Mean field Limits with discontinuous drifts

Nicolas Gast and Bruno Gaujal

Grenoble University **INRIA**

Warwick – May, 2012

Outline

Scaled Markov chain

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.
the *drift* of the chain is denoted g^N:

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

$f^N(y)$ the drift rescaled by $I(N)$:

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

Stochastic approximation

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

$$Y^N(k+1) = Y^N(k) + I(N)\Big(f^N(Y^N(k)) + U^N(k+1)\Big),$$

where $U^N(k+1) \stackrel{\text{def}}{=} \left(Y^N(k+1) - Y^N(k)\right) / I(N) - f^N(Y^N(k))$ is a zero mean process that captures the random innovation of the chain between steps k and $k+1$.
$U^N(k)$ is a martingale difference sequence w. r. t. 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, $\mathbb{E}(U^N(k+1) \mid Y^N(k)) = \mathbb{E}(U^N(k+1) \mid \mathcal{F}_k) = 0$.

Stochastic approximation (II)

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:

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

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.

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

is the limit differential inclusion (DI) of $Y^N(\cdot)$.
The set of solutions up to T is denoted $\mathscr{S}_T(y_0)$.

Main theorem

Theorem

Let $Y^N(\cdot)$ be a Markov process on $\mathbb{R}^d$ as above. Assume that

- *The drift g^N vanishes with speed $I(N)$:*

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

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

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

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

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

Population Dynamics

Let D^N be a continuous time Markov chain on $\mathbb{Z}^d/N$ ($d \geq 1$) for $N \geq 1$. 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 $N\beta_\ell(y) \geq 0$, where $\beta_\ell(.)$ does not depend on N.

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 $\sum_{\ell\in\mathcal{L}} \|\ell\| \sup_y \beta_\ell(y) < \infty$. 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: $D^N(t) = Y^N(\Lambda^N(t))$.

For all $T > 0$:

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

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

Steady state distribution

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:

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

Theorem

Under the conditions of Theorem 1, if the DI 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.

Corollary

If 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^* : $\lim_{N\to\infty} \Pi^N = \delta_{y_*}$.

Convergence result (II)

Previous theorem:

Interests

- Constructive limit.
- General convergence result.
- Convergence holds for all f.

But

- no uniqueness of solution.
- no speed of convergence

F satisfies a one-sided Lipschitz (OSL) with constant L if for all $f(x) \in F(x), f(y) \in F(y)$:

$$\langle x - y, f(x) - f(y) \rangle \leq L \|x - y\|^2 .$$

When F is OSL, we can say more

Convergence result (III): OSL condition

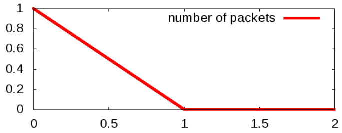
Theorem ([GG])

Assume that F is OSL with constant L, then for all T, there exists A, B s.t. for all $\epsilon > 0$,

$$\mathbb{P}\left(\sup_{0\leq t\leq T}\left\|M^N(t)-m(t)\right\| \geq \left\|M_0^N - m_0\right\| e^{LT} + \sqrt{I(N)}A + \epsilon\right) \leq \frac{I(N)}{\epsilon^2}B$$

Remarks:

- Applicability of OSL is big in dim 1:



- Similar to the Lipschitz case
  - The solution is unique.
  - Similar bounds except for $\sqrt{I(N)}$.
  - $\sqrt{I(N)}$ can become $I(N)$ if the solution is smooth enough.

Sketch of the Proof

We use a stochastic approximation method.

By definition of the drift, $M^N(k+1)$ can be written:

$$M^N(k+1) = M^N(k) + I(N)\left(\text{drift} + \underbrace{\text{noise}}_{\mathbb{E}(.)=0}\right)$$

Sketch of the Proof

We use a stochastic approximation method.

By definition of the drift, $M^N(k+1)$ can be written:

$$M^N(k+1) = M^N(k) + I(N)\Big(\text{drift} + \underbrace{\text{noise}}_{\mathbb{E}(.)=0}\Big)$$

At time $t = kI(N)$, $M^N(t)$ can be written:

$$M^N(k) = M_0^N + \sum_{i=0}^{k-1} I(N)\text{drift} \quad + \quad I(N)\sum_{i=0}^{k}\text{noise}.$$

Sketch of the Proof

We use a stochastic approximation method.

By definition of the drift, $M^N(k+1)$ can be written:

$$M^N(k+1) = M^N(k) + I(N)\left(\text{drift} + \underbrace{\text{noise}}_{\mathbb{E}(.)=0}\right)$$

At time $t = kI(N)$, $M^N(t)$ can be written:

$$\underbrace{M^N(k) = M_0^N + \sum_{i=0}^{k-1} I(N)\text{drift}}_{\text{Euler approximation of DI}} \quad + \quad \underbrace{I(N)\sum_{i=0}^{k} \text{noise}}_{\text{Converges to 0}}.$$

Sketch of the Proof

We use a stochastic approximation method.

By definition of the drift, $M^N(k+1)$ can be written:

$$M^N(k+1) = M^N(k) + I(N)\Big(\text{drift} + \underbrace{\text{noise}}_{\mathbb{E}(.)=0}\Big)$$

At time $t = kI(N)$, $M^N(t)$ can be written:

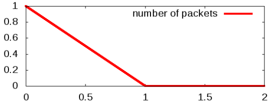
$$\underbrace{M^N(k) = M_0^N + \sum_{i=0}^{k-1} I(N)\text{drift}}_{\text{Euler approximation of DI}} \quad + \quad \underbrace{I(N)\sum_{i=0}^{k}\text{noise}}_{\text{Converges to 0}}.$$

- Convergence of the Euler approximation based on analytical arguments.
- When F is OSL, we have explicit bounds.

The One-side Lipschitz condition

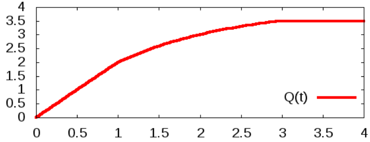
The one-side Lipschitz condition = bound the spreading of two trajectories.

- $\frac{\partial Q}{\partial t} = \begin{cases} -1 & \text{if } Q(t) > 0 \\ 0 & \text{if } Q(t) = 0 \end{cases}$



$$\langle x-y, f(x)-f(y)\rangle = \begin{cases} 0 \leq \|x-y\|^2 & \text{if } x,y>0 \\ -|x| \leq 0 \leq \|x-y\|^2 & \text{if } x>0, y=0 \end{cases}$$

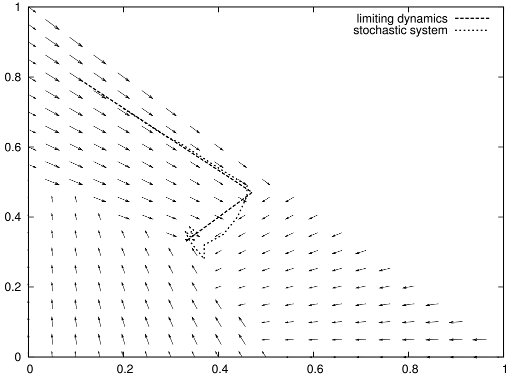
- Threshold policies: buffer $Q(t)$
  - Accept all packets if $Q < Q_{\min}$
  - Reject with prob. $p\frac{Q-Q_{\min}}{Q_{\max}-Q_{\min}}$ for $Q_{\min} \leq Q < Q_{\max}$
  - Reject all if $Q_{\max} \leq Q$.



- OSL = natural discontinuous differential equations.

Uniqueness of solution

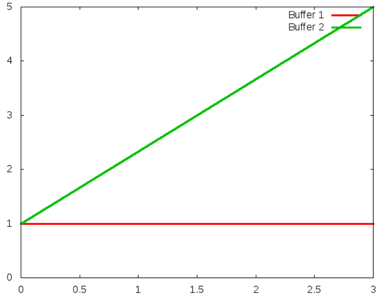
- The one-side Lipschitz condition is not necessary to show uniqueness.
- For example, in best-response dynamics (Rock-scissor-paper)

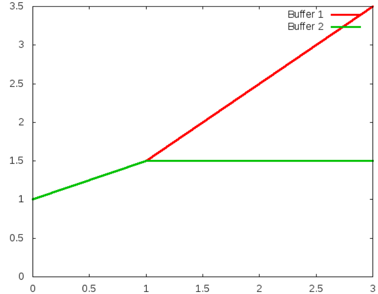


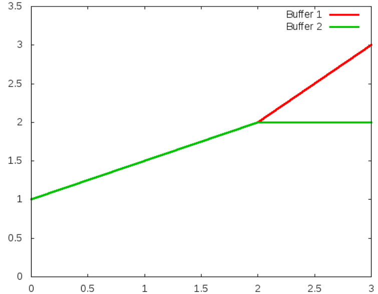
Multiple solutions: Join the longest Queue.

Two queues

- Arrival rate 1.
- No service rate.
- Packets are routed using Join Longest Queue.





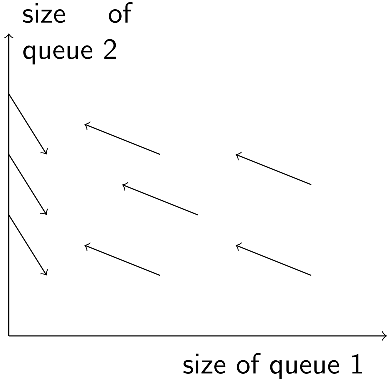


The power of Differential Inclusion

- Two queue, 1 server
- Queue 1 has full priority.

The power of Differential Inclusion

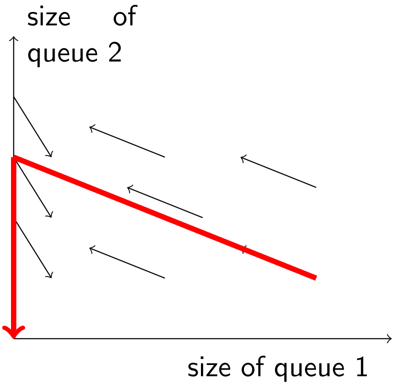
- Two queue, 1 server
- Queue 1 has full priority.



- Drift is easy to compute.
- But: differential equation has no solution.

The power of Differential Inclusion

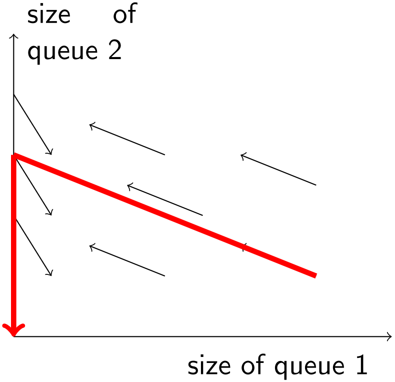
- Two queue, 1 server
- Queue 1 has full priority.



- Drift is easy to compute.
- But: differential equation has no solution.
- Convexification makes the inclusion easy to solve.

The power of Differential Inclusion

- Two queue, 1 server
- Queue 1 has full priority.



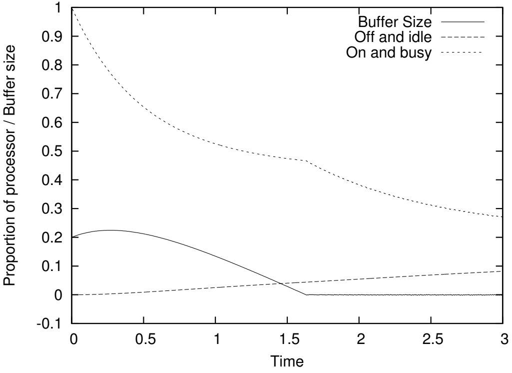
- Drift is easy to compute.
- But: differential equation has no solution.
- Convexification makes the inclusion easy to solve.

- The differential inclusion has a unique solution.

Volunteer Computing : fast simulation

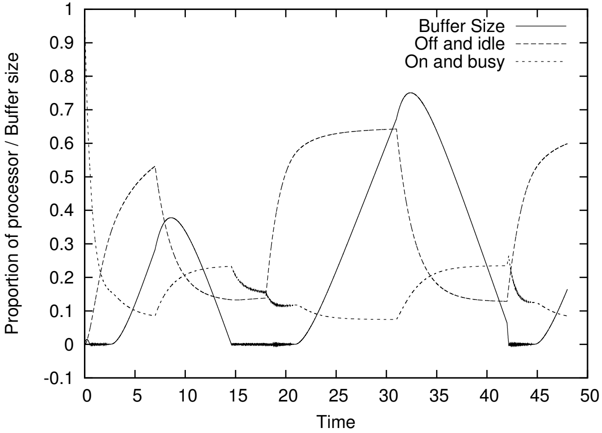
One buffer, N volunteers that are OFF or ON (busy or idle).

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



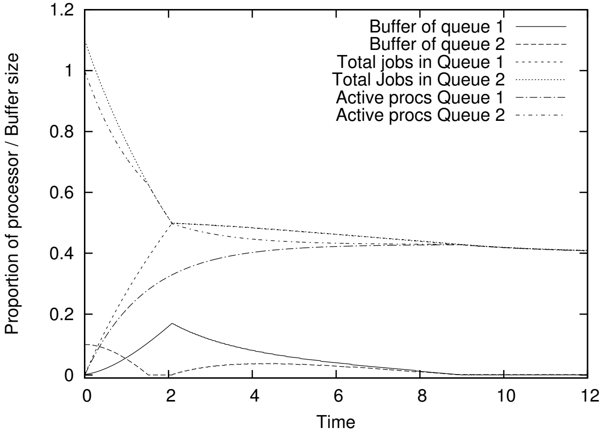
Volunteer Computing : day and night hours

- Same model with day and night behavior.
- $\frac{\partial(m,t)}{\partial t} \in F(m,t)$.



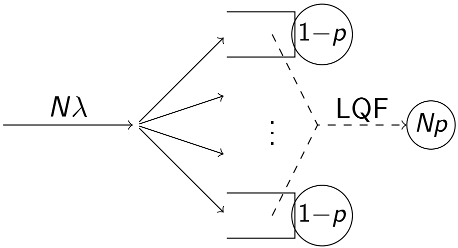
Volunteer Computing : two clusters using JSQ

- Two volunteer clusters with JSQ.



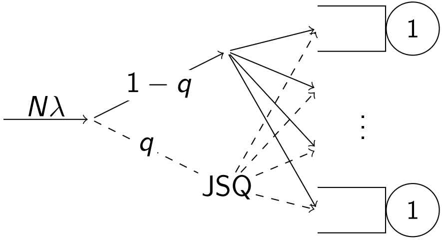
Push and pull strategies in server farms

The goal of our first example is to show how our framework can help the study of discontinuities due to centralized decisions. We consider the following model of a server farm, with a pulling server.



Push and pull strategies in server farms (II)

Now we consider the same farm with a pushing router.



Drift set

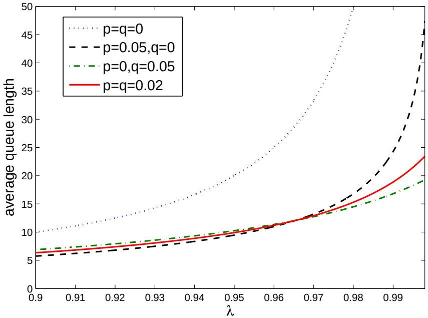
$$F_i(s) = \lambda(1-q)(s_{i-1} - s_i) - (1-p)(s_i - s_{i+1}) + u_i q - v_i p$$

with

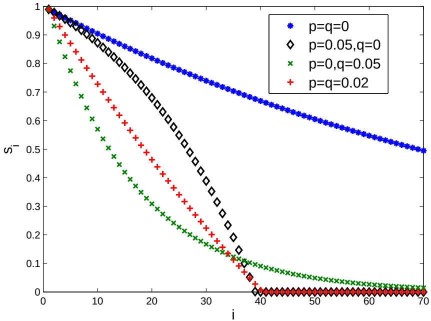
$$\left| \begin{array}{l} u_i = 0 \text{ if } s_{i-1} < 1; \\ v_i = 0 \text{ if } s_{i+1} > 0; \\ \sum_{i \geq 0} u_i = \sum_{i \geq 0} v_i = 1 \end{array} \right.$$

the differential inclusion $\dot{s} \in F(s)$ has a unique solution, that is a global attractor of all trajectories.

Performance measures



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



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

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

Fluid limits

Let $X(.)$ be a discrete time 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:

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

the fluid limits of Y^N is the set E if for all $T > 0$:

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

Proposition

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*

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

Stability and Harris recurrence

$\dot{y} \in F(y)$ is stable if there exists $T > 0$ and $\rho < 1$ such that:

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

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.

Stability and Harris recurrence (II)

Theorem

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 (1):

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

If the differential inclusion $\dot{y} \in F(y)$ is stable in the sense of Equation (6), then X is positive Harris recurrent.

Opportunistic scheduling policies in wireless networks

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 $\mathbb{E}(A_k(t)) = \lambda_k$.
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 $q_{k,i} \neq 0$. 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 $\mu_{k,i}$.
Let us consider the following policy (called "Best Rate" policy in

- if there are n users $x_1 \dots x_n$ 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 x_1).
- if there are no users in their best channel condition, serve a user at random.

Opportunistic scheduling policies in wireless networks (II)

Let us compute the set-value function F at point $y = (0, \dots 0, y_\ell, \dots y_K)$, with $y_\ell > 0$. 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.
The value of the drift at Ny^N is equal to

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

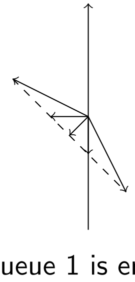
F is the following set-valued function:

$$F(y) = \begin{cases} \vec{u_1} & \text{if } x_1 > 0 \\ \text{conv}(\vec{u_1}, \dots, \vec{u_k}) & \text{if } y_1 = \cdots = y_{k-1} = 0, y_k > 0. \end{cases}$$

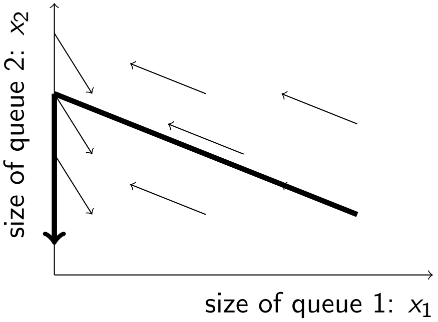
However, notice that when y_i^N goes to zero as N goes to infinity, the sequence p_i^N does not necessarily converge as N goes to infinity. This implies that the rescaled drift f^N does not converge to any single-valued function (continuous or not) in that case.

Opportunistic scheduling policies in wireless networks (III)

The condition $\sum_k \lambda_k/\mu_k < 1$ implies that the unique solution of the DI goes to 0 in finite time.



(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.