ZOOMNOTES FOR
LINEAR ALGEBRA

GILBERT STRANG
Massachusetts Institute of Technology

WELLESLEY - CAMBRIDGE PRESS
Box 812060 Wellesley MA 02482
ZoomNotes for Linear Algebra
Copyright ©2021 by Gilbert Strang
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

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.
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 $Ax = b$ 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 $F(x, v)$ from training data $v$ (like images of handwritten numbers) and matrix weights $x$. The piecewise linear “ReLU function” plays a mysterious but crucial part in $F$. Then $F(x, v_{\text{new}})$ can come close to new data that the system has never seen.

The learning function $F(x, v)$ 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

<table>
<thead>
<tr>
<th>Textbook/Resource</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction to Linear Algebra, 5th Ed. (2016)</td>
<td><a href="http://math.mit.edu/linearalgebra">math.mit.edu/linearalgebra</a></td>
</tr>
<tr>
<td>Linear Algebra and Learning from Data (2019)</td>
<td><a href="http://math.mit.edu/learningfromdata">math.mit.edu/learningfromdata</a></td>
</tr>
<tr>
<td>Linear Algebra for Everyone (2020)</td>
<td><a href="http://math.mit.edu/everyone">math.mit.edu/everyone</a></td>
</tr>
<tr>
<td>Differential Equations and Linear Algebra (2014)</td>
<td><a href="http://math.mit.edu/dela">math.mit.edu/dela</a></td>
</tr>
<tr>
<td>ZoomNotes for Linear Algebra (2021)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Video Lectures</th>
<th>OpenCourseWare</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math 18.06 and 18.06SC</td>
<td>Linear Algebra at MIT</td>
<td><a href="http://ocw.mit.edu/courses">ocw.mit.edu/courses</a></td>
</tr>
<tr>
<td>(added to 18.06)</td>
<td>A 2020 Vision of Linear Algebra</td>
<td><a href="http://youtube/mitocw">youtube/mitocw</a></td>
</tr>
<tr>
<td>Math 18.065</td>
<td>Linear Algebra and Learning from Data</td>
<td></td>
</tr>
<tr>
<td>Math 18.085 and 18.086</td>
<td>Computational Science and Engineering</td>
<td></td>
</tr>
<tr>
<td>Strang and Moler</td>
<td>Differential Equations and Linear Algebra</td>
<td></td>
</tr>
</tbody>
</table>

**Interview with Lex Fridman** | [https://www.youtube.com/watch?v=lEZPfmGCEk0](https://www.youtube.com/watch?v=lEZPfmGCEk0)

---

**Wellesley - Cambridge Press**  
Box 812060, Wellesley MA 02482 USA  
[www.wellesleycambridge.com](http://www.wellesleycambridge.com)  
Orders: [math.mit.edu/weborder.php](http://math.mit.edu/weborder.php)  
Outside US/Canada: [www.cambridge.org](http://www.cambridge.org)  
Select books, India: [www.wellesleypublishers.com](http://www.wellesleypublishers.com)

**Gilbert Strang’s page**: [math.mit.edu/~gs](http://math.mit.edu/~gs)
Three Great Factorizations: \(LU, QR, \text{SVD}\)

Orthogonal matrix
\[ Q^T Q = I \]
Square \( QQ^T = I \)

Triangular matrix
\[ R_{ij} = 0 \text{ for } i > j \]
\[ R_{jj} \neq 0 \text{ on diagonal} \]

Orthogonal basis
\[ Q = \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix} \]

Triangular basis
\[ R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{22} & \ddots & \ddots \\ \vdots & \ddots & \ddots \\ r_{nn} \end{bmatrix} \]

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

---

**Output Basis Diagram:**

- **Row Space of \(A\)**
  - Input Basis: \(U\) and \(R\) and \(V\)
  - Output Basis: \(L\) and \(Q\) and \(U\)

- **Column Space of \(A\)**
Part 1

Basic Ideas of Linear Algebra

1.1 Linear Combinations of Vectors
1.2 Dot Products $v \cdot w$ and Lengths $||v||$ and Angles $\theta$
1.3 Matrices Multiplying Vectors
1.4 Column Space and Row Space of $A$
1.5 Dependent and Independent Columns
1.6 Matrix-Matrix Multiplication $AB$
1.7 Factoring $A$ into $CR$: Column rank $= r =$ Row rank
1.8 Rank one matrices $A = (1 \text{ column}) \times (1 \text{ row})$
Part 1: Basic Ideas of Linear Algebra

1.1 Linear Combinations of Vectors

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

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

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

\[
\begin{bmatrix}
3 \\
4 \\
5
\end{bmatrix} +
\begin{bmatrix}
2 \\
0 \\
-2
\end{bmatrix} =
\begin{bmatrix}
5 \\
4 \\
3
\end{bmatrix} \\
\begin{bmatrix}
3 \\
4 \\
5
\end{bmatrix} =
\begin{bmatrix}
12 \\
16 \\
20
\end{bmatrix} \\
\begin{bmatrix}
2 \\
0 \\
-2
\end{bmatrix} =
\begin{bmatrix}
0 \\
0 \\
0
\end{bmatrix} = \text{zero vector}
\]

Linear combinations \( 2v - 3w \) and \( cv + dw \) and \( w - 2z + u \)

\[
\begin{bmatrix}
3 \\
4 \\
5
\end{bmatrix} - 3
\begin{bmatrix}
1 \\
2 \\
3
\end{bmatrix} =
\begin{bmatrix}
3 \\
2 \\
1
\end{bmatrix} \\
\begin{bmatrix}
1 \\
2 \\
3
\end{bmatrix} - 2
\begin{bmatrix}
4 \\
5 \\
6
\end{bmatrix} =
\begin{bmatrix}
1 \\
2 \\
3
\end{bmatrix} \\
\begin{bmatrix}
7 \\
8 \\
9
\end{bmatrix} =
\begin{bmatrix}
0 \\
0 \\
0
\end{bmatrix}
\]

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

All \( c \begin{bmatrix}
3 \\
4 \\
5
\end{bmatrix} + d \begin{bmatrix}
1 \\
2 \\
3
\end{bmatrix} \) fill a plane 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 \( \mathbb{R}^3 \)

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 \( \begin{bmatrix}
-3 \\
-4 \\
-5
\end{bmatrix} \)

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 $v \cdot w$ and Lengths $|v|$ and Angles $\theta$

Dot product $v \cdot w = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} = \frac{3 \times 2}{5 \times 1} + \frac{4 \times 0}{5 \times 1} + \frac{5 \times 1}{5 \times 1} = 11$

$v \cdot w = w \cdot v$

Length squared of $v = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}$ is $|v|^2 = 3^2 + 4^2 = 9 + 16$. This is Pythagoras $c^2 = a^2 + b^2$

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

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

$||v + w||^2 = ||v||^2 + ||w||^2 + 2v \cdot w$

Length squared of $v + w$ is $48$ is $50 + 2 + 2v \cdot w$

Length squared of $v + w$ is $4^2 + 4^2 + 4^2$

Triangle has edges $v, w, v - w$

The dot product $v \cdot w$ reveals the angle $\theta$ between $v$ and $w$

$|\cos \theta| \leq 1$ is one way to see the Schwarz inequality

The angle between $v = \begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix}$ and $w = \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}$ is $\theta = 90^\circ$ because $v \cdot w = 0$ : Perpendicular

The angle between $v = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ and $w = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ is $\theta = 45^\circ$ because $v \cdot w = 1$ and $|v||w| = \sqrt{2}$. 
1.3 Matrices Multiplying Vectors

There is a row way to multiply $Ax$ and also a column way to compute the vector $Ax$.

**Row way** = Dot product of vector $x$ with each row of $A$

\[
A \mathbf{x} = \begin{bmatrix} 2 & 5 \\ 3 & 7 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 2v_1 + 5v_2 \\ 3v_1 + 7v_2 \end{bmatrix} = \begin{bmatrix} 2 & 5 \\ 3 & 7 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 7 \\ 10 \end{bmatrix}
\]

**Column way** = $Ax$ is a combination of the columns of $A$

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

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 $Ax$ from the same multiply-adds. Just in a different order.

**$C(A)$ = Column space of $A$ = all combinations of the columns = all outputs $Ax$**

The identity matrix has $I \mathbf{x} = \mathbf{x}$ for every $\mathbf{x}$

\[
\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}
\]

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

If all columns are multiples of column 1 (not zero), the column space $C(A)$ 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 $A\mathbf{x}$ (for all $\mathbf{x}$) fill the column space $\mathbf{C}(A)$: line or plane or . . .

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

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

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 $= 2$D 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: $A\mathbf{x} = \mathbf{0}$ for some vector $\mathbf{x}$ (other than $\mathbf{x} = \mathbf{0}$)

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

Reasons: Column 2 of $A_1 = 2$ (Column 1) $A_2$ times $\mathbf{x} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ 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: $A\mathbf{x} = \mathbf{0}$ only when $\mathbf{x} = \mathbf{0}$

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

\[
\begin{bmatrix}
1 & 2 \\
3 & 4 \\
\end{bmatrix}
\begin{bmatrix}
5 & 7 \\
6 & 8 \\
\end{bmatrix}
= 
\begin{bmatrix}
17 \\
\cdot \\
\cdot \\
\end{bmatrix}
\]

because
\[
\begin{bmatrix}
1 & 2 \\
\end{bmatrix}
\begin{bmatrix}
5 \\
6 \\
\end{bmatrix}
= 17
\]

Dot product

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

\[
\begin{bmatrix}
1 & 2 \\
3 & 4 \\
\end{bmatrix}
\begin{bmatrix}
5 & 7 \\
6 & 8 \\
\end{bmatrix}
= 
\begin{bmatrix}
17 \\
39 \\
\cdot \\
\cdot \\
\end{bmatrix}
\]

because
\[
5
\begin{bmatrix}
1 \\
\end{bmatrix}
+ 6
\begin{bmatrix}
2 \\
\end{bmatrix}
= 
\begin{bmatrix}
17 \\
39 \\
\end{bmatrix}
\]

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 $B$

\[
\begin{bmatrix}
1 & 2 \\
3 & 4 \\
\end{bmatrix}
\begin{bmatrix}
5 & 7 \\
6 & 8 \\
\end{bmatrix}
= 
\begin{bmatrix}
17 & 23 \\
\cdot & \\
\cdot & \\
\end{bmatrix}
\]

because
\[
1
\begin{bmatrix}
5 \\
\end{bmatrix}
+ 2
\begin{bmatrix}
6 \\
8 \\
\end{bmatrix}
= 
\begin{bmatrix}
17 & 23 \\
\end{bmatrix}
\]

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

\[
\begin{bmatrix}
1 \\
3 \\
\end{bmatrix}
\begin{bmatrix}
5 & 7 \\
15 & 21 \\
\end{bmatrix}
\quad \text{and} \quad
\begin{bmatrix}
2 \\
4 \\
\end{bmatrix}
\begin{bmatrix}
6 & 8 \\
12 & 16 \\
\end{bmatrix}
= 
\begin{bmatrix}
17 & 23 \\
39 & 53 \\
\end{bmatrix}
\]

NOW ADD

\[
\begin{bmatrix}
17 & 23 \\
39 & 53 \\
\end{bmatrix}= AB
\]

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 | $A$ times $BC = AB$ times $C$ | 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 $C$

$A = \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = CR$

2 columns in $C$

2 rows in $R$

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**

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 \( A = (1 \text{ column}) \text{ times } (1 \text{ row}) \)

**Example**

\[
\begin{bmatrix}
2 & 4 & 6 \\
3 & 6 & 9
\end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} = CR
\]

Suppose all columns of \( A \) are multiples of one column. Then all rows of \( A \) 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 \( r \) matrix is the sum of \( r \) 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*

\[
A = CR \quad A = LU \quad A = QR \quad S = Q\Lambda Q^T \quad A = U\Sigma V^T
\]

Those 5 factorizations are described in *Parts 1+3, 2, 4, 6, 7* of these ZoomNotes
Part 2
Solving Linear Equations
\[ Ax = b : A \text{ is } n \text{ by } n \]

2.1 Inverse Matrices \( A^{-1} \) and Solutions \( x = A^{-1}b \)

2.2 Triangular Matrix and Back Substitution for \( Ux = c \)

2.3 Elimination: Square Matrix \( A \) to Triangular \( U \)

2.4 Row Exchanges for Nonzero Pivots: Permutation \( P \)

2.5 Elimination with No Row Exchanges: Why is \( A = LU \)?

2.6 Transposes / Symmetric Matrices / Dot Products
Part 2: Solving Linear Equations

\(Ax = b: A\) is \(n\) by \(n\)

2.1 Inverse Matrices \(A^{-1}\) and Solutions \(x = A^{-1}b\)

The inverse of a square matrix \(A\) has \(A^{-1}A = I\) and \(AA^{-1} = I\)

\[
2 \text{ by } 2 \quad A^{-1} = \begin{bmatrix} 2 & 1 \\ 5 & 4 \end{bmatrix}^{-1} = \frac{1}{7} \begin{bmatrix} 4 & -1 \\ -3 & 2 \end{bmatrix} \quad \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}
\]

\(A\) has no inverse if \(ad - bc = 0\)

\[
A = \begin{bmatrix} 1 & 2 \\ 4 & 8 \end{bmatrix}
\]

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

14
2.2 Triangular Matrix and Back Substitution for $Ux = c$

Solve $Ux = \begin{bmatrix} 2 & 3 & 4 \\ 0 & 5 & 6 \\ 0 & 0 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 19 \\ 17 \\ 14 \end{bmatrix} = c$ 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 $x_3 = 2$
- **Work upwards** The next equation $5x_2 + 6(2) = 17$ gives $x_2 = 1$
- **Upwards again** The first equation $2x_1 + 3(1) + 4(2) = 19$ gives $x_1 = 4$
- **Conclusion** The only solution to this example is $x = (4, 1, 2)$
- **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

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

Calculus: Inverse of derivative is integral

$$\int_0^x \frac{df}{dx} \, dx = f(x) - f(0)$$
2.3 Elimination: Square Matrix $A$ to Triangular $U$

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

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

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

To invert elimination, add

$\ell_{ij}$ times row $j$ back to row $i$

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

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

This $A$ needs 3 elimination steps to a beautiful result

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

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

For $Ax = b$ Add extra column $b$ Elimination and back substitution

$$\begin{bmatrix} A & b \end{bmatrix} = \begin{bmatrix} 2 & 3 & 4 & 19 \\ 4 & 11 & 14 & 55 \\ 2 & 8 & 17 & 50 \end{bmatrix} \rightarrow \begin{bmatrix} U & c \end{bmatrix} = \begin{bmatrix} 2 & 3 & 4 & 19 \\ 0 & 5 & 6 & 17 \\ 0 & 0 & 7 & 14 \end{bmatrix} \text{ backsob in 2.2 } \rightarrow x = \begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix}$$
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

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

Permutation matrix $P = \text{Rows of } I \text{ in any order}$

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

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

Can you find all six 3 by 3 permutations? Is every $P_1P_2 = P_2P_1$?

If $A$ is invertible then some $PA$ has no zero pivots and

$PA = LU$

Reverse the order by $P$

Circular shift by $P$
2.5 Elimination with No Row Exchanges: Why is \( A = LU \)?

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

\[
\begin{bmatrix}
1 \text{ (row 1)} \\
\ell_{21} \text{ (row 1)} \\
\ell_{31} \text{ (row 1)} \\
\ell_{41} \text{ (row 1)}
\end{bmatrix}
\]

from \( A \) to leave \( 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 \( \ell_1 = (1, \ell_{21}, \ell_{31}, \ell_{41}) \) times row 1 of \( A \)—the first pivot row \( u_1 \).

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

\[
\begin{bmatrix}
0 \text{ (row 2 of } A_2) \\
1 \text{ (row 2 of } A_2) \\
\ell_{32} \text{ (row 2 of } A_2) \\
\ell_{42} \text{ (row 2 of } A_2)
\end{bmatrix}
\]

from \( A_2 \) to leave \( 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 \( u_2 \) of \( U \). We removed a column \( \ell_2 = (0, 1, \ell_{32}, \ell_{42}) \) times \( u_2 \). Continuing this way, every step removes a column \( \ell_j \) times a pivot row \( u_j \) of \( U \). Now put those pieces back:

\[
A = \ell_1 u_1 + \ell_2 u_2 + \cdots + \ell_n u_n = \begin{bmatrix} \ell_1 \cdots \ell_n \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} = LU
\]

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 \( A \):

\[
L_{32} L_{31} L_{21} A = U \quad \text{and} \quad A = L_{21}^{-1} L_{31}^{-1} L_{32}^{-1} U = LU
\]
2.6 Transposes / Symmetric Matrices / Dot Products

Transpose of $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 4 \end{bmatrix}$ is $A^T = \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 3 & 4 \end{bmatrix}$

$(A^T)_{ij} = A_{ji}$

Rules for the sum and product

Transpose of $A + B$ is $A^T + B^T$

Transpose of $AB$ is $B^T A^T$

A symmetric matrix has $S^T = S$ This means that every $s_{ij} = s_{ji}$

The matrices $A^T A$ and $AA^T$ are symmetric / usually different

$\begin{bmatrix} 2 & 3 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 13 \\ 4 + 9 = 13 \end{bmatrix}$

$\begin{bmatrix} 2 & 3 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 6 & 9 \\ 6 & 6 \end{bmatrix}$

$S = LU$ is improved to symmetric $S = LDL^T$ (pivots in $U$ go into $D$)

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

Dot product

Work $= \text{Movements} \cdot \text{Forces} = x^T f$

Inner product

Heat $= \text{Voltage drops} \cdot \text{Currents} = e^T y$

$x \cdot y = x^T y$

Income $= \text{Quantities} \cdot \text{Prices} = q^T p$
Part 3
Vector Spaces and Subspaces
Basis and Dimension

3.1 Vector Spaces and Four Fundamental Subspaces

3.2 Basis and Dimension of a Vector Space $S$

3.3 Column Space and Row Space: Bases by Elimination

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

3.5 Four Fundamental Subspaces $C(A), C(A^T), N(A), N(A^T)$

3.6 Graphs, Incidence Matrices, and Kirchhoff’s Laws

3.7 Every Matrix $A$ Has a Pseudoinverse $A^+$
Part 3: Vector Spaces and Subspaces
Basis and Dimension

3.1 Vector Spaces and Four Fundamental Subspaces

**Vector space**: Linear combinations of vectors in \( S \) must stay in \( S \)

\[ S = \mathbb{R}^n \] or Subspace of \( \mathbb{R}^n \), \( S = \) matrices \( \mathbb{R}^{m \times n} \) or functions \( ax + b \)

Not vector spaces: Half-line \( x \geq 0 \), invertible matrices, singular matrices

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

<table>
<thead>
<tr>
<th>Four Subspaces</th>
<th>Column space ( C(A) ) = all vectors ( Ax )</th>
<th>Row space ( C(A^T) ) = all vectors ( A^Ty )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Column space = “range”</td>
<td>Nullspace ( N(A) ) = all ( x ) with ( Ax = 0 )</td>
<td>Left nullspace ( N(A^T) ) = all ( y ) with ( A^Ty = 0 )</td>
</tr>
</tbody>
</table>

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 \( A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \) is the \( x-y \) plane in \( \mathbb{R}^3 \)

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 \( \mathbb{R}^3 \): \( \mathbf{N}(A) = \text{all} \begin{bmatrix} 0 \\ 0 \\ z \end{bmatrix} \)

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 1. Notice \( 2 + 1 = \) dimension of \( \mathbb{R}^3 \)

**Matrix spaces**  The vector space of 3 by 3 matrices has dimension 9

The subspace of upper triangular matrices has dimension 6

If \( v_1, \ldots, v_m \) and \( w_1, \ldots, w_n \) are bases then \( m \) must equal \( n \)

Proof \( W = VA \begin{bmatrix} w_1 & \cdots & w_n \end{bmatrix} = \begin{bmatrix} v_1 & \cdots & v_m \end{bmatrix} \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \)

If \( m < n \) then \( Ax = 0 \) would have a nonzero solution \( x \)

Then \( Wx = VAx = 0 \). \( W \) would have dependent columns!
3.3 Column Space and Row Space: Bases by Elimination

Every pivot = 1  
Eliminate below  
Eliminate above  

\[
\begin{bmatrix}
1 & 2 & 4 \\
4 & 9 & 0 \\
3 & 7 & 0 \\
4 & 9 & 0 \\
\end{bmatrix}
\rightarrow
\begin{bmatrix}
1 & 2 & 4 \\
0 & 1 & 2 \\
0 & 1 & 2 \\
0 & 0 & 1 \\
\end{bmatrix}
\rightarrow R_0 = I
\]

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

\(r = \text{rank } \ A\) has \(r\) independent columns and \(r\) independent rows

\[
A = \begin{bmatrix}
2 & 4 \\
3 & 7 \\
4 & 9 \\
\end{bmatrix}
\rightarrow
\begin{bmatrix}
1 & 2 \\
3 & 7 \\
4 & 9 \\
\end{bmatrix}
\rightarrow
\begin{bmatrix}
1 & 2 \\
0 & 1 \\
0 & 1 \\
\end{bmatrix}
\rightarrow
\begin{bmatrix}
1 & 0 \\
0 & 0 \\
\end{bmatrix}
= \begin{bmatrix} I \\
0 \end{bmatrix} = R_0
\]

\(I\) locates \(r\) independent columns  
\(P\) permutes the columns if needed  

\[
R_0 = \begin{bmatrix} I & F \\
0 & 0 \end{bmatrix}
\]

Basis for the column space  
\(C = \text{First \(r\) independent columns of } A\)

Basis for the row space  
\(R = \begin{bmatrix} I & F \end{bmatrix}\)  
\(P = \text{\(r\) rows and } A = CR\)

\[
A = \begin{bmatrix}
1 & 1 & 2 \\
1 & 2 & 3 \\
\end{bmatrix}
\rightarrow
\begin{bmatrix}
1 & 1 & 2 \\
0 & 1 & 1 \\
\end{bmatrix}
\rightarrow
\begin{bmatrix}
1 & 0 & 1 \\
0 & 1 & 1 \\
\end{bmatrix}
= \begin{bmatrix} I & F \end{bmatrix} = R_0 = R
\]

\(A = CR\) is  
\[
\begin{bmatrix}
1 & 1 & 2 \\
1 & 2 & 3 \\
\end{bmatrix}
= \begin{bmatrix} 1 & 1 \\
1 & 2 \end{bmatrix}\begin{bmatrix} 1 & 0 & 1 \\
0 & 1 & 1 \end{bmatrix}
\]

Row rank = 2  
Column rank = 2
3.4 \( Ax = 0 \) and \( Ax = b \): \( x_{\text{nullspace}} \) and \( x_{\text{particular}} \)

\[
A = \begin{bmatrix} 1 & 2 & 1 & 4 \\ 2 & 4 & 3 & 9 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 1 & 4 \\ 0 & 0 & 1 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 0 & 3 \\ 0 & 0 & 1 & 1 \end{bmatrix} = R
\]

2 columns of \( I \)

4 – 2 columns of \( F \)

Basis for nullspace

Special solutions

\( s_1 = \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad s_2 = \begin{bmatrix} -3 \\ 0 \\ -1 \\ 1 \end{bmatrix} \)

2, 3, 1 in \( R \)

1 and 0

\( As_1 = 0 \) and \( As_2 = 0 \)

\( \left[ \begin{array}{c} -2 \\ 1 \\ 0 \\ 0 \end{array} \right] \)

\( \left[ \begin{array}{c} -3 \\ 0 \\ -1 \\ 1 \end{array} \right] \)

2, 3, 1 in \( R \)

1 and 0

“Special” Dependent columns 2 and 4 = combination of independent columns 1 and 3

Elimination from \( A \) to \( R \) reveals the \( n – r \) special solutions

\( Ax = 0 \) and \( Rx = 0 \)

\( Rx = \begin{bmatrix} I & F \end{bmatrix} P x = 0 \)

\( S = [s_1 \cdots s_{n-r}] = \begin{bmatrix} -F \\ I_{n-r} \end{bmatrix} \)

\( PP^T = I \) leads to \( RS = 0 \)

\( r \) equations \quad \( n – r \) solutions

Complete solution to \( Ax = b \)

\( x = x_{\text{nullspace}} + x_{\text{particular}} = \) above \( \uparrow + \) below \( \downarrow \)

\[
\begin{bmatrix} A & b \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 & 4 & b_1 \\ 2 & 4 & 3 & 9 & b_2 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 0 & 3 & d_1 \\ 0 & 0 & 1 & 1 & d_2 \end{bmatrix} = \begin{bmatrix} R & d \end{bmatrix} \quad x_{\text{particular}} = \begin{bmatrix} d_1 \\ 0 \\ d_2 \\ 0 \end{bmatrix}
\]

\[
\begin{bmatrix} A_{\text{new}} & b_{\text{new}} \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 & 4 & b_1 \\ 2 & 4 & 3 & 9 & b_2 \\ 3 & 6 & 4 & 13 & b_3 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 0 & 3 & d_1 \\ 0 & 0 & 1 & 1 & d_2 \\ 0 & 0 & 0 & 0 & d_3 \end{bmatrix}
\]

No solution if \( d_3 \neq 0 \)

No solution if \( b_1 + b_2 \neq b_3 \)

Elimination must give \( 0 = 0 \)
3.5 Four Fundamental Subspaces \( C(A), C(A^T), N(A), N(A^T) \)

The big picture

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 \( Ax = 0 \)?

\( m \) equations, \( n \) unknowns, rank \( r \) \( \Rightarrow \) \( Ax = 0 \) 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

This graph has 5 edges and 4 nodes. \( A \) is its 5 by 4 incidence matrix.

\( b = b_1 \) to \( b_5 = \) currents. \( x = x_1 \) to \( x_4 = \) voltages

| Edges 1, 2, 3 form a loop in the graph | Dependent rows 1, 2, 3 |
| Edges 1, 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 \( m - n + 1 \):

- **\( \text{N}(A) \)** The constant vectors \((c, c, \ldots, c)\) make up the nullspace of \( A \) : \( \dim = 1 \).
- **\( \text{C}(A^T) \)** The edges of any spanning tree give \( r \) independent rows of \( A \) : \( r = n - 1 \).
- **\( \text{C}(A) \)** **Voltage Law**: The components of \( Ax \) add to zero around all loops: \( \dim = n - 1 \).
- **\( \text{N}(A^T) \)** **Current Law**: \( A^T y = (\text{flow in}) - (\text{flow out}) = 0 \) is solved by loop currents.

There are \( m - r = m - n + 1 \) independent small loops in the graph.

Current law \( A^T y = 0 \) at each node is fundamental to applied mathematics.
3.7 Every $A$ Has a Pseudoinverse $A^+$ 

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

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

$A$ has a right inverse $A^+ = A^T (AA^T)^{-1}$ 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

$AA^+ b$: project $b$ onto column space

$A^+ Ax$: project $x$ onto row space

$A^+ e = 0$

$A^+ = V \Sigma^+ U^T$ is computed from $A = U \Sigma V^T$: $\Sigma^+$ has $1/\sigma$’s and 0’s
Example \( A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = uv^T = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} = CR \)

The pseudoinverse \( A^+ \) has the same rank \( r = 1 \)

Row space of \( A = \text{line in } \mathbb{R}^3 \) \hspace{1cm} Column space of \( A = \text{line in } \mathbb{R}^2 \)

Reverse the column space and row space: \( v \) and \( u \)

\[
A^+ = \frac{vu^T}{\|v\|^2 \|u\|^2} = \frac{1}{6} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}
\]

Check \( A^+A = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \text{projection onto row space of } A \)

\( AA^+ = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \text{projection onto column space of } A \)
Part 4
Orthogonal Matrices and Least Squares

4.1 Orthogonality of the Four Subspaces

4.2 Projections onto Subspaces

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

4.4 Orthogonal Matrices and Gram-Schmidt
Part 4: Orthogonal Matrices and Least Squares

4.1 Orthogonality of the Four Subspaces

Vectors \( x \) and \( y \) are orthogonal if \( x^T y = 0 \) (Complex vectors: \( \bar{x}^T y = 0 \))

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

Orthogonal subspaces: Every \( v \) in \( V \) is orthogonal to every \( w \) 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 \( C(A) \) and nullspace \( N(A^T) \): Also orthogonal

Clear from \( Ax = 0 \)

\[
Ax = \begin{bmatrix}
\text{row 1} \\
\text{row m}
\end{bmatrix}
\begin{bmatrix}
x
\end{bmatrix} = \begin{bmatrix}
0 \\
0
\end{bmatrix}
\]

All rows are orthogonal to \( x \) \( \Rightarrow \) whole row space is orthogonal to \( x \)

Big Picture of Linear Algebra: Two pairs of orthogonal subspaces

More: \( r + (n - r) = \) full dimension \( n \) so every \( x \) equals \( x_{\text{row}} + x_{\text{null}} \)
4.2 Projections onto Subspaces

\[ \hat{x} = a + \frac{a^T b}{a^T a} \]

\[ p = \hat{x} a = A(\hat{x}) = A \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \]

\[ e = b - p \]

\[ p = A\hat{x} = A \left( A^T A \right)^{-1} A^T b = Pb \]

\[ b \] is projected onto line through \( a \) and onto column space of \( A \)

Error vector \( e = b - p \) is **orthogonal** to the line and subspace

**Projection matrix \( P \)**

\[
\begin{align*}
P_{\text{line}} &= \frac{a a^T}{a^T a} \\
P_{\text{subspace}} &= A \left( A^T A \right)^{-1} A^T
\end{align*}
\]

Notice \( P^2 = P \) and \( P^T = P \)  

Second projection: \( p \) doesn’t move!

Project \( b = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} \)  

Line: \( a = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \) and \( A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \)

\[ P_{\text{line}} = a \frac{a^T b}{a^T a} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \frac{6}{3} = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix} \]

\[ \text{Error } e = b - p = \begin{bmatrix} 4 \\ -2 \\ -2 \end{bmatrix} \]

Plane: Column spaces of \( a = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \) \[ A^T \hat{x} = A^T b \]

\[ \begin{bmatrix} 3 & 3 & 3 \\ 3 & 5 & 5 \end{bmatrix} \hat{x} = \begin{bmatrix} 6 \\ 0 \end{bmatrix} \]

gives \( \hat{x} = \begin{bmatrix} 5 \\ -3 \end{bmatrix} \)

Projection \( p = A\hat{x} = \begin{bmatrix} 5 \\ -2 \\ -1 \end{bmatrix} \)

\[ \text{Error } e = b - p = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \]

What is the 3 by 3 projection matrix \( P = A(A^T A)^{-1} A^T \)? Then \( p = Pb \)
4.3 Least Squares Approximations (Regression): $A^T A \hat{x} = A^T b$

If $Ax = b$ has no solution then minimize $E = \|b - Ax\|^2 = x^T A^T Ax - 2x^T A^T b + b^T b$

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

**Linear algebra**  $Ax$ is in the column space of $A$  Best $A\hat{x}$ = projection of $b$ on $\text{C}(A)$

“Normal equations”  $A^T A \hat{x} = A^T b$ and then the projection is $p = A\hat{x}$

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

$m > 2$  $m$ equations  $2$ unknowns  No solution

\[
A = \begin{bmatrix}
1 & t_1 \\
. & . \\
. & . \\
1 & t_m
\end{bmatrix},
\quad
x = \begin{bmatrix}
C \\
D
\end{bmatrix},
\quad
b = \begin{bmatrix}
b_1 \\
. \\
b_m
\end{bmatrix}
\]

\[
A^T A \hat{x} = A^T b \Rightarrow \begin{bmatrix}
m \\
\sum t_i \\
\sum t_i^2
\end{bmatrix}\begin{bmatrix}
C \\
D
\end{bmatrix} = \begin{bmatrix}
3 & 3 \\
3 & 5 \\
-3 & -3
\end{bmatrix}\begin{bmatrix}
5 \\
0
\end{bmatrix} = \begin{bmatrix}
6 \\
-3
\end{bmatrix} = \begin{bmatrix}
\sum b_i \\
\sum t_i b_i
\end{bmatrix}
\]

errors $e = \text{vertical distances } b - p = (1, -2, 1)$ to best line
4.4 Orthogonal Matrices and Gram-Schmidt

Orthogonal columns

\[ Q = \begin{bmatrix} q_1 & \cdots & q_n \end{bmatrix} \]

\[ q_i^T q_i = 1 \quad \text{unit vectors} \]

\[ q_i^T q_j = 0 \quad \text{orthogonal} \]

\[ Q^T Q = I \]

Important case = Square matrix

Then \( QQ^T = I \quad Q^T = Q^{-1} \) “Orthogonal matrix”

\[ Q = \frac{1}{3} \begin{bmatrix} 2 & -1 & 2 \\ 2 & 2 & -1 \\ -1 & 2 & 2 \end{bmatrix} \]

has \( Q^T Q = I \)

\( QQ^T \neq I \)

Orthogonal matrix

\[ Q_1 \text{ times } Q_2 \text{ is orthogonal because } (Q_1 Q_2)^{-1} = Q_2^{-1} Q_1^{-1} = Q_2^T Q_1^T = (Q_1 Q_2)^T \]

\[ v = c_1 q_1 + \cdots + c_n q_n \] leads to \[ c_k = q_k^T v \]

\[ v = Q c \] leads to \[ c = Q^T v \]

Gram-Schmidt

Start with independent \( a, b, c \)

Create orthogonal vectors \( q_1, q_2, q_3 \)

\[ q_1 = \frac{a}{||a||} \quad Q_2 = b - (q_1^T b)q_1 \quad q_2 = \frac{Q_2}{||Q_2||} \quad Q_3 = c - (q_1^T c)q_1 - (q_2^T c)q_2 \quad q_3 = \frac{Q_3}{||Q_3||} \]

Gram-Schmidt \( A = QR \)

\( A = (\text{orthogonal})(\text{triangular}) \)

\[ A = \begin{bmatrix} \cos \theta & a_{12} \\ \sin \theta & a_{22} \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} 1 & r_{12} \\ 0 & r_{22} \end{bmatrix} = QR \]
Part 5

Determinant of a Square Matrix

5.1 3 by 3 and $n$ by $n$ Determinants

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

5.3 $\det AB = (\det A)(\det B)$ and Cramer’s Rule

5.4 Volume of Box = $|\text{Determinant of Edge Matrix } E|$
Part 5: Determinant of a Square Matrix

5.1 3 by 3 and n by n Determinants

\[
\begin{pmatrix}
1 & 1 & 1 \\
-1 & 1 & 1 \\
1 & +1 & -1 \\
\end{pmatrix}
\]

Even permutations have \( \det P = +1 \) Odd permutations have \( \det P = -1 \)

Three defining properties 1 Row exchange reverses sign of \( \det \) 2 \( \det \) is linear in each row separately 3 \( \det I = 1 \)

\[
\begin{pmatrix}
a & b & c \\
p & q & r \\
x & y & z \\
\end{pmatrix}
\]

Linearity separates \( \det A \) into \( n! = 3! = 6 \) simple determinants

\[
\det = \left| \begin{array}{ccc}
a & b & c \\
p & q & r \\
x & y & z \\
\end{array} \right|
\]

Combine 6 simple determinants into \( \det A \)

\[+ aqz + brx + cpy - ary - bpz - cqx\]

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

\[
\det A = \sum (\det P) a_{1j} a_{2k} \ldots a_{nz} \text{ as in } +a_{11}a_{22} - a_{12}a_{21}
\]
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 \[ \det A = a(qz - ry) + b(rx - pz) + c(py - qx) \]

$n$ factors $a, b, c$  \( n \) cofactors = determinants of size $n - 1$

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

Cofactors along row 1 \[ \det A = a_{11}C_{11} + \cdots + a_{1n}C_{1n} \]

Inverse formula \[ A^{-1} = (\text{transpose of } C) / (\text{determinant of } A) \]

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

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

$n = 3$ \[ A = \begin{bmatrix} a & b & c \\ p & q & r \\ x & y & z \end{bmatrix} \quad C = \begin{bmatrix} qz - ry & rx - pz & py - qx \\ bq - cy & az - cx & qx - py \\ br - cq & cp - ar & az - bx \end{bmatrix} \]

$AC^T = \begin{bmatrix} \det A & 0 & 0 \\ 0 & \det A & 0 \\ 0 & 0 & \det A \end{bmatrix} = (\det A)I$  This explains $A^{-1} = \frac{C^T}{\det A}$
5.3 Det $AB = (\det A) (\det B)$ and Cramer’s Rule

- $\det A = \det A^T$
- $\det AB = (\det A) (\det B)$
- $\det A^{-1} = \frac{1}{\det A}$

Orthogonal matrix \( \det Q = \pm 1 \) because \( Q^T Q = I \) gives \( (\det Q)^2 = 1 \)

Triangular matrix \( \det U = u_{11}u_{22} \cdots u_{nn} \)

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

Cramer’s Rule to Solve $Ax = b$

Start from

\[
\begin{bmatrix}
A \\
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3 \\
\end{bmatrix}
= \begin{bmatrix}
b_1 & a_{12} & a_{13} \\
b_2 & a_{22} & a_{23} \\
b_3 & a_{32} & a_{33} \\
\end{bmatrix} = B_1
\]

Use \( (\det A) (x_1) = (\det B_1) \) to find \( x_1 \)

Same idea

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

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

\[
\begin{bmatrix}
3 & 2 \\
5 & 4 \\
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
\end{bmatrix}
= \begin{bmatrix}
12 \\
22 \\
\end{bmatrix}
\]

\[
B_1 = \begin{bmatrix}
12 & 2 \\
22 & 4 \\
\end{bmatrix}
\]

\[
B_2 = \begin{bmatrix}
3 & 12 \\
5 & 22 \\
\end{bmatrix}
\]

\[
x_1 = \frac{\det B_1}{\det A} = \frac{4}{2} x_2 = \frac{2}{2}
\]

\[
\begin{bmatrix}
x_1 \\
x_2 \\
\end{bmatrix}
= \begin{bmatrix}
12 \\
22 \\
\end{bmatrix}
\]

\[
B_1 = \begin{bmatrix}
12 & 2 \\
22 & 4 \\
\end{bmatrix}
\]

\[
B_2 = \begin{bmatrix}
3 & 12 \\
5 & 22 \\
\end{bmatrix}
\]

\[
x_1 = \frac{\det B_1}{\det A} = \frac{4}{2} x_2 = \frac{2}{2}
\]
5.4 Volume of Box = | Determinant of Edge Matrix $E$ |

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

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

2D area Gram-Schmidt in 4.4 gives $E = QR = (\text{orthogonal}) (\text{triangular})$

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:
\[ Ax = \lambda x \text{ and } A^nx = \lambda^n x \]

6.1 Eigenvalues \( \lambda \) and Eigenvectors \( x \): \( Ax = \lambda x \)

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

6.3 Symmetric Positive Definite Matrices: Five Tests

6.4 Linear Differential Equations \( \frac{du}{dt} = Au \)

6.5 Matrices in Engineering: Second differences
Part 6: Eigenvalues and Eigenvectors: 
\[ Ax = \lambda x \text{ and } A^n x = \lambda^n x \]

6.1 Eigenvalues \( \lambda \) and Eigenvectors \( x \): \( Ax = \lambda x \)

\( Ax \) is on the same line as \( x \) / \( Ax = \lambda x \) means \((A - \lambda I)x = 0\)

Then \( A^2x = A(\lambda x) = \lambda(Ax) = \lambda^2x \) \[ A^n x = \lambda^n x \]

\[ A^{-1}x = \frac{1}{\lambda} x \]

**Determinant of** \( A - \lambda I = 0 \)  
Solutions \( \lambda_1 \) to \( \lambda_n \): \( A \) has \( n \) eigenvalues

\( A = \begin{bmatrix} \cdot8 & \cdot3 \\ \cdot2 & \cdot7 \end{bmatrix} \)

\( A - \lambda I = \begin{bmatrix} \cdot8 - \lambda & \cdot3 \\ \cdot2 & \cdot7 - \lambda \end{bmatrix} \)

\[ \det(A - \lambda I) = \lambda^2 - 1.5\lambda + .56 - .06 = (\lambda - 1) \left(\lambda - \frac{1}{2}\right) \]

**Eigenvector** \( x_2 \)  
for \( \lambda_2 = \frac{1}{2} \)

\( (A - \frac{1}{2}I)x_2 = \begin{bmatrix} \cdot3 & \cdot3 \\ \cdot2 & \cdot2 \end{bmatrix} \begin{bmatrix} x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \) gives \( x_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \)

**Eigenvector** \( x_1 \)  
for \( \lambda_1 = 1 \)

\( (A - I)x_1 = \begin{bmatrix} -\cdot2 & \cdot3 \\ \cdot2 & -\cdot3 \end{bmatrix} \begin{bmatrix} x_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \) gives \( x_1 = \begin{bmatrix} \cdot6 \\ \cdot4 \end{bmatrix} \)

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

\[ \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \cdot6 & \cdot4 \\ \cdot4 & -\cdot4 \end{bmatrix} + \begin{bmatrix} \cdot4 & -\cdot4 \\ \cdot4 & \cdot4 \end{bmatrix} \]

\[ A^{10} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 1^{10} \begin{bmatrix} \cdot6 & \cdot4 \\ \cdot4 & \cdot4 \end{bmatrix} + \left(\frac{1}{2}\right)^{10} \begin{bmatrix} \cdot4 & -\cdot4 \\ \cdot4 & \cdot4 \end{bmatrix} \]

**Useful Facts**  
sum of \( \lambda \)'s = \( \lambda_1 + \cdots + \lambda_n \) = trace of \( A = a_{11} + a_{22} + \cdots + a_{nn} \)

Product of \( \lambda \)'s = \( \lambda_1 \cdots (\lambda_n) \) = determinant of \( A \)

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

40
6.2 Diagonalizing a Matrix: \( X^{-1}AX = \Lambda = \text{eigenvalues} \)

Key idea / Follow each eigenvector separately / \( n \) simple problems

**Eigenvector matrix** \( X \)

- Assume independent \( x \)'s
- Then \( X \) is invertible

\[
\begin{align*}
AX &= X\Lambda \\
X^{-1}AX &= \Lambda \\
A &= X\Lambda X^{-1}
\end{align*}
\]

- \( A^k \) becomes easy
- \( A^k = (X\Lambda X^{-1})(X\Lambda X^{-1})\cdots(X\Lambda X^{-1}) \)
- Same eigenvectors in \( X \)

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

2 Question: **When does** \( A^k \rightarrow \text{zero matrix} \)?

Answer: **All** \( |\lambda_i| < 1 \)

3 Some matrices are not diagonalizable

- They don’t have \( n \) independent vectors

\[
A = \begin{bmatrix} 3 & 6 \\ 0 & 3 \end{bmatrix} \text{ has } \lambda = 3 \text{ and } 3
\]

- That \( A \) has double eigenvalue, single eigenvector

\[
\text{Only one } x = \begin{bmatrix} 1 \\ 0 \end{bmatrix}
\]

4 All the “similar matrices” \( BAB^{-1} \) have the same eigenvalues as \( A \)

If \( Ax = \lambda x \) then \( (BAB^{-1})(Bx) = BAx = B\lambda x = \lambda (Bx) \)
6.3 Symmetric Positive Definite Matrices: Five Tests

If $S = S^T$  

- **Eigenvalues** $\lambda$ are real  
- **Eigenvectors** $x$ are orthogonal

$$S = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix} = S^T$$ has $S \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 9 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $S \begin{bmatrix} -1 \\ -1 \end{bmatrix} = -9 \begin{bmatrix} -1 \\ -1 \end{bmatrix}$  

Notice $x_1 \cdot x_2 = 0$  

$$q = \frac{x}{||x||} = \frac{\text{eigenvectors}}{\text{length} = 1}$$  

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

$$S = QAQ^{-1} = QAQ^T$$  

Spectral theorem

$$S = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 9 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \frac{1}{\sqrt{2}} = QAQ^T$$

**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** $x^T S x$ is positive for every vector $x \neq 0$
- **Test 3** The **pivots** in elimination on $S$ are all positive
- **Test 4** The upper left **determinants** of $S$ are all positive
- **Test 5** $S = A^T A$ for some matrix $A$ with independent columns

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

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

$$\begin{bmatrix} 2 & b \\ b & 4 \end{bmatrix}$$  

Positive definite if $b^2 < 8$  

Semidefinite if $b^2 \leq 8$

**Second difference matrix**  

Positive definite in 6.5  

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

Positive semidefinite  

$Sx = 0$ for $x = (1, 1, 1)$  

$$S = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}$$
6.4 Linear Differential Equations $\frac{du}{dt} = Au$

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

$n \geq 1$ \quad $\frac{du}{dt} = Au$ is solved by eigenvectors as in $u(t) = c_1e^{\lambda_1 t}x_1$

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

Check: If $u = e^{\lambda t}x$ then $\frac{du}{dt} = \lambda e^{\lambda t}x = Ae^{\lambda t}x = Au$ as required

$A = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$ has $\lambda_1 = 9$ \quad $u_1 = e^{9t} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ \quad $\frac{du_1}{dt} = 9e^{9t} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = e^{9t}A \begin{bmatrix} 1 \\ 1 \end{bmatrix} = Au_1$

Initial condition $u(0)$ at $t = 0$ \quad Split $u(0)$ into eigenvectors $x$ \quad $u(0) = c_1x_1 + \cdots + c_nx_n$

Each eigenvector goes its own way \quad Combine solutions \quad $u(t) = c_1e^{\lambda_1 t}x_1 + \cdots + c_ne^{\lambda_n t}x_n$

Special case $\lambda_1 = \lambda_2$ with one eigenvector $x$ \quad $u(t) = c_1e^{\lambda_1 t}x_1 + c_2te^{\lambda_1 t}x_1$

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

Weak stability \quad $u(t) \to$ steady state if one $\lambda$ moves up to $\lambda = 0$

Matrix $e^{At}$ is the solution matrix \quad $e^{At}u(0) = u(t)$ \quad $e^{At}x = e^{\lambda t}x$

Exponential Series $e^{At}$ \quad $e^x = 1 + x + x^2/2 + \cdots + x^n/n! + \cdots$ \quad $e^A = I + A + \frac{A^2}{2} + \cdots + \frac{A^n}{n!} + \cdots$
6.5 Matrices in Engineering: Second differences

Centered difference\[ \frac{du}{dx} \approx \frac{u(x+h)-u(x-h)}{2h} \]

Second differences\[ \frac{d^2 u}{dx^2} \approx \frac{u(x+h)-2u(x)+u(x-h)}{h^2} \]

Second differences with \( u_0 = u_4 = 0 \)

Eigenvalues \( 2 - \sqrt{2}, 2, 2 + \sqrt{2} \)

Pivots \( \frac{2}{1}, \frac{3}{2}, \frac{4}{3} \)

Determinants 2, 3, 4

Energy\[ x^T K x \]

\[ K = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \]

\[ \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \frac{1}{h^2} K u \]

\[ 2(x_1^2 - x_1 x_2 + x_2^2 - x_2 x_3 + x_3^2) > 0 \]

\( K \) is positive definite

Steady state

Oscillation

Spring forces

Balance gravity

\( K u = g \)

Displacements \( u \)

\( u_0 = 0 \)

pulling up

pulling down

Masses move

Use eigenvalues of \( K \)

Newton’s Law \( F = ma \)

\( \frac{d^2 u}{dt^2} \)
Part 7

Singular Values and Vectors:
\[ Av = \sigma u \quad \text{and} \quad A = U \Sigma V^T \]

7.1 Singular Vectors in \( U \), \( V \) and Singular Values in \( \Sigma \)

7.2 Reduced SVD / Full SVD / Construct \( U \Sigma V^T \) from \( A^T A \)

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

7.4 \( A_k \) is Closest to \( A \): Principal Component Analysis PCA

7.5 Computing Eigenvalues of \( S \) and Singular Values of \( A \)

7.6 Compressing Images by the SVD

7.7 The Victory of Orthogonality
Part 7: **Singular Values and Vectors:**
\[ Av = \sigma u \text{ and } A = U \Sigma V^T \]

### 7.1 Singular Vectors in \( U, V \) and Singular Values in \( \Sigma \)

An example shows orthogonal vectors \( v \) going into orthogonal vectors \( u \)

\[ Av_1 = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 9 \end{bmatrix} \text{ and } Av_2 = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -3 \\ 1 \end{bmatrix} \]

\[ v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \text{ is orthogonal to } v_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \begin{bmatrix} 3 \\ 9 \end{bmatrix} \text{ is orthogonal to } \begin{bmatrix} -3 \\ 1 \end{bmatrix} \]

Divide both inputs \( v \) by \( \sqrt{2} \)

Divide both outputs \( u \) by \( \sqrt{10} \)

Matrix form
\[ AV = U \Sigma \]

\[ \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \begin{bmatrix} 3\sqrt{5} & 0 \\ 0 & \sqrt{5} \end{bmatrix} \]

\( V \) and \( U \) = orthogonal matrices

\[ V^T V = I \quad U^T U = I \quad A = U \Sigma V^T \]

\( v_1, v_2 = \) orthogonal basis for the **row space** = **inputs**

\( u_1, u_2 = \) orthogonal basis for the **column space** = **outputs**

\( \sigma_1 = 3\sqrt{5} \) and \( \sigma_2 = \sqrt{5} \) are the **singular values** of this \( A \)
7.2 Reduced SVD / Full SVD / Construct $U \Sigma V^T$ from $A^T A$

Reduced SVD : Stop at $u_r$ and $v_r$  \hspace{1cm} \text{Full SVD} : Go on to $u_m$ and $v_n$

\[ A = U_r \Sigma_r V_r^T = \begin{bmatrix} u_1 \text{ to } u_r \\ \text{column space} \\ m \times r \end{bmatrix} \begin{bmatrix} \sigma_1 \\ \vdots \\ \sigma_r \end{bmatrix} \begin{bmatrix} v_1^T \text{ row} \\ \text{to space} \\ r \times n \end{bmatrix} \]

\[ A = U \Sigma V^T = \begin{bmatrix} u_1 \text{ to } u_m \\ \text{columns} \\ m \times m \end{bmatrix} \begin{bmatrix} \sigma_1 \\ \vdots \\ \sigma_r \end{bmatrix} \begin{bmatrix} v_1^T \text{ row space} \\ v_r^T \text{ n x n} \\ m \times n \end{bmatrix} \]

\[ v_n^T \text{ nullspace} \]

Key ideas \hspace{1cm} $A^T A = V \Sigma^T U^T \Sigma V^T = V \Sigma^T \Sigma V^T$ \hspace{1cm} $AA^T = U \Sigma \Sigma^T U^T$

\[ \text{Eigenvectors!} \quad A^T A v = \sigma^2 v \quad \text{and} \quad AA^T u = \sigma^2 u \]

The $u$’s are chosen so that $A v_k = \sigma_k u_k$  \hspace{1cm} $\sigma_1 \geq \cdots \geq \sigma_r > 0$

\[ k \leq r \quad v_k \text{ and } \sigma_k^2 \text{ from } A^T A \]

\[ u_k = \frac{A v_k}{\sigma_k} \]

\[ u_j^T u_k = \left( \frac{A v_j}{\sigma_j} \right)^T \frac{A v_k}{\sigma_k} = \frac{\sigma_k}{\sigma_j} v_j^T v_k = 0 \]

\[ \text{Square } A \]

\[ |\lambda| \leq \sigma_1 \]

\[ |\lambda| \quad ||x|| = ||Ax|| = ||U \Sigma V^T x|| = ||\Sigma V^T x|| \leq \sigma_1 \quad ||V^T x|| = \sigma_1 \quad ||x|| \]

\[ A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 8 \\ 0 & 0 & 0 \end{bmatrix} \quad \lambda = 0, 0, 0 \quad \text{rank } r = 2 \quad 2 \text{ singular values} \]

\[ \sigma = 8, 1, (0) \quad A = u_1 \sigma_1 v_1^T + u_2 \sigma_2 v_2^T \]

\[ A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} \quad \lambda = 5, 3 \quad \sigma = 3\sqrt{5}, \sqrt{5} \quad u_1 \sigma_1 v_1^T + u_2 \sigma_2 v_2^T = \frac{3}{2} \begin{bmatrix} 1 & 1 \\ 3 & 3 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 3 & -3 \\ -1 & 1 \end{bmatrix} = A \]
7.3 The Geometry of the SVD: Rotate – Stretch – Rotate

\[ A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} \sigma_1 \\ \sigma_2 \end{pmatrix} \begin{pmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix} \]

Four numbers \(a, b, c, d\) in \(A\) produce four numbers \(\theta, \sigma_1, \sigma_2, \phi\) in the SVD

3 \(\times\) 3: Nine numbers in \(A\) produce which 9 numbers for \(U \Sigma V^T\)?

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

Inputs \(x = \) unit circle \hspace{1cm} Outputs \(Ax = \) stretched ellipse

Radius vectors \(v_1\) and \(v_2\) \hspace{1cm} Axis vectors \(\sigma_1 u_1\) and \(\sigma_2 u_2\)
7.4 $A_k$ is Closest to $A$: Principal Component Analysis PCA

SVD $A = UΣV^T = u_1σ_1v_1^T + \cdots + u_rσ_rv_r^T$ $A$ has rank $r$

$A_k = U_kΣ_kV_k^T = u_1σ_1v_1^T + \cdots + u_kσ_kv_k^T$ any $k \leq r$

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

“Eckart-Young” $||A - A_k|| \leq ||A - B||$ if $B$ has rank $k$

Matrix norms $||A||_{\ell^2 \text{ norm}} = σ_1$ $||A||_{\text{Frobenius}} = \sqrt{σ^2_1 + \cdots + σ^2_r}$

$A_0 = \text{matrix of data}$ $A = \text{subtract row average from each row of } A_0$

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

Line closest to data is $u_1$: The key to PCA

Straight line fit using perpendicular distances

$u_1 = \text{eigenvector of } S = \text{first principal component}$

$= \text{singular vector of } A = \text{captures most variance}$

Total variance of the data = Add the eigenvalues of $S = σ^2_1 + \cdots + σ^2_r$
7.5 Computing Eigenvalues of $S$ and Singular Values of $A$

**Step 1** Produce zeros in the matrix $S \rightarrow Q^{-1}SQ = S_0$  
Same $\lambda$’s  
$Q, Q_1, Q_2 = $ orthogonal matrix  
$A \rightarrow Q_1^{-1}AQ_2 = A_0$  
Same $\sigma$’s  

New $S_0$ has only 3 nonzero diagonals  
$A_0$ has only 2 nonzero diagonals

**Step 2** “$QR$ method” uses Gram-Schmidt to orthogonalize columns  
$S = (\text{Orthogonal } Q)(\text{Upper triangular } R)$ at every step

| Factor $S_0 = Q_0R_0$  | Reverse $S_1 = R_0Q_0$  | Repeat $S_1 = Q_1R_1$ and $S_2 = R_1Q_1$ |

Amazing: The off-diagonal entries get small: **Watch** $\sin \theta \rightarrow -\sin^3 \theta$

$$
S_k = Q_kR_k \begin{bmatrix} \cos \theta & \sin \theta \\ \sin \theta & 0 \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} 1 & \sin \theta \cos \theta \\ 0 & -\sin^2 \theta \end{bmatrix}
$$

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

$S_k$ **approaches** $\Lambda$: The eigenvalues $\lambda$ begin to appear on the diagonal

Similar idea for SVD=Golub-Kahan algorithm: $\sigma$’s appear on the diagonal
7.6 Compressing Images by the SVD

Flag with 3 stripes
\[
\begin{bmatrix}
B & B & B & B & B & B \\
B & B & B & B & B & B \\
W & W & W & W & W & W \\
W & W & W & W & W & W \\
R & R & R & R & R & R \\
R & R & R & R & R & R \\
\end{bmatrix} =
\begin{bmatrix}
B \\
B \\
W \\
W \\
R \\
R \\
\end{bmatrix} \begin{bmatrix}
1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 \\
\end{bmatrix}
\]

Triangular flag
\[
\begin{bmatrix}
1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 \\
1 & 1 & 1 & 0 \\
1 & 1 & 1 & 1 \\
\end{bmatrix}
\]
This has all \( \sigma > \frac{1}{2} \)

Compress photographs [link]

Uncompressed image \( = 600 \times 600 = 360,000 \) pixels

Compressed image \( U\Sigma V^T = 600 \times 100 + 100 + 100 \times 600 = 120,000 \)

Serious compression \( U\Sigma V^T = 600 \times 20 + 20 + 20 \times 600 = 24,020 \)

Compression is highly developed See Wikipedia for Eigenfaces
### 7.7 The Victory of Orthogonality

1. Length of $Qx = \text{Length of } x$ $\quad ||Qx||^2 = ||x||^2$ $(Qx)^T(Qy) = x^Ty$
2. All powers $Q^n$ and all products $Q_1Q_2$ remain orthogonal
3. Reflection $H = I - 2uu^T$ is orthogonal + symmetric when $u^Tu = 1$
4. Symmetric matrices have orthogonal eigenvectors $SQ = Q\Lambda$
5. All matrices have orthogonal singular vectors $v$’s and $u$’s: $AV = U\Sigma$
6. The pseudoinverse of $U\Sigma V^T$ is $V\Sigma^+U^T$ Nonzeros in $\Sigma^+$ are $\frac{1}{\sigma}$
7. Polar decomposition $A = QS =$ (orthogonal) (symm positive definite)
8. Gram-Schmidt = Highly valuable $A = QR =$ (orthogonal) (triangular)
9. Orthogonal functions for Fourier series $f(x) = \sum a_k \cos kx + b_k \sin kx$

Big picture of the SVD
Orthogonal rows $\rightarrow$ Orthogonal columns
Part 8
Linear Transformations and Their Matrices

8.1 Examples of Linear Transformations

8.2 Derivative Matrix $D$ and Integral Matrix $D^+$

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

\( T \) is a linear transformation from \( V \) to \( Y \) (inputs to outputs)

Test for linearity

\[
T(cv + dw) = cT(v) + dT(w)
\]

for all \( v, w \) in \( V \)

Example 1  \( V = x-y \) plane  Rotate the plane \( \mathbb{R}^2 \) by a fixed angle \( \theta \)

Straight lines rotate into straight lines (required by linearity)

Center point \( 0 = (0, 0) \) stays put  \( T(0+0) = T(0) + T(0) \) requires \( T(0) = 0 \)

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

Example 2  Input space \( V = \) all 3 by 3 matrices = output space \( Y \)

\( T \) 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  \( T_2(T_1x) = (T_2T_1)x \)

Example 3  \( V = \) all functions \( a + bx + cx^2 \)  \( Y = \) all functions \( d + ex \)

\( T(a + bx + cx^2) = \text{derivative} \) 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 $v_1=1, v_2=x, v_3=x^2$

Express outputs $T(v_1) = 0, T(v_2) = 1, T(v_3) = 2x$ in the output basis

$$T(v_1) = 0 \quad T(v_2) = \frac{dx}{dx} = 1 = u_1 \quad T(v_3) = \frac{d}{dx}(x^2) = 2x = 2u_2$$

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

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

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

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

$$S(1) = x = v_2 \quad S(x) = \frac{1}{2}x^2 = \frac{1}{2}v_3 \quad \text{Integral matrix } E = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$$

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

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

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

$E = \text{pseudoinverse } D^+ \text{ of } D$ Not the inverse because derivative of 1 is 0
8.3 Basis for V and Basis for Y ⇒ Matrix for \( T : V \to Y \)

Every linear transformation \( T : V \to Y \) can be expressed by a matrix \( A \). That matrix \( A \) depends on the basis for \( V \) and the basis for \( Y \).

To construct \( A \): Apply \( T \) to the input basis vectors \( v_1 \) to \( v_n \).

Then \( T(v_j) = a_{1j}y_1 + a_{2j}y_2 + \cdots + a_{mj}y_m \) gives column \( j \) of \( A \).

Input \( v = c_1v_1 + \cdots + c_nv_n \)  
Output \( y = c_1T(v_1) + \cdots + c_nT(v_n) \)

That output \( y \) has coefficients \( Ac \) in the output basis for \( Y \).

Main point! **Multiplication by \( A \) copies the linear transformation \( T \).**

Both linear and both correct for basis ⇒ both correct for every input.

Change input basis to \( V_1, \ldots, V_n \)  
Change output basis to \( Y_1, \ldots, Y_m \)

**The matrix for the same \( T \) in these new bases is** \( M = Y^{-1}AV \)

\( V \) = identity on input space: but basis change from \( v \)’s to \( V \)’s

\( Y \) = identity on output space: but basis change from \( y \)’s to \( Y \)’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 $S = S^T$ and Unitary $Q^{-1} = Q^T$

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

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

$1$ 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 $\overline{z} = x - iy$. $|z|^2 = x^2 + y^2 = z \overline{z}$ and $\frac{1}{z} = \frac{\overline{z}}{|z|^2}$.

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

$\mathbb{R}^n$: vectors with $n$ real components $\leftrightarrow \mathbb{C}^n$: vectors with $n$ complex components

- length: $\|x\|^2 = x_1^2 + \cdots + x_n^2 \leftrightarrow$ length: $\|z\|^2 = |z_1|^2 + \cdots + |z_n|^2$
- transpose: $(A^T)_{ij} = A_{ji} \leftrightarrow$ conjugate transpose: $(A^H)_{ij} = \overline{A_{ji}}$

- dot product: $x^T y = x_1 y_1 + \cdots + x_n y_n \leftrightarrow$ inner product: $u^H v = \overline{u_1 v_1} + \cdots + \overline{u_n v_n}$
- reason for $A^T$: $(Ax)^T y = x^T (A^T y) \leftrightarrow$ reason for $A^H$: $(Au)^H v = u^H (A^H v)$
- orthogonality: $x^T y = 0 \leftrightarrow$ orthogonality: $u^H v = 0$

- symmetric matrices: $S = S^T \leftrightarrow$ Hermitian matrices: $S = S^H$
- $S = QAQ^{-1} = QAQ^T$ (real $\Lambda$) $\leftrightarrow$ $S = U\Lambda U^{-1} = U\Lambda U^H$ (real $\Lambda$)
- orthogonal matrices: $Q^T = Q^{-1} \leftrightarrow$ unitary matrices: $U^H = U^{-1}$

$(Qx)^T (Qy) = x^T y$ and $\|Qx\| = \|x\| \leftrightarrow (Ux)^H (Uy) = x^H y$ and $\|Uz\| = \|z\|$
Part 9: **Complex Numbers and the Fourier Matrix**

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

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

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

**Euler’s Formula**  
$z = re^{i\theta} = r \cos \theta + ir \sin \theta = \frac{x}{r} + ir \frac{y}{r}$

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

Add $\bar{z}_1 + \bar{z}_2 = \bar{z}_1 + \bar{z}_2$  
Multiply $(\bar{z}_1)(\bar{z}_2) = \bar{z}_1 \bar{z}_2$  
Divide $\frac{1}{z} = \frac{\bar{z}}{|z|^2} = \frac{a - ib}{a^2 + b^2}$

---

**Real axis**

**Unit circle**  
$|z| = 1$

**Complex plane**

**On the circle**  
$|z| = |e^{i\theta}| = 1$

Add angles  
$(re^{i\theta})(Re^{i\phi}) = rR e^{i(\theta + \phi)}$  
$(i)^4 = (e^{i\pi/2})^4 = e^{i2\pi} = 1$
9.2 Complex Matrices: Hermitian $S = \overline{S}^T$ and Unitary $Q^{-1} = \overline{Q}^T$

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

Automatic for computer systems like MATLAB and Julia

Inner product = Dot product = $\overline{x}^T y = x_1 y_1 + \cdots + x_n y_n$

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

Hermitian matrix $S = \begin{bmatrix} 2 & 3 - 3i \\ 3 + 3i & 5 \end{bmatrix} = \overline{S}^T$

$S$ has real eigenvalues 8, $-1$ and perpendicular eigenvectors

$$\det(S - \lambda I) = \lambda^2 - 7\lambda + 10 - |3 + 3i|^2 = (\lambda - 8)(\lambda + 1)$$

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

Unitary matrix $Q = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 & 1 - i \\ 1 + i & -1 \end{bmatrix}$

Orthonormal columns $Q = Q^{-1}$

The Fourier matrix $\frac{F}{\sqrt{N}}$ is the most important unitary matrix
9.3 Fourier Matrix $F$ and the Discrete Fourier Transform

Fourier Matrix $F_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & i^2 & i^3 \\ 1 & i^2 & i^4 & i^6 \\ 1 & i^3 & i^6 & i^9 \end{bmatrix}$

DFT Matrix $\overline{F}_4 = \text{powers of } -i$

$F_N$ and $\overline{F}_N$

$N$ by $N$ matrices

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

$F_{jk} = w^{jk} = e^{2\pi ijk/N}$

Columns $k = 0$ to $N - 1$

$F_N = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & i^2 & i^3 \\ 1 & i^2 & i^4 & i^6 \\ 1 & i^3 & i^6 & i^9 \end{bmatrix}$

$\overline{F}_N = N I$

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

$N = 2$

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

$F_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$

$\overline{F}_2 F_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = N I$

Discrete Fourier Transform $f \rightarrow c$ $c = F_N^{-1} f$

Inverse Fourier Transform $c \rightarrow f$ $f = F_N c$
9.4 Cyclic Convolution and the Convolution Rule

\[ c = (1, 2, 3) \quad \text{Convolution}\quad c \ast d = (5, 10, 19, 8, 12) \]
\[ d = (5, 0, 4) \quad \text{Cyclic}\quad c \ast\circ d = (5 + 8, 10 + 12, 19) \]
\[
(1 + 2x + 3x^2)(5 + 4x^2) = 5 + 10x + 19x^2 + 8x^3 + 12x^4
\]
\[ x^3 = 1 \text{ for cyclic}\quad (5 + 8) + (10 + 12)x + 19x^2 \]

\[
\begin{array}{ccc}
5 & 10 & 15 \\
5 & 10 & 19 & 8 & 12
\end{array}
\]

Convolution matrices
\[ C = \begin{bmatrix} 1 & 3 & 2 \\ 2 & 1 & 3 \\ 3 & 2 & 1 \end{bmatrix} \quad D = \begin{bmatrix} 5 & 4 & 0 \\ 0 & 5 & 4 \\ 4 & 0 & 5 \end{bmatrix} \]
\[ CD = \begin{bmatrix} 13 & 19 & 22 \\ 22 & 13 & 19 \\ 19 & 22 & 13 \end{bmatrix} \quad DC = \begin{bmatrix} 13 & 19 & 22 \\ 22 & 13 & 19 \\ 19 & 22 & 13 \end{bmatrix} \]

Eigenvectors of \( C \) and \( D \) = columns of the Fourier matrix \( F \)

Eigenvalues of \( C = F^{-1}c \) \quad Eigenvalues of \( D = F^{-1}d \)

<table>
<thead>
<tr>
<th>Convolve vectors</th>
<th>Multiply transforms</th>
<th>Convolution Rule</th>
<th>Key to Signal Processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x \ast y = (x_1y_1, \ldots, x_ny_n) )</td>
<td></td>
<td>( F(c \ast\circ d) = (Fc) \ast (Fd) )</td>
<td></td>
</tr>
</tbody>
</table>
9.5 FFT: The Fast Fourier Transform

Direct matrix multiplication of \( c \) by \( F_N \) needs \( N^2 \) multiplications

FFT factorization with many zeros: \( \frac{1}{2} N \log_2 N \) multiplications

\[ N = 2^{10} = 1024 \quad \log_2 N = 10 \quad 1 \text{ million reduced to } 5000 \]

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

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

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 \( N \): \( \log_2 N \) steps to Fast Transform

Each step has \( N \) multiplications from the diagonal matrices \( D \)

One overall permutation = product of the \( P \)'s

**FFTW** is hardwired in many computers/bases other than 2
Part 10
Learning from Data by Gradient Descent

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

10.2 Counting Flat Pieces in the Graph of $F$

10.3 Minimizing the Loss: Stochastic Gradient Descent

10.4 Slow Convergence with Zigzag: Add Momentum

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

10.6 Backpropagation: Chain Rule for $\nabla F$
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 $v_0$

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

Level $k$ from $k - 1$: Multiply each vector by $A_k$, add $b_k$, apply ReLU

$$v_k = F_k(v_{k-1}) = \text{ReLU}(A_k v_{k-1} + b_k)$$

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

ReLU = ramp function

$=$ Rectified Linear Unit

This gives the nonlinearity that learning functions need

Levels 0 to $L$ Output $v_L = F_L(\ldots(F_2(F_1(v_0)))) = F(v_0)$

$F(x, v_0) =$ Composition of $L$ piecewise linear $F_k : F$ is piecewise linear

Each level contributes a weight matrix $A_k$ and a vector $b_k$ to $x$

Neural Net $$(Av)_1$$ \hspace{1cm} [(Av + b)_1]_+$$ \hspace{1cm} \text{One hidden layer}$$

$$(Av)_q$$ \hspace{1cm} [(Av + b)_q]_+$$

$p = 3, q = 4$ \hspace{1cm} $p = 3, q = 4$ $$(Av)_q$$ \hspace{1cm} [(Av + b)_q]_+$

$pq + 2q = 20$ weights

$r(4, 3) = 15$ linear pieces in the graph of $w = F(v)$

65
10.2 Counting Flat Pieces in the Graph of $F$

The weight matrices $A_k$ and bias vectors $b_k$ produce $F = \text{learning function}$

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

\[
\begin{array}{c}
N = 3 \text{ folds in } \\
p = 2 \text{ dimensions }
\end{array}
\]

Start with 2 folds

\[r(2, 2) = 4\]

Add new fold $H$

\[r(2, 1) = 3\]

Total 7 regions

The $r(2, 1) = 3$ pieces of the new fold $H$ create new regions $1b, 2b, 3b$. Then the count becomes $r(3, 2) = 4 + 3 = 7$ flat regions in the continuous piecewise linear surface $v_2 = F(v_0)$. A fourth fold would cross all three existing folds and create 4 new regions, so $r(4, 2) = 7 + 4 = 11$.

The count $r$ 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 $v$ in $\mathbb{R}^p$, suppose the graph of $F(v)$ 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 $r(N, p)$:

\[
r(N, p) = \binom{N}{0} + \binom{N}{1} + \cdots + \binom{N}{p}.
\]

These binomial coefficients are

\[
\binom{N}{i} = \frac{N!}{i!(N - i)!} \quad \text{with } 0! = 1 \quad \text{and} \quad \binom{N}{0} = 1 \quad \text{and} \quad \binom{N}{i} = 0 \quad \text{for } i > N.
\]

With more layers: $N$ folds from $N$ ReLU’s: still $\approx r(N, p) \approx cN^p$ pieces
10.3 Minimizing the Loss : Stochastic Gradient Descent

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

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

The graph of $y = F(x) = F(x_1, \ldots, x_N)$ 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 $F = ||x||^2 + 5$ is $F_{\text{min}} = 5$ at the point $x = \text{arg min } F = 0$

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

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

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

Recompute the gradient $\nabla F(x_1)$ at the new point $x_1$

At every step follow the gradient $\nabla F(x_k)$ to $x_{k+1} = x_k - s_k \nabla F(x_k)$

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

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

**Error** **Square loss** $||\text{output layer } L - \text{known output} ||^2$ or “Cross-entropy loss”

**Solution 1** Use error in only one randomly chosen sample / one $v_0$

**Solution B** Use sum of errors in only $B$ random samples : **minibatch**

**Stochastic gradient descent** has new sampling at every step. **Successful**
10.4 Slow Convergence with Zigzag: Add Momentum

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

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

\[
\begin{align*}
  x_k &= b \left( \frac{b - 1}{b + 1} \right)^k \\
  y_k &= \left( \frac{1 - b}{1 + b} \right)^k \\
  F(x_k, y_k) &= \left( \frac{1 - b}{1 + b} \right)^{2k} F(b, 1)
\end{align*}
\]

Crucial ratio $\left( \frac{1 - b}{1 + b} \right)^2$ 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** $z_{k+1}$ remembers $z_k$

\[
(x, y)_{k+1} = (x, y)_k - s z_k \\
 z_{k+1} - \nabla F(x, y)_{k+1} = \beta z_k
\]

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

Optimal $\beta$

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

**ADAM** $G_k$ combines all earlier gradients by $G_k = \delta G_{k-1} + (1 - \delta) \nabla F(x_k)$

**Question** Why do the weights (matrices $A_k$) work well for unseen data?
10.5 Convolutional Neural Nets: CNN in 1D and 2D

Convolution matrix = Moving average

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

A has constant diagonals

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

2D Convolution for Images
Windows move across and down
Nine \( 3 \times 3 \) windows in \( 5 \times 5 \) square
Center points are marked 1 to 9
Only \( 3^2 = 9 \) weights to choose

\[
\begin{array}{ccc}
1 & 2 & 3 \\
4 & 5 & 6 \\
7 & 8 & 9 \\
\end{array}
\]

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 \( w_k \) to probabilities \( p_k = e^{w_k} / \sum e^{w_k} \)

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 $\nabla F$

$F(x) = \text{minimum}$  \hspace{2mm} $\nabla F = \text{partial derivatives}$  \hspace{2mm} $\frac{\partial \text{errors}}{\partial \text{weights}} = \text{zero}$

Chain rule  \hspace{2mm} $\frac{d}{dx} \left( F_2(F_1(x)) \right) = \left( \frac{dF_2}{dF_1} (F_1(x)) \right) \left( \frac{dF_1}{dx} (x) \right)$

Multivariable chain rule  \hspace{2mm} $\frac{\partial w}{\partial u} = \left( \frac{\partial w}{\partial v} \right) \left( \frac{\partial v}{\partial u} \right)$  \hspace{2mm} $L$ layers in chain

Multiply $L$ matrices

$$\frac{\partial w}{\partial v} = \begin{bmatrix} \frac{\partial w_1}{\partial v_1} & \cdots & \frac{\partial w_1}{\partial v_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial w_p}{\partial v_1} & \cdots & \frac{\partial w_p}{\partial v_n} \end{bmatrix} \quad \frac{\partial v}{\partial u} = \begin{bmatrix} \frac{\partial v_1}{\partial u_1} & \cdots & \frac{\partial v_1}{\partial u_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial v_n}{\partial u_1} & \cdots & \frac{\partial v_n}{\partial u_m} \end{bmatrix}$$

At each layer

Derivatives before ReLU  \hspace{2mm} $w = Av + b$  \hspace{2mm} $\frac{\partial w_i}{\partial b_j} = \delta_{ij} = 0 \text{ or } 1$  \hspace{2mm} $\frac{\partial w_i}{\partial A_{jk}} = \delta_{ij}v_k$

Product of matrices $ABC$  \hspace{2mm} $AB$ first or $BC$ first?  \hspace{2mm} Forward or back?

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

Example  \hspace{2mm} $A = m \times n$  \hspace{2mm} $B = n \times p$  \hspace{2mm} $C = p \times 1$ vector  \hspace{2mm} **Don’t multiply $AB$**!

Backpropagation = Automatic Differentiation = the key to speed
Part 11

Basic Statistics :
Mean, Variance, Covariance

11.1 Mean and Variance : Actual and Expected
11.2 Probability Distributions : Binomial, Poisson, Normal
11.3 Covariance Matrices and Joint Probabilities
11.4 Three Basic Inequalities of Statistics
11.5 Markov Matrices and Markov Chains
Part 11: Basic Statistics: Mean, Variance, Covariance

11.1 Mean and Variance: Actual and Expected

The sample mean \( \mu \) is the average of outputs from \( N \) trials

The expected mean \( m \) is based on probabilities \( p_1, \ldots, p_n \) of outputs \( x_1, \ldots, x_n \)

\[ \text{Expected value } m = E[x] = p_1 x_1 + \cdots + p_n x_n \]

Law of Large Numbers: With probability 1, sample mean \( \rightarrow m \) as \( N \rightarrow \infty \)

The sample variance measures the spread around the sample mean \( \mu \)

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

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

\[ \sigma^2 = E[(x - m)^2] = p_1 (x_1 - m)^2 + \cdots + p_n (x_n - m)^2 \]

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

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

Continuous probability distributions: Sums change to integrals

\[ \int p(x) \, dx = 1 \quad m = \int x \, p(x) \, dx \quad \sigma^2 = \int (x - m)^2 \, p(x) \, dx \]
11.2 Probability Distributions: Binomial, Poisson, Normal

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

\[
p_{1,1} = p \quad p_{n,n} = p^n \quad p_{k,n} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \quad (0! = 1)
\]

Mean \( m \) in \( n \) trials = \( np \) \quad Variance \( \sigma^2 \) in \( n \) trials = \( np \, (1 - p) \)

2 Poisson: Rare events \( p \to 0 \), many trials \( n \to \infty \)  Keep \( np = \lambda \) constant

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

Poisson mean = \( \lambda \) \quad variance \( \sigma^2 = \lambda \) \quad Limits of binomial \( np \) and \( np \, (1 - p) \)

3 Normal distribution: \( N(m, \sigma^2) \) has

\[
p(x) = \frac{1}{\sqrt{2\pi} \, \sigma} e^{-(x - m)^2 / 2\sigma^2}
\]

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

Shifted and scaled \( X = \frac{x - m}{\sigma} \) \quad Centered and normalized

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

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

Normal \( p(x) \) for \( n \) variables \quad Means \( m = (m_1, \ldots, m_n) \) \quad Covariance matrix \( V \)

\[
p(x) = p(x_1, \ldots, x_n) = \frac{1}{(\sqrt{2\pi})^n \sqrt{\det V}} e^{-\frac{(x - m)^T V^{-1} (x - m)}{2}}
\]
11.3 Covariance Matrices and Joint Probabilities

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

**Mean** \( \mathbf{m} = (m_x, m_y) = (\text{average age, average height}) \)

<table>
<thead>
<tr>
<th>Joint probabilities ( p_{ij} ) = probability that age = $i$ and height = $j$</th>
</tr>
</thead>
</table>

\[
p_i = \sum_j p_{ij} = \text{probability that age} = i \text{ allowing all heights} j
\]

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

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

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

\( \mathbf{V} \) = sum of positive semidefinite rank 1 matrices = semidefinite or definite

\( \mathbf{V} \) is positive definite unless age tells you the exact height (dependent case)

\( \mathbf{V} \) is a diagonal matrix if age and height are **independent** : covariance = 0

**Coin flip**

Glue 2 coins together \( \mathbf{V} = \begin{bmatrix} 1/4 & 1/4 \\ 1/4 & 1/4 \end{bmatrix} = \text{dependent case: semidefinite } \mathbf{V} \)

Separate the coins \( \mathbf{V} = \begin{bmatrix} 1/4 & 0 \\ 0 & 1/4 \end{bmatrix} = \text{independent: diagonal } \mathbf{V} \)
11.4 Three Basic Inequalities of Statistics

Markov’s inequality when $x \geq 0$: No negative samples

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

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

Markov’s inequality says probability $p_3 + p_4 + p_5 + \cdots \leq \frac{m}{3}$

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

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

Chebyshev’s inequality  The probability of $|x - m| \geq a$ is at most $\frac{\sigma^2}{a^2}$

Proof  Apply Markov’s inequality to the new variable $y = |x - m|^2$

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

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

Chernoff’s inequality  $S = X_1 + \cdots + X_n$ independent random variables

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

\[
\begin{align*}
\Pr \left( S \geq (1 + \delta)\bar{S} \right) &\leq e^{-\bar{S}\delta^2/(2+\delta)} \\
\Pr \left( S \leq (1 - \delta)\bar{S} \right) &\leq e^{-\bar{S}\delta^2/2}
\end{align*}
\]

Exponential dropoff!

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

Reason: A large sum $S$ usually needs several $X_i$ to be large / unlikely!
11.5 Markov Matrices and Markov Chains

Markov matrix \( M_{ij} \geq 0 \) All columns add to 1

Perron-Frobenius \( \lambda_{\text{max}} = 1 \) | other \( \lambda \leq 1 \) if all \( M_{ij} \geq 0 \)

Eigenvalues of \( M \) | other \( \lambda \leq 1 \) if all \( M_{ij} > 0 \)

Markov chain \( p_{n+1} = Mp_n \) Probabilities at times \( n + 1 \) and \( n \)

\[
M = \begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix}
\]
has \( \lambda = 1 \) and \( \lambda = \frac{1}{2} \)

\( M^n \) has \( \lambda = 1 \) and \( \lambda = \left( \frac{1}{2} \right)^n \)

\[
\begin{bmatrix}
\text{Rental cars in Chicago} \\
\text{Rental cars in Denver}
\end{bmatrix}_{n+1}
= M
\begin{bmatrix}
\text{in Chicago} \\
\text{in Denver}
\end{bmatrix}_n
\]

\( y_{n+1} = M y_n \)

Start in Chicago \( y_0 = \begin{bmatrix} 100 \\ 0 \end{bmatrix} \) \( y_1 = \begin{bmatrix} 80 \\ 20 \end{bmatrix} \) \( y_2 = \begin{bmatrix} 70 \\ 30 \end{bmatrix} \) \( y_3 = \begin{bmatrix} 65 \\ 35 \end{bmatrix} \)

Start in Denver \( y_0 = \begin{bmatrix} 0 \\ 100 \end{bmatrix} \) \( y_1 = \begin{bmatrix} 30 \\ 70 \end{bmatrix} \) \( y_2 = \begin{bmatrix} 45 \\ 55 \end{bmatrix} \) \( y_\infty = \begin{bmatrix} 60 \\ 40 \end{bmatrix} \)

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