Linear Algebra

Lecture slides for Chapter 2 of *Deep Learning*
Ian Goodfellow
2016-06-24

Adapted by m.n. for CMPS 392

About this chapter

- Not a comprehensive survey of all of linear algebra
- Focused on the subset most relevant to deep learning
- For a full introductory linear algebra course, I recommend the course and book of **Gilbert strang**.
- https://ocw.mit.edu/faculty/gilbert-strang/

Scalars

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

$$a, n, x$$

(Goodfellow 2016)

Vectors

- A vector is a 1-D array of numbers:

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}. \tag{2.1}$$

- Can be real, binary, integer, etc.

- Example notation for type and size:

$$\mathbb{R}^n$$

(Goodfellow 2016)

Matrices

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

Row

$$\begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix}. \tag{2.2}$$

Column

- Example notation for type and shape:

$$\boldsymbol{A} \in \mathbb{R}^{m \times n}$$

(Goodfellow 2016)

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.

(Goodfellow 2016)

Matrix Transpose

$$(\boldsymbol{A}^\top)_{i,j} = A_{j,i}. \tag{2.3}$$

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

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

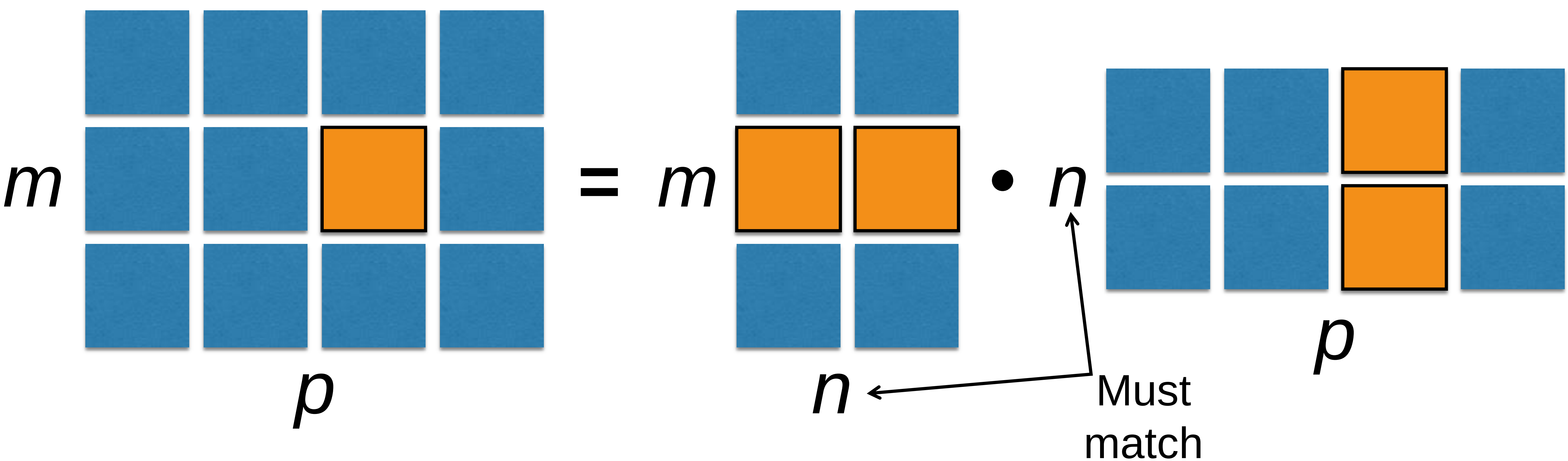
$$(\boldsymbol{AB})^\top = \boldsymbol{B}^\top \boldsymbol{A}^\top. \tag{2.9}$$

(Goodfellow 2016)

Matrix (Dot) Product

$$C = AB. \tag{2.4}$$

$$C_{i,j} = \sum_k A_{i,k} B_{k,j}. \tag{2.5}$$



(Goodfellow 2016)

Identity Matrix

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

Figure 2.2: *Example identity matrix*: This is $\boldsymbol{I}_3$.

$$\forall \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{I}_n \boldsymbol{x} = \boldsymbol{x}. \tag{2.20}$$

(Goodfellow 2016)

Systems of Equations

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \tag{2.11}$$

expands to

$$\boldsymbol{A}_{1,:}\boldsymbol{x} = b_1 \tag{2.12}$$

$$\boldsymbol{A}_{2,:}\boldsymbol{x} = b_2 \tag{2.13}$$

$$\dots \tag{2.14}$$

$$\boldsymbol{A}_{m,:}\boldsymbol{x} = b_m \tag{2.15}$$

(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

(Goodfellow 2016)

Matrix Inversion

- Matrix inverse:

$$A^{-1}A = I_n. \tag{2.21}$$

- Solving a system using an inverse:

$$Ax = b \tag{2.22}$$

$$A^{-1}Ax = A^{-1}b \tag{2.23}$$

$$I_n x = A^{-1}b \tag{2.24}$$

- Numerically unstable, but useful for abstract analysis

(Goodfellow 2016)

Invertibility

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

(Goodfellow 2016)

Norms

- Functions that measure how "large" a vector is
- Similar to a distance between zero and the point represented by the vector

  - $f(\boldsymbol{x}) = 0 \Rightarrow \boldsymbol{x} = \boldsymbol{0}$
  - $f(\boldsymbol{x} + \boldsymbol{y}) \leq f(\boldsymbol{x}) + f(\boldsymbol{y})$ (the *triangle inequality*)
  - $\forall \alpha \in \mathbb{R}, f(\alpha \boldsymbol{x}) = |\alpha| f(\boldsymbol{x})$

(Goodfellow 2016)

Norms

- L^p norm

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

- Most popular norm: L2 norm, p=2
- L1 norm, p=1: $$||\boldsymbol{x}||_1 = \sum_i |x_i|. \tag{2.31}$$
- Max norm, infinite p: $$||\boldsymbol{x}||_\infty = \max_i |x_i|. \tag{2.32}$$

(Goodfellow 2016)

Special Matrices and Vectors

- Unit vector:

$$||\boldsymbol{x}||_2 = 1. \tag{2.36}$$

- Symmetric Matrix:

$$\boldsymbol{A} = \boldsymbol{A}^\top. \tag{2.35}$$

- Orthogonal matrix:

$$\boldsymbol{A}^\top \boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{I}. \tag{2.37}$$

$$\boldsymbol{A}^{-1} = \boldsymbol{A}^\top$$

(Goodfellow 2016)

Eigendecomposition

- Eigenvector and eigenvalue:

$$\boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v}. \tag{2.39}$$

- Eigendecomposition of a diagonalizable matrix:

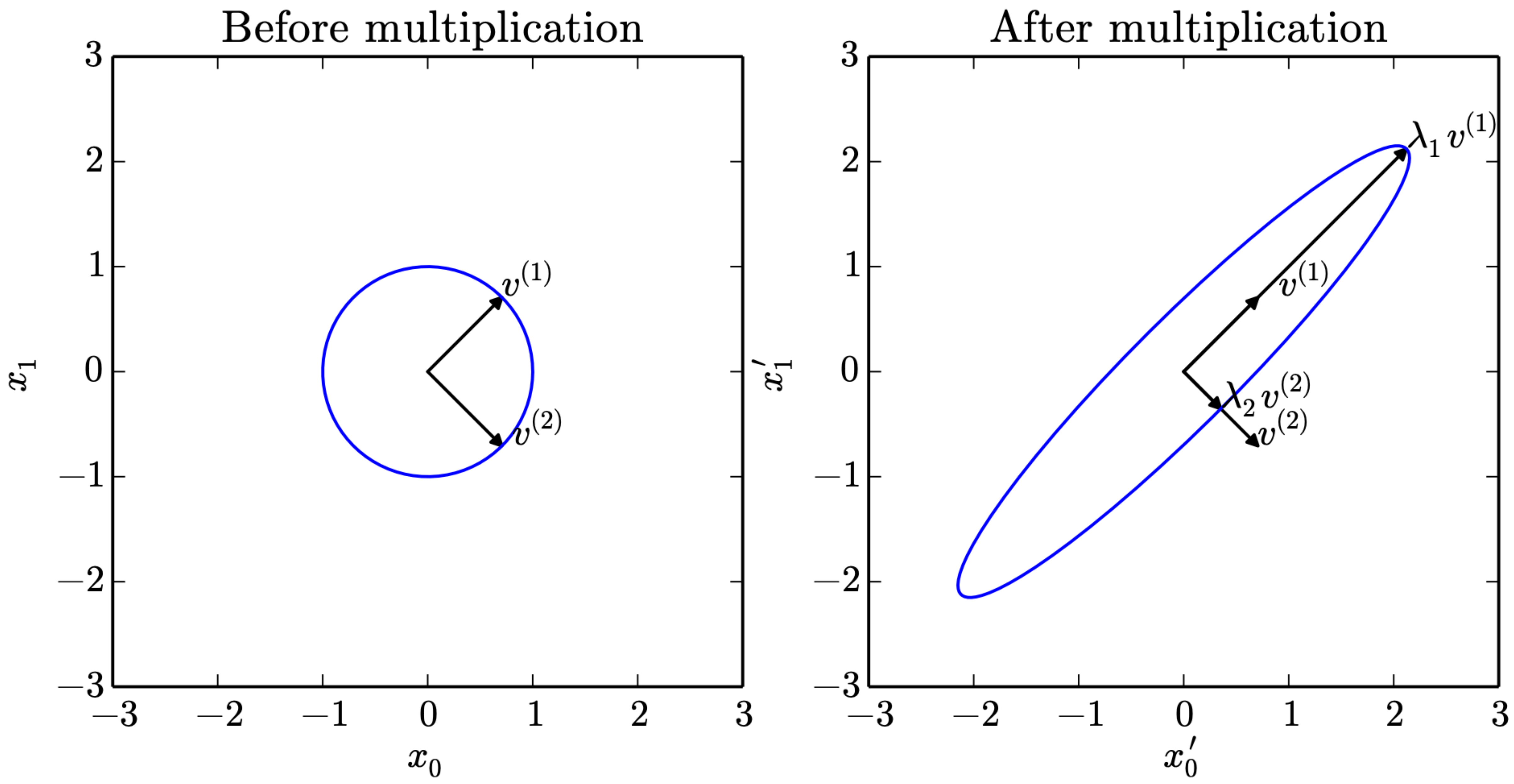
$$\boldsymbol{A} = \boldsymbol{V} \operatorname{diag}(\boldsymbol{\lambda}) \boldsymbol{V}^{-1}. \tag{2.40}$$

- Every real symmetric matrix has a real, orthogonal eigendecomposition:

$$\boldsymbol{A} = \boldsymbol{Q} \boldsymbol{\Lambda} \boldsymbol{Q}^{\top} \tag{2.41}$$

(Goodfellow 2016)

Effect of Eigenvalues



(Goodfellow 2016)

Singular Value Decomposition

- Similar to eigendecomposition
- More general; matrix need not be square

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\top. \tag{2.43}$$

Moore-Penrose Pseudoinverse

$$\boldsymbol{x} = \boldsymbol{A}^{+}\boldsymbol{y}$$

- If the equation has:
  - Exactly one solution: this is the same as the inverse.
  - No solution: this gives us the solution with the smallest error $||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}||_2$.
  - Many solutions: this gives us the solution with the smallest norm of $\boldsymbol{x}$.

(Goodfellow 2016)

Computing the Pseudoinverse

The SVD allows the computation of the pseudoinverse:

$$\boldsymbol{A}^+ = \boldsymbol{V}\boldsymbol{D}^+\boldsymbol{U}^\top, \tag{2.47}$$

Take reciprocal of non-zero entries

(Goodfellow 2016)

Trace

$$\mathrm{Tr}(\boldsymbol{A}) = \sum_i \boldsymbol{A}_{i,i}. \tag{2.48}$$

$$\mathrm{Tr}(\boldsymbol{ABC}) = \mathrm{Tr}(\boldsymbol{CAB}) = \mathrm{Tr}(\boldsymbol{BCA}) \tag{2.51}$$

(Goodfellow 2016)

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