



Graduate Statistical Mechanics

8.333 MIT Physics

Julien Tailleur

These lecture notes follow the structure of the Graduate Statistical Mechanics course given in 2024 at MIT by the author. The lectures follow closely the book by Kardar “Statistical Physics of Particles”, which is a recommended complementary reading. As always, the notes are likely to contain an extensive number of typos and feedback is very much appreciated. The author is infinitely grateful to Amer Al-Hiyasat, Antti Asikainen, Bahar Shakerifakher, Jessica Metzger, Quan Nguyen, Rosalind Huang, and Sara Dal Cengio their help in preparing and proof-reading these notes.

The lecture notes are organized as follows. Chapter 1 is an introduction to classical equilibrium statistical mechanics: what is the purpose of this theory? What are the essential ideas? It introduces the concepts of statistical ensembles and discusses possible reasons why the overall approach is reasonable. Appendix ?? is a brief reminder on probability theory, which covers a content close to Chapter 2 of [1].

Chapter 2 then considers the dynamics of a dilute gas of interacting particles. Starting from classical mechanics, we build the BBGKY hierarchy and show how an explicit coarse-graining of the system on the scale intermediate between the particle size and the system size allows one to justify a closure in the form of the Boltzmann equation. From there, we derive a large-scale description of the relaxation of the density, velocity and temperature fields towards an equilibrium description. This chapter thus legitimates using equilibrium statistical mechanics to describe ideal gases.

Chapter 3 then covers in more detail the standard equilibrium ensembles. Change of ensembles and ensemble equivalence are discussed and most concepts are illustrated using either the ideal gas or two-level systems. Thermodynamic relations are recovered as a large- N limit of the equilibrium ensembles.

Chapter 4 then turns to the static properties of interacting particle systems beyond the dilute limit. We present a cluster expansion that allows one to derive the leading-order correction to the ideal gas law. This expansion is compatible with the van der Waals equation, which is constructed as a mean-field approximation. This equation is shown to predict a liquid-gas phase separation, whose mean-field properties are discussed. We show how a similar closure allows one to understand the emergence of ferromagnetism in the Ising model.

Chapter 5 then deals with quantum statistical mechanics. Quantum statistics are presented from a density-matrix perspective, whose validity is discussed in the light of Eigenstate Thermalization Hypothesis. We start the chapter by discussing how quantization of energy allows one to solve a number of experimental puzzles: the heat capacity of diatomic gases, the heat capacity of solids, and the black-body radiation. Finally, we consider quantum ideal gases, the Fermi statistics and the Bose-Einstein condensation.

Bibliography

[1] Mehran Kardar. *Statistical physics of particles*. Cambridge University Press, 2007.

Contents

1	Introduction	1
1.1	What is statistical mechanics?	2
1.2	Equilibrium Statistical Mechanics — The ensemble approach	3
1.2.1	The microcanonical ensemble	4
1.2.2	The principle of minimal information	6
1.2.3	The canonical ensemble	7
1.2.4	Liouville theorem, conserved quantities, and the statistical independence of macroscopic volumes	8
1.2.5	Key concepts and results to remember	11
2	The kinetic theory of gases	13
3	Equilibrium Statistical Mechanics	15
4	Interacting Particle Systems	17
5	Quantum Statistical Mechanics	19
	Appendices	19
	Bibliographie	21

Chapter 1

Introduction

1.1 What is statistical mechanics?

The beginning of the twentieth century has witnessed three scientific revolutions. Quantum mechanics and relativity have taught us that physics at the *infinitely small* and *infinitely large* scales is fundamentally different from what we experience at our scale. Statistical mechanics, on the other hand, addresses the question of the *infinitely complex*: how can we understand the diversity of the world around us, given the relatively limited diversity of atoms that makes up most of the materials we encounter? The answer to that question lies in two apparently contradicting ideas.

On the one hand, paraphrasing Anderson [1], “**more is different**”. This quote captures the fact that a large number of interacting particles are able to acquire properties inaccessible to the individual units. Consider for instance a molecule of silica. Its dynamics is given by Newton’s first law so that any force you exert on it leads to its motion through space. Gently pushing on a heavy glass table tells a completely different story: the amorphous packing of silica molecules into a tempered glass confers them the collective ability to resist motion and deformation. Collectively, the silica molecules have created a solid. The process through which one goes from a free-flowing silica sand to a solid glass involves a number of complex phase transitions, the understanding of which is a central goal of statistical mechanics. Starting from the microscopic scale, statistical physics thus aims at understanding how new properties emerge from the collective behaviors of individual units.

On the second hand, if “more is different”, then understanding the macroscopic world may seem a hopeless goal: All forms of matter would acquire system-dependent properties, turning the beauty of the universal laws of physics into a depressing catalogue of material properties that would need to be studied on a case-by-case basis. Luckily, **different is often the same**, in the sense that emerging phenomena often share profound similarities. For instance, hexagonal lattices are observed in a multitude of situations: in a honeycomb, on the pineapple fruit’s surface, in the Giant causeway in Ireland, in the arrangement of cell tissues (e.g., in the hexagonal patterning of the *Drosophila* eye), etc. While our world is complex, its complexity respects rules and symmetries that allow categorizing collective states of matter based on their properties: solid, liquid, and gaseous phases for instance describe a large part of the materials you find around you. At the collective ‘macroscopic’ scale, statistical physics aims at understanding what determines the types of patterns and forms of organizations that can be observed in nature.

In addition, **more can also be simpler**. Consider the three-body problem in classical mechanics. As we know since Poincaré, there is no closed-form solution to the dynamics of three interacting bodies and, for most initial conditions, the resulting dynamics is chaotic. In short, this is a horribly complicated problem on which we cannot predict much. Now, consider one cubic meter of air, which contains $\mathcal{O}(10^{23})$ molecules of N_2 and O_2 . If I tell you that its temperature is 300 K, its pressure is 1 bar, and its humidity is 50%, you will have a very clear idea of what I am talking about. Three numbers are thus enough to give you all the informa-

tion you need to characterize the collective state of $\mathcal{O}(10^{23})$ degrees of freedom! Statistical mechanics aims at understanding this emerging simplicity.

In short, the goal of statistical mechanics is to derive a micro-to-macro approach that allows predicting the large-scale properties of a system starting from its microscopic description. On the way, it will determine which microscopic details are crucial or irrelevant, and identify the relevant variables to describe a system at the macroscopic scale. While this may sound like a fundamental and theoretical question, the applications of statistical mechanics are everywhere around you: from the control of the rheological properties of your toothpaste to the engineering of LCDs, statistical physics plays a major role in material design. Note also that nothing so far restricts the principles of statistical mechanics to physical systems and its methodological toolbox can be applied (with, or without wisdom) to economics, sociology, epidemiology, and computer science.

In this course, we will focus on the properties of large collections of particles in thermal equilibrium. This course is complemented by others at MIT: 8.334 in the spring term focuses on phase transitions and the emerging phenomenology they induce, 8.08/8.S308 in IAP focuses on stochastic and nonequilibrium dynamics, with application to the thriving field of active matter, and finally 8.592 in fall focuses on applications to biophysics.

1.2 Equilibrium Statistical Mechanics — The ensemble approach

In most fields of physics, the fundamental laws are given as dynamical equations: Newton’s laws of classical mechanics, Einstein’s equations for GR, Schrödinger’s equation for QM, Navier-Stokes for fluid mechanics, Maxwell’s equation for electromagnetism, etc. Solving such equations in a simple context is already a challenging task, but doing so for a system with 10^{23} degrees of freedom is completely out of reach¹. Furthermore, it is also quite useless since these equations predict the state of a system at a time t given the perfect knowledge of initial conditions at a time $t' < t$. Any error on the initial condition will lead to a lack of precision on the state of the system at time t . What classical mechanics taught us in the 20th century is that most complex systems are *chaotic*: any initial uncertainty increases exponentially in time at a rate given by the largest so-called *Lyapunov exponent* [5].

Instead of trying to face this daunting task, equilibrium statistical mechanics proposes to take a completely different stance: since the air around us looks right now much like it looked like a few seconds ago, maybe solving these dynamical equations is useless and the only thing we need to do is to characterize the most probable state of a system and the typical fluctuations around it. Assuming that we can describe a classical system by a set of positions and momenta \vec{q}, \vec{p} , with $\vec{q} = (q_1, \dots, q_N)$ and $\vec{p} = (p_1, \dots, p_N)$, we can try to characterize the probability $P(\{\vec{q}, \vec{p}\}) = \rho(\{\vec{q}, \vec{p}\})d\Gamma$ to observe the system within a given phase-space volume $d\Gamma = \prod_{i=1}^N d\vec{q}_i d\vec{p}_i$ around \vec{q}, \vec{p} . Of course, this is only viable if one can guess what the

¹Note that simulations are getting closer and closer to being able to solve such complicated systems, but only on very short time scales.

right probability density $\rho(\{\vec{q}, \vec{p}\})$ is, and it is only useful if this distribution is not horribly complicated.

1.2.1 The microcanonical ensemble

Classical equations of motion. The simplest way to proceed is to start from what we know on our system: if the energy is given by a time-independent Hamiltonian $H(\vec{q}, \vec{p})$, the trajectories $\vec{q}(t), \vec{p}(t)$ will be solutions of Hamilton's equations of motion:

$$\frac{d}{dt}\vec{q}(t) = \frac{\partial H}{\partial \vec{p}}, \quad \frac{d}{dt}\vec{p}(t) = -\frac{\partial H}{\partial \vec{q}}, \quad (1.1)$$

where $\frac{\partial H}{\partial \vec{x}}$ denotes the vector whose coordinates are $\frac{\partial H}{\partial x_i}$. All the symmetries and conservation laws that are encoded in the dynamics (1.1) must be reflected in ρ .

Conservation of energy. A first important property of dynamics (1.1) is that it conserves the value of the Hamiltonian along a trajectory:

$$H(\vec{q}(t), \vec{p}(t)) = H(\vec{q}(0), \vec{p}(0)) \equiv E. \quad (1.2)$$

The energy is a constant of motion. To show this, we can use the chain rule:

$$\frac{dH(\vec{q}(t), \vec{p}(t))}{dt} = \frac{\partial H}{\partial t} + \sum_i \left(\frac{\partial H}{\partial q_i} \frac{dq_i}{dt} + \frac{\partial H}{\partial p_i} \frac{dp_i}{dt} \right) = 0 + \sum_i \left(\frac{\partial H}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial H}{\partial p_i} \frac{\partial H}{\partial q_i} \right) = 0, \quad (1.3)$$

where we have used that $\partial_t H = 0$ since the Hamiltonian does not explicitly depend on time.

Note of caution. We note here a difficulty commonly encountered in physics textbooks (and in these lecture notes). In Eq. (1.2), we have written $H(\vec{q}(t), \vec{p}(t)) \equiv E$ to explicitly state that E is the value of the function $H(\vec{q}, \vec{p})$ evaluated with arguments $\vec{q}(t), \vec{p}(t)$ that are solutions of (1.1), instead of simply writing $H(\vec{q}, \vec{p}) \equiv E$. This degree of ‘precision’ ensures clarity, even if we have already been somewhat sloppy: since the values of $\vec{q}(t), \vec{p}(t)$ depend on the initial conditions, we should really write something like $H(\vec{q}(t), \vec{p}(t)|\vec{q}(0), \vec{p}(0))$. Clearly, notations will soon become unpractically heavy and obscure the meaning of our equations. As often in physics, we will trade rigor for clarity and often silently omit arguments, dependencies, initial conditions, etc. It is not uncommon to find in textbooks $E \equiv H(\vec{q}, \vec{p})$ and it is only when Eq. (1.3) is written that the fact that \vec{q}, \vec{p} are assumed to be solutions of (1.2)—and not constant, arbitrary vectors—becomes explicit. Such concise yet sloppy notations allow us to deal with situations arguably richer than the systems typically studied in more rigorous contexts, but it comes at a price: ambiguity, which is a common source of mistakes.

Lecture 1 stopped here

The microcanonical hypothesis: An important consequence of Eq. (1.2) is that, if the energy of the system at time 0 is E_0 , then we know that $\rho(\vec{q}, \vec{p}) = 0$ if $H(\vec{q}, \vec{p}) \neq E_0$. Since the solutions of Hamilton's equations of motion are trajectories confined within an energy surface, then the support of the distribution we are looking for must coincide with the energy surface.

But how do we go forward? How do we know which region of the energy surface are more favorable than others? An answer is suggested by the theory of dynamical systems: we don't.

And if we don't know that some regions are more favorable than others, then we should not bias our probability distribution in any way. If there are no clear reasons that our system should explore some regions more often than others, then we can assume that the chaotic dynamics induced by the complexity of our 10^{23} interacting degrees of freedom makes the system explore the energy surface ergodically and uniformly.

This idea predates significantly the theory of chaos, in the form of the microcanonical hypothesis. For simplicity, we present the microcanonical hypothesis in the context of discrete systems. We will come back to its expression for continuous systems in Chapter 3². Consider a classical complex system described by a discrete set of configurations $\{\mathcal{C}\}$, then all configurations with the same energy E have the same probability of occurrence and

$$P_E(\mathcal{C}) = \frac{1}{\Omega(E)} \delta_{H(\mathcal{C}), E} , \quad (1.4)$$

where $\Omega(E)$ is the number of configurations with energy E and $\delta_{a,b}$ is the Kronecker delta, defined by $\delta_{a,b} = 1$ if $a = b$ and $\delta_{a,b} = 0$ otherwise.

Microcanonical entropy. In Eq. (1.4), $\Omega(E)$ is simply a normalization constant such that $\sum_{\mathcal{C}} P_E(\mathcal{C}) = 1$. As the energy of the system changes, so does the number of accessible configurations $\Omega(E)$. It turns out that the rate of change of Ω with the energy E or the number of constituents N is typically exponential, so that a more natural observable to quantify the accessible space of configurations is provided by Boltzmann microcanonical entropy:

$$S_m(E) = k_B \ln \Omega(E) , \quad (1.5)$$

where $k_B = 1.380649 \times 10^{-23} \text{ J} \cdot \text{K}^{-1}$ is the Boltzmann constant.

Microcanonical temperature: How Ω and S_m vary with E is quantified by the microcanonical temperature:

$$\frac{1}{T_m} \equiv \frac{\partial S_m}{\partial E} . \quad (1.6)$$

At this stage, why the definition (1.6) is a good definition for the temperature, and how to relate it to what you measure using a thermometer, is far from obvious. This will become clearer in chapter 3 when we establish ensemble equivalence.

Summary: Starting from the only constraint we had on the classical dynamics describing our system, we have suggested in Eq. (1.4) a probability distribution to describe the probability to observe a given configuration \mathcal{C} . Is it simple? Yes. Is it practical? Not really, because counting the number of configurations of a given energy is often a combinatorial challenge. Furthermore, this only allows describing systems in which the total energy is strictly conserved, which are thus isolated from their environments. While such systems can be engineered, e.g. using ultrahigh vacuum, this is far from the standard situation that we encounter around us.

This suggests studying another situation, in which a system is allowed to exchange energy with the environment. In such a case, the energy is not a constant of motion anymore

²If $\{\mathcal{C}\}$ is a continuous space, then P_E is a probability density, and $\Omega(E)$ is the area of the hyper-surface of configurations with energy E .

and, instead, we may simply know that the average energy of the system is fixed: $\langle E \rangle \equiv \sum_{\mathcal{C}} E(\mathcal{C})p(\mathcal{C}) = E_0$. Can we then guess what a reasonable choice for $p(\mathcal{C})$ is? Yes, following a proposal by Jaynes [2, 3], which makes use of Shannon entropy [6].

Notations: Throughout these lecture notes, the brackets $\langle \dots \rangle$ represent averages with respect to the relevant probability distribution.

1.2.2 The principle of minimal information

Consider a random variable $n \in \{1, \dots, N\}$, that you can sample freely, characterized by a probability distribution p that you do not know. A first natural question is ‘how many times do you have to sample this distribution to have an accurate estimate of $p(n)$?’ If $p(n) = \delta_{n,n_0}$, then a single sample allows you to determine p . If p is more complex, then you may need many different samples to estimate the value of p . An alternate but equivalent question is ‘how surprising are the results of sampling p ?’ If p is a delta function, there is no surprise in the outcome. If p is uniform, then you cannot predict anything about the result of sampling p and the surprise is maximal. The ideas of Shannon on information theory allow us to quantify the surprise contained in a distribution using a surprise function s .

Consider a sample n , if $p(n) = 1$, then there is no surprise in observing n and $s(p(n)) = 0$. If you look at two independent samples, n_1 and n_2 , then the surprise contained in the joint observation of n_1 and n_2 should be additive: $s(p(n_1, n_2)) = s(p(n_1)p(n_2)) = s(p(n_1)) + s(p(n_2))$. Finally, the surprise should be decreasing function of $p(n)$: the less likely n , the more surprised you should be to observe it. Putting this together, one can show that the surprise function has to be written as $s(p) = -k \ln(p)$, with $k > 0$. The average surprise contained in a distribution is then

$$S_s(p) = - \sum_{\mathcal{C}} p(\mathcal{C}) \ln p(\mathcal{C}) , \quad (1.7)$$

which is called the Shannon entropy.

Gibbs’ entropy Note that, another definition of the microcanonical entropy, due to Gibbs, reads:

$$S_G(E) = -k_B \sum_{\mathcal{C}} P_E(\mathcal{C}) \ln P_E(\mathcal{C}) = -k_B \sum_{\mathcal{C}|H(\mathcal{C})=E} \frac{1}{\Omega(E)} \ln \frac{1}{\Omega(E)} = k_B \ln \Omega(E) . \quad (1.8)$$

As you can see, the definitions of Boltzmann and Gibbs coincide in the microcanonical ensemble. Boltzmann’s definition has a strong appeal because of its direct connection to a physical observable ($\Omega(E)$). Thanks to Shannon [6], we now understand that, up to the constant k_B , Gibbs & Boltzmann entropy also measure the (lack of) information contained in the distribution p . Equation (1.8) thus tells us that, when sampling a system in the microcanonical ensemble, our surprise is proportional to $\ln \Omega(E)$. This is consistent with our intuition: the larger $\Omega(E)$, the more configurations are accessible to the system.

Let us now turn back to the choice of distribution we should use to try and describe the statistics of a system in contact with a thermostat, whose average energy is fixed.

1.2.3 The canonical ensemble

We thus consider a system whose average energy is fixed, due to energy exchange with a thermostat. What other information do we have on the system? None, and we thus do not have any good reason to add more information or bias in the distribution $p(\mathcal{C})$. We thus want a distribution that contains the information that

$$\langle E \rangle = \sum_{\mathcal{C}} E(\mathcal{C})p(\mathcal{C}) = E_0 \quad \text{and} \quad \sum_{\mathcal{C}} p(\mathcal{C}) = 1, \quad (1.9)$$

but nothing else. We thus want to minimize the information contained in the distribution $p(\mathcal{C})$, under the constraints (1.9). To do so, we maximize the average surprise contained in the distribution (i.e. the Shannon entropy), under the constraints (1.9). To proceed, we introduce a function

$$L(p) = - \sum_{\mathcal{C}} p(\mathcal{C}) \ln p(\mathcal{C}) + \beta [E_0 - \sum_{\mathcal{C}} p(\mathcal{C}) E(\mathcal{C})] + \alpha [1 - \sum_{\mathcal{C}} p(\mathcal{C})], \quad (1.10)$$

where α and β are two Lagrange multipliers. Let us consider a configuration \mathcal{C}_i and extremize L with respect to $p(\mathcal{C}_i)$ ³:

$$\frac{\partial L(p)}{\partial p(\mathcal{C}_i)} = -\ln p(\mathcal{C}_i) - 1 - \beta E(\mathcal{C}_i) - \alpha = 0. \quad (1.11)$$

This equation can be solved as

$$p(\mathcal{C}_i) = e^{-1-\alpha-\beta E(\mathcal{C}_i)}. \quad (1.12)$$

The Lagrange multipliers α and β can then be determined to satisfy $\langle E \rangle = E_0$ and $\sum_{\mathcal{C}} p(\mathcal{C}) = 1$. Using the normalization to fix α leads to the celebrated canonical distribution:

$$p(\mathcal{C}) = \frac{1}{Z} e^{-\beta E(\mathcal{C})} \quad \text{where} \quad Z = \sum_{\mathcal{C}} e^{-\beta E(\mathcal{C})}, \quad (1.13)$$

is the so-called partition function in the canonical ensemble. The value of β is then fixed by noting that

$$E_0 = \langle E \rangle = \frac{1}{Z} \sum_{\mathcal{C}} E(\mathcal{C}) e^{-\beta E(\mathcal{C})} = -\frac{1}{Z} \partial_{\beta} \sum_{\mathcal{C}} e^{-\beta E(\mathcal{C})} = -\partial_{\beta} \ln Z. \quad (1.14)$$

Summary: Starting from the observation that the dynamics of complex macroscopic systems is both irrelevant and impractical to describe their typical states, we have suggested to replace this problem by a formally simpler one: to characterize directly the probability distribution of observing the configurations of a system. We decided to put as little information as we can in these distributions: enforcing the constraints given by the laws of physics, and nothing else. Shannon information theory told us that this can be done by a constrained extremization of the Gibbs-Shannon entropy. For an isolated system, you will show in the problem set that

³We note that $\frac{\partial^2 L(p)}{\partial p(\mathcal{C}_i) \partial p(\mathcal{C}_j)} = -\frac{1}{p(\mathcal{C}_i)} \delta_{ij}$ so that any local extremum is necessarily a maximum.

this gives back the microcanonical ensemble. For a system whose average energy is fixed, it yields the canonical distribution.

This construction of the various statistical ensembles, which can be generalized beyond the cases discussed here is simple and appealing, but it is not completely satisfactory. Indeed, we have extremized the Gibbs-Shannon entropy under the constraints that we are aware of, but nature follows its own rules whether we know them or not: a physical theory should not be based on our knowledge or ignorance. A partial answer comes from classical mechanics and chaos: H is a constant of motion but we can expect that in complex systems all other quantities are “stirred” by the chaotic dynamics so that the energy surface is visited uniformly. There are many issues with this statement. For instance, even when it is correct, it is hard to prove. And there are cases in which we know that it is wrong: For instance, the KAM theorem tells us that it is wrong at the level of perturbation theory close to integrability [5]⁴. Furthermore, even when the chaotic dynamics of the system would lead to an ergodic sampling of the energy surface, the time for the dynamics to do so diverges exponentially with the number of degrees of freedom, so this is not why statistical mechanics works in practice. Many other reasons have been put forward to “justify” equilibrium statistical mechanics, and we will review some of them (e.g. justifying statistical mechanics at the quantum level and deriving classical statistical mechanics as its high temperature limit). A rigorous proof of equilibrium statistical mechanics remains an open research problem for most systems. We will close this introductory chapter by mentioning another nice argument, that is discussed in the series of books by Landau and Lifschitz [4].

1.2.4 Liouville theorem, conserved quantities, and the statistical independence of macroscopic volumes

Let us derive Liouville’s equation and Liouville’s theorem to provide an appealing justification for the form of the Boltzmann weight.

Liouville’s equation. We consider N classical particles in a $3d$ -space whose positions and momenta are given by

$$\vec{q}_i = \begin{pmatrix} q_{i,x} \\ q_{i,y} \\ q_{i,z} \end{pmatrix} \quad \text{and} \quad \vec{p}_i = \begin{pmatrix} p_{i,x} \\ p_{i,y} \\ p_{i,z} \end{pmatrix}. \quad (1.15)$$

We then write the corresponding equations of motion as

$$\dot{\vec{q}}_i = \frac{\partial H}{\partial \vec{p}_i} \quad \text{and} \quad \dot{\vec{p}}_i = -\frac{\partial H}{\partial \vec{q}_i}, \quad (1.16)$$

⁴Consider N independent harmonic oscillators, that admit N constants of motion, and add a random non-linear coupling between them that scales with a control parameter ε . As $\varepsilon \rightarrow 0$, the measure of the energy surface that corresponds to chaotic dynamics vanishes and most of the energy surface is regular. The microcanonical hypothesis is thus violated for such a system.

where we have introduced the operators

$$\frac{\partial}{\partial \vec{q}_i} \equiv \begin{pmatrix} \frac{\partial}{\partial q_{i,x}} \\ \frac{\partial}{\partial q_{i,y}} \\ \frac{\partial}{\partial q_{i,z}} \end{pmatrix} \quad \text{and} \quad \frac{\partial}{\partial \vec{p}_i} \equiv \begin{pmatrix} \frac{\partial}{\partial p_{i,x}} \\ \frac{\partial}{\partial p_{i,y}} \\ \frac{\partial}{\partial p_{i,z}} \end{pmatrix}. \quad (1.17)$$

Let us denote $\rho(\{\vec{q}_i, \vec{p}_i\}, t)$ the probability density of finding the system at $\{\vec{q}_i, \vec{p}_i\}$ at time t . We thus choose an initial condition $\{\vec{q}_i(0), \vec{p}_i(0)\}$ at random, such that it lies within an infinitesimal volume element $d\Gamma$ near $\{\vec{q}_i^0, \vec{p}_i^0\}$, where

$$d\Gamma \equiv \prod_{i=1}^N \prod_{\alpha=x,y,z} dq_{i,\alpha} dp_{i,\alpha} \equiv \prod_{i=1}^N d\vec{q}_i d\vec{p}_i, \quad (1.18)$$

with probability

$$\rho(\{\vec{q}_i^0, \vec{p}_i^0\}, t=0) d\Gamma. \quad (1.19)$$

For any phase-space coordinate $\{\vec{q}_i, \vec{p}_i\}$, the time evolution of the probability density of finding the system at $\{\vec{q}_i, \vec{p}_i\}$ is then given by

$$\left. \frac{\partial \rho(\{\vec{q}_i, \vec{p}_i\}, t)}{\partial t} \right|_{\{\vec{q}_i, \vec{p}_i\} \text{ held fixed}} \quad (1.20)$$

We stress the difference with Eq. (1.2) where the function H was evaluated with arguments $\vec{q}(t), \vec{p}(t)$ that evolve in time. The conservation of probability then reads ⁵

$$\partial_t \rho(\{\vec{q}_i, \vec{p}_i\}, t) = - \sum_i \left[\frac{\partial}{\partial \vec{q}_i} \cdot [\dot{\vec{q}}_i \rho(\{\vec{q}_i, \vec{p}_i\}, t)] + \frac{\partial}{\partial \vec{p}_i} \cdot [\dot{\vec{p}}_i \rho(\{\vec{q}_i, \vec{p}_i\}, t)] \right]. \quad (1.21)$$

Using the dynamics (1.16), one then finds the celebrated Liouville's equation:

$$\partial_t \rho(\{\vec{q}_i, \vec{p}_i\}, t) = \sum_i \left[\frac{\partial H}{\partial \vec{q}_i} \cdot \frac{\partial \rho(\{\vec{q}_i, \vec{p}_i\}, t)}{\partial \vec{p}_i} - \frac{\partial H}{\partial \vec{p}_i} \cdot \frac{\partial \rho(\{\vec{q}_i, \vec{p}_i\}, t)}{\partial \vec{q}_i} \right]. \quad (1.22)$$

Note that our vectorial notation is simply a way to save a summation over the coordinates x, y, z , since expanding the dot products leads to

$$\partial_t \rho(\{\vec{q}_i, \vec{p}_i\}, t) = \sum_{i=1}^N \sum_{\alpha=x,y,z} \left[\frac{\partial H}{\partial q_{i,\alpha}} \frac{\partial \rho(\{\vec{q}_i, \vec{p}_i\}, t)}{\partial p_{i,\alpha}} - \frac{\partial H}{\partial p_{i,\alpha}} \frac{\partial \rho(\{\vec{q}_i, \vec{p}_i\}, t)}{\partial q_{i,\alpha}} \right]. \quad (1.23)$$

⁵This footnote is a brief reminder of the derivation of a conservation law. Consider a variable \vec{x} that evolves as $\dot{\vec{x}}(t) = \vec{u}(\vec{x}(t))$ and a (probability) density $\rho(x, t)$ advected by this flow. In any volume V and dimension d , the variation of ρ inside V is due to the flux of the mass/probability current $\vec{j}(x, t) = \rho(x, t) \vec{u}(\vec{x})$ through the surface ∂V of V : $\frac{d}{dt} \int_V \rho(\vec{x}, t) d^d \vec{x} = - \int_{\partial V} \vec{j} \cdot d^{d-1} \vec{S} = - \int_V \nabla \cdot \vec{j} d^d \vec{x}$, where the last equality stems from the divergence theorem. Since this holds for any volume V , one has that $\partial_t \rho(\vec{x}, t) = -\nabla \cdot \vec{j} = -\nabla \cdot [\rho \vec{u}(\vec{x})]$. For a more direct derivation, you can work in $1d$ and compare the mass/probability in $[x, x+dx]$ between time t and $t+dt$, and then send dt and dx to zero. Equation (1.21) is a direct application of this theorem with the flow given in Eq. (1.16).

Liouville's equation is often compactly rewritten as

$$\partial_t \rho(\{\vec{q}_i, \vec{p}_i\}, t) = -\{\rho, H\} , \quad (1.24)$$

where we have introduced the Poisson bracket

$$\{A, B\} \equiv \sum_i \left[\frac{\partial A}{\partial \vec{q}_i} \cdot \frac{\partial B}{\partial \vec{p}_i} - \frac{\partial B}{\partial \vec{q}_i} \cdot \frac{\partial A}{\partial \vec{p}_i} \right] . \quad (1.25)$$

Liouville's theorem Let us now show that ρ is constant along a trajectory, i.e. that $\rho(\{\vec{q}_i(t), \vec{p}_i(t)\}, t)$ is a constant of motion. Using the chain rule, we find:

$$\frac{d}{dt} \rho(\{\vec{q}_i(t), \vec{p}_i(t)\}, t) = \partial_t \rho + \sum_i [\dot{\vec{q}}_i \cdot \partial_{\vec{q}_i} \rho + \dot{\vec{p}}_i \cdot \partial_{\vec{p}_i} \rho] . \quad (1.26)$$

Using Eqs. (1.22) and (1.16), one finds that **the probability density is a constant of motion**:

$$\frac{d}{dt} \rho(\{\vec{q}_i(t), \vec{p}_i(t)\}, t) = 0 . \quad (1.27)$$

Conserved quantity and statistical ensembles. Liouville's theorem offers an elegant justification of the form of statistical ensemble, which is discussed in detail in Ref [4]. The underlying idea is to assume that, in a macroscopic system, mesoscopic volumes are independent.

Consider a very large system S and two subsystem S_1 and S_2 that describe two sets of degrees of freedom $\{\vec{q}_i, \vec{p}_i\}_{i \in S_1}$ and $\{\vec{q}_i, \vec{p}_i\}_{i \in S_2}$. If these subsystems are very large compared to any correlation length, then, in the steady state, they are statistically independent and the steady-state density distributions should satisfy

$$\rho_{1 \cup 2}(\{\vec{q}_i, \vec{p}_i\}_{i \in S_1 \cup S_2}) \approx \rho_1(\{\vec{q}_i, \vec{p}_i\}_{i \in S_1}) \rho_2(\{\vec{q}_i, \vec{p}_i\}_{i \in S_2}) \quad (1.28)$$

Taking the logarithm of this equality leads to

$$\ln \rho_{1 \cup 2} = \ln \rho_1 + \ln \rho_2 , \quad (1.29)$$

so that **$\ln \rho$ is additive**.

Justification of the functional form of the Boltzmann weight. One then note that the energy is also an additive constant of motion. If it is the only one, then $\ln \rho \propto E$. Denoting by $-\beta$ the proportionality constant leads back to the Boltzmann weight given in Eq. (1.13)⁶. Again, there are many caveats in this reasoning (the energy is not strictly additive due to surface terms, there might be other additive constants of motions, etc.), but this is another intuitive and appealing justification for why the canonical ensemble is relevant to describe the statistics of systems in equilibrium.

A natural question is whether we can do better to justify the use of these ensembles. In the next chapter, we answer this question positively and show explicitly that the statistics of a dilute gas of interacting particles can be described by the canonical ensemble.

⁶This reasoning can be extended to consider other conserved additive quantities like the particle number, the linear and angular momentum, etc. This then leads to other statistical ensembles.

1.2.5 Key concepts and results to remember

- The notion of statistical ensemble.
- Conservation of energy along classical mechanics trajectories.
- Microcanonical ensemble: probability distribution, definition of entropy and temperature.
- Canonical ensemble: distribution and partition function.
- Shannon entropy and the notion of information contained in a distribution.
- Liouville equation and Liouville theorem.

Bibliography

- [1] Philip W Anderson. More is different: Broken symmetry and the nature of the hierarchical structure of science. *Science*, 177(4047):393–396, 1972.
- [2] Edwin T Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4):620, 1957.
- [3] Edwin T Jaynes. Information theory and statistical mechanics. ii. *Physical review*, 108(2):171, 1957.
- [4] Lev Davidovich Landau, Evgenii Mikhailovich Lifshitz, Evgeniĭ Mikhaĭlovich Lifshitz, and LP Pitaevskii. *Statistical physics: theory of the condensed state*, volume 9. Butterworth-Heinemann, 1980.
- [5] Edward Ott. *Chaos in dynamical systems*. Cambridge university press, 2002.
- [6] Claude Elwood Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.

Chapter 2

The kinetic theory of gases

Chapter 3

Equilibrium Statistical Mechanics

Chapter 4

Interacting Particle Systems

Chapter 5

Quantum Statistical Mechanics

Bibliography

- [1] Philip W Anderson. More is different: Broken symmetry and the nature of the hierarchical structure of science. *Science*, 177(4047):393–396, 1972.
- [2] Edwin T Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4):620, 1957.
- [3] Edwin T Jaynes. Information theory and statistical mechanics. ii. *Physical review*, 108(2):171, 1957.
- [4] Mehran Kardar. *Statistical physics of particles*. Cambridge University Press, 2007.
- [5] Lev Davidovich Landau, Evgenii Mikhailovich Lifshitz, Evgeniĭ Mikhaĭlovich Lifshitz, and LP Pitaevskii. *Statistical physics: theory of the condensed state*, volume 9. Butterworth-Heinemann, 1980.
- [6] Edward Ott. *Chaos in dynamical systems*. Cambridge university press, 2002.
- [7] Claude Elwood Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.