6th St.Petersburg Workshop on Simulation (2009) 1-6

A sufficient condition for stability of fluid limit models[1]

Rosario Delgado[2]

Abstract

We prove that in the subcritical case, a fluid limit model is stable if *state space collapse* with a "lifting" matrix that verifies a restriction holds.

1 Introduction

We consider a fluid model which consists of J stations, with a single server and an infinite-capacity buffer at each one, that process K different fluid classes, with $K \geq J \geq 1$. Each fluid class can be processed at only one station and feedback is allowed. We assume a work-conserving service discipline. This fluid model can be considered as the fluid approximation of an associated queueing network that works under any head-of-the-line work-conserving service discipline and with inter-arrival and service times not necessarily exponential. It is known that the stability of the queueing network (the *positive Harris recurrence* of the underlying Markov process describing the network dynamics) is ensured if the fluid model is stable (see Theorem 4.2 [2]). *Stability* of a fluid limit model means that the queue process reaches zero in finite time and stays there regardless of the initial fluid levels. It is known that sub-criticality (traffic intensity strictly less than one at each station) is not a sufficient condition, although necessary, for stability.

In this work we establish a sufficient condition for the stability of the fluid limit model (in the subcritical case): it is a kind of *state space collapse* assumption with a "lifting" matrix that verifies a technical restriction. *State space collapse* condition has turned out to be a key ingredient in the proof of *heavy-traffic* limits for multi-class queueing networks in the light-tailed as well as in the heavy-tailed environment. See for instance [5], [4], [1], [7] and [3]. As far as we know, this is the first time that this kind of condition has been related with the study of stability (*light-traffic*).

[1]This work was supported by grant MEC-FEDER ref. MTM2006-06427 and 2005SGR-01043

[2]Universitat Autònoma de Barcelona (Spain), E-mail: delgado@mat.uab.cat

2 The fluid limit model

Consider a fluid model consisting of $J \geq 1$ single-server stations with a single server and an infinite-capacity buffer at each one. There are K fluid classes with $K \geq J$, each one processed at only one station (but at each station more than one fluid class can be processed), $s(k)$ being the station where class k fluid is processed, and $s^{-1}(j)$ the set of fluid classes served at station j. We introduce the $J \times K$ (deterministic) *constituency matrix* $C = \left(C_{jk}\right)_{j,k}$ by $C_{jk} = 1$ if $j = s(k)$ and 0 otherwise.

Let $\alpha_k \geq 0$ be the exogenous inflow rate and $\mu_k > 0$ be the potential outflow rate, for class k fluid, and define $m_k \stackrel{\text{def}}{=} \frac{1}{\mu_k}$ and matrix $M \stackrel{\text{def}}{=} diag(m_1, \ldots, m_k)$. Upon being processed at station $s(k)$, a proportion $P_{k\ell}$ of class k fluid leaving station $s(k)$ goes next to station $s(\ell)$ to be processed there as a class ℓ fluid. The *"flow-transfer" matrix* $P = (P_{k\ell})_{k,\ell=1}^{K}$ is assumed to be sub-stochastic and to have spectral radius strictly less than one. Hence, $Q \stackrel{\text{def}}{=} \left(I - P^T\right)^{-1}$ is well defined. We assume that fluid at each station is processed following a *work-conserving* service discipline by arrival order into each class.

The fluid model is described by elements $\alpha = (\alpha_1, \ldots, \alpha_k)^T$, M, C, P and $z = (z_1, \ldots, z_K)^T \geq 0$, z_k being the initial amount of class k fluid in the system. We refer to it by (α, M, C, P, z).

We define λ to be the unique $K-$dimensional vector solution to the *traffic equation* $\lambda = \alpha + P^T \lambda$, that is, $\lambda = Q\,\alpha$, and introduce the *fluid traffic intensity* at station j as $\rho_j \stackrel{\text{def}}{=} \sum_{k \in s^{-1}(j)} \lambda_k \, m_k$ (in matrix form, $\rho = C\,M\,\lambda$). We will assume throughout the paper that $\rho < e$, with $e = (1, \ldots, 1)^T$ (sub-criticality).

Processes A, D, T, Z, W and Y will be used to measure the performance of the fluid network: $A(t)$ is the cumulative amount of fluid arrived (from outside and by feedback) by time t (to each fluid class) and $D(t)$ is the cumulative amount of fluid departing from each class (to other classes or to outside). $T(t)$ is the cumulative amount of processing time spent on each fluid class by time t. $Z(t)$ is the amount of fluid of any class in the system at time t. All the above processes are $K-$dimensional and the rest are $J-$dimensional: $W(t)$ denotes the workload or amount of time required by any server to complete processing of all fluid in queue, at time t, and $Y(t)$ is the cumulative amount of time that the server at each station has been idle in the interval $[0, t]$. By definition, T and Y are nondecreasing processes which depend on the specific service discipline, and $A(0) = D(0) = T(0) = Y(0) = 0$.

These processes are related by means of the following *fluid model equations*:

$$
\begin{aligned}
A(t) &= \alpha\, t + P^T D(t)\,, && (1)\\
Z(t) &= z + A(t) - D(t) = z + \alpha\, t - (I - P^T)\, M^{-1}\, T(t)\,, && (2)\\
D(t) &= M^{-1}\, T(t)\,, && (3)\\
C\, T(t) + Y(t) &= e\, t\,, && (4)\\
\int_0^\infty W_j(t)\, d\, Y_j(t) &= 0 \quad \text{for all } j = 1, \dots, J\,, && (5)\\
W(t) &= C\, M\, (z + A(t)) - C\, T(t)\,, && (6)
\end{aligned}
$$

Note that equation (5) expresses that for any station j, idle time Y_j can only increase when workload W_j is zero, that is exactly the meaning of a work-conserving discipline.

Let $\Psi(\cdot) \stackrel{\text{def}}{=} \big(A(\cdot),\, D(\cdot),\, T(\cdot),\, Z(\cdot),\, W(\cdot),\, Y(\cdot) \big)$ be any solution of the fluid model equations (1)-(6), which may not have in general a unique solution.

Definition 1 (Stability of the fluid limit model)**.** *We say that the fluid limit model* $(\alpha,\, M,\, C,\, P,\, z)$ *is stable if there exists* $t_0 > 0$ *such that for any solution* $\Psi(\cdot)$ *of the fluid model equations,* $Z(t) = 0 \;\; \forall t \geq t_0\, |z|$ *, where* $|z| \stackrel{\text{def}}{=} \sum_{k=1}^{K} z_k$ *.*

3 The main result

Note that from (6), (2) and (3) we can express the workload in terms of the queue process by means of

$$
W(t) = C\, M\, \big(z + A(t) - M^{-1}\, T(t) \big) = C\, M\, Z(t)\,, \tag{7}
$$

that is, for any j, $W_j(t) = \displaystyle\sum_{k \in s^{-1}(j)} m_k\, Z_k(t)$, which expresses workload at station j in terms of fluid amount for each class processed at that station. Next definition introduces a condition establishing that Z, in its turn, can be expressed in terms of W by means of a *"lifting"* deterministic matrix.

Definition 2. *Given a solution* $\Psi(\cdot)$ *of the fluid model equations associated to a fluid limit model* $(\alpha,\, M,\, C,\, P,\, z)$*, we say that the fluid limit model satisfies* **state space collapse** *with "lifting" matrix* Δ *if*

$$
\boxed{Z = \Delta\, W}
$$

where $\Delta = \big(\Delta_{kj}\big)_{kj}$ *with* $\Delta_{kj} = \delta_k > 0$ *if* $k \in s^{-1}(j)$ *and* 0 *otherwise. And we say that the "lifting" matrix* Δ *is* **regular** *if accomplishes the following technical restriction:* $C\, M\, Q\, \Delta$ *is invertible and matrix* R *defined by* $R \stackrel{\text{def}}{=} \big(C\, M\, Q\, \Delta\big)^{-1}$ *verifies assumption* **(HR)***:* R *can be expressed as* $I + \Theta$*, with* Θ *a square matrix such that the matrix obtained from* Θ *by replacing its elements by their absolute values, has spectral radius strictly less than* 1.

Roughly speaking, *state space collapse* assumption expresses that any fluid class k contributes a fixed portion δ_k to the workload at station $s(k)$. That is, the fluid classes processed at the same station are mixed in a fixed way in the station's queue.

Remark 1. *In the particular case $K = J$, if we assume for convenience (and without loss of generality) that $s(j) = j$ for any $j = 1, \ldots, J$, then $C = I$, (7) becomes $W = M\,Z$ and we trivially obtain state space collapse with regular "lifting" matrix $\Delta = M^{-1}$.*

Now we establish our main result. Recall that we assume $\rho < e$.

Theorem 1. *The fluid limit model is stable if verifies state space collapse with a regular "lifting" matrix Δ.*

The proof of the theorem is based on two lemmas formulated below. For the sake of completeness we introduce a known definition:

Definition 3 (R-regularization or Skorokhod problem)**.** *Let $\tilde{X}$ be a $J-$dim. stochastic process with continuous paths, defined on some probability space, with $\tilde{X}(0) \geq 0$, and $\tilde{R}$ a $J \times J$ matrix. We say that the pair $(\tilde{W}, \tilde{Y})$ of $J-$dim. stochastic processes defined on the same probability space and with continuous paths is a solution of the $\tilde{R}-$Skorokhod problem of $\tilde{X}$ in the first orthant $\mathbb{R}^J_+$ if:*

$$\tilde{W}(t) \in \mathbb{R}^J_+ \quad \text{for all } t \geq 0\,, \quad \tilde{W} = \tilde{X} + \tilde{R}\,\tilde{Y} \text{ a.s.}$$

$$\tilde{Y} \text{ has non}-\text{decreasing paths}\,,\ \tilde{Y}(0) = 0 \text{ and for any } j,\ \tilde{Y}_j \text{ only increases}$$

$$\text{if } \tilde{W} \text{ is on face } \{w \in \mathbb{R}^J_+ \,:\, w_j = 0\}\,,\ \text{that is}\,,\ \int_0^\infty \tilde{W}_j(t)\, d\,\tilde{Y}_j(t) = 0\,.$$

Remark 2. *Proposition 4.2 [6] shows that condition* **(HR)** *on a matrix $\tilde{R}$ is sufficient to ensure strong path-wise uniqueness of the solution.*

Lemma 1 (Lemma 5.1 [2])**.** *Assume $\rho < e$. Let $(\tilde{W}, \tilde{Y})$ be the (unique) solution of the $\tilde{R}-$Skorokhod problem on the first orthant of a process $\tilde{X}$, with $\tilde{R}$ verifying assumption* **(HR)***. If*

$$\tilde{W}(s) + \tilde{X}(t+s) - \tilde{X}(s) \geq \theta\, t \quad \text{for all } s,\, t \geq 0\,,$$

with $\theta = \tilde{R}\,(\rho - e)$, then we have that $\tilde{Y}(t+s) - \tilde{Y}(s) \leq (e-\rho)\, t$ for all $s,\, t \geq 0$, and hence $\tilde{Y}'(s) \leq (e - \rho)$ if $\tilde{Y}(\cdot)$ is differentiable at s and $\tilde{Y}'(\cdot)$ denotes its derivative.

Lemma 2 (Lemma 5.2 [2])**.** *Let $f \,:\, [0, +\infty) \longrightarrow [0, +\infty)$ be a nonnegative function that is absolutely continuous and let $\kappa > 0$ be a constant. Suppose that for almost surely (with respect to the Lebesgue measure on $[0, +\infty)$) all regular points t, $f'(t) \leq -\kappa$ whenever $f(t) > 0$. Then f is nonincreasing and $f(t) \equiv 0$ for $t \geq \frac{f(0)}{\kappa}$.*

Proof of Theorem 1: Consider a *fluid limit model* (α, M, C, P, z) with $\rho < e$ and satisfying *state space collapse* with a regular "lifting" matrix Δ. We want to prove the existence of some $t_0 > 0$ such that for any solution of the fluid model equations, $\Psi(\cdot) = \big(A(\cdot),\, D(\cdot),\, T(\cdot),\, Z(\cdot),\, W(\cdot),\, Y(\cdot)\big)$, $Z(t) = 0 \;\; \forall t \geq t_0\, |z|$.

Step 1: We will see that (W, Y) is the unique solution of the $R-$Skorokhod problem of X on the first orthant, X being defined by $X(t) \overset{\text{def}}{=} W(0) + \theta\, t$, and $R = (C\, M\, Q\, \Delta)^{-1}$. Indeed, from (2) we obtain $D(t) = z + A(t) - Z(t)$, which can be substituted in (1) giving

$$A(t) = Q\, \alpha\, t + Q\, P^T\, z - Q\, P^T\, Z(t)\,. \tag{8}$$

By *state space collapse* assumption with *regular* "lifting" matrix Δ, we can replace in (8) Z by $\Delta\, W$, and by substituting into (6) obtain

$$W(t) = W(0) + C\, M\, \big(Q\, \alpha\, t + Q\, P^T\, \Delta\, W(0) - Q\, P^T\, \Delta\, W(t)\big) - e\, t + Y(t)\,,$$

by using (4) and the fact that $W(0) = C\, M\, z$. By isolating $W(t)$ in its turn from this expression and taking into account the definition of R and the fact that $I + C\, M\, Q\, P^T\, \Delta = C\, M\, Q\, \Delta$, and that $\rho = C\, M\, Q\, \alpha$, we finally have that

$$W(t) = W(0) + R\, (\rho - e)\, t + R\, Y(t)\,. \tag{9}$$

If we denote $R\, (\rho - e)$ by θ as in Lemma 1, we have by (9) and (5) that (W, Y) is a solution of the $R-$Skorokhod problem of X on the first orthant. Assumption **(HR)** on matrix R given by the regularity of Δ, ensures the uniqueness of the solution. Therefore we can apply Lemma 1 because

$$W(s) + X(t + s) - X(s) \geq \theta\, t \quad \text{for all } s,\, t \geq 0\,,$$

which is easy to check since

$$W(s) + X(t+s) - X(s) = W(s) + (W(0) + \theta\, (t+s)) - (W(0) + \theta\, s) = W(s) + \theta\, t\,,$$

and $W \geq 0$. As a consequence, if Y is differentiable at point s,

$$Y'(s) \leq e - \rho\,. \tag{10}$$

Step 2: Take the Lyapunov function

$$g(t) = e^T\, R^{-1}\, W(t)\,, \tag{11}$$

to which we will apply Lemma 2. By substituting (9) into (11),

$$g(t) = e^T\, R^{-1}\, W(0) + e^T\, (\rho - e)\, t + e^T\, Y(t) = g(0) + \sum_{j=1}^{J} \big((\rho_j - 1)\, t + Y_j(t)\big)\,.$$

Then, the points of differentiability of $Y_j(\cdot)$ coincide with those of $g(\cdot)$, and if t is one of these points,

$$g'(t) = \sum_{j=1}^{J} \big((\rho_j - 1) + Y_j'(t)\big)\,, \tag{12}$$

$g'(t)$ being non positive by (10). We finish the proof using Lemma 2. To this end, let $t \geq 0$ be a point such that $g(t) > 0$ (if any). By definition of g and nonnegativity of all elements of R^{-1}, there exists i such that $W_i(t) > 0$. Then, by (5), $Y_i'(t) = 0$, and by (12),

$$g'(t) \leq \rho_i - 1 \leq \max_{j=1,\dots,J} \rho_j - 1 \quad (< 0 \quad \text{because } \rho < e) . \tag{13}$$

Thus, we have proved that $g'(t) \leq -\kappa$, with $\kappa = 1 - \max_{j=1,\dots,J} \rho_j > 0$, at any point t of differentiability of $g(\cdot)$ such that $g(t) > 0$. Lemma 2 ensures that, in this situation, $g(\cdot)$ is non-increasing and that $g(t) \equiv 0$ for $t \geq \frac{g(0)}{\kappa}$.

Finally we have that

$$g(0) = e^T R^{-1} W(0) = e^T R^{-1} C M z \leq e^T R^{-1} C M e |z| ,$$

and therefore, by Lemma 2,

$$g(t) \equiv 0 \text{ for any } t \geq t_0 |z| , \quad \text{with } t_0 = \frac{e^T R^{-1} C M e}{1 - \max_{j=1,\dots,J} \rho_j} > 0 .$$

On account of (11) and the nonnegativity of the elements of R^{-1} we also obtain that $W(t) \equiv 0$ for any $t \geq t_0 |z|$, and the same applies for Z. □

Remark 3. *In the particular case $J = 1$ (a $\bigvee -$system), (12) becomes $g'(t) = (\rho_1 - 1) + Y_1'(t)$, and (13) in its turn, $g'(t) = \rho_1 - 1 < 0$. The rest of the proof follows similarly with $\kappa = 1 - \rho_1 > 0$. Note that we do not use (10) in this situation, so actually we do not need Lemma 1. As a consequence, if $J = 1$ we have that $\rho_1 < 1$ is sufficient to ensure stability (see Theorem 6.1 [2]).*

References

[1] Bramson, M.: State space collapse with application to heavy traffic limits for multiclass queueing networks. Queueing Syst. **30**, 89-148 (1998).

[2] Dai, J. G.: On positive Harris recurrence of multiclass queueing networks: a unified approach via fluid limit models. Ann. Appl. Prob. **5(1)**, 49-77 (1995).

[3] Delgado, R.: State space collapse for asymptotically critical multi-class fluid networks. Queueing Syst. **59**, 157-184 (2008).

[4] Peterson, W. P.: Diffusion approximations for networks of queues with multiple customers types. Math. Oper. Res. **16**, 90-118 (1991).

[5] Reiman, M. I.: Open queueing networks in heavy traffic. Math. Oper. Res. **9(3)**, 441-458 (1984).

[6] Williams, R. J.: An invariance principle for semimartingale reflecting Brownian motions in an orthant. Queueing Syst. **30**, 5-25 (1998).

[7] Williams, R. J.: Diffusion approximations for open multi-class queueing networks: sufficient conditions involving state space collapse. Queueing Syst. **30**, 27-88 (1998).