Optimization on Manifolds, Appendix: Linear Algebra

Ibrahim Akbar

April 8, 2018

1 Definitions

This is an abbreviation of the material provided in the Appendix A1 of *Optimization on Manifolds* and will provide definitions of various terms that appear throughout the notes. For proofs and further insight into the concepts refer to the Linear Algebra notes. Note that the conventions used between separate texts may not be the same. I will make sure to be as explicit as possible unless it is obvious from context.

Let A, B be $n \times n$ matrices, the **commutator** of A, B is: $[A, B] = AB - BA$.
The commutator of two symmetric/skew-symmetric matrices is symmetric and skew-symmetric for a symmetric and skew-symmetric matrix.

The **trace** of $A \in \mathbb{R}^{n \times p}$ is

$$tr(A) = \sum_{i=1}^{\min n,p} = (A)_{ii}$$

Properties of Trace

- $tr(A) = tr(A^T$
- $tr(AB) = tr(BA)$
- $tr([A, B]) = 0$
- $B^T = -B \Rightarrow tr(B) = 0$
- $(A^T = A \wedge B^T = -B) \Rightarrow tr(AB) = 0$

Let A be an $n \times n$ matrix. A is **invertible** if there exists a matrix, B, such that $AB = I_n$.

A matrix A is **orthonormal** if $A^T A = I$. If A is a square orthonormal matrix then it is termed **orthogonal**.

A **vector space**, $\mathcal{V}$, is a collection over a field, $\mathbb{F}$ that is endowed with the operators, $+ : V \times V \to \mathbb{V}$ and $\cdot : \mathbb{F} \times \mathcal{V} \to \mathcal{V}$.
Note any n-dimensional real vector space $\mathcal{E}$ is isomorphic to $\mathbb{R}^n$. A diffeomorphism requires determining a basis for $\mathcal{E}$ but may be computationally intractable, given it is a continuous space. Therefore the material considers abstract vector spaces and only finite-dimensional vector spaces over $\mathbb{R}$. To understand this statement better refer to the *Morphisms and Functions* Appendix.

A **norm** is an operator that maps from the vector space to the real field, $\|\cdot\| : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$.

Properties of Norm

- $\|\mathbf{v}\| = 0 \iff \mathbf{v} = 0 \ \forall \mathbf{v} \in \mathcal{V}$
- $\|\alpha \mathbf{v}\| = |\alpha| \, \|\mathbf{v}\|$
- $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$

This mapping allows for the definition of a **normed vector space**, $\mathcal{V}$, which is endowed with a norm.

A **linear transformation**, $\mathcal{T} : \mathcal{V} \to U$, is linear if $\mathcal{T}(\alpha \mathbf{v}_1 + \beta \mathbf{v}_2) = \alpha T(\mathbf{v}_1) + \beta T(\mathbf{v}_2) \;\; \forall \mathbf{v}_1, \mathbf{v}_2 \in \mathcal{V}$.

We can define the set $\mathcal{L}(\mathcal{E};\mathcal{F})$ as all operators $A : \mathcal{E} \to \mathcal{F}$ which is also a vector space. Furthermore, if we let $\mathcal{E}, \mathcal{F}, \mathcal{G}$ be normed vector spaces we can define norms as: $\|\cdot\|_{\mathcal{L}(\mathcal{E};\mathcal{F})}, \|\cdot\|_{\mathcal{L}(\mathcal{E};\mathcal{G})}, \|\cdot\|_{\mathcal{L}(\mathcal{F};\mathcal{G})}$.
These are **mutually consistent** if for all $A \in \mathcal{L}(\mathcal{E};\mathcal{F})$ and $B \in \mathcal{L}(\mathcal{F};\mathcal{G})$:

$$\|B \circ A\|_{\mathcal{L}(\mathcal{E};\mathcal{G})} \leq \|A\|_{\mathcal{L}(\mathcal{E};\mathcal{F})} \|B\|_{\mathcal{L}(\mathcal{F};\mathcal{G})}$$

A **consistent** or **submultiplicative norm** is mutually consistent with itself.

Furthermore the **operator norm** or **induced norm** of $A \in \mathcal{L}(\mathcal{E};\mathcal{F})$ is

$$\|A\|_{\mathcal{L}(\mathcal{E};\mathcal{F})} = \sup_{x \in \mathcal{E}} \frac{\|Ax\|}{\|x\|}$$

Operator norms are mutually consistent.

A **bilinear** operator is a mapping $A : \mathcal{E} \times \mathcal{F} \to \mathcal{G}$ over normed vector spaces such that it is linear operator with respect to both spaces $\mathcal{E}, \mathcal{F}$. We will denote the space of such operators as $\mathcal{L}(\mathcal{E}, \mathcal{F}; \mathcal{G})$ or $\mathcal{L}_2(\mathcal{E}; \mathcal{F}$ if the two input spaces are identical. A simple example of a bilinear operator would be matrix multiplication. A bilinear operator is denoted as **symmetric** if $A[x, y] = A[y, x]$.

A symmetric bilinear operator $A \in \mathcal{L}_2(\mathcal{E}; \mathbb{R})$ is **positive-definite** if $A[\mathbf{x}, \mathbf{x}] > 0 \;\; \forall \mathbf{x} \in \mathcal{E}$.

A **Euclidean space** is a finite-dimensional vector space endowed with an inner product (bilinear, symmetric, positive definite operator) $\langle \cdot, \cdot \rangle$.

An **orthonormal basis** of an n-dimensional Euclidean space $\mathcal{E}$ is a sequence $(e_1, e_2, \ldots, e_n)$ of elements such that,

$$\langle e_i, e_j \rangle = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

An operator $\mathcal{T} : \mathcal{E} \to \mathcal{E}$ is **symmetric** if $\langle \mathcal{T}[x], y \rangle = \langle x, \mathcal{T}[y] \rangle$.

Given an operator $\mathcal{T} : \mathcal{E} \to \mathcal{F}$ maps between two Euclidean spaces, the **adjoint** of T, denoted $\mathcal{T}^* : \mathcal{F} \to \mathcal{E}$ is such that $\langle \mathcal{T}[x], y \rangle = \langle x, \mathcal{T}^*[y] \rangle \;\; \forall x \in \mathcal{E} \;\; y \in \mathcal{F}$.

The **kernel** of an operator is $ker(\mathcal{T}) := \{x \in \mathcal{E} \mid \mathcal{T}[x] = 0\}$.

The **range** of an operator is $range(\mathcal{T}) := \{\mathcal{T}[x] \mid x \in \mathcal{E}\}$.

The **orthogonal compliment** of a subspace, $\mathcal{S} \subset \mathcal{E}$ is denoted $\mathcal{S}^\perp := \{x \mid \langle x, y \rangle = 0 \;\; y \in \mathcal{S}\}$.

Given that $x \in \mathcal{E}$ it can be uniquely decomposed as $x = x_1 + x_2$ where $x_1 \in \mathcal{S}$ and $x_2 \in \mathcal{S}^\perp$. x_1, x_2 are **orthogonal projections** of x into their respective space. Let $\Pi_{\mathcal{S}}(x)$ denote the orthogonal projection of x onto $\mathcal{S}$.

The **Moore-Penrose Inverse** also known as the pseudo-inverse is used to determine the orthogonal projection that satisfies the least squares problem.

$$T^\dagger = (T^* T)^{-1} T^* y$$

The **Euclidean norm** on a Euclidean space, $\mathcal{E}$, is $\|x\|_2 := \sqrt{\langle x, x \rangle}$, whereas on $\mathbb{R}^n$ it is $\|x\|_2 := \sqrt{x^T x}$.

Given an inner product space $\mathbb{R}^{n \times p}$ that has the inner product $\langle X, Y \rangle := tr(X^T Y)$ the Euclidean norm is the **Forbenius norm** denoted by $\|A\|_F := \left(\sum_{i,j} (A)^2_{i,j} \right)^{\frac{1}{2}}$.

An operator norm on $\mathbb{R}^{n \times n}$ with $\mathbb{R}^n$ has the Euclidean norm is known as the **spectral norm**

$$\|A\|_2 := \sqrt{\lambda_{max}(A^T A)}$$