1

Preliminaries

1.1 Function spaces

1.1.1 Functions of the space variables

Let Q be a domain in $\mathbb{R}^d$ (i.e., a connected open subset of $\mathbb{R}^d$) or the torus $\mathbb{T}^d = \mathbb{R}^d/2\pi\mathbb{Z}^d$. We shall say that a domain Q is *Lipschitz* if its boundary ∂Q is locally Lipschitz.[1] We shall need *Lebesgue* and *Sobolev* spaces on Q and some embedding and interpolation theorems.

Lebesgue spaces

We denote by $L^p(Q;\mathbb{R}^n)$, $1 \le p \le \infty$, the usual Lebesgue space of vector-valued functions and abbreviate $L^p(Q;\mathbb{R}) = L^p(Q)$. We write $\langle \cdot, \cdot \rangle$ for the L_2 scalar product and $|\cdot|_p$ for the standard norm in $L^p(Q;\mathbb{R}^n)$.

Sobolev spaces

We denote by $C_0^\infty(Q;\mathbb{R}^n)$ the space of infinitely smooth functions $\varphi : Q \to \mathbb{R}^n$ with compact support. Let u and v be two locally integrable scalar functions on Q and let $\alpha = (\alpha_1, \dots, \alpha_d)$ be a multi-index. We say that v *is the* α^{th} *weak partial derivative of* u if

$$\int_Q u D^\alpha \varphi \, dx = (-1)^{|\alpha|} \int_Q v\varphi \, dx \quad \text{for all } \varphi \in C_0^\infty(Q;\mathbb{R}),$$

where $|\alpha| := \alpha_1 + \cdots + \alpha_d$ and $D^\alpha = \partial_1^{\alpha_1} \cdots \partial_d^{\alpha_d}$. In this case, we write $D^\alpha u = v$.

Let $m \ge 0$ be an integer. The space $H^m(Q,\mathbb{R}^n)$ consists of all locally integrable functions $u : Q \to \mathbb{R}^n$ such that the derivative $D^\alpha u$ exists in the weak

[1] This means that ∂Q can be represented locally as the graph of a Lipschitz function.

sense for each multi-index α with $|\alpha| \le m$ and belongs to $L^2(Q;\mathbb{R}^n)$. We write $H^m(Q;\mathbb{R}) = H^m(Q)$ and define the norm in $H^m(Q;\mathbb{R}^n)$ as

$$\|u\|_m := \Big(\sum_{|\alpha|\le m} |D^\alpha u|_2^2\Big)^{1/2}.$$

In the case $Q = \mathbb{T}^d$, it is easy to define $H^m(\mathbb{T}^d;\mathbb{R}^n)$ for all $m \in \mathbb{R}$. To this end, let us expand a function $u \in L^2(\mathbb{T}^d, \mathbb{R}^n)$ in a Fourier series:

$$u(x) = \sum_{s\in\mathbb{Z}^d} u_s e^{isx}.$$

Define the following norm, which is equivalent to $\|\cdot\|_m$ for non-negative integers m:

$$\|u\|_m = \Big(\sum_{s\in\mathbb{Z}^d} \big(1+|s|^2\big)^m |u_s|^2\Big)^{1/2}. \tag{1.1}$$

The space $H^m(\mathbb{T}^d;\mathbb{R}^n)$ is defined as the closure of $C^\infty(\mathbb{T}^d,\mathbb{R}^n)$ with respect to the norm $\|\cdot\|_m$. It is easy to see that if $m \ge 0$ is an integer, then the two definitions of $H^m(\mathbb{T}^d;\mathbb{R}^n)$ give the same function space. The following result is a simple consequence of the definition of $\|\cdot\|_m$.

Lemma 1.1.1 *For any $m \in \mathbb{R}$ and any multi-index α, the linear map D^α is continuous from $H^m(\mathbb{T}^d;\mathbb{R}^n)$ to $H^{m-|\alpha|}(\mathbb{T}^d;\mathbb{R}^n)$. Accordingly, the Laplace operator $\Delta : H^m(\mathbb{T}^d;\mathbb{R}^n) \to H^{m-2}(\mathbb{T}^d;\mathbb{R}^n)$ is continuous. Similar assertions are true for any open domain $Q \subset \mathbb{R}^d$ and any integer $m \ge 0$.*

Now let $u \in H^m(\mathbb{T}^d;\mathbb{R}^n)$ be a function with zero mean value, that is,

$$\langle u\rangle := (2\pi)^{-d}\int_{\mathbb{T}^d} u(x)\,dx = 0, \tag{1.2}$$

where the integral is understood in the sense of the theory of distributions if $m < 0$. In this case, the first Fourier coefficient of u is zero, $u_0 = 0$, and therefore the norm

$$\|u\|_m = \Big(\sum_{s\ne 0} |s|^{2m}|u_s|^2\Big)^{1/2}$$

is equivalent to (1.1) on the space

$$\dot H^m(\mathbb{T}^d;\mathbb{R}^n) = \{u \in H^m(\mathbb{T}^d;\mathbb{R}^n) : \langle u\rangle = 0\}.$$

In particular, $\|u\|_1^2 = |\nabla u|_2$ is a norm on $\dot H^1(\mathbb{T}^d;\mathbb{R}^n)$.

Finally, let us define the *Sobolev space* $H^m(Q;\mathbb{R}^n)$ in a bounded Lipschitz domain $Q \subset \mathbb{R}^d$ for an arbitrary $m \ge 0$. Namely, without loss of generality, we can assume that $Q \subset \mathbb{T}^d$. We shall say that a function $u \in L^2(Q,\mathbb{R}^n)$ belongs to $H^m(Q;\mathbb{R}^n)$ if there is a function $\tilde u \in H^m(\mathbb{T}^d;\mathbb{R}^n)$ whose restriction to Q coincides with u. In this case, we define $\|u\|_m$ as the infimum of $\|\tilde u\|_m$ over all possible extensions $\tilde u \in H^m(\mathbb{T}^d;\mathbb{R}^n)$ for u.

Property 1.1.2 *Sobolev embeddings*. Let Q be either a Lipschitz domain in $\mathbb{R}^d$ or the torus $\mathbb{T}^d$.

1. If $m \le \frac{d}{2}$ and $2 \le q \le \frac{2d}{d-2m}$, $q < \infty$, then
$$H^m(Q;\mathbb{R}^n) \subset L^q(Q;\mathbb{R}^n). \tag{1.3}$$
2. If $m \ge \frac{d}{2} + \alpha$ with $0 < \alpha < 1$, then
$$H^m(Q;\mathbb{R}^n) \subset C_b^\alpha(Q;\mathbb{R}^n), \tag{1.4}$$
where $C_b^\alpha(Q)$ denotes the space of functions that are bounded and Hölder continuous with exponent α. In particular, if $m > \frac{d}{2}$, then $H^m(Q;\mathbb{R}^n)$ is continuously embedded into the space $C_b(Q;\mathbb{R}^n)$ of bounded continuous functions.
3. If Q is bounded, then we have the compact embedding
$$H^{m_1}(Q;\mathbb{R}^n) \Subset H^{m_2}(Q;\mathbb{R}^n) \quad \text{for } m_1 > m_2. \tag{1.5}$$
It follows that embeddings (1.3) and (1.4) are compact for $q < \frac{2d}{d-2m}$ and $m > \frac{d}{2} + \alpha$, respectively.

Property 1.1.3 *Duality*. The spaces $H^m(\mathbb{T}^d;\mathbb{R}^n)$ and $H^{-m}(\mathbb{T}^d;\mathbb{R}^n)$ are dual with respect to the L^2-scalar product $\langle\cdot,\cdot\rangle$. That is,
$$\|u\|_m = \sup_v |\langle u, v\rangle| \quad \text{for any } u \in C^\infty(\mathbb{T}^d;\mathbb{R}^n), \tag{1.6}$$
where the supremum is taken over all $v \in C^\infty(\mathbb{T}^d;\mathbb{R}^n)$ such that $\|v\|_{-m} \le 1$. Relation (1.6) implies that the scalar product in L^2 extends to a continuous bilinear map from $H^m(\mathbb{T}^d;\mathbb{R}^n) \times H^{-m}(\mathbb{T}^d;\mathbb{R}^n)$ to $\mathbb{R}$.

Property 1.1.4 *Interpolation inequality*. Let $Q \subset \mathbb{R}^d$ be a Lipschitz domain, let $a < b$ be non-negative integers, and let $0 \le \theta \le 1$ be a constant. Then
$$\|u\|_{\theta a+(1-\theta)b} \le \|u\|_a^\theta \|u\|_b^{1-\theta} \quad \text{for any } u \in H^b(Q;\mathbb{R}^n). \tag{1.7}$$
In the case of the torus, inequality (1.7) holds for any real numbers $a < b$ and any $\theta \in [0, 1]$.

Proof for the case of a torus We have
$$\begin{aligned}\|u\|^2_{\theta a+(1-\theta)b} &= \sum_{s\in\mathbb{Z}^d} \bigl(1+|s|^2\bigr)^{\theta a+(1-\theta)b}|u_s|^2 \\ &= \sum_{s\in\mathbb{Z}^d}\Bigl(\bigl(1+|s|^2\bigr)^{\theta a}|u_s|^{2\theta}\Bigr)\Bigl(\bigl(1+|s|^2\bigr)^{(1-\theta)b}|u_s|^{2(1-\theta)}\Bigr) \\ &\le \Bigl(\sum_{s\in\mathbb{Z}^d}\bigl(1+|s|^2\bigr)^a|u_s|^2\Bigr)^\theta \Bigl(\sum_{s\in\mathbb{Z}^d}\bigl(1+|s|^2\bigr)^b|u_s|^2\Bigr)^{1-\theta},\end{aligned}$$
where we used Hölder's inequality in the last step. □

Example 1.1.5 Let Q be either a Lipschitz domain in $\mathbb{R}^2$ or the torus $\mathbb{T}^2$. Then the Sobolev embedding (1.3), with $m = 1/2$ and $q = 4$, and the interpolation inequality (1.6) with $a = 0$, $b = 1$, and $\theta = \frac{1}{2}$, imply that

$$|u|_4 \le C_1 \|u\|_{1/2} \le C_2 \sqrt{|u|_2 \|u\|_1} \quad \text{for any } u \in H^1(Q; \mathbb{R}^n). \tag{1.8}$$

This is *Ladyzhenskaya's inequality*.

A proof of Properties 1.1.2–1.1.4 can be found in [BIN79; Ste70; Tay97].

1.1.2 Functions of space and time variables

Solutions of the equations mentioned in the introduction are functions depending on the time t and the space variables x. We fix any $T > 0$ and view a solution $u(t, x)$ with $0 \le t \le T$ as a map

$$[0, T] \longrightarrow \text{“space of functions of } x\text{”}, \qquad t \mapsto u(t, \cdot)\,.$$

Let us introduce the corresponding functional spaces.

For a Banach space X, we denote by $C(0, T; X)$ the space of continuous functions $u : [0, T] \to X$ and endow it with the norm

$$\|u\|_{C(0,T;X)} = \sup_{0 \le t \le T} \|u(t)\|_X,$$

where $\|\cdot\|_X$ stands for the norm in X. We denote by $S(0, T; X)$ the space of functions of the form

$$u(t) = \sum_{k=1}^{N} u_k \mathbb{I}_{\Gamma_k}(t),$$

where $N \ge 1$ is an integer depending on the function, $u_k \in X$ are some vectors, Γ_k are Borel-measurable subsets of $[0, T]$ (see Section 1.2.1), and $\mathbb{I}_\Gamma$ stands for the indicator function of Γ. If X is separable, then for $p \in [1, \infty]$ define $L^p(0, T; X)$ as the completion of the space $S(0, T; X)$ with respect to the norm

$$\|u\|_{L^p(0,T;X)} = \begin{cases} \left(\int_0^T \|u(t)\|_X^p dt \right)^{1/p} & \text{for } 1 \le p < \infty, \\ \operatorname*{ess\,sup}_{0 \le t \le T} \|u(t)\|_X & \text{for } p = \infty. \end{cases}$$

Note that, in view of Fubini's theorem, we have

$$L^p\big(0, T; L^p(Q; \mathbb{R}^n)\big) = L^p\big((0, T) \times Q; \mathbb{R}^n\big) \quad \text{for } p < \infty.$$

A more detailed discussion of these spaces can be found in [Lio69; Yos95].

We shall also need the space of continuous functions on an interval with range in a metric space. Namely, let $J \subset \mathbb{R}$ be a closed interval and let X be a Polish space, that is, a complete separable metric space with a distance dist_X.

We denote by $C(J;X)$ the space of continuous functions from J to X. When J is bounded, $C(J;X)$ is a Polish space with respect to the distance

$$\|u-v\|_{C(J;X)} = \max_{t\in J} \operatorname{dist}_X\big(u(t), v(t)\big).$$

In the case of an unbounded interval J, we endow $C(J;X)$ with the metric

$$\operatorname{dist}(u,v) = \sum_{k=1}^{\infty} 2^{-k} \frac{\|u-v\|_{C(J_k;X)}}{1+\|u-v\|_{C(J_k;X)}}, \tag{1.9}$$

where $J_k = J\cap[-k,k]$. Note that, for a sequence $\{u_j\}\subset C(J,X)$, we have $\operatorname{dist}(u_j,u)\to 0$ as $j\to\infty$ if and only if $\|u_j-u\|_{C(J_k;X)}$ for each k. That is, (1.9) is the *metric of uniform convergence on bounded intervals*. When $J=\mathbb{Z}$ (or J is a countable subset of $\mathbb{Z}$), formula (1.9) may be used to define a distance on X^J. This distance corresponds to the *Tikhonov topology* on X^J.

Exercise 1.1.6 Prove that if $J\subset\mathbb{R}$ is an unbounded closed interval, then $C(J;X)$ is a Polish space. Prove also that if X is a separable Banach space, then $C(J;X)$ is a separable Fréchet space.

1.2 Basic facts from measure theory

In this section, we first recall the concept of a σ-algebra, together with some related definitions, and formulate without proof three standard results on the passage to the limit under Lebesgue's integral. We next discuss various metrics on the space of probability measures on a Polish space and establish some results on (maximal) couplings of measures.

1.2.1 σ-algebras and measures

Let Ω be an arbitrary set and let $\mathcal{F}$ be a family of subsets of Ω. Recall that $\mathcal{F}$ is called a *σ-algebra* if it contains the sets $\varnothing$ and Ω, and is invariant under taking the complement and countable union of its elements. Any pair $(\Omega,\mathcal{F})$ is called a *measurable space*. If $(\Omega_i,\mathcal{F}_i)$, $i=1,2$, are measurable spaces, then a mapping $f:\Omega_1\to\Omega_2$ is said to be *measurable* if $f^{-1}(\Gamma)\in\mathcal{F}_1$ for any $\Gamma\in\mathcal{F}_2$. If μ is a (positive) measure on $(\Omega_1,\mathcal{F}_1)$, then its *image* under f is the measure $f_*(\mu)$ on $(\Omega_2,\mathcal{F}_2)$ defined by $f_*(\mu)(\Gamma)=\mu(f^{-1}(\Gamma))$ for any $\Gamma\in\mathcal{F}_2$. Note that f_* is a linear mapping on the space of positive measures:

$$f_*(c_1\mu_1+c_2\mu_2) = c_1 f_*(\mu_1)+c_2 f_*(\mu_2) \quad \text{for any } c_1,c_2\geq 0.$$

The *product* of two measurable spaces $(\Omega_i,\mathcal{F}_i)$, $i=1,2$, is defined as the set $\Omega_1\times\Omega_2$ endowed with the minimal σ-algebra $\mathcal{F}_1\otimes\mathcal{F}_2$ generated by subsets of the form $\Gamma_1\times\Gamma_2$ with $\Gamma_i\in\mathcal{F}_i$. The product of finitely or countably many σ-algebras is defined in a similar way.

Given a probability measure μ on a measurable space $(\Omega,\mathcal{F})$, we denote by $\mathcal{N}_\mu$ the family of subsets $A\subset\Omega$ such that $A\subset B$ for some $B\in\mathcal{F}$ with

$\mu(B) = 0$. A σ-algebra $\mathcal{F}$ is said to be *complete* with respect to a measure μ if it contains all sets from $\mathcal{N}_\mu$. The *completion of $\mathcal{F}$ with respect to μ* is defined as the minimal σ-algebra generated by $\mathcal{F} \cup \mathcal{N}_\mu$ and is denoted by $\mathcal{F}_\mu$. This is the minimal complete σ-algebra which contains $\mathcal{F}$. A subset $\Gamma \subset \Omega$ is said to be *universally measurable* if it belongs to $\mathcal{F}_\mu$ for any probability measure μ on $(\Omega, \mathcal{F})$. If μ is a measure on $(\Omega_1, \mathcal{F}_1)$, $\mathcal{F}_1$ is complete with respect to μ, and a map $f : \Omega_1 \to \Omega_2$ is a μ-almost sure limit of a sequence of measurable maps, then f is measurable.

Now let X be a Polish space, that is, a complete separable metric space. We denote by dist_X the metric on X. The *Borel σ-algebra* $\mathcal{B} = \mathcal{B}(X)$ is defined as the minimal σ-algebra containing all open subsets of X. The pair $(X, \mathcal{B}(X))$ is called a *measurable Polish space*. If X_1 and X_2 are Polish spaces, then a map $f : X_1 \to X_2$ is said to be *measurable* if $f^{-1}(\Gamma) \in \mathcal{B}(X_1)$ for any $\Gamma \in \mathcal{B}(X_2)$. In particular, a function $f : X \to \mathbb{R}$ is called *measurable* if it is measurable with respect to the Borel σ-algebras on X and $\mathbb{R}$. An important property of Polish spaces is that any probability measure on it is *regular*. Namely, *Ulam's theorem* says that, for any probability measure μ on a Polish space X and any $\varepsilon > 0$, there is a compact set $K \subset X$ such that $\mu(K) \geq 1 - \varepsilon$. A proof of this result can be found in [Dud02] (see theorem 7.1.4).

Recall that, for any probability measure $\mathbb{P}$ on a measurable space $(\Omega, \mathcal{F})$, the triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a probability space. A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is said to be *complete* if $\mathcal{F}_\mathbb{P} = \mathcal{F}$. We shall often consider a probability space together with a family $\{\mathcal{F}_t \subset \mathcal{F}\}$ of σ-algebras that depend on a parameter t varying either in $\mathbb{R}_+$ or in $\mathbb{Z}_+$. In this case, we shall always assume that $\mathcal{F}_t$ is non-decreasing with respect to t. The quadruple $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ is called a *filtered probability space*. We shall say that $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ satisfies the *usual hypotheses* if $(\Omega, \mathcal{F}, \mathbb{P})$ is complete and $\mathcal{F}_t$ contains all $\mathbb{P}$-null sets of $\mathcal{F}$.

If X is a Polish space, then an X-valued *random variable* is a measurable map ξ from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ into X. The *law* or the *distribution* of ξ is defined as the image of $\mathbb{P}$ under ξ and is denoted by $\mathcal{D}(\xi)$, i.e., $\mathcal{D}(\xi) = \xi_*(\mathbb{P})$. If we need to emphasise that the distribution of a random variable is considered with respect to a probability measure μ, then we write $\mathcal{D}_\mu(\xi)$. An X-valued *random process* is defined as a collection of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a family of X-valued random variables $\{\xi_t\}$ on Ω (where t varies in $\mathbb{R}_+$ or $\mathbb{Z}_+$). If the underlying probability space is equipped with a filtration $\mathcal{F}_t$, then we shall say that the process ξ_t is *adapted to* $\mathcal{F}_t$ if ξ_t is $\mathcal{F}_t$-measurable for any $t \geq 0$. Finally, a random process ξ_t defined on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ is said to be *progressively measurable* if for any $t \geq 0$ the map $(s, \omega) \mapsto \xi_s(\omega)$ from $[0, t] \times \Omega$ to X is measurable. It is clear that if t varies in $\mathbb{Z}_+$, then these two concepts coincide.

1.2.2 Convergence of integrals

In what follows, we shall systematically use well-known results on the passage to the limit under Lebesgue's integrals. For the reader's convenience, we state them here without proof, referring the reader to section 4.3 of [Dud02].

Let $(\Omega, \mathcal{F})$ be a measurable space, let μ be an arbitrary σ-finite measure on it (so $\mu(\Omega) \le \infty$), and let $f_n : \Omega \to \mathbb{C}$ be a sequence of integrable functions. The following result, called *Lebesgue's theorem on dominated convergence*, gives a sufficient condition for the convergence of the integrals of f_n to that of the limit function.

Theorem 1.2.1 *Assume that $\{f_n\}_{n\ge 1}$ is a sequence of functions that converge μ-almost surely and satisfy the inequality*

$$|f_n(\omega)| \le g(\omega) \quad \text{for } \mu\text{-almost every } \omega \in \Omega, \tag{1.10}$$

where $g : \Omega \to \mathbb{R}_+$ is a μ-integrable function. Then

$$\lim_{n\to\infty} \int_\Omega f_n d\mu = \int_\Omega \Big(\lim_{n\to\infty} f_n\Big)\, d\mu. \tag{1.11}$$

In the case when the functions f_n are real-valued and form a monotone sequence, the bound (1.10) can be replaced by a weaker condition, which a posteriori turns out to be equivalent to the former. Namely, we have the following *monotone convergence theorem*.

Theorem 1.2.2 *Let $f_n : \Omega \to \mathbb{R}$ be a non-decreasing (or non-increasing) sequence that converges μ-almost surely and satisfies the condition*

$$\sup_{n\ge 1} \left| \int_\Omega f_n d\mu \right| < \infty.$$

Then relation (1.11) *holds.*

Finally, the following result, called *Fatou's lemma*, is useful when estimating the integral of the limit for a sequence of non-negative functions.

Theorem 1.2.3 *Let $f_n : \Omega \to \mathbb{R}_+$ be an arbitrary sequence of μ-integrable functions. Then*

$$\int_\Omega \Big(\liminf_{n\to\infty} f_n\Big)\, d\mu \le \liminf_{n\to\infty} \int_\Omega f_n d\mu.$$

In particular, the three theorems above apply if Ω is the set $\mathbb{N}$ of non-negative integers with the counting measure. In this case, they describe passage to the limit for sums of infinite series.

1.2.3 Metrics on the space of probabilities and convergence of measures

In what follows, we denote by X a Polish space with a metric d_X. Define $C_b(X)$ as the space of bounded continuous functions $f : X \to \mathbb{R}$ endowed with the norm

$$\|f\|_\infty = \sup_{u \in X} |f(u)|,$$

and denote by $L_b(X)$ the space of bounded Lipschitz functions on X. That is, of functions $f \in C_b(X)$ for which

$$\mathrm{Lip}(f) := \sup_{u_1, u_2 \in X} \frac{|f(u_1) - f(u_2)|}{\mathrm{dist}_X(u_1, u_2)} < \infty.$$

The space $L_b(X)$ is endowed with the norm

$$\|f\|_L = \|f\|_\infty + \mathrm{Lip}(f).$$

Note that $C_b(X)$ and $L_b(X)$ are Banach spaces with respect to the corresponding norms. The following exercise summarises some further properties of these spaces.

Exercise 1.2.4 Let X be a Polish space.

(i) Prove that $C_b(X)$ is separable if and only if X is compact.
(ii) Prove that $L_b(X)$ is not separable for the space $X = [0, 1]$ with the usual metric.

Hint: To prove that $C_b(X)$ is separable for a compact metric space X, use the existence of a finite ε-net and a partition of unity on X. To show that if X is not compact, then $C_b(X)$ is not separable, use the existence of a sequence $\{x_k\} \subset X$ such that $\mathrm{dist}_X(x_k, x_m) \geq \varepsilon > 0$. Finally, to prove (ii), construct a continuum $\{\varphi_\alpha\} \subset L^\infty(X)$ such that the distance between any two functions is equal to 1, and use the integrals of φ_α.

Let us denote by $\mathcal{P}(X)$ the set of probability measures on $(X, \mathcal{B}(X))$ and by $\mathcal{P}_1(X)$ the subset of those measures $\mu \in \mathcal{P}(X)$ for which

$$\mathfrak{m}_1(\mu) := \int_X \mathrm{dist}_X(u, u_0)\, \mu(du) < \infty, \tag{1.12}$$

where $u_0 \in X$ is an arbitrary point. The triangle inequality implies that the class $\mathcal{P}_1(X)$ does not depend on the choice of u_0. We shall need the following three metrics.

Total variation distance:

$$\|\mu_1 - \mu_2\|_{\mathrm{var}} := \frac{1}{2} \sup_{\substack{f \in C_b(X) \\ \|f\|_\infty \leq 1}} \big|(f, \mu_1) - (f, \mu_2)\big|, \quad \mu_1, \mu_2 \in \mathcal{P}(X). \tag{1.13}$$

This is the distance induced on $\mathcal{P}(X)$ by its embedding into the space dual to $C_b(X)$. It can be extended to probability measures on an *arbitrary* measurable space; see Remark 1.2.8 below.

Dual-Lipschitz distance:

$$\|\mu_1 - \mu_2\|_L^* := \sup_{\substack{f \in L_b(X) \\ \|f\|_L \le 1}} \big|(f, \mu_1) - (f, \mu_2)\big|, \quad \mu_1, \mu_2 \in \mathcal{P}(X). \tag{1.14}$$

This is the distance induced on $\mathcal{P}(X)$ by its embedding into the space dual to $L_b(X)$.

Kantorovich distance:

$$\|\mu_1 - \mu_2\|_K := \sup_{\substack{f \in L_b(X) \\ \mathrm{Lip}(f) \le 1}} \big|(f, \mu_1) - (f, \mu_2)\big|, \quad \mu_1, \mu_2 \in \mathcal{P}_1(X). \tag{1.15}$$

Exercise 1.2.5 Show that the symmetric functions (1.13)–(1.15) define metrics on the sets $\mathcal{P}(X)$ and $\mathcal{P}_1(X)$. *Hint:* The only non-trivial point is that if the measures μ_1 and μ_2 satisfy the relation $\|\mu_1 - \mu_2\|_L^* = 0$, then $\mu_1 = \mu_2$. This can be done with the help of the monotone class technique; see Corollary A.1.3 in the appendix.

An immediate consequence of definitions (1.13)–(1.15) and the inequalities $\|f\|_\infty \le \|f\|_L$ and $\mathrm{Lip}(f) \le \|f\|_L$ is that

$$\|\mu_1 - \mu_2\|_L^* \le 2\|\mu_1 - \mu_2\|_{\mathrm{var}} \quad \text{for } \mu_1, \mu_2 \in \mathcal{P}(X), \tag{1.16}$$

$$\|\mu_1 - \mu_2\|_L^* \le \|\mu_1 - \mu_2\|_K \quad \text{for } \mu_1, \mu_2 \in \mathcal{P}_1(X). \tag{1.17}$$

Furthermore, if the space X is bounded, that is, there is an element $u_0 \in X$ and a constant $d_0 > 0$ such that

$$\mathrm{dist}_X(u, u_0) \le d_0 \quad \text{for all } u \in X,$$

then, for any function $f \in C_b(X)$ vanishing at $u_0 \in X$, we have

$$\|f\|_\infty \le d_0 \,\mathrm{Lip}(f),$$

where the right-hand side may be infinite. It follows that in this case

$$\|\mu_1 - \mu_2\|_K \le 2d_0\|\mu_1 - \mu_2\|_{\mathrm{var}} \quad \text{for } \mu_1, \mu_2 \in \mathcal{P}_1(X).$$

It turns out that the distance $\|\cdot\|_L^*$ is equivalent to the one obtained by replacing $L_b(X)$ in (1.14) with the space of bounded Hölder-continuous functions. Namely, for $\gamma \in (0, 1)$ we denote by $C_b^\gamma(X)$ the space of continuous functions

$f : X \to \mathbb{R}$ such that

$$|f|_\gamma := \|f\|_\infty + \sup_{0<\mathrm{dist}_X(u,v)\le 1} \frac{|f(u)-f(v)|}{\mathrm{dist}_X(u,v)^\gamma} < \infty.$$

Let us set

$$\|\mu_1-\mu_2\|_\gamma^* := \sup_{\substack{f\in C_b^\gamma(X)\\ |f|_\gamma\le 1}} \big|(f,\mu_1)-(f,\mu_2)\big|, \quad \mu_1,\mu_2\in\mathcal{P}(X). \tag{1.18}$$

Proposition 1.2.6 *For any $\gamma\in(0,1)$ and $\mu_1,\mu_2\in\mathcal{P}(X)$, we have*

$$\|\mu_1-\mu_2\|_L^* \le \|\mu_1-\mu_2\|_\gamma^* \le 5\big(\|\mu_1-\mu_2\|_L^*\big)^{\frac{1}{2-\gamma}}.$$

Proof The lower bound of the inequality is obvious, and therefore we shall confine ourselves to the proof of the upper bound. For any continuous function $f : X \to \mathbb{R}$, we define an approximation for it by the relation

$$f_\varepsilon(u) = \inf_{v\in X}\big(\varepsilon^{-1}d(u,v)+f(v)\big), \quad u\in X, \tag{1.19}$$

where $\varepsilon > 0$ is an arbitrary constant. It is a matter of direct verification to show that if $f\in C_b^\gamma(X)$ and $\|f\|_\gamma\le 1$, then

$$\|f_\varepsilon\|_L \le 1+\varepsilon^{-1}, \qquad 0\le f(u)-f_\varepsilon(u)\le \varepsilon^{\frac{1}{1-\gamma}} \quad \text{for } u\in X. \tag{1.20}$$

We now fix $\delta>0$ and find a function $f\in C_b^\gamma(X)$ with $\|f\|_\gamma\le 1$ such that

$$\|\mu_1-\mu_2\|_\gamma^* \le |(f,\mu_1)-(f,\mu_2)|+\delta. \tag{1.21}$$

It follows from (1.20) that, for any $\varepsilon>0$, we have

$$\begin{aligned}|(f,\mu_1)-(f,\mu_2)| &\le |(f_\varepsilon-f,\mu_1)|+|(f_\varepsilon-f,\mu_2)|+|(f_\varepsilon,\mu_1)-(f_\varepsilon,\mu_2)|\\ &\le 2\varepsilon^{\frac{1}{1-\gamma}}+\big(1+\varepsilon^{-1}\big)\|\mu_1-\mu_2\|_L^*.\end{aligned}$$

Choosing $\varepsilon=(\|\mu_1-\mu_2\|_L^*)^{\frac{1-\gamma}{2-\gamma}}$ and noting that $\|\mu_1-\mu_2\|_L^*\le 2$, we get

$$|(f,\mu_1)-(f,\mu_2)| \le 5\big(\|\mu_1-\mu_2\|_L^*\big)^{\frac{1}{2-\gamma}}.$$

Combining this with (1.21) and recalling that $\delta>0$ was arbitrary, we arrive at the required assertion. □

The following proposition gives an alternative description of the total variation distance and provides some formulas for calculating it.

Proposition 1.2.7 *For any $\mu_1,\mu_2\in\mathcal{P}(X)$, we have*

$$\|\mu_1-\mu_2\|_{\mathrm{var}} = \sup_{\Gamma\in\mathcal{B}(X)} |\mu_1(\Gamma)-\mu_2(\Gamma)|. \tag{1.22}$$

Furthermore, if μ_1 and μ_2 are absolutely continuous with respect to a given measure $m\in\mathcal{P}(X)$, then

$$\|\mu_1-\mu_2\|_{\mathrm{var}} = \frac{1}{2}\int_X \big|\rho_1(u)-\rho_2(u)\big|\,dm = 1-\int_X(\rho_1\wedge\rho_2)(u)\,dm, \tag{1.23}$$

where $\rho_i(u)$ is the density of μ_i with respect to m.