CS 143 Linear Algebra Review

Stefan Roth

September 29, 2003

Introductory Remarks

- This review does not aim at mathematical rigor very much, but instead at ease of understanding and conciseness.
- Please see the course webpage on background material for links to further linear algebra material.

Overview

- Vectors in $\mathbb{R}^n$
- Scalar product
- Bases and transformations
- Inverse Transformations
- Eigendecomposition
- Singular Value Decomposition

Warm up: Vectors in $\mathbb{R}^n$

- We can think of vectors in two ways
  - As points in a multidimensional space with respect to some coordinate system
  - As a translation of a point in a multidimensional space, of the coordinate origin.
- Example in two dimensions ($\mathbb{R}^2$):

Vectors in $\mathbb{R}^n$ (II)

- Notation:

$$\mathbf{x} \in \mathbb{R}^n, \qquad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \qquad \mathbf{x} = (x_1, x_2, \ldots, x_n)$$

- We will skip some rather obvious things: Adding and subtracting of vectors, multiplication by a scalar, and simple properties of these operations.

Scalar Product

- A product of two vectors
- Amounts to projection of one vector onto the other
- Example in 2D:

The shown segment has length $\langle \mathbf{x}, \mathbf{y} \rangle$, if $\mathbf{x}$ and $\mathbf{y}$ are unit vectors.

Scalar Product (II)

- Various notations:
  - $\langle \mathbf{x}, \mathbf{y} \rangle$
  - $\mathbf{x}\mathbf{y}$
  - $\mathbf{x}\ \mathbf{y}$ or $\mathbf{x} \cdot \mathbf{y}$
- We will only use the first and second one!
- Other names: dot product, inner product

Scalar Product (III)

- Built on the following axioms:
  - Linearity:
  $$\langle \mathbf{x} + \mathbf{x}', \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{x}', \mathbf{y} \rangle$$
  $$\langle \mathbf{x}, \mathbf{y} + \mathbf{y}' \rangle = \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{x}, \mathbf{y}' \rangle$$
  $$\langle \lambda \mathbf{x}, \mathbf{y} \rangle = \lambda \langle \mathbf{x}, \mathbf{y} \rangle$$
  $$\langle \mathbf{x}, \lambda \mathbf{y} \rangle = \lambda \langle \mathbf{x}, \mathbf{y} \rangle$$
  - Symmetry: $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$
  - Non-negativity:
  $\forall \mathbf{x} \neq 0: \ \langle \mathbf{x}, \mathbf{x} \rangle > 0 \qquad \langle \mathbf{x}, \mathbf{x} \rangle = 0 \Leftrightarrow \mathbf{x} = 0$

Scalar Product in $\mathbb{R}^n$

- Here: Euclidean space
- Definition:

$$\mathbf{x}, \mathbf{y} \in \mathbb{R}^n : \; \langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} x_i \cdot y_i$$

- In terms of angles:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \|\mathbf{x}\| \cdot \|\mathbf{y}\| \cdot \cos(\angle(\mathbf{x}, \mathbf{y}))$$

- Other properties: commutative, associative, distributive

Norms on $\mathbb{R}^n$

- Scalar product induces Euclidean norm (sometimes called 2-norm):

$$\|\mathbf{x}\| = \|\mathbf{x}\|_2 = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \sqrt{\sum_{i=1}^{n} x_i^2}$$

- The length of a vector is defined to be its (Euclidean) norm. A unit vector is one of length 1.

- Non-negativity properties also hold for the norm:
  $\forall \mathbf{x} \neq 0: \ \|\mathbf{x}\|^2 > 0 \qquad \|\mathbf{x}\|^2 = 0 \Leftrightarrow \mathbf{x} = 0$

Bases and Transformations

We will look at:

- Linear dependence
- Bases
- Orthogonality
- Change of basis (Linear transformation)
- Matrices and matrix operations

Linear dependence

- Linear combination of vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$:

$$\alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2 + \cdots + \alpha_n \mathbf{x}_n$$

- A set of vectors $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ is linearly dependant if $\mathbf{x}_i \in X$ can be written as a linear combination of the rest, $X \setminus \{\mathbf{x}_i\}$.

Linear dependence (II)

- Geometrically:

- In $\mathbb{R}^n$ it holds that
  - a set of 2 to n vectors *can* be linearly dependant
  - sets of $n + 1$ or more vectors are *always* linearly dependant

Bases

- A basis is a linearly independent set of vectors that spans the "whole space". we can write every vector in our space as linear combination of vectors in that set.
- Every set of n linearly independent vectors in $\mathbb{R}^n$ is a basis of $\mathbb{R}^n$.
- *Orthogonality*: Two non-zero vectors $\mathbf{x}$ and $\mathbf{y}$ are orthogonal if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$.
- A basis is called
  - *orthogonal*, if every basis vector is orthogonal to all other basis vectors
  - *orthonormal*, if additionally all basis vectors have length 1.

Bases (Examples in $\mathbb{R}^2$)

Bases continued

- Standard basis in $\mathbb{R}^n$ (also called unit vectors):

$$\{\mathbf{e}_i \in \mathbb{R}^n \ : \ \mathbf{e}_i = (\underbrace{0, \ldots, 0}_{i-1 \text{ times}}, 1, \underbrace{0, \ldots, 0}_{n-i-1 \text{ times}})\}$$

- We can write a vector in terms of its standard basis,

$$\begin{pmatrix} 4 \\ 7 \\ -3 \end{pmatrix} = 4 \cdot \mathbf{e}_1 + 7 \cdot \mathbf{e}_2 - 3 \cdot \mathbf{e}_3$$

- Important observation: $x_i = \langle \mathbf{e}_i, \mathbf{x} \rangle$, to find the coefficient for a particular basis vector, we project our vector onto it.

Change of basis

- Suppose we have a basis $B = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}, \mathbf{b}_i \in \mathbb{R}^m$ and a vector $\mathbf{x} \in \mathbb{R}^m$ that we would like to represent in terms of B. Note that m and n can be equal, but don't have to.
- To that end we project $\mathbf{x}$ onto all basis vectors:

$$\tilde{\mathbf{x}} = \begin{pmatrix} \langle \mathbf{b}_1, \mathbf{x} \rangle \\ \langle \mathbf{b}_2, \mathbf{x} \rangle \\ \vdots \\ \langle \mathbf{b}_n, \mathbf{x} \rangle \end{pmatrix} = \begin{pmatrix} \mathbf{b}_1 \mathbf{x} \\ \mathbf{b}_2 \mathbf{x} \\ \vdots \\ \mathbf{b}_n \mathbf{x} \end{pmatrix}$$

Change of basis (Example)

Change of basis continued

- Let's rewrite the change of basis:

$$\tilde{\mathbf{x}} = \begin{pmatrix} \mathbf{b}_1\mathbf{x} \\ \mathbf{b}_2\mathbf{x} \\ \vdots \\ \mathbf{b}_n\mathbf{x} \end{pmatrix} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_n \end{pmatrix} \mathbf{x} = \mathbf{B}\mathbf{x}$$

- Voila, we got a $n \times m$ matrix $\mathbf{B}$ times a m-vector $\mathbf{x}$!

- The matrix-vector product written out is thus:
  $\tilde{x}_i = (\mathbf{B}\mathbf{x})_i = \mathbf{b}_i\mathbf{x} = \sum_{j=1}^{m} b_{ij}x_j$

Some properties

- Because of the linearity of the scalar product, this transformation is *linear* as well:
  - $\mathbf{B}(\lambda \cdot \mathbf{x}) = \lambda \cdot \mathbf{B}\mathbf{x}$
  - $\mathbf{B}(\mathbf{x} + \mathbf{y}) = \mathbf{B}\mathbf{x} + \mathbf{B}\mathbf{y}$
- The *identity* transformation / matrix maps a vector to itself:

$$\mathbf{I} = \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_n \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}$$

Matrix Multiplication

- We can now also understand what a matrix-matrix multiplication is. Consider what happens when we change the basis first to B and then from there change the basis to A:
  - $\hat{\mathbf{x}} = \mathbf{A}\mathbf{B}\mathbf{x} = \mathbf{A}(\mathbf{B}\mathbf{x}) = \mathbf{A}\tilde{\mathbf{x}}$

$$\begin{aligned} \hat{\mathbf{x}}_i &= (\mathbf{A}\tilde{\mathbf{x}})_i = \sum_{j=1}^{n} a_{ij}\tilde{x}_j = \sum_{j=1}^{n} a_{ij} \sum_{k=1}^{m} b_{jk} x_k \\ &= \sum_{k=1}^{m} \left(\sum_{j=1}^{n} a_{ij} b_{jk} \right) x_k = \sum_{k=1}^{m} (\mathbf{A}\mathbf{B})_{ik} x_k \end{aligned}$$

- Therefore $\mathbf{A} \cdot \mathbf{B} = \left(\sum_{j=1}^{n} a_{ij} b_{jk} \right)_{ik}$

More matrix properties

- The product of a $l \times n$ and a $n \times m$ matrix is a $l \times m$ matrix.
- The matrix product is associative, distributive, but *not commutative*.
- We will sometimes use the following rule: $(\mathbf{AB}) = \mathbf{BA}$
- Addition, subtraction, and multiplication by a scalar are defined much like for vectors. Matrices are associative, distributive, and commutative regarding these operations.

One more example - Rotation matrices

- Suppose we want to rotate a vector in $\mathbb{R}^2$ around the origin. This amounts to a change of basis, in which the basis vectors are rotated:

$$\mathbf{R} = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}$$

Inverse Transformations - Overview

- Rank of a Matrix
- Inverse Matrix
- Some properties

Rank of a Matrix

- The rank of a matrix is the number of linearly independent rows or columns.
- Examples: $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ has rank 2, but $\begin{pmatrix} 2 & 1 \\ 4 & 2 \end{pmatrix}$ only has rank 1.
- Equivalent to the dimension of the range of the linear transformation.
- A matrix with full rank is called *non-singular*, otherwise it is singular.

Inverse Matrix

- A linear transformation can only have an inverse, if the associated matrix is non-singular.
- The inverse $\mathbf{A}^{-1}$ of a matrix $\mathbf{A}$ is defined as:

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{I} \quad \left(= \mathbf{A}\mathbf{A}^{-1}\right)$$

- We cannot cover here, how the inverse is computed. Nevertheless, it is similar to solving ordinary linear equation systems.

Some properties

- Matrix multiplication $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.
- For orthonormal matrices it holds that $\mathbf{A}^{-1} = \mathbf{A}$.
- For a diagonal matrix $\mathbf{D} = \{d_1, \ldots, d_n\}$:

$$\mathbf{D}^{-1} = \{d_1^{-1}, \ldots, d_n^{-1}\}$$

Determinants

- Determinants are a pretty complex topic; we will only cover basic properties.
- Definition ($\Pi(n) =$ Permutations of $\{1, \ldots, n\}$):

$$\det(\mathbf{A}) = \sum_{\pi \in \Pi(n)} (-1)^{|\pi|} \prod_{i=1}^{n} a_{i\pi(i)}$$

- Some properties:
  - $\det(\mathbf{A}) = 0$ iff $\mathbf{A}$ is singular.
  - $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B}), \qquad \det(\mathbf{A}^{-1}) = \det(\mathbf{A})^{-1}$
  - $\det(\lambda\mathbf{A}) = \lambda^n \det(\mathbf{A})$

Eigendecomposition - Overview

- Eigenvalues and Eigenvectors
- How to compute them?
- Eigendecomposition

Eigenvalues and Eigenvectors

- All non-zero vectors $\mathbf{x}$ for which there is a $\lambda \in \mathbb{R}$ so that

$$\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$$

  are called *eigenvectors* of $\mathbf{A}$. λ are the associated *eigenvalues*.

- If $\mathbf{e}$ is an eigenvector of $\mathbf{A}$, then also $c \cdot \mathbf{e}$ with $c \neq 0$.

- Label eigenvalues $\lambda_1 \geq \lambda_1 \geq \cdots \geq \lambda_n$ with their eigenvectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ (assumed to be unit vectors).

How to compute them?

- Rewrite the definition as

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

- $\det(\mathbf{A} - \lambda \mathbf{I}) = 0$ has to hold, because $\mathbf{A} - \lambda \mathbf{I}$ cannot have full rank.
- This gives a polynomial in λ, the so-called *characteristic polynomial*.
- Find the eigenvalues by finding the roots of that polynomial.
- Find the associated eigenvector by solving linear equation system.

Eigendecomposition

- Every real, square, symmetric matrix $\mathbf{A}$ can be decomposed as:

$$\mathbf{A} = \mathbf{VDV},$$

  where $\mathbf{V}$ is an orthonormal matrix of $\mathbf{A}$'s eigenvectors and $\mathbf{D}$ is a diagonal matrix of the associated eigenvalues.

- The eigendecomposition is essentially a restricted variant of the Singular Value Decomposition.

Some properties

- The determinant of a square matrix is the product of its eigenvalues: $\det(\mathbf{A}) = \lambda_1 \cdot \ldots \cdot \lambda_n$.
- A square matrix is singular if it has some eigenvalues of value 0.
- A square matrix $\mathbf{A}$ is called positive (semi-)definite if all of its eigenvalues are positive (non-negative). Equivalent criterion: $\mathbf{x}\mathbf{A}\mathbf{x} > 0 \quad \forall \mathbf{x} \qquad (\geq$ if semi-definite$)$.

Singular Value Decomposition (SVD)

- Suppose $\mathbf{A} \in \mathbb{R}^{m \times n}$. Then a $\lambda \geq 0$ is called a *singular value* of $\mathbf{A}$, if there exist $\mathbf{u} \in \mathbb{R}^m$ and $\mathbf{v} \in \mathbb{R}^n$ such that

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{u} \quad \text{and} \quad \mathbf{A}\mathbf{u} = \lambda\mathbf{v}$$

- We can decompose *any* matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ as

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V},$$

where $\mathbf{U} \in \mathbb{R}^{m \times m}$ and $\mathbf{V} \in \mathbb{R}^{n \times n}$ are orthonormal and Σ is a diagonal matrix of the singular values.

Relation to Eigendecomposition

- The columns of $\mathbf{U}$ are the eigenvectors of $\mathbf{A}\mathbf{A}$, and the (non-zero) singular values of $\mathbf{A}$ are the square roots of the (non-zero) eigenvalues of $\mathbf{A}\mathbf{A}$.
- If $\mathbf{A} \in \mathbb{R}^{m \times n}$ ($m << n$) is a matrix of m i.i.d. samples from a n-dimensional Gaussian distribution, we can use the SVD of $\mathbf{A}$ to compute the eigenvectors and eigenvalues of the covariance matrix without building it explicitly.