Diffusion Approximations of Markovian Solutions to Discontinuous ODEs

Alberto Bressan(1), Marco Mazzola(2), and Khai T. Nguyen(3)

(1) Department of Mathematics, Penn State University,
(2) IMJ-PRG, CNRS, Sorbonne Université,
(3) Department of Mathematics, North Carolina State University.

e-mails: axb62@psu.edu, marco.mazzola@imj-prg.fr, khai@math.ncsu.edu

June 12, 2021

Abstract

In a companion paper, the authors have characterized all deterministic semigroups, and all Markov semigroups, whose trajectories are Carathéodory solutions to a given ODE $\dot{x} = f(x)$, with f possibly discontinuous. The present paper establishes two approximation results. Namely, every deterministic semigroup can be obtained as the pointwise limit of the flows generated by a sequence of ODEs $\dot{x} = f_n(x)$ with smooth right hand sides. Moreover, every Markov semigroup can be obtained as limit of a sequence of diffusion processes with smooth drifts and with diffusion coefficients approaching zero.

1 Introduction

Consider the Cauchy problem for a scalar ODE with possibly discontinuous right hand side:

$$\dot{x} \;=\; f(x), \tag{1.1}$$

$$x(0) \;=\; x_0\,. \tag{1.2}$$

By definition, a map $t \mapsto x(t)$ is a *Carathéodory solution* of (1.1)-(1.2) if

$$x(t) \;=\; x_0 + \int_0^t f(x(s))\,ds \qquad\qquad \text{for all } t \geq 0\,. \tag{1.3}$$

When the function f is not Lipschitz continuous, it is well known that this Cauchy problem can admit multiple solutions. Because of this non-uniqueness, in [7, 9] it was proposed to study "generalized flows", described by a probability measure on the set of all Carathéodory solutions. In this direction, in the companion paper [3] the authors have characterized all deterministic semigroups, and all Markov semigroups, whose trajectories are solutions to the ODE in (1.1)-(1.2).

Aim of the present paper is to show that all of these semigroups can be obtained as limits of smooth approximations. To explain these results more precisely, we recall

Definition 1.1 *A* **deterministic semigroup** *compatible with the ODE (1.1) is a map* $S : \mathbb{R} \times \mathbb{R}_+ \mapsto \mathbb{R}$*, with the properties*

(i) $S_t(S_s(x_0)) \;=\; S_{t+s}(x_0),\;\; S_0(x_0) \;=\; x_0.$

(ii) For each $x_0 \in \mathbb{R}$*, the map* $t \mapsto S_t x_0$ *is a solution to the Cauchy problem (1.1)-(1.2).*

Notice that here we do not require any continuity w.r.t. the initial point x_0. Of course, if f is Lipschitz continuous, then the solution $t \mapsto x(t) = S_t x_0$ of (1.1)-(1.2) is unique, and this uniquely determines the semigroup.

Throughout this paper, as in [3] we shall assume:

(A1) *The function* $f : \mathbb{R} \mapsto \mathbb{R}$ *is bounded and regulated. Namely,* f *admits left and right limits* $f(x-)$*,* $f(x+)$ *at every point* x*, while* $M \doteq \sup_x |f(x)| < +\infty$*.*

(A2) *If* y *is a point where either* $f(y-)\cdot f(y+) = 0$ *or else* $f(y-) > 0 > f(y+)$*, then* $f(y) = 0$*.*

As in [2], the "no jam" assumption **(A2)** guarantees that the Cauchy problem (1.1)-(1.2) has at least one solution. Indeed, in the present setting every solution in the sense of Filippov [8] is a Carathéodory solution as well. Under the above assumptions, the set of discontinuities

$$D_f \;=\; \Big\{x \in \mathbb{R}\,;\; f(x-) \neq f(x+) \;\;\text{or}\;\; f(x-) = f(x+) \neq f(x)\Big\} \tag{1.4}$$

is at most countable. Moreover, the set of zeroes

$$f^{-1}(0) \;=\; \big\{x \in \mathbb{R}\,;\; f(x) = 0\big\} \tag{1.5}$$

is closed. Given a regulated function $f : \mathbb{R} \mapsto \mathbb{R}$, we can complete its graph by adding a vertical segment at each point x where f has a jump. This yields a multifunction with closed graph and compact, convex values:

$$F(x) \;\doteq\; \text{co}\big\{f(x),\, f(x+),\, f(x-)\big\}, \tag{1.6}$$

where "co" denotes the convex closure. Throughout the sequel, $B(V, r) = \{x\,;\; d(x, V) < r\}$ denotes the open neighborhood of radius r around the set V.

Definition 1.2 *Let* f *be a regulated function and let* F *be the corresponding multifunction in (1.6). We say that a sequence of smooth functions* f_n *converges* f *in the sense of the graph if, for every given* $\varepsilon > 0$ *and* $N > 0$*, one has*

$$\text{Graph}\Big(f_n\Big|_{[-N,N]}\Big) \;\subset\; B\big(\text{Graph}(F)\,,\; \varepsilon\big) \tag{1.7}$$

for all n *sufficiently large.*

In other words, for all n sufficiently large, every point $(x, f_n(x))$, with $|x| \leq N$, is contained in an ε-neighborhood of the graph of F. We recall that this approach has been used in the literature, in the analysis of upper semicontinuous multifunctions with convex values [1, 4].

Our first main result shows that every deterministic semigroup S can be approximated by the semigroups S^n generated by a sequence of smooth ODEs.

$$\dot{x} = f_n(x), \qquad x(0) = x_0 . \tag{1.8}$$

Theorem 1.1 *Let f satisfy the assumptions* **(A1)-(A2)**. *Let $S : \mathbb{R} \times \mathbb{R}_+ \mapsto \mathbb{R}$ be a semigroup compatible with the ODE (1.1). Then there exists a sequence of smooth functions $(f_n)_{n \geq 1}$, converging to f in the sense of the graph, such that the following holds. For each $x_0 \in \mathbb{R}$, calling $t \mapsto S^n_t x_0$ the unique solution to the Cauchy problem (1.8), one has the convergence*

$$\lim_{n\to\infty} \sup_{t\geq 0} \left| S^n_t x_0 - S_t x_0 \right| = 0. \tag{1.9}$$

We remark that, since the map $x_0 \mapsto S_t x_0$ can be discontinuous, one can only achieve the pointwise convergence for each fixed initial datum $x_0 \in \mathbb{R}$. On the other hand, we can achieve uniform convergence w.r.t. the time variable $t \in [0, +\infty[$.

The second part of the paper is concerned with Markov semigroups. We consider a family of transition kernels $P_t(x_0, A)$, denoting the probability that a solution starting at x_0 reaches a point in the Borel set $A \subset \mathbb{R}$ at time t. It is assumed that these transition kernels satisfy the Chapman-Kolmogorov equation

$$P_{t+s}(x_0, A) = \int P_t(z, A)\, P_s(x_0, dz). \tag{1.10}$$

Definition 1.3 *We say that a Markov semigroup with transition kernels $P_t(\cdot,\cdot)$ is compatible with the ODE (1.1) if the following holds. There exists a probability space $\mathcal{W}$ and a Markov process $X(t,\omega)$, $\omega \in \mathcal{W}$, such that, for every $x_0 \in \mathbb{R}$ and every Borel set $A \subset \mathbb{R}$,*

$$\mathit{Prob.}\Big\{ X(t,\omega) \in A \;\Big|\; X(0,\omega) = x_0 \Big\} = P_t(x_0, A), \tag{1.11}$$

and moreover all sample paths $t \mapsto X(t,\omega)$, $\omega \in \mathcal{W}$, are Carathéodory solutions to the ODE (1.1).

Our second main result shows that every one of these Markov semigroups can be obtained as a limit of a sequence of diffusion processes, of the form

$$dX = f_n(X)\, dt + \sigma_n dW. \tag{1.12}$$

Here $(f_n)_{n\geq 1}$ is a sequence of smooth functions, converging to f in the sense of the graph, while the diffusion coefficients σ_n decrease to zero, as $n \to \infty$. As usual, by W we denote the standard one-dimensional Wiener process.

It is well known that the transition probability kernel $P^{(n)}$ for (1.12) can be obtained as follows. Let $\Gamma_n(t, x; x_0)$ be the fundamental solution to the linear parabolic equation

$$v_t + \big(f_n(x) v\big)_x = \frac{\sigma_n^2}{2}\, v_{xx} , \tag{1.13}$$

where the initial data is a unit mass at the point x_0, namely

$$\lim_{t\to 0+} \Gamma(t,x,x_0) \;=\; \delta_{x_0} \tag{1.14}$$

in distributional sense. Then, for every $t>0$ and every Borel set $A\subset\mathbb{R}$,

$$P_t^{(n)}(x_0,A) \;=\; \int_A \Gamma_n(t,y;x_0)\,dy. \tag{1.15}$$

The next theorem shows that, by a suitable choice of the drifts f_n, one can achieve the convergence $P^{(n)}\to P$ in distribution.

Theorem 1.2 *Let f satisfy the assumptions* **(A1)-(A2)**. *Let $P=P_t(x_0,A)$ be a family of transition kernels, defining a Markov semigroup compatible with the ODE (1.1). Then there exists a sequence of diffusion processes of the form (1.12) with $f_n\in\mathcal{C}^\infty$, such that*

- *(i) $f_n\to f$ in the sense of the graph,*
- *(ii) $\sigma_n\to 0$,*
- *(iii) for every $x_0\in\mathbb{R}$ and $t>0$, the transition kernels converge in distribution. Namely, for every bounded continuous function $\varphi\in\mathcal{C}^0(\mathbb{R})$, one has the convergence of the expected values*

$$\lim_{n\to\infty}\int \varphi(x)P_t^{(n)}(x_0,dx) \;=\; \int \varphi(x)P_t(x_0,dx). \tag{1.16}$$

Here the heart of the matter is to construct a sequence of piecewise constant drifts $f_n\in\mathbf{L}^\infty$ and diffusion coefficients $\sigma_n>0$ which satisfy (i)–(iii). The additional property $f_n\in\mathcal{C}^\infty$ is then achieved by suitable mollifications. We again remark that, since in general the limit semigroup has no continuous dependence w.r.t. the initial point x_0, the convergence $P_t^{(n)}(x_0,\cdot)\to P_t(x_0,\cdot)$ can only be attained in a pointwise sense, for each initial point x_0.

The remainder of the paper is organized as follows. Section 2 reviews some earlier results on discontinuous ODEs, while Section 3 contains a proof of Theorem 1.1. Finally, Theorem 1.2 is proved in Sections 4 to 6.

To explain the main ideas involved in these proofs we recall that, as proved in [3], to single out a unique deterministic semigroup compatible with (1.1), three additional ingredients are needed:

(Q1) A continuum (i.e., atomless) positive measure μ supported on the set of zeroes (1.5) of f. Strictly increasing trajectories $t\mapsto x(t)=S_t(x_0)$ of the semigroup are then implicitly defined by the identity

$$t \;=\; \int_{x_0}^{x(t)}\frac{dy}{f(y)}+\mu\big([x_0,x(t)]\big),$$

while a similar formula holds for decreasing ones. Notice that, if $f^{-1}(0)$ is countable, then necessarily $\mu=0$. However, in [3] an example was constructed where $f^{-1}(0)\subset[0,1]$ is the Cantor set, while f is Hölder continuous and strictly positive at all other points. In this case, different choices of the measure μ lead to infinitely many different semigroups, all compatible with the ODE.

(Q2) A countable set of points $\mathcal{S} \subseteq f^{-1}(0)$, where the dynamics is forced to stop. Among the (possibly many) solutions of (1.1) starting from $x_0 \in \mathcal{S}$, this means that we are selecting the stationary one: $S_t(x_0) = x_0$.

(Q3) A map $\Phi : \Omega^* \mapsto \{-1, 1\}$, defined on a set $\Omega^* \subset \mathbb{R}$ of isolated points from where both an increasing and a decreasing solution of (1.1) can originate. For $x_0 \in \Omega^*$, setting $\Phi(x_0) = 1$ selects the increasing solution, while $\Phi(x_0) = -1$ selects the decreasing one.

To prove Theorem 1.1, we first identify maximal open intervals $J_k =]x_{k-1}, x_k[$ where trajectories of the semigroup S are strictly increasing, or strictly decreasing. Then, to achieve the convergence (1.9), we construct a sequence of smooth functions f_n with the following property. For each interval J_k where the dynamics is increasing and every two points $a, b \in J_k$, if $b = S_\tau a$ for some $\tau = \tau(a, b)$, then

$$\lim_{n\to\infty} \int_a^b \frac{1}{f_n(x)}\, dx \;=\; \tau. \tag{1.17}$$

Notice that (1.17) yields the convergence of the times needed for trajectories to move from a to b. We choose the f_n so that the same property also holds on intervals where the dynamics is strictly decreasing.

Toward a proof of Theorem 1.2 we recall that, still by the results in [3], a general Markov semigroup compatible with (1.1) is determined by adding two more items to the list **(Q1)**–**(Q3)**. Namely:

(Q4) A countable set $\mathcal{S}^* \subset f^{-1}(0)$ and a map $\Lambda : \mathcal{S}^* \mapsto [0, +\infty]$, describing the random waiting time of a trajectory which reaches a point $x_k \in \mathcal{S}^*$. More precisely, a solution initially at $x_k \in \mathcal{S}^*$ remains at x_k for a random time $T_k \geq 0$, then starts moving. All these random waiting times are mutually independent, with Poisson distribution:

$$\text{Prob.}\{T_k > s\} \;=\; e^{-\lambda_k s} \qquad \text{with} \qquad \lambda_k \;=\; \Lambda(x_k). \tag{1.18}$$

(Q5) A map $\Theta : \Omega^* \mapsto [0, 1]$, defined on the countable set $\Omega^* \subset \mathbb{R}$ of points from which both an increasing and a decreasing solution of (1.1) can originate. For $x_k \in \Omega^*$, the value $\theta_k = \Theta(x_k)$ gives the probability that, when the solution starting from x_k begins to move, it will be increasing. Of course, $1 - \theta_k$ is then the probability that the solution will be decreasing.

After various approximations and reductions, discussed in Sections 4 and 5, we are led to study a discontinuous ODE having the basic form

$$\dot{x} \;=\; f(x) \;=\; \begin{cases} a & \text{if } x < 0, \\ 0 & \text{if } x = 0, \\ b & \text{if } x > 0, \end{cases} \tag{1.19}$$

for some $a, b \neq 0$. For a Markov semigroup compatible with (1.19), three main cases must be considered, see Fig. 1.

CASE 1: $a, b > 0$ and all trajectories reaching the origin remain at the origin forever after.

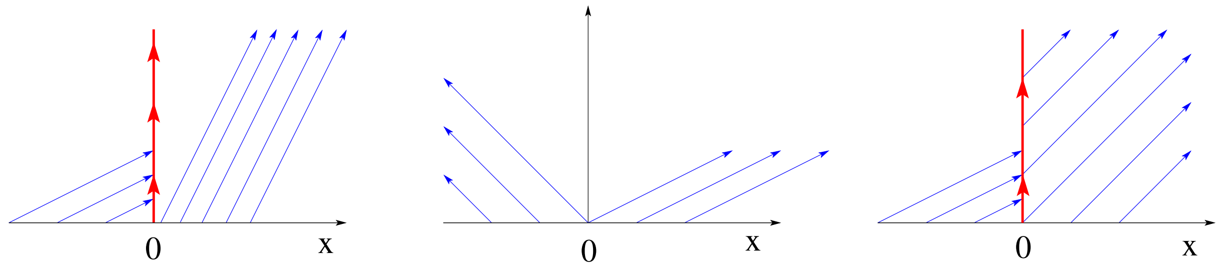

Figure 1: Left: Case 1, where all trajectories stop at the origin. Center: Case 2, where trajectories starting at the origin move to the right or to the left with probabilities θ and $1-\theta$. Right: Case 3, where trajectories reaching the origin wait for a random time, then start moving again to the right.

CASE 2: $a < 0 < b$ and all trajectories starting at the origin move to the right with probability $\theta \in [0,1]$ and to the left with probability $1-\theta$.

CASE 3: $a, b > 0$ and all trajectories reaching the origin wait for a random time T with Poisson distribution

$$\text{Prob.}\{T > s\} = e^{-\lambda s}, \tag{1.20}$$

then start moving to the right with speed b.

For the above three cases, sequences of piecewise constant drifts g_n achieving the desired convergence are constructed in (6.2), (6.6), and (6.19), respectively.

The approximation of a random waiting time (1.20) requires a more careful analysis. This is based on the preliminary construction of a family of eigenfunctions u_n of a corresponding rescaled problem on the interval $[0,1]$, with eigenvalues $\lambda_n \to \lambda$. To help the reader, the distribution function for the limit Markov semigroup is shown in Fig. 7. In addition, a lower and and an upper solution to the parabolic equation describing the approximating diffusion process are shown in Figures 8 and 9, respectively.

The convergence of a sequence of stochastic processes to a diffusion process is a classical topic in probability theory. See for example [10, 11] or Chapter 11 in [14]. The present result goes in the converse direction, approximating a Markov process with smooth diffusions. In a way, this is in the same spirit as [5]. Indeed, we show that the existence of diffusion approximations is not a selection principle for any of the Markov semigroups compatible with (1.1). A survey of results on solutions to generalized ODEs can be found in [6].

2 Review of scalar discontinuous ODEs

As a preliminary, we collect here some basic results on the existence and properties of solutions to a Cauchy problem with possibly discontinuous right hand side. In the following theorem, the first two statements are proved in somewhat greater generality in [2]. On the other hand, the closure of the solution set strongly relies on the assumption **(A1)** that the function f is regulated. See [3] for a proof.

Theorem 2.1 *Consider the Cauchy problem (1.1)-(1.2), assuming that* **(A1)-(A2)** *hold. Then*

(i) For every $x_0 \in \mathbb{R}$ there exists at least one Carathéodory solution, defined for all times $t \geq 0$.

(ii) Every solution is monotone (either increasing or decreasing).

(iii) If $x_n : [0, \tau] \mapsto \mathbb{R}$ is a sequence of solutions of (1.1), and the pointwise convergence $x_n(t) \to x(t)$ holds for every $t \in [0, \tau]$, then $x(\cdot)$ is a solution as well.

The general form of a semigroup S compatible with the ODE (1.1) was described in [3]. To uniquely determine one of these semigroups, together with the ODE one needs to assign a positive, atomless measure μ supported on $f^{-1}(0)$, a countable set $\mathcal{S} \subseteq f^{-1}(0)$, and a map $\Phi : \Omega^* \mapsto \{-1, 1\}$, as described at **(Q1)–(Q3)** in the Introduction.

We briefly review the construction in [3]. Given f, consider the regulated functions

$$f^+(x) \doteq \max\{0, f(x)\}, \qquad f^-(x) \doteq \min\{0, f(x)\}. \tag{2.1}$$

Definition 2.1 *We say that an open interval $]a, b[$ is a* **domain of increase** *if*

$$]a, b[\ \cap\ \mathcal{S} = \emptyset, \qquad \mu([c, d]) + \int_c^d \frac{dx}{f^+(x)} < +\infty \qquad \textit{for all } [c, d] \subset]a, b[\,. \tag{2.2}$$

Similarly, we say that $]a, b[$ is a **domain of decrease** *if*

$$]a, b[\ \cap\ \mathcal{S} = \emptyset, \qquad \mu([c, d]) - \int_c^d \frac{dx}{f^-(x)} < +\infty \qquad \textit{for all } [c, d] \subset]a, b[\,. \tag{2.3}$$

If $]a, b[$ and $]a', b'[$ are two intervals of increase having non-empty intersection, then their union $]a, b[\cup]a', b'[$ is also an interval of increase. We can thus identify countably many, disjoint maximal intervals of increase $]\alpha_i, \beta_i[$, $i \in \mathcal{I}^+$. Similarly, we can identify countably many disjoint maximal intervals of decrease $]\gamma_i, \delta_i[$, $i \in \mathcal{I}^-$.

It remains to analyze what happens at the endpoints of these intervals.

- If $\alpha_i \notin \mathcal{S}$ and, for some $\varepsilon > 0$

  $$\mu([\alpha_i,\, \alpha_i + \varepsilon]) + \int_{\alpha_i}^{\alpha_i + \varepsilon} \frac{dx}{f^+(x)} < +\infty, \tag{2.4}$$

  we then consider the half-open interval $I_i^+ \doteq [\alpha_i, \beta_i[\,$. Otherwise, we let I_i^+ be an open interval: $I_i^+ \doteq]\alpha_i, \beta_i[\,$.

- If $\delta_i \notin \mathcal{S}$ and, for some $\varepsilon > 0$

  $$\mu([\delta_i - \varepsilon,\, \delta_i]) - \int_{\delta_i - \varepsilon}^{\delta_i} \frac{dx}{f^-(x)} < +\infty, \tag{2.5}$$

  we then consider the half-open interval $I_i^- \doteq]\gamma_i, \delta_i]\,$. Otherwise, we let I_i^- be an open interval: $I_i^- \doteq]\gamma_i, \delta_i[\,$.

For each $i \in \mathcal{I}^+$, we now describe the increasing dynamics on the intervals I_i^+. Given $x_0 \in I_i^+$, we consider the time

$$\tau^+(x_0) \doteq \mu([x_0, \beta_i]) + \int_{x_0}^{\beta_i} \frac{dy}{f^+(y)} \in \,]0, +\infty].$$

We then set

$$S_t^+(x_0) \doteq x(t), \tag{2.6}$$

where $x(t)$ is implicitly defined by

$$\begin{cases} \mu([x_0, x(t)]) + \displaystyle\int_{x_0}^{x(t)} \frac{dy}{f^+(y)} = t & \text{if } t < \tau^+(x_0), \\ x(t) = \beta_i & \text{if } t \geq \tau^+(x_0). \end{cases} \tag{2.7}$$

The construction of the decreasing dynamics on the intervals I_i^- is entirely similar. Given $x_0 \in I_i^-$, we consider the time

$$\tau^-(x_0) \doteq \mu([\gamma_i, x_0]) - \int_{\gamma_i}^{x_0} \frac{dy}{f^+(y)} \in \,]0, +\infty].$$

We then set

$$S_t^-(x_0) \doteq x(t), \tag{2.8}$$

where $x(t)$ is implicitly defined by

$$\begin{cases} \mu([x(t), x_0]) - \displaystyle\int_{x(t)}^{x_0} \frac{dy}{f^-(y)} = t & \text{if } t < \tau^-(x_0), \\ x(t) = \beta_i & \text{if } t \geq \tau^-(x_0). \end{cases} \tag{2.9}$$

We can now combine together the above solutions, and define the semigroup S on the whole real line, as follows.

$$S_t(x_0) = \begin{cases} x_0 & \text{if } \; x_0 \notin \left(\bigcup_i I_i^+\right) \cup \left(\bigcup_i I_i^-\right), \\ S_t^+(x_0) & \text{if } \; x_0 \in \left(\bigcup_i I_i^+\right) \setminus \left(\bigcup_i I_i^-\right), \\ S_t^-(x_0) & \text{if } \; x_0 \in \left(\bigcup_i I_i^-\right) \setminus \left(\bigcup_i I_i^+\right). \end{cases} \tag{2.10}$$

To complete the definition, it remains to define $S_t(x_0)$ in the case $x_0 \in I_i^+ \cap I_j^-$, for some $i \in \mathcal{I}^+$, $j \in \mathcal{I}^-$. Notice that this can happen only if

$$x_0 = \alpha_i = \delta_j,$$

where $I_i^+ = [\alpha_i, \beta_i[$ and $I_j^- =]\gamma_j, \delta_j]$ are half-open intervals where the dynamics is increasing and decreasing, respectively. By our definitions, this implies $x_0 \in \Omega^*$. Recalling **(Q3)**, we thus define

$$S_t(x_0) = \begin{cases} S_t^+(x_0) & \text{if } \; \Phi(x_0) = 1, \\ S_t^-(x_0) & \text{if } \; \Phi(x_0) = -1. \end{cases} \tag{2.11}$$

The first main result in [3] shows that every semigroup compatible with the ODE (1.1) has the above form.

Theorem 2.2 *Let $f : \mathbb{R} \mapsto \mathbb{R}$ satisfy the assumptions* **(A1)-(A2)**. *The following statements are equivalent.*

(i) The map $S : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ is a deterministic semigroup compatible with the ODE (1.1).

(ii) There exist a positive atomless measure μ supported on the set $f^{-1}(0)$, a countable set of points $\mathcal{S} \subseteq f^{-1}(0)$, and a map $\Phi : \Omega^* \mapsto \{-1, 1\}$ as in **(Q1)–(Q3)** *such that S coincides with the corresponding semigroup constructed at (2.6)–(2.11).*

Next, we consider a Markov semigroup whose sample paths satisfy the ODE (1.1) with probability one. To uniquely determine such a Markov process, in addition to **(Q1)-(Q2)** one needs to assign:

- A countable set $\mathcal{S}^* \subseteq f^{-1}(0)$ and, for each $y_j \in \mathcal{S}^*$, a number $\lambda_j > 0$ characterizing the Poisson waiting time, when trajectories reach the point $y_j \in \mathcal{S}^*$, as in **(Q4)**.
- A map $\Theta : \Omega^* \mapsto [0, 1]$ determining the probability of moving upwards or downwards, from an initial point $z_k \in \Omega^*$, as in **(Q5)**.

We start by selecting an underlying probability space $\mathcal{W}$ such that, as $\omega \in \mathcal{W}$, the countably many random variables $Y_j(\omega) \in \mathbb{R}_+$ and $Z_k(\omega) \in \{-1, 1\}$ are independent, with distributions

$$\text{Prob.}\{Y_j > s\} \ = \ e^{-\lambda_j s} \quad \text{for all } s > 0, \qquad\qquad \text{Prob.}\{Z_k = 1\} \ = \ \Theta(z_k). \tag{2.12}$$

Next, we recall that, given the measure μ and the countable set $\mathcal{S}$, as in Section 3 one can uniquely determine the sets

$$\Omega^+ \ = \ \bigcup_{i \in \mathcal{I}^+} I_i^+ , \qquad\qquad \Omega^- \ = \ \bigcup_{i \in \mathcal{I}^-} I_i^- \tag{2.13}$$

consisting of countable unions of disjoint intervals where the dynamics can increase or decrease, respectively. For $x_0 \in \Omega^+$, a trajectory $t \mapsto S_t^+(x_0)$ was defined at (2.6)-(2.7), while for $x_0 \in \Omega^-$, a trajectory $t \mapsto S_t^-(x_0)$ was defined at (2.8)-(2.9).

To construct a Markov process with sample paths $t \mapsto X(t, \omega)$ satisfying (1.1), for every $t > 0$ we need to define the transition probabilities

$$P_t(x_0, A) \ = \ \text{Prob.}\Big\{ X(t, \omega) \in A \ \Big| \ X(0, \omega) = x_0 \Big\},$$

for any initial point $x_0 \in \mathbb{R}$ and any Borel set $A \subset \mathbb{R}$. Recalling our construction of a deterministic flow at (2.10)-(2.11), this can be done as follows.

(i) If $x_0 \notin \Omega^+ \cup \Omega^-$, then $X(t, x_0, \omega) = x_0$ for every $t \geq 0$ and $\omega \in \Omega$. Hence

$$P_t(x_0, \{x_0\}) \ = \ 1 \qquad \text{for all } \ t \geq 0. \tag{2.14}$$

In other words, all trajectories starting at x_0 remain constant.

(ii) If $x_0 \in \Omega^+ \setminus \Omega^-$, a random trajectory starting at x_0 will have the form

$$X(t, x_0, \omega) \;=\; S^+_{T^+(t,x_0,\omega)}(x_0)\,, \tag{2.15}$$

where the time T^+ along the trajectory is a random variable with distribution

$$\text{Prob.}\big\{T^+(t,x_0,\omega) < s\big\} \;=\; \text{Prob.}\left\{ s + \sum_{y_j \in \mathcal{S}^* \cap [x_0,\, S_s^+(x_0)]} Y_j(\omega) \;>\; t \right\}. \tag{2.16}$$

Note that (2.16) accounts for the (possibly countably many) waiting times when the trajectory crosses one of the points $y_j \in \mathcal{S}^*$. The transition probabilities are thus given by

$$\begin{aligned} P_t\big(x_0;\ ]-\infty, x_0[\big) \;&=\; 0,\\ P_t\big(x_0;\ [x_0, S_s^+(x_0)]\big) \;&=\; \text{Prob.}\left\{ s + \sum_{y_j \in \mathcal{S}^* \cap [x_0,\, S_s^+(x_0)]} Y_j(\omega) \;\geq\; t \right\}. \end{aligned} \tag{2.17}$$

(iii) Similarly, for an initial state $x_0 \in \Omega^- \setminus \Omega^+$, a random trajectory starting at x_0 will have the form

$$X(t, x_0, \omega) \;=\; S^-_{T^-(t,x_0,\omega)}(x_0)\,, \tag{2.18}$$

where the time T^- along the trajectory is a random variable with distribution

$$\text{Prob.}\big\{T^-(t,x_0,\omega) < s\big\} \;=\; \text{Prob.}\left\{ s + \sum_{y_j \in \mathcal{S}^* \cap [S_s^-(x_0),\, x_0]} Y_j(\omega) \;>\; t \right\}. \tag{2.19}$$

The transition probabilities are thus given by

$$\begin{aligned} P_t\big(x_0;\ ]x_0, +\infty[\big) \;&=\; 0,\\ P_t\big(x_0;\ [S_s^-(x_0), x_0]\big) \;&=\; \text{Prob.}\left\{ s + \sum_{y_j \in \mathcal{S}^* \cap [S_s^-(x_0),\, x_0]} Y_j(\omega) \;\geq\; t \right\}. \end{aligned} \tag{2.20}$$

(iv) To complete the definition, it remains to define the transition probabilities in the case $x_0 \in \Omega^+ \cap \Omega^- \subseteq \Omega^*$. In this case, by construction we have $x_0 = z_k$ for some k. We then define the random variable $X(t, x_0, \omega)$ by setting

$$X(t, x_0, \omega) \;=\; \begin{cases} S^+_{T^+(t,x_0,\omega)}(x_0) & \text{if} \quad Z_k(\omega) = 1\,,\\[2mm] S^-_{T^-(t,x_0,\omega)}(x_0) & \text{if} \quad Z_k(\omega) = -1\,. \end{cases} \tag{2.21}$$

By (2.12), its distribution satisfies

$$P_t(x_0, A) \;=\; \Theta(x_0)\cdot \text{Prob.}\Big\{S^+_{T^+(t,x_0,\omega)} \in A\Big\} + (1-\Theta(x_0))\cdot \text{Prob.}\Big\{S^-_{T^-(t,x_0,\omega)} \in A\Big\} \tag{2.22}$$

for every Borel set $A \subseteq \mathbb{R}$.

From the above construction it is clear that, in all cases (i)–(iv), every sample path $t \mapsto X(t, x_0, \omega)$ is a Carathéodory solution of the Cauchy problem (1.1)-(1.2). Indeed, we are only adding a countable number of waiting times, when the random trajectories reach one of the points $y_j \in \mathcal{S}^* \subseteq f^{-1}(0)$. We conclude by stating the second main result proved in [3].

Theorem 2.3 *Let f be a function satisfying* **(A1)-(A2)**. *The following statements are equivalent.*

- *(I) The random variables $X(t, x_0, \omega)$ yield a Markov process whose sample paths are solutions to the ODE (1.1)-(1.2).*
- *(II) There exist: (i) a positive, atomless Borel measure μ supported on $f^{-1}(0)$, (ii) a countable set $\mathcal{S} \subseteq f^{-1}(0)$ of stationary points, (iii) a countable set $\mathcal{S}^* = \{y_j : j \geq 1\} \subseteq f^{-1}(0)$ and corresponding numbers $\lambda_j > 0$ determining the Poisson waiting times, and (iv) a map $\Theta : \Omega^* \mapsto [0, 1]$, such that the transition kernels $P_t(x_0, A) = Prob.\{X(t, x_0, \omega) \in A\}$ coincide with the corresponding ones constructed at (2.14), (2.17), (2.20), (2.22).*

3 Approximating a deterministic semigroup by smooth flows

In this section we give a proof of Theorem 1.1, in several steps.

1. Consider the decomposition

$$\mathbb{R} \;=\; \Omega_S^+ \cup \Omega_S^- \cup \Omega_S^0 \tag{3.1}$$

where

- $\Omega_S^+ = \bigcup_{j \in \mathcal{J}^+} J_j^+$ is the union of countably many disjoint intervals $J_j^+ = [a_j, b_j[$ or $J_j^+ =]a_j, b_j[$ open to the right, where the dynamics is strictly increasing and b_j cannot be crossed from the left, i.e. $S_t(b_j - \varepsilon) \leq b_j$ for all $t \geq 0$, $\varepsilon \in]0, b_j - a_j[$.
- $\Omega_S^- = \bigcup_{j \in \mathcal{J}^-} J_j^-$ is the union of countably many disjoint intervals $J_j^- =]c_j, d_j]$ or $J_j^- =]c_j, d_j[$ open to the left, where the dynamics is strictly decreasing and c_j cannot be crossed from the right, i.e. $S_t(c_j + \varepsilon) \geq c_j$ for all $t \geq 0$, $\varepsilon \in]0, d_j - c_j[$.
- $\Omega_S^0 \subseteq f^{-1}(0)$ is the set of points x_0 such that $S_t(x_0) = x_0$ for all $t > 0$.

Let μ be the positive atomless measure supported on the set $f^{-1}(0)$, described at **(Q1)**. Recall that μ is finite in any compact subset of J_i^+ or J_i^-.

2. Let $\varepsilon > 0$ be given. As an intermediate step, we shall approximate f with a piecewise constant function g_ε. Consider the two sets of indices

$$\mathcal{J}_\varepsilon^+ \;\doteq\; \Big\{ i \in \mathcal{J}^+ \,;\; |J_i^+| = b_i - a_i \geq \varepsilon \Big\}, \qquad \mathcal{J}_\varepsilon^- \;\doteq\; \Big\{ i \in \mathcal{J}^- \,;\; |J_i^-| = d_i - c_i \geq \varepsilon \Big\}, \tag{3.2}$$

labelling the intervals with length $\geq \varepsilon$. For each $i \in \mathcal{J}_\varepsilon^+$, on the interval J_i^+ where the dynamics is increasing we perform the following construction.

CASE 1: If $J_i^+ =]a_i, b_i[$ is bounded and open, we insert points $a_i < x_{i,1} < \cdots < x_{i,N_i} < b_i$ so that

$$\frac{\varepsilon}{3} < x_{i,1} - a_i < \varepsilon, \qquad \frac{\varepsilon}{3} < b_i - x_{i,N_i} < \varepsilon, \qquad x_{i,k} - x_{i,k-1} \ < \ \varepsilon \tag{3.3}$$

$$\tau_{i,k} \ \doteq \ \int_{x_{i,k-1}}^{x_{i,k}} \frac{dy}{f(y)} + \mu\big([x_{i,k-1}, x_{i,k}]\big) \ < \ \varepsilon \quad \text{for all } k \in \{2, \ldots, N_i\}. \tag{3.4}$$

We then define the piecewise constant function $g_\varepsilon : \Big]a_i + \dfrac{\varepsilon}{3}\,,\ b_i - \dfrac{\varepsilon}{3}\Big[\mapsto \mathbb{R}$ by setting

$$g_\varepsilon(x) \ = \ 0 \qquad \text{for all } x \in \Big]a_i + \frac{\varepsilon}{3},\, x_{i,1}\Big[\,\cup\, \Big[x_{i,N_i},\, b_i - \frac{\varepsilon}{3}\Big[,$$

$$g_\varepsilon(x) \ = \ \frac{x_{i,k} - x_{i,k-1}}{\tau_{i,k}} \qquad \text{if} \quad x \in \big[x_{i,k-1}, x_{i,k}\big[\quad \text{for some } k \in \{2, \ldots, N_i\}\,.$$

CASE 2: If $J_i^+ = [a_i, b_i[$ is bounded and half-open, we insert points $a_i = x_{i,0} < x_{i,1} < \cdots < x_{i,N_i} < b_i$ so that

$$\frac{\varepsilon}{3} < b_i - x_{i,N_i} < \varepsilon, \qquad x_{i,k} - x_{i,k-1} < \varepsilon \tag{3.5}$$

$$\tau_{i,k} \ \doteq \ \int_{x_{i,k-1}}^{x_{i,k}} \frac{dy}{f(y)} + \mu\big([x_{i,k-1}, x_{i,k}]\big) < \varepsilon \quad \text{for all } k \in \{1, \ldots, N_i\}\,. \tag{3.6}$$

We then define the piecewise constant function $g_\varepsilon : \Big[a_i - \dfrac{\varepsilon}{3}\,,\ b_i - \dfrac{\varepsilon}{3}\Big[\mapsto \mathbb{R}$ by setting

$$g_\varepsilon(x) \ = \ 0 \qquad \text{for all } x \in \Big[x_{i,N_i}, b_i - \frac{\varepsilon}{3}\Big[,$$

$$g_\varepsilon(x) \ = \ \frac{x_{i,k} - x_{i,k-1}}{\tau_{i,k}} \qquad \text{if} \quad x \in \big[x_{i,k-1}, x_{i,k}\big[\quad \text{for some } k \in \{2, \ldots, N_i\}\,,$$

$$g_\varepsilon(x) \ = \ \frac{x_{i,1} - x_{i,0}}{\tau_{i,1}} \qquad \text{if} \quad x \in \Big[a_i - \frac{\varepsilon}{3}, x_{i,1}\Big[\,.$$

If J_i^+ is unbounded, we can perform the same construction as above by inserting countably many points. In this case, $N_i = \infty$.

Next, for each $i \in \mathcal{J}_\varepsilon^-$, we perform an entirely similar construction on the interval $J_i^- =]c_i, d_i[$ or $J_i^- =]c_i, d_i]$ where the dynamics is decreasing. Notice that the choice (3.3), (3.5) of the nodes $x_{i,k}$ of each partition imply that the domains where the functions g_ε are defined are strictly disjoint.

We then extend the function g_ε to the whole real line $\mathbb{R}$, by defining $g_\varepsilon(x) \ = \ 0$ on the remaining set.

3. In this step, we approximate g_ε with a smooth function $f_\varepsilon : \mathbb{R} \mapsto \mathbb{R}$, with the following properties.

(i) Introducing the convex valued multifunction $G_\varepsilon(x) \ \doteq \ \mathrm{co}\{g_\varepsilon(x+),\, g_\varepsilon(x-)\}$, one has $\mathrm{Graph}(f_\varepsilon) \subseteq B\,(\mathrm{Graph}(G_\varepsilon), \varepsilon)$. Moreover,

$$f_\varepsilon(x) \ = \ 0 \quad \text{for all } x \in \mathbb{R} \setminus \left[\Big(\bigcup_{i \in \mathcal{J}_\varepsilon^+} \Big[a_i - \frac{\varepsilon}{2},\, b_i\Big[\Big) \cup \Big(\bigcup_{i \in \mathcal{J}_\varepsilon^i} \Big]c_i,\, d_i + \frac{\varepsilon}{2}\Big]\Big)\right]. \tag{3.7}$$

(ii) For every $i \in \mathcal{J}_\varepsilon^+$, if $J_i^+ =]a_i, b_i[$ is open then $f_\varepsilon(a_i) = 0$ and

$$\int_{x_{i,k-1}}^{x_{i,k}} \frac{dy}{f_\varepsilon^+(y)} \;=\; \tau_{i,k} \quad \text{for all } k = 2, \ldots, N_i; \tag{3.8}$$

On the other hand, if $J_i^+ = [a_i, b_i[$ is half open, then (3.8) holds for all $k = 1, \ldots, N_i$.

(iii) The analogous conditions hold for the intervals J_i^- with $i \in \mathcal{J}_\varepsilon^-$.

Notice that all the above properties can be achieved by a suitable mollification.

4. We claim that, as $\varepsilon \to 0+$, the functions f_ε converge to f in the sense of the graph. Since $\text{Graph}(f_\varepsilon) \subseteq B\,(\text{Graph}(G_\varepsilon), \varepsilon)$, it will be sufficient to show that

$$(x, g_\varepsilon(x)) \;\in\; B\left(\text{Graph}(F), \frac{4\varepsilon}{3}\right) \qquad \text{for all } x \in \mathbb{R}, \tag{3.9}$$

since this yields $\text{Graph}(G_\varepsilon) \;\subseteq\; B\left(\text{Graph}(F), \dfrac{4\varepsilon}{3}\right)$. Let us prove (3.9) for $g_\varepsilon(x) \neq 0$.

- If $x \in \left[a_i - \dfrac{\varepsilon}{3}, b_i\right[$ for some $i \in \mathcal{J}_\varepsilon^+$ then

$$x \;\in\; \left[a_i - \frac{\varepsilon}{3}, x_{i,1}\right[\qquad \text{or} \qquad x \in \left[x_{i,k-1}, x_{i,k}\right[\quad \text{for some } k \in \{2, \ldots, N_i\}.$$

Assume that $x \in \left[a_i - \dfrac{\varepsilon}{3}, x_{i,1}\right[$. Then

$$g_\varepsilon(x) \;=\; \frac{1}{\tau_{i,1}} \cdot \int_0^{\tau_{i,1}} f(S_s(x_{i,0}))ds \;\in\; \left[\inf_{y \in [x_{i,0}, x_{i,1}]} f(y), \sup_{y \in [x_{i,0}, x_{i,1}]} f(y)\right].$$

Thus, since $0 < x_{i,1} - a_i < \varepsilon$, it follows that $(x, g_\varepsilon(x)) \in B\left(\text{Graph}(F), \frac{4\varepsilon}{3}\right)$. Otherwise, if $x \in \left[x_{i,k-1}, x_{i,k}\right[$ for some $k \in \{2, \ldots, N_i\}$, then

$$g_\varepsilon(x) \;=\; \frac{1}{\tau_{i,k}} \cdot \int_0^{\tau_{i,k}} f(S_s(x_{i,k-1}))ds \;\in\; \left[\inf_{y \in \left[x_{i,k-1}, x_{i,k}\right]} f(y), \sup_{y \in \left[x_{i,k-1}, x_{i,k}\right]} f(y)\right],$$

and this implies that $(x, g_\varepsilon(x)) \in (\text{Graph}(F), \varepsilon)$.

- Similarly, one can show that (3.9) holds if $x \in \left]c_i, d_i + \dfrac{\varepsilon}{3}\right]$ for some $i \in \mathcal{J}_\varepsilon^-$.

Let us now prove (3.9) in the case where $g_\varepsilon(x) = 0$. This is trivial when $x \in \Omega_S^0$. On the other hand, if $x \notin \Omega_S^0 \cup \left(\bigcup_{i \in \mathcal{J}_\varepsilon^+} I_i^+\right) \cup \left(\bigcup_{i \in \mathcal{J}_\varepsilon^-} I_i^-\right)$, then

$$|x - c_i| \;<\; \varepsilon \quad \text{for some } i \notin \mathcal{J}_\varepsilon^+ \qquad \text{or} \qquad |x - b_i| \;<\; 0 \quad \text{for some } i \notin \mathcal{J}_\varepsilon^-.$$

Recalling that $f(c_i) = f(b_i) = 0$, we obtain (3.9). Now, assuming that $x \in J_i^+$ for some $i \in \mathcal{J}_\varepsilon^+$, we have

$$|x - b_i| \;\leq\; \frac{4}{3} \cdot \varepsilon \qquad \text{or} \qquad |x - a_i| \;\leq\; \varepsilon, \;\; a_i \notin I_i^+.$$

If $|x-b_i| \le \dfrac{4}{3}\varepsilon$ then (3.9) holds because $f(b_i)=0$. Otherwise, since $a_i \notin J_i^+$, one has $f(a_i)=0$ or $f(a_i-) < 0 < f(a_i+)$ and this also implies (3.9).

Similarly, one can show that (3.9) holds if $x \in J_i^-$ for some $i \in J_\varepsilon^-$.

5. To complete the proof, it remains to show that, for each $x_0 \in \mathbb{R}$, the unique solution $t \mapsto S_t^\varepsilon x_0$ to the Cauchy problem

$$\dot{x} \;=\; f_\varepsilon(x), \qquad\qquad x(0) \;=\; x_0,$$

converges uniformly to $t \mapsto S_t x_0$, i.e.

$$\lim_{\varepsilon\to 0+} \left(\sup_{t\ge 0} \left| S_t^\varepsilon(x_0) - S_t(x_0) \right| \right) \;=\; 0. \tag{3.10}$$

Several cases will be considered:

CASE 1: $x_0 \in J_i^+$, where $i \in \mathcal{J}_\varepsilon^+$ and $J_i^+ =]a_i, b_i[$ is open. For every ε such that

$$0 \;<\; \varepsilon \;<\; \min\left\{ x_0 - a_i,\; b_i - x_0,\; \frac{3}{7} \right\},$$

one then has

$$\|f_\varepsilon\|_\infty \;\le\; \|f\|_{\mathbf{L}^\infty} + \frac{7\varepsilon}{3} \;\le\; M+1 \qquad \text{and} \qquad x_0 \in \big[x_{i,k-1}, x_{i,k}\big[\;\text{ for some } k \in \{2,\dots,N_i\}.$$

Let $\tau_0, \tau_0^\varepsilon \in]0,\varepsilon[$ be such that $S_{\tau_0}(x_0) = S^\varepsilon_{\tau_0^\varepsilon}(x_0) = x_{i,k}$. For any $0 < t < \max\{\tau_0, \tau_0^\varepsilon\}$ one has

$$\left| S_t^\varepsilon(x_0) - S_t(x_0) \right| \;\le\; (M+1)\varepsilon.$$

On the other hand, if $t \ge \max\{\tau_0, \tau_0^\varepsilon\}$, we have

$$\begin{aligned} \left| S_t^\varepsilon(x_0) - S_t(x_0) \right| \;&=\; \left| S^\varepsilon_{t-\tau_0^\varepsilon}\big(x_{i,k}\big) - S_{t-\tau_0}\big(x_{i,k}\big) \right| \\ &\le\; \left| S^\varepsilon_{t-\tau_0^\varepsilon}\big(x_{i,k}\big) - S^\varepsilon_{t-\tau_0}\big(x_{i,k}\big) \right| + \left| S^\varepsilon_{t-\tau_0}\big(x_{i,k}\big) - S_{t-\tau_0}\big(x_{i,k}\big) \right| \\ &\le\; (M+1)\varepsilon + \left| S^\varepsilon_{t-\tau_0}\big(x_{i,k}\big) - S_{t-\tau_0}\big(x_{i,k}\big) \right|. \end{aligned}$$

To estimate the second term $\left| S^\varepsilon_{t-\tau_0}\big(x_{i,k}\big) - S_{t-\tau_0}\big(x_{i,k}\big) \right|$, we consider two subcases:

- if $S_{t-\tau_0}\big(x_{i,k}\big) \in \big[x_{i,k_0-1}, x_{i,k_0}\big[$ for some $k_0 \in \{k+1,\dots,N_i\}$, then

$$\begin{aligned} \left| S^\varepsilon_{t-\tau_0}\big(x_{i,k}\big) - S_{t-\tau_0}\big(x_{i,k}\big) \right| \;&=\; \left| S^\varepsilon_{t-\tau_0-\sum_{k'=k+1}^{k_0-1}\tau_{i,k'}}\big(x_{i,k_0-1}\big) - S_{t-\tau_0-\sum_{k'=k+1}^{k_0-1}\tau_{i,k'}}\big(x_{i,k_0-1}\big) \right| \\ &\le\; x_{i,k_0} - x_{i,k_0-1} \;\le\; \varepsilon\,; \end{aligned}$$

- Otherwise, if $S_{t-\tau_0}\big(x_{i,k}\big) \in \big[x_{i,N_i}, b_i\big]$ then

$$\begin{aligned} \left| S^\varepsilon_{t-\tau_0}\big(x_{i,k}\big) - S_{t-\tau_0}\big(x_{i,k}\big) \right| \;&=\; \left| S^\varepsilon_{t-\tau_0-\sum_{k'=k+1}^{N_i}\tau_{i,k'}}\big(x_{i,N_i}\big) - S_{t-\tau_0-\sum_{k'=k+1}^{N_i}\tau_{i,k'}}\big(x_{i,N_i}\big) \right| \\ &\le\; b_i - x_{i,N_i} \;<\; \varepsilon. \end{aligned}$$

In all cases we conclude

$$\left|S_t^\varepsilon(x_0) - S_t(x_0)\right| \;\leq\; (M+2)\varepsilon \quad \text{for all } t \geq 0\,. \tag{3.11}$$

CASE 2: $x_0 \in J_i^+$, where $i \in \mathcal{J}_\varepsilon^+$ and $J_i^+ = [a_i, b_i[$ is half-open. Choosing $\varepsilon < b_i - x_0$, we can obtain (3.11) in a similar way as in Case 1.

CASE 3: $x_0 \in J_i^-$, with $i \in \mathcal{J}_\varepsilon^-$. The estimates are analogous to the previous ones.

CASE 4: $x_0 \in \Omega_S^0 \bigcup \left(\bigcup_{i \notin \mathcal{J}_\varepsilon^+} I_i^+\right) \bigcup \left(\bigcup_{i \notin \mathcal{J}_\varepsilon^-} I_i^-\right)$. In this case, we have that

$$|S_t(x_0) - x_0| \;<\; \varepsilon \qquad\qquad \text{for all } t \geq 0. \tag{3.12}$$

If $x_0 = a_i$ for some $i \in \mathcal{J}_\varepsilon^+$ or $x_0 = d_i$ for some $i \in \mathcal{J}_\varepsilon^-$, then $S_t^\varepsilon(x_0) = x_0 = S_t(x_0)$ for any $t, \varepsilon > 0$. Otherwise, choosing $\varepsilon > 0$ such that

$$x_0 \notin \left[a_i - \frac{\varepsilon}{2}\,,\; a_i\right] \quad \text{for all } i \in \mathcal{J}_\varepsilon^+ \quad \text{and} \quad x_0 \notin \left[d_i\,,\; d_i + \frac{\varepsilon}{2}\right] \quad \text{for all } i \in \mathcal{J}_\varepsilon^-,$$

we obtain $S_t^\varepsilon(x_0) = x_0$ for all $t \geq 0$. Hence (3.12) yields (3.10). □

4 Markov semigroups with a simpler dynamics

According to Theorem 2.3, a Markov semigroup compatible with the ODE (1.1) is uniquely determined by the positive, atomless measure μ supported on $f^{-1}(0)$, the countable sets $\mathcal{S}$, $\mathcal{S}^*$, and the maps $\Lambda : \mathcal{S}^* \mapsto \mathbb{R}_+$, $\Theta : \Omega^* \mapsto [0,1]$, following the construction in Section 2.

The next lemma shows that this random dynamics can be approximated with a simpler one, where the function f is piecewise constant, the measure μ vanishes, and the sets $\mathcal{S}, \mathcal{S}^*, \Omega^*$ contain finitely many points. Here and throughout the following, on the set of Lipschitz functions $\varphi : \mathbb{R} \mapsto \mathbb{R}$, we use the norm

$$\|\varphi\|_{Lip} \;=\; \|\varphi\|_{W^{1,\infty}} \;=\; \max\left\{\sup_x |\varphi(x)|\;,\; \sup_{x \neq y} \frac{|\varphi(x) - \varphi(y)|}{|x-y|}\right\}. \tag{4.1}$$

Lemma 4.1 *Let f satisfy the assumptions* **(A1)-(A2)**, *and let $P_t = P_t(x_0, A)$ be the transition kernels for a Markov semigroup, compatible with (1.1). Then, for any given $T, R > 0$, there exists a sequence of transition kernels $P_t^n = P_t^n(x_0, A)$, such that*

$$\lim_{n\to\infty} \sup_{t\in[0,T], |x|\leq R} \sup_{\|\varphi\|_{Lip}\leq 1} \left|\int \varphi(x) P_t^n(x_0, dx) - \int \varphi(x) P_t(x_0, dx)\right| \;=\; 0. \tag{4.2}$$

For each $n \geq 1$, the transition kernels $P_t^n(\cdot,\cdot)$ define a Markov semigroup compatible with an ODE

$$\dot{x} \;=\; f_n(x), \tag{4.3}$$

where f_n is piecewise constant. The functions f_n converge to f in the sense of the graph. Moreover, the corresponding sets $\mathcal{S}_n, \mathcal{S}_n^*, \Omega_n^*$ are finite and disjoint, while all the measures μ_n vanish.

Proof. For every fixed $N \geq 1$, consider the Markov process X^N obtained by stopping the motion as soon as it exits from the interval $[-N, N]$. This is achieved by inserting a stopping time τ_N:

$$X^N(t, x_0, \omega) \;=\; X\big(\tau_N, x_0, \omega\big),$$

where

$$\tau_N(t, x_0, \omega) \;\doteq\; \min\Big\{t \geq 0\,;\;\; \big|X(t, x_0, \omega)\big| \;\geq\; N\Big\}.$$

The corresponding transition kernels $P_t^N(x_0, \cdot)$ define a Markov semigroup compatible with the ODE $\dot{x} = f^N(x)$, where

$$f^N(x) \;=\; \begin{cases} f(x) & \text{if } \;|x| < N, \\ 0 & \text{if } \;|x| \geq N. \end{cases}$$

For every given $x_0 \in \mathbb{R}$ and $t \geq 0$, as soon as $N \geq Mt + |x_0|$ we have the identity

$$P_t^N(x_0, A) \;=\; P_t(x_0, A) \qquad \text{for all Borel set } A \subseteq \mathbb{R},$$

Each f_N has a compact support, and moreover the functions f_N converge to f in the sense of the graph.

By replacing f with f^N, it thus suffices to consider the case where f has compact support. The proof will be given in several steps.

1. We begin by identifying the set of points where the dynamics stops forever:

$$\Omega_X^0 \;=\; \big\{x \in \mathbb{R}\,;\;\; \text{Prob.}\{X(1, x, \omega) = x\} = 1\big\} \;\subseteq\; f^{-1}(0).$$

We also introduce:

Definition 4.1 *We say that a right-open interval $J^+ =]a, b[$ or $J^+ = [a, b[$ is an* **interval of increase** *if*

$$\lim_{t \to \infty} \textit{Prob.}\{X(t, x_1, \omega) > x_2\} \;=\; 1 \qquad \textit{for all } x_1, x_2 \in J^+ \;\; \text{with} \;\; x_1 < x_2.$$

Similarly, a left-open interval $J^- =]c, d[$ or $J^- =]c, d]$ is an **interval of decrease** *if*

$$\lim_{t \to \infty} \textit{Prob.}\{X(t, x_1, \omega) < x_2\} \;=\; 1 \qquad \textit{for all } x_1, x_2 \in J_k^- \;\; \text{with} \;\; x_1 > x_2.$$

If J_1^+, J_2^+ are two intervals of increase, with $J_1^+ \cap J_2^+ \neq \emptyset$, then the union $J_1^+ \cup J_2^+$ is also an interval of increase. The same of course is true for intervals of decrease. We can thus identify countably many maximal intervals of increase $J_k^+ =]a_k, b_k[$ or $[a_k, b_k[$, $k \in \mathcal{J}^+$, and countably many maximal intervals of decrease $J_k^- =]c_k, d_k[$ or $]c_k, d_k]$, $k \in \mathcal{J}^-$.

As in Theorem 2.3, the set of points where the dynamics stops for a random waiting time is

$$\mathcal{S}^* \;=\; \big\{x \in \mathbb{R}\,;\;\; 0 < \text{Prob.}\{X(1, x, \omega) = x\} < 1\big\}.$$

We also consider the countable set of stationary points

$$\mathcal{S} \;=\; \Big(\{a_k, b_k\,;\;\; k \in \mathcal{J}^+\} \cup \{c_k, d_k\,;\;\; k \in \mathcal{J}^-\}\Big) \cap \Omega_X^0\,.$$

2. Next, given $\varepsilon > 0$, among all the maximal intervals of increase or decrease, we select a finite number, chosen as follows:

(i) All maximal intervals whose length satisfies $b_k - a_k > \varepsilon$, or $d_k - c_k > \varepsilon$.

(ii) All maximal intervals to the left $]z_k^-, z_k[$ or to the right $]z_k, z_k^+[$ of a point $z_k \in \Omega^*$, from where both a decreasing and an increasing trajectory can initiate, whose length satisfies $z_k^+ - z_k^- > \varepsilon$.

It is clear that the above rules select a finite set of intervals J_k, $k = 1, \ldots, N$. We now define a new dynamics, setting

$$\begin{cases} f_\varepsilon(x) &= 0 \qquad \text{for all } x \notin \Omega^* \cup J_1 \cup \cdots \cup J_N\,, \\ X^\varepsilon(t, x_0, \omega) &= x_0 \qquad \text{for all } t \geq 0, \quad x_0 \notin \Omega^* \cup J_1 \cup \cdots \cup J_N\,. \end{cases} \tag{4.4}$$

By construction, for each random trajectory we have

$$\big|X^\varepsilon(t, x_0, \omega) - X(t, x_0, \omega)\big| \leq \varepsilon \tag{4.5}$$

for all t, x_0, ω. Hence, calling P_t^ε the corresponding transition probability kernels, for every given $(t, x_0) \in [0, \infty[\times\mathbb{R}$ and every continuous function φ, we have

$$\begin{aligned} \Big|E[\varphi(X^\varepsilon(t, x_0, \omega))] - E[\varphi(X(t, x_0, \omega))]\Big| &\leq \int_{\mathcal{W}} |\varphi(X^\varepsilon(t, x_0, \omega)) - \varphi(X(t, x_0, \omega))|\, d\mathbb{P}(\omega) \\ \leq \sup\Big\{|\varphi(s_1) - \varphi(s_2)|\,;\ \ s_1, s_2 \in [x_0 - tM,\ x_0 + tM],\ |s_1 - s_2| \leq \varepsilon\Big\}. \end{aligned} \tag{4.6}$$

Here M is the upper bound on $|f|$, introduced in the assumption **(A1)**. Since φ is uniformly continuous on bounded intervals, as $\varepsilon \to 0$ the right hand side of (4.6) approaches zero. This implies the convergence

$$\lim_{\varepsilon \to 0} \int \varphi(x) P_t^\varepsilon(x_0, dx) = \int \varphi(x) P_t(x_0, dx). \tag{4.7}$$

Thanks to the previous arguments, we can now assume that the dynamics is stationary outside a finite number of intervals of increase or decrease. As shown in Fig. 2, three cases must be considered:

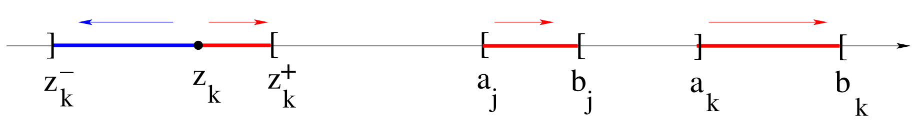

Figure 2: The three cases considered in the proof of Lemma 4.1.

Case 1: an open maximal interval of increase $I_k \doteq]a_k, b_k[$, of length $b_k - a_k > \varepsilon$. In this case, we consider the smaller interval $I_{k,\varepsilon} \doteq]a_k + \varepsilon/2,\ b_k - \varepsilon/2[$.

Case 2: a half-open maximal interval of increase $I_j \doteq [a_j, b_j[$, of length $b_j - a_j > \varepsilon$. In this case, we consider the smaller interval $I_{j,\varepsilon} \doteq [a_j,\ b_j - \varepsilon/2[$.

Case 3: a point $z_k \in \Omega^*$, from where both increasing and decreasing trajectories can originate, with probability $\theta \in [0, 1]$ and $1 - \theta$, respectively. We then call $]z_k^-, z_k[$ and $]z_k, z_k^+[$ the corresponding maximal intervals of decrease and of increase, to the left and to the right of z_k such that $z_k^+ - z_k^- > \varepsilon$. In this case, we set $I_k \doteq]z_k^-, z_k^+[$ and consider the smaller interval $I_{k,\varepsilon} \doteq]z_k^- + \varepsilon(z_k - z_k^-),\ z_k^+ - \varepsilon(z_k^+ - z_k)[$.

3. In all three cases, we modify the dynamics by setting

$$f_\varepsilon(x) \;=\; 0 \qquad \text{if} \quad x \in I_k \setminus I_{k,\varepsilon}\,. \tag{4.8}$$

In other words, the dynamics is stopped outside $I_{k,\varepsilon}$. This implies

$$X^\varepsilon(t, x_0, \omega) \;=\; x_0 \quad \text{for all} \;\; x_0 \in I_k \setminus I_{k,\varepsilon}\,. \tag{4.9}$$

Moreover, for $x_0 \in I_{k,\varepsilon}$, in Cases 1 and 2 we have

$$X^\varepsilon(t, x_0, \omega) \;=\; \min\bigl\{X(t, x_0, \omega),\, b_k - \varepsilon/2\bigr\}\,,$$

while in Case 3:

$$X^\varepsilon(t, x_0, \omega) \;=\; \begin{cases} X(t, x_0, \omega) & \text{if} \quad X(t, x_0, \omega) \in [z_k^- + \varepsilon(z_k - z_k^-)\,,\; z_k^+ - \varepsilon(z_k^+ - z_k)]\,, \\ z_k^- + \varepsilon(z_k - z_k^-) & \text{if} \quad X(t, x_0, \omega) \;\leq\; z_k^- + \varepsilon(z_k - z_k^-), \\ z_k^+ - \varepsilon(z_k^+ - z_k) & \text{if} \quad X(t, x_0, \omega) \;\geq\; z_k^+ - \varepsilon(z_k^+ - z_k). \end{cases}$$

Since the new dynamics is stationary on each $I_k \setminus I_{k,\varepsilon}$, the set $\mathcal{S}^*$ of points where a random waiting time occurs, as well as the measure μ, can now be replaced by $\mathcal{S}^*_\varepsilon$ and μ_ε respectively. Here

$$\mathcal{S}^*_\varepsilon \;\doteq\; \mathcal{S}^* \cap \Bigl(\bigcup_k I_{k,\varepsilon}\Bigr), \qquad\qquad \mu_\varepsilon(A) \;\doteq\; \mu\left(A \cap \Bigl(\bigcup_k I_{k,\varepsilon}\Bigr)\right), \tag{4.10}$$

for every Borel set A.

We observe that, for a fixed x_0, as $\varepsilon \to 0$ we have

$$\lim_{\varepsilon\to 0}\; \sup_{t,\omega} \bigl|X^\varepsilon(t, x_0, \omega) - X(t, x_0, \omega)\bigr| \;\leq\; \varepsilon. \tag{4.11}$$

Therefore, the same argument used at (4.6) implies the convergence (4.7).

4. For a fixed $\varepsilon > 0$, consider the approximate dynamics constructed in the previous step. We claim that the measure μ^ε in (4.10) has finite total mass.

Indeed, consider an interval $I_k =]a_k, b_k[$, as in Case 1. By assumption there are some random trajectories starting at $a_k + \varepsilon/2$ and reaching $b_k - \varepsilon/2$ in finite time. But, for every trajectory, the time needed to cross the whole interval $I_{k,\varepsilon}$ is $\geq \mu(I_{k,\varepsilon})$. Hence this value must be finite. The other two cases are entirely similar.

Since the intervals $I_{k,\varepsilon}$ are finitely many, this implies

$$\mu^\varepsilon(\mathbb{R}) \;=\; \sum_{k=1}^N \mu(I_{k,\varepsilon}) \;<\; +\infty.$$

5. Next, by a further approximation, we can modify the dynamics so that each interval I_k, $k = 1, \ldots, N$, contains at most finitely many points $y_j \in \mathcal{S}^*$ and thus the set $\mathcal{S}^*$ is finite. To justify this, let us consider a maximal interval of increase $I_k =]a_k, b_k[$, as in Case 1. By construction, there is a positive probability of moving from $a_k + \varepsilon$ to $b_k - \varepsilon$ in finite time. Hence, for some $t > 0$ one has

$$\delta \doteq \text{Prob.}\{X(t, a_k + \varepsilon, \omega) \leq b_k - \varepsilon\} < 1.$$

By the Markov property, we have

$$\text{Prob.}\{X(mt, a_k + \varepsilon, \omega) < b_k - \varepsilon\} \leq \delta^m.$$

This implies that the sum of all random waiting times $Y_j \in I_{k,\varepsilon}$ satisfies

$$\text{Prob.}\left\{ \sum_{y_j \in [a_k+\varepsilon, b_k-\varepsilon]} Y_j \geq mt \right\} \leq \delta^m \qquad \text{for all } k \geq 1.$$

In particular,

$$E\left[\sum_{y_j \in [a_k+\varepsilon, b_k-\varepsilon[} Y_j \right] = \sum_{y_j \in [a_k+\varepsilon, b_k-\varepsilon[} \frac{1}{\lambda_j} < +\infty.$$

Given an auxiliary constant $\delta > 0$, we can now

- remove all waiting times Y_j with

$$y_j \in [a_k, a_k + \varepsilon] \cup [b_k - \varepsilon, b_k],$$

- remove all waiting times Y_j, with $y_j \in I_{k,\varepsilon} =]a_k + \varepsilon,\, b_k - \varepsilon[$ and $j > N_k$, choosing N_k large enough so that

$$E\left[\sum_{y_j \in I_{k,\varepsilon},\ j > N_k} Y_j \right] = \sum_{y_j \in I_{k,\varepsilon},\ j > N_k} \frac{1}{\lambda_j} < \delta. \tag{4.12}$$

This will produce a new family of random trajectories $\widetilde{X}^\varepsilon(\cdot, x_0, \omega)$. Since we are removing some of the waiting times, within the interval $I_{k,\varepsilon}$ the new trajectories will be shifted forward, by an amount

$$0 \leq \widetilde{X}^\varepsilon(\cdot, x_0, \omega) - X^\varepsilon(\cdot, x_0, \omega) \leq M \cdot \left(\sum_{y_j \in I_{k,\varepsilon},\ j > N_k} Y_j(\omega) \right), \tag{4.13}$$

where M is the upper bound on f, introduced in **(A1)**.

For any given $\delta > 0$, we now set

$$\mathcal{W}_\delta \doteq \left\{ \omega \in \mathcal{W};\ \ M \cdot \left(\sum_{y_j \in I_{k,\varepsilon},\ j > N_k} Y_j(\omega) \right) \geq \sqrt{\delta} \right\}.$$

By Chebyshev's inequality, it follows

$$\mathbb{P}(\mathcal{W}_\delta) \leq \frac{M}{\sqrt{\delta}} \cdot E\left[\left(\sum_{y_j \in I_{k,\varepsilon},\ j > N_k} Y_j(\omega) \right) \right] \leq M\sqrt{\delta}.$$

Therefore, given any $x_0 \in I_{k,\varepsilon}$ and $t > 0$, and any bounded continuous function φ, using (4.13) we estimate

$$\begin{aligned}\Big|E[\varphi(\widetilde X^\varepsilon(t,x_0,\omega))] - E[\varphi(X(t,x_0,\omega))]\Big| &\leq \int_{\mathcal W}\Big|\varphi(\widetilde X^\varepsilon(t,x_0,\omega)) - \varphi(X^\varepsilon(t,x_0,\omega))\Big|\, d\mathbb{P}(\omega)\\ &\leq 2\|\varphi\|_\infty\cdot\mathbb{P}(\mathcal W_\delta) + \int_{\mathcal W\setminus\mathcal W_\delta}\Big|\varphi(\widetilde X^\varepsilon(t,x_0,\omega)) - \varphi(X^\varepsilon(t,x_0,\omega))\Big|\, d\mathbb{P}(\omega)\\ &\leq 2M\|\varphi\|_\infty\cdot\sqrt{\delta} + \sup_{s_1,s_2\in I_{k,\varepsilon},\ |s_1-s_2|\leq\sqrt{\delta}}\big|\varphi(s_1)-\varphi(s_2)\big|.\end{aligned} \tag{4.14}$$

By choosing $\delta > 0$ in (4.12) sufficiently small, the difference of the expectations in (4.14) can be rendered as small as we like.

The same ideas apply to intervals such as $I_k = [a_k, b_k[$ or $]z_k^-, z_k^+[$, described in Cases 2 and 3 respectively. Of course, intervals of decrease can be handled in the same way.

6. In general, the semigroup may contain an exponential waiting time also at points $z_k \in \Omega^*$ where trajectories can move both upward or downward, with given probabilities. However, we can construct an approximation so that the two sets $\mathcal{S}^*$ and Ω^* are disjoint.

Indeed, consider a point in the intersection: $z_k = y_j \in \Omega^* \cap \mathcal{S}^*$, for some indices k, j. Since $\mathcal{S}^*$ is finite, we can choose $\varepsilon > 0$ small enough so that the interval

$$[y_j', y_j''] \doteq [y_j - \varepsilon\,,\ y_j + \varepsilon]$$

does not contain any other point of S^*. We now redefine f at these two points, by setting $f(y_j') = f(y_j'') = 0$. Moreover, we construct a new family of transition kernels $P_t^\varepsilon(\cdot,\cdot)$ by removing the random waiting time Y_j at the point $y_j = z_k$, and inserting two waiting times Y_j', Y_j'' with the same exponential distribution as Y_j, at the two points y_j' and y_j''. In this case, for each random trajectory we have

$$\big|X^\varepsilon(t,z_k,\omega) - X(t,z_k,\omega)\big| \leq \max\big\{|y_j'-y_j|, |y_j'', -y_j|\big\} < \varepsilon$$

and

$$X^\varepsilon(t,x,\omega) = X(t,x,\omega) \qquad \text{for all } x \notin [y_j-\varepsilon,\ y_j+\varepsilon],$$

for every $t > 0$, $\omega \in \mathcal{W}$. Therefore, as $\varepsilon \to 0$, the same argument used at (4.6) implies the convergence in distribution.

Summarizing the previous analysis, we can approximate the original Markov semigroup with a new semigroup where:

(i) The dynamics is stationary outside finitely many intervals of increase or decrease.

(ii) The measure μ is finite.

(iii) The sets $\mathcal{S}, \mathcal{S}^*, \Omega^*$ are finite and disjoint.

7. To achieve the proof, we still need to approximate the dynamics with another Markov semigroup where f is piecewise constant, and μ vanishes. To fix ideas, we will show how to modify f on a maximal interval of increase J_j. The construction is entirely similar in the case of a maximal interval of decrease.

Given $\varepsilon > 0$ sufficiently small, two cases will be considered:

CASE 1: The interval $J_j =]a_j, b_j[$ is open.

By the same technique used in Section 3, for the proof of Theorem 1.1, (step **2**, CASE 1), we can approximate f by a function f_ε such that

(i) f_ε is piecewise constant and

$$\mathrm{Graph}(f_\varepsilon) \ \subseteq \ B(\mathrm{Graph}(f_\varepsilon), \varepsilon), \qquad f_\varepsilon(x) \ = \ f(x) \qquad \text{for all } x \in (\mathbb{R}\backslash J_j) \cup \mathcal{S}^*;$$

(ii) For every $[x_1, x_2] \subseteq \big[a_j + \varepsilon, b_j - \varepsilon\big]$, one has

$$\left| \int_{x_1}^{x_2} \frac{dy}{f(y)} + \mu([x_1, x_2]) - \int_{x_1}^{x_2} \frac{dy}{f_\varepsilon(y)} \right| \ \leq \ \varepsilon. \tag{4.15}$$

We then take μ_ε to be the zero measure.

In view of (i)–(ii), for every $x_0 \in \mathbb{R}\backslash J_j$ one has

$$X(t, x_0, \omega) \ = \ X^\varepsilon(t, x_0, \omega) \qquad \text{for all } (t, \omega) \in [0, \infty[\times \mathcal{W}. \tag{4.16}$$

On the other hand, for a fixed $x_0 \in J_j$, when $\varepsilon > 0$ is sufficiently small so that $x_0 \geq a_j + \varepsilon$, if $\min\{X(t, x_0, \omega), X^\varepsilon(t, x_0, \omega)\} \geq b_j - \varepsilon$, we have

$$\big|X^\varepsilon(t, x_0, \omega) - X(t, x_0, \omega)\big| \ \leq \ |(b_j - \varepsilon) - b_j| \ = \ \varepsilon.$$

Otherwise, without loss of generality, assume that

$$X(t, x_0, \omega) \ > \ X^\varepsilon(t, x_0, \omega) \qquad \text{and} \qquad X^\varepsilon(t, x_0, \omega) \ < \ b_j - \varepsilon.$$

Let $\tau^\varepsilon(\omega), \tau_1(\omega), \tau(\omega)$ be such that

$$S_{\tau(\omega)}(x_0) \ = \ X(t, x_0, \omega) \qquad \text{and} \qquad S_{\tau_1(\omega)}(x_0) \ = \ S^\varepsilon_{\tau^\varepsilon(\omega)}(x_0) \ = \ X^\varepsilon(t, x_0, \omega).$$

From (4.15), we have

$$|\tau_1(\omega) - \tau^\varepsilon(\omega)| \ = \ \left| \int_{x_0}^{S_{\tau_1(\omega)}(x_0)} \frac{dy}{f(y)} + \mu([x_0, S_{\tau_1(\omega)}(x_0)]) - \int_{x_0}^{S_{\tau_1(\omega)}(x_0)} \frac{dy}{f_\varepsilon(y)} \right| \ \leq \ \varepsilon.$$

Thus,

$$\begin{aligned} 0 \ < \ \tau(\omega) - \tau_1(\omega) \ &\leq \ \left(t - \sum_{y_j \in [x_0, S_{\tau(\omega)}(x_0)[} Y_j(\omega) \right) - \tau^\varepsilon(\omega) + \varepsilon \\ &\leq \ \left(t - \sum_{y_j \in \left[x_0, S^\varepsilon_{\tau^\varepsilon(\omega)}(x_0)\right]} Y_j(\omega) - \tau^\varepsilon(\omega) \right) + \varepsilon \ \leq \ \varepsilon \end{aligned}$$

and this implies

$$|X^\varepsilon(t,x_0,\omega) - X(t,x_0,\omega)| \;=\; |S_{\tau(\omega)}(x_0) - S_{\tau_1(\omega)}(x_0)| \;\leq\; M\varepsilon. \tag{4.17}$$

CASE 2: The interval $J_j = [a_j, b_j[$ is half open.

By the same technique use in Section 3 (step **2**, CASE 2), we can approximate f by a function f_ε such that

(i) f_ε is piecewise constant, and

$$\mathrm{Graph}(f_\varepsilon) \;\subseteq\; B(\mathrm{Graph}(f_\varepsilon), \varepsilon), \qquad f_\varepsilon(x) \;=\; f(x) \qquad \text{for all } x \in (\mathbb{R}\backslash[a_j - \varepsilon, b_j[)\cup\mathcal{S}^*;$$

(ii) For every $x_1 < x_2 \in [a_j, b_j - \varepsilon]$, it holds

$$\left|\int_{x_1}^{x_2} \frac{dy}{f(y)} + \mu([x_1,x_2]) - \int_{x_1}^{x_2} \frac{dy}{f_\varepsilon(y)}\right| \;\leq\; \varepsilon. \tag{4.18}$$

Again, we take μ_ε to be the zero measure.

For every $x_0 \in \mathbb{R}\backslash J_j$, the identity (4.16) again holds, provided that $\varepsilon > 0$ is sufficiently small so that $x_0 \notin [a_j - \varepsilon, b_j[$.

On the other hand, if $x_0 \in J_j$, the same argument used in the previous case yields (4.17).

In both cases, given any $x_0 \in \mathbb{R}$, for each random trajectory we have

$$\bigl|X^\varepsilon(t,x_0,\omega) - X(t,x_0,\omega)\bigr| \;\leq\; M\varepsilon \qquad \text{for all } t \geq 0, \omega \in \mathcal{W}$$

for $\varepsilon > 0$ sufficiently small. Therefore, the same argument used at (4.6) implies the convergence (4.7). □

5 An approximation lemma

The main goal of the next two sections is to prove Theorem 1.2, showing that every Markov semigroup whose trajectories are solutions to the ODE (1.1) can be approximated by a sequence of diffusion processes with smooth coefficients (1.12).

Thanks to Lemma 4.1, we can assume that the function f is piecewise constant, the sets $\mathcal{S}, \mathcal{S}^*, \Omega^*$ are finite, and the measure μ vanishes. To fix the ideas, let

$$x_0 \;<\; x_1 \;<\; x_2 \;<\; \cdots \;< x_N \tag{5.1}$$

be a list of all points in $\mathcal{S}, \mathcal{S}^*, \Omega^*$, together with all points where f has a jump. Consider the midpoints

$$y_1 \;<\; y_2 \;<\; \cdots \;< y_N\,, \qquad\qquad y_j = \frac{x_{j-1} + x_j}{2}\,. \tag{5.2}$$

For any random trajectory starting at a point $\bar{x}$, say $s \mapsto X(s,\bar{x},\omega)$ we define the stopping time

$$\tau(\bar{x},\omega) \;=\; \min\Bigl\{s \geq 0\,;\; X(s,\bar{x},\omega) = y_j \quad \text{for some } y_j \neq \bar{x}\Bigr\}. \tag{5.3}$$

In other words, if $y_{j-1} < \bar{x} < y_j$, then τ is the first time when the random trajectory hits either y_{j-1} or y_j. In the special case where $\bar{x} = y_k$, then τ is the first time when the trajectory hits either y_{k-1} or y_{k+1}.

Similarly, given a sequence of Markov processes X_n, $n \geq 1$, we define

$$\tau_n(\bar{x}, \omega) \;=\; \min\Big\{ s \geq 0\,;\; X_n(s, \bar{x}, \omega) = y_j \quad \text{for some } y_j \neq \bar{x} \Big\}. \tag{5.4}$$

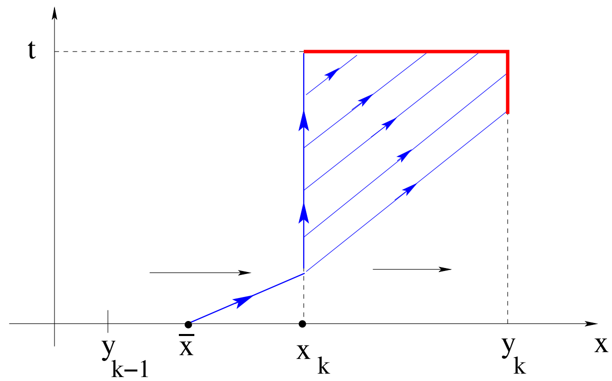

Figure 3: Random trajectories of the Markov semigroup, starting at $\bar{x}$. Here the speed $f(x)$ is constant for $x < x_k$ and $x > x_k$. At x_k there is a random waiting time. The thick red lines denote the support of the measure $\mu^{t,\bar{x}}$.

Throughout the following, we adopt the notation $a \wedge b \doteq \min\{a, b\}$. Using the above stopping times, given an initial point $\bar{x}$, for every $t > 0$ we consider the probability measures $\mu^{t,\bar{x}}$, $\mu_n^{t,\bar{x}}$ on $\mathbb{R}_+ \times \mathbb{R}$ defined as the push-forward of the maps

$$\omega \;\mapsto\; \big(\tau(\bar{x}, \omega) \wedge t,\; X(\tau(\bar{x}, \omega) \wedge t,\, \bar{x}, \omega)\big), \qquad \omega \;\mapsto\; \big(\tau_n(\bar{x}, \omega) \wedge t,\; X_n(\tau_n(\bar{x}, \omega) \wedge t,\, \bar{x}, \omega)\big).$$

More precisely, for every open set $V \subset \mathbb{R}_+ \times \mathbb{R}$, we define

$$\mu^{t,\bar{x}}(V) \;=\; \text{meas}\Big\{\omega\,;\;\; \big(\tau(\bar{x}, \omega) \wedge t,\; X(\tau(\bar{x}, \omega) \wedge t, \bar{x}, \omega)\big) \in V \Big\}, \tag{5.5}$$

$$\mu_n^{t,\bar{x}}(V) \;=\; \text{meas}\Big\{\omega\,;\;\; \big(\tau_n(\bar{x}, \omega) \wedge t,\; X_n(\tau_n(\bar{x}, \omega) \wedge t, \bar{x}, \omega)\big) \in V \Big\}. \tag{5.6}$$

As shown in Fig. 3, all measures $\mu^{t,\bar{x}}$ and $\mu_n^{t,\bar{x}}$ are supported within the set

$$\Sigma \;\doteq\; \{t\} \times \mathbb{R} \;\cup\; [0, t] \times \{y_1, y_2, \ldots, y_N\}. \tag{5.7}$$

The next lemma establishes a "local to global" result. If the random dynamics converge separately on each subinterval $[y_{j-1}, y_j]$ (i.e., after introducing the stopping times at each point y_j), then the weak convergence of the transition kernels (1.16) holds as well, for every initial point $\bar{x}$ and every $t > 0$.

Lemma 5.1 *Consider a Markov semigroup compatible with the ODE (1.1), where f is piecewise constant, the sets $\mathcal{S}, \mathcal{S}^*, \Omega^*$ are finite, and the measure μ vanishes. In addition, consider a sequence of diffusions processes $X^{(n)} = X^{(n)}(s, \bar{x}, \omega)$ as in (1.12), with $f_n \to f$ in the sense of the graph and $\sigma_n \downarrow 0$. Assume that, for every $\bar{x} \in \mathbb{R}$ and $t > 0$, we have the weak convergence of the corresponding mesures:*

$$\mu_n^{t,\bar{x}} \rightharpoonup \mu^{t,\bar{x}}, \tag{5.8}$$

defined as in (5.5)-(5.6). Then the weak convergence of the transition kernels (1.16) holds.

Proof. **1.** Given $(t,\bar{x})\in\mathbb{R}^+\times\mathbb{R}$, for every $p=1,2,3,\ldots$, define the p-th random stopping time by induction, in the obvious way. Namely, $\tau_1(\omega)$ is the first time when the trajectory $X(s,\omega)$ starting from $\bar{x}$ hits one of the points y_j. By induction, the p-th stopping time is

$$\tau_p(\omega) \;=\; \min\Big\{s>\tau_{p-1}(\omega)\,;\quad X(s,\omega)=y_j\quad\text{for some}\quad y_j\neq X(\tau_{p-1}(\omega),\omega)\Big\}.$$

The corresponding measures $\mu_p^{t,\bar{x}}$ are defined as in (5.5), replacing τ with τ_p, i.e.,

$$\mu_p^{t,\bar{x}}(V) \;=\; \text{meas}\Big\{\omega\,;\quad \big(\tau^p(\bar{x},\omega)\wedge t,\; X(\tau^p(\bar{x},\omega)\wedge t,\bar{x},\omega)\big)\in V\Big\}.$$

The same construction can of course be repeated for each of the Markov processes X_n, thus defining a sequence of measures $\mu_{p,n}^{t,\bar{x}}$, $n\geq 1$.

2. We observe that, for a sequence of probability measures μ_n on the real line, the weak convergence $\mu_n\rightharpoonup\mu$ is equivalent to the convergence of the distribution functions $F_n\to F$ in $\mathbf{L}^1_{loc}(\mathbb{R})$, where

$$F_n(x) \;=\; \mu_n\big(\,]-\infty,x]\big),\qquad F(x) \;=\; \mu\big(\,]-\infty,x]\big)\qquad\text{for all } x\in\mathbb{R}.$$

In the case we are presently considering, since all of measures $\mu_{p,n}^{t,\bar{x}}$ are supported on the 1-dimensional set Σ at (5.7), the weak convergence

$$\mu_{p,n}^{t,\bar{x}} \;\rightharpoonup\; \mu_p^{t,\bar{x}} \tag{5.9}$$

is equivalent to the $\mathbf{L}^1$ convergence of a finite family of distribution functions.

More precisely, for every $j=1,\ldots,N$, consider the nondecreasing functions

$$F_{p,n,j}^{t,\bar{x}}(s) \;\doteq\; \mu_{p,n}^{t,\bar{x}}\Big([0,s]\times\{y_j\}\Big),\qquad F_{p,j}^{t,\bar{x}}(s) \;\doteq\; \mu_p^{t,\bar{x}}\Big([0,s]\times\{y_j\}\Big),\qquad s\in[0,t]. \tag{5.10}$$

Moreover, for $y\in\mathbb{R}$, consider the nondecreasing functions

$$\begin{aligned} G_{p,n}^{t,\bar{x}}(y) \;&\doteq\; \mu_{p,n}^{t,\bar{x}}\Big(\{t\}\times]-\infty,y]\Big)+\sum_{y_j\leq y}\mu_{p,n}^{t,\bar{x}}\Big([0,t]\times\{y_j\}\Big),\\ G_p^{t,\bar{x}}(y) \;&\doteq\; \mu_p^{t,\bar{x}}\Big(\{t\}\times]-\infty,y]\Big)+\sum_{y_j\leq y}\mu_p^{t,\bar{x}}\Big([0,t]\times\{y_j\}\Big). \end{aligned} \tag{5.11}$$

Then the weak convergence (5.9) is then equivalent to the convergence

$$F_{p,n,j}^{t,\bar{x}}\to F_{p,j}^{t,\bar{x}}\quad\text{in}\quad \mathbf{L}^1([0,t])\qquad\text{and}\qquad G_{p,n}^{t,\bar{x}}\to G_p^{t,\bar{x}}\quad\text{in}\quad \mathbf{L}^1_{loc}(\mathbb{R}). \tag{5.12}$$

3. By induction, assuming the weak convergence

$$\mu_{p-1,n}^{t,\bar{x}} \;\rightharpoonup\; \mu_{p-1}^{t,\bar{x}}\,, \tag{5.13}$$

we need to prove the weak convergence (5.9).

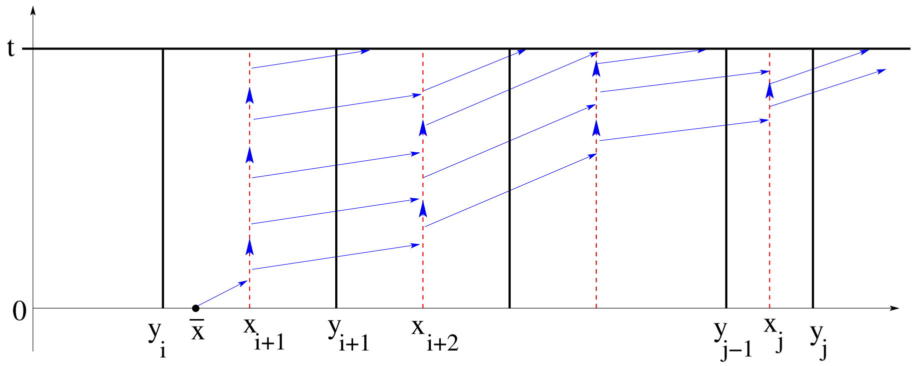

Figure 4: The thick lines describe the set Σ in (5.7). In this picture, trajectories of the Markov semigroup start at $\bar x$ and move with positive speed, with random waiting times at the points $x_{i+1}, x_{i+2}, \ldots$ If $j = i+p$, the p-th stopping time occurs when a trajectory reaches the point y_j, with $j = i+p$.

To fix ideas, assume that all trajectories of the Markov semigroup starting from $\bar x$ are non-decreasing. Moreover, assume $y_i \le \bar x < y_{i+1}$. Setting $j = i+p$, this implies

$$\begin{aligned} Supp\big(\mu_{p-1}^{t,\bar x}\big) &\subseteq \Sigma_{j-1} &\doteq \{t\}\times\mathbb{R} \,\cup\, [0,t]\times\{y_{j-1}\}\,, \\ Supp\big(\mu_{p}^{t,\bar x}\big) &\subseteq \Sigma_{j} &\doteq \{t\}\times\mathbb{R} \,\cup\, [0,t]\times\{y_{j}\}\,. \end{aligned} \tag{5.14}$$

We start by proving the convergence $F_{p,n,j} \to F_{p,j}$ in $\mathbf{L}^1\big([0,t]\big)$. Let $\varepsilon > 0$ be given. By the inductive assumption, we have the convergence $F_{p-1,n,j-1} \to F_{p-1,j-1}$ in $\mathbf{L}^1\big([0,t]\big)$.

Call

$$\begin{aligned} \Gamma_n(\tau) &\doteq \mu_{1,n}^{t,y_{j-1}}\big([0,\tau]\times\{y_j\}\big), \\ \Gamma(\tau) &\doteq \mu_{1}^{t,y_{j-1}}\big([0,\tau]\times\{y_j\}\big). \end{aligned} \tag{5.15}$$

the probability that a random trajectory starting at y_{j-1} reaches y_j within time τ (without ever touching y_{j-2}, in the case of a diffusion). The function $F_{j,n}$ can now be computed by the Stieltjes integral

$$F_{j,n}(\tau) \;=\; \int_0^\tau \Gamma_n(\tau-\zeta)\, dF_{j-1,n}(\zeta). \tag{5.16}$$

We observe that all functions $\Gamma_n, \Gamma, F_{j-1,n}F_{j-1}$ are non-decreasing with values in $[0,1]$. Moreover

$$\Gamma_n(0) \;=\; \Gamma(0) \;=\; F_{j-1,n}(0) \;=\; F_{j-1}(0) \;=\; 0.$$

By the convergence $\Gamma_n \to \Gamma$ and $F_{j-1,n} \to F_{j-1}$ in $\mathbf{L}^1\big([0,t]\big)$ it thus follows the convergence of the integral functions in (5.16), namely $F_{j,n} \to F_j$ in $\mathbf{L}^1\big([0,t]\big)$.

4. It remains to prove the convergence $G_{p,n}^{t,\bar x} \to G_p^{t,\bar x}$ in $\mathbf{L}^1_{loc}(\mathbb{R})$.

As a preliminary, we remark that the diffusion semigroups some trajectories can move backward, from some y_k to y_{k-1}. However, the probability that this happens goes to zero as $n \to \infty$.

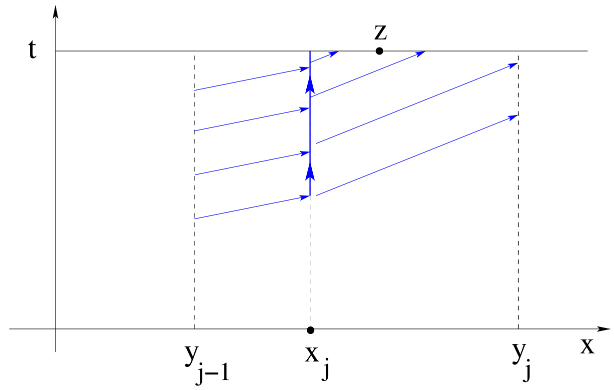

Figure 5: Starting with the distribution $\mu^{t,\bar{x}}_{n,p-1}$ restricted to the vertical segment $[0,t]\times\{y_{j-1}\}$, by the formula (5.20) one can compute the probability that trajectories of the semigroups reach a point $y\le z$ at time t.

Next, we observe that our previous assumption $y_i \le \bar{x} < y_{i+1}$ implies the identities

$$\begin{aligned} G^{t,\bar{x}}_p(y) &= G^{t,\bar{x}}_{p-1}(y) && \text{for} \quad y\le y_{j-1}\,, \\ G^{t,\bar{x}}_{p,n}(y) &= G^{t,\bar{x}}_p(y) = 1 && \text{for} \quad y\ge y_j\,, \end{aligned} \tag{5.17}$$

where $j=i+p$.

To prove the convergence in $\mathbf{L}^1\big([y_{j-1},y_j]\big)$, for any point $z\in\,]y_{j-1},y_j[$ and any $\varepsilon>0$, we will show that

$$\limsup_{n\to\infty}\ G^{t,\bar{x}}_{p,n}(z-\varepsilon)-\varepsilon\ \le\ G^{t,\bar{x}}_p(z)\ \le\ \liminf_{n\to\infty}\ G^{t,\bar{x}}_{p,n}(z+\varepsilon)+\varepsilon. \tag{5.18}$$

In analogy with (5.15), we now set

$$\begin{aligned} \Lambda_n(\tau) &\doteq \mu^{\tau,y_{j-1}}_{1,n}\big([0,\tau]\times[z,y_j]\big), \\ \Lambda(\tau) &\doteq \mu^{\tau,y_{j-1}}_{1}\big([0,\tau]\times[z,y_j]\big). \end{aligned} \tag{5.19}$$

In other words, $\Lambda_n(\tau)$ is the probability that a trajectory, starting at y_{j-1}, either stops at y_j or reaches a point $\ge z$ at time τ.

We now have

$$G^{t,\bar{x}}_{p,n}(z)\ =\ 1-\int_0^t \Lambda_n(t-\zeta)\,dF_{j-1,n}(\zeta), \tag{5.20}$$

and an entirely similar formula holds for $G^{t,\bar{x}}_p(z)$. For any $\varepsilon>0$, the convergence $F_{j-1,n}(\zeta)\to F_{j-1}(\zeta)$ and $\Lambda_n\to\Lambda$ implies (5.18). This proves the desired $\mathbf{L}^1$ convergence $G^{t,\bar{x}}_{p,n}\to G^{t,\bar{x}}_p$ in $\mathbf{L}^1\big([y_{j-1},y_j]\big)$.

5. Finally we observe that, for a fixed $t>0$, when p is sufficiently large one has $\tau^p(\bar{x},\omega)>t$ with probability 1, for all $\bar{x}\in\mathbb{R}$. Indeed, since trajectories of the semigroup are monotone, taking $p=N$ would suffice. Hence $\mu^{t,\bar{x}}_N$ is supported on $\{t\}\times\mathbb{R}$, and the weak convergence $\mu^{t,\bar{x}}_{N,n}\rightharpoonup\mu^{t,\bar{x}}_N$ implies the weak convergence (1.16). □

6 Approximating a Markov semigroup with smooth diffusions

In this section we complete the proof of Theorem 1.2. This will require several steps.

1. It will suffice to construct diffusions of the form

$$dX_t \;=\; g_n(X_t)\,dt + \sigma_n\,dW_t \tag{6.1}$$

where g_n is piecewise constant. Indeed, when this is done, we can then take a mollification: $f_n = \phi_\delta * g_n$, where the mollifier satisfies

$$\phi_\delta \in \mathcal{C}^\infty_c\,, \qquad \mathrm{Supp}(\phi_\delta) = [-\delta, \delta], \qquad\qquad \delta << \sigma_n\,.$$

Thanks to Lemma 4.1, without loss of generality we can assume that the function f is piecewise constant with compact support, the sets $\mathcal{S}, S^*, \Omega^*$ are finite and disjoint, and μ is the zero measure. As in (5.1), we call $x_0 < \cdots < x_N$ the points in $\mathcal{S} \cup S^* \cup \Omega^*$, together with all points where f has a jump.

2. Thanks to Lemma 5.1, it suffices to show that, after inserting the stopping times at all points y_j in (5.2), the corresponding transition kernels converge, for every initial point $\bar{x}$ and every $t > 0$. We are thus left with the task of constructing a sequence of diffusion approximations (6.1) on an interval $J = [y_{j-1}, y_j]$ where the function f is piecewise constant, with a single jump at the interior point x_j. In the easy case where $x_j \notin \mathcal{S} \cup S^* \cup \Omega^*$, it suffices to choose $g_n(x) = f(x)$ for every $x \in J$ and $n \geq 1$. The three main remaining cases will be discussed in the following steps.

3. CASE 1: $x_j \in \mathcal{S}$. Without loss of generality, we assume $x_j = 0$. Also, to fix ideas, assume that the restriction of f to the interval J is given by (1.19), with $a, b > 0$. See Fig. 1, left. Other cases can be treated similarly.

To make sure that the trajectories of the vanishing diffusion limit stop at the origin, for $x \in J$ we then define

$$g_n(x) \;=\; \begin{cases} a & \text{if} \quad x < -\sqrt{\sigma_n} \\ 0 & \text{if} \quad |x| \leq \sqrt{\sigma_n}\,, \\ b & \text{if} \quad x > \sqrt{\sigma_n}\,. \end{cases} \tag{6.2}$$

To prove that this sequence of diffusions achieves the desired limit (1.16), let τ_n be the first time when a random trajectory of (6.1) starting at the origin reaches one of the two points $\pm\sqrt{\sigma_n}$. For every given $t > 0$, one has

$$\begin{aligned} \mathrm{Prob.}\big\{\tau_n \leq t\big\} \;&\leq\; \mathrm{Prob.}\left(\sup_{s\in[0,t]} |W(s)| \;\geq\; \frac{1}{\sqrt{\sigma_n}}\right) \\ &\leq\; 4\cdot\left(1 - \frac{1}{\sqrt{2\pi t}}\int_0^{\frac{1}{\sqrt{\sigma_n}}} e^{\frac{-x^2}{2t}}\,dx\right) \;\xrightarrow{\sigma_n\to 0}\; 0. \end{aligned} \tag{6.3}$$

In the case where the motion starts at $x_0 = 0$, this yields

$$\lim_{n\to\infty} P_t^{(n)}\Big(0\,;\;[-\sqrt{\sigma_n}\,,\;\sqrt{\sigma_n}]\Big) \;=\; 1.$$

Therefore (1.16) holds for $x_0 = 0$.

Next, assume $x_0 \in J$ but $x_0 > 0$. In this case, the result can be achieved by standard results in large deviation theory. Since we are assuming $f(x) = b > 0$ for $x > 0$, on the interval $[\sqrt{\sigma}_n, y_j]$ the diffusion process takes the form

$$dX_t = b\,dt + \sigma_n\,dW_t\,, \qquad X_0 = x_0\,. \tag{6.4}$$

For any $\delta \in]0, x_0]$ and $T > 0$, as $\sigma_n \to 0$ the probability that a random trajectory $X_t(\omega)$ starting at x_0 reaches a distance $> \delta$ from the limit solution $x(t) = x_0 + bt$ satisfies

$$\text{Prob.}\Big\{|X_t(\omega) - x(t)| \ > \delta \quad \text{for some } t \in [0,T]\Big\} \ \xrightarrow{\sigma_n \to 0} \ 0.$$

This already implies (1.16).

Finally, consider the case where $x_0 < 0$. For any given $0 < \delta < |x_0|$, consider the time

$$T = \frac{-\delta - x_0}{a}\,,$$

As before, the probability that a random trajectory $X_t(\omega)$ starting at x_0 reaches a distance $> \delta$ from the limit solution $x(t) = x_0 + at$ satisfies

$$\text{Prob.}\Big\{|X_t(\omega) - x(t)| \ > \delta \quad \text{for some } t \in [0,T]\Big\} \ \xrightarrow{\sigma_n \to 0} \ 0.$$

Next, for a random trajectory such that

$$\big|X_T(\omega) - x(T)\big| = \big|X_T(\omega) + \delta\big| < \delta,$$

let $\tau_n > T$ be the first time where $X_t(\omega)$ touches a point in the set $\{-3\delta, \delta\}$. As soon as $\sqrt{\sigma_n} < \delta$, the same argument used at (6.3) yields

$$\text{Prob.}\big\{\tau_n \leq T + t\big\} \ \xrightarrow{\sigma_n \to 0} \ 0. \tag{6.5}$$

Since $\delta > 0$ was arbitrary, this again yields the weak convergence (1.16).

4. CASE 2: $x_j \in \Omega^*$ is a point from which both an increasing and a decreasing trajectory initiate. Again, we assume that $x_j = 0$, and that the restriction of f to the interval J is given by (1.19), with $a < 0 < b$. Moreover, let $\theta \in [0,1]$ be the probability that a random trajectory starting at the origin moves to the right. For each given $\sigma_n > 0$, we consider the diffusion process (6.1), where

$$g_n(x) = \begin{cases} a & \text{if } x < \xi_n\,, \\ b & \text{if } x > \xi_n\,, \end{cases} \tag{6.6}$$

for a suitable point ξ_n, with $\xi_n \to 0$ as $n \to \infty$.

To achieve the weak convergence (1.16) for the initial point $x_0 = 0$, the points ξ_n must be carefully chosen. Toward this goal, consider again the piecewise constant function f in (1.19). Performing the change of variables $y = x/\sqrt{\sigma_n}$, the diffusion process

$$dX_t = f(X_t)\,dt + \sigma_n\,dW_t \tag{6.7}$$

becomes

$$d\widetilde{X}_t \;=\; \widetilde{f}_n(\widetilde{X}_t)dt + \sqrt{\sigma_n}\,dW_t\,, \tag{6.8}$$

where the new drift is

$$\widetilde{f}_n(y) \;=\; \begin{cases} a/\sqrt{\sigma_n} & \text{if} \quad y < 0, \\ b/\sqrt{\sigma_n} & \text{if} \quad y > 0. \end{cases} \tag{6.9}$$

Let $\widetilde{\tau}_n^y$ be the first time when a random trajectory of (6.8) starting from a point $y \in [-1, 1]$ hits one of the two points in the set $\{-1, 1\}$. Consider the expected values

$$\begin{cases} \widetilde{v}_n(y) \;\doteq\; E\left(\widetilde{\tau}_n^y\right) \\ \widetilde{u}_n(y) \;\doteq\; E\left(\widetilde{X}_{\widetilde{\tau}_n^y}\right). \end{cases} \tag{6.10}$$

It is well known that the functions $\widetilde{v}_n, \widetilde{u}_n$ provide the unique solutions to the boundary value problems

$$\begin{cases} \widetilde{f}_n(y)v_y + \dfrac{\sigma_n}{2}v_{yy} \;=\; -1 & \text{if} \;\; y \in]-1, 1[\,, \\ v \;=\; 0 & \text{if} \;\; y \in \{-1, 1\}, \end{cases} \tag{6.11}$$

$$\begin{cases} \widetilde{f}_n(y)u_y + \dfrac{\sigma_n}{2}u_{yy} \;=\; 0 & \text{if} \;\; y \in]-1, 1[\,, \\ u(y) \;=\; y & \text{if} \;\; y \in \{-1, 1\}. \end{cases} \tag{6.12}$$

For a proof, see for example Chapter 9 in [12].

The solutions of (6.11) and (6.12) have the form

$$\widetilde{v}_n(y) \;=\; \begin{cases} -\dfrac{(y+1)\cdot\sqrt{\sigma_n}}{a} + c_{1,n}\cdot\left(e^{-a_n y} - e^{a_n}\right) & \text{if} \quad -1 < y \le 0, \\ \dfrac{(1-y)\cdot\sqrt{\sigma_n}}{b} + c_{2,n}\cdot\left(e^{-b_n y} - e^{-b_n}\right) & \text{if} \quad 0 \le y < 1, \end{cases} \tag{6.13}$$

$$\widetilde{u}_n(y) \;=\; \begin{cases} -1 + 2\cdot\dfrac{b\left(e^{a_n} - e^{-a_n y}\right)}{b(e^{a_n}-1) + a(1-e^{-b_n})} & \text{if} \quad -1 < y \le 0, \\ 1 + 2\cdot\dfrac{a\left(e^{-b_n} - e^{-b_n y}\right)}{b(e^{a_n}-1) + a(1-e^{-b_n})} & \text{if} \quad 0 \le y < 1, \end{cases} \tag{6.14}$$

and where the constants $a_n, b_n, c_{1,n}, c_{2,n}$ satisfy

$$a_n \;\doteq\; \frac{2a}{\sigma_n^{3/2}}, \qquad b_n \;\doteq\; \frac{2b}{\sigma_n^{3/2}} \qquad \text{and} \qquad |c_{1,n}|, |c_{2,n}| \;\le\; \frac{C}{\sqrt{\sigma_n}}. \tag{6.15}$$

We observe that $\widetilde{u}_n$ is strictly increasing in $[-1, 1]$. Moreover, as $n \to +\infty$ we have the limits

$$\sup\left\{\widetilde{u}_n(x)\,;\; x \in \left[-1, -\frac{1}{3}\sqrt{\sigma_n}\right]\right\} \;\to\; 0, \qquad \inf\left\{\widetilde{u}_n(x)\,;\; x \in \left[\frac{1}{3}\sqrt{\sigma_n}\,,\; 1\right]\right\} \;\to\; 1. \tag{6.16}$$

Therefore, for any given $0 < \theta < 1$ and every $n \ge 1$ large enough, we can uniquely choose an intermediate point $\zeta_n \in \left[-\sqrt{\sigma_n}, \sqrt{\sigma_n}\right]$ such that

$$\widetilde{u}_n(\zeta_n) \;=\; 2\theta - 1. \tag{6.17}$$

Moreover, from (6.15) and (6.13) it follows

$$\lim_{n\to\infty} \widetilde{v}_n(x) = 0 \qquad \text{for all } x \in [-1,1]. \tag{6.18}$$

Indeed, since the drifts in (6.8)-(6.9) become very large, the average time needed for a random trajectory to exit from the interval $[-1,1]$ approaches zero.

Going back to the original space variable x, we now set $\xi_n = \sqrt{\sigma_n}\,\zeta_n$. According to the previous analysis, for a random trajectory of (6.1), (6.6), starting at the origin, the following holds.

(i) The random time $\tau_n(\omega)$ at which this trajectory reaches one of the points $\xi_n \pm \sqrt{\sigma_n}$ approaches zero as $\sigma_n \downarrow 0$.

(ii) The probabilities of reaching the right or the left point are given by

$$\begin{aligned} \text{Prob.}\Big\{X_{\tau_n}(\omega) = \xi_n + \sqrt{\sigma_n}\Big\} &= \theta, \\ \text{Prob.}\Big\{X_{\tau_n}(\omega) = \xi_n - \sqrt{\sigma_n}\Big\} &= 1-\theta. \end{aligned}$$

Thanks to (6.16), for all n large enough we can assume $|\xi_n| \le \frac{1}{3}\sqrt{\sigma_n}$. A computation entirely similar to (6.3) now shows that, as soon as a random trajectory has reached one of the points $\xi_n \pm \sqrt{\sigma_n}$, the probability that it goes back to the point ξ_n, approaches zero as $\sigma_n \downarrow 0$.

To achieve the proof of weak convergence of the transition kernels starting at the origin, let $\delta > 0$ be given. Consider any random trajectory $X_t(\omega)$ of (6.6) that starts at the point $\xi_n + \sqrt{\sigma_n}$, and then never touches the point ξ_n. As an easy consequence of the theory of large deviations, for any $T > 0$ we have

$$\text{Prob.}\Big\{\sup_{t\in[0,T]} \big|X_t(\omega) - \xi_n + \sqrt{\sigma_n} + bt\big| > \delta \,\Big|\, X_0(\omega) = \xi_n + \sqrt{\sigma_n}\Big\} \xrightarrow{\sigma_n\to 0} 0.$$

Combining this fact with the above properties (i)-(ii), the weak convergence of the transitions kernels starting from $x_0 = 0$ is proved.

For every other initial point $x_0 \in [y_{j-1}, y_j] \setminus \{x_0\}$, the weak convergence of the transition kernels is trivial.

4. CASE 3: $x_j \in \mathcal{S}^*$ is a point where trajectories of the Markov semigroup stop for a random time $T(\omega) \ge 0$ with Poisson distribution

$$\text{Prob.}\{T(\omega) > s\} = e^{-\lambda s},$$

and then start moving again. As usual, we assume that $x_j = 0$, while the drift f is piecewise constant, as in (1.19). To fix ideas, let $a, b > 0$. The case $a, b < 0$ is entirely similar.

We claim that this process can be approximated by a sequence of diffusions as in (6.1), where each g_n has the form

$$g_n(x) = \begin{cases} a & \text{if } x < 0, \\ -\eta_n & \text{if } x \in [0, \varepsilon_n], \\ b & \text{if } x > \varepsilon_n, \end{cases} \tag{6.19}$$

for a suitable choice of the constants ε_n, η_n, with

$$0 \;<\; \varepsilon_n \;<<\; \sigma_n \;<<\; \eta_n. \tag{6.20}$$

Introducing the rescaled space variable $y = \dfrac{x}{\varepsilon_n}$, the equation (6.1) becomes

$$dY_t \;=\; \tilde g_n(Y_t)dt + \tilde\sigma_n dW_t, \tag{6.21}$$

with

$$\tilde g_n(y) \;=\; \begin{cases} a/\varepsilon_n & \text{if} \quad y < 0,\\ -\tilde\eta_n & \text{if} \quad y \in [0,1],\\ b/\varepsilon_n & \text{if} \quad y > 1, \end{cases} \qquad\qquad (\tilde\eta_n, \tilde\sigma_n) \;=\; \left(\frac{\eta_n}{\varepsilon_n}, \frac{\sigma_n}{\varepsilon_n}\right). \tag{6.22}$$

5. Toward the analysis of (6.21)-(6.22), i n this step we perform a preliminary computation. Let Y_t be the random solution to the diffusion process with constant coefficients, on the unit interval:

$$dY_t \;=\; -\,\eta\, Y_t dt + \sigma\, dW_t \tag{6.23}$$

starting at the origin, with the following boundary conditions.

- **Reflecting at the point** $y = 0$**:** when $Y_t(\omega) = 0$, the particle is reflected back inside the domain $]0,1[$;
- **Absorbing at the point** $y = 1$**:** when $Y_t(\omega) = 1$, its motion stops forever.

It is well known that in this case the distribution function

$$u(t,y) \;=\; \text{Prob.}\{Y_t(\omega) \le y\}$$

satisfies the parabolic equation

$$u_t \;=\; \eta u_y + \frac{\sigma^2}{2} u_{yy}\,, \tag{6.24}$$

with initial and boundary conditions

$$u(0,y) \;=\; 1, \qquad\qquad \begin{cases} u(t,0) &=\; 0,\\ u_x(t,1) &=\; 0. \end{cases} \tag{6.25}$$

We seek a lower and an upper solution of (6.24)-(6.25) in terms of suitable eigenfunctions. This leads us to the boundary value problem

$$\frac{\sigma^2}{2} w''(y) + \eta w'(y) \;=\; -\lambda w(y), \qquad w(0) = 0, \quad w'(1) = 0\,. \tag{6.26}$$

Explicit solutions of (6.26) are found by solving

$$\frac{\sigma^2}{2} r^2 + \eta r + \lambda \;=\; 0, \qquad\qquad r \;=\; \frac{-\eta \pm \sqrt{\eta^2 - 2\sigma^2\lambda}}{\sigma^2} \;\doteq\; \gamma \pm s.$$

For a given $\lambda > 0$, choose any $s > 0$ and let $\eta, \sigma > 0$, $\gamma < 0$ be the constants such that

$$\frac{\sigma^2}{2} = \frac{(e^{2s}-1)^2}{4s^2 e^{2s}} \cdot \lambda, \qquad \eta = s \cdot \frac{e^s + e^{-s}}{e^s - e^{-s}} \cdot \sigma^2 \qquad \text{and} \qquad \gamma = \frac{-\eta}{\sigma^2} = -s \cdot \frac{e^{2s}+1}{e^{2s}-1}. \quad (6.27)$$

In this case, an increasing solution to (6.26), normalized so that $w(1) = 1$, is explicitly found to be (see Fig. 6)

$$w(y) = \frac{e^{(\gamma+s)y} - e^{(\gamma-s)y}}{e^{\gamma+s} - e^{\gamma-s}} \qquad \text{for all } y \in [0,1]. \quad (6.28)$$

The function

$$w^-(t,y) = e^{-\lambda t} w(y) \quad (6.29)$$

thus provides a lower solution to the parabolic problem (6.24)-(6.25).

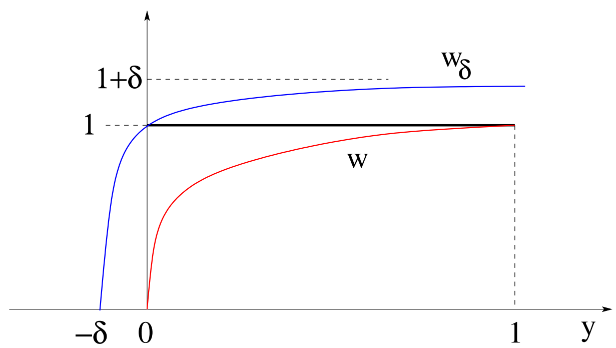

Figure 6: The function w at (6.28), used to construct a lower solution to (6.38), and the functions w_δ at (6.30)-(6.31), used to construct upper solutions.

We now work toward the construction of an upper solution. We claim that, for every $\delta > 0$, one can find constants $\sigma_\delta, \eta_\delta$ large enough, and a function $w = w_\delta(x)$ such that (see Fig. 6)

$$w(-\delta) = 0, \qquad w(0) = 1, \qquad w(1) \leq 1 + \delta, \qquad w'(1) > 0, \quad (6.30)$$

and moreover

$$\frac{\sigma_\delta^2}{2} w'' + \eta_\delta w' + (\lambda - \delta) w \leq 0 \qquad x \in [-\delta, 1]. \quad (6.31)$$

To construct the upper solution w_δ, let $\lambda, s > 0$ be given, and let σ, η, γ be as in (6.27). For any small $\delta > 0$, define the constants $s_\delta, \lambda_\delta > 0$ in terms of the two equations

$$\gamma = -s_\delta \cdot \frac{e^{2s_\delta(1+2\delta)}+1}{e^{2s_\delta(1+2\delta)}-1} \qquad \text{and} \qquad \lambda_\delta = \frac{\gamma^2 - s_\delta^2}{2\sigma^2}. \quad (6.32)$$

Then, on the larger interval $[-\delta, 1]$, the problem

$$\frac{\sigma^2}{2} w''(y) + \eta w'(y) = -\lambda_\delta w(y), \qquad w(-\delta) = 0, \qquad w(0) = 1, \quad (6.33)$$

has a unique increasing solution (see Fig. 6), namely

$$w_\delta(y) = \frac{e^{(\gamma+s_\delta)(y+\delta)} - e^{(\gamma-s_\delta)(y+\delta)}}{e^{(\gamma+s_\delta)\delta} - e^{(\gamma-s_\delta)\delta}}, \qquad y \in [-\delta, 1]. \quad (6.34)$$

Using (6.32), we compute

$$w_\delta(1) \;=\; \frac{e^{(\gamma+s_\delta)(1+\delta)}-e^{(\gamma-s_\delta)(1+\delta)}}{e^{(\gamma+s_\delta)\delta}-e^{(\gamma-s_\delta)\delta}} \;\leq\; \frac{e^{\gamma+s_\delta}}{1-e^{-2s_\delta\delta}},$$

$$w'_\delta(1) \;=\; \frac{e^{(\gamma-s_\delta)(1+\delta)}}{e^{(\gamma+s_\delta)\delta}-e^{(\gamma-s_\delta)\delta}}\cdot(\gamma-s_\delta)\cdot\Big(e^{-2s_\delta\delta}-1\Big) \;>\; 0. \tag{6.35}$$

Since $s\cdot\dfrac{e^{2s}+1}{e^{2s}-1}=s_\delta\cdot\dfrac{e^{2s_\delta(1+2\delta)}+1}{e^{2s_\delta(1+2\delta)}-1}$, one has

$$s\cdot\frac{e^{2s}+1}{e^{2s}-1}\leq s_\delta\cdot\frac{e^{2s_\delta}+1}{e^{2s_\delta}-1},\qquad s_\delta\;\leq\; s\cdot\left(1+\frac{2}{e^{2s}-1}\right)\;\leq\; s+1.$$

In particular, for $s\geq 1$ and $0<\delta\leq 1/2$, we have

$$s\;\leq\; s_\delta\;\leq\; 2s,\qquad \gamma+s_\delta \;=\; -\,\frac{2s_\delta}{e^{2s_\delta(1+2\delta)}-1}\;\leq\; -\,2se^{-8s},$$

and hence

$$w_\delta(1)\;\leq\;\frac{e^{-2se^{-8s}}}{1-e^{-2s\delta}},\qquad\qquad \frac{-3s^4e^{2s}}{(e^{2s}-1)^2\lambda}\;\leq\;\frac{s^2-s_\delta^2}{2\sigma^2}\;=\;\lambda_\delta-\lambda\;\leq\;0.$$

Therefore, by choosing $s=\delta^{-2}$, so that $\delta=1/\sqrt{s}$, one obtains

$$w_\delta(1)\;\leq\;1+\delta,\qquad\qquad \lambda-\delta\;\leq\;\lambda_\delta\;\leq\;\lambda, \tag{6.36}$$

provided that $s>0$ is sufficiently large.

6. We are now ready to construct a sequence of diffusion processes with piecewise constant drift, which approximate a Poisson waiting time at $x_0=0$. For each $t>0$, the transition kernel we need to approximate is (see Fig. 7)

$$U(t,x)\;=\;\text{Prob.}\{X_t(\omega)\leq x\}\;=\;\begin{cases} 0 & \text{if}\quad x\leq 0,\\ \exp\Big\{-\lambda\Big(t-\dfrac{x}{b}\Big)\Big\} & \text{if}\quad x\in[0,bt],\\ 1 & \text{if}\quad x\geq bt.\end{cases} \tag{6.37}$$

Indeed, since particles travel with constant speed $b>0$, a particle will reach the point $x>0$ after time t if and only if it departs from the origin after time $t-(x/b)$. The probability of this event is $\exp\big\{-\lambda\big(t-\frac{x}{b}\big)\big\}$.

Next, we compare the kernel (6.37) with the solution $u=u_n(t,x)$ of the parabolic problem

$$u_t+g_n(x)u_x\;=\;\frac{\sigma_n^2}{2}u_{xx}\,,\qquad x\in\mathbb{R}, \tag{6.38}$$

with g_n as in (6.19), and with initial data

$$u(0,x)\;=\;\begin{cases}0 & \text{if}\quad x<0\,,\\ 1 & \text{if}\quad x\geq 0\,.\end{cases} \tag{6.39}$$

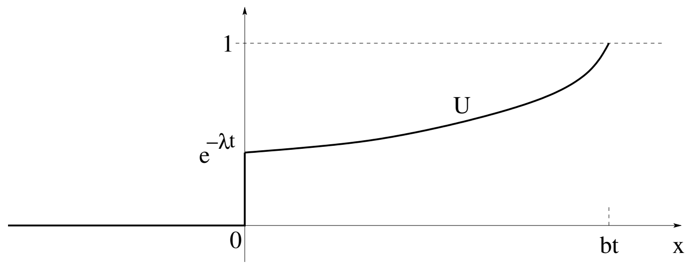

Figure 7: The transition kernel (6.37), corresponding to a random Poisson waiting time at $x = 0$, followed by motion with constant speed $b > 0$.

We claim that, for a suitable choice of the sequences, $\sigma_n, \varepsilon_n, \eta_n \to 0$, one can achieve the weak convergence of the transition kernels. In other words, the sequence of solutions $u_n = u_n(t, \cdot)$ of the parabolic problems (6.38)-(6.39) converge to $U(t, \cdot)$ in $\mathbf{L}^1_{loc}(\mathbb{R})$, for every $t > 0$. The claim will be proved by constructing suitable sequences of upper and lower solutions.

The sequences $\varepsilon_n, \eta_n \to 0$, are chosen as follows. First, we take a sequence $s_n \to +\infty$. Then, in view of (6.22) and (6.27), we define

$$\frac{\widetilde{\sigma}_n^2}{2} \;=\; \frac{(e^{2s_n-1})^2}{4s_n^2 e^{2s_n}} \cdot \lambda, \qquad \widetilde{\eta}_n \;=\; s_n \cdot \frac{e^{s_n} + e^{-s_n}}{e^{s_n} - e^{-s_n}} \cdot \widetilde{\sigma}_n^2\,, \tag{6.40}$$

$$\eta_n \;=\; \varepsilon_n \widetilde{\eta}_n, \qquad \sigma_n \;=\; \varepsilon_n \widetilde{\sigma}_n\,. \tag{6.41}$$

Notice that here $\widetilde{\sigma}_n, \widetilde{\eta}_n \to +\infty$. Therefore, we need to choose a sequence $\varepsilon_n \to 0$ converging to zero fast enough so that $\eta_n, \sigma_n \to 0$.

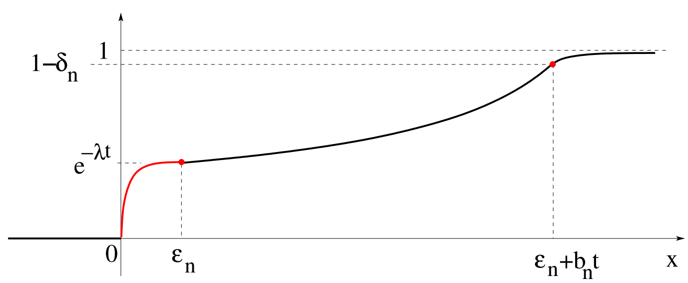

Figure 8: The lower solution defined at (6.42).

Let w_n^- be the solution to (6.26) with $\sigma = \tilde{\sigma}_n$, $\eta = \tilde{\eta}_n$ and $w_n^-(1) = 1$. We now set $\delta_n = 1/\sqrt{s_n}$. For suitable sequences $\kappa_n \to +\infty$ and $b_n \downarrow b$, we define

$$u_n^-(t,x) \;=\; \begin{cases} 0 & \text{if} \quad x < 0, \\ (1-\delta_n)e^{-\lambda t} w_n^-(x/\varepsilon_n) & \text{if} \quad x \in [0, \varepsilon_n], \\ (1-\delta_n)\exp\left\{-\lambda\Big(t - \dfrac{x-\varepsilon_n}{b_n}\Big)\right\} & \text{if} \quad x \in \big[\varepsilon_n, \varepsilon_n + b_n t\big], \\ 1 - \delta_n \exp\Big\{ -\kappa_n\big(x - \varepsilon_n - b_n t\big)\Big\} & \text{if} \quad x \geq \varepsilon_n + b_n t. \end{cases} \tag{6.42}$$

This lower solution is illustrated in Fig. 8. Here we choose $\kappa_n \to +\infty$ fast enough so that $\delta_n \kappa_n \to +\infty$. Finally, we choose the decreasing sequence $b_n \downarrow b$ converging slowly enough so that the traveling profile

$$v(t,x) \;=\; 1-\delta_n e^{-\kappa_n(x-\varepsilon_n-b_n t)}$$

is a lower solution. This will be the case if

$$v_t \;\leq\; \frac{\sigma_n^2}{2} v_{xx} + b_n v_x \,.$$

That is, iff

$$-b_n\kappa_n \;\leq\; -\frac{\sigma_n^2}{2}\kappa_n^2 - b\kappa_n \,.$$

We thus need

$$b_n - b \;\geq\; \frac{\sigma_n^2 \kappa_n}{2} \,.$$

Notice that, since the product $\delta_n\kappa_n \to +\infty$, this ensures that the derivative of the solution at the matching point $x=\varepsilon_n + b_n t$ jumps upward. The fact that this derivative also jumps upward at $x=0$ and at $x=\varepsilon_n$ is obvious.

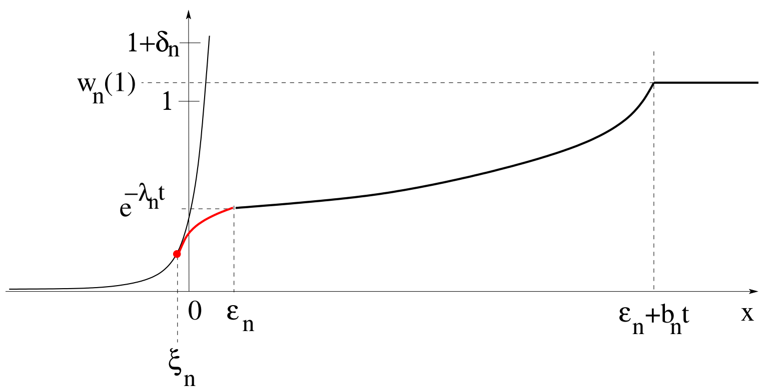

Figure 9: The upper solution defined at (6.43).

Next, an upper solution is constructed as follows. As before, set $\delta_n = 1/\sqrt{s_n}$. Let $w_n = w_{\delta_n}$ be a sequence of functions satisfying (6.30)-(6.31).

We then consider the rescaled functions $x\mapsto w_n(\varepsilon_n^{-1}x)$, choosing a sequence $\varepsilon_n \downarrow 0$ fast enough so that $\varepsilon_n^{-1} w_n'(1)\to +\infty$.

Next, we take a sequence $\kappa_n \to +\infty$. For each $n\geq 1$, we choose the values x_n and $\xi_n \in [-\delta_n\varepsilon_n, 0]$ so that at $x=\xi_n$ the two functions

$$x\mapsto w_n(\varepsilon_n^{-1}x) \qquad \text{and} \qquad x\mapsto e^{\kappa_n(x-x_n)}$$

are tangent:

$$w_n(\varepsilon^{-1}\xi_n) \;=\; e^{\kappa_n(\xi_n - x_n)}, \qquad \varepsilon^{-1}w_n'(\varepsilon^{-1}\xi_n) \;=\; \kappa_n e^{\kappa_n(\xi_n-x_n)} \,.$$

Finally, we choose an increasing sequence of speeds $b_n^- \uparrow b$ and define (see Fig. 9)

$$u_n^+(t,x) \;=\; \begin{cases} e^{-\lambda_n t} e^{\kappa_n(x-x_n)} & \text{if} \quad x<\xi_n, \\ e^{-\lambda_n t} w_n(x/\varepsilon_n) & \text{if} \quad x\in[\xi_n,\varepsilon_n], \\ w_n(1)\cdot \exp\left\{-\lambda_n\Big(t-\dfrac{x-\varepsilon_n}{b_n^-}\Big)\right\} & \text{if} \quad x\in\big[\varepsilon_n,\ \varepsilon_n + b_n^- t\big], \\ w_n(1) & \text{if} \quad x\geq \varepsilon_n + b_n^- t. \end{cases} \tag{6.43}$$

The condition $\varepsilon_n^{-1} w_n(1) \to +\infty$ guarantees that at the point $x = \varepsilon_n$ the derivative $\partial_x u_n^+$ has a downward jump. On the other hand, this derivative is continuous at the matching point $x = \xi_n$, and has a downward jump at $x = \varepsilon_n + b_n^- t$. We thus conclude that u_n^+ is an upper solution.

Since the sequences u_n^- and u_n^+ at (6.42) and (6.43) both converge to the distribution function U at (6.37), for all $t \geq 0$, this proves the weak convergence of the transition kernels, starting at the origin.

For a starting point $x_0 \neq 0$, the convergence of the corresponding transition kernels is a straightforward.

7. To cover the set of all possibities, one should also consider a dynamics of the form (1.19) in two additional cases:

CASE 4: $a, b > 0$, and there is no stopping time at $x = 0$.

CASE 5: $b < 0 < a$, so that all trajectories stop forever, as they reach the origin.

In both of these cases, the approximation of the Markov semigroup with a diffusion process is trivial. Indeed, taking $g_n(x) = f(x)$ for all $n \geq 1$, as the coefficients $\sigma_n \downarrow 0$, the transition kernels for the diffusion process converge to the corresponding transition kernels of the Markov semigroup.

In view of Lemma 5.1, this completes the proof of Theorem 1.2. □

References

[1] J. P. Aubin and A. Cellina, *Differential inclusions. Set-valued maps and viability theory.* Springer-Verlag, Berlin, 1984.

[2] P. Binding, The differential equation $\dot{x} = f \circ x$, *J. Differential Equations* **31** (1979), 183–199.

[3] A. Bressan, M. Mazzola, and K. T. Nguyen, Markovian solutions to discontinuous ODEs, *J. Dynam. Diff. Equat.* (2021).

[4] A. Cellina, Approximation of set valued functions and fixed point theorems. *Ann. Mat. Pura Appl.* **82** (1969), 17–24.

[5] G. Crippa, and S. Spirito, Smooth approximation is not a selection principle for the transport equation with rough vector field. *Calc. Var. Partial Diff. Equat.* **59** (2020), no. 1, Paper No. 13.

[6] D. C. Biles, M. Federson, and R. L. Pouso, A survey of recent results for the generalizations of ordinary differential equations. *Abstr. Appl. Anal.* Art. ID 260409, (2014).

[7] W. E and E. Vanden-Eijnden, A note on generalized flows. *Phys. D* **183** (2003), 159–174.

[8] A. Filippov, Differential equations with discontinuous right-hand side, *Amer. Math. Soc. Transl.* (1964), 199–231.

[9] F. Flandoli and J. Langa, Markov attractors: a probabilistic approach to multivalued flows. *Stoch. Dyn.* **8** (2008), 59–75.

[10] I. I. Gihman and A. V. Skorohod, *The Theory of Stochastic Processes*, Vols. I, II, and III, Springer-Verlag, Berlin, 1974, 1975, and 1979.

[11] S. Karlin and H. M. Taylor, *A Second Course in Stochastic Processes.* Academic Press, New York-London, 1981.

[12] B. Oksendal, *Stochastic Differential Equations. An Introduction with Applications.* Fifth Edition. Springer, 2003.

[13] D. Stroock, *An introduction to the theory of large deviations.* Universitext. Springer-Verlag, New York, 1984.

[14] D. W. Stroock and S. R. S. Varadhan, *Multidimensional Diffusion Processes.* Springer, Berlin, 1979.

[15] S. R. S. Varadhan, *Large deviations.* Courant Lecture Notes in Mathematics, 27. American Mathematical Society, Providence, RI, 2016.