Supplementary material of "Hybrid Limits of Continuous Time Markov Chains"

This is an online supplementary document for the paper "Hybrid Limits of Continuous Time Markov Chains", published in the proceeding of QEST 2011. In this document, we present detailed proofs of the theorems stated in the paper.

We recall here, for simplicity, all the assumptions made on the PWS system defined by the sequence $\hat{\mathcal{X}}^{(N)}$ of CTMC models and the conditions of regularity that the solutions of these systems must satisfy. The discontinuity surfaces of the PWS system F are defined by the inequalities that are conjuncts of the guard predicates φ_i of transitions of $\hat{\mathcal{X}}^{(N)}$. We assumed that there are m_0 non-trivial and distinct inequalities, each of the form $h_j(\hat{\mathbf{X}}) \geq 0$ or $h_j(\hat{\mathbf{X}}) > 0$. Functions h_j are sufficiently smooth (they must have continuous second order derivatives, cf. Section 3). Furthermore, we consider only *regular* solutions of the PWS system starting from $\mathbf{x_0}$: they must be unique in $[0, T]$, they must be non-Zeno, and they can intersect only one manifold $\mathcal{H}_j$ at a time. Furthermore, sliding motion must always terminate with first order exit conditions (not both vector fields on the two sides of $\mathcal{H}_j$ can become tangent to $\mathcal{H}_j$ simultaneously).

First, we prove the convergence of the sequence $\hat{\mathcal{X}}^{(N)}$ of models to the solution of the PWS in case of sliding motion (Theorem 4.2). Then we deal with the main theorem (Theorem 4.1)

Dealing with sliding motion

We focus now on sliding motion. In particular, we prove that if the sequence $\hat{\mathcal{X}}^{(N)}$ of CTMC starts (in the limit) in a point of a discontinuity surface $\mathcal{H}$ from which the (unique) solution of the PWS system follows the sliding vector field, then the sequence of CTMC will converge to this trajectory. The regularity conditions imposed on solutions of the PWS system and the local nature of limit theorems allow us to reason on a simpler setting than the one of the PWS system constructed from the CTMC sequence. In fact, we will consider a PWS system composed only of two regions, separated by the manifold $\mathcal{H}$ defined by $\mathcal{H} = \{\mathbf{x} \in \mathbb{R}^n \mid h(\mathbf{x}) = 0\}$, for a smooth function h (at least $\mathcal{C}_2$). Then, we will assume that the initial points $\hat{\mathbf{X}}^{(N)}(0)$ of the CTMC sequence converge (in probability) to a point $\mathbf{x_0} \in \mathcal{H}$, such that $n^T(\mathbf{x_0})f_1(\mathbf{x_0}) < 0$ and $n^T(\mathbf{x_0})f_2(\mathbf{x_0}) > 0$, where f_1 is the vector field in the region $\mathcal{R}_1 = \{\mathbf{x} \in \mathbb{R}^n \mid h(x) \geq 0\}$ and f_2 the vector field in the region $\mathcal{R}_2 = \{\mathbf{x} \in \mathbb{R}^n \mid h(x) < 0\}$. Notice that this condition guarantees that the solution of the PWS starting in $\mathbf{x_0}$ is unique and it will be sliding along $\mathcal{H}$ up to time T_S, the time in which the trajectory leaves the surface $\mathcal{H}$ with first-order exit conditions.

First of all, we need the following lemma, adapted from [Kur70] by specializing it to our scaling conditions:

Lemma .1. *Let $\hat{\mathcal{X}}^{(N)}$ be a sequence of CTMC models satisfying the scaling assumptions of Section 2.2, and let $\hat{\mathbf{X}}^{(N)}(t)$ be the associated CTMC processes. Let the drift for $\hat{\mathbf{X}}^{(N)}(t)$ be F, defined in $E \subseteq \mathbb{R}^n$. Then, for every $\delta > 0$ and any finite $T > 0$,*

$$\lim_{N\to\infty} \sup_{\hat{\mathbf{X}}_0 \in \hat{\mathcal{D}}^{(N)}} \mathbb{P}\left\{ \sup_{t \leq T} \left\| \hat{\mathbf{X}}^{(N)}(t) - \hat{\mathbf{X}}^{(N)}(0) - \int_0^t F(\hat{\mathbf{X}}^{(N)}(s))ds \right\| > \delta \mid \hat{\mathbf{X}}^{(N)}(0) = \hat{\mathbf{X}}_0 \right\} = 0. \blacksquare$$

The previous lemma essentially proves that in the limit of $N \to \infty$, the process $\hat{\mathbf{X}}^{(N)}(t)$ becomes deterministic, i.e. fluctuations vanish. In fact, the lemma states that the process

$\hat{\mathbf{X}}^{(N)}(t)$ converges in the limit to the solution of an integral equation, defined by the drift F of the sequence, if such a solution exists. This is essentially the core of the proof of Theorem 2.1. In fact, once this lemma is established, the next step to prove the theorem is just to show that the solution of this integral equation is exactly the solution of the fluid ODE. This is the point in which the Lipschitz continuity is used (applying Gronwall inequality). Here, however, we need to do some extra work, as the drift that defines the integral equation is not the vector field that defines the sliding motion along $\mathcal{H}$, and the limit integral equation may not even have a solution. Essentially, we have to show that we can replace F by the sliding vector field g in this integral equation. Recall that g is defined as a convex combination of f_1 and f_2: $g(\mathbf{x}) = \lambda(\mathbf{x})f_1(\mathbf{x}) + (1-\lambda(\mathbf{x}))f_2(\mathbf{x})$, $\lambda(\mathbf{x}) = \frac{f_2^{\perp}(\mathbf{x})}{f_2^{\perp}(\mathbf{x})-f_1^{\perp}(\mathbf{x})}$, with $f_i^{\perp}(\mathbf{x}) = n^T(\mathbf{x})f_i(\mathbf{x})$. The conditions on the sliding motion (i.e. the relative orientation of f_1 and f_2 with respect to $\mathcal{H}$) imply that $f_2^{\perp}(\mathbf{x}) - f_1^{\perp}(\mathbf{x}) > \varepsilon > 0$ (in an open neighbourhood of the solution $\mathbf{x}(t)$, for $t \leq T_S$), hence λ and g are Lipschitz.[1]

We can now prove convergence of the sequence $\hat{\mathbf{X}}^{(N)}(t)$ of CTMC to the trajectory sliding along the surface $\mathcal{H}$.

Theorem .1. *Let the sequence $\hat{\mathcal{X}}^{(N)}$ of CTMC models satisfy the scaling assumptions of Section 2.2 and consider the PWS system $\frac{d}{dt}\mathbf{x} = F(\mathbf{x})$, assumed to have only two continuity regions separated by a smooth manifold. Let $\mathcal{H}$ be the discontinuity surface of the PWS system, and assume $\mathbf{x}_0 \in \mathcal{H}$ is such that $f_1^{\perp}(\mathbf{x}_0) := n^T(\mathbf{x}_0)f_1(\mathbf{x}_0) < 0$ and $f_2^{\perp}(\mathbf{x}_0) := n^T(\mathbf{x}_0)f_2(\mathbf{x}_0) > 0$. Let $\mathbf{x}(t)$, $t \leq T_S$ be the unique solution of the PWS system starting from $\mathbf{x}_0$, defined as the solution of the ODE $\frac{d}{dt}\mathbf{x} = g(\mathbf{x})$, where g is the sliding vector field defined as in Section 3, and $T_S \leq \infty$ is the time at which sliding motion terminates, with first order exit conditions. Fix $T \leq T_S$, $T < \infty$. If $\mathbf{X}^{(N)}(0) \to \mathbf{x}_0$ in probability, then*

$$\lim_{N\to\infty} \sup_{t\leq T} \left\| \hat{\mathbf{X}}^{(N)}(t) - \mathbf{x}(t) \right\| = 0 \quad \textit{in probability.}$$

Proof. Applying triangular inequality and recalling that $\mathbf{x}(t) = \mathbf{x}(0) + \int_0^t g(\mathbf{x}(s))ds$, it holds that:

$$\begin{aligned}
\sup_{t\leq T} \left\| \mathbf{X}^{(N)}(t) - \mathbf{x}(t) \right\| &\leq \sup_{t\leq T} \left\| \mathbf{X}^{(N)}(t) - \mathbf{X}^{(N)}(0) - \int_0^t F(\mathbf{X}^{(N)}(s))ds \right\| && (2)\\
&+ \left\| \mathbf{X}^{(N)}(0) - \mathbf{x}(0) \right\| && (3)\\
&+ \sup_{t\leq T} \left\| \int_0^t F(\mathbf{X}^{(N)}(s)) - g(\mathbf{X}^{(N)}(s))ds \right\| && (4)\\
&+ \sup_{t\leq T} \left\| \int_0^t g(\mathbf{X}^{(N)}(s)) - g(\mathbf{x}(s))ds \right\|. && (5)
\end{aligned}$$

The term (2) goes to zero in probability, thanks to Lemma .1. The term (3) goes to zero in probability by hypothesis. Term (4) goes to zero in probability thanks to Lemma .3 below. Hence, for each $\varepsilon > 0$, with probability $(1-\delta_N)$, $\delta_N \to 0$, it holds that

$$\sup_{t\leq T} \left\| \mathbf{X}^{(N)}(t) - \mathbf{x}(t) \right\| \leq \varepsilon + \int_0^T \sup_{s\leq t} \| g(\mathbf{X}^{(N)}(s)) - g(\mathbf{x}(s)) \| dt.$$

[1] Due to the local nature of convergence theorems, it is not restrictive to localize properties of functions, like Lipschitzness, in an open neighbourhood of the solution, cf also Remark .1.

Now the proof uses Lipschitz continuity of g and the classical argument of Gronwall inequality, as in Theorem 2.1[2], obtaining

$$\sup_{t \leq T} \left\| \mathbf{X}^{(N)}(t) - \mathbf{x}(t) \right\| \leq \varepsilon e^T,$$

which proves convergence in probability of $\sup_{t \leq T} \|\mathbf{X}^{(N)}(t) - \mathbf{x}(t)\|$ to zero, due to the arbitrariness of ε. ■

Inspecting the proof of the previous theorem, and comparing it with the proof of Theorem 2.1 [Dar02], we can notice that the main difference is in the presence of term (4) in the main inequality of the proof. In order to show that the sequence of CTMC converges to the sliding motion, in fact, we essentially prove that $\hat{\mathbf{X}}^{(N)}(t)$, in the limit, remains on the surface $\mathcal{H}$ and that the integral of its drift along a trajectory becomes indistinguishable in the limit from the integral of the sliding vector field.

In order to prove that term (4) go to zero in probability, we may exploit the essence of Lemma .1. In fact, this lemma shows that in the limit the CTMC sequence $\hat{\mathbf{X}}^{(N)}(t)$ becomes indistinguishable in $[0, T]$ from the solution of the integral equation $\mathbf{y}(t) = \mathbf{y}(0) + \int_0^t F(\mathbf{y}(s))ds$. Therefore, it holds that $\sup_{t \leq T} \| \int_0^t F(\hat{\mathbf{X}}^{(N)}(s)) - g(\hat{\mathbf{X}}^{(N)}(s))ds\| \to \sup_{t \leq T} \| \int_0^t F(\mathbf{y}(s)) - g(\mathbf{y}(s))ds\|$ in probability, hence one can conclude that term (4) goes to zero by showing that $\sup_{t \leq T} \| \int_0^t F(\mathbf{y}(s)) - g(\mathbf{y}(s))ds\| = 0$. Even if this reasoning seems appealing, it relies on the *existence* of the solution of the integral equation $\mathbf{y}(t) = \mathbf{y}(0) + \int_0^t F(\mathbf{y}(s))ds$, which is not guaranteed due to the discontinuity of F, even if the integral of F along $\hat{\mathbf{X}}^{(N)}(t)$ is always defined.

Hence, we need a different approach. The basic idea is to replace the integral equation $\mathbf{y}(t) = \mathbf{y}(0) + \int_0^t F(\mathbf{y}(s))ds$ by a discrete version of it, and then show that the supremum of the integral in term (4) goes to zero along the solution of this discrete equation. The correctness of this kind of reasoning relies on the fact that the sequence of CTMC and the sequence of DTMC constructed from the former using the uniformization construction (and ignoring the Poisson process) become indistinguishable in the limit of $N \to \infty$. This allows us to use the discrete version of Lemma .1, in which the integral equation for $\mathbf{y}(t)$ is replaced by a discrete equation. We properly formalize this approach in the following lemma.

Lemma .2. *Let $\hat{\mathcal{X}}^{(N)}$ be a sequence of CTMC models and let $\hat{\mathbf{X}}^{(N)}(t)$ be its sequence of CTMC, with drift F defined in $E \subseteq \mathbb{R}^n$. Let $\Lambda \geq \sup_{\mathbf{x} \in E} \sum_i f_i(\mathbf{x}) I_{\varphi_i}(\mathbf{x})$, so that ΛN is an upper bound for the exit rate of $\hat{\mathbf{X}}^{(N)}(t)$ in each state, and define $\mathbf{z}^{(N)}(k) = \mathbf{z}^{(N)}(0) + \sum_{i=0}^{k-1} \frac{1}{\Lambda N} F(\mathbf{z}^{(N)}(i))$ and $\mathbf{y}^{(N)}(t) = \mathbf{z}^{(N)}(\lfloor \Lambda N t \rfloor)$. Then*

$$\sup_{t \leq T} \left\| \hat{\mathbf{X}}^{(N)}(t) - \mathbf{y}^{(N)}(t) \right\| \to 0 \quad \textit{in probability.}$$

Proof. We first recall the definition of uniformization [Jen53] of CTMC $\hat{\mathbf{X}}^{(N)}(t)$. Essentially, $\hat{\mathbf{X}}^{(N)}(t)$ is decoupled in two independent processes: a Discrete Time Markov Chain (DTMC) $\mathbf{Z}^{(N)}(k)$ and a Poisson process $\mathcal{N}$. Let ΛN be the upper bound on the exit rate of $\hat{\mathbf{X}}^{(N)}(t)$, as defined above. The DTMC is constructed by allowing the same set of transitions out of $\hat{\mathbf{X}}$ as in the CTMC $\hat{\mathbf{X}}^{(N)}(t)$, with probabilities $p_i(\hat{\mathbf{X}}) = \frac{f_i(\hat{\mathbf{X}})}{\Lambda}$. The drift of this DTMC is defined [BB08] by $\sum_i \Lambda N \frac{1}{N} \mathbf{v}_i p_i(\hat{\mathbf{X}}) = F(\hat{\mathbf{X}})$, i.e. it equals that of the CTMC. The uniformized CTMC is the process $\mathbf{Y}_c^{(N)}(t) = \mathbf{Z}^{(N)}(\mathcal{N}(\Lambda N t))$, and it

[2] Gronwall inequality states that, if $y(t) \leq a + b \int_0^t y(s)ds$, then $y(t) \leq a e^{bt}$.

equals $\hat{\mathbf{X}}^{(N)}(t)$. Furthermore, let $\mathbf{Y}_d^{(N)}(t) = \mathbf{Z}^{(N)}(\lfloor \Lambda N t \rfloor)$ be the DTMC $\mathbf{Z}^{(N)}$ in which steps last $\frac{1}{\Lambda N}$ time units. A known result in stochastic processes literature (see [Kal02], proof of Theorem 17.28) is that $\mathbf{Y}_c^{(N)}$ converges to $\mathbf{Y}_d^{(N)}$ weakly. Furthermore, if $\mathbf{Y}_d^{(N)}(t)$ converges in probability to a deterministic process $\mathbf{y}^{(N)}(t)$ uniformly in $[0, T]$, then so does $\mathbf{Y}_c^{(N)}(t)$. Now we combine this result with a discrete-time version of Lemma .1 holding for DTMC, see Proposition 4.1 of [Kur70], which essentially states that

$$\sup_{t \leq T} \left\| \mathbf{Y}_d^{(N)}(t) - \mathbf{y}^{(N)}(t) \right\| \to 0 \text{ in probability.}$$

■

We can now prove that term (4) goes to zero in probability.

Lemma .3. *Let $\hat{\mathbf{X}}^{(N)}(t)$, F, $\mathcal{H}$, g, and Λ be defined as above. Then*

$$\lim_{N \to \infty} \sup_{t \leq T} \left\| \int_0^t F(\hat{\mathbf{X}}^{(N)}(s)) - g(\hat{\mathbf{X}}^{(N)}(s)) ds \right\| = 0 \quad \textit{in probability.}$$

Proof. First of all, notice that

$$\begin{aligned} \sup_{t \leq T} \left\| \int_0^t F(\hat{\mathbf{X}}^{(N)}(s)) - F(\mathbf{y}^{(N)}(s)) ds \right\| &\leq \sup_{t \leq T} \left\| \hat{\mathbf{X}}^{(N)}(0) + \int_0^t F(\hat{\mathbf{X}}^{(N)}(s)) ds - \hat{\mathbf{X}}^{(N)}(t) \right\| \\ &+ \sup_{t \leq T} \left\| \hat{\mathbf{X}}^{(N)}(t) - \mathbf{y}^{(N)}(t) \right\| \\ &+ \left\| \mathbf{y}^{(N)}(0) - \hat{\mathbf{X}}^{(N)}(0) \right\| \end{aligned}$$

and all the terms on the right hand side go to zero in probability. Furthermore,

$$\begin{aligned} \sup_{t \leq T} \left\| \int_0^t g(\hat{\mathbf{X}}^{(N)}(s)) - g(\mathbf{y}^{(N)}(s)) ds \right\| &\leq \sup_{t \leq T} \int_0^t L_g \left\| \hat{\mathbf{X}}^{(N)}(s) - \mathbf{y}^{(N)}(s) \right\| ds \\ &\leq L_g T \sup_{t \leq T} \left\| \hat{\mathbf{X}}^{(N)}(t) - \mathbf{y}^{(N)}(t) \right\|, \end{aligned}$$

where L_g is the Lipschitz constant of the function g. Now, it holds that

$$\begin{aligned} \sup_{t \leq T} \left\| \int_0^t F(\hat{\mathbf{X}}^{(N)}(s)) - g(\hat{\mathbf{X}}^{(N)}(s)) ds \right\| &\leq \sup_{t \leq T} \left\| \int_0^t F(\hat{\mathbf{X}}^{(N)}(s)) - F(\mathbf{y}^{(N)}(s)) ds \right\| \\ &+ \sup_{t \leq T} \left\| \int_0^t g(\mathbf{y}^{(N)}(s)) - g(\hat{\mathbf{X}}^{(N)}(s)) ds \right\| \\ &+ \sup_{t \leq T} \left\| \int_0^t F(\mathbf{y}^{(N)}(s)) - g(\mathbf{y}^{(N)}(s)) ds \right\|. \end{aligned}$$

Therefore, we just need to prove that

$$\sup_{t \leq T} \left\| \int_0^t F(\mathbf{y}^{(N)}(s)) - g(\mathbf{y}^{(N)}(s)) ds \right\|$$

goes to zero in the limit to conclude the proof.

Now, recall that $\mathbf{y}^{(N)}(t) = \mathbf{z}^{(N)}(\lfloor \Lambda N t \rfloor)$, hence we can focus the attention on the sequence $\mathbf{z}^{(N)}(k)$.

First of all, observe that $F(\mathbf{x}) - g(\mathbf{x})$ equals $(1 - \lambda(\mathbf{x}))(f_1(\mathbf{x}) - f_2(\mathbf{x}))$ for $\mathbf{x} \in \mathcal{R}_1$ and $-\lambda(\mathbf{x})(f_1(\mathbf{x}) - f_2(\mathbf{x}))$ for $\mathbf{x} \in \mathcal{R}_2$. Therefore, on the two sides of the surface, the vectors that we are integrating are collinear but have opposite orientations. In particular, close to $\mathcal{H}$, $f_1(\mathbf{x}) - f_2(\mathbf{x})$ points towards $\mathcal{R}_2$, as $n^T(\mathbf{x})f_1(\mathbf{x}) - n^T(\mathbf{x})f_2(\mathbf{x}) < 0$. Therefore, $F(\mathbf{x}) - g(\mathbf{x})$ points towards $\mathcal{R}_2$ if $\mathbf{x} \in \mathcal{R}_1$ and towards $\mathcal{R}_1$ if $\mathbf{x} \in \mathcal{R}_2$.

Now, the intuition is that $\mathbf{z}^{(N)}(k)$ will remain closer and closer to the surface $\mathcal{H}$, as it is pushed towards $\mathcal{R}_1$ while in $\mathcal{R}_2$, and viceversa. Now, it holds that

$$\int_0^t F(\mathbf{y}^{(N)}(s)) - g(\mathbf{y}^{(N)}(s))ds = \sum_{i=0}^{\lfloor \Lambda N t \rfloor - 1} \frac{1}{\Lambda N}(F(\mathbf{z}^{(N)}(i)) - g(\mathbf{z}^{(N)}(i))).$$

Therefore, while computing the value of $\int_0^t F(\mathbf{y}^{(N)}(s)) - g(\mathbf{y}^{(N)}(s))ds$ along $\mathbf{z}^{(N)}$, we will alternatively add and subtract the vector $\frac{1}{\Lambda N}(f_1(\mathbf{x}) - f_2(\mathbf{x}))$ (multiplied by the relative weights), so that the norm that we accumulate along the integral will remain bounded by a vanishing constant.

We turn now this intuitive argument into a formal one. First of all, we need to investigate the behaviour of $\mathbf{z}^{(N)}$ with respect to the surface $\mathcal{H}$, in order to show that $\mathbf{z}^{(N)}$ jumps from one side to the other, and that its distance from the surface is bounded by a constant vanishing with N. In order to measure the distance of $\mathbf{z}^{(N)}$ from $\mathcal{H}$, we consider its projection $\bar{\mathbf{z}}^{(N)}(k)$ along the normal vector n of $\mathcal{H}$: $\bar{\mathbf{z}}^{(N)}(k) = n^T(\mathbf{z}^{(N)}(k))\mathbf{z}^{(N)}(k)$. Due to the choice of the initial conditions, $\bar{\mathbf{z}}^{(N)}(0) = 0$.

To investigate the behaviour of $\bar{\mathbf{z}}^{(N)}$, suppose $\mathbf{z}^{(N)}(k)$ is a point in $\mathcal{R}_1$ close to the surface $\mathcal{H}$. Now, by the conditions of sliding motion, $f_1^{\perp}(\mathbf{z}^{(N)}(k)) := n^T(\mathbf{z}^{(N)}(k))f_1(\mathbf{z}^{(N)}(k)) < 0$, hence the value of $\bar{\mathbf{z}}^{(N)}(k+1)$ will keep decreasing until $\mathbf{z}^{(N)}$ enters the region $\mathcal{R}_2$. At this stage the value of $\bar{\mathbf{z}}^{(N)}$ will increase until $\mathbf{z}^{(N)}$ enters again region $\mathcal{R}_1$. Clearly, as long as we take steps within the same region, the absolute value of $\bar{\mathbf{z}}^{(N)}$ cannot increase. In fact, it can increase only during steps in which the sequence $\mathbf{z}^{(N)}$ changes region. But after these steps, $|\bar{\mathbf{z}}^{(N)}| \leq \frac{M}{\Lambda N}$, where the right hand side is an upper bound on the increase of a single step, with $M = \max\{\sup_{\mathbf{x}\in E} f_1(\mathbf{x}), \sup_{\mathbf{x}\in E} f_2(\mathbf{x})\}$. Now, let k be a step in which $\mathbf{z}^{(N)}$ just entered $\mathcal{R}_1$ ($k = 0$ will also do, as we start in $\mathcal{H}$, in which the vector field is f_1).Then, consider a consecutive sequence of $l = l_1 + l_2$ steps, such that at step $k + l_1$ we leave $\mathcal{R}_1$ to enter $\mathcal{R}_2$ and at step $k + l_1 + l_2$ we leave $\mathcal{R}_2$ and enter back again in $\mathcal{R}_1$. We call this sequence a *cycle*. Assuming $|\bar{\mathbf{z}}^{(N)}(k)| \leq \frac{M}{\Lambda N}$ (a condition holding for $k = 0$), then also $|\bar{\mathbf{z}}^{(N)}(k + j)| \leq \frac{M}{\Lambda N}$, for each $j \leq l$, i.e. $|\bar{\mathbf{z}}^{(N)}(k)|$ remains bounded by $\frac{M}{\Lambda N}$ in a cycle. Reasoning inductively, it immediately follows that $\sup_{k\leq\lfloor \Lambda N T\rfloor} |\bar{\mathbf{z}}^{(N)}(k)| \leq \frac{M}{\Lambda N}$, proving that $\sup_{k\leq\lfloor \Lambda N T\rfloor} |\bar{\mathbf{z}}^{(N)}(k)| \to 0$. This means that $\mathbf{z}^{(N)}$ (hence $\mathbf{y}^{(N)}$) is essentially stuck to the surface $\mathcal{H}$ as N grows to infinity.

Notice that the length of a cycle starting in a given point $\mathbf{x} \in \mathcal{R}_1$ is determined only by the value of $f_1^{\perp}(\mathbf{x})$ and $f_2^{\perp}(\mathbf{x})$, if N is large enough[3]. This is true because the variation of functions $f_i^{\perp}$ in a cycle becomes arbitrary small for N large. In particular, the length of a cycle starting at k depends on the relative magnitude of $|f_1^{\perp}(\mathbf{z}^{(N)}(k))|$ and $|f_2^{\perp}(\mathbf{z}^{(N)}(k))|$, i.e. it is bounded by $\max\left\{\frac{|f_1^{\perp}(\mathbf{z}^{(N)}(k))|}{|f_2^{\perp}(\mathbf{z}^{(N)}(k))|}, \frac{|f_1^{\perp}(\mathbf{z}^{(N)}(k))|}{|f_2^{\perp}(\mathbf{z}^{(N)}(k))|}\right\} + 2$ (for N large). In fact, if $|f_1^{\perp}(\mathbf{z}^{(N)}(k))| > |f_2^{\perp}(\mathbf{z}^{(N)}(k))|$, then for large N we will need just one step to jump to $\mathcal{R}_2$, and one or more steps to return to $\mathcal{R}_1$, but no more than $\left\lceil\frac{|f_1^{\perp}(\mathbf{z}^{(N)}(k))|}{|f_2^{\perp}(\mathbf{z}^{(N)}(k))|}\right\rceil$.

[3]Recall that in our setting, functions f_i are defined on the whole state space E, so that the quantities $f_i^{\perp,0}$ are defined. Actually, as $\bar{\mathbf{z}}^{(N)} \to 0$, we just need functions f_i to be defined in an open neighborhood of $\mathcal{R}_i$.

We finally deal with the full integral. Consider the sequence $\tilde{\mathbf{z}}^{(N)}(k+1) = \tilde{\mathbf{z}}^{(N)}(k) + \frac{1}{\Lambda N}(F(\mathbf{z}^{(N)}(k)) - g(\mathbf{z}^{(N)}(k)))$. We need to prove that $\|\tilde{\mathbf{z}}^{(N)}(k)\|$ is uniformly bounded by a vanishing constant, for $k \leq \lfloor \Lambda NT \rfloor$.

In order to do this, consider the value of $\tilde{\mathbf{z}}_i^{(N)}$ during a cycle starting at step k (cycles are defined as before, i.e. according to $\bar{\mathbf{z}}^{(N)}$). In such a cycle, letting $\mu(\mathbf{x}) = \frac{f_2(\mathbf{x}) - f_1(\mathbf{x})}{f_2^{\perp}(\mathbf{x}) - f_1^{\perp}(\mathbf{x})}$, $\mu^j = \mu(\mathbf{z}^{(N)}(k+j))$, and $f_i^{\perp,j} = f_i^{\perp}(\mathbf{z}^{(N)}(k+j))$, and applying the definitions of $F-g$ and λ, we have

$$\tilde{\mathbf{z}}^{(N)}(k+s) = \tilde{\mathbf{z}}^{(N)}(k) + \frac{1}{\Lambda N}\left(\sum_{j=0, j\leq s-1}^{l_1-1} f_1^{\perp,j}\mu^j + \sum_{j=l_1}^{s-1} f_2^{\perp,j}\mu^j\right),$$

Now, if we approximate the value of μ at step $k+j$ with the value of μ at step k, it holds that

$$\begin{aligned}
\left\|\tilde{\mathbf{z}}^{(N)}(k+s)\right\| &\leq \left\|\tilde{\mathbf{z}}^{(N)}(k) + \frac{\mu^0}{\Lambda N}\left(\sum_{j=0, j\leq s-1}^{l_1-1} f_1^{\perp,j} + \sum_{j=l_1}^{s-1} f_2^{\perp,j}\right)\right\| \\
&+ \left\|\frac{1}{\Lambda N}\left(\sum_{j=0, j\leq s-1}^{l_1-1} f_1^{\perp,j}(\mu^j - \mu^0) + \sum_{j=l_1}^{s-1} f_2^{\perp,j}(\mu^j - \mu^0)\right)\right\| \\
&\leq \left\|\tilde{\mathbf{z}}^{(N)}(k) + \frac{\mu^0}{\Lambda N}\left(\sum_{j=0, j\leq s-1}^{l_1-1} f_1^{\perp,j} + \sum_{j=l_1}^{s-1} f_2^{\perp,j}\right)\right\| + \frac{LM^2 l^2}{(\Lambda N)^2},
\end{aligned}$$

where L is the Lipschitz constant of the function $\mu(\mathbf{x})$,[4] M is the upper bound on f_1 and f_2 on E, l is the cycle length. In the previous inequality, we used the fact that $|\mu^j - \mu^0| \leq L|\mathbf{z}^{(N)}(k+j) - \mathbf{z}^{(N)}(k)| \leq \frac{lLM}{\Lambda N}$, as $|\mathbf{z}^{(N)}(k+1) - \mathbf{z}^{(N)}(k)| = |\frac{1}{\Lambda N}F(\mathbf{z}^{(N)}(k))| \leq \frac{M}{\Lambda N}$.

Therefore, along the direction of the vector μ^0, we are essentially replicating the behaviour of $\bar{\mathbf{z}}^{(N)}$, a part from a multiplicative factor (the norm of μ^0), which is bounded by a constant $M' < \infty$. Hence, reasoning as for $\bar{\mathbf{z}}^{(N)}$, it holds that $\|\tilde{\mathbf{z}}^{(N)}(k)\|$ is bounded for all k by $\frac{MM'}{\Lambda N}$, plus the error that we commit in approximating μ^j by μ^0, which adds up at a rate of $\frac{LM^2l^2}{(\Lambda N)^2}$ per cycle. Adding this term on all cycles, observing that $\sum_{cycles} l = \Lambda NT$, and letting $\hat{l}_N$ be the maximum cycle length for $\mathbf{z}^{(N)}$, we obtain that

$$\sup_{k\leq\lfloor \Lambda NT\rfloor} \|\tilde{\mathbf{z}}^{(N)}(k)\| \leq \frac{MM'}{\Lambda N} + \frac{LM^2T\hat{l}_N}{\Lambda N}.$$

If $\hat{l}_N$, the upper bound on a cycle length, is bounded by $\hat{l}$ independently of N, then we are done, as the quantity on the right hand side converges to zero. However, we must deal also with the case in which $T = T_S$. In $\mathbf{x}(T_S)$ one of the two functions $f_1^{\perp}$ and $f_2^{\perp}$ equals zero (due to first order exit conditions of the sliding motion), so that $\max\left\{\frac{|f_1^{\perp}(\mathbf{x})|}{|f_2^{\perp}(\mathbf{x})|}, \frac{|f_1^{\perp}(\mathbf{x})|}{|f_2^{\perp}(\mathbf{x})|}\right\}$, and hence $\hat{l}_N$, is unbounded while approaching T_S (along $\mathbf{x}(t)$). Fix $\varepsilon > 0$, then there is a $\delta_\varepsilon > 0$ such that the functions $f_1^{\perp}(\mathbf{y}^{(N)}(t))$ and $f_2^{\perp}(\mathbf{y}^{(N)}(t))$ are greater than $\delta_\varepsilon > 0$

[4] This function is bounded and Lipschitz continuous in a neighborhood of the solution $\mathbf{x}(t)$, because $f_2^{\perp}(\mathbf{x}) - f_1^{\perp}(\mathbf{x})$ is Lipschitz and greater than a constant $\delta > 0$ in a neighborhood of the solution $\mathbf{x}(t)$, and we can always restrict to such a neighborhood, cf. the argument of footnote 5.

for $t \leq T - \varepsilon$ and $N \geq N_\varepsilon$, [5] and so $\hat{l}_N \leq \hat{l}_\varepsilon < \infty$ is bounded independently from N, at least up to time $t \leq T - \varepsilon$. In the last ε units of time, however, $\frac{l}{\Lambda N} \leq \varepsilon$, hence $\|\tilde{\mathbf{z}}^{(N)}(k)\| \leq \frac{MM'}{\Lambda N} + \frac{LM^2T\hat{l}_\varepsilon}{\Lambda N} + LM^2T\varepsilon$, meaning that $\lim_{N\to\infty} \sup_{k\leq\lfloor \Lambda NT \rfloor} \|\tilde{\mathbf{z}}^{(N)}(k)\| \leq \varepsilon'$ for $\varepsilon' = LM^2T\varepsilon > 0$. For the arbitrariness of ε, $\sup_{k\leq\lfloor \Lambda NT\rfloor} \|\tilde{\mathbf{z}}^{(N)}(k)\| \to 0$. ∎

This lemma shows that the sequence of stochastic processes is essentially confined to the surface $\mathcal{H}$. In fact, the sequence behaves, in the limit, like the solution of an integral equation defined by the sliding vector field. Intuitively, the CTMC keeps jumping from one side to the other of the surface $\mathcal{H}$, with jumps becoming smaller and more frequent as N grows. Hence, it keeps following either the vector field f_1 or the vector field f_2. As N grows, the probability of being on the R_1 side of the surface $\mathcal{H}$ converges to $\frac{f_2^\perp(\mathbf{x})}{f_2^\perp(x) - f_1^\perp(\mathbf{x})}$, which is exactly the definition of the weights of the sliding vector field, obtained by geometric reasoning.

Main theorem

We are finally ready to prove the main theorem of the paper:

Theorem .2. *Let the sequence $\hat{\mathcal{X}}^{(N)}$ of CTMC models satisfy the scaling assumptions of Section 2.2, and suppose $\hat{\mathbf{X}}^{(N)}(0) \to \mathbf{x}_0$. Fix a set $S \subseteq E$ and let $\mathbf{x}(t)$ be the solution of the PWS system $\frac{d}{dt}\mathbf{x} = F(\mathbf{x})$ starting in point $\mathbf{x}(0) = \mathbf{x}_0 \in S$. Fix a finite time horizon $T < \zeta(S)$, where $\zeta(S)$ is the exit time of $\mathbf{x}(t)$ from S, and assume that $\mathbf{x}(t)$ is a regular trajectory in $[0, \zeta(S)]$. Then*

$$\lim_{N\to\infty} \sup_{t\leq T} \left\| \hat{\mathbf{X}}^{(N)}(t) - \mathbf{x}(t) \right\| = 0 \quad \textit{in probability.}$$

Proof. Consider the trajectory $\mathbf{x}(t)$ and the times $T_1 < \ldots < T_k$ in which it switches vector field, i.e. the times in which the trajectory hits a discontinuity surface, coming from the interior of a region, or in which the sliding motion terminates. Let $T_0 = 0$ and $T_{k+1} = T$. The idea to prove the theorem is to apply in an interval $[T_i, T_{i+1}]$ either Theorem 2.1, if the solution evolves in the interior of a region $\mathcal{R}_j$, or Theorem 4.2, in case of sliding motion. The only technical detail we have to deal with is that, in order to apply Theorem 2.1, we have to restrict our attention to the interior of the region $\mathcal{R}_j$, where F is Lipschitz. Hence, we need to use points 3 and 4 of Theorem 2.1, taking care explicitly of exit times. Consider the time interval $[0, T_1]$, and suppose that during this time interval the solution $\mathbf{x}(t)$ is within a region $\mathcal{R}_j$, and hence follows the solution of a Lipschitz ODE. Let $\zeta_1^{(N)} = \zeta_1^{(N)}(\mathcal{R}_j)$ be the exit time of $\hat{\mathbf{X}}^{(N)}(t)$ from $\mathcal{R}_j$, and apply Theorem 2.1 to prove that $\sup_{t\leq T_1} \|\hat{\mathbf{X}}^{(N)}(\min\{\zeta_1^{(N)}, t\}) - \mathbf{x}(t)\| \to 0$ in probability and $\zeta_1^{(N)} \to T_1$ in probability. In case we are dealing with sliding motion and not with motion in the interior of a region where the vector field is Lipschitz, we can ignore such a stopping time, simply setting $\zeta_1^{(N)} = T_1$.

[5] More precisely, it holds that the functions $f_1^\perp(\mathbf{x}), f_2^\perp(\mathbf{x}) \geq \delta'_\varepsilon > 0$ along $\mathbf{x}(t)$, for $t \leq T - \varepsilon$. Suppose now that, along $\mathbf{y}^{(N)}(t)$, one of the two functions $f_1^\perp$ and $f_2^\perp$, say $f_2^\perp$, approaches 0 for $\bar{t} < T - \varepsilon$ and for $N \geq N'_\varepsilon$. Then, by continuity of $f_1^\perp$ and $f_2^\perp$ and by the fact that $f_1^\perp(\mathbf{y}^{(N)}(0)) < 0$ and $f_2^\perp(\mathbf{y}^{(N)}(0)) > 0$, there is a time t^* such that $\delta'_\varepsilon > f_2^\perp(\mathbf{y}^{(N)}(t^*)) > \frac{\delta'_\varepsilon}{2}$ and $|f_1^\perp(\mathbf{y}^{(N)}(t))|, |f_2^\perp(\mathbf{y}^{(N)}(t))| > \frac{\delta'_\varepsilon}{2}$, for each $t \leq t^*$ (and each N sufficiently large). Therefore, the maximum cycle length is bounded and independent of N up to t^*. Furthermore, $\mathbf{y}^{(N)}(t^*)$ does not converge to $\mathbf{x}(t^*)$, as $f_2^\perp(\mathbf{y}^{(N)}(t^*)) < \delta'_\varepsilon$ and $f_2^\perp(\mathbf{x}(t^*)) \geq \delta'_\varepsilon$. However, as the maximum cycle length is bounded and independent of N for $t \leq t^*$, the lemma and, *a fortiori*, Theorem 4.2 hold up to time t^*. Therefore, using Lemma .2, $\mathbf{y}^{(N)}(t^*) \to \mathbf{x}(t^*)$, a contradiction. It follows that $f_1^\perp$ and $f_2^\perp$ are definitively bounded away from zero along $\mathbf{y}^{(N)}(t)$, for each $t \leq T - \varepsilon$.

Consider the stopping time $\bar{\zeta}_1^{(N)} = \min\{\zeta_1^{(N)}, T_1\}$: Clearly, $\bar{\zeta}_1^{(N)} \to T_1$ and $\hat{\mathbf{X}}^{(N)}(\bar{\zeta}_1^{(N)}) \to \mathbf{x}(T_1)$ in probability (also in case of sliding motion). Now, stop the process $\hat{\mathbf{X}}^{(N)}(t)$ at time $\bar{\zeta}_1^{(N)}$ and restart it. The restarted process $\hat{\mathbf{X}}_1^{(N)}(t) = \hat{\mathbf{X}}^{(N)}(\bar{\zeta}_1^{(N)} + t)$ is again a Markov process due to the Strong Markov Property [Nor97]. Letting $\mathbf{x}_1(t) = \mathbf{x}(T_1 + t)$, we can use Theorem 2.1 or Theorem 4.2 to prove that $\sup_{t \leq T_2 - T_1} \|\hat{\mathbf{X}}_1^{(N)}(\min\{\zeta_2^{(N)}, t\}) - \mathbf{x}_1(t)\| \to 0$ in probability. Iterating this reasoning and defining restarted processes $\hat{\mathbf{X}}_i^{(N)}$ and $\mathbf{x}_i$ in a similar way, we can prove that, for each $i \geq 0$, $\sup_{t \leq T_{i+1} - T_i} \|\hat{\mathbf{X}}_i^{(N)}(\min\{\zeta_{i+1}^{(N)}, t\}) - \mathbf{x}_i(t)\| \to 0$ in probability.[6] Furthermore, as $\bar{\zeta}_i^{(N)} \to T_i$ and $\min\{\zeta_{i+1}^{(N)}, t\} \to T_{i+1}$ in probability, then both $\sup_{t \leq T_{i+1} - T_i} \|\hat{\mathbf{X}}^{(N)}(T_i + t) - \mathbf{X}^{(N)}(\bar{\zeta}_i^{(N)} + t)\|$ and $\sup_{t \leq T_{i+1} - T_i} \|\hat{\mathbf{X}}_i^{(N)}(t) - \mathbf{X}_i^{(N)}(\min\{\zeta_{i+1}^{(N)}, t\})\|$ converge to zero in probability. [7] Hence,

$$
\begin{aligned}
\sup_{t \leq T} \|\hat{\mathbf{X}}^{(N)}(t) - \mathbf{x}(t)\| \quad \leq \quad & \max_{i=0,\ldots k} \left\{ \sup_{t \leq T_{i+1} - T_i} \|\hat{\mathbf{X}}^{(N)}(T_i + t) - \hat{\mathbf{X}}^{(N)}(\bar{\zeta}_i^{(N)} + t)\| \right. \\
& + \sup_{t \leq T_{i+1} - T_i} \|\hat{\mathbf{X}}_i^{(N)}(t) - \hat{\mathbf{X}}_i^{(N)}(\min\{\zeta_{i+1}^{(N)}, t\})\| \\
& \left. + \sup_{t \leq T_{i+1} - T_i} \|\hat{\mathbf{X}}_i^{(N)}(\min\{\zeta_{i+1}^{(N)}, t\}) - \mathbf{x}_i(t)\| \right\},
\end{aligned}
$$

and so $\sup_{t \leq T} \|\hat{\mathbf{X}}^{(N)}(t) - \mathbf{x}(t)\| \to 0$ in probability. ■

Remark .1. The hypothesis on boundedness and Lipschitzness of functions f_i can be relaxed to local boundedness and local Lipschitzness (i.e. functions f_i must be bounded and Lipschitz in a neighbourhood of each point). In fact, if the trajectory $\mathbf{x}(t)$ do not diverge in time $[0, T]$, we can always find a compact set S, with non-empty interior, containing the trajectory. Boundedness and Lipschitzness hold globally in this set.

[6] We are assuming $\bar{\zeta}_0^{(N)} = 0$.

[7] The first fact is proved as follows. Fix $\varepsilon > 0$, then $|T_i - \bar{\zeta}_i^{(N)}| < \varepsilon$ with probability $1 - \delta_N$, $\delta_N \to 0$. Now, in ε units of time, the difference between $\hat{\mathbf{X}}^{(N)}(t)$ and $\hat{\mathbf{X}}^{(N)}(t + \varepsilon)$ is bounded by $\frac{v}{N}\mathcal{N}(\Lambda N \varepsilon)$, where $v = \max\{\|\mathbf{v}_i\|\}$, $\mathcal{N}$ is a Poisson random variable and Λ is defined as in Lemma .2. Now, $\frac{\mathcal{N}(\Lambda N \varepsilon)}{N}$ converges almost surely to $\Lambda\varepsilon$ (this is the law of large numbers), and so $\sup_{t \leq T_{i+1} - T_i} \|\hat{\mathbf{X}}^{(N)}(T_i + t) - \mathbf{X}^{(N)}(\bar{\zeta}_i^{(N)} + t)\| \leq 2v\Lambda\varepsilon$ for N large enough (with probability $1 - \delta_N$).