Ann. Inst. H. Poincaré, Probab. et Stat. 37, 2 (2001) 155–194

S0246-0203(00)01064-5/FLA

ON THE STABILITY OF INTERACTING PROCESSES WITH APPLICATIONS TO FILTERING AND GENETIC ALGORITHMS

Pierre DEL MORAL [a], **Alice GUIONNET** [b]

[a] *LSP, CNRS-UMR C5583, Bat.1R1, Univ. Paul Sabatier, 31062 Toulouse, France*
[b] *URA 743, CNRS, Bat. 425, Université de Paris Sud, 91405 Orsay, France*

Received in 18 March 1998, revised 28 October 1999

ABSTRACT. – The stability properties of a class of interacting measure valued processes arising in nonlinear filtering and genetic algorithm theory is discussed.

Simple sufficient conditions are given for exponential decays. These criteria are applied to study the asymptotic stability of the nonlinear filtering equation and infinite population models as those arising in Biology and evolutionary computing literature.

On the basis of these stability properties we also propose a uniform convergence theorem for the interacting particle numerical scheme of the nonlinear filtering equation introduced in a previous work. In the last part of this study we propose a refinement genetic type particle method with periodic selection dates and we improve the previous uniform convergence results. We finally discuss the uniform convergence of particle approximations including branching and random population size systems. © 2001 Éditions scientifiques et médicales Elsevier SAS

Keywords: Interacting random processes; Nonlinear filtering; Asymptotic stability; Measure valued processes; Genetic algorithms; Stochastic approximation

AMS classification: 60B10; 62E25; 62G09; 60G35; 93E11; 62L20

RÉSUMÉ. – Cet article porte sur la stabilité d'une classe de processus à valeurs mesures et en interaction liée au filtrage non linéaire et à la théorie des algorithmes génétiques.

Nous présentons des conditions suffisantes simples permettant d'obtenir des taux de convergence et des décroissances exponentielles. Nous illustrons notre démarche en appliquant ces critères à l'étude de la stabilité asymptotique des équations du filtrage non linéaire et d'une classes de systèmes à population infinie utilisés en biologie et dans la littérature sur les algorithmes de type génétique. En s'appuyant sur ces propriétés de stabilité nous démontrons un théorème de convergence uniforme dans le temps de schémas numériques basés sur des systèmes de particules en interaction. Dans la partie finale nous proposons un algorithme génétique plus performant associé à des sélections périodiques et pour lequel il est possible d'améliorer les précédentes estimations. Nous clôturons notre étude en étudiant le comportement en temps long d'une classe de méthodes particulaires basées sur des mécanismes de branchements avec des tailles de population aléatoires. © 2001 Éditions scientifiques et médicales Elsevier SAS

E-mail addresses: delmoral@cict.fr (P. Del Moral), guionnet@stats.matups.fr (A. Guionnet).

1. Introduction

In this paper, we study the long time behavior of a class of interacting measure valued processes arising in biology (and particularly in genetic models) and evolutionary computing, in Physics, in advanced signal processing and particularly in nonlinear filtering problems. The main results of the present work is to establish a theorem on the stability properties of the limiting process and a uniform convergence result with respect to the time parameter for the finite particle approximating model.

We shall consider the following discrete time evolution on the space $\mathcal{P}(E)$ of probability measures on a measurable space $(E,\mathcal{E})$ given by

$$\pi_n = \phi_n(\pi_{n-1}), \quad \forall n \geqslant 1, \tag{1}$$

where $\pi_0 \in \mathcal{P}(E)$ and the one step mapping $\phi_n : \mathcal{P}(E) \to \mathcal{P}(E)$ associates to any $\pi \in \mathcal{P}(E)$, the probability measure $\phi_n(\pi)$ given for any $f : E \to \mathbb{R}$ in the set of bounded measurable functions $\mathcal{B}_b(E)$ by

$$\phi_n(\pi)(f) = \frac{\pi(g_n(K_n f))}{\pi(g_n)} \quad \text{with } (K_n f)(x) \overset{\text{def.}}{=} \int_E K_n(x, dy) f(y)$$

with some transition probability kernels $(K_n, n \geqslant 0)$ and nonnegative measurable functions $(g_n, n \geqslant 0)$.

Such discrete evolutions arise in the analysis of some conditioned Markov processes (and in particular in those showing up in nonlinear filtering problems) and in biology.

Let us present first its connection with nonlinear filtering theory. Recall that the nonlinear filtering problem consists in computing the conditional distributions of internal states in dynamical systems when partial observations are made, and random perturbations are present in the dynamics as well as in the sensors.

Several presentations of the discrete time filtering model are available in the literature, here we follow rather closely [10]. Some collateral readings such as [4,24–26] will be useful in appreciating the relevance of our study.

In discrete time settings the signal $\{X_n;\ n \geqslant 0\}$ is an E valued nonhomogeneous Markov chain with one step transition probabilities $\{K_n;\ n \geqslant 1\}$ and initial law π_0. The observation sequence $\{Y_n;\ n \geqslant 1\}$ takes its values in $\mathbb{R}^d$, $d \geqslant 1$, and it takes the form

$$Y_n = H_n(X_{n-1}, V_n), \quad n \geqslant 1,$$

for some measurable function $H_n : E \times \mathbb{R}^d \to \mathbb{R}^d$. The sequence $V = \{V_n;\ n \geqslant 1\}$ are $\mathbb{R}^d$-valued, independent of X, and independent random variables. For each $x \in E$ and $n \geqslant 0$ we assume that the variable $H_n(x, V_n)$ admits a density $y \to \overline{g}_n(y, x)$ with respect to Lebesgue measure on $\mathbb{R}^d$.

Let $E(\boldsymbol{.})$ denotes the expectation on the original space on which the chain (X, Y) is Markov with the prescribed initial condition.

For any fixed observation sequence $y_1, y_2, \ldots$ Bayes' formula gives a Feynman–Kac expression for the desired conditional distributions, namely for any bounded measurable function f we have

$$\pi_n^y(f) = E\big(f(X_n) \mid Y_1 = y_1, \dots, Y_n = y_n\big) = \frac{E(f(X_n)Z_n(X,y))}{E(Z_n(X,y))}$$

with

$$Z_n(X,y) \stackrel{\text{def.}}{=} \prod_{k=1}^{n} \overline{g}_k(y_k, X_{k-1}).$$

It is transparent from this description that

$$\pi_n^y(f) = \phi_n\big(y_n, \pi_{n-1}^y\big)(f) \stackrel{\text{def.}}{=} \frac{\pi_{n-1}^y(g_n^y(K_n f))}{\pi_{n-1}^y(g_n^y)} \quad \text{with } g_n^y \stackrel{\text{def.}}{=} \overline{g}_n(y_n, x). \tag{2}$$

In these settings the dynamical system (2) is called the nonlinear filtering equation. In some particular situations it can be solved explicitly but in general its simulation requires extensive calculations.

The discrete dynamics (1) also appears when one considers a killed inhomogeneous Markov process. Indeed, let $\{X_n;\ n \geqslant 0\}$ be a Markov chain taking values in E with initial distribution $\pi_0 \in \mathcal{P}(E)$ and transition probability kernels

$$\forall A \in \mathcal{E} \quad P(X_n \in A \mid X_{n-1}) = K_n(X_{n-1}, A).$$

Using the Markov property we can verify that the Feynman–Kac type distributions given for any $f \in \mathcal{B}_b(E)$ by

$$\pi_n(f) = \frac{E(f(X_n)Z_n(X))}{E(Z_n(X))}, \quad \text{with } Z_n(X) \stackrel{\text{def.}}{=} \prod_{k=1}^{n} g_k(X_{k-1}) \tag{3}$$

(with the usual convention $\prod_{\emptyset} = 1$) satisfy the desired recursion (1), that is

$$\pi_n(f) = \frac{E(g_n(X_{n-1})(K_n f)(X_{n-1})Z_{n-1}(X))/E(Z_{n-1}(X))}{E(g_n(X_{n-1})Z_{n-1}(X))/E(Z_{n-1}(X))} = \frac{\pi_{n-1}(g_n(K_n f))}{\pi_{n-1}(g_n)} = \phi_n(\pi_{n-1})(f).$$

These "un-normalized" Feynman–Kac formulae arise naturally in the study of the distributions laws of a Markov process killed particle. We begin by noting that there is no loss of generality to assume that the fitness functions g_n take values in $(0, 1]$. Next, the fitness functions will be regarded as the killing rates of a nonhomogeneous Markov particle. We adjoin classically to E a cemetery point Δ and we define the Markov transitions kernels $\{\tilde{K}_n;\ n \geqslant 1\}$ by setting for any $A \in \mathcal{E}$ and $n \geqslant 1$ and $x \in E$

$$\tilde{K}_n(x, A) = g_n(x)\,.\,K_n(x, A)$$

and

$$\forall x \in E \quad \tilde{K}_n(x, \{\Delta\}) = 1 - g_n(x) \quad \text{and} \quad \tilde{K}_n(\{\Delta\}, \{\Delta\}) = 1.$$

If $\{\tilde{X}_n;\ n \geqslant 0\}$ denotes the corresponding Markov process on $E \cup \{\Delta\}$ then we have for any subset $A \in \mathcal{E}$

$$P\big(\tilde{X}_n \in A \mid T^{\Delta} > n\big) = \frac{E(f(X_n)Z_n(X))}{E(Z_n(X))},$$

where $T^{\Delta} = \inf\{n \geqslant 0;\ \tilde{X}_n = \Delta\}$ is the life time of $\tilde{X}$.

We shall now interpret (1) as the limit of a finite, weakly interacting particle systems encountered in discrete generation genetic models. This discrete approximation of (1) can be as well used as a numerical approximating model of (2). The advantage of this approximation scheme among the numerous existing ones is that it guarantees an occupation of the probability space regions proportional to their fitness thus providing a well behaved adaptative and stochastic grid approximating model.

To our knowledge the Feynman–Kac interpretation (3) of the limiting system (1) of the genetic algorithm has never been covered in genetic and evolutionary computing literature but only in numerical filtering papers [12,13].

The class of measure valued processes described by (1) arises naturally as the deterministic limit of the empirical measures of a finite interacting particle system (abbreviate **IPS**). To make this more precise we recall that the N-IPS approximating model associated to (1) is the Markov process $(\Omega, (F_n)_{n\geqslant 0}, (\xi_n)_{n\geqslant 0}, P)$ taking values in the product space E^N and defined by

$$P(\xi_0 \in dz) = \prod_{p=1}^{N} \pi_0\big(dz^p\big)$$

and

$$P(\xi_n \in dz/\xi_{n-1} = x) = \prod_{p=1}^{N} \phi_n\left(\frac{1}{N}\sum_{i=1}^{N} \delta_{x^i}\right)\big(dz^p\big), \tag{4}$$

where $dz \stackrel{\text{def.}}{=} dz^1 \times \cdots \times dz^N$ is an infinitesimal neighborhood of the point $z = (z^1, \ldots, z^N) \in E^N$ and $x = (x^1, \ldots, x^N) \in E^N$.

Since

$$\phi_n\left(\frac{1}{N}\sum_{i=1}^{N} \delta_{x^i}\right) = \sum_{i=1}^{N} \frac{g_n(x^i)}{\sum_{j=1}^{N} g_n(x^j)} K_n\big(x^i, \boldsymbol{\cdot}\big)$$

we see that the motion of the particles is decomposed into two stages

$$\xi_{n-1} = \big(\xi_{n-1}^1, \ldots, \xi_{n-1}^N\big) \longrightarrow \widehat{\xi}_{n-1} = \big(\widehat{\xi}_{n-1}^1, \ldots, \widehat{\xi}_{n-1}^N\big) \longrightarrow \xi_n = \big(\xi_n^1, \ldots, \xi_n^N\big).$$

The first one updates the positions in accordance with the fitness functions $\{g_n;\ n \geqslant 1\}$ and the current configuration. More precisely, at each time $n \geqslant 1$, each particle examines the system of particles $\xi_{n-1} = (\xi_{n-1}^1, \ldots, \xi_{n-1}^N)$ and chooses randomly a site ξ_{n-1}^i, $1 \leqslant i \leqslant N$, with a probability which depends on the entire configuration ξ_{n-1}, namely

$$\frac{g_n(\xi_{n-1}^i)}{\sum_{j=1}^{N} g_n(\xi_{n-1}^j)}.$$

This mechanism is called the Selection/Updating transition as the particles are selected for reproduction the more fit individuals being more likely to be selected. In other words this transition allows particles to give birth to some particles at the expense of light particle which die.

The second mechanism is called the Mutation/Prediction. During this stage each particle evolves randomly according a given transition probability kernel.

This scheme is clearly a system of interacting particles undergoing adaptation in a time non-homogeneous environment represented by the fitness functions $\{g_n;\ n \geqslant 1\}$.

In biology and evolutionary computing literature this N-IPS model corresponds to a discrete generation genetic model with haploid selection. The term, *haploid*, refers to the fact that the selection function depends only on the type of one "parent" rather than two. The N-IPS model (4) is sometimes referred as a simple or canonical genetic algorithm mainly because it does not involve cross-over transitions and it simply uses a proportional selection mechanism. In this connection the limiting system (1) is also referred, in genetic terminology, as the infinite population model.

Let us briefly survey some different works related to our subject and motivate our work.

In view of the previous discussion the measure valued dynamical system (1) and its N-IPS approximating model (4) arise in a variety of research areas and similar problems have been studied by many authors.

In evolutionary computing literature the connections between the long time behavior of the finite and infinite population models are discussed for instance in [28,31–36] but no satisfactory analysis was done to obtain precise stability properties and/or convergence results. Some authors have made the sanguine assumption that the one step mappings ϕ_n are contractive (see for instance [32], p. 401). As we shall see in the further development of Section 2 this assumption is very restrictive and the resulting stability results do not apply to many situations of interest.

In measure valued processes and IPS literature the connections between the stability properties of the limiting system (1) and the convergence of the N-IPS scheme have also been studied in [23]. The author gives an ergodicity criterion on the limiting system under which the N-IPS scheme converges uniformly with respect to the time parameter. This result also applies to related genetic algorithms such as the Wright–Fisher model but it relies on the fact that the limiting system is homogeneous and uniformly converges to a distribution (with respect to its initial conditions) and it does not discuss any rates of convergence.

In nonlinear filtering settings the study of the long time behavior of the filtering equation is a more active research area. The motivations here come from the fact that the initial law of the signal is usually unknown and it is therefore essential to check whether or not the nonlinear filtering equation "forgets" any erroneous initial distribution. The papers [21,22,29,30] are mainly concerned with the existence of invariant probability measures and in [25] the authors prove that the filtering equation "forgets" any erroneous initial condition if the unknown initial law of the signal is absolutely continuous with respect to this new starting point. In the following chain of papers [1–3,5,9,18] the authors discuss the stability properties of (2) using Hilbert projective metrics and/or

Oseledec's Theorem when the state space is finite. The first place in which such Hilbert projective metrics have been used in filtering settings seems to be [9].

In this paper we propose a new approach based on semi-group techniques and Dobrushin ergodic coefficient (cf. [19]). In contrast to the previous referenced papers an advantage of our methodology is that it is not restricted to filtering or genetic algorithms and it allows to study any nonhomogeneous systems of the form (1). The other main result of the paper is to connect the long time behavior of (1) with the uniform convergence of the N-IPS approximating model.

Let us state the main results of this paper more explicitly. Our first goal is to study the long time behavior of the limiting process (1). Let $\{\phi_{n/p};\ 0 \leqslant p \leqslant n\}$ be the nonlinear semi-group associated to (1) and defined by the composite mappings

$$\phi_{n/p} = \phi_n \circ \cdots \circ \phi_{p+1}, \quad 0 \leqslant p \leqslant n,$$

with the convention $\phi_{n,n} = Id$. If we denote by

$$\|\mu - \nu\|_{tv} \overset{\text{def.}}{=} \sup_{A \in \mathcal{E}} \left|\mu(A) - \nu(A)\right|, \quad \forall \mu, \nu \in \mathcal{P}(E),$$

the total variation distance on $\mathcal{P}(E)$ then our main result will then be basically stated under the following form.

THEOREM 1.1. – *If the functions $\{g_n,\ n \geqslant 1\}$ and the Markov operators $\{K_n,\ n \geqslant 1\}$ are "good enough" then*

$$\forall \mu, \nu \in \mathcal{P}(E), \quad \lim_{n \to \infty} \|\phi_{n/0}(\mu) - \phi_{n/0}(\nu)\|_{tv} = 0. \tag{5}$$

The crucial point is to specify the assumptions needed on the fitness functions $\{g_n,\ n \geqslant 1\}$ and the Markov operators $\{K_n,\ n \geqslant 1\}$ for such result to hold. Throughout this paper we shall weaken these hypotheses as much as we can in order to include as many examples encountered in nonlinear filtering and genetic algorithm theory.

We will also propose a mixing type condition on the transition probability kernels K_n under which the convergence in (5) takes place exponentially fast in the sense that there exists some positive constant $\lambda > 0$ such for any $0 \leqslant p \leqslant n$

$$\sup_{\mu, \nu \in \mathcal{P}(E)} \|\phi_{n/p}(\mu) - \phi_{n/p}(\nu)\|_{tv} \leqslant \mathrm{e}^{-\lambda \cdot (n-p)}.$$

The other main result of the paper is to connect the stability properties of the limiting system (1) with the long time behavior of the empirical measures π_n^N associated to the N-IPS scheme (4). Recall that it was proven in [12,13,15] that the empirical measure

$$\pi_n^N = \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi_n^i}$$

converges in finite time intervals towards the desired distribution $\{\pi_n, n \geqslant 0\}$ as N goes to infinity. The large deviations and the fluctuations for this convergence are developed in [16,17], but we let open the question of the long time behavior of this particle approximations. Here, we shall prove that

THEOREM 1.2. – *When the functions* $\{g_n;\ n \geqslant 1\}$ *and the Markov transitions* $\{K_n;\ n \geqslant 1\}$ *are "sufficiently regular" there exists a convergence exponent* $\alpha > 0$ *such that for any bounded measurable function* f

$$\sup_{n\geqslant 0} E\left(\left|\pi_n^N f - \pi_n f\right|\right) \leqslant \frac{\mathrm{Cte}}{N^\alpha}\|f\|.$$

1.1. Notations and terminology

In view of the previous development the interpretation of (1) and the corresponding particle scheme (4) may vary considerably. Most of the terminology we will use is drawn from mean field IPS and genetic algorithm theory.

The deterministic nonlinear evolution equation (1) will be regarded as the limiting measure-valued process associated with a sequence of N-IPS schemes. With reference to genetic algorithm theory the functions $\{g_n;\ n \geqslant 1\}$ will be called the fitness functions and the transitions $\{K_n;\ n \geqslant 1\}$ will be referred as the mutation transitions.

With a slight abuse of the classical mathematical terminology, we shall say that the limiting system is asymptotically stable when its long time time behavior does not depend on its initial condition.

To describe more precisely the dynamical structure of (1) we need to introduce some additional notations. We first recall that any Markov transition kernel $K(x, dy)$ on E generates two integral operators one acting on functions $f \in \mathcal{B}_b(E)$ and the other on probability distributions $\mu \in \mathcal{P}(E)$

$$(Kf)(x) = \int_E K(x, dy) f(y), \qquad (\mu K)(dy) = \int_E \mu(dx) K(x, dy).$$

As usual $\mathcal{B}_b(E)$ is regarded as a Banach space endowed with the supremum norm

$$\|f\| = \sup_{x\in E} |f(x)|.$$

If K_1, K_2 are two Markov transition kernels on E we write K_1K_2 the composite Markov transition kernel given by

$$K_1K_2(x, dz) = \int_E K_1(x, dy) K_2(y, dz).$$

As announced our approach is based on semi-group techniques and we will use the powerful tools developed by R.L. Dobrushin to study central limit Theorems for nonstationary Markov chains [19]. We recall that if K is a Markov transition on E then the ergodic coefficient of K is the quantity $\alpha(K) \in [0, 1]$ given by

$$\alpha(K) = 1 - \beta(K) \quad \text{with } \beta(K) \stackrel{\text{def.}}{=} \sup_{x,y\in E, A\in\mathcal{E}} |K(x, A) - K(z, A)|.$$

We shall call the number $\alpha(K)$ the Dobrushin ergodic coefficient of K (see [19]). It can also be defined as

$$\alpha(K) = \inf \sum_{i=1}^{m} \min\big(K(x, A_i), K(z, A_i)\big), \tag{6}$$

where the infimum is taken over all $x, z \in E$ and all finite partitions of E, $\{A_i;\ 1 \leqslant i \leqslant m\}$ with $m \geqslant 1$.

The quantity $\alpha(K)$ is a measure of contraction of the distance of probability measures induced by the Markov operator K. Namely, for any $\mu, \nu \in \mathcal{P}(E)$ we have the well known formula (see [19])

$$\beta(K) = \sup_{\mu,\nu\in\mathcal{P}(E)} \frac{\|\mu K - \nu K\|_{tv}}{\|\mu - \nu\|_{tv}}. \tag{7}$$

An outline of the development of the article is as follows:

The main result on the asymptotic stability of the limiting system (1) is given in Section 2. In a preliminary Section 2.1 we examine in more details the dynamical structure of (1). In a short Section 2.2 we present a very basic proof of Theorem 1.1 in the situation where the state space E is finite. In Section 2.3 we present a semi-group technique to study the stability properties of (1) when the state space E is an arbitrary measurable space. Several examples including nonlinear filtering problems are studied in some details in Section 2.4. The uniform convergence result of the N-IPS scheme as $N \to \infty$ is discussed in Section 3. In Section 3.1 we state and prove our main theorem. In Section 3.2 we present a novel genetic algorithm with periodic selections and we improve the uniform convergence decays given in Section 3.1. The last Section 3.3 we propose a comparison of genetic type variants recently suggested in nonlinear filtering literature including branching transitions and random population size models.

2. Asymptotic stability theorems

2.1. Introduction

In this section we discuss the asymptotic stability properties of the limiting system (1). Before getting into the details it may be useful to make a couple of remarks regarding the dynamical structure of (1).

In the first place it should be recalled that the limiting system (1) is a two stage process. More precisely, the one step mappings ϕ_n can be rewritten as follows

$$\forall \pi \in \mathcal{P}(E), \quad \phi_n(\pi) = \psi_n(\pi) K_n, \quad \text{with } \forall f \in \mathcal{B}_b(E), \quad \psi_n(\pi)(f) \overset{\text{def.}}{=} \frac{\pi(g_n f)}{\pi(g_n)}$$

with a selection ψ_n and a mutation K_n.

This first observation already indicates that the resulting system (1) may have completely different kinds of long time behavior.

For instance, if the fitness functions are constant functions then (1) is simply based on mutation transitions and it describes the time evolution of the distributions of the Markov process $\{X_n;\ n \geqslant 0\}$ with transition $(K_n, n \geqslant 0)$. In this very special case the theory of Markov processes and stochastic stability can be applied.

On the other hand, if the transition probability kernels $\{K_n;\ n \geqslant 1\}$ are trivial, that is $K_n = Id$ for any $n \geqslant 1$, then (1) is only based on selection transitions and its long time behavior is strongly related on its initial value. For instance, if $g_n = \exp(-U)$, for some $U : E \to \mathbb{R}_+$, then for any bounded continuous function $f : E \to \mathbb{R}_+$ with compact support

$$\pi_n(f) = \frac{\pi_0(f\,\mathrm{e}^{-nU})}{\pi_0(\mathrm{e}^{-nU})} \xrightarrow[n\to\infty]{} \frac{\pi_0(f 1_{U^\star})}{\pi_0(U^\star)},$$

where, at least if $\pi_0(U^*) > 0$,

$$U^\star \stackrel{\text{def.}}{=} \{x \in E;\ U(x) = \operatorname{essinf}_{\pi_0} U\}.$$

The second remark is the fact that the one step mappings ϕ_n usually fail to be contractive and the classical tools of dynamical system theory cannot be used to study the stability properties of the system (1).

By way of example, let us suppose that the one step mappings ϕ_n are time homogeneous (i.e., $g_n = g$ and $K_n = K$) and the Markov operator K is a strict contraction with respect (w.r.t.) to the total variation norm, that is

$$\forall \mu, \nu \in \mathcal{P}(E) \quad \|\mu K - \nu K\|_{tv} \leqslant \beta(K) \|\mu - \nu\|_{tv} \quad \text{with } \beta(K) < 1.$$

Using the above inequality we obtain

$$\|\phi(\mu) - \phi(\nu)\|_{tv} = \|\psi(\mu)K - \psi(\nu)K\|_{tv} \leqslant \beta(K)\|\psi(\mu) - \psi(\nu)\|_{tv}.$$

It is therefore tempting to check that ψ is nonexpansive w.r.t. the total variation norm. Unfortunately, this property is intimately related to the form of the function g. Let us examine a situation in which ψ is not contractive.

Namely, let us assume that

$$E = \{0, 1\}, \quad \mu = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1, \quad \text{and} \quad \nu = \delta_0. \tag{8}$$

In this situation it is easily checked that

$$\|\mu - \nu\|_{tv} = \frac{1}{2} \quad \text{and} \quad \|\psi(\mu) - \psi(\nu)\|_{tv} = \frac{g(1)}{g(1) + g(0)}$$

so that

$$g(1) > g(0) \Longrightarrow \|\psi(\mu) - \psi(\nu)\|_{tv} > \|\mu - \nu\|_{tv}$$

and consequently ψ is not contractive. Let us also remark that in this simple example we have that

$$\|\mu K - \nu K\|_{tv} = \beta(K)\|\mu - \nu\|_{tv} \quad \text{with } \beta(K) = |K(0, 0) - K(1, 0)|,$$

for any Markov operator K on E. When the distributions μ and ν are given by (8) it does follow that

$$\|\phi(\mu)-\phi(\nu)\|_{tv}=2\beta(K)\frac{g(1)}{g(1)+g(0)}\|\mu-\nu\|_{tv},$$

and hence that

$$\left(\frac{1}{2}<\beta(K)\quad\text{and}\quad g(0)<2(\beta(K)-1/2)g(1)\right)\Longrightarrow\|\phi(\mu)-\phi(\nu)\|_{tv}>\|\mu-\nu\|_{tv}.$$

In the last example, the homogeneous one step mappings $(\phi_n,\ n\geqslant 0)$ are not contractive but we shall see that the composite mappings $\{\phi_{n/p};\ 0\leqslant p\leqslant n\}$ can still be stable. To do so, we need to investigate more closely the dynamical structure of the limiting system (1). Our analysis will be based on the following lemma which roughly says that the composite mappings $\{\phi_{n/p};\ 0\leqslant p\leqslant n\}$ have essentially the same form as the one step mappings $\{\phi_n;\ n\geqslant 1\}$.

LEMMA 2.1 [12]. – *For any* $0\leqslant p\leqslant n$, $\mu\in\mathcal{P}(E)$ *and* $f\in\mathcal{B}_b(E)$ *we have*

$$\phi_{n/p}(\mu)f=\frac{\mu(g_{n/p}\,(K_{n/p}f))}{\mu(g_{n/p})},$$

where the fitness functions $\{g_{n/p};\ 0\leqslant p\leqslant n\}$ *and the Markov transitions* $\{K_{n/p};\ 0\leqslant p\leqslant n\}$ *satisfy the backward formulae*

$$K_{n/p-1}f=\frac{K_p(g_{n/p}(K_{n/p}f))}{K_p(g_{n/p})},\quad g_{n/p-1}=g_pK_p(g_{n/p}),\tag{9}$$

with the conventions $g_{n/n}=1$ *and* $K_{n/n}=Id$.

For the convenience of the reader we indicate that this lemma can be proved using a clear backward induction on the parameter $p(\leqslant n)$.

It is transparent from the backward recursions (9) that the Markov operators $\{K_{n/p};\ 0\leqslant p\leqslant n\}$ are composite operators of time-inhomogeneous but linear Markov operators. More precisely it can be checked directly that

$$K_{n/p-1}=S_{n/p}K_{n/p}=S_{n/p}S_{n/p+1}\cdots S_{n/n-1}S_{n/n},\tag{10}$$

with

$$S_{n/p}f=\frac{K_p(g_{n/p}f)}{K_p(g_{n/p})},\quad 0\leqslant p\leqslant n.$$

2.2. Finite state space

In this short subsection we consider the case where E is finite. We introduce this result in the article since it shows in a natural and very simple way how some simple properties on the functions $\{g_n;\ n\geqslant 1\}$ and the stochastic matrices $\{K_n;\ n\geqslant 1\}$ combine to give the desired asymptotic stability of (1). However, the present result will be strengthened in the next subsection. For simplicity we will always assume that

(A) *There exists an* $0 < \varepsilon \leqslant 1$ *such that*

$$\forall n \geqslant 1, x, z \in E \quad \varepsilon \leqslant g_n(x) \leqslant 1 \quad \varepsilon \leqslant K_n(x, z) \leqslant 1.$$

In this situation a clear combination of (9) yields that

$$\frac{g_{n/p}(z)K_p(x,z)}{g_{n/p}(z')K_p(x,z')} = \frac{g_{p+1}(z)K_p(x,z)K_{p+1}(g_{n/p+1})(z)}{g_{p+1}(z')K_p(x,z')K_{p+1}(g_{n/p+1})(z')} \leqslant 1/\varepsilon^3.$$

Thus

$$S_{n/p}(x,z) = \frac{g_{n/p}(z)K_p(x,z)}{\sum_{z'} g_{n/p}(z')K_p(x,z')} \geqslant \lambda, \tag{11}$$

with

$$\lambda^{-1} = 1 + (|E| - 1)\varepsilon^{-3}.$$

This gives us the following theorem

THEOREM 2.2. – *Under assumption* (A) *we have that*

$$\forall x \in E, \mu, \nu \in \mathcal{P}(E) \quad |\phi_{n/0}(\mu)(x) - \phi_{n/0}(\nu)(x)| \leqslant (1-\lambda)^n.$$

Proof. – If we put

$$K^+_{n/p}(x) = \sup_{z \in E} K_{n/p}(z, x), \qquad K^-_{n/p}(x) = \inf_{z \in E} K_{n/p}(z, x),$$

and

$$K^+_{n/p}(x) = K_{n/p}(x^+_{n/p}, x), \qquad K^-_{n/p}(x) = K_{n/p}(x^-_{n/p}, x)$$

using the decomposition (10) we find that

$$K^+_{n/p-1}(x) \leqslant K^-_{n/p}(x) S_{n/p}(x^+_{n/p-1}, x^-_{n/p}) + K^+_{n/p}(x)\big(1 - S_{n/p}(x^+_{n/p-1}, x^-_{n/p})\big)$$

and

$$K^-_{n/p-1}(x) \geqslant K^+_{n/p}(x) S_{n/p}(x^-_{n/p-1}, x^+_{n/p}) + K^-_{n/p}(x)\big(1 - S_{n/p}(x^-_{n/p-1}, x^+_{n/p})\big).$$

Thus, from (11) one gets the inequality

$$\big(K^+_{n/p-1}(x) - K^-_{n/p-1}(x)\big) \leqslant (1-\lambda)\big(K^+_{n/p}(x) - K^-_{n/p}(x)\big).$$

It is then an elementary matter to prove that

$$\sup_{z,z'} |K_{n/p}(z,x) - K_{n/p}(z',x)| \leqslant (1-\lambda)^{n-p}.$$

The end of proof of the theorem is now straightforward. □

2.3. Measurable state space

The purpose of this section is to present a natural sufficient condition for asymptotic stability of (1) when the state space E is an arbitrary measurable space. The starting point to do it is to use the Dobrushin ergodic coefficient (6) to obtain the following pivotal lemma

LEMMA 2.3. – *For any* $0 \leqslant p \leqslant n$ *we have*

$$\sup_{\mu,\nu\in\mathcal{P}(E)} \|\phi_{n/p}(\mu)-\phi_{n/p}(\nu)\|_{tv} = \beta(K_{n/p}) \leqslant \prod_{q=p+1}^{n} \big(1-\alpha(S_{n/q})\big)$$

and therefore for any $p \geqslant 0$

- $\lim_{n\to\infty} \sum_{q=p+1}^{n} \alpha(S_{n/q}) = \infty \Longrightarrow \lim_{n\to\infty} \beta(K_{n/p}) = 0.$
- $\lim_{n\to\infty} \frac{1}{n} \sum_{q=p+1}^{n} \alpha(S_{n/q}) \stackrel{\text{def.}}{=} \overline{\alpha}(S) \Longrightarrow \limsup_{n\to\infty} \frac{1}{n} \log \beta(K_{n/p}) \leqslant -\overline{\alpha}(S).$
- $\inf_{0\leqslant p_1\leqslant p_2} \alpha(S_{p_2/p_1}) \stackrel{\text{def.}}{=} \alpha(S) \Longrightarrow \forall n \geqslant p \quad \beta(K_{n/p}) \leqslant \mathrm{e}^{-\alpha(S).(n-p)}.$

Proof. – In the first place, note that for any $\mu \in \mathcal{P}(E)$ and $0 \leqslant p \leqslant n$

$$\phi_{n/p}(\mu) = \mu_{n/p} K_{n/p} \quad \text{with } \mu_{n/p}(f) = \frac{\mu(g_{n/p} f)}{\mu(g_{n/p})}.$$

Recalling that

$$K_{n/p} = S_{n/p+1} S_{n/p+2} \cdots S_{n/n}$$

and using (7) we obtain

$$\begin{aligned} \|\phi_{n/p}(\mu)-\phi_{n/p}(\nu)\|_{tv} &= \|\mu_{n/p}K_{n/p} - \nu_{n/p}K_{n/p}\|_{tv} \\ &\leqslant \beta(K_{n/p}) \|\mu_{n/p} - \nu_{n/p}\|_{tv}. \end{aligned} \tag{12}$$

Since for any $x \in E$

$$\phi_{n/p}(\delta_x) = K_{n/p}(x, \boldsymbol{.})$$

it follows that

$$\begin{aligned} \beta(K_{n/p}) &= \sup_{x,y} \|K_{n/p}(x,\boldsymbol{.}) - K_{n/p}(y,\boldsymbol{.})\|_{tv} \\ &\leqslant \sup_{\mu,\nu\in\mathcal{P}(E)} \|\phi_{n/p}(\mu)-\phi_{n/p}(\nu)\|_{tv}. \end{aligned}$$

The reverse inequality is a consequence of (12). Taking into account that

$$K_{n/q-1} = S_{n/q} K_{n/q}, \quad \forall 1 \leqslant q \leqslant n,$$

it is easily seen that

$$\beta(K_{n/q-1}) \leqslant \big(1-\alpha(S_{n/q})\big)\beta(K_{n/q}), \quad \forall 1 \leqslant q \leqslant n,$$

from which one concludes that

$$\beta(K_{n/p}) \leqslant \prod_{q=p+1}^{n} \big(1-\alpha(S_{n/q})\big)$$

and the first part of the lemma is proved. On the other hand if we use the inequality

$$\prod_{q=p+1}^{n} \big(1-\alpha(S_{n/q})\big) \leqslant \exp -\left(\sum_{q=p+1}^{n} \alpha(S_{n/q})\right)$$

the end of the proof of the lemma is straightforward. □

In view of the previous considerations we see that the collection of Markov transitions $\{S_{n/p};\ 0 \leqslant p \leqslant n\}$ plays a pivotal role in the study of the asymptotic stability properties of the limiting system (1).

To control the Dobrushin coefficients of $\{S_{n/p};\ 0 \leqslant p \leqslant n\}$, we shall now make the following natural assumptions

(B) *For any time* $n \geqslant 1$, *there exists a reference probability measure* $\lambda_n \in \mathcal{P}(E)$ *and a positive number* $\varepsilon_n \in (0,1]$ *so that* $K_n(x,\boldsymbol{.}) \sim \lambda_n$ *for any* $x \in E$ *and*

$$\varepsilon_n \leqslant \frac{dK_n(x,\boldsymbol{.})}{d\lambda_n} \leqslant \frac{1}{\varepsilon_n}.$$

In contrast to (A) condition (B) doesn't depend anymore on the fitness functions $\{g_n;\ n \geqslant 1\}$. One way to relax (B) is to take advantage of the specific structure of the fitness functions $\{g_{n/p};\ 0 \leqslant p \leqslant n\}$ defined in Lemma 2.1. The price to pay is that the resulting condition now depends on the boundedness of the fitness functions. In this situation we will use the next condition.

(C) *For any* $n \geqslant 1$ *there exists an* $a_n \in [1,\infty)$ *such that*

$$\frac{1}{a_n} \leqslant g_n(x) \leqslant a_n, \quad \forall x \in E,\ \forall n \geqslant 1.$$

In addition, the mutation transitions are homogeneous (*that is* $K_n = K$) *and there exists an* $m \geqslant 1$ *and* $\varepsilon > 0$ *and a reference probability measure* $\lambda \in \mathcal{P}(E)$ *such that*

$$\varepsilon \leqslant \frac{dK^m(x,\boldsymbol{.})}{d\lambda} \leqslant \frac{1}{\varepsilon}, \quad \forall x \in E.$$

In the last condition the mutation transitions are assumed to be homogeneous, the generalization to nonhomogeneous Markov transitions will be straightforward.

We can now show the

THEOREM 2.4. –

- *If condition* (B) *holds for some sequence of nonnegative numbers* $\{\varepsilon_n;\ n \geqslant 1\}$ *such that*

$$\sum_{n\geqslant 1} \varepsilon_n^2 = \infty,$$

then we have

$$\lim_{n\to\infty} \sup_{\mu,\nu\in\mathcal{P}(E)} \|\phi_{n/0}(\mu) - \phi_{n/0}(\nu)\|_{tv} = 0.$$

In addition, suppose that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{p=1}^{n} \varepsilon_p^2 \stackrel{\text{def.}}{=} \varepsilon^2,$$

then

$$\limsup_{n\to\infty} \frac{1}{n} \log \sup_{\mu,\nu\in\mathcal{P}(E)} \|\phi_{n/0}(\mu) - \phi_{n/0}(\nu)\|_{tv} \leqslant -\varepsilon^2.$$

Moreover when

$$\inf_{n\geqslant 1} \mathrm{e}_n \stackrel{\text{def.}}{=} \varepsilon$$

one concludes that for any $0 \leqslant p \leqslant n$

$$\sup_{\mu,\nu\in\mathcal{P}(E)} \|\phi_{n/p}(\mu) - \phi_{n/p}(\nu)\|_{tv} < \exp -\big(\varepsilon^2 . (n-p)\big).$$

- *When the fitness functions and the mutation transitions satisfy condition* (C) *for some* $m \geqslant 1$, $\varepsilon > 0$ *and* $a \stackrel{\text{def.}}{=} \sup_n a_n < \infty$ *then we have for any* $n \geqslant m \geqslant 0$

$$\frac{1}{n} \log \sup_{\mu,\nu\in\mathcal{P}(E)} \|\phi_{n/0}(\mu) - \phi_{n/0}(\nu)\|_{tv} \leqslant -\left(1 - \frac{m}{n}\right)\alpha(K)\left(\frac{\varepsilon}{a^m}\right)^2. \tag{13}$$

Proof. – Under (B) we first note that each coefficient $\alpha(S_{n/p})$ is lower bounded by ε_p^2. To see this claim, note that for any $x \in E$ and $A \in \mathcal{E}$

$$S_{n/p}(x, A) = \frac{K_p(g_{n/p}1_A)(x)}{K_p(g_{n/p})(x)} \geqslant \varepsilon_p^2 \frac{\lambda_p(g_{n/p}1_A)}{\lambda_p(g_{n/p})}$$

so that (6) implies that

$$\alpha(S_{n/p}) \geqslant \varepsilon_p^2. \tag{14}$$

The end of the proof of the first part of the theorem is then is clear consequence of Lemma 2.1. Let us assume that condition (C) holds for some $m \geqslant 1$, $\varepsilon > 0$ and $a \stackrel{\text{def.}}{=} \sup_n a_n < \infty$. An induction on the parameter m yields

$$a^{-2m} \frac{K(K^m(g_{n/p+m})\varphi)}{K(K^m(g_{n/p+m}))} \leqslant S_{n/p}\varphi \leqslant a^{2m} \frac{K(K^m(g_{n/p+m})\varphi)}{K(K^m(g_{n/p+m}))}$$

for any $0 \leqslant p+m \leqslant n$ and any bounded nonnegative function $\varphi : E \to \mathbb{R}_+$. This in turn implies that

$$\left(\frac{\varepsilon}{a^m}\right)^2 \alpha(K) \leqslant \alpha(S_{n/p}) \leqslant \left(\frac{a^m}{\varepsilon}\right)^2 \alpha(K)$$

and inequality (13) is again a consequence of Lemma 2.1. □

Next we consider a corollary of Theorem 2.4 in the time homogeneous situation which is particularly important in genetic algorithm theory. When the fitness functions and the mutation transitions are homogeneous (that is $g_n = g$ and $K_n = K$) the resulting one step mapping of the limiting system (1) is again homogeneous and it is defined by

$$\phi(\mu) = \psi(\mu)K \quad \text{with } \psi(\mu)(f) = \frac{\mu(gf)}{\mu(g)}$$

for any $f \in \mathcal{B}_b(E)$ and $\mu \in \mathcal{P}(E)$.

COROLLARY 2.5. – *Let E be a finite state space. Assume that the fitness functions and the Markov transitions are homogeneous.*

- *If condition* (B) *holds for some $\varepsilon > 0$ then the corresponding homogeneous one step mapping ϕ has a unique fixed point, that is*

$$\exists! \pi_\infty \in \mathcal{P}(E) \colon \ \pi_\infty = \phi(\pi_\infty),$$

and for any $0 \leqslant p \leqslant n$ we have that

$$\sup_{\mu \in \mathcal{P}(E)} \|\phi_{n/p}(\mu) - \pi_\infty\|_{tv} < \exp -\big(\varepsilon^2 \,.\, (n-p)\big).$$

- *When the fitness functions and the mutation transitions satisfy condition* (C) *for some $m \geqslant 1$, $\varepsilon > 0$ and $a_n = a < \infty$ then the corresponding homogeneous one step mapping ϕ has a unique fixed point $\pi_\infty = \phi(\pi_\infty)$ and for any $n \geqslant m \geqslant 0$ we have that*

$$\frac{1}{n} \log \sup_{\mu \in \mathcal{P}(E)} \|\phi_{n/0}(\mu) - \pi_\infty\|_{tv} \leqslant -\left(1 - \frac{m}{n}\right) \alpha(K) \left(\frac{\varepsilon}{a^m}\right)^2.$$

Proof. – Under our assumptions the one step mapping ϕ is clearly continuous on $\mathcal{P}(E)$ (for the weak topology). Since the state space E is assumed to be finite the set $\mathcal{P}(E)$ is a compact and convex subset of $\mathbb{R}^{|E|}$ and Brouwer fixed-point theorem tells us that there exist one fixed point $\pi_\infty = \phi(\pi_\infty) \in \mathcal{P}(E)$. The end of the proof of the corollary is a straightforward application of Theorem 2.4. □

We conclude this section with an asymptotic stability result when the fitness function g_n tends to 1 as n tends to infinity, corresponding to the degenerate situation in filtering theory where the noise can become huge.

THEOREM 2.6. – *Let* $\{g_n;\ n \geqslant 1\}$ *be a collection of measurable and positive functions such that*

$$l(g) \stackrel{\text{def.}}{=} \sum_{n\geqslant 1} \|\log g_n\| < \infty. \tag{15}$$

If the Dobrushin coefficients $\{\alpha(K_n);\ n \geqslant 1\}$ *of the mutations transitions are such that*

$$\sum_{n\geqslant 1} \alpha(K_n) = \infty,$$

then

$$\lim_{n\to\infty} \sup_{\mu,\nu\in\mathcal{P}(E)} \|\phi_{n/0}(\mu) - \phi_{n/0}(\nu)\|_{tv} = 0.$$

In addition, if

$$\lim_{n\to\infty} \frac{1}{n}\sum_{p=1}^{n} \alpha(K_p) \stackrel{\text{def.}}{=} \overline{\alpha}(K),$$

then

$$\limsup_{n\to\infty} \frac{1}{n} \log \sup_{\mu,\nu\in\mathcal{P}(E)} \|\phi_{n/0}(\mu) - \phi_{n/0}(\nu)\|_{tv} \leqslant -\overline{\alpha}(K) \exp -l(g).$$

Moreover when

$$\inf_{n\geqslant 1} \alpha(K_n) \stackrel{\text{def.}}{=} \alpha(K)$$

one concludes that for any $0 \leqslant p \leqslant n$

$$\sup_{\mu,\nu\in\mathcal{P}(E)} \|\phi_{n/p}(\mu) - \phi_{n/p}(\nu)\|_{tv} < \exp -\big(\overline{\alpha}(K) e^{-l(g)} (n-p)\big).$$

Proof. – Under our assumptions we first notice that

$$\forall n \geqslant 0 \quad \|\log g_n\| < \infty,$$

and therefore for any $n \geqslant 0$ and $x \in E$ we clearly have

$$\frac{1}{a_n} \leqslant g_n(x) \leqslant a_n \quad \text{with } a_n = \exp\|\log g_n\|.$$

By definition of the fitness functions $\{g_{n/p};\ 0 \leqslant p \leqslant n\}$ we also have for any $0 \leqslant p \leqslant n$ and $x \in E$

$$e^{-l(g)} \leqslant g_{n/p}(x) \leqslant e^{l(g)},$$

from which one concludes that

$$S_{n/p}(\varphi)(x) = \frac{K_p(g_{n/p}\varphi)(x)}{K_p(g_{n/p})(x)} \geqslant e^{-2l(g)} K_p(\varphi)(x) \tag{16}$$

for any bounded nonnegative function $\varphi : E \to \mathbb{R}_+$. Recalling that (16) implies that

$$\alpha(S_{n/p}) \geqslant e^{-2l(g)} \alpha(K_p)$$

the end of the proof is now a consequence of Lemma 2.1. □

Remark 2.7. – It is noteworthy that when (15) is satisfied then the condition

$$\sum_{n \geqslant 1} \alpha(K_n) = \infty$$

is a sufficient condition for the asymptotic stability of the limiting system (1). This condition is in fact a necessary and sufficient condition for a nonhomogeneous Markov transition to be strongly ergodic (see for instance [19], part II, p. 76).

2.4. Applications

As we said in the introduction the study of the asymptotic behavior of the limiting system (1) is motivated by nonlinear filtering and genetic algorithm theory. The purpose of this section is to indicate some consequences of the previous asymptotic stability properties. In order to illustrate the assumptions of Theorems 2.4 and 2.6 we start with some simple examples.

Example 1. – Suppose $E = \mathbb{R}$ and K_n, $n \geqslant 1$, are given by

$$K_n(x, dz) = \frac{1}{2} \alpha_n \exp\bigl(-\alpha_n |z - b_n(x)|\bigr)\, dz, \quad \alpha_n > 0,\ b_n \in \mathcal{C}_b(\mathbb{R}).$$

Note that K_n may be written

$$K_n(x, dz) = \exp\bigl(\alpha_n (|z| - |z - b_n(x)|)\bigr) \lambda_n(dz)$$

with

$$\lambda_n(dz) = \frac{1}{2} \alpha_n \exp(-\alpha_n |z|)\, dz.$$

It follows that (B) holds with $\log \varepsilon_n = -\alpha_n \|b_n\|$.

Example 2. – In nonlinear filtering settings the system (1) represents the dynamical structure of the conditional distribution of a Markov process given its noisy observations. For instance the unknown Markov process may be a noncooperative target evolving randomly and the filtering problem is concerned with estimating its position at each time.

In some practical situations such as the so-called proportional navigation for manoeuvring targets the number of strategies used by the target may be finite. In this situation $|E| < \infty$ and the action of the process can be modeled with a transition probability kernel of the form

$$K_n(x, z) = \sum_{m=1}^{M} \alpha_n(m, x)\, 1_{F_m(x)}(z),$$

where $\{\alpha_n(m,\boldsymbol{.});\ n \geqslant 1,\ 1 \leqslant m \leqslant M\}$ is a sequence of positive functions satisfying $\sum_{m=1}^M \alpha_n(m,x) = 1$ and $\{F_m;\ 1 \leqslant m \leqslant M\}$ is a sequence of transformations on E.

At each time n and in a new position $x \in E$ the system uses a new strategy $\alpha_n(m,x)$ for choosing the next direction $F_m(x)$. In this case a sufficient condition for the exponential stability of (1) is to assume that, for any $1 \leqslant m \leqslant M$, $n \geqslant 1$ and $x \in E$ the following hold

$$\bigcup_{m=1}^M F_m(x) = E \quad \text{and} \quad \alpha_n(m,x) > 0.$$

Example 3. – In gene analysis each individual represents a chromosome and it is modeled by a binary string of a fixed length L. In this setting the fitness functions represent the performance of the set of genes in a chromosome.

The corresponding genetic model can be defined as in (4) with the finite state space $E = \{0,1\}^L$ with cardinality $|E| = 2^L$. As a parenthesis, we recall that if the state space is finite then $\mathcal{P}(E)$ coincide with the unit $|E|$-simplex Δ with

$$\Delta = \left\{ p \in \mathbb{R}^{|E|};\ p_i \geqslant 0 \text{ and } \sum_{i=1}^{|E|} p_i = 1 \right\}.$$

For time homogeneous fitness and mutations Corollary 2.5 gives sufficient conditions under which the homogeneous one step mapping ϕ has a unique fixed point and it also presents exponential decays.

By way of example, if the mutation transition matrix is such that

$$\forall x, y \in E \quad K(x,y) > 0,$$

then (B) holds for the uniform distribution λ on the finite set E and

$$\varepsilon = \min\left(|E| \min_{x,y} K(x,y),\ \frac{1}{|E| \max_{x,y} K(x,y)} \right).$$

Example 4. – Assume that the fitness functions take the form

$$g_n(x) = \exp -\big(\beta_n U(x)\big)$$

for some bounded nonnegative function $U : E \to \mathbb{R}_+$ and some sequence of parameters $\beta_n \to 0$ as $n \to \infty$. It is easily verified that condition (15) of Theorem 2.6 holds as soon as

$$\sum_{n \geqslant 1} \beta_n < \infty.$$

Next we present an easily verifiable sufficient condition for (C).

(C)′ *For any* $n \geqslant 1$ *there exists an* $a_n \in [1,\infty)$ *such that*

$$\frac{1}{a_n} \leqslant g_n(x) \leqslant a_n, \quad \forall x \in E,\ \forall n \geqslant 1.$$

In addition, the mutation transitions are time homogeneous (*that is* $K_n = K$) *and there exist a subset* $A \in \mathcal{E}$, *a reference probability measure* $\lambda \in \mathcal{P}(E)$ *and a positive number* $\varepsilon \in (0,1)$ *such that*

$$\varepsilon \leqslant \frac{dK(x,\boldsymbol{.})}{d\lambda}(z) \leqslant \frac{1}{\varepsilon}, \quad \forall x \in E,\ \forall z \in A.$$

In addition there exists a decomposition $A^c = B_1 \cup \cdots \cup B_m$, $m \geqslant 1$ *and* $2m$ *reference probability measures* $\lambda_1, \ldots, \lambda_m, \gamma_1, \ldots, \gamma_m \in \mathcal{P}(E)$ *such that for any* $1 \leqslant k \leqslant m$

$$\varepsilon \leqslant \frac{dK(x,\boldsymbol{.})}{d\lambda_k}(z) \leqslant \frac{1}{\varepsilon}, \quad \forall x \in B_k,\ \forall z \in E$$

and

$$\gamma_k(B_k) > 0 \quad \text{and} \quad \frac{dK(x,\boldsymbol{.})}{d\gamma_k}(z) \geqslant \varepsilon, \quad \forall x \in E,\ \forall z \in B_k.$$

Under (C)′ one can check that (C) holds with $m = 2$. To see that this sufficient condition is a reasonable assumption let us present an example of Gaussian transition which can be handled in this framework.

Example 5. – Let K be the Markov transition on $E = \mathbb{R}$ given by

$$K(x,dz) = \frac{1}{\sqrt{2\pi}} \exp -\frac{1}{2}\big(z - f(x)\big)^2\, dz, \tag{17}$$

where the drift function $f : \mathbb{R} \to \mathbb{R}$ is bounded and satisfies

$$f(x) = f(\operatorname{sign}(x) M), \quad \forall |x| \geqslant M.$$

This transition corresponds to the Markov chain determined by

$$X_{n+1} = f(X_n) + W_n,$$

where the W_n are independent and standard normal. In this situation it is not difficult to check that the mixing type conditions in (C)′ hold with

$$A = [-M, M], \quad B_1 = (-\infty, -M), \quad B_2 = (M, +\infty),$$

and

$$\begin{aligned} \lambda_1 &= \delta_{-M} K, & \lambda_2 &= \delta_{-M} K, \\ \gamma_1(dz) &= \frac{1}{\sqrt{2\pi}} \exp -\frac{1}{2}(z - M)^2\, dz, & \gamma_2(dz) &= \frac{1}{\sqrt{2\pi}} \exp -\frac{1}{2}(z + M)^2\, dz. \end{aligned}$$

When (C)′ holds for some $\varepsilon > 0$ and some sequence of positive numbers $\{a_n;\ n \geqslant 1\}$ such that $a \stackrel{\text{def.}}{=} \sup_n a_n < \infty$ then Theorem 2.4 applies to study the stability properties of the limiting system (1).

Much more is true. We can use the full force of (C)′ to prove useful asymptotic stability results even when $a_n \to \infty$.

PROPOSITION 2.8. – *Assume that* (C)′ *holds. Then, for any* $\mu, \nu \in \mathcal{P}(E)$ *we have that*

$$\sum_{n\geqslant 1} a_n^{-2} = \infty \Longrightarrow \lim_{n\to\infty} \|\phi_{n/0}(\mu) - \phi_{n/0}(\nu)\|_{tv} = 0,$$

$$\lim_{n\to\infty} \frac{1}{n}\sum_{p=1}^{n} a_p^{-2} > 0 \Longrightarrow \limsup_{n\to\infty} \frac{1}{n}\log \|\phi_{n/0}(\mu) - \phi_{n/0}(\nu)\|_{t} < 0.$$

Proof. – For any bounded positive function φ we first notice that

$$K(g_{n/p}\varphi) \geqslant \varepsilon\lambda(g_{n/p}\varphi 1_A) + \frac{\varepsilon^2}{a_{p+1}}\sum_{k=1}^{m}\lambda_k(g_{n/p+1})\gamma_k(\varphi 1_{B_k})$$

and

$$K(g_{n/p}) \leqslant \frac{1}{\varepsilon}\lambda(g_{n/p}1_A) + \frac{a_{p+1}}{\varepsilon}\sum_{k=1}^{m}\lambda_k(g_{n/p+1}).$$

This yields

$$S_{n/p}\varphi \geqslant \left(\frac{\varepsilon^3}{a_{p+1}^2}\right)\frac{\lambda(g_{n/p}\varphi 1_A) + \sum_{k=1}^m \lambda_k(g_{n/p+1})\gamma_k(\varphi 1_{B_k})}{\lambda(g_{n/p}1_A) + \sum_{k=1}^m \lambda_k(g_{n/p+1})},$$

which in turns implies that

$$\alpha(S_{n/p}) \geqslant \left(\frac{\varepsilon^3}{a_{p+1}^2}\right)\frac{\lambda(g_{n/p}1_A) + \varepsilon\sum_{k=1}^m \lambda_k(g_{n/p+1})}{\lambda(g_{n/p}1_A) + \sum_{k=1}^m \lambda_k(g_{n/p+1})} \geqslant \frac{\varepsilon^4}{a_{p+1}^2}$$

as soon as ε is chosen so that

$$\inf_{1\leqslant k\leqslant m} \gamma_k(B_k) \geqslant \varepsilon.$$

The end of the proof is straightforward. □

Until the end of this section we investigate more closely the consequences of the asymptotic stability results of Section 2 in the study of the long time behavior of the nonlinear filtering equation (2) when the observation sequence takes the form

$$Y_n = h_n(X_{n-1}) + V_n, \quad n \geqslant 1, \tag{18}$$

for some bounded measurable function $h_n : E \times \mathbb{R}^d \to \mathbb{R}^d$. The sequence $V = \{V_n;\ n \geqslant 1\}$ are $\mathbb{R}^d$-valued, independent of X, and independent random variables with continuous and positive densities $\{\tilde{g}_n;\ n \geqslant 1\}$ with respect to Lebesgue measure. In this situation and using the notations of Section 1 one can check that the functions $\tilde{g}_n, \overline{g}_n$ and h_n are connected by

$$\forall (x,y) \in E \times \mathbb{R}^d \quad \overline{g}_n(y,x) = \tilde{g}_n\big(y - h_n(x)\big).$$

We start with some comments on how the previous results can be used to study the "memory length" of the optimal filter. The so-called optimal filter of fixed memory length T is the measure valued process $\{\pi_n^{Y,T};\ n \geqslant 0\}$ given by

$$\pi_n^{Y,T}(f) = \begin{cases} \pi_n^Y(f) & \text{if } n \leqslant T, \\ E(f(X_n) \mid Y_{n-T+1}, \ldots, Y_n) & \text{otherwise} \end{cases}$$

for any bounded test function f. In other words $\pi_n^{Y,T}$ is the conditional distribution of X_n given the last current observations $\{Y_{n-p};\ p = 0, \ldots, T-1\}$. For practical and theoretical reasons (see for instance [11,14,25]), it is natural to seek conditions which ensures that the optimal filter of memory length T will converge in a sense to be defined to the optimal filter as $T \to \infty$ and uniformly w.r.t. time. The following corollary of Theorem 2.4 answers to this question.

COROLLARY 2.9. – *Assume that the signal transition probability kernels* $\{K_n;\ n \geqslant 1\}$ *satisfy condition* (B) *for a sequence of positive numbers* $\{\varepsilon_n;\ n \geqslant 1\}$ *so that* $\inf_n \varepsilon_n = \varepsilon \in (0,1)$. *Then we have*

$$\sup_{n \geqslant 0} \frac{1}{T} \log \left\| \pi_n^{Y,T} - \pi_n^Y \right\|_{tv} \leqslant -\varepsilon^2.$$

Proof. – Let us fix the observation process Y. To clarify the presentation we also suppress the observation parameter Y so that we simply note ϕ_n and π_n and π_n^T instead of $\phi_n(Y_n, \boldsymbol{.})$ and π_n^Y and $\pi_n^{Y,T}$. Coming back to the definitions of the composite mappings $\{\phi_{n/p};\ 0 \leqslant p \leqslant n\}$ and the distributions $\{\pi_n, \pi_n^T;\ n \geqslant 0\}$ it is easy to see that

$$\pi_n = \phi_{n/n-T}(\pi_{n-T}) \quad \text{and} \quad \pi_n^T = \phi_{n/n-T}\left(\pi_0 K^{n-T}\right), \quad \forall 0 \leqslant T \leqslant n.$$

As a consequence of the results of Section 2 we have that

$$\sup_{n \geqslant 0} \left\| \pi_n^T - \pi_n \right\|_{tv} \leqslant \left(1 - \varepsilon^2\right)^T \leqslant \exp -\left(T\varepsilon^2\right). \qquad \square$$

In our setting the fitness functions are random in the observation parameter. Instead of (C) we will use the following assumption

(C)″ *For any time* $n \geqslant 1$ *there exists a positive function* $a_n : \mathbb{R}^d \to [1, \infty)$ *and a nondecreasing function* $\theta : \mathbb{R} \to \mathbb{R}$ *such that*

$$\frac{1}{a_n(y)} \leqslant \frac{g_n(y - h_n(x))}{g_n(y)} \leqslant a_n(y), \quad \forall x \in E,\ y \in \mathbb{R}^d \tag{19}$$

and

$$|\log a_n(y+u) - \log a_n(y)| \leqslant \theta(\|u\|).$$

In addition, the mutation transitions are homogeneous (*that is* $K_n = K$) *and there exists an* $m \geqslant 1$ *and* $\varepsilon > 0$ *and a reference probability measure* $\lambda \in \mathcal{P}(E)$ *such that*

$$\varepsilon \leqslant \frac{dK^m(x, \boldsymbol{.})}{d\lambda} \leqslant \frac{1}{\varepsilon}, \quad \forall x \in E.$$

As usual and to clarify the presentation we suppress the observation parameter Y so that we simply note ϕ_n and π_n instead of $\phi_n(Y_n, \bullet)$ and π_n^Y. Let us select another initial condition $\mu \in \mathcal{P}(E)$ and denote by $\{\pi_n^\mu;\ n \geqslant 0\}$ the solution of (2) starting at μ (i.e. $\pi_0^\mu = \mu$). Next results are simple corollaries of Theorem 2.4 and Proposition 2.8.

COROLLARY 2.10. – *Assume that* (C)$''$ *holds for some* $m \geqslant 1$. *If* $\alpha(K) > 0$ *and* $\sup_{n\geqslant 1} \|h_n\| < \infty$ *then for any* $\mu \in \mathcal{P}(E)$ *we have*

$$\sum_{p\geqslant 0}\prod_{q=1}^{m} a_{p+q}^{-2}(V_{p+q}) = \infty \Longrightarrow \lim_{n\to\infty} \left\|\pi_n^\mu - \pi_n\right\|_{tv} = 0,$$

$$\lim_{n\to\infty} \frac{1}{n}\sum_{p=0}^{n-1}\prod_{q=1}^{m} a_{p+q}^{-2}(V_{p+q}) > 0 \Longrightarrow \limsup_{n\to\infty} \frac{1}{n}\log\left\|\pi_n^\mu - \pi_n\right\|_{tv} < 0$$

and

$$\sup_{n\geqslant 1} E\big(\log a_n(V_n)\big) < \infty \Longrightarrow \lim_{n\to\infty} E\big(\left\|\pi_n^\mu - \pi_n\right\|_{tv}\big) = 0.$$

Proof. – It is useful here to replace the functions $g_n(y - h_n(\bullet))$ by the "normalized" ones $\frac{g_n(y-h_n(\bullet))}{g_n(y)}$. This choice does not alter the structure of (2).

Let us fix the observation sequence Y. Using the same arguments as in the proof of Theorem 2.4 we first note that for any $0 \leqslant m + p \leqslant n$

$$\varepsilon^2\alpha(K)\prod_{q=1}^{m} a_{p+q}^{-2}(Y_{p+q}) \leqslant \alpha(S_{n/p}) \leqslant \frac{\alpha(K)}{\varepsilon^2}\prod_{q=1}^{m} a_{p+q}^{2}(Y_{p+q}).$$

Since $Y_n = h(X_n) + V_n$, it follows that

$$\left\|\pi_n^\mu - \pi_n\right\|_{tv} \leqslant \prod_{p=0}^{n-m}\left(1 - \varepsilon\alpha(K)\mathrm{e}^{-2m\theta(M)}\prod_{q=1}^{m} a_{p+q}^{-2}(V_{p+q})\right)$$

for any $\mu \in \mathcal{P}(E)$ and $n \geqslant m \geqslant 1$ as soon as $\sup_{n\geqslant 1}\|h_n\| \leqslant M$. The end of proof of the first two implications is now straightforward. Let us prove the third and last one. Writing

$$C = \varepsilon\alpha(K)\mathrm{e}^{-2m\theta(M)}, \qquad \varphi^{(p)}(V) = \prod_{q=1}^{m} a_{p+q}^{-2}(V_{p+q})$$

we first notice that for any $t > 0$

$$E\big(\big(1 - C\varphi^{(p)}(V)\big)^{n-m+1}\big) \leqslant \sum_{q=1}^{m} P\left(\log a_{p+q}(V_{p+q}) \geqslant \frac{t}{2m}\right) + \big(1 - C\mathrm{e}^{-t}\big)^{n-m+1}.$$

Hence

$$E\big(\big(1 - C\varphi^{(p)}(V)\big)^{n-m+1}\big) \leqslant \frac{2m^2}{t}\sup_{n\geqslant 1} E\big(\log a_n(V_n)\big) + \big(1 - C\mathrm{e}^{-t}\big)^{n-m+1}. \qquad (20)$$

The final step is to note that repeated use of Holder inequality gives

$$E\left(\left\|\pi_n^\mu - \pi_n\right\|_{tv}\right) \leqslant \prod_{p=0}^{n-m} E\left(\left(1 - C\varphi^{(p)}(V)\right)^{n-m+1}\right)^{\frac{1}{n-m+1}}. \tag{21}$$

Combining (20) and (21) one concludes that

$$\lim_{n\to\infty} E\left(\left\|\pi_n^\mu - \pi_n\right\|_{tv}\right) \leqslant \frac{2m^2}{t} \sup_{n\geqslant 1} E\left(\log a_n(V_n)\right) \quad \forall t > 0.$$

Letting $t \to \infty$ we end the proof of the corollary. □

COROLLARY 2.11. – *Assume that* (C)″ *holds. If* $\sup_{n\geqslant 1} \|h_n\| < \infty$ *then we have*

$$\sum_{n\geqslant 1} E\left(a_n^{-2}(V_n)\right) = \infty \Longrightarrow \lim_{n\to\infty} E\left(\|\pi_n^\mu - \pi_n\|_{tv}\right) = 0,$$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{p=1}^{n} E\left(a_p^{-2}(V_p)\right) > 0 \Longrightarrow \limsup_{n\to\infty} \frac{1}{n} \log E\left(\|\pi_n^\mu - \pi_n\|_{tv}\right) < 0.$$

Proof. – The basic ideas of the proof were already given in the proof of Proposition 2.8. Under our assumptions one can check that

$$\alpha(S_{n/p}) \geqslant \frac{\varepsilon^4}{a_{p+1}^2(Y_{p+1})} \geqslant \frac{\varepsilon^4}{a_{p+1}^2(V_{p+1})} \mathrm{e}^{-2\theta(M)}$$

as soon as $\sup_{n\geqslant 1} \|h_n\| \leqslant M$. Thus we have

$$E\left(\|\pi_n^\mu - \pi_n\|_{tv}\right) \leqslant \prod_{p=1}^{n} \left(1 - \varepsilon^4 \mathrm{e}^{-2\theta(M)} E\left(a_{p+1}^{-2}(V_{p+1})\right)\right)$$

from which the end of the proof is straightforward. □

Let us now investigate assumption (C)″ through some examples of nonlinear sensor.

Example 6. – As a typical example of nonlinear filtering problem assume the functions $h_n : E \to \mathbb{R}^d$, $n \geqslant 1$, are bounded continuous and the densities g_n given by

$$\tilde{g}_n(v) = \frac{1}{((2\pi)^d |R_n|)^{1/2}} \exp\left(-\frac{1}{2} v' R_n^{-1} v\right),$$

where R_n is a $d \times d$ symmetric positive matrix. This corresponds to the situation where the observations are given by

$$Y_n = h_n(X_{n-1}) + V_n, \quad \forall n \geqslant 1, \tag{22}$$

where $(V_n)_{n\geqslant 1}$ is a sequence of $\mathbb{R}^d$-valued and independent random variables with Gaussian densities.

After some easy manipulations one gets the bounds (19) with

$$\log a_n(y) = \frac{1}{2}\|R_n^{-1}\|\|h_n\|^2 + \|R_n^{-1}\|\|h_n\||y|,$$

where $\|R_n^{-1}\|$ is the spectral radius of R_n^{-1}. In addition we have

$$|\log a_n(y+u) - \log a_n(y)| \leqslant L_n|u| \quad \text{with } L_n = \|R_n^{-1}\|\|h_n\|.$$

It is therefore not difficult to check that the assumptions of Theorem 2.11 are satisfied when

$$\sup_{n\geqslant 1}\big(\|h_n\|, \|R_n^{-1}\|\big) < \infty.$$

To see this claim it suffices to note that Jensen's inequality yields that

$$\log E\big(a_n^{-2}(V_n)\big) \geqslant -\|R_n^{-1}\|\|h_n\|^2 - 2\|R_n^{-1}\|\|h_n\|E(|V_n|).$$

Finally we note that the last assertion of Theorem 2.10 holds since we have

$$E\big(\log a_n(V_n)\big) = \frac{1}{2}\|R_n^{-1}\|\|h_n\|^2 + \|R_n^{-1}\|\|h_n\|E(|V_n|).$$

Example 7. – Our result is not restricted to Gaussian noise sources. For instance, let us assume that $d = 1$ and g_n is a bilateral exponential density

$$\tilde{g}_n(v) = \frac{\alpha_n}{2}\exp -(\alpha_n|v|), \quad \alpha_n > 0.$$

In this case one gets the bounds (19) with

$$\log a_n(y) = \alpha_n\|h_n\|$$

which is independent of the observation parameter y. One concludes easily that the conditions of Theorems 2.10 and 2.11 are satisfied as soon as

$$\sup_{n\geqslant 1}\{\alpha_n, \|h_n\|\} < \infty.$$

Finally, if

$$\sum_{n\geqslant 1}(\alpha_n\|h_n\|) < \infty$$

one can also check that condition (15) of Theorem 2.6 is satisfied.

3. Uniform convergence of genetic algorithm

3.1. A uniform convergence theorem

In the present section the asymptotic stability results presented in Section 2 are applied to prove uniform convergence results for the finite *IPS* model.

THEOREM 3.1. – *Let* $\{g_n;\ n \geqslant 1\}$ *be a collection of bounded and positive functions on* E *such that for any* $n \geqslant 1$ *there exists an* $a_n \in [1, \infty)$ *such that for any* $x \in E$ *and* $n \geqslant 1$

$$\frac{1}{a_n} \leqslant g_n(x) \leqslant a_n. \tag{23}$$

If the sequence $\{a_n;\ n \geqslant 1\}$ *is uniformly bounded, so that*

$$a \stackrel{\text{def.}}{=} \sup_{n \geqslant 1} a_n < \infty$$

and the limiting system (1) *satisfies the following asymptotic stability assumption*

$$\forall f \in \mathcal{B}_b(E) \quad \lim_{T \to \infty} \sup_{\mu, \nu \in \mathcal{P}(E)} \sup_{p \geqslant 0} \left| \phi_{p+T/p}(\mu)(f) - \phi_{p+T/p}(\nu)(f) \right| = 0, \tag{24}$$

then we have the following uniform convergence result with respect to the time parameter

$$\forall f \in \mathcal{B}_b(E) \quad \lim_{N \to \infty} \sup_{n \geqslant 0} E\left(|\pi_n^N f - \pi_n f|\right) = 0. \tag{25}$$

In addition, if the limiting system is exponentially asymptotically stable in the sense that there exists some $T_0 \geqslant 1$ *and* $\gamma > 0$ *such that for any* $f \in \mathcal{B}_b(E)$, $\|f\| \leqslant 1$, $\mu, \nu \in \mathcal{P}(E)$ *and* $T \geqslant T_0$

$$\sup_{p \geqslant 0} \left| \phi_{p+T/p}(\mu)(f) - \phi_{p+T/p}(\nu)(f) \right| \leqslant \mathrm{e}^{-\gamma . T}, \tag{26}$$

then we have for any $f \in \mathcal{B}_b(E)$, $\|f\| \leqslant 1$, *the following uniform bounds*

$$\sup_{n \geqslant 0} E\left(\left|\pi_n^N f - \pi_n f\right|\right) \leqslant \frac{5 \exp(2\gamma')}{N^{\alpha/2}} \tag{27}$$

for any $N \geqslant 1$ *so that*

$$T(N) \stackrel{\text{def.}}{=} \left[\frac{1}{2} \frac{\log N}{\gamma + \gamma'}\right] + 1 \geqslant T_0,$$

where α *and* γ' *are given by*

$$\alpha = \frac{\gamma}{\gamma + \gamma'} \quad \text{with } \gamma' = 1 + 2 \log a.$$

Proof. – In what follows we denote by $f : E \to \mathbb{R}$ a bounded measurable function such that $\|f\| \leqslant 1$. To prove our result we will use repeatedly formula (9). For instance and for later use we immediately notice that

$$\frac{1}{a_{n/p}} \leqslant g_{n/p}(x) \leqslant a_{n/p}, \quad \forall x \in E,\ \forall 0 \leqslant p \leqslant n,$$

with

$$a_{n/p}=\prod_{q=p+1}^{n}a_q$$

and the usual convention $\prod_{\emptyset}=1$. On the other hand, using the above simplified notations we have the decomposition

$$\pi_n^N f-\pi_n f=\sum_{p=0}^{n}\bigl(\phi_{n/p}(\pi_p^N)f-\phi_{n/p}(\phi_p(\pi_{p-1}^N))f\bigr)$$

with the convention $\phi_0(\pi_{-1}^N)=\pi_0$. Therefore we also have the inequality

$$\bigl|\pi_n^N f-\pi_n f\bigr|\leqslant\sum_{p=0}^{n}\bigl|\phi_{n/p}(\pi_p^N)f-\phi_{n/p}(\phi_p(\pi_{p-1}^N))f\bigr|. \tag{28}$$

Using (9) we see that each term

$$\bigl|\phi_{n/p}(\pi_p^N)f-\phi_{n/p}(\phi_p(\pi_{p-1}^N))f\bigr|$$

is bounded by

$$a_{n/p}^2\bigl(\bigl|\pi_p^N f_1-\phi_p(\pi_{p-1}^N)f_1\bigr|+\bigl|\pi_p^N f_2-\phi_p(\pi_{p-1}^N)f_2\bigr|\bigr) \tag{29}$$

with

$$f_1=\frac{g_{n/p}}{a_{n/p}}K_{n/p}(f),\qquad f_2=\frac{g_{n/p}}{a_{n/p}}$$

so that $\|f_1\|,\|f_2\|\leqslant 1$.

By recalling that π_p^N is the empirical measure associated to N conditionally independent random variables with common law $\phi_p(\pi_{p-1}^N)$ we clearly have the estimate

$$E\bigl(\bigl|\pi_p^N f-\phi_p(\pi_{p-1}^N)f\bigr|\bigr)\leqslant\frac{1}{\sqrt{N}}.$$

Collecting the above inequalities one concludes that

$$E\bigl(\bigl|\pi_n^N f-\pi_n f\bigr|\bigr)\leqslant\frac{2}{\sqrt{N}}\bigl(1+n\,a_{n/0}^2\bigr).$$

This yields for any $T\geqslant 0$

$$\sup_{n=0,\dots,T}E\bigl(\bigl|\pi_n^N f-\pi_n f\bigr|\bigr)\leqslant\frac{4T\,a_{T/0}^2}{\sqrt{N}}.$$

Under our assumptions this implies that

$$\sup_{n=0,\dots,T}E\bigl(\bigl|\pi_n^N f-\pi_n f\bigr|\bigr)\leqslant\frac{4T\,a^{2T}}{\sqrt{N}}$$

and

$$\sup_{n=0,\dots,T} E\left(\left|\pi_n^N f - \pi_n f\right|\right) \leqslant \frac{4e^{\gamma' T}}{\sqrt{N}} \quad \text{with } \gamma' = 1 + 2\log a. \tag{30}$$

For any $n \geqslant 0$ we also have the decomposition

$$\pi_n^N f - \pi_n f = \sum_{p=n-T+1}^{n} \left(\phi_{n/p}\left(\pi_p^N\right) f - \phi_{n/p}\left(\phi_p\left(\pi_{p-1}^N\right)\right) f\right) + \left(\phi_{n/n-T}\left(\pi_{n-T}^N\right) f - \phi_{n/n-T}(\pi_{n-T}) f\right).$$

Under our assumptions this implies that

$$\left|\pi_n^N f - \pi_n f\right| \leqslant \sum_{p=n-T+1}^{n} \left|\phi_{n/p}\left(\pi_p^N\right) f - \phi_{n/p}\left(\phi_p\left(\pi_{p-1}^N\right)\right) f\right| + \varepsilon_T(f),$$

where

$$\varepsilon_T(f) \stackrel{\text{def.}}{=} \sup_{\mu,\nu \in \mathcal{P}(E)} \sup_{p \geqslant 0} \left|\phi_{p+T/p}(\mu)(f) - \phi_{p+T/p}(\nu)(f)\right|.$$

In the same way that we deduce (30) from (28) we can establish that for any $n \geqslant T$

$$E\left(\left|\pi_n^N f - \pi_n f\right|\right) \leqslant \frac{4e^{\gamma' T}}{\sqrt{N}} + \varepsilon_T(f). \tag{31}$$

If we combine (30) with (31) we arrive at

$$\sup_{n \geqslant 0} E\left(\left|\pi_n^N f - \pi_n f\right|\right) \leqslant \frac{4e^{\gamma' T}}{\sqrt{N}} + \varepsilon_T(f)$$

for any $T \geqslant T_0$. Letting $N \to \infty$ and then $T \to \infty$ we prove (25).

If the exponential bound (26) holds then using the same line of arguments as before one can check that for any $T \geqslant T_0$

$$\sup_{n \geqslant 0} E\left(\left|\pi_n^N f - \pi_n f\right|\right) \leqslant \frac{4e^{\gamma' T}}{\sqrt{N}} + e^{-\gamma \cdot T}.$$

Now, if we put

$$T = T(N) \stackrel{\text{def.}}{=} \left[\frac{1}{2} \frac{\log N}{\gamma + \gamma'}\right] + 1,$$

where $[a]$ denotes the integer part of $a \in \mathbb{R}$, we find that

$$\sup_{n \geqslant 0} E\left(\left|\pi_n^N f - \pi_n f\right|\right) \leqslant \frac{5\exp(2\gamma')}{N^{\alpha/2}} \quad \text{with } \alpha = \frac{\gamma}{\gamma + \gamma'}$$

as soon as $T(N) \geqslant T_0$. $\square$

The preceding theorem shows that under some mild assumptions on the signal semigroup the confidence intervals

$$\left[\pi_n^N(A) - \frac{\lambda b}{N^{\alpha/2}},\ \pi_n^N(A) + \frac{\lambda b}{N^{\alpha/2}}\right], \quad A \in \mathcal{E},$$

where $b = 5\mathrm{e}^{2\gamma'}$, have reliability $1 - \frac{1}{\lambda}$ for any time $n \geqslant 0$. Namely

$$P\left(\left|\pi_n^N(A) - \pi_n(A)\right| \leqslant \frac{\lambda b}{N^{\alpha/2}}\right) \geqslant 1 - \frac{1}{\lambda}, \quad \forall A \in \mathcal{E},\ \forall n \geqslant 0.$$

For instance we can say at each time $n \geqslant 0$, with a probability greater than 0, 9 that the exact values of $\pi_n(A)$, $A \in \mathcal{E}$, are in the intervals

$$\left[\pi_n^N(A) - \frac{10b}{N^{\alpha/2}},\ \pi_n^N(A) + \frac{10b}{N^{\alpha/2}}\right], \quad A \in \mathcal{E}.$$

This statement is usually written in the symbolic form

$$\forall n \geqslant 0,\ \forall A \in \mathcal{E}, \quad \pi_n^N(A) \simeq \pi_n(A) \pm \frac{10b}{N^{\alpha/2}} \quad (\geqslant 0, 9).$$

Condition (26) guarantees that the measure valued process (1) is asymptotically stable and corrects with a exponential rate any erroneous initial condition. Sufficient conditions for (26) to hold are given in Section 2. For instance, if (B) holds for some $\varepsilon \in (0, 1)$ then we have

$$\sup_{p \geqslant 0} \left|\phi_{p+T/p}(\mu) f - \phi_{p+T/p}(\nu) f\right| \leqslant \left(1 - \varepsilon^2\right)^T$$

for any $T \geqslant 1$, $\mu, \nu \in \mathcal{P}(E)$ and for any bounded test function f so that $\|f\| \leqslant 1$.

Let us discuss some consequences of Theorem 3.1 when the state space E is a Polish space (that is a complete separable metric space). In this situation we first recall that $\mathcal{P}(E)$ with the topology of weak convergence can be considered as a metric space with metric d defined for $\mu, \nu \in \mathcal{P}(E)$ by

$$d(\mu, \nu) = \sum_{m \geqslant 1} 2^{-(m+1)} |\mu f_m - \nu f_m|,$$

where $\{f_m;\ m \geqslant 1\}$ is a suitable sequence of uniformly continuous functions such that $\|f_m\| \leqslant 1$ for any $m \geqslant 1$ (see for instance Theorem 6.6 p. 47 in [27]). Now the Kantorovitch–Rubinstein or Vaserstein metric on $\mathcal{P}(\mathcal{P}(E))$ and associated to the metric d is defined by

$$D(\Pi_1, \Pi_2) = \inf E\big(d(\mu, \nu)\big), \tag{32}$$

where the infimum is taken over all pair of random variables (μ, ν) with values in $\mathcal{P}(E)$ and such that μ has distribution Π_1 and ν has distribution Π_2. The metric d being a bounded function, formula (32) defines a complete metric on $\mathcal{P}(\mathcal{P}(E))$ which gives to $\mathcal{P}(\mathcal{P}(E))$ the topology of weak convergence (see Theorem 2 in [20]).

If we note Π_n^N the law of the random measures π_n^N and Π_n the Dirac distribution at the point π_n Theorem 3.1 leads to

$$\sup_{n\geqslant 0} D\left(\Pi_n^N, \Pi_n\right) \leqslant \frac{\mathrm{Cte}}{N^{\alpha/2}}.$$

In nonlinear filtering settings the fitness functions depend on the observation delivered by the sensors and the previous theorems cannot be applied directly. Let us come back to the nonlinear filtering problem described in Section 2.4 with the observation sequence given by (18). To clarify the presentation we will also suppress the observation parameter Y and we simply note ϕ_n and π_n instead of $\phi_n(Y_n, \bullet)$ and π_n^Y. As in Section 2.4, it is convenient to replace the boundedness assumptions (23) by the following condition

(G) *For any time $n \geqslant 1$ there exists a positive function $a_n : \mathbb{R}^d \to [1, \infty)$ and a nondecreasing function $\theta : \mathbb{R} \to \mathbb{R}$ such that for any $(x, y) \in E \times \mathbb{R}^d$*

$$\frac{1}{a_n(y)} \leqslant \frac{g_n(y - h_n(x))}{g_n(y)} \leqslant a_n(y), \tag{33}$$

$$|\log a_n(y+u) - \log a_n(y)| \leqslant \theta(\|u\|)$$

and

$$\sup_{n\geqslant 1} \log E\left(a_n^2(V_n)\right)^{1/2} \stackrel{\text{def.}}{=} L < \infty \quad \text{and} \quad \sup_{n\geqslant 1} \|h_n\| \stackrel{\text{def.}}{=} M < \infty.$$

Under the boundedness condition (G), if we replace the limiting condition (24) by

$$\lim_{T\to\infty} \sup_{\mu,\nu\in\mathcal{P}(E)} \sup_{p\geqslant 0} E\left(\left|\phi_{p+T/p}(\mu)(f) - \phi_{p+T/p}(\nu)(f)\right| \mid Y_1, \ldots, Y_p\right) = 0 \tag{34}$$

then, using the same line of arguments as in the proof of Theorem 3.1, one can check that (25) holds. In much the same way, if we replace in Theorem 3.1 the exponential bound (26) by the following inequality

$$\sup_{p\geqslant 0} E\left(\left|\phi_{p+T/p}(\mu)(f) - \phi_{p+T/p}(\nu)(f)\right| \mid Y_1, \ldots, Y_p\right) \leqslant \mathrm{e}^{-\gamma.T}, \tag{35}$$

then one can check that the uniform convergence result (27) holds with

$$\log a = L + \theta(M).$$

Example 8. – It can be directly checked that the Gaussian and bi-exponential examples of noise sources given in Examples 6 and 7 satisfy condition (G).

3.2. A genetic algorithm with periodic selections

Our present purpose is to understand why the selection mechanism plays a very special role in the behavior of the particle filter. What is important is that each particle interacts selectively with the system in accordance with the environment represented by the fitness functions.

This remark underlines the very interesting role played by the updating/selection transition. In nonlinear filtering settings each fitness function is related to the current observation data and intuitively the selection mechanism stabilizes the particles' motion around certain values of the real signal which are determined by the noisy observation and thus provides a well behaved adaptative stochastic grid.

We also remark that the updating transition is used at each time and only depends on the current fitness. Another idea is to use the selection mechanism from time to time. In this case the interaction depends on the series of fitness functions and on the path particles between two selection dates. The choice of the selection/updating times then requires a criterion for optimality. It will now be shown that one can take advantage of the stability properties of the limiting system (1) to develop a more efficient genetic algorithm.

In the last part of this paper the genetic type scheme presented in Section 2 are generalized. The prediction/mutation mechanisms of the former will include exploration paths of a given length $T \geqslant 1$ and the corresponding updating/selection procedure will be used every T steps and it will consider T fitness functions.

This new algorithm with periodic selection is particularly important in nonlinear filtering settings since in this situation each selection transition depends on T observation values and the resulting genetic algorithm appears to be more efficient in practice.

Our immediate goal is to show that the former genetic algorithm can be reduced to the latter through a suitable state space basis. To this end we need to introduce some additional notations. To any $p \in \{1, \ldots, T\}$ and $T \geqslant 1$ we associate a sequence of meshes $\{t_n^{(T,p)};\ n \geqslant 0\}$ by setting

$$t_0^{(T,p)} = 0, \quad t_n^{(T,p)} = (n-1)T + p, \quad \forall n \geqslant 1.$$

The parameter T will be the selection/updating period, n will denote the time steps and the parameter p will only be used to cover all the time space basis so that

$$\bigcup_{1 \leqslant p \leqslant T} \{t_n^{(T,p)};\ n \geqslant 0\} = \mathbb{N}.$$

The construction below will depend on the pair parameter (T, p) and on the observations Y. To clarify the presentation the mutation transition is assumed to be time-homogeneous, that is $K_n = K$ and we simplify the notations suppressing the pair parameter (T, p) so that we simply note t_n instead of $t_n^{(T,p)}$. If, we write for any $n \geqslant 0$

$$\Delta_n = t_n - t_{n-1}, \quad n \geqslant 1,$$

then we clearly have that

$$\Delta_1 = p, \quad \text{and} \quad \Delta_n = T \quad \forall n > 1.$$

We also notice that the distributions given by

$$\eta_n = \pi_{t_n} \otimes \underbrace{K \otimes \cdots \otimes K}_{\Delta_{n+1}-1} \in \mathcal{P}\big(E^{\Delta_{n+1}}\big), \quad n \geqslant 0, \tag{36}$$

are solution of the measure valued process

$$\eta_n = \Phi_n(\eta_{n-1}), \quad n \geqslant 1, \tag{37}$$

where $\Phi_n : \mathcal{P}(E^{\Delta_n}) \to \mathcal{P}(E^{\Delta_{n+1}})$ is the continuous function given by

$$\Phi_n(\eta) = \Psi_n(\eta)\mathcal{K}_n$$

and

- $\Psi_n : \mathcal{P}(E^{\Delta_n}) \to \mathcal{P}(E^{\Delta_n})$ is the continuous function defined by

$$\forall f \in \mathcal{C}_b\big(E^{\Delta_n}\big) \quad \Psi_n(\eta)(f) = \frac{\eta(\mathcal{G}_n\, f)}{\eta(\mathcal{G}_n)}$$

  with

$$\mathcal{G}_n(x) = \prod_{q=1}^{\Delta_n} g_{t_n+q}(x_q).$$

- $\mathcal{K}_n$ is a Markov transition probability kernel from E^{Δ_n} to $E^{\Delta_{n+1}}$ given by

$$\mathcal{K}_n\big((x_1,\dots,x_{\Delta_n}), d(z_1,\dots,z_{\Delta_{n+1}})\big) = K(x_{\Delta_n}, dz_1) \times \cdots \times K(z_{\Delta_{n+1}-1}, dz_{\Delta_{n+1}}).$$

Remark 3.2. – To check that this model generalizes the one given in Section 2 observe that it coincide with the previous one when $T = 1$ and $p = 1$.

It is interesting to note that in nonlinear filtering settings each fitness function g_n is related to the current observation data, that is

$$\forall x \in E \quad g_n(x) = \overline{g}_n(Y_n, x)$$

and the distributions (36) represent the conditional distribution of $\mathcal{X}_n$ given the random variables $\mathcal{Y}_1,\dots,\mathcal{Y}_n$ where

$$\mathcal{X}_n = (X_{t_n},\dots,X_{t_{n+1}-1}) \quad \text{and} \quad \mathcal{Y}_{n+1} = (Y_{t_n+1},\dots,Y_{t_{n+1}}), \quad n \geqslant 0.$$

Now, the genetic type algorithm associated to (37) is a Markov chain $\{\zeta_n;\ n \geqslant 1\}$ with product state spaces $\{(E^{\Delta_{n+1}})^N;\ n \geqslant 0\}$ where N is the number of particles and $\{\Delta_{n+1};\ n \geqslant 0\}$ the selection/updating periods.

The initial particle system $\zeta_0 = (\zeta_0^1,\dots,\zeta_0^N)$ takes values in $(E^{\Delta_1})^N = (E^p)^N$ and it is given by

$$P_Y(\zeta_0 \in dx) = \prod_{q=1}^N \eta_0\big(dx^q\big) = \prod_{q=1}^N \pi_0\big(dx_1^q\big) K\big(x_1^q, dx_2^q\big) \dots K\big(x_{\Delta_1-1}^q, dx_{\Delta_1}^q\big)$$

and the transition of the chain is given by

$$P_Y(\zeta_n \in dx \mid \zeta_{n-1} = z) = \prod_{q=1}^N \Phi_n\left(\frac{1}{N}\sum_{i=1}^N \delta_{z^i}\right)\big(dx^q\big)$$

$$= \prod_{q=1}^{N} \sum_{i=1}^{N} \frac{\mathcal{G}_n(z^i)}{\sum_{j=1}^{N} \mathcal{G}_n(z^j)} K\left(z^i_{\Delta_n}, dx^q_1\right) \cdots K\left(x^q_{T-1}, dx^q_T\right),$$

where $dx = dx^1 \times \cdots \times dx^N$ is an infinitesimal neighborhood of the point $x = (x^1, \ldots, x^N) \in (E^{\Delta_{n+1}})^N$ and for any $1 \leqslant i \leqslant N$, $z^i = (z^i_1, \ldots, z^N_{\Delta_n}) \in E^{\Delta_n}$.

If we denote

$$\zeta_n = (\xi_{t_n}, \ldots, \xi_{t_{n+1}-1}), \quad \forall n \geqslant 0$$

we see that the former algorithm is a genetic type algorithm with T-periodic selection/updating transitions:

Between the dates t_n and t_{n+1} the particles evolve randomly according to the mutation transition and the selection mechanism takes place at each time t_n, $n \geqslant 1$.

The approximation of the desired conditional distributions $\{\pi_{t_n};\ n \geqslant 0\}$ by the particle density profiles

$$\pi^{N,T}_{t_n} = \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi^i_{t_n}}$$

is guaranteed by the following theorem

THEOREM 3.3. – *Let $\{g_n;\ n \geqslant 1\}$ be a collection of bounded and positive functions on E such that for any $n \geqslant 1$ there exists an $a_n \in [1, \infty)$ such that for any $x \in E$ and $n \geqslant 1$*

$$\frac{1}{a_n} \leqslant g_n(x) \leqslant a_n.$$

If the sequence $\{a_n;\ n \geqslant 1\}$ is uniformly bounded, so that

$$a \stackrel{\text{def.}}{=} \sup_{n \geqslant 1} a_n < \infty$$

and the limiting system (1) *is exponentially asymptotically stable in the sense that there exists some $T_0 \geqslant 1$ and $\gamma > 0$ such that for any $f \in \mathcal{B}_b(E)$, $\|f\| \leqslant 1$, $\mu, \nu \in \mathcal{P}(E)$ and $T \geqslant T_0$*

$$\sup_{p \geqslant 0} \left|\phi_{p+T/p}(\mu)(f) - \phi_{p+T/p}(\nu)(f)\right| \leqslant \mathrm{e}^{-\gamma . T},$$

then we have for any $f \in \mathcal{B}_b(E)$, $\|f\| \leqslant 1$, the following uniform convergence rates

$$\sup_{n \geqslant 0} E\left(\left|\pi^{N,T(N)}_{t_n} f - \pi_{t_n} f\right|\right) \leqslant \frac{5 \exp(2\gamma')}{N^{\beta/2}} \tag{38}$$

for any $N \geqslant 1$ so that

$$T(N) \stackrel{\text{def.}}{=} \left[\frac{1}{2} \frac{\log N}{\gamma + \gamma''}\right] + 1 \geqslant T_0,$$

where β and γ'' are given by

$$\beta = \frac{\gamma}{\gamma + \gamma''} \quad \textit{with } \gamma'' = 2\log a.$$

Proof. – The proof will only be sketched since we will follow essentially the same line of proof of Theorem 3.1.

First we note that the definition of

$$\eta_{n-1}^N = \frac{1}{N}\sum_{i=1}^N \delta_{\zeta_{n-1}^i}, \qquad \zeta_{n-1} = (\xi_{t_{n-1}}, \ldots, \xi_{t_n-1}) \in \left(E^N\right)^{\Delta_n}$$

and the weak law of large numbers yield

$$E\left(\left|\eta_{n-1}^N(\varphi) - \overline{\eta}_{n-1}^N(\varphi)\right|\right) \leqslant \frac{\|\varphi\|}{\sqrt{N}}, \quad \forall \varphi \in \mathcal{C}_b\left(E^{\Delta_n}\right),$$

where $\overline{\eta}_n^N$ is the random measure

$$\overline{\eta}_{n-1}^N = \frac{1}{N}\sum_{i=1}^N \delta_{\xi_{t_{n-1}}^i} \otimes \underbrace{K \otimes \cdots \otimes K}_{\Delta_n - 1} \in \mathbf{M}_1\left(E^{\Delta_n}\right).$$

Similar to (29) and (30) we obtain for any $\varphi \in \mathcal{C}_b(E^{\Delta_{n+1}})$

$$\begin{aligned} E\left(\left|\Phi_n\left(\eta_{n-1}^N\right)(\varphi) - \Phi_n\left(\overline{\eta}_{n-1}^N\right)(\varphi)\right|\right) &= E\left(\left|\Psi_n\left(\eta_{n-1}^N\right)(\mathcal{K}_n\varphi) - \Psi_n\left(\overline{\eta}_{n-1}^N\right)(\mathcal{K}_n\varphi)\right|\right) \\ &\leqslant \frac{2\|\varphi\|}{\sqrt{N}} \exp\left(2(\log a)\Delta_n\right). \end{aligned} \tag{39}$$

On the other hand we have the decomposition

$$\pi_{t_n}^N(f) - \pi_{t_n}(f) = \eta_n^N(f_1) - \eta_n(f_1) = I_N^1 + I_N^2 + I_N^3,$$

where

$$f_1 = f \otimes \underbrace{1 \otimes \cdots \otimes 1}_{\Delta_{n+1} - 1}$$

and

$$\begin{aligned} I_N^1 &= \eta_n^N(f_1) - \Phi_n\left(\eta_{n-1}^N\right)(f_1), \\ I_N^2 &= \Phi_n\left(\eta_{n-1}^N\right)(f_1) - \Phi_n\left(\overline{\eta}_{n-1}^N\right)(f_1), \\ I_N^3 &= \Phi_n\left(\overline{\eta}_{n-1}^N\right)(f_1) - \Phi_n(\eta_{n-1})(f_1). \end{aligned}$$

In order to derive a bound for I_N^3 we simply note that

$$\Phi_n(\pi \otimes \underbrace{K \otimes \cdots \otimes K}_{\Delta_n})(f_1) = \phi_{t_{n+1}/t_n}(\pi)(f).$$

Indeed, under our assumptions, this yields

$$E\left(\left|I_N^3\right|\right) = E\left(\left|\phi_{t_n/t_{n-1}}\left(\pi_{t_{n-1}}^N\right)(f) - \phi_{t_n/t_{n-1}}\left(\pi_{t_{n-1}}\right)(f)\right|\right) \leqslant \mathrm{e}^{-\gamma\Delta_n}. \tag{40}$$

Finally, by definition of η_n^N we have that

$$E\left(\left|I_N^1\right|\right) \leqslant \frac{1}{\sqrt{N}}. \tag{41}$$

Combining (39), (40) and (41) one can check that

$$\sup_{n\geqslant 0} E\left(\left|\pi_n^{N,T(N)} f - \pi_n f\right|\right) \leqslant \frac{1+2\mathrm{e}^{\gamma' T}}{N^{1/2}} + \mathrm{e}^{-\gamma T} \quad \text{with } \gamma' = 2\log a$$

from which the end of the proof is straightforward. □

Remark 3.4. – We note that the error bound (38) is an improvement of (27). More precisely, using the notations of Theorem 3.3 we have that

$$\frac{\beta}{\alpha} = 1 + \frac{1}{\gamma + 2\log a} > 1.$$

In view of the preceding construction the genetic type algorithm with periodic selection $T(N)$ depends on a parameter $p = 1, \ldots, T(N)$ so that we need $T(N)$ genetic algorithms to describe the conditional distributions $\{\pi_n;\ n \geqslant 0\}$.

In other words we need $NT(N)$ particles to approximate the whole solution of the limiting system (1) with an error bound (38). It is therefore natural to ask how rapid the approach is in (27) when we use $NT(N)$ particles. In this last situation it is clear that the convergence rate is proportional to

$$D_\alpha(N) = \frac{1}{(NT(N))^{\alpha/2}}.$$

If we write $D_\beta(N) = N^{-\beta/2}$ the decay rate of the genetic scheme with periodic selection $T(N)$ we find that

$$\frac{D_\alpha(N)}{D_\beta(N)} = \left(\frac{N^{\frac{1}{\gamma+2\log a}}}{T(N)}\right)^{\alpha/2} \xrightarrow[N\to\infty]{} \infty.$$

Remark 3.5. – Theorem 3.3 also applies to the nonlinear filtering problem described in Section 2.4 with the observation sequence given by (18).

Arguing as before, under (G) and (35) the uniform convergence result (38) holds with

$$\log a = L + \theta(M).$$

3.3. Comparison of genetic-type schemes

In this work we have presented a way to combine the stability properties of the limiting system (1) with the long time behavior of a class of genetic algorithms.

It remains to discuss the extensions and limitations of genetic-type approximating schemes. Several variants of the particle scheme studied in this paper have been recently suggested to approximate the nonlinear filtering equation (see for instance [6–8] and references therein).

These variants are less "time consuming" mainly because they use independent branching corrections but as a result the size of the system is no longer fixed but random.

To be more precise, let us briefly recall these constructions. As before the natural and classical idea is to approximate the two steps transition of the limiting dynamical system (1)

$$\pi_n \xrightarrow{\text{Updating}} \widehat{\pi}_n \stackrel{\text{def.}}{=} \Psi_n(\pi_n) \xrightarrow{\text{Prediction}} \pi_{n+1} = \widehat{\pi}_n K_n$$

by a two steps Markov chain taking values in the set of discrete and finite measures. Namely,

$$\pi_n^N = \frac{1}{N_0}\sum_{i=1}^{N_n} \delta_{\xi_n^i} \xrightarrow{\text{Branching}} \widehat{\pi}_n^N = \frac{1}{N_0}\sum_{i=1}^{\widehat{N}_n} \delta_{\widehat{\xi}_n^i} \xrightarrow{\text{Mutation}} \pi_{n+1} = \frac{1}{N_0}\sum_{i=1}^{N_{n+1}} \delta_{\xi_{n+1}^i},$$

where $\{(N_n, \xi_n), (\widehat{N}_n, \widehat{\xi}_n);\ n \geqslant 0\}$ is a suitably chosen Markov chain with state space $\mathcal{E} = \bigcup_{\alpha\in\mathbb{N}}(\{\alpha\} \times E^\alpha)$ (with the convention $E^0 = \{\Delta\}$ a cemetery point). Here the parameter $\alpha \in \mathbb{N}$ represents the size of the system and the initial number of particles $N_0 \in \mathbb{N}$ is a fixed nonrandom number which represents the precision parameter of the scheme.

To check that this abstract formulation contains the genetic algorithm presented in Section 1 it suffices to note that it coincides with (5) when the size of the population $N_0 = N_n$ is fixed and

$$P(\widehat{\xi}_n \in dx/\xi_n = z) = \prod_{p=1}^{N_0}\sum_{i=1}^{N_0} \frac{g_n(z^i)}{\sum_{j=1}^{N_0} g_n(z^j)} \delta_{z^i}(dx^p),$$

$$P(\xi_{n+1} \in dx/\widehat{\xi}_n = z) = \prod_{p=1}^{N_0} K_{n+1}(z^p, dx^p).$$

In this situation, if we put

$$M_n^i = \operatorname{Card}\{1 \leqslant p \leqslant N_n \colon \widehat{\xi}_n^p = \xi_n^i\} \quad \text{then } \widehat{\pi}_n^N = \frac{1}{N_0}\sum_{i=1}^{N_n} M_n^i \delta_{\xi_n^i},$$

where, conditionaly on $F_n = \sigma(N_n, \xi_n)$

$$(M_n^1, \ldots, M_n^{N_n}) = \text{Multinomial}(N_n; W_n^1, \ldots, W_n^{N_n}),$$

$$W_n^i = \frac{g_n(\xi_n^i)}{\sum_{j=1}^{N_n} g_n(\xi_n^j)} = \frac{1}{N_n} E(M_n^i/F_n). \tag{42}$$

Sampling according to a multinomial branching law may be "time consuming" mainly because the random numbers $(M_n^1, \ldots, M_n^{N_n})$ are negatively correlated in order to keep fixed the size of the system. Another idea is to use independent numbers (conditionally with respect to F_n) with a suitable law so that the nonbias condition (42) still holds. Let us present some classical examples of independent branching numbers

Poisson branching numbers:

$$\forall k \geqslant 0 \quad P\left(M_n^i = k \mid F_n\right) = \exp\left(-N_n W_n^i\right) \frac{(N_n W_n^i)^k}{k!}.$$

Binomial branching numbers:

$$\forall 0 \leqslant k \leqslant N_n \quad P\left(M_n^i = k \mid F_n\right) = C_{N_n}^k \left(W_n^i\right)^k \left(1 - W_n^i\right)^{N_n - k}.$$

Bernoulli branching numbers:

$$P\left(M_n^i = k \mid F_n\right) = \begin{cases} \{N_n W_n^i\} & \text{if } k = [N_n W_n^i] + 1, \\ 1 - \{N_n W_n^i\} & \text{if } k = [N_n W_n^i], \end{cases}$$

where $[a]$ is the integer part of $a \in \mathbb{R}$ and $\{a\} = a - [a]$.

In the resulting particle schemes the mutation transition is unchanged and consists on sampling independent transitions according to the kernels $\{K_n;\ n \geqslant 1\}$.

As a results during the mutation transition the size of the system is unchanged and we have $N_{n+1} = \widehat{N}_n$.

We note that the multinomial genetic scheme arises by conditioning the branching particle scheme with Poisson branching to have constant population size (see [6]). Theses three variants of the genetic algorithm are known to approximate the desired distribution at each time $n \geqslant 0$ but their long time behavior is still an open question.

In view of the preceding development it is tempting to apply our approach to prove uniform convergence with respect to time. Unfortunately when we move from the genetic scheme with constant size to the branching schemes with random population size we find that we no longer have a uniform convergence with respect to time. More precisely, in view of (42) the total size process $\{N_n;\ n \geqslant 0\}$ is an F-martingale with predictable quadratic variation

$$A_n = N_0^2 + \sum_{p=1}^{n} E\left(|N_p - N_{p-1}|^2 / F_{p-1}\right) = N_0^2 + \sum_{p=0}^{n-1} \sum_{i=1}^{N_p} E\left(\left(M_p^i - N_p W_p^i\right)^2 / F_p\right).$$

To see that the integrability of this increasing process completely determines the long time behavior of such schemes it suffices to note that

$$E\left(\left(\pi_n^N(1) - \pi_n(1)\right)^2\right) = E\left(|1 - N_n/N_0|^2\right) = \frac{1}{N_0^2} E\left((A_n - A_0)^2\right)$$

and therefore a uniform convergence result will take place if and only if

$$\sup_{n \geqslant 0} E\left(A_n^2\right) = \sum_{p=1}^{\infty} E\left(|N_p - N_{p-1}|^2\right) < \infty.$$

The increasing process $\{A_n;\ n \geqslant 0\}$ is usually not uniformly integrable and therefore one cannot expect to obtain a uniform convergence result. For instance, when we use Poisson branching numbers the following basic result

$$E\big((M_p^i - N_p W_p^i)^2/F_p\big) = N_p W_p^i, \quad \forall 1 \leqslant i \leqslant N_{p-1}$$

describes a "typical situation" of independent branching numbers in which the uniform convergence fails. To see this claim we simply observe that

$$E\big((\pi_n^N(1) - \pi_n(1))^2\big) = \frac{n}{N_0} \xrightarrow[n\to\infty]{} \infty.$$

This simple example shows that the particle scheme with independent branchings presented in [7] for solving the nonlinear filtering equation does not converge uniformly with respect to time.

We continue our discussion and examine the long time behavior of the particle scheme with binomial branching numbers. In this situation one can check that

$$E\big((M_p^i - N_p W_p^i)^2/F_p\big) = N_p W_p^i (1 - W_p^i).$$

If we assume that

$$\frac{1}{a} \leqslant g_n(x) \leqslant a, \quad \forall n \geqslant 1,\ \forall x \in E$$

for some $a \geqslant 1$, then one gets

$$\sum_{i=1}^{N_p} N_p W_p^i (1 - W_p^i) \geqslant N_p - a^2$$

which in turns implies that

$$E\big((\pi_n^N(1) - \pi_n(1))^2\big) \geqslant n\left(\frac{1}{N_0} - \frac{a^2}{N_0^2}\right) \xrightarrow[n\to\infty]{} \infty$$

as soon as $N_0 > a^2$.

The Bernoulli branching law seems to be the most efficient one since the independent random variables $(M_p^1, \ldots, M_p^{N_p})$ have minimal variance and the population size cannot vanish (see [6–8]). Nevertheless the following simple example shows that even in this case one cannot expect to approximate the desired system (1) uniformly with respect to time.

Let us assume that the state space $E = \{0, 1\}$, the fitness functions $\{g_n;\ n \geqslant 1\}$ and the transition kernels $\{K_n;\ n \geqslant 1\}$ are time homogeneous and given by

$$g(1) = 3g(0) > 0, \quad K(x, dz) = \nu(dz) \stackrel{\text{def.}}{=} \frac{1}{2}\delta_0(dz) + \frac{1}{2}\delta_1(dz).$$

In this simple situation the system $\xi_p = (\xi^1, \ldots, \xi_p^{N_p})$ consists of N_p i.i.d. particles with common law ν. This yields

$$\forall \varepsilon > 0 \quad P\left(\left|\frac{1}{N_p}\sum_{i=1}^{N_p} g(\xi_p^i) - \nu(g)\right| \geqslant \varepsilon g(0)/N_p\right) \leqslant \frac{5}{\varepsilon^2 N_p}. \tag{43}$$

Noticing that $\nu(g)/g(0) = 2 = 3\nu(g)/g(1)$ and $g(0) \leqslant g(1)$, on the set

$$\Omega_\varepsilon = \left\{\left|\frac{1}{N_p}\sum_{i=1}^{N_p} g(\xi_p^i) - \nu(g)\right| \leqslant \varepsilon g(0)\right\}$$

we have that

$$\left|\frac{g(0)}{\frac{1}{N_p}\sum_{i=1}^{N_p} g(\xi_p^i)} - \frac{1}{2}\right| \leqslant \frac{\varepsilon}{2(2-\varepsilon)} \leqslant \frac{\varepsilon}{2} \quad \text{and}$$
$$\left|\frac{g(1)}{\frac{1}{N_p}\sum_{i=1}^{N_p} g(\xi_p^i)} - \frac{3}{2}\right| \leqslant \frac{3\varepsilon}{2(2/3-\varepsilon)} \leqslant \frac{9\varepsilon}{2}$$

as soon as $\varepsilon \in (0, 1/9)$. This in turns implies that

$$\left[\frac{g(0)}{\frac{1}{N_p}\sum_{i=1}^{N_p} g(\xi_p^i)}\right] = 0 \quad \text{and} \quad \left[\frac{g(1)}{\frac{1}{N_p}\sum_{i=1}^{N_p} g(\xi_p^i)}\right] = 1,$$

and

$$\xi_p^i = 0 \Longrightarrow \{N_p W_p^i\}\left(1 - \{N_p W_p^i\}\right) \geqslant \frac{1}{4}(1-\varepsilon)^2,$$
$$\xi_p^i = 1 \Longrightarrow \{N_p W_p^i\}\left(1 - \{N_p W_p^i\}\right) \geqslant \frac{1}{4}(1-9\varepsilon)^2.$$

It is then clear that on the set Ω_ε we have the lower bounds

$$E\left(\left(M_p^i - N_p W_p^i\right)^2 / F_p\right) = \{N_p W_p^i\}\left(1 - \{N_p W_p^i\}\right) \geqslant \frac{1}{4}(1-9\varepsilon)^2.$$

This, together with (43), shows that

$$E\left(\sum_{i=1}^{N_p}\left(M_p^i - N_p W_p^i\right)^2\right) \geqslant \frac{1}{4}(1-9\varepsilon)^2\left(N_0 - \frac{5}{\varepsilon^2}\right).$$

As soon as $N_0 > 5\varepsilon^{-2}$ one concludes that

$$E\left(\left(\pi_n^N(1) - \pi_n(1)\right)^2\right) \geqslant \frac{n}{4}(1-9\varepsilon)^2\left(\frac{1}{N_0} - \frac{5}{N_0^2\varepsilon^2}\right).$$

Hence it follows that for sufficiently large N_0

$$E\left(\left(\pi_n^N(1)-\pi_n(1)\right)^2\right)\geqslant\frac{nN_0}{5}\xrightarrow[n\to\infty]{}\infty.$$

In contrast to the situation described above in this simple case the genetic algorithm (5) will consist at each time of N_0 i.i.d. particles with common law ν and

$$\forall n\geqslant 0\quad E\left(\left(\pi_n^N(f)-\pi_n(f)\right)^2\right)\leqslant\frac{1}{N_0}$$

for any bounded test function such that $\|f\|\leqslant 1$.

Acknowledgements

We gratefully acknowledge the European Community for the research fellowships CEC Contract No ERB-FMRX-CT96-0075 and INTAS-RFBR 95-0091. We also thank the anonymous referee for his comments and suggestions that improve the presentation of the paper.

REFERENCES

[1] Atar R., Exponential stability for nonlinear filtering of diffusion processes in a noncompact domain, Ann. Probab. 26 (4) (1998) 1552–1574.

[2] Atar R., Zeitouni O., Liapunov exponents for finite state space nonlinear filtering, SIAM J. Control Optim. 35 (1) (1997) 36–55.

[3] Atar R., Zeitouni O., Exponential stability for nonlinear filtering, Ann. Inst. H. Poincare 33 (6) (1997) 697–725.

[4] Bucy R.S., Lectures on discrete time filtering, Signal Processing and Digital Filtering, Springer Verlag, 1994.

[5] Budhiraja A., Ocone D., Exponential stability of discrete time filters for bounded observation noise, Systems and Control Letters 30 (1997) 185–193.

[6] Crisan D., Del Moral P., Lyons T.J., Discrete filtering using branching and interacting particle systems, Markov Processes and Related Fields 5 (3) (1999) 293–319.

[7] Crisan D., Lyons T.J., Nonlinear filtering and measure valued processes, Probab. Theory Related Fields 109 (1997) 217–244.

[8] Crisan D., Gaines J., Lyons T.J., A particle approximation of the solution of the Kushner–Stratonovitch equation, SIAM J. Appl. Math. 58 (5) (1998) 1568–159.

[9] Da Prato G., Furhman M., Malliavin P., Asymptotic ergodicity for the Zakai filtering equation, C.R. Acad. Sci. Paris, Série I 321 (1995) 613–616.

[10] Del Moral P., Jacod J., Interacting Particle Filtering With Discrete Observations, Publications du Laboratoire de Statistiques et Probabilités, Université Paul Sabatier, No 11-99, 1999.

[11] Del Moral P., Nonlinear filtering using random particles, Theor. Prob. Appl. 40 (4) (1995).

[12] Del Moral P., Non-linear filtering: interacting particle solution, Markov Processes and Related Fields 2 (4) (1996) 555–581.

[13] Del Moral P., Measure valued processes and interacting particle systems. Application to nonlinear filtering problems, Ann. Appl. Probab. 8 (2) (1998) 438–495.

[14] Del Moral P., A uniform theorem for the numerical solving of nonlinear filtering problems, J. Appl. Probab. 35 (1998) 873–884.
[15] Del Moral P., Filtrage non linéaire par systèmes de particules en intéraction, C.R. Acad. Sci. Paris, Série I 325 (1997) 653–658.
[16] Del Moral P., Guionnet A., Large deviations for interacting particle systems. Applications to nonlinear filtering problems, Stochastic Processes and their Applications 78 (1998) 69–95.
[17] Del Moral P., Guionnet A., A central limit theorem for nonlinear filtering using interacting particle systems, Ann. Appl. Probab. 9 (2) (1999) 275–297.
[18] Delyon B., Zeitouni O., Liapunov exponents for filtering problems, in: Davis M.H.A., Elliot R.J. (Eds.), Applied Stochastic Analysis, 1991, pp. 511–521.
[19] Dobrushin R.L., Central limit theorem for nonstationnary Markov chains, I,II, Theory Probab. Appl. 1 (1,4) (1956) 66–80 and 330–385.
[20] Dobrushin R.L., Prescribing a system of random variables by conditional distributions, Theor. Prob. Appl. 15 (3) (1970).
[21] Kunita H., Asymptotic behavior of the nonlinear filtering errors of Markov processes, J. Multivariate Analysis 1 (1971) 365–393.
[22] Kunita H., Ergodic properties nonlinear filtering processes, in: Alexander K.C., Watkins J.C. (Eds.), Spatial Stochastic Processes, 1991.
[23] Norman M.F., Ergodicity of diffusion and temporal uniformity of diffusion approximations, J. Appl. Prob. 14 (1977) 399–404.
[24] Ocone D.L., Topics in nonlinear filtering theory, Ph.D. Thesis, MIT Press, Cambridge, MA, 1980.
[25] Ocone D., Pardoux E., Asymptotic stability of the optimal filter with respect to its initial condition, SIAM J. Control Optim. 34 (1996) 226–243.
[26] Pardoux E., Filtrage Non Linéaire et Equations aux Dérivés Partielles Stochastiques Associées, Ecole d'été de Probabilités de Saint-Flour XIX-1989, Lecture Notes in Mathematics, Vol. 1464, Springer-Verlag, 1991.
[27] Parthasarathy K.R., Probability Measures on Metric Spaces, Academic Press, New York, 1968.
[28] Rowe J.R., Population fixed points for functions of unitation, in: Barzhaf W., Reeves C. (Eds.), Foundations of Genetic Algorithms 5, Morgan Kauffmann, 1999, pp. 69–84.
[29] Stettner L., On invariant measures of filtering processes, in: Helmes K., Kohlmann N. (Eds.), Stochastic Differential Systems, Proc. 4th Bad Honnef Conference, 1988, Lecture Notes in Control and Inform. Sci., 1989, pp. 279–292.
[30] Stettner L., Invariant measures of pair state/approximate filtering process, Colloq. Math. LXII (1991) 347–352.
[31] Van Nimwegen E., Crutchfield J.P., Michell M., Finite populations induce metastability in evolutionary search, Physics Letters A 229 (2) (1997) 144–150.
[32] Vose M.D., Logarithmic convergence of random heuristic search, Evolutionary Computation 4 (4) (1997) 395–404.
[33] Vose M.D., Modelling simple genetic algorithms, in: Foundations of Genetic Algorithms, Morgan Kaufmann, 1993.
[34] Vose M.D., Modelling simple genetic algorithms, Elementary Computations 3 (4) (1995) 453–472.
[35] Vose M.D., Liepins G.E., Punctuated equilibra in genetic search, Complex Systems 5 (1993) 31–44.
[36] Vose M.D., Wright A.H., Simple genetic algorithms with linear fitness, Elementary Computations 2 (4) (1995) 347–368.