Asymptotic Results for Genetic Algorithms with Applications to Non Linear Estimation

P. Del Moral L. Miclo

Abstract

Genetic algorithms are stochastic search methods based on natural evolution processes. They are defined as a system of particles (or individuals) evolving randomly and undergoing adaptation in a time non necessarily homogeneous environment represented by a collection of fitness functions. The purpose of this work is to study the long time behavior as well as large population asymptotic of genetic algorithms. Another side topic is to discuss the applications of genetic algorithms in numerical function analysis, Feynman-Kac formulae approximations and in non linear filtering problems. Several variations and refinements will also be presented including continuous time and branching particle models with random population size.

Keywords : Genetic algorithms, Branching and Interacting particle systems, Feynman-Kac formula, non linear Filtering, Numerical function Optimization.

1 Introduction

In [37] J. H. Holland introduced genetic algorithms as a kind of universal and global search method based on natural evolution processes. During the last two decades they have been used as an optimization tool in a variety of research areas, to name a few: machine learning [34], control systems [33], electromagnetics [39, 53], economics and finance [42, 49], aircraft landing [1, 16], topological optimum design [40] and identification of mechanical inclusions [51, 52].

More recently genetic algorithms have appeared naturally in the study of Feynman-Kac formulas and non linear filtering problems (the reader is recommended to consult the survey paper [20] and references therein).

These particle interpretations of Feynman-Kac models has had numerous applications in many nonlinear filtering problems; to name a few, radar signal processing ([28, 30]), Global Positioning System ([6, 7]) as well as in tracking problems ([41, 45, 46, 35]). Other numerical experiments are also given in [11] and [17].
In contrast to the applications in numerical function analysis genetic algorithms are not used here to approximate the extrema of a given numerical function but a flow of conditional distributions. In addition the genetic structure of the algorithm (such as the mutation and the selection transitions) is not only designed as an instrumental tool to mimic natural evolution but it is in fact dictated by the dynamics structure of the so-called non linear filtering equations.

The main purpose of this article is to introduce the reader to the asymptotic theory of genetic algorithms. We also explain the use of these stochastic methods for the numerical solving of non linear filtering problems and in numerical function optimization problems. We also give a detailed discussion on several variations and refinements algorithms recently proposed in the literature of non linear estimation problems.

This work is essentially divided into two main parts devoted respectively to the applications of genetic algorithms for the numerical solving of the so-called non linear filtering equations and the convergence of genetic algorithms towards the global minima of a given numerical function. Our presentation of this material has relied heavily on the two papers [20] and [21].

In the opening section 2 we introduce the two step mutation-selection procedure and the time inhomogeneous Markov model of genetic algorithms treated in this work. As announced above this model will then be regarded as a global stochastic search method for studying the set of the global minima of a given numerical function or as a stochastic adaptative grid approximation of a flow of conditional distributions in non linear filtering settings.

To each of these applications correspond a specific asymptotic analysis. In section 2 we lay the foundations of the work that follows by explaining the general methodologies needed to study the large population asymptotic and the long time behavior of the algorithm. In section 2.1 we give an alternative description of the genetic model presented in section 2 in terms of an N-interacting particle system approximating model associated to a measure valued dynamical system. This formulation enables us to identify the limit of the empirical measures associated to the genetic algorithm in terms of a Feynman-Kac

formula. The modeling impact of this approach will be illustrated in non linear filtering in section 3.

Section 2.2 is devoted to the study of the long time behavior of the genetic model presented in section 2. The idea here is to connect genetic algorithms with the so-called generalized simulated annealing. We describe a general methodology to conclude that a genetic algorithm converges in probability as time tends to infinity to the set of global minima of a virtual energy function. We will combine this general convergence result with a natural test set approach in section 4 to prove that the resulting stochastic algorithm converges towards the set of global minima of the desired fitness function as the time parameter tends to infinity and when the population size is sufficiently large.

The genetic algorithm presented in section 2 and further developed in section 3 and section 4 is the crudest of the evolutionary particle methods. In section 5 we discuss several variations and refinements arising in the literature about non linear filtering and generalized simulated annealing.

The final section discusses continuous time genetic algorithms. In contrast to the classical Moran-type genetic model commonly used in genetic algorithms literature, our interacting particle system model converges to a deterministic distribution flow. This model has been proposed in [23] for solving continuous time non linear filtering problems. Several variants based on an auxiliary time discretization procedure can be found in [11] and [13]. The fundamental difference between the Moran-type particle scheme and the algorithmss presented in [11, 13]. lies in the fact that in the former competitive interactions occur at random times. The resulting scheme is therefore a genuine continuous time particle approximating model.
The interested reader is referred to [20] for a detailed description of the robust and pathwise filter and for a complete proof of the convergence results in a context more general than we have given. Here we have chosen to restrict our attention to continuous time Feynman-Kac formulae. We will also discuss the connections between this scheme and the generalized and spatially homogeneous Boltzmann models presented in [36, 44].
We end this paper with a novel branching genetic algorithm in which the size of the population is not necessarily fixed but random.
Only a selection of existing results is presented here. Deeper information should be available in [19, 20, 21].

2 Description of the Models, Statement of Some Results

The simplest genetic algorithm is a two stages and time in-homogeneous Markov chain given for each $n \geq 0$ by setting

$$\xi_n \overset{\text{def.}}{=} \left(\xi_n^1, \ldots, \xi_n^N\right) \xrightarrow{\text{Selection}} \widehat{\xi}_n \overset{\text{def.}}{=} \left(\widehat{\xi}_n^1, \ldots, \widehat{\xi}_n^N\right) \xrightarrow{\text{Mutation}} \xi_{n+1}$$

and taking values in a product space E^N where $N \geq 1$ and E is an abstract topological space. The coordinates of points of E^N are seen as positions of N particles and the integer parameter N represents the size of the population.

- The initial system $\xi_0 = \left(\xi_0^1, \ldots, \xi_0^N\right)$ consists of N independent random particles with a common law η_0 on E.
- In the selection transition the particles $\widehat{\xi}_n = \left(\widehat{\xi}_n^1, \ldots, \widehat{\xi}_n^N\right)$ are chosen randomly and independently in the previous configuration $\xi_n = \left(\xi_n^1, \ldots, \xi_n^N\right)$ according to a given non necessarily homogeneous fitness function

  $$g_n : E \to \mathbb{R}_+,$$

  namely

  $$\mathbb{P}\left(\widehat{\xi}_n \in dx \,|\xi_n = y\right) = \prod_{p=1}^{N} \sum_{i=1}^{N} \frac{g_n(y^i)}{\sum_{j=1}^{N} g_n(y^j)} \; \delta_{y^i}(dx^p) \tag{1}$$

  where $dx \overset{\text{def}}{=} dx^1 \times \cdots \times dx^N$ is an infinitesimal neighborhood of the point $x = (x^1, \ldots, x^N) \in E^N$, $y = (y^1, \ldots, y^N) \in E^N$ and δ_a stands for the Dirac measure at $a \in E$.
- The mutation transition is modelled by independent motions of each particle that is

  $$\mathbb{P}\left(\xi_{n+1} \in dx \,\middle|\, \widehat{\xi}_n = y\right) = \prod_{p=1}^{N} K_{n+1}\left(y^p, dx^p\right) \tag{2}$$

  where $\{K_n \,;\, n \geq 1\}$ is a collection on Markov transition kernels from E into itself.

The study of the convergence as $n \to \infty$ or as $N \to \infty$ of this algorithm requires specific developments.
To explain and motivate the organization of our work in the next two subsections we describe the main ideas involved in the study of these different asymptotics as well as some of their consequences in the study

of non linear estimation problems.

Before turning to further details it is convenient at this point to give a couple of remarks.

As we said above, the previous selection-mutation Markov chain is the crudest of the genetic type methods. They are in fact a number of ways to construct variations on this model (see for instance [12, 20, 21] and section 5). In particular the definition of the initial system as N i.i.d. particles is not really essential. In numerical function analysis the asymptotic results as $n \to \infty$ (and fixed N) presented here (see also [8] and [21]) are valid for any choice of N starting points. In non linear filtering settings we will be interested in the asymptotic behavior of the empirical measures of the system as $N \to \infty$. In this framework the initial distribution η_0 is not arbitrarily chosen but it represents the initial law of the state signal. Therefore the initial configuration of the particle systems will be chosen so that the associated empirical measure is an N-approximating measure of η_0.

Another more general remark is that in filtering problems the choice of quantities (η_0, g_n, K_n) is dictated by the problem at hand. In some situations the initial law η_0, the transitions of the state signal K_n and/or the corresponding fitness functions g_n are not explicitly known and/or we cannot simulate random variables exactly according to η_0 and/or K_n. Therefore we need to introduce additional approximating quantities $(\eta_0^{(M)}, g_n^{(M)}, K_n^{(M)})$, where the parameter $M \geq 1$ is a measure of the quality of the approximation so that in some sense $(\eta_0^{(M)}, g_n^{(M)}, K_n^{(M)}) \to (\eta_0, g_n, K_n)$ as $M \to \infty$. The way the two asymptotics $N \to \infty$ and $M \to \infty$ combine are studied in all details in [18].

2.1 Large Population Asymptotic

To show one of the central roles played by the selection/mutation transitions (1) and (2) we start with the study of the asymptotic behavior of the empirical measures associated to the systems of particles ξ_n and $\widehat{\xi}_n$,

$$m(\xi_n) \stackrel{\text{def.}}{=} \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi_n^i} \qquad \text{and} \qquad m(\widehat{\xi}_n) \stackrel{\text{def.}}{=} \frac{1}{N} \sum_{i=1}^{N} \delta_{\widehat{\xi}_n^i}, \qquad n \geq 0$$

as the number of particles N tends to infinity. It is transparent from the previous construction that the pair selection/mutation transition

can be summarized, for each each $n \geq 0$, as follows

$$\mathbb{P}\left(\xi_{n+1} \in dx \,|\xi_n = y\,\right) = \prod_{p=1}^{N} \sum_{i=1}^{N} \frac{g_n(y^i)}{\sum_{j=1}^{N} g_n(y^j)} \; K_{n+1}\left(y^i, dx^p\right). \quad (3)$$

In order to obtain a more tractable description of (3) in terms of a transition which only depends on the empirical measure of the system $m(\xi_n)$ it is convenient to introduce some additional notations. We recall that any transition probability kernel $K(x, dy)$ on E generates two integral operators. One is acting on the set $\mathcal{B}_b(E)$ of bounded Borel test functions $f : E \to \mathbb{R}$ endowed with the supremum norm, defined by

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

and the other on the set $\mathbf{M}_1(E)$ of probability measures μ on E

$$K(f)(x) \stackrel{\text{def.}}{=} \int K(x, dz)\, f(z)$$

and

$$(\mu K)(f) \stackrel{\text{def.}}{=} \mu(Kf) = \int \mu(dx)\, K(x, dz)\, f(z)$$

If K_1 and K_2 are two integral operators on $\mathcal{B}_b(E)$ we denote by K_1K_2 the composite operator on $\mathcal{B}_b(E)$ defined for any $f \in \mathcal{B}_b(E)$ by

$$K_1K_2f(x) = \int_E K_1(x, dy)\, K_2(y, dz)\, f(z)$$

Using these notations (3) can be rewritten as

$$\mathbb{P}\left(\xi_{n+1} \in dx \,|\xi_n = y\,\right) = \prod_{p=1}^{N} \Phi_{n+1}\left(\frac{1}{N}\sum_{i=1}^{N} \delta_{y^i}\right)(dx^p), \qquad n \geq 0 \quad (4)$$

where for all $n \geq 0$, $\Phi_{n+1} : \mathbf{M}_1(E) \to \mathbf{M}_1(E)$ is the mapping defined by

$$\Phi_{n+1}(\eta) \stackrel{\text{def.}}{=} \Psi_n(\eta)K_{n+1} \quad \text{with} \quad \Psi_n(\eta)(f) \stackrel{\text{def.}}{=} \frac{\eta(g_n\, f)}{\eta(g_n)} \qquad \forall f \in \mathcal{B}_b(E) \quad (5)$$

We note that the one-step mapping Φ_{n+1} involves two separate transitions: The first one $\eta \mapsto \Psi_n(\eta)$ is nonlinear and it will be called the updating step and the second one $\eta \mapsto \eta K_{n+1}$ will be called the prediction transition with reference to filtering theory.

With this formulation it also becomes quite clear that the flow of empirical measures $\{m(\xi_n)\ ;\ n \geq 0\}$ converge in some sense as $N \to$ to the solution $\{\eta_n\ ;\ n \geq 0\}$ of the following measure valued process

$$\eta_n = \Phi_n\left(\eta_{n-1}\right), \qquad n \geq 1 \tag{6}$$

Intuitively speaking if $m(\xi_{n-1})$ is close to the desired distribution η_{n-1} then one expect that $\Phi_n(m(\xi_{n-1}))$ is a nice approximating measure for η_n. Therefore at the next step the particle system $\xi_n = (\xi_n^1, \ldots, \xi_n^N)$ looks like a sequence of independent random variables with common law η_n and therefore $m(\xi_n)$ is close to the desired distribution η_n...

As a parenthesis and along the same ideas we can associate to any abstract measure valued process (6) an N-interacting particle approximating model as in (4). In other words the previous algorithm is a particular example of particle approximating model and the mutation/selection transitions are dictated by the form of the limiting measure valued dynamical system (6).

In our situation, the preceding scheme is clearly a system of interacting particles undergoing adaptation in a time non-homogeneous environment represented by the fitness functions $\{g_n;\ n \geq 0\}$ and the selection/mutation transitions are dictated by the nature of the two step mappings $\{\Phi_n\ ;\ n \geq 1\}$. Roughly speaking the natural idea is to approximate the two step transitions

$$\eta_n \xrightarrow{\text{Updating}} \widehat{\eta}_n \stackrel{\text{def}}{=} \psi_n(\eta_n) \xrightarrow{\text{Prediction}} \eta_n = \widehat{\eta}_n K_{n+1}, \qquad n \geq 0$$

of the system (6) by a two step Markov chain taking values in the set of finitely discrete probability measures with atoms of size some integer multiple of $1/N$. Namely, for each $n \geq 0$,

$$\eta_n^N \stackrel{\text{def.}}{=} \frac{1}{N}\sum_{i=1}^{N} \delta_{\xi_n^i} \xrightarrow{\text{Selection}} \widehat{\eta}_n^N \stackrel{\text{def.}}{=} \frac{1}{N}\sum_{i=1}^{N} \delta_{\widehat{\xi}_n^i} \xrightarrow{\text{Mutation}} \eta_{n+1}^N = \frac{1}{N}\sum_{i=1}^{N} \delta_{\xi_{n+1}^i}. \tag{7}$$

These constructions first appeared in [26] and [27] and they were developed in [24]. In [20] the authors present an exposé of the mathematical theory that it is useful in analyzing the convergence of such particle approximating models including law of large numbers, large deviations principles, fluctuations and empirical process theory as well as semigroup techniques and limit theorems for processes. In section 3 we briefly indicate some of the main directions explored in this recent research and we will introduce the reader to some mathematical tools upon which the theory dwells.

Anticipating on section 3.1 we also mention that the measure valued dynamical system (6) can be explicitly solved. More precisely if

$$X = \{X_n \; ; \; n \geq 0\}$$

denotes a time in-homogeneous Markov chain with transition probability kernels $\{K_n \; ; \; n \geq 1\}$ and initial distribution η_0 and if $\gamma_n(f)$, $f \in \mathcal{B}_b(E)$, represents the Feynman-Kac formula

$$\gamma_n(f) = \mathbb{E}\left(f(X_n) \prod_{k=0}^{n-1} g_k(X_k) \right)$$

(with the convention $\prod_{\emptyset} = 1$) then the distribution flow $\{\eta_n; \; n \geq 0\}$ defined for any $n \geq 0$ and for any test function $f \in \mathcal{B}_b(E)$ as the ratio

$$\eta_n(f) = \frac{\gamma_n(f)}{\gamma_n(1)} \tag{8}$$

is solution of the measure valued dynamical system (6). In fact, as we shall see in the further development of section 3 the classical non linear filtering problem can be summarized as to find distributions of the form (8). In this framework the probability kernels $\{K_n \; ; \; n \geq 1\}$ represent the transitions of the signal process and the fitness functions $\{g_n \; ; \; n \geq 0\}$ depend on the observation data and on the density of the noise source.

2.2 Large Time Behavior

Our next objective is to initiate the study of the long time behavior of the genetic type algorithms. In contrast to the situation presented in section 2.1 the size N of the particle systems is fixed and the genetic model is thought as a global search procedure for studying the set $U^\star$ of global minima of a given numerical function $U : E \to \mathbb{R}_+$ and the state space E is assumed to be finite, namely

$$U^\star \stackrel{\text{def.}}{=} \left\{ x \in E \; ; \; U(x) = \min_E U \right\}$$

To clarify the notations we shall use the following notations

$$\mathcal{Q}_n^{(1)}(x, dy) = \prod_{p=1}^{N} K_n(x^p, dy^p)$$

and

$$\mathcal{Q}_n^{(2)}(x, dy) = \prod_{p=1}^{N} \sum_{i=1}^{N} \frac{g_n(y^i)}{\sum_{j=1}^{N} g_n(y^j)} \; \delta_{y^i}(dx^p)$$

Thus $\xi = \{\xi_n \ ; \ n \geq 0\}$ is a time in-homogeneous Markov chain with transition probability kernel

$$\mathcal{Q}_n = \mathcal{Q}_{n-1}^{(2)} \mathcal{Q}_n^{(1)}$$

and $\widehat{\xi} = \{\widehat{\xi}_n \ ; \ n \geq 0\}$ is a time in-homogeneous Markov chain with transition probability kernel

$$\widehat{\mathcal{Q}}_n = \mathcal{Q}_n^{(1)} \mathcal{Q}_n^{(2)} \tag{9}$$

In time homogeneous settings (that is if $K_n = K$ and $g_n = g$) the general theory of time homogeneous Markov chains can be used to study the long time behavior of these two chains but to our knowledge the stochastic stability results which can be stated are not really useful to calibrate the convergence of genetic algorithms to the desired extrema of a given numerical function.
One of the apparent difficulty in establishing a useful convergence result as $n \to \infty$ is finding a candidate invariant measure which enable us to describe some interesting aspects of the limiting behavior of the algorithm.

The key idea is to introduce an inverse cooling schedule parameter $\beta : \mathbb{R}_+ \to \mathbb{R}_+$ with $\lim_{t \to \infty} \beta(t) = \infty$ so that to reduce the analysis to the study of a generalized simulated annealing. This idea has been initiated in [8, 9, 10] and it has been simplified and further extended in [21].
As the time parameter is growing the arbitrary exploration of the path space by the particles during the mutation step will progressively disappear. The precise choice of the mutation transitions K_n in terms of the parameter $\beta(n)$ will be given in section 4.1. We already mention that in the selection transitions the fitness functions g_n will take the form

$$g_n(x) = e^{-\beta(n)\, U(x)}, \qquad n \geq 1$$

and, as the time is growing, the randomness in the selection will also tend to disappear so that the particles with below pick fitness will progressively not be selected.

The purpose of this paper is to present some theoretical background needed to analyze the convergence of the algorithm. The results presented here will be restricted to the transition probability kernel (9) and they can be found with complete proof in [21].
In this opening section we describe the basic but general idea which is in fact quite simple. This methodology will be used in several part of this paper. It is also quite general and it can be used in other contexts.

We have tried to present easily verifiable conditions and results at a relevant level of generality. Our claim that this description of the mutation and selection transitions is the natural framework for formulating and studying the long time behavior of genetic algorithms in numerical function analysis will be amply justified by the results that will follows.

We provide no examples in this short section, this choice is deliberate. In section 4 we will show how to obtain the transitions $Q_\beta^{(1)}$ and $Q_\beta^{(2)}$ in terms of the mutation kernels and the fitness functions. We will also use this framework in section 5 for studying a related genetic algorithm in which the mutation and the selection stage take place randomly at each time step. We also believe that it is possible to use this formulation to analyze the convergence of the branching genetic type variants presented in section 5.

To commence to formalize this we first chose the mutation/selection transitions $\mathcal{Q}_n^{(1)}$ and $\mathcal{Q}_n^{(2)}$ as governed by $\beta(n)$, that is

$$\mathcal{Q}_n^{(1)} = Q_{\beta(n)}^{(1)} \qquad \text{and} \qquad \mathcal{Q}_n^{(2)} = Q_{\beta(n)}^{(2)} \tag{10}$$

and so that for any $\beta > 0$, $Q_\beta^{(1)}$ and $Q_\beta^{(1)}$ take the form

$$Q_\beta^{(1)}(x,y) = q_\beta^{(1)}(x,y)\; e^{-\beta\, V^{(1)}(x,y)}$$

and

$$Q_\beta^{(2)}(x,y) = q_\beta^{(2)}(x,y)\; e^{-\beta\, V^{(2)}(x,y)}$$

for some numerical functions $q_\beta^{(1)}$, $q_\beta^{(2)} : E^N \times E^N \longrightarrow \mathbb{R}_+$ and $V^{(1)}$, $V^{(2)} : E^N \times E^N \longrightarrow \overline{\mathbb{R}}_+$ ($\overline{\mathbb{R}}_+ \stackrel{\text{def.}}{=} \mathbb{R}_+ \cup \{+\infty\}$).

It is then straightforward to check that the transition probability kernels

$$\widehat{Q}_\beta(x,y) \stackrel{\text{def.}}{=} Q_\beta^{(1)} Q_\beta^{(2)}(x,y)$$

take the form

$$\widehat{Q}_\beta(x,y) = \sum_{v \in \mathcal{V}} \widehat{q}_\beta(x,v,y)\; e^{-\beta\, \widehat{V}(x,v,y)} \tag{11}$$

with $\mathcal{V} = E^N$ and

$$\widehat{q}_\beta(x,v,y) = q_\beta^{(1)}(x,v) q_\beta^{(2)}(v,y)$$

and

$$\widehat{V}(x,v,y) = V^{(1)}(x,v) + V^{(2)}(v,y)$$

This kind of mathematical models naturally arise when studying the long time behavior of stochastic algorithms such as the generalized simulated annealing. The parameter β in (10) will be regarded as the inverse freezing schedule in classical simulated annealing and it will be used to control the random perturbations of the stochastic algorithm. When $\beta \to \infty$ the random perturbations will progressively disappear and the two different cost functions $V^{(1)}$ and $V^{(2)}$ will be regarded respectively as the mutation and selection costs to communicate from a population to another.

The objective is to prove that the law of a well chosen time inhomogeneous genetic particle scheme concentrates as times tends to infinity to the set $U^\star$ of global minima of a desired numerical function $U : E \to \mathbb{R}_+$. In order to prove this asymptotic result we need to characterize more explicitly the long time behavior of the algorithm in terms of the communication cost functions $V^{(1)}$ and $V^{(2)}$.

As traditionally, under some nice conditions, the first step consists in proving that the algorithm converges to the set of the global minima $W^\star$ of a virtual energy function $W : E^N \to \mathbb{R}_+$ defined explicitly in terms of the communication cost functions $V^{(1)}$ and $V^{(2)}$.
The second subtle step will be to find conditions on the population size which ensures that $W^\star$ is contained into the subset $U^\star \times \ldots \times U^\star (\subset E^N)$. We will settle this question in section 4 and 5 by using a natural test set approach.

Under appropriate continuity and irreductibility conditions the first step can be solved using quite general results on the generalized simulated annealing. Anticipating on section 5.1 we also notice that the transition probability kernel $\tilde{Q}_\beta$ defined by

$$\tilde{Q}_\beta = \alpha_1 \, Q_\beta^{(1)} + \alpha_2 \, Q_\beta^{(2)} \qquad \alpha_1 + \alpha_2 = 1 \quad (\alpha_1, \alpha_2 \in (0,1)) \qquad (12)$$

can be written as in (11) with $\mathcal{V} = \{1, 2\}$.

The precise continuity and irreductibility condition needed to handle to the first step are summarized in the following assumption.

(H) The transition probability kernels $\widehat{Q}_\beta$ take the form

$$\widehat{Q}_\beta(x, y) = \sum_{v \in \mathcal{V}} \widehat{q}_\beta(x, v, y) \, e^{-\beta \, \widehat{V}(x,v,y)}$$

where $\mathcal{V}$ is a finite set and there exists a non negative function

$$\widehat{q} : E^N \times \mathcal{V} \times E^N \longrightarrow \mathbb{R}_+$$

so that

- **For any** $x, y \in E^N$ **and** $v \in \mathcal{V}$ **and** $\beta > 0$ **we have**

$$\lim_{\beta \to +\infty} \widehat{q}_\beta(x, v, z) = \widehat{q}(x, v, z)$$

and

$$\widehat{q}_\beta(x, v, z) > 0 \Longleftrightarrow \widehat{q}(x, v, z) > 0$$

- **For every** $\widehat{q}(x, v, z) > 0$ **and for some** $\beta_0 \geq 0$

$$\sup_{\beta \geq \beta_0} |\frac{d \log \widehat{q}_\beta}{d\beta}(x, v, z)| < +\infty$$

- **For any** $x, y \in E^N$ **there exists a integer** $r \geq 1$ **and sequence of elements** $(p_k, v_k)_{0 \leq k \leq r}$ **in** $E^N \times \mathcal{V}$ **such that**

$$p_0 = x \qquad \text{and} \quad \widehat{q}(p_k, v_k, p_{k+1}) > 0 \quad \forall 0 \leq k < r \quad \text{and} \quad p_r = y$$

It is transparent from these conditions that we have some suitable function $\epsilon(\beta) \to 0$, as $\beta \to +\infty$, such that

$$(1 - \epsilon(\beta))\ Q_\beta(x, y) \ \leq \ \widehat{Q}_\beta(x, y) \ \leq \ (1 + \epsilon(\beta))\ Q_\beta(x, y) \qquad (13)$$

where

$$Q_\beta(x, y) = \sum_{v \in \mathcal{V}(x,y)} \widehat{q}(x, v, y)\ e^{-\beta \widehat{V}(x,v,y)}$$

and

$$\mathcal{V}(x, y) = \{v \in \mathcal{V} \ : \ \widehat{q}(x, v, y) > 0\}$$

But, if we write

$$\begin{aligned} V(x, y) &= \min_{v \in \mathcal{V}(x,y)} \widehat{V}(x, v, y) \\ q(x, y) &= \sum_{v \in \mathcal{V}^\star(x,y)} \widehat{q}(x, v, y) \\ \mathcal{V}^\star(x, y) &= \{v \in \mathcal{V}(x, y) \ : \ \widehat{V}(x, v, y) = V(x, y)\} \end{aligned}$$

then we also have that

$$\begin{aligned} & Q_\beta(x, y) \\ &= \sum_{v \in \mathcal{V}^\star(x,y)} \widehat{q}(x, v, y)\ e^{-\beta V(x,y)} \\ &\qquad + e^{-\beta V(x,y)} \sum_{v \in \mathcal{V}(x,y) - \mathcal{V}^\star(x,y)} \widehat{q}(x, v, y)\ e^{-\beta(\widehat{V}(x,v,y) - V(x,y))} \\ &= q(x, y)\ e^{-\beta V(x,y)} \\ &\qquad + e^{-\beta V(x,y)} \sum_{v \in \mathcal{V}(x,y) - \mathcal{V}^\star(x,y)} \widehat{q}(x, v, y)\ e^{-\beta(\widehat{V}(x,u,y) - V(x,y))} \end{aligned}$$

Note that condition (**H**) implies that q is irreducible. Furthermore, if we write,

$$\begin{aligned} I &= \{(x,y) \in E^2 \ : \ \mathcal{V}(x,y) \neq \emptyset\} \\ J &= \{(x,v,y) \in E^N \times \mathcal{V} \times E^N \ : \ (x,y) \in I \ \ v \in \mathcal{V}(x,y)\} \end{aligned}$$

and

$$\begin{aligned} h_1 &= \min_{(x,y)\in I} \sum_{v \in \mathcal{V}(x,y) - \mathcal{V}^\star(x,y)} \widehat{q}(x,v,y)/q(x,y) \\ h_2 &= \min_{(x,v,y) \, : \, v \notin \mathcal{V}^\star(x,y)} (\widehat{V}(x,v,y) - V(x,y)) \end{aligned}$$

using (13) we get the system of inequalities

$$(1-\epsilon(\beta)) \ q(x,y) \ e^{-\beta V(x,y)} \ \leq \ \widehat{Q}_\beta(x,y)$$

and

$$\widehat{Q}_\beta(x,y) \ \leq \ (1+\epsilon(\beta)) \ (1+h_1 \ e^{-\beta h_2}) \ q(x,y) \ e^{-\beta V(x,y)} \tag{14}$$

As a parenthesis if we choose $q(x,y) > 0$, after some elementary computations, then we find that

$$\left| \frac{d \log \widehat{Q}_\beta(x,y)}{d\beta} \right| \ \leq \ \sup_{v \in \mathcal{V}(x,y)} |\frac{d \log \widehat{q}_\beta}{d\beta}(x,v,y)| + \sup_{v \in \mathcal{V}(x,y)} \widehat{V}(x,v,y)$$

The inequality (14) shows that the transition probability kernels

$$\{\widehat{Q}_\beta \ ; \ \beta > 0\}$$

are of the general form of generalized simulated annealing models studied in [54] and [21].
In [54] the author studies the asymptotic behavior of such chain using large deviation techniques and in [21] the authors propose an alternative approach based on semi-group techniques. Both approaches give a precise study of the convergence of the time-inhomogeneous Markov process controlled by a suitably chosen cooling schedule and associated to the family of Markov transitions $\widehat{Q}_\beta$ of the form (11) when $\mathcal{V}$ is an auxiliary finite set.

The first method in [54] is developed for discrete time models whereas the convergence analysis in [21] is centered around continuous time models. There is a vast literature on discrete time simulated annealing (see for instance [54] and references therein). For this reason we have chosen to give a more detailed description of the second approach.

It is now convenient to introduce some additional notations. In discrete time or continuous time settings the asymptotic behavior of the desired time-inhomogeneous Markov processes will be strongly related to the virtual energy function $W : E^N \to \overline{\mathbb{R}}_+$ defined as follows

$$W(x) = \min_{g \in G(x)} \sum_{(y \to z) \in g} V(y,z) - \min_{x' \in E^N} \min_{g \in G(x')} \sum_{(y \to z) \in g} V(y,z) \quad (15)$$

where $G(x)$ is the set of x-graphs over E^N (we recall that an x-graph is an oriented tree over the vertice set E^N such that for any $x \neq y$ there exists a unique path in the x-graph leading from x to y. See also [5] or [32] for more details), and $V : E^N \times E^N \to \widehat{\mathbb{R}}_+$ is the virtual communication cost function given by

$$V(x,y) = \min \{ \widehat{V}(x,v,y) \ ; \ v \in \mathcal{V} \ \ \widehat{q}(x,v,y) > 0 \}$$

We will also use the notation

$$W^\star = \{ x \in E^N \ : \ W(x) = \min_E W \}$$

As announced the first approach presented in [54] gives a complete answer for the convergence in discrete time settings and in the time-inhomogeneous case when the parameter $\beta(n)$ is an increasing function of the time parameter n. With some obvious abusive notations let us denote by $\{\widehat{\xi}_n \ ; \ n \geq 0\}$ the discrete time and time-inhomogeneous Markov chain starting at some point $x \in E^N$ and associated to the collection of time-inhomogeneous transitions $\{\widehat{Q}_{\beta(n)} \ ; \ n \geq 1\}$.

Theorem 2.1 ([54]) *There exists a constant C_0 (which can be explicitly described in terms of V) such that if $\beta(n)$ takes the parametric form $\beta(n) = \frac{1}{C} \log n$ for sufficiently large n and $C > C_0$ then*

$$\lim_{n \to \infty} \mathbb{P} \left(\widehat{\xi}_n \in W^\star \right) = 1$$

The semi-group approach presented in [21] is based on log-Sobolev inequalities and on the notion of relative entropy. We recall that the relative entropy $\mathrm{Ent}_\pi(\mu)$ of a measure μ with respect to a measure π (charging all the points) is defined by

$$\mathrm{Ent}_\pi(\mu) = \sum_{x \in E} \mu(x) \ \log \left(\mu(x) / \pi(x) \right)$$

In contrast to the latter the former approach is based entirely on considerations of the time continuous semi-group associated to the Markov kernels $\widehat{Q}_\beta$, $\beta > 0$. Namely define, for $f : E^N \to \mathbb{R}$

$$L_\beta(f)(x) = \sum_{y \in E^N} \left(f(y) - f(x) \right) \widehat{Q}_\beta(x,y)$$

Instead of the discrete time model introduced above we are now concerned with the continuous time Markov process defined as follows. For a probability measure μ on E, an inverse-freezing schedule $\beta \in C^1(\mathbb{R}_+, \mathbb{R}_+)$ we slight abuse notations and we write $\{\widehat{\xi}_t \ ; \ t \in \mathbb{R}_+\}$ the canonical process associated to the family of generators

$$(L_{\beta(t)})_{t\geq 0} = (\widehat{Q}_{\beta(t)} - I)_{t\geq 0}$$

and whose initial condition is $\mu_0 = \mu$. We also write $\mu(t)$ the distribution of $\widehat{\xi}_t$.
Before we turn to the long time behavior of ξ_t we first give a more tractable description of this process. Let $\Delta = \{\Delta_k \ ; \ k \geq 0\}$ be independent and exponentially distributed random variables with parameter 1 and, given Δ, let $\widehat{\zeta} = \{\widehat{\zeta}_n \ ; \ n \geq 0\}$ be a time inhomogeneous Markov chain on E^N with initial distribution μ and time inhomogeneous transition probability kernels

$$\widehat{K}_n \stackrel{\text{def.}}{=} \widehat{Q}_{\beta(T_n)}, \qquad n \geq 1$$

where

$$T_n \stackrel{\text{def.}}{=} \sum_{k=0}^{n} \Delta_k, \qquad n \geq 0$$

Then

$$\widehat{\xi}_t = \begin{cases} \widehat{\zeta}_0 & 0 \leq t < T_0 \\ \widehat{\zeta}_n & T_{n-1} \leq t < T_n \end{cases}$$

defines a time in-homogeneous Markov process $\widehat{\xi} = \{\widehat{\xi}_t \ ; \ t \in \mathbb{R}_+\}$ with initial law μ and infinitesimal generators $\{L_{\beta(t)} \ ; \ t \in \mathbb{R}_+\}$.
Whenever $\widehat{\xi}$ is time homogeneous (i.e. $\beta(t) = \beta$) it is well known that L_β has a unique invariant probability measure π_β so that

$$\forall f \in \mathcal{B}_b(E^N) \qquad \pi_\beta(L_\beta(f)) = 0$$

and π_β charges all the points. Asymptotically the behavior of the invariant measure π_β as $\beta \to \infty$ depends principally on the virtual energy function W defined in (15). To be more precise we recall that π_β can be written as follows

$$\pi_\beta(x) = \frac{\widehat{R}_\beta(x)}{\sum_{z\in E^N} \widehat{R}_\beta(z)} \quad \text{where} \quad \widehat{R}_\beta(x) = \sum_{g\in G(x)} \prod_{(y\to z)\in g} \widehat{Q}_\beta(y,z)$$

Now, from the inequality (14) one concludes that

$$\epsilon_1(\beta)\ R_\beta(x) \ \leq \ \widehat{R}_\beta(x) \leq \ \epsilon_2(\beta)\ R_\beta(x)$$

where $\epsilon_i(\beta)$, $i = 1, 2$, are some functions such that

$$\lim_{\beta\to\infty} \epsilon_i(\beta) = 1, \qquad i = 1, 2$$

and

$$R_\beta(x) = \sum_{g\in G(x)} \prod_{(y\to z)\in g} q(y,z)\ e^{-\beta V(y,z)}$$

This can also be rewritten in the form

$$R_\beta(x) = \sum_{g\in G(x)} q(g)\ e^{-\beta V(g)}$$

with

$$q(g) = \prod_{(y\to z)\in g} q(y,z) \quad \text{and} \quad V(g) = \sum_{(y\to z)\in g} V(y,z)$$

Therefore we clearly have the estimate

$$\begin{aligned}
\lim_{\beta\to\infty} -\frac{1}{\beta}\log \pi_\beta(x) &= \lim_{\beta\to\infty} -\frac{1}{\beta}\log R_\beta(x) - \lim_{\beta\to\infty} -\frac{1}{\beta}\log \sum_{z\in E^N} R_\beta(z) \\
&= \min_{g\in G(x)} V(g) - \min_{z\in E^N} \min_{g\in G(z)} V(g) \\
&= W(x)
\end{aligned}$$

Due to this estimate, for each $\beta > 0$, if $\{\widehat{\xi}_{\beta,t}\ ;\ t \geq 0\}$ denotes the time homogeneous Markov process associated to L_β then we have that

$$\lim_{\beta\to\infty} \lim_{t\to\infty} \mathbb{P}\left(\widehat{\xi}_{\beta,t} \in W^\star\right) = 1$$

In the time-inhomogeneous situation the convergence of the algorithm to $W^\star$ is guaranteed by the following result.

Theorem 2.2 ([21]) *Let $\{\widehat{Q}_\beta\ ;\ \beta > 0\}$ be a collection of general Markov kernels of the form*

$$\widehat{Q}_\beta(x,y) = \sum_{v\in\mathcal{V}} \widehat{q}_\beta(x,v,y)\ e^{-\beta\widehat{V}(x,v,y)}$$

*where $\mathcal{V}$ is a given finite set, $\widehat{V} : E^N \times \mathcal{V} \times E^N \to \mathbb{R}_+$ and $\widehat{q}_\beta : E^N \times U \times E^N \to \mathbb{R}_+$, $\beta \in \mathbb{R}_+$, is a family of functions satisfying condition (**H**).*
There exist a constant C_0 (which can be explicitly described in terms of V) such that if $\beta(t)$ takes the parametric form $\beta(t) = \frac{1}{C}\log t$ for sufficiently large t and $C > C_0$ then

$$\lim_{t\to\infty} \mathrm{Ent}_{\pi_{\beta(t)}}(\mu(t)) = 0 \qquad \text{and} \qquad \lim_{t\to\infty} \mathbb{P}\left(\widehat{\xi}_t \in W^\star\right) = 1$$

This theorem is quite general and it will be used to study the convergence of genetic algorithm when the corresponding transitions have the form (11) or (12). It is also powerful enough to allow one to treat the classical simulated annealing algorithm. In this situation $N = 1$ and $\widehat{Q}_\beta$ takes the form

$$\widehat{Q}_\beta(x, y) = q(x, y)\ e^{-\beta\, V(x,y)}$$

with

$$V(x, y) = \max\left(U(y) - U(x), 0\right)$$

for $x \neq y$, where q is an irreducible transition probability kernel on E and $U : E \to \mathbb{R}_+$. In the special case where q is symmetric (that is $q(x, y) = q(y, x)$) it is also well known that the corresponding virtual energy function $W = U$ and the previous theorem implies that converges to the desired subset of the global minima $U^\star$. For genetic type algorithms the virtual energy function depends on the function U in a more subtle way and we need to work harder to check that $W^\star$ is contained into the desired subset of global minima.

The results developed here are in fact a particular form of those in [21] which also apply to study the convergence of generalized simulated annealing with random and time inhomogeneous communication cost functions. Although this subject is tangential to the main object of this article let us discuss how these results may be useful in solving mean cost optimization problems.
In some practical problems the object is to find the global minima of a function $U : E \to \mathbb{R}_+$ defined by

$$U(x) = \mathbb{E}\left(\mathcal{U}(x, Z)\right) = \int_F \mathcal{U}(x, z)\ \nu(dz)$$

where Z is a random variable taking values in a finite set F with distribution ν and $\mathcal{U} : E \times F \to \mathbb{R}_+$. The essential problem is to compute at each time step the mean cost function U and the huge size of the set F often precludes the use of the previous stochastic algorithms.
To solve this problem an additional level of approximation is needed. The natural idea proposed in [21] consists in replacing at each moment of time in the description of the stochastic algorithm the function U by the time inhomogeneous and random function

$$U_t(x) \stackrel{\text{def.}}{=} \frac{1}{t^A} \int_0^{t^A} \mathcal{U}(x, Z_s)\ ds$$

where $A > 0$ and $\{Z_t\ ;\ t \geq 0\}$ is a given time homogeneous Markov process associated to the generator $\mathcal{G} = \mathcal{K} - Id$ where $\mathcal{K}$ is an irreducible

transition probability kernel on E with invariant measure ν. A full discussion of the convergence of the resulting stochastic algorithm to the desired subset $U^\star$ is outside the scope of this work, the interested reader is referred to [21].

3 Feynman-Kac and Non Linear Filtering Models

3.1 Description of the Models

The nonlinear filtering problem consists in computing the conditional distribution 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. In discrete time settings the state signal $X = \{X_n \ ; \ n \geq 0\}$ is a discrete time Markov chain taking values in a Polish space E (i.e. a complete separable metric space) with transition probabilities $\{K_n \ ; \ n \geq 1\}$ and initial distribution η_0. The observation sequence $Y = \{Y_n \ ; \ n \geq 0\}$ are $\mathbb{R}^q$-valued random variables and take the form

$$Y_n = H_n(X_n, V_n)$$

where the V_n are independent and q-dimensional variables, independent of X and with a law having a known density, and H_n is a measurable function from $E \times \mathbb{R}^q$ into $\mathbb{R}^q$. For any $x \in E$ we assume that the variable $Y_n = H_n(x, V_n)$ admits a positive density $y \mapsto \varphi_n(x, y)$ and the function φ_n is bounded. To clarify the notations we fix the observations $Y_n = y_n$, $n \geq 0$ and we write

$$g_n(x) \stackrel{\text{def.}}{=} \varphi_n(x, y_n)$$

For somewhat technical reasons we will assume that H_n and φ_n and the observation sequence $\{y_n \ ; \ n \geq 0\}$ are chosen so that g_n is a positive and bounded function on E. These assumptions can be relaxed considerably, a more complete and general set of assumptions is formulated in [17] and [18].
Given the stochastic nature of the pair signal/observation process the nonlinear filtering problem consists in computing recursively in time the one step predictor conditional probabilities η_n and the filter conditional distributions $\widehat{\eta}_n$ given for any bounded Borel test function f by

$$\begin{aligned} \eta_n(f) &= \mathbb{E}(f(X_n)|Y_0 = y_0, \ldots, Y_{n-1} = y_{n-1}) \\ \widehat{\eta}_n(f) &= \mathbb{E}(f(X_n)|Y_0 = y_0, \ldots, Y_{n-1} = y_{n-1}, Y_n = y_n) \end{aligned}$$

As usually the n-step filter $\widehat{\eta}_n$ is written in terms of η_n as

$$\widehat{\eta}_n(f) = \Psi_n(\eta_n)(f) = \frac{\eta_n\,(f\ g_n)}{\eta_n\,(g_n)} \tag{16}$$

and the n-step predictor is defined in terms of the Feynman-Kac type formula

$$\eta_n(f) = \frac{\gamma_n(f)}{\gamma_n(1)} \qquad \text{with} \quad \gamma_n(f) = \mathbb{E}\left(\ f(X_n)\ \prod_{k=0}^{n-1} g_k(X_k)\ \right) \tag{17}$$

with the convention $\prod_{\emptyset} = 1$. By (16), the n-step filter $\widehat{\eta}_n$ may also be expressed as the ration

$$\widehat{\eta}_n(f) = \frac{\widehat{\gamma}_n(f)}{\widehat{\gamma}_n(1)} \tag{18}$$

with

$$\widehat{\gamma}_n(f) = \gamma_n(g_n f) = \mathbb{E}\left(\ f(X_n)\ \prod_{k=0}^{n} g_k(X_k)\ \right)$$

It is also not difficult to check that γ_n and $\widehat{\gamma}_n$ are connected by

$$\gamma_n = \widehat{\gamma}_{n-1} K_n, \qquad n \geq 1 \tag{19}$$

It then follows from the relations (16) and (19) that for any $n \geq 1$

$$\eta_n = \Phi_n\left(\eta_{n-1}\right), \tag{20}$$

with

$$\Phi_n(\eta) = \Psi_{n-1}(\eta) K_n.$$

Another interest feature of the genetic algorithm defined by (3) is that it can be used to approximate the Feynman-Kac formulas $\gamma_n(f)$ and $\widehat{\gamma}_n(f)$ defined in (17) and (18). One of the best ways for introducing the corresponding particle approximating models is through the following observation. By definition it is easy to establish that for any $n \geq 0$

$$\eta_n(g_n) = \frac{\gamma_n(g_n)}{\gamma_n(1)} = \frac{\gamma_{n+1}(1)}{\gamma_n(1)}$$

This yields that

$$\gamma_n(1) = \prod_{p=0}^{n-1} \eta_p(g_p) \qquad \text{and} \qquad \gamma_n(f) = \eta_n(f) \prod_{p=0}^{n-1} \eta_p(g_p)$$

with the usual convention $\prod_{\emptyset} = 1$. Taking into consideration these relations we define a natural N-approximating measure γ_n^N for γ_n by setting

$$\gamma_n^N(1) = \prod_{p=0}^{n-1} \eta_p^N(g_p) \qquad \text{and} \qquad \gamma_n^N(f) = \eta_n^N(f)\ \gamma_n^N(1) \tag{21}$$

In view of (7) and using the same line of ideas we can defined the corresponding N-approximating measures of $\widehat{\gamma}_n$ and $\widehat{\eta}_n$. We have chosen here to restrict our attention to the distributions γ_n and η_n.

3.2 Asymptotic Behavior

One of the simplest way for studying the asymptotic behavior as $N \to \infty$ of the genetic algorithm presented in section 2 is through the analysis of the un-normalized distributions $\{\gamma_n \; ; \; n \geq 0\}$. This approach has been initiated in [26] and it has been further developed in [19] and [20]. Here we follow line by line the synthetic presentation given in [19]. This approach is based on the observation that the dynamics structure of the latter is linear and one might expect that the analysis of the corresponding approximating measures will be simplified. In view of (17) and (18) we have that

$$\forall 0 \leq p \leq n, \qquad \gamma_n = \gamma_p \, L_{p,n}, \qquad (\gamma_0 = \eta_0) \tag{22}$$

where $\{L_{p,n} \; ; \; 0 \leq p \leq n\}$ is the time inhomogeneous semi-group defined by the relations

$$L_{p,n} = L_{p+1} L_{p+2} \ldots L_n \qquad \text{with} \qquad L_n(f) = g_{n-1} . K_n(f)$$

and the convention $L_{n,n} = Id$. Using these notations one can also check that the one step mappings Φ_n can be rewritten as

$$\Phi_n(\eta)(f) = \frac{\eta(L_n(f))}{\eta(L_n(1))}$$

for any $\eta \in \mathbf{M}_1(E)$ and $f \in \mathcal{B}_b(E)$. Using these notations we notice that for any $n \geq 0$ and $f \in \mathcal{B}_b(E)$ the stochastic process

$$\{M_q^N(f) \; ; \; 0 \leq q \leq n\}$$

defined as

$$\begin{aligned} M_q^N(f) &\stackrel{\text{def.}}{=} \gamma_q^N(L_{q,n} f) - \gamma_q(L_{q,n} f) \\ &= \sum_{p=0}^{q} \left(\gamma_p^N(L_{p,n} f) - \gamma_{p-1}^N(L_p L_{p,n} f) \right) \\ &= \sum_{p=0}^{q} \gamma_p^N(1) \; \left(\eta_p^N(L_{p,n} f) - \Phi_p\left(\eta_{p-1}^N\right)(L_{p,n} f) \right), \end{aligned} \tag{23}$$

with the convention $\Phi_0\left(\eta_{-1}^N\right) = \eta_0$, is a martingale with respect to the natural filtration $F^N = \{F_n^N \; ; \; n \geq 0\}$ associated with the N-particle

system $\{\xi_n \ ; \ n \geq 0\}$ and its angle bracket is given by

$$\langle M^N(f)\rangle_q = \frac{1}{N}\sum_{p=0}^{q} \ \left(\gamma_p^N(1)\right)^2 \ \Phi_p(\eta_{p-1}^N)\left(\ \left(L_{p,n}f - \Phi_p(\eta_{p-1}^N)L_{p,n}f\right)^2\right) \tag{24}$$

One concludes easily that γ_n^N is an approximating measure of γ_n without any bias, that is for any bounded Borel test function f

$$\mathbb{E}\left(\gamma_n^N(f)\right) = \gamma_n(f) \tag{25}$$

and

$$\mathbb{E}\left(\left(\gamma_n^N(f) - \gamma_n(f)\right)^2\right)$$

$$= \frac{1}{N}\sum_{p=0}^{n} \mathbb{E}\left(\left(\gamma_p^N(1)\right)^2 \ \Phi_p(\eta_{p-1}^N)\left(\ \left(L_{p,n}f - \Phi_p(\eta_{p-1}^N)L_{p,n}f\right)^2\right)\right) \tag{26}$$

Under our assumptions it is also clear that there exist some finite constants $C(n) < \infty$ such that

$$\mathbb{E}\left(\left(\gamma_n^N(f) - \gamma_n(f)\right)^2\right)^{1/2} \leq \frac{C(n)}{\sqrt{N}} \ \|f\|$$

Exponential bounds can also be obtained using the decomposition (23). For instance, by definition of η_t^N, Hoeffding's inequality implies that for each $0 \leq p \leq n$ and for any $\epsilon > 0$

$$\mathbb{P}\left(\left|\eta_p^N(L_{p,n}f) - \Phi_p\left(\eta_{p-1}^N\right)(L_{p,n}f)\right| > \epsilon \left|\eta_{n-1}^N\right.\right) \leq 2 \ e^{-\frac{N}{8}\frac{\epsilon^2}{\|L_{p,n}f\|^2}}$$

¿From which one concludes that

$$\mathbb{P}\left(\sup_{0\leq p\leq n}\left|\eta_p^N(L_{p,n}f) - \Phi_p\left(\eta_{p-1}^N\right)(L_{p,n}f)\right| > \epsilon\right) \leq 2 \ \sum_{p=0}^{n} \ e^{-\frac{N}{8}\frac{\epsilon^2}{\|L_{p,n}f\|^2}}$$

Since the fitness functions are assumed to be bounded this exponential bound implies that there exists some finite constants $C_1(n)$ and $C_2(n)$ such that for any bounded Borel function f, $\|f\| \leq 1$ and for every $\epsilon > 0$ we have that

$$\mathbb{P}\left(\sup_{0\leq p\leq n}\left|\gamma_p^N(L_{p,n}f) - \gamma_p(L_{p,n}f)\right| > \epsilon\right) \leq C_1(n) \ \exp -\frac{N\epsilon^2}{C_2(n)}$$

We now give a brief indication of how these results can be used to obtain useful estimates for the N-approximating measures η_n^N and $\widehat{\eta}_n^N$.

¿From the previous displayed exponential rate one can also prove that there exists some finite constants $C_1(n), C_2(n)$ such that for any $\epsilon > 0$ and for any bounded Borel test function f, $\|f\| \leq 1$

$$\mathbb{P}\left(\left|\eta_n^N(f) - \eta_n(f)\right| > \epsilon\right) \leq C_1(n)\ \exp -\frac{N\epsilon^2}{C_2(n)} \tag{27}$$

Precise estimates of these exponential rates are studied in [24] using large deviations techniques. The previous exponential rates also imply $\mathbb{L}^p$ mean errors

$$\forall p \geq 1 \qquad \mathbb{E}\left(\left|\eta_n^N(f) - \eta_n(f)\right|^p\right)^{1/p} \leq \frac{C(p,n)}{\sqrt{N}}\,\|f\|$$

for some constant $C(p,n) < \infty$ which only depends on the parameters p and n. With some little work one can use (25) and (26) to prove that there exists some finite constants $C(n)$ such that for any bounded Borel function f such that $\|f\| \leq 1$

$$\left|\mathbb{E}\left(\eta_n^N(f)\right) - \eta_n(f)\right| \leq \frac{C(n)}{N} \tag{28}$$

Taking into consideration this inequality, by the exchangeability of the particles and the definition of the total variation distance of probability measures one can check that for each $1 \leq i \leq N$

$$\|\mathcal{L}aw(\xi_n^i) - \eta_n\|_{\mathrm{tv}} \leq \frac{C(n)}{N} \tag{29}$$

The precise magnitude of variability of these mean errors is given by central limit theorems. A full discussion on these fluctuations would be too great digression here but as the form of the angle bracket (24) indicates one can prove that the sequence of random fields

$$U_n^N(f) \stackrel{\text{def.}}{=} \sqrt{N}\ \left(\gamma_n^N(f) - \gamma_n(f)\right), \qquad f \in \mathcal{B}_b(E)$$

converges in law as $N \to \infty$ to a centered Gaussian field

$$\{U_n(f)\ ;\ f \in \mathcal{B}_b(E)\}$$

satisfying

$$\mathbb{E}\left(U_n(f)^2\right) = \sum_{p=0}^{n}\ (\gamma_p(1))^2\ \eta_p\left(\ (L_{p,n}f - \eta_p L_{p,n}f)^2\right)$$

for any $f \in \mathcal{B}_b(E)$ (in the sense of convergence of finite dimensional distributions). The previous fluctuations imply that the sequence of random fields

$$W_n^N(f) \stackrel{\text{def.}}{=} \sqrt{N}\ \left(\eta_n^N(f) - \eta_n(f)\right), \qquad f \in \mathcal{B}_b(E)$$

converges in law as $N \to \infty$ to the centered Gaussian field

$$W_n(f) \stackrel{\text{def.}}{=} U_n\left(\frac{1}{\gamma_n(1)}\left(f - \eta_n(f)\right)\right), \qquad f \in \mathcal{B}_b(E)$$

We conclude this section with some comments on the long time behavior of the N-interacting particle system approximating models. If the measure valued dynamical system (20) is sufficiently stable in the sense that it forgets any erroneous initial condition, then one can prove uniform convergence results with respect to the time parameter (see for instance [20, 22] and [25] and references therein). For instance, with some suitable stability properties for the Markov kernels $\{K_n\ ;\ n \geq 1\}$ one can find some coefficient $\alpha \in (0, 1/2)$ such that for any bounded Borel test function f, $\|f\| \leq 1$,

$$\forall p \geq 1 \qquad \sup_{n \geq 0} \mathbb{E}\left(\left|\eta_n^N(f) - \eta_n(f)\right|^p\right)^{1/p} \leq \frac{c(p)}{N^\alpha}$$

for some constant $c(p) < \infty$ which only depends on the parameter p. This uniform convergence result with respect to the time parameter leads us to hope that maybe we can construct an asymptotic method to study the convergence of genetic algorithms in numerical function optimization in a more general settings than the one treated in section 2.2 and in the next section.

4 Numerical Function Analysis

4.1 Description of the Models

The objective of this section is to formulate more precisely the mutation and selection transitions (10) so that the resulting empirical measures of the genetic algorithm presented in section 2.2 will concentrate in probability, as the time parameter tends to infinity, on the set $U^\star$ of the global minima of a given numerical function $U : E \to \mathbb{R}_+$.

As in section 2.2 we assume that E is a finite state space and

$$\beta : \mathbb{N} \to \mathbb{R}_+$$

is a inverse cooling schedule. Let $a : E \times E \to \mathbb{R}_+$ be a numerical function which induces an equivalence relation on E defined by

$$x \sim y \Longleftrightarrow a(x, y) = 0$$

This leads us naturally to consider the partition

$$S_1, \ldots, S_{n(a)}, \qquad n(a) \geq 1,$$

induced by $\sim$.
If x is a typical element of E then the equivalence class of x will be denoted by $S(x)$

$$S(x) = \{y \in E \ : \ x \sim y\}$$

We further require that

$$a(x, y) = 0 \Longrightarrow U(x) = U(y)$$

A trivial example of equivalence relation satisfying this condition is given by the following function a

$$a(x, y) = a_0 \ (1 - 1_x(y)), \qquad a_0 > 0$$

In this case we clearly have $a(x, y) = 0 \Longleftrightarrow x = y$.
The mutation kernels K_n and the fitness functions g_n are related to $\beta(n)$ as

$$g_n(x) = e^{-\beta(n)\, U(x)} \qquad \text{and} \qquad K_n(x, y) = k_{\beta(n)}(x, y)$$

with for any $\beta > 0$

$$k_\beta(x, y) = \begin{cases} k(x, y)\, e^{-\beta\, a(x,y)} & \text{if} \quad a(x, y) > 0 \\ \frac{1}{|S(x)|} \left(1 - \sum_{z \notin S(x)} \ k(x, z)\, e^{-\beta\, a(x,z)}\right) & \text{otherwise} \end{cases}$$

where $k : E \times E \to \mathbb{R}_+$ is an irreductible Markov kernel, that is for any $x \in E$

$$\sum_{y \in E} k(x,y) = 1$$

and for any $(x,y) \in E \times E$ there exists a sequence $x_0, x_1, \ldots, x_r \in E$, $r \geq 1$ such that

$$x_0 = x, \qquad k(x_k, x_{k+1}) > 0 \qquad (\forall 0 \leq k < r), \qquad x_r = y$$

We now describe a general construction which allows us to find the asymptotics of the desired transition kernels

$$\widehat{Q}_\beta \stackrel{\text{def.}}{=} Q_\beta^{(1)} Q_\beta^{(2)} \tag{30}$$

where

$$Q_\beta^{(1)}(x,y) = \prod_{p=1}^{N} k_\beta(x^p, y^p)$$

and

$$Q_\beta^{(2)}(x,y) = \prod_{p=1}^{N} \sum_{i=1}^{N} \frac{e^{-\beta\, U(x^i)}}{\sum_{j=1}^{N} e^{-\beta\, U(x^j)}}\ 1_{x^i}(y^p)$$

It can be directly checked that

$$\begin{aligned}
& Q_\beta^{(1)}(x,y) \\
&= \left(\prod_{p:a(x^p,y^p)=0} k_\beta(x^p,y^p)\right) \\
&\qquad \times \left(\prod_{p:a(x^p,y^p)>0} k(x^p,y^p)\right)\ e^{-\beta \sum_{p=1}^{N} a(x^p,y^p)} \\
&= \theta_\beta^{(1)}(x,y) \quad q^{(1)}(x,y) \quad e^{-\beta\, V^{(1)}(x,y)}
\end{aligned}$$

with

$$\begin{aligned}
\theta_\beta^{(1)}(x,y) &= \prod_{p:a(x^p,y^p)=0} k_\beta(x^p,y^p)|S(x^p)| \\
q^{(1)}(x,y) &= \left(\prod_{p:a(x^p,y^p)>0} k(x^p,y^p)\right)\left(\prod_{p:a(x^p,y^p)=0} |S(x^p)|^{-1}\right)
\end{aligned}$$

and

$$V^{(1)}(x,y) = \sum_{p=1}^{N} a(x^p,y^p)$$

We also notice that

$$\theta_\beta^{(1)}(x,y) \to 1 \ \text{ as } \ \beta \to \infty$$

To describe the asymptotic of $Q_\beta^{(2)}$ as $\beta \to \infty$ we need to recall some terminology introduced in [8]. We will use the superscript $f^\star$ to denote the set of global minima of a given numerical function $f : \mathcal{E} \to \overline{\mathbb{R}}$ on a given finite state space $\mathcal{E}$ so that

$$f^\star \stackrel{\text{def.}}{=} \left\{ x \in \mathcal{E} \ ; \ f(x) = \min_{\mathcal{E}} f \right\}$$

The cardinality of a finite set $\mathcal{E}$ will be denoted by $|\mathcal{E}|$ and if x and y belongs to E^N and $z \in E$ we write

$$x(z) = |\{p \ : \ 1 \leq p \leq N \ , \ x_p = z\}|$$

and

$$\widehat{x} = \{p \ : \ 1 \leq p \leq N \ , \ U(x_p) = \widehat{U}(x)\} \quad \text{and} \quad \widehat{U}(x) = \min_{1 \leq p \leq N} U(x_p)$$

A similar discussion to that above leads to the decomposition

$$\begin{aligned} Q_\beta^{(2)}(x,y) &= \prod_{p=1}^N \sum_{i : x_i = y_p} \frac{e^{-\beta \, U(x^i)}}{\sum_{j=1}^N e^{-\beta \, U(x^j)}} \\ &= \prod_{p=1}^N \frac{x(y_p)}{|\widehat{x}|} \, \frac{e^{-\beta \left(U(y^p) - \widehat{U}(x) \right)}}{1 + |\widehat{x}|^{-1} \sum_{j \notin \widehat{x}} e^{-\beta \left(U(x^j) - \widehat{U}(x) \right)}} \\ &= \theta_\beta^{(2)}(x,y) \quad q^{(2)}(x,y) \quad e^{-\beta \, V^{(2)}(x,y)} \end{aligned} \tag{31}$$

with

$$\begin{aligned} \theta_\beta^{(2)}(x,y) &= [1 + |\widehat{x}|^{-1} \sum_{j \notin \widehat{x}} e^{-\beta \left(U(x^j) - \widehat{U}(x) \right)}]^{-N} \\ q^{(2)}(x,y) &= \prod_{p=1}^N \frac{x(y_p)}{|\widehat{x}|} \qquad \text{and} \qquad V^{(2)}(x,y) = \sum_{p=1}^N \left(U(y^p) - \widehat{U}(x) \right) \end{aligned}$$

As before we also notice that

$$\theta_\beta^{(2)}(x,y) \to 1 \ \text{ as } \ \beta \to \infty$$

If we combine (4.1) and (31) one concludes that the transition (30) has the same form as in (11), namely

$$\widehat{Q}_\beta(x,z) = \sum_{y \in E^N} \widehat{q}_\beta(x,y,z) \, e^{-\beta \, \widehat{V}(x,y,z)}$$

with

$$\begin{aligned} \widehat{q}_\beta(x,y,z) &= \widehat{q}(x,y,z) \, \theta_\beta(x,y,z), \ \widehat{V}(x,y,z) = V^{(1)}(x,y) + V^{(2)}(y,z) \\ \theta_\beta(x,y,z) &= \theta_\beta^{(1)}(x,y) \, \theta_\beta^{(2)}(y,z), \quad \widehat{q}(x,y,z) = q^{(1)}(x,y) \, q^{(2)}(y,z) \end{aligned}$$

Using the fact that $q^{(1)}$ is irreducible, $q^{(2)}(x,x) > 0$ and using the form of $\theta_\beta^{(1)}$, $\theta_\beta^{(2)}$ one can also check that the assumption (**H**) introduced on page 11 is satisfied and therefore theorem 2.2 applies to our situation with

$$W(x) = \min_{g \in G(x)} \sum_{(y \to z) \in g} V(y,z) - \min_{x' \in E^N} \min_{g \in G(x')} \sum_{(y \to z) \in g} V(y,z) \quad (32)$$

and

$$V(x,z) = \min \Big\{ V^{(1)}(x,y) + V^{(2)}(y,z) \; ; \; q^{(1)}(x,y) q^{(2)}(y,z) > 0 \Big\} \quad (33)$$

Furthermore we proved in [21] that there exists a critical population size $N(a,U)$ depending on the function U and on the equivalence relation a such that

$$N \geq N(a,U) \quad \Longrightarrow \quad W^\star \subset \mathbf{U}^\star \cap \mathbf{A}$$

where

$$\mathbf{A} \stackrel{\text{def.}}{=} \{ x \in E^N \; : \; x_i \sim x_j \quad \forall 1 \leq i,j \leq N \}$$

and

$$\mathbf{U}^\star \stackrel{\text{def.}}{=} \{ x \in E^N \; : \; \widehat{U}(x) = \min_E U \}$$

4.2 A Test Set Method

To be more precise about this critical population size we need to investigate more closely the properties of the virtual energy function W. We now describe a natural test set approach to study the set its global minima. This approach is based on the following concept of λ-stability

Definition 4.1 *Let λ be a non negative real number.*
A subset $H \subset E^N$ is called λ-stable with respect to a communication cost function V when the following conditions are satisfied:

1. $\forall x \in H \;\; \forall y \notin H \qquad V(x,y) > \lambda$
2. $\forall x \notin H \;\; \exists y \in H \qquad V(x,y) \leq \lambda$

The importance of the notion of λ-stability resides in the following result which extends lemma 4.1 of Freidlin-Wentzell [32].

Proposition 4.2 ([21])
Let λ be a non negative real number and $H \subset E^N$. Any λ-stable subset H with respect to V contains $W^\star$

One remark is that the subset **A** is 0-stable with respect to the communication cost function V defined in (33). From this observation one concludes that the canonical process $\{\widehat{\xi}_t \; ; \; t \geq 0\}$ associated to the

family of generators $\{L_{\beta_t} = \widehat{Q}_{\beta_t} - Id\ ;\ t \geq 0\}$, converges as $t \to \infty$ in probability to the set $\mathbf{A}$ and

$$\min_{x \in \mathbf{A}} W(x) = \min_{x \in E^N} W(x) = 0 \tag{34}$$

Using classical arguments (34) implies that for any $x \in \mathbf{A}$

$$\begin{array}{l} W(x) = W_{\mathbf{A}}(x) \\ \stackrel{\text{def.}}{=} \min\limits_{g \in G_{\mathbf{A}}(x)} \sum\limits_{(y \to z) \in g} V_{\mathbf{A}}(y,z) - \min\limits_{x' \in \mathbf{A}} \min\limits_{g \in G_{\mathbf{A}}(x')} \sum\limits_{(y \to z) \in g} V_{\mathbf{A}}(y,z) \end{array}$$

where $G_{\mathbf{A}}(x)$ is the set of x-graphs over $\mathbf{A}$ (here the starting and end points of the x-graphs are in $\mathbf{A}$) and $V_{\mathbf{A}} : \mathbf{A} \times \mathbf{A} \to \overline{\mathbb{R}}_+$ is the *taboo* communication cost function defined by setting for any $x, y \in \mathbf{A}$

$$\begin{array}{l} V_{\mathbf{A}}(x,y) \\ = \min \left\{ \sum_{k=0}^{|p|-1} V(p_k, p_{k+1})\ ;\ p \in C_{x,y} \ \text{ with } \ \forall 0 < k < |p| \ \ p_k \notin \mathbf{A} \right\} \end{array}$$

where $C_{x,y}$ is the set of all paths $p = (p_0, \ldots, p_{|p|})$, with some length $|p|$, admissible for q (that is $q(p_k, p_{k+1}) > 0$ for each $0 \leq k < |p|$) leading from x to y (that is $p_0 = x$ and $p_{|p|} = y$).
Let $\mathcal{A} = \{A_1, \ldots, A_{n(a)}\}$ be the partition of $\mathbf{A}$ induced by the partition $\mathcal{S} = \{S_1, \ldots, S_{n(a)}\}$ of E associated to the relation $\sim$

$$\forall 1 \leq i \leq n(a) \qquad A_i \stackrel{\text{def.}}{=} \mathbf{A} \cap S_i^N = S_i^N$$

with

$$S_i^N \stackrel{\text{def.}}{=} \underbrace{S_i \times \ldots \times S_i}_{\text{N times}}$$

We observe that for any $1 \leq i \leq n(a)$ and $x, y \in A_i$, $V(x,y) = 0$. Using this observation one can prove that for any $x \in \mathbf{A}$

$$\begin{array}{rl} W_{\mathbf{A}}(x) = W_{\mathcal{A}}(x) \stackrel{\text{def.}}{=} & \min\limits_{g \in G_{\mathbf{A}}(x)} \sum\limits_{(y \to z) \in g} V_{\mathcal{A}}(y,z) \\ & - \min\limits_{x' \in \mathbf{A}} \min\limits_{g \in G_{\mathbf{A}}(x')} \sum\limits_{(y \to z) \in g} V_{\mathcal{A}}(y,z) \end{array}$$

where $V_{\mathcal{A}}$ is the communication cost function defined by setting for any $x \in A_i$ and $y \in A_j$ and $1 \leq i, j \leq n(a)$

$$\begin{array}{rcl} V_{\mathcal{A}}(x,y) & = & \min\{ \sum\limits_{k=0}^{|p|-1} V(p_k, p_{k+1})\ :\ p \in C_{x,y}, \qquad \exists 0 \leq n_1 < n_2 \leq |p|, \\ & & \forall\ 0 \leq k \leq n_1,\ \ p_k \in A_i, \quad \forall n_1 < k < n_2,\ \ p_k \notin \mathbf{A}, \\ & & \qquad\qquad \forall n_2 \leq k \leq |p|,\ \ p_k \in A_j \} \end{array} \tag{35}$$

As it is easily seen $V_{\mathcal{A}}(x,y)$ does not depend on the choice of $x \in A_i$ and $y \in A_j$. Another remark is that

$$W^\star = W^\star_{\mathbf{A}} = W^\star_{\mathcal{A}}$$

and therefore the following implication holds for any subset $H \subset \mathbf{A}$

$$\exists \lambda \geq 0 \quad : \quad H \ \ \lambda - \text{stable w.r.t. } V_{\mathcal{A}} \Longrightarrow W^\star \subset H \tag{36}$$

In other words $\widehat{\xi_t}$ converges in probability as $t \to \infty$ to any λ-stable subset $H \subset \mathbf{A}$ with respect to $V_{\mathcal{A}}$. The technical trick now is to find a critical size $N(a,U)$ and a non negative constant $\lambda(a,U)$ such that the subset $\mathbf{U}^\star \cap \mathbf{A}$ is $\lambda(a,U)$-stable with respect to $V_{\mathcal{A}}$.
To describe precisely $N(a,U)$ and $\lambda(a,U)$ we need to introduce some additional notations. By $\Gamma_{x,y}$, $x,y \in E$, we denote the paths q in E joining x and y, that is

$$\forall 0 \leq l < |q| \qquad k(x_l, x_{l+1}) > 0 \qquad q_0 = x \quad q_{|q|} = y$$

We will also note $R(a)$ the smallest integer such that for every $x,y \in E$ in two different classes there exists a path joining x and y with length $|q| \leq R(a)$, namely

$$R(a) = \max_{1 \leq i,j \leq n(a)} \min_{(x_i,x_j) \in S_i \times S_j} \min_{q \in \Gamma_{x_i,x_j}} |q|$$

It will be also convenient to use the following definitions

$$\begin{aligned} \triangle a &= \min \{a(x,y) \ : \ x,y \in E \ \ a(x,y) \neq 0\} \\ \delta(a) &= \sup \{a(x,y) \ : \ x,y \in E\} \end{aligned}$$

and

$$\begin{aligned} \triangle U &= \min \{|U(x) - U(y)| \ : \ x,y \in E \ \ U(x) \neq U(y)\} \\ \delta(U) &= \sup \{|U(x) - U(y)| \ : \ x,y \in E\} \end{aligned}$$

To formulate precisely our convergence result we need the following lemma.

Lemma 4.3 ([21]) *For every $x \in \mathbf{A}$ there exists a state $y \in \mathbf{U}^\star \cap \mathbf{A}$ such that*

$$V_{\mathcal{A}}(x,y) \leq (\delta(a) + \delta(U))\ R(a)$$

For every $x,y \in A$ such that $\widehat{U}(x) < \widehat{U}(y)$ we have

$$V_{\mathcal{A}}(x,y) \geq \min(\Delta a, \Delta U)\ N$$

Let us write

$$\lambda(a,U) \stackrel{\text{def.}}{=} (\delta(a)+\delta(U))\ R(a)$$

and

$$N(a,f) \stackrel{\text{def.}}{=} \lambda(a,U)/\min(\Delta a, \Delta U)$$

Using the above lemma one concludes that

$$N > N(a,f) \Longrightarrow \mathbf{U}^\star \cap \mathbf{A} \ \text{ is } \ \lambda(a,U) - \text{stable with respect to } \ V_{\mathcal{A}} \tag{37}$$

If we combine (36) and (37), with theorem 2.2 one concludes that

Theorem 4.4 ([21]) *There exist a constant C_0 (which can be explicitly described in terms of V) such that if $N \geq N(a,U)$ and if $\beta(t)$ takes the parametric form $\beta(t) = \frac{1}{C}\log t$ for sufficiently large t and $C > C_0$ then*

$$\lim_{t\to\infty} \mathbb{P}\left(\widehat{\xi}_t \in \mathbf{U}^\star \cap \mathbf{A}\right) = 1$$

5 Refinements and Variants

The research literature abounds with variations of the genetic algorithm described in section 2. Each of these variants is intended to make the selection and/or the mutation more efficient in some sense. The convergence analysis of all these alternative schemes is far from being complete. We also emphasize that these variations come from different sources of inspiration. Some of them are strongly related to traditional weighted re-sampling plans in weighted bootstrap theory (see [4] and references therein). Another source of inspiration was provided by branching and interacting particle system theory. The aim of this section is to introduce the reader to these recently established connections between branching and interacting particle systems, genetic algorithms, simulated annealing and bootstrap theory.

We begin our program with an alternative genetic algorithm which transitions are obtained through choosing randomly at each step the selection or the mutation transition. This variation has been presented for the first time in [21] to improve the convergence results of the classical genetic algorithm studied in section 4.
We will use the general methodology presented in section 2.2 and the test set approach of section 4 to prove that the corresponding genetic-type algorithm converges towards the set of the global minima of a desired numerical function. These results can be found with complete proof in [21]. We will give some comments on how these results improve the one of section 4.
The second variation has been presented in [27] for solving non linear filtering problems. The main difference with the classical genetic

algorithm of section 2 lies in the fact that in the former the mutation kernels also depend on the fitness function. The corresponding mutation transition has in fact a natural interpretation in non linear filtering and it can be regarded as a conditional transition probability. By reference with non linear filtering we will call this kind of mutation a conditional mutation.
We end this section with a brief presentation of several branching genetic type algorithms. These branching strategies are strongly related to weighted bootstrap techniques [4].
There are many open problems concerning these variations such as finding a way to study the convergence in global optimization problems.

5.1 Random Selection/Mutation Transitions

The setting here is exactly as in section 2.2 and section 4 but the genetic type algorithm is now described by the transition probability kernels

$$\tilde{Q}_\beta = \alpha_1\, Q_\beta^{(1)} + \alpha_2\, Q_\beta^{(2)} \qquad \alpha_1 + \alpha_2 = 1 \quad (\alpha_1, \alpha_2 \in (0,1))$$

Returning to the definition of $Q_\beta^{(1)}$ and $Q_\beta^{(2)}$ given in (4.1) and (31) and using the same notations as in there one concludes that $\tilde{Q}_\beta$ has the same form as in (11)

$$\tilde{Q}_\beta(x,z) = \sum_{v\in\mathcal{V}} \widehat{q}_\beta(x,v,z)\; e^{-\beta\,\widehat{V}(x,v,z)}$$

with $\mathcal{V} = \{1,2\}$ and for any $v \in \mathcal{V}$

$$\begin{aligned} \widehat{q}_\beta(x,v,y) &= \theta_\beta(x,v,y)\;\widehat{q}(x,v,y) & \theta_\beta(x,v,y) &= \theta_\beta^{(v)}(x,y) \\ \widehat{q}(x,v,y) &= \alpha_v\; q^{(v)}(x,y) & \widehat{V}(x,v,y) &= V^{(v)}(x,y) \end{aligned}$$

To clarify the presentation we use the superscript $\tilde{(\cdot)}$ to denote the communication cost function $\tilde{V}$, the critical height constant $\tilde{C}_0$ arising in theorem 2.2 and the virtual energy function $\tilde{W}$ associated to the transition probability kernels $\tilde{Q}_\beta$. ¿From the above observations and theorem 2.2, choosing β of the form

$$\beta(t) = \frac{1}{C}\,\log t \quad \text{where} \quad C > \tilde{C}_0$$

for t sufficiently large, yields that the canonical process

$$(\Omega, P, (F_t)_{t\geq 0}, (\tilde{\xi}_t)_{t\geq 0})$$

associated to the family of generators

$$L_{\beta(t)} = \tilde{Q}_{\beta(t)} - Id$$

converges in probability to the set of the global minima $\tilde{W}^\star$ of the virtual energy $\tilde{V}$ associated to $\tilde{Q}_\beta$ and defined as in (32) by replacing the communication cost functions V by $\tilde{V}$ where

$$\tilde{V}(x,y) = \min\Big\{V^{(v)}(x,y)\ ;\ v \in \mathcal{V},\ \ q^{(v)}(x,y) > 0\Big\}$$

By the same test set approach we used in section 4 the technical trick here is to find a critical size $\tilde{N}(a,U)$ and a non negative constant $\tilde{\lambda}(a,U)$ such that the subset $\mathbf{U}^\star \cap \mathbf{A}$ is $\tilde{\lambda}(a,U)$-stable with respect to $\tilde{V}_{\mathcal{A}}$, where $\tilde{V}_{\mathcal{A}}$ is defined as in (35) by replacing the communication cost function V by $\tilde{V}$. In this setting the analogue of lemma 4.3 is the following

Lemma 5.1 ([21]) *For every $x, y \in \mathbf{A}$ such that $\widehat{U}(x) \geq \widehat{U}(y)$ we have*

$$\tilde{V}_{\mathcal{A}}(x,y) \leq \delta(a)\ R(a)$$

For every $x, y \in A$ such that $\widehat{U}(x) < \widehat{U}(y)$ we have

$$\tilde{V}_{\mathcal{A}}(x,y) \geq \min(\Delta a, \Delta U)\ N$$

Now, if we write

$$\tilde{\lambda}(a,U) = \delta(a)\ R(a) \qquad \text{and} \qquad \tilde{N}(a,U) = \tilde{\lambda}(a,U)/\min(\Delta a, \Delta U)$$

one concludes that

$$N > \tilde{N}(a,U) \Longrightarrow \mathbf{U}^\star \cap \mathbf{A} \ \text{ is } \ \tilde{\lambda}(a,U) - \text{stable with respect to } \ V_{\mathcal{A}}$$

Using the same line of arguments as in the end of section 4 one gets finally

Theorem 5.2 ([21]) *If $N \geq \tilde{N}(a,U)$ and if $\beta(t)$ takes the parametric form $\beta(t) = \frac{1}{C}\log t$ for sufficiently large t and $C > \tilde{C}_0$ then*

$$\lim_{t\to\infty} \mathbb{P}\left(\tilde{\xi}_t \in \mathbf{U}^\star \cap \mathbf{A}\right) = 1$$

Several comments are in order. The first remark is that in contrast to $\lambda(a,U)$, the constant $\tilde{\lambda}(a,U)$ does not depend any more on U. Furthermore the critical population size $\tilde{N}(a,U)$ does not depend on $\delta(U)$. In addition, the bound

$$\lambda(a,U) > \tilde{\lambda}(a,U)$$

seems to indicate that it is more difficult for the algorithm associated to the communication cost function V to move from one configuration to a better one. This observation also implies that for the critical size values we obtained we have that

$$N(a,U) = \left(1 + \frac{\delta(U)}{\delta(a)}\right) \ \tilde{N}(a,U) > \ \tilde{N}(a,U)$$

Let us see what happens when our this alternative genetic type model specializes to the case where the state is

$$E = \{-1,+1\}^{\mathcal{S}} \qquad \mathcal{S} = [-n,n]^p \qquad p \geq 1$$

and the fitness function U is given by

$$U(x) = \sum_{s \in \mathcal{S}} \sum_{s' \in V_s} \mathcal{I}_{s,s'} \ x(s) \ x(s') + \sum_{s \in \mathcal{S}} h(s) \ x(s)$$

where $\mathcal{I}_{s,s'}, h(s) \in Z$, and

$$\forall s \in \mathcal{S} \qquad V_s = \{s' \in \mathcal{S} \ : \ |s_k - s'_k| \leq 1, \ \ 1 \leq k \leq p\}$$

Let k be the Markovian mutation kernel on S given by

$$k(x,y) = \frac{1}{|\mathcal{V}(x)|} \ 1_{\mathcal{V}(x)}(y)$$

with

$$\mathcal{V}(x) \stackrel{\text{def}}{=} \{y \in E \ : \ \text{Card}\{s \in \mathcal{S} \ : \ x(s) \neq y(s)\} \leq 1\}$$

Suppose that the function a is given by

$$a(x,y) = (1 - 1_x(y)) \qquad \forall (x,y) \in E^2$$

Then, one can check that

$$R(a) \leq \max_{x,y} \min_{q \in C_{x,y}} |q| = \text{card}(\mathcal{S}) = (2n+1)^p$$

and

$$\delta(a) = \Delta(a) = 1$$

Let $\mathcal{I}_{s,s'}$ and $h(s)$ be chosen so that $\Delta U \geq 1$ and let N be an integer that $N > (2n+1)^p$. The above theorem shows that N individuals will solve the optimization problem when using the genetic algorithm associated to $\tilde{Q}_\beta$.

5.2 Conditional Mutations

We now present some genetic-type variants arising in non linear filtering literature (see [12, 20] and references therein). For the sake of unity and to highlight issues in both non linear filtering and numerical function analysis we place ourselves in the abstract setting of section 2 and section 3.

The first variation is based on the observation that the distribution flow $\{\widehat{\eta}_n \ ; \ n \geq 0\}$ is solution of a measure valued dynamical system defined as in (5) by replacing the transitions K_n and the fitness functions g_n by the transitions $\widehat{K}_n$ and the fitness functions $\widehat{g}_n$ defined for any $f \in \mathcal{B}_b(E)$ by setting

$$\widehat{K}_n(f) \stackrel{\text{def.}}{=} \frac{K_n(g_n f)}{K_n(g_n)} \qquad \text{and} \qquad \widehat{g}_n \stackrel{\text{def.}}{=} K_n(g_n)$$

More precisely one can check that

$$\widehat{\eta}_n = \widehat{\Phi}_n(\widehat{\eta}_{n-1}), \qquad n \geq 1 \tag{38}$$

with

$$\widehat{\Phi}_n(\eta) \stackrel{\text{def.}}{=} \widehat{\Psi}_n(\eta)\widehat{K}_n$$

and

$$\widehat{\Psi}_n(\eta)(f) \stackrel{\text{def.}}{=} \frac{\eta(\widehat{g}_n \, f)}{\eta(\widehat{g}_n)} \qquad \forall f \in \mathcal{B}_b(E)$$

As in section 2.1 we can associate to (38) an N-interacting particle system $\{\zeta_n \ ; \ n \geq 0\}$ which is a Markov chain in E^N with transitions

$$\mathbb{P}\left(\zeta_{n+1} \in dx \,|\zeta_n = y\,\right) = \prod_{p=1}^{N} \widehat{\Phi}_{n+1}\left(\frac{1}{N}\sum_{i=1}^{N} \delta_{y^i}\right)(dx^p), \qquad n \geq 0$$

and initial law $\widehat{\eta}_0 = \Psi(\eta_0)$, where as usual $dx \stackrel{\text{def}}{=} dx^1 \times \cdots \times dx^N$ is an infinitesimal neighborhood of the point $x = (x^1, \ldots, x^N) \in E^N$, $y = (y^1, \ldots, y^N) \in E^N$. Arguing as in section 2.1 it is transparent that this transition is decomposed into two separate mechanisms, namely for each $n \geq 0$

$$\zeta_n \stackrel{\text{def.}}{=} \left(\zeta_n^1, \ldots, \zeta_n^N\right) \stackrel{\text{Selection}}{\longrightarrow} \widehat{\zeta}_n \stackrel{\text{def.}}{=} \left(\widehat{\zeta}_n^1, \ldots, \widehat{\zeta}_n^N\right) \stackrel{\text{Mutation}}{\longrightarrow} \zeta_{n+1}.$$

The selection transition is now defined by

$$\mathbb{P}\left(\widehat{\zeta}_n \in dx \,|\zeta_n = y\right) = \prod_{p=1}^{N} \sum_{i=1}^{N} \frac{\widehat{g}_n(y^i)}{\sum_{j=1}^{N} \widehat{g}_n(y^j)} \ \delta_{y^i}(dx^p)$$

and the mutation step

$$\mathbb{P}\left(\zeta_{n+1} \in dx \,\middle|\, \widehat{\zeta}_n = y\right) = \prod_{p=1}^{N} \widehat{K}_{n+1}\left(y^p, dx^p\right)$$

We emphasize that in contrast to the latter genetic model this genetic particle scheme involves mutation transitions that depend on the fitness functions. The study of this variant has been initiated in [27], large population asymptotic are described in [24] and [20].

5.3 Branching Genetic-type Algorithms

We end this section with a brief description of branching and genetic-type variants presented in [20]. Here again we place ourselves in the abstract setting of section 2 and section 3.

All these branching strategies are based on the same natural idea. Namely, how to approximate an updated empirical measure of the following form

$$\Psi_n\left(\frac{1}{N}\sum_{i=1}^{N}\delta_{\xi_n^i}\right) = \sum_{i=1}^{N}\frac{g_n(\xi_n^i)}{\sum_{j=1}^{N} g_n(\xi_n^j)}\ \delta_{\xi_n^i} \tag{39}$$

by a new probability measure with atoms of size integers multiples of $1/N$? In the genetic algorithm presented in section 2.1 this approximation is done by sampling N-independent random variables

$$\{\widehat{\xi}_n^i\ ;\ 1 \leq i \leq N\}$$

with common law (39) and the corresponding approximating measure is given by

$$\frac{1}{N}\sum_{i=1}^{N}\delta_{\widehat{\xi}_n^i} = \sum_{i=1}^{N}\frac{M_n^i}{N}\ \delta_{\xi_n^i}$$

where

$$\left(M_n^1, \ldots, M_n^N\right) \overset{\text{def.}}{=} \text{Multinomial}\left(N, W_n^1, \ldots, W_n^N\right)$$

and for any $1 \leq i \leq N$

$$W_n^i \overset{\text{def.}}{=} \frac{g_n(\xi_n^i)}{\sum_{j=1}^{N} g_n(\xi_n^j)}$$

Using these notations the random and $\mathbb{N}$-valued random variables

$$\left(M_n^1, \ldots, M_n^N\right)$$

can be regarded as random numbers of offsprings created at the positions $(\xi_n^1, \ldots, \xi_n^N)$. The above question is strongly related to weighted bootstrap and genetic algorithms theory (see for instance [4] and references therein). In this connection the above multinomial approximating strategy can be viewed as a Weighted Efron bootstrap.

Let us present several examples of branching laws The first one is known as the *Remainder Stochastic Sampling* in genetic algorithms literature. It has been presented for the first time in [2, 3]. From a pure practical point of view this sampling technique seems to be the more efficient since it is extremely time saving and if the branching particle model is only based on this branching selection scheme then the size of the system remains constant.

In what follows we denote by $[a]$ (resp. $\{a\} = a - [a]$) the integer part (resp. the fractional part) of $a \in \mathbb{R}$.

1. **Remainder Stochastic Sampling**
   At each time $n \geq 0$, each particle ξ_n^i branches directly into a fixed number of offsprings

$$\overline{M}_n^i \stackrel{\text{def.}}{=} [NW_n^i] \qquad \forall 1 \leq i \leq N$$

   so that the intermediate population consists of $\overline{N}_n \stackrel{\text{def.}}{=} \sum_{i=1}^N \overline{M}_n^i$ particles. To prevent extinction and to keep the size of the system fixed it is convenient to introduce in this population $\tilde{N}_n$ additional particles with

$$\tilde{N}_n \stackrel{\text{def.}}{=} N - \overline{N}_n = \sum_{i=1}^N NW_n^i - \sum_{i=1}^N [NW_n^i] = \sum_{i=1}^N \{NW_n^i\}$$

   One natural way to do this is to introduce the additional sequence of branching numbers

$$\begin{aligned} &\left(\tilde{M}_n^1, \ldots, \tilde{M}_n^N\right) \\ &\stackrel{\text{def.}}{=} \text{Multinomial}\left(\tilde{N}_n, \frac{\{NW_n^1\}}{\sum_{j=1}^N \{NW_n^j\}}, \ldots, \frac{\{NW_n^N\}}{\sum_{j=1}^N \{NW_n^j\}}\right) \end{aligned} \tag{40}$$

   More precisely, if each particle ξ_n^i again produce a number of $\tilde{M}_n^i$ additional offsprings, $1 \leq i \leq N$, then the total size of the system is kept constant.
   At the end of this stage, the particle system $\widehat{\xi}_n$ again consists of N particles denoted by

$$\widehat{\xi}_n^i \quad = \quad \xi_n^k$$

with

$$1 \leq k \leq N, \qquad \sum_{l=1}^{k-1} \overline{M}_n^l + 1 \leq i \leq \sum_{l=1}^{k-1} \overline{M}_n^l + \overline{M}_n^k$$

and for

$$1 \leq k \leq N, \qquad \sum_{l=1}^{k-1} \tilde{M}_n^l + 1 \leq i \leq \sum_{l=1}^{k-1} \tilde{M}_n^l + \tilde{M}_n^k$$

$$\widehat{\xi}_n^{\overline{N}_n+i} \quad = \quad \xi_n^k$$

The multinomial (40) can also be defined as follows

$$\tilde{M}_n^k = \text{Card}\left\{1 \leq j \leq \tilde{N}_n \; ; \; \tilde{\xi}_n^j = \xi^k\right\} \qquad 1 \leq k \leq N$$

where $(\tilde{\xi}_n^1, \ldots, \tilde{\xi}_n^{\tilde{N}_n})$ are $\tilde{N}_n$ independent random variables with common law

$$\sum_{i=1}^{N} \frac{\{NW_n^i\}}{\sum_{j=1}^{N} \{NW_n^j\}} \, \delta_{\xi_n^i}$$

2. **Independent Branching Numbers**
In the next examples the branching numbers are, at each time step, independent one of each other (conditionally on the past). As a result the size of the population at each time n is not fixed but random. The corresponding branching genetic type algorithms can be regarded as a two step Markov chain

$$(N_n, \xi_n) \overset{Branching}{\longrightarrow} (\widehat{N}_n, \widehat{\xi}_n) \overset{Mutation}{\longrightarrow} (N_{n+1}, \xi_{n+1}) \tag{41}$$

with product state space $\mathcal{E} = \bigcup_{\alpha \in \mathbb{N}} (\{\alpha\} \times E^\alpha)$ with the convention $E^\alpha = \{\triangle\}$ a cemetery if $\alpha = 0$. We will note

$$\mathcal{F} = \{F_n, \widehat{F}_n \; : \; n \geq 0\}$$

the canonical filtration associated to (41) so that

$$F_n \; \subset \; \widehat{F}_n \; \subset F_{n+1}$$

(a) **Bernoulli branching numbers**
The Bernoulli branching numbers were introduced in [11] and further developed in [12]. They are defined as a sequence $M_n = (M_n^i, \; 1 \leq i \leq N_n)$ of conditionally independent random numbers with respect to F_n with distribution given for any $1 \leq i \leq N_n$ by

$$P(M_n^i = k | F_n) = \left\{ \begin{array}{rll} \{N_n W_n^i\} & \text{if} & k = [N_n W_n^i] + 1 \\ 1 - \{N_n W_n^i\} & \text{if} & k = [N_n W_n^i] \end{array} \right.$$

In addition it can be seen from the relation

$$\sum_{i=1}^{N_n}(N_n W_n^i) = N_n$$

that at least one particle has one offspring (cf. [11] for more details).
Therefore using the above branching correction the particle system never dies.

It is also worth observing that the Bernouilli branching numbers are defined as in the *Remainder Stochastic Sampling* by replacing the multinomial remainder branching law (40) by a sequence of N_n independent Bernouilli random variables $\left(\tilde{M}_n^1, \ldots, \tilde{M}_n^{N_n}\right)$ given by

$$P(\tilde{M}_i^{N_n} = 1|F_n) = 1 - P(\tilde{M}_i^{N_n} = 0|F_n) = \{N_n W_n^i\}$$

(b) **Poisson branching numbers:**
The Poisson branching numbers are defined as a sequence $M_n = (M_n^i,\ 1 \leq i \leq N_n)$ of conditionally independent random numbers with respect to F_n with distribution given for any $1 \leq i \leq N_n$ by

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

(c) **Binomial branching numbers:**

The binomial branching numbers are defined as a sequence $M_n = (M_n^i,\ 1 \leq i \leq N_n)$ of conditionally independent random numbers with respect to F_n with distribution given for any $1 \leq i \leq N_n$ by

$$P(M_n^i = k|F_n) = \left(\begin{array}{c} N_n \\ k \end{array}\right)\ (W_n^i)^k\ (1 - W_n^i)^{N_n - k}$$

for any $0 \leq k \leq N_n$

The previous models are described in full details in [12]. In particular it is shown that the genetic algorithm with multinomial branching laws arises by conditioning a genetic algorithm with Poisson branching laws.
The law of large numbers and large deviations for the genetic model with Bernouilli branching laws are studied in [12] and [14]. The convergence analysis of these particle approximating schemes is still in progress.

6 Continuous time Genetic Algorithms

We shall now describe the continuous time version of the genetic algorithm discussed in section 2. This particle algorithm has been introduced in [23] for solving a flow of distributions defined by the ratio

$$\eta_t(f) = \frac{\gamma_t(f)}{\gamma_t(1)} \qquad \forall f \in \mathcal{B}_b(E) \qquad t \in \mathbb{R}_+ \tag{42}$$

where $\gamma_t(f)$ is defined through a Feynman-Kac formula of the following form

$$\gamma_t(f) = \mathbb{E}\left(f(X_t) \ \exp\left(\int_0^t U_s(X_s)\, ds \right) \right)$$

where $\{X_t \ ; \ t \in \mathbb{R}_+\}$ is a càdlàg and time in-homogeneous Markov process taking values in a Polish space E and $\{U_t \ ; \ t \in \mathbb{R}_+\}$ is a measurable collection of locally bounded (in time) and measurable non-negative functions. Here we merely content ourselves in describing the mathematical models of such particle numerical schemes. The detailed convergence analysis as the size of the system tends to infinity can be founded in [20] or [23]. In order to illustrate the idea in a simple form we will also make the sanguine assumption that X is a time homogeneous Markov process with initial law η_0, its infinitesimal generator is a bounded linear operator on the set on bounded Borel test functions $\mathcal{B}_b(E)$ and $U_t = U$ is a time homogeneous function. The interested reader is referred to [20] for a more general presentation including Riemannian or Euclidean diffusions X.

To motivate our work we also mention that the Feynman-Kac model (42) has different interpretations coming from quite distinct research areas. Firstly it can be regarded as the distributions of a random particle X killed at a given rate and conditioned by non-extinction (see for instance [48]). Secondly the previous Feynman-Kac formula may serve to model the robust version of the optimal filter in non linear filtering settings (see [20] and [23]). Finally, as pointed out in [20], the ratio distributions (42) can also be regarded as the solution flow of a simple generalized and spatially homogeneous Boltzmann equation as defined in [36, 44].

As for the discrete time models discussed in section 2.1 and section 3 one of the best way to define the genetic particle approximating models of (42) is through the dynamical structure of (42). By definition one can easily check that for any bounded Borel test function $f \in \mathcal{B}_b(E)$

$$\frac{d}{dt}\eta_t(f) = \eta_t(L(f)) + \eta_t(fU) - \eta_t(f)\eta_t(U) = \eta_t(L_{\eta_t}(f)) \tag{43}$$

where L_η, for any fixed distribution η on E, is the bounded linear operator on $\mathcal{B}_b(E)$ defined by

$$L_\eta(f)(x) = L(f)(x) + \int (f(z) - f(x))\, U(z)\, \eta(dz) \tag{44}$$

As its discrete time analogue (20), we want to solve a nonlinear and measure valued dynamical system (43) and the associate generator $\mathcal{L}_\eta$ is decomposed into two separate generators.

To highlight the quadratic nature of (43) and the connections with spatially homogeneous Boltzmann equations we also notice that (43) can be rewritten as

$$\frac{d}{dt}\,\eta_t(f) = \eta_t(L(f))$$

$$+\frac{1}{2}\int \eta_t(dx)\,\eta_t(dy) \times ((f(x^\star) - f(x)) + (f(y^\star) - f(y)))\; Q\,(x, y; dx^\star, dy^\star)$$

with

$$Q\,(x, y; .) = U(y)\; \delta_{(y,y)} + U(x)\; \delta_{(x,x)}$$

In the first section 6.1 we discuss a Moran-type particle approximation of the Feynman-Kac formula (42). In section 6.1.2 we also give an illustration of the semi-group techniques introduced in [20] for proving useful convergence results as the size of the population tends to infinity including central limit theorem and exponential bounds. In the final section 6.2 we propose a branching and interacting particle approximating scheme. To the best of our knowledge this branching-type particle approximation of the Feynman-Kac formula (42) has not been covered in the literature. We will also give the connections between this particle scheme and the previous Moran particle model.

6.1 A Moran Particle Model

6.1.1 Description of the Model

As traditionally, starting from a family $\{\mathcal{L}_\eta\ ;\ \eta \in \mathbf{M}_1(E)\}$, we consider an interacting N-particle system $(\xi_t)_{t\geq 0} = ((\xi_t^1, \ldots, \xi_t^N))_{t\geq 0}$, which is Markov process on the product space E^N, $N \geq 1$, whose infinitesimal generator acts on bounded Borel functions $f : E^N \to \mathbb{R}$ by setting for any $x = (x_1, \ldots, x_N) \in E^N$

$$\mathcal{L}(f)(x) = \sum_{i=1}^{N} L^{(i)}_{m(x)}(f)(x) \qquad \text{with} \quad m(x) \stackrel{\text{def.}}{=} \frac{1}{N}\sum_{i=1}^{N} \delta_{x_i}$$

and where the notation $\mathcal{G}^{(i)}$ have been used instead of $\mathcal{G}$ when an operator $\mathcal{G}$ on $\mathcal{B}_b(E)$ acts on the i-th variable of $f(x_1, \ldots, x_N)$. This abstract and general formulation is well known in mean field interacting particle system literature (the interested reader is for instance referred to [44] and [50] and references therein).
Taking into consideration the definition (44) we get

$$\mathcal{L} = \tilde{\mathcal{L}} + \widehat{\mathcal{L}} \tag{45}$$

where

$$\tilde{\mathcal{L}}(f)(x) = \sum_{i=1}^{N} L^{(i)}(f)(x)$$

and

$$\widehat{\mathcal{L}}(f)(x) = \sum_{i=1}^{N} \sum_{j=1}^{N} \Big(f(x^{(i,j)}) - f(x) \Big) \ \frac{1}{N} \, U(x_j)$$

and where for $1 \leq i, j \leq N$ and $x = (x_1, \cdots, x_N) \in E^N$, $x^{(i,j)}$ is the element of E^N given by

$$\forall \ 1 \leq k \leq N, \qquad x_k^{(i,j)} \quad = \quad \left\{ \begin{array}{ll} x_k & , \text{ if } k \neq i \\ x_j & , \text{ if } k = i \end{array} \right.$$

In order to describe more explicitly the time evolution of the E^N-valued Markov process $\{\xi_t \ ; \ t \geq 0\}$ with infinitesimal generator $\mathcal{L}$ it is convenient to write (45) as follows

$$\begin{aligned} \mathcal{L}(f)(x) &= \tilde{\mathcal{L}}(f)(x) + \lambda(x) \int_{E^N} (f(y) - f(x)) \ \mathcal{Q}(x, dy) \\ &= \tilde{\mathcal{L}}(f)(x) + \widehat{\lambda} \int_{E^N} (f(y) - f(x)) \ \widehat{\mathcal{Q}}(x, dy) \end{aligned} \tag{46}$$

with

$$\lambda(x) = \sum_{i=1}^{N} U(x_i) = N \ m(x)(U) \qquad \text{and} \qquad \widehat{\lambda} = N \ \|U\|$$

and

$$\begin{aligned} \mathcal{Q}(x, dy) &= \sum_{i,j=1}^{N} \frac{1}{N} \, \frac{U(x_i)}{\sum_{k=1}^{N} U(x_k)} \, \delta_{x^{(i,j)}}(dy) \\ \widehat{\mathcal{Q}}(x, dy) &= \left(1 - m(x) \left(\frac{U}{\|U\|} \right) \right) \delta_x(dy) + m(x) \left(\frac{U}{\|U\|} \right) \ \mathcal{Q}(x, dy) \end{aligned}$$

The construction of $\{\xi_t \ ; \ t \geq 0\}$ on an explicit probability space is now classical (see for instance [23] or [31]). For the convenience of the

reader we propose a basic construction based on the second decomposition (46).

Let $\{X^{(k,i)}(a)\ ;\ (k,i) \in \mathbb{N}^2,\ a \in E\}$ be a collection of independent copies of $\{X(a)\ ;\ a \in E\}$ where for any $a \in E$, $X(a)$ denotes the process X starting at a. Let $\{T_k\ ;\ k \in \mathbb{N}\}$ $(T_0 = 0)$ be a sequence of independent and identically distributed random variables on $\mathbb{R}_+$ with common exponential law with parameter $N\,\|U\|$.

The random times $\{T_k\ ;\ k \in \mathbb{N}\}$ $(T_0 = 0)$ will be regarded as the random dates at which competitive interaction occurs. The initial particle system $\xi_0 = (\xi_0^1, \ldots, \xi_0^N)$ consists of N independent random variables with common law η_0.

1. **Mutation** :
   Between the dates T_{k-1} and T_k the particles evolve randomly and independently according the law of the time-inhomogeneous Markov process X. That is for any $1 \leq i \leq N$
   $$\xi_t^i = X^{(k,i)}_{t-T_{k-1}}\left(\xi^i_{T_{k-1}}\right), \qquad \forall t \in [T_{k-1}, T_k[\qquad k \geq 1$$
2. **Competitive Selection** :
   At the time $t = T_k$, $\xi_{T_k} = (\xi^1_{T_k}, \ldots, \xi^N_{T_k})$ is an E^N-valued random variable with law $\widehat{\mathcal{Q}}\left(\xi_{T_k-}, .\right)$.

The important difference between this Moran-type particle model and the classical one is that for the former N-particles system, the total rate of selection jumps $\widehat{\lambda}$ is of order N, while for the classical N- particle Moran model it is of order N^2. It is that difference of scaling, with comparatively less frequent selections, which enables us to end up with a deterministic process in the limit.

Furthermore, even if we would have multiplied by N the rate of selection, the limit exists (as a right continuous measure valued stochastic process) only if the weight of replacing the particle ξ_t^i by the particle ξ_t^j is symmetrical in ξ_t^i and ξ_t^j, condition which is not satisfied here, since due to the fitness functions, its value is $U_t(\xi_t^j)/N$. In our case more frequent selections would oblige the limit measure-valued process to jump instantaneously from a probability to another one better suited for the maximization of U. In fact, an asymmetrical weighted sampling needs a selection total rate of order N (this can be deduced from the calculations given in the section 5.7.8 of [15]), if one wants to end up with a bounded selection generator. Then one can add the natural non-weighted sampling selection (cf. section 2.5 of [15], or more generally, any other symmetrical weighted sampling selection) with a total rate of order N^2, to obtain in the limit a Fleming-Viot process

with selection, as it is defined directly in the level of measure valued process (and not at the particles system approximation level) in the section 10.1.1 of [15] (or more generally p. 175 of this review).

6.1.2 Asymptotic Behavior

The interpretation of the distribution flow $\{\eta_t \ ; \ t \geq 0\}$ in terms of the limit of the empirical measures

$$\eta_t^N \stackrel{\text{def.}}{=} \frac{1}{N} \sum_{i=1}^{N} \ \delta_{\xi_t^i} \tag{47}$$

as $N \to \infty$ is given in [20, 23] including central limit theorem and exponential bounds, see also [36, 44] for an alternative approach using coupling techniques. To see that (47) is a reasonable approximation of η_t observe that for any bounded Borel function $\varphi \in \mathcal{B}_b(E)$ if

$$f(x_1, \ldots, x_N) \stackrel{\text{def.}}{=} \frac{1}{N} \sum_{i=1}^{N} \ \varphi(x_i)$$

then for any $x = (x_1, \ldots, x_N) \in E^N$

$$\mathcal{L}(f)(x) = m(x) \left(L_{m(x)}(\varphi) \right)$$

Our aim is now to give some comments on the semi-group approach presented in [20] to study the asymptotic behavior of η_t^N as the population size N tends to infinity.
Under our assumption it is well known (see lemma 3.68, p 446 in [38]) that for any bounded Borel test function $f \in \mathcal{B}_b(E^N)$ the stochastic process

$$M_t(f) \stackrel{\text{def.}}{=} f(\xi_t) - f(\xi_0) - \int_0^t \ \mathcal{L}(f)(\xi_s) \ ds$$

is a square integrable martingale and its angle bracket is given by

$$\langle M(f) \rangle_t = \int_0^t \ \Gamma(f, f)(\xi_s) \ ds$$

where Γ is the "carré du champ" associated to $\mathcal{L}$

$$\forall f \in \mathcal{B}_b(E^N), \qquad \Gamma(f, f) = \mathcal{L}(f^2) - 2f \ \mathcal{L}(f)$$

Using the decomposition (45) and the definition of $\tilde{\mathcal{L}}$ and $\widehat{\mathcal{L}}$ it is easy to establish that

$$\Gamma(f, f) = \tilde{\Gamma}(f, f) + \widehat{\Gamma}(f, f)$$

with

$$\tilde{\Gamma}(f, f) = \tilde{\mathcal{L}}(f^2) - 2f \ \tilde{\mathcal{L}}(f) \qquad \widehat{\Gamma}(f, f) = \widehat{\mathcal{L}}(f^2) - 2f \ \widehat{\mathcal{L}}(f)$$

and if $f \in \mathcal{B}_b(E^N)$ is chosen so that

$$f(x) = m(x)(\varphi)$$

for some $\varphi \in \mathcal{B}_b(E)$ then

$$\tilde{\Gamma}(f,f)(x) \quad = \quad \frac{1}{N}\, m(x)\left(\Gamma_L(\varphi,\varphi)\right),$$

with

$$\Gamma_L(\varphi,\varphi) = L(\varphi^2) - 2\varphi\; L(\varphi)$$

and

$$\widehat{\Gamma}(f,f)(x) \quad = \quad \frac{1}{N}\, m(x)\left(\left(\varphi - m(x)(\varphi)\right)^2 (U + m(x)(U))\right)$$

Using these notations one concludes that

$$d\eta_t^N(\varphi) = \eta_t^N(L_{\eta_t^N}(\varphi))\; dt + dM_t(f)$$

with

$$|\langle M(f)\rangle_t| \leq \frac{C_t}{N}\; \|\varphi\|^2, \qquad C_t < \infty \qquad \forall t \geq 0$$

One can use this result to check that the sequence of distributions $\{\eta_t^N\ ;\ t \geq 0\}$ is weakly compact and any weak limit point is concentrated on the set of solutions of (43). Using the continuity of the angle bracket and the construction of ξ_t one can check that there exists some finite constant $C_t' < \infty$ such that the jumps $\Delta M_t(f)$ of the previously defined martingale are bounded by $C_t'\|\varphi\|/N$, that is $\mathbb{P}$-a.s.

$$|\Delta M_t(f)| \leq \frac{C_t'}{N}\; \|\varphi\|$$

Let us recall a classical exponential inequality for martingales M_t starting at 0 and whose jumps are bounded uniformly by $a \in]0,\infty[$: for all $0 < \epsilon \leq \dfrac{b}{a}$ and $t > 0$

$$\mathbb{P}\left(\sup_{s\in[0,t]} |M_s| > \epsilon\,, \quad \langle M\rangle_t \leq b\right) \leq 2\; \exp -\frac{\epsilon^2}{4b} \tag{48}$$

This inequality may be established using calculations from the section 4.13 of [43] (see corollary 3.3 in [47]). Now, if we apply this inequality to the martingale $M_t(f)$ one obtain the following result

Proposition 6.1 *For any bounded Borel test function $\varphi \in \mathcal{B}_b(E)$ and $T > 0$ and $0 < \epsilon \leq \|\varphi\|$ we have that*

$$\mathbb{P}\left(\sup_{t\in[0,T]} |\eta_t^N(\varphi) - \eta_0^N(\varphi) - \textstyle\int_0^t\; \eta_s^N(L_{\eta_s^N}(\varphi))\; ds| > \epsilon\right)$$
$$\leq 2\; \exp -\frac{N\epsilon^2}{C(t)\|\varphi\|^2}$$

for some finite constant $C(t) < \infty$.

To get some more precise estimates we proceed as in discrete time settings. We start by noting that

$$\gamma_t(1) = \exp \int_0^t \eta_s(U)\, ds$$

and therefore for any $\varphi \in \mathcal{B}_b(E)$

$$\gamma_t(\varphi) = \eta_t(\varphi) \ \exp \int_0^t \eta_s(U)\, ds$$

As in section 3 we introduce the N-approximating measures

$$\gamma_t^N(\varphi) \stackrel{\text{def.}}{=} \eta_t^N(\varphi) \ \exp \int_0^t \eta_s^N(U)\, ds$$

On the other hand using the Markovian property of X we observe the simple but essential fact that

$$\gamma_t(\varphi) = \gamma_s(K_{t-s}(\varphi))$$

where $\{K_\tau \ ; \ \tau \geq 0\}$ is the semi-group defined by

$$\forall \varphi \in \mathcal{B}_b(E), \qquad (K_\tau(\varphi))(x) = \mathbb{E}\left(\varphi\left(X_\tau(x)\right) \ \exp \int_0^\tau U(X_s(x))\, ds\right)$$

where $\{X_\tau(x) \ ; \ \tau \geq 0\}$ is the time homogeneous Markov process with infinitesimal generator L and starting at $x \in E$. From this simple observation one concludes that for any fixed $T > 0$ and for any $t \in [0, T]$, $x \in E$ and $\varphi \in \mathcal{B}_b(E)$

$$\frac{d}{dt}(K_{T-t}(\varphi))(x) = -L(K_{T-t}(\varphi))(x) - U(x)\ (K_{T-t}(\varphi))(x)$$

By definition of $\gamma_t^N(1)$ and using the same kind of arguments as before one can check that the stochastic process

$$\mathcal{M}_t(\varphi, T) \stackrel{\text{def.}}{=} \sqrt{N}\ \left(\gamma_t^N\left(K_{T-t}(\varphi)\right) - \gamma_0^N\left(K_T(\varphi)\right)\right), \qquad 0 \leq t \leq T$$

is a martingale and its angle bracket is given by

$$\begin{aligned}
&\langle \mathcal{M}(\varphi, T)\rangle_t \\
&= \int_0^t \gamma_s^N(1)\ \left\{\eta_s^N\left(\Gamma_L(K_{T-s}(\varphi), K_{T-s}(\varphi))\right) + \right. \\
&\qquad \left. \eta_s^N\left(\left(K_{T-s}(\varphi) - \eta_s^N(K_{T-s}(\varphi))\right)^2 \left(U + \eta_s^N(U)\right)\right)\right\}\, ds
\end{aligned} \tag{49}$$

Recalling that $\gamma_t(K_{T-t}(\varphi)) = \gamma_T(\varphi) = \gamma_0(K_T(\varphi))$ and $\gamma_0^N = \eta_0^N$, $\gamma_0 = \eta_0$ one concludes that

$$\begin{aligned}&\gamma_t^N\left(K_{T-t}(\varphi)\right) - \gamma_t\left(K_{T-t}(\varphi)\right)\\ &= \eta_0^N\left(K_T(\varphi)\right) - \eta_0\left(K_T(\varphi)\right) + \frac{1}{\sqrt{N}}\,\mathcal{M}_t(\varphi, T)\end{aligned}$$

¿From which it becomes clear that

Proposition 6.2 *For any $N \geq 1$ and for any bounded Borel test function $\varphi \in \mathcal{B}_b(E)$ we have that*

$$\mathbb{E}\left(\gamma_T^N(\varphi)\right) = \gamma_T(\varphi)$$

and

$$\mathbb{E}\left(\left(\gamma_T^N(\varphi) - \gamma_T(\varphi)\right)^2\right)^{1/2} \leq \frac{C_T}{\sqrt{N}}\,\|\varphi\| \tag{50}$$

for some finite constant C_T which do not depend on the test function.

Using the same line of arguments as the one we used in discrete time settings (see section 3) it is possible to obtain central limit theorems for the N-approximating measures γ_T^N and η_T^N as well as errors bounds for the total variation distance. For instance using the decomposition

$$\begin{aligned}&\eta_T^N(\varphi) - \eta_T(\varphi) = \frac{\gamma_T^N(\varphi)}{\gamma_T^N(1)} - \frac{\gamma_T(\varphi)}{\gamma_T(1)}\\ &= \frac{1}{\gamma_T(1)}\left(\left(\gamma_T^N(\varphi) - \gamma_T(\varphi)\right) + \eta_T^N(\varphi)\ \left(\gamma_T(1) - \gamma_T^N(1)\right)\right)\end{aligned} \tag{51}$$

and (50) one gets the following result.

Proposition 6.3 *For any $N \geq 1$ and for any bounded Borel test function $\varphi \in \mathcal{B}_b(E)$*

$$\mathbb{E}\left(\left(\eta_T^N(\varphi) - \eta_T(\varphi)\right)^2\right)^{1/2} \leq \frac{C_T}{\sqrt{N}}\,\|\varphi\|$$

for some finite constant C_T which do not depend on the test function.

Using the decomposition (51) and proposition 6.2 one obtain that

$$\begin{aligned}\mathbb{E}\left(\eta_T^N(\varphi)\right) - \eta_T(\varphi) &= \mathbb{E}\left(\ \eta_T^N(\varphi)\ \left(1 - \frac{\gamma_T^N(1)}{\gamma_T(1)}\right)\right)\\ &= \mathbb{E}\left(\ \left(\eta_T^N(\varphi) - \eta_T(\varphi)\right)\ \left(1 - \frac{\gamma_T^N(1)}{\gamma_T(1)}\right)\right)\end{aligned}$$

Thus, a simple application of Cauchy-Schwartz's inequality yields that for any test function φ, $\|\varphi\| \leq 1$,

$$\left|\mathbb{E}\left(\eta_T^N(\varphi)\right) - \eta_T(\varphi)\right| \leq \frac{C_T}{N}$$

for some finite constant C_T which only depends on the time parameter T. By exchangeability of the particles and the definition of the total variation distance of probability measures this implies that

$$\|\mathcal{L}aw(\xi_t^i) - \eta_t\|_{\mathrm{tv}} \leq \frac{C_T}{N}$$

Finally, as the form of the angle bracket (49) indicates one can prove the following result

Theorem 6.4 ([20]) *The sequence of random fields*

$$U_T^N(f) \stackrel{\mathrm{def.}}{=} \sqrt{N}\ \left(\gamma_T^N(f) - \gamma_T(f)\right), \qquad f \in \mathcal{B}_b(E)$$

converges in law as $N \to \infty$, in the sense of finite distributions, to a centered Gaussian field $\{U_n(f)\ ;\ f \in \mathcal{B}_b(E)\}$ satisfying

$$\begin{aligned}
&\mathbb{E}\left(U_T(f)^2\right) = \eta_0\left(\left(K_T(\varphi) - \eta_0(K_T(\varphi))\right)^2\right) \\
&+ \int_0^t \gamma_s(1)\ \{\eta_s\left(\Gamma_L(K_{T-s}(\varphi), K_{T-s}(\varphi))\right) \\
&\qquad + \eta_s\left(\left(K_{T-s}(\varphi) - \eta_s(K_{T-s}(\varphi))\right)^2 (U + \eta_s(U))\right)\Big\}\ ds
\end{aligned}$$

Arguing as in discrete time settings the previous fluctuation result implies that the sequence of random fields

$$W_T^N(f) \stackrel{\mathrm{def.}}{=} \sqrt{N}\ \left(\eta_T^N(f) - \eta_T(f)\right), \qquad f \in \mathcal{B}_b(E)$$

converges in law as $N \to \infty$ to the centered Gaussian field

$$W_T(f) \stackrel{\mathrm{def.}}{=} U_T\left(\frac{1}{\gamma_T(1)}\left(f - \eta_T(f)\right)\right), \qquad f \in \mathcal{B}_b(E)$$

Finally, setting

$$M_t(T,\varphi) \stackrel{\mathrm{def.}}{=} \gamma_t^N(K_{T-t}(\varphi)) - \gamma_0^N(K_T(\varphi)), \qquad 0 \leq t \leq T$$

and using the same reasoning as before one can prove that for any $0 \leq t \leq T$

$$| < M(T,\varphi) >_t | \leq \frac{1}{N}\ C_T\ \|\varphi\|^2 \qquad \text{and} \qquad |\Delta M_t(T,\varphi)| \leq \frac{1}{N}\ C_T\ \|\varphi\|$$

for some finite constant $C_T < \infty$. Thus the exponential bound (48) implies that for any $0 < \epsilon \leq \|\varphi\|$

$$\mathbb{P}\left(\sup_{t\in[0,T]} |\gamma_t^N(K_{T-t}(\varphi)) - \gamma_0^N(K_T(\varphi))| > \epsilon\right) \leq 2\ \exp -\frac{N\epsilon^2}{C(T)\|\varphi\|^2}$$

for some finite constant $C(T) < \infty$. On the other hand, using Hoeffding's inequality we have that

$$\mathbb{P}\left(|\gamma_0^N(K_T(\varphi)) - \gamma_0(K_T(\varphi))| > \epsilon\right) \leq 2\ \exp - \frac{N\epsilon^2}{C'(T)\|\varphi\|^2}$$

for some finite constant $C'(T) < \infty$. If we combine these two bounds one concludes that

$$\begin{aligned}&\mathbb{P}\left(\sup_{t\in[0,T]} |\gamma_t^N(K_{T-t}(\varphi)) - \gamma_t(K_{T-t}(\varphi))| > \epsilon\right)\\ &\leq 4\ \exp - \frac{N\epsilon^2}{\max(C(T), C'(T))\|\varphi\|^2}\end{aligned}$$

and therefore

$$\mathbb{P}\left(|\eta_t^N(\varphi) - \eta_t(\varphi)| > \epsilon\right) \leq 4\ \exp - \frac{N\epsilon^2}{C''(T)\|\varphi\|^2}$$

for some finite constant $C''(T) < \infty$.

Uniform convergence results are developed in [20, 22]. These papers provide various stability conditions on the process X underwhich one can find (as in discrete time settings, see page 23) some coefficient $\alpha \in (0, 1/2)$ such that for any $1 \leq i \leq N$

$$\sup_{t\geq 0} \|\mathcal{L}aw(\xi_t^i) - \eta_t\|_{\mathrm{tv}} \leq \frac{C}{N^\alpha}, \qquad C < \infty$$

and, for any bounded Borel test function f, $\|f\| \leq 1$,

$$\forall p \geq 1 \qquad \sup_{t\geq 0} \mathbb{E}\left(\left|\eta_t^N(f) - \eta_t(f)\right|^p\right)^{1/p} \leq \frac{c(p)}{N^\alpha}$$

for some constant $c(p) < \infty$ which only depends on the parameter p.

6.2 A Branching Particle Model

We end this paper with a presentation of a novel genetic type model based on branching selection transitions. To our knowledge this model has not been covered by the literature and its convergence analysis is still in progress. We also believe that the semi-group approach presented in [20] applies to study the convergence of this branching algorithm to the distributions (42).

In contrast to the previous Moran-type genetic algorithm the size of the population here will not be necessarily fixed but random. As a

result the corresponding branching particle system will be regarded as a continuous time process taking values in the state space

$$\mathcal{E} \stackrel{\text{def.}}{=} \cup_{p\geq 0}\ E^p$$

with the convention $E^p = \{\triangle\}$ a cemetery point if $p = 0$. The point $\triangle$ will be isolated and by convention all bounded Borel test functions $f \in \mathcal{B}_b(\mathcal{E} - \{\triangle\})$ will be extended to $\mathcal{E}$ by setting $f(\triangle) = 0$.
It will be also convenient to adjoin $\triangle$ to the state space E and we set $E_\triangle = E \cup \{\triangle\}$. Again the test functions $\varphi \in \mathcal{B}_b(E)$ will be extended to $E_\triangle$ by setting $\varphi(\triangle) = 0$.

The infinitesimal generator $\mathcal{G}$ of this branching scheme is defined by

$$\mathcal{G} = \tilde{\mathcal{G}} + \widehat{\mathcal{G}} \tag{52}$$

where for any $f \in \mathcal{B}_b(\mathcal{E})$ and $x = (x_1, \ldots, x_p) \in E^p$, $p \geq 1$

$$\tilde{\mathcal{G}} f(x) = \sum_{i=1}^{p} L^{(i)}(f)(x)$$

and

$$\widehat{\mathcal{G}} f(x) = \lambda(x) \int_{\mathcal{E}} (f(y) - f(x))\ \mathcal{Q}(x, dy)$$

with

$$\lambda(x) = \sum_{i=1}^{p} U(x_i)$$

the transition probability kernel $\mathcal{Q}$ on $\mathcal{E}$ given by

$$\begin{aligned} \mathcal{Q}f(x) &= \int f(y)\ \mathcal{Q}(x, dy) \\ &= \frac{1}{p} \sum_{i=1}^{p} \sum_{q\geq 0} \left\{ \int_E f\left(x(i,q,u)\right)\ \mathcal{S}(x, x_i, du) \right\} \mathcal{B}(x, x_i, q) \end{aligned}$$

where

$$x(i,q,u) = (x_1, \ldots, x_{i-1}, \underbrace{u, \ldots, u}_{q \text{ times}}, x_{i+1}, \ldots, x_p)$$

and for any $x \in \mathcal{E}$, $\mathcal{S}(x, x_i, du)$ and $\mathcal{B}(x, x_i, q)$ are distributions on E and on $\mathbb{N}$. In our construction the point $\triangle$ will be an absorbing point in the sense that if the process started at $\triangle$ it will stay in $\triangle$. Therefore for $p = 0$ we will also use the convention $\sum_\emptyset = 0$ and $\mathcal{Q}(\triangle, \{\triangle\}) = 1$. With this convention if $p = 0$ (i.e. $x = \triangle$) we have that $\tilde{\mathcal{G}} f(\triangle) = 0$ and $\widehat{\mathcal{G}} f(\triangle) = 0$.

The distributions $\mathcal{S}(x, x_i, du)$ and $\mathcal{B}(x, x_i, q)$ will be chosen so that the following equality holds true

$$\frac{1}{p}\sum_{i=1}^{p} \overline{\mathcal{B}}(x, x_i)\ \overline{\mathcal{S}}(\varphi)(x, x_i) = \sum_{i=1}^{p} \frac{U(x_i)}{\sum_{j=1}^{p} U(x_j)}\ \varphi(x_i) \tag{53}$$

for any $\varphi \in \mathcal{B}_b(E)$ where $\overline{\mathcal{B}}(x, x_i)$ and $\overline{\mathcal{S}}(\varphi)(x, x_i)$ are defined by

$$\overline{\mathcal{B}}(x, x_i) = \sum_{q \geq 0}\ q\ \mathcal{B}(x, x_i, q)$$

and

$$\overline{\mathcal{S}}(\varphi)(x, x_i) = \int_E\ \mathcal{S}(x, x_i, du)\ \varphi(u)$$

We now make this condition more precise by noting that if $f \in \mathcal{B}_b(\mathcal{E})$ is defined for any $x = (x_1, \ldots, x_p) \in E^p$, $p \geq 1$, by

$$f(x) = p.m(x)(\varphi) \qquad \text{where} \qquad m(x) = \frac{1}{p}\sum_{i=1}^{p}\ \delta_{x_i} \tag{54}$$

for some $\varphi \in \mathcal{B}_b(E)$ then

$$\mathcal{G}(f)(x) = p.m(x)\left(L_{m(x)}(\varphi)\right)$$

To see this claim we first observe that for such a bounded test function f and for any $x = (x_1, \ldots, x_p) \in E^p$, $p \geq 1$,

$$\begin{aligned}
&\widehat{\mathcal{G}}(f)(x) \\
&= \lambda(x)\ \frac{1}{p}\sum_{i=1}^{p}\sum_{q \geq 0}\ \left\{\int_E\ (q\varphi(u) - \varphi(x_i))\ \mathcal{S}(x, x_i, du)\right\}\mathcal{B}(x, x_i, q) \\
&= \lambda(x)\ \frac{1}{p}\sum_{i=1}^{p}\sum_{q \geq 0}\ \left(\overline{\mathcal{B}}(x, x_i)\,\overline{\mathcal{S}}(\varphi)(x, x_i) - \varphi(x_i)\,\mathcal{B}(x, x_i, q)\right)
\end{aligned}$$

Using (53) one concludes that for any $x = (x_1, \ldots, x_p) \in E^p$, $p \geq 1$,

$$\begin{aligned}
\widehat{\mathcal{G}}(f)(x) &= \sum_{j=1}^{p} U(x_j)\ \left(\sum_{i=1}^{p}\ \frac{U(x_i)}{\sum_{j=1}^{p} U(x_j)}\ \varphi(x_i) - m(x)(\varphi)\right) \\
&= p.\ (\ m(x)(\varphi U) - m(x)(\varphi)\, m(x)(U)\ )
\end{aligned}$$

Recalling that for any bounded test function f of the form (54) and for any $x = (x_1, \ldots, x_p) \in E^p$, $p \geq 1$, we have that

$$\tilde{\mathcal{G}}(f)(x) = \sum_{i=1}^{p}\ L^{(i)}(f)(x) = \sum_{i=1}^{p}\ L(\varphi)(x_i) = p.m(x)(L(\varphi))$$

one concludes that for any bounded test function f of the form (54) and for any $x = (x_1, \ldots, x_p) \in E^p$, $p \geq 1$,

$$\begin{aligned} \mathcal{G}(f)(x) &= p.\,(\; m(x)(L(\varphi)) + m(x)\,(\varphi U) - m(x)\,(\varphi)\,m(x)(U)\;) \\ &= p.m(x)\left(L_{m(x)}(\varphi)\right) \end{aligned}$$

Along the same line of ideas as before it is possible to construct inductively the branching (with interaction) particle system with generator $\mathcal{G}$. In contrast to the previous situation the size of the population is not necessarily fixed and it will be denoted by N_t at each time t, in other words

$$\xi_t = (\xi_t^1, \ldots, \xi_t^{N_t}) \in E^{N_t}$$

We also need to introduce a sequence $\{t_k\,;\; k \geq 1\}$ of independent random variables with a common exponential law on $\mathbb{R}_+$ with parameter 1. If we write $\{T_k\,;\; k \geq 0\}$ the random times at which the competitive branching interaction occurs the inductive description is as follows.
Initially $T_0 = 0$ and the particle system $\xi_0 = (\xi_0^1, \ldots, \xi_0^{N_0})$ consists of N_0 independent random variables with common law η_0. The initial size N_0 is a non random integer and it represents the precision parameter of the scheme.
Now we assume that we have defined the branching process up to time T_{k-1} (included) for some $k \geq 1$.
If $N_{T_{k-1}} = 0$ the particle system dies and we let $N_t = 0$ and $\xi_t = \triangle$ for any $t \geq T_{k-1}$. Otherwise the mutation/branching selection transition is defined as follows.

1. **Mutation** :
   Between the dates T_{k-1} and T_k the particles evolve randomly and independently according the law of the time-inhomogeneous Markov process X. That is for any $1 \leq i \leq N_{T_{k-1}}$

   $$\xi_t^i = X_{t-T_{k-1}}^{(k,i)}\left(\xi_{T_{k-1}}^i\right) \qquad \forall t \in [T_{k-1}, T_k[$$

   where $T_k = T_{k-1} + \tau_k$ and τ_k is defined by setting

   $$t_k = \int_{T_{k-1}}^{T_{k-1}+\tau_k} \sum_{i=1}^{N_{T_{k-1}}} U\left(X_{s-T_{k-1}}^{(k,i)}\left(\xi_{T_{k-1}}^i\right)\; ds\right)$$

   (recall that t_k is a random variable with exponential law on $\mathbb{R}_+$ with parameter 1). During this stage the size of the system remains constant and we set

   $$N_t = N_{T_{k-1}} \qquad \forall t \in [T_{k-1}, T_k[$$

2. **Competitive branching selection** :
   At the time $t = T_k$ a label i is chosen uniformly on $\{1, \ldots, N_{T_{k-1}}\}$

and the particle with label i dies and is replaced by a random number of offsprings q_k^i with law

$$\mathcal{B}\left(\xi_{T_k-}, \xi^i_{T_k-}, .\right)$$

and independently, these offsprings are randomly given a location u_k^i with law

$$\mathcal{S}\left(\xi_{T_k-}, \xi^i_{T_k-}, .\right)$$

At the end of this stage the particle system ξ_{T_k} is defined as

$$\xi_{T_k} = (\xi^1_{T_k-}, \ldots, \xi^{i-1}_{T_k-}, \underbrace{u_k^i, \ldots, u_k^i}_{q_k^i \text{ times}}, \xi^{i+1}_{T_k-}, \ldots, \xi^{N_{T_{k-1}}}_{T_k-})$$

and the resulting population size is defined as

$$N_{T_k} = \left(N_{T_{k-1}} - 1\right) + q_k^i$$

Let us give some examples of branching selection laws satisfying condition (53). We assume that $x = (x_1, \ldots, x_p) \in E^p$ for some $p \geq 1$.

1. If $\mathcal{B}(x, x_i, .)$ and $\mathcal{S}(x, x_i, .)$ are defined by

$$\mathcal{B}(x, x_i, .) = \delta_1$$

and

$$\mathcal{S}(x, x_i, .) = \sum_{j=1}^{p} \frac{U(x_j)}{\sum_{k=1}^{p} U(x_k)}\, \delta_{x_j} \tag{55}$$

then, since for any $\varphi \in \mathcal{B}_b(E)$ and $1 \leq i \leq p$

$$\overline{\mathcal{B}}(x, x_i) = 1 \qquad \text{and} \qquad \overline{\mathcal{S}}(\varphi)(x, x_i) = \sum_{j=1}^{p} \frac{U(x_j)}{\sum_{k=1}^{p} U(x_k)}\, \varphi(x_j)$$

condition (53) clearly holds. This example corresponds to the Moran-type genetic scheme presented in the previous section. Indeed, in this situation we clearly have for any $f \in \mathcal{B}_b(\mathcal{E})$ and for any $x = (x_1, \ldots, x_p) \in E^p$, $p \geq 1$,

$$\begin{aligned}
&\widehat{\mathcal{G}}(f)(x) \\
&= \left(\sum_{k=1}^{p} U(x_k)\right) \; \frac{1}{p} \sum_{i=1}^{p} \sum_{j=1}^{p} \left(f(x^{(i,j)}) - f(x)\right) \; \frac{U(x_j)}{\sum_{l=1}^{p} U(x_l)} \\
&= \frac{1}{p} \sum_{i=1}^{p} \sum_{j=1}^{p} \left(f(x^{(i,j)}) - f(x)\right) \; U(x_j)
\end{aligned}$$

where for $1 \leq i, j \leq p$ and $x = (x_1, \cdots, x_p) \in E^p$, $x^{(i,j)}$ is as usual the element of E^p given by

$$\forall\ 1 \leq k \leq p, \qquad x_k^{(i,j)} \quad = \quad \left\{ \begin{array}{ll} x_k & , \text{ if } k \neq i \\ x_j & , \text{ if } k = i \end{array} \right.$$

2. If $\mathcal{B}(x, x_i, .)$ and $\mathcal{S}(x, x_i, .)$ are defined by

$$\mathcal{S}(x, x_i, .) = \delta_{x_i} \qquad \text{and} \qquad \overline{\mathcal{B}}(x, x_i) = \frac{p\ U(x_i)}{\sum_{k=1}^{p}\ U(x_k)} \tag{56}$$

then for any $\varphi \in \mathcal{B}_b(E)$ we have $\overline{\mathcal{S}}(\varphi)(x, x_i) = \varphi(x_i)$ and condition (53) is again met. In this situation the size of the population may not be fixed. To highlight the connections with the discrete time branching schemes presented in section 5.3 the reader may check that condition (56) holds for the Bernoulli and Poisson branching laws

$$\mathcal{B}(x, x_i, .)$$

$$= (1 - \{\overline{\mathcal{B}}(x, x_i)\})\ 1_{[\overline{\mathcal{B}}(x,x_i)]}(.) + \{\overline{\mathcal{B}}(x, x_i)\}\ 1_{[\overline{\mathcal{B}}(x,x_i)]+1}(.)$$

and

$$\mathcal{B}(x, x_i, .) = e^{-\overline{\mathcal{B}}(x,x_i)} \sum_{q \geq 0} \frac{(\overline{\mathcal{B}}(x, x_i))^q}{q!}\ 1_q(.)$$

We recall that $[a]$ (resp. $\{a\} = a - [a]$) the integer part (resp. the fractional part) of $a \in \mathbb{R}$.

Now we return to the probabilistic analysis of this branching particle model. We have study the asymptotic behavior of this scheme but the corresponding publication still isn't ready. For the convenience of the reader, here we only formulate a few basic result to illustrate how the methodology used for the Moran type genetic algorithm can be used in this more general framework. As usually we start by noting that for any bounded Borel test function $f \in \mathcal{B}_b(\mathcal{E})$ the stochastic process

$$M_t(f) \stackrel{\text{def.}}{=} f(\xi_t) - f(\xi_0) - \int_0^t \mathcal{G}(f)(\xi_s)\ ds$$

is a local martingale and its angle bracket is given by

$$\langle M(f) \rangle_t = \int_0^t \Gamma(f, f)(\xi_s)\ ds$$

where Γ is the "carré du champ" associated to $\mathcal{G}$

$$\Gamma(f, f) = \mathcal{G}(f^2) - 2f\ \mathcal{G}(f)$$

Now, using the decomposition (52) and the definition of $\tilde{\mathcal{G}}$ and $\widehat{\mathcal{G}}$ one can check that

$$\Gamma(f,f) = \tilde{\Gamma}(f,f) + \widehat{\Gamma}(f,f)$$

with

$$\tilde{\Gamma}(f,f) = \tilde{\mathcal{G}}(f^2) - 2f\ \tilde{\mathcal{G}}(f) \qquad \widehat{\Gamma}(f,f) = \widehat{\mathcal{G}}(f^2) - 2f\ \widehat{\mathcal{G}}(f)$$

If $f \in \mathcal{B}_b(\mathcal{E})$ is chosen so that for any $x = (x_1, \ldots, x_p) \in E^p$, $p \geq 1$, and for some $\varphi \in \mathcal{B}_b(E)$

$$f(x) = p.m(x)(\varphi) \tag{57}$$

then for any $x = (x_1, \ldots, x_p) \in E^p$, $p \geq 1$,

$$\tilde{\Gamma}(f,f)(x) = p.\ m(x)\left(\Gamma_L(\varphi,\varphi)\right),$$

with

$$\Gamma_L(\varphi,\varphi) = L(\varphi^2) - 2\varphi\ (L(\varphi)),$$

and

$$\begin{aligned}
&\widehat{\Gamma}(f,f)(x) = \widehat{\mathcal{G}}\left(\left(f(.) - f(x)\right)^2\right)(x) \\
&= \lambda(x)\ \frac{1}{p}\sum_{i=1}^{p}\sum_{q\geq 0}\ \int_E \left(q\varphi(u) - \varphi(x_i)\right)^2\ S(x,x_i,du)\ B(x,x_i,q)
\end{aligned}$$

Let us notice that if distributions $S(x,x_i,du)$ and $B(x,x_i,q)$ are defined by (55) then for any $x = (x_1, \ldots, x_p) \in E^p$, $p \geq 1$, we have that

$$\widehat{\Gamma}(f,f)(x) = p.m(x)\left(\left(\varphi - m(x)(\varphi)\right)^2\left(U + m(x)(U)\right)\right)$$

and if these distributions satisfy (56) one gets that

$$\widehat{\Gamma}(f,f)(x) = p.m(x)(U).m(x)\left(\varphi^2\ \sum_{q\geq 0}(q-1)^2\ B(x,.,q)\right)$$

In contrast to the previous Moran type genetic model the carré du champ corresponding to the selection procedure is not necessarily bounded and we need to introduce some auxiliary assumption on the mass variation of the systems, namely we will assume that

$$\sup_{x\in\mathcal{E}}\ m(x)\left(\sum_{q\geq 0}\ q^2\ B(x,.,q)\right) < \infty$$

In the special case where the test function f is given by (57) we have that the stochastic process

$$N_t\ m(\xi_t)(\varphi) - N_0\ m(\xi_0)(\varphi) - \int_0^t\ N_s\ m(\xi_s)\left(L_{m(\xi_s)}\varphi\right)\ ds, \qquad t \geq 0$$

is a square integrable martingale. Of course, if $\varphi = 1$ this implies that the total mass process $\{N_t\ ;\ t \geq 0\}$ is a square integrable martingale starting at N_0.

References

[1] J. Abela, D. Abramson, M. Krishnamoorthy, A. De Silval and G. Mills. Computing Optimal Schedules for Landing Aircraft. Technical Report, Department of Computer Systems Eng. R.M.I.T., Melbourne, May 25, 1993.

[2] J. Baker, (1985) Adaptative Selection Methods for Genetic Algorithms. *Proc. International Conf. on Genetic Algorithms and their Applications.* J. Grefenstette, ed. Lawrence Erlbaum.

[3] J. Baker, (1987) Reducing Bias and Inefficiency in the Selection Algorithm. *Second Proc. International Conf. on Genetic Algorithms and their Applications.* J. Grefenstette, ed. Lawrence Erlbaum.

[4] P. Barbe and P. Bertail. The Weighted Bootstrap. *Lecture Notes in Statistics*, No 98, Springer-Verlag, 1995.

[5] R. Bott, J.P. Mayberry, *Matrices and trees*, Economics Activity Analysis, Wiley, New York, 1954.

[6] H. Carvalho, Filtrage optimal non linéaire du signal GPS NAVSTAR en racalage de centrales de navigation, Thèse de L'Ecole Nationale Supérieure de l'Aéronautique et de l'Espace, Sept. 1995.

[7] H. Carvalho, P. Del Moral, A. Monin, G. Salut, *Optimal non-linear filtering in GPS/INS Integration*, IEEE Trans. on Aerospace and Electronic Systems, Vol. 33, (3), pp. 835-850, 1997.

[8] R. Cerf (1994) Asymptotic convergence of a genetic algorithm. C. R. Acad. Sci. Paris Sr. I Math. 319 (1994), no. 3, 271–276.

[9] R. Cerf (1996) The dynamics of mutation-selection algorithms with large population sizes. Ann. Inst. H. Poincar Probab. Statist. 32 (1996), no. 4, 455–508.

[10] R. Cerf (1998) Asymptotic convergence of genetic algorithms. Adv. in Appl. Probab. 30 (1998), no. 2, 521–550.

[11] D. Crisan, J. Gaines and T.J. Lyons. *A Particle Approximation of the Solution of the Kushner-Stratonovitch Equation.* SIAM J. Appl. Math. 58 (1998), no. 5, 1568–1590

[12] D. Crisan, P. Del Moral and T.J. Lyons. *Non Linear Filtering Using Branching and Interacting Particle Systems.* To appear in Markov Processes and Related Fields, vol.3 (1999).

[13] D. Crisan, P. Del Moral and T. Lyons, *Interacting Particle Systems Approximations of the Kushner Stratonovitch Equation.* To appear in Advances in Applied Probability, vol.31, no.3, September 1999.

[14] D. Crisan and M. Grunwald, Large Deviation Comparison of Branching Algorithms versus Re-sampling Algorithm. Preprint, Imperial College, London, 1998.

[15] D. Dawson. Measure-valued Markov processes. In P.L. Hennequin, editor, *Ecole d'Eté de Probabilités de Saint-Flour XXI-1991*, Lecture Notes in Mathematics 1541. Springer-Verlag, 1993.

[16] D. Delahaye, J.-M. Alliot, M. Schoenauer, and J.-L. Farges. Genetic algorithms for automatic regrouping of air traffic control sectors. J. R. McDonnell, R. G. Reynolds, and D. B. Fogel, editors, Proceedings of the 4th Annual Conference on Evolutionary Programming, pages 657–672. MIT Press, March 1995.

[17] P. Del Moral, J. Jacod and Ph. Protter, *The Monte-Carlo Method for filtering with discrete time observations.* Publications du Laboratoire de Probabilités Paris VI, No 453 , june 1998.

[18] P. Del Moral, J. Jacod, *The Monte-Carlo Method for filtering with discrete time observations. Central Limit Theorems* Publications du Laboratoire de Statistiques et Probabilités, Toulouse III, No 07-99.

[19] P. Del Moral, J. Jacod, *Interacting Particle Filtering With Discrete Observations*
Publications du Laboratoire de Statistiques et Probabilités, Toulouse III, No 08-99.

[20] P. Del Moral and L. Miclo, *Branching and Interacting Particle Systems Approximations of Feynman-Kac Formulae with Applications to Non-Linear Filtering* , Publications du Laboratoire de Statistiques et Probabilités, 05-99, 1999.

[21] P. Del Moral and L. Miclo, *On the Convergence and the Applications of the Generalized Simulated Annealing*, SIAM Journal on Control and Optimization, Volume 37, Number 4, pp. 1222-1250.

[22] P. Del Moral and L. Miclo. Asymptotic Stability of nonlinear Semigroups of Feynman-Kac type. Publications du Laboratoire de Statistiques et Probabilités No 04-99 (1999).

[23] P. Del Moral and L. Miclo. A Moran particle system approximation of Feynman-Kac formulae. Publications du Laboratoire de Statistiques et Probabilités No 11-98 (1998).

[24] P. Del Moral and A. Guionnet, Large Deviations for Interacting Particle Systems. Applications to nonlinear Filtering Problems. *Stochastic Processes and their Applications* **78**, 69-95, (1998).

[25] P. Del Moral and A. Guionnet. On the Stability of Measure Valued Processes. Applications to nonlinear Filtering and Interacting Particle Systems, *Publications du Laboratoire de Statistiques et Probabilités*, Université Paul Sabatier, No 03-98, 1998.

[26] P. Del Moral. nonlinear Filtering: Interacting particle resolution. *Markov Processes and Related Fields, Vol.2*, 1996.

[27] P. Del Moral. Measure Valued Processes and Interacting Particle Systems. Application to nonlinear Filtering Problems. *Ann. Appl. Probab.* 8 (1998), no. 2, 438–495.

[28] P. Del Moral, J.C. Noyer, G. Salut, *Résolution particulaire et traitement non-linéaire du signal : Application Radar/Sonar*, Traitement du signal, Septembre 1995.

[29] P. Del Moral, G. Rigal, J.C. Noyer, G. Salut, *Traitement non-linéaire du signal par reseau particulaire : Application Radar*, quatorzième colloque GRETSI, Juan les Pins, 13-16 Septembre 1993.

[30] P. Del Moral, G. Rigal and G. Salut, *Estimation et commande optimale non lineaire*, Contract D.R.E.T.-DIGILOG-LAAS/CNRS, SM.MCY/685.92/A, 89.34.553.00.470.75.01, Report No.2, 18 Mars 1992.

[31] S. Ethier and T. Kurtz. *Markov Processes, Characterization and Convergence*. Wiley series in probability and mathematical statistics. John Wiley and Sons, New York, 1986.

[32] M.I. Freidlin, A.D. Wentzell, *Random Perturbations of Dynamical Systems*, Springer Verlag, 1984.

[33] D.E. Goldberg (1985), Genetic Algorithms and Rule Learning in Dynamic Control Systems, Proceedings of the First International Conference on Genetic Algorithms, Lawrence Erlbaum Associates Hillsdale, NJ, pp. 8-15.

[34] D.E. Goldberg (1989), Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley, Reading, MA.

[35] N.J. Gordon, D.J. Salmon, A.F.M. Smith, *Novel Approach to non-linear/non-Gaussian Bayesian state estimation*, IEE 1993.

[36] C. Graham and S. Méléard. Stochastic particle approximations for generalized Boltzmann models and convergence estimates. *The Annals of Probability*, 25(1):115–132, 1997.

[37] J.H. Holland (1975) *Adaptation in Natural and Artificial Systems.* University of Michigan Press, Ann Arbor.

[38] J.Jacod and A.N. Shiryaev. Limit Theorem for Processes *Springer-Verlag*.

[39] J.M. Johnson and Y. Rahmat-Samii. Genetic algorithms in electromagnetics. IEEE Antennas and Propagation Society International Symposium Digest, Vol. 2, pp. 1480-1483.

[40] C. Kane and M. Schoenauer Topological Optimum Design using Genetic Algorithms Control and Cybernetics - 25(5) - 1996.

[41] G. Kitagawa, Monte-Carlo Filter and Smoother for Non-Gaussian nonlinear State Space Models, *Journal on Computational and Graphical Statistics*, Vol. 5., no.1, pp.1-25, 1996.

[42] J.W. Kwiatkowski, Algorithms for Index Tracking, Department of Business Studies, The University of Edinburgh, 1991.

[43] R.S. Liptser and A.N. Shiryayev. Theory of Martingales. Dordrecht: Kluwer Academic Publishers, 1989.

[44] S. Méléard. Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models. In D. Talay and L. Tubaro, editors, *Probabilistic Models for nonlinear Partial Differential Equations, Montecatini Terme, 1995*, Lecture Notes in Mathematics 1627. Springer-Verlag, 1996.

[45] C. Musso and N. Oudjane, Regularized Particle Schemes applied to the Tracking Problem, ONERA Chatillon (preprint 1998).

[46] C. Musso and N. Oudjane, Regularization Schemes for Branching Particle Systems as a Numerical Solving Method of the nonlinear Filtering Problem.

[47] Y. Nishiyama, Some central limit theorems for l^{∞}-valued semimartingales and their applications. Probability Theory and Related Fields, Vol 108, pp. 459–494, 1997.

[48] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*. Springer, 1991.

[49] J. Shapcott. Index Tracking: Genetic Algorithms for Investment Portfolio Selection. EPCC-SS92-24, September 1992.

[50] T. Shiga and H. Tanaka (1985) ; Central limit theorem for a system of Markovian particles with mean field interaction. Z.f. Wahrscheinlichkeitstheorie verw. Geb., 69, 439-459.

[51] M. Schoenauer, L. Kallel and F. Jouve (1996); Mechanical Inclusions Identification by Evolutionary Computation. *European Journal of Finite Elements*, 5 (5-6), 619-648.

[52] M. Schoenauer, F. Jouve and L. Kallel (1997); Identification of Mechanical Inclusions. *In D. Dagsgupta and Z. Michalewicz, Editors, Evolutionary Computation in Engeneering*, 477-494, Springer Verlag.

[53] D. Treyer, D.S. Weile and E. Michielsen, The Application of novel genetic algorithms to electromagnetic problems. Applied Computational Electromagnetics Symposium Digest. Monterey, CA, Marsh 17-21, 1997, Vol. 2, pp. 1382-1386.

[54] Trouvé, *Parallélisation massive du recuit simulé*, Thèse de Doctorat, Université Paris XI, Janvier 1993.

Laboratoire de Statistiques et Probabilités
Université Paul Sabatier
118, route de Narbonne
31062 Toulouse cedex, France
E-mail: delmoral@cict.fr
E-mail: miclo@cict.fr