Mean field Limit of Non-Smooth Systems and Differential Inclusions

Nicolas Gast
Grenoble University and LIG
51, avenue Jean Kuntzmann
Montbonnont, France
nicolas.gast@imag.fr

Bruno Gaujal
INRIA and LIG
51, avenue Jean Kuntzmann
Montbonnont, France
bruno.gaujal@inria.fr

This paper studies the limiting behavior of a system composed by a large number of objects when the dynamics is non-smooth. Under classical smoothness assumptions, there exist general results that show that the limiting system can be described by a system of ordinary differential equations

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

See [4] and the references therein for examples of such convergence results. In most cases, the limiting function f is assumed to be Lipschitz. This condition limits the applicability of these results in many practical cases, in particular, for systems exhibiting thresholds dynamics or with boundary conditions.

Let us consider a simple queuing system with one buffer and many processors that can serve one packet per unit of time in average. If y denotes the size of the queue, then the average decrease of the size of the queue is one packet per unit of time (under a proper rescaling of time) if the queue is non-empty (*i.e.* $y > 0$) and zero if the queue is empty. This leads to a deterministic limit behavior:

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

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

Actually, most work using mean field for networks do not involve queues or when they do, the number of queues scale with the number of objects (as in [5]), or convergence is obtained using *ad hoc* proofs (see for example [2]).

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

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

where F is a set-valued mapping. Of course a differential inclusion problem may have multiple solutions.

In the following, we will provide generic convergence results that show that under few conditions on the initial system, its behavior converges to the solutions of (3)(Theorem 1). This result is generic 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 gap with the limiting dynamics with explicit bound (Theorem 3). The one-sided Lipschitz condition is satisfied in most systems involving queues.

The rest of the paper is organized as follows. In Section 1, we briefly describe the model and give some examples. We will briefly recall some definitions and properties of DIs in Section 1.1. Section 2 provides the main theoretical results and Section 3 give some application example. Finally, we show in Section 3.2 that these results can also be applied to optimization problems.

1. MEAN-FIELD MODEL

We consider of system of N *objects* evolving in a finite state space $\mathcal{S} = \{1 \ldots S\}$. Time is discrete and the state of object n at time step k is denoted $X_n^N(k)$. The objects all evolve in a common environment, called the *context*. The context at time step k is denoted $C^N(k) \in \mathbb{R}^d$. The state of the global system at time k is $(\mathcal{X}^N(k), C^N(k))$. We denote by $M^N(k)$ the empirical measure associated with the N objects:

$$M^N(k) \stackrel{\text{def}}{=} \frac{1}{N} \sum_{n=1}^{N} \delta_{X_n^N(k)}.$$

Since an object has S possible states, $M^N(k)$ can be represented by a vector with S components, its ith component being the proportion of objects in state i. The system $(M^N(k), C^N(k))_k$ is assumed to be a Markov chain. In particular, this is true if the evolution of the context is deterministic and if the evolution of the system is invariant by any permutation of the N objects. The state space of this Markov chain is included in $\mathbb{R}^{S+d}$. To simplify the notations, we call $Y^N(k) \stackrel{\text{def}}{=} (M^N(k), C^N(k))$.

The expected difference between $Y^N(k+1)$ and $Y^N(k)$ is called the *drift* and is denoted $f^N(y)$:

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

We assume that as N grows, the drift vanishes with speed $I(N)$. For all N, $I(N) > 0$ is called the *intensity* of the model and we assume that $\lim_{N \to \infty} I(N) = 0$. We further assume that there exists a function f defined on $\mathbb{R}^{S+d}$ such

that

$$f^N(y) = I(N)f(y). \tag{4}$$

Finally , the difference between $Y^N(k+1)$ and $Y^N(k)$ is assumed to be of the same order as the drift: There exists $b > 0$ such that

$$\mathbb{E}\left(\left\|Y^N(k+1) - Y^N(k)\right\|^2\right) \leq bI(N)^2. \tag{5}$$

All these assumptions are more or less necessary to use a mean field approach. The notable important feature of our model is that we do not assume any regularity property of the function f.

1.1 Differential inclusions

Here are 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 $\|x\| = \langle x, x \rangle$ (L^2 norm) and $\|A\| = \sup_{x \in A} \|x\|$.

We consider the following differential inclusion problem:

$$\dot{y}(t) \stackrel{\text{def}}{=} \frac{\mathrm{d}}{\mathrm{d}t} y(t) \in F(y(t)), \quad y(0) = y_0 \tag{6}$$

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 $y : I \rightarrow \mathbb{R}^d$ is called a solution of the DI (6) with initial condition $y(0) = y_0$ if there exists a function $f : I \rightarrow \mathbb{R}^d$ such that:

(i) for all $t \in I$: $y(t) = y_0 + \int_0^t f(s)ds$;
(ii) for almost every (a.e.) $t \in I$: $f(t) \in F(y(t))$

In particular, this implies that y is differentiable at almost every $t \in I$ with $\dot{y}(t) \in F(y(t))$.

In the following, we assume that for all $y \in \mathbb{R}^d$, $F(y)$ is a non-empty closed, convex and bounded set.

We say that F is upper semi-continuous (USC) if for any $y \in \mathbb{R}^d$ and any open set O containing $F(y)$, there exists a neighborhood V of y such that $F(V) \subset O$. We say that F is one-sided Lipschitz (OSL) with constant L if for all $y, \bar{y} \in D$ and for all $z \in F(y)$ $\bar{z} \in F(\bar{y})$:

$$\langle y - \bar{y}, z - \bar{z} \rangle \leq L \langle y - \bar{y}, y - \bar{y} \rangle. \tag{7}$$

These two conditions gives necessary and sufficient conditions for the existence and the uniqueness of solution of the DI (6). If F is USC with $\|F(x)\| \leq c(1 + \|x\|)$ then there exists solutions of (6) on all $[0; \infty)$. If F is OSL, there exists at most one solution of (6).

In our framework, if f is the limit of the drift as defined by (4), we define the set-valued function F by

$$F(y) = \bigcap_{\epsilon > 0} \overline{\text{conv}}\left(f(\{z \in \mathbb{R}^S : \|z - y\| \leq \epsilon\})\right)$$

where $\overline{\text{conv}}(A)$ is the closure of the convex hull of a set A.

If the original function f is continuous in a point y, $F(y) = \{f(y)\}$ while if f is discontinuous in y, $F(y)$ is a set-valued function. Moreover, if the original $f(y)$ is bounded on all sufficiently small open sets, F is USL.

2. CONVERGENCE RESULTS

The state of the system at time $k + 1$ can be written

$$\begin{aligned}
Y^N(k+1) &= Y^N(k) + f^N(Y^N(k)) \\
&\quad + Y^N(k+1) - Y^N(k) - f^N(Y^N(k)).
\end{aligned}$$

Let $U^N(k+1) \stackrel{\text{def}}{=} \frac{1}{I(N)}\left(Y^N(k+1) - Y^N(k) - f^N(Y^N(k))\right)$ so that Equation 4 can be rewritten

$$Y^N(k+1) = Y^N(k) + I(N)\left(f(Y^N(k)) + U^N(k+1)\right). \tag{8}$$

$U^N(k+1)$ is a martingale difference sequence with respect to the filtration $\mathcal{F}_k$ associated with the Markov chain (Y_k). (i.e. $\mathbb{E}\left(U^N_{k+1} \big| \mathcal{F}^N_k\right) = 0$).

Let us denote $\bar{Y}^N(t)$ the piecewise affine interpolation of $Y^N(k)$ when we shrink time by a factor $I(N)$: $\bar{Y}^N(t)$ is affine between $kI(N)$ and $(k+1)I(N)$ and $Y^N(kI(N)) = Y^N(k)$. $\bar{Y}^N(.)$ is a *stochastic approximation* associated with the DI (6). It can be seen as an Euler discretization of the DI (6) plus a random error term $U^N(k+1)$.

Using an approach similar to [3] from stochastic approximation, if we denote $D_T(y_0)$ the set of the solutions of the DI (6) starting from $y(0) = y_0$, we can show that $\bar{Y}(.)$ converges (in probability) to $D_T(y_0)$. More precisely, the following theorem holds.

THEOREM 1. *Assume that:*

- *F is USC and $\exists c > 0$ s.t. $\|F(y)\| \leq c(1 + \|y\|)$,*
- *$Y^N(0) \stackrel{\mathcal{P}}{\rightarrow} y_0$.*

Then $\forall T > 0$, $\displaystyle\inf_{y \in D_T(y_0)} \sup_{0 \leq t \leq T} \left\|\bar{Y}^N(t) - y(t)\right\| \stackrel{\mathcal{P}}{\rightarrow} 0$.

In general a differential inclusion may have multiple solutions. Here, $\bar{Y}^N$ may converge to any solution of the DI, depending on the value of $U^N(.)$, making all this inefficient for performance evaluation. This result is of greater interest if the DI starting from y_0 has a unique solution: $D_T(y_0) = \{y\}$. In that case, as a direct corollary of the preceding result, $\bar{Y}^N$ converges in probability to y on all intervals $[0; T]$.

COROLLARY 2. *Under the conditions of Theorem 1 and if the DI (6) has a unique solution y, then for all T:*

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

In some cases, like the example of Boinc in Section 3, the limiting differential inclusion clearly has a unique solution which makes the preceding corollary directly applicable. The main drawback of the previous theorem is that it does not give any insight on the speed of convergence on the stochastic system towards its limit.

This limitation can be overcome when the function F satisfies a one-sided Lipschitz condition (7). Firstly, this ensures the uniqueness of the solution. Secondly, one can get precise bounds on the gap between the stochastic system and its limit in that case.

THEOREM 3. *Under the conditions of Theorem 1 and if F is OSL with constant L, the DI has a unique solution y and there exist constants A_T, B_C, C_T depending only on T, L and c such that for all ϵ,*

$$\mathcal{P}\left(\sup_{0 \leq t \leq T} \left\|Y^N(t) - y(t)\right\| \geq \left\|Y^N(0) - y(0)\right\| e^{LT}\right.$$
$$\left. + \sqrt{I(N)}A_T + I(N)B_T + \epsilon\right) \leq I(N)\frac{C_T}{\epsilon^4}.$$

The constants A_T, B_T and C_T are given by

$$A_T \stackrel{\text{def}}{=} \left(c + (cT + \left\|y_0^N\right\|)e^{cT}\right)\left(\frac{LI(N)}{3} + 1\right)^{1/2}\frac{e^{LT}}{L^{1/2}};$$

$$B_T \stackrel{\text{def}}{=} \frac{e^{(L+c)T}}{\sqrt{Lc}}\left(\left\|y^N(0)\right\| + \left\|Y^N(0)\right\| + 2cT + 1\right);$$

$$C_T \stackrel{\text{def}}{=} b + be^{2LT}(B_T^2 + 2b\epsilon/L).$$

where c is defined in Theorem 1 and b in Equation (5).

These bounds are of a greater order than bounds that can be obtained in the case where f is Lipschitz (see [5]). The convergence speed with respect to N is in $O(\sqrt{I(N)})$ (compared with $O(I(N))$ is the Lipschitz case). When the unique solution $y(.)$ is piecewise Lipschitz with a finite number of pieces, this can be reduced to order $O(I(N))$ using the results of [6].

3. AN EXAMPLE: DESKTOP GRIDS

We consider a model of a volunteer computing system such as BOINC `http://boinc.berkeley.edu/`. In such systems, each volunteer offer their computing resource (called a *host*) to a distributed computing system. When a host is not used by its owner, its resources are available for the distributed computing system. However, as soon the owner of a host wants to use it, she preempts it and the host becomes unavailable for the computing system. The BOINC project is based on a push/pull model. The distributed applications *push* jobs to a central server that stores it in a buffer. Whenever a host becomes available, it *pulls* a job from the buffer and executes it.

This system fits our framework: The context $C(t)$ represents the size of the buffer while the N objects represent both the application 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 one-sided Lipschitz condition (7). Therefore, we can apply Theorem 3 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 $I(N) = 1/N$ and an application sends a job to the system with probability λ/N while a job is completed with probability μ/N. To represent the communication cost, a host gets a job with probability γ. It becomes unavailable with probability p_u/N, available with probability p_a/N 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:

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

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

3.1 Controlled dynamics

Yet an other application is for controlled systems. Consider a model similar to the previous one where, to avoid contention on the server that dispatches the jobs, a controller allocates multiple servers to this task. This system can be represented by multiple buffers, with respective sizes $C_1(t)\dots C_k(t)$ and each host is associated to a single server.

When designing such a system, a natural question is how to balance the load among servers. A natural strategy is to send incoming jobs to the server with the smaller current number of jobs. This strategy, called join the shortest queue, is highly non-smooth. If two queues almost have the same number of jobs $C_1 \approx C_2$, the job will be sent to queue 1 or 2 depending on their precise comparison.

Many practical policies used for controlling systems have such threshold dynamics. Moreover, in many cases, policies are based on heuristics on the fluid limit and lead to a unique fluid trajectory. In such cases, Corollary 2 can be used and shows that the stochastic system indeed converges to its fluid approximation.

3.2 Extension to optimal control

Yet another application of these convergence results is the optimal control of stochastic system. Again, let us consider the previous model with multiple servers where the controller wants to minimize the expected sojourn time of the jobs. The optimal control of this stochastic system can be modeled by a Markov decision process: a controller chooses an action a_k at each time step k trying to minimize the expectation of a cost that both depends on the action taken and the current state of the system. Markov decision process theory shows that if the action set is compact, then there exists an optimal policy π_*^N for the stochastic system of finite size N.

If the action is fixed ($a_k = a$ for all k), the corresponding limiting dynamics of the system is $\dot{f}(t) \in F_a(y(t))$. Using Theorem 3 and similar ideas as in [5], one can show that if all multi-valued functions F_a are OSL, then the optimal control of the limiting system π_* is asymptotically optimal for the stochastic system. Moreover, one can bound the gap between the performances of π_* and the optimal policy of the stochastic system, using similar bounds as in Theorem 3.

4. REFERENCES

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

[2] F. Baccelli, D. McDonald, and J. Reynier. A mean field model for multiple tcp connections through a buffer implementing red. *Performance Evaluation*, 49, 2002.

[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*, 65:823–838, 2008.

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

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