About this chapter

• Not a comprehensive survey of all of linear algebra

• Focused on the subset most relevant to deep learning

• Larger subset: e.g., *Linear Algebra* by Georgi Shilov
Scalars

• A scalar is a single number
• Integers, real numbers, rational numbers, etc.
• We denote it with italic font:

\[ a, n, x \]
Vectors

• A vector is a 1-D array of numbers:

\[
x = \begin{bmatrix}
  x_1 \\
  x_2 \\
  \vdots \\
  x_n
\end{bmatrix}.
\]  \hspace{1cm} (2.1)

• Can be real, binary, integer, etc.

• Example notation for type and size:

\(\mathbb{R}^n\)
Matrices

- A matrix is a 2-D array of numbers:

\[
\begin{bmatrix}
A_{1,1} & A_{1,2} \\
A_{2,1} & A_{2,2}
\end{bmatrix}.
\] (2.2)

- Example notation for type and shape:

\[A \in \mathbb{R}^{m \times n}\]
Tensors

• A tensor is an array of numbers, that may have
  • zero dimensions, and be a scalar
  • one dimension, and be a vector
  • two dimensions, and be a matrix
  • or more dimensions.
Matrix Transpose

\[(A^\top)_{i,j} = A_{j,i}. \quad (2.3)\]

\[
A = \begin{bmatrix}
A_{1,1} & A_{1,2} \\
A_{2,1} & A_{2,2} \\
A_{3,1} & A_{3,2}
\end{bmatrix} \Rightarrow A^\top = \begin{bmatrix}
A_{1,1} & A_{2,1} & A_{3,1} \\
A_{1,2} & A_{2,2} & A_{3,2}
\end{bmatrix}
\]

Figure 2.1: The transpose of the matrix can be thought of as a mirror image across the main diagonal.

\[(AB)^\top = B^\top A^\top. \quad (2.9)\]
Matrix (Dot) Product

\[ C = AB. \]  
\[ C_{i,j} = \sum_k A_{i,k} B_{k,j}. \]

\[ \text{Must match} \]
Identity Matrix

\[
\begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}
\]

Figure 2.2: Example identity matrix: This is $I_3$.

\[\forall \mathbf{x} \in \mathbb{R}^n, I_n \mathbf{x} = \mathbf{x}.\]  \hspace{1cm} (2.20)
Systems of Equations

\[ Ax = b \]  \hspace{1cm} (2.11)

expands to

\[ A_1 \cdot x = b_1 \]  \hspace{1cm} (2.12)

\[ A_2 \cdot x = b_2 \]  \hspace{1cm} (2.13)

...  \hspace{1cm} (2.14)

\[ A_m \cdot x = b_m \]  \hspace{1cm} (2.15)

(Goodfellow 2016)
Solving Systems of Equations

- A linear system of equations can have:
  - No solution
  - Many solutions
  - Exactly one solution: this means multiplication by the matrix is an invertible function
Matrix Inversion

- Matrix inverse:
  \[ A^{-1}A = I_n. \]  \hspace{1cm} (2.21)
- Solving a system using an inverse:
  \[ Ax = b \]  \hspace{1cm} (2.22)
  \[ A^{-1}Ax = A^{-1}b \]  \hspace{1cm} (2.23)
  \[ I_nx = A^{-1}b \]  \hspace{1cm} (2.24)
- Numerically unstable, but useful for abstract analysis
Invertibility

- Matrix can’t be inverted if…
  - More rows than columns
  - More columns than rows
  - Redundant rows/columns ("linearly dependent", "low rank")
Norms

• Functions that measure how “large” a vector is

• Similar to a distance between zero and the point represented by the vector

\[ f(x) = 0 \Rightarrow x = 0 \]

\[ f(x + y) \leq f(x) + f(y) \text{ (the triangle inequality)} \]

\[ \forall \alpha \in \mathbb{R}, f(\alpha x) = |\alpha| f(x) \]
Norms

- $L^p$ norm

$$\|x\|_p = \left(\sum_i |x_i|^p\right)^{\frac{1}{p}}$$

- Most popular norm: L2 norm, $p=2$

- L1 norm, $p=1$: $\|x\|_1 = \sum_i |x_i|$. \hspace{1cm} (2.31)

- Max norm, infinite $p$: $\|x\|_\infty = \max_i |x_i|$. \hspace{1cm} (2.32)
Special Matrices and Vectors

- **Unit vector:**

  \[ \| \mathbf{x} \|_2 = 1. \]  

- **Symmetric Matrix:**

  \[ A = A^\top. \]  

- **Orthogonal matrix:**

  \[ A^\top A = AA^\top = I. \]

  \[ A^{-1} = A^\top \]  

(Goodfellow 2016)
Trace

\[ \text{Tr}(A) = \sum_i A_{i,i}. \]  

(2.48)

\[ \text{Tr}(ABC) = \text{Tr}(CAB) = \text{Tr}(BCA) \]  

(2.51)
Learning linear algebra

• Do a lot of practice problems

• Start out with lots of summation signs and indexing into individual entries

• Eventually you will be able to mostly use matrix and vector product notation quickly and easily
Linear Algebra - Part II
Projection, Eigendecomposition, SVD

Punit Shah

(Adapted from Sargur Srihari’s slides)
Brief Review from Part 1

- **Symmetric Matrix:**
  \[ A = A^T \]

- **Orthogonal Matrix:**
  \[ A^T A = AA^T = I \quad \text{and} \quad A^{-1} = A^T \]

- **L2 Norm:**
  \[ \|x\|_2 = \sqrt{\sum_{i} x_i^2} \]
Angle Between Vectors

- Dot product of two vectors can be written in terms of their L2 norms and the angle $\theta$ between them.

$$ a^T b = \|a\|_2 \|b\|_2 \cos(\theta) $$
Cosine Similarity

- Cosine between two vectors is a measure of their similarity:
  \[ \cos(\theta) = \frac{a \cdot b}{||a|| \ ||b||} \]

- **Orthogonal Vectors**: Two vectors \(a\) and \(b\) are orthogonal to each other if \(a \cdot b = 0\).
Vector Projection

- Given two vectors $\mathbf{a}$ and $\mathbf{b}$, let $\hat{\mathbf{b}} = \frac{\mathbf{b}}{|\mathbf{b}|}$ be the unit vector in the direction of $\mathbf{b}$.
- Then $\mathbf{a}_1 = a_1 \hat{\mathbf{b}}$ is the orthogonal projection of $\mathbf{a}$ onto a straight line parallel to $\mathbf{b}$, where

$$a_1 = ||\mathbf{a}|| \cos(\theta) = \mathbf{a} \cdot \hat{\mathbf{b}} = \mathbf{a} \cdot \frac{\mathbf{b}}{|\mathbf{b}|}$$

Image taken from wikipedia.
Diagonal Matrix

- Diagonal matrix has mostly zeros with non-zero entries only in the diagonal, e.g. identity matrix.

- A square diagonal matrix with diagonal elements given by entries of vector \( \mathbf{v} \) is denoted:
  \[
  \text{diag}(\mathbf{v})
  \]

- Multiplying vector \( \mathbf{x} \) by a diagonal matrix is efficient:
  \[
  \text{diag}(\mathbf{v})\mathbf{x} = \mathbf{v} \odot \mathbf{x}
  \]

- Inverting a square diagonal matrix is efficient:
  \[
  \text{diag}(\mathbf{v})^{-1} = \text{diag}\left(\left[\frac{1}{v_1}, \ldots, \frac{1}{v_n}\right]^T\right)
  \]
Zero Determinant

If \( \det(A) = 0 \), then:

- \( A \) is linearly dependent.
- \( Ax = b \) has no solution or infinitely many solutions.
- \( Ax = 0 \) has a non-zero solution.
Matrix Decomposition

- We can decompose an integer into its prime factors, e.g. $12 = 2 \times 2 \times 3$.

- Similarly, matrices can be decomposed into factors to learn universal properties:

  $A = V \text{diag}(\lambda) V^{-1}$
Eigenvectors

- An eigenvector of a square matrix $A$ is a nonzero vector $v$ such that multiplication by $A$ only changes the scale of $v$.

\[ Av = \lambda v \]

- The scalar $\lambda$ is known as the eigenvalue.

- If $v$ is an eigenvector of $A$, so is any rescaled vector $sv$. Moreover, $sv$ still has the same eigenvalue. Thus, we constrain the eigenvector to be of unit length:

\[ ||v|| = 1 \]