HAL
open science

Construction of Lyapunov functions via relative entropy with application to caching

Nicolas Gast

To cite this version:

Nicolas Gast. Construction of Lyapunov functions via relative entropy with application to caching. The 18th Workshop on MAthematical performance Modeling and Analysis, Jun 2016, Nice, France. hal-01321017v1

HAL Id: hal-01321017

https://inria.hal.science/hal-01321017v1

Submitted on 24 May 2016 (v1), last revised 30 Jun 2016 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Construction of Lyapunov functions via relative entropy with application to caching*

Nicolas Gast
Inria
Univ. Grenoble Alpes
CNRS, LIG, F-38000 Grenoble, France
nicolas.gast@inria.fr

ABSTRACT

We consider a system of interacting objects that is a generalization of the model of the cache-replacement policy RAND($\mathbf{m}$) policy introduced in [6]. We provide a mean-field approximation of this system. We show how to use relative entropy to construct a Lyapunov function for this model. This guarantees that the mean-field model converges to its unique fixed point.

1. INTRODUCTION

Mean-field approximation is a powerful tools for studying systems of interacting objects. Many papers have established that, under mild condition, a system of interacting objects converges to a deterministic dynamical system $\dot{\mathbf{x}} = f(\mathbf{x})$ over any finite time interval. However, characterizing the asymptotic behavior of a differential equation is often a difficult task: in particular, showing that the differential equation has a unique fixed point does not guarantee the convergence of the stochastic system to this fixed point [1]. A sufficient condition is that all trajectories of this dynamical system converge to a fixed point of the equation, which is often difficult to show. Recently, it has been shown in [3] that relative entropy can be used to construct a Lyapunov function that guarantees the convergence of all trajectories to a fixed point. In this paper, we use a similar idea to construct a Lyapunov function for a generalization of the cache-replacement policies RAND($\mathbf{m}$) introduced in [6].

This paper makes two contributions. First, we finish the analysis of the mean-field model of [6] by providing a Lyapunov function for the mean-field model. Second, we illustrate a method for constructing explicitly a Lyapunov function by considering a relative entropy plus a corrective term. We believe that this method is generic and can be applied to other mean-field dynamics.

2. MODEL AND APPROXIMATION

We consider a system of n objects that move between boxes. There are h boxes and the ith box contains m_i objects. If the object k is in box i, then at rate $\lambda_{k,i}$ it is exchanged with an object (taken at random) from box $i+1$. A configuration of the system is a vector $\mathbf{c} = (c_1 \dots c_n)$ where c_k is the box in which object k is. The set of possible configurations $\mathcal{C}$ is the set of vectors $\mathbf{c} = (c_1 \dots c_n)$ such that there are exactly m_i objects in box i: $|\{k : c_k = i\}| = m_i$.

This model is a generalization of the model of the cache-replacement policy RAND($\mathbf{m}$) introduced in [6], where the authors study a cache that is decomposed into boxes. At each request, an item that is in box i is exchanged with another item from box $i+1$. In our setting, this correspond to $\lambda_{k,i}$ independent of i and equal to the request-rate of object k.

It can be verified that this system is reversible and that its stationary measure satisfies, for each $c \in \mathcal{C}$:

$$\pi(\mathbf{c}) = \kappa \prod_{k=1}^{n} \prod_{i=1}^{c_k - 1} \lambda_{k,i},$$

where $\kappa = (\sum_{\mathbf{c}\in\mathcal{C}} \prod_{k=1}^{n} \prod_{i=1}^{c_k-1} \lambda_{k,i})^{-1}$ is the normalizing constant.

The constant κ can be computed by a dynamic programming algorithm in a time $O(n \prod_{i=1}^{h}(m_i + 1))$ by using a similar algorithm as the one of [6]. While this complexity is polynomial in n and m, it is exponential in the number of boxes. This calls for an approximation.

2.1 The mean-field approximation

We consider the following set of ODEs (for $k \in \{1 \dots n\}, i \in \{1 \dots h\}$):

$$\dot{x}_{k,i}(t) = \mathbf{1}_{\{i>1\}} \left(\lambda_{k,i} x_{k,i-1}(t) - \frac{\sum_\ell \lambda_{\ell,i-1} x_{\ell,i-1}(t)}{m_i} x_{k,i}(t) \right) + \mathbf{1}_{\{i<h\}} \left(\frac{\sum_\ell \lambda_{\ell,i} x_{\ell,i}(t)}{m_{i+1}} x_{k,i+1}(t) - \lambda_{k,i} x_{k,i}(t) \right) \quad (1)$$

with initial conditions $x_{k,i}(0) = \mathbf{1}_{\{c_k(0)=i\}}$, where $\mathbf{1}_{\{A\}} = 1$ if A is true and 0 otherwise.

The quantity $x_{k,i}$ is meant to be an approximation of the probability for an object k to be in box i. For object k, this ODE corresponds to the probability measure of a birth-death process of birth rate $\lambda_{k,i}$ and of death rate $\mu_i(\mathbf{x}) = \sum_{\ell=1}^{n} \lambda_{\ell,i-1} x_{\ell,i-1}/m_i$. This birth-death process is represented in Figure 1. The intuition behind this ODE is the following. If object k is in box i, it jumps to box $i+1$ at rate $\lambda_{k,i}$. It jumps to box $i-1$ when an object from box $i-1$ jumps to box i and is exchanged with this object. This occurs at rate $\sum_{\ell=1}^{n} \mathbf{1}_{\{c_\ell = i-1\}} \lambda_{\ell,i-1}$.

The following result guarantees that $x_{k,i}$ is indeed a good approximation of the stochastic process. It can be proven by adapting the proof of [6, Theorem 6], that is based on stochastic approximation arguments.

*This work is partially supported by the EU project QUANTICOL, 600708.

Copyright is held by author/owner(s).

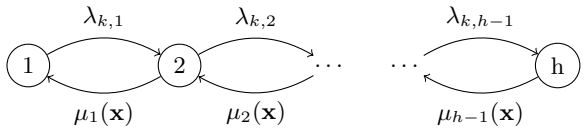

Figure 1: The birth-death process approximation of the box in which object k is, where $\mu_i(\mathbf{x}) = \sum_{\ell=1}^n \lambda_{\ell,i-1} x_{\ell,i-1}/m_i$.

THEOREM 1. *For any T, there exists a constant C independent of the parameters $\lambda_{i,k}$ and $m_{i,k}$ such that*

$$\mathbf{E}\left[\sup_{t\le T/\varepsilon, i, \alpha\ge 0} \left\| \sum_k (\lambda_{k,i})^\alpha (x_{k,i}(t) - \mathbf{1}_{\{c_k(t)=i\}}) \right\|_\infty \right] \le C\varepsilon,$$

where $\varepsilon = \max_{i,k} \lambda_{i,k} + \max_i (m_i)^{-1}$.

This theorem guarantees that the content of the boxes of the stochastic process converges to the one of the ODE approximation as the values $(\lambda_{k,i})$s go to 0 and the box sizes (m_i)s go to infinity. Note that the values of $(\lambda_{i,k})$s can be made arbitrarily small by multiplying all of them by a constant $\tau < 1$. Hence, the main assumption for the convergence to hold is that the box sizes (m_i)s go to infinity.

2.2 Fixed point analysis

The differential equation (1) has the following form:

$$\frac{d}{dt} x_{k,i}(t) = \sum_{j=1}^h x_{k,j}(t) \Gamma_{j,i}^{(k)}(\mathbf{x}(t)),$$

where $\Gamma^{(k)}(\mathbf{x})$ is the transition kernel of a birth-death process with a birth rate $\lambda_{k,i}$ and a death rate that depends on $\mathbf{x}$: $\mu_i(\mathbf{x}) = \sum_{\ell=1}^n \lambda_{\ell,i-1} x_{\ell,i-1}/m_i$.

For a fixed $\mathbf{x}$, the stationary measure of the birth-death process of kernel $\Gamma^{(k)}(\mathbf{x})$ is $\pi_{k,\cdot}(\mathbf{x})$, where for $i \in \{1 \dots h\}$:

$$\pi_{k,i}(\mathbf{x}) = \frac{\prod_{j=1}^{i-1} \lambda_{k,j}/\mu_j(\mathbf{x})}{\sum_{j'=1}^h \prod_{j=1}^{j'-1} \lambda_{k,j}/\mu_j(\mathbf{x})} \tag{2}$$

A fixed point of the ODE (1) satisfies $\mathbf{x} = \pi(\mathbf{x})$. By monotonicity arguments similar to the proof of [6, Theorem 7], it can be shown that this equation has a unique fixed point, that we denote $\mathbf{x}^*$. This fixed point can be efficiently computed by an iterative procedure (see [6, Section 5.3]).

3. LYAPUNOV FUNCTION

In this section, we show that, any solution of (1) converges to the unique fixed point $\mathbf{x}^*$. We construct explicitly a Lyapunov function, by using relative entropy.

3.1 The reason to use relative entropy

As we recall below, relative entropy between a Markov chain and its stationary measure decreases with time. We will see that this implies that it is natural to consider the relative entropy between $\mathbf{x}$ and $\pi(x)$ plus a corrective term.

Recall that for two discrete distributions p and q over a finite set $\{1 \dots h\}$, the relative entropy of p with respect to q (also called Kullback-Leibler divergence) is

$$R(p\|q) := \sum_{i=1}^h p_i \log \frac{p_i}{q_i}$$

If $\mathbf{x}(t)$ is the probability distribution at time t of a continuous time Markov process that has kernel Γ and if π is its steady-state distribution, then the relative entropy of $\mathbf{x}(t)$ with respect to π decreases with time:

$$\frac{d}{dt} R(\mathbf{x}(t)\|\pi) = \sum_{i=1}^h \left(\frac{d}{dt} x_i(t)\right)\left(1 + \log \frac{x_i(t)}{\pi_i}\right) \le 0, \tag{3}$$

where $\frac{d}{dt} x_i(t) = \sum_j x_{k,j}(t) \Gamma_{j,i}$. The equality holds only when $\mathbf{x}(t) = \pi$. A simple proof of this result is in [3, Lemma 3.1].

For a given object k, the process $x_{k,\cdot}$ is a Markovian model with kernel $\Gamma^{(k)}(\mathbf{x})$ that depends on $\mathbf{x}$. The stationary measure of this kernel is $\pi_{k,\cdot}(\mathbf{x})$, defined in (2).

The derivative with respect to time of $R(\mathbf{x}_{k,\cdot}, \pi_{k,\cdot}(\mathbf{x}))$ is

$$\begin{aligned}\frac{d}{dt} R(\mathbf{x}_{k,\cdot}(t)\|\pi_{k,\cdot}(\mathbf{x}(t))) &= \sum_{i=1}^h \left(\frac{d}{dt} x_{k,i}(t)\right)\left(1 + \log \frac{x_{k,i}(t)}{\pi_{k,i}(\mathbf{x})}\right) \\ &\quad - \sum_{i=1}^h x_{k,i}(t) \frac{d}{dt} \log \pi_{k,i}(\mathbf{x}(t)) \\ &\le - \sum_{i=1}^h x_{k,i}(t) \frac{d}{dt} \log \pi_{k,i}(\mathbf{x}(t)) \end{aligned} \tag{4}$$

This inequality holds because of Equation (3).

If $h(\mathbf{x})$ be a primitive integral of (4), then the function $t \mapsto R(\mathbf{x}(t)\|\pi(\mathbf{x}(t))) - h(\mathbf{x}(t))$ decreases with time. When $h(\mathbf{x})$ is upper bounded, this function decreases to some value, which by Equation (3), implies that $\mathbf{x}(t)$ converges to a fixed point. Hence, exhibiting a bounded primitive integral of (4) suffices to guarantee the convergence of the ODE to a fixed point.

3.2 Main result and proofs

We say that a vector $\mathbf{x}$ is admissible if $x_{k,i} \ge 0$ and if $\sum_{k=1}^n x_{k,i} = m_i$ and $\sum_{i=1}^h x_{k,i} = 1$. The first condition corresponds to having m_i objects per box and the second that the object k is in one of the h boxes.

THEOREM 2. *For any admissible initial condition $\mathbf{x}(0)$, the ODE (1) has a unique solution $\mathbf{x}$. Moreover, for all t, $\mathbf{x}(t)$ is admissible and $\lim_{t\to\infty} \mathbf{x}(t) = \mathbf{x}^*$.*

PROOF. The existence and uniqueness of the solution is due to the Lipchitz-continuity of the right-hand side of Equation (1). The fact that $\mathbf{x}(t)$ remains admissible when $\mathbf{x}(0)$ is admissible can be verified by showing that $\sum_{i=1}^h \dot{x}_{k,i} = \sum_{k=1}^n \dot{x}_{k,i} = 0$ when $\mathbf{x}$ is admissible.

We prove in Lemma 1 and 2 that the function (5) is a Lyapunov function: it is non increasing and lower bounded. Hence, $\lim_{t\to\infty} \frac{d}{dt} L(\mathbf{x}(t)) = 0$. By Lemma 1 and the continuity of $\pi(\mathbf{x})$, this shows that $\lim_{t\to\infty} \mathbf{x}(t) = \mathbf{x}^*$. □

We define the function $L(\mathbf{x})$ be by:

$$\begin{aligned} L(\mathbf{x}) = \sum_{k=1}^n R(x_k\|\pi_k(\mathbf{x})) - \sum_{i=1}^h m_i \log\left(\prod_{j=1}^{i-1} \mu_j(\mathbf{x})\right) \\ - \sum_{k=1}^n \log\left(\sum_{i=1}^h \prod_{j=1}^{i-1} \lambda_{k,j}/\mu_j(\mathbf{x})\right) \end{aligned} \tag{5}$$

The first term is the relative entropy and the other two form a primitive integral of Equation (4).

LEMMA 1. *Let $\mathbf{x}$ be a solution of the ODE (1). Then $\frac{d}{dt} L(\mathbf{x}(t)) \le 0$ with equality if and only if $\mathbf{x}(t) = \pi(\mathbf{x}(t))$.*

Proof. By definition of $\pi(\mathbf{x})$ in Equation (2), we have

$$\log \pi_{k,i}(\mathbf{x}) = \log\left(\prod_{j=1}^{i-1}\lambda_{k,j}\right) - \log\left(\prod_{j=1}^{i-1}\mu_j(\mathbf{x})\right) - \log\left(\sum_{i=1}^{h}\prod_{j=1}^{i-1}\lambda_{k,j}/\mu_j(\mathbf{x})\right) \quad (6)$$

As $\frac{d}{dt}\log(\prod_{j=1}^{i-1}\lambda_{k,j}) = 0$, this implies that

$$\sum_{i=1}^{h}\sum_{k=1}^{n} x_{k,i}(t)\frac{d}{dt}\log(\pi_{k,i}(\mathbf{x}(t))) = \sum_{i=1}^{h}\underbrace{\sum_{k=1}^{n} x_{k,i}(t)}_{=m_i}\frac{d}{dt}\log\left(\prod_{j=1}^{i-1}\mu_j(\mathbf{x})\right) \sum_{k=1}^{n}\underbrace{\sum_{i=1}^{h} x_{k,i}(t)}_{=1}\frac{d}{dt}\log\left(\sum_{j'=1}^{h}\prod_{j=1}^{j'-1}\lambda_{k,j}/\mu_j(\mathbf{x})\right)$$

This is equal to the the derivative with respect to time of the last two additional terms of Equation (5).

Applying this and Equation (4) implies that

$$\frac{d}{dt}L(\mathbf{x}(t)) = \sum_{i=1}^{h}\sum_{k=1}^{n}\left(\frac{d}{dt}x_{k,i}(t)\right)\left(1+\log\frac{x_{k,i}(t)}{\pi_{k,i}(\mathbf{x}(t))}\right).$$

By Equation (3), this implies that $\frac{d}{dt}L(\mathbf{x}(t)) \le 0$ with equality only if $\mathbf{x}(t) = \pi(\mathbf{x}(t))$. □

Lemma 2. *For any admissible* $\mathbf{x}$*, we have*

$$L(\mathbf{x}) \ge -n\log(h) - \sum_{k=1}^{n}\sum_{i=1}^{h}\left|\log\prod_{j=1}^{i-1}\lambda_{k,j}\right|.$$

Proof. Expanding the definition of the relative entropy, we have:

$$L(\mathbf{x}) = \sum_{k=1}^{n}\sum_{i=1}^{h} x_{k,i}\log x_{k,i} - \sum_{k=1}^{n}\sum_{i=1}^{h} x_{k,i}\log\pi_{k,i}(\mathbf{x}) - \sum_{i=1}^{h} m_i\log\left(\prod_{j=1}^{i-1}\mu_j(\mathbf{x})\right) - \sum_{k=1}^{n}\log\left(\sum_{i=1}^{h}\prod_{j=1}^{i-1}\lambda_{k,j}/\mu_j(\mathbf{x})\right)$$

By using the same transformation as in (6), we get:

$$\sum_{k=1}^{n}\sum_{i=1}^{h} x_{k,i}\log\pi_{k,i}(\mathbf{x}) = \sum_{k=1}^{n}\sum_{i=1}^{h} x_{k,i}\log\left(\prod_{j=1}^{i-1}\lambda_{k,j}\right) - \sum_{k=1}^{n}\sum_{i=1}^{h} x_{k,i}\log\left(\prod_{j=1}^{i-1}\mu_j(\mathbf{x})\right) - \sum_{k=1}^{n}\sum_{i=1}^{h} x_{k,i}\log\left(\sum_{j'=1}^{h}\prod_{j=1}^{j'-1}\lambda_{k,j}/\mu_j(\mathbf{x})\right)$$

This shows that

$$L(\mathbf{x}) = \sum_{k=1}^{n}\sum_{i=1}^{h} x_{k,i}\log x_{k,i} + \sum_{k=1}^{n}\sum_{i=1}^{h} x_{k,i}\log\left(\prod_{j=1}^{i-1}\lambda_{k,j}\right) \ge -n\log h - \sum_{i=1}^{h}\sum_{k=1}^{n}\left|\log\prod_{j=1}^{i-1}\lambda_{k,j}\right|,$$

where the first part of inequality comes from the entropy the second from the fact that $|x_{i,k}(t)| \le 1$. □

4. CONCLUSION AND DISCUSSION

In this paper, we constructed a Lyapunov function for a generalization of the RAND($\mathbf{m}$) model of [6]. This Lyapunov function is composed of a term of relative entropy term $R(\mathbf{x}\|\pi(\mathbf{x}))$ plus an additive correction term that compensate for the variation of $\pi(\mathbf{x})$. This function guarantees that the ODE approximation converges to its unique fixed point and hence justifies the fixed point analysis.

The construction of the Lyapunov function proposed in this paper is generic. We believe that similar ideas can be applied to many other mean-field systems. The two additive terms in the expression of the Lyapunov function (5) form a primitive integral of Equation (4). Their simple expression comes form the fact that in the stationary measure $\pi(\mathbf{x})$, the quantity $\mu_i(\mathbf{x})$ does not depend on the object k while $\lambda_{k,i}$ is does not depend on time. Note that the Lyapunov function of the model of [5, Section 3] has the same form.

The stationary measure of the original system is of Gibbs type since it is a product form measure on the set of possible configuration $\mathcal{C}$. However, this measure has not the same form as the Gibbs type measure of Section 4 of the papers [3, 4]. Hence, it seems that their Lyapunov function [3, Equation (4.14)] is not directly applicable here although the expression are close.

Finally, it is shown in [2] that, a sequence of reversible process that converges to the solution of an ODE has a sequence of stationary measures that concentrates on the fixed point of the ODE, regardless of the asymptotic behavior of this ODE. A benefit of exhibiting a Lyapunov function is that we are also able to characterize the asymptotic behavior of the ODE. It is not clear if the reversibility has an impact on the particular form of the Lyapunov function L.

References

[1] M. Benaïm and J. Le Boudec. A class of mean field interaction models for computer and communication systems. *Performance evaluation*, 65(11-12):823–838, 2008.

[2] J.-Y. L. Boudec. The stationary behaviour of fluid limits of reversible processes is concentrated on stationary points. *Arxiv:1009.5021*, 2010.

[3] A. S. Budhiraja, P. Dupuis, M. Fischer, and K. Ramanan. Limits of relative entropies associated with weakly interacting particle systems. *Electronic journal of probability*, 20, 2015.

[4] A. S. Budhiraja, P. Dupuis, M. Fischer, and K. Ramanan. Local stability of kolmogorov forward equations for finite state nonlinear markov processes. *Electronic journal of probability*, 20, 2015.

[5] C. Fricker and N. Gast. Incentives and redistribution in homogeneous bike-sharing systems with stations of finite capacity. *Euro journal on transportation and logistics*:1–31, 2014.

[6] N. Gast and B. Van Houdt. Transient and steady-state regime of a family of list-based cache replacement algorithms. In *Proceedings of the 2015 ACM SIGMETRICS international conference on measurement and modeling of computer systems*. ACM, 2015, pp. 123–136.