Chapter 1
Classical physics on a finite phase space

Throughout this chapter, X is a *finite set*, playing the role of the configuration space of some physical system, or, equivalently (as we shall see), of its pure state space (in the continuous case, X will be the phase space rather than the configuration space). One should not frown upon finite sets: for example, the configuration space of N bits is given by $X = \underline{2}^{\underline{N}}$, where for arbitrary sets Y and Z, the set Y^Z consists of all functions $x : Z \to Y$, and for any $N \in \mathbb{N}$ we write $\underline{N} = \{1, 2, \ldots, N\}$ (although, following the computer scientists, $\underline{2}$ usually denotes $\{0, 1\}$). More generally, if one has a lattice $\Lambda \subset \mathbb{Z}^d$ and each site is the home of some classical object (say a "spin") that may assume N different configurations, then $X = \underline{N}^{\Lambda}$, in that $x : \Lambda \to \underline{N}$ describes the configuration in which the "spin" at site $\mathbf{n} \in \Lambda$ takes the value $x(\mathbf{n}) \in \underline{N}$.

Although the setting is *a priori* deterministic, in that (knowing) some point $x \in X$ in its guise as a pure state at least in principle determines everything (there is to say), the mathematical language will be probabilistic. Even within the confines of classicality this allows one to do statistical physics, and as such it also sheds light on e.g. the special status of x as an extreme probability measure (see below). Furthermore, the use of this language may be motivated by the goal of describing classical and quantum mechanics as analogously as possible at this elementary level.

The following concepts play a central role in this chapter. Recall that the power set $\mathscr{P}(X)$ of X is the set of all subsets of X (for finite X, these are all measurable).

Definition 1.1. *1. An* **event** *is a subset $U \subseteq X$, i.e., $U \in \mathscr{P}(X)$.*

2. A **probability distribution** *on X is a function $p : X \to [0,1]$ such that $\sum_x p(x) = 1$.*

3. A **probability measure** *on X is a function $P : \mathscr{P}(X) \to [0,1]$ such that $P(X) = 1$ and $P(U \cup V) = P(U) + P(V)$ whenever $U \cap V = \emptyset$.*

4. For a given probability measure P on X, and an event $V \subseteq X$ such that $P(V) > 0$, the **conditional probability** *$P(U|V)$ of U given V is defined by*

$$P(U|V) = \frac{P(U \cap V)}{P(V)}. \tag{1.1}$$

5. A **random variable** *on X is a function $f : X \to \mathbb{R}$.*

6. The **spectrum** *of a random variable f is the subset $\sigma(f) = \{f(x) \mid x \in X\}$ of $\mathbb{R}$.*

K. Landsman, *Foundations of Quantum Theory*,
Fundamental Theories of Physics 188, DOI 10.1007/978-3-319-51777-3_1

1.1 Basic constructions of probability theory

Probability distributions p and probability measures P determine each other by

$$P(U) = \sum_{x \in U} p(x); \tag{1.2}$$

$$p(x) = P(\{x\}), \tag{1.3}$$

but this is peculiar to finite sets (in general, probability *measures* will be primary). Two special classes of probability measures and of random variables stand out:

- Each $y \in X$ defines a probability distribution p_y by $p_y(x) = \delta_{xy}$, or explicitly $p_y(x) = 1$ if $x = y$ and $p_y(x) = 0$ if $x \neq y$; for the corresponding probability measure one has $P_y(U) = 1$ if $y \in U$ and $P_y(U) = 0$ if $y \notin U$.
- Each event $U \subset X$ defines a random variable 1_U (i.e., the ***characteristic function*** of U) by $1_U(x) = 1$ if $x \in U$ and $1_U(x) = 0$ if $x \notin U$. Clearly, $\sigma(1_U) = \{0\}$ when $U = \emptyset$, $\sigma(1_U) = \{1\}$ when $U = X$, and $\sigma(1_U) = \{0,1\}$ otherwise. Note that $1_U(x) = P_x(U)$. Conversely, any random variable f with spectrum $\sigma(f) \subseteq \{0,1\}$ is given by $f = 1_U$ for some $U \subseteq X$; just take $U = \{x \in X \mid f(x) = 1\}$. Such functions may be construed as yes-no questions to the system (i.e. $f = 1$ versus $f = 0$) and will lie at the basis of the logical interpretation of the theory (cf. §1.4).

The single most important construction in probability theory is as follows.

Theorem 1.2. *A probability distribution p on X and a random variable $f : X \to \mathbb{R}$ jointly yield a probability distribution p_f on the spectrum $\sigma(f)$ by means of*

$$p_f(\lambda) = \sum_{x \in X \mid f(x) = \lambda} p(x). \tag{1.4}$$

In terms of the corresponding probability measure P on X, one has

$$p_f(\lambda) = P(f = \lambda), \tag{1.5}$$

where $f = \lambda$ denotes the event $\{x \in X \mid f(x) = \lambda\}$ in X. Similarly, the probability measure P_f on $\sigma(f)$ corresponding to the probability distribution p_f is given by

$$P_f(\Delta) = P(f \in \Delta), \tag{1.6}$$

where $\Delta \subseteq \sigma(f)$ and $f \in \Delta$ denotes the event $\{x \in X \mid f(x) \in \Delta\}$ in X.

The proof is trivial. Instead of $f = \lambda$, the notation $f^{-1}(\{\lambda\})$ might be used, and similarly, $f^{-1}(\Delta)$ is the same as $f \in \Delta$. If $\lambda \in \sigma(f)$ is non-degenerate in that there is exactly one $x_\lambda \in X$ such that $f(x_\lambda) = \lambda$, then one simply has $P(f = \lambda) = p(x_\lambda)$.

For example, combining both our special cases $P = P_y$ and $f = 1_U$ above yields

$$P_y(1_U = 1) = 1 \text{ and } P_y(1_U = 0) = 0 \text{ if } y \in U; \tag{1.7}$$

$$P_y(1_U = 1) = 0 \text{ and } P_y(1_U = 0) = 1 \text{ if } y \notin U. \tag{1.8}$$

Given some probability measure P, the ***expectation value*** $E_P(f)$ and the ***variance*** $\Delta_P(f)$ of a random variable f with respect to P are defined by, respectively,

$$E_P(f) = \sum_{x\in X} f(x)p(x); \tag{1.9}$$

$$\Delta_P(f) = E_P(f^2) - E_P(f)^2. \tag{1.10}$$

A simple calculation shows that E_P may be written directly in terms of P itself as

$$E_P(f) = \sum_{\lambda\in\sigma(f)} P(f=\lambda)\cdot\lambda. \tag{1.11}$$

Note that $\Delta_P(f) \geq 0$. The special role of the point measures P_y may now be clarified:

Proposition 1.3. *A probability measure P takes the form $P = P_y$ for some $y \in X$ iff $\Delta_P(f) = 0$ for all random variables $f : X \to \mathbb{R}$.*

Proof. For "⇒", we compute $E_{P_y}(f) = f(y)$, and hence $E_{P_y}(f^2) = f(y)^2$. In the opposite direction, take $f = p_y$, so that $f^2 = f$ and hence $\Delta_P(f) = p(y) - p(y)^2$. The assumption $\Delta_P(f) = 0$ for each f implies that either $p(y) = 0$ or $p(y) = 1$ for each $y \in X$. Definition 1.1.2 then implies that $p(y) = 1$ for exactly one $y \in X$. □

More generally, a collection $f_1, \ldots, f_n$ of n random variables and a (single) probability distribution p on X jointly define a probability distribution $p_{f_1,\ldots,f_n}$ on the product $\sigma(f_1) \times \cdots \times \sigma(f_n)$ of the individual spectra by

$$p_{f_1\cdots f_n}(\lambda_1, \ldots, \lambda_n) = \sum_{x\in X|f_1(x)=\lambda_1,\ldots,f_n(x)=\lambda_n} p(x). \tag{1.12}$$

Once again, this may be rewritten as

$$p_{f_1\cdots f_n}(\lambda_1, \ldots, \lambda_n) = P(f_1 = \lambda_1, \ldots, f_n = \lambda_n), \tag{1.13}$$

where the argument of P denotes the intersection $\cap_{k=1}^{n}(f_k = \lambda_k)$, i.e.,

$$P(f_1 = \lambda_1, \ldots, f_n = \lambda_n) = \{x \in X \mid f_1(x) = \lambda_1, \ldots, f_n(x) = \lambda_n\}. \tag{1.14}$$

Simple calculations then yield results for the so-called ***marginal distributions***, like

$$\sum_{\lambda_{l+1}\in\sigma(f_{l+1}),\ldots,\lambda_n\in\sigma(f_n)} P(f_1 = \lambda_1, \ldots, f_n = \lambda_n) = P(f_1 = \lambda_1, \ldots, f_l = \lambda_l), \tag{1.15}$$

where $1 \leq l < n$. The above constructions also apply to the corresponding conditional probabilities: given m additional random variables $a_1, \ldots, a_m$, one has

$$\sum_{\lambda_{l+1}\in\sigma(f_{l+1}),\ldots,\lambda_n\in\sigma(f_n)} P(f_1 = \lambda_1, \ldots, f_n = \lambda_n | a_1 = \alpha_1, \ldots a_m = \alpha_m) \tag{1.16}$$

$$= P(f_1 = \lambda_1, \ldots, f_l = \lambda_l | a_1 = \alpha_1, \ldots a_m = \alpha_m). \tag{1.17}$$

1.2 Classical observables and states

Given a finite set X, we may form the set $C(X)$ of all complex-valued functions on X, enriched with the structure of a complex vector space under pointwise operations:

$$(\lambda \cdot f)(x) = \lambda f(x) \ (\lambda \in \mathbb{C}); \tag{1.18}$$

$$(f+g)(x) = f(x)+g(x). \tag{1.19}$$

We use the notation $C(X)$ with some foresight, anticipating the case where X is no longer finite, but in any case, since for the moment it is, every function is continuous. Moreover, the vector space structure on $C(X)$ may be extended to that of a commutative algebra (where, by convention, all our algebras are associative and are defined over the complex scalars) by defining multiplication pointwisely, too:

$$(f \cdot g)(x) = f(x)g(x). \tag{1.20}$$

Note that this algebra has a unit 1_X, i.e., the function identically equal to 1.

For finite X, this structure suffices for X to be recovered from $C(X)$, as follows.

Definition 1.4. *The* **Gelfand spectrum** $\Sigma(A)$ *of a (complex) algebra A is the set of all nonzero linear maps $\omega : A \to \mathbb{C}$ that satisfy $\omega(fg) = \omega(f)\omega(g)$.*

These are, of course, precisely the nonzero algebra homomorphisms from A to $\mathbb{C}$.

Proposition 1.5. *The Gelfand spectrum $\Sigma(C(X))$ is isomorphic (as a set) to X.*

Proof. Each $x \in X$ defines a map $\omega_x : C(X) \to \mathbb{C}$ by $\omega_x(f) = f(x)$. One obviously has $\omega_x \in \Sigma(C(X))$, so we have a map $X \to \Sigma(C(X))$, $x \mapsto \omega_x$. We show that this map is a bijection. Injectivity is easy: if $\omega_x = \omega_y$, then $f(x) = f(y)$ for each $f \in C(X)$, so taking $f = \delta_z$ for each $z \in X$ gives $x = y$ (here $\delta_z(x) = \delta_{xz}$). To prove surjectivity, we note that since $C(X)$ is finite-dimensional as a vector space, with basis $(\delta_y)_{y\in X}$, each linear functional $\omega : C(X) \to \mathbb{C}$ takes the form

$$\omega(f) = \sum_x \mu(x) f(x), \tag{1.21}$$

for some function $\mu : X \to \mathbb{C}$. For $\omega \in \Sigma(C(X))$, find some $z \in X$ for which $\mu(z) \neq 0$ (this has to exist, as $\omega \neq 0$). For arbitrary $w \in X$, imposing $\omega(\delta_w \delta_z) = \omega(\delta_w)\omega(\delta_z)$ enforces $\mu = \delta_z$ (which also shows that z is unique), and hence $\omega = \omega_z$. □

The physically relevant set $R(X)$ of all real-valued functions on X is obviously a real vector space inside $C(X)$. To recover it algebraically, we equip $C(X)$ with an ***involution***, which on an arbitrary (not necessarily commutative) algebra A is defined as an anti-linear anti-homomorphism that squares to id_A, i.e., a linear map $* : A \to A$ (written $a \mapsto a^*$) that satisfies $(\lambda a)^* = \overline{\lambda} a^*$, $(ab)^* = b^* a^*$, and $a^{**} = a$. In our case $A = C(X)$, which is commutative, the latter property simply becomes $(fg)^* = f^* g^*$. In any case, we define this involution by pointwise complex conjugation, i.e.,

$$f^*(x) = \overline{f(x)}. \tag{1.22}$$

We evidently recover the real-valued functions in the involutive algebra $C(X)$ as

$$R(X) \equiv C(X)_{\text{sa}} = \{f \in C(X) \mid f^* = f\}. \tag{1.23}$$

Finally, although we do not need this yet, we note that $C(X)$ has a natural ***norm***

$$\|f\|_\infty = \sup_{x \in X}\{|f(x)|\}. \tag{1.24}$$

These structures turn $C(X)$ into a ***commutative C*-algebra*** (cf. Definition C.1).

Definition 1.6. *The* **algebra of observables** *of the physical system described by the phase space X is $C(X)$, seen as a (commutative)* C*-algebra *in the above way.*

Thence elements of $C(X)$ are called ***observables*** (a term that really should be applied only to its self-adjoint elements, i.e., those satisfying $f^* = f$).

We have thus equipped the *random variables* on X with enough structure to recover X itself, and now turn to the other side of the coin, viz. the *probability measures* on X. Here the relevant mathematical structure is that of a *compact convex set*, a concept we only need to define in the context of an ambient (real) vector space.

Definition 1.7. *A subset K of a (real or complex) vector space V is called* **convex** *if the straight line segment between any two points on K lies in K. Expressed formally, this means that whenever $v, w \in K$ and $t \in (0,1)$, one has $tv + (1-t)w \in K$.*

The following probabilistic reformulation of this notion is very useful.

Proposition 1.8. *A set $K \subset V$ is convex iff for any k, given k probabilities $(t_1, \ldots, t_k)$ (i.e., $t_i \geq 0$ and $\sum_i t_i = 1$) and k points $(v_1, \ldots, v_k)$ in K, one has $\sum_{i=1}^k t_i \cdot v_i \in K$.*

Proof. Taking $k = 2$ recovers Definition 1.7 from its probabilistic version. Conversely, one uses induction on k, using the identity (assuming $0 < t_k < 1$):

$$t_1 v_1 + \cdots + t_k v_k = (1 - t_k)\left(\frac{t_1}{1 - t_k} v_1 + \cdots + \frac{t_{k-1}}{1 - t_k} v_{k-1}\right) + t_k v_k. \qquad \square$$

Any linear subspace of V is trivially convex, as is any translate thereof (i.e., any ***affine*** subspace of V). Another, much more important example is the ***convex hull*** $\text{co}(S)$ of any subset $S \subset V$; noting that the intersection of any family of convex sets is again convex, $\text{co}(S)$ may be defined as the intersection of all convex subsets of V that contain S, or, equivalently, as the smallest convex subset of V that contains S (whose existence is guaranteed by the previous remark). Proposition 1.8 then yields

$$\text{co}(S) = \left\{\sum_{i=1}^k t_i \cdot v_i \mid k \in \mathbb{N}, (v_1, \ldots, v_k) \in S^k, t_i \geq 0, \sum_i t_i = 1\right\}. \tag{1.25}$$

In particular, if $S = \{v_1, \ldots, v_k\}$ is a finite set, then one simply has

$$\text{co}(\{v_1, \ldots, v_k\}) = \left\{\sum_{i=1}^k t_i \cdot v_i \mid t_i \geq 0, \sum_i t_i = 1\right\}. \tag{1.26}$$

The convex hull of any finite set of points in $\mathbb{R}^{n+1}$ is called a ***convex polytope***. Such convex sets are closed and bounded (since none of the $t_i \geq 0$ can walk away too far without violating the condition $\sum_i t_i = 1$), and hence are compact. In particular,

$$\Delta_n = \{x \in \mathbb{R}^{n+1} \mid x_i \geq 0, \sum_i x_i = 1\} \tag{1.27}$$

is a convex polytope called a ***simplex***. For example, Δ_1 is the line segment from $(0,1)$ to $(1,0)$ in $\mathbb{R}^2$. We would like to say that Δ_1 is "isomorphic" to the unit interval $[0,1]$, so we define two convex sets K_1, K_2 to be ***isomorphic*** (as such) if there is a bijection $f : K_1 \to K_2$ that is ***affine***, in that for $t \in (0,1)$ and $v_1, v_2 \in K_1$, we have

$$f(tv_1 + (1-t)v_2) = tf(v_1) + (1-t)f(v_2). \tag{1.28}$$

Then the function $f : \Delta_1 \to [0,1]$ given by $f(\lambda, 1-\lambda) = \lambda$, where $\lambda \in [0,1]$, will do. Similarly, $\Delta_2 \subset \mathbb{R}^3$ is isomorphic to any equilateral triangle in $\mathbb{R}^2$ with sides of unit length, whereas Δ_3 is just the tetrahedron (which is one of the five Platonic solids).

There are many other convex polytopes (cf. §B.11), but simplices are of prime importance for us, since Δ_n is isomorphic to the set $\text{Pr}(X)$ of all probability distributions on a set $X = \{0, \ldots, n\}$ with $n+1$ points; the identification $\text{Pr}(X) \ni p \leftrightarrow x \in \Delta_n$ is given by $x_i = p(i+1)$. In particular, we see that for any finite set X, $\text{Pr}(X)$ is a compact convex set. This is also clear from Definitions 1.1 and 1.7 (and will even be true for general compact phase spaces X, cf. Corollary B.17 and §C.25).

Definition 1.9. *The* **state space** *of the physical system described by a (finite) space X is the set* $\text{Pr}(X)$ *of all probability measures on X (or, equivalently, of all probability distributions on X), seen as a* compact convex set.

Thus a probability measure (or distribution) on X is often called a ***state*** (of the physical system described by X). The operation of passing from states $P, Q \in \text{Pr}(X)$ to a new state $tP + (1-t)Q \in \text{Pr}(X)$, where $t \in (0,1)$ as usual, or, more generally, from a (finite) family of states (P_i) and a set (t_i) of probabilities (i.e., $t_i \geq 0$ and $\sum_i t_i = 1$) to the convex sum $\sum_i t_i P_i$, is called ***mixing***.

It is possible to recover X from its associated state space $\text{Pr}(X)$, as follows.

Definition 1.10. *The* **(extreme) boundary** $\partial_e K$ *of a convex set K consists of all points $v \in K$ satisfying the following condition:*

if $v = tw + (1-t)x$ for certain $w, x \in K$ and $t \in (0,1)$, then $v = w = x$.

Elements $v \in \partial_e K$ of the boundary are called **extreme points** *of K.*

We will now compute the boundary of $\text{Pr}(X)$. The result may be expressed by

$$\partial_e \Delta_n = \{\mathbf{e}_1, \ldots . \mathbf{e}_{n+1}\}, \tag{1.29}$$

where $(\mathbf{e}_1, \ldots . \mathbf{e}_{n+1})$ is the standard basis of $\mathbb{R}^{n+1}$ (i.e., $\mathbf{e}_1 = (1, 0, \ldots, 0)$, etc.). However, we will give a direct probabilistic proof. We already noted the special probability measures P_x, $x \in X$. The association $x \mapsto P_x$ defines a map from X to $\text{Pr}(X)$.

Proposition 1.11. *The set X is isomorphic to the boundary $\partial_e\mathrm{Pr}(X)$ through $x \mapsto P_x$.*

Proof. It is convenient to work with probability distributions p rather than probability measures P. First, $x \mapsto p_x$ is trivially injective from X to $\mathrm{Pr}(X)$: if $x \neq y$ then $p_x(x) = 1$ whereas $p_y(x) = 0$, so $p_x \neq p_y$. Second, $p_x \in \partial_e\mathrm{Pr}(X)$. For suppose one has $p_x = tp + (1-t)q$ for some $p, q \in \mathrm{Pr}(X)$ and $t \in (0,1)$. Hence $p_x(y) = tp(y) + (1-t)q(y)$. Taking $y \neq x$ yields $p(y) = q(y) = 0$, so that $p = q = p_x$. Consequently, $X \subseteq \partial_e\mathrm{Pr}(X)$.

The converse inclusion is (contrapositively) equivalent to the property that for any $p \neq p_x$ (for all x), there are q and r, $q \neq r$, and $t \in (0,1)$, with $p = tq + (1-t)r$. Indeed, if $p \neq p_x$, there is some $x_0 \in X$ with $0 < p(x_0) < 1$. Now define q, r, and t by $q(x_0) = 1$ and $q(x) = 0$ for all $x \neq x_0$, $r(x_0) = 0$ and $r(x) = p(x)/(1 - p(x_0))$, and finally $t = p(x_0)$. Then $p = tq + (1-t)r$ and $q \neq r$. □

The simplest example would be $X = \{0,1\}$, so that $\mathrm{Pr}(X) \cong [0,1]$ by mapping the distribution $p \in \mathrm{Pr}(X)$ to $p(1)$. Since one may directly verify that $\partial_e[0,1] = \{0,1\}$, under the above isomorphism one therefore has $\partial_e\mathrm{Pr}(X) \cong \{0,1\}$. Analogously, $\partial_e(0,1) = \emptyset$, so that the boundary of a convex set may apparently be empty. Hence we see that one remarkable ingredient of Proposition 1.11 lies in the claim that the convex set $\mathrm{Pr}(X)$ actually *has* a (nonempty) boundary! This is no accident: by the Krein-Milman Theorem (cf. §B.10), this is true for any *compact* convex set (which is consistent with the counterexample just given). For example in quantum mechanics we will encounter the case of $K = B^3$ (i.e. the closed unit ball in $\mathbb{R}^3$) as the state space of a qubit, whose (extreme) boundary is the two-sphere S^2, cf. Proposition 2.9. Something similar is true in any dimension, but beware of surprises: if $K = \Delta_2$ is an equilateral triangle in the plane, then its *extreme* boundary $\partial_e K$ consists of the *vertices* of K (whereas its *faces* form the *geometric* boundary of the triangle).

The general problem arises whether some point $v \in K$ of a compact convex set K may be written as a convex sum (or, more generally, an integral) of extreme points of K, and if so, to what extent this ***extremal decomposition***

$$v = \sum_{i \in I} t_i v_i,\ t_i \geq 0, \sum_i t_i = 1,\ v_i \in \partial_e K, \tag{1.30}$$

which for simplicity has been assumed to be a finite sum here, is unique. Without proof, we state a general result of convexity theory, called ***Caratheodory's Theorem***:

Theorem 1.12. *If K is a nonempty compact convex subset of $\mathbb{R}^n$, then $\partial_e K \neq \emptyset$, and each point of K is a convex sum of at most $n+1$ points in $\partial_e K$.*

If $K = \Delta_n$, then this sum generically has $n+1$ points and is unique. Probabilistically:

Proposition 1.13. *If X is finite, then any probability measure $P \in \mathrm{Pr}(X)$ may be written in a unique way as a* finite *mixture of extreme probability measures, viz.*

$$P = \sum_{x \in X} t_x P_x. \tag{1.31}$$

Proof. Take $t_x = P(\{x\})$ in the sense of Definition 1.1, or, equivalently, $t_x = E_P(\delta_x)$ in the sense of (1.9). To see that this decomposition is unique, use Proposition 1.11, i.e. $\partial_e \Pr(X) \cong X$, in (1.30) to force $I = X$ and apply both sides of (1.31) to δ_x. □

The state space and the algebra of observables may also be defined in terms of each other. We start with the (re)construction of states from observables, where the following definition and proposition may leave a hybrid impression. The rationale behind our approach is that for many purposes it is easier to work with the *complex* algebra $C(X)$, but on the other hand, compact convex sets are most naturally defined in terms of *real* vector spaces. Fortunately, it is easy to switch between the two: we already know how to obtain the real part $R(X)$ from $C(X)$, see (1.23), and conversely, $C(X)$ is simply the complexification of the real vector space $R(X)$.

Definition 1.14. *A* **state** *on $C(X)$ is a linear map $\omega : C(X) \to \mathbb{C}$ that satisfies:*

1. $\omega(f^2) \geq 0$ *for each $f \in C(X)$ with $f^* = f$ (***positivity***);*
2. $\omega(1_X) = 1$ *(***normalization***).*

The first condition obviously comes down to $\omega(f) \geq 0$ whenever $f \geq 0$ pointwise.

Equivalently, we may define a state on $R(X)$ as a real-linear map $\omega_{\mathbb{R}} : R(X) \to \mathbb{R}$ that satisfies the very same conditions. Indeed, a state $\omega_{\mathbb{R}}$ on $R(X)$ defines a complex-linear map $\omega : C(X) \to \mathbb{C}$ by $\omega(f + ig) = \omega_{\mathbb{R}}(f) + i\omega_{\mathbb{R}}(g)$, where $f, g \in R(X)$. This map satisfies the same conditions of positivity and normalization. Conversely, ω may be restricted to the real part $R(X)$ of $C(X)$, so that there is no real (sic) difference between ω and $\omega_{\mathbb{R}}$. Hence we will use these interchangeably, often even dropping the suffix $\mathbb{R}$ on ω. One advantage of this ability to switch is that a state ω on $C(X)$ may be regarded as an element of the *real* vector space $R(X)^*$. Doing so shows that the terminology of Definitions 1.9 and 1.14 is consistent:

Theorem 1.15. *There is a bijective correspondence between states ω on $C(X)$ and probability measures P on X, given by $\omega \leftrightarrow E_P$, cf.* (1.9) *and* (1.11). *Therefore, as a subset of the (real) vector space $R(X)^*$ of all (real-) linear maps from $R(X)$ to $\mathbb{R}$, the set $S(C(X))$ of all states on $C(X)$ coincides with the set* $\Pr(X)$ *of all probability measures on X. In particular, the state space $S(C(X))$ of $C(X)$ is a compact convex set in $R(X)^*$ (as a finite-dimensional vector space with its usual topology).*

Proof. Given a state ω, define a function $p : X \to \mathbb{R}$ by $p(x) = \omega(\delta_x)$. Since $\delta_x \geq 0$ pointwise, positivity of ω yields $p(x) \geq 0$. Noting that $1_X = \sum_x \delta_x$, normalization then forces $\sum_x p(x) = 1$, so that p is a probability distribution on X. Hence $P \in \Pr(X)$, where P is the probability measure corresponding to p. Conversely, $P \in \Pr(X)$ defines a map $E_P : R(X) \to \mathbb{R}$ by (1.9), which is positive and normalized. Note that compactness and convexity of the set $S(C(X))$ in $R(X)^*$ follow directly from its definition, i.e., even without knowing that it equals $\Pr(X)$. □

Consequently, we may refer to $S(C(X))$ as *the* ***state space*** of $C(X)$ without any ambiguity, and we will always regard state spaces of (unital) C*-algebras A (cf. Appendix C) as compact convex sets $S(A)$, where in the present case $A = C(X)$.

1.3 Pure states and transition probabilities

For any C*-algebra A (with unit), and hence in particular for $A = C(X)$, elements of the boundary $\partial_e S(A)$ are called ***pure states***, and we call

$$P(A) \equiv \partial_e S(A) \tag{1.32}$$

the ***pure state space*** of A. States that are not pure are called ***mixed***.

Theorem 1.16. *One has $P(C(X)) \cong X$, in that the following map is an isomorphism:*

$$X \to P(C(X)),\ x \mapsto \omega_x,\ \omega_x(f) = f(x). \tag{1.33}$$

Proof. Combine Proposition 1.11 and Theorem 1.15. □

For finite X this isomorphism is merely meant as a bijection between sets (and for general compact Hausdorff spaces X it will be a homeomorphism of topological spaces), but we will now introduce some additional structure on pure state spaces that will enrich Theorem 1.16 to an isomorphism of so-called ***sets with a transition probability***. This will be necessary in order to reconstruct the observables from the pure states, but it also clarifies the general probabilistic structure of physics (note that the following definition is unusual in probability theory!).

Definition 1.17. *1. A* **transition probability** *on a set X is a function*

$$\tau : X \times X \to [0,1] \tag{1.34}$$

*that satisfies $\tau(x,y) = 1$ iff $x = y$ and $\tau(x,y) = \tau(y,x)$ (***symmetry***).*

The simplest example of a transition probability (on any set X) is obviously

$$\tau(x,y) = \delta_{xy}. \tag{1.35}$$

The point is that this transition probability may be derived from the classical C*-algebra of observables $C(X)$ by the following formula (assuming X finite):

$$\delta_{xy} = \inf\{f(x) \mid f \in C(X), 0 \leq f \leq 1_X, f(y) = 1\}. \tag{1.36}$$

Indeed, for $x = y$ this is a tautology, whereas for $x \neq y$ the infimum (which is zero) is attained by $f = \delta_y$. In terms of the pure state space $P(C(X))$, which is *isomorphic* to but not *equal* to X, cf. Theorem 1.16, this formula may be written as

$$\delta_{xy} = \inf\{\omega_x(f) \mid f \in C(X), 0 \leq f \leq 1_{C(X)}, \omega_y(f) = 1\}. \tag{1.37}$$

Furthermore (and this is the *real* point, so that we already have to mention it here, ahead of a more detailed treatment in the context of quantum mechanics), the right-hand side of (1.37) may be generalized to any finite-dimensional C*-algebra A by

$$\tau^A(\omega, \omega') = \inf\{\omega(a) \mid a \in A, 0 \leq a \leq 1_A, \omega'(a) = 1\}, \tag{1.38}$$

where $\omega, \omega' \in P(A)$. Since (1.38) clearly generalizes (1.37), for $A = C(X)$ we have

$$\tau^{C(X)}(\omega_x, \omega_y) = \delta_{xy}. \tag{1.39}$$

Note that the symmetry property in Definition 1.17 is not obvious from (1.38), but in the classical case $A = C(C)$ it is true by computation, and the same will hold in quantum theory. To motivate these definitions, we recall that f in (1.37), and likewise a in (1.38), are yes-no question to the system, so that the transition probability $\tau^A(\omega, \omega')$ monitors to what extent the states ω and ω' may be sharply distinguished by asking such questions. If they can, there should be some question a for which $\omega'(a) = 1$ and $\omega(a) = 0$, so that $\tau^A(\omega, \omega')$ (if $\omega \neq \omega'$, of course). As we have seen, in the classical case this can always be done. However, we shall see this is no longer the case in quantum mechanics, where pure states may be thus distinguished iff they correspond to orthogonal unit vectors in Hilbert space. Further motivation for the expression (1.38) is *post hoc*, as it turns out to allow a reconstruction of the vector space of observables A, supplemented by the part of its algebraic structure that determines its logical and probabilistic structure (viz. the ability to form squares, $a \mapsto a^2$) from $P(A)$ *with its associated transition probability*. See Theorem C.179.

First, we develop some theory that puts both classical and quantum mechanics into a more general setting. Notwithstanding the formal incorporation of the former, the underlying Hilbert space thinking will be obvious throughout.

Definition 1.18. *Let (X, τ) be a set with a transition probability.*

1. *A subset $O \subset X$ is* **orthonormal** *if $\tau(x,y) = \delta_{xy}$ for all $x, y \in O$.*
2. *A* **basis** *of a set X with a transition probability τ is an orthonormal family $B \subset X$ such that for each $x \in X$ one has*

$$\sum_{u \in B} \tau(x,u) = 1. \tag{1.40}$$

A basis of a subset $S \subset X$ is an orthonormal family $B \subset S$ such that (1.40) *holds for each $x \in S$. Relative to such a basis B of S, we define $\tau_S : X \to \mathbb{R}$ by*

$$\tau_S(x) = \sum_{u \in B} \tau(x,u). \tag{1.41}$$

As a special case, for $S = \{u\}$ we write $\tau_{\{u\}} \equiv \tau_u$, so that

$$\tau_u(x) = \tau(x,u). \tag{1.42}$$

3. *The* **orthocomplement** *$S^{\perp}$ of some subset $S \subset X$ is defined as*

$$S^{\perp} = \{y \in X \mid \tau(x,y) = 0 \forall x \in S\}. \tag{1.43}$$

4. *A subset $S \subset X$ is* **orthoclosed** *if $S^{\perp\perp} = S$ (where $S^{\perp\perp} = (S^{\perp})^{\perp}$).*
5. *A* **resolution of the identity** *in X is a family of orthogonal orthoclosed subsets $(S_j)_j$ (i.e., $\tau(x_i,x_j) = 0$ if $x_i \in S_i$, $x_j \in S_j$, and $i \neq j$), for which $\sum_j \tau_{S_j} = 1_X$.*

6. *An* **observable** *for the pair (X,τ) is a bounded function $f : X \to \mathbb{R}$ of the form*

$$f = \sum_i c_i \cdot \tau_{y_i},\ c_i \in \mathbb{R},\ y_i \in X. \tag{1.44}$$

The real vector space of such observables is called $\ell^\infty(X,\tau)$.

7. *A* **spectral resolution** *of an observable $f \in \ell^\infty(X,\tau)$ is a decomposition*

$$f = \sum_\lambda \lambda \cdot \tau_{S_\lambda}, \tag{1.45}$$

where $(S_\lambda)_\lambda$ is a resolution of the identity and each $\lambda \in \mathbb{R}$ occurs at most once.

In the present section X is finite, whilst in the following section on quantum mechanics on finite-dimensional Hilbert spaces at least all bases will be finite, so that there are no convergence issues. In general, B may be infinite, in which case (1.40) is defined as the least upper bound of all finite partial sums, and all sums in Definition 1.18 are defined pointwise (i.e., in x). In that case, eq. (1.45) may need to be adapted through limit constructions. Furthermore, one may worry about the basis-dependence of τ_S in (1.41), but fortunately it turns out that in all sets with a transition probability that arise as pure state spaces defined by C*-algebras according to (1.38), the function τ_S is independent of the basis B whenever S is orthoclosed. In that case, spectral resolutions exists and are unique, and one may turn the real vector space $\ell^\infty(X,\tau)$ of part 6 into a ***Jordan algebra*** by defining a product $\circ$ through

$$f^2 = \sum_\lambda \lambda^2 \cdot \tau_{S_\lambda}; \tag{1.46}$$

$$f \circ g = \tfrac{1}{4}((f+g)^2 - (f-g)^2). \tag{1.47}$$

In the classical case this yields the pointwise product (1.20), whereas in quantum mechanics it recovers the anti-commutator. Both are examples of ***Jordan products*** (cf. §C.25), i.e., commutative products $\circ$ satisfying the curious axiom (C.619).

All this trivializes if $\tau = \tau^{C(X)}$ is given by (1.35), where X need not even be finite:

1. Any subset $O \subset X$ is orthonormal.
2. The set $B = X$ itself is the only basis of (X,τ), and analogously $B = S$.
3. The orthocomplement $S^\perp$ is the set-theoretic complement $S^c \equiv X \backslash S$.
4. Hence any subset $S \subset X$ is orthoclosed.
5. Any partition $X = \bigsqcup_j S_j$ yields a resolution of the identity.
6. Any bounded function $f : X \to \mathbb{R}$ is an observable, so that when X is finite,

$$\ell^\infty(X,\tau) = R(X) \equiv C(X,\mathbb{R}); \tag{1.48}$$

7. The spectral resolution (1.45) of f is given (analogously to operator theory) by

$$f = \sum_{\lambda \in \sigma(f)} \lambda \cdot \tau_{f=\lambda}, \tag{1.49}$$

cf. Definition 1.1.5. In particular, spectral resolutions in (1.48) are unique.

1.4 The logic of classical mechanics

Whatever one's route to $C(X,\mathbb{R})$ as the algebra of observables, i.e. either as a starting point or as a derived concept as in (1.48), it determines the logical structure of classical mechanics (we here restrict ourselves to propositional logic). According to the general scheme reviewed in §D.2, apart from the usual logical connectives $\neg$, $\wedge$, $\vee$, and $\rightarrow$ for *not*, *and*, *or*, and *implies*, a propositional theory needs a set Σ_X of *atomic propositions*. These are provided by $C(X,\mathbb{R})$, and Σ_X consist of all expressions $f \in \Delta$ (we expect no confusion between this notation for both *propositions* in logic and *events* in probability theory), where $f : X \to \mathbb{R}$ is a function, and Δ is some subset of $\mathbb{R}$. As we shall see, $f \in \Delta$ is always false if $\Delta \cap \sigma(f) = \emptyset$, so we might as well assume that $\Delta \subseteq \sigma(f)$. We write $f = \lambda$ for $f \in \{\lambda\}$. From these elementary propositions, propositions are constructed inductively using the iterative rules of propositional logic (see §D.2). This produces a set $B_X \equiv B_{\Sigma_X}$ of propositions.

Of course, there are logical relations between our atomic propositions (and hence between elements of B_X). For example, if $\Delta \subset \Delta'$, then $f \in \Delta$ should imply $f \in \Delta'$. Such relations may be formulated as axioms of some propositional theory $\mathscr{T}_X$ describing the logic of classical mechanics. These axioms take the following form:

$$(f \in \Gamma) \rightarrow (g \in \Delta) \ \text{ iff } \ f^{-1}(\Gamma) \subseteq g^{-1}(\Delta). \tag{1.50}$$

This may also be formulated through the notion of ***semantic entailment***. For each $x \in X$, we define a valuation $V_x : \Sigma_X \to \{0,1\}$ (cf. §D.2) by

$$V_x(f \in \Delta) = 1 \ \text{ iff } \ f(x) \in \Delta, \tag{1.51}$$

extended to a map $V_x : B_X \to \{0,1\}$ through the recursive use of truth tables. Defining the semantic entailment relation $\models_X$ on B_X by $\alpha \models_X \beta$ iff $V_x(\alpha) = 1$ implies $V_x(\beta) = 1$ for all $x \in X$, it is easy to see that $\alpha \rightarrow \beta$ as defined in (1.50) iff $\alpha \models_X \beta$.

In order to compute the ensuing Lindenbaum algebra $L_X \equiv L_{\Sigma_X}$, we note that

$$(f \in \Gamma) \leftrightarrow (g \in \Delta) \ \text{ iff } \ f^{-1}(\Gamma) = g^{-1}(\Delta). \tag{1.52}$$

Writing $\sim_X$ for $\sim_{\mathscr{T}_X}$ (which is the equivalence relation given by $\models_X$, too), we find

$$(f \in \Delta) \sim_X (1_{f^{-1}(\Delta)} = 1), \tag{1.53}$$

where we recall that 1_A is the characteristic (or indicator) function of A. Using the truth tables for $\wedge$ and for $\neg$, we also obtain (in terms of the complement $\Delta^c = \mathbb{R}\backslash\Delta$):

$$(f \in \Gamma) \wedge (g \in \Delta) \sim_X (1_{f^{-1}(\Gamma)\cap g^{-1}(\Delta)} = 1); \tag{1.54}$$

$$(\neg f \in \Delta) \sim_X (f \in \Delta^c) \sim_X (1_{f^{-1}(\Delta^c)} = 1). \tag{1.55}$$

Finally, the truth tables yield logical (and hence semantic) equivalences like

$$\alpha \vee \beta \sim_X \neg(\neg\alpha \wedge \neg\beta). \tag{1.56}$$

Combining the specific and the general equivalences (1.53) - (1.56), we have:

Lemma 1.19. *Any proposition in B_X is logically (and semantically) equivalent (relative to X) to one of the form $1_U = 1$, for some event $U \subset X$. Furthermore,*

$$(\neg 1_U = 1) \sim_X (1_{U^c} = 1); \tag{1.57}$$
$$(1_U = 1) \wedge (1_V = 1) \sim_X (1_{U\cap V} = 1); \tag{1.58}$$
$$(1_U = 1) \vee (1_V = 1) \sim_X (1_{U\cup V} = 1). \tag{1.59}$$

Theorem 1.20. *The Lindenbaum algebra L_X is isomorphic (as a Boolean algebra) to the power set $\mathscr{P}(X)$ of X under the map $\varphi : L_X \to \mathscr{P}(X)$ induced by*

$$\varphi([f \in \Delta]_X) = f^{-1}(\Delta). \tag{1.60}$$

In particular, the logical connectives $\neg$, $\wedge$ and $\vee$ (descended to L_X) turn into set-theoretic complementation $(-)^c$, intersection $\cap$, and union $\cup$, respectively, in that

$$\varphi([\neg\alpha]_X) = \varphi([\alpha]_X)^c; \tag{1.61}$$
$$\varphi([\alpha \wedge \beta]_X) = \varphi([\alpha]_X) \cap \varphi([\beta]_X; \tag{1.62}$$
$$\varphi([\alpha \vee \beta]_X) = \varphi([\alpha]_X) \cup \varphi([\beta]_X), \tag{1.63}$$

and φ maps the partial order $\leq$ on L_X into set-theoretic inclusion $\subseteq$, i.e.,

$$[\alpha]_X \leq [\beta]_X \textit{ iff } \varphi([\alpha]_X) \subseteq \varphi([\beta]_X). \tag{1.64}$$

This is immediate from Lemma 1.19. Interestingly, the *Boolean* algebra structure just derived as the governor of the (propositional) logic of classical mechanics may be reformulated in terms of the *Jordan* algebraic structure (1.46) - (1.47) of $\ell^\infty(X\tau)$, or, when X is finite, of the C*-algebra of observables $C(X)$ itself:

- Events $U \subseteq X$ (and hence, by Theorem 1.20, logical equivalence classes of propositions) correspond bijectively to characteristic functions 1_U on X, that is, with ***yes-no questions*** (having spectrum in $\{0,1\}$). Algebraically, these are precisely the ***idempotents*** in $\ell^\infty(X,\tau)$, i.e., those functions e satisfying $e^2 = e$.
- In terms of those, the partial ordering and the logical connectives are given by

$$e \leq f \text{ iff } e \circ f = e; \tag{1.65}$$
$$\neg e = 1_X - e; \tag{1.66}$$
$$e \wedge f = e \circ f; \tag{1.67}$$
$$e \vee f = e + f - e \circ f. \tag{1.68}$$

Indeed, in this case $\circ$ is pointwise multiplication (1.20). Using $1_U \cdot 1_V = 1_{U\cap V}$ yields (1.67), (1.65) comes down to $U \subseteq V$ iff $U \cap V = U$, (1.66) is $1_X - 1_U = 1_{U^c}$, and (1.68) follows by writing its right-hand side as $1_X - (1_X - e) \wedge (1_X - f)$.

1.5 The GNS-construction for $C(X)$

As a bridge from classical to quantum mechanics (as well as a good exercise), we finally inject some Hilbert space theory into classical physics by discussing the GNS-***construction*** of C*-algebra theory for the special case of $C(X)$, where X remains finite. In general, for each state ω on a C*-algebra A, the GNS-construction canonically yields a ***Hilbert space*** H_ω (which is finite-dimensional for $A = C(X)$ with finite X) and a ***representation*** of A on H_ω, in the sense of a (complex) linear map

$$\pi_\omega : A \to B(H_\omega) \tag{1.69}$$

that satisfies

$$\pi_\omega(ab) = \pi_\omega(a)\pi_\omega(b); \tag{1.70}$$
$$\pi_\omega(a^*) = \pi_\omega(a)^*. \tag{1.71}$$

Furthermore, H_ω contains a special *unit vector* Ω_ω that is ***cyclic*** for π_ω in that

$$\pi_\omega(A)\Omega_\omega \equiv \{\pi_\omega(a)\Omega_\omega,\, a \in A\} = H_\omega, \tag{1.72}$$

at least in the relevant case where $\dim(H_\omega) < \infty$; otherwise, the left-hand side is merely dense in H_ω and one needs to take the (norm) closure to obtain H_ω. Furthermore, Ω_ω realizes the state ω as a quantum-mechanical expectation value by

$$\omega(a) = \langle \Omega_\omega, \pi_\omega(a)\Omega_\omega \rangle_{H_\omega}. \tag{1.73}$$

Given $\omega \in S(A)$, the GNS-construction starts with the vector spaces

$$N_\omega = \{a \in A \mid \omega(a^*a) = 0\}; \tag{1.74}$$
$$H_\omega = A/N_\omega. \tag{1.75}$$

Now, if $b \in N_\omega$ and $a \in A$, then $ab \in N_\omega$, because of the important inequality

$$\omega(b^*a^*ab) \le \|a\|^2 \omega(b^*b). \tag{1.76}$$

This is true for any C*-algebra A, but below we prove it only for our example. Assuming (1.76) for the moment, the action of A on itself by left multiplication descends to a well-defined action on H_ω, which we call π_ω. In other words, if $b_\omega \in H_\omega$ is the image of $b \in A$ under the canonical projection $A \to A/N_\omega$, then

$$\pi_\omega(a)b_\omega = (ab)_\omega. \tag{1.77}$$

Crucially, this vector space H_ω is equipped with a canonical inner product

$$\langle a_\omega, b_\omega \rangle = \omega(a^*b). \tag{1.78}$$

Indeed, this form is well defined, and is positive definite because ω is a state.

In general, H_ω as defined by (1.75) with inner product (1.78) is merely a pre-Hilbert space, which needs to be completed in the associated norm, and it takes some effort to check that the operators defined by (1.77) are bounded. In our example, on the other hand, H_ω is finite-dimensional and hence complete. In any case, it is easy to verify the properties (1.70) - (1.73), whilst (1.72) holds with the unit $1 = 1_H$.

We now prove (1.76) for $A = C(X)$. Fom Theorem 1.15 we have $\omega = E_P$, and by (1.9) and (1.24), the inequality (1.76) comes down to the obviously correct result

$$\sum_x |f(x)g(x)|^2 \leq \|f\|_\infty^2 \sum_x |g(x)|^2. \tag{1.79}$$

Writing $N_{E_P} \equiv N_P$, we may also check directly that if $g \in N_P$ and $f \in C(X)$, then $fg \in N_P$. Indeed, in terms of the set $\mathrm{supp}\,(P) \subseteq X$ defined by

$$\mathrm{supp}\,(P) = \{x \in X \mid p(x) > 0\}, \tag{1.80}$$

we have

$$N_P = \{f \in C(X) \mid f(x) = 0 \forall x \in \mathrm{supp}\,(P)\}, \tag{1.81}$$

and clearly $g = 0$ on $\mathrm{supp}\,(P)$ implies $fg = 0$ on $\mathrm{supp}\,(P)$. We now compute H_P and π_P. From (1.81) we have $f - g \in N_P$ and hence $f \sim g$ iff $f(x) = g(x)$ for all $x \in \mathrm{supp}\,(P)$, where $\sim$ is the equivalence relation whose equivalence classes f_P define elements of $H_P = C(X)/N_P$. Hence f_P is simply the restriction of f to $\mathrm{supp}\,(P)$, and

$$H_P = \ell^2(X,P) \tag{1.82}$$

is the Hilbert space that consists of these restriction, with inner product

$$\langle f_P, g_P \rangle = \sum_{x \in \mathrm{supp}\,(P)} p(x)\overline{f(x)}g(x). \tag{1.83}$$

The representation (1.77) then trivially gives

$$\pi_P(f)g_P = f_P g_P, \tag{1.84}$$

so that $\pi_P(f)$ is the ***multiplication operator*** defined by f on $\ell^2(X,P)$. In functional analysis one often denotes elements $g_P \in \ell^2(X,P)$ by the functions g themselves, and similarly writes $\pi_P(f)$ as f, so that (1.84) simply reads $\pi_P(f)g = fg$.

The operator norm of $\pi_P(f)$ is easily computed to be

$$\|\pi_P(f)\| = \sup\{|f(x)|, x \in \mathrm{supp}\,(P)\} = \|f_{|\mathrm{supp}\,(P)}\|_\infty. \tag{1.85}$$

Indeed, the bound $\|\pi_P(f)\| \leq \|f_{|\mathrm{supp}\,(P)}\|_\infty$ is immediate from the definition

$$\|\pi_P(f)\| = \sup\{\|\pi_P(f)g_P\|, g_P \in H_P, \|g_P\| = 1\}, \tag{1.86}$$

and equality in this bound follows from applying the operator $\pi_P(f)$ to the function $g = 1_U$, where $U \subset X$ is any set where $|f|$ attains its maximum $\|f_{|\mathrm{supp}\,(P)}\|_\infty$.

Notes

§1.1. **Basic constructions of probability theory**

§1.2. **Classical observables and states**

For (advanced) treatments of convexity theory and probability theory in contexts relevant to mathematical physics we recommend Israel (1979), Alfsen & Shultz (2001), and Simon (2001).

§1.3. **Pure states and transition probabilities**

Transition probabilities (in the abstract sense meant here) were introduced by von Neumann, but his manuscript from 1937 was only published in 1981 as von Neumann (1981/1937). This remarkable paper has remained largely unused (or even unknown) in both mathematical physics and operator algebras; Mielnik (1968), Shultz (1982), and Landsman (1996, 1997) are exceptions. An extensive discussion with further references may be found in Landsman (1998a).

§1.4. **The logic of classical mechanics**

Unless one counts Boole (1847), it seems that the logical analysis of classical mechanics was initiated by the famous paper of Birkhoff & von Neumann (1936), which was primarily concerned with quantum logic (cf. §2.10). Our use of semantic implication (also in the quantum case) was inspired by Rédei (1998).

§1.5. **The GNS-construction for $C(X)$**

See §C.12 for the GNS-construction in general.