Michael Schönlein

Stability and Robustness of Fluid Networks: A Lyapunov Perspective

Stability and Robustness of Fluid Networks: A Lyapunov Perspective

Dissertationsschrift zur Erlangung des
naturwissenschaftlichen Doktorgrades
der Julius-Maximilians-Universität Würzburg

vorgelegt von

Michael Schönlein

aus

Werneck

Würzburg 2012

Eingereicht am 23. September 2011

1. Gutachter: Prof. Dr. Fabian Wirth

2. Gutachter: Dr. Erjen Lefeber

Tag der mündlichen Prüfung: 13. April 2012

To Miriam,
my parents, Thomas, Franziska,
Daniel and Markus.

Contents

Introduction

The nature of many dynamical systems is the interaction of various components each of which is governed by dynamics admitting inputs from or producing output for other components. The possible interaction is described by a topological structure, which can be interpreted in a natural way as a graph or network. This is in essence what is called a complex system, although there is no rigorous definition [18]. In a kind of minimal definition a system is called *complex* if

(i) it exhibits complications and heterogeneity that extend virtually on all scales allowed by physical size of the system,
(ii) these features are spontaneous outcome of the interactions among the many units of the system.

In recent years the complexity of logistics systems, such as manufacturing or production networks as well as globally distributed supply chains, has increased enormously.

The network approach to the analysis of supply chains has recently attracted considerable attention, cf. [83].

Typical processes taking place in logistics networks are production, storage and shipment of different commodities. Material, information and monetary flows connect the facilities of a logistics network. The structure of a logistics network is given by the connections between the individual facilities. The performance analysis of logistics networks has been an active research area with respect to different aspects in logistics networks. There are different approaches in the literature to model such networks. In general, the incorporation of the interconnection structure of the network leads to coupled systems of equations describing the behavior of the whole network. Often do these equations represent some kind of balance equations.

Some modeling approaches are based on conservation laws by means of partial differential equations [1, 2]. Another modeling approach, inspired by physics of interconnected oscillators, has been investigated in [50, 65, 64]. There, a supply chain is described as a physical transport problem, where the flows of products are considered. But then, in logistics systems especially stochastic influences have an important effect on the systems behavior. For instance, the demand for products and the manufacturing process are subject to randomize variations. Thus, to embrace such kind of phenomenon one must count on stochastic processing networks.

Multiclass queueing networks

Multiclass queueing networks are an effective tool for modeling complex manufacturing networks [58] as they reproduce two main features of such systems. On the one hand, multiple product lines as well as highly reentrant routes through which products visit the machines can be modeled, which is very important, for instance in the wafer-fabrication. On the other hand, enable multiclass queueing networks to directly implement various production policies. Popular examples are 'first-in-first out', 'processor sharing' and 'priority' disciplines.

A multiclass queueing networks consists of J stations that serve K different classes of jobs or customers. The dynamics of the network can be described by the following stochastic processes. The process $E_k(t)$ describes the number of external arrivals in the time period $[0, t]$. The service process $S_k(t)$ reflects the number of possibly finished jobs of class k during the first t time units. For convenience we assume that each job class is served exclusively at one station. The mapping c from classes to stations determines which job class is served at which station. The set of job classes that are served at station j is denoted by $C(j)$. After being served, the jobs randomly either change their class or leave the network. The routing process $R_k^l(n)$ denotes the number of class l jobs among the first n class l jobs that become jobs of class k after service completion. As each station can serve various classes a policy determines in which order the jobs are served. The mean values of the counting processes E_k, S_k and R_l^k are denoted by α_k, μ_k and P_{lk}, respectively, and all of them are supposed to be finite. The allocation process $T_k(t)$ denotes the total amount of time that station $c(k)$ has devoted on serving class k jobs. The initial amount of class k jobs is $Q_k(0)$. So the evolution of the amount of class k jobs, denoted by $Q_k(\cdot)$, is given by the following balance equation

$$Q_k(t) = Q_k(0) + E_k(t) + \sum_{l=1}^{K} R_k^l(S_l(T_l(t))) - S_k(T_k(t)).$$

To obtain a complete description of the network dynamics further conditions on Q and T that depend on the service discipline have to be taken into account.

Despite all that possibilities to map reality in an appropriate fashion a multiclass queueing network remains an approximation of a real system. Hence, to be able to give insights for the real system, the crudest property a model should have is to be stable. The capability to catch the random changes bears the tough challenge of analyzing the stability of stochastic processes. Roughly speaking, a multiclass queueing network is stable if its long-run input rate equals its long-run output rate. In this context the nominal workload ρ_j (also called traffic intensity) of each station j is of interest. To define the nominal workload we first describe the effective arrival rate λ_k of class k jobs defined as the solution of

$$\lambda_k = \alpha_k + \sum_{l=1}^{K} P_{lk}\lambda_l,$$

where the spectral radius of the transition matrix P is assumed to be strictly less than one. The nominal workload ρ_j of station j is then given by the sum of the quotients of the effective arrival rates λ_k and the service rates μ_k over all job classes present at the station, i.e.

$$\rho_j = \sum_{k \in C(j)} \frac{\lambda_k}{\mu_k}.$$

For a long period a common belief was that a sufficient condition for stability is that the nominal workload of every station is strictly less then one. However, in 1990 Kumar and Seidman [56] presented a network with two stations processing four classes of jobs which is unstable although the nominal workload at each station is less than one. This example inspired a number of examples with different service disciplines, like first-in-first-out (FIFO) and priority, that have surprising properties. In the literature they are known as the Lu-Kumar network, the Rybko-Stolyar network or the Bramson network, see e.g. [13], [17], [34] and [71], respectively. In recent years further disciplines like maximum pressure and join-the-shortest-queue have been investigated [33], [36].

Rybko and Stolyar [71], Stolyar [81], and Dai [28] pursued the strategy of rescaling the stochastic processes that describe the dynamics of a multiclass queueing network and considered the limit of the scaling. More precisely, a fluid limit model is obtained through replacing the stochastic processes by their mean values, i.e. as $t \to \infty$ almost surely we have that

$$\frac{1}{t}E_k(t) \to \alpha_k, \qquad \frac{1}{t}S_k(t) \to \mu_k, \qquad \frac{1}{t}R_k^l(t) \to P_{lk}.$$

The corresponding balance equation for the limit takes the form

$$Q_k(t) = Q_k(0) + \alpha_k t + \sum_{l=1}^{K} P_{lk}\mu_l T_l(t) - \mu_k T_k(t).$$

Again there are additional conditions on Q and T that are specific to the service discipline. This limit is called the fluid limit model for the queueing network. Moreover, a totally deterministic model is given by the set of solutions to the latter balance equation and the additional equations related to the discipline. Of course, deterministic models are much easier to investigate. The great benefit of this approach is, that the stability of the corresponding deterministic fluid model is sufficient for the stability of a multiclass queueing network [28, 81]. In addition, there are conditions for instability of queueing networks in terms of their fluid limit model [30, 60, 67]. A discussion of the relationship between queueing networks and fluid models can be found in [17].

Due to this fact the question arises, under which conditions the deterministic fluid networks are stable. A fluid model is called stable if the fluid level process $Q(\cdot)$ with unit initial level is drained to zero in a uniform finite time τ and remains zero beyond τ. Explicite stability conditions may clearly depend on the particular discipline of the network.

The main focus of this thesis is on the Lyapunov theory for fluid networks. In addition, we always draw a comparison to the well established Lyapunov theory for dynamical systems modeled by ordinary differential equations.

Moreover, in any case it is of fundamental importance to keep in mind that the multiclass queueing network and the fluid network are simply a model for a real system. Consequently, the behavior of the real system might be quite different from the behavior of the model. There are many reasons for possible discrepancies and in the systems theory literature they are collectively referred to as model uncertainties. For instance, by observing a real manufacturing system over a certain time period one must be quite lucky or experienced to determine the mean values of e.g. the service times precisely. This means the mean value of the random variables might be uncertain. The model which is chosen on the basis of a "best guess" will be referred to as the *nominal model*. To handle the uncertainty a set of parameterized systems is considered where the parameter vectors lie in a given neighborhood around the parameters of the nominal model.

In this thesis we will also consider the robust stability issue for fluid networks. In doing so, we seize up the consideration of the stability region of a fluid network defined by Dai [29]. Precisely, we will focus on a quantitative theory for robust stability. That is, we aim to derive bounds and quantitative information about shifts in the parameters of the fluid network that lead to an unstable network. Following the catalog of Meyn [58] our robustness analysis is based on perturbations of the arrival and service times parameters. Since shifts of the routing parameters concerns the *flexibility* of the network.

Existing results

In the stability analysis of fluid networks researchers have been able to develop fairly sophisticated tools, such as Lyapunov functions, to analyze the stability of fluid networks under various disciplines. Chen states necessary and sufficient conditions for stability of general work-conserving fluid networks [19]. Stability conditions for fluid networks under FIFO and priority discipline have been derived by Chen and Zhang, cf. [23] and [24], respectively. Often the strategy to prove such conditions is to use Lyapunov functions. In this context a locally Lipschitz function $V : \mathbb{R}_+^K \to \mathbb{R}_+$ such that $V(x) = 0$ if and only if $x = 0$ is called a Lyapunov function if there exists a constant $\varepsilon > 0$ such that for each fluid level process $Q(\cdot)$ it holds that

$$\tfrac{\mathrm{d}}{\mathrm{d}t} V(Q(t)) \leq -\varepsilon$$

whenever $Q(t) \neq 0$ and the derivative of $V(Q(t))$ exists at time t. Within this framework linear Lyapunov functions of the form

$$V(x) = h^{\mathsf{T}} x, \qquad x \in \mathbb{R}_+^n,$$

where h is some positive vector in $\mathbb{R}_+^K$ are used to establish a sufficient condition for the stability of fluid network models under a priority discipline [24]. The special case $h = (1 \ldots 1)^\mathsf{T}$ is used to show that a fluid model of a re-entrant line operating under last-buffer-first-served (LBFS) discipline is stable if the usual traffic condition $\rho_j < 1$ is satisfied for all stations j [31]. The special case $h = (1 \ldots 1)^\mathsf{T}$ is also used to prove a stability condition for fluid networks under the join-the-shortest-queue discipline [33]. Ye and Chen investigated fluid networks under priority disciplines by using piecewise linear Lyapunov functions of the form

$$V(x) = \max_{1 \leq j \leq N} h_j^\mathsf{T} x$$

for some nonnegative vectors $h_1, ..., h_N$, for details see [22]. This approach yields a sharper stability condition for fluid networks under priority discipline than in [24]. Furthermore, in the verification of a stability condition for fluid networks under general work-conserving disciplines a quadratic Lyapunov function

$$V(x) = x^\mathsf{T} A\, x$$

is used, where A is a strictly copositive matrix [19]. All this results have in common that the existence of Lyapunov functions is used to establish sufficiency of the proposed criteria.

In order to obtain a commonly known converse Lyapunov theorem for fluid networks Ye and Chen followed a different, more general approach [86]. They collected characteristic properties of fluid networks and defined a generic fluid network (GFN) model Φ as a set of functions $Q : \mathbb{R}_+ \to \mathbb{R}_+^K$ that are Lipschitz continuous and satisfy a scaling and shift property. In addition, if the set of functions Φ is closed with respect to the topology of uniform convergence on compact sets, Φ is called a closed GFN model. In the work mentioned above, Ye and Chen proved that stability of a GFN model is equivalent to the property that for every function $Q(\cdot) \in \Phi$ a functional $v : \mathbb{R}_+ \to \mathbb{R}_+$ is decaying along $Q(\cdot)$. In particular, v can be chosen as

$$v(t) = \int_t^\infty \|Q(s)\| \,\mathrm{d}s.$$

This result falls short of a converse Lyapunov theorem in that no state dependent Lyapunov function is constructed. Rather in principle the whole solution set has to be known in order to even define a Lyapunov functional. The strength and basis of applicability of the classic second method of

Lyapunov, however, is that it can be checked without the knowledge of solutions, whether a given state-dependent function is indeed a Lyapunov function.

Robust stability analysis has received a lot of attraction in mathematical systems theory over the last 15 years. The interested reader is referred to the book Hinrichsen and Pritchard [53], and the references therein. The fundamental idea of the quantitative approach to robust stability is based on the introduction of a measure for the perturbations that is indicated by a single real number, which is called the stability radius. More precisely, given a nominal system and set of feasible perturbations the stability radius represents the smallest magnitude of a perturbation for which the perturbed system is no longer stable. This is a sort of worst case measure in the sense that there might be perturbations larger in magnitude than the stability radius for which the perturbed system is stable. However, the crucial point is that for any perturbation strictly smaller in magnitude the system is definitely stable.

Contribution of the thesis

The contribution of this thesis consists mainly of deriving a reasonable Lyapunov theory for GFN models. Further, we show that this result also applies to fluid networks under general work-conserving, priority, and head-of-the-line proportional processor sharing disciplines.

We define state-dependent Lyapunov functions for GFN models and consider the following Lyapunov function candidate $V : \mathbb{R}_+^K \to \mathbb{R}_+$ defined by

$$V(x) = \sup \left\{ \int_0^\infty \|Q(s)\| \, \mathrm{d}s \, : \, Q(\cdot) \in \Phi, \, Q(0) = x \right\}.$$

Using counterexamples we emphasize that the class of (closed) GFN models is too general to provide a converse Lyapunov theorem with state-dependent Lyapunov functions.

To resolve this gap we introduce the class of strict GFN models by forcing the closed GFN models to satisfy additionally

(i) a *concatenation property*: If $Q_1(\cdot), Q_2(\cdot)$ are trajectories of Φ such that $Q_1(t^*) = Q_2(t^*)$ for some $t^* \geq 0$, then $Q_1 \diamond_{t^*} Q_2(\cdot) \in \Phi$, where

$$Q_1 \diamond_{t^*} Q_2(t) := \begin{cases} Q_1(t) & t \leq t^*, \\ Q_2(t) & t \geq t^*. \end{cases}$$

(ii) a *lower semicontinuity property*: For every initial fluid level $x \in \mathbb{R}_+^K$ it holds that the set-valued map $x \rightsquigarrow \{Q(\cdot) \in \Phi : Q(0) = x\}$ is lower semicontinuous.

Within this framework we show that the stability of a strict GFN model is equivalently characterized by the existence of a state-dependent continuous Lyapunov function. It will turn that the concatenation property is essential for the existence of state-dependent Lyapunov functions, whereas the lower semicontinuity gives the additional benefit of continuity of the Lyapunov function.

Furthermore, inspired by results concerning smooth converse Lyapunov theorems for differential inclusions, cf. [26, 82], we will adapt the technique of convoluting a continuous Lyapunov function with mollifiers to obtain conditions on the strict GFN model that allow for a smooth converse Lyapunov theorem.

Moreover, we use results from differential inclusions to show that general work-conserving, priority, head-of-the-line proportional processor sharing fluid networks define strict GFN models. In addition, we show that the Lyapunov theory for fluid networks under the aforenamed disciplines allows for an alternative proof that if the fluid network associated with a multiclass queueing networks is stable and defines a strict GFN model then the multiclass queueing network is stable. Also, we will explain why the approach of strict GFN models is not immediately applicable to FIFO fluid networks. Moreover, we discuss the relation of fluid limit models to GFN models.

Apart from the Lyapunov theory we adapt the framework of robust stability analysis provided in [53] to the framework of fluid networks. Precisely, we aim to obtain bounds on the shifts of the mean values of the interarrival and service times such that the stability of the associated fluid network and, thus, of the multiclass queueing network is preserved.

Acknowledgments

I wish to extend my thanks to everybody who supported me in the development of this thesis in one way or another. Certainly, this is such a long list that I do not dare to mention them all, since I fear to forget anyone. Still a few of them can't go unmentioned.

First, I am deeply grateful to my adviser Fabian Wirth for long-lasting encouragement and many useful discussions and remarks.

I wish to thank all my colleagues in the Dynamics and Control group at the University of Würzburg for the useful discussions and the nice atmosphere during lunches and coffee breaks. In particular, I thank Jens Jordan for drawing my attention to the PhD position which made this thesis possible and special thanks go to Rudolf Sailer for countless discussions and the very nice atmosphere while sharing an office, hotel rooms, etc., over the last years.

Furthermore, I thank my colleagues Thomas Makuschewitz, Sergey Dashkovskiy and Michael Kosmykov within the research project *Stability, Robustness and Approximation of Dynamic Large-Scale Networks - Theory and Applications in Logistics Networks*. In this context I also thank the Volkswagen Stiftung for the funding of this research.

Also, I would like to thank the DAAD for the possibility of two research stays at the Universidad de Valladolid and Mustapha Ait Rami for the very nice and fruitful atmosphere and discussions during the stays.

Many thanks to Esther Hornung, who read earlier versions of this dissertation, for very useful comments to improve the readability of this thesis.

Finally, I would like to take the opportunity to thank Miriam, my family and all my good friends for a lot of things, in particular for their support and (re)-encouragement, the lasting friendship, and the very cheerful and relaxed atmosphere everytime when we are together.

1 Preliminaries

In this thesis we are concerned with the stability theory of multiclass queueing networks. In recent decades much effort has been devoted on this starting from Jackson networks, Kelly networks and multiclass queueing networks, see [17] and [21]. An extention to multiclass queueing networks where setup times are incorporated has been obtained in [35]. However, this is beyond the scope of this work. We stick to multiclass queueing networks.

In this chapter we present the basic notations and results from probability theory that form the basis for the investigation of multiclass queueing networks. In the first Section 1.1 we recall some fundamental vocabulary and introduce the concept of stochastic processes and Markov processes, respectively. Further, we provide a procedure to define a family of Markov processes on a canonical space, namely the Skorokhod space of càdlàg functions. In Section 1.3 we will introduce briefly piecewise deterministic processes and Borel right processes. The final Section 1.4 is devoted to the theoretical framework of recurrence theory for Markov processes that forms the foundation for the stability analysis of multiclass queueing networks.

1.1 Probability Theory

In this section we provide a sketch of the basic notations and some results from probability theory which will be used in this thesis.

Let Ω be a set, called the basic space, that contains points ω representing the possible realizations of some random phenomenon; events are subsets of Ω.

We denote an σ-algebra in Ω by $\mathcal{F}$ and a measurable space is a pair $(\Omega, \mathcal{F})$. A probability measure on $(\Omega, \mathcal{F})$ is denoted by $\mathbb{P}$. The triple

$(\Omega, \mathcal{F}, \mathbb{P})$ is called a probability space. Further, a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called *complete* if $A \subset B \in \mathcal{F}$ and $\mathbb{P}[B] = 0$ imply that $A \in \mathcal{F}$ and thus, $\mathbb{P}[A] = 0$. Two events A and B contained in the σ-algebra $\mathcal{F}$ are called *independent* if $\mathbb{P}[A \cap B] = \mathbb{P}[A] \cdot \mathbb{P}[B]$. The sub σ-algebras $\mathcal{G}, \mathcal{G}' \subset \mathcal{F}$ are called independent if G, G' are independent for all $G \in \mathcal{G}$ and $G' \in \mathcal{G}'$.

For the Euclidean space $(\mathbb{R}^N, \|\cdot\|)$ the σ-algebra generated by the open sets is called the *Borel σ-algebra* and is denoted by $\mathcal{B}$. A *random variable* X with values in $\mathbb{R}^N$ is a measureable function $X : (\Omega, \mathcal{F}) \to (\mathbb{R}^N, \mathcal{B})$. The expectation of X is denoted by $\mathbb{E}[X]$. For $\mathcal{G} \subset \mathcal{F}$ the *conditional expectation* of X given $\mathcal{G}$ is denoted by $\mathbb{E}[\, X \,|\, \mathcal{G}\,]$. In the following, whenever a random variable with values in $\mathbb{R}^N$ is considered, we simply call it a random variable. The random variables X_1, X_2 are called independent if $\mathcal{G}_1$ and $\mathcal{G}_2$ are independent, where $\mathcal{G}_i = X_i^{-1}(\mathcal{B})$ for $i = 1, 2$. A sequence $(X_n)_{n\in\mathbb{N}}$ of real valued random variables is called independent and identically distributed (i.i.d.) if the elements are independent and have a common distribution. We say a property holds almost surely (a.s.) on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ if it holds for every $\omega \in \Omega \setminus A$, where A is a set with probability zero, i.e. $\mathbb{P}[A] = 0$. The following statement is called the *strong law of large numbers* (SLLN).

Theorem 1.1.1 *Let $(X_n)_{n\in\mathbb{N}}$ be an i.i.d. sequence of random variables such that X_1 is integrable. Then, a.s.*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_k = \mathbb{E}[X_1]\,.$$

Proof. See [25] Theorem 5.4.2. ∎

Moreover, a sequence $(X_n)_{n\in\mathbb{N}}$ of random variables is said to *converge a.s.* to a random variable X if

$$\lim_{n\to\infty} \mathbb{P}[\, \|X_n - X\| = 0\,] = 1.$$

Besides, $(X_n)_{n\in\mathbb{N}}$ is said to *converge in distribution* to X if for every bounded continuous function $f : \mathbb{R}^N \to \mathbb{R}$ it holds that

$$\lim_{n\to\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)].$$

The subsequent statement contains conditions for the convergence of the expectation of the random variables.

Theorem 1.1.2 *Suppose that* $(X_n)_{n\in\mathbb{N}}$ *and* $(Y_n)_{n\in\mathbb{N}}$ *are sequences of random variables with values in* $\mathbb{R}$ *such that* $(X_n)_{n\in\mathbb{N}}$ *and* $(Y_n)_{n\in\mathbb{N}}$ *converge in distribution to* X *and* Y*, respectively,* $|X_n| \leq Y_n$ *for all* $n \in \mathbb{N}$*, and* $\lim_{n\to\infty} \mathbb{E}[Y_n] = \mathbb{E}[Y]$*. Then,*

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X] .$$

Proof. See [43] Appendixes Theorem 1.2. ∎

A collection of real-valued random variables $\{X_i, i \in I\}$ is said to be *uniformly integrable* if

$$\lim_{c\to\infty} \sup_{i\in I} \mathbb{E}[|X_i| \, \mathbb{1}_{\{|X_i|>c\}}] = 0.$$

Here, $\mathbb{1}_{\{A\}}$ denotes the indicator function of the set A. The following statement considers the expectations corresponding to a uniformly integrable sequence of random variables.

Proposition 1.1.3 *Let* $(X_n)_{n\in\mathbb{N}}$ *be an uniformly integrable sequence of random variables converging to* X *in distribution, then* $\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X]$*. Conversely, if the* X_n *are integrable, converge to* X *in distribution, and it holds that* $\lim_{n\to\infty} \mathbb{E}[|X_n|] = \mathbb{E}[|X|]$*, then the sequence* $(X_n)_{n\in\mathbb{N}}$ *is uniformly integrable.*

Proof. See [43] Appendixes Proposition 2.3. ∎

1.2 Markov processes

The description and analysis of multiclass queueing networks relies essentially on the theory of Markov processes. This section is devoted to lay the foundation. To this end, the material in this section is presented more detailed.

Let (E, d) denote a complete separable metric space. Recall that a metric space is called *separable* if it contains a countable dense set, i.e. there is a sequence $(x_n)_{n\in\mathbb{N}}$ such that for every $x \in X$ and for every $\varepsilon > 0$ there is a $m \in \mathbb{N}$ such that $d(x, x_m) < \varepsilon$. The σ-algebra generated by the open sets with respect to the metric d is denoted by $\mathcal{B}(E)$. Also, we denote by $\mathcal{P}(E)$ the set of probability measures defined on E.

Definition 1.2.1 *A stochastic process X with index set $I \subset \mathbb{R}$ and state space $(E, \mathcal{B}(E))$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a function defined on $I \times \Omega$ with values in E such that $X(t, \cdot) : \Omega \to E$ is a random variable for all $t \in I$.*

That is, $\{\omega : X(t, \omega) \in A\} \in \mathcal{F}$ for every $A \in \mathcal{B}(E)$. For any fixed $\omega \in \Omega$, a function $t \mapsto X(t, \omega)$ is called a *sample path* or *realization* of the stochastic process. In this thesis, we mostly consider stochastic processes where the index set is given by $I = \mathbb{R}_+ := \{t \in \mathbb{R} : t \geq 0\}$.

A collection $\{\mathcal{F}_t\} := \{\mathcal{F}_t, t \in \mathbb{R}_+\}$ of σ-algebras of sets in $\mathcal{F}$ is called a *filtration* if for $s, t \in \mathbb{R}_+$ it holds that $\mathcal{F}_t \subset \mathcal{F}_{t+s}$. A filtration is said to be *right continuous* if $\mathcal{F}_t = \bigcap_{\varepsilon > 0} \mathcal{F}_{t+\varepsilon}$. Further, a filtration is called *complete* if $(\Omega, \mathcal{F}, \mathbb{P})$ is complete and $\{A \in \mathcal{F} : \mathbb{P}[A] = 0\} \subset \mathcal{F}_0$.

Given a stochastic process X, the *natural filtration* $\mathcal{F}_t^X := \sigma(X(s), s \in [0, t])$ describes the information that is available to an observer of the stochastic process up to time t. In addition, $\mathcal{F}_t^X$ is the smallest σ-algebra in $\mathcal{F}$ with respect to which the random variables $\{X(s), s \in [0, t]\}$ are measureable. By definition it follows that

$$s \leq t \Rightarrow \mathcal{F}_s^X \subset \mathcal{F}_t^X.$$

So, $(\Omega, \mathcal{F}_t^X, \mathbb{P})$ defines for each $t \in \mathbb{R}_+$ a probability space that can be completed by adjoining to $\mathcal{F}_t^X$ all subsets of null sets.

A tuple $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ is called a filtered probability space if $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $\{\mathcal{F}_t\}$ is a filtration. A stochastic process X is said to be *adapted to a filtration* $\{\mathcal{F}_t\}$ if $X(t)$ is an $\{\mathcal{F}_t\}$-measurable random variable for each $t \in \mathbb{R}_+$. Moreover, a stochastic process is called $\{\mathcal{F}_t\}$-*progressive* if for each $t \in \mathbb{R}_+$ the restriction of X to $[0, t] \times \Omega$ is measurable with respect to $\mathcal{B}([0, t]) \times \mathcal{F}_t$.

Next, we look at a notion of equivalence for stochastic processes. To this end, for $0 \leq t_1 \leq \ldots \leq t_m$ consider the probability measure $\mathbb{P}_m$ on $\mathcal{B}(E) \times \ldots \times \mathcal{B}(E)$ defined by

$$\mathbb{P}_m[A] := \mathbb{P}[(\, X(t_1), \ldots, X(t_m)\,) \in A],$$

where $A \in \mathcal{B}(E) \times \ldots \times \mathcal{B}(E)$. Then,

$$\{\, \mathbb{P}_m \,:\, m \geq 1\,,\, 0 \leq t_1 \leq \ldots \leq t_m \,\} \tag{1.1}$$

are called the *finite-dimensional distributions* of the stochastic process X. For two stochastic processes X and Y, not necessarily defined on the same probability space, we say that Y is a *version* of X, or X and Y are *equal in distribution*, if they have the same finite-dimensional distributions.

A stochastic process X with index set $\mathbb{R}_+$ and values in $\mathbb{R}$ which is adapted to a filtration $\{\mathcal{F}_t\}$ such that $\mathbb{E}[\,|X(t)|\,] < \infty$ for all $t \geq 0$ is called

(i) an $\{\mathcal{F}_t\}$-*martingale* if for all $s, t \geq 0$ we have that

$$\mathbb{E}[\,X(t+s) \,|\, \mathcal{F}_t] = X(t),$$

(ii) an $\{\mathcal{F}_t\}$-*submartingale* if for all $s, t \geq 0$ we have that

$$\mathbb{E}[\,X(t+s) \,|\, \mathcal{F}_t] \geq X(t),$$

(iii) an $\{\mathcal{F}_t\}$-*supermartingale* if for all $s, t \geq 0$ we have that

$$\mathbb{E}[\,X(t+s) \,|\, \mathcal{F}_t] \leq X(t).$$

A *stopping time* T on a filtered probability space is a random variable taking values in $\mathbb{R}_+ \cup \{\infty\}$ such that $\{T \leq t\} := \{\omega \in \Omega : T(\omega) \leq t\} \in \mathcal{F}_t$ for all $t \in \mathbb{R}_+$. Thus, given $\mathcal{F}_t$ it is known whether T has happend by time t or not. The following result is called optional sampling theorem.

Theorem 1.2.2 *Let X be a right-continuous supermartingale of a filtration $\{\mathcal{F}_t\}$ and suppose there exists an integrable random variable Y such that $X(t) \geq \mathbb{E}[\,Y \,|\, \mathcal{F}_t\,]$ for all $t \geq 0$. Let S, T be $\{\mathcal{F}_t\}$-stopping times such that $S \leq T$ a.s. and define $X(T) = Y$ when $T = \infty$, $X(S) = Y$ when $S = \infty$. Then, the random variables $X(S), X(T)$ are integrable, and a.s. we have*

$$\mathbb{E}[\,X(T) \,|\, \mathcal{F}_S] \leq X(S).$$

Proof. See [43] Chapter 2, Theorem 2.13. ■

Now we turn the attention to Markov processes.

Definition 1.2.3 *A stochastic process X is called a* Markov process *if*

$$\mathbb{P}[X(t+s) \in A \,|\, \mathcal{F}_t^X] = \mathbb{P}[X(t+s) \in A \,|\, X(t)] \tag{1.2}$$

for all $s, t \geq 0$ and $A \in \mathcal{B}(E)$.

In the sequel, we consider the relation of Markov processes to transition functions. It will be shown that a transition function and an initial distribution uniquely determine a Markov process. A function $P : \mathbb{R}_+ \times E \times \mathcal{B}(E) \to [0, 1]$ is called a *time homogeneous transition function* if it satisfies the following conditions.

(i) For fixed t, x the function $A \mapsto P(t, x, A)$ is a probability measure on $(E, \mathcal{B}(E))$.

(ii) For fixed t, A the function $x \mapsto P(t, x, A)$ is E-measureable.

(iii) For all x, A it holds that $P(0, x, A) = \mathbb{1}_A(x)$.

(iv) For all $s, t \geq 0, x \in E$ and $A \in \mathcal{B}(E)$ it holds that

$$P(s + t, x, A) = \int_A P(s, y, A)\, P(t, x, \mathrm{d}y).$$

Furthermore, P is called a transition function for a time homogeneous Markov process X if

$$\mathbb{P}[X(t + s) \in A \,|\, \mathcal{F}_t^X] = P(s, X(t), A). \tag{1.3}$$

Given a Markov process X, a probability measure $\mathbb{P}_0 \in \mathcal{P}(E)$ is called an *initial distribution* of X if $\mathbb{P}_0[A] = \mathbb{P}[X(0) \in A]$ for all $A \in \mathcal{B}(E)$.

To examine the uniqueness issue, we consider the finite-dimensional distributions of X. Given a transition function P and an initial distribution $\mathbb{P}_0$, the finite-dimensional distributions of the Markov process are determined by

$$\begin{aligned}\mathbb{P}[X(0) \in A_0, X(t_1) \in A_1, \ldots, X(t_n) \in A_n] :=&\\ \int_{A_0} \int_{A_1} \cdots \int_{A_{n-1}} P(t_n - t_{n-1}, x_{n-1}, A_n) P(t_{n-1} - t_{n-2}, x_{n-2}, \mathrm{d}x_{n-1}) \cdots&\\ \cdot P(t_1, x_0, dx_1)\, \mathrm{d}\mathbb{P}_0(x_0).&\end{aligned} \tag{1.4}$$

The subsequent theorem states that for any transition function P and initial distribution $\mathbb{P}_0$ there is a uniquely determined Markov process (up to versions of it).

Theorem 1.2.4 *Let P be a time-homogeneous transition function and let $\mathbb{P}_0$ be a measure on E. Then there exists a Markov process X in E whose finite-dimensional distributions are uniquely determined by* (1.4).

Proof. See [43] Chapter 4, Theorem 1.1. ■

In particular, using the Dirac measure δ_x, defined for $x \in E$ by

$$\delta_x(A) := \begin{cases} 1 & \text{for } x \in A \in \mathcal{B}(E) \\ 0 & \text{for } x \notin A \in \mathcal{B}(E), \end{cases} \tag{1.5}$$

it is possible to construct a Markov process $X = \{X(t), t \geq 0\}$ with transition function P satisfying $\mathbb{P}[X(0) = x] = 1$. In this context, a *Markov family* is a collection

$$(\,\Omega,\, \mathcal{F},\, \{\mathcal{F}_t\},\, \{X(t),\, t \geq 0\},\, \{\mathbb{P}_x,\, x \in E\}\,)\,, \tag{1.6}$$

where $(\Omega, \mathcal{F})$ is a measurable space, $\{\mathcal{F}_t\}$ a filtration, and $\{X(t),\, t \in \mathbb{R}_+\}$ is a family of random variables with values in E satisfying the following conditions:

(i) For each t the random variable $X(t)$ is $\{\mathcal{F}_t\}$-measurable.

(ii) For each $x \in E$ the triple $(\Omega, \mathcal{F}, \mathbb{P}_x)$ is a probability space such that $\{X(t),\, t \in \mathbb{R}_+\}$ is a Markov process on $(\Omega, \mathcal{F}, \mathbb{P}_x)$ with transition function P satisfying $\mathbb{P}_x[X(0) = x] = 1$.

The probability measures $\{\mathbb{P}_x,\, x \in E\}$ and the transition function P are related as follows. For all $(t, x, A) \in \mathbb{R}_+ \times E \times \mathcal{B}(E)$ it holds that

$$\mathbb{P}_x[X(t) \in A] = P(t, x, A).$$

Further, in a Markov family only the probability measure $\mathbb{P}_x$ depends on the initial point $x \in E$. Hence, the Markov property can also be expressed by

$$\mathbb{P}_x[X(t+s) \in A \,|\, \mathcal{F}_s] = \mathbb{P}_z[X(t) \in A]|_{z=X(s)}\,.$$

A Markov family $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \{X(t), t \geq 0\}, \{\mathbb{P}_x, x \in E\})$ is called *strong Markov* if the Markov property (1.2) holds for every stopping time of the σ-algebra. That is, the collection (1.6) is called a strong Markov family if $X(t)$ is $\{\mathcal{F}_t\}$-progressive and

$$\mathbb{P}_x[X(t+T) \in A \,|\, \mathcal{F}_T] = P(t, X(T), A).$$

for all $x \in E$, $A \in \mathcal{B}(E)$, $\{\mathcal{F}_t\}$-stopping times T, and $t \geq 0$.

The Skorokhod Space $D(\mathbb{R}_+, E)$

In the following we show that it is possible to realize a Markov process on a *canonical space.* To this end, we consider right continuous functions having left limits. Based on the French translation "continue à droite, limites à gauche" we refer to these functions as càdlàg functions. A function $x : \mathbb{R}_+ \to E$ is called *càdlàg* if for $h > 0$ it holds that

$$x(t^+) := \lim_{h\to 0} x(t+h) = x(t) \quad \text{and} \quad x(t^-) := \lim_{h\to 0} x(t-h) \text{ exists.}$$

The set of càdlàg functions $x : \mathbb{R}_+ \to E$ is denoted by $D(\mathbb{R}_+, E)$. The set of càdlàg functions $D(\mathbb{R}_+, E)$ can be equipped with a metric d such that it becomes a separable complete metric space provided that (E, d) is complete and separable, see [43, Chapter 3, Proposition 5.6]. Here, we will not introduce the metric explicitly that induces the so called Skorokhod topology. As we are mostly concerned with convergence issues we define the Skorokhod topology via their convergence characteristics. For this reason, let $\mathcal{K}_\infty$ denote the set of continuous strictly increasing functions $\lambda : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying $\lambda(0) = 0$ and $\lim_{t\to\infty} \lambda(t) = \infty$. The following definition of the Skorokhod topology is based on Proposition 5.3 in Section 3.5 in [43].

Definition 1.2.5 *A sequence $(x_n)_{n\in\mathbb{N}}$ in $D(\mathbb{R}_+, E)$ is said to converge in the Skorokhod topology to $x \in D(\mathbb{R}_+, E)$, denoted by $x_n \to_s x$, if for each $T > 0$ there is a sequence $(\lambda_n)_{n\in\mathbb{N}}$ in $\mathcal{K}_\infty$ such that*

$$\lim_{n\to\infty} \sup_{t\in[0,T]} |\lambda_n(t) - t| = 0 \quad \text{and} \quad \lim_{n\to\infty} \sup_{t\in[0,T]} d(x_n(\lambda_n(t)), x(t)) = 0.$$

Proposition 1.2.6 *If a sequence $(x_n)_{n\in\mathbb{N}}$ in $D(\mathbb{R}_+, E)$ converges to $x \in D(\mathbb{R}_+, E)$ in the Skorokhod topology, it holds that*

$$\lim_{n\to\infty} x_n(t) = \lim_{n\to\infty} x_n(t^-) = x(t)$$

for all points of continuity of x.

Proof. See [43] Chapter 3, Proposition 5.2. ■

In the analysis of multiclass queueing networks the uniform convergence on compact subsets (u.o.c.) plays an important role. In addition, for most of the purposes it is sufficient to consider $E = \mathbb{R}^N$.

Definition 1.2.7 *A sequence $(x_n)_{n\in\mathbb{N}}$ in $D(\mathbb{R}_+, \mathbb{R}^N)$ is said to converge uniformly on compact sets (u.o.c.) to $x \in D(\mathbb{R}_+, \mathbb{R}^N)$ if for each $T > 0$ it holds that*

$$\lim_{n\to\infty} \sup_{t\in[0,T]} \|x_n(t) - x(t)\| = 0.$$

Here, we note that if a sequence $(x_n)_{n\in\mathbb{N}}$ in $D(\mathbb{R}_+, \mathbb{R}^N)$ converges u.o.c. it also converges in the Skorokhod topology. However, the converse is false in general.

In the following we provide the background for the limit theorems that are used in Chapter 2 and Chapter 5. Let $(\Omega, \mathcal{F}, \mathbb{P})$ and $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ be probability spaces and let X and X_n be stochastic processes with values in $D(\mathbb{R}_+, \mathbb{R}^N)$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and $(\Omega, \mathcal{F}_n, \mathbb{P}_n)$, respectively. We say that X_n converges to X *in distribution* if for every bounded continuous function $f : D(\mathbb{R}_+, \mathbb{R}^N) \to \mathbb{R}$ it holds that

$$\lim_{n\to\infty} \mathbb{E}_n[f(X_n)] = \mathbb{E}[f(X)],$$

where $\mathbb{E}_n$ and $\mathbb{E}$ denote the expectation with respect to $\mathbb{P}_n$ and $\mathbb{P}$, respectively.

Theorem 1.2.8 *Let X be a stochastic process and $(X_n)_{n\in\mathbb{N}}$ be a sequence of stochastic processes with values in $D(\mathbb{R}_+, \mathbb{R}^N)$. Suppose that X_n converges to X in distribution. Then, there exists a probability space $(\Omega', \mathcal{F}', \mathbb{P}')$ on which versions of $(X_n)_{n\in\mathbb{N}}$ and X, denoted by $(X'_n)_{n\in\mathbb{N}}$ and X', respectively, are defined such that X'_n converges almost surely to X' in the Skorokhod topology, i.e. $\mathbb{P}'[X_n \to_s X] = 1$.*

Proof. See [43] Chapter 3 Theorem 1.8. ■

At the end of this section we consider two fundamental processes that will be very useful for the examination of multiclass queueing networks. Let $(a_n)_{n\in\mathbb{N}}$ be an i.i.d. sequence of nonnegative real-valued random variables. For $t \in \mathbb{R}_+$ let $\lfloor t \rfloor$ denote the largest integer less than or equal to t. Further, let

$$X(t) := \sum_{n=1}^{\lfloor t \rfloor} a_n \qquad \text{for } t \geq 1,$$

and $X(t) = 0$ for all $t \in [0,1)$. Moreover, we consider the associated counting process

$$Y(t) = \sup\{s \geq 0 : \; X(s) \leq t\}.$$

The following convergence result is commonly known as the *functional strong law of large numbers* (FSLLN).

Theorem 1.2.9 *Suppose that $(a_n)_{n\in\mathbb{N}}$ is an i.i.d. sequence of nonnegative real-valued random variables with finite mean $m > 0$. Then, as $n \to \infty$ almost surely,*

$$\tfrac{1}{n} X(n\,t) \to m\,t \qquad \textit{u.o.c.} \qquad \textit{and} \qquad \tfrac{1}{n} Y(n\,t) \to \tfrac{1}{m}\,t \quad \textit{u.o.c.}$$

Proof. See [21] Theorem 5.10. ■

1.3 Piecewise Deterministic Processes

The essence of a piecewise deterministic process (PDP) is that its evolution is deterministic despite of jumps occurring at random times. In this section we provide a description for PDPs, which is taken from [38].

As an initial step let K be a countable set that indexes the jumps in the state of the process and let $n : K \to \mathbb{N}$ be a mapping that assigns the dimensions of the corresponding state space of the deterministic evolution between the jumps. Precisely, the deterministic motion of the process under consideration is described by *flows* $\varphi_v(t,\xi)$ with related state space E_v^0, where $v \in K$, $t \in \mathbb{R}$ expresses time, and $\xi \in E_v^0$. We suppose that the state space E_v^0 of the flow $\varphi_v(\cdot,\xi)$ is a subset of $\mathbb{R}^{n(v)}$. The union of all spaces for the flows is denoted by

$$E^0 = \{x = (v,\xi) \; : \; v \in K, \xi \in E_v^0\}.$$

Further, for $x = (v,\xi) \in E^0$ we define the boundary hitting time

$$t_*(x) := \begin{cases} \inf\{t > 0 \; : \; \varphi_v(t,\xi) \in \partial E_v^0\} \\ \infty \text{ if no such time exists.} \end{cases}$$

For $x = (v,\xi) \in E^0$ let $t_\infty(x)$ denote the *life span* of the flow $\varphi_v(\cdot,\xi)$. To avoid a finite life span we assume that $t_\infty(x) = \infty$ if $t_*(x) = \infty$. To outline the state space we use the boundaries $\partial E_v^0 := \overline{E}_v^0 \setminus E_v^0$ of the spaces on which the flows exist. We denote the subset of boundary points that are attained by the flows by

$$\begin{aligned}\partial^{\pm} E_v^0 &= \{z \in \partial E_v^0 \,:\, z = \varphi_v(\pm t, \xi) \text{ for some } \xi \in E_v^0, t > 0\},\\ \partial_1 E_v^0 &= \partial^{-} E_v^0 \setminus \partial^{+} E_v^0,\\ \Gamma^* &= \bigcup_{v \in K} \partial^{+} E_v^0.\end{aligned}$$

The state space E is defined by

$$E := \bigcup_{v \in K} E_v^0 \cup \partial_1 E_v^0.$$

Let A_v be a Borel subset of $E_v^0 \cup \partial_1 E_v^0$, which exists since $E_v^0 \cup \partial_1 E_v^0$ is a subset of the Euclidean space $\mathbb{R}^{n(v)}$, then the sets $A = \bigcup_v A_v$ generate a σ-algebra on E which will be denoted by $\mathcal{E}$. It is possible to endow E with a metric d such that the Borel sets generated by d coincide with $\mathcal{E}$ and $(E, \mathcal{E})$ is a Borel subset of complete separable metric space. The jump mechanism is described by the *jump rate* λ and the *transition measure* Q. The jump rate is defined by a measurable mapping $\lambda : E \to \mathbb{R}_+$ with the following properties: For each $x = (v, \xi) \in E$ there exists an $\varepsilon(x) > 0$ such that $t \mapsto \lambda(v, \varphi_v(t, \xi))$ is integrable on $[0, \varepsilon(x))$. Moreover, let $\mathcal{P}(E)$ denote the set of all probability measures on $(E, \mathcal{E})$. A function $Q : E \cup \Gamma^* \to \mathcal{P}(E)$ is called a *transition measure* if it satisfies

$$\begin{aligned} Q(\{x\}, x) = 0 &\quad \text{for each } x \in E,\\ x \mapsto Q(A, x) &\quad \text{is measurable for all } A \in \mathcal{E}.\end{aligned}$$

A collection of the latter yields the following definition of a piecewise deterministic process.

Definition 1.3.1 *A process X defined on E is called a piecewise deterministic process (PDP) if the following conditions are satisfied.*

(*i*) *For each $v \in K$, $\varphi_v(t, \xi)$ is a continuous flow. The life span $t_\infty(\xi) = +\infty$ whenever $t_*(x) = \infty$.*

(*ii*) *The* jump rate *$\lambda : E \to \mathbb{R}_+$ is a measurable function such that $t \mapsto \lambda(v, \varphi_v(t, \xi))$ is integrable on $[0, \varepsilon(x))$ for some $\varepsilon(x) > 0$, for each $x \in E$.*

(*iii*) *The* transition measure $Q : E \cup \Gamma \to \mathcal{P}(E)$ *is a measurable function such that* $Q(\{x\}, x) = 0$ *for each* $x \in E$.

(*iv*) $\mathbb{E}_x[N_t] < \infty$ *for each* $(t, x) \in \mathbb{R}_+ \times E$.

The content of the following result concerns the relation of PDPs and Markov processes.

Theorem 1.3.2 *Let* $X = \{X(t),\, t \geq 0\}$ *be a PDP, then* X *is a homogeneous strong Markov process.*

Proof. See [38] Theorem 25.5. ■

Another common class of stochastic processes are Borel right processes. The state space is a *Lusin* space, that is, a topological space which is homeomorphic to a Borel subset of a compact metric space.

Definition 1.3.3 *A stochastic process* $X = \{X(t),\, t \geq 0\}$ *is called a Borel right process if*

(*i*) *the state space* E *is a Lusin space,*

(*ii*) *the transition function* $P(t, \cdot, \cdot)$ *maps bounded measurable functions to bounded measureable functions,*

(*iii*) *the sample paths* $X(\cdot, \omega)$ *are right continuous,*

(*iv*) *the process is strong Markov.*

Here, we note that a locally compact Hausdorff space with countable base is a Lusin space, cf. [78, p. 370]. Finally, the last statement of this section relates PDPs to Borel right processes.

Theorem 1.3.4 *Let* $X = \{X(t),\, t \geq 0\}$ *be a PDP then* X *is a Borel right process.*

Proof. See [38] Theorem 27.8. ■

1.4 Harris Recurrence

In this section we present the foundation for the stability analysis of multi-class queueing networks. To this end, we consider a strong Markov process X with values in a separable and locally compact metric space E. The Borel σ-algebra induced by the metric is denoted by $\mathcal{B}(E)$. Furthermore, let P be the transition function associated with X such that for $A \in \mathcal{B}(E)$ we have that $P(t, x, A) = \mathbb{P}_x[X(t) \in A]$, where x denotes the initial value.

A measure ν on $(E, \mathcal{B}(E))$ is called *invariant* for X if it is σ-finite and it holds that

$$\nu(A) = \int_E P(t, x, A)\nu(\mathrm{d}x)$$

for all $A \in \mathcal{B}(E)$ and $t \geq 0$. For a set $A \in \mathcal{B}(E)$ let

$$\tau_A := \inf\{t \geq 0 \,:\, X(t) \in A\}$$

denote the *first entrance time*, and for $\delta > 0$ the *first entrance time past* δ is given by $\tau_A(\delta) := \inf\{t \geq \delta \,:\, X(t) \in A\}$. By the Début Theorem, cf. [78, Theorem A 5.1], the first entrance time defines a stopping time. Furthermore, for $A \in \mathcal{B}(E)$ we consider the *occupation time* η_A, describing the number of visits by X to A, given by

$$\eta_A := \int_0^\infty \mathbb{1}_{\{X(t) \in A\}} \,\mathrm{d}t.$$

A Markov process X is called *Harris recurrent* if there exists a nontrivial σ-finite measure ν such that whenever $\nu(A) > 0$ and $A \in \mathcal{B}(E)$ it holds that

$$\mathbb{P}_x\left[\eta_A = \infty\right] = 1$$

for all $x \in E$. Based on this characterization an interpretation of Harris recurrence is that sets with positive measure are visited infinitely often. Getoor has shown that for Harris recurrent processes there is an unique invariant measure (up to a constant multiple), see [47]. If the unique invariant measure can be normalized to a probability measure, the Markov process X is called *positive Harris recurrent*.

Next, we state a characterization of positive Harris recurrence that is easier to apply. Suppose that a is probability measure on $(0, \infty)$ and consider the Markov process X_a with transition function

$$T_a(x, A) = \int_0^\infty P(t, x, A)a(\mathrm{d}t),$$

where $x \in E$ and $A \in \mathcal{B}(E)$. A probabilistic interpretation of X_a is the following: X_a is the Markov process X sampled at time points drawn successively according to the distribution a, or more precisely, at time points of an independent renewal process with increment distribution a, cf. [59]. Let μ be some nontrivial measure on $(E, \mathcal{B}(E))$. A nonempty set $A \in \mathcal{B}(E)$ is called *petite* if there is a nontrivial measure μ on $(E, \mathcal{B}(E))$ and a probability measure a on $(0, \infty)$ such that the transition function $T_a(x, B)$ of the sample process satisfies

$$T_a(x, B) \geq \mu(B)$$

for all $x \in A$ and for all $B \in \mathcal{B}(E)$. A petite set A has the property that each set $B \in \mathcal{B}(E)$ is equally accessible from all points $x \in A$ with respect to the measure μ. The following result contains conditions that will be very useful for the Harris recurrence analysis of multiclass queueing networks.

Theorem 1.4.1 *Let X be a Markov process.*

(1) *X is Harris recurrent if and only if there exists a closed petite set A such that for all $x \in E$ it holds that*

$$\mathbb{P}_x[\tau_A < \infty] = 1.$$

(2) *If x is Harris recurrent. Then, X is positive Harris recurrent if and only if there is a closed petite set A such that for some $\delta > 0$,*

$$\sup_{x \in A} \mathbb{E}_x[\tau_A(\delta)] < \infty. \tag{1.7}$$

Proof. See [17] Section 4.5. ■

1.5 Notes and References

A detailed presentation of the first Section 1.1 can be found in many places. For instance [10, 25] are good textbooks, just to mention two.

The main reference for the outline of stochastic processes and Markov processes in Section 1.2 is the book of Ethier and Kurtz [43]. This reference is a comprehensive collection of material related to Markov processes. But, the version of the optional sampling theorem 1.2.2 stated here is due to [38]. This formulation of the optional sampling theorem for supermartingales is appropriate for our purpose, see Section 5.3. In addition, a full description of the Skorokhod space and its topology can be found in the books [11, 43]. Another acronym for càdlàg functions is RCLL which is short for right continuous left limits. The definition of uniform convergence on compact sets used in this section is derived from [31]. Moreover, the formulation of Skorokhod's representation theorem as well as the short display of the fundamental processes, including the functional law of large numbers, can be found in the text book [21]. Also, we note that the notation $\mathcal{K}_\infty$ for the set of strictly increasing unbounded continuous functions differs from [43]. In probability literature the set $\mathcal{K}_\infty$ is usually denoted by Λ. We use $\mathcal{K}_\infty$ for two reasons. On the hand, the letter Λ will be used in Chapter 3 with a different meaning. On the other hand, the notion $\mathcal{K}_\infty$ is standard in dynamical systems literature and we will use the similar notion $\mathcal{K}$ to denote the set of functions with the same properties despite of the fact that they are not necessarily unbounded.

The standard reference for piecewise deterministic processes is the text book [38]. A comprehensive discussion of Borel right processes can be found in Section 20 of [78]. The definition of a Borel right process here differs from the one in [78]. It is taken from [17] and is based on Theorem 9.4 (i) in [47].

The short outline of Harris recurrence in Section 1.4 follows a collection given by Bramson in [17]. Harris recurrence has been introduced by Harris for discrete time Markov chains [49]. Later, Harris recurrence for continuous time Markov processes was first considered by Azéma, Kaplan-Dulfo and Revuz [6]. A full description of (positive) Harris recurrence for Markov chains can be found in [59, 61]. The papers [62, 63] extended the results to Markov processes and contain so called Foster-Lyapunov criteria for positive Harris recurrence. We present a Foster-Lyapunov theorem in Setion 5.3. Another useful sufficient condition for a Markov process being positive Harris recurrent was given by Dai in [28]. In the following chapter we present a special type of stochastic processing networks, namely we investigate multiclass queueing networks. In particular, we turn our attention to a technique, called fluid approximation, to obtain a deterministic criterion to conclude that a multiclass queueing network is positive Harris recurrent.

2 Multiclass Queueing Networks

In this chapter we lay the conceptual foundation for the remainder of this thesis. The main topic in this thesis is the stability analysis of fluid networks. In this chapter we recapitulate that, based on the remarkable results developed by Rybko, Stolyar and Dai, see [71, 81] and [28], respectively, the stability analysis of fluid networks is a powerful tool for the investigation of positive Harris recurrence of multiclass queueing networks.

To this end, we provide a description of multiclass queueing networks, the underlying Markov process, and the dynamic equations of multiclass queueing networks which stems from [17] and [28]. Moreover, we discuss an approach to reduce the stability problem of a multiclass queueing networks to the examination of the stability of a purely deterministic network, called the associated fluid network. In addition, we collect known properties of the associated fluid network that will play an important role throughout this thesis.

Section 2.1 starts with the basic notation of multiclass queueing networks and the standing assumptions are stated. Afterwards, the state space and several queueing disciplines are described, e.g. first-in-first-out (FIFO), priority, and processor sharing disciplines. The state space is a separable and locally compact metric space. That is, the Harris recurrence theory from Section 1.4 applies.

In Section 2.2 we present the underlying Markov process that is defined by a multiclass queueing network. Furthermore, it is outlined that the Markov process defines a piecewisewise deterministic process. In particular, the process is a Borel right process and satisfies the strong Markov property.

In Section 2.3 we turn our attention to a description of the evolution of multiclass queueing networks which is based on a set of equations, commonly known as the basic queueing network equations. For instance, the evolution of the queue length can be modelled by a balance equation that depends on counting processes and a process which describes the particular service discipline. It turns out that the basic queueing network equations embrace four equations which are complemented by further conditions specific to the discipline.

The final Section 2.4 is dedicated to the characteristics of the processes that appear in the basic queueing network equations. In particular, we regard the behavior of the primitive cumulatives under taking limits of a sequence of rescaled versions of it. Moreover, the relation of the obtained limits and the basic queueing network equations is considered. The great benefit of this technique is that it enables us to conclude stability of the stochastic multiclass queueing network by investigating an associated fluid network, which is a continuous and deterministic model.

2.1 Model Description

Typically a multiclass queueing network consists of objects, for instance jobs or customers, that are waiting for service in buffers in front of diverse stations. After service completion, a served job either moves to another buffer at some further station or leaves the network.

A multiclass queueing network consists of J service stations and K classes of customers. The *interarrival times* for customers of class $k \in \{1, ...K\}$ are given by positive random variables $a_k(n),$ with $n = 1, 2, 3, \dots,$ and the *service times* of class k customers are given by positive random variables $s_k(n)$, with $n = 1, 2, 3\dots$. Each customer class is exclusively served at a certain station. The many-to-one mapping $c : \{1, ..., K\} \to \{1, ..., J\}$ determines which customer class is served at which station. The corresponding $J \times K$ matrix C, called the *constituency matrix*, is defined by

$$C_{jk} := \begin{cases} 1 & \text{if } c(k) = j, \\ 0 & \text{else.} \end{cases}$$

For station $j \in \{1, ..., J\}$, the set

$$C(j) := \{k \in \{1, ..., K\} : c(k) = j\}$$

is the collection of all customer classes that are served at the station j.

After a class k customer received service at the station $c(k)$ its routing is given by a K dimensional *Bernoulli random variable* ϕ^k. To be precise, each component of $\phi^k(n)$ is either 0 or 1, but the entry 1 appears at most once. Let e_k denote the kth standard basis vector for $\mathbb{R}^K$. Then, the nth served class k customer at station $c(k)$ becomes a class l customer after service completion if $\phi^k(n) = e_l$, and the customer leaves the network if $\phi^k(n) = 0$.

Remark 2.1.1 *For some customer class k the interarrival time may be $a_k(n) = \infty$ for all n. Then, the exogenous arrival process is null. The corresponding notation is the following*

$$\mathcal{E} := \{k \in \{1, ..., K\} : a_k(n) < \infty,\, n \geq 1\}.$$

Further, the buffer at each station is assumed to have infinite capacity.

Throughout this thesis we pose the following general assumptions on the interarrival and the service times. To this end, a distribution ν of the interarrival times is called *unbounded* if for each class $k \in \mathcal{E}$ and for all $t \geq 0$ it holds that

$$\mathbb{P}_\nu[a_k(1) \geq t] := \int_t^\infty a_k(1)\nu(\mathrm{d}s) > 0.$$

Unboundedness expresses that arbitrarily large interarrival times appear with positive probability. Moreover, the distribution of the interarrival times of customer class $k \in \mathcal{E}$ is said to *spread out* if there exists some $l_k \in \{1, 2, 3, ...\}$ and some nonnegative function $q_k : \mathbb{R}_+ \to \mathbb{R}_+$ with $\int_0^\infty q_k(s)\,\mathrm{d}s > 0$ such that for all $0 \leq a < b$,

$$\mathbb{P}_\nu\left[\sum_{i=1}^{l_k} a_k(i) \in [a, b]\right] \;\geq\; \int_a^b q_k(s)\,\mathrm{d}s.$$

Before summarizing the standing assumptions, we recall that the spectral radius of a matrix $M \in \mathbb{R}^{K\times K}$ is defined by $\varrho(M) := \sup\{|\lambda| \,|\, \exists\, x \in \mathbb{R}^K : Ax = \lambda x\}$.

Assumptions 2.1.2

(1) *The sequences $a_1, ..., a_K, s_1, ..., s_K$ and $\phi^1, ..., \phi^K$ are identically and independently distributed and mutually independent.*

(2) *The first moments satisfy*

$$\begin{aligned} \alpha_k &:= \mathbb{E}[\, a_k(1)\,]^{-1} < \infty, \ \textit{for} \ k \in \mathcal{E}, \\ \mu_k &:= \mathbb{E}[\, s_k(1)\,]^{-1} < \infty, \ \textit{for} \ k \in \{1, ..., K\}, \\ P_k &:= \mathbb{E}[\, \phi^k(1)\,] \geq 0, \qquad \textit{for} \ k \in \{1, ..., K\}, \end{aligned}$$

and the spectral radius of the matrix $P = (P_1 \dots P_K)$ *is strictly less than one.*

(3) *The distributions of the interarrival times are unbounded and spread out.*

Throughout the thesis we refer to the triple (a, s, ϕ) as the *primitive increments* of the multiclass queueing network. The lth component of the parameter P_k of the Bernoulli routing ϕ^k reflects the probability that a class k customer becomes a class l customer. Hence,

$$1 - \sum_{l=1}^{K} P_{kl}$$

represents the probability that a class k customer is leaving the network after service completion. Since the transition matrix $P = (P_{kl})$ is assumed to have spectral radius strictly less than one the Neumann series converges, i.e.

$$(I + P + P^2 + \dots)^{\mathsf{T}} = (I - P^{\mathsf{T}})^{-1}$$

exists. As a consequence, almost surely every customer visits only finitely many stations before leaving the network. In open multiclass queueing networks the *effective arrival rate* λ_k of class k customers is given by

$$\lambda_k = \alpha_k + \sum_{l=1}^{K} P_{lk}\, \lambda_l.$$

Since the spectral radius of P is strictly less than one, the vector form of the effective arrival rate is

$$\lambda = (I - P^{\mathsf{T}})^{-1} \alpha.$$

Further, using $M = \operatorname{diag}(\mu_1, ..., \mu_K)$, the *nominal workload* of the stations per time unit is represented by the J-dimensional vector

$$\rho = C\, M^{-1} \lambda. \tag{2.1}$$

State Space

In this thesis we restrict ourselves to *head-of-the-line (HL)* queueing networks. The term HL indicates that within each class the customers are served in *First-In-First-Out* (FIFO) order. The evolution of the HL multiclass queueing network will be described by a stochastic process $\mathsf{X} = \{\mathsf{X}(t),\, t \geq 0\}$ and its corresponding state space is denoted by $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, where

$$\mathcal{X} = \{x \,:\, x = (y, z) = (((k, w), u, v), z) \in (\mathbb{Z} \times \mathbb{R})^{\infty} \times \mathbb{R}^{|\mathcal{E}|} \times \mathbb{R}^{K} \times \mathbb{R}^{K}\} \tag{2.2}$$

and $\mathcal{X}$ is endowed a metric d. Here $(\mathbb{Z} \times \mathbb{R})^{\infty}$ denotes the set of finitely terminated sequences taking values in $\mathbb{Z} \times \mathbb{R}$, and $\mathcal{B}(\mathcal{X})$ denotes the standard Borel σ-algebra of $\mathcal{X}$ induced by the metric d. Before introducing the metric, we consider the state space $\mathcal{X}$ itsself. The meaning of each of its components are the following.

$(k, w) \in (\mathbb{Z} \times \mathbb{R})^{\infty}$: The global order of the customers in the network is given by the pair of sequences $(k, w) = ((k_1, w_1), (k_2, w_2), ..., (k_l, w_l))$. The first entry $k_i \in \{1, ..., K\}$ denotes the current class of the ith customer and the second entry $w_i \geq 0$ reflects the elapsed time since the customer i entered the class k_i. The order of elements (k_i, w_i) in the sequence (k, w) is defined by the property that it is descending in w_i. That is, within each class the customer with largest second component represents the oldest one and so appears first in the sequence (k, w). In the case two or more customers have an identical second coordinate, the ordering is ascending with respect to the customer class. The number of customers of each class is denoted by $q = (q_1, \ldots, q_K)$ and $\|q\|_1 = \sum_{k=1}^{K} q_k$ is the total number of customers in the state.

$u \in \mathbb{R}^{|\mathcal{E}|}$: This component denotes the *residual interarrival time*. That is, the coordinate $u_k > 0$ denotes the remaining time before the next arrival of a class $k \in \mathcal{E}$ customer from outside the network.

$v \in \mathbb{R}^{K}$: This component represents the *residual service time*, i.e. the coordinate v_k denotes the remaining service time for the oldest class k customer, where $v_k \geq 0$ and $v_k = 0$ only if $q_k = 0$.

$z \in [0, 1]^{K}$: The component z_k denotes the proportion of the service effort of station $c(k)$ that the oldest class k customer receives, while other class k customers do not receive any service. This represents the HL property. For each station j we have that if $\sum_{k\in C(j)} q_k > 0$, it holds that $\sum_{k\in C(j)} z_k = 1$, where $z_k = 0$ if $q_k = 0$. If station j is empty, i.e. $\sum_{k\in C(j)} q_k = 0$, then $\sum_{k\in C(j)} z_k = 0$.

Now we introduce a metric d on $\mathcal{X}$ that sums the difference of each component. That is, for $x, x' \in \mathcal{X}$ we define

$$d(x,x') := \sum_{i=1}^{\infty} \min\{|k_i - k_i'| + |w_i - w_i'| + |z_i - z_i'|, 1\} + \sum_{k\in\mathcal{E}} |u_k - u_k'| + \sum_{k=1}^{K} |v_k - v_k'|.$$

The metric space $(\mathcal{X}, d)$ has the following properties.

Proposition 2.1.3 *The metric space $(\mathcal{X}, d)$ is separable and locally compact. In particular, $\mathcal{X}$ is a Lusin space.*

Proof. See [17] Section 4.1. ■

Remark 2.1.4 *The metric space $(\mathcal{X}, d)$ is not complete. A complete metric can be obtained by adding an appropriate term in each of the second and third sum in d, cf. [17, Section 4.1].*

So, the state space $(\mathcal{X}, \mathcal{B})$ is measurable. For most of the issues in this thesis we do not require the full metric d. With a slight abuse of notation we call

$$|x| := \|q\| + \|u\| + \|v\|$$

a norm on $\mathcal{X}$. Further, let $\mathcal{X}$ be equipped with the natural induced topology then $\{x \in \mathcal{X} : |x| \le \kappa\}$ is a compact subset of $\mathcal{X}$ for every $\kappa > 0$.

Various Queueing Disciplines

In the following we focus on some popular HL disciplines and consider the resulting state space $\mathcal{X}$. We will see that the state space of specific HL queueing networks can be simplified by removing redundant information. Furthermore, a restriction to exponential interarrival and service times allows to drop the coordinates u and v from the state [17].

FIFO disciplines

In first-in-first-out (FIFO) queueing networks the customers are served according to the time spent in the queue, where the oldest customers are served first. Let $N_j(t) = \sum_{k \in C(j)} Q_k(t)$ denote the queue length at time t present at station j. The customers at station j are ordered according to their age, i.e. for $N_j(t) > 0$ we consider the list

$$k_j(t) := (k_{j,1}, k_{j,2}, ..., k_{j,N_j(t)}),$$

where $k_{j,i}$ denotes the number of the ith customer at station j. If $N_j(t) = 0$ the list is set to be the empty list. Consequently, the customers age can be removed from the state space description. Moreover, since the queueing network is HL, the coordinate z in (2.2) can also be removed and the evolution of the Markov process can be described by

$$\mathsf{X}(t) = (\,(k_1(t), ..., k_J(t)), U(t), (V_{k_{j,1}(t)}(t), j = 1, \ldots, J)\,).$$

Thus, the state space satisfies

$$\mathcal{X} \subset (\mathbb{Z}_K^\infty)^J \times \mathbb{R}_+^{|\mathcal{E}|+J},$$

where $\mathbb{Z}_K^\infty$ denotes the set of terminating sequences taking values in $\mathbb{Z}_K := \{1, 2, ..., K\}$.

Priority disciplines

Under priority disciplines each station ranks the customer classes and serves them accordingly. That is, for each station j there is a one-to-one mapping

$$\pi_j : C(j) \to \{1, \ldots, |C(j)|\}$$

and the class k' with $\pi_j(k') = 1$ is said to be of highest priority. If the station has completed a customer, it picks the oldest customer of the class with the highest priority from the queue. So, within each class the customers are served in FIFO order. If the queue is empty, the station idles. The priority service disciplines can be distinguished in two subdisciplines.

Preemptive priority disciplines

Here, if during the time a customer receives service a customer of a class with higher priority enters the queue, the service is stopped immediately, and the customer with the higher priority is served. The service does not continue until there is no further customer with higher priority present in the queue. Following the same arguments as in the FIFO case the Markov process can be described by

$$\mathsf{X}(t) = (\,(Q_1(t), \ldots, Q_K(t)), U(t), V(t)\,)\,,$$

and the state space can be taken as

$$\mathcal{X} \subset \mathbb{Z}_+^K \times \mathbb{R}_+^{|\mathcal{E}|+K}.$$

Nonpreemptive priority disciplines

In this case, even if there arrives a customer of a class with higher priority, the station has to finish the service of the current customer. Hence, at each station j there is at most one customer being served at time t. So, let $k_j(t)$ denote the customer class that is served at time t, then the development of the Markov process can be described by

$$\mathsf{X}(t) = \big(\,(Q_1(t), \ldots, Q_K(t)), U(t), k_j(t), V_{k_{j,1}(t)}(t), j = 1, ..., J\big)\,.$$

If there are no customers present at station j at time t, then $k_j(t)$ is set to zero and $V_{k_j(t)}(t) = 0$. Consequently, the Markov process is completely described by elements of the state space

$$\mathcal{X} \subset \mathbb{Z}_+^K \times \mathbb{R}_+^{|\mathcal{E}|+K} \times \mathbb{Z}_K^J.$$

HL processor sharing disciplines

The essence of HL processor sharing (HLPS) disciplines is that the oldest customers of each class receive service simultaneously at the station defined by the constituency matrix. Within each class the customers are served in FIFO order. The state is then described by

$$\mathsf{X}(t) = (Q(t), U(t), V(t))\,,$$

and the state space is

$$\mathcal{X} \subset \mathbb{Z}_+^K \times \mathbb{R}_+^{|\mathcal{E}|+K}\,.$$

2.2 The Underlying Markov Process

In this section we outline the stochastic process which describes the evolution of a multiclass queueing network. First, we describe the evolution of the process $\mathsf{X} = \{\mathsf{X}(t) = (Y(t), Z(t)) \in \mathcal{X}; t \geq 0\}$ in between arrivals and departures of customers, where the process develops according to the piecewise constant service rates $Z(t)$. Note that, based on the service discipline, the relation of $Z(t)$ and $Y(t)$ can be described by a measurable function.

Second, to describe the progress of the Markov process, when an arrival or a departure took place somewhere in the network, we consider the primitive increments $\{a_k(n), s_k(n), \phi^k(n) : k \in \{1, ..., K\}, n \geq 1\}$ to construct the evolution of X inductively.

For time t in between arrivals and departures, the decrease rate of residual service times $V(t) = v$ is given by $Z(t)$. There are two possibilities that are allowed to happen:

(i) *Service completion*: Here $V_k(t^-) = 0$ for some $k \in \{1, ..., K\}$. The transition of the oldest class k customer is then given by ϕ^k. Hence, we set $V_k(t) = s_k(i)$ or $V_k(t) = 0$ if $q_k(t^-) > 0$ or $q_k(t^-) = 0$, respectively. Not before a customer either leaves the network or transits to another class, its age $W_i(t) = w_i$ increases with rate 1. Also, the components of the residual arrival rates $U(t) = u$ decrease at rate 1 until hitting 0.

(ii) *New arrival*: Here $U_k(t^-) = 0$ for some $k \in \mathcal{E}$. In this case, a pair $(k, 0)$ is added to the state $Y(t)$ and the residual interarrival time is set to $U_k(t) = a_k(i)$, where i represents the number of the next unused class k customer up to time t. If the customer of class k arrives at an empty queue, i.e. $q_k(t^-) = 0$, it holds that $V_k(t^-) = 0$. Then, we set $V_k(t) = s_k(i')$, where i' is the index of the first unused service time at time t.

Moreover, the underlying stochastic process X has the following properties.

Theorem 2.2.1 ([28]) *The stochastic process X is piecewise deterministic with state space $\mathcal{X}$. In particular, X is a Borel right process and satisfies the strong Markov property.*

We will not prove this statement. The interested reader is referred to [17] Section 4.1. Based on the last statement the process is equipped with the basic ingredients defining a strong Markov process. So, the process X is Borel right and is defined on a measurable space $(\Omega, \mathcal{F})$ with values in the measurable space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. Furthermore, the process is adapted to a filtration $\{\mathcal{F}_t\}$ and $\{\mathbb{P}_x, x \in \mathcal{X}\}$ are probability measures on $(\Omega, \mathcal{F})$ such that for every $x \in \mathcal{X}$ we have $\mathbb{P}_x[\mathsf{X}(0) = x] = 1$. As outlined in Section 1.2 the collection

$$(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \{X(t), t \geq 0\}, \{\mathbb{P}_x, x \in \mathcal{X}\})$$

defines a strong Markov family with transition function defined by

$$P(t, x, A) = \mathbb{P}_x[X(t) \in A]$$

for all $t \geq 0, x \in \mathcal{X}$, and $A \in \mathcal{B}(\mathcal{X})$.

2.3 Dynamic Equations of Queueing Networks

In this section we introduce dynamic equations that describe the evolution of the performance processes of multiclass queueing networks. Based on the primitive increments (a, s, ϕ) we define the *primitive cumulatives* (E, S, R) of the queueing network. By convention, we assume that $E(0) = S(0) = R(0) = 0$. The process $E(t)$ is called *external arrival process*. The arrival time of the nth customer of class k is given by $a_k(1) + ... + a_k(n)$. The process $E_k(t)$ counts the arrivals of class k customers from outside up to time t. That is, $E_k(t)$ is defined by

$$E_k(t) := \max\{n \in \mathbb{Z}_+ : a_k(1) + ... + a_k(n) \leq t\}.$$

The process $S(t)$ is called the *cumulative service process*. The component $S_k(t)$ counts the service completions of class k customers in the time period $[0, t]$ in the case the station allocated its entire capacity to that customer class, i.e.

$$S_k(t) = \max\{n \in \mathbb{Z}_+ : s_k(1) + ... + s_k(n) \leq t\}.$$

The process $R(n)$ is called the *routing process*. For each $n \in \mathbb{Z}_+$ and for each customer class $k \in \{1, ..., K\}$ the routing process is defined by

$$R^k(n) := \sum_{i=1}^{n} \phi^k(i).$$

Certainly, the processes E, S and R are càdlàg. Before we are able to state the dynamic equations that describe the evolution of the queueing network, we have to introduce the *allocation process*, denoted by $T := \{T(t), t \geq 0\}$. This process contains the information how much time a station has devoted to serve the customer classes present at this station. To be precise, $T_k(t)$ denotes the cumulative amount of time that station $c(k)$ has spent on serving class k customers in the time period $[0, t]$. The allocation process is determined by the service discipline. From the definition it follows that $T(\cdot)$ is nondecreasing. At the end of this section we will specify T for the particular service disciplines that have been introduced in Section 2.1.

Given the allocation process T, the number of service completions of class k customers up to time t is given by $S_k(T_k(t))$. Further, the number of class l customers that have routed from to class l to class k in the time period from $[0, t]$ is given by $R_k^l(S_l(T_l(t)))$. Thus, the queue length of class k customers at time t can be described by the following balance equation

$$Q_k(t) = Q_k(0) + E_k(t) + \sum_{l=1}^{K} R_k^l(S_l(T_l(t))) - S_k(T_k(t)),$$

where $Q_k(0)$ denotes the number of customers that are already in the queue at time zero. Denoting the total arrivals of class k customers up to time t by

$$A_k(t) := E_k(t) + \sum_{l=1}^{K} R_k^l(S_l(T_l(t)))$$

and the departures in $[0, t]$ by

$$D_k(t) := S_k(T_k(t)),$$

the above balance equation can be written in the following simple form

$$Q_k(t) = Q_k(0) + A_k(t) - D_k(t).$$

Moreover, there are additional processes that are used to describe the evolution of the multiclass queueing network. The process $W := \{W(t), t \geq 0\}$ is called the *immediate workload process*. The component $W_j(t)$ reflects the residual time that is needed to serve all customers that are currently waiting to be served at station j. Further, for $n = \left(n_1 \,...\, n_K\right)^{\mathsf{T}}$ let $\Gamma(n) = \left(\Gamma_1(n_1) \,...\, \Gamma_K(n_K)\right)^{\mathsf{T}}$ denote the cumulative service defined by

$$\Gamma_k(n_k) := \sum_{i=1}^{n_k} s_k(i).$$

Thus, the immediate workload can be characterized by

$$W(t) = C\,\Gamma\left(Q(0) + A(t)\right) - C\,T(t),$$

where C denotes the constituency matrix. Moreover, the J dimensional process $I := \{I(t), t \geq 0\}$, called the *idle time process*, denotes the total time that the stations were not working in the time period $[0, t]$. The idle process can be described by the following condition

$$I(t) = et - C\,T(t),$$

where $e = (1, ..., 1)^{\mathsf{T}}$. Since $T(\cdot)$ is nondecreasing, it follows that $I(\cdot)$ is also nondecreasing. Based on these processes another essential property of multiclass queueing networks which are considered in this thesis can be stated. This property is called the *non-idling property*, which is also called the *work-conserving property*. This means that a station can idle at time t only if currently there are no customers in the queue. That is, if for station j it holds that $I_j(t_2) > I_j(t_1)$ for $t_1 < t_2$, then there exists a $t \in [t_1, t_2]$ such that $W_j(t) = 0$. As $I(\cdot)$ is continuous, the non-idling property can also be written as

$$\int_0^\infty W_j(t)\,\mathrm{d}I_j(t) = 0.$$

Finally, before we put the dynamic equations in a nutshell, we note that the HL property of a queueing network can be expressed in the following condition $\Gamma(D(t)) \leq T(t) < \Gamma(D(t) + e)$, where the inequalities have to be understood componentwise. In vector form the dynamic equations of a multiclass queueing network can be summarized by

$$Q(t) = Q(0) + A(t) - D(t) \geq 0, \tag{2.3}$$

$$T(\cdot) \text{ is nondecreasing, with } T(0) = 0, \tag{2.4}$$

$$W(t) = C\,\Gamma(Q(0) + A(t)) - C\,T(t), \tag{2.5}$$

$$I(t) = et - CT(t), \quad I(\cdot) \text{ is nondecreasing}, \tag{2.6}$$

$$I_j(t) \text{ can only increase when } W_j(t) = 0,\ j \in \{1, ..., J\}, \tag{2.7}$$

$$\text{additional conditions on } (Q(\cdot), T(\cdot)), \text{ specific to the discipline.} \tag{2.8}$$

In the following we refer to (2.3)-(2.8) as the *queueing network equations*. The processes defining the basic queueing network equation determine evolution of the queueing network. Due to this we simply call

$$\mathsf{X}(t) = (A(t), D(t), T(t), W(t), I(t), Q(t))$$

the queueing network process. In the reminder of the section we state the additional conditions (2.8) for the services discipline introduced in Section 2.1.

FIFO disciplines

In FIFO queueing networks the customers are served in the order of their arrivals. So, the allocation process is determined by

$$D_k(t + W_j(t)) = Q_k(0) + A_k(t), \quad j = c(k).$$

for all $t \geq 0$. The role of the initial data is served by $Q(0)$ and

$$\{D_k(s) \text{ for } s \leq W_j(0),\ j = c(k),\ k \in \{1, ..., K\}\}.$$

Priority disciplines

Let the priority ordering be defined by the permutation π. For each customer class k let

$$\Pi_k := \{k' \in \{1, ..., K\} \,:\, c(k') = c(k),\ \pi(k') \leq \pi(k)\}$$

denote the set of customer classes that are served at the some station $j = c(k)$ and have priority at least as k. Correspondingly, we define

$$T_k^+(t) := \sum_{l \in \Pi_k} T_l(t).$$

Preemptive priority disciplines

For priority queueing networks with preemption the additional condition is then

$$t - T_k^+(t) \text{ can only increase if } Q_k^+ := \sum_{l \in H_k} Q_l(t) = 0, \qquad k \in \{1, ..., K\}$$

for all $t \geq 0$. This can be rewritten as

$$\int_0^\infty Q_k^+(t)\ \mathrm{d}(t - T_k^+(t)) = 0, \qquad k \in \{1, ..., K\}.$$

The role of the initial data is served by $Q(0)$.

Nonpreemptive priority disciplines

In what follows, we state the allocation process for the case that the priority discipline is without preemption. For a jump function $f : \mathbb{R} \to \mathbb{Z}_+$ the time instant of the last jump can be characterized by

$$l(t; f) := \sup\{s \leq t\ :\ |f(s^-) - f(s)| \geq 1\}.$$

Let n_j denote the number of customer classes that are served at station $j \in \{1, ..., J\}$ and let $C(j) = \{k_{j1}, \dots, k_{jn_j}\}$ denote the set of customer classes that are served at station j, the ordering is decreasing with respect to the priorities, i.e. k_{jl} has priority over $k_{j,l+1}$. The indicator set for station j and customer class k_{jl} is

$$\widetilde{\mathcal{I}}_{k_{jl}}(t) = \{\exists\, u \,|\, l\,(t; Q_{k_{jl}}) \leq u \leq t\ :\ Q_{k_{jl'}}(u) = 0,\ l' < l,\ Q_{k_{jl}}(u) > 0\},$$

where $l = 1, \dots, n_j$. Further, using the superscript c to denote the complement, we define the sets

$$\begin{aligned} \mathcal{I}_{k_{jn_j}}(t) &:= \widetilde{\mathcal{I}}_{k_{jn_j}}(t) \\ \mathcal{I}_{k_{jl}}(t) &:= \widetilde{\mathcal{I}}_{k_{jl}}(t) \cap (\mathcal{I}_{k_{j,l+1}}(t))^c, \quad l = n_j - 1, \dots, 1. \end{aligned}$$

Then, the allocation process can then be defined by

$$T_{k_{jl}}(t) = \int_0^t \mathbb{1}_{\{\mathcal{I}_{k_{jl}}(s)\}}\, \mathrm{d}s$$

for $j \in \{1, ..., J\}$ and $l = 1, ..., n_j$. The initial data is given by $Q(0)$.

HLPS disciplines

Under the HL processor sharing discipline the oldest customers of each nonempty class are served simultaneously and the amount of service is shared equally among them. That is, the allocation of class k customers up to time t is given by

$$T_k(t) := \int_0^t \frac{\mathbb{1}_{\{Q_k(s)>0\}}}{\sum_{l\in C(c(k))} \mathbb{1}_{\{Q_l(s)>0\}}} \, \mathrm{d}s.$$

It can be checked that the evolution of the process $\mathsf{X}(t)$ is completely determined by the primitive cumulatives and the initial queue length. So, in this case the initial data is given by $Q(0)$.

HLPPS disciplines

In HL proportional processor sharing disciplines a station serves the oldest customers of each nonempty class simultaneously, where the service effort is not shared equally, but proportional the queue length. That is, the allocation in the time period $[0, t]$ that class k customers received from station $j = c(k)$ is given by

$$T_k(t) := \int_0^t \frac{Q_k(s)}{\sum_{l\in C(c(k))} Q_l(s)} \, \mathrm{d}s.$$

The primitive cumulatives as well as the initial queue length determine evolution of the process $\mathsf{X}(t)$ for all $t \geq 0$.

2.4 Fluid Approximation and Stability

In this section we provide an approach that was first considered by Rybko and Stolyar in 1992 [71] and was further developed by Stolyar [81] and Dai [28]. The idea is to investigate scaled versions of the Markov process. To this end, let $(r_n, x_n)_{n\in\mathbb{N}}$ be a sequence of pairs, where $r_n \in \mathbb{R}_+$ and $x_n \in \mathcal{X}$ is a sequence of initial states. We assume that the sequence of pairs satisfies the following conditions

$$\lim_{n\to\infty} r_n = \infty, \quad \limsup_{n\to\infty} \tfrac{\|q_n\|}{r_n} < \infty, \quad \lim_{n\to\infty} \tfrac{\|u_n\|}{r_n} = \lim_{n\to\infty} \tfrac{\|v_n\|}{r_n} = 0, \tag{2.9}$$

where q_n, u_n and v_n denote the queue length, the residual interarrival time, and the residual service time, respectively. In the sequel, we consider the family $\mathsf{X}' := \{\mathsf{X}_n(t) : t \geq 0, n \in \mathbb{N}\}$ of Markov processes defined by

$$\mathsf{X}_n(t) := \tfrac{1}{r_n} \mathsf{X}^{x_n}(r_n t),$$

where the superscript x_n expresses the dependence on the initial state $x_n \in \mathcal{X}$. In order to investigate the family X' we start to focus on the primitive cumulatives (E, S, R). In the next lemma we recall the convergence results of the scaled versions of the primitive cumulatives.

Lemma 2.4.1 ([28]) *Assume that the sequence of pairs $(r_n, x_n)_{n\in\mathbb{N}}$ satisfies* (2.9). *Then, as $n \to \infty$, almost surely*

$$\begin{aligned} \tfrac{1}{r_n} E_k^{x_n}(\, r_n\, t\,) &\longrightarrow \alpha_k t \quad \textit{u.o.c.}, \\ \tfrac{1}{r_n} S_k^{x_n}(\, r_n\, t\,) &\longrightarrow \mu_k t \quad \textit{u.o.c.}, \\ \tfrac{1}{r_n} R^k(\, [r_n\, t]\,) &\longrightarrow P_k\, t \quad \textit{u.o.c.}, \end{aligned} \tag{2.10}$$

where $[a]$ denotes the integer part of $a \in \mathbb{R}$.

Proof. By Assumptions 2.1.2 (1) and (2), the primitive increments satisfy the strong law of large numbers, i.e. as $n \to \infty$ almost surely it holds for each $k \in \{1, ..., K\}$ that

$$\tfrac{1}{n}\sum_{i=1}^{n} a_k(i) \longrightarrow \tfrac{1}{\alpha_k}, \quad \tfrac{1}{n}\sum_{i=1}^{n} s_k(i) \longrightarrow \tfrac{1}{\mu_k}, \quad \tfrac{1}{n}\sum_{i=1}^{n} \phi^k(i) \longrightarrow P_k. \tag{2.11}$$

The assertion then follows from Theorem 1.2.9. ∎

We denote by G the set of all sample paths satisfying (2.11) and use G as the set on which we will take fluid limits. Note that $\mathbb{P}[G] = 1$ and that the fluid limits can be taken on any set G' with $\mathbb{P}[G'] = 1$ such that (2.11) is satisfied.

Theorem 2.4.2 ([28]) *For each HL queueing network, $(r_n, x_n)_{n\in\mathbb{N}}$ satisfying* (2.9)*, and $\omega \in G$, there is a subsequence of pairs $(r_{n_i}, x_{n_i})_{i\in\mathbb{N}}$ such that*

$$\lim_{i\to\infty} \tfrac{1}{r_{n_i}} \mathsf{X}^{x_{n_i}}(r_{n_i} t, \omega) = \overline{\mathsf{X}}(t, \omega) \quad \textit{u.o.c.} \tag{2.12}$$

Proof. Let $\omega \in G$ be a fixed sample path. For (x_n, r_n) and $k \in \{1, ..., K\}$, as $\frac{1}{r_n} I_k^{x_n}(r_n \cdot, \omega)$ is nondecreasing, it follows from (2.6) that

$$\frac{1}{r_n} T_k^{x_n}(r_n t, \omega) - \frac{1}{r_n} T_k^{x_n}(r_n s, \omega) \le t - s$$

for all $0 \le s \le t$. That is, the family $\{\frac{1}{r_n} T^{x_n}(r_n t, \omega), n \in \mathbb{N}\}$ is equicontinuous. Hence, by the Arzelà-Ascoli Theorem A.1 and a diagonal argument there is a subsequence of pairs $(r_{n_i}, x_{n_i})_{i \in \mathbb{N}}$ such that

$$\lim_{i\to\infty} \frac{1}{r_{n_i}} T_k^{x_{n_i}}(r_{n_i} t, \omega) = \overline{T}_k(t, \omega) \qquad u.o.c.$$

for some process $\overline{T}_k(t, \omega)$. Applying (2.6) again yields that

$$\lim_{i\to\infty} \frac{1}{r_{n_i}} I_k^{x_{n_i}}(r_{n_i} t, \omega) = \overline{I}_k(t, \omega) \qquad u.o.c.$$

and with Lemma 2.4.1 we conclude that

$$\lim_{i\to\infty} \frac{1}{r_{n_i}} A_k^{x_{n_i}}(r_{n_i} t, \omega) = \overline{A}_k(t) \quad u.o.c.\,,$$
$$\lim_{i\to\infty} \frac{1}{r_{n_i}} D_k^{x_{n_i}}(r_{n_i} t, \omega) = \overline{D}_k(t) \quad u.o.c.$$

Moreover, it follows from Lemma 2.4.1 and the balance equation (2.3) that

$$\lim_{i\to\infty} \frac{1}{r_{n_i}} Q_k^{x_{n_i}}(r_{n_i} t, \omega) = \overline{Q}_k(t, \omega) \qquad u.o.c.$$

Finally, the convergence of the workload follows from (2.5). ■

Any limit $\overline{\mathsf{X}}(\cdot) = (\overline{A}(\cdot), \overline{D}(\cdot), \overline{T}(\cdot), \overline{W}(\cdot), \overline{I}(\cdot), \overline{Q}(\cdot)$ obtained from a scaling (2.12) is a called a *fluid limit* of the discipline. The set of all fluid limits associated with the sample path ω is denoted by $\mathcal{FL}(\omega)$. So, whenever a fluid limit is considered it is always assumed that $\overline{\mathsf{X}}(\cdot) \in \mathcal{FL}(\omega)$ for some $\omega = \{a(n), s(n), \phi(n), n \in \mathbb{N}\}$. Hence, after taking limits some randomness may remain.

Definition 2.4.3 *The collection of all fluid limits for all sample paths ω is called the fluid limit model, i.e.* $\mathcal{FLM} = \{\mathsf{X}(\omega) : \omega \in G\}$.

Next, we show that any fluid limit satisfies a set of dynamic equations which is analog to the queueing network equations (2.3)-(2.8), which are obtained by replacing the primitive cumulatives by their limits of the scaling. That is, using $M = \mathrm{diag}(\mu)$,

$$\overline{A}(t) = \alpha t + P^{\mathsf{T}} M\overline{T}(t), \tag{2.13}$$

$$\overline{Q}(t) = \overline{Q}(0) + \overline{A}(t) - M\overline{T}(t) \geq 0, \tag{2.14}$$

$$\overline{T}(0) = 0 \text{ and } \overline{T}(\cdot) \text{ is nondecreasing}, \tag{2.15}$$

$$\overline{W}(t) = C\, M^{-1}\, (\overline{Q}(0) + \overline{A}(t)) - C\,\overline{T}(t), \tag{2.16}$$

$$\overline{I}(t) = et - C\,\overline{T}(t) \text{ and } \overline{I}(\cdot) \text{ is nondecreasing}, \tag{2.17}$$

$$\overline{I}_j(t) \text{ can only increase when } \overline{W}_j(t) = 0,\ j \in \{1, ..., J\}, \tag{2.18}$$

$$\text{additional conditions on } (\overline{Q}(\cdot), \overline{T}(\cdot)),\ \text{ specific to the discipline.} \tag{2.19}$$

For convenience, we refer to this set of equations in the following as the *fluid equations*.

Definition 2.4.4 *A pair* $(\overline{Q}(\cdot), \overline{T}(\cdot))$ *is called a fluid solution if it satisfies the fluid equations. In addition, the set of all fluid solutions to the equations* (2.13)-(2.19) *is called the associated fluid network, denoted by* $\mathcal{FN}$.

The associated fluid network is a purely deterministic network which is based on the mean values of the primitive increments of the stochastic queueing network.

Theorem 2.4.5 ([28]) *Let* $\overline{\mathsf{X}}(\cdot)$ *be a fluid limit. Then,* $(\overline{Q}(\cdot), \overline{T}(\cdot))$ *is a fluid solution.*

Proof. The equations (2.13)-(2.17) follow immediately from the proof of Theorem 2.4.2. To prove (2.18) we assume that for some $j \in \{1, ..., J\}$ there is an interval $[r, s]$ such that $\overline{W}_j(t) > 0$ for all $t \in [r, s]$. Since $\overline{W}_j(\cdot)$ is continuous and the workload process $\frac{1}{r_{n_i}} W_j^{x_{n_i}}(r_{n_i} t)$ converges u.o.c. to $\overline{W}_j(t)$, for i sufficiently large it holds that $\frac{1}{r_{n_i}} W_j^{x_{n_i}}(r_{n_i} t) > 0$ for all $t \in [r, s]$. Then, by equation (2.7) it holds that

$$\frac{1}{r_{n_i}} I_j^{x_{n_i}}(r_{n_i} r) = \frac{1}{r_{n_i}} I_j^{x_{n_i}}(r_{n_i} s).$$

The assertion then follows since the limit $i \to \infty$ preserves the equality. ■

An immediate consequence of the above theorem is that $\mathcal{FLM} \subset \mathcal{FN}$.

Remark 2.4.6 *A counterexample in [31, Section 2.7] shows that, in general, the inclusion $\mathcal{FLM} \subset \mathcal{FN}$ is strict. Furthermore, the fluid solutions are not unique in general. A related counterexample can be found in [17] Example 1 in Section 4.3.*

In the following we address a fundamental question in queueing theory, namely to derive, for a given queueing network, conditions that characterize the fact that the network has an unique equilibrium in the sense that there is an attractive invariant probability measure. The following definition gives a precise statement when a queueing network is called stable.

Definition 2.4.7 *A multiclass queueing network is called stable if the underlying Markov process X is positive Harris recurrent.*

Based on the definition of positive Harris recurrence and its equivalent characterization in Theorem 1.4.1, the first step towards a sufficient condition of the stability of a multiclass queueing network is to investigate closed petite sets. The following lemma shows that the Assumptions 2.1.2 provide a closed petite set.

Lemma 2.4.8 ([28]) *If the interarrival times satisfy the Assumptions 2.1.2, then for all $\kappa > 0$ the set*

$$A = \{x \in \mathcal{X} : |x| \leq \kappa\} \tag{2.20}$$

is closed and petite.

Proof. See [17] Section 4.2. ∎

In what follows, we state the main result of this section, namely a sufficient criterion for the stability of a multiclass queueing network in terms of the fluid limit model. To this end, we have to introduce the concept of stability for the fluid limit model. This is done in the following definition.

Definition 2.4.9 *A fluid limit model of a queueing discipline is said to be stable if there is a $\tau > 0$ such that for any fluid limit $\overline{\mathsf{X}}(\cdot) \in \mathcal{FLM}$ the $\overline{Q}(\cdot)$ component satisfies $\overline{Q}(t) = 0$ for all $t \geq \tau \|\overline{Q}(0)\|$.*

Now, we state the main theorem of this chapter.

Theorem 2.4.10 ([28]) *Let a queueing discipline be fixed. Assume that Assumptions 2.1.2 are satisfied. If the fluid limit model is stable, then the queueing network is stable. In particular, if the associated fluid network is stable, then the queueing network is stable.*

In general, it is not easy to work with fluid limits. Due to the fact that the associated fluid network is a deterministic model, we will work with the associated fluid network.

Remark 2.4.11 *Lemma 2.4.8 is the only place where Assumption 2.1.2* (3) *appears explicitly. These conditions put proper restrictions on the distributions of the interarrival times. To this end, one can allow for general distributions of the interarrival times and show directly that condition* (2.20) *is satisfied for the particular situation.*

The relation of multiclass queueing networks to its fluid limit models and the associated fluid networks is the content of the following remark.

Remarks 2.4.12

(a) *A partial converse to Theorem 2.4.10 due to [30] is the following. If the fluid limit model (resp. the associated fluid network) is weakly unstable, i.e. for each sample path $\omega \in G$ there is a $\delta > 0$ that may depend on ω such that $\overline{Q}(\delta) \neq 0$ for each $\overline{Q}(\cdot) \in \mathcal{FL}(\omega)$ (respectively $\overline{Q}(\cdot) \in \mathcal{FN}$) with $\overline{Q}(0) = 0$, then the queueing network is unstable in the sense that a.s. we have that*

$$\lim_{t\to\infty} \|\overline{Q}(t)\| = \infty.$$

(b) *Bramson has shown by a counterexample that there are stable multiclass queueing networks where the associated fluid network is unstable. That is, a converse to the second statement in Theorem 2.4.10 may not hold, see [17].*

(c) *In addition, a counterexample by Dai, Hasenbein, Vande Vate shows that the stability of a multiclass queueing network may depend on the distributions of the primitive increments, see [34] . As a matter of fact, knowing the mean values of the primitive increments may not be sufficient to conclude stability. Hence, in general the associated fluid network is not able to completely describe the stability of the queueing network.*

(d) *Further partial converse results can be found in [60], [67].*

The remainder of this chapter is devoted to the properties of the class of fluid networks associated to work-conserving queueing networks. To this end, we consider the set of solutions to the *basic fluid equations* (2.13)-(2.18). So, for simplicity we omit the overline symbol in the following.

Proposition 2.4.13 ([19] Lipschitz continuity)
The fluid solutions $(Q(\cdot), T(\cdot))$ are Lipschitz continuous with a global Lipschitz constant. In particular, the pair is differential almost everywhere with respect to the Lebesgue measure on $[0, \infty)$.

Proof. Let $(Q(\cdot), T(\cdot))$ be a fluid solution. By (2.17) the idle process $I(\cdot)$ is nondecreasing. Hence, by (2.15) for any $k \in \{1, ..., K\}$ and for all $0 \leq s \leq t$,

$$T_k(t) - T_k(s) \leq \sum_{l \in C(j)} T_l(t) - T_l(s) = t - s - (I_j(t) - I_j(s)) \leq t - s.$$

Thus, $T(\cdot)$ is Lipschitz continuous with constant $L = 1$. Further, the balance equation (2.14) shows that the fluid level process $Q(\cdot)$ is Lipschitz continuous and that there is a global Lipschitz constant. Further, by Rademacher's Theorem the processes are differential almost everywhere cf. [44, Theorem 5.8.6]. ■

In order to state further properties we introduce a scaling and a shift operator. Given a function $f : \mathbb{R}_+ \to \mathbb{R}_+^K$, for $r > 0$ the *scaling operator* σ_r is defined by

$$\sigma_r f(t) := \tfrac{1}{r} f(r\,t), \tag{2.21}$$

and for $s \geq 0$ the *shift operator* δ_s is defined by

$$\delta_s f(t) := f(t + s). \tag{2.22}$$

The subsequent statement is a scaling and shift property of the fluid solutions.

Proposition 2.4.14 ([19] Scaling and shift property)
Let $(Q(\cdot), T(\cdot))$ be a fluid solution with initial fluid level $Q(0)$.

(1) For each $r > 0$ the pair $(\sigma_r Q(\cdot), \sigma_r T(\cdot))$ is a fluid solution with initial fluid level $\sigma_r Q(0)$.

(2) For each $s \geq 0$ the pair $(\delta_s Q(\cdot), \delta_s T(\cdot) - \delta_s T(0))$ is a fluid solution with initial fluid level $\delta_s Q(0)$.

Proof. The assertion follows directly from the basic fluid equations. ■

The next property is about the convergence of the fluid level processes. We note that for fluid limits the shift property and the convergence is not immediately clear, see Section 5.2.

Proposition 2.4.15 ([86] Closedness in the uniform topology) *Let $(Q_n(\cdot), T_n(\cdot))_{n\in\mathbb{N}}$ be a sequence of fluid solutions with initial fluid level $Q_n(0)$. Suppose $(Q_n(\cdot), T_n(\cdot))_{n\in\mathbb{N}}$ converges u.o.c. to $(Q_*(\cdot), T_*(\cdot))$ as $n \to \infty$. Then, $(Q_*(\cdot), T_*(\cdot))$ is a fluid solution with initial fluid level $Q_*(0)$.*

Proof. To avoid any double subscripts we use superscripts to denote the sequences. Suppose that $(Q^n(\cdot), T^n(\cdot))_{n\in\mathbb{N}}$ is a sequence of fluid solutions that converges u.o.c. to some $(Q^*(\cdot), T^*(\cdot))$ as $n \to \infty$. Then, it follows directly that the limit $(Q^*(\cdot), T^*(\cdot))$ satisfies the equations (2.13)-(2.17). To verify (2.18) it suffices to show that if $\sum_{k\in C(j)} Q_k^*(\cdot) > 0$ on some interval implies this implies that $I_j^*(\cdot)$ is constant on this interval. So, consider a station $j \in \{1, ..., J\}$ such that $\sum_{k\in C(j)} Q_k^*(t) > 0$. Hence, there are $\varepsilon, \delta > 0$ such that for all $s \in [t-\delta, t+\delta]$,

$$\sum_{k\in C(j)} Q_k^*(s) \geq 2\varepsilon. \tag{2.23}$$

By the convergence hypothesis it holds that $\sum_{k\in C(j)} Q_k^n(t)$ converges u.o.c. to $\sum_{k\in C(j)} Q_k^*(t)$. and, thus, n sufficiently large we have

$$|\sum_{k\in C(j)} Q_k^n(s) - \sum_{k\in C(j)} Q_k^*(s)| \leq \varepsilon$$

for all $s \in [t-\delta, t+\delta]$. Further, by (2.23) and the triangular inequality it holds that

$$\sum_{k\in C(j)} Q_k^n(s) \geq \sum_{k\in C(j)} Q_k^*(s) - \varepsilon \geq \varepsilon > 0.$$

Then, by (2.18), for all $s \in [t-\delta, t+\delta]$ we have that $I_j^n(s) = I_j^n(t-\delta)$. Hence, for $n \to \infty$ this yields for all $s \in [t-\delta, t+\delta]$ that

$$I_j^*(s) = I_j^*(t-\delta)$$

and so $\dot{I}_j^*(t) = 0$. This shows the assertion. ■

2.5 Notes and References

This chapter introduces multiclass queueing networks and collects results from the literature which will be useful in the remainder of this thesis. The intention of this summary embraces mainly two issues. First, it motivates why the stability analysis of fluid networks is an interesting topic. To this end, it is recalled how fluid networks emerge in the stability analysis of multiclass queueing networks. In addition, this chapter displays that the basic fluid network equations have their origin in the dynamic equations describing the evolution of the performance processes of a multiclass queueing network. Second, the degree of accuracy results, on the one hand, from the fact that the properties of the fluid network stated in the Propositions 2.4.13, 2.4.14 and 2.4.15 play a fundamental role in the derivation of a Lyapunov theory for fluid networks in Chapter 4. On the other hand, in Chapter 5 we provide an alternative proof of Theorem 2.4.10 in terms of the Lyapunov theory that we will develop in Chapter 4.

The model description given in Section 2.1 is taken from [17] and [28]. The Assumptions 2.1.2, which are appropriate to enable a stability analysis of multiclass queueing networks by means of the associated fluid network, stem from [28]. Also, the state space description for the disciplines considered in this chapter come from [28]. Whereas, the outline of the general state space has its origin in [17]. Further, the comprehensive depiction of the underlying Markov process describing the evolution of the HL queueing network can also be found in [17]. The fact that the underlying process X defines a PDP, and hence a Borel right process, has also been addressed in [17, 28, 54].

The basic dynamic equations characterizing the evolution of the queueing network under certain disciplines can be found in many places, see for instance [16, 17, 19, 23, 24, 28, 31, 33].

The development of the fluid approximation is related to [17]. Applying a SLLN scaling of the stochastic processes, as presented in Section 2.4, avoids dealing with a delayed fluid model, as does the approach in [28]. By Theorem 5.3 in [19] this is no restriction since the stability of the undelayed fluid model implies the stability of the delayed fluid model.

Finally, these results provide the origin for the fluid limit approach to tackle the stability issue. This framework was first introduced by Rybko and Stolyar in 1992 to investigate the stability of a two station queueing network, cf. [71]. In 1995 Dai generalized this to multiclass queueing networks, see [28].

The derivation of the properties of the fluid networks associated to HL queueing networks are taken from [19, 81, 86]. In the subsequent chapter we turn our attention to the derivation of stability conditions for associated fluid networks under the disciplines introduced in this chapter. In doing so, it will turn out that the use of Lyapunov theory is a powerful tool.

3 Stability of Fluid Networks

The fluid approximation method for the stability analysis of multiclass queueing networks leads to the objective of deriving stability conditions for fluid networks. Hence, in this chapter we focus on stability conditions for fluid networks operating under the disciplines introduced in the previous chapter. It will be shown that for each individual discipline there are separate sufficient stability criteria.

In a first step, we will consider fluid networks under general work-conserving disciplines. The only specification for this class is the non-idling property, i.e. a station can only idle if there are no customers waiting to be served, see condition (2.18). It will be shown that a necessary stability condition is that the nominal workload at each station is strictly less than one. As this class is a superset for all the disciplines considered in this thesis the nominal workload condition is also necessary for all other disciplines.

Furthermore, in this chapter we run through the sufficient stability conditions for each discipline that are available in the literature. As the focus of this thesis is to see the stability analysis from a Lyapunov perspective, we recall those proofs from the literature that are based on Lyapunov arguments.

In the first section we recall the definition of stability for fluid networks. Further, the section contains an equivalent characterization for stability, which is apparently weaker. In addition, Section 3.1 contains the Lyapunov framework that lays the foundation for the stability analysis of the individual disciplines. In the subsequent sections the fluid networks under the particular disciplines are examined in detail. The final section of this chapter provides a comparison to the stability theory for dynamical systems based on differential equations.

3.1 Generalities on Stability for Fluid Networks

A fluid network, as introduced in Section 2.4, is a continuous deterministic analog to a multiclass queueing network. For this reason, the variables describing fluid networks are used accordingly. However, the interpretation is different. A fluid network consists of J stations that serve K different classes of fluids. For each fluid class k the variable α_k is interpreted as the rate at which fluids flow into the network. The corresponding vector $\alpha \in \mathbb{R}_+^K$ is called the *exogenous inflow rate.* Likewise, the variable $\mu_k \in \mathbb{R}_+$ is interpreted as the *potential outflow rate* of class k fluids and the substochastic matrix $P \in [0,1]^{K\times K}$ is considered as the *flow transfer matrix.* Also, it is assumed that $\varrho(P) < 1$. However, to keep the analogy and for simplicity we refer to α, μ and P as the arrival rate, service capacity, and routing matrix, respectively.

The many-to-one map c from classes to stations and the corresponding *constituency matrix* C are completely analog to the queueing network case. The initial fluid level of the network and the *fluid level process* are denoted by $Q(0) \in \mathbb{R}_+^K$ and $Q := \{Q(t) \in \mathbb{R}_+^K,\, t \geq 0\}$, respectively. Given the parameters α and μ, the structure P and C, and the initial fluid level $Q(0)$, the evolution of the fluid network is determined by the discipline. On the analogy of queueing network the *allocation process*, denoted by $T = \{T(t) \in \mathbb{R}_+^K, t \geq 0\}$, represents the discipline, where $T_k(t)$ denotes the cumulative amount of time in the time period $[0,t]$ that station $c(k)$ has allocated to serving fluids of class k. Hence, the quantity $\mu_k T_k(t)$ reflects the cumulative outflow of class k fluids up to time t.

Recall that the processes are Lipschitz and so differential almost everywhere. Further, the set of equations, called the basic fluid equations, describing a fluid network is the following

$$Q(t) = Q(0) + \alpha\, t - (I - P^{\mathsf{T}})MT(t) \geq 0, \tag{3.1}$$

$$T(\cdot) \text{ is nondecreasing, with } T(0) = 0, \tag{3.2}$$

$$I(t) = et - CT(t), \quad I(\cdot) \text{ is nondecreasing}, \tag{3.3}$$

$$0 = (C\,Q(t))^{\mathsf{T}}\, \dot{I}(t) \quad \text{ for almost all } t \geq 0, \tag{3.4}$$

where $M = \operatorname{diag}(\mu)$ and $I = \{I(t),\, t \geq 0\}$ is the *cumulative idle process.* To specify the network for a particular discipline you have to add at least one more equation describing the discipline. Since a fluid network is defined by the parameters α, μ, P, C and the discipline π we denote a fluid network by (α, μ, P, C, π). Recall that the formal definition of stability of fluid networks is the following.

Definition 3.1.1 *A fluid network* (α, μ, P, C, π) *is said to be stable if there exists a* $\tau > 0$ *such that* $Q(t) \equiv 0$ *for all* $t \geq \tau \|Q(0)\|$ *and for all fluid level processes* $Q(\cdot)$*.*

The following result contains an apparently weaker condition for stability of fluid networks. It turns out that the subsequent characterization of stability is useful to provide a comparison to Lyapunov stability for dynamical systems. For simplicity we denote by $\Phi(1)$ the set of fluid level processes with total initial level one, i.e. $\|Q(0)\| = 1$.

Theorem 3.1.2 ([81]) *A fluid network is stable if and only if for any fluid level process* $Q(\cdot) \in \Phi(1)$ *it holds that*

$$\inf_{t \geq 0} \|Q(t)\| < \|Q(0)\| = 1. \tag{3.5}$$

Proof. It suffices to show that (3.5) implies Definition 3.1.1. In a first step we show that there is a $r \in [0,1)$ such that for all fluid level processes $Q(\cdot) \in \Phi(1)$ it holds that

$$\inf_{t \geq 0} \|Q(t)\| < r. \tag{3.6}$$

Suppose that such an r does not exist. Then, there is a sequence $(r_n)_{n \in \mathbb{N}}$ with $\lim_{n \to \infty} r_n = 1$ and a sequence $(Q_n(\cdot))_{n \in \mathbb{N}}$ with $Q_n(\cdot) \in \Phi(1)$ such that

$$\inf_{t \geq 0} \|Q_n(t)\| \geq r_n.$$

Since the family $(Q_n(\cdot))_{n \in \mathbb{N}}$ is equicontinuous and by a diagonal argument, there is a convergent subsequence $(Q_{n_l}(\cdot))_{l \in \mathbb{N}}$ that satisfies $\lim_{l \to \infty} Q_{n_l}(\cdot) = Q(\cdot) \in \Phi(1)$ and $\inf_{t \geq 0} \|Q(t)\| = 1$, which contradicts (3.5).

Let $r \in [0,1)$ be fixed such that (3.6) holds. Then, given any $Q(\cdot) \in \Phi(1)$ there exists a $t_1 > 0$ such that

$$t_1 = \min\{t \geq 0 : ||Q(t)|| = r\} < \infty. \tag{3.7}$$

The latter, the scaling, and the shift property imply that there are times $t_k \in \mathbb{R}_+$ such that for $k = 1, 2, ...$ we have that

$$t_k = \min\{t \geq 0 : ||Q(t)|| = r^k\} < \infty\,. \tag{3.8}$$

Hence, for every $Q(\cdot) \in \Phi(1)$ it holds that

$$\inf_{t \geq 0} \|Q(t)\| = 0. \tag{3.9}$$

Moreover, for any $r \in (0, 1)$ there is a $T_r \in \mathbb{R}_+$ such that

$$\sup_{Q(\cdot) \in \Phi(1)} \min\{t \geq 0 : ||Q(t)|| = r\} \leq T_r < \infty. \tag{3.10}$$

To see this, suppose that this is not true. Then, there is a fluid level process $Q(\cdot) \in \Phi(1)$ such that

$$\inf_{t \geq 0} \|Q(t)\| \geq r,$$

which provides a contradiction to (3.9).

Let r be constant as given in (3.7). In the sequel, we construct for any given $Q(\cdot) \in \Phi(1)$ a sequence $0 = \tau_0 < \tau_1 < \tau_2 < ...$ such that for $k = 1, 2, ...$ we have

$$||Q(\tau_k)|| = r^k\,.$$

Initially, we take $\tau_1 = t_1$ defined by (3.7). Then, by the scaling and shift property we have $\frac{1}{r} Q(rt + t_1) \in \Phi(1)$ and, hence, there is a $t_2 < \infty$ such that

$$t_2 = \min\{t \geq 0 \,:\, Q(rt + t_1) = r^2\}.$$

Then, defining

$$\tau_k := r^{k-1} t_k + r^{k-2} t_{k-1} + ... + t_1$$

yields that $||Q(\tau_k)|| = r^k$. Further, the sequence $(\tau_k)_{k \in \mathbb{N}}$ is Cauchy since

$$\tau_k - \tau_{k-1} = r^{k-1}\, t_k \leq r^{k-1}\, T_r \qquad \text{for } k = 1, 2, ... \,.$$

Moreover, (3.10) implies that

$$\lim_{k\to\infty} \tau_k = \lim_{k\to\infty} \sum_{i=1}^{k} r^{i-1}\, t_i \le T_r \sum_{i=1}^{\infty} r^{i-1} = \frac{T_r}{1-r} < \infty.$$

Thus, the sequence $(\tau_k)_{k\in\mathbb{N}}$ has a finite limit, denoted by τ^*, that satisfies $\tau^* \le \frac{T_r}{1-r}$. In addition, the continuity of $Q(\cdot)$ and the norm implies that

$$||Q(\tau^*)|| = ||Q(\lim_{k\to\infty} \tau_k)|| = \lim_{k\to\infty} ||Q(\tau_k)|| = \lim_{k\to\infty} r^k = 0.$$

Consequently, there is a $\tau > 0$ such that

$$\sup_{Q(\cdot)\in\Phi(1)} \min\{t \ge 0 : ||Q(t)|| = 0\} \le \tau < \infty. \tag{3.11}$$

Thus, it remains to show that $Q(\cdot)$ stays zero for all $t \ge \tau^*$. Suppose there is a $t_* > \tau^*$ such that $\|Q(t_*)\| > 0$. Then, by the continuity of $\|Q(\cdot)\|$ there are t', t'' with $t' < t''$ such that $\|Q(t')\| = 0$ and $\|Q(s)\| > 0$ for all $s \in (t', t'']$. For any $\varepsilon > 0$ such that $\varepsilon < \max\{\,\|Q(s)\| : s \in [t', t'']\,\}$ we consider $t_\varepsilon := \min\{\,s \ge t' : \|Q(s)\| = \varepsilon\,\}$. Moreover, let $\bar{t} := \min\{\,s \ge t_\varepsilon : \|Q(s)\| = 0\,\}$, which clearly satisfies $\bar{t} > t''$. By the scaling and shift property we have $\frac{1}{\varepsilon}Q(\varepsilon\cdot + t_\varepsilon) \in \Phi(1)$. Thus, by (3.11) we have $\bar{t} \le t_\varepsilon + \varepsilon\tau$. Now, for $\varepsilon \to 0$ it follows by definition that $t_\varepsilon \to t'$ and $\bar{t} \to t'$, which yields a contradiction to $t_1 < t_2$. This shows the assertion. ■

Before presenting a Lyapunov method for fluid networks we consider the following auxiliary lemma.

Lemma 3.1.3 ([31]) *Let $f : \mathbb{R}_+ \to \mathbb{R}_+$ be absolutely continuous and let $\varepsilon > 0$ be fixed. Suppose that if $f(t) > 0$ and f is differentiable at t, it holds that $\dot{f}(t) \le -\varepsilon$. Then, $f(t) = 0$ for all $t \ge \varepsilon^{-1} f(0)$.*

Proof. Let $t^* := \inf\{\,t > 0 : f(t) = 0\,\}$. Suppose that there is a $t' > t^*$ such that $f(t') > 0$. Then, by the continuity of f there exists a $\delta > 0$ such that $f(t' - \delta) = 0$ and $f(s) > 0$ for all $s \in (t' - \delta, t']$. Since f is absolutely continuous it holds that

$$f(t') = f(t') - f(t' - \delta) = \int_{t'-\delta}^{t'} \dot{f}(s)\, \mathrm{d}s \le -\varepsilon \cdot \delta,$$

which is a contradiction to the choice of t'. Therefore, such a t' cannot exist. In particular, the function f is nonincreasing and if $f(t) = 0$ and f is differentiable at t it holds that $\dot{f}(t) = 0$. If $f(0) > 0$, then as f is absolutely continuous we have for all $t \in [0, t^*)$ that

$$f(t) - f(0) = \int_0^t \dot{f}(s)\,\mathrm{d}s \le -\varepsilon\, t$$

and hence $f(t) \le f(0) - \varepsilon\, t$. Let $t_0 := \frac{f(0)}{\varepsilon}$, then we have $t^* \le t_0$. Since f is nonincreasing, t_0 is the time when f will definitely become zero and stay zero afterwards. ■

Based on the latter result, we recall a Lyapunov criterion for the stability of fluid networks. It will turn out in the subsequent sections that this condition is an effective tool for establishing stability conditions for individual disciplines.

Theorem 3.1.4 ([31]) *Consider a fluid network (α, μ, P, C, π). Let $V : \mathbb{R}_+^K \to \mathbb{R}_+$ be locally Lipschitz such that $V(x) = 0$ if and only if $x = 0$ and let $\varepsilon > 0$. Assume that for each fluid level process $Q(\cdot)$ it holds that*

$$\dot{V}(Q(t)) := \tfrac{\mathrm{d}}{\mathrm{d}t}\left[V(Q(t))\right] \le -\varepsilon, \tag{3.12}$$

whenever $Q(t) \neq 0$ and t is regular for the map $s \mapsto V(Q(s))$. Then, the fluid network is stable.

Proof. According to the scaling and shift property it suffices to consider fluid level processes $Q(\cdot) \in \Phi(1)$. Since the fluid level processes $Q(\cdot)$ are Lipschitz and V is locally Lipschitz it follows that the function $t \mapsto V(Q(t))$ is locally Lipschitz and hence differentiable almost everywhere. Then, as $V(Q(t)) = 0$ if and only if $Q(t) = 0$ and by Lemma 3.1.3, we have $V(Q(t)) = 0$, and thus $Q(t) = 0$ for all $t \ge \varepsilon^{-1} V(Q(0))$. For

$$\delta := \max\{V(x) \,:\, x \in \mathbb{R}_+^K,\ \|x\| = 1\}$$

we have $V(Q(0)) \le \delta$ and so $Q(t) = 0$ for all $t \ge \varepsilon^{-1}\,\delta$. This shows the assertion. ■

The properties of the function V, namely the positive definiteness and the decrease along each fluid level process with uniform rate, resemble the characteristics of a Lyapunov function for an ordinary differential equation.

For this reason, we refer to V as a *Lyapunov function.* A precise definition and a discussion of the relation to the theory of ordinary differential equations is provided at the end of this chapter.

In the subsequent sections we investigate necessary and sufficient conditions for the stability of fluid networks under the disciplines introduced in Chapter 2. The strategy to evaluate sufficient stability conditions will be to find appropriate Lyapunov functions and apply Theorem 3.1.4.

3.2 General Work-Conserving Fluid Networks

In the first place we consider stability properties of fluid networks defined by the basic fluid equations. As the only put proper restriction is the non-idling condition (3.4), these networks are called *general work-conserving fluid networks*. Following Proposition 2.4.13 the processes are differential almost everywhere. Hence, the work-conserving property in its differentiation reads as

$$\sum_{k\in C(j)} Q_k(t)\cdot \dot{I}_j(t) = \Big(\sum_{k\in C(j)} Q_k(t)\Big)\cdot\Big(1-\sum_{k\in C(j)} \dot{T}_k(t)\Big) = 0.$$

Every pair $(Q(\cdot),T(\cdot))$ satisfying the basic fluid equations (3.1)-(3.4) is called a *work-conserving* pair. The following theorem guarantees the existence of a work-conserving pair, given the parameters (α,μ,P,C) and the initial fluid level $Q(0)$. This statement was shown by Chen, cf. appendix in [19]. We provide a new and elegant proof using the theory of differential inclusions in Section 5.2.

Theorem 3.2.1 ([19]) *For any (α,μ,P,C) and $Q(0)$ there exists at least one work-conserving allocation process $T(\cdot)$.*

Next, we show that the validity of the nominal workload condition $\rho < e$ is necessary for stability, where $\rho = CM^{-1}(I-P^{\mathsf{T}})^{-1}\alpha$.

Theorem 3.2.2 ([19]) *Suppose that the general work-conserving fluid network (α,μ,P,C) is stable, then the nominal workload condition $\rho < e$ holds.*

Proof. As the fluid network is stable there is a $\tau > 0$ such that $Q(\tau+\cdot)\equiv 0$ for every fluid level process with initial level $\|Q(0)\| = 1$. Multiplying the flow balance equation (3.1) by $C\,M^{-1}\,(I-P^{\mathsf{T}})^{-1}$ yields for $t=\tau$ that

$$\begin{aligned}0 = CM^{-1}(I-P^{\mathsf{T}})^{-1}Q(\tau) &= CM^{-1}(I-P^{\mathsf{T}})^{-1}Q(0) + \rho\,\tau - CT(\tau)\\ &= CM^{-1}(I-P^{\mathsf{T}})^{-1}Q(0) + (\rho - e)\,\tau + I(\tau)\\ &\geq CM^{-1}(I-P^{\mathsf{T}})^{-1}Q(0) + (\rho - e)\,\tau > (\rho - e)\tau,\end{aligned}$$

where the strict inequality follows by choosing $Q(0) = \frac{1}{K}e$ and the fact that $CM^{-1}(I-P^{\mathsf{T}})^{-1}$ is nonnegative and in each row there is at least one positive entry. ■

Here we note that the converse of the latter theorem is valid if $J=1$ or $K=J$. In order to formulate a sufficient condition for the stability of a fluid network under general work-conserving disciplines, we have to introduce strictly copositive matrices. A symmetric matrix $A \in \mathbb{R}^{K\times K}$ is called *strictly copositive* if for every $x \in \mathbb{R}^K_+$ it holds that $x^{\mathsf{T}}Ax \geq 0$ and $x^{\mathsf{T}}Ax = 0$ if and only if $x=0$. Moreover, for $c \in \mathbb{R}$ we make use of the notation $c^- := \min\{c,0\}$ and $c^+ := \max\{c,0\}$.

Theorem 3.2.3 ([19]) *Consider a work-conserving fluid network. Suppose there exists a $K\times K$ symmetric strictly copositive matrix A such that for $k=1,...,K$ it holds that*

$$\theta_k := -\left(\sum_{i=1}^{K}\alpha_i a_{ik} - \min_{i\in C(c(k))} h_{ik} - \sum_{\substack{j=1\\ j\neq c(k)}}^{J}\left(\min_{i\in C(j)} h_{ik}\right)^{-}\right) > 0,$$

where $H = M(I-P)A$. Then, the network is stable.

Proof. Let $Q(\cdot)$ be a fluid level process of the work-conserving fluid network such that $\|Q(0)\| = 1$. Consider the Lyapunov function $V : \mathbb{R}^K_+ \to \mathbb{R}$ defined by

$$V(x) = x^{\mathsf{T}}\,A\,x.$$

Clearly, V is locally Lipschitz, satisfies $V(\cdot)\geq 0$, and $V(x)=0$ if and only if $x=0$. Then, we consider the derivate of V along a fluid level process

$Q(\cdot)$. That is,

$$
\begin{aligned}
\tfrac{1}{2}\dot{V}(Q(t)) = \Big(\alpha^{\mathsf{T}} - \dot{T}(t)^{\mathsf{T}}\, M\, (I-P)\Big)\, A\, Q(t) \\
= \alpha^{\mathsf{T}}\, A\, Q(t) - \sum_{k=1}^{K}\sum_{i=1}^{K} \dot{T}_i(t)\, h_{ik}\, Q_k(t) \\
= \alpha^{\mathsf{T}}\, A\, Q(t) - \sum_{k=1}^{K}\sum_{j=1}^{J}\sum_{i\in C(j)} h_{ik}\, Q_k(t)\, \dot{T}_i(t). \qquad (3.13)
\end{aligned}
$$

To conclude the decrease condition, we consider the very last term in (3.13) individually. Using condition (3.3) it holds that

$$
\begin{aligned}
\sum_{i\in C(j)} h_{ik}\, Q_k(t)\, \dot{T}_i(t) &\geq \min_{i\in C(j)} h_{ik} \sum_{i\in C(j)} Q_k(t)\, \dot{T}_i(t) \\
&= \min_{i\in C(j)} h_{ik} Q_k(t) \cdot (1 - \dot{I}_j(t)).
\end{aligned}
$$

Further, by the work-conserving property (3.4) we have that

$$
\begin{aligned}
\sum_{j=1}^{J} \min_{i\in C(j)} h_{ik} Q_k(t) \cdot \dot{I}_j(t) &= \sum_{\substack{j=1\\ j\neq c(k)}}^{J} \min_{i\in C(j)} h_{ik} Q_k(t) \cdot \dot{I}_j(t) \\
&\leq \sum_{\substack{j=1\\ j\neq c(k)}}^{J} \Big(\min_{i\in C(j)} h_{ik}\Big)^{+} Q_k(t).
\end{aligned}
$$

Hence, we obtain the following estimate for the very last term in (3.13)

$$
\begin{aligned}
&\sum_{k=1}^{K}\sum_{j=1}^{J} \min_{i\in C(j)} h_{ik} Q_k(t) \cdot (1 - \dot{I}_j(t)) \\
&\qquad \geq \sum_{k=1}^{K} \Bigg[\min_{i\in C(c(k))} h_{ik} + \sum_{\substack{j=1\\ j\neq c(k)}}^{J} \Big(\min_{i\in C(j)} h_{ik}\Big)^{-} \Bigg] Q_k(t).
\end{aligned}
$$

Consequently, this yields

$$\tfrac{1}{2}\dot{V}(Q(t)) \leq \sum_{k=1}^{K} \left(\sum_{i=1}^{K} \alpha_i \, a_{ik} - \min_{i \in C(c(k))} h_{ik} - \sum_{\substack{j=1 \\ j \neq c(k)}}^{J} \left(\min_{i \in C(j)} h_{ik} \right)^{-} \right) Q_k(t).$$

Using the introduced notation this reads as

$$\dot{V}(Q(t)) \leq -\, 2\, \theta^{\mathsf{T}}\, Q(t) \leq 0.$$

Further, we define

$$2\,\gamma := \inf_{\substack{x \geq 0, \\ x \neq 0}} \frac{\theta^{\mathsf{T}}\, x}{\sqrt{x^{\mathsf{T}} A x}} = \inf_{\substack{x \geq 0, \\ \|x\| = 1}} \frac{\theta^{\mathsf{T}}\, x}{\sqrt{x^{\mathsf{T}} A x}}$$

and note that $\gamma > 0$. Moreover, using $W(x) := \sqrt{V(x)}$ it follows that

$$\dot{W}(Q(t)) = \frac{\dot{V}(Q(t))}{2\sqrt{Q^{\mathsf{T}}(t)\, A\, Q(t)}} \leq \frac{-\theta^{\mathsf{T}} Q(t)}{\sqrt{Q^{\mathsf{T}}(t)\, A\, Q(t)}} \leq -\gamma,$$

whenever $\|Q(t)\| > 0$. Hence, an application of Lemma 3.1.3 yields that $Q(t) = 0$ for all $t \geq \gamma^{-1} W(Q(0))$. ∎

The class of general work-conserving fluid networks contains all non-idling fluid networks since the dynamic equations for a fluid network under a certain work-conserving discipline are a specification of the dynamic equations (3.1)-(3.4). So, if the general work-conserving fluid network is stable the network is stable under every work-conserving discipline. For this reason, the stability of general work-conserving fluid networks is also called *global stability*.

3.3 Priority Fluid Networks

In a priority regime the various fluid classes are served at the stations according to a predefined priority ordering. The *priority* is determined by a permutation $\pi : \{1, ..., K\} \to \{1, ..., K\}$. Given fluid classes $l, k \in \{1, ..., K\}$ served at the same station $c(l) = c(k)$, fluids of class l are said to have a higher priority than fluids of class k if $\pi(l) < \pi(k)$. So, fluids of class k are not served as long as the fluid level of class l is greater than zero. For each fluid class $k \in \{1, ..., K\}$ the set

$$\Pi_k := \{l : l \in C(c(k)), \pi(l) \leq \pi(k)\}$$

contains all fluid classes which are served at the same station $c(k)$ and have a higher priority than fluids of class k. In the following the symbol Π is used to express the priority discipline. To describe the dynamic equations of a priority fluid network (α, μ, P, C, Π), we introduce the *unused capacity process* $Y(t)$, where $Y_k(t)$ denotes the cumulative remaining capacity of station $c(k)$ for serving fluids of classes that have a lower priority than class k fluids. The dynamic equations can be summarized as follows

$$Q(t) = Q(0) + \alpha\, t - (I - P^{\mathsf{T}})MT(t) \geq 0, \tag{3.14}$$

$$T(\cdot) \text{ is nondecreasing, with } T(0) = 0, \tag{3.15}$$

$$Y_k(t) = t - \sum_{l \in \Pi_k} T_l(t) \text{ and } Y(\cdot) \text{ is nondecreasing,} \tag{3.16}$$

$$0 = Q_k(t)\, \dot{Y}_k(t) \quad \text{for almost all } t \geq 0, \quad k \in \{1, ..., K\}. \tag{3.17}$$

This representation is close to the dynamic equations of a general work-conserving fluid network. In the following we want to derive conditions such that the fluid network (α, μ, P, C, Π) is stable. To this end, we present an alternative description of the dynamic equations that will be more appropriate for the stability analysis.

Let $\Pi_k^+ := \Pi_k \backslash \{k\}$ and for the case that k is not the fluid class with highest priority at station $c(k)$, let $h(k)$ denote the fluid class that has next higher priority, which is also served at station $c(k)$. That is,

$$h(k) := \arg\max\{\pi(l) : l \in \Pi_k^+\}$$

and $h(k) := 0$ if $\Pi_k^+ = \emptyset$. Also, consider the matrix $B \in \mathbb{R}^{K \times K}$ defined by $b_{lk} = 1$ if $h(k) = l$ and $b_{lk} = 0$ else. Let e^{Π} be defined by $e_k^{\Pi} = 1$ if $\Pi_k^+ = \emptyset$ and $e_k^{\Pi} = 0$ else. With this notation the allocation process $T(\cdot)$ can be expressed by

$$T(t) = -(I - B)Y(t) + e^{\Pi}\, t.$$

In addition, using $\theta = \alpha - (I - P^{\mathsf{T}})\, M\, e^{\Pi}$ and $R = (I - P^{\mathsf{T}})\, M\, (I - B)$ the flow balance equation (3.14) reads as

$$Q(t) = Q(0) + \theta\, t + R\, Y(t) \geq 0. \tag{3.18}$$

Further, the priority regime can also be expressed by the following condition

$$\forall\, l, k \in \{1, ..., K\}, \forall\, t \geq 0 \text{ regular}:$$
$$c(l) = c(k),\ \pi(l) < \pi(k)\ \Rightarrow\ \dot{Y}_l(t) \geq \dot{Y}_k(t). \quad (3.19)$$

To get further insight into the priority regime we consider partitions (a, b) of set of fluid classes $\{1, ..., K\}$ that takes the priority discipline into account in the sense that if fluid class l is in a and k is a fluid class that is also served at the station $c(l)$ and has higher priority than l, then it is also contained in a. More formally, we define the following.

Definition 3.3.1 *A partition (a, b) of $\{1, ..., K\}$ is said to be hierarchical with respect to π, if $l \in a$ and $k \in \Pi_l^+$, then $k \in a$. The set of all hierarchical partitions of $\{1, ..., K\}$ is denoted by $\mathcal{H}$.*

Let $(a, b) \in \mathcal{H}$ be a hierarchical partition, then the flow balance equation (3.18) in block form reads as

$$\begin{pmatrix} Q_a(t) \\ Q_b(t) \end{pmatrix} = \begin{pmatrix} Q_a(0) \\ Q_b(0) \end{pmatrix} + \begin{pmatrix} \theta_a \\ \theta_b \end{pmatrix} t + \begin{pmatrix} R_a & R_{ab} \\ R_{ba} & R_b \end{pmatrix} \begin{pmatrix} Y_a(t) \\ Y_b(t) \end{pmatrix}. \quad (3.20)$$

In addition, for $(a, b) \in \mathcal{H}$ with $b \neq \emptyset$ we define

$$S(b) := \{u \geq 0 \,:\, \theta_b + R_b u = 0 \text{ and } u \leq e\}.$$

Theorem 3.3.2 ([24]) *Consider the fluid network (α, μ, P, C, Π). Assume $\rho < e$ and that there is an $\varepsilon > 0$ and a vector $h \in \mathbb{R}^K_{>0}$ such that for each partition $(a, b) \in \mathcal{H}$ the condition*

$$h_a^\mathsf{T}(\theta_a + R_{ab} x_b) < -\varepsilon \quad (3.21)$$

holds for $x_b \in S(b)$ if $b \neq \emptyset$, and for $x_b = 0$ if $b = \emptyset$. In addition, if $S(b) = \emptyset$, then condition (3.21) is also assumed to hold. Then, the fluid network is stable.

Proof. Let $(Q(\cdot), Y(\cdot))$ be a pair that satisfies the dynamic equations. Since both are Lipschitz, their derivative exists almost everywhere. Let the Lyapunov function be defined by

$$V(x) := h^\mathsf{T} x.$$

Clearly, V is locally Lipschitz, satisfies $V(\cdot) \geq 0$, and $V(x) = 0$ if and only if $x = 0$. Following Lemma 3.1.3 it suffices to prove that for all regular t and $Q(t) \neq 0$ it holds that $\dot{V}(Q(t)) \leq -\varepsilon$. So, let t be a regular point and let the partition (a, b) be defined by

$$a = \{k \in \{1, ..., K\} : \dot{Y}_k(t) = 0\} \text{ and } b = \{k \in \{1, ..., K\} : \dot{Y}_k(t) > 0\}.$$

By condition (3.19) it follows that $(a, b) \in \mathcal{H}$. Suppose that $b \neq \emptyset$. Then, taking derivatives and using that $\dot{Y}_a(t) = 0$ the balance equation (3.20) yields

$$\begin{aligned} \dot{Q}_a(t) &= \theta_a + R_{ab}\dot{Y}_b(t) \\ \dot{Q}_b(t) &= \theta_b + R_b\dot{Y}_b(t). \end{aligned}$$

By the definition of b, for any $k \in b$ it holds that $\dot{Y}_k(t) > 0$. Then, by (3.17) it follows that $Q_k(t) = 0$. In addition, on the one hand it holds that $\dot{Q}_k(t^-) \leq 0$ and, since the fluid levels are nonnegative, on the other hand we have that $\dot{Q}_k(t^+) \geq 0$. Thus, it follows that $\dot{Q}_k(t) = 0$. That is, we have that $\dot{Q}_b(t) = 0$ and this in turn implies that

$$\dot{V}(Q(t)) = h_a^{\mathsf{T}}\dot{Q}_a(t) = h_a^{\mathsf{T}}(\theta_a + R_{ab}\dot{Y}_b(t)).$$

Since $\dot{Y}_b(\cdot) \geq 0$ and $Y_b(t) - Y_b(s) \leq e(t-s)$ it follows that $0 \leq \dot{Y}_b(t) \leq e$ and $\dot{Y}_b(t) \in S(b)$. Consequently, by (3.21) it follows that $\dot{V}(Q(t)) < -\varepsilon$. Also, if $b = \emptyset$ it holds that $\dot{Q}(t) = \theta$ and since condition (3.21) holds for $x = 0$ we have that $\dot{V}(Q(t)) = h^{\mathsf{T}}\theta < -\varepsilon$. This shows the assertion. ■

The sufficient condition of the previous theorem can be weakend, as the following shows. To describe the more general sufficient condition some notations are in order. For $b \subseteq \{1, ..., K\}$ the set

$$\Pi(b) := \{k \in \{1, ..., K\} : l \in b \text{ and } c(k) = c(l) \Rightarrow \pi(k) < \pi(l)\}$$

defines the set of highest priority fluid classes in each station that serves at least one fluid class in b. For each partition $(a, b) \in \mathcal{H}$ the set

$$S(a, b) := \{q \in \mathbb{R}^K : q \geq 0,\ q_a = 0 \text{ and } q_{\Pi(b)} > 0\}$$

defines the fluid state. Furthermore, we consider the condition

$$0 \leq x \leq e \quad \text{and} \quad \theta_a + R_a x = 0, \tag{3.22}$$

and define the set of regular flow rates by

$$F(a,b) := \{d = (0\ d_b)^\mathsf{T} \in \mathbb{R}^K\ :\ d_b = \theta_b + R_{ba}y_a \text{ where } y_a \text{ satisfies (3.22)}\}.$$

That is, the dynamic equations imply for every partition $(a,b) \in \mathcal{H}$ that if t is a regular point and $Q(t) \in S(a,b)$, then we have that $\dot{Q}(t) \in F(a,b)$. In fact, $F(a,b)$ is exactly the set of all such derivatives.

Theorem 3.3.3 ([22]) *Suppose there is an $\varepsilon > 0$, an integer $N \geq 1$, and N nonnegative K dimensional vectors $h_1, \ldots, h_N$, such that the following conditions hold.*

(a) Associated to each partition $(a,b) \in \mathcal{H}$ with $b \neq \emptyset$ and $F(a,b) \neq \emptyset$, there is an Index set $I(a,b) \subseteq \{1,\ldots,N\}$ such that for all $i \in I(a,b)$

$$\sup_{d \in F(a,b)} h_i^\mathsf{T} d \leq -\varepsilon. \tag{3.23}$$

b) For any partion $(a,b) \in \mathcal{H}$ with $b \neq \emptyset$ and $F(a,b) \neq \emptyset$, and any $j \notin I(a,b)$, there is an $i \in I(a,b)$ such that

$$(h_j)_b \leq (h_i)_b. \tag{3.24}$$

Then, $V(x) := \max\limits_{1 \leq i \leq N} h_i^\mathsf{T} x$ is a piecewise linear Lyapunov function. In particular, the fluid network (α, μ, P, C, Π) is stable.

Proof. See [22] Theorem 3.1. ■

3.4 HLPPS Fluid Networks

In fluid networks under the *Head-of-the-Line Proportional Processor Sharing* discipline (HLPPS) all nonempty fluid classes present at a station are served simultaneously proportional to their current fluid level. The total fluid level at station $j = c(k)$ at time t is given by

$$Q_j^\Sigma(t) := {e_j}^\mathsf{T} C\, Q(t) = \sum_{l \in C(j)} Q_l(t).$$

If the overall fluid level at station j is positive, the *allocation rate* $\dot{T}_k(t)$ of class k fluids served at station $j = c(k)$ is defined by

$$\dot{T}_k(t) = \frac{Q_k(t)}{Q^{\Sigma}_{c(k)}(t)}. \tag{3.25}$$

Note that even if $Q^{\Sigma}_j(t) = 0$ the allocation rate $\dot{T}_k(t)$ for $k \in C(j)$ may still be positive to keep the total fluid level at station j zero. The idle time process is denoted by $I = \{I(t) : t \geq 0\}$, where $I_j(t)$ denotes the cumulative time that station $j = c(k)$ idles in the interval $[0, t]$. The dynamic equations of the fluid network under HLPPS discipline can be summarized as follows

$$Q(t) = Q(0) + \alpha\, t - (I - P^{\mathsf{T}})MT(t) \geq 0, \tag{3.26}$$

$$W(t) = C\, M^{-1}\, Q(t), \tag{3.27}$$

$$I(t) = et - C\, T(t), \tag{3.28}$$

$$\dot{T}_k(t) = \frac{Q_k(t)}{Q^{\Sigma}_{c(k)}(t)}, \qquad \text{if} \quad Q^{\Sigma}_{c(k)}(t) > 0. \tag{3.29}$$

$$I_j(t) \text{ can only increase if } W_j(t) = 0,\, j \in \{1, ..., J\}, \tag{3.30}$$

Before addressing stability, we consider the behavior of HLPPS fluid level processes present at the same station.

Lemma 3.4.1 *Given a HLPPS fluid network, suppose that at station j there is at least one initial fluid level nonempty, i.e. $Q^{\Sigma}_j(0) > 0$. Further, let t_0 be defined by*

$$t_0 := \inf\{\, t > 0 \,|\, \exists\, k \in C(j) \,:\, Q_k(t) = 0 \,\}.$$

Then, it holds that $t_0 > 0$ and if $t_0 < \infty$ we have that $Q^{\Sigma}_j(t_0) = 0$.

Proof. If $Q_k(0) > 0$ for all $k \in C(j)$ there is nothing to show. Suppose there is a fluid class $l \in C(j)$ with $Q_l(0) = 0$. To prove that $t_0 > 0$ we note that, since $Q^{\Sigma}_j(0) > 0$, it holds that

$$\dot{Q}_l(0) = \alpha_l + \sum_{\substack{k \in C(j),\\ k \neq l}} p_{kl}\ \mu_k\ \frac{Q_k(0)}{Q^{\Sigma}_j(0)} + \sum_{k \notin C(j)} p_{kl}\ \mu_k\, \dot{T}_k(0) > 0. \tag{3.31}$$

Hence, for $l \in C(j)$ the fluid level process $Q_l(\cdot)$ is strictly increasing at time zero. Furthermore, as $Q(\cdot)$ is Lipschitz there is a $T > 0$ such that $Q_l(t) > 0$ for all $t \in (0, T]$. Thus, it follows that $t_0 > 0$.

To show the second claim, suppose that $t_0 < \infty$. Let $Q(\cdot)$ be a fluid level process. By assumption, its initial level $Q(0)$ satisfies $Q_j^{\Sigma}(0) > 0$. Without loss of generality let the fluid classes and stations be numbered such that $j = 1$, $Q_k(0) > 0$ for all $k \in C(1)$, and $Q_1(t_0) = 0$. To show the assertion, we assume contrary that the claim does not hold, i.e.

$$\exists\, \varepsilon > 0,\; l \in C(1) \text{ such that } Q_l(t) \geq \varepsilon > 0 \quad \forall\, t \in [0, t_0]. \tag{3.32}$$

Then, for fluid class 1 it holds for all $t \in [0, t_0]$ that

$$\dot{Q}_1(t) = \alpha_1 - (\mu_1 - p_{11}) \cdot \frac{Q_1(t)}{Q_1^{\Sigma}(t)} + \sum_{k=2}^{K} p_{k1}\, \mu_k\, \dot{T}_k(t),$$

where $\dot{T}(\cdot)$ is a feasible allocation rate corresponding to the fluid level process $Q(\cdot)$. On one hand, it holds that $\alpha_1 + \sum_{k=2}^{K} p_{k1}\, \mu_k\, \dot{T}_k(t) \geq 0$. On other hand, since $Q_1^{\Sigma}(t) \geq Q_l(t) \geq \varepsilon$ for all $t \in [0, t_0]$, we have

$$(\mu_1 - p_{11}) \frac{Q_1(t)}{Q_1^{\Sigma}(t)} \leq \frac{\mu_1}{\varepsilon} Q_1(t)$$

for all $t \in [0, t_0]$. Hence, it holds that

$$\dot{Q}_1(t) \geq -\tfrac{\mu_l}{\varepsilon} Q_1(t). \tag{3.33}$$

Any solution to $\dot{x}(t) = -c\, x(t)$, with $c > 0$ and $x(0) > 0$ satisfies $x(t) = x(0)\, e^{-ct} > 0$. Thus, we have that $Q_1(t) > 0$ for all $t \geq 0$, which is a contradiction to (3.32). ■

HLPPS fluid networks have the nice property that the nominal workload condition $\rho < e$ is also sufficient for stability, cf. [16]. We discuss Bramson's stability condition for HLPPS fluid network from a Lyapunov perspective in terms of Theorem 3.1.4.

Theorem 3.4.2 *A HLPPS fluid network is stable if and only if $\rho < e$.*

The necessity of the nominal workload condition follows from Theorem 3.2.2.

To investigate the reverse implication we consider the Lyapunov function V defined by

$$V(x) = e^{\mathsf{T}} C M^{-1} (I - P^{\mathsf{T}})^{-1} x.$$

Then, V is locally Lipschitz and, since the spectral radius of P is strictly less than one, we also have that $V(x) = 0$ if and only if $x = 0$. To show that the decrease condition (3.12) is satisfied, we define $\varepsilon := \min_{j=1,\dots,J} 1 - \rho_j$. Let $Q(t) \neq 0$ and let t be regular for $V(Q(\cdot))$. The derivative of V along each fluid level processes is given by

$$\dot{V}(Q(t)) = e^{\mathsf{T}} C M^{-1} (I - P^{\mathsf{T}})^{-1} \dot{Q}(t) = e^{\mathsf{T}} \rho - e^{\mathsf{T}} C \dot{T}(t).$$

If the total fluid level at every station is positive, i.e. $Q_j^{\Sigma}(t) > 0$ for every $j = 1, \dots, J$, it follows from (3.29) that $C\dot{T}(t) = e$. Hence, we have that

$$\dot{V}(Q(t)) = e^{\mathsf{T}} (\rho - e) \leq -\varepsilon.$$

Suppose there is a station i that is empty, i.e. $Q_i^{\Sigma}(t) = 0$, while $Q_j^{\Sigma}(t) > 0$ for all $j \neq i$. Without loss of generality let $i = 1$. In this case the allocation rates for classes $k \in C(1)$ are not determined by (3.29). However, the allocation rates are still positive to keep the fluid level equally zero, which is possible since $\rho_1 < 1$. Let $a := C(1)$ and $b := \{1, \dots, K\} \setminus a$. Then, the flow balance equation in its differential form reads as

$$\begin{aligned} 0 &= \alpha_a + P_a^{\mathsf{T}} M_a \dot{T}_a(t) + P_{ab}^{\mathsf{T}} M_b \dot{T}_b(t) - M_a \dot{T}_a(t) \\ \dot{Q}_b(t) &= \alpha_b + P_{ba}^{\mathsf{T}} M_a \dot{T}_a(t) + P_b^{\mathsf{T}} M_b \dot{T}_b(t) - M_b \dot{T}_b(t). \end{aligned}$$

On one hand the allocation rates for the fluid classes in b are determined by (3.29). On other hand, as the spectral radius $\varrho(P) < 1$ the Neuman series converges $\sum_{n=0}^{\infty} (P_a^{\mathsf{T}})^n$ converges. Thus, the allocation rates for the fluid classes in a are given

$$\dot{T}_a(t) = M_a{}^{-1} (I_a - P_a^{\mathsf{T}})^{-1} \left(\alpha_a + P_{ab}^{\mathsf{T}} M_b \dot{T}_b(t) \right), \tag{3.34}$$

which yield that $Q_1^{\Sigma}(t + \cdot) = 0$. In order to investigate the allocation rates for fluid classes a further, we consider the according partition the traffic equation. That is,

$$\begin{aligned} \lambda_a &= \alpha_a + P_a^{\mathsf{T}} \lambda_a + P_{ab}^{\mathsf{T}} \lambda_b \\ \lambda_b &= \alpha_b + P_{ba}^{\mathsf{T}} \lambda_a + P_b^{\mathsf{T}} \lambda_b \end{aligned}$$

and therefore we have

$$(I_a - P_a^\mathsf{T})^{-1}\alpha_a = \lambda_a - (I_a - P_a^\mathsf{T})^{-1}P_{ab}^\mathsf{T}\lambda_b.$$

Hence, the allocation rates for fluid classes a are given by

$$\dot{T}_a(t) = {M_a}^{-1}\Big(\lambda_a - P_{ab}^\mathsf{T}\,\lambda_b + (I_a - P_a^\mathsf{T})^{-1}P_{ab}^\mathsf{T}\,M_b\dot{T}_b(t)\Big). \tag{3.35}$$

Thus, using $e_a^\mathsf{T} := (1, ..., 1)^\mathsf{T} \in \mathbb{R}^{|a|}$ and if $Q_1^\Sigma(t) = 0$ we have

$$\rho_1 - \sum_{l\in C(1)} \dot{T}_l(t) = e_a^\mathsf{T}{M_a}^{-1}\Big(P_{ab}^\mathsf{T}\,\lambda_b - (I_a - P_a^\mathsf{T})^{-1}P_{ab}^\mathsf{T}\,M_b\dot{T}_b(t)\Big).$$

Hence, in this case the derivative of V along $Q(\cdot)$ equals

$$\begin{aligned}\dot{V}(Q(t)) = e^\mathsf{T}(\rho - C\,\dot{T}(t)) &= \rho_1 - \sum_{l\in C(1)} \dot{T}_l(t) + \sum_{j=2}^{J}\rho_j - 1\\ &= e_a^\mathsf{T}{M_a}^{-1}\Big(P_{ab}^\mathsf{T}\,\lambda_b - (I_a - P_a^\mathsf{T})^{-1}P_{ab}^\mathsf{T}\,M_b\dot{T}_b(t)\Big) + \sum_{j=2}^{J}\rho_j - 1.\end{aligned} \tag{3.36}$$

The following example, however, shows that this expression is nonnegative in general.

Example 3.4.3 *We consider a two station network serving three classes. Let the arrival rates be given by* $\alpha = \begin{pmatrix}1 & 1 & 0\end{pmatrix}^\mathsf{T}$ *and let the service capacities be given by* $\mu = \begin{pmatrix}4 & 4 & 1+\varepsilon\end{pmatrix}^\mathsf{T}$ *with* $\varepsilon > 0$*. The routing of the network is defined in Figure 3.1. The effective arrival rates are* $\lambda = \begin{pmatrix}1 & 1 & 1\end{pmatrix}^\mathsf{T}$ *and the nominal workload is* $\rho = \begin{pmatrix}\frac{1}{2} & \frac{1}{1+\varepsilon}\end{pmatrix}^\mathsf{T}$*. Further, it holds that*

$$M^{-1}(I - P^\mathsf{T})^{-1} = \begin{pmatrix}\frac{1}{4} & 0 & 0\\ 0 & \frac{1}{4} & 0\\ 0 & \frac{1}{1+\varepsilon} & \frac{1}{1+\varepsilon}\end{pmatrix}.$$

Let the initial fluid level $Q(0) = \begin{pmatrix}Q_1 & Q_2 & Q_3\end{pmatrix}^\mathsf{T}$ *be such that* $Q_1 >> Q_2 > 0$ *and* $Q_3 = 0$*. Then, for small* t *to keep the fluid level at station* 2 *equal to*

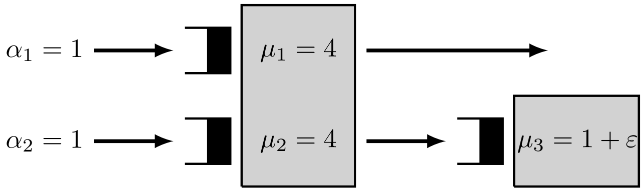

Figure 3.1: A two station network serving three fluid classes.

zero the allocation rate of fluid class 3 *is*

$$\dot{T}_3(t) = \tfrac{4}{1+\varepsilon}\dot{T}_2(t) = \tfrac{4}{1+\varepsilon}\tfrac{Q_2(t)}{Q_1^\Sigma(t)}.$$

Hence, for small $t \geq 0$ *equation* (3.36) *reads as*

$$\dot{V}(Q(t)) = \tfrac{1}{2} - 1 + \tfrac{1}{1+\varepsilon} - \dot{T}_3(t) = -\tfrac{1}{2} + \tfrac{1}{1+\varepsilon} - \tfrac{4}{1+\varepsilon}\tfrac{Q_2(t)}{Q_1^\Sigma(t)}.$$

Since $Q_1 >> Q_2 > 0$ *we have for* ε *and* Q_2 *sufficiently small that*

$$\dot{V}(Q(t)) = -\tfrac{1}{2} + \tfrac{1}{1+\varepsilon} - \tfrac{4}{1+\varepsilon}\tfrac{Q_2(t)}{Q_1^\Sigma(t)} > 0.$$

The significance of the previous example is that the Lyapunov function candidate $V(x) = e^\mathsf{T} C M^{-1}(I - P^\mathsf{T})^{-1}x$ yields the desired decrease condition as long as all stations are nonempty. Unfortunately, once there is some empty station in the network the decrease of V along $Q(\cdot)$ can no longer be concluded.

3.5 FIFO Fluid Networks

For fluid networks under FIFO disciplines the fluids are served in the order of their arrivals. To describe the evolution of class k fluids we have to consider the immediate workload given by $W(t) = C\,M^{-1}\,Q(t)$. Any fluids that arrive after time t will have a lower priority in the FIFO discipline. So, fluids that arrive at time t are served at time $t + W_j(t)$, where $W_j(t)$ denotes the workload of the station $j = c(k)$. The total arrivals up to time t are given by

$$A(t) = \alpha t + P^\mathsf{T}\,M\,T(t).$$

For each class $k \in \{1, ..., K\}$ the FIFO regime can be represented by the relation

$$T_k(t + W_j(t)) = m_k(Q_k(0) + A_k(t)), \tag{3.37}$$

where $m_k = \mu_k^{-1}$. So, the dynamic equations describing a FIFO fluid network may be summarized as follows

$$Q(t) = Q(0) + \alpha\, t - (I - P^{\mathsf{T}})MT(t) \geq 0, \tag{3.38}$$

$$T(\cdot) \text{ is nondecreasing, with } T(0) = 0, \tag{3.39}$$

$$I(t) = et - CT(t), \quad I(\cdot) \text{ is nondecreasing}, \tag{3.40}$$

$$0 = (C\, Q(t))^{\mathsf{T}}\, \dot{I}(t) \quad \text{for almost all } t \geq 0, \tag{3.41}$$

$$T_k(t + W_j(t)) = m_k(Q_k(0) + A_k(t)), \quad \forall\, k \in \{1, ..., K\}. \tag{3.42}$$

We note that a FIFO fluid network is not completely determined by the initial fluid level $Q(0)$. This is because it has to be specified how the initial fluid level is served in $[0, W_j(0)]$. So, the initial data for each class $k \in \{1, ..., K\}$ is given by $Q(0)$ and

$$\{T_k(s) \,:\, s \in [0, W_j(0)]\,\}. \tag{3.43}$$

Remark 3.5.1 *Bramson and Seidman have shown the surprising result that FIFO fluid networks satisfying the nominal workload condition are not stable in general, see [13] and [77]. Moreover, Bramson has also shown that FIFO fluid networks of Kelly type, i.e. the service capacities of those classes present at the same station are equal, are stable if and only if the nominal workload condition is satisfied, see [15, Theorem 1].*

To state sufficient stability conditions, derived by Chen and Zhang, recall that the vector of effective arrival rates is given by $\lambda = (I - P^{\mathsf{T}})^{-1}\alpha$ and that the spectral radius is denoted by ϱ. Further, let $\Lambda = \mathrm{diag}(\lambda)$.

Theorem 3.5.2 ([23]) *Suppose the FIFO fluid network satisfies $\rho < e$ and one of the following conditions*

(*i*) $\varrho(\, P^{\mathsf{T}} + (I - P^{\mathsf{T}})\, \Lambda C^{\mathsf{T}} CM\,) < 1$,

(*ii*) $\varrho(\, CMP^{\mathsf{T}}(I - P^{\mathsf{T}})\Lambda C^{\mathsf{T}}\,) < 1$.

Then, the FIFO fluid network is stable.

We do not the recall the proof for this statement since it does not rely on Lyapunov arguments. The interested reader is referred to [23]. Besides, this indicates that the FIFO discipline is special among the disciplines considered in this thesis. We will establish this in Section 5.2.

3.6 Stability Theory of Differential Equations

The notion of stability introduced in Definition 3.1.1 may also be interpreted as the zero fluid level process $Q_0(\cdot) \equiv 0$ being the unique stable and attractive fixed point of the shift operator $\delta_\tau Q(\cdot) := Q(\tau + \cdot)$ defined on the set fluid level processes. To see this, suppose that $Q_*(\cdot)$ is another fixed point. Then, for all $t \geq 0$ it holds that

$$Q_*(t) - Q_0(t) = \delta_\tau Q_*(t) = 0,$$

where the last equality is valid since the network is stable.

Next, we recall briefly the definition of stability and Lyapunov functions from the theory of dynamical systems. For a detailed description the reader is referred e.g. to [7], [53]. Consider an ordinary differential equation

$$\dot{x}(t) = f(x(t)), \qquad x \in \mathbb{R}^n,\, t \in [0, \infty) \tag{3.44}$$

with initial condition $x(0) = x_0$ and continuous f, where the origin is an equilibrium, i.e. $f(0) = 0$. The origin is said to be *globally asymptotically stable* in the sense of Lyapunov provided that

***Stability*:** For every $\varepsilon > 0$ there is a $\delta > 0$ such that $\|x_0\| < \delta$ implies that $\|x(t)\| < \varepsilon$ for all $t \geq 0$ and for all $x(\cdot)$ with $x(0) = x_0$.

***Attractivity*:** There is an $\eta > 0$ such that for all $\|x_0\| < \eta$ it holds for all $x(\cdot)$ with $x(0) = x_0$ that $\lim_{t\to\infty} \|x(t)\| = 0$ and η can be taken as large as desired.

Another way to describe stability is by means of a Lyapunov function and comparison functions. A function $f : \mathbb{R}_+ \to \mathbb{R}_+$ is of class $\mathcal{K}$ if f is continuous, strictly increasing, and satisfies $f(0) = 0$. A real-valued map $V : \mathbb{R}^n \to \mathbb{R}$ is called a global strict *Lyapunov function* for (3.44) if

(i) it is *positive definite* and *proper*[1], i.e. there exist class $\mathcal{K}$ functions

[1] Properness can also be defined by the fact that the sublevel sets $\{x \in \mathbb{R}^n : V(x) \leq c\}$ are bounded for all $c > 0$ [26].

a, b such that for all $x \in \mathbb{R}^n$ we have that

$$a(\|x\|) \leq V(x) \leq b(\|x\|), \tag{3.45}$$

(ii) there exists a class $\mathcal{K}$ function w such that for every solution $x(\cdot)$ and each interval $I \subset [0, \infty)$ we have that

$$V(x(t)) - V(x(s)) \leq -\int_s^t w(\|x(r)\|)\,\mathrm{d}r \tag{3.46}$$

for each $s < t \in I$ provided that $x(\cdot)$ is defined on I.

Remark 3.6.1 *It is well known that the origin is globally asymptotically stable if and only if there is a global strict Lyapunov function [7].*

From that perspective the definition of stability for fluid networks seems to deviate from the asymptotic stability in the Lyapunov sense. The following lemma, however, shows that the definitions of stability are in fact equivalent.

Lemma 3.6.2 *A fluid network is stable if and only if the zero fluid level process is globally asymptotically stable in the sense of Lyapunov.*

Proof. Suppose that the fluid network is stable. Then there is a $\tau < \infty$ such that $Q(t) = 0$ for all $t \geq \tau\|Q(0)\|$ and for all fluid level processes $Q(\cdot)$. This implies $\lim_{t\to\infty} \|Q(t)\| = 0$ and attractivity holds true. To conclude stability let $\varepsilon > 0$ and $\delta := \frac{\varepsilon}{L\tau}$. Let $Q(\cdot)$ be a fluid level process with $\|Q(0)\| < \delta$. Then, by assumption it holds that $Q(t) = 0$ for all $t \geq \tau\,\|Q(0)\|$. Further, by the Lipschitz continuity of $Q(\cdot)$ we have that

$$\|Q(t)\| = \|Q(\tau\|Q(0)\|) - Q(t)\| \leq L|\tau\|Q(0)\| - t| \leq L\tau\|Q(0)\|$$

for all $t \in [0, \tau\|Q(0)\|]$. Moreover, as $Q(\tau\|Q(0)\| + \cdot) \equiv 0$ the latter estimate holds for all $t \geq 0$. Thus, we have $\|Q(t)\| \leq L\tau\|Q(0)\| < \varepsilon$.

Conversely, let $Q(\cdot) \equiv 0$ be asymptotically stable in the sense of Lyapunov. Due to the scaling property it suffices to consider a fluid level process $Q(\cdot)$ with $\|Q(0)\| = 1$. Then, by attractivity it holds that $\|Q(t)\| \to 0$ as $t \to \infty$ for any fluid level process $Q(\cdot)$. Hence, $\inf\{\|Q(t)\| : t \geq 0\} = 0$ for any $Q(\cdot)$ with $\|Q(0)\| = 1$. The assertion then follows from the proof of Theorem 3.1.2 starting with (3.9). ∎

In view of Theorem 3.1.4 and the definition of Lyapunov functions in dynamical systems theory we now provide a precise definition for a Lyapunov function for the class of fluid networks.

Definition 3.6.3 *Given a fluid network, a positive definite and proper function* $V : \mathbb{R}_+^K \to \mathbb{R}$ *is called a Lyapunov function if there is a function* $w \in \mathcal{K}$ *such that*

$$V(Q(t)) - V(Q(s)) \leq -\int_s^t w(||Q(r)||) \, \mathrm{d}r \tag{3.47}$$

for all $0 \leq s \leq t$ *and all fluid level processes* $Q(\cdot)$.

The function V considered in Theorem 3.1.4 is assumed to be locally Lipschitz, positive definite, and to satisfy the uniform decrease condition

$$\dot{V}(Q(t)) \leq -\varepsilon \tag{3.48}$$

if $Q(t) \neq 0$ and t regular for $s \mapsto V(Q(s))$. We note that if V is locally Lipschitz the decrease condition in Definition 3.6.3 also reads as

$$\dot{V}(Q(t)) \leq -w(\|Q(t)\|), \tag{3.49}$$

where the decrease depends on the fluid level at the current time instant, rather than being uniform. In any case the Lyapunov function V in Theorem 3.1.4 is not forced to be proper. In the Lyapunov theory for dynamical systems this condition is essential for an equilibrium to be stable as this enforces the boundedness of the sublevel sets of the Lyapunov function. This in turn guarantees that the trajectories, along which the Lyapunov function decreases, will converge towards the equilibrium and not diverge. However, due to the scaling property this behavior is ruled out for fluid networks by the request that the decrease of the Lyapunov function V is uniform as described in Theorem 3.1.4.

3.7 Notes and References

Based on the fact that the stability of a multiclass queueing network is implied by the stability of the associated fluid network, starting from 1995, much effort has been spent to derive conditions that guarantee stability of fluid networks under various disciplines.

General-work-conserving fluid networks were investigated in [19]. The stability conditions for priority fluid networks are taken from [24] and [22]. The fact that HLPPS fluid networks are stable if and only if $\rho < e$ was shown by Bramson [16]. His proof does not rely on a state dependent Lyapunov function. Instead, an entropy type functional was used that depends on the trajectories of the departure process $M\,T(\cdot)$ and the immediate workload process $W(\cdot)$. Here we provide a investigation of the result from a Lyapunov perspective. It turns out that as long as all stations in the network have positive total fluid level the decrease condition (3.12) is satisfied. Unfortunately, the case, where at least one station has total fluid level equal to zero has shown some resilience towards attempts of proof. Moreover, a related result to Lemma 3.4.1 has been shown by Bramson, cf. [16, Lemma 4.2 (b)]. The stability conditions for FIFO fluid networks are stated without providing a proof, since the proof is not based on Lyapunov arguments. The interested reader is referred to [23]. We will show in Chapter 5 that FIFO fluid networks take a special position among the fluid networks considered in this thesis.

The equivalence of condition (3.5) and the stability of fluid networks was obtained by Stolyar, see [81]. In practice this characterization is not easy to check for particular fluid networks. However, from a theoretical point of view condition (3.5) is very useful to see that the stability notions of fluid networks and asymptotic stability in the sense of Lyapunov are in fact equivalent.

Moreover, this reinforces the pursuit of the development of a Lyapunov theory for fluid networks. In addition, this is strengthened by the efficiency of the stability condition stated in Theorem 3.1.4. At this point we note that, to the best of the author's knowledge, Lyapunov arguments based on Theorem 3.1.4 have only been used to verify proposed stability conditions.

In the theory of dynamical systems a lot of effort has been spent to conclude that the converse also holds true. The content of the subsequent chapters is to consider the question whether a converse of Theorem 3.1.4 in terms of Definition 3.6.3 is valid. That is, the fundamental question of the next chapter is: Does a stable fluid network necessarily admit a Lyapunov function?

4 Converse Lyapunov Theorems for Fluid Networks

This chapter is devoted to the question whether the existence of a Lyapunov function is also necessary for the stability of a fluid network. The approach to conclude a commonly known converse Lyapunov theorem differs from Chapter 3. There we used piecewise linear or quadratic Lyapunov functions to prove stability of fluid networks under a particular discipline. This chapter is based on the characteristic properties of fluid networks discussed at the end of Chapter 2, namely the scaling and shift property, the Lipschitz continuity of the processes, and the fact the set of fluid level processes is closed under the topology of uniform convergence on compact sets. We will define a generic fluid model as a set of processes having exactly these properties. This approach captures in essence the behavioral approach to mathematical systems theory proposed by J. C. Willems, see [84].

The first work in this direction was done by Ye and Chen, who introduced the notion generic fluid network (GFN) model. The class of generic fluid networks models covers a wide variety of fluid network models. Within this framework Ye and Chen propose a Lyapunov method for stability. In [86] they proved that stability of a GFN model is equivalent to the existence of a functional on the set of trajectories that is decaying along trajectories. This result falls short of a converse Lyapunov theorem in that no state dependent Lyapunov function is constructed. Rather in principle the whole solution set has to be known in order to even define a Lyapunov functional. The strength and basis of applicability of the classic second method of Lyapunov, however, is that it can be checked without the knowledge of solutions, whether a given state-dependent function is indeed a Lyapunov function.

In this chapter we construct state-dependent Lyapunov functions in contrast to trajectory-wise functionals. We first show by counterexamples that closed GFN models do not provide sufficient information that allow for a converse Lyapunov theorem. In this sense the class of GFN models is too wide. To resolve this problem we introduce the class of strict GFN models by forcing the closed GFN model to satisfy a concatenation and a semicontinuity condition. We show that for the class of strict GFN models a converse Lyapunov theorem holds. As in other converse theorems the proof is not constructive, so that the question of finding a Lyapunov function for a particular case is not solved by this result.

Section 4.3 is devoted to the construction of smooth Lyapunov functions. That is, we start from a stable strict GFN model and, thus, there exists a continuous Lyapunov function, denoted by V. To obtain a smooth Lyapunov function we consider the Lyapunov function V and its convolution with mollifiers. Our approach to obtain a smooth Lyapunov function is based on the derivation of smooth Lyapunov functions in the dynamical systems literature. However, the absence of a differential equation or a differential inclusion brings along the disadvantage that a Gronwall-like argument is not available that provides an estimate for the evolution of the difference of two trajectories. To overcome this problem we will present a condition, in terms of strict GFN models, which assures that the constructions works.

4.1 Generic Fluid Models

In this section we start to change our perspective in the investigation of fluid networks to a more abstract point of view. In doing so, we aim to analyze the stability independently of the service discipline. We recall from the very last section of Chapter 2 that fluid networks share certain properties regardless of the particular discipline, see Propositions 2.4.13, 2.4.14, 2.4.15. We will define generic fluid network (GFN) models as a set of trajectories having exactly this properties. Further, we exhibit the idea of the behavioral approach to look at dynamical systems.

Moreover, we will present a Lyapunov method for characterizing the stability of GFN models. By counterexamples we will explain why GFN models are too general to provide a state-dependent Lyapunov theory. To define GFN models recall that the scaling operator σ_r and the shift operator δ_s are defined by $\sigma_r\, Q(t) = \frac{1}{r} Q(rt)$ and $\delta_s\, Q(t) = Q(t+s)$, respectively.

Definition 4.1.1 *A nonempty set* Φ *of functions* $Q\colon \mathbb{R}_+ \to \mathbb{R}_+^K$ *is said to be a GFN model, if*

(*a*) *there is a* $L > 0$*, such that for any* $Q \in \Phi$ *and* $t, s \in \mathbb{R}_+$ *it holds that*

$$\|Q(t) - Q(s)\| \leq L\,|t - s|.$$

(*b*) $Q \in \Phi$ *implies* $\sigma_r\, Q \in \Phi$ *for all* $r > 0$.

(*c*) $Q \in \Phi$ *implies* $\delta_s\, Q \in \Phi$ *for all* $s \geq 0$.

If the following condition holds in addition, then we call Φ *a closed GFN model*

(*d*) *if a sequence* $(Q_n)_{n\in\mathbb{N}}$ *in* Φ *converges to* Q_* *u.o.c., then* $Q_* \in \Phi$.

The conditions defining a closed GFN model are in line with the Propositions 2.4.13, 2.4.14, and 2.4.15. At this point we note that this approach captures the essence of the behavioral approach to dynamical systems proposed by J. C. Willems, where a dynamical system is thought of as a triple $\Sigma = (\mathbb{T}, \mathbb{W}, \mathbb{B})$ such that

(i) $\mathbb{T}$ denotes the time axis, e.g. $\emptyset \neq \mathbb{T} \subset \mathbb{R}$

(ii) $\mathbb{W}$ denotes the space of signal values, e.g. $\emptyset \neq \mathbb{W} \subset \mathbb{R}^n$

(iii) $\mathbb{B} = \{w : \mathbb{T} \to \mathbb{W}\}$ denotes the *behavior* consisting of the set of trajectories $w(\cdot)$ that are feasible to describe the evolution of the system.

Interpreting the time axis is $\mathbb{T} = \mathbb{R}_+$, the set of possible outcomes $\mathbb{W} = \mathbb{R}_+^K$ as the set of all possible fluid levels, and the behavior $\mathbb{B} = \Phi$, we see that the class of GFN models falls into this framework. In particular, the shift property is in one to one correspondence to the notion of *time-invariance* of a dynamical system $\Sigma = (\mathbb{T}, \mathbb{W}, \mathbb{B})$ which is defined by the property that if $w(\cdot) \in \mathbb{B}$ and $s \in \mathbb{T}$, then the shift $\delta_s w\,(\cdot) = w(s + \cdot) \in \mathbb{B}$, cf. [84, Definition II.3]. Since the focus of this chapter is on stability we quote from [85, Definition 11] the concept of stability in the behavioral framework.

Definition 4.1.2 *A dynamical system* $\Sigma = (\mathbb{R}, \mathbb{W}, \mathbb{B})$ *is called* stable*, if for any* $w(\cdot) \in \mathbb{B}$ *it holds that* $\lim_{t\to\infty} w(t) = 0$.

Hence, a fluid network can be written as a system $\Sigma = (\mathbb{R}_+, \mathbb{R}^K_+, \Phi)$. However, we will stick to the notation Φ for simplicity. Moreover, for future use we also introduce some notation. Any element $Q \in \Phi$ is called a trajectory (of Φ) and the set of trajectories with total initial level one is denoted by $\Phi(1) := \{Q \in \Phi : \|Q(0)\| = 1\}$. For $x \in \mathbb{R}^K_+$ the set of trajectories starting in x is denoted by $\Phi_x := \{Q \in \Phi : Q(0) = x\}$. The definition of stability for GFN models is analog to the definition of stability for fluid networks except that, due to scaling property, it is sufficient to consider trajectories whose initial values have norm one.

Definition 4.1.3 *A GFN model Φ is said to be stable if there exists a $\tau > 0$ such that $\delta_\tau Q \equiv 0$ for all trajectories $Q \in \Phi(1)$.*

The notion of stability of a GFN may also be expressed by saying that the zero fluid level process $Q_0 \equiv 0$ is the unique stable and attractive fixed point of the shift operator $\delta_\tau Q(\cdot) = Q(\cdot + \tau)$ defined on Φ. Following the proof of Lemma 3.6.2 it can be seen that the notion of stability for a GFN model Φ is equivalent to the notion of stability for a system $\Sigma = (\mathbb{R}_+, \mathbb{R}^K_+, \Phi)$.

Remark 4.1.4 *For a stable GFN model Φ with Lipschitz constant L and $\tau > 0$ it holds that $L\tau \geq 1$, as for any $Q \in \Phi(1)$ we have that*

$$1 = \|Q(0)\| = \|Q(\tau) - Q(0)\| \leq L\,\tau.$$

To begin with the stability analysis of GFN models from a Lyapunov perspective we recall the known results from the literature. Ye and Chen presented a Lyapunov method to characterize stability of closed GFN models, where the *L-condition* plays a key role, which is defined as follows.

Definition 4.1.5 *A GFN model Φ is said to satisfy the L-condition if there exist class $\mathcal{K}$ functions w_i, $i = 1, 2, 3$, such that for any trajectory $Q \in \Phi$ there exists an absolutely continuous function $v : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying*

$$w_1(\|Q(t)\|) \leq v(t) \leq w_2(\|Q(t)\|), \tag{4.1}$$

$$\dot{v}(t) \leq -w_3(\|Q(t)\|) \tag{4.2}$$

for almost all $t \geq 0$.

The content of the subsequent statement is that the L-condition is equivalent to the stability of a GFN model. This result was derived by Ye and Chen. We do not recall its proof here. The interested reader is referred to [86].

Theorem 4.1.6 ([86]) *A GFN model Φ is stable if and only if the L-condition is satisfied. In particular, given $Q \in \Phi$ the function v can be chosen as*

$$v(t) := \int_t^\infty \|Q(s)\| \, \mathrm{d}s. \tag{4.3}$$

Given a trajectory Q of a stable closed GFN model, $v(t)$ may be interpreted as the total remaining fluid mass of Q from time t on. We note that an equivalent way of interpreting v is as a functional $\bar{v} : \Phi \to \mathbb{R}_+$ on the GFN model with the following properties. There are comparison function such that for each trajectory $Q \in \Phi$ its value under the functional $\bar{v}$ can be estimated from below and above by its initial value. That is, for any $Q \in \Phi$ it holds that

$$w_1(\|Q(0)\|) \leq \bar{v}(Q) \leq w_2(\|Q(0)\|).$$

Furthermore, the evolution of $\bar{v}(Q)$ can also be estimated in terms of a comparison function. Precisely, the mapping $t \mapsto \bar{v}(Q)(t + \cdot)$ satisfies

$$\frac{d}{\mathrm{d}t}\bar{v}(Q)(t + \cdot) \leq -w_3(\|Q(t)\|).$$

For this reason we refer to v, interpreted as $v(0) =: \bar{v}(Q)$, as a Lyapunov functional on the set of trajectories. It can be seen that this approach differs from the one taken in the theory of dynamical systems in which Lyapunov functions are state-dependent. The dependence on solutions is undesirable, because the benefit of Lyapunov's second method is that trajectories need not be known to be able to determine stability, whereas the Theorem 4.1.6 requires the knowledge of all solutions. In the following we present a way to overcome this drawback. To this end, we define a Lyapunov function that does not depend on the trajectory of the closed GFN model, which is in line with Definition 3.6.3.

Definition 4.1.7 *Given a GFN model Φ, a function $V : \mathbb{R}_+^K \to \mathbb{R}_+$ is said to be a Lyapunov function if there exist class $\mathcal{K}$ functions w_i, $i = 1, 2, 3$*

such that

$$w_1(\|x\|) \leq V(x) \leq w_2(\|x\|), \tag{4.4}$$

$$V(Q(t)) - V(Q(s)) \leq -\int_s^t w_3(\|Q(r)\|)\,\mathrm{d}r \tag{4.5}$$

for all $0 \leq s \leq t \in \mathbb{R}_+$ *and all trajectories* $Q \in \Phi$.

We denote by $\mathcal{A}(\Phi) = \{x \in \mathbb{R}_+^K : \exists\, Q \in \Phi,\, Q(0) = x\}$ and consider for our purpose the following candidate $V : \mathcal{A}(\Phi) \to \mathbb{R}_+ \cup \{\infty\}$ defined by

$$V(x) = \sup_{Q \in \Phi_x} \int_0^\infty \|Q(s)\|\,\mathrm{d}s. \tag{4.6}$$

In the sequel, we assume that $\mathcal{A}(\Phi) = \mathbb{R}_+^K$. The function V can be interpreted as a measurement of the state x in the sense that $V(x)$ represents the total possible fluid mass that the network has to deal with being in state x.

Remark 4.1.8 *The Lyapunov function candidate has the following radial property: For any scalar* $r > 0$ *it holds that*

$$V(rx) = r^2\, V(x).$$

To see this, let $r > 0$ *be a scalar. Then, the scaling property implies that*

$$\begin{aligned} V(rx) &= \sup_{Q \in \mathcal{Q}_{rx}} \int_0^\infty \|Q(s)\|\,\mathrm{d}s = \sup_{Q \in \mathcal{Q}_x} \int_0^\infty \|\, r\, Q(r^{-1}\, s)\|\,\mathrm{d}s \\ &= r \sup_{Q \in \mathcal{Q}_x} \int_0^\infty \|Q(u)\|\, r\ \mathrm{d}u = r^2 \sup_{Q \in \mathcal{Q}_x} \int_0^\infty \|Q(u)\|\ \mathrm{d}u = r^2\, V(x). \end{aligned}$$

An interesting question concerns the regularity of V. Of course, we aim for continuous dependence on the state, as this would entail robustness of stability, see [55], [82]. Note that for stable closed GFN models the supremum in (4.6) is actually attained because of the closedness condition (d) in Definition 4.1.1. We say that a function $g : \mathbb{R}_+^K \to \mathbb{R}$ is upper semicontinuous in $a \in \mathbb{R}_+^K$ if $g(a) \geq \limsup_{x \to a} g(x)$. Certainly, g is called upper semicontinuous if it is upper semicontinuous for every $a \in \mathbb{R}_+^K$. Further, a function $g : \mathbb{R}_+^K \to \mathbb{R}_+$ is lower semicontinuous at $a \in \mathbb{R}_+^K$ if $-g$ is upper semicontinuous in $a \in \mathbb{R}_+^K$ and g is called lower semicontinuous if g is lower semicontinuous in every point.

Proposition 4.1.9 *If Φ is a stable closed GFN model, then V given by* (4.6) *is well defined and upper semicontinuous.*

Proof. It is an easy consequence of the Lipschitz continuity and stability that V as defined in (4.6) is finite. Let $x \in \mathbb{R}_+^K$ and $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}_+^K$ which converges to x. As Φ is stable the set $\{V(x_n) : n \in \mathbb{N}\}$ is bounded. Hence, there exists a subsequence $(x_{n_l})_{l\in\mathbb{N}}$ and a sequence $(Q_{n_l})_{l\in\mathbb{N}}$ with $Q_{n_l} \in \Phi_{x_{n_l}}$ such that

$$\limsup_{n\to\infty} V(x_n) = \lim_{l\to\infty} V(x_{n_l}) = \lim_{l\to\infty} \int_0^\infty \|Q_{n_l}(s)\| \,\mathrm{d}s.$$

Now we consider the family $\{Q_{n_l}(\cdot) : l \in \mathbb{N}\}$. Since Φ is stable the family $\{Q_{n_l}(\cdot) : l \in \mathbb{N}\}$ is bounded. By the Lipschitz condition (a) in Definition 4.1.1 there is a single Lipschitz constant for any trajectory $Q_{n_l}(\cdot)$ of the family $\{Q_{n_l}(\cdot) : l \in \mathbb{N}\}$ and so the family is equicontinuous. By the theorem of Arzelà-Ascoli there exists a subsequence which converges u.o.c. to some Q_* with $Q_*(0) = x$. Since the model is closed it follows that $Q_* \in \Phi$. Thus, by the definition of V it holds that

$$\limsup_{n\to\infty} V(x_n) = \lim_{l\to\infty} \int_0^\infty \|Q_{n_l}(s)\| \,\mathrm{d}s = \int_0^\infty \|Q_*(s)\| \,\mathrm{d}s \le V(x).$$

This shows the assertion. ∎

As we are interested in the continuity of V the question remains whether V is also lower semicontinuous. The subsequent example indicates that in general this is not the case.

Example 4.1.10 *Let $K = 2$ and consider the GFN model Φ defined by*

$$\Phi = \left\{ \begin{pmatrix} (x_1 - t)^+ \\ (x_2 - t)^+ \end{pmatrix}, \begin{pmatrix} (c - \frac{1}{2}t)^+ \\ (c - \frac{1}{2}t)^+ \end{pmatrix} : x_1, x_2, c \in \mathbb{R}_+ \right\}.$$

It is easy to check that Φ is a stable closed GFN model. We consider $x_0 = (1\ 1)^\mathsf{T}$ and $x_n = (1 + \frac{1}{n}\ 1 - \frac{1}{n})^\mathsf{T}$. It holds that

$$\begin{aligned} \lim_{n\to\infty} V(x_n) &= \lim_{n\to\infty} \int_0^\infty (1 + \tfrac{1}{n} - t)^+ + (1 - \tfrac{1}{n} - t)^+ dt \\ &= \lim_{n\to\infty} \tfrac{1}{2} \left((1 + \tfrac{1}{n})^2 + (1 - \tfrac{1}{n})^2 \right) = 1 \\ &< 2 = \int_0^2 2(1 - \tfrac{1}{2}t) \,\mathrm{d}t = V(x_0). \end{aligned}$$

So, V defined by (4.6) is not necessarily lower semicontinuous for stable closed GFN models.

Remark 4.1.11 *The example shows that in the frame of Definition 4.1.1 our candidate V is not continuous in general. The problem with this example is that along the diagonal a particular trajectory exists which is not approximated by solutions starting close to but not on the diagonal.*

The key property of a Lyapunov function V for a dynamical system is that it is decreasing along trajectories. The next example addresses this issue for closed GFN models.

Example 4.1.12 *Let $K = 2$ and define for given $x_1, x_2 \in \mathbb{R}_+$ the trajectories*

$$Q_1(t) = \begin{cases} \begin{pmatrix} x_1 - t \\ x_2 + t \end{pmatrix} & \text{if} \quad t \leq x_1, \\ \begin{pmatrix} 0 \\ 2x_1 + x_2 - t \end{pmatrix}^+ & \text{else,} \end{cases}$$

and

$$Q_2(t) = \begin{cases} \begin{pmatrix} x_1 + t \\ x_2 - t \end{pmatrix} & \text{if} \quad t \leq x_2, \\ \begin{pmatrix} x_1 + 2x_2 - t \\ 0 \end{pmatrix}^+ & \text{else.} \end{cases}$$

Then, we consider the closed GFN model Φ defined by

$$\Phi = \{Q_1(\cdot), Q_2(\cdot) \,:\, x_1, x_2 \in \mathbb{R}_+\}\,.$$

Clearly, Φ is stable. In Figure 4.1 a schematic illustration of the feasible trajectories is provided. In this GFN model it is obvious that trajectories cannot be concatenated. However, let us assume that there is a state-dependent Lyapunov function V which is decaying along trajectories. The closed GFN model Φ has the following property. For every state $z = (z_1, z_2)$ there is a state $y = (y_1, y_2)$ such that there are two trajectories that go to zero, where one trajectory starts in z and passes y and the other trajectory starts in y and passes z, see Figure 4.1. As V is decaying along trajectories it follows that $V(z) < V(y)$ and $V(y) < V(z)$, which is a contradiction.

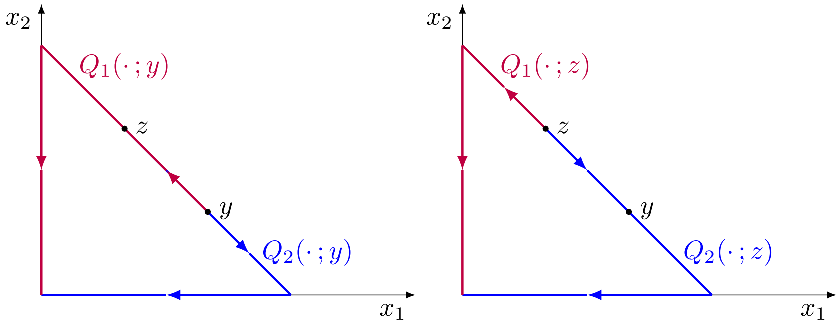

Figure 4.1: This figure illustrates the trajectories $Q_i(\cdot\,; z)$ and $Q_i(\cdot\,; y)$, $i = 1, 2$ of the GFN model Φ for the initial values $y \in \mathbb{R}^2_+$ and $z \in \mathbb{R}^2_+$, respectively. The components of the initial values y and z satisfy $y_k > 0$ and $z_k > 0$, respectively.

Remark 4.1.13 *The vital issue of Example 4.1.12 is that for (closed) GFN models a Lyapunov function in the sense of Definition 4.1.1 cannot exist. Moreover, the crucial aspect of the previous examples resembles a known phenomenon of fluid networks under FIFO disciplines: To determine the evolution a FIFO trajectory after a starting point $Q(0)$ it is necessary to fix additionally the initial direction in terms of the initial allocation $\{T(s) : s \in [0, W(0)]\}$, see also* (3.43) *and Example 5.2.15.*

4.2 Strict Generic Fluid Models

In this section we present a way out of the dilemma. We specify the class of closed GFN models by adding two conditions, namely a concatenation property and a lower semicontinuity property and denote them strict GFN models. The concatenation of trajectories is defined as follows.

Definition 4.2.1 *Let Φ be a closed GFN model. Suppose that Q_1 and Q_2 are trajectories of Φ such that $Q_1(t^*) = Q_2(0)$ for some $t^* \geq 0$. Then, $Q_1 \diamond_{t^*} Q_2$ is called the concatenation of Q_1 and Q_2 at t^*, which is defined by*

$$Q_1 \diamond_{t^*} Q_2(t) := \begin{cases} Q_1(t) & t \leq t^*, \\ Q_2(t - t^*) & t \geq t^*. \end{cases}$$

In the behavioral approach to dynamical systems the concatenation of trajectories plays a key role for the understanding of *state*. Transferred to the present setting a GFN model Φ has the *property of state* if and only if it satisfies the concatenation property. That is, all the information that is needed to decide if two trajectories $Q_1, Q_2 \in \Phi$ can be concatenated within Φ at time t is the knowledge whether $Q_1(t)$ and $Q_2(t)$ coincide or not, see e.g. [66, Definition 4.3.3 and Remark 4.3.4]. Similar to GFN models we denote $\mathcal{Q}(1) := \{Q \in \mathcal{Q} : \|Q(0)\| = 1\}$. Moreover, let $\mathcal{Q}_x := \{Q \in \mathcal{Q} : Q(0) = x\}$ denote the set of trajectories in $\mathcal{Q}$ starting in $x \in \mathbb{R}^n_+$.

To overcome the difficulties from the previous section, we consider the closed GFN models with additional properties, which are defined as follows.

Definition 4.2.2 *A subset $\mathcal{Q} \subset \Phi$ is said to be a strict GFN model if*

(e) *it satisfies the concatenation property,*

(f) *there is a $T > 0$ such that the set-valued map $x \rightsquigarrow \mathcal{Q}_x\big|_{[0,T]}$ is lower semicontinuous.*

A definition of lower semicontinuous set-valued maps is given in Appendix B. It is possible that a closed GFN model satisfies (e) and not (f). We do not introduce yet another name for such GFN models but simply speak of a closed GFN model satisfying (e).

Remark 4.2.3 *Given a closed GFN model, the uniform Lipschitz continuity condition* (a) *implies that if a sequence $(Q_n)_{n\in\mathbb{N}}$ in $\mathcal{Q}$ converges uniformly on $[0, T)$ with $T < \infty$ then it converges uniformly on the closed interval $[0, T]$.*

We note for further reference, that for closed GFN models satisfying (e) the semicontinuity condition can be stated the following.

Proposition 4.2.4 *Given a closed GFN model satisfying* (e), *then the semicontinuity condition* (f) *holds if and only if*

(f') *for each x_0 so that $\mathcal{Q}_{x_0} \neq \emptyset$ there exists a $T(x_0) > 0$ such that the set-valued map $x \rightsquigarrow \mathcal{Q}_x\big|_{[0,T(x_0)]}$ is lower semicontinuous at x_0.*

Proof. It is clear that (f) implies (f'). Conversely, fix any $T > 0$, x_0 and a $T_0 := T(x_0)$ such that (f') holds. Choose $Q \in \mathcal{Q}_{x_0}$ and a sequence

$x_n \to x_0$. We have to construct a sequence $(Q_n)_{n\in\mathbb{N}}$ with $Q_n \in \mathcal{Q}_{x_n}$ so that $Q_n(\cdot) \to Q(\cdot)$ uniformly on $[0,T]$. We may assume that $T_0 < T$ as otherwise there is nothing to show.

By assumption (f') there exist $Q_n^0 \in \mathcal{Q}_{x_n}$ such that $Q_n^0(\cdot) \to Q(\cdot)$ uniformly on $[0,T_0]$. In particular, $Q_n^0(T_0) \to Q(T_0)$. By the shift property it holds that $\delta_{T_0} Q \in \mathcal{Q}_{Q(T_0)}$ and so for $T_1 := T(Q(T_0))$ we may by (f') choose a sequence $Q_n^1 \in \mathcal{Q}_{Q_n^0(T_0)}$ such that $Q_n^1(\cdot) \to \delta_{T_0} Q(\cdot)$ uniformly on $[0,T_1]$. Now define the concatenation $\bar{Q}_n^1 := Q_n^0 \diamond_{T_0} Q_n^1$ and note that $\bar{Q}_n^1(\cdot) \to Q(\cdot)$ uniformly on $[0, T_0 + T_1]$.

Repeating this step countably often, we can construct an open interval $[0,\bar{T})$ such that there exist $\bar{Q}_n \in \mathcal{Q}_{x_n}$ satisfying $\bar{Q}_n(\cdot) \to Q(\cdot)$ u.o.c. on $[0,\bar{T})$. Assume that $\bar{T} < \infty$ is chosen as the maximal real for which this u.o.c. convergence is possible.

Then, by Remark 4.2.3 it follows that $\bar{Q}_n(\cdot) \to Q(\cdot)$ uniformly on $[0,\bar{T}]$. Hence, we can repeat the argument and extend the uniform convergence to the interval $[0, \bar{T} + T(Q(\bar{T}))]$. This contradicts the assumption that $\bar{T}$ was chosen to be maximal. This shows the equivalence, as $\bar{T}$ can be arbitrarily large and so chosen to be bigger than T. ■

Moreover, the following proposition is also a consequence of the conditions (e) and (f).

Proposition 4.2.5 *The trajectories of a strict GFN model depend lower semicontinuously on the initial value, i.e.*

- (g) *for each x, $Q \in \mathcal{Q}_x$ and $(x_n)_{n\in\mathbb{N}}$ converging to x, there is a sequence of trajectories $(Q_n)_{n\in\mathbb{N}}$ with $Q_n \in \mathcal{Q}_{x_n}$ which converges to Q u.o.c.*

Proof. Let $x \in \mathbb{R}_+^K$, $Q \in \mathcal{Q}_x$ and $(x_n)_n$ be a sequence converging to x. By condition (f) in Definition 4.2.2 there exists a $T > 0$ and a sequence $(Q_n)_{n\in\mathbb{N}}$ with $Q_n \in \mathcal{Q}_{x_n}$ that converges to $Q \in \mathcal{Q}$ uniformly on $[0,T]$. In particular, we have $Q_n(T)$ converges to $Q(T)$ as $n \to \infty$. Moreover, for $Q^1 \in \mathcal{Q}_{Q(T)}$ such that $Q^1(\cdot)|_{[0,T]} = Q(\cdot)|_{[T,2T]}$ condition (f) implies the existence of a sequence $Q_n^1 \in \mathcal{Q}_{Q_n(T)}$ satisfying $\lim_{n\to\infty} Q_n^1 = Q^1$ uniformly on $[0,T]$.

By the concatenation property (e) we have a sequence $(Q_n)_{n\in\mathbb{N}}$ with $Q_n \in \mathcal{Q}_{x_n}$ that converges uniformly to $Q \in \mathcal{Q}_x$ on $[0,2T]$. A successive continuation in this manner yields the existence of a sequence $(Q_n)_{n\in\mathbb{N}}$ with $Q_n \in \mathcal{Q}_{x_n}$ converging uniformly on compact sets to $Q_* \in \mathcal{Q}_{x_*}$. ■

Next, we show that condition (f) is the appropriate condition for continuity of our candidate V. To be precise, condition (g) closes the gap from upper semicontinuity to continuity.

Proposition 4.2.6 *If $\mathcal{Q}$ is a stable strict GFN model, then V defined in (4.6) is continuous.*

Proof. We show that V is lower semicontinuous as the continuity of V then follows together with Proposition 4.1.9. Let $x \in \mathbb{R}_+^K$ and $Q \in \mathcal{Q}_x$ be such that

$$V(x) = \int_0^\infty \|Q(s)\| \,\mathrm{d}s.$$

Further, let $(x_n)_n$ be a sequence converging to x. As $\mathcal{Q}$ is stable and using the same arguments as in the proof of Proposition 4.1.9,

$$V(x) = \int_0^\infty \|Q(s)\| \,\mathrm{d}s = \lim_{n\to\infty} \int_0^\infty \|Q_n(s)\| \,\mathrm{d}s \leq \liminf_{n\to\infty} V(x_n).$$

That is, V is lower semicontinuous. ∎

Next, we show that for the class of strict GFN models, in fact, a converse Lyapunov theorem holds.

Theorem 4.2.7 *A strict GFN model $\mathcal{Q}$ is stable if and only if it admits a Lyapunov function. In particular, V can be chosen as*

$$V(x) = \sup_{Q\in\mathcal{Q}_x} \int_0^\infty \|Q(s)\| \,\mathrm{d}s$$

and V is continuous.

Proof. First, we show that the existence of a Lyapunov function is sufficient for stability. Let V be a Lyapunov function for $\mathcal{Q}$. From (4.4) it follows that $V(Q(t)) \geq 0$ and inequality (4.5) implies that

$$V(Q(t_2)) - V(Q(t_1)) \leq 0$$

for all $t_1 \leq t_2 \in \mathbb{R}_+$. So, $V(Q(\cdot))$ is monotone decreasing and bounded. In order to show that $V(Q(t))$ tends to zero as t goes to infinity assume that

$$\lim_{t\to\infty} V(Q(t)) =: c > 0.$$

Then, for all $t \geq 0$ it holds that

$$0 < c \leq V(Q(t)) \leq w_2(\|Q(t)\|) \tag{4.7}$$

and further $0 < w_2^{-1}(c) \leq \|Q(t)\|$. It also holds that

$$0 < w_3(w_2^{-1}(c)) \leq w_3(\|Q(t)\|).$$

Now observe that from (4.5) it follows that

$$\begin{aligned} V(Q(t)) - V(Q(0)) &\leq -\int_0^t w_3(\|Q(s)\|)\,\mathrm{d}s \leq -\int_0^t w_3(w_2^{-1}(c))\,\mathrm{d}s \\ &\leq -\,w_3(w_2^{-1}(c))\,t \end{aligned}$$

and, hence, $\lim_{t\to\infty} V(Q(t)) = -\infty$, which is a contradiction to (4.7). So,

$$\lim_{t\to\infty} V(Q(t)) = 0. \tag{4.8}$$

By (4.4) it follows that

$$\lim_{t\to\infty} \|Q(t)\| = 0.$$

Consequently, the origin is asymptotically stable and by Lemma 3.6.2 this implies the stability of the strict GFN model $\mathcal{Q}$.

Conversely, suppose that $\mathcal{Q}$ is stable. Then, there is a $\tau > 0$ such that $\delta_\tau Q \equiv 0$ for all trajectories $Q \in \mathcal{Q}(1)$. We define the following comparison functions

$$w_1(r) := \tfrac{r^2}{2L}, \quad w_2(r) := r^2\,(1 + L\tau)\,\tau, \quad w_3(r) := r$$

and show that our candidate

$$V(x) = \sup_{Q\in\mathcal{Q}_x} \int_0^\infty \|Q(s)\|\,\mathrm{d}s$$

is a Lyapunov function. As $\mathcal{Q}$ satisfies the Lipschitz condition (a) it follows that

$$\|Q(s)\| \geq \|Q(t)\| - L(s-t) \tag{4.9}$$

for all $Q \in \mathcal{Q}$ and $s \geq t$. In particular, for $t = 0$ this implies

$$\|Q(s)\| \geq \|Q(0)\| - Ls. \tag{4.10}$$

Using the last inequality we get the following estimate from below

$$
\begin{aligned}
V(x) = \sup_{Q\in\mathcal{Q}_x} \int_0^{\infty} \|Q(s)\| \; \mathrm{d}s &\geq \sup_{Q\in\mathcal{Q}_x} \int_0^{\frac{\|x\|}{L}} \|Q(s)\| \; \mathrm{d}s \\
&\geq \sup_{Q\in\mathcal{Q}_x} \int_0^{\frac{\|x\|}{L}} \left(\|x\| - Ls\right) \; \mathrm{d}s \\
&= \sup_{Q\in\mathcal{Q}_x} \left\{ \|x\| \frac{\|x\|}{L} - \frac{\|x\|^2}{2L} \right\} = \frac{\|x\|^2}{2L} = w_1(\|x\|).
\end{aligned}
$$

To obtain an estimate from above consider $Q \in \mathcal{Q}_x$. Note that by the scaling property it follows that $\sigma_{\|x\|} Q \in \mathcal{Q}(1)$ and further the stability of $\mathcal{Q}$ implies that

$$
Q(t) = 0 \qquad \forall\, t \geq \|x\|\tau. \tag{4.11}
$$

The triangle inequality together with the Lipschitz condition imply that for all $t \in [0, \|x\|\tau]$ it holds that

$$
\|Q(t)\| \leq \|Q(0)\| + Lt \leq \|x\| \,(1 + L\tau). \tag{4.12}
$$

With (4.11) and (4.12) an estimate from above is derived as follows

$$
\begin{aligned}
V(x) = \sup_{Q\in\mathcal{Q}_x} \int_0^{\|x\|\tau} \|Q(s)\| \; \mathrm{d}s &\leq \sup_{Q\in\mathcal{Q}_x} \int_0^{\|x\|\tau} \|x\| \,(1 + L\,\tau) \; \mathrm{d}s \\
&= \|x\|^2 \,(1 + L\tau)\, \tau = w_2(\|x\|).
\end{aligned}
$$

Now consider the decrease condition

$$
V(Q(t_2)) - V(Q(t_1)) = \sup_{Q\in\mathcal{Q}_{Q(t_2)}} \int_0^{\infty} \|Q(s)\| \; \mathrm{d}s - \sup_{Q\in\mathcal{Q}_{Q(t_1)}} \int_0^{\infty} \|Q(s)\| \; \mathrm{d}s.
$$

From condition (e) it follows that

$$
\begin{aligned}
V(Q(t_1)) &= \sup_{Q\in\mathcal{Q}_{Q(t_1)}} \int_0^{\infty} \|Q(s)\| \; \mathrm{d}s \\
&\geq \int_{t_1}^{t_2} \|Q(s)\| \; \mathrm{d}s + \sup_{Q\in\mathcal{Q}_{Q(t_2)}} \int_0^{\infty} \|Q(s)\| \; \mathrm{d}s \\
&= \int_{t_1}^{t_2} \|Q(s)\| \; \mathrm{d}s + V(Q(t_2)).
\end{aligned}
$$

and, hence,

$$V(Q(t_2)) - V(Q(t_1)) \leq -\int_{t_1}^{t_2} \|Q(s)\| \, \mathrm{d}s = -\int_{t_1}^{t_2} w_3(\|Q(s)\|) \, \mathrm{d}s.$$

Thus, together with Proposition 4.2.6 we see that V is a continuous Lyapunov function. ■

From the proof of the previous theorem we see that the semicontinuity property (f) is only needed to conclude continuity of V. Thus, we have also proved the following.

Corollary 4.2.8 *A closed GFN model Φ satisfying the concatenation property* (e) *is stable if and only if it admits a Lyapunov function. In particular, V can be chosen as in* (4.6) *and V is upper semicontinuous.*

In addition, the proof of Theorem 4.2.7 shows the following equivalence.

Corollary 4.2.9 *A strict GFN model $\mathcal{Q}$ is stable if and only if it admits a continuous Lyapunov function and the comparison functions are of class $\mathcal{K}_\infty$.*

Remark 4.2.10 *Continuity of Lyapunov functions is of interest because this would ensure robustness properties of the network subject to unknown parameters or external perturbations. Since upper semicontinuous Lyapunov function do not imply robustness statements the benefit of Lyapunov functions that are upper semicontinuous is restricted compared to continuous Lyapunov functions, see [55], [82].*

4.3 Construction of Smooth Lyapunov Functions

Another essential point in the Lyapunov theory for dynamical systems is the construction of smooth Lyapunov functions. In this section we regard this issue in the context of strict GFN models. To this end, we consider a strict GFN models $\mathcal{Q}$ which is supposed to be stable. According to Theorem 4.2.7 there is a continuous Lyapunov function V defined by

$$V(x) = \sup_{Q \in \mathcal{Q}_x} \int_0^\infty \|Q(s)\| \, \mathrm{d}s$$

and a comparison function w such that the decrease condition

$$V(Q(t)) - V(Q(s)) \leq -\int_s^t w(\|Q(r)\|)\,\mathrm{d}r \tag{4.13}$$

is satisfied. The key issue of this section is the following. Given a stable strict GFN model, we aim to construct a C^∞-smooth Lyapunov function V_s and a C^∞-smooth comparison function w_s such that for all $Q \in \mathcal{Q}$ we have

$$\dot{V}_s(Q(t)) := \lim_{h\to 0} \frac{V_s(Q(t+h)) - V_s(Q(t))}{h} \leq -w_s(\|Q(t)\|). \tag{4.14}$$

Our construction will be based on C^∞-smooth mollifiers and their convolution with the Lyapunov function V admitted by the stable strict GFN model. This technique is well-known in the theory of partial differential equations and for the construction of smooth Lyapunov functions for ordinary differential equations and differential inclusions, see [44] and [26, 82], respectively, and the references therein.

For $r > 0$ and $x \in \mathbb{R}^n$ let $B(x,r) := \{y \in \mathbb{R}^n : \|x-y\| \leq r\}$. A function $k \in C^\infty(\mathbb{R}^K, \mathbb{R}_+)$ is called a *mollifier* if $\operatorname{supp} k = B(0,1)$ and

$$\int_{\mathbb{R}^K} k(x)\,\mathrm{d}x = 1.$$

Furthermore, the support of a mollifier can be scaled in the following way. For $r > 0$ consider

$$k_r(x) := \tfrac{1}{r^n}\, k(r^{-1}x).$$

Then, it follows that $k_r \in C^\infty(\mathbb{R}^K, \mathbb{R}_+)$, $\operatorname{supp} k_r = B(0,r)$, and

$$\int_{\mathbb{R}^K} k_r(x)\,\mathrm{d}x = 1.$$

Moreover, to consider the convolution of a function $f \in C(\mathbb{R}^n, \mathbb{R})$ and a mollifier k_r, let U be an open subset of $\mathbb{R}^K$ and $U_r = \{x \in U : \operatorname{dist}(x, \partial U) > r\}$, where $\operatorname{dist}(x, A) = \inf\{\|x-a\| : a \in A\}$. Then, the *convolution*, denoted by $f_r : U_r \to \mathbb{R}$, is defined by

$$x \mapsto f_r(x) := f * k_r\,(x) = \int_{B(0,r)} f(x-y)\, k_r(y)\,\mathrm{d}y.$$

By standard convolution results it follows that $f_r \in C^\infty(U_r, \mathbb{R}_+)$, see for instance [44, Theorem 6 Appendix C.4]. Furthermore, if f is continuous in U, it holds that $f_r \to f$ uniformly on compact subsets of U as $r \to 0$.

In order to construct a C^∞-smooth Lyapunov function and a C^∞-smooth comparison function we extend V and w to $\mathbb{R}^K$ as an initial step. To this end, let $|\cdot|_{\text{vec}}$ denote the map that takes componentwise absolute values, i.e.

$$|x|_{\text{vec}} := (|x_1|, ..., |x_K|)^\mathsf{T}.$$

The extention V^e, defined by the composition

$$V^e(x) := V(|x|_{\text{vec}})$$

of the continuous functions V and $|\cdot|_{\text{vec}}$, is also continuous. Applying the latter to the comparison function yields

$$w^e(x) := w(\||x|_{\text{vec}}\|).$$

Before starting the construction of a C^∞-smooth Lyapunov function we make the following standing assumption for this section. In the following we often denote trajectories by $Q(\cdot; x)$ to emphasize the starting point x.

Assumption 4.3.1 *For any $x \in \mathbb{R}^K$, $Q \in \mathcal{Q}_{|x|_{vec}}$, $\varepsilon > 0$, and $T > 0$ there is a constant $c > 0$ such that if $\|x - y\| < \varepsilon$, there exists a trajectory $R \in \mathcal{Q}_{|x-y|_{vec}}$ such that for all $t \in [0, T]$ it holds that*

$$\| Q(t; |x|_{vec}) - y - R(t; |x - y|_{vec}) \| \leq c \|y\| t.$$

As a first consequence of the Assumption 4.3.1 we shall show that V^e is locally Lipschitz. To this end, we consider the Dini subderivative. Let U be an open subset of $\mathbb{R}^n$. The Dini subderivative of a function $f : U \to \mathbb{R}$ at a point $x \in U$ in the direction $v \in \mathbb{R}^n$ is defined by

$$Df(x; v) := \liminf_{\varepsilon \to 0, w \to v} \frac{f(x + \varepsilon w) - f(x)}{\varepsilon}.$$

Based on the Dini subderivative the following result is useful to establish the local Lipschitz continuity of V^e. For further details the interested reader is referred to [27].

Lemma 4.3.2 ([27]) *Suppose that the function $f : U \to (-\infty, \infty]$ is lower semicontinuous. Let $U' \subset U$ be open and convex. Then f is Lipschitz with constant M on U' if and only if for all $x \in U'$ and for all $v \in \mathbb{R}^n$ it holds that*

$$Df(x; v) \leq M\|v\|.$$

Based on this result we will conclude that under the Assumption 4.3.1 the extended Lyapunov function V^e is locally Lipschitz.

Lemma 4.3.3 *Let $\mathcal{Q}$ be a stable strict GFN model. Suppose that $\mathcal{Q}$ satisfies Assumption 4.3.1, then V^e is locally Lipschitz on $\mathbb{R}^K$.*

Proof. Let $B \subset \mathbb{R}^K$ be open, convex, and bounded and let $x \in B$. Further, let $Q \in \mathcal{Q}_{|x|_{\text{vec}}}$ be a trajectory of the strict GFN model satisfying

$$V^e(x) = \int_0^\infty \|Q(s; |x|_{\text{vec}})\| \,\mathrm{d}s.$$

Given any $w \in \mathbb{R}^K$ we consider a sequence $(\varepsilon_k)_{k\in\mathbb{N}}$ of positive real numbers which converges towards 0 so that $x + \varepsilon_k w \to x$ as $k \to \infty$. The continuity of $|\cdot|_{\text{vec}}$ implies that $|x + \varepsilon_k w|_{\text{vec}} \to |x|_{\text{vec}}$ as $k \to \infty$. So, for every $\varepsilon > 0$ there exists an $N \in \mathbb{N}$ such that $\|x - (x + \varepsilon_k w)\| = \varepsilon_k\|w\| < \varepsilon$ for all $k \geq N$. Furthermore, for N and $T := \sup\{\|x\|\,\tau, \sup_{k\geq N}\|x + \varepsilon_k w\|\,\tau\}$, by Assumption 4.3.1, there exists a $c > 0$ and trajectories $R(\cdot\,; |x + \varepsilon_k w|_{\text{vec}})$, with $k \geq N$, such that for all $t \in [0, T]$,

$$\begin{aligned}
&\|\,Q(t; |x|_{\text{vec}})\| - \varepsilon_k\|w\| - \|R(t; |x + \varepsilon_k w|_{\text{vec}})\,\| \\
&\qquad\leq \|\,Q(t; |x|_{\text{vec}}) + \varepsilon_k w - R(t; |x + \varepsilon_k w|_{\text{vec}})\,\| \leq c\,\varepsilon_k\,\|w\|\,t,
\end{aligned}$$

where the first inequality follows from the triangular inequality. This also reads as

$$\|\,Q(t; |x|_{\text{vec}})\,\| - \|\,R(t; |x + \varepsilon_k w|_{\text{vec}})\,\| \leq \varepsilon_k\,\|w\|\,(1 + c\,t) \tag{4.15}$$

for all $t \in [0, T]$. Besides, by the definition of V^e and the stability of $\mathcal{Q}$ it holds that

$$\begin{aligned}
V^e(x + \varepsilon_k w) &\geq \int_0^\infty \|\,R(s; |x + \varepsilon_k w|_{\text{vec}})\,\| \,\mathrm{d}s \\
&= \int_0^{\|x+\varepsilon_k w\|\tau} \|\,R(s; |x + \varepsilon_k w|_{\text{vec}})\,\| \,\mathrm{d}s.
\end{aligned}$$

Also, the stability of $\mathcal{Q}$ implies that

$$V^e(x) = \int_0^\infty \|Q(s; |x|_{\text{vec}})\| \,\mathrm{d}s = \int_0^{\|x\|\tau} \|Q(s; |x|_{\text{vec}})\| \,\mathrm{d}s.$$

Moreover, on one hand if $\|x\| \le \|x + \varepsilon_k w\|$ it holds that

$$\begin{aligned} V^e(x) - V^e(x + \varepsilon_k w) &\le \int_0^{\|x\|\tau} \|\, Q(s; |x|_{\text{vec}}) \,\| \,\mathrm{d}s \\ &\quad - \int_0^{\|x+\varepsilon_k w\|\tau} \|\, R(s; |x + \varepsilon_k w|_{\text{vec}}) \,\| \,\mathrm{d}s \\ &\le \int_0^{\|x\|\tau} \|\, Q(s; |x|_{\text{vec}})\| - \|R(s; |x + \varepsilon_k w|_{\text{vec}}) \,\| \,\mathrm{d}s \\ &\le \int_0^{\|x\|\tau} \varepsilon_k \,\|w\| \,(1 + c\, s) \,\mathrm{d}s = \varepsilon_k \,\|w\| \,\|x\| \,\tau \,(1 + \tfrac{c}{2}\|x\|\tau). \end{aligned}$$

On other hand, if $\|x\| > \|x + \varepsilon_k w\|$ we have that

$$\begin{aligned} V^e(x) - V^e(x + \varepsilon_k w) &\le \int_0^{\|x\|\tau} \|\, Q(s; |x|_{\text{vec}}) \,\| \,\mathrm{d}s \\ &\quad - \int_0^{\|x+\varepsilon_k w\|\tau} \|\, R(s; |x + \varepsilon_k w|_{\text{vec}}) \,\| \,\mathrm{d}s \\ &\le \int_0^{\|x+\varepsilon_k w\|\tau} \|\, Q(s; |x|_{\text{vec}})\| - \|R(s; |x + \varepsilon_k w|_{\text{vec}}) \,\| \,\mathrm{d}s \\ &\quad + \int_{\|x+\varepsilon_k w\|\tau}^{\|x\|\tau} \|\, Q(s; |x|_{\text{vec}}) \,\| \,\mathrm{d}s \\ &\le \int_0^{\|x+\varepsilon_k w\|\tau} \varepsilon_k \,\|w\| \,(1 + c\, s) \,\mathrm{d}s \\ &\quad + \tau(\, \|x\| - \|x + \varepsilon_k w\| \,) \cdot \sup_{s \in [\|x+\varepsilon_k w\|\tau, \|x\|\tau]} \|\, Q(s; |x|_{\text{vec}}) \,\| \\ &\le \varepsilon_k \,\|w\| \,\|x + \varepsilon_k w\| \,\tau \,(1 + \tfrac{c}{2}\|x + \varepsilon_k w\|\tau) \,+\, \tau \,\varepsilon_k \,\|w\| \,\|x\| \,(1 + L\tau). \end{aligned}$$

The last inequality but one follows from (4.15) and the last inequality follows from (4.12) and the triangular inequality. Consequently, taking limits implies

$$D(-V^e)(x;v) = \liminf_{k\to\infty, w\to v} \frac{V^e(x) - V^e(x+\varepsilon_k w)}{\varepsilon_k}$$
$$\le \tau\left(2 + (\tfrac{\varepsilon}{2}\|x\| + L)\tau\right) \|x\| \cdot \|v\|.$$

Since $-V^e$ is lower semicontinuous and $\|x\|$ is bounded the assertion follows from Lemma 4.3.2. ■

In the sequel, we proceed with the construction of a smooth Lyapunov function and a smooth comparison function. Let U be an open subset of $\mathbb{R}^K$ and consider the convolution of V^e and the mollifier k_r defined by

$$V_r^e(x) := V^e * k_r\,(x) = \int_{\mathbb{R}^K} V^e(x-y)\,k_r(y)\;\mathrm{d}y = \int_{\mathbb{R}^K} V(|x-y|_{\mathrm{vec}})\,k_r(y)\;\mathrm{d}y.$$

Also, we consider the convolution of the extended comparison function $w^e(x)$ given by

$$w_r^e(x) := w^e * k_r\,(x) = \int_{\mathbb{R}^K} w^e(x-y)\,k_r(y)\;\mathrm{d}y.$$

By standard convolution results it follows that $V_r^e \in C^\infty(U, \mathbb{R}_+)$ and $w_r^e \in C^\infty(U, \mathbb{R}_+)$. Furthermore, since V^e is continuous in U it holds that $V_r^e \to V^e$ uniform on compact subsets of U as $r \to 0$. Consequently, for every $\varepsilon > 0$ there is an r_0 such that for all $r \in (0, r_0)$ we have that V_r^e and w_r are smooth on U and

$$|V_r^e(x) - V^e(x)| \le \varepsilon \qquad \text{and} \qquad |w_r^e(x) - w^e(x)| \le \tfrac{\varepsilon}{2} \tag{4.16}$$

for all $x \in U$. The subsequent statement addresses the decrease condition of the convolution along trajectories of the strict GFN model $\mathcal{Q}$.

Lemma 4.3.4 *Let $U \subset \mathbb{R}_+^K$ be compact, V^e and $w \in \mathcal{K}$ be such that* (4.13) *holds. Suppose that Assumption 4.3.1 is valid. Then, for every $\varepsilon > 0$ there exists a $r_0 > 0$ such that for all $r \in (0, r_0)$ we have*

$$\dot{V}_r^e(Q(t)) \le -w^e(Q(t)) + \varepsilon \tag{4.17}$$

for all $Q \in \mathcal{Q}$ and $t \in [0,T]$ with $Q(\cdot)|_{[0,T]} \subset U$.

Proof. Due to the shift property it suffice to consider the case $t = 0$. So, let $x \in U$ and let $Q \in \mathcal{Q}_x$. Then,

$$\begin{aligned}
V_r^e(Q(t;x)) - V_r^e(x) &= \int_{\mathbb{R}^K} \Big(V^e(Q(t;x) - y) - V^e(x - y) \Big)\, k_r(y)\,\mathrm{d}y \\
&\le \int_{\mathbb{R}^K} \Big| V^e(Q(t;x) - y) \;-\; V^e(R(t;|x-y|_{\mathrm{vec}})) \Big|\, k_r(y)\,\mathrm{d}y \\
&\quad + \int_{\mathbb{R}^K} \Big(V(\,R(t;|x-y|_{\mathrm{vec}})\,) - V(|x-y|_{\mathrm{vec}}) \Big)\, k_r(y)\,\mathrm{d}y, \qquad (4.18)
\end{aligned}$$

where $R(\cdot;|x-y|_{\mathrm{vec}})$ is a trajectory corresponding to Assumption 4.3.1. By the local Lipschitz continuity of V with constant L_U and by Assumption 4.3.1, the first term on the right hand side in the above inequality can be estimated as follows

$$\begin{aligned}
&\int_{\mathbb{R}^K} \Big| V^e(Q(t;x) - y) \;-\; V^e(R(t;|x-y|_{\mathrm{vec}})) \Big|\, k_r(y)\,\mathrm{d}y \\
&\qquad \le \int_{\mathbb{R}^K} L_U \big\|\, Q(t;x) - y \;-\; R(t;|x-y|_{\mathrm{vec}})\,\big\|\, k_r(y)\,\mathrm{d}y \\
&\qquad\qquad \le t\,c\,L_U \int_{\mathbb{R}^K} \|y\| \cdot k_r(y)\,\mathrm{d}y.
\end{aligned}$$

Further, it holds that $\int \|y\|\,k_r(y)\,\mathrm{d}y \le \int r\,k_r(y)\,\mathrm{d}y = r$ and for $r_0 := \frac{\varepsilon}{2\,c\,L_U}$ it follows that

$$\int_{\mathbb{R}^K} \big| V^e(Q(t;x) - y) \;-\; V^e(R(t;|x-y|_{\mathrm{vec}})) \big| \cdot k_r(y)\,\mathrm{d}y \le t\,\tfrac{\varepsilon}{2}.$$

The stability of the strict GFN model $\mathcal{Q}$ implies that the very last term in (4.18) can be estimated by means of the convolution of the comparison function $w \in \mathcal{K}$ and the mollifier,

$$\begin{aligned}
&\int_{\mathbb{R}^K} \big(V(R(t;|x-y|_{\mathrm{vec}})) - V(|x-y|_{\mathrm{vec}}\,) \big) \cdot k_r(y)\,\mathrm{d}y \\
&\qquad \le \int_{\mathbb{R}^K} \left(- \int_0^t w\big(\|R(s;|x-y|_{\mathrm{vec}})\|\big)\,\mathrm{d}s \right) \cdot k_r(y)\,\mathrm{d}y \\
&\qquad\qquad = - \int_0^t \left(\int_{\mathbb{R}^K} w\big(\|R(s;|x-y|_{\mathrm{vec}})\|\big) \cdot k_r(y)\,\mathrm{d}y \right) \mathrm{d}s.
\end{aligned}$$

Next, we show that the function

$$s \mapsto \int_{\mathbb{R}^K} w\big(\| R(s; |x-y|_{\mathrm{vec}}) \|\big)\, k_r(y)\, \mathrm{d}y$$

is continuous in $[0,t]$.

To see this, consider the *modulus of continuity* of the function

$$s \mapsto w\big(\| R(s; |x-y|_{\mathrm{vec}}) \|\big),$$

defined for $\delta \in [0,t]$ by

$$\mathrm{m}\Big(\delta, w\big(\| R(\cdot; |x-y|_{\mathrm{vec}}) \|\big)\Big) := \sup_{|s-s'|\le\delta} \Big| w\big(\| R(s; |x-y|_{\mathrm{vec}}) \|\big) - w\big(\|R(s'; |x-y|_{\mathrm{vec}})\|\big)\Big|.$$

Then, for $s, s' \in [0,t]$ it holds that

$$w\big(\| R(s; |x-y|_{\mathrm{vec}}) \|\big) - w\big(\| R(s'; |x-y|_{\mathrm{vec}}) \|\big) \le \mathrm{m}\Big(t, w\big(\| R(\cdot; |x-y|_{\mathrm{vec}}) \|\big)\Big).$$

The stability of the strict GFN model $\mathcal{Q}$ yields that $\|R(\cdot; |x-y|_{\mathrm{vec}})\|$ is bounded and, hence, $w(\| R(\cdot; |x-y|_{\mathrm{vec}}) \|)$ is uniformly continuous. Thus, we have

$$\lim_{t\to 0} \mathrm{m}\,\Big(t, w\big(\| R(\cdot; |x-y|_{\mathrm{vec}})\|\big)\Big) = 0.$$

That is, for every $\varepsilon' > 0$ there is a $\delta_{\varepsilon'} > 0$ such that for all $t \le \delta_{\varepsilon'}$ it holds that $\mathrm{m}(t, w(\| R(\cdot; |x-y|_{\mathrm{vec}}) \|)) \le \varepsilon'$. Now, for $\varepsilon' > 0$ choose $\delta > 0$ such that $|s-s'| < \delta < \delta_{\varepsilon'}$. Then,

$$\begin{aligned}
&\int_{\mathbb{R}^K} \Big(w\big(\| R(s; |x-y|_{\mathrm{vec}}) \|\big) - w\big(\| R(s'; |x-y|_{\mathrm{vec}}) \|\big)\Big)\, k_r(y)\, \mathrm{d}y \\
&\qquad \le \int_{\mathbb{R}^K} \mathrm{m}\Big(\delta, w\big(\| R(\cdot; |x-y|_{\mathrm{vec}}) \|\big)\Big)\, k_r(y)\, \mathrm{d}y \\
&\qquad\qquad \le \int_{\mathbb{R}^K} \varepsilon'\, k_r(y)\, \mathrm{d}y = \varepsilon'.
\end{aligned}$$

Besides, by conditions (4.16) we have that $-w_r^e(x) + \frac{\varepsilon}{2} \le -w^e(x) + \varepsilon$. Finally, applying Assumption 4.3.1 and collecting the above relations yields

$$
\begin{aligned}
\dot{V}_r^e(Q(0;x)) &= \lim_{t\to 0} \frac{V_r^e(Q(t;x)) - V_r^e(x)}{t} \\
&\le \tfrac{\varepsilon}{2} - \lim_{t\to 0} \frac{1}{t} \int_{\mathbb{R}^K} \left(\int_0^t w(\|R(s;|x-y|_{\text{vec}})\|)\, k_r(y)\, dy \right) \mathrm{d}s \\
&\le - \lim_{t\to 0} \frac{1}{t} \int_0^t \left(\int_{\mathbb{R}^K} w\big(\|R(s;|x-y|_{\text{vec}})\|\big)\, k_r(y)\, dy \right) \mathrm{d}s + \tfrac{\varepsilon}{2} \\
&= - \int_{\mathbb{R}^K} w\big(\|R(0;|x-y|_{\text{vec}})\|\big)\, k_r(y)\ \mathrm{d}y + \tfrac{\varepsilon}{2} \\
&= - \int_{\mathbb{R}^K} w^e(x-y)\, k_r(y)\ \mathrm{d}y + \tfrac{\varepsilon}{2} \\
&= -w_r^e(Q(0;x)) + \tfrac{\varepsilon}{2} \le -w^e(Q(0;x)) + \varepsilon.
\end{aligned}
$$

This shows the assertion. ■

Now, let $\mathcal{U} = \{U_i\}_{i=1}^\infty$ be a locally finite open cover of $\mathbb{R}^n$ such that for every i the closure $\bar{U}_i$ is compact. Further, let $\{\varphi_i\}_{i=1}^\infty$ be a smooth partition of unity that is subordinate to $\mathcal{U}$. Details are provided in Appendix A. Define

$$
\varepsilon_i = \tfrac{1}{4} \min\{\min_{x\in\bar{U}_i} V^e(x), \min_{x\in\bar{U}_i} w^e(x)\} \qquad \text{and} \qquad q_i = \max_{x\in\bar{U}_i} \|\nabla\varphi_i(x)\|. \tag{4.19}
$$

Then, by Lemma 4.3.4 for every i there are C^∞-functions V_i^e and w_i^e such that for every $x \in U_i$,

$$
|V^e(x) - V_i^e(x)| < \frac{\varepsilon_i}{2^{i+1}(1+q_i)} \qquad \text{and} \qquad |w^e(x) - w_i^e(x)| < \varepsilon_i. \tag{4.20}
$$

Moreover, by the conditions (4.17) and (4.19) we have that

$$
\dot{V}_i^e(Q(t;x)) \le -w^e(Q(t;x)) + 2\,\varepsilon_i \le -\tfrac{1}{2} w^e(Q(t;x)). \tag{4.21}
$$

Next, we define

$$
V_s^e(x) := \sum_{i=1}^\infty \varphi_i(x)\, V_i^e(x).
$$

To see that V_s^e is proper and positive definite we note that the following estimate holds true

$$\begin{aligned}|V_s^e(x) - V^e(x)| &\leq \sum_{i=1}^{\infty} \varphi_i(x)\left|V_i^e(x) - V^e(x)\right| \\ &\leq \frac{V^e(x)}{4}\sum_{i=1}^{\infty}\frac{\varphi_i(x)}{2^{i+1}(1+q_i)} \leq \tfrac{1}{8}V^e(x).\end{aligned}$$

The next step is to derive that V_s^e is decaying along trajectories of $\mathcal{Q}$. To this end, we consider

$$\begin{aligned}\tfrac{\mathrm{d}}{\mathrm{d}t}[V_s^e(Q(t))] &= \tfrac{\mathrm{d}}{\mathrm{d}t}\left[V^e(Q(t)) + V_s^e(Q(t)) - V^e(Q(t))\right] \\ &= \tfrac{\mathrm{d}}{\mathrm{d}t}[V^e(Q(t))] + \tfrac{\mathrm{d}}{\mathrm{d}t}\left[\sum_{i=1}^{\infty}\varphi_i(Q(t))\Big(V_i^e(Q(t)) - V^e(Q(t))\Big)\right] \\ &= \dot{V}^e(Q(t)) + \sum_{i=1}^{\infty}\varphi_i(Q(t))\ \Big(\dot{V}_i^e(Q(t)) - \dot{V}^e(Q(t))\Big) \\ &\quad + \sum_{i=1}^{\infty}\dot{\varphi}_i(Q(t))\ \Big(V_i^e(Q(t)) - V^e(Q(t))\Big) \\ &\leq \sum_{i=1}^{\infty}\varphi_i(Q(t))\ \left(\dot{V}_i^e(Q(t)) + \sum_{j=1}^{\infty}\dot{\varphi}_j(Q(t))\left|V_j^e(Q(t)) - V^e(Q(t))\right|\right).\end{aligned}$$

Using the conditions (4.20) and (4.21) we get the following estimate

$$\dot{V}_s^e(Q(t)) \leq \sum_{i=1}^{\infty}\varphi_i(Q(t))\ \left(-\tfrac{1}{2}\,w^e(Q(t)) + \sum_{j=1}^{\infty}\frac{q_j\varepsilon_j}{2^{j+1}(1+q_j)}\right).$$

Defining $\widetilde{\varepsilon}_i := \max\{\varepsilon_j\ :\ x \in U_i \cap U_j \neq \emptyset\}$ we have that

$$\begin{aligned}\dot{V}_s^e(Q(t)) &\leq \sum_{i=1}^{\infty}\varphi_i(Q(t))\ \left(-\tfrac{1}{2}\,w^e(Q(t)) + \widetilde{\varepsilon}_i\sum_{j=1}^{\infty}\tfrac{1}{2^{j+1}}\right) \\ &= \sum_{i=1}^{\infty}\varphi_i(Q(t))\big(-\tfrac{1}{2}w^e(Q(t)) + \widetilde{\varepsilon}_i\big).\end{aligned}$$

Using (4.19) and the triangular inequality applied to the second inequality in (4.20), it holds that

$$-\tfrac{1}{2}w^e(Q(t)) + \widetilde{\varepsilon}_i \leq -\tfrac{1}{4}w^e(Q(t)) \leq -\tfrac{1}{5}w_i^e(Q(t)).$$

Finally, we have that

$$\dot{V}_s^e(Q(t)) \leq -\tfrac{1}{5}\sum_{i=1}^{\infty}\varphi_i(Q(t))\, w_i^e(Q(t)) =: -w_s^e(Q(t)).$$

The collection of the material in this section and the converse Lyapunov Theorem 4.2.7 yields the following smooth converse Lyapunov theorem.

Theorem 4.3.5 *Suppose the strict GFN model $\mathcal{Q}$ satisfies Assumption 4.3.1. Then, $\mathcal{Q}$ is stable if and only if there is a C^∞-smooth Lyapunov function V and a C^∞-smooth comparison function w such that*

$$\dot{V}(Q(t)) \leq -w(\|Q(t)\|)$$

for every trajectory $Q \in \mathcal{Q}$.

4.4 Notes and References

Much of the material presented in Sections 4.1 and 4.2 has already been published in [75], [76]. The notion of GFN models has been introduced by Ye and Chen, see [86]. The conditions defining GFN models have been shown for fluid limit models by Stolyar in [81]. Furthermore, in [19] these properties have been stated as remarks for work-conserving fluid networks.

An introduction to the behavioral approach to mathematical systems theory can be found in the text book [66]. However, the stability analysis therein is restricted to systems which are described by differential equations and a so-called kernel representation. The Definition 4.1.2 for stability in the behavioral framework is taken from [85].

A stability analysis of GFN models based on Lyapunov arguments was previously only considered by Ye and Chen. Their converse result, in terms of the L-condition, falls short of a converse Lyapunov theorem as stated in Theorem 4.2.7. The proof of this result is not recalled as all its steps are contained in this thesis, although they are distributed across the Chapters 3 and 4, cf. [86].

In the dynamical systems and control literature converse Lyapunov theorems have been widely studied since the 1950s. In particular, the construction of smooth Lyapunov functions has been an active field for a long time. During the last 15 years smooth Lyapunov functions have also been constructed for differential inclusions with appropriate right-hand sides, cf. [26, 82]. Especially, Teel and Praly provide a good historical overview. In 2007 a metric approach to construct smooth Lyapunov was published by Siconolfi and Terrone [79].

In the queueing literature the smoothing technique based on mollifiers was considered by Dupuis and Williams [41] and Down and Meyn [39].

The presented strategy to gain a smooth converse Lyapunov theorem is based on the approach of Clarke, Ledyaev and Stern [26]. For strict GFN models a major difference is, however, that in the abstract setting a strict GFN model is not given as the set of solutions of a differential inclusion. In particular, the conditions (a)-(f) in Definition 4.2.2 do not provide Gronwall lemma-like estimates for the evolution of the difference of two trajectories with different initial points. This exactly is the content of Assumption 4.3.1, which is appropriate that the technique works. For additional comments on Assumption 4.3.1 and situations in which it is satisfied, we refer to the notes and references at the end of the subsequent chapter. Moreover, the domain of the Lyapunov function V, defined in (4.6), is $\mathbb{R}_+^K$ as a strict GFN model contains only nonnegative trajectories. This causes the problem of defining the convolution of the continuous Lyapunov function V and a mollifier for all points on the boundary $\partial\mathbb{R}_+^K$ of the nonnegative orthant. To overcome the difficulty we extend V to $\mathbb{R}^K$ by taking componentwise absolute values. In [41] Dupuis and Williams this problem is solved by a shift of the nonnegative orthant through some constant.

In order to conclude converse Lyapunov theorems for fluid networks under the disciplines introduced in Chapter 3 the question remains whether the results obtained in this chapter are applicable to the individual fluid networks discussed in Chapter 3. This will be the content of the subsequent chapter.

5 Applications of the Converse Theorems

In this Chapter we focus again on fluid networks under the disciplines considered in Chapter 3. There, by means of a Lyapunov function, we established sufficient stability conditions for fluid networks under general work-conserving, priority, and HLPPS disciplines. One aim of this chapter is to show that for fluid networks under these disciplines a converse Lyapunov theorem indeed holds. That is, we will show that a stable fluid network admits a Lyapunov function. The strategy to obtain the desired results will be to verify that the fluid network under consideration defines a strict GFN model and apply the converse Lyapunov Theorem 4.2.7.

At the end of Chapter 2 it is demonstrated that general work-conserving fluid networks define closed GFN models. Hence, what is left to show is the concatenation property and the lower semicontinuous dependency on the initial value. For this reason, since the involved fluid processes are Lipschitz continuous, we will outline that a fluid network can be considered in terms of a differential inclusion. A precise description will be given in Section 5.1.

In Section 5.2 we will establish that general work-conserving, priority, and HLPPS fluid networks define strict GFN models. In particular, in the framework of differential inclusions we give a new proof of the existence Theorem 3.2.1 of an allocation process. Moreover, we also discuss the relation of strict GFN models to FIFO fluid networks and the fluid limit model of a HL queueing network.

In Section 5.3 we point out the impact of the converse Lyapunov theorem for the stability of HL queueing networks. Based on a Foster-Lyapunov criterion for positive Harris recurrence, we provide an alternative proof of Theorem 2.4.10 in the case that the associated fluid network defines a strict GFN model. We will use the continuous Lyapunov function of the fluid network to construct a Foster-Lyapunov function for the multiclass queueing network.

Besides, the fluid approximation of a HL queueing network, which is based on scaling associated to the strong law of large numbers, the diffusion approximation of an HL queueing network is also of interest. This approximation is based on the functional central limit theorem. Especially for the analysis of heavily loaded queueing networks the diffusion approximation is a powerful tool. Within this context the linear Skorokhod problem is of the same significance as fluid networks. In fact, the fluid limit model approach of Dai was inspired by the work of Dupuis and Williams [41]. In Section 5.4 we show that the set of solutions to the linear Skorokhod problem defines a closed GFN model which satisfies the concatenation property. Thus, for linear Skorokhod problems a converse Lyapunov theorem holds true. Precisely, we will show that a linear Skorokhod problem is stable if and only if it admits an upper semicontinuous Lyapunov function.

5.1 Fluid Networks as Differential Inclusions

To apply the converse Lyapunov Theorem 4.2.7 to fluid network models that work under one of the disciplines introduced in Chapter 3, we need to show that it defines a strict GFN model. Hence, we have to demonstrate that the concatenation property (e) and the lower semicontinuity property (f) in the Definition 4.2.2 are satisfied in each case. In order to obtain the concatenation property we make use of concepts from the theory of differential inclusions. In this section we describe that differential inclusions provide a natural framework for the analysis of fluid networks. Clearly, a detailed description of the dynamics of a fluid network depends on the specific discipline that is used. But one part of the dynamics of fluid network models that all service disciplines have in common is the flow balance relation

$$Q(t) = Q(0) + \alpha t - (I - P^{\mathsf{T}})\, MT(t), \tag{5.1}$$

where the precise description depends on the allocation process $T(\cdot)$, and hence, on the service discipline. According to Proposition 2.4.13 the processes are Lipschitz continuous and by Rademacher's Theorem differentiable almost everywhere. Hence, for almost all $t \in \mathbb{R}_+$ the flow balance relation (5.1) can also be written as

$$\dot{Q}(t) = \alpha - (I - P^{\mathsf{T}})\, M\dot{T}(t), \qquad Q(0) = Q_0. \tag{5.2}$$

Now we consider the derivative of the allocation process as variable, i.e. we define $u(t) = \dot{T}(t)$ almost everywhere and note that u is measurable. The corresponding differential equation can be written as

$$\dot{Q}(t) = f(Q(t), u(t)) := \alpha - (I - P^{\mathsf{T}})\, Mu(t). \tag{5.3}$$

The allocation rate $u(\cdot)$ is determined through the service discipline. So, each service discipline has a set of admissible values $U(Q)$, where $u \in U(Q)$ if and only if $u \in \mathbb{R}_+^K$ satisfies some allocation conditions that are specific to the discipline. As mentioned in Remark 2.4.6 the solutions to the fluid network equations are not unique and, thus, the allocation process need not be unique as well. Hence, for every $Q \in \mathbb{R}_+^K$ there are different choices of u possible, where the admissible values u depend on the fluid level process $Q(t)$ through the allocation conditions. Thus, the flow balance relation (5.2) can also be expressed by a differential inclusion of the form

$$\dot{Q}(t) \in \{f(Q(t), u(t)) \,:\, u(t) \in U(Q(t))\}, \quad Q(0) = Q_0. \tag{5.4}$$

Sometimes the set-valued map $U : Q \rightsquigarrow u$ is referred to as the feedback map. Defining

$$F(Q(t)) := \{\alpha - (I - P^{\mathsf{T}})\, Mu(t) \,:\, u(t) \in U(Q(t))\}, \tag{5.5}$$

we rewrite (5.4) as a closed loop differential inclusion

$$\dot{Q}(t) \in F(Q(t)), \quad Q(0) = Q_0. \tag{5.6}$$

In the following section we will investigate the differential inclusions and the set-valued map U for the particular service disciplines introduced in the Chapters 2 and 3. For this reason, our analysis is based on general results from the theory of differential inclusions. These are summarized in Appendix B. However, the so called Filippov Lemma is stated here, since it fits to the special structure of the differential inclusion defined by (5.5).

Lemma 5.1.1 ([45]) *Let $f(t, u)$ be continuous and let the set-valued map $t \rightsquigarrow U(t)$ be upper semicontinuous with closed convex values. Let $y(\cdot)$ be a measurable function such that $y(t) \in f(t, U(t))$ for almost all t. Then, there is a measurable function $u(\cdot)$ such that $y(t) = f(t, u(t))$ for almost all t.*

Proof. See [45] pp.78/79. ∎

5.2 Applications to Some Fluid Networks

In this section we discuss the differential inclusions that arise from the consideration of the particular disciplines. We will start our analysis for general work-conserving fluid networks. Subsequently, we examine fluid networks under priority and HLPPS disciplines. In addition, we contemplate the relation of fluid limit models to strict GFN models and point out why FIFO fluid networks do not fall immediately into the framework of differential inclusions.

General Work-Conserving Fluid Networks

First, we recall the dynamic equations for fluid networks under a general work-conserving service discipline. That is,

$$Q(t) = Q(0) + \alpha\, t - (I - P^{\mathsf{T}})MT(t) \geq 0, \tag{5.7}$$

$$T(0) = 0 \text{ and } \dot{T}(\cdot) \geq 0 \tag{5.8}$$

$$I(t) = et - C\,T(t) \text{ and } \dot{I}(\cdot) \geq 0 \tag{5.9}$$

$$0 = (C\,Q(t))^{\mathsf{T}}\ \dot{I}(t), \qquad \text{for almost all } t \geq 0. \tag{5.10}$$

Second, we bring the conditions (5.8)-(5.10) into the context of the differential inclusions (5.6). So, for $Q \in \mathbb{R}^K_+$ the conditions defining the admissible values for u are

$$u \geq 0, \qquad e - Cu \geq 0, \qquad (CQ)^{\mathsf{T}} \cdot (e - Cu) = 0. \tag{5.11}$$

The conditions in (5.11) are immediate consequences of (5.8), (5.9) and (5.10) in their differentiation. Hence, the set of admissible values of u is given by

$$U_C(Q) := \left\{ u \in \mathbb{R}^K_+ \ :\ (5.11) \text{ is satisfied } \right\}.$$

The set of allocation rates u that satisfy $u \geq 0$ and $e - Cu \geq 0$ is a compact and convex subset of $\mathbb{R}^K_+$. In addition, the condition $(CQ)^{\mathsf{T}} \cdot (e - Cu) = 0$ defines a hyperplane. Hence, the set $U_C(Q)$, given by the intersection of a compact and convex set and a hyperplane, is compact and convex. That is, the set of all feasible directions being in the state $Q \in \mathbb{R}^K$ is given by

$$F(Q) = \left\{ \alpha - (I - P^{\mathsf{T}})\, Mu \ :\ u \in U_C(Q) \right\}. \tag{5.12}$$

So, we consider the differential inclusion defined by

$$\dot{Q}(t) \in F(Q(t)) \tag{5.13}$$

and the initial condition $Q(0) = Q_0$. Note that the fluid level processes defined by (5.7)-(5.10) are forced to be nonnegative. However, for fluid levels on the boundary of the nonnegative orthant the conditions in (5.11) do not provide the nonnegativity of the fluid level. Hence, for fluid levels on the boundary of the nonnegative orthant we have to rule out the directions that would cause an abandonment of the nonnegative orthant. For this reason, according to Theorem B.3 in Appendix B, for every $Q \in \mathbb{R}_+$ we have to consider intersection of $F(Q)$ in (5.12) and the contingent cone $\mathcal{T}_{\mathbb{R}^K_+}(Q)$ to the nonnegative orthant. The existence of solutions that remain in the nonnegative orthant is the content of the next result, which is an elegant method to prove Theorem 2.1 in [19].

Theorem 5.2.1 *For any work-conserving fluid network* (α, μ, P, C) *with an initial level* Q_0 *there exists a work-conserving allocation* T.

Proof. From the conditions (5.11) it follows that the set $U_C(Q)$ is compact and convex and U_C is upper semicontinuous. Further, the map $(Q, u) \mapsto \alpha - (I - P^\mathsf{T})\, M\, u$ is continuous. Hence, by Proposition B.1 the set-valued map F is upper semicontinuous. Moreover, F has closed convex values that are contained in some ball with radius $b > 0$. Also the conditions (5.11) imply that $F(Q) \cap \mathcal{T}_{\mathbb{R}^K_+}(Q) \neq \emptyset$ for all $Q \in \mathbb{R}^K_+$. Then, by Theorem B.3 there exists a solution $Q(\cdot)$ to (5.6) such that $Q(t) \in \mathbb{R}^K_+$ for all $t \geq 0$.

Furthermore, let $Q(\cdot)$ be a solution that remains in the nonnegative orthant. Note that $f(Q, u)$ is continuous in u and that $U(t) := \{u \in \mathbb{R}^K_+ : e - Cu \geq 0,\ (CQ(t))^\mathsf{T}(e - Cu) = 0\}$ is closed and bounded. Also, $t \rightsquigarrow U(t)$ is upper semicontinuous. Then, by the Filippov Lemma 5.1.1, there is a measurable selection $u(\cdot)$ of $t \mapsto U_C(Q(t))$ such that $u(t) \in U_C(Q(t))$ for almost all $t \geq 0$ and

$$\dot{Q}(t) = \alpha - (I - P^\mathsf{T})M\, u(t) \qquad \text{for almost all } t \geq 0.$$

Thus, integrating the latter yields that given Q_0 the pair $(Q(\cdot), T(\cdot))$ with $T(t) := \int_0^t u(s)\, \mathrm{d}s$ is a fluid solution. ■

Remark 5.2.2 *In general, the set-valued map* $Q \rightsquigarrow F(Q) \cap \mathcal{T}_{\mathbb{R}^K_+}(Q)$ *is not upper semicontinuous as the following example shows.*

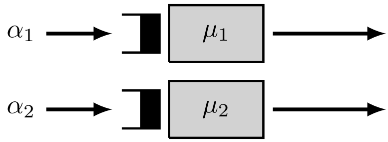

Figure 5.1: A two station network serving two fluid classes.

Example 5.2.3 *We consider a two station fluid network serving two fluid classes as depicted in Figure 5.1. Let* $\varepsilon > 0$ *and let the service capacities be* $\mu_i = \alpha_i + \varepsilon$, $i = 1, 2$. *Further, we consider the initial fluid level* $Q_0 = 0$ *and the sequence* $Q_n = \big(0 \ \tfrac{1}{n}\big)^\mathsf{T}$ *of initial levels. Then, for all* $n \in \mathbb{N}$ *and for small* $t \geq 0$, *the allocation rates are given by* $\dot{T}_n(t) = \big(\tfrac{\alpha_1}{\alpha_1+\varepsilon} \ 1\big)^\mathsf{T}$. *So, for the set-valued map* F *defining the right-hand side of the differential inclusion we have* $F(Q_n) = \big(0 \ -\varepsilon\big)^\mathsf{T}$ *and, thus,*

$$F(Q_n) \cap \mathcal{T}_{\mathbb{R}^2_+}(Q_n) = \begin{pmatrix} 0 \\ -\varepsilon \end{pmatrix}.$$

Now we pick the sequence $\big(Q_n, (0 \ -\varepsilon)^\mathsf{T}\big)$ *on the graph of* $F \cap \mathcal{T}_{\mathbb{R}^2_+}$ *which converges to* $\big(0, (0 \ -\varepsilon)^\mathsf{T}\big) \notin \operatorname{graph}(F \cap \mathcal{T}_{\mathbb{R}^2_+})$. *Hence, the graph is not closed and by Proposition B.2 the set-valued map* $F \cap \mathcal{T}_{\mathbb{R}^2_+}$ *is not upper semicontinuous.*

We use $AC(\mathbb{R}_+, \mathbb{R}^K_+)$ to denote the set of absolutely continuous functions $x : \mathbb{R}_+ \to \mathbb{R}^K_+$. Let $\mathcal{S} : \mathbb{R}^K_+ \rightsquigarrow AC(\mathbb{R}_+, \mathbb{R}^K_+)$ denote the solution map of the differential inclusion

$$\dot{Q}(t) \in F(Q(t)) \cap \mathcal{T}_{\mathbb{R}^K_+}(Q(t))$$

describing general work-conserving fluid networks. We omit any sub- and superscripts since it will be clear from the context to which service discipline and differential inclusion the solution map refers to. Thus, we can represent the set of work-conserving fluid level processes by

$$\mathcal{Q}_C = \{Q(\cdot) \,:\, Q(\cdot) \in \mathcal{S}(Q_0),\, Q_0 \in \mathbb{R}^K_+\}.$$

Following the Propositions 2.4.13, 2.4.14, and 2.4.15 it holds that $\mathcal{Q}_C$ defines a closed GFN model. So, we only have to prove that the concatenation and the lower semicontinuity property are satisfied.

Proposition 5.2.4 *The set of fluid level processes $\mathcal{Q}_C$ satisfies the concatenation property (e) in Definition 4.2.2.*

Proof. Since solutions of differential inclusions are by definition absolutely continuous functions, and concatenation preserves absolute continuity. This shows the assertion. ■

The next statement shows that the solution map depends lower semicontinuously on the initial condition.

Proposition 5.2.5 *The set of fluid level processes $\mathcal{Q}_C$ satisfies the lower semicontinuity property (f) in Definition 4.2.2.*

Proof. To show condition (f) we have to verify the existence of a $T > 0$ such that $Q_0 \rightsquigarrow \mathcal{S}(Q_0)|_{[0,T]}$ is lower semicontinuous. In view of Proposition 4.2.4 and Proposition 5.2.4 it is sufficient to construct for each Q_0 a $T(Q_0) > 0$ such that (f') holds. To this end, let $Q_0 \in \mathbb{R}_+^K$ be fixed, $Q(\cdot) \in \mathcal{S}(Q_0)$. Then, by the proof of Theorem 3.2.1 there exists a function $u(\cdot) \in U(Q(\cdot))$ such that

$$Q(t) = Q_0 + \alpha\, t - (I - P^{\mathsf{T}})M \int_0^t u(s)\,\mathrm{d}s. \tag{5.14}$$

We distinguish the following situations.

First, suppose that $Q_0 \in \mathbb{R}_+^K$ is such that all stations have some nonempty queues, i.e. $CQ_0 > 0$. Hence, there is a $T(Q_0) > 0$ such that $CQ(t) > 0$ for all $t \in [0, T(Q_0)]$. We note that $(CQ)^{\mathsf{T}} \cdot (e - Cu) = 0$ from (5.11) also reads as

$$\sum_{j=1}^{J} \left(\sum_{l \in C(j)} Q_l \cdot \left(1 - \sum_{l \in C(j)} u_l \right) \right) = 0.$$

Since both factors are nonnegative and, in fact, $CQ(t) > 0$ for $t \in [0, T(Q_0)]$, it holds that

$$1 = \sum_{l \in C(j)} u_l(\cdot)|_{[0,T(Q_0)]} =: e_j^{\mathsf{T}} C u(\cdot)|_{[0,T(Q_0)]} \tag{5.15}$$

for all $j = 1, ..., J$. Let $(Q_0^n)_{n\in\mathbb{N}}$ be a sequence of initial values converging to Q_0. Consider the functions

$$Q^n(t) := Q_0^n + \alpha\, t - (I - P^\top)M \int_0^t u(s)\,\mathrm{d}s = Q(t) + (Q_0^n - Q_0)\,. \quad (5.16)$$

The selection u clearly satisfies the constraint (5.11), so if Q^n is nonnegative on $[0, T(Q_0)]$, it defines a fluid solution on that interval. We claim that this is the case for n large enough. Indeed (5.16) shows for the indices k for which $Q_{0,k} = 0$ that $Q_k^n(t) \geq Q_k(t) \geq 0$, while if $Q_{0,k} > 0$, then by the choice of $T(Q_0)$ we have $Q_k(t) > 0$ on $[0, T(Q_0)]$. And so for n sufficiently large $Q^n(t) > 0$. As $Q^n \to Q$ uniformly on $[0, T(Q_0)]$ we obtain that $Q \rightsquigarrow \mathcal{S}_F(Q)|_{[0,T(Q_0)]}$ is lower semicontinuous at Q_0.

Second, suppose that the initial fluid level at some stations is zero. We first treat the case of a single station with empty queues. Without loss of generality let this station be $j = 1$ and let a denote the set of classes which are served at station 1. Then, the last constraint in (5.11) is not active for station 1 and therefore the constraints for fluid classes $k \in a$ are given by

$$u_k \geq 0, \qquad 1 - \sum_{l\in a} u_l \geq 0. \quad (5.17)$$

However, since $Q(\cdot)$ is a solution to the differential inclusion (5.6) potentially only a proper subset of (5.17) is feasible. If this condition enforces equality in the second constraint in (5.17), then we can argue as in (5.15) on a sufficiently small time interval and the previous argument applies again. The interesting case is when there is idle capacity at station $j = 1$. Here $u_k(\cdot) \geq 0$ are such that $\sum_{l\in a} u_l(\cdot) < 1$ and that the fluid levels of classes $k \in a$ remain nonnegative. Using $b := \{1, ..., K\}\backslash a$ the differential form of the flow balance equation (5.7) can be expressed in block form by

$$\begin{aligned}\dot{Q}_a(t) &= \alpha_a + P_a^\top M_a\, u_a(t) + P_{ab}^\top M_b\, u_b(t) - M_a\, u_a(t)\\ \dot{Q}_b(t) &= \alpha_b + P_{ba}^\top M_a\, u_a(t) + P_b^\top M_b\, u_b(t) - M_b\, u_b(t).\end{aligned}$$

The nonnegativity of the fluid levels for classes $l \in a$ yields the following condition

$$0 \leq \alpha_a + P_{ab}^\top M_b\, u_b(\cdot) - (I_a - P_a^\top)\, M_a\, u_a(\cdot),$$

which also reads as

$$u_a(\cdot) \leq M_a^{-1}\, (I_a - P_a^\top)^{-1}\, (\alpha_a + P_{ab}^\top M_b\, u_b(\cdot)).$$

As $e_j^T C Q_0 > 0$ for $j \neq 1$ then, arguing as in (5.15) there is a $T(Q_0) > 0$ such that the allocation rates corresponding to fluid classes present at the stations $j \neq 1$ satisfy $\sum_{l \in C(j)} u_l(\cdot)|_{[0,T(Q_0)]} = 1$. Let $\varepsilon > 0$ be fixed, so that if $\|Q_0 - Q\| < \varepsilon$ then $Q_k > 0$ when $Q_{0,k} > 0$. Now, for another initial value Q_0^1 with $\|Q_0 - Q_0^1\| < \varepsilon$ we consider $u^1(\cdot) := (u_a(\cdot) + v(\cdot)\,,\, u_b(\cdot))^\mathsf{T}$, where $v(\cdot)$ takes values in $\mathbb{R}^{|a|}$ such that

$$\sum_{l \in a} u_l(t) + v_l(t) = 1 \quad \text{if} \quad e_1^T C Q^1(t) > 0, \tag{5.18}$$

and $v(t) = 0$ otherwise. Then, we consider the solution $Q^1(\cdot)$ associated with $u^1(\cdot)$ and Q_0^1, i.e.

$$\begin{aligned} Q^1(t) &= Q_0^1 + \alpha\, t - (I - P^\mathsf{T})M \int_0^t \begin{pmatrix} u_a(s) + v(s) \\ u_b(s) \end{pmatrix} \mathrm{d}s \\ &= Q_0^1 - Q_0 + Q(t) - (I - P^\mathsf{T})M \int_0^t \begin{pmatrix} v(s) \\ 0 \end{pmatrix} \mathrm{d}s. \end{aligned}$$

So, the difference between the solutions $Q(\cdot)$ and $Q^1(\cdot)$ is given by

$$Q^1(t) - Q(t) = Q_0^1 - Q_0 + \int_0^t \begin{pmatrix} P_a^\mathsf{T} M_a v(s) \\ P_{ba}^\mathsf{T} M_a v(s) \end{pmatrix} - \begin{pmatrix} M_a v(s) \\ 0 \end{pmatrix} \mathrm{d}s$$

In particular, as $Q_{0,a} = 0$ we have that

$$Q_a^1(t) - Q_a(t) = Q_{0,a}^1 - (I - P_a^\mathsf{T})M_a \int_0^t v(s)\ \mathrm{d}s.$$

Hence, if $Q_{0,a}^1 > 0$ the nonnegativity of $(I - P_a^\mathsf{T})M_a$ and $v(\cdot)$ imply that there is a $r \geq 0$ such that

$$Q_{0,a}^1 - (I - P_a^\mathsf{T})M_a \int_0^r v(s)\ \mathrm{d}s = 0. \tag{5.19}$$

We will assume that $v(\cdot)$ is chosen so that the time in which (5.19) is achieved is minimal.

Thus, given a sequence of initial values $(Q_0^n)_{n \in \mathbb{N}}$ converging to Q_0 and in particular $Q_{0,a}^n$ converging to zero, we define

$$r_n := \min\{r \geq 0\ :\ v^n(\cdot) \text{ satisfies (5.18) and (5.19)}\}$$

and

$$u^n(t) := \begin{cases} (\, u_a(t) + v^n(t)\,,\, u_b(t)\,)^\mathsf{T} & \text{for} \quad 0 \le t \le r_n, \\ (\, u_a(t)\,,\, u_b(t)\,)^\mathsf{T} & \text{for} \quad t > r_n. \end{cases}$$

Further, we note that (5.19) implies that if $Q^n_{0,a}$ converges to $Q_{0,a} = 0$ it holds that r_n converges to zero as well. Hence, we have that $u^n(\cdot)$ converges to $u(\cdot)$ and consequently $Q^n(\cdot)$ converges uniformly to $Q(\cdot)$ on $[0, T(Q_0)]$, i.e. $Q(\cdot)|_{[0,T(Q_0)]}$ depends lower semicontinuously on Q_0.

The cases where more than one stations have empty queues follows the same line of reasoning. The assertion follows from Proposition 4.2.4. ∎

In a nutshell, the latter results yield the desired converse Lyapunov theorem for general work-conserving fluid networks.

Theorem 5.2.6 *A general work-conserving fluid network defines a strict GFN model. In particular, it is stable if and only if it admits a continuous Lyapunov function.*

Priority Fluid Networks

For convenience we recall the basic notations for priority fluid networks. A priority discipline is determined by a permutation

$$\pi : \{1, ..., K\} \to \{1, ..., K\},$$

where for fluid classes l, k present at the same station $j = c(k) = c(l)$ fluids of class l have priority over fluids of class k if and only if $\pi(l) < \pi(k)$. For $k \in \{1, ..., K\}$ the set of all fluid classes served at the same station $j = c(k)$ that have priority over k is denoted by

$$\Pi_k := \{l : l \in C(c(k)),\, \pi(l) \le \pi(k)\}.$$

The dynamic equations describing the evolution are given by

$$Q(t) = Q(0) + \alpha\, t - (I - P^\mathsf{T}) M T(t) \ge 0, \tag{5.20}$$

$$T(0) = 0 \text{ and } \dot{T}(\cdot) \ge 0, \tag{5.21}$$

$$Y_k(t) = t - \sum_{l \in \Pi_k} T_l(t) \text{ and } \dot{Y}(\cdot) \ge 0, \quad k \in \{1, ..., K\} \tag{5.22}$$

$$0 = Q_k(t)\, \dot{Y}_k(t) \quad \text{for almost all } t \ge 0, \quad k \in \{1, ..., K\}. \tag{5.23}$$

The subsequent statement follows similar to the Propositions 2.4.13, 2.4.14, and 2.4.15 directly from the dynamic equations.

Lemma 5.2.7 ([86]) *A priority fluid network defines a closed GFN model.*

To show that priority fluid networks also define strict GFN models we rewrite the dynamic equations in the context of differential inclusions by setting $\dot{T} = u$. To determine the set of feasible allocation rates u we consider the conditions (5.21)-(5.23) in their differentiation, i.e. for $k \in \{1, ..., K\}$,

$$u_k \geq 0, \qquad 1 - \sum_{l \in \Pi_k} u_l \geq 0, \qquad Q_k \cdot (1 - \sum_{l \in \Pi_k} u_l) = 0. \tag{5.24}$$

Likewise, for a given fluid level $Q \in \mathbb{R}_+^K$ we define

$$U_\Pi(Q) := \left\{ u \in \mathbb{R}_+^K \, : \, (5.24) \text{ is satisfied for all } k \in \{1, ..., K\} \right\}.$$

The constraints (5.24) show that $U_\Pi(Q)$ is compact. To see that $U_\Pi(Q)$ is also convex, let u and v be in $U_\Pi(Q)$ and let $\lambda \in (0, 1)$. Then, it follows immediately that $\lambda u + (1 - \lambda) v \geq 0$. Further, we have that

$$\sum_{l \in \Pi_k} \lambda u_l + (1 - \lambda) v_l = \lambda \sum_{l \in \Pi_k} u_l + (1 - \lambda) \sum_{l \in \Pi_k} v_l \leq 1$$

and the second constraint in (5.24) is also satisfied. To verify the third condition, it suffices to regard the case when $Q_k > 0$ since otherwise the condition is already included in the other constraints. Note that in this case we have that $\sum_{l \in \Pi_k} u_l = \sum_{l \in \Pi_k} v_l = 1$. Consequently, it holds that

$$\begin{aligned} Q_k \cdot \Big(1 - \sum_{l \in \Pi_k} \lambda u_l + (1 - \lambda) v_l \Big) &= Q_k \cdot \Big(1 - (\lambda \sum_{l \in \Pi_k} u_l + (1 - \lambda) \sum_{l \in \Pi_k} v_l) \Big) \\ &= Q_k \cdot (1 - (\lambda + (1 - \lambda))\,)= 0. \end{aligned}$$

In addition, the set-valued map $U_\Pi : Q \rightsquigarrow U_\Pi(Q)$ is upper semicontinuous.

Thus, a priority fluid network is described by the set of solutions to the following differential inclusion

$$\dot{Q}(t) \in F(Q(t)) := \{ \alpha - (I - P^\mathsf{T})\, M u : \; u \in U_\Pi(Q(t))\, \}.$$

The set-valued map F has nonempty compact and convex values and, by Proposition B.1, it is upper semicontinuous. Since $F(Q) \cap \mathcal{T}_{\mathbb{R}_+^K}(Q) \neq \emptyset$ for all $Q \in \mathbb{R}_+^K$, the same line of arguments as for general work-conserving fluid networks implies the subsequent converse Lyapunov theorem for priority networks.

Theorem 5.2.8 *A priority fluid network defines a strict GFN model. In particular, it is stable if and only if it admits a continuous Lyapunov function.*

HLPPS Fluid Networks

We recall briefly the description for HLPPS fluid networks. At stations with positive total fluid level, the fluid classes present at a station are served simultaneously proportional to their current fluid level. The total fluid level at station j at time t is given by

$$Q_j^{\Sigma}(t) := e_j^{\mathsf{T}} C\, Q(t) := \sum_{l \in C(j)} Q_l(t).$$

The dynamic equations of fluid networks under HLPPS disciplines can be summarized as follows

$$Q(t) = Q(0) + \alpha\, t - (I - P^{\mathsf{T}}) M T(t) \geq 0, \tag{5.25}$$

$$W(t) = C\, M^{-1}\, Q(t), \tag{5.26}$$

$$I(t) = et - C\, T(t), \quad \dot{I}(\cdot) \geq 0, \tag{5.27}$$

$$0 = \dot{I}_j(t)\, W_j(t) \quad \text{for almost all } t \geq 0,\ \forall\, j = 1, ..., J, \tag{5.28}$$

$$\dot{T}_k(t) = \frac{Q_k(t)}{Q_{c(k)}^{\Sigma}(t)} \quad \text{if} \quad Q_{c(k)}^{\Sigma}(t) > 0. \tag{5.29}$$

The subsequent statement follows similar to the Propositions 2.4.13, 2.4.14, and 2.4.15 directly from the dynamic equations. Thus, we obtain the following result.

Lemma 5.2.9 *A HLPPS fluid network defines a closed GFN model.*

To conclude that HLPPS fluid networks satisfy the concatenation and the lower semicontinuity property, we define $\dot{T}(t) =: u(t)$ and transfer the conditions (5.26)-(5.29) to the context of differential inclusions. Further, we regard the conditions that determine the allocation rates and, hence, the set-valued map U_{PPS} mapping to each fluid level the set of feasible allocation rates. If $Q_j^{\Sigma}(t) > 0$, the allocation rate u_k of fluid class $k \in C(j)$ is given by

$$u_k(t) = \frac{Q_k(t)}{Q_{c(k)}^{\Sigma}(t)}. \tag{5.30}$$

If $Q_j^\Sigma(t) = 0$ for some $j \in \{1, ..., J\}$, according to the basic fluid network equation (5.27) the allocation rates u_k for $k \in C(j)$ have to satisfy

$$u_k \geq 0, \quad 1 - e_j^\top C\, u \geq 0. \tag{5.31}$$

Hence, for $Q \in \mathbb{R}_+^K$ the set of feasible allocation rates is given by

$$U_{PPS}(Q) := \left\{ u \in \mathbb{R}_+^K \;:\; (5.30) \text{ or } (5.31) \text{ is satisfied } \right\}.$$

To see that, given $Q \in \mathbb{R}_+^K$, the set of feasible allocation rates is convex observe that if the total fluid level of a station is positive, the set of feasible allocation rates is a singleton. Otherwise, let $e_j^\top CQ = 0$ for some $j \in \{1, ..., J\}$ and let u and v be feasible allocation rates. Then,

$$e_j^\top C\,(\lambda u + (1-\lambda)v) = \lambda e_j^\top Cu + (1-\lambda)e_j^\top Cv \leq \lambda + (1-\lambda) = 1.$$

We consider the differential inclusion

$$\dot{Q}(t) \in F(Q(t)) := \left\{ \alpha - (I - P^\top)\, Mu \;:\; u \in U_{PPS}(Q(t)) \right\}, \tag{5.32}$$

with initial condition $Q(0) = Q_0$. The right-hand side F is upper semicontinuous and has nonempty, compact and convex values. In addition, we have that $F(Q) \cap \mathcal{T}_{\mathbb{R}_+^K}(Q) \neq \emptyset$ for all $Q \in \mathbb{R}_+^K$. Consequently, we obtain the desired result.

Theorem 5.2.10 *A HLPPS fluid network defines a strict GFN model. It is stable if and only if it admits a continuous Lyapunov function.*

Fluid Limit Models

A further class of interest are fluid limit models of queueing networks. Queueing networks are discussed in Chapter 2. We assume that the Assumptions 2.1.2 are satisfied. Furthermore, the evolution of the queueing network is described by the tuple $\mathsf{X}(t) = (A(t), D(t), T(t), W(t), I(t), Q(t))$ which contains the queue length process $Q(t)$. As in Chapter 2, G denotes the set on which the strong law of large numbers holds for the interarrival times, the service times, and the routing. Furthermore, we consider sequences of pairs $(r_n, x_n)_{n\in\mathbb{N}}$ that satisfy condition (2.9), where x_n are initial states and $r_n \in \mathbb{R}_+$. By Theorem 2.4.2 for every $\omega \in G$ and for every sequence of pairs $(r_n, x_n)_{n\in\mathbb{N}}$ satisfying condition (2.9), there exist a subsequence such that

$$\frac{1}{r_{n_j}}\, \mathsf{X}^{x_{n_j}}\,(r_{n_j}t) \longrightarrow \overline{\mathsf{X}}(t) \qquad \text{u.o.c. as } j \to \infty.$$

In the following, we focus on the set of fluid level limits $\overline{Q}$, which is denoted by $\mathcal{Q}_L$. That is, whenever a fluid limit $\overline{Q}(\cdot)$ is considered there is a sample path $\omega \in G$ and a sequence of pairs satisfying (2.9) such that

$$\frac{1}{r_n}\, Q^{x_n}(r_n t, \omega) \longrightarrow \overline{Q}(t) \qquad \text{u.o.c. as } n \to \infty. \tag{5.33}$$

Next, we investigate the Lipschitz, the scaling, and shift property. The subsequent statement was derived by Ye and Chen. However, in the proof provided in [86] there is a gap in the verification of the shift property. We will fill this gap in the following proof.

Proposition 5.2.11 ([86]) *The fluid limit model $\mathcal{Q}_L$ defines a GFN model.*

Proof. Let $\overline{Q} \in \mathcal{Q}_L$ be a fluid limit. The Lipschitz continuity of the fluid limit processes follows immediately from Theorem 2.4.5 and Proposition 2.4.13.

To see that $\mathcal{Q}_L$ satisfies the scaling property let ω be a sample path that satisfies the strong law of large numbers and $(r_n, x_n)_{n\in\mathbb{N}}$ be a sequence of pairs satisfying (2.9). For $r > 0$ and the sequence of pairs $(r \cdot r_n, x_n)_{n\in\mathbb{N}}$ it holds that

$$\frac{1}{r\cdot r_n}\, Q^{x_n}((r\cdot r_n)t, \omega) = \frac{1}{r}\frac{1}{r_n}\, Q^{x_n}(r_n \cdot (r\,t), \omega) \longrightarrow \frac{1}{r}\overline{Q}(r\,t) \qquad \text{u.o.c.}$$

as $n \to \infty$. Hence, the scaling property is valid.

To show the shift property, we follow an idea that is due to Robert, see [68]. Let $\overline{Q} \in \mathcal{Q}_L$ be so that $\overline{Q}(0) = \overline{q}$ and let $\mathcal{FL}(\overline{q})$ denote the set of fluid limits with initial level $\overline{q}$, i.e.

$$\mathcal{FL}\,(\overline{q}) := \{\,\overline{Q} : \mathbb{R}_+ \to \mathbb{R}^K_+ \;:\; \overline{Q}(t) = \lim_{n\to\infty} \frac{1}{r_n} Q^{x_n}(r_n\, t)\,,\; \overline{Q}(0) = \overline{q}\,\}.$$

Fix a sequence of pairs $(r_n, x_n)_{n\in\mathbb{N}}$ satisfying (2.9) so that $\lim_{n\to\infty} \frac{x_n}{r_n} = (\overline{q}, 0, 0)$. Then, by the Skorokhod's Representation Theorem 1.2.8 we have along a subsequence

$$\lim_{k\to\infty} \frac{1}{r_{n_k}} Q^{x_{n_k}}(r_{n_k} t) = \overline{Q}^{n_k}(t, \overline{q}) \in \mathcal{FL}(\overline{q})$$

a.s. in the Skorokhod topology. The superscript to the fluid limit expresses the dependence on the particular sequence. Moreover, for any $s \geq 0$ by the Markov property we have the following equality in distribution

$$Q^{x_{n_k}}(r_{n_k}(t+s)) \stackrel{d}{=} Q^{Q^{x_{n_k}}(r_{n_k} s)}(r_{n_k} t). \tag{5.34}$$

Also, by Proposition 1.2.6 and since $t \mapsto \overline{Q}^{n_k}(t, \overline{q})$ is continuous, it holds that

$$\lim_{k\to\infty} \frac{1}{r_{n_k}} Q^{x_{n_k}}(r_{n_k} s) = \overline{Q}^{n_k}(s, \overline{q}) \quad \text{a.s.}$$

Consequently, dividing (5.34) by r_{n_k} and taking limits yields that

$$\overline{Q}^{n_k}(t+s, \overline{q}) \stackrel{d}{=} \overline{Q}^{n_k}(t, \overline{Q}^{n_k}(s))$$

and, hence, we have

$$\overline{Q}^{n_k}(\cdot + s, \overline{q}) \in \mathcal{FL}\left(\overline{Q}^{n_k}(s, \overline{q})\right).$$

This shows the assertion. ■

For this class the open question remained whether they define closed GFN models [86]. As we will see, taking the closure with respect to uniform convergence on compact sets does not change the stability properties. In this way, we obtain from fluid limit models closed GFN models. We define the fluid limit model as the closure of $\mathcal{Q}_L$ with respect to uniform convergence on compact intervals, and denote it by $\overline{\mathcal{Q}}_L$.

Lemma 5.2.12 *The fluid limit model $\overline{\mathcal{Q}}_L$ is stable if and only if $\mathcal{Q}_L$ is stable.*

Proof. Obviously, if $\overline{\mathcal{Q}}_L$ is stable, then $\mathcal{Q}_L$ is stable. Conversely, assume that $\mathcal{Q}_L$ is stable. Let $Q_* \in \overline{\mathcal{Q}}_L \setminus \mathcal{Q}_L$ and $Q_n \in \mathcal{Q}_L$ a sequence such that $Q_n(\cdot) \to Q_*(\cdot)$ u.o.c. as $n \to \infty$. Since $\mathcal{Q}_L$ is stable there is a uniform

$\tau > 0$ such that $Q_n(\tau + \cdot) \equiv 0$ for all $n \in \mathbb{N}$. By the Lipschitz continuity property it follows for all $t \geq \tau$ that

$$Q_*(t) = \lim_{n\to\infty} Q_n(t) = 0$$

and the proof is completed. ∎

In the following we consider queueing networks under disciplines that are *memoryless* in the sense that the allocation process $\overline{T}$ of the fluid limit model at a time t does only depend on the queue length at that time t. In particular, it does not require information of the past. In terms of the fluid limit models described in [28] this means that only the fluid level at a given time is needed to describe the evolution of the fluid level process. Note that this explicitly excludes a number of disciplines as e.g. FIFO networks. We will comment on FIFO fluid networks later in this section. We also note that the problem of concatenating fluid limits was also addressed by Stolyar [81] and Robert [68, Section 9.2.3]. In [81] it is shown that if the queueing disciplines in every station satisfy a certain 'uniqueness condition' on the disciplines of the individual servers the concatenation property holds. However, there the definition of state is different, because the state as used in [81] includes the past trajectory of the queue. Furthermore, in [68] concatenation is possible if the fluid limits going through a certain queue level Q are unique.

Remark 5.2.13 *Consider a queueing network with a memoryless discipline. We conjecture that in this case the fluid limit model* $\overline{\mathcal{Q}}_L$ *satisfies the concatenation property. Unfortunately, this claim has shown some resilience towards attempts of proof.*

Due to fact that $\overline{\mathcal{Q}}_L$ is closed by definition we would obtain the following result.

Conjecture *The fluid limit model of a "memoryless" discipline defines a GFN model satisfying* (e). *It is stable if and only if it admits an upper semicontinuous Lyapunov function.*

The conjecture holds true for the systems considered in [68], but unfortunately, the interesting fluid limits do not have unique trajectories. As to the question of under which conditions fluid limit models satisfy condition (f) we dare not venture a conjecture.

FIFO Fluid Networks

Recall that in FIFO fluid networks the fluids are served in the order of their arrivals. To describe the evolution of class k fluids we have to consider the immediate workload $W(t) = C\,M^{-1}\,Q(t)$ of the stations. For any time t all jobs that arrive later than t have lower priority in the FIFO discipline. Hence, fluids arriving at time t are served at time $t + W_j(t)$. The total arrivals up to time t are given by

$$A(t) = \alpha t + P^{\mathsf{T}}\,M\,T(t).$$

The characteristic equation of FIFO fluid networks can be represented for each class $k \in \{1, ..., K\}$ by the following relation of the allocation process and the immediate workload process

$$T_k(t + W_j(t)) = m_k(Q_k(0) + A_k(t)), \tag{5.35}$$

where $j = c(k)$ and $m_k = \mu_k^{-1}$. Note that the fluid network is not completely determined by the initial fluid level $Q(0)$ as it has to be specified in which order the initial fluid level is served in the time period $[0, W_j(0)]$. So, the initial data for each class $k \in \{1, ..., K\}$ is given by $Q(0)$ and

$$\{T_k(s)\ :\ s \in [0, W_j(0)]\,\}.$$

The dynamic equations describing FIFO fluid networks are given by

$$Q(t) = Q(0) + \alpha\,t - (I - P^{\mathsf{T}})MT(t) \geq 0, \tag{5.36}$$

$$T(0) = 0 \text{ and } \dot{T}(\cdot) \geq 0, \tag{5.37}$$

$$I(t) = et - C\,T(t) \text{ and } \dot{I}(\cdot) \geq 0, \tag{5.38}$$

$$0 = (C\,Q(t))^{\mathsf{T}}\ \dot{I}(t), \ \text{ for almost all } \ t \geq 0, \tag{5.39}$$

$$T_k(t + W_{c(k)}(t)) = m_k(Q_k(0) + A_k(t)),\ \forall\, t \geq 0,\ k = 1, ..., K. \tag{5.40}$$

In the following we will investigate whether the results of Chapter 4 are applicable to FIFO fluid networks. For this reason, in a first step we recapitulate that FIFO fluid network define closed GFN models.

Proposition 5.2.14 ([86]) *A FIFO fluid network defines a closed GFN model.*

Proof. The Lipschitz property, the scaling and shift property also holds true for (5.40) follow immediately. Thus, what is left to show is the closedness property. To avoid double subscripts we denote the sequence by superscripts.

Let $(Q^n(\cdot), T^n(\cdot))_{n\in\mathbb{N}}$ be a sequence of fluid solutions to (5.36)-(5.40) that converges u.o.c. to $(Q^*(\cdot), T^*(\cdot))$. Then, since all processes are continuous we have for all $t \geq 0$ that

$$\lim_{n\to\infty} T_k^n(t + W_j^n(t)) = T_k^*(t + W_j^*(t))$$

as well as

$$\lim_{n\to\infty} A^n(t) = \lim_{n\to\infty} \alpha t + P^\mathsf{T} M\, T^n(t) = \alpha t + P^\mathsf{T} M\, T^*(t) = A^*(t)$$

and the limits satisfy (5.40). ■

However, the fluid networks under FIFO discipline differ from the previous fluid models. One reason for this is the following. Consider the flow balance equation in its differential form,

$$\dot{Q}(t) = \alpha - (I - P^\mathsf{T})\, M\dot{T}(t).$$

The corresponding differential version of the FIFO equation (5.40) is given by

$$\dot{T}_k(t + W_j(t))\ (1 + \dot{W}_j(t)) = m_k\alpha_k - m_k \sum_{l=1}^{K} p_{lk}\mu_l \dot{T}_l(t).$$

That is, allocation process has to satisfy a functional differential equation of neutral type [48]. Consequently, FIFO fluid networks do not fit into the framework of differential inclusions. In fact, functional differential inclusions are appropriate for FIFO fluid networks.

Example 5.2.15 *We consider a single station network serving two classes. The arrival rates α, the service capacities μ and the routing are defined in Figure 5.2. The flow balance equation in this case is given by*

$$Q(t) = Q(0) + \begin{pmatrix} \frac{1}{2} \\ 0 \end{pmatrix} t - \begin{pmatrix} 1 & 0 \\ -1 & 2 \end{pmatrix} T(t),$$

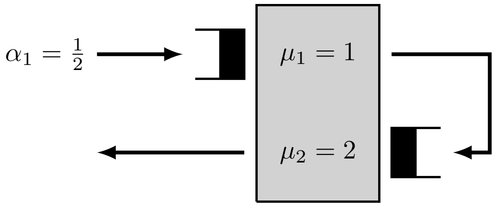

Figure 5.2: A single station network serving two fluid classes.

and the immediate workload is $W(t) = Q_1(t) + \frac{1}{2}\,Q_2(t)$. *Let the initial fluid level be* $Q(0) = \begin{pmatrix}1 & 0\end{pmatrix}^{\mathsf{T}}$. *Then, the initial immediate workload is* $W(0) = 1$. *To describe the dynamic behavior we need to determine the allocation process on the interval* $[0, W(0)] = [0,1]$. *For instance, let the initial allocation be given by*

$$T(t) = \begin{pmatrix}1\\0\end{pmatrix} t, \qquad t \in [0,1].$$

Then, the fluid level process is on the interval $[0,1]$ *given by*

$$Q(t) = \begin{pmatrix}1\\0\end{pmatrix} + \begin{pmatrix}\frac{1}{2}\\0\end{pmatrix} t - \begin{pmatrix}1 & 0\\-1 & 2\end{pmatrix}\begin{pmatrix}t\\0\end{pmatrix} = \begin{pmatrix}1-\frac{1}{2}t\\t\end{pmatrix},$$

and at time $t = 1$ *we have* $Q(1) = \begin{pmatrix}\frac{1}{2} & 1\end{pmatrix}^{\mathsf{T}}$.

Moreover, we consider another initial fluid level $Q'(0) = \frac{1}{2}\begin{pmatrix}1 & 3\end{pmatrix}^{\mathsf{T}}$. *The corresponding initial immediate workload is* $W'(0) = \frac{5}{4}$. *In this case let the initial allocation process be given by*

$$T'(t) = \tfrac{1}{2}\begin{pmatrix}1\\1\end{pmatrix} t, \qquad t \in [0,\tfrac{5}{4}].$$

Hence, the fluid level process on the interval $[0,\frac{5}{4}]$ *is given by*

$$Q'(t) = \tfrac{1}{2}\begin{pmatrix}1\\3\end{pmatrix} + \begin{pmatrix}\frac{1}{2}\\0\end{pmatrix} t - \begin{pmatrix}1 & 0\\-1 & 2\end{pmatrix}\tfrac{1}{2}\begin{pmatrix}t\\t\end{pmatrix} = \tfrac{1}{2}\begin{pmatrix}1\\3-t\end{pmatrix},$$

and at time $t = 1$ *we have that* $Q'(1) = \begin{pmatrix} \frac{1}{2} & 1 \end{pmatrix}^{\mathsf{T}}$. *However, despite the fact* $Q(\cdot)$ *and* $Q'(\cdot)$ *coincide at* $t = 1$*, they have different history. Thus, a concatenation of the trajectories* $Q(\cdot)$ *and* $Q'(\cdot)$ *at* $t = 1$ *is not possible.*

This example shows that for FIFO fluid networks the concatenation is not immediately possible. In this context the initial data $\{T_k(s) : s \in [0, W_j(0)]\}$ plays a key role, see also Example 4.1.12.

5.3 Stability of Queueing Networks Revisited

In this section we exhibit how the converse Lyapunov Theorem 4.2.7 can be used to give a new proof of Theorem 2.4.10 if the associated fluid network defines a strict GFN model. The road map is as follows. In this section we consider HL queueing networks whose associated fluid network defines a strict GFN model. Then, if the associated fluid network is stable it admits a continuous Lyapunov function. Based on the Lyapunov function of the associated fluid network we define a Foster-Lyapunov function for the underlying Markov process of the HL queueing network. Finally, we apply a version of the Foster-Lyapunov criterion and conclude that the underlying Markov process is positive Harris recurrent. But before beginning, we recall some properties of the components for the scaled versions of the strong Markov process $\mathsf{X}(\cdot)$.

Lemma 5.3.1 ([28]) *Given a HL queueing network whose fluid limit model is stable, then it holds that*

$$\lim_{n\to\infty} \tfrac{1}{r_n} \|U^{x_n}(r_n\tau)\| = \lim_{n\to\infty} \tfrac{1}{r_n} \|V^{x_n}(r_n\tau)\| = 0$$

and the following families

$$\left\{\tfrac{1}{r_n} Q^{x_n}(r_n\tau),\, n \in \mathbb{N}\right\},\ \left\{\tfrac{1}{r_n} U_k^{x_n}(r_n\tau),\, n \in \mathbb{N}\right\},\ \left\{\tfrac{1}{r_n} V_k^{x_n}(r_n\tau),\, n \in \mathbb{N}\right\}$$

are uniformly integrable.

Proof. See [17] Section 4.4. ∎

We will use the subsequent formulation of the Foster-Lyapunov criterion for positive Harris recurrence, which is close to Proposition 4.5 in [17]. Here, we consider the decrease condition for the expectation evaluated at $W(X(cW(x)))$, rather than the expectation of the norm of $X(W(x))$. The proof is almost along the lines of the one for Proposition 4.5 in [17].

Theorem 5.3.2 *Suppose that X is continuous time Markov process, such that there exist $\varepsilon > 0$, $\kappa > 0$, $c > 0$, and measurable function $W : \mathcal{X} \to \mathbb{R}$ with $W(x) \geq \delta > 0$ and*

$$\mathbb{E}_x[\, W(X(c\,W(x)))\,] \leq \max\{W(x), \kappa\} - \varepsilon W(x) \tag{5.41}$$

for all $x \in \mathcal{X}$. Then, for all x,

$$\mathbb{E}_x[\, \tau_B(\delta)\,] \leq \tfrac{1}{\varepsilon} \max\{W(x), \kappa\},$$

where $B := \{x \in \mathcal{X} \,:\, W(x) \leq \kappa\}$. In particular, if B is a closed petite set, X is positive Harris recurrent.

Proof. Let $\kappa > 0, \varepsilon > 0$ and $c > 0$ and let $(T_n)_{n \in \mathbb{N}}$ denote a sequence of stopping times defined by $T_0 := 0$ and

$$T_{n+1} := T_n + c\,W(X(T_n)). \tag{5.42}$$

We abbreviately denote by $\mathcal{F}_n := \mathcal{F}_{T_n}$ the σ-algebra corresponding to the stopping time T_n. The strong Markov property and condition (5.41) imply that

$$\begin{aligned}\mathbb{E}_x[W(X(T_n)) \,|\, \mathcal{F}_{n-1}] &= \mathbb{E}_x[\, W(\, X(\, T_{n-1} + c\,W(X(T_{n-1})\,)\,)\,)|\, \mathcal{F}_{n-1}] \\ &= \mathbb{E}_{X(T_{n-1})}[\, W(X(\, c\,W(X(T_{n-1})\,)\,)\,)] \\ &\leq \max\{W(X(T_{n-1})), \kappa\} - \varepsilon W(X(T_{n-1})).\end{aligned} \tag{5.43}$$

Further, let $M(0) := \max\{W(x), \kappa\}$ and for $n \geq 1$ we define

$$M(n) := cW(X(T_n)) + \varepsilon T_n. \tag{5.44}$$

Also, we note that $T_n \in \mathcal{F}_{n-1}$.

Next, we show that for $N = \inf\{n \in \mathbb{N} \,:\, M(n) \in B\}$ and for all $n \leq N$ we have

$$\mathbb{E}_x[\, M(n) \,|\, \mathcal{F}_{n-1}\,] \leq M(n-1).$$

To see this, first note that for $n \leq N$ it holds that $\max\{W(X(T_{n-1})), \kappa\} = W(X(T_{n-1}))$ as well as $T_n \in \mathcal{F}_{n-1}$. Moreover, using (5.42) and (5.43) it holds that

$$\begin{aligned}\mathbb{E}_x[\, M(n) \,|\, \mathcal{F}_{n-1}\,] &= \mathbb{E}_x[\, cW(X(T_n)) + \varepsilon T_n \,|\, \mathcal{F}_{n-1}\,] \\ &= c\,\mathbb{E}_x[\, W(X(T_n)) \,|\, \mathcal{F}_{n-1}\,] + \varepsilon \mathbb{E}_x[\, T_n \,|\, \mathcal{F}_{n-1}\,] \\ &\leq c \max\{W(X(T_{n-1})), \kappa\} - \varepsilon W(X(T_{n-1})) + \varepsilon \mathbb{E}_x[\, T_n \,|\, \mathcal{F}_{n-1}\,] \\ &= cW(X(T_{n-1})) - \varepsilon (T_n - T_{n-1}) + \varepsilon \mathbb{E}_x[\, T_n \,|\, \mathcal{F}_{n-1}\,] = M(n-1).\end{aligned}$$

The validity of the last equality follows from the basic properties of stopping times and expectations. Hence, $M(\min\{n, N\})$ is a supermartingale on $\mathcal{F}_n$. Moreover, by the Optional Sampling Theorem 1.2.2 we have that

$$\mathbb{E}_x[\, M(N)\,] \leq \mathbb{E}_x[\, M(0)\,] = \max\{W(x), \kappa\}. \tag{5.45}$$

Besides, since $W(X(T_n)) \leq M(n)$ it also holds that $\tau_B(\delta) \leq T_N$ and together with (5.44), (5.45) it follows that for all $x \in \mathcal{X}$ we have

$$\varepsilon\, \mathbb{E}_x[\, \tau_B(\delta)\,] \leq \mathbb{E}_x[\, M(N)\,] \leq \max\{W(x), \kappa\}$$

and, hence, we have $\mathbb{P}_x[\tau_B < \infty] = 1$ for all $x \in \mathcal{X}$ and

$$\sup_{x \in B} \mathbb{E}_x[\, \tau_B(\delta)\,] \leq \tfrac{\kappa}{\varepsilon}.$$

The assertion then follows from Theorem 1.4.1. ■

In the sequel, we give a new proof of the fact that the stability of the associated fluid network is sufficient for the positive Harris recurrence of the multiclass queueing network, which is the content of the second part in Theorem 2.4.10. The main tools for the verification are the Lyapunov function admitted by the stable fluid network and the Foster-Lyapunov theorem stated above. The line of reasoning is similar to [46].

Theorem 5.3.3 *Let a queueing discipline be fixed such that the associated fluid network defines a strict GFN model. Assume that the Assumptions 2.1.2 hold. If the associated fluid network is stable, then the queueing network is stable.*

Proof. Since $\mathcal{FLM} \subset \mathcal{FN}$ and the associated fluid network is stable, there is a $\tau > 0$ such that for all $\overline{Q} \in \mathcal{FLM}$ it holds that $\overline{Q}(t) = 0$ for all $t \geq \tau \|\overline{Q}(0)\|$. Also, from Corollary 4.2.9 there is a continuous Lyapunov function V_L and class $\mathcal{K}_\infty$ functions w_i, $i = 1, 2, 3$ such that for $q \in \mathbb{R}^K_+$ we have that

$$\begin{aligned} w_1(\|q\|) \leq V_L(q) &\leq w_2(\|q\|) \\ \dot{V}_L(\overline{Q}(t)) &\leq -\, w_3(\|\overline{Q}(t)\|). \end{aligned}$$

Further, let $(r_n, x_n)_{n \in \mathbb{N}}$ be a sequence of pairs satisfying (2.9). Then, along a subsequence, which is also indexed by n, it holds that

$$\tfrac{1}{r_n} Q^{x_n}(r_n t) \to \overline{Q}(t) \qquad \text{u.o.c.}$$

as $n \to \infty$. In particular, the stability of the fluid limit model implies that

$$\tfrac{1}{r_n} Q^{x_n}(r_n \tau) \to 0.$$

That is, for any $\tilde{\varepsilon} \in (0,1)$ there is a $N \in \mathbb{N}$ such that for all $n > N$ we have

$$\tfrac{1}{r_n} w_2^{-1}(V_L(Q^{x_n}(r_n \tau))) \leq \tilde{\varepsilon}.$$

Moreover, since $w_2^{-1}(V_L(Q^{x_n}(r_n t))) \leq \|Q^{x_n}(r_n t)\|$ and by Lemma 5.3.1 we have that

$$\left\{ \tfrac{1}{r_n} w_2^{-1}(V_L(Q^{x_n}(r_n \tau))),\, n \in \mathbb{N} \right\}$$

is uniformly integrable. In addition, it also holds that

$$\lim_{n\to\infty} \tfrac{1}{r_n} \|U^{x_n}(r_n \tau)\| = \lim_{n\to\infty} \tfrac{1}{r_n} \|V^{x_n}(r_n \tau)\| = 0.$$

The families $\left\{ \tfrac{1}{r_n} U_k^{x_n}(r_n \tau),\, n \in \mathbb{N} \right\}$ and $\left\{ \tfrac{1}{r_n} V_k^{x_n}(r_n \tau),\, n \in \mathbb{N} \right\}$ are for each $k \in \{1, ..., K\}$ uniformly integrable by Lemma 5.3.1. Hence, we have

$$\limsup_{n\to\infty} \tfrac{1}{r_n} \mathbb{E}\left[w_2^{-1}(V_L(Q^{x_n}(r_n \tau))) + \|U_k^{x_n}(r_n \tau)\| + \|V_k^{x_n}(r_n \tau)\| \right] \leq \tilde{\varepsilon}.$$

Thus, there is a $\kappa > 0$ such that for all $r_n > \kappa$ we have

$$\mathbb{E}\left[w_2^{-1}(V_L(Q^{x_n}(r_n \tau))) + \| U^{x_n}(r_n \tau) \| + \| V^{x_n}(r_n \tau) \| \right] \leq r_n \tilde{\varepsilon}.$$

We define the Foster-Lyapunov function $W : \mathcal{X} \to \mathbb{R}_+$ by

$$W(x) := w_2^{-1}(V_L(q)) + \|u\| + \|v\|.$$

Then, for all x with $W(x) > \kappa$ it follows that

$$\mathbb{E}_x \left[W(\mathsf{X}(W(x)\, \tau)) \right] \leq \tilde{\varepsilon}\, W(x)$$

and consequently,

$$\mathbb{E}_x \left[W(\mathsf{X}(\tau\, W(x))) \right] \leq \max\{W(x), \kappa\} - \varepsilon\, W(x).$$

Finally, we have to show that $B = \{x : W(x) \leq \kappa\}$ is closed and petite. To see that B is petite, note that by Lemma 2.4.8 the set $A = \{x \in \mathcal{X} : |x| \leq \kappa\}$ is closed and petite. Furthermore, since $w_2^{-1}(V_L(q)) \leq \|q\|$ it holds that $B \subset A$. In addition, the continuity of w_2^{-1} and V_L imply that B is closed. As subsets of petite sets are petite, the assertion then follows from Theorem 5.3.2. ∎

5.4 The Linear Skorokhod Problem

Another approximation of a multiclass queueing network is the so called diffusion limit. This limit can be regarded as a *semi-martingale reflected Brownian motion* (SRBM). Similar to the fluid limit, a sufficient condition for the stability of the SRBM is the stability of the *linear Skorokhod problem* (LSP) [41]. The following description of the linear Skorokhod problem is taken from [20] and [86]. Let R be a $J \times J$ matrix, $\theta \in \mathbb{R}^J$ and $Z_0 \in \mathbb{R}^J_+$. The pair $(Z, Y) \in C(\mathbb{R}_+, \mathbb{R}^J_+) \times C(\mathbb{R}_+, \mathbb{R}^J_+)$ is said to solve the LSP (θ, R) with initial state Z_0 if they jointly satisfy

$$Z(t) = Z_0 + \theta t + RY(t) \geq 0, \tag{5.46}$$

$$Y(0) = 0 \text{ and } Y(\cdot) \text{ is nondecreasing,} \tag{5.47}$$

$$0 = \int_0^\infty Z_j(t)\, \mathrm{d}Y_j(t), \qquad j = 1, ..., J. \tag{5.48}$$

The first question that arises is, which conditions guarantee the existence of a solution of the LSP(θ, R). In order to state such a condition recall that a $J \times J$ matrix R is said to be an *S-matrix* if there exists an $x \geq 0$ such that $Rx > 0$. Further, R is said to be *completely-S* if all of its principal submatrices are S-matrices.

Theorem 5.4.1 ([8]) *The LSP(θ, R) has a solution $(Z(\cdot), Y(\cdot))$ if and only if the matrix R is completely-S.*

Proof. See [8] Theorem 1. ∎

We define

$$\mathcal{Q}_{LSP} := \{Z(\cdot) : \exists Y(\cdot) \text{ such that } (Z(\cdot), Y(\cdot)) \text{ satisfy } (5.46) - (5.48)\}.$$

Note that Theorem 5.4.1 states only the existence of a solution. In general, the solutions to the LSP(θ, R) need not to be unique, for a counterexample see e.g. [8].

Definition 5.4.2 *A solution $Z \in C(\mathbb{R}_+, \mathbb{R}^J_+)$ is said to be attracted to the origin if for any $\varepsilon > 0$ there is a $T < \infty$ such that $t \geq T$ implies $\|Z(t)\| < \varepsilon$. A LSP(θ, R) is said to be stable if for each $Z_0 \in \mathbb{R}^J_+$ any solution $Z(\cdot)$ with $Z(0) = Z_0$ is attracted to the origin.*

To ensure that $\mathcal{Q}_{LSP}$ is nonempty, Theorem 5.4.1 states that R has to be completely-S. In [86, Theorem 5.2] it is shown that in this case Definition 5.4.2 is equivalent to Definition 4.1.3. To derive a necessary and sufficient condition for stability of the linear Skorokhod problem we have to show that $\mathcal{Q}_{LSP}$ defines a strict GFN model. The next lemma states that $\mathcal{Q}_{LSP}$ defines a closed GFN model.

Lemma 5.4.3 ([8]) *If the matrix R is completely-S, then $\mathcal{Q}_{LSP}$ defines a closed GFN model.*

Proof. A proof of the Lipschitz property can be found in [8] Lemma 1. The scaling and shift property is shown in [42] Section 2. The closedness is shown in Proposition 1 in [8]. ■

Hence, in order to conclude a converse Lyapunov theorem it remains to investigate whether $\mathcal{Q}_{LSP}$ satisfies the concatenation and the lower semicontinuity property. Again, we bring the linear Skorokhod problem into the context of differential inclusions. That is, let $\dot{Y}(t) = u$ and

$$F(Z) = \{\theta + Ru \,:\, u \in U_{LSP}(Z)\}\,, \tag{5.49}$$

where the set U_{LSP} of admissible values of $u \in \mathbb{R}^J$ is determined by the conditions

$$u \geq 0, \quad Z^{\mathsf{T}} u = 0. \tag{5.50}$$

While it is clear that the set described by (5.50) is unbounded on the boundary of the positive orthant, Lemma 5.4.3 may be used to see that the effective set of values is bounded. Indeed from the Lipschitz continuity of solutions, only values of u below a certain bound need to be considered in (5.49). The corresponding differential inclusion is of the form

$$\dot{Z}(t) \in F(Z(t)), \qquad Z(0) = Z_0. \tag{5.51}$$

It can be seen that the right-hand side is upper semicontinuous and the set $F(Z)$ is nonempty, convex and compact. The latter is a consequence of the Lipschitz continuity of the solutions of linear Skorokhod problem. In addition, it holds that $F(Z) \cap \mathcal{T}_{\mathbb{R}^J_+}(Z) \neq \emptyset$ for all $Z \in \mathbb{R}^J_+$. Again arguments from the theory of differential inclusions show the validity of the concatenation property. Hence, Corollary 4.2.8 is applicable for the linear Skorokhod problem.

Theorem 5.4.4 *The LSP(θ, R) defines a closed GFN model satisfying the concatenation property. It is stable if and only if it admits an upper semicontinuous Lyapunov function.*

We note that it is not obvious whether the right-hand side is also lower semicontinuous.

Remark 5.4.5 *The consequence of the above theorem is, that the main theorem is applicable for the linear Skorokhod problem. However, for the provided Lyapunov function we can only show upper semicontinuity.*

5.5 Notes and References

The transfer of fluid networks to the framework of differential inclusions is very natural. Standard references for differential inclusions and set-valued analysis are, for instance, [3, 4, 5] and [80]. On one hand the consideration fluid networks as differential inclusions allows to give a new and elegant proof for the existence of a work-conserving allocation process, given the parameters (α, μ, P, C) of the network and the initial condition $Q(0)$. In addition, the validity of the concatenation property follows immediately by using known results from differential inclusions. On other hand, the lower semicontinuity property needs to be established directly from the definition.

In the case that a fluid network under a certain discipline can be modelled in terms of a differential inclusion such that $Q \rightsquigarrow F(Q) \cap \mathcal{T}_{\mathbb{R}_+^K}(Q)$ is upper semicontinuous and has nonempty, compact and convex values, we can apply the smooth converse Lyapunov theorem B.4 to conclude that in this case a smooth converse Lyapunov theorem holds. However, as stated in Remark 5.2.2 the set-valued map $Q \rightsquigarrow F(Q) \cap \mathcal{T}_{\mathbb{R}_+^K}(Q)$ is not upper semicontinuous in general.

Despite all that, there is an interesting relation to the construction of smooth Lyapunov functions we presented in Section 4.3. That is, we have to suppose that the strict GFN model satisfies Assumption 4.3.1 to undertake that the construction will work. The content of Assumption 4.3.1 is that for any trajectory $Q(\cdot) \in \mathcal{Q}$ with initial value $Q(0) = x \in \mathbb{R}_+^K$ and any $y \in \mathbb{R}^K$ there is a trajectory $R(\cdot) \in \mathcal{Q}$ with initial value $|x - y|_{\text{vec}}$ such that the difference $\|Q(t) - y - R(t)\|$ of the trajectories grows locally at most linear in $\|y\|$. This resembles the content of the well known Gronwall inequality.

To prove the smooth converse Lyapunov Theorem B.4 for differential inclusions Clarke, Ledyaev and Stern constructed, for a given right-hand side F, a locally Lipschitz set-valued map F_L that keeps the asymptotic stability and satisfies $F(x) \subset F_L(x)$ for all x. The smooth Lyapunov function is then constructed for F_L. In doing so, for the set-valued map F_L, based on a Gronwall argument, an estimate for the difference between trajectories is available. The desired decrease condition and finally the smooth converse Lyapunov Theorem B.4 for differential inclusions is then obtained from this estimate. In particular, it follows from [26] that if for fluid network the right-hand side $Q \rightsquigarrow F(Q) \cap \mathcal{T}_{\mathbb{R}^K_+}(Q)$ is upper semicontinuous, then Assumption 4.3.1 is satisfied.

Besides that, the differential inclusions approach strengthens the special role of FIFO fluid networks among the disciplines considered in this thesis. As stated in Section 5.2 the appropriate differential framework for FIFO fluid networks are functional differential inclusions, cf. [3]. Already for functional differential equations it is general not possible to define Lyapunov functions as in Definition 3.6.3.

In the literature there are various formulations of Foster-Lyapunov criteria, see e.g. [17, 46, 61, 62, 63] and the references therein. This list is far from being complete. The Foster-Lyapunov theorem 5.3.2 is a slightly modified version of Proposition 4.5 in [17]. This adaptation is appropriate to provide a different proof for the second part of Theorem 2.4.10 in terms of the Lyapunov function admitted by the stable associated fluid network. The line of argument is similar to the one for Theorem 2 in [46].

Parts of this chapter have already been published in [75] and [76].

6 Robust Stability of Fluid Networks

In this chapter we will stress the fact that if a real system is considered, the distributions and their mean values may not be known precisely. The model for the real system obtained from a best guess of the parameters is called the *nominal network*. The analysis of the nominal associated fluid network is based on the mean values of the primitive increments of the multiclass queueing network. So, the basic question of this chapter is, how perturbations in the mean values representing the arrival rates and service capacities influence the stability of the network. Throughout this chapter we make the standing assumption that the topology of the nominal network is known precisely, i.e. the constituency matrix C and the routing matrix P are fixed.

In Chapter 3 we discussed stability conditions for fluid networks under general work-conserving, priority, HLPPS and FIFO disciplines. Given a certain discipline π the set of arrival rates and service capacities that lead to a stable fluid network is denoted by

$$\mathcal{D}_\pi := \{\, (\alpha, \mu) \in \mathbb{R}_+^K \times \mathbb{R}_{>0}^K \;:\; \text{the fluid network } (\alpha, \mu, P, C, \pi) \text{ is stable} \,\}.$$

In the following, we call $\mathcal{D}_\pi$ the *stability region* of a fluid network with topology (P, C) under the discipline π. According to Theorem 3.2.2 we call

$$\mathcal{D}_0 := \{\, (\alpha, \mu) \in \mathbb{R}_+^K \times \mathbb{R}_{>0}^K \;:\; \rho < e \,\}$$

the *nominal stability region*. Moreover, for general work-conserving fluid networks we call $\mathcal{D}_\infty$ the *global stability region*. The stability results in Chapter 3 show that the stability regions of the disciplines considered in this thesis are related as follows.

$$\mathcal{D}_\infty \subset \mathcal{D}_\pi \subsetneq \mathcal{D}_{\pi'} = \mathcal{D}_0, \tag{6.1}$$

where $\pi = \{\text{FIFO, priority}\}$ and $\pi' = \{\text{HLPPS, FIFO of Kelly type}\}$.

In the present chapter we will derive a quantitative approach to robust stability. To this end, we introduce a measure for the perturbations that is indicated by a single real number. This number will be called the *stability radius.* More precisely, given a nominal fluid network (α, μ, P, C) and a set of feasible perturbations, the stability radius represents the magnitude of the smallest shift of the arrival rates and/or service capacities so that the fluid network looses the property of stability. A formal definition is given in Section 6.1. In addition, the first section contains some results from convex analysis that will be used in this chapter.

In Section 6.2 we consider fluid networks where the arrival rates are subject to perturbations. The third section is devoted to the investigation of the situation where the service capacities are disturbed. In the subsequent section we combine the latter cases. That is, we consider fluid networks that are subject to perturbations of the arrival rates and the service capacities.

6.1 First Steps

To analyze the robustness of a fluid network (α, μ, P, C) under a discipline π we consider a perturbed fluid network $(\alpha, \mu, P, C, \pi, \Delta)$, where $\Delta \in \mathbb{R}^n_+$ denotes a feasible perturbation. At this point we do not yet specify the perturbation. In order to define a measure for the perturbation we define for $\gamma \in \mathbb{R}^K_{>0}$ the γ-weighted norm by $\|x\|_\gamma := \sum_{k=1}^K |\gamma_k\, x_k|$. Here $\mathbb{R}^K_{>0}$ denotes the positive orthant, i.e. $\mathbb{R}^K_{>0} = \{x \in \mathbb{R}^K_+ \,:\, x_i > 0\, \forall\, i = 1, ..., K\}$. For the special case $\gamma = e$ we simply write $\|\cdot\|$.

Definition 6.1.1 *Let $\gamma \in \mathbb{R}^K_{>0}$ be fixed. The γ-weighted stability radius of the fluid network (α, μ, P, C, π) is defined by*

$$r_\gamma(\alpha, \mu, P, C, \pi) := \inf\{\|\Delta\|_\gamma : \Delta \textit{ is feasible, } (\alpha, \mu, P, C, \pi, \Delta) \textit{ is unstable}\}.$$

Here we note that specifying the stability region is challenging. Furthermore, in many cases analytic descriptions of the stability region are not available. Therefore, a calculation of the stability radius for such disciplines is not possible. Nevertheless, for some disciplines and networks of special structure there are conditions at hand which determine the stability region precisely. For instance, the nominal workload condition constitutes the stability region of fluid networks under HLPPS and FIFO of Kelly type disciplines, see Theorem 3.4.2 and the subsequent remark.

In the queueing literature the analysis of the global stability region has received considerable attraction over the last 15 years. Chen showed that the global stability region is monotone with respect to the arrival rates, cf. [19, Theorem 3.6]. This means that if the fluid network (α, μ, P, C) is globally stable, then the fluid network (α', μ, P, C) is globally stable provided that $\alpha' \leq \alpha$. For two station fluid networks the global stability region is monotone with respect to the arrival rates and service capacities, see [37, Corollary 1]. However, counterexamples given by Dumas and Bramson show that, in general, the stability region is neither convex nor monotone with respect to the service rates, see [40] and [14], respectively.

Apart from that, for HLPPS fluid networks, FIFO fluid networks of Kelly type stability can be equivalently characterized by the nominal workload condition

$$\rho < e. \tag{6.2}$$

Thus, in the remainder of this chapter we concentrate on the stability region $\mathcal{D}_0$ determined by the nominal workload condition (6.2). As stated in Theorem 3.2.2 the latter provides a necessary condition for stability for all work-conserving disciplines. Thus, the nominal workload condition allows for a derivation of an upper bound of the stability radius that is valid for any work-conserving discipline. In particular, this bound is tight for disciplines, in which condition (6.2) is also sufficient for stability. Referring to the nominal stability region $\mathcal{D}_0$ we will denote the upper bound by $r^0_\gamma(\alpha, \mu, P, C)$ and call it the *nominal stability radius*.

As the nominal stability radius is not affected by the particular discipline, we omit the letter π whenever the nominal stability radius is considered. In terms of the nominal workload condition (6.2) the nominal stability radius is given by

$$r^0_\gamma(\alpha, \mu, P, C) = \inf \{ \ \|\Delta\|_\gamma \ : \ \Delta \text{ is feasible and } \rho(\Delta) \not< e \ \}, \tag{6.3}$$

where $\rho(\Delta)$ denotes the nominal workload of the fluid network $(\alpha, \mu, P, C, \Delta)$ that is subject to a perturbation Δ. Moreover, for $x, y \in \mathbb{R}^K_+$ the notion $x \not< y$ means that $x_i \geq y_i$ for at least one component $i \in \{1, ..., K\}$. This reflects the fact that for some station $j \in \{1, ..., J\}$ the nominal workload is at least one and, thus, the network is unstable.

Before we will present a scheme for the calculation of the nominal stability radius r^0_γ, we have to collect some notations and preliminary results concerning convex analysis. A set $A \subset \mathbb{R}^n$ is called *convex* if $(1-c)x + cy \in A$ whenever $x, y \in A$ and $c \in (0,1)$. A point z of a convex set A is called an *extreme point* if there is no way to express z as a convex combination $(1-c)x + cy$ such that $x, y \in A$ are distinct and $c \in (0,1)$. The set of all extreme points of A is denoted by $\mathrm{ext}(A)$. The *convex hull* of a set A, denoted by $\mathrm{conv}(A)$, is the smallest convex set that contains A. The *boundary* of a set A is denoted by ∂A. The *interior* of a set A is denoted by $\mathrm{int}(A) = \overline{A} \setminus \partial A$.

The subsequent result provides a relation of convex compact sets and their extreme points.

Theorem 6.1.2 *Any compact convex set $C \subset \mathbb{R}^n$ is the convex hull of its extreme points.*

This result is called Minkowski's Theorem in finite dimensions, cf. [12, Theorem 4.1.8]. We will use Minkowski's Theorem to show the following auxiliary result.

Proposition 6.1.3 *Given a closed and convex set $B \subset \mathbb{R}^n$ and let $A \subset B$ be convex and compact such that $\partial A \cap \partial B \neq \emptyset$. Then $ext(A) \cap \partial B \neq \emptyset$.*

Proof. The assertion is shown by contradiction. So assume that $\mathrm{ext}(A) \cap \partial B = \emptyset$. This implies that $\mathrm{ext}(A) \subset \mathrm{int}(B)$. Further, by Theorem 6.1.2 it holds that $A = \mathrm{conv}(\mathrm{ext}(A)) \subset \mathrm{int}(B)$, which is a contradiction to $\partial A \cap \partial B \neq \emptyset$. ■

Let C be a convex set in $\mathbb{R}^n$. A function $f : C \to (-\infty, \infty]$ is *convex* if and only if for every x and y in C and $\lambda \in (0,1)$ it holds that

$$f(\,\lambda x + (1-\lambda)y\,) \leq \lambda f(x) + (1-\lambda) f(y).$$

The subsequent results can be useful for establishing convexity of certain sets of interest, see [69, Theorem 4.6].

Theorem 6.1.4 *For any convex function f and any $c \in [-\infty, \infty]$, the level sets $\{x : f(x) < c\}$ and $\{x : f(x) \leq c\}$ are convex.*

Throughout the chapter, we consider an academic example consisting of three fluid classes, which are served at two stations.

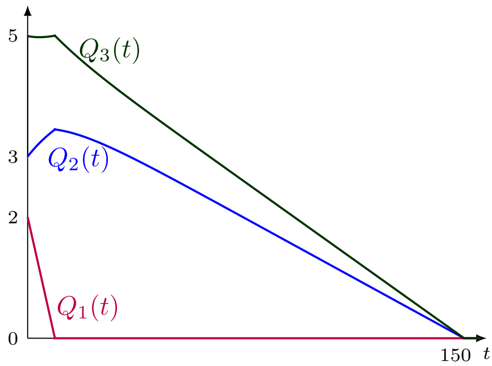

Figure 6.1: Fluid level processes of the HLPPS fluid network with parameters defined in Example 6.1.5. The initial level is $Q(0) = (2\ 3\ 5)^{\mathsf{T}}$.

Example 6.1.5 *We consider a two station fluid network that serves three fluid classes, where fluid class 1 is served at station 1 and station 2 serves the fluid classes 2 and 3. The parameters for the example network are the following. The arrival rates and the service capacities are given by $\alpha = (\,0.15\ 0.15\ 0.10\,)^{\mathsf{T}}$ and $\mu = (\,0.6\ 0.9\ 0.5\,)^{\mathsf{T}}$, respectively. The structure of the network is given by the constituency matrix C and the routing matrix P,*

$$C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}, \quad P = \begin{pmatrix} 0.25 & 0.15 & 0.20 \\ 0.05 & 0.25 & 0.15 \\ 0.20 & 0.25 & 0.10 \end{pmatrix}.$$

The nominal workload of the network (α, μ, P, C) is $\rho = (\,0.472\ 0.829\,)^{\mathsf{T}}$. According to Theorem 3.4.2 the network is stable under the HLPPS discipline. Figure 6.1 shows the fluid level processes to the fluid network $(\alpha, \mu, P, C, HLPPS)$, where the initial levels are $Q(0) = (2\ 3\ 5)^{\mathsf{T}}$. Moreover, Figure 6.1 also illustrates the property that the fluid level processes of classes 2 and 3 empty at the same time, which is in line with Lemma 3.4.1.

6.2 Perturbations of Arrival Rates

In this section we focus on the situation where the service capacities μ, the routing matrix P and the constituency matrix C are fixed, while the arrival rates α are subject to perturbations of the form $\Delta \in \mathbb{R}^K_+$. So, we investigate the stability of the fluid network $(\alpha+\Delta, \mu, P, C)$ in terms of the perturbation Δ. We will derive estimates on the size of Δ that characterize the nominal stability region of the fluid network $(\alpha + \Delta, \mu, P, C)$. The nominal workload of the fluid network $(\alpha + \Delta, \mu, P, C)$ is given by

$$\rho(\Delta) = C\, M^{-1}\,(I - P^{\mathsf{T}})^{-1}\,(\alpha + \Delta).$$

By Theorem 3.2.2, the nominal stability radius can be expressed by

$$r^0_\gamma(\alpha, \mu, P, C) = \min \left\{ \, \|\Delta\|_\gamma \;:\; \Delta \in \mathbb{R}^K_+ \text{ and } \rho(\Delta) \not< e \, \right\}. \tag{6.4}$$

To describe the geometric perspective of the nominal stability radius consider a stable single station fluid network (α, μ, P, C) that serves two classes of fluid. According to Theorem 3.2.2 and the subsequent remark the nominal workload is strictly less than one. In Figure 6.2 the light grey domain represents the nominal stability region $\mathcal{D}_0$. Let $B_\gamma(x, r) = \{z \in \mathbb{R}^K : \|x - z\|_\gamma \leq r\}$. Geometrically the nominal stability radius can be illustrated as the largest neighborhood $B_\gamma(\alpha, r)$ around α that is completely contained in the nominal stability region $\mathcal{D}_0$, where the size of the neighborhood is measured by the γ-weighted norm $\|\cdot\|_\gamma$. Precisely, for arrival rates α' in the interior of $B_\gamma(\alpha, r)$ the fluid network (α', μ, P, C) might be stable, depending on the discipline, while for arrival rates $\alpha' \in \partial B_\gamma(\alpha, r) \cap \partial \mathcal{D}_0$ the fluid network (α', μ, P, C) is definitely unstable no matter which discipline is considered.

In the following, we will derive a scheme for the calculation of the nominal stability radius. Based on the geometric interpretation our approach to calculate (6.4) is by means of an optimization problem. For this reason, one constraint on the perturbation $\Delta \in \mathbb{R}^K_+$ is that for at least one station $j \in \{1, ..., J\}$ of the fluid network the nominal workload $\rho_j(\Delta)$ is at least one. In terms of the fluid network $(\alpha + \Delta, \mu, P, C)$ this can be expressed as

$$\max_{j=1,\ldots,J} \rho_j(\Delta) = 1. \tag{6.5}$$

Consequently, the nominal stability radius can be calculated by an opti-

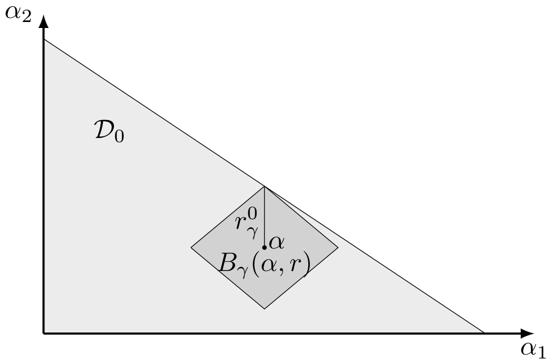

Figure 6.2: Illustration of the nominal stability radius for a single station fluid network serving two classes of fluids.

mization problem of the form

$$\begin{aligned} r_\gamma^0(\alpha, \mu, P, C) = \min \; & \sum_{k=1}^{K} \gamma_k \Delta_k \\ \text{subject to } & \max_{j=1,\dots,J} \rho_j(\Delta) = 1, \\ & \Delta \geq 0. \end{aligned} \tag{6.6}$$

The following statement characterizes the solutions of the optimization problem.

Theorem 6.2.1 *If* $r_\gamma^0 = \|\Delta_*\|_\gamma > 0$ *is a solution to* (6.6)*, then* Δ_* *can be chosen such that at most one component is strictly positive. That is,* $\Delta_* = r_\gamma^0 {\gamma_k}^{-1} e_k$ *for some* $k \in \{1, ..., K\}$*.*

Proof. For brevity, we denote $N := C\, M^{-1}(I - P^\mathsf{T})^{-1}$. Then, the workload condition $\rho(\Delta) \not< e$ can be written as $N\Delta \not< e - \rho$. Since the $J \times K$ matrix N contains only nonnegative entries the left hand side of the above condition is a weighted sum in Δ with positive weights. Hence, a perturbation with minimal γ-norm is of the form $\Delta_* = r_\gamma^0 {\gamma_k}^{-1} e_k$ for some $k \in \{1, ..., K\}$. ■

Using $N(\mu) := C\, M(\mu)^{-1}(I - P^\mathsf{T})^{-1}$ and the fact that a destabilizing perturbation is of the form $\Delta_* = r_\gamma^0 {\gamma_k}^{-1} e_k$ for some $k \in \{1, ..., K\}$, it

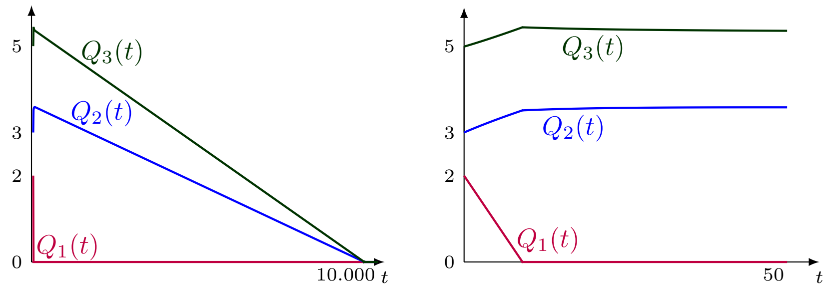

Figure 6.3: The figure on the left shows the fluid level processes to the fluid network $(\alpha + \Delta, \mu, P, C)$ under HLPPS discipline where the parameters (α, μ, P, C) and the perturbation are chosen as in Example 6.1.5 and $\Delta = (0\,0\,0.54)^{\mathsf{T}}$, respectively. On the right the transient behavior of the fluid level processes is shown.

holds for some $j \in \{1, ..., J\}$ that

$$r_\gamma^0 \, {\gamma_k}^{-1} \cdot e_j^{\mathsf{T}} \, N(\mu) \, e_k = 1 - (N(\mu)\alpha)_j \, .$$

Thus, a representation of the nominal stability radius that expresses the dependency on μ is given by

$$r_\gamma^0(\mu) = \min_{j=1,\dots,J} \left\{ \frac{1 - (N(\mu)\alpha)_j}{\max\limits_{k=1,\dots,K} \frac{N(\mu)_{jk}}{\gamma_k}} \right\}. \tag{6.7}$$

Example 6.2.2 *We consider the fluid network given in Example 6.1.5. Using $\gamma = e$, a calculation based on* `fmincon` *in MATLAB yields the nominal stability radius $r^0(\alpha, \mu, P, C) = 0.05478$. For HLPPS fluid networks the nominal stability radius coincides with the stability radius. This situation is illustrated in Figure 6.3 for a perturbation of the form $\Delta = (0\ 0\ 0.054)^{\mathsf{T}}$. Figure 6.3 shows that the fluid levels of classes 2 and 3 are increasing until the fluid level of class 1 is positive. Not before the fluid level of class 1 has reached zero, the fluid levels of class 2 and 3 are decreasing. Moreover, it can be seen, related to Lemma 3.4.1, that the fluid levels of classes 2 and 3 empty at the same time.*

Since $r^0(\alpha, \mu, P, C) = 0.05478$ is the nominal stability radius the question raises what measure of quality it yields. The next example considers the global stability region for the Example 6.1.5.

Example 6.2.3 *Consider the fluid network given in Example 6.1.5. For the perturbation* $\Delta' = (\,0\ 0\ 0.05428\,)^{\mathsf{T}}$ *the matrix*

$$A = \begin{pmatrix} 10 & 0.374 & 0.53 \\ 0.374 & 0.645 & 0.916 \\ 0.53 & 0.916 & 1.3 \end{pmatrix}$$

satisfies the sufficient condition of Theorem 3.2.3. Consequently, the fluid network $(\alpha + \Delta', \mu, P, C)$ *is globally stable.*

6.3 Perturbations of Service Capacities

In this section, we will measure the robustness of a fluid network (α, μ, P, C) with respect to perturbations $\delta \in \mathbb{R}_+^K$ of the service capacities of the form $0 \le \delta < \mu$. The inequalities have to be understood componentwise. For the perturbed fluid network $(\alpha, \mu - \delta, P, C)$ the matrix of service capacities and the nominal workload are denoted by $M(\delta) = \mathrm{diag}(\mu - \delta)$ and $\rho(\delta) = C\, M(\delta)^{-1}\, (I - P^{\mathsf{T}})^{-1}\, \alpha$, respectively. Then, the nominal stability radius can be expressed as

$$r^0_\gamma(\alpha, \mu, P, C) = \min\{\ \|\delta\|_\gamma\ :\ 0 \le \delta < \mu \text{ and } \max_{j=1,\dots,J} \rho_j(\delta) = 1\ \}. \quad (6.8)$$

To come up with a geometric illustration for this scenario we consider once again a single station fluid network serving two classes of fluids. Given the arrival rates α and routing matrix P the effective arrival rates λ are determined. The light grey domain in Figure 6.4 represents the nominal stability region $\mathcal{D}_0$ for fixed arrival rates. By Theorem 3.2.2 service capacities in the interior of the nominal stability region might provide a stable network. Contrary, for service capacities on the boundary $\partial\mathcal{D}_0$ of the nominal stability region the corresponding fluid network will be unstable. Using $B_\gamma(\mu, r) = \{\bar\mu \in \mathbb{R}^2_{>0} : \|\bar\mu - \mu\|_\gamma \le r\}$ the nominal stability radius can be described as the radius of the largest neighborhood $B_\gamma(\mu, r)$ around μ that is completely contained in the nominal stability region $\mathcal{D}_0$ such that at least one edge of $\partial B_\gamma(\mu, r)$ intersects the boundary $\partial\mathcal{D}_0$ of the nominal stability region. In Figure 6.4 the neighborhood $B_\gamma(\mu, r)$ is illustrated by

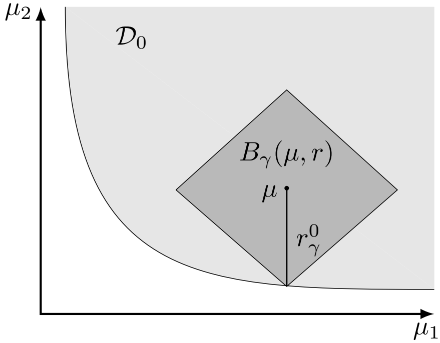

Figure 6.4: Illustration of the nominal stability radius of a single station fluid network with two classes.

the dark grey domain. The nominal stability radius can be calculated by the following optimization problem

$$
\begin{aligned}
r_\gamma^0(\alpha, \mu, P, C) = \min \quad & \sum_{k=1}^{K} \gamma_k \, \delta_k \\
\text{subject to} \quad & \max_{j=1,\dots,J} \rho_j(\delta) = 1, \\
& 0 \le \delta < \mu.
\end{aligned}
\tag{6.9}
$$

To calculate the nominal stability radius we split the problem into J subproblems as follows. We consider the optimization problem (6.9) for each location $j \in \{1, ..., J\}$ individually. The corresponding solution is denoted by r_j. The smallest solution in magnitude represents the solution to (6.9). In the sequel, we describe how to solve the optimization problem for a single location $j \in \{1, ..., J\}$. That is, for each j the task is of the form

$$
\begin{aligned}
r_j := \min \quad & \sum_{k \in C(j)} \gamma_k \, \delta_k \\
\text{subject to} \quad & \sum_{k \in C(j)} \frac{\lambda_k}{\mu_k - \delta_k} = 1, \\
& 0 \le \delta_k < \mu_k, \qquad k \in C(j).
\end{aligned}
\tag{6.10}
$$

Consequently, the nominal stability radius of the fluid network (α, μ, P, C) is given by

$$r^0_\gamma(\alpha, \mu, P, C) = \min_{j=1,\ldots,J} r_j. \tag{6.11}$$

For some $c > 0$ the general form of the optimization problem (6.10) for a single location is

$$\begin{aligned} r_\gamma = \min \quad & \gamma_1\,\delta_1 + \ldots + \gamma_n\,\delta_n \\ \text{subject to} \quad & \frac{\lambda_1}{x_1 - \delta_1} + \ldots + \frac{\lambda_n}{x_n - \delta_n} = c, \\ & 0 \le \delta_k < x_k, \qquad k = 1, 2, \ldots, n. \end{aligned} \tag{6.12}$$

The subsequent statement characterizes the structure of the solutions of the optimization problem (6.12).

Theorem 6.3.1 *If* $r_\gamma = \|\delta_*\|_\gamma > 0$ *is a solution to* (6.12)*, then* δ_* *can be chosen such that at most one component is strictly positive. That is,* $\delta_* = r_\gamma\,\gamma_k^{-1}\,e_k$ *for some* $k \in \{1, 2, \ldots, n\}$*.*

Proof. Assume that $\delta = (\delta_1\,\delta_2\,\ldots\,\delta_n)^\mathsf{T}$ is a solution to (6.12), i.e. $r_\gamma = \|\delta\|_\gamma$ and

$$\frac{\lambda_1}{x_1 - \delta_1} + \frac{\lambda_2}{x_2 - \delta_2} + \ldots + \frac{\lambda_n}{x_n - \delta_n} = c.$$

The statement is shown by induction. For $n = 2$ consider the optimization problem

$$\begin{aligned} \min \quad & \gamma_1\,\delta_1 + \gamma_2\,\delta_2 \\ \text{subject to} \quad & \frac{\lambda_1}{x_1 - \delta_1} + \frac{\lambda_2}{x_2 - \delta_2} = c, \\ & 0 \le \delta_k < x_k, \qquad k = 1, 2. \end{aligned} \tag{6.13}$$

For given λ_1, λ_2 and $c > 0$ the set

$$\mathcal{M} = \{y = (y_1, y_2) \in \mathbb{R}^2_{>0} \,:\, \frac{\lambda_1}{y_1} + \frac{\lambda_2}{y_2} \le c\}$$

is closed, convex and the boundary $\partial\mathcal{M}$ equals the set of extreme points $\operatorname{ext}(\mathcal{M})$. The minimal distance r_γ from $x = (x_1, x_2) \in \mathcal{M}$ to the boundary $\partial\mathcal{M}$ can be described by

$$r_\gamma = \max\{r \,:\, B_\gamma(x, r) \subset \mathcal{M}\}.$$

Further, for every $\varepsilon > 0$ it holds that $B_\gamma(x, r_\gamma + \varepsilon) \not\subset \mathcal{M}$ and this implies that $\partial\mathcal{M} \cap \partial B_\gamma(x, r_\gamma) \neq \emptyset$. By Proposition 6.1.3 we have that

$$\partial\mathcal{M} \cap \operatorname{ext}\left(B_\gamma(x, r_\gamma)\right) \neq \emptyset.$$

Hence, δ can be chosen as $\delta = r_\gamma\, \gamma_k^{-1}\, e_k$ for some $k \in \{1, 2\}$.

In the inductive step we assume that the claim is valid for n and consider the case $n + 1$. So, consider the condition

$$\frac{\lambda_1}{x_1 - \delta_1} + \frac{\lambda_2}{x_2 - \delta_2} + \ldots + \frac{\lambda_n}{x_n - \delta_n} = c - \frac{\lambda_{n+1}}{x_{n+1} - \delta_{n+1}}.$$

By hypothesis it holds that $r_\gamma = \gamma_i\, \delta_i$ for some $i = 1, 2, ..., n$. Without loss of generality, let $\gamma_1\, \delta_1 = r_\gamma$. Hence,

$$\frac{\lambda_1}{x_1 - \delta_1} + \frac{\lambda_{n+1}}{x_{n+1} - \delta_{n+1}} = c - \frac{\lambda_2}{x_2} - \ldots - \frac{\lambda_n}{x_n}.$$

Thus, by induction hypothesis, it follows that $\gamma_1\, \delta_1 = r_\gamma$ or $\gamma_{n+1}\, \delta_{n+1} = r_\gamma$, which shows the assertion. ∎

Example 6.3.2 *We consider again the fluid network given in Example 6.1.5. For $\gamma = e$ a calculation based on* `fmincon` *in MATLAB yields the nominal stability radius $r^0(\alpha, \mu, P, C) = 0.13534$. In Figure 6.5 the fluid level processes of the fluid network $(\alpha, \mu - \delta, P, C)$ under HLPPS discipline is illustrated for the perturbation $\delta = (0\ 0\ 0.1325)^\mathsf{T}$. Furthermore, Figure 6.5 shows that the fluid level processes of classes 2 and 3 are increasing in the beginning. They start to decrease not before station 1 has reduced its fluid level to zero. The behavior looks like that of Example 6.2.2 except that the fluid levels empty earlier.*

Since $r^0(\alpha, \mu, P, C) = 0.13534$ is the nominal stability radius the question raises what measure of quality does it yields. The next example considers the global stability region for the Example 6.1.5.

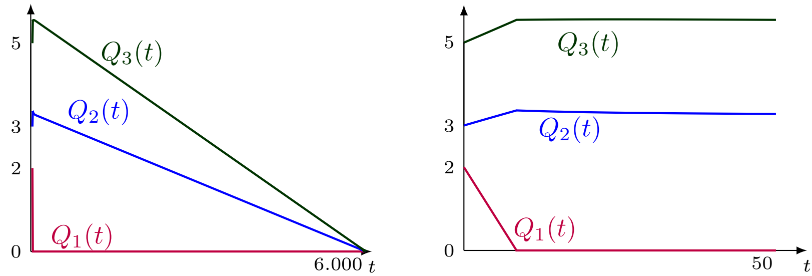

Figure 6.5: The figure on the left shows the fluid level processes to the fluid network $(\alpha, \mu-\delta, P, C)$ under HLPPS discipline where the parameters (α, μ, P, C) and the perturbation are chosen as in Example 6.1.5 and $\delta = (0\,0\,0.1325)^{\mathsf{T}}$, respectively. On the right the transient behavior of the fluid level processes is shown.

Example 6.3.3 *Consider the fluid network given in Example 6.1.5. For the perturbation* $\delta' = (\,0\ 0\ 0.1325\,)^{\mathsf{T}}$ *the matrix*

$$A = \begin{pmatrix} 10 & 0.3893 & 0.655 \\ 0.3893 & 0.599 & 1.01 \\ 0.655 & 1.01 & 1.703 \end{pmatrix}$$

satisfies the sufficient condition of Theorem 3.2.3. Consequently, the fluid network $(\alpha, \mu - \delta', P, C)$ *is globally stable.*

6.4 Perturbations of Arrival Rates and Service Capacities

In this section we will consider fluid networks that are subject to perturbations of the arrival rates as well as perturbations of the service capacities. That is, given a nominal fluid network (α, μ, P, C) we consider fluid networks $(\alpha + \Delta, \mu - \delta, P, C)$. The disturbances Δ and δ are of the form $\Delta \in \mathbb{R}_+^K$ and $0 \le \delta < \mu$, respectively. The nominal workload of the perturbed fluid network is given by

$$\rho(\Delta, \delta) = CM(\delta)^{-1}\,(I - P^{\mathsf{T}})^{-1}(\alpha + \Delta).$$

Following the approach of the previous sections we consider the nominal stability radius that corresponds to the nominal stability region, i.e.

$$r_\gamma^0(\alpha,\mu,P,C) = \min_{j=1,\ldots,J} \quad \min\{\|\Delta\|_\gamma + \|\delta\|_\gamma : \rho_j(\Delta,\delta)=1\}.$$

So, the nominal stability radius can be calculated by an optimization problem of the following from. For brevity, let $A := (I-P^\mathsf{T})^{-1}$ and consider

$$r_\gamma^0(\alpha,\mu,P,C) = \min_{j\in\{1,\ldots,J\}} r_j,$$

where r_j is a solution to

$$\begin{aligned} r_j = \min \quad & \sum_{k=1}^{K} \gamma_k(\delta_k+\Delta_k) \\ \text{subject to} \quad & \sum_{k\in C(j)} \frac{\lambda_k + \sum_{l=1}^{K} a_{kl}\,\Delta_l}{\mu_k-\delta_k} = 1, \\ & 0 \le \Delta, \\ & 0 \le \delta_k < \mu_k, \qquad k\in C(j). \end{aligned} \tag{6.14}$$

For each station $j\in\{1,...,J\}$ the general form of the latter setting is of the following form

$$\begin{aligned} r_j = \min \quad & \sum_{k=1}^{K} \gamma_k(\delta_k+\Delta_k) \\ \text{subject to} \quad & \sum_{k\in C(j)} \frac{\lambda_k + \sum_{l=1}^{K} a_{kl}\,\Delta_l}{\mu_k-\delta_k} = c, \\ & 0 \le \Delta, \\ & 0 \le \delta_k < \mu_k, \qquad k\in C(j). \end{aligned} \tag{6.15}$$

In the sequel, we will characterize the perturbations δ^*, Δ^* that provide solutions to (6.15). This is the content of the following theorem.

Theorem 6.4.1 *If $r_\gamma = \|\Delta_*\|_\gamma + \|\delta_*\|_\gamma > 0$ is a solution to* (6.15)*, then a destabilizing perturbation of minimal γ-norm can be chosen such that either for Δ_* or δ_* at most one component is strictly positive. That is, either $\Delta_* = r_\gamma\,\gamma_k^{-1}\,e_k$ and $\delta_* = 0$ or $\Delta_* = 0$ and $\delta_* = r_\gamma\,\gamma_k^{-1}\,e_k$ for some $k\in\{1,...,K\}$.*

Proof. Suppose that the perturbations (Δ_*, δ_*) are a solution to (6.15). Then, by Theorem 6.2.1 and 6.3.1, it follows that $\Delta_* = \Delta_m\, e_m$ for some $m \in \{1, ..., K\}$ and $\delta_* = \delta_n\, e_n$ for some $n \in \{1, ..., K\}$. If $n \neq m$, the optimization problem is given by

$$\begin{aligned}
\min \quad & \gamma_m\, \Delta_m + \gamma_n\, \delta_n \\
\text{subject to} \quad & \sum_{\substack{k\in C(j),\\ k\neq n}} \frac{\lambda_k}{\mu_k} + \frac{\lambda_n}{\mu_n - \delta_n} + \sum_{\substack{k\in C(j),\\ k\neq n}} \frac{a_{km}\, \Delta_m}{\mu_k} + \frac{a_{nm}\, \Delta_m}{\mu_n - \delta_n} = c, \\
& 0 \leq \delta_n < \mu_n, \\
& 0 \leq \Delta_m.
\end{aligned} \tag{6.16}$$

In the following, we consider the first constraint in (6.16), which can be written as

$$\frac{\lambda_n}{\mu_n - \delta_n} + \Delta_m \sum_{\substack{k\in C(j),\\ k\neq n}} \frac{a_{km}}{\mu_k} + \frac{a_{nm}\, \Delta_m}{\mu_n - \delta_n} = c - \sum_{\substack{k\in C(j),\\ k\neq n}} \frac{\lambda_k}{\mu_k}. \tag{6.17}$$

Since $\mu_n - \delta_n > 0$ and using for brevity $q := \sum_{k\in C(j),\, k\neq n} \frac{a_{km}}{\mu_k}$ and $\bar{c} := c - \sum_{k\in C(j),\, k\neq n} \frac{\lambda_k}{\mu_k}$, equation (6.17) reads as

$$\bar{c}\, \delta_n + (a_{nm} + q\mu_n)\, \Delta_m - q\, \Delta_m\, \delta_n = \bar{c}\mu_n - \lambda_n. \tag{6.18}$$

Moreover, using $\bar{d} := a_{nm} + q\mu_n$ and $p := \bar{c}\mu_n - \lambda_n$ condition (6.18) can be written in compact form as

$$f(\Delta_m, \delta_n) := \bar{c}\, \delta_n + \bar{d}\, \Delta_m - q\, \delta_n\, \Delta_m - p = 0. \tag{6.19}$$

By the definition of $\bar{d}, q$ it follows that $\frac{\bar{d}}{q} > \mu_n > \delta_n$. So, (6.19) can be written as

$$\Delta_m = h(\delta_n) := \frac{p - \bar{c}\, \delta_n}{\bar{d} - q\, \delta_n} = \frac{\bar{c}}{q} \cdot \frac{\frac{p}{\bar{c}} - \delta_n}{\frac{\bar{d}}{q} - \delta_n}, \qquad \delta_n \in [0, \mu_n).$$

The following instances may occur.

Suppose that $p\, q > \bar{c}\, \bar{d}$. Then, it holds that $\frac{\mathrm{d}h}{\mathrm{d}\delta}(\delta) = \frac{p\, q - \bar{c}\, \bar{d}}{(\bar{d} - q\delta)^2} > 0$. Rewritting the problem (6.16) as

$$\begin{aligned}
\min \quad & m(\delta_n) := \gamma_m\, h(\delta_n) + \gamma_n\, \delta_n \\
\text{subject to} \quad & 0 \leq \delta_n < \mu_n
\end{aligned} \tag{6.20}$$

and using that $\frac{\mathrm{d}m}{\mathrm{d}\delta}(\delta) = \gamma_m \frac{\mathrm{d}h}{\mathrm{d}\delta}(\delta) + \gamma_n > 0$, the minimum of m is attained for $\delta_n = 0$. Thus, the minimum of (6.16) is attained for $(\Delta_m, \delta_n) = (\frac{p}{\bar{d}}, 0)$.

If $\frac{\bar{c}}{q} = \frac{p}{\bar{d}}$, it holds that

$$\Delta_m = h(\delta_n) = \frac{\bar{c}}{q} \cdot \frac{\frac{p}{\bar{c}} - \delta_n}{\frac{\bar{d}}{q} - \delta_n} \equiv \frac{\bar{c}}{q} = \frac{p}{\bar{d}}, \qquad \delta_n \in [0, \mu_n).$$

So, the solution to

$$\begin{aligned} \min \qquad & m(\delta_n) := \gamma_m \tfrac{\bar{c}}{q} + \gamma_n \delta_n \\ \text{subject to} \quad & 0 \leq \delta_n < \mu_n \end{aligned} \tag{6.21}$$

is $(\Delta_m, \delta_n) = (\frac{p}{\bar{d}}, 0)$.

Finally, suppose that $pq < \bar{c}\,\bar{d}$ or equivalently $\frac{p}{\bar{c}} < \frac{\bar{d}}{q}$. Then, the constraint set is given by

$$\{(\Delta_m, \delta_n) \,:\, \Delta_m = h(\delta_m),\, \delta_m \in [0, \tfrac{p}{\bar{c}}]\,\}.$$

Further, $\gamma_m \Delta_m + \gamma_n \delta_n = r_\gamma$ defines a straight line, where (Δ_m, δ_n) also satisfy $\Delta_m = h(\delta_n)$. So, $\delta_n \in [0, \frac{p}{\bar{c}}]$ and $\Delta_m \in [0, \frac{p}{\bar{d}}]$. Hence, we have that either $\Delta_m = \frac{p}{\bar{d}}$ and $\delta_n = 0$ or $\Delta_m = 0$ and $\delta_n = \frac{p}{\bar{c}}$.

The case $n = m$ follows the same line of reasoning. This shows the assertion. ■

Remark 6.4.2 *The significance of the Theorems 6.2.1, 6.3.1 and 6.4.1 is that they allow for a calculation of the nominal stability radius as follows:*

1. *Consider each station $j \in \{1, ..., J\}$ separately.*
2. *Perturb either the arrival rate or service capacity of one fluid class and solve for each $k \in \{1, ..., K\}$ the optimization problem:*

$$\begin{aligned} r^{\chi}_{j,k} &= \min \gamma_k \chi_k \\ \textit{subject to} \quad & \rho_j(\chi_k) = 1. \end{aligned}$$

3. *Take the minimum of all obtained results, i.e.*

$$r^0_\gamma(\alpha, \mu, P, C) = \min \Big\{ r^{\chi}_{j,k} \,:\, j = 1, ..., J\,,\, k = 1, ..., K\,, \chi \in \{\Delta, \delta\} \Big\}.$$

6.5 Notes and References

In this chapter we derived bounds on possible shifts of the arrival rates and service capacities such that the property of stability is preserved. The terminology stability region of a discipline and global stability region has been introduced by Dai, cf. [29]. The monotonicity of the global stability region with respect to arrival rates follows from Theorem 3.6 in [19]. Later on Dai, Hasenbein and Vande Vate considered, for fixed arrival rates α the *monotone global stability region* $\mathcal{M}_\infty$ with respect to the service capacities, defined by the largest monotone subset of the stability region, see [32]. In this work it is also shown that, in general, $\mathcal{M}_\infty \neq \mathcal{D}_\infty$. For the counterexamples mentioned in Section 6.1 the interested reader is referred to [17, 32, 40] and the references therein. However, Dai and Vande Vate have shown that for two station networks the global stability region is monotone with respect to service rates, see [37]. This work is based on virtual stations and push starts and extends a linear programming approach to characterize the global stability region for two stations networks given in [9]. A comprehensive discussion of the global stability region is provided in Section 5.4 in [17].

The notion stability radius originates from the dynamical systems literature and was introduced by Hinrichsen and Pritchard in 1986, cf. [51, 52]. A comprehensive exposition to that can be found in [53]. The weighted stability radius for fluid networks as introduced in Definition 6.1.1 is an adaptation of the conceptual idea in [51], to measure the distance to instability, to the setting of fluid networks. The notations and results about convex analysis are taken from [69]. Therein, Theorem 6.1.2 is also stated, see Corollary 18.5.1, but it is not called Minkowski's Theorem.

In this chapter we concentrate on the nominal stability radius since for most disciplines, in general, analytic stability conditions are not available. For this reason, given a stable fluid network under a discipline π, defined by the parameters α, μ and P, and the constituency matrix C, the nominal stability radius provides an upper bound on the distance of (α, μ) to the boundary of the stability region $\mathcal{D}_\pi$. That is, we have that

$$B(\,(\alpha,\mu)\,,\, r_\gamma(\alpha,\mu,P,C,\pi)\,) \subset B(\,(\alpha,\mu)\,,\, r^0_\gamma(\alpha,\mu,P,C)\,).$$

Consequently, the insights of the results obtained in this chapter may be interpreted as follows. If a fluid network (α, μ, P, C, π) is subject to a disturbance Δ and it holds $\|\Delta\| < r^0(\alpha, \mu, P, C, \pi)$, the network might be stable and the stability analysis needs to be done in terms of the particular discipline. Parts of this Chapter have already been published in [72, 73, 74].

Finally, we note that the Lyapunov theory we developed in Chapter 4 and Chapter 5 is supposed to be an appropriate tool for future research, since the converse Lyapunov theorems if a Lyapunov function is available, provide necessary and sufficient stability conditions.

A Some Analysis

In this chapter we recall some definitions and results from analysis that are used throughout this thesis. The Banach space of continuous functions $x : [0,T] \to \mathbb{R}^n$ with the norm $\|x\|_C := \max_{t\in[0,T]} \|x(t)\|$is denoted by $C([0,T],\mathbb{R}^n)$. A set $X \subset C([0,T],\mathbb{R}^n)$ is called *equicontinuous* if for every $\varepsilon > 0$ there is a $\delta > 0$ such that for each $x \in X$ and $s,t \in [0,T]$ such that $|t-s| < \delta$ it holds that $\|x(t) - x(s)\| < \varepsilon$. The content of the Theorem of Arzelà-Ascoli is that the compact sets in $C([0,T],\mathbb{R}^n)$ are exactly those that are closed, bounded and equicontinuous.

Theorem A.1 (Arzelà-Ascoli) *A set $X \subset C([0,T],\mathbb{R}^n)$ is relatively compact if and only if it is bounded and equicontinuous.*

A proof can be found in [70] Section 9.8. In particular, this implies that every bounded and equicontinuous sequence $(x_n)_{n\in\mathbb{N}}$ in $C([0,T],\mathbb{R}^n)$ contains a convergent subsequence. Further, a function $x : \mathbb{R} \to \mathbb{R}^n$ is called *absolutely continuous* if for every finite interval $[a,b]$ and $\varepsilon > 0$ there is a $\delta > 0$ such that for any disjoint intervals $]a_i, b_i[\subset [a,b]$, $i = 1,2,...,n$ with $\sum_{i=1}^n (b_i - a_i) < \delta$ it holds that

$$\sum_{i=1}^n \|x(b_i) - x(a_i)\| < \varepsilon.$$

The following property of absolutely continuous functions is well known, a proof can be found in [70] Section 5.4.

Theorem A.2 *Suppose that x is absolutely continuous, then x is differential almost everywhere.*

A very useful property of absolutely continuous functions is that its derivative is Lebesgue integrable. That is, for all $s < t$,

$$x(t) - x(s) = \int_s^t \dot{x}(r)\,\mathrm{d}r.$$

Consequently, the space $AC([0,T],\mathbb{R}^n)$ of absolutely continuous functions $x : [0,T] \to \mathbb{R}^n$, equipped with the norm $\|\cdot\|_{AC}$ defined by

$$\|x\|_{AC} := \|x(0)\| + \int_0^T \|\dot{x}(t)\|\,\mathrm{d}t,$$

is a Banach space. A function $f : \mathbb{R}^m \to \mathbb{R}^n$ is called *locally Lipschitz* if for any bounded set $B \subset \mathbb{R}^m$ there is a $L_B > 0$ such that for all $x, y \in B$ we have that $\|f(x) - f(y)\| \leq L_B\|x - y\|$. Further, f is called Lipschitz if the last condition holds for $B = \mathbb{R}^m$.

Partition of Unity

Let $M \subset \mathbb{R}^n$, a family $\mathcal{U} = \{U_\alpha\}_{\alpha\in A}$ of open set is called an open cover of $M \subset \mathbb{R}^n$ if each $x \in M$ is in some $U_\alpha \in \mathcal{U}$. That is, $M \subset \bigcup_{\alpha\in A} U_\alpha$. An open cover is called *locally finite* if for every $\alpha \in A$ the set $\{\alpha' \in A : U_\alpha \cap U_{\alpha'} \neq \emptyset\}$ is finite.

Definition A.3 *A family $\Psi = \{\varphi_\alpha\}_{\alpha\in A}$ of C^∞ functions $\varphi_\alpha : M \to \mathbb{R}$ is called a* smooth partition of unity subordinate *to $\mathcal{U}$ if*

(1) $0 \leq \varphi_\alpha(x) \leq 1$ *for all $x \in M$ and for all $\alpha \in A$,*

(2) *supp $\varphi_\alpha \subset U_\alpha$,*

(3) *the set of supports $\{$supp $\varphi_\alpha\}_\alpha$ is locally finite,*

(4) *for all $x \in M$ it holds that* $\displaystyle\sum_{\alpha\in A} \varphi_\alpha(x) = 1$.

An important consequence of condition (3) in Definition A.3 is that the sum in condition (4) has actually only finitely many nonzero terms in a neighborhood of each point. So, there is no issue of convergence.

Theorem A.4 *Let $M \subset \mathbb{R}^n$ and $\mathcal{U}$ be an open cover of M. Then, there is a smooth partition of unity subordinate to $\mathcal{U}$.*

Proof. See [57] Theorem 2.25. ∎

B Differential Inclusions

Let X and Y denote metric spaces. A set-valued map $F : X \rightsquigarrow Y$ is a mapping that maps every $x \in X$ into a set $F(x)$ called the *value* of F at x. The domain of a set-valued map $F : X \rightsquigarrow Y$ is the subset of elements $x \in X$ such that the values $F(x)$ are non empty, i.e. $\mathrm{dom}(F) = \{x \in X \,:\, F(x) \neq \emptyset\}$. The *image* of F is the union of all values $F(x)$ for all $x \in X$. The *graph* of a set-valued map F is $\mathrm{graph}(F) := \{(x, y) \in X \times Y \,:\, y \in F(x)\}$. A set-valued map F is said to be *closed-valued* if the values of F are closed, i.e. for every $x \in X$ the set $F(x)$ is closed. Accordingly, F is said to be *convex* if the images are convex. Moreover, a set-valued map $F : X \rightsquigarrow Y$ is called *lower semicontinuous* at $x \in \mathrm{dom}(F)$ if for any $y \in F(x)$ and for any sequence of elements $(x_n)_{n\in\mathbb{N}} \in \mathrm{dom}(F)$ converging to x, there exists a sequence $(y_n)_{n\in\mathbb{N}}$ with $y_n \in F(x_n)$ converging to y. F is said to be lower semicontinuous if it is lower semicontinuous at every point $x \in \mathrm{dom}(F)$. In addition, F is called *upper semicontinuous* at $x \in \mathrm{dom}(F)$, if for any open neighborhood $U \supset F(x)$ there is an $\varepsilon > 0$ such that for all $x' \in B(x, \varepsilon) \cap \mathrm{dom}(F)$ it holds that $F(x') \subset U$. Again F is said to be upper semicontinuous if it is upper semicontinuous at every point $x \in \mathrm{dom}(F)$.

A useful criterion to conclude upper semicontinuity of a parameterized set-valued map is the following.

Proposition B.1 *Let X, Y and Z be metric spaces and $U : X \rightsquigarrow Z$ be a set-valued map. Assume that $f : \mathrm{graph}(U) \to Y$ is continuous. If U is upper semicontinuous with compact values, then $F : X \rightsquigarrow Y$ defined by*

$$F(x) := \{f(x, u) \,:\, u \in U(x)\}.$$

is upper semicontinuous.

Proof. See [5] Proposition 1.4.14. ■

Proposition B.2 *The graph of an upper semicontinuous set-valued map $F : X \rightsquigarrow Y$ with closed values is closed.*

Proof. See [4] Section 1.1 Proposition 2. ■

Let $K \subset \mathbb{R}^n$ and consider the differential inclusion

$$\dot{x}(t) \in F(x(t)), \quad x(0) = x_0. \tag{B.1}$$

A solution to (B.1) is an absolutely continuous function $x \in \mathrm{AC}(\mathbb{R}_+, \mathbb{R}^n)$ with $x(0) = x_0$ such that (B.1) is satisfied almost everywhere. The set of solutions to (B.1) starting at $x_0 \in K$ is denoted by $\mathcal{S}_F(x_0)$. The existence theorem is as follows.

Theorem B.3 *Let $K \subset \mathbb{R}^n$ be a closed set. Assume that the set-valued map $F : K \rightsquigarrow \mathbb{R}^n$ with closed convex values contained in a ball of radius $b > 0$ is upper semicontinuous. Then, the following conditions are equivalent.*

(1) For any $x_0 \in K$ there is a solution $x(\cdot) \in \mathcal{S}_F(x_0)$ satisfying $x(t) \in K$ for all $t \geq 0$.

(2) For any $x \in K$ it holds that $F(x) \cap \mathcal{T}_K(x) \neq \emptyset$.

Proof. See [80] Theorem 5.2. ■

Here, $\mathcal{T}_K(x)$ denotes the *contingent cone* to $K \subset \mathbb{R}^n$ at x, which is defined as the set of $v \in \mathbb{R}^n$ such that there is a sequence $(h_n)_{n\in\mathbb{N}} \subset \mathrm{int}(\mathbb{R}_+)$ converging to 0 and a sequence $(v_n)_{n\in\mathbb{N}} \subset \mathbb{R}^n$ converging to v such that for all $n \in \mathbb{N}$ it holds that $x + h_n\, v_n \in K$.

A differential inclusion (B.1) is called *strongly asymptotically stable* if no solution exhibits finite time blow-up, and

***Lyapunov Stability*:** For any $\varepsilon > 0$ there is a $\delta > 0$ such that any solution $x(\cdot)$ with $\|x_0\| < \delta$ satisfies $\|x(t)\| < \varepsilon$ for all $t \geq 0$.

***Attractiveness*:** For each individual solution $x(\cdot)$ one has $x(t)$ converges to 0 as $t \to \infty$.

The following statement is a smooth converse Lyapunov theorem for differential inclusions.

Theorem B.4 *Let $F : \mathbb{R}^n \rightsquigarrow \mathbb{R}^n$ be an upper semicontinuous set-valued map with nonempty, compact and convex values. Then, the differential inclusion* (B.1) *is strongly asymptotically stable if and only if there is a positive definite and proper function $V \in C^\infty(\mathbb{R}^n, \mathbb{R})$ and positive definite function $W \in C^\infty(\mathbb{R}^n \backslash \{0\}, \mathbb{R})$ such that for all $x \neq 0$ the following decrease condition is satisfied*

$$\max_{v \in F(x)} \langle \nabla V(x), v \rangle \leq -W(x). \tag{B.2}$$

Proof. See [26]. ∎

The content of the next statement is about the limit of a convergent of sequence of solutions.

Theorem B.5 *Assume that the set-valued map $F : \mathbb{R}^n \rightsquigarrow \mathbb{R}^n$ is upper semicontinuous with closed convex values contained in a ball of radius $b > 0$. If $(x_k(\cdot))_{k \in \mathbb{N}}$ is a sequence of solutions with $x_k(\cdot) \in \mathcal{S}_F$ that converges u.o.c. to a function $x(\cdot)$, then $x(\cdot) \in \mathcal{S}_F$.*

Proof. See [80] Theorem 4.6. ∎

A set-valued map $F : \mathbb{R}^n \rightsquigarrow \mathbb{R}^n$ is called *locally Lipschitz* on $\mathbb{R}^n \backslash \{0\}$ provided that for every compact set $U \in \mathbb{R}^n \backslash \{0\}$ there corresponds $L > 0$ such that

$$F(x_1) \subset F(x_2) + L \, \|x_1 - x_2\| \, B(0,1)$$

for all $x_1, x_2 \in U$, where $B(0,1) = \{x \in \mathbb{R}^n : \|x\| \leq 1\}$. The following statement from [26] will be very useful to prove a smooth converse Lyapunov theorem for differential inclusions.

Theorem B.6 *Let $F : \mathbb{R}^n \rightsquigarrow \mathbb{R}^n$ be an upper semicontinuous set-valued map with nonempty, compact and convex values. Suppose that F is strongly asymptotically stable. Then, there exists an upper semicontinuous set-valued map F_L with nonempty, compact and convex values which is locally Lipschitz on $\mathbb{R}^n \backslash \{0\}$, strongly asymptotically stable, and satisfies $F(x) \subset F_L(x)$.*

Proof. See [26]. ∎

Theorem B.7 *Assume that $F : X \rightsquigarrow X$ is Lipschitz with constant λ and has closed values on the interior of its domain. Let $y(\cdot) \in \mathcal{S}_F(y_0)$ be a given absolutely continuous function. Then, for all $t \geq 0$ it holds that*

$$d(y(t), \mathcal{S}_F(x_0)(t)) \leq ||x_0 - y_0||\, e^{\lambda t},$$

so that the solution map $\mathcal{S}_F$ is lower semicontinuous.

Proof. See [3] Section 5.3. ■

Bibliography

[1] D. Armbruster, P. Degond, and C. Ringhofer. A model for the dynamics of large queuing networks and supply chains. *SIAM J. Appl. Math.*, 66(3):896–920, 2006.

[2] D. Armbruster, K. Kaneko, and A. S. Mikhailov (Eds.). *Networks of interacting machines. Production organization in complex industrial systems and biological cells.* World Scientific Lecture Notes in Complex Systems 3. Hackensack, NJ: World Scientific, 2005.

[3] J.-P. Aubin. *Viability theory.* Systems and Control: Foundations and Applications. Birkhäuser, Boston, 1991.

[4] J.-P. Aubin and A. Cellina. *Differential inclusions. Set-valued maps and viability theory.* Grundlehren der Mathematischen Wissenschaften, 264. Springer, Berlin, 1984.

[5] J.-P. Aubin and H. Frankowska. *Set-valued analysis.* Systems and Control: Foundations and Applications, 2. Birkhäuser, Boston, 1990.

[6] J. Azéma, M. Kaplan-Duflo, and D. Revuz. Mesure invariante sur les classes recurrentes des processus de Markov. *Z. Wahrscheinlichkeitstheor. Verw. Geb.*, 8:157–181, 1967.

[7] A. Bacciotti and L. Rosier. *Liapunov functions and stability in control theory. 2nd ed.* Communications and Control Engineering. Springer, Berlin Heidelberg, 2005.

[8] A. Bernard and A. El Kharroubi. Régulations déterministes et stochastiques dans le premier "orthant" de $\mathbb{R}^n$. *Stochastics Stochastics Rep.*, 34:149–167, 1991.

[9] D. Bertsimas, D. Gamarnik, and J. N. Tsitsiklis. Stability conditions for multiclass fluid queueing networks. *IEEE Trans. Autom. Control*, 41(11):1618–1631, 1996.

[10] P. Billingsley. *Probability and measure. 3rd ed.* John Wiley & Sons, Inc., 1995.

[11] P. Billingsley. *Convergence of probability measures. 2nd ed.* John Wiley & Sons, Inc., 1999.

[12] J. Borwein and A. Lewis. *Convex analysis and nonlinear optimization: theory and examples.* Springer, 2006.

[13] M. Bramson. Instability of FIFO queueing networks. *Ann. Appl. Probab.*, 4(2):414–431, 1994.

[14] M. Bramson. Instability of FIFO queueing networks with quick service times. *Ann. Appl. Probab.*, 4(3):693–718, 1994.

[15] M. Bramson. Convergence to equilibria for fluid models of FIFO queueing networks. *Queueing Syst.*, 22(1-2):5–45, 1996.

[16] M. Bramson. Convergence to equilibria for fluid models of head-of-the-line proportional processor sharing queueing networks. *Queueing Syst.*, 23(1-4):1–26, 1996.

[17] M. Bramson. *Stability of queueing networks.* Lecture Notes in Mathematics 1950, Springer, Berlin Heidelberg, 2008.

[18] G. Caldarelli and A. Vespignani. *Large scale structure and dynamics of complex networks. From information technology to finance and natural science.* Complex Systems and Interdisciplinary Science 2, World Scientific, Hackensack, 2007.

[19] H. Chen. Fluid approximations and stability of multiclass queueing networks: Work-conserving disciplines. *Ann. Appl. Probab.*, 5(3):637–665, 1995.

[20] H. Chen. A sufficient condition for the positive recurrence of a semimartingale reflecting Brownian motion in an orthant. *Ann. Appl. Probab.*, 6(3):758–765, 1996.

[21] H. Chen and D. D. Yao. *Fundamentals of queueing networks. Performance, asymptotics, and optimization.* Applications of Mathematics. 46. Springer, New York, 2001.

[22] H. Chen and H. Ye. Piecewise linear Lyapunov function for the stability of multiclass priority queueing networks. *IEEE Trans. Autom. Control*, 47(4):564–575, 2002.

[23] H. Chen and H. Zhang. Stability of multiclass queueing networks under FIFO service discipline. *Math. Oper. Res.*, 22(3):691–725, 1997.

[24] H. Chen and H. Zhang. Stability of multiclass queueing networks under priority service disciplines. *Oper. Res.*, 48(1):26–37, 2000.

[25] K. Chung. *A course in probability theory.* Academic Press, San Diego, 2001.

[26] F. Clarke, Y. Ledyaev, and R. Stern. Asymptotic stability and smooth Lyapunov functions. *J. Differ. Equations*, 149(1):69–114, 1998.

[27] F. Clarke, R. Stern, and P. Wolenski. Subgradient criteria for monotonicity, the Lipschitz condition, and convexity. *Canad. J. Math*, 45:1167–1183, 1993.

[28] J. Dai. On positive Harris recurrence of multiclass queueing networks: A unified approach via fluid limit models. *Ann. Appl. Probab.*, 5(1):49–77, 1995.

[29] J. Dai. Stability of open multiclass queueing networks via fluid models. In *Kelly, F. and Williams, R. (Eds.), Stochastic Networks. Volume 71 of IMA Volumes in Mathematics and Its Applications*, pp. 71–90, Springer New York, 1995.

[30] J. Dai. A fluid limit model criterion for instability of multiclass queueing networks. *Ann. Appl. Probab.*, 6(3):751–757, 1996.

[31] J. Dai. *Stability of fluid and stochastic processing networks.* MaPhySto. Miscellanea. 9. Aarhus: Univ. of Aarhus, 1999.

[32] J. Dai, J. Hasenbein, and J. Vande Vate. Stability of a three-station fluid network. *Queueing Syst.*, 33(4):293–325, 1999.

[33] J. Dai, J. J. Hasenbein, and B. Kim. Stability of join-the-shortest-queue networks. *Queueing Syst.*, 57(4):129–145, 2007.

[34] J. Dai, J. J. Hasenbein, and J. H. Vande Vate. Stability and instability of a two-station queueing network. *Ann. Appl. Probab.*, 14(1):326–377, 2004.

[35] J. Dai and O. Jennings. Stabilizing queueing networks with setups. *Math. Oper. Res.*, 29(4):891–922, 2004.

[36] J. Dai and W. Lin. Maximum pressure policies in stochastic processing networks. *Oper. Res.*, 53(2):197–218, 2005.

[37] J. G. Dai and J. H. Vande Vate. The stability of two-station multitype fluid networks. *Oper. Res.*, 48(5):721–744, 2000.

[38] M. Davis. *Markov models and optimization.* Chapman & Hall, 1993.

[39] D. Down and S. Meyn. Piecewise linear test functions for stability and instability of queueing networks. *Queueing Syst.*, 27(3-4):205–226, 1997.

[40] V. Dumas. A multiclass network with non-linear, non-convex, non-monotonic stability conditions. *Queueing Syst.*, 25(1-4):1–43, 1997.

[41] P. Dupuis and R. J. Williams. Lyapunov functions for semimartingale reflecting Brownian motions. *Ann. Probab.*, 22(2):680–702, 1994.

[42] A. El Kharroubi, A. Ben Tahar, and A. Yaacoubi. On the stability of the linear Skorohod problem in an orthant. *Mathmatical Methods Operations Research.*, 56:243–258, 2002.

[43] S. Ethier and T. Kurtz. *Markov processes. Characterization and convergence.* John Wiley & Sons, Inc., New York, 1986.

[44] L. C. Evans. *Partial differential equations.* Graduate Studies in Mathematics 19. American Mathematical Society, Providence, Rhode Island, 1998.

[45] A. Filippov. On certain questions in the theory of optimal control. *J. Soc. Ind. Appl. Math., Ser. A, Control*, 1:76–84, 1962.

[46] S. Foss and T. Konstantopoulos. An overview of some stochastic stability methods. *J. Oper. Res. Soc. Japan*, 47(4):275–303, 2004.

[47] R. Getoor. Transience and recurrence of Markov processes. In Azéma, J. and Yor, M. (Eds.), Séminaire de Probabilités XIV, 1978/79, Lecture Notes in Mathematics 784, pp. 397–409, Springer-Verlag Berlin Heidelberg, 1980.

[48] J. K. Hale and S. M. Verduyn Lunel. *Introduction to functional differential equations.* Applied Mathematical Sciences. 99. Springer, New York, 1993.

[49] T. Harris. The existence of stationary measures for certain markov processes. In *Proc. 3rd Berkeley Sympos. Math. Statist. Probability 2*, pages 113–124, Univ. of Calif. Press, 1956.

[50] D. Helbing and S. Lämmer. Supply and production networks: from the bullwhip effect to business cycles. In *Armbruster, D., Kaneko, K. and Mikhailov, A. (Eds.), Networks of interacting machines: Production organization in complex industrial systems and biological cells.*, pages 33–66, World Scientific, Singapore, 2005.

[51] D. Hinrichsen and A. Pritchard. Stability radii of linear systems. *Syst. Control Lett.*, 7:1–10, 1986.

[52] D. Hinrichsen and A. Pritchard. Stability radius for structured perturbations and the algebraic Riccati equation. *Syst. Control Lett.*, 8:105–113, 1986.

[53] D. Hinrichsen and A. J. Pritchard. *Mathematical systems theory. I. Modelling, state space analysis, stability and robustness.* Texts in Applied Mathematics 48. Springer, Berlin Heidelberg, 2005.

[54] H. Kaspi and A. Mandelbaum. Regenerative closed queueing networks. *Stochastics Stochastics Rep.*, 39(4):239–258, 1992.

[55] C. M. Kellett and A. R. Teel. Smooth Lyapunov functions and robustness of stability for difference inclusions. *Syst. Control Lett.*, 52(5):395–405, 2004.

[56] P. Kumar and T. I. Seidman. Dynamic instabilities and stabilization methods in distributed real-time scheduling of manufacturing systems. *IEEE Trans. Autom. Control*, 35(3):289–298, 1990.

[57] J. M. Lee. *Introduction to smooth manifolds.* Graduate Texts in Mathematics 218. Springer, New York, 2003.

[58] S. Meyn. *Control techniques for complex networks.* Cambridge University Press, New York, 2008.

[59] S. Meyn and R. Tweedie. *Markov chains and stochastic stability.* Springer, London, 1993.

[60] S. P. Meyn. Transience of multiclass queueing networks via fluid limit models. *Ann. Appl. Probab.*, 5(4):946–957, 1995.

[61] S. P. Meyn and R. Tweedie. Stability of Markovian processes. I: Criteria for discrete-time chains. *Adv. Appl. Probab.*, 24(3):542–574, 1992.

[62] S. P. Meyn and R. Tweedie. Stability of Markovian processes. II: Continuous-time processes and sampled chains. *Adv. Appl. Probab.*, 25(3):487–517, 1993.

[63] S. P. Meyn and R. Tweedie. Stability of Markovian processes. III: Foster-Lyapunov criteria for continuous-time processes. *Adv. Appl. Probab.*, 25(3):518–548, 1993.

[64] Y. Ouyang and C. Daganzo. Characterization of the bullwhip effect in linear, time-invariant supply chains: Some formulae and tests. *Management Science*, 52(10):1544–1556, 2006.

[65] Y. Ouyang and C. Daganzo. Robust tests for the bullwhip effect in supply chains with stochastic dynamics. *Eur. J. Oper. Res.*, 185(1):340–353, 2008.

[66] J. Polderman and J. Willems. *Introduction to mathematical systems theory: a behavioral approach.* Springer, New York, 1998.

[67] A. Pukhalskij and A. Rybko. Nonergodicity of a queueing network under nonstability of its fluid model. *Probl. Inf. Transm.*, 36(1):23–41, 2000.

[68] P. Robert. *Stochastic networks and queues.* Applications of Mathematics 52. Springer, Berlin Heidelberg, 2003.

[69] R. Rockafellar. *Convex analysis.* Princeton University Press, 1970.

[70] H. Royden. *Real analysis. 2nd ed.* The Macmillan Company, London, 1968.

[71] A. Rybko and A. Stolyar. Ergodicity of stochastic processes describing the operation of open queueing networks. *Probl. Inf. Transm.*, 28(3):199–220, 1992.

[72] B. Scholz-Reiter, F. Wirth, T. Makuschewitz, and M. Schönlein. An approach to supply chain design based on robust capacity allocation of production locations. In *Proceedings of the POMS 21st Annual Conference, edited by Mark D. Hanna*, paper 015–0639, Vancouver, Cananda, May 7-10 2010.

[73] B. Scholz-Reiter, F. Wirth, T. Makuschewitz, and M. Schönlein. Robust capacity allocation in dynamic production networks. *CIRP Annals-Manufacturing Technology*, 60(1):445–448, 2011.

[74] M. Schönlein, T. Makuschewitz, F. Wirth, and B. Scholz-Reiter. Supply network engineering: An approach to robust capacity allocation for stochastic production processes. In *Preprints of the 18th IFAC World Congress 2011*, pp. 441–446, Milano, Italy, August 28 - September 2 2011.

[75] M. Schönlein and F. Wirth. Stability of fluid network models and Lyapunov functions. In *Proceedings of the 19th International Symposium on Mathematical Theory of Networks and Systems*, pp. 871–878, Budapest, Hungary, 5-9 July 2010.

[76] M. Schönlein and F. Wirth. On converse Lyapunov theorems for fluid network models. *Queueing systems*, 70:339–367, 2012.

[77] T. I. Seidman. ‘First come, first served’ can be unstable! *IEEE Trans. Autom. Control*, 39(10):2166–2171, 1994.

[78] M. Sharpe. *General theory of Markov processes.* Academic Press, Inc., San Diego, 1988.

[79] A. Siconolfi and G. Terrone. A metric approach to the converse Lyapunov theorem for continuous multivalued dynamics. *Nonlinearity*, 20(5):1077–1093, 2007.

[80] G. V. Smirnov. *Introduction to the theory of differential inclusions.* Graduate Studies in Mathematics. 41. American Mathematical Society, Providence, Rhode Island, 2002.

[81] A. Stolyar. On the stability of multiclass queueing networks: A relaxed sufficient condition via limiting fluid processes. *Markov Process. Relat. Fields*, 1(4):491–512, 1995.

[82] A. Teel and L. Praly. A smooth Lyapunov function from a class-$\mathcal{KL}$ estimate involving two positive semidefinite functions. *ESAIM: Control, Optimisation and Calculus of Variations*, 5:313–367, 2000.

[83] H. Wiendahl and S. Lutz. Production in networks. *CIRP Annals-Manufacturing Technology*, 51(2):573–586, 2002.

[84] J. C. Willems. Paradigms and puzzles in the theory of dynamical systems. *IEEE Trans. Automat. Control*, 36(3):259–294, 1991.

[85] J. C. Willems and K. Takaba. Dissipativity and stability of interconnections. *Int. J. Robust Nonlinear Control*, 17(5-6):563–586, 2007.

[86] H. Q. Ye and H. Chen. Lyapunov method for the stability of fluid networks. *Operations Research Letters*, 28(3):125–136, 2001.

List of Notations

(a, b)	Partition of $\{1, ..., K\}$	62
$\lvert x\rvert_{\text{vec}}$	Vector with componentwise absolute values	91
$\delta_s f(\cdot)$	Shift operator	47
$\mathbb{E}[\cdot]$	Expectation	12
$\mathcal{H}$	Set of hierarchical partitions of the set of classes	62
$\mathcal{K}$	Set of continuous strictly increasing functions mapping $\mathbb{R}_+$ to $\mathbb{R}_+$ and being zero at zero	71
$\lfloor t \rfloor$	The largest integer less than or equal to $t \in \mathbb{R}_+$	19
$\mathcal{FL}(\overline{q})$	The set of fluid limits with initial value $\overline{q}$.	114
$\mathcal{E}$	Set of customer classes with finite interarrival times	29
$\mathcal{FLM}$	Fluid limit model to a multiclass queueing network	43
$\mathcal{FL}(\omega)$	Set of fluid limits associated with the sample path ω	43
$\mathcal{FN}$	Fluid network associated with a multiclass queueing network	44
$\mathcal{K}_\infty$	Set of unbounded strictly increasing continuous functions mapping $\mathbb{R}_+$ to $\mathbb{R}_+$ and being zero at zero	18
$\mathcal{P}(M)$	Set of probability measures defined on M	13
$\mathcal{S}_F$	Solution map to the differential inclusion defined by F	150
$\mathcal{T}_K(x)$	Contingent cone of the set K at the point x	150
$\mathbb{1}_{\{A\}}$	Indicator function of a set A	13
$\mathrm{int}(A)$	Interior of the set A	132
Ω	Basic space of possible realizations	11
$\overline{\mathsf{X}}$	Fluid limit of a multiclass queueing network	43
∂A	Boundary of a set A	132
Φ	(closed) GFN model	77
Φ_x	Set of trajectories of the GFN model Φ starting at x	78
Π_k	Set of classes served at station $c(k)$ that have priority over class k	61

Index