Информационные процессы, Том 3, № 2, 2003, стр. 138–150.

MATHEMATICAL MODELS

Lipschitz Continuity and Unique Solvability of Fluid Models of Queueing Networks [1]

I. Nedaiborshch*, K. Nikolaev*, and A. Vladimirov**

**Moscow Institute of Physics and Technology, Moscow, Russia*
***Institute for Information Transmission problems,*
Moscow, Bol'shoi Karetnyi, 19
e-mail: vladim@iitp.ru
Received September 1, 2003

Abstract—Deterministic fluid models imitate the long-term behavior of stochastic queueing networks. We address the problem of unique solvability of these models under arbitrary time-dependent workloads. This property is stronger than the stability under a constant workload; in particular, it ensures the "smooth" behavior of the network under supernominal inflows. Sufficient conditions of unique solvability may be expressed in terms of Lipschitz constants of open networks in various norms. We find precise Lipschitz constants of a single multiclass server station under FIFO, generalized processor sharing, and priority service disciplines. The best possible Lipschitz constant for a two-class server is proved to be given by a queue-equalizing discipline. For proofs we use the technique of hysteresis operators, in particular, polyhedral Skorokhod problems.

1. INTRODUCTION

The problem of stability of fluid models of stochastic queueing networks was drawing considerable attention ever since the appearance of surprising examples of unstable networks that satisfy the nominal stability conditions, see [2, 13, 16, 17]. These counterexamples demonstrate the existence of diverging solutions of the fluid model while the trivial solution with empty queues also exists. This is a particular case of multiple solutions of a fluid model under a linear inflow but not the only one, see [9].

The property of stability of a fluid model under a subnominal workload is, under mild assumptions, equivalent to the property of unique solvability of this model under the same workload. Indeed, a trivial solution with empty queues exists in this case and the existence of any nontrivial solution implies instability. Vice-versa, for an unstable model, by simple compactness and homogeneity considerations, one can construct a non-trivial solution from zero.

The property of unique solvability under any continuous inflow is a stronger one and it can give us more insight into the behavior of the queueing network. Namely, this property would guarantee that during periods of supernominal inflow the queues behave in a predictable manner demonstrating, for instance, no auto-oscillations under a smoothly changed workload.

The goal of this research is to find maximally tight sufficient conditions of unique solvability of several common fluid models such as FIFO, priority, and generalized processor sharing (GPS) ones. We realize that this problem, in general, might be a hard one (apart from special cases like Jackson models) as is also the case with the stability problem. The importance of the unique solvability property, however, justifies the attempt.

[1] This work was partially supported by grant 03-01-00258 of the Russian Foundation for Fundamental Research, by grant for Scientific Schools 1532.2003.1 of the President of Russian Federation, and by the project "Development of mathematical methods of analysis of distributed asynchronous computational networks" of the program of fundamental research of OITVS RAN (Branch of Information Technologies and Computer Systems of Russian Academy of Sciences) "New physical and structural methods in infocommunications".

Some progress for FIFO networks was reported in [9], where sufficient conditions for the uniqueness in the restricted class of linear solutions of a reentrant line were derived.

We suggest several methods of research. First, for finite-dimensional disciplines like the priority disciplines, processor sharing disciplines, and their combinations, the dynamics of the fluid model is described by a *Skorokhod problem* on the positive orthant and one may use known criteria of unique solvability of Skorokhod problems and, more generally, *sweeping processes* that are, in turn, particular cases of *hysteresis operators*, see [15, 14, 10, 12].

Such service disciplines as FIFO and LIFO, however, cannot be modelled by a finite-dimensional Skorokhod problem because their state spaces are essentially infinite-dimensional. Another option is to find the Lipschitz constant of a single server input-output operator in an appropriate norm and then to find conditions on the routing matrix R that would ensure the unique solvability of the network. For instance, $L_W L_R < 1$ is such a condition, where L_W is the Lipschitz constant of the network without feedback, that is, of the parallel combination of all servers, and L_R is the norm of the linear operator R in an appropriate space.

We demonstrate that a single server with two classes of customers of equal viscosity and the FIFO discipline has the precise Lipschitz constant 3 for the max-norm $\|x(\cdot)\| = \max_{t,i} |x_i(t)|$. For k classes, we get $L = 2k - 1$, which is also true for $k = 1$.

The Lipschitz constants in the max-norm for the priority and the GPS disciplines happen to be exactly the same as for the FIFO discipline, that is, $2k - 1$ for k classes of customers. We do not know if this is just a coincidence or there exists a special property of these disciplines that ensures this value of the Lipschitz constant.

The theoretical lower bound for the Lipschitz constant among all work-conserving disciplines is found to be 2, and it is realized by the queue-equalizing discipline that always gives priority to the class of customers with the longest queue.

Fluid models of queueing networks become particular cases of so called hysteresis operators (possibly, multivalued) if the notion of variable service effort is included into the model (this is the maximal amount of customers that can be served up to time t). The input-output fluid operators, apart from characteristic hysteresis properties of rate-independence and causality, possess also the homogeneity property (characteristic for unrestricted fluid models) and, in some cases, the short-memory property. For the research of Lipschitz continuity of fluid operators we use basic properties of short-memory hysteresis operators.

The authors wish to thank A. Rybko and S. Pirogov for valuable discussions and remarks concerning this work.

2. FLUID MODELS OF QUEUEING NETWORKS

2.1. Queueing Networks

A *queueing network* (QN) consists of a finite number of *server stations* that serve incoming customers and then reroute them to other stations (including itself) or to the exit. The throughput of the server at each station is limited, so the customers that have arrived at the station and cannot be served immediately wait in the *queue*. The whole process is, in general, stochastic.

For a *multiclass* QN, customers of a finite number of classes $i \in J = \{1, \dots, k\}$ are circulating in the network. Each class may be present at a unique server station, but at a given station more than one class may be present. A *service discipline* at a given station is used for the choice of the next customer from the queue as soon as the previous one is released. For each class of customers, its service time distribution is given. After being served at the station, the customer of class $i \in J$ changes its class to $j \in \bar{J} = \{0, 1, \dots, k\}$ according to the *routing policy* of the station and goes to the corresponding station or to the exit if $j = 0$. For a general description of multiclass queueing networks see [3, 4, 7, 8].

A *fluid model* (FM) is a deterministic analogue of a QN whose set of solutions contains all the Euler limits of the original stochastic model, see, for instance, [16, 4, 18]. In this paper we study the properties of fluid models.

In order to describe the fluid dynamics of the network, it suffices to do that for each node separately. This is the same as to describe the service discipline plus the routing rule for this node or to specify the set of all *solutions* at the node, that is, of all possible combinations of flows $v_j(\cdot)$, $f_{ji}(\cdot)$, $j \in J(i)$, through node i for any given initial queue $\mathbf{x}_i^0 = (x_1^0, \ldots, x_{n_i}^0)$. Here $v_j(t)$ is the mass of fluid j that came to the server up to time t (hence, a nondecreasing function) and $f_{ji}(t)$ is the mass of fluid that has changed class from j to i up to time t.

The routing rule is usually given by a $k \times k$ routing matrix R such that $r_{ij} \geq 0, \sum_j r_{ij} \leq 1, 1 \leq i, j \leq k$. Some of the most popular service disciplines are described below. Fluid limits and fluid models were introduced in [16], see also [4, 18].

2.2. The FIFO Discipline

The work of a single FIFO server with incoming fluids $u_i(t)$ of k classes and a single service effort $s(t)$ can be formalized as follows. First, let us assume that the server capacity is the same for all classes of fluid (one also says that their viscosities are equal). The cumulative outflow $w(t) = w_1(t) + \cdots + w_k(t)$ is found explicitly by the formula

$$w(t) = W[u(\cdot), s(\cdot)](t) = \inf_{\tau \leq t} [u(\tau) + (s(t) - s(\tau))], \tag{2.1}$$

where $u(t) = u_1(t) + \cdots + u_k(t)$. This is a standard way to describe the deterministic service process at a single server, see [1]. Below we derive (2.1) as a limit of discrete-time processes with vanishing lengths of customers.

The functions $u_i(t)$ and $s(t)$ are nondecreasing at $[0, T]$ and we will usually assume them to be right-continuous. For discontinuous inflows, the FIFO service rule should be supplemented by a rule that would handle the case of simultaneous arrivals. Another option is to allow multiple outputs for the same input and, hence, to consider a multivalued input-output map.

The value of $u(t)$ is equal to the cumulative mass of fluid that has arrived at the server station up to time t. Thus, we consider the situation with empty queues at $t = 0$. If all the flows are continuous, the range of $u(t), 0 \leq t \leq T$, is the segment $[0, u(T)]$. For each $u \in [0, u(T)]$, denote by $U_i(u)$ the value of $u_i(t)$, where $u(t) = u$. Such a t is not unique, in general, but the values $U_i(u)$ are uniquely determined on $[0, u(T)]$, nondecreasing, and Lipschitz continuous with constant 1:

$$0 \leq U_i(u_+ - u_-) \leq u_+ - u_-, \qquad 0 \leq u_- \leq u_+ \leq u(T).$$

The specific outflows $w_i(t)$ are found now as

$$w_i(t) = U_i(w(t)), \qquad w(t) = W[u(\cdot), s(\cdot)](t), \qquad 0 \leq t \leq T, \;\; i = 1, \ldots, k.$$

They are right-continuous (continuos) functions whenever $u(\cdot)$ and $s(\cdot)$ are right-continuous (continuos).

2.3. Priority Disciplines

Let, again, the inflows $u_i(t)$ of k classes of fluid be given. Let the lower indices have priority over the higher ones. The outflows $w_i(t)$ under the service effort $s(t)$ are then defined as follows. The first fluid behaves as there were no other fluids: $w_1(t) = W[u_1(\cdot), s_1(\cdot)](t)$, $t \geq 0$, where $s_1(t) = s(t)$, $t \geq 0$, by definition. Then we set $s_2(t) = s_1(t) - w_1(t)$ and define recursively

$$w_i(t) = W[u_i(\cdot), s_i(\cdot)](t), \quad s_i(t) = s_{i-1}(t) - w_{i-1}(t), \quad i = 2, \ldots, k.$$

2.4. Generalized Processor Sharing

The generalized processor sharing (GPS) service discipline is characterized by the weight vector $p = (p_1, \ldots, p_k)$, $p_i \geq 0$, $\sum_{i=1,\ldots,k} p_i = 1$. Normally, the processor (server) shares its service effort among the k classes proportionally to their weights. In each class, the customer with the earliest arrival time is served. If one or more classes are empty at a given time, their shares of service effort are distributed among the other ones proportionally to their weights.

This discipline was thoroughly studied in [6]. The fluid model for a single server is described by a Skorokhod problem with oblique reflection on the positive orthant. The input-output operator is Lipschitz continuous in the max-norm $\|u(\cdot)\| = \sup_{t\in[0,T]} |u(t)|$ for continuous inputs and outputs, and this result can be extended to the space of llrc (left-limit right-continuous) functions with the same norm.

2.5. General Uniqueness Theorems

The contraction principle can be used for the proof of unique solvability of a fluid model under a given initial condition. Let us first consider the set of all nodes of the network without feedback and suppose that we know the Lipschitz constant L_W of the corresponding input-output operator W with respect to some norm $\|\cdot\|$ in the k-dimensional functional space of flows. Suppose we also know the norm L_R of the routing matrix R as a linear operator in the same space. We have

$$w(\cdot) = W(u(\cdot) + Rw(\cdot)) \tag{2.2}$$

for any solution $w(\cdot)$ of the fluid model, where $u(\cdot)$ is the exogenous inflow.

Suppose two different solutions $w(\cdot)$ and $w'(\cdot)$ exist for the same inflow $u(\cdot)$. Then from (2.2) we get

$$\|w'(\cdot) - w(\cdot)\| \leq L_W L_R \|w'(\cdot) - w(\cdot)\|.$$

This gives us a simple sufficient condition of unique solvability of the fluid model: $L_W L_R < 1$.

Thus, a construction of a sufficient condition of unique solvability for a fluid network may proceed as follows. First, an appropriate norm is chosen in the space of flows. Then upper bounds for the Lipschitz constants of single servers should be found. From these, an upper bound L_W for the Lipschitz constant of the whole network without feedback, that is, for the the parallel composition of all the nodes, is derived. Then we know that the model is uniquely solvable under each inflow if $L_W L_R < 1$.

The above condition is rather rough and far from being tight in many cases. Its advantage is, however, in its universality. In particular, it can be applied to networks with mixed service disciplines and variable service effort. Our next goal is to find Lipschitz constants of a single server under common service disciplines in various norms.

3. SINGLE-CLASS SERVERS

3.1. Time-dependent Single-Class Servers

Here we study the simplest elements of a queueing network, that is, single-class server stations. The necessity to consider time-dependent inflow and service effort intensities is well justified by occurrence of such phenomena in real networks. The situation can be modelled by a leaky bucket with a hole of variable diameter. Another, formal reason in favor of time-dependent service efforts is that in this generality the fluid operator becomes a *rate-independent* one, that is, a hysteresis operator which can often be studied by special methods of nonlinear analysis [10, 11].

The dynamics of the server will be described by an input-output operator $W : \{u(\cdot), s(\cdot)\} \to \{w(\cdot), x(\cdot)\}$ (the initial queue length will be included in the input). Both the inflow $u(t)$ and the service effort $s(t)$ are nondecreasing (deterministic) functions of t on the domain $[0, T]$. Moreover, we assume $u(0) \geq 0$ (the value of $u(0)$ is interpreted as the initial queue at the server) and $s(0) = 0$. Without loss of generality we will assume here $T = 1$. This can be achieved by a monotone change of time because of the rate-independence property of operator W (see below).

The outflow $w(t)$ is, again, a nondecreasing function on $[0, 1]$, $w(0) = 0$, and $x(\cdot)$ is a difference of two nondecreasing functions because of the obvious balance relation $x(t) = u(t) - w(t)$. Hence, $x(t)$ is a function of bounded variation on $[0, 1]$ (its variation is bounded from above by, for instance, $2u(1)$). The same is, of course, true for the remaining components of input and output. A function of bounded variation possesses finite left and right limits at each point of the interior of its domain $[0, 1]$.

Now, we are ready to give a formal definition of the operator W. First, we will define the action of W for step functions $u(\cdot)$ and $s(\cdot)$ and then extend the definition to the general case. Let a finite partition $0 = t_0 < t_1 < \cdots < t_k = 1$ of the interval $[0, 1]$ be given and let both $u(\cdot)$ and $s(\cdot)$ be step-functions whose break points are t_i, $i = 1, \ldots, k$. For definiteness, let us assume them to be right-continuous (this, in particular, means that $u(\cdot)$ and $s(\cdot)$ are continuous at $t_0 = 0$). Both $u(\cdot)$ and $s(\cdot)$ are determined by their values $u_i = u(t_i)$, $s_i = s(t_i)$, $i = 1, \ldots, k$. We will also use analogous notation for the outputs $x(\cdot)$ and $w(\cdot)$.

The evolution of the system proceeds as follows. At each time instant t_i, the amount $u_i - u_{i-1}$ of customers is added to the queue and the amount $s_i - s_{i-1}$ is subtracted from the queue if the length of queue x_{i-1} allows that, that is, if $x_{i-1} + u_i - u_{i-1} \geq s_i - s_{i-1}$. Otherwise, all the customers leave the system at $t = t_i$. This recursive rule can be written as

$$x_i = \max\{0, x_{i-1} + (u_i - u_{i-1}) - (s_i - s_{i-1})\}, \qquad i = 1, \ldots, k, \tag{3.1}$$

and we set $x_0 = u_0$. According to the balance relation, we set $w_i = u_i - x_i$, $i = 0, \ldots, k$. Thus, the operator W is well defined for all right-continuous piecewise constant (step) inputs, and the outputs are, again, piecewise constant step functions.

We will extend the definition of operator W to the space BV of functions of bounded variation as follows. Using the notation $p(t) = u(t) - s(t)$, $q(t) = \min\{0, p(t)\}$, we set

$$x(t) = p(t) - \inf_{t_0 \leq \tau \leq t} q(\tau), \tag{3.2}$$

$$w(t) = u(t) - x(t) = s(t) + \min\{0, \inf_{t_0 \leq \tau \leq t} (u(\tau) - s(\tau))\}. \tag{3.3}$$

Obviously, for nondecreasing step functions $u(\cdot)$ and $s(\cdot)$, the definition (3.2)-(3.3) coincides with (3.1). Note that the outputs are well defined for any pair $(u(\cdot), s(\cdot))$ of functions of bounded variation on $[0, 1]$ and that $w(\cdot)$ is nondecreasing on $[0, 1]$, and $x(\cdot)$ is a function of bounded variation on $[0, 1]$, though not necessarily monotone.

In order to make formulas (3.2) and (3.3) more transparent, we will reduce them to the input-output relations of elementary hysteresis operators, namely, of one-sided *play* and *stop*. The one-sided play operator for the input $u(t)$, $0 \leq t \leq 1$, is defined as

$$Pu(t) = \min\{0, \inf_{\tau \in [0,t]} u(\tau)\}, \tag{3.4}$$

and the stop output is equal to

$$Su(t) = u(t) - Pu(t). \tag{3.5}$$

Obviously, $Pu(\cdot)$ is always a nonincreasing function of t and $Su(\cdot)$ is nonnegative. Note that the operators P and S are well defined not only for inputs of bounded variation but for any $u(\cdot)$ such that $\inf_{t \in [0,1]} u(t) > -\infty$. For our needs, it will be sufficient to consider P and S on the space M of bounded functions defined on $t \geq 0$.

The one-sided play and stop are examples of hysteresis operators, that is, they are *causal* (the output on any initial time interval depends only on the input constrained to this interval, and not on the future) and rate-independent (the pairs input-output withstand any monotone increasing continuous change of time, that is, after such a change the resulting output is, again, the output of the corresponding hysteresis operator with the resulting input).

The causality and rate-independence follow immediately from definitions. It is also important to notice that, as a consequence of rate-independence, the operators P and S depend essentially only on the order structure of the real axis, that is, they can be naturally extended to any linearly ordered set. In particular, it is often instructive to study the properties of P and S on the set $\mathbb{Z}$ of integers.

The intuitive interpretation of these operators is as follows. The play operator models the motion of a point on a straight line that is pushed by another point (the input) from the right, without inertia. The stop is the same as the original (one-dimensional) Skorokhod problem, where the output point follows the motion of the input point as long as the output is on the positive half-axis, and it stays equal to zero whenever the motion of input point tries to drive it into the negative half-axis.

Now, for a single-class server, we may write that $w(t) = s(t) + Pp(t)$ and $x(t) = Sp(t)$, where $p(t) = u(t) - s(t)$. Let us study the basic properties of operators P and S. Note first that, apart from rate-independence and causality that are inherent to all hysteresis operators, an additional property of *homogeneity* holds for P and S, that is, if $Z(\cdot) = \{u(\cdot), s(\cdot), x(\cdot), w(\cdot)\}$ is an admissible flow on the server, then the same is true for

$$\alpha Z(\cdot) = \{\alpha u(\cdot), \alpha s(\cdot), \alpha x(\cdot), \alpha w(\cdot)\}, \qquad \alpha \geq 0.$$

3.2. Lipschitz Constants

Let us address the continuity properties of P and S, in particular, their Lipschitz continuity constants. The inputs and outputs of both operators will be considered as elements of two normed spaces M and BV in all possible combinations. The space M consists of functions bounded on $[0, 1]$. The norm in M is given by

$$\|u(\cdot)\| = \sup_{t\in[0,1]} |u(t)| \tag{3.6}$$

The space BV consists of functions of bounded variation on $[0, 1]$ and the norm in BV is

$$\|u(\cdot)\|_V = |u(0)| + \mathrm{Var}_{[0,1]} u(\cdot). \tag{3.7}$$

(C1–C2) First, as operators from $M \cap BV$ with the induced norm from M to BV, both P and S are discontinuous. As an example, let us consider $u_1(t) = -t$ and $u_2^m(t) = -1/m[mt]$, where $[x]$ is the maximal integer $i \leq x$. As $m \to \infty$, the series of functions $w_2^m(\cdot) = Pu_2^m(\cdot) \equiv u_2^m(\cdot)$ converges to $w_1(\cdot) = Pu_1(\cdot) \equiv u_1(\cdot)$ in M but not in BV because the variation of $w_1(\cdot) - w_2^m(\cdot)$ on $[0, 1]$ remains equal to 1 for each m.

(C3) The play is Lipschitz continuous with constant 1 for P acting from M to M. Indeed, this is true for the map $u(\cdot) \to b(\cdot)$, where $b(t) = \inf_{\tau\in[0,t]} u(\tau)$ and then, for the map $b(\cdot) \to p(\cdot)$, where $p(t) = \min\{0, b(t)\}$.

(C4) The stop is Lipschitz continuous with constant 2 as an operator acting from M to M. The value of 2 is an upper bound for the constant because $Su(t) \equiv u(t) - Pu(t)$ and P is Lipschitz continuous with constant 1. The following example demonstrates that this bound is precise, see Fig. 3.2.

(C5–C6) Let us prove that the play P is Lipschitz continuous with constant 1 from BV to both BV and M. It suffices to demonstrate that

$$\mathrm{Var}_{[0,1]}(x(\cdot) - x'(\cdot)) \leq \mathrm{Var}_{[0,1]}(u(\cdot) - u'(\cdot)) \tag{3.8}$$

for any pair of inputs $u(\cdot)$, $u'(\cdot)$, $u(0) = u'(0)$, and their outputs $x(\cdot)$, $x'(\cdot)$, $x(0) = x'(0)$.

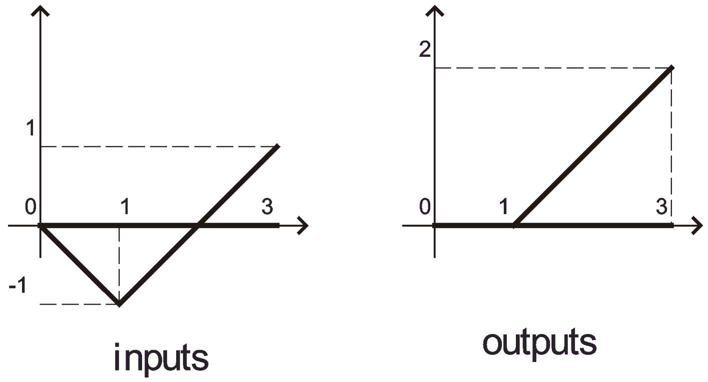

Fig. 1. Lipschitz constant for $S : M \to M$

Let us denote $a(t) = u(t) - u'(t)$, $b(t) = x(t) - x'(t)$ and choose a finite sequence of times $0 < t_1 < t_2 < \cdots < t_k \leq 1$ such that

$$0 < b(t_1) > b(t_2) < b(t_3) > \dots b(t_k) \quad \text{or} \quad 0 > b(t_1) < b(t_2) > b(t_3) < \dots b(t_k).$$

For definiteness, we will consider the first case.

Since $0 \geq x(t_1) > x'(t_1)$, for any $\varepsilon > 0$ there exists a $t_1^* \in [0, t_1]$ such that $u'(t_1^*) < x'(t_1) + \varepsilon$. From $u(t_1^*) \geq x(t_1^*)$ we get

$$a(t_1^*) = u(t_1^*) - u'(t_1^*) > b(t_1) - \varepsilon.$$

Then, since $x(t_2) < x(t_1)$, there exists a $t_2^* \in (t_1, t_2]$ such that $u(t_2^*) < x(t_2^*) + \varepsilon$. We also have $u'(t_2^*) \geq x'(t_2^*) \geq x'(t_2)$ and, hence,

$$a(t_2^*) = u(t_2^*) - u'(t_2^*) < b(t_2) + \varepsilon.$$

Proceeding in the same manner we conclude that

$$\mathrm{Var}_{[0,1]} a(\cdot) \geq \mathrm{Var}_{[0,1]} b(\cdot) - n\varepsilon - \delta,$$

where

$$\delta = \mathrm{Var}_{[0,1]} b(\cdot) - \sum_{i=1}^{k} |b(t_i) - b(t_{i-1})|.$$

Since both $\varepsilon > 0$ and $\delta > 0$ can be taken arbitrarily small, we get inequality (3.8). The constant 1 is precise, as follows from trivial examples.

(C7) It follows from above that the stop is Lipschitz continuous with constant 2 from BV to BV. The example in Fig. 3.2 demonstrates that this constant is precise.

(C8) The stop is Lipschitz continuous from BV to M with constant 1. To prove it, let us show that the inequality

$$|x(t) - x'(t)| \leq \mathrm{Var}_0^t (u(\cdot) - u'(\cdot))$$

holds for any $t \in [0, 1]$. It obviously holds for $t = 0$. Then, suppose that it holds for t_i and prove it for t_{i+1}. It suffices to notice that $|x'(t) - x(t)|$ does not increase if the inputs $u(\cdot)$ and $u'(\cdot)$ are parallel, that is, if $u'(t) \equiv u(t) + c$. The Lipschitz constant 1 is precise, as follows from trivial examples.

The Lipschitz constants for the server operators $u(\cdot) \to x(\cdot)$, $u(\cdot) \to w(\cdot)$ with the fixed service effort $s(\cdot)$ and for $s(\cdot) \to x(\cdot)$, $s(\cdot) \to w(\cdot)$ with the fixed inflow $u(\cdot)$ can be easily derived from (C1)–(C8). One of these results will be formulated as a lemma.

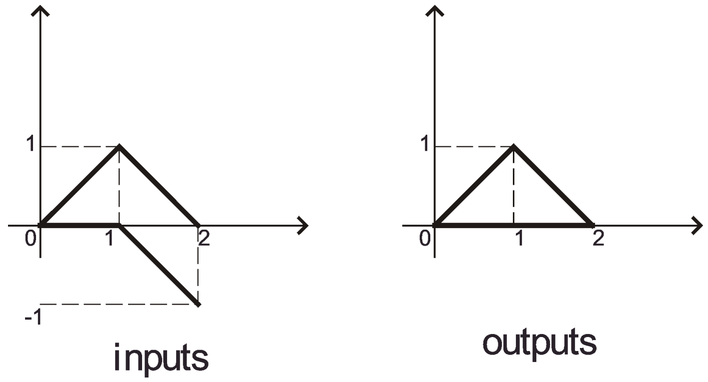

Fig. 2. Lipschitz constant for $S : BV \to BV$

Lemma 3.1. *The operator* $u(\cdot) \to w(\cdot)$ *is a contraction in both* M *and* BV. *The Lipschitz constant* 1 *is tight in both cases.*

For operators of type $(u(\cdot), s(\cdot)) \to (x(\cdot), w(\cdot))$ with different combinations of norms for the input and output, one should combine the results (C1)–(C8) according to the norms chosen.

4. LIPSCHITZ CONTINUITY OF 2-CLASS SERVER STATIONS

4.1. FIFO Servers

Let us set $s(t) = t$, $t \geq 0$, and study the properties of the input-output operator $F : u(\cdot) \to w(\cdot)$. First, we assume $k = 2$. Denote by C the space of continuous functions $x(\cdot) : [0, T] \to \mathbb{R}^2$ with the max-norm

$$\|x(\cdot)\| = \max_{t\in[0,T]} \max_{i=1,2} |x_i(t)| \tag{4.1}$$

and consider F as an operator from C to C.

The aggregate input-output operator $W : (u_1(\cdot) + u_2(\cdot)) \to (w_1(\cdot) + w_2(\cdot))$ also acts from C to C; it is Lipschitz continuous with precise constant 1 according to Lemma 3.1.

Theorem 4.1. *The operator* F *is Lipschitz continuous with the precise constant* 3.

Proof. Since the class of solutions $(u(\cdot), w(\cdot))$ is invariant to transformations of the form $\tilde{u}(t) = \varepsilon u(t/\varepsilon)$, $\tilde{w}(t) = \varepsilon w(t/\varepsilon)$ for any $\varepsilon > 0$, it suffices to prove that $\|u'(\cdot) - u(\cdot)\| \leq 1$ implies $\|w'(\cdot) - w(\cdot)\| \leq 3$. For each $t^* \in [0, T]$, there exist $t, t' \leq t^*$ such that $w(t^*) = u(t)$ and $w'(t^*) = u'(t')$. Without loss of generality we assume $t' \geq t$. Let us prove, say, that

$$|w_1'(t^*) - w_1(t^*)| \leq 3. \tag{4.2}$$

By Lemma 3.1, we have $u(t) - 2 \leq u'(t') \leq u(t) + 2$. By monotonicity, we also get $u_2'(t') \geq u_2'(t)$ and, hence, $u_2'(t') \geq u_2(t) - 1$. Since $w(t^*) = w_1(t^*) + w_2(t^*) = u_1(t) + u_2(t)$ and $w'(t^*) = w_1'(t^*) + w_2'(t^*) = u_1'(t) + u_2'(t)$, it follows that

$$w_1'(t^*) = u'(t') - u_2'(t') \leq u(t) + 2 - u_2(t) + 1 = w_1(t^*) + 3.$$

On the other hand,

$$w_1'(t^*) = u_1'(t') \geq u_1'(t) \geq u_1(t) - 1 = w_1(t^*) - 1,$$

and, hence, (4.2) follows.

As an example, where the upper bound 3 is attained, let us consider a pair of piecewise linear inputs $u(\cdot)$, $u'(\cdot)$ (their values are given at integer points $t = 0, 2, 4, 6, 8 = T$):

$$\begin{aligned} u_1(\cdot) &= (0, 1, 1, 3, 5), \quad & u_2(\cdot) &= (0, 1, 1, 3, 3), \\ u'_1(\cdot) &= (0, 2, 2, 4, 6), \quad & u'_2(\cdot) &= (0, 2, 2, 2, 2), \end{aligned}$$

that is, $\|u(\cdot) - u'(\cdot)\| \le 1$. The outputs are

$$\begin{aligned} w_1(\cdot) &= (0, 1, 1, 2, 3), \quad & w_2(\cdot) &= (0, 1, 1, 2, 3), \\ w'_1(\cdot) &= (0, 1, 2, 4, 6), \quad & w'_2(\cdot) &= (0, 1, 2, 2, 2), \end{aligned}$$

and $w'_1(8) - w_1(8) = 3$.

4.2. Another Norm

Let us consider the norm $\|x(\cdot)\|_s = \max_{t\in[0,T]}(|x_1(t)| + |x_2(t)|)$.

Theorem 4.2. *The operator F is Lipschitz continuous with the precise constant* 3 *with respect to the norm* $\|\cdot\|_s$.

Proof. Let $\|u'(\cdot) - u(\cdot)\|_s \le 1$. Again, we choose an arbitrary $t^* \in [0, T]$ and $t \le t' \le t^*$ such that $w(t^*) = u(t)$ and $w'(t^*) = u'(t')$. We have $|(u_1(t) + u_2(t)) - (u'_1(t) + u'_2(t))| \le 1$, $t \in [0, T]$, and, by Lemma 3.1,

$$|(w_1(t^*) + w_2(t^*)) - (w'_1(t^*) + w'_2(t^*))| \le 1.$$

Suppose that $|w'_1(t^*) - w_1(t^*)| + |w'_2(t^*) - w_2(t^*)| > 3$. Then either $w'_1 < w_1 - 1$ or $w'_2 < w_2 - 1$. This, however, is impossible because $u'_i(t') \ge u'_i(t)$ and $|u_i(t) - u'_i(t)| \le 1$, $i = 1, 2$. Hence, the operator F is Lipschitz continuous in the norm $\|\cdot\|_s$ with constant at most 3.

In order to prove that this constant is precise, let us consider the piecewise linear inputs (the values are taken at $t = 0, 2, 4$).

$$\begin{aligned} u_1(\cdot) &= (0, 1, 3), \quad & u_2(\cdot) &= (0, 0 \to 2, 2), \\ u'_1(\cdot) &= (0, 1, 3), \quad & u'_2(\cdot) &= (0, 1, 1), \end{aligned}$$

where $u_2(t)$ changes its value from 0 to 2 in a small neighborhood of $t = 2$.

Note that the norm of any routing matrix R induced by the norm $\|\cdot\|_s$ is less than or equal to 1, and the upper bound for this norm is the maximal sum of elements in the columns of R.

4.3. The BV-norm for FIFO

Let us consider the norm $\|x\|_{BV} = \mathrm{Var}(x_1) + \mathrm{Var}(x_2)$. In this norm the flow operator is not Lipschitz continuous. Indeed, let us consider the inflow $u_1(t) = (t + \sin t)/2$, $u_2(t) = t - u_1(t)$ to the unit capacity server. The outflow $w(\cdot)$ coincides with $u(\cdot)$. Then we add a small amount of any fluid at some finite time interval (this disturbance has a small BV-norm). The effect on the outflow at the long range is a small delay of the periodic intensity function, but this implies the infinite BV-difference with the original output on $[0, +\infty)$.

4.4. Priority Servers

Let $k = 2$, $s(t) = t$, $t \in [0, T]$, and $\|u'(\cdot) - u(\cdot)\| \le 1$. Then, by Lemma 3.1, $|w'_1(t) - w_1(t)| \le 1$ and, hence,

$$|s'_2(t) - s_2(t)| \le 1, \qquad 0 \le t \le T. \tag{4.3}$$

We will need the following assertion which is an immediate consequence of (C3).

Lemma 4.1. *For a server with inflows $u(t), u'(t)$, the service efforts $s(t), s'(t)$, and the respective outflows $w(t), w'(t)$, the inequalities*

$$|u'(t) - u(t)| \le a, \qquad |s'(t) - s(t)| \le b, \quad 0 \le t \le T,$$

imply

$$|w'(t) - w(t)| \le a + 2b, \quad 0 \le t \le T.$$

Theorem 4.3. *The operator G is Lipschitz continuous with the precise constant* 3.

Proof. By (4.3) and Lemma 4.1, we get an upper bound 3 for the Lipschitz constant. The tightness of this bound is proved by the following piecewise linear example:

$$u_1(t) = u_2(t) = \frac{t}{2}, \ \ t \ge 0,$$

$$u_1'(t) = \begin{cases} 0 & t \le 2 - e, \\ 2\frac{t-2}{\varepsilon} & 2 - \varepsilon < t \le 2, \\ \frac{t}{2} + 1, & t > 2, \end{cases} \qquad u_2'(t) = \begin{cases} 0 & t \le 2, \\ \frac{t}{2} - 1, & t > 2. \end{cases}$$

Indeed, $w_3(6) = 3$, $w_3'(6) = \varepsilon/2$, and we get the tightness of the bound 3 as $\varepsilon \to 0$.

4.5. GPS Servers

Here, again, we get the tight constant 3 for two classes of fluid with equal viscosities and shares. Let us do it by a Skorokhod problem method. We will also use this method for an alternative proof of the upper bound 3 for the priority discipline.

The Skorokhod problem for the GPS discipline on $\mathbb{R}_+^2$ has the reflection vectors $d_1 = (-1, 1)$, $d_2 = -d_1$, and $d_3 = (1, 1)$ for the virtual face $F = \{x : \langle x, d_3 \rangle = 0\}$, see [6]. In order to find an upper bound for the Lipschitz constant in some norm $\|\cdot\|$, one should consider a special set of broken lines $\{0 = x^0, x^1, \dots, x^k\}$ from 0 in $\mathbb{R}^2$ (not just in $\mathbb{R}_+^2$!). This set is defined as follows.

First, let us define three sets

$$S_i = \{x \in \mathbb{R}^2 : \inf_{y \in \mathbb{R}^2,\ y_i = 0} \|x - y\| \le 1\}, \quad i = 1, 2,$$

and

$$S_3 = \{x \in \mathbb{R}^2 : \inf_{y \in \mathbb{R}^2,\ y_1 = y_2} \|x - y\| \le 1\}.$$

We will say that the broken line $\{0 = x^0, x^1, \dots, x^k\}$ is a 1-path of the SP $\{d_1, d_2, d_3\}$ if, for each $j = 1, \dots k$, we have $x_j \in S_i$ and $x_j - x_{j-1} \in \mathbb{R} \cdot d_i$ for some $i = 1, 2, 3$. The supremum of $\|x_k\|$ for all 1-paths of the SP is an upper bound for the Lipschitz constant [19, 5]. Actually, the maximal norm of the endpoint of a 1-path is the maximal possible max-distance between $w(\cdot) - s(\cdot)$ and $w'(\cdot) - s'(\cdot)$ if the max-distance between $u(\cdot) - s(\cdot)$ and $u'(\cdot) - s'(\cdot)$ does not exceed 1. This is the required Lipschitz constant.

This supremum is equal to 3 for the GPS discipline and the max-norm in $\mathbb{R}^2$. Namely, a path $\{(0, 0), (1, 1), (3, -1)\}$ is a 1-path of the SP. It is also clear that $|x_1 + x_2| \le 2$ for the endpoint $x = (x_1, x_2)$ of any 1-path. Finally, $\min_{i=1,2} |x_i| \le 1$ for such an endpoint and, hence, $|x_i| \le 3$, $i = 1, 2$.

For the priority discipline, the SP has two reflection vectors $d_1 = (0, 1)$ and $d_2 = (1, 1)$. By similar argument we conclude that the upper bound for the max-norm is, again, 3.

5. LIPSCHITZ CONTINUITY OF MULTICLASS SERVER STATIONS

5.1. FIFO Servers

Let us consider a FIFO server with k classes of fluid with equal viscosities. In this case we get $L = 2k - 1$ (this is the precise constant). First, let us demonstrate that $L \geq 2k - 1$ by an example.

At the interval $[t_0, t_1]$ there is no inflow to the first server and, for the second server, $u'_i(t_1) = w'_i(t_1) = 1$. Then, at the interval $[t_1, t_2]$, $t_2 = t_1 + k - 1$, the input to the first server is linear and $u_i(t_2) = 3$, $i = 2, \dots, k$, $u_1(t_2) = 0$. We get, hence, $x_i(t_2) = 2$, $i = 2, \dots, k$. At the second server the input at this interval is also linear and $u'_i(t_2) = 2$, $i = 2, \dots, k$, and, again, $u'_1(t_2) = u'_1(t_1) = 1$. The queue $x'(t_2)$ is empty. Finally, at $[t_2, t_3]$, $t_3 = t_2 + 2k - 2$, we set $u'_1(t_2 + t) = 1 + t$ and $u_1(t_2 + t) = t$. We get then $w_1(t_3) = 0$, $w'_1(t_3) = 2k - 1$.

Now, let us show that $L \leq 2k - 1$. The proof is analogous to that for $k = 2$. We assume without loss of generality that $w(t_3) = u(t_2)$, $w'(t_3) = u'(t_3)$, $t_3 > t_2$, and note that $w'(t_3) \leq w(t_3) + k$ and that $w'_i(t_3) \geq w_i(t_3) - 1$ for $i = 2, \dots, k$. This implies immediately $w'_1(t_3) \leq w_1(t_3) + 2k - 1$.

5.2. Priority Disciplines

The Lipschitz constant is, again, $L = 2k - 1$. In order to prove that, let us denote $u_s(t) = u_1(t) + \dots + u_{k-1}(t)$ and, analogously, define w_s, u'_s, and w'_s. The outflow w_s is equal to the outflow of a single-class server with the inflow u_s, and the same is true for w'_s and u'_s. Hence, $|w_s(t) - w'_s(t)| \leq k - 1$ for each t. This implies the inequality $q(t) - q'(t) \leq k - 1$ for each t, where $q(t)$ and $q'(t)$ are the remaining service efforts for the last class of fluid u_k and u'_k, respectively. Form Lemma 4.1 we conclude that $|w_k(t) - w'_k(t)| \leq (2k - 2) + 1 = 2k - 1$ and the same inequality holds for all indices $1, \dots, k - 1$ by the induction hypothesis.

An example demonstrating that this constant is precise can be easily constructed in the same manner as for the case $k = 2$ because of the independence of $w_s(\cdot)$ from $u_k(\cdot)$.

5.3. Generalized Processor Sharing

Here, again, $L = 2k - 1$. The inequality $L \leq 2k - 1$ is derived from the Lipschitz criterion for the corresponding Skorokhod problem. Again, as in the case $k = 2$ considered above, we see that $|x_1 + \dots + x_k| \leq k$ for the endpoint x of any 1-path, and within this domain the 1-path may reach, say, the endpoint $(-1, -1, \dots, -1, 2k - 1)$.

Let us construct an example demonstrating that this bound is tight. For simplicity, we will use discontinuous inflows, but this example can easily be adapted also to continuous inflows.

At the initial time interval $[0, k)$, we set $u(t) \equiv 0$ and $u'(t) = (t/k, \dots, t/k)$. Then we set $u(k) = (2k, 2, 2, \dots, 2)$ and $u'(k) = (2k + 1, 1, 1, \dots, 1)$. Then we stop the inflows and check the outflows at $t = 3k$. We get $x_1(3k) = 2k - 2$ and $x'_1(3k) = 0$. Hence, $w_1(3k) = u_1(3k) - x_1(3k) = 2$ and $w'_1(3k) = 2k + 1$, which finishes the proof.

6. MINIMAL LIPSCHITZ CONSTANT

6.1. The Upper Bound

The problem is to find the best possible minimal Lipschitz constant for the class of work-conserving disciplines on a server with two classes of customers. Clearly, without the work-conserving property, the best Lipschitz constant is 0 (the server does not work at all).

Let a work-conserving discipline at the two servers be the same. The initial inflow $u(t_1)$ to the first server will be $(1, 1)$; t_1 is assumed large enough, so the queue is empty at time t_1. There is no inflow to the second server up to t_1. Then there is a fast linear inflow to the second server: $u'(t_1+\varepsilon) = (2+\varepsilon, 2+\varepsilon)$. The inflow to the first server is also linear at $[t_1, t_1+\varepsilon]$: $u(t_1+\varepsilon) = (1+\varepsilon, 1+\varepsilon)$.

Let us look at the situation at time $t_2 = t_1+\varepsilon+2$. Either for $i = 1$ or for $i = 2$, we have $w'_i(t_2) \le 1+\varepsilon$. Let, for definiteness, $i = 1$. Then we choose a linear inflow to the first server at $[t_1+\varepsilon, t_2]$ such that $u(t_2) = (3+\varepsilon, 1+\varepsilon)$ and note that $w_1(t_2) = 3+\varepsilon$. We also have $\|u(\cdot) - u'(\cdot)\|_{[0,t_2]} = 1$ and, hence, $L \ge w_1(t_2) - w'_1(t_2) \ge 2$.

6.2. The Equalizing Discipline

Lemma 6.1. *The constant* 2 *is attained at a server with the equalizing discipline, that is, the discipline that gives priority to the class with the maximal length of queue.*

Proof. Suppose that

$$w'_1(t) = w_1(t) + 2 + \varepsilon \tag{6.1}$$

for some $\varepsilon > 0$, and demonstrate that the difference $w'_1 - w_1$ cannot increase further. Suppose the contrary, that is, in any right-hand neighborhood of t there exists a t' such that $w'_1(t') > w_1(t') + 2 + \varepsilon$. Then we have $x'_1(t) \ge x'_2(t)$ (otherwise w'_1 does not increase in some right-hand neighborhood of t) and $x_1(t) \le x_2(t)$ (otherwise w_1 is growing at maximal possible rate in some right-hand neighborhood of t and w'_1 cannot grow faster than that). Hence

$$x'_1(t) - x_1(t) \ge x'_2(t) - x_2(t).$$

Moreover, we have

$$w'_1(t) = u'_1(t) - x'_1(t) = u_1(t) - x_1(t) + 2 + \varepsilon,$$

which gives us the inequality

$$u'_1(t) - u_1(t) \ge x'_2(t) - x_2(t) + 2 + \varepsilon = u'_2(t) - u_2(t) - (w'_2(t) - w_2(t)) + 2 + \varepsilon,$$

and, hence,

$$w'_2(t) - w_2(t) \ge \varepsilon. \tag{6.2}$$

Finally, summing up (6.1) and (6.2) we get $w'(t) - w(t) > 2$ which is a contradiction with the contraction property of the single-class server input-output operator.

REFERENCES

1. F. Baccelli, G. Cohen, Olsder G. J., and J.-P. Quadrat. *Synchronization and Linearity, An Algebra for Discrete Event Systems*. Wiley, New York, 1992.
2. M. Bramson. Instability of FIFO queueing networks. *Annals of Applied Probability*, 4:414–431, 1994.
3. M. Bramson. Stability of two classes of fluid networks and a discussion of fluid limits. *Queueing Systems*, 28:7–31, 1998.
4. J. G. Dai. Stability of open multiclass queueing networks via fluid models. In *Stochastic Networks*, volume 71 of *IMA Volumes in Mathematics and Its Applications*, pages 71–90. Springer-Verlag, New York, 1995.
5. P. Drabek, P. Krejci, and P. Takac. *Nonlinear differential equations*, volume 404 of *Research Notes in Mathematics*. Chapman & Hall/CRC, London, 1999.
6. P. Dupuis and K. Ramanan. Convex duality and the Skorokhod problem, i, ii. *Probab. Theory Relat. Fields*, 115:153–195, 197–236, 1999.
7. J. M. Harrison. Brownian models of queueing networks with heterogeneous customer population. In *Proc. of the IMA Workshop on Stoch. Differential Systems*. Springer-Verlag, 1988.

8. J. M. Harrison and V. Nguyen. Brownian models of multiclass queueing networks of queues: Current status and open problems. *Queueing Systems Theory Appl.*, 13:5–40, 1993.

9. K. Khanin, D. Khmelev, A. Rybko, and A. Vladimirov. Steady solutions for FIFO networks. *Moscow. Math. J.*, 4:407–419, 2001.

10. M. A. Krasnosel'skii and A. V. Pokrovskii. *Systems with hysteresis*. Springer-Verlag, Berlin, 1988.

11. P. Krejci. *Hysteresis, convexity and dissipation in hyperbolic equations*. Gakkotosho, Tokyo, 1996.

12. P. Krejci and A. Vladimirov. Polyhedral sweeping processes with oblique reflection in the space of regulated functions. *Set-Valued Anal.*, 11:91–110, 2003.

13. S. H. Lu and P. R. Kumar. Distributed scheduling based on due dates and buffer priorities. *IEEE Trans. Autom. Control*, 36:1406–1416, 1991.

14. M. D. P. Monteiro Marques. *Differential inclusions in nonsmooth mechanical problems—shocks and dry friction*. Birkhauser, Basel, 1993.

15. J. J. Moreau. Evolution problem associated with a moving convex set in a Hilbert space. *Journ. of Dif. Eq.*, 20:347–374, 1977.

16. A. N. Rybko and A. L. Stolyar. Ergodicity of stochastic processes describing the operation of open queueing networks. *Problems of Information Transmission*, 28:199–220, 1992.

17. T. I. Seidman. 'First in, first serve' can be unstable! *IEEE Trans. Automat. Control*, 39:2166–2171, 1994.

18. A. L. Stolyar. On the stability of multiclass queueing networks: a relaxed sufficient condition via limiting fluid processes. *Markov Process. Related Fields*, 1:491–512, 1995.

19. A. A. Vladimirov and A. F. Kleptsyn. On some hysteresis elements. *Avtomat. i Telemekh.*, (7):165–169, 1982. Russian.

This paper was recommended for publication by V.I. Venets, a member of the Editorial Board