Markov chains with discontinuous drifts have differential inclusions limits. Application to stochastic stability and mean field approximation.

Nicolas Gast, Bruno Gaujal

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Markov chains with discontinuous drifts have differential inclusions limits

Nicolas Gast — Bruno Gaujal

N° 7315 — version 4

initial version June 2010 — revised version March, 2012

Thème NUM

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--7315--FR+ENG

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

INRIA

centre de recherche
GRENOBLE - RHÔNE-ALPES

Markov chains with discontinuous drifts have differential inclusions limits

Nicolas Gast, Bruno Gaujal

Thème NUM — Systèmes numériques
Équipe-Projet MESCAL

Rapport de recherche n° 7315 — version 4 — initial version June 2010 — revised version March, 2012 — 29 pages

Abstract: In this paper, we study deterministic limits of Markov processes having discontinuous drifts. While most results assume that the limiting dynamics is continuous, we show that these conditions are not necessary to prove convergence to a deterministic system. More precisely, we show that under mild assumptions, the stochastic system is a stochastic approximation algorithm with constant step size that converges to a differential inclusion. This differential inclusion is obtained by convexifying the rescaled drift of the Markov chain.

This generic convergence result is used to compute stability conditions of stochastic systems, via their fluid limits. It is also used to analyze systems where discontinuous dynamics arise naturally, such as queueing systems with boundary conditions or with threshold control policies, via mean field approximations.

Key-words: Mean field, fluid limit, stability, differential inclusion, non-smooth dynamics, queueing systems.

Centre de recherche INRIA Grenoble – Rhône-Alpes
655, avenue de l'Europe, 38334 Montbonnot Saint Ismier
Téléphone : +33 4 76 61 52 00 — Télécopie +33 4 76 61 52 52

Les chaînes de Markov ayant une dérive discontinue ont pour limite une inclusion différentielle.

Résumé : Ce document étudie des limites d'échelle de chaînes de Markov ayant une dérive discontinue. Alors que beaucoup de travaux sur le sujet supposent que la dérive est continue, nous montrons que cette condition n'est pas nécessaire pour obtenir une limite déterministe.

Nous montrons que sous des hypothèses faibles, un passage à l'échelle d'une chaîne de Markov peut être vu comme un algorithme d'approximation stochastique à pas constant, qui converge vers l'ensemble des solutions d'une inclusion différentielle. Cette inclusion est obtenue à partir d'une convexification de la dérive du processus initial.

Cette méthode est générique et permet de calculer la région de stabilité de nombreux systèmes stochastique, en étudiant leur limite fluide. Elle permet aussi d'étendre les techniques d'approximation champ moyen à des systèmes où les discontinuités apparaissent naturellement, comme des réseaux de files d'attente ou des systèmes contrôlés par une politique à seuil.

Mots-clés : Champ Moyen, Limite fluides, stabilité, Inclusions différentielles, dynamiques non-continues, réseaux de files d'attente.

1 Introduction

The use of ordinary differential equations has proved useful for performance evaluation of computing systems and communication networks. Here are a few striking examples: Fluid limits have been used to prove stability of a large class of queuing systems [12, 11]; The performance of the wifi protocol 802.11b has been analyzed using a mean field approximation in [6, 7] and distributed algorithms such as work stealing [27, 18] have also been studied using the famed population dynamics approach introduced by Kurtz [23].

In this paper, we show that both scalings (fluid limit and mean field) can be studied within a common framework, by seeing a Markovian stochastic system as a *stochastic approximation* of a deterministic differential system driven by the rescaled *drift* of the initial system. Under classical smoothness assumptions on the drift, there exist general results that show that the limiting system (when the scaling parameter goes to infinity) can be described by a system of deterministic ordinary differential equations

$$\dot{y}(t) = f(y(t)). \tag{1}$$

See [23, 4] and the references therein for examples of such convergence results. In most cases, the limiting drift function f in (1) is assumed to have a Lipschitz property. This strong condition restricts the applicability of these results in many practical cases, in particular, for systems exhibiting threshold dynamics or with boundary conditions.

The purpose of this paper is to study the limiting behavior of such a system when the drift f is not continuous. Let us consider a simple queuing system with one buffer and N processors that can serve two packets each, per unit of time, on average. If packets arrive at rate N , and if y denotes the number of packets in the queue, then the average decrease of y is one packet per unit of time under a proper rescaling of time) if the queue is non-empty (*i.e.* $y > 0$) and the average increase is one if the queue is empty. This leads to a deterministic limit behavior:

$$\dot{y}(t) = -1 \text{ if } y(t) > 0 \quad \text{and} \quad \dot{y}(t) = 1 \text{ if } y(t) = 0. \tag{2}$$

This dynamics is not continuous and therefore not Lipschitz which makes the classical approach inapplicable in this case.

In the case of a non-continuous right-hand side, the differential equation (2) is not well-defined since there exists no function y that is differentiable and that satisfies (2). The proper way to define solutions of (2) is to use *differential inclusions* (DI) instead. Equation (1) is replaced by the following equation

$$\dot{y}(t) \in F(y(t)), \tag{3}$$

where F is a set-valued mapping, defined as the convex hull of the accumulation points of the drift. In the above example, if $y \neq 0$ then $F(y) = \{-1\}$ and $F(0) = [-1, 1]$. Of course a differential inclusion problem may (or may not) have multiple solutions. The main result of the paper is that over any finite time interval, the trajectory of the initial system converges to one element in the set of the solutions of the differential inclusion, when the scaling parameter N goes to infinity, (Theorem 1). This result is rather general and does not require any Lipschitz property on the function F. In particular, it shows that when (3) has a unique solution, the behavior of the system converges to it. Moreover, we also show that when F satisfies a one-sided Lipschitz condition (7), we can bound the difference with the limiting dynamics with explicit bounds (Theorem 4).

This generic result is put to practice in several applications. First (in Section 3), we show how it can be used to compute the fluid limit of a system and to provide sufficient conditions for the stability of the system. Several well-known papers have established that the stability of the fluid limit implies the stability of the initial stochastic system, and this result has been used extensively to prove stability of many stochastic systems where the main difficulty has often been to characterize the fluid limit. Our approach has two advantages: It provides a generic way of constructing the limit even with non-continuous drifts, and this construction is explicit enough so that it can be used to give stability conditions in closed form. We illustrate this by establishing the stability condition of opportunistic scheduling policies whose original proofs are rather involved.

The second application concerns mean field limits (in Section 4). We show that a stochastic system composed of N indistinguishable objects with a non-continuous drift can be seen as a stochastic approximation of a differential inclusion. This result is used to compute the mean field approximation of several systems that could not be studied this way before. We illustrate this by two examples where discontinuities arise naturally: a parallel server system in which a centralized controller tries to improve load balancing and a volunteer computing system with boundary constraints.

2 Stochastic Approximations and differential inclusions

This section presents a generic result that will be used as the methodological basis for the rest of the paper. We first show that a family of Markov chains with a vanishing drift can be seen as a stochastic approximation with a constant step size of a differential inclusion and we state the main convergence result (§2.1). A more precise convergence result is established when the limit differential inclusion has the one-sided Lipschitz property (§2.2). The result is also extended to the important case of continuous time Markov chains (§2.3).

2.1 Construction of the stochastic approximation algorithm and main result

Let us consider a *discrete time Markov chain* $Y^N(k)$ with values in $\mathbb{R}^d$. The index N is used to denote some scaling parameter of the chain and will have a clear meaning in applications (for example N could be the number of objects forming the system). The expected difference between $Y^N(k+1)$ and $Y^N(k)$ is called the *drift* of the chain and is denoted g^N:

$$g^N(y) \stackrel{\text{def}}{=} \mathbb{E}\left(Y^N(k+1) - Y^N(k) | Y^N(k) = y\right)$$

The main feature of the chains studied here, concerns their scaling with N. This translates as one essential assumption on the drift: we assume that the drift vanishes at speed γ^N as N goes to infinity. More precisely, this means that we assume that there exists a sequence γ^N, called the *intensity* of the chain, such that $\lim_{N\to\infty} \gamma^N = 0$ and such that for all y: $\left\|g^N(y)\right\| \leq c(1+\|y\|)\cdot\gamma^N$, for some constant c. We also denote by $f^N(y)$ the drift rescaled by γ^N:

$$f^N(y) \stackrel{\text{def}}{=} \frac{g^N(y)}{\gamma^N}.$$

Using these definitions, one can write the evolution of the Markov chain $Y^N(k)$ as a *stochastic approximation* algorithm with constant step size γ^N:

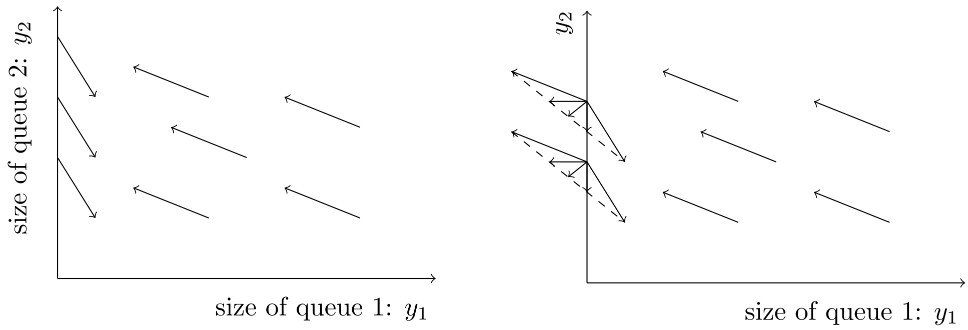
$$Y^N(k+1) = Y^N(k) + \gamma^N\Big(f^N(Y^N(k)) + U^N(k+1)\Big), \tag{4}$$

where $U^N(k+1) \stackrel{\text{def}}{=} \left(Y^N(k+1) - Y^N(k))\right)/\gamma^N - f^N(Y^N(k))$ is a zero mean process that captures the random innovation of the chain between steps k and $k+1$. U^N is a martingale difference sequence with respect to the filtration $\mathcal{F}_k$ associated with the process $Y^N(k)$. In particular, it has zero mean conditionally to $Y^N(k)$: by the Markov property, $\mathbb{E}(U^N(k+1) \mid Y^N(k)) = \mathbb{E}(U^N(k+1) \mid \mathcal{F}_k) = 0$.

Under mild conditions on U^N, when f^N converges uniformly to a Lipschitz continuous function f, the behavior of $Y^N(t/\gamma^N)$ is known to converge to the solution of an ODE $dy/dt = f(y)$ as N goes to infinity. However, when f is not continuous, this result does not hold and this differential system cannot be defined properly. To deal with the general case, we introduce a set-valued function F to replace f and the ODE is replaced by the *differential inclusion* $dy/dt \in F(y)$ (see A for a brief introduction on differential inclusions). The set-valued function F associated with the rescaled drift f^N, at point y, is defined as the convex closure of the set of the accumulation points of $f^N(y^N)$ as N goes to infinity, for all sequences y^N converging to y:

$$F(y) \stackrel{\text{def}}{=} \text{conv}\left(\left\{\underset{N\to\infty}{\text{acc}} f^N(y^N) \text{ for all sequences } y^N \underset{N\to\infty}{\longrightarrow} y\right\}\right). \tag{5}$$

where $\mathrm{acc}_{N\to\infty} x^N$ denotes the set of accumulation points of the sequence x^N as N goes to infinity and $\mathrm{conv}(A)$ is the convex hull of set A. The construction of F from f^N is illustrated in Figure 1 in an example in $\mathbb{R}^2$.



(a) Drift of the system of size N: f^N

(b) Corresponding set-valued drift F

Figure 1: Example of construction of the set-valued function F from the non-continuous drift f^N. This example is taken from the fluid limit of two queues with priority, developed in Section 3.2 with parameters $\lambda_1 = \lambda_2 = 0.1$ and $\mu_1 = \mu_2 = .3$. For all N, $f^N(y_1, y_2) = (-0.2, +0.1)$ if $y_1 > 0$ and $f(0, y_2) = (+0.1, 0.2)$ if $y_2 > 0$. Therefore, $f^N(y)$ is independent of N and is discontinuous in $y_1 = 0$. Since f^N is continuous for $y_1 > 0$, one has $F(y_1, y_2) = \{(-0.2, +0.1)\}$ for $y_1 > 0$. When $y_1 = 0$, $F(0, y_2)$ is the convex closure of $(-0.2, +0.1)$ and $(+0.1, -0.2)$.

We are now ready to state the main theorem of this section. Let us define the continuous function $\bar{Y}^N(t)$ as the piecewise-linear interpolation of $\{Y^N(k)\}_{k\in\mathbb{N}}$ whose time has been accelerated by $1/\gamma^N$: for all $k \in \mathbb{N}$, $\bar{Y}^N(k \cdot \gamma^N) = Y^N(k)$ and $\bar{Y}^N$ is linear on $[k\gamma^N, (k+1)\gamma^N]$. Let us denote by $\mathscr{S}_T(y_0)$ the set of the solutions of the differential inclusion (DI)

$$\dot{y}(t) \in F(y(t)), \quad y(0) = y_0, \tag{6}$$

where a solution of the DI (6) is an absolutely continuous function y such that $\dot{y}(t) \in F(y(t))$ almost everywhere. Starting from y_0, we can show that $\bar{Y}^N(.)$ converges (in probability) to $\mathscr{S}_T(y_0)$. More precisely, the following theorem holds.

Theorem 1. *Let $Y^N(\cdot)$ be a Markov process on $\mathbb{R}^d$ satisfying (4). Assume that*

- *The drift g^N vanishes with speed γ^N: there exists a sequence γ^N and a constant c such that*

$$\lim_{N\to\infty} \gamma^N = 0 \quad \text{and} \quad \forall y \in \mathbb{R}^d : \left\|f^N(y)\right\| \stackrel{\text{def}}{=} \left\|\frac{g^N(y)}{\gamma^N}\right\| \le c(1 + \|y\|).$$

- *U^N is a martingale difference sequence which is uniformly integrable[1]:*

$$\mathbb{E}\left(U^N(k+1) \mid Y^N(k)\right) = 0 \quad \text{and} \quad \lim_{R\to\infty} \sup_k \mathbb{E}\left(\left\|U^N(k+1)\right\| \mathbf{1}_{\|U^N(k+1) \ge R\|} \mid Y^N(k)\right) = 0.$$

If $Y^N(0) \xrightarrow{\mathcal{P}} y_0$ (convergence in probability), then for all $T > 0$:

$$\inf_{y\in\mathscr{S}_T(y_0)} \sup_{0\le t\le T} \left\|\bar{Y}^N(t) - y(t)\right\| \xrightarrow{\mathcal{P}} 0.$$

where $\mathscr{S}_T(y_0)$ is the set of solutions of the DI (6) and F is defined by (5).

[1]The uniform integrability can be obtained by assuming that the second moment is bounded: if there exists b such that $\mathbb{E}(\left\|U^N(k+1)\right\|^2 \mid Y^N(k)) \le b < \infty$ for all k, then U^N is uniformly integrable.

Proof. The proof is given in B.1. □

This theorem shows that if N is large enough, the trajectory of the stochastic system $\bar{Y}^N$ on T/γ^N steps is close to a solution of the differential inclusion (6) over $[0,T]$. This theorem does not assume any regularity condition on the drift function f^N and only requires that the drift vanishes as N grows. In particular, it does not assume that f^N converges uniformly to a function f. It also provides a constructive definition of the set-valued drift F.

The price to be paid for this generality is that a differential inclusion may have multiple solutions. In that case, $\bar{Y}^N$ may converge to any solution of the DI, depending on its random innovations, making this result rather inefficient for performance evaluation. This result is of greater interest if the DI starting from y_0 has a unique solution: $\mathscr{S}_T(y_0)=\{y\}$. In that case, as a direct corollary of the preceding result, $\bar{Y}^N$ converges in probability to y.

Corollary 2. *Under the conditions of Theorem 1 and if the DI (6) has a unique solution y on $[0;T]$:*

$$\sup_{0\le t\le T}\left\|\bar{Y}^N(t)-y(t)\right\| \xrightarrow{\mathcal{P}} 0.$$

In many cases, the limiting differential inclusion clearly has a unique solution which makes the preceding corollary directly applicable. In particular, this is the case for all the examples presented in this paper except for the example §3.4.

2.2 Speed of convergence under OSL condition

The main drawback of the previous theorem is that it does not give any insight in the speed of convergence of the stochastic system toward its limit. In fact, without further conditions, the convergence may be arbitrarily slow. This limitation can be overcome when the function F satisfies the one-sided Lipschitz (OSL) condition.

2.2.1 The one-sided Lipschitz (OSL) condition

A set-valued function F is said to be OSL if there exists a constant L such that for all points $y,y'\in\mathbb{R}^d$ and $z\in F(y), z'\in F(y')$:

$$\langle y-y', z-z'\rangle \le L\left\|y-y'\right\|^2. \tag{7}$$

where $\langle x,y\rangle$ denotes the classical inner product on $\mathbb{R}^d$. OSL conditions are commonly assumed in the non-smooth system literature [10, 24], mainly because it ensures the uniqueness of the solution of a DI $\dot{y}(t)\in F(y(t)), y(0)=y_0$.

It should be clear that if F is a single-valued Lipschitz function of constant L, F is also OSL with constant L. The term *one-sided Lipschitz* comes from the fact that a Lipschitz function F would satisfy $-L\left\|y-y'\right\|^2\le\langle y-y',z-z'\rangle\le L\left\|y-y'\right\|^2$. A simple example of OSL function is $F(y)=-1$ if $y>0$ and $F(0)=[-1;0]$. In that case, F is OSL of constant zero. Moreover, if $f:\mathbb{R}^d\to\mathbb{R}^d$ is a single-valued function and F be the convex set-valued function associated with f, defined by $F(y)=\mathrm{conv}(\mathrm{acc}_{z\to y}f(z))$, then F is OSL with constant L *iff* f is OSL with constant L. Finally, the sum of two OSL functions with constant L_1 and L_2, is OSL with constant L_1+L_2, where F_1+F_2 is defined by: $(F_1+F_2)(y)=\{u_1+u_2 : u_1\in F_1(y), u_2\in F_2(y)\}$.

Although OSL can be seen as a natural condition that extends Lipschitz continuity to set-valued dynamics, several examples presented in this paper will not satisfy the OSL conditions, while others will. In the fluid case, the DI derived in the examples of Sections 3.2 and 3.3 do not satisfy the OSL condition but can be transformed into DI with the OSL property by using a change of variables. However, in Section 3.4, the DI cannot be made OSL because it has several solutions.

As for the mean field examples, the DI of Section 4.3 is not OSL but has a unique solution. For a range of parameter, we were able to find a change of variable that makes the dynamics OSL but not in the general case. Similarly, the example in §4.4 is not OSL and we could not find any transformation into an OSL DI even though it has a unique solution.

2.2.2 Explicit bound on the stochastic approximation

The set-valued function F, defined by (5) represents a limit of the functions f^N as N goes to infinity. To be able to bound the quality of the approximation of the DI $\dot{y} \in F(y)$, we need a bound on the speed of convergence of f^N to F. For that purpose, we define the distance $d(f, F)$ between a function f^N and a set-valued function F by:

$$d(f^N, F) = \sup_{x\in\mathbb{R}^d} \inf_{y\in\mathbb{R}^d} \max\left(\|x-y\|, \inf_{z\in F(y)} \left\|f^N(x) - z\right\|\right). \quad (8)$$

The next lemma shows that, by construction of F, $d(f^N, F)$ converges to 0.

Lemma 3. *Let F be defined by Equation (5). Then, there exists a sequence δ^N such that $\lim_{N\to\infty} \delta^N = 0$ and $d(f^N, F) \le \delta^N$.*

Proof. This result is a direct consequence of Lemma 15, given in B.1. □

Although δ^N is not explicit in the lemma, it can be computed very easily in many cases. In particular, if f^N converges uniformly to a function f at speed δ^N, the same sequence δ^N satisfies $d(f^N, F) \le \delta^N$. This is the case in the examples of Section 4 where the drifts are constant in N.

This lemma guarantees that even if f^N does not converge uniformly to a function f, we always have $\lim_{N\to\infty} d(f^N, F) = 0$. For example, this is the case for the drift of the model of opportunistic scheduling developed in §3.3. When f^N does not converge uniformly, the existence of δ^N is guaranteed but its computation may depend on the example considered.

Theorem 4. *Let $Y^N(k)$ be a Markov chain on $\mathbb{R}^d$ satisfying (4). Assume that the assumptions of Theorem 1 hold and that*

- *$U^N(k+1)$ is bounded in second moment: $\mathbb{E}\left(\left\|U^N(k+1)\right\|^2 \mid Y^N(k)\right) \le b$.*
- *F is OSL of constant L and $d(f^N, F) \le \delta^N$.*

then the DI (6) has a unique solution y and there exist constants A_T, B_T, C_T depending only on T, L and c such that for all ε:

$$\mathcal{P}\left(\sup_{0\le t\le T} \left\|Y^N(t) - y(t)\right\| \ge \left\|Y^N(0) - y(0)\right\| e^{LT} + \min\left\{T, \frac{e^{LT}}{\sqrt{2L}}\right\} \sqrt{\gamma^N A_T + \delta^N B_T + \varepsilon C_T}\right) \le \frac{\gamma^N bT}{\varepsilon^2}.$$

Proof. The proof is given in B.2. □

The constants A_T, B_T, C_T and the sequence δ^N are given in B.2. These constants are of a similar order as bounds that can be obtained in the case where f is Lipschitz (see [19]). However, the convergence speed with respect to N is $O(\sqrt{\gamma^N})$ (compared with $O(\gamma^N)$ in the Lipschitz case).

2.3 Density Dependent Population Processes

In this section, we show that our results can be adapted to the case of continuous time Markov chains and in particular to the well-known model of density dependent population processes of Kurtz [23].

Let D^N be a continuous time Markov chain on $\mathbb{Z}^d/N$ ($d \ge 1$) for $N \ge 1$. D^N is called a *density dependent population process* if there exists a set $\mathcal{L} \subset \mathbb{Z}^d$ (with $0 \notin \mathcal{L}$), such that for each $\ell \in \mathcal{L}$ and $y \in \mathbb{Z}^d/N$, the rate of transition from y to $y + \ell/N$ is $N\beta_\ell(y) \ge 0$, where $\beta_\ell(.)$ does not depend on N. The ith component of $D^N(t)$, $D^N_i(t)$ can be seen as the density of individuals of a population that are in state i, hence the name, and a transition ℓ changes the number of individuals in state i by the quantity ℓ_i.

Let us assume that the transition rate from a state y is bounded: $\tau \stackrel{\text{def}}{=} \sup_{y\in\mathbb{Z}^d} \sum_{\ell\in\mathcal{L}} \beta_\ell(y) < \infty$ and that $\sum_{\ell\in\mathcal{L}} \|\ell\| \sup_y \beta_\ell(y) < \infty$. The expectation of the change of the system during a small interval dt is $f(y)dt$ where $f(y)$ is the drift of the system, defined by $f(y) = \sum_{\ell\in\mathcal{L}} \beta_\ell(y)\ell$. If f is Lipschitz, it is well-known that $D^N(.)$ goes to the solution of the ODE $\dot{y} = f(y)$ as N grows [23]. Under these assumptions, the continuous time Markov chain $D^N(t)$ can be seen as a composition of a Poisson counting process $\Lambda^N(t)$ whose rate is $N\tau$ with a discrete time Markov chain Y^N: $D^N(t) = Y^N(\Lambda^N(t))$. This is called the uniformization of the Markov chain. Using Theorems 1 and 4, we show that this convergence still holds for general drifts, replacing f by its set-valued counterpart F, defined in (5).

Theorem 5. *Assume that* $\sup_{y\in\mathbb{Z}^d}\sum_{\ell\in\mathcal{L}}\beta_\ell(y) < \infty$ *and that* $\sum_{\ell\in\mathcal{L}}\|\ell\|\sup_y\beta_\ell(y) < \infty$. *Let* f *be defined by* $f(y) = \sum_\ell \ell\beta_\ell(y)$. *For all* $T > 0$:

$$\inf_{d\in\mathscr{S}_T(y_0)} \sup_{0\le t\le T} \left\|D^N(t) - d(t)\right\| \xrightarrow{\mathcal{P}} 0,$$

where $\mathscr{S}_T(y_0)$ *is the set of solutions of the DI* (6) *starting in* y_0.

Moreover, if F *is OSL of constant* L *and* $\sup_y\sum_{\ell\in\mathcal{L}}\|\ell\|^2\sup_y\beta_\ell(y) \le b$ *then the differential inclusion* (6) *has a unique solution* d *and there exist constants* A_T, B_T, C_T *depending only on* T, L *and* c *such that for all* ε:

$$\mathcal{P}\left(\sup_{0\le t\le T}\left\|D^N(t) - d(t)\right\| \ge \left\|D^N(0) - d(0)\right\| e^{LT} + \min\left\{T, \frac{e^{LT}}{\sqrt{2L}}\right\}\sqrt{\frac{A_T}{N} + \varepsilon C'_T}\right) \le \frac{b + 1/\tau}{N\varepsilon^2}T.$$

Proof. The proof is based on uniformization of the Markov chain D (a sampling technique for continuous-time Markov processes with bounded jump rates): We construct a discrete time Markov chain Y^N that satisfies the assumptions of Theorem 1. A detailed proof is given in B.3. □

The constant A_T is the same as in Theorem 4. The constant C'_T is given in B.3.

3 Application 1: Fluid limits and Stability Issues

Fluid limits have become an important tool for studying stochastic stability of queuing networks. For a large class of queuing networks, when the initial state of the system is rescaled by a factor $N \to \infty$ and the time is accelerated by the same factor N, the system is shown to satisfy a system of deterministic equations, called the *fluid limit model*. The link between stability of fluid limit and stochastic stability has a long history of research. The results obtained can be mainly categorized in two types. On the one hand, many people have studied specific queuing models with general arrival process and service rate and constructed explicitly the fluid model equations corresponding to these systems, see [12, 11] and the references therein. More recently, structural properties have been studied but only for continuous drifts [17]. Theorem 7 makes the link between the two approaches by showing that generic results can be obtained even for non-continuous dynamics.

In §3.1, we define the fluid limit model of a system and show that the fluid limit model satisfies a differential inclusion. We further show that stability of this differential inclusion implies stability of the stochastic system. These results are illustrated in §3.2 and §3.3 to study the stability of opportunistic scheduling policies. We conclude by some limitations of the approach in §3.4.

3.1 Definition of fluid limits and stability

Let $X(.)$ be a discrete time[2] Markov chain in $\mathbb{R}^d$. For any $y_0 \in \mathbb{R}^d$ and $N > 0$, we consider the rescaled process $\bar{Y}^N$ for which the state has been scaled by a factor $1/N$ and the time accelerated by N:

$$\bar{Y}^N(t) = \frac{1}{N}X(\lfloor N\cdot t\rfloor) \qquad\qquad \bar{Y}^N(0) = \frac{1}{N}X(0) = y_0.$$

[2]For readability, we restrict our presentation to discrete time models. However, these results can be extended directly to continuous time Markov chains using uniformization as in §2.3.

The next result shows that fluid limits are solutions of a differential inclusion that can be constructed directly from the drift.

We say that a set E of functions from $\mathbb{R}^+$ to $\mathbb{R}^d$ *contains the fluid limits* of Y^N if for all $T > 0$:

$$\inf_{y \in E} \sup_{0 \le t \le T} \left\| \bar{Y}^N(t) - y(t) \right\| \xrightarrow{\mathcal{P}} 0. \tag{9}$$

Such a set E is a superset of the limiting behaviors of X^N/N when the initial condition and the time are rescaled by a factor N. This definition is in accordance with classical definitions of fluid limits in the literature. For example in [17], a fluid limit is defined as a weak limit of the process $\bar{Y^N}$. The support of such a measure contains the fluid limits according to our definition.

The following theorem shows that the differential inclusion corresponding to (10) describes a superset of the limiting behavior of $\bar{Y}^N$.

Proposition 6. *Assume that the drift $f(x) = \mathbb{E}\left(X(t+1) - X(t) \mid X(t) = x\right)$ is bounded and that $\lim_{R\to\infty} \mathbb{E}(\|X(t+1) - X(t)\| \mathbf{1}_{\|X(t+1)-X(t)\| \ge R} \mid X(t) = x) = 0$. Let F be a set-valued function defined as*

$$F(y) \stackrel{\text{def}}{=} \operatorname{conv}\left(\operatorname*{acc}_{N\to\infty} f(N \cdot y^N) \text{ with } \lim_{N\to\infty} y^N = y\right). \tag{10}$$

Then, the set of solutions $\mathscr{S}_T(y_0)$ of the differential inclusion $\dot{y} \in F(y)$ starting in x contains the fluid limits of Y^N (in the sense of (9)).

Proof. This result is a direct consequence of Theorem 1. To fit into the framework of Theorem 1, let us call $f^N(y) \stackrel{\text{def}}{=} f(Ny)$. For all $t \in \frac{1}{N} \cdot \mathbb{N}$, $\bar{Y}^N(t + \frac{1}{N})$ satisfies $\bar{Y}^N(t + \frac{1}{N}) = \bar{Y}(t) + \frac{1}{N}\left(f^N(\bar{Y}^N(t)) + U(t + \frac{1}{N})\right)$ with $\mathbb{E}\left(U(t + \frac{1}{N}) \mid X(t)\right) = 0$. The function F defined by Equation (10) is the same as in Equation (5). As f is bounded, f^N is bounded. Moreover, the assumption implies that $X(t+1) - X(t)$ is uniformly integrable. This shows that Y^N satisfies assumptions of Theorem 1. □

This theorem does not require any continuity assumption on f and provides a characterization of the fluid limit in term of differential inclusions. This theorem can be viewed as a generalization of Proposition 1.5 of [17] that assumes that f^N goes to some continuous function f. If the differential inclusion has a unique solution y on $[0;T]$, then y is called the fluid limit of Y^N and Proposition 6 implies that $\bar{Y}^N$ converges to y:

$$\sup_{0 \le t \le T} \left\| \bar{Y}^N(t) - y(t) \right\| \xrightarrow{\mathcal{P}} 0.$$

In turn, this result can be viewed as a generalization of Theorem 1.6 of [17].

There are several ways to define the stability of the fluid limit model. We follow the definition of [30, 17] and say that the differential inclusion $\dot{y} \in F(y)$ is stable if there exists $T > 0$ and $\rho < 1$ such that:

$$\text{For any } y \text{ solution of } \dot{y} \in F(y) \text{ with } \|y(0)\| = 1 : \quad \inf_{0 \le t \le T} \|y(t)\| \le \rho < 1. \tag{11}$$

As expressed by the next proposition, stability of the fluid limit in the sense of (11) implies the stability of the stochastic model. Before stating the main theorem, we recall the definitions of φ-irreducibility and petite set that are useful to show stability of a Markovian process on a non-countable set. We refer to [26] for a more detailed presentation of these notions.

A discrete time Markov chain X on $\mathbb{R}^d$ is said to be φ-irreducible if there exists a σ-finite measure φ such that for any set $A \subset \mathbb{R}^d$, $\varphi(A) > 0$ implies $\sum_{k\ge 0} \mathcal{P}(X(k) \in A \mid X(0) = x) > 0$. Moreover, a set $A \subset \mathbb{R}^d$ is said to be *petite* if for some fixed probability measure a on $\mathbb{Z}^+$ and some non-trivial measure ν on $\mathbb{R}^d$, $\nu(B) \le \sum_{k\ge 0} \mathcal{P}(X(k) \in B \mid X(0) = x) a(k)$ for all $x \in A$ and $B \subset \mathbb{R}^d$. Finally, X is said to be positive Harris recurrent if X has a unique stationary probability distribution π and $P^k(x, .)$ converges to π. In particular, if the state space of X is included in $\mathbb{Z}^d$ and if X is irreducible and aperiodic, then X is φ-irreducible and all compact sets are petite.

Theorem 7. *Assume that X is an aperiodic, φ-irreducible Markov chain such that all compact sets are petite. Assume that the drift $f(x) = \mathbb{E}\left(X(t+1) - X(t) \mid X(t) = x\right)$ is bounded and that $\lim_{R\to\infty} \mathbb{E}(\|X(t+1) - X(t)\| \mathbf{1}_{X(t+1)-X(t)\geq R} \mid X(t) = x) = 0$ and let F be defined as in Equation (10):*

$$F(y) \stackrel{\text{def}}{=} \operatorname{conv}\left(\underset{N\to\infty}{\operatorname{acc}} f(N \cdot y^N) \quad \text{for all } \{y^N\}_{N\in\mathbb{N}} \text{ s.t. } \lim_{N\to\infty} y^N = y \right).$$

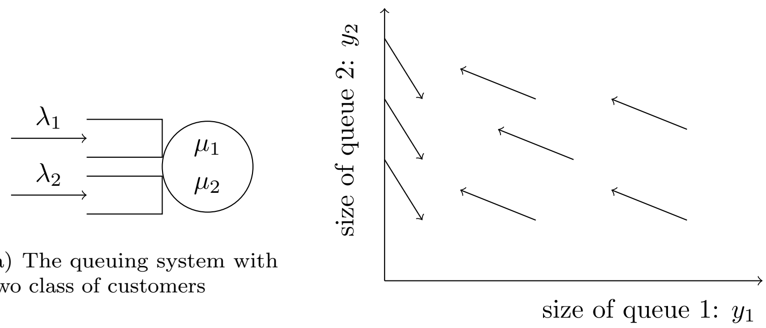
If the differential inclusion $\dot{y} \in F(y)$ is stable in the sense of Equation (11), then X is positive Harris recurrent.

Proof. Theorem 1.4 of [17] shows that if all functions y of a set containing the fluid limits of $\bar{Y}^N$ are stable in the sense $\inf_{0\leq t\leq T} \|y(t)\| \leq \rho$, then the process X is Harris recurrent. Proposition 6 shows that the solutions of the differential inclusion $\dot{y} \in F(y)$ contains the fluid limits. Therefore, the stability of the DI given by Equation (11) implies the Harris recurrence of X. □

This theorem shows that the constructive definition of F allows one to obtain sufficient conditions for stability. In the following, we illustrate this result with some examples. When the DI has a unique solution, the stability conditions obtained by Theorem 7 are generally also necessary. However, when the DI has multiple solutions, the DI may describe a superset of the fluid limits and stability conditions obtained may be too strong, as in the example of §3.4.

3.2 Fluid limit of a system of parallel queues with static priority.

We consider a time-slotted model of a queuing system composed of one server serving multiple classes of users. There are K classes of customers. At time step t, $A_k(t)$ customers of class k arrive. A_k are *i.i.d.* with $\mathbb{E}(A_k) = \lambda_k$. Let $X_k(t)$ be the number of customers of class k in the system at time t. For $k < k'$, customers of class k have preemptive priority over customers of class k'. When the system serves a customer of class k, it leaves the system in the same time slot with probability μ_k. This means that if there are one or more customers of class 1 present in the system, a customer of class 1 leaves the system with probability μ_1 and no other customer departs in the same time slot. When there are no customer of class $1 \dots k-1$ and one or more customers of class k, a customer of class k departs with probability μ_k. This model with two classes of customers is depicted in Figure 2(a) where each queue corresponds to a class of customers.



(a) The queuing system with two class of customers

(b) Drift of the system

Figure 2: The system and the corresponding drift

The drift of the system is easy to compute:

$$f(x) = \begin{cases} (\lambda_1 - \mu_1, \lambda_2, \dots, \lambda_K) & \text{if } x_1 > 0 \\ (\lambda_1, \dots, \lambda_{k-1}, \lambda_k - \mu_k, \lambda_{k+1}, \dots, \lambda_K) & \text{if } x_1 = \dots = x_{k-1} = 0 \text{ and } x_k > 0. \end{cases} \tag{12}$$

For example, when $\lambda_1 = \lambda_2 = 0.1$ and $\mu_1 = \mu_2 = 0.3$, the drift is $f(x_1, x_2) = (-0.2, +0.1)$ if $x_1 > 0$ and $(+0.1, -0.2)$ if $x_1 = 0$ and $x_2 > 0$. This drift is depicted in Figure 2(b). As shown on

Figure 2(b), the drift is constant for all $x_1 > 0$ but is discontinuous for $x_1 = 0$. Because of this discontinuity, there is no function x differentiable almost everywhere such that $\dot{x}(t) = f(x)$: the axis $x_1 = 0$ both attracts the trajectories from $x_1 > 0$ and repulses the trajectories starting from $x_1 = 0$.

It should be clear that the model satisfies all the assumptions of Proposition 6. Let us compute the set-valued function F corresponding to the drift f defined as in Equation (10). For all k, let us define $u_k \stackrel{\text{def}}{=} (\lambda_1, \dots, \lambda_{k-1}, \lambda_k - \mu_k, \lambda_{k+1}, \dots, \lambda_K)$. When $x_1 > 0$, all points x' in a small neighborhood of x are such that $x'_1 > 0$. Thus, f is locally constant and $F(x)$ is single-valued: $F(x) = \{u_1\}$. Because of the discontinuity at $x_1 = 0$, when $x_2 > 0$, in a neighborhood of $(0, x_2, x_3 \dots)$, there are points x' with $x'_1 > 0$ and points x' with $x'_1 = 0$ (and $x'_2 > 0$). Thus, $F(0, x_2, x_3 \dots)$ is the convex hull of the vectors $\{u_1, u_2\}$. In the two class case, this convex hull corresponds to the dashed line of Figure 3(a). Thus, the drift has the expression:

$$F(x) = \begin{cases} u_1 & \text{if } x_1 > 0 \\ \operatorname{conv}(u_1, \dots, u_k) & \text{if } x_1 = \dots = x_{k-1} = 0, x_k > 0. \end{cases} \tag{13}$$

Let us assume that $\sum_k \lambda_k/\mu_k < 1$ and let us show that the differential inclusion associated with F has a unique solution when starting from $x = (x_1 \dots x_K)$. The function F is not OSL. To show that, let $x = (\varepsilon/\mu_1, 3\varepsilon/\mu_2, 0 \dots 0)$ and $x' = (0, \epsilon/\mu_2, 0 \dots, 0)$, then:

$$\langle x - x', f(x) - f(x') \rangle = \sum_i x_i (f(x_i) - f(x'_i)) = \frac{\varepsilon}{\mu_1}(-\mu_1) + \frac{2\varepsilon}{\mu_2}\mu_2 = \varepsilon.$$

Thus, there does not exist an L such that for all ε, $\langle x - x', f(x) - f(x') \rangle \leq L \left\| x - x' \right\|^2 = O(\varepsilon^2)$.

However, a change of variable makes F an OSL function. Let $y_k = \sum_{i=1}^k x_i/\mu_i$ and let $g_k(y) = \sum_{i=1}^k f_i(x)/\mu_i(x)$ be the associated drift. A straightforward computation shows that $g_k(y) = \sum_{i=1}^k \lambda_k/\mu_k - \mathbf{1}_{y_k > 0}$ which implies that g is an OSL function. Thus, its associated set-valued function G is OSL. Therefore, the differential inclusion $\dot{y} \in G(y)$ has a unique solution, given by:

$$\dot{y}_k = \begin{cases} \sum_{i=1}^k \frac{\lambda_k}{\mu_k} - 1 & \text{if } y_k > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

Since the change of variable from x to y is a bijection, the differential inclusion $x \in F(x)$ also has a unique solution. Equation (14), along with the condition $\sum_k \lambda_k/\mu_k < 1$ shows that there exists a sequence $0 \leq T_1 \leq \dots \leq T_K < \infty$ such that for all $t \in [T_{k-1}, T_k]$, the derivative of x satisfies

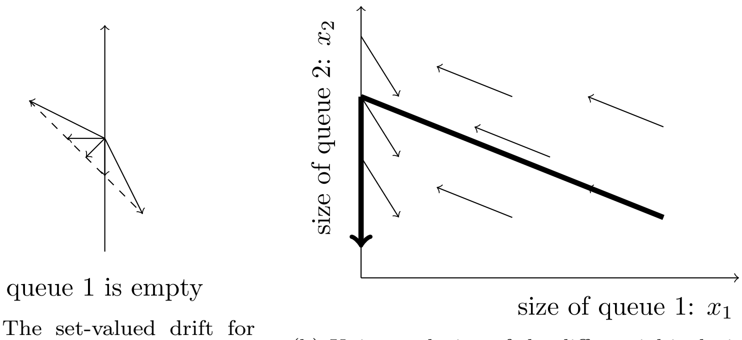
$$\dot{x}(t) = \left(0, \dots, 0, \lambda_k - \left(1 - \sum_{i<k} \frac{\lambda_i}{\mu_i}\right) \mu_k, \lambda_{k+1}, \dots, \lambda_K\right).$$

If the condition $\sum_k \lambda_k/\mu_k < 1$ were not verified, then let k be the minimal k such that $\sum_{i \leq k} \lambda_i/\mu_i > 1$. In that case, the fluid limit would diverge to an infinite number of customers of type $k \dots K$ while the number of customers of type $1 \dots k-1$ would remain zero for the fluid limit.

This trajectory is depicted in Figure 3(b). Moreover, the solution of the differential inclusion goes to 0 in finite time and this system satisfies the assumptions of Theorem 7. This shows that the system is stable. Although this result can be shown directly, our framework provides an easy way to construct the fluid limit and prove the convergence of the original process.

3.3 Stability of opportunistic scheduling policies in wireless networks

In this section, we show how Proposition 6 and Theorem 7 can be used to characterize the stability of opportunistic scheduling policies in a wireless setting with flow-level dynamics. The stability of such policies has been first studied in [2, 34]. Because of the discontinuity of the dynamics, generic approaches such as the ones introduced in [17] fail and ad-hoc methods have been developed. Our



(a) The set-valued drift for $x_1 = 0$ is the set of vectors displayed by the dashed line.

(b) Unique solution of the differential inclusion.

Figure 3: Convex hull of the drift at $C_2 = 0$ and unique solution of the fluid limit.

framework shows that a systematic generic approach can also be used in that case to compute easily the limiting dynamics and show stability.

We consider the model studied in [2]. Transmissions occur in a time-slotted channel. There are K classes of users. At time slot t, $A_k(t)$ new users of type k arrive in the system. The $A_k(t)$ are *i.i.d.* with $\mathbb{E}(A_k(t)) = \lambda_k$, $\mathbb{E}(A_k^2(t)) < \infty$. The condition of the channel is varying over time and at time slot t, a user of type k has condition $i \in \{1 \dots I_k\}$ with probability $q_{k,i} \neq 0$. The channel condition of a user is independent of other users and of the channel history. At each time slot, a server observes the channel condition of all users and chooses to serve one user. If this user is of type k and has a channel condition i, this user leaves the system with probability $\mu_{k,i}$. Without loss of generality, we may assume $\mu_{k,1} > \mu_{k,2} \dots$ The quantity $\mu_{k,i}$ represents the rate at which at user k with condition i is served. At best, a user of type k is served at rate $\mu_k^{\max} \stackrel{\text{def}}{=} \mu_{k,1}$.

The design of an efficient policy for scheduling the users has received a considerable attention in the past (see [2, 34] and the reference therein). When building efficient policies for such system, a first requirement is that this policy stabilizes the system, *i.e.* such that the number of users in each class is a positive recurrent Markov chain. We next show how our framework can be used to prove the following results (originally proved in [2] by ad-hoc arguments).

Proposition 8 (Theorem 5.2 of [2])**.** *There exists a scheduling policy that stabilizes the system if and only if*

$$\sum_{k=1}^{K} \frac{\lambda_k}{\mu_k^{\max}} < 1. \tag{15}$$

Proof. It should be clear that (15) is a necessary condition for stability. Therefore, we only show that (15) is a sufficient condition and we assume that (15) holds.

Let us consider the following policy (called "Best Rate" policy in [2]):

- if there are n users $u_1 \dots u_n$ of class $k_1 \leq \dots \leq k_n$ that are in their best channel condition, serve the user with the smallest class (*i.e.* user u_1).
- if there are no users in their best channel condition, serve a user at random.

For all k, let $X_k(t)$ be the number of users in class k at time t when applying this policy. Since the channel conditions are independent, the process $X(.)$ is a Markov chain.

Let us compute the set-value function F at point $y = (0, \dots 0, y_\ell, \dots y_K)$, with $y_\ell > 0$. For that, we must study the limit of the sequence $f^N(y^N) \stackrel{\text{def}}{=} f(Ny^N)$ where $(y^N)_{N\in\mathbb{N}}$ is any sequence with $\lim_{N\to\infty} y^N = y$ and $f(x) = \mathbb{E}(X(t+1) - X(t) \mid X(t) = x)$ is the drift of the system. Let $p_i^N = (1 - q_{i,1})^{Ny_i^N}$ be the probability that there are no user of class i in its best state when the number of users in each class is Ny^N. Notice that if $y_i > 0$, this quantity goes to 0 as N goes to infinity.

If the server is serving a user of type i which is in its best state, the drift of the system is $u_i = (\lambda_1, \dots, \lambda_{i-1}, \lambda_i - \mu_i^{\max}, \lambda_{i+1}, \dots, \lambda_K)$. This occurs if there is a user of class i in its best state but no user of class $1 \dots i-1$ in its best state, which occurs with probability $p_1^N \dots p_{i-1}^N(1-p_i^N)$. Therefore, the value of the drift at Ny^N is equal to

$$f(Ny^N) = (1-p_1^N)u_1 + \dots + p_1^N \dots p_{\ell-1}^N(1-p_\ell^N)u_\ell + o(1).$$

This shows that $\lim_{N\to\infty} d(f(Ny^N), F(y)) = 0$ where $d(\cdot,\cdot)$ is the distance defined by Eq.(8) and F is the following set-valued function:

$$F(x) = \begin{cases} u_1 & \text{if } x_1 > 0 \\ \operatorname{conv}(u_1, \dots, u_k) & \text{if } x_1 = \dots = x_{k-1} = 0, x_k > 0. \end{cases} \tag{16}$$

However, notice that when y_i^N goes to zero as N goes to infinity, the sequence p_i^N does not necessarily converge as N goes to infinity. This implies that the rescaled drift f^N does not converge to any single-valued function (continuous or not) in that case.

Equation (16) is the same as (13). Therefore, up to a change of variable, F is OSL and the differential inclusion has a unique solution that goes to 0 in finite time under condition (15). This shows that (15) implies the stability of the stochastic system. □

3.4 Limitations of the differential inclusion approach

In many cases, the differential inclusion approach allows one to characterize exactly what will be the fluid limits of a stochastic system. In particular, this is the case when the differential inclusion has a unique solution. However, in other cases, the construction of the set-valued function F by convexification leads to a super-set of the fluid limits, so that the stability conditions provided by this approach are only sufficient but not necessary.

Let us consider the example of three weakly coupled queues, presented in §5.1 of [9]. The system is composed of 3 queues. Customers arrive at queue i with rate λ_i. If x_i, x_j, x_k are the numbers of customers present in queues $i \neq j \neq k \in \{1,2,3\}$, a customer of queue i is served with rate $\psi_i(x)$:

$$\psi_i(x) = \begin{cases} a_i & \text{if } x_j = x_k = 0 \\ a_{ij} & \text{if } x_j > 0, x_k = 0 \\ 1 & \text{if } x_j > 0, x_k > 0, \end{cases}$$

where $a_i \leq a_{ij} \leq 1$.

The drift of the system is $f(x) = (\lambda_1 - \psi_1(x), \lambda_2 - \psi_2(x), \lambda_3 - \psi_3(x))$. Let us compute the solutions of the corresponding differential inclusion starting from a point (x_1, x_2, x_3) with $x_1, x_2, x_3 > 0$. Let $x(.)$ be a solution of the differential inclusion. For t small enough, the derivative of x is $\dot{x}(t) = (\lambda_1 - 1, \lambda_2 - 1, \lambda_3 - 1)$. Let us assume (w.l.o.g.) that

$$\lambda_1 < 1, \tag{17}$$

and that $x_1(t)$ reaches 0 before $x_2(t)$ and $x_3(t)$.

Let T_1 be the time when $x_1(t)$ reaches 0. For $t > T_1$ and as long as $x_2(t) > 0$ and $x_3(t) > 0$, using the convex closure of the drift of the system implies that there exists $0 \leq \theta \leq 1$ such that $\dot{x}(t) = (\lambda_1 - \theta, \lambda_2 - \theta - (1-\theta)a_{23}, \lambda_3 - \theta - (1-\theta)a_{32})$. Since $\dot{x}_1(t) = 0$ for $t > T_1$ then $\theta = \lambda_1$ and the drift is:

$$\dot{x}(t) = (0, \lambda_2 - \lambda_1 - a_{23}(1-\lambda_1), \lambda_3 - \lambda_1 - a_{32}(1-\lambda_1)).$$

One of the two components of this drift has to be negative for the system to be stable. Thus, we may assume w.l.o.g. that

$$\lambda_2 < \lambda_1 + a_{23}(1-\lambda_1), \tag{18}$$

and that $x_2(t)$ reaches 0 before $x_3(t)$.

When $x_2(t)$ reaches 0, F is the convex closure of 4 vectors u_0, u_1, u_2, u_{12} corresponding respectively to the drift when $(x_1 > 0, x_2 > 0)$, $(x_1 = 0, x_2 > 0)$, $(x_1 > 0, x_2 = 0)$, and $(x_1 = x_2 = 0)$. Using the fact that the actual drift is in F and that $\dot{x_1} = \dot{x_2} = 0$, there exist $\theta_0, \theta_1, \theta_2 \in [0; 1]$ with $\theta_1 + \theta_2 + \theta_0 \leq 1$ such that:

$$0 = \lambda_1 - \theta_0 - a_{13}\theta_2 \tag{19}$$
$$0 = \lambda_2 - \theta_0 - a_{23}\theta_1 \tag{20}$$
$$\dot{x}_3(t) = \lambda_3 - \theta_0 - a_{31}\theta_2 - a_{32}\theta_1 - (1 - \theta_0 - \theta_1 - \theta_2)a_3. \tag{21}$$

In general, there are multiple triplets $(\theta_0, \theta_1, \theta_2)$ such that (19–20) are satisfied[3]. If for all $(\theta_0, \theta_1, \theta_2)$ such that (19–20) are verified, (21) is negative, then the system is stable. Conversely, if for all $(\theta_0, \theta_1, \theta_2)$ satisfying (19–20), (21) is positive, then the fluid limit is unstable. However, in general, one cannot compute the stability condition of the fluid system only using Equations (19–20–21) since the sign of (21) may depend on $(\theta_0, \theta_1, \theta_2)$.

In [9], the exact stability conditions are given. Equations (17–18) are similar while the conditions on θ (19–20) are expressed as a function of the stationary distribution of X_1, X_2 conditioned by the fact that $X_3 > 0$. However, the proof of this result is much more involved and these equations cannot be solved in closed form whereas the present approach gives upper bounds in closed form. In [21], a simpler approximation method is also applied to the same problem. However, this leads to looser bounds than ours.

4 Application 2: Mean Field Limits

In this section, we show how our framework allows one to extend the expressive power of mean field models to study models with discontinuous dynamics. Although there is no commonly admitted definition of what is exactly a mean field model, they all share the same principle. The main idea is to study the behavior of a system composed by a large number N of objects evolving in a common environment. When N is finite, the behavior of each object depends on its interactions with others. However, as N goes to infinity, one can show that in many cases, objects become independent and interact only through aggregate quantities.

Convergence results for mean field models have received considerable attention in the past. Many results concern the convergence of the occupancy measure (see Eq.(22) for a more formal definition). This is often done by showing that it asymptotically satisfies a deterministic differential equation as N goes to infinity [23, 13, 4]. This can be used to obtain both transient and steady-state dynamics of the proportion of objects in a given state [4] or to prove propagation of chaos [31]: under some conditions, the steady-state distribution of objects has asymptotically a product form (*e.g.* Corollary 2 of [4]).

A powerful extension of these results is to study asymptotic properties of the trajectories of the objects. It many cases, it can be shown that objects become asymptotically independent as N grows and that their behavior only depends on the occupancy measure [20, 7]. This allows one to get an efficient simulation of the behavior of an object. These results are stronger that convergence of the occupancy measure and they imply the later. However, proving the convergence is usually quite challenging and although there exists generic convergence results, the assumptions are not easy to verify [8]. In the rest of this section, we will only focus on the convergence of the occupancy measure since it suffices for our results. An extension of our results to study the individual behavior of objects would be useful but is left for future work.

The stochastic approximation framework is a powerful tool to show these types of convergence results. However, except in particular cases where ad-hoc proofs are presented, convergence results for mean field models in the literature [23, 4, 13] always assume the Lipschitz continuity of the drift. Our framework shows that these models can be extended to characterize the limiting behavior of systems with discontinuous dynamics and therefore simplify their study.

[3]Remark that no change of variable can make this dynamics OSL since it has multiple solutions.

This section is organized as follows. We first present a generic mean field model and show how our framework can be adapted in §4.1. Then, we show how our result can be used to characterize the steady state distribution in §4.2. Then, we focus on two examples. In the first, we compare the gain obtained by taking centralized decisions in parallel servers. This example shows how to handle discontinuities that arise because of centralized actions. The second one shows a more realistic example that illustrates discontinuities that arise because of buffers.

4.1 Mean field model and its convergence

We consider of system composed of N *objects* evolving in a finite state space $\mathcal{S} = \{1 \dots d\}$. Time is discrete and the state of object n at time step k is denoted by $X_n^N(k)$. The state of the global system at time k is $(X_1^N(k) \dots X_N^N(k))$. We denote by $Y^N(k)$ the empirical measure associated with the N objects. Since an object has d possible states, $Y^N(k)$ can be represented by a vector with d components, its ith component being the proportion of objects in state i:

$$Y_i^N(k) \stackrel{\text{def}}{=} \frac{1}{N} \sum_{n=1}^{N} \mathbf{1}_{X_n^N(k)=i}. \tag{22}$$

The system $(Y^N(k))_k$ is assumed to be a Markov chain. In particular, this is true if the objects all have a Markovian dynamics and if the law of the whole system is invariant by any permutation of the N objects. The state space of this Markov chain is included in $\mathbb{R}^d$. To match the notation introduced in Section 2, we denote by f^N the rescaled drift of Y^N and by F the convex hull of its accumulation points.

There are multiple situations in which assumptions of Theorem 1 are satisfied, for example, if the number of objects that perform a transition at each time slot is bounded by a deterministic constant c, the intensity is $\gamma^N = 1/N$. If $Y^N(0) \to m_0$, then Y^N can be approximated by the solutions of the differential inclusion as N grows. for all $T > 0$:

$$\inf_{y \in \mathscr{S}_T} \sup_{0 \le t \le T} \left\| \bar{Y}^N(t) - y(t) \right\| \xrightarrow{\mathcal{P}} 0,$$

where $\mathscr{S}_T$ denotes the set of solutions of the DI $\dot{y} \in F(y)$ with $y(0) = y_0$.

This model can be modified to study continuous time dynamics, following §2.3. It can also be easily extended if objects evolve in a common environment, called the context $C(t) \in \mathbb{R}^{d'}$. An example will be provided in §4.4 where the context represents a shared buffer in which packets are stored. In that case, the quantity of interest is $(Y^N(t), C(t)) \in \mathbb{R}^{d+d'}$. If the number of objects that perform a transition during one time slot is bounded and if the evolution of the context is deterministic and there exists a constant k_1 such that for all y, c, $f^N(y, c) \le k_1$, then the assumptions of Theorem 1 still hold.

4.2 Stationary regime and steady state distribution

Mean field limits also provide a way to compute an approximation of the stationary distribution. When the drift f is continuous, it can be shown that if all trajectories of the differential equation $\dot{y} = f(y)$ converges to a point y^*, then the stationary distribution of the system of size N concentrates on y^* as N grows. In this section, we show the analog of this results for discontinuous dynamics, under the condition that the differential inclusion has a unique solution.

Let us assume that for any starting point $y(0)$, the differential inclusion $y \in F(y)$ has a unique solution on $[0; \infty)$. We denote this solution $t \mapsto \phi_t(y)$. We define the Birkhoff center of ϕ by:

$$R = \{x \in \mathbb{R}^d : \liminf_{t \ge 0} \|x - \phi_t(x)\| = 0\}.$$

The next theorem shows that the support of the stationary measures of the stochastic system Y^N concentrates on the Birkhoff center of the differential inclusion.

Theorem 9. *Under the conditions of Theorem 1, if the DI (6) has a unique solution y on $[0;T]$ and if for each N, Y^N has a stationary measure Π^N, then, any limit point of Π^N (for the weak convergence topology) has support in R.*

Computing the set R is often a hard problem, even for a differential equation. R contains all fixed points $\{y : 0 \in F(y)\}$ but may also contains limit cycles or chaotic behavior. This result is of particular interest when the DI has a unique point to which all trajectories converge:

Corollary 10. *If moreover Π^N is tight and there is a unique point y^* to which all trajectory converge, then $R = \{y^*\}$ and Π^N converges weakly to the Dirac measure in y^* : $\lim_{N\to\infty} \Pi^N = \delta_{y_*}$.*

Proof. Let us assume that from any starting point $y \in \mathbb{R}^d$, the differential inclusion $\dot{y} \in F(y)$ has a unique solution on $[0;\infty)$, denoted by $t \mapsto \phi_t(y)$. ϕ is clearly a semi-flow. Moreover, because of the first assumption of Theorem 1, $F(y(t))$ is bounded for $t \in [0;T]$, hence, ϕ_t is continuous in t. Moreover, let y_n be a sequence that converges to some $y \in \mathbb{R}^d$ and $z(t)$ be a limit point of $\phi_t(y_n)$ for $t \in [0:T]$. Then, similarly to the end of the proof of Theorem 1, it can be shown that z is the solution of a differential inclusion which shows that $\lim_{n\to\infty} \phi_t(y_n) = \phi_t(y)$. This shows that the deterministic process ϕ is a semi-flow continuous in t and y. Theorem 1 of [5] shows that any limit point of Π^N is an invariant probability for ϕ. Since ϕ is a continuous semi-flow, the Poincaré's recurrence theorem [25] shows that the invariant probabilities of ϕ have support in R. □

This result is the exact analog of Theorem 3 in [4] for continuous dynamics. It is similar to the decreasing step size case (when the step size γ^N depends on t instead of N), although in the latter, the stochastic system converges with probability one to R [16, 3].

In the next section, we show how this result can be applied to compare the steady-state performance of two load balancing strategies in server farms.

4.3 Comparison of push and pull strategies in server farms

The goal of our first example is to show how our framework can help the study of discontinuities due to centralized decisions. We consider a model of a server farm depicted on Figure 4. The system is composed of N identical servers. Jobs arrive in a system according to a Poisson process of rate $N\lambda \in [0;1)$ and have a size exponentially distributed of mean 1. Each server can buffer up to B jobs[4]. If all processors process jobs at rate 1 and jobs are routed uniformly at random, the average waiting time would be $1/(1-\lambda)$, independently of N. To reduce the waiting time, we consider two strategies that improve load balancing:

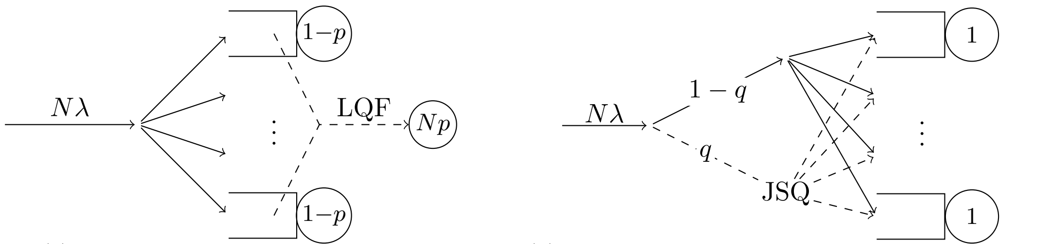
(a) pull strategy – we add a centralized server that serves jobs at rate Np. It chooses to serve jobs from the longest queue first (LQF). To provide a fair comparison, we consider that the total computing capacity remains N, *i.e.* the new speed of the N servers is set to $1-p$. This model is depicted on Figure 4(a). It is similar to the model studied in [32].

(b) push strategy – with probability q, a job is *pushed* to the server with the shortest queue (JSQ). With probability $1-q$, it is routed to a server at random (uniformly). This model is depicted on Figure 4(b).

Since these two strategies require costly synchronizations, our goal is to compare them when p and q are small. We will consider three cases: case (a) with $p = 5\%$, $q = 0\%$ case (b) with $p = 0\%$, $q = 5\%$ and a mix of both strategies with $p = q = 2\%$.

The system is composed of N objects. Each object is a queue and can have a state in $\{0 \dots B\}$. We denote by $Y_i^N(t)$ the proportion of servers having i jobs. Y^N is a Markov chain and its transitions are described in Table 1. In the column "modification of Y_i^N", the vector $\mathbf{e}_i$ denotes the vector with only the ith component equal to 1 and the others equal to 0. For example, an arrival in a queue with i jobs modifies Y^N by removing $1/N$ to Y_i^N and adding $1/N$ to Y_{i+1}^N.

[4] To avoid the dependence in B, in all our numerical examples, the computation are done with $B = 10^5$ which in practice is equivalent to $B = \infty$.



(a) Adding a centralized pulling server (b) Pushing a fraction of jobs to the shortest queue

Figure 4: Comparison of two strategies to reduce the response time. In both cases, the load of the system is λ. We compare a system with an additional fast centralized server that serves jobs from the longest queue (Figure 4(a)) and a system where a small fraction of jobs is routed to the server with the shortest queue (Figure 4(b)).

Transition	Rate	Modification of Y^N
arrival (if $i < B$)	$N\lambda(1-q)Y_i^N$	$-\frac{1}{N}\mathbf{e}_i + \frac{1}{N}\mathbf{e}_{i+1}$
due to JSQ	qN if no server with $< i$ jobs and $Y_i^N > 0$.	
departure (if $i > 0$)	$(1-p)Y_i^N$	$-\frac{1}{N}\mathbf{e}_i + \frac{1}{N}\mathbf{e}_{i-1}$
due to LQF	pN if no server with $> i$ jobs and $Y_i^N > 0$.	

Table 1: Transition and rate of the Markov chain associated with the model of server farm depicted on Figure 4(b).

To simplify the notation, let s_i be the proportion of servers having i jobs or more. By definition, we have $s_0 = 1$, s_i is decreasing and $s_{B+1} = 0$. Let s be a state and $i \in \{1 \dots B-1\}$. The drift of the system of size N can be computed directly using Table 1. It is independent of N and its projection on the ith coordinate is:

$$f_i(s) = \lambda(1-q)(s_{i-1} - s_i) + (1-p)(s_{i+1} - s_i) - g_i^{pull}(s) + g_i^{push}(s), \tag{23}$$

where g^{pull} and g^{push} are defined by:

$$g_i^{pull}(s) = \begin{cases} 0 & \text{if } s_{i+1} > 0 \text{ or } s_i = 0 \\ p & \text{otherwise.} \end{cases} \quad \text{and} \quad g_i^{push}(s) = \begin{cases} 0 & \text{if } s_{i-1} < 1 \text{ or } s_i = 1 \\ \lambda q & \text{otherwise.} \end{cases}$$

The drift when $i = B$ is similar except that the term λs_i should be removed. The drift for $i = 0$ is zero. The first two terms of Eq. (23) are due to randomly routed arrivals and non-centralized departures of jobs. The two terms g^{push} and g^{pull} are due to the centralized actions. The first two terms are continuous in s while the last two are discontinuous. Because of the discontinuity of g, the differential equation $\dot{s} = f(s)$ has no solutions in general.

The only discontinuities of f comes from the functions g^{pull} and g^{push}. Let us compute G^{pull} be the set-valued function corresponding to g^{pull}, defined by Eq. (5). Let s be a state of the system and $i \in \{1 \dots B-1\}$. We distinguish two cases. If $s_{i+1} > 0$, then the function g_i^{pull} is locally continuous in s and is equal to 0 and we have $G_i^{pull}(s) = \{0\}$. If $s_{i+1} = 0$, then for all neighborhoods of s, there are both points s' such that $s'_{i+1} > 0$ and other points such that $s'_{i+1} = 0$ and $s'_i > 0$. In that case the drift g_i^{pull} can be either 0 or p. This shows that the set-valued G^{pull} is the convex hull of the vectors $\{p\mathbf{e}_i \mid i \text{ s.t. } s_{i+1} > 0\}$. The computation of the set-valued function G^{push} corresponding to g^{push} is similar.

Combined with Eq. (23), this shows that the set-valued drift F is defined by:

$$F_i(s) = \left\{ \lambda(1-q)(s_{i-1} - s_i) - (1-p)(s_i - s_{i+1}) + u_i q - v_i p \;\middle|\; \begin{array}{l} u_i = 0 \text{ if } s_{i-1} < 1; \\ v_i = 0 \text{ if } s_{i+1} > 0; \\ \sum_{i \geq 0} u_i = \sum_{i \geq 0} v_i = 1 \end{array} \right\} \tag{24}$$

Again, the term for $i = B$ is not written but is similar except that the term λs_i should be removed and the term for $i = 0$ is zero.

The function F is not OSL. If $p = 0$, a change of variable $w_i = \sum_{j\leq i} s_j$ makes the dynamics OSL. If $q = 0$, the change of variable $v_i = \sum_{j\leq i} s_j$ makes the dynamics OSL. If both p and q are positive, then none of these changes of variable makes the dynamics OSL. Nevertheless, one can show that the DI has a unique solution. Let $i < B$ is such that $s_{i-1} > 0$ and $s_i = 0$. Combining Eq. (23) and Eq.(24), we get:

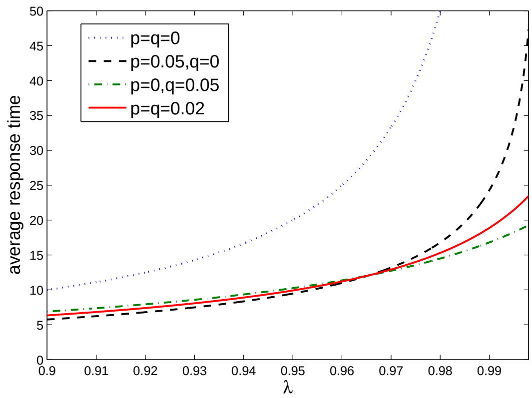
$$\sum_{k\geq i} \dot{s_k} = \max(0, \lambda(1-q)s_{i-1} - p) = \lambda(1-q)s_{i-1} - (1-u_{i-1})p \quad \text{and} \quad \sum_{k\geq i+1} \dot{s_k} = 0 = -p \sum_{k\geq i+1} u_i.$$

In particular, this shows that $pu_i = p(1 - u_{i-1}) = \min(p, \lambda(1 - q)s_{i-1})$. Similarly, if $j > 0$ is such that $s_j = 1$ and $s_{j+1} < 1$, then $pu_j = p(1 - u_{j+1}) = \min(\lambda q, (1 - p)(s_i - 1))$. Therefore, the differential inclusion $\dot{s} \in F(s)$ has a unique solution. Moreover, a direct computation shows that if $\lambda > p$, the Eq. (24) has a unique fixed point s, given by, for $i \in \{1, \dots, B\}$:

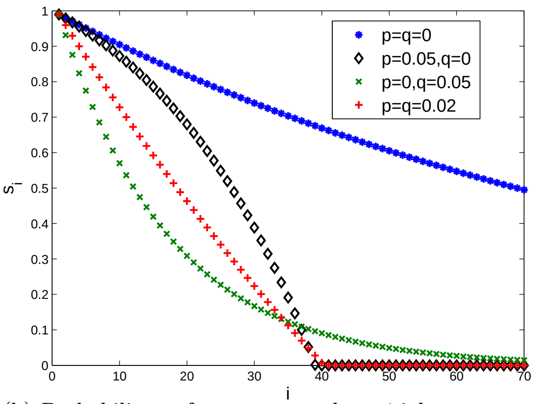
$$s_i = \begin{cases} \max\left(0,\ \alpha\left(\lambda\frac{1-q}{1-p}\right)^i + \beta\right) & \text{if } \lambda\frac{1-q}{1-p} \neq 1 \\ \max\left(0,\ \alpha i + \beta\right) & \text{if } \lambda\frac{1-q}{1-p} = 1 \end{cases} \tag{25}$$

where α and β are constants that can be computed using that $\lambda(1 - s_B) = s_1(1 - p) + p$ and $s_2 = s_1(1 + \lambda\frac{1-q}{1-p}) - \frac{\lambda}{1-p}$. If $\lambda \leq p$, the fixed point is $s_i = 0$ for $i \geq 1$. Moreover, this point is a global attractor of all trajectories. This fact is technical and can be shown by a careful examination of the differential inclusion corresponding to Eq. (24), using similar techniques as in Section 7.3 of [33]. Therefore, Theorem 9 shows that the stationary measure of the system concentrates on the fixed point given by Equation (25).

These results allow us to easily compare the gain obtained by using a centralized pulling system versus a centralized pushing system. A numerical evaluation of the fixed point is reported on Figure 5 on which we compare four situations: a scenario with no centralization at all, the scenario with a centralized server at speed $pN = .05N$, a scenario in which $q = 5\%$ of the jobs are routed to the shortest queue and a scenario with $p = q = 2\%$. To avoid the dependence in B, the computation are done with $B = 10^5$ which is equivalent to $B = \infty$ in practice.



(a) Average response time as a function of λ.



(b) Probability s_i for a server to have i jobs or more.

Figure 5: Average response time and steady state distribution of occupancy for the model of parallel servers of Figure 4. The four curves corresponds to different parameters: in blue, $p = q = q$ (N independent M/M/1 queues); in black: $p = 0.05, q = 0$ (model of Figure 4(a)); in green: $p = 0, q = 0.05$ (model of Figure 4(b)); in red: $p = q = .02$.

Figure 5(a) shows the average number of jobs per server as a function of the load λ. As pointed out in [32], when $q = 0$, the average number of jobs goes from $\lambda/(1 - \lambda)$ to $O(\log(1/(1 - \lambda))$

which provides a large gain in term of waiting time, even for $p = 5\%$. When $p = 0$ and $B = \infty$, Equation (25) shows that for $i \geq 1$, $s_i = (\lambda(1-q))^{i-1}$. Thus, the average number of jobs is $\lambda/(1-\lambda(1-q))$ which is bounded by $1/q$ regardless of the load. This shows that when the load is high, a judicious routing of the packets decreases the average response time more efficiently than adding a centralized server.

On Figure 5(b) is reported the distribution s_i as a function of i for a highly loaded system $\lambda = .99$. When $p > 0$ and $B = \infty$, the constant β of Eq.(25) is negative and there exists $i^* = \lfloor \log_{\lambda \frac{1-q}{1-p}}(-\beta/\alpha) \rfloor$ such that the probability for a server to have more than i^* jobs goes to zero as N goes to infinity. For example, Figure 5(b) shows that when $\lambda = .99$ and $(p, q) = (.05, 0)$ (or $p = q = 2\%$), then $i^* = 40$ (or $i^* = 41$): there are almost no queues with more than i^* jobs. However, when $p = 0$, $\beta \geq 0$ and $s_i > 0$ for all i. This shows that to avoid big queues, adding a centralized server helps more. Both figures show that adding both a centralized server and a judicious routing, even for the very small values $p = q = 2\%$ allows one to get the better of the two worlds: a low response time and a tail distribution equal to zero.

4.4 Volunteer Computing and boundary constraints

Here, we consider a model of a volunteer computing system, such as BOINC http://boinc.berkeley.edu/. This model is less schematic than the previous one and show how our framework can be used to accelerate the numerical simulation of such system: at the limit, we only have to integrate numerically a differential inclusion, which can be done very efficiently [1].

The system is composed of a single buffer and N desktop machines, offered by their owners (volunteers), that serve the packets of this buffer. However, as soon as the owner of a processor wants to use it, she preempts it and the processor becomes unavailable for the computing system. As for the incoming packets, they are assumed to arrive in the buffer according to a Poisson process at rate λ. These kinds of systems are often called push/pull models: The distributed applications *push* jobs to a central server that stores them in a buffer and whenever a processor becomes available, it *pulls* a job from the buffer and executes it.

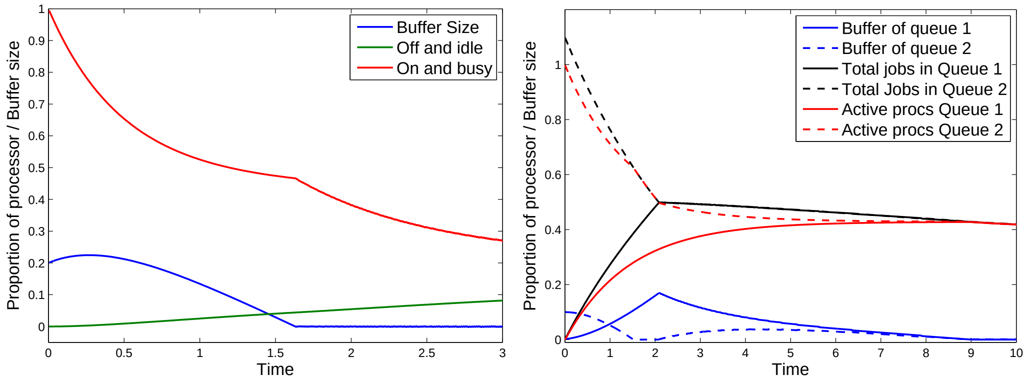
Such systems fit into our density dependent population process framework. The context $C(t)$ represents the size of the buffer while the N objects represent both the applications sending jobs and the hosts executing them. The state of a host is its availability and its idleness (whether it is executing a job or not). The non-smooth part of the dynamics comes from the buffer size. When $C(t) > 0$, if a host asks for a job, it gets it with probability one while when $C(t) = 0$, a host asking for a job will get nothing. In that case, one can show that this dynamics satisfies the conditions of Theorem 5 that can be used to study the limiting behavior of the system when the number of hosts and applications grows.

In the simplest case, the intensity of the system is $\gamma^N = 1/N$ and an application sends a job to the system at rate λ while jobs are completed at rate μ by each server. To represent the communication delays, every host gets jobs at rate γ. It becomes unavailable with rate p_u, and available with rate p_a if $C(t) > 0$ and 0 otherwise. If b, a, u denote respectively the proportion of busy, available and unavailable hosts, the limiting system is described by a DI:

$$\begin{aligned}
\dot{b}(t) &= -\mu b(t) + \gamma a(t)\mathbf{1}_{C(t)>0} \\
\dot{a}(t) &= \mu(t)b(t) + p_a u(t) - p_a a(t) - \gamma a(t)\mathbf{1}_{C(t)>0} \\
\dot{u}(t) &= -p_a u(t) + p_u a(t) \\
\dot{C}(t) &= -\gamma a(t)\mathbf{1}_{C(t)>0} + \lambda \mathbf{1}_{C(t)<C_{\max}}.
\end{aligned}$$

The formal DI is obtained by replacing $a(t)\mathbf{1}_{C(t)>0}$ by the singleton $\{\gamma a(t)\}$ if $C(t) > 0$ and the interval $[0; \gamma a(t)]$ when $C(t) = 0$.

The behavior of the system is represented in Figure 6(a). At time $t = 0$, we consider that the size of the buffer is $C(0) = .2$ and that all processors are available and are serving a job. One can see that there is a point of non-differentiability in the behavior of the system when the size of the buffer reaches 0. For this example, we used a forward Euler discretization of the differential inclusion for numerical integration , which is very simple but also very accurate in this case. In



(a) Initial model of volunteer computing (b) Model of volunteer with two clusters and JSQ

Figure 6: Limit dynamics of a volunteer computing system. A non-differentiable point occurs when the buffer becomes empty.

more complex examples, other solutions exist to improve the accuracy of numercial integration of differential inclusions [1]. One of the main limitation of our method is that it cannot be applied to study the behavior of one resource. For example, the proportion of time when a server is busy can be easily derived from $b(t)$ but our analysis does not allow us to compute the length of a busy period. This limitation could be overcome by studying limiting properties of the individual behavior of objects, (*e.g.* following Sec 4.4 of [13]).

Figure 6(b) depicts a simulation of a model with two identical time-homogeneous volunteer systems. Each time a packet arrives, it is routed to the system with the smallest number of packets. Here, the scheduling of packets introduces a new cause of non-smoothness: there is a threshold in the dynamics of the system when both backlogs are equal. Figure 6(b) shows the behavior of the limit differential inclusion. Once again, the limit behavior is unique once the initial condition is given. As expected, new non-differential points appear when both buffers are equal.

4.4.1 Remark on the OSL condition

Let $y = (b, a, u, C)$ and $\bar{y} = (\bar{b}, a, u, 0)$, then $\left\langle y - \bar{y}, f(y) - f(\bar{(y)})\right\rangle = -\mu|b-\bar{b}|^2 + \gamma a(b-\bar{b}) - \gamma a$. If f were OSL, this would be less than $L\left\|y-\bar{y}\right\|^2$. However, when $b - \bar{b}$ is small enough and positive, this expression is of order $\gamma a(b-\bar{b})$ which is greater than $L\|y-\bar{y}\|^2 = L(|b-\bar{b}|^2 + C^2)$. In fact, there are two types of non-smoothness in these equations. The first one is that the dynamics of C depends on C in a discontinuous manner but in a OSL way. The second type of discontinuity is that the dynamics of b depends on C in a discontinuous manner. This latter discontinuity leads to a term of order $(b-\bar{b})$ which is greater than $L\left\|b-\bar{b}\right\|^2$ whenever $b-\bar{b}$ is small enough.

5 Conclusion and future work

In this paper, we studied the asymptotic properties of a family of Markov processes Y^N evolving on subset of $\mathbb{R}^d$. We showed that if their drift f^N vanishes as N goes to infinity, then the behavior of Y^N converges to the set of solutions of a deterministic differential inclusion $\dot{y} \in F(y)$. In particular, this result holds even if f^N does not converge to a single-valued function f as N grows. Using this result, we developed two applications. We first show how to prove stability results using a fluid approximation described by a differential inclusion. Then, we show how to handle discontinuities that arise in mean field models due to centralized actions or boundary conditions. The examples provided illustrate that one can easily retrieve results from the literature (§3.3 or §3.4), but also extend existing models (§4.3) or develop new examples (§4.4).

Several perspectives remain open. First, a natural extension of the mean field model would be to obtain properties on the individual behavior of objects. We believe similar results as the ones of Sec 4.4 of [13] could be adapted to our case to show that if differential inclusion has a unique solution y, the behavior of a collection k objects is asymptotically a continuous time inhomogeneous Markov chain with k independent components with kernel at time t depending on $y(t)$. A second important question concerns the quality of the approximation of the steady-state distribution. Theorem 9 shows that if the differential inclusion has unique attractor, then the stationary distribution concentrates on this point, it does not provide a bound on the speed of convergence. Simulations on our examples indicates that this convergence occurs at rate $1/\sqrt{N}$ but this remains a conjecture. Finally, checking the applicability of the OSL condition is an open issue. In all our examples, the original drift is not OSL but for most of them, we were able to find a change of variable in which the dynamics was OSL. It would be helpful to find a more direct way to show if a dynamic can be transformed in an OSL dynamics or a simpler condition to check to guarantee the speed of convergence.

Acknowledgment

The authors would like to thank David McDonald for his precious advice and his careful proof reading of the paper.

A Differential inclusions

In this appendix, we recall the main concepts on differential inclusions. For a more complete description, the reader is referred to [1]. In all that follows, $\langle x, y\rangle$ denotes the classical inner-product on $\mathbb{R}^d$ and $\|x\| = \sqrt{\langle x, x\rangle}$ (L^2 norm) and for a set $A \subset \mathbb{R}^d$, $\|A\| = \sup_{x\in A} \|x\|$.

Definition 11. *Consider a differential inclusion problem:*

$$\dot{y}(t) \in F(y(t)), \quad y(0) = y_0,$$

where F is a set-valued function mapping each point $y \in \mathbb{R}^d$ to a set $F(y) \subset \mathbb{R}^d$. Let $I \subset \mathbb{R}$ be an interval with $0 \in I$. A function $y : I \to \mathbb{R}^d$ is a solution of the DI $\dot{y} \in F(y)$ with initial condition $y(0) = y_0$ if there exists a function $\varphi : I \to \mathbb{R}^d$ such that:

(i) for all $t \in I$: $y(t) = y_0 + \int_0^t \varphi(s)ds$;

(ii) for almost every (a.e.) $t \in I$: $\varphi(t) \in F(y(t))$.

In particular, *(i)* is equivalent to saying that y is absolutely continuous. *(i)* and *(ii)* imply that y is differentiable at almost every $t \in I$ with $\dot{y}(t) \in F(y(t))$.

Definition 12 (Upper Semi-Continuous (USC))**.** *The function F is upper semi-continuous (USC) if for any $y \in \mathbb{R}^d$, $F(y)$ is a non-empty closed, convex and bounded set and if for any open set O containing $F(y)$, there exists a neighborhood V of y such that $F(V) \subset O$.*

Definition 13 (One-Sided Lipschitz (OSL))**.** *A set-valued function F is one-sided Lipschitz (OSL) with constant L if for all $y, \bar{y} \in \mathbb{R}^d$ and for all $u \in F(y)$ $\bar{u} \in F(\bar{y})$:*

$$\langle y - \bar{y}, u - \bar{u}\rangle \le L \|y - \bar{y}\|^2 .$$

These two definitions give sufficient conditions for the existence (resp. uniqueness) of solutions for the differential inclusion. We recall the following results.

Proposition 14 (Theorems 2.2.1 and 2.2.2 of [22])**.**

- *If F is USC and if there exists c such that $\|F(x)\| \le c(1 + \|x\|)$ then for any initial condition y_0, $\dot{y} \in F(y)$ has at least one solution on $[0; \infty)$ with $y(0) = y_0$.*
- *If F is OSL, then for all $T > 0$, there exists at most one solution of $\dot{y} \in F(y)$ on $[0; T]$.*

Of course (USC) and (OSL) combined ensure that the DI has a unique solution.

B Proofs of Theorem 1 and Theorem 4

This section is devoted to the proof of Theorems 1 and 4. We first recall some notation of Section 2 before jumping into the proofs.

Let us recall that Y^N is defined by

$$Y^N(k+1) \quad = \quad Y^N(k) + \gamma^N \left(f^N(Y^N(k)) + U^N(k+1) \right). \tag{26}$$

This equation can be seen as an Euler discretization of the DI $\dot{y} \in F(y)$ plus two error terms:

- A random error term caused by $U^N(k+1)$ which is such that $\mathbb{E}\left(U^N(k+1) \middle| Y^N(k)\right) = 0$ and is either uniformly integrable (Theorem 1) or bounded in second moment (Theorem 4).
- A "deterministic" error term coming from the fact that $f^N(y)$ is not necessarily in $F(y)$ but converges to F in the sense of Equation (5) (see also Lemma 15):

$$F(y) \stackrel{\text{def}}{=} \text{conv}\left(\underset{N\to\infty}{\text{acc}} f^N(y^N) \ \text{ with } \ \lim_{N\to\infty} y^N = y \right).$$

Equation (26) is called a *stochastic approximation* algorithm with *constant step size* associated with the DI (6). The term *constant step size* comes from the fact that γ^N does not vary with time. Both proofs of Theorem 1 and Theorem 4 are based on the convergence of such stochastic approximation (26) as N goes to infinity. However, the two proofs are radically different. The first one is based on compactness argument while the second one focuses on computing explicit error terms.

B.1 Proof of Theorem 1

The classical approach to prove convergence of a stochastic approximation to the solution of the associated differential system uses Gronwall's lemma [14]. Here, we use a different approach, based on compactness properties of the trajectories of the stochastic system. This proof is inspired by several results on differential inclusions, in particular the proof of Theorem 2.2.1 of [22]. However, it is different from Theorem 4.2 of [3] since we need to deal with *constant step sizes* instead of vanishing step sizes (often easier) and we are interested in the convergence over a finite time-horizon. Also, we do not need any *a priori* assumption on the boundedness of the stochastic process.

The idea of the proof is to show that for any sub-sequence of $\bar{Y}^N$, there exists a sub-sequence $\bar{Y}^{\sigma(N)}$ (of this sub-sequence) such that the distance between $\bar{Y}^{\sigma(N)}$ and the set of solution of the differential inclusion $\mathscr{S}_T(y_0))$ goes to 0 almost surely. In all that follows, let $\bar{Y}^{\sigma(N)}$ be a sub-sequence of $\bar{Y}^N$. In order to simplify the notations and because we will take several sub-sequences of sub-sequences, we omit the σ in the notation and we denote all sub-sequences by $\bar{Y}^N$. In the first part of the proof, we consider the problem from a probabilistic point of view to make sure that the random part of the process goes almost surely to 0. Then we consider the problem from a trajectorial point of view using analytic arguments.

We first start with two technical lemmas that show that f^N converges to F uniformly on all compact:

Lemma 15. *Let f^N be such that $\left\|f^N(y)\right\| \leq c(1 + \|y\|)$. Let F be defined by Equation (5) and for all $\varepsilon > 0$, define F^ε by:*

$$F^\varepsilon(y) = \left\{ z \quad \text{s.t } \exists u \in \mathbb{R}^d, \exists v \in F(u) \text{ with } \|u - y\| \leq \varepsilon \wedge \|v - z\| \leq \varepsilon \right\}. \tag{27}$$

Then:

(i) for all compact $K \subset \mathbb{R}^d$, there exists a sequence $\delta^N \to 0$ such that, for all $N \geq N_0$ and for all $y \in K$: $f^N(y) \in F^{\delta^{N_0}}(y)$.

(ii) F is USC, i.e., *for all y: $\bigcap_{\varepsilon>0} F^\varepsilon(y) \subset F(y)$.*

Proof. We prove *(i)* by contradiction. Assume that *(i)* does not hold. Then, there exists a compact K and $\varepsilon > 0$ such that for all N_0, there exists $N > N_0$ with $y_N \in K$ and $y_N \notin F^\varepsilon(y_N)$. Since K is compact, there exists a sub-sequence of y_N that converges to some y. This implies that for N large enough, $\|y_N - y\| \leq \varepsilon$. Since we assumed that $f^N(y^N) \notin F^\varepsilon(y_N)$ and by definition of $F^\varepsilon(y^N)$, this implies that for all $v \in F(y)$, $\left\|f^N(y^N) - v\right\| \geq \varepsilon$. This contradicts the definition of $F(y)$ which contains the set of limit points of $f^N(y^N)$.

Proof of (ii). Let $v \in \bigcap_{\varepsilon>0} F^\varepsilon(y)$. This implies that there exists a sequence $y_k \to y$ with $v_k \in F(y_k)$ and $v_k \to v$. By definition of F, v_k is a convex combination of points $\{w_{k,\ell}\}_\ell$ with $f^N(y^N_{k,\ell}) \to w_{k,\ell}$ and $y^N_{k,\ell} \to y_k$. By setting $z_{N,\ell} = y^N_{N,\ell}$, we have $z_{N,\ell} \to y$ and $f^N(z_{N,\ell})$ converges to $w_\ell = \lim_{k\to\infty} w_{k,\ell}$. This shows that $w_\ell \in F(y)$. Therefore, any convex combination of w_ℓ also belongs to $F(y)$. □

Lemma 16. *Let $(U^N(.))_{k\geq 0}$ be a uniformly integrable martingale difference sequence with respect to a filtration $\{\mathcal{F}_k\}$ and let γ^N be a sequence with $\gamma^N \to 0$. Then for all $T > 0$,*

$$\sup_{0\leq t\leq T} \left\| \gamma^N \sum_{k=0}^{T/\gamma^N} U^N(k) \right\| \xrightarrow{\mathcal{P}} 0.$$

Proof. Let $\varepsilon, \nu > 0$ and let $V^N(i) = \sum_{k=0}^{i} U^N(k)$. We prove that for N large enough, we have $\mathcal{P}(\sup_{0\leq i\leq T/\gamma^N} \left\|V^N(i)\right\| \geq \varepsilon) \leq \nu$.

Let $\delta = \nu\varepsilon/8$. As U^N is uniformly integrable, there exists R such that $\mathbb{E}(U^N(k)\mathbf{1}_{U^N(k)\geq R}) \leq \delta$. Define $V^N_+(k)$ and $V^N_-(k)$ as:

$$\begin{aligned} V^N_+(k) &= U^N(k)\mathbf{1}_{U^N(k)\geq R} - \mathbb{E}\left(U(k)\mathbf{1}_{U^N(k)\geq R} \mid \mathcal{F}_{k-1}\right) \\ V^N_-(k) &= U^N(k)\mathbf{1}_{U^N(k)< R} - \mathbb{E}\left(U(k)\mathbf{1}_{U^N(k)< R} \mid \mathcal{F}_{k-1}\right) = U^N(k) - U^N_+(k) \end{aligned}$$

Applying Kolmogorov's inequality for martingales, we get:

$$\begin{aligned} \mathcal{P}\left(\sup_{0\leq i\leq T} \left\|V^N(i)\right\| \geq \varepsilon\right) &\leq \mathcal{P}\left(\sup_{0\leq i\leq T} \left\|\gamma^N \sum_{k=0}^{i} U^N_-(k)\right\| \geq \frac{\varepsilon}{2}\right) + \mathcal{P}\left(\sup_{0\leq i\leq T} \left\|\gamma^N \sum_{k=0}^{i} U^N_+(k)\right\| \geq \frac{\varepsilon}{2}\right) \\ &\leq \frac{4}{\varepsilon^2}\mathbb{E}\left(\left\|\gamma^N \sum_{k=0}^{T/\gamma^N} U^N_-(k)\right\|^2\right) + \frac{2}{\varepsilon}\mathbb{E}\left(\left\|\gamma^N \sum_{k=0}^{T/\gamma^N} U^N_+(k)\right\|\right) \\ &\leq 16\frac{R^2}{N\varepsilon^2} + \frac{4\delta}{\varepsilon} \leq 16\frac{R^2}{N\varepsilon^2} + \frac{\nu}{2}. \end{aligned}$$

Therefore, for all $N \geq 32R^2/(\varepsilon^2\nu)$, this quantity is less than ν. □

Developing the recurrence (4), the value of $Y^N(k+1)$ is equal to:

$$Y^N(k+1) = Y^N(0) + \sum_{i=0}^{k} \gamma^N f^N(Y^N(i)) + \gamma^N \sum_{i=0}^{k} U^N(i+1). \tag{28}$$

We define two functions $Z^N(t)$, and $V^N(t)$ to be piecewise linear functions such that for all $t = k\gamma^N$, $Z^N(t) = Y^N(0) + \sum_{i=0}^{k-1} \gamma^N f^N(Y^N(i))$ and $V^N(t) = \sum_{i=0}^{k-1} \gamma^N U^N(i+1)$.

By Lemma 16, since U^N is a martingale difference sequence uniformly integrable, $\sup_{0\leq t\leq T} \left\|V^N(t)\right\|$ converges in probability to 0. Therefore, there exists a sub-sequence of V^N such that $\sup_{t\leq T} \left\|V^N(t)\right\|$ converges almost surely to 0.

We now reason from a trajectorial point of view. Let us now consider a trajectory $\omega \in \Omega$ of the system such that $\sup_{t\leq T} \left\|V^N(t)\right\|$ converges to 0. In particular, this implies that $\left\|V^N(t)\right\|$ is bounded for all N and t: $\sup_{N,0\leq t\leq T}\left\|V^N(t)\right\| \leq d < \infty$. Using (28) and since $\left\|f^N(y)\right\| \leq c(1+\|y\|)$, for all $k \leq T/\gamma^N$, $\left\|Y^N(k+1)\right\|$ can be bounded by:

$$\begin{aligned}
\left\|Y^N(k+1)\right\| &\leq \left\|Y^N(0)\right\| + \sum_{i=0}\gamma^N c(1+\left\|Y^N(i)\right\|) + \sup_{N,t}\left\|V^N(t)\right\| \\
&\leq \left\|Y^N(0)\right\| + ck\gamma^N + d + \sum_{i=0}^{k}\gamma^N\left\|Y^N(i)\right\| \\
&\leq \left(\left\|Y^N(0)\right\| + cT + d\right)\exp\left(cT\right)/c, \qquad (29)
\end{aligned}$$

where we used the discrete Gronwall's lemma and the fact that $k\gamma^N \leq T$.

Once we know that $\sup_{N,0\leq t\leq T}\left\|Y^N(t)\right\|$ is bounded, the rest of the proof can be adapted from classical results on the convergence of the Euler approximation for differential inclusions, see [22] for example. There exists $e > 0$ such that $\sup_{N,0\leq t\leq T}\left\|Y^N(t)\right\| \leq e$. Thus $\left\|f(Y^N(k))\right\| < c(1+e) < \infty$. This shows that the functions Z^N are Lipschitz with constant $c(1+e)$. Thus the sequence of functions $(Z^N)_N$ are equicontinuous and bounded. Therefore by the Arzéla-Ascoli theorem, for all sub-sequences of $(Z^N)_N$, there exists a sub-sequence that converges to some $z : [0;T] \to \mathbb{R}^d$. In the following, we will show that z is a solution of (3) which shows that the distance between Z^N and the set of solutions $\mathscr{S}_T(y_0))$ goes to 0 as N goes to infinity. As $\left\|Z^N - Y^N\right\| = \left\|V^N\right\| \to 0$, this implies that the distance between Y^N and $\mathscr{S}_T(y_0)$ goes to 0. To prove this, we will construct a function φ such that:

(i) for all t: $z(t) = z(0) + \int_0^t \varphi(s)ds$;

(ii) for almost every t: $\varphi(t) \in F(z(t))$.

Let $\varphi^N(t)$ be a step function, constant on the intervals $[k\gamma^N, (k+1)\gamma^N)$ and such that for $t = k\gamma^N$, $\varphi^N(t) = f(Y^N(k))$. Therefore, the sequence φ^N is bounded in $L_2([0;T],\mathbb{R}^d)$. Thus, there exists a sub-sequence of φ^N converging weakly in L_2 to a function φ. Since L_2 is a reflexive space, if a sequence of functions φ^N converges to φ, this means that for all functions v, there exists a sub-sequence of φ^N such that $\langle v, \varphi^N\rangle \to \langle v, \varphi\rangle$. Let $\xi \in \mathbb{R}^d$ and $t \in [0;T]$. Let the function v be defined by $v(s) \stackrel{\text{def}}{=} \xi$ for $s < t$ and $v(s) \stackrel{\text{def}}{=} 0$ for $t \geq s$. Since φ^N converges weakly to φ and $Z^N(t) \to z(t)$, we have:

$$\begin{aligned}
\left\langle Z^N(t), \xi\right\rangle &\to \left\langle z(t), \xi\right\rangle; \\
\left\langle Z^N(t), \xi\right\rangle &= \left\langle Z^N(0), \xi\right\rangle + \left\langle \int_0^t \varphi^N(s)ds, \xi\right\rangle \\
&= \left\langle Z^N(0), \xi\right\rangle + \left\langle \varphi^N, v\right\rangle \\
&\to \left\langle z(0), \xi\right\rangle + \left\langle \varphi, v\right\rangle \\
&= \left\langle z(0) + \int_0^t \varphi(s)ds, \xi\right\rangle.
\end{aligned}$$

As this is true for all $\xi \in \mathbb{R}^d$, this shows that z is absolutely continuous: $z(t) = \int_0^t \varphi(s)ds$.

It remains to show that for a.e. t, $\varphi(t) \in F(z(t))$. Let t^N denote the greater multiple of γ^N less than t ($t^N \stackrel{\text{def}}{=} \lfloor t/\gamma^N\rfloor\gamma^N$). Using that $f^N(Y^N(k)) \leq c(1+e)$ and that z^N converges uniformly to y, for all $\delta > 0$, there exists N_0 such that $N \geq N_0$ implies $\left\|z(t) - Y^N(t^N)\right\| \leq \delta$. By Lemma 15*(i)*, this shows that for N large enough, $\varphi^N(t) \in F^{2\delta}(z(t))$. Since F is convex and z bounded, $\{\alpha \in L^2 : \alpha(t) \in F^\delta(z(t))\}$ is convex and closed. This shows that this set is weakly closed (see [28], Theorem 3.12). Therefore, for all t, $\varphi(t) \in F^\delta(t)$. As this is true for all δ and because of Lemma 15*(ii)*, this shows $\varphi(t) \in \cap_{\delta>0}F^\delta(t) = F(z(t))$. Thus, z is a solution of the DI.

□

B.2 Proof of Theorem 4

The constants A_T, B_T and C_T of Theorem 4 are given by:

$$A_T = M_T \left(M_T^2 + \frac{14M_T}{3} + 2K_T \right)$$
$$B_T = 2M_T^2 + 4L\delta^N + 12K_T$$
$$C_T = 2M_T^2 + 4L\varepsilon + 8K_T,$$

with $K_T = \left(\max\left\{\left\|Y^N(0)\right\|, \|y(0)\|\right\} + (cT+\varepsilon)\right)e^{cT}/c$ and $M_T = \sup_{0\le t\le T} f^N(Y^N(t)) \le c(1 + K_T)$. If $F(.)$ is bounded by some M, the constant M_T is just M and is in particular independent of T. This is true for example if Y^N is constrained to stay in a compact space of $\mathbb{R}^d$ or if the drift is bounded for all $y \in \mathbb{R}^d$. The existence of the sequence δ^N is given by the definition of F in Equation (5) (see Lemma 15*(i)*).

By definition, $Y^N(k+1)$ can be written:

$$Y^N(k+1) = Y^N(0) + \gamma^N \sum_{i=0}^{k} f^N(Y^N(i)) + \gamma^N \sum_{i=0}^{k} U^N(i+1). \tag{30}$$

Let us define two random sequences Z and V by:

$$Z(k) \stackrel{\text{def}}{=} Y^N(0) + \gamma^N \sum_{i=0}^{k} f^N(Y^N(i)) \quad \text{and} \quad V(k) \stackrel{\text{def}}{=} \gamma^N \sum_{i=0}^{k} U^N(i+1).$$

We first start with two lemmas. The first one shows that $V(k)$ is small while the second one computes bounds on the growth of Y^N and the solution of the DI y.

Lemma 17. *For all T and all $\varepsilon > 0$,*

$$\mathcal{P}\left(\sup_{i \le T/\gamma^N} \left\|V^N(i)\right\| \ge \varepsilon \right) \le \frac{\gamma^N T}{\varepsilon^2}.$$

Proof. Since $\mathbb{E}(U^N(k+1) \mid Y^N(k)) = 0$ and $\mathbb{E}(\left\|U^N(k+1)\right\|^2 \mid Y^N(k)) \le b$, we have $\mathbb{E}(\|V(k)\|^2) \le k{\gamma^N}^2 b \le Tb\gamma^N$ for all $k \le T/\gamma^N$. Applying Kolmogorov's inequality for martingales to the martingale V leads to the bound of the lemma. □

Lemma 18. *Let Y^N be a sequence satisfying (30) with $\left\|f^N(y)\right\| \le c\,\|1 + \|y\|\|$. Let y denote the solution of the differential equation associated with F.*

Then, if $\sup_{i\le k}\left\|V^N(i)\right\| \le \varepsilon$, there exists a constant K_T such that

$$\max\left\{ \sup_{0\le k\le T/\gamma^N} \left\|Y^N(k)\right\|, \sup_{0\le t\le T} \|y(t)\| \right\} \le K_T.$$

The constant K_T is given by:

$$K_T \stackrel{\text{def}}{=} \left(\max\left\{\left\|Y^N(0)\right\|, \|y(0)\|\right\} + (cT+\varepsilon)\right)e^{cT}/c.$$

Proof. By definition of $Y^N(k+1)$, we have:

$$\left\|Y^N(k+1)\right\| \le \left\|Y^N(0)\right\| + \gamma^N \sum_{i=0}^{k} c(1 + \left\|Y^N(i)\right\|) + \varepsilon$$
$$= \left\|Y^N(0)\right\| + k\gamma^N c + \varepsilon + \gamma^N c \sum_{i=0}^{k} \left\|Y^N(i)\right\|.$$

Therefore, by the discrete Gronwall's lemma, we have $\left\|Y^N(k)\right\| \leq (\left\|Y^N(0)\right\| + (cT+\varepsilon)e^{cT}/c$ for $k \leq T/\gamma^N$.

The proof for y is similar, replacing the discrete Gronwall's inequality by the continuous Gronwall's inequality. □

Let $T>0$ and $\varepsilon>0$. Assume that $\|V^N(k)\| \leq \varepsilon$ for all $k \leq T/\gamma^N$ and let K_T be defined as in Lemma 18. Since F is OSL, there exists a unique solution y of the DI $\dot{y} \in F(y)$ with $y(0)=y_0$. Therefore, $y(t) = y(0) + \int_0^t f(s)ds$ with $f(s) \in F(y(s))$ *a.e.*

Let $k \leq T/\gamma^N$ and denote $t_N = k\gamma^N$.

$$\begin{aligned}
\left\|Z^N(k+1) - y(t_N+\gamma^N)\right\|^2 &= \left\|Z^N(k) - y(t_N) + \int_0^{\gamma^N} f^N(Y^N(k)) - f(t_N+s)ds\right\|^2 \\
&= \left\|Z^N(k) - y(t_N)\right\|^2 + \left\|\int_0^{\gamma^N} f^N(Y^N(k)) - f(t_N+s)ds\right\|^2 \\
&\quad + \int_0^{\gamma^N} 2\left\langle Z^N(k) - y(t_N), f^N(Y^N(k)) - f(t_N+s)\right\rangle ds \\
&\leq \left\|Z^N(k) - y(t_N)\right\|^2 + {\gamma^N}^2 4M_T^2 + 2\int_0^{\gamma^N} w(s)ds,
\end{aligned}$$

where $w(s) \stackrel{\text{def}}{=} \left\langle Z^N(k) - y(t_N), f^N(Y^N(k)) - f(t_N+s)\right\rangle$. To prove the last inequality, we used Lemma 18 that shows that $\left\|Y^N(k)\right\|$ and $\|y(k)\|$ are bounded by K_T. Therefore, there exists a constant M_T such that $\left\|f^N\right\|$ and $\|f\|$ are bounded by M_T.

Because of Lemma 15 that guarantees the speed of convergence of f^N to F , there exists $u \in \mathbb{R}^d$ and $v \in F(v)$ with $\left\|u - Y^N(k)\right\| \leq \delta^N$ and $\left\|v - f^N(Y^N(k))\right\| \leq \varepsilon$. Thus, $w(s)$ is equal to:

$$\begin{aligned}
w(s) &= \left\langle Z^N(k) - u + u - y(t_N+s) + y(t_N) - y(t_N+s), f^N(Y^N(k)) - v + v - f(t_N+s)\right\rangle \\
&= \left\langle Z^N(k) - u + y(t_N) - y(t_N+s), f^N(Y^N(k)) - f(t_N+s)\right\rangle \\
&\quad + \left\langle u - y(t_N+s), f^N(Y^N(k)) - v\right\rangle + \left\langle u - y(t_N+s), v - f(t_N+s)\right\rangle,
\end{aligned}$$

where we expanded the inner product using $\langle a+b+c, d+e\rangle = \langle a+c, d+e\rangle + \langle b,d\rangle + \langle b,e\rangle$.

By assumption on u and V, one has $\|Z^N(k) - u\| \leq \|Z^N(k) - Y^N(k)\| + \|Y^N(k) - u\| \leq \varepsilon + \delta^N$. Moreover, since $\|f\| \leq M_T$, one has $\|y(t_N) - y(t_N+s)\| \leq sM_T$. Combining with the fact that F is OSL of constant L, this gives:

$$w(s) \leq (\varepsilon + \delta^N + sM_T)M_T^2 + 2K_T\delta^N + L\left\|u - y(t_N+s)\right\|^2.$$

Finally, $\|u - y(t_N+s)\|^2$ can be bounded by:

$$\begin{aligned}
\|u - y(t_N+s)\|^2 &= \left\|u - Z^N(k)\right\|^2 + \left\|Z^N(k) - y(t_N)\right\|^2 + \|y(t_N) - y(t_N+s)\|^2 \\
&\quad + 2\left\langle u - Z^N(k), Z^N(k) - y(t_N+s)\right\rangle + 2\left\langle Z^N(k) - y(t_N), y(t_N) - y(t_N+s)\right\rangle \\
&\leq \left\|Z^N(k) - y(t_N)\right\|^2 + (\delta^N+\varepsilon)^2 + s^2M_T^2 + 2(\delta^N+\varepsilon)2K_T + 2K_TsM_T.
\end{aligned}$$

This shows that $\int_0^{\gamma^N} w(s)ds$ can be bounded by γ^N times:

$$\begin{aligned}
L\left\|Z^N(k) - y(t_N)\right\|^2 + (\varepsilon + \delta^N + \frac{\gamma^N}{2}M_T)M_T^2 + 2K_T\delta^N + L(\delta^N+\varepsilon)^2 + \frac{{\gamma^N}^2 M_T^2}{3} \\
+2(\delta^N+\varepsilon)2K_T + K_T\gamma^N M_T.
\end{aligned}$$

Therefore, ${\gamma^N}^2 4M_T^2 + 2\int_0^{\gamma^N} w(s)ds$ is bounded by $2L\left\|Z^N(k) - y(t_N)\right\|^2$ plus γ^N times

$$\gamma^N M_T\left(M_T^2 + \frac{14M_T}{3} + 2K_T\right) + \delta^N\left(2M_T^2 + 4L\delta^N + 12K_T\right) + \varepsilon\left(2M_T^2 + 4L\varepsilon + 8K_T\right).$$

If a sequence a_k satisfies $a_{k+1} \leq (1+2\gamma^N L)a_k + b$ with $L \neq 0$, one has:

$$a_k = (1+2\gamma^N L)^k a_0 + \frac{(1+2\gamma^N L)^k - 1}{2\gamma^N L} b \leq \mathbf{e}^{2L\gamma^N k} a_0 + \frac{\mathbf{e}^{2L\gamma^N k}}{2\gamma^N L} b.$$

If $L = 0$ and a_k satisfies the recurrence, then $a_k \leq a_0 + kb$. This concludes the proof of the theorem.

B.3 Proof of Theorem 5

Since $\tau < \infty$, the rate of transition of $D^N(.)$ is bounded by $N\tau$. Using uniformization of continuous time Markov chain (see [29] for example), there exists a Poisson process Λ^N of rate $N\tau$ and a discrete time Markov chain $Y^N(.)$ such that $D^N(t) = Y^N(\Lambda^N(t))$ and Y^N and Λ^N are independent. Moreover, for all y and $\ell \in \mathcal{L}$,

$$\begin{aligned}\mathcal{P}\left(Y^N(k+1) = y + \frac{\ell}{N} | Y^N(k) = y\right) &= \frac{1}{\tau}\beta_\ell(y),\\ \mathcal{P}\left(Y^N(k+1) = y | Y^N(k) = y\right) &= 1 - \frac{1}{\tau}\sum_{\ell\in\mathcal{L}}\beta_\ell(y).\end{aligned}$$

For all $y \in \mathbb{R}^d$, the drift of $Y^N(.)$ is $\mathbb{E}\left(Y^N(k+1) - Y^N(k) | Y^N(k) = y\right) = (N\tau)^{-1} f(y)$ and $Y^N(k+1)$ can be written $Y^N(k+1) = Y^N(k) + (N\tau)^{-1}(f(y) + U^N(k+1))$. By assumption $\sum_{\ell\in\mathcal{L}} \|\ell\| \sup_y \beta_\ell(y) < \infty$, U^N is uniformly integrable. Therefore, $Y^N(k)$ satisfies the conditions of Theorem 1. This shows that $\inf_{y\in\mathscr{S}_T(y_0)} \sup_{t\leq T} \left\|Y^N(tN) - y(t)\right\| = 0$.

As Λ^N is a Poisson process of rate $N\tau$, $|\Lambda^N(t) - tN\tau|^2$ is a sub-martingale and by Doob's inequality ([15] p 250), $\mathcal{P}(\sup_{t\leq T} |\Lambda^N(t) - tN\tau| \geq N\tau\varepsilon) \leq \mathbb{E}(|\Lambda^N(T) - TN\tau|^2)/(N\tau\varepsilon)^2 = (TN\tau)/(N\tau\varepsilon)^2 = T/(N\tau\varepsilon^2)$. If y is a solution of the DI (6) on $[0;T]$, for all $t,s \in [0,T]$, $\|y(t) - y(s)\| \leq c(1+K_T)\,|t-s|$ where K_T is defined in Lemma 18. This shows that if y is a solution of the differential inclusion, with probability greater than $1 - T/(N\tau\varepsilon^2)$, we have:

$$\begin{aligned}\left\|D^N(t) - y(t)\right\| &= \left\|Y^N(\Lambda^N(t)) - y(t)\right\|\\ &\leq \left\|Y^N(\Lambda^N(t)) - y\left(\frac{\Lambda^N(t)}{N\tau}\right)\right\| + \left\|y\left(\frac{\Lambda^N(t)}{N\tau}\right) - y(t)\right\|\\ &\leq \left\|Y^N(\Lambda^N(t)) - y\left(\frac{\Lambda^N(t)}{N\tau}\right)\right\| + c(1+K_T)\varepsilon.\end{aligned}$$

By Theorem 1, for all $\varepsilon > 0$, for all N large enough, there exists a solution y of the DI such that the first term of the last inequality is less than ε.

In the OSL case, since f^N does not depend on N, we have $d(f^N, F) = 0$ and the sequence δ^N is equal to 0. The constant C'_T is given by $C'_T = C_T + c(1+K_T)\varepsilon$ where C_T is the same as in the previous section (B.2): $C_T = 2M_T^2 + 4L\varepsilon + 8K_T$.

References

[1] V. Acary and B. Brogliato. *Numerical methods for nonsmooth dynamical systems: applications in mechanics and electronics*. Springer Verlag, 2008.

[2] U. Ayesta, M. Erausquin, M. Jonckheere, and IM Verloop. Scheduling in a random environment: stability and asymptotic optimality. *Arxiv preprint arXiv:1101.5794*, 2011.

[3] M. Benaim, J. Hofbauer, and S. Sorin. Stochastic approximation and differential inclusions. *SIAM J. Control and Optimization*, 44(1):328–348, 2005.

[4] M. Benaim and J.-Y. Le Boudec. A Class Of Mean Field Interaction Models for Computer and Communication Systems. *Performance Evaluation*, 65(11-12):823–838, 2008.

[5] M. Benaim and J.-Y. Le Boudec. On mean field convergence and stationary regime. *Arxiv preprint arXiv:1111.5710*, 2011.

[6] G. Bianchi. Performance analysis of the IEEE 802.11 distributed coordination function. *IEEE Journal on selected areas in communications*, 18(3):535–547, 2000.

[7] C. Bordenave, D. Mcdonald, and A. Proutière. Random multi-access algorithms: A mean field analysis. In *Proceedings of Allerton conference*. Citeseer, 2005.

[8] C. Bordenave, D. McDonald, and A. Proutière. A particle system in interaction with a rapidly varying environment: Mean field limits and applications. *Networks and Heterogeneous Media (NHM)*, 5(1):31–62, 2010.

[9] S. Borst, M. Jonckheere, and L. Leskelä. Stability of parallel queueing systems with coupled service rates. *Discrete Event Dynamic Systems*, 18(4):447–472, 2008.

[10] F. Bouchut and F. James. One-dimensional transport equations with discontinuous coefficients. *Nonlinear Analysis*, 32(7):891, 1998.

[11] M. Bramson. Stability of queueing networks. *Probability Surveys*, 5(169-345):1, 2008.

[12] J.G. Dai. On positive Harris recurrence of multiclass queueing networks: A unified approach via fluid limit models. *The Annals of Applied Probability*, 5(1):49–77, 1995.

[13] RWR Darling and J.R. Norris. Differential equation approximations for Markov chains. *Probability surveys*, 5:37–79, 2008.

[14] J. Dieudonne. *Foundations of modern analysis*. Academic Press, 1960.

[15] R. Durrett. *Probability: Theory and Examples*. Duxbury Press, 1996.

[16] M. Faure and G. Roth. Stochastic approximations of set-valued dynamical systems: Convergence with positive probability to an attractor. *Arxiv preprint arXiv:0905.1858*, 2009.

[17] G. Fort, S. Meyn, E. Moulines, and P. Priouret. The ODE method for stability of skip-free Markov chains with applications to MCMC. *Annals of Applied Probability*, 18(2):664–707, 2008.

[18] N. Gast and B. Gaujal. A mean field model of work stealing in large-scale systems. *ACM SIGMETRICS Performance Evaluation Review*, 38(1):13–24, 2010.

[19] N. Gast, B. Gaujal, and J.-Y. Le Boudec. Mean field for Markov Decision Processes: from Discrete to Continuous Optimization. *INRIA RR*, 7239, 2010.

[20] C. Graham. Chaoticity on path space for a queueing network with selection of the shortest queue among several. *Journal of Applied Probability*, 37(1):198–211, 2000.

[21] M. Jonckheere and S. Shneer. *Stability of multi-dimensional birth-and-death processes with state-dependent 0-homogeneous jumps*. EURANDOM report 2009-029, 2009.

[22] M. Kunze. *Non-smooth dynamical systems*. Springer Berlin Heidelberg, 2000.

[23] T.G. Kurtz. *Approximation of population processes*. Society for Industrial Mathematics, 1981.

[24] F. Lempio. Euler's method revisited. *Proceedings of the Steklov Institute of Mathematics*, 211(4):429–449, 1995.

[25] R. Mané and S. Levy. *Ergodic theory and differentiable dynamics*, volume 258. Springer-Verlag Berlin, 1987.

[26] S.P. Meyn, R.L. Tweedie, and JL Hibey. *Markov chains and stochastic stability*. Springer London et al., 1993.

[27] M. Mitzenmacher. Studying balanced allocations with differential equations. *Combinatorics, Probability and Computing*, 8(5):473–482, 1999.

[28] W. Rudin. *Functional Analysis*. McGraw-Hill Science, 1991.

[29] W.J. Stewart. *Numerical solution of Markov chains*. CRC, 1991.

[30] A.L. Stolyar. On the stability of multiclass queueing networks: a relaxed sufficient condition via limiting fluid processes. *Markov Processes and Related Fields*, 1(4):491–512, 1995.

[31] A.S. Sznitman. Topics in propagation of chaos. *Ecole d'Eté de Probabilités de Saint-Flour XIX—1989*, pages 165–251, 1991.

[32] J.N. Tsitsiklis and K. Xu. On the power of (even a little) centralization in distributed processing. *ACM Sigmetrics, San Jose*, 2011.

[33] J.N. Tsitsiklis and K. Xu. On the power of (even a little) centralization in distributed processing. *Technical Report*, 2011. http://web.mit.edu/jnt/www/Papers/TXSIG11.pdf.

[34] P. van de Ven, S. Borst, and S. Shneer. Instability of MaxWeight scheduling algorithms. In *INFOCOM 2009, IEEE*, pages 1701–1709. IEEE, 2009.

Centre de recherche INRIA Grenoble – Rhône-Alpes
655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399