Convergence of ODE approximations and bounds on performance models in the steady-state

Richard A. Hayden

Imperial College London
London, United Kingdom

rh@doc.ic.ac.uk

We present a limiting convergence result for differential equation approximations of continuous-time Markovian performance models in the stationary (steady-state) regime. This extends existing results for convergence up to some finite time. We show how, for a large class of performance models, this result can be inexpensively exploited to make strong statements about the stationary behaviour of massive continuous-time Markov chains. Furthermore, we present a new technique based on Lyapunov functions which has the potential to allow the efficient computation of tight guaranteed bounds on the stationary distribution.

1 Introduction

Fluid-analysis of performance models offers the exciting potential of analysing massive state spaces at small computational cost. In the case of stochastic process algebra models, fluid-analysis involves approximating the underlying discrete state space with continuous real-valued variables and describing the time-evolution of these variables with ordinary differential equations (ODEs). This approach was first applied to a subset of the stochastic process algebra PEPA [1] by Hillston [2], and has since been extended and developed in a number of different directions in the literature [3, 4, 5]. More recently, in terms of the Grouped PEPA (GPEPA) extension to PEPA, similar approximations have been developed for higher-order moments of performance models [6]. These are supported by the Grouped PEPA Analyser tool (GPA) [7, 8]. Furthermore, similar ideas have been applied in other stochastic process algebra [9, 10] and continuous stochastic Petri net [11] formalisms. Additionally, the *mean field analysis* as developed by Benaïm and Le Boudec [12] and Haverkort *et al.* [13, 14] for discrete-time performance models; and Gribaudo *et al.* [15] for continuous-time performance models is very closely related.

Despite the successful and widespread application of these techniques, many questions still exist regarding the relationship of the approximation to the original stochastic model, which, for the purposes of this paper, is its underlying continuous-time Markov chain (CTMC).

It can be shown that in the limit of large populations, the relative error between the fluid approximation and the underlying CTMC *up to some finite time* converges to zero almost surely [16, 17]. However, these results do not extend immediately to the stationary distribution of the CTMC. The first contribution of this paper is to prove a convergence result for the stationary measure of the CTMC to certain limiting quantities of the corresponding system of ODEs. The proof is based on techniques drawn from the field of stochastic approximation algorithms, in particular the work of Benaïm [18] and extends the mean-field stationary distribution convergence results of Benaïm and Le Boudec [12] to the continuous-time case. We will further show that for a subclass of performance models — those with piecewise linear approximating ODEs[1], we can often make strong statements about convergence in the stationary regime.

Consideration of limit results in the stationary regime requires naturally that one consider the stability properties of the phase space of the approximating system of ODEs. This is in direct contrast to the

[1]This includes all continuous stochastic Petri net models, a large subclass of GPEPA models and a large class of many-server queueing networks.

Submitted to:
PASTA 2010

transient, finite time convergence results already established by Kurtz *et al.* These tend to employ rather coarse devices such as Grönwall's inequality [19, pg. 498], which, in some sense, assume worst case stability properties for the system under consideration. This tends to result in bounds which grow worse exponentially with time and that are more often than not, of little direct use in practice. We will show in Section 4, how information about the stability profile of the ODEs in the form of a *Lyapunov function* allows us to obtain guaranteed and potentially tight bounds on the stationary measure of the CTMC.

2 Example performance model in grouped PEPA

We will proceed by presenting a straightforward, but non-trivial case study of a client–server interaction. This will be specified in the grouped PEPA [6] (GPEPA) stochastic process algebra, but similar models could alternatively be specified in other formalisms, for example, as a continuous stochastic Petri net.

In the GPEPA model:

$$\boldsymbol{CS}(N,M) \stackrel{def}{=} \mathbf{Clients}\{\boldsymbol{Client}[N]\} \underset{L}{\bowtie} \mathbf{Servers}\{\boldsymbol{Server}[M]\}$$

where $L = \{request, data\}$, we have a population of N clients and a population of M servers, whose local behaviour is given by:

$$\begin{aligned} \boldsymbol{Client} &\stackrel{def}{=} (request, r_r).\boldsymbol{Client_{waiting}} \\ \boldsymbol{Client_{waiting}} &\stackrel{def}{=} (data, r_d).\boldsymbol{Client_{think}} + (timeout, r_{tmt}).\boldsymbol{Client} \\ \boldsymbol{Client_{think}} &\stackrel{def}{=} (think, r_t).\boldsymbol{Client} \end{aligned}$$

$$\begin{aligned} \boldsymbol{Server} &\stackrel{def}{=} (request, r_r).\boldsymbol{Server_{get}} + (break, r_b).\boldsymbol{Server_{broken}} \\ \boldsymbol{Server_{get}} &\stackrel{def}{=} (data, r_d).\boldsymbol{Server} + (break, r_b).\boldsymbol{Server_{broken}} \\ \boldsymbol{Server_{broken}} &\stackrel{def}{=} (reset, r_{rst}).\boldsymbol{Server} \end{aligned}$$

The system uses a 2-stage fetch mechanism: a client requests data from the pool of servers; one of the servers receives the request, another server may then fetch the data for the client. At any stage, a server in the pool may fail. Clients may also timeout when waiting for data after their initial request.

Consider the three integer-valued stochastic processes which count the number of the N clients in each of the three possible derivative states of ***Client***. Let these be $N_{\boldsymbol{C}}(t)$, $N_{\boldsymbol{C_w}}(t)$ and $N_{\boldsymbol{C_t}}(t)$ respectively. Similarly, define for the servers, $N_{\boldsymbol{S}}(t)$, $N_{\boldsymbol{S_g}}(t)$ and $N_{\boldsymbol{S_b}}(t)$. Together, these stochastic processes form the *underlying aggregated CTMC*[2]

2.1 Fluid analysis

The idea of the fluid-analysis and related mean-field techniques is to define, by means of ODEs, deterministic, real-valued approximations, $x_\cdot(t)$, to the integer stochastic processes, $N_\cdot(t)$. In order to construct the ordinary differential equation which governs the evolution of each $x_\cdot(t)$, we consider the aggregate CTMC rate at which copies of the component are lost in the model and the rate at which they are gained, balancing the two quantities in terms of the fluid approximations $x_\cdot(t)$:

$$\begin{aligned} \dot{x}_{\boldsymbol{C}}(t) &= -\min(x_{\boldsymbol{C}}(t), x_{\boldsymbol{S}}(t))r_r + x_{\boldsymbol{C_w}}(t)r_{tmt} + x_{\boldsymbol{C_t}}(t)r_t \\ \dot{x}_{\boldsymbol{C_w}}(t) &= -\min(x_{\boldsymbol{C_w}}(t), x_{\boldsymbol{S_g}}(t))r_d - x_{\boldsymbol{C_w}}(t)r_{tmt} + \min(x_{\boldsymbol{C}}(t), x_{\boldsymbol{S}}(t))r_r \\ \dot{x}_{\boldsymbol{C_t}}(t) &= -x_{\boldsymbol{C_t}}(t)r_t + \min(x_{\boldsymbol{C_w}}(t), x_{\boldsymbol{S_g}}(t))r_d \end{aligned}$$

$$\begin{aligned} \dot{x}_{\boldsymbol{S}}(t) &= \min(x_{\boldsymbol{C_w}}(t), x_{\boldsymbol{S_g}}(t))r_d - \min(x_{\boldsymbol{C}}(t), x_{\boldsymbol{S}}(t))r_r - x_{\boldsymbol{S}}(t)r_b + x_{\boldsymbol{S_b}}(t)r_{rst} \\ \dot{x}_{\boldsymbol{S_g}}(t) &= -\min(x_{\boldsymbol{C_w}}(t), x_{\boldsymbol{S_g}}(t))r_d - x_{\boldsymbol{S_g}}(t)r_b + \min(x_{\boldsymbol{C}}(t), x_{\boldsymbol{S}}(t))r_r \\ \dot{x}_{\boldsymbol{S_b}}(t) &= -x_{\boldsymbol{S_b}}(t)r_{rst} + x_{\boldsymbol{S}}(t)r_b + x_{\boldsymbol{S_g}}(t)r_b \end{aligned}$$

[2] The aggregated state space of this model consists of potentially $\frac{1}{4}(2+N)(1+N)(2+M)(1+M)$ states. For $N = 100$ and $M = 50$, this is $6,830,226$ states. While an improvement on the unaggregated state space (around 3^{N+M} states), aggregation does not in general solve the state space explosion problem — the size of the aggregated space grows quickly as the number of possible derivative states increases.

These can then be inexpensively integrated to obtain the $x_\cdot(t)$ as deterministic, real-valued functions. To see in more detail how differential equations may be derived from an arbitrary grouped PEPA model, the reader is directed to [6].

3 Convergence results

In this section, we will review the known result for convergence of a sequence of appropriately-scaled CTMCs to their approximating system of ODEs *up to any finite point in time*. We call this *convergence in the transient regime*. We will then present the new result for the stationary regime. We first present the setup we will consider in the next two sections. For brevity, the conditions we impose are not the most general possible, but do include the class of performance models with which this paper is concerned.

Let $\mathbf{X}_n(t)$ be a sequence of continuous-time Markov chains each of whose state space, say $\mathscr{X}_n$, is a finite subset of $[0, 1]^D$ for some positive integer D. We assume that they are each irreducible and thus have a unique stationary measure, say μ_n on $(\mathscr{X}_n, \mathscr{P}(\mathscr{X}_n))$, which we will treat as a measure on $([0, 1]^D, \mathfrak{B}([0, 1]^D))^\dagger$. We make the following further assumptions:

1. The jump rate of $\mathbf{X}_n(t)$ is bounded above by $k_1 n$ for some $k_1 \in \mathbb{R}_+$
2. The second moment of the jump size of $\mathbf{X}_n(t)$ is bounded above by $k_2 n^{-2}$ for some $k_2 \in \mathbb{R}_+$
3. There exists some Lipschitz continuous $\mathbf{f}(\mathbf{x}) : [0, 1]^D \to \mathbb{R}^D$, such that for all n and all $\mathrm{x} \in \mathscr{X}_n$:

$$\mathbf{f}(\mathrm{x}) = \lim_{s \to 0} \frac{\mathbb{E}[\mathbf{X}_n(t+s) - \mathbf{X}_n(t) \,|\, \mathbf{X}_n(t) = \mathrm{x}]}{s}$$

With such a sequence of CTMCs, we may associate a single system of ODEs: $\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t))$. Write also $\mathbf{x}_\mathrm{x}(t)$ for the solution to this system when the initial condition is $\mathrm{x} \in [0, 1]^D$. We assume also that $\mathbf{x}_\mathrm{x}(t)$ can be shown to remain within $[0, 1]^D$ for all t if $\mathrm{x} \in [0, 1]^D$.

As an example, consider using the example GPEPA specification of Section 2 to create a sequence of models with an increasing total component population size, $N + M$. Furthermore, rescale each counting process by the total component population size. The resulting sequence of CTMCs then fits into the above framework with the system of ODEs given in Section 2.1.

3.1 Convergence in the transient regime

By applying the methods of Kurtz [16] or of Darling *et al.* [17], we may obtain the following result. For brevity, we do not give a proof here.

Theorem 3.1. *Let $T \geq 0$ and $\delta > 0$. Then for all $\varepsilon > 0$, there exists some N, such that for all $n \geq N$, we have:*

$$\mathbb{P}^n_\mathrm{x}\left\{\sup_{0 \leq t \leq T} \|\mathbf{X}_n(t) - \mathbf{x}_\mathrm{x}(t)\| \geq \delta\right\} \leq \varepsilon \tag{3.1}$$

for all $\mathrm{x} \in \mathscr{X}_n$. $\mathbb{P}^n_\mathrm{x}$ represents the transient probability measure of the CTMC, $\mathbf{X}_n(t)$, conditioned on $\mathbf{X}_n(0) = \mathrm{x}$.

†We work with $([0, 1]^D, \mathfrak{B}([0, 1]^D))$, the measure space formed by the Borel sets under the usual topology of $[0, 1]^D$.

3.2 Convergence in the stationary regime

In order to present this result, we first require a few dynamical systems definitions. The *omega limit set* of $\mathrm{x} \in [0,1]^D$, denoted $\omega(\mathrm{x})$, is the set of $\mathrm{y} \in [0,1]^D$, such that $\lim_{k\to\infty} \mathbf{x}_{\mathrm{x}}(t_k) = \mathrm{y}$ for some sequence $\{t_k\}_{k=0}^{\infty}$ with $\lim_{k\to\infty} t_k = \infty$. The *Birkhoff centre* of $\mathbf{x}(t)$, denoted $\mathscr{B}(\mathbf{x}(t))$, is the closure of the set of $\mathrm{x} \in [0,1]^D$ with $\mathrm{x} \in \omega(\mathrm{x})$. It is clear that this set contains any equilibrium points and periodic orbits. An *invariant measure* for $\mathbf{x}(t)$ is a probability measure, μ, on $([0,1]^D, \mathfrak{B}([0,1]^D))$ such that $\mu(\mathbf{x}_A^{-1}(t)) = \mu(A)$ for every $t \geq 0$ and $A \in \mathfrak{B}([0,1]^D)$, where $\mathbf{x}_A^{-1}(t) := \{\mathrm{x} \in [0,1]^D : \mathbf{x}_{\mathrm{x}}(t) \in A\}$. Then by the *Poincaré recurrence theorem*, $\mu(\mathscr{B}(\mathbf{x}(t))) = 1$. For more details, see [18].

The convergence result for the stationary regime now follows. The proof is based on results from the area of stochastic approximation algorithms, specifically, the work of Benaïm [18].

Theorem 3.2. *Let $K \subset [0,1]^D$ be closed and disjoint from $\mathscr{B}(\mathbf{x}(t))$, then $\lim_{n\to\infty} \mu_n(K) = 0$.*

Proof. See Appendix A. □

3.2.1 The case of a globally asymptotically stable fixed point

If the system $\mathbf{x}(t)$ has a globally asymptotically stable fixed point, that is, if all trajectories converge to this fixed point, then we can improve Theorem 3.2. In this case the Birkhoff centre consists of just one point, say, $\mathscr{B}(\mathbf{x}(t)) = \{\mathrm{x}^*\}$. For any $\varepsilon > 0$, let U_ε be the open ball around x^*. Then, clearly by Theorem 3.2, $\mu_n([0,1]^D \setminus U_\varepsilon) \to 0$ as $n \to \infty$. Thus the sequence of measures μ_n converges in probability and thus also weakly in distribution to the point mass at x^*.

We now apply this result to the example in Section 2. In order to proceed, we will perform a dimensionality reduction on the system of ODEs[3]. Note that for both the clients and servers, any one of the three counting processes is redundant and determined by the values of the other two and N and M. That is, for example, $x_{\boldsymbol{C_t}}(t) = N - x_{\boldsymbol{C}}(t) - x_{\boldsymbol{C_w}}(t)$ and $x_{\boldsymbol{S_b}}(t) = M - x_{\boldsymbol{S}}(t) - x_{\boldsymbol{S_g}}(t)$. We then need only differential equations for the other four components, but note that they will require the constant terms N and M to appear in their right-hand sides. This means that two rescaled instances of this model will now only have the same system of ODEs if the ratios $N/(N+M)$ and $M/(N+M)$ are the same in each case. Thus in order to fit into the framework of Section 3, the kinds of sequences of models we will be interested in are those in which the total population of clients and servers increase, but the ratio of clients to servers remains fixed. For example, the reduced system of ODEs associated to any grouped PEPA model of the form $\boldsymbol{CS}(2n, n)$, where the component counting processes are rescaled by the total population size, $3n$, is:

$$
\begin{aligned}
\dot{\bar{x}}_{\boldsymbol{C}}(t) &= -\min(\bar{x}_{\boldsymbol{C}}(t), \bar{x}_{\boldsymbol{S}}(t))r_r + \bar{x}_{\boldsymbol{C_w}}(t)r_{tmt} + (2/3 - \bar{x}_{\boldsymbol{C}}(t) - \bar{x}_{\boldsymbol{C_w}}(t))r_t \\
\dot{\bar{x}}_{\boldsymbol{C_w}}(t) &= -\min(\bar{x}_{\boldsymbol{C_w}}(t), \bar{x}_{\boldsymbol{S_g}}(t))r_d - \bar{x}_{\boldsymbol{C_w}}(t)r_{tmt} + \min(\bar{x}_{\boldsymbol{C}}(t), \bar{x}_{\boldsymbol{S}}(t))r_r \\
\dot{\bar{x}}_{\boldsymbol{S}}(t) &= -\min(\bar{x}_{\boldsymbol{C}}(t), \bar{x}_{\boldsymbol{S}}(t))r_r - \bar{x}_{\boldsymbol{S}}(t)r_b + \min(\bar{x}_{\boldsymbol{C_w}}(t), \bar{x}_{\boldsymbol{S_g}}(t))r_d + (1/3 - \bar{x}_{\boldsymbol{S}}(t) - \bar{x}_{\boldsymbol{S_g}}(t))r_{rst} \\
\dot{\bar{x}}_{\boldsymbol{S_g}}(t) &= -\min(\bar{x}_{\boldsymbol{C_w}}(t), \bar{x}_{\boldsymbol{S_g}}(t))r_d - \bar{x}_{\boldsymbol{S_g}}(t)r_b + \min(\bar{x}_{\boldsymbol{C}}(t), \bar{x}_{\boldsymbol{S}}(t))r_r
\end{aligned}
$$

where we can determine the other two rescaled component counts by $\bar{x}_{\boldsymbol{C_t}}(t) = 2/3 - \bar{x}_{\boldsymbol{C}}(t) - \bar{x}_{\boldsymbol{C_w}}(t)$ and $\bar{x}_{\boldsymbol{S_b}}(t) = 1/3 - \bar{x}_{\boldsymbol{S}}(t) - \bar{x}_{\boldsymbol{S_g}}(t)$. This is now a piecewise affine system and can be written in the form $\dot{\bar{\mathbf{x}}}(t) = A_{\sigma(\bar{\mathbf{x}}(t))}\bar{\mathbf{x}}(t) + (\frac{2}{3}r_t, 0, \frac{1}{3}r_{rst}, 0)^T$, where $\sigma(\mathrm{x}) : [0,1]^D \to \{1,2,3,4\}$ selects one of four 4×4 matrices, A_i, depending on the state of the two $\min(\cdot,\cdot)$ terms.

[3] This is necessary since the unreduced system of ODEs is the same for a whole family of grouped PEPA models (i.e. for different values of N and M), each with potentially a different stationary measure. They can thus not be expected to have only one fixed point.

In order to find the fixed point(s) of the reduced system, we solve, for each i, the linear equation $A_i\mathtt{x} = -(\frac{2}{3}r_t, 0, \frac{1}{3}r_{rst}, 0)^T$. For example, if $r_r = 2.0$, $r_{tmt} = 0.3$, $r_t = 0.2$, $r_b = 0.1$, $r_d = 1.0$ and $r_{rst} = 1.0$, each of these linear equations has exactly one solution. Only one of these four solutions, $\mathtt{x}^* = (0.1685, 0.0830, 0.0540, 0.2491)^T$, is actually a valid fixed point of the system. It now remains to show that $\mathtt{x}^*$ is actually globally asymptotically stable, that is, that all trajectories converge to it. The usual approach is to construct an appropriate Lyapunov function witnessing the global asymptotic stability of the fixed point, see e.g. [20, 21]. For general non-linear systems, this is a very difficult problem, but, for piecewise linear and affine systems, there are some inexpensive methods which allow the automatic computation of a such a function.

One approach is to construct a *common quadratic Lyapunov function* [22] by finding a solution to a set of *linear matrix inequalities* (LMIs). Fortunately, determining feasible solutions to a system of LMIs is a convex optimisation problem [23], and thus, can be done very efficiently. We do not have space to describe this technique in detail, but, for the above example, we can construct the quadratic form $V(\mathtt{x}) := (\mathtt{x}-\mathtt{x}^*)^T P(\mathtt{x}-\mathtt{x}^*)$, where P is a 4×4 symmetric, positive definite matrix solving[4] the system of LMIs, $PA_i + A_i^T P < 0$, for $i = 1, 2, 3, 4$. Feasible solutions can be found very quickly (in under a second) using, for example, the MATLAB® LMI toolbox [24]. Specifically, we used the `feasp` function to find the following solution[5]:

$$P = 10^3 \begin{pmatrix} 1.0374 & -0.1433 & -0.3085 & 0.5549 \\ -0.1433 & 3.2110 & 0.2893 & -2.8222 \\ -0.3085 & 0.2893 & 0.8283 & 0.1725 \\ 0.5549 & -2.8222 & 0.1725 & 3.7930 \end{pmatrix}$$

Then since P is positive definite, we have $V(\mathtt{x}^*) = 0$ and $V(\mathtt{x}) > 0$ for all $\mathtt{x} \neq \mathtt{x}^*$. Furthermore, the fact P solves the above system of LMIs can be shown [22] to ensure that $\frac{dV(\bar{x}(t))}{dt} < 0$ everywhere apart from $\mathtt{x}^*$, proving global asymptotic stability of $\mathtt{x}^*$, as required. So, having established this, Theorem 3.2 implies that the stationary measures of the sequence of GPEPA models, $\boldsymbol{CS}(2n, n)$, where component counts are rescaled by $3n$, converge in probability to the point mass at $\mathtt{x}^*$. So, for example, for large enough populations and, in the stationary regime, we can expect the broken servers to make up $(1/3 - 0.0540 - 0.2491)\times 3 \approx 9\%$ of the population of servers.

3.2.2 A more complicated example

In this section, we very briefly detail an example GPEPA model whose ODEs have a much more complicated Birkhoff centre. Specifically, we consider the GPEPA model:

$$(\mathbf{G}\{\boldsymbol{P}_1[N] \parallel \boldsymbol{P}_2[0]\} \underset{\{c\}}{\bowtie} \mathbf{H}\{\boldsymbol{P}_5[K]\}) \underset{\{a,b\}}{\bowtie} (\mathbf{J}\{\boldsymbol{P}_3[M] \parallel \boldsymbol{P}_4[0]\} \underset{\{d\}}{\bowtie} \mathbf{K}\{\boldsymbol{P}_6[J]\})$$

$$\boldsymbol{P}_1 \stackrel{def}{=} (a, r_r).\boldsymbol{P}_2 \qquad \boldsymbol{P}_3 \stackrel{def}{=} (a, r_r).\boldsymbol{P}_3 + (b, r_q).\boldsymbol{P}_4$$

$$\boldsymbol{P}_2 \stackrel{def}{=} (b, r_q).\boldsymbol{P}_2 + (c, r_s).\boldsymbol{P}_1 \qquad \boldsymbol{P}_4 \stackrel{def}{=} (d, r_z).\boldsymbol{P}_3$$

$$\boldsymbol{P}_5 \stackrel{def}{=} (c, r_s).\boldsymbol{P}_5 \qquad \boldsymbol{P}_6 \stackrel{def}{=} (d, r_z).\boldsymbol{P}_6$$

Since the population of $\boldsymbol{P}_5$ and $\boldsymbol{P}_6$ components are constant at K and J components, respectively; and, since the number of $\boldsymbol{P}_2$ and $\boldsymbol{P}_3$ components are determined by the number of $\boldsymbol{P}_1$ and $\boldsymbol{P}_4$ components, respectively, we may describe the fluid approximation to this model using only two coupled ODEs:

[4] Inequalities for matrices, e.g. $Q < 0$ or $Q > 0$ are interpreted as statements of negative or positive definiteness, respectively.

[5] Instead of giving the constraint that P be positive definite, we used the stronger condition, $P - I > 0$, which is more numerically stable.

$$\dot{x}_{P_1}(t) = -\min(x_{P_1}(t), M - x_{P_4}(t))r_r + \min(N - x_{P_1}(t), K)r_s$$
$$\dot{x}_{P_4}(t) = -\min(x_{P_4}(t), J)r_z + \min(M - x_{P_4}(t), N - x_{P_1}(t))r_q$$

Again, this is a piecewise affine system, described by 16 pairs of 2×2 matrices and 1×2 vectors. For the case, $N = M = 100$, $K = J = 15$, $r_r = r_q = 1.0$, $r_s = 3.0$ and $r_z = 2.0$, it can be shown that there is only one fixed point, at $\mathbf{x}^* = (70, 55)^T$. However, Fig. 1 shows that it is certainly not globally asymptotically stable — the Birkhoff centre consists also of periodic orbits. For any finite time, Theorem 3.1 guarantees that the population can be scaled high enough to ensure that, with high probability, a rescaled CTMC trace tracks the ODE closely up to that time. However, as can be verified with simulation, the stationary measure of the rescaled model *does not converge to a point mass as the populations are increased*, at $\mathbf{x}^*$, or anywhere else. In fact, the steady-state standard deviation of the component counts remain of the order of the population size. Theorem 3.2 would only allow us to say something formally about what happens in the stationary regime if we are first somehow able to characterise the Birkhoff centre formally. How this might be possible is an avenue for future work. Informally, observing many traces of the ODEs, started from different initial conditions, might convince oneself of the approximate Birkhoff centre. Then Theorem 3.2 can be used to argue that the stationary measure is concentrated there.

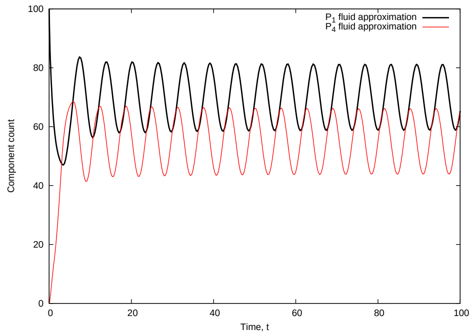

Figure 1: *An example of a GPEPA fluid approximation with a periodic orbit.*

4 Bounding the stationary distribution

In this section, we will show how the existence of a Lyapunov function witnessing global asymptotic stability of the system of ODEs can be used to get guaranteed bounds on the stationary distribution of the underlying CTMC. In some cases, these can be very tight. This can, in some sense, be viewed as a complementary approach to Theorem 3.2 in the case of a globally asymptotically stable fixed point.

Since we will be dealing with bounds for a specific model, we no longer need to explicitly consider a *sequence* of models. So, we will drop the n subscript from the quantities introduced in Section 3 — apart from this, we use the same notation. We begin by assuming the existence of the Lyapunov function and applying Dynkin's formula [25, pg. 382], where $\mathbf{X}(t)$ is the CTMC evolving in the stationary regime:

$$\mathbb{E}[V(\mathbf{X}(t))] - \mathbb{E}[V(\mathbf{X}(0))] = \int_0^t \mathbb{E}[\mathbf{Q}V(\mathbf{X}(u)]\,du \tag{4.1}$$

where $\mathbf{Q}$ is the infinitesimal generator of the CTMC[6]. Differentiating Eq. (4.1) and noting that $\frac{d}{dt}\mathbb{E}[V(\mathbf{X}(t))] = 0$ in the stationary regime, we obtain, $\mathbb{E}[\mathbf{Q}V(\mathbf{X}(t))] = 0$.

Now since $\mathbf{Q}V(\mathbf{X}(t))$ represents, in some sense, the stochastic analog of $\frac{dV(\bar{\mathbf{x}}(t))}{dt}$, we would expect it to be negative, at least on a large portion of the state space. Furthermore, it can be shown [22] that, for some $\alpha > 0$, $\frac{dV(\bar{\mathbf{x}}(t))}{dt} < -\alpha V(\bar{\mathbf{x}}(t))$, so we would actually expect $\mathbf{Q}V(\mathtt{x})$ to get more negative as we get

[6] In the discrete state space case, $\mathbf{Q}$ is the infinitesimal generator matrix and $\mathbf{Q}V(\mathtt{x})$ is interpreted as matrix multiplication if the function $V(\cdot)$ is represented as a vector, one entry for its value on each CTMC state, $\mathtt{x}$.

further away from x^*. To exploit this, we can compute the maximal value of $\mathbf{Q}V(\mathrm{x})$ on the entire state space of the CTMC. Let this be $V_+ := \max\{\mathbf{Q}V(\mathrm{x}) : \mathrm{x} \in \mathscr{X}\}$. Now assume that $\mathscr{X}' \subseteq \mathscr{X}$ is some subset of the state space of interest, and let $V_- := \max\{\mathbf{Q}V(\mathrm{x}) : \mathrm{x} \in \mathscr{X}'\}$ be the largest value taken on this subset. Then we have, $\mathbb{E}[\mathbf{Q}(V(\mathbf{X}(t)))] = 0 \leq (1-\mu(\mathscr{X}'))V_+ + \mu(\mathscr{X}')V_-$, and, if $V_- - V_+ < 0$, $\mu(\mathscr{X}') \leq -\frac{V_+}{V_- - V_+}$, a bound on the stationary measure, μ.

For our example of the last section, we have:

$$\begin{aligned}\mathbf{Q}V(\mathrm{x}) = & -(\min(\mathrm{x}_1,\mathrm{x}_3)r_r + \min(\mathrm{x}_2,\mathrm{x}_4)r_d + \mathrm{x}_2 r_{tmt} + (N-\mathrm{x}_1-\mathrm{x}_2)r_t + \mathrm{x}_3 r_b + \mathrm{x}_4 r_b + (M-\mathrm{x}_3-\mathrm{x}_4)r_{rst})V(\mathrm{x}) \\ & + \min(x_1,x_3)r_r V((\mathrm{x}_1-1,\mathrm{x}_2+1,\mathrm{x}_3-1,\mathrm{x}_4+1)^T) + \min(x_2,x_4)r_d V((\mathrm{x}_1,\mathrm{x}_2-1,\mathrm{x}_3+1,\mathrm{x}_4-1)^T) \\ & + \mathrm{x}_2 r_{tmt} V((\mathrm{x}_1+1,\mathrm{x}_2-1,\mathrm{x}_3,\mathrm{x}_4)^T) + (N/(N+M)-\mathrm{x}_1-\mathrm{x}_2)r_t V((\mathrm{x}_1+1,\mathrm{x}_2,\mathrm{x}_3,\mathrm{x}_4)^T) + \mathrm{x}_3 r_b V((\mathrm{x}_1,\mathrm{x}_2,\mathrm{x}_3-1,\mathrm{x}_4)^T) \\ & + \mathrm{x}_4 r_b V((\mathrm{x}_1,\mathrm{x}_2,\mathrm{x}_3,\mathrm{x}_4-1)^T) + (M/(N+M)-\mathrm{x}_3-\mathrm{x}_4)r_{rst} V((\mathrm{x}_1,\mathrm{x}_2,\mathrm{x}_3+1,\mathrm{x}_4)^T)\end{aligned}$$

Maximising this over the state space of the CTMC is a non-convex polynomial integer programming problem, which will be very expensive for large state spaces. So, instead, we relax the maximisations discussed above to find suprema over subsets of the real set $[0,1]^D$ which include the subset of $\mathscr{X}$ of interest — of course this still results in a valid bound because we have only potentially increased the values of V_+ and V_-. We have not yet investigated efficient methods for solving such real polynomial programming problems, but we have blindly applied the power of the `Maximize` function in `Mathematica`®, which, at least for the above expression, is able to return guaranteed global maxima in under a second (for the case $N = 100$, $M = 50$).

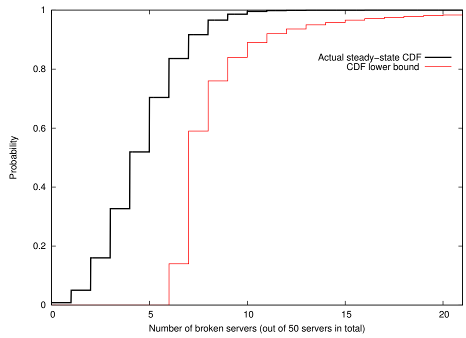

Figure 2: *Comparison of the actual steady-state CDF for the number of broken servers in the* $\mathbf{CS}(100,50)$ *model (computed by stochastic simulation) with a lower bound computed using the Lyapunov function technique. The rates used are those given earlier in Section 3.2.1.*

Fig. 2 gives an example CDF bound computed for the client/server model, specifically for the number of broken servers in the steady-state. In this case, the bound is quite tight and potentially very useful. However, we do note that it is not possible to get such impressive bounds for all of the component counts in that model using the same Lyapunov function. Further work is certainly needed to properly explore this very promising avenue. In particular, there is a whole family of different Lyapunov functions witnessing the global asymptotic stability of the fixed point in the earlier example — some may well perform better for the computation of certain bounds than others. It may also make a difference if we reduce the dimensionality of the ODEs by removing a different variable.

5 Conclusion

In this paper, we have extended some theoretical discrete-time results regarding steady-state convergence of Markov chains to their approximating system of differential equations to the continuous-time setting. We have also shown that for a large class of performance models (those with piecewise affine ODE systems), this result can be inexpensively applied in practice to validate the use of the ODE's fixed point as an approximation to the CTMC's steady-state measure. We have also given an example where things are not quite so straightforward. Furthermore, we have shown how a Lyapunov function can be used to provide useful guaranteed bounds on the stationary measure. Future work involves investigating this avenue more thoroughly. Also worthy of consideration is whether differential equation stability considerations can somehow lead to tighter bounds on the approximation error in the transient regime. Finally, while preparing this work, it has come to our attention that similar ideas regarding stationary bounds on Markov chains using Lyapunov functions have been considered very recently by Dayer *et al.* These authors have a preliminary technical report applying similar techniques to biochemical models [26].

References

[1] J. Hillston, *A Compositional Approach to Performance Modelling*. Cambridge University Press, 1996.

[2] J. Hillston, "Fluid flow approximation of PEPA models," in *Proceedings of the Second International Conference on the Quantitative Evaluation of Systems*, (Torino, Italy), pp. 33–43, IEEE Computer Society Press, Sept. 2005.

[3] J. T. Bradley, S. T. Gilmore, and J. Hillston, "Analysing distributed internet worm attacks using continuous state-space approximation of process algebra models," *Journal of Computer and System Sciences*, vol. 74, pp. 1013–1032, September 2008.

[4] N. Geisweiller, J. Hillston, and M. Stenico, "Relating continuous and discrete PEPA models of signalling pathways," *Theoretical Computer Science*, vol. 404, pp. 97–111, November 2008.

[5] R. Hayden and J. Bradley, "Evaluating fluid semantics for passive stochastic process algebra cooperation," *Performance Evaluation*, vol. 67, pp. 260–284, March 2009.

[6] R. Hayden and J. T. Bradley, "A fluid analysis framework for a Markovian process algebra," *Theoretical Computer Science*, vol. 411, pp. 2260–2297, May 2010. `doi://10.1016/j.tcs.2010.02.001`.

[7] A. Stefanek, R. Hayden, and J. T. Bradley, "A new tool for the performance analysis of massively parallel computer systems," in *Eighth Workshop on Quantitative Aspects of Programming Languages (QAPL 2010)*, Electronic Proceedings in Theoretical Computer Science, March 2010.

[8] A. Stefanek, R. Hayden, and J. T. Bradley, "GPA - Tool for rapid analysis of very large scale PEPA models," in *26th UK Performance Engineering Workshop (UKPEW 2010)*, July 2010.

[9] L. Bortolussi and A. Policriti, "Stochastic concurrent constraint programming and differential equations," in *QAPL'07, 5th Workshop on Quantitative Aspects of Programming Languages*, vol. 190 of *Electronic Notes in Theoretical Computer Science*, pp. 27–42, September 2007.

[10] L. Cardelli, "From processes to ODEs by Chemistry," in *TCS 2008, Fifth IFIP International Conference on Theoretical Computer Science*, (Milan), Springer, 2008.

[11] J. Júlvez, E. Jiménez, L. Recalde, and M. Silva, "On observability in timed continuous petri net systems," in *QEST'04, 1st International Conference on Quantitative Evaluation of Systems*, vol. 266, pp. 60–69, IEEE, September 2004.

[12] M. Benaïm and J.-Y. Le Boudec, "A class of mean field interaction models for computer and communication systems," *Performance Evaluation*, vol. 65, no. 11-12, pp. 823–838, 2008.

[13] R. Bakhshi, L. Cloth, W. Fokkink, and B. Haverkort, "Mean-field analysis for the evaluation of gossip protocols," in *Sixth International Conference on Quantitative Evaluation of Systems, QEST 2009*, pp. 247–256, September 2009.

[14] R. Bakhshi, J. Endrullis, S. Endrullis, W. Fokkink, and B. Haverkort, "Automating the mean-field method for large dynamic gossip networks," in *Seventh International Conference on Quantitative Evaluation of Systems, QEST 2010*, September 2010.

[15] A. Bobbio, M. Gribaudo, and M. Telek, "Analysis of large scale interacting systems by mean field method," in *Fifth International Conference on Quantitative Evaluation of Systems, QEST 2008*, pp. 215 – 224, September 2008.

[16] T. Kurtz, "Solutions of ordinary differential equations as limits of pure jump Markov processes," *Applied Probability*, vol. 7, pp. 49–58, April 1970.

[17] R. W. R. Darling and J. R. Norris, "Differential equation approximations for Markov chains," *Probability surveys*, vol. 5, pp. 37–79, 2008.

[18] M. Benaïm, "Recursive algorithms, urn processes and chaining number of chain recurrent sets," *Ergodic Theory and Dynamical Systems*, vol. 18, no. 1, pp. 53–87, 1998.

[19] T. Kurtz and S. Ethier, *Markov Processes Characterisation and Convergence*. Wiley, 1986.

[20] R. K. Miller and A. N. Michel, *Ordinary differential equations*. Academic Press, 1982.

[21] D. W. Jordan and P. Smith, *Nonlinear ordinary differential equations*. Oxford University Press, 1996.

[22] A. Pavlov, N. van de Wouw, and H. Nijmeijer, "Convergent piecewise affine systems: analysis and design, Part I: continuous case," in *44th IEEE Conference on Decision and Control, CDC-ECC '05*, pp. 5391 – 5396, December 2005.

[23] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear matrix inequalities in system and control theory*. Society for Industrial and Applied Mathematics, 1994.

[24] The MathWorks, "Introduction to linear matrix inequalities (Robust Control Toolbox™)." `http://www.mathworks.com/access/helpdesk/help/toolbox/robust/ug/f8-2138.html`.

[25] O. Kallenberg, *Foundations of Modern Probability*. Springer, 2002.

[26] T. Dayar, H. Hermanns, D. Spieler, and V. Wolf, "Bounding the equilibrium distribution of Markov population models," July 2010, arXiv:1007.3130.

[27] P. Billingsley, *Convergence of Probability Measures*. John Wiley & Sons, 1968.

A Proof of Theorem 3.2

Proof. Throughout this proof, we work with the compact metric space $[0,1]^D$ under its usual topology.

First we note that the sequence of measures μ_n is tight (see e.g. Billingsley [27, pg. 37] for a definition of tightness of measures) since the space $[0,1]^D$ is compact. Therefore the sequence μ_n is also relatively compact (e.g. [27, Theorem 6.1]), meaning that every subsequence of μ_n, say μ_{n_i} contains a further subsequence, say $\mu_{n_{i_k}}$ which converges weakly. These limit points are not necessarily equal, that is, the sequence μ_n does not necessarily converge weakly itself. However, we will show that any limit of a subsequence of μ_n is at least an invariant measure for $\mathbf{x}(t)$.

Let μ be some weak limit point, say $\mu_{n_i} \xrightarrow{w} \mu$. So we wish to show that if $A \in \mathfrak{B}([0,1]^D)$ and $t \geq 0$, we have $\mu(\mathbf{x}_A^{-1}(t)) = \mu(A)$. To see this, it is sufficient (e.g. [27, Theorem 1.3] and [25, Lemma 1.22]) to verify that for any bounded and continuous $g : [0,1]^D \to \mathbb{R}$:

$$\int_{[0,1]^D} g(\mathbf{x}_{\mathrm{x}}(t))\,\mu(d\mathrm{x}) = \int_{[0,1]^D} g(\mathrm{x})\,\mu(d\mathrm{x}) \tag{A.1}$$

Before we proceed, we observe that for any n:

$$\int_{[0,1]^D} g(\mathrm{x})\,\mu_n(d\mathrm{x}) = \int_{[0,1]^D} \mathbb{E}^n_{\mathrm{x}}[g(\mathbf{X}_n(t))]\,\mu_n(d\mathrm{x}) \tag{A.2}$$

where we write $\mathbb{E}^n_{\mathrm{x}}[g(\mathbf{X}_n(t))]$ for the expectation of $g(\mathbf{X}_n(t))$ conditioned on $\mathbf{X}_n(0) = \mathrm{x}$. This is true since the μ_n are the stationary measures of the $\mathbf{X}_n(t)$.

To show Eq. (A.1), fix $\delta > 0$, then combining Eq. (A.2) with the fact that $\mu_{n_i} \xrightarrow{w} \mu$, we may choose I sufficiently large such that for all $i \geq I$:

$$\left| \int_{[0,1]^D} g(\mathrm{x})\,\mu(d\mathrm{x}) - \int_{[0,1]^D} \mathbb{E}^{n_i}_{\mathrm{x}}[g(\mathbf{X}_{n_i}(t))]\,\mu_{n_i}(d\mathrm{x}) \right| < \delta/4 \tag{A.3}$$

We now proceed to bound the term $|\mathbb{E}^n_{\mathrm{x}}[g(\mathbf{X}_n(t))] - g(\mathbf{x}_{\mathrm{x}}(t))|$ for large enough n, uniformly for any $\mathrm{x} \in \mathscr{X}_n$. We note that g is uniformly continuous since $[0,1]^D$ is compact. Therefore we can find $\alpha > 0$ independent of x such that $\|\mathrm{y} - \mathbf{x}_{\mathrm{x}}(t)\| < \alpha \Rightarrow |g(\mathrm{y}) - g(\mathbf{x}_{\mathrm{x}}(t))| < \delta/4$. This allows us to employ the following straightforward upper bound, where $\|g\|$ is an upper bound on the magnitude of g:

$$|\mathbb{E}^n_{\mathrm{x}}[g(\mathbf{X}_n(t))] - g(\mathbf{x}_{\mathrm{x}}(t))| \leq \delta/4 + 2\|g\|\mathbb{P}^n_{\mathrm{x}}\left\{ \sup_{0\leq s\leq t} \|\mathbf{X}_n(s) - \mathbf{x}_{\mathrm{x}}(s)\| \geq \alpha \right\}$$

Then applying Theorem 3.1 with $\varepsilon = \delta/(8\|g\|)$, we may choose M sufficiently large such that for all $m \geq M$, $|\mathbb{E}^m_{\mathrm{x}}[g(\mathbf{X}_m(t))] - g(\mathbf{x}_{\mathrm{x}}(t))| \leq \delta/2$ uniformly for any $\mathrm{x} \in \mathscr{X}_m$. Choose J sufficiently large such that $J \geq I$ and $n_J \geq M$, then, using this bound and that of Eq. (A.3), we obtain for all $j \geq J$:

$$\begin{aligned}\left| \int_{[0,1]^D} g(\mathrm{x})\,\mu(d\mathrm{x}) - \int_{[0,1]^D} g(\mathbf{x}_{\mathrm{x}}(t))\,\mu_{n_j}(d\mathrm{x}) \right| &\leq \left| \int_{[0,1]^D} g(\mathrm{x})\,\mu(d\mathrm{x}) - \int_{[0,1]^D} \mathbb{E}^{n_j}_{\mathrm{x}}[g(\mathbf{X}_{n_j}(t))]\,\mu_{n_j}(d\mathrm{x}) \right| \\ &\quad + \int_{[0,1]^D} |\mathbb{E}^{n_j}_{\mathrm{x}}[g(\mathbf{X}_{n_j}(t))] - g(\mathbf{x}_{\mathrm{x}}(t))|\,\mu_{n_j}(d\mathrm{x}) \\ &\leq 3\delta/4 \end{aligned} \tag{A.4}$$

Also since $g(\mathbf{x}_{\mathrm{x}}(t))$ is bounded and continuous as a function of x (see e.g. [20] for the relevant arguments regarding continuous dependence on initial conditions), we can use the fact that $\mu_{n_i} \xrightarrow{w} \mu$ to find K sufficiently large such that for all $k \geq K$:

$$\left| \int_{[0,1]^D} g(\mathbf{x}_{\mathrm{x}}(t))\, \mu(d\mathrm{x}) - \int_{[0,1]^D} g(\mathbf{x}_{\mathrm{x}}(t))\, \mu_{n_k}(d\mathrm{x}) \right| < \delta/4$$

Then combining this with Eq. (A.4), we can obtain:

$$\left| \int_{[0,1]^D} g(\mathrm{x})\, \mu(d\mathrm{x}) - \int_{[0,1]^D} g(\mathbf{x}_{\mathrm{x}}(t))\, \mu(d\mathrm{x}) \right| < \delta$$

Since δ was arbitrary, this verifies Eq. (A.1) and thus shows that μ is an invariant measure for $\mathbf{x}(t)$.

Now let $K \subset [0,1]^D$ be closed and disjoint from $\mathscr{B}(\mathbf{x}(t))$. Then by the *Poincaré recurrence theorem*, $\mu(K) = 0$. This holds for all limit points, μ, of μ_n.

Finally, we assume for a contradiction that $\lim_{n\to\infty} \mu_n(K) \neq 0$. Then there is some $\varepsilon > 0$ such that $\mu_{n_l}(K) > \varepsilon$ for the subsequence μ_{n_l}. By relative compactness, we can find a further subsequence, say $\mu_{n_{l_q}}$, such that $\mu_{n_{l_q}} \xrightarrow{w} \mu$ for some weak limit point μ. But since K is closed, we have $\limsup \mu_{n_{l_q}}(K) \leq \mu(K) = 0$, a contradiction, as required. □