9.520 Math Camp 2010

Functional Analysis Review

Andre Wibisono

February 8, 2010

These notes present a brief summary of some of the basic definitions from functional analysis that we will need in this class. Although we will not actually use all the material presented here, these notes can serve as a useful reference for the definitions and relationships between properties that we encounter in the class later.

These notes are organized as follows. We first discuss finite dimensional vector spaces and additional structures that we can impose on them. We then introduce Hilbert spaces, which play a central role in this class, and talk about the subtleties that we have to deal with when working in infinite dimension. Finally, we talk about matrices and linear operators, in the context of linear mappings between vector spaces. We will see how matrices represent linear functions between finite dimensional vector spaces, and develop a parallel theory on linear operators between general Hilbert spaces.

Throughout these notes, we assume that we are working with the base field $\mathbb{R}$.

1 Structures on Vector Spaces

A **vector space** V is a set with a linear structure. This means we can add elements of the vector space or multiply elements by scalars (real numbers) to obtain another element. A familiar example of a vector space is $\mathbb{R}^n$. Given $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ in $\mathbb{R}^n$, we can form a new vector $x + y = (x_1 + y_1, \dots, x_n + y_n) \in \mathbb{R}^n$. Similarly, given $r \in \mathbb{R}$, we can form $rx = (rx_1, \dots, rx_n) \in \mathbb{R}^n$.

Every vector space has a basis. A subset $B = \{v_1, \dots, v_n\}$ of V is called a **basis** if every vector $v \in V$ can be expressed uniquely as a linear combination $v = c_1 v_1 + \cdots + c_m v_m$ for some constants $c_1, \dots, c_m \in \mathbb{R}$. The cardinality (number of elements) of V is called the **dimension** of V. This notion of dimension is well defined because while there is no canonical way to choose a basis, all bases of V have the same cardinality. For example, the standard basis on $\mathbb{R}^n$ is $e_1 = (1, 0, \dots, 0), e_2 = (0, 1, 0, \dots, 0), \dots, e_n = (0, \dots, 0, 1)$. This shows that $\mathbb{R}^n$ is an n-dimensional vector space, in accordance with the notation. In this section we will be working with finite dimensional vector spaces only.

We note that any two finite dimensional vector spaces over $\mathbb{R}$ are isomorphic, since a bijection between the bases can be extended linearly to be an isomorphism between the two vector spaces. Hence, up to isomorphism, for every $n \in \mathbb{N}$ there is only one n-dimensional vector space, which is $\mathbb{R}^n$. However, vector spaces can also have extra structures that distinguish them from each other, as we shall explore now.

A **distance (metric)** on V is a function $d\colon V \times V \to \mathbb{R}$ satisfying:

1. (positivity) $d(v, w) \geq 0$ for all $v, w \in V$, and $d(v, w) = 0$ if and only if $v = w$.
2. (symmetry) $d(v, w) = d(w, v)$ for all $v, w \in V$.
3. (triangle inequality) $d(v, w) \leq d(v, x) + d(x, w)$ for all $v, w, x \in V$.

The standard distance function on $\mathbb{R}^n$ is given by $d(x, y) = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}$. Note that the notion of metric does not require a linear structure, or any other structure, on V; a metric can be defined on any set.

A similar concept that requires a linear structure on V is **norm**, which measures the "length" of vectors in V. Formally, a norm is a function $\|\cdot\|\colon V \to \mathbb{R}$ that satisfies the following three properties:

1. (positivity) $\|v\| \geq 0$ for all $v \in V$, and $\|v\| = 0$ if and only if $v = 0$.
2. (homogeneity) $\|rv\| = |r|\|v\|$ for all $r \in \mathbb{R}$ and $v \in V$.
3. (subadditivity) $\|v + w\| \leq \|v\| + \|w\|$ for all $v, w \in V$.

For example, the standard norm on $\mathbb{R}^n$ is $\|x\|_2 = \sqrt{x_1^2 + \cdots + x_n^2}$, which is also called the ℓ_2-norm. Also of interest is the ℓ_1-norm $\|x\|_1 = |x_1| + \cdots + |x_n|$, which we will study later in this class in relation to sparsity-based algorithms. We can also generalize these examples to any $p \geq 1$ to obtain the ℓ_p-norm, but we will not do that here.

Given a normed vector space $(V, \|\cdot\|)$, we can define the **distance (metric) function** on V to be $d(v, w) = \|v - w\|$. For example, the ℓ_2-norm on $\mathbb{R}^n$ gives the standard distance function

$$d(x, y) = \|x - y\|_2 = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2},$$

while the ℓ_1-norm on $\mathbb{R}^n$ gives the Manhattan/taxicab distance,

$$d(x, y) = \|x - y\|_1 = |x_1 - y_1| + \cdots + |x_n - y_n|.$$

As a side remark, we note that all norms on a finite dimensional vector space V are **equivalent**. This means that for any two norms μ and ν on V, there exist positive constants C_1 and C_2 such that for all $v \in V$, $C_1\mu(v) \leq \nu(v) \leq C_2\mu(v)$. In particular, continuity or convergence with respect to one norm implies continuity or convergence with respect to any other norms in a finite dimensional vector space. For example, on $\mathbb{R}^n$ we have the inequality $\|x\|_1/\sqrt{n} \leq \|x\|_2 \leq \|x\|_1$.

Another structure that we can introduce to a vector space is the inner product. An **inner product** on V is a function $\langle \cdot, \cdot \rangle\colon V \times V \to \mathbb{R}$ that satisfies the following properties:

1. (symmetry) $\langle v, w \rangle = \langle w, v \rangle$ for all $v, w \in V$.
2. (linearity) $\langle r_1 v_1 + r_2 v_2, w \rangle = r_1 \langle v_1, w \rangle + r_2 \langle v_2, w \rangle$ for all $r_1, r_2 \in \mathbb{R}$ and $v_1, v_2, w \in V$.
3. (positive-definiteness) $\langle v, v \rangle \geq 0$ for all $v \in V$, and $\langle v, v \rangle = 0$ if and only if $v = 0$.

For example, the standard inner product on $\mathbb{R}^n$ is $\langle x, y \rangle = x_1 y_1 + \cdots + x_n y_n$, which is also known as the *dot product*, written $x \cdot y$.

Given an inner product space $(V, \langle \cdot, \cdot \rangle)$, we can define the norm of $v \in V$ to be $\|v\| = \sqrt{\langle v, v \rangle}$. It is easy to check that this definition satisfies the axioms for a norm listed above. On the other hand, not every norm arises from an inner product. The necessary and sufficient condition that has to be satisfied for a norm to be induced by an inner product is the **paralellogram law**:

$$\|v+w\|^2 + \|v-w\|^2 = 2\|v\|^2 + 2\|w\|^2.$$

If the parallelogram law is satisfied, then the inner product can be defined by **polarization identity**:

$$\langle v, w\rangle = \frac{1}{4}\big(\|v+w\|^2 - \|v-w\|^2\big).$$

For example, you can check that the ℓ_2-norm on $\mathbb{R}^n$ is induced by the standard inner product, while the ℓ_1-norm is not induced by an inner product since it does not satisfy the parallelogram law.

A very important result involving inner product is the following **Cauchy-Schwarz inequality**:

$$\langle v, w\rangle \le \|v\|\|w\| \quad \text{for all } v, w \in V.$$

Inner product also allows us to talk about orthogonality. Two vectors v and w in V are said to be **orthogonal** if $\langle v, w\rangle = 0$. In particular, an **orthonormal basis** is a basis $v_1, \dots, v_n$ that is orthogonal ($\langle v_i, v_j\rangle = 0$ for $i \neq j$) and normalized ($\langle v_i, v_i\rangle = 1$). Given an orthonormal basis $v_1, \dots, v_n$, the decomposition of $v \in V$ in terms of this basis has the special form

$$v = \sum_{i=1}^{n} \langle v, v_n\rangle v_n.$$

For example, the standard basis vectors $e_1, \dots, e_n$ form an orthonormal basis of $\mathbb{R}^n$. In general, a basis $v_1, \dots, v_n$ can be orthonormalized using the Gram-Schmidt process.

Given a subspace W of an inner product space V, we can define the **orthogonal complement** of W to be the set of all vectors in V that are orthogonal to W,

$$W^\perp = \{v \in V \mid \langle v, w\rangle = 0 \text{ for all } w \in W\}.$$

If V is finite dimensional, then we have the **orthogonal decomposition** $V = W \oplus W^\perp$. This means every vector $v \in V$ can be decomposed uniquely into $v = w + w'$, where $w \in W$ and $w' \in W^\perp$. The vector w is called the **projection** of v on W, and represents the unique vector in W that is closest to v.

2 Hilbert Space

A central object of study in this class, also in the development of learning theory in general, is a Hilbert space – a complete inner product space – and that is the theme of this section. The difficulty arises from the possibility of working in infinite dimension, for example when we are dealing with function spaces. Most of the discussion in Section 1 carry over easily to the case of infinite dimensional vector spaces, but we have to be a bit careful about the notion of a basis, since now we have to deal with infinite sums. In particular, we will only be concerned with Hilbert spaces $\mathcal{H}$ which have countable orthonormal basis $(v_n)_{n=1}^{\infty}$, so that we can write every element $v \in \mathcal{H}$ as

$$v = \sum_{n=1}^{\infty} \langle v, v_n\rangle v_n.$$

We will now talk about a few concepts that will allow us to make sense of these properties.

We first discuss Cauchy sequence and completeness. Recall that a sequence $(v_n)_{n\in\mathbb{N}}$ in a normed space V **converges** to $v \in V$ if $\|v_n - v\| \to 0$ as $n \to \infty$, or equivalently, if for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that $\|v_n - v\| < \epsilon$ whenever $n \geq N$. Intuitively, this means that v_n becomes arbitrarily close to v as we go further down the sequence. A similar condition on a sequence is Cauchy. A sequence $(v_n)_{n\in\mathbb{N}}$ in V is a **Cauchy sequence** if the distance between any pair of elements in the sequence becomes arbitrarily small as we go further in the sequence. More formally, $(v_n)_{n\in\mathbb{N}}$ is a Cauchy sequence if for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that $\|v_m - v_n\| < \epsilon$ whenever $m, n \geq N$. Clearly every convergent sequence is Cauchy, by the triangle inequality, but the converse needs not be true.

A normed vector space is **complete** if every Cauchy sequence converges. Intuitively, this means that there are no "holes" in the space. For example, $\mathbb{Q}$ is not complete since it is missing the irrationals. More concretely, the sequence $1.4142, 1.41421, 1.414213, \dots$ converges to $\sqrt{2}$ in $\mathbb{R}$, but $\sqrt{2} \notin \mathbb{Q}$. On the other hand, $\mathbb{R}$ is complete by definition (in fact, $\mathbb{R}$ is the **completion** of $\mathbb{Q}$), and it can be shown that $\mathbb{R}^n$ is complete for every $n \in \mathbb{N}$. Moreover, every finite dimensional normed vector space (over $\mathbb{R}$) is complete. This is because, as we saw in Section 1, every n-dimensional real vector space V is isomorphic to $\mathbb{R}^n$, and any two norms in $\mathbb{R}^n$ are equivalent. Therefore, V is complete if and only if $\mathbb{R}^n$ is complete under the standard norm, which it is.

We are now ready to define Hilbert spaces.

Definition. A **Hilbert space** is a complete inner product space.

The remark at the end of the previous paragraph shows that $\mathbb{R}^n$ and any finite dimensional inner product space are examples of Hilbert spaces. The archetypical example of an infinite dimensional Hilbert space is the space of square-summable sequences,

$$\ell_2 = \{(a_n)_{n=1}^{\infty} \mid a_n \in \mathbb{R},\ \sum_{n=1}^{\infty} a_n^2 < \infty\},$$

where addition and scalar multiplication are defined componentwise, and inner product is defined by $\langle a, b\rangle = \sum_{n=1}^{\infty} a_n b_n$. We can also consider the continuous analogue of ℓ_2, namely, the space of square-integrable functions

$$L_2([0,1]) = \{f \colon [0,1] \to \mathbb{R} \mid \int_0^1 f(x)^2 \, dx < \infty\},$$

where the integral is Lebegue integration with respect to the Lebesgue measure dx on $[0,1]$. Addition and scalar multiplication on this space are defined pointwise, and inner product is given by $\langle f, g\rangle_{L_2} = \int_0^1 f(x)g(x)\, dx$. With this structure, $L_2([0,1])$ is also an infinite dimensional Hilbert space.

A Hilbert space always has an orthonormal basis, but it might be uncountable. Typically we are only concerned with Hilbert spaces with countable orthonormal basis, and a natural condition that can be imposed to ensure this property is separability. Intuitively, a space is separable if it can be approximated by a countable subset of it, in a sense to be made precise shortly. For example, $\mathbb{R}$ can be approximated by $\mathbb{Q}$, which is countable, so $\mathbb{R}$ is separable.

First, recall that given a topological space X, a subset Y is **dense** in X if $\overline{Y} = X$. Informally, Y is dense in X every point $x \in X$ is a limit of a sequence of points $(y_n)_{n\in\mathbb{N}}$ in Y. For example, $\mathbb{Q}$ is dense in $\mathbb{R}$ since every real number can be approximated by its truncated decimal representations, and $\mathbb{Q}^n$ is dense in $\mathbb{R}^n$ for every $n \in \mathbb{N}$. As a further example, Weierstrass approximation theorem states that the space of polynomials is dense in the space of continuous functions on a compact domain, under the supremum norm.

A topological space X is **separable** if it has a countable dense subset. In some sense, separability is a limitation on the size of X, since this condition means that X can be approximated by a countable subset of it, even if X itself is uncountable. As we mentioned in the example above, $\mathbb{Q}$ is a countable dense subset of $\mathbb{R}$, so $\mathbb{R}$ is separable. More generally, $\mathbb{Q}^n$ is a countable dense subset of $\mathbb{R}^n$, so $\mathbb{R}^n$ is separable for any $n \in \mathbb{N}$.

Separability is an essential condition because we have the following result.

Theorem. A Hilbert space has a countable orthonormal basis if and only if it is separable.

In this class, whenever we encounter Hilbert spaces we will always assume that they are separable, so we can work with countable orthonormal basis. Given a countable orthonormal basis $(v_n)_{n\in\mathbb{N}}$ of a Hilbert space $\mathcal{H}$, we can write the decomposition

$$v = \sum_{n=1}^{\infty} \langle v, v_n \rangle v_n \quad \text{for all } v \in \mathcal{H}.$$

For example, the spaces ℓ_2 and $L_2([0,1])$ from the examples above are separable Hilbert spaces. An orthonormal basis for ℓ_2 is given by $(e_n)_{n\in\mathbb{N}}$, where e_n is an ℓ_2 sequence that is 1 at the n-th position and 0 everywhere else. An orthonormal basis for $L_2([0,1])$ is given by the functions

$$\{1, 2\sin 2\pi nx, 2\cos 2\pi nx \mid n \in \mathbb{N}\},$$

which is precisely the Fourier basis. In fact, just as every finite dimensional Hilbert space is isomorphic to $\mathbb{R}^n$, it is also true that every infinite dimensional separable Hilbert space is isomorphic to ℓ_2, simply by linearly extending the bijection between the orthonormal basis.

Since a Hilbert space is also an inner product space, we can still talk about the orthogonal complement, as in Section 1. Given a Hilbert space $\mathcal{H}$ and a subspace V, the orthogonal complement $V^\perp$ is always a closed subspace of $\mathcal{H}$ (note that every finite dimensional subspace of a normed space is always closed, but if the subspace is infinite dimensional it can be not closed). In addition, if V is closed in $\mathcal{H}$, then we again have the **orthogonal decomposition** $\mathcal{H} = V \oplus V^\perp$.

Remarks

We now add a few remarks about the development of the concepts that we introduced in the previous section.

Remark 1. We noted that any finite dimensional normed vector space is complete. The situation is more interesting in infinite dimension. Consider the space $\mathcal{C}([0,1])$ of continuous functions $f\colon [0,1] \to \mathbb{R}$ with elementwise addition and scalar multiplication. This space is complete under the supremum norm $\|f\|_\infty = \sup_{0\le x\le 1} |f(x)|$ (sketch of proof: given a Cauchy sequence $(f_n)_{n\in\mathbb{N}}$, show that for each x the sequence $(f_n(x))_{n\in\mathbb{N}}$ is Cauchy in $\mathbb{R}$, and hence converges; form a function f whose values are the pointwise limit in $\mathbb{R}$, and show that the original sequence converges to $f \in \mathcal{C}([0,1])$). On the other hand, $\mathcal{C}([0,1])$ is *not* complete under the L_1-norm, $\|f\|_1 = \int_0^1 |f(x)|\,dx$. As a counter example, take f_n to be the function that takes the value 0 for $x = 0$, value 1 for $x \in [1/n, 1]$, and linear in between. It is easy to check that $(f_n)_{n\in\mathbb{N}}$ is a Cauchy sequence, but it converges to the function

$$f(x) = \begin{cases} 0, & \text{if } x = 0, \\ 1, & \text{if } 0 < x \le 1, \end{cases}$$

which is not continuous.

Remark 2. The construction of $L_2([0,1])$ above is not technically correct. This is because two functions f and g can have the same Lebesgue integral as long as they differ on a set of measure zero (for instance, on countably many points). Therefore, the norm given above is not well-defined since we can have nonzero functions f with $\|f\| = 0$. To take care of this problem, we declare two functions f and g to be equivalent, $f \sim g$, if $\int_0^1 (f(x) - g(x))^2 \, dx = 0$, and define $L_2([0,1])$ to be the quotient space

$$L_2([0,1]) = \{f \colon [0,1] \to \mathbb{R} \mid \int_0^1 f(x)^2 \, dx < \infty\} / \sim .$$

This means elements of $L_2([0,1])$ are not functions, but rather equivalence classes of functions. In particular, we cannot evaluate "functions" in $L_2([0,1])$ on points. We will see in the next lecture a nice function space in which we can evaluate functions at points.

Remark 3. As a side note, a **Banach space** is a complete normed space. For example, $\mathcal{C}([0,1])$ with the supremum norm is a Banach space. As another example of a Banach space, consider the space of absolutely-summable sequences,

$$\ell_1 = \{(a_n)_{n=1}^{\infty} \mid a_n \in \mathbb{R}, \sum_{n=1}^{\infty} |a_n| < \infty\},$$

where addition and scalar multiplication are again defined componentwise, and norm is given by $\|a\|_1 = \sum_{n=1}^{\infty} |a_n|$.

3 Matrices and Operators

In addition to talking about vector spaces, we can also talk about operators on those spaces. A **linear operator** is a function $L \colon V \to W$ between two vector spaces that preserves the linear structure. In finite dimension, every linear operator can be represented by a matrix by choosing a basis in both the domain and the range, i.e. by working in coordinates. For this reason we focus the first part of our discussion on matrices.

If V is n-dimensional and W is m-dimensional, then a linear map $L \colon V \to W$ is represented by an $m \times n$ matrix A whose columns are the values of L applied to the basis of V. The **rank** of A is the dimension of the image of A, and the **nullity** of A is the dimension of the kernel of A. The **rank-nullity theorem** states that $\text{rank}(A) + \text{nullity}(A) = m$, the dimension of the domain of A. Also note that the transpose of A is an $n \times m$ matrix $A^\top$ satisfying

$$\langle Av, w \rangle_{\mathbb{R}^m} = (Av)^\top w = v^\top A^\top w = \langle v, A^\top w \rangle_{\mathbb{R}^n}$$

for all $v \in \mathbb{R}^n$ and $w \in \mathbb{R}^m$.

Let A be an $n \times n$ matrix with real entries. Recall that an **eigenvalue** $\lambda \in \mathbb{R}$ of A is a solution to the equation $Av = \lambda v$ for some nonzero vector $v \in \mathbb{R}^n$, and v is the **eigenvector** of A corresponding to λ. If A is symmetric, i.e. $A^\top = A$, then the eigenvalues of A are real. Moreover, in this case the **spectral theorem** tells us that there is an orthonormal basis of $\mathbb{R}^n$ consisting of the eigenvectors of A. Let $v_1, \dots, v_n$ be this orthonormal basis of eigenvectors, and let $\lambda_1, \dots, \lambda_n$ be the corresponding eigenvalues. Then we can write

$$A = \sum_{i=1}^{n} \lambda_i v_i v_i^\top,$$

which is called the **eigendecomposition** of A. We can also write this as

$$A = V \Lambda V^\top,$$

where V is the $n \times n$ matrix with columns v_i, and Λ is the $n \times n$ diagonal matrix with entries λ_i. The orthonormality of $v_1, \ldots, v_n$ makes V an orthogonal matrix, i.e. $V^{-1} = V^\top$.

A symmetric $n \times n$ matrix A is **positive definite** if $v^\top Av > 0$ for all nonzero vectors $v \in \mathbb{R}^n$. A is **positive semidefinite** if the inequality is not strict (i.e. ≥ 0). A positive definite (resp. positive semidefinite) matrix A has positive (resp. nonnegative) eigenvalues.

Another method for decomposing a matrix is the **singular value decomposition** (SVD). Given an $m \times n$ real matrix A, the SVD of A is the factorization

$$A = U\Sigma V^\top,$$

where U is an $m \times m$ orthogonal matrix ($U^\top U = I$), Σ is an $m \times n$ diagonal matrix, and V is an $n \times n$ orthogonal matrix ($V^\top V = I$). The columns $u_1, \ldots, u_m$ of U form an orthonormal basis of $\mathbb{R}^m$, and the columns $v_1, \ldots, v_n$ of V form an orthonormal basis of $\mathbb{R}^n$. The diagonal elements $\sigma_1, \ldots, \sigma_{\min\{m,n\}}$ in Σ are nonnegative and called the **singular values** of A. This factorization corresponds to the decomposition

$$A = \sum_{i=1}^{\min\{m,n\}} \sigma_i u_i v_i^\top.$$

This decomposition shows the relations between σ_i, u_i, and v_i more clearly: for $1 \leq i \leq \min\{m, n\}$,

$$\begin{aligned} Av_i &= \sigma_i u_i & \qquad AA^\top u_i &= \sigma_i^2 u_i \\ A^\top u_i &= \sigma_i v_i & \qquad A^\top Av_i &= \sigma_i^2 v_i \end{aligned}$$

This means the u_i's are eigenvectors of $AA^\top$ with corresponding eigenvalues σ_i^2, and the v_i's are eigenvectors of $A^\top A$, also with corresponding eigenvalues σ_i^2.

Given an $m \times n$ matrix A, we can define the **spectral norm** of A to be largest singular value of A,

$$\|A\|_{\text{spec}} = \sigma_{\max}(A) = \sqrt{\lambda_{\max}(AA^\top)} = \sqrt{\lambda_{\max}(A^\top A)}.$$

Another common norm on A is the **Frobenius norm**,

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2} = \sqrt{\text{trace}(AA^\top)} = \sqrt{\text{trace}(A^\top A)} = \sqrt{\sum_{i=1}^{\min\{m,n\}} \sigma_i^2}.$$

However, since the space of all matrices can be identified with $\mathbb{R}^{m\times n}$, the discussion in Section 1 still holds and all norms on A are equivalent.

In general, let $\mathcal{H}_1$ and $\mathcal{H}_2$ be (possibly infinite dimensional) Hilbert spaces. A linear operator $L \colon \mathcal{H}_1 \to \mathcal{H}_2$ is continuous if and only if it is **bounded**: there exists a constant $C > 0$ such that

$$\|Lv\|_{\mathcal{H}_2} \leq C\|v\|_{\mathcal{H}_1} \;\text{ for all } v \in \mathcal{H}_1.$$

In other words, a continuous linear operator maps bounded sets into bounded sets[1]. The smallest such C is called the **operator norm** of L. If $\mathcal{H}_1$ and $\mathcal{H}_2$ are finite dimensional, then every linear operator from $\mathcal{H}_1$ to $\mathcal{H}_2$ is continuous, but in general this needs not be the case.

A linear **functional** is a linear operator $L \colon \mathcal{H} \to \mathbb{R}$. An important result about linear functionals is **Riesz' representation theorem**: for every bounded linear functional $L \colon \mathcal{H} \to \mathbb{R}$ there exists

[1] Note that "bounded" here does *not* mean $\|Lv\|_{\mathcal{H}_2} \leq C$ for all $v \in \mathcal{H}_1$. In fact, it is easy to convince yourself that this condition cannot be true for a linear operator L.

a unique element $v \in \mathcal{H}$ such that $Lw = \langle v, w \rangle_{\mathcal{H}}$ for all $w \in \mathcal{H}$. That is, every continuous linear functional can be "represented" by an element $v \in \mathcal{H}$.

The generalization of the concept of transpose is adjoint. Given a bounded linear operator $L\colon \mathcal{H}_1 \to \mathcal{H}_2$, the **adjoint** of L is the unique bounded linear operator $L^*\colon \mathcal{H}_2 \to \mathcal{H}_1$ satisfying

$$\langle Lv, w \rangle_{\mathcal{H}_2} = \langle v, L^* w \rangle_{\mathcal{H}_1} \quad \text{for all } v \in \mathcal{H}_1, w \in \mathcal{H}_2.$$

An operator $L\colon \mathcal{H} \to \mathcal{H}$ is **self-adjoint** if $L^* = L$. Self-adjoint operators have real eigenvalues.

A self-adjoint linear operator L is **positive definite** if $\langle Lv, v \rangle_{\mathcal{H}} > 0$ for all $v \in \mathcal{H}$, $v \neq 0$. Similarly, L is **positive** (or **positive semidefinite**) if $\langle Lv, v \rangle_{\mathcal{H}} \geq 0$ for all $v \in \mathcal{H}$, $v \neq 0$. A positive definite (resp. positive semidefinite) operator has positive (resp. nonnegative) eigenvalues.

A bounded linear operator $L\colon \mathcal{H}_1 \to \mathcal{H}_2$ is a **finite rank operator** if the range of L in $\mathcal{H}_2$ is finite dimensional. Finally, perhaps the most interesting class of operators is compact operators, which are direct generalization of matrices to the theory of operators. A bounded linear operator $L\colon \mathcal{H}_1 \to \mathcal{H}_2$ is **compact** if the image of unit ball in $\mathcal{H}_1$ is pre-compact in $\mathcal{H}_2$, i.e. has compact closure. It can be shown that the space of compact operators is the closure of the space of finite rank operators.

Compact operators are nice because we also have the **spectral theorem**: Let $L\colon \mathcal{H} \to \mathcal{H}$ be a compact self-adjoint operator on a separable Hilbert space $\mathcal{H}$. Then there exists a countable orthonormal basis of $\mathcal{H}$ consisting of the eigenfunctions v_i of L,

$$Lv_i = \lambda_i v_i,$$

and the only possible limit point of λ_i as $i \to \infty$ is 0. As in the case of symmetric matrix, we can write the **eigendecomposition** of L:

$$L = \sum_{i=1}^{\infty} \lambda_i \langle v_i, \cdot \rangle_{\mathcal{H}} v_i.$$

When $L\colon \mathcal{H}_1 \to \mathcal{H}_2$ is a compact but not self-adjoint operator, we can still perform the **singular value decomposition** of L:

$$L = \sum_{i=1}^{\infty} \sigma_i \langle v_i, \cdot \rangle_{\mathcal{H}_1} u_i,$$

where $(u_i)_{i \in \mathbb{N}}$ is an orthonormal basis of $\overline{\text{range}(L)} \subseteq \mathcal{H}_2$, $(v_i)_{i \in \mathbb{N}}$ is an orthonormal basis of $\overline{\text{range}(L^*)} \subseteq \mathcal{H}_1$, and $(\sigma_i)_{i \in \mathbb{N}}$ consists of the nonzero singular values of L. Another way of looking at this decomposition is via the following singular system:

$$\begin{aligned} Lv_i &= \sigma_i u_i & \qquad\qquad LL^* u_i &= \sigma_i^2 u_i \\ L^* u_i &= \sigma_i v_i & \qquad\qquad L^* L v_i &= \sigma_i^2 v_i \end{aligned}$$

From the relations above, it is clear that u_i is an eigenfunction of LL^* with eigenvalue σ_i^2, and similarly v_i is an eigenfunction of L^*L with eigenvalue σ_i^2.