ZOOMNOTES FOR LINEAR ALGEBRA

GILBERT STRANG
Massachusetts Institute of Technology

WELLESLEY - CAMBRIDGE PRESS
Box 812060 Wellesley MA 02482

ZoomNotes for Linear Algebra

ISBN 978-1-7331466-4-7

LaTeX typesetting by Ashley C. Fernandes

Printed in the United States of America 9 8 7 6 5 4 3 2 1

Texts from Wellesley - Cambridge Press

Linear Algebra for Everyone, 2020, Gilbert Strang ISBN 978-1-7331466-3-0
Linear Algebra and Learning from Data, 2019, Gilbert Strang ISBN 978-0-6921963-8-0
Introduction to Linear Algebra, 5th Ed., 2016, Gilbert Strang ISBN 978-0-9802327-7-6
Computational Science and Engineering, Gilbert Strang ISBN 978-0-9614088-1-7
Differential Equations and Linear Algebra, Gilbert Strang ISBN 978-0-9802327-9-0
Wavelets and Filter Banks, Gilbert Strang and Truong Nguyen ISBN 978-0-9614088-7-9
Introduction to Applied Mathematics, Gilbert Strang ISBN 978-0-9614088-0-0
Calculus Third Edition, Gilbert Strang ISBN 978-0-9802327-5-2
Algorithms for Global Positioning, Kai Borre & Gilbert Strang ISBN 978-0-9802327-3-8
Essays in Linear Algebra, Gilbert Strang ISBN 978-0-9802327-6-9
An Analysis of the Finite Element Method, 2008 edition, Gilbert Strang and George Fix ISBN 978-0-9802327-0-7

Wellesley - Cambridge Press
Box 812060, Wellesley MA 02482 USA
www.wellesleycambridge.com

Gilbert Strang's page : **math.mit.edu/∼gs**
For orders : **math.mit.edu/weborder.php**
Outside US/Canada : **www.cambridge.org**
Select books, India : **www.wellesleypublishers.com**

The textbook websites are **math.mit.edu/linearalgebra** and **math.mit.edu/everyone**.

Those sites link to 18.06 course materials and video lectures on YouTube and OCW.
Solution Manuals can be printed from those sites and **math.mit.edu/learningfromdata**.

Linear Algebra is included in MIT's OpenCourseWare site **ocw.mit.edu/courses**.
This provides video lectures of the full linear algebra courses 18.06 and 18.06 SC and 18.065.

ZoomNotes for Linear Algebra : Gilbert Strang

Preface

The title **"ZoomNotes"** indicates that these pages were created in 2020 and 2021. But they are not limited to online lectures. I hope these notes will help instructors and students to see linear algebra in an organized way, from vectors to matrices to subspaces to bases. "Linear independence" is a crucial idea for this subject, so it comes early—for vectors of integers.

I hope that faculty who are planning a linear algebra course and students who are reading for themselves will see these notes.

A happy part of linear algebra is the wonderful variety of matrices—diagonal, triangular, symmetric, orthogonal, and many more. The organizing principles have become matrix factorizations like $A = LU$ (lower triangular times upper triangular). The idea of elimination—to simplify the equations $A\boldsymbol{x} = \boldsymbol{b}$ by introducing zeros in the matrix—appears early as it must. Please don't spend forever on those computations. Linear algebra has so many more good ideas.

The reader may know my video lectures on OpenCourseWare : Math 18.06 is on **ocw.mit.edu** and on **Youtube/mitocw**. I am so grateful that those have been helpful. Now I have realized that lecture notes can help in a different way. You will quickly gain a picture of the whole course—the structure of the subject, the key topics in a natural order, the connecting ideas that make linear algebra so beautiful. This structure is the basis of two textbooks from Wellesley-Cambridge Press :

Introduction to Linear Algebra **Linear Algebra for Everyone**

I don't try to teach every topic in those books. I do try to reach eigenvalues and singular values ! A basis of eigenvectors for square matrices—and of singular vectors for all matrices—takes you to the heart of a matrix in a way that elimination cannot do.

The last chapters of these notes extend to a third book and a second math course 18.065 with videos on OpenCourseWare :

Linear Algebra and Learning from Data (Wellesley-Cambridge Press 2019)

This is "Deep Learning" and it is not entirely linear. It creates a learning function $\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{v})$ from training data $\boldsymbol{v}$ (like images of handwritten numbers) and matrix weights $\boldsymbol{x}$. The piecewise linear "ReLU function" plays a mysterious but crucial part in $\boldsymbol{F}$. Then $\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{v}_{\text{new}})$ can come close to new data that the system has never seen.

The learning function $\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{v})$ grows out of linear algebra and optimization and statistics and high performance computing. Our aim is to understand (in part) why it succeeds.

Above all, I hope these ZoomNotes help you to teach linear algebra and learn linear algebra. This subject is used in so many valuable ways. And it rests on ideas that everyone can understand.

Thank you. **Gilbert Strang**

Textbooks, ZoomNotes, and Video Lectures

Introduction to Linear Algebra, 5th Ed. (2016)	**math.mit.edu/linearalgebra**
Linear Algebra and Learning from Data (2019)	**math.mit.edu/learningfromdata**
Linear Algebra for Everyone (2020)	**math.mit.edu/everyone**
Differential Equations and Linear Algebra (2014)	**math.mit.edu/dela**
ZoomNotes for Linear Algebra (2021)	

Video Lectures **OpenCourseWare** **ocw.mit.edu/courses** **youtube/mitocw**

Math 18.06 and 18.06SC	**Linear Algebra at MIT**
(added to 18.06)	**A 2020 Vision of Linear Algebra**
Math 18.065	**Linear Algebra and Learning from Data**
Math 18.085 and 18.086	**Computational Science and Engineering**
Strang and Moler	**Differential Equations and Linear Algebra**

Interview with Lex Fridman **https://www.youtube.com/watch?v=lEZPfmGCEk0**

Wellesley - Cambridge Press
Box 812060, Wellesley MA 02482 USA
www.wellesleycambridge.com
Orders : **math.mit.edu/weborder.php**

Gilbert Strang's page : **math.mit.edu/∼gs**

Outside US/Canada : **www.cambridge.org**
Select books, India : **www.wellesleypublishers.com**

Three Great Factorizations : LU, QR, SVD

Orthogonal matrix
$Q^{\mathrm{T}}Q = I$
Square $QQ^{\mathrm{T}} = I$

$$Q = \begin{bmatrix} \\ q_1 & q_2 & \cdots & q_n \\ \\ \end{bmatrix}$$

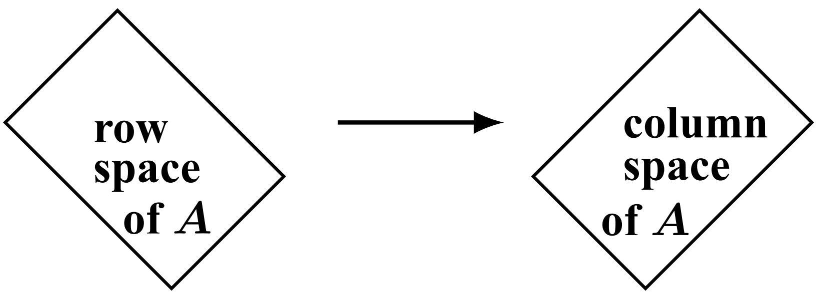
Orthogonal basis

Triangular matrix
$R_{ij} = 0$ **for** $i > j$
$R_{jj} \neq 0$ **on diagonal**

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdot & r_{1n} \\ & r_{22} & \cdot & r_{2n} \\ & & \cdot & \cdot \\ & & & r_{nn} \end{bmatrix}$$

Triangular basis

1. $A = LU =$ **(lower triangular) (upper triangular): Elimination**
2. $A = QR =$ **(orthogonal) (upper triangular): Gram-Schmidt**
3. $A = U\Sigma V^{\mathrm{T}} =$ **(orthogonal)(diagonal)(orthogonal): Singular values**
   Chapters 2, 4, 7



U **and** R **and** V **input basis**

L **and** Q **and** U **output basis**

Part 1

Basic Ideas of Linear Algebra

Part 1 : Basic Ideas of Linear Algebra

1.1 Linear Combinations of Vectors

A 3-dimensional vector $\boldsymbol{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$ has 3 components v_1, v_2, v_3 as in $\boldsymbol{v} = \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix}$

$\boldsymbol{v}$ gives a point in 3-dimensional space $\mathbf{R}^3$. Think of an arrow from $(0,0,0)$ to $(2,4,1)$.

We add vectors $\boldsymbol{v} + \boldsymbol{w}$. We multiply them by numbers like $c = \mathbf{4}$ and $d = \mathbf{0}$ (called scalars)

$$\begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} + \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix} = \begin{bmatrix} \mathbf{5} \\ \mathbf{4} \\ \mathbf{3} \end{bmatrix} \qquad \mathbf{4} \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} = \begin{bmatrix} \mathbf{12} \\ \mathbf{16} \\ \mathbf{20} \end{bmatrix} \qquad \mathbf{0} \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} = \textbf{zero vector}$$

Linear combinations $\quad 2\boldsymbol{v} - 3\boldsymbol{w}$ **and** $c\boldsymbol{v} + d\boldsymbol{w}$ **and** $\boldsymbol{w} - 2\boldsymbol{z} + \boldsymbol{u}$

$$\mathbf{2} \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} - \mathbf{3} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} \mathbf{3} \\ \mathbf{2} \\ \mathbf{1} \end{bmatrix} \qquad \mathbf{1} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \mathbf{2} \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} + \mathbf{1} \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

$\boxed{\textbf{Allow every } \boldsymbol{c, d} \text{ or } \textbf{all } \boldsymbol{c, d, e}}$ All combinations of $\boldsymbol{v}$ and $\boldsymbol{w}$ usually (!) fill a plane in $\mathbf{R}^3$

All $c \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} + d \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ fill a **plane** $\qquad$ All $c \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + d \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} + e \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ fill **3D space** $\mathbf{R}^3$

Sometimes a combination gives the zero vector. Then the vectors are **dependent**.

All $c \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} + d \begin{bmatrix} 6 \\ 8 \\ 10 \end{bmatrix}$ only fill a line. They are all multiples of $\begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}$. This includes $\begin{bmatrix} \mathbf{-3} \\ \mathbf{-4} \\ \mathbf{-5} \end{bmatrix}$

All $c \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + d \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} + e \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix}$ only fill a plane and not 3D space. That third vector is nothing new. $\begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix}$ is $2 \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} - \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

1.2 Dot Products $\boldsymbol{v} \cdot \boldsymbol{w}$ and Lengths $||\boldsymbol{v}||$ and Angles $\boldsymbol{\theta}$

Dot product $\boldsymbol{v} \bullet \boldsymbol{w} = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} \bullet \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} = \begin{matrix} 3 \times 2 \\ 4 \times 0 \\ \underline{5 \times 1} \\ \textbf{add} \end{matrix} = \mathbf{11}$

$$\boxed{\boldsymbol{v} \bullet \boldsymbol{w} = \boldsymbol{w} \bullet \boldsymbol{v}}$$

$$\begin{bmatrix} a \\ b \end{bmatrix} \bullet \begin{bmatrix} c \\ d \end{bmatrix} = \boldsymbol{ac + bd}$$

Length squared of $\boldsymbol{v} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$ is $||\boldsymbol{v}||^2 = 3^2 + 4^2 = \mathbf{9 + 16}$. This is Pythagoras $c^2 = a^2 + b^2$

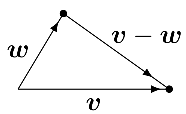
Length squared of $\boldsymbol{v}$ is $||\boldsymbol{v}||^2 = \boldsymbol{v} \cdot \boldsymbol{v} = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} = 9 + 16 + 25$ (Pythagoras in $\mathbf{3}$D)

Length squared of $\boldsymbol{v} + \boldsymbol{w}$ is $(\boldsymbol{v} + \boldsymbol{w}) \cdot (\boldsymbol{v} + \boldsymbol{w}) = \boldsymbol{v} \cdot \boldsymbol{v} + \boldsymbol{v} \cdot \boldsymbol{w} + \boldsymbol{w} \cdot \boldsymbol{v} + \boldsymbol{w} \cdot \boldsymbol{w}$

$$\boxed{||\boldsymbol{v} + \boldsymbol{w}||^2 = ||\boldsymbol{v}||^2 + ||\boldsymbol{w}||^2 + 2\boldsymbol{v} \cdot \boldsymbol{w}} \qquad \boxed{||\boldsymbol{v} + \boldsymbol{w}|| \le ||\boldsymbol{v}|| + ||\boldsymbol{w}||}$$

$\boldsymbol{v} = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}$ $\quad \boldsymbol{w} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$ $\quad \boldsymbol{v} + \boldsymbol{w} = \begin{bmatrix} 4 \\ 4 \\ 4 \end{bmatrix}$ Length squared of $\boldsymbol{v} + \boldsymbol{w}$ is $4^2 + 4^2 + 4^2$ $\quad 48$ is $\mathbf{50 + 2 + 2}\, \boldsymbol{v} \cdot \boldsymbol{w}$

Triangle has edges $\boldsymbol{v}, \boldsymbol{w}, \boldsymbol{v} - \boldsymbol{w}$



$$||\boldsymbol{v} - \boldsymbol{w}||^2 = ||\boldsymbol{v}||^2 + ||\boldsymbol{w}||^2 - 2\boldsymbol{v} \cdot \boldsymbol{w}$$

The dot product $\boldsymbol{v} \cdot \boldsymbol{w}$ reveals the angle $\boldsymbol{\theta}$ between $\boldsymbol{v}$ and $\boldsymbol{w}$
$|\cos \boldsymbol{\theta}| \le \mathbf{1}$ is one way to see the Schwarz inequality

$$\boxed{\begin{aligned} \boldsymbol{v} \cdot \boldsymbol{w} &= ||\boldsymbol{v}||\, ||\boldsymbol{w}|| \cos \theta \\ |\boldsymbol{v} \cdot \boldsymbol{w}| &\le ||\boldsymbol{v}||\, ||\boldsymbol{w}|| \end{aligned}}$$

The angle between $\boldsymbol{v} = \begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix}$ and $\boldsymbol{w} = \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}$ is $\boldsymbol{\theta} = \mathbf{90°}$ because $\boxed{\boldsymbol{v} \cdot \boldsymbol{w} = \mathbf{0} : \textbf{Perpendicular}}$

The angle between $\boldsymbol{v} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\boldsymbol{w} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is $\boldsymbol{\theta} = \mathbf{45°}$ because $\boldsymbol{v} \cdot \boldsymbol{w} = 1$ and $||\boldsymbol{v}||\, ||\boldsymbol{w}|| = \sqrt{2}$.

1.3 Matrices Multiplying Vectors

There is a **row way** to multiply $A\boldsymbol{x}$ and also a **column way** to compute the vector $A\boldsymbol{x}$

$$\boxed{\textbf{Row way} = \textbf{Dot product} \text{ of vector } \boldsymbol{x} \text{ with each row of } A}$$

$$A\boldsymbol{x} = \begin{bmatrix} 2 & 5 \\ 3 & 7 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 2v_1 + 5v_2 \\ 3v_1 + 7v_2 \end{bmatrix} \qquad \begin{bmatrix} 2 & 5 \\ 3 & 7 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 7 \\ 10 \end{bmatrix}$$

$$\boxed{\textbf{Column way} = A\boldsymbol{x} \text{ is a } \textbf{combination of the columns of } \boldsymbol{A}}$$

$$A\boldsymbol{x} = \begin{bmatrix} 2 & 5 \\ 3 & 7 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = v_1 \begin{bmatrix} \text{column} \\ 1 \end{bmatrix} + v_2 \begin{bmatrix} \text{column} \\ 2 \end{bmatrix} \qquad \begin{bmatrix} 2 & 5 \\ 3 & 7 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 5 \\ 7 \end{bmatrix} = \begin{bmatrix} \mathbf{7} \\ \mathbf{10} \end{bmatrix}$$

Which way to choose? Dot products with **rows** or combination of **columns**?

For computing with numbers, I use the row way : dot products

For understanding with vectors, I use the column way : combine columns

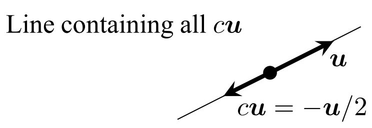
Same result $A\boldsymbol{x}$ from the same multiply-adds. Just in a different order

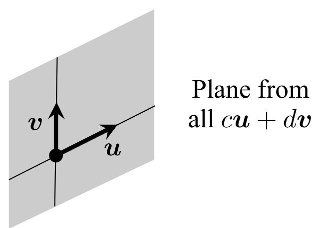
$$\boxed{\mathbf{C}(A) = \textbf{Column space of } \boldsymbol{A} = \textbf{all combinations of the columns} = \textbf{all outputs } \boldsymbol{A\boldsymbol{x}}}$$

The **identity matrix** has $I\boldsymbol{x} = \boldsymbol{x}$ for every $\boldsymbol{x}$ $\quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

The column space of the 3 by 3 identity matrix I is the whole space $\mathbf{R}^3$.

If all columns are multiples of column 1 (not zero), the column space $\mathbf{C}(A)$ is a line.





1.4 Column Space and Row Space of A

The column space of A contains all linear combinations of the columns of A

All the vectors $A\boldsymbol{x}$ (for all $\boldsymbol{x}$) fill the column space $\mathbf{C}(A)$: line or plane or . . .

If $\boldsymbol{v}$ is in $\mathbf{C}(A)$ so is every $c\,\boldsymbol{v}$. [Reason : $\boldsymbol{v} = A\boldsymbol{x}$ gives $c\boldsymbol{v} = A(c\,\boldsymbol{x})$]

If $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are in $\mathbf{C}(A)$ so is $\boldsymbol{v}_1 + \boldsymbol{v}_2$ [$\boldsymbol{v}_1 = A\boldsymbol{x}_1$ and $\boldsymbol{v}_2 = A\boldsymbol{x}_2$ give $\boldsymbol{v}_1 + \boldsymbol{v}_2 = A(\boldsymbol{x}_1 + \boldsymbol{x}_2)$]

The column spaces of $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$ and $\begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}$ are the whole $\mathbf{R}^2$

The column spaces of $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}$ are lines inside 2-dimensional space

The column space of $Z = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ has $\mathbf{C}(Z) =$ only one point $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

The **row space** of A contains all combinations of the rows of A

To stay with column vectors, **transpose A to make its rows into columns of A^{T}**

Then the row space of A is the column space of A^{T} (A transpose)

The column space of $A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 6 & 9 \end{bmatrix}$ is an infinite line in the direction of $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$

The row and column spaces of $A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 4 \\ 1 & 4 & 5 \end{bmatrix}$ are infinite planes. **Not all of $\mathbf{R}^3$**

The row and column spaces of $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}$ are the whole $\mathbf{R}^3$.

$A = \begin{bmatrix} 1 & 2 & 5 & 1 \\ 3 & 4 & 6 & 7 \end{bmatrix}$ has column space $= \mathbf{R}^2$
has row space $=$ 2D plane in $\mathbf{R}^4$

1.5 Dependent and Independent Columns

The columns of A are **"dependent"** if one column is a combination of the other columns

Another way to describe dependence : $A\boldsymbol{x} = \mathbf{0}$ for some vector $\boldsymbol{x}$ (other than $\boldsymbol{x} = \mathbf{0}$)

$$A_1 = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 1 & 2 \end{bmatrix} \text{ and } A_2 = \begin{bmatrix} 1 & 4 & 0 \\ 2 & 5 & 0 \\ 3 & 6 & 0 \end{bmatrix} \text{ and } A_3 = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \text{ have } \textbf{dependent columns}$$

Reasons : Column 2 of $A_1 = 2\,$(Column 1) $\quad A_2$ times $\boldsymbol{x} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ gives $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

A_3 has 3 columns in 2-dimensional space. Three vectors in a plane : **Dependent !**

The columns of A are **"independent"** if no column is a combination of the other columns

Another way to say it : $A\boldsymbol{x} = \mathbf{0}$ **only when** $\boldsymbol{x} = \mathbf{0}$

$$A_4 = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 9 \end{bmatrix} \text{ and } A_5 = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } A_6 = I \text{ have independent columns}$$

What about the **rows** of A_1 to A_6 ? A_1, A_2, A_4 have dependent rows. Possibly also A_3.

For any square matrix : Columns are independent if and only if rows are independent.

A key idea is coming	Number of independent rows = Number of independent columns

1.6 Matrix-Matrix Multiplication AB

There are **4 ways** to multiply matrices. The first way is usually best for hand computation. The other three ways produce whole vectors instead of just one number at a time.

1. (Row i of A) · (Column j of B) produces one number : row i, column j of AB

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} 5 & 7 \\ 6 & 8 \end{bmatrix} = \begin{bmatrix} \mathbf{17} & \cdot \\ \cdot & \cdot \end{bmatrix} \text{ because } \begin{bmatrix} 1 & 2 \end{bmatrix}\begin{bmatrix} 5 \\ 6 \end{bmatrix} = \mathbf{17} \quad \textbf{Dot product}$$

2. (Matrix A) (Column j of B) produces column j of AB : **Combine columns of A**

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} \mathbf{5} & 7 \\ \mathbf{6} & 8 \end{bmatrix} = \begin{bmatrix} \mathbf{17} & \cdot \\ \mathbf{39} & \cdot \end{bmatrix} \text{ because } \mathbf{5}\begin{bmatrix} 1 \\ 3 \end{bmatrix} + \mathbf{6}\begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} \mathbf{17} \\ \mathbf{39} \end{bmatrix}$$

This is the best way for understanding : **Linear combinations**. "Good level"

3. (Row i of A) (Matrix B) produces row i of AB : **Combine rows of B**

$$\begin{bmatrix} \mathbf{1} & \mathbf{2} \\ 3 & 4 \end{bmatrix}\begin{bmatrix} 5 & 7 \\ 6 & 8 \end{bmatrix} = \begin{bmatrix} \mathbf{17} & \mathbf{23} \\ \cdot & \cdot \end{bmatrix} \text{ because } \mathbf{1}\begin{bmatrix} 5 & 7 \end{bmatrix} + \mathbf{2}\begin{bmatrix} 6 & 8 \end{bmatrix} = \begin{bmatrix} \mathbf{17} & \mathbf{23} \end{bmatrix}$$

4. (Column k of A) (Row k of B) produces a simple matrix : Add these simple matrices !

$$\begin{bmatrix} 1 \\ 3 \end{bmatrix}\begin{bmatrix} 5 & 7 \end{bmatrix} = \begin{bmatrix} 5 & 7 \\ 15 & 21 \end{bmatrix} \text{ and } \begin{bmatrix} 2 \\ 4 \end{bmatrix}\begin{bmatrix} 6 & 8 \end{bmatrix} = \begin{bmatrix} 12 & 16 \\ 24 & 32 \end{bmatrix} \quad \begin{matrix} \textbf{NOW} \\ \textbf{ADD} \end{matrix} \quad \begin{bmatrix} \mathbf{17} & \mathbf{23} \\ \mathbf{39} & \mathbf{53} \end{bmatrix} = \boldsymbol{AB}$$

Dot products in **1** are "inner products". Column-row products in **4** are "outer products".

All four ways use the **same mnp multiplications** if A is m **by** n and B is n **by** p.

If A and B are square n by n matrices then AB uses n^3 **multiply-adds in 1, 2, 3, 4**.

Associative Law	$\boldsymbol{A}$ **times** $\boldsymbol{BC} = \boldsymbol{AB}$ **times** $\boldsymbol{C}$	Most important rule !

Block multiplication
Block sizes must fit

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}\begin{bmatrix} E \\ F \end{bmatrix} = \begin{bmatrix} AE + BF \\ CE + DF \end{bmatrix}$$

1.7 Factoring A into CR: Column rank $= r =$ Row rank

Step 1 C contains the first r independent columns of A (delete dependent columns of A)

1. If column 1 of A is not zero, put it into C
2. If column 2 of A is not a multiple of column 1 of A, put it into C
3. If column 3 of A is not a combination of columns 1 and 2 of A, put it into C

n. If column n of A is not a combination of the first $n-1$ columns, put it into C

Step 2 Column j of CR expresses column j of A as a combination of the columns of C

Example $A = \begin{bmatrix} 1 & 2 & 4 \\ 1 & 3 & 5 \end{bmatrix}$ Columns 1 and 2 of A go directly into C
Column 3 $=$ **2** (Column 1) $+$ **1** (Column 2) Not in C

$$A = \begin{bmatrix} 1 & 2 & \mathbf{4} \\ 1 & 3 & \mathbf{5} \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 & \mathbf{2} \\ 0 & 1 & \mathbf{1} \end{bmatrix} = CR \qquad \begin{array}{c} \text{2 columns in } C \\ \text{2 rows in } R \end{array}$$

These matrices A, C, R all have **column rank 2** (2 independent columns)

By the theorem A, C, R also have **row rank 2** (2 independent rows)

First great theorem $\boxed{\textbf{Every matrix has column rank = row rank}}$

Dimension r of the column space $=$ Dimension r of the row space $=$ Rank of matrix A

$$A = (m \text{ by } n) = CR = (m \text{ by } r)\,(r \text{ by } n)$$

1.8 Rank one matrices $\quad A = (1 \text{ column}) \text{ times } (1 \text{ row})$

Rank one Example $\quad \boldsymbol{A} = \begin{bmatrix} \mathbf{2} & \mathbf{4} & \mathbf{6} \\ \mathbf{3} & 6 & 9 \end{bmatrix} = \begin{bmatrix} \mathbf{2} \\ \mathbf{3} \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{2} & \mathbf{3} \end{bmatrix} = \boldsymbol{CR}$

> **Suppose all columns of $\boldsymbol{A}$ are multiples of one column.**
> **Then all rows of $\boldsymbol{A}$ are multiples of one row. Rank $= 1$.**

Row space is a line Column space is a line

If all columns of A are multiples of column 1, it goes into C.

If all rows of A are multiples of row 1, that row (divided by a_{11}) goes into R.

Every rank 1 **matrix factors into one column times one row**.

Every rank $\boldsymbol{r}$ matrix is the sum of $\boldsymbol{r}$ rank one matrices.

This comes from *column times row multiplication* of C times R.

$*$ If A starts with a row or column of zeros, look at row 2 or column 2

$**$ Rank 1 matrices are the building blocks of all matrices

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 3 & 4 \\ 2 & 6 & 8 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 3 & 4 \end{bmatrix}$$

All the key factorizations of linear algebra add **columns times rows**

$$\boxed{\boldsymbol{A = CR} \qquad \boldsymbol{A = LU} \qquad \boldsymbol{A = QR} \qquad \boldsymbol{S = Q\Lambda Q^{\mathrm{T}}} \qquad \boldsymbol{A = U\Sigma V^{\mathrm{T}}}}$$

Those 5 factorizations are described in **Parts $\mathbf{1{+}3, 2, 4, 6, 7}$** of these ZoomNotes

Part 2

Solving Linear Equations
$Ax = b$: A is n by n

Part 2 : Solving Linear Equations $Ax = b$: A is n by n

2.1 Inverse Matrices A^{-1} and Solutions $x = A^{-1}b$

$$\boxed{\text{The inverse of a square matrix } A \text{ has } \boldsymbol{A^{-1}A = I} \text{ and } \boldsymbol{AA^{-1} = I}}$$

2 by 2 $\quad \boldsymbol{A^{-1}} = \begin{bmatrix} \mathbf{2} & \mathbf{1} \\ \mathbf{5} & \mathbf{4} \end{bmatrix}^{-1} = \frac{1}{3}\begin{bmatrix} \mathbf{4} & \mathbf{-1} \\ \mathbf{-3} & \mathbf{2} \end{bmatrix} \qquad \begin{bmatrix} \boldsymbol{a} & \boldsymbol{b} \\ \boldsymbol{c} & \boldsymbol{d} \end{bmatrix}^{-1} = \frac{1}{\boldsymbol{ad - bc}}\begin{bmatrix} \boldsymbol{d} & \boldsymbol{-b} \\ \boldsymbol{-c} & \boldsymbol{a} \end{bmatrix}$

$\boldsymbol{A}$ has no inverse if $\boldsymbol{ad - bc = 0}$ $\quad A = \begin{bmatrix} \mathbf{1} & \mathbf{2} \\ \mathbf{4} & \mathbf{8} \end{bmatrix}$ has no inverse matrix
has dependent rows
has dependent columns

1. **Invertible $\Leftrightarrow$ Rows are independent $\Leftrightarrow$ Columns are independent**
2. **No zeros on the main diagonal $\Leftrightarrow$ Triangular matrix is invertible**
3. If $BA = I$ and $AC = I$ then $\boldsymbol{B} = B(AC) = (BA)C = \boldsymbol{C}$
4. Invertible $\Leftrightarrow$ **The only solution to $\boldsymbol{Ax = b}$ is $\boldsymbol{x = A^{-1}b}$**
5. Invertible $\Leftrightarrow$ determinant is not zero $\Leftrightarrow$ $\boldsymbol{A^{-1}} = [\text{cofactor matrix}]^{\mathrm{T}} / \det A$
6. **Inverse of $\boldsymbol{AB} = \boldsymbol{B^{-1}}$ times $\boldsymbol{A^{-1}}$** (need both inverses) $\quad \boxed{ABB^{-1}A^{-1} = I}$
7. Computing A^{-1} is *not efficient* for $A\boldsymbol{x} = \boldsymbol{b}$. Use 2.3: *elimination*.

2.2 Triangular Matrix and Back Substitution for $Ux = c$

Solve $Ux =$ $\begin{bmatrix} \mathbf{2} & 3 & 4 \\ 0 & \mathbf{5} & 6 \\ 0 & 0 & \mathbf{7} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 19 \\ 17 \\ 14 \end{bmatrix} = \boldsymbol{c}$ **without finding U^{-1}**

Upper triangular U / Pivots $2, 5, 7$ are not zero / Go from bottom to top

Back substitution The last equation $7x_3 = 14$ gives $\boldsymbol{x_3 = 2}$

Work upwards The next equation $5x_2 + 6(2) = 17$ gives $\boldsymbol{x_2 = 1}$

Upwards again The first equation $2x_1 + 3(1) + 4(2) = 19$ gives $\boldsymbol{x_1 = 4}$

Conclusion The only solution to this example is $\boldsymbol{x} = (\mathbf{4}, \mathbf{1}, \mathbf{2})$

Special note To solve for x_3, x_2, x_1 we divided by the pivots $7, 5, 2$

A **zero pivot in U** produces dependent rows, dependent columns, **no U^{-1}**

Inverse of this difference matrix = sum matrix

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

Calculus : Inverse of derivative is integral $\displaystyle\int_0^x \frac{df}{dx}\,dx = f(x) - f(0)$

2.3 Elimination : Square Matrix A to Triangular U

$$\boldsymbol{A} = \begin{bmatrix} \mathbf{2} & 3 & 4 \\ 4 & 11 & 14 \\ 2 & 8 & 17 \end{bmatrix} \rightarrow \begin{bmatrix} \mathbf{2} & 3 & 4 \\ \mathbf{0} & \mathbf{5} & 6 \\ 2 & 8 & 17 \end{bmatrix} \rightarrow \begin{bmatrix} \mathbf{2} & 3 & 4 \\ \mathbf{0} & \mathbf{5} & 6 \\ \mathbf{0} & 5 & 13 \end{bmatrix} \rightarrow \begin{bmatrix} \mathbf{2} & 3 & 4 \\ \mathbf{0} & \mathbf{5} & 6 \\ \mathbf{0} & \mathbf{0} & \mathbf{7} \end{bmatrix} = \boldsymbol{U}$$

One elimination step subtracts $\boldsymbol{\ell_{ij}}$ times row $\boldsymbol{j}$ from row $\boldsymbol{i}$ $(i > j)$

Each step produces a zero below the diagonal of U : $\ell_{21} = 2, \ell_{31} = \ell_{32} = 1$

To invert elimination, add $\boldsymbol{\ell_{ij}}$ times row $\boldsymbol{j}$ back to row $\boldsymbol{i}$

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

$$\boxed{\boldsymbol{A} = \boldsymbol{LU} = (\textbf{Lower triangular } \boldsymbol{L}) \text{ times } (\textbf{Upper triangular } \boldsymbol{U})}$$

This A needs 3 elimination steps to a beautiful result

$$\boldsymbol{A} = \begin{bmatrix} 2 & 3 & 4 \\ 4 & 11 & 14 \\ 2 & 8 & 17 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ \mathbf{2} & 1 & 0 \\ \mathbf{1} & \mathbf{1} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{2} & \mathbf{3} & \mathbf{4} \\ 0 & \mathbf{5} & \mathbf{6} \\ 0 & 0 & \mathbf{7} \end{bmatrix} = \boldsymbol{L} \textbf{ times } \boldsymbol{U}$$

Elimination produced no zeros on the diagonal and created 3 zeros in U

For $\boldsymbol{Ax = b}$ Add extra column $\boldsymbol{b}$ Elimination and back substitution

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{b} \end{bmatrix} = \begin{bmatrix} 2 & 3 & 4 & \mathbf{19} \\ 4 & 11 & 14 & \mathbf{55} \\ 2 & 8 & 17 & \mathbf{50} \end{bmatrix} \rightarrow \begin{bmatrix} \boldsymbol{U} & \boldsymbol{c} \end{bmatrix} = \begin{bmatrix} \mathbf{2} & 3 & 4 & \mathbf{19} \\ 0 & \mathbf{5} & 6 & \mathbf{17} \\ 0 & 0 & \mathbf{7} & \mathbf{14} \end{bmatrix} \xrightarrow[\textbf{in 2.2}]{\textbf{backsub}} \boldsymbol{x} = \begin{bmatrix} \mathbf{4} \\ \mathbf{1} \\ \mathbf{2} \end{bmatrix}$$

2.4 Row Exchanges for Nonzero Pivots : Permutation P

If a diagonal pivot is zero or small : Look below it for a better pivot
Exchange rows

$$A = \begin{bmatrix} \mathbf{0} & \mathbf{2} \\ \mathbf{3} & \mathbf{4} \end{bmatrix} \text{ goes to } PA = \begin{bmatrix} \mathbf{0} & \mathbf{1} \\ \mathbf{1} & \mathbf{0} \end{bmatrix} \begin{bmatrix} 0 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} \mathbf{3} & \mathbf{4} \\ \mathbf{0} & \mathbf{2} \end{bmatrix} \quad \begin{array}{c} \text{Nonzero pivots} \\ \text{3 and 2} \end{array}$$

$$\boxed{\textbf{Permutation matrix } \boldsymbol{P} = \textbf{Rows of } \boldsymbol{I} \textbf{ in any order}}$$

There are $n\,!$ row orders and $n\,!$ permutations of size n (this includes $P = I$)

The inverse of P is the transpose of P
Exchange rows with columns

$$\begin{bmatrix} 0 & \mathbf{1} & 0 \\ 0 & 0 & \mathbf{1} \\ \mathbf{1} & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & \mathbf{1} \\ \mathbf{1} & 0 & 0 \\ 0 & \mathbf{1} & 0 \end{bmatrix} = \boldsymbol{I}$$

Can you find all six 3 by 3 permutations ? Is every $P_1 P_2 = P_2 P_1$?

If A is invertible then some PA has no zero pivots and $\boxed{\boldsymbol{PA = LU}}$

$$\begin{array}{l} \text{Reverse} \\ \text{the order} \\ \text{by } \boldsymbol{P} \end{array} \begin{bmatrix} 0 & 0 & 0 & \mathbf{1} \\ 0 & 0 & \mathbf{1} & 0 \\ 0 & \mathbf{1} & 0 & 0 \\ \mathbf{1} & 0 & 0 & 0 \end{bmatrix} \qquad \begin{array}{l} \text{Circular} \\ \text{shift} \\ \text{by } \boldsymbol{P} \end{array} \begin{bmatrix} 0 & 0 & 0 & \mathbf{1} \\ \mathbf{1} & 0 & 0 & 0 \\ 0 & \mathbf{1} & 0 & 0 \\ 0 & 0 & \mathbf{1} & 0 \end{bmatrix}$$

2.5 Elimination with No Row Exchanges : Why is $A = LU$?

Reason : Each step removes a column of L times a row of U

$$\text{Remove} \begin{bmatrix} 1 \text{ (row 1)} \\ \ell_{21} \text{ (row 1)} \\ \ell_{31} \text{ (row 1)} \\ \ell_{41} \text{ (row 1)} \end{bmatrix} \text{ from } A \text{ to leave } \boldsymbol{A_2} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & \times & \times & \times \end{bmatrix}$$

We removed a rank-one matrix : column times row. It was the column $\boldsymbol{\ell_1} = (1, \ell_{21}, \ell_{31}, \ell_{41})$ times row 1 of A—the first pivot row $\boldsymbol{u}_1$.

We face a similar problem for A_2. We take a similar step to A_3 :

$$\text{Remove} \begin{bmatrix} 0 \text{ (row 2 of } A_2) \\ 1 \text{ (row 2 of } A_2) \\ \ell_{32} \text{ (row 2 of } A_2) \\ \ell_{42} \text{ (row 2 of } A_2) \end{bmatrix} \text{ from } \boldsymbol{A_2} \text{ to leave } \boldsymbol{A_3} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \times & \times \\ 0 & 0 & \times & \times \end{bmatrix}$$

Row 2 of A_2 was the second pivot row = second row $\boldsymbol{u}_2$ of U. We removed a column $\boldsymbol{\ell}_2 = (0, 1, \ell_{32}, \ell_{42})$ times $\boldsymbol{u}_2$. Continuing this way, every step removes a column $\boldsymbol{\ell}_j$ times a pivot row $\boldsymbol{u}_j$ of U. Now put those pieces back :

$$\boxed{\boldsymbol{A} = \boldsymbol{\ell}_1\boldsymbol{u}_1 + \boldsymbol{\ell}_2\boldsymbol{u}_2 + \cdots + \boldsymbol{\ell}_n\boldsymbol{u}_n = \begin{bmatrix} \boldsymbol{\ell}_1 \cdots \boldsymbol{\ell}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{u}_1 \\ \vdots \\ \boldsymbol{u}_n \end{bmatrix} = \boldsymbol{LU}}$$

That last step was column-row multiplication (see 1.6) of L times U.

Column k of L and row k of U begin with $k-1$ zeros. Then L is lower triangular and U is upper triangular. Here are the separate elimination matrices—inverted and in reverse order to bring back the original A :

$$\boxed{\boldsymbol{L_{32}L_{31}L_{21}\,A = U} \quad \text{and} \quad \boldsymbol{A = L_{21}^{-1}L_{31}^{-1}L_{32}^{-1}\,U = LU}}$$

2.6 Transposes / Symmetric Matrices / Dot Products

Transpose of $\boldsymbol{A} = \begin{bmatrix} \mathbf{1} & \mathbf{2} & \mathbf{3} \\ 0 & 0 & 4 \end{bmatrix}$ is $\boldsymbol{A}^{\mathbf{T}} = \begin{bmatrix} \mathbf{1} & 0 \\ \mathbf{2} & 0 \\ \mathbf{3} & 4 \end{bmatrix}$ $\boxed{(\boldsymbol{A}^{\mathbf{T}})_{\boldsymbol{ij}} = \boldsymbol{A}_{\boldsymbol{ji}}}$

Rules for the sum and product

Transpose of $A + B$ is $\boldsymbol{A}^{\mathbf{T}} + \boldsymbol{B}^{\mathbf{T}}$
Transpose of AB is $\boldsymbol{B}^{\mathbf{T}}\boldsymbol{A}^{\mathbf{T}}$

A symmetric matrix has $\boldsymbol{S}^{\mathbf{T}} = \boldsymbol{S}$ This means that every $\boldsymbol{s}_{\boldsymbol{ij}} = \boldsymbol{s}_{\boldsymbol{ji}}$

The matrices $A^{\mathrm{T}}A$ and AA^{T} are symmetric / usually different

$$\begin{bmatrix} \mathbf{2} & \mathbf{3} \end{bmatrix}\begin{bmatrix} \mathbf{2} \\ \mathbf{3} \end{bmatrix} = \begin{bmatrix} \mathbf{13} \end{bmatrix} \qquad \begin{bmatrix} \mathbf{2} \\ \mathbf{3} \end{bmatrix}\begin{bmatrix} \mathbf{2} & \mathbf{3} \end{bmatrix} = \begin{bmatrix} \mathbf{4} & \mathbf{6} \\ \mathbf{6} & \mathbf{9} \end{bmatrix} \qquad \begin{aligned} \mathbf{4} + \mathbf{9} &= \mathbf{13} \\ \mathbf{6} &= \mathbf{6} \end{aligned}$$

$S = LU$ is improved to symmetric $\boldsymbol{S} = \boldsymbol{LDL}^{\mathbf{T}}$ (pivots in U go into D)

$$\boldsymbol{S} = \begin{bmatrix} 2 & 4 \\ 4 & 11 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 2 & 4 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} = \boldsymbol{LDU}^{\mathbf{T}}$$

Dot product	**Work**	= Movements • Forces = $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{f}$
Inner product	**Heat**	= Voltage drops • Currents = $\boldsymbol{e}^{\mathrm{T}}\boldsymbol{y}$
$\boldsymbol{x} \cdot \boldsymbol{y} = \boldsymbol{x}^{\mathbf{T}}\boldsymbol{y}$	**Income**	= Quantities • Prices = $\boldsymbol{q}^{\mathrm{T}}\boldsymbol{p}$

Part 3

Vector Spaces and Subspaces
Basis and Dimension

Part 3 : Vector Spaces and Subspaces Basis and Dimension

3.1 Vector Spaces and Four Fundamental Subspaces

$\boxed{\textbf{Vector space}: \text{Linear combinations of vectors in } \boldsymbol{S} \textbf{ must stay in } \boldsymbol{S}}$

$\boldsymbol{S} = \mathbf{R}^{\boldsymbol{n}}$ or **Subspace of** $\mathbf{R}^{\boldsymbol{n}}$, $\boldsymbol{S} =$ **matrices** $\mathbf{R}^{\boldsymbol{m \times n}}$ or **functions** $\boldsymbol{ax+b}$

Not vector spaces Half-line $x \geq 0$, invertible matrices, singular matrices

Subspaces All of $\mathbf{R}^3$, planes or lines through $(0,0,0)$, one point $(0,0,0)$

Four Subspaces	**Column space** $\mathbf{C}(A)$ $=$ all vectors $A\boldsymbol{x}$
	Row space $\mathbf{C}(A^{\mathrm{T}})$ $=$ all vectors $A^{\mathrm{T}}\boldsymbol{y}$
Column space $=$ **"range"**	**Nullspace** $\mathbf{N}(A)$ $=$ all $\boldsymbol{x}$ with $A\boldsymbol{x} = \mathbf{0}$
Nullspace $=$ **"kernel"**	**Left nullspace** $\mathbf{N}(A^{\mathrm{T}})$ $=$ all $\boldsymbol{y}$ with $A^{\mathrm{T}}\boldsymbol{y} = \mathbf{0}$

Any set of vectors ***spans*** a vector space. It contains ***all their combinations***

3.2 Basis and Dimension of a Vector Space S

Basis $=$ A set of **independent vectors** that **span** the space S

Every vector in S is a **unique combination** of those basis vectors

Dimension of S $=$ The number of vectors in any basis for S

All bases contain the same number of vectors

The column space of $A = \begin{bmatrix} \mathbf{1} & 0 & 0 \\ 0 & \mathbf{1} & 0 \\ 0 & 0 & 0 \end{bmatrix}$ is the x-y plane in $\mathbf{R}^3$

The first two columns are a basis for $\mathbf{C}(A)$

Another basis for $\mathbf{C}(A)$ consists of the vectors $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix}$

Infinitely many bases   Always 2 vectors   Dimension of $\mathbf{C}(A) = 2$

The **nullspace** of this A is the z-axis in $\mathbf{R}^3$: $\mathbf{N}(A) =$ all $\begin{bmatrix} 0 \\ 0 \\ \boldsymbol{z} \end{bmatrix}$

Every basis for that nullspace $\mathbf{N}(A)$ contains one vector like $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

The **dimension** of $\mathbf{N}(A)$ is $\mathbf{1}$. Notice $\mathbf{2} + \mathbf{1} =$ dimension of $\mathbf{R}^3$

Matrix spaces   The vector space of 3 by 3 matrices has dimension $\mathbf{9}$

The subspace of upper triangular matrices has dimension $\mathbf{6}$

If $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m$ and $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n$ are bases then $\boldsymbol{m}$ **must equal** $\boldsymbol{n}$

Proof $\boldsymbol{W} = \boldsymbol{V}\boldsymbol{A}$ $\begin{bmatrix} \boldsymbol{w}_1 & \cdots & \boldsymbol{w}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_m \end{bmatrix} \begin{bmatrix} a_{11} & \cdot & a_{1n} \\ \cdot & \cdot & \cdot \\ a_{m1} & \cdot & a_{mn} \end{bmatrix}$

If $\boldsymbol{m} < \boldsymbol{n}$ then $\boldsymbol{A}\boldsymbol{x} = \mathbf{0}$ would have a nonzero solution $\boldsymbol{x}$

Then $W\boldsymbol{x} = VA\boldsymbol{x} = \mathbf{0}$. $\boldsymbol{W}$ **would have dependent columns !**

3.3 Column Space and Row Space : Bases by Elimination

$$\begin{array}{l}\textbf{Every pivot} = \mathbf{1}\\ \textbf{Eliminate below}\\ \textbf{Eliminate above}\end{array}\ \begin{bmatrix}\mathbf{1}&4&9\\0&2&4\\0&3&7\end{bmatrix}\rightarrow\begin{bmatrix}\mathbf{1}&4&9\\0&\mathbf{1}&\mathbf{2}\\0&3&7\end{bmatrix}\rightarrow\begin{bmatrix}\mathbf{1}&4&9\\0&\mathbf{1}&\mathbf{2}\\0&\mathbf{0}&\mathbf{1}\end{bmatrix}\rightarrow\begin{bmatrix}\mathbf{1}&\mathbf{0}&\mathbf{1}\\0&\mathbf{1}&\mathbf{2}\\0&\mathbf{0}&\mathbf{1}\end{bmatrix}\rightarrow \boldsymbol{R_0} = \boldsymbol{I}$$

$$\boldsymbol{R_0} = \text{“reduced row echelon form”} = \textbf{rref}(A) = \begin{bmatrix}\boldsymbol{r}\ \textbf{rows start with 1}\\ \boldsymbol{m-r}\ \textbf{rows of zeros}\end{bmatrix}$$

$\boldsymbol{r} = \textbf{rank}$ $\quad$ $\boldsymbol{A}$ **has** $\boldsymbol{r}$ **independent columns and** $\boldsymbol{r}$ **independent rows**

$$\boldsymbol{A} = \begin{bmatrix}2&4\\3&7\\4&9\end{bmatrix}\rightarrow\begin{bmatrix}\mathbf{1}&2\\3&7\\4&9\end{bmatrix}\rightarrow\begin{bmatrix}\mathbf{1}&2\\\mathbf{0}&\mathbf{1}\\0&1\end{bmatrix}\rightarrow\begin{bmatrix}\mathbf{1}&2\\\mathbf{0}&\mathbf{1}\\\mathbf{0}&\mathbf{0}\end{bmatrix}\rightarrow\begin{bmatrix}\mathbf{1}&\mathbf{0}\\\mathbf{0}&\mathbf{1}\\\mathbf{0}&\mathbf{0}\end{bmatrix}=\begin{bmatrix}\boldsymbol{I}\\\mathbf{0}\end{bmatrix}=\boldsymbol{R_0}$$

$\boldsymbol{I}$ locates r independent columns
$\boldsymbol{P}$ permutes the columns if needed

$$\boldsymbol{R_0}=\begin{bmatrix}\boldsymbol{I}&\boldsymbol{F}\\0&0\end{bmatrix}\boldsymbol{P}=\begin{bmatrix}\boldsymbol{r}\text{ rows with }\boldsymbol{I}\\\boldsymbol{m-r}\text{ zero rows}\end{bmatrix}$$

Basis for the column space $\quad$ $\boldsymbol{C}$ = First $\boldsymbol{r}$ independent columns of $\boldsymbol{A}$

Basis for the row space $\quad$ $\boldsymbol{R} = \begin{bmatrix}\boldsymbol{I} & \boldsymbol{F}\end{bmatrix}\boldsymbol{P} = \boldsymbol{r}$ rows and $\boldsymbol{A} = \boldsymbol{CR}$

$$\boldsymbol{A} = \begin{bmatrix}1&1&2\\1&2&3\end{bmatrix}\rightarrow\begin{bmatrix}1&1&2\\0&1&1\end{bmatrix}\rightarrow\begin{bmatrix}\mathbf{1}&0&\mathbf{1}\\0&\mathbf{1}&\mathbf{1}\end{bmatrix}=\begin{bmatrix}\boldsymbol{I}&\boldsymbol{F}\end{bmatrix}=\boldsymbol{R_0}=\boldsymbol{R}$$

$$\boldsymbol{A}=\boldsymbol{CR}\text{ is }\begin{bmatrix}1&1&2\\1&2&3\end{bmatrix}=\begin{bmatrix}1&1\\1&2\end{bmatrix}\begin{bmatrix}\mathbf{1}&0&1\\0&\mathbf{1}&1\end{bmatrix}\qquad \begin{array}{l}\text{Row rank} = 2\\ \text{Column rank} = 2\end{array}$$

3.4 $Ax = 0$ and $Ax = b$: $x_{\text{nullspace}}$ and $x_{\text{particular}}$

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 1 & 4 \\ 2 & 4 & 3 & 9 \end{bmatrix} \to \begin{bmatrix} \mathbf{1} & 2 & 1 & 4 \\ \mathbf{0} & 0 & 1 & 1 \end{bmatrix} \to \begin{bmatrix} \mathbf{1} & 2 & \mathbf{0} & 3 \\ \mathbf{0} & 0 & \mathbf{1} & 1 \end{bmatrix} = \boldsymbol{R}$$

2 columns of I
$4-2$ columns of F

Basis for nullspace

Special solutions

$A\boldsymbol{s}_1 = \mathbf{0}$ and $A\boldsymbol{s}_2 = \mathbf{0}$

$$\boldsymbol{s}_1 = \begin{bmatrix} -2 \\ \mathbf{1} \\ 0 \\ \mathbf{0} \end{bmatrix} \qquad \boldsymbol{s}_2 = \begin{bmatrix} -3 \\ \mathbf{0} \\ -1 \\ \mathbf{1} \end{bmatrix}$$

$2, 3, 1$ in R
$\leftarrow \mathbf{1}$ and $\mathbf{0}$
$-2, -3, -1$ in S
$\leftarrow \mathbf{0}$ and $\mathbf{1}$

> "Special" Dependent columns 2 and 4 = combination of independent columns 1 and 3

> **Elimination** from A to R reveals the $n-r$ special solutions
>
> $\boldsymbol{A}\boldsymbol{x} = \mathbf{0}$ and $\boldsymbol{R}\boldsymbol{x} = \mathbf{0}$ $\qquad \boldsymbol{R}\boldsymbol{x} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{F} \end{bmatrix} \boldsymbol{P}\boldsymbol{x} = \mathbf{0}$ $\qquad \boldsymbol{S} = \begin{bmatrix} \boldsymbol{s}_1 & \cdots & \boldsymbol{s}_{n-r} \end{bmatrix} = \boldsymbol{P}^{\mathrm{T}} \begin{bmatrix} -\boldsymbol{F} \\ \boldsymbol{I}_{\boldsymbol{n-r}} \end{bmatrix}$
>
> $PP^{\mathrm{T}} = I$ leads to $\boldsymbol{RS} = \mathbf{0}$ $\qquad$ $\boldsymbol{r}$ **equations** $\qquad$ $\boldsymbol{n-r}$ **solutions**

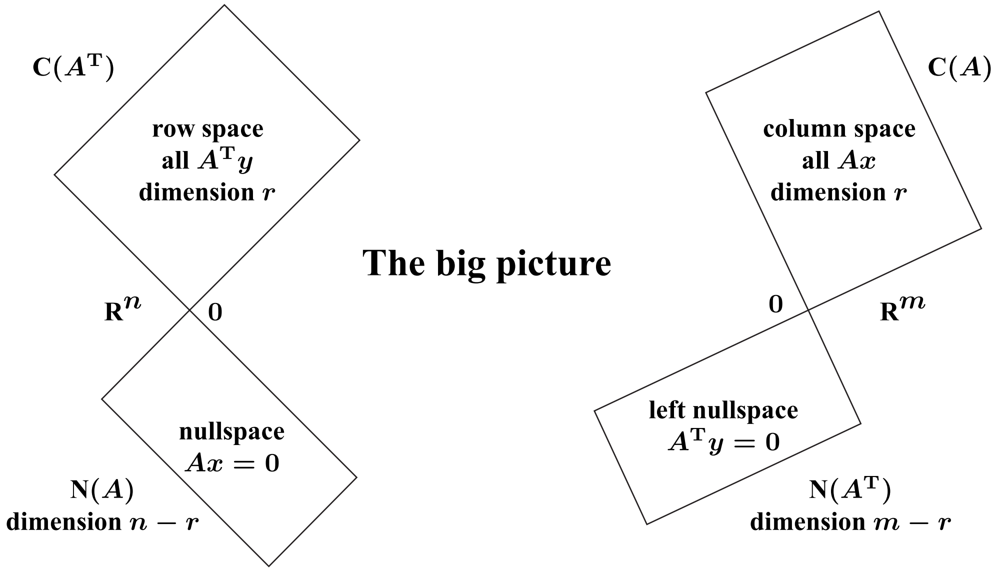
Complete solution to $\boldsymbol{Ax} = \boldsymbol{b}$ $\quad \boldsymbol{x} = \boldsymbol{x}_{\text{nullspace}} + \boldsymbol{x}_{\text{particular}} =$ above $\uparrow +$ below $\downarrow$

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{b} \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 & 4 & b_1 \\ 2 & 4 & 3 & 9 & b_2 \end{bmatrix} \to \begin{bmatrix} \mathbf{1} & 2 & \mathbf{0} & 3 & d_1 \\ \mathbf{0} & 0 & \mathbf{1} & 1 & d_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{R} & \boldsymbol{d} \end{bmatrix} \qquad \boldsymbol{x}_{\text{particular}} = \begin{bmatrix} \boldsymbol{d_1} \\ 0 \\ \boldsymbol{d_2} \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} \boldsymbol{A}_{\text{new}} & \boldsymbol{b}_{\text{new}} \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 & 4 & b_1 \\ 2 & 4 & 3 & 9 & b_2 \\ 3 & 6 & 4 & 13 & b_3 \end{bmatrix} \to \begin{bmatrix} \mathbf{1} & 2 & 0 & 3 & \boldsymbol{d_1} \\ 0 & 0 & \mathbf{1} & 1 & \boldsymbol{d_2} \\ 0 & 0 & 0 & 0 & \boldsymbol{d_3} \end{bmatrix}$$

No solution if $d_3 \neq 0$
No solution if $b_1 + b_2 \neq b_3$
Elimination must give $\mathbf{0} = \mathbf{0}$

3.5 Four Fundamental Subspaces $\mathbf{C}(A), \mathbf{C}(A^{\mathrm{T}}), \mathbf{N}(A), \mathbf{N}(A^{\mathrm{T}})$



Fundamental Theorem of Linear Algebra, **Part 1**
The column space and row space both have dimension r.
The nullspaces have dimensions $n - r$ and $m - r$.

This tells us the **Counting Theorem** : How many solutions to $Ax = \mathbf{0}$?

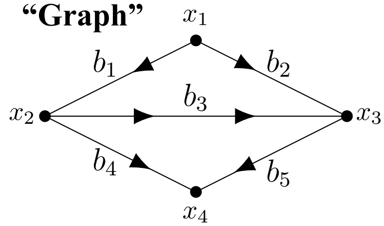
m equations, n unknowns, rank $r \Rightarrow A\boldsymbol{x} = \mathbf{0}$ has $n - r$ independent solutions

At least $n - m$ solutions. More for dependent equations (then $r < m$)

There is always a nonzero solution x to $Ax = 0$ if $n > m$

3.6 Graphs, Incidence Matrices, and Kirchhoff's Laws

"Graph"



$$\begin{array}{c} \textbf{nodes} \\ \\ \boldsymbol{A} = \end{array} \begin{array}{c} \begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \end{array} \\ \begin{bmatrix} \mathbf{-1} & \mathbf{1} & 0 & 0 \\ \mathbf{-1} & 0 & \mathbf{1} & 0 \\ 0 & \mathbf{-1} & \mathbf{1} & 0 \\ 0 & \mathbf{-1} & 0 & \mathbf{1} \\ 0 & 0 & \mathbf{-1} & \mathbf{1} \end{bmatrix} \end{array} \begin{array}{c} \textbf{edges} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array}$$

This graph has 5 edges and 4 nodes. $\boldsymbol{A}$ **is its 5 by 4 incidence matrix**.

$\boldsymbol{b} = b_1$ to b_5 = currents. $\boldsymbol{x} = x_1$ to x_4 = voltages

Edges $\mathbf{1}, \mathbf{2}, \mathbf{3}$ form a **loop** in the graph	**Dependent rows** $1, 2, 3$
Edges $\mathbf{1}, \mathbf{2}, 4$ form a **tree**. **Trees have no loops!**	**Independent rows** $1, 2, 4$

The incidence matrix A comes from a connected graph with n nodes and m edges. The row space and column space have dimensions $r = n - 1$. The nullspaces of A and A^{T} have dimensions 1 and $m - n + 1$:

$\mathbf{N}(A)$ The constant vectors $(c, c, \ldots, c)$ make up the nullspace of A : $\dim = 1$.

$\mathbf{C}(A^{\mathrm{T}})$ The edges of any spanning tree give r independent rows of A : $r = n - 1$.

$\mathbf{C}(A)$ ***Voltage Law***: The components of $A\boldsymbol{x}$ add to zero around all loops: $\dim = n - 1$.

$\mathbf{N}(A^{\mathrm{T}})$ ***Current Law***: $A^{\mathrm{T}}\boldsymbol{y} =$ (**flow in**)$-$(**flow out**) $= \mathbf{0}$ is solved by loop currents.
There are $m - r = m - n + 1$ independent small loops in the graph.

Currrent law $A^{\mathrm{T}}\boldsymbol{y} = \mathbf{0}$ at each node is fundamental to applied mathematics

3.7 Every A Has a Pseudoinverse A^+ $\quad \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}^+ = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{bmatrix}$

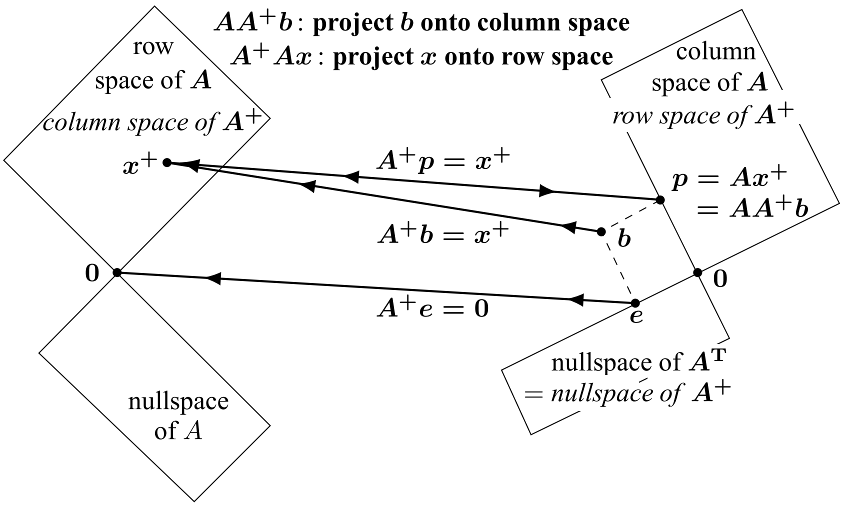
A is invertible if and only if $m = n = r$ (rank). Then $A^+ = A^{-1}$

A has a **left inverse** $A^+ = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ when $r = n$: $A^+A = I_n$

A has a **right inverse** $A^+ = A^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}$ when $r = m$: $AA^+ = I_m$

$A = CR$ has a **pseudoinverse** $A^+ = R^+C^+$. Reverse the 4 subspaces

A Row space to column space
A^+ Column space to row space



$\boxed{\textbf{Pseudoinverse } A^+}$ $\quad A^+A = \begin{bmatrix} I & 0 \\ 0 & \mathbf{0} \end{bmatrix} \begin{matrix} \text{row space} \\ \text{nullspace} \end{matrix} \qquad AA^+ = \begin{bmatrix} I & 0 \\ 0 & \mathbf{0} \end{bmatrix} \begin{matrix} \mathbf{C}(A) \\ \mathbf{N}(A^{\mathrm{T}}) \end{matrix}$

$A^+ = V\Sigma^+U^{\mathrm{T}}$ is computed from $A = U\Sigma V^{\mathrm{T}}$: Σ^+ has $1/\sigma$'s and 0's

Example $\quad \boldsymbol{A} = \begin{bmatrix} \mathbf{1} & \mathbf{1} & \mathbf{1} \\ \mathbf{1} & \mathbf{1} & \mathbf{1} \end{bmatrix} = \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} = \boldsymbol{CR}$

The pseudoinverse A^+ has the same rank $r = 1$

Row space of $A =$ line in $\mathbf{R}^3$ $\qquad$ Column space of $A =$ line in $\mathbf{R}^2$

Reverse the column space and row space : $\boldsymbol{v}$ and $\boldsymbol{u}$

$$\boldsymbol{A}^+ = \frac{\boldsymbol{v}\boldsymbol{u}^{\mathrm{T}}}{||\boldsymbol{v}||^2\,||\boldsymbol{u}||^2} = \frac{1}{6}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \end{bmatrix} = \frac{\mathbf{1}}{\mathbf{6}}\begin{bmatrix} \mathbf{1} & \mathbf{1} \\ \mathbf{1} & \mathbf{1} \\ \mathbf{1} & \mathbf{1} \end{bmatrix}$$

Check $\quad A^+A = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} =$ **projection onto row space of $\boldsymbol{A}$** $\qquad AA^+ = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} =$ **projection onto column space of $\boldsymbol{A}$**

Part 4

Orthogonal Matrices and Least Squares

Part 4 : Orthogonal Matrices and Least Squares

4.1 Orthogonality of the Four Subspaces

Vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are orthogonal if $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = 0$ (Complex vectors : $\overline{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{y} = 0$)

Then $||\boldsymbol{x}+\boldsymbol{y}||^2 = ||\boldsymbol{x}||^2 + ||\boldsymbol{y}||^2 = ||\boldsymbol{x}-\boldsymbol{y}||^2$ (Right triangles : Pythagoras)

Orthogonal subspaces : Every $\boldsymbol{v}$ in V is orthogonal to every $\boldsymbol{w}$ in W

Two walls of a room are not orthogonal. Meeting line is in both subspaces !

The row space and the nullspace of any matrix are orthogonal
The column space $\mathbf{C}(A)$ and nullspace $\mathbf{N}(A^{\mathrm{T}})$: **Also orthogonal**

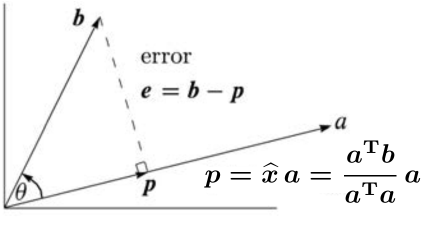
Clear from $A\boldsymbol{x} = \mathbf{0}$ $\qquad A\boldsymbol{x} = \begin{bmatrix} \textbf{row 1} \\ \text{---} \\ \textbf{row } \boldsymbol{m} \end{bmatrix} \begin{bmatrix} \\ \boldsymbol{x} \\ \\ \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \cdot \\ \mathbf{0} \end{bmatrix}$

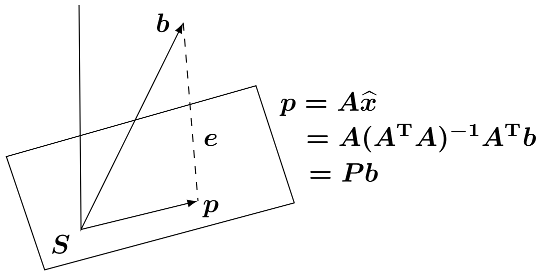
All rows are orthogonal to $\boldsymbol{x}$ $\Rightarrow$ whole row space is orthogonal to $\boldsymbol{x}$

Big Picture of Linear Algebra : Two pairs of orthogonal subspaces

More : $r + (n - r) =$ full dimension n so every $\boldsymbol{x}$ equals $\boldsymbol{x}_{\text{row}} + \boldsymbol{x}_{\text{null}}$

4.2 Projections onto Subspaces





$\boldsymbol{b}$ is projected onto line through $\boldsymbol{a}$ and onto column space of A

Error vector $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p}$ is **orthogonal** to the line and subspace

Projection matrix $\boldsymbol{P}$ $\quad \boxed{P_{\text{line}} = \dfrac{\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}} \qquad P_{\text{subspace}} = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}}$

Notice $\boldsymbol{P^2 = P}$ and $\boldsymbol{P^{\mathrm{T}} = P}$ $\qquad$ Second projection : $\boldsymbol{p}$ doesn't move !

Project $\boldsymbol{b}$ $= \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$ Line : Plane : Column spaces of $\boldsymbol{a} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$

$$\boldsymbol{p}_{\textbf{line}} = \boldsymbol{a}\,\frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \frac{6}{3} = \begin{bmatrix} \mathbf{2} \\ \mathbf{2} \\ \mathbf{2} \end{bmatrix} \qquad \text{Error } \boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p} = \begin{bmatrix} 4 \\ -2 \\ -2 \end{bmatrix}$$

$$\boldsymbol{p}_{\textbf{plane}} \quad \textbf{Solve } \boldsymbol{A^{\mathrm{T}}\widehat{x} = A^{\mathrm{T}}b} \qquad \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \widehat{\boldsymbol{x}} = \begin{bmatrix} 6 \\ 0 \end{bmatrix} \quad \text{gives } \widehat{\boldsymbol{x}} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}$$

$$\text{Projection } \boldsymbol{p} = A\widehat{\boldsymbol{x}} = \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix} \text{ Error } \boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} \mathbf{1} \\ \mathbf{-2} \\ \mathbf{1} \end{bmatrix}$$

What is the 3 by 3 projection matrix $\boldsymbol{P = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}}$? Then $\boldsymbol{p = Pb}$

4.3 Least Squares Approximations (Regression) : $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$

If $A\boldsymbol{x} = \boldsymbol{b}$ has no solution then minimize $E = ||\boldsymbol{b} - A\boldsymbol{x}||^2 = \boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}A\boldsymbol{x} - 2\boldsymbol{x}^{\mathrm{T}}A^{\mathrm{T}}\boldsymbol{b} + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{b}$

Calculus Partial derivatives $\partial E/\partial x$ of that error E are zero

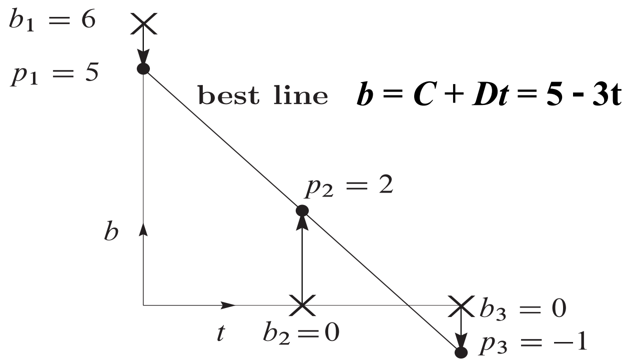
Linear algebra $A\boldsymbol{x}$ is in the column space of A Best $A\widehat{\boldsymbol{x}}$=projection of $\boldsymbol{b}$ on $\mathbf{C}(A)$

"Normal equations" $A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$ and then the projection is $\boldsymbol{p} = A\widehat{\boldsymbol{x}}$

Key example Closest straight line $y = C + Dt$ to m points (t_i, b_i)

$\boldsymbol{m > 2}$ **$\boldsymbol{m}$ equations** **2 unknowns** **No solution**

$$C + Dt_1 = b_1 \quad \cdots\cdots\cdots \quad C + Dt_m = b_m \qquad A = \begin{bmatrix} 1 & t_1 \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & t_m \end{bmatrix} \quad \boldsymbol{x} = \begin{bmatrix} C \\ D \end{bmatrix} \quad \boldsymbol{b} = \begin{bmatrix} b_1 \\ \cdot \\ \cdot \\ b_m \end{bmatrix}$$



errors $\boldsymbol{e}$ = vertical distances $\boldsymbol{b}$ - $\boldsymbol{p}$ = $(1, -2, 1)$ to best line

$$A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b} \qquad \begin{bmatrix} m & \sum t_i \\ \sum t_i & \sum t_i^2 \end{bmatrix}\begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix}\begin{bmatrix} \mathbf{5} \\ \mathbf{-3} \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix} = \begin{bmatrix} \sum b_i \\ \sum t_i b_i \end{bmatrix}$$

4.4 Orthogonal Matrices and Gram-Schmidt

Orthogonal columns $\quad \boldsymbol{Q} = \begin{bmatrix} \boldsymbol{q}_1 & \cdots & \boldsymbol{q}_n \end{bmatrix}$ $\quad \boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{q}_i = 1$ unit vectors $\quad \boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{q}_j = 0$ orthogonal $\quad \boxed{\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{I}}$

Important case = Square matrix Then $\boldsymbol{QQ}^{\mathrm{T}} = \boldsymbol{I}$ $\quad \boldsymbol{Q}^{\mathrm{T}} = \boldsymbol{Q}^{-1}$ **"Orthogonal matrix"**

$$Q = \frac{1}{3}\begin{bmatrix} 2 & -1 \\ 2 & 2 \\ -1 & 2 \end{bmatrix} \text{ has } \begin{matrix} \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{I} \\ \boldsymbol{QQ}^{\mathrm{T}} \neq \boldsymbol{I} \end{matrix} \qquad Q = \frac{1}{3}\begin{bmatrix} 2 & -1 & 2 \\ 2 & 2 & -1 \\ -1 & 2 & 2 \end{bmatrix} \begin{matrix} \text{Now } Q^{\mathrm{T}} = Q^{-1} \\ \text{Orthogonal matrix} \end{matrix}$$

Q_1 times Q_2 is orthogonal because $(Q_1Q_2)^{-1} = Q_2^{-1}Q_1^{-1} = Q_2^{\mathrm{T}}Q_1^{\mathrm{T}} = (Q_1Q_2)^{\mathrm{T}}$

$\boldsymbol{v} = c_1\boldsymbol{q}_1 + \cdots + c_n\boldsymbol{q}_n$ leads to $c_k = \boldsymbol{q}_k^{\mathrm{T}}\boldsymbol{v}$ $\qquad \boldsymbol{v} = Q\boldsymbol{c}$ leads to $\boldsymbol{c} = Q^{\mathrm{T}}\boldsymbol{v}$

Gram-Schmidt Start with independent $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ Create orthogonal vectors $\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3$

$$\boldsymbol{q}_1 = \frac{\boldsymbol{a}}{||\boldsymbol{a}||} \quad \boldsymbol{Q}_2 = \boldsymbol{b} - (\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b})\boldsymbol{q}_1 \quad \boldsymbol{q}_2 = \frac{\boldsymbol{Q}_2}{||\boldsymbol{Q}_2||} \quad \boldsymbol{Q}_3 = \boldsymbol{c} - (\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{c})\boldsymbol{q}_1 - (\boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{c})\boldsymbol{q}_2 \quad \boldsymbol{q}_3 = \frac{\boldsymbol{Q}_3}{||\boldsymbol{Q}_3||}$$

Gram-Schmidt $\boldsymbol{A} = \boldsymbol{QR}$

$\boldsymbol{A} =$ **(orthogonal)(triangular)**

$$\boldsymbol{A} = \begin{bmatrix} \cos\theta & a_{12} \\ \sin\theta & a_{22} \end{bmatrix} = \begin{bmatrix} \cos\boldsymbol{\theta} & -\sin\boldsymbol{\theta} \\ \sin\boldsymbol{\theta} & \cos\boldsymbol{\theta} \end{bmatrix} \begin{bmatrix} \mathbf{1} & \boldsymbol{r}_{12} \\ \mathbf{0} & \boldsymbol{r}_{22} \end{bmatrix} = \boldsymbol{QR}$$

Part 5

Determinant of a Square Matrix

Part 5 : Determinant of a Square Matrix

5.1 3 by 3 and n by n Determinants

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

$$\mathbf{det} = \boldsymbol{+1} \qquad \boldsymbol{-1} \qquad \boldsymbol{+1} \qquad \boldsymbol{-1} \qquad \boldsymbol{+1} \qquad \boldsymbol{-1}$$

Even permutations have $\det P = \boldsymbol{+1}$ **Odd permutations** have $\det P = \boldsymbol{-1}$

Three defining properties **1** Row exchange reverses sign of det

2 det is linear in each row separately **3** $\det I = 1$

$$\begin{bmatrix} \boldsymbol{a} & & \\ & \boldsymbol{q} & \\ & & \boldsymbol{z} \end{bmatrix} \begin{bmatrix} & \boldsymbol{b} & \\ \boldsymbol{p} & & \\ & & \boldsymbol{z} \end{bmatrix} \begin{bmatrix} & \boldsymbol{b} & \\ & & \boldsymbol{r} \\ \boldsymbol{x} & & \end{bmatrix} \begin{bmatrix} & & \boldsymbol{c} \\ & \boldsymbol{q} & \\ \boldsymbol{x} & & \end{bmatrix} \begin{bmatrix} & & \boldsymbol{c} \\ \boldsymbol{p} & & \\ & \boldsymbol{y} & \end{bmatrix} \begin{bmatrix} \boldsymbol{a} & & \\ & & \boldsymbol{r} \\ & \boldsymbol{y} & \end{bmatrix}$$

$$\mathbf{det} = \boldsymbol{+aqz} \qquad \boldsymbol{-bpz} \qquad \boldsymbol{+brx} \qquad \boldsymbol{-cqx} \qquad \boldsymbol{+cpy} \qquad \boldsymbol{-ary}$$

Linearity separates $\det \boldsymbol{A}$ **into** $\boldsymbol{n! = 3! = 6}$ **simple determinants**

$$\mathbf{det} = \begin{bmatrix} \boldsymbol{a} & \boldsymbol{b} & \boldsymbol{c} \\ \boldsymbol{p} & \boldsymbol{q} & \boldsymbol{r} \\ \boldsymbol{x} & \boldsymbol{y} & \boldsymbol{z} \end{bmatrix} \begin{matrix} a & b \\ p & q \\ x & y \end{matrix}$$

$$- \quad - \quad - \qquad + \quad + \quad +$$

Combine 6 simple determinants into $\det \boldsymbol{A}$

$$+ \boldsymbol{aqz} + \boldsymbol{brx} + \boldsymbol{cpy} - \boldsymbol{ary} - \boldsymbol{bpz} - \boldsymbol{cqx}$$

Each term takes 1 number from each row and each column

BIG FORMULA = Sum over all $n!$ orders $P = (j, k, \ldots, z)$ of the columns

$$\boxed{\mathbf{det}\,\boldsymbol{A} = \sum (\mathbf{det}\,\boldsymbol{P})\, \boldsymbol{a_{1j}\, a_{2k} \ldots a_{nz}} \ \text{ as in } \ \boldsymbol{+a_{11}a_{22} - a_{12}a_{21}}}$$

5.2 Cofactors and the Formula for A^{-1}

3 by 3 determinant: 2 terms start with a and with b and with c

Cofactor formula $\det A = a\,(\boldsymbol{qz - ry}) + b\,(\boldsymbol{rx - pz}) + c\,(\boldsymbol{py - qx})$

$\boldsymbol{n}$ **factors** a, b, c $\boldsymbol{n}$ **cofactors** = determinants of size $n-1$

Remove row i and column j from A $\boxed{\textbf{Cofactor } \boldsymbol{C_{ij}} = \textbf{det} \text{ times } \boldsymbol{(-1)^{i+j}}}$

Cofactors along row 1 $\boxed{\textbf{det}\,\boldsymbol{A = a_{11}C_{11} + \cdots + a_{1n}C_{1n}}}$

Inverse formula $\boxed{\boldsymbol{A^{-1}} = (\textbf{transpose of } \boldsymbol{C})/(\textbf{determinant of } \boldsymbol{A})}$

$$\text{Every entry of } A^{-1} = \frac{\text{cofactor}}{\det A} = \frac{\det \text{ of size } n-1}{\det \text{ of size } n}$$

$$n = 2 \qquad \boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad \text{Cofactors } \boldsymbol{C} = \begin{bmatrix} d & -c \\ -b & a \end{bmatrix} \qquad \boldsymbol{A^{-1}} = \frac{\boldsymbol{C}^{\mathrm{T}}}{\boldsymbol{ad - bc}}$$

$$n = 3 \qquad \boldsymbol{A} = \begin{bmatrix} a & b & c \\ p & q & r \\ x & y & z \end{bmatrix} \qquad \boldsymbol{C} = \begin{bmatrix} qz - ry & rx - pz & py - qx \\ bz - cy & az - cx & qx - py \\ br - cq & cp - ar & aq - bp \end{bmatrix}$$

$$\boldsymbol{AC}^{\mathrm{T}} = \begin{bmatrix} \det A & 0 & 0 \\ 0 & \det A & 0 \\ 0 & 0 & \det A \end{bmatrix} = (\textbf{det}\,\boldsymbol{A})\boldsymbol{I} \quad \text{This explains } \boldsymbol{A^{-1}} = \frac{\boldsymbol{C}^{\mathrm{T}}}{\textbf{det}\,\boldsymbol{A}}$$

5.3 Det AB = (Det A) (Det B) and Cramer's Rule

$$\boxed{\det \boldsymbol{A} = \det \boldsymbol{A}^{\mathrm{T}}} \quad \boxed{\det \boldsymbol{AB} = (\det \boldsymbol{A})\,(\det \boldsymbol{B})} \quad \boxed{\det \boldsymbol{A}^{-1} = \frac{\mathbf{1}}{\det \boldsymbol{A}}}$$

Orthogonal matrix $\det \boldsymbol{Q} = \pm \mathbf{1}$ because $Q^{\mathrm{T}}Q = I$ gives $(\det Q)^2 = 1$

Triangular matrix $\det \boldsymbol{U} = \boldsymbol{u_{11}u_{22}\cdots u_{nn}}$

$\det \boldsymbol{A} = \det \boldsymbol{LU} = (\det \boldsymbol{L})\,(\det \boldsymbol{U}) =$ **product of the pivots** $\boldsymbol{u_{ii}}$

Cramer's Rule to Solve $\boldsymbol{Ax = b}$ Start from

$$\begin{bmatrix} & & \\ & \boldsymbol{A} & \\ & & \end{bmatrix} \begin{bmatrix} \boldsymbol{x_1} & 0 & 0 \\ \boldsymbol{x_2} & 1 & 0 \\ \boldsymbol{x_3} & 0 & 1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{b_1} & a_{12} & a_{13} \\ \boldsymbol{b_2} & a_{22} & a_{23} \\ \boldsymbol{b_3} & a_{32} & a_{33} \end{bmatrix} = \boldsymbol{B_1}$$

Use $(\det \boldsymbol{A})\,(\boldsymbol{x}_1) = (\det \boldsymbol{B_1})$ to find $\boldsymbol{x}_1$

Same idea

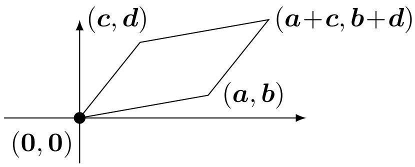
$$\begin{bmatrix} & & \\ & \boldsymbol{A} & \\ & & \end{bmatrix} \begin{bmatrix} 1 & \boldsymbol{x_1} & 0 \\ 0 & \boldsymbol{x_2} & 0 \\ 0 & \boldsymbol{x_3} & 1 \end{bmatrix} = \begin{bmatrix} & & \\ \boldsymbol{a_1} & \boldsymbol{b} & \boldsymbol{a_3} \\ & & \end{bmatrix} = \boldsymbol{B_2}$$

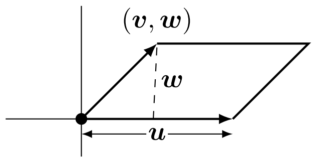
$$\boxed{\begin{aligned} \boldsymbol{x}_1 &= \frac{\det \boldsymbol{B_1}}{\det \boldsymbol{A}} \\ \boldsymbol{x}_2 &= \frac{\det \boldsymbol{B_2}}{\det \boldsymbol{A}} \end{aligned}}$$

Cramer's Rule is usually not efficient ! Too many determinants

$$\begin{bmatrix} 3 & 2 \\ 5 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 12 \\ 22 \end{bmatrix} \quad \boldsymbol{B}_1 = \begin{bmatrix} 12 & 2 \\ 22 & 4 \end{bmatrix} \quad \boldsymbol{B}_2 = \begin{bmatrix} 3 & 12 \\ 5 & 22 \end{bmatrix} \quad x_1 = \frac{\det B_1}{\det A} = \frac{4}{2} \quad x_2 = \frac{2}{2}$$

5.4 Volume of Box = | Determinant of Edge Matrix E |





Rotate to align
Area $= uw$

Edge matrix $\quad E = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = QR =$ (orthogonal Q) (triangular R) $\quad R = \begin{bmatrix} u & v \\ 0 & w \end{bmatrix}$

To Prove: Area of a parallelogram is $|\det E| = |ad - bc| = |\det R| = uw$

2 D area Gram-Schmidt in 4.4 gives $E = QR =$ **(orthogonal) (triangular)**

Orthogonal Q: Rotates the shape = No change in area!

Triangular R: u = base, w = height, uw = **area** $= |\det R| = |\det E|$

3D volume Edges of box = Rows of E Volume of box $= |\det E| = |\det R|$

Orthogonal Q: No volume change **Rotate box to see volume** $= r_{11}r_{22}r_{33}$

If the box is a unit cube: E = identity matrix and volume = 1

Any shape Multiply all points by A Volume multiplies by $\det A$

Part 6

Eigenvalues and Eigenvectors : $Ax = \lambda x$ and $A^n x = \lambda^n x$

Part 6 : Eigenvalues and Eigenvectors : $\boldsymbol{Ax = \lambda x}$ and $\boldsymbol{A^n x = \lambda^n x}$

6.1 Eigenvalues $\boldsymbol{\lambda}$ and Eigenvectors $\boldsymbol{x}$: $\boldsymbol{Ax = \lambda x}$

$A\boldsymbol{x}$ is on the same line as $\boldsymbol{x}$ / $A\boldsymbol{x} = \lambda\boldsymbol{x}$ means $(\boldsymbol{A - \lambda I})\boldsymbol{x} = \boldsymbol{0}$

Then $\boldsymbol{A^2 x} = \boldsymbol{A}(\lambda\boldsymbol{x}) = \lambda(\boldsymbol{Ax}) = \boldsymbol{\lambda}^2\boldsymbol{x}$ $\quad \boxed{A^n\boldsymbol{x} = \lambda^n\boldsymbol{x}} \quad \boxed{A^{-1}\boldsymbol{x} = \dfrac{1}{\lambda}\,\boldsymbol{x}}$

Determinant of $\boldsymbol{A - \lambda I} = \boldsymbol{0}$ Solutions λ_1 to λ_n : A has n eigenvalues

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{.8} & \boldsymbol{.3} \\ \boldsymbol{.2} & \boldsymbol{.7} \end{bmatrix} \qquad \boldsymbol{A - \lambda I} = \begin{bmatrix} .8-\lambda & .3 \\ .2 & .7-\lambda \end{bmatrix}$$

$$\textbf{det}(\boldsymbol{A - \lambda I}) = \boldsymbol{\lambda^2 - 1.5\lambda + .56 - .06} = (\lambda - \mathbf{1})\left(\lambda - \tfrac{\mathbf{1}}{\mathbf{2}}\right)$$

Eigenvector $\boldsymbol{x_2}$ for $\boldsymbol{\lambda_2 = \frac{1}{2}}$ $\quad (A - \tfrac{1}{2}I)\,\boldsymbol{x}_2 = \begin{bmatrix} .3 & .3 \\ .2 & .2 \end{bmatrix}\begin{bmatrix} \boldsymbol{x}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ gives $\boldsymbol{x}_2 = \begin{bmatrix} \mathbf{1} \\ \mathbf{-1} \end{bmatrix}$

Eigenvector $\boldsymbol{x_1}$ for $\boldsymbol{\lambda_1 = 1}$ $\quad (A - I)\boldsymbol{x}_1 = \begin{bmatrix} -.2 & .3 \\ .2 & -.3 \end{bmatrix}\begin{bmatrix} \boldsymbol{x}_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ gives $\boldsymbol{x}_1 = \begin{bmatrix} \mathbf{0.6} \\ \mathbf{0.4} \end{bmatrix}$

What is $A^{10}\begin{bmatrix} 1 \\ 0 \end{bmatrix}$? Separate into eigenvectors / Follow each eigenvector

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.6 \\ 0.4 \end{bmatrix} + \begin{bmatrix} 0.4 \\ -0.4 \end{bmatrix} \qquad A^{10}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \mathbf{1^{10}}\begin{bmatrix} \mathbf{0.6} \\ \mathbf{0.4} \end{bmatrix} + \left(\frac{\mathbf{1}}{\mathbf{2}}\right)^{\mathbf{10}}\begin{bmatrix} \mathbf{0.4} \\ \mathbf{-0.4} \end{bmatrix}$$

Useful	**Sum of $\boldsymbol{\lambda}$'s $= \boldsymbol{\lambda_1 + \cdots + \lambda_n} =$ trace of $\boldsymbol{A = a_{11} + a_{22} + \cdots + a_{nn}}$**
facts	**Product of $\boldsymbol{\lambda}$'s $= (\boldsymbol{\lambda_1})\cdots(\boldsymbol{\lambda_n}) =$ determinant of $\boldsymbol{A}$**

Eigenvalues of $A + B$ and AB are usually not $\lambda(A) + \lambda(B)$ and $\lambda(A)\lambda(B)$

6.2 Diagonalizing a Matrix : $X^{-1}AX = \Lambda$ = eigenvalues

Key idea / **Follow each eigenvector separately / n simple problems**

Eigenvector matrix X
Assume independent $\boldsymbol{x}$'s
Then X is invertible

$$AX = A\begin{bmatrix} \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n \end{bmatrix} = \begin{bmatrix} \lambda_1\boldsymbol{x}_1 & \cdots & \lambda_n\boldsymbol{x}_n \end{bmatrix}$$

$$\boxed{\begin{aligned} \boldsymbol{AX} &= \boldsymbol{X\Lambda} \\ \boldsymbol{X^{-1}AX} &= \boldsymbol{\Lambda} \\ \boldsymbol{A} &= \boldsymbol{X\Lambda X^{-1}} \end{aligned}} \qquad \begin{bmatrix} \lambda_1\boldsymbol{x}_1 & \cdots & \lambda_n\boldsymbol{x}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n \end{bmatrix}\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

1 A^k becomes easy $\qquad A^k = (X\Lambda X^{-1})\,(X\Lambda X^{-1})\cdots(X\Lambda X^{-1})$

Same eigenvectors in X $\qquad \boxed{\boldsymbol{A^k = X\Lambda^k X^{-1}}} \qquad \boldsymbol{\Lambda^k}$ = (eigenvalues)$^{\boldsymbol{k}}$

$$\begin{bmatrix} \mathbf{1} & \mathbf{2} \\ \mathbf{0} & \mathbf{3} \end{bmatrix}^{\mathbf{4}} = X\Lambda^4X^{-1} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \mathbf{1^4} & 0 \\ 0 & \mathbf{3^4} \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 81 \\ 0 & 81 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{1} & \mathbf{80} \\ \mathbf{0} & \mathbf{81} \end{bmatrix}$$

2 Question : **When does $\boldsymbol{A^k} \to$ zero matrix ?**

Answer : $\boxed{\textbf{All } |\boldsymbol{\lambda_i}| < \mathbf{1}}$

3 Some matrices are not diagonalizable

They don't have n independent vectors $\qquad A = \begin{bmatrix} 3 & 6 \\ 0 & 3 \end{bmatrix}$ has $\lambda = \mathbf{3}$ and $\mathbf{3}$

That A has double eigenvalue, single eigenvector $\qquad$ Only one $\boldsymbol{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

4

All the "similar matrices" $\boldsymbol{BAB^{-1}}$ have the **same eigenvalues as $\boldsymbol{A}$**

$$\text{If } A\boldsymbol{x} = \lambda\boldsymbol{x} \text{ then } (BAB^{-1})\,(B\boldsymbol{x}) = BA\boldsymbol{x} = B\lambda\boldsymbol{x} = \lambda(B\boldsymbol{x})$$

6.3 Symmetric Positive Definite Matrices : Five Tests

If $\boldsymbol{S}=\boldsymbol{S}^{\mathrm{T}}$ $\boxed{\textbf{Eigenvalues } \boldsymbol{\lambda} \textbf{ are real}}$ $\boxed{\textbf{Eigenvectors } \boldsymbol{x} \textbf{ are orthogonal}}$

$$\boldsymbol{S}=\begin{bmatrix}\mathbf{5} & \mathbf{4}\\ \mathbf{4} & \mathbf{5}\end{bmatrix}=\boldsymbol{S}^{\mathrm{T}} \text{ has } S\begin{bmatrix}1\\1\end{bmatrix}=\mathbf{9}\begin{bmatrix}1\\1\end{bmatrix} \text{ and } S\begin{bmatrix}-1\\1\end{bmatrix}=\begin{bmatrix}-1\\1\end{bmatrix} \quad \begin{matrix}\text{Notice}\\ \boldsymbol{x}_1\cdot\boldsymbol{x}_2=\mathbf{0}\end{matrix} \begin{bmatrix}1\\1\end{bmatrix}\cdot\begin{bmatrix}-1\\1\end{bmatrix}=\mathbf{0}$$

$\boldsymbol{q}=\dfrac{\boldsymbol{x}}{||\boldsymbol{x}||}=$ eigenvectors length $=1$

Eigenvector matrix Q is an orthogonal matrix : $Q^{\mathrm{T}}=Q^{-1}$

$$\boxed{\begin{matrix}\boldsymbol{S}=\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{-1}=\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{\mathrm{T}}\\ \text{Spectral theorem}\end{matrix}}$$

$$\boldsymbol{S}=\begin{bmatrix}\mathbf{5} & \mathbf{4}\\ \mathbf{4} & \mathbf{5}\end{bmatrix}=\frac{1}{\sqrt{2}}\begin{bmatrix}1 & -1\\ 1 & 1\end{bmatrix}\begin{bmatrix}\mathbf{9} & 0\\ 0 & \mathbf{1}\end{bmatrix}\begin{bmatrix}1 & 1\\ -1 & 1\end{bmatrix}\frac{1}{\sqrt{2}}=\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{\mathrm{T}}$$

Positive definite matrices are the best. How to test S for $\lambda_i>0$?

Test **1** Compute the **eigenvalues** of S : All eigenvalues positive

Test **2** The **energy** $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{x}$ is positive for every vector $\boldsymbol{x}\neq\mathbf{0}$

Test **3** The **pivots** in elimination on S are all positive

Test **4** The upper left **determinants** of S are all positive

Test **5** $\boldsymbol{S}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ for some matrix A with independent columns

Positive **semidefinite** matrices can be singular : Test 5 is $\boldsymbol{S}=$ **any** $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$

Eigenvalues and energy and pivots and determinants of S can be zero

$\begin{bmatrix}\mathbf{2} & \boldsymbol{b}\\ \boldsymbol{b} & 4\end{bmatrix}$ Positive definite if $\boldsymbol{b}^{\mathbf{2}}<\mathbf{8}$ Semidefinite if $\boldsymbol{b}^{\mathbf{2}}\leq\mathbf{8}$

Second difference matrix
Positive definite in 6.5

$$\boldsymbol{S}=\begin{bmatrix}2 & -1 & 0\\ -1 & 2 & -1\\ 0 & -1 & 2\end{bmatrix}$$

Positive semidefinite
$S\boldsymbol{x}=\mathbf{0}$ **for** $\boldsymbol{x}=(1,1,1)$

$$\boldsymbol{S}=\begin{bmatrix}\mathbf{1} & -1 & 0\\ -1 & 2 & -1\\ 0 & -1 & \mathbf{1}\end{bmatrix}$$

6.4 Linear Differential Equations $\dfrac{d\boldsymbol{u}}{dt} = A\boldsymbol{u}$

$n = 1$ $\quad \dfrac{d\boldsymbol{u}}{dt} = \boldsymbol{au}$ is solved by $\boldsymbol{u(t) = Ce^{at} = u(0)e^{at}}$

$n \geq 1$ $\quad \dfrac{d\boldsymbol{u}}{dt} = \boldsymbol{Au}$ is solved by eigenvectors as in $\boldsymbol{u(t) = c_1 e^{\lambda_1 t} x_1}$

The key is constant matrix $A \Leftrightarrow$ exponential solution $e^{\lambda t}\boldsymbol{x}$ when $A\boldsymbol{x} = \lambda \boldsymbol{x}$

Check : If $\boldsymbol{u} = e^{\lambda t}\boldsymbol{x}$ then $\dfrac{d\boldsymbol{u}}{dt} = \lambda e^{\lambda t}\boldsymbol{x} = Ae^{\lambda t}\boldsymbol{x} = A\boldsymbol{u}$ as required

$A = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$ has $\lambda_1 = 9$ $\quad \boxed{\boldsymbol{u}_1 = \boldsymbol{e^{9t}} \begin{bmatrix} \mathbf{1} \\ \mathbf{1} \end{bmatrix}}$ $\quad \dfrac{d\boldsymbol{u}_1}{dt} = 9e^{9t}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = e^{9t}A\begin{bmatrix} 1 \\ 1 \end{bmatrix} = A\boldsymbol{u}_1$

Initial condition $\boldsymbol{u}(0)$ at $t = 0$	Split $\boldsymbol{u}(0)$ into eigenvectors $\boldsymbol{x}$	$\boldsymbol{u}(0) = c_1\boldsymbol{x}_1 + \cdots + c_n\boldsymbol{x}_n$
Each eigenvector goes its own way	Combine solutions	$\boxed{\boldsymbol{u}(t) = \boldsymbol{c_1 e^{\lambda_1 t} x_1} + \cdots + \boldsymbol{c_n e^{\lambda_n t} x_n}}$
Special case	$\lambda_1 = \lambda_2$ with one eigenvector $\boldsymbol{x}$	$\boldsymbol{u}(t) = c_1\boldsymbol{e^{\lambda_1 t}}\boldsymbol{x}_1 + c_2\boldsymbol{te^{\lambda_1 t}x_1}$

Stability $\quad \boldsymbol{u}(t) \to \mathbf{0}$ if all eigenvalues $\lambda = a + ib$ have real part $\boldsymbol{a} < \mathbf{0}$

Weak stability $\quad \boldsymbol{u}(t) \to$ steady state if one λ moves up to $\boldsymbol{\lambda} = \mathbf{0}$

Matrix Exponential	$\boldsymbol{e}^{At}$ is the solution matrix	$\boxed{e^{At}\,\boldsymbol{u}(0) = u(t)}$	$e^{At}\boldsymbol{x} = e^{\lambda t}\boldsymbol{x}$
Exponential Series $\boldsymbol{e^{At}}$	$\boldsymbol{e^x} = 1 + x + \dfrac{x^2}{2} + \cdots + \dfrac{x^n}{n!} + \cdots$	$\boldsymbol{e^A = I + A + \dfrac{A^2}{2} + \cdots + \dfrac{A^n}{n!}} + \cdots$	

6.5 Matrices in Engineering : Second differences

Centered difference $\dfrac{du}{dx} \approx \dfrac{u\,(x+h)-u\,(x-h)}{2h}$ **Second difference** $\dfrac{\boldsymbol{d^2u}}{\boldsymbol{dx^2}} \approx \dfrac{\boldsymbol{u\,(x+h)-2u\,(x)+u\,(x-h)}}{\boldsymbol{h^2}}$

Second differences with $u_0 = u_4 = 0$
$$\frac{-\,d^2u}{dx^2} \approx \frac{1}{h^2}\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \frac{1}{h^2}K\boldsymbol{u}$$

Eigenvalues $\boldsymbol{2-\sqrt{2}, 2, 2+\sqrt{2}}$ Pivots $\boldsymbol{\frac{2}{1}, \frac{3}{2}, \frac{4}{3}}$ Determinants $\boldsymbol{2, 3, 4}$

Energy $\boldsymbol{x^{\mathrm{T}}Kx}$
$$\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}\begin{bmatrix} & & \\ & \boldsymbol{K} & \\ & & \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 2(x_1^2 - x_1x_2 + x_2^2 - x_2x_3 + x_3^2) > 0$$
$\boldsymbol{K}$ **is positive definite**

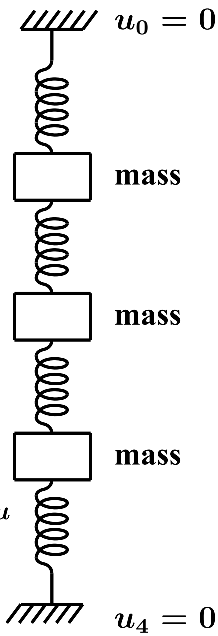
Steady state

Spring forces

Balance gravity

$\boldsymbol{Ku = g}$

Displacements $\boldsymbol{u}$



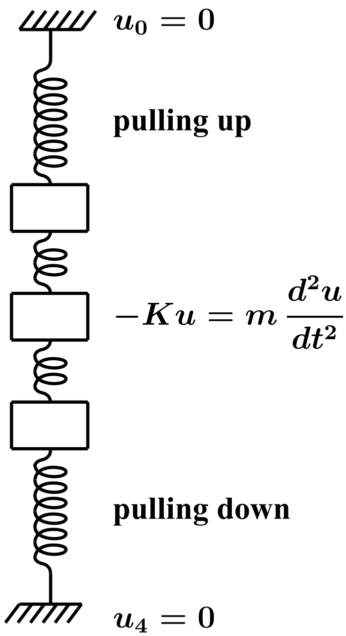
Oscillation

Springs pull

Masses move

Newton's Law $\boldsymbol{F = ma}$

Use eigenvalues of $\boldsymbol{K}$



Part 7

Singular Values and Vectors : $Av = \sigma u$ and $A = U\Sigma V^{\mathrm{T}}$

Part 7 : Singular Values and Vectors : $\boldsymbol{Av = \sigma u}$ and $\boldsymbol{A = U\Sigma V^{\mathrm{T}}}$

7.1 Singular Vectors in $\boldsymbol{U}$, $\boldsymbol{V}$ and Singular Values in $\boldsymbol{\Sigma}$

An example shows orthogonal vectors $\boldsymbol{v}$ going into orthogonal vectors $\boldsymbol{u}$

$$A\boldsymbol{v}_1 = \begin{bmatrix} \mathbf{3} & 0 \\ \mathbf{4} & \mathbf{5} \end{bmatrix} \begin{bmatrix} \mathbf{1} \\ \mathbf{1} \end{bmatrix} = \begin{bmatrix} \mathbf{3} \\ \mathbf{9} \end{bmatrix} \quad \text{and} \quad A\boldsymbol{v}_2 = \begin{bmatrix} \mathbf{3} & 0 \\ \mathbf{4} & \mathbf{5} \end{bmatrix} \begin{bmatrix} \mathbf{-1} \\ \mathbf{1} \end{bmatrix} = \begin{bmatrix} \mathbf{-3} \\ \mathbf{1} \end{bmatrix}$$

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \text{ is orthogonal to } \boldsymbol{v}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \qquad \begin{bmatrix} 3 \\ 9 \end{bmatrix} \text{ is orthogonal to } \begin{bmatrix} -3 \\ 1 \end{bmatrix}$$

Divide both inputs $\boldsymbol{v}$ by $\sqrt{2}$ Divide both outputs $\boldsymbol{u}$ by $\sqrt{10}$

Matrix form
$\boldsymbol{AV} = \boldsymbol{U\Sigma}$

$$\begin{bmatrix} \mathbf{3} & 0 \\ \mathbf{4} & \mathbf{5} \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 \end{bmatrix} \begin{bmatrix} \mathbf{3\sqrt{5}} & 0 \\ 0 & \mathbf{\sqrt{5}} \end{bmatrix}$$

$\boldsymbol{V}$ and $\boldsymbol{U}$ = orthogonal matrices $\boxed{\boldsymbol{V^{\mathrm{T}}V = I} \quad \boldsymbol{U^{\mathrm{T}}U = I} \quad \boldsymbol{A = U\Sigma V^{\mathrm{T}}}}$

$\boldsymbol{v}_1, \boldsymbol{v}_2$ = orthogonal basis for the **row space** = **inputs**

$\boldsymbol{u}_1, \boldsymbol{u}_2$ = orthogonal basis for the **column space** = **outputs**

$\boldsymbol{\sigma}_1 = 3\sqrt{5}$ and $\boldsymbol{\sigma}_2 = \sqrt{5}$ are the **singular values** of this A

7.2 Reduced SVD / Full SVD / Construct $U\Sigma V^{\mathrm{T}}$ from $A^{\mathrm{T}}A$

Reduced SVD : Stop at $\boldsymbol{u_r}$ and $\boldsymbol{v_r}$ **Full SVD** : Go on to $\boldsymbol{u_m}$ and $\boldsymbol{v_n}$

$$\boldsymbol{A} = \boldsymbol{U}_r\Sigma_r\boldsymbol{V}_r^{\mathrm{T}} = \begin{bmatrix} \boldsymbol{u}_1 \text{ to } \boldsymbol{u}_r \\ \text{column space} \\ m \times r \end{bmatrix} \begin{bmatrix} \boldsymbol{\sigma}_1 & & \\ & \ddots & \\ r \text{ by } r & & \boldsymbol{\sigma}_r \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_1^{\mathrm{T}} & \text{row} \\ \text{to} & \text{space} \\ \boldsymbol{v}_r^{\mathrm{T}} & r \times n \end{bmatrix}$$

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}} = \begin{bmatrix} \boldsymbol{u}_1 \text{ to } \boldsymbol{u}_m \\ \\ \text{columns} \\ \\ m \times m \end{bmatrix} \begin{bmatrix} \boldsymbol{\sigma}_1 & & \\ & \cdot & \\ & & \boldsymbol{\sigma}_r \\ m \text{ by } n & & \mathbf{0}\ \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_1^{\mathrm{T}} & \text{row space} \\ \boldsymbol{v}_r^{\mathrm{T}} & n \times n \\ \boldsymbol{v}_n^{\mathrm{T}} & \text{nullspace} \end{bmatrix}$$

Key ideas $\boldsymbol{A^{\mathrm{T}}A} = V\Sigma^{\mathrm{T}}U^{\mathrm{T}}U\Sigma V^{\mathrm{T}} = \boldsymbol{V\Sigma^{\mathrm{T}}\Sigma V^{\mathrm{T}}}$ $\boldsymbol{AA^{\mathrm{T}}} = \boldsymbol{U\Sigma\Sigma^{\mathrm{T}}U^{\mathrm{T}}}$

Eigenvectors ! $\boldsymbol{A^{\mathrm{T}}Av = \sigma^2 v}$ and $\boldsymbol{AA^{\mathrm{T}}u = \sigma^2 u}$	n $\boldsymbol{v}$'s and m $\boldsymbol{u}$'s
The $\boldsymbol{u}$'s are chosen so that $A\boldsymbol{v}_k = \sigma_k \boldsymbol{u}_k$	$\boldsymbol{\sigma_1} \geq \cdots \geq \boldsymbol{\sigma_r} > 0$

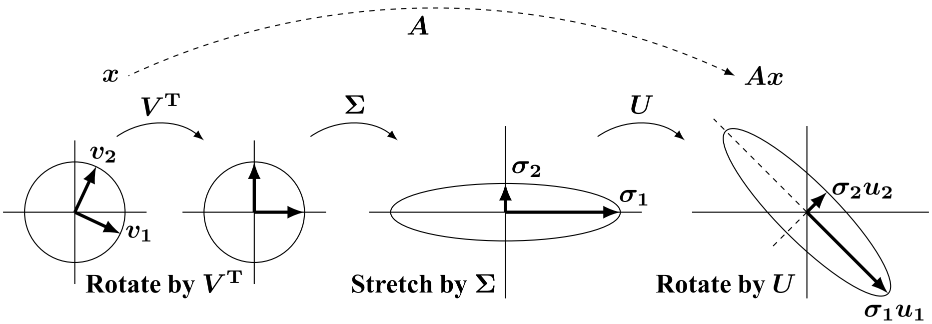
$\boldsymbol{k \leq r}$ $\boldsymbol{v_k}$ and $\boldsymbol{\sigma_k^2}$ from $\boldsymbol{A^{\mathrm{T}}A}$ $\boxed{\boldsymbol{u}_k = \dfrac{\boldsymbol{A}\boldsymbol{v}_k}{\boldsymbol{\sigma}_k}}$ $\boldsymbol{u}_j^{\mathrm{T}}\boldsymbol{u}_k = \left(\dfrac{A\boldsymbol{v}_j}{\sigma_j}\right)^{\mathrm{T}} \dfrac{A\boldsymbol{v}_k}{\sigma_k} = \dfrac{\sigma_k}{\sigma_j}\boldsymbol{v}_j^{\mathrm{T}}\boldsymbol{v}_k = \boldsymbol{0}$

$\boxed{\begin{array}{c}\textbf{Square}\,\boldsymbol{A} \\ \text{has } |\lambda| \leq \sigma_1\end{array}}$ $|\boldsymbol{\lambda}|\ ||\boldsymbol{x}|| = ||A\boldsymbol{x}|| = ||U\Sigma V^{\mathrm{T}}\boldsymbol{x}|| = ||\Sigma V^{\mathrm{T}}\boldsymbol{x}|| \leq \sigma_1\ ||V^{\mathrm{T}}\boldsymbol{x}|| = \boldsymbol{\sigma_1}\ ||\boldsymbol{x}||$

$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 8 \\ 0 & 0 & 0 \end{bmatrix}$ $\begin{array}{l} \lambda = 0, 0, 0 \\ \sigma = 8, 1, (0) \end{array}$ $A = \boldsymbol{u}_1\sigma_1\boldsymbol{v}_1^{\mathrm{T}} + \boldsymbol{u}_2\sigma_2\boldsymbol{v}_2^{\mathrm{T}}$ A has rank $r = 2$
2 singular values

$A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}$ $\begin{array}{l} \lambda = 5, 3 \\ \sigma = 3\sqrt{5}, \sqrt{5} \end{array}$ $\begin{array}{c} \boldsymbol{u}_1\sigma_1\boldsymbol{v}_1^{\mathrm{T}} \\ + \\ \boldsymbol{u}_2\sigma_2\boldsymbol{v}_2^{\mathrm{T}} \end{array} = \dfrac{3}{2}\begin{bmatrix} 1 & 1 \\ 3 & 3 \end{bmatrix} + \dfrac{1}{2}\begin{bmatrix} 3 & -3 \\ -1 & 1 \end{bmatrix} = \boldsymbol{A}$

7.3 The Geometry of the SVD : Rotate – Stretch – Rotate



$\boldsymbol{A} =$ (**Orthogonal**) (**Diagonal**) (**Orthogonal**)

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a} & \boldsymbol{b} \\ \boldsymbol{c} & \boldsymbol{d} \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \sigma_1 & \\ & \sigma_2 \end{bmatrix} \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix}$$

Four numbers $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{d}$ in A produce four numbers $\boldsymbol{\theta}, \boldsymbol{\sigma_1}, \boldsymbol{\sigma_2}, \boldsymbol{\phi}$ in the SVD

3×3: Nine numbers in A produce which 9 numbers for $U\Sigma V^{\mathrm{T}}$?

$n \times n$: An orthogonal matrix comes from $\frac{1}{2}n(n-1)$ simple rotations

Inputs $\boldsymbol{x} =$ unit circle	Outputs $A\boldsymbol{x} =$ stretched ellipse
Radius vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$	Axis vectors $\boldsymbol{\sigma_1 u_1}$ and $\boldsymbol{\sigma_2 u_2}$

7.4 A_k is Closest to A: Principal Component Analysis PCA

$$\textbf{SVD}\quad A = U\Sigma V^{\mathrm{T}} = \boldsymbol{u}_1\boldsymbol{\sigma}_1\boldsymbol{v}_1^{\mathrm{T}} + \cdots + \boldsymbol{u}_r\boldsymbol{\sigma}_r\boldsymbol{v}_r^{\mathrm{T}} \quad A \text{ has rank } r$$

$$A_k = U_k\Sigma_k V_k^{\mathrm{T}} = \boldsymbol{u}_1\boldsymbol{\sigma}_1\boldsymbol{v}_1^{\mathrm{T}} + \cdots + \boldsymbol{u}_k\boldsymbol{\sigma}_k\boldsymbol{v}_k^{\mathrm{T}} \quad \text{any } \boldsymbol{k} \leq \boldsymbol{r}$$

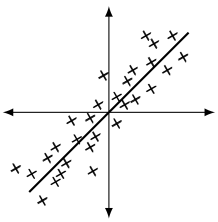
Great fact **This A_k from the SVD is the closest rank k matrix to A**

"Eckart-Young" $\boxed{||A - A_k|| \leq ||A - B|| \text{ if } B \text{ has rank } k}$

Matrix norms $||\boldsymbol{A}||_{\ell^2 \textbf{ norm}} = \boldsymbol{\sigma}_1 \quad ||\boldsymbol{A}||_{\textbf{Frobenius}} = \sqrt{\boldsymbol{\sigma}_1^2 + \cdots + \boldsymbol{\sigma}_r^2}$

$\boldsymbol{A}_0$ = matrix of data $\quad \boldsymbol{A}$ = subtract row average from each row of A_0

$$\boldsymbol{S} = \frac{\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}}{n-1} = \text{sample covariance matrix is symmetric positive definite}$$



Line closest to data is $\boldsymbol{u}_1$: The key to PCA
Straight line fit using perpendicular distances

$\boldsymbol{u}_1$ = eigenvector of S = first principal component
= singular vector of A = captures most variance

Total variance of the data = Add the eigenvalues of $S = \sigma_1^2 + \cdots + \sigma_r^2$

7.5 Computing Eigenvalues of S and Singular Values of A

Step 1 Produce zeros in the matrix $S \to Q^{-1}SQ = \boldsymbol{S}_0$ **Same $\boldsymbol{\lambda}$'s**

$Q, Q_1, Q_2 =$ orthogonal matrix $A \to Q_1^{-1}AQ_2 = \boldsymbol{A}_0$ **Same $\boldsymbol{\sigma}$'s**

New $\boldsymbol{S_0}$ has only 3 nonzero diagonals $\boldsymbol{A_0}$ has only 2 nonzero diagonals

Step 2 **"$\boldsymbol{QR}$ method"** uses Gram-Schmidt to orthogonalize columns

$S =$ (Orthogonal Q) (Upper triangular R) at every step

$$\boxed{\textbf{Factor } \boldsymbol{S_0 = Q_0 R_0} \quad \textbf{Reverse } \boldsymbol{S_1 = R_0 Q_0} \quad \textbf{Repeat } \boldsymbol{S_1 = Q_1 R_1} \text{ and } \boldsymbol{S_2 = R_1 Q_1}}$$

Amazing : The off-diagonal entries get small : **Watch $\boldsymbol{\sin\theta \to -\sin^3\theta}$**

$$\boldsymbol{S_k = Q_k R_k} \qquad \begin{bmatrix} \cos\theta & \sin\theta \\ \boldsymbol{\sin\theta} & 0 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} 1 & \sin\theta\cos\theta \\ 0 & -\sin^2\theta \end{bmatrix}$$

$$\boldsymbol{S_{k+1} = R_k Q_k} \qquad \begin{bmatrix} 1 & \sin\theta\cos\theta \\ 0 & -\sin^2\theta \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} * & * \\ \boldsymbol{-\sin^3\theta} & * \end{bmatrix}$$

$\boldsymbol{S_k}$ **approaches $\boldsymbol{\Lambda}$** : The eigenvalues λ begin to appear on the diagonal

Similar idea for SVD = Golub-Kahan algorithm : σ's appear on the diagonal

7.6 Compressing Images by the SVD

Flag with 3 stripes $\begin{bmatrix} B & B & B & B & B & B \\ B & B & B & B & B & B \\ W & W & W & W & W & W \\ W & W & W & W & W & W \\ R & R & R & R & R & R \\ R & R & R & R & R & R \end{bmatrix} = \begin{bmatrix} B \\ B \\ W \\ W \\ R \\ R \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$ **Rank one matrix Great compression**

Triangular flag $\begin{bmatrix} \mathbf{1} & 0 & 0 & 0 \\ \mathbf{1} & \mathbf{1} & 0 & 0 \\ \mathbf{1} & \mathbf{1} & \mathbf{1} & 0 \\ \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} \end{bmatrix}$ This has all $\sigma > \dfrac{1}{2}$ Poor compression

Compress photographs **https://timbaumann.info/svd-image-compression-demo/**

Uncompressed image $= 600 \times 600 = 360,000$ pixels

Compressed image $U\Sigma V^{\mathrm{T}} = 600 \times 100 + 100 + 100 \times 600 = 120,000$

Serious compression $U\Sigma V^{\mathrm{T}} = 600 \times 20 + 20 + 20 \times 600 = 24,020$

Compression is highly developed See Wikipedia for Eigenfaces

7.7 The Victory of Orthogonality

1 Length of $Q\boldsymbol{x}$ = Length of $\boldsymbol{x}$ $\quad ||Q\boldsymbol{x}||^2 = ||\boldsymbol{x}||^2 \quad (Q\boldsymbol{x})^{\mathrm{T}}(Q\boldsymbol{y}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$

2 All powers $\boldsymbol{Q}^n$ and all products $\boldsymbol{Q_1Q_2}$ remain orthogonal

3 Reflection $\quad \boldsymbol{H} = \boldsymbol{I} - 2\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}$ = orthogonal + symmetric when $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u} = 1$

4 Symmetric matrices have orthogonal eigenvectors $\boldsymbol{SQ} = \boldsymbol{Q\Lambda}$

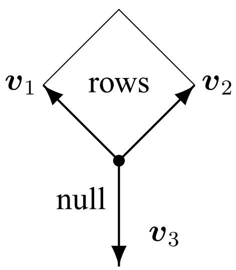
5 All matrices have orthogonal singular vectors $\boldsymbol{v}$'s and $\boldsymbol{u}$'s : $\boldsymbol{AV} = \boldsymbol{U\Sigma}$

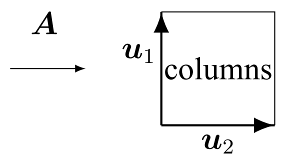
6 The pseudoinverse of $U\Sigma V^{\mathrm{T}}$ is $\boldsymbol{V\Sigma^{+}U^{\mathrm{T}}}$ $\quad$ Nonzeros in Σ^{+} are $\dfrac{1}{\sigma}$

7 Polar decomposition $\quad \boldsymbol{A} = \boldsymbol{QS}$ = (orthogonal) (symm positive definite)

8 Gram-Schmidt = Highly valuable $\quad \boldsymbol{A} = \boldsymbol{QR}$ = (orthogonal) (triangular)

9 Orthogonal functions for Fourier series $\quad \boldsymbol{f}(x) = \sum a_k \mathbf{cos}\, \boldsymbol{kx} + b_k \mathbf{sin}\, \boldsymbol{kx}$





$$A\boldsymbol{v}_1 = \sigma_1\boldsymbol{u}_1$$
$$A\boldsymbol{v}_2 = \sigma_2\boldsymbol{u}_2$$

Big picture of the SVD

Orthogonal rows $\rightarrow$ Orthogonal columns

Part 8
Linear Transformations and Their Matrices

Part 8 : Linear Transformations and Their Matrices

8.1 Examples of Linear Transformations

V and **Y** are vector spaces (the vectors can be matrices or functions !)

$\boldsymbol{T}$ is a linear transformation from **V** to **Y** (inputs to outputs)

Test for linearity $\boxed{\boldsymbol{T}(c\boldsymbol{v} + d\boldsymbol{w}) = c\,\boldsymbol{T}(\boldsymbol{v}) + d\,\boldsymbol{T}(\boldsymbol{w})}$ for all $\boldsymbol{v}, \boldsymbol{w}$ in $\boldsymbol{V}$

Example 1 $\mathbf{V} = x$-y plane Rotate the plane $\mathbf{R}^2$ by a fixed angle θ

Straight lines rotate into straight lines (required by linearity)

Center point $\mathbf{0} = (0, 0)$ stays put $T(\mathbf{0}+\mathbf{0}) = T(\mathbf{0}) + T(\mathbf{0})$ requires $T(\mathbf{0}) = \mathbf{0}$

This $\boldsymbol{T}$ has an inverse $\boldsymbol{T}^{-1}$: Rotate by $-\theta$ is another linear transformation

Example 2 Input space $\mathbf{V}$ = all 3 by 3 matrices = output space $\mathbf{Y}$

T sets all off-diagonal entries to zero T (matrix) = (diagonal matrix)

T^2 will be the same as T : T is like a projection on matrix space

Multiply transformations T_2T_1 Output space for T_1 = Input space for T_2

T_2T_1 obeys the same rule as matrix multiplication $T_2(T_1\boldsymbol{x}) = (T_2T_1)\boldsymbol{x}$

Example 3 $\mathbf{V}$ = all functions $a + bx + cx^2$ $\mathbf{Y}$ = all functions $d + ex$

$\boldsymbol{T}(a + bx + cx^2)$ = **derivative** of the input function = output $b + 2cx$

"Derivative" is a linear transformation ! Otherwise calculus would fail

"Integral" is also a linear transformation on a space of functions

8.2 Derivative Matrix D and Integral Matrix D^+

Choose basis $1, x, x^2$ for input space **V** : Quadratic functions

Choose basis $1, x$ for output space **Y** : Linear functions

Apply **derivative transformation** to the input basis $\boldsymbol{v}_1 = 1, \boldsymbol{v}_2 = x, \boldsymbol{v}_3 = x^2$

Express outputs $T(\boldsymbol{v}_1) = 0, T(\boldsymbol{v}_2) = 1, T(\boldsymbol{v}_3) = 2x$ in the output basis

$$T(\boldsymbol{v}_1) = \mathbf{0} \qquad T(\boldsymbol{v}_2) = \frac{dx}{dx} = 1 = \boldsymbol{u}_1 \qquad T(\boldsymbol{v}_3) = \frac{d}{dx}(x^2) = 2x = 2\boldsymbol{u}_2$$

The columns of D show those derivatives with respect to the bases

$$\boldsymbol{D} = \begin{bmatrix} \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{2} \end{bmatrix} = \text{matrix form of the derivative } T = \frac{d}{dx}$$

$$\boldsymbol{D} \text{ times } \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} b \\ 2c \end{bmatrix} \text{ tells us the outputs from the inputs } a, bx, cx^2$$

Integral transformation $\boldsymbol{S}$ from Y back to V Inputs $1, x$ Outputs $1, x, x^2$

$$\boldsymbol{S}(1) = x = \boldsymbol{v}_2 \qquad \boldsymbol{S}(x) = \tfrac{1}{2}x^2 = \tfrac{1}{2}\boldsymbol{v}_3 \qquad \text{Integral matrix } \boldsymbol{E} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \frac{\mathbf{1}}{\mathbf{2}} \end{bmatrix}$$

Fundamental Theorem of Calculus : Derivative of integral of f is f

$$DE = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \textbf{identity} \text{ transformation on } \mathbf{Y}$$

$$ED = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \text{only a } \textbf{projection} \text{ on } \mathbf{V}$$

$\boldsymbol{E}$ = pseudoinverse $\boldsymbol{D}^+$ of $\boldsymbol{D}$ Not the inverse because derivative of 1 is 0

8.3 Basis for V and Basis for Y $\Rightarrow$ Matrix for T : V$\to$Y

Every linear transformation $\boldsymbol{T}$: $\mathbf{V} \to \mathbf{Y}$ can be expressed by a matrix A

That matrix A depends on the basis for $\mathbf{V}$ and the basis for $\mathbf{Y}$

To construct A: Apply T to the input basis vectors $\boldsymbol{v}_1$ to $\boldsymbol{v}_n$

Then $T(\boldsymbol{v}_j) = a_{1j}\boldsymbol{y}_1 + a_{2j}\boldsymbol{y}_2 + \cdots + a_{mj}\boldsymbol{y}_m$ gives **column $\boldsymbol{j}$** of A

Input $\boldsymbol{v} = c_1\boldsymbol{v}_1 + \cdots + c_n\boldsymbol{v}_n$ Output $\boldsymbol{y} = c_1T(\boldsymbol{v}_1) + \cdots + c_nT(\boldsymbol{v}_n)$

$\boxed{\text{That output } \boldsymbol{y} \text{ has coefficients } A\boldsymbol{c} \text{ in the output basis for } \mathbf{Y}}$

Main point! **Multiplication by $\boldsymbol{A}$ copies the linear transformation $\boldsymbol{T}$**

Both linear and both correct for basis $\Rightarrow$ both correct for every input

Change input basis to $\boldsymbol{V}_1, \ldots, \boldsymbol{V}_n$ Change output basis to $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m$

The matrix for the same $\boldsymbol{T}$ in these new bases is $\boxed{\boldsymbol{M} = \boldsymbol{Y}^{-1}\boldsymbol{A}\boldsymbol{V}}$

$\boldsymbol{V}$ = identity on input space: but basis change from $\boldsymbol{v}$'s to $\boldsymbol{V}$'s

$\boldsymbol{Y}$ = identity on output space: but basis change from $\boldsymbol{y}$'s to $\boldsymbol{Y}$'s

Part 9
Complex Numbers and the Fourier Matrix

9.1 **Complex Numbers $x + iy = re^{i\theta}$: Unit circle $r = 1$**

9.2 **Complex Matrices : Hermitian $S = \overline{S}^{\mathrm{T}}$ and Unitary $Q^{-1} = \overline{Q}^{\mathrm{T}}$**

9.3 **Fourier Matrix F and the Discrete Fourier Transform**

9.4 **Cyclic Convolution and the Convolution Rule**

9.5 **FFT : The Fast Fourier Transform**

$\mathbf{R}$ = line of all real numbers $-\infty < x < \infty$ $\leftrightarrow$ $\mathbf{C}$ = plane of all complex numbers $z = x + iy$

$|x|$ = absolute value of x $\leftrightarrow$ $|\boldsymbol{z}| = \sqrt{\boldsymbol{x^2 + y^2}} = \boldsymbol{r}$ = absolute value (or modulus) of z

1 and -1 solve $x^2 = 1$ $\leftrightarrow$ $z = 1, w, \ldots, w^{n-1}$ solve $z^n = 1$ where $w = e^{2\pi i/n}$

The **complex conjugate** of $z = x + iy$ is $\overline{\boldsymbol{z}} = \boldsymbol{x - iy}$. $|z|^2 = x^2 + y^2 = z\overline{z}$ and $\dfrac{1}{z} = \dfrac{\overline{z}}{|z|^2}$.

The **polar form** of $z = x + iy$ is $|z|e^{i\theta} = \boldsymbol{re^{i\theta}} = \boldsymbol{r\cos\theta + ir\sin\theta}$. The angle has $\tan\theta = \dfrac{y}{x}$.

$\mathbf{R}^n$: vectors with n real components	$\leftrightarrow$	$\mathbf{C}^n$: vectors with n complex components
length: $\\|\boldsymbol{x}\\|^2 = x_1^2 + \cdots + x_n^2$	$\leftrightarrow$	length: $\\|\boldsymbol{z}\\|^2 = \|z_1\|^2 + \cdots + \|z_n\|^2$
transpose: $(A^{\mathrm{T}})_{ij} = A_{ji}$	$\leftrightarrow$	conjugate transpose: $(A^{\mathrm{H}})_{ij} = \overline{A_{ji}}$
dot product: $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = x_1y_1 + \cdots + x_ny_n$	$\leftrightarrow$	inner product: $\boldsymbol{u}^{\mathrm{H}}\boldsymbol{v} = \overline{u}_1v_1 + \cdots + \overline{u}_nv_n$
reason for A^{T}: $(A\boldsymbol{x})^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{x}^{\mathrm{T}}(A^{\mathrm{T}}\boldsymbol{y})$	$\leftrightarrow$	reason for A^{H}: $(A\boldsymbol{u})^{\mathrm{H}}\boldsymbol{v} = \boldsymbol{u}^{\mathrm{H}}(A^{\mathrm{H}}\boldsymbol{v})$
orthogonality: $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = 0$	$\leftrightarrow$	orthogonality: $\boldsymbol{u}^{\mathrm{H}}\boldsymbol{v} = 0$
symmetric matrices: $S = S^{\mathrm{T}}$	$\leftrightarrow$	Hermitian matrices: $S = S^{\mathrm{H}}$
$\boldsymbol{S = Q\Lambda Q^{-1}} = \boldsymbol{Q\Lambda Q}^{\mathrm{T}}$(real Λ)	$\leftrightarrow$	$\boldsymbol{S = U\Lambda U^{-1}} = \boldsymbol{U\Lambda U}^{\mathrm{H}}$ (real Λ)
orthogonal matrices: $Q^{\mathrm{T}} = Q^{-1}$	$\leftrightarrow$	unitary matrices: $U^{\mathrm{H}} = U^{-1}$
$(Q\boldsymbol{x})^{\mathrm{T}}(Q\boldsymbol{y}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$ and $\\|Q\boldsymbol{x}\\| = \\|\boldsymbol{x}\\|$	$\leftrightarrow$	$(U\boldsymbol{x})^{\mathrm{H}}(U\boldsymbol{y}) = \boldsymbol{x}^{\mathrm{H}}\boldsymbol{y}$ and $\\|U\boldsymbol{z}\\| = \\|\boldsymbol{z}\\|$

Part 9 : Complex Numbers and the Fourier Matrix

9.1 Complex Numbers $x + iy = re^{i\theta}$: Unit circle $r = 1$

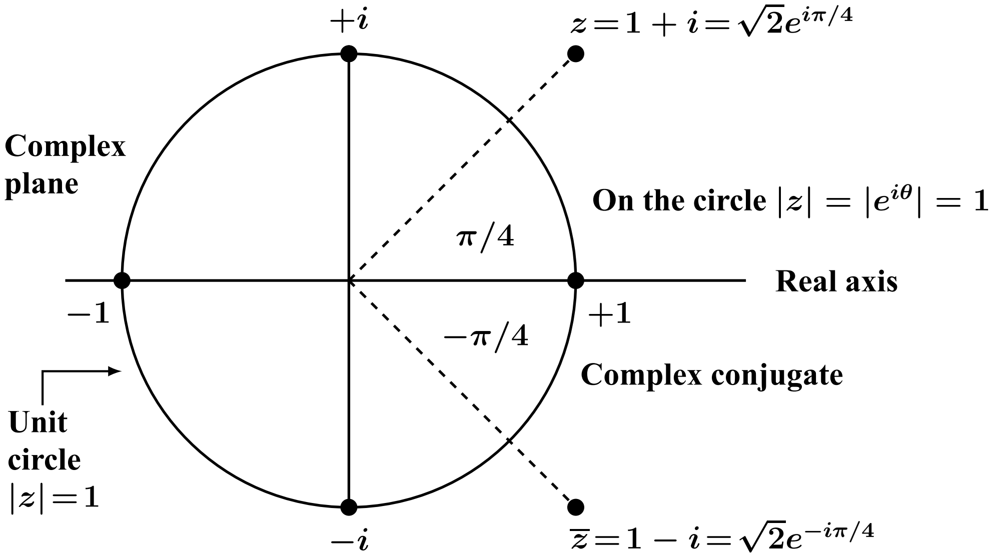
Complex numbers $z = x + iy$ $\quad$ x = real part $\quad$ y = imaginary part

Magnitude $|z| = r = \sqrt{x^2 + y^2}$ $\quad$ **Angle** $\tan\theta = \dfrac{y}{x}$

$$\boxed{\textbf{Euler's Formula} \quad z = re^{i\theta} = r\cos\theta + ir\sin\theta = r\,\frac{x}{r} + ir\,\frac{y}{r}}$$

Complex conjugate $\overline{z} = x - iy = re^{-i\theta}$ $\quad$ Then $\overline{z}z = x^2 + y^2 = r^2$

Add $\overline{z}_1 + \overline{z}_2 = \overline{z_1 + z_2}$ $\quad$ **Multiply** $(\overline{z}_1)(\overline{z}_2) = \overline{z_1 z_2}$ $\quad$ **Divide** $\dfrac{1}{z} = \dfrac{\overline{z}}{|z|^2} = \dfrac{a - ib}{a^2 + b^2}$



Add angles $\quad (re^{i\theta})(Re^{i\phi}) = rR\,e^{i(\theta+\phi)}$ $\qquad (i)^4 = (e^{i\pi/2})^4 = e^{i2\pi} = 1$

9.2 Complex Matrices : Hermitian $S = \overline{S}^{\mathrm{T}}$ and Unitary $Q^{-1} = \overline{Q}^{\mathrm{T}}$

Rule When you transpose, take complex conjugates : $\overline{\boldsymbol{x}}^{\mathrm{T}}\ \overline{\boldsymbol{A}}^{\mathrm{T}}$

Automatic for computer systems like MATLAB and Julia

Inner product = Dot product = $\overline{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{y} = \overline{x}_1 y_1 + \cdots + \overline{x}_n y_n$

Length squared $= ||\boldsymbol{x}||^2 = \overline{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{x} = |x_1|^2 + \cdots + |x_n|^2 = ||Re\,\boldsymbol{x}||^2 + ||Im\,\boldsymbol{x}||^2$

Hermitian matrix $\boldsymbol{S} = \begin{bmatrix} 2 & \mathbf{3-3}\boldsymbol{i} \\ \mathbf{3+3}\boldsymbol{i} & 5 \end{bmatrix} = \overline{\boldsymbol{S}}^{\mathrm{T}}$ $\boxed{\begin{array}{c}\text{Real diagonal}\\ \boldsymbol{S_{ji}} = \overline{\boldsymbol{S}}_{\boldsymbol{ij}}\end{array}}$

S has **real eigenvalues** 8, $\mathbf{-1}$ and **perpendicular eigenvectors**

$\det(S - \lambda I) = \lambda^2 - 7\lambda + 10 - \left|3 + 3i\right|^2 = (\boldsymbol{\lambda} - \mathbf{8})\,(\boldsymbol{\lambda} + \mathbf{1})$

$$(\boldsymbol{S} - \mathbf{8}\boldsymbol{I})\begin{bmatrix} \mathbf{1} \\ \mathbf{1} + \boldsymbol{i} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad (\boldsymbol{S} + \boldsymbol{I})\begin{bmatrix} \mathbf{1} - \boldsymbol{i} \\ \mathbf{-1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Unitary matrix
Orthonormal columns $\boldsymbol{Q} = \dfrac{1}{\sqrt{3}}\begin{bmatrix} 1 & 1-i \\ 1+i & -1 \end{bmatrix}$ $\boxed{\overline{\boldsymbol{Q}}^{\mathrm{T}} = \boldsymbol{Q}^{-1}}$ $||Q\boldsymbol{z}|| = ||\boldsymbol{z}||$, $|\boldsymbol{\lambda}| = \mathbf{1}$

The Fourier matrix $\dfrac{\boldsymbol{F}}{\sqrt{N}}$ is the most important unitary matrix

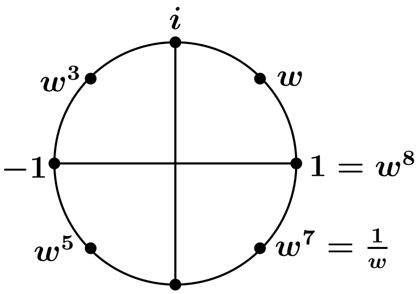
9.3 Fourier Matrix F and the Discrete Fourier Transform

Fourier Matrix $F_4 = \begin{bmatrix} 1 & \mathbf{1} & 1 & 1 \\ 1 & \boldsymbol{i} & i^2 & i^3 \\ 1 & \boldsymbol{i^2} & i^4 & i^6 \\ 1 & \boldsymbol{i^3} & i^6 & i^9 \end{bmatrix}$ **DFT Matrix** $\overline{F_4} =$ **powers of** $-i$

F_N and $\overline{F_N}$
N by N **matrices**

Replace $i = e^{2\pi i/4}$
by $w = e^{2\pi i/N}$

$\boxed{\begin{array}{l} F_{jk} = w^{jk} = e^{2\pi ijk/N} \\ \text{Columns } k = 0 \text{ to } N-1 \end{array}}$



$w = e^{2\pi i/8}$ $w^8 = 1$

$1 + w + w^2 + \cdots + w^{N-1} = 0$

$\boxed{\overline{F_N}F_N = NI}$ Then $F_N/\sqrt{N}$ is a unitary matrix. It has orthonormal columns

$N = 2$
$w = e^{\pi i} = -1$

$F_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ $\overline{F}_2 F_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = NI$

Discrete Fourier Transform f **to** c $c = F_N^{-1} f$

Inverse Fourier Transform c **to** f $f = F_N c$

9.4 Cyclic Convolution and the Convolution Rule

$\boldsymbol{c} = (1,2,3)$ Convolution $\boldsymbol{c} * \boldsymbol{d} = (\mathbf{5}, \mathbf{10}, \mathbf{19}, \mathbf{8}, \mathbf{12})$

$\boldsymbol{d} = (5,0,4)$ Cyclic $\boldsymbol{c} \circledast \boldsymbol{d} = (\mathbf{5} + \mathbf{8}, \mathbf{10} + \mathbf{12}, \mathbf{19})$

$(1 + 2x + 3x^2)\,(5 + 4x^2) = 5 + 10x + 19x^2 + 8x^3 + 12x^4$

$\boldsymbol{x^3 = 1}$ **for cyclic** $(5+8) + (10+12)x + 19x^2$

$$
\begin{array}{rrrrr}
 & & \mathbf{1} & \mathbf{2} & \mathbf{3} \\
 & & \mathbf{5} & \mathbf{0} & \mathbf{4} \\
\hline
 & & 4 & 8 & 12 \\
5 & 10 & 15 & & \\
\hline
\mathbf{5} & \mathbf{10} & \mathbf{19} & \mathbf{8} & \mathbf{12}
\end{array}
$$

Convolution matrices
Constant diagonals

$$C = \begin{bmatrix} \mathbf{1} & 3 & 2 \\ \mathbf{2} & 1 & 3 \\ \mathbf{3} & 2 & 1 \end{bmatrix} \quad D = \begin{bmatrix} \mathbf{5} & 4 & 0 \\ \mathbf{0} & 5 & 4 \\ \mathbf{4} & 0 & 5 \end{bmatrix} \quad \begin{matrix} CD = \\ DC = \end{matrix} \begin{bmatrix} \mathbf{13} & 19 & 22 \\ \mathbf{22} & 13 & 19 \\ \mathbf{19} & 22 & 13 \end{bmatrix}$$

Eigenvectors of $\boldsymbol{C}$ and $\boldsymbol{D}$ $=$ columns of the Fourier matrix F

Eigenvalues of $\boldsymbol{C} = \boldsymbol{F^{-1}c}$ **Eigenvalues of $\boldsymbol{D} = \boldsymbol{F^{-1}d}$**

Convolve vectors **Multiply transforms**	**Convolution Rule** $\boldsymbol{F}(\boldsymbol{c} \circledast \boldsymbol{d}) = (\boldsymbol{Fc}) \mathbin{.*} (\boldsymbol{Fd})$
$\boldsymbol{x} \mathbin{.*} \boldsymbol{y} = (x_1y_1, \ldots, x_ny_n)$	**Key to Signal Processing**

9.5 FFT : The Fast Fourier Transform

Direct matrix multiplication of $\boldsymbol{c}$ by F_N needs $\boldsymbol{N^2}$ multiplications

FFT factorization with many zeros : $\dfrac{\mathbf{1}}{\mathbf{2}}\boldsymbol{N}\,\mathbf{log_2}\,\boldsymbol{N}$ multiplications

$\boldsymbol{N} = \mathbf{2^{10}} = \mathbf{1024}$ $\qquad \mathbf{log_2}\,\boldsymbol{N} = \mathbf{10}$ $\quad$ 1 million reduced to 5000

Step 1 of the FFT : From 1024 to 512 (Cooley-Tukey)

$$\begin{bmatrix} \boldsymbol{F_{1024}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{D} \\ \boldsymbol{I} & -\boldsymbol{D} \end{bmatrix} \begin{bmatrix} \boldsymbol{F_{512}} & 0 \\ 0 & \boldsymbol{F_{512}} \end{bmatrix} \begin{bmatrix} \boldsymbol{P_{1024}} \end{bmatrix}$$

Permutation P_{1024} puts columns $0, 2, \ldots, 1022$ ahead of $1, 3, \ldots, 1023$

Two zero blocks reduce the computing time nearly by 50%

Step 2 of the FFT : 512 to 256 (same factorization of F_{512})

Recursion continues to small N : $\mathbf{log_2}\,\boldsymbol{N}$ **steps** to Fast Transform

Each step has $\boldsymbol{N}$ multiplications from the diagonal matrices $\boldsymbol{D}$

One overall permutation $=$ product of the $\boldsymbol{P}$'s

FFTW is hardwired in many computers / bases other than 2

Part 10
Learning from Data by Gradient Descent

Part 10 : Learning from Data by Gradient Descent

10.1 Learning Function $F(x, v_0)$: Data v_0 and Weights x

Training data $=$ p features for N samples $=$ N vectors $\boldsymbol{v}_0$

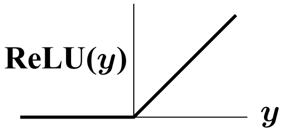
Each of those N vectors enters level zero of the neural net

Level k from $k-1$: Multiply each vector by A_k, add $\boldsymbol{b}_k$, apply ReLU

$$\boxed{\boldsymbol{v}_k = F_k(\boldsymbol{v}_{k-1}) = \text{ReLU}(A_k \boldsymbol{v}_{k-1} + \boldsymbol{b}_k)}$$

ReLU applies to each component of each vector : $\textbf{ReLU}(\boldsymbol{y}) = \textbf{max}(\boldsymbol{y}, \boldsymbol{0})$

ReLU $=$ ramp function
$=$ **Rec**tified **L**inear **U**nit

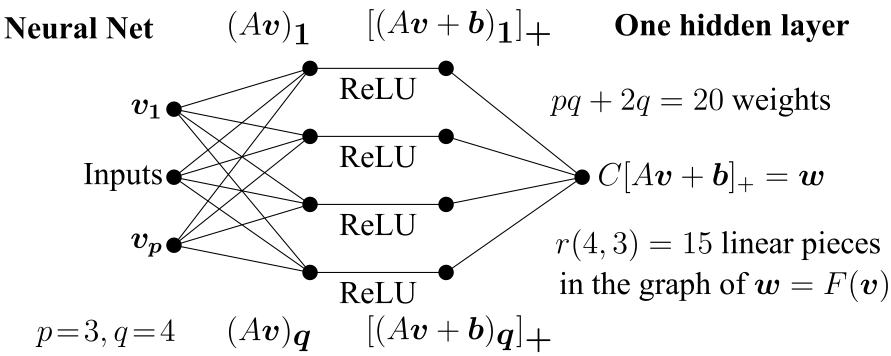


This gives the nonlinearity that learning functions need

Levels 0 to L $\boxed{\text{Output } \boldsymbol{v}_L = F_L(\ldots(F_2(F_1(\boldsymbol{v}_0)))) = F(\boldsymbol{v}_0)}$

$\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{v}_0)$ $=$ Composition of L piecewise linear F_k : F is piecewise linear

Each level contributes a weight matrix A_k and a vector $\boldsymbol{b}_k$ to $\boldsymbol{x}$

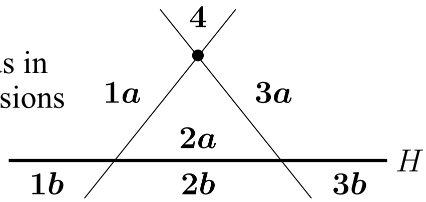


10.2 Counting Flat Pieces in the Graph of F

The weight matrices A_k and bias vectors $\boldsymbol{b}_k$ produce $F =$ learning function

Each application of ReLU creates a fold in the graph of F

$\boldsymbol{N} = \mathbf{3}$ folds in
$\boldsymbol{p} = \mathbf{2}$ dimensions



Start with 2 folds
$\leftarrow r(2,2) = 4$
Add new fold H
$\leftarrow r(2,1) = 3$
Total 7 regions

The $r(2,1) = 3$ pieces of the new fold H create new regions $\mathbf{1}\boldsymbol{b}, \mathbf{2}\boldsymbol{b}, \mathbf{3}\boldsymbol{b}$. Then the count becomes $r(3,2) = 4+3 = 7$ flat regions in the continuous piecewise linear surface $\boldsymbol{v}_2 = F(\boldsymbol{v}_0)$. A fourth fold would cross all three existing folds and create 4 new regions, so $r\,(4,2) = 7 + 4 = 11$.

The count $\boldsymbol{r}$ of linear pieces of F will follow from the recursive formula

$$\boxed{r(N,p) = r(N-1,p) + r(N-1,p-1)}$$

Theorem For $\boldsymbol{v}$ in $\mathbf{R}^p$, suppose the graph of $F(\boldsymbol{v})$ has folds along N hyperplanes $H_1, \ldots, H_N$. Those come from ReLU at N neurons. Then the number of regions bounded by the N hyperplanes is $\boldsymbol{r(N,p)}$:

$$\boldsymbol{r(N,p)} = \begin{pmatrix} \boldsymbol{N} \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \boldsymbol{N} \\ \mathbf{1} \end{pmatrix} + \cdots + \begin{pmatrix} \boldsymbol{N} \\ \boldsymbol{p} \end{pmatrix}.$$

These binomial coefficients are

$$\begin{pmatrix} \boldsymbol{N} \\ \boldsymbol{i} \end{pmatrix} = \frac{N\,!}{i\,!(N-i)\,!} \text{ with } 0\,! = 1 \text{ and } \begin{pmatrix} \boldsymbol{N} \\ \mathbf{0} \end{pmatrix} = 1 \text{ and } \begin{pmatrix} \boldsymbol{N} \\ \boldsymbol{i} \end{pmatrix} = 0 \text{ for } i > N.$$

With more layers: N folds from N ReLU's: **still** $\approx \boldsymbol{r(N,p)} \approx \boldsymbol{cN^p}$ **pieces**

10.3 Minimizing the Loss : Stochastic Gradient Descent

The **gradient** of a function $\boldsymbol{F}(x_1, \ldots, x_p)$ is a vector $\boldsymbol{\nabla F} = \text{grad}\,\boldsymbol{F}$

$\boldsymbol{\nabla F} = (\partial F/\partial x_1, \ldots, \partial F/\partial x_n)$ points in the steepest direction for $F(\boldsymbol{x})$

The graph of $\boldsymbol{y} = \boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{F}(x_1, \ldots, x_N)$ is a surface in $N+1$ dimensions

The graph of $F = x_1^2 + x_2^2 + 5$ is a bowl in $2+1=3$ dimensions

Minimum of $F = ||\boldsymbol{x}||^2 + 5$ is $F_{\text{min}} = 5$ at the point $\boldsymbol{x} = \textbf{arg min}\,\boldsymbol{F} = \boldsymbol{0}$

We want the minimizing point $\boldsymbol{x} = \textbf{arg min}\,\boldsymbol{F}$ for a complicated $F(\boldsymbol{x})$

Gradient descent starts from a point $\boldsymbol{x_0}$. Go down along the gradient $\boldsymbol{\nabla F}(\boldsymbol{x_0})$

Stop at a point $\boldsymbol{x_1} = \boldsymbol{x_0} - s\boldsymbol{\nabla F}(\boldsymbol{x_0})$. **Stepsize**$=$**learning rate**$=s=$maybe .001

Recompute the gradient $\boldsymbol{\nabla F}(\boldsymbol{x}_1)$ at the new point $\boldsymbol{x}_1$

$$\boxed{\text{At every step follow the gradient } \boldsymbol{\nabla F}(\boldsymbol{x}_k) \text{ to } \boldsymbol{x}_{k+1} = \boldsymbol{x}_k - s_k \boldsymbol{\nabla F}(\boldsymbol{x}_k)}$$

Big Problem 1 Many unknowns x_1 to x_N : all weights in all L layers

Big Problem 2 $\boldsymbol{F}(\boldsymbol{x}) =$ sum of errors in all training samples : many terms

Error **Square loss** $\left\| \begin{matrix}\text{output}\\ \text{layer}L\end{matrix} - \begin{matrix}\text{known}\\ \text{output}\end{matrix} \right\|^2$ or **"Cross-entropy loss"**

Solution 1 Use error in only **one** randomly chosen sample / one $\boldsymbol{v}_0$

Solution $\boldsymbol{B}$ Use sum of errors in only $\boldsymbol{B}$ random samples : **minibatch**

Stochastic gradient descent has new sampling at every step. **Successful**

10.4 Slow Convergence with Zigzag : Add Momentum

Test example : Minimize $\boldsymbol{F}(\boldsymbol{x},\boldsymbol{y}) = \frac{\mathbf{1}}{\mathbf{2}}(\boldsymbol{x}^{\mathbf{2}} + \boldsymbol{b}\boldsymbol{y}^{\mathbf{2}})$ with small $\boldsymbol{b} > 0$

Gradient $\boldsymbol{\nabla F} = (x, by)$. Exact search $(x_{k+1}, y_{k+1}) = (x_k, y_k) - (\text{best } s)\boldsymbol{\nabla F}$

$$\boxed{\boldsymbol{x}_k = b\left(\frac{b-1}{b+1}\right)^k \qquad \boldsymbol{y}_k = \left(\frac{1-b}{1+b}\right)^k \qquad \boldsymbol{F}(\boldsymbol{x}_k, \boldsymbol{y}_k) = \left(\frac{1-b}{1+b}\right)^{2k} F(b,1)}$$

Crucial ratio $\left(\dfrac{1-b}{1+b}\right)^2$ is near 1 for small b : **Slow convergence !**

The path **zig-zags** across a narrow valley : moves slowly down to $(0,0)$

Heavy ball **Add momentum** **Direction $\boldsymbol{z_{k+1}}$ remembers $\boldsymbol{z_k}$**

$$\boxed{(x,y)_{k+1} = (x,y)_k - s\boldsymbol{z}_k \qquad \boldsymbol{z}_{k+1} - \boldsymbol{\nabla F}(x,y)_{k+1} = \beta \boldsymbol{z}_k}$$

Optimal s
Optimal β give fast descent : ratio $\dfrac{1-b}{1+b}$ changes to $\dfrac{1-\sqrt{b}}{1+\sqrt{b}}$

$$b = \frac{1}{100} \qquad \left(\frac{1-b}{1+b}\right)^2 = \left(\frac{.99}{1.01}\right)^2 \approx \mathbf{.96} \text{ changes to } \left(\frac{0.9}{1.1}\right)^2 \approx \mathbf{.67}\,!$$

ADAM G_k combines all earlier gradients by $\boldsymbol{G}_k = \delta \boldsymbol{G}_{k-1} + (1-\delta)\boldsymbol{\nabla F}(\boldsymbol{x}_k)$

Question **Why do the weights (matrices $\boldsymbol{A_k}$) work well for unseen data ?**

10.5 Convolutional Neural Nets : CNN in 1D and 2D

Convolution matrix = Moving average

A has constant diagonals

Sliding window : same weights $\frac{1}{2}, \frac{1}{2}$ in each window

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

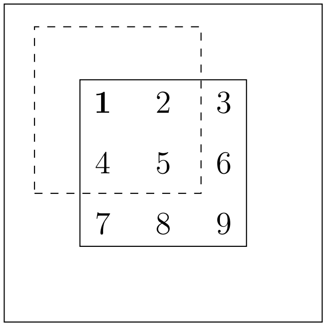
2D Convolution for Images

Windows move across and down

Nine 3×3 windows in 5×5 square

Center points are marked 1 to 9

Only $3^2 = 9$ weights to choose



A convolutional filter treats all positions the same

1. Many weights repeated—distant weights are zero
2. **$3^2 = 9$ weights copied in every window**
3. No reason to treat positions differently—**"shift invariant"**

Recognizing digits (like Zip codes) in MNIST : Basic test data

Max-pooling Reduce dimensions Take max from each block of outputs

Softmax Convert outputs $\boldsymbol{w_k}$ to probabilities $\boldsymbol{p_k} = \boldsymbol{e}^{\boldsymbol{w_k}} / \sum \boldsymbol{e}^{\boldsymbol{w_k}}$

Residual network Add skip connections that jump several layers

Batch normalization Reset the input variance at each new layer

10.6 Backpropagation : Chain Rule for ∇F

$F(x) =$ minimum $\quad \nabla F =$ partial derivatives $\quad$ **∂ errors / ∂ weights = zero**

Chain rule $\quad \dfrac{d}{dx}\Big(F_2(F_1(x))\Big) = \left(\dfrac{dF_2}{dF_1}(F_1(x))\right)\left(\dfrac{dF_1}{dx}(x)\right)$

Multivariable chain rule $\quad \dfrac{\partial \boldsymbol{w}}{\partial \boldsymbol{u}} = \left(\dfrac{\partial \boldsymbol{w}}{\partial \boldsymbol{v}}\right)\left(\dfrac{\partial \boldsymbol{v}}{\partial \boldsymbol{u}}\right)$ $\quad$ L layers in chain
Multiply L matrices

$$\underset{p \times n}{\frac{\partial \boldsymbol{w}}{\partial \boldsymbol{v}}} = \begin{bmatrix} \dfrac{\partial w_1}{\partial v_1} & \cdots & \dfrac{\partial w_1}{\partial v_n} \\ & \cdots & \\ \dfrac{\partial w_p}{\partial v_1} & \cdots & \dfrac{\partial w_p}{\partial v_n} \end{bmatrix} \qquad \underset{n \times m}{\frac{\partial \boldsymbol{v}}{\partial \boldsymbol{u}}} = \begin{bmatrix} \dfrac{\partial v_1}{\partial u_1} & \cdots & \dfrac{\partial v_1}{\partial u_m} \\ & \cdots & \\ \dfrac{\partial v_n}{\partial u_1} & \cdots & \dfrac{\partial v_n}{\partial u_m} \end{bmatrix}$$

At each layer
Derivatives before ReLU $\quad \boldsymbol{w} = A\boldsymbol{v} + \boldsymbol{b} \quad \dfrac{\partial w_i}{\partial b_j} = \delta_{ij} = 0$ or $1 \quad \dfrac{\partial w_i}{\partial A_{jk}} = \delta_{ij} v_k$

Product of matrices $\boldsymbol{ABC}$ $\qquad AB$ first or BC first ? $\quad$ Forward or back ?

For $\boldsymbol{\nabla F}$ in deep learning, going backward is faster : **Reverse mode $\boldsymbol{BC}$ first**

Example $\quad A = \boldsymbol{m} \times \boldsymbol{n} \quad B = \boldsymbol{n} \times \boldsymbol{p} \quad C = \boldsymbol{p} \times \mathbf{1}$ vector $\quad$ **Don't multiply $\boldsymbol{AB}$!**

Backpropagation = Automatic Differentiation = the key to speed

Part 11

Basic Statistics : Mean, Variance, Covariance

Part 11 : Basic Statistics : Mean, Variance, Covariance

11.1 Mean and Variance : Actual and Expected

The **sample mean** $\boldsymbol{\mu}$ is the average of outputs from N trials

The **expected mean** $\boldsymbol{m}$ is based on probabilities $p_1, \ldots, p_n$ of outputs $x_1, \ldots, x_n$

$$\boxed{\textbf{Expected value } \boldsymbol{m} = \mathbf{E}[\boldsymbol{x}] = \boldsymbol{p_1 x_1 + \cdots + p_n x_n}}$$

Law of Large Numbers : With probability 1, sample mean $\to m$ as $N \to \infty$

The **sample variance** measures the spread around the sample mean μ

$$\boldsymbol{S^2} = \frac{1}{N-1}\left[(x_1-\mu)^2 + \cdots + (x_N-\mu)\right]^2$$

The **variance** is the expected value of $(x-m)^2$ based on probabilities

$$\boxed{\boldsymbol{\sigma^2} = \mathrm{E}[(x-m)^2] = p_1(x_1-m)^2 + \cdots + p_n\,(x_n-m)^2}$$

Second formula for this important number : $\sigma^2 = \sum p_i\, x_i^2 - m^2$

Fair coin flip $\quad x=0$ or 1, $p_1=p_2=\frac{1}{2}$: Mean $\boldsymbol{m=\frac{1}{2}}$ $\quad$ Variance $\boldsymbol{\sigma^2=\frac{1}{2}-\frac{1}{4}=\frac{1}{4}}$

Continuous probability distributions : Sums change to integrals

$$\boxed{\boldsymbol{\int p(x)\,dx = 1 \quad m = \int x\,p(x)\,dx \quad \sigma^2 = \int (x-m)^2\,p(x)\,dx}}$$

11.2 Probability Distributions : Binomial, Poisson, Normal

1 **Binomial** : $\boldsymbol{p_{k,n}} =$ **probability of $\boldsymbol{k}$ heads in $\boldsymbol{n}$ trials** (coin flips)

$$p_{1,1} = p \qquad p_{n,n} = p^n \qquad p_{k,n} = \frac{n!}{k!\,(n-k)!}\, p^k (1-p)^{n-k} \qquad (0\,! = 1)$$

Mean m in n trials $= \boldsymbol{np}$ Variance σ^2 in n trials $= \boldsymbol{np\,(1-p)}$

2 **Poisson** : Rare events $\boldsymbol{p \to 0}$, many trials $\boldsymbol{n \to \infty}$ Keep $\boldsymbol{np = \lambda}$ constant

No successes $\boldsymbol{p_{0,n}} = (1-p)^n = \left(1 - \dfrac{\lambda}{n}\right)^n \to \boldsymbol{e^{-\lambda}}$ **$\boldsymbol{k}$ successes** $\boldsymbol{p_{k,n}} \to \dfrac{\boldsymbol{\lambda^k}}{\boldsymbol{k!}}\, \boldsymbol{e^{-\lambda}}$

Poisson mean $= \boldsymbol{\lambda}$ **variance $\boldsymbol{\sigma^2}$** $= \boldsymbol{\lambda}$ Limits of binomial np and $np\,(1-p)$

3 **Normal distribution** : $\mathbf{N}(m, \sigma^2)$ has $\boxed{\boldsymbol{p(x) = \dfrac{1}{\sqrt{2\pi}\,\sigma}\, e^{-(x-m)^2/2\sigma^2}}}$

Bell-shaped curve / Symmetric around mean / Standard $\mathbf{N}(0,1)$ is $\dfrac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$

Shifted and scaled $\boldsymbol{X = \dfrac{x-m}{\sigma}}$ Centered and normalized

Central Limit Theorem for any distribution $p(x)$ Average many samples

The probabilities for the average $\overline{X}$ of X_1 to X_M approaches $\mathbf{N}(0,1)$ as $M \to \infty$

Normal $\boldsymbol{p(x)}$ for $\boldsymbol{n}$ variables Means $\boldsymbol{m} = (m_1, \ldots, m_n)$ Covariance matrix $\boldsymbol{V}$

$$\boldsymbol{p}(\boldsymbol{x}) = \boldsymbol{p}(x_1, \ldots, x_n) = \frac{1}{\left(\boldsymbol{\sqrt{2\pi}}\right)^{\boldsymbol{n}} \boldsymbol{\sqrt{\det V}}}\, \boldsymbol{e}^{\boldsymbol{-(x-m)^{\mathrm{T}} V^{-1} (x-m)/2}}$$

11.3 Covariance Matrices and Joint Probabilities

M experiments at once $\qquad M = 2$ for (age x, height y)

Mean $\boldsymbol{m} = (m_x, m_y) =$ (average age, average height)

$\boxed{\textbf{Joint probabilities} \quad \boldsymbol{p_{ij}} = \text{probability that age} = \boldsymbol{i} \text{ and height} = \boldsymbol{j}}$

$$\boldsymbol{p_i} = \sum_j p_{ij} = \text{probability that } \textbf{age} = \boldsymbol{i} \text{ allowing all heights } \boldsymbol{j}$$

Expected value of $(x - m_x)^2 = \boldsymbol{\sigma_{11}^2} = \sum \boldsymbol{p_i}\,(x_i - m_x)^2 =$ usual variance

Expected value of $(x - m_x)(y - m_y) = \boldsymbol{\sigma_{12}} = \sum_i \sum_j \boldsymbol{p_{ij}}(x_i - m_x)(y_j - m_y)$

$$\textbf{Covariance matrix } \boldsymbol{V} = \sum_i \sum_j p_{ij} \begin{bmatrix} (x_i - m_x)^2 & (x_i - m_x)(y_j - m_y) \\ (x_i - m_x)(y_j - m_y) & (y_j - m_y)^2 \end{bmatrix}$$

$V =$ sum of positive semidefinite rank 1 matrices $=$ semidefinite or definite

V is positive definite unless age tells you the exact height (**dependent case**)

V is a diagonal matrix if age and height are **independent** : covariance $= 0$

Coin flip

Glue 2 coins together $\quad \boldsymbol{V} = \begin{bmatrix} 1/4 & 1/4 \\ 1/4 & 1/4 \end{bmatrix} =$ dependent case : semidefinite V

Separate the coins $\quad \boldsymbol{V} = \begin{bmatrix} 1/4 & 0 \\ 0 & 1/4 \end{bmatrix} =$ independent : diagonal V

11.4 Three Basic Inequalities of Statistics

Markov's inequality when $\boldsymbol{x \geq 0}$: No negative samples

$$\boxed{\text{The probability of } x \geq a \text{ is at most } \frac{\mathrm{E}[x]}{a} = \frac{\text{mean } m}{a}}$$

Suppose $\boldsymbol{a} = 3$ and mean $\boldsymbol{m} = \sum x_i p_i = 0p_0 + 1p_1 + 2p_2 + \cdots$

Markov's inequality says probability $\boldsymbol{p_3 + p_4 + p_5 +} \cdots \leq \dfrac{\boldsymbol{m}}{\boldsymbol{3}}$

Write $\boldsymbol{m} = p_1 + 2p_2 + \boldsymbol{3(p_3 + p_4 + p_5 + \cdots)} + p_4 + 2p_5 + \cdots$

No negative terms so $\boldsymbol{m} \geq \boldsymbol{3(p_3 + p_4 + p_5} + \cdots)$ THIS IS MARKOV

$$\boxed{\textbf{Chebyshev's inequality} \quad \text{The probability of } |x - m| \geq a \text{ is at most } \frac{\sigma^2}{a^2}}$$

Proof Apply Markov's inequality to the new variable $\boldsymbol{y = |x - m|^2}$

The mean value E$[\boldsymbol{y}]$ for $\boldsymbol{y}$ is the variance $\boldsymbol{\sigma^2}$ for $\boldsymbol{x}$

Apply Markov ! The probability of $y \geq a^2$ is at most $\dfrac{\mathrm{E}[y]}{a^2} = \dfrac{\sigma^2}{a^2}$

Chernoff's inequality $\boldsymbol{S = X_1} + \cdots + \boldsymbol{X_n}$ independent random variables

What is the probability that S is far from its mean $\overline{S}$?

$$\boxed{\begin{array}{l} \text{Prob}\left(\boldsymbol{S \geq (1+\delta)\overline{S}}\right) \leq e^{-\overline{S}\delta^2/(2+\delta)} \\ \text{Prob}\left(\boldsymbol{S \leq (1-\delta)\overline{S}}\right) \leq e^{-\overline{S}\delta^2/2} \end{array}}$$

Exponential dropoff !

Bound for $\boldsymbol{2\delta}$ = (Bound for $\boldsymbol{\delta}$)$^{\boldsymbol{4}}$

Reason : A large sum S usually needs several X_i to be large / unlikely !

11.5 Markov Matrices and Markov Chains

Markov matrix **All $M_{ij} \geq 0$** **All columns add to 1**

Perron-Frobenius
Eigenvalues of M $\boldsymbol{\lambda_{\max} = 1}$ $|\,\text{other }\lambda\,| < 1$ if all $M_{ij} > 0$
$|\,\text{other }\lambda\,| \leq 1$ if all $M_{ij} \geq 0$

Markov chain $\boldsymbol{p_{n+1} = Mp_n}$ Probabilities at times $n+1$ and n

$$\boldsymbol{M} = \begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix} \text{ has } \lambda = 1 \text{ and } \lambda = \frac{1}{2} \qquad M^n \text{ has } \lambda = 1 \text{ and } \lambda = \left(\frac{1}{2}\right)^n$$

$$\begin{bmatrix} \text{Rental cars in Chicago} \\ \text{Rental cars in Denver} \end{bmatrix}_{\boldsymbol{n+1}} = \boldsymbol{M} \begin{bmatrix} \text{in Chicago} \\ \text{in Denver} \end{bmatrix}_{\boldsymbol{n}} \qquad \boldsymbol{y}_{n+1} = \boldsymbol{M}\boldsymbol{y}_n$$

Start in Chicago $\boldsymbol{y}_0 = \begin{bmatrix} 100 \\ 0 \end{bmatrix}$ $\boldsymbol{y}_1 = \begin{bmatrix} 80 \\ 20 \end{bmatrix}$ $\boldsymbol{y}_2 = \begin{bmatrix} 70 \\ 30 \end{bmatrix}$ $\boldsymbol{y}_3 = \begin{bmatrix} 65 \\ 35 \end{bmatrix}$

Start in Denver $\boldsymbol{y}_0 = \begin{bmatrix} 0 \\ 100 \end{bmatrix}$ $\boldsymbol{y}_1 = \begin{bmatrix} 30 \\ 70 \end{bmatrix}$ $\boldsymbol{y}_2 = \begin{bmatrix} 45 \\ 55 \end{bmatrix}$ $\boldsymbol{y}_\infty = \begin{bmatrix} \mathbf{60} \\ \mathbf{40} \end{bmatrix}$

Steady state from every start : Eigenvector of M for $\lambda = 1$ is $\begin{bmatrix} 60 \\ 40 \end{bmatrix}$

Other eigenvalue $\lambda = \frac{1}{2}$: Distance to $\begin{bmatrix} 60 \\ 40 \end{bmatrix}$ is halved at every step