Variational Modelling:
Energies, gradient flows, and large deviations

Mark Peletier

July 13, 2013
Version 1.0

Preface

Acknowledgements

Thanks to David Bourne, Alex Cox, Manh Hong Duong, Joep Evers, Alexander Mielke, Georg Prokert, and Michiel Renger for comments on earlier drafts of these notes.

Notation

$\lVert \cdot \rVert_{1,\rho}$ and $(\cdot,\cdot)_{1,\rho}$	dual Wasserstein norm and inner product	Sec. 3.4.2
$\lVert \cdot \rVert_{-1,\rho}$ and $(\cdot,\cdot)_{-1,\rho}$	Wasserstein norm and inner product	Sec. 3.4.2
$C_b^k(\Omega)$	space of functions with k ct. and bounded derivatives	
Ent	entropy $\rho \mapsto \int \rho \log \rho$	(4.2)
$\mathcal{H}(\rho\lvert\mu)$	relative entropy of ρ with respect to μ	Def. 4.1
$\mathcal{L}^d$	d-dimensional Lebesgue measure	
$\mathcal{M}(\Omega)$	Borel regular, locally finite measures on Ω	Appendix A
$\mathscr{P}(\Omega)$	probability measures on Ω	Appendix A
$P_z\mathcal{Z}$ and $P\mathcal{Z}$	process space at $z \in \mathcal{Z}$ and process bundle	Sec. 3.3
$\mathcal{P}$	mapping from process space to tangent space	Sec. 3.3
R	universal gas constant, $R = kN_A \simeq 8.314\,\mathrm{J\,K^{-1}\,mol^{-1}}$	
$T_z\mathcal{Z}, T\mathcal{Z}$	tangent space at $z \in \mathcal{Z}$ and tangent bundle	Sec. 3.1
$W_2(\rho_0,\rho_1)$	Wasserstein distance between ρ_0 and ρ_1	Sec. 3.4.4
$\mathcal{Z}$	abstract name of the state space	

Contents

Chapter 1

Introduction

1.1 Variational modelling

Modelling is the art of taking a real-world situation and constructing some mathematics to describe it. It's an art rather than a science, because it involves choices that can not be rationally justified. Personal preferences are important, and 'taste' plays a major role.

Having said that, some choices are more rational than others. And we often also need to explain our choices to ourselves and to our colleagues. In these notes we describe a framework for making and explaining such choices, a framework that we call *variational modelling*, which provides both a workflow for deriving a model and a body of theory to support it.

We restrict ourselves to the class of *dissipative systems*, which are systems driven by dissipation—dissipation of entropy or free energy, for instance—and more precisely, to the subclass of these systems with the mathematical structure of *gradient flows*. In its simplest form, and in abstract terms, variational modelling of such a system consists of choosing a state space $\mathcal{Z}$, a driving functional $\mathcal{F}$, and a *dissipation mechanism* $\mathcal{D}$. Once these three are chosen, the evolution equations can be derived, typically in the form of differential or differential-integral equations. The crucial point is that the triple $(\mathcal{Z}, \mathcal{F}, \mathcal{D})$ contains *all the modelling choices*: it completely determines the model.

Therefore all discussion about the modelling necessarily focuses on the choice of $(\mathcal{Z}, \mathcal{F}, \mathcal{D})$:

> *What are the modelling assumptions that underlie these choices?*

In these notes we will explain various possible choices, discuss their features, and provide the background and theory that we will need to give a good answer to this question. We will also discuss various generalizations of the simple version above, for instance to 'generalized' gradient flows and systems with inertia. Finally, throughout the notes we will give many examples to illustrate the concepts.

1.2 Gradient flows

Gradient flows are systems that can be written in one of the forms

$$\dot{z} = -\mathcal{K}(z)\mathcal{F}'(z) \qquad \text{or} \qquad (1.1a)$$

$$\mathcal{G}(z)\dot{z} = -\mathcal{F}'(z). \qquad (1.1b)$$

Here $z \in \mathcal{Z}$ is the state of the system, $\mathcal{F}'(z)$ is the (Fréchet) derivative of $\mathcal{F}$ at z, and the linear operators $\mathcal{K}$ and $\mathcal{G}$ are two alternative ways of characterizing the dissipation mechanism $\mathcal{D}$ mentioned above. Both are assumed to be non-negative and symmetric (all these notions are made concrete in Chapter 3). Whenever $\mathcal{K}$ and $\mathcal{G}$ are invertible we assume $\mathcal{K} = \mathcal{G}^{-1}$, and the two formulations above are exactly equivalent.

Such systems are driven by the functional $\mathcal{F}$: the functional $\mathcal{F}$ decreases along solutions, since the non-negativity of $\mathcal{K}$ and $\mathcal{G}$ implies that

$$\frac{d}{dt}\mathcal{F}(z(t)) = \mathcal{F}' \cdot \dot{z} = -\mathcal{F}' \cdot \mathcal{K}(z)\mathcal{F}' = -\dot{z} \cdot \mathcal{G}(z)\dot{z} \leq 0. \qquad (1.2)$$

In fact the evolution can be assumed to decrease $\mathcal{F}$ as fast as possible, where the meaning of *as fast as possible* is given by the operators $\mathcal{K}$ and $\mathcal{G}$, i.e. by the dissipation mechanism (see Chapter 3).

In the modelling of real-world systems, $\mathcal{F}$ is often an *energy*, which makes equation (3.1b) a *force balance:* the right-hand side is the derivative of the energy, i.e. a generalized force. The left-hand side can be interpreted as the (generalized) force necessary for the instantaneous (generalized) velocity $\dot{z}$.

The right-hand side of (1.2) contains two quadratic forms that will often return. These are related to the *dissipation potential* and the *dual dissipation potential*

$$\Psi(z,s) := \frac{1}{2}s \cdot \mathcal{G}(z)s \qquad \text{and} \qquad \Psi^*(z,\xi) := \frac{1}{2}\xi \cdot \mathcal{K}(z)\xi.$$

For fixed z, these are dual quadratic forms in s and ξ in the sense of Legendre transforms (see (3.3)),

$$\Psi^*(z,\xi) = \sup_s \big[\xi \cdot s - \Psi(z,s)\big]$$

$$\Psi(z,s) = \sup_\xi \big[\xi \cdot s - \Psi^*(z,\xi)\big].$$

This relation between Ψ and Ψ^* gives a natural generalization of the relationship between $\mathcal{G}$ and $\mathcal{K}$ to the case of non-invertible operators. It also shows, in a very general sense, how it is equivalent to choose any one of the four objects $\mathcal{G}$, $\mathcal{K}$, Ψ, or Ψ^*—the other three follow automatically.

A central tool in these notes is a reformulation of (1.1b) in terms of Ψ. At any given point z, equation (1.1b) is the stationarity condition associated with the minimization problem

$$\inf_s \Psi(z,s) + \mathcal{F}'(z) \cdot s. \qquad (1.3)$$

5

If $\mathcal{G}$ is invertible, then this expression is strictly convex in s for fixed z, and minimizers of (1.3) are unique. The minimizer s is then equal to the unique solution $\dot{z}$ of (1.1b). On the other hand, when $\mathcal{G}$ is not invertible, then (1.3) suggests a natural extension (see Section 3.2).

1.3   Making choices

The modelling procedure that we follow in these notes consists of four steps:

1. Choose *state space* $\mathcal{Z}$;

2. Choose a functional $\mathcal{F}$ that *drives* the evolution;

3. Choose the *dissipation mechanism*, expressed by $\mathcal{G}$, $\mathcal{K}$, Ψ, or Ψ^*;

4. Derive the equations.

Often, Step 3 consists of two parts:

3a. Choose the space of all *processes* that participate in changing the state;

3b. Choose the *dissipation potential* Ψ as a function of these processes.

The purpose of these notes is to describe how to do this for a reasonable class of systems, and to explain how one can understand the choices that are made.

One important choice is not yet listed: the choice to use a gradient-flow, or a generalized gradient flow (Section 3.2) to describe the system. In practice this choice means two things: first, that inertia plays no role, and second, that the coordinate frame is inertial.

Inertia becomes unimportant when viscous or frictional forces dominate the inertial forces. We work out an example of this in the first system in Chapter 2. Gradient-flow-like systems that *do* include inertia have been developed in various forms; currently the most popular version is called GENERIC (General Equation for Non-Equilibrium Reversible-Irreversible Coupling [Ött05]. It includes generalized gradient flows as a special case, and like generalized gradient flows, provides a form of inherent thermodynamic consistency.

Choosing an inertial frame is important in order to avoid ficticious forces. When the energy is minimal, its derivative is zero, and by (1.1) therefore the system should be stationary. For instance, an object that is stationary in an inertial frame is rotating when viewed from a rotating frame, and therefore experiences a ficticious, centripetal force—even though the energy might be minimal.

Note: Often the functional $\mathcal{F}$ that we choose will be the 'free energy' that is well known in thermodynamics and statistical mechanics. Many modellers, indeed, would start by postulating that this free energy drives the evolution, without further comment. In these notes we want to presume very little, and therefore we will show *when*, *how*, and *why* the free energy drives the evolution.

1.4   Overview

These notes are structured as follows.

- Chapter 2 describes the application of the variational-modelling framework to three examples: a spring-dashpot system, the slow flow of a viscous fluid, and the diffusion of solutes in water.

- In Chapter 3 we recall some of the basic facts of gradient flows, including the Wasserstein metric and Wasserstein gradient flows.

- The third example in Chapter 2, on diffusion, involves the concepts of entropy, free energy, and Wasserstein dissipation. In Chapter 4 we study entropy and free energy, and create an understanding for why they appear in the driving functional of a gradient-flow system. In Chapter 5 we do the same for the Wasserstein dissipation.

- In Chapter 6 we discuss further examples of the theory.

- In Chapter 7 we comment on the relation between the variational-modelling framework on one hand and thermodynamics on the other.

We conclude with an appendix on the concepts of measure theory that we use.

Chapter 2

Examples

In this first chapter we discuss three examples—models whose main purpose is to provide illustrations of the general principles in a simple setting. The variational-modelling framework is overpowered for these simple systems, and the modelling may therefore feel slightly clumsy. The benefit of the simplicity is of course that we can concentrate on the essence, and show in a simple, well-known setting, how the framework functions.

The examples are (a) a spring-dashpot system, (b) slow viscous flow, and (c) diffusion of solutes in a fluid. The third example, on diffusion, will figure as a central example in the rest of these notes, since it illustrates the important concepts of entropy and Wasserstein dissipation.

Dimensions: Throughout these notes, we work with mathematical objects that have dimensions. As a result all the usual physical constants will appear, and these constants also help in understanding the origin and interpretation of various terms. Of course, once a model has been constructed, it is often wise to scale the variables and parameters in a different way to facilitate analysis and numerical approximation.

2.1 A spring-dashpot system

As described in the previous chapter, in the variational-modelling framework the construction of a model consists of choosing three entities: the state space, the driving functional, and the dissipation mechanism. In our first example we illustrate this on a very simple system: a spring and a dashpot that are connected—see Figure 2.1.
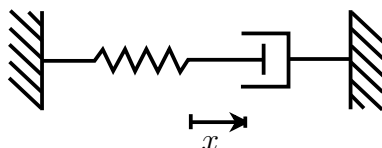

Figure 2.1: A simple spring-dashpot system. The variable x is the extension of the spring and at the same time the position of the dashpot.

State space: We characterize the state of the system by the single variable $x \in \mathbb{R}$, the extension of both the spring and the dashpot.

Energy: We choose for the energy of this system the potential energy of the spring $\mathcal{F}(x)$; for a linear spring this would be $kx^2/2$, where k is the spring constant.

Dissipation: We assume that energy is dissipated in the dashpot, and only there. We assume that the dashpot is linear, which means that the force f on the dashpot and the velocity v of the plunger are related by $f = \eta v$, where η is a constant characterizing the dashpot. Since $\mathcal{G}\dot{z}$ should be interpreted as the force necessary to move the system with velocity $\dot{z}$ (see Section 1.2), the relation $f = \eta v$ is the same as $\mathcal{G}v = \eta v$ for any $v \in \mathbb{R}$. Correspondingly $\mathcal{K}\xi = \xi/\eta$ for all ξ. In this case $\Psi(x, s) = \eta s^2/2$ and $\Psi^*(x, \xi) = \xi^2/2\eta$.

Equation: Applying (1.1a) we find the equation $\dot{x} = -\mathcal{K}\mathcal{F}'(x) = -\eta^{-1}\mathcal{F}'(x)$. For the linear spring $\mathcal{F}(x) = kx^2/2$ it becomes $\dot{x} = -\eta^{-1}kx$.

Discussion

Assumptions and conclusions. The list of boldfaced terms above highlights the assumptions that we made. In this context, the concepts and assumptions are fairly standard: the force of the spring is the (negative) derivative of the energy, and the velocity of the plunger in the dashpot is linearly related to the force on the plunger.

No inertia. As we mentioned above, an essential assumption is that inertia plays no role—or more accurately, that inertia can be neglected. In practice this comes down to requiring that the viscous forces dominate the inertial forces in the system.

As an example, consider the case of a spring-dashpot system with mass, for instance (Figure 2.2). Assuming a linear spring (with constant k) and a linear dashpot (with
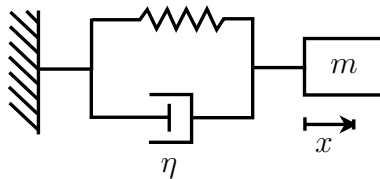
Figure 2.2: A non-gradient-flow system: a mass connected to a spring and a dashpot

constant η), the equation for the position x of the mass is

$$m\ddot{x} + \eta\dot{x} + kx = 0.$$

Rescaling this equation by defining a dimensionless time τ in terms of the old time t, $\tau = t\,\eta/k$, we find

$$\frac{mk}{\eta^2}\partial_{\tau\tau}x + \partial_\tau x + x = 0.$$

When the dimensionless combination mk/η^2 is small, this equation is a singular perturbation of the viscous equation $\partial_\tau x + x = 0$, or in dimensional terms, $\eta\dot{x} + kx = 0$. Apart from

a fast accommodation at time zero, the solutions of the full system follow the solution of the 'outer equation' $\eta\dot{x} + kx = 0$ (see e.g. [Hin91, Ch. 5]).

2.2   Slow viscous flow

The system is that of the slow flow of an incompressible, highly viscous liquid, with variable mass density, driven by gravity. The word 'slow' means that inertia will be neglected, and a gradient-flow structure is natural.

State space: We choose a bounded set $\Omega \subset \mathbb{R}^3$ to be the container with the fluid. The state variable will be the mass density $\rho : \Omega \to [0, \infty)$ (in $\mathrm{kg\,m^{-3}}$). The state space is therefore $\mathcal{Z} := \{\rho \in L^1(\Omega) : \rho \geq 0\}$. Positions in Ω are labeled x.

Energy: We choose as driving functional for this system the gravitational potential

$$\mathcal{F} : \mathcal{Z} \to \mathbb{R}, \qquad \mathcal{F}(\rho) := g \int_\Omega x_3 \rho(x)\, dx,$$

where $g \approx 9.8 \mathrm{\ m\,s^{-2}}$ is the gravitational acceleration and where we choose the last coordinate x_3 of $x \in \mathbb{R}^3$ to be the vertical coordinate.

Processes: The system can evolve through a *flow field* $u : \Omega \to \mathbb{R}^3$ satisfying the incompressibility condition $\mathrm{div}\, u = 0$ in Ω and the no-slip condition $u = 0$ at $\partial\Omega$. For a given flow field u, the density ρ evolves according to

$$\dot{\rho} + \mathrm{div}\, \rho u = 0 \qquad \text{in } \Omega. \tag{2.1}$$

Note that since $u = 0$ on the boundary $\partial\Omega$, this evolution preserves the total mass $\int_\Omega \rho$. We call the flow field u a *process vector*, since it characterizes all possible processes in this model; it generates a corresponding rate of change in the state, $\dot{\rho}$, through (2.1).

Dissipation mechanism: Energy is dissipated by viscous friction: the layers of liquid sliding over each other generate heat. This dissipation is characterized by the dissipation potential

$$\widetilde{\Psi}(u) := \frac{\eta}{2} \int_\Omega |\varepsilon(u)(x)|^2\, dx,$$

where $\eta > 0$ is a parameter (the 'dynamic viscosity' of the fluid, in $\mathrm{kg\,m^{-1}\,s^{-1}}$), and $\varepsilon(u) := \frac{1}{2}(\nabla u + \nabla u^T)$ is the symmetric part of ∇u. This function $\widetilde{\Psi}$ is one-half of the actual dissipation of energy (see e.g. [LL87, (16.3)]).

Equations: The resulting equations are now determined by solving the minimization problem

$$\inf_u \widetilde{\Psi}(u) + \langle \mathcal{F}'(\rho), -\mathrm{div}\, \rho u \rangle, \tag{2.2}$$

We introduced this minimization problem in Section 1.2, in the case of a dissipation potential Ψ defined on tangent vectors $\dot{c}$, instead of the potential $\widetilde{\Psi}$ defined on flow fields u.

We discuss the difference between these two in Section 3.3. The dual pairing is the pairing of the Fréchet derivative $\mathcal{F}'$ with a tangent vector (Section 3.1).

Writing the expression in (2.2) as

$$\frac{\eta}{2}\int_\Omega |\varepsilon(u)|^2 - g\int_\Omega x_3 \operatorname{div} \rho(x)u(x)\,dx = \frac{\eta}{2}\int_\Omega |\varepsilon(u)|^2 + g\int_\Omega e_3\rho u,$$

where e_3 is the unit vector for the third coordinate, we calculate the stationarity condition for (2.2) to be

$$\forall \tilde{u}: \qquad 0 = \eta\int_\Omega \varepsilon(u):\varepsilon(\tilde{u}) + g\int_\Omega e_3\rho\tilde{u} - \int_\Omega p\operatorname{div}\tilde{u},$$

where p is a Lagrange multiplier associated with the condition $\operatorname{div} u = 0$, and the identity holds for all $\tilde{u}$ that vanish on $\partial\Omega$ (even those that are not divergence-free). By applying partial integration we deduce that

$$
\begin{aligned}
-\operatorname{div}[\eta\varepsilon(u) - pI] &= -e_3\rho & &\text{in } \Omega,\\
\operatorname{div} u &= 0 & &\text{in } \Omega,\\
u &= 0 & &\text{on } \partial\Omega.
\end{aligned}
$$

Discussion

Modelling choices.   As in all the examples of these notes, the lack of inertial effects is important. In this case it brings us to the Stokes equations instead of the Navier-Stokes equations.

Note how in this case we choose the state space $\mathcal{Z}$ separately from the space of *processes*. This is a feature that we will encounter more often, since from a modelling point of view the *state* often has little in common with the *way the state can change*.

Quantifying the dissipation mechanism.   In the spring-dashpot example we could specify either $\mathcal{G}$ or $\mathcal{K}$; both were relatively simple to characterize. In this case it is more straightforward to specify Ψ than any of the other three possibilities $\mathcal{G}$, $\mathcal{K}$, and Ψ^*.

In the case of quadratic dissipation potentials, as in this example, there is a useful connection between the dissipation *potential* and the actual dissipation D of energy. This deserves its own discussion.

Dissipation and dissipation potential   A consequence of the gradient-flow, or dissipative structure that we use here, is the free-energy-dissipation identity

$$\mathcal{F}(\rho_T) + \int_0^T D(u_t)\,dt = \mathcal{F}(\rho_0),$$

or the differential version,

$$\partial_t\mathcal{F}(\rho_t) = -D(u_t), \tag{2.3}$$

where we use subscripts to indicate dependence on time. Here u_t is assumed to be the minimizer of (2.2), and the *dissipation* D is defined as

$$D(u) := \eta \int_\Omega |\varepsilon(u)|^2 \, dx.$$

By (2.3) D is indeed the dissipation of energy $\mathcal{F}$ per unit of time.

The *dissipation* D and the *dissipation potential* Ψ are closely related, but different. The dissipation D quantifies the actual conversion of energy into heat; the dissipation potential Ψ is the object in the minimization problem (2.2). The general formula is $D(u) = \langle D_u \Psi(u), u \rangle$, and when D and Ψ are quadratic in u, this expression simplifies to $D = 2\Psi$.

2.3 Diffusion of particles in a fluid

Our third example will be central in the discussion about 'how' and 'why': it will be the main example on which we illustrate the methods. As in the previous examples, we already know the equations for this system, but we will re-derive them using the variational framework. We will make choices for the state space, the driving functional, and the dissipation structure, and we will show how they reproduce the known equations for this system.

For this system, however, the choice of driving functional and dissipation are not immediately clear—and we will not explain them in this chapter. Indeed, this is exactly the *raison d'être* of these notes: the variational modelling philosophy is powerful, but the assumptions underlying it are not obvious. The purpose of this example is therefore to provide a test system, on which we will develop and test the understanding in Chapters 4 and 5.

We model the diffusion of a collection of particles in a stationary viscous fluid under the influence of gravity and thermal agitation.

State space: We choose a bounded set $\Omega \subset \mathbb{R}^d$ to be the container with the fluid; the particles are represented by their concentration $c : \Omega \to [0, \infty)$ (in moles per m^3). The state space is therefore $\mathcal{Z} := \{c \in L^1(\Omega) : c \geq 0\}$. Positions in Ω are labeled x.

This choice represents many assumptions at the same time: the particles are indistinguishable from each other, there are many of them and we are only interested in their density, and the fluid remains stationary; and the fact that we did not include temperature as a variable implies that temperature is held constant.

Energy: We choose as driving functional for this system the following combination of *entropy* and *gravitational energy*:

$$\mathcal{F} : \mathcal{Z} \to \mathbb{R}, \qquad \mathcal{F}(c) := RT \int_\Omega c(x) \log \frac{c(x)}{c_0} \, dx + \rho g \int_\Omega x_3 c(x) \, dx. \qquad (2.4)$$

Here $c_0 > 0$ is an arbitrary reference concentration (since it is arbitrary, it should not appear in the final equations; and indeed it does not), $R \simeq 8.314 \, \text{J K}^{-1} \text{mol}^{-1}$ is the *universal*

gas constant and T is the temperature; ρ is the mass density contrast of the particles (the difference between the density of the particles and that of the fluid, in kg/m^3), and $g \simeq 9.8\,\mathrm{m/s}^2$ is the gravitational acceleration.

This choice follows from assuming that the particles have no spatial preference, that they are thermally agitated at temperature T, that there are many of them, and that they are independent in a certain way. How these aspects can be recognized in this choice is indeed a non-trivial question, which is treated in detail in Chapters 4 and 5. For now we continue with the next part of the modelling.

Processes: Similarly to (2.1) we allow the state $c \in \mathcal{Z}$ to change through the effects of a flux $w : \Omega \to \mathbb{R}^d$ by

$$\dot{c} + \operatorname{div} cw = 0 \quad \text{in } \Omega, \qquad w \cdot n = 0 \quad \text{on } \partial\Omega, \tag{2.5}$$

or in weak form,

$$\forall \varphi \in C_b^1(\Omega) : \qquad \partial_t \int_\Omega \varphi(x) c(t, x)\, dx - \int_\Omega c(t, x) w(t, x) \nabla\varphi(x)\, dx = 0.$$

This is again a modelling choice: for instance, the divergence form of (2.5) implies that no particles can be created or removed. If one wants to allow for chemical reactions, for instance, which create and destroy partices, then the characterization (2.5) should also contain terms that are not in divergence form. The boundary condition $w \cdot n = 0$ is the modelling assumption that no particles enter or leave through the boundary.

The set of all *process vectors* is therefore the set $\{w : \Omega \to \mathbb{R}^d : w \cdot n = 0 \text{ on } \partial\Omega\}$. Each process vector w generates a *tangent vector* $\dot{c}$ through the relation (2.5).

Now that we have characterized the processes that change the state, we can choose the dissipation:

Dissipation potential: We define the *dissipation potential* on the set of process vectors w as the functional

$$\widetilde{\Psi}(c; w) := \frac{\eta}{2} \int_\Omega c(x) |w(x)|^2\, dx. \tag{2.6}$$

This choice follows from assuming that the fluid remains stationary and is linearly viscous, that the particles are sufficiently symmetric, and that they are far enough from each other to prevent interference between the flow fields surrounding them. Here the constant $\eta > 0$ is a *friction* coefficient: it characterizes the ratio between a velocity of a particle through the fluid and the corresponding force exerted by the fluid on the particle. As in the case of the entropy, the origin of this choice will be discussed at length in Chapter 5.

As in the previous example, the dissipation potential is one half of the *actual dissipation* of free energy; see the discussion on page 11.

Equations: To derive the equations, we postulate that at a given time t and at state c, the process vector w is characterized by the minimization problem

$$\min_{w, \dot{c}} \left\{ \widetilde{\Psi}(c; w) + \langle \mathcal{F}'(c), \dot{c} \rangle : \ \dot{c} \text{ and } w \text{ connected by (2.5)} \right\}, \tag{2.7}$$

as we also did in the previous example.

We find the equation for c from the stationarity condition: writing the expression in braces as

$$\widetilde{\Psi}(c; w) + \langle \mathcal{F}'(c), -\operatorname{div} w \rangle = \widetilde{\Psi}(c; w) + RT \int_\Omega \nabla\left(\log \frac{c}{c_0} + 1\right) \cdot w,$$

the stationarity condition is seen to be

$$\forall \tilde{w}: \qquad 0 = \int_\Omega cw\tilde{w}\, dx + RT \int_\Omega \tilde{w}\nabla c\, dx,$$

which implies $cw = -RT/\eta\, \nabla c$ and leads to the equation

$$\dot{c} = \frac{RT}{\eta}\Delta c \qquad \text{in } \Omega, \tag{2.8a}$$

$$\partial_n c = 0 \qquad\qquad \text{on } \partial\Omega. \tag{2.8b}$$

Discussion

Comparison to classical derivations. A common way of deriving the diffusion equation (2.8) is by combining the conservation law (2.5) with a *constitutive law* $cw = -D\nabla c$, known under the name of *Fick's law*. This leads to the same equation, but with RT/η replaced by D.

In this form Fick's law is beautifully simple, and indeed there is little reason to use the overpowered variational-modelling route to model this simple system. However, for more complex systems it is often not clear how to generalize Fick's law, for instance when species interfere with each other's mobility, as in the case of cross-diffusion, or when species diffuse on lower-dimensional structures such as surfaces or linear structures. In addition, for some systems the species concentration has an energetic impact that is not easy to quantify in a simple constitutive law, such as the case when in vesicle membranes the species changes the rigidity of those membranes.

The motivation is not yet complete. As indicated, we have not yet explained which modelling choices are represented by the chosen free energy $\mathcal{F}$ and the dissipation potential Ψ. After we provide this motivation in Chapters 4 and 5, this derivation, and those that follow, will be a self-contained modelling route.

The derivation is formal. The description of all allowable changes $\dot{c}$ is formal; we did not specify conditions on $\dot{c}$ and w in (2.5). In many other situations the situation is even worse, since $\mathcal{Z}$ is often a set without a differentiable structure, and therefore a 'smooth curve' might not even exist. It is to handle this type of difficulty that the theory of metric-space gradient flows has been developed [AGS05].

In fact it is very common that modelling derivations are not rigorous, and often would even be very hard to make rigorous. The difficulty usually lies in the assumed regularity of the components of the model: if the solutions are regular, then the arguments are rigorous. Of course there are several well-known examples in which this approach fails to produce complete models; for instance, in the case of weak solutions of scalar conservation laws, uniqueness of such solutions only holds under additional entropy conditions. One should think of these model derivations as thought experiments, whose outcome should later still be checked for mathematical consistency.

The process space has redundant information. The field w contains more information than $\dot{c}$, since adding a divergence-free vector field to cw does not change $\dot{c}$. This is not a problem; the minimization (2.7) deals with this redundancy in a natural way. In the context of Wasserstein gradient flows this phenomenon is well known, where it leads to the characterization of the Wasserstein tangent in terms of gradients (see e.g. [Ott01] or [AGS05, Th. 9.3.2]).

Such redundancies are very common, and they arise from the fact that it is often convenient to characterize the state space and the space of processes in different ways. We will see more examples of this below.

Chapter 3

Gradient-flow facts

3.1 Forces and velocities, tangents and cotangents

Variational modelling is helped by adopting the geometric terminology of tangents and cotangents, even though we do this in a non-rigorous manner.

Let $\mathcal{Z}$ be a subset of a linear space. The space of tangent vectors at a point $z \in \mathcal{Z}$, noted $T_z\mathcal{Z}$, is the set of all derivatives $\dot{\boldsymbol{z}}(0)$ of smooth curves $\boldsymbol{z} : (-1, 1) \to \mathcal{Z}$ such that $\boldsymbol{z}(0) = z$. The space of cotangent vectors, noted $T_z^*\mathcal{Z}$, is the set of all linear functionals on $T_z\mathcal{Z}$. The set of all pairs (z, τ) where $z \in \mathcal{Z}$ and $\tau \in T_z\mathcal{Z}$ is called the *tangent bundle* and is denoted $T\mathcal{Z}$, and similarly the set of all points and cotangent vectors is written $T^*\mathcal{Z}$. If $\tau \in T_z\mathcal{Z}$ and $\xi \in T_z^*\mathcal{Z}$, then we write $\langle \xi, \tau \rangle = \langle \tau, \xi \rangle \in \mathbb{R}$ for the value of ξ applied to τ.

In mechanical terminology, the space of tangents is the space of *generalized velocities* or *infinitesimal displacements*. The cotangent space can be identified with *generalized forces*, if one gives $T\mathcal{Z}$ and $T^*\mathcal{Z}$ dimensions such that $\langle \xi, \tau \rangle$ has dimensions of energy per unit time (power); then the duality is sometimes called a *work couple*. The classical example is when $\mathcal{Z} = \mathbb{R}$ consists of distances, measured in m, and $T^*\mathcal{Z}$ has dimensions of force [N]; then τ is a velocity [m/s] and $\langle \xi, \tau \rangle$ has dimensions Nm/s. But many other combinations are possible, such as when z is an angle, τ is a rotational velocity [1/s], and ξ is a torque [Nm].

This division into tangents and cotangents is useful from many points of view. Maybe the most important is that correct use of this structure guarantees coordinate invariance, as is the case in Riemannian geometry.

If $E : \mathcal{Z} \to \mathbb{R}$ has dimensions of energy, then the derivative of E with respect to z is a cotangent vector with the dimensions of a generalized force; for any $\tau \in T_z\mathcal{Z}$, $\langle E'(z), \tau \rangle$ has dimensions of power, and if $z : (-1, 1) \to \mathcal{Z}$, then $\partial_t E(z(t))|_{t=0} = \langle E'(z(0)), \dot{z}(0) \rangle$. This derivative is mathematically the Fréchet derivative.

The operators $\mathcal{K}$ and $\mathcal{G}$ that we mentioned in Section 1.2 are examples of *duality maps*, that map tangents to cotangents or vice versa: for each $z \in \mathcal{Z}$, $\mathcal{G}(z)$ is an operator that maps $T_z\mathcal{Z}$ to $T_z^*\mathcal{Z}$, and similarly $\mathcal{K}(z)$ maps $T_z^*\mathcal{Z}$ to $T_z\mathcal{Z}$. (Heed the notation: for $z \in \mathcal{Z}$,

$\mathcal{G}(z)$ is an operator that takes a tangent $\tau \in T_z\mathcal{Z}$ and maps it to a cotangent vector written as $\mathcal{G}(z)\tau$.) Because of this property, they give rise to bilinear forms on $T_z\mathcal{Z}$ and $T_z^*\mathcal{Z}$, respectively:

$$(\tau_1, \tau_2)_{\mathcal{G}, z} := \langle \tau_1, \mathcal{G}(z)\tau_2 \rangle$$
$$(\xi_1, \xi_2)_{\mathcal{K}, z} := \langle \xi_1, \mathcal{K}(z)\xi_2 \rangle.$$

These bilinear forms are symmetric if $\mathcal{G}$ and $\mathcal{K}$ are symmetric, in the sense that $\langle \tau_1, \mathcal{G}(z)\tau_2 \rangle = \langle \tau_2, \mathcal{G}(z)\tau_1 \rangle$.

Such duality maps are necessary in equations such as (1.1), when one wishes to equate $\dot{z}$ with the derivative of an energy, since these two objects live in different spaces. The duality maps implement the mapping from one to the other.

3.2 Gradient flows

We now revisit the informal definition of a gradient flow that we gave in Section 1.2. The starting point is to think of gradient flows as a set of equations of the form

$$\dot{z} = -\mathcal{K}(z)\mathcal{F}'(z) \qquad \text{or} \tag{3.1a}$$
$$\mathcal{G}(z)\dot{z} = -\mathcal{F}'(z). \tag{3.1b}$$

As we described earlier, the interpretation of $\mathcal{G}$ and $\mathcal{K}$ is as a characterization of dissipation, through the energy-dissipation identity

$$\frac{d}{dt}\mathcal{F}(z(t)) = \langle \mathcal{F}', \dot{z} \rangle = -\langle \mathcal{F}', \mathcal{K}(z)\mathcal{F}' \rangle = -\langle \dot{z}, \mathcal{G}(z)\dot{z} \rangle. \tag{3.2}$$

However, we need to allow for several generalizations:

1. The duality operators $\mathcal{K}$ and $\mathcal{G}$ can be degenerate or infinite; for instance, certain changes of state, i.e. certain tangent vectors, might not lead to dissipation of energy, implying that $\mathcal{G}$ vanishes in those directions; similarly, certain directions might be inadmissible, implying that '$\mathcal{G} = \infty$' in those directions. By duality, these two possibilities translate into $\mathcal{K}$ being 'infinity' in the first case and 'zero' in the second, in the corresponding dual directions.

2. The duality operators might not be linear in their arguments; a typical example of this is a rate-independent system, in which case $\mathcal{G}(z)(\lambda s) = \mathcal{G}(z)s$ for all $\lambda > 0$.

3. As we saw in Sections 2.2 and 2.3, it is natural to model changes of state through a auxiliary variables such as flow fields. This suggests to define dissipation in terms of a *process space*, which can be different than the tangent space.

In the next section we discuss how to deal with point 3. To deal with 1 and 2, we redefine the concept the concept of a gradient flow in terms of a pair of dissipation functionals

$\Psi : T\mathcal{Z} \to \mathbb{R}$ and $\Psi^* : T^*\mathcal{Z} \to \mathbb{R}$, which are assumed to be dual in the sense that for all $z \in \mathcal{Z}$,

$$\Psi^*(z, \xi) = \sup_{s \in T_z\mathcal{Z}} \big[\langle \xi, s \rangle - \Psi(z, s)\big]$$
$$\Psi(z, s) = \sup_{\xi \in T_z^*\mathcal{Z}} \big[\langle \xi, s \rangle - \Psi^*(z, \xi)\big].$$

Because of this duality, both Ψ and Ψ^* are convex in their second argument, and we have the inequality

$$\Psi(z, s) + \Psi^*(z, \xi) \geq \langle \xi, s \rangle \qquad \text{for all } s \in T_z\mathcal{Z} \text{ and } \xi \in T_z^*\mathcal{Z},$$

with equality if and only if $s \in \partial_\xi \Psi^*(z, \xi) \iff \xi \in \partial_s \Psi(z, s)$.

Given (linear) duality operators $\mathcal{G}$ and $\mathcal{K}$, we define corresponding (quadratic) potentials Ψ and Ψ^* by $\Psi(z, s) := \frac{1}{2}\langle s, \mathcal{G}(z)s \rangle$ and $\Psi^*(z, \xi) := \frac{1}{2}\langle \xi, \mathcal{K}(z)\xi \rangle$. The fact that these are a dual pair can be seen as follows:

$$
\begin{aligned}
\Psi^*(z, \xi) &= \sup_{s \in T_z\mathcal{Z}} \big[\langle \xi, s \rangle - \Psi(z, s)\big] \\
&= \sup_{s \in T_z\mathcal{Z}} \big[\langle \xi, s \rangle - \tfrac{1}{2}\langle s, \mathcal{G}(z)s \rangle\big] \\
&= \sup_{\sigma \in T_z^*\mathcal{Z}} \big[\langle \xi, \mathcal{K}(z)\sigma \rangle - \tfrac{1}{2}\langle \mathcal{K}(z)\sigma, \sigma \rangle\big] \qquad \text{by substituting } s = \mathcal{K}(z)\sigma \\
&= \tfrac{1}{2}\langle \mathcal{K}(z)\xi, \xi \rangle + \sup_{\sigma \in T_z^*\mathcal{Z}} -\tfrac{1}{2}\langle \mathcal{K}(z)(\xi - \sigma), \xi - \sigma \rangle = \tfrac{1}{2}\langle \mathcal{K}(z)\xi, \xi \rangle.
\end{aligned}
\tag{3.3}
$$

For this case the identity conditions $s \in \partial_\xi \Psi^*(x, \xi) \iff \xi \in \partial_s \Psi(z, s)$ simplify to $s = \mathcal{K}(z)\xi \iff \xi = \mathcal{G}(z)s$.

We now give a defintion of a generalized gradient flow that also applies to the case of non-quadratic potentials Ψ and Ψ^*.

Definition 3.4 (Global formulation). *A function $z : [0, T] \to \mathcal{Z}$ is a solution of the generalized gradient flow associated with $(\mathcal{Z}, \mathcal{F}, \Psi)$ if it satisfies the inequality*

$$\mathcal{F}(z(T)) - \mathcal{F}(z(0)) + \int_0^T \Big[\Psi\big(z(t), \dot{z}(t)\big) + \Psi^*\big(z(t), -\mathcal{F}'(z(t))\big)\Big] dt \leq 0. \tag{3.5}$$

This definition extends the previous ones:

Lemma 3.6. *Assume that $\mathcal{Z}$ is a Hilbert space. Let $\mathcal{F} : \mathcal{Z} \to \mathbb{R}$ be smooth, and assume that for each $z \in \mathcal{Z}$ the duality operators $\mathcal{G}(z)$ and $\mathcal{K}(z)$ are invertible with $\mathcal{G}(z) = \mathcal{K}(z)^{-1}$. Define $\Psi(z, s) := \frac{1}{2}\langle s, \mathcal{G}(z)s \rangle$ and $\Psi^*(z, \xi) := \frac{1}{2}\langle \xi, \mathcal{K}(z)\xi \rangle$. Let $z : [0, T] \to \mathcal{Z}$ be smooth. Then the following four properties are equivalent:*

    1. $\dot{z} = -\mathcal{K}(z)\mathcal{F}'(z)$ for all $t \in [0, T]$ (i.e., z satisfies (3.1a));

18

2. $\mathcal{G}(z)\dot{z} = -\mathcal{F}'(z)$ *for all* $t \in [0, T]$ *(i.e., z satisfies (3.1b))*;

3. *z satisfies (3.5);*

4. *At each time t, and for given $z(t)$, $\dot{z}(t)$ is given by the property that it minimizes* $\Psi(z(t), \dot{z}) + \langle \mathcal{F}'(z(t)), \dot{z} \rangle$.

Note that the conditions of this lemma exclude all but the most trivial of examples, since in nearly all cases $\mathcal{F}$ will not be smooth with respect to the topology generated by $\mathcal{G}$; this is exactly the origin of the parabolic behaviour that infinite-dimensional gradient flows typically show. Therefore either $\mathcal{F}$ will not be smooth in the chosen topology, or $\mathcal{G}$ will not be invertible with respect to that topology.

The reason for this lemma therefore is not to apply it directly, but to provide a case in which the usually formal connections can be made rigorous.

Proof. The equivalence (3.1a) $\iff$ (3.1b) is immediate. To show (3.1) $\iff$ (3.5), note that for *any* smooth curve $z : [0, T] \to \mathcal{Z}$ we have

$$\frac{d}{dt}\mathcal{F}(z(t)) = \langle \mathcal{F}'(z(t)), \dot{z}(t) \rangle \overset{(*)}{\geq} -\Psi\big(z(t), \dot{z}(t)\big) - \Psi^*\big(z(t), -\mathcal{F}'(z(t))\big).$$

Therefore inequality (3.5) holds with the *opposite* inequality for any curve z; imposing the inequality (3.5) is equivalent to imposing equality, which in turn is equivalent to imposing equality in $(*)$ above; this in turn is equivalent to $\dot{z} = \mathcal{K}(z)(-\mathcal{F}'(z))$. This proves (3.1) $\iff$ (3.5).

Finally, the equation $\mathcal{G}(z)\dot{z} = -\mathcal{F}'(z)$ is the stationarity condition for the minimization problem $\min_s \Psi(z, s) + \langle \mathcal{F}'(z), s \rangle$, and therefore part 4 is equivalent with (3.1b). $\square$

Part 4 is so important that we give it a name:

Definition 3.7 (Local formulation)**.** *A function $z : [0, T] \to \mathcal{Z}$ is a* local *solution at time t of the generalized gradient flow associated with $(\mathcal{Z}, \mathcal{F}, \Psi)$ if $\dot{z}(t)$ has the property*

$$\dot{z}(t) \text{ solves } \quad \min_s \Psi(z(t), s) + \langle \mathcal{F}'(z(t)), s \rangle. \tag{3.8}$$

3.3 Tangents and processes

We now make the second generalization of the gradient-flow concept, in order to deal with point 3 on page 17.

A simplistic reformulation of the variational modelling method is '(a) choose the state space and the energy on this space, and (b) choose how the state can change, and how much energy that costs'. However, the requirements on the state space are different than the requirements on the changes of state: the state space needs to be rich enough to define the energy $\mathcal{F}$ on this space, and the description of the potential changes of state needs to be rich enough to define the dissipation.

In practice, from a modelling point of view, the state changes as the result of physical *processes*, which need not be related to the states themselves in a one-to-one manner. It therefore makes sense to define a separate *space of processes* $P_z\mathcal{Z}$. For instance,

19

- The state space could be that of the volume $z = V > 0$ of fluid in a container, and this volume might change because of an inflow $f_i > 0$ and an outflow $f_o > 0$. Then the process space is
$$P_V \mathcal{Z} = \{(f_o, f_i) \in \mathbb{R}^2\},$$
and the tangent $\dot{V}$ is given in terms of (f_o, f_i) by $\dot{V} = f_i - f_o$.

- In the example of diffusion of solutes (Section 2.3), $z = c$, the concentration field, and the tangent space is the space of derivatives $\dot{c}$ of smooth curves, which one can formally identify with the subset of $L^1(\Omega)$ with zero integral; the process space is
$$P_c \mathcal{Z} = \{w : \Omega \to \mathbb{R}^3 : w \cdot n = 0 \text{ on } \partial\Omega\}$$

Given $w \in P_c \mathcal{Z}$, the corresponding tangent is given by $\dot{c} = -\operatorname{div} cw$.

The minimization problem (3.8) provides a natural way to formulate the dissipation in terms of the process space. Assume that we have chosen a state space $\mathcal{Z}$, a functional $\mathcal{F}$, a process space $P_z \mathcal{Z}$, and a dissipation potential $\Psi(z, w)$ defined on pairs (z, w) of $z \in \mathcal{Z}$ and $w \in P_z \mathcal{Z}$. Assume that for each $z \in \mathcal{Z}$ an operator $\mathcal{P}(z) : P_z \mathcal{Z} \to T_z \mathcal{Z}$ is given that maps the process space to the tangent space, such that for each $z \in \mathcal{Z}$ and $w \in P_z \mathcal{Z}$, $\mathcal{P}(z)w$ is the corresponding tangent vector. (In the examples above, for the container volume $\mathcal{P}(V)(f_o, f_i) = f_i - f_o$, and for the solute diffusion $\mathcal{P}(c)w = -\operatorname{div} cw$.) We can then define a dissipation potential $\widetilde{\Psi}$ on the tangent space as
$$\widetilde{\Psi}(z, s) := \inf_w \{\Psi(z, w) : s = \mathcal{P}(z)w\}.$$

This dissipation potential $\widetilde{\Psi}(z, s)$ corresponds to selecting the best process w that gives rise to the tangent vector s: that is, 'best' in the sense of having the smallest dissipation potential.

Both the global Definition 3.4 and the local Definition 3.7 can now be applied to $\widetilde{\Psi}$. Note that the local definition (3.8) can be rewritten as
$$\dot{z} \in \operatorname*{argmin}_{s \in T_z \mathcal{Z}} \widetilde{\Psi}(z, s) + \langle \mathcal{F}'(z), s \rangle$$
$$\iff \dot{z} \in \operatorname*{argmin}_{s \in T_z \mathcal{Z}} \min_{w \in P^{-1}s} \Psi(z, w) + \langle \mathcal{F}'(z), Pw \rangle$$
$$\iff w \in \operatorname*{argmin}_{w \in P_z \mathcal{Z}} \Psi(z, w) + \langle \mathcal{F}'(z), Pw \rangle \quad \text{and } \dot{z} = \mathcal{P}(z)w.$$

We summarize this discussion in the form that we use it throughout these notes:

If $P_z \mathcal{Z}$ is the process space at $z \in \mathcal{Z}$, with dissipation $\Psi(z, w)$, and $\mathcal{P} : P\mathcal{Z} \to T\mathcal{Z}$ maps processes onto tangents, then we can find the evolution equations by solving the minimization problem
$$\min_{w \in P_z \mathcal{Z}} \Psi(z, w) + \langle \mathcal{F}'(z), \mathcal{P}(z)w \rangle.$$
Then $\dot{z}$ is given by the projected minimizer $\mathcal{P}(z)w$.

3.4 Wasserstein distance and Wasserstein gradient flows

The example of Section 2.3 is also known as a *Wasserstein gradient flow*, and this class of gradient flows is central in these notes.

3.4.1 Wasserstein gradient flows

Wasserstein gradient flows are gradient flows in the state space $\mathcal{Z} = \mathcal{M}_2$ of non-negative finite-mass measures with finite second moments, i.e. the space

$$\mathcal{M}_2(\mathbb{R}^d) := \left\{ \rho \in \mathcal{M}(\mathbb{R}^d) : \rho \geq 0 \text{ and } \int_{\mathbb{R}^d} (1 + |x|^2)\, \rho(dx) < \infty \right\}.$$

See Appendix A for some background on measure-theoretical concepts and notation.

We define the gradient flow by defining the operator $\mathcal{K}$. For $\rho \in \mathcal{M}_2(\mathbb{R}^d)$ and a function $\xi : \mathbb{R}^d \to \mathbb{R}$ with weak derivative $\nabla \xi \in L^2(\rho; \mathbb{R}^d)$, let $\mathcal{K}(\rho)\xi$ be the distribution

$$\mathcal{K}(\rho)\xi := -\operatorname{div} \rho \nabla \xi.$$

We interpret this as an operator on the cotangent space of $\mathcal{Z}$ by applying it not to the Fréchet derivative $\mathcal{F}'$ but to the variational derivative $\mathrm{D}\mathcal{F}$, defined by

$$\int_{\mathbb{R}^d} \mathrm{D}\mathcal{F}(\rho)(x) f(x)\, dx = \langle \mathcal{F}'(\rho), f \rangle = \lim_{h \to 0} \frac{1}{h} \big[\mathcal{F}(\rho + hf) - \mathcal{F}(\rho) \big] \qquad \text{for all } f.$$

The Wasserstein gradient flow of a functional $\mathcal{F}$ therefore is the evolution equation

$$\dot{\rho} = \operatorname{div} \rho \nabla \mathrm{D}\mathcal{F}(\rho).$$

Many well-known diffusive partial differential equations have the structure of a gradient flow with respect to the Wasserstein metric or a related metric. These are a few examples of energies and corresponding equations:

$$\mathcal{F}(\rho) = \int [\rho(x) \log \rho(x) + \rho(x)V(x)]\, dx \qquad\qquad \partial_t \rho = \Delta \rho + \operatorname{div} \rho \nabla V$$

$$\mathcal{F}(\rho) = \int [U(\rho(x)) + \rho(x)V(x)]\, dx + \iint \rho(x)\rho(y)W(x-y)\, dxdy \qquad \partial_t \rho = \operatorname{div} \rho \nabla \big[U'(\rho) + V + W * \rho \big]$$

$$\mathcal{F}(\rho) = \int |\nabla \rho|^2(x)\, dx \qquad\qquad \partial_t \rho = -\operatorname{div} \rho \nabla \Delta \rho$$

3.4.2 The Wasserstein metric tensor

As we discussed in Sections 1.2 and 3.2, the operator $\mathcal{K}$ gives rise to an inner product, which is actually a weighted Sobolev inner product,

$$(\xi_1, \xi_2)_{1,\rho} := \int_{\mathbb{R}^d} \xi_1 \mathcal{K}(\rho)\xi_2 = \int_{\mathbb{R}^d} \nabla \xi_1 \nabla \xi_2\, d\rho,$$

with corresponding norm

$$\|\xi\|_{1,\rho}^2 = \int_{\mathbb{R}^d} |\nabla \xi|^2 \, d\rho. \tag{3.9}$$

In Wasserstein terminology this is called the *dual* norm. The corresponding *primal* Wasserstein norm $\| \cdot \|_{-1,\rho}$ is defined for any distribution $s \in \mathcal{D}'(\mathbb{R}^d)$,

$$\|s\|_{-1,\rho}^2 := \inf_v \left\{ \int_{\mathbb{R}^d} |v|^2 \, d\rho : v \in L^2(\rho; \mathbb{R}^d), \ s + \operatorname{div} \rho v = 0 \text{ in } \mathcal{D}'(\mathbb{R}^d) \right\}. \tag{3.10}$$

If s has no representation as $\operatorname{div} \rho v$ for any $v \in L^2(\rho_t)$, then the norm is given the value $+\infty$.

Note how definition (3.10) contains the same expression as the dissipation potential $\widetilde{\Psi}$ in (2.6).

The infimum in (3.10) is achieved, and the corresponding vector field v is an element of the set

$$\overline{\{\nabla \varphi : \varphi \in C_c^\infty(\mathbb{R}^d)\}}^{L^2(\rho)}. \tag{3.11}$$

The distribution s and the corresponding velocity field v are two different ways of describing the same tangent vector. Introducing the notation

$$(s, v) \in \operatorname{Tan}_\rho \qquad \Longleftrightarrow \qquad s + \operatorname{div} \rho v = 0 \quad \text{and} \quad v \in \overline{\{\nabla \varphi : \varphi \in C_c^\infty(\mathbb{R}^d)\}}^{L^2(\rho)},$$

this optimal vector field allows us to define a local metric tensor on the tangent space, the corresponding local inner product:

$$(s_1, s_2)_{-1,\rho_t} := \int_{\mathbb{R}^d} v_1 \cdot v_2 \, d\rho_t \qquad \text{where } (s_i, v_i) \in \operatorname{Tan}_{\rho_t}, \tag{3.12}$$

such that the norm (3.10) satisfies $\|s\|_{-1,\rho}^2 = (s, s)_{-1,\rho}$.

A consequence of the construction above is that $\mathcal{K}\mathrm{D}\mathcal{F}$ can be viewed as the *Wasserstein gradient* of $\mathcal{F}$, in the sense that along a curve $t \mapsto \rho_t$,

$$\partial_t \mathcal{F}(\rho_t) = \langle \mathcal{F}'(\rho_t), \dot{\rho}_t \rangle = \big(\mathcal{K}\mathrm{D}\mathcal{F}(\rho_t), \dot{\rho}_t\big)_{-1,\rho_t}. \tag{3.13}$$

3.4.3 Formulation in terms of a process space

The definitions above can be simply summarized in terms of the process-space concept introduced above, as follows:

State space: $\qquad \mathcal{Z} = \mathcal{M}_2(\mathbb{R}^d)$

Process space : $\qquad P_\rho \mathcal{Z} = \overline{\{\nabla \varphi : \varphi \in C_c^\infty(\mathbb{R}^d)\}}^{L^2(\rho)}$

Mapping $\mathcal{P}$: $\qquad \mathcal{P}(\rho)w = -\operatorname{div} \rho w$

Tangent space : $\qquad T_\rho \mathcal{Z} = \{\mathcal{P}(\rho)w : w \in P_\rho \mathcal{Z}\} = \{-\operatorname{div} \rho w : w \in P_\rho \mathcal{Z}\}$

Cotangent space: $\qquad T_\rho^* \mathcal{Z} = \overline{C_c^\infty(\mathbb{R}^d)}^{\|\cdot\|_{1,\rho}}.$

The Wasserstein metric tensor is indeed a good example of how the process space gives a representation of the tangent space, through the operator $\mathcal{P}$, in terms of which the dissipation is more directly characterized.

3.4.4 The Wasserstein distance

Based on the norm $\|\cdot\|_{-1,\rho}$, the *Wasserstein distance* between two measures of equal mass can then be defined as the infimum of the Wasserstein norm integrated along curves [BB00],

$$W_2(\rho_0, \rho_1)^2 = \inf\left\{\int_0^1 \|\dot{\rho}_t\|_{-1,\rho_t}^2\, dt : \rho : [0,1] \to \mathcal{M}_2(\mathbb{R}^d),\ \rho_t\big|_{t=0,1} = \rho_{0,1}\right\}. \tag{3.14}$$

The distance function W_2 is a complete separable metric on each subspace of $\mathcal{M}_2(\mathbb{R}^d)$ of functions of a given mass; the distance between measures of different mass is not defined, since the condition $\dot{\rho}_t + \operatorname{div}\rho_t v_t = 0$ implies that the total mass of ρ_t is independent of t. Convergence in Wasserstein metric is equivalent to the combination of (a) narrow convergence of measures (i.e. in duality with continuous and bounded functions) and (b) convergence of the second moments [AGS05, Remark 8.1.5-7].

3.4.5 The Wasserstein distance for empirical measures

For the special case of empirical measures the Wasserstein norms and distance take a particularly simple form. This case also explains some of the relevance of this metric for modelling.

Let $x_i : [0,1] \to \mathbb{R}^d$, $i = 1, \ldots n$ be n time-courses of positions in $\mathbb{R}^d$, and define the time-parametrized empirical measure $\rho_t := \frac{1}{n}\sum_{i=1}^n \delta_{x_i(t)}$. Then the local Wasserstein norm is

$$\|\dot{\rho}_t\|_{-1,\rho_t}^2 = \frac{1}{n}\sum_{i=1}^n \dot{x}_i(t)^2. \tag{3.15}$$

If $n = 1$, for instance, then the Wasserstein norm is the same as the Euclidean norm of the velocity $\dot{x}$. For general n, the Wasserstein norm is proportional to the Euclidean norm of the velocity vector $(\dot{x}_1, \ldots, \dot{x}_n) \in \mathbb{R}^{nd}$. As n becomes large, the Wasserstein norm can be seen as a version of the Euclidean norm of the n-particle velocity that remains meaningful in the limit $n \to \infty$.

Turning to the distance W_2, it follows from the discussion above that for single delta functions W_2 reproduces the Euclidean distance:

$$W_2(\delta_x, \delta_y) = |x - y|.$$

For multiple delta functions the situation is similar; however, the infimum in the definition translates into a combinatorial optimization:

$$W_2\left(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}, \frac{1}{n}\sum_{j=1}^n \delta_{y_j}\right)^2 = \inf_{\sigma \in S_n} \frac{1}{n}\sum_{i=1}^n |x_i - y_{\sigma(i)}|^2.$$

In constructing curves connecting the two empirical measures, there is freedom in the matching between initial and final points; this matching results in the minimization over the set of permuations S_n of n objects.

3.5   Comments

Characterization of tangent and cotangent spaces.   Our introduction of tangent and cotangent vectors implicitly assumed that $\mathcal{Z}$ is a space with a differentiable structure, in which time derivatives $\dot{z}$ and state-space derivatives E' can be defined in a meaningful way. In addition, our implicit definition of the cotangent space as the topological dual of the tangent space requires a topology on the tangent space, which we didn't define. In many cases of importance, including the important example of the Wasserstein metric space, these aspects are non-trivial, and appropriate constructions and generalizations are necessary; the book by Ambrosio, Gigli and Savaré [AGS05] is an example of such generalizations in the case of gradient flows. Nonetheless, we feel that the formal differentiable way of writing provides the right intuition, and therefore we maintain this way of writing the system.

Chapter 4

Entropy and free energy, in stationary situations

In this chapter I focus on the concepts of *entropy*, *free energy*, and *large-deviation theory*, and their interconnections. The main modelling insights that I want to establish are

M1 *Entropy is best understood as the rate functional of the empirical measure of a large number of identical stochastic particles; it describes the probability distribution of that system* (Section 4.4);

M2 *Entropy arises from the indistinguishability of the particles* (Section 4.2);

M3 *Free energy arises from the tilting of a particle system by exchanging energy with a heat bath* (Section 4.5).

These insights imply the following modelling guidelines:

M4 Any *concentration* c (in moles per m³) contributes an entropy $S(c)$ to the free energy of the system. For particles without interaction or with weak interaction the entropy is $S(\rho) = -R \int c \log(c/c_0)$ (Section 4.8).

M5 The free energy of a system with energy E and concentrations $c_1, \ldots, c_n$ is $E - \sum_i TS(c_i)$ (Section 4.8).

M6 The equilibrium states of such a system are the global minimizers of the free energy (Section 4.5).

We start by introducing entropy (Sections 4.1 and 4.2) and large-deviation theory (Section 4.3).

4.1 Entropy

What is entropy? This question has been asked an unimaginable number of times, and has received a wide variety of answers. Here I will not try to summarize the literature, but only mention that I personally like the treatments in [LY97, Eva01].

Instead we define one version of entropy, the *relative* entropy of two probability measures. As always we assume we are working in a complete metrizable separable space Ω with a σ-algebra Σ that contains the Borel sets (see Appendix A for the concepts used here).

Definition 4.1. *Let $\mu, \nu \in \mathcal{M}(\Omega)$, with $\mu, \nu \geq 0$. The* relative entropy *of μ with respect to ν is*

$$\mathcal{H}(\mu|\nu) := \begin{cases} \displaystyle\int_\Omega f \log f \, \mathrm{d}\nu & \text{if } \mu \ll \nu \text{ and } f = \dfrac{\mathrm{d}\mu}{\mathrm{d}\nu} \\ +\infty & \text{otherwise.} \end{cases}$$

A very useful special case will be the relative entropy with respect to the Lebesgue measure: for $\rho \in \mathscr{P}(\Omega)$,

$$\mathsf{Ent}(\rho) := \mathcal{H}(\rho|\mathcal{L}^d), \tag{4.2}$$

where $\mathcal{L}^d$ is the Lebesgue measure on $\mathbb{R}^d$.

Before we discuss the interpretation of this object in the next section, we first mention a number of properties.

Theorem 4.3.   *1. If $\mu(\Omega) = \nu(\Omega)$, then $\mathcal{H}(\mu|\nu) \geq 0$, and $\mathcal{H}(\mu|\nu) = 0$ if and only if $\mu = \nu$;*

   2. If $\mu(\Omega) = \nu(\Omega)$, then $2\|\mu - \nu\|_{TV}^2 \leq \mathcal{H}(\mu|\nu)$ (Csiszár-Kullback-Pinsker inequality);

   3. $\mathcal{H}$ is invariant under transformations of the underlying space, i.e. if $\varphi : \Omega \to \Omega$ is a one-to-one measurable mapping, and $\varphi_\#\mu$ and $\varphi_\#\nu$ are the push-forwards of μ and ν, then $\mathcal{H}(\mu|\nu) = \mathcal{H}(\varphi_\#\mu|\varphi_\#\nu)$.

Proof. The first part of the theorem can be understood by writing $\mathcal{H}$ as

$$\mathcal{H}(\mu|\nu) = \int (f \log f - f + 1) \, d\nu \qquad \text{if } f = \frac{\mathrm{d}\mu}{\mathrm{d}\nu},$$

and using the fact that $g(s) = s \log s - s + 1$ is non-negative and only zero at $s = 1$. For the Csiszár-Kullback-Pinsker inequality we refer to [Csi67, Kul67].

To prove the invariance under transformations, note that for all $\omega \in \Omega$,

$$\frac{\mathrm{d}\varphi_\#\mu}{\mathrm{d}\varphi_\#\nu}(\varphi(\omega)) = \frac{\mathrm{d}\mu}{\mathrm{d}\nu}(\omega),$$

so that

$$\mathcal{H}(\mu|\nu) = \int \log \frac{\mathrm{d}\mu}{\mathrm{d}\nu}(\omega) \, \mu(d\omega) = \int \log \frac{\mathrm{d}\varphi_\#\mu}{\mathrm{d}\varphi_\#\nu}(\varphi(\omega)) \, \mu(d\omega) =$$

$$= \int \log \frac{\mathrm{d}\varphi_\#\mu}{\mathrm{d}\varphi_\#\nu}(\omega') \, \varphi_\#\mu(d\omega') = \mathcal{H}(\varphi_\#\mu|\varphi_\#\nu).$$

$\square$

The properties in this theorem are relevant for the central role that this relative entropy plays. The non-negativity and Csiszár-Kullback-Pinsker inequality show that when μ and ν have the same mass (e.g. if $\mu, \nu \in \mathscr{P}(\Omega)$) then $\mathcal{H}$ acts like a measure of distance between μ and ν. It's not a distance function, since it is not symmetric ($\mathcal{H}(\mu|\nu) \neq \mathcal{H}(\nu|\mu)$), but via the Csiszár-Kullback-Pinsker inequality it does generate the topology of total variation on the space of probability measures.

The fact that $\mathcal{H}$ is invariant under transformations of space is essential for the modelling, for the following reason. Every modelling process involves a choice of coordinates, and this choice can often be made in many different ways. Nevertheless, if we believe that there is a well-defined energy that drives the evolution, then the value of this energy should not depend on which set of coordinates we have chosen to describe it in.

Note that Ent is *not* invariant under change of coordinates. This implies that this functional corresponds to a specific choice of coordinates.

In part 3 of the theorem above, if φ is a general mapping from Ω to a set Ω', then we still have an inequality: $\mathcal{H}(\varphi_{\#}\mu|\varphi_{\#}\nu) \leq \mathcal{H}(\mu|\nu)$. In fact, the discrepancy between the left- and right-hand sides of this inequality can be fully characterized by a *tensorization identity*

$$\mathcal{H}(\varphi_{\#}\mu|\varphi_{\#}\nu) = \mathcal{H}(\mu|\nu) + \int_{\Omega'} \mathcal{H}(\mu_y|\nu_y)\, \varphi_{\#}\mu(dy).$$

Here μ_y and ν_y are defined through the disintegration theorem [AGS05, Th. 5.3.1].

4.2 Entropy as measure of degeneracy

This mathematical definition of the relative entropy $\mathcal{H}$ above does not explain why it might appear in a model. One way to give an interpretation to the relative entropy $\mathcal{H}$ is by a counting argument, that we explain here for the case of a finite state space. It involves *microstates* and *macrostates*, where multiple microstates correspond to a single macrostate. The result will be that the entropy characterizes two concepts: one is the degree of degeneracy, that is the number of microstates that corresponds to a given macrostate, and the other is the probability of observing a macrostate, given a probability distribution over microstates. The two are closely connected.

Take a finite state space I consisting of $|I|$ elements. If $\mu \in \mathscr{P}(I)$, then μ is described by $|I|$ numbers μ_i, and the relative entropy with respect to μ is

$$\mathcal{H}(\rho|\mu) = \sum_{i \in I} \rho_i \log \frac{\rho_i}{\mu_i}, \qquad \text{for } \rho \in \mathscr{P}(I).$$

Consider N particles on the lattice described by I, i.e. consider a mapping $x : \{1, \ldots, N\} \to I$. We think of x as the *microstate*. Define an *empirical measure* $\rho \in \mathscr{P}(I)$ by

$$k_i := \#\big\{j \in \{1, \ldots, N\} : x(j) = i\big\} \quad \text{and} \quad \rho_i := \frac{k_i}{N}, \qquad \text{for } i \in I. \tag{4.4}$$

In going from x to ρ there is loss of information; multiple mappings x produce the same empirical measure ρ. The degree of degeneracy, the number of unique mappings x that

correspond to a given ρ, is $N! \left(\prod_{i \in I} k_i!\right)^{-1}$. The fact that the particles are identical, *indistinguishable*, is important here—this is required for the description in terms of integers k_i. Because of this loss of information, we think of ρ as the *macrostate*.

We now determine the behaviour of this 'degree of degeneracy' in the limit $N \to \infty$. Using Stirling's formula in the form

$$\log n! = n \log n - n + o(n) \qquad \text{as } n \to \infty,$$

we estimate

$$\log N! \left(\prod_{i \in I} k_i!\right)^{-1} = \log N! - \sum_{i \in I} \log k_i!$$

$$= N \log N - N - \sum_{i \in I} (k_i \log k_i - k_i) + o(N)$$

$$= -N \sum_{i \in I} \rho_i \log \rho_i + o(N) \qquad \text{as } N \to \infty.$$

One interpretation of the relative entropy therefore is as follows. Take for the moment μ_i to be the uniform measure, i.e. $\mu_i = |I|^{-1}$; then

$$\mathcal{H}(\rho|\mu) = \sum_{i \in I} \rho_i \log \rho_i + \log |I|.$$

Then

$$\mathcal{H}(\rho|\mu) = -\lim_{N \to \infty} \frac{1}{N} \log \#\text{realizations of } \rho \ + \ \log |I|. \tag{4.5}$$

This shows that if the number of microscopic realizations x of the macroscopic object ρ is large, then $\mathcal{H}(\rho|\mu)$ is small, and vice versa.[1] This is the interpretation in terms of a counting argument.

We now switch to the probabilistic point of view. If we allocate particles at random with the same, independent, probability for each microstate x, then the probability of obtaining each microstate is $|I|^{-N}$, and the probability of a macrostate ρ satisfies

$$\log \text{Prob}(\rho) = \log |I|^{-N} N! \left(\prod_{i \in I} k_i!\right)^{-1} = -N \mathcal{H}(\rho|\mu) + o(N) \qquad \text{as } N \to \infty.$$

We can do the same with non-equal probabilties: we place each particle at an $i \in I$ with probability μ_i. Then the probability of a microstate x is

$$\prod_{j=1}^{N} \mu_{x(j)},$$

---

[1]There is a problem here, though; the ρ in the right-hand side can not be independent of N, since it is a vector with components of the form k/N, while the ρ in the left-hand side expression can not depend on N. This contradiction will be resolved when we discuss large deviations in the next section.

and now the probability of a macrostate ρ satisfies

$$
\begin{aligned}
\log \operatorname{Prob}(\rho) &= \log \Big(\prod_{j=1}^{N} \mu_{x(j)}\Big) N! \Big(\prod_{i \in I} k_i!\Big)^{-1} \\
&= \sum_{j=1}^{N} \log \mu_{x(j)} + \log N! \Big(\prod_{i \in I} k_i!\Big)^{-1} \\
&= N \sum_{i \in I} \rho_i \log \mu_i - N \sum_{i \in I} \rho_i \log \rho_i + o(N) \\
&= -N \mathcal{H}(\rho|\mu) + o(N) \qquad \text{as } N \to \infty.
\end{aligned}
$$

The common element in both points of view is the degeneracy, the number of microstates that is mapped to a single macrostate. Of course, this degeneracy only arises if the particles can not be distinguished from each other. Therefore I like to summarize this section like this:

> *Entropy arises from the indistinguishability of the particles in an empirical measure.*

We will return to this issue in Section 4.4.

4.3   Large deviations

For our purposes, the main role of the relative entropy is in the characterization of *large deviations of empirical measures*.

Large deviations are best explained by an example. We toss a balanced coin n times, and we call S_n the number of heads. Well-known properties of S_n are[2]

- $\dfrac{1}{n} S_n \overset{n \to \infty}{\longrightarrow} \dfrac{1}{2}$ almost surely (the law of large numbers)

- $\dfrac{2}{\sqrt{n}} \big(S_n - \dfrac{n}{2}\big) \overset{n \to \infty}{\longrightarrow} Z$ in law, where Z is a standard normal random variable (the central limit theorem).

The second (which contains the first) states that S_n is typically $n/2$ plus a random deviation of order $O(1/\sqrt{n})$. Deviations of this size from the expectation are called normal. Large deviations are those that are larger than normal, such as for instance the event that $S_n \geq an$ for some $a > 1/2$. Such large-deviation events have a probability that vanishes as $n \to \infty$, and a *large-deviation principle* characterizes exactly how fast it vanishes. A typical example is

$$
\text{For any } a \geq \frac{1}{2}, \qquad \operatorname{Prob}(S_n \geq an) \sim e^{-nI(a)} \qquad \text{as } n \to \infty, \tag{4.6}
$$

---

[2]This section draws heavily from the introduction in [dH00].

where

$$I(a) := \begin{cases} a \log a + (1-a) \log(1-a) + \log 2 & \text{if } 0 \le a \le 1 \\ +\infty & \text{otherwise} \end{cases}$$

The characterization (4.6) states that the probability of such a rare event decays exponentially in n, for large n, for each $a \ge 1/2$. The function I is called the *rate function*, since it characterizes the constant in the exponential decay.
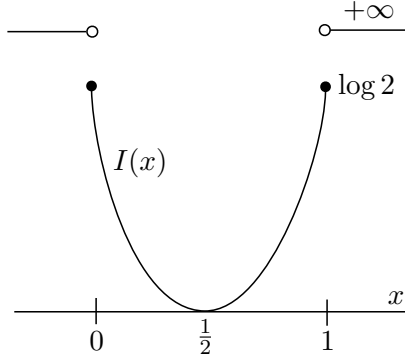

Figure 4.1: The function I

In order to explain what exactly the symbol $\sim$ in (4.6) means we give a precise definition of a large-deviation principle.

Definition 4.7. *A sequence $\mu_n \in \mathscr{P}(\Omega)$ satisfies a* large-deviation principle *with speed n and rate function I iff*

$$\forall O \subset \Omega \text{ open}, \qquad \liminf_{n \to \infty} \frac{1}{n} \log \mu_n(O) \ge -\inf_O I$$

$$\forall C \subset \Omega \text{ closed}, \qquad \limsup_{n \to \infty} \frac{1}{n} \log \mu_n(C) \le -\inf_C I.$$

Let us make a few remarks.

- The definition of the large-deviation principle is rather cumbersome. We often write it, formally, as

$$\text{Prob}(X_n \approx x) \sim e^{-nI(x)},$$

which is intended to mean exactly the same as Definition 4.7.

- The characterization (4.6) can be deduced from Definition 4.7 as follows. Take $\Omega = \mathbb{R}$, the set in which $\frac{1}{n} S_n$ takes its values, and define

$$\mu_n(A) := \text{Prob}\left(\frac{1}{n} S_n \in A\right).$$

Then $\text{Prob}(S_n \ge an) = \mu_n([a, \infty))$, and note that $\inf_{[a,\infty)} I = I(a)$ whenever $a \ge 1/2$. Supposing that the large-deviation principle has been proved for μ_n with rate

function I (see e.g. [dH00, Th. I.3]), we then find that for all $1/2 \leq a < 1$

$$-I(a) = -\inf_{(a,\infty)} I \ \leq \ \liminf_{n\to\infty} \text{Prob}(S_n > an)$$

$$\leq \ \limsup_{n\to\infty} \text{Prob}(S_n \geq an) \ \leq \ -\inf_{[a,\infty)} I \ = \ -I(a).$$

Therefore, if $1/2 \leq a < 1$ then the liminf and limsup coincide, and we have that

$$\lim_{n\to\infty} \frac{1}{n} \log \text{Prob}(S_n \geq an) = \lim_{n\to\infty} \frac{1}{n} \log \text{Prob}(S_n > an) = -I(a).$$

This is the precise version of (4.6).

- In the example of the coin, the two inequalities in Definition 4.7 reduced to one and the same equality at all points between 0 and 1, by the continuity of I at those points. In general a rate function need not be continuous, as the example of I above shows; neither is the function I unique. However, we can always assume that I is lower semi-continuous, and this condition makes I unique.

- Looking back at the discussion in Section 4.2, we see that for instance the limit (4.5) is a large-deviation description, at least formally. We also remarked there that the characterization (4.5) can not be true as it stands. In Definition 4.7 we see how this is remedied: instead of a single macrostate ρ, we consider open and closed sets of macrostates, which may contain states ρ of the form k/N for different values of N.

Remark 4.8. The definition of the large-deviation principle has close ties to two other concepts of convergence.

- A sequence of probability measures μ_n converges *narrowly* (in duality with continuous and bounded functions) to μ if

$$\forall O \subset \Omega \text{ open}, \qquad \liminf_{n\to\infty} \mu_n(O) \geq \mu(O)$$

$$\forall C \subset \Omega \text{ closed}, \qquad \limsup_{n\to\infty} \mu_n(C) \leq \mu(C).$$

Apparently, the large-deviation principle corresponds to a statement like 'the measures $\frac{1}{n} \log \mu_n$ converge narrowly'.

- The definition in terms of two inequalities also recalls the definition of Gamma-convergence, and indeed we have the equivalence

$$\mu_n \text{ satisfies a large-deviation principle with rate function } I \quad \Longleftrightarrow \quad \frac{1}{n} \mathcal{H}(\cdot \,|\mu_n) \xrightarrow{\Gamma} \hat{I},$$

where $\hat{I}(\nu) := \int I \, d\nu$. A proof is given in [Mar12].

$\square$

A property of large deviations that will come back later is the following. Suppose that we have a large-deviation result for a sequence of probability measures μ_n on a space $\mathcal{X}$ with rate functional I. Suppose that we now *tilt* the probability distribution μ_n by a functional $F : \mathcal{X} \to \mathbb{R}$, by defining the new measure

$$\tilde{\mu}_n(A) = \frac{\displaystyle\int_A e^{-nF(x)}\,\mu_n(dx)}{\displaystyle\int_{\mathcal{X}} e^{-nF(x)}\,\mu_n(dx)}.$$

This increases the probability of events x with lower $F(x)$, with respect to events x with higher $F(x)$, with a similar exponential rate (the prefactor n) as a large-deviation result.

The large-deviation behaviour of $\tilde{\mu}_n$ is now given by

Theorem 4.9 (Varadhan's Lemma (e.g. [dH00, Th. III.13])). *Let $F : \mathcal{X} \to \mathbb{R}$ be continuous and bounded from below. Then $\tilde{\mu}_n$ satisfies a large deviation principle with rate function*

$$\tilde{I}(x) := I(x) + F(x) - \inf_{\mathcal{X}}(I + F).$$

The final term in this expression is only a normalization constant that makes sure that $\inf \tilde{I} = 0$. The important part is that $\tilde{I}$ is the sum of the two functions I and F. In words: if we modify a probability distribution by tilting it with an exponential factor e^{-nF}, then that tilting function F ends up being *added* to the original rate function I.

4.4 Entropy as large-deviation rate function

Now to the question why relative entropy appears in the context of thermodynamics. Consider the following situation. We place n independent particles in a space $\mathcal{X}$ according to a distribution $\mu \in \mathscr{P}(\mathcal{X})$, i.e. the probability that the particle is placed in a set $A \subset \mathcal{X}$ is $\mu(A)$. We now consider the empirical measure of these n particles, which is the measure

$$\rho_n := \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i},$$

where X_i is the position of the ith particle. The empirical measure ρ_n is a random element of $\mathscr{P}(\mathcal{X})$, and the law of large numbers gives us that with probability one ρ_n converges weakly (in the sense of measures) to the law μ. (This is of course the standard way of determining μ if one only has access to the sample points X_i).

In this situation the large deviations of ρ_n are given by Sanov's theorem (see e.g. [DZ98, Th. 6.2.10]). The random measure ρ_n satisfies a large-deviation principle with rate n and rate function

$$I(\rho) := \mathcal{H}(\rho|\mu),$$

or in the shorthand notation that we used earlier,

$$\boxed{\text{Prob}(\rho_n \approx \rho) \sim e^{-n\mathcal{H}(\rho|\mu)} \qquad \text{as } n \to \infty.} \tag{4.10}$$

This is such an important result that I state it separately:

> *The relative entropy is the rate functional of the empirical measure of a large number of identical particles.*

Note the stress on 'empirical measure': the appearance of the relative entropy is intimately linked to the fact that we are considering empirical measures. Section 4.2 gives an insight into why this is: when passing from a vector of positions to the corresponding empirical measure, there is loss of information, since particles at the same position are indistinguishable.

4.5 Free energy and the Boltzmann distribution

In many books one encounters in various forms the following claim. Take a system of particles living in a space $\mathcal{X}$, and introduce an 'energy' $E : \mathcal{X} \to \mathbb{R}$ (in Joules, J) depending on the position $x \in \mathcal{X}$. Bring the system of particles into contact with a 'heat bath' of temperature T (in Kelvin, K), and let it settle into equilibrium. Then the probability distribution of the particles will be given by

$$\text{Prob}(A) = \frac{\displaystyle\int_A e^{-E(x)/kT}\,dx}{\displaystyle\int_{\mathcal{X}} e^{-E(x)/kT}\,dx}. \tag{4.11}$$

This is known as the *Boltzmann distribution*, or *Boltzmann statistics*, and the *Boltzmann constant* k has the value $1.4 \cdot 10^{-23}$ J/K. (Note that it only exists if the exponentials are integrable, which is equivalent to sufficient growth of E for large x. We will assume this for this discussion).

Where does this distribution come from? The concept of entropy turns out to give us the answer.

Since we need a system and a heat bath, we take two systems, called S and S_B (for 'bath'). Both are probabilistic systems of particles; S consists of n independent particles $X_i \in \mathcal{X}$, with probability law $\mu \in \mathscr{P}(\mathcal{X})$; similarly S_B consists of m independent particles $Y_j \in \mathcal{Y}$, with law $\nu \in \mathscr{P}(\mathcal{Y})$. The total state space of the system is therefore $\mathcal{X}^n \times \mathcal{Y}^m$.

The *coupling* between these systems will be via an *energy constraint*. We assume that there are energy functions $e : \mathcal{X} \to \mathbb{R}$ and $e_B : \mathcal{Y} \to \mathbb{R}$, and we will constrain the joint system to be in a state of fixed total energy, i.e. we will only allow states in $\mathcal{X}^n \times \mathcal{Y}^m$ that satisfy

$$\sum_{i=1}^{n} e(X_i) + \sum_{j=1}^{m} e_B(Y_j) = \text{constant}. \tag{4.12}$$

The physical interpretation of this is that energy (in the form of heat) may flow freely from one system to the other, but no other form of interaction is allowed.

Similar to the example above, we describe the total states of systems S and S_B by empirical measures $\rho_n = \frac{1}{n} \sum_i \delta_{X_i}$ and $\zeta_m = \frac{1}{m} \sum_j \delta_{Y_j}$. We define the average energies $E(\rho_n) := \frac{1}{n} \sum_i e(X_i) = \int_{\mathcal{X}} e \, d\rho_n$ and $E_B(\zeta_m) := \int_{\mathcal{Y}} e_B \, d\zeta_m$, so that the energy constraint above reads $nE(\rho_n) + mE_B(\zeta_m) = \text{constant}$.

By Section 4.4, each of the systems *separately* satisfies a large-deviation principle with rate functions $I(\rho) = \mathcal{H}(\rho|\mu)$ and $I_B(\zeta) = \mathcal{H}(\zeta|\nu)$. However, instead of using the explicit formula for I_B, we are going to assume that I_B can be written as a function of the energy E_B of the heat bath alone, i.e. $I_B(\zeta) = \tilde{I}_B(E_B(\zeta))$. For the coupled system we derive a joint large-deviation principle by choosing that (a) $m = nN$ for some large $N > 0$, and (b) the constant in (4.12) scales as n, i.e.

$$nE(\rho_n) + nNE_B(\zeta_{nN}) = n\overline{E} \qquad \text{for some } \overline{E}.$$

The joint system satisfies then a large-deviation principle[3]

$$\text{Prob}\left((\rho_n, \zeta_{nN}) \approx (\rho, \zeta) \;\middle|\; E(\rho_n) + NE_B(\zeta_{nN}) = \overline{E} \right) \sim \exp\left(-nJ(\rho, \zeta) \right),$$

with rate functional

$$J(\rho, \zeta) := \begin{cases} \mathcal{H}(\rho|\mu) + N\tilde{I}_B(E_B(\zeta)) + \text{constant} & \text{if } E(\rho) + NE_B(\zeta) = \overline{E}, \\ +\infty & \text{otherwise.} \end{cases}$$

Here the constant is chosen to ensure that $\inf J = 0$.

The functional J can be reduced to a functional of ρ alone,

$$J(\rho) = \mathcal{H}(\rho|\mu) + N\tilde{I}_B\left(\frac{\overline{E} - E(\rho)}{N} \right) + \text{constant}.$$

In the limit of large N, one might approximate

$$N\tilde{I}_B\left(\frac{\overline{E} - E(\rho)}{N} \right) \approx N\tilde{I}_B(\overline{E}) - \tilde{I}_B'(\overline{E})E(\rho).$$

The first term above is absorbed in the constant, and we find

$$J(\rho) \approx \mathcal{H}(\rho|\mu) - E(\rho)\tilde{I}_B'(\overline{E}) + \text{constant}.$$

In many cases I_B' is negative, since larger energies typically lead to higher probabilities and therefore smaller values of I_B. Now we simply define $kT := -1/\tilde{I}_B'(\overline{E})$, and we find

$$J(\rho) \approx \mathcal{H}(\rho|\mu) + \frac{1}{kT}E(\rho) + \text{constant}.$$

---

[3]This statement is formal; I haven't yet worked out how to formulate this rigorously for general state spaces.

Compare this to the typical expression for free energy $E - TS$; if we interpret S as $-k\mathcal{H}(\cdot|\mu)$, then we find the expression above, up to a factor kT.

Note that the right-hand side can be written as $\mathcal{H}(\rho|\tilde{\mu})$, where $\tilde{\mu}$ is the tilted distribution

$$\tilde{\mu}(A) = \frac{\displaystyle\int_A e^{-e(x)/kT}\,\mu(dx)}{\displaystyle\int_{\mathcal{X}} e^{-e(x)/kT}\,\mu(dx)}$$

Here we recognize the expression (4.11) for the case when μ is the Lebesgue measure.

This derivation shows that the effect of the heat bath is to *tilt* the system S: a state ρ of S with larger energy $E(\rho)$ implies a smaller energy E_B of S_B, which in turn reduces the probability of ρ. The role of temperature T is that of an exchange rate, since it characterizes the change in probability (as measured by the rate function I_B) per unit of energy. When T is large, the exchange rate is low, and then larger energies incur only a small probabilistic penalty. When temperature is low, then higher energies are very expensive, and therefore more rare. From this point of view, the Boltzmann constant k is simply the conversion factor that converts our Kelvin temperature scale for T into the appropriate 'exchange rate' scale.

One consequence of the discussion above is that the expression

$$\mathcal{F}(\rho) := \mathcal{H}(\rho|\mu) + \frac{1}{kT}E(\rho) + \text{constant} \tag{4.13}$$

is the rate function for a system of particles in contact with a heat bath. Let us make this statement precise, because this will explain some of the questions that we started this chapter with.

We recall the experiment that we just formulated. For each n, draw $n \times m$ particle positions $\{X_i^n\}i = 1^n$ and $\{Y_j^n\}_{j=1}^{nN}$ from the joint state space $\mathcal{X}^n \times \mathcal{Y}^{nN}$, with distribution μ (for X_i^n) and ν (for Y_j^n), conditioned on the equality (4.12) with constant equal to $n\overline{E}$. Then, writing again $\rho_n := n^{-1}\sum_i \delta_{X_i^n}$, we have seen that

$$\text{Prob}(\rho_n \approx \rho) \sim \exp(-n\mathcal{F}(\rho)) \qquad \text{as } n \to \infty.$$

For large n, this characterization implies that the distribution of ρ_n is strongly concentrated around the global minimizers of $\mathcal{F}$—and in the limit $n \to \infty$, it collapses onto this set. This explains 'modelling insight [M6]' on page 25, *the equilibrium states of this system are global minimizers of the free energy.*

In books on thermodynamics one often encounters the identity (or definition) $T = dS/dE$. This is formally the same as our definition of kT as $-dI_B/dE$, if one interprets I_B as an entropy and adopts the convention to multiply the non-dimensional quantity I_B with $-k$.

Yet another insight that this calculation gives is the following. It might be puzzling that in defining a free energy (e.g. $E - TS$, or $kT\,\mathsf{Ent} + E$), one adds two rather different objects: the energy of a system seems to be a completely different type of object than the entropy. This derivation of Boltzmann statistics shows that it's not exactly energy and entropy that one's adding; it really is more like adding two entropies ($\mathcal{H}$ and I_B, in the notation above). The fact that we write the second entropy as a constant times energy follows from the coupling and the approximation allowed by the assumption of a large heat bath.

35

4.6 Generalizations

Interaction energy In fact, the argument in the previous section applies just as well with an energy E that is not a simple sum over particles, but includes interaction effects. A typical example would be a pair-interaction energy of the type

$$\frac{1}{n}\sum_{i=1}^{n} V_b(X_i) + \frac{1}{n^2}\sum_{i,j=1}^{n} V_i(X_i - X_j) = \int_{\mathcal{X}} V_b \, d\rho_n + \iint_{\mathcal{X}\times\mathcal{X}} V_i(x-y)\,\rho_n(dx)\rho_n(dy).$$

Both electrostatic interaction energies and gravitaional energies are of this type. We come back to this type of energy in Section 5.6.

Multiple species The case of multiple species of particle can be treated by assuming that each particle has not only a position $X_i \in \Omega$ but also a *type* $T_i \in \{1,\ldots,m\}$. This corresponds to taking the set $\mathcal{X}$ above as $\mathcal{X} := \Omega \times \{1,\ldots,m\}$. We define the empirical measure ρ_n similarly as $\frac{1}{n}\sum_i \delta_{(X_i,T_i)}$, which is now a probability measure on the space $\mathcal{X}$ with variables $x \in \Omega$ and $t \in \{1,\ldots,m\}$.

Following the same line of reasoning, we again find an approximate rate functional

$$\mathcal{H}(\rho|\mu) + \frac{1}{kT}E(\rho) + \text{constant},$$

where μ is now the original distribution of the particles on the state space $\mathcal{X}$. If we write

$$\rho_i := \rho\big|_{\{t=i\}}, \qquad \mu_i := \mu\big|_{\{t=i\}},$$

then

$$\mathcal{H}(\rho|\mu) = \sum_i \mathcal{H}(\rho_i|\mu_i),$$

so that the rate functional becomes

$$\sum_i \mathcal{H}(\rho_i|\mu_i) + \frac{1}{kT}E(\rho) + \text{constant}. \tag{4.14}$$

4.7 Dimensional versions

In applications we work with concentrations that have dimensions, typically of moles per m^3 or per liter, and traditionally free energies have dimensions of energy rather than being dimensionless. To convert a dimensionless version such as (4.14) into a dimensional form we first multiply with an energy quantity to make the dimension that of energy. Since E is an energy, it is natural to multiply with kT.

We next connect the dimensionless measures ρ_i with dimensional concentrations c_i in a physical domain Ω by

$$\rho_i = \bar{c}^{-1}\, c_i\, \mathcal{L}^3\big|_{\Omega} \qquad \text{with} \qquad \bar{c} := \sum_i \int_{\Omega} c_i(x)\,dx. \tag{4.15}$$

The natural choice for μ in (4.14) is to be proportional to the Lebesgue measure,

$$\mu = \mu_0 \mathcal{L}^3\big|_\Omega \qquad \text{for some } \mu_0 > 0.$$

This corresponds to assuming no preference of the particles for one place above another—apart from that which might be encoded in the energy E. We then calculate that

$$kT \cdot (4.14) \;=\; kT \sum_i \bar{c}^{-1} \int_\Omega c_i(x) \log \frac{c_i(x)}{\bar{c}\mu_0} \, dx \;+\; E(c_1, \ldots, c_m) \;+\; \text{constant},$$

where we leave the dependence of E on $(c_1, \ldots, c_m)$ unspecified for the moment.

This expression can be thought of as the free energy *per particle*: the energy E in Section 4.5 is defined as an energy per particle, and the entropy terms are invariant under changes in particle number. The total free energy is this expression multiplied by the total number of particles $n\bar{c}$. If c is indeed measured in moles per cubic meter, then n is known exactly to be *Avogadro's number*, $N_A \simeq 6.022 \cdot 10^{23}$. Multiplying by $N_A \bar{c}$ we find

$$RT \sum_i \int_\Omega c_i(x) \log \frac{c_i(x)}{c_0} \, dx \;+\; \tilde{E}(c_1, \ldots, c_m) \;+\; \text{constant}.$$

Here the constant $R = kN_A \simeq 8.314 \,\text{J}\,\text{K}^{-1}\,\text{mol}^{-1}$ is the same universal gas constant as in Section 2.3. In this expression we have replaced the energy-per-particle E by the total energy $\tilde{E}$, and we replaced $\bar{c}\mu_0$ by a reference concentration c_0. Note that the value of c_0 only changes the value of the constant in this expression, and therefore we can choose it arbitrarily.

4.8 Modelling free energy

Let us now summarize. The discussions above give an algorithm how to construct the total free energy in systems described by concentrations. (For completeness: the total free energy is equal to the large-deviation rate functional, multiplied by kT and by the total number of particles.) Take a system with m concentrations $c_1, \ldots, c_m : \Omega \to [0, \infty)$ (in moles per m^3), which can exchange energy with a heat bath with temperature T, assuming that we know the total energy $E(c_1, \ldots, c_m)$ as a function of $(c_1, \ldots, c_m)$. Apart from the interaction described by the energy, the particles are assumed to have no interaction with each other. By the arguments above we find that the free energy of this system is

$$\mathcal{F}(c_1, \ldots, c_m) := E(c_1, \ldots, c_m) + RT \sum_{i=1}^m \int_\Omega c_i(x) \log \frac{c_i(x)}{c_0} \, dx + \text{constant}. \qquad (4.16)$$

We saw in the sections above how the entropy terms in (4.16) arise from the indistinguishability of the particles of the same species: each species contributes a term $RT \int c \log c/c_0$ to the free energy.

While modelling the system of diffusing particles, in Section 2.3, we chose a functional of this type, (2.4). We now recognize that this is a free energy, i.e. a dimensional version of the large-deviation rate functional associated with concentrations of independent particles. This argument gives an interpretation of this functional, but it does not yet explain why it should be a driving force in a gradient flow. That aspect we discuss in the next chapter.

Chapter 5

Free energy dissipated through Wasserstein metrics

In this chapter I want to establish the following modelling insights.

M7 Particles moving through a viscous fluid dissipate energy. If the starting and ending positions of n particles are characterized by two empirical measures ρ_0 and ρ_1, then the minimally dissipated energy in a stationary Newtonian fluid is

$$\frac{n\eta}{\tau} W_2(\rho_0, \rho_1)^2,$$

where τ is the duration of the motion, η is a friction parameter, and W_2 is the Wasserstein distance.

M8 The Wasserstein distance W_2 also characterizes the mobility of empirical measures of Brownian particles, in the sense of large deviations.

M9 In an SDE with Brownian noise, in the many-particle limit, free energy is dissipated through the Wasserstein distance. Therefore in many cases the many-particle limit is a Wasserstein gradient flow of the free energy.

Taken together, these will explain the modelling choices of Section 2.3.

5.1   Dissipation in a viscous fluid

When a spherical particle is dragged through a Newtonian viscous fluid, the relative velocity v is linearly related to the force f [N] that is required to keep it moving:

$$f = \eta v. \tag{5.1}$$

(Stokes' law gives an explicit expression for the *friction coefficient* η: $\eta = 6\pi\nu R$, where R is the radius of the particle and ν [Ns/m^2] the dynamic viscosity of the fluid; see [Sto51,

Eq. (126)] or [LL87, §20].) Therefore, if $X_1(t), \ldots, X_n(t)$ are trajectories of n particles, for $t \in [0, \tau]$, then the total work done by the particles on the fluid is equal to

$$\sum_{i=1}^{n} \int_0^\tau f_i(t) \cdot \dot{X}_i(t) \, dt = \eta \sum_{i=1}^{n} \int_0^\tau \dot{X}_i(t)^2 \, dt = \frac{\eta}{\tau} \sum_{i=1}^{n} \int_0^1 \left(\partial_s X_i(\tau s) \right)^2 ds.$$

This work is converted into heat by friction with the molecules of the fluid.

Writing $\rho_s = n^{-1} \sum_i \delta_{X_i(\tau s)}$ for the empirical measure of these particles, we have, using the formula for the Wasserstein tensor of atomic measures,

$$\frac{\eta}{\tau} \sum_{i=1}^{n} \int_0^1 \left(\partial_s X_i(\tau s) \right)^2 ds \stackrel{(3.15)}{=} \frac{\eta n}{\tau} \int_0^1 \|\dot{\rho}_s\|_{-1,\rho_s}^2 \, ds \stackrel{(3.14)}{\geq} \frac{\eta n}{\tau} W_2(\rho_0, \rho_1)^2.$$

This lower bound is achieved by a particle motion along straight lines with constant velocity.

This observation shows us that

> the Wasserstein distance for empirical measures of particles can be interpreted as the minimal energy dissipated by moving those particles through a viscous fluid.

A similar statement holds locally in time: the Wasserstein tensor $\|\dot{\rho}\|_{-1,\rho}^2$, multiplied by the coefficient $n\eta$, is the power (work per unit of time) expended by the movement of n particles. The dimensional version of this statement is found by again connecting a concentration c with a particle empirical measure ρ by $\rho = c/\bar{c}\,\mathcal{L}|_\Omega$ and $n = \bar{c} N_A$ (see Section 4.7). Then, writing $\dot{\rho} = -\operatorname{div} \rho v = -\operatorname{div}(cv/\bar{c})$, we have

$$n\eta \|\dot{\rho}\|_{-1,\rho}^2 = n\eta \int \rho |v|^2 = \frac{n\eta}{\bar{c}} \int |w|^2 c \, dx = \eta N_A \int |w|^2 c \, dx. \tag{5.2}$$

This observation provides a first motivation of the choice of the Wasserstein metric tensor in (2.6),

$$\widetilde{\Psi}(c, w) = \frac{\tilde{\eta}}{2} \int_\Omega |w(x)|^2 c(x) \, dx, \qquad \text{where } \dot{c} = -\operatorname{div} cw.$$

By the argument above, $\widetilde{\Psi}(c, w)$ is one-half of the energy dissipation associated with moving the particles of concentration c according to w, where $\tilde{\eta} = \eta N_A$ is a macroscopic friction coefficient. If one is willing to accept that the free energy $\mathcal{F}$ in (2.4) is the driving force in a gradient flow, then modelling the dissipation with $\widetilde{\Psi}$ corresponds to assuming that dissipation of the free energy happens through viscous dissipation generated by the particles, as they move through a stagnant fluid.

While this argument has the advantage of being simple, it neglects the fact—put forward in the previous chapter—that the free energy, and especially the entropy terms, can only be understood from a stochastic point of view. In the next section we therefore develop a stochastic view on the Wasserstein distance.

5.2 Brownian particles and Wasserstein dissipation

The previous section gave an interpretation of the Wasserstein distance in terms of viscous dissipation—conversion of work into heat by particles moving through a fluid. We now focus on the role of the Wasserstein distance in *stochastic* particle systems, and we first consider a simple case.

Consider the system of n particles $X_{\varepsilon,i}$ in $\mathbb{R}^d$ ($i = 1, \ldots, n$) defined as a rescaled Brownian motion:

$$dX_{\varepsilon,i}(t) = \sqrt{2}\,\varepsilon\,dW_i(t), \qquad X_{\varepsilon,i}|_{\{t=0\}} = x_{0,i}. \tag{5.3}$$

where for each i, W_i is a Brownian motion in $\mathbb{R}^d$, $\sigma > 0$ is a mobility coefficient, and the vector of positions $x_0 = (x_{0,1}, \ldots, x_{0,n}) \in \mathbb{R}^{nd}$ is fixed. If we fix $\tau > 0$, then by Schilder's theorem (e.g. [DZ98, Th. 5.2.3]), the process $\{X_\varepsilon : t \in [0, \tau]\}$, satisfies a large-deviation principle for fixed n and small ε,

$$\mathrm{Prob}\big(X_\varepsilon|_{[0,\tau]} \approx x|_{[0,\tau]}\big) \sim \exp\Big[-\frac{1}{\varepsilon^2}I(x)\Big], \qquad \text{as } \varepsilon \to 0,$$

with rate functional

$$I(x) = \frac{1}{4}\int_0^\tau |\dot{x}(t)|^2\,dt \qquad \text{provided } x|_{t=0} = x_0.$$

Now note that the exponent can be rewritten, again using the notation $\rho_n = \frac{1}{n}\sum_i \delta_{x_i}$, as

$$\frac{1}{4\varepsilon^2}\int_0^\tau |\dot{x}(t)|^2\,dt = \frac{1}{4\varepsilon^2}\sum_{i=1}^n \int_0^\tau |\dot{x}_i(t)|^2\,dt = \frac{n}{4\varepsilon^2}\int_0^\tau \|\dot{\rho}_n(t)\|_{-1,\rho_n(t)}^2\,dt.$$

By the same argument as in the previous section, this expression has a sharp lower bound

$$\frac{1}{\varepsilon^2}I(x) \geq \frac{n}{4\varepsilon^2\tau}W_2\big(\rho_n(0), \rho_n(\tau)\big)^2. \tag{5.4}$$

Note how the this expression equals the minimal dissipated energy, when η is equal to $1/4\varepsilon^2$.

Léonard [Léo12] was the first to prove the corresponding statement for the many-particle limit of empirical measures of (constant-mobility) Brownian particles, as follows. Let $X_{1,i}$ be the set of Brownian particles defined in (5.3) with $\varepsilon = 1$, and again let $\rho_n = \frac{1}{n}\sum_i \delta_{X_{1,i}}$ be the empirical measure. Then

$$\mathrm{Prob}\big(\rho_n(t) \approx \rho^1 | \rho_n(0) \approx \rho^0\big) \sim \exp\big[-nI_t(\rho^1|\rho^0)\big] \qquad \text{as } n \to \infty, \tag{5.5}$$

and the rate function $I_t(\cdot\,|\,\cdot)$ satisfies

$$tI_t(\cdot\,|\rho^0) \to \frac{1}{4}W_2(\rho^0, \rho^1)^2 \qquad \text{as } t \downarrow 0$$

in the sense of Gamma-convergence (see also [ADPZ11, DLR12, PR11]). Written informally, this result states that

$$\text{Prob}\left(\rho_n(t) \approx \rho^1 | \rho_n(0) \approx \rho^0\right) \sim \exp\left[-\frac{n}{4t}W_2(\rho^0, \rho^1)^2\right] \qquad \text{as } n \to \infty \text{ and then } t \downarrow 0.$$
(5.6)

This is the sense in which

> *The Wasserstein distance characterizes the stochastic mobility of empirical measures of systems of Brownian particles.*

5.3 Brownian particles and Wasserstein dissipation, take 2

The large-deviation characterization (5.6) shows the connection between the Wasserstein distance and Brownian particles, but it doesn't yet explain how the whole gradient-flow arises. For this, it turns out, we need to include 'the next term' in the small-t asymptotic expansion of I_t. In [ADPZ11] we first proved that

$$I_t(\,\cdot\,|\rho^0) = \frac{1}{4t}W_2(\rho^0, \rho^1)^2 + \frac{1}{2}\,\text{Ent}(\,\cdot\,) - \frac{1}{2}\,\text{Ent}(\rho^0) + o(1) \qquad \text{as } t \downarrow 0, \qquad (5.7)$$

in the sense of Gamma-asymptotic developments [AB93]. This was later generalized to a larger class of systems in [DLR12].

The function on the right-hand side of (5.7) is well known in the theory of gradient flows, as the basis for a discrete-time approximation of a gradient flow. Construct a sequence ρ^k in an iterative fashion: fix $h > 0$, and for each ρ^{k-1} let ρ^k be the solution of

$$\min_\rho \frac{1}{2h}W_2(\rho, \rho^{k-1})^2 + \text{Ent}(\rho),$$

starting from some ρ^0. Then the piecewise-constant interpolation of this sequence with time step h, i.e. the function $t \mapsto \rho^{\lfloor t/h \rfloor}$, is an approximation of the Wasserstein gradient flow of Ent [JKO98].

5.4 Interpretation

Equations (5.5) and (5.7) together illustrate how the dissipation metric W_2 and the entropy Ent together create a macroscopic gradient-flow behaviour. We have two competing phenomena:

1. For short time, Brownian particles prefer to stay put: this is represented by the large-deviation contribution

$$\exp\left[-\frac{n}{4t}W_2(\rho^1, \rho_0)^2\right].$$

2. Macrostates with lower entropy Ent contain more microstates; this leads to the large-deviation contribution

$$\exp\left[-\frac{n}{2}\left(\mathsf{Ent}(\rho^1) - \mathsf{Ent}(\rho^0)\right)\right]$$

The two effects independently behave like large-deviation effects; because both exponents scale with n, they are able to compensate each other, and the resulting behaviour is a mixture of the movement-aversion characterized by W_2 and the movement-preference characterized by Ent. This mixture is the gradient flow.

5.5 Revisit the derivation of solute diffusion

Let us now walk through the derivation of the convection-diffusion equation in Section 2.3, and interpret the choices made there in the light of the last two chapters. We repeat the relevant parts.

State space: We choose a bounded set $\Omega \subset \mathbb{R}^d$ to be the container with the fluid; the particles are represented by their concentration $c : \Omega \to [0, \infty)$ (in moles per m^3). The state space is therefore $\mathcal{Z} := \{c \in L^1(\Omega) : c \geq 0\}$. Positions in Ω are labeled x.

Energy: We choose as driving functional for this system the *free energy*

$$\mathcal{F} : \mathcal{Z} \to \mathbb{R}, \qquad \mathcal{F}(c) := RT \int_\Omega c(x) \log \frac{c(x)}{c_0} \, dx + \rho g \int_\Omega x_3 c(x) \, dx, \qquad (5.8)$$

where $c_0 > 0$ is an arbitrary reference concentration, and ρ is the mass density contrast with the fluid.

Processes: We allow the state $c \in \mathcal{Z}$ to change through the effects of a velocity $w : \Omega \to \mathbb{R}^d$ by

$$\dot{c} + \operatorname{div} cw = 0 \quad \text{in } \Omega, \qquad w \cdot n = 0 \quad \text{on } \partial\Omega, \qquad (5.9)$$

or in weak form,

$$\forall \varphi \in C_b^1(\Omega) : \qquad \partial_t \int_\Omega \varphi(x) c(t, x) \, dx - \int_\Omega c(t, x) w(t, x) \nabla \varphi(x) \, dx = 0.$$

Dissipation potential: We define the *dissipation potential* on the set of process vectors w as the functional

$$\widetilde{\Psi}(c, w) := \frac{\eta}{2} \int_\Omega |w(x)|^2 c(x) \, dx. \qquad (5.10)$$

We now understand that these choices mean the following.

- The free energy (5.8) arises because we consider *concentrations*; there is no energy term because (a) the particles are independent, and (b) there is no distinction between the points of Ω, from the point of view of the particles.

- The dissipation potential is (one-half of) the heat generated by a particle movement w (see page 11 for the 'one-half').

- The combination of the two can be motivated by the fact that both are large-deviation rate functionals at the same speed n.

5.6   Large deviations for SDEs

We can take the explanation of Section 5.3 a step further by considering large deviations of the time course of empirical measures. At the same time we extend the previous example by including interaction of the particles with a background potential V_b and with each other via an interaction potential V_i. Specifically, we take a system of n particles described by

$$dX_i(t) = -\nabla V_b(X_i(t))\, dt - \frac{1}{n} \sum_{j=1}^{n} \nabla V_i(X_i(t) - X_j(t))\, dt + \sqrt{2}\, dW_i(t). \tag{5.11}$$

The continuum limit, as $n \to \infty$, of this system is the equation

$$\dot{\rho} = \Delta \rho + \operatorname{div} \rho \nabla \big[V_b + \rho * V_i \big]. \tag{5.12}$$

The large-deviation rate functional describing fluctuations of the system is given by (see [FK06, Theorem 13.37] and [DG87])

$$I(\rho) := \frac{1}{4} \int_0^T \left\| \dot{\rho} - \Delta \rho - \operatorname{div} \rho \nabla \big[V_b + \rho * V_i \big] \right\|_{-1,\rho}^2 dt. \tag{5.13}$$

We now connect this expression to the gradient-flow structures defined in Section 3.2.

1. First, define the free energy $\mathcal{F}$ as the sum of entropy and potential energy for this system, as in Sections 4.5 and 4.6:

$$\mathcal{F}(\rho) := \mathsf{Ent}(\rho) + \int_{\mathbb{R}^d} \left[\rho V_b + \frac{1}{2} \rho(\rho * V_i) \right]. \tag{5.14}$$

For this functional $\mathcal{F}$, the Wasserstein gradient $\mathcal{K}\mathrm{D}\mathcal{F}$ is equal to (minus) the right-hand side of (5.12):

$$\mathcal{K}(\rho)\mathrm{D}\mathcal{F}(\rho) = \mathcal{K}(\rho)\big[\log \rho + 1 + V_b + \rho * V_i \big] = -\operatorname{div} \rho \nabla \big[\log \rho + 1 + V_b + \rho * V_i \big].$$

2. Second, expanding the square in (5.13) and using the chain rule (3.13), I can therefore be written as

$$2I(\rho) = \mathcal{F}(\rho(T)) - \mathcal{F}(\rho(0)) + \frac{1}{2} \int_0^T \big[\|\dot{\rho}\|_{-1,\rho}^2 + \| - \mathrm{D}\mathcal{F}(\rho)\|_{1,\rho}^2 \big]\, dt,$$

Note that $\|\dot{\rho}\|_{-1,\rho}$ is the Wasserstein dissipation, and $\| \cdot \|_{1,\rho}$ is the dual Wasserstein norm. Therefore the rate function I is *exactly* the 'global definition' of the gradient flow of Definition 3.4 (actually, up to a factor 2).

Finally, note that if we scale time by setting $t = Ts$, for $s \in [0, 1]$, then in terms of the rescaled time s we can write

$$I(\rho) = \frac{1}{4T} \int_0^1 \|\partial_s \rho\|^2_{-1,\rho} \, ds + \frac{1}{2} \big[\mathcal{F}(\rho(T)) - \mathcal{F}(\rho(0)) \big] + \frac{T}{4} \int_0^1 \|-\mathrm{D}\mathcal{F}(\rho)\|^2_{1,\rho}, \, ds.$$

Compare this with the expression (5.7): the term $W_2(\rho^1, \rho^0)^2$ is the infimum over all curves connecting ρ^0 to ρ^1 in time 1, and therefore corresponds to the first term above; for the example of purely Brownian particles of Section (5.3), $\mathcal{F} = \mathsf{Ent}$, and therefore the second terms of the two expressions are identical; and a separate argument shows that the third term above indeed is expected to be small as $T \to 0$ [DLR12].

5.7   Geometry and reversibility

There are interesting connections between the geometry of the Brownian noise, the reversibility of the stochastic process, and the question whether the resulting evolution equation is a gradient flow or not.

This becomes apparent when we modify the system of the previous section by introducing a diffusion matrix $A \in \mathbb{R}^{d \times d}$ and a mobility matrix $\sigma \in \mathbb{R}^{d \times d}$, as follows:

$$dX_i(t) = -A\nabla V_b(X_i(t)) \, dt - \frac{1}{n} \sum_{j=1}^n A\nabla V_i(X_i(t) - X_j(t)) \, dt + \sqrt{2}\, \sigma \, dW_i(t). \qquad (5.15)$$

The large-deviation rate functional of the system is similarly given by

$$I(\rho) := \frac{1}{4} \int_0^T \left\| \dot{\rho} - \mathrm{div}\, \sigma\sigma^T \nabla \rho - \mathrm{div}\, \rho A \nabla \big[V_b + \rho * V_i \big] \right\|^2_{-1,D(\rho)} dt, \qquad (5.16)$$

where the norm $\| \cdot \|_{-1,D(\rho)}$ is induced by the inner product

$$(s_1, s_2)_{-1,D(\rho)} := \int w_1 \cdot w_2 \, D(\rho) \, dx,$$

with $D(\rho) = \rho\sigma\sigma^T$. As before, the hydrodynamic limit of this system is the minimiser of I,

$$\dot{\rho} = \mathrm{div}\, \sigma\sigma^T \nabla \rho + \mathrm{div}\, \rho A \nabla \big[V_b + \rho * V_i \big]. \qquad (5.17)$$

With the additional parameter freedom in A and σ, it is not always possible to write (5.16) in the form of (3.5). This depends on whether the cross term in (5.16) is an exact differential, i.e., whether there exists a functional $\mathcal{E}$ such that

$$\Big(\dot{\rho}, -\mathrm{div}\, \sigma\sigma^T \nabla \rho - \mathrm{div}\, \rho A \nabla \big[V_b + \rho * V_i \big] \Big)_{-1,D(\rho)} = \partial_t \mathcal{E}(\rho).$$

This is the case if and only if $\sigma\sigma^T$ is a positive multiple of A, a condition that is familiar from the fluctuation-dissipation theorem, also known as the Einstein relation. In that case, and writing $\sigma\sigma^T = kTA$ for some 'temperature' $T > 0$ and the Boltzmann constant k,

$$-\operatorname{div}\sigma\sigma^T\nabla\rho - \operatorname{div}\rho A\nabla\big[V_b + \rho * V_i\big] = \mathcal{K}(\rho)\mathrm{D}\mathcal{F}(\rho),$$

where $\mathcal{K}(\rho)\xi$ now is defined as $-\operatorname{div}D(\rho)\nabla\xi$ and the free energy $\mathcal{F}$ is a modification of (5.14),

$$\mathcal{F}(\rho) := \mathsf{Ent}(\rho) + \frac{1}{kT}\int_{\mathbb{R}^d}\left[\rho V_b + \frac{1}{2}\rho(\rho * V_i)\right].$$

Then the rate functional I can be written in the form (3.5) as

$$2I(\rho) = \mathcal{F}(\rho(T)) - \mathcal{F}(\rho(0)) + \frac{1}{2}\int_0^T\left[\|\dot\rho\|_{-1,D(\rho)}^2 + \|-\mathrm{D}\mathcal{F}(\rho)\|_{1,D(\rho)}^2\right]\,dt$$

and the evolution equation (5.17) is the (modified, D-) Wasserstein gradient flow of $\mathcal{F}$.

Our freedom to choose A and σ separately gives us the insight that for this system the following four statements are equivalent:

1. $\sigma\sigma^T = kTA$ for some $T > 0$;

2. The evolution (5.17) is a $D(\rho)$-Wasserstein gradient flow of $\mathcal{F}$;

3. The rate functional I can be written in the form (3.5);

4. For any finite number n of particles, the system (5.15) is reversible.

This equivalence, which holds for this specific system, suggests a much deeper connection between reversibility and gradient-flow structure, that we comment on in detail in [MPR].

5.8   Comments

Validity of Stokes' law.   A natural criticism of the law (5.1) with $\eta = 6\pi\nu R$ would be that it is derived for a macroscopic spherical particle in a continuum viscous fluid, with no-slip boundary conditions. Real particles are not spherical and they are embedded in a sea of other molecules, which may even be of similar size. This is a valid point.

Surprisingly, Stokes' law is fairly robust under such generalizations. Molecular-dynamics simulations of hard-sphere particles in a hard-sphere 'fluid' show the same law, regardless of the size of the particles (see e.g. [BHP94, Fig. 5]). In general, the law holds, with the same coefficient, for particles no smaller than a few times the size of the surrounding particles; below this size the coefficient may be different. See e.g. [Li09] and the references therein.

Chapter 6

Further Examples

6.1 A moving vesicle in a viscous fluid with diffusing solutes

The strength of the variational-modelling framework lies in the modelling of more complex systems. Therefore we now discuss an example with multiple components that interact.

In this example $\Omega \subset \mathbb{R}^d$ is not fixed, but moves around in some viscous fluid. A typical case is a vesicle (a biological container bounded by a membrane of lipids) immersed in water. Inside Ω some chemical is dissolved in the fluid, which can diffuse freely inside Ω, but can not pass the boundary $\partial\Omega$. The surrounding fluid can pass through $\partial\Omega$ albeit with some resistance.

We now go through the same modelling steps as before.

State space: The state is a set of pairs:

$$\mathcal{Z} := \{(\Omega, c) : \Omega \subset \mathbb{R}^d, \ c \in L^1(\mathbb{R}^d), \ \operatorname{supp} c \subset \Omega\}.$$

The requirement that the solute stays within Ω is encoded in the definition of $\mathcal{Z}$.

Energy: The boundary of the vesicle is assumed to have surface tension, which implies that the domain Ω has *surface energy*. The driving force should therefore be the sum of (negative) entropy and surface energy, the *free energy*

$$\mathcal{F} : \mathcal{Z} \to \mathbb{R}, \qquad \mathcal{F}(\Omega, c) := RT \int_\Omega c(x) \log \frac{c(x)}{c_0} \, dx + \alpha |\partial\Omega|,$$

where we write $|\partial\Omega|$ for the perimeter of Ω (the surface area in three dimensions, and the length of the boundary in two dimensions). The parameter $\alpha > 0$ is the energy per unit area, which is often called the surface tension.

Processes: We allow the state $(\Omega, c) \in \mathcal{Z}$ to change through the effects of three objects, w, u, and v_n:

1. The evolution of Ω is characterized by the normal velocity v_n of the boundary $\partial\Omega$ (the normal n points outwards);

2. The water moves with velocity $u : \mathbb{R}^d \to \mathbb{R}^d$, which is required to be divergence-free, $\operatorname{div} u = 0$;

3. The solute concentration evolves (as above) through a solute velocity w with $\operatorname{supp} w \subset \Omega$:

$$\dot{c} + \operatorname{div} cw = 0 \quad \text{in } \mathcal{D}'(\mathbb{R}^d), \tag{6.1}$$

or again in weak form,

$$\forall \varphi \in C_b^1(\mathbb{R}^d): \qquad \partial_t \int_{\mathbb{R}^d} \varphi(x) c(t, x)\, dx - \int_{\mathbb{R}^d} c(t, x) w(t, x) \nabla \varphi(x)\, dx = 0. \tag{6.2}$$

Note that we now consider w defined on the whole of $\mathbb{R}^d$, and the no-flux boundary condition of the previous case has vanished. Indeed the normal flux need not be zero on the boundary: if the boundary moves, then it will typically happen that nearby solutes move with it. A formal calculation with the weak form (6.2) and the condition $\operatorname{supp} c, \operatorname{supp} w \subset \Omega$ shows that it indeed implies a weak version of the boundary condition

$$w \cdot n = v_n \qquad \text{on } \partial\Omega. \tag{6.3}$$

Again these assumptions contain a number of modelling choices, such as the fact that the water is assumed to be incompressible. Another choice is hidden in the fact that v_n may be different from $u \cdot n$, i.e. that the boundary does not necessarily move with the fluid—this implies that the fluid can move through the boundary.

In this case a process vector is thus a triple (w, u, v_n).

Dissipation potential: We define the dissipation potential on the set of process vectors (w, u, v_n) as the functional

$$\widetilde{\Psi}(\Omega, c; w, u, v_n) := \frac{\eta_c}{2} \int_\Omega c|w - u|^2 + \frac{\eta_u}{2} \int_{\mathbb{R}^d} |\varepsilon(u)|^2 + \frac{\eta_b}{2} \int_{\partial\Omega} (u \cdot n - v_n)^2. \tag{6.4}$$

Here η_c, η_u, and η_b are three friction-type constants; $\varepsilon(u) = \frac{1}{2}(\nabla u + \nabla u^T)$ is the symmetric part of the gradient ∇u. Note the different domains of integration.

The first term is the same as in the previous example, with a twist: we first subtract cu from w, before we penalize the result. This has the effect of measuring this dissipation in terms of the difference with the convective flux cu. The idea is that simply convecting the solute along with the flow u of the fluid, i.e. with flux cu, should not lead to diffusive dissipation; differences with respect to cu should. The second term is a measure of dissipation in the viscous fluid, as a result of shear in the velocity field u. The inclusion of this term therefore reflects the choice that fluid movement is accompanied by dissipation of energy through viscous friction. The third term similarly measures dissipation due to friction, but in the boundary, in terms of the relative velocity $u \cdot n - v_n$ between fluid and

boundary. The inclusion of this term therefore implies that water can move through the boundary, but doing so requires dissipation of energy.

Derive the equations: We now apply the algorithm to find the equations of motion, by minimizing

$$(w, u, v_n) \mapsto \widetilde{\Psi}(\Omega, c; w, u, v_n) + \langle \mathcal{F}'(\Omega, c), (w, u, v_n) \rangle. \tag{6.5}$$

The term in angle brackets requires some specification. If $t \mapsto (\Omega_t, c_t)$ is a curve in $\mathcal{Z}$, with process vector (w, u, v_n) at $t = \tau$, then by the Reynolds transport theorem

$$\partial_t \int_{\Omega_t} c_t \log \frac{c_t}{c_0} \Big|_{t=\tau} = \int_{\Omega_\tau} \left(\log \frac{c_\tau}{c_0} + 1 \right) \dot{c}_t \big|_{t=\tau} + \int_{\partial\Omega_\tau} c_\tau \log \frac{c_\tau}{c_0} v_n.$$

The first integral on the right-hand side can be rewritten using (6.1) and (6.3) into

$$-\int_{\Omega_\tau} \left(\log \frac{c_\tau}{c_0} + 1 \right) \operatorname{div} w = \int_{\Omega_\tau} w \cdot \nabla c_\tau - \int_{\partial\Omega_\tau} \left(\log \frac{c_\tau}{c_0} + 1 \right) c_\tau w \cdot n$$

$$= \int_{\Omega_\tau} w \cdot \nabla c_\tau - \int_{\partial\Omega_\tau} \left(\log \frac{c_\tau}{c_0} + 1 \right) c_\tau v_n$$

By the properties of the total curvature H of the surface $\partial\Omega$ (i.e. the sum of principal curvatures, or equivalently d times the mean curvature),

$$\partial_t |\partial\Omega_t| \Big|_{t=\tau} = -\int_{\partial\Omega_\tau} H v_n.$$

Therefore, and this is the definition of the angle-bracket term,

$$\langle \mathcal{F}'(\Omega_\tau, c_\tau), (w, u, v_n) \rangle := \partial_t \mathcal{F}(\Omega_t, c_t) \big|_{t=\tau} = RT \int_{\Omega_\tau} w \cdot \nabla c_\tau - \int_{\partial\Omega_\tau} \left[RT c_\tau + \alpha H \right] v_n.$$

Minimization of (6.5) with respect of each of the three components leads to the system of equations

$$\forall \tilde{w}: \quad 0 = \eta_c \int_\Omega c(w - u) \cdot \tilde{w} + RT \int_\Omega \tilde{w} \cdot \nabla c$$

$$\forall \tilde{u}: \quad 0 = \eta_c \int_\Omega c(u - w) \cdot \tilde{u} + \eta_u \int_{\mathbb{R}^d} \varepsilon(u) : \varepsilon(\tilde{u}) + \eta_b \int_{\partial\Omega} (u \cdot n - v_n) \tilde{u} \cdot n - \int_{\mathbb{R}^d} p \operatorname{div} \tilde{u}$$

$$\forall \tilde{v}_n: \quad 0 = \eta_b \int_{\partial\Omega} (v_n - u \cdot n) \tilde{v}_n - \int_{\partial\Omega} \left[RT c + \alpha H \right] \tilde{v}_n.$$

Note the appearance of the pressure p as a Lagrange multiplier associated with the constraint $\operatorname{div} u = 0$. Because of this Lagrange multiplier the second equation holds for each $\tilde{u}$, even for those that are not divergence-free.

By performing similar manipulations on these equations as in the previous example, we derive that

$$cw = cu - \frac{RT}{\eta_c}\nabla c \quad \text{in } \Omega, \qquad w \cdot n = v_n \text{ on } \partial\Omega,$$

$$v_n = u \cdot n + \frac{1}{\eta_b}[RTc + \alpha H] \text{ on } \partial\Omega,$$

and for u we find the set of equations and boundary conditions

$$
\begin{aligned}
-\operatorname{div}\sigma &= -RT\nabla c && \text{in } \Omega \\
-\operatorname{div}\sigma &= 0 && \text{in } \mathbb{R}^d \setminus \Omega \\
\operatorname{div} u &= 0 && \text{in } \mathbb{R}^d \\
[\sigma] \cdot n &= \eta_b(u \cdot n - v_n)n && \text{on } \partial\Omega.
\end{aligned}
$$

Here $\sigma = \eta_u\varepsilon(u) - pI$, and $[\sigma]$ is the jump in σ over $\partial\Omega$, i.e. $\sigma_{\text{ext}} - \sigma_{\text{int}}$.

Combining these expressions with (6.1) and (6.3) we find

$$-\operatorname{div}\sigma = -RT\nabla c \qquad\qquad \text{in } \Omega \tag{6.6a}$$

$$-\operatorname{div}\sigma = 0 \qquad\qquad \text{in } \mathbb{R}^d \setminus \Omega \tag{6.6b}$$

$$\operatorname{div} u = 0 \qquad\qquad \text{in } \mathbb{R}^d \tag{6.6c}$$

$$\dot{c} = -\operatorname{div}(cu) + \frac{RT}{\eta_c}\Delta c \qquad \text{in } \Omega \tag{6.6d}$$

$$[\sigma] \cdot n = -RTc - \alpha H \qquad \text{on } \partial\Omega \tag{6.6e}$$

$$cv_n = \left(\frac{RT}{\eta_c}\nabla c - cu\right) \cdot n \qquad \text{on } \partial\Omega \tag{6.6f}$$

$$v_n = u \cdot n + \frac{1}{\eta_b}[RTc + \alpha H] \qquad \text{on } \partial\Omega. \tag{6.6g}$$

6.2   Discussion

Interpretation.   The equations (6.6) combine a number of effects. The first three equations describe Stokes flow in the two domains Ω and $\mathbb{R}^d \setminus \Omega$, driven by boundary forcing. Note that the term RTc is the same as the *osmotic pressure* given by the Van 't Hoff equation (see e.g. [AdP06, Sec. 5.4]). If necessary, it can be incorporated into the pressure p, since this pressure is a Lagrange multiplier and therefore not known explicitly. In that case the osmotic pressure disappears from equations (6.6a) and (6.6e); it still remains present in the boundary conditions (6.6f) and (6.6g).

Equation (6.6d) is a traditional convection-diffusion very similar to the one we derived in Section 2.3; the only difference is the convective term $\operatorname{div} cu$, which arises as the result of the large-scale flow field u.

Equation (6.6g) gives a direct illustration of the phenomenon of osmotic swelling: on the right-hand side, apart from the passive convection term $u \cdot n$, there is the balance

of the two pressures on the membrane. The osmotic pressure RTc forces the membrane outwards, and for convex shapes the curvature term $-\alpha H$ acts in the opposite direction.

Extensions. In the case of a vesicle surrounded by a lipid bilayer, the assumed linear dependence of the energy on the surface area is rather crude. Actual lipid bilayers can extend only very slightly, and a better assumption would be a dependence of the type $\alpha'(|\partial\Omega| - A_0)^2$. With this choice, the derivation would similar with only α replaced by $2\alpha'(|\partial\Omega| - A_0)$.

A further improvement could be to treat the lipid bilayer as a two-dimensional viscous incompressible fluid. In that case one would define the vectorial membrane velocity v_{mem}, replace the normal velocity v_n above by $v_{\mathrm{mem}} \cdot n$, and include a dissipation term of the form $\int_{\partial\Omega} |d|^2$, where d is an appropriate measure of in-plane strain (see [AD09]).

Limiting cases. It is instructive to consider setting constants to 0 or $+\infty$, and compare the effect (a) on the free energy and the dissipation potential, and (b) on the equation (6.6). For instance, if $\eta_b = \infty$, which means in (6.4) that trans-membrane fluxes require infinite amounts of dissipated energy, then indeed equation (6.6g) reduces to $v_n = u \cdot n$ and there is no flux through the boundary. If $\eta_b = 0$, on the other hand, implying that the fluid can move through the membrane without any friction, then equation (6.6g) reduces to the force balance $RTc = -\alpha H$. This latter boundary condition is known as the *Gibbs-Thomson law* [Bir08, Sec. 7.3.1.3], and this discussion therefore gives some insight into the type of modelling choices that lead to this law.

Similarly, if $\alpha = 0$, then the boundary can increase without limit, and indeed in (6.6e) and (6.6g) the curvature then no longer appears. Or if $\eta_u = \infty$, then the fluid is infinitely viscous, and indeed equations (6.6a), (6.6b), and (6.6e) give $u = 0$.

Chapter 7

Remarks and connections with thermodynamics

7.1 Boundary conditions

The examples in these notes all focus on boundary conditions for which the system is *closed*—in the sense that the boundary conditions do not add or take away energy. For other boundary conditions this will be different.

Boundary conditions in models can play two roles. First there is the type of boundary conditions that reflects the interaction with the part of the world that is not represented in the model. For such boundary conditions it is very hard to give general guidelines, since practically anything is possible, and in general the driving functional need no longer be decreasing along solutions.

The second type of boundary condition interfaces with another part of the model. Examples of this are two domains connected by their boundaries, or free- and moving-boundary problems. In this case the natural approach is to include the boundary and its properties in the model, in the same way as the other aspects. As an example, in Section 6.1 we consider a moving-boundary problem, in which the moving boundary is the wall of a vesicle. Fluid can move through the boundary, but in doing so it dissipates energy, and therefore the cross-boundary flow enters in the process vector and in the dissipation. As a result, boundary conditions at this moving interface arise in the derivation of the equations.

7.2 Free energy from a thermodynamic point of view

For this section [Sch07, Sec. 5.1] is a good reference.

7.2.1 The 'free' in free energy: *available* work

One thermodynamic situation in which 'free energy' arises is that of zero pressure and constant temperature. The traditional thermodynamic definition of *Gibbs free energy* in this context is $G = E - TS$, where E is the energy of the system, S the entropy (with the physically traditional sign) and T the temperature. This is the same formula as we found in Section 4.8.

Imagine that we create this system by starting with nothing, and slowly (quasistatically) adding material, performing work, and transferring heat, until we reach final energy E and final entropy S. Since we do this under constant temperature, the traditional thermodynamic relation $dQ = TdS$ for quasistatic evolutions implies that we need to transfer exactly TS heat from the surroundings into the system, if we normalize the empty system to have zero entropy. The remaining energy $E - TS$ then needs to be supplied in the form of work on the system, in such a way that the entropy does not increase. Therefore $E - TS$ is that amount of work that we need to do to create the system at energy E and entropy S, taking into account that the environment can supply heat at temperature T.

The word 'free' actually comes from performing the opposite operation, annihilating the system by performing work on the environment and transferring the entropy into the environment by heat transfer. 'Free' then refers to the fact that the transfer of entropy requires transfer of heat, and therefore 'costs' energy; the remaining energy $E - TS$ is 'available' to perform work on the environment.

7.2.2 Free energy as driving force

Section 4.5 explains the combination $E - TS$ in a different way: E/T is an approximation of the change in the entropy of the heat bath (or more generally, the change in the entropy of the rest of the universe) as a result of withdrawing energy E from the universe and putting it into the system. The sum $S - E/T$ then is the total energy of the system and the rest of the universe, i.e. of the whole universe.

This shows that entropy and free energy are really the same concept: the free energy is simply an approximation of the entropy of the universe taking into account only the behaviour of the system and the exchange of energy.

The question why free energy should be a driving force for any given process is therefore the same as the question why entropy should be one. For this question I know of only one answer, which is the one given in Chapters 4 and 5.

Appendix A

Elements of measure theory

Throughout these notes we work with a measurable space (Ω, Σ) consisting of a topological set Ω with a σ-algebra Σ that contains the Borel σ-algebra $\mathcal{B}(\Omega)$. We always assume that Ω is complete, metrizable, and separable. $\mathcal{M}(\Omega)$ is the set of all (finite or infinite) signed measures on $(\Omega, \mathcal{B}(\Omega))$, and $\mathscr{P}(\Omega)$ is the set of all non-negative measures $\mu \in \mathcal{M}(\Omega)$ with $\mu(\Omega) = 1$. For $\rho \in \mathscr{P}(\Omega)$, $L^2(\rho)$ is the set of measurable functions on Ω with finite $L^2(\rho)$-norm:

$$\|f\|_{L^2(\rho)}^2 := \int_\Omega f^2 \, d\rho.$$

A sequence of measures $\mu_n \in \mathcal{M}(\Omega)$ is said to converge *narrowly* to $\mu \in \mathcal{M}(\Omega)$ if

$$\int_\Omega f \, d\mu_n \overset{n\to\infty}{\longrightarrow} \int_\Omega f \, d\mu \qquad \text{for all } f \in C_b(\Omega).$$

The Lebesgue measure on $\mathbb{R}^d$ is indicated by $\mathcal{L}^d$.

The *push-forward* of a measure μ by a Borel measurable mapping $\varphi : \Omega \to \Omega$ is the measure $\varphi_\# \mu$ defined by

$$\varphi_\# \mu(A) := \mu(\varphi^{-1}(A)) \qquad \text{for any } A \in \mathcal{B}(\Omega).$$

It satisfies the identity

$$\int_\Omega f(y) \, \varphi_\# \mu(dy) = \int_\Omega f(\varphi(x)) \, \mu(dx) \qquad \text{for all } f \in C_b(\Omega).$$

Bibliography

[AB93]   G Anzellotti and S. Baldo. Asymptotic development by Γ-convergence. *Appl.
           Math. Optim.*, 27:105–123, 1993.

[AD09]   M. Arroyo and A. DeSimone. Relaxation dynamics of fluid membranes. *Physical
           Review E*, 79(3):031915, 2009.

[AdP06]   P. Atkins and J. de Paula. *Atkins' Physical Chemistry*. Oxford University Press,
           2006.

[ADPZ11]   S. Adams, N. Dirr, M. A. Peletier, and J. Zimmer. From a large-deviations
           principle to the Wasserstein gradient flow: A new micro-macro passage. *Com-
           munications in Mathematical Physics*, 307:791–815, 2011.

[AGS05]   L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows in Metric Spaces and
           in the Space of Probability Measures*. Lectures in mathematics ETH Zürich.
           Birkhäuser, 2005.

[BB00]   J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the
           Monge-Kantorovich mass transfer problem. *Numer. Math.*, 84:375–393, 2000.

[BHP94]   L. Bocquet, J.-P. Hansen, and J. Piasecki. On the brownian motion of a massive
           sphere suspended in a hard-sphere fluid. ii. molecular dynamics estimates of the
           friction coefficient. *Journal of statistical physics*, 76(1-2):527–548, 1994.

[Bir08]   K. S. Birdi. *Handbook of Surface and Colloid Chemistry*. CRC Press, 2008.

[Csi67]   I. Csiszár. Information-type measures of difference of probability distributions
           and indirect observations. *Studia Sci. Math. Hungar.*, 2:299–318, 1967.

[DG87]   D. A. Dawson and J. Gartner. Large deviations from the McKean-Vlasov limit
           for weakly interacting diffusions. *Stochastics*, 20(4):247–308, 1987.

[dH00]   F. den Hollander. *Large Deviations*. American Mathematical Society, Provi-
           dence, RI, 2000.

[DLR12]   M. H. Duong, V. Laschos, and D. R. M. Renger. Wasserstein gradient flows
           from large deviations of thermodynamic limits. 2012. Submitted.

[DZ98]   A. Dembo and O. Zeitouni. *Large deviations techniques and applications.* Springer Verlag, 1998.

[Eva01]   L. C. Evans. Entropy and partial differential equations. Technical report, UC Berkeley, 2001.

[FK06]   J. Feng and T. G. Kurtz. *Large deviations for stochastic processes*, volume 131 of *Mathematical Surveys and Monographs*. American Mathematical Society, 2006.

[Hin91]   E. J. Hinch. *Perturbation Methods.* Cambridge University Press, 1991.

[JKO98]   R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker-Planck Equation. *SIAM Journal on Mathematical Analysis*, 29(1):1–17, 1998.

[Kul67]   S. Kullback. A lower bound for discrimination information in terms of variation. *IEEE Transactions on Information Theory*, 13(1):126–127, 1967.

[Léo12]   C. Léonard. From the schrödinger problem to the monge–kantorovich problem. *Journal of Functional Analysis*, 262(4):1879–1920, 2012.

[Li09]   Zhigang Li. Critical particle size where the Stokes-Einstein relation breaks down. *Physical Review E*, 80(6):061204, 2009.

[LL87]   L. D. Landau and E. M. Lifshitz. Fluid mechanics, vol. 6. *Course of Theoretical Physics*, pages 227–229, 1987.

[LY97]   E. H. Lieb and J. Yngvason. A guide to entropy and the second law of thermodynamics. *Notices of the AMS*, 45(5), 1997.

[Mar12]   M. Mariani. A Gamma-convergence approach to large deviations. *Arxiv preprint arXiv:1204.0640*, 2012.

[MPR]   A. Mielke, M. A. Peletier, and D. R. M. Renger. Gradient flows, large deviation, and the role of detailed balance, with applications to Markov chains and diffusion. In preparation.

[Ott01]   F. Otto. The geometry of dissipative evolution equations: The porous medium equation. *Communications in Partial Differential Equations*, 26:101–174, 2001.

[Ött05]   H. C. Öttinger. *Beyond Equilibrium Thermodynamics.* Wiley-Interscience, 2005.

[PR11]   M. A. Peletier and D. R. M. Renger. Variational formulation of the Fokker-Planck equation with decay: A particle approach. *Arxiv preprint arXiv:1108.3181*, 2011.

[Sch07]   D. V. Schroeder. *An Introduction to Thermal Physics*. Pearson Education India, 2007.

[Sto51]   G. G. Stokes. *On the effect of the internal friction of fluids on the motion of pendulums*, volume 9. Pitt Press, 1851.