

Lecture notes for Collective and Emergent Phenomena
(PHYS 504)

Daniel M. Sussman

January 8, 2026

Contents

Preface	7
Sources	9
I Phenomenological descriptions of phases and phase changes	11
1 Thermodynamics	5
1.1 Introduction	5
1.2 The laws of thermodynamics	6
1.3 Thermodynamics of simple systems	11
1.4 Less standard thermodynamic systems	14
1.5 Various thermodynamic potentials	14
1.6 Extensivity, Gibbs-Duhem relations, and Maxwell relations	19
1.7 Geometry, thermodynamics, and classical mechanics	23
1.8 Problems	25
2 The phenomenology of phase transitions	31
2.1 Discontinuous transitions	31
2.2 The law of corresponding states	34
2.3 The van der Waals critical point	36
2.4 Another mean field theory, more critical exponents	38
2.5 Problems	42
3 Landau theory	45
3.1 Landau's phenomenological theory	45
3.2 Correlations and fluctuations	49
3.3 Critical exponents	52
3.4 Scaling hypothesis	55
3.5 Problems	61
II Statistical routes to thermodynamic descriptions of macroscopic systems	75
4 Probability, big numbers, information, and entropy!	69
4.1 A funny observation	69

4.2	Moment and cumulant generating functions	71
4.3	The Central limit theorem	74
4.4	Adding exponentially large quantities together	77
4.5	Information, Entropy, and unbiased estimation	79
4.6	Problems	84
5	Kinetic Theory: from Liouville to the H-theorem	91
5.1	Elements of ensemble theory	92
5.2	BBGKY hierarchy	96
5.3	The H-Theorem	100
5.4	Introduction to hydrodynamics	103
5.5	Problems	110
6	Classical Statistical Mechanics	113
6.1	Thermodynamics from Statistical Mechanics	113
6.2	The canonical ensemble	119
6.3	The partition function and thermodynamics	121
6.4	Non-interacting systems and mean field calculations	124
6.5	Gibbs canonical ensemble	126
6.6	The grand canonical ensemble	127
6.7	Problems	132
III	Interacting systems, phase transitions, and of out-of-equilibrium statistical physics	147
7	Interacting systems: Perturbative approaches	141
7.1	From moment expansions...	141
7.2	... to cumulant expansions...	142
7.3	...to cluster expansions!	146
7.4	Virial expansion for a dilute gas	149
7.5	The van der Waals equation	152
7.6	Problems	155
8	Interacting systems: Variational approaches	161
8.1	Mean field Magnets	161
8.2	The van der Waals equation as a mean field theory	162
8.3	A variational view of mean field theories	165
8.4	Problems	169
9	Critical behavior: the renormalization group approach	173
9.1	Kadanoff's block spin argument	174
9.2	The renormalization group framework	177
9.3	The 1-dimensional Ising chain	181
9.4	The 2-dimensional Ising lattice	185
9.5	Universality and a momentum-space formulation	192

9.6 Problems	193
10 Fluctuations and non-equilibrium physics	195
10.1 Brownian motion – Einstein, Langevin, and Smoluchowski	195
10.2 Fokker-Planck	195
10.3 Fluctuation-dissipation	195
10.4 Onsager – reciprocity and non-reciprocity	195
10.5 Problems	196
IV Appendices	205
A Thermodynamics	201
A.1 2nd Law: from fridges to entropy	201
B Probability	207
B.1 Basic Definitions	207
B.2 Properties of single random variables	208
B.3 Important distributions	209
B.4 Properties of multiple random variables	211
C Kinetic Theory – from BBGKY to Boltzmann	215
C.1 Boltzmann a la BBGKY	215
D Quantum Statistical Mechanics	221
D.1 Classical statistical mechanics: We’ve got some issues	221
D.2 Microstates, observables, and dynamics	225
D.3 The density matrix and macroscopic observables	227
D.4 Quantum ensembles	228
D.5 The classical limit of a quantum partition function	231
D.6 Quantum indistinguishability	233
D.7 The canonical ensemble density matrix for non-interacting identical particles	236
D.8 GCE for ideal quantum particles	240
D.9 Ideal quantum gases	242
D.10 Ideal Bose gases	246
D.11 Ideal Fermi gases	250
D.12 Problems	253
Bibliography	264

Preface

In most of our physics courses we focus on the fundamental laws of Nature and how they dictate the properties of elementary particles, the forces that they interact according to, and the way they evolve in time. We learn via Newton/Lagrange/Hamilton how to describe the motion of classical particles described by their positions and momenta, and how to use the Schrödinger equation to understand the evolution of a wave function, and how electromagnetic fields propagate, and so on. It is all extremely satisfying.

And yet... as we go about our lives the physical things we want to understand about the world are often on a very different level of description. As I wait for an electric kettle to heat the water for my coffee every morning, I don't really care about the positions and momenta of the water molecules – I want to know things like “Is that water hot yet, and how much more energy do I need to add to get it to boil?” When I'm feeling particularly tired perhaps I catch myself asking even more basic questions: “What do I really mean by ‘hot’? Will that steam burn my hand? Why, in fact, are some of these molecules organized into ‘water’ and others ‘steam’? What physics leads to me calling water ‘wet’? Also...”

How can we even begin to answer these kinds of questions? If I hand you a blob containing Avogadro's number of particles and I tell you all of the masses and positions and momenta and interactions, would you be able to tell me anything about that blob? Could we understand what would happen if we poked it? Melted it? Would we know what color it was? Whether it was sticky? Magnetic? We like to think that the basic perspective of physics is that the world around us is understandable¹, so giving up is not an option. Dutifully writing down the coupled evolution equations for 10^{23} particles doesn't really feel like an option either.

The goal of this course, as schematically illustrated in Fig. 1, is to begin to make maps that connect our understanding of the microscopic world with the behavior of matter at the macroscopic scale. As suggested by its name, the broad field of *Statistical Physics* is in the business of building *probabilistic* descriptions, and relies on the limit of large numbers of particles involved so that these probabilistic maps nevertheless correspond to incredibly precise, deterministic predictions. In this class we'll learn how to use these maps to understand the collective properties of matter, and how rich and interesting phenomena can arise from even simple building blocks.

This very-much-work-in-progress set of notes is currently divided into three parts. The first part deals with the basic phenomenology of macroscopic systems, and will begin with a review of Thermodynamics. Having understood how we expect bulk phases of matter to behave, and the interrelations we expect of various quantities when a system is “in equilibrium” (whatever that means), we will discuss mean field versions of what happens when a system

¹John J. Hopfield, [1]

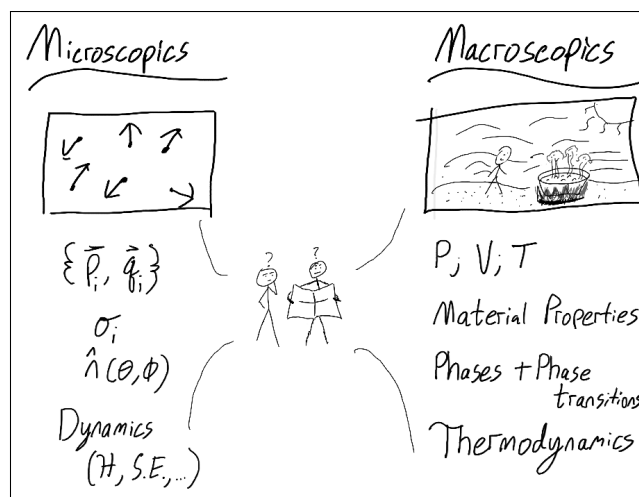


Figure 1: Making probabilistic maps between the microscopic and macroscopic world

experiences a phase transition. We will talk about both discontinuous phase transitions (as when water abruptly freezes in to solid ice) and continuous ones (as when a magnet is brought below a critical temperature in the absence of an external field), and we will get a glimpse of a remarkable *universal behavior* that seemingly disparate physical systems exhibit when they get close to their critical points. We will talk about Landau's phenomenological approach to explaining this universality, we will learn about critical scaling near phase transitions, and we will close the first part of the course with our first tentative steps towards figuring out what might be going on (via the static scaling hypothesis and Kadanoff's block spin transformation idea).

The second part covers fairly standard concepts in statistical physics, albeit in a somewhat compressed form. We will introduce a bit of probability theory so that we have not only the language but the mathematical foundations we'll need to understand where thermodynamic descriptions might come from. From there we'll chart a traditional path through kinetic theory to a classical ensemble picture of statistical mechanics. We will mirror that logic in developing quantum statistical mechanics, developing the technology to (a) understand classical statistical mechanics as a limit of the quantum version, (b) be able to describe non-interacting Bose and Fermi systems, and (c) see our first example of a phase transition – Bose-Einstein condensation.

The third part returns to classical systems, but now in the context of interacting degrees of freedom. We will look at cluster and virial expansions, and think about perturbative approaches to systems with interactions. We will then introduce the essential ideas of the RG approach to understanding critical behavior. The lecture notes will close with a brief look at fluctuations and out-of-equilibrium behavior.

Sources and Comments

This is a set of lecture notes prepared while teaching PHYS 526: Statistical Physics (Emory University, Spring 2020 - 2023), and then adapted for this course. It is more verbose than what I will actually write on the board, but far from a comprehensive textbook. I am sure there are both typos and errors in this document, so please email any corrections to:

daniel.m.sussman@emory.edu

The content of these notes is obviously not original to me. They represent a merging of some of the sources that I learned stat mech from, along with many other resources. As I say on the course syllabus, “Graduate-level statistical physics is a subject with many available textbooks and wide disagreements about which one(s) to use.” For these notes I have particularly drawn from:

1. Pathria & Beale (*Statistical Mechanics, 3rd edition*; Primary source) [2],
2. Kardar (*Statistical physics of particles* [3] and *fields* [4]; Primary source(s)),
3. Goldenfeld (*Lectures on Phase Transitions and the Renormalization Group* [5]; general secondary source, especially for the chapter on phase transitions),
4. Preskill (Chapter 10 of his Quantum Information notes for discussion on information entropy and mutual information.) [6]
5. David Tong (Chapter 2 of his lecture notes on Kinetic Theory [7], Chapter 1 of his notes on Statistical Physics [8])
6. Huang (Chapter 5 for some parts of hydrodynamics. Also, the structure of Huang’s text – which is, not surprisingly, echoed in Kardar – inspired the original order in which these lecture notes covered topics. That order is now... quite different.) [9]
7. Sethna (*Entropy, Order Parameters, and Complexity* [10]; general source)
8. Kadanoff (*Statistical Physics: Statics, Dynamics and Renormalization*; [11] general source)

As I write I sometimes leave little notes to myself for parts to improve / add / elaborate on. Those comments [will appear like this](#), so if you see that kind of text you can safely ignore it.

[Here’s an example: I know I should be globally more consistent about \$\vec{q}\$ vs \$\vec{r}\$, especially the contrast between how I define coordinates for Kinetic Theory vs. the section on Interacting Stat Mech.](#)

Part I

Phenomenological descriptions of phases and phase changes

Chapter 1

Thermodynamics

1.1 Introduction

Our education in physics is so steeped in conservation laws that it's a bit odd to think about how mysterious the concepts of “energy” and “heat” have been for so much of human history. Were all things just different manifestations of fire, as at some ancient Greeks allegedly thought? Was heat itself a material substance — caloric fluid — that could flow? Was cold the absence of caloric, or was it a separate material substance — frigorific — that flowed in its own right? Historically, the founding documents of thermodynamics² — Carnot's “Reflections on the Motive Power of Fire” and Kelvin's “Account of Carnot's Theory of the Motive Power of Heat” — grew out of a 19th century desire to understand steam engines and improve their efficiency³. Steam engines and gas laws and Carnot Cycles might be where thermodynamics came from, but they are far from what it *is*. Just as calculus grew to be so much more than counting pebbles, the framework of thermodynamics is vastly larger and more important than it's grimy, coal-fired origins suggest. It has, amusingly, been called everything from a “village witch [13] to a “consistency machine” [14].

So, what *is*— thermodynamics? For our purposes, we'll define it as a phenomenological description of the equilibrium properties of macroscopic systems [3]. The “equilibrium” part of this qualification means that we are trying to understand systems whose properties don't change over whatever amount of time we choose to observe them. This qualification should feel extremely subjective, because it is. The “phenomenological” part is crucial to appreciate: thermodynamics is a fantastically powerful approach, but it is *not* a theory built up from first principles. It is, instead, fundamentally an attempt to build a self-consistent mathematical framework that describes empirically observed behavior of the systems in the world around us. It encodes this mathematical framework in the context of four key laws that together suffice to explain those observations.

Finally, what kinds of “properties” of “macroscopic” systems will we think about? Macroscopic systems are those with a very large number of degrees of freedom — so many that, for instance, despite the very small size of an atom we can often see these systems with our

²Don't be tricked by our modern understanding of “dynamics” in the etymology of this word — the name of the subject is supposed to be a reflection of $\tau\epsilon\rho\mu\eta$ (“heat”) and $\delta\nu\nu\alpha\mu\iota\zeta$ (“power”). The word seems to have been first used — and assiduously hyphenated as “thermo-dynamic” — in Ref. [12]

³If you'd rather trace the history of thermodynamics farther back to things like Boyle's law, be my guest.

eyes. What low-dimensional set of descriptors will we use to characterize systems? We will choose various appropriate *thermodynamic coordinates* for different systems — the length L and tension F in a wire, or the magnetization M and field B for a magnet, or the volume V and the pressure P for a gas⁴ — along with some coordinates we will use to talk about the thermal state of the system. In the rest of this quick review of thermodynamics, we’ll introduce the laws of thermodynamics, see the beautiful mathematical framework they lead to, and we’ll think about how the various thermodynamic coordinates depend on each other (or co-evolve).

1.2 The laws of thermodynamics

1.2.1 0th Law

The “0th” law is not so much part of the original three laws of thermodynamics as an axiom added later to make them logically complete⁵, the zeroth law is a statement of the transitivity of (thermal) equilibrium.

0th law: *If two systems, A and B , are separately in equilibrium with system C , then they are in equilibrium with each other*

The zeroth law suggests that different kinds of thermal equilibrium are not specialized to the details of the physical systems under study — A could be a fluid, B could be a piston filled with gas, and C could be a magnet — and apparently we need some additional thermodynamic coordinate to describe an equilibrium system that tells us the “type” of equilibrium the system is in. We don’t know what properties this additional thermodynamic coordinate has yet, but let’s provisionally agree to call that coordinate the *temperature* of the system. If one really wanted to, one could also dress this more formally. Something like: The zeroth law establishes an equivalence relation between mathematical sets, partitioning the space of all possible thermodynamic coordinates into mutually distinct subsets; we *label* those subsets by something, and we’ll call that something “temperature.” That sounds kind of nice, but is it obvious at this stage that this label has the properties that we intuitively associate with a temperature? That it is even a scalar value and not something more exotic, say, a tensor [18]?

Going back to the phenomenological roots of thermodynamics, the coarse properties of dilute gases were known well before the formulation of thermodynamics — at least to Boyle in the 1660’s — suggesting that even if we don’t yet know the properties of temperature, the zeroth law already gives us a way to *measure* it: just use a sufficiently dilute gas as a thermometer, and compare the result to some known reference system. Empirically we

⁴A word whose etymology I enjoy: “...therefore, by the Licence of a Paradox, for want of a name, I have called that vapour, Gas, being not far severed from the Chaos of the Auncients.” J.B. van Helmont [15]

⁵Maxwell got things started by asserting that there is only one kind of heat; the zeroth law was only more formally introduced as “Characteristic 4” by Caratheodory [16] or as the “zeroth law” by Fowler [17]

might notice that the coexistence of ice, water, and steam can only occur at a single P and T , and then decide to let a “dilute-gas thermometer” — a thermometer in which we measure the product of P and V of our dilute gas in the limit $P \rightarrow 0$ — come into thermal equilibrium with our system and compare it to the value when it comes into equilibrium with an ice-water-steam system:

$$T(^{\circ}K) \equiv 273.16 \left(\lim_{P \rightarrow 0} (PV)_{\text{system}} \right) / \left(\lim_{P \rightarrow 0} (PV)_{\text{ice-water-steam}} \right). \quad (1.1)$$

Clunky, and perhaps only empirically supported as a good definition, but unambiguous!

1.2.2 1st Law

The first law is really just a statement about the conservation of energy, adapted to the fact that “heat” is a kind of energy. To walk through a quasi-historical formulation, let’s first think about “adiabatically isolating” a system, by which we mean somehow preventing the system from exchanging heat with the outside world. We’ll then write the

1st law: *If the state of an adiabatically isolated system is changed by work, the amount of work is only a function of the initial and final coordinates of the system.*

From the fact that there are conditions under which the work done, ΔW , is path- and mechanism-independent, we infer the existence of *another* state function⁶, the internal energy. Letting \mathbf{X} refer to the state of the system, we now have the internal energy of the system, $E(\mathbf{X})$, the first law says that for an adiabatically isolated system

$$\Delta W = E(\mathbf{X}_f) - E(\mathbf{X}_i) \quad (1.2)$$

When thinking about a gas, we might have originally thought that we needed to care about two basic macroscopic properties — the pressure and volume. Thanks to the first two laws we now know we should also care about temperature and internal energy, but we have yet to understand the potential interrelations between all of these quantities.

Note, by the way, that some of the powerful content of the first law comes when we violate the conditions it comes with. If we do work on a system that is *not* adiabatically isolated we now expect $\Delta W \neq E_f - E_i$. Since we still believe that energy is a perfectly nice conserved quantity, this lets us define heat, Q , via

$$\Delta Q = (E_f - E_i) - \Delta W. \quad (1.3)$$

Although E might be a state function, it is clear that ΔQ and ΔW are not *independent* functions of state, so we will adopt the following notation:

$$dE(\mathbf{X}) = \delta Q + \delta W, \quad (1.4)$$

⁶Intuition from classical mechanics: the path independence of the work we do in pushing a frictionless object up and down a hill lets us deduce the existence of a potential energy

Where differentials like dE indicate an exact differential and ones like δQ indicate an inexact differential — i.e., indicate that Q is path-dependent⁷. We will typically think about systems being operated on in a simplified setting where the work is, in fact, path-independent. Before we see that (in the next section), let's finish meeting the other laws of thermodynamics.

1.2.3 2nd Law

Why does heat flow from hot to cold? Why are there no perpetual motion machines that function by turning water into ice while doing work? I think there is something kind of delightful in how practical concerns about burning coal leads directly to the more abstract concept of entropy and implies its inevitable increase. The original formulations of the 2nd Law were due to Clausius and Kelvin, and they wrote it as

Kelvin's statement of the 2nd law: *No process is possible whose sole result is the complete conversion of heat to work (“No ideal engines”)*

Clausius' statement of the 2nd law: *No process is possible whose sole result is the transfer of heat from cold to hot (“No ideal refrigerators”)*

It is probably not clear why these statements are equivalent to each other, or what they have to do with entropy. In Appendix A we trace a path from these statements, through Carnot cycles, to:

Clausius' Theorem: For any cyclic process, with path parameterized by s

$$\oint \frac{\delta Q(s)}{T(s)} \leq 0, \quad (1.5)$$

where the heat $\delta Q(s)$ is an amount of heat delivered to the system by a reservoir or machine at temperature $T(s)$.

A cyclic process is one in which the state of the system at the beginning and end of some process is the same (i.e., the system can be described by the same set of thermodynamic coordinates before and after). Some of the power of Clausius' theorem come from applying it not to a general cyclic process but to a *reversible* process — during such a process one can reverse the direction of the process (e.g., compressing rather than expanding a gas) by reversing all of the inputs and outputs of and on the system (e.g., by cooling rather than heating the system, or doing work on it rather than letting it do work on its environment). During a reversible process the system stays in thermodynamic equilibrium — any changes

⁷Sign convention for this class: work and heat *add* energy to the system.

must be done so slowly that after an infinitesimal change in the inputs or outputs the system can adjust to stay in equilibrium. There are three main conceptual things that this careful setup buys us:

The first is the *existence of entropy*. We first apply Clausius' Theorem to a **reversible** transformation, as promised. By applying it to both the forward and the reversed version of the cycle, we learn that

$$\pm \oint \frac{\delta Q_{rev}(s)}{T(s)} \leq 0. \quad (1.6)$$

Since both the directions give an integral that is less than or equal to zero, the integral must vanish. If we consider breaking up the cycle into segments from state A to state B and back, we see that the integral corresponding to traversing part of the cycle from A to B , the one corresponding to traversing the other part of the cycle from B to A “backwards” must then be equal:

$$\int_A^B \frac{\delta Q_{rev,1}(s)}{T_1(s)} = \int_A^B \frac{\delta Q_{rev,2}(s)}{T_2(s)}. \quad (1.7)$$

That is: the value of the integral as we go from A to B along two different paths is the same! This path independence tells us that its value must depend only on the value of some function evaluated at the endpoints:

$$\int_A^B \frac{\delta Q_{rev}(s)}{T(s)} = S_B - S_A, \quad (1.8)$$

where we've used the letter S to defined some quantity that is a function of the state at the end points. Let's call that quantity “entropy,”⁸ which we have defined up to some arbitrary constant of integration⁹.

The second thing we get from this setup comes by considering an *irreversible* change as we go from state A to B followed by a reversible transformation from B back to A . This is some kind of cyclic process, so Clausius' theorem tells us that

$$\int_A^B \frac{\delta Q}{T} + \int_B^A \frac{\delta Q_{rev}}{T} \leq 0 \Rightarrow \int_A^B \frac{\delta Q}{T} \leq S_B - S_A. \quad (1.9)$$

This tells us, in differential form, that $\delta Q \leq TdS$ for any transformation. If that transformation is adiabatic ($\delta Q = 0$) we've actually learned that $dS \geq 0$. Thus, as a system approaches equilibrium, apparently the arrow of time points in the direction of increasing entropy, since changes in a system's internal state can only increase S . This “entropy increases” principle is probably what most people think the statement of the 2nd law is.

⁸“We might call S the *transformational content* of the body, just as we termed the magnitude U its *thermal and ergonal content*. But as I hold it to be better to borrow terms for important magnitudes from the ancient languages, so that they may be adopted unchanged in all modern languages, I propose to call the magnitude S the *entropy* of the body, from the Greek word $\tau\rho\omicron\pi\eta$, transformation. I have intentionally formed the word *entropy* so as to be as similar as possible to the word *energy*; for the two magnitudes to be denoted by these words are so nearly allied in their physical meanings, that a certain similarity in designation appears to be desirable.” — Clausius [19]. For all that talk of borrowing terms for important magnitudes from ancient languages, note that Clausius tried to name the unit of entropy “the Clausius,” a calorie per degree Celsius.

⁹Actually: a *function* of intergration...[See Jaynes and Pauli](#)

Finally, from the above we know that a reversible process corresponds to heat $\delta Q_{rev} = TdS$. Thus, for reversible processes we can write the first law as

$$dE = \delta W + TdS. \quad (1.10)$$

1.2.4 3rd Law

The 3rd law has a different flavor than the other laws of thermodynamics. The 1st law is something we recognize as a statement about the conservation of energy (along with “heat” being a form of energy), and neatly fits into the framework of what we know about how the microscopic world works. The 0th and 2nd laws are ultimately about an irreversible march towards equilibrium states — they don’t have microscopic analogs, but in the later chapters we’ll try to derive them as a consequence of the statistical properties of large numbers of particles. The 3rd law is different, in that its validity actually depends on what the underlying microscopic laws are — if classical mechanics was all there was, the 3rd law would not be true! In this course we will not deal too much with quantum systems, and so we will be a bit briefer in our treatment of it.

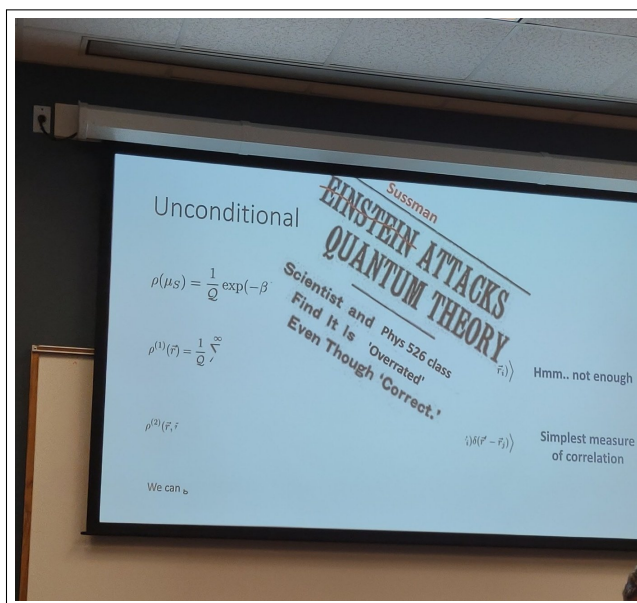


Figure 1.1: Thoughts about quantum mechanics, as communicated in a student presentation in Spring 2023

To set it up the final law of thermodynamics, let's use what we learned from the second law to compute the difference in entropy between two state points at some constant temperature, doing reversible transformations to compute $\Delta S = \int \delta Q_{rev}/T$. With this, the 3rd law is:

Nernst-Simon statement of the 3rd law: The change in entropy associated with a system undergoing a reversible, isothermal process goes to zero as the temperature

approaches 0K:

$$\lim_{T \rightarrow 0} \Delta S(T) \rightarrow 0. \quad (1.11)$$

“Stronger” Nernst statement: The entropy of all systems at absolute zero is a universal constant, which we will define to be the zero point of the entropy scale: $\lim_{T \rightarrow 0} S(\mathbf{X}, \mathbf{T}) = 0$.

The third law has several important consequences, e.g.: not only the entropy by derivatives of the entropy vanish at zero temperature, various response functions we are about to meet (heat capacities, thermal expansivities) vanish at zero temperature, and the need for an infinite number of operations to actually reach zero temperature in a real system.

The “vanishing heat capacity” consequence mentioned above was (basically) Nernst’s original formulation of the 3rd law. Much disagreement ensued, and the third law’s validity and proper framing was hotly contested. This is because of what we said above: the validity of the third law actually depends on what the microscopic laws are. For instance, we’ll soon see statistical mechanical expressions like $S = k \log g$, where g is a measure of the degeneracy of states. The third law tells us that we must have, from the vanishing of entropy, $g = \mathcal{O}(1)$ as $T \rightarrow 0$. But, in classical mechanics, that is simply not true (trues me: we’ll think more about the degeneracy of a classical ideal gas soon enough). On the other hand, we know that as $T \rightarrow 0$, CM is not an appropriate description of the microscopic behavior of particles. It should be hardly surprising, I suppose, that a law whose validity actually rests on quantum mechanics was not well-understood or properly justified before QM itself was.

1.3 Characterizing simple physical systems thermodynamically

Just as thinking about reversible processes let us equate heat, δQ , with a quantity TdS , we will define a slightly less powerful idealization — that of a *quasi-static transformation*: one which is done so slowly that the system is in equilibrium everywhere along the path from its initial to final states¹⁰. Since the system is maintained in equilibrium the whole time, the work done loses any path dependence. That means the work can be directly related to changes in the system’s thermodynamic coordinates.

To be a bit more concrete, let’s divide the state functions, \mathbf{X} , into *generalized displacements* \mathbf{x} and *generalized forces* \mathbf{J} . Some particularly common generalized coordinates are listed in Table 1.3; note that the *displacements are generally extensive* (scale with the size of the system) and the *forces are generally intensive* (are not proportional to system size).

¹⁰A reversible process must also be quasi-static, but a quasi-static process is not necessarily reversible.

System	generalized force		generalized displacement	
Wire	tension	F	length	L
Film	surface tension	σ	area	A
Fluid	pressure	$-P$	volume	V
Magnet	field	B	magnetization	M
Solid	stress tensor	σ_{ij}	strain tensor	ϵ_{ij}

So, for a quasi-static transformation we have

$$\delta W = \sum_i J_i dx_i; \quad (1.12)$$

combined with Eq. 1.10, we can write for a reversible transformation

$$dE = \sum_i J_i dx_i + TdS. \quad (1.13)$$

No big deal, just (perhaps) the most important expression in thermodynamics! Note that this already tells us quite a lot about how different thermodynamic coordinates are interrelated. In general, if the quantities f , x , and y are mutually related¹¹, we can write

$$df(x, y) = \left. \frac{\partial f}{\partial x} \right|_y dx + \left. \frac{\partial f}{\partial y} \right|_x dy. \quad (1.14)$$

In the context of this form of the first law, we see that derivatives of entropy with respect to other coordinates tell us about the temperature and the various conjugate forces:

$$\left. \frac{\partial S}{\partial E} \right|_{\{x_i\}} = \frac{1}{T}, \quad \left. \frac{-\partial S}{\partial x_i} \right|_{\{E, x_{j \neq i}\}} = \frac{J_i}{T}. \quad (1.15)$$

From Eq. 1.13 we see that if there n ways of doing mechanical and chemical work to a system (the n pairs $\{J_i, x_i\}$), then we need $n + 1$ independent coordinates to describe its equilibrium states: entropy is one more extensive generalized “displacement” and it has T as its conjugate generalized “force”. The fact that we need $n + 1$ coordinates (one from each *pair* of coordinates) and not $2(n + 1)$ is because the displacements and forces are interrelated via the internal energy and entropy, as indicated in Eq. 1.15 above. We have substantial freedom in choosing which of each displacement/force pair we pick, and we will later exploit that freedom in defining different thermodynamic ensembles — Do we want to run an experiment on this gas at fixed volume, or at fixed pressure? How about while specifying the temperature, or specifying the entropy?

¹¹With caveats about the derivatives existing, and the functions being 1-1 when holding one independent variable fixed, etc.

1.3.1 Physical properties and response functions

From the mathematical structure described above, thermodynamics quantifies the behavior of a system — for instance, how its thermodynamic coordinates change in response to external perturbations — in several flavors of “response functions.”

Force constants measure how much generalized displacements change in response to generalized forces. An example of a force constant is the isothermal compressibility of a substance:

$$\kappa_T = -\frac{1}{V} \left. \frac{\partial V}{\partial P} \right|_T,$$

where here we have adopted the (thermodynamically) common notation where a subscripted variable reminds us what we are holding constant as we measure or define some quantity. This, “isothermal” = $|_T$ = “constant temperature.”

Thermal responses quantify how a system responds to changes in temperature. An example is the isobaric (constant pressure) expansivity of a substance:

$$\alpha_P = \frac{1}{V} \left. \frac{\partial V}{\partial T} \right|_P.$$

Heat capacities quantify how much heat is needed to change the temperature of a system, $C = \frac{\delta Q}{dT}$. As with all of these response functions, what you choose to hold constant when you measure a heat capacity is quite important. Let’s walk through the difference, for instance, between the heat capacity of a gas at constant V and at constant P :

$$\begin{aligned} C_V &= \left. \frac{\delta Q}{dT} \right|_V = \left. \frac{dE - \delta W}{dT} \right|_V = \left. \frac{dE + PdV}{dT} \right|_V = \left. \frac{\partial E}{\partial T} \right|_V. \\ C_P &= \left. \frac{\delta Q}{dT} \right|_P = \left. \frac{dE - \delta W}{dT} \right|_P = \left. \frac{dE + PdV}{dT} \right|_P = \left. \frac{\partial E}{\partial T} \right|_P + P \left. \frac{\partial V}{\partial T} \right|_P. \end{aligned}$$

Further simplifying to the case of an ideal gas — for which the energy in fact depends only on its temperature and $\left. \frac{\partial E}{\partial T} \right|_V = \left. \frac{\partial E}{\partial T} \right|_P$ — we see that C_P is systematically larger than C_V (in this case, some of the heat in the constant pressure case will be used in changing the volume).

Comment: Joule’s free expansion experiment Joule observed¹² that if you take an extremely dilute, adiabatically isolated gas, and let it expand from V_i to V_f , then the initial and final temperatures are the same: $T_f = T_i = T$. The system was adiabatically isolated, so $\Delta Q = 0$, and no work was done, so $\Delta W = 0$, hence the energy of the system is unchanged: $\Delta E = 0$. From this, we can conclude that in an ideal gas the internal energy (which could, in principle, have depended on its pressure and volume and temperature) actually depends only on temperature: $E(P, V, T) = E(T)$.

For an ideal gas, and knowing already the ideal gas law¹³, then, we have

$$C_P - C_V = P \left. \frac{\partial V}{\partial T} \right|_P = \frac{PV}{T} = Nk_B. \quad (1.16)$$

¹²Before Joule, though, this observation was previously (first?) presented Gay-Lussac in 1806 [20].

¹³And if you don’t not to worry: we’ll derive it soon enough.

where I've used the ideal gas law to, for the first time, introduce our old friend: Boltzmann's constant, $k_B = 1.38064852 \times 10^{-23} \text{ J/K}$. Some people like to say that the smallness of k_B in SI units reflects one of the fundamental facts about the world, namely: atoms are small.

1.4 Less standard thermodynamic systems

Before moving on, a brief comment about the kinds of physical systems we'll talk about is in order. Throughout not only this chapter but this entire set of notes I'll mostly be referring to simple gases and magnets as our go-to example systems. This is because these systems have intuitive physical behavior, are simple to build both microscopic and macroscopic descriptions of, and are rich enough to exhibit interesting physical properties that we can study. That is to say: they are useful for illustrating the fundamental laws and concepts of both thermodynamics and statistical mechanics. However, they are just the tip of the proverbial iceberg: the frameworks we'll learn about in this class are incredibly powerful, and can be applied to a vast array of systems at the frontiers of research.

Soft Matter is home to a number of such examples [21, 22]. “Active” soft matter considers systems composed of degrees of freedom which consume energy in order to either self-propel or exert forces on their environment. These systems are inherently out of equilibrium — and as such challenge many of our traditional notions of pressure, temperature, and other thermodynamic quantities — and researchers are actively exploring how to extend statistical and thermodynamic concepts to these dynamic systems [23, 24, 25].

[granular matter and tensorial temperature](#)

[black holes?](#)

[information, engines, landauer, \[26\]](#)

[biology, non-eq thermo, flies, etc.](#)

All just a few ideas to whet the appetite. Thermodyn and stat phys of complex systems is a vibrant and still evolving field, with many open questions.

1.5 Various thermodynamic potentials

Mechanical equilibrium occurs at a minimum of a potential energy. Similarly, thermal equilibrium similarly occurs at the extremum of an appropriately defined thermodynamic potential. For example, in our discussion of Clausius' Theorem and the 2nd Law (Sec. 1.2.3) we found that the entropy of an adiabatically isolated system increases after any change until it reaches a maximum in equilibrium — although we usually work in the energy ensemble, one could work with an “entropy potential” and define equilibrium states via its maxima. What about systems that are not adiabatically isolated? Or systems which are subject to mechanical work? In this section we will define a handful of thermodynamic potentials that we will frequently want to think about. Briefly, before doing so, it may be useful to think about a purely mechanical analog of the manipulations we are about to do.

So, suppose we have a mass on a spring connected to a fixed all. Let x be the deviation of the masses position away from the equilibrium rest length of the spring, and let the spring potential be $U(x) = \frac{1}{2}kx^2$. Clearly, the state of mechanical equilibrium is at the spring's rest

length, $x = 0$. What happens if we impose an external force, J , on the mass — what will the new state of mechanical equilibrium position be?

I trust you have already thought of at least one of several ways of answering that question. One convenient approach is to define a kind of net energy that encompasses both the spring energy and the external work:

$$H(x, J) = \frac{1}{2}kx^2 - Jx,$$

and find where the variation of this energy with respect to x vanishes:

$$\frac{\partial H}{\partial x} = 0 \Rightarrow x_{eq} = J/k. \quad (1.17)$$

In equilibrium, then, we have a kind of augmented energy (the “*enthalpy*”) whose value is

$$H(x_{eq}, J) = \frac{-J^2}{2k}.$$

Had I already known this functional form of the enthalpy for our system in mechanical equilibrium, I see that I could have computed the equilibrium position by computing

$$x_{eq} = -\frac{\partial H(J)}{\partial J}.$$

We are about to see this pattern repeatedly, but for functions of more variables.

1.5.1 The geometry of Legendre transformations¹⁴

How does this “mechanical analog” of how we will think about different thermodynamic potentials generalize, and what is the right mathematical framework to use? We saw earlier that in the limit of reversible and quasi-static transformations the first law of thermodynamics gave us an expression like $df(x_1, x_2, \dots) = \frac{\partial f}{\partial x_1} \Big|_{\{x_{i \neq 1}\}} dx_1 + \frac{\partial f}{\partial x_2} \Big|_{\{x_{i \neq 2}\}} dx_2 + \dots$. Here f (say, the internal energy) is a function of x_1 and x_2 (say, the volume and entropy of a gas), and in equilibrium terms like $\frac{\partial f}{\partial x_1} \Big|_{x_2}$ represent the conjugate generalized force of the system for particular values of x_1 and x_2 . Sometimes, though, it is not so convenient to always use the generalized displacements as the independent variables. The energy has *entropy* as an independent variable, and if you’re sitting in a lab doing any experiment I challenge you to measure anything as a function of entropy! Consider, in contrast, the much more modest challenge of measuring something as a function of temperature (a feat even the author of this text can manage that once in a while).

Let’s consider the case of a convex function of just one variable: the challenge is to start with a function $f(x)$ which has a derivative $p(x_0) \equiv \frac{\partial f}{\partial x} \Big|_{x=x_0}$, and construct a new function which (a) takes as an independent variable the derivative p rather than the original variable

¹⁴You have likely already encountered this notion, both in your thermodynamics coursework and also in classical mechanics (in deriving the Hamiltonian formalism from Lagrangian mechanics, or vice versa). If you want a quick reminder, or if you haven’t thought about the geometrical interpretation of Legendre transformations, then this subsection might be of interest.

x and (b) contains all of the information that the original function f did. It should be clear that it is *not* sufficient to merely keep track of the derivative of f at every value of x — the resulting information would correspond to an entire family of functions that were shifted relative to each other.

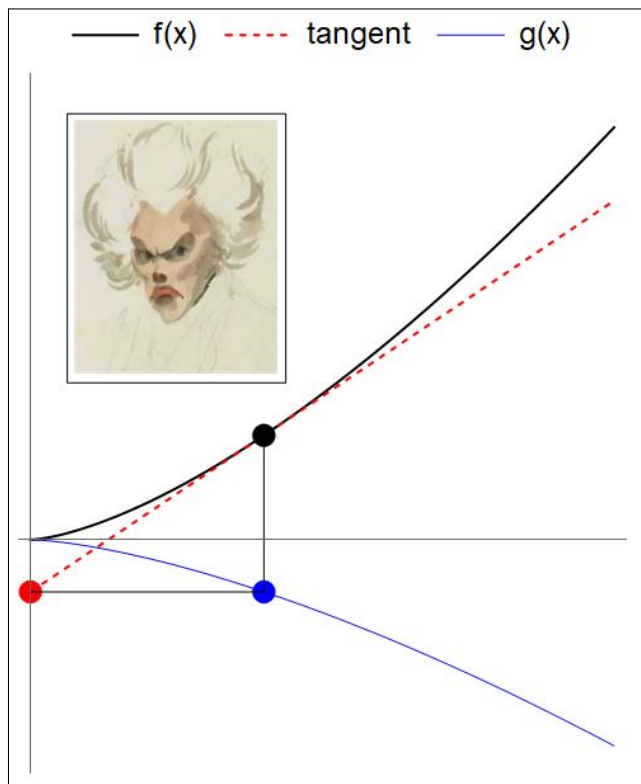


Figure 1.2: A convex function $f(x)$, the tangent line to the curve at some point, and a function $g(x) = f(x) - xp$, where $p = df/dx$ evaluated at that point. The function $g(x)$ clearly encodes information about the value of the original independent variable and the intercept of the tangent line at that point (which itself encodes information both about the slope and the value of the function). Inset: Apparently (and hilariously), the only known portrait of Legendre the mathematician [27, 28].

When we think geometrically about what we are trying to do, as illustrated in Fig. 1.2, I think it becomes clear what is missing. To take the slope of the function as the independent variable *while keeping all of the information encoded in $f(x)$* , we also need to know something equivalent to “where does each of the tangent curves to $f(x)$ intercept the axis?” As illustrated in the figure, a relation for the intercept g as a function of the slope $p(x)$ would allow us to completely and uniquely reconstruct the curve $f(x)$, and hence let us use p as our independent variable without losing any information encoded in the original equation. That is:

$$g = g(p) \quad \text{is equivalent to} \quad f = f(x).$$

All that remains, then, is to figure out what the relation $g(p)$ actually *is* given $f(x)$. The mathematical operation we want is, after all, *the Legendre transformation!* Given a tangent

line through the point $\{x, f(x)\}$ with slope p and intercept g , clearly

$$p = \frac{f - g}{x - 0} \Rightarrow g = f - px, \quad (1.18)$$

and we take this¹⁵ to be our definition of g , the Legendre transform of f . The inverse problem — recovering $f = f(x)$ if we are given $g = g(p)$ — is quite symmetric. Taking the differential of the equation above and using $df = pdx$ we have

$$\begin{aligned} dg &= df - pdx - xdp = -x dp \\ \Rightarrow -x &= \frac{dg}{dp} \end{aligned} \quad (1.19)$$

Hopefully you see the immediate connection between Eq. 1.18 and what we did in our “mechanical analogy” above. In the context of thermodynamics, different thermodynamic potentials (which consider various combinations of intensive and extensive coordinates as the independent variables) simply correspond to different Legendre transformations of the 1st Law’s expression for the internal energy.

1.5.2 Thermodynamic potentials

We have the machinery via Legendre transforms to quickly and easily move between different thermodynamic potentials. Each different potential is just a mix-and-match of choosing one from each conjugate pair of generalized forces and displacements to be the independent variable. Given a physical system in which there are n ways of doing mechanical or chemical work and 1 way of doing thermal work, we would seem to have 2^{n+1} candidate thermodynamic potentials that we could choose between. Are they all equally valid choices, though? In Sec. 1.6.1 we’ll see that the answer is “almost, but not quite”. In the meantime, we briefly review some of the most common potentials that we’ll see in the rest of this text. It is useful to remember that all of these potentials contain equivalent information about the physical system — thinking about one is, in principle, as good as thinking about any of them — and so we are typically free to choose the one which is easiest to conduct an experiment in or perform a theoretical calculation.

Enthalpy

Precisely in analogy with our mechanical example above, if a system is adiabatically isolated ($\delta Q = 0$) but comes to equilibrium under a constant external force, we define our potential of interest to be the *enthalpy*:

$$H = E - \mathbf{J} \cdot \mathbf{x}. \quad (1.20)$$

Variations in this quantity for quasistatic and reversible processes are given by

$$dH = dE - d(\mathbf{J} \cdot \mathbf{x}) = TdS + \mathbf{J} \cdot d\mathbf{x} - \mathbf{x} \cdot d\mathbf{J} - \mathbf{J} \cdot d\mathbf{x} = TdS - \mathbf{x} \cdot d\mathbf{J}. \quad (1.21)$$

Note that in general the work added to the system at constant \mathbf{J} is $\delta W \leq \mathbf{J} \cdot \delta \mathbf{x}$ (where equality occurs for reversible processes). By the first law using our $\delta Q = 0$ condition, we have $dE \leq \mathbf{J} \cdot d\mathbf{x}$, which means that $\delta H \leq 0$ as a system approaches equilibrium — in considering equilibrium states in this condition we are looking for *minima* of the enthalpy.

¹⁵A definition with an overall sign difference is commonly used in classical mechanics

Helmholtz Free energy:

What if we consider a system in the absence of any mechanical work ($\delta W = 0$), but which is allowed to exchange energy with the rest of the universe so that it undergoes transformations at constant *temperature*? For this situation we define the Helmholtz free¹⁶ energy¹⁷

$$F = E - TS, \quad (1.22)$$

which has variations for quasistatic and reversible processes given by

$$dF = dE - d(TS) = TdS + \mathbf{J} \cdot d\mathbf{x} - SdT - TdS = -SdT + \mathbf{J} \cdot d\mathbf{x}. \quad (1.23)$$

Note that Clausius' theorem says that at constant T the heat added to the system is constrained by $\delta Q \leq TdS$. Making use of our $\delta W = 0$ condition, we have $dE = \delta Q \leq TdS$, so $\delta F \leq 0$ — in considering equilibrium states in this set up we are looking for *minima* of the free energy.

Gibbs Free Energy:

What if the system is undergoing an isothermal transformation in the presence of mechanical work done at constant external force? We define the Gibbs free energy by

$$G = E - TS - \mathbf{J} \cdot \mathbf{x}, \quad (1.24)$$

which has variations given by

$$dG = dE - d(TS) - d(\mathbf{J} \cdot \mathbf{x}) = \dots = -SdT - \mathbf{x} \cdot d\mathbf{J}. \quad (1.25)$$

Note that in this case, we have both $\delta W \leq \mathbf{J} \cdot \delta\mathbf{x}$ and $\delta Q \leq TdS$, so $\delta G \leq 0$ — again, we are looking for minima.

Grand Potential:

Traditionally “chemical work” — adding more of some chemical species, or reacting species together and releasing energy in the process, etc. — is treated separately from mechanical work... for chemical equilibrium in the case of no mechanical work, we define the Grand potential by

$$\mathcal{G} = E - TS - \mu \cdot \mathbf{N}, \quad (1.26)$$

where the \mathbf{N} refer to the number of particles of different chemical species, and μ refers to their chemical potentials. Variations in \mathcal{G} satisfy

$$d\mathcal{G} = -SdT + \mathbf{J} \cdot d\mathbf{x} - \mathbf{N} \cdot d\mu. \quad (1.27)$$

¹⁶Certainly this has dimensions of energy, but why are we calling it a “free” energy? In this case, free means “available to do work”, as follows: Suppose you have an engine operating with a hot source and a cold sink at a particular temperature, T_C . It takes in an amount of energy $E = Q_H$, and must discharge Q_C into the sink, which we can write as discharging an entropy $S = Q_C/T_C$. In the process it can do useful work $W = Q_H - Q_C = E - T_C S$, a combination we see is the same as the Helmholtz free energy.

¹⁷Many textbooks — and the IUPAC [29], suggest using the symbol A for this quantity. The A stands for “arbeit,” which translates to “work.” Seeing as we are already using the word “work” for something else, this seems silly. Perhaps if every other free energy was a different language’s translation of the “work” I’d find this convention sufficiently amusing, or at least self-consistent, and follow it.

1.5.3 Thermodynamic stability

We saw above that in equilibrium we are extremizing different thermodynamic potentials. For instance, in the mechanical analog of a particle equilibrating on some potential energy landscape, it would not be enough to find points of stable mechanical equilibrium by $\frac{\partial H}{\partial x} = 0$ — maxima of the enthalpy would be unstable! — but we also require that we are at a minima: $\frac{\partial^2 H}{\partial x^2} > 0$.

The thermodynamic generalization of this involves simultaneously considering variations in each pair of conjugate thermodynamic coordinates we are using to define our system, leading to

$$\delta \mathbf{J} \cdot \delta \mathbf{x} + \delta \mathbf{N} \cdot \delta \boldsymbol{\mu} + \delta T \delta S \geq 0 \quad (1.28)$$

as the condition for thermodynamic stability.

This has some useful consequences. For instance, we immediately are told the *sign* of all of the response functions for equilibrium systems! If we hold everything fixed except for, say, one of the generalized displacements, the equation above immediately says that

$$\left. \frac{\partial x_i}{\partial J_i} \right|_{T, x_{j \neq i}} \geq 0.$$

For instance: in the context of the isothermal compressibility of a equilibrium gas state, we instantly know from thermodynamic stability that

$$\kappa_{T,N} = -\frac{1}{V} \left. \frac{\partial V}{\partial P} \right|_{T,N}$$

must be positive.

Some consequences are a bit more subtle. Later in this course we'll look for the critical point of a gas, which we will identify as a point where $\left. \frac{\partial P}{\partial V} \right|_T = 0$ at some particular value of temperature. If we assume that the isotherm can be written as an analytic expansion¹⁸ then the variation in pressure on the critical isotherm could be written as

$$\delta P = \left. \frac{\partial P}{\partial V} \right|_T \delta V + \frac{1}{2} \left. \frac{\partial^2 P}{\partial V^2} \right|_T \delta V^2 + \dots \quad (1.29)$$

Apparently the condition $-\delta P \delta V \geq 0$ tells us that if we find point where $\left. \frac{\partial P}{\partial V} \right|_T = 0$, thermodynamic stability requires that $\left. \frac{\partial^2 P}{\partial V^2} \right|_T = 0$, too (because it multiplies a δV^3 term, which can take either sign). We'll use this pair of equations in Ch. 2 to begin our exploration of critical points and phase transitions.

1.6 Extensivity, Gibbs-Duhem relations, and Maxwell relations

[re-write / clean up](#) Understanding the thermodynamic properties of systems sometimes involves an intricate dance — transforming from one potential to another in an effort to

¹⁸Which seems innocuous enough — we write down Taylor series approximations all the time, right? Right?

relate the quantity or derivative or response function one wants to know about to one which is (for one reason or another) easier to calculate. As a simple example: it's often easier to calculate C_V for some system theoretically, but easier to actually *measure* C_P experimentally. In this section we meet a few of the standard tricks of the trade.

1.6.1 Extensivity and Gibbs-Duhem

Let's look at the differential for E , including chemical work:

$$dE = TdS + \mathbf{J} \cdot d\mathbf{x} + \mu \cdot d\mathbf{N}. \quad (1.30)$$

In general the extensive quantities are proportional to the size of the system, which we can write mathematically as

$$E(\lambda S, \lambda \mathbf{x}, \lambda \mathbf{N}) = \lambda E(S, \mathbf{x}, \mathbf{N}). \quad (1.31)$$

Let's take the above and differentiate with respect to λ and then evaluate at $\lambda = 1$. This gives

$$\left. \frac{\partial E}{\partial S} \right|_{\mathbf{x}, \mathbf{N}} S + \left. \frac{\partial E}{\partial x_i} \right|_{S, x_{j \neq i}, \mathbf{N}} x_i + \left. \frac{\partial E}{\partial N_\alpha} \right|_{S, \mathbf{x}, N_{\beta \neq \alpha}} N_\alpha = E(S, \mathbf{x}, \mathbf{N}). \quad (1.32)$$

Note that the partial derivatives here are (in order) T , J_i , and μ_α . This leads to what some people write as the fundamental equation of thermodynamics:

$$E = TS + \mathbf{J} \cdot \mathbf{x} + \mu \cdot \mathbf{N}. \quad (1.33)$$

Please note that this is not a requirement — not all systems have energies that are extensive — and so this equation does not have the same standing as the differential form we write the 1st law in; it is simply a statement about the behavior of “most things” (where most things often means systems governed by sufficiently short ranged interactions). If a system is extensive, though we can combine equations 1.30 and 1.33 to get a constraint on allowed variations of the intensive coordinates:

$$SdT + \mathbf{x} \cdot d\mathbf{J} + \mathbf{N} \cdot d\mu = 0, \quad (1.34)$$

which is the *Gibbs-Duhem relation*. Notice that this puts a constraint on the ensembles you are allowed to transform into: for extensive systems you cannot use only the intensive coordinates, because the intensive coordinates are not all independent.

Isothermal compressibility and the grand potential

As an example of the kind of connections between different thermodynamic quantities that we can derive, let's suppose we have a gas in a system in which we allow particle number to fluctuate. The differential of the grand potential is $d\mathcal{G} = -SdT - PdV - Nd\mu$, so we know that the particle number is $N = -\left. \frac{\partial \mathcal{G}}{\partial \mu} \right|_{T, V}$. Perhaps surprisingly, derivatives of the particle number with respect to μ can give you the same information as the isothermal compressibility, which we usually think of as characterizing how the volume of the system changes in response to changes in pressure at constant temperature.

To see this, let's first define $v = V/N$, in terms of which we can write

$$\left. \frac{\partial N}{\partial \mu} \right|_{T,V} = \left. \frac{\partial(V \cdot N/V)}{\partial \mu} \right|_{T,V} = V \left. \frac{\partial v^{-1}}{\partial \mu} \right|_{T,V} = \frac{-V}{v^2} \left. \frac{\partial v}{\partial \mu} \right|_{T,V}.$$

Now, for this system the Gibbs-Duhem relation for this system tells us that

$$0 = SdT + Nd\mu - VdP,$$

which implies that at constant T we have $d\mu = vdP$. Thus, we can relate partial derivatives with respect to μ to those with respect to P :

$$\left. \frac{\partial N}{\partial \mu} \right|_T = \frac{-V}{v^2} \frac{N}{V} \left. \frac{\partial v}{\partial P} \right|_T.$$

We can express the isothermal compressibility in terms of v as

$$\kappa_T = \frac{-1}{v} \left. \frac{\partial v}{\partial p} \right|_T,$$

finally giving us a connection between the curvature of the grand potential and one of these response functions:

$$-\left. \frac{\partial^2 \mathcal{G}}{\partial \mu^2} \right|_T = \left. \frac{\partial N}{\partial \mu} \right|_T = \frac{N^2}{V} \kappa_T \quad (1.35)$$

1.6.2 Maxwell relations

“Maxwell relations” follow from combining thermodynamic relationships with the basic properties of partial derivatives. We already saw in Eq. 1.14 that for mutually related f , x , and y , we have

$$df(x, y) = \left. \frac{\partial f}{\partial x} \right|_y dx + \left. \frac{\partial f}{\partial y} \right|_x dy;$$

we will then combine this with symmetry of second derivations,

$$\frac{\partial}{\partial x} \left. \frac{\partial f}{\partial y} \right|_x = \frac{\partial}{\partial y} \left. \frac{\partial f}{\partial x} \right|_y, \quad (1.36)$$

to relate various thermodynamic derivatives.

Example

For instance, let's start with $dE = TdS + J_i dx_i$. We can immediately write

$$T = \left. \frac{\partial E}{\partial S} \right|_{\mathbf{x}} \quad \text{and} \quad J_i = \left. \frac{\partial E}{\partial x_i} \right|_{S, x_j \neq i}, \quad (1.37)$$

and the equality of mixed partials then tells us

$$\frac{\partial^2 E}{\partial S \partial x_i} = \left. \frac{\partial T}{\partial x_i} \right|_S = \left. \frac{\partial J_i}{\partial S} \right|_{\mathbf{x}}. \quad (1.38)$$

The last equality there is usually what is called a Maxwell relation.

Strategy for deriving Maxwell relations

There are several tricks to remembering how to rapidly find the Maxwell relation relevant to a particular expression. Logically, though, it's not so hard to always construct them on the fly. Suppose someone asks you to find a Maxwell relation for

$$\left. \frac{\partial A}{\partial B} \right|_C, \quad (1.39)$$

as long as A and B are not conjugate pairs (in which case, see the Gibbs-Duhem example above) we'll find a Maxwell relation by: (1) write down the fundamental expression for dE , (2) transform it so that B and C are differentials whereas A is not, and (3) profit.

Worked example of finding Maxwell relations: I want to know $(\partial\mu/\partial P)|_T$ for an ideal gas.

Step 1 We write

$$dE = TdS - PdV + \mu dN. \quad (1.40)$$

Step 2 We note that μ is already in a position to appear in first derivatives. Moving on,

$$d(E + PV) = TdS + VdP + \mu dN \quad (1.41)$$

$$d(E + PV - ST) = -SdT + VdP + \mu dN. \quad (1.42)$$

We did not really care what the name of $(E + PV - ST)$ was (the Gibbs Free Energy, I guess), let's just call it Y . Clearly

$$\mu = \left. \frac{\partial Y}{\partial N} \right|_{T,P} \quad \text{and} \quad V = \left. \frac{\partial Y}{\partial P} \right|_{T,N}, \quad (1.43)$$

so

$$\left. \frac{\partial \mu}{\partial P} \right|_N = \left. \frac{\partial V}{\partial N} \right|_P. \quad (1.44)$$

Step 3 We're done.

1.7 Geometry, thermodynamics, and classical mechanics

It's outside the scope of the course to really go into this, but you deserve to have at least been pointed in the direction of a very different, much more geometrical¹⁹ view of what is happening in all of these thermodynamic manipulations. This whole section needs to be written!

A common complaint — spanning generations of scientists [31] — is that there is something fundamentally ugly and fussy about the mathematical notation we constantly carry around in our thermodynamic calculations. In contrast to, say, the beautiful mathematical structure of classical mechanics, thermodynamic manipulations feel simultaneously mathematically simple yet obscure. This is probably because in the course of deriving various thermodynamic relations we have a shifting set of variables whose (in)dependence seems to change over the course of a calculation, requiring us to decorate our equations with reminders of what

¹⁹Algebra is the offer made by the devil to the mathematician. The devil says: I will give you this powerful machine, it will answer any question you like. All you need to do is give me your soul: give up geometry and you will have this marvelous machine" [30]

thermodynamic coordinates are being held fixed at every step. This is a shame, as there are in fact deep links between the mathematical structure of thermodynamics and that of classical mechanics, all with a beautiful geometrical underpinning. I won't explain all of this, but let me merely provoke you with the following.

First, let's stare at a pair of easily derived Maxwell relations:

$$\left. \frac{\partial T}{\partial V} \right|_S = - \left. \frac{\partial P}{\partial S} \right|_V; \quad \left. \frac{\partial S}{\partial V} \right|_T = \left. \frac{\partial P}{\partial T} \right|_V. \quad (1.45)$$

We might not love the notation, but we know what it means. For instance, we look at the left-most partial derivative above and internally translate it as “at this moment we are thinking of the temperature as a function of the independent variables volume and the entropy, and we take the derivative of the temperature with respect to the volume holding entropy fixed.” Clunky, but fine.

Next, let's stare at something we (perhaps) feel more kinship with: Hamilton's equations for a particle on a line with position q , momentum p , and governed by a Hamiltonian H :

$$\left. \frac{\partial p}{\partial t} \right|_q = - \left. \frac{\partial H}{\partial q} \right|_t; \quad \left. \frac{\partial q}{\partial t} \right|_p = \left. \frac{\partial H}{\partial p} \right|_t. \quad (1.46)$$

Do you find your self objecting to this? Is your mind rebelling²⁰ at the unfamiliar notation? You would have rather I had written things like $\frac{dp}{dt}$ rather than $\left. \frac{\partial p}{\partial t} \right|_q$, but that's only because in classical mechanics you are used to unambiguously knowing what the independent variables are at all times! After getting over your discomfort and staring for longer, you may notice that if we simple relabel the variables in Eq. 1.46, sending $p \rightarrow T$, $t \rightarrow V$, $q \rightarrow S$, and $H \rightarrow P$ we recover *exactly* the Maxwell relations above.

Is this a coincidence? Would your answer to that change if I were to remind you that Hamilton's principle function²¹ A for this one-dimensional problem has as its total differential

$$dA = \frac{\partial A}{\partial q} dq + \frac{\partial A}{\partial t} dt = pdq - Hdt?$$

It's almost as though we could invent a Legendre transformation to the function $X = A - pq$ and derive Hamilton's equations using *exactly* the same logic we used to derive the Maxwell relation [32]!

There are indeed rich connections here, and although exploring them is tangential to the goals of this course you deserve to know it is there! The history here goes back (at least) to Carathéodory's use of Pfaffians in his attempt to axiomatize thermodynamics [16, 33], and continues with Poisson bracket formulations of thermodynamics [34] and approaches based on contact geometry [35]. If you are the kind of person who finds these connections interesting, dive in!

²⁰Mine certainly is: I'm writing this book and I still think it looks weird!

²¹Which is just the classical action plus an undetermined constant; I'm using the very non-standard notation “ A ” instead of S for it here only because it is appearing so close to the entropy, for which we are already using S .

1.8 Problems

1.8.1 Simple equations of state

An “equation of state” is an equation that relates the different thermodynamic coordinates of a physical system. Perhaps the single most famous one is the *ideal gas equation of state*: $PV = Nk_B T$.

Start from the 1st Law for a gas in which there is only one way of doing mechanical work (i.e., start from $dE = TdS - PdV$), and think of both the energy and the entropy as functions of T and V . Show that, in fact, the ideal gas law implies that E is only a function of T .

1.8.2 Deriving simple equations of state

An “equation of state” is an equation that relates the different thermodynamic coordinates of a physical system. Perhaps the single most famous one is the *ideal gas equation of state*: $PV = Nk_B T$.

(A) Suppose that in a certain gas the internal energy can be expressed as

$$E = \alpha N \left(\frac{N}{V} \right)^\psi f \left(\frac{\psi S}{Nk_B} \right),$$

for some constants α and ψ and some function f . For what function $f(x)$ will this system obey the ideal gas equation of state?

(B) Considering an isentropic (constant entropy) process at fixed number of particles, for what exponent γ will this system (i.e., with the function f you found above) satisfy $PV^\gamma = (\text{constant})$? What is the ratio C_P/C_V ?

1.8.3 A true pearl...

Your friend hands you a box of...well, you're not exactly sure what. But the box has a piston on it, and your friend tells you that the relationship between the pressure and the energy density $u = E/V$ is $u = 3p$. Using the first law of thermodynamics, and perhaps a relevant Maxwell relation, what is the functional form for the energy density as a function of temperature, $u(T)$?

1.8.4 A black hole, thermodynamically

Hawking famously calculated how black holes emit radiation, finding that they are perfect black-bodies with a temperature of

$$T = \frac{\hbar c^3}{8\pi G M k_B},$$

where G is the gravitational constant. (A) Assuming, a la Einstein, that the energy of a black hole is $E = mc^2$, what is the specific heat of a black hole? How do you interpret the

sign of your answer? **(B)** Assuming the entropy of a black hole with zero mass is zero (seems reasonable!), what is the entropy of a black hole with finite mass? Please write your answer in terms of the surface area of the black hole (using the Schwarzschild radius for a black hole with no charge and no angular momentum, $R_s = 2GM/c^2$), taking the unit of area to be the Planck length squared ($l_p^2 = \hbar G/c^3$).

1.8.5 Black holes and Maxwell relations

It turns out there is a long history of people trying to use thermodynamics to understand the behavior of black holes. One can²² write an equivalent of the first law of thermodynamics for a charged, rotating black hole as

$$dE = \frac{T_H}{4} dA + \Omega dJ + \Phi dQ,$$

where E is the internal energy, T_H is a temperature associated with the Hawking radiation the black hole is emitting, A is the surface area of the event horizon²³, Ω is the angular velocity of the rotating black hole, J is its angular momentum, Φ is the electrostatic potential, and Q is the black hole's electric charge. Phew!

I am curious about the equilibrium properties of these thermodynamic black holes. For instance, suppose I want to know how the “temperature” of a black hole varies with the angular velocity at fixed A and Q ... write down a Maxwell relation to fill in:

$$\left. \frac{\partial T_H}{\partial \Omega} \right|_{A,Q} = ?$$

1.8.6 Differences in specific heats

The specific heats at constant volume and pressure can be written (somewhat differently from what's in the lecture notes — can you see why the expressions are related) as

$$C_V = T \left. \frac{\partial S}{\partial T} \right|_V, \quad C_P = T \left. \frac{\partial S}{\partial T} \right|_P.$$

The isothermal compressibility and the isobaric expansivity are defined, respectively, as

$$\kappa_T = \frac{-1}{V} \left. \frac{\partial V}{\partial P} \right|_T, \quad \alpha_P = \frac{1}{V} \left. \frac{\partial V}{\partial T} \right|_P.$$

Write an expression for $(C_P - C_V)$ in terms of the quantities T , V , α , and κ . It may help to think of S as a function of T and V and to think of V itself as a function of T and P .

²²perhaps...there's lots of debate

²³Notice that here, interestingly, the surface area of the black hole is playing the role of an entropy-like quantity (in that it is the conjugate variable to a temperature-like variable). This has led to all sorts of interesting speculations about holography.

1.8.7 Thermodynamic interrelations

In class I defined thermodynamics as “A phenomenological description of the equilibrium properties of macroscopic systems.” An alternate definition lurking both in the literature and in many of our minds might be that it is “a zoo of partial derivatives, transformations, and relations” [10] We often have to make choices about what energy or free energy to work with, what variables we want to hold fixed, and while doing all of this we have to keep in mind that because of the interrelations between thermodynamic free energies, generalized displacements, and generalized forces there are certain mathematical relationships we have to keep track of²⁴.

(A) First derivative triangles: Consider a general mathematical interrelationship of the form $A dx + B dy + C df = 0$. Separately calculate $\left. \frac{\partial f}{\partial x} \right|_y$, $\left. \frac{\partial y}{\partial f} \right|_x$, and $\left. \frac{\partial x}{\partial y} \right|_f$; use this to then show the so-called “triple product relation,” $\left. \frac{\partial f}{\partial x} \right|_y \left. \frac{\partial y}{\partial f} \right|_x \left. \frac{\partial x}{\partial y} \right|_f = -1$. Next, write a relevant expression for $dS(E, V)$ and use a triple-product relation to show that $\left. \frac{\partial E}{\partial V} \right|_S = -P$

(B): Symmetric second derivatives What are usually called Maxwell relations are simple consequences of the symmetry of second derivatives: $\frac{\partial^2}{\partial x \partial y} = \frac{\partial^2}{\partial y \partial x}$. Show that $\left. \frac{\partial V}{\partial T} \right|_P = - \left. \frac{\partial S}{\partial P} \right|_T$.

1.8.8 Applied thermodynamic interrelations

Recall that *enthalpy* of a gas is defined as by the Legendre transform of the internal energy $H = E + PV$, with $dH = TdS + VdP$. In this and the next problem we’re going to think about a response function which describes how the temperature of a adiabatically isolated fluid changes when it is forced through a valve. For this problem, define

$$\mu \equiv \left. \frac{\partial T}{\partial P} \right|_H$$

(A) First, use the triple-product rule (or any other thermodynamic manipulation) to show that

$$\mu = - \frac{\left. \frac{\partial H}{\partial P} \right|_T}{\left. \frac{\partial H}{\partial T} \right|_P}$$

(B) One of those terms (which one?) is just the heat capacity at constant pressure, C_P . Do a few manipulations (probably involving a Maxwell relation) to show that we can write what is sometimes called the “Joule-Thomson” coefficient as

$$\mu = \frac{1}{C_P} \left(T \left. \frac{\partial V}{\partial T} \right|_P - V \right)$$

1.8.9 Heat capacity of a specific system

Suppose your friend conducts an experiment and finds that, over some range of T , the a generalized displacement x requires a generalized force

$$J = ax + bT + cTx,$$

²⁴or, more positively, can exploit to calculate things!

where a , b , c are constants. In a separate experiment she finds that the heat capacity of the system at constant displacement is proportional to temperature:

$$C_x = T \left. \frac{\partial S}{\partial T} \right|_x = A(x)T,$$

for some function $A(x)$.

(A) Given that we have an expression for J above, find an appropriate Maxwell relation and use it to evaluate

$$\left. \frac{\partial S}{\partial x} \right|_T.$$

(B) Does the function $A(x)$ actually have any dependence on the displacement x ?

(C) Define the entropy at $T = 0$ and $x = 0$ to be $S(T = 0, x = 0) \equiv S_0$. Assuming that the “experimental” results described above hold over the entire range of parameters that are relevant, write down an expression for $S(T, x)$.

(D) Use all of the above results to predict what the heat capacity at constant generalized force,

$$C_J = T \left. \frac{\partial S}{\partial T} \right|_J,$$

should be as a function of T , J , and the constants a , b , c .

1.8.10 The shape of a phase boundary

Whenever two different phases of a substance meet along an equilibrium phase boundary the temperatures, pressures, and chemical potentials of those phases have to match (or else the physical boundary between the phases would be unstable, or there would be a directed flux of heat and/or particles, meaning the system would be out of equilibrium). Let’s use this fact to derive a relation for the shape of the phase boundary in the P - T plane.

So, suppose you have two phases of a substance meeting in equilibrium at a phase boundary. Apply the Gibbs-Duhem relation to each phase in the case that there is a small shift in temperature, ΔT , along the phase boundary. Derive the following equation for the slope of the coexistence line:

$$\frac{dP}{dT} = \frac{s_1 - s_2}{v_1 - v_2},$$

where $s_i = S_i/N_i$ and $v_i = V_i/N_i$. Since it is typically difficult to experimentally measure something like the entropy per particle, connect the difference in entropies to the heat flow, Q , needed to induce a phase change, and write your formula in terms of the latent heat of the phase change.

1.8.11 Stability of the grand potential

Suppose you are considering a system whose volume is fixed but which can exchange both energy and particle number with a reservoir. By considering the thermodynamic stability of your system, show that:

$$(A) C_{V,N} \geq 0 \quad \text{and} \quad (B) \left. \frac{\partial N}{\partial \mu} \right|_{V,S} \geq 0$$

1.8.12 Confined ideal gas thermodynamics

A devious experimental physicist has confined a gas of N classical particles of mass m that don't interact with each other in a harmonic trap. The potential energy associated with this trap is

$$U(r) = \frac{1}{2}kr^2,$$

where r is the distance from the center of the trap and k is the strength of confinement. There is no “volume” of the gas, nor is there a homogeneous pressure. But we can still talk about a mechanical work done if the experimenter were to vary the strength of confinement quasi-statically: $dW = Adk$, where by dimensional analysis A is something with units of area. Thus, the analog of our first law of thermodynamics is

$$dE = TdS + Adk.$$

(A) Argue on physical grounds that $A = \frac{1}{2}N \langle r^2 \rangle$, where $\langle r^2 \rangle$ is the average mean squared displacement of a particle from the center of the trap.

(B) Suppose this experiment is being carried out in a temperature-controlled setting. If I tell you that the relevant free energy is²⁵

$$\beta F = -\log \left[\frac{1}{N!} \left(\frac{\pi}{\beta h} \sqrt{\frac{2m}{k}} \right)^{3N} \right],$$

where $\beta = (k_B T)^{-1}$ and h is Planck's constant, evaluate A by taking an appropriate partial derivative. Compare your result to the ideal gas law.

(C) Write an expression for the difference between the heat capacity at constant k and the heat capacity at constant A .

1.8.13 Hyperelasticity and Reynolds dilatancy

In the 1880s Osborne Reynolds pointed out the general tendency of soils, sands, and other granular materials to dilate when subject to stress [36] (You may have noticed a consequence of this if you've ever had the experience of walking on a wet beach and seeing the sand seem to swell and dry out around each of your footsteps). This dilation is (a) a nonlinear effect and (b) not a feature of elastic solids in general – for instance, sheared fiber network contract rather than expand under constant-volume strains – in general it is a quite complex problem to understand at the microscopic level. Let's see if we can, however, use some thermodynamic-like consistency machinery to see if this Reynolds dilatancy is related to other properties of the material.

First, to set up the problem we define several variables for a system schematically depicted in Fig. 1.3. Let h_0 be the original height of an isotropic solid of cross sectional area A initially under some pressure p . It is subject to a shear strain γ which results in both a shear stress σ

²⁵We'll derive this in a few chapters!

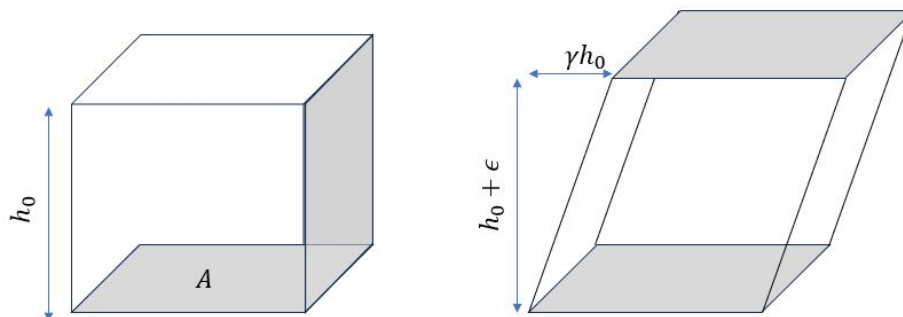


Figure 1.3: An isotropic granular solid of height h_0 and cross-sectional area h is subject to a shear strain γ . An expansion of the material by an amount $\epsilon \propto \gamma^2$ is observed.

and an observed dilation of the material, which causes the height to change to $h = (1 + \epsilon)h_0$. The shear modulus (G) and Young's modulus (Y) of the material are defined as

$$G = \left. \frac{\partial \sigma}{\partial \gamma} \right|_{\epsilon} \quad \text{and} \quad Y = \left. \frac{\partial p}{\partial \epsilon} \right|_{\gamma},$$

and we expect that we can write the dilation at constant pressure as a low-order expansion in γ :

$$\epsilon_p = \frac{R_p}{2} \gamma^2 + \dots$$

(A) Why doesn't the above equation contain a term which is linear in γ ? Give a short answer that includes the phrase "By symmetry..."

Thanks to your answer in part (A), we know that we will quantify the Reynolds dilatancy by the material parameter

$$R_p = \left. \frac{\partial^2 \epsilon}{\partial \gamma^2} \right|_p.$$

Nonlinear elasticity is hard, but apparently something more than a linear elastic theory is needed to understand R_p . One simplification is to assume that our system *hyperelastic* – a material whose behavior can be modeled as coming from an elastic potential energy. Suppose the internal energy of our system is

$$dE = -pAdh - \sigma Ahd\gamma,$$

where the first term looks exactly like a " $-pdV$ " term and the other represents a way of doing mechanical work by shearing the system – all that is missing from our usual thermodynamic set up of the first law is a thermal work term²⁶.

(B) From all of this: derive a relationship between the R_p , the shear and Young's modulus, and a derivative of one of the moduli (assuming you are working in the limit $\gamma \rightarrow 0$, and keep only the leading terms)

²⁶Missing because for granular systems we typically think of the system as effectively athermal.

Chapter 2

The phenomenology of phase transitions

[ADD complete introduction, full of wonder](#)

In the last chapter we did phases in equilibrium...what about transitions between phases? We will use the tools of statistical physics to understand these phase transitions and the emergence of collective phenomena, too.

[Add figure](#) with phase transition diagram of water (triple-point named by James Thomson [37], Kelvin's older (and completely overshadowed) brother), and of something else.

2.1 Discontinuous transitions

Short version: many phase transitions around us happen quite abruptly (water freezing, etc), and are characterized by kinks in the free energy and discontinuities in their first (and higher-order) derivatives. They also feature things like coexistence of phases at some state points, hysteresis, etc etc...

2.1.1 Abrupt transitions in fluids

Suppose we think about the Gibbs potential for a gas:

$$G(T, P, N); \quad dG = -SdT + VdP + \mu dN, \quad (2.1)$$

where, e.g., $V = \left. \frac{\partial G}{\partial P} \right|_T$. What does the property V look like as we vary P and T ? We show this in Fig. 2.1 on the left. Note that for isotherms below some critical temperature, V suddenly jumps as P is varied. On the right of the figure we see that if we go to a different ensemble where we are controlling T and V , there is a range of volumes (or densities) at which there is liquid-gas *coexistence*.

How could we come up with a theoretical model that reproduces this kind of behavior? Consider the van der Waals equation of state for a fluid (which we will *derive* later in this course):

$$P = \frac{k_B T}{V/N - b} - a \frac{N^2}{V^2}. \quad (2.2)$$

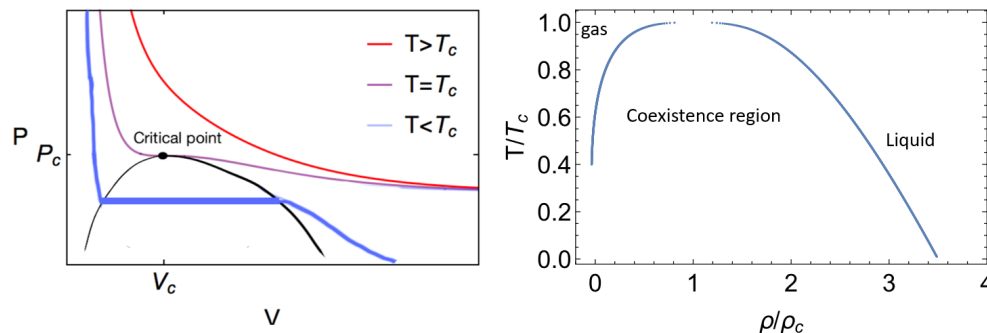


Figure 2.1: **Phase and coexistence curves of a liquid-gas system** (Left) Pressure vs volume curves at constant temperature for a liquid-gas system. (Right) Controlling T and V , there is a coexistence region in which it is favorable for the system to separate into a mixture of phases rather than remain homogeneous.

This equation is like a generalization of the ideal gas law that accounts for both weak attractions between the particles (in that there is a term proportional to ρ^2 where ρ is the density which tends to lower the pressure) and very short-ranged harsh repulsion between the particles (in that the total volume is available to the system is effectively reduced by bN). A plot of this equation of state for several temperatures is shown in Fig. 2.2. At high temperatures and large volumes these curves look fine, but something is clearly unphysical about this equation at low temperatures. Namely, it predicts a region in which increasing pressure actually *increases* the equilibrium volume of the system. This violates basic considerations of thermodynamic stability.

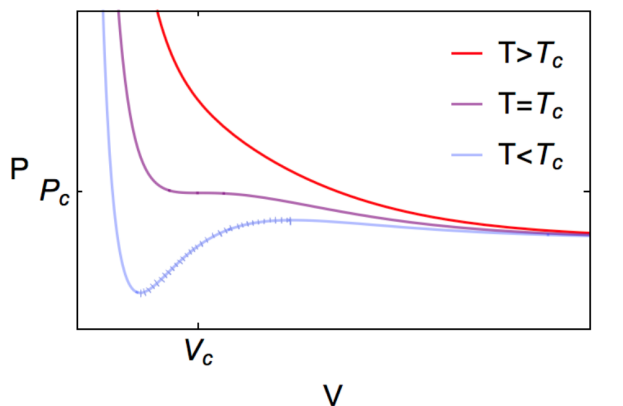


Figure 2.2: **Compressibility vs reduced pressure for several substances** The van der Waals equation of state for several temperatures. Hatched lines for the coldest curve indicate an *unphysical* part of the equation of state.

2.1.2 Maxwell construction

Our phenomenological strategy for patching things up and getting rid of the unphysical nature will be the “Maxwell construction,” in which we replace part of the equation of state

along an isotherm with a straight line... This will make the curves look like real life (cf. Fig. 2.1), but where should we put it?

In the Gibbs free energy, coexistence will occur when $G_{liquid} = G_{gas}$, and if we fix both T and N we know that $\frac{\partial G}{\partial P}|_{T,N} = V$. Let's integrate both sides of this equation from P_{liquid} to P_{gas} to get

$$\Delta G = \int_{P_l}^{P_g} V(P)dP. \quad (2.3)$$

We know from the coexistence condition that $\Delta G = G_{liquid} - G_{gas} = 0$. On the other hand, let's re-write the integral as

$$\int_{P_l}^{P_g} V(P)dP = \int_{P_l}^{P_{min}} V(P)dP + \int_{P_{min}}^{P_{unst}} V(P)dP + \int_{P_{unst}}^{P_{max}} V(P)dP + \int_{P_{max}}^{P_g} V(P)dP \quad (2.4)$$

This has a nice geometrical interpretation (see Fig. 2.3), which tells us that the RHS vanishes when there is “equal area” around parts of the original equation of state: apparently *that* is how we decide to draw our new equation of state so that things are thermodynamically self-consistent.

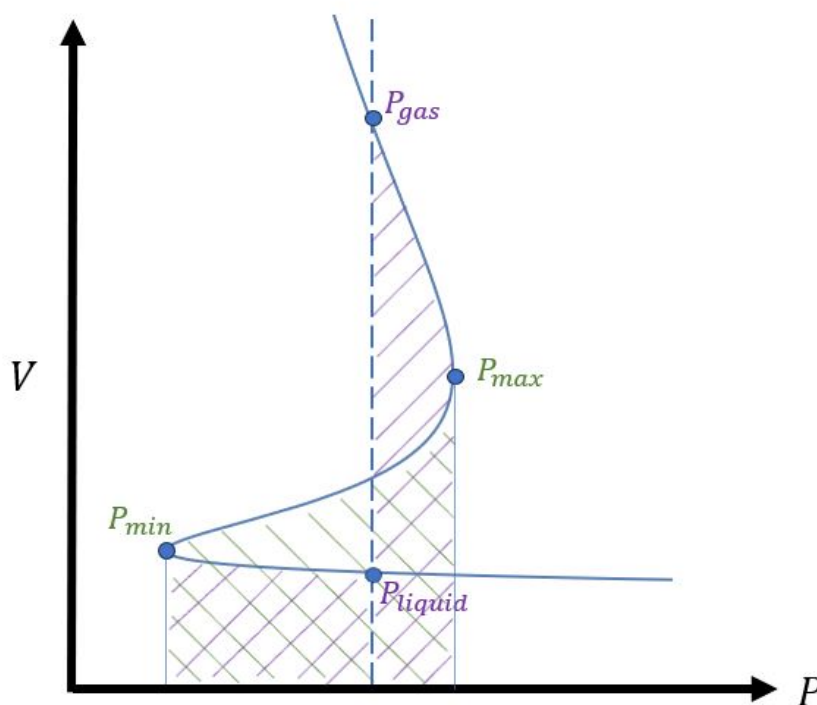


Figure 2.3: **Maxwell's equal area construction** We geometrically find the position of the discontinuity in pressure by an equal area construction on Eq. 2.4.

2.2 The law of corresponding states

Let's again write the van der Waals equation,

$$P = \frac{k_B T}{v - b} - \frac{a}{v^2}, \quad (2.5)$$

where $v = V/N$ is just the volume per particle, and a and b are parameters that depend on the microscopic details of the systems under study (generically related to the characteristic size of the interacting particles, and how strong the attractive interactions are).

We have seen that this equation goes from being monotonic to non-monotonic at some critical value of temperature, and that at T_c there is a point where the critical isotherm is flat. We know from our discussion of thermodynamic stability that, at this point, not only the first but also the second derivative of the pressure with respect to volume must vanish. Where is this critical point? We find it by²⁷ writing

$$\frac{\partial P}{\partial v} = 0 = -\frac{k_B T}{(v - b)^2} + \frac{2a}{v^3} \quad (2.6)$$

$$\frac{\partial^2 P}{\partial v^2} = 0 = \frac{2k_B T}{(v - b)^3} - \frac{6a}{v^4}. \quad (2.7)$$

Dividing the first equation by the second equation gives us $\frac{v_c - b}{2} = \frac{v_c}{3} \Rightarrow v_c = 3b$, and we can then plug this critical volume per particle into the above expressions to find the critical point on the critical isotherm:

$$v_c = 3b, \quad k_B T_c = \frac{8a}{27b}, \quad P_c = \frac{a}{27b^2}. \quad (2.8)$$

Suppose we wanted to know if the van der Waals equation was actually any good, and we convinced someone to do an experiment. In it's basic form, comparing experimental data to the van der Waals equation would simultaneously test both the quality of the van der Waals equation *and the ability to measure the microscopic parameters* that enter the van der Waals equation, a and b . But, in general, measuring the microscopic Hamiltonian is really hard! Better would be to try to locate the critical point of the fluid in question, noting that the van der Waals equation predicts that the ratio

$$\frac{P_c v_c}{k_B T_c} = \frac{3}{8} = 0.375 \quad (2.9)$$

is independent of the microscopic details, and should be the same for all fluids! This combination is often called the “critical compressibility factor”, and experimentally, this estimate

²⁷A more elegant approach – which I will do in class – is to note that the van der Waals equation can be written as a cubic polynomial in v : $pv^3 - (pb + k_B T)v^2 + av - ab = 0$. For $T > T_c$ there is just one real root of this equation, and for $T < T_c$ there are three real roots; precisely at the critical point all three roots must be real and identical, so we must be able to write the equation as $p_c(v - v_c)^3 = 0$. Comparing the coefficients of this to the more general cubic expression immediately gives us the critical pressure, volume, and temperature.

is okay²⁸ Many real substances have a range for this ratio between 0.28 and 0.33 (carbon tetrachloride is about 0.27, argon is about 0.29, He⁴ is about 0.31, water is pretty low at 0.23) but similar substances seem to differ from the 3/8 ratio in similar ways; this is encouraging when we think about the types of approximations we made in deriving the van der Waals equation in the first place. For instance, we assumed the attractive potential was isotropic, so the fact that diatomic gases have a particular ratio while monatomic gases a different one feels consistent with the way we constructed our theory.

Even better, this suggests that we write down a *dimensionless version* of the van der Waals equation, in which we measure everything relative to the value at the critical point. Defining $P_r = P/P_c$, $v_r = v/v_c$, and $T_r = T/T_c$, we can write the equation of state as

$$P_r = \frac{8}{3} \frac{T_r}{v_r - \frac{1}{3}} - \frac{3}{v_r^2}, \quad (2.10)$$

a universal equation of state that all fluids obeying the van der Waals equation of state would be expected to obey. Remarkable! With no other parameters, all fluids are expected to have the same equation of state, and similarly all thermodynamic properties which are derived from the equation of state should also be in correspondence. So, if two different systems are in states that are characterized by the same values of v_r and T_r , one would predict their P_r would match as well: this is the *law of corresponding states*.

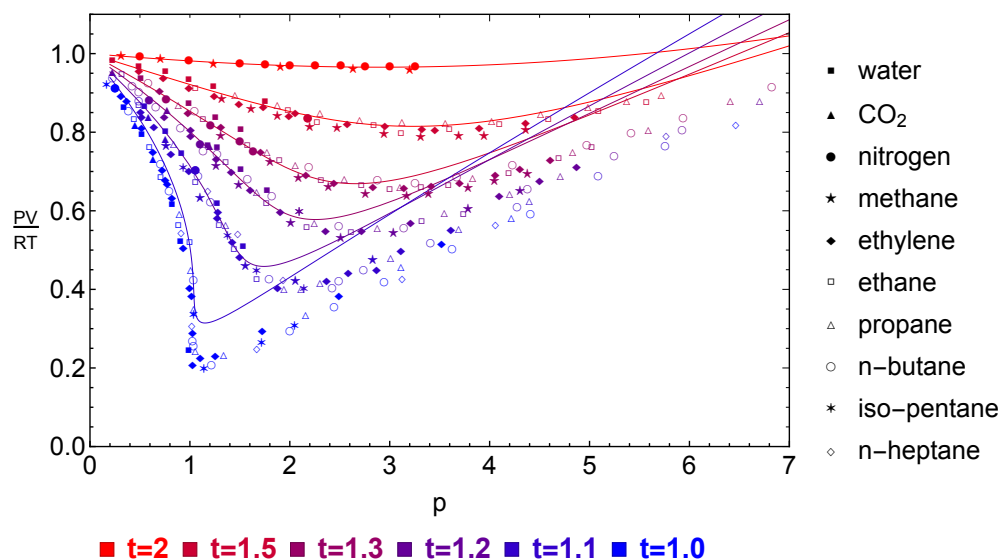


Figure 2.4: **Compressibility vs reduced pressure for several substances** Thin lines are curves generated by the van der Waals equation of state at various reduced temperatures; reduced temperature for the actual measurements is indicated by color. Data taken from Ref. [38].

²⁸Different textbooks will compare 3/8 to the experimental numbers I'm about to quote and declare the agreement “a little high” [8] or “reasonable” [5] or “not good” [3] or “a long way from the van der Waals value” [2]. Agreement with experiment, and what we think that should mean for a physical theory, is in the eye of the beholder.

This is our first look at a type of *universality*, albeit of a different character than we will see when we look at universality near a critical point. The law of corresponding states is expected to hold not only close to the critical point, but in fact *everywhere in the phase diagram*. Experimentally the law of corresponding states is often well satisfied, even by fluids that do not obey the van der Waals equation! Generally, the expectation set forth by the corresponding states hypothesis is that we should be able to write the equation of state as

$$\frac{P}{P_c} = f\left(\frac{T}{T_c}, \frac{v}{v_c}\right) \quad (2.11)$$

for some function f which might be related to the van der Waals derivation, but might be completely different. A demonstration of this correspondence is shown in Fig. 2.4, showing the ratio Pv/T vs reduced pressure for several substances. None of them follow the van der Waals equation particularly well, but do demonstrate the expectation that, for instance, this ratio when plotted against reduced pressure at a fixed reduced temperature should be independent of the fluid being measured.

2.3 Continuous transitions, and the critical point behavior of a van der Waals fluid

We want to investigate why we should expect the law of corresponding states to hold, *at least near a critical point*. So let's imagine expanding the equation of state close to the critical point²⁹. One of the things we'll see is that as T gets close to T_c the discontinuities in the phase transition go away, introducing us to the idea of continuous phase transitions. We start by defining the reduced variables

$$\pi = P_r - 1 = \frac{P - P_c}{P_c}, \quad \nu = v_r - 1, \quad \tau = T_r - 1,$$

in terms of which the reduced equation of state becomes

$$3\nu^3 + \pi(1 + \nu)^2(2 + 3\nu) = 8\tau(1 + \nu)^2. \quad (2.12)$$

We'll use this simplified form to study the relationship between our reduced variables and various thermodynamic quantities, and the compare those relationships to experimentally obtained results.

Critical isotherm: Along the critical isotherm, for which $\tau = 0$, we can easily expand Eq. 2.12 for close to the critical point (i.e., for small π and ν), and we find

$$\pi \approx -\frac{3}{2}\nu^3 \Rightarrow (P - P_c) \sim (v - v_c)^3. \quad (2.13)$$

²⁹afterwards, we'll see why this approach has made a subtle but tremendously important mistake!

Volume differences: We next look at how ν depends on τ as the critical point is approached from low temperatures. Rewriting Eq. 2.12 as

$$3\nu^3 + 8(\pi - \tau)\nu^2 + (7\pi - 16\tau)\nu + 2(\pi - 4\tau) = 0,$$

we next recognize that near the critical point the coexistence curve is symmetric. This means that near the critical point, when we consider the polynomial in ν , we should have two roots that are approximately equal in magnitude (with opposite signs), and a third root which is very close to zero. Comparing this consideration to the above equation means that

$$\pi - 4\tau \approx 0 \Rightarrow \pi \approx 4\tau, \quad (2.14)$$

and so for the other possible values of ν we need to solve $\nu^2 + 8\tau\nu + 4\tau = 0$. Doing so gives

$$\lim_{T \rightarrow T_c^-} \nu \approx \pm 2|\tau|^{1/2} \Rightarrow \lim_{T \rightarrow T_c^-} (v_{gas} - v_{liquid}) \sim (T_c - T)^{1/2}. \quad (2.15)$$

Critical isochore: Finally, for this comparison, we look at the isothermal compressibility, which is essentially determined by

$$\kappa_\tau \sim - \left. \frac{\partial \nu}{\partial \pi} \right|_\tau \approx \frac{2}{7\pi + 9\nu^2 - 16\tau}.$$

If we approach the critical point along the critical isochore ($\nu = 0$) from the high-temperature side, we get (again, using $\pi \approx 4\tau$)

$$\lim_{T \rightarrow T_c^+} - \left. \frac{\partial \nu}{\partial \pi} \right|_\tau \approx \frac{1}{6\tau} \Rightarrow \lim_{T \rightarrow T_c^+} \kappa_T(v_c) \sim (T - T_c)^{-1} \quad (2.16)$$

We did these calculations in the context of the van der Waals equation of state, but actually all we have really assumed is that our system is (a) mechanically stable and (b) analytic, in that close to the critical point we could expand the pressure like

$$P(T, v) = P_c + \alpha(T - T_c) - a(T - T_c)(v - v_c) + \frac{b}{2}(T - T_c)(v - v_c)^2 - \frac{c}{6}(v - v_c)^3 + \dots,$$

where mechanical stability tells us that the coefficient $a > 0$ above T_c and $c > 0$ at T_c . What we arrive at is a prediction of *universal singular behavior* near the critical point, with various exponents characterizing the strength of the singularities.

Writing the above results using conventional names for the exponents, we can write

$$\begin{aligned} \lim_{T \rightarrow T_c^+} \kappa_T(v_c) &\sim (T - T_c)^{-\gamma} \\ \lim_{T \rightarrow T_c^-} (v_{gas} - v_{liquid}) &\sim (T_c - T)^\beta \\ (P - P_c) &\sim (v - v_c)^\delta. \end{aligned} \quad (2.17)$$

In our *mean field* calculation we had

$$\gamma = 1, \quad \delta = 3, \quad \beta = 1/2.$$

Experimentally, it is found that there *are* singularities in all of these quantities near the critical point, but the exponents are

$$\gamma \approx 1.24, \quad \delta \approx 4.79, \quad \beta \approx 0.33$$

On the one hand, that's not too bad! The δ exponent doesn't look great, but β and γ aren't so far off. On the other hand, usually in physics we are willing to make “spherical cow” approximations as we build our theories – i.e. capturing the basic, often qualitative behavior of some phenomenon of interest – as long as we are convinced that systematically improving our approximations will lead to systematically closer quantitative agreement with reality. The (perhaps surprising!) fact is that until the development of the theory of modern critical phenomena it was not possible *even in principle* to account for the difference between, say, $\gamma = 1$ in the theory and $\gamma \approx 1.2$ in real life. After all, as emphasized a few paragraphs above, all we have really done to come up with our theoretical critical exponents was assume stability and analyticity near the critical point! The physical systems are indeed stable, so apparently we made some subtly but mightily wrong assumption when we treated the equation of state as analytic near the critical point.

2.4 Another mean field theory, more critical exponents

2.4.1 Mean-field Ising model

So far we have been focusing our set of examples heavily on fluids (ideal gases, dilute gases, van der Waals fluids below T_c ...). Let's take what will feel like a detour and briefly talk about the the Ising model, which is typically introduced as a toy model for magnetic systems.

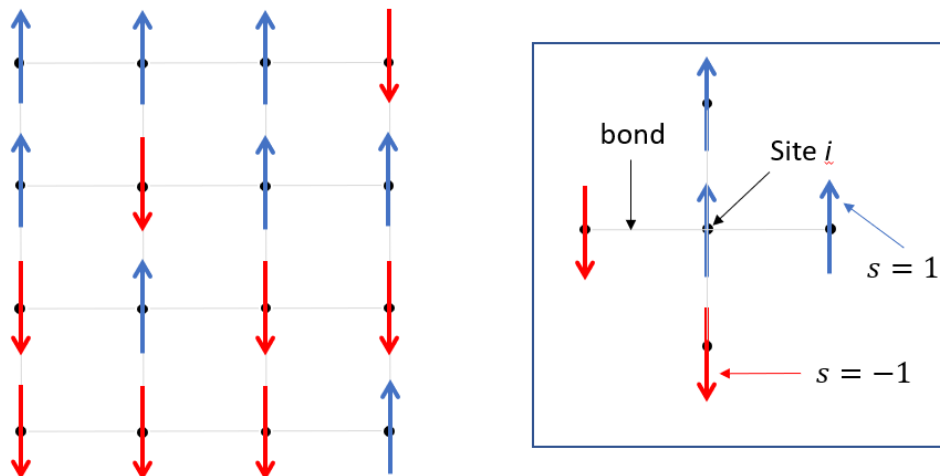


Figure 2.5: **Schematic representation of the Ising model** (Left) a representation of the square lattice Ising model. (Right) On each site “spins” are either up or down, corresponding to $s_i = \pm 1$. Sites with a bond between them are included in the term which sums over neighboring sites in the Hamiltonian.

Schematically depicted in Fig. 2.5, the Ising model consists of a set of N spins, $\{s\}$, each of which can take the values $s_i = \pm 1$, arranged on a lattice. These spins interact with their neighbors, and may also be coupled to an external magnetic field. The Hamiltonian governing the spins is

$$\mathcal{H} = -J \sum_{\langle ij \rangle} s_i s_j - B \sum_i s_i, \quad (2.18)$$

where J determines the strength of the spin-spin interaction, B is the external field, and $\sum_{\langle ij \rangle}$ indicates a sum over all spins i and j that are neighbors of each other. This model has many variations – Is J the same for all pairs of spins? What lattice do the spins live on? When do we count two spins as being neighbors of each other? – but we will focus on the simplest case of nearest-neighbor spins with constant J on a hyper-cubic lattice. We will be interested in thinking about the magnetization per spin,

$$m = \frac{1}{N} \langle s_i \rangle, \quad (2.19)$$

as a function of model parameters and temperature. Although seemingly simple, this model turns out to contain deep insights into the way we construct physical theories and choose the level at which we want to describe a system of interest. A mean field approximation³⁰ for this model can be quickly written down – we will derive it in less than a page in the upcoming chapter on classical statistical mechanics – yielding a self-consistent expression for the relationship between the magnetization per spin and the external field:

$$m = \tanh(\beta B + \beta J q m), \quad (2.20)$$

where $\beta = (k_B T)^{-1}$ and q is the number of neighbors each spin has (e.g., if the lattice is a hypercubic lattice in spatial dimension d , then $q = 2d$).

2.4.2 Critical point behavior

First, in the absence of an external field, our mean field equation is just

$$m = \tanh(\beta J q m).$$

Taylor expanding, $\tanh x \approx x - x^3/3 + \dots$, so the the slope of m near the origin is $\beta J q$. Illustrated in Fig. 2.6 by plotting the left- and right-hand sides of the above equation simultaneously, if $\beta J q < 1$, the only solution to the mean field equation is $m = 0$; if, however, $\beta J q > 1$ there are three solutions: $m = 0$ or $m = \pm m_0$ (and the $m = 0$ solution turns out to be unstable, just like the putative middle solution for the volume at a particular pressure when $T < T_c$ in the van der Waals equation).

Given this zero-field critical temperature, $k_B T_c = Jq$, let's define the reduced *inverse* temperature $\tau = T_c/T$, and also apply the hyperbolic trig identity

$$\tanh(a + b) = \frac{\tanh a + \tanh b}{1 + \tanh a \tanh b}$$

³⁰Just a comment: mean field theories are not unique descriptions of a system. There are often many ways to generate a mean field theory and the results need not be identical. In general, all mean field theories for the same system will share the same scaling near a critical point, but will typically differ in their calculations of non-universal features.

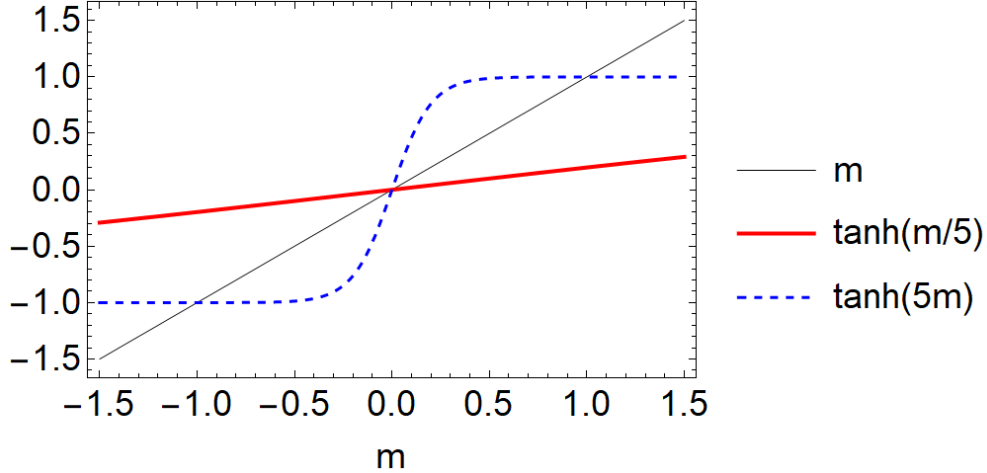


Figure 2.6: **Solution to the self-consistent Ising model magnetization at zero field** For $\beta Jq < 1$ (red solid curve) the only intersection between the tanh and linear functions is the solution $m = 0$. For $\beta Jq > 1$ (blue dashed curve) solutions at $\pm m_0$ exist.

to Eq. 2.20 to obtain

$$\tanh(\beta B) = \frac{m - \tanh(m\tau)}{1 - m \tanh(m\tau)}. \quad (2.21)$$

Close to the critical point, where we expect both the field and the magnetization to be small, we can expand this to obtain

$$\beta B \approx m(1 - \tau) + m^3 \left(\tau - \tau^2 + \frac{\tau^3}{3} \right) + \mathcal{O}(m^5) \quad (2.22)$$

From this, we can read off some critical exponents. In the absence of an external field we can approach T_c from below, and we see

$$m^2 \sim \frac{T_c - T}{T_c} + \dots \Rightarrow m \sim \pm (T_c - T)^{1/2}. \quad (2.23)$$

Conventionally, this critical exponent is called β , and here $\beta = 1/2$.

We can also look at how the external field and the magnetization are related along the critical isotherm (here, denoted $\tau = 1$). We immediately see

$$B \sim m^3; \quad (2.24)$$

this critical exponent is conventionally called δ , and here $\delta = 3$.

We can also easily look at the isothermal magnetic susceptibility, $\chi_T = \frac{\partial m}{\partial B} \Big|_T$, and study how this changes as we vary T . Differentiating our series expansion of the equation of state with respect to B gives

$$\frac{1}{k_B T} = \chi_T(1 - \tau) + 3m^2 \chi_T \left(\tau - \tau^2 + \frac{\tau^3}{3} \right). \quad (2.25)$$

For $T > T_c$ the only solution is $m = 0$, so immediately

$$\chi_T = \frac{1}{k_B (T - T_c)}. \quad (2.26)$$

For $T < T_c$, we substitute in our result that $m \sim (T_c - T)^{1/2}$, and similarly find $\chi_T \sim |T - T_c|^{-1}$. This critical exponent is conventionally called $-\gamma$, and here $\gamma = 1$.

Comparison with exact results

You'll notice that the critical exponents we've just computed do not depend on the dimension, d , of the lattice. In $d = 1$ the mean field theory is disastrously wrong, as in 1D it turns out that there isn't even a phase transition! In higher dimensions the qualitative features of our calculation are correct (there is a phase transition; there are power-law divergences of the quantities we've studied; etc.). In $d \geq 4$, the mean field calculation turns out to give the *correct* critical exponents, so that's neat!

What about $d = 2$ and $d = 3$? Here's a table³¹ (in $d = 2$ there is an exact solution for the Ising model; in $d = 3$ they have been determined via a great deal of numerical effort):

	Mean field	Exact result ($d = 2$)	Numerical result ($d = 3$), and experiments
β	1/2	1/8	≈ 0.33
δ	3	15	≈ 4.79
γ	1	7/4	≈ 1.24

What is going on, here?! The mean field results look the same as the mean field theory for the van der Waals equation, but perhaps you suspect that's just because I've shuffled variable names around to make things look good. But by comparing the $d = 3$ results with the experimental results for real liquid-gas critical behavior you should be convinced that we really are talking about the same critical exponents. Apparently, then, mean field models of ferromagnetism and mean field models of fluids give the same critical point behavior, *and get the answer wrong in the same non-obvious way!*

This is evidence for *universality at the critical point*, where apparently there is a single theory which describes the essential physics at the critical point for magnets, for the liquid-gas transition, for the 3D Ising model, and many other seemingly unrelated systems. We're living the dream! In physics we're always trying to strip away as much of the unnecessary detail from a system as we can, and apparently near a critical point "Nature" does all the hard work for us!

As an aside, the pattern we saw above as we varied the dimension of our model under study is pretty generic: when you write down a mean field theory there is a dimension at or below which which the theory fails completely (called the *lower critical dimension*, d_l), a dimension at or above which the theory gives the right answers (called the *upper critical dimension*, d_c). For $d_l < d < d_c$ mean field theory often returns crudely correct phase diagrams that are wrong or, worse, misleading near critical points. Sadly for mean field theory, life tends to happen in between d_l and d_c .

³¹There are other critical exponents, which we'll meet in the next chapter, that fit the pattern. There's α (for the divergence of the heat capacity), there's ν (for how the correlation length depends on temperature close to the critical point), and there's η (which describes the long-range behavior of the two-point correlation function *precisely* at the critical point).

2.5 Problems

2.5.1 The Dieterici equations of state

The Dieterici equation is a modified version of the van der Waals equation which accounts for the fact that interactions can create gradients in pressure at the boundary of a gas. The equation of state is

$$\beta P(v - b) = \exp\left(-\beta \frac{a}{v}\right),$$

where as in the notes $v = V/N$.

- (A) Find the location of the critical point, $\{P_c, T_c, v_c\}$, and compute the ratio $\frac{P_c v_c}{k_B T_c}$
- (B) Find the critical exponent that characterizes the pressure in terms of $v - v_c$ on the critical isotherm.
- (C) Find the critical exponent that characterizes the isothermal compressibility, κ_T as a function of $T - T_c$ on the critical isochore.

2.5.2 Critical point behavior of a gas

In class we showed three power-law relationships between quantities close to the critical point predicted by the van der Waals equation of state for fluids and the Curie-Weiss equation of state for a ferromagnet. The similarity of the answers should make you suspicious. In fact, our results really only depended on the pressure being an analytic function in the volume and temperature. Let's see this in action. Let the density $\rho = N/V$, and suppose that we can write the pressure of the fluid as

$$P(\rho, T) = k_B T \rho - b \rho^2 + c \rho^3$$

where b and c are positive constants.

(A) Stability argument: Thermodynamic stability demands that at the critical point $\frac{\partial P}{\partial V} \Big|_T = \frac{\partial^2 P}{\partial V^2} \Big|_T = 0$. What does this imply about derivatives of the pressure with respect to the density?

(B) Critical point: Find the critical temperature, T_c , below which this formula cannot be the valid equation of state, as well as the corresponding values of P_c and ρ_c . Calculate the dimensionless ratio

$$\frac{P_c}{k_B T_c \rho_c}$$

(C) Scaling near the critical point: (i) At $T = T_c$, how does $(P - P_c)$ depend on $(\rho - \rho_c)$?

(ii) What is the isothermal compressibility,

$$\kappa_T = -\frac{1}{V} \frac{\partial V}{\partial P} \Big|_T,$$

as a function of T on the critical isochore (i.e., for $\rho = \rho_c$).

(D) More terms in the expansion: Suppose in our expression above we had included another power of density:

$$P(\rho, T) = k_B T \rho - b\rho^2 + c\rho^3 + d\rho^4.$$

Is the dimensionless ratio characterizing the location of the critical point, $\frac{P_c}{k_B T_c \rho_c}$, still independent of the constants b , c , and d ? If no, what kinds of interesting asymptotic limits does the ratio have when b , c , and d are large or small relative to each other?

2.5.3 Non-analytic equations of state

We've looked at the critical exponents associated with various analytic, mean-field equations of state. Suppose instead you managed to derive an equation of state for a system near the critical point that looked like

$$h \approx am(t + bm^2)^\kappa,$$

where h is the external field, m is the order parameter, a, b are positive constants, and $1 < \kappa < 2$. What are the critical exponents β and δ of this model? What about the exponent γ , evaluated for both $t > 0$ and $t < 0$? Do the critical exponents that you found satisfy the Widom relation we expect (i.e., is $\gamma = \beta(\delta - 1)$)?

Chapter 3

Landau theory, universality, and scaling near a critical point

We just saw some universal behavior. [Figure](#) / [example](#) / [more examples go here!](#)

3.1 Landau's phenomenological theory

If seemingly very different mean field theories – for very different physical systems! – give rise to the same essential behavior near their respective critical points, your instinct is that there ought to be a unified way of looking at phase transitions that reveals *why* this should be the case. Landau's phenomenological theory³² of phase transitions serves exactly this role, focusing on understanding the universal behavior of physical systems based on two general considerations: analyticity and symmetry. The Landau approach to phase transitions is typically only qualitatively correct (as we are about to see, it gives exactly the same critical exponents as mean field theory), but it (a) let's us understand universality and (b) is extremely straightforward, letting you compute mean field critical exponents for systems in different universality classes with typically great rapidity.

Landau theory postulates the existence of an object, \mathcal{L} , called the Landau free energy density (it is not a thermodynamic free energy density, and it need not be convex). The theory starts by identifying an *order parameter* of the system, η , which is a quantity that is typically zero in a high-temperature or disordered phase, and non-zero in an ordered phase: For the liquid-gas transition one can take the difference in densities between the phases ($\eta = v_{gas} - v_{liquid}$), for the Ising model magnet we can take the magnetization ($\eta = m$), for a superconductor it is related to off-diagonal long-range order in the one-particle density matrix, etc. Order parameters for a particular system need not be unique, and depending on the nature of the problem they can be scalars, vectors, etc., although here we will assume it is a scalar.

Once we have identified an order parameter, we want to construct an \mathcal{L} that will act like a free energy density (and has dimensions of energy per volume), insofar as we will compute

³²To be distinguished from a version of Landau theory which is motivated by systematic calculation starting with a microscopic Hamiltonian. This more complicated version is typically not more insightful.

thermodynamic quantities by taking appropriate derivatives of \mathcal{L} . We construct \mathcal{L} via the following constraints:

1. The state of the system is specified by the *global minimum* of \mathcal{L} with respect to η .
2. \mathcal{L} *must be* consistent with the symmetries of the system.
3. Near the critical point, \mathcal{L} is an analytic function of η and any coupling constants, which I'll denote as $\{K\}$ here. Thus, for instance, for a spatially uniform system we can write

$$\mathcal{L} = \sum_{n=0} a_n(\{K\})\eta^n \quad (3.1)$$

4. In the disordered phase the order parameter should be $\eta = 0$, while it should be small and non-zero in the ordered phase.

The Landau free energy density for a magnetic system

Let's see how these constraints let us build \mathcal{L} for a particular example: we'll choose the kind of Ising model we just discussed in the last section. By the third and the fourth constraint, near T_c we can expand \mathcal{L} as a Taylor series, and since we expect η to be small, we don't need to go to very high order. We'll write

$$\mathcal{L} = \sum_{n=0}^4 a_n \eta^n, \quad (3.2)$$

where the coefficients a_n could depend on the Ising model coupling term J , the external field B , the temperature T , or whatever else happens to be in our model.

Additionally, from the first constraint we have that \mathcal{L} is extremized by solving

$$\frac{\partial \mathcal{L}}{\partial \eta} = a_1 + 2a_2\eta + 3a_3\eta^2 + 4a_4\eta^3 = 0. \quad (3.3)$$

Since we want $\eta = 0$ to be the solution for $T > T_c$, we have $a_1 = 0$.

Finally, what about the symmetry constraint? Consider our Ising model in the absence of any external field, $B = 0$: certainly the Hamiltonian here is invariant under the flipping of every spin³³, and we expect that the probability of finding the system with a particular value of the magnetization has the property $P(\eta) = P(-\eta)$. We expect, since we want \mathcal{L} to behave like a free energy, that $P \sim \exp(-\beta\mathcal{L})$, so *we require* for this model that $\mathcal{L}(\eta) = \mathcal{L}(-\eta)$. The fact that \mathcal{L} is even implies that in our Taylor series $a_i = 0$ for every odd value of i . Thus, we have

$$\mathcal{L} = a_0 + a_2\eta^2 + a_4\eta^4 + \mathcal{O}(\eta^6), \quad (3.4)$$

with the additional constraint that $a_4 > 0$ (if it is not, then \mathcal{L} above would be minimized by $\eta \rightarrow \infty$, and we want η to be finite and small... if a_4 is negative for a particular system, one needs to include higher order terms to stabilize the system).

³³Including the field, the Ising model has \mathbb{Z}_2 symmetry associated with $\mathcal{H}(B, J, \{s_i\}) = \mathcal{H}(-B, J, -\{s_i\})$

What remains is to ask, for the Ising ferromagnet, about the temperature dependence of the coefficients a_i . First³⁴, a_0 is the value of \mathcal{L} in the high temperature phase, and in general we expect this coefficient to vary smoothly (i.e., without divergence) through the critical temperature. It represents, in a sense, degrees of freedom which are not described by (and are certainly not coupled to) the order parameter; it may be important for some detailed calculations, but we will typically set it to zero.

Next, we expand the fourth-order coefficient as

$$a_4 = a_4^{(0)} + \frac{T - T_c}{T_c} a_4^{(1)} + \dots, \quad (3.5)$$

where the notation indicates the part of the coefficient associated with a particular order of this series expansion. As it turns out, the temperature dependence in a_4 does not control the overall behavior of the system near T_c , so we will just assume that a_4 is some positive constant.

That leaves only a_2 , which we similarly expand:

$$a_2 = a_2^{(0)} + \frac{T - T_c}{T_c} a_2^{(1)} + \dots.$$

Once again, though, we want to find $\eta = 0$ for $T > T_c$ and $\eta \neq 0$ for $T < T_c$. Comparing with Eq. 3.3, whose solution (for $a_1 = a_3 = 0$) is

$$\eta = 0 \quad \text{or} \quad \eta = \pm \sqrt{\frac{-a_2}{2a_4}}, \quad (3.6)$$

we see that we want to set $a_2^{(0)} = 0$ and $a_2^{(1)}$ to some positive constant to ensure that the order parameter is non-zero below T_c . Thus

$$a_2 = \frac{T - T_c}{T_c} a_2^{(1)} + \dots, \quad (3.7)$$

and as with a_4 , it is this lowest order term which dominates behavior near the critical point.

Finally, we now reintroduce the possibility of an external field. This breaks the even symmetry of the system, and we know from the Ising model Hamiltonian that it adds an $a_1 = -B$ term. Dropping some of the cumbersome notation and introducing a and b as phenomenological constants, we have our final expression for the Landau free energy density for the Ising model universality class in the absence of spatial variations:

$$\mathcal{L} = -B\eta + a t \eta^2 + \frac{b}{2} \eta^4, \quad \text{where } t = \frac{T - T_c}{T_c} \quad (3.8)$$

In principle we are allowed by symmetry to now also add an a_3 term; a calculation reveals that it is not a leading term near the critical point, so we've neglected it for now. In general, \mathcal{L} is constructed by writing down *all possible scalar terms which are powers of the order parameter* (or the order parameter components, if the order parameter itself is more complicated than a scalar) *that are consistent with the symmetry of the system.*

³⁴Zeroth?

Critical exponents

For the above \mathcal{L} , we now compute a few critical exponents. The β exponent is the one that characterizes the divergence of m with t below T_c . From the above, we already know that

$$\eta = \sqrt{\frac{-at}{2b}}, \quad (3.9)$$

so we read off $\beta = 1/2$.

We next differentiate \mathcal{L} with respect to η to find the magnetic equation of state:

$$B = 2at\eta + 2b\eta^3. \quad (3.10)$$

On the critical isotherm ($t = 0$) we immediately see $B \sim \eta^3$, and this is our $\delta = 3$ exponent.

The isothermal magnetic susceptibility is

$$\chi_T = \left. \frac{\partial \eta(B)}{\partial B} \right|_T = \frac{1}{2(at + 3b\eta(B)^2)}, \quad (3.11)$$

where $\eta(B)$ is the value of the order parameter in the presence of the external field, i.e., the solution to Eq. 3.10. The exponent γ characterizes the divergence of the susceptibility at zero field. For $t > 0$ we know $\eta = 0$ so $\chi_T = (2at)^{-1}$. For $t < 0$ we know $\eta = (-at/b)^{1/2}$ so $\chi_T = (-4at)^{-1}$. In either case, $\gamma = 1$.

Ignoring fluctuations... not variations

We very briefly note that Landau theory neglects the (important!) effects of fluctuations, but it does not only apply to homogeneous systems. It is straightforward to generalize the ideas above to the case where the order parameter can be a spatially varying one: $\eta = \eta(\mathbf{r})$, and if this were a class on statistical field theory we would spend a lot of time thinking about these cases (this would also let us write down critical exponents related to the divergence of *correlation lengths* in the system). When we do treat spatially inhomogeneous systems, we need to add the constraint that

5. \mathcal{L} should be a *local* function, depending only on a finite number of spatial gradients of the order parameter.

As with the earlier construction, the gradient terms we write down must be consistent with the symmetry of the system; for the Ising model where we keep only even terms, the lowest order term is the square of the gradient, which would give the Landau free energy of

$$L = \int d\mathbf{r} [\mathcal{L}(\eta(\mathbf{r})) + \zeta (\nabla\eta(\mathbf{r}))^2], \quad (3.12)$$

where ζ is some new positive constant and \mathcal{L} is the homogenous \mathcal{L} of the preceding notes. In general for this symmetry there are also terms like $(\nabla\eta)^4$ and $(\nabla^2\eta)^2$, etc., and here we've just written the lowest order gradient term³⁵.

³⁵I hear you asking, "Why isn't there a term like $\eta\nabla^2\eta$, which is of the same order as the term we did use, is perfectly isotropic, and has the right $m \rightarrow -m$ symmetry?" The above form is customary because we

3.2 Correlations and fluctuations

Having introduced the idea that the order parameter might have spatial variations, we now introduce a dimensionless two-point correlation function:

$$G(\mathbf{r} - \mathbf{r}') = \frac{1}{\bar{\eta}^2} [\langle \eta(\mathbf{r})\eta(\mathbf{r}') \rangle - \bar{\eta}^2] = \frac{1}{\bar{\eta}^2} \langle (\eta(\mathbf{r}) - \bar{\eta})(\eta(\mathbf{r}') - \bar{\eta}) \rangle. \quad (3.13)$$

This quantity measures correlations between the fluctuations of the order parameter at different distances³⁶, and it is deeply linked to some of the thermodynamic properties of the system.

For instance, suppose as the order parameter we chose the spatially varying density field, $\rho(\mathbf{r})$. The total number of particles is $N = \int d^d r \rho(\mathbf{r})$, and the integral over G is

$$\int d^d r d^d r' G(\mathbf{r} - \mathbf{r}') = \frac{1}{\rho^2} \int d^d r d^d r' [\langle \rho(\mathbf{r})\rho(\mathbf{r}') \rangle - \rho^2] = \frac{1}{\rho^2} [\langle N^2 \rangle - \langle N \rangle^2]. \quad (3.14)$$

Translational symmetry gives us one of the integrals over G for free, and we recognize a connection between the number fluctuations and the isothermal compressibility of the system. Combing everything, then, gives us

$$\int d^d r G(\mathbf{r}) = k_B T \kappa_T. \quad (3.15)$$

This is an example of a *fluctuation-susceptibility* relation, and is the equilibrium limit of the more general *fluctuation-dissipation relation*.

3.2.1 Correlation function for a specific model

To be definite, let's calculate the two-point correlation function for a system described by the Ising universality class Landau theory. The spatially varying order parameter will be $\eta(\mathbf{r})$, the external field will be $B(\mathbf{r})$, and we'll write the Landau free energy as

$$L = \int d^d r \left[(at)\eta^2 + \frac{1}{2}b\eta^4 - B\eta + \frac{c}{2}(\nabla\eta)^2 \right], \quad (3.16)$$

where a , b , c are phenomenological parameters and t is the reduced distance to the critical point. If the system was uniform we would relate the typical value of the order parameter to the appropriate derivative of the free energy: $\langle \eta \rangle = -\frac{\partial L}{\partial B}$. In the presence of spatial variations, these partial derivatives get replaced by *functional derivatives*.

know the identity

$$\nabla \cdot (\eta \nabla \eta) = \eta \nabla^2 \eta + (\nabla \eta)^2,$$

which implies

$$\int d\mathbf{r} \eta \nabla^2 \eta = - \int d\mathbf{r} (\nabla \eta)^2 + \int d\mathbf{S} \cdot \eta \nabla \eta.$$

In the thermodynamic limit we neglect the surface term, so in general we pick either $(\nabla \eta(\mathbf{r}))^2$ or $\eta \nabla^2 \eta$, but not both.

³⁶We have assumed translational symmetry in writing $G(\mathbf{r} - \mathbf{r}')$; more generally we would have $G(\mathbf{r}, \mathbf{r}')$

Functional differentiation

Suppose $F[\eta(\mathbf{r})]$ is a functional³⁷ of $\eta(\mathbf{r})$. The functional derivative of F with respect to the function η is defined as

$$\frac{\delta F}{\delta \eta(\mathbf{r}')} = \lim_{\epsilon \rightarrow 0} \frac{F[\eta(\mathbf{r}) + \epsilon \delta(\mathbf{r} - \mathbf{r}')] - F[\eta(\mathbf{r})]}{\epsilon}. \quad (3.17)$$

This generalizes the definition of the usual derivative, and the operation satisfies properties like

$$\begin{aligned} \frac{\delta}{\delta \eta(\mathbf{r})} \int d^d r' \eta(\mathbf{r}') &= 1 \\ \frac{\delta}{\delta \eta(\mathbf{r})} \eta(\mathbf{r}') &= \delta(\mathbf{r} - \mathbf{r}') \\ \frac{\delta}{\delta \eta(\mathbf{r})} \int d^d r' \frac{1}{2} (\nabla \eta(\mathbf{r}'))^2 &= -\nabla^2 \eta(\mathbf{r}). \end{aligned} \quad (3.18)$$

That last expression, which you probably can see that we're about to use, involves an integration by parts that neglects the surface term.

Linear response

With those definitions, we compute the expectation value of the order parameter via functional differentiation:

$$\langle \eta(\mathbf{r}) \rangle = -\frac{\delta L}{\delta B(\mathbf{r})}. \quad (3.19)$$

This implies that a small change in the external field would cause a small change in the Landau energy

$$\delta L = - \int d^d r' \langle \eta(\mathbf{r}') \rangle \delta B(\mathbf{r}'). \quad (3.20)$$

Since the susceptibility is a measure of how the order parameter changes when the field changes, we have

$$\chi_T(\mathbf{r}, \mathbf{r}') = -\frac{\delta}{\delta B(\mathbf{r}')} \left(\frac{\delta L}{\delta B(\mathbf{r})} \right). \quad (3.21)$$

Treating L as if it is related to a partition function, $L = -k_B T \log Z$ (for the purposes of remembering which thermodynamic derivatives we want to take), we re-write this as

$$\begin{aligned} \chi_T(\mathbf{r}, \mathbf{r}') &= k_B T \left[\frac{1}{Z} \frac{\delta^2 Z}{\delta B(\mathbf{r}) \delta B(\mathbf{r}')} - \frac{1}{Z} \frac{\delta Z}{\delta B(\mathbf{r})} \frac{1}{Z} \frac{\delta Z}{\delta B(\mathbf{r}')} \right] \\ &= \beta [\langle \eta(\mathbf{r}) \eta(\mathbf{r}') \rangle - \langle \eta(\mathbf{r}) \rangle \langle \eta(\mathbf{r}') \rangle] \\ &= \beta G(\mathbf{r} - \mathbf{r}'), \end{aligned} \quad (3.22)$$

³⁷Just as a function is a map – accepting, say, a set of numbers and returning an output number – a functional is a map that accepts a function of a set of numbers and returns an output number

again connecting response functions with correlation functions. Note that, using tildes to denote Fourier transforms, one connects isothermal susceptibility with the wave-vector dependent susceptibility $\tilde{\chi}(\mathbf{k}) = \beta \tilde{G}(\mathbf{k})$ as

$$\chi_T \equiv \lim_{\mathbf{k} \rightarrow 0} \tilde{\chi}(\mathbf{k}) = \beta \tilde{G}(\mathbf{k}) \Big|_{\mathbf{k}=0} = \beta \int d^d r G(\mathbf{r}), \quad (3.23)$$

as in Eq. 3.15.

The correlation function

Enough dithering, let's calculate the two-point correlation function from Eq. 3.16. To do this, we extremize the Landau free energy to find the spatially varying field that the system will adopt in equilibrium. I.e., we set $\frac{\delta L}{\delta \eta(\mathbf{r})} = 0$. This tells us that $\eta(\mathbf{r})$ must satisfy

$$2at\eta(\mathbf{r}) + 2b\eta^3(\mathbf{r}) - B(\mathbf{r}) - c\nabla^2\eta(\mathbf{r}) = 0. \quad (3.24)$$

We then (functionally) differentiate this expression with respect to the spatially varying field:

$$\begin{aligned} \frac{\delta}{\delta B(\mathbf{r}')} [2at\eta(\mathbf{r}) + 2b\eta^3(\mathbf{r}) - B(\mathbf{r}) - c\nabla^2\eta(\mathbf{r})] &= 0 \\ \Rightarrow \beta [-c\nabla^2 + 2at + 6b\eta^2(\mathbf{r})] G(\mathbf{r} - \mathbf{r}') &= \delta(\mathbf{r} - \mathbf{r}'). \end{aligned} \quad (3.25)$$

Well, would you look at that! The correlation function is a Green's function; seems like a retroactively good reason to have used G for it...

We can simplify the above expression by noting that for translationally invariant systems the order parameter η we want is just the equilibrium value from the homogenous Landau theory calculation we did earlier: $\eta = 0$ for $t > 0$ and $\eta = \pm\sqrt{-at/b}$ for $t < 0$. We now introduce the *correlation length*, ξ , noting that $G(\mathbf{r} - \mathbf{r}')$ satisfies

$$(-\nabla^2 + \xi^{-2}(t)) G(\mathbf{r} - \mathbf{r}') = \frac{k_B T}{c} \delta(\mathbf{r} - \mathbf{r}'), \quad (3.26)$$

where

$$\xi(t) = \begin{cases} \sqrt{\frac{c}{2at}} & \text{for } t > 0 \\ \sqrt{\frac{c}{-4at}} & \text{for } t < 0 \end{cases} \sim |t|^{-1/2}. \quad (3.27)$$

Great, given that $G(\mathbf{r} - \mathbf{r}')$ satisfies such an equation, what does it look like? The Fourier transform of Eq. 3.26 gives

$$\tilde{G}(\mathbf{k}) = \frac{k_B T}{c} \frac{1}{k^2 + \xi^{-2}}. \quad (3.28)$$

Evaluated at $\mathbf{k} = 0$, this gives us the isothermal susceptibility (a *measurable* quantity for the system) in terms of the microscopic parameter c and the correlation length:

$$\chi_T = \beta \tilde{G}(0) = \frac{\xi^2}{c},$$

with which we can write the two-point correlation function as

$$\tilde{G}(\mathbf{k}) = \frac{k_B T \chi_T(t)}{1 + k^2 \xi^2}. \quad (3.29)$$

In real space one can take the inverse Fourier transform (i.e. “look it up in a table”), or one can solve the real-space differential equation in polar coordinates. I’ll spare you the details for now. Using the correlation length as our unit of length and defining $\rho = r/\xi$, the result is:

$$\frac{c}{k_B T \xi^{d-2}} G(\rho) = \begin{cases} e^{-r/\xi} & \text{for } d = 1 \\ \frac{K_{\frac{d-2}{2}}(r/\xi)}{(2\pi)^{d/2} (r/\xi)^{(d-2)/2}} & \text{for } d \geq 2 \end{cases}, \quad (3.30)$$

where the K_n are modified spherical Bessel functions of the third kind. We really are most interested in the short- and long-range behavior of these functions, which are

$$\begin{aligned} K_n(x) &\sim \left(\frac{\pi}{2x}\right)^{1/2} e^{-x}, & \text{for } x \rightarrow \infty \\ K_n(x) &\sim \frac{\Gamma(n)}{2} \left(\frac{x}{2}\right)^{-n}, & \text{for } x \rightarrow 0 \\ K_0(x) &\sim -\log x, & \text{for } x \rightarrow 0. \end{aligned} \quad (3.31)$$

Combining everything, very close to T_c the correlation length has diverged, so we just need the $r \ll \xi$ limit of the K_n . For $d = 2$ we use the special log form appropriate to K_0 , and for $d > 2$ all of the powers of ξ conveniently cancel out, giving us

$$G(r) \sim \begin{cases} \log(\xi/r) & \text{for } d = 2 \\ r^{2-d} & \text{for } d > 2 \end{cases}. \quad (3.32)$$

Far from the critical point, though, ξ is small; taking the $r \gg \xi$ limit of the K_n gives (for $d \geq 2$)

$$G(r) \sim \frac{k_B T \exp(-r/\xi)}{c} \frac{1}{r^{(d-1)/2} \xi^{(d-3)/2}}, \quad (3.33)$$

where I’ve dropped prefactors related to 2’s and π ’s for simplicity.

Comments on the correlation length

From the preceding analysis, and from the definition of $G(\mathbf{r})$, we see that ξ is a measure of the spatial extent over which correlations extend. In mean field models of the Ising type we predicted $\xi \sim |t|^{-1/2}$; far from the critical temperature the correlation length will be on the same scale as something microscopic (distance between spins, or the range of the attractive part of the Lennard-Jones interaction, or...), and any fluctuations in the local order parameter away from the average value will quickly wash out.

As $T \rightarrow T_c$, though, $\xi \rightarrow \infty$, and correlations can extend over the entire system. Even though the actual interactions in the system are short-ranged, *long-range order* can propagate. Since the correlations are essentially macroscopic, the microscopic differences that distinguish one system from another likely are irrelevant at such macroscopic scales; this is an important clue for building a general understanding of critical phenomena.

3.3 Critical exponents

We pause to briefly summarize (and comment on) the most important critical exponents that characterize various systems. In all of these cases we are investigating the idea that, close

to the critical point, some thermodynamic quantity has the limiting form of a power law. Letting $t = (T - T_c)/T_c$ and looking a quantity f , when we write

$$f(t) \sim t^\lambda$$

what we *mean* is

$$\lambda = \lim_{t \rightarrow 0} \frac{\log f(t)}{\log t}. \quad (3.34)$$

This is a particularly relevant point when remembering (a) there may be other, non-dominant power law behavior near the exponent; we are capturing just the leading order term, and (b) sometimes we quote an exponent as having the value zero. This can either mean that the thermodynamic quantity has a discontinuity *or* that it has a logarithmic divergence rather than a power-law one. This last possibility comes from using the identity

$$\log t = \lim_{\lambda \rightarrow 0} \left[\frac{1 - e^{-\lambda \log t}}{\lambda} \right] = \lim_{\lambda \rightarrow 0} \left[\frac{1 - t^{-\lambda}}{\lambda} \right]. \quad (3.35)$$

Suppose we are considering a system with Landau free energy L , order parameter m , and ordering field B .

δ : All critical exponents *except* for this one are evaluated at $B = 0$. In the presence of the field, though, δ characterizes the relationship between the field and the order parameter:

$$m \sim B^{1/\delta}. \quad (3.36)$$

α : The divergence of the heat capacity is measured via α . In principle, the divergence could be different on the two sides of the transition, which is commonly denoted by writing α for the divergence above T_c and α' for the divergence below T_c .

$$C = -T \frac{\partial^2 \mathcal{L}}{\partial T^2} \sim |t|^{-\alpha}. \quad (3.37)$$

In the Ising model universality class, mean field predicts $\alpha = 0$ (in the form of a discontinuity); the 2d Ising model has $\alpha = 0$ (in the log divergence form), and the 3d model has $\alpha \approx 0.11$.

β : Below the critical temperature, where the order parameter is non-zero, it diverges like

$$m \sim |t|^\beta. \quad (3.38)$$

γ : The divergence of the low-field susceptibility is measured by γ ; it, too, can in principle have different values above and below the transition:

$$\chi = \left. \frac{\partial m}{\partial B} \right|_{B=0} \sim |t|^{-\gamma}. \quad (3.39)$$

ν : The last two critical exponents on this list are related to the behavior of the two point correlation function. We did a mean-field calculation that suggested the correlation length diverged near $t = 0$ like $\xi \sim |t|^{-1/2}$. There's no reason to expect for real systems that the mean field prediction here is correct, so we introduce ν :

$$\xi \sim |t|^{-\nu}. \quad (3.40)$$

In the Ising model universality class, mean field predicts $\nu = 1/2$, the 2d model has $\nu = 1$ and the 3d model has $\nu = 0.63$.

η : Finally, we want to characterize how the two-point correlation function behaves precisely at the critical point ($t = 0$). Our mean field theory (using the $r \ll \xi$ limit of the K_n) predicted that $G(r) \sim r^{-(d-2)}$, and again, we expect real experiments could differ from this. We introduce the exponent η to measure how wrong mean field is at $t = 0$:

$$G(r) \sim r^{-(d-2+\eta)}. \quad (3.41)$$

In the Ising model universality class, mean field predicts $\eta = 0$, the 2d model has $\eta = 1/4$ and the 3d model has $\eta = 0.032$. Experimentally these small exponents are hard to measure

3.3.1 Dimensional analysis and mean field theory

[Tighten the argument here...](#) [Goldenfeld's treatment](#) To emphasize just how surprising it should be that mean field theory gets the critical exponents wrong, and the surprising sense in which a diverging correlation length *does not* mean that the system has completely forgotten about microscopic length scales, let's write a dimensionless version of the Landau free energy of the Ising type in the absence of an external field:

$$\beta L = \int d^d r \frac{1}{2} (\nabla \phi)^2 + \frac{r_0}{2} \phi^2 + \frac{1}{4} u_0 \phi^4. \quad (3.42)$$

In terms of our earlier expressions, this is just writing

$$\phi = \eta \sqrt{\beta c}, \quad r_0 = \frac{2at}{c}, \quad u_0 = \frac{2b}{\beta c^2}. \quad (3.43)$$

Let's think about the correlation function in terms of basic dimensional analysis. Using the bracket notation to denote the dimension of various quantities, we have $[\beta L] = 1$ (i.e., is dimensionless), so each separate term in the integrand must be dimensionless, too. Thus:

$$\left[\int d^d r (\nabla \phi)^2 \right] = 1 \Rightarrow (x^d)(x^{-2}) [\phi]^2 = 1, \quad (3.44)$$

where x denotes units of length. From this, we immediately get

$$[\phi] = x^{1-d/2}, \quad [r_0] = x^{-2}, \quad [u_0] = x^{d-4}. \quad (3.45)$$

So, what does dimensional analysis say about the correlation length? Clearly $[\xi] = x$, but from the above equation *the only independent quantity* with units of length is r_0 , with dimensions $[r_0] = x^{-2}$. So, apparently, we're done: by dimensional analysis we get that

$$\xi \sim r_0^{-1/2} \sim |t|^{-1/2},$$

where in the last line we remember that $r_0 \propto t$. In some sense, then, the deep mystery is *how could this have been wrong?*

The answer is that *we ignored one other source of an independent length scale* in our dimensional analysis! In particular, there is a *microscopically* small length scale embedded in the problem – say, the spacing between lattice sites in our Ising model – an apparently if mean field is wrong we need to include this length scale.

How does this solve our problem? Let's call this microscopic length λ , with $[\lambda] = x$, of course. By dimensional analysis we can conclude that

$$\xi = r_0^{-1/2} f(r_0 \lambda^2), \tag{3.46}$$

where f is some function we know nothing about (yet). We've written it this way, though, because $r_0 \propto t$ and λ is independent of t , so near the critical point we are interested in $\lim_{z \rightarrow 0} f(z)$. If, *for whatever reason*, it so happens that

$$f(z) \sim z^\theta \tag{3.47}$$

as $z \rightarrow 0$, then as the critical temperature is approached we get

$$\xi \sim t^{-1/2+\theta} \lambda^{2\theta}. \tag{3.48}$$

The exponent θ , characterizing the difference between an observed divergence of the correlation length and the prediction from Landau theory, is called the anomalous dimension.

Notice something remarkable that's happened here: we have been emphasizing that near a critical point the diverging correlation length sweeps away any microscopic details, and we are used to assuming that when trying to explain phenomena at one scale we can disregard phenomena at much shorter scales³⁸. Apparently near the critical point this idea is not quite right: the *very existence* of a microscopic length scale allows for an anomalous dimension / departure from mean field theory. In general, although not a priori guaranteed, the *value* of the anomalous dimension does not depend on the microscopic length scale itself. Thus, near a critical point we expect that the *scaling* of various quantities, the way they diverge near the critical point, to be universal, but there is no reason to expect the prefactors of the scaling relations to share that universal character.

3.4 Scaling hypothesis

We just described a set of critical exponents that characterize the behavior of systems near their critical point; let's stare at a table³⁹ for a moment:

³⁸“Don't model bulldozers with quarks” [39]

³⁹Anything that is not an integer should be assumed to be approximate. Sources are wikipedia (where there are tables of critical exponents for many more models) [40], plus [41] for the Potts model

	Ising (MF)	Ising ($d = 2$)	Ising ($d = 3$)	Percolation ($d = 3$)	4-State Potts ($d = 2$)	Directed percolation (MF)	XY model ($d = 3$)
α	0	0	0.11	-0.625	2/3	-1	-0.015
β	1/2	1/8	0.327	0.418	1/12	1	0.349
δ	3	15	4.79	5.3	15	2	4.78
γ	1	7/4	1.24	1.793	7/6	1	1.32
ν	1/2	1	0.63	0.88	2/3	1/2	0.67
η	0	1/4	0.036	0.046	1/4	0	0.038

In the early 60s, as mean field results were known and as some finite-dimensional estimates were obtained by numerical analysis, people started noticing that these exponents did not seem to be independent. For instance, the exponent values seem to satisfy

$$\alpha + 2\beta + \gamma = 2 \quad (3.49)$$

$$\gamma - \beta(\delta - 1) = 0 \quad (3.50)$$

$$\gamma - \nu(2 - \eta) = 0, \quad (3.51)$$

which are often called the Rushbrooke, Widom, and Fisher identities, respectively. Since then, there have been various proofs that thermodynamics requires (via, say, the convexity of the free energy, or the relative sizes of C_P and C_V certain *inequalities* to be satisfied, for instance

$$\alpha + 2\beta + \gamma \geq 2,$$

but that in the actual data these thermodynamic inequalities are *saturated*.

3.4.1 The static scaling hypothesis

Thermo was presented as “what is a consistent mathematical framework that matches phenomenology?” So: what framework near the CP could give both non-MF exponents and exponent relationships? The static scaling hypothesis⁴⁰ is an attempt to encode multiple features of the behavior of a system near the critical point in a single expression; as we will see, assuming this hypothesis allows one to strengthen the thermodynamic inequalities mentioned above for the relationships between exponents into *equalities* – which is good, because the equalities are satisfied. Let’s focus on a magnetic system, where just to keep you on your toes I’ll write m for the magnetization and h for the external field (in units of $k_B T$); t will still stand for the reduced distance to the critical point.

Static scaling of the magnetization

We want to simultaneously encode two things we have already seen:

$$m(t = 0, h) = \pm C_1 |h|^{1/\delta} \quad \text{and} \quad m(t, h = 0) = \begin{cases} 0 & t > 0 \\ \pm C_2 |t|^\beta & t < 0 \end{cases}. \quad (3.52)$$

⁴⁰Widom, in the modestly titled “Some topics in the theory of fluids” [42]

One of Widom's many insights was that, staying in a regime where $|t| \ll 1$ and $|h| \ll 1$, both of these results can be expressed as

$$m(t, h) = \begin{cases} t^\beta F^+(h/t^\Delta) & t > 0 \\ \pm(-t)^\beta F^-(h/(-t)^\Delta) & t < 0 \end{cases}, \quad (3.53)$$

where we assume that β and the *gap exponent* Δ are universal, as are the *scaling functions* above and below the critical temperature, F^+ and F^- . At first glance we don't know very much about these scaling functions, but the requirement that Eq. 3.53 reproduce the results in Eq. 3.52 will let us both put some constraints on F^+ and F^- and also derive relationships between the critical exponents.

First, we'll show that the gap exponent is actually not a new critical exponent at all! Consider the susceptibility in the low-field limit, $\chi_T \propto \left. \frac{\partial m}{\partial h} \right|_{h=0}$. Differentiating Eq. 3.53 with respect to h gives

$$\chi_T \propto |t|^\beta \left. \frac{\partial F^\pm(h/|t|^\Delta)}{\partial h} \right|_{h=0} \propto \left. \frac{|t|^\beta}{|t|^\Delta} \frac{\partial F^\pm(x)}{\partial x} \right|_{x=0} \sim |t|^{\beta-\Delta} F'^\pm(0). \quad (3.54)$$

So, as long as the derivative of these scaling functions doesn't either diverge or vanish as the argument goes to zero, we can connect the gap exponent to a critical exponent we've already encountered:

$$\beta - \Delta = -\gamma. \quad (3.55)$$

Next, let's see what happens by requiring that Eq. 3.53 reproduces the results in Eq. 3.52. First, we take the *small field* limit of Eq. 3.53 and match it to the zero field result

$$m(t, h) = \begin{cases} t^\beta F^+(0) & t > 0 \\ (-t)^\beta F^-(0) & t < 0 \end{cases} = \begin{cases} 0 & t > 0 \\ \pm C_2 (-t)^\beta & t < 0 \end{cases} \quad (3.56)$$

This already gives us some constraints on the scaling functions, namely

$$F^+(0) = 0 \quad \text{and} \quad F^-(0) = (\text{some finite constant}). \quad (3.57)$$

We also want to reproduce the result for the magnetization along the isotherm in the presence of a field ($m \sim h^{1/\delta}$), so we look at the limit $t \rightarrow 0$ while keeping h small and finite. This amounts to looking at the scaling functions in the limit that their argument diverges, but we also know that in this limit m is well-behaved. How do we reconcile the situation? We assume that both scaling functions adopt a power-law form for large argument:

$$F^\pm(x) \sim x^\lambda. \quad (3.58)$$

Making this assumption, the magnetization as the critical isotherm is approached is

$$m(t \rightarrow 0, h) \sim |t|^\beta \left(\frac{h}{|t|^\Delta} \right)^\lambda \sim |t|^{\beta-\lambda\Delta} h^\lambda. \quad (3.59)$$

On the critical isotherm, though, we need the t -dependence of this expression to cancel out (otherwise we would get the incorrect result that either $m = 0$ or $m = \infty$, depending on the sign of the exponent of $|t|$). We therefore have the simultaneous requirements

$$\beta = \lambda\Delta \quad \text{and} \quad \lambda = 1/\delta. \quad (3.60)$$

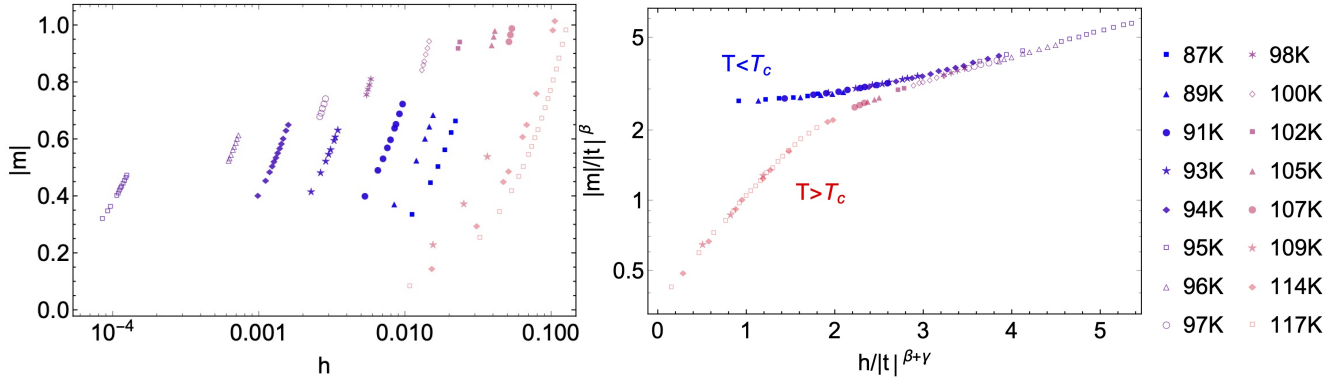


Figure 3.1: **Scaling plot for magnetization vs field near the critical point** for the “anhydrous compound with the formula $\{[\text{Mn}_2^{\text{II}}(\text{pydz})][\text{Nb}^{\text{IV}}(\text{CN})_8]\}_n$,” taken at different temperatures, with estimates of $T_c \approx 95.3$, $\beta \approx 0.41$, and $\gamma \approx 1.39$. Data roughly extracted from Ref. [43]. Units on the plot have been suppressed.

This both tells us that the scaling functions behave like

$$F^\pm(x \rightarrow \infty) \sim x^{1/\delta} \quad (3.61)$$

and

$$\Delta = \beta/\lambda = \beta\delta. \quad (3.62)$$

Combining the above result with Eq. 3.55 gives the Widom identity:

$$\beta\delta = \beta + \gamma. \quad (3.63)$$

The static scaling hypothesis, if true, suggests two powerful things. First, by derivations like the above, it places constraints on the critical exponent that should be measured in experiments or in exact theories (and, thus, also provides targets for approximate theories, if you’re into that kind of thing). Second, the assumption that the scaling functions are also universal *suggests an important way of plotting data!*

For instance: suppose you were to measure the magnetization of a system as a function of temperature and external field and then make a plot of m vs h , you’d get different curves for each value of the field t . The static scaling hypothesis says that *even if you have no idea* what the functions F^\pm are, if you plot $m/|t|^\beta$ vs $|h|/|t|^{\beta+\gamma}$ then you should observe *data collapse*, with all of your data falling onto two branches. To observe this collapse you would need to know the right values for T_c , β , and γ (which you usually don’t know ahead of time). An example of this kind of data collapse is shown in Fig. 3.1, where I pulled data from an arXiv paper studying magnetization near the critical point of... let’s just say “some material.” Manipulating data in order to observe scaling collapse is a way of estimating both the location of the critical point and the values of the critical exponents, although it must be done with care.

Static scaling for the free energy

Above we wrote down a scaling hypothesis for the magnetization; we can do the same thing for the free energy (or, at least, the part of the free energy which is singular at the critical

point), f_s . This would look like

$$f(t, h) = |t|^{2-\alpha} F_s^\pm \left(\frac{h}{|t|^\Delta} \right). \quad (3.64)$$

Defining the magnetization by differentiating this with respect to the field in the limit that h is small gives a relationship between the exponents β , α , and Δ . Defining the isothermal susceptibility by differentiating again gives a relationship between the exponents α , Δ , and γ . Combining these two lets one derive the Rushbrooke scaling law.

Hyperscaling

Before closing, we briefly mention that in addition to scaling laws relating the critical exponents, there are also *hyperscaling laws* which involve both the critical exponents *and* the dimension of space. These can be derived from, e.g., assuming that the correlation function has a scaling form:

$$G(\mathbf{r}, t, h) = \frac{1}{r^{d-2+\eta}} F_G^\pm \left(r|t|^\nu, \frac{h}{|t|^\Delta} \right),$$

where we've added the subscript to F_G since there is no reason that the scaling function for the two-point correlations should be the same as the scaling function for the magnetization. This scaling hypothesis leads to laws like the Josephson relation,

$$2 - \alpha = d\nu.$$

Hyperscaling laws are on a different footing than the scaling laws – they must be, and clearly something must be afoot in mean field models, where the critical exponents stay the same for all $d > d_c$. We also mention that just as there are *static scaling hypotheses* there are *dynamic scaling hypotheses*. These relate the *relaxation time* – the way that fluctuations from the typical order parameter decay away – to the correlation length, and bring with them both new critical exponents and new relationships between those exponents.

Finally, you may be wondering where these scaling laws in fact derive from. The main physical idea is the diverging correlation length, ξ , is responsible for all of the singular behavior near the critical point. We've already indicated that this vague statement cannot be completely true, but the idea is the following: consider just the part of the free energy density which has the singularity, and by dimensional analysis write it as

$$\frac{\mathcal{L}_s}{k_B T} \sim \xi^{-d} \left(c_0 + c_1 \left(\frac{\lambda_1}{\xi} \right)^{\pi_1} + c_2 \left(\frac{\lambda_2}{\xi} \right)^{\pi_2} + \dots \right). \quad (3.65)$$

Here the λ_i are a list of any microscopic length scales in the problem, the c_i are coefficients that depend at most weakly on temperature, and the $\pi_i > 0$.

With those *assumptions*, in the $t \rightarrow 0$ limit the leading behavior is dominated by

$$\mathcal{L}_s \sim \xi^{-d} \sim |t|^{d\nu},$$

and this immediately leads to the Josephson relation:

$$t^{-\alpha} \sim C = -T \frac{\partial^2 \mathcal{L}_s}{\partial T^2} \sim |t|^{d\nu-2} \Rightarrow d\nu = 2 - \alpha. \quad (3.66)$$

This, perhaps, feels a bit unsatisfying: we do not really know a priori that, given a quantity A with $[A] = x^y$ we should always take the dimensionless quantity $A\xi^{-y}$ in our scaling theory – why might it not be the case that we should take $A\xi^{z-y}\lambda^{-z}$, mixing and matching the diverging length scale with one of the microscopic ones? Ultimately, the answers to all of these questions rely on the theory of the renormalization group: RG lets us see where scaling hypotheses come from, it gives criteria for which measurables acquire anomalous dimensions, allows us to *calculate* those anomalous dimensions, and (with a little bit of help) lets us derive the particular forms of scaling functions.

3.5 Problems

3.5.1 Dotted some i's and crossing some t's on Landau expansions

During lectures I will occasionally say that we are dropping terms because they contribute at sub-leading order in the final answer, so for ease of calculating on the blackboard I ignore them. You should not trust me when I say things like this, but rather, you should check for yourself that that's how the calculation works out! In this problem, let's double check some statements from lecture – who knows, maybe we'll learn that I was leading you astray?

(A) Consider a Landau free energy density written in terms of an order parameter η , a temperature $t = (T - T_c)/T_c$, an external field B , and constants a, b, c which is

$$\mathcal{L} = -B\eta + at\eta^2 - cB\eta^3 + \frac{b}{2}\eta^4.$$

- (i) What is the critical exponent δ , which characterizes how B varies with η at $T = T_c$ when both B and η are small? Writing B as a power series in η , what is the next higher order term that comes from including this η^3 term?
- (ii) What is the critical exponent γ , which characterizes

$$\chi_T = \frac{\partial \eta(B)}{\partial B} \sim |t|^{-\gamma}$$

when $B=0$?

(B) Consider a different Landau free energy density in the absence of an external field, $\mathcal{L} = a_2\eta^2 + a_4\eta^4$. In class we wrote, $a_2 = a_2^{(0)} + ta_2^{(1)} + \dots$ and $a_4 = a_4^{(0)} + ta_4^{(1)} + \dots$, we argued that $a_2^{(0)} = 0$, and I claimed that we only needed to keep the $a_2^{(1)}$ and $a_4^{(0)}$ terms to get the right scaling. Keep terms corresponding to both $a_2^{(2)}$ and $a_4^{(1)}$, and then solve for the equilibrium values of the order parameter. What is the critical exponent β , which characterizes $\eta \sim |t|^\beta$ below T_c ? What is the order of the first correction?

3.5.2 Flavors of various transitions in a Landau Theory

Suppose your favorite system is described by the following Landau free energy density:

$$\mathcal{L} = \frac{a}{2}\eta^2 + \frac{b}{4}\eta^4 + \frac{c}{6}\eta^6,$$

where both a and b could be of either sign (i.e., they might depend on whether you are above or below the critical temperature), and $c > 0$ for stability⁴¹. In this problem we want to understand a little bit more about how Landau theory can capture some aspects of both continuous *and* discontinuous transitions.

⁴¹Such a free energy density, where a and b are both functions of $t = (T - T_c)/T_c$, and $p = (P - P_c)/P_c$, can be used to try to model a mixture of Helium-3 and -4, for instance

(A): Behavior of the order parameter In this part, we'll first derive expressions for the equilibrium value of the order parameter everywhere in the $a - b$ plane of parameter space by extremizing \mathcal{L} – this will tell us where the order parameter is zero, where it is not, where it changes, etc.

- (i) Consider the case in which $a < 0$: what is η^2 ?
- (ii) Consider the case in which $a > 0$ and $b > 0$: what is η ?
- (iii) Consider (with care!) the case in which $a > 0$ and $b < 0$: what is going on with η here?

(B): Zero-field phase diagram Use the results of part (A) to create a reasonable sketch of the $a - b$ plane of the phase diagram. This sketch should include:

- (i) The positions of the phase boundaries
- (ii) A representative sketch of form of \mathcal{L} in each region
- (iii) An indication, for each phase boundary, whether the order parameter changes continuously or discontinuously across the transition

3.5.3 Landau theory (but not for a magnet)!

Liquid crystals are composed of long, rod-like molecules. At high temperatures these molecules point every which way, but at low temperatures the molecules tend to align in the same direction. But the molecules do not have a true vector character, so if we think of them like vectors they don't care about whether they are parallel or anti-parallel. A good order parameter for this type of ordering (called “nematic⁴²” order) is not a scalar but a *matrix*, Q . Things can get more complicated (and less diagonal-looking), but for the basic isotropic-nematic transition the typical average value of Q is simple:

$$Q = \begin{pmatrix} \frac{2}{3}S & 0 & 0 \\ 0 & \frac{-1}{3}S & 0 \\ 0 & 0 & \frac{-1}{3}S \end{pmatrix},$$

where I've chosen a coordinate system where one axis is in the direction the molecules are aligned in, and the scalar S is a measure of how aligned the molecules are⁴³.

Let's build a Landau free energy density for this! Because \mathcal{L} is a scalar, and it must not depend on the direction the molecules happen to be lined up in, and because Q transforms like a tensor under rotations, we are forced to build \mathcal{L} out of traces of powers of Q (recall: the trace of a matrix is the sum of its diagonal elements; this is invariant under rotations of the coordinate system the matrix is written in). To fourth order:

$$\mathcal{L} = \frac{3r}{4} \text{Tr}\{Q^2\} - \frac{3w}{2} \text{Tr}\{Q^3\} + \frac{9u}{8} \text{Tr}\{Q^4\}.$$

⁴²From a greek word for “thread”

⁴³For instance, if θ is the angle each molecule makes with respect to the average direction of alignment, then $S = \frac{1}{2} \langle 3 \cos^2 \theta - 1 \rangle$

Notice that, physically, the cubic term is non-zero precisely because the order parameter doesn't depend on alignment vs anti-alignment (as it would if we picked a vector order parameter). In our case, since Q is so simple, explicitly re-write this as a Landau free energy just in terms of powers of S . Let w , and u be positive constants, and r be positive for high temperatures and negative at low temperatures (i.e., something like $r = a(T - T_c)$ for positive constant a).

(A) Qualitatively sketch the function $\mathcal{L}(S)$ for temperatures above, below, and at the critical temperature.

(B) The critical temperature, r_c , and the value of the order parameter at the critical temperature, S_c , can be found by simultaneously (a) extremizing \mathcal{L} to find the equilibrium value of S and (b) requiring that the value of the equilibrium \mathcal{L} be continuous (although obviously not smooth) as you go across the transition from the isotropic ($S = 0$) to the nematic ($S = S_c$) phase⁴⁴. What are r_c and S_c ?

(C) Because of that cubic term in \mathcal{L} , we have a model of the isotropic-nematic transition which is discontinuous in the order parameter – this means that there will be a latent heat associated with the transition. To determine it, look extremely close to the transition, writing $\mathcal{L}(S_c)$ and letting $r = a(T - T_c)$. You can calculate the entropy density in the nematic phase relative to the isotropic phase as $s = -\frac{\partial \mathcal{L}}{\partial T}$, and the latent heat that must be absorbed as $q = -T_c s$. What is q ?

3.5.4 Droplets and dimensionality near phase transitions Landau theory (but not for a magnet)!

3.5.5 Data collapse

The other day I was reading a paper⁴⁵ in which the authors were using a technique called “active microrheology” to study the mechanical properties of a disordered system: basically, they simulated the process of dragging a big probe particle through the system at different velocities, and then measured the friction experienced by that probe particle. (The idea, in case you're curious, is that you can relate these measurements of what happens at the microscopic scale to macroscopic properties of the bulk system – for instance, relating the friction acting on the probe to the viscosity of the substance.

They did these probe-dragging experiments for various system temperatures and probe velocities, and got something like the data in Fig. 3.2. Go to [this link](#) to find a file called “scalingData.csv,” a comma-separated collection of data corresponding to that figure. Each row of that file corresponds to a different temperature of the system. Each row is separated into pairs, corresponding to a bunch of “(probe velocity, friction on probe)” data at that temperature. The temperatures studied were

$$0.15, 0.2, 0.225, 0.25, 0.275, 0.3, 0.325, 0.35, 0.375, 0.4, 0.425, 0.45, 0.466$$

where 0.15 corresponds to the first row of scalingData.csv, and 0.466 corresponds to the last. The temperatures corresponding to each of the rows can be found in the “Ts.csv” file.

⁴⁴This should certainly inform something about your sketch in part A

⁴⁵Citation omitted, so you can't just look up the answer – ask me if you're interested!

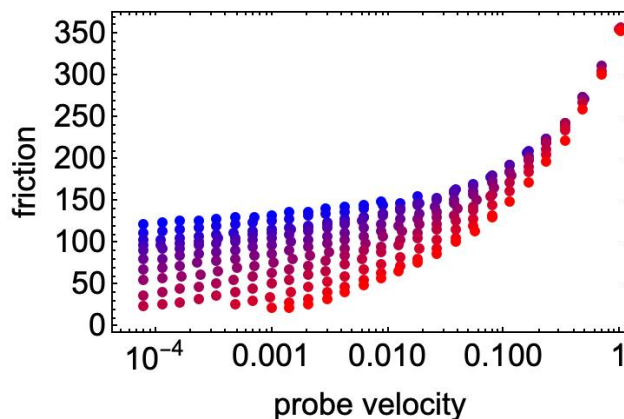


Figure 3.2: **Friction vs probe velocity** The friction acting on a probe particle as it is dragged through a disordered system as a function of the velocity of the probe. Different colors correspond to different temperatures of the system (blue-to-red→cold-to-hot).

First, we’re going to be manipulating this data, so make sure you can read those files and reproduce Fig. 3.2.

Next, the authors claim that their data suggests the existence of a critical point. From the figure above it kind of seems like everything is smooth and continuous as the temperature of the system is varied, but use the data to predict the critical temperature of the system, T_c , and scaling exponents which you may call β and γ . (The analogy isn’t perfect, but feel free to think of “friction” as an order parameter and “velocity” as an imposed external field.) Part of your submission should include a plot that justifies your predictions.

3.5.6 More Data collapse

The other day I was reading a paper in which the authors were trying to figure out if a dense monolayer of cells cultured from dog kidneys⁴⁶ were going through a phase transition associated with rigidity (i.e., turning from a disordered fluid to a disordered solid). Why not just, you know, poke at the cells and see if it’s rigid or not? Great question – I guess we have to assume that doing these experiments on cells is really really hard. Anyway, what they did instead was watch long videos of the monolayer of cells, and quantify how long they had to wait to see a rearrangement of cells. They did these experiments while varying a bunch of regulatory pathways that controlled various things (myosin contractility, organization of adhesive proteins inside the cell, and lots of other things I didn’t understand). They got something like the data in Fig. 3.3. We want to know if they observed something that is more like a phase transition, or something that is more like how the viscosity of liquids vary continuously with temperature

We’re going to be manipulating this data, which you will find a link to on the canvas page for this assignment. The top row contains 11 numbers (corresponding to the rows for experiments 1-11), and lists the effective temperature (“jiggle strength”) at which each experiment was done). The second row contains a list of 40 entries (corresponding to the

⁴⁶I don’t know, and I’m not going to ask

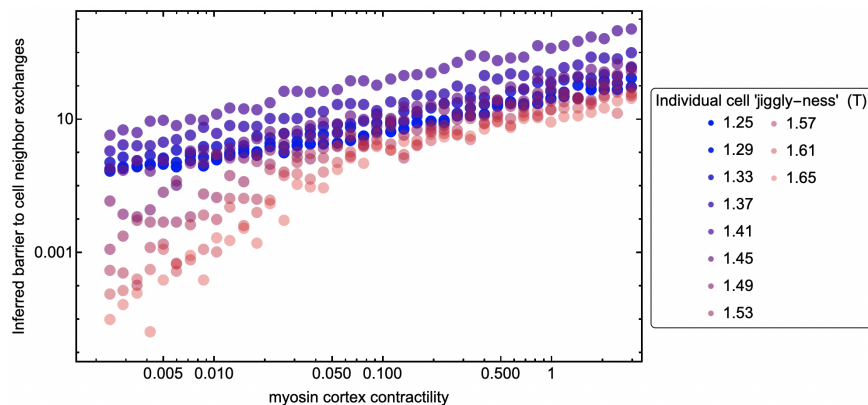


Figure 3.3: Something like the energy barriers to cell-cell rearrangements (inferred from how long one has to wait to observe them) on the y-axis, as a function of controlling some regulatory pathway in cells that helps determine the overall shape (more rounded, more irregular) on the x-axis. Each point is a different experimental result, and points have been color-coded by how much individual cells are just jiggling around most of the time (which is kind of like an effective temperature).

x-axis of the plot above). Finally, there are rows for each “Experiment.” Each of the entries is the observed/inferred energy barrier (corresponding to the y-axis on the plot above)⁴⁷.

(A) Use the spreadsheet and your program of choice (Mathematica? Python? Matlab? Excel? Whatever...) to reproduce Fig. 3.3 above (I obviously don’t care about all of the details of the styling – just load the data and plot it). Include in your submission the code to generate the plot and the plot itself.

(B) We want to know if all of this data suggests the existence of a critical point. The authors suggest that “barriers to cell rearrangement” are like an order parameter that is “thermodynamically” conjugate to this cellular shape parameter, in the same way that m and h are conjugate to each other. They furthermore assert that observing the magnitude of cell motion in between rearrangements is like a measurement of an effective temperature. Given all of that, write down something analogous to the Widom scaling hypothesis for these variables, and see if you can collapse the data and infer the existence and properties of a critical point. This will require you to try to vary T_c , β , and Δ to achieve a collapse.

You *could* do this by blindly guessing and checking different numbers, but you could also think about the limits of the scaling functions and try to be more systematic. In any event, your submission should include: (i) a description of your strategy for finding the best collapse of the data, (ii) a plot showing the quality of the data collapse you achieve, and (iii) your estimate of T_c and of the exponents β , γ , and δ .

⁴⁷So, for instance: one of the points on the plot at jiggle strength 1.53 (Expt5) and cell shape parameter 0.0041472 (column E) has coordinates in the plot corresponding to the $\{E2, E11\}$ cells in the spreadsheet.

Part II

**Statistical routes to thermodynamic
descriptions of macroscopic systems**

Chapter 4

Information, Entropy, and the Logic of Inference

In the last chapters we looked at thermodynamics as a phenomenological theory in which empirical observations were elevated into a set of “laws” that formed a consistent mathematical framework, a framework which was appropriate for treating various “black-box” systems. We then studied phases and phase transitions through a similarly phenomenological lens, letting ourselves be guided by symmetry principles. Ultimately, we want to see how that framework arises from the actual microscopic rules that real systems evolve according to, and this will rely on the statistical properties of having large numbers of interacting degrees of freedom. In this brief chapter we will cover the parts of probability theory that we will be using. Some of the chapter will be standard “here are some mathematical definitions” stuff, some of it will be calculational tools and tricks to use down the road, and some of it will be an alternate, information-centric approach to what we mean by “entropy.”

Note that Appendix B contains the basic definitions (of random variables, probability distributions, characteristic functions, cumulants, and so on) that we use in this chapter. [In the lectures I will typically do one of the observation sections immediately below, and then review the appendix in a lecture as needed.](#)

4.1 A funny observation

Just the other day⁴⁸, I was generating configurations of the standard 2D Ising model using [this web simulation applet](#). I played around at various temperatures and at field $h = 0$, letting the simulation reach what seemed like a point where the average energy stopped changing in time. I saved a bunch of these images of snapshots at different temperatures, and compressed those images using an off-the-shelf algorithm on my computer, *gzip*, and compared the size in bytes of the compressed image the size in bytes of one of the high-temperature images. I plotted the results in Fig. 4.1, and I was shocked! Just looking at “how much can I compress this image” I started to see something that looked suspiciously like the behavior of some of the static scaling plots we met in the last section! What is the connection between entropy

⁴⁸for fun, you know?

– which so far seems like a thermodynamic concept – and information and our ability to compress data? Let’s find out!

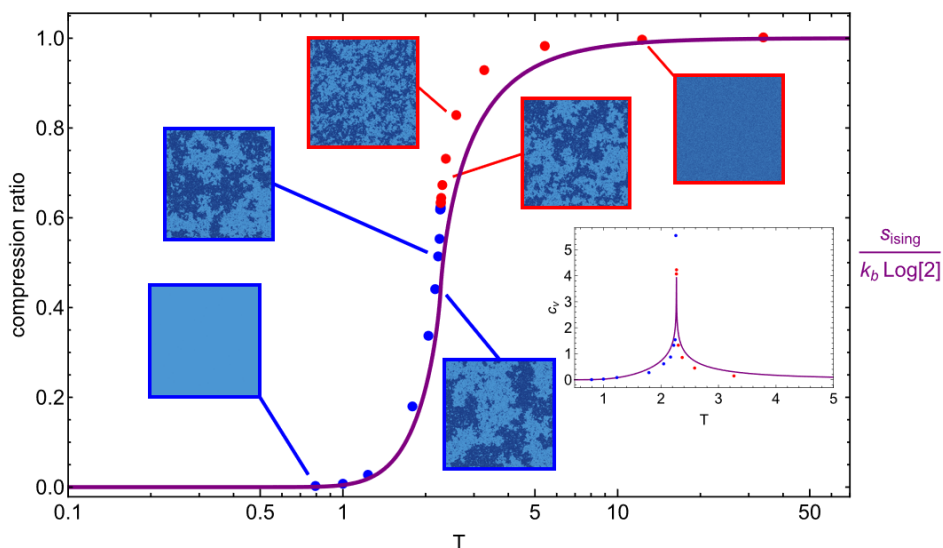


Figure 4.1: **Compressing pictures of the 2d ising model with no field** This figure was generated by saving snapshots of the 2D ising model, computed using [this link](#) at various temperatures and at field $h = 0$. Mathematica was used to take each image and compress it (using the default “Compress[]” function). The y -axis shows the size of the compressed image (using the “ByteCount[]” function) divided by the size of the compressed image at high temperature – i.e., the plot shows as a function of T how much more Mathematica was able to compress the various images relative to the most disordered image. Note how in the main plot there is a blue dot just to the left of the lowest red dots (where the coloring corresponds to above vs. below the critical temperature of the model). The inset shows a scaling-like plot of the log compression ratio vs the log of $|T - T_c|^{-1}$.

4.2 Moment and cumulant generating functions

In Chapter 5 we will introduce probability density functions (pdf – see Appendix B for definitions) whose variables are the *classical microstate* of a system – i.e., the complete list of particle positions and momenta – these ensemble densities will then be the building blocks of our statistical mechanical description. It will then be quite helpful to have in mind two related mathematical constructions.

The first is the *characteristic function*, which is simply the Fourier transform of the pdf. For a pdf for a single random variable, this is:

$$\tilde{p}(k) = \langle e^{-ikx} \rangle = \int dx p(x) e^{-ikx}. \quad (4.1)$$

Likewise, if you know the characteristic function, the pdf is the inverse FT⁴⁹:

$$p(x) = \frac{1}{2\pi} \int dk \tilde{p}(k) e^{ikx}. \quad (4.2)$$

Assuming that the moments of the distribution exist, we can expand the exponential⁵⁰ in the definition, $\exp(-ikx) = \sum_{n=0}^{\infty} \frac{(-ik)^n}{n!} x^n$, so that

$$\tilde{p}(k) = \left\langle \sum_{n=0}^{\infty} \frac{(-ik)^n}{n!} x^n \right\rangle = \sum_{n=0}^{\infty} \frac{(-ik)^n}{n!} \langle x^n \rangle. \quad (4.3)$$

We see from here that derivatives at $k = 0$ encode the moments of our original distribution:

$$\left. \frac{d^n \tilde{p}(k)}{d(-ik)^n} \right|_{k=0} = \langle x^n \rangle. \quad (4.4)$$

For this reason the characteristic function is also known as the *moment generating function*. It also allows for manipulations to get, e.g., relative moments pretty easily:

$$e^{ikx_0} \tilde{p}(k) = \langle \exp(-ik(x - x_0)) \rangle = \sum_{n=0}^{\infty} \frac{(-ik)^n}{n!} \langle (x - x_0)^n \rangle. \quad (4.5)$$

We'll also define the *cumulant generating function* as the log of the characteristic function⁵¹. We will define the *cumulants* of the distribution by in terms of the prefactors of the expansion of this function:

$$\log \tilde{p}(k) \equiv \sum_{n=1}^{\infty} \frac{(-ik)^n}{n!} \langle x^n \rangle_c. \quad (4.6)$$

From the definition of the characteristic function, and then by expanding the log as $\log(1 + x) = \sum_{n=1}^{\infty} \frac{-(-x)^n}{n}$, we have

$$\begin{aligned} \log \tilde{p}(k) &= \log \left(1 + \sum_{n=1}^{\infty} \frac{(-ik)^n}{n!} \langle x^n \rangle \right) \\ &= \left(\sum_{n=1}^{\infty} \frac{(-ik)^n}{n!} \langle x^n \rangle \right) - \frac{1}{2} \left(\sum_{n=1}^{\infty} \frac{(-ik)^n}{n!} \langle x^n \rangle \right)^2 + \frac{1}{3} \left(\sum_{n=1}^{\infty} \frac{(-ik)^n}{n!} \langle x^n \rangle \right)^3 + \dots \end{aligned} \quad (4.7)$$

By comparing the definition of the cumulants from Eq. 4.6 with the expansion in Eq. 4.7 (i.e., matching terms of order k^n), we can relate the cumulants to the moments. For example, the first two are pretty easy to see:

$$\langle x \rangle_c = \langle x \rangle \quad (4.8)$$

$$\langle x^2 \rangle_c = \langle x^2 \rangle - \langle x \rangle^2 = \langle (x - \langle x \rangle)^2 \rangle \quad (4.9)$$

$$\langle x^3 \rangle_c = \langle x^3 \rangle - 3 \langle x^2 \rangle \langle x \rangle + 2 \langle x \rangle^3 \quad (4.10)$$

⁴⁹Appendix B has an example of a distribution where it is straightforward to work out the characteristic function first and from it obtain the pdf.

⁵⁰We'll be seeing a lot of factorials from here on out; we apparently owe Christian Kramp for the now-standard $n!$ notation [44]. This was introduced in 1808, the same year that Gay-Lussac announced his law concerning the volumes of chemically reacting gases. Coincidence [45]?

⁵¹Some people would call this the “second characteristic function.” The difference is whether we end up with something which is defined independent of whether the moments themselves are well-defined.

Graphical connection between moments and cumulants The combinatorics of coefficient matching above might seem difficult to parse; there is a fun graphical way of remembering how to connect moments and cumulants. Notationally, let's represent the n th cumulant as a bag with n points inside of it (conveniently, since $\langle x \rangle_c = \langle x \rangle$, a bag with one point is the same as one point on its own). Then, the m th moment can be graphically expressed as the sum of all ways of distributing m points among bags. See Fig. 4.3.

$$\begin{aligned}
 \langle x \rangle &= \text{⊙} = \langle x \rangle_c \\
 \langle x^2 \rangle &= \text{⊙⊙} + \text{⊙⊙} = \langle x^2 \rangle_c + \langle x \rangle_c^2 \\
 \langle x^3 \rangle &= \text{⊙⊙⊙} + 3 \text{⊙⊙} + \text{⊙⊙⊙} \\
 &= \langle x^3 \rangle_c + 3 \langle x^2 \rangle_c \langle x \rangle_c + \langle x \rangle_c^3
 \end{aligned}$$

Figure 4.2: **Graphical expansion of the first three moments**

Slightly more formally, one might say you represent the n th cumulant as a connected cluster of points, and obtain the m th moment by adding together all subdivisions of m points into groupings of connected or disconnected clusters. The contribution of each subdivision to the sum is then the product of the connected cumulants it represents. This graphical connection between moments and cumulants is the basis for several diagrammatic computations (in statistical physics, and in field theory)... Will we see it again in this class? Stay tuned!

4.2.1 A comment about generating functions and cumulants

The definitions above probably just feel like a lot of definitional manipulations – what is the point? It is first worth noting that sometimes there are tricks that let you figure out what the moment generating function of a distribution will be, from which you can then figure out what all of the moments are. Section B.3.2 has such an example.

But what about the cumulants? First, it might be helpful to think of the cumulants as a hierarchical collection of the most important ways of describing a distribution. That is, you might first want to know about the most typical result coming from some distribution is (the mean), and if you know then mean what the typical spread around that value is (the variance), and then whether the distribution is symmetric about the mean or not, and so on. Well, first cumulant *is* the mean, and the second cumulant *is* the variance, and it has the nice property that it doesn't depend on where the distribution is centered (whereas just measuring the second moment wraps both the mean and the spread about the mean together). The third cumulant is the same as the third central moment, $\langle x^3 \rangle_c = \langle (x - \langle x \rangle)^3 \rangle$, and so is closely related to the usual definition of the skewness of a distribution. But at fourth order this pattern breaks down, and the cumulants and central moments are *not* the same. It's nice that there is apparently a one-to-one mapping between moments and cumulants – and thus that we are ultimately capturing the same amount of information about the distribution –

but why should we particularly care about cumulants instead of, say, central moments (or even just the raw moments)?

The short answer to that is twofold. On the one hand, cumulants have some nice mathematical properties. For instance, the sum of two independent random variables have a cumulant generating function which is just the sum of each variable's cumulant generating function (and, hence, the cumulants of the sum are just the sum of the cumulants). On the other hand, in the next few chapters we'll develop a deep connection between cumulants and various thermodynamic properties, and between partition functions and characteristic functions. Cumulants will further turn out to be an especially helpful calculational tool when we try to perturbatively describe interacting systems.

4.3 Adding random variables together and the central limit theorem

We typically think of statistical mechanics as being a relevant framework when the number of microscopic degrees of freedom, N , becomes very large. Indeed, in the *thermodynamic limit*, $N \rightarrow \infty$, a number of mathematical simplifications become available to our analysis of how systems behave. Let's first discuss what probability distributions emerge when you *sum* random variables together, which will lead us to the classical⁵² central limit theorem. The central limit theorem is an absolutely core engine in allowing us to make *precise* statements – of the sort we encountered in thermodynamics – in the context of an inherently *statistical* framework. More explicitly, for instance, we observed (empirically) that heat flows from hot to cold – not sometimes, or most of the time, but *all* the time. If we're going to start probabilistic arguments at the microscopic core of our theory, how do we end up with precise, essentially deterministic thermodynamic statements? In the coming chapters we'll see how the CLT gets us where we want to go.

4.3.1 Adding independent random variables

Suppose you have two pdfs of different (continuous) independent (but not necessarily identical) random variables, $p_X(x)$ and $p_Y(y)$, and are interested in the pdf associated with the random variable $Z = X + Y$ – how could we derive this? One way is to think about the cumulative probabilities:

$$P_Z(z) = \text{probability}(Z \in [-\infty, z]) = \text{probability}(X + Y \leq z).$$

We can write an integral expression for this, asking for the probability that Y takes a value in the range $[-\infty, z - x]$ for all possible values of x :

$$P_Z(z) = \int_{-\infty}^{\infty} dx \int_{-\infty}^{z-x} dy p_X(x)p_Y(y) = \int_{-\infty}^{\infty} dx \int_{-\infty}^z dw p_X(x)p_Y(w-x),$$

⁵²As opposed to the Lyapunov or other versions with weaker conditions.

where we have changed variables, $y = w - x$. From here, we can find the pdf by simply taking the derivative of the cumulative probability function:

$$p_Z(z) = \frac{dP_Z(z)}{dz} = \int_{-\infty}^{\infty} dx p_X(x)p_Y(z-x). \quad (4.11)$$

To get an intuition, let's think about what this looks like for simple, uniform distributions. For instance, suppose both X and Y are drawn from a uniform distribution in the domain $[0, 1]$:

$$p(x) = \begin{cases} 1 & x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Using this as both p_X and p_Y in Eq. 4.11, we can evaluate the integral to find

$$p_Z(z) = \begin{cases} z & \text{for } 0 < z \leq 1 \\ 2 - z & \text{for } 1 < z \leq 2 \\ 0 & \text{otherwise.} \end{cases}$$

What if we consider a new random variable formed by summing three random variable from this uniform distribution rather than two? We can iterate on our solution, combining the uniform distribution for $p_X(x)$ and the piecewise linear function above for the distribution associated with the sum of two variables. Let's call our new random variable Z_3 , as it represents the sum of three variables; by direct integration we find

$$p_{Z_3}(z) = \frac{1}{2} \cdot \begin{cases} z^2 & \text{for } 0 < z \leq 1 \\ -2z^2 + 6z - 3 & \text{for } 1 < z \leq 2 \\ (3-z)^2 & \text{for } 2 < z \leq 3 \\ 0 & \text{otherwise.} \end{cases}$$

We can repeat this iterative procedure as we sum more and more of these uniform random variables together, doing ever more piecewise integrations of polynomial terms. The result, considering the sum of $n > 1$ summed variables each drawn from the unit interval, can be written as

$$p_{Z_n}(z) = \frac{1}{(n-1)!} \sum_{k=0}^n (-1)^k \frac{n!}{k!(n-k)!} (x-k)_+^{n-1}, \quad (4.12)$$

where

$$(x-k)_+ = \begin{cases} x-k & \text{for } x-k \geq 0 < 1 \\ 0 & \text{otherwise.} \end{cases}$$

These pdfs are known as the uniform sum distributions or the Irwin-Hall family of distributions, and examples of them are plotted for various choices of n in Fig. 4.4. The mean and variance of these distributions are $n/2$ and $n/12$ respectively, and the figure also shows a Gaussian distribution with the same mean and variance. I suspect you are not surprised to see that $p_{Z_n}(z)$ starts approaching a Gaussian shape, even for relatively modest values of n .

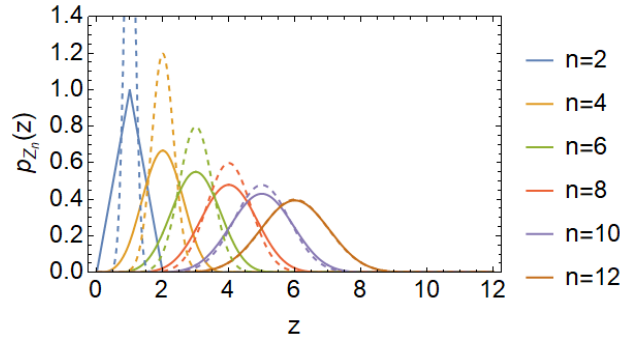


Figure 4.3: **The uniform sum distribution for various choices of n .** Dashed lines show the comparison with a Gaussian distribution having the same mean and variance as the corresponding uniform sum distribution.

4.3.2 The Central Limit Theorem

Clean this up

With that as a warm up, let's be more general and consider the sum of N random variables, $X = \sum_{i=1}^N x_i$, where the random variables x_i have some joint PDF $p(\mathbf{x})$. What is the cumulant generating function of the sum, $\log \tilde{p}_X(k)$? Well,

$$\log \tilde{p}_X(k) = \log \langle e^{-ikX} \rangle = \log \left\langle \exp \left(-ik \sum_{i=1}^N x_i \right) \right\rangle = \log \tilde{p}_x(k_1 = k, k_2 = k, \dots, k_N = k). \quad (4.13)$$

That is, it is the same as the log of the *joint characteristic function* of the x_i , but evaluated at the same k . Let's expand each side of the above equation, writing things so we can easily match powers of k :

$$\sum_{n=1}^{\infty} \frac{(-ik)^n}{n!} \langle X^n \rangle_c = (-ik) \sum_{i=1}^N \langle x_i \rangle_c + \frac{(-ik)^2}{2!} \sum_{i,j=1}^N \langle x_i x_j \rangle_c + \dots \quad (4.14)$$

Matching terms of order k^n , we see that

$$\langle X \rangle_c = \sum_i \langle x_i \rangle_c, \quad \langle X^2 \rangle_c = \sum_{i,j} \langle x_i x_j \rangle_c, \dots \quad (4.15)$$

Now, we specialize to the case of the classical central limit theorem by supposing that the x_i are both *independent*, so that $p(x_1, \dots, x_N) = p_1(x_1)p_2(x_2) \cdots p_N(x_N)$ and *identically distributed*, i.e., each of the labeled probability distributions $p_i(x_i)$ are the same⁵³, so that

$$p(x_1, \dots, x_N) = \prod_{i=1}^N p(x_i). \quad (4.16)$$

Now, the fact that the variables are independent means that the cross-correlation cumulants in Eq. 4.15 vanish, so that in the double sum only terms with $i = j$ contribute. Thus, we get

⁵³This combination of conditions, independent and identically distributed, is often abbreviated *iid*.

that

$$\langle X^m \rangle_c = \sum_{i=1}^N \langle x_i^m \rangle_c. \quad (4.17)$$

The condition of identically distributed takes us back to the case we looked at with the binomial distribution: the cumulants of N repeated but independent draws from the same distribution are easily related to the cumulants of the single-random-variable distribution:

$$\langle X^m \rangle_c = \sum_{i=1}^N \langle x_i^m \rangle_c = N \langle x^m \rangle_c. \quad (4.18)$$

The (classical) Central Limit Theorem follows directly. Define a new random variable to be

$$y = \frac{X - N \langle x \rangle_c}{\sqrt{N}}, \quad (4.19)$$

and then one computes its cumulants:

$$\langle y \rangle_c = 0, \quad \langle y^2 \rangle_c = \frac{\langle X^2 \rangle_c}{(\sqrt{N})^2} = \langle x^2 \rangle_c, \quad \langle y^m \rangle = \frac{N \langle x^m \rangle_c}{N^{m/2}}. \quad (4.20)$$

In words: as N becomes large, the distribution for a sum of *iid* random variables with mean μ and variance σ^2 converges to a distribution that itself has a finite mean, a variance that only grows as \sqrt{N} , and higher-order cumulants that all decay to zero as $N \rightarrow \infty$. Thus, sums of random variables converge to normal distributions, ignoring the details of what the original random variables looked like (up to some point). Note that the condition for this all working is really (a) the existence of the moments in question, and (b) a condition on how correlated the variables are allowed to be:

$$\sum_{i_1, \dots, i_m}^N \langle x_{i_1} \dots x_{i_m} \rangle_c \ll \mathcal{O}(N^{m/2}).$$

4.4 Adding exponentially large quantities together

The central limit theorem is a core result in how the thermodynamic limit works; another crucial piece of mathematical technology associated with $N \rightarrow \infty$ is the (surprising?) way that very large numbers behave when you add them together. What do I mean by “very large numbers,” there? In statistical physics we quite frequently run into (1) *intensive variables* (like T , P , etc.), which are independent of system size ($\mathcal{O}(N^0)$), (2) *extensive variables* (like S , E , etc.), which scale linearly with system size ($\mathcal{O}(N^1)$), and (3) *exponential variables* (like volumes of phase space), which fantastically large: ($\mathcal{O}(V^N) = \mathcal{O}(e^{aN})$). Exponentially large quantities behave a bit differently under addition than “normal” numbers, and this helps make calculating thermodynamic limits possible.

Summing exponentials

To warm up, let's suppose we want to sum together a large number of quantities which are themselves both positive and exponentially large:

$$S = \sum_{i=1}^{\mathcal{N}} \mathcal{E}_i, \quad (4.21)$$

where the terms $\mathcal{E}_i \sim \mathcal{O}(\exp(Na_i))$ and there are at most $\mathcal{N} \sim \mathcal{O}(N^P)$, i.e. the number of terms is at most polynomially (rather than exponentially) large in N .

This might seem tough, but actually: *we can approximate the entire sum by just picking out the largest term!* That is,

$$S \approx \mathcal{E}_{max}. \quad (4.22)$$

Here's the specific sense in which we can use this claim. First, it is clear that we can bound the sum by saying that the sum is surely in between the largest term and \mathcal{N} times the largest term:

$$\mathcal{E}_{max} \leq S \leq \mathcal{N}\mathcal{E}_{max}. \quad (4.23)$$

Next, let's take some natural logs and divide by N :

$$a_{max} \leq \frac{\log S}{N} \leq a_{max} + \frac{P \log \mathcal{N}}{N}, \quad (4.24)$$

but that last term on the right we can write (by assumption about the number of terms) as

$$\frac{\log \mathcal{N}}{N} = \frac{p \log N}{N}, \quad (4.25)$$

and this goes to zero as $N \rightarrow \infty$! Thus, as $N \rightarrow \infty$ we are bounding our sum by two terms which become the same, and we can write

$$\lim_{N \rightarrow \infty} \frac{\log S}{N} = \frac{\log \mathcal{E}_{max}}{N} = a_{max}. \quad (4.26)$$

So, even if the second-largest a_i is only *ever-so-slightly* less than the maximum one, upon exponentiation N times it gets completely dominated by the larger term.

Integrating exponentials (Saddle-point integrals)

We generalize this result in order to make a similar claim about integrating over a function that appears in an exponential multiplied by N , $\mathcal{I} = \int dx \exp(N\phi(x))$. We will see that this integral is dominated by the place where the function $\phi(x)$ itself is maximized.

To see that, we Taylor expand ϕ about its *maximum* x_m

$$\mathcal{I} = \int dx \exp \left(N\phi(x_m) - \frac{N}{2} |\phi''(x_m)| (x - x_m)^2 + \dots \right) \quad (4.27)$$

This term has two types of corrections encoded in that set of trailing dots. First, of course, there are the higher-order terms in the expansion of the function $\phi(x)$ about its maximum

value; these terms (if calculated) lead to a series of additional terms in powers of $1/N$. Second, there are contributions to this sum from any additional local maxima the function $\phi(x)$ might have. However, by arguments similar to those made in the previous subsection, any such contribution will be completely subdominant! Thus, we truncate the series at quadratic order as above and write

$$\mathcal{I} = e^{N\phi(x_m)} \int dx \exp\left(-\frac{N}{2}|\phi''(x_m)|(x-x_m)^2\right). \quad (4.28)$$

This is just a Gaussian integral, albeit one without its normalization factor, so we get

$$\mathcal{I} = e^{N\phi(x_m)} \sqrt{\frac{2\pi}{N|\phi''(x_m)|}} \Rightarrow \lim_{N \rightarrow \infty} \frac{\log \mathcal{I}}{N} = \phi(x_m). \quad (4.29)$$

Example: deriving Stirling’s formula The machinery of saddle point integration can be used to derive Stirling’s approximation for the factorial. Start by noting that

$$N! = \int_0^\infty x^N e^{-x}, \quad (4.30)$$

which can itself be seen by starting with $\int_0^\infty \exp(-\alpha x) = 1/\alpha$, taking N derivatives, and setting $\alpha = 1$. Some rearrangements of the above equation (writing $\phi(x) = \log x - x/N$), expanding about $x_m = N$, and doing the Gaussian integral gets you to

$$N! = N^N e^{-N} \sqrt{2\pi N} \left(1 + \mathcal{O}\left(\frac{1}{N}\right)\right), \quad (4.31)$$

the log of which is Stirling’s formula. Filling in the missing steps should be straightforward, and also an excellent way to make sure you understand the machinery of this method.

4.5 Information, Entropy, and unbiased estimation

We’ll close out this chapter by taking an information-based view of what we mean by entropy, one introduced by Shannon⁵⁴ in a groundbreaking 1948 paper⁵⁵. We will discuss the connection between information and entropy, and by thinking about ‘unbiased’ ways of assigning probabilities, we will formalize the *subjective* procedure of assigning probabilities discussed at the beginning of this chapter.

⁵⁴A massively important figure often cited as “father of information theory,” but also important for much more trivial accomplishments, such as co-inventing the very first wearable computer (to try to cheat at roulette) and creating a computer that used Roman numerals both on its keyboard and in its internal representation.

⁵⁵Ref. [46]. Note that Shannon named the symbol of entropy H , after Boltzmann’s H -theorem, which we’ll encounter in the very next chapter.

4.5.1 Shannon entropy

Let's change our focus from one in which we're characterize properties of the macroscopic world around us to a setting which seems very different: the problem of sending messages over a wire. We begin by imagining a source trying to send us a message from an "alphabet" of k characters, $\{a_1, \dots, a_k\}$ that have an discrete associated probability distribution $p(a_i)$, (here by "alphabet" we mean the characters in the alphabet together with their single-character probability distribution, $\{a_i, p(a_i)\}$). This isn't too much of a stretch: the actual alphabet has characters, and indeed, some letters appear more frequently than others in real messages.

In a bit of a simplification, let's assume that the characters are *iid* (in real messages there are, of course, correlations; we neglect them in this idealized setting). With this assumption, the probability that the source sends the n -character message $x = x_1 x_2 \cdots x_n$ is just

$$p(x) = \prod_{i=1}^n p(x_i). \quad (4.32)$$

Let's use the symbol X^n to denote the entire ensemble of n -length messages that are chosen under the assumption that the x_i are *iid*.

Compressing messages If one wanted to naively transmit the entire message, one would have to send $\sim \log k$ bits per symbol in the message, or $\sim n \log k$ for the entire message. Suppose the length of the message, n , grows very large. Is it possible to *compress* the message into a shorter string that conveys the same "information"? As long as $p(a_i)$ is not uniform, then yes! The total number of possible messages is k^n , but for large n most of those messages are *extremely* unlikely. In a "typical" long message we expect each character to occur about $n_i = np(a_i) + \mathcal{O}(n^{1/2})$ times. So the number of *typical* strings, g , is not k^n but rather

$$g = \frac{n!}{\prod_{i=1}^k (np(a_i))!}, \quad (4.33)$$

which is the number of ways of partitioning the set so that each letter in the alphabet occurs the typical number of times. Applying Stirling's formula, we find that

$$\log_2 g \equiv nH(X) \approx -n \sum_{i=1}^k p(a_i) \log_2 p(a_i), \quad (4.34)$$

where $H(X)$ is the *Shannon entropy* of the ensemble $X = \{a_i, p(a_i)\}$. If we imagine adopting a (completely impractical, but imaginable) code for messages of length n where the integers label typical messages of such length, a typical n -letter string could be communicated using about $nH(X)$ bits. To be extra explicit, for discrete probability distributions with values $\{p_i\}$ we will be defining the entropy in this way:

$$S = H(X) = -\langle \log p \rangle = - \sum_i p_i \log p_i. \quad (4.35)$$

One is, of course, tempted to generalize this to the continuous case by writing $S = -\langle \log p(x) \rangle = \int dx p(x) \log p(x)$, but... this seems to depend in a non-trivial way on what units one measures

x in. The resolution to this – that there is a specific choice of units that makes the information entropy defined this way agree with the thermodynamic one – is something we’ll tackle in upcoming chapters.

Compressing binary messages Let’s briefly show this in more detail for a binary alphabet: each character is either zero with probability p or one with probability $1 - p$, so the ensemble X is completely specified by the single value p . Well, for large values of n there are going to be about np zeros and $n(1 - p)$ ones, and the number of distinct strings of this form is given by the binomial coefficient. So, using $\log x! = x \log x - x + \mathcal{O}(\log x)$, we have:

$$\log g = \log \binom{n}{np} = \log \left(\frac{n!}{(np)!(n(1-p))!} \right) \quad (4.36)$$

$$\approx n \log n - n - (np \log(np) - np + n(1-p) \log(n(1-p)) - n(1-p)) \quad (4.37)$$

$$= nH(p), \quad (4.38)$$

$$\text{for } H(p) = -p \log p - (1-p) \log(1-p). \quad (4.39)$$

What about actual compression? Again, we make up an integer code that labels every typical message. There are about $2^{nH(p)}$ messages, and a priori *typical* messages occur with equal frequency, so we need to specify a given message by a binary string whose length is about $nH(p)$. If $p = 1/2$ (and thus $H(p) = 1$ for \log_2) we haven’t done anything: we need as many bits to communicate the message as there are in the message. But if the probability $p \neq 1/2$, our new code shortens typical messages. The insight here is that we don’t need a codeword for every message, just typical ones, since the probability of atypical messages is negligible!

4.5.2 Information, conditional entropy, and mutual information

Shannon entropy is a way of quantifying our ignorance (per letter) about the output of a source operating with alphabet X : if the source sends an n -character message, we need about $nH(X)$ bits to know the message. *Information* quantifies how much knowledge you gain by knowing the probability distribution the characters came from, i.e., “if you already know the alphabet, how many fewer bits do I need to transmit to communicate a (typical) message?”. Well, the total reduction in the number of bits for a n length message from the alphabet of k characters is

$$n \log_2 k - \left(-n \sum_i p_i \log_2 p_i \right) = n \left(\log_2 k + \sum_i p_i \log_2 p_i \right). \quad (4.40)$$

Given a knowledge of the p_i , we define the *information per bit*⁵⁶ as

$$I(X) = \log_2 k + \sum_i p_i \log_2 p_i, \quad (4.41)$$

and we see that information and entropy are one and the same (up to signs and constants).

Information and entropy of the uniform distribution As a quick example, suppose we have a uniform distribution of k characters, $p_i = 1/k$. Well:

$$S = -k \left(\frac{1}{k} \log_2 \frac{1}{k} \right) = \log_2 k \quad (4.42)$$

$$I = \log_2 k + \log_2 \frac{1}{k} = 0. \quad (4.43)$$

So, the entropy is the log of the number of equal-probability characters (sound familiar, from the microcanonical ensemble, perhaps?), and there is no information in the distribution.

Information and entropy of a delta function distribution The opposite extreme is also trivial to work out. Suppose the distribution is such that a particular event definitely happens: $p_i = \delta_{\alpha,i}$. Well:

$$S = 0 \quad (4.44)$$

$$I = \log_2 k. \quad (4.45)$$

By knowing the distribution you already know *everything* about the outcome of an n -length message, and the entropy (a quantification of ignorance) is zero.

Finally, suppose we have two *correlated* sources of information, X and Y (for uncorrelated sources we would have $p(x, y) = p_X(x)p_Y(y)$). Then, if I know these correlations then upon reading a message in Y^n I can further reduce my ignorance about a message generated by X^n , which means I should be able to further compress messages in X^n than I could without access to Y . This is captured by the *conditional entropy*,

$$H(X|Y) = H(XY) - H(Y) = \langle -\log p(x, y) + \log p(y) \rangle = \langle -\log p(x|y) \rangle, \quad (4.46)$$

⁵⁶Check if Kardar's convention in calling this info per bit is echoed elsewhere, and reference it. Some students are confused by discrepancy with, e.g., Wikipedia.

where we see the conditional probability distribution introduced earlier in this chapter. Unsurprisingly (given the connection between information and entropy above), the *mutual information* (in words: “the number of fewer bits per letter needed to specify X when Y is known”) is closely connected: The information about X gained when you learn about Y is

$$\begin{aligned} I(X; Y) &= H(X) - H(X|Y) \\ &= H(X) + H(Y) - H(XY) \\ &= H(Y) - H(Y|X), \end{aligned} \tag{4.47}$$

which is a quantification of the degree to which X and Y are correlated, and is symmetric under the interchange of X and Y as we see above.

4.5.3 Unbiased estimation of probabilities

We can now *use* the entropy as a way to quantify subjective assignments of probabilities! To start, if we know absolutely nothing about a probability distribution, the *unbiased estimate* is that every outcome is equally likely. Earlier, we saw that a special feature of a uniform distribution is that it (a) encodes no information and (b) maximizes the entropy. So, we generalize the idea of constructing an unbiased estimate as one that *maximizes entropy subject to the constraints* of whatever we happen to know.

A specific example should make the idea clear. Suppose we observe a discrete random variable for a while; we don’t pay too much attention to how often different outcomes occur, but we manage to observe that it has a specific mean value, $\langle F(x) \rangle = f$. We want make an unbiased estimate of the probabilities p_i , and we will use Lagrange multipliers (α , β) to impose the two constraints we now know about the probability distribution: we know it is normalized and we know it has a specific mean:

$$S(\{p_i\}, \alpha, \beta) = - \sum_i p_i \log p_i - \alpha \left(\sum_i p_i - 1 \right) - \beta \left(\sum_i p_i F(x_i) - f \right). \tag{4.48}$$

Maximizing the entropy with respect to the possible probabilities p_i :

$$0 = \frac{dS}{dp_i} = -\log p_i - 1 - \alpha - \beta F(x_i) \tag{4.49}$$

$$\Rightarrow p_i = e^{-(1+\alpha)} e^{-\beta F(x_i)}. \tag{4.50}$$

Given this form (which should remind you of Boltzmann weights!), we can then solve for α and β that satisfy the given constraints.

Make broader points about maxEnt formalism [47]. Note (for later reference) the Kubo derivation on page 91 (Near Equilibrium ensembles). [48] has a nice Gibbs-to-Shannon-to-Jaynes-to(self-consistent inference) path

4.6 Problems

4.6.1 Simple mutual information calculation

Suppose you have random variables X_1 and X_2 , with a joint PDF of

$$p(x_1, x_2) \propto \exp\left(-\frac{a_{11}x_1^2}{2} - \frac{a_{22}x_2^2}{2} - \frac{a_{12}x_1x_2}{2}\right).$$

Calculate the mutual information, $I(X, Y)$.

4.6.2 Biased random walks

Consider a biased random walk that starts at the origin and is composed of N steps of length l . The direction of each step is picked by choosing the azimuthal angle, $0 \leq \phi \leq 2\pi$ from a uniform probability distribution, and choosing the polar angle, $0 \leq \theta \leq \pi$ from the probability density function $p(\theta) = \frac{2}{\pi} \sin^2\left(\frac{\theta}{2}\right)$. Note that I've already absorbed the factor of $\sin\theta$ to account for the solid angle: $\int_{\pi}^{\pi} p(\theta)d\theta = 1$. Each step is independent of the other steps. I want to know “how big” this random walk is:

- **Find the mean position of the end of the random walk by calculating $\langle x \rangle$, $\langle y \rangle$, and $\langle z \rangle$.**
- **Find the characteristic size of the random walk by calculating $\langle x^2 \rangle$, $\langle y^2 \rangle$, and $\langle z^2 \rangle$.**
- **Since each step chooses a direction on the sphere of radius l , the variables x , y , and z are not independent (for step i , $x_i^2 + y_i^2 + z_i^2 = l^2$). Why didn't I ask you to calculate the covariances $\langle xy \rangle$, $\langle yz \rangle$, and $\langle xz \rangle$?**

4.6.3 Unbiased estimation

In this section we'll be deriving probability distributions associated with velocities (say, of a gas particle) under various circumstances, by generalizing the discrete “biased estimation of probabilities section of the notes. In both cases, let's think of a random variable v that can take any value $-\infty \leq v \leq \infty$.

Constrained speed:

Find the unbiased estimate of the probability distribution, $p_1(v)$, subject to the constraint that you know the speed: $\langle |v| \rangle = c$

Constrained kinetic energy:

Find the unbiased estimate of the probability distribution, $p_2(v)$, subject to the constraint that you know the average kinetic energy: $\langle mv^2/2 \rangle = mc^2/2$.

Generalization [worth zero points! Just think about it]:

It sure seems like, from the above two examples, that when the first n moments of a distribution are specified, the unbiased estimate is an exponential of an n th-order polynomial. Is this the case?

4.6.4 A saddle-point problem

I know that you know that the Taylor expansion of e^x is $\sum_n \frac{1}{n!} x^n$. Use the method of saddle-point integration to derive an asymptotic formula for the coefficient $\frac{1}{n!}$ in the limit of large n . How does this compare to Stirling's formula for $n!$?

4.6.5 New distributions from old

The lecture notes will tell you how to perform changes of variables on pdfs⁵⁷... but what if we want to find pdfs associated with *combinations* of other random variables? In some cases we know what will happen (like, “summing a large number of random variables will tend towards a Gaussian” from the central limit theorem), but what about more generally? In this problem, you may want to think about both pdfs *and/or* cumulative probabilities in order to get at the answer:

- (A): Suppose you have two pdfs of different (continuous) independent random variables, $p_X(x)$ and $p_Y(y)$. Write an integral expression for the pdf for the random variable $Z = X + Y$?
- (B): Are addition and subtraction equivalent, here? Write instead a expression for the pdf of the random variable $Z = X - Y$.
- (C): Same setup, but write an integral expression for the product of the random variables, $Z = XY$
- (D): Show that your expressions derived above are reasonable via a little numerical experiment. Assume that X and Y are *iid* variables uniformly distributed in the range $[-1, 1]$. What do your integral expressions simplify to? Use any compute program / programming language to generate a reasonably large number of random numbers; manipulate them appropriately and by make some histograms to demonstrate that your expressions fit your data.
- (E): What about the division of random variables? Consider the random variable $Z = X/Y$... Is a compact integral formula for this easy to obtain? Do more numerical experiments, supposing X and Y are *iid* variables uniformly distributed in the range $[1, 2]$, make a histogram of Z . Do the same supposing that X and Y are both in the range $[0, 1]$, and then also in the range $[-1/2, 1/2]$. Do you observe (empirically) any patterns?

⁵⁷Like, what if you know $p(x)$ by you really want the probability distribution $p(y)$ where $y = f(x)$

4.6.6 Manipulating random variables

I've been working on a problem where I want to figure out the elastic properties (bulk modulus, shear modulus, etc.) of a disordered colloidal solid just by looking through a microscope and watching the colloids fluctuate in space. After working hard I come up with an argument that if I measure a particular funny looking quantity, Λ_{xy} (defined below) over windows of size $L = 2R$, then the variance in that variable will be proportional to the inverse shear modulus:

$$\text{var}\Lambda_{xy} = \frac{kT}{4L^2}(G(L))^{-1},$$

and if I let L get bigger and bigger $G(L)$ will approach the shear modulus of the system. (You can tell there was hard work involved by the factor of 4: somebody did more than dimensional analysis for this problem!).

You have no problem with the derivation to get the above expression, but you think in my actual experiment I'm measuring pure noise. You set out to prove it:

Needed definitions

The definition of the quantity of interest is:

$$\Lambda_{xy} = \sum_i^n \Delta_j (Ay_i + Bx_i),$$

where $\{x_i, y_i\}$ is the spatial position of particle i , there are n particles in the observation window, Δ_j is the x -component of the displacement of particle i from one time point to the next, and A and B are structural quantities related to the relative arrangements of particles in the observation window:

$$A = \frac{a}{ac - b^2} \quad ; \quad B = \frac{-b}{ac - b^2}$$

$$a = \sum_i^n (x_i)^2 \quad b = \sum_i^n (x_i y_i) \quad c = \sum_i^n (y_i)^2$$

Part A: Warm-up

Let X be a random variable with a Gaussian probability density function with mean $\lambda = 0$ and variance $\sigma^2 = 1$. Consider the random variable $Y = X^2$. Using the change-of-variables formula, write down the probability density function, $p_Y(y)$. Plot it on a log-linear scale.

Part B: Structural quantities

You first want to think about A and B as random variables, and decide to slowly build up to what their distributions are:

Step 1: Assume that each of the particle positions are (1) uncorrelated with each other in the observation window and (2) uniformly distributed in the range $[-R, R]$, i.e.

$$p_{x_i}(x) = \begin{cases} \frac{1}{2R} & |x| \leq R \\ 0 & |x| > R \end{cases} .$$

What is the probability distribution of the quantities x^2 and y^2 ? What about the quantity xy ? Using the central limit theorem, write down an estimate of the probability density function for a , b , and c . Write your answers in terms of Gaussians with n , R , and all the right numerical factors.

Step 2: You next tackle the denominators, $ac - b^2$. Since this appears in the denominator, you decide to make some simplifying assumptions (can you think of any justifications?) and approximate

$$ac - b^2 = \frac{(a+c)^2}{4} - \frac{(a-c)^2}{4} - b^2 \approx \frac{(a+c)^2}{4} .$$

If the mean of $a+c$ had been zero I would have asked you to do some extra derivations, but it's not. Instead, I'll tell you that $(a+c)^2/4$ is a "non-central χ^2 variable with number of summed parameters $k=1$ " ([wiki link](#)). Given that information, what is your approximation for the following:

$$\begin{aligned} \langle ac - b^2 \rangle &= ? \\ \text{var}(ac - b^2) &= ? \end{aligned}$$

Step 3: You're finally ready to estimate properties of A and B , which involve ratios of random variables. [Assuming the numerator and denominator have vanishing covariance](#), you approximate:

$$\begin{aligned} \left\langle \frac{X}{Y} \right\rangle &= \frac{\langle X \rangle}{\langle Y \rangle} + \frac{\langle X \rangle \text{var} Y}{\langle Y \rangle^3} + \dots \\ \text{var} \frac{X}{Y} &= \frac{\text{var} X}{\langle Y \rangle^2} + \frac{\langle X \rangle^2 \text{var} Y}{\langle Y \rangle^4} + \dots \end{aligned}$$

From all of that, what are your estimates for:

$$\begin{aligned} \langle A \rangle &= ? & \langle B \rangle &= ? \\ \text{var} A &= ? & \text{var} B &= ? \end{aligned}$$

Part C: Combining everything!

You are firmly convinced that the Δ_j are just Gaussian distributed with zero mean and variance (set by the temperature, basically) σ_Δ^2 , uncorrelated with the

structural random variables. Using the facts for a collection of random variables X_i the variance behaves like

$$\begin{aligned}\text{var} \sum_i X_i &= \sum_i \text{var} X_i \\ \text{var} \prod_i X_i &= \prod_i (\text{var} X_i + \langle X_i \rangle^2) - \prod_i \langle X_i \rangle^2,\end{aligned}$$

and assuming that the contribution from each of the j particles contributes identically and independently, give an expression for

$$\text{var} \Lambda_{xy} = \text{var} \sum_i^n \Delta_j (Ay_i + Bx_i)$$

4.6.7 “Gaussian” cumulants

The Maxwell distribution (describing, say, the speed of a particle in a gas) is

$$p(x) = \sqrt{\frac{2}{\pi}} \frac{x^2}{a^3} \exp\left(-\frac{x^2}{2a^2}\right),$$

where in a gas a is related to the mass and temperature and where x is in the interval $0 \leq x \leq \infty$. I’m not going to ask you to take the Fourier transform to find the characteristic function⁵⁸, but I’m sure you remember the cumulants of a Gaussian distribution, and how to connect moments to cumulants (*hint: graphically? Or from the definitions if needs be?*) **What is the variance (second cumulant) of the Maxwell distribution?**

4.6.8 Information, compression, and entropy

During class I showed the result of a numerical experiment where I generated a bunch of random strings, of total length n and made up of 26 distinct symbols (an “alphabet”), drawing random numbers so that the probability of picking the i th letter of the English alphabet was

$$p_i \propto i^{-\alpha}.$$

I used “gzip,” a universal lossless compression tool, to compress these random strings (“messages”), and showed a plot of the ratio of the length of the original string to the compressed string (on average) on the y -axis vs. the parameter α on the x -axis. I showed that in the limit that n got large (but where, in practice, it didn’t have to be that large), the properly normalized compression ratio approached the Shannon entropy associated with the probability distribution of the alphabet, $H(\alpha)$.

⁵⁸life would be easier if x ran from negative infinity to positive infinity instead of starting from zero. Alas.

Conduct your own numerical experiment!

Choose a discrete alphabet made up of some number of characters, and think of a one-parameter family of probability distributions by which you could choose letters in your randomly generated messages:

$$p_i \propto F(i, \alpha)$$

for some function F (as in the above, but choose a different distribution). Try to compress⁵⁹ randomly generated messages for different α and message length n , and show that, after proper normalization, your compression efficiency is bounded by the Shannon entropy.

To get full credit on this problem, you should turn in (a) whatever code you wrote, (b) a plot showing some measure of “compression ratio” and Shannon entropy vs α for various n , and (c) an explanation of how you had to define and normalize the “compression ratio” (length of alphabet? choice of \ln vs \log_2 ? etc) so that compression approaches the bound set by the Shannon entropy in the $n \rightarrow \infty$ limit.

If you’re interested in learning more about how these ideas are being used, here are an interesting pair of papers from 2019: [49, 50]

4.6.9 Eigenvalues of matrices with random elements

Theoretical results on the properties of “random” matrices have been used to describe a (perhaps) surprising number of physical systems – including the mechanical response of jammed packings and the energy level spacing in heavy nuclei. In this problem we’ll consider square $N \times N$ matrices that are formed by a two-step process:

1. Generate a square $N \times N$ matrix, A , where every element of A are independent random numbers drawn from a Gaussian distribution of zero mean and standard deviation $\sigma = 1$.
2. Create the symmetric matrix you actually want to think about, $B = A + A^T$, i.e., by adding A to its transpose.

In this problem you’ll be doing both some analytical work and some numerical work: for the numerical work you can use any programming language or software you like (python, matlab, mathematica, fortran, assembly, whatever).

(A): Numerically generate a few ensembles of this type of matrix for $N = 2, 4, 6, 8, 10\dots$ something like 1000 samples for each matrix size should suffice⁶⁰. For each size, calculate the eigenvalues λ_n of each matrix, and sort them

⁵⁹Python has a `gzip.py` module, Mathematica has `ExportString[string, "GZIP"]` commands, etc.

⁶⁰On even a simple modern laptop this shouldn’t take much computing time, so feel free to do larger N or more samples per N if you’d like... but do at least this much. For reference: on the 2019 laptop I am typing on, once I had the code written this took Mathematica about 1.85 seconds to finish.

from smallest to largest. Create a list of the difference between neighboring eigenvalues, $\Delta\lambda = \lambda_{n+1} - \lambda_n$. Plot $\langle\Delta\lambda\rangle$ as a function of N , and plot the histograms of $\Delta\lambda/\langle\Delta\lambda\rangle$.

(B): Let's do some theory to try to understand our plots. Let's work in the $N = 2$ ensemble, where each of our matrices look like $B = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$. Defining $d = (c - a)/2$, write an expression for the difference of eigenvalues of this matrix. Given how these quantities are defined, what are the probability distributions that characterize the numbers a , b , c , and d ? Finally, calculate a formula for the probability distribution of the difference in eigenvalues. Compare this formula to the histograms in part (A). **Note:** Feel free to numerically calculate anything else as well, either as a test of the results you are getting or as a hint to what you should be getting!

(C): Hopefully you found that your result in (B) did an excellent job in describing *all* of the histograms you made in part (A)... I wonder how important our 2-step process for making matrices really was? Conduct two more numerical experiments. **First**, what if we ignore the symmetrization step? Create a bunch of non-symmetric matrices A composed of random gaussians, and compute their eigenvalues. What's going on, here? **Second**, what if we keep the symmetrization but drop the condition that the initial matrix elements are all drawn from a Gaussian distribution? Instead, generate a bunch of A matrices where each element is just randomly chosen to be either ± 1 , compute their $\Delta\lambda$, and make some histograms as N increases.

Chapter 5

Kinetic Theory: from Liouville to the H-theorem

Perspective and questions for this chapter

At the beginning of class we talked about thermodynamics as a phenomenological theory that encodes “black-box” observations of material systems into a coherent set of physical laws – laws of thermodynamics. But, as you might protest, “We know what’s inside of the box! It’s molecules interacting with each other⁶¹!” So, we should be able to derive the structure of thermodynamics from what we already know about equations of motion and the mathematics we learned in the chapter on probability.

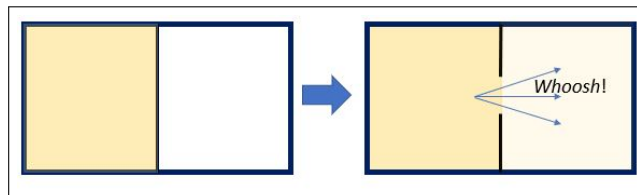


Figure 5.1: **A dilute gas rushes to fill the available space after a partition is removed** The gas is composed of microscopic degrees of freedom evolving according to completely time-reversible dynamics, but “Entropy increases.” What is entropy and how is it time irreversible for such a process?

In this chapter we will explore the ideas of ensembles of microstates that share the same macroscopic description – this will be something of a microscopic look at the rules associated with building a low-dimensional of a physical system that is actually composed of a very large number of degrees of freedom. We will also spend some time exploring the classical mechanics of dilute gases. Major questions we want to think about (ref Fig. 5.1):

1. How do we even *define* the idea of “equilibrium” for a system made out of particles?
2. Do such systems evolve towards equilibrium? How could they?! Every microscopic

⁶¹(either classically or quantum mechanically, as the scale of the problem demands)

equation we want to write down is *time reversible*, but if a system evolves from a non-equilibrium to an equilibrium state it is picking out a direction of time.

The kinetic theory of fluids was quite important historically, and many textbooks take it as a foundational component of any core statistical physics course. It not only explains macroscopic properties of gases in equilibrium (P, T , etc.), but also can be used to understand transport properties (thermal conductivity, viscosity), and the way a system *approaches* equilibrium from a non-equilibrium state. It is steadily being de-emphasized in many standard treatments of statistical mechanics (cf Pathria, or Sethna's *Entropy, Order Parameters, and Complexity*), but I think there is still something compelling about the subject. I have included sections that use the BBGKY hierarchy to derive the Boltzmann equation, discuss Boltzmann's H-theorem, and cover basic hydrodynamics, but these sections are for your interest only (i.e., I won't test you on them. But they're cool).

5.1 Elements of ensemble theory

5.1.1 Phase space of a classical system

We want to connect a thermodynamic description of a system – which is captured by only a handful of coordinates, like V, N, T , etc. – with a *classical microstate*. Classical microstates are specified by the positions and momenta of each of the N particles, $\mu = \{\mathbf{q}_1, \mathbf{p}_1, \dots, \mathbf{q}_N, \mathbf{p}_N\}$, which is a point in a $6N$ -dimensional phase space⁶². Since in this chapter we will be working entirely within a classical framework, we'll assume the system is governed by some Hamiltonian, \mathcal{H} , and we'll write down Hamilton's equations:

$$\begin{aligned}\frac{d\mathbf{p}_i}{dt} &= -\frac{\partial\mathcal{H}}{\partial\mathbf{q}_i} = \dot{\mathbf{p}}_i \\ \frac{d\mathbf{q}_i}{dt} &= \frac{\partial\mathcal{H}}{\partial\mathbf{p}_i} = \dot{\mathbf{q}}_i.\end{aligned}\tag{5.1}$$

Note that time-reversal invariance means that if we reverse the direction of time we transform $\mathbf{p} \rightarrow -\mathbf{p}$ and $\mathbf{q}(t) \rightarrow \mathbf{q}(-t)$.

Now, given that there are a handful of thermodynamic coordinates describing the system and $\sim 10^{23}$ describing the classical microstate, it will not surprise us that there is a many-to-one mapping between microstates and equilibrium states. How do we formalize this? Let's start by imagining we have \mathcal{N} “copies” of the same macrostate of our system, each corresponding to a different representative microstate, and we'll think about the *ensemble density function*,

$$\rho(\mathbf{p}, \mathbf{q}, t) = \lim_{\mathcal{N} \rightarrow \infty, d\Gamma \rightarrow 0} \frac{d\mathcal{N}(\mathbf{p}, \mathbf{q}, t)}{\mathcal{N}d\Gamma},\tag{5.2}$$

where

$$d\Gamma = \prod_{i=1}^N d^3\mathbf{p}_i d^3\mathbf{q}_i$$

⁶²Quite hard to draw

is how we'll be writing classical phase space volume differentials and $d\mathcal{N}$ is the number of microstates corresponding to the target macrostate in our differential box. Note that if we integrate ρ over all of phase space we get

$$\int d\Gamma \rho(\mathbf{p}, \mathbf{q}) = \int \frac{d\mathcal{N}}{\mathcal{N}} = 1, \quad (5.3)$$

so we see that between positivity (we're counting numbers of points, so it's not negative) and the above, the ensemble density is actually a probability density function. We already know things, then, from our work in the last chapter. E.g., to compute expectation values using this ensemble density we take some operator we're interested in, A , and do this:

$$\langle A \rangle = \int d\Gamma \rho(\mathbf{p}, \mathbf{q}, t) A(\mathbf{p}, \mathbf{q}).$$

5.1.2 Liouville's theorem and its consequences

Incompressibility of phase space volumes

Liouville's theorem, which I'm sure you encountered in classical mechanics, is a characterization of the evolution of ρ with time, and basically states that ρ behaves like an incompressible fluid. So, first, how does the phase space density evolve? Let's look at Fig. 5.2, which shows two dimensions of phase space around some representative point p_α, q_α , as well as where those points have moved after a short time dt .

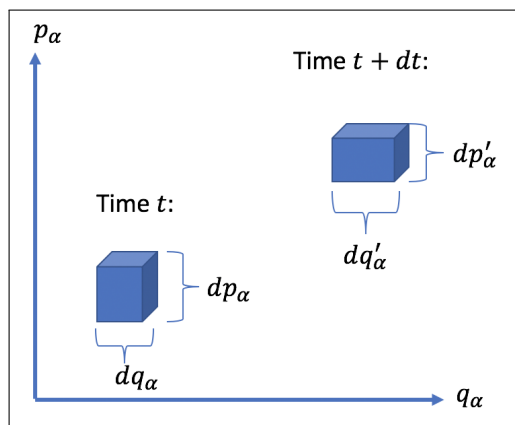


Figure 5.2: **Evolution of phase space volumes**

First, note that the representative point itself flows like

$$\begin{aligned} q'_\alpha &= q_\alpha + \dot{q}_\alpha dt + \mathcal{O}(dt^2) \\ p'_\alpha &= p_\alpha + \dot{p}_\alpha dt + \mathcal{O}(dt^2). \end{aligned} \quad (5.4)$$

How does the little volume element near this phase space point flow and distort? The points composing ends of the edge of the volume element may be moving at different velocities –

that difference depends, e.g., on $\frac{\partial \dot{q}_\alpha}{\partial q_\alpha} dq_\alpha$ (“how fast are the velocities changing along that direction multiplied by the separation of the points along that edge”), so that

$$\begin{aligned} dq'_\alpha &= dq_\alpha + \frac{\partial \dot{q}_\alpha}{\partial q_\alpha} dq_\alpha dt + \dots \\ dp'_\alpha &= dp_\alpha + \frac{\partial \dot{p}_\alpha}{\partial p_\alpha} dp_\alpha dt + \dots \end{aligned} \quad (5.5)$$

Thus, for each pair of conjugate coordinates we see that

$$dq'_\alpha dp'_\alpha = dq_\alpha dp_\alpha \left[1 + dt \left(\frac{\partial \dot{q}_\alpha}{\partial q_\alpha} + \frac{\partial \dot{p}_\alpha}{\partial p_\alpha} \right) + \mathcal{O}(dt^2) \right] = dq_\alpha dp_\alpha, \quad (5.6)$$

where the last equality holds because the term proportional to dt vanishes by equality of mixed partials after using Hamilton’s equation. This is just the familiar statement that Hamiltonian dynamics preserves phase space volumes: $d\Gamma = d\Gamma'$.

Liouville’s theorem

The above has consequences for our ensemble density. All of the states $d\mathcal{N}$ that were originally near (\mathbf{p}, \mathbf{q}) move to the neighborhood of $(\mathbf{p}', \mathbf{q}')$, and we just learned that they also occupy the same phase space volume. Thus, $d\mathcal{N}/d\Gamma$ is unchanged (“ ρ behaves like an incompressible fluid”), and we can write

$$\rho(\mathbf{p}, \mathbf{q}, t) = \rho(\mathbf{p}', \mathbf{q}', t + dt). \quad (5.7)$$

Expanding out this expression, we can write

$$\rho(\mathbf{p}, \mathbf{q}, t) = \rho(\mathbf{p} + \dot{\mathbf{p}}dt, \mathbf{q} + \dot{\mathbf{q}}dt, t + dt) \quad (5.8)$$

$$= \rho(\mathbf{p}, \mathbf{q}, t) + \left[\sum_\alpha \dot{p}_\alpha \frac{\partial \rho}{\partial p_\alpha} + \dot{q}_\alpha \frac{\partial \rho}{\partial q_\alpha} + \frac{\partial \rho}{\partial t} \right] dt + \mathcal{O}(dt^2). \quad (5.9)$$

That quantity in brackets has the form of something we’ll meet many times. Let’s define the total derivative (or “streamline derivative⁶³”) of a function f as

$$\frac{d}{dt} f(\mathbf{p}, \mathbf{q}, t) = \frac{\partial f}{\partial t} + \sum_\alpha \frac{\partial f}{\partial p_\alpha} \dot{p}_\alpha + \frac{\partial f}{\partial q_\alpha} \dot{q}_\alpha, \quad (5.10)$$

where the interpretation is that d/dt is the derivative “as you flow” (following the evolution of the volume of phase space as it itself moves through phase space), whereas the partial derivative ∂_t is like sitting a fixed position in space and watching the changes in f in time at that location. Anyway, the incompressibility of phase space flow, Eq. 5.7, tells us that the total time derivative of ρ vanishes. Thus, from Eq. 5.9 ρ says that

$$\frac{d\rho}{dt} = 0 = \frac{\partial \rho}{\partial t} + \sum_\alpha \left(\frac{-\partial \mathcal{H}}{\partial q_\alpha} \right) \frac{\partial \rho}{\partial p_\alpha} + \left(\frac{\partial \mathcal{H}}{\partial p_\alpha} \right) \frac{\partial \rho}{\partial q_\alpha} \quad (5.11)$$

$$\Rightarrow \frac{\partial \rho}{\partial t} = \{\mathcal{H}, \rho\}, \quad (5.12)$$

where we have defined the Poisson bracket⁶⁴

$$\{A, B\} = \sum_i \frac{\partial A}{\partial q_i} \frac{\partial B}{\partial p_i} - \frac{\partial A}{\partial p_i} \frac{\partial B}{\partial q_i} \quad (5.13)$$

⁶³Sometimes also denoted as $\frac{D}{Dt}$ rather than $\frac{d}{dt}$

⁶⁴Using curly braces, and with Pathria’s / Goldstein’s sign convention here instead of Landau’s, though

5.1.3 Equilibrium ensemble densities

Clean this old text up! Individual microstates will evolve in time (microscopically we certainly know that if we take snapshots of the state of our system at time t and $t + \Delta t$ the microstates will look measurably different!). At the same time, we want our equilibrium *macrostates* to not be changing in time, and in particular we want anything we care to measure to not be changing with time. What does this imply?

Time dependence of observables

First do the direct $\langle dA/dt \rangle$, then come back to this Given an observable A , its total time derivative is

$$\frac{dA}{dt} = \left[\sum_{\alpha} \frac{\partial A}{\partial q_{\alpha}} \dot{q}_{\alpha} + \frac{\partial A}{\partial p_{\alpha}} \dot{p}_{\alpha} \right] + \frac{\partial A}{\partial t}. \quad (5.14)$$

Assuming that A doesn't explicitly depend on time **a la pressure, kinetic energy, etc**, we see that the time dependence of A is just

$$\frac{dA}{dt} = \left[\sum_{\alpha} \frac{\partial A}{\partial q_{\alpha}} \frac{\partial \mathcal{H}}{\partial p_{\alpha}} - \frac{\partial A}{\partial p_{\alpha}} \frac{\partial \mathcal{H}}{\partial q_{\alpha}} \right] = \{A, \mathcal{H}\}. \quad (5.15)$$

Indeed, observables in individual microstates may be evolving in time, but we want their *ensemble average* to be stationary in equilibrium.

Thus, ensemble average of the time derivative is just

$$\left\langle \frac{dA}{dt} \right\rangle = \langle \{A, \mathcal{H}\} \rangle. \quad (5.16)$$

We can get the same result by taking the time derivative of the ensemble average:

$$\frac{d}{dt} \langle A \rangle = \int d\Gamma \frac{\partial \rho(\mathbf{p}, \mathbf{q}, t)}{\partial t} A(\mathbf{p}, \mathbf{q}) \quad (5.17)$$

$$= \sum_{i=1}^{3N} \int d\Gamma A(\mathbf{p}, \mathbf{q}) \{ \mathcal{H}, \rho \}, \quad (5.18)$$

where in the first line we have used the Leibniz integral rule to turn the total time derivative outside the integral sign into a partial derivative inside it. From this expression, one can then write out the Poisson bracket and use integration by parts to take the partial derivatives off of the ρ . This leads to:

$$\frac{d}{dt} \langle A \rangle = \sum_{i=1}^{3N} \int d\Gamma \rho \left[\left(\frac{\partial A}{\partial p_i} \frac{\partial \mathcal{H}}{\partial q_i} - \frac{\partial A}{\partial q_i} \frac{\partial \mathcal{H}}{\partial p_i} \right) + A \left(\frac{\partial^2 \mathcal{H}}{\partial p_i \partial q_i} - \frac{\partial^2 \mathcal{H}}{\partial q_i \partial p_i} \right) \right] \quad (5.19)$$

$$= - \int d\Gamma \rho \{ \mathcal{H}, A \} = \langle \{A, \mathcal{H}\} \rangle \quad (5.20)$$

Thus, to get ensemble averages that don't depend on time, we will look for stationary ensemble densities, i.e., ones that satisfy

$$\frac{\partial \rho_{eq}}{\partial t} = 0. \quad (5.21)$$

We can combine Liouville's theorem above with our criteria that in equilibrium ρ is stationary to come up with a criteria for equilibrium ensemble densities:

$$\frac{\partial \rho_{eq}(\mathbf{q}, \mathbf{p})}{\partial t} = 0 = \{\mathcal{H}, \rho_{eq}\}. \quad (5.22)$$

In principle, we now have the task of solving the partial differential equation of $6N$ variables coming from setting the Poisson bracket to zero. In practice, we'll guess! We have already required that in equilibrium ρ has no explicit time derivative; we could also assume it has no dependence on \mathbf{q} or \mathbf{p} at all: $\rho_{eq} = \text{const.}$ is certainly a valid solution of the above (which is like saying the ensemble of systems corresponding to the equilibrium macrostate are uniformly distributed through phase space).

More generally, we could come up with more interesting ensemble densities by allowing an *implicit* dependence on the phase space coordinates by letting ρ be some arbitrary function of \mathcal{H} , because

$$\{\rho_{eq}(\mathcal{H}), \mathcal{H}\} = \sum_i \left(\frac{\partial \rho}{\partial \mathcal{H}} \frac{\partial \mathcal{H}}{\partial q_i} \right) \frac{\partial \mathcal{H}}{\partial p_i} - \left(\frac{\partial \rho}{\partial \mathcal{H}} \frac{\partial \mathcal{H}}{\partial p_i} \right) \frac{\partial \mathcal{H}}{\partial q_i} = 0. \quad (5.23)$$

Different choices of this function of \mathcal{H} lead to different *ensembles* in statistical physics. For example, choosing

$$\rho(\mathcal{H}) = \delta(\mathcal{H} - E),$$

i.e., equally weighting all microstates that have a particular energy, gives us the microcanonical ensemble. Choosing instead

$$\rho(\mathcal{H}) \propto \exp(-\beta \mathcal{H}(\mathbf{q}, \mathbf{p}))$$

gives us the canonical ensemble.

Even more generally, while we're in the business of simply guessing solutions to Eq. 5.22, we can assume that ρ is an arbitrary function of both \mathcal{H} itself *and* of any conserved quantities. To see why, we first note that if some quantity A is conserved under the Hamiltonian, $\{A, \mathcal{H}\} = 0$. Then, using the same sort of manipulations as above, we can write

$$\{\rho_{eq}(\mathcal{H}, A), \mathcal{H}\} = \frac{\partial \rho}{\partial \mathcal{H}} \{\mathcal{H}, \mathcal{H}\} + \frac{\partial \rho}{\partial A} \{A, \mathcal{H}\} = 0, \quad (5.24)$$

satisfying our condition for ρ_{eq} to be an equilibrium distribution.

5.2 BBGKY⁶⁵ hierarchy

Pattern matching for the course: how do we start with too much information and create reduced-dimensional descriptions? ρ definitely has too much info (joint PDF of $6N$ variables?), so how do we organize that information more hierarchically?

⁶⁵Bogoliubov-Born-Green-Kirkwood-Yvon. Some aspects of this were introduced by Yvon in 1935; the actual hierarchy was written out by Bogoliubov in 1945; Kirkwood did relevant work on kinetic transport in 1945/1946; Born and Green used an analogous structure for the kinetic theory of liquids in 1946

Starting comments

We are about to embark on a sequence of formal manipulations, so it is worth emphasizing where we are going today (and why). In the last lecture we addressed the first question posed at the start of the chapter – “How do we define the idea of equilibrium for a system described microscopically as a collection of particles?” – by thinking about stationary ensemble density functions.

The rest of the chapter will be devoted to the second question: if we start from a non-equilibrium density, can we show it evolves towards equilibrium (a la Fig. 5.1 – an experiment which we know reproducibly leads from one equilibrium state to another)? Where does time irreversibility come from?

We have already seen “entropy” defined in two seemingly very different ways: entropy as a thermodynamic state function somehow related to heat and temperature, and entropy in the information theoretic sense as a quantification of our ignorance about the output of a probabilistic source. By the end of the chapter we will see yet another version of entropy: the Boltzmann version of entropy, capturing information about one-particle densities.

BBGKY

In the last lecture we encountered the ensemble density, $\rho(\mathbf{p}_1, \dots, \mathbf{p}_N, \dots, \mathbf{q}_N, t)$, but in general this contains *far* more microscopic information than we would ever need to compute or describe equilibrium properties (for instance – knowledge of the typical behavior of just a single particle would be sufficient to calculate the pressure of a gas. Let’s define the *one-particle density* as the expectation value of finding any of the N particles at some particular location with some particular momentum:

$$f_1(\mathbf{p}, \mathbf{q}, t) = \left\langle \sum_{i=1}^N \delta^3(\mathbf{p} - \mathbf{p}_i) \delta^3(\mathbf{q} - \mathbf{q}_i) \right\rangle = N \int \prod_{i=2}^N dV_i \rho(\mathbf{p}_1 = \mathbf{p}, \mathbf{q}_1 = \mathbf{q}, \mathbf{p}_2, \dots, \mathbf{p}_N, \mathbf{q}_N, t), \quad (5.25)$$

where $dV_i = d^3\mathbf{p}_i d^3\mathbf{q}_i$. We are going to be writing integrals over products of dV_i quite a lot, so let’s introduce some compact notation for our measures of integration in phase space. Let’s use the symbol

$$d\Gamma_a^b = \prod_{i=a}^b dV_i. \quad (5.26)$$

Here, $d\Gamma_1^N$ is simply what we were calling $d\Gamma$ in the previous section.

The general s -particle density is defined similarly, integrating over the $s + 1$ through N th variables:

$$f_s(\mathbf{p}_1, \dots, \mathbf{q}_s, t) = \frac{N!}{(N-s)!} \rho_s(\mathbf{p}_1, \dots, \mathbf{q}_s, t), \quad (5.27)$$

where

$$\rho_s(\mathbf{p}_1, \dots, \mathbf{q}_s, t) = \int d\Gamma_{s+1}^N \rho(\mathbf{p}_1, \dots, \mathbf{q}_N, t) \quad (5.28)$$

is just the unconditional pdfs for the coordinates of s of the particles. These f_s functions differ from the unconditional pdfs by a simple normalization factor, and we’ll see why each is vaguely preferable in different situations.

Now, this is a lovely set of definitions. What we really want is a way of expressing the time evolution of these s -particle densities (e.g., if we can tell the pressure from f_1 , then knowing $f_1(t)$ will let us study how pressure equilibrates in time as we go from one equilibrium state to the other; again, c.f. Fig. 5.1). Well,

$$\frac{\partial f_s}{\partial t} = \frac{N!}{(N-s)!} \int d\Gamma_{s+1}^N \frac{\partial \rho}{\partial t} = \frac{N!}{(N-s)!} \int d\Gamma_{s+1}^N \{ \mathcal{H}, \rho \}. \quad (5.29)$$

It would be quite difficult to make any interesting headway here for a truly arbitrary Hamiltonian, so let's consider the case of up to two-body interactions:

$$\mathcal{H} = \sum_{i=1}^N \left[\frac{\mathbf{p}_i^2}{2m} + U(\mathbf{q}_i) \right] + \frac{1}{2} \sum_{i,j=1}^N V(\mathbf{q}_i - \mathbf{q}_j), \quad (5.30)$$

where U is some external potential and V is some pairwise inter-particle potential. What's our strategy going to be? Well, we have to integrate over a Poisson bracket, so integration by parts will be extremely useful whenever we're taking a derivative with respect to one of the variables we're integrating over. Because of this, let's partition the sums we have to deal with into (a) one which covers *only* the first s particle coordinates, (b) ones which cover *only* the last $N-s$ particle coordinates, and (c) ones which couple the first s and the last $N-s$ particles together:

$$\mathcal{H} = \mathcal{H}_s + \mathcal{H}_{N-s} + \mathcal{H}_x \quad (5.31)$$

$$\mathcal{H}_s = \sum_{i=1}^s \left[\frac{\mathbf{p}_i^2}{2m} + U(\mathbf{q}_i) \right] + \frac{1}{2} \sum_{i,j=1}^s V(\mathbf{q}_i - \mathbf{q}_j) \quad (5.32)$$

$$\mathcal{H}_{N-s} = \sum_{i=s+1}^N \left[\frac{\mathbf{p}_i^2}{2m} + U(\mathbf{q}_i) \right] + \frac{1}{2} \sum_{i,j=s+1}^N V(\mathbf{q}_i - \mathbf{q}_j) \quad (5.33)$$

$$\mathcal{H}_x = \sum_{i=1}^s \sum_{j=s+1}^N V(\mathbf{q}_i - \mathbf{q}_j). \quad (5.34)$$

This lets us write⁶⁶

$$\frac{\partial \rho_s}{\partial t} = \int d\Gamma_{s+1}^N \{ \mathcal{H}_s + \mathcal{H}_{N-s} + \mathcal{H}_x, \rho \}, \quad (5.35)$$

and we'll take each part of the Poisson bracket in turn.

Self term, \mathcal{H}_s Notice that the variables we are integrating over do not show up in ρ_s or \mathcal{H}_s , so we can interchange the order of integrations and differentiations and write

$$\int d\Gamma_{s+1}^N \{ \mathcal{H}_s, \rho \} = \{ \mathcal{H}_s, \left(\int d\Gamma_{s+1}^N \rho \right) \} = \{ \mathcal{H}_s, \rho_s \}. \quad (5.36)$$

This has a natural interpretation / connection to what we already know: *if there are only s particles, we have just re-written the Liouville equation.*

⁶⁶This is just a factor of $\frac{N!}{(N-s)!}$ different from the equation above – we can always freely trade this factor back and forth to go from f_s to ρ_s .

Just the non- s particle terms, \mathcal{H}_{N-s} First, we write out the Poisson bracket involved:

$$\int d\Gamma_{s+1}^N \sum_{j=1}^N \left(\frac{\partial \rho}{\partial \mathbf{p}_j} \cdot \frac{\partial \mathcal{H}_{N-s}}{\partial \mathbf{q}_j} - \frac{\partial \rho}{\partial \mathbf{q}_j} \cdot \frac{\partial \mathcal{H}_{N-s}}{\partial \mathbf{p}_j} \right). \quad (5.37)$$

Well, any of the terms coming from the Poisson bracket with $j < s + 1$ vanish because \mathcal{H}_{N-s} doesn't depend on those particle positions and momenta. All of the rest, though, we notice that we are integrating over a coordinate which appears as a derivative. We use integration by parts to pull the derivatives off of the ρ and onto the \mathcal{H}_{N-s}

$$= [\text{surface terms}] + \int d\Gamma_{s+1}^N \sum_{j=s+1}^N \rho \left(\frac{\partial^2 H_{N-s}}{\partial \mathbf{p}_i \partial \mathbf{q}_i} - \frac{\partial^2 H_{N-s}}{\partial \mathbf{q}_i \partial \mathbf{p}_i} \right) = 0. \quad (5.38)$$

Cross terms, \mathcal{H}_x Apparently if there is going to be anything interesting it will come from these cross terms. We're making progress, here:

$$\int d\Gamma_{s+1}^N \sum_{j=1}^N \left[\frac{\partial \rho}{\partial \mathbf{p}_j} \cdot \frac{\partial \mathcal{H}_x}{\partial \mathbf{q}_j} - \frac{\partial \rho}{\partial \mathbf{q}_j} \cdot \frac{\partial \mathcal{H}_x}{\partial \mathbf{p}_j} \right] \quad (5.39)$$

$$= \int d\Gamma_{s+1}^N \left(\left[\sum_{k=1}^s \frac{\partial \rho}{\partial \mathbf{p}_k} \cdot \sum_{j=s+1}^N \frac{\partial V(\mathbf{q}_k - \mathbf{q}_j)}{\partial \mathbf{q}_k} \right] + \left[\sum_{j=s+1}^N \frac{\partial \rho}{\partial \mathbf{p}_j} \cdot \sum_{k=1}^s \frac{\partial V(\mathbf{q}_j - \mathbf{q}_k)}{\partial \mathbf{q}_j} \right] \right) \quad (5.40)$$

where we have used the fact that \mathcal{H}_x is independent of any \mathbf{p}_i , and then split the first term depending on whether the derivatives are variables that are integrated over or not. We've indexed things so that the k always refer to one of the first s particles, and the J refer to one of the last $(N - s)$ particles. We now note that we can integrate by parts again, bringing a $\frac{\partial}{\partial p_j}$ onto the sum over the potential terms in \mathcal{H}_x and hence causing it to vanish. Doing this, and interchanging the order of integration and summing, gives

$$\int d\Gamma_{s+1}^N \{ \mathcal{H}_x, \rho \} = \sum_{k=1}^s \int d\Gamma_{s+1}^N \frac{\partial \rho}{\partial \mathbf{p}_k} \cdot \sum_{j=s+1}^N \frac{\partial V(\mathbf{q}_k - \mathbf{q}_j)}{\partial \mathbf{q}_k} \quad (5.41)$$

Physically, we expect that we can treat all of the $j = s + 1, \dots, N$ particles equivalently, so we change the labels of our sums and replace the sum by $(N - s)$ equivalent integrations. We also separate out the integral over the $(s + 1)$ particle and the others:

$$\int d\Gamma_{s+1}^N \{ \mathcal{H}_x, \rho \} = (N - s) \sum_{k=1}^s \int dV_{s+1} \frac{\partial V(\mathbf{q}_k - \mathbf{q}_{s+1})}{\partial \mathbf{q}_k} \cdot \frac{\partial}{\partial \mathbf{p}_k} \left(\int d\Gamma_{s+2}^N \rho \right). \quad (5.42)$$

Notice that the final term in parentheses, above, is just ρ_{s+1} . This, at last, is the key connection we wanted to make.

Combining everything: the hierarchy Using either ρ_s or f_s , we have a connection between different s -body densities:

$$\boxed{\frac{\partial f_s}{\partial t} - \{\mathcal{H}_s, f_s\} = \sum_{i=1}^s \int dV_{s+1} \frac{\partial V(\mathbf{q}_i - \mathbf{q}_{s+1})}{\partial \mathbf{q}_i} \cdot \frac{\partial f_{s+1}}{\partial \mathbf{p}_i}}. \quad (5.43)$$

This, at last, is the BBGKY Hierarchy! Physically, what have we done? We started with a very complicated function of $\mathcal{O}(10^{23})$ variables and replaced it with a set (for $s = 1, 2, \dots, N$) of $\mathcal{O}(10^{23})$ coupled equations. This may not feel like progress, but a key insight is that this set of coupled equations forms a hierarchy that isolates the simplest (and presumably most physically relevant) densities: we have an equation that gives us f_1 provided we know f_2 , and f_2 provided we know f_3 , and so on. This provides us with a powerful physical tool: given a particular problem we want to solve, we can *inject physical approximations* to decide what terms in this hierarchy can be truncated or simplified or ignored altogether.

For example, in our motivating scenario of the gas whooshing to fill a suddenly-larger chamber, we're certainly not interested in the joint probability of all of the particles together; most physical observables (the pressure of the gas, or the average velocity of the gas flow through the opening) can be written in terms of *just* the f_1 – this is another example of our perspective of ultimately aiming for very reduced-dimensionality descriptions of our microscopic states. Perhaps, then, we could take the BBGKY hierarchy and, for a dilute gas, imagine that three-body densities (and three-particle collisions) are much less important than binary collisions. We could attempt to make progress by *truncating* the whole hierarchy, keeping only the full equation for the time evolution of f_1 and a simplified equation for the time evolution of f_2 ⁶⁷. This is precisely the path that we will follow in Sec. C.1 and 5.3, allowing us to derive *directly in the framework of classical mechanics* the existence of an entropy-like quantity.

5.3 The H-Theorem

If you indulged yourself in the derivation of Appendix C.1 you've already seen the Boltzmann equation:

$$\boxed{\begin{aligned} \frac{\partial f_1}{\partial t} - \{\mathcal{H}_1, f_1\} &= \\ &= \int d^3 p_2 d^3 p'_1 d^3 p'_2 \omega(\mathbf{p}'_1, \mathbf{p}'_2 | \mathbf{p}, \mathbf{p}_2) [f_1(\mathbf{q}, \mathbf{p}'_1) f_1(\mathbf{q}, \mathbf{p}'_2) - f_1(\mathbf{q}, \mathbf{p}) f_1(\mathbf{q}, \mathbf{p}_2)], \end{aligned}} \quad (5.44)$$

where the *scattering function* $\omega(\mathbf{p}, \mathbf{p}_2 | \mathbf{p}'_1, \mathbf{p}'_2)$ keeps track of the details of the particle-particle collision process (and hence which is sensitive to all of the microscopic details of the

⁶⁷Leaving us with a still-daunting but now at least approachable set of coupled equations to solve

system); using this function we can write the rate at which collision processes happen as

$$\text{rate} = \omega(\mathbf{p}, \mathbf{p}_2 | \mathbf{p}'_1, \mathbf{p}'_2) f_2(\mathbf{q}_1 = \mathbf{q}, \mathbf{q}_2 = \mathbf{q}, \mathbf{p}_1 = \mathbf{p}, \mathbf{p}_2, t) d^3 p_2 d^3 p'_1 d^3 p'_2. \quad (5.45)$$

That is, the rate is proportional to the details of the scattering process times f_2 , which itself gives us the probability of having particles with (\mathbf{q}, \mathbf{p}) and $(\mathbf{q}, \mathbf{p}_2)$ in the first place.

Shoemakers and tailors⁶⁸, indeed. The Boltzmann equation combines derivatives and integrals and complicated nonlinearities; exact solutions to it are not so easy to come by. But using it we will ultimately be able to show that systems actually *do* reach equilibrium if they start out of it. We have been dancing around the question of how is it that thermodynamics tells us that systems will eventually settle into equilibrium states – which involves an arrow of time that distinguishes past from future – even though the equations of motion are fundamentally invariant under the reversal of time. Specifically, we’ll first show that within the framework of the Boltzmann equation, entropy does indeed increase and systems do indeed equilibrate (this section and the next section, respectively).

This first accomplishment will come via the famous H-Theorem⁶⁹ To state the theorem, let’s first define a (possibly) time-dependent quantity, $H(t)$, as

$$H(t) = \int d^3 q d^3 p f_1(\mathbf{q}, \mathbf{p}, t) \log (f_1(\mathbf{q}, \mathbf{p}, t)). \quad (5.46)$$

For those with a short memory this this might seem like an odd-looking choice; fortunately we encountered similar constructions in Chapter 4 and so suspect that it is actually a surprisingly natural object to consider. Up to a factor of normalization f_1 is a *probability density function* and we will recognize H as something like $\langle \log f_1 \rangle$ – intimately related to the Shannon entropy associated with a the probability function⁷⁰. The H-Theorem is just:

H-Theorem: If f_1 satisfies the Boltzmann equation, then

$$\frac{dH}{dt} \leq 0. \quad (5.47)$$

We have added the frame around this remarkable theorem because it was a stunning early success in showing the power of statistical thinking in understanding thermodynamics, here as an attempted derivation of the second law. These increasingly probabilistic interpretations of thermodynamics eventually led to Gibbs’ ensemble interpretations for the statistical mechanics of general systems.

⁶⁸cf. the quote in Fig. C.1

⁶⁹Historical oddity: I can’t figure out if this “ H ” is a capital roman character or the capital Greek Eta, “ H ,” which in handwriting and all typesetting I know of looks identical. Boltzmann first used E for entropy, and later switched to H . So, probably Eta.

⁷⁰Boltzmann’s ideas on deriving entropy-like functions from physical laws enormously influenced Shannon’s work.

Proof of the H-Theorem: We take the time derivative in the same way we took the time derivative of ensemble average quantities earlier in the chapter: H only has a time dependence through explicit time-dependences:

$$\frac{dH}{dt} = \int d^3q d^3p (\log f_1 + 1) \frac{\partial f_1}{\partial t} = \int d^3q d^3p \log f_1 \frac{\partial f_1}{\partial t}, \quad (5.48)$$

where we exploited the fact that $\int d^3r d^3p f_1 = N$ is independent of time, so $\int \partial_t f_1 = 0$. Using the fact that f_1 satisfies the Boltzmann equation, we can write the above as

$$\frac{dH}{dt} = \int d^3q d^3p \log f_1 \left(\frac{\partial U}{\partial \mathbf{q}_1} \cdot \frac{\partial f_1}{\partial \mathbf{p}_1} - \frac{\mathbf{p}_1}{m} \cdot \frac{\partial f_1}{\partial \mathbf{q}_1} + \left(\frac{\partial f_1}{\partial t} \right)_{coll} \right). \quad (5.49)$$

Actually, though, the first two terms in the Eq. 5.49 vanish: to see this we integrate by parts twice, first moving the derivative from f_1 onto the $\log f_1$ term, and then from the $\log f_1$ term back onto the f_1 . Thus, the time evolution of H is entirely governed by the collision terms:

$$\begin{aligned} \frac{dH}{dt} &= \int d^3q d^3p \log f_1 \left(\frac{\partial f_1}{\partial t} \right)_{coll} \\ &= \int d^3q d^3p_1 d^3p_2 d^3p'_1 d^3p'_2 \omega(\mathbf{p}'_1, \mathbf{p}'_2 | \mathbf{p}_1, \mathbf{p}_2) \log f_1(\mathbf{p}_1) [f_1(\mathbf{p}'_1) f_1(\mathbf{p}'_2) - f_1(\mathbf{p}_1) f_1(\mathbf{p}_2)], \end{aligned} \quad (5.50)$$

where in this expression I've suppressed all of the \mathbf{q} and t arguments, and I've named the dummy variable we're integrating over \mathbf{p}_1 .

To make progress, we play with the dummy indices. First, let's relabel $1 \leftrightarrow 2$, which only changes the argument of the log. Adding the resulting (equivalent) expression and averaging it with the original expression gives a more symmetric expression:

$$\frac{dH}{dt} = \frac{1}{2} \int d^3q d^3p_1 d^3p_2 d^3p'_1 d^3p'_2 \omega(\mathbf{p}'_1, \mathbf{p}'_2 | \mathbf{p}_1, \mathbf{p}_2) \log [f_1(\mathbf{p}_1) f_1(\mathbf{p}_2)] [f_1(\mathbf{p}'_1) f_1(\mathbf{p}'_2) - f_1(\mathbf{p}_1) f_1(\mathbf{p}_2)]. \quad (5.51)$$

We can play the same trick with the incoming and outgoing momenta, swapping $\mathbf{p} \leftrightarrow \mathbf{p}'$, while simultaneously making use of the symmetry properties of the scattering processes⁷¹. This gives us

$$\frac{dH}{dt} = \frac{-1}{2} \int d^3q d^3p_1 d^3p_2 d^3p'_1 d^3p'_2 \omega(\mathbf{p}'_1, \mathbf{p}'_2 | \mathbf{p}_1, \mathbf{p}_2) \log [f_1(\mathbf{p}'_1) f_1(\mathbf{p}'_2)] [f_1(\mathbf{p}'_1) f_1(\mathbf{p}'_2) - f_1(\mathbf{p}_1) f_1(\mathbf{p}_2)]. \quad (5.52)$$

Finally, we average the above two numbered equations to get

$$\begin{aligned} \frac{dH}{dt} &= \frac{-1}{4} \int d^3q d^3p_1 d^3p_2 d^3p'_1 d^3p'_2 \omega(\mathbf{p}'_1, \mathbf{p}'_2 | \mathbf{p}_1, \mathbf{p}_2) \\ &\quad \times [\log [f_1(\mathbf{p}'_1) f_1(\mathbf{p}'_2)] - \log [f_1(\mathbf{p}_1) f_1(\mathbf{p}_2)]] [f_1(\mathbf{p}'_1) f_1(\mathbf{p}'_2) - f_1(\mathbf{p}_1) f_1(\mathbf{p}_2)]. \end{aligned} \quad (5.53)$$

At this point you may be wondering what exactly we're driving at with this sequence of manipulations (i.e., averaging different versions of the same expression to get something

⁷¹ $\int d^3p'_1 d^3p'_2 \omega(p'_1, p'_2 | p_1, p_2) = \int d^3p'_1 d^3p'_2 \omega(p_1, p_2 | p'_1, p'_2)$

symmetric-looking), but we're actually done! Let's think about the terms in the integral. First, the scattering rate, ω , is definitionally a positive quantity. Second, the terms involving f_1 have been manipulated into the form

$$(\log a - \log b)(a - b),$$

which is positive as long as a and b are themselves positive – which they are here, since f_1 has the property of positivity! Thus,

$$\boxed{\frac{dH}{dt} \leq 0 \quad \Leftrightarrow \quad \frac{dS}{dt} \geq 0}$$

A few comments are in order:

1. The arrow of time, again, emerges from the assumption of molecular chaos. If we had decided that the rate of scattering was proportional to f_2 *after* the collision instead of before, and still kept $f_2 \sim f_1 f_1$, we would have found $\frac{dH}{dt} \geq 0$, suggesting entropy *decreases* as we move into the future. Clearly some real subtleties are in the assumptions we made!
2. Note also that the H-theorem permits the time derivative of H to vanish (i.e., it's not a strict inequality). One of the homework problems deals with distributions that satisfy a notion of "local equilibrium" by satisfying a condition of *detailed balance*, making $\frac{dH}{dt}$ vanish by virtue of:

$$f_1(\mathbf{p}'_1) f_1(\mathbf{p}'_2) = f_1(\mathbf{p}_1) f_1(\mathbf{p}_2).$$

These distributions are not quite in equilibrium, as they do not satisfy the streaming terms, but they *do* make the collision terms vanish. These systems have things like densities, temperatures, drift velocities, etc., varying over space. We'll learn more about this in the next section.

5.4 Introduction to hydrodynamics

The equilibrium properties of a macroscopic system are governed by thermodynamics, but we said at the outset of this chapter that we also care about, e.g., the common situation shown in Fig. 5.1. What happens if you start with an equilibrium system and perturb it (perhaps in a large way, as by suddenly and radically expanding the volume available for a gas)? *Hydrodynamics* provides a systematic way to think about characteristically long-wavelength, low-energy excitations of a system. Phenomenologically, one can write down hydrodynamic equations based on the symmetries of a system, but here (in the context of the Boltzmann equation) we'll see that you can also explicitly derive hydrodynamic descriptions by starting with the microscopic dynamics of a system.

To motivate a bit of what follows, let us think about the *equilibrium condition* in the context of the Boltzmann equation, $\frac{dH}{dt} = 0$. This sets up the following tension⁷²: One way

⁷²which we will resolve in this section

to satisfy this condition is to satisfy the constraints of local equilibrium, writing a candidate one-body distribution which takes the form

$$f(\mathbf{p}, \mathbf{r}) = \exp\left(-\alpha(\mathbf{r}) - \beta(\mathbf{r})\frac{(\mathbf{p} - \pi(\mathbf{r}))^2}{2m}\right), \quad (5.54)$$

where α , β , and π are functions of the spatial coordinates. A distribution of this form sets the time-derivative of H to zero, but *it does not* satisfy Boltzmann's equation itself! We saw quite generically when discussing the Liouville equation that the left hand side of the Boltzmann equation, even if there is no explicit time dependence, requires the Poisson bracket of the one-particle distribution and the one-particle Hamiltonian to vanish, $\{H_1, f_1\} = 0$; given the set of conserved quantities this tells us that in global equilibrium f_1 should just be a function of H_1 :

$$f(\mathbf{p}, \mathbf{r}) \propto \exp\left(\beta\left(\frac{p^2}{2m} + U(\mathbf{r})\right)\right). \quad (5.55)$$

These, in general, are not the same. The key is that the Boltzmann equation is built on a separation of time scales that we can physically interpret. At the fastest, collisional time scale, we approximate $f_2 \sim f_1 f_1$, where there are correlations in these quantities. On time scales related to the mean time between collisions, τ_x , f_1 relaxes to a local equilibrium form, and quantities that are conserved in collisions reach this state of local equilibrium. Finally, there is a subsequent slow relaxation to the global equilibrium state, governed not by the collision terms or integrals over the collision terms, but by the streaming terms on the LHS of the Boltzmann equation.

5.4.1 Collision-conserved quantities

Let's think about a function over the single-particle phase space, $A(\mathbf{r}, \mathbf{p})$ – this could be density, or kinetic energy, or... but importantly, we are thinking of quantities that do not have explicit time-dependences. Now, we want to think about the typical way that A varies with space, so we will integrate over momentum⁷³ So, let's define averages of A as

$$\langle A(\mathbf{r}, t) \rangle = \frac{\int d^3p A(\mathbf{r}, \mathbf{p}) f_1(\mathbf{r}, \mathbf{p}, t)}{\int d^3p f_1(\mathbf{r}, \mathbf{p}, t)}.$$

Note that the denominator of the above expression is just a local number density of particles,

$$n(\mathbf{r}, t) = \int d^3p f_1(\mathbf{r}, \mathbf{p}, t), \quad (5.56)$$

so

$$\langle A(\mathbf{r}, t) \rangle = \frac{1}{n(\mathbf{r}, t)} \int d^3p A(\mathbf{r}, \mathbf{p}) f_1(\mathbf{r}, \mathbf{p}, t) \quad (5.57)$$

⁷³This could be motivated by saying it is more common to experimentally measure spatial dependence than momentum dependence in the kind of systems we're studying, but I would rather say something like the following: When we started with the Liouville equation there was complete symmetry between \mathbf{p} and \mathbf{q} , but in the derivation of the Boltzmann equation we started treating the two inequivalently. To emphasize the fact that \mathbf{p} and \mathbf{q} are no longer on the same footing, in this section I'm using the variable \mathbf{r} instead of \mathbf{q} to represent position.

Importantly, the time-dependence of these average quantities only come through the fact that f_1 can evolve in time.

We're not going to be interested in arbitrary choices of A : ultimately we want to think about the kind of slowly-varying quantities that are relevant as we are approaching equilibrium, and we know from our discussion above that typically terms involved in the collision integral will vary over a fast time scale. So, we want quantities A that will vanish when integrated against the collision part of the Boltzmann equation. I.e., we want A 's that have the property

$$\int d^3p A(\mathbf{r}, \mathbf{p}) \left(\frac{\partial f_1}{\partial t} \right)_{coll} = 0.$$

We can insert the expression for the collision term and go through precisely the same kind of manipulations we did in Sec. 5.3 to find that the A 's that have this property obey

$$A(\mathbf{r}, \mathbf{p}_1) + A(\mathbf{r}, \mathbf{p}_2) = A(\mathbf{r}, \mathbf{p}'_1) + A(\mathbf{r}, \mathbf{p}'_2). \quad (5.58)$$

That is, *they are properties that are conserved in the course of a collision.*

Time evolution of collision-conserved quantities: Before we investigate the particular A 's of interest, let's write down the general way that A changes with time if f_1 satisfies the Boltzmann equation. To do this, we start with the Boltzmann equation, which we will from now on occasionally write as

$$\mathcal{L}f_1 = C[f_1, f_1], \quad (5.59)$$

where $\mathcal{L} = \partial_t + \frac{p_\alpha}{m} \partial_\alpha + F_\alpha \partial_{p_\alpha}$ and $C[f_1, f_1] = \left(\frac{\partial f_1}{\partial t} \right)_{coll}$. Let's multiply by a collision-invariant $A(\mathbf{r}, \mathbf{p})$ and integrate both sides $\int d^3p$. The RHS will vanish (by definition of how we're choosing the A), leaving us with

$$\int d^3p A(\mathbf{r}, \mathbf{p}) \left(\frac{\partial}{\partial t} + \frac{p_\alpha}{m} \partial_\alpha + F_\alpha \partial_{p_\alpha} \right) f_1(\mathbf{r}, \mathbf{p}, t) = 0, \quad (5.60)$$

where \mathbf{F} is the external force, $\mathbf{F} = -\nabla U$ (and the rest of the notation should be clear... i.e. with the summation convention $\frac{\mathbf{p}}{m} \cdot \frac{\partial}{\partial \mathbf{r}} = \frac{p_\alpha}{m} \partial_\alpha$). We can simplify this expression (for instance by integrating the term involving the external potential by parts) and various simple manipulations. Making use of how we defined the angle brackets here to be averages over momenta, we can write the above expression as⁷⁴

$$\frac{\partial}{\partial t} \langle nA \rangle + \frac{\partial}{\partial \mathbf{r}} \cdot \langle n\mathbf{v}A \rangle - n \left\langle \mathbf{v} \cdot \frac{\partial A}{\partial \mathbf{r}} \right\rangle - n \left\langle \mathbf{F} \cdot \frac{\partial A}{\partial \mathbf{p}} \right\rangle = 0. \quad (5.61)$$

This general equation becomes quite powerful when we start replacing the generic A with *specific* collision-conserved quantities, as we'll now see.

Specific collision-conserved quantities

Let's apply this general expression to quantities which we *know* are conserved in a collision.

⁷⁴suppressing dependencies, and with $\mathbf{v} = \mathbf{p}/m$

Particle number The most trivial thing conserved in a collision between particles is simply *the number of particles!*. If we insert the choice $A = 1$ into Eq. 5.61, we simply get

$$\partial_t n + \partial_\alpha (n u_\alpha) = 0. \quad (5.62)$$

This result is often written using the particle *current*, $\mathbf{J}(\mathbf{r}, t) = n(\mathbf{r}, t)\mathbf{u}(\mathbf{r}, t)$. It is the *continuity equation*, expressing the fact that if particle number is conserved, variations in local particle density are related to particle currents. Note that we have introduced a new quantity, $\mathbf{u} = \langle v \rangle$

Momentum Linear functions of the momentum are also conserved during collisions, and it is convenient to look not at $\mathbf{A} = \mathbf{p}$, but rather at the momentum relative to the mean local velocity we just defined. Choosing $\mathbf{A} = \mathbf{c} \equiv \mathbf{p}/m - \mathbf{u}$, substituting it into Eq. 5.61, and exploiting the fact that we defined things so that $\langle c_\alpha \rangle = 0$, we find

$$\partial_t u_\alpha + u_\beta \partial_\beta u_\alpha = \frac{F_\alpha}{m} - \frac{1}{mn} \partial_\beta P_{\alpha\beta}, \quad (5.63)$$

where the *pressure tensor* for the fluid is

$$P_{\alpha\beta} \equiv mn \langle c_\alpha c_\beta \rangle. \quad (5.64)$$

This expression is like a generalized Newton's law, telling us that the fluid elements experience accelerations that come both from the external forces and also gradients in the pressure tensor.

Kinetic energy Finally, we look at the kinetic energy of the particles as our last collision-conserved quantity. As before, it is slightly easier to work with the relative kinetic energy, $A = \frac{m}{2}(\mathbf{v} - \mathbf{u})^2 = \frac{m}{2}\mathbf{c}^2$. We substitute this into Eq. 5.61, and go through some simplifications to find

$$\partial_t \epsilon + u_\alpha \partial_\alpha \epsilon = -\frac{1}{n} \partial_\alpha h_\alpha - \frac{1}{n} P_{\alpha\beta} u_{\alpha\beta}, \quad (5.65)$$

where I have just introduced the *average local kinetic energy*:

$$\epsilon \equiv \left\langle \frac{mc^2}{2} \right\rangle,$$

the *local heat flux*:

$$h_\alpha \equiv \frac{nm}{2} \langle c_\alpha c^2 \rangle,$$

and the *strain rate tensor*:

$$u_{\alpha\beta} = \frac{1}{2} (\partial_\alpha u_\beta + \partial_\beta u_\alpha).$$

Equations 5.62, 5.63, and 5.65 form a set of coupled equations for the time evolution of the particle density n the local average velocity \mathbf{u} , and the local average kinetic energy ϵ (which is itself going to be related to the temperature!). But the equations are not closed, as to calculate those three quantities we would need expressions for the pressure tensor and the heat flux. We next show how we can build up approximations, using the Boltzmann equation, for these two quantities to finally get a simple, hydrodynamic description of how systems approach equilibrium.

5.4.2 Zeroth-order hydrodynamics

Let's make progress by straight-up guessing a functional form for the one-body distribution function f_1 . Our logic in making this guess is that we know the collision term will induce fast relaxations, so if we want a distribution function which only varies slowly, a good place to start would be with a distribution function that satisfies $C[f_1, f_1] = 0$. Let's take one such solution which we already met in Eq. 5.54; getting the normalization correct and introducing a new variable that we'll suspiciously label T , we'll take our zeroth-order guess, denoted f_1^0 , to be

$$f_1^0(\mathbf{p}, \mathbf{r}, t) = \frac{n(\mathbf{r}, t)}{(2\pi m k_B T(\mathbf{r}, t))^{3/2}} \exp \left[-\frac{(\mathbf{p} - m\mathbf{u}(\mathbf{r}, t))^2}{2m k_B T(\mathbf{r}, t)} \right] \quad (5.66)$$

This Gaussian form is clearly normalized so that, e.g., $\int d^3p f_1^0 = n$, $\langle \mathbf{p}/m \rangle^0 = \mathbf{u}$, and $\langle c_\alpha c_\beta \rangle^0 = \frac{k_B T}{m} \delta_{\alpha\beta}$. This lets us calculate the approximations for the pressure tensor, energy density, and heat flux:

$$P_{\alpha\beta}^0 = n k_B T \delta_{\alpha\beta}, \quad \epsilon = \frac{3}{2} k_B T, \quad \mathbf{h}^0 = 0.$$

The equations we derived by thinking about collision conserved quantities are very simple in this approximation. Defining the *material derivative*

$$D_t \equiv \partial_t + u_\beta \partial_\beta$$

we get:

$$\begin{aligned} D_t n &= -n \partial_\alpha u_\alpha \\ m D_t u_\alpha &= F_\alpha - \frac{1}{n} \partial_\alpha (n k_B T) \\ D_t T &= -\frac{2}{3} T \partial_\alpha u_\alpha. \end{aligned} \quad (5.67)$$

The inadequacy of zeroth-order hydrodynamics: Sadly, these very simple equations do a terrible job of describing the relaxation of a system to equilibrium. Let's imagine starting with a system with initial $\mathbf{u}_0 = 0$ in the absence of external forces $\mathbf{F} = 0$, and making a small perturbation:

$$n(\mathbf{r}, t) = \bar{n} + \delta n(\mathbf{r}, t), \quad T(\mathbf{r}, t) = \bar{T} + \delta T(\mathbf{r}, t). \quad (5.68)$$

We want to study what happens to these small perturbations, so we expand Eqs. 5.67 to first order in $(\delta n, \delta T, \mathbf{u})$, where we note that to first order the material derivative is just $D_t = \partial_t + \mathcal{O}(u)$, so our linearized zeroth order equations become:

$$\begin{aligned} \partial_t \delta n &= -\bar{n} \partial_\alpha u_\alpha \\ m \partial_t u_\alpha &= -\frac{k_B \bar{T}}{\bar{n}} \partial_\alpha \delta n - k_B \partial_\alpha \delta T \\ \partial_t \delta T &= -\frac{2}{3} \bar{T} \partial_\alpha u_\alpha. \end{aligned} \quad (5.69)$$

The easiest way to investigate the effect of our perturbations is to take Fourier transforms,

$$\tilde{A}(\mathbf{k}, \omega) = \int d^3q dt A(\mathbf{r}, t) \exp [i(\mathbf{k} \cdot \mathbf{r} - \omega t)],$$

where A is any of $(\delta n, \delta T, \mathbf{u})$. This gives us the matrix equation

$$\omega \begin{pmatrix} \tilde{\delta n} \\ \tilde{u}_\alpha \\ \tilde{\delta T} \end{pmatrix} = \begin{pmatrix} 0 & \bar{n}k_\beta & 0 \\ \frac{k_B \bar{T}}{m \bar{n}} \delta_{\alpha\beta} k_\beta & 0 & \frac{k_B}{m} \delta_{\alpha\beta} k_\beta \\ 0 & \frac{2}{3} \bar{T} k_\beta & 0 \end{pmatrix} \begin{pmatrix} \tilde{\delta n} \\ \tilde{u}_\beta \\ \tilde{\delta T} \end{pmatrix}. \quad (5.70)$$

This matrix has the following modes: There are two modes describing a transverse shear flow in a uniform and isothermal gas ($n = \bar{n}, T = \bar{T}$), with the velocity field varying transverse to its orientation, e.g. $\mathbf{u} = f(x, t)\hat{y}$, and both have $\omega = 0$. There is another $\omega = 0$ mode describing a gas with uniform pressure $nk_B T$, where n and T may be spatially varying but with a constant product. Lastly, there is a mode with variations along the direction of \mathbf{k} ; the eigenvector looks like:

$$\mathbf{v}_{longitudinal} = \begin{pmatrix} \bar{n}|\mathbf{k}| \\ \omega(\mathbf{k}) \\ \frac{2}{3}\bar{T}|\mathbf{k}| \end{pmatrix}, \quad \omega(\mathbf{k}) = \pm \sqrt{\frac{5k_B \bar{T}}{3m}} |\mathbf{k}|.$$

Well, shoot. Apparently within this approximation *none* of our conserved quantities relax to equilibrium if we apply a little perturbation: shear flow persists forever, sound modes have undamped oscillations, etc.

5.4.3 First-order hydrodynamics

Perhaps this should not have surprised us: we picked something that satisfied local equilibrium, but it is straightforward to check that $\mathcal{L}f_1^0 \neq 0$. Actually, it's more straightforward to show that if n, T, u_α satisfy the zeroth-order hydrodynamic equations, then the effect of \mathcal{L} on the log of the zeroth-order approximation for f_1 is

$$\mathcal{L} [\log f_1^0] = \frac{m}{k_B T} \left(c_\alpha c_\beta - \frac{\delta_{\alpha\beta}}{3} c^2 \right) u_{\alpha\beta} + \left(\frac{mc^2}{2k_B T} - \frac{5}{2} \right) \frac{c_\alpha}{T} \partial_\alpha T. \quad (5.71)$$

Our instinct for moving forward is to note that, although the RHS of the above is not zero, it depends on gradients of temperature and on the local velocity. In a sense, then, if we stick to long-wavelength variations in T and \mathbf{u} we are “close” to a solution. Thus, we will try to construct first-order hydrodynamics by adding a little something extra to the distribution:

$$f_1^1 = f_1^0 + \delta f_1. \quad (5.72)$$

Relaxation time approximation What happens if we act on f_1^1 with our collision operator?

$$\begin{aligned} C[f_1^1, f_1^1] &= \int d^3p_2 d^3p'_1 d^3p'_2 \omega(\mathbf{p}'_1, \mathbf{p}'_2 | \mathbf{p}_1, \mathbf{p}_2) [f_1(\mathbf{p}'_1) f_1(\mathbf{p}'_2) - f_1(\mathbf{p}_1) f_1(\mathbf{p}_2)] \\ &= \int d^3p_2 d^3p'_1 d^3p'_2 \omega(\mathbf{p}'_1, \mathbf{p}'_2 | \mathbf{p}_1, \mathbf{p}_2) [f_1^0(\mathbf{p}'_1) \delta f_1(\mathbf{p}'_2) + \delta f_1(\mathbf{p}'_1) f_1^0(\mathbf{p}'_2) - f_1^0(\mathbf{p}_1) \delta f_1(\mathbf{p}_2) - \delta f_1(\mathbf{p}_1) f_1^0(\mathbf{p}_2)] \end{aligned}$$

where we have used the fact that f_1^0 vanishes in the collision integral and dropped any terms of order $(\delta f_1)^2$. We now have a linear function of δf_1 , albeit a messy one to work with. At this point there is a proper way to proceed⁷⁵, and a physically just fine way to proceed, which has the great virtue of being much easier while still capturing the dissipational piece that was missing from our zeroth-order description. We simply approximate

$$C[f_1^1, f_1^1] \approx \frac{-\delta f_1}{\tau_x}, \quad (5.74)$$

which is called the *relaxation time approximation* or the *single collision time approximation* or using the *BGK operator*⁷⁶. With this approximation the Boltzmann equation is

$$\frac{\partial f_1^0 + \delta f_1}{\partial t} - \{H_1, f_1^0 + \delta f_1\} = \frac{-\delta f_1}{\tau_x}, \quad (5.75)$$

but we assume that $\delta f_1 \ll f_1^0$, so we ignore the δf_1 on the LHS. We can then explicitly work out the additional contribution to our improved estimate:

$$\delta f_1 = -\tau_x \left[\left(\frac{mc^2}{2k_B T} - \frac{5}{2} \right) \frac{\mathbf{c}}{T} \cdot \frac{\partial T}{\partial \mathbf{r}} + \frac{m}{k_B T} \left(c_\alpha c_\beta - \frac{\delta_{\alpha\beta}}{3} c^2 \right) u_{\alpha\beta} \right] f_1^0. \quad (5.76)$$

One can use this improved description to calculate corrections to various quantities. For example, the first-order estimate of the pressure tensor becomes

$$P_{\alpha\beta}^1 = nk_B T \delta_{\alpha\beta} - 2nk_B T \tau_x \left(u_{\alpha\beta} - \frac{\delta_{\alpha\beta}}{3} u_{\gamma\gamma} \right), \quad (5.77)$$

and the heat flux acquires a dependence on spatial gradients in the temperature. These are important: they say that shear flows get opposed by off-diagonal terms in the pressure tensor, and that spatial variations in temperature generate heat flows that in turn smooth out those variations! These are the sorts of effects that cause the relaxation to equilibrium.

In case you're curious, if I've T_EX'ed this correctly the matrix equation corresponding to the first-order hydrodynamic equations after a Fourier transformation look like

$$\omega \begin{pmatrix} \tilde{\delta n} \\ \tilde{u}_\alpha \\ \tilde{\delta T} \end{pmatrix} = \begin{pmatrix} 0 & \bar{n} \delta_{\alpha\beta} k_\beta & 0 \\ \frac{k_B \bar{T}}{m \bar{n}} \delta_{\alpha\beta} k_\beta & -i \frac{\mu}{m \bar{n}} \left(k^2 \delta_{\alpha\beta} + \frac{k_\alpha k_\beta}{3} \right) & \frac{k_B}{m} \delta_{\alpha\beta} k_\beta \\ 0 & \frac{2}{3} \bar{T} \delta_{\alpha\beta} k_\beta & -i \frac{2K k^2}{3k_B \bar{n}} \end{pmatrix} \begin{pmatrix} \tilde{\delta n} \\ \tilde{u}_\beta \\ \tilde{\delta T} \end{pmatrix}, \quad (5.78)$$

where $K = (5\bar{n}k_B^2 \bar{T} \tau_x)/(2m)$ and $\mu = \bar{n}k_B \bar{T} \tau_x$ is a viscosity coefficient. The important point of writing this is the ability to verify that now all of the modes have an imaginary component (either they are strictly imaginary eigenvalues (for variations in pressure and for transverse shear modes) or complex ones (for longitudinal sound modes), so that we know that over long time scales perturbations to the gas die away, and the gas eventually reaches its equilibrium state.

⁷⁵The Chapman-Enskog expansion, doing a careful expansion in δf_1

⁷⁶Bhatnagar-Gross-Krook, [51]

5.5 Problems

5.5.1 Detailed balance

Let's consider functions which satisfy the Boltzmann equation, and for simplicity let's assume that the external potential $U(\mathbf{q}) = 0$. As we saw in class, one way to make the collision integral vanish is by ensuring that f_1 satisfies the *detailed balance* condition:

$$f_1^{eq}(\mathbf{r}, \mathbf{p}'_1) f_1^{eq}(\mathbf{r}, \mathbf{p}'_2) = f_1^{eq}(\mathbf{r}, \mathbf{p}_1) f_1^{eq}(\mathbf{r}, \mathbf{p}_2),$$

where the challenge is to figure out how to guarantee this equality for all momenta.

The Boltzmann distribution:

By taking the log of the detailed balance condition, argue based on your knowledge of conserved quantities during collisions that f_1^{eq} takes the form of the Maxwell-Boltzmann distribution. (i.e., interestingly, having the collision term in the Liouville equation makes the f_1 sit in the Boltzmann distribution at equilibrium.).

5.5.2 Momentum moments

Let's consider a gas of N particles of mass m , in thermal equilibrium at temperature T in a box of volume V .

- (a) Write down the appropriate equilibrium one-particle density, $f_{eq}(\mathbf{q}, \mathbf{p})$
- (b) What is the joint characteristic function, $\langle \exp(-i\mathbf{k} \cdot \mathbf{p}) \rangle$?
- (c) Calculate the joint cumulants, $\langle p_x^l p_y^m p_z^n \rangle_c$, for any integers l, m, n .
- (d) Calculate the joint moment, $\langle p_\alpha p_\beta \mathbf{p} \cdot \mathbf{p} \rangle$, for any choice of cartesian directions α, β (hint: the answer is given in the next question... just show me how to get there).

5.5.3 Local equilibrium

If we ignore the “streaming term” $\{H_1, f_1\}$ then there is a much larger class of distributions that satisfy detailed balance, and they are said to be in *local equilibrium*; they take the form of a distribution with a Boltzmann distribution, but where the number density, temperature, and velocity all become functions of the spatial coordinates⁷⁷.

⁷⁷Take, for instance, a glass of water with a ice cube melting in it. This is clearly not an equilibrium state, but if you look at the distribution of velocities of water molecules as a function of distance from the ice cube, you will find to a good approximation that at every point in space and at any moment in time the distribution is close to Maxwell-Boltzmann

Let's use the Boltzmann formalism to calculate the thermal conductivity of a dilute gas. Suppose we have such a gas between two plates, which are a distance h apart. The first plate (at $z = 0$) is fixed at temperature T_1 , and the second plate (at $z = h$) is fixed at temperature T_2 . There is no net drift velocity of the gas (i.e., there are no net flows), and so our zeroth-order approximation for the one-particle density is

$$f_1^0(\mathbf{p}, x, y, z) = \frac{n(z)}{(2\pi m k_B T(z))^{3/2}} \exp\left[-\frac{\mathbf{p} \cdot \mathbf{p}}{2m k_B T(z)}\right],$$

where the superscript 0 indicates that this is the zeroth-order approximation.

Local pressure

If the gas initially has no drift velocity, what is needed for a relationship between $n(z)$ and $T(z)$ to ensure that the gas velocity remains zero? You will need this relationship in later parts.

First order approximation

The zeroth-order approximation we started with does not allow for the relaxation of density and temperature variations. Let's linearize the Boltzmann equation in the single-collision-time approximation as

$$\mathcal{L}[f] \approx \left[\frac{\partial}{\partial t} + \frac{p_z}{m} \frac{\partial}{\partial z} \right] f_1^0 \approx -\frac{f_1^1 - f_1^0}{\tau_K},$$

where τ_K is again of the order of the typical time between collisions. What is your improved approximation for f_1^1 , the first-order approximate solution to the Boltzmann equation (assuming it is independent of time)?

Heat transfer

What is the z -component of the heat transfer vector, h_z , defined as

$$h_z = n \left\langle c_z \frac{mc^2}{2} \right\rangle,$$

where $\mathbf{c} = \mathbf{p}/m - \mathbf{u}$? Give the answer using both the zeroth order approximation, f_1^0 , and using the first order approximation you derived, f_1^1 . It may help you to know that

$$\langle p^2 \rangle^0 = 3mk_B T, \quad \langle p^4 \rangle^0 = 15(mk_B T)^2, \quad \langle p_z^2 p^4 \rangle^0 = 35(mk_B T)^3,$$

where $\langle A \rangle^0$ means the local average using f_1^0 . Your answer should include a spatial derivative of the temperature.

Thermal conductivity and temperature profile

From the connection between thermal conductivity and temperature gradient, $\mathbf{h} = -K\nabla T$, and your answer above, what is your first-order approximation for the thermal conductivity of a dilute gas? What is the temperature profile in the steady state, $T(z)$?

5.5.4 BBGKY for a plasma

Anatoly Vlasov wanted to understand plasmas consisting of charged particles that had long-ranged (Coulomb) interactions, and the usual Boltzmann approach based on pair collisions seemed problematic⁷⁸. Let's make use of the BBGKY hierarchy to derive one of Vlasov's celebrated equations.

(a) Start by assuming that the full N -body density is just a product of the 1-body densities ($\rho = \prod_{i=1}^N \rho_1(\mathbf{p}_i, \mathbf{q}_i, t)$) Write down the definition of f_s , and calculate its normalization in the limit $s \ll N$.

(b) Assume, like Vlasov did, that the problem comes from the collisional terms in the BBGKY hierarchy, i.e., the terms that are like $\frac{\partial \Phi}{\partial \mathbf{q}} \cdot \frac{\partial}{\partial \mathbf{p}}$. Write down a version of the BBGKY hierarchy (which you can write as an equation for the time evolution of f_s) in which all of the collisional terms are set to zero. Assume you are looking at a level of the hierarchy $s \ll N$ to make any simplifications to your expression. You've derived the Vlasov equation!

(c) Finally, consider a box of N particles, with no additional external potential. Show that

$$f_1(\mathbf{q}, \mathbf{p}) = \frac{1}{V} g(\mathbf{p})$$

is a stationary solution to the Vlasov equation (that is: you can take *any* function of only the momenta of the particle, and it will be a stationary solution). Said another way: momentum fluctuations never relax to equilibrium in this framework. Why don't they?

⁷⁸Among other things, the long-ranged nature of the interaction potential led to divergences in some of the kinetic terms

Chapter 6

Classical Statistical Mechanics

Statistical mechanics is connected with the phenomenological, “thermodynamical” view of macroscopic properties we saw in the first part of this course via a *probabilistic description of large numbers of degrees of freedom*. In this chapter we will focus not on microscopic theories for which we can study both equilibrium and approach-to-equilibrium dynamics as in Chapter 2, but rather on attempting to provide probability distributions that connect microstates to macrostates. We will use the idea of unbiased estimates of probability to assign these probability distributions for different equilibrium ensembles, and use the mathematics of the large- N limit to show that the ensembles are equivalent in the thermodynamic limit.

6.1 The microcanonical ensemble and the laws of thermodynamics

We begin, just as we did in the chapter on Thermodynamics, with a simplified version of our system of interest, taking an adiabatically isolated state. In the absence of adding heat or work to the system, the *macrostate* M is specified completely by the internal energy E , the set of generalized coordinates \mathbf{x} , and the number of particles N : $M(E, \mathbf{x}, N)$. In the absence of any other information, or any knowledge of other conserved quantities, we say that at a minimum the Hamiltonian evolution equations conserve the total energy of the system, so that the evaluation of the Hamiltonian on a microstate μ is $\mathcal{H}(\mu) = E$.

The central postulate of statistical mechanics is that the equilibrium probability distribution is

$$p_{(E, \mathbf{x}, N)}(\mu) = \frac{1}{\Omega(E, \mathbf{x}, N)} \cdot \begin{cases} 1 & \text{if } \mathcal{H}(\mu) = E \\ 0 & \text{otherwise} \end{cases} \quad (6.1)$$

This is often called “the assumption of equal a priori probabilities,” and we see that it is the same as the unbiased estimate of probability given only a constraint of constant energy E . Certainly we saw that this is one of the allowed assignments consistent with Liouville’s theorem, although it is not the only one! Given our work in the chapter on Probability we might not be surprised by this assignment, but it is nevertheless a deep assumption.

There is a potential subtlety in defining the normalization factor⁷⁹ Ω here: as written, in order to make p a properly normalized probability density function (so that it integrates to unity), we want Ω to be the area of the relevant surface of fixed energy E over the microscopic phase space. You might be worried about defining probability densities that are non-zero only on a surface, and so we sometimes *define* a microstate in the microcanonical ensemble to be within Δ of the specified energy: $E - \frac{1}{2}\Delta \leq \mathcal{H}(\mu) \leq E + \frac{1}{2}\Delta$. The normalization Ω' is now the volume of a shell rather than the area of a surface, and $\Omega' \approx \Omega\Delta$. We will see that, since Ω is typically exponentially large in E , which is itself typically proportional to N , that the difference between the surface area and the shell volume is negligible, so we'll go back and forth between Ω and Ω' freely.

We define the entropy of the uniform probability distribution exactly as you by now expect:

$$S(E, \mathbf{x}, N) = k_B \log \Omega(E, \mathbf{x}, N), \quad (6.2)$$

where we have introduced a factor called “ k_B ” so that entropy has units of energy divided by temperature. Note, by the way, that we know from Liouville that under *canonical transformations* volumes in phase space are invariant, and for the transformed probability distribution stays uniform on the transformed phase-space surface of constant energy. This tells us that both Ω and S are invariant under canonical coordinate changes.

To highlight the interconnections with the beginning of this class, we now show that Eq. 6.1 can be used to derive the laws of thermodynamics (with the exception of Nernst's theorem, which as we hinted at the time requires quantum mechanics).

6.1.1 0th Law

Let's think about bringing two previously isolated microcanonical systems, which originally had energies E_1 and E_2 , into contact in a way that lets them exchange *heat*, but not *work*. Certainly the combined system has energy $E = E_1 + E_2$, and we assume (by assuming that the interactions between the systems are small) that the microstate of the combined system corresponds to a pair of microstates of the components. We'll write this as $\mu = \mu_1 \otimes \mu_2$, assuming $\mathcal{H}(\mu) = \mathcal{H}_1(\mu_1) + \mathcal{H}_2(\mu_2)$. We thus write the fundamental postulate, Eq. 6.1 as

$$p_E(\mu) = \frac{1}{\Omega(E)} \cdot \begin{cases} 1 & \text{if } \mathcal{H}_1(\mu_1) + \mathcal{H}_2(\mu_2) = E \\ 0 & \text{otherwise} \end{cases}. \quad (6.3)$$

We have a fixed total energy, so we can compute the normalization factor as

$$\Omega(E) = \int dE_1 \Omega_1(E_1)\Omega_2(E - E_1) = \int dE_1 \exp \left[\frac{S_1(E_1) + S_2(E_2)}{k_B} \right]. \quad (6.4)$$

We have written the normalization factor for our probability this way to make contact with our “sums and integrals of exponentials” discussion in a previous chapter! We think (and will see later) that entropy is extensive, so that S_1 and S_2 are proportional to the number

⁷⁹We must learn how to count the number of states it is possible for a system to have or, more precisely, how to avoid having to count that number” – David Goodstein [52]

of particles in the system. We use the simple (i.e., non-complex) saddle-point method to approximate the integral by the maximum of the integrand, giving us

$$S(E) = k_B \log \Omega(E) \approx S_1(E_1^*) + S_2(E_2^*), \quad (6.5)$$

where the starred quantities are the values that maximize the value of the integrand. We can calculate these values by setting the first derivative (w/r/t E_1) of the exponent above to zero, which gives us:

$$\left. \frac{\partial S_1}{\partial E_1} \right|_{\mathbf{x}_1, N_1} = \left. \frac{\partial S_2}{\partial E_2} \right|_{\mathbf{x}_2, N_2}. \quad (6.6)$$

In words, there are many joint microstates which sum to a total energy of E , but there are exponentially *more* microstates sitting near (E_1^*, E_2^*) , and so the system eventually flows from (E_1, E_2) to (E_1^*, E_2^*) . We get no information about the time dynamics of this process; just the end result, which is a place where Eq. 6.6 is satisfied.

Also... ah ha!! We have found a condition satisfied by two systems that come into equilibrium with each other: they have equal partial derivatives of entropy with respect to energy (holding everything else fixed)! To be consistent with what we had from phenomenological thermodynamics, we'll choose a particular name for the state function corresponding to this condition:

$$\left. \frac{\partial S}{\partial E} \right|_{\mathbf{x}} = \frac{1}{T}. \quad (6.7)$$

6.1.2 1st Law

Let's now bring together two previously isolated microcanonical systems and separate them by an insulated partition that lets them exchange generalized displacements – and thus do work on each other – but that does not permit direct heat exchange (For instance, think of an insulated partition between two gases that can slide back and forth). We let the two systems equilibrate, and then ask what happens to the entropy of the first system when there is a spontaneous fluctuation by some amount⁸⁰ by $\delta \mathbf{x}$. This fluctuation both changes the extensive variable \mathbf{x} and does work on the system $\delta W = \mathbf{J} \cdot \delta \mathbf{x}$; since the system is otherwise isolated its internal energy changes to $E + \delta W$. To first order, the entropy $S(E, \mathbf{x})$ of the system varies by

$$\delta S = S(E + \mathbf{J} \cdot \delta \mathbf{x}, \mathbf{x} + \delta \mathbf{x}) - S(E, \mathbf{x}) = \left(\left. \frac{\partial S}{\partial E} \right|_{\mathbf{x}} \mathbf{J} + \left. \frac{\partial S}{\partial \mathbf{x}} \right|_E \right) \cdot \delta \mathbf{x}. \quad (6.8)$$

Now, what does it mean to be in equilibrium? Even though all consistent microstates are equally probable, we expect (analogous to the arguments above) that there are exponentially many more consistent states sitting near the extrema of the entropy function.

So, we expect that when we let the systems equilibrate then the first system is at a value of E and \mathbf{x} . We expect that the counting factor.... double check argument in my lecture notes

⁸⁰for the remainder of this section I'm going to be T_EX-lazy and stop distinguishing mechanical work from chemical work.... it's all just \mathbf{x} now.

We say we have an equilibrium state at some value of E and some set of generalized displacements \mathbf{x} , and we have said that all consistent microstates are equally probable. Well, the above equation says that spontaneous changes in the system will occur, taking us into more likely states, *unless the terms in the parentheses vanish!* Thus, a condition to be in equilibrium is that

$$\left. \frac{\partial S}{\partial x_i} \right|_{E, x_{j \neq i}} = -\frac{J_i}{T}, \quad (6.9)$$

where we have used the relationship between temperature and $\partial S/\partial E$ from the zeroth law. With this constraint on the variations, we get that in general

$$dS(E, \mathbf{x}) = \frac{\partial S}{\partial E} dE + \frac{\partial S}{\partial \mathbf{x}} d\mathbf{x} = \frac{dE}{T} - \frac{\mathbf{J} \cdot d\mathbf{x}}{T} \Rightarrow dE = TdS + \mathbf{J} \cdot d\mathbf{x}. \quad (6.10)$$

Wow – it’s the first law! And where we’ve, of course, identified $\delta Q = TdS$.

6.1.3 2nd Law

The second law⁸¹ is almost obvious, almost by construction, given what we have set up so far. Indeed, you already know from the section of unbiased estimates that we are assigning probabilities in a way that *maximizes the entropy* subject to the constraint on the energy. For instance, consider our example from the zeroth law of bringing two equilibrium states at E_1 and E_2 into contact. Well, clearly

$$S(E) \equiv k_B \log \Omega(E_1 + E_2) \geq S_1(E_1) + S_2(E_2),$$

which must be true since the two states of the original system are a *subset of the possible combined joint microstates*.

Note that we can make additional mathematical statements by considering variations of entropy. When two systems are first brought into contact but have not yet reached equilibrium, the equality in Eq. 6.6 does not yet hold. Instead we have

$$\delta S = \left(\left. \frac{\partial S_1}{\partial E_1} \right|_{\mathbf{x}_1} - \left. \frac{\partial S_2}{\partial E_2} \right|_{\mathbf{x}_2} \right) \delta E_1 = \left(\frac{1}{T_1} - \frac{1}{T_2} \right) \delta E_1 \geq 0. \quad (6.11)$$

Thus we recover Clausius’ statement of the second law: we know the variations in S are positive as we move towards a new equilibrium state, and we see that heat flows from the hotter to the colder system.

Note that, in principle, these are all probabilistic statements: it is merely *much more likely* that a combined system ends up at (E_1^*, E_2^*) rather than its initial and (E_1, E_2) . This obscures just how much work “much” is doing in the previous sentence, though: the number of

⁸¹“The law that entropy always increases holds, I think, the supreme position among the laws of Nature. If someone points out to you that your pet theory of the universe is in disagreement with Maxwell’s equations - then so much the worse for Maxwell’s equations. If it is found to be contradicted by observation - well, these experimentalists do bungle things sometimes. But if your theory is found to be against the Second Law of Thermodynamics I can give you no hope; there is nothing for it to collapse in deepest humiliation.” – Arthur Eddington [53]

microstates available grows exponentially with system size, and so if we were to ask how long we would have to wait before seeing our combined system at (E_1, E_2) rather than (E_1^*, E_2^*) , the answer would be related to an exponential of that exponential. Needing to wait this long to see something is a practical way of saying “you will never see that.”

6.1.4 The ideal gas in the microcanonical ensemble

We return to our favorite toy system for illustrating concepts, the ideal gas of N particles. We ignore any particle interactions, and say we have

$$\mathcal{H} = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m} + U(\mathbf{q}_i),$$

where the potential U simply imposes strict confinement to a box of volume V (i.e., we take $U(\mathbf{q}_i) = 0$ if particle i is inside the box and ∞ if it is outside the box). Explicitly, the microcanonical ensemble has a probability density function

$$p(\mu) = \frac{1}{\Omega(E, V, N)} \cdot \begin{cases} 1 & \text{if } \sum_i (\mathbf{p}_i^2 / (2m)) = E \pm \frac{1}{2} \Delta_E \text{ and } \mathbf{q}_i \in \text{box} \\ 0 & \text{otherwise} \end{cases}. \quad (6.12)$$

We can calculate Ω by the requirement that $p(\mu)$ is properly normalized, i.e., $\int \prod_i dV_i p(\mu) = 1$. The integrals over \mathbf{q}_i are all trivial: each of those N integrals just gives a factor of V . The integral over the momenta is down by noting we are constraining the momenta to a (finite-thickness shell around a) surface of a hypersphere, given by $\sum_i^N \mathbf{p}_i^2 = 2mE$. We thus need to know the area of a $3N$ -dimensional sphere of radius $R = \sqrt{2mE}$. The relevant formula for a d -dimensional sphere is

$$\mathcal{A}_d = \frac{2\pi^{d/2}}{(d/2 - 1)!} R^{d-1}, \quad (6.13)$$

where it is easy to check that this gives the right result in $2D$ (and $3D$, since $(1/2)! = \sqrt{\pi}/2$). Putting this together, with $d = 3N$, we calculate:

$$\Omega(E, V, N) = V^N \frac{2\pi^{3N/2}}{(3N/2 - 1)!} (2mE)^{(3N-1)/2} \Delta_E \quad (6.14)$$

The entropy of the ideal gas is then just the log of this normalization factor. Using Stirling’s approximation:

$$\begin{aligned} S(E, V, N) &= k_B \left[N \log V + \frac{3N}{2} \log(2\pi mE) - \frac{3N}{2} \log \frac{3N}{2} + \frac{3N}{2} \right] \\ &= Nk_B \log \left[V \left(\frac{4\pi emE}{3N} \right)^{3/2} \right], \end{aligned} \quad (6.15)$$

where in the first line we have dropped terms of order 1 and of order $\log E \propto \log N$ by arguing the latter are small compared to these terms of order N in the thermodynamic limit.

With the entropy in hand, and writing $dS = \frac{1}{T}dE + \frac{P}{T}dV - \frac{\mu}{T}dN$ we can get the usual properties of the ideal gas by differentiating the entropy as appropriate. For instance,

$$\frac{1}{T} = \left. \frac{\partial S}{\partial E} \right|_{N,V} = \frac{3}{2} \frac{Nk_B}{E} \Rightarrow E = \frac{3}{2} Nk_B T,$$

the usual equipartition result in the absence of a potential. Similarly,

$$\frac{P}{T} = \left. \frac{\partial S}{\partial V} \right|_{N,E} = \frac{Nk_B}{V} \Rightarrow PV = Nk_B T,$$

the ideal gas equation of state.

As a final example of a simple calculation, what about the probability distribution for finding a particle with some momentum vector \mathbf{p}_1 ? Well, we can calculate the unconditional PDF by integrating out everything else:

$$\begin{aligned} p(\mathbf{p}_1) &= \int d^3\mathbf{q}_1 d\Gamma_2^N p(\mu) \\ &= V \frac{\Omega(E - \frac{\mathbf{p}_1^2}{2m}, V, N - 1)}{\Omega(E, V, N)}. \end{aligned} \quad (6.16)$$

One can explicitly check (by plugging in the result of Eq. 6.14, using Stirling's formula, and noting that \mathbf{p}_1^2/E is a number of order $1/N$) that this explicitly gives you a Maxwell-Boltzmann distribution for the momentum:

$$p(\mathbf{p}_1) = \frac{1}{(2\pi mk_B T)^{3/2}} \exp\left(\frac{-\mathbf{p}_1^2}{2mk_B T}\right).$$

Show one more line above, then foreshadow this calculation v/v the reservoir stuff we're about to do in the canonical ensemble.

A comment about the form of the Maxwell-Boltzmann distribution We derived the Maxwell-Boltzmann distribution result for the velocities in a dilute gas of non-interacting particles, but it is actually much more general, and Maxwell's original derivation relies on nothing by rotational invariance. It's cool, so I reproduce it here: Consider the distribution of velocities in, say, the x direction, and call that distribution $p(v_x)$. By rotational symmetry we have the same distribution in the y and z directions. Rotational symmetry *also* guarantees that the full distribution cannot depend on the particular direction for the momentum, but only on the speed $c = \sqrt{\mathbf{v} \cdot \mathbf{v}}$. So, we want functions $p_c(c)$ and $p(v_x)$ that satisfy

$$p_c(c)dv_x dv_y dv_z = p(v_x)p(v_y)p(v_z)dv_x dv_y dv_z.$$

Remarkably, there is only one solution which satisfies this, and it is

$$p(v_x) = A \exp(-Bv_x^2),$$

for some constants A and B . Thus, the distribution of speeds must be

$$p_c(c)dv_x dv_y dv_z = 4\pi c^2 p_c(c)dc = 4\pi A^3 c^2 e^{-Bc^2} dc.$$

Nifty. Determining that the coefficients have specific values, like $B = \frac{m}{2k_B T}$ as in the ideal gas, requires a harder microscopic calculation of the sort we just did.

6.1.5 Gibbs' Paradox: What's up with mixing entropy?

You will notice that Eq. 6.15, giving the entropy of an ideal gas, has a major flaw: we expect entropy to be extensive, but under a transformation $(E, V, N) \rightarrow (\lambda E, \lambda V, \lambda N)$ the entropy actually changes to $\lambda(S + Nk_B \log \lambda)$ rather than λS . There are these extra contributions that come from the integration over positions, like V^N , and this additional contribution is related to the entropy of mixing distinct gases. Gibbs' paradox, the fact that this expression for the entropy suggests an increase even if two *identical* gases are mixed, is subtle, with some to this day arguing that its resolution must be quantum mechanical⁸² and others arguing that the paradox is toothless and can be resolved classically⁸³. We will come back to this issue when we talk about quantum statistical mechanics.

As our yet-to-be-understood resolution to the paradox, we will from now on modify our phase-space measure for identical particles to be

$$d\Gamma = \frac{1}{h^{3N} N!} \prod_{i=1}^N d^3 \mathbf{q}_i d^3 \mathbf{p}_i. \quad (6.17)$$

6.2 The canonical ensemble

In the last section we defined the microcanonical ensemble by considering a macrostate with a specified energy, E , and we were able to derive an expression for the temperature: $T^{-1} = \left. \frac{\partial S}{\partial E} \right|_{\mathbf{x}}$. In the chapter on thermodynamics, though, we saw that E and T are both functions of state; because of this we think it should be possible to reverse this chain, *starting* with a statistical description in which the temperature of a macrostate is prescribed and an expression for the energy is derived.

Welcome to the canonical ensemble⁸⁴, in which we specify the macrostate $M(T, \mathbf{x})$. We think of prescribing the temperature of the system by putting it in thermal contact with *the reservoir*, which is another macroscopic system that is so large that its own temperature does not change as a result of interacting with the system of interest. Intuitively, this should feel like a reasonable possibility: if we imagine tossing a warm pebble into the ocean, the

⁸²see, e.g., Kardar [3]

⁸³see Jayne's discussion [54], or that of Frenkel's more recent article [55]

⁸⁴"We consider especially ensembles of systems in which the index (or logarithm) of probability of phase is a linear function of the energy. This distribution, on account of its unique importance in the theory of statistical equilibrium, I have ventured to call canonical..." J. W. Gibbs [56]

temperature of the pebble will surely equilibrate with the temperature of the ocean, and no thermometer reading the temperature of the ocean will notice the difference.

The goal of the statistical mechanical formulation of thermodynamics is to write down microstate probabilities that we can associate with a given macrostate, $p_M(\mu)$ (and from there derive thermodynamic consequences). You may have expected that our strategy would be to assume a form for this probability and derive consequences – in fact, some treatments follow the kind of reasoning we did above (and in Sec. 4.5.3) for each different ensemble, assigning probabilities by maximizing entropy subject to the relevant known features of the system.

Here, instead, we rely on the central postulate of statistical mechanics, Eq. 6.1, and *derive more consequences from it!* Let's define the system to be in microstate μ_S with energy $\mathcal{H}_S(\mu_S)$ and the reservoir to be in microstate μ_R with energy $\mathcal{H}_R(\mu_R)$, satisfying $E_{total} = \mathcal{H}_S(\mu_S) + \mathcal{H}_R(\mu_R)$. The joint probability of the microstates $\mu_S \oplus \mu_R$ is then assumed to be

$$p(\mu_S \oplus \mu_R) = \frac{1}{\Omega_{S \oplus R}(E_{total})} \cdot \begin{cases} 1 & \text{if } \mathcal{H}_S(\mu_S) + \mathcal{H}_R(\mu_R) = E_{total} \\ 0 & \text{otherwise} \end{cases} \quad (6.18)$$

Ultimately, though, we are not interested in this microcanonical joint-probability distribution for the combined system, we want the unconditional PDF for the microstates. But this we can obtain (as expected) by summing the joint PDF over the microstates of the reservoir:

$$p(\mu) = \sum_{\mu_R} p(\mu_S \otimes \mu_R) \quad (6.19)$$

We make progress, here, by doing something similar to how we calculated $p(\mathbf{p}_1)$ in the microcanonical ensemble in Eq. 6.16: we say that by looking at a specific system microstate μ , the sum over reservoir states in the above equation is restricted to run over microstates with $\mathcal{H}_R(\mu_R) = E_{total} - \mathcal{H}_S(\mu_S)$:

$$p(\mu_S \otimes \mu_R) = \frac{\Omega_R(E_{total} - \mathcal{H}_S(\mu_S))}{\Omega_{S \oplus R}(E_{total})}. \quad (6.20)$$

Let's focus on the numerator for now (noting that the denominator, which just provides an overall normalization, can always be effectively recovered by imposing a normalization condition of $p(\mu)$ when we're done). Well, Ω_R is related to the entropy of the reservoir:

$$\frac{\Omega_R(E_{total} - \mathcal{H}_S(\mu_S))}{\Omega_{S \oplus R}(E_{total})} \propto \exp\left(\frac{S_R(E_{total} - \mathcal{H}_S(\mu_S))}{k_B}\right). \quad (6.21)$$

Let's write that entropy expression as

$$S_R\left(E_{total} \left(1 - \frac{\mathcal{H}_s(\mu_S)}{E_{total}}\right)\right) \approx S_R(E_{total}) - \mathcal{H}_s(\mu_S) \frac{\partial S_R}{\partial E_R} = S_R(E_{total}) - \frac{\mathcal{H}_S(\mu_S)}{T},$$

where we have finally used our assumption that the *reservoir is (energetically) humongous relative to the system*. Combining this approximation with our evaluation of the unconditional PDF, and dropping the subscript S since now nothing about the reservoir is left, we ultimately

arrive at the canonical probability distribution for a microstate with a specified temperature and set of generalized displacements:

$$p(\mu) = \frac{1}{Z(T, \mathbf{x})} e^{-\beta \mathcal{H}(\mu)}, \quad (6.22)$$

where again $\beta = (k_B T)^{-1}$, and we've defined the **partition function**⁸⁵

$$\boxed{Z(T, \mathbf{x}) = \sum_{\mu} e^{-\beta \mathcal{H}(\mu)},} \quad (6.23)$$

which plays a central role in the statistical mechanical description of macroscopic systems.

6.3 The partition function and thermodynamics

Our instinct – speculating from our work in the microcanonical ensemble – is that this partition function, this normalization factor of our probability distribution, will be connected to some⁸⁶ thermodynamically interesting quantities. To see how that comes about, remember that we've specified the temperature of the system, and it is exchanging energy back and forth with the reservoir to maintain that temperature. The energy of the system is now a random variable, which we'll call \mathcal{E} , and we associate the thermodynamic energy E with the expectation value of this quantity, $E = \langle \mathcal{E} \rangle$. What is the probability distribution associated with \mathcal{E} ? Glad you asked! We change variables from μ to $\mathcal{H}(\mu)$ and get

$$p(\mathcal{E}) = \sum_{\mu} p(\mu) \delta(\mathcal{H}(\mu) - \mathcal{E}) = \frac{1}{Z} e^{-\beta \mathcal{E}} \sum_{\mu} \delta(\mathcal{H}(\mu) - \mathcal{E}),$$

but the sum over delta functions just picks out the number of microstates with the appropriate energy $\Omega(\mathcal{E})$, which is related to the entropy, so

$$p(\mathcal{E}) = \frac{\Omega(\mathcal{E}) e^{-\beta \mathcal{E}}}{Z} = \frac{1}{Z} \exp \left[\frac{S(\mathcal{E})}{k_B} - \frac{\mathcal{E}}{k_B T} \right] = \frac{1}{Z} \exp [-\beta F(\mathcal{E})], \quad (6.24)$$

where $F = \mathcal{E} - TS(\mathcal{E})$ is obviously going to be related to the Helmholtz free energy. We can further simplify by noting that both S and E should be extensive, so we should be able to approximate sums of exponentials by the dominant term... we expect the probability to be sharply peaked about some most probably energy, E^* , and so we approximate the partition function as

$$Z = \sum_{\mu} e^{-\beta \mathcal{H}(\mu)} = \sum_{\mathcal{E}} e^{-\beta F(\mathcal{E})} \approx e^{-\beta F(E^*)}. \quad (6.25)$$

⁸⁵So-called because of its description of how probabilities are partitioned among all of the states with various energies. As far as I can tell the terminology dates to Darwin & Fowler's 1922 article, "On the Partition of Energy," in which they connect their partition function to the "zustandssumme" of Planck's 1921 article. I think the German term (roughly, the "sum-over-states") is more transparent, which is why I've always liked the notation here.

⁸⁶All?

This so-called “method of most probable values” suggests a logarithmic relationship between F and Z . This is reinforced by the “method of mean values,” where we compute the average energy of the system as

$$\langle \mathcal{H} \rangle = \sum_{\mu} \mathcal{H}(\mu) \frac{\exp(-\beta \mathcal{H}(\mu))}{Z} = -\frac{1}{Z} \frac{\partial}{\partial \beta} \sum_{\mu} e^{-\beta \mathcal{H}} = -\frac{\partial \log Z}{\partial \beta}. \quad (6.26)$$

Thermodynamically we recall that we could write

$$E = F + TS = F - T \left. \frac{\partial F}{\partial T} \right|_{\mathbf{x}} = \frac{\partial(\beta F)}{\partial \beta},$$

so this approach suggests the same identification of the Helmholtz free energy as

$$F(T, \mathbf{x}) = -k_B T \log Z(T, \mathbf{x}). \quad (6.27)$$

6.3.1 Moment and cumulant generating functions in thermodynamics

With this suggested connection between the partition function and the free energy, it’s interesting to notice that the partition function is *almost* a moment generating function of the sort we encountered back in Sec. 4.2. So, what’s the relationship between them? Suppose we have a system which can occupy a discrete set of states with energy $\mathcal{H}(\mu_k)$, where we’ve established that each state has a probability of $\exp(-\beta \mathcal{H}(\mu_k))/Z$ of being observed. We’ll simplify our lives slightly by assuming that the moments of the energy exist and viewing the characteristic function’s variable “ k ” as the combination “ ik ” we had before – with this the characteristic function for the energy is

$$\tilde{p}_{\mathcal{H}}(k) = \langle e^{k\mathcal{H}} \rangle = \sum_i e^{k\mathcal{H}(\mu_k)} \frac{\exp(-\beta \mathcal{H}(\mu_k))}{Z(\beta)}, \quad (6.28)$$

where i indexes the states. A simple manipulation of the product of the exponentials gives

$$\tilde{p}_{\mathcal{H}} = \frac{1}{Z(\beta)} \sum_i e^{-(\beta-k)\mathcal{H}(\mu_k)} \quad (6.29)$$

$$= \frac{Z(\beta - k)}{Z(\beta)}. \quad (6.30)$$

So: the moment generating function of the energy is the ratio of different values of the partition function⁸⁷

To get moments of the energy we can differentiate with respect to k and then evaluate our expression at $k = 0$. For instance, we readily see that

$$\langle \mathcal{H} \rangle = \frac{-1}{Z(\beta)} \frac{\partial Z}{\partial \beta}. \quad (6.31)$$

⁸⁷By the way: the “Wick rotation” of k here connects the seemingly unrelated fields of quantum mechanics and statistical mechanics: where inverse temperature formally behaves like an imaginary time coordinate in quantum mechanical problems.

We also know that with a moment generating function comes a cumulant generating function, which here is apparently

$$\log \tilde{p}_{\mathcal{H}}(k) = \log Z(\beta - k) - \log Z(\beta), \quad (6.32)$$

which in some ways looks close to our expression for the free energy, but is clearly not identical. And yet: to find the cumulants of the energy we want to take derivatives with respect to k and then evaluate at $k = 0$ – this will be mathematically the same as just taking derivatives of the free energy with respect to the inverse temperature! Thus, taking care of all the signs, we see that the n th cumulant of the energy is just

$$\langle \mathcal{H}^n \rangle_c = (-1)^n \frac{\partial^n \log Z}{\partial \beta^n}. \quad (6.33)$$

6.3.2 Thermodynamic connections and thermodynamic consistency

Before our little digression about the link between partition functions and generating functions, we did two separate calculations that suggested a link between the partition function and the Helmholtz free energy. For them to be mutually consistent we need the mean and most probable values to actually coincide in the thermodynamic limit – do they? Well, we just developed some easy technology for computing cumulants, let's use it. We know that the mean value is $\langle \mathcal{H} \rangle = \frac{-1}{Z(\beta)} \frac{\partial Z}{\partial \beta}$, and of course the energy is an extensive quantity. The second cumulant tells us about the spread of energies around the mean, and explicitly it is

$$\langle \mathcal{H}^2 \rangle_c = \frac{\partial^2 \log Z}{\partial \beta^2} = -\frac{\partial \langle \mathcal{H} \rangle_c}{\partial \beta} = k_B T^2 \left. \frac{\partial \langle \mathcal{H} \rangle}{\partial T} \right|_{\mathbf{x}} \Rightarrow \langle \mathcal{H}^2 \rangle_c = k_B T^2 C_{\mathbf{x}}.$$

We have identified that the variance of the energy is related to the heat capacity – another extensive quantity (actually, we can just as easily see that *every* cumulant of \mathcal{H} is proportional to N). This tells us that the “relative error,” $\sqrt{\langle \mathcal{H}^2 \rangle_c} / \langle \mathcal{H} \rangle_c \sim N^{-1/2}$; i.e., it vanishes in the thermodynamic limit. Apparently, indeed, the mean energy and the most probable energy are identical as $N \rightarrow \infty$.

Finally, do the expressions we have derived above connecting the partition function with the Helmholtz free energy make sense with our earlier expectations from thermodynamics? Thermodynamically, the canonical ensemble should correspond to a Legendre transform of the internal energy, $F = E - TS$. Let's deduce the *entropy* in the canonical ensemble via two equivalent means to show that this is, indeed, what we get. The first is to define the (Shannon) entropy associated with the probability distribution in Eq. 6.22 directly:

$$S = k_B \langle -\log p(\mu) \rangle = k_B \langle (\beta \mathcal{H} + \log Z) \rangle = \frac{E - F}{T}, \quad (6.34)$$

recovering once again the familiar expression for the Helmholtz free energy, $F = E - TS$. The second uses our knowledge of thermodynamics, where $dF = -SdT + \mathbf{J} \cdot d\mathbf{x}$. This gives

$$\begin{aligned} S &= - \left. \frac{\partial F}{\partial T} \right|_{\mathbf{x}} = k_B \frac{\partial (T \log Z)}{\partial T} = k_B \left(\log Z + T \frac{\partial \log Z}{\partial T} \right) \\ &= k_B \left(\log Z - \frac{1}{k_B T Z} \frac{\partial Z}{\partial \beta} \right) = \frac{-F + \langle \mathcal{H} \rangle}{T}. \end{aligned}$$

6.4 Non-interacting systems and mean field calculations

To get more practice with partition functions and how to manipulate them to get at the physical behavior of macroscopic systems, let's pause and work through some examples.

6.4.1 The ideal gas in the canonical ensemble

We first think about the ideal gas, and we'll see that the canonical ensemble returns the familiar relations we've come to expect. The probability distribution is⁸⁸

$$p(\mu) = \frac{1}{Z} \exp\left(-\beta \sum_{i=1}^N \frac{p_i^2}{2m}\right) \cdot \begin{cases} 1 & \text{if } \mathbf{q}_i \in \text{box} \\ 0 & \text{otherwise} \end{cases}, \quad (6.35)$$

and as always our task is basically to compute the partition function: " $Z = \sum_{\mu} e^{-\beta\mathcal{H}}$ ". In this case, what we *mean* by the sum over all possible states is the integral (with the correction to the phase space measure from Eq. 6.17!) over all possible positions and momenta the N gas particles could have:

$$Z(T, V, N) = \int \frac{1}{N!} \prod_{i=1}^N \frac{d^3\mathbf{q}_i d^3\mathbf{p}_i}{h^3} \exp\left(-\beta \sum_{i=1}^N \frac{p_i^2}{2m}\right). \quad (6.36)$$

This integral contains N copies of an integral over the volume and a bunch of Gaussian integrals. As such, we can read off the answer:

$$Z(T, V, N) = \frac{V^N}{N!} \left(\frac{2\pi m k_B T}{h^2}\right)^{3N/2}. \quad (6.37)$$

Let's take a moment to define the thermal de Broglie wavelength⁸⁹ $\lambda = h/\sqrt{2\pi m k_B T}$. With this bit of notation, the partition function is the more aesthetically pleasing

$$Z(T, V, N) = \frac{1}{N!} \left(\frac{V}{\lambda(T)^3}\right)^N,$$

Where I've written $\lambda(T)$ here to remind us that when we take derivatives of this partition function with respect to temperature, all the temperature dependence is buried in λ .

With the partition function in hand, we write the free energy as

$$F = -k_B T \log Z = -N k_B T \left(\log\left(\frac{V}{N}\right) + \frac{3}{2} \log\left(\frac{2\pi m k_B T}{h^2}\right) \right), \quad (6.38)$$

⁸⁸Note that the very form of the probability distribution here immediately gives us the microcanonical result we had for $p(\mathbf{p}_1)$.

⁸⁹Anticipating future developments in this course, λ helps characterize the regime in which classical statistical mechanics is typically valid: if λ is roughly the same order of the typical separation between particles then quantum effects become important.

or, using the more compact expression with λ ,

$$\begin{aligned} F &= -k_B T \log \left(\frac{1}{N!} \left(\frac{V}{\lambda^3} \right)^N \right) = Nk_B T \left(\log N - 1 - \log \left(\frac{V}{\lambda^3} \right) \right) \\ &= Nk_B T \left(\log \left(\frac{N\lambda^3}{V} \right) - 1 \right). \end{aligned} \quad (6.39)$$

From here, and using the thermodynamic expression $dF(T, V, N) = -SdT - PdV + \mu dN$, we can extract the usual properties of the ideal gas. For instance, we quickly have that

$$P = - \left. \frac{\partial F}{\partial V} \right|_{T,N} = \frac{Nk_B T}{V},$$

so the familiar ideal gas law is again derived. We can likewise re-derive the equipartition of energy – and get an explicit expression for the entropy of the canonical ideal gas, by noting that

$$S = - \left. \frac{\partial F}{\partial T} \right|_{V,N} = -Nk_B \left(\left[\log \left(\frac{N\lambda^3}{V} \right) - 1 \right] - 3T \frac{\lambda'(T)}{\lambda(T)} \right).$$

Noting that $\lambda'/\lambda = -(2T)^{-1}$, this simplifies to

$$S = \frac{5}{2} k_B N - k_B N \log \left(\frac{N\lambda^3}{V} \right).$$

Combining this with the thermodynamic $E = F + TS$ gives $E = 3NK_B T/2$, as expected,

It is interesting to note – or perhaps completely expected – that it seems not to matter which ensemble the gas is in: we recover the same macroscopic properties of the system either way. This will be a general theme: with one massive exception (having to do with the behavior of systems near phase transitions), different ensembles return the same physical results once you take the thermodynamic, $N \rightarrow \infty$ limit and identify the means of fluctuating quantities with the thermodynamic variables (just as we identified the average energy, $\langle \mathcal{H} \rangle$ with the thermodynamic internal energy E).

6.4.2 Non-interacting spins in a field

Earlier, in our discussions of phase transitions, we talked about the Ising model. As a warm-up, consider a variation of the Ising model in which the spins don't actually interact with each other, and only couple with an external field:

$$\mathcal{H} = -B \sum_i s_i,$$

where $s_i = \pm 1$. We again want to compute the partition function, and in this context the “sum over states” is a sum over all possible up-down configurations of each spin:

$$\begin{aligned}
 Z &= \sum_{\mu} e^{-\beta\mathcal{H}} = \sum_{\{s_i=\pm 1\}} \exp\left(-\beta B \sum_i s_i\right) \\
 &= \sum_{\{s_i=\pm 1\}} \prod_i e^{-\beta B s_i} \\
 &= \prod_{i=1}^N (e^{-\beta B} + e^{\beta B}) \\
 &= 2^N \cosh^N(\beta B).
 \end{aligned} \tag{6.40}$$

From this we immediately get the free energy,

$$F = -k_B T \log Z = -N k_B T \log [2 \cosh(\beta B)].$$

What is the magnetization, $M = \sum_i s_i$? Note that the probability of a microstate here is

$$p(\mu) = \frac{1}{Z} e^{-\beta B M},$$

so

$$\langle M \rangle = \sum_{\mu} M p(\mu) = \sum_{\mu} \frac{M}{Z} e^{-\beta B M}.$$

That is: just as we saw above that Z generates moments of \mathcal{H} , because of the particularly simple, single-term in the Hamiltonian in this case we also see that Z can be used to generate moments of M , since $\frac{\partial Z}{\partial B} = \sum_{\mu} -\beta M e^{-\beta B M}$. Thus, for this non-interacting model,

$$\langle M \rangle = \frac{1}{\beta} \frac{\partial \log Z}{\partial B} = -\frac{\partial F}{\partial B} = N \tanh(\beta B). \tag{6.41}$$

6.5 Gibbs canonical ensemble

Briefly⁹⁰, it is sometimes more convenient to work in an ensemble where the internal energy can change by exchanging heat with a reservoir (as in the canonical ensemble) but *also* by exchanging work. The macrostate $M(T, \mathbf{J})$ is now specified by the temperature and forces acting on the system, and we view both the energy and the generalized coordinates \mathbf{x} as random variables (but *not* including chemical work, which we will treat in the grand canonical ensemble shortly... thus, N is fixed). We proceed just as above, and find that the microstates of the system have a probability distribution

$$p(\mu) = \frac{\exp(-\beta\mathcal{H}(\mu) + \beta\mathbf{J} \cdot \mathbf{x})}{\mathcal{Z}(T, N, \mathbf{J})}, \tag{6.42}$$

⁹⁰i.e., entirely by analogy with the last section

where the Gibbs partition function is

$$\mathcal{Z}(T, N, \mathbf{J}) = \sum_{\mu, \mathbf{x}} \exp(-\beta\mathcal{H}(\mu) + \beta\mathbf{J} \cdot \mathbf{x}). \quad (6.43)$$

We can use, again, either the “most probable value” or “mean value” method to relate the expectation values of the generalized displacements to the Gibbs partition function, like

$$\langle \mathbf{x} \rangle = k_B T \frac{\partial \log \mathcal{Z}}{\partial \mathbf{J}}, \quad (6.44)$$

and we use the thermodynamic result that $\mathbf{x} = -\frac{\partial G}{\partial \mathbf{J}}$ to make the identification

$$G(T, N, \mathbf{J}) = -k_B T \log \mathcal{Z}, \quad (6.45)$$

where we have again written the Gibbs free energy encountered in Chapter 1 as $G = E - TS - \mathbf{x} \cdot \mathbf{J}$. One can, for instance, extract the enthalpy $H = E - \mathbf{x} \cdot \mathbf{J} = -\frac{\partial \log \mathcal{Z}}{\partial \beta}$, or the heat capacity at constant force as $C_{\mathbf{J}} = \frac{\partial H}{\partial T}$

6.6 The grand canonical ensemble

We now generalize once more from the canonical ensemble to the *grand* canonical ensemble. For the canonical ensemble we said that even though energy was a conserved quantity, it often makes more sense to put a system in contact with a reservoir of heat so that *temperature* is the fixed or controlled quantity; we let E become a random variable as the system and the reservoir exchange heat, and identified the thermodynamic E as the expectation value of the Hamiltonian⁹¹. In the same way, it often behooves us to generalize yet further and allow our system to also exchange *particle number* with the reservoir – now both N and E are variables and the expectation values, $\langle N \rangle$ and $\langle E \rangle$ are identified with thermodynamically interested quantities.

I’ll be honest and tell you: when I first took thermodynamics, I failed to appreciate why the grand canonical ensemble was interesting... “Chemical potential” sounded a lot like “chemistry” to me, and I saw this topic as a bit of esoterica. As we will see *repeatedly* in the coming chapters, the grand canonical ensemble is secretly central to our ability to describe systems undergoing phase transitions, and it often (paradoxically?) makes it easier to calculate various thermodynamic quantities.

So, we will turn to a reservoir for our system providing a fixed value of temperature T and chemical potential μ , and we want to find the probability distribution corresponding for the grand canonical ensemble, in which the macrostates are functions of $M(T, \mu, \mathbf{x})$. Notationally, since μ is the chemical potential, I’ll try to be careful about using μ_S to refer to a microstate of the system we’re interested in. The probability density function for μ_S is again derived by invoking the fundamental postulate and summing over all states of the reservoir, ultimately giving

$$p(\mu_S) = \frac{1}{\mathcal{Q}(T, \mu, \mathbf{x})} \exp[-\beta\mathcal{H}(\mu_S) + \beta\mu N(\mu_S)], \quad (6.46)$$

⁹¹Our attitude could be perhaps summarized as, “We measure temperature all the time, but when, after all, was the last time you precisely measured the energy of a macroscopically large system?”

where $N(\mu_S)$ is the number of particles in microstate μ_S , and where \mathcal{Q} is the *grand partition function*:

$$\mathcal{Q}(T, \mu, \mathbf{x}) = \sum_{\mu_S} \exp [\beta\mu N(\mu_S) - \beta\mathcal{H}(\mu_S)]. \quad (6.47)$$

We can usefully rearrange the above summation by first grouping together all of the microstates with the same number of particles in them. Letting \mathcal{H}_N refer to the Hamiltonian associated with the N -particle system, we write **next iteration: choose a different symbol for fugacity**⁹²

$$\mathcal{Q}(T, \mu, \mathbf{x}) = \sum_{N=0}^{\infty} e^{\beta\mu N} \sum_{\mu_S} e^{-\beta\mathcal{H}_N(\mu_S)} = \sum_{N=0}^{\infty} z^N \sum_{\mu_S} e^{-\beta\mathcal{H}_N(\mu_S)}, \quad (6.48)$$

where we have additionally defined the *fugacity*⁹³ $z = \exp(\beta\mu)$. Notice, by the way, that the second sum – over microstates with a particular number of particles – are the canonical partition functions associated with an N -particle system, so we can additionally write

$$\mathcal{Q}(T, z, \mathbf{x}) = \sum_{N=0}^{\infty} z^N Z_N(T, \mathbf{x}), \quad \text{defining } Z_0 \equiv 1. \quad (6.49)$$

This form makes it seem that to calculate the grand canonical partition function we need to have *already* calculated the canonical partition function. In principle we indeed do, but in later chapters⁹⁴ we'll see how we can sometimes make great progress in evaluating \mathcal{Q} even in situations where evaluating Z is very difficult.

6.6.1 Number fluctuations in the grand canonical ensemble

Earlier we showed the generic equivalence of the canonical and microcanonical ensembles as $N \rightarrow \infty$ by establishing that the average value of the energy and the most probable value of the energy became indistinguishable in the thermodynamic limit. Here we do the same thing by considering both the *mean* and *most typical* value of the number of particles. We note that from the definition of \mathcal{Q} we can read off the total weight of the microstates with N particles, the probability for finding the system with N particles is

$$p(N) = \frac{e^{\beta\mu N} Z(T, N, \mathbf{x})}{\mathcal{Q}(T, \mu, \mathbf{x})}. \quad (6.50)$$

We write the average as

$$\langle N \rangle = \frac{1}{\mathcal{Q}} \frac{\partial \mathcal{Q}}{\partial(\beta\mu)} = \frac{\partial \log \mathcal{Q}}{\partial(\beta\mu)}, \quad (6.51)$$

and the variance as

$$\langle N^2 \rangle_c = \langle N^2 \rangle - \langle N \rangle^2 = \frac{1}{\mathcal{Q}} \frac{\partial^2 \mathcal{Q}}{\partial(\beta\mu)^2} - \left(\frac{\partial \log \mathcal{Q}}{\partial(\beta\mu)} \right)^2 = \frac{\partial^2 \log \mathcal{Q}}{\partial(\beta\mu)^2} = \frac{\partial \langle N \rangle}{\partial(\beta\mu)}. \quad (6.52)$$

⁹²Common choices for the symbol used for fugacity include f, z, λ , all of which I want to use for other things...maybe we go with ϕ ?

⁹³Derived from *fugere*, to flee. The term was popularized in an early textbook by Gilbert Lewis and Merle Randall as an “escaping tendency,” referring to the flow of matter between phases, and playing a similar role to temperature in the flow of heat [57].

⁹⁴both on quantum stat mech and on interacting systems

We again see that the variance is proportional to $\langle N \rangle$, so we again get that the relative fluctuations, $\sigma_N / \langle N \rangle$ vanish in the thermodynamic limit. This suggests the equivalence of the grand canonical ensemble with the others, but have we been a little too glib, here?

From the above we see that $\beta \langle N^2 \rangle_c = \left. \frac{\partial \langle N \rangle}{\partial \mu} \right|_{T,V}$, recall that in Sec. 1.6.1 we showed that $\left. \frac{\partial N}{\partial \mu} \right|_{T,V}$ was itself related to the isothermal compressibility. We can use that result to write a simple *fluctuation-response* relationship, connecting the relative fluctuations in particle number to the compressibility (one of our thermodynamic response functions):

$$\frac{\langle N^2 \rangle_c}{\langle N \rangle^2} = \frac{1}{\beta N^2} \left. \frac{\partial N}{\partial \mu} \right|_{T,V} = \frac{k_B T}{V} \kappa_T. \quad (6.53)$$

From this we see that ordinarily the compressibility is just some value, and in the thermodynamic limit of $N \rightarrow \infty, V \rightarrow \infty$ the relative RMS fluctuations in particle number just scale like $\mathcal{O}\left(\frac{1}{\sqrt{V}}\right) \sim \mathcal{O}(N^{-1/2})$. Near phase transitions, though, more interesting things can happen! For instance, near a liquid-vapor phase transition at T_c , experiments suggest that the isothermal compressibility is one of these critically scaling quantities, $\kappa_T(T_c) \sim N^{0.63}$, suggesting unusually large fluctuations of particle density near the critical point. Such fluctuations can be seen in experiments⁹⁵.

So, under these sorts of circumstances the formalism associated with the grand canonical ensemble could in principle give non-identical answers as the formalism associated with the canonical ensemble. In these cases, we have no choice but to use the grand canonical ensemble.

6.6.2 Thermodynamics in the grand canonical ensemble

From the above, we now once again approximate the sum in Eq. 6.47 by its single largest term, which corresponds to the typical value of N (note that we feel free to go back and forth between fugacity and chemical potential representations as we see fit):

$$\mathcal{Q}(T, \mu, \mathbf{x}) = \lim_{N \rightarrow \infty} \sum_{N=0}^{\infty} [e^{\beta \mu N} Z(T, N, \mathbf{x})] \approx e^{\beta \mu N^* - \beta F} = e^{-\beta \mathcal{G}}, \quad (6.54)$$

where

$$\mathcal{G}(T, \mu, \mathbf{x}) = E - TS - \mu N = -k_B T \log \mathcal{Q}$$

is the *grand potential* (which we first met in Chapter 1), which is up to a factor of $-k_B T$ what Pathria defines as the q -potential. We can recover typical thermodynamic relationships using $d\mathcal{G} = -SdT - Nd\mu + \mathbf{J} \cdot d\mathbf{x}$, and extract pressures or heat capacities by usual derivative manipulations.

Entropy in the grand canonical ensemble can be defined by exactly the same two paths we saw when considering the canonical ensemble. We can either define the Shannon entropy

⁹⁵check out, for instance, some of the [critical opalescence](#) videos that are easy to find on youtube

associated with the probability distribution in Eq. 6.46, or take the thermodynamic path of $S = -\left.\frac{\partial \mathcal{G}}{\partial T}\right|_{\mu, \mathbf{x}}$, and we would ultimately find expressions like

$$S = \frac{E - F}{T} = k_B T \left.\frac{\partial \log \mathcal{Q}}{\partial T}\right|_{\mu, \mathbf{x}} - N k_B \log z + k_b \log \mathcal{Q}.$$

6.6.3 The ideal gas in the grand canonical ensemble

We once again churn through our standard example, the ideal gas, for the grand canonical ensemble. We have a macrostate $M(T, \mu, V)$, where the corresponding microstates are over particle positions and momenta with an indefinite number of particles in the specified volume. Again using the thermal de Broglie wavelength $\lambda = h/\sqrt{2\pi m k_B T}$, the grand partition function is

$$\begin{aligned} \mathcal{Q}(T, \mu, V) &= \sum_{N=0}^{\infty} e^{\beta \mu N} \frac{1}{N!} \int \left(\prod_{i=1}^N \frac{dV_i}{h^3} \right) \exp \left[-\beta \sum_{i=1}^N \frac{p_i^2}{2m} \right] \\ &= \sum_{N=0}^{\infty} \frac{1}{N!} e^{\beta \mu N} \left(\frac{V}{\lambda^3} \right)^N \\ &= \exp \left(e^{\beta \mu} \frac{V}{\lambda^3} \right) \end{aligned} \quad (6.55)$$

$$\Rightarrow \mathcal{G}(T, \mu, V) = -k_B T \log \mathcal{Q} = -k_B T e^{\beta \mu} \frac{V}{\lambda^3} \quad (6.56)$$

We can immediately read off things like⁹⁶

$$P = -\left.\frac{\partial \mathcal{G}}{\partial V}\right|_{T, \mu} = k_B T \frac{z}{\lambda^3}$$

or

$$N = -\left.\frac{\partial \mathcal{G}}{\partial \mu}\right|_{T, V} = \frac{zV}{\lambda^3} \Rightarrow PV = N k_B T,$$

as we have come to expect the equation of state for the ideal gas to look.

Notice, also, that \mathcal{G} for an ideal gas only depends on a single extensive variable: V . Since we expect \mathcal{G} to itself be extensive⁹⁷, it must be that $\mathcal{G} \propto V$. We have a name for the constant of proportionality: “pressure,” so $\mathcal{G}(T, V, \mu) = -P(T, \mu)V$. This makes for an easy method of computing pressures of systems that depend only on one extensive variable⁹⁸.

It is worth saying a bit more about the chemical potential, here. Rearranging that last expression for N gives

$$\mu = k_B T \log \left(\frac{\lambda^3 N}{V} \right), \quad (6.57)$$

⁹⁶switching back to the fugacity, largely to save on \TeX time

⁹⁷i.e., satisfying $\mathcal{G}(T, \lambda V, \mu) = \lambda \mathcal{G}(T, V, \mu)$

⁹⁸As we’ll see in the chapter on Quantum statistical mechanics, for example, when we calculate the grand partition function of ideal quantum gases.

and if $\lambda^3 < V/N$, as we certainly expect if we are to be doing classical calculations in the first place, then the chemical potential is negative! Is it clear why this makes sense? By writing μ as conjugate to N , we intuitively think of it as the energy cost associated with adding an extra particle to the system, but we need to look more carefully at the definition:

$$\mu = \left. \frac{\partial E}{\partial N} \right|_{S,V}.$$

That is, μ is the energy cost of adding a particle *at fixed entropy and volume*. In general, adding a particle will in fact increase the entropy (since there are more ways of partitioning the available energy), so if we are holding *entropy* fixed then the system must be lowering its energy when adding a particle. Hence, $\mu < 0$ for the classical ideal gas.

6.7 Problems

6.7.1 Momentum correlations in an ideal gas

In class we (hastily!) talked about deriving the unconditional probability of an ideal gas particle having some momentum by integrating over the full probability distribution (which came directly from the central postulate). What about two-particle correlations?

- (A) Find the unconditional probability for two particle positions, $\rho(\vec{q}_1, \vec{q}_2)$.
- (B) Find the unconditional probability for two particle momenta, $\rho(\vec{p}_1, \vec{p}_2)$.

6.7.2 Microcanonical two-level system

Consider N non-interacting two-state particles (for instance, impurity atoms trapped in a solid matrix), where each particle has a fixed position (so that they have no kinetic energy, and they are distinguishable). Each particle can be in either the “spin up”/“excited” state, which costs energy ϵ , or in the “spin down”/“ground” state, which costs zero energy. So, at energy E there are $N_1 = E/\epsilon$ excited states.

What is the microcanonical normalization factor, $\Omega(E, N)$? What is the entropy? What is the heat capacity, $C = \frac{dE}{dT}$?

Note, by the way, that the heat capacity you derived vanishes extremely quickly as $T \rightarrow 0$. In most materials, the contributions of spins to the heat capacity is dwarfed by other contributions (e.g., phonons, or conduction electrons in metals). In some materials, though, one can detect contributions of the form you derived; the extra contribution is called the Schottky anomaly.

6.7.3 Harmonic oscillators

Your friend hands you a collection of N harmonic oscillators⁹⁹, each of which has energy eigenvalues $(n + 1/2)\hbar\omega$ for $n = 0, 1, 2, \dots$

Part A:

In the limit that N is large, derive an expression for the number of ways of partitioning a total amount of energy E among the N harmonic oscillators.

Part B:

What is the volume of phase space corresponding to states with energy in the range $E \pm 1/2$? In the asymptotic limit, compute the volume of the relevant shell in $2N$ -dimensional space, and show that the volume per state is h^N .

⁹⁹“friend” – who would do such a thing?!

6.7.4 Energy fluctuations in the microcanonical ensemble

In deriving the laws of thermodynamics we exploited the fact that when bringing two subsystems together with energies E_1 and $E_2 = E - E_1$ we could assume that an in-principle complicated sum could be approximated by just picking out the single biggest term:

$$\Omega(E) = \int dE_1 \Omega_1(E_1)\Omega_2(E - E_1) \approx \exp\left(\frac{1}{k_B}(S_1(E_1^*) + S_2(E_2^*))\right).$$

(A) An expression for the equilibrium fluctuations: Consider the probability that subsystem 1 has energy E_1 , $p(E_1) = \frac{\Omega_1(E_1)\Omega_2(E-E_1)}{\Omega(E)}$. Taylor expand the numerator¹⁰⁰ about the value E_1^* , using the fact that when the two subsystems are in equilibrium they have the same temperature, to show that the energy fluctuations of system 1 are Gaussian. What is the variance, $\sigma_{E_1}^2$, of that Gaussian? Assuming that both S and E are proportional to N , how do the energy fluctuations per particle, σ_{E_1}/N , scale with N ?

(B) Fluctuation-response formula Juggle some derivatives around to show that

$$\frac{1}{k_B} \frac{\partial^2 S}{\partial E^2} = -\frac{1}{k_B T} \frac{1}{N c_v T},$$

where $c_v = \frac{1}{N} \frac{\partial E}{\partial T}|_{V,N}$, the energy per particle needed to change the temperature. This should be thermodynamically straightforward. But it's interesting! This is an example of a "fluctuation-response" relation – normally you think of measuring the specific heat (C_V) by adding some energy to your system and measuring the temperature. This kind of formula suggests that you can *also* measure it by simply measuring the equilibrium fluctuations of the energy of the system¹⁰¹; it's something we'll see more of later in the course.

6.7.5 Ideal relativistic gas

Consider an extreme, but ideal, relativistic gas composed of N *indistinguishable* particles whose kinetic energy per particle is not $p_i^2/(2m)$ but rather pc . **Show that** the partition function in the canonical distribution is

$$Z(T, V, N) = \frac{1}{N!} \left[8\pi V \left(\frac{k_B T}{hc} \right)^3 \right]^N.$$

From the partition function you can derive thermodynamic relationships. **Show that**

$$PV = E/3, \quad E/N = 3k_B T, \quad PV^\gamma = \text{constant}.$$

Note that, given a partition function, one can calculate the *density of states* $g(E)$, where $g(E)dE$ counts the number of states in a range dE near the energy E for the system, as the inverse Laplace transform of the partition function $Z(\beta)$:

$$g(E) = \frac{1}{2\pi i} \int_{\beta' - i\infty}^{\beta' + i\infty} e^{\beta E} Z(\beta) d\beta = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{(\beta' + i\beta'')E} Z(\beta' + i\beta'') d\beta'',$$

¹⁰⁰Consider how exponentially large numbers work – should you do a Taylor expansion of the exponential or the argument of the exponential?

¹⁰¹The fluctuations might be teeny-tiny if N is Avogadro's number, but this can be incredibly useful when doing smaller computer simulations, or when looking at (e.g.), colloidal experiments.

where we're treating $\beta = \beta' + i\beta''$ as a complex variable and, since β' is positive the integration path is parallel and to the right of the imaginary axis. For this problem **what is the density of states?**

6.7.6 Biased random walk

Consider an ideal random walk in which each step direction is i.i.d. uniformly distributed over the unit sphere, in which one end fixed is fixed (at the origin) and there is someone pulling on the other end with force $\mathbf{F} = f\hat{z}$. Letting \mathbf{R} stand for the end-to-end vector of the chain and \mathbf{r}_i for the vector corresponding to step i , the the work due to this external force is $W = \mathbf{F} \cdot \mathbf{R} = f\hat{z} \cdot \sum_i \mathbf{r}_i$. Every conformational state of the random walk costs zero energy otherwise.

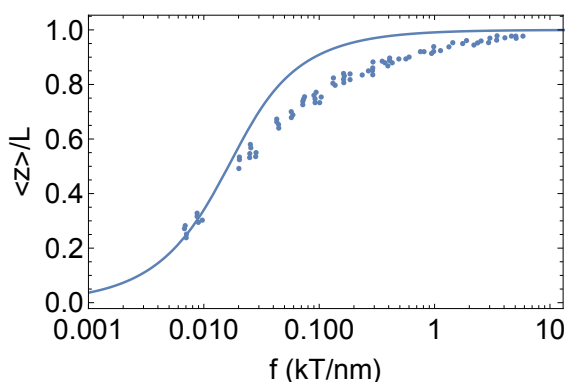


Figure 6.1: **Force vs extension curves** for 97004-basepair DNA in 10 nM Na+ (points), and the functional form of what what your answer should look like (curve). Figure modeled after Ref. [58].

- Write down an expression for the partition function in the ensemble where you specify N and the force F . Given the independence of the steps, you may find it helpful to first find the partition function for a single step.
- Now that you have a partition function, use it to find the average end-to-end distance for a given external force: $\frac{\langle R \rangle}{R_{max}}$. Your function should behave something like the curve in Fig. 6.1; this simple estimate, together with an improved theoretical treatment, is the core of a famous paper in biophysics.

6.7.7 Differences in specific heats, again

Show that in general we can write

$$C_P - C_V = -k \frac{\left[\frac{\partial}{\partial T} \left(T \frac{\partial \log Z}{\partial V} \Big|_T \right) \right]_V^2}{\frac{\partial^2 \log Z}{\partial V^2} \Big|_T}.$$

From there, prove that $(C_P - C_V)$ is always positive. Verify that the value of this difference for an ideal gas is $k_B N$.

6.7.8 Equilibration between positive- and negative-temperature systems

A systems of N_1 spins at negative temperature and positive energy is brought into thermal contact with an ideal gas of N_2 particles. What is the character of the state to which this combined system eventually comes to equilibrium? Is the final temperature positive or negative? Does your answer depend on the ratio N_1/N_2 ?

6.7.9 Statistics of sub-systems

Suppose you have a system of total volume V_0 which contains N_0 particles. In this problem we'll think about the behavior of sub-sets of this system.

First, assuming that the positions of the particles are uncorrelated with each other, calculate the probability $P(N, V)$ that a region of volume V located somewhere inside V_0 has exactly N particles.

Part A:

Show that $\langle N \rangle = N_0 p$ and $\sqrt{\langle N^2 \rangle - \langle N \rangle^2} = \sqrt{N_0 p(1-p)}$ for $p = V/V_0$

Part B:

Show that if both $N_0 p$ and $N_0(1-p)$ are large numbers, expand about the average value of N to show that $P(N, V)$ has approximately a Gaussian form.

Part C:

If $p \ll 1$ and $N/N_0 \ll 1$, show that $P(N, V)$ has the form of a Poisson distribution,

$$P(N) = \exp(-\langle N \rangle) \frac{\langle N \rangle^N}{N!}.$$

6.7.10 Adsorption of surfactant molecules

Dilute solutions of surfactants can be approximated as ideal gases. The surfactant molecules can reduce their energy by coming into contact with air (or other porous media – such as polymers and gels – that have an affinity for them), and at a solution-air interface the fraction of molecules at the surface then behave like a two-dimensional gas. Let's think about this dimensionally-dependent problem a bit more

Part A:

Consider an ideal gas of indistinguishable classical particles of mass m in d dimensions, in a *uniform* external potential of strength ϵ_d . **What is the canonical partition function, $Z_d(N, V, T)$?** Show that the chemical potential is

$$\mu_d = \epsilon_d - k_B T \log \left[\frac{V_d}{N_d \lambda^d} \right], \quad \text{where } \lambda = \frac{h}{\sqrt{2\pi m k_B T}}$$

Part B:

Suppose a surfactant lowers its energy by ϵ_0 by moving from the bulk of the solution to the surface. What is the ratio of particles at the $d = 2$ surface relative to the $d = 3$ bulk?

Hint: it may help to write the d -dimensional grand canonical partition function \mathcal{Q} , and then to use the relation $\langle N_d \rangle = k_B T \frac{\partial \log \mathcal{Q}}{\partial \mu}$.

Part C:

Gels are typically formed by cross-linking linear polymer molecules together, and it is often asserted that porous gels should be viewed as having a fractal-like structure. This suggests viewing surfactant adsorption as a gas in d_f -dimensional space, where d_f is some non-integer dimension. From your result to Part B above, is it possible to test the above assertion by comparing as a function of temperature the relative amount of adsorption of (a) surfactants to the gel and (b) surfactants to the same amount of individual (linear) polymers before cross-linking?

6.7.11 A gas of magnetic atoms

Suppose you have an ideal gas of “magnetic” particles that don’t interact with each other but have a usual kinetic energy and are also coupled to an external potential so that each atom can have magnetic potential energy equal to $\mu_B H$ or $-\mu_B H$, depending on the orientation of the particle relative to the applied field H .

What is the grand partition function? What is the expression for the magnetization of the system? What is the entropy? Also, given your expression for the entropy, **how much heat will be given off** by the system if the field were reduced from H to zero at constant V and constant T ?

6.7.12 An ideal gas in a confining potential

In this problem we are studying a gas of N classical, indistinguishable, non-interacting particles in a gravitational field, with

$$U_g(x, y, z \geq 0) = mgz,$$

where the constant $g > 0$, and the gas is confined in a semi-infinite vertical container of cross-sectional area A (i.e., the z -coordinate of every particle must be between zero and positive infinity, and the x and y coordinates are confined to some shape of area A).

For this system at temperature T , **what is the canonical partition function? What is the Helmholtz free energy? What is the entropy? What is the internal energy? What is the heat capacity?**

6.7.13 A different ideal gas, in a different confining potential

This time we have have N classical, indistinguishable, non-interacting particles in a harmonic trap. The Hamiltonian is

$$\mathcal{H} = \sum_i \left(\frac{\vec{p}_i^2}{2m} + ar_i^2 \right).$$

- (A) What is the canonical partition function? What is the internal energy?
- (B) What is the grand canonical partition function? What is the expectation value for the number of particles, $\langle N \rangle$?
- (C) What is the isothermal compressibility, $\kappa_T = -\frac{1}{V} \frac{\partial V}{\partial P} \Big|_T$? In answering this question, it may be helpful to find a Maxwell relation connecting the partial derivatives in the definition of κ_T with partial derivatives that are easier to calculate. The Gibbs-Duhem relation may be helpful.

6.7.14 Number fluctuations in the grand canonical ensemble

When studying number fluctuations we said we could use “some thermodynamic relationships” to write

$$\frac{\langle N^2 \rangle_c}{\langle N \rangle^2} = \frac{k_B T}{V} \kappa_T, \quad (6.58)$$

where κ_T is the isothermal compressibility. Fill in the missing steps to actually derive this equation.

6.7.15 Triplet oxygen partition function

“Triplet oxygen” refers to the most stable form of an O_2 molecule. O_2 has two unpaired electrons and hence four independent spin states, and exchange interactions split these four spin states into the singlet state (with total spin $s = 0$) and the triplet state (with total spin $s = 1$); the triplet state is the energetically favorable one.

So, in the triplet state the spin s^z is quantized to be either $s^z = -1$, $s^z = 0$, or $s^z = 1$. Suppose we put an ideal (non-interacting) gas of N molecules of triplet oxygen in a magnetic field of strength B pointing in the \hat{z} direction, all in a box of total volume V . The Hamiltonian is

$$\mathcal{H} = \sum_i \left(\frac{\vec{p}_i^2}{2m} - \mu B s_i^z \right),$$

where μ here is just the magnetic permeability (i.e., has no bearing to the μ which is the chemical potential).

Part A:

Treating the positions and momenta of the molecules classically, keeping track of the quantized spin, and ignoring any other degrees of freedom that might exist in the problem, calculate the canonical partition function $Z(N, V, T, B)$.

Part B:

At a given temperature, what fraction of the molecules do you expect to find in each of the $s^z = \{-1, 0, 1\}$ states?

Part C:

What is the average magnetic moment, $\langle M \rangle$, where $M = \mu \sum_i s_i^z$?

Part D:

What is the magnetic susceptibility, $\chi_M = \frac{\partial \langle M \rangle}{\partial B}$, in the limit of zero field?

Part III

Interacting systems, phase transitions,
and of out-of-equilibrium statistical
physics

Chapter 7

Interacting systems: Perturbative approaches

The truth is that most things that are interesting are not ideal, but that makes it hard to calculate.¹⁰²

So far we have almost exclusively focused on ideal systems, that is, systems in which the degrees of freedom did not interact with each other via interparticle potentials. This simplification helped us more clearly understand the structure of our statistical descriptions of macroscopic systems (and, helpfully, let us solve everything analytically) – and for quantum mechanical systems we even saw that non-interacting Bose and Fermi systems can exhibit rich/interesting behaviors.

However, interactions are responsible for the amazing variety of phases of matter and material behaviors! Since most physical systems that we encounter cannot be described without considering interactions, in this chapter and the next we'll figure out how to incorporate interactions into our statistical mechanical formalism. This chapter will focus on systematic expansions, in which an idealized, non-interacting system serves as a useful starting point. Implicitly, for instance, throughout the next sections you can imagine that we're trying to move from thinking about an ideal classical gas to a dilute classical gas.

7.1 From moment expansions...

Let's return to the idea of a classical system¹⁰³ and start with a general Hamiltonian in the absence of an external potential:

$$\mathcal{H} = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m} + \mathcal{U}(\{\mathbf{r}\}),$$

where we've written a general interaction potential, \mathcal{U} , which could be an arbitrary function involving the spatial coordinates of the particles.

¹⁰²With thermodynamics, one can calculate almost everything crudely; with kinetic theory, one can calculate fewer things, but more accurately; and with statistical mechanics, one can calculate almost nothing exactly” – Eugene Wigner, [59].

¹⁰³With our corrected phase-space measure

In the canonical ensemble the partition function would be

$$\begin{aligned} Z(N, V, T) &= \frac{1}{N! h^{3N}} \int \left(\prod_i d^3 p_i d^3 r_i \right) \exp \left(-\beta \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m} \right) e^{-\beta \mathcal{U}(\{\mathbf{r}\})} \\ &= \frac{1}{N!} \left(\frac{V}{\lambda^3} \right)^N \int \prod_i \frac{d^3 r_i}{V} e^{-\beta \mathcal{U}(\{\mathbf{r}\})}, \end{aligned} \quad (7.1)$$

where in the second line we've done the integral over momenta and taken the liberty of multiplying and dividing by N copies of the system volume V . We've done this because now we see terms directly related to ideal, non-interacting quantities. The prefactor is just the ideal gas partition function, and the integral is like doing an average over particle positions where there are no correlations between the particle positions – exactly as if the positions were those from an ideal gas with no interactions. Using the notation where a subscript (0) refers to these ideal-gas like quantities or averages, we can write the canonical partition function as

$$Z(N, V, T) \equiv Z_{(0)}(N, V, T) \langle e^{-\beta \mathcal{U}(\{\mathbf{r}\})} \rangle_{(0)} \quad (7.2)$$

$$= Z_{(0)} \sum_l \frac{(-\beta)^l}{l!} \langle \mathcal{U}^l \rangle_{(0)}. \quad (7.3)$$

This looks like a moment-based perturbative description of a system: when $\mathcal{U} = 0$ we recover the ideal gas which we know how to solve, and when $\mathcal{U} \neq 0$ we can perhaps calculate corrections systematically. As we'll see shortly, this direct moment-based expansion is often not especially useful: at short ranges there are often strong repulsions (Pauli exclusion, or hard-core repulsion between particles, or...), so the moments of \mathcal{U} need not be small. Nevertheless, working with this will lead us to an expansion which *is* useful.

7.2 ... to cumulant expansions...

We know that what we often want to work with is the log of the partition function, so let's replace our moment expansion with a cumulant expansion: we see above that Z is acting like a generator of moments, so taking the log gives us a generator of cumulants¹⁰⁴:

$$\log Z = \log Z_{(0)} + \sum_{l=1}^{\infty} \frac{(-\beta)^l}{l!} \langle \mathcal{U}^l \rangle_{c(0)}. \quad (7.4)$$

At this point we specialize the form of the interparticle potential away from complete generality, and study potentials, ϕ that are pairwise in nature,

$$\mathcal{U} = \sum_{i < j} \phi(\mathbf{r}_i - \mathbf{r}_j) \equiv \phi(\mathbf{r}_{ij}), \quad (7.5)$$

¹⁰⁴We'll need an occasional gross double subscript as a result of this notational choice, but we'll reap the benefits overall.

where we've introduced the notation \mathbf{r}_{ij} to represent vector separations between pairs of particles, which we'll be seeing a lot of in the next sections.

Writing out the first few terms of this cumulant expansion, we have

$$\log Z = \log Z_{(0)} - \beta \langle \mathcal{U} \rangle_{(0)} + \frac{\beta^2}{2} \left(\langle \mathcal{U}^2 \rangle_{(0)} - \langle \mathcal{U} \rangle_{(0)}^2 \right) + \dots \quad (7.6)$$

Let's evaluate these first few cumulants for our pairwise potential.

7.2.1 First cumulant

The first cumulant is quite straightforward, recognizing that for each pair of particles we pick as the interacting pair we're going to get copies of the same result:

$$\begin{aligned} \langle \mathcal{U} \rangle_{(0)} &= \sum_{i < j} \int \left(\prod_{\alpha} \frac{d^3 r_{\alpha}}{V} \right) \phi(\mathbf{r}_i - \mathbf{r}_j) \\ &= \frac{N(N-1)}{2} \int \frac{d^3 r_1}{V} \frac{d^3 r_2}{V} \dots \frac{d^3 r_N}{V} \phi(\mathbf{r}_1 - \mathbf{r}_2) \\ &= \frac{N(N-1)}{2V} \int d^3 r \phi(\mathbf{r}), \end{aligned} \quad (7.7)$$

where in the last line we've let $\mathbf{r} = \mathbf{r}_{12}$. There you go: someone hands you a particular interparticle potential (Lennard-Jones, or screened Coulomb, or...), and you go off, calculate an integral, and you've got the first term in the cumulant expansion.

7.2.2 Second cumulant

Let's write out the second cumulant as

$$\langle \mathcal{U}^2 \rangle_{c_{(0)}} = \sum_{\substack{i < j \\ k < l}} \left[\langle \phi(\mathbf{r}_{ij}) \phi(\mathbf{r}_{kl}) \rangle_{(0)} - \langle \phi(\mathbf{r}_{ij}) \rangle_{(0)} \langle \phi(\mathbf{r}_{kl}) \rangle_{(0)} \right]. \quad (7.8)$$

This is a sum over $\binom{N(N-1)}{2}$ total terms, and it is helpful to divide those into three classes of terms:

All particle labels are distinct, i.e., i, j, k, l are all different indices. In this case we can look at the second moment and see

$$\begin{aligned} \langle \phi(\mathbf{r}_{ij}) \phi(\mathbf{r}_{kl}) \rangle_{(0)} &= \int \left(\prod_{\alpha} \frac{d^3 r_{\alpha}}{V} \right) \phi(\mathbf{r}_{ij}) \phi(\mathbf{r}_{kl}) \\ &= \left(\int \frac{d^3 r_i}{V} \frac{d^3 r_j}{V} \phi(\mathbf{r}_{ij}) \right) \left(\int \frac{d^3 r_k}{V} \frac{d^3 r_l}{V} \phi(\mathbf{r}_{kl}) \right) \\ &= \langle \phi(\mathbf{r}_{ij}) \rangle_{(0)} \langle \phi(\mathbf{r}_{kl}) \rangle_{(0)}. \end{aligned} \quad (7.9)$$

Thus, these terms do not make any contribution to $\langle \mathcal{U}^2 \rangle_{(0)}$.

One particle label is shared i.e., we have $i, j, k = i, l$ as the four labels. We can make a similar argument for neglecting this class of terms, too. We first write

$$\langle \phi(\mathbf{r}_{ij})\phi(\mathbf{r}_{il}) \rangle_{(0)} = \int \frac{d^3r_i d^3r_j d^3r_l}{V^3} \phi(\mathbf{r}_{ij})\phi(\mathbf{r}_{il}), \quad (7.10)$$

and then change variables in the integration from $\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_l$ to $\mathbf{r}_i, \mathbf{r}_{ij}, \mathbf{r}_{il}$. This leaves us with

$$\langle \phi(\mathbf{r}_{ij})\phi(\mathbf{r}_{il}) \rangle_{(0)} = \int \frac{d^3r_i d^3r_{ij} d^3r_{il}}{V^3} \phi(\mathbf{r}_{ij})\phi(\mathbf{r}_{il}) = \langle \phi(\mathbf{r}_{ij}) \rangle_{(0)} \langle \phi(\mathbf{r}_{il}) \rangle_{(0)}, \quad (7.11)$$

which again means these terms do not contribute to $\langle \mathcal{U}^2 \rangle_{(0)}$.

The same pair is considered twice i.e., $i = k, j = l$. Well, the second cumulant doesn't vanish, and these are the remaining terms which we do need to keep track of. There are $\frac{N(N-1)}{2}$ terms of this form, and they all contribute identically. Thus, our complete expression for the second cumulant is

$$\langle \mathcal{U}^2 \rangle_{c(0)} = \frac{N(N-1)}{2} \left[\int \frac{d^3r_{ij}}{V} \phi^2(\mathbf{r}_{ij}) - \left(\int \frac{d^3r_{ij}}{V} \phi(\mathbf{r}_{ij}) \right)^2 \right]. \quad (7.12)$$

7.2.3 Cumulant expansion of the partition function

So far, we have the following expansion for the log of the partition function:

$$\log Z = \log Z_{(0)} + \frac{N(N-1)}{2} \left[-\beta \int \frac{d^3r}{V} \phi(\mathbf{r}) + \frac{\beta^2}{2} \left(\int \frac{d^3r}{V} \phi^2(\mathbf{r}) - \left(\int \frac{d^3r}{V} \phi(\mathbf{r}) \right)^2 \right) + \dots \right] + \dots \quad (7.13)$$

We now consider the thermodynamic limit, $V, N \rightarrow \infty$, and we find

$$\log Z \approx N \log \left(\frac{Ve}{N\lambda^3} \right) + \frac{N^2}{2V} \left[-\beta \int d^3r \phi(\mathbf{r}) + \frac{\beta^2}{2} \int d^3r \phi^2(\mathbf{r}) + \dots \right]. \quad (7.14)$$

If we now, for instance, want to know the pressure of our system, we take the appropriate derivative and find

$$\begin{aligned} \beta P &= \frac{\partial \log Z}{\partial V} \\ &= \frac{N}{V} - \frac{1}{2} \left(\frac{N}{V} \right)^2 \left[-\beta \int d^3r \phi(\mathbf{r}) + \frac{\beta^2}{2} \int d^3r \phi^2(\mathbf{r}) + \dots \right] + \mathcal{O} \left(\left(\frac{N}{V} \right)^3 \right). \end{aligned} \quad (7.15)$$

What have we done? We've basically written the equation of state of the system as a perturbation organized in powers of density, where the contribution at each order in density is a sum over a series of terms involving integrals of powers of the pairwise potential.

As written, this sort of cumulant expansion is still not very helpful. Why? Because for typical interactions there are large forces keeping molecules apart. For instance, one common potential is the Lennard-Jones potential:

$$\phi(\mathbf{r}) = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right], \quad (7.16)$$

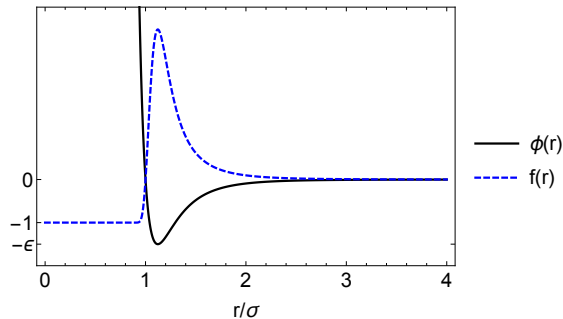


Figure 7.1: **Lennard-Jones potential and the corresponding Mayer f-function** The potential, ϕ , is shown as a solid black line and the corresponding $f = \exp(-\beta\phi) - 1$ is shown as a dashed blue line.

where ε is related to the depth of the potential well and σ to the range over which the repulsion is felt.¹⁰⁵

With such a potential, we see that *each integral in the series of series we wrote about diverges!* This is, on its face, not such a good perturbation theory. Before we integrate each term and despair, though, let's do two things. First, let's assume that the perturbative series in *density* is still okay; for a dilute gas we'll go ahead and truncate at order $(\frac{N}{V})^2$. Second, let's sum the series first:

$$\begin{aligned}
 -\beta \int d^3r \phi(\mathbf{r}) + \frac{\beta^2}{2} \int d^3r \phi^2(\mathbf{r}) + \dots &= \int d^3r \left[-\beta\phi + \frac{\beta^2}{2}\phi^2 - \frac{\beta^3}{3!}\phi^3 + \dots \right] \\
 &= \int d^3r [e^{-\beta\phi(\mathbf{r})} - 1] \\
 &\equiv \int d^3r f(\mathbf{r}), \tag{7.17}
 \end{aligned}$$

where in the final line we defined a new function (the *Mayer f function*) to stand for the combination $e^{-\beta\phi(\mathbf{r})} - 1$. In Figure 7.1 I show a plot of the Lennard-Jones potential together with the associated $f(\mathbf{r})$: At short distances (where the potential is diverging) the f function converges to a value of -1 , and at large distances (where the potential is vanishing) the f function converges to 0. Indeed, for reasonable potentials integrals over these f functions are perfectly well behaved, and we see that in our cumulant expansion we wrote down a series in which every term individually diverges, but the sum of the series is something we can evaluate!

Surveying our work in this section, we see that we tried to write down *cumulant expansion* – a perturbative expansion in the potential – but that for reasonable potentials we ended up

¹⁰⁵Where does this sort of potential come from? The attractive r^{-6} part comes from fluctuating dipoles of electrically neutral atoms. Recall that if there were two interacting permanent dipole moments, p_1 and p_2 , the potential energy would scale as $p_1 p_2 / r^3$. There are no permanent dipoles for neutral atoms, but atoms can acquire transient dipoles through quantum fluctuations. If the first atom has a transient dipole p_1 it will induce an electric field, which will in turn induce a dipole in the second atom $p_2 \sim E \sim p_1 / r^3$. The resulting energy thus scales as $p_1 p_2 / r^3 \sim r^{-6}$; this is typically called the van der Waals attraction. The r^{-12} term is meant to reflect the rapid transition to strong repulsion as the atoms get very close. The exact form is not so important, and the common choice of a term like r^{-12} is simply a mathematical convenience.

having to re-express our result in terms of these Mayer f functions. It seems like we would have been better off – and could perhaps have made more systematic progress – had we been able to *expand in powers of f* , rather than in powers of ϕ . The *cluster expansion* introduced in the next section allows us to do exactly that!

7.3 ...to cluster expansions!

Now that we have a *reason* that we might want to find an expansion of the partition function in powers of these f functions, let's do so. We continue working with our Hamiltonian with only pairwise interparticle potential,

$$\mathcal{H} = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m} + \sum_{i<j} \phi(\mathbf{r}_{ij}),$$

for which the canonical partition function is

$$Z(N, V, T) = \frac{1}{N! \lambda^{3N}} \int \left(\prod_{\alpha} d^3 r_{\alpha} \right) \prod_{i<j} e^{-\beta \phi(\mathbf{r}_{ij})}. \quad (7.18)$$

In order to understand how to manipulate this expression into a useful expansion, our initial goal of this section is going to be to transcribe Eq. 7.18 from math to pictures. We begin by introducing the additional bit of notation, writing the Mayer f functions as

$$f_{ij} \equiv f(\mathbf{r}_{ij}) = e^{-\beta \phi(\mathbf{r}_i - \mathbf{r}_j)} - 1.$$

Using this notation, let's organize the terms in Z by how many powers of f they contain:

$$\begin{aligned} Z(N, V, T) &= \frac{1}{N! \lambda^{3N}} \int \left(\prod_{\alpha} d^3 r_{\alpha} \right) \prod_{i<j} (1 + f_{ij}) \\ &= \frac{1}{N! \lambda^{3N}} \int \left(\prod_{\alpha} d^3 r_{\alpha} \right) \left[1 + \sum_{i<j} f_{ij} + \sum_{\substack{i<j \\ k<l}} f_{ij} f_{kl} + \dots \right] \end{aligned} \quad (7.19)$$

7.3.1 Diagrammatic representation of the canonical partition function

To help us organize the many terms that appear in Eq. 7.19, let's start representing integrals with diagrams¹⁰⁶. To my knowledge ([9], Chapter 10), this represents one of the earlier uses of diagrams to organize terms in perturbative calculations [60, 61], pre-dating Feynman diagrams by a handful of years. The way we'll draw these diagrams is by (1) drawing N points, and then (2) representing f_{ij} by a line connecting points i and j . According to this prescription, an n th-order term in f corresponds to diagrams with n lines drawn.

¹⁰⁶Recall, from the section on probability, how we used a similar approach to deal with the combinatorics of relating cumulants to moments.

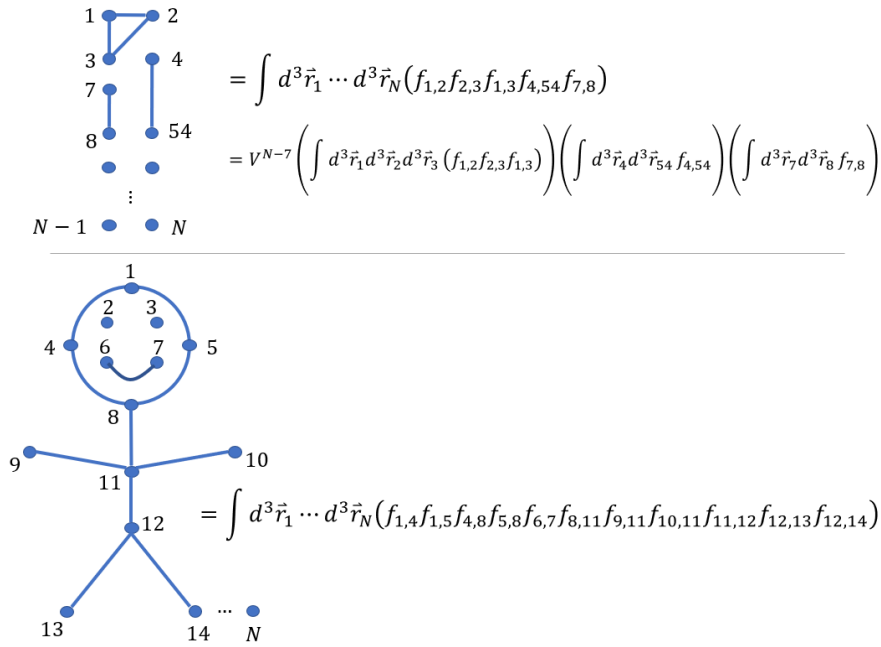


Figure 7.2: **The contribution of a graph is a product of the linked clusters**, where each unlinked point gives a factor of the volume. The geometrical details of how we draw these is irrelevant, as indicated with the stick-person diagram.

An example of a diagram with 5 copies of f is shown in the top portion of Fig. 7.2. An important observation is that the contribution of a particular diagram can be written as the product of the linked clusters it contains (and where each point without a connecting line contributes one power of the volume, V). Another important observation is that every diagram with the same structure contributes identically, i.e., independently of the *labels*; we'll ultimately only need to care about the number of 1-clusters, 2-clusters, triangles, etc., in a diagram. The bottom portion of the figure further emphasizes that the way we draw our diagrams (relative positions of points, geometric details) is completely irrelevant.

We now have a prescription for writing the terms in Eq. 7.19 as pictures, and now we will *organize* those pictures in a particular way. We define *cluster integrals*,

$$b_l \equiv \text{Sum over contributions of all linked clusters of } l \text{ points.} \tag{7.20}$$

Ultimately, these b_l are the namesake of the *cluster expansion* we will construct by the end of this section. The first few cluster integrals are shown in Fig. 7.3:

7.3.2 The cluster expansion

Let's see why we care about these b_l . First, consider a diagram in which we take N points and partition it into n_1 clusters of size 1, n_2 clusters of size 2, and so forth (so, in the example in Fig. 7.2, we would have $n_1 = N - 7$, $n_2 = 2$, $n_3 = 1$, and $n_4 = n_5 = \cdots = n_N = 0$). This defines a set of numbers of clusters of different sizes, $\{n_l\}$. We noted above, though, that every diagram that looks the same contributes the same amount, regardless of the labels

$$\begin{aligned}
b_1 &= \bullet = \int d^3\vec{r} = V \\
b_2 &= \text{---} = \int d^3\vec{r}_1 d^3\vec{r}_2 f(\vec{r}_{12}) = V \int d^3\vec{r} f(\vec{r}) \\
b_3 &= \triangle + \text{---} + \text{---} + \text{---} \\
&= \int d^3\vec{r}_1 d^3\vec{r}_2 d^3\vec{r}_3 (f_{12}f_{23}f_{13} + f_{12}f_{13} + f_{12}f_{23} + f_{13}f_{23}) \\
b_4 &= \text{---} + \text{---} + \text{---} + \text{---} + \text{---} + \dots
\end{aligned}$$

Figure 7.3: Graphical representations of the first few cluster integrals

on the f 's, so we are interested in the number of ways of choosing these sets of clusters of various sizes¹⁰⁷. Let's call this number $W(\{n_l\})$, which we can compute as

$$W(\{n_l\}) = \frac{N!}{\prod_l n_l!(l!)^{n_l}}. \quad (7.21)$$

How did we arrive at this expression? Our strategy was essentially to first pick any particular partitioning into $\{n_l\}$ l -cluster. We first *overcount everything* by considering every possible permutation of the labels, giving us the $N!$ in the numerator above. Having overcounted, we now divide out to account for diagram symmetries. First, for each individual l -cluster, we note that the permutations of labels within that cluster give the same diagram (e.g., a line from point 1 to point 2 is unchanged when the labels 1 and 2 are permuted); this gives us a factor of $(l!)$ for each of the n_l clusters of size l . Additionally, we have to consider permutations of labels that replace all of the labels of one cluster of size l with a different cluster of size l (e.g., a line from point 1 to 2 and from 5 to 6 is unchanged when labels (1 and 5) and (2 and 6) are permuted); this gives us the additional factor of $n_l!$ in the denominator¹⁰⁸.

With this counting of the number of different types of diagrams, we can write the canonical partition function as

$$Z = \frac{1}{N!\lambda^{3N}} \sum_{\{n_l\}_{\text{restricted}}} W(\{n_l\}) \prod_l b_l^{n_l}, \quad (7.22)$$

where the restriction is that every point has to be in a cluster, $\sum_l l n_l = N$. From our discussion of ideal quantum gases, this problem should feel familiar: we find ourself trying to evaluate a canonical partition function involving a restricted sum; the restriction means we cannot deal with the different cluster sizes independently, and hence is making our life more difficult. In exactly the same way as before, this difficulty is lifted by moving to the grand canonical ensemble.

¹⁰⁷Note that this problem of partitioning a large number, N , into sets of integers is an interesting math problem. Look up, for instance, the Hardy-Ramanujan Asymptotic Partition Formula if you're interested.

¹⁰⁸I highly encourage you to get a feel for this by playing around. Try drawing, say, 5 or 6 points, break them up into clusters of different sizes, and get a feel for the combinatorics.

So, we write down the grand partition function, noting that the sum over all possible N of a restricted sum on the $\{n_l\}$ is equivalent to performing an *unrestricted sum* over the $\{n_l\}$:

$$\begin{aligned} \mathcal{Q} &= \sum_{N=0}^{\infty} e^{\beta\mu N} Z_N \\ &= \sum_{\{n_l\}} \frac{1}{N! \lambda^{3N}} (e^{\beta\mu})^{\sum_l n_l} N! \prod_l \frac{b_l^{n_l}}{n_l! (l!)^{n_l}} \\ &= \prod_l \sum_{n_l=0}^{\infty} \left(\frac{e^{\beta\mu}}{\lambda^3} \right)^{l n_l} \frac{b_l^{n_l}}{n_l! (l!)^{n_l}}. \end{aligned} \quad (7.23)$$

In the last line we recognize that we have written something like

$$\prod_l \sum_{n_l} \frac{1}{n_l!} [\dots]^{n_l},$$

which is just the expansion of an exponential. Thus, we find¹⁰⁹

$$\mathcal{Q} = \prod_l \exp \left[\left(\frac{e^{\beta\mu}}{\lambda^3} \right)^l \left(\frac{b_l}{l!} \right) \right], \quad (7.24)$$

i.e.,

$$\Rightarrow \log \mathcal{Q} = \sum_{l=1}^{\infty} \left(\frac{e^{\beta\mu}}{\lambda^3} \right)^l \left(\frac{b_l}{l!} \right) \quad (7.25)$$

Before we make use of this expression for the grand canonical partition function, it is worth thinking for a moment about the graphical expressions we wrote for cumulants and moments. The graphical interpretation of the above result is that the log of the sum over all graphs is equivalent to the sum over only the connected clusters; this is known as the “linked cluster theorem,” and frequently appears in diagrammatic perturbation techniques.

7.4 Virial expansion for a dilute gas

We consider, now, a dilute gas, and our goal for this section is to write a *virial expansion*¹¹⁰ for the equation of state of the dilute gas, i.e., something of the form

$$\beta P = B_1(T) \frac{N}{V} + B_2(T) \left(\frac{N}{V} \right)^2 + B_3(T) \left(\frac{N}{V} \right)^3 + \dots, \quad (7.26)$$

¹⁰⁹if you’re comparing these expression with Pathria [2], or other texts, please note that different authors use slightly different conventions in the definitions of the cluster integrals; in particular, sometimes the b_l account for the factors of $l!$ in that divide it in the expressions I’m writing. Here I’m following Kardar’s convention [3]

¹¹⁰So-called because the virial equation of state can also be derived by thinking about the contribution of the *virial*, $G = \sum_i^N \mathbf{q}_i \cdot \mathbf{f}_i$, to the pressure. This technical definition of the virial (related to the latin word for “force” or “energy”) was given by Clausius in 1870, and the virial expansion of the equation of state is typically credited to Kammerlingh Onnes, of superconductivity fame.

where we've written temperature-dependent coefficients B_i for the powers of density. For a dilute gas we can certainly exploit the extensivity condition to write the grand potential as

$$-\beta\mathcal{G} = \log \mathcal{Q} = -\beta(E - TS - \mu N) = \beta PV. \quad (7.27)$$

Comparing the with Eq. 7.25 we note that, in fact, each cluster integral $b_l \propto V$. We thus define a version of the cluster integrals with this volume dependence removed,

$$\bar{b}_l \equiv \frac{b_l}{V},$$

in terms of which we can write the pressure as

$$\beta P = \sum_l \left(\frac{e^{\beta\mu}}{\lambda^3} \right)^l \frac{\bar{b}_l}{l!}. \quad (7.28)$$

At the same time, we know that, specifying the chemical potential μ in the grand canonical ensemble we can compute the expected number density of particles as

$$n = \frac{N}{V} = \frac{1}{V} \frac{\partial \log \mathcal{Q}}{\partial (\beta\mu)} = \sum_l l \left(\frac{e^{\beta\mu}}{\lambda^3} \right)^l \frac{\bar{b}_l}{l!}. \quad (7.29)$$

Simplifying the notation by writing a dimensional version of the fugacity, $x \equiv e^{\beta\mu}/\lambda^3$, we arrive at the equation of state for a dilute gas:

$$\begin{aligned} \beta P &= \sum_l x^l \frac{\bar{b}_l}{l!} \\ n &= \sum_l l x^l \frac{\bar{b}_l}{l!}. \end{aligned} \quad (7.30)$$

This is... not exactly what we typically want: rather than an expression for pressure in terms of number density, we have pressure in terms of the fugacity *and* number density in terms of the fugacity. The situation is exactly the same as when we encountered the high-temperature limit of ideal quantum gases, and our way forward will *also* be exactly the same.

So, we first write out the series expansion for n ,

$$n = \bar{b}_1 x + \bar{b}_2 x^2 + \frac{\bar{b}_3}{2} x^3 + \dots \quad (7.31)$$

Rearranging this, and noting that $\bar{b}_1 = 1$, we have

$$x = n - \bar{b}_2 x^2 - \frac{\bar{b}_3}{2} x^3 - \dots, \quad (7.32)$$

which we can self-consistently solve to any desired order by substituting the i th-order solution for x into the $(i + 1)$ th-order expression. Starting with the first order approximation $x_1 \approx n$,

we get

$$\begin{aligned}
 x_1 &= n + \mathcal{O}(n^2) \\
 x_2 &= n - \bar{b}_2 n^2 + \mathcal{O}(n^3) \\
 x_3 &= n - \bar{b}_2 (n - \bar{b}_2 n^2)^2 - \frac{\bar{b}_3}{2} (n - \bar{b}_2 n^2)^3 + \mathcal{O}(n^4) \\
 &= n - \bar{b}_2 n^2 + \left(2\bar{b}_2^2 - \frac{\bar{b}_3}{2} \right) n^3 + \mathcal{O}(n^4), \tag{7.33}
 \end{aligned}$$

and so on. We can now substitute this into the equation for the pressure; to get an expansion correct to order i we use the x_i expression obtained by the above procedure. So, for instance, to third order we write (collecting all terms up to third order in the number density)

$$\begin{aligned}
 \beta P &= \bar{b}_1 x + \frac{\bar{b}_2}{2} x^2 + \frac{\bar{b}_3}{3!} x^3 \\
 &= n - \frac{1}{2} \bar{b}_2 n^2 + \left(\bar{b}_2^2 - \frac{\bar{b}_3}{3} \right) n^3 + \mathcal{O}(n^4). \tag{7.34}
 \end{aligned}$$

In general, we have a result of the form

$$\beta P = n + \sum_{l=2}^{\infty} B_l(T) n^l, \tag{7.35}$$

as desired! The first term, $B_1 = 1$, reproduces the ideal gas result. The second term is

$$B_2(T) = -\frac{\bar{b}_2}{2} = -\frac{1}{2} \int d^3 r [e^{-\beta \phi(\mathbf{r})} - 1], \tag{7.36}$$

which diagrammatically is $-1/2$ times what we drew as a line between two points. We calculate the third term diagrammatically, as illustrated in Fig. 7.4, and find

$$B_3(T) = -\frac{1}{3} \int d^3 r_{12} d^3 r_{13} f(\mathbf{r}_{12}) f(\mathbf{r}_{13}) f(\mathbf{r}_{12} - \mathbf{r}_{13}) \tag{7.37}$$

$$\begin{aligned}
 B_3 &= \bar{b}_2^2 - \frac{\bar{b}_3}{3} = (\text{---}\bullet\text{---}\bullet\text{---})^2 - \frac{1}{3} (\text{---}\triangle\text{---} + \text{---}\nabla\text{---} + \text{---}\triangleleft\text{---} + \text{---}\triangleright\text{---}) \\
 &= -\frac{1}{3} (\text{---}\triangle\text{---})
 \end{aligned}$$

Figure 7.4: **Diagrammatic calculation of the third virial coefficient**

That calculation – in which many of the diagrams cancel out – is representative of a much more general result:

$$B_l(T) = -\frac{l-1}{l!} \bar{d}_l, \tag{7.38}$$

where the \bar{d}_l represents a sum over all “irreducible clusters of size l .” These irreducible l -clusters¹¹¹ are those in which *there are at least two independent and non-intersecting paths that link every pair of nodes in the diagram*. We can find these by considering the set of all l -clusters and then discarding the “1-particle reducible diagrams,” those diagrams in which the removal of one point leads to disjoint clusters.

We pause to note that actually computing B_l for $l > 2$ is typically¹¹² a fairly unpleasant¹¹³ and not particularly rewarding task. We know (from Chapter 3) that computing higher order terms in the virial equation of state does not improve our ability to understand what’s going on near the phase transition, but it is worth commenting that it *does* allow one to systematically estimate the critical point, i.e., where the phase transition actually happens (in addition, of course, to more and more accurately describing dilute gases).

In the next section we’ll evaluate these sorts of virial coefficients for a “typical” gas, i.e., one governed by a potential similar to a Lennard-Jones form.

7.5 The van der Waals equation

To see how (or, indeed, if!) the sort of virial expansion we just worked out is useful, let’s consider just the first correction to the ideal gas equation of state – the term involving $B_2(T)$ – for the Lennard-Jones potential,

$$\phi(\mathbf{r}) = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right], \quad (7.39)$$

which, again, is a simple but reasonable approximation to the interaction between two neutral molecules separated by some distance r .

7.5.1 The second virial coefficient for a Lennard-Jones interaction

The most straightforward thing we could do is simply to use the definition of the second virial coefficient we just derived:

$$\begin{aligned} B_2(T) &= \frac{-1}{2} \int d^3r (e^{-\beta\phi(r)} - 1) \\ &= -2\pi \int dr r^2 (e^{-\beta\phi(r)} - 1). \end{aligned} \quad (7.40)$$

¹¹¹Also called biconnected graphs, or a graph which is connected and non-separable: i.e., a graph in which the removal of any one vertex leaves a connected graph behind.

¹¹²For a long time, B_5 was basically the limit of what could be done for the Lennard-Jones potential. Recent advances in Monte Carlo sampling techniques of the relevant integrals [62] have let people accurately go beyond that – at least to B_{16} – across a range of temperatures.

¹¹³After all, how many different integrals do you need to evaluate for \bar{d}_l ? [Here’s a link](#) to the integer sequence with the answer to that question for $3 \leq l \leq 19$. Notice that, apparently, \bar{d}_{19} requires evaluating of order 2.46×10^{34} terms!

That integral looks unfamiliar, but one can express the answer in terms of the Kummer confluent hypergeometric function¹¹⁴ ${}_1F_1(a; b; z)$, and the Gamma function:

$$B_2(T) = \frac{-\pi\sigma^3(\beta\varepsilon)^{1/4}}{3\sqrt{2}} \left[\Gamma\left(\frac{-1}{4}\right) {}_1F_1\left(\frac{-1}{4}; \frac{1}{2}; \beta\varepsilon\right) + 2\sqrt{\beta\varepsilon}\Gamma\left(\frac{1}{4}\right) {}_1F_1\left(\frac{1}{4}; \frac{3}{2}; \beta\varepsilon\right) \right] \quad (7.41)$$

Is this helpful? This is, indeed, what you get by plugging things into, e.g., Mathematica, and you could plot the behavior of this second virial coefficient as you vary the temperature for various values of the Lennard-Jones parameters. This “plug-and-chug” approach, while technically correct, obscures the physical picture of how we expect a dilute gas to behave.

7.5.2 Approximate but physical treatment of B_2

Let’s do a better job of capturing the physics by making a *rougher* approximation for the molecular interactions. In particular, let’s replace the Lennard-Jones potential with the function

$$\phi(r) = \begin{cases} \infty & r < \sigma \\ -\varepsilon \left(\frac{\sigma}{r}\right)^6 & r > \sigma \end{cases}. \quad (7.42)$$

This approximate potential captures the following main features of the Lennard-Jones potential: it has the van der Waals r^{-6} at large distances, it has an attractive well of depth ε at short distances, and it is strongly (in fact, *infinitely!*) repulsive within a characteristic size of the molecular core. Note that this simpler form makes it clear that σ is like the characteristic *diameter* of one of the interacting molecules.

With this simpler potential we can compute the second virial coefficient as

$$\begin{aligned} B_2(T) &= -\frac{1}{2} \int d^3r (e^{-\beta\phi(r)} - 1) \\ &= -\frac{1}{2} \left[-\frac{4\pi\sigma^3}{3} + 4\pi \int_{\sigma}^{\infty} dr r^2 \left(e^{\beta\varepsilon\left(\frac{\sigma}{r}\right)^6} - 1 \right) \right]. \end{aligned} \quad (7.43)$$

The first term above reflects an *excluded volume* in the potential we have written down. Keeping in mind that we are trying to describe a dilute gas at reasonably high temperatures, let’s further approximate the second term by assuming $\beta\varepsilon \ll 1$, that is, that we are at high T relative to the well depth. We will use this approximation to write a series expansion for the exponential of the potential:

$$e^{\beta\phi} - 1 \approx \beta\varepsilon \left(\frac{\sigma}{r}\right)^6 + \mathcal{O}(\beta\varepsilon)^2.$$

With this approximation, the integral can be easily carried out:

$$\begin{aligned} B_2(T) &= -\frac{1}{2} \left[\left(-\frac{4\pi\sigma^3}{3} \right) + \frac{4\pi\sigma^3}{3} \beta\varepsilon \right] \\ &= \frac{\Omega}{2} [1 - \beta\varepsilon], \quad \text{where } \Omega = \frac{4\pi\sigma^3}{3}. \end{aligned} \quad (7.44)$$

¹¹⁴As a series expansion, ${}_1F_1(a; b; z) = \sum_{n=0}^{\infty} \frac{a^{(n)} z^n}{b^{(n)} n!}$, where $a^{(n)}$ is the “rising factorial,” $a^{(0)} = 1$, $a^{(1)} = a$, $a^{(n)} = a(a+1)\cdots(a+n-1)$, etc.

This is a nice, compact expression for the second virial coefficient: there is a small perturbative effect at high temperatures that corrects an overall scale set by the excluded volume of the molecules.

A few comments are in order, directed at the question of “*when should this approximation be reasonable?*” Two obvious conditions stand out.

First, this is clearly meant to be a low-density expansion, but what we do mean by “low” in this context? As with any series expansion, we want the ratio of consecutive terms to be small, and the ratio of the correction to the ideal term is $\sim B_2 \frac{\rho^2}{\rho} \sim \frac{\Omega}{\rho^{-1}}$. Note that Ω is close to an atomic volume, so one can think of the above ratio as $\sim \frac{\rho_{gas}}{\rho_{liquid}}$. That is, when the density of the system is such that the gas is close to liquifying, we should assume our series approximation is insufficient.

Second, in our expansion of the attractive tail, we assumed that $\beta\varepsilon \ll 1$. If $\beta\varepsilon \gtrsim 1$ for these attractive potentials, we could not have done a reasonable series expansion in the first place. In general, for low temperatures the ground state of a system with attractive interactions is not a gas, but a dense collection of molecules sticking together.

There is an additional consideration, which speaks to our ability to factor out the Ω term in the series expansion we wrote. The fact that we could do this speaks to the sense that the *short-ranged* part of the potential dominates the integral we had to do. In this context there is a sense in which potentials which fall off faster than r^{-3} are “short ranged;” for potentials like $\phi \sim r^{-3}$ there are terms which are logarithmic in the size of the system, and for even longer-ranged potentials the whole expansion fails.

7.5.3 The van der Waals equation

Finally, we can write down the equation of state based on our approximate treatment of the second virial coefficient:

$$\begin{aligned} \beta P &= n + \frac{n^2}{2} \Omega (1 - \beta\varepsilon) \\ \Rightarrow \beta \left(P + \frac{n^2 \Omega}{2} \varepsilon \right) &= n \left(1 + \frac{n\Omega}{2} + \dots \right). \end{aligned} \quad (7.45)$$

Since we’re only being accurate to order n^2 in the above expression, we’re free to re-write the right-hand side as

$$n \left(1 + \frac{n\Omega}{2} + \dots \right) \approx \frac{n}{1 - \frac{n\Omega}{2}} + \dots$$

Using this rearranging, we arrive at the van der Waals equation of state:

$$\boxed{P = \frac{Nk_B T}{V - \frac{N\Omega}{2}} - \frac{\varepsilon\Omega}{2} \left(\frac{N}{V} \right)^2}, \quad (7.46)$$

which is often presented as

$$\left(P + a \left(\frac{N}{V} \right)^2 \right) (V - bN) = Nk_B T.$$

7.6 Problems

7.6.1 Every picture an integral

In Section 7.4 of the lecture notes we showed that the second virial coefficient,

$$B_2(T) = -\frac{\bar{b}_2}{2}.$$

This meant we only needed to evaluate a single diagram (the integral over a single f -function). The third virial coefficient looked like it had more terms,

$$B_3(T) = \bar{b}_2^2 - \frac{\bar{b}_3}{3},$$

but we showed that it could be re-written in terms of a single diagram (negative one-third times “triangle” diagram, which stood for a particular integral). Expressed in terms of the cluster integrals, the direct expression for the 4th virial coefficient is

$$B_4(T) = -\frac{\bar{b}_4}{8} + \frac{3}{2}\bar{b}_2\bar{b}_3 - \frac{5}{2}\bar{b}_2^3.$$

That looks like a lot of pictures – i.e., integrals – to evaluate!

(A) Fortunately, we learned that when evaluating physical expressions that a lot of those pictures cancel with each other and we really only need the sum over *one-particle irreducible clusters of size l* . So, without doing any work beyond drawing pictures, tell me the number of different diagrams you actually need to write down in order to evaluate $B_4(T)$. Draw them for me.

Note: If you choose to answer this question “up to permutations of the labels,” make sure you tell me how many diagrams of each type you get. For instance, in the lecture notes I drew \bar{b}_3 as 4 diagrams; I also could have said it was the complete triangle plus three permutations of “triangle missing one edge.”

(B) Choose any one of your diagrams from part (A), and write down the integral expressions that this diagram corresponds to. (*this answer should be a single line – no evaluation or explanation needed*)

(C) We’ve seen that $B_1(T) = 1$, and that both $B_2(T)$ and $B_3(T)$ require evaluating just a single diagram. In part (A) you calculated the number of diagrams needed to calculate $B_4(T)$. Based on this, and thinking about how these one-particle irreducible clusters can be formed, please speculate on the general trend here. How bad is the growth in the number of terms you need to calculate if you wanted to know higher and higher virial coefficients, $B_i(T)$?

7.6.2 Terms in the cumulant expansion

In Section 7.2 of the lecture notes we looked at an interacting system and saw how the first two cumulants contributed to a perturbation series for the equation of state as a function of

density, where the coefficient of each term in the density expansion was itself an expansion in powers of β . This gave rise to terms that could be of order β^2 or $(N/V)^2$.

Write an expression for the third cumulant. By carefully considering the various cases involving similar or different particle indices, determine how the third cumulant contributes to the expansion of the partition function in powers of density and inverse temperature. There is an exhaustive way to do this problem, which involves checking a large number of individual cases. I am hoping you all will start seeing a pattern in the kinds of terms that do and do not cancel out. Please make use of intermediate results — including, e.g., what cancels in the second cumulant calculation — and know that you can get full credit without including every last manipulation of the integrals once it is clear how they behave.

7.6.3 Virial expansion for a repulsive gas

Consider a classical gas of N indistinguishable particles that interact according to the following purely repulsive pairwise:

$$\phi(r) = \begin{cases} k \left(1 - \frac{r}{\sigma}\right) & r \leq \sigma \\ 0 & r > \sigma \end{cases}.$$

That is: when particles are farther than σ away from each other they don't interact, and otherwise they feel a linearly increasing potential as they approach.

(A)

I would like to know the equation of state of this system at high temperatures and relatively low densities. Please perform a virial expansion, keeping terms up to order ρ^2 and β^2 (where $\rho = N/V$ and $\beta = (k_B T)^{-1}$, as usual).

(B)

Suppose you wanted to find the critical point (for the liquid-gas transition) for this model — i.e., the simultaneous values of the critical density, critical pressure, and critical temperature — do you already have enough information? Why or why not?

(C)

In class (and in the lecture notes) we calculated the second virial coefficient for an approximate Lennard-Jones interaction and then did some manipulations (with the help of an uncontrolled approximation) to the van der Waals equation of state. For this model, if you didn't want to make uncontrolled approximations, what additional terms (if any) would you calculate so that you could find the critical point?

7.6.4 Joule-Thomson for a dilute gas

The virial equation of state is

$$\frac{P}{k_B T} = \frac{N}{V} + B_2(T) \frac{N^2}{V^2} + B_3(T) \frac{N^3}{V^3} + \dots,$$

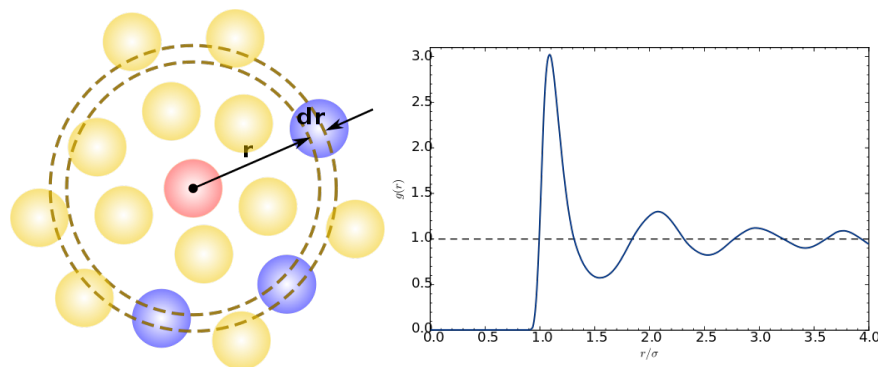


Figure 7.5: **Radial distribution function** (Left) Counting particles with a shell of thickness dr at a distance r away from a reference particle. In 3D, the probability of finding a particle in such a shell, given that a particle is at the origin, is $4\pi\rho r^2g(r)dr$. (Right) The radial distribution function for a Lennard-Jones fluid at some particular density and temperature. [Both figures from [wikipedia](#).]

where the B_i were temperature-dependent coefficients that can be computed for any particular inter-particle potential. By doing perturbation theory (what we called an “order-by-order expansion”), write a second-order expression for V as a function of T , P , N , and B_2 , to show that the *Joule-Thomson coefficient*, μ equiv $\left. \frac{\partial T}{\partial P} \right|_H$, can be written to this order as

$$\mu = \frac{N}{C_P} \left(T \frac{\partial B_2}{\partial T} - B_2 \right).$$

Hint: the thermodynamic manipulations of Problem 1.8.8 may be helpful!

7.6.5 Radial distribution function and virial expansions

The *radial distribution function*, $g(r)$, describes the probability of finding a particle in your system at some position relative to that of a given reference particle. It is defined by first considering the two-particle density,

$$\rho_2(\mathbf{r}, \mathbf{r}') = \left\langle \sum_{i \neq j} \delta(\mathbf{r} - \mathbf{r}_i) (\mathbf{r}' - \mathbf{r}_j) \right\rangle.$$

Assuming the system is translationally and rotationally symmetric, $\rho_2(\mathbf{r}, \mathbf{r}') \rightarrow \rho_2(|\mathbf{r} - \mathbf{r}'|)$, and we define the radial distribution function (also called the “pair correlation function”) by normalizing this by the square of the density $\rho = N/V$:

$$g(|\mathbf{r} - \mathbf{r}'|) = \frac{\rho_2(\mathbf{r}, \mathbf{r}')}{\rho^2}.$$

A schematic cartoon and a plot of $g(r)$ are shown in Fig. 7.5 This quantity plays an important role, both as a characterization of the structure of fluids and because one can expressions for thermodynamic quantities (pressure, energy, etc) in terms of the pair correlation function:

this permits an explicit link between microscopic structure and macroscopic behavior. In this problem we will explore a virial expansion approach to computing $g(r)$ for dilute gases¹¹⁵.

(A) Working in the grand canonical ensemble and from the definition of the ensemble average, show that you can write the two-particle density function as

$$\rho_2(\mathbf{r}_{12}) = \frac{1}{\mathcal{Q}(\mu, T, V)} \sum_{N=2}^{\infty} \frac{x^N}{(N-2)!} \int d^3r_3 \cdots d^3r_N \exp(-\beta\mathcal{U}(\{\mathbf{r}\})),$$

where $x = \exp(\beta\mu)/\lambda^3$

(B) Approximate the two-particle density by the first two terms in this sum (i.e., the $N = 2$ and $N = 3$ terms) and assume that $\mathcal{U} = \sum_{i<j} \phi(\mathbf{r}_{ij})$ for some pair potential ϕ . Noting that \mathcal{Q} can also be written as a power series in x , show that you can write this in the form

$$\rho_2(\mathbf{r}_{12}) = e^{-\beta\phi(\mathbf{r}_{12})} w(\mathbf{r}_{12}),$$

where $w(r)$ is a function of x and integrals over various Mayer f functions. Make sure you keep all terms up to order x^3 .

(C) Finally, using a low density expansion for x , show that you can write the radial distribution function as

$$g(r) = \exp(-\beta\phi(\mathbf{r})) y(r),$$

Your function $y(r)$ should be written in terms of the density and integrals over various Mayer f functions. Keep all terms up to those linear in ρ .

7.6.6 A one-dimensional gas

In this problem we explore a 1d interacting model that can be solved *exactly*. Consider a gas of (classical) particles confined to a line of total length L . We'll imagine that each particle has diameter σ , and that the particles have hard-core interactions – that is, $\phi(r) = 0$ if $r > \sigma$ and $\phi(r) = \infty$ if $r < \sigma$. The particles are indistinguishable, and their positions can be labeled with coordinates $\{x_i\}$, where $\sigma/2 \leq x_1 \leq x_2 \leq \cdots \leq x_N \leq L - \sigma/2$.

(A) Write down the definition of the canonical partition function, and be sure to carefully write down the part associated with the allowed positions of the particles, indicating the allowed range of integration.

(B) Consider a change of variables such as $y_j = x_j - (2j - 1)\sigma/2$; by using this (or something similar) to assist in evaluating integrals, what is the canonical partition function $Z(N, L, T)$? What is the Helmholtz free energy?

¹¹⁵For working in dense regimes where this kind of perturbative approach doesn't seem particularly appealing, see Hansen and McDonald, *Theory of Simple Liquids* [63] for a variety of ways for writing down approximate theoretical treatments for the structure of dense liquid phases if you're interested.

(C) What is the pressure of this gas? If you wanted to write the equation of state as a virial expansion, what would the virial coefficients, $B_l(T)$, be? What is the isothermal compressibility?

(D) You have just done an *exact* calculation for the statistical mechanics and thermodynamics of a 1D gas. Compare what you have learned about the phase behavior of this system to the *mean field* behavior we studied for the van der Waals gas.

Chapter 8

Interacting systems: Variational approaches

Introductory text goes here. Basic idea: Most of the models that can be solved exactly according to the “partition function, thermodynamic potential, properties” pattern of the previous chapter are all non-interacting ones in which very little of interest happens. There are two particularly fruitful ways forward: we can choose to try to systematically compute better and better approximations to the partition function — Chapter 7 — or we can try to make physically motivated by uncontrolled approximations for either the partition function or, sometimes, directly for the free energy. In this chapter, we’ll take that latter approach and learn a bit about writing mean field approximations for F .

8.1 Mean field Magnets

We saw in chapter 6 that in the canonical ensemble, calculating the properties of non-interacting systems becomes quite straightforward. Perhaps this is not so surprising: since the particles are non-interacting, calculating the N -particle partition function reduces to calculating a 1-particle partition function, and single-particle problems in physics are usually quite tractable. This inspires a strategy for analyzing more complicated, interacting problems by trying to model them *as a non-interacting* systems governed by some effective average (or “mean”) field.

Let’s see this in action. As an intuitive introduction to the idea, in this subsection we’ll present a rather hand-waving argument for defining a mean field theory. If you find this unsatisfying, wait until section 8.3, in which we will derive exactly the same theory from a more principled variational argument.

We yet again write the Ising model Hamiltonian,

$$\mathcal{H} = -J \sum_{\langle ij \rangle} s_i s_j - B \sum_i s_i, \quad (8.1)$$

where J determines the strength of the spin-spin interaction, B is the external field, and $\sum_{\langle ij \rangle}$ indicates a sum over all spins i and j that are neighbors of each other. To make progress,

let's approximate the spin-spin interactions as

$$\begin{aligned} s_i s_j &= ([s_i - m] + m) ([s_j - m] + m) \\ &= (s_i - m)(s_j - m) + m((s_i - m) + (s_j - m)) + m^2 \\ &\approx m(s_i + s_j) - m^2. \end{aligned} \tag{8.2}$$

That is, we assume that the fluctuations of spins away from the average spin (m) are small on average. Of course, it is not literally the case that for each i, j we consider that $(s_i - m)(s_j - m)$ is small – it certainly need not be, since the s variables are ± 1 but the value of m might be anywhere between 0 and 1 – but rather than when we sum over all interactions, $\sigma_{\langle ij \rangle}$, these terms can be neglected.

With this approximation, we can write our mean field Hamiltonian as

$$\mathcal{H} \approx \frac{JNqm^2}{2} - (Jqm + B) \sum_i s_i, \tag{8.3}$$

where q is the number of nearest neighbors ($q(d) = 2d$ for the hypercubic lattice). And we've done it! We've transformed our problem into a *non-interacting* Ising model in an effective external field, where

$$B_{eff} = B + Jqm.$$

We then take the free energy F to be

$$F = \log \left(2^N e^{-\beta JNqm^2/2} \cosh^N (\beta B_{eff}) \right). \tag{8.4}$$

Finally, we self-consistently determine what the value of m is by connecting the definition with the thermodynamic result: $m = \frac{1}{N} \langle s_i \rangle = \frac{1}{N\beta} \frac{\partial F}{\partial B}$.

$$m = \tanh(\beta B + \beta Jqm). \tag{8.5}$$

This is precisely the Curie-Weiss equation of state for the magnet that we investigated when we looked at phase transitions in section 2.4.1!

8.2 The van der Waals equation as a mean field theory

We met the van der Waals equation of state,

$$P = \frac{k_B T}{\rho^{-1} - b} - a\rho^2,$$

in Chapter 2, where you might have felt a certain tension in how we motivated it and how we used it. On the one hand, we thought of it as an extension of an ideal-gas like equation of state that captured information about particle-particle interactions (hence the inclusion of a ρ^2 term). The functional form is perhaps not exactly what we might expect from a perturbative approach (since, in particular, the first term on the right hand side of the above equation is not just proportional to the density), but it certainly feels like it's in that spirit¹¹⁶.

¹¹⁶We'll look forward to precisely such a calculation in Section 7.5.3!

On the other hand: we used the van der Waals equation (with the help of a Maxwell-construction adjustment) to describe the dilute gas phase, the dense fluid phase, *and* as a mean field description of liquid-gas critical behavior. Those latter two regimes feel quite far from the kind of limit where a perturbative approach is supposed to work at all, let alone work well. To see why we might nevertheless think the van der Waals equation might be a reasonable theory for both a dilute gas and a dense liquid, let's approach it from a different perspective — similar in spirit to the mean-field discussion of magnets in Sec. 8.1.

Our approach will be not to focus on an expansion in the density, but rather to think of both liquids and gases as having a *uniform density*. Our goal is to start with our usual expression for the partition function,

$$Z = \int d\Gamma \exp \left(-\beta \left(\sum_i \frac{p_i^2}{2m} + \sum_{ij} \phi_{i,j} \right) \right),$$

where ϕ is some pairwise interaction, and see how an assumption of uniform density simplifies the otherwise intractable problem of computing the position integrals. The actual distribution of particle density can be written as a collection of delta functions picking out the location of all of the particles in the system,

$$\rho(\mathbf{q}) = \sum_i \delta(\mathbf{q} - \mathbf{q}_i),$$

and our uniform density approximation will be to simply set

$$\rho(\mathbf{q}) \approx \rho = N/V.$$

Notice that we can write the potential energy part of the Hamiltonian in terms of the distribution of particle density,

$$U(\{\mathbf{q}_i\}) = \frac{1}{2} \int d^3q d^3q' \rho(\mathbf{q})\rho(\mathbf{q}')\phi(\mathbf{q} - \mathbf{q}'). \quad (8.6)$$

Let's let the actual pairwise potential be quite general, asking only that it has three features. First, we'll assume that it has some sort of "hard-core" character for $r < \sigma$ that gives the particles a length scale over which they cannot overlap. Second, we'll assume that the potential is sufficiently short-ranged that any integrals we work with converge. Finally, we'll assume that there is some attractive part of the potential — we'll quantify this by saying the integrated volume of the potential well is $-u$. With all of that, we can write the mean-field "typical energy", \tilde{U} , as

$$\tilde{U} = \frac{\rho^2}{2} \int d^3q d^3q' \phi(\mathbf{q} - \mathbf{q}') = \frac{\rho^2 V}{2} \int d^3r \phi(r) = \frac{-u\rho^2 V}{2}. \quad (8.7)$$

The fact that this is a constant means we can pull terms involving it out of the integrals defining the partition function, and of course we already know how to do the Gaussian integrals associated with the momenta. What we have left is

$$Z = \frac{1}{N!\lambda^{3N}} \exp \left(\frac{\beta u N^2}{2V} \right) \int \prod_i d^3q_i. \quad (8.8)$$

The quotation marks are meant to indicate that we were a little bit sloppy in evaluating eq. (8.7), as we forgot to account for the hard-core interactions preventing particle overlaps. If the repulsions were absent we would be done, with each of the remaining integrals contributing one power of V . Let's instead approximately treat the excluded volume effects to linear order in a quantity Ω/V , where $\Omega = \frac{4\pi\sigma^3}{3}$ represents a characteristic particle volume.

The naive approach to packing particles in a volume (“the first particle can go anywhere, the second particle can go anywhere except in the excluded volume of the first, the third...”) would lead to

$$\int \prod_i d^3r_i \approx V(V - \Omega)(V - 2\Omega) \cdots (V - (N - 1)\Omega). \quad (8.9)$$

This clearly neglects higher-order effects (i.e., consider two particles which are just outside of each others' excluded volume – in this configuration $V - 2\Omega$ is an incorrect estimate of the volume available to a third particle), but is a good starting point when the densities are uniformly distributed. We can group together terms in the above product in pairs chosen about the center of the series, looking at the product of the first and the last terms, and the second and second-to-last terms, etc. Each of these products will look like

$$V^2 - \Omega V(N - 1) + \mathcal{O}(\Omega^2) \approx \left(V - \frac{N\Omega}{2}\right)^2,$$

where we play with and then ignore any term of order Ω^2 , and set $N \approx N - 1$. Since there are $N/2$ such pairs, we approximate the whole excluded volume contribution by

$$\int \prod_i d^3r_i \approx \left(V - \frac{N\Omega}{2}\right)^N. \quad (8.10)$$

Combining this estimate of the available volume with our mean-field estimate of the effect of the attractions (and the usual momentum piece), we have our approximated partition function:

$$Z(V, T, N) \approx \frac{\left(V - \frac{N\Omega}{2}\right)^N}{N! \lambda^{3N}} \exp\left(\frac{\beta u N^2}{2V}\right). \quad (8.11)$$

With this in hand we can compute the pressure as

$$\beta P = k_B T \frac{\partial \log Z}{\partial V} = \frac{N}{V - \frac{N\Omega}{2}} - \beta \frac{u N^2}{2V^2} \quad (8.12)$$

or, rearranging terms, as

$$P = \frac{k_B T}{\rho^{-1} - \frac{\Omega}{2}} - \frac{u}{2} \rho^2. \quad (8.13)$$

This is precisely the van der Waals equation, and here we see it not in the context of expanding about a small parameter (perhaps density, or inverse temperature), but in the context of an *approximation of uniform density*. Certainly a uniform density approximation is the sort of thing we expect to be equally valid in both the gas and liquid states (and probably *not* at the transition between the two!), hence why we expect that we ought to be able to describe multiple phases with the same Hamiltonian and the same partition function.

But now we can check the self-consistency of our condition! We assume density is uniform, but we can use vdW to estimate the number fluctuations! (density fluctuations are prop to κ_T , and from chapter 2 κ_T diverges at the critical point, etc).

8.3 A variational view of mean field theories

The derivation of mean field theories often has this flavor of physically motivated by a priori uncontrolled approximations. There is another way of thinking about and deriving mean field theories, which brings in the rich history of exactly solvable models in statistical physics. You probably (and correctly!) have begun to suspect that actually carrying out the idealized program of statistical mechanics — taking a set of degrees of freedom and their governing Hamiltonian, exactly evaluating the partition function, and deriving thermodynamic quantities from there — ranges in difficulty for interesting problems from extraordinarily hard to impossible. There are, indeed, only a precious few models that we know how to solve exactly [citations and expand this a bit](#).

These exactly solvable models have historically played an important role in statistical physics, in no small part because these exactly solvable models can be used as *building blocks* in approximate solutions to the systems we actually care about. Back in section 8.1 we presented a quite hand-waving “derivation” of the Curie-Weiss mean field equation of state for a magnet. With can get rid of the hand-waving and show this same mean field theory can in fact be understood more rigorously: there is a specific sense in which that mean field theory is the *best possible* mapping from the system we care about (here, the Ising Hamiltonian with nearest-neighbor spin-spin interactions) to a system we can solve (in this case, non-interacting spins in an external field). We’ll demonstrate this in two steps.

8.3.1 Bogoliubov inequality

Our strategy is going to be to approximate the true Hamiltonian governing our system, \mathcal{H} , with a simpler Hamiltonian that we are able to calculate quantities of interest, \mathcal{H}_0 , in a systematic way. Here we’ll derive a classical version of the so-called “Bogoliubov variational theorem¹¹⁷” – a more general discussion and derivation that applies when \mathcal{H} and \mathcal{H}_0 do not commute is in Feynman’s text [65].

[Choose different symbol instead of \$\lambda\$](#)

Let’s imagine defining a new Hamiltonian that interpolates between \mathcal{H}_0 and \mathcal{H} as we vary some parameter λ from zero to one:

$$\mathcal{H}(\lambda) = \mathcal{H}_0 + \lambda(\mathcal{H} - \mathcal{H}_0). \quad (8.14)$$

The canonical free energy associated with this new Hamiltonian is, of course,

$$F(\lambda) = -\beta^{-1} \log Z = -\beta^{-1} \log \sum_{\mu} e^{-\beta \mathcal{H}(\lambda, \mu)},$$

where we have adopted the notation $\mathcal{H}(\lambda, \mu)$ to mean the value of $\mathcal{H}(\lambda)$ in state μ . What can we say about the behavior of $F(\lambda)$ as λ is varied? Well, the first derivative is

$$\frac{dF(\lambda)}{d\lambda} = \frac{1}{Z} \sum_{\mu} ((\mathcal{H} - \mathcal{H}_0) \exp[-\beta \mathcal{H}(\lambda, \mu)]) = \langle \mathcal{H} - \mathcal{H}_0 \rangle, \quad (8.15)$$

¹¹⁷Seemingly named by Griffiths [64]. I think it’s [Bogoliubov Jr.](#)

where we emphasize (to contrast with a result we are about to get to) that $\langle \dots \rangle$ refers to an average with respect to states weighted by the Boltzmann factor associated with $\mathcal{H}(\lambda)$. Proceeding on, the second derivative is

$$\frac{d^2 F(\lambda)}{d\lambda^2} = -\beta \left(\frac{\sum_{\mu} ((\mathcal{H} - \mathcal{H}_0)^2 \exp[-\beta \mathcal{H}(\lambda, \mu)])}{Z} - \frac{\left(\sum_{\mu} ((\mathcal{H} - \mathcal{H}_0) \exp[-\beta \mathcal{H}(\lambda, \mu)]) \right)^2}{Z^2} \right).$$

A bit unwieldy, but this second derivative is just a difference between the second moment and the square of the first moment, i.e., a variance!

$$\frac{d^2 F(\lambda)}{d\lambda^2} = -\beta [\langle (\mathcal{H} - \mathcal{H}_0)^2 \rangle - \langle \mathcal{H} - \mathcal{H}_0 \rangle^2] = -\beta \langle (\mathcal{H} - \mathcal{H}_0)^2 \rangle_c. \quad (8.16)$$

We immediately deduce that this second derivative is non-positive for *any* value of λ . Among other things, this tells us that as a function of λ , $F(\lambda)$ can never be greater than the straight line which is tangent to $F(\lambda)$ at $\lambda = 0$:

$$F(\lambda) \leq F(0) + \lambda \left[\frac{dF(\lambda)}{d\lambda} \right]_{\lambda=0}.$$

In particular, if we now *evaluate* this inequality at $\lambda = 1$ we recover Bogoliubov's¹¹⁸ inequality¹¹⁹:

$$F \leq \mathcal{F}_0 + \langle \mathcal{H} - \mathcal{H}_0 \rangle_0 \equiv \mathcal{F}. \quad (8.17)$$

That is: the free energy associated with a given Hamiltonian \mathcal{H} is less than or equal to the free energy associated with a (typically simpler) Hamiltonian \mathcal{H}_0 plus the average value of the difference *as calculated for states that are weighted according to the Hamiltonian \mathcal{H}_0* . It is this last clause that gives the inequality its applicability: for a sufficiently simple trial Hamiltonian (say, one which contains no interactions between the degrees of freedom) it is often much easier to evaluate ensemble averages with respect to \mathcal{H}_0 than it is with respect to \mathcal{H} .

8.3.2 Mean field magnets redux

The inequality above provides what is often the most elegant way of deriving mean field theories. It says that the true free energy of the system is always less than (or, at best, equal to) this funny combination of approximate free energy associated with a trial Hamiltonian. Said another way, this gives us a prescription for choosing the best approximate Hamiltonian for our system: choose an \mathcal{H}_0 that lets you calculate things analytically, and then tune any free parameters in \mathcal{H}_0 so as to minimize \mathcal{F} . This makes the inequality in Eq. 8.17 as tight a bound as possible, providing a quantitative measure of choosing an optimal simpler theory. That is, our mean field free energy will be

$$\mathcal{F}_{mean \ field} = \min_{\mathcal{H}_0} (\mathcal{F}_0 + \langle \mathcal{H} - \mathcal{H}_0 \rangle_0). \quad (8.18)$$

¹¹⁸Sometimes, in slightly different form, called Gibbs' inequality

¹¹⁹Shifting notation slightly, with $\mathcal{F}_0 \equiv F(0)$.

Let's see this program in action as we approximate the nearest-neighbor Ising model Hamiltonian with the best possible non-interacting spin model. Our system Hamiltonian is

$$\mathcal{H} = -J \sum_{\langle ij \rangle} s_i s_j - B \sum_i s_i,$$

and we will use the trial Hamiltonian

$$\mathcal{H}_0 = -B_{eff} \sum_i s_i.$$

We know from our work in Sec. 6.4 that these non-interacting systems are easy to work with, and we immediately can write down the trial Hamiltonian's free energy

$$\mathcal{F}_0 = -Nk_B T \log [2 \cosh (\beta B_{eff})]$$

and the trial Hamiltonian's magnetization per spin (for any of the spins) is

$$m_0 = \langle s_i \rangle_0 = \tanh \beta B_{eff}.$$

We next need to evaluate $\langle \mathcal{H} - \mathcal{H}_0 \rangle_0$ which, again, denotes an average taken in the ensemble defined by \mathcal{H}_0 . We start by writing this definitionally:

$$\langle \mathcal{H} - \mathcal{H}_0 \rangle_0 = \frac{\sum_{\{s_i = \pm 1\}} \left(-J \sum_{\langle ij \rangle} s_i s_j - B \sum_i s_i + B_{eff} \sum_i s_i \right) \exp [\beta B_{eff} \sum_i s_i]}{\sum_{\{s_i = \pm 1\}} \exp [\beta B_{eff} \sum_i s_i]}. \quad (8.19)$$

We see that this all is just

$$\langle \mathcal{H} - \mathcal{H}_0 \rangle_0 = -J \sum_{\langle ij \rangle} \langle s_i s_j \rangle_0 + (B_{eff} - B) \sum_i \langle s_i \rangle_0, \quad (8.20)$$

but we further note that \mathcal{H}_0 contains no interaction terms, so in fact $\langle s_i s_j \rangle_0 = \langle s_i \rangle_0 \langle s_j \rangle_0 = m_0^2$. For a lattice of N spins where each site has q nearest neighbors we count up the number of terms in $\sum_{\langle ij \rangle}$, and get

$$\langle \mathcal{H} - \mathcal{H}_0 \rangle_0 = -\frac{qJN}{2} m_0^2 + N(B_{eff} - B) m_0. \quad (8.21)$$

Substituting our expression for m_0 , we finally have the free energy

$$\mathcal{F} = -Nk_B T \log [2 \cosh (\beta B_{eff})] - \frac{qJN}{2} \tanh^2 \beta B_{eff} + N(B_{eff} - B) \tanh \beta B_{eff}. \quad (8.22)$$

All that remains is to minimize this with respect to our choice of B_{eff} . Taking a derivative of \mathcal{F} with respect to B_{eff} and setting the result to zero, we find that the effective field must satisfy

$$B_{eff} = B + Jq \tanh \beta B_{eff}, \quad (8.23)$$

From here we could insert this result into the expression for \mathcal{F} above and simplify to get our mean field free energy $\mathcal{F}_{meanfield}$. We could also cleverly exploit the definition of the

trial Hamiltonian's magnetization per spin, $m_0 = \tanh \beta B_{eff}$, to each side of the above self-consistency condition:

$$B_{eff} = B + Jq \tanh \beta B_{eff} \quad (8.24)$$

$$\Rightarrow B + Jqm_0 = B + Jq \tanh \beta B_{eff} \quad (8.25)$$

$$\Rightarrow m_0 = \tanh (\beta B + \beta Jqm_0), \quad (8.26)$$

exactly the mean field equation of state we derived by quite different means back in Sec. 8.1. These variational approaches elegantly show how to map systems you care about to the best possible non-interacting system. Very satisfying.

8.4 Problems

8.4.1 Computers, partition functions, and thermodynamics

We're about to start working on the statistical mechanics of *interacting* particles, which are much harder to analytically work with than their non-interacting counterparts. This is one of the areas of computational physics where computer simulations have long been a crucial tool.

In this problem we'll be using an online computer simulation of the 2D Ising model to get a flavor for this second mode of working with numerics. The Ising model that we'll look at here consists of many spins, $\sigma_i = \pm 1$, arranged on a square lattice. The Hamiltonian is

$$\mathcal{H} = -J \sum_{\langle ij \rangle} \sigma_i \sigma_j - H \sum_i \sigma_i,$$

where " $\sum_{\langle ij \rangle}$ " indicates a sum over the four nearest neighbors on the lattice of each spin, J is a coupling constant, and H is an external field. Thermodynamically, H is a generalized force which is conjugate to the magnetization, $M = \sum_i \sigma_i$. As is typical, we'll work in units where $J = 1$ and $k_B = 1$ (i.e., we're measuring both energies and temperatures in units of the coupling constant). As you know, the partition function is

$$Z = \sum_{\mu} e^{-\beta \mathcal{H}(\mu)}.$$

Since we are fixing the spins to their lattice sites, each classical microstate is simply a specification of every spin...thus, there are 2^N microstates that in principle we need to sum over. Perhaps you begin to see the issue: how can we possibly make analytic progress in evaluating this sum of Boltzmann factors over an exponentially large number of terms¹²⁰?

So, instead of solving this analytically, we'll explore tools used to *statistically sample* the configurations which contribute the most to the partition function. The methods used will not be our focus (if this were a computational physics class they would be), but if you want to know more I'm happy to chat! For this whole exercise you'll be using the javascript simulation that I've shown a few times in class: <https://www.dmsussman.org/resources/isingmodelsimulator/>

A [Computational] First, just spend a few minutes playing around with the simulation. Next, as a first exploration of the simulation, give me an estimate the *critical temperature at zero field*, which is to say, the largest temperature at $H = 0$ for which there is still a net magnetization. This should be straightforward (especially since the correct number is, in fact, listed on the page!). Also go ahead and tell me your estimate for the critical temperature at $H = 0.1$.

B [Pencil and paper] From the definition of the partition function, show that the average magnetization is given by

$$M = - \left. \frac{\partial F}{\partial H} \right|_T,$$

¹²⁰Actually, in 1 and 2 dimensions there are clever tricks you can play to *analytically* solve this model! Onsager's 2D square lattice Ising model solution is the calculation I have most frequently heard referred to as a "mathematical tour de force." But in general... It should feel daunting, and it is.

where $F = -kT \log Z$. Derive a formula for the susceptibility,

$$\chi = \left. \frac{\partial M}{\partial H} \right|_T,$$

that can be expressed in terms of the moments and/or cumulants of M . Do a similar calculation connecting the specific heat and moments and/or cumulants of the energy.

C [Computational] Test your formulas for $H = 0$ and $T = 4$ by first computing the relevant moments and/or cumulants at those values of field and temperature. Next, compare your answer to just directly taking a numerical derivative (i.e., vary the field by some small ΔH or the temperature by some small ΔT). In this problem, you should definitely be making use of the fact that there is a “Download Data” button; this will download a txt file where each row is a time point, an energy per spin, and a magnetization per spin corresponding to the current state of the simulation.

Some notes and comments. First, when when analyzing data, make sure you let the system equilibrate first. Second, notice that the website is outputting *intensive* versions of the energy and magnetization ($e = E/N$, $m = M/N$) – adjust your formulas / expectations accordingly! Finally, notice that the web simulator has a drop-down menu controlling the system size. Do your answers depend sensitively on what you choose here?

Full credit on this problem will involve turning in both any hand-written/typed work for the “pencil and paper” question, a summary of the results for the computational questions, and a print-out of all code that you used. You do not need to (in fact, please do not) turn in any of the raw data you downloaded and processed.

8.4.2 Surfactants and surface tension in the mean field approximation

Suppose we have a fluid interface of area A with N surfactant molecules – how does that affect the surface tension of the fluid? Real surfactants are complicated molecules with complicated interactions, but the essential mechanism of reducing surface tension is straightforward. This question will lead you through a path very similar to the mean-field version of the van der Waals equation, and remind you of the basic workings of conjugate thermodynamic coordinates.

If we ignore the coupling of the surfactants with the fluid and assume the interface is flat, the surfactants have a Hamiltonian

$$\mathcal{H} = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m} + \sum_{i<j} \phi(\mathbf{r}_{ij}),$$

where the position and momentum vectors are in 2D.

(A) Write down the canonical partition function $Z(T, A, N)$ in terms of integrals over the \mathbf{r} and \mathbf{p} (with the corrected phase space measure). The momentum integrals should be pretty standard by now, so carry them out.

(B) For the rest of the problem, let's assume that the pair potential is infinite for $\mathbf{r}_{ij} < a$ and that $\int_a^\infty dr 2\pi r \phi(r) = -\varepsilon_0$. Separately estimate the total amount of non-excluded area at the interface and the total potential energy of the system. For the total potential energy, make use of the uniform density approximation, $n(\mathbf{r}) \approx n = N/A$.

(C) From the above, write down the mean field approximation for the canonical partition function, and from that the free energy.

(D) Holding other thermodynamic coordinates fixed, the work done in changing the surface area of the system is $dW = \sigma dA$. Assume that the total surface tension of the fluid is a sum of σ_0 (the surface tension of the fluid without surfactants, which we will approximate as temperature-independent) and σ_s (the contribution from the surfactant molecules). What is the total n - and T -dependent surface tension, σ ?

8.4.3 Problem 2

8.4.4 Problem 3

8.4.5 Problem 4

Chapter 9

Critical behavior: the renormalization group approach

All the way back in Chapter 3 we saw –although did not fully understand why – that near a phase transition something remarkable happens. When considering, say, a transition between the liquid and gas state we saw that totally different substances, made of completely different microscopic constituents, exhibited the same *behavior*. For instance, all gases have an isothermal susceptibility that diverges as they approach the temperature along the critical isochore – $\kappa_T \sim |T - T_c|^\gamma$ – with the same value $\gamma = 1.237075\dots$ characterizing *all* gases. Even more remarkably, the same set of critical exponents describe the behavior of not only all gases close to their critical point, but also ferromagnetic materials close to their Curie point. Somehow these critical exponents define *universality classes*, allowing us to ignore whole swaths of microscopic detail to collapse very different physical systems into the same mathematical description.

RG etymology... clever people say “semi-group”, but it’s actually a commutative monoid. I guess “renormalization commutative monoid” doesn’t quite roll of the tongue in the same way.

matrix M as Jacobian

Two major (and perhaps one “minor”) questions seemed especially pressing: Why do we see this universality across different physical systems? How can we compute the critical exponents defining a universality class? And relatedly, are the critical exponents all independent from each other (so that we have to learn to calculate all of them), or are there relationships between them that have to be satisfied (and if so, why)?

We got a hint from Landau that the fundamental reason for this universality is related to different systems sharing the same fundamental symmetries. But while the Landau approach tentatively¹²¹ helped us make progress on that front, it predicted the mean field values of the critical exponents and – even worse – with *no* way to systematically improve upon the calculation to make more accurate predictions.

We also got a hint from Widom’s scaling hypothesis that the critical exponents should *not* be thought of as independent... but why is this hypothesis valid? What new technique

¹²¹“Tentatively,” because it still gave us quantitatively the wrong answers, so how confident in it should we really be?

/ idea / approach do we need in order to not only *derive* the scaling hypothesis, but also compute the forms of the scaling functions and values gap exponents? We'll begin answering this question by discussing Kadanoff's "block-spin" argument – a deep insight that teaches us that in the face of a diverging correlation length we should shift our perspective substantially: away from trying to understand the behavior of a specific Hamiltonian and instead thinking about the relationship between the *length scale* over which we define our order parameter and the *coupling constants* defining the Hamiltonian of the systems.

9.1 Kadanoff's block spin argument

Up until now, we've been moving the goal posts back on what we need to understand to really tackle critical phenomena. Widom taught us that we should figure out how to derive a kind of static scaling hypothesis, and now Kadanoff¹²² is going to swoop in with a bold change of perspective: rather than focus on choosing a particular level of description for a problem¹²³, we'll think about how physics changes as we *vary* the level of description we choose for our problem.

Here's the basic idea. Suppose we start with a nearest-neighbor Ising model characterized by N spins sitting on a lattice with lattice spacing λ , with Hamiltonian

$$\beta\mathcal{H}_1 = -K_1 \sum_{\langle ij \rangle} \sigma_i \sigma_j - h_1 \sum_i \sigma_i. \quad (9.1)$$

Here we define $K_1 = \beta J$, $h_1 = \beta B$, and as usual $\sigma_i = \pm 1$ (and we'll see why we wrote " \mathcal{H}_1 , K_1 , and h_1 " in just a moment). Now, if we take this model close to the critical point we know that the spins will be correlated over some range given by the correlation length, ξ . In some sense, if we pick a length scale which is some multiple of the lattice spacing, $l\lambda$, then we can kind of think of spins on that larger length scale as "acting together," as long as $\lambda \ll l\lambda \ll \xi$.

Since we aren't doing the full version, should probably replace the block spin definition with the more intuitive one. This motivates what we will call a *block spin transformation*, schematically illustrated in Fig. 9.1. We imagine averaging together l^d spins at a time into new "block spin" variables, which we will denote by using capital subscripts:

$$\sigma_I = \frac{1}{\langle \sigma_i \rangle_I l^d} \sum_{i \in I} \sigma_i, \quad (9.2)$$

where

$$\langle \sigma_i \rangle_I = \frac{1}{l^d} \left(\sum_{i \in I} \sum_{j \in I} \langle \sigma_i \sigma_j \rangle \right)^{1/2}.$$

¹²²From a time when people would write things like, "On the other hand, \tilde{K} is perhaps a somewhat more subtle beast" [66]

¹²³e.g.: we decide on the level of description we want to apply to a problem – say, we'll have a fluid made up of classical point particles, or a gas made up of molecular dipoles, or we'll think of the hard quantum chemistry problems involved in magnetism and decide to treat it as discrete spins on a lattice – and we'll take that to describe our microscopic degrees of freedom, and (perhaps) not think too hard about how we got to those degrees of freedom or whether we should have coarse-grained further

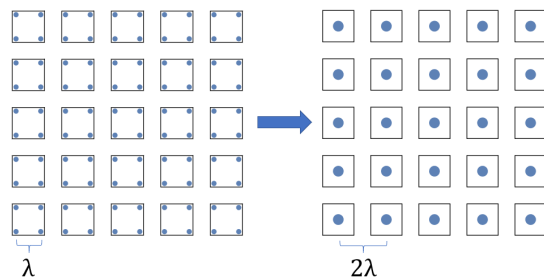


Figure 9.1: **A block spin transformation** Here, 2^d spins on the original lattice (left) are grouped together into block spins, leading to $N/2^d$ new degrees of freedom on a lattice whose lattice spacing is twice as big as the original lattice's (right).

From here, we will progress by making some courageous assumptions that are definitely incorrect, but which are incorrect in a helpful way. In the words of Wilson, “In short the Kadanoff block picture, although absurd, will be the basis for generalizations which are not absurd¹²⁴.”.

Courageous assumption 1

First, we boldly assume that we can write down a Hamiltonian for the block spin system which looks the same as the original Hamiltonian:

$$\beta\mathcal{H}_l = -K_l \sum_{\langle IJ \rangle} \sigma_I \sigma_J - h_l \sum_I \sigma_I, \quad (9.3)$$

where $\sigma_I = \pm 1$. This is clearly wrong, in several different ways. For instance, we have assumed that the spins within the block are *all* aligned either up or down, so that our block spins are two-valued in the same way that the original spins were. Another issue: block spins that are *diagonally adjacent* with each other have spins inside them that interact with a common spin – this suggests that after block spinning we should also have some *next* nearest neighbor interactions in addition to the nearest-neighbor ones in Eq. 9.3. That is, it sure seems like the form of the Hamiltonian should change after the block spin transformation.

Well, we're not calling it a courageous assumption for nothing. We next notice that the new lattice spacing for our block spins is $l\lambda$, but the underlying physical system is the same. This implies that ξ_l , measured in units of its own lattice spacing, is smaller than ξ_1 :

$$\xi_l = \frac{\xi}{l}. \quad (9.4)$$

What does this mean? If the original system is close to the critical point and has, consequently, some very large correlation length, then the block spin system will look like it is farther from its critical point! That is, the block spin system will behave as though it is at some different (larger) value of the reduced temperature t_l and some new value of the effective field strength h_l (the latter because of the way the field interacts with the combination of spins forming

¹²⁴Ref. [67]

the block, the former because of the way the block spins couple together with a different K_l , which itself sets T_c).

Finally, we note that if \mathcal{H}_1 and \mathcal{H}_l have the same functional form, then the functional form of the free energy density¹²⁵ should be the same, too! But, again, the underlying physical system hasn't changed – just how hard we are squinting our eyes when we look at it – so we can go further and say

$$N/l^d f_s(t_l, h_l) = N f_s(t_1, h_1) \quad \Rightarrow \quad f(t_1, h_1) = l^{-d} f(t_l, h_l). \quad (9.5)$$

Courageous assumption 2

Simplifying the actual argument in Kadanoff's paper, we'll frame the next courageous assumption as: "We want to understand power-law behavior near the critical point, so... maybe near the critical point we have

$$t_l = t l^{y_t} \quad (9.6)$$

$$h_l = h l^{y_h}, \quad (9.7)$$

where both exponents y_t and y_h are positive?" Substituting the assumption into Eq. 9.5 we get

$$f_s(t, h) = l^{-d} f_s(t l^{y_t}, h l^{y_h}). \quad (9.8)$$

Up to now, we haven't actually specified the scale of the transformation. Let's make one more assumption – that we have complete freedom in choosing this lengthscale l , and select something which simplifies our expressions. We make the specific choice

$$l = |t|^{-1/y_t},$$

so that

$$f_s(t, h) = |t|^{d/y_t} f_s(1, h |t|^{-y_h/y_t}). \quad (9.9)$$

From here, we define $\Delta = y_h/y_t$, $\alpha = 2 - d/y_t$, and $F_s(x) = f_s(1, x)$; doing so we see that the above equation is actually

$$f_s(t, h) = |t|^{2-\alpha} F_s\left(\frac{h}{|t|^\Delta}\right), \quad (9.10)$$

i.e., the starting point of the static scaling hypothesis!

Comments

Kadanoff's block spin argument motivates the form of the scaling hypothesis, but we still don't know (a) just how bold our assumptions were, (b) what the scaling functions actually are, or (b) how to actually calculate the exponents y_t and y_h that we just encountered. We also haven't really done anything to explain universality: in fact, the degree of validity of some of those courageous assumptions above might actually depend on the microscopic details; this feels like a step backwards in some ways.

¹²⁵Or, at least, the part of the free energy density which is singular at the critical point

The biggest assumption we made – that the transformed system’s Hamiltonian, \mathcal{H}_l , looked like it had the same form as the original one, \mathcal{H}_1 – both feels the most uncertain *and* contains the key insight of the renormalization group. Namely: we should shift our focus from trying to solve the entire problem all at once to, instead, trying to understand what happens as we progressively remove more and more degrees of freedom from the problem and watch how the Hamiltonian of the system evolves.

9.2 General formulation of the renormalization group approach

We will now turn to Wilson’s renormalization group approach, generalizing and reformulating Kadanoff’s picture; we will do some explicit calculations in the context of the 1-dimensional Ising model and, following Niemeyer and van Leeuwen, on a specific 2-dimensional Ising model.

Goal is not really to learn RG – we want to understand why it is an approach that leads to scaling, can give non-trivial critical exponents, and why it could lead to singular behavior, etc. We’ll have two simple real-space calculations, and gaussian momentum shell formulation. Better to do this and then the 1D Ising model calculation, or the other way around? For now I like this order, but in any event: clean this up, and remind the reader what we did back when we talked about Kadanoff block spins. In the next section we’ll do an explicit RG transformation for the nearest-neighbor Ising model in 1D. Before that, we pause to paint the picture of the general formulation of the approach.

Our goal is to find a renormalization group transformation, \mathcal{R}_l , which coarse-grains over some length scale l and in so doing reduces the number of degrees of freedom in the problem: if there were originally N degrees of freedom, now there are only

$$N' = \frac{N}{l^d}.$$

We, furthermore, want to do this in such a way that the actual free energy is preserved. We might accomplish this trick by some sort of “block spin” procedure suggested by Kadanoff, or by moving to a Fourier space description of a model and integrating out any components with large momenta (doing something like $\int_{k/l}^k dq q^{d-1}$, where k represents some upper momentum cutoff), or by some other method. This typically reduces the density of the degrees of freedom, so we additionally rescale all lengths. For the simple lattice spin systems we are about to study, this can be thought of as agreeing to measure lengths in units of the lattice spacing, regardless of whether we have the original lattice or the “thinned out” lattice. In any case, measuring distances as $x' = x/l$ means that the new correlation length will be trivially related to the correlation length before the transformation: $\xi' = \xi/l$.

Schematically, after doing this procedure we have arrived at something like

$$Z = \sum_{\{\sigma\}} e^{\beta\mathcal{H}} = \sum_{\{\sigma'\}} e^{\beta\mathcal{H}'}. \quad (9.11)$$

If we suppose the Hamiltonian for our system depends on a collection of coupling constants \vec{k} , then the new effective Hamiltonian depends on the set $\vec{k}' = \mathcal{R}_l(\vec{k})$. In general, an RG transformation corresponds to some unpleasantly nonlinear function of these coupling constants, which might not generically preserve the form of the Hamiltonian. That is: a typical case might be that coupling constants which were zero in the original \mathcal{H} flow to nonzero values for \mathcal{H}' . However, since the partition function Z is unchanged, we can still say that the free energy is not changed by the transformation, and hence from our rescaling of lengths the free energy density transforms simply as

$$f(\vec{k}) = l^{-d} f(\vec{k}') \quad (9.12)$$

Suppose one is able to find a *fixed point* of R_l , i.e., a set of \vec{k}^* such that $\vec{k}^* = R_l(\vec{k}^*)$. As we have indicated before, R_l has scaled lengths by a factor of l , so the correlation length behaves as

$$\xi(\vec{k}') = \frac{\xi(\vec{k})}{l} \Rightarrow \xi(\vec{k}^*) = \frac{\xi(\vec{k}^*)}{l} \Rightarrow \xi(\vec{k}^*) = 0 \text{ or } \infty. \quad (9.13)$$

Fixed points with a vanishing correlation are called *trivial* fixed points, whereas fixed points with a diverging correlation length are *critical* fixed points.

9.2.1 Local flows near a fixed point

Suppose, having found a fixed point, we investigate how coupling constants changes if you start not *at* that fixed point but rather somewhere nearby, at $\vec{k} = \vec{k}^* + \delta\vec{k}$. Under the action of an RG transformation, we have $\vec{k}' = R_l(\vec{k}) = \vec{k}^* + \delta\vec{k}'$, where we can Taylor expand to find

$$\delta k'_n = \sum_m \left. \frac{\partial k'_n}{\partial k_m} \right|_{\vec{k}=\vec{k}^*} \delta k_m. \quad (9.14)$$

Change notation so that the matrix takes l as an argument Call \mathbf{M}_l the matrix whose (n, m) th element is this $\left. \frac{\partial k'_n}{\partial k_m} \right|_{\vec{k}=\vec{k}^*}$, and denote its eigenvalues and eigenvectors as $\lambda_i(l)$ and $\vec{e}_i(l)$, respectively (here we are emphasizing that the eigenvalues and eigenvectors depend on the scale of the RG transformation; we will suppress this dependence much of the time).

One of the things we require of the transformation R_l is that it is *monoid*¹²⁶, by which we mean that successive transformations with $l = l_1$ and $l = l_2$ are equivalent to a single transformation of scale $l_1 l_2$. Thus, if $\vec{k}' = R_{l_1}(\vec{k})$ and $\vec{k}'' = R_{l_2}(\vec{k}')$, then we also have $\vec{k}'' = R_{l_1 l_2}(\vec{k})$. The RG transformation being a monoid means that the matrix \mathbf{M}_l has the same property, and so do its eigenvalues (commutative – simultaneous diagonalizability):

$$\mathbf{M}_{l_1} \mathbf{M}_{l_2} = \mathbf{M}_{l_1 l_2} \quad \text{and} \quad \lambda_i(l_1) \lambda_i(l_2) = \lambda_i(l_1 l_2). \quad (9.15)$$

Thinking of the eigenvalues as functions of the lengthscale over which the transformation is defined, what kinds of functions have this property? We can answer this by taking the partial

¹²⁶Almost a group, but more than a semi-group. We require $l \geq 1$, so there is an identity transformation ($l = 1$), but there are no inverse transformations that let you recover information at a finer scale. I guess “renormalization monoid” is not as catchy a name.

derivative with respect to l_2 on both sides of the equation above, giving us

$$\lambda_i(l) \frac{\partial \lambda_i(l_2)}{\partial l_2} = l_1 \frac{\partial \lambda_i(l_1 l_2)}{\partial l_1 l_2}. \quad (9.16)$$

If we now set $l_2 = 1$ – for which we know that $\lambda_i(1) = 1$, since a transformation with $l = 1$ corresponds to keeping all of the original degrees of freedom and hence not changing anything about the Hamiltonian – we have a differential equation that reads (letting $l_1 = l$ now, for notational convenience)

$$y_i \lambda_i(l) = l \frac{\partial \lambda_i}{\partial l}, \quad \text{where} \quad y_i = \left. \frac{\partial \lambda_i(l)}{\partial l} \right|_{l=1}, \quad (9.17)$$

which (given the boundary condition $\lambda_i(1) = 1$) has the solution

$$\lambda_i(l) = l^{y_i}. \quad (9.18)$$

We will make use of this functional form shortly. First, though, let's think about how coupling constants evolve in terms of the possible values that these eigenvalues can take. To do so, we'll write $\delta \vec{k}$ in the basis of \mathbf{M}_l 's eigenvectors: $\delta \vec{k} = \sum_i a_i \vec{e}_i$, where a_i is just the projection of $\delta \vec{k}$ onto \vec{e}_i . Well, we can now compactly write

$$\delta \vec{k}' = \mathbf{M}_l \delta \vec{k} = \mathbf{M}_l \sum_i a_i \vec{e}_i = \sum_i (a_i \lambda_i) \vec{e}_i, \quad (9.19)$$

where we see now that $(a_i \lambda_i)$ represents how the projection of $\delta \vec{k}$ onto each eigendirection *changes* after each RG step. As a bit of nomenclature¹²⁷, there are three important cases to consider:

- (a) $|\lambda_i| > 1$, which means the projection *increases* (i.e., in this direction the coupling constants flow *away* from the fixed point). These are said to be “relevant” directions.
- (b) $|\lambda_i| < 1$, which means the projection *decreases* (i.e., in this direction the coupling constants flow *towards* from the fixed point). These are said to be “irrelevant” directions.
- (c) $|\lambda_i| = 1$, which means the projection *stays the same*. These are said to be “marginal” directions (and are often associated with log corrections to scaling, and are often important if d is either the upper or lower critical dimension).

9.2.2 Origin of scaling

To explicitly see how all of this machinery lets us understand critical behavior, consider the simple case of there being a single relevant variable near a fixed point. Let's suppose the

¹²⁷Importantly, these terms refer to the behavior of the different directions *at a particular fixed point*. The same flow equations can, of course, have more than one fixed point, and there is nothing that says (e.g.) that a relevant direction at one fixed point must also be a relevant direction at a different fixed point.

relevant variable is just the temperature of the system, T , and that we start at $T = T^* + \delta T$. We know that after our RG transformation this will flow to $T' = R_l(T) = T^* + \delta T'$, where

$$\delta T' = (T' - T^*) \approx \lambda(l)(T - T^*), \quad \text{for} \quad \lambda(l) = \left. \frac{\partial R_l}{\partial T} \right|_{T=T^*}. \quad (9.20)$$

We already know what $\lambda(l)$ has to look like; it is $\lambda(l) = l^{y_t}$. Let's divide both sides of Eq. 9.20 by T^* and reintroduce our friend $t = (T - T^*)/T^*$, the reduced temperature. We have that $t' = tl^{y_t}$ after a single RG transformation, and after n iterations we have

$$t^{(n)} = tl^{ny_t}, \quad \text{and} \quad \xi(t) = l^n \xi t^n = l^n \xi(tl^{ny_t}). \quad (9.21)$$

Let's choose a particular scale for our transformation, selecting

$$l^n = \left(\frac{b}{t} \right)^{1/y_t},$$

where b is just some arbitrary number of order one. We finally see that the correlation length close to the critical point behaves like

$$\xi(t) = \left(\frac{t}{b} \right)^{-1/y_t} \xi(b). \quad (9.22)$$

That is, we have just computed the critical exponent associated with the divergence of the correlation length:

$$\nu = \frac{1}{y_t}, \quad \text{where} \quad y_t = \frac{\log(\lambda(l))}{\log(l)}. \quad (9.23)$$

That is: the critical exponent is associated with computing the eigenvalues of the linearized transformation that describes how relevant coupling constants flow near a fixed point of the transformation. These can take quite general values, and hence are not at all restricted to coincide with the mean field values we repeatedly encountered in the last chapter. Even more importantly, the critical exponents calculated according to this prescription *coincide with those of actual physical systems!*

9.2.3 Origin of singular behavior

[ball rolling on a hill picture of wilson → discontinuous jump in position after infinite amount of time](#) ¹²⁸. A basic insight of RG: rather than tackle the whole problem all at once, it might be easier to see how much the problem changes if you adjust the resolution you choose to study the problem at. Looking at the whole bulk material / taking the thermodynamic limit then corresponds to performing an RG transformation \mathcal{R}_l an infinite number of times. This gives a more pleasing intuition for why the thermodynamic limit permits singularities (compared, that is, to the earlier, “well, the infinite sum of analytic functions need not necessarily be analytic” argument we had to resort to earlier).

it might be easier to

¹²⁸Ref. [67]

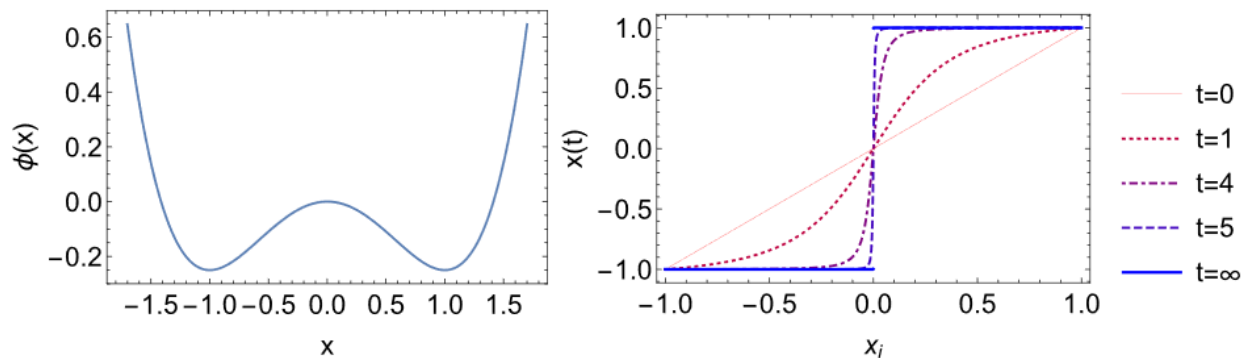


Figure 9.2: **Singular behavior in a classical system** Suppose dynamics are governed by the equation $dx/dt = -d\phi/dx$, with our favorite double well potential (left). At any finite time, the final position of a particle is a continuous function of its initial position, but if we take the $t \rightarrow \infty$ limit we get a discontinuity.

Explicit example: $\phi(x) = -x^2/2 + x^4/4$, then solution to differential equation is

$$x(t) = \text{Sign}(x_i) \frac{e^t}{\sqrt{e^{2t} + x_i^{-2} - 1}}$$

so infinite time limit is just which side of the saddle point you started on.

9.3 The 1-dimensional Ising chain

In the rest of this chapter, we will see explicit examples of this renormalization group approach put into practice¹²⁹. We begin by writing down the canonical partition function associated with a 1D Ising model with N sites and only nearest neighbor interactions:

$$Z = \sum_{\{\sigma_i\}} e^{-\beta\mathcal{H}}, \quad \text{where } -\beta\mathcal{H} = Nk_0 + k_1 \sum_i \sigma_i + k_2 \sum_i \sigma_i \sigma_{i+1}. \quad (9.24)$$

We have once again changed notation to emphasize that we are writing some general Hamiltonian that might have terms containing different numbers of coupled spins; relative to what we've written in the last chapter we have $k_2 = \beta J$, $k_1 = \beta B$, and $k_0 = 0$ (we will see shortly why we needed to pay attention to the k_0 term!).

Recalling the spirit of Kadanoff's block spin argument, our goal is to find a transformation which (a) reduces the number of spins we are considering but (b) describes the same physical system and, hence, leaves the partition function unchanged while (c) not changing the form of the effective Hamiltonian describing the new degrees of freedom. That is, we want

$$Z = \sum_{\{\sigma'_i\}} e^{-\beta\mathcal{H}'}, \quad \text{where } -\beta\mathcal{H}' = N'k'_0 + k'_1 \sum_i \sigma'_i + k'_2 \sum_i \sigma'_i \sigma'_{i+1} \text{ and } N' < N. \quad (9.25)$$

¹²⁹Recalling, M. E. Fisher's remark that "The actual process of explicitly constructing a useful renormalization group is not trivial" [68], in this section we'll focus on a "trivial" example which, alas, is not as useful as we might like.

One idea might be to average neighboring pairs of spins together into a new effective spin, but this has the issue that the average of two spins is a *three*-valued quantity, whereas the original spins were two-valued. We can avoid this by picking one of the two spins in each pair to be the “tie-breaker,” say, the spin with i an odd number; this procedure is equivalent to saying the value of the new block spin is in fact whatever the spin of the tie-breaker is.

9.3.1 Decimation of the 1D Ising Chain

That is: we are going to separate our original set of spins σ_i into the set where i is odd vs even. Let $\sigma'_i = \sigma_{2i-1}$, the i th odd spin of the original chain, and let $s_i = \sigma_{2i}$, the i th even spin of the original chain. We will now manipulate the original partition function by summing over (“integrating out”) the even spins. First, let’s write the original partition function in a more convenient form:

$$Z = \sum_{\{\sigma_i\}} \exp \left(\sum_{i=1}^N \left[k_0 + \frac{k_1}{2} (\sigma_i + \sigma_{i+1}) + k_2 \sigma_i \sigma_{i+1} \right] \right) \quad (9.26)$$

$$\begin{aligned} &= \sum_{\{\sigma'_i\}} \sum_{\{s_i\}} \exp \left(\sum_{i=1}^{N/2} \left[k_0 + \frac{k_1}{2} (\sigma'_i + s_i) + k_2 \sigma'_i s_i + k_0 + \frac{k_1}{2} (s_i + \sigma'_{i+1}) + k_2 s_i \sigma'_{i+1} \right] \right) \\ &= \sum_{\{\sigma'_i\}} \sum_{\{s_i\}} \prod_{i=1}^{N/2} \exp \left(2k_0 + \frac{k_1}{2} [\sigma'_i + 2s_i + \sigma'_{i+1}] + k_2 [\sigma'_i s_i + s_i \sigma'_{i+1}] \right). \end{aligned} \quad (9.27)$$

Now that we have Z in a form where all of the contributions of each s_i are accounted for, we explicitly carry out the sum over $s_i = -1, +1$:

$$\begin{aligned} Z &= \sum_{\{\sigma'_i\}} \prod_{i=1}^{N/2} \left[e^{2k_0 + \frac{k_1}{2} (\sigma'_i + \sigma'_{i+1} + 2) + k_2 (\sigma'_i + \sigma'_{i+1})} + e^{2k_0 + \frac{k_1}{2} (\sigma'_i + \sigma'_{i+1} - 2) + k_2 (-\sigma'_i - \sigma'_{i+1})} \right] \\ &\quad \sum_{\{\sigma'_i\}} \prod_{i=1}^{N/2} \left[\exp(2k_0) \exp \left[\frac{k_1}{2} (\sigma'_i + \sigma'_{i+1}) \right] \cdot 2 \cosh [k_1 + k_2 (\sigma'_i + \sigma'_{i+1})] \right]. \end{aligned} \quad (9.28)$$

We compare this with Eq. 9.25, and see that to meet our criteria of having a Hamiltonian in the same form as the original one, we need to satisfy

$$\exp \left[k'_0 + \frac{k'_1}{2} (\sigma'_i + \sigma'_{i+1}) + k'_2 \sigma'_i \sigma'_{i+1} \right] = \exp(2k_0) \exp \left[\frac{k_1}{2} (\sigma'_i + \sigma'_{i+1}) \right] (2 \cosh [k_1 + k_2 (\sigma'_i + \sigma'_{i+1})]). \quad (9.29)$$

The two sides look quite different, but remember: we only need them to be equal for the values that the degrees of freedom can actually take! That is, we need the equation to hold for the allowed values of $\sigma_i = \pm 1$ and $\sigma_{i+1} = \pm 1$.

So, let’s introduce the notation

$$x = e^{k_0}, \quad x' = e^{k'_0}, \quad y = e^{k_1}, \quad y' = e^{k'_1}, \quad z = e^{k_2}, \quad z' = e^{k'_2} \quad (9.30)$$

and enumerate the four relevant cases and corresponding statement of Eq. 9.29 to hold:

$$\begin{array}{cc} \sigma'_i & \sigma'_{i+1} & \text{equation} \\ + & + & x'y'z' = x^2y(yz^2 + y^{-1}z^{-2}) \end{array} \quad (9.31)$$

$$+ \quad - \quad x'(z')^{-1} = x^2(y + y^{-1}) \quad (9.32)$$

$$- \quad + \quad x'(z')^{-1} = x^2(y + y^{-1}) \quad (9.33)$$

$$- \quad - \quad x'(y')^{-1}z' = x^2y^{-1}(y^{-1}z^2 + yz^{-2}) \quad (9.34)$$

Note that the middle two equations are identical, leaving us with three equations and three unknowns, with the solution¹³⁰

$$\begin{aligned} (x')^4 &= x^8(yz^2 + y^{-1}z^{-2})(y^{-1}z^2 + yz^{-2})(y + y^{-1})^2, \\ (y')^2 &= y^2 \frac{(yz^2 + y^{-1}z^{-2})}{(y^{-1}z^2 + yz^{-2})}, \\ (z')^4 &= \frac{(yz^2 + y^{-1}z^{-2})(y^{-1}z^2 + yz^{-2})}{(y + y^{-1})^2}, \end{aligned} \quad (9.35)$$

or, in terms of the coupling constants themselves,

$$\begin{aligned} k'_0 &= 2k_0 + \log 2 + \frac{1}{4} \log [\cosh^2(k_1) \cosh(2k_2 + k_1) \cosh(2k_2 - k_1)], \\ k'_1 &= k_1 + \frac{1}{2} \log [\cosh(2k_2 + k_1) \operatorname{sech}(2k_2 - k_1)], \\ k'_2 &= \frac{1}{4} \log [\operatorname{sech}^2(k_1) \cosh(2k_2 + k_1) \cosh(2k_2 - k_1)]. \end{aligned} \quad (9.36)$$

9.3.2 Flow of the coupling constants

Equations 9.36 dictate how the coupling constants of the effective Hamiltonian evolve for this particular renormalization group transformation (which blocks together pairs of spins at a time). Upon *repeated* application of this transformation, do these equations have a fixed point? First, we note that the k'_0 terms reflects the part of the free energy which does not couple to our degrees of freedom – near a phase transition we expect this to vary smoothly rather than contribute to the singular part of the free energy. We also see more directly that the value of k_0 does not enter in the expressions for how k_1 and k_2 evolve. As such, even though it was needed to have a Hamiltonian whose form was unchanged after our renormalization group transformation, we will henceforth neglect it.

Case 1: No external field First, we notice that if there is no external field, so that $k_1 = 0$, the flow equations tell us that the new effective Hamiltonian *also* has no field: $k'_1 = 0$. Our equation for the nearest neighbor spin-spin coupling constant becomes

$$k'_2 = \frac{1}{2} \log \cosh(2k_2). \quad (9.37)$$

¹³⁰The x' equation comes from multiplying the four cases together, the y' equation comes from taking the ratio of $(++)/(- -)$, and the z' equation comes from taking the ratio of $(++)(- -)/(+-)(-+)$.

Recalling that $k_2 = \beta J$, we see that as $T \rightarrow \infty$, $k_2 \rightarrow 0$ and this equation becomes $k'_2 = k_2^2$. That is, we see that if k_2 is a small number, k'_2 is an even smaller number. Thus, the $k_2 = 0$ fixed point is a stable one. It represents the high-temperature, disordered phase of the Ising spin chain, and has a vanishing correlation length.

At the other extreme, $T \rightarrow 0$, $k_2 \rightarrow \infty$ and this equation becomes $k'_2 = \frac{1}{2} \log(\exp(2k_2)/2) \approx k_2 - \log(2)/2$. That is, if k_2 is a very large number, then k'_2 is a very slightly smaller number. Thus, the $k_2 = 0$ fixed point is an *unstable* one. It represents the zero-temperature ordered state, but because it is unstable we see that if the system is at any finite temperature the system will slowly flow away from that fixed point. There are no other fixed points for this equation, and so on long enough length scales the 1D Ising model in the absence of a field always looks disordered.

Case 2: Finite external field When $k_1 \neq 0$, one can calculate a line of stable fixed points along $k_2 = 0$ and any value of k_1 . Here, for any value of the external field, the 1D Ising models will (on long enough length scales) look like a collection of independent spins in an external field.

ABOVE: FINISH CALCULATIONS TO SHOW, e.g., how the flow near unstable FP relates back to the scaling of the correlation length.

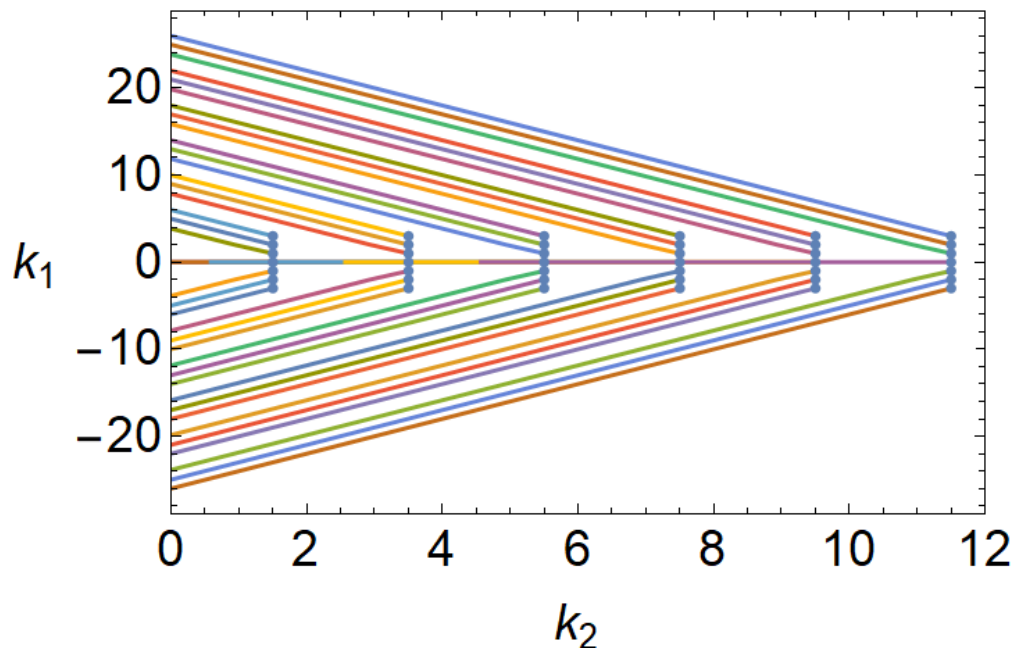


Figure 9.3: **Evolution of coupling constants of the 1D Ising model under decimation** Solutions to Eq. 9.36, where the starting point of each flow curve is indicated by a dot. Flows that start at $k_1 = 0$ stay at $k_1 = 0$, whereas nonzero values of the external field flow to fixed points with larger effective values of k_1 and with no spin-spin coupling.

9.3.3 Critical surfaces, flow, etc

[69], also Plischke & Bergen 7.2

9.4 Approximate real space calculation: The 2-dimensional Ising model on a triangular lattice

“In 1D we could exactly solve for critical exponents, but there was no finite temperature critical point; what happens in 2D?” Niemeijer and van Leeuwen, basically¹³¹. Need to write the connective tissue.

below, move the “spins s conditioned on sigmas” from the hamiltonians in the exponentials to the sums themselves. clean up the notation

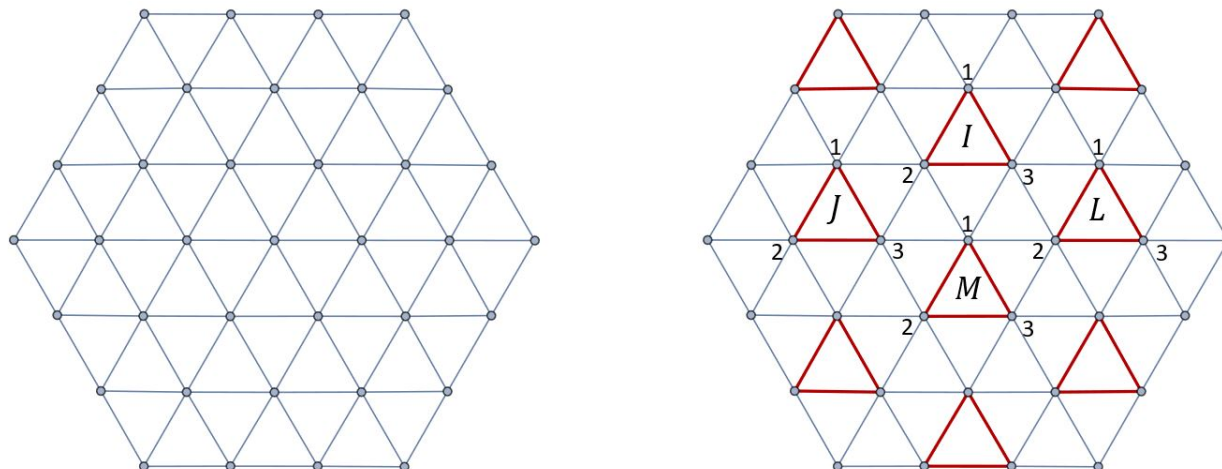


Figure 9.4: **Block spins for the Niemeijer-van Leeuwen cumulant expansion** (Left) A small patch of a triangular lattice. (Right) Grouping spins into a new triangular lattice of block spins, each of which is formed from three microscopic spins.

We’re now on a triangular lattice (Fig. 9.4), and for the purpose of keeping our calculations as simple as possible, let’s consider the Ising model Hamiltonian with only nearest-neighbor interactions (no k_0 , no external field),

$$-\beta\mathcal{H} = k \sum_{\langle ij \rangle} \sigma_i \sigma_j,$$

where $\langle ij \rangle$ is our standard notation for “spins i and j which are nearest neighbors.” You may wonder why we are not including the constant (k_0) term; as before it turns out not to contribute to the flow of the more interesting coupling constants, and we will soon see that we have more pressing problems to worry about when it comes to maintaining the form of the Hamiltonian under the action of a block spin transformation. As shown in Fig. 9.4, we will consider block spins composed of three microscopic spins each. We adopt a notation in which block spins are identified σ_s and capital roman subscripts and, since they will be averaged over, we’ll use a s_i^I to index each of the three microscopic spins corresponding to a given block spin. Whereas in the 1D case we used a “tie-breaker” spin (which amounted to

¹³¹Ref. [70]

completely integrating out every other degree of freedom), here each block spin contains an odd number of spins, so the block spins will be determined by majority rule:

$$\sigma_I = \text{sign} (s_1^I + s_2^I + s_3^I). \quad (9.38)$$

Note that, as enumerated in Table 9.1, each value for a block spin corresponds to four configurations of the underlying microscopic spins.

σ_I	s_1^I	s_2^I	s_3^I
+1	+	+	+
	+	+	-
	+	-	+
	-	+	+
-1	-	-	-
	-	-	+
	-	+	-
	+	-	-

Table 9.1: Microscopic spin configurations corresponding to a given block spin

9.4.1 Niemeijer-van Leeuwen Cumulant Expansion

With this set-up and choice of a block spin, how should we proceed? Our goal is to identify how this block spin transformation causes the coupling constant(s) of the Hamiltonian to flow. Anticipating an aspect of this problem which we saw even in the case of decimation of the 1D chain feature we found (in Eq. 9.29, where the form of the renormalized Hamiltonian looked quite different from the original, but which we only needed to agree with the form of the original Hamiltonian on the allowed values of the block spins), we'll write our condition that the partition function be the same before and after a block spin transformation as

$$\exp(-\beta\mathcal{H}'\{\sigma_I\}) = \sum_{\{s_i\}} \exp(-\beta\mathcal{H}\{s_i^I|\sigma_I\}). \quad (9.39)$$

We will try to follow a “perturbative” approach to this problem, albeit a perturbative approach of a very strang flavor. We'll write $\mathcal{H} = \mathcal{H}_0 + V$, where \mathcal{H}_0 corresponds to all interactions between spins entirely within a single block, and V corresponds to interactions between blocks. That is, we start with

$$-\beta\mathcal{H}_0 = k \sum_I \sum_{i \neq j \in I} s_i^I s_j^I \quad (9.40)$$

$$-\beta V = k \sum_{I \neq J} \sum_{i \in I, j \in J} s_i^I s_j^J. \quad (9.41)$$

Something of a ludicrous perturbation theory! It is the same constant, k , that sets the scale of both \mathcal{H}_0 and V . This is clearly not so much a perturbative calculation in the sense of expanding via a small parameter, but rather just a scheme with which to calculate.

In this scheme, we'll adopt the notation from the last chapter in which $\langle A \rangle_{(0)}$ means the average with respect to the \mathcal{H}_0 part of the Hamilton. For observables which depend on the value of the block spins, this average corresponds to averaging over spins consistent with the value of the block spins, weighted by the relevant exponential factor of \mathcal{H}_0 . That is,

$$\langle A(\sigma_I) \rangle_{(0)} = \frac{\sum_{\{s_i\}} \exp(-\beta \mathcal{H}_0 \{s_i^I | \sigma_I\}) A(\sigma_I)}{\sum_{\{s_i\}} \exp(-\beta \mathcal{H}_0 \{s_i^I | \sigma_I\})}. \quad (9.42)$$

We can now write Eq. 9.39 as

$$\exp(-\beta \mathcal{H}' \{ \sigma_I \}) = \langle e^{-\beta V} \rangle_{(0)} \sum_{s_i} \exp(-\beta \mathcal{H}_0(s_i^I | \sigma_I)). \quad (9.43)$$

The blue sum on the right-hand side corresponds to what we'll call Z_0 , the partition function for one block given a specific (± 1) value of σ_I :

$$Z_0 = \sum_{s_1^I, s_2^I, s_3^I | \sigma_I} \exp[k(s_1^I s_2^I + s_2^I s_3^I + s_1^I s_3^I)] = e^{3k} + 3e^{-k}, \quad (9.44)$$

where a direct calculation of the sum given Table 9.1 confirms that Z_0 is the same regardless of whether σ_I is plus or minus one. We will get a copy of Z_0 for every block spin in the system, leaving us to solve

$$e^{-\beta \mathcal{H}'} = Z_0^{N/3} \langle e^{-\beta V} \rangle_{(0)}. \quad (9.45)$$

We have seen the idea of using the log to go from a generator of moments to a generator of cumulants, and that's exactly how we proceed here, transforming the above into an expression for the new Hamiltonian as a cumulant expansion of averages with respect to \mathcal{H}_0 :

$$-\beta \mathcal{H}' = \frac{N}{3} \log Z_0 + \langle -\beta V \rangle_{(0)} + \frac{1}{2} \left[\langle (-\beta V)^2 \rangle_{(0)} - \langle -\beta V \rangle_{(0)}^2 \right] + \dots \quad (9.46)$$

9.4.2 Critical behavior at 1st-order

We begin by calculating the first moment of V defined in Eq. 9.41. For convenience, we define

$$\beta V = \sum_{I \neq J} -\beta V_{IJ},$$

where up to labels and indexing (c.f. Fig. 9.4) the inter-block interactions are

$$-\beta V_{IJ} = k s_2^I (s_1^J + s_3^J).$$

We note that the spins within a block are all equivalent to each other, and that \mathcal{H}_0 does *not* couple spins in different blocks together. This fact powers the cumulant expansion we are trying to carry out, as it implies that s_i^I and s_j^J are independent when averaging with respect to the weights given by \mathcal{H}_0 :

$$\langle -\beta V_{IJ} \rangle_{(0)} = 2k \langle s_2^I s_1^J \rangle_{(0)} = 2k \langle s_2^I \rangle_{(0)} \langle s_1^J \rangle_{(0)}. \quad (9.47)$$

To evaluate the average of a specific spin within a block, we combine the table of configurations in Table 9.1 with the definition of an ensemble average

$$\langle s_1^I \rangle_{(0)} = \frac{1}{Z_0} \sum_{\{s_i | \sigma_I\}} s_1^I \exp [k(s_1^I s_2^I + s_2^I s_3^I + s_1^I s_3^I)]. \quad (9.48)$$

Explicitly performing the sum we see that

$$\langle s_1^I \rangle_{(0)} = \begin{cases} \frac{1}{Z_0} (e^{3k} + e^{-k}) & \text{for } \sigma_I = +1 \\ \frac{1}{Z_0} (-e^{3k} - e^{-k}) & \text{for } \sigma_I = -1 \end{cases} \quad (9.49)$$

Defining for convenience the function

$$\phi_1(k) = \frac{e^{3k} + e^{-k}}{e^{3k} + 3e^{-k}} \quad (9.50)$$

we see that the average of 1 spin within a block is simply $\langle s_1^I \rangle_{(0)} = \phi_1(k)\sigma_I$. Combining the above, we find

$$\beta V = \sum_{I \neq J} -\beta V_{IJ} = 2k\phi_1^2(k) \sum_{\langle IJ \rangle} \sigma_I \sigma_J. \quad (9.51)$$

Fantastic! This relation $-k' = 2k\phi_1^2(k)$ – is precisely the kind of flow of coupling constants we had hoped to calculate, and we can now ask what it implies about the potential critical points. We first look for fixed points of this equation. When k is small we Taylor expand to find $k' \approx k/2 + k^2 + \dots$ – i.e., when k is small k' becomes smaller still, so $k = 0$ is a stable fixed point (corresponding again to the high temperature disordered phase). In the opposite limit, for very large k we have $k' \approx 2k$, so in this calculation $k \rightarrow \infty$ is also a stable fixed point (a low-temperature ordered phase).

With stable fixed points at both zero and infinity, there must be some nontrivial *unstable* or critical fixed point at some finite k (or, equivalently, at some finite temperature). We are looking for some k^* such that

$$k^* = 2k^* \left(\frac{e^{3k^*} + e^{-k^*}}{e^{3k^*} + 3e^{-k^*}} \right)^2.$$

Defining $x = \exp(4k^*)$, we see that this nontrivial fixed point is at

$$x = 1 + 2\sqrt{2} \Rightarrow k^* = \frac{1}{4} \log(1 + 2\sqrt{2}) \approx 0.3356 \dots \quad (9.52)$$

Compared to Onsager's exact result of $k^* \approx 0.274$ [71] this is certainly an improvement on the mean field result of $k^* = 1/6$ (as seen in Sec. 2.4.2). However: the location of the critical point is sensitive to all of the details of a theory, and much more important is to understand the critical exponents. We follow the program outlined earlier in this chapter, and first evaluate how the coupling constant flows near the critical point:

$$\lambda_k = \left. \frac{\partial k'}{\partial k} \right|_{k=k^*} = \left[\frac{2(1+x)(3+x^2+4x)(1+4k)}{(3+x)^3} \right]_{k=k^*} \approx 1.6235 \dots \quad (9.53)$$

We connect this derivative to the critical exponent describing the correlation length ($\xi = |t|^{-\nu}$ for $\nu = 1/y_t$) as usual:

$$\lambda_k = l^{y_t} \Rightarrow y_t = \frac{\log \lambda_k}{\log l}. \quad (9.54)$$

The length scale of our block spin transformation (i.e., the lattice spacing of the block spins compared to the microscopic lattice spacing) $l = \sqrt{3}$, giving us

$$y_t \approx 0.8822 \dots \Rightarrow \nu \approx 1.1335 \dots \quad (9.55)$$

The exact result is $\nu = 1$, and our 1st-order estimate is indeed an improvement over the mean-field result of $\nu = 1/2$.

9.4.3 Critical behavior at 2nd-order

Foundational to a standard definition of a successful physical theory is not merely that it gives a more satisfactory answer than another theory, but that we have some confidence that were we to work harder we would find ever-closer agreement with experimental results. With that perspective, we continue the Niemeijer-van Leeuwen calculation to the next order. Fundamentally we need to calculate the second cumulant of $(-\beta V)$, for which we write out all of the terms:

$$\frac{k^2}{2} \left[\sum_{\substack{I \neq J \\ i \in I, j \in J}} \sum_{\substack{M \neq L \\ m \in M, l \in L}} \left(\langle s_i^I s_j^J s_m^M s_l^L \rangle_{(0)} - \langle s_i^I s_j^J \rangle_{(0)} \langle s_m^M s_l^L \rangle_{(0)} \right) \right]. \quad (9.56)$$

We already know that $\langle s_i^I s_j^J \rangle_{(0)} = \phi_1^2(k) \sigma_I \sigma_J$, and that furthermore if the pair of blocks I and J are completely distinct from blocks M and L then

$$\langle s_i^I s_j^J s_m^M s_l^L \rangle_{(0)} = \langle s_i^I s_j^J \rangle_{(0)} \langle s_m^M s_l^L \rangle_{(0)},$$

and hence all of these terms will have their contribution canceled by the analogous terms coming from the square of the first moment. What remains are contributions from terms in which one label between block-spin pairs is shared, with representative examples shown in Fig. 9.5.

Let's first calculate the contribution from the left-most graph in Fig. 9.5, for which $L = J$. The square of the first moment looks like

$$\langle s_i^I s_j^J \rangle_{(0)} \langle s_m^M s_j^J \rangle_{(0)} = \phi_1^4(k) \sigma_I \sigma_M,$$

where we have noted that $\sigma_j^2 = 1$. We need only now calculate the corresponding second moment terms that appear:

$$\langle s_2^I (s_1^J + s_3^J) s_3^J (s_1^M + s_2^M) \rangle_{(0)} = \langle s_2^I (s_1^M + s_2^M) + s_2^I s_1^J s_3^J (s_1^M + s_2^M) \rangle_{(0)} \quad (9.57)$$

$$= 2\phi_1^2 \left(1 + \langle s_1^J s_3^J \rangle_{(0)} \right) \sigma_I \sigma_M. \quad (9.58)$$

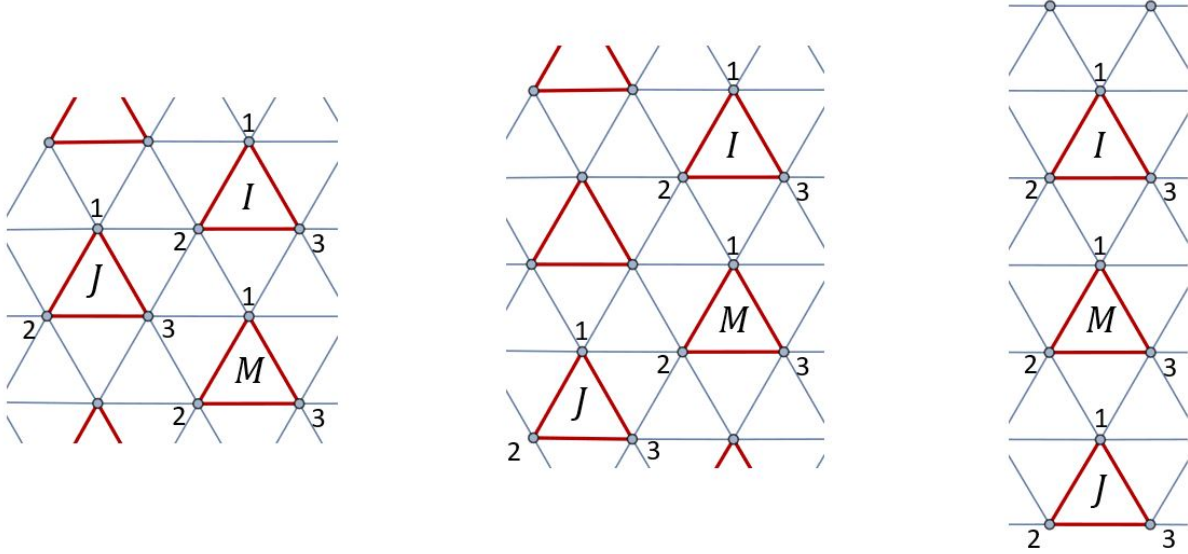


Figure 9.5: **Graphs for the second-order cumulant expansion** (Left) A graph contributing to nearest-neighbor block-spin interactions (Middle) A graph contributing to next-nearest-neighbor block-spin interactions. (Right) A graph contributing to more distant next-nearest-neighbor block-spin interactions.

We again directly evaluate the remaining two-spin-on-the-same-block expectation value, discovering that regardless of the sign of σ_J we get

$$\phi_2(k) \equiv \langle s_1^J s_3^J \rangle_{(0)} = \frac{e^{3k} - e^{-k}}{e^{3k} + 3e^{-k}}. \quad (9.59)$$

This is not so bad: what we've been calling k' gets a contribution from this second-order calculation that are like

$$2k^2 \phi_1^2 (1 + \phi_2 - 2\phi^2),$$

further renormalizing the nearest-neighbor coupling.

The problem is that Fig. 9.5 contains other graphs, too, corresponding to two *new* kinds of block-spin interactions (both next-nearest-neighbor, and which we'll label k_2 and k_3 for the middle and right graphs, respectively). For example, consider the right-most graph in the figure; the second-moment term contains terms like

$$\langle s_1^M (s_2^I + s_3^I) s_1^J (s_2^M + s_3^M) \rangle_{(0)} \rightarrow 4 \langle s_2^I \rangle_{(0)} \langle s_1^J \rangle_{(0)} \langle s_1^M s_2^M \rangle_{(0)}. \quad (9.60)$$

Subtracting of the square of the first moment contributions, we find that at this order the next-nearest-neighbor coupling constant gets a contribution like

$$k'_3 = 4k^2 (\phi_1^2 \phi_2 - \phi^4). \quad (9.61)$$

What are we to do? We have discovered that our block spin transformations starts generating new interactions as we work to higher order in our cumulant expansion (and which we certainly did *not* take into account when we were working at first order). How

can we organize all of these terms so that we are self-consistently working to a given order of our expansion? Niemeijer and van Leeuwen’s strategy was to assume that the coupling constants generated at different orders in the cumulant expansion have different orders of magnitude: k is a first order term and always appears, $k_2, k_3 \ll k$ are second order terms, and more and more smaller and smaller coupling constants appear as we move to higher orders in the cumulant expansion. None of this changes \mathcal{H}_0 (since all new interactions couple different blocks together), but we need to accommodate them in the inter-block piece,

$$-\beta V = k \sum_{\langle I \neq J \rangle} \sigma_I \sigma_J + k_2 \sum_{\langle I \neq J \rangle_2} \sigma_I \sigma_J + k_3 \sum_{\langle I \neq J \rangle_3} \sigma_I \sigma_J + \dots,$$

where in an abuse of notation we use $\langle I \neq J \rangle_i$ to stand for the collection of spins corresponding to the k_i interaction.

Working consistently to second order (i.e., in this case, including $\{k, k_2, k_3\}$ terms in the calculation of the first cumulant but including only k terms in the calculation of the second cumulant). The result is a system of equations for how the coupling constants evolve:

$$k' = 2k\phi_1^2 + 4k^2(\phi_1^2 + \phi_1^2\phi_2 - 2\phi_1^4) + 3k_2\phi_1^2 + 2k_3\phi_1^2 \quad (9.62)$$

$$k'_2 = k^2(7\phi_1^2\phi_2 + \phi_1^2 - 8\phi_1^4) + k_3\phi_1^2 \quad (9.63)$$

$$k'_3 = 4k^2(\phi_1^2\phi_2 - \phi_1^4). \quad (9.64)$$

Does this equation have a non-trivial fixed point (and, moreover, one that satisfies the underlying assumption that near the critical point different coupling constants are of different orders)? Indeed, it does! One can find that

$$\vec{k}^* \approx \{0.278869\dots, -0.014248\dots, -0.0152313\dots\},$$

a fixed points in which k is roughly an order of magnitude larger than the “second order” terms k_2 and k_3 . We can compute the matrix of derivatives of \vec{k}' with respect to \vec{k} evaluated at this fixed point¹³²,

$$\bar{M}_l \approx \begin{pmatrix} 1.831\dots & 1.345\dots & 0.896\dots \\ -0.00519\dots & 0 & 0.448\dots \\ -0.0781\dots & 0 & 0 \end{pmatrix}. \quad (9.65)$$

So, is this better than the first-order calculation? The critical point is at $k^* \approx 0.257\dots$, which is indeed closer than our first-order estimate to the exact value. The largest eigenvalue of the above matrix gives

$$\lambda_k \approx 1.7728 \Rightarrow y_t \approx 1.0424 \Rightarrow \nu \approx 0.959\dots, \quad (9.66)$$

which is again closer to the exact exponent of one than the 1st order calculation’s result!

Kardar “likely accidental”, Hemmer “asymptotic series”, Goldenfeld “results are converging but non-uniformly”, Reichl “works well”, Plischke & Bergersen “not encouraging.” Tough to go to much higher order: start generating more and more long-range interactions, 4- and 6-spin interactions, etc etc. Not how one really computes exponent values anyway, but good intuition?

[69] for “does it converge or not”, together with other texts

¹³²The particularly eagle-eyed reader will notice some small numerical differences between this matrix and the corresponding one in Niemeijer and van Leeuwen’s work [70] – this is entirely due to having a slightly better numerical estimate of the fixed point.

9.5 Universality and a momentum-space formulation

Gaussian model. motivate with “how do we get universality back?”

9.6 Problems

Chapter 10

Fluctuations and non-equilibrium physics

Intro – in equilibrium, we've already seen some fluctuation-response relationships... etc...
whole chapter needs to be added!

- 10.1 Brownian motion – Einstein, Langevin, and Smoluchowski
- 10.2 Fokker-Planck
- 10.3 Fluctuation-dissipation
- 10.4 Onsager – reciprocity and non-reciprocity

10.5 Problems

Part IV
Appendices

“Murder your darlings,” better writers than I have advised [72]. The less cold-hearted of us merely shuffle them off to an appendix.

The appendices are mostly where I’ve shuffled TeX’d notes that used to be a core part of the class, but which I don’t typically include in lectures anymore... some will eventually remain, some are sufficiently duplicative of existing text that I’ll just say “Go look at Kardar, Chapter X for details”

Appendix A

Thermodynamics

A.1 2nd Law: from fridges to entropy¹³³

We'll start quantifying these statements by defining what engines and refrigerators are, along with “figures of merit” for how they are performing. An *engine* is a machine which takes Q_H of heat from a source, converts some of it to work W , and dumps some Q_C of it into a heat sink. Its *efficiency*, η , is

$$\eta = \frac{W}{Q_H} = \frac{Q_H - Q_C}{Q_H}, \quad (\text{A.1})$$

and Kelvin says that $\eta < 1$.

A *refrigerator* is just an engine running in reverse: work is done in order to take some amount of heat Q_C out of a cold place and dumps Q_H as exhaust into a hotter place (moving heat from cold to hot). The *performance* of a refrigerator, ω , is

$$\omega = \frac{Q_C}{W} = \frac{Q_C}{Q_H - Q_C} \quad (\text{A.2})$$

To see that Kelvin's and Clausius' formulations are equivalent, one can play games involving connecting engines and refrigerators together. If you use an ideal engine to run a refrigerator, you end up with an ideal refrigerator (i.e., “not Kelvin's statement implies not

¹³³Kardar, Stat Phys Particles., Ch.1

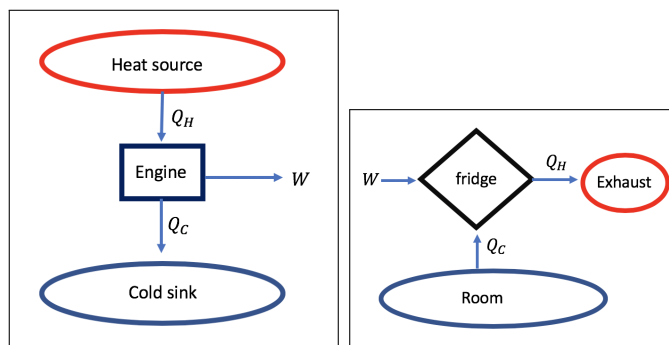


Figure A.1: Idealized engine (left) and refrigerators (right)

Clausius’ statement”). Similarly, if you use the heat exhausted from an ideal refrigerator to power an engine you get an ideal engine (i.e., “not Clausius’ statement implies not Kelvin’s statement”). Taken together, you get Kelvin’s statement \iff Clausius’ statement.

These definitions probably seem pretty trivial¹³⁴. And yet, with an excursion through thinking about Carnot Engines, we’ll see that these formulations let us answer the question I thought you should pose in Section 1.3, when we wrote:

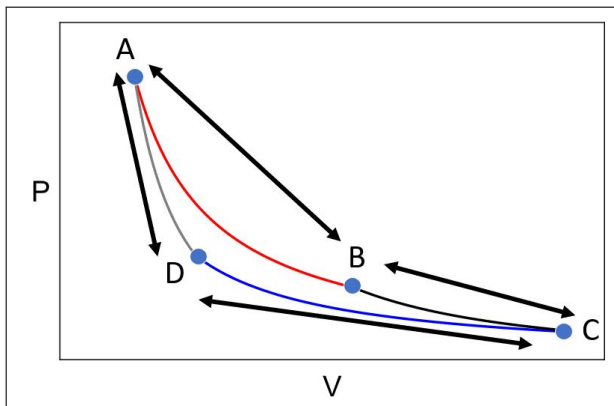
$$dE = \sum_i J_i dx_i + \boxed{?}$$

A.1.1 Carnot Engines

A *Carnot Engine* (CE) is any engine that (1) is reversible, (2) runs in a cycle, and (3) operates by exchanging heat with a source temperature T_H and a sink temperature T_C . The reversibility of a CE is like a generalization of a kind of “frictionless” condition in mechanics, letting us go forward and backwards along a path just by reversing the inputs and outputs of the CE. The cyclic condition means that we can think of the start and end points of the work-path the CE traces out as the same. Finally, the last condition is more precise than the figure we drew in A.1: we insist that the heat sources and sinks have well-defined thermodynamic temperatures.

Example: Ideal gas Carnot Cycle

As an example of a system that could satisfy the Carnot Engine conditions, let’s imagine making a CE with an ideal gas. The task is essentially to chose the paths through Pressure-Volume space we will take that will form a cycle and let us run everything reversibly, while operating between exactly two temperatures. We will select two isotherms to serve as the heat exchanges that the engine will operate between, but how can we pick reversible adiabatic paths *between* the temperatures?



¹³⁴Perhaps, even, a bit dull!

Figure A.2: **Schematic of a Carnot Cycle for an ideal gas**, where two of the path segments are along isotherms (blue and red curves), and two adiabatic paths between the isotherms (black and grey curves).

A reversible transformation must be quasi-static, and for an ideal gas we know $E = \frac{3}{2}Nk_B T = \frac{3}{2}PV$. Along a quasi-static path we have

$$\delta Q = dE - \delta W = d\left(\frac{3}{2}PV\right) + pdV = \frac{5}{2}PdV + \frac{3}{2}VdP,$$

and the adiabatic ($\delta Q = 0$) gives us

$$0 = \frac{dP}{P} + \frac{5}{3} \frac{dV}{V} \Rightarrow PV^{5/3} = \text{constant}. \quad (\text{A.3})$$

A schematic of such an ideal gas CE cycle is illustrated in Fig. A.2. In the “forward” cycle ($A \rightarrow B \rightarrow C \rightarrow D \rightarrow A$) heat is used from A to B to expand the gas, the gas does work as it now continues to (adiabatically) expand B to C , is isothermally compressed from C to D , and finally it is adiabatically compressed from D back to A . Work (positive and negative) is done by the gas along each segment, with the total work done equal to the area enclosed by the cycle in the P - V plane.

Now, *Carnot’s Theorem* states that, of all imaginable engines operating between T_H and T_C , a Carnot engine is the most efficient¹³⁵. This can be proved by playing the usual kind of engine games (you, the reader, should draw the following pictures!). First, take a Carnot Engine, and use a non-Carnot-Engine’s output to run the CE as a refrigerator. Let primes refer to heat connected to the Carnot engine, and unprimes to the NCE. The net effect is to transfer heat $Q_H - Q'_H = Q_C - Q'_C$ from T_H to T_C . Clausius’ formulation of the 2nd law tells us you can’t transfer negative heat, so $Q_H \geq Q'_H$. But the amount of work, W , is the same for both engines, so

$$\frac{W}{Q_H} \leq \frac{W}{Q'_H} \Rightarrow \eta_{CE} \geq \eta_{NCE}. \quad (\text{A.4})$$

Thermodynamic Temperature Scale

We saw that the 0th law let us define an empirical temperature scale, and now we will see that the 2nd law lets us define a thermodynamic temperature scale. We established in the example above that we can (in theory) build a Carnot engine using an ideal gas, and it is straightforward to show that *all* Carnot engines operating between the same T_H and T_C have the same efficiency¹³⁶.

¹³⁵I never met Carnot, but apparently he was the kind of guy to write theorems like “Carnot is the best.”

¹³⁶Play the usual sorts of games: use one Carnot Engine to run the other in reverse. Then switch which engine is running forward vs backwards, demonstrating that $\eta_{CE_1} = \eta_{CE_2}$.

This is already quite interesting: since the efficiency of a CE is independent of the workings of the machine itself, the only thing the efficiency can depend on are the temperatures it operates between, i.e., we have $\eta(T_H, T_C)$. Already, *if* you can build any CE and if you know its efficiencies at different T , you can define T independent of any material properties. We can make even more progress by running two engines in series; one between T_1 and T_2 , and the other between T_2 and T_3 , as in Fig. A.3.

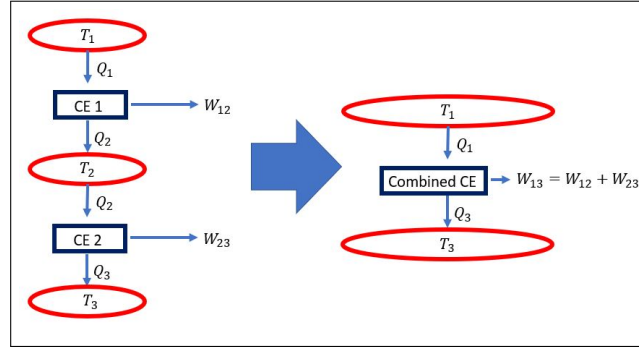


Figure A.3: **Schematic of Carnot engines in series**

In that figure, the working of CE1 tells us that

$$Q_2 = Q_1 - W_{12} = Q_1(1 - \eta(T_1, T_2)), \quad (\text{A.5})$$

and the working of the second CE tells us

$$Q_3 = Q_2 - W_{23} = Q_2(1 - \eta(T_2, T_3)) = Q_1(1 - \eta(T_1, T_2))(1 - \eta(T_2, T_3)). \quad (\text{A.6})$$

But we also know, viewing the series as a single engine, that

$$Q_3 = Q_1 - W_{13} = Q_1(1 - \eta(T_1, T_3)). \quad (\text{A.7})$$

Comparing our two expressions for Q_3 we see that

$$(1 - \eta(T_1, T_3)) = (1 - \eta(T_1, T_2))(1 - \eta(T_2, T_3)); \quad (\text{A.8})$$

this is a *constraint on the functional form* that η can take. We postulate, as something that satisfies the constraint, that

$$(1 - \eta(T_1, T_2)) = \frac{Q_2}{Q_1} \equiv \frac{f(T_2)}{f(T_1)}. \quad (\text{A.9})$$

By convention, we choose $f(T) = T$, giving us

$$\eta(T_H, T_C) = \frac{T_H - T_C}{T_H}. \quad (\text{A.10})$$

We've done it! Up to a constant of proportionality, Eq. A.10 defines a thermodynamic temperature (and, if we like, we can again set the arbitrary constant with the help of the

measured triple point of water-ice-steam). If we felt like it, we could run a Carnot cycle for an ideal gas and show that the ideal gas scale and the thermodynamic temperature scale are, in fact, identical¹³⁷ As a fun note in closing: unlike it statistical mechanics, in thermodynamics *temperatures must be positive*, otherwise we see that Kelvin’s formulation of the 2nd law could be violated, breaking the structure of much of what we’ve written above.

A.1.2 Clausius’ Theorem

Why (you may be wondering) are we pretending that these games – hooking up engines to refrigerators, and defining Carnot Engines, and thinking about efficiencies – are deep or interesting? Why did I claim that they would help us understand the thermodynamic coordinate we should think of as conjugate to temperature? We only need one more piece, and that is:

Clausius’ Theorem: For any cyclic process, with path parameterized by s

$$\oint \frac{\delta Q(s)}{T(s)} \leq 0, \quad (\text{A.11})$$

where the heat $\delta Q(s)$ is an amount of heat delivered to the system by a reservoir or machine at temperature $T(s)$.

To prove this, we will (of course) hook the system up to a Carnot Engine. The reversibility of the Carnot Engine is quite helpful: $\delta Q(s)$ can be of any sign, so we can think of the whole cyclic process as a series of infinitesimal cycles in which the CE delivers whatever $\delta Q(s)$ needs to be. Let the Carnot engine be at a given reference temperature T_0 , as in Fig. A.4A. To prove the theorem, simply reinterpret the setup by pretending the Carnot engine and the system are a single “device,” giving us the picture in Fig. A.4B.

This looks dumb, but we’re done! To deliver heat at a specified temperature, our efficiency functions from above tell us that $\delta Q_0 = T_0 \delta Q / T(s)$. From this view the net extracted heat is

$$\oint \delta Q_0(s) = \oint T_0 \frac{\delta Q(s)}{T(s)} \leq 0, \quad (\text{A.12})$$

where the last inequality is simply Kelvin’s formulation of the 2nd law: we can’t convert 100% of some amount of heat to work.

As with so many of our Carnot engine manipulations, this probably feels trivial. However: there are major consequences!

¹³⁷This is not especially *useful*, but it does conceptually show that for both scales the temperature is not something that depends on the properties of a particular material.

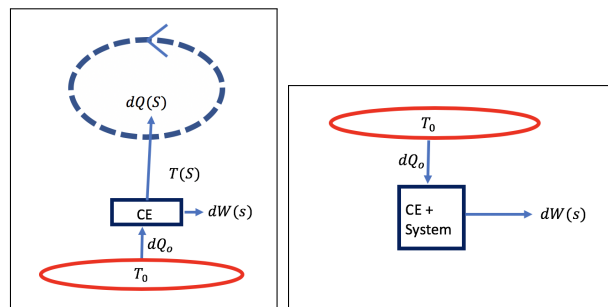


Figure A.4: Clausius theorem setup (left) and reinterpretation (right)

Appendix B

Probability

B.1 Basic Definitions

Random variable: A random variable x is a measurable variable described by a set of possible outcomes, S . This could be a discrete set, as for a coin: $S_{\text{coin}} = \{\text{heads, tails}\}$, or a continuous range, as for a particle’s velocity: $S_{v_x} = \{-\infty \leq v_x \leq \infty\}$. We call each element of such a set as an *event* $E \subset S$

Probability We will assign a value, called the *probability* to each event, denoted $p(E)$, which has the following properties:

1. *positivity:* $p(E) \geq 0$
2. *additivity:* $p(A \text{ or } B) = p(A) + p(B)$ if A and B are distinct.
3. *normalization:* $p(S) = 1$.

Now, we’re not not mathematicians, and so we won’t be starting from this definition and proving things. You might wonder, as a practical matter, “how will we be determining $p(E)$ as we go about our lives?” By one of two ways:

1. *Objectively* (or, we might say, “experimentally,” as frequentists): $p(E)$ is the frequency of outcome in many trials:
$$p(E) = \lim_{N \rightarrow \infty} N_E / N$$
2. *Subjectively* (or, we might say, “theoretically,” as good Bayesians): We will determine $p(E)$ based on our uncertainty among all outcomes.

We will be using this “subjective” version of things *repeatedly* in statistical physics¹³⁸, so we will formalize this way of assigning probabilities later in the chapter.

¹³⁸And, anyway, who has time to repeatedly measure the same thing over and over and over again?

B.2 Properties of single random variables

Let's focus on random variables which are continuous and real-valued (the specialization to discrete ones is straightforward; we'll see an example in the next section). Here are some essential definitions:

Cumulative probability function: $P(x) = \text{probability}(E \subset [-\infty, x])$. This must be a monotonically increasing function, with $P(-\infty) = 0$ and $P(\infty) = 1$.

Probability density function (pdf): $p(x) = dP(x)/dx$, so that $p(x)dx = \text{probability}(E \subset [x, x + dx])$. The pdf must be everywhere positive, and normalized so that $\int_{-\infty}^{\infty} p(x)dx = 1$. *Note:* $P(x)$ is dimensionless, which means $p(x)$ has dimensions of $[x^{-1}]$.

Expectation value $\langle F(x) \rangle = \int dx p(x)F(x)$

Change of variables Suppose x is a random variable and we have a function of it, $F(x)$. We can regard $F(x)$ itself as a new random variable. The probability that we find F in some interval between f and $f + df$ is written $p_F(f)df$, where¹³⁹ $|p_F(f)df| = |p_X(x)dx|$, since the probability in some differential *area* must should be invariant under a change of variables. (see Fig. B.1).

Something that sometimes trips people up when changing variables: we need to add up all the contributions of x that contribute. Thus, we first solve $F(x) = f$, and call the set of solutions x_i . We can then write:

$$p_F(f) = \sum_i p(x_i) \left| \frac{dx}{df} \right|_{x_i}. \quad (\text{B.1})$$

In standard texts you will usually see the $\frac{dx}{df}$ written in terms of inverse functions, but the meaning is clear (and it is the slope). Note that the factors of $\left| \frac{dx}{df} \right|$ are the *Jacobians* used to change variables.

Simple example Suppose we have a random variable x with pdf $p_X(x) = 3x^2$ defined in the interval $0 < x < 1$, and we look at a new random variable $Y = X^2$. This is easily invertible in the range, and we can write $x(y) = \sqrt{y}$, and $dx/dy = y^{-1/2}/2$. Thus

$$p_Y(y) = p_X(x) \left| \frac{1}{2\sqrt{y}} \right| = \frac{3}{2} (\sqrt{y})^2 \frac{1}{\sqrt{y}} = \frac{3}{2} \sqrt{y},$$

defined in the range $0 < y < 1$.

¹³⁹One can go back to the definition of the cumulative probability function to show this is true for monotonically increasing or decreasing functions; it is more work to show it for the sorts of piecewise monotonic functions we might want to consider

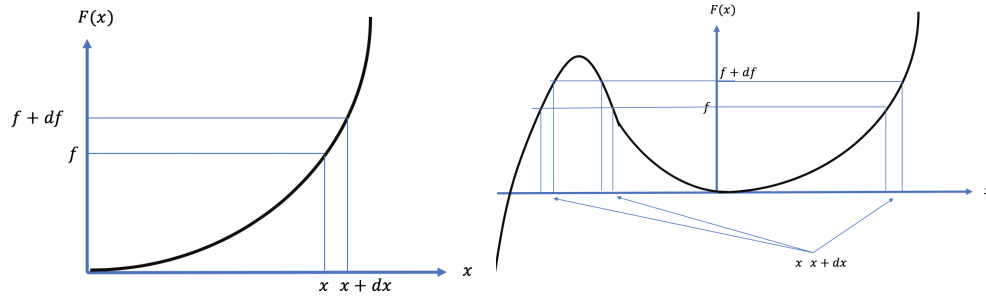


Figure B.1: Changing variables from x to $F(x)$

2-valued example Suppose instead that we have a random variable x where

$$p(x) = \frac{\lambda}{2} \exp(-\lambda|x|),$$

defined for any x on the real line. We want to know the probability density function for the random variable $F(x) = x^2$. There are, by inspection, two solutions to $F(x) = f$ (when f is positive!), and they are $x = \pm\sqrt{f}$. The derivatives we need are $|dx/df| = |\pm \frac{1}{2\sqrt{f}}|$. Thus, we have:

$$p_F(f) = \frac{\lambda}{2} \exp(-\lambda|\sqrt{f}|) \left| \frac{1}{2\sqrt{f}} \right| + \frac{\lambda}{2} \exp(-\lambda|-\sqrt{f}|) \left| \frac{-1}{2\sqrt{f}} \right| = \frac{\lambda \exp(-\lambda\sqrt{f})}{2\sqrt{f}},$$

for any $f > 0$ (and $p_F(f) = 0$ for $f < 0$).

Moments We define the n th moment to be $m_n \equiv \langle x^n \rangle = \int dx x^n p(x)$

B.3 Important distributions

B.3.1 Binomial distribution

Given a discrete random variable with two outcomes, which occur with probability p_A and $p_B = 1 - p_A$, the binomial distribution gives the probability that event A occurs exactly N_A times out of N trials. It is equal to

$$P_N(N_A) = \binom{N}{N_A} p_A^{N_A} p_B^{N-N_A}, \quad \binom{N}{N_A} = \frac{N!}{N_A!(N-N_A)!}. \quad (\text{B.2})$$

The characteristic function for the discrete distribution is

$$\tilde{p}_N(k) = \langle e^{-ikN_A} \rangle = \sum_{N_A=0}^N \frac{N!}{N_A!(N-N_A)!} p_A^{N_A} p_B^{N-N_A} e^{-ikN_A} = (p_A e^{-ik} + p_B)^N. \quad (\text{B.3})$$

This has the properties that we can easily relate the cumulant generating function for the N -trial case to that of the 1-trial case:

$$\log \tilde{p}_N(k) = N \log (P_A e^{-ik} + p_B) = N \log \tilde{p}_1(k). \quad (\text{B.4})$$

For a single trial, N_A can only be either zero or one, which means that we must have $\langle N_A^m \rangle = p_A$ for all powers m . Combining this property of the moments with the above feature of the cumulants, we learn that the cumulants for the N -trial case are

$$\langle N_A \rangle_c = N p_A, \quad \langle N_A^2 \rangle_c = N (p_A - p_A^2) = N p_A p_B, \quad (\text{B.5})$$

and higher order cumulants can be easily calculated. We'll see that this type of feature – where there is a trivial relation between an independent thing repeated N times and the case of an individual trial – will be of great use as we build up statistical mechanics.

B.3.2 Poisson distribution

We'll get at the Poisson distribution, a continuous pdf, relating it to the binomial distribution. Consider a process in time where two properties hold. First, the probability of observing (exactly) one event in the interval $[t, t + dt]$ is proportional to dt in the limit $dt \rightarrow 0$. Second, suppose the probability of observing an event in different intervals is uncorrelated¹⁴⁰. Then, the *Poisson distribution* is the probability of observing exactly M events in the interval T .

We get the details of the distribution by imagining dividing up the interval T into many segments of length dt , say $N = T/dt \gg 1$ such that dt is so small the probability of observing more than one event is negligible. So, in each segment we have an event occurring with probability $p = \alpha dt$ and no event occurring with probability $q = 1 - p$. From our expression for the binomial distribution, we immediately know the characteristic function for this process:

$$\tilde{p}(k) = (p e^{-ik} + q)^N = \lim_{dt \rightarrow 0} (1 + \alpha dt (e^{-ik} - 1))^{T/dt} = \exp(\alpha (e^{-ik} - 1)T), \quad (\text{B.6})$$

where the last equality is an example of the famous Euler limit formula. Knowing the characteristic function, we can take the inverse Fourier transform to get the pdf:

$$p(x) = \int_{-\infty}^{\infty} \frac{dk}{2\pi} e^{ikx} \exp(\alpha (e^{-ik} - 1)T). \quad (\text{B.7})$$

This can be solved¹⁴¹ by expanding the exponential, and using

$$\int_{-\infty}^{\infty} \frac{dk}{2\pi} e^{-ik(x-M)} = \delta(x - M)$$

to get the probability of M events in a time T for a process characterized by α as

$$p_{\alpha T}(M) = e^{-\alpha T} \frac{(\alpha T)^M}{M!}. \quad (\text{B.8})$$

¹⁴⁰An example might be found, e.g., in radioactive decay

¹⁴¹As you, student in this class, should explicitly verify!

Additionally, the cumulants can be read off of the expansion of the log characteristic function

$$\log \tilde{p}_{\alpha T}(k) = \alpha T(e^{-ik} - 1) = \alpha T \sum_{n=1}^{\infty} \frac{(-ik)^n}{n!} \Rightarrow \langle M^n \rangle_c = \alpha T. \quad (\text{B.9})$$

That is, while the binomial distribution has the property that every moment is the same, the Poisson distribution has the property that every *cumulant* is the same.

B.3.3 Gaussian distribution

We will definitely, definitely, definitely be using Gaussians in this class. I hope you are not surprised to hear this. Define the gaussian pdf as

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\lambda)^2}{2\sigma^2}\right). \quad (\text{B.10})$$

The characteristic function is computed by the usual means of completing the square inside the integral, a trick I believe we all know:

$$\begin{aligned} \tilde{p}(k) &= \int \frac{dx}{\sqrt{2\pi\sigma^2}} \exp\left(-ikx - \frac{(x-\lambda)^2}{2\sigma^2}\right) \\ &= e^{-ik\lambda} \int \frac{dy}{\sqrt{2\pi\sigma^2}} \exp\left(-iky - \frac{y^2}{2\sigma^2} + \frac{k^2\sigma^2}{2} - \frac{k^2\sigma^2}{2}\right), \quad \text{for } y = x - \lambda \\ &= e^{-ik\lambda - \frac{k^2\sigma^2}{2}} \int \frac{dz}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-z^2}{2\sigma^2}\right), \quad \text{for } z = y + ik\sigma^2 \\ &= \exp\left(-ik\lambda - \frac{k^2\sigma^2}{2}\right). \end{aligned} \quad (\text{B.11})$$

This manipulation shows that the Fourier transform of a Gaussian is, itself, a Gaussian. The cumulants of this are easily identified:

$$\log \tilde{p}(k) = -ik\lambda - \frac{k^2\sigma^2}{2}, \quad (\text{B.12})$$

immediately showing that

$$\langle x \rangle_c = \lambda, \quad \langle x^2 \rangle_c = \sigma^2, \quad \langle x^{n>2} \rangle_c = 0. \quad (\text{B.13})$$

So, the Gaussian is completely specified by its first two cumulants, and all moments involve only products of one- and two-point clusters.

B.4 Properties of multiple random variables

Joint probability density function: We define, by analogy, the joint pdf $p(x_1, x_2, \dots, x_N)$ as

$$p(\mathbf{x}) = \lim_{\{dx_i \rightarrow 0\}} \frac{\text{prob. of outcome in } \{(x_1, x_1 + dx_1), \dots, (x_N, x_N + dx_N)\}}{dx_1 dx_2 \cdots dx_N}. \quad (\text{B.14})$$

The normalization of the joint PDF is

$$p_{\mathbf{x}}(S) = 1 = \int d^N \mathbf{x} p(\mathbf{x}), \quad (\text{B.15})$$

and *iff* the N random variables are *independent*, then the joint pdf simplifies to the product of the individual probability density functions:

$$p(\mathbf{x}) = \prod_{i=1}^N p_i(x_i). \quad (\text{B.16})$$

Joint characteristic function: This is just the N -dimensional Fourier transform:

$$\tilde{p}(\mathbf{k}) = \langle \exp(-i\mathbf{k} \cdot \mathbf{x}) \rangle = \int \left(\prod_i dx_i e^{-ik_i x_i} \right) p(x_1, \dots, x_N). \quad (\text{B.17})$$

Joint moments and cumulants: These are defined perfectly analogously with the moments and cumulants of single random variable distributions. We previously noted that moments are related to the coefficient of the relevant power of k in the expansion of the characteristic function and its log; more generally, we can express these as the following derivatives:

$$\langle x_1^{m_1} x_2^{m_2} \dots x_N^{m_N} \rangle = \left[\frac{\partial}{\partial(-ik_1)} \right]^{m_1} \dots \left[\frac{\partial}{\partial(-ik_N)} \right]^{m_N} \tilde{p}(\mathbf{k})|_{\mathbf{k}=0} \quad (\text{B.18})$$

$$\langle x_1^{m_1} x_2^{m_2} \dots x_N^{m_N} \rangle_c = \left[\frac{\partial}{\partial(-ik_1)} \right]^{m_1} \dots \left[\frac{\partial}{\partial(-ik_N)} \right]^{m_N} \log \tilde{p}(\mathbf{k})|_{\mathbf{k}=0}. \quad (\text{B.19})$$

As a simple – but perhaps the most important – example, the “co-variance” between two random variables is

$$\langle x_1 x_2 \rangle_c = \langle x_1 x_2 \rangle - \langle x_1 \rangle \langle x_2 \rangle. \quad (\text{B.20})$$

Note that the graphical expansion we wrote earlier still applies. All that is needed is to label the points under consideration by the variables they correspond to. See Fig. B.2, in which this is done via color.

$$\begin{aligned} \langle x_1^2 x_2 \rangle &= \begin{array}{c} \bullet \\ \bullet \end{array} + \begin{array}{c} \bullet \\ \bullet \end{array} + 2 \begin{array}{c} \bullet \\ \bullet \end{array} + \begin{array}{c} \bullet \\ \bullet \end{array} \\ &= \langle x_1 \rangle_c^2 \langle x_2 \rangle_c + \langle x_1 \rangle_c \langle x_2 \rangle_c + 2 \langle x_1 x_2 \rangle_c \langle x_1 \rangle_c + \langle x_1^2 x_2 \rangle_c \end{aligned}$$

Figure B.2: Graphical expansion of a joint moment

Unconditional probability: The unconditional pdf describes the pdf for a subset of random variables independent of what the others are doing:

$$p(x_1, \dots, x_m) = \int \left(\prod_{i=m+1}^N dx_i \right) p(x_1, \dots, x_N). \quad (\text{B.21})$$

For example, particles in a gas would generically have a pdf over both its position and velocity, $p(\mathbf{x}, \mathbf{v})$, but we might only care about the distribution of positions, so we would integrate out the velocities:

$$p(\mathbf{x}) = \int d^3\mathbf{v} p(\mathbf{x}, \mathbf{v})$$

Conditional probability: The conditional pdf describes the behavior of a subset of the random variables *given specified values for the other random variables*. Following the above example, suppose we are interested in the conditional probability of velocities *given* a set of position, denoted $p(\mathbf{v}|\mathbf{x})$. This should be proportional to the full joint PDF:

$$p(\mathbf{v}|\mathbf{x}) = \frac{p(\mathbf{x}, \mathbf{v})}{\mathcal{A}},$$

where the constant of proportionality is just the probability of having that value of position in the first place:

$$\mathcal{A} = \int d^3\mathbf{v} p(\mathbf{x}, \mathbf{v}) = p(\mathbf{x}).$$

Note that in the case of *independent* random variables, the conditional probability is the same as the unconditional probability.

Joint Gaussian distribution You might have thought the natural generalization of Eq. B.10 was

$$p(\mathbf{x}) = \frac{1}{\sqrt{\prod_{n=1}^N 2\pi\sigma_n^2}} \exp\left(\frac{-1}{2} \sum_{n=1}^N \frac{-(x_n - \lambda_n)^2}{\sigma_n^2}\right). \quad (\text{B.22})$$

but this neglects the potential for cross-correlations! The most general form is, instead,

$$p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^N \det \mathbf{C}}} \exp\left(\frac{-1}{2} \sum_{n,m=1}^N (x_n - \lambda_n)(x_m - \lambda_m) (\mathbf{C})_{nm}^{-1}\right), \quad (\text{B.23})$$

where the matrix \mathbf{C} is symmetric, and for $p(\mathbf{x})$ to be a well-defined probability the matrix \mathbf{C} must be positive definite. We can write this more compactly as

$$p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^N \det \mathbf{C}}} \exp\left(\frac{-1}{2} (\mathbf{x} - \boldsymbol{\lambda})^T \mathbf{C}^{-1} (\mathbf{x} - \boldsymbol{\lambda})\right). \quad (\text{B.24})$$

The matrix \mathbf{C} is called the covariance matrix. If one goes through and performs the fourier transform on the above joint pdf, one finds

$$\tilde{p}(\mathbf{k}) = \exp\left(-i\mathbf{k} \cdot \boldsymbol{\lambda} - \frac{1}{2} \mathbf{k}^T \mathbf{C} \mathbf{k}\right), \quad (\text{B.25})$$

or, in index notation,

$$\tilde{p}(\mathbf{k}) = \exp \left(-ik_m \lambda_m - \frac{1}{2} k_m C_{mn} k_n \right). \quad (\text{B.26})$$

The latter re-writing lets us immediately read off the joint cumulants of the joint Gaussian distribution:

$$\langle x_m \rangle_c = \lambda_m, \quad \langle x_m x_n \rangle_c = C_{mn}, \quad (\text{B.27})$$

with all higher-order cumulants vanishing.

Wick's theorem: Note that there is an important special case of joint Gaussian distributions when $\lambda = \mathbf{0}$. Consider the joint cumulant

$$\langle x_1^{n_1} x_2^{n_2} \cdots x_N^{n_N} \rangle,$$

and think about the combinatorics of the graphical expansion we've seen above.

First, if the sum of the n_i is *odd*, then in the graphical expansion there is no way to avoid a term with an odd-power cumulant, and in this special case of the joint Gaussian distribution with $\lambda = \mathbf{0}$, *all such terms are zero!*

Second, if the sum is even, we know that there will only be contributions from *combinations of covariances*: all even-power cumulants with power greater than two vanish because, again, we are dealing with the joint Gaussian. Thus, the cumulant can be obtained by all ways of summing over pairs of the random variables. For example,

$$\langle x_i x_j x_k x_l \rangle = C_{ij} C_{kl} + C_{ik} C_{jl} + C_{il} C_{jk},$$

where it didn't matter if the i, j, k, l were distinct. For instance:

$$\langle x_1^2 x_2 x_3 \rangle = C_{11} C_{23} + 2C_{12} C_{13}.$$

This property of the joint Gaussian distribution is sometimes summarized as:

$$\langle x_1^{n_1} x_2^{n_2} \cdots x_N^{n_N} \rangle = \begin{cases} 0 & \text{if } \sum_{\alpha} n_{\alpha} \text{ is odd} \\ \sum (\text{all pairwise contractions of covariances}) & \text{else} \end{cases}$$

In this formulation, we see the analogy of *Wick's Theorem* applied to fields.

Appendix C

Kinetic Theory – from BBGKY to Boltzmann¹⁴²

C.1 Boltzmann a la BBGKY

So, now that we know where we are heading, let's derive the Boltzmann equation with a bit more rigor. After having done so, we'll ask ourselves about the *consequences* of the Boltzmann equation. What do its solutions tell us about the behavior of equilibrium states? What does it say about the origin of irreversibility?

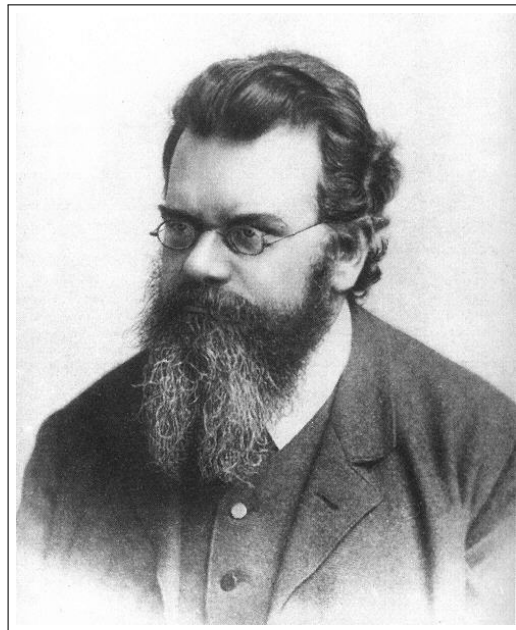


Figure C.1: “**Eleganz sei die Sache der Schuster und Schneider,**” surely what every student hopes to hear before a lecture

¹⁴²Kardar, Tong, Huang

C.1.1 BBGKY for a dilute gas

You may have noticed that so far we have not really used the fact that we are studying a dilute gas – that is about to change as we are finally in a position to make some physically motivated approximations to the BBGKY hierarchy. Let's start by explicitly writing the first two levels of the hierarchy, where for notational ease we'll write the derivative of the pairwise potential as a force: $\frac{\partial V(\mathbf{q}_i - \mathbf{q}_j)}{\partial \mathbf{q}_i} = \frac{\partial \Phi_{ij}}{\partial \mathbf{q}_i}$, which is the contribution to the force on i from j . Our first two levels are:

$$\left[\frac{\partial}{\partial t} - \frac{\partial U}{\partial \mathbf{q}_1} \cdot \frac{\partial}{\partial \mathbf{p}_1} + \frac{\mathbf{p}_1}{m} \cdot \frac{\partial}{\partial \mathbf{q}_1} \right] f_1 = \int dV_2 \frac{\partial \Phi_{12}}{\partial \mathbf{q}_1} \cdot \frac{\partial}{\partial \mathbf{p}_1} f_2, \quad (\text{C.1})$$

$$\left[\frac{\partial}{\partial t} - \frac{\partial U}{\partial \mathbf{q}_1} \cdot \frac{\partial}{\partial \mathbf{p}_1} - \frac{\partial U}{\partial \mathbf{q}_2} \cdot \frac{\partial}{\partial \mathbf{p}_2} + \frac{\mathbf{p}_1}{m} \cdot \frac{\partial}{\partial \mathbf{q}_1} + \frac{\mathbf{p}_2}{m} \cdot \frac{\partial}{\partial \mathbf{q}_2} - \frac{\partial \Phi_{12}}{\partial \mathbf{q}_1} \cdot \left(\frac{\partial}{\partial \mathbf{p}_1} - \frac{\partial}{\partial \mathbf{p}_2} \right) \right] f_2 = \int dV_3 \left[\frac{\partial \Phi_{13}}{\partial \mathbf{q}_1} \cdot \frac{\partial}{\partial \mathbf{p}_1} + \frac{\partial \Phi_{23}}{\partial \mathbf{q}_2} \cdot \frac{\partial}{\partial \mathbf{p}_2} \right] f_3. \quad (\text{C.2})$$

Relative importance of terms

Would you like to write/L^AT_EX the next level? Neither would I. Let's think physically about the terms in the above two levels: We've arranged things so that every term in the square brackets has dimensions of inverse time, so lets *estimate the typical magnitudes* of the various terms! We're studying a gas, and a reasonable speed for a gas particle at room temperature is on the order of $c = 100m/s$; to make a characteristic time, the typical length scale will depend on the nature of the term in question.

1. First, there is a characteristic time related to the external potential, like $\tau_U^{-1} \sim \frac{\partial U}{\partial \mathbf{q}} \cdot \frac{\partial}{\partial \mathbf{p}}$: these are spatial variations in the external potential, which we will typically think of taking place over basically *macroscopic* distances, where the characteristic length L is at least a millimeter. Very roughly, that would give us:

$$\tau_U \sim L/v \sim 10^{-5} s$$

2. Next there are terms that scale like a typical collision duration, like $\tau_c^{-1} \sim \frac{\partial \Phi}{\partial \mathbf{q}} \cdot \frac{\partial}{\partial \mathbf{p}}$; that is, these terms have a magnitude which should be commensurate with the duration over which two particles are within a characteristic effective range of the potential, d . If we restrict ourselves to reasonably short-range interaction potentials (van der Waals, or, say, Lennard-Jones interactions), this effective distance is on the scale of angstroms, $d \sim 10^{-10} m$. Very roughly, that would give us:

$$\tau_c \sim d/v \sim 10^{-12} s$$

3. Finally, there are collisional terms like

$$\tau_x^{-1} \sim \int dV \frac{\partial \Phi}{\partial \mathbf{q}} \cdot \frac{\partial}{\partial \mathbf{p}} N \frac{\rho_{s+1}}{\rho_s}$$

The integral has some non-zero contribution over a volume that scales like the characteristic volume of the potential, d^3 , and the ratio $N\rho_{s+1}/\rho_s$ is like a probability of finding an additional particle in the vicinity of the s particles, which should be of the order of the particle number density, $n = N/V \sim 10^{26}m^{-3}$. Combining this and the above gives, very roughly:

$$\tau_x \sim \frac{\tau_c}{nd^3} \sim \frac{1}{nvd^2} \sim 10^{-8}s$$

What does all of this buy us? Well, we see that the second level (and all higher- s levels) of the hierarchy is balance between three competing terms: something like τ_U^{-1} , something like τ_c^{-1} , and something like τ_x^{-1} . So, as long as we are in the *dilute limit*, we see that the terms on the RHS of Eq. C.2 are orders of magnitude smaller than the terms on the LHS, and so we approximate these levels of the hierarchy just by the balance of terms that are like τ_U^{-1} and τ_c^{-1} .

In contrast, the first level is different. It has no terms that are like τ_c^{-1} , and so we have no choice but to keep all of the terms.

C.1.2 Truncating the hierarchy, and continued simplifications

This illustrates a common strategy in kinetic (and other) theories...at first glance the hierarchy does not seem helpful, because it is not *closed* (i.e., to solve for ρ_s one needs to know ρ_{s+1}). However, we can try to come up with a model/theory/approximation that governs the higher-order levels of the hierarchy; the quality of our predictions will then (hopefully) be related to the quality of our closure of the theory. In the present case, explicitly, we have

$$\left[\frac{\partial}{\partial t} - \frac{\partial U}{\partial \mathbf{q}_1} \cdot \frac{\partial}{\partial \mathbf{p}_1} + \frac{\mathbf{p}_1}{m} \cdot \frac{\partial}{\partial \mathbf{q}_1} \right] f_1 = \int dV_2 \frac{\partial \Phi_{12}}{\partial \mathbf{q}_1} \cdot \frac{\partial}{\partial \mathbf{p}_1} f_2, \quad (\text{C.3})$$

$$\left[\frac{\partial}{\partial t} - \frac{\partial U}{\partial \mathbf{q}_1} \cdot \frac{\partial}{\partial \mathbf{p}_1} - \frac{\partial U}{\partial \mathbf{q}_2} \cdot \frac{\partial}{\partial \mathbf{p}_2} + \frac{\mathbf{p}_1}{m} \cdot \frac{\partial}{\partial \mathbf{q}_1} + \frac{\mathbf{p}_2}{m} \cdot \frac{\partial}{\partial \mathbf{q}_2} - \frac{\partial \Phi_{12}}{\partial \mathbf{q}_1} \cdot \left(\frac{\partial}{\partial \mathbf{p}_1} - \frac{\partial}{\partial \mathbf{p}_2} \right) \right] f_2 = 0. \quad (\text{C.4})$$

You may wonder if this closure is sufficient to break the time-reversal symmetry of the underlying equations of motion. It is not! At this stage we have the collisions which will in principle allow us to relax to equilibrium, but everything is still time-reversible. So we proceed with a sequence of physically motivated approximations...

Let's focus on the evolution of the two-body term. In particular, we expect that most of the changes we are interested in are those that are due to the changes wrought by collisions, rather than the slower changes of evolutions under the effect of the external potential. So, Eqs. C.3,C.4 we will ignore the terms related to $\frac{\partial U}{\partial \mathbf{q}}$. Additionally, we see that in Eq. C.4 the collision term depends not on absolute positions but on *relative* positions¹⁴³, so let's switch coordinates to the center of mass, relative position frame (and similarly for momenta):

$$\mathbf{R} = \frac{1}{2}(\mathbf{q}_1 + \mathbf{q}_2), \mathbf{r} = (\mathbf{q}_1 - \mathbf{q}_2), \mathbf{P} = (\mathbf{p}_1 + \mathbf{p}_2), \mathbf{p} = \frac{1}{2}(\mathbf{p}_1 - \mathbf{p}_2).$$

¹⁴³(which makes sense – they're collisions!)

We now have a distribution function $f_2(\mathbf{R}, \mathbf{r}, \mathbf{P}, \mathbf{p}, t)$, where the distribution function depends on the center of mass variables, \mathbf{R}, \mathbf{P} “slowly”, and has a much faster dependence on the the relative coordinates \mathbf{r}, \mathbf{p} , which vary over the small distance d and the time scale τ_c .

Since the relative distributions in f_2 vary so quickly, we assume that in a since f_2 reaches equilibrium and then enters the dynamics of f_1 . That is, we focus in on time intervals that are long compared to τ_c (but perhaps short compared to τ_U), to get the “steady state” behavior of f_2 at small relative distances that are relevant to the collision term. Combining the approximations in the above paragraph, we have

$$\left(\frac{\mathbf{p}}{m} \cdot \frac{\partial}{\partial \mathbf{r}} - \frac{\partial \Phi(\mathbf{r})}{\partial \mathbf{r}} \cdot \frac{\partial}{\partial \mathbf{p}} \right) f_2 \approx 0. \quad (\text{C.5})$$

This is the right form to allow us to start massaging the collision term in the RHS of the equation for f_1 :

$$\begin{aligned} \left(\frac{\partial f_1}{\partial t} \right)_{coll} &= \int dV_2 \frac{\partial \Phi_{12}}{\partial \mathbf{q}_1} \cdot \frac{\partial}{\partial \mathbf{p}_1} f_2 = \int dV_2 \frac{\partial \Phi(\mathbf{r})}{\partial \mathbf{r}} \cdot \left[\frac{\partial}{\partial \mathbf{p}_1} - \frac{\partial}{\partial \mathbf{p}_2} \right] f_2 \\ &= \frac{1}{m} \int_{|\mathbf{r}| \leq d} dV_2 (\mathbf{p}_1 - \mathbf{p}_2) \cdot \frac{\partial f_2}{\partial \mathbf{r}}. \end{aligned} \quad (\text{C.6})$$

In the first equality (where we put in an extra $\frac{\partial}{\partial \mathbf{p}_2}$, we’re just noting that if we integrate by parts that term vanishes (we’ve added a derivative of something we’re integrating over...), and in the next line we’re plugging in the results of Eq. C.5.

C.1.3 Scattering theory and the Boltzmann Equation

This part is not crucial to our conceptual discussion, but it is what allows us to massage the above expression into the Boltzmann equation form. Let’s think more about classical two-particle collisions, which begin with momenta $\mathbf{p}_i = m\mathbf{v}_i$ and end with momenta $\mathbf{p}'_i = m\mathbf{v}'_i$. We proceed to transform into the rest frame of the first particle, so that it is being bombarded with oncoming particles that have velocity $\mathbf{v}_2 - \mathbf{v}_1$, and these oncoming particles are uniformly distributed over the plan normal to that oncoming velocity. We define several relevant quantities in Fig. C.2. Geometrically, we see that the solid angles are $d\sigma = b db d\phi$ and $d\Omega = \sin \theta d\theta d\phi$. The number of particles scattered *into* $d\Omega$ per unit time is related to the flux of particles hitting the plane and the other solid angle, $I d\theta$, typically written $I \frac{d\sigma}{d\Omega} d\Omega = I b db d\phi$, where the *differential cross section* is

$$\left| \frac{d\sigma}{d\Omega} \right| = \frac{b}{\sin \theta} \left| \frac{db}{d\theta} \right| = \frac{1}{2} \left| \frac{d(b^2)}{d \cos \theta} \right|.$$

What we are really saying here is that for a fixed relative incoming velocity there is a particular relationship between the impact parameter, b , and the scattering angle, θ , and this is something you can figure out for any particular classical pair potential Φ .

If we compare these types of scattering expressions to what we had in the “intuitive” version’s

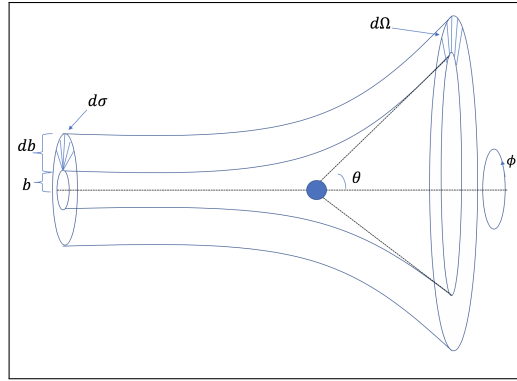


Figure C.2: **Differential cross section for a scattering process** b is the impact parameter, i.e. the distance from the asymptotic trajectory to the central line, which denotes a head-on collision with the particle (shown as a blue disk here); b and the polar angle ϕ together parameterize the plane normal to the incoming particle. The scattering angle θ is the angle by which the incoming particle is deflected. The solid angles $d\sigma$ and $d\Omega$ are illustrated, with relations between them in the text.

expressions Eq. 5.45, we see that when we talked about the rate of scattering *into* some small area of momentum space we can express this in terms of the differential cross section:

$$\omega(\mathbf{p}, \mathbf{p}_2, \mathbf{p}'_1, \mathbf{p}'_2) d^3 p'_1 d^3 p'_2 = |\mathbf{v}_1 - \mathbf{v}_2| \left| \frac{d\sigma}{d\Omega} \right| d\Omega$$

Great. Let's go back to our collision integral:

$$\left(\frac{\partial f_1}{\partial t} \right)_{coll} = \frac{1}{m} \int_{|\mathbf{r}| \leq d} dV_2 (\mathbf{p}_1 - \mathbf{p}_2) \cdot \frac{\partial f_2}{\partial \mathbf{r}}. \quad (C.7)$$

Let's transform to the coordinate system illustrated in Fig. C.3: the direction parallel to the relative velocity is parameterized by x , we have our interaction range of the potential d , and the plane normal to the relative velocity is still parameterized by ϕ and b . Using all of the

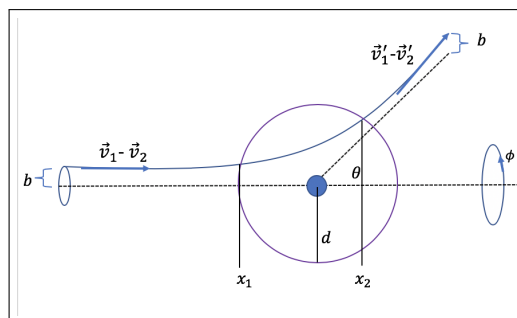


Figure C.3: **Coordinate system for two-particle collision**

above we can write

$$\begin{aligned} \left(\frac{\partial f_1}{\partial t}\right)_{coll} &= \int d^3 p_2 |\mathbf{v}_1 - \mathbf{v}_2| \int d\phi db b \int_{x_1}^{x_2} \frac{\partial f_2}{\partial x} \\ &= \int d^3 p_2 d^3 p'_1 d^3 p'_2 \omega(\mathbf{p}'_1, \mathbf{p}'_2, \mathbf{p}, \mathbf{p}_2) [f_2(x_2) - f_2(x_1)]. \end{aligned} \quad (\text{C.8})$$

Getting to Boltzmann

Working out the classical scattering theory to massage the collision term into the form of Eq. C.8, all that's left is to decide on the same simplifications for f_2 itself. To finish the derivation of the Boltzmann equation and write a closed equation for f_1 , we make a final, big approximation (“*the assumption of molecular chaos*”) that the momenta of the two particles are uncorrelated:

$$f_2(\mathbf{q}, \mathbf{q}, \mathbf{p}, \mathbf{p}_2) = f_1(\mathbf{q}, \mathbf{p}) f_1(\mathbf{q}, \mathbf{p}_2). \quad (\text{C.9})$$

This, perhaps, doesn't look so strong an assumption on its surface, but looking at how f_2 enters the rate of collision expressions we've written down, we see that this amounts to explicitly assuming that the momenta are uncorrelated *before* the collision (and then the outgoing momenta follow from conservation and details of the scattering process). This has, quite sneakily, been a means to smuggle in an arrow of time. Finally, we imagine coarse graining over space (on the scale of d) so that we evaluate $f_2(x_1)$ and $f_2(x_2)$ at the same location, \mathbf{q} . We arrive at Eq. 5.44,

$$\frac{\partial f_1}{\partial t} - \{\mathcal{H}_1, f_1\} = \int d^3 p_2 d^3 p'_1 d^3 p'_2 \omega(\mathbf{p}'_1, \mathbf{p}'_2 | \mathbf{p}, \mathbf{p}_2) [f_1(\mathbf{q}, \mathbf{p}'_1) f_1(\mathbf{q}, \mathbf{p}'_2) - f_1(\mathbf{q}, \mathbf{p}) f_1(\mathbf{q}, \mathbf{p}_2)]. \quad (\text{C.10})$$

Appendix D

Quantum Statistical Mechanics

In Chapter 6 we drew out a logical chain of reasoning that combined ideas from probability theory and kinetic theory, and made one massive assumption about the nature of the microcanonical ensemble, in order to derive the structure of thermodynamics (in the so-called $N \rightarrow \infty$ “thermodynamic limit”). Here we’ll see what happens if the laws governing microscopic systems are quantum mechanical rather than classical; we will see that all of the *logical* steps are the same as in the classical case, but some of the details turn out to be different.

So, in recapitulating the same logic given different microscopic goals, part of the aim of this chapter is simply to solidify our understanding about what statistical mechanics is trying to do. We will also see the traditional resolution to Gibbs’ paradox, finally understanding why there are factors of h and $N!$ floating around otherwise classical expressions. We will *also* see that collections of ideal (non-interacting) quantum particles can be quite a bit richer than their classical counterparts. In that context, we will get our first taste of a phase transition, and we will meet some of the same mathematical technology we’ll use in our study of interacting systems later in the course.

D.1 Classical statistical mechanics: We’ve got some issues

Classical statistical mechanics is an incredibly powerful framework for understanding the physical properties of many systems, but it certainly has limitations. We have already seen the Gibbs paradox and our current, somewhat *ad hoc* solution of throwing in a factor of $N!$ to account for classically indistinguishable particles. We have also seen that in going from $S = -\sum_i p_i \log p_i$ to $S = -\int dx p(x) \log p(x)$ there is a problem of choosing units, and in the last few sections we have introduced a random factor of h to fix the unit problem without understanding why a particular scale of coarse-graining is appropriate (i.e., why the thing we called h was the Planck constant). In this section we also emphasize that there are observable, low-temperature phenomena for which classical mechanics makes incorrect predictions, further necessitating the introduction of quantum mechanical to our description. The following subsections provide a few concrete illustrations.

D.1.1 Dilute diatomic gases

Actually, classical statistical mechanics fails to predict heat capacities both in the gaseous phase *and* in the solid phase... not a great track record. Let's see what happens for a dilute gas of diatomic molecules.

We'll take a simplified view and consider diatomic molecules consisting of two atoms in a bound state, and let's write down an idealized, classical model for such a molecule: two equal masses connected by a spring. So, in addition to the usual translational modes we've been considering, the molecule can also move via (a) rotations in which the molecule rigidly rotates about either of the two axes¹⁴⁴ normal to the symmetry axis, with moment of inertia I , and (b) vibrations in which the molecule oscillates along the axis of symmetry.

We first assume that the gas is sufficiently dilute that the molecules behave independently, so that the total partition function is

$$Z(N) = \frac{Z_1^N}{N!},$$

Where Z_1 is the partition function for a single diatomic molecule. We further assume that three modes of molecular motion are all independent, in which case the single-molecule partition function factorizes into the contribution due to each term:

$$Z_1 = Z_{trans}Z_{vib}Z_{rot}.$$

We already know what the translational partition function looks like, what about the other two terms?

From your classical mechanics course, the Lagrangian for the rotational degrees of freedom is

$$L_{rot} = \frac{1}{2}I \left(\dot{\theta}^2 + \dot{\phi}^2 \sin^2 \theta \right),$$

with conjugate momenta

$$p_\theta = \frac{\partial L_{rot}}{\partial \dot{\theta}} = I\dot{\theta}, \quad p_\phi = \frac{\partial L_{rot}}{\partial \dot{\phi}} = I\dot{\phi} \sin^2 \theta.$$

The hamiltonian for the rotational piece is therefore

$$\mathcal{H}_{rot} = \dot{\theta}p_\theta + \dot{\phi}p_\phi - L = \frac{p_\theta^2}{2I} + \frac{p_\phi^2}{2I \sin^2 \theta}.$$

So, the rotational contribution to the partition function is

$$Z_{rot} = \frac{1}{h^2} \int d\theta d\phi dp_\theta dp_\phi e^{-\beta \mathcal{H}_{rot}} = \frac{8\pi^2 I k_B T}{h^2}. \quad (\text{D.1})$$

What about the vibrational mode? It's just a harmonic oscillator. Denoting the displacement away from the equilibrium position of the molecular "spring" by ζ and the vibrational frequency by ω , the Hamiltonian is

$$\mathcal{H}_{vib} = \frac{p_\zeta^2}{2m} + \frac{1}{2}m\omega^2\zeta^2,$$

¹⁴⁴we neglect the rotation about the axis of symmetry, arguing that it has a low moment of inertia compared to the other two. We'll see that this is hardly the problem.

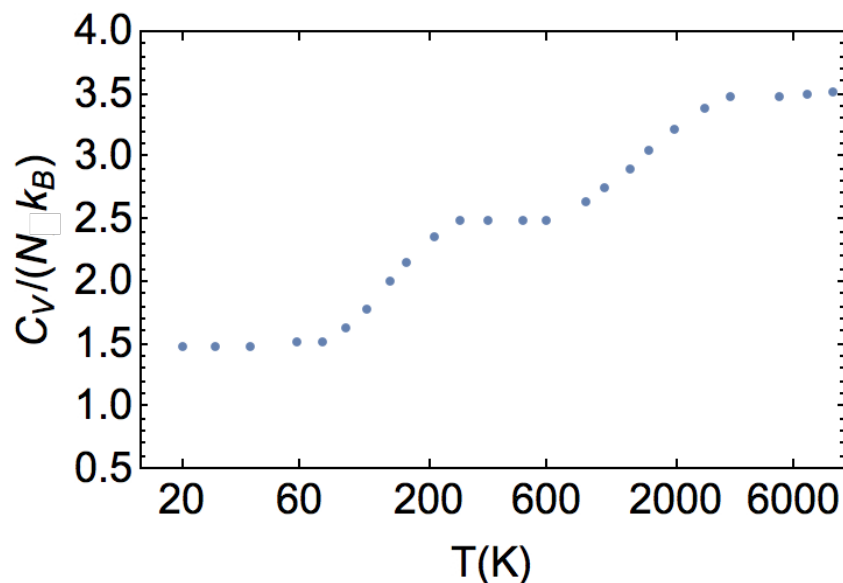


Figure D.1: (Schematic) heat capacity of hydrogen gas vs logarithmically scaled temperature.

from which we find the partition function contribution to be

$$Z_{vib} = \frac{1}{h} \int d\zeta dp_{\zeta} e^{-\beta \mathcal{H}_{vib}} = \frac{2\pi k_B T}{h\omega}. \quad (\text{D.2})$$

Putting together all of these ingredients (or, by your expectations from equipartition of energy among all of the quadratic modes in the combined Hamiltonian), we expect that the heat capacity and constant volume for our diatomic gas is

$$C_V = \frac{7}{2} N k_B, \quad (\text{D.3})$$

an end result which does not depend on the precise value of I or the stiffness of the bond between the atoms. The only problem is that our prediction for the heat capacity is... not borne out in the experimental data. In Fig. D.1 I've schematically¹⁴⁵ plotted the heat capacity of H_2 – the simplest diatomic gas – over a broad range of temperatures. At very high temperatures we do see the heat capacity we expected, but at the lowest temperatures the system seems to behave like a *monatomic* gas, so apparently the diatomic molecules are neither rotating nor vibrating. Even away from the typical “very low temperatures” we expect to see quantum effects in – that is, even at room temperature! – there is a large discrepancy between our prediction and the actual heat capacity, and apparently there are rotations but not vibrations (we'll see how we picked out that particular mode later).

This behavior was arguably the first time that quantum mechanics revealed itself in experiments, and scientists at the end of the 19th century were increasingly unsettled.

¹⁴⁵Actual data available from NIST, if you're interested

D.1.2 Black-body radiation

The classical failure of the calculation of the black-body spectrum (i.e., what light is emitted from a source at a particular temperature) and its quantum resolution is a story that I suspect is familiar from previous courses in quantum mechanics¹⁴⁶. Very briefly, then, we consider our first quantum gas: a gas of photons. In principle we are interested in the emission spectrum of an idealized substance that can absorb photons of any wavelength and reflects none of them. In a real atomic system there would be a (potentially) interesting pattern of absorption and emission lines, but we ignore such details for now and consider our idealized substance which, at zero temperature, would appear black (hence the name).

So, we know a photon is characterized by its wavelength λ or its frequency $\omega = 2\pi c/\lambda = kc$ for speed of light c and wavevector k , with energy $E = \hbar\omega$. An important fact about photons is that they are *not conserved*: there is no reason that the walls of our black-body substance couldn't absorb a photon and then emit two¹⁴⁷. Thus, when we calculate quantities in the canonical ensemble we need to make sure that we sum over possible states with different numbers of photons, since they are allowed states. Equivalently, we can imagine that we work in the grand canonical ensemble, but with chemical potential $\mu_{\text{photon}} = 0$.

To build up the partition function, let's first consider photons with a particular frequency ω ; N such photons would have energy $E = N\hbar\omega$, and summing over the allowed N gives a partial partition function

$$Z_\omega = 1 + e^{-\beta\hbar\omega} + e^{-2\beta\hbar\omega} + e^{-3\beta\hbar\omega} + \dots = \frac{1}{1 - e^{-\beta\hbar\omega}}. \quad (\text{D.4})$$

We assume that the different frequencies are independent, and the total partition function is a product of independent partition functions, so we can write the logarithm as a sum:

$$\log Z = \int_0^\infty d\omega g(\omega) \log Z_\omega,$$

where $g(\omega)$ is the *density of states*: $g(\omega)d\omega$ counts the number of states in the frequency range between ω and $\omega + d\omega$. We can calculate this by, say, assuming periodic boundary conditions in a box of linear size L , which then permits wavevectors $\mathbf{k} = 2\pi\{n_x, n_y, n_z\}$, where $\{n_x, n_y, n_z\}$ are all integers. Planck suggested that the allowed values of energy must be quantized,

$$\mathcal{H}_{EM} = \sum_{\mathbf{k}, \alpha} \hbar ck \left(n_\alpha(\mathbf{k}) + \frac{1}{2} \right), \quad n_\alpha(\mathbf{k}) = 0, 1, 2, \dots,$$

where α refers to the polarization of the photon. One can use this to compute the associated density of states for the photon gas. For now I'll ignore terms relating to the factor of 1/2 above, since we usually only care about (or can detect!) energy differences. Taking into account the fact that photons can come in two polarization states, one eventually gets

$$g(\omega)d\omega = \frac{V\omega^2}{\pi^2 c^3} d\omega.$$

¹⁴⁶It was an act of desperation. For six years I had struggled with the blackbody theory. I knew the problem was fundamental and I knew the answer. I had to find a theoretical explanation at any cost, except for the inviolability of the two laws of thermodynamics." – Max Planck [73]

¹⁴⁷Of course, you *know* that photons aren't conserved, since you demonstrate their non-conservation every time you change the state of a (functional) light switch

Combing these results, we get that

$$\log Z = \int_0^\infty d\omega g(\omega) \log Z_\omega = -\frac{V}{\pi^2 c^3} \int_0^\infty d\omega \omega^2 \log(1 - e^{-\beta\hbar\omega}). \quad (\text{D.5})$$

From this we can get, e.g., the energy stored in the photon gas:

$$E = -\frac{\partial \log Z}{\partial \beta} = \frac{V\hbar}{\pi^2 c^3} \int_0^\infty d\omega \frac{\omega^3}{e^{\beta\hbar\omega} - 1} = \frac{V}{\pi^2 c^3} \frac{(k_B T)^4}{\hbar^3} \int_0^\infty dx \frac{x^3}{e^x - 1}. \quad (\text{D.6})$$

That last integral can be explicitly evaluated (with some work), with the end result

$$\frac{E}{V} = \frac{\pi^2 k_B^4}{15 \hbar^3 c^3} T^4. \quad (\text{D.7})$$

The free energy $F = -k_B T \log Z$ can be calculated; integration by parts (to take care of the log inside the integral) gets us to

$$F = -\frac{V\pi^2}{45\hbar^3 c^3} (k_B T)^4, \quad (\text{D.8})$$

from which we can calculate, e.g., the pressure due to electromagnetic radiation as

$$P = -\left. \frac{\partial F}{\partial V} \right|_T = \frac{E}{3V} = \frac{4\sigma}{3c} T^4,$$

where

$$\sigma = \frac{\pi^2 k_B^4}{60 \hbar^3 c^2} = 5.67 \times 10^{-8} \frac{\text{J}}{\text{sm}^2 \text{K}^4}$$

is the Stefan¹⁴⁸ constant.

I haven't really emphasized the precise way that the classical version of the above calculation fails, as I'm sure you've seen it before. Briefly, classically we should think that the Hamiltonian for the EM field can be written in terms of the normal modes characterized by \mathbf{k} , so that the energy looks like a collection of independent harmonic oscillators corresponding to photons of different wave-number and polarization. Classically, though, there is no limit on the size of \mathbf{k} , leading to the *ultraviolet catastrophe*: we assign $k_B T/2$ of equipartitioned energy to each independent quadratic mode of \mathcal{H} , leading to an infinite amount of energy stored in the high-frequency modes. This is... not physical.

D.2 Microstates, observables, and dynamics

D.2.1 Quantum microstates

As we begin to trace out the logic of statistical mechanics again, let's write down a few definitions¹⁴⁹ to get us more properly started on quantum statistical mechanics. Classically,

¹⁴⁸Josef Stefan, who used his law to (for the first time?) calculate a reasonable temperature for the surface of the sun [74]

¹⁴⁹I will, naturally, assume you already know quantum mechanics, so here we're just dotting some i's

we started with microstates for N particles that were specified by a point in $6N$ -dimensional phase space and governed by Hamiltonian evolution equations. Quantum mechanically, of course, positions and momenta are not independently observables, so this is a poor choice of microstate. Instead, a quantum system is specified by a *unit vector in a Hilbert space*¹⁵⁰, $|\psi\rangle$. Given a set of orthonormal basis vectors $|n\rangle$ we can write the microstate as

$$|\psi\rangle = \sum_n \langle n|\psi\rangle |n\rangle, \quad (\text{D.9})$$

where the $\langle n|\psi\rangle$ are complex numbers, where we keep in mind that $\langle\psi|n\rangle = \langle n|\psi\rangle^*$, the complex conjugate. Probably the first basis we learn about is composed of the spatial coordinates, $\{\mathbf{q}_i\}$, for which

$$\psi(\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_N) \equiv \langle\{\mathbf{q}_i\}|\psi\rangle$$

is the wavefunction.

Finally, of course: as a unit vector, ψ is normalized so that

$$\langle\psi|\psi\rangle = \sum_n \langle\psi|n\rangle \langle n|\psi\rangle = 1, \quad (\text{D.10})$$

D.2.2 Quantum observables

In a previous chapter we introduced classical observable functions of phase space, $A(\{\mathbf{p}, \mathbf{q}\})$. Quantum mechanically, observables get promoted to *operators* by substituting the position and momentum operators for the position and momentum variables in the classical expressions¹⁵¹, $A(\{\hat{p}, \hat{q}\})$. Just as classically we had the Poisson bracket, $\{\mathbf{p}_i, \mathbf{q}_j\} = \delta_{ij}$, here we have the commutation relation $[\hat{p}_a, \hat{q}_b] = \hat{p}_a \hat{q}_b - \hat{q}_b \hat{p}_a = \frac{\hbar}{i} \delta_{ab}$.

Unlike in classical mechanics, in addition to the (classical) probabilistic nature of our ensembles, quantum mechanically our *observables themselves* are matrices that don't have definite values; i.e., they are not uniquely determined for a particular microstate. This additional randomness means the observables are themselves random variables, so we must content ourselves with their expectation values, defined as

$$\langle A \rangle = \langle\psi|A|\psi\rangle = \sum_{m,n} \langle\psi|m\rangle \langle m|A|n\rangle \langle n|\psi\rangle. \quad (\text{D.11})$$

Since we demand real observables, the operators A must be Hermitian¹⁵²: $A^\dagger = A$.

D.2.3 Time evolution of states

Classically we got a lot of mileage out of the hamiltonian evolution of the phase space coordinates. The quantum mechanical state vector has a time evolution given by

$$i\hbar \frac{\partial}{\partial t} |\psi(t)\rangle = H |\psi(t)\rangle. \quad (\text{D.12})$$

¹⁵⁰A generalization of Euclidean space: a Hilbert space H is a vector space equipped with an inner product whose induced distance function makes H a complete metric space

¹⁵¹after, of course, properly symmetrizing products, for instance $pq \rightarrow (pq + qp)/2$

¹⁵²which is why we worried about symmetrizing position and momentum when going from classical to quantum operators.

It is often convenient to work in the basis which diagonalizes the Hamiltonian (i.e., the basis formed by the energy eigenstates), satisfying $H|n\rangle = E_n|n\rangle$ where E_n are the “eigen-energies.” In such a basis, exploiting the orthonormality of the basis $\langle m|n\rangle = \delta_{mn}$ lets us write the time evolution of the state as

$$i\hbar \frac{d}{dt} \langle n|\psi(t)\rangle = E_n \langle n|\psi(t)\rangle \Rightarrow \langle n|\psi(t)\rangle = \exp\left(\frac{-iE_n t}{\hbar}\right) \langle n|\psi(0)\rangle. \quad (\text{D.13})$$

D.3 The density matrix and macroscopic observables

Classically, macrostates are specified by just a few thermodynamic coordinates, and we studied ensembles of large numbers of microstates μ_s , which were equipped with a probability, $p_s \equiv p(\mu_s)$ and which corresponded to a given macrostate. We often don't have precise knowledge of the microstate (i.e., the system is not a *pure state*); more generally we expect it to be a *mixed state*, existing as an incoherent mixture of being in a variety of quantum states¹⁵³.

We similarly start out with a mixed state, an incoherent mixture of states $|\psi_\alpha\rangle$ with probabilities p_α . The ensemble average of the expectation value of an observable in such a mixed state is

$$\begin{aligned} \langle \bar{A} \rangle &= \sum_{\alpha} p_{\alpha} \langle \psi_{\alpha} | A | \psi_{\alpha} \rangle = \sum_{\alpha, m, n} p_{\alpha} \langle \psi_{\alpha} | m \rangle \langle n | \psi_{\alpha} \rangle \langle m | A | n \rangle \\ &= \sum_{m, n} \langle m | A | n \rangle \langle n | \rho | m \rangle = \text{Tr}\{\rho A\}, \end{aligned} \quad (\text{D.14})$$

where we have introduced the *density matrix* ρ , which in a given basis is

$$\rho = \sum_{\alpha} p_{\alpha} |\psi_{\alpha}\rangle \langle \psi_{\alpha}|, \quad (\text{D.15})$$

and where the trace of an operator is the sum over the diagonal elements, $\text{Tr}\{M\} = \sum_{\alpha} \langle \Phi_{\alpha} | M | \Phi_{\alpha} \rangle$, which is independent of which basis $\{\Phi_{\alpha}\}$ you use.

D.3.1 Basic properties of the density matrix

1. *Sufficient*: All measurements in quantum mechanics involve expectation values of operators. Thus, the density matrix contains sufficient information for anything we might want to do.
2. *Positive definite*: The eigenvalues of ρ are all positive, since for any state

$$\langle \phi | \rho | \phi \rangle = \sum_n p_n \langle \phi | \psi_n \rangle \langle \psi_n | \phi \rangle = \sum_n p_n |\langle \phi | \psi_n \rangle|^2 \geq 0.$$

¹⁵³Not in a superposition, of states, by the way. A quick example: take a spin in the up-down basis. The superposition state $2^{-1/2}(|\uparrow\rangle + |\downarrow\rangle)$ is a diagonally polarized state. An unpolarized spin is a mixture of half up and half down, described by a density matrix $\frac{1}{2}(|\uparrow\rangle\langle\uparrow| + |\downarrow\rangle\langle\downarrow|)$. The latter is what we mean.

3. *Normalized*: Since the ψ_α are themselves normalized, we have

$$\text{Tr}\{\rho\} = \sum_n p_n \langle \psi_n | \psi_n \rangle = \sum_n p_n = 1.$$

4. *Hermitian*: By inspection, the density matrix is Hermitian, with $\rho^\dagger = \rho$.

Time evolution of the density matrix

Classically we had Liouville's theorem for the evolution of the density, $\frac{d\rho}{dt} = \frac{\partial\rho}{\partial t} - \{\mathcal{H}, \rho\}$; what do we have here? Again working in the energy eigenbasis, we can write

$$\begin{aligned} i\hbar\partial_t\rho &= \sum_n p_n i\hbar\partial_t (|\psi_n\rangle\langle\psi_n|) \\ &= \sum_n p_n [H|\psi_n\rangle\langle\psi_n| - |\psi_n\rangle\langle\psi_n|H] \\ &= H\rho - \rho H = [H, \rho]. \end{aligned} \tag{D.16}$$

D.4 Quantum ensembles

With this structure, we can follow the same logic that we did in the classical case: we define equilibrium by having none of the averages of the observables vary with time, which can be satisfied if we choose an equilibrium density matrix so that $\partial_t\rho = 0$. Just as when we were dealing with Poisson brackets, we accomplish this by having the density matrix be a function of the Hamiltonian itself, along with any conserved quantities A , $\rho(H, A_1, \dots)$, that satisfy $[H, A_i] = 0$.

D.4.1 Quantum microcanonical ensemble

We define the microcanonical ensemble, specified by (E, \mathbf{x}, N) , but enforcing a fixed value for the ensemble average energy. We choose our density matrix

$$\rho(E) = \frac{\delta(H - E)}{\Omega(E)},$$

where in the energy eigen-basis we can write this as

$$\langle m | \rho | n \rangle = \sum_\alpha p_\alpha \langle m | \psi_\alpha \rangle \langle \psi_\alpha | n \rangle = \begin{cases} \Omega^{-1} & \text{if } E_n = E, \text{ and } m = n \\ 0 & \text{otherwise} \end{cases} \tag{D.17}$$

The first of those two conditions we recognize as the equivalent of the assumption of equal a priori probabilities. The second, quantum mechanical condition is the *assumption of random a priori phases*, in which we don't get contributions from off-diagonal terms (even if they have degenerate and correct energies) because we assume the mixed state is in an *incoherent* superposition of the basis states. Finally, note that from the normalization condition on the density matrix, $\Omega(E)$ is again just counting the number of (eigen)states of H with the correct energy E .

D.4.2 Quantum canonical ensemble

You know what's coming: Now we're fixing temperature $T = \beta^{-1}$ by putting our quantum system in contact with a reservoir. Considering the above two assumptions for the combined system, we find that the density matrix for the system of interest is

$$\rho(\beta) = \frac{e^{-\beta H}}{Z(\beta)}, \quad (\text{D.18})$$

where the normalization condition on the density matrix leads to the quantum canonical partition function for N particles,

$$Z_N(\beta) = \text{Tr}\{e^{-\beta H}\} = \sum_n e^{-\beta E_n}. \quad (\text{D.19})$$

As one would expect from the above formulas, the expectation value of a physical observable is given by

$$\langle A \rangle = \text{Tr}\{\hat{\rho}\hat{A}\} = \frac{1}{Z(\beta)} \text{Tr}\{\hat{A}e^{-\beta\hat{H}}\} = \frac{\text{Tr}\{\hat{A}e^{-\beta\hat{H}}\}}{\text{Tr}\{e^{-\beta\hat{H}}\}} \quad (\text{D.20})$$

D.4.3 Quantum grand canonical ensemble

For completeness – and because we'll see it again soon – in the grand canonical ensemble we no longer fix the number of particles¹⁵⁴. The density matrix is

$$\rho(\beta, \mu) = \frac{e^{-\beta H + \beta\mu N}}{\mathcal{Q}}, \quad (\text{D.21})$$

where the grand canonical partition function is

$$\mathcal{Q}(\beta, \mu) = \text{Tr}\{e^{-\beta H + \beta\mu N}\} = \sum_{N=0}^{\infty} e^{\beta\mu N} Z_N(\beta). \quad (\text{D.22})$$

D.4.4 Example: Free particle in a box

Suppose we care about the quantum canonical ensemble version of a single particle in a box of volume V . Working in the coordinate basis, the Hamiltonian is

$$H = \frac{\mathbf{p}^2}{2m} = -\frac{\hbar^2}{2m}\nabla^2, \quad (\text{D.23})$$

which has energy eigenstates $|\mathbf{k}\rangle$ specified by

$$\langle \mathbf{r} | \mathbf{k} \rangle = \frac{e^{i\mathbf{k}\cdot\mathbf{r}}}{\sqrt{V}}, \quad E_{\mathbf{k}} = \frac{\hbar^2 k^2}{2m}. \quad (\text{D.24})$$

¹⁵⁴Microstates with an indefinite number of particles span a *Fock space*, which is a set of Hilbert spaces associated with zero or more quantum particles

What are the allowed \mathbf{k} ? Assuming for simplicity periodic boundary conditions for a cube of side length L , we can have $\mathbf{k} = \frac{2\pi}{L}(l_x, l_y, l_z)$, where the l_α are integers. So, the space of microstates is enormously larger than in the classical case: rather than 6 degrees of freedom per particle, one can have countably infinite numbers of states per particle. In the limit $L \rightarrow \infty$ the partition function becomes

$$\begin{aligned} Z &= \sum_{\mathbf{k}} e^{-\frac{\beta \hbar^2 k^2}{2m}} = V \int \frac{d^3 \mathbf{k}}{(2\pi)^3} e^{-\frac{\beta \hbar^2 k^2}{2m}} \\ &= \frac{V}{(2\pi)^3} \left(\frac{\sqrt{2\pi m k_B T}}{\hbar} \right)^3 = \frac{V}{\lambda^3}, \end{aligned} \quad (\text{D.25})$$

for $\lambda = h/\sqrt{2\pi m k_B T}$, which indeed coincides with our classical calculation when we use the right (adjusted) phase space measure.

What about the elements of the density matrix itself? We can compute

$$\begin{aligned} \langle \mathbf{r}' | \rho | \mathbf{r} \rangle &= \sum_{\mathbf{k}} \langle \mathbf{r}' | \mathbf{k} \rangle \frac{e^{-\beta E_{\mathbf{k}}}}{Z} \langle \mathbf{k} | \mathbf{r} \rangle \\ &= \frac{\lambda^3}{V} \int V \frac{d^3 \mathbf{k}}{(2\pi)^3} \frac{e^{-i\mathbf{k} \cdot (\mathbf{r} - \mathbf{r}')}}{V} e^{-\frac{\beta \hbar^2 k^2}{2m}} \\ &= \frac{1}{V} \exp \left(-\frac{\pi (\mathbf{r} - \mathbf{r}')^2}{\lambda^2} \right). \end{aligned} \quad (\text{D.26})$$

What does this mean? The diagonal elements are all $\langle \mathbf{r} | \rho | \mathbf{r} \rangle = V^{-1}$, the common expectation that the probability for finding the particle is uniform throughout the box. The off-diagonal terms are a quantum-mechanical effect, measuring the “spontaneous transition” between coordinates \mathbf{r} and \mathbf{r}' , giving a measure of the “intensity” of the wave packet some distance from the center of the packet. Said another way, the spatial extent of the packet is a measure of the uncertainty involved in locating the particle position. This quantum mechanical effect vanishes in the $\beta \rightarrow 0$ limit, as the density matrix elements approach delta functions.

Finally, we can compute the expectation value of the Hamiltonian itself, $\langle H \rangle = \text{Tr} \{ \hat{H} \hat{\rho} \}$.

We’ve already calculated the partition function $Z = \text{Tr} \{ e^{-\beta \hat{H}} \}$, so the result is the last line of:

$$\begin{aligned} \langle H \rangle &= \text{Tr} \{ \hat{H} \hat{\rho} \} = \frac{\text{Tr} \{ \hat{H} e^{-\beta \hat{H}} \}}{\text{Tr} \{ e^{-\beta \hat{H}} \}} \\ &= -\frac{\partial}{\partial \beta} \log \text{Tr} \{ e^{-\beta \hat{H}} \} \\ &= \frac{3}{2} k_B T. \end{aligned} \quad (\text{D.27})$$

This result should be...expected.

D.4.5 Example: An electron in a magnetic field

Suppose we care about the quantum canonical ensemble version of a single electron in a magnetic field. The electron has spin $\hbar\hat{\sigma}/2$ and a magnetic moment $\mu_B = \frac{e\hbar}{2mc}$ (nothing to do with the chemical potential; it's just standard notation to use μ here, too), where $\hat{\sigma}$ is the Pauli spin operator.

When we apply a magnetic field, \mathbf{B} , the electron can have either spin up or spin down. If we take the applied field to be along \hat{z} , the configurational part of the Hamiltonian is

$$\hat{H} = -\mu_B \hat{\sigma} \cdot \mathbf{B}. \quad (\text{D.28})$$

Life is easier when we work in the basis in which the Hamiltonian is diagonal, i.e.,

$$\hat{H} = -\mu_B B \hat{\sigma}_z, \quad (\text{D.29})$$

where

$$\hat{\sigma}_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \hat{\sigma}_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \hat{\sigma}_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (\text{D.30})$$

From this it is straightforward to calculate the density matrix:

$$\begin{aligned} \hat{\rho} &= \frac{e^{-\beta\hat{H}}}{\text{Tr}\{e^{-\beta\hat{H}}\}} \\ &= \frac{1}{e^{\beta\mu_B B} + e^{-\beta\mu_B B}} \begin{pmatrix} e^{\beta\mu_B B} & 0 \\ 0 & e^{-\beta\mu_B B} \end{pmatrix}, \end{aligned} \quad (\text{D.31})$$

from which we can calculate, e.g., the expectation value for σ_z :

$$\langle \sigma_z \rangle = \text{Tr}\{\hat{\rho}\hat{\sigma}_z\} = \frac{e^{\beta\mu_B B} - e^{-\beta\mu_B B}}{e^{\beta\mu_B B} + e^{-\beta\mu_B B}} = \tanh(\beta\mu_B B), \quad (\text{D.32})$$

an expression I'm quite sure you've seen in other classes.

D.5 The classical limit of a quantum partition function

With some of the above definitions in hand, but before we move on to quantum statistical mechanics, it is interesting to note that we already have enough to be able to figure out at least part of the corrected phase space measure that we introduced in the last chapter. In this short section let's leave aside the question of where $1/N!$ comes from and see why Planck's constant appears even in classical statistical mechanics. Recall that for a single classical particle the canonical partition function can be written as an integral over phase space:

$$Z_1 = \frac{1}{h^3} \int d^3\mathbf{q} d^3\mathbf{p} e^{-\beta H(p,q)}.$$

At the time the $1/h$ was required to get the units right (i.e., so that Z is dimensionless), but where there was a particular value of h to use: Planck's constant $h = 2\pi\hbar \approx 6.6 \times 10^{-34} Js$.

Why is there this quantum-mechanical number in our classical formulas? We don't need to wave our hands, let's derive it.

To keep things simple, let's consider a single particle in one dimension, so that its Hamiltonian is

$$\hat{H} = \frac{\hat{p}^2}{2m} + V(\hat{q}),$$

where \hat{p} is the momentum operator and \hat{q} is the position operator.

Writing the eigenstates in the energy basis, in which state $|n\rangle$ has energy E_n , the quantum partition function is

$$Z = \sum_n e^{-\beta E_n} = \sum_n \langle n | e^{-\beta \hat{H}} | n \rangle, \quad (\text{D.33})$$

where by the exponential of the operator we mean

$$e^{-\beta \hat{H}} = \sum_{n=0}^{\infty} (-1)^n \frac{(\beta \hat{H})^n}{n!}.$$

As usual we are free to insert the identity operator, constructed by summing over any complete basis of states. We'll do this with both the position eigenvectors and the momentum eigenvectors:

$$\mathbf{1} = \int dq |q\rangle \langle q|, \quad \mathbf{1} = \int dp |p\rangle \langle p|.$$

We first add two copies of the position-eigenstate identity to the partition function, on either side of the $e^{-\beta \hat{H}}$:

$$\begin{aligned} Z &= \sum_n \langle n | \int dq |q\rangle \langle q| e^{-\beta \hat{H}} \int dq' |q'\rangle \langle q'| n \rangle \\ &= \int dq dq' \langle q | e^{-\beta \hat{H}} | q' \rangle \sum_n \langle q' | n \rangle \langle n | q \rangle \\ &= \int dq \langle q | e^{-\beta \hat{H}} | q \rangle, \end{aligned} \quad (\text{D.34})$$

where in the last line we replaced $\sum_n |n\rangle \langle n|$ with the identity operator, used $\langle q' | q \rangle = \delta(q' - q)$, and integrated over q' . So far the result of this manipulation is to replace a sum over energy eigenstates with an integral over position eigenstates. This is really just another expression of the fact that we can write the partition function in a basis-independent way:

$$Z = \text{Tr} \left\{ e^{-\beta \hat{H}} \right\}, \quad (\text{D.35})$$

but is perhaps a helpful warm-up.

Let's try to take the classical limit of this partition function, by which we mean that we'll neglect terms that are of order \hbar . We exploit this by trying to factorize $e^{-\beta \hat{H}}$ into a position and a momentum piece, using the Baker-Campbell-Hausdorff formula¹⁵⁵

$$e^{\hat{A}} e^{\hat{B}} = \exp \left(\hat{A} + \hat{B} + \frac{1}{2} [\hat{A}, \hat{B}] + \dots \right). \quad (\text{D.36})$$

¹⁵⁵In general, $e^{\hat{A}} e^{\hat{B}} \neq \exp(\hat{A} + \hat{B})$, and the BCH formula lets you work out the correction in terms of a series of nested commutators of \hat{A} and \hat{B} .

What does that mean for us? Assuming that $V(\hat{q})$ is a reasonably well-behaved potential and recalling that $[\hat{q}, \hat{p}] = i\hbar$, Taking the classical limit means we neglect any corrections to the naive factorization, writing

$$e^{-\beta\hat{H}} = e^{-\beta\hat{p}^2/(2m)} e^{-\beta V(\hat{q})} e^{\mathcal{O}(\hbar)} \approx e^{-\beta\hat{p}^2/(2m)} e^{-\beta V(\hat{q})}. \quad (\text{D.37})$$

From here, we proceed by turning exponentiated operators into ordinary functions (again, recalling what the exponentials of operators mean). We first take care of the position operators:

$$\begin{aligned} Z &= \int dq \langle q | e^{-\beta\hat{p}^2/(2m)} e^{-\beta V(\hat{q})} | q \rangle \\ &= \int dq e^{-\beta V(q)} \langle q | e^{-\beta\hat{p}^2/(2m)} | q \rangle. \end{aligned}$$

To proceed, we insert two complete sets of momentum states, use $\langle q | p \rangle = \frac{1}{\sqrt{2\pi\hbar}} e^{ipq/\hbar}$, and notice that objects like $\langle p | e^{-f(\hat{p})} | p' \rangle$ will give us delta functions to relate p' to p :

$$\begin{aligned} Z &= \int dq dp dp' e^{-\beta V(q)} \langle q | p \rangle \langle p | e^{-\beta\hat{p}^2/(2m)} | p' \rangle \langle p' | q \rangle \\ &= \int dq dp dp' e^{-\beta V(q)} \langle p | e^{-\beta\hat{p}^2/(2m)} | p' \rangle \frac{1}{2\pi\hbar} e^{i(p-p')q/\hbar} \\ &= \frac{1}{h} \int dq dp e^{-\beta H(p,q)}. \end{aligned} \quad (\text{D.38})$$

Thus, we see the natural consequence of the underlying quantum mechanical description even when we take the classical limit and do our best to ignore \hbar .

D.6 Quantum indistinguishability

The previous section motivated why Planck's constant appears in the classical partition function. In this and the next section we will work out a position-space representation of the canonical density matrix for indistinguishable quantum particles, and show precisely how the classical limit of this expression lets us see where the corrected Gibbs phase-space measure actually comes from. The first step, which we tackle in this section, is to write down eigenstates of an N -particle Hamiltonian that have either Bosonic or Fermionic symmetry.

D.6.1 Two identical particles

Suppose we were to write down a simple two-particle Hamiltonian, for two particles of equal mass and with an interaction potential that depended only on the relative separation:

$$\mathcal{H}(1, 2) = \frac{\mathbf{p}_1^2}{2m} + \frac{\mathbf{p}_2^2}{2m} + V(|\mathbf{r}_1 - \mathbf{r}_2|).$$

Clearly this Hamiltonian is symmetric under the exchange of particle label, $\mathcal{H}(1, 2) = \mathcal{H}(2, 1)$. Classically, our labeling of particles is meaningful, but for identical quantum mechanical particles these labels are arbitrary and convey no physical meaning.

For instance, the probability of finding two identical particles at positions \mathbf{r}_1 and \mathbf{r}_2 is given by $|\psi(\mathbf{r}_1, \mathbf{r}_2)|^2 = |\psi(\mathbf{r}_2, \mathbf{r}_1)|^2$. As long as the wavefunction is single-valued this leads to two distinct possibilities¹⁵⁶:

$$\psi(\mathbf{r}_1, \mathbf{r}_2) = \begin{cases} \psi(\mathbf{r}_2, \mathbf{r}_1) & \text{for bosons} \\ -\psi(\mathbf{r}_2, \mathbf{r}_1) & \text{for fermions} \end{cases} \quad (\text{D.39})$$

D.6.2 N identical particles

Starting with a wavefunction for N particles, $\psi(\mathbf{r}_1, \dots, \mathbf{r}_N)$, we generalize the above by introducing a *permutation operator* P , of which there are $N!$ possible permutation operators we might consider for our set of particles. We'll adopt the notation

$$P\psi(\mathbf{r}_1, \dots, \mathbf{r}_N) = \begin{cases} \psi(\mathbf{r}_1, \dots, \mathbf{r}_N) & \text{for bosons} \\ (-1)^P \psi(\mathbf{r}_1, \dots, \mathbf{r}_N) & \text{for fermions} \end{cases} \quad (\text{D.40})$$

to represent the two classes of Hilbert spaces we might find ourselves in. Here we take $(-1)^P$ to represent the *parity*¹⁵⁷ of the permutation under question: if P can be represented by an even number of pairwise particle exchanges then $(-1)^P = 1$, and if it can be represented by an odd number of pairwise exchanges the $(-1)^P = -1$.

Note that the Hamiltonian for the particles, \mathcal{H} , must be symmetric: $P\mathcal{H} = \mathcal{H}$, but the Hamiltonian can admit eigenstates of different symmetries under the action of the permutation operator. The same Hamiltonian will thus allow eigenstates of either total symmetry or total anti-symmetry: the statistics one wants to study must be specified independently of the Hamiltonian, and so one studies only a subspace (either the fermionic subspace, in which eigenstates are anti-symmetric, or the bosonic subspace, in which eigenstates are symmetric) of the total Hilbert space. Let's see a convenient way of representing these subspaces.

D.6.3 Product states for distinguishable, non-interacting particles

We consider an N -particle Hamiltonian which is just a collection of single-particle Hamiltonians¹⁵⁸ for free particles in a box of volume V :

$$\mathcal{H} = \sum_{\alpha=1}^N \frac{\mathbf{p}_\alpha^2}{2m} = \sum_{\alpha=1}^N -\frac{\hbar^2}{2m} \nabla_\alpha^2. \quad (\text{D.41})$$

Each of the single-particle Hamiltonians can be diagonalized by writing it in the energy basis, $|\mathbf{k}_\alpha\rangle$ with energy $\hbar^2 \mathbf{k}_\alpha^2 / (2m)$, and we will build our N -particle wavefunction out of these one-particle eigen-pieces.

¹⁵⁶Because the square of the exchange operator must be the identity matrix. For single-valued functions this restriction means that there can only be a complex phase shift under the operation of a single application of the exchange operator, so the square being the identity means the phase shift can only be 0 (bosons) or π (fermions). You may have heard, though, of *anyons*! These have multi-valued wavefunctions, and in two dimensions one can find other allowed statistics without this constraint on the value of the phase shift.

¹⁵⁷In Pathria's notation, what I will eventually write as η^P is denoted $\delta_P = (\pm 1)^{[P]}$

¹⁵⁸“Hamiltonia”?

We define a *product state* as

$$|\mathbf{k}_1, \dots, \mathbf{k}_N\rangle_{\times} \equiv |\mathbf{k}_1\rangle |\mathbf{k}_2\rangle \cdots |\mathbf{k}_N\rangle, \quad (\text{D.42})$$

where in the coordinate representation the product state is

$$\langle \mathbf{r}_1, \dots, \mathbf{r}_N | \mathbf{k}_1, \dots, \mathbf{k}_N \rangle_{\times} = \frac{1}{V^{N/2}} \exp \left(-i \sum_{\alpha} \mathbf{k}_{\alpha} \cdot \mathbf{r}_{\alpha} \right), \quad (\text{D.43})$$

and of course

$$\mathcal{H} |\mathbf{k}_1, \dots, \mathbf{k}_N\rangle_{\times} = \left(\sum_{\alpha} \frac{\hbar^2 \mathbf{k}_{\alpha}^2}{2m} \right) |\mathbf{k}_1, \dots, \mathbf{k}_N\rangle_{\times}. \quad (\text{D.44})$$

These product states are very convenient to work with, but they are too general! That is, they are appropriate for *distinguishable* particles, but for indistinguishable particles they do not have the correct symmetry for either bosons *or* fermions.

To show how we can compactly write either fermionic or bosonic states, let's start by defining a symbol

$$\eta = \begin{cases} 1 & \text{for bosons} \\ -1 & \text{for fermions} \end{cases}, \quad (\text{D.45})$$

and we will write things like $|\{\mathbf{k}\}\rangle_{+}$ and $|\{\mathbf{k}\}\rangle_{-}$ for bosonic and fermionic states, respectively.

Fermionic eigenstates

We build the set of possible fermionic states by summing over all possible permutations of the product state, but including the appropriate anti-symmetrizing factor:

$$|\mathbf{k}_1, \dots, \mathbf{k}_N\rangle_{-} = \frac{1}{\sqrt{N_{-}}} \sum_P (-1)^P P |\mathbf{k}_1, \dots, \mathbf{k}_N\rangle_{\times}, \quad (\text{D.46})$$

where $N_{-} = N!$ is a factor that ensures proper normalization of our fermionic eigenstate. Because of the anti-symmetrization, if there are any value \mathbf{k}_{α} appears more than once the whole eigenstate vanishes, and so anti-symmetrization is only possible if there are N distinct \mathbf{k}_{α} . This is why we know there are as many distinct terms in the sum as there are particles, and thus why $N_{-} = N!$. For example, a three-particle anti-symmetrized state is¹⁵⁹

$$|123\rangle_{-} = \frac{|123\rangle_{\times} + |231\rangle_{\times} + |312\rangle_{\times} - |213\rangle_{\times} - |321\rangle_{\times} - |132\rangle_{\times}}{\sqrt{6}} \quad (\text{D.47})$$

Bosonic eigenstates

Formally, we write the bosonic states similarly, as sum of possible permutations of the product state with a factor of $(+1)^P = 1$ accounting for the parity of each permutation:

$$|\mathbf{k}_1, \dots, \mathbf{k}_N\rangle_{+} = \frac{1}{\sqrt{N_{+}}} \sum_P (+1)^P P |\mathbf{k}_1, \dots, \mathbf{k}_N\rangle_{\times}. \quad (\text{D.48})$$

¹⁵⁹for convenience, let $\mathbf{k}_1 = 1$, etc.

Bosons, though, are allowed to have states in which the same \mathbf{k}_α appears multiple times, so computing the normalization factor is slightly more complicated. To see this, consider the bosonic state $|121\rangle_+$ (i.e., a state in which there are two identical \mathbf{k} and one unlike \mathbf{k} . We see that

$$\begin{aligned} |121\rangle_+ &= \frac{1}{\sqrt{N_+}} (|112\rangle_\times + |121\rangle_\times + |211\rangle_\times + |112\rangle_\times + |121\rangle_\times + |211\rangle_\times) \\ &= \frac{2}{\sqrt{N_+}} (|112\rangle_\times + |121\rangle_\times + |211\rangle_\times), \end{aligned} \quad (\text{D.49})$$

so for proper normalization of $|121\rangle_+$ we need $N_+ = 12$. The combinatorial generalization is that if each k is repeated n_k times in the N -particle bosonic state, then $N_+ = N! \prod_k n_k!$. We can see this by requiring

$$\begin{aligned} 1 &= {}_+ \langle \{\mathbf{k}\} | \{\mathbf{k}\} \rangle_+ = \frac{1}{N_+} \sum_{P, P'} \times \langle P' \{\mathbf{k}\} | P \{\mathbf{k}\} \rangle_\times \\ &= \frac{N!}{N_+} \sum_P \times \langle \{\mathbf{k}\} | P \{\mathbf{k}\} \rangle_\times, \end{aligned} \quad (\text{D.50})$$

but the $\langle \{\mathbf{k}\} | P \{\mathbf{k}\} \rangle$ vanish by orthogonality unless the permuted set of wavevectors matches the original, which happens $n_k!$ times for each repeated k . Thus,

$$1 = \frac{N! \prod_k n_k!}{N_+} \Rightarrow N_+ = N! \prod_k n_k!. \quad (\text{D.51})$$

Compact notation

Actually, though, since for a fermionic state n_k can only be zero or one (again, the antisymmetrization gets rid of any states with multiply repeated k), we can combine the bosonic and fermionic notation above into:

$$|\mathbf{k}_1, \dots, \mathbf{k}_N\rangle_\eta = \frac{1}{\sqrt{N_\eta}} \sum_P \eta^P P |\mathbf{k}_1, \dots, \mathbf{k}_N\rangle_\times. \quad (\text{D.52})$$

For both bosons and fermions $N_\eta = N! \prod_{\mathbf{k}} n_{\mathbf{k}}!$, and note that the states end up being uniquely specified by the set of occupation numbers, $n_{\mathbf{k}}$, with the constraint

$$\sum_{\mathbf{k}} n_{\mathbf{k}} = N. \quad (\text{D.53})$$

D.7 The canonical ensemble density matrix for non-interacting identical particles

In the main text we argued that the $h \rightarrow 0$ limit of quantum statistical mechanics would imply the existence of h 's in the phase space measure *even* for strictly classical systems. Let's use the product-state technology to show more formally not just the presence of Planck's

constant but also the factor of $1/N!$ that should be there when dealing with indistinguishable particles.

From our Bosonic and Fermionic eigenstates, we will begin by writing down the canonical density matrix for non-interacting sets of identical particles; we will eventually extract the canonical partition function by using the fact that ρ is normalized. Just as in classical statistical mechanics, once we have Z we can extract all of thermodynamics. But just like the classical $p(\mu_s)$ contains more information than *just* thermodynamics, ρ gives us additional information (for instance, about multi-particle correlations). This section will be a bit of a technical calculation; at the end we will understand where the classical partition function comes from, and also how the quantum statistics of non-interacting identical particles are approximately equivalent to introducing either attractive or repulsive classical interactions which are felt over distances comparable to the thermal de Broglie wavelength.

Let's make our lives messy by writing the elements of the density matrix in the position basis: $\langle \{\mathbf{r}'\} | \rho_N | \{\mathbf{r}\} \rangle_\eta$, where we know ρ_N will be diagonal in the energy basis:

$$\langle \mathbf{r}'_1, \dots, \mathbf{r}'_N | \rho_N | \mathbf{r}_1, \dots, \mathbf{r}_N \rangle_\eta = \sum_{\{\mathbf{k}\}_{\text{restricted}}} \left[\frac{1}{N_\eta} \sum_{P, P'} \eta^P \eta^{P'} \langle \{\mathbf{r}'\} | P' \{\mathbf{k}\} \rangle_\times \rho(\{\mathbf{k}\}) \langle P \{\mathbf{k}\} | \{\mathbf{r}\} \rangle_\times \right]. \quad (\text{D.54})$$

The density matrix is (c.f. Eq. D.18)

$$\rho_N(\{\mathbf{k}\}) = \frac{\exp\left(-\beta \sum_{\alpha=1}^N \frac{\hbar^2 k_\alpha^2}{2m}\right)}{Z_N}, \quad (\text{D.55})$$

and the “restricted” sum above makes sure that every unique indistinguishable particle state appears exactly once (correctly accounting for either bosonic or fermionic statistics). That restriction is, in fact, a bit cumbersome, so it is more convenient to sum over *all* $\{\mathbf{k}\}$ and then correct for any over-counting. Since the states are specified by the occupation numbers, and since (again) for fermions the $\eta^P \eta^{P'}$ cancels all contributions from terms with $n_k > 1$, we can do this via

$$\sum_{\{\mathbf{k}\}_{\text{restricted}}} = \sum_{\{\mathbf{k}\}} \frac{\prod_{\mathbf{k}} n_{\mathbf{k}}!}{N!} \quad (\text{D.56})$$

Making this change (combined with the factor of $1/N_\eta = 1/(N! \prod_{\mathbf{k}} n_{\mathbf{k}}!)$), gives

$$\langle \{\mathbf{r}'\} | \rho_N | \{\mathbf{r}\} \rangle_\eta = \sum_{\{\mathbf{k}\}} \frac{1}{(N!)^2} \sum_{P, P'} \frac{\eta^P \eta^{P'}}{Z_N} e^{-\beta \sum_{\alpha=1}^N \frac{\hbar^2 k_\alpha^2}{2m}} \langle \{\mathbf{r}'\} | P' \{\mathbf{k}\} \rangle \langle P \{\mathbf{k}\} | \{\mathbf{r}\} \rangle. \quad (\text{D.57})$$

We reorder the sums and replace the sum over $\{\mathbf{k}\}$ with an integral to get

$$\langle \{\mathbf{r}'\} | \rho_N | \{\mathbf{r}\} \rangle_\eta = \frac{1}{(N!)^2 Z_N} \sum_{P, P'} \eta^P \eta^{P'} \int \left(\prod_{\alpha=1}^N \frac{V d^3 \mathbf{k}_\alpha}{(2\pi)^3} e^{-\beta \frac{\hbar^2 k_\alpha^2}{2m}} \right) \frac{e^{-i \sum_{\alpha} (\mathbf{k}_{P\alpha} \cdot \mathbf{r}_\alpha - \mathbf{k}_{P'\alpha} \cdot \mathbf{r}'_\alpha)}}{V^N}. \quad (\text{D.58})$$

Perhaps you feel that we have made things worse rather than better; fortunately, we are undeterred. Let's introduce a new label $\gamma = P\alpha$, $\alpha = P^{-1}\gamma$ to keep track of permutations.

We'll make use of the fact that for functions / operators / variables f and g we can sum over indices $\sum_{\alpha} f(P\alpha)g(\alpha) = \sum_{\gamma} f(\gamma)g(P^{-1}\gamma)$; this allows us to focus on a particular wavevector:

$$\langle \{\mathbf{r}'\} | \rho_N | \{\mathbf{r}\} \rangle_{\eta} = \frac{1}{(N!)^2 Z_N} \sum_{P, P'} \eta^P \eta^{P'} \prod_{\alpha=1}^N \int \frac{d^3 \mathbf{k}_{\alpha}}{(2\pi)^3} e^{-i\mathbf{k}_{\alpha} \cdot (\mathbf{r}_{P^{-1}\alpha} - \mathbf{r}'_{(P')^{-1}\alpha}) - \beta \frac{\hbar^2 k_{\alpha}^2}{2m}}. \quad (\text{D.59})$$

The Gaussian integrals in this expression give

$$\int \frac{d^3 \mathbf{k}_{\alpha}}{(2\pi)^3} e^{-i\mathbf{k}_{\alpha} \cdot (\mathbf{r}_{P^{-1}\alpha} - \mathbf{r}'_{(P')^{-1}\alpha}) - \beta \frac{\hbar^2 k_{\alpha}^2}{2m}} = \frac{1}{\lambda^3} \exp\left(-\frac{\pi}{\lambda^2} \left(\mathbf{r}_{P^{-1}\alpha} - \mathbf{r}'_{(P')^{-1}\alpha}\right)^2\right). \quad (\text{D.60})$$

Using this result and setting $\mu = P^{-1}\alpha$ we get

$$\langle \{\mathbf{r}'\} | \rho_N | \{\mathbf{r}\} \rangle_{\eta} = \frac{1}{Z_N \lambda^{3N} (N!)^2} \sum_{P, P'} \eta^P \eta^{P'} \exp\left(-\frac{\pi}{\lambda^2} \sum_{\mu=1}^N \left(\mathbf{r}_{\mu} - \mathbf{r}'_{(P')^{-1}P\mu}\right)^2\right). \quad (\text{D.61})$$

The last step is to do one of the two sums over the permutations. We define $Q = (P')^{-1}P$, and since $\eta^P = \eta^{P^{-1}}$ we can write $\eta^P \eta^{P'} = \eta^{(P')^{-1}P} = \eta^Q$. With this, and summing over one set of $N!$ permutations, we arrive at

$$\langle \{\mathbf{r}'\} | \rho_N | \{\mathbf{r}\} \rangle_{\eta} = \frac{1}{Z_N \lambda^{3N} N!} \sum_Q \eta^Q \exp\left(-\frac{\pi}{\lambda^2} \sum_{\mu=1}^N \left(\mathbf{r}_{\mu} - \mathbf{r}'_{Q\mu}\right)^2\right). \quad (\text{D.62})$$

We can *finally* get the canonical partition function by enforcing the normalization of the density matrix:

$$\text{Tr}\{\rho\} = 1 \quad \Rightarrow \quad \int \prod_{\alpha=1}^N d^3 \mathbf{r}_{\alpha} \langle \{\mathbf{r}\} | \rho_N | \{\mathbf{r}\} \rangle_{\eta} = 1 \quad (\text{D.63})$$

$$\Rightarrow \quad Z_N = \frac{1}{\lambda^{3N} N!} \int \prod_{\alpha=1}^N d^3 \mathbf{r}_{\alpha} \sum_Q \eta^Q \exp\left(-\frac{\pi}{\lambda^2} \sum_{\mu=1}^N \left(\mathbf{r}_{\mu} - \mathbf{r}_{Q\mu}\right)^2\right). \quad (\text{D.64})$$

We see that the quantum mechanical partition function has within it a sum over the $N!$ permutations of identical particles, the classical result,

$$Z_N = \frac{1}{N!} \left(\frac{V}{\lambda^3}\right)^N,$$

corresponds to the term in which there are no exchanges, that is, where Q is the identity. We see that there are lots of other terms involving products of terms like

$$\exp\left(-\frac{\pi}{\lambda^2} (\mathbf{r}_1 - \mathbf{r}_2)^2\right),$$

but as $T \rightarrow \infty$, $\lambda \rightarrow 0$ and these quantum corrections vanish.

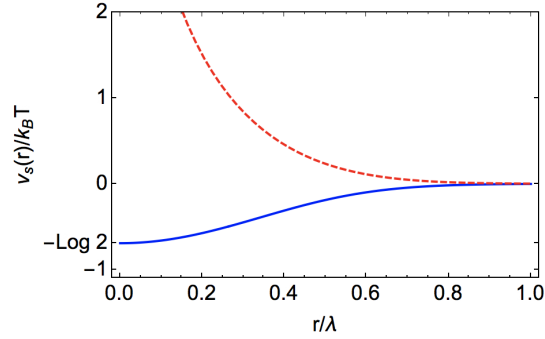


Figure D.2: **The effective statistical interparticle potential is attractive (bosons) or repulsive (fermions)** The solid line corresponds to v_s in Eq. D.67 with $\eta = +1$, as in the case of bosons, and the dashed line corresponds to the case of $\eta = -1$, as for particles with fermionic statistics.

D.7.1 Statistical interparticle potential

Before we try to evaluate Eq. D.64, let's pause to think about perturbations away from the infinite temperature limit. Clearly the lowest order correction from the classical canonical partition function involves permutations which just exchange two particles. So, let's consider the simplest possible (non-trivial) case, where $N = 2$ and \sum_Q is a sum over the identity and either a symmetric or antisymmetric exchange of particles. The partition function is

$$Z_2 = \frac{1}{2!} \left(\frac{V}{\lambda^3} \right)^2 \left[1 + \frac{\eta}{2^{3/2}} \frac{\lambda^3}{V} \right], \quad (\text{D.65})$$

and the diagonal elements of the density matrix are

$$\langle \mathbf{r}_1, \mathbf{r}_2 | \rho_2 | \mathbf{r}_1, \mathbf{r}_2 \rangle \approx \frac{1}{V^2} \left[1 + \eta \exp \left(-2\pi \frac{(\mathbf{r}_1 - \mathbf{r}_2)^2}{\lambda^2} \right) \right]. \quad (\text{D.66})$$

This tells us that when the interparticle separation is comparable to λ the probability density for the two *non-interacting* identical particles is different from the classical result of $1/V^2$, on account of the statistics of identical particle permutations. For $\eta = 1$ there is a *greater* likelihood of finding the particles close together, and for $\eta = -1$ there is a reduction in the probability (all the way down to zero for $|\mathbf{r}_1 - \mathbf{r}_2| \rightarrow 0$).

We can express these different statistical correlations by pretending that we have classical particles interacting according to a *statistical interparticle potential* $v_s(r)$. We'll see these kinds of statistical potentials again later in the course; the idea is that the Boltzmann weight associated with a particular separation should be exactly equal to change in the pair correlation away from the infinite temperature limit above:

$$\exp(-\beta v_s(r)) = 1 + \eta \exp \left(-2\pi \frac{r^2}{\lambda^2} \right). \quad (\text{D.67})$$

This effective potential is shown in Fig. D.2, again, it is a classical potential that mimics the effect of quantum correlations at high temperatures.

D.8 The grand canonical ensemble for non-interacting identical particles

In the Sec. (D.7) we work with these combinations of product states in the position basis to show the relationship between quantum and classical partition functions, and we demonstrate the traditional resolution to Gibbs' paradox for identical particles¹⁶⁰. However, explicitly calculating the kinds of sums over all possible permutations of particle indices that appear in Eq. D.64 for either Bosonic or Fermionic states is... daunting, to say the least. We can first make our life a little bit easier by working with the canonical partition function in the basis for which the Hamiltonian is diagonal. Here

$$Z_N = \text{Tr}\{e^{-\beta H_N}\} = \sum_{\{\mathbf{k}\}_{\text{restricted}}} \eta \langle \{\mathbf{k}\} | e^{-\beta \sum_{\alpha=1}^N E_{\mathbf{k}_\alpha}} | \{\mathbf{k}\} \rangle_\eta. \quad (\text{D.68})$$

We've switched from the particular single-particle Hamiltonian with only kinetic energy in the last section to arbitrary single-particle Hamiltonians that have some set of energy levels characterized by energies $E_{\mathbf{k}_\alpha}$ (so, these could be free particles in a box, or quantum harmonic oscillators, etc.). We still have this restriction on the $\{\mathbf{k}\}$, which we work around as follows.

We recall that the allowed states can be specified by the occupation numbers, $n_{\mathbf{k}}$, for each \mathbf{k} :

$$\begin{aligned} Z_N &= \sum_{\{\mathbf{k}\}_{\text{restricted}}} e^{-\beta \sum_{\alpha=1}^N E_{\mathbf{k}_\alpha}} \\ &= \sum_{\{n_{\mathbf{k}}\}_{\text{restricted}}} \exp\left(-\beta \sum_{\mathbf{k}} E_{\mathbf{k}_\alpha} n_{\mathbf{k}}\right). \end{aligned} \quad (\text{D.69})$$

Furthermore, we have gone from sums over symmetry-restricted sets of \mathbf{k} to restricted sums over the occupation numbers:

$$\sum_{\mathbf{k}} n_{\mathbf{k}} = N, \quad \text{and } n_{\mathbf{k}} = \begin{cases} 0 \text{ or } 1 & \text{fermions} \\ 0, 1, 2, \dots & \text{bosons} \end{cases} \quad (\text{D.70})$$

Performing this restricted sum over occupations numbers is *still* difficult, so we move to the grand canonical ensemble:

$$\mathcal{Q}_\eta = \sum_{N=0}^{\infty} z^N Z_N = \sum_{N=0}^{\infty} e^{\beta \mu N} \sum_{\{n_{\mathbf{k}}\}_{\text{restricted}}} \exp\left(-\beta \sum_{\mathbf{k}} E_{\mathbf{k}_\alpha} n_{\mathbf{k}}\right) \quad (\text{D.71})$$

$$= \sum_{N=0}^{\infty} \left[\sum_{\{n_{\mathbf{k}}\}_{\text{restricted}}} \prod_{\mathbf{k}} \exp[-\beta (E_{\mathbf{k}} - \mu) n_{\mathbf{k}}] \right], \quad (\text{D.72})$$

where the subscript η reminds us that the restriction on the sum implicitly depends on the quantum statistics in question. *At last* we see the utility of moving to the grand canonical

¹⁶⁰i.e., we recover the factors of h and $N!$ we introduced in an ad-hoc way in Sec. 6.1.5

ensemble: the double summation above – first over a restricted set of occupation numbers at fixed N (where the sum of the occupation numbers equals N), and then over all N – is equivalent to simply *summing over all values of the different occupation numbers independently for each \mathbf{k}* .

We now have

$$\mathcal{Q}_\eta = \sum_{\{n_{\mathbf{k}}\}_\eta} \prod_{\mathbf{k}} \exp[-\beta(E_{\mathbf{k}} - \mu)n_{\mathbf{k}}], \quad (\text{D.73})$$

where $\{n_{\mathbf{k}}\}_\eta$ reminds us that the sum over occupation numbers is either $\sum_{n_{\mathbf{k}}=0}^1$ for fermions or $\sum_{n_{\mathbf{k}}=0}^{\infty}$ for bosons. We can evaluate the sums over the occupation numbers independently for each \mathbf{k} above; explicitly, we can write

$$\begin{aligned} \mathcal{Q}_\eta &= \sum_{n_{\mathbf{k}_1}, n_{\mathbf{k}_2}, \dots} [(ze^{-\beta E_{\mathbf{k}_0}})^{n_0} (ze^{-\beta E_{\mathbf{k}_1}})^{n_1} \dots] \\ &= \left[\sum_{n_0} (ze^{-\beta E_{\mathbf{k}_0}})^{n_0} \right] \left[\sum_{n_1} (ze^{-\beta E_{\mathbf{k}_1}})^{n_1} \right] \dots \end{aligned} \quad (\text{D.74})$$

for fermions this just gives us two terms per \mathbf{k} , and for bosons we get a simple geometric series *as long as that series converges*:

$$\mathcal{Q}_- = \prod_{\mathbf{k}} [1 + \exp(\beta\mu - \beta E_{\mathbf{k}})], \quad (\text{D.75})$$

$$\mathcal{Q}_+ = \prod_{\mathbf{k}} [1 - \exp(\beta\mu - \beta E_{\mathbf{k}})]^{-1}, \quad \text{with } E_{\mathbf{k}} - \mu > 0 \forall \mathbf{k}. \quad (\text{D.76})$$

Thermodynamically we usually want to take derivatives of the log of the above expressions, so we combine them compactly as

$$\log \mathcal{Q}_\eta = -\eta \sum_{\mathbf{k}} \log [1 - \eta \exp(\beta\mu - \beta E_{\mathbf{k}})]. \quad (\text{D.77})$$

From this we can calculate the usual suspects. For instance, we typically want to know how many particles we actually have for a given value of the chemical potential. Recall that classically we know that the unconditional probability of finding N particles in the system is

$$p(N) = \frac{e^{\beta\mu N} Z_N}{\mathcal{Q}},$$

and here we have products of independent single-particle states. So, we very similarly can write down the probability of having a particular set of occupation numbers:

$$p_\eta(\{n_{\mathbf{k}}\}) = \frac{1}{\mathcal{Q}_\eta} \prod_{\mathbf{k}} \exp[-\beta(E_{\mathbf{k}} - \mu)n_{\mathbf{k}}]. \quad (\text{D.78})$$

From this we can pull down the average occupation number of a particular state with energy $E_{\mathbf{k}}$, as

$$\langle n_{\mathbf{k}} \rangle_\eta = -\frac{\partial \log \mathcal{Q}_\eta}{\partial (\beta E_{\mathbf{k}})} = \frac{1}{\exp(\beta E_{\mathbf{k}} - \beta\mu) - \eta}. \quad (\text{D.79})$$

From this the average number of particles at fixed μ is

$$N_\eta = \sum_{\mathbf{k}} \langle n_{\mathbf{k}} \rangle_\eta = \sum_{\mathbf{k}} \frac{1}{z^{-1} e^{\beta E_{\mathbf{k}}} - \eta} \quad (\text{D.80})$$

and the average energy is

$$E_\eta = \sum_{\mathbf{k}} E_{\mathbf{k}} \langle n_{\mathbf{k}} \rangle_\eta = \sum_{\mathbf{k}} \frac{E_{\mathbf{k}}}{z^{-1} e^{\beta E_{\mathbf{k}}} - \eta} \quad (\text{D.81})$$

D.9 Ideal quantum gases

D.9.1 Initial comments

In the rest of the chapter we will specialize the above results to the case of ideal quantum gases, looking at some of the thermodynamic properties of both ideal Fermi and Bose gases. Before we dive into the details, it is worth saying a few words about what we will get out of this exploration. Classically, the ideal gas is a prime example we keep coming back to, largely because it is extremely analytically tractable. It does provide a starting point for understanding real gases, but it's a pretty poor starting point for just about any other system (solids, dense liquids, etc).

Quantum ideal gases turn out to be much richer in their phenomenology and much more closely connected to the physics of real systems. Non-interacting systems of bosons are surprisingly accurate descriptors of photons¹⁶¹, phonons¹⁶², and also (of course) actual dilute gases of bosons. Additionally, ideal Bose gases are the first system in this class which exhibit a phase transition¹⁶³ – certainly I find it interesting that a collection of “non-interacting” particles can exhibit different collective forms of organization.

As it happens, thinking about ideal collections of fermions is *also* a good starting point for understanding a wide range of physical systems! This is perhaps surprising – electrons are charged, and the interactions between electrons in an atom, or in a whole material, always make a large contribution to the energy. Nevertheless, a gas of non-interacting fermions is a powerful description of atoms, metals, insulators, neutron stars, etc. Not a *free* gas of fermions: the trick is that interacting fermions often act like collections of non-interacting fermions sitting in a modified external potential¹⁶⁴.

To begin thinking about these ideal quantum gases, let's first just specialize the expressions we derived at the end of Sec. D.8 to the case of non-relativistic idealized quantum gases. That is, we'll make the specific choice for the energies $E_{\mathbf{k}} = \hbar^2 k^2 / (2m)$, where the energy levels have a *degeneracy* g (associated with the spin s of the particles, e.g., $g = 2s + 1$), and where we assume the particles are in a large enough box that we can safely replace the sum over \mathbf{k} by an integral: $\sum_{\mathbf{k}} \rightarrow V \int d^3\mathbf{k} / (2\pi)^3$. The evaluation of the grand canonical partition

¹⁶¹Which might have been anticipated from the discussion of blackbody radiation

¹⁶²And, hence, gives insight into the heat capacity of most crystalline solids

¹⁶³To the Bose-Einstein condensate

¹⁶⁴See Ref. [10] for a slightly expanded discussion of this point.

function gives us the following results for the pressure, number density, and energy density:

$$\beta P_\eta = \frac{\log \mathcal{Q}_\eta}{V} = -\eta g \int \frac{d^3 \mathbf{k}}{(2\pi)^3} \log \left[1 - \eta z e^{-\beta \frac{\hbar^2 k^2}{2m}} \right], \quad (\text{D.82})$$

$$n_\eta = \frac{N_\eta}{V} = g \int \frac{d^3 \mathbf{k}}{(2\pi)^3} \left(z^{-1} e^{\beta \frac{\hbar^2 k^2}{2m}} - \eta \right)^{-1}, \quad (\text{D.83})$$

$$\varepsilon_\eta = \frac{E_\eta}{V} = g \int \frac{d^3 \mathbf{k}}{(2\pi)^3} \frac{\hbar^2 k^2}{2m} \left(z^{-1} e^{\beta \frac{\hbar^2 k^2}{2m}} - \eta \right)^{-1}. \quad (\text{D.84})$$

D.9.2 Mathematical simplifications

One of my goals is to show how we can continue to talk about both bosons and fermions in the same language, and to do so we'll introduce some mathematical functions (which you may or may not have already encountered).

First, we define the **polylogarithm of order m** by the series

$$\text{Li}_m(z) = \sum_{\alpha=1}^{\infty} \frac{z^\alpha}{\alpha^m}; \quad (\text{D.85})$$

this expression is fine for any complex value of z with $|z| \leq 1$, and for other z the polylogarithm is defined by analytic continuation. Note that $\text{Li}_m(1)$ is the Riemann zeta function, and that $\text{Li}_1(z) = -\log(1-z)$. The name of the function stems from how different polylogs can be recursively related to each other, e.g.,

$$\text{Li}_{m+1}(z) = \int_0^z \frac{\text{Li}_m(t)}{t} dt,$$

or

$$\frac{d}{dz} \text{Li}_m(z) = \frac{1}{z} \text{Li}_{m-1}(z). \quad (\text{D.86})$$

Next, we define the following families of functions that are closely related to what are typically called Bose-Einstein¹⁶⁵ and Fermi-Dirac integrals:

$$f_m^\eta(z) \equiv \frac{1}{(m-1)!} \int_0^\infty dx \frac{x^{m-1}}{z^{-1} e^x - \eta}, \quad (\text{D.87})$$

where we will often want m to take non-integer values, so when we write “ $(m-1)!$ ” we really mean the gamma function¹⁶⁶. We defined the $\text{Li}_m(z)$ because the polylog is closely related

¹⁶⁵Note that $f_m^+(1) = \zeta(m)$, the Riemann zeta function

¹⁶⁶ $\Gamma(m) = (m-1)!$, with, e.g., $(1/2)! = \sqrt{\pi}/2$, etc.

to these physical integrals. Thinking of z as a small parameter:

$$\begin{aligned}
 f_m^\eta(z) &= \frac{1}{(m-1)!} \int_0^\infty dx \frac{x^{m-1}}{z^{-1}e^x - \eta} \\
 &= \frac{1}{(m-1)!} \int_0^\infty dx x^{m-1} \sum_{\alpha=1}^\infty \eta^{\alpha+1} z^\alpha e^{-\alpha x} \\
 &= \sum_{\alpha=1}^\infty \left(\eta^{\alpha+1} z^\alpha \frac{1}{(m-1)!} \int_0^\infty dx x^{m-1} e^{-\alpha x} \right) \\
 &= \sum_{\alpha=1}^\infty \eta^{\alpha+1} \frac{z^\alpha}{\alpha^m} \\
 &= \eta \text{Li}_m(\eta z)
 \end{aligned} \tag{D.88}$$

Back to physics! Let's look at our thermodynamic expressions for ideal gases of fermions or bosons, and make the obvious change of variables to $x = \beta \hbar^2 k^2 / (2m)$, i.e., $k = \frac{2\sqrt{\pi}}{\lambda} x^{1/2}$. Let's explicitly evaluate our expression for, say, pressure¹⁶⁷:

$$\begin{aligned}
 \beta P_\eta &= -\eta g \int \frac{d^3\mathbf{k}}{(2\pi)^3} \log \left[1 - \eta z e^{-\beta \frac{\hbar^2 k^2}{2m}} \right] \\
 &= -\frac{2\eta g}{\lambda^3 \sqrt{\pi}} \int_0^\infty dx x^{1/2} \log(1 - \eta z e^{-x}) \\
 &= \frac{4g}{3\lambda^3 \sqrt{\pi}} \int_0^\infty dx \frac{x^{3/2}}{z^{-1}e^x - \eta} \\
 &= \frac{g}{\lambda^3} f_{5/2}^\eta(z).
 \end{aligned} \tag{D.89}$$

Similarly, we can now compactly write the three thermodynamic expressions above as

$$\begin{aligned}
 \beta P_\eta &= \frac{g}{\lambda^3} f_{5/2}^\eta(z) \\
 n_\eta &= \frac{g}{\lambda^3} f_{3/2}^\eta(z) \\
 \varepsilon_\eta &= \frac{3}{2} P_\eta.
 \end{aligned} \tag{D.90}$$

Beautiful. These equations are a complete description of the thermodynamics of ideal Fermi and Bose gases, neatly encoded in the properties of unusual integral expressions involving the fugacity¹⁶⁸.

Of course, these equations are also somewhat different than what we might actually want. Usually we want *equations of state* – equations that compactly relate all of the thermodynamic

¹⁶⁷Using an integration by parts to get to line 3

¹⁶⁸Or, you might say that in picking a name for a complicated integral we've just parameterized our ignorance, without gaining any understanding yet

coordinates we use to characterize our system. Here we have an *implicit* relationships: we get an equation for P as a function of z , and also $n = N/V$ as a function of z . We would like, therefore, to invert the middle equation and know z in terms of the density; to do so let's work to better understand the behavior of the $f_m^\eta(z)$.

D.9.3 High-temperature and low-density limit of ideal quantum gases

The simplest limit to consider – in which we can continue to treat Fermions and Bosons simultaneously – is the high- T and small- n limit of our gases, for which z is small. For small z we make use of the series expansion to write

$$f_m^\eta(z) = \eta \text{Li}_m(\eta z) = \sum_{\alpha=1}^{\infty} \eta^{\alpha+1} \frac{z^\alpha}{\alpha^m} = z + \eta \frac{z^2}{2^m} + \frac{z^3}{3^m} + \dots, \quad (\text{D.91})$$

which makes it clear that for $z \ll 1$ we also have $f_m^\eta(z) \ll 1$. Hence, n_η and P_η are all small, too, and we are reassured that our calculation here is going to be nicely self-consistent.

So, we take our equation for n in terms of z (Eq. D.90) and find instead a relationship for z in terms of n :

$$z = d - \eta \frac{z^2}{2^{3/2}} - \frac{z^3}{3^{3/2}} - \dots, \quad (\text{D.92})$$

where we have defined the *degeneracy factor*, $d = n_\eta \lambda^3 / g$, which is another way of characterizing the regime in which quantum effects become important (i.e., when $n_\eta \lambda^3 > g$, quantum mechanical effects become crucial to keep track of). From here, we can compute z as a power series to any order in n (plus corrections of higher order) by recursively substituting lower order solutions. Let me show you what I mean:

Explicitly, to lowest order $z \approx z_1 = d$, where I'm using subscripts to temporarily note the order of the solution we have found. That was straightforward.

To improve this to next order, we substitute z_1 into the right hand side of the power series above, and keep all terms of order d^2 , getting

$$z \approx z_2 = d - \frac{\eta}{2^{3/2}} d^2.$$

To get the next order term, we substitute this improved approximation into the right hand side of the power series (keeping all terms up to third order), giving

$$\begin{aligned} z \approx z_3 &= d - \frac{\eta}{2^{3/2}} (z_2)^2 - \frac{1}{3^{3/2}} (z_2)^3 \\ &= d - \frac{\eta}{2^{3/2}} \left(d - \frac{\eta}{2^{3/2}} d^2 \right)^2 - \frac{1}{3^{3/2}} \left(d - \frac{\eta}{2^{3/2}} d^2 \right)^3 \\ &= d - \frac{\eta}{2^{3/2}} d^2 + \left(\frac{1}{4} - \frac{1}{3^{3/2}} \right) d^3 - \dots, \end{aligned}$$

and so on. The point is *not* to work out the numerical values of these prefactors in this power series¹⁶⁹, but rather that *we can systematically and self-consistently* rearrange $n_\eta(z)$ into $z(n_\eta)$,

¹⁶⁹Unless you actually want to calculate certain precise quantities, of course.

at least in the limit we're considering. Having done that, we can then substitute this power series back into Eq. D.90 and get a systematic expansion describing the high-temperature, low-density ideal quantum gas:

$$P_\eta = n_\eta k_B T \left(1 - \frac{\eta}{2^{5/2}} \left(\frac{n_\eta \lambda^3}{g} \right) + \left[\frac{1}{8} - \frac{2}{3^{5/2}} \right] \left(\frac{n_\eta \lambda^3}{g} \right)^2 + \dots \right) \quad (\text{D.93})$$

We'll see in the next chapter that this is our first look at a *virial expansion* of the equation of state.

D.10 Ideal Bose gases

At higher temperature we were able to work out a power series representation for the equation of state for a Bose gas; as the temperature is reduced and $d = n_+ \lambda^3 / g$ grows that approach is no longer useful and we must work directly with the $f_m^\eta(z)$ functions. Recall that the average occupation number of a particular energy eigenstate, Eq. D.79, is

$$\langle n_{\mathbf{k}} \rangle_+ = \frac{1}{\exp[\beta(E_{\mathbf{k}} - \mu)] - 1},$$

which clearly cannot be a negative number. This means that $\mu < E_{\mathbf{k}}$ for any choice of k ; given our investigation of $E_{\mathbf{k}} = \hbar^2 k^2 / (2m)$ here that means that $\mu < 0$ and hence $0 \leq z \leq 1$.

Thus, since the $f_m^+(z)$ are monotonically increasing functions for $0 \leq z \leq 1$, we see that the density of excited states for the ideal Bose gas is bounded. From Eq. D.90:

$$n_+ = \frac{g}{\lambda^3} f_{3/2}^+(z) \leq \frac{g}{\lambda^3} f_{3/2}^+(1), \quad (\text{D.94})$$

where

$$f_{3/2}^+(1) = 1 + \frac{1}{2^{3/2}} + \frac{1}{3^{3/2}} + \dots \approx 2.612.$$

Since there is a bound on the number of excited states, what happens if we take a fixed number of particles in a fixed volume and start cooling them down? At high temperatures, the bound above is not relevant and the density of excited states is the same as the number density. Writing out the factors of λ above, though, we can see that there is a critical temperature at which the bound becomes relevant:

$$\begin{aligned} \frac{n_+ \lambda^3}{g} &= \frac{n_+}{g} \left(\frac{h}{\sqrt{2\pi m k_B T_c}} \right)^3 = f_{3/2}^+(1) \\ \Rightarrow T_c(n) &= \frac{h^2}{2\pi m k_B} \left(\frac{n}{g f_{3/2}^+(1)} \right)^{2/3}. \end{aligned} \quad (\text{D.95})$$

Below this temperature the fugacity is stuck at $z = 1$; the limiting density of excited states, $n^* = g f_{3/2}^+(1) / \lambda^3$, is less than the total number density, and the rest of the particles are forced to occupy the $\mathbf{k} = 0$ zero-energy ground state. This is *Bose-Einstein condensation*: having a macroscopically large number of particles accumulating in just one single-particle state. The schematic growth of the number of particles in the ground state at low temperature is shown in Fig. D.3.

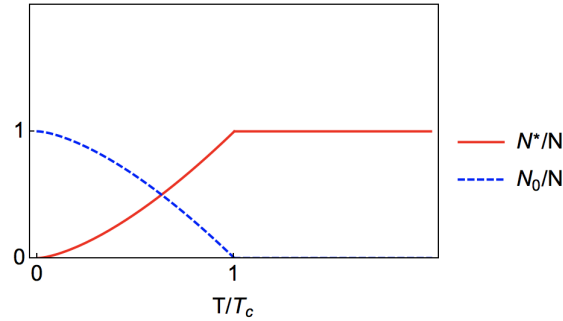


Figure D.3: **Fraction of the normal phase and the condensed phase in an ideal Bose gas.** Since the density of excited states $n^* = gf_{3/2}^+(1)/\lambda^3 \propto T^{3/2}$, we can schematically draw the figure as shown.

D.10.1 Pressure

We now turn to the pressure of the low-temperature phase of this system. For $T < T_c$ we have $z = 1$, and our thermodynamic expression given earlier gives us

$$\beta P_+ = \frac{g}{\lambda^3} f_{5/2}^+(1) \approx 1.341 \frac{g}{\lambda^3}, \quad (\text{D.96})$$

a pressure which is *independent of n* and proportional to $T^{5/2}$. We can use our expression for the critical temperature to note, by the way, that *at T_c* we have

$$P(T_c)V = \frac{f_{5/2}^+(1)}{f_{3/2}^+(1)} (Nk_B T_c) \approx 0.5134 N K_B T_c, \quad (\text{D.97})$$

and we see that right at the transition the pressure of an ideal Bose gas is about half of what would be expected from a classical gas.

D.10.2 Specific heat

Let's look a little closer at the transition from high- to low-temperature Bose gases by studying the heat capacity. Looking again at Eq. D.90 to get an expression for the energy, we see the heat capacity (at constant volume and particle number) is

$$C_{V,N} = \left. \frac{\partial E}{\partial T} \right|_{V,N} = \frac{15Vgk_B}{4\lambda^3} f_{5/2}^+(z) + \frac{3Vgk_B T}{2\lambda^3} \frac{df_{5/2}^+(z)}{dz} \frac{dz}{dT}. \quad (\text{D.98})$$

Note that the first term contributes for the entire range of T , but the second term does not: z only appreciably varies above the critical temperature, so the second term only contributes for $T > T_c$. What are the limits here?

Low temperatures Below T_c we set $z = 1$, ignore the second term, and have

$$C_{V,N} = \frac{15Vgk_B}{4\lambda^3} f_{5/2}^+(1) = Nk_B \frac{15}{4} \frac{f_{5/2}^+(1)}{f_{3/2}^+(1)} \left(\frac{T}{T_c} \right)^{3/2}.$$

So, below T_c the specific heat per particle simply scales as $T^{3/2}$ and it approaches T_c with a slope

$$\frac{d}{dT} \left[\frac{C_{V,N}}{Nk_B} \right]_{T=T_c^-} = \frac{45}{8T_c} \frac{f_{5/2}^+(1)}{f_{3/2}^+(1)} \approx \frac{2.889}{T_c}.$$

High temperatures Above T_c we need to account for the second term in the heat capacity. The derivative with respect to z is easy enough: recalling Eq. D.86, we have

$$\frac{d}{dz} f_m^+(z) = \frac{1}{z} f_{m-1}^+(z),$$

so our heat capacity is

$$C_{V,N} = \frac{3Vgk_B T}{2\lambda^3} \left(\frac{5}{2T} f_{5/2}^+(z) + \frac{f_{3/2}^+(z)}{z} \frac{dz}{dT} \right). \quad (\text{D.99})$$

All that remains is to find how the fugacity changes with temperature. We do this by invoking our condition of fixed particle number in Eq. D.90:

$$\begin{aligned} \left. \frac{dN}{dT} \right|_V = 0 &\Rightarrow 0 = \frac{gV}{\lambda^3} \left(\frac{3}{2T} f_{3/2}^+(z) + \frac{f_{1/2}^+(z)}{z} \frac{dz}{dT} \right) \\ &\Rightarrow \frac{dz}{dT} = -\frac{3z}{2T} \frac{f_{3/2}^+(z)}{f_{1/2}^+(z)}. \end{aligned} \quad (\text{D.100})$$

Substituting this in, we get

$$C_{V,N} = \frac{3Vgk_B}{2\lambda^3} \left[\frac{5f_{5/2}^+(z)}{2} - \frac{3}{2} \frac{(f_{3/2}^+(z))^2}{f_{1/2}^+(z)} \right]. \quad (\text{D.101})$$

How does this behave as T (and, hence, z) varies? At very high temperatures we can use our technique of recursively expanding and solving for z to see that

$$C_V/(Nk_B) = \frac{3}{2} \left(1 + \frac{n\lambda^3}{2^{7/2}} + \dots \right).$$

That is: as the temperature approaches infinity, the specific heat approaches the classical limit *from above*. What about at temperatures close to the transition? We first note that as $z \rightarrow 1$ the function $f_{3/2}^+(z)$ approaches some number, but the function $f_{1/2}^+(z)$ *diverges*. Thus, as $T \rightarrow T_c^+$ the second term in the heat capacity actually vanishes, and we find

$$\left[\frac{C_{V,N}}{Nk_B} \right]_{T \rightarrow T_c^+} = \frac{15}{4} \frac{f_{5/2}^+(1)}{f_{3/2}^+(1)} \approx 1.926.$$

That is: the specific heat is *continuous* across the transition. Above T_c how does the specific heat approach this value? To answer that we just need to evaluate

$$\frac{d}{dT} \left[\frac{C_{V,N}}{Nk_B} \right] = \frac{3gV}{2N} \frac{d}{dT} \left[\frac{5}{2\lambda^3} f_{5/2}^+(z) - \frac{3}{2\lambda^3} \frac{\left(f_{3/2}^+(z) \right)^2}{f_{1/2}^+(z)} \right].$$

This is a bit fussy, but we have the technology: we already know how to relate derivatives of the f_m^+ functions to lower-order ones, and we have already evaluated dz/dT . Churning through these manipulations gives

$$\left[\frac{C_{V,N}}{Nk_B} \right]_{T \rightarrow T_c^+} = \frac{1}{T_c} \left(\frac{45}{8} \frac{f_{5/2}^+(1)}{f_{3/2}^+(1)} - \frac{27}{8} \frac{\left(f_{3/2}^+(1) \right)^2 f_{-1/2}^+(1)}{\left(f_{1/2}^+(1) \right)^3} \right) \approx \frac{-0.777}{T_c}.$$

Note that even though the specific heat itself is continuous across the transition, its derivative discontinuously jumps (from positive to negative!) as the temperature is scanned from below to above T_c . The behavior – from low temperature, to a cusp at T_c , to the classical limit as $T \rightarrow \infty$ – is schematically shown in Fig. D.4. Importantly, we also just explicitly calculated our first exact partition function that eventually led us to a discontinuity in some derivative of a free energy (here in the context of the derivative of a response function) – our first calculated phase transition!

D.10.3 Ground state occupation

We just found explicitly calculated a quantity that suggests a discontinuity in a higher derivative of a response function – the specific heat itself was continuous across the transition, but its derivative wasn't. This does *not* automatically tell you the order of the phase transition, it just sets a bound on a number of derivatives in which you will first see something non-analytic. Actually, we could have seen a non-analyticity much soon by looking at number of particles in the excited state above and below T_c . From Eq. D.95 we get that $N_{excited}/N = (T/T_c)^{3/2}$, and so we expect the relative occupation in the ground state to be

$$\frac{N_{ground}}{N} = \frac{N_{excited}}{N} = 1 - \left(\frac{T}{T_c} \right)^{3/2}, \quad T < T_c \quad (\text{D.102})$$

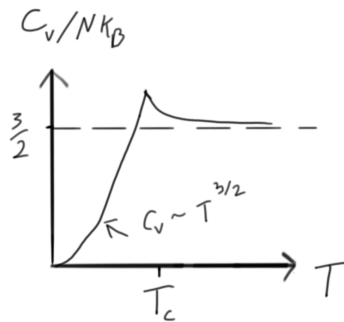


Figure D.4: **Heat capacity of an ideal Bose Gas** (note the cusp at T_c .)

and $\frac{N_{ground}}{N} \rightarrow 0$ in the thermodynamic limit for $T > T_c$.

First vs third-order transition; Landau, SSB, order parameter (like superfluidity, complex number associated with macroscopically occupied GS), etc

In the following chapter we'll more systematically look at interacting systems and phase transitions (in the simpler context of classical systems). First, we wrap up the chapter with a quick look at degenerate Fermi gases.

D.11 Ideal Fermi gases

Just as in the case of the ideal Bose gas, when $d = n_- \lambda^3 / g$ approaches unity we can no longer usefully rely on the power series expansions of the $f_m^n(z)$ in Eq. D.90, and we must start working with the full set of equations there. In the limit that $T \rightarrow 0$ we can look at the average fermi occupation number for states associated with \mathbf{k} :

$$\langle n_{\mathbf{k}} \rangle_- = \frac{1}{e^{\beta(E_{\mathbf{k}} - \mu)} + 1} = \begin{cases} 1 & E_{\mathbf{k}} < \mu \\ 0 & \text{otherwise} \end{cases} . \quad (\text{D.103})$$

At $T = 0$ this is just a step function, so at zero temperature all of the single-particle states up to $E_{\mathbf{k}} = \varepsilon_F$, the *fermi energy*, are completely filled, forming the so-called *fermi sea*. The corresponding wavenumber is¹⁷⁰ referred to as the *fermi wavenumber*, k_F . For an ideal gas with $E_{\mathbf{k}} = \hbar^2 k^2 / (2m)$, these are related by

$$N = \sum_{|\mathbf{k}| \leq k_F} g = gV \int_{k \leq k_f} \frac{d^3 \mathbf{k}}{(2\pi)^3} = \frac{gV}{6\pi^2} k_F^3, \quad (\text{D.104})$$

so

$$k_F = \left(\frac{6\pi^2 n}{g} \right)^{1/3}, \quad \varepsilon_F = \frac{\hbar^2}{2m} \left(\frac{6\pi^2 n}{g} \right)^{2/3} \quad (\text{D.105})$$

Schematically the behavior of the occupation numbers are shown in Fig. D.5.

We need to do some work to investigate the finite-temperature behavior (i.e., for large z). Here we follow Sommerfeld's approach, and first take Eq. D.87 and perform an integration by parts:

$$f_m^-(z) = \frac{1}{m!} \int_0^\infty dx x^m \frac{d}{dx} \left(\frac{-1}{e^{x - \log z} + 1} \right), \quad (\text{D.106})$$

where we've just written z^{-1} as $e^{-\log z}$ for convenience. We then say that we know the fermi occupation itself changes very rapidly from 1 to zero across ε_F , so the derivative above must be sharply peaked. We expand about this peak of the derivative (which occurs when $x = \log z$), by setting $x = \log z + t$ and taking the new integration variable $-\infty \leq t \leq \infty$.

¹⁷⁰shockingly

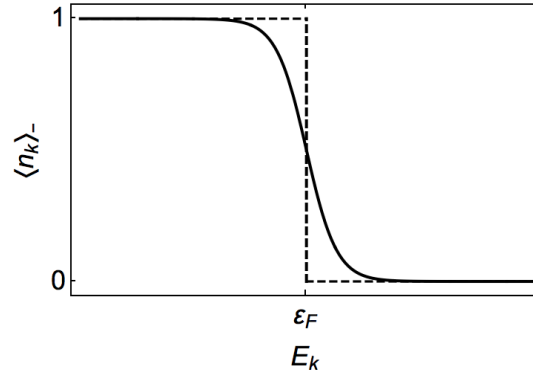


Figure D.5: **Fermi occupation numbers** The dotted line shows the zero-temperature limit, and the solid curve shows the finite temperature result.

This trick gives us:

$$\begin{aligned}
 f_m^-(z) &\approx \frac{1}{m!} \int_{-\infty}^{\infty} dt (\log z + t)^m \frac{d}{dt} \left(\frac{-1}{e^t + 1} \right) \\
 &= \frac{1}{m!} \int_{-\infty}^{\infty} dt \sum_{\alpha=0}^{\infty} \left[\binom{m}{\alpha} t^\alpha (\log z)^{m-\alpha} \right] \frac{d}{dt} \left(\frac{-1}{e^t + 1} \right) \\
 &= \frac{(\log z)^m}{m!} \sum_{\alpha=0}^{\infty} \frac{m!}{\alpha!(m-\alpha)!} (\log z)^{-\alpha} \int_{-\infty}^{\infty} dt t^\alpha \frac{d}{dt} \left(\frac{-1}{e^t + 1} \right). \quad (\text{D.107})
 \end{aligned}$$

The last type of integral appearing above can be manipulated (exploiting the anti-symmetry of the integrand under exchange of sign of t , etc.) to give:

$$\frac{1}{\alpha!} \int_{-\infty}^{\infty} dt t^\alpha \frac{d}{dt} \left(\frac{-1}{e^t + 1} \right) = \begin{cases} 0 & \text{if } \alpha \text{ is odd} \\ \frac{2}{(\alpha-1)!} \int_0^{\infty} dt \frac{t^{\alpha-1}}{e^t + 1} & \text{if } \alpha \text{ is even} \end{cases}, \quad (\text{D.108})$$

and in that last expression we recognize an expression which is just $2f_\alpha^-(1)$. So, we combine the above two equations, and exploit the fact that other people have computed the integrals associated with $f_m^-(1)$, to give the *Sommerfeld expansion*:

$$\begin{aligned}
 \lim_{z \rightarrow \infty} f_m^-(z) &= \frac{(\log z)^m}{m!} \sum_{\alpha \text{ even}}^{\infty} 2f_\alpha^-(1) \frac{m!}{(m-\alpha)!} (\log z)^{-\alpha} \\
 &= \frac{(\log z)^m}{m!} \left(1 + \frac{\pi^2}{6} \frac{m(m-1)}{(\log z)^2} + \dots \right). \quad (\text{D.109})
 \end{aligned}$$

To first approximation, explicitly, we have

$$\begin{aligned} f_{5/2}^-(z) &\approx \frac{8(\log z)^{5/2}}{15\sqrt{\pi}} \left(1 + \frac{5\pi^2}{8(\log z)^2} + \dots \right) \\ f_{3/2}^-(z) &\approx \frac{4(\log z)^{3/2}}{3\sqrt{\pi}} \left(1 + \frac{\pi^2}{8(\log z)^2} + \dots \right) \\ f_{1/2}^-(z) &\approx \frac{2(\log z)^{1/2}}{\sqrt{\pi}} \left(1 - \frac{\pi^2}{24(\log z)^2} + \dots \right) \end{aligned} \quad (\text{D.110})$$

We can now plug these results into Eq. D.90. When $z \gg 1$ the degeneracy factor is

$$\frac{n_- \lambda^3}{g} = f_{3/2}^-(z) \approx \frac{4(\log z)^{3/2}}{3\sqrt{\pi}} \left(1 + \frac{\pi^2}{8(\log z)^2} + \dots \right). \quad (\text{D.111})$$

The leading term reproduces our earlier result for the fermi energy:

$$\beta \varepsilon_F = \frac{\beta \hbar^2}{2m} \left(\frac{6\pi^2 n}{g} \right)^{2/3} = \left(\frac{3n\lambda^3}{4\sqrt{\pi}} \right)^{2/3} = \log z \quad (\text{D.112})$$

which gives a chemical potential of

$$\mu = k_B T \log z \approx \varepsilon_F \left(1 - \frac{\pi^2}{12} \left(\frac{k_B T}{\varepsilon_F} \right)^2 \right). \quad (\text{D.113})$$

Note that this is positive at low temperatures and negative at high temperatures, suggesting that the chemical potential changes sign at a fermi temperature $T_F \sim \varepsilon_F/k_B$. The energy density $E/V = 3P/2$, and the low-temperature pressure is

$$\beta P_- = \frac{g}{\lambda^3} \frac{8(\log z)^{5/2}}{15\sqrt{\pi}} \left(1 + \frac{5\pi^2}{8(\log z)^2} + \dots \right) \quad (\text{D.114})$$

With the help of our previous expression (relating the fermi energy to $\log z$, for instance), we can write this as a power series in temperature, as

$$P_- = \frac{2}{5} \varepsilon_F n_- \left(1 + \frac{5\pi^2}{12} \left(\frac{k_B T}{\varepsilon_F} \right)^2 + \dots \right). \quad (\text{D.115})$$

Unlike a classical gas, the degenerate fermi gas has finite pressure (and internal energy) even at zero temperature. Additionally, the heat capacity,

$$\frac{C_V}{Nk_B} = \frac{1}{Nk_B} \frac{\partial E}{\partial T} = \frac{\pi^2}{2} \frac{k_B T}{\varepsilon_F} \quad (\text{D.116})$$

varies as the first power of temperature at low T (something which is quite general in fermi gases, regardless of the dimension). This reflects the fact that only a small fraction (of order T/T_F) of the particles are excited at temperature T , and most of the particles do not feel the effects of the finite temperatures. Each of those excited particles gains about $k_B T$ of energy, hence $C_V \sim (T/T_F)$. Additionally, we see that the heat capacity at low temperature is much smaller than the classical expectation of $3Nk_B/2$.

D.12 Problems

D.12.1 Practice with density matrix formalism

We ended section D.4 of the notes with a few examples in which we calculated the density matrix for simple examples. In this problem we'll do another:

Consider the hamiltonian for a simple two-dimensional quantum mechanical rotor,

$$\mathcal{H} = -\frac{\hbar^2}{2I} \frac{d^2}{d\theta^2},$$

where $0 \leq \theta \leq 2\pi$, and eigenstates must satisfy $\psi(\theta) = \psi(\theta + 2\pi)$.

(A): What are the eigenstates of this Hamiltonian? What are the energy levels?

(B): What are the elements of the density matrix, $\langle \theta' | \rho | \theta \rangle$, in the canonical ensemble? Give simplified expressions for these elements in the high- and low-temperature limits.

D.12.2 Density matrices, Entropy, and unbiased estimation

In this question I basically want to know if the method of unbiased estimation makes sense in quantum mechanical settings. Suppose a quantum mechanical system has a hamiltonian \mathcal{H} , is at temperature T , and is described by a (potentially time-evolving) density matrix $\rho(t)$. Define the entropy associated with the density matrix as

$$S(t) = -\text{Tr} [\rho \log \rho].$$

(A) We already have expressions for the time evolution of density matrices; what is $\frac{dS}{dt}$?

(B) By using Lagrange multipliers (and perhaps thinking back to what we did in the chapter on probability), find the density matrix that maximizes the entropy, subject to the constraint that you know the average energy $\langle \mathcal{H} \rangle = \text{Tr}(\rho \mathcal{H}) = E$.

(C) Is your solution for this maximum-entropy density matrix stationary (i.e., is $\frac{\partial \rho}{\partial t} = 0$)?

D.12.3 Canonical ensemble density matrix for non-interacting, distinguishable particles

In Section D.6 we said that the product states are appropriate for *distinguishable* particles, and in Section D.7 we looked at the canonical ensemble density matrix for non-interacting particles where the eigenstates were either the fermionic or bosonic (anti)-symmetrizations of the product states. Study the density matrix and canonical partition function using the product states. **Show that if you do so, you do not get Gibbs' correction factor of $\frac{1}{N!}$.**

Additionally, **show that there are no spatial correlations between the particles in the system, and so there is no statistical interparticle potential.**

D.12.4 Discontinuity of the specific heat in Bose-Einstein condensation

Earlier I asserted that above T_c the slope of the specific heat of a BEC approaching the transition was

$$\left[\frac{C_{V,N}}{Nk_B} \right]_{T \rightarrow T_c^+} = \frac{1}{T_c} \left(\frac{45 f_{5/2}^+(1)}{8 f_{3/2}^+(1)} - \frac{27 \left(f_{3/2}^+(1) \right)^2 f_{-1/2}^+(1)}{\left(f_{1/2}^+(1) \right)^3} \right) \approx \frac{-0.777}{T_c}.$$

Derive this expression. Note that both $f_{-1/2}^+(z)$ and $f_{1/2}^+(z)$ individually diverge as $z \rightarrow \infty$, so take some care in evaluating the ratio of the powers of them that appear above.

D.12.5 Bose condensation in other dimensions

Consider an ideal gas of non-interacting, spinless bosons in a box whose generalized volume is $V \equiv L^d$ in d dimensions. We'll be working in the grand canonical ensemble.

(A): Given the value of the chemical potential μ , Calculate both the grand potential $\mathcal{G} = -k_B T \log \mathcal{Q}$, the number density $n = N/V$. Write your answer in terms of d and the $f_m^+(z)$ functions. *Hints: integration by parts will help with the expression for $\log \mathcal{Q}$. The lecture notes have expressions for surface areas and volumes of d -dimensional unit spheres*

(B): Calculate the ratio PV/E , and compare with the classical result for an ideal gas.

(C): Calculate the critical temperature, $T_c(n)$, for Bose-Einstein condensation in d dimensions.

(D): Calculate the heat capacity at low temperatures, $C(T)$ for $T < T_c(n)$.

(E): Calculate the heat capacity at high temperature. What is the ratio $C_{max}/C(T \rightarrow \infty)$?

(F): How does the ratio you calculated in the last part behave as $d \rightarrow 2$? What does this say about the dimensions for which your calculation was valid?

D.12.6 Specific heat for two-dimensional ideal quantum gases

In this problem we'll be considering the specific heat of ideal Fermi and Bose gases in two dimensions. Working directly with the expressions from Sec. D.8 of the notes will (probably) help, after being appropriately generalized for two rather than three dimensions.

Part A:

For a given N , T , and V , show that the fugacities of the Bose and Fermi systems, z_B and z_F , can be mutually related to each other:

$$(1 - z_B)(1 + z_F) = 1 \quad \Rightarrow \quad z_B = \frac{z_F}{1 + z_F}. \quad (\text{D.117})$$

Part B:

Use properties of the $f_m^\eta(z)$ to derive the following relationship:

$$f_2^{-1}(z_F) = f_2^{+1} \left(\frac{z_F}{1+z_F} \right) + \frac{1}{2} \log^2(1+z_F). \quad (\text{D.118})$$

Part C:

Combining your results, show that the energy of the Fermi system is equal to the energy of the Bose system plus a constant (and tell me what that constant is), i.e.

$$E_F(N, T) = E_B(N, t) + \text{constant}. \quad (\text{D.119})$$

You have just shown, among other things, that the specific heat of an ideal Fermi gas is *identical* to the specific heat of an ideal Bose gas for *all* T and N . Apparently the properties of two-dimensional ideal quantum gases are qualitatively different from the properties of three-dimensional ones!

D.12.7 Heat capacities of an ideal Fermi gas

(A): Show that for an ideal Fermi gas one has

$$\frac{1}{z} \frac{\partial z}{\partial T} \Big|_P = - \frac{5}{2T} \frac{f_{5/2}^-(z)}{f_{3/2}^-(z)}$$

(B): From the lecture notes we know the grand partition function for this system. Take the appropriate thermodynamic derivatives to derive the entropy of an ideal Fermi gas.

(C): Find an expression for $\gamma \equiv C_P/C_V$, the ratio of the heat capacities at constant pressure and at constant volume. Your answer should be given in terms of combinations of $f_m^-(z)$.

Check that the low temperature limit of your expression is

$$\gamma \approx 1 + \frac{\pi^2}{3} \left(\frac{k_B T}{\varepsilon_F} \right)^2$$

Bibliography

- [1] John Hopfield. Reflections on the aps and the evolution of physics. *Bulletin of the American Physical Society*, 52, 2007.
- [2] Raj Kumar Pathria. *Statistical mechanics*. Elsevier, 2016.
- [3] Mehran Kardar. *Statistical physics of particles*. Cambridge University Press, 2007.
- [4] Mehran Kardar. *Statistical physics of fields*. Cambridge University Press, 2007.
- [5] Nigel Goldenfeld. *Lectures on phase transitions and the renormalization group*. CRC Press, 2018.
- [6] John Preskill. Lecture notes for physics 229: Quantum information and computation. *California Institute of Technology*, 16(1):1–8, 1998.
- [7] David Tong. Kinetic theory. *Graduate Course, University of Cambridge, Cambridge, UK*, 2012.
- [8] David Tong. *Statistical physics*. University of Cambridge, 2011.
- [9] Kerson Huang. *Introduction to statistical physics*. CRC press, 2009.
- [10] James P Sethna. *Statistical mechanics: entropy, order parameters, and complexity*, volume 14. Oxford University Press, USA, 2021.
- [11] Leo P Kadanoff. *Statistical physics: statics, dynamics and renormalization*. World Scientific, 2000.
- [12] William Thomson. Xxxvi.—an account of carnot’s theory of the motive power of heat;* with numerical results deduced from regnault’s experiments on steam. *Earth and Environmental Science Transactions of The Royal Society of Edinburgh*, 16(5):541–574, 1849.
- [13] John Goold, Marcus Huber, Arnau Riera, Lídia Del Rio, and Paul Skrzypczyk. The role of quantum information in thermodynamics—a topical review. *Journal of Physics A: Mathematical and Theoretical*, 49(14):143001, 2016.
- [14] James H Luscombe. *Statistical mechanics: from thermodynamics to the renormalization group*. CRC Press, 2021.

- [15] Jan Baptist van Helmont. *Oriatrike or, physick refined. The common errors therein refuted, and the whole art reformed [and] rectified: being a new rise and progress of phylosophy and medicine, for the destruction of diseases and prolongation of life. Written By that most Learned, Famous, Profound and Acute Phylosopher, and Chemical Physitian.* Printed for Lodowick Loyd, London, [and] to be sold at his Shop next the Castle in Cornhill, 1662.
- [16] Constantin Carathéodory. Untersuchungen über die grundlagen der thermodynamik. *Mathematische Annalen*, 67(3):355–386, 1909.
- [17] R.H. Fowler and E.A. Guggenheim. *Statistical Thermodynamics: A Version of Statistical Mechanics for Students of Physics and Chemistry.* Macmillan, 1939.
- [18] Dapeng Bi, Silke Henkes, Karen E Daniels, and Bulbul Chakraborty. The statistical physics of athermal materials. *Annu. Rev. Condens. Matter Phys.*, 6(1):63–83, 2015.
- [19] R Clausius. Über verschiedene für die abwendung bequeme formen der haupt-gleichungen der mechanischen wärmetheorie. vierteiljahrschrift der naturforschenden gessellschaft (zürich) 1865 [chapter x (on several convenient forms of the fundamental equations of the mechanical theory of heat)]. *The mechanical theory of heat: with its applications to the steam engine and to physical properties of bodies.* London: John van Voorst, pages 1–59, 1867.
- [20] Joseph Louis Gay-Lussac. *The Free Expansion of Gases: Memoirs by Gay-Lussac, Joule, and Joule and Thomson*, volume 1. Harper, 1898.
- [21] Pierre-Gilles De Gennes. Soft matter. *Science*, 256(5056):495–497, 1992.
- [22] Wim Van Saarloos, Vincenzo Vitelli, and Zorana Zeravcic. *Soft Matter: Concepts, Phenomena, and Applications.* Princeton University Press, 2024.
- [23] Sriram Ramaswamy. The mechanics and statistics of active matter. *Annu. Rev. Condens. Matter Phys.*, 1(1):323–345, 2010.
- [24] Étienne Fodor, Robert L Jack, and Michael E Cates. Irreversibility and biased ensembles in active matter: Insights from stochastic thermodynamics. *Annual Review of Condensed Matter Physics*, 13(1):215–238, 2022.
- [25] Julien Tailleur, Gerhard Gompper, M Cristina Marchetti, Julia M Yeomans, and Christophe Salomon. *Active Matter and Nonequilibrium Statistical Physics: Lecture Notes of the Les Houches Summer School: Volume 112, September 2018*, volume 112. Oxford University Press, 2022.
- [26] Charles H Bennett. The thermodynamics of computation—a review. *International Journal of Theoretical Physics*, 21:905–940, 1982.
- [27] Peter Duren. Changing faces: the mistaken portrait of legendre. *Notices of the AMS*, 56(11):1440–1443, 2009.

- [28] Notices of the ams: About the cover. <http://www.ams.org/notices/200911/200911-about-the-cover.pdf>.
- [29] Terry Renner. *Quantities, units and symbols in physical chemistry*. Royal Society of Chemistry, 2007.
- [30] Michael Atiyah. Mathematics in the 20th century. *Bulletin of the London Mathematical Society*, 34(1):1–15, 2002.
- [31] E.T. Jaynes, *Unpublished notes on thermodynamics* Chapter II: Use of jacobians in thermodynamics. <https://bayes.wustl.edu/etj/thermo/stat.mech.2.pdf>. Accessed: 2023-11-26.
- [32] J.C. Baez, classical mechanics versus thermodynamics. https://math.ucr.edu/home/baez/classical_vs_thermo/classical_vs_thermo_web.pdf. Accessed: 2023-11-26.
- [33] Subrahmanyan Chandrasekhar and Subrahmanyan Chandrasekhar. *An introduction to the study of stellar structure*, volume 2. Courier Corporation, 1957.
- [34] Mark A Peterson. Analogy between thermodynamics and mechanics. *American Journal of Physics*, 47(6):488–490, 1979.
- [35] SG Rajeev. A hamilton–jacobi formalism for thermodynamics. *Annals of Physics*, 323(9):2265–2285, 2008.
- [36] Osborne Reynolds. Lvii. on the dilatancy of media composed of rigid particles in contact. with experimental illustrations. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 20(127):469–481, 1885.
- [37] James Thomson. I. a quantitative investigation of certain relations between the gaseous, the liquid, and the solid states of water-substance. *Proceedings of the Royal Society of London*, 22(148-155):27–36, 1874.
- [38] H Eugene Stanley. *Phase transitions and critical phenomena*, volume 7. Clarendon Press, Oxford, 1971.
- [39] Nigel Goldenfeld and Leo P Kadanoff. Simple lessons from complexity. *science*, 284(5411):87–89, 1999.
- [40] Wikipedia: Universality class. https://en.wikipedia.org/wiki/Universality_class.
- [41] Richard J Creswick and Seung-Yeon Kim. Critical exponents of the four-state potts model. *Journal of Physics A: Mathematical and General*, 30(24):8785, 1997.
- [42] Ben Widom. Some topics in the theory of fluids. *The Journal of Chemical Physics*, 39(11):2808–2812, 1963.

- [43] R Pelka, Piotr Konieczny, Magdalena Fitta, Mariusz Czapla, PM Zielinski, M Balanda, T Wasiutynski, Yuji Miyazaki, Akira Inaba, Dawid Pinkowicz, et al. Magnetic systems at criticality: different signatures of scaling. *arXiv preprint arXiv:1312.0756*, 2013.
- [44] C Kramp. *Elements d'arithmétique universelle*. L'imprimerie de Th. F. Thiriart, Cologne, 1808.
- [45] Yes.
- [46] Claude E Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.
- [47] Edwin T Jaynes. Where do we stand on maximum entropy? *The maximum entropy formalism*, 1979.
- [48] Steve Pressé, Kingshuk Ghosh, Julian Lee, and Ken A Dill. Principles of maximum entropy and maximum caliber in statistical physics. *Reviews of Modern Physics*, 85(3):1115, 2013.
- [49] Ram Avinery, Micha Kornreich, and Roy Beck. Universal and accessible entropy estimation using a compression algorithm. *Physical review letters*, 123(17):178102, 2019.
- [50] Stefano Martiniani, Paul M Chaikin, and Dov Levine. Quantifying hidden order out of equilibrium. *Physical Review X*, 9(1):011031, 2019.
- [51] Prabhu Lal Bhatnagar, Eugene P Gross, and Max Krook. A model for collision processes in gases. i. small amplitude processes in charged and neutral one-component systems. *Physical review*, 94(3):511, 1954.
- [52] David L Goodstein. *States of matter*. Courier Corporation, 2014.
- [53] Arthur Stanley Eddington. *New pathways in science*. Read Books Ltd, 2013.
- [54] Edwin T Jaynes. The gibbs paradox. *Maximum Entropy and Bayesian Methods: Seattle, 1991*, pages 1–21, 1992.
- [55] Daan Frenkel. Why colloidal systems can be described by statistical mechanics: some not very original comments on the gibbs paradox. *Molecular Physics*, 112(17):2325–2329, 2014.
- [56] Josiah Willard Gibbs. *Elementary principles in statistical mechanics: developed with especial reference to the rational foundations of thermodynamics*. C. Scribner's sons, 1902.
- [57] Gilbert Newton Lewis, Merle Randall, Kenneth S Pitzer, and Leo Brewer. *Thermodynamics*. Courier Dover Publications, 2020.
- [58] John F Marko and Eric D Siggia. Stretching dna. *Macromolecules*, 28(26):8759–8770, 1995.

- [59] Edward Bernard Stuart, Benjamin Gal-Or, and Alan J Brainard. *A Critical review of thermodynamics: the proceedings of an international symposium, held at the University of Pittsburgh, Pittsburgh, Pa., April 7-8, 1969*. Mono Book Corp., 1970.
- [60] Joseph Edward Mayer and Maria Goeppert Mayer. *Statistical Mechanics*. John Wiley & Sons, 1940.
- [61] Elliott W Montroll and Joseph E Mayer. Statistical mechanics of imperfect gases. *The Journal of Chemical Physics*, 9(8):626–637, 1941.
- [62] Richard J Wheatley. Calculation of high-order virial coefficients with applications to hard and soft spheres. *Physical review letters*, 110(20):200601, 2013.
- [63] Jean-Pierre Hansen and Ian Randal McDonald. *Theory of simple liquids: with applications to soft matter*. Academic press, 2013.
- [64] Robert B Griffiths. A proof that the free energy of a spin system is extensive. *Journal of Mathematical Physics*, 5(9):1215–1222, 1964.
- [65] Richard P Feynman. *Statistical mechanics: a set of lectures*. CRC press, 2018.
- [66] Leo P Kadanoff. Scaling laws for ising models near t c. *Physics Physique Fizika*, 2(6):263, 1966.
- [67] Kenneth G Wilson. Renormalization group and critical phenomena. i. renormalization group and the kadanoff scaling picture. *Physical review B*, 4(9):3174, 1971.
- [68] Friedrich Johannes Wilhelm Hahne. *Critical Phenomena: Proceedings of the Summer School Held at the University of Stellenbosch, South Africa January 18–29, 1982*, volume 186. Springer, 2005.
- [69] DJ Wallace and RKP Zia. The renormalisation group approach to scaling in physics. *Reports on Progress in Physics*, 41(1):1, 1978.
- [70] Th Niemeyer and JMJ Van Leeuwen. Wilson theory for 2-dimensional ising spin systems. *Physica*, 71(1):17–40, 1974.
- [71] Lars Onsager. Crystal statistics. i. a two-dimensional model with an order-disorder transition. *Physical Review*, 65(3-4):117, 1944.
- [72] Arthur Thomas Quiller-Couch. *On the art of writing*. GP Putnam’s Sons, 1916.
- [73] Max Planck. Letter from m. *Planck to RW Wood. (See Hermann, A., 1971, The Genesis of Quantum Theory (1899-1913), MIT Press, Cambridge Mass., pp. 23-24)*, 1931.
- [74] John Crepeau. Josef stefan: His life and legacy in the thermal sciences. *Experimental thermal and fluid science*, 31(7):795–803, 2007.