1

Preliminaries

In this Chapter we collect some definitions and results which will be useful later in the book. Virtually no proofs are given, but we provide references to works where these matters are dealt with in a comprehensive way. All vector spaces which will be mentioned will be assumed to be over the complex field, unless otherwise stated.

1.1 Hausdorff and Minkowski dimensions

Definition 1.1.1. *Given any $s \geq 0$, $\varepsilon > 0$ and $E \subset \mathbb{R}^n$, we put*

$$H^s_\varepsilon(E) = \inf \left\{ \sum_{j=1}^{\infty} \omega_s 2^{-s} (\text{diam } A_j)^s : E \subset \bigcup_{j=1}^{\infty} A_j, \text{diam } A_j < \varepsilon \right\},$$

where

$$\omega_s = \pi^{s/2} / \Gamma\left(\frac{s}{2} + 1\right).$$

Note that H^s_ε is an outer measure on $\mathbb{R}^n$. Since $H^s_\varepsilon(E)$ is non-decreasing in ε, we define

$$H^s(E) = \lim_{\varepsilon \to 0} H^s_\varepsilon(E)$$

and call this the $s-$**dimensional Hausdorff outer measure of** E. The restriction of H^s to the $\sigma-$field of H^s-measurable sets (which can be shown to include the Borel sets) is called the $s-$**dimensional Hausdorff measure**

The reason for introducing H^s is to provide a means of distinguishing between various lower-dimensional subsets of $\mathbb{R}^n$. We summarise some of the more important properties of H^s as follows and refer to [90], [93] and [79] for proofs and further information:

(i) H^s is a Borel regular measure;
(ii) H^0 is counting measure;
(iii) H^n coincides with $n-$dimensional Lebesgue measure μ_n on $\mathbb{R}^n$;

(iv) if $s > n$, H^s is the zero measure on $\mathbb{R}^n$;
(v) $H^s(\lambda E) = \lambda^s H^s(E)$ for all $s \geq 0$, all $\lambda > 0$ and all $E \subset \mathbb{R}^n$;
(vi) if $f : \mathbb{R}^m \to \mathbb{R}^n$ is Lipschitz-continuous on $\mathbb{R}^m$ (that is, there exists $C > 0$ such that for all $x, y \in \mathbb{R}^m$, $|f(x) - f(y)| \leq C\,|x - y|$, where $|\cdot|$ denotes the corresponding Euclidean distance), then for all $E \subset \mathbb{R}^n$ and all $s \geq 0$,

$$H^s(f(E)) \leq C^s H^s(E).$$

To help us define the Hausdorff dimension of subsets of $\mathbb{R}^n$ the following Lemma will be very useful.

Lemma 1.1.2. *Let $E \subset \mathbb{R}^n$ and suppose that $0 \leq s < t < \infty$. If $H^s(E) < \infty$, then $H^t(E) = 0$; if $H^t(E) > 0$, then $H^s(E) = \infty$.*

Proof. First suppose that $H^s(E) < \infty$, and let $\delta > 0$. Then for some sets A_j $(j \in \mathbb{N})$ with diam $A_j \leq \delta$ and $E \subset \cup_{j=1}^{\infty} A_j$ we have

$$\sum_{j=1}^{\infty} \omega_s 2^{-s} (\text{diam } A_j)^s \leq H^s_\delta(E) + 1 \leq H^s(E) + 1.$$

Hence

$$H^t_\delta(E) \leq \sum_{j=1}^{\infty} \omega_t 2^{-t} (\text{diam } A_j)^t$$

$$= \frac{\omega_t}{\omega_s} 2^{s-t} \sum_{j=1}^{\infty} \omega_s 2^{-s} (\text{diam } A_j)^s (\text{diam } A_j)^{t-s}$$

$$\leq \frac{\omega_t}{\omega_s} 2^{s-t} \delta^{t-s} \left\{ H^s(E) + 1 \right\}.$$

Now let $\delta \to 0$: it follows that $H^t(E) = 0$. The rest is obvious. □

Definition 1.1.3. *The* ***Hausdorff dimension*** *of a subset E of $\mathbb{R}^n$ is defined to be*

$$d(E) = \inf \left\{ s \in [0, \infty) : H^s(E) = 0 \right\}.$$

Note that in view of Lemma 1.1.2,

$$H^s(E) = \begin{cases} 0 & \text{if } s > d(E), \\ \infty & \text{if } s < d(E). \end{cases}$$

Moreover, $0 \leq d(E) \leq n$; $d(E)$ need not be an integer, and even if $d(E) = k \in \mathbb{N}$, E need not be a "$k-$dimensional surface" in any reasonable sense.

Now we turn to the Minkowski dimension. Let E be a compact subset of $\mathbb{R}^n$, let $\varepsilon > 0$ and put

$$E_\varepsilon = \left\{ x \in \mathbb{R}^n : d(x, E) < \varepsilon \right\}.$$

Definition 1.1.4. *Given any compact subset E of $\mathbb{R}^n$ and any $d \geq 0$,*

$$M^d(E) := \limsup_{\varepsilon \to 0^+} \varepsilon^{-(n-d)} |E_\varepsilon|_n$$

*is the $d-$**dimensional upper Minkowski content** of E. Here $|E|_n$ is the Lebesgue n-measure of E. The **Minkowski dimension** of E is defined to be*

$$d_M(E) = \inf \left\{ d \geq 0 : M^d(E) = 0 \right\} = \sup \left\{ d \geq 0 : M^d(E) = \infty \right\}.$$

If $0 < M^{d_M(E)}(E) < \infty$ and

$$M^{d_M(E)}(E) = \lim_{\varepsilon \to 0^+} \varepsilon^{-(n-d)} |E_\varepsilon|_n ,$$

*then E is said to be **Minkowski-measurable** and $M^{d_M(E)}(E)$ is called the **Minkowski measure** of E. It is known (see, for example, [153]) that if Ω is a non-empty open subset of $\mathbb{R}^n$ with boundary $\partial\Omega$, then $n-1 \leq d_M(\partial\Omega) \leq n$.*

1.2 The area and coarea formulae

These important formulae relate to functions $f : \mathbb{R}^m \to \mathbb{R}^n$, where $m, n \in \mathbb{N}$, which are **Lipschitz-continous** so that there is a constant $C > 0$ such that for all $x, y \in \mathbb{R}^m$,

$$|f(x) - f(y)| \leq C |x - y| .$$

The weaker notion of a **locally Lipschitz** function is also useful: by this we mean that for each compact set $K \subset \mathbb{R}^m$, there is a constant $C(K)$ such that for all $x, y \in K$,

$$|f(x) - f(y)| \leq C(K) |x - y| .$$

An important result for such functions is

Theorem 1.2.1. *(Rademacher's theorem) Let $f : \mathbb{R}^m \to \mathbb{R}^n$ be locally Lipschitz and let μ_m be Lebesgue $m-$measure on $\mathbb{R}^m$. Then f is differentiable μ_m-a.e. on $\mathbb{R}^m$.*

From this it is not difficult to prove

Corollary 1.2.2. *Let $f : \mathbb{R}^m \to \mathbb{R}^n$ be locally Lipschitz. Then its derivative $Df(x)$ is zero for μ_m-a.e. $x \in \ker(f)$.*

For these results see [79].

We shall also need the Jacobian of a Lipschitz map $f : \mathbb{R}^m \to \mathbb{R}^n$. For this we recall some basic facts concerning linear maps $L : \mathbb{R}^m \to \mathbb{R}^n$. Such a map is called **orthogonal** if $(Lx, Ly)_n = (x, y)_m$ for all $x, y \in \mathbb{R}^m$; it is **symmetric** if $m = n$ and $(x, Ly)_m = (Lx, y)_m$ for all $x, y \in \mathbb{R}^m$. Here $(\cdot, \cdot)_k$ denotes the inner product in $\mathbb{R}^k$. If $m \leq n$, there are a symmetric map $S : \mathbb{R}^m \to \mathbb{R}^m$ and an orthogonal map $O : \mathbb{R}^m \to \mathbb{R}^n$ such that $L = O \circ S$; if $m \geq n$, there are a symmetric map $S : \mathbb{R}^n \to \mathbb{R}^n$ and an orthogonal map $O : \mathbb{R}^n \to \mathbb{R}^m$ such that $L = S \circ O^*$, where O^* is the adjoint of O. In both cases we define the **Jacobian** of L to be $|\det S|$. For details of all this, and for a proof that the Jacobian is well-defined (that is, independent of the particular choices of O and S) we refer to [79].

Returning to our Lipschitz map $f : \mathbb{R}^m \to \mathbb{R}^n$, we note that by Rademacher's theorem, f is differentiable μ_m-a.e., so that its derivative $Df(x)$ exists and corresponds to a linear map from $\mathbb{R}^m$ to $\mathbb{R}^n$ for μ_m-a.e. $x \in \mathbb{R}^m$. The **Jacobian** of f at μ_m-a.e. $x \in \mathbb{R}^m$ is defined to be the Jacobian of this linear map and is denoted by $Jf(x)$.

After these preliminaries we can give the *area theorem*:

Theorem 1.2.3. *Let $m, n \in \mathbb{N}$, $m \leq n$, let $f : \mathbb{R}^m \to \mathbb{R}^n$ be Lipschitz-continuous and let A be a μ_m-measurable subset of $\mathbb{R}^m$. Then*

$$\int_A Jf dx = \int_{\mathbb{R}^n} H^0 \left(A \cap f^{-1}(y) \right) dH^m(y).$$

The corresponding result when $m \geq n$ is the *coarea theorem* :

Theorem 1.2.4. *Let $m, n \in \mathbb{N}$, $m > n$, let $f : \mathbb{R}^m \to \mathbb{R}^n$ be Lipschitz-continuous, let $g \in L_1(\mathbb{R}^m)$ and let A be a μ_m-measurable subset of $\mathbb{R}^m$. Then*

$$\int_A g(x) Jf(x) dx = \int_{\mathbb{R}^n} \int_{A \cap f^{-1}(y)} g(x) dH^{m-n}(x) dy.$$

For a proof of these important theorems we refer to [79] and [93].

From the special case $n = 1$, $m > 1$, $A = \mathbb{R}^m$ of the coarea theorem we have a result of particular interest.

Corollary 1.2.5. *Let $f : \mathbb{R}^m \to \mathbb{R}$ be Lipschitz-continuous and let $g \in L_1(\mathbb{R}^m)$. Then*

$$\int_{\mathbb{R}^m} g(x) \left| \nabla f(x) \right| dx = \int_0^\infty \int_{\{x \in \mathbb{R}^m : |f(x)| = t\}} g(x) dH^{m-1}(x) dt.$$

Proof. Just observe that $Jf = |\nabla f|$ and use Theorem 1.2.4. □

A more general version of the coarea formula will be useful. To explain this, let Ω be an open subset of $\mathbb{R}^n$ and let $f \in L_1(\Omega)$. We say that f is of **bounded variation** in Ω if its first-order distributional partial derivatives are signed Radon measures with finite total variation in Ω. The family of all functions of bounded variation on Ω is denoted by $BV(\Omega)$. If $u \in BV(\Omega)$, the distributional gradient Du of u is a vector-valued measure whose total variation $\|Du\|(\Omega)$ is a finite measure on Ω, and

$$\|Du\|(\Omega) = \sup\left\{\int_\Omega u \text{ div } \phi\, dx : \phi \in C_0^\infty(\Omega, \mathbb{R}^n), |\phi(x)| \le 1 \text{ for all } x \in \Omega\right\}.$$

Given any $u \in BV(\Omega)$, the measure Du can be split into a part which is absolutely continuous with respect to Lebesgue measure, and a singular part. The density of the absolutely continuous part will be denoted by ∇u : thus if $u \in W_1^1(\Omega)$, $dDu = \nabla u dx$ and

$$\|Du\|(\Omega) = \int_\Omega |\nabla u|\, dx.$$

A set $E \subset \mathbb{R}^n$ is said to have **finite perimeter** if its characteristic function χ_E is in $BV(\mathbb{R}^n)$, in which case the **perimeter** of E is defined to be

$$P(E) = \|D\chi_E\|(\mathbb{R}^n).$$

It can be shown that sets with minimally smooth boundary, such as Lipschitz domains, have finite perimeter. The version of the coarea theorem which we shall need involves the perimeter of sets of the form

$$E_t := \{x \in \Omega : u(x) > t\},\ t > 0,\ u \in BV(\Omega).$$

With this notation, the theorem reads as follows.

Theorem 1.2.6. *Let Ω be an open subset of $\mathbb{R}^n$ and let $u \in BV(\Omega)$. Then*

$$\|Du\|(\Omega) = \int_{\mathbb{R}} \|D\chi_{E_t}\|(\Omega) dt.$$

Moreover, if $u \in W_1^1(\mathbb{R}^n)$ (the Sobolev space consisting of functions which, together with their first-order distributional derivatives, are in $L_1(\mathbb{R}^n)$) and f is any Borel function on $\mathbb{R}^n$,

$$\int_{\mathbb{R}^n} f\, |\nabla u|\, dx = \int_{\mathbb{R}} \int_{\{u=t\}} f dH^{n-1}(x) dt.$$

For proofs of results of this nature we refer to the books of Giusti [100], Maz'ya [171] and Ziemer [231].

In conjunction with the coarea theorem we shall sometimes need the classical isoperimetric inequality:

Let E be a subset of $\mathbb{R}^n$ with finite $n-$measure $|E|_n$ and finite perimeter. Then

$$P(E) \ge n\omega_n^{1/n}\, |E|_n^{1-1/n}. \tag{1.2.1}$$

The books just mentioned may be consulted for details of this famous result. It implies that if E is a subset of $\mathbb{R}^n$ with finite $n-$measure and appropriate boundary, then

$$H^{n-1}(\partial E) \ge n\omega_n^{1/n}\, |E|_n^{1-1/n}. \tag{1.2.2}$$

Often it is applied in the situation where $E = \{x \in \mathbb{R}^n : |u(x)| > t\}$, $t > 0$, where u is a smooth function on $\mathbb{R}^n$ with compact support.

1.3 Approximation numbers

First, it may be helpful to give some information about quasi-normed spaces. A **quasi-norm** on a linear space X is a map $\|\cdot \mid X\| : X \to [0,\infty)$ which has the following three properties:
(i) $\|x \mid X\| = 0$ if, and only if, $x = 0$;
(ii) $\|\lambda x \mid X\| = |\lambda| \, \|x \mid X\|$ for all scalars λ and all $x \in X$;
(iii) there is a constant C such that for all $x, y \in X$;

$$\|x + y \mid X\| \leq C \left(\|x \mid X\| + \|y \mid X\|\right).$$

It is clear that $C \geq 1$. If it is possible to take $C = 1$, then (iii) is the familiar triangle inequality and $\|\cdot \mid X\|$ is a norm on X. A quasi-norm $\|\cdot \mid X\|$ defines a topology on X which is compatible with the linear structure of X : this topology has a basis of (not necessarily open) neighbourhoods of any point $x \in X$ given by the sets $\{y \in X : \|x - y \mid X\| < 1/n\}$, $n \in \mathbb{N}$. The pair $(X, \|\cdot \mid X\|)$ is said to be a **quasi-normed space** and is a special type of metrisable topological vector space. The notions of convergence and of Cauchy sequences are defined in the obvious way, and if every Cauchy sequence in X converges, to a point in X, then X is called a **quasi-Banach** space.

Let $p \in (0,1]$. By a $p-$**norm** on a linear space X is meant a map $\|\cdot \mid X\| : X \to [0,\infty)$ which has properties (i) and (ii) above and instead of (iii) satisfies (iii') $\|x + y \mid X\|^p \leq \|x \mid X\|^p + \|x \mid X\|^p$ for all $x, y \in X$.

Two quasi-norms or $p-$norms $\|\cdot \mid X\|_1$ and $\|\cdot \mid X\|_2$ on X are called **equivalent** if there is a constant $c \geq 1$ such that for all $x \in X$,

$$c^{-1} \|x \mid X\|_1 \leq \|x \mid X\|_2 \leq c \|x \mid X\|_1 .$$

It can be shown that if $\|\cdot \mid X\|_1$ is a quasi-norm on X, then there exist $p \in (0,1]$ and a $p-$norm $\|\cdot \mid X\|_2$ on X which is equivalent to $\|\cdot \mid X\|_1$; the connection between p and the constant C in (iii) is given by $C = 2^{\frac{1}{p}-1}$. Conversely, any $p-$norm is a quasi-norm with $C = 2^{\frac{1}{p}-1}$.

The standard examples of quasi-Banach spaces which are not Banach spaces are l_p and L_p, with $0 < p < 1$.

Let X, Y be quasi-Banach spaces and let $T : X \to Y$ be linear. As in the Banach space case, T is called **bounded** or **continuous** if

$$\|T\| := \sup \{\|Tx \mid Y\| : x \in X, \|x \mid X\| \leq 1\} < \infty.$$

Let X and Y be Banach spaces and let $\mathcal{B}(X,Y)$ be the space of all bounded linear maps from X to Y. If $T \in \mathcal{B}(X,Y)$ and $k \in \mathbb{N}$, the k^{th} **approximation number** of T, denoted by $a_k(T)$, is defined by

$$a_k(T) = \inf \ \{\|T - L\| : L \in \mathcal{B}(X,Y), \ \text{rank } L < k\},$$

where rank $L = \dim L(X)$. The same definition can be used for the situation in which X and Y are quasi-Banach spaces.

It is easy to verify that if X, Y and Z are Banach spaces and $S, T \in \mathcal{B}(X,Y)$, $R \in \mathcal{B}(Y,Z)$, then
(i) $\|T\| = a_1(T) \geq a_2(T) \geq ... \geq 0$;
(ii) for all $k, l \in \mathbb{N}$,

$$a_{k+l-1}(S+T) \leq a_k(S) + a_l(T)$$

and

$$a_{k+l-1}(R \circ S) \leq a_k(R) a_l(S);$$

(iii) $a_k(T) = 0$ if, and only if, rank $T < k$;
(iv) if dim $X \geq n$ and $id : X \to X$ is the identity map, then $a_k(id) = 1$ for $k = 1, ..., n$.

With more effort (see [46], Prop. II.2.5), it can be shown that
(v) if T is compact, then $a_k(T) = a_k(T^*)$ for all $k \in \mathbb{N}$.

In view of (i) above, it is clear that

$$\alpha(T) := \lim_{k \to \infty} a_k(T)$$

exists. If $\alpha(T) = 0$, then T is the limit (in the operator norm sense) of a sequence of finite-dimensional maps and so is compact. However, if T is compact it does not follow that $\alpha(T) = 0$: this is a consequence of Per Enflo's work on the approximation problem (see [78]). Compactness of T *does* imply that the approximation numbers converge to zero if Y has the **bounded approximation property** (see [228]): we recall that this means that there is a constant C such that for every finite subset F of Y and every $\varepsilon > 0$, there is a bounded linear map $L : Y \to Y$ with finite rank such that $\|Ly - y \mid Y\| \leq \varepsilon$ for all $y \in F$, and $\|L\| \leq C$. This is so if, for example, Y is a Hilbert space or $Y = L_p(\Omega)$, where $1 \leq p < \infty$ and Ω is an open subset of $\mathbb{R}^n$; in both these cases (see [46], Corollary V.5.4),

$$\alpha(T) := \inf \{\|T - K\| : K \text{ is a compact linear map from } X \text{ to } Y\}.$$

An important property of the approximation numbers is their connection with eigenvalues, in a Hilbert space setting. Thus if H is a complex Hilbert space and T is a compact linear map from H to itself, then T^*T has a positive compact square root $|T|$, which accordingly has a sequence $\{\lambda_k(|T|)\}$ of positive eigenvalues, each repeated according to multiplicity and ordered so that

$$\lambda_1(|T|) \geq \lambda_2(|T|) \geq ... \geq 0.$$

If T has only a finite number of distinct positive eigenvalues and M is the sum of their multiplicities, we put $\lambda_k(|T|) = 0$ for all $k > M$. The eigenvalues $\lambda_k(|T|)$ of $|T|$ are called the **singular values** of T. It turns out (see, for example, Theorem II.5.10 of [46]) that for all $k \in \mathbb{N}$,

$$a_k(T) = \lambda_k(|T|).$$

In particular, if T is compact and positive (hence self-adjoint), then for all $k \in \mathbb{N}$,

$$a_k(T) = \lambda_k(T).$$

It is plain that for a compact map $T \in \mathcal{B}(X, Y)$ the approximation numbers may be thought of as providing a means of measuring 'how compact' it is, at least under some restrictions on Y. There are other sequences of numbers which perform the same function: here we single out the entropy numbers for special mention. Let X and Y be Banach spaces and let $U_X = \{x \in X : \|x \mid X\| \leq 1\}$. Given $T \in \mathcal{B}(X, Y)$ and $k \in \mathbb{N}$, the k^{th} **entropy number** of T, denoted by $e_k(T)$, is defined by

$$e_k(T) = \inf \left\{ \varepsilon > 0 : T(U_X) \text{ can be covered by } 2^{k-1} \text{ balls in } Y \text{ of radius } \varepsilon \right\}.$$

It may be easily checked that properties (i) and (ii) above of the approximation numbers are also enjoyed by the entropy numbers. This is not so for (iii)-(v), however. Moreover,

$$\beta(T) := \lim_{k \to \infty} e_k(T)$$

is the (**ball**) **measure of non-compactness** of T; and T is compact if, and only if, $\beta(T) = 0$.

If T is a compact linear map from a Banach space X to itself, its spectrum, apart from the point 0, consists of eigenvalues of finite algebraic multiplicity: we let $\{\lambda_k(T)\}$ be the sequence of all non-zero eigenvalues of T, repeated according to algebraic multiplicity and ordered by decreasing modulus. If T has only a finite number of distinct eigenvalues and M is the sum of their algebraic multiplicities, then just as before we put $\lambda_k(T) = 0$ for all $k > M$. A most useful connection between the spectral properties of T and its geometrical characteristics as expressed by the entropy numbers is provided by Carl's inequality (see [28]):

$$|\lambda_k(T)| \leq \sqrt{2} e_k(T) \text{ for all } k \in \mathbb{N}.$$

For another proof of this and a more general inequality see [30]; an extension to quasi-Banach spaces is given in [74].

Two-sided estimates of the approximation numbers of embeddings between Sobolev spaces (and much more general spaces) are available. To illustrate this, let Ω be a bounded domain in $\mathbb{R}^n$ with smooth boundary, and for any $k \in \mathbb{N}$ and any $p \in (0, \infty]$ let $W_p^k(\Omega)$ be the Sobolev space of all functions u which, together with their distributional derivatives of all orders up to and including k, are in $L_p(\Omega)$. When endowed with the quasi-norm

$$\left\| u \mid W_p^k(\Omega) \right\| = \left(\sum_{|\alpha| \leq k} \|D^\alpha u \mid L_p(\Omega)\|^p \right)^{1/p}$$

(with the natural interpretation when $p = \infty$), this is a quasi-Banach space. Now suppose that

$$s_1, s_2 \in \mathbb{N};\ p_1, p_2 \in (0, \infty] \text{ and that } \delta^+ := s_1 - s_2 - n\left(\frac{1}{p_1} - \frac{1}{p_2}\right)_+ > 0.$$

Then $W^{s_1}_{p_1}(\Omega)$ is compactly embedded in $W^{s_2}_{p_2}(\Omega)$; denote the embedding map by id. It turns out that if in addition $0 < p_1 \le p_2 \le 2$, or $2 \le p_1 \le p_2 \le \infty$, or $0 < p_2 \le p_1 \le \infty$, then

$$a_k(id) \approx k^{-\delta^+/n}.$$

The situation when p_1 and p_2 lie on opposite sides of 2, with $p_1 < p_2$, is more complicated, but it can be shown that if in addition to the hypothesis that $\delta^+ > 0$ we have $0 < p_1 < 2 < p_2 < \infty$ (or $1 < p_1 < 2 < p_2 = \infty$) and $\delta^+ < n/\min\{p_1', p_2\}$, then

$$a_k(id) \approx k^{-\frac{\delta^+}{2n}\min\{p_1', p_2\}}.$$

For these results we refer to [74], Chapter 3, and [26].

Additional results relating to the material in this section, and in particular concerning comparisons between approximation and entropy numbers, may be found in [29], [46], [74], [143], [198] and [199].

1.4 Inequalities

Here we give some inequalities which will be of help in the text. The first is of Minkowski type.

Theorem 1.4.1. *Let (S_1, μ_1) and (S_2, μ_2) be positive measure spaces and let K be a $\mu_1 \times \mu_2-$measurable function on $S_1 \times S_2$. Then if $1 \le p < \infty$,*

$$\begin{aligned} &\left\{\int_{S_1}\left[\int_{S_2}|K(s_1, s_2)|\, d\mu_1(s_1)\right]^p d\mu_2(s_2)\right\}^{1/p} \\ &\le \int_{S_1}\left\{\int_{S_2}|K(s_1, s_2)|^p\, d\mu_2(s_2)\right\}^{1/p} d\mu_1(s_1). \end{aligned}$$

For this we refer to [44], Vol. 1, p. 530.

The next is Jensen's inequality.

Theorem 1.4.2. *Let (X, μ) be a finite measure space, let I be an interval in $\mathbb{R}$, let $\Phi : I \to \mathbb{R}$ be convex and suppose that $f \in L_1(X, \mu)$ is such that $f(X) \subset I$ and $\Phi \circ f \in L_1(X, \mu)$. Then*

$$\Phi\left(\frac{1}{\mu(X)}\int_X f d\mu\right) \le \frac{1}{\mu(X)}\int_X (\Phi \circ f) d\mu.$$

We refer to [126], p.202, for this.