2

Three examples of coupling techniques

In this chapter I shall present three applications of coupling methods. The first one is classical and quite simple, the other two are more original but well-representative of the topics that will be considered later in these notes. The proofs are extremely variable in difficulty and will only be sketched here; see the references in the bibliographical notes for details.

Convergence of the Langevin process

Consider a particle subject to the force induced by a potential $V \in C^1(\mathbb{R}^n)$, a friction and a random white noise agitation. If X_t stands for the position of the particle at time t, m for its mass, λ for the friction coefficient, k for the Boltzmann constant and T for the temperature of the heat bath, then Newton's equation of motion can be written

$$m \frac{d^2 X_t}{dt^2} = -\nabla V(X_t) - \lambda \, m \, \frac{dX_t}{dt} + \sqrt{kT} \, \frac{dB_t}{dt}, \qquad (2.1)$$

where $(B_t)_{t \geq 0}$ is a standard Brownian motion. This is a second-order (stochastic) differential equation, so it should come with initial conditions for both the position X and the velocity $\dot{X}$.

Now consider a large cloud of particles evolving independently, according to (2.1); the question is whether the distribution of particles will converge to a definite limit as $t \to \infty$. In other words: Consider the stochastic differential equation (2.1) starting from some initial distribution $\mu_0(dx \, dv) = \text{law} \, (X_0, \dot{X}_0)$; is it true that law (X_t), or law $(X_t, \dot{X}_t)$, will converge to some given limit law as $t \to \infty$?

Obviously, to solve this problem one has to make some assumptions on the potential V, which should prevent the particles from all escaping at infinity; for instance, we can make the very strong assumption that V is uniformly convex, i.e. there exists $K > 0$ such that the Hessian $\nabla^2 V$ satisfies $\nabla^2 V \geq K I_n$. Some assumptions on the initial distribution might also be needed; for instance, it is natural to assume that the Hamiltonian has finite expectation at initial time:

$$\mathbb{E}\left(V(X_0) + \frac{|\dot{X}_0|^2}{2}\right) < +\infty$$

Under these assumptions, it is true that there is exponential convergence to equilibrium, at least if V does not grow too wildly at infinity (for instance if the Hessian of V is also bounded above). However, I do not know of any simple method to prove this.

On the other hand, consider the limit where the friction coefficient is quite strong, and the motion of the particle is so slow that the acceleration term may be neglected in front of the others: then, up to resetting units, equation (2.1) becomes

$$\frac{dX_t}{dt} = -\nabla V(X_t) + \sqrt{2}\,\frac{dB_t}{dt}, \tag{2.2}$$

which is often called a Langevin process. Now, to study the convergence of equilibrium for (2.2) there is an extremely simple solution by coupling. Consider another random position $(Y_t)_{t \geq 0}$ obeying the same equation as (2.2):

$$\frac{dY_t}{dt} = -\nabla V(Y_t) + \sqrt{2}\,\frac{dB_t}{dt}, \tag{2.3}$$

where the random realization of the Brownian motion is *the same* as in (2.2) (this is the coupling). The initial positions X_0 and Y_0 may be coupled in an arbitrary way, but it is possible to assume that they are independent. In any case, since they are driven by the same Brownian motion, X_t and Y_t will be correlated for $t > 0$.

Since B_t is not differentiable as a function of time, neither X_t nor Y_t is differentiable (equations (2.2) and (2.3) hold only in the sense of solutions of stochastic differential equations); but it is easily checked that $\alpha_t := X_t - Y_t$ is a continuously differentiable function of time, and

$$\frac{d\alpha_t}{dt} = -\big(\nabla V(X_t) - \nabla V(Y_t)\big),$$

so in particular

$$\frac{d}{dt}\frac{|\alpha_t|^2}{2} = -\left\langle \nabla V(X_t) - \nabla V(Y_t),\ X_t - Y_t \right\rangle \leq -K\left|X_t - Y_t\right|^2 = -K\left|\alpha_t\right|^2.$$

It follows by Gronwall's lemma that

$$|\alpha_t|^2 \leq e^{-2Kt}\,|\alpha_0|^2.$$

Assume for simplicity that $\mathbb{E}\,|X_0|^2$ and $\mathbb{E}\,|Y_0|^2$ are finite. Then

$$\mathbb{E}\,|X_t - Y_t|^2 \leq e^{-2Kt}\,\mathbb{E}\,|X_0 - Y_0|^2 \leq 2\left(\mathbb{E}\,|X_0|^2 + \mathbb{E}\,|Y_0|^2\right)e^{-2Kt}. \quad (2.4)$$

In particular, $X_t - Y_t$ converges to 0 almost surely, and this is independent of the distribution of Y_0.

This in itself would be essentially sufficient to guarantee the existence of a stationary distribution; but in any case, it is easy to check, by applying the diffusion formula, that

$$\nu(dy) = \frac{e^{-V(y)}\,dy}{Z}$$

(where $Z = \int e^{-V}$ is a normalization constant) is stationary: If $\mathrm{law}\,(Y_0) = \nu$, then also $\mathrm{law}\,(Y_t) = \nu$. Then (2.4) easily implies that $\mu_t := \mathrm{law}\,(X_t)$ converges weakly to ν; in addition, the convergence is exponentially fast.

Euclidean isoperimetry

Among all subsets of $\mathbb{R}^n$ with given surface, which one has the largest volume? To simplify the problem, let us assume that we are looking for a bounded open set $\Omega \subset \mathbb{R}^n$ with, say, Lipschitz boundary $\partial\Omega$, and that the measure of $|\partial\Omega|$ is given; then the problem is to maximize the measure of $|\Omega|$. To measure $\partial\Omega$ one should use the $(n-1)$-dimensional Hausdorff measure, and to measure Ω the n-dimensional Hausdorff measure, which of course is the same as the Lebesgue measure in $\mathbb{R}^n$.

It has been known, at least since ancient times, that the solution to this "isoperimetric problem" is the ball. A simple scaling argument shows that this statement is equivalent to the Euclidean **isoperimetric inequality**:

$$\frac{|\partial \Omega|}{|\Omega|^{\frac{n}{n-1}}} \geq \frac{|\partial B|}{|B|^{\frac{n}{n-1}}},$$

where B is any ball.

There are very many proofs of the isoperimetric inequality, and many refinements as well. It is less known that there is a proof by coupling.

Here is a sketch of the argument, forgetting about regularity issues. Let B be a ball such that $|\partial B| = |\partial \Omega|$. Consider a random point X distributed uniformly in Ω, and a random point Y distributed uniformly in B. Introduce the Knothe–Rosenblatt coupling of X and Y: This is a deterministic coupling of the form $Y = T(X)$, such that, at each $x \in \Omega$, the Jacobian matrix $\nabla T(x)$ is triangular with nonnegative diagonal entries. Since the law of X (resp. Y) has uniform density $1/|\Omega|$ (resp. $1/|B|$), the change of variables formula yields

$$\forall x \in \Omega \qquad \frac{1}{|\Omega|} = \big(\det \nabla T(x)\big) \frac{1}{|B|}. \tag{2.5}$$

Since ∇T is triangular, the Jacobian determinant of T is $\det(\nabla T) = \prod \lambda_i$, and its divergence $\nabla \cdot T = \sum \lambda_i$, where the nonnegative numbers $(\lambda_i)_{1 \leq i \leq n}$ are the eigenvalues of ∇T. Then the arithmetic–geometric inequality $(\prod \lambda_i)^{1/n} \leq (\sum \lambda_i)/n$ becomes

$$\big(\det \nabla T(x)\big)^{1/n} \leq \frac{\nabla \cdot T(x)}{n}.$$

Combining this with (2.5) results in

$$\frac{1}{|\Omega|^{1/n}} \leq \frac{(\nabla \cdot T)(x)}{n \, |B|^{1/n}}.$$

Integrate this over Ω and then apply the divergence theorem:

$$|\Omega|^{1-\frac{1}{n}} \leq \frac{1}{n \, |B|^{\frac{1}{n}}} \int_\Omega (\nabla \cdot T)(x) \, dx = \frac{1}{n \, |B|^{\frac{1}{n}}} \int_{\partial\Omega} (T \cdot \sigma) \, d\mathcal{H}^{n-1}, \tag{2.6}$$

where $\sigma : \partial\Omega \to \mathbb{R}^n$ is the unit outer normal to Ω and $\mathcal{H}^{n-1}$ is the $(n-1)$-dimensional Hausdorff measure (restricted to $\partial\Omega$). But T is valued in B, so $|T \cdot \sigma| \leq 1$, and (2.6) implies

$$|\Omega|^{1-\frac{1}{n}} \leq \frac{|\partial\Omega|}{n \, |B|^{\frac{1}{n}}}.$$

Since $|\partial\Omega| = |\partial B| = n|B|$, the right-hand side is actually $|B|^{1-\frac{1}{n}}$, so the volume of Ω is indeed bounded by the volume of B. This concludes the proof.

The above argument suggests the following problem:

Open Problem 2.1. *Can one devise an optimal coupling between sets (in the sense of a coupling between the uniform probability measures on these sets) in such a way that the total cost of the coupling decreases under some evolution converging to balls, such as mean curvature motion?*

Caffarelli's log-concave perturbation theorem

The previous example was about transporting a set to another, now the present one is in some sense about transporting a whole space to another.

It is classical in geometry to compare a space $\mathcal{X}$ with a "model space" $\mathcal{M}$ that has nice properties and is, e.g., less curved than $\mathcal{X}$. A general principle is that certain inequalities which hold true on the model space can automatically be "transported" to $\mathcal{X}$. The theorem discussed in this section is a striking illustration of this idea.

Let F, G, H, J, L be nonnegative continuous functions on $\mathbb{R}$, with H and J nondecreasing, and let $\ell \in \mathbb{R}$. For a given measure μ on $\mathbb{R}^n$, let $\lambda[\mu]$ be the largest $\lambda \geq 0$ such that, for all Lipschitz functions $h : \mathbb{R}^n \to \mathbb{R}$,

$$\int_{\mathbb{R}^n} L(h)\, d\mu = \ell \quad \Longrightarrow \quad F\left(\int_{\mathbb{R}^n} G(h)\, d\mu\right) \leq \frac{1}{\lambda}\, H\left(\int_{\mathbb{R}^n} J(|\nabla h|)\, d\mu\right).$$
$$(2.7)$$

Functional inequalities of the form (2.7) are variants of Sobolev inequalities; many of them are well-known and useful. Caffarelli's theorem states that *they can only be improved by log-concave perturbation of the Gaussian distribution*. More precisely, if γ is the standard Gaussian measure and $\mu = e^{-v}\gamma$ is another probability measure, with v convex, then

$$\lambda[\mu] \geq \lambda[\gamma].$$

His proof is a simple consequence of the following remarkable fact, which I shall call **Caffarelli's log-concave perturbation theorem**: *If $d\mu/d\gamma$ is log-concave, then there exists a 1-Lipschitz change*

of variables from the measure γ to the measure μ. In other words, there is a deterministic coupling $\big(X,\ Y = \mathcal{C}(X)\big)$ of (γ, μ), such that $|\mathcal{C}(x) - \mathcal{C}(y)| \leq |x - y|$, or equivalently $|\nabla \mathcal{C}| \leq 1$ (almost everywhere). It follows in particular that

$$\big|\nabla(h \circ \mathcal{C})\big| \leq |(\nabla h) \circ \mathcal{C}|, \tag{2.8}$$

whatever the function h.

Now it is easy to understand why the existence of the map $\mathcal{C}$ implies (2.7): On the one hand, the definition of change of variables implies

$$\int G(h)\, d\mu = \int G(h \circ \mathcal{C})\, d\gamma, \qquad \int L(h)\, d\mu = \int L(h \circ \mathcal{C})\, d\gamma;$$

on the other hand, by the definition of change of variables again, inequality (2.8) and the nondecreasing property of J,

$$\int J(|\nabla h|)\, d\mu = \int J\big(|\nabla h \circ \mathcal{C}|\big)\, d\gamma \geq \int J\big(|\nabla(h \circ \mathcal{C})|\big)\, d\gamma.$$

Thus, inequality (2.7) is indeed "transported" from the space $(\mathbb{R}^n, \gamma)$ to the space $(\mathbb{R}^n, \mu)$.

Bibliographical notes

It is very classical to use coupling arguments to prove convergence to equilibrium for stochastic differential equations and Markov chains; many examples are described by Rachev and Rüschendorf [696] and Thorisson [781]. Actually, the standard argument found in textbooks to prove the convergence to equilibrium for a positive aperiodic ergodic Markov chain is a coupling argument (but the null case can also be treated in a similar way, as I learnt from Thorisson). Optimal couplings are often well adapted to such situations, but definitely not the only ones to apply.

The coupling method is not limited to systems of independent particles, and sometimes works in presence of correlations, for instance if the law satisfies a nonlinear diffusion equation. This is exemplified in works by Tanaka [777] on the spatially homogeneous Boltzmann equation with Maxwell molecules (the core of Tanaka's argument is reproduced in my book [814, Section 7.5]), or some recent papers [138, 214, 379, 590].

Cattiaux and Guillin [221] found a simple and elegant coupling argument to prove the exponential convergence for the law of the stochastic process

$$dX_t = \sqrt{2}\, dB_t - \widetilde{\mathbb{E}}\, \nabla V(X_t - \widetilde{X}_t)\, dt,$$

where $\widetilde{X}_t$ is an independent copy of X_t, the $\widetilde{\mathbb{E}}$ expectation only bears on $\widetilde{X}_t$, and V is assumed to be a uniformly convex C^1 potential on $\mathbb{R}^n$ satisfying $V(-x) = V(x)$.

It is also classical to couple a system of particles with an auxiliary artificial system to study the limit when the number of particles becomes large. For the Vlasov equation in kinetic theory this was done by Dobrushin [309] and Neunzert [653] several decades ago. (The proof is reproduced in Spohn [757, Chapter 5], and also suggested as an exercise in my book [814, Problem 14].) Later Sznitman used this strategy in a systematic way for the propagation of chaos, and made it very popular, see e.g. his work on the Boltzmann equation [767] or his Saint-Flour lecture notes [768] and the many references included.

In all these works, the "philosophy" is always the same: Introduce some nice coupling and see how it evolves in a certain asymptotic regime (say, either the time, or the number of particles, or both, go to infinity).

It is possible to treat the convergence to equilibrium for the complete system (2.1) by methods that are either analytic [301, 472, 816, 818] or probabilistic [55, 559, 606, 701], but all methods known to me are much more delicate than the simple coupling argument which works for (2.2). It is certainly a nice open problem to find an elementary coupling argument which applies to (2.1). (The arguments in the above-mentioned probabilistic proofs ultimately rely on coupling methods via theorems of convergence for Markov chains, but in a quite indirect way.)

Coupling techniques have also been used recently for proving rather spectacular uniqueness theorems for invariant measures in infinite dimension, see e.g. [321, 456, 457].

Classical references for the isoperimetric inequality and related topics are the books by Burago and Zalgaller [176], and Schneider [741]; and the survey by Osserman [664]. Knothe [523] had the idea to use a "coupling" method to prove geometric inequalities, and Gromov [635, Appendix] applied this method to prove the Euclidean isopetrimetric inequality. Trudinger [787] gave a closely related treatment of the same inequality and some of its generalizations, by means of a clever use of the Monge–Ampère equation (which more or less amounts to the construction of an optimal coupling with quadratic cost function, as will

be seen in Chapter 11). Cabré [182] found a surprising simplification
of Trudinger's method, based on the solution of just a linear elliptic
equation. The "proof" which I gave in this chapter is a variation on
Gromov's argument; although it is not rigorous, there is no real diffi-
culty in turning it into a full proof, as was done by Figalli, Maggi and
Pratelli [369]. These authors actually prove much more, since they use
this strategy to establish a sharp quantitative *stability* of the isoperi-
metric inequality (if the shape of a set departs from the optimal shape,
then its isoperimetric ratio departs from the optimal ratio in a quantifi-
able way). In the same work one can find a very interesting comparison
of the respective performances of the couplings obtained by the Knothe
method and by the optimal transport method (the comparison turns
very much to the advantage of optimal transport).

Other links between coupling and isoperimetric-type inequalities are
presented in Chapter 6 of my book [814], the research paper [587],
the review paper [586] and the bibliographical notes at the end of
Chapters 18 and 21.

The construction of Caffarelli's map $\mathcal{C}$ is easy, at least conceptually:
The optimal coupling of the Gaussian measure γ with the measure $\mu =
e^{-v}\gamma$, when the cost function is the square of the Euclidean distance,
will do the job. But proving that $\mathcal{C}$ is indeed 1-Lipschitz is much more of
a sport, and involves some techniques from nonlinear partial differential
equations [188]. An idea of the core of the proof is explained in [814,
Problem 13]. It would be nice to find a softer argument.

Üstünel pointed out to me that, if v is convex and symmetric
$(v(-x) = v(x))$, then the Moser transport T from γ to $e^{-v}\gamma$ is con-
tracting, in the sense that $|T(x)| \leq |x|$; it is not clear however that T
would be 1-Lipschitz.

Caffarelli's theorem has many analytic and probabilistic applica-
tions, see e.g. [242, 413, 465]. There is an infinite-dimensional version
by Feyel and Üstünel [361], where the Gaussian measure is replaced
by the Wiener measure. Another variant was recently studied by
Valdimarsson [801].

Like the present chapter, the lecture notes [813], written for a CIME
Summer School in 2001, present some applications of optimal transport
in various fields, with a slightly impressionistic style.