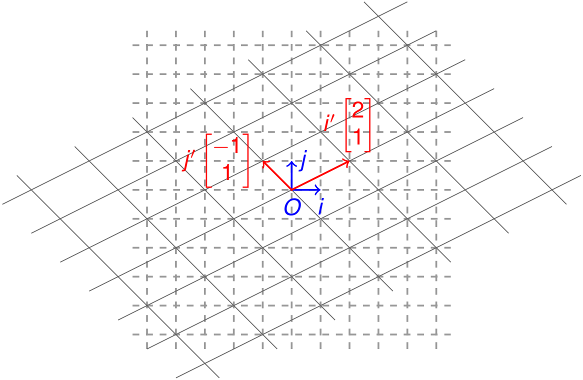
(a) An Example Transformation

- The blue basis vectors define a original space whereas the red basis vectors define the transformed space
- What is the transformation matrix ?



(a) An Example Transformation

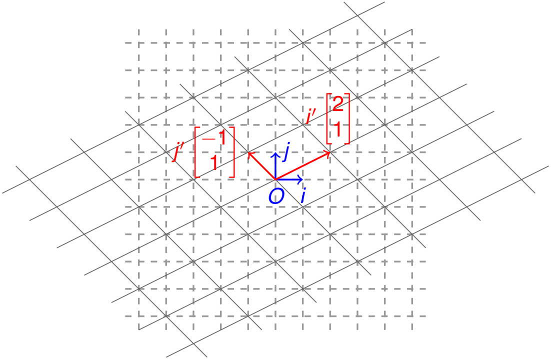
- Look at how the transformed basis vectors i' and j' be represented in the untransformed space.



(a) An Example Transformation

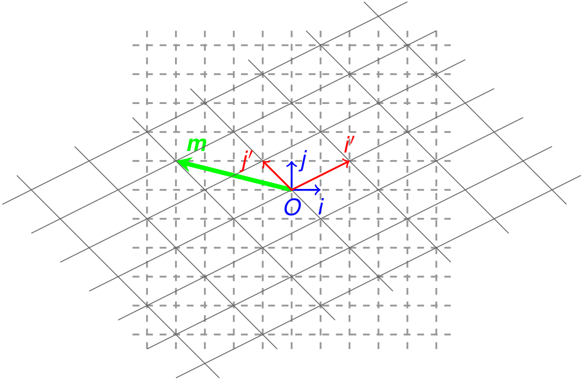
- Look at how the transformed basis vectors i' and j' be represented in the untransformed space.

$$\mathbf{A} = \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix}$$



(a) An Example Transformation

- How can we represent m with red basis vectors ?
- How can we represent m with blue basis vectors ?



(a) Example Transformation

- Now let's try to see the same analytically ?

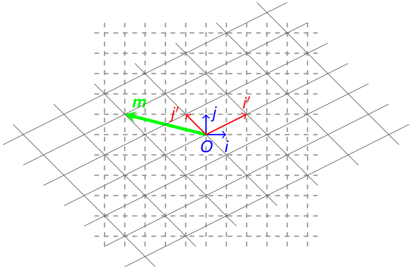
$$\mathbf{A} = \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix}$$

- Now let's try to see the same analytically ?

$$\mathbf{A} = \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix}$$

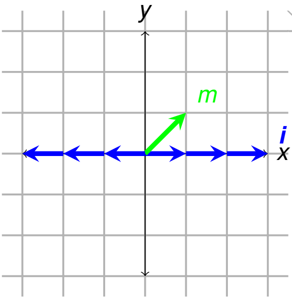
- If you multiply the transformation matrix with the coordinates of m according to the red basis, you get the coordinates of m according to blue basis

$$\underbrace{\begin{bmatrix} -4 \\ 1 \end{bmatrix}}_{\text{Coordinates in blue basis}} = \underbrace{\begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix}}_{\text{Transformation from blue to red basis}} \underbrace{\begin{bmatrix} -1 \\ 2 \end{bmatrix}}_{\text{Coordinates in red basis}}$$

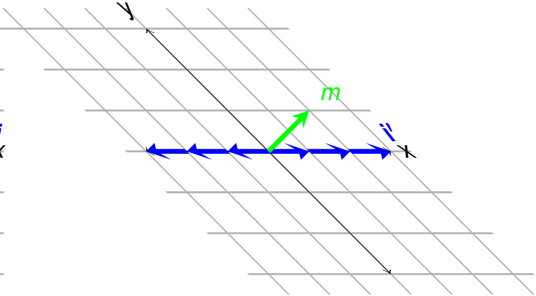


(a) An Example Transformation

- There are special vectors in the original space which **do not change direction** even after transformation. However, they may change scale
- As an example, for the transformation below the vectors in the x direction remain in the same direction after the transformation



(a) Original Space



(b) Transformed Space

- Thinking of eigenvectors of a matrix is bit non-intuitive
- Thinking it in terms of the **vectors which do not change direction** under a linear transformation is much more simple
- As we already know, every linear transformation can be represented as a matrix
- Algebraically, we can represent these vectors as follows :-

$$\mathbf{Av} = \lambda \mathbf{v}$$

Converting the scalar-vector product on the right hand side to matrix vector product

$$\mathbf{Av} = \lambda \mathbf{Iv}$$

$$\mathbf{Av} - \lambda \mathbf{Iv} = \mathbf{0}$$

$$(\mathbf{A} - \lambda \mathbf{I})\mathbf{v} = \mathbf{0}$$

Either

- Thinking of eigenvectors of a matrix is bit non-intuitive
- Thinking it in terms of the **vectors which do not change direction** under a linear transformation is much more simple
- As we already know, every linear transformation can be represented as a matrix
- Algebraically, we can represent these vectors as follows :-

$$\mathbf{Av} = \lambda \mathbf{v}$$

Converting the scalar-vector product on the right hand side to matrix vector product

$$\mathbf{Av} = \lambda \mathbf{Iv}$$

$$\mathbf{Av} - \lambda \mathbf{Iv} = \mathbf{0}$$

$$(\mathbf{A} - \lambda \mathbf{I})\mathbf{v} = \mathbf{0}$$

Either

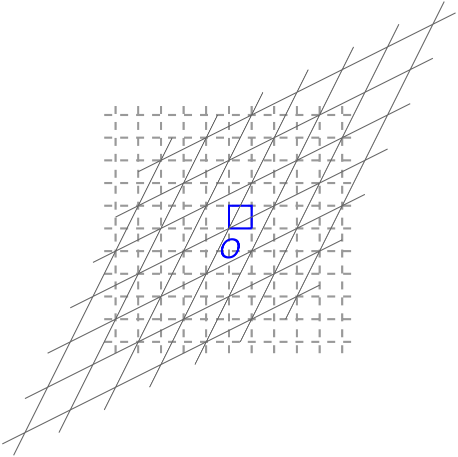
- $\mathbf{v}$ is Null vector or
- $(\mathbf{A} - \lambda \mathbf{I})$ is Null matrix or
- $(\mathbf{A} - \lambda \mathbf{I})\mathbf{v} = \mathbf{0}$

$$det(\mathbf{A} - \lambda \mathbf{I}) = 0$$ Link

The above equation is called the **characteristic equation** of the matrix $\mathbf{A}$

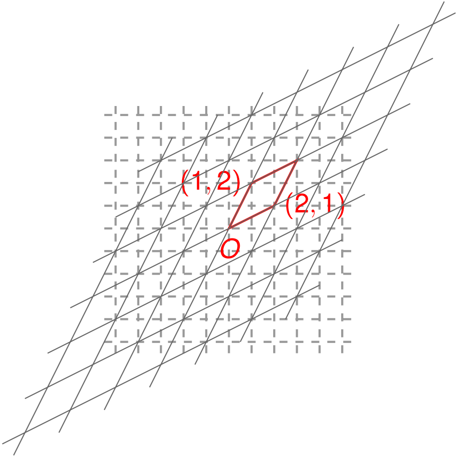
*The purpose of computation is **insight**, not **numbers**.*

- Richard Hamming



*The purpose of computation is **insight**, not **numbers**.*

- Richard Hamming



*The purpose of computation is **insight**, not **numbers**.*

- Richard Hamming

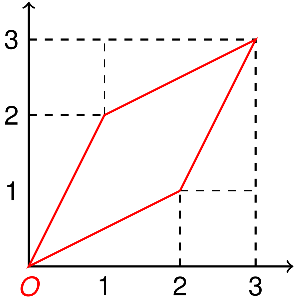


Figure: Finding the value of determinant

- It is important to understand that why $(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0} \implies det(\mathbf{A} - \lambda\mathbf{I}) = 0$. Note that $\mathbf{0}$ is the **Null Vector** and 0 is a scalar.
- Algebraically, let's say that $\mathbf{B} = (\mathbf{A} - \lambda\mathbf{I})$. We have $\mathbf{Bv} = \mathbf{0}$.

$$\implies \begin{bmatrix} \vdots & \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \mathbf{c}_1 & \mathbf{c}_2 & \mathbf{c}_3 & \cdots & \mathbf{c}_n \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \cdots & \vdots \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix} \tag{1}$$

where c_i is the i^{th} column of $\mathbf{B}$.

$$\implies \sum_{i=0}^{n} \mathbf{c}_i v_i = \mathbf{0}$$

$$\implies \sum_{i=0}^{n-1} \mathbf{c}_i v_i = -\mathbf{c}_n v_n$$

As all v_i are scalars, we define $w_i = -\frac{\mathbf{v}_i}{v_n}$

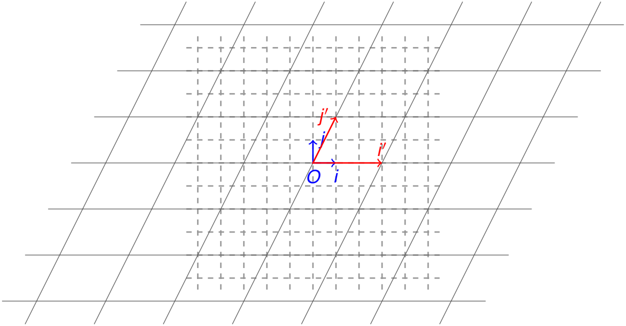
$$\implies \sum_{i=0}^{n-1} \mathbf{c}_i w_i = \mathbf{c}_n$$

Putting in back in equation (1), we see that one column of the matrix c_n can be represented as linear combination of other columns. Using $C_n \to C_n - \sum_{i=0}^{n-1} \mathbf{C}_i w_i$, the last column of the determinant becomes $\mathbf{0}$. Hence the value of the determinant of $\mathbf{B}$ is 0.

- Analytically : It is just breaking one matrix into three parts where the middle matrix contains non-zero elements at it's diagonal

$$\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$$

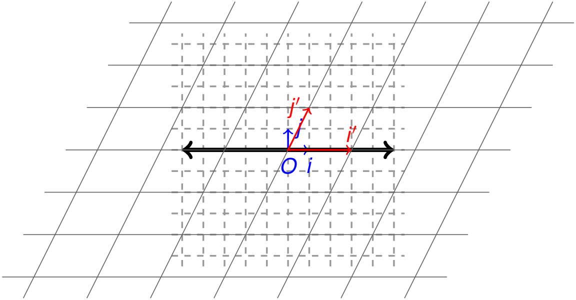
- How can we interpret this it geometrically ?
  - Can eigenvectors help us in interpreting the transformation geometrically ?



(a) Eigenvector

- Is there any other vector except i which gets scaled under this transformation?

$$(1,0) \to (3,0)$$



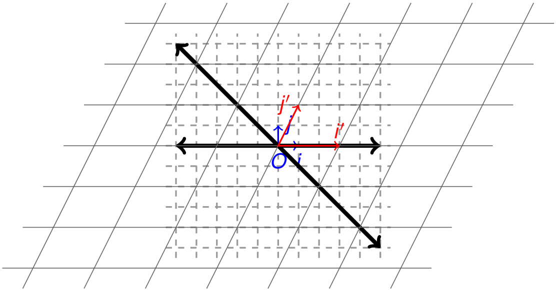
(a) Example Transformation

- Is there any other vector except i which gets scaled under this transformation?

$$(1,0) \rightarrow (3,0)$$

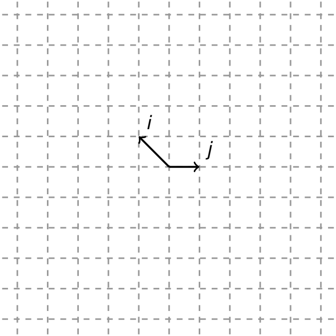
$$(-1,1) \rightarrow (-2,2)$$

- These are the eigenvectors under the transform $\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$



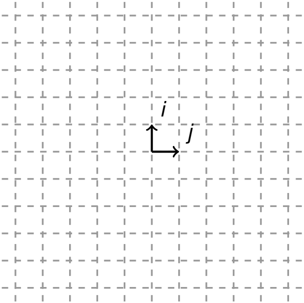
(a) Two Eigenvectors

- Imagine what would happen if we choose our basis as the set of eigenvectors !!
  - Under the given transform, the basis vectors would just **scale**
  - If we can **express our the coordinates in this basis**, the original transform $\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$ (shear and scale) can be re-written as a scaling transform $\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$ (only scale)

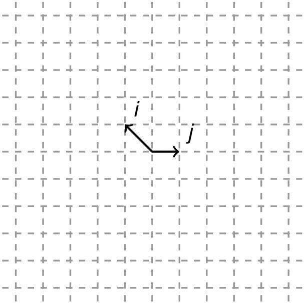


(a) Changing the basis to Eigenvectors

- Imagine what would happen if we choose our basis as the set of eigenvectors !!
  - Under the given transform, the basis vectors would just **scale**
  - If we can **express our the coordinates in this basis**, the original transform $\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$ (shear and scale) can be re-written as a scaling transform $\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$ (only scale)



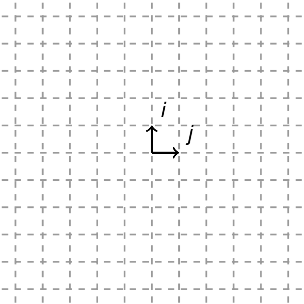
(a) Original Basis



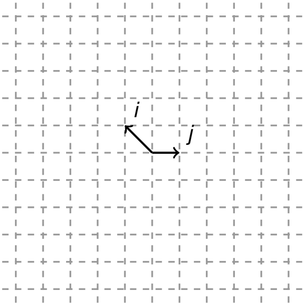
(b) Changing the basis to Eigenvectors

- Component 1
  - We must multiply with the **inverse** of the matrix which contains basis vectors as it's columns.

$$\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}^{-1}$$



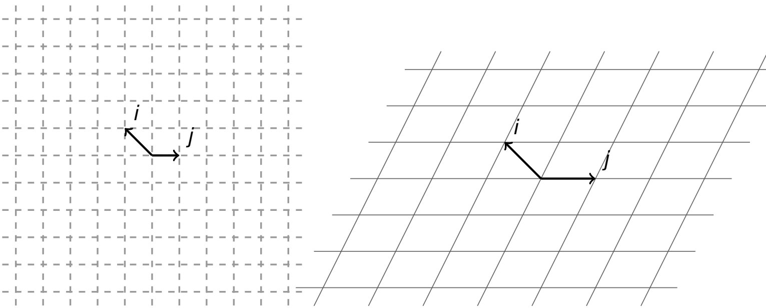
(a) Original Basis



(b) Changing the basis to Eigenvectors

- Component 2
  - Now, multiply with a diagonal matrix with the elements as the corresponding eigenvalues

$$\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$$



(a) New Basis (b) After Transformation

- Component 3
  - As we changed the basis vectors initially to make things easier, we need to revert back to the original basis vectors

$$\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$$

(a) After Transformation (b) Original Basis

- The whole transform can be represented in a space with eigenbasis. This is known as **eigenspace** of a transformation.

$$\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}}_{\text{Component 3}} \underbrace{\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}}_{\text{Component 2}} \underbrace{\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}^{-1}}_{\text{Component 1}}$$

$$\implies \underbrace{\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}}_{\text{Component 2}} = \underbrace{\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}^{-1}}_{\text{Component 1}} \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} \underbrace{\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}}_{\text{Component 3}}$$

- Can we diagonalize every matrix ?

- The whole transform can be represented in a space with eigenbasis. This is known as **eigenspace** of a transformation.

$$\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}}_{\text{Component 3}} \underbrace{\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}}_{\text{Component 2}} \underbrace{\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}^{-1}}_{\text{Component 1}}$$

$$\implies \underbrace{\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}}_{\text{Component 2}} = \underbrace{\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}^{-1}}_{\text{Component 1}} \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} \underbrace{\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}}_{\text{Component 3}}$$

- Can we diagonalize every matrix ? No
- Can we diagonalize every square matrix ?

- The whole transform can be represented in a space with eigenbasis. This is known as **eigenspace** of a transformation.

$$\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}}_{\text{Component 3}} \underbrace{\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}}_{\text{Component 2}} \underbrace{\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}^{-1}}_{\text{Component 1}}$$

$$\implies \underbrace{\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}}_{\text{Component 2}} = \underbrace{\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}^{-1}}_{\text{Component 1}} \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} \underbrace{\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}}_{\text{Component 3}}$$

- Can we diagonalize every matrix ? No
- Can we diagonalize every square matrix ? No
  - **Diagonalizability Theorem**: An $n \times n$ matrix $\mathbf{A}$ is diagonalizable iff it has n linearly independent eigenvectors.
- Can Eigen-decomposition be used for non-square matrices ?

- The whole transform can be represented in a space with eigenbasis. This is known as **eigenspace** of a transformation.

$$\begin{bmatrix}3 & 1\\0 & 2\end{bmatrix} = \underbrace{\begin{bmatrix}1 & -1\\0 & 1\end{bmatrix}}_{\text{Component 3}} \underbrace{\begin{bmatrix}3 & 0\\0 & 2\end{bmatrix}}_{\text{Component 2}} \underbrace{\begin{bmatrix}1 & -1\\0 & 1\end{bmatrix}^{-1}}_{\text{Component 1}}$$

$$\implies \underbrace{\begin{bmatrix}3 & 0\\0 & 2\end{bmatrix}}_{\text{Component 2}} = \underbrace{\begin{bmatrix}1 & -1\\0 & 1\end{bmatrix}^{-1}}_{\text{Component 1}} \begin{bmatrix}3 & 1\\0 & 2\end{bmatrix} \underbrace{\begin{bmatrix}1 & -1\\0 & 1\end{bmatrix}}_{\text{Component 3}}$$

- Can we diagonalize every matrix ? No
- Can we diagonalize every square matrix ? No
  - **Diagonalizability Theorem**: An $n \times n$ matrix $\mathbf{A}$ is diagonalizable iff it has n linearly independent eigenvectors.
- Can Eigen-decomposition be used for non-square matrices ? No
  - Singular Value Decomposition (SVD)

Pokemon Reproduction

Let's assume for now that bearing a child is not contingent upon any other activity. Assume two fixed rules

- Each adult gives birth to a child every month
- Each child becomes an adult in a month
- No one dies

The equation can be written as

$$\begin{bmatrix} A^{m+1} \\ C^{m+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} A^{m} \\ C^{m} \end{bmatrix}$$

Calculating the adult and child population after 20 months

$$\begin{bmatrix} A^{20} \\ C^{20} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \cdots \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}}_{\text{20 times}} \begin{bmatrix} A^{0} \\ C^{0} \end{bmatrix}$$

The matrix is **non-diagonal**, what can be done to get around the issue ?

- Diagonalize the matrix

Pokemon Reproduction - Not at all Surprising

Let us take a matrix $\mathbf{S} = \begin{bmatrix} \frac{\sqrt{5}+1}{2} & \frac{1-\sqrt{5}}{2} \\ 1 & 1 \end{bmatrix}$. Observe that

$$\mathbf{\Lambda} = \mathbf{S}^{-1} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \mathbf{S} = \begin{bmatrix} \frac{1}{\sqrt{5}} & -\frac{(1-\sqrt{5})}{2\sqrt{5}} \\ -\frac{1}{\sqrt{5}} & \frac{1+\sqrt{5}}{2\sqrt{5}} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{5}+1}{2} & \frac{1-\sqrt{5}}{2} \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{5}+1}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{bmatrix}$$

is a diagonal matrix. How can this information help you ?

Pokemon Reproduction - Not at all Surprising

Let us take a matrix $\mathbf{S} = \begin{bmatrix} \frac{\sqrt{5}+1}{2} & \frac{1-\sqrt{5}}{2} \\ 1 & 1 \end{bmatrix}$. Observe that

$$\mathbf{\Lambda} = \mathbf{S}^{-1} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \mathbf{S} = \begin{bmatrix} \frac{1}{\sqrt{5}} & -\frac{(1-\sqrt{5})}{2\sqrt{5}} \\ -\frac{1}{\sqrt{5}} & \frac{1+\sqrt{5}}{2\sqrt{5}} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{5}+1}{2} & \frac{1-\sqrt{5}}{2} \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{5}+1}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{bmatrix}$$

is a diagonal matrix. How can this information help you ?

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^{20} = \underbrace{\mathbf{S\Lambda S}^{-1}\mathbf{S\Lambda S}^{-1} \cdots \mathbf{S\Lambda S}^{-1}}_{\text{20 times}} = \mathbf{S\Lambda}^{20}\mathbf{S}^{-1}$$

where $\mathbf{\Lambda}$ is a diagonal matrix.

If we are able to decompose the original matrix in such a way $\mathbf{S\Lambda S}^{-1}$, then we call the matrix **diagonalizable**, otherwise it's called **defective**.

Pokemon Reproduction - Not at all Surprising

Let us take a matrix $\mathbf{S} = \begin{bmatrix} \frac{\sqrt{5}+1}{2} & \frac{1-\sqrt{5}}{2} \\ 1 & 1 \end{bmatrix}$. Observe that

$$\mathbf{\Lambda} = \mathbf{S}^{-1} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \mathbf{S} = \begin{bmatrix} \frac{1}{\sqrt{5}} & -\frac{(1-\sqrt{5})}{2\sqrt{5}} \\ -\frac{1}{\sqrt{5}} & \frac{1+\sqrt{5}}{2\sqrt{5}} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{5}+1}{2} & \frac{1-\sqrt{5}}{2} \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{5}+1}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{bmatrix}$$

is a diagonal matrix. How can this information help you ?

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^{20} = \underbrace{\mathbf{S\Lambda S}^{-1}\mathbf{S\Lambda S}^{-1} \cdots \mathbf{S\Lambda S}^{-1}}_{\text{20 times}} = \mathbf{S\Lambda}^{20}\mathbf{S}^{-1}$$

where $\mathbf{\Lambda}$ is a diagonal matrix.

If we are able to decompose the original matrix in such a way $\mathbf{S\Lambda S}^{-1}$, then we call the matrix **diagonalizable**, otherwise it's called **defective**.

Pokemon Reproduction - Not at all Surprising

The columns of $\mathbf{S}$ contain eigenvectors and the diagonal values of $\mathbf{\Lambda}$ are eigenvalues corresponding to the eigenvectors.

Epidemic in PokeTown

- Let's alter the problem a bit. An epidemic has spread in PokeTown.
  - A a fraction of adult population catches it every month
  - A c fraction of children catch it every month
  - A d fraction of pokemons who have the epidemic die every month, and the rest recover
  - Pokemons who catch the disease cannot reproduce

- How can we tell if the Pokemon population explodes/vanishes/remains the same ?

Epidemic in PokeTown

- Let's alter the problem a bit. An epidemic has spread in PokeTown.
  - A a fraction of adult population catches it every month
  - A c fraction of children catch it every month
  - A d fraction of pokemons who have the epidemic die every month, and the rest recover
  - Pokemons who catch the disease cannot reproduce

- How can we tell if the Pokemon population explodes/vanishes/remains the same ?

$$\begin{bmatrix} A^{m+1} \\ C^{m+1} \\ S^{m+1} \end{bmatrix} = \begin{bmatrix} 1-a & 1-c & 1-d \\ 1 & 0 & 0 \\ a & c & 0 \end{bmatrix} \begin{bmatrix} A^{m} \\ C^{m} \\ S^{m} \end{bmatrix}$$

Epidemic in PokeTown

- Let's alter the problem a bit. An epidemic has spread in PokeTown.
  - A a fraction of adult population catches it every month
  - A c fraction of children catch it every month
  - A d fraction of pokemons who have the epidemic die every month, and the rest recover
  - Pokemons who catch the disease cannot reproduce

- How can we tell if the Pokemon population explodes/vanishes/remains the same ?

$$\begin{bmatrix} A^{m+1} \\ C^{m+1} \\ S^{m+1} \end{bmatrix} = \begin{bmatrix} 1-a & 1-c & 1-d \\ 1 & 0 & 0 \\ a & c & 0 \end{bmatrix} \begin{bmatrix} A^m \\ C^m \\ S^m \end{bmatrix}$$

- Examples
  - As an example, take $a = 0.5$, c=0.5, d=0.8
  - Another example, take $a = 0.7$, c=0.7, d=0.8
  - Another one, take $a = 0.2$, c=0.2, d=0.8

Epidemic in PokeTown

- Let's alter the problem a bit. An epidemic has spread in PokeTown.
  - A a fraction of adult population catches it every month
  - A c fraction of children catch it every month
  - A d fraction of pokemons who have the epidemic die every month, and the rest recover
  - Pokemons who catch the disease cannot reproduce

- How can we tell if the Pokemon population explodes/vanishes/remains the same ?

$$\begin{bmatrix} A^{m+1} \\ C^{m+1} \\ S^{m+1} \end{bmatrix} = \begin{bmatrix} 1-a & 1-c & 1-d \\ 1 & 0 & 0 \\ a & c & 0 \end{bmatrix} \begin{bmatrix} A^m \\ C^m \\ S^m \end{bmatrix}$$

- Examples
  - As an example, take $a = 0.5$, c=0.5, d=0.8
  - Another example, take $a = 0.7$, c=0.7, d=0.8
  - Another one, take $a = 0.2$, c=0.2, d=0.8
- Hints
  - What does the **largest eigenvalue** of the transformed matrix mean ?
  - What if the eigenvalue is between ($-$ inf,1) or (1, inf) ?
  - What if the eigenvalue is between (-1 to 1) ?
  - What if it is in the set -1,1 ?

- To remember
  - Eigendecomposition is a process of simplifying a matrix transformation based on the eigenvectors of the transformation
  - Eigendecomposition is only possible for **square** matrices. However, not all square matrices can be decomposed in this way
  - In case eigenvectors are orthogonal to each other, $\mathbf{A} = \mathbf{A}^T$
- Homework
  - If **A** is a symmetric matrix and **u** and **v** are two eigenvectors of **A**, show that **u** and **v** are orthogonal to each other.

- To remember
  - Eigendecomposition is a process of simplifying a matrix transformation based on the eigenvectors of the transformation
  - Eigendecomposition is only possible for **square** matrices. However, not all square matrices can be decomposed in this way
  - In case eigenvectors are orthogonal to each other, $\mathbf{A} = \mathbf{A}^T$
- Homework
  - If $\mathbf{A}$ is a symmetric matrix and $\mathbf{u}$ and $\mathbf{v}$ are two eigenvectors of $\mathbf{A}$, show that $\mathbf{u}$ and $\mathbf{v}$ are orthogonal to each other.
    - **Analytically**, show that $\mathbf{u}.\mathbf{v} = 0$
    - Try to ponder **geometrically** too

$$\begin{bmatrix}3 & 1\\0 & 2\end{bmatrix} = \underbrace{\begin{bmatrix}1 & -1\\0 & 1\end{bmatrix}}_{\text{Reverse Change of Basis}} \underbrace{\begin{bmatrix}3 & 0\\0 & 2\end{bmatrix}}_{\text{Scaling of Eigenbasis}} \underbrace{\begin{bmatrix}1 & -1\\0 & 1\end{bmatrix}^{-1}}_{\text{Change of Basis}}$$

- Any matrix **A** can be factorized into three matrices **P**, **D** and $\mathbf{P}^{-1}$ if it follows certain conditions. We will revise them today.

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1} \tag{2}$$

- Another interpretation of the same can be given as :-

$$\mathbf{A}\mathbf{P} = \mathbf{P}\mathbf{D} \tag{3}$$

  - Defining **P** through it's columns p_i and **D** through it's diagonal entries

$$\mathbf{P} = \begin{bmatrix}p_1 & p_2 & \dots & p_n\end{bmatrix}$$
$$\mathbf{D} = \begin{bmatrix}\lambda_1 & 0 & \dots & 0\\ 0 & \lambda_2 & \dots & 0\\ \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & \dots & \lambda_n\end{bmatrix}$$

  - The diagonal entries in **D** must be the eigenvalues as all it is doing is scaling the matrix **P** whose columns constitute an eigenbasis

$$\mathbf{P}\mathbf{D} = \begin{bmatrix}\lambda_1 p_1 & \lambda_2 p_2 & \dots & \lambda_n p_n\end{bmatrix}$$

- What are the eigenvalues of these matrices ? Can they be diagonalized ?

a) $\begin{bmatrix} 2 & -2 \\ 2 & -2 \end{bmatrix}$ b) $\begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$ c) $\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$

- What are the eigenvalues of these matrices ? Can they be diagonalized ?

  a) $\begin{bmatrix} 2 & -2 \\ 2 & -2 \end{bmatrix}$ b) $\begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$ c) $\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$

  - Matrix a) has a single eigenvalue ie. 0. Can it be diagonalized using eigendecomposition ? No

- What are the eigenvalues of these matrices ? Can they be diagonalized ?

  a) $\begin{bmatrix} 2 & -2 \\ 2 & -2 \end{bmatrix}$ b) $\begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$ c) $\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$

  - Matrix a) has a single eigenvalue ie. 0. Can it be diagonalized using eigendecomposition ? No
  - Although Matrix b) has two distinct eigenvalues ie. 0 and 1, it squishes the whole $\mathbb{R}^2$ space to the x-axis. We cannot find an inverse to this transform. Hence, it is not possible to diagonalize this matrix.

- What are the eigenvalues of these matrices ? Can they be diagonalized ?

  a) $\begin{bmatrix} 2 & -2 \\ 2 & -2 \end{bmatrix}$ b) $\begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$ c) $\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$

  - Matrix a) has a single eigenvalue ie. 0. Can it be diagonalized using eigendecomposition ? No
  - Although Matrix b) has two distinct eigenvalues ie. 0 and 1, it squishes the whole $\mathbb{R}^2$ space to the x-axis. We cannot find an inverse to this transform. Hence, it is not possible to diagonalize this matrix.
  - The eigenbasis ((1,0) and (1,-1)) for matrix c) spans the whole space. Hence, it can be diagonalized.
  - Can we come to the conclusion that the 2×2 matrices with a non-zero single eigenvalue are defective ? Or in general can we say that $n \times n$ matrices with less than n non-zero eigenvalues are defective ?
- Eigenvectors should form a basis that spans the whole space !!
- What about this matrix ?

$$\begin{bmatrix} 3 & 5 & 0 \\ 2 & 1 & -1 \end{bmatrix}$$

  - The transform is squishing 3-D space to 2-D space

- Prove that if **A** is a symmetric matrix, then the eigenvectors form an orthonormal basis.
  - Let **u** and **v** be the two eigenvectors and λ_1 and λ_2 be the corresponding eigenvalues

$$\mathbf{A}\mathbf{u} = \lambda_1 \mathbf{u}$$

$$\mathbf{A}\mathbf{v} = \lambda_2 \mathbf{v}$$

- Prove that if **A** is a symmetric matrix, then the eigenvectors form an orthonormal basis.
  - Let **u** and **v** be the two eigenvectors and λ_1 and λ_2 be the corresponding eigenvalues

$$\mathbf{Au} = \lambda_1 \mathbf{u}$$
$$\mathbf{Av} = \lambda_2 \mathbf{v}$$

  - Taking transpose both sides

$$\implies \mathbf{u}^T \mathbf{A}^T = \lambda_1 \mathbf{u}^T$$
$$\implies \mathbf{u}^T \mathbf{A} = \lambda_1 \mathbf{u}^T$$

- Prove that if **A** is a symmetric matrix, then the eigenvectors form an orthonormal basis.
  - Let **u** and **v** be the two eigenvectors and λ_1 and λ_2 be the corresponding eigenvalues

$$\mathbf{A}\mathbf{u} = \lambda_1 \mathbf{u}$$

$$\mathbf{A}\mathbf{v} = \lambda_2 \mathbf{v}$$

  - Taking transpose both sides

$$\implies \mathbf{u}^T \mathbf{A}^T = \lambda_1 \mathbf{u}^T$$

$$\implies \mathbf{u}^T \mathbf{A} = \lambda_1 \mathbf{u}^T$$

  - Multiplying with **v** on both sides

$$\implies \mathbf{u}^T \mathbf{A} \mathbf{v} = \lambda_1 \mathbf{u}^T \mathbf{v}$$

- Prove that if $\mathbf{A}$ is a symmetric matrix, then the eigenvectors form an orthonormal basis.
  - Let $\mathbf{u}$ and $\mathbf{v}$ be the two eigenvectors and λ_1 and λ_2 be the corresponding eigenvalues

$$\mathbf{Au} = \lambda_1 \mathbf{u}$$
$$\mathbf{Av} = \lambda_2 \mathbf{v}$$

  - Taking transpose both sides

$$\implies \mathbf{u}^T \mathbf{A}^T = \lambda_1 \mathbf{u}^T$$
$$\implies \mathbf{u}^T \mathbf{A} = \lambda_1 \mathbf{u}^T$$

  - Multiplying with $\mathbf{v}$ on both sides

$$\implies \mathbf{u}^T \mathbf{A} \mathbf{v} = \lambda_1 \mathbf{u}^T \mathbf{v}$$

  - By the definition of eigenvectors

$$\implies \mathbf{u}^T (\lambda_2 \mathbf{v}) = \lambda_1 \mathbf{u}^T \mathbf{v}$$
$$\implies \mathbf{u}^T \mathbf{v} (\lambda_2 - \lambda_1) = 0$$

- Prove that if $\mathbf{A}$ is a symmetric matrix, then the eigenvectors form an orthonormal basis.
  - Let $\mathbf{u}$ and $\mathbf{v}$ be the two eigenvectors and λ_1 and λ_2 be the corresponding eigenvalues

$$\mathbf{A}\mathbf{u} = \lambda_1 \mathbf{u}$$
$$\mathbf{A}\mathbf{v} = \lambda_2 \mathbf{v}$$

  - Taking transpose both sides

$$\implies \mathbf{u}^T \mathbf{A}^T = \lambda_1 \mathbf{u}^T$$
$$\implies \mathbf{u}^T \mathbf{A} = \lambda_1 \mathbf{u}^T$$

  - Multiplying with $\mathbf{v}$ on both sides

$$\implies \mathbf{u}^T \mathbf{A}\mathbf{v} = \lambda_1 \mathbf{u}^T \mathbf{v}$$

  - By the definition of eigenvectors

$$\implies \mathbf{u}^T (\lambda_2 \mathbf{v}) = \lambda_1 \mathbf{u}^T \mathbf{v}$$
$$\implies \mathbf{u}^T \mathbf{v} (\lambda_2 - \lambda_1) = 0$$

  - As the eigenvalues must be distinct, $\mathbf{u}^T \mathbf{v} = 0$

- What is the effects of a linear transformation on ?
  - A space ?

- What is the effects of a linear transformation on ?
  - A space ? Rotation, Scaling, Shear or Combination
  - A vector ?

- What is the effects of a linear transformation on ?
  - A space ? Rotation, Scaling, Shear or Combination
  - A vector ? Rotation, Scaling or Combination
- Take the transformation

$$\begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$$

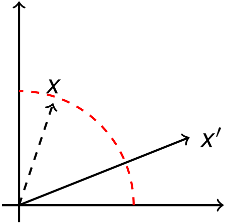


Figure: Example Transformation

- A transformation matrix **A** rotates and scales the vector.

- Give me a matrix **B** constructed using **A** which is a $m \times n$ matrix such that **B** is guaranteed to be symmetric.

- Give me a matrix $\mathbf{B}$ constructed using $\mathbf{A}$ which is a $m \times n$ matrix such that $\mathbf{B}$ is guaranteed to be symmetric.

$$(\mathbf{A}\mathbf{A}^T)^T = (\mathbf{A}^T)^T\mathbf{A}^T = \mathbf{A}\mathbf{A}^T$$

- How can we we use the eigenvectors of $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$ to aid the decomposition of $\mathbf{A}$? Assume that it's a fact that the eigenvectors of $\mathbf{A}\mathbf{A}^T$ span the whole space.

- For a symmetric matrix, what is the relation between eigenvalues and singular values ?

$$\mathbf{A}\mathbf{A}^T\mathbf{U} = \mathbf{U}\boldsymbol{\Sigma}^2$$

As $\mathbf{U}$ is an orthonormal matrix $\mathbf{U}^T = \mathbf{U}^{-1}$. Post multiplying with $\mathbf{U}^{-1}$, we have

$$\mathbf{A}\mathbf{A}^T = (\mathbf{U}\boldsymbol{\Sigma}^2\mathbf{U}^T)$$

$$\mathbf{A}\mathbf{A}^T = (\mathbf{U}\boldsymbol{\Sigma}\mathbf{I}\boldsymbol{\Sigma}\mathbf{U}^T)$$

As $\mathbf{V}^T\mathbf{V} = \mathbf{I}$, we can replace the identity matrix

$$\mathbf{A}\mathbf{A}^T = (\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T)(\mathbf{V}\boldsymbol{\Sigma}\mathbf{U}^T)$$

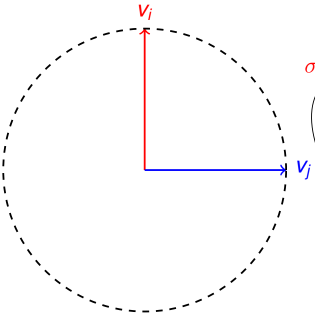
$$\mathbf{A}\mathbf{A}^T = (\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T)(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T)^T$$

- Similarly, we have

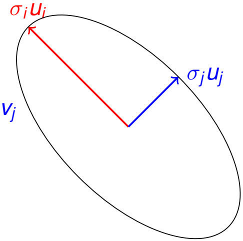
$$\mathbf{A}^T\mathbf{A}\mathbf{V} = \boldsymbol{\Sigma}^2\mathbf{V}$$

- This means that we can decompose any $m \times n$ matrix $\mathbf{A}$ through SVD

- Let's take an example in 2-D space. And we have an orthonormal basis $\mathbf{v}_1$ and $\mathbf{v}_2$

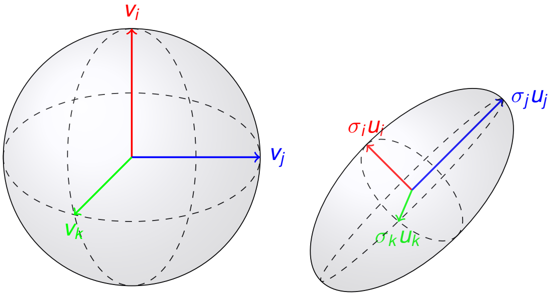


(a) Before transform



(b) After transform

- Let's take an example in 3-D space. And we have an orthonormal basis $\mathbf{v}_1$ and $\mathbf{v}_2$



(a) Before transform (b) After transform

- For any n-dimensional vector space, can we define a transformation $\mathbf{A}$ such that it will take the unit orthonormal basis $v_1, v_2, \ldots, v_n$ to a unit basis $u_1, u_2, \ldots, u_n$ with the scaling factor $\sigma_1, \sigma_2, \ldots, \sigma_n$

- For a n-dimensional transformation, we can write

$$\mathbf{A}\mathbf{v}_i = \sigma_i \mathbf{u}_i, j = 1, 2, \ldots, n$$

$$\begin{bmatrix} A \end{bmatrix} \underbrace{\begin{bmatrix} v_1 & v_2 & \ldots v_n \end{bmatrix}}_{\text{Orthonormal Basis of } \mathbb{R}^n} = \underbrace{\begin{bmatrix} u_1 & u_2 & \ldots u_n \end{bmatrix}}_{\text{Orthonormal basis of } \mathbb{R}^n} \begin{bmatrix} \sigma_1 & 0 & \ldots & 0 \\ 0 & \sigma_2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & \sigma_n \end{bmatrix}$$

- For any two spaces $\mathbb{R}^n$ and $\mathbb{R}^m$

$$\begin{bmatrix} A \end{bmatrix} \underbrace{\begin{bmatrix} v_1 & v_2 & \ldots v_n \end{bmatrix}}_{\text{Orthonormal Basis of } \mathbb{R}^n} = \underbrace{\begin{bmatrix} u_1 & u_2 & \ldots u_m \end{bmatrix}}_{\text{Orthonormal basis of } \mathbb{R}^m} \begin{bmatrix} \sigma_1 & 0 & \ldots & 0 \\ 0 & \sigma_2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & \sigma_n \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & \ldots & 0 \end{bmatrix}$$

$$\mathbf{A} \underbrace{\mathbf{V}}_{\text{Rotation}} = \underbrace{\mathbf{U}}_{\text{Rotation}} \underbrace{\mathbf{\Sigma}}_{\text{Scaling}}$$

- For orthonormal matrices, $\mathbf{V}^{-1} = \mathbf{V}^T$. Post-multiplying with $\mathbf{V}^{-1}$

$$\mathbf{A} = \underbrace{\mathbf{U}}_{\text{Left Singularvectors}} \mathbf{\Sigma} \underbrace{\mathbf{V}^T}_{\text{Right Singularvectors}}$$

- Let us take a shearing transformation

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \qquad \implies \mathbf{A}^T\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix} \qquad \mathbf{A}\mathbf{A}^T = \begin{bmatrix} 5 & 2 \\ 2 & 1 \end{bmatrix}$$

- Find the eigenvectors of **correlation matrices** $\mathbf{AA}^T$ and $\mathbf{A}^T\mathbf{A}$, $\lambda'_1 = 5.82$ and $\lambda'_2 = 0.17$
- Use them to get the singular values of $\mathbf{A}$, $\sigma_1 = 2.41$ and $\sigma_2 = 0.41$
- Find the eigenvectors of $\mathbf{A}^T\mathbf{A}$, they are the column vectors of $\mathbf{V}$

$$\mathbf{V} = \begin{bmatrix} 0.38 & -0.92 \\ 0.92 & 0.38 \end{bmatrix}$$

- Find the eigenvectors of $\mathbf{AA}^T$, they are the column vectors of $\mathbf{U}$

$$\mathbf{U} = \begin{bmatrix} 0.92 & -0.38 \\ 0.38 & 0.92 \end{bmatrix}$$

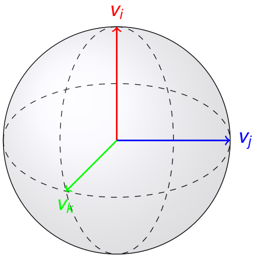
- Singular Value Decomposition of $\mathbf{A}$ is

$$\underbrace{\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}}_{\text{Shear}} = \underbrace{\begin{bmatrix} 0.92 & -0.38 \\ 0.38 & 0.92 \end{bmatrix}}_{\text{Rotation}} \underbrace{\begin{bmatrix} 2.41 & 0 \\ 0 & 0.41 \end{bmatrix}}_{\text{Scaling}} \underbrace{\begin{bmatrix} 0.38 & -0.92 \\ 0.92 & 0.38 \end{bmatrix}}_{\text{Rotation}}$$

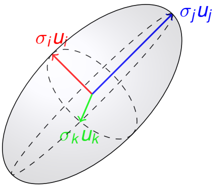
- Any matrix $\mathbf{A}$ Singular Value Decomposition can be factorized into a set of two orthogonal matrices ($\mathbf{U}$,$\mathbf{V}$) and a diagonal matrix $\mathbf{\Sigma}$. It follows directly from the fact that the matrix $\mathbf{A}\mathbf{A}^T$ is symmetric positive semi-definite

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$

- Geometrically, any linear transform can be broken down into a rotation followed by scaling followed by yet another rotation. Note that the first rotation and second rotation are in different vector space.



(a) Before transform



(b) After transform

- Let us take an equation $\mathbf{Ax} = \mathbf{b}$

$$\begin{bmatrix} 1 & 1 \\ 1 & 1.001 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

- Let us take an equation $\mathbf{Ax} = \mathbf{b}$

$$\begin{bmatrix} 1 & 1 \\ 1 & 1.001 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \qquad \begin{bmatrix} 1 & 1 \\ 0 & 0.001 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \qquad \Longrightarrow \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

- Let us perturb **b** a little bit.

$$\begin{bmatrix} 1 & 1 \\ 1 & 1.001 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ \mathbf{2.001} \end{bmatrix}$$

- Let us take an equation $\mathbf{Ax} = \mathbf{b}$

$$\begin{bmatrix}1 & 1\\ 1 & 1.001\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix} = \begin{bmatrix}2\\ 2\end{bmatrix} \qquad \begin{bmatrix}1 & 1\\ 0 & 0.001\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix} = \begin{bmatrix}2\\ 0\end{bmatrix} \qquad \implies \begin{bmatrix}x\\ y\end{bmatrix} = \begin{bmatrix}2\\ 0\end{bmatrix}$$

- Let us perturb **b** a little bit.

$$\begin{bmatrix}1 & 1\\ 1 & 1.001\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix} = \begin{bmatrix}2\\ \mathbf{2.001}\end{bmatrix} \qquad \begin{bmatrix}1 & 1\\ 0 & 0.001\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix} = \begin{bmatrix}2\\ 0.001\end{bmatrix} \qquad \implies \begin{bmatrix}x\\ y\end{bmatrix} = \begin{bmatrix}\mathbf{1}\\ \mathbf{1}\end{bmatrix}$$

- Notice that a **tiny change** in **b** resulted in **large change** in solution
- If we calculate the eigenvectors of **A**, we will find something interesting

$$det\left(\begin{bmatrix}1-\lambda & 1\\ 1 & 1.001-\lambda\end{bmatrix}\right) = 0 \qquad \implies \lambda^2 - 2.001\lambda + 0.001 = 0$$

$$\implies \lambda = \frac{2.001 + -\sqrt{(2.001)^2 - 4 * (0.001)}}{2} \qquad \implies \lambda = \frac{2.001 + -2}{2} \qquad \implies \lambda \cong 0, 2$$

- Note that the eigenvalues are not exactly 0 and 2. In fact, they are slightly more than 0 and 2 respectively. Because of this, the ratio between the two eigenvalues is very large.

$$\frac{\lambda_1}{\lambda_2} = \frac{2}{0.0000\ldots}$$

Condition Number

- If **A** is a square matrix with real eigenvalues such that $|\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_n|$. If $\frac{\lambda_1}{\lambda_n}$ is huge, then **A** is a bad matrix.
- The ratio between the largest and the smallest eigenvalue of a square matrix is called it's **condition number**

- What about non-symmetric matrices ? Given a $m \times n$ matrix **A** with eigenvalues σ_i's such that $|\sigma_1| \geq |\sigma_2| \geq \ldots \geq |\sigma_n|$, then the condition number $C(A)$ is $\frac{\sigma_1}{\sigma_n}$

$$C(\mathbf{A}) = \sqrt{C(\mathbf{A}\mathbf{A}^T)}$$

- Takeaway : Small $C(\mathbf{A})$ ensures that solving $\mathbf{Ax} = \mathbf{b}$ is not too sensitive on the value of **b**
- Read : Why batchnorm really works ? Link

- Storing an image so that it takes up lesser space or transferring an image so that it takes less bandwidth

$$\mathbf{I} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$

- Let us arrange the diagonal entries such that $\Sigma_{11} \geq \Sigma_{22} \geq \ldots \geq \Sigma_{nn}$. Further, rearrange the corresponding rows and columns within $\mathbf{U}$ and $\mathbf{V}^T$ respectively

$$\mathbf{I} = \underbrace{\mathbf{U}_1\mathbf{\Sigma}_1\mathbf{V}_1^T}_{\text{Most Significant}} + \mathbf{U}_2\mathbf{\Sigma}_2\mathbf{V}_2^T + \ldots + \underbrace{\mathbf{U}_n\mathbf{\Sigma}_n\mathbf{V}_n^T}_{\text{Least Significant}}$$

- Just omit the term with higher indices.

$$\mathbf{I} = \mathbf{U}_1\mathbf{\Sigma}_1\mathbf{V}_1^T + \mathbf{U}_2\mathbf{\Sigma}_2\mathbf{V}_2^T + \ldots + \mathbf{U}_5\mathbf{\Sigma}_5\mathbf{V}_5^T$$

- Other applications
  - 3^{rd} Prize in Netflix Challenge ▸ Link ▸ Link
  - Truncated SVD ▸ Link
  - Calculating the pseudo-inverse ▸ Link
  - PageRank using SVD ▸ Link
  - General Document on SVD ▸ Link
- Homework
  - Is SVD unique for a given matrix $\mathbf{A}$?
  - Why are the singular values of $\mathbf{AA}^T$ always greater than or equal to 0 ?

- Other applications
  - 3^{rd} Prize in Netflix Challenge Link Link
  - Truncated SVD Link
  - Calculating the pseudo-inverse Link
  - PageRank using SVD Link
  - General Document on SVD Link
- Homework
  - Is SVD unique for a given matrix **A** ?
  - Why are the singular values of $\mathbf{AA}^T$ always greater than or equal to 0 ?
    It is positive definite Link
  - How can you determine the rank of matrix using SVD ?

Thank You