Stochastic approximation, differential inclusion and stability.

Nicolas Gast [1]

EPFL

Algo GT 2011 – June 20-21

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

1. Joint work with Bruno Gaujal (INRIA)

Outline

Outline

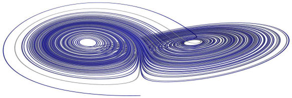
What is stochastic stability ?

$X(t)$ is a stochastic system $X(t)$ (*e.g.* number of customers waiting.)

Stability $\approx$ $X(t)$ *does not go to infinity* as t goes to infinity.

Stability is a weak property

- needed to have a stationary regime.
- no bound on waiting time.

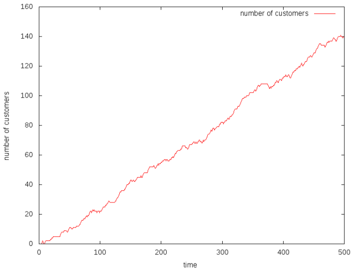


This is stable.

Stochastic stability on an example

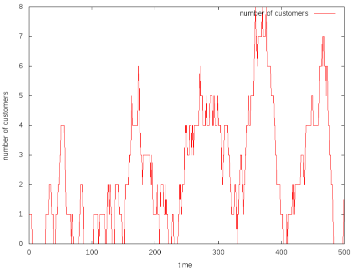
M/M/1 queue :

- Arrival rate of customers λ.
- Departure rate μ.



$\lambda/\mu = 2$

Unstable



$\lambda/\mu = .8$

stable

Mathematical definition

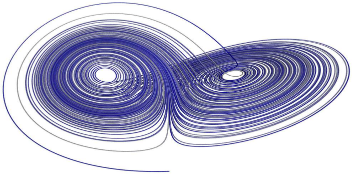
Definition (stability)

Let $X(t)$ be a Markov chain on a discrete space $\mathcal{S}$.

- $X(t)$ is stable if $X(t)$ is positive recurrent (has a stationary distribution).

Example :

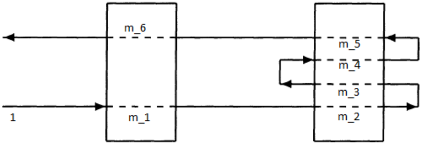
- The M/M/1 queue is stable *iff* $\lambda/\mu < 1$.
- A stable system can oscillate.



This is stable.

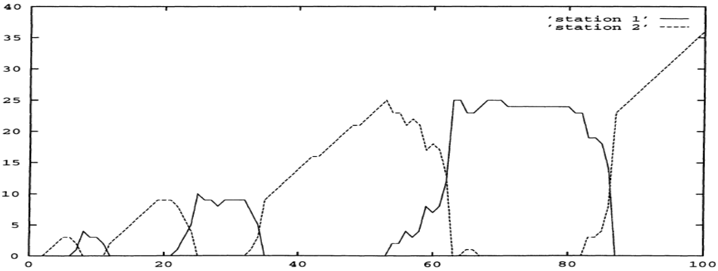
Stability : a difficult problem

Example : reentrant line with FIFO disciplines. Rybko,Stolyar (1992), Daï (1995).



m_1	m_2	m_3	m_4	m_5	m_6
.001	.897	.001	.001	.001	.899

The load on each server is $\rho_1 = \rho_2 = .90$ but :



Outline

Foster's Criteria

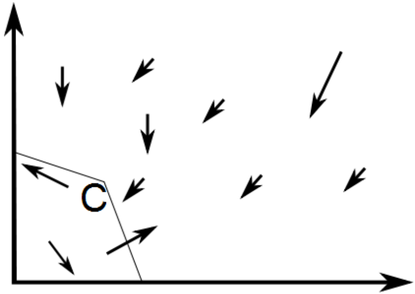
Idea : average decrease of the norm for large state implies stability.

Foster's Criterion (taken from Bramson [2008])

Let $X(n)$, be a Markov chain on which all states communicate. If there is a finite space C s.t.

- $\mathbb{E}\left[\|X(n+1)\| \mid X(n) = x\right] < \infty$ for all $x \in C$
- $\mathbb{E}\left[\|X(n+1)\| \mid X(n) = x\right] \leq \|x\| - \epsilon$ for $x \notin C$.

Then, X is stable.



Foster's Criteria : example and limits

Example : M/M/1 queue (arrival w. proba. λ, departure w. proba. μ.)

- $\mathbb{E}\left[|X(n+1)|X(n)=0\right] = \lambda < \infty$.
- $\mathbb{E}\left[|X(n+1)|X(n)=x\right] = x + \lambda - \mu$ if $x > 0$.

Thus : λ/μ implies stability.
(in that case, there exists simpler methods).

Powerful tool but :

- Hard to find a Lyapounov
- Little intuition on speed of convergence.

Fluid limit : definition and example

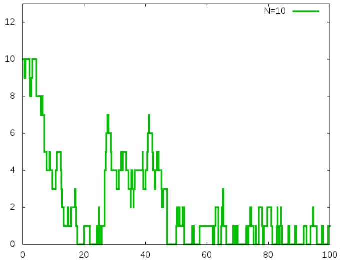
Consider $X^N(t)$:

- Scale the initial state by N.
- Scale the time by N : $X^N(t) = X(N \cdot t)$.

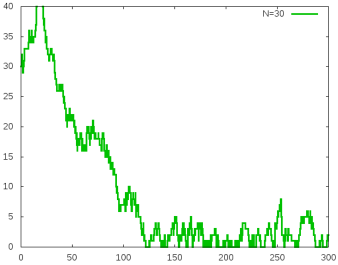
A fluid limit is a limit of $X^N(t)$ as N goes to infinity.

Example : $M/M/1$ queue.

- $X(t) \in \mathbb{N}$ = number of packets in queue.



$N = 10$



$N = 30$

Fluid limit and stability of queing networks.

Bransom (2008) (generalization of Daï (1995)) showed that

Fluid stability implies stochastic stability for HL.

For head-of-the-line queuing networks :

- Fluid limits are described by differential equations.
- If $\exists c$ such that all solutions of the ODE satisfies :

$$X(t) = 0 \quad \text{for } t > c\,\|X(0)\|\,,$$

  then *the system is stable.*

The converse might false (see Bramson (1999))
However :

-
- restricted to queuing models.
  - Models not valid, *e.g.* wireless networks (interferences).

Outline

Drift and stochastic approximation

Recall : fluid limit

- behavior when space and time is scaled by N : $X^N(t+\frac{1}{N}) = \frac{1}{N}X(N\cdot t)$

The drift of a discrete time Markov chain $X(n)$ is :

$$g(x) = \mathbb{E}\left[X(t+1) - X(t) \mid X(t) = x\right].$$

We have $X(t+1) = X(t) \quad + \quad \underbrace{g(x)}_{\text{drift}} \quad + \quad \underbrace{\text{noise}}_{\mathbb{E}[\cdot]=0}.$

$$X^N(t+\frac{1}{N}) = X^N(t) + \frac{1}{N}\left(g(N\cdot X^N(t)) + \text{noise}\right). \quad (1)$$

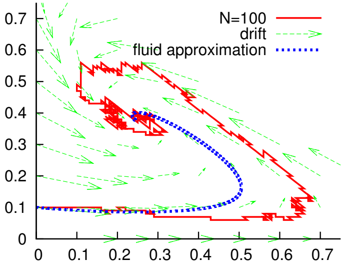
(1) is a stochastic approximation algorithm.

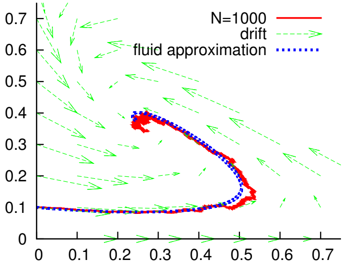
Convergence of stochastic approximation

$$X^N(t + \frac{1}{N}) = X^N(t) + \frac{1}{N}\left(g(N \cdot X^N(t)) + \text{noise}\right). \quad (1)$$

When $\lim_{N\to\infty} g(N \cdot x) = f(x)$ is lipschitz continuous :

- X^N satisfies the ODE $\dot{x} = f(x)$. (started with Robbins and Monro (1951), see also Benaïm (1999).)





Discontinuous drifts

However, most system have discontinuous drift. Example : M/M/1 queue. The drift is :

$$f(x) = \begin{cases} \lambda - \mu & \text{if } x > 0. \\ \lambda & \text{if } x = 0. \end{cases}$$

Problem : $\dot{x} = f(x)$ has no solutions on $[0; \infty)$ starting from $x = 1$.

Solution : transform the ODE into a differential inclusion (DI) :

$$\dot{x} \in F(x) \quad \text{with } F(x) = \begin{cases} \lambda - \mu & \text{if } x > 0. \\ [\lambda - \mu; \lambda] & \text{if } x = 0. \end{cases}$$

The DI has a unique solution.

Construction of set valued drift

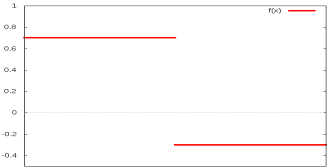
If f is the drift, then we define F :

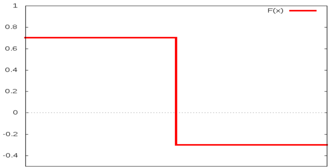
$$F(x) = \textbf{convex\_hull}\left(\overline{\lim}\ g(N \cdot x^N) \mid x^N \to x\right).$$

Example : $M/M/1$.

$$f(x) = \begin{cases} -.3 & \text{if } x > 0. \\ +.7 & \text{if } x \leq 0. \end{cases}$$

$$F(x) = \begin{cases} -.3 & \text{if } x > 0. \\ [-.3 : .7] & \text{if } x = 0 \\ +.7 & \text{if } x \leq 0. \end{cases}$$





Convergence

Theorem (G, Gaujal 10)

Assume that f is bounded, that the variance of the noise is bounded and that F is the set-valued drift. If $X^N(0) \to x$, then for all T :

$$\inf_{x \in S(x_0)} \sup_{0 \le t \le T} \left\| X^N(t) - x(t) \right\| \xrightarrow{\mathcal{P}} 0.$$

where $S(x_0)$ = set of solutions of $\dot{x} \in F(x)$ with $x(0) = x_0$.

- Generalization of decreasing step-size studied by Benaïm et al. (06).
- Similar results obtained by Faure and Roth (10).

In words : for any drift f, fluid limits are characterized by the DI $\dot{x} \in F(x)$.

Theorem (G, Gaujal 10)

If the DI associated with f is stable, then the stochastic system is stable.

Outline

Opportunistic routing in wireless networks

- An antenna serves mobiles.
- Channel condition is varying with time (fading).

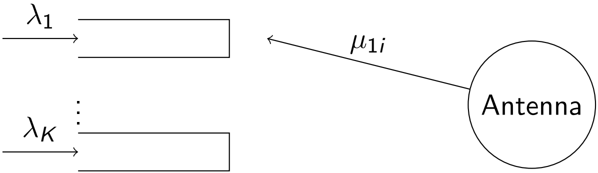
At each time step, the antenna chooses which user to serve :

- knowing the channel condition of everyone.

Goal : find a “good” policy for serving the users.

Opportunistic routing, flow-level model from (Ayesta et al. 10)

- Mobiles can be divided in K classes.
  - Users of class k arrive at rate λ_k
- At each time step :
  - a mobile of class k has a channel condition (*i.i.d*).
  - The antenna chooses one user j. It is served with rate $\mu(\text{class}, \text{channel\_condition})$



- $X(t) = (X_1(t) \dots X_K(t)) = \#$ users of each class in the system.

Questions :
- What is the stability condition of a policy ?
- Is there a maximum stable policy ?

Stability condition

Theorem (Ayesta et al.)

There exists a policy that stabilize the system iff $\sum_{k=1}^{K} \frac{\lambda_k}{\mu_k^{\max}} < 1$

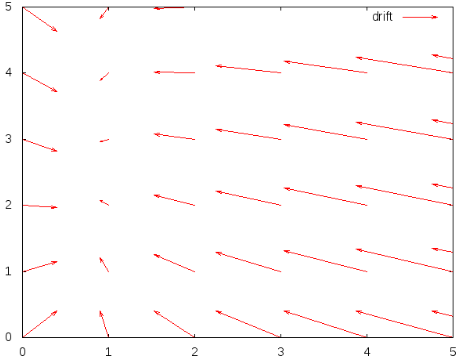
Consider the policy :

- Serve in priority users that are in their best state.
- Serve in priority class 1 users, then class 2, etc.
- Any tie breaking rule.

Drift of the stochastic system

When x is small, the drift is complex.

Example with 2 classes of users and 2 channel conditions.



- When x is large, there is a user in its best state with proba. ≈ 1.

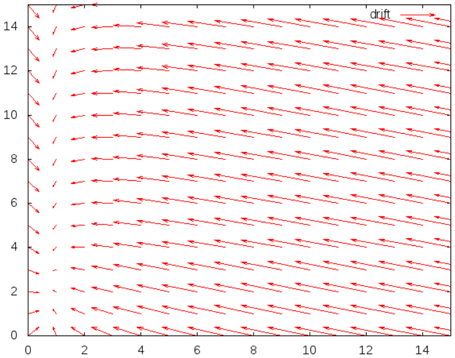
If a user of class k in its best state is served, the drift is :

$$f(x) = (\lambda_1, \ldots, \lambda_{k-1}, \lambda_k - \mu_k^{\max}, \lambda_{k+1}, \ldots, \lambda_K) = \mathbf{v}_k.$$

Drift of the stochastic system

When x is small, the drift is complex.

Example with 2 classes of users and 2 channel conditions.



- When x is large, there is a user in its best state with proba. ≈ 1.

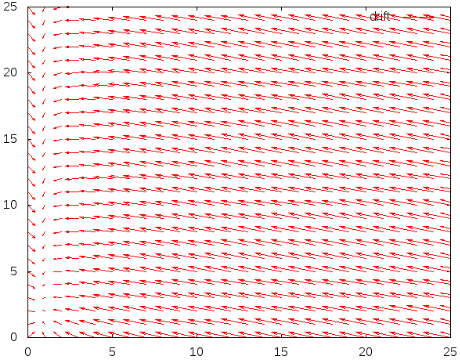
If a user of class k in its best state is served, the drift is :

$$f(x) = (\lambda_1, \ldots, \lambda_{k-1}, \lambda_k - \mu_k^{\max}, \lambda_{k+1}, \ldots, \lambda_K) = \mathbf{v}_k.$$

Drift of the stochastic system

When x is small, the drift is complex.

Example with 2 classes of users and 2 channel conditions.



- When x is large, there is a user in its best state with proba. ≈ 1.

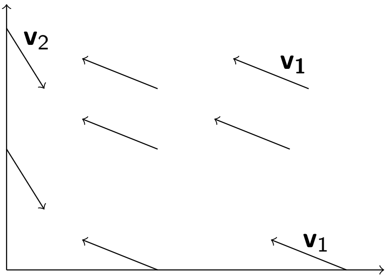
If a user of class k in its best state is served, the drift is :

$$f(x) = (\lambda_1, \ldots, \lambda_{k-1}, \lambda_k - \mu_k^{\max}, \lambda_{k+1}, \ldots, \lambda_K) = \mathbf{v}_k.$$

Stability of the fluid limit

For, x "large", the drift is :

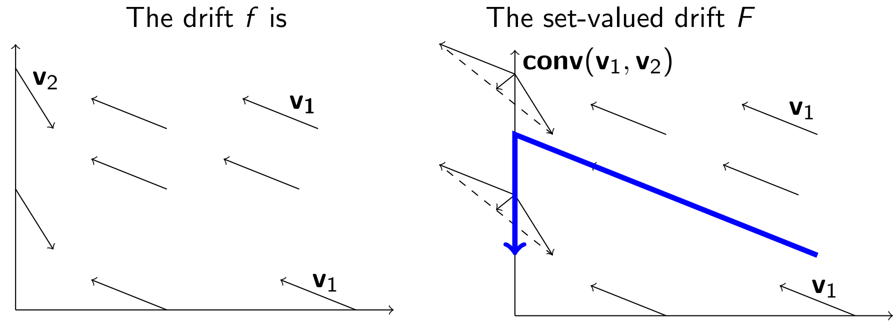
The drift f is



- The ODE corresponding to f has no solution.

Stability of the fluid limit

For, x "large", the drift is :



- The ODE corresponding to f has no solution.
- The DI can be solved in closed form.

Therefore, if $\sum_{k=1}^{K} \frac{\lambda_k}{\mu_k^{\max}} < 1$, then :

- The solution of the DI converges to 0.
- Thus, the opportunistic scheduling is stable.

Limit of the approach

The set of solution of the DI is a super-set of the fluid limits.
Example : three weakly coupled queues (Borst et al. 2008).

- Arrival rate at queue i is λ_i
- 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 $\phi_i(x)$

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

where $a_i \geq a_{ij} \geq 1$.
Stability conditions given by the convex hull (up to a permutation of $1,2,3$) :

- $\lambda_1 < 1$ and $\lambda_2 < \lambda_1 + a_{23}(1-\lambda_1)$ (also necessary conditions).
- $+$ a linear equation on λ_3 stronger than necessary.

The DI has multiple solutions.

Outline

Conclusion

Stochastic approximation to construct fluid limits.

- Differential inclusion to handle discontinuity.
- Constructive definition of the limit.
- Facilitate the computation of closed-form behavior.

Other applications :

- Study quantitative properties (mean field approximation).
  - Need for speed of convergence (OSL, R-OSL).
- Hitting time and stochastic approximation.
  - Extinction time.

Some references

Paper corresponding to this talk :

- [Gast,Gaujal 10] – *Markov chains with discontinuous drifts have differential inclusions limits. Application to stochastic stability and mean field approximation.*

Other references cited :

- [Benaïm et al. 99] – *Dynamics of stochastic approximation algorithms*
- [Benaïm et al. 06] – *Stochastic approximations and differential inclusions*
- [Bramson 99] – *A stable queueing network with unstable fluid model*
- [Bramson 08] – *Stability of queuing networks*
- [Borst et al. 08] – *Stability of parallel queueing systems with coupled service rates*
- [Daï 95] – *On positive Harris recurrence of multiclass queueing networks : a unified approach via fluid limit models*
- [Rybko,Stolyar 92] – *Ergodicity of stochastic processes describing the operation of open queueing networks*
- [Ayesta et al. 10] – *Scheduling in a random environment : stability and asymptotic optimality*