

Parabolic equations and ergodicity

Luca Tamanini

Université Paris Ouest Nanterre La Défense
SISSA

August 4, 2015

Contents

1	Lecture 1 - 7 November	2
2	Lecture 2 - 14 November	18
3	Lecture 3 - 21 November	28
4	Lecture 4 - 28 November	28
5	Lecture 5 - 5 December	28
6	Lecture 6 - 19 December	28
7	Lecture 7 - 9 January	28
8	Lecture 8 - 16 January	28
9	Lecture 9 - 6 February	28
A	Auxiliary results	28
A.1	Liouville's theorem and problem	28
A.2	Tauberian theorems	28
A.3	A couple of theorems	28
A.4	The Hille-Yosida theorem	28

These notes are based on the course *Équations paraboliques et ergodicité* held by Pierre-Louis Lions at the Collège de France in Paris during the academic year 2014-15. The structure of the course is the following:

- (0) ergodicity;
- (1) necessary and sufficient conditions for ergodicity;
- (2) convergence rate and Poisson problem (also known as cell problem);
- (3) properties of *the* invariant measure (indeed, we will see that ergodicity entails the uniqueness of the invariant measure);
- (4) extensions (time dependence, weak regularity, etc.);
- (5) non-linear extensions (in the sense of PDEs); the generalizations we will consider are essentially two: on the one hand, we will extend the parabolic-type equations for the laws of stochastic processes to Bellman-type equations (ergodic control) and, on the other hand, we will extend to non-linear cases the Fokker-Planck equation (describing the propagation of the law of a stochastic process) and its “adjoint”, both playing a crucial role in the study of asymptotic behaviours in the theory of Mean Field Games (MFG).

1 Lecture 1 - 7 November

Abstract

In this first lecture, we present an intuitive introduction to ergodic theory, by first providing several examples. The study of the different behaviours appearing in these problems will lead us to two possible definitions of ergodicity and to the problem of characterizing these notions, in order to check them concretely. By the way, several digressions on auxiliary items and multidisciplinary motivations will be provided. At the end of the lecture, the problems we are really interested in will be formulated.

In this lecture we will first focus our attention on the word “ergodicity”, because we can find it in many different domains (such as analysis, probability and dynamical systems) and for this reason half a dozen of different definitions can be found (uniform ergodicity, geometric ergodicity, etc.). Hence, we will try to clarify all these aspects by means of analytical methods, which have also a strong probabilistic interpretation and apply to asymptotic behaviour of Markovian processes. Indeed, we will mainly work with diffusion processes, but the essential aspects of the methods we will adopt fit also to jump processes, associated to integro-differential operators. A very good reference for the discrete time case (i.e. for Markov chains) is [4], a classic of the literature in probability; on the contrary, in the literature on PDEs only some comments are spent on ergodicity. Nevertheless, in this course the analytic point of view, together with the probabilistic one, will be predominant. On the contrary, we will rarely deal with ergodic theory from the dynamical systems viewpoint.

Let us now begin the discussion on ergodicity in an informal way. Let us simply consider a deterministic or stochastic system (e.g. the evolution of a particle) described by a differential system (again, deterministic or stochastic) and evolving within a certain domain D , possibly the whole space. An ergodic phenomenon appears when there exists a sub-region of D such that the process approaches it in finite time and then the evolution of our system is essentially contained in such sub-region and essentially explores all of its points. In this case, “essentially” means that the evolution of the system is contained in a neighbourhood of the sub-region and comes sufficiently near to every point of the same sub-region. Thus, if the sub-region is a curve, then the process always stays within a tubular neighbourhood of such curve and comes very near to every point of it. Note that the sub-region having such ergodic property is possibly the whole domain or, on the contrary, may

reduce to a single point. After this qualitative description, a second way to present ergodicity is the following obscure statement: “time averages become space averages”; we will see later what this concretely means.

Let us also say that in this course we will often refer to examples in the periodic case and more precisely on the torus; from a PDE point of view, this amounts to consider differential equations on a periodic domain. In more details, some of the examples we will consider are the following.

Example 1.1. Let Q be a periodic domain (think to the unit cube) and let us describe the behaviour of the Brownian motion on it. This amounts to consider the heat equation

$$\begin{cases} \frac{\partial u}{\partial t} - \frac{1}{2}\Delta u = 0 \\ u|_{t=0} = u_0 \end{cases} \quad (1.1)$$

so that the natural probabilistic representation

$$u(x, t) = \mathbb{E}[u_0(x + W_t)]$$

for the solution to (1.1) holds (W_t is the standard Brownian motion with zero mean and unit variance). In the case of a d -valued Brownian motion, it is well-known that the process goes to infinity; in the case $Q = \mathbb{R}^d/\mathbb{Z}^d$, i.e. Q is the unit cube, our domain is bounded and therefore it makes no sense to say that the process goes to infinity. What happens is this: two points differing by integer coordinates are identified and therefore at any time the Brownian motion is essentially and uniformly everywhere. This means that the process evolves in any direction and, thanks to the identification modulo \mathbb{Z}^d , it returns sufficiently near to any point, without privileged areas. Thus, the law of the Brownian motion (exponentially) converges to a uniform distribution on Q , i.e.

$$u(x, t) \rightarrow \int_Q u_0$$

where the convergence is uniform in x ; in a more precise way

$$\lim_{t \rightarrow +\infty} \left\| u(x, t) - \int_Q u_0 \right\|_{Q, \infty} = 0 \quad (1.2)$$

This can be seen by decomposition of $u(x, t)$ in its Fourier series: the non-constant modes exponentially decay, whereas only the constant mode survives. In more details, if the wavelength of a non-constant mode is k , then such mode decays in time with speed $\exp(-k^2/2t)$, so that the uniform convergence (1.2) is dominated by an explicit exponential, which corresponds to the second eigenvalue of the Laplacian on Q , namely the first non-zero natural frequency (fréquence propre). Thus, in this first example the law of the process becomes stable and converges to a uniform distribution. \diamond

Example 1.2. Let us still consider a periodic framework, even if 1-dimensional, and replace the previous diffusion problem by a deterministic one. Let $Q = [0, 1]/\sim$, where \sim is the equivalence relation identifying 0 and 1, and on this domain look at the first order PDE

$$\frac{\partial u}{\partial t} - \frac{\partial u}{\partial x} = 0$$

In this case, the solution is explicitly given by $u(x, t) = u_0(x + t)$ for any C^1 function u_0 ; this means that the initial profile u_0 moves with constant speed along the interval and when it reaches the end, it starts again at the beginning of the interval. Thus, we explore in a systematic and very regular way the whole domain Q and this is coherent with the intuitive description we gave of ergodicity.

But on the other hand, a behaviour like the one in (1.2) is not possible. More generally, it is not even possible any uniform convergence of $u(x, t)$ to something; think for instance to $u_0(x) = \sin(x)$: in this case $u(x, t)$ has the same profile of u_0 , because it is the shifted sine, and so no damping happens. Nevertheless, we can still say that

$$\frac{1}{t} \int_0^t u(x, s) ds \rightarrow \int_0^1 u_0(y) dy \quad (1.3)$$

and also in this case the convergence is uniform, because u_0 is bounded. Observe that in (1.3) we actually have a time average transforming into a space average, but we will better explain this phenomenon later. \diamond

In these examples we have seen two “ergodic” behaviours, in the sense that they match with what we said in the heuristic introductory part. However, from a mathematical point of view, they can not be formulated in the same manner, because of the different type of asymptotic stabilization: (1.2) vs. (1.3). In the next example, we are going to mix the two different behaviours and we will investigate which one is prevailing.

Example 1.3. Let $Q = \mathbb{R}^2/\mathbb{Z}^2$ and on this periodic domain let us consider the following PDE:

$$\frac{\partial u}{\partial t} - \frac{\partial u}{\partial x} - \frac{1}{2} \frac{\partial^2 u}{\partial y^2} = 0 \quad (1.4)$$

What happens is clear: pure transport in the x -direction and pure diffusion in the y -one; however, which is the predominant behaviour? If we take a solution u independent of y , we see that such solution solves a transport equation and therefore a (1.2)-like stabilization is impossible; this means that, also in general, the transport aspect prevails for any solution to (1.4) and the ergodicity type is (1.3).

Now let us modify this example by applying a rotation by an angle ϑ to Q and leaving (1.4) unchanged. This amounts to consider the periodicity w.r.t. a different lattice. In an equivalent way, we can represent the rotation as follows: we leave Q unchanged and we modify the first order part in (1.4) by adding a first order partial derivative in the y -direction. More explicitly, we rewrite (1.4) as

$$\frac{\partial u}{\partial t} - \frac{\partial u}{\partial x} - \alpha \frac{\partial u}{\partial y} - \frac{1}{2} \frac{\partial^2 u}{\partial y^2} = 0$$

for a suitable coefficient α depending on ϑ . The situation becomes much more interesting and two cases are possible:

- if $\alpha \in \mathbb{Q}$, then the asymptotic behaviour is essentially the same we have in the case of pure transport and therefore we need a Cesare-type average for the stabilization;
- if $\alpha \notin \mathbb{Q}$, then the situation is the same we have for the Brownian motion and so the system spreads all over the cube.

Note that α can be arbitrarily small and thus for arbitrarily small rotations of Q the qualitative behaviour of the system is completely different from the limit case with no rotation: this means that ergodicity is a pure matter of periodicity. \diamond

This example shows that very similar systems may have completely different asymptotic behaviours and for this reason we would like to be able to distinguish them, i.e. we would like to find necessary and sufficient conditions for the first and the second type of ergodicity, at least in periodic domains.

Now let us spend some words on the obscure statement we said above, namely “time averages

become space averages”. Let us assume that our system is described by a differential or integro-differential equation, associated to a given Markov process; then, the solution of the equation for a given initial condition u_0 can be explicitly written as

$$u(x, t) = \mathbb{E}[u_0(X_t) | X_0 = x] \quad (1.5)$$

and this gives a 1-parameter semigroup $(P_t)_{t \geq 0}$ such that $u(x, t) = (P_t u_0)(x)$. With this premise, the change of time averages into spatial ones can be expressed in a first way as follows: for a sufficiently large class of test functions χ ,

$$\frac{1}{t} \int_0^t \chi(X_s^x) ds \rightarrow \int \chi d\mathbf{m} \quad (1.6)$$

in probability, where the apex x in X_s^x recalls us the initial condition, that is $X_0 = x$, and \mathbf{m} is the invariant measure, roughly speaking the distribution at an infinite time. Note that if χ is the characteristic function of a certain set E , then this property tells us that the mean time the process spends in E converges to the volume of E w.r.t. the invariant measure \mathbf{m} . The convergence

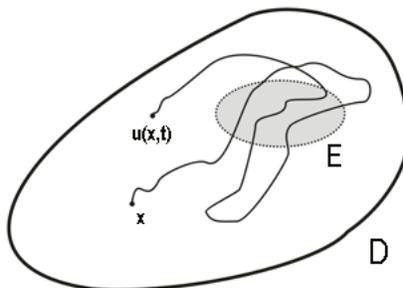


Figure 1: Time spent by the process in E .

in probability can be refined and by technical arguments (via maximal functions in analysis and via martingale-type arguments in probability) it is possible to obtain a.e. convergence. On the other hand, there is a second way to express the fact that “time averages become space averages” and this is the following

$$\mathbb{E}[\chi(X_t) | X_0 = x] \rightarrow \int \chi d\mathbf{m} \quad (1.7)$$

or, in a weaker form and still involving time Cesaro means,

$$\frac{1}{t} \int_0^t \mathbb{E}[\chi(X_s) | X_0 = x] ds \rightarrow \int \chi d\mathbf{m} \quad (1.8)$$

Thus, the class of test functions χ is nothing but the class of admissible initial conditions for the PDE describing our system and this last condition means that the time Cesaro mean of the solution to the PDE describing the system converges to the spatial mean (w.r.t. the invariance measure) of the initial datum. In a while we will see that, under the only hypothesis that the process (X_t) is Markovian (so that we have the semigroup property), (1.8) entails (1.6). For the moment, let us point out that in this discussion we have been a little superficial, because we have not made precise if a stronger version of (1.6) is possible, as it is for (1.8). The answer is given in the next remark.

Remark 1.4. In the previous discussion about time and spatial averages, the time Cesaro mean in (1.6) is necessary. Indeed, even if we assume the strongest stabilization behaviour, that is (1.7), we

can not say that

$$\chi(X_t^x) \rightarrow \int \chi \, d\mathbf{m}$$

in probability as $t \rightarrow +\infty$. Actually, this is almost never true, because a non-zero variance is involved. In fact, recall that $\text{Var}(\chi) = \mathbb{E}(\chi^2) - \mathbb{E}(\chi)^2$ and observe that, by virtue of (1.7) and so by the fact that the solution to the PDE converges to a constant function as $t \rightarrow +\infty$, we have

$$\text{Var}(\chi(X_t^x)) \rightarrow \int \chi^2 \, d\mathbf{m} - \left(\int \chi \, d\mathbf{m} \right)^2$$

However, the right-hand side above is strictly positive in general. The only way to have a zero right-hand side is that \mathbf{m} is a Dirac measure and from an intuitive point of view this means that there is a (unique) point of attraction, i.e. a point such that the whole system evolves towards it. The evolution may happen also with noise, but in this case the noise is singular at the attraction point. \diamond

The fact that we can not forget the time Cesaro mean in (1.6) can be seen as a sort of decorrelation in time of the process and the time Cesaro mean $t^{-1} \int_0^t$ can be regarded as a sort of law of large numbers, as we will better see later, because thanks to the decorrelation property we can really think to $t^{-1} \int_0^t$ as an averaged sum of independent random variables. Furthermore, we will also study the following question: if we have the law of large numbers, under which assumptions do we have the central limit theorem too? The interest in this question is due to the fact that the central limit theorem is strictly connected to the convergence rate (point (2) of the summary).

Now let us talk about the equivalence between (1.6) and (1.8). Indeed, we have already anticipated the fact that, under Markov hypothesis, (1.8) implies (1.6), but also the converse implication is true and it is also the easier one to prove (in the case χ is bounded, it is really straightforward). For this reason, ergodicity is often defined by (1.6): it seems a stronger notion; however, *a posteriori* it is not. In addition, the choice of (1.6) as definition of ergodicity is not the best one, because it hides the fundamental aspect of time Cesaro means: in (1.6) such average is necessary and it is not possible an alternative version without it; on the contrary, (1.7) and (1.8) allow us to appreciate the relevance of time Cesaro means. After these theoretical considerations, let us show in a formal way that (1.8) entails (1.6).

Proof. Since we already know by hypothesis that

$$\frac{1}{t} \int_0^t \mathbb{E}[\chi(X_s^x)] \, ds \rightarrow \int \chi \, d\mathbf{m} \tag{1.9}$$

it is sufficient to show that

$$\mathbb{E} \left[\left(\frac{1}{t} \int_0^t \chi(X_s^x) \, ds \right)^2 \right] \rightarrow \left(\int \chi \, d\mathbf{m} \right)^2$$

because, coupled with (1.9), this fact entails that

$$\text{Var} \left(\frac{1}{t} \int_0^t \chi(X_s^x) \, ds \right) \rightarrow 0$$

and this is really the convergence in probability we want to prove. Hence, let us assume for sake of simplicity that convergence in (1.9) is uniform in x and let us observe that

$$\frac{1}{t^2} \mathbb{E} \left[\int_0^t \int_0^t \chi(X_{s_1}^x) \chi(X_{s_2}^x) \, ds_1 \, ds_2 \right] = \frac{2}{t^2} \mathbb{E} \left[\int_0^t \left(\int_0^{s_1} \chi(X_{s_1}^x) \chi(X_{s_2}^x) \, ds_2 \right) \, ds_1 \right]$$

$$\begin{aligned}
&= \frac{2}{t^2} \mathbb{E} \left[\int_0^t \left(\int_0^{s_1} \mathbb{E}[\chi(X_{s_1-s_2}) | X_{s_2}^x] \chi(X_{s_2}^x) ds_2 \right) ds_1 \right] \\
&= \frac{2}{t^2} \mathbb{E} \left[\int_0^t \left(\int_{s_2}^t \mathbb{E}[\chi(X_{s_1-s_2}) | X_{s_2}^x] ds_1 \right) \chi(X_{s_2}^x) ds_2 \right] \\
&= \frac{2}{t^2} \mathbb{E} \left[\int_0^t \left(\int_0^{t-s_2} \mathbb{E}[\chi(X_\tau) | X_{s_2}^x] d\tau \right) \chi(X_{s_2}^x) ds_2 \right] \\
&\approx \frac{2}{t^2} \mathbb{E} \left[\int_0^t (t-s_2) \underbrace{\int \chi d\mathbf{m}}_{c_0} \chi(X_{s_2}^x) ds_2 \right] \\
&= \frac{2c_0}{t} \mathbb{E} \left[\int_0^t \chi(X_{s_2}^x) ds_2 \right] - \frac{2c_0}{t^2} \mathbb{E} \left[\int_0^t \chi(X_{s_2}^x) s_2 ds_2 \right]
\end{aligned}$$

The first identity is simply motivated by symmetry. The second one is more subtle and strongly relies on the fact that X is Markovian, because by the semigroup property $\chi(X_{s_1}^x) = \mathbb{E}[\chi(X_{s_1-s_2}) | X_{s_2}^x]$ (pay attention that $X_{s_1-s_2}$ does not have the apex x , because it is not issued from x but from X_{s_2}). The third step is an application of Fubini's theorem and the following one comes from a change of variable. Finally, thanks to the simplifying assumption on the convergence in (1.9) we can uniformly estimate the inner integral by $c_0(t-s_2)$. Now, in order to go on note that, again by uniform convergence in (1.9),

$$\frac{2c_0}{t} \mathbb{E} \left[\int_0^t \chi(X_{s_2}^x) ds_2 \right] \approx 2c_0^2$$

For the second summand an integration by parts is required. More precisely, by the fact that

$$\int_0^t \chi(X_{s_2}^x) s_2 ds_2 = s_2 \int_0^{s_2} \chi(X_r^x) dr \Big|_0^t - \int_0^t \int_0^{s_2} \chi(X_r^x) dr ds_2$$

it follows that

$$\begin{aligned}
\frac{2c_0}{t^2} \mathbb{E} \left[\int_0^t \chi(X_{s_2}^x) s_2 ds_2 \right] &= \frac{2c_0}{t} \int_0^t \mathbb{E}[\chi(X_r^x)] dr - \frac{2c_0}{t^2} \int_0^t \int_0^{s_2} \mathbb{E}[\chi(X_r^x)] dr ds_2 \\
&\approx 2c_0^2 - \frac{2c_0^2}{t^2} \int_0^t s_2 ds_2 = c_0^2
\end{aligned}$$

Therefore, recalling the definition of c_0 , we have shown that

$$\frac{1}{t^2} \mathbb{E} \left[\int_0^t \int_0^t \chi(X_{s_1}^x) \chi(X_{s_2}^x) ds_1 ds_2 \right] \approx \left(\int \chi d\mathbf{m} \right)^2$$

and this concludes the proof. \square

Notice that (1.9) has been widely used in the proof above, but the most crucial aspect is the Markov property, even if we exploited it only once. In fact, it is this property that enabled us to “cut” time and in this property the decorrelation in time we previously talked about is hidden: with $\chi(X_{s_1}^x) = \mathbb{E}[\chi(X_{s_1-s_2}) | X_{s_2}^x]$ we are saying that $\chi(X_{s_1}^x)$, the evolution of the system at time s_1 issued from x , can be recovered by making the system evolve until $t = s_2$ and then restart it until $t = s_1 - s_2$, averaging on all the possible restarting point $X_{s_2}^x$. From a probabilistic point of view, this is not surprising at all, because, roughly speaking, in the Markovian case the mean value of a product is the product of the respective mean values.

Hence, one more time, when we have a Markov process (or equivalently a semigroup), (1.6) is for free once we have an apparently weaker condition, namely (1.8), which involves the solution to the

PDE governing our system. For this reason, from now on we can essentially forget about (1.6) and try to formulate every interesting result and definition in terms of semigroups. The first notion we are interested in is the one of *ergodicity*, which can be formulated in two different ways, but before we provide the definitions, let us describe the framework we are going to work within.

Remark 1.5. For the moment, for sake of simplicity we will consider the time-homogeneous case, but, as we will see later, the techniques we will adopt in the homogeneous case will give us a huge amount of results that fit in with the non-homogeneous case. The second simplifying assumption we ask is the following: the domain D is compact; this may be for instance the case if:

- D is periodic;
- we ask some reflection conditions on the boundary, as for instance Neumann conditions; on the contrary, it does not make sense to ask Dirichlet conditions, because if we ask the Brownian motion to be zero at the boundary, then it converges to zero exponentially fast in time also in the interior of D .

As a quick digression, observe that from a probabilistic point of view a Neumann condition corresponds to a reflected process, whereas a null Dirichlet condition is equivalent to a killed process. Both hypotheses will be gradually removed, but as a first step let us work within this framework. \diamond

Let us consider the following equation

$$\begin{cases} \frac{\partial u}{\partial t} + Au = 0 \\ u|_{t=0} = u_0 \end{cases} \quad (1.10)$$

where A is a linear operator (as anticipated in the summary, the non-linear case will be approached only at the end of the course), independent of time, as we have assumed. Using the standard exponential notation, the solution $u = u(x, t)$ will be also denoted by

$$u(x, t) = (e^{-tA}u_0)(x)$$

where e^{-tA} is the C_0 -semigroup (strongly continuous one-parameter semigroup) whose infinitesimal generator is $-A$; the negative sign comes from the fact that if we write (1.10) in normal form, that is the time derivative is the only term on the left, then a negative sign appears in front of A . However, in this discussion we are a bit informal, because for A to be the infinitesimal generator of a C_0 -semigroup, some technical assumptions are required and they are part of the content of the Hille-Yosida theorem (see Section A.4). Depending on such linear operator, (1.10) may be a PDE as well as an integro-differential equation. A typical example is given by a second order differential operator of the form

$$A = -a_{ij}\partial_{ij} - b_i\partial_i \quad (1.11)$$

with associated SDE

$$\begin{cases} dX_t^i = b_i(X_t)dt + \sum_{j=1}^d a_{ij}(X_t)W_t^j \\ X_0 \sim u_0 \end{cases} \quad (1.12)$$

In this case, A is a (possibly degenerate) elliptic operator, it is the generator of a semigroup associated to a parabolic PDE and it is said to be in *divergence form*. The matrix $a = (a_{ij})_{i,j}$ has non-negative coefficients and a nice way to express the ellipticity/positivity of a is to say that the matrix factorizes in the form

$$a = \frac{1}{2}\sigma\sigma^t$$

Furthermore, the coefficients a_{ij}, b_i are regular, so that even in the degenerate case the regularity of the coefficients ensures existence and uniqueness of the solution, provided u_0 is continuous. Hence, let us assume that $u_0 \in C(D)$, the system is well-posed and admits a probabilistic interpretation in terms of a suitable Markov process, in the sense that the solution to (1.10) can be expressed as in (1.5). The classical theory of SDEs tells us that such Markov process is not only unique in law, but also pathwise unique (trajectory by trajectory).

Remark 1.6. The assumption on the probabilistic representation of the solution to (1.10) needs a necessary condition to hold and it is the following: constant functions must belong to the kernel of A , because if u_0 is constant, then so is u by (1.5) and a constant c is a solution to (1.10) if and only if $Ac = 0$. On the other hand, the probabilistic representation (1.5) entails the conservation of positivity and the maximum principle, in the sense that

$$u_0 \geq 0 \quad \Rightarrow \quad u \geq 0$$

and

$$\|u(\cdot, t)\|_\infty \leq \|u_0\|_\infty, \quad \forall t > 0$$

because

$$\sup_x u(x, t) \leq \sup_x \mathbb{E}[\|u_0\|_\infty \mid X_0 = x] = \|u_0\|_\infty$$

This is for instance the case if A is given by (1.11), but we can actually say something more, since by classical results in potential theory if A enjoys the maximum principle, then it can be written as a sum of a parabolic part of type (1.11) and an integro-differential operator satisfying good properties of positivity. \diamond

Now we are finally ready to give the two notions of ergodicity we will consider in this course.

Definition 1.7 (Cesaro mean ergodicity). *We say that the system described by (1.10) is ergodic in the sense of Cesaro if A admits a unique invariant measure \mathbf{m} , that is a probability measure such that*

$$\int e^{-tA} u_0 \, d\mathbf{m} = \int u_0 \, d\mathbf{m}$$

for every $t > 0$ and the solution $u(x, t)$ to (1.10) satisfies

$$\lim_{t \rightarrow +\infty} \frac{1}{t} \int_0^t u(x, s) \, ds = \int u_0 \, d\mathbf{m}$$

uniformly in x .

Definition 1.8 (Ergodicity). *We say that the system described by (1.10) is ergodic if A admits a unique invariant measure \mathbf{m} , that is a probability measure such that*

$$\int e^{-tA} u_0 \, d\mathbf{m} = \int u_0 \, d\mathbf{m}$$

for every $t > 0$ and the solution $u(x, t)$ to (1.10) satisfies

$$\lim_{t \rightarrow +\infty} u(x, t) = \int u_0 \, d\mathbf{m}$$

uniformly in x .

It is worth saying that the definition of invariant measure we gave is fairly general, because no regularity for A is needed. However, if A is regular enough, the condition below

$$\int e^{-tA} u_0 \, d\mathbf{m} = \int u_0 \, d\mathbf{m} \quad (1.13)$$

can be equivalently restated in a different way, because by differentiation we get

$$\int Au_0 \, d\mathbf{m} = 0$$

which can be rewritten as $\langle Au_0, \mathbf{m} \rangle = 0$; by definition of adjoint and by the arbitrariness of u_0 , this becomes

$$A^* \mathbf{m} = 0 \quad (1.14)$$

Conversely, if (1.14) is assumed, then let us notice that

$$\frac{d}{dt} \int e^{-tA} u_0 \, d\mathbf{m} = - \int Au_0 \, d\mathbf{m} = - \int u_0 \, dA^* \mathbf{m} = 0$$

whence we deduce that

$$t \mapsto \int e^{-tA} u_0 \, d\mathbf{m}$$

is constant and (1.13) follows. From a probabilistic point of view, this means that if \mathbf{m} is the initial law of the Markov process $(X_t)_{t \geq 0}$ associated to (1.10), then at any time t the law of X_t is still \mathbf{m} . After this digression, observe that both definitions above are quite technical and demanding. Hence it is natural to question if, in the same way (1.6) can be replaced by (1.8), they can be made much more lighter. The answer is affirmative and the results are the following (apparently weaker) definitions.

Definition 1.9 (Cesaro mean ergodicity). *We say that the system described by (1.10) is ergodic in the sense of Cesaro if there exists a constant c such that the solution $u(x, t)$ to (1.10) satisfies*

$$\lim_{t \rightarrow +\infty} \frac{1}{t} \int_0^t u(x, s) \, ds = c \quad (1.15)$$

uniformly in x .

Definition 1.10 (Ergodicity). *We say that the system described by (1.10) is ergodic if there exists a constant c such that the solution $u(x, t)$ to (1.10) satisfies*

$$\lim_{t \rightarrow +\infty} u(x, t) = c \quad (1.16)$$

uniformly in x .

Let us prove immediately that the latter definitions entail the previous ones. As a first remark, the constant c can be determined. Indeed, by the linearity of (1.10) c linearly depends on u_0 and, by the maximum principle (see Remark 1.6), it preserves positivity: hence there exists a unique non-negative measure \mathbf{m} such that

$$c = \int u_0 \, d\mathbf{m} \quad (1.17)$$

Secondly, since the domain D is compact, \mathbf{m} has finite mass; actually, it is a probability measure, still by (1.5), because if we take $u_0 \equiv 1$, then $u \equiv 1$ and

$$\mathbf{m}(D) = \int u_0 \, d\mathbf{m} = \lim_{t \rightarrow +\infty} \frac{1}{t} \int_0^t u(x, s) \, ds = 1$$

Thirdly, \mathbf{m} is an invariant measure for A . The proof in the case of (strong) ergodicity is direct, because if (1.16) holds, then by the semigroup property $e^{-(T+t)A}u_0 = e^{-TA}e^{-tA}u_0$ and by passing to the limit as $T \rightarrow +\infty$ we obtain

$$e^{-(T+t)A}u_0 \rightarrow \int u_0 d\mathbf{m}, \quad e^{-TA}(e^{-tA}u_0) \rightarrow \int e^{-tA}u_0 d\mathbf{m}$$

whence

$$\int e^{-tA}u_0 d\mathbf{m} = \int u_0 d\mathbf{m}$$

In the case of Cesaro mean ergodicity, the proof is more delicate, since for any fixed $t > 0$ we have

$$\begin{aligned} \frac{1}{T} \int_0^T e^{-tA}u(s)ds &= \frac{1}{T} \int_0^T e^{-tA}(e^{-sA}u_0)ds = \frac{1}{T} \int_0^T e^{-(t+s)A}u_0ds \\ &= \frac{1}{T} \int_t^{T+t} e^{-sA}u_0ds = \frac{1}{T} \int_0^T e^{-sA}u_0ds \\ &\quad + \left(\frac{1}{T} \int_T^{T+t} e^{-sA}u_0ds - \frac{1}{T} \int_0^t e^{-sA}u_0ds \right) \\ &= \frac{1}{T} \int_0^T e^{-sA}u_0ds + O(1/T) \end{aligned} \tag{1.18}$$

and the last identity come from the fact that

$$\frac{1}{T} \left\| \int_T^{T+t} e^{-sA}u_0ds - \int_0^t e^{-sA}u_0ds \right\|_{L^\infty} \leq \frac{2t}{T} \sup_{s \geq 0} \|e^{-sA}u_0\|_{L^\infty} \leq \frac{2t}{T} \|u_0\|_{L^\infty}$$

where the second inequality is an immediate consequence of the probabilistic representation (1.5) or, equivalently, of the fact that the semigroup associated to A preserves positivity and constant functions belong to its kernel, in other words because of the maximum principle. By passing to the limit as $T \rightarrow +\infty$ in (1.18) and by the arbitrariness of $t > 0$, the conclusion follows.

Finally, A admits a unique invariant measure, since if $\hat{\mathbf{m}}$ is a second (non-negative probability) invariant measure, then by definition

$$\int u_0 d\hat{\mathbf{m}} = \int u(t)d\hat{\mathbf{m}} \tag{1.19}$$

for any $t > 0$. In the case (1.16) holds, by the identity above and by the uniform convergence in (1.16) we entail that

$$\int u_0 d\hat{\mathbf{m}} = \int u(t)d\hat{\mathbf{m}} \rightarrow \int c d\hat{\mathbf{m}} = c$$

independently of the regularity of $\hat{\mathbf{m}}$ (i.e. this is true even if $\hat{\mathbf{m}}$ is a Dirac mass). Thus, by (1.17) and by the arbitrariness of u_0 , $\mathbf{m} = \hat{\mathbf{m}}$ follows. In the case (1.15) holds, let us first observe that (1.19) implies

$$\int u_0 d\hat{\mathbf{m}} = \frac{1}{t} \int_0^t \left(\int u(s)d\hat{\mathbf{m}} \right) ds$$

for any $t > 0$. By this identity and by the uniform convergence in (1.15), the fact that $\hat{\mathbf{m}} = \mathbf{m}$ follows also in this case. Hence, thanks to this digression we can really look at Definitions 1.9 and 1.10 when we want to know whether a given system is ergodic or not; on the contrary, Definitions 1.7 and 1.8 tell us a lot of information about the properties of ergodic systems.

Again, but in other words, in the study of ergodicity what is really important is the convergence of the semigroup (or of the Cesaro means of the semigroup) to a constant, independent of the starting

point x (but clearly dependent on the initial datum u_0). In some cases, as for instance Example 1.3, we can study by hands the behaviour of the system, but this is not really interesting. What really cares is the development of mathematical methods that allow us to treat generally the different situations we may find.

Roughly speaking and up to technical assumptions, the main idea is to consider a certain class of test problems and to know if the solutions to these problems are constant or not, so that, depending on the answer, we will be able to say if the system is ergodic or not. In particular, two cases will be considered.

Problem 1.11 (Test for Cesaro mean ergodicity). In the setting of Remark 1.5, is it true that

$$Au = 0 \quad \Rightarrow \quad u = \text{constant}$$

provided u has a certain degree of regularity? \diamond

Problem 1.12 (Test for ergodicity). In the setting of Remark 1.5, if we consider equation (1.10) on the whole \mathbb{R} instead of the only positive axis and $u = u(x, s)$ is a solution, is it true that u is constant, provided u is bounded and uniformly continuous in both variables? \diamond

We have just stated these problems within the framework of Remark 1.5, because the definitions of ergodicity we gave strongly rely on the fact that A is time-homogeneous; ergodicity in the non-homogeneous case, as anticipated in the summary, will be discussed later.

Observe that the second test can be seen as a kind of Liouville's problem, because we have both directions of time and we want to show that the only solutions of a given evolution problem are constant. A first natural question arises: are these test problems coherent with the examples we gave? Let us only consider Examples 1.1 and 1.2, since the third one is a mixture of the previous ones. For the ergodicity in the sense of Cesaro, observe that:

- if A is the Laplacian on a periodic domain, then $\Delta u = 0$ entails $u = \text{constant}$, provided u is continuous: it is a particular case of Liouville's theorem and it can be obtained by decomposition of u in Fourier series;
- if $A = \partial_x$, then $Au = 0$ trivially implies that u is constant in x .

On the other hand, for the ergodicity in the sense of Definition 1.10 notice that:

- if A is the Laplacian, then by considering any irreversible property (that is, a property affected by the time direction, as for instance energy dissipation) and by the fact that

$$\frac{\partial u}{\partial t}(x, t) + \Delta u(x, t) = 0$$

holds on the whole \mathbb{R} , i.e. in both time directions, we can infer that u has to be constant;

- on the contrary, if $A = \partial_x$ and $u(x, s)$ is a solution to the second test problem, then u needs not be constant; indeed, for any initial datum u_0 , $u_0(x - s)$ defines a (non-constant) solution to the transport equation not only for positive times, but for any $s \in \mathbb{R}$.

Hence, the test problems are coherent with the initial examples, since we have shown that Example 1.1 describes an ergodic system, whereas the one of Example 1.2 is ergodic in the sense of Cesaro but not ergodic *tout court*. The next natural step is to single out some necessary and sufficient conditions in order to answer the test problems, but before we approach this topic, let us explain the link between the two notions of ergodicity we gave and the test problems themselves.

For ergodicity, we can see that up to technical difficulties Problem 1.12 is equivalent to Definition 1.10, namely if $u = u(x, t)$ is a solution to (1.10) defined for $t \geq 0$ and $\bar{u} = \bar{u}(x, s)$ is a solution to

(1.10) defined for every $s \in \mathbb{R}$, then (1.16) holds (and so the system is ergodic) if and only if \bar{u} is constant, that is Problem 1.12 has positive answer. Indeed, if we consider

$$u_T(x, s) := u(x, s + T) \quad (1.20)$$

where $T > 0$ is thought as a large time in the future, then u_T is still globally defined in the future and it is still a solution to (1.10) up to $s = -T$ in the past, thanks to the fact that we are under a time-homogeneity assumption on A . From an intuitive point of view, we expect that, under some compactness assumption and up to extract a suitable subsequence, by ω -limit set arguments we have $u_T \rightarrow \bar{u}$ as $T \rightarrow +\infty$, so that the equivalence is established.

On the contrary, the link between Cesaro mean ergodicity and Problem 1.11 is more subtle, since we start from an evolution equation and we get the stationary problem $Au = 0$. The connection relies on the following theorem, which belongs to mathematical folklore, but before the statement let us slightly modify the notation of the equation (1.10): the initial datum u_0 will be denoted by f .

Theorem 1.13. *The following are equivalent:*

(i) *there exists a constant c such that*

$$\frac{1}{t} \int_0^t u(x, s) ds \rightarrow c$$

uniformly in x ;

(ii) *$\varepsilon u_\varepsilon$ uniformly converges to a constant as $\varepsilon \downarrow 0$, where u_ε is the solution to the problem*

$$Au_\varepsilon + \varepsilon u_\varepsilon = f \quad (1.21)$$

for $\varepsilon > 0$.

A rough look at the statement tells us that the parameter ε plays the same role of $1/t$ in the time Cesaro mean and the operators

$$\frac{1}{t} \int_0^t e^{-sA} ds, \quad (\varepsilon \text{Id} + A)^{-1}$$

behave more or less in the same way. In addition, for sake of information, let us say that this result belongs to the class of the so-called *Tauberian theorems* (see Section A.2 for more details). Together with Abelian ones, these theorems give conditions for for two methods of summing divergent series to give the same result, named after Niels Henrik Abel and Alfred Tauber. The original examples are Abel's theorem showing that if a series converges to some limit then its Abel sum is the same limit, and Tauber's theorem showing that if the Abel sum of a series exists and the coefficients are sufficiently small (say $o(1/n)$) then the series converges to the Abel sum. More general abelian and tauberian theorems give similar results for more general summation methods, even in integral form (as it is the case for Theorem 1.13).

Proof. Let us prove the two implications.

(i) \Rightarrow (ii) By using the fact that

$$(\varepsilon \text{Id} + A)^{-1} = \int_0^{+\infty} e^{-\varepsilon s} e^{-sA} ds$$

(and this identity follows by the semigroup property) and by integration by parts, we can observe that

$$\varepsilon u_\varepsilon = \varepsilon \int_0^{+\infty} e^{-\varepsilon s} u(s) ds = \varepsilon^2 \int_0^{+\infty} e^{-\varepsilon s} \left(\int_0^s u(r) dr \right) ds$$

$$\approx \varepsilon^2 c \int_0^{+\infty} s e^{-\varepsilon s} ds = \varepsilon c \int_0^{+\infty} e^{-\varepsilon s} ds = c$$

where the approximation is motivated by the assumption on time Cesaro means.

(ii) \Rightarrow (i) Since u_ε does not depend on time, (1.21) can be restated as

$$\frac{\partial u_\varepsilon}{\partial t} + A u_\varepsilon = f - \varepsilon u_\varepsilon \quad (1.22)$$

whence we deduce that

$$u_\varepsilon = \int_0^t e^{-sA} (f - \varepsilon u_\varepsilon) ds + e^{-tA} u_\varepsilon \quad (1.23)$$

Indeed, $e^{-tA} u_\varepsilon$ is, by definition of semigroup, the solution to (1.10) with initial datum u_ε and evaluated at time t . Since on the one hand u_ε is a solution to (1.21) and on the other hand (1.22) is a non-homogeneous version of (1.10), by the constants variation method we infer that u_ε is precisely given by (1.23). If we divide this expression by t and we rearrange the terms, then we have

$$\frac{1}{t} \int_0^t e^{-sA} f ds = \frac{\varepsilon}{t} \int_0^t e^{-sA} u_\varepsilon ds + \frac{1}{t} u_\varepsilon + \frac{1}{t} e^{-tA} u_\varepsilon$$

and the first term at the right-hand side can be easily estimated, since by assumption $\varepsilon u_\varepsilon$ uniformly converges to a constant c as $\varepsilon \downarrow 0$ and the semigroup e^{-sA} preserves the bounds. Therefore

$$\frac{1}{t} \int_0^t e^{-sA} f ds = c + o(\varepsilon) + \frac{1}{t} (u_\varepsilon + e^{-tA} u_\varepsilon)$$

As a further step, u_ε can be controlled by a constant depending on ε , so that

$$\left| \frac{1}{t} \int_0^t e^{-sA} f ds - c \right| \leq o(\varepsilon) + \frac{C(\varepsilon)}{t}$$

and this shows that

$$\left| \lim_{t \rightarrow +\infty} \frac{1}{t} \int_0^t e^{-sA} f ds - c \right| \leq o(\varepsilon)$$

By the arbitrariness of ε , the desired convergence for the time Cesaro mean follows. \square

Since this result belongs to the mathematical folklore, its statement can be found in many different versions: for instance, the uniform convergence to a constant can be replaced both in (i) and (ii) by uniform convergence to any function belonging to the kernel of A . Finally, the statement would be false if we considered u_ε instead of $\varepsilon u_\varepsilon$, because u_ε may explode, as shown in the example below.

Example 1.14. Let us consider the PDE

$$-\Delta u_\varepsilon + \varepsilon u_\varepsilon = f$$

If $f \equiv 1$, then u_ε is given by $1/\varepsilon$ and this function clearly explodes as $\varepsilon \downarrow 0$. More generally, if $f \leq C$, then $u_\varepsilon \leq C/\varepsilon$. On the other hand, $\varepsilon u_\varepsilon$ is bounded. \diamond

As already observed, on the one hand the time Cesaro mean can be rewritten in terms of generator as

$$\frac{1}{t} \int_0^t e^{-sA} u_0 ds \rightarrow c \quad \text{uniformly in } x$$

while on the other hand u_ε involves the resolvent operator of A , because $u_\varepsilon = (\varepsilon \text{Id} + A)^{-1} u_0$. Thus, thanks to the theorem above, (1.15) can be replaced by the uniform convergence (to a constant) of

the solutions to (1.21) and this explains why in Problem 1.11 we have to look at $Au = 0$: in some sense this is the limit case of (1.21). In more details, we have to notice that if u_ε is the solution to (1.21), then $\varepsilon u_\varepsilon$ solves the following equation

$$A(\varepsilon u_\varepsilon) + \varepsilon(\varepsilon u_\varepsilon) = \varepsilon f \tag{1.24}$$

Moreover, it is not difficult to prove that $\varepsilon u_\varepsilon$ is uniformly bounded, provided f is bounded (the *a priori* estimate that proves this claim will be studied later, but in the case $A = -\Delta$ it has already been noticed in Example 1.14). Therefore, up to pass to a suitable subsequence, we can say that $\varepsilon u_\varepsilon$ uniformly converges to a suitable limit function \bar{u} and by passing to the limit as $\varepsilon \downarrow 0$ in (1.24) we see that

$$A\bar{u} = 0$$

Thus, also the equivalence between Definition 1.9 and Problem 1.11 is proved and strongly relies on Theorem 1.13.

However, in the whole discussion above we have been rather superficial on the existence of solutions to the resolvent equation. The maximum principle grants it, but the problem is very delicate, even in the non-degenerate case if A is rather general; in the case A is of (1.11)-type but degenerate, the problem is still difficult, although existence of solutions to the resolvent equation is still ensured; the situation becomes much easier if A is uniformly elliptic, since in this case the resolvent set is in $[0, +\infty[$ (indeed, constants are the first eigenfunctions in the sense that $\lambda = 0$ is the first eigenvalue of A). It is worth saying that, at least formally, this problem is strictly related to well-known facts, such as the Krein-Rutman theorem and the Perron-Frobenius one; these results may help us to understand why Problem 1.11 is formulated as it is. The interested reader is addressed to Section A.3 for more details, but for the moment let us stress the strong analogy between our problem and the Perron-Frobenius theorem. Also in the latter there is a sign-preserving operator (more precisely a matrix) with simple and positive first eigenvalue and the thesis of the theorem is the existence of the invariant measure, under suitable further assumptions on the matrix. However, we will not spend more time on this abstract result, since as just said it holds for matrices and in this course we are not interested in Markov chains. Let us only point out an important application of it.

Remark 1.15. Just open your favorite search engine, like Google, AltaVista, Yahoo, type in the key words and the search engine will display the pages relevant for your search. But how does a search engine really work? Modern search engines employ methods of ranking the results to provide the best results first that are more elaborate than just plain text ranking (as it was in the early 90s). One of the most known and influential algorithms for computing the relevance of web pages is the Page Rank algorithm used by the Google search engine. It was invented by Larry Page and Sergey Brin while they were graduate students at Stanford and it became a Google trademark in 1998. The idea that Page Rank brought up was that, the importance of any web page can be judged by looking at the pages that link to it. If we create a web page i and include a hyperlink to the web page j , this means that we consider j important and relevant for our topic. If there are a lot of pages that link to j , this means that the common belief is that page j is important. If on the other hand, j has only one backlink, but that comes from an authoritative site k , we say that k transfers its authority to j ; in other words, k asserts that j is important. Whether we talk about popularity or authority, we can iteratively assign a rank to each web page, based on the ranks of the pages that point to it. To this aim, we picture the Web net as a directed graph, with nodes represented by web pages and edges represented by the links between them (see for instance Figure 2). This situation can be modeled very well in terms of (very big) matrices, satisfying the assumptions of the Perron-Frobenius theorem; hence, the invariant measure exists and can be computed. It is precisely this measure that gives the desired Page Rank, displayed by your favorite search engine.¹ \diamond

¹Adapted from [6].

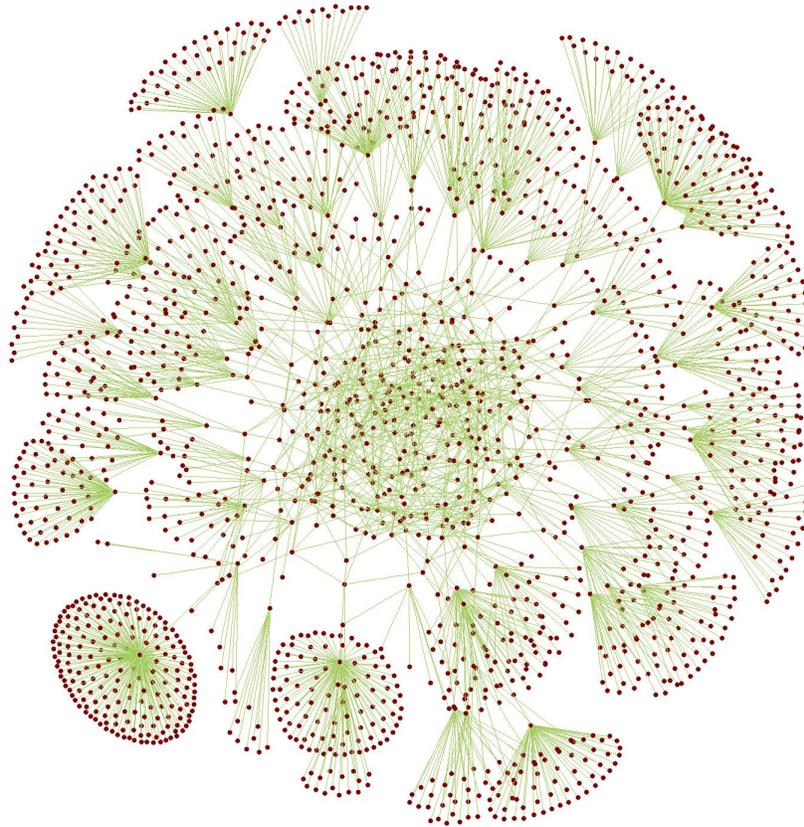


Figure 2: The Internet nodes are red dots and linkages are green lines.

On the contrary, the Krein-Rutman theorem holds also for operators but the flip side is the requirement of compactness, in the sense that we ask the semigroup to be compact. Unfortunately, this hypothesis is false in many situations: it is sufficient to consider the transport equation

$$\frac{\partial u}{\partial t} - \frac{\partial u}{\partial x} = 0$$

we have already discussed in Example 1.2, because in this case the norm of the solution is constant in time. Hence we can not expect any help from this theorem in general situations.

During this first lecture we have thus underlined many different motivations for the content of the course: the fact that asymptotic behaviour is a classical problem in the theory of Markov chains, diffusion processes and parabolic PDEs and the link between ergodic theory and MFG (just cited in the summary), for instance, but there is also a strong connection with stochastic control. Further links may be found with other sciences, e.g. biology and economics, and it is precisely on the connection with the latter that we would like to spend some more words.

Remark 1.16. The French economist Thomas Piketty has recently published the great best-seller [5], focused on wealth and income inequality in Europe and the United States since the 18th century. The book's central thesis is that when the rate of return on capital r is greater than the rate of economic growth g over the long term, the result is an increasing concentration of wealth (the book actually provides macroeconomic evidences of this claim) and this unequal distribution of wealth causes social and economic instability; thus, the quantity $r - g$ is an amplification factor of social inequalities. Hence inequality is not an accident, but rather a feature of capitalism and can only be

reversed through state interventionism. The book thus argues that, unless capitalism is reformed, the very democratic order will be threatened. Piketty proposes a global system of progressive wealth taxes to help reduce inequality and avoid the vast majority of wealth coming under the control of a tiny minority. From a rough mathematical point of view, by using generally accepted discrete-time

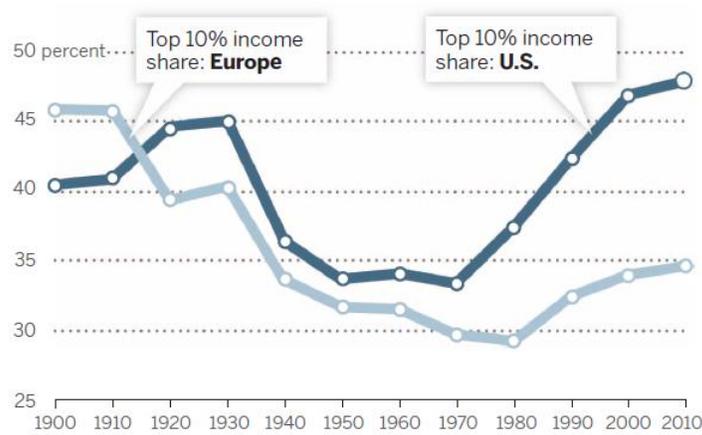


Figure 3: Share of top income decile in total pretax income.

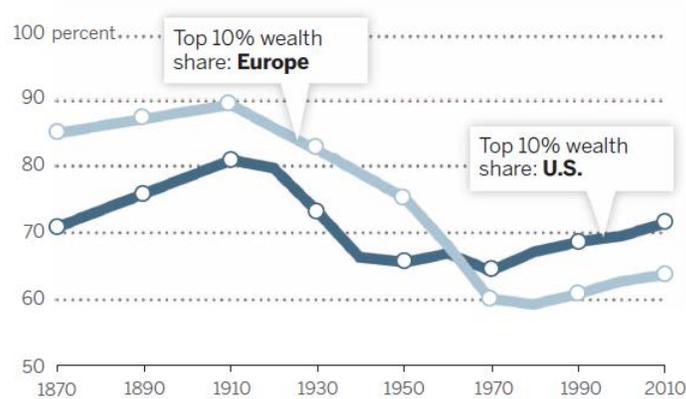


Figure 4: Share of top wealth decile in total net wealth.

models (i.e. models used by governmental institutions, such as the Fed) Piketty proves that there exists an invariant measure and the law of the income or wealth distribution converges towards it. By computing explicitly the invariant measure, it is seen that in the case $r - g$ is large, social inequalities become bigger and bigger.

Xavier Gabaix, Jean-Michel Lasry, Pierre Louis Lions and Benjamin Moll ask whether such theory can also explain the dynamics of income and wealth inequality observed in Europe and in the United States, namely the convergence rate of the law of the income and wealth distribution towards the invariant measure. This convergence is known to be of exponential type, but this information is useless unless we know the characteristic time, that is the inverse of the constant appearing in the exponential, representing the time needed for the invariant measure to be “essentially” reached.

Gabaix, Lasry, Lions and Moll gave in [3] the following analytic formula for the speed of convergence

$$\text{speed} = \frac{1}{8} \frac{\sigma^2}{\eta^2}$$

where η denotes the top inequality and σ the randomness of growth rates. As a consequence, gradual rise in top wealth inequality can potentially be explained by theories based on random growth processes (see Figure 4), but on the contrary this is not possible for the rapid rise in top income inequality, observed in the United States in the past 40 years and shown in Figure 3 (this result is independent of Piketty's theory). Moreover, they proved that the characteristic time is about 30 years, which is a very long period, during which everything changes from a socio-economic point of view. Therefore, although there are stationary solutions, we can not exploit them and not even invariant measures. This does not mean that Piketty's thesis is false (the empirical observations he made are not at all rejected), but just that we are always in a transient situation, so that long term forecasts are rather uncertain. \diamond

Thanks to the machinery we have just developed, from the next lecture we will begin to attack the problems we formulated.

2 Lecture 2 - 14 November

Abstract

Aim of the second lecture is to understand

As we have explained in the previous lecture, main aim of the course is the understanding of the ergodicity of Markov processes or, in an equivalent way, the asymptotic behaviour of the solutions to (1.10), where the linear operator A enjoys the properties pointed out in Remark 1.6, which are consequences of the probabilistic representation (1.5): $Ac = 0$ for any constant c and the maximum principle. This problem has been fixed with Definitions 1.9 and 1.10 and later restated in Problems 1.11 and 1.12, but the situation is far from being satisfactory, because we are not able to answer easily the test problems yet. For this reason, we would like to single out some necessary and sufficient conditions for Problems 1.11 and 1.12 to hold.

In order to solve this problem, let us first assume that the coefficients of A are regular and the domain Q we are working within is compact and periodic (later we will gradually remove this hypothesis). If A is of the form (1.11), this amounts to say that a_{ij} and b_i are regular, but if A is general this requirement is not much harder to formulate, since we know by Remark 1.6 that if A enjoys the maximum principle, then A splits into a parabolic part of (1.11)-type and an integro-differential one. By classical and well-known facts, this assumption ensures the semigroup to be regularizing, even in the degenerate case, in the sense that if u_0 is regular, then so is u in both variables. On the contrary, it is worth stressing that the regularity of u_ε is more delicate. Recall that u_ε is the solution to the resolvent equation (1.21), that is

$$u_\varepsilon = (\varepsilon \text{Id} + A)^{-1} f \tag{2.1}$$

and its importance is due to Theorem 1.13, because it enables us to replace the time Cesaro mean by the resolvent operator. But in general u_ε is not smooth, even if so is f (this is for instance the case if A is degenerate): in fact, its regularity decreases as $\varepsilon \downarrow 0$. Thus, also the meaning of (1.21) is delicate, since there might be a lack of regularity. However, even in the worst case all u_ε are Hölder continuous, so that (1.21) has to be meant in the sense of distributions or equivalently, at least for the cases we are interested in, in the sense of viscosity solutions. The importance of the resolvent equation (1.21) is not only linked to the notion of Cesaro mean ergodicity, but also to the fact that if we are able to solve such equation, then we can prove the existence of at least one

invariant measure for A under a compactness assumption on the domain Q , as shown in the remark below. Pay attention to the fact that, however, such a measure need not be unique.

For sake of information, let us say that the existence of invariant measures for sufficiently “nice” maps defined on “nice” domains is a classical and fundamental result in the theory of dynamical systems, known as Krylov-Bogolyubov theorem, named after the Russian-Ukrainian mathematicians and theoretical physicists Nikolay Krylov and Nikolay Bogolyubov. The interested reader can find the proof in [1].

Remark 2.1. Let us assume we can solve the resolvent equation (1.21) for any f ; then we are also able to solve the dual problem for the uniform measure $(1/|Q|)\mathcal{L}^d$, where \mathcal{L}^d denotes the d -dimensional Lebesgue measure, $|Q|$ denotes the mass of the domain w.r.t. it and we assume $Q \subset \mathbb{R}^d$. This means that there exists a measure \mathbf{m}_ε such that

$$A^* \mathbf{m}_\varepsilon + \varepsilon \mathbf{m}_\varepsilon = \varepsilon \frac{1}{|Q|} \mathcal{L}^d \quad (2.2)$$

Taking again into account Remark 1.6, by the maximum principle $\mathbf{m}_\varepsilon \geq 0$ and by the very definition of adjoint

$$\int A^* \mathbf{m}_\varepsilon = \int A(1) d\mathbf{m}_\varepsilon = 0$$

Therefore, by integrating (2.2) on Q we get that \mathbf{m}_ε is a probability measure. Now the compactness of Q is fundamental, because it entails the compactness w.r.t. weak convergence of the family $\mathcal{P}(Q)$ of probability measures on Q and, in particular, of $\{\mathbf{m}_\varepsilon\}$ too. Hence, there exist a suitable subsequence $(\varepsilon_n)_{n \in \mathbb{N}}$ with $\varepsilon_n \rightarrow 0$ and a probability measure \mathbf{m} such that $\mathbf{m}_{\varepsilon_n} \rightharpoonup \mathbf{m}$, that is

$$\int \varphi d\mathbf{m}_{\varepsilon_n} \rightarrow \int \varphi d\mathbf{m}, \quad \forall \varphi \in C_b(Q)$$

as $n \rightarrow \infty$. This immediately implies that \mathbf{m} is an invariant measure, because by the fact that all $\mathbf{m}_{\varepsilon_n}$ is invariant and by choosing $u_0 \in C_b(Q)$ we get

$$\int u_0 d\mathbf{m} = \lim_{n \rightarrow \infty} \int u_0 d\mathbf{m}_{\varepsilon_n} = \lim_{n \rightarrow \infty} \int e^{-tA} u_0 d\mathbf{m}_{\varepsilon_n} = \int e^{-tA} u_0 d\mathbf{m}$$

It is worth stressing that \mathbf{m} is still a probability measure, that is we do not lose mass, because we are in a compact setting. On the contrary, if the domain we are working within is not, then there might be a loss of mass towards the boundary: for instance in \mathbb{R}^d it is very easy to build a sequence of probability measures convergent to 0. Thus, when we will move to the non-compact case, it will be fundamental to single out some good conditions ensuring us that the mass does not accumulate at the boundary. \diamond

If the system is ergodic either in the sense of Definition 1.9 or 1.10, we have already stressed that the invariant measure is unique.

After this remark, let us now approach the problem of ergodicity and its characterization we mentioned in the previous lecture and let us try to be more precise. Indeed, technical difficulties were avoided in the statement of Problem 1.11 and 1.12 and the notation was also a little bit ambiguous, because the reader might have thought u to be a solution to (1.10), while it was not. But aim of the previous lecture was to give a rough idea of the link between Problem 1.11 and Definition 1.9 (resp. Problem 1.12 and Definition 1.10). For this reason, we now restate the test problems in a proper way, keeping in mind that Q is always assumed to be compact and A is time-homogeneous.

Problem 2.2 (Test for Cesaro mean ergodicity). In the setting of Remark 1.5, is it true that

$$A\bar{u} = 0 \quad \Rightarrow \quad \bar{u} = \text{constant}$$

provided \bar{u} is continuous? \diamond

Problem 2.3 (Test for ergodicity). In the setting of Remark 1.5, if we consider equation (1.10) on the whole \mathbb{R} instead of the only positive axis and $\bar{u} = \bar{u}(x, s)$ is a uniformly continuous and bounded solution (thus, it is defined for all $s \in \mathbb{R}$) such that

$$\max_{x \in Q} \bar{u}(x, s), \quad \min_{x \in Q} \bar{u}(x, s) \quad (2.3)$$

are both independent of s , is it true that \bar{u} is constant? \diamond

A quick remark on the assumptions we are making. The compactness of the domain is the easiest one to lighten; on the contrary, the time-homogeneity is much more delicate and in the future we will work with gradually general operators (periodic, quasi-periodic, stationary and finally arbitrary). Furthermore, pay attention to the fact that in the formulation of Problem 2.2 we simply add a precise regularity assumption on \bar{u} w.r.t. Problem 1.11 and we do the same for Problem 2.3 w.r.t. Problem 1.12, but in the latter case there is also a further hypothesis, which is not completely intuitive, at least for the requirement on the minimum (we will discuss it later). On the contrary, the assumption on the maximum can be easily explained by analogy and the reader can simply think to the heat equation: in that case, the maximum strictly decreases as $s \rightarrow +\infty$, unless \bar{u} is constant. As we will see later, the independence of the maximum and of the minimum allows us to obtain a necessary and sufficient condition for ergodicity.

Before going further in the discussion, let us anticipate or repeat some basic observations on the equations we are interested in, namely equation (1.10), which describes our system, and the resolvent equation (1.21). Let us begin with the first one.

- (i) For the problem (1.10) there is continuous dependence of u on the initial datum u_0 in the sense that

$$\|u(\cdot, t)\|_{L^\infty} \leq \|u_0\|_{L^\infty} \quad (2.4)$$

for any $t \geq 0$ and this is a consequence of the probabilistic representation (1.5) or, equivalently, of the fact that the constant functions belong to the kernel of $-A$ and the semigroup associated to it preserves positivity (both properties are in turn consequences of the maximum principle). Thus, $-A$ is the generator of a contraction semigroup in L^∞ .

- (ii) As a consequence, in order to understand what happens to u as $t \rightarrow +\infty$ it is sufficient to know the behaviour of u_0 on a dense subset of Q . In other words, from the point of view of Definitions 1.9 and 1.10 assuming u_0 to be continuous is completely equivalent to assuming it to be smooth, since any continuous function can be arbitrarily well approximated and the difference between u_0 and the smooth approximation is controlled for any t by (2.4).
- (iii) If u_0 is regular (thanks to the previous remark, it is not important to specify the degree of regularity), then $A^k u_0 \in L^\infty$ for every $k \in \mathbb{N}$ and in addition $A^k u$ solves the same equation of u , that is

$$\frac{\partial}{\partial t}(A^k u) + A(A^k u) = 0 \quad (2.5)$$

This is simply due to the fact that A is linear and commutes with e^{-tA} , whence

$$\frac{\partial}{\partial t}(A^k u) = \frac{\partial}{\partial t} A^k e^{-tA} u_0 = -A e^{-tA} A^k u_0 = -A(A^k u)$$

but we are implicitly using the fact that Q is a periodic domain (otherwise, further assumptions on u_0 at the boundary are required). As a consequence of (2.5), of the fact that $A^k u_0 \in L^\infty$ and of property (i) above, we have that $\|A^k u(\cdot, t)\|_{L^\infty}$ is uniformly bounded in t for any $k \in \mathbb{N}$; however, *a priori* the bound might depend on k and increase with it, as it is for instance the case if $A = \partial_x$ and u_0 does not belong to a particular class of analytic functions. What is really important to stress is the big gain of information on orbits and the fact that if A provides any sort of regularity, then this regularity is uniformly preserved in time.

For the resolvent equation, the same properties are still true under slight modifications, say replace u and u_0 by $\varepsilon u_\varepsilon$ and f respectively.

- (a) The solution to (1.21) continuously depends on f in the sense that

$$\|\varepsilon u_\varepsilon\|_{L^\infty} \leq \|f\|_{L^\infty} \quad (2.6)$$

for any $\varepsilon > 0$ and this is again a consequence of the maximum principle, but unlike property (i) some refined manipulations on the resolvent equation are needed. For the reader who is not at ease with them, it is sufficient to observe that

$$u_\varepsilon = \int_0^{+\infty} e^{-\varepsilon t} e^{-tA} f dt$$

From a formal viewpoint, this integral representation formula can be guessed by looking at (2.1) and at

$$\frac{1}{\varepsilon + a} = \int_0^{+\infty} e^{-(\varepsilon+a)t} dt$$

but the rigorous proof relies on the Hille-Yosida theorem (see Section A.4). By means of this formula, (2.6) follows, because

$$\|\varepsilon u_\varepsilon\|_{L^\infty} \leq \varepsilon \int_0^{+\infty} e^{-\varepsilon t} \|e^{-tA} f\|_{L^\infty} dt \leq \varepsilon \int_0^{+\infty} e^{-\varepsilon t} \|f\|_{L^\infty} dt = \|f\|_{L^\infty}$$

where the last inequality is due to (2.4).

- (b) As for property (ii), in order to understand what happens to $\varepsilon u_\varepsilon$ as $\varepsilon \downarrow 0$ it is sufficient to know the behaviour of f on a dense subset of Q and so, for the same reason described at (ii), assuming f to be continuous is completely equivalent to assuming it to be smooth.
- (c) If f is regular (thanks to the previous remark, it is not important to specify the degree of regularity), then $A^k f \in L^\infty$ for every $k \in \mathbb{N}$ and in addition $A^k u_\varepsilon$ solves the following equation

$$A(A^k u_\varepsilon) + \varepsilon A^k u_\varepsilon = A^k f$$

whence we deduce that $\|\varepsilon A^k u_\varepsilon\|_{L^\infty}$ is uniformly bounded in ε for any $k \in \mathbb{N}$, even if *a priori* the bound might again depend on k .

For the moment this is all what we need to know on (1.10) and (1.21) and actually we can not expect more interesting preliminary considerations without further assumptions, e.g. auto-adjoint hypothesis. On the contrary, a further observation on (2.3) is possible.

Remark 2.4. Let u be the solution to (1.10) with initial condition u_0 ; then

$$\begin{aligned} \max_{x \in Q} u(x, t) &\searrow && \text{as } t \rightarrow +\infty \\ \min_{x \in Q} u(x, t) &\nearrow && \text{as } t \rightarrow +\infty \end{aligned}$$

but growth and degrowth need not be strict. Let us prove the first claim (for the second one, it will be sufficient to change u_0 into $-u_0$); main tool is the probabilistic representation (1.5), because by means of it we deduce that

$$\max_{x \in Q} u(x, t) = \max_{x \in Q} \mathbb{E}[u_0(X_t) | X_0 = x] \leq \mathbb{E} \left[\max_{x \in Q} u_0(X_t) \mid X_0 = x \right]$$

$$\leq \max_{z \in Q} u_0(z) \tag{2.7}$$

Hence, the maximum of u at time t can not be larger than the maximum at the initial time and since we have the semigroup property, this immediately entails that u is non-increasing in time. So let us stress that with no hypothesis (except for the maximum principle) we have an interesting stabilization property: the maximum and the minimum converge as $t \rightarrow +\infty$, although in general we can not say anything on the asymptotic behaviour of u . \diamond

From an intuitive point of view, this remark allows us to understand why in Problem 2.3 we asked the maximum and the minimum of \bar{u} to be both independent of s . Indeed, let us first recall how \bar{u} is built: by defining u_T as in (1.20) for any $T > 0$, we obtain a family of functions defined on $[-T, +\infty[$ and under suitable compactness assumptions we expect that it is possible to extract a subsequence converging to a function \bar{u} defined on $Q \times \mathbb{R}$; in more details but still roughly speaking, if we strengthen the compactness assumptions so that the convergence of u_{T_n} towards \bar{u} is uniform in space and on compact subsets of time, that is for $x \in Q$ and s bounded, then

$$\max_{x \in Q} u(x, s + T_n) = \max_{x \in Q} u_{T_n}(x, s) \rightarrow \max_{x \in Q} \bar{u}(x, s)$$

and, by virtue of the previous remark, the maximum on the left is also convergent to a quantity which clearly does not depend on time, whence the assumption in Problem 2.3.

As a further digression, let us notice that both in ergodicity and Cesaro mean ergodicity uniform convergence is involved: for $\varepsilon u_\varepsilon$ in the first case and for u_T in the latter. The importance of this assumption is due to the fact that it ensures the continuity of the limit function, but why is so important for the limit function to be continuous? We will briefly explain it via a deterministic example.

Example 2.5. Let us consider $Q = [-1, 1]$ equipped with the restriction of the Lebesgue measure and A of type (1.11) with $a \equiv 0$ and $b(x) = -x$; with these choices, (1.10) becomes

$$\begin{cases} \frac{\partial u}{\partial t}(x, t) = -x \frac{\partial u}{\partial x}(x, t) \\ u|_{t=0} = u_0 \end{cases}$$

and the SDE (1.12) associated to A is $dX_t = -X_t dt$, that is a deterministic differential equation, since there is no noise. Hence if we solve it with the initial condition $X_0 \sim 1$, then $X_t = e^{-t}$ and this allows us to check immediately that ergodicity holds, because the solution to (1.10) is given by

$$u(x, t) = u_0(e^{-t}x) \tag{2.8}$$

and it is clear that as $t \rightarrow +\infty$, $u(x, t) \rightarrow u_0(0)$ uniformly in x since u_0 is assumed to be continuous and Q is bounded, so that $\{e^{-t}x\}_{x \in Q}$ is bounded too. Thus, ergodicity in the sense of Definition 1.10 can be directly checked without passing through Problem 2.3 and in particular it entails Cesaro mean ergodicity.

However, if we want to check weak/strong ergodicity by means of Problem 2.2 and Problem 2.3 some difficulties arise. In the case we consider (1.10) on the whole \mathbb{R} , as in Problem 2.3, and we assume \bar{u} to be continuous, the solution \bar{u} is still given by (2.8). Thus, by the fact that $\max_{x \in Q} \bar{u}(x, s)$ and $\min_{x \in Q} \bar{u}(x, s)$ are both supposed independent of s , we deduce that \bar{u} is constant, because u_0 is continuous and so $\bar{u}(x, s) \rightarrow u_0(0)$ as $s \rightarrow +\infty$ for any $x \in Q$. As a consequence,

$$\max_{x \in Q} \bar{u}(x, s) = \min_{x \in Q} \bar{u}(x, s) = u_0(0)$$

and this exactly means that \bar{u} is constant, i.e. the system is ergodic. The invariant measure is easily seen to be $\mathfrak{m} = \delta_0$. However, if \bar{u} is not continuous, we can not say that $\bar{u}(x, s) = u_0(e^{-s}x)$ and we

can not prove that \bar{u} is constant.

In the even easier case of Cesaro mean ergodicity, an analogous problem is present, because in this case we have to look at the static problem $A\bar{u} = 0$, namely

$$-x\bar{u}'(x) = 0, \quad \forall x \in Q$$

and this condition implies that \bar{u} is constant on $Q \setminus \{0\}$. If \bar{u} is continuous, then we can say that \bar{u} is constant on Q *tout court* and $\bar{u} \equiv \bar{u}(0)$, that is Problem 2.2 has positive answer and the system is ergodic in the sense of Cesaro; but if we do not ask \bar{u} to be continuous, then there might be a jump at $t = 0$, as it is the case for the Heaviside step function (see Figure 5), which belongs to the kernel of A . In both cases (strong and weak ergodicity) the discontinuity of \bar{u} is nothing but

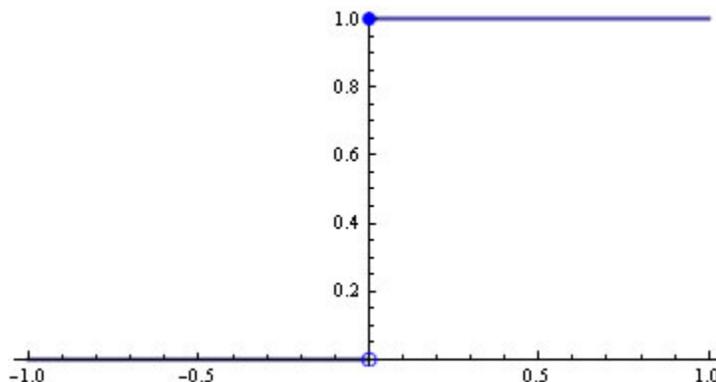


Figure 5: The Heaviside step function.

a consequence of a discontinuous initial datum u_0 , as it can be seen by (2.8). In the direct proof we did at the beginning of the example, we explicitly said that u_0 is continuous and this allowed us to prove ergodicity; on the contrary, in Problem 2.2 and 2.3 u_0 is not involved: for this reason, the continuity of u_0 must be replaced by the continuity of \bar{u} . \diamond

This example shows that the approach to ergodicity based on Problems 2.2 and 2.3 makes sense only in the continuous setting, in the sense that initial data have to be at least continuous (the problem is much more delicate for initial data which are only L^1 or L^2 regular and it can be studied only in particular cases). A different question is the following: given a continuous initial condition, is it possible to replace uniform convergence in Definition 1.9 and 1.10 (and in all equivalent formulations) by more general types of convergence? For the moment we will not explore this aspect, but pay attention to the different nature of the problem. What is now really important to stress is the role played by the continuity of u_0 and \bar{u} and by the uniform convergence and the reason why it is important to work within this setting.

However, there is a flip side: in order to work properly in the continuous framework we have to make the following strong assumptions.

Hypothesis 2.6 (for Cesaro mean ergodicity). In the case of Definition 1.9, given a regular initial condition f , let us assume that the family $(\varepsilon u_\varepsilon)_{\varepsilon \in]0,1]}$ is relatively compact in the Banach space $(C(Q), \|\cdot\|_\infty)$. \diamond

Hypothesis 2.7 (for ergodicity). In the case of Definition 1.10, given a regular initial condition u_0 , let us assume that the orbit $(u(t))_{t \geq 0}$ generated by u_0 under the semigroup e^{-tA} is relatively compact in the Banach space $(C(Q), \|\cdot\|_\infty)$. \diamond

These hypotheses are not technical, but really demanding, contain a lot of information and in practical situations it might be hard to prove them. The forthcoming necessary and sufficient conditions for ergodicity rely on them. It is worth saying that, since Q is assumed to be compact,

the assumptions above can be equivalently restated as follows: the orbit of u_0 and $(\varepsilon u_\varepsilon)_{\varepsilon \in]0,1]}$ are equicontinuous families of functions. However, how can we really check that $(u(t))_{t \geq 0}$ is equicontinuous (and $(\varepsilon u_\varepsilon)_{\varepsilon \in]0,1]}$ in an analogous way)? There are several situations where this condition is easily seen to hold, as for instance the case when the semigroup compactifies the orbits: uniformly elliptic and even hypoelliptic operators, jump processes with small jumps and, more generally, not too degenerate operators. A second case consists of preserving the uniform continuity along the whole orbit and we will refer to it as *stabilization*; this is for instance the case described in Example 2.5, because by formula (2.8) we have that

$$|u(x, t) - u(y, t)| = |u_0(e^{-t}x) - u_0(e^{-t}y)| \leq \omega_0(|e^{-t}x - e^{-t}y|) \leq \omega_0(|x - y|)$$

where ω_0 is a modulus of continuity for u_0 (indeed, u_0 is continuous on the compact set Q , hence it is uniformly continuous and without loss of generality a modulus of continuity can always be supposed increasing). This example is extremely stable and more general situations will be studied. For instance, for an elliptic operator it will be interesting to single out the conditions that the coefficients of the operator must satisfy in order to preserve the uniform continuity. There is a third possible case, neither hypoelliptic or stable, but still equicontinuity holds: it is the case of quasi-periodic homogenization; an example is given by the unit cube $\mathbb{R}^d/\mathbb{Z}^d$ where the diffusion is assumed to happen only along some prescribed parallel lines with irrational direction: this situation is more degenerate than the hypoelliptic case and it is not stable, unless the operator has constant coefficients along each line, but nevertheless Hypothesis 2.7 is satisfied.

The list of examples we have just presented is not at all exhaustive and the previous assumptions can not be characterized in terms of conditions easier to check. Hence the reader might be tempted to say that such assumptions are too strong, but this is false: they are really necessary, in the sense that if the system is ergodic (resp. ergodic in the sense of Cesaro), then Hypothesis 2.7 (resp. Hypothesis 2.6) is satisfied. In fact, if the system is ergodic, then $u(t)$ is continuous for any $t \geq 0$ and thus $(u(t))_{t \in [0, T]}$ is equicontinuous for any $T > 0$; moreover, $u(x, t) \rightarrow c$ uniformly in x as $t \rightarrow +\infty$, so that we can pass to the limit as $T \rightarrow +\infty$ and say that $(u(t))_{t \geq 0}$ is equicontinuous *tout court*. However, the previous assumptions are not sufficient and so are not equivalent to weak/strong ergodicity.

From now on, Hypothesis 2.6 and 2.7 will be always assumed, unless differently precised, and by means of them we can actually prove that ergodicity in the sense of Cesaro (resp. ergodicity) can be reduced to Problem 2.2 (resp. Problem 2.3); the equivalence between definitions and problems has been already approached in a rough way, but thanks to these assumptions all the missing details can be made precise. Indeed, in the case of Cesaro mean ergodicity, by Hypothesis 2.6 we know that $(\varepsilon u_\varepsilon)_{\varepsilon \in]0,1]}$ is equicontinuous and then there exists a subsequence $\varepsilon_n \downarrow 0$ such that $\varepsilon_n u_{\varepsilon_n}$ uniformly converges to a suitable function \bar{u} as $n \rightarrow \infty$. If we look at (1.24), that is the equation solved by $\varepsilon_n u_{\varepsilon_n}$, this allows us to pass to the limit as $n \rightarrow \infty$ and conclude that $A\bar{u} = 0$ and also that \bar{u} is continuous (uniform limit of continuous functions). As a final (but necessary) consideration, observe that if \bar{u} is constant, say $\bar{u} \equiv c$, then such constant might *a priori* depend on the choice of the convergent subsequence, but it is easy to check that this is not the case. In fact, let \mathbf{m} be the invariant measure and let us integrate the resolvent equation (1.21) w.r.t. it, namely

$$\int Au_{\varepsilon_n} d\mathbf{m} + \varepsilon_n \int u_{\varepsilon_n} d\mathbf{m} = \int f d\mathbf{m}$$

By passing to the limit as $n \rightarrow \infty$, we get

$$c = \int f d\mathbf{m}$$

and thus the proof of the claim. Thanks to what we have already said it is now completely clear why Problem 2.2 is equivalent to ergodicity in the sense of Cesaro.

In the case of ergodicity, the argument is more or less the same, because by Hypothesis 2.7 we know that $(u(T + s, \cdot))_{T \geq 0, s \geq -T}$ is equicontinuous and then there exists a subsequence $T_n \rightarrow +\infty$ such that $u(T_n + s, \cdot)$ converges uniformly in space and on compact subsets of time to a suitable function \bar{u} as $n \rightarrow \infty$; let us stress that the dependence on s is linked to the L^∞ -contractivity property (i) we discussed above. A different way to prove the existence of a subsequence converging uniformly in x and locally uniformly in t is the following: by property (iii) above we know that $\|A^k u(\cdot, t)\|_{L^\infty}$ is uniformly bounded in t for any $k \in \mathbb{N}$, so that just using the information for $k = 1$ we have

$$\left\| \frac{\partial u}{\partial s}(\cdot, T_n + s) \right\|_{L^\infty} \stackrel{(1,10)}{=} \|Au(\cdot, T_n + s)\|_{L^\infty} \leq C$$

for a suitable constant $C > 0$ independent of n . This means that in time we have not only equicontinuity, but even equi-Lipschitzianity, so that if we further assume equicontinuity (only) in space, then Ascoli-Arzelà theorem allows us to produce the desired subsequence. We can thus summarize all this preliminary discussion in the following result.

Theorem 2.8. *We have proved that:*

- *Definition 1.9 is equivalent to Problem 2.2 + Hypothesis 2.6;*
- *Definition 1.10 is equivalent to Problem 2.3 + Hypothesis 2.7.*

This means that the problem of ergodicity consists of two steps: the validity of an equicontinuous hypothesis and a positive answer for a certain number of harmonic functions to a test problem, which encodes most of the intuitive and geometric information about ergodicity; as a quick remark, the adjective *harmonic* is correct for both test problems: in the case of Problem 2.2, \bar{u} is harmonic for the stationary operator A , while in the case of Problem 2.3, \bar{u} is harmonic for $\partial_s + A$. For this reason, let us now investigate in more details Problem 2.2 and Problem 2.3, still in a compact framework, starting with the Cesaro mean ergodicity. That is, given

$$\begin{cases} A\bar{u} = 0 \\ \bar{u} \in C(Q) \end{cases} \quad (2.9)$$

what can we deduce?

From a stochastic point of view, as a main tool we will use the probabilistic representation (1.5), which reads for a harmonic function \bar{u} as

$$\bar{u}(x) = \mathbb{E}[\bar{u}(X_t) \mid X_0 = x]$$

However, in order to invoke it some technical results on the support of the process are required. It is the *support theorem*, which belongs to the probabilistic folklore, but for people who are not at ease with stochastic analysis a different approach can be found in [2], so that the probabilistic approach can be avoided. More details on this topic will be given later. Since Q is a compact domain and \bar{u} is continuous, there exists at least a maximum point $x_0 \in Q$ and for this point we have

$$\max_{x \in Q} \bar{u}(x) = \bar{u}(x_0) = \mathbb{E}[\bar{u}(X_t) \mid X_0 = x_0] \leq \mathbb{E} \left[\max_{x \in Q} \bar{u}(x) \mid X_0 = x_0 \right] = \max_{x \in Q} \bar{u}(x_0)$$

As a consequence,

$$\bar{u}(X_t) = \bar{u}(x_0), \quad \text{a.s.}$$

i.e. \bar{u} is constant along almost every trajectory issued from the maximum point x_0 and this is true for any maximum point. Actually, from what we have just presented we can only infer that, given t , $\bar{u}(X_t) = \bar{u}(x_0)$ almost surely, but by standard arguments for Markov processes it is possible to get a uniform in t almost certainty and so the previous claim is correct. If we replace x_0 by a minimum

point y_0 , the same conclusion holds, so that we can also say that \bar{u} is constant along almost every trajectory issued from any minimum point. Now the situation is clear: we have two families of curves, starting from x_0 and y_0 respectively, such that \bar{u} is constant along them and if one of these trajectories intersects a curve belonging to the other family, then

$$\max_{x \in Q} \bar{u}(x) = \min_{x \in Q} \bar{u}(x)$$

whence \bar{u} is constant, i.e. Problem 2.2 has positive answer. The intersection condition can actually be weakened, because \bar{u} is continuous; thus, it is sufficient for two curves issued from x_0 and y_0 respectively to be arbitrarily close. For this reason, another way to approach the problem is the following: the supports of the two families can be closed (i.e. for both families we close the union of the supports of the curves belonging to such family in the topology of Q) and this provides us with two compact subsets of Q . If the intersection of these sets is non-empty, then we conclude that \bar{u} is constant and Problem 2.2 has positive answer; otherwise, Cesaro mean ergodicity always fails, because it is sufficient to take as initial condition a function which is identically equal to 0 on one of the two compact subsets and 1 on the other one (by classical results of general topology, a function defined in this way always admits a continuous extension out of the two compact subsets): with this choice, the solution to our system (1.10) is constantly equal to 0 or 1 on the two compact subsets and then its Cesaro time mean can not converge towards a constant.

Hence, although we kept the discussion informal, the reader can see that Problem 2.2 turns out to be equivalent to decide whether the intersection of two compact sets is empty or not. Now the problem is: how do these sets look like?

Before approaching this new question, let us say that in the case of jump processes (notice that above we have considered continuous processes) the problem can be treated in the same fashion, as it is the support of the process that really matters. As a further digression, let us study (2.9) also from an analytical point of view. In this case, main tool is the strong maximum principle, which we are going to describe in a while; first, let us recall the (weak) maximum principle: it tells us that at a maximum point x_0 for \bar{u} we have

$$\nabla \bar{u}(x_0) = 0, \quad D^2 \bar{u}(x_0) \leq 0$$

In the case A is a local operator of the form (1.11), $A\bar{u} = 0$ becomes

$$a_{ij} \partial_{ij} \bar{u}(x_0) = 0$$

and since $(a_{ij})_{i,j}$, $(\partial_{ij})_{i,j}$ are two symmetric matrices, the former being positive and the latter negative, this yields

$$D^2 \bar{u}(x_0) \xi = 0$$

for any $\xi \in \mathbb{R}^d$ such that $\xi \cdot a\xi > 0$, where $a = (a_{ij})_{i,j}$. Notice that ξ is a non-degenerate direction, because the condition $\xi \cdot a\xi > 0$ means that ξ is an admissible diffusion direction for the underlying process. Thus, the (weak) maximum principle entails that at a maximum point x_0 the first and the second derivatives of \bar{u} along any non-degenerate direction vanish; it is easy to see that also the third derivative must be zero. After this consideration, one could be tempted to say that \bar{u} is constant along such directions, but this is not true in general, since we have no information on the fourth derivative. A complete information on all higher order derivatives follows by the strong maximum principle, which was first formulated by Hopf; it tells us that if A is uniformly elliptic (so that all $\xi \in \mathbb{R}^d$ is a non-degenerate direction) and \bar{u} is a harmonic function attaining its maximum at an interior point of the domain, then \bar{u} is constant. The result of Jean-Michel Bony [2] we were talking about lies between the weak and the strong formulation of the maximum principle, in the sense that it points out the conditions under which \bar{u} is constant along a given direction ξ . In order to present such result, let us come back to the probabilistic approach and more precisely to the support theorem, still assuming that A is a local operator of the form (1.11); the most general case,

where A is an integro-differential operator with a local part and a jump part, will not be considered, although it can be found in the literature (on the contrary, some examples of jump processes will be considered for Problem 2.3). But for the support theorem we first need to rewrite properly the operator A .

In the study of parabolic PDEs, there are essentially two viewpoints, given by

$$A = -a_{ij}\partial_{ij} + \text{first order terms} \quad (2.10)$$

$$A = -\partial_i a_{ij}\partial_j + \text{first order terms} \quad (2.11)$$

In (2.10) we have written the second order part of A as it appears in (1.11) and this approach is commonly adopted in stochastic calculus and optimal control theory. On the other hand, in (2.11) A is written in a so-called conservative form (linked to classical mechanics) but, under some regularity assumptions, (2.11) is easily seen to be equivalent to (2.10), up to first order terms. However, there is also a third point of view, which differs from (2.10) and (2.11) by first order terms; such approach suits particularly well for the geometric analysis of PDEs and for this reason we will adopt it. Roughly speaking, in this third situation we will try to express A in a squared form, as it had been done in Hörmander's work on the classification of linear PDEs. In order to be more precise, let us recall that we are assuming the matrix a to factorize as

$$a = \frac{1}{2}\sigma\sigma^t$$

so that

$$a_{ij} = \frac{1}{2}\sigma_i^\alpha\sigma_j^\alpha$$

where $\sigma_i^\alpha, \sigma_j^\alpha$ replace the standard notation $\sigma_{i\alpha}, \sigma_{j\alpha}$; this choice is motivated by the fact that for any fixed α we have a vector field, given by $\sigma^\alpha(x) := (\sigma_1^\alpha(x), \dots, \sigma_d^\alpha(x))^t$, and to any vector field one can easily associate a derivation operator, in this case

$$X^\alpha := \sigma_i^\alpha\partial_i$$

With this premise, the third possible way to express A is then the following

$$A = -\frac{1}{2}(X^\alpha)^2 + \text{first order terms} \quad (2.12)$$

It is worth saying that, from a probabilistic point of view, (2.10) corresponds to Itô's formulation of the associated SDE, while (2.12)

3 Lecture 3 - 21 November

4 Lecture 4 - 28 November

5 Lecture 5 - 5 December

6 Lecture 6 - 19 December

7 Lecture 7 - 9 January

8 Lecture 8 - 16 January

9 Lecture 9 - 6 February

A Auxiliary results

A.1 Liouville's theorem and problem

A.2 Tauberian theorems

A.3 A couple of theorems

Theorem A.1 (Perron-Frobenius).

Theorem A.2 (Krein-Rutman).

A.4 The Hille-Yosida theorem

References

- [1] N. Bogolyubov, N. Krylov, *La théorie générale de la mesure dans son application à l'étude des systèmes dynamiques de la mécanique non linéaire*, Ann. of Math., Second Series, vol. 38 (1937) pp. 65-113
- [2] J.-M. Bony, *Principe du maximum, inégalité de Harnack et unicité du problème de Cauchy pour les opérateurs elliptiques dégénérés*, Annales de l'Institut Fourier, vol. 19 (1969) pp. 277-304
- [3] X. Gabaix, J.-M. Lasry, P.-L. Lions, B. Moll, *The Dynamics of Inequality*, work in progress
- [4] S. P. Meyn, R. L. Tweedie, *Markov Chains and Stochastic Stability*, Springer Verlag (1993)
- [5] T. Piketty, *Le Capital au XXIe siècle*, Éditions du Seuil (2013)
- [6] R. Radu, R. Tanase, *The Mathematics of Web Search*, lecture notes available on <http://www.math.cornell.edu/mec/Winter2009/RalucaRemus/index.html>