The Baccelli-Foss Saturation Rule for stability for continuous time input processes

Eitan Altman
INRIA Sophia Antipolis, B.P. 93
06902 Sophia Antipolis, France

Abstract

We analyze the stability of systems that are driven by input processes that are right continuous and have left limits. These include stochastic fluid models frequently used in telecommunications for modeling variable rate inputs, as well as jump processes. Our approach is an adaptation of the Saturation Rule developed recently by Baccelli and Foss. The analysis is carried under the assumption that the input process is stationary ergodic, or that it has some stationary periodic structure.

Keywords: Saturation Rule, stationarity, Ergodicity.

1 Introduction

A very powerful and elegant new approach for studying the stability of non-Markovian stochastic systems has been recently developed by Baccelli and Foss in [4]. It was very successfully applied to generalized Jackson networks [3], stochastic Petri nets [5], and polling systems [7]. This method considers systems as operating on some exogenous marked point processes. For example, in queueing applications, the times at which jumps occur may correspond to arrival of customers, and the marks could correspond to service times and routing information. One then focuses on the daters, which are random variables that indicate the time at which some local or global activity has terminated, given that the input process is restricted to some finite time. The methodology is based on spearable and monotone axioms that are assumed to be satisfied by the daters. They are shown to imply sub-additivity properties. Under some fairly general stochastic assumptions on the input processes, Baccelli and Foss obtained a first order ergodic theorem, by applying Kingman's sub-additive ergodic theorem, and some refined second-order ergodic theorems.

In this note we extend the above framework to general exogenous input processes that are right continuous and have left limits. This generalization is partly motivated by the growing use of stochastic fluid models in telecommunication and production systems [8, 9]. Our results are based on adaptation of the approach in [4] (see also [2]), and by introducing accelerating and decelerating time change functions, that allows to compare daters that correspond to different processes. An additional monotone axiom needs to be satisfied in our framework, which we call the "dual monotonicity", in order to apply an extention of Kingman's ergodic theorem for continuous time ([6] p. 177). We further obtain stability results for input processes that have some periodic behavior. An example of a G/G/1 type queue is given finally.

2 The Model

We begin by presenting some axioms that refer to the behavior of the system for given fixed input processes $N : \mathbb{R} \to G$, where G is some complete separable metric space, endowed with the standard Borel field $\mathcal{G}$. N is assumed to be right continuous with left limits. Let D be the set of paths which are right continuous with left limits, endowed with the standard Borel σ-Algebra $\mathcal{D}$. We assume that there exists some element $0 \in G$ that corresponds to no input, or, to an "inactive" input process (for example, no arrivals of customers). Define

- $\theta : \mathbb{R} \times D \to D$ the operation of shifting by a constant, given by $\theta_s N(t) = N(t+s)$. We shall use (with some abuse of notation) θ also as a shift over sets, i.e. $\theta : \mathbb{R} \times \mathcal{D} \to \mathcal{D}$, where $\theta_s A := \{\theta_s N, N \in A\}$ for any $A \in \mathcal{D}$.
- The operation of scaling N by a constant c is written as cN, i.e. $cN(t) := N(t/c)$.
- $N_{r,s}(t) := N \cdot 1\{t \in (r, s]\}$ is the restriction of N to the interval $(r, s]$.

- $\mathcal{T} : \mathbb{R} \to \mathbb{R}$:= the set of monotone functions which are right continuous. Functions in $\mathcal{T}$ will be identified as "time-changes".
- $\Theta : \mathcal{T} \times D \to D$: the operation of time change; given $\pi \in \mathcal{T}$, $N \in D$, we define $\Theta(\pi, N)(t) = N(\pi(t))$.

We say that a time change π is an acceleration if $\pi(t) \geq t$ for all t, and it is deceleration if $\pi(t) \leq t$ for all t. We say that N is slower than N' (or N' is faster than N), and denote it by $N \prec N'$, if there exists some acceleration π such that $N' = \Theta(\pi, N)$. Finally, we note that

$$\begin{gathered} c(N_{s,t}) = (cN)_{cs,ct} \text{ and } \theta_r(N_{s,t}) = (\theta_r N)_{s-r,t-r}, \\ \text{if } N' = \Theta(\pi, N) \text{ then } N'_{r,s}(t) = N_{\pi(r),\pi(s)}(\pi(t)) \text{ and } \Theta(\pi, N_{r,s}) = N'_{\pi^{-1}(r),\pi^{-1}(s)} \end{gathered} \tag{1}$$

whenever π^{-1} is well defined.

Following [3], we define the maximal dater $X(N)$ to be the time of the last activity in the network, when it is fed by N (given that the system is empty at $t = -\infty$). X thus describes some property of the system to which process N is fed. Define for any $s \leq t$ $Z_{s,t}(N) := X(N_{s,t}) - t$. We assume that X has the following properties. For all s, t, N, with $s \leq t$:

1. Causality: $X(N_{s,t}) \geq t$.
2. External monotonicity: $X(N') \geq X(N_{s,t})$ whenever $N' \prec N_{s,t}$.
3. Homogeneity: For any $r \in \mathbb{R}$, $X(\theta_r N) = X(N) + r$.
4. Separability: For any $r \in \mathbb{R}$ such that $s \leq r < t$, if $X(N_{s,r}) \leq r$, then $X(N_{s,t}) = X(N_{r,t})$.

Lemma 2.1 *Under conditions 1-4, X and $Z_{s,t}$ satisfy the following:*

5. Internal monotonicity property: *For all r, s, t with $s \leq r < t$ and all N,*

$$X(N_{s,t}) \geq X(N_{r,t}), \qquad Z_{s,t}(N) \geq Z_{r,t}(N). \tag{2}$$

Proof: The proof is given in [1]. It is based on choosing arbitrary r, s, t with $s \leq r < t$ and N, and then defining the process $N' = \Theta(\pi^1, N_{s,t})$, where the accelerating time change $\pi^1 \in \mathcal{T}$ is given by

$$\pi^1(t) = \begin{cases} t + Z_{s,r}(N) & \text{if } t < r - Z_{s,r}(N) \\ r & \text{if } r - Z_{s,r}(N) \leq t < r \\ t & \text{if } t \geq r \end{cases} \tag{3}$$

The remaining of the proof follows from ideas similar to those in [4]. ∎

Lemma 2.2 *Under conditions 1-4, $Z_{s,t}$ satisfies the following*
6. Sub-additive property: *For all r, s, t with $s \leq r < t$ and all N,*

$$Z_{s,t}(N) \leq Z_{s,r}(N) + Z_{r,t}(N). \tag{4}$$

Proof: The proof is given in [1]. It is based on introducing the following decelerating time change π^2:

$$\pi^2(t) = \begin{cases} t & \text{if } t < r \\ r & \text{if } r \leq t < r + Z_{s,r}(N) \\ t - Z_{s,r}(N) & \text{if } r + Z_{s,r}(N) \leq t \end{cases} \tag{5}$$

and considering the processes $N^1 = \Theta(\pi^1, N)$ and $N^2 = \Theta(\pi^2, N)$. The remaining of the proof follows from ideas similar to those in [4]. ∎

3 The stochastic framework

We assume in this section that N is defined on the canonical probability space $(D, \mathcal{D}, P)$, and is stationary ergodic under the shift θ_t. We further assume that it satisfies conditions 1-4 on all trajectories (with probability 1). We need the following additional assumption.

$$E\left[\sup_{0<s\leq 1} Z_{0,s}(N)\right] < \infty. \tag{6}$$

By the internal monotonicity this implies in particular

$$E\left[\sup_{0\leq r<s\leq 1} Z_{r,s}(N)\right] < \infty. \tag{7}$$

Since sup is a convex function, (6) also implies (by Jensen's inequality) that

$$\sup_{0<s\leq 1} EZ_{0,s}(N) < \infty. \tag{8}$$

(8) implies that for any r and s such that $r \leq s$, $EZ_{r,s}(N) < \infty$. Indeed, by expressing s as $s = K\epsilon$ for some $0 < \epsilon \leq 1$ with K being some integer, we have

$$EZ_{r,s}(N) \leq_5 EZ_{0,s}(N) \leq_6 KEZ_{0,\epsilon}(N) < \infty. \tag{9}$$

(7), (9) and the nonnegativity of $Z_{0,t}(N)$ allow us to use a generalization of Kingman's subadditive ergodic theorem ([6] p. 177), which yields

Lemma 3.1 *Assume Conditions 1-4 and (6). Then there exists some finite non-negative constant γ such that*

$$\lim_{t\to\infty} \frac{Z_{-t,0}(N)}{t} = \lim_{t\to\infty} \frac{Z_{0,t}(N)}{t} = \lim_{t\to\infty} \frac{EZ_{0,t}(N)}{t} = \lim_{t\to\infty} \frac{EZ_{-t,0}(N)}{t} = \gamma \tag{10}$$

P-a.s.

Note that Lemma 3.1 implies that $\lim_{t\to\infty} X(N_{0,t})/t = \gamma + 1$. The proof of the following Lemma is the same as in [2] Chapter 9 and [4].

Lemma 3.2 *Assume Conditions 1-4 and (6). Then*
(i) Define $A = \{lim_{t\to\infty} Z_{-t,0}(N) = \infty\}$. Then $P(A)$ is either one or zero.
(ii) $Z_{-ct,0}(cN)$ is decreasing in c and $Z_{0,ct}(cN)$ is increasing in c; for all $c > 0$, there exists a non-negative constant $\gamma(c)$ such that the following hold for the scaled processes:

$$\lim_{t\to\infty} \frac{Z_{-ct,0}(cN)}{t} = \lim_{t\to\infty} \frac{Z_{0,ct}(cN)}{t} = \lim_{t\to\infty} \frac{EZ_{-ct,0}(cN)}{t} = \lim_{t\to\infty} \frac{EZ_{0,ct}(cN)}{t} = \gamma(c), \qquad a.s.; \tag{11}$$

$\gamma(c)$ is decreasing in c and $\gamma(c) + c$ is increasing in c. In particular, $\gamma + 1 \geq \gamma(0)$ where $\gamma = \gamma(1)$.

Remark 3.1 *The process cN is not well defined if $c = 0$ (it is not in $\mathcal{D}$ anymore) unless some extra assumptions are made. Therefore, unlike the framework studied in [4], we do not allow $c = 0$ in the above Lemma. However, since $\gamma(c)$ is decreasing in c, the limit $\gamma(0) := \lim_{c\to 0} \gamma(c)$ is well defined. This problem requires special attention in the proof of the second part of the following main result.*

Theorem 3.1 *Assume conditions 1-4 and (6).*
(i) If $\gamma(0) > 1$ *then* $Z_0(N) := \lim_{t\to\infty} Z_{-t,0}(N) = \infty$ *a.s.*
(ii) If $\gamma(0) < 1$ *then* $Z_0(N) < \infty$ *a.s.*

Proof: The proof is given in [1]. The proof of (ii) requires special attention due to Remark 3.1. ■

We thus obtain a stationary ergodic process $Z_t(N) := \theta_t Z_0(N)$ under the assumptions that $\gamma(0) < 1$.

Finally, we present a simple sufficient condition for (6):
7. Dual monotonicity: For any $r \le s \le t$, $X_{r,s}(N) \le X_{r,t}(N)$.

Lemma 3.3 *[1] Assume conditions 1-4 and 7, and that* $EZ_{0,1}(N) < \infty$*. Then (6) holds.*

4 The stochastic periodic framework

Assume that there is some period T, such that the input process is stationary ergodic under the shift $\theta'_n := \theta_{nT}$, $n = 0, 1, 2, ...$ (instead of the shift θ_t, $t \in \mathbb{R}$). This includes in particular the case of periodic deterministic input process, with period T. Assume further that the input process satisfies conditions 1-4, as well as the integrability assumption (6). It follows that Lemma 3.1, 3.2 and Theorem 3.1 still hold. The process $Z_t(N) := \theta_t Z_0(N)$ thus obtained, is stationary ergodic under the shift θ'_n, if $\gamma(0) < 1$. Again, The dual monotonicity can be used as a sufficient condition to establish (6).

5 Example: a G/G/1 type queue

We consider a queue where the server operates at a unit speed. The arrival process N is right continuous with left limits; it is stationary ergodic (with respect to the shift θ_t, $t \in \mathbb{R}$, or with respect to the shift $\theta'_n = \theta_{nT}$, $n \in \mathbb{N}$, where T is the period). $N(t)$ represents the total accumulated work that arrived to the system till time t. The standard G/G/1 queue is obtained as special case when work arrives at discrete times. N then becomes a jump process which can be described as a marked point processes, where the times of jumps represent the arrival epochs, and the jump sizes (or the marks) represent the service time required by a customer.

Let W_t denote the amount of work in the system. It evolves according to:

$$W_t = \max\left\{\int_s^t (dN(t) - dt), t_0 < s \le t, W_{t_0} + \int_{t_0}^t (dN(t) - dt), 0\right\}. \tag{12}$$

If we identify an inactive system with an empty system and take $X(N)$ to denote the time till the system becomes inactive, then it turns out that $X(N)$ does not satisfy the causality condition, since the system may be empty even if positive amount of work arrives to the system at any time, provided that the input rate is smaller than 1. However, if we identify an inactive system as an empty system with idle server then conditions 1-4 and 7 hold. We note that $Z_t(N)$ is simply the work present at the system at time t, so that our results provide the stability conditions (existence of a stationary regime) for the work in the system.

Finally, it follows from (12) that $\lim_{c\to 0} Z_{-ct,0}(cN) = \int_{-t}^0 dN(s)$, so that $\gamma(0)$ is the average rate at which work arrives to the system per unit of time. More precisely, in the stochastic framework $\gamma(0)$ is the a.s. limit $\gamma(0) = \lim_{t\to\infty} t^{-1}\int_{-t}^0 dN(s)$. For the special case that N is deterministic with period T, it is given by

$$\gamma(0) = \lim_{t\to\infty} t^{-1}\int_{-t}^0 dN(s) = T^{-1}\int_{-T}^0 dN(s). \tag{13}$$

References

[1] E. Altman, “On the Baccelli-Foss Saturation Rule for stochastic stability”, manuscript.

[2] F. Baccelli and P. Brémaud, *Elements of Queueing theory*, Springer-Verlag, 1994.

[3] F. Baccelli, and S. Foss, “Stability of Jackson-type Queueing Networks, I”, INRIA Report n. 1845, To appear in QUESTA.

[4] F. Baccelli and S. Foss, “On the Saturation Rule for the Stability of Queues”, INRIA Report n. 2015, 1993.

[5] F. Baccelli and B. Gaujal, "Stationary Regime and Stability of Free-Choice Petri Nets", *Proc. 11th Int. Conf. on Analysis and Optimization of systems.*

[6] R. Carmona and J. Lacroix, *Spectral Theory of Random Scrœdinger Operators*, Birkhäuser, 1990.

[7] N. Chernova and S. Foss, "On the boundedness of queue-length process in polling systems", In preparation.

[8] D. Mitra, "Stochastic theory of a fluid model of producers and consumers coupled by a buffer", *Adv. of Appl. Prob.* **20**, pp. 646-676, 1988.

[9] T. E. Stern and A. I. Elwalid, "Analysis of separable Markov-modulated rate models for information-handling systems", *Adv. of Appl. Prob.* **23**, pp. 105-139, 1991.