MODELING GENETIC ALGORITHMS WITH INTERACTING PARTICLE SYSTEMS

P. Del Moral[*]– L. Kallel[†]– J. Rowe[‡]

Abstract

We present in this work a natural Interacting Particle System (IPS) approach for modeling and studying the asymptotic behavior of Genetic Algorithms (GAs). In this model, a population is seen as a distribution (or measure) on the search space, and the Genetic Algorithm as a measure valued dynamical system. This model allows one to apply recent convergence results from the IPS literature for studying the convergence of genetic algorithms when the size of the population tends to infinity.

We first review a number of approaches to Genetic Algorithms modeling and related convergence results. We then describe a general and abstract discrete time Interacting Particle System model for GAs, and we propose a brief review of some recent asymptotic results about the convergence of the N-IPS approximating model (of finite N-sized-population GAs) towards the IPS model (of infinite population GAs), including law of large number theorems, $\mathbb{L}^p$-mean and exponential bounds as well as large deviations principles.

Finally, the impact of modeling Genetic Algorithms with our interacting particle system approach is detailed on different classes of generic genetic algorithms including mutation, cross-over and proportionate selection. We explore the connections between Feynman-Kac distribution flows and the simple genetic algorithm. This Feynman-Kac representation of the infinite population model is then used to develop asymptotic stability and uniform convergence results with respect to the time parameter.

Keywords: Genetic algorithms, Interacting particle systems, asymptotical convergence, Feynman-Kac formula, measure valued processes.

Resumen

[*]UMR C55830, CNRS, Bat.1R1, Université Paul Sabatier, 31062 Toulouse cedex, France; E-Mail: delmoral@cict.fr.

[†]CMAP, Ecole Polytechnique, 91128 Palaiseau cedex, France; E-Mail: Leila.Kallel@polytechnique.fr

[‡]University of Birmingham, Birmingham B15 2TT, United Kingdom; E-Mail: J.E.Rowe@cs.bham.ac.uk

En este trabajo presentamos un enfoque natural de Sistemas de Partículas Interactuantes (IPS) para modelar y estudiar el comportamiento asintótico de Algoritmos Genéticos (GAs). En este modelo, una población es vista como una distribución (o medida) en el espacio de búsqueda, y el Algoritmo Genético como un sistema dinámico valuado en medida. Este modelo permite aplicar resultados recientes sobre convergencia de la literatura sobre IPS para estudiar la convergencia de GAs cuando el tamaño de la población tiende al infinito.

Primero revisamos algunos enfoques para modelar GAs y resultados relacionados con la convergencia. Enseguida describimos un modelo general y de tiempo discreto abstracto para GAs, basado en un IPS, y proponemos una breve revisión de algunos resultados asintóticos recientes acerca de la convergencia de N-IPS modelos de aproximación (de GAs de población finita de tamaño N), que conducen al modelo IPS (de GAs de población infinita), incluyendo teoremas de leyes de los grandes números, LL^p-media y cota exponencial, así como principios de grandes desviaciones.

Finalmente, se detalla el impacto de modelar Algoritmos Genéticos con nuestro enfoque de IPS sobre diferentes clases de algoritmos genéticos genéricos que incluyen mutación, cruzamiento y selección proporcional. Exploramos las conexiones entre los flujos de distribución de Feynman-Kac y el algoritmo genético simple. Esta representación de Feynman-Kac del modelo de población infinita es usada luego para desarrollar resultados de estabilidad asintótica y convergencia con respecto al parámetro de tiempo.

Palabras clave: Algoritmos genéticos, sistemas de partículas interactuantes, convergencia asintótica, fórmula de Feynman-Kac, procesos valuados en medida.

1   Introduction

Evolutionary algorithms (abbreviate EAs) are a class of stochastic optimization techniques that have been successfully applied in diverse areas, such as machine learning, combinatorial problems, and numerical optimization. This success, has initiated the development of various EA variants, and stimulated the theoretical research about convergence properties of these algorithms.

The history of EAs goes back to the sixties where genetic algorithms were introduced for solving biological adaptation problems [54]. Today, EAs include Evolution Strategies [22] first designed for real-valued parameter search spaces, Evolutionary Programming [41, 40], Genetic Programming [59] designed for tree search spaces, simulated annealing [91, 89] and genetic algorithms [47] first designed for discrete parameter search spaces.

While most differ slightly in their actual implementations, all these evolutionary techniques use the same metaphor of mapping problem solving onto a model of evolution, where a candidate solution is represented by an *individual* and the solution quality is determined by a *fitness* function. The evolutionary process consists in evolving a vector (or population) of individuals, under stochastic modification and selection operators (see figure 1).

In contrast to many papers in the literature of Evolutionary Algorithms, we don't follow the above described historical classification of algorithms. Rather, we review convergence results classified according to the model (e.g. Markov, schema, allele distributions),
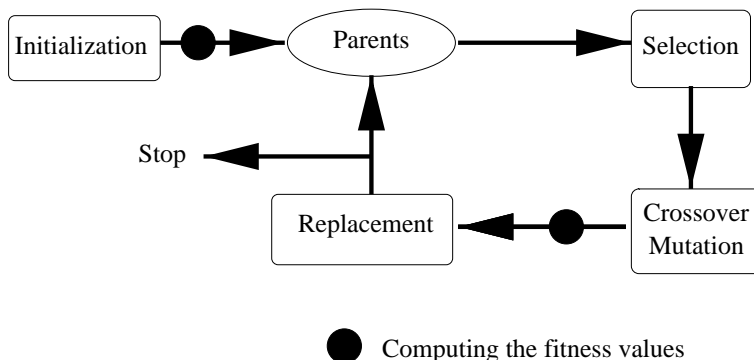
Figure 1: A general description of Evolutionary Algorithms. – *Initialization* consists in selecting, usually at random and uniformly, a population of N individuals from the search space. – *Selection* consists in choosing a number (in the range $[1, N]$) of high fitness members that undergo Mutation and/or Crossover. – *Mutation* is an elementary modification operator. – *Crossover* combines the properties of two parents to produce an offspring. – In the *replacement* step, a new population is selected among the current and modified individuals, such that the fittest individuals are more likely to survive.

space (finite, infinite) and operators (mutation, crossover). Then, we detail an alternative Markov model GAs valid for finite or infinite search spaces, static or dynamic fitness environment, homogeneous or inhomogeneous operators. An overview of our approach is presented later on in this introduction.

1.1   Summary of the paper

We first review the modeling of homogeneous GAs in the binary space $E = \{0, 1\}^\ell$, where an individual is a sequence of binary bits (called *alleles*). We describe approaches studying the evolution of population schemata ; population mean and variance in fitness ; population alleles marginal distribution.

  We then point out the limitations of these approaches: the first approaches applies to infinite populations, the second applies to particular fitness functions, the third only models a particular type of GAs – basicly selection and crossover algorithms with no mutation.

  We also review strong convergence results, about the complexity of the time to absorption, that have been obtained for simplified and rudimentary GAs (selection only or mutation only).

  The main general theoretical results in the field of evolutionary function optimization have been obtained using Markov chain analyses in finite spaces. These results prove the asymptotical absorption into optimal states w.r.t. time for the homogeneous GA ; and prove a finite time visits to the optimal set for the inhomogeneous GA. In this Markov framework, the state space is the set of all populations.

  An alternative Markov approach considers the state space of all probability distributions of the populations, rather than the space of populations. It proves the existence of

fix points for population distributions, for an infinite population size. In this framework, a distribution is a vector of length 2^ℓ, describing the proportion of each point of $E = \{0,1\}^\ell$ in the population.

In this paper we propose an alternative Markov approach on the space of population distributions. But, in contrast to the latter approach we consider the empirical measures (2) associated to the population (1), resulting in distributions having a small variable support, which length is equal to the population size.

A first advantage of this approach is that it also applies to infinite and continuous spaces. A second advantage is that it allows to build a bridge between the fields of genetic algorithms and that of interacting particle systems.

An important part of this paper is then devoted to applying some recent convergence results from the IPS literature to the simple mutation-only genetic algorithm. Note that the term convergence here is twofold. We (briefly) consider the long time behavior of GAs, and (widely) investigate the convergence quality and rate of the genetic algorithm, seen as a measure valued process, when the population size tends to infinity.

For example, when the population size tends to infinity, in some sense to be defined, the difference between the finite population empirical measure at time t and the measure it would have if the population size were infinite, goes to zero in $\mathbb{L}^p$-mean. Analogous results also apply to the path space, involving a sequence of measures taken from time 0 to n, and provide large deviation bounds.

On the other hand, when the time parameter goes to infinity, we show that under some conditions, the distribution of GA (infinite) population at time n converges exponentially fast to a unique fix point.

Last but not least, we generalize the IPS approach so as to cope with different variants of genetic algorithms.

For example, we consider mutations where the transition probabilities depend on the whole current population (and not only the point to be mutated). Such transition kernels naturally include special crossover mechanisms. We also investigate the convergence of GAs with commonly used crossover operators.

We end the paper by proposing a new dynamical system, corresponding to GAs with an original generational scheme: selection-or-mutation transitions rather than the common selection-and-mutation ones.

1.2   The IPS approach: preliminaries

In the following we outline the IPS approach, before presenting the related sections' contents.

Genetic algorithms can be defined as a system of particles (or individuals) evolving randomly in a given measurable space $(E, \mathcal{E})$ and undergoing adaptation in an environment, not necessarily time homogeneous, represented by a collection of fitness functions.

In discrete time settings these stochastic algorithms can be modeled as a Markov chain

$$\xi_n = \left(\xi_n^1, \ldots, \xi_n^N \right), \qquad n \geq 0, \tag{1}$$

taking values in the product state space E^N where $N \geq 1$ denotes the number of particles. Although it is not really essential the initial particle system ξ_0 usually consists in N independent particles with common law $\eta_0 \in \mathcal{P}(E)$, where $\mathcal{P}(E)$ denotes the set of all probability measures on E endowed with the weak topology.

Rather than studying the dynamic and limiting process of the so defined Markov chain (1), we propose to study the flow of the empirical measures associated with the systems of particles ξ_n

$$m(\xi_n) \stackrel{\text{def.}}{=} \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi_n^i} \tag{2}$$

where δ_a stands for the Dirac measure at $a \in E$.

The transition from the population ξ_{n-1} at time $(n-1)$ to the next population ξ_n only depends on the empirical measure $m(\xi_{n-1})$. More precisely, given the configuration ξ_{n-1} at time $(n-1)$ the next generation ξ_n consists of N (conditionally) independent random variables with common law

$$\Phi_n \left(m(\xi_{n-1}) \right)$$

where $\Phi_n : \mathcal{P}(E) \to \mathcal{P}(E)$, $n \geq 1$, is a given collection of sufficiently regular functions, that are dictated by the GA operators and parameters (see section 7.1).

This quite general interacting particle system model has been introduced in [10, 11]. Its asymptotic behavior as $N \to \infty$ is now well understood. In some sense to be defined the empirical measures

$$\eta_n^N \stackrel{\text{def}}{=} m(\xi_n), \qquad n \geq 0,$$

converge as the number of particles $N \to \infty$ to a deterministic flow of distributions

$$\eta_n \in \mathcal{P}(E), \qquad n \geq 0,$$

solution of the measure valued dynamical system

$$\eta_n = \Phi_n \left(\eta_{n-1} \right), \qquad n \geq 1, \tag{3}$$

In measure valued processes literature the system (3) is usually called the limiting process. Such measure valued dynamical systems have arisen in such diverse scientific disciplines as physics, biology, evolutionary computing, nonlinear filtering and elsewhere. In GA settings this measure valued dynamical system is sometimes referred as the infinite population model and it is used to predict the behavior of the finite population model. In advanced signal processing and particularly in filtering literature (3) represents the evolution in time of the conditional distributions of a signal given the observation record. Incidently the N-IPS approximating scheme of the nonlinear filtering equation is defined in terms of a time inhomogeneous and simple GA.

The continuous time version of this model is described in [7, 8] revealing a very striking analogy between GA, the robust and path-wise nonlinear filter, Feynman-Kac formulae and spatially homogeneous Boltzmann equation.

The objective of this paper is not to develop the details of all these connections between these research fields, the interested reader is referred to the monograph [7] and to [8]. Our

aim is to introduce the reader to IPS approximations of nonlinear measure valued processes of the form (3) and to present the mathematical theory that it is useful in studying the asymptotic behavior of GAs.

1.3   Organization of the paper

This paper is organized as follows.

Section 2 summarizes a number of approaches to GA modeling, so far encountered during the last decade, excluding the Markov models that are described in a separate section. These approaches does not lead necessarily to strong convergence results, but give insights into some aspects of GA dynamics under some approximations or in particular fitness cases.

Section 3 reviews results obtained on simplified and rudimentary processes, such as selection only or mutation only algorithms. These simplifications have the advantage of making theoretical analysis very amenable and allow one to compare complexity of time to convergence with different selection or mutation schemes. On the other hand, results obtained with such simplifications cannot be generalized to more complex and realistic (working) GA instances.

Section 4 presents different exact Markov models leading to the proof of the convergence of different classes of Genetic Algorithms towards the optima of the fitness function. While these results put almost no restrictions on the fitness function, they cannot give the order or complexity of convergence time.

Having introduced different models for Genetic Algorithms, related results and limitations, we introduce the IPS model for infinite populations, which can be seen as a generalization of the infinite population model of Vose [96], to the case of infinite search spaces. We then show how the N-IPS approximating model can be used to study N-sized-population GAs.

A first advantage of the IPS model is that it allows us to model a wide variety of Genetic Algorithms, as shown in section 7. A second advantage is the existence of a number of recent theorems from probability literature that guarantee the convergence of the finite population dynamic of GAs (N-IPS approximating model) towards the infinite population dynamic (IPS model), for large populations uniformly with respect to time (these results are summarized in section 6 for an abstract and general IPS model). Finally, the Feynman-Kac interpretation allows us to derive the exact analytical distribution of GA population members.

In section 6 we describe precisely the IPS approximating model and we review recent results on the convergence of such IPS methods when the number of particles (or population size) goes to infinity, including weak convergence theorems, $\mathbb{L}^p$-mean error estimates, exponential rates and large deviation principles.

Most of the limit theorems presented in this preliminary section 6 result from collaborations of one of the authors with Alice Guionnet, Michel Ledoux and Laurent Miclo. Only a selection of existing results is presented. More information and detailed proofs can be founded in the set of referenced papers [1, 2, 3, 4, 6, 7, 10, 11].

All these developments offer the appropriate theoretical background to analyze the asymptotic behavior of GA. They also permit us to construct new genetic-type methods.

In section 7 we present several generic GAs which fit into our framework including the simple GA with classical mutation and proportionate selection, but also GAs with interacting mutation, cross-over transitions and GAs with randomly ordered selection/mutation. We will also give a detailed illustration of how the convergence results presented in section 6 may be applied to study the asymptotic behavior of such GAs.

It turns out that many GAs including cross-over transitions fit into the simple GA framework. For these reasons, and because of the importance of the simple GA in practice, section 7.1 is built around this theme. We complement the limit theorems of section 6 and we present several additional asymptotic theorems including a Central Limit Theorem, a limit theorem on the long time behavior of the limiting system as well as a uniform convergence theorem for the N-IPS approximating model with respect to the time parameter.

We hope this paper will be useful to our colleagues working on GAs and evolutionary computing.

2   Different approaches to GA modeling

This section briefly presents a number of tentative approaches to GA modeling, excluding the Markov models that are described in the next section. We also present some alternative algorithms that arose from some of these modeling approaches. Such alternative algorithms also intend to achieve a stochastic optimization task, but at a reduced cost compared to the standard Evolutionary Algorithms described in figure 1. Most of these approaches consider search spaces of binary strings $E = \{0,1\}^\ell$, where a string is viewed as a vector of ℓ binary *alleles*.

Section 2.1 recalls and comments on the schema theory, modeling the evolution of schemata. Section 2.2 presents and comments on an approach borrowed from physics, approximating the evolution of population fitness distributions by some of their moments. Section 2.3 focuses on the evolution of allelic marginal distributions. This latter point of view has been given much attention in the last 5 years, and led to a number of alternative algorithms, that are not guaranteed to find the global optimum, but which might converge quicker than GAs.

2.1   The Evolution of Schemata

Initial investigations about binary Genetic Algorithms (GAs) concentrated on a macroscopic view of the algorithms, based on the evolution of schemata (rather than strings or proportions of strings, as in the Markov framework). A first version of the schema theorem was proved in [55] then an extension in [47].

The framework is the following. Consider the binary space $E = \{0,1\}^\ell$, a proportionate selection, a 1-point crossover and a $1/\ell$ mutation.

A schema is defined as a subspace $\mathcal{H}$ of E where a number, denoted $o(\mathcal{H})$, of string positions have fixed values. For example, the schema $(1*11) = \{1011, 1111\}$ has 3 fixed

values.

Theorem 2.1 ([55, 47]) *Let $\mathcal{N}(\mathcal{H}, n)$ be the number of representatives of $\mathcal{H}$ in the population at time step n, then, neglecting the possible reconstruction of new representatives of $\mathcal{H}$ by mutation or crossover,*

$$\mathbb{E}[\mathcal{N}(\mathcal{H}, \mathbf{n+1})] \geq \mathcal{N}(\mathcal{H}, \mathbf{n}) \frac{\overline{\mathbf{f}}(\mathcal{H}, \mathbf{n})}{\overline{\mathbf{f}}(\xi_{\mathbf{n}})} \left[\mathbf{1} - \mathbf{p_m o}(\mathcal{H}) - \mathbf{p_c} \frac{\mathbf{l}(\mathcal{H})}{\ell - \mathbf{1}}) \right], \qquad (4)$$

where $\bar{f}(\mathcal{H}, n)$ denotes the average fitness of representatives of $\mathcal{H}$ present in the population ξ_n, and $l(\mathcal{H})$ denotes the distance (number of positions) between the first and last fixed positions of $\mathcal{H}$ ($l(\mathcal{H}) = 4$ in the previous example), p_m and p_c denote the mutation and crossover probabilities.

Under infinite population approximation, the expectation of the number of schema representatives can be replaced by its actual value in (4), one obviously conclude that schemata with above average fitness ($\bar{f}(\mathcal{H}, n) > \bar{f}(\xi_n)$) tend to colonize the population.

The usefulness of the Schema Theory has been seriously questioned [52, 93, 71, 63, 19, 97, 36, 22]. One of the main limitations of the Schema theory is that it is only valid for one generation prediction and cannot be iterated more than once (as it maps a schemata into the next expected one only), hence cannot be used for studying the long term behavior.

Recent work [17, 84, 85, 96] prove a new relation for the evolution of schemata, taking into account the effect of schema reconstruction by crossover and mutation (and so accounting for possible low fitted schemata that yield high fitted offspring), but still falling under the limitation of relating the expectation of $\mathcal{N}(\mathcal{H}, n+1)$ to $\mathcal{N}(\mathcal{H}, n)$ as explained above.

Note finally that, at its current stage, the schema theory still doesn't lead to any strong theoretical result about convergence, or complexity of the algorithm itself.

2.2   Cumulants of Population Fitnesses

In the statistical mechanics formulation [29, 99, 72], a population is described in terms of a few macroscopic quantities: the *cumulants* of the population *magnetization* distribution. The *cumulants* denote the mean k_1, variance k_2 (and eventually moments of higher order) of the population *magnetization* distribution. The magnetization of a string is defined as

$$M(x) = \sum_{i=1}^{\ell} x_i \quad \text{where} \quad x_i = \{0, 1\}. \qquad (5)$$

For most problems the fitness distribution is not related to the magnetization distribution. However for simple examples such as one-max, magnetization equals fitness. Thus many papers study the evolution of expected population cumulants, providing recurrence relations between the expected cumulants when mutation (resp. selection, or crossover) is applied to population members, in a generational scheme (see [14] for a detailed review of the approach and simplification hypothesis).

Under the usual assumption that only the – 2,3 or n – first order cumulants are non null, [99, 100] prove the following relation when N individuals are selected following Boltzmann selection with small strength β, the change in cumulants – when terms of higher powers of β are truncated – follows from [99, 100].

$$
\begin{aligned}
< \kappa_1 >_S &\simeq\ < \kappa_1 > + < \kappa_2 > \beta, \\
< \kappa_2 >_S &\simeq\ < \kappa_2 > - \frac{< \kappa_2 >}{N} + < \kappa_3 > \beta.
\end{aligned}
$$

where the subscript S denotes that this is the effect of selection.

When mutation is applied to all population members with bit-wise-flipping probability c/ℓ.

$$
\begin{aligned}
< \kappa_1 >_M &= (1 - 2c) < \kappa_1 >, & (6) \\
< \kappa_2 >_M &= (1 - 2c)^2 < \kappa_2 > + \ell(1 - (1 - 2c)^2), & (7)
\end{aligned}
$$

where the subscript M denotes that this is the effect of mutation, and $< . >$ denotes the expected cumulant value over an *ensemble* (infinite set) of populations evolving in parallel.

Combining these expressions, to find the expected cumulants after selection, mutation (and eventually crossover (15)) is straight-forward for the onemax function, but can be unwieldy for more complex functions, in such cases the formulae are simulated numerically. This approximated model has been extended to other generational and selection schemes and also demonstrated to be quite close to the actual behavior of GAs averaged over many runs [75, 14].

This approach has also been applied to study the dynamics of GAs on functions different from one-max, but in most cases, the fitness is a function of the magnetization (or Hamming distance to all-ones optimum) [82, 76].

At the cost of some approximations, this statistical mechanics model allows us to gain significant insights into the population dynamics on a number of fitness functions. These results would be difficult to obtain at this stage using exact models of GAs such as Markov chains. Indeed, apart from particular cases [57, 43], Markov chain studies of GA exact dynamic consider populations of size one and no crossover, on simple functions (see section 5).

2.3   Allelic Marginal Distributions

This modeling approach is in some ways similar to the one we are presenting in the paper – namely, the interacting particle systems model (IPS). However, the goals are different. The IPS model is used to study the convergence of the dynamic of finite population GAs towards that of the infinite population GAs, whereas the following model is used to construct alternative stochastic algorithms, with a smaller time to convergence than GAs.

The basic idea is to transform the initial search problem over E into a search problem over probability distributions on E. The target probability distributions are then mixtures of Dirac distributions charging optimal solutions of the initial search problem over E.

The model we describe in this section was first proposed in practical numerical simulations [23]: selected population members are no longer mutated and mated, but used to update the allelic marginal distributions p_i, as follows.

$$p_i(x_i, t+1) = p_i(x_i, t) + \lambda\big(p_i^s(x_i, t) - p_i(x_i, t)\big), \tag{8}$$

where $p_i^s(x_i, t)$ are the marginal frequencies of allele x_i in the selected population members. The new population is then constructed by drawing N strings $x = (x_1 \ldots x_\ell)$, according to the probability profile

$$p(x, t+1) = \prod_{i=1}^{\ell} p_i(x_i, t+1).$$

The convergence speed of this algorithm (8) critically depends on λ. For $\lambda = 0$ we obtain a random search, for $\lambda = 1$ we obtain the so called UMDA algorithm [65]. Mühlenbein formalizes the UMDA algorithm, in the infinite population framework, with proportionate selection (10) and complete crossover, but no mutation. He shows that this algorithm consists in performing a gradient search with the potential $W(t) = \sum_x p(x, t)f(x)$ [65], (that is, the average fitness),

$$p(x_i, t+1) = p(x_i, t) + p(x_i, t)(1 - p(x_i, t))\frac{\partial W/\partial p(x_i)}{W(t)} \tag{9}$$

where $p(x_i, t)$ denotes the probability that $x_i = 1$ at time t. However, this gradient is not easy to implement directly as $W(t)$ requires the knowledge of 2^ℓ parameters.

In order to search directly in the space of distributions, another approach consists in reducing the complexity of the distributions space. This can be achieved by restricting the search to a family of parametrized distributions, allowing one to simply search in the parameter space. This is a typical approach in the learning field.

This approach is illustrated in [24] where the vector $x = (x_1, \ldots, x_\ell)$ of E is replaced with a vector of ℓ random variables with a vector of respective parametrized density profiles

$$p(x, \theta) = (p_1(x_1, \theta_1), \ldots p_\ell(x_\ell, \theta_\ell)),$$

where $p_i(., \theta_i)$ defines the distribution of allele x_i over $\{0, 1\}$, parametrized by θ_i. For example, numerical simulations of [24] assume $p_i(., \theta_i)$ as independent Bernoulli laws (resp. gaussian) with average θ_i (resp. average and variance given by θ_i components).

The discrete optimization problem in E is then transformed into a continuous minimization problem of the Kullback-Leibler (KL) divergence [101] between p and the Gibbs distribution p_T^*, which charges the high fitness states of E for low temperatures T. It can be shown that this minimization can be achieved with a gradient in the space of the free energy of the system $F(\theta)$ [24]. Yet as for $W(t)$ above (9), computing the free energy requires the knowledge of 2^ℓ parameters.

In order to make this gradient minimization computationally amenable, Berny [24] proposes approximate (and stochastic) update (gradient following) rules for θ, allowing him to successfully minimize the free energy on a number of fitness functions, and for

different temperatures. However, due to the approximations, the asymptotical convergence to the global optima is not guaranteed on a given fitness landscape.

Note finally that more robust algorithms could be obtained, if dependent random variables were considered for allelic distributions, as pointed out in [24] and [65].

3   Convergence of simplified algorithms

In any evolutionary algorithm, the strong interactions between selection and evolution operators (crossover, mutation, ...) make it difficult to efficiently compare the effects of different operators without taking into account the characteristics of the selection (i.e. the fitness function).

A standard method for avoiding such a bias, is to start by considering the effects of each operator alone: Selection without evolution operators (see the studies of *takeover time* [49, 34]), or *innovation time* of crossover or mutation in the absence of selection [88, 48], or the evolution of random walks directed by some mutation [45] or crossover [31] operators.

In the following, for the sake of clarity, the reported results do not always cover the whole set of extensions and generalizations that could have been achieved in the papers, but mainly focus on the techniques used to find the results (for example we might omit to mention extensions of formulae to r-ary alphabets).

3.1   Selection only algorithms

Consider a finite population of N members $\xi \in E^N$ and suppose some members of the population ξ are selected possibly more than once in order to form a new population of size N. This defines a rudimentary dynamical process and we are usually interested in calculating the time required to reach a population with N copies of a same member.

$$\xi_{n-1} = \left(\xi_{n-1}^1, \ldots, \xi_{n-1}^N\right) \xrightarrow{\text{Selection}} \xi_n = \left(\xi_n^1, \ldots, \xi_n^N\right)$$

This is a widely studied in both the literatures of genetic algorithms and populations genetics. However, the approaches and selections involved differ significantly.

In the context of genetic algorithms, we are interested in problem solving, hence selection directly depends on the fitness of population members: high fitted members are more likely to be selected than low fitted ones. Typically, with a proportionate selection the probability to select the i-th individual of ξ_{n-1} is given by

$$\frac{f\left(\xi_{n-1}^i\right)}{\sum_{j=1}^N f\left(\xi_{n-1}^j\right)} \tag{10}$$

At each iteration, N individuals are selected following (10), that build up the population ξ_n (see Table 3.1 for alternative typical iterations). Other typical selections are Ranking selection (same as (10) replacing fitness values by their rank in the population), and tournament selection (p members are chosen uniformly, and the fittest is selected).

In the genetic algorithms literature, early studies use Markov chains to numerically calculate the expected absorption time. This is achieved by calculating the visitation matrix $V = (I - Q)^{-1}$, after partitioning the states of the absorbing Markov chain (ξ_n) associated with the selection algorithm,

$$P = \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix}.$$

Summing over the set of transient states TS, $\sum_{j \in TS} V_{ij}$, yields the expected absorption time for a process starting at state i. This method was first applied in the space $E = \{0, 1\}$ with proportionate selection [51].

The impact of selection with sharing [1] has been investigated using the same method and space $E = \{0, 1\}$ [56], demonstrating that the expected absorption time is significantly larger with sharing (simulated for various small population sizes and various ratios $f(1)/f(0)$).

An original issue is also addressed in the same paper [56]. The author studies the "quasi-equilibrium state" of the chain before absorption (as absorption time is very large with sharing). This corresponds to finding the quasi-stationary distribution of matrix Q (as defined in [35] for near-ergodic absorbing Markov chains).

Also in the Markov framework, [60] considers the general case of $E = \{0, 1\}^\ell$ and shows, under some simplifications, that *Boltzmann tournament selection* have the same effect of maintaining the diversity in the populations resulting in a larger expected absorption time.

Still from the GA literature, a different approach to selection algorithms approximates the recurrence between the proportion p_n of the optimum x^* at each generation (valid for infinite populations),

$$p_n \simeq \frac{f(x^*) \, p_{n-1}}{\bar{f}(\xi_{n-1})}$$

In the space $E = \{0, 1\}$ [18] solves this recurrence, showing that starting from a random uniform population, expected time to have a proportion $1 - 1/N$ of the optimum (say individual 1) is

$$\frac{\ln(N - 1)}{\ln(f(1)/f(0))} \tag{11}$$

Again under the infinite population approximation, different types of selections are studied and compared in the continuous space $E = [0, 1]$ [49]. Results are summarized below, showing that starting from a population containing one optimum, the expected time to reach a proportion $1 - 1/N$ of the optimum, have a complexity ranging from $O(\ln N)$ to $O(N \ln N)$.

---

[1]Sharing [50] tends to spread the population out over multiple peaks in proportion to the height of the peaks. For example, if there are m individuals y_k in a Hamming radius of σ from local optimum x, their fitness is divided by $\sum_{k=1,\dots,m}(1 - d(x, y_k)/\sigma)$

Generational Selection	Takeover time	
Proportionate	$\dfrac{1}{c}(N \ln N - 1),$	for $f(x) = x^c$
	$\dfrac{1}{c}(N \ln N),$	for $f(x) = \exp(c\,x)$
Linear ranking	$\dfrac{1}{\ln 2}(\ln N + \ln \ln N),$	for $c = 2$
by $r(f(x)) = c - 2(c-1)x,$	$\dfrac{2}{c-1}log(N-1),$	for $1 < c < 2$
p-**Tournament**	$\dfrac{1}{\ln p}(\ln N + \ln \ln N)$	

In the population genetics literature, selection is usually uniform on the population. The progressive loss of diversity in the population is then called *genetic drift*. Analysis of genetic drift is often performed by calculating the Markov chain transition matrices in order to analytically approximate the time to absorption [62, 58].

An alternative approach to genetic drift consider the rate of decrease in population fitness variance under uniform selection [68, 67, 72], and lead to an exact analytical approach in many cases (see [74] for a quick overview). The ratio between population variances after one generation is given below (the first is exact and the others are approximations accurate up to terms of order $1/N$).

Iteration loop	ratio of fitness variance of successive populations
Generational (choose N members independently and uniformly in ξ_{n-1} to form ξ_n)	$1 - 1/N$
Generational gap G (choose GN members uniformly, that replace GN members uniformly chosen)	$1 - (2-G)/N,$ for $0 \leq G \leq 1$
CHC type (duplicate each member of ξ_{n-1}, then delete N members uniformly chosen)	$1 - 1/(2N)$

These studies are half way between population genetics (as the selection is fitness-independent) and genetic algorithms as different typical iterations are compared for the same selection. These results show that genetic drift for a CHC iteration is slower (at half the rate) than for the traditional generational iteration, which in turn is slower than generational gap and steady-state (obtained for $G = 1/N$) iterations.

Genetic drift is usually seen as the responsible for the often observed problem of premature convergence. That's why such results obtained for selection algorithms can be helpful in guiding GA practicers in the choice of selection schemes and iteration loop according to the problem at hand.

3.2 Random walks on operator-neighborhood graph

Consider a finite population of N members, $\xi \in E^N$, and suppose each member of the population undergoes a mutation, transforming ξ_{n-1}^k into ξ_n^k, for $k = 1, \ldots, N$.

$$\xi_{n-1} = \left(\xi_{n-1}^1, \ldots, \xi_{n-1}^N\right) \xrightarrow{\text{Mutation}} \xi_n = \left(\xi_n^1, \ldots, \xi_n^N\right) \tag{12}$$

In the binary space $E = \{0,1\}^\ell$, two typical mutations are compared [45]. For large ℓ values, it is shown that the hitting time of solution obeys an exponential distribution with respective means

$$\frac{2^\ell}{N(1 - e^{-c})}, \quad \text{for the } c/\ell \text{ mutation (flips each bit with probability } c/\ell),$$

$$\frac{2^\ell}{N}, \qquad \text{for the 1-bit mutation (flips exactly 1 bit uniformly chosen).}$$

In other words, this result shows that the widely used (positive) c/ℓ mutation is clearly outperformed by the simple 1-bit mutation (in terms of the probability that the optimum has already been hit) at each time step.

Again in the binary space $E = \{0,1\}^\ell$, but for large population sizes and only for the c/ℓ mutation, the whole dynamic of the algorithm can be described in terms of the distribution of the population magnetization (average number of ones in population members, as detailed in section 2.2). Equations (6) and (7) provide an approximate recurrence relation between the expected average and variance of the population magnetization at successive generations.

The case of populations evolving under crossover has also been addressed. Consider N random couples selected uniformly and independently in ξ_{n-1}. Each couple is mated, resulting in N offspring that build up the next population.

$$\xi_{n-1} = \left(\xi_{n-1}^1, \ldots, \xi_{n-1}^N\right) \xrightarrow{\text{Crossover}} \xi_n = \left(\xi_n^1, \ldots, \xi_n^N\right)$$

In the context of infinite populations, general results can be found in [46, 30] which prove that all complete[2] crossover schemes lead to the same limit distribution, given by Robbins proportions of individuals in the limiting population.

$$\pi(x) = \prod_{i=1}^{\ell} p(x_i), \tag{13}$$

where $p(x_i)$ is the constant marginal distribution of the allele at position i. Note that these marginal distributions do not change when crossover is applied.

Recently, [31] proved that finite populations of r-ary strings under crossover converge in average to the populations with Robbins proportions. This result assumes that the application of crossover can produce any combination of parents alleles.

---

[2]Assuming that the application of crossover can produce any combination of parents' alleles.

The case of finite populations has been widely investigated [99, 100] using statistical mechanics techniques in the binary framework, introduced in section 2.2. Averaging over all the ways of performing bit-wise crossover, the expected mean and variance of population magnetization is given by the recurrence relations

$$\mu_c = \mu, \tag{14}$$

$$\sigma_c^2 = (a^2 + (1-a)^2)\sigma^2 + 2a(1-a)\frac{\ell(1-q)}{4}, \tag{15}$$

where a is the probability we choose an offspring allele from parent x and $(1-a)$ is the probability that we choose an offspring allele from parent y, and

$$q = \frac{1}{N(N-1)} \sum_{x \neq y} \left(\frac{1}{\ell} \sum_{i=1}^{\ell} (2x_i - 1)(2y_i - 1) \right),$$

is the pair-wise correlation between population members. Hence crossover increases the population magnetization variance towards $\ell(1-q)/4$ at a rate which is maximized for $a = 0.5$, and does not change the mean magnetization of the population.

As a conclusion, these results obtained for rudimentary versions of algorithms certainly allow one to obtain theoretical bounds on convergence for different operators and selections, but these results cannot be generalized to general GA instances using mutation, crossover and selection, nor do they allow us to prove the convergence of GAs as function optimizers, in a general function context.

4   Three Markov approaches to GA convergence

The main theoretical results in the field of evolutionary binary function optimization have been obtained using Markov chain analyses. The two first approaches are Markov chains in the space of the GA populations; the third approach is a Markov chain in the space of the distributions of population members in E.

4.1   Homogeneous Genetic Algorithms in finite search spaces

Historically, it has been quite natural to model GAs as Markov chains where the "state of the GA" at time step n is given by the current population (viewed as N individuals)

$$\xi_n = \left(\xi_n^1, \ldots, \xi_n^N \right). \tag{16}$$

The state space is then the product space E^N, where $N \geq 0$ denotes the number of individuals in a population, and the initial population ξ_0 usually consists of N random individuals uniformly chosen in E.

$$\xi_{n-1} \xrightarrow{\text{Selection}} \widehat{\xi}_{n-1} \xrightarrow{\text{Mutation}} \widehat{\widehat{\xi}}_{n-1} \xrightarrow{\text{Crossover}} \xi_n$$

We shall give the transition probabilities in the case of a generational GA, with proportionate selection, fitness function f, and mutation kernel K from E into itself.

At each iteration, N individuals are selected following (10), that build up the population $\widehat{\xi}_{n-1}$. Each member of the selected population $\widehat{\xi}_{n-1}$ undergoes a mutation. The transition can be summarized, for each $n \geq 0$, as follows

$$\mathbb{P}\left(\xi_{n+1} = x \,|\, \xi_n = y\right) = \prod_{p=1}^{N} \sum_{i=1}^{N} \frac{f\left(y^i\right)}{\sum_{j=1}^{N} f(y^j)} \; K\left(y^i, x^p\right), \tag{17}$$

where $K\left(y^i, x^p\right)$ gives the probability that individual $y^i \in E$ is mutated into x^p.

The mutation operator is at the core of all convergence results obtained so far in the GA literature, as irreducible mutation is often responsible for the reachability of all states in a finite time. More precisely, we have the following theorem by well known Markov Chain results, proving the convergence with probability one towards optimal states.

Theorem 4.1 ([53]) *If the sequence of populations-best-fitness is monotone (increasing if maximizing) with respect to time; and if any point of E is reachable by means of mutation and recombination in a finite number of steps, then*

$$\mathbb{P}(lim_{n \to \infty} x^* \in \xi_n) = 1. \tag{18}$$

Note that the monotonicity can easily be achieved by considering an elitist selection or an adequate iteration loop guaranteeing the survival of the best individual at each generation. For example, with a steady-state iteration, at each iteration an individual (chosen with any selection) undergoes mutation and crossover, then replaces the individual with the worst fitness in the population.

Analogous general theorems exist for large classes of evolutionary algorithms in infinite real-valued search spaces, as Evolutionary Programming and Evolution Strategies. In fact, these algorithms can be seen as particular cases of the global random search algorithm which convergences with probability one towards optimal states [98]. The reader is referred to [22] (pages 48–51) for a discussion of the connection between these continuous space algorithms and the uniform and global random search algorithms.

This Markov chain model has been adopted in many papers, where some convergence proofs are provided for specific GA instances mainly in the space $E = \{0, 1\}^{\ell}$ and for the widely used positive $1/\ell$ mutation.

For example, Rudolph's convergence results [78, 77, 80] directly rely on the positivity[3] of the mutation operator, while their extension by Agapie [16] relaxes the strong hypothesis of positive mutations to the irreducible[4] and diagonal positive transition mutation matrix. Both these authors emphasize that the monotonicity condition can be fulfilled virtually by maintaining and updating the best so far individual in the population without it taking place in the evolution process. In other words, the sequence of best fitnesses found so far by GA populations converges almost surely to the optimal value of the fitness function,

---

[3]there exists a mutation that links any two points of the space with non-zero probability
[4]Existence of a finite chain of mutations linking any two points of the space with non-zero probability.

when an ergodic GA is used. On the other hand, Agapie [16] shows that the monotonicity condition is necessary for the convergence with probability one (as in equation 18) of the homogeneous GA with proportionate selection and irreducible mutation.

Note finally that nearly all convergence results for the homogeneous GA prove that the best so far individual of the population enters the optimal set with probability one (18). On the other hand, the inhomogeneous parameters of Cerf's GA (next section), allows him to prove first, the absorption of all population members towards the optimal set almost surely, and second, finite time visits to the optimal set.

4.2   Inhomogeneous Genetic Algorithms in finite search spaces

Also in the Markov framework, other results have been obtained with techniques, similar to those used to prove the convergence of simulated annealing [89, 90] : a genetic algorithms is seen as a stochastic perturbation (mutation, crossover) of a deterministic dynamical system (selection only). The rate of these perturbations goes to zero with time going to infinity.

Some preliminary results obtained by Davis and al. [37, 38], consider populations as a vector describing the number of occurrences of each individual in the population, (the same as the population vector described in next section for homogeneous GAs (21)). The authors prove the convergence of an inhomogeneous GA towards populations with identical individuals. More recently, Suzuki [87] builds on Davis model, and proves the asymptotical convergence to optimal states.

These results are largely extended by Cerf [32, 33] using the same Markov model as last section, where a population is defined as a vector of N individuals (16).

Using the Friedlin-Wentzell theory of stochastic perturbation of dynamical systems [61], Cerf obtains a lower bound for the population size allowing him to prove finite time convergence results for the inhomogeneous Genetic Algorithm detailed below. His results are primarily based on mutation (but he shows that an acceleration of convergence might be brought by crossover, in the sense that the lower bound on population size is smaller when crossover is used).

Cerf's results emphasize a decreasing mutation rate, tightly coupled to an increasing selection pressure, to reach finite time convergence (absorption of all the members of the population in the set of optimal points). More precisely, this result holds if the probability of mutating x into y is given by the kernel

$$\alpha_k(x,y) = \begin{cases} \alpha(x,y)k^{-a}, & \text{if } x \neq y \\ 1 - \sum_{z \in E,\, z \neq x} \alpha_k(x,z) & \text{if } x = y \end{cases}$$

and the probability of selecting an individual y_i in a population ξ is

$$\beta_k(y_i,\xi) = \frac{e^{c\,f(y_i)\,ln(k)}}{\sum_{j=1}^{m} e^{c\,f(x_j)\,ln(k)}} \#\{j, x_j = y_i\},$$

where α is an irreducible mutation kernel on E, and k is a parameter that tends to infinity such that λ, the radius of convergence of the series $S = \sum_{t \in \mathbf{N}} k(t)^{-\theta}$ is in the range $\theta 1 < \lambda < \theta 2$ ($\theta 2$ and $\theta 2$ are dictated by the fitness function).

Finally, Cerf provides a function N^* such that $N > N^*(\alpha, f, a, c)$ ensures the following results.

$$\forall x \in E^N, \qquad \lim_{n \to \infty} \mathbb{P}([\xi_n] \in f^* \,|\xi_0 = x) = 1, \tag{19}$$

$$\forall x \in E^N, \quad \mathbb{P}(\exists N, \; \forall n \geq N \;, \; [\xi_{T_n}] \in f^* \,|\xi_0 = x) = 1, \tag{20}$$

where T_n is the time of the n-th visit to the set of populations with equi-fitness values, f^* being the class of optimal populations. Equation (19) ensures the absorption in the optimal states with probability one, while (20) ensures finite time recurrent visits to populations with optimal states only.

An alternative inhomogeneous algorithm has been proposed [42], where the first $K(n)$ points of the population vector undergo a random mutation, and the others are replaced by the optimal point at each generation. The proof of convergence relies on the Perron-Frobenius theorem and shows that if K follows a binomial law with a parameter going to zero at the rate $p = n^{-1/D}$ (where D denotes the diameter of the mutation graph), and if the population size is large enough (a sufficient condition is $N > D$), then

$$\forall x \in E^N \quad \lim_{n \to \infty} \mathbb{P}([\xi_n] \in f^* \,/\xi_0 = x) = 1.$$

Note that these inhomogeneous GAs are not used in practice as the rate of decay of the perturbations is too slow to be implemented, and the required population size is too big. However, one can expect these results to be improved under some assumptions on the fitness function.

4.3   Dynamical Systems model for Homogeneous GAs

When the search space E is finite, and the fitness function is constant (homogeneous with respect to time), then there is a model describing the simple genetic algorithm (selection, mutation, and cross-over) called the *dynamical systems* model. The analysis is best developed for GAs with proportional selection, although others (rank-based, tournament) can be described within the framework. There has also been some work done on analysing fitness functions which are periodic with respect to time (see, for example, [107, 103, 106])

The existing theory is largely due to Michael Vose, and is described in detail in [96] (and see [104] for a simple introduction). The idea is to represent populations as vectors

$$p_n = (p_n^1, \ldots, p_n^d), \qquad n \geq 0, \tag{21}$$

taking values in the unit d-simplex $\Delta^d \subset \mathbb{R}^d$ with

$$d = \mathrm{Card}(E) \quad \text{and} \quad \Delta^d = \left\{ x \in \mathbb{R}^d \,;\; x_i \geq 0 \quad \sum_{i=1}^d x_i = 1 \right\},$$

where p_n^k denotes the proportion of the k-th individual of the search space in the population at time step n. We are then interested in the trajectory of the population through the simplex, which defines a Markov chain (p_n), $n \geq 0$. Notice that a vector $p \in \Delta^d$ has a dual interpretation. It represents an actual (finite) population, but it can be viewed as a sampling distribution for choosing a new population. In the infinite population framework, both these points of view can be confused.

Given a population p_n at time step n, we define an operator $\mathcal{G} : \Delta^d \to \Delta^d$ such that $\mathcal{G}(p_n)$ is the sampling distribution for the population at time step $n + 1$, so that in the inifinite population framework we have,

$$p_{n+1} = \mathcal{G}(p_n) \tag{22}$$

and in the finite population framework we have,

$$\mathbb{E}[p_{n+1}] = \mathcal{G}(p_n).$$

This operator $\mathcal{G}$ can be broken down into three parts, one for each of selection $(\mathcal{F})$, mutation $(\mathcal{U})$ and cross-over $(\mathcal{C})$, with

$$\mathcal{G} = \mathcal{C} \circ \mathcal{U} \circ \mathcal{F}$$

For certain typical definitions of these operators, it turns out that mutation and cross-over commute. The combined effect of these operators is called the *mixing* operator, denoted $\mathcal{M}$.

Given a population p_n at time step n, $\mathcal{G}(p_n)$ defines the sampling distribution for choosing a new population. Such a population may be chosen by sampling the k-th individual with probability $\mathcal{G}(p_n)^k$. This is a multipnomial distribution, hence the probability that this results in a specified population q, is

$$\mathbb{P}(p_{n+1} = q \,|\, p_n = p) = N! \prod_{j=0}^{s-1} \frac{(\mathcal{G}(p)_j)^{Nq_j}}{(Nq_j)!} \tag{23}$$

where N is the population size.

In addition to representing the sampling probabilities for the next population, $\mathcal{G}(p_n)$ is the *expected* next population. As $N \to \infty$, the variance of the sampling shrinks to zero. In effect, the trajectory of the markov chain tends (in some sense to be defined in Theorems 6.3 and 7.6 using the IPS model of GAs) to the deterministic sequence $p_0, \mathcal{G}(p_0), \mathcal{G}^2(p_0), \mathcal{G}^3(p_0), \ldots$. For this reason, this model (22) is sometimes referred to as the *infinite population model*.

It should be noted, however, that properties of this (infinite population) sequence do give us information about finite populations. In particular, regions in which $\|\mathcal{G}(p) - p\|$ are small enough[5] will, by a continuity argument, be regions in which a finite population will spend some period of time. Such regions (for example, those around fixed points of $\mathcal{G}$) are referred to as *metastable* states [102, 108].

---

[5]Whether this is possible – to closely approach any real numbers by fractions of N – is out of the topic of this paper.

An other recent direction of research has been proposed by [31]. He investigates the link between an abstract property π of a finite population and that of an infinite population. He proves that a sufficient condition for having $\pi(\mathbb{E}[p_{n+2}]) = \pi(\mathcal{G}^2(p_n))$ is that π be a linear operator on the finite populations vectors and that for all finite populations p_n, $\pi(\mathcal{G}(p_n)) = \alpha\pi(p_{n+1})$.

Each of the three GA operators (selection, mutation and cross-over) are thought of in terms of their action on the simplex Δ^d. Proportional selection is a scaled linear operator:

$$\mathcal{F}(p) = \frac{Sp}{\|Sp\|_1}$$

where S is a diagonal matrix with

$$S_{k,k} = f(k)$$

for each individual k. Mutation is a linear operator. It is given by a matrix U, in which $U_{i,j}$ is the probability that individual j mutates to become individual i. Then

$$\mathcal{U}(p) = Up$$

Suppose we have a bilinear operator $B : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$, so that B is linear in both parameters. Then a mapping $Q : \mathbb{R}^d \to \mathbb{R}^d$ defined by

$$Q(x) = B(x,x)$$

is called a *quadratic* operator. Cross-over is given by just such an operator. It can be represented by a set of d matrices, $C^1, \ldots, C^d$, in which the i,j-th component of C^k is the probability that individuals i and j will cross-over to form k.

When the GA has no cross-over, its analysis is relatively straight-forward, as it is a (scaled) linear system. For example, a fixed-point of $\mathcal{G}$ satisfies

$$USp = \lambda p$$

where λ is the mean fitness of p. Thus fixed-points are eigenvectors of the matrix US. The Perron-Frobenius theorem ensures that there is eactly one such eigenvector in Δ^d, corresponding to the largest eigenvalue. Other eigenvectors may be significant, for example by creating metastable regions, as described above, if they are near to the simplex.

When there is cross-over, the situation becomes harder to analyse. The fixed-point equations, for example, form a set of simultaneous quadratic equations. However, the situation can often be dramatically simplified. This happens when crossover and mutation are designed to respect structural symmetries that exist in the search space. This is the case, for example, when the search space comprises fixed-length binary strings, with the usual forms of cross-over and mutation [96]. However, the theory can be generalised to apply to any finite search space with a given group structure [105]. As an example of the kind of results that can be obtained, it is possible to diagonalise the cross-over matrices $C^1, \ldots, C^d$ by applying a Fourier transform, if and only if the underlying group structure is abelian.

It will become clear, as the paper progresses, that the dynamical systems model (22) is a special case of the IPS model (24). It corresponds exactly to the case where E is finite, and fitness is homogeneous with respect to time. For this special case, there are a number of theorems which describe the behaviour of the simple GA. However, there are also a number of open questions. One of the most intriguing issues is the extent to which results which have been proved for finite E can be generalised to infinite E, and also the extent to which results which have been proved for infinite populations apply to finite populations. Recasting the dynamical systems model as an N-IPS model will help to establish both these questions. In particular, convergence properties of the finite population GA when N tends to infinity, can be addressed directly using the N-IPS model, and these results are presented in Theorems 6.1 and 6.2 for a general IPS model, and in Theorem 7.2 and 7.6 for a Genetic-type IPS model.

5   GA behavior and Complexity results on particular fitness functions

None of the results of last section make any particular assumption[6] on the fitness function. But on the other hand, they do not give any indication about the complexity of the algorithm: the expectation of the number of generations required to reach the optimum remains unknown.

5.1   Finite spaces

In binary spaces $E = \{0,1\}^\ell$, the first results about complexity give the expected first hitting time of the solution [20, 64, 21, 81] on the simple, well-studied, OneMax problem [15] ($f(x) = \sum_{i=1}^\ell x_i$).

Recently, these results have been extended by Markov chain techniques, giving the variance of convergence time with different mutations [45]. Another well-studied problem is the long k-path (optimizable with a 1-bit-flip hill climber in exponential time w.r.t. string length) where expected convergence time [79, 39] and variance of convergence time have also been analytically derived [44] as a function of string length and population size. This study examines the simplified situation where the optimal population size is one. Furthermore, the GA with c/ℓ mutation and proportionate or ranking selection takes a super-polynomial time to solve a sub-class of long k-path problems, although these functions present only one local optimum in the Hamming space. This latter paper also answers the challenging issue of existence of *controlled paths* for c/ℓ mutation GAs.

By definition, GAs are stochastic population-based algorithms. It makes little sense to exactly control the composition of the successive populations. Rather, we say that a GA follows a *controlled path* when the best individual of the population visits a number of predefined points in the search space in a predefined order during the course of its run.

This issue is addressed in [57]: A family of *controlled-path* functions is constructed, based on an arbitrary choice of *keys* which define a path in the search space. Experiments

---

[6]The only prerequisite, is that the fitness is supposed to be totally known.

show that the trajectory of elitist GAs with crossover and mutation, follow this path at each run.

5.2   Infinite spaces

In real-valued spaces $E = [-b, b]^\ell$, an exhaustive set of results have been obtained for the corridor $(f_1(x) = c_0 + c_1 x_1)$ and sphere $(f_2(x) = \sum_{i=1}^{\ell} (x_i - x_i^*)^2)$ fitness functions, since the early days of Evolution Strategies (abbreviate ES) [73, 83]. In the case of the sphere model (f_2), it is shown that for large ℓ, $log(f_t/f_0) \simeq t$, where f_t denotes the smallest fitness found at time t.

The most recent and powerful results provide bounds for the Euclidean distance to the solution in different situations. Beyer [25, 26, 27, 28] handles sophisticated variants of ES (i.e. based on Gaussian mutations with self-adaptive mutation parameters), while Rudolph [80, 81] studies the local convergence speed for a larger class of mutation operators.

6   The Interacting Particle System model of GAs

As discussed in the introduction, we can associate with the Markov chain modeling GAs on the space of populations $(\xi_n)_{n\geq)}$ (1) with transitions (17), a measure valued process $(m(\xi_n))_{n\geq 0}$ where $m(\xi)$ is the empirical measures associated with the population ξ_n, defining a uniform distribution over the population members.

Alternatively, we can associate with any abstract measure valued process (24) an N-interacting particle approximating model, defining a finite population GA, as in (25). In this case, the mutation/selection transitions are dictated by the form of the limiting measure valued dynamical system (24). We illustrate this fact in the following section.

We first describe the Interacting Particle model, and limiting behavior with detailed convergence results. Then, we show how these results impact on standard GAs (section 7.1), GAs with crossover (section 7.3) and interacting (fitness dependent) mutations (section 7.2). We finally show how these developments permit us to construct and study new genetic algorithms (section 7.4).

6.1   Description of interacting particle processes

Consider a measure valued process of the form

$$\eta_n = \Phi_n (\eta_{n-1}), \qquad n \geq 1, \tag{24}$$

where $\{\Phi_n \; ; \; n \geq 1\}$ is a given collection of continuous functions from $\mathcal{P}(E)$ into $\mathcal{P}(E)$. The N-IPS approximating model associated with this measure valued process is defined in terms of an E^N-valued Markov chain

$$\xi_n = \left(\xi_n^1, \ldots, \xi_n^N\right), \qquad n \geq 0,$$

with transition probability kernels

$$P\left(\xi_n \in dy \,|\, \xi_{n-1} = x\right) = \prod_{p=1}^{N} \Phi_n(m(x))(dy^p), \tag{25}$$

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

$$\forall x = (x^1, \ldots, x^N) \in E^N, \qquad m(x) \;=\; \frac{1}{N} \sum_{i=1}^{N} \delta_{x^i} \;\in \mathcal{P}(E), \tag{26}$$

The initial system $\xi_0 = (\xi_0^1, \ldots, \xi_0^N)$ consists of N independent particles with common law η_0.

This IPS algorithm can be described as follows:

- At the time $n = 0$ the particle system consists of N independent random particles $\xi_0^1, \ldots, \xi_0^N$. with common law η_0.

- At the time $n \geq 1$ the empirical measure $m(\xi_{n-1})$ associated with the particle system ξ_{n-1} enters in the plant equation (3) so that the resulting measure $\Phi_n(m(\xi_{n-1}))$ depends on the configuration of the system at the previous time $n - 1$.

- Finally, the particle system at the time n consists of N independent (conditionally to ξ_{n-1}) particles $\xi_n^1, \ldots, \xi_n^N$ with common law $\Phi_n(m(\xi_{n-1}))$.

The above description enables us to consider the particle density profiles

$$\eta_n^N = m(\xi_n) = \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi_n^i}, n \geq 0,$$

as a measure valued Markov process with transition probability kernel given by

$$\mathbb{E}\left(F(\eta_n^N)\,\big|\,\eta_{n-1}^N\right) = \int_{E^N} F\left(\frac{1}{N}\sum_{i=1}^{N}\delta_{x^i}\right)\,\Phi_n\left(\eta_{n-1}^N\right)(dx^1)\ldots\Phi_n\left(\eta_{n-1}^N\right)(dx^N),$$

for any $F : \mathcal{P}(E) \to \mathbb{R}$ in the set $\mathcal{C}_b(\mathcal{P}(E))$ of all bounded continuous test functions on $\mathcal{P}(E)$.

6.2   Asymptotic Behavior

This section is concerned with the asymptotic behavior of the N-IPS approximating model as the number of particles N tends to infinity.

It is divided into three parts. Section 6.2.1 covers weak law of large numbers results. In section 6.2.2 we propose an easily verifiable Lipschitz condition on the one step mappings $\{\Phi_n \;;\; n \geq 1\}$ under which one can derive $\mathbb{L}^p$-mean error estimates and useful exponential rates. Section 6.2.3 focuses on large deviations principles. These results complement and strengthen the distributional limit theorems presented in section 6.2.1.

The investigation of these asymptotic results require quite specific mathematical tools. To each of these approaches and techniques correspond an appropriate set of conditions on the one step mappings $\{\Phi_n \;;\; n \geq 1\}$. We provide no examples in this section; this choice

is deliberate. Several detailed and precise examples of GAs satisfying these conditions will be given in the further development of section 7.

We also emphasize that when the state space E is finite most of these conditions are easy to interpret mainly because in this situation the set of all probability measures $\mathcal{P}(E)$ coincide with the unit d-simplex $\Delta^d \subset \mathbb{R}^d$ with

$$d = \text{Card}(E) \quad \text{and} \quad \Delta^d = \left\{ p \in \mathbb{R}^d \ ; \ p_i \geq 0 \quad \sum_{i=1}^d p_i = 1 \right\}.$$

6.2.1 Weak Law of Large Numbers

The approximation of dynamical system (24) by the N-IPS approximating model (25) involving the empirical measures η_n^N, is guaranteed by the following Theorem.

Theorem 6.1 ([11], p. 448.) *If E is compact then we have that*

$$\forall F \in \mathcal{C}_b(\mathcal{P}(E)), \quad \forall n \geq 0, \qquad \lim_{N \to \infty} \mathbb{E}\left(F(\eta_n^N)\right) = F(\eta_n). \tag{27}$$

More generally (27) holds when E is locally compact and the mappings Φ_n, $n \geq 1$, are uniformly continuous.

We now give some comments on Theorem 6.1.
By $\mathcal{B}_b(E)$ we denote the space of all bounded Borel measurable functions $f : E \to \mathbb{R}$ and by $\mathcal{C}_b(E)$ we denote the sub-space of all bounded continuous functions. For any $f \in \mathcal{B}_b(E)$ and $\mu \in \mathcal{P}(E)$ we write

$$\mu(f) = \int_E f(x) \ \mu(dx).$$

One consequence of (27) is that for any $\overline{F} \in \mathcal{C}_b(\mathbb{R}^d)$, $d \geq 1$, and $f_1, \ldots, f_d \in \mathcal{C}_b(E)$ and for any $n \geq 0$

$$\lim_{N \to \infty} \mathbb{E}\left(\overline{F}(\eta_n^N(f_1), \ldots, \eta_n^N(f_d))\right) = \overline{F}(\eta_n(f_1), \ldots, \eta_n(f_d)).$$

Applying this result one can obtain the limit of the $\mathbb{L}^p$-moments of the particle density profiles errors. More precisely for any $n \geq 0$ and $p \geq 1$ and $f \in \mathcal{C}_b(E)$ we have that

$$\lim_{N \to \infty} \mathbb{E}\left(\left|\eta_n^N(f) - \eta_n(f)\right|^p\right) = 0.$$

6.2.2 $\mathbb{L}^p$ and Exponential Bounds

In the next Theorem we propose an easily verifiable condition with regard to the one step mappings Φ_n, $n \geq 1$, which enables us to develop useful estimates. For any finite subset $\mathcal{F} \subset \mathcal{B}_b(E)$ and $\mu, \nu \in \mathcal{P}(E)$ we will use the notations

$$\|\mu - \nu\|_{\mathcal{F}} = \sup\left\{|\mu(f) - \nu(f)| \ ; \ f \in \mathcal{F}\right\}$$

Theorem 6.2 ([11], pp. 451-458.) *Suppose that for any $n \geq 1$ and $\eta \in \mathcal{P}(E)$ and for any finite subset $\mathcal{F} \subset \mathcal{B}_b(E)$ there exists a finite constant $C_{n,\eta}(\mathcal{F}) < \infty$ and a finite subset $\Theta_{n,\eta}(\mathcal{F}) \subset \mathcal{B}_b(E)$ such that*

$$\forall \mu \in \mathcal{P}(E) \qquad \|\Phi_n(\mu) - \Phi_n(\eta)\|_{\mathcal{F}} \leq C_{n,\eta}(\mathcal{F}) \ \|\mu - \nu\|_{\Theta_{n,\eta}(\mathcal{F})} . \tag{28}$$

Then, for any finite subset $\mathcal{F} \subset \mathcal{B}_b(E)$ and $n \geq 0$ and $p \geq 1$ there exists a finite constant $A_{n,p}(\mathcal{F}) < \infty$ such that

$$\mathbb{E}\left(\left\|\eta_n^N - \eta_n\right\|_{\mathcal{F}}^p\right)^{\frac{1}{p}} \leq \frac{A_{n,p}(\mathcal{F})}{\sqrt{N}} . \tag{29}$$

In addition, for any finite subset $\mathcal{F} \subset \mathcal{B}_b(E)$ and for any $n \geq 0$ there exists a finite constant $B_n(\mathcal{F}) < \infty$ such that

$$\forall \epsilon > 0, \qquad \mathbb{P}\left(\left\|\eta_n^N - \eta_n\right\|_{\mathcal{F}} > \epsilon\right) \leq B_n(\mathcal{F}) \ \exp-\frac{N \ \epsilon^2}{B_n(\mathcal{F})} . \tag{30}$$

Theorem 6.2 is a combination of Theorem 2, p. 451, and Proposition 2, p. 458, in [11]. It might seem difficult to check but we shall see in fact that it is in many cases fulfilled.

The proof of (29) is given in [11] for $p = 1, 2$. To prove this inequality for any $p \geq 1$ we simply combine Marcinkiewicz-Zygmund's inequality (cf. [12] p. 498) with the inductive proof presented in [11] (see also [7]).

The extension of Theorem 6.2 to countable and uniformly bounded classes $\mathcal{F}$ can be found in [6] and [7]. These extensions are proved for Feynman-Kac type limiting processes (3) using empirical process techniques. They also depend on appropriate conditions on the covering and entropy numbers associated with the class of function.

The exact magnitude of variability of the $\mathbb{L}^p$-mean errors (29) for general one step mappings Φ_n is still in progress. Central limit theorems for the N-IPS associated with a Feynman-Kac type limiting processes (3) can be founded in [1, 5, 7] including a Donsker type theorem and central limit theorem for the empirical measures on path space. The interested reader is recommended to consult the survey paper [7].

6.2.3   Large Deviation Principles

Precise estimates of the exponential bounds (30) are described in [2] using large deviations techniques. As in this work we will assume that the state space is a Polish space (that is E is a complete separable metric space).

To get some flavor of these precise estimates let us present some quite simple but general large deviations results. To state these results we need to introduce some additional notations.

By $I(\mu|\nu)$ we denote the relative entropy of μ with respect to ν, that is the function

$$I(\mu|\nu) = \int \log \frac{d\mu}{d\nu} \ d\mu$$

if $\mu \ll \nu$ and $+\infty$ otherwise. We will write $Q_n^{(N)}$ the law of the empirical measures

$$\eta_{[0,n]}^N \overset{\text{def.}}{=} \frac{1}{N} \sum_{i=1}^{N} \delta_{(\xi_0^i, \ldots, \xi_n^i)} \in \mathcal{P}(E^{n+1})$$

on the path space E^{n+1} and by μ_k we denote the marginal at time $k \in \{0, \ldots, n\}$ of a measure $\mu \in \mathcal{P}(E^{n+1})$.

Let $\Phi_{[0,n]} : \mathcal{P}(E^{n+1}) \to \mathcal{P}(E^{n+1})$ be the function defined for any $\mu \in \mathcal{P}(E^{n+1})$ by setting

$$\Phi_{[0,n]}(\mu) = \eta_0 \otimes \Phi_1(\mu_0) \otimes \ldots \otimes \Phi_n(\mu_{n-1})$$

then we have the following Theorem.

Theorem 6.3 ([2]) *Suppose the functions $\{\Phi_k \; ; \; 1 \leq k \leq n\}$ satisfy the condition*
($\mathcal{H}$) **For any time $1 \leq k \leq n$ there exists a reference probability measure $\lambda_k \in \mathcal{P}(E)$ such that $\Phi_k(\mu) \sim \lambda_k$, $\forall \mu \in \mathcal{P}(E)$ and**

- **For any $1 \leq k \leq n$ the function**

$$(x, \nu) \mapsto \log \frac{d\Phi_k(\nu)}{d\lambda_k}(x)$$

   is uniformly continuous w.r.t. x and continuous w.r.t. ν.

- **There exist $\mathbf{B}(E)$-measurable functions φ and ψ and constants $\alpha, \beta \in]1, \infty]$ and $\epsilon > 0$ such that $\frac{1}{\alpha} + \frac{1}{\beta} < 1$ and for any $1 \leq k \leq n$**

$$\left| \log \frac{d\Phi_k(\mu)}{d\lambda_k}(x) \right| \leq \varphi(x) + \mu(\psi)$$

   and

$$\int \exp\left(\alpha \varphi^{1+\epsilon}\right) d\lambda_k \vee \int \exp\left(\beta \psi^{1+\epsilon}\right) d\lambda_k < \infty$$

Then, $\{Q_n^{(N)} : N \geq 1\}$ satisfies the LDP with good rate function

$$J_n : \mu \in \mathcal{P}(E^{n+1}) \longrightarrow J_n(\mu) \stackrel{\text{def.}}{=} I(\mu | \Phi_{[0,n]}(\mu)) \in [0, +\infty].$$

and

$$J_n(\mu) = 0 \iff \mu = \eta_0 \otimes \ldots \otimes \eta_n$$

In other words for any closed subset $A \subset \mathcal{P}(E^{n+1})$ and for any open subset $B \subset \mathcal{P}(E^{n+1})$

$$\limsup_{N \to \infty} \frac{1}{N} \log \mathbb{P}\left(\eta_{[0,n]}^N \in A\right) \leq -\inf_{\mu \in A} J_n(\mu)$$

$$\liminf_{N \to \infty} \frac{1}{N} \log \mathbb{P}\left(\eta_{[0,n]}^N \in B\right) \geq -\inf_{\mu \in B} J_n(\mu)$$

Theorem 6.3 makes it possible to estimate in a simple way the probability of the events

$$B_{n,\epsilon} = \left\{\mu \in \mathcal{P}(E^{n+1}) \; ; \; d(\mu, \Phi_{[0,n]}(\mu)) < \epsilon\right\}$$

where d is a metric on $\mathcal{P}(E^{n+1})$ compatible with the weak topology

$$d(\mu, \nu) = \sum_{n \geq 1} 2^{-(n+1)} \; |(\mu(f_n) - \nu(f_n)|$$

where $\{f_n \; ; \; n \geq 1\}$ is a suitably chosen sequence of uniformly continuous functions uniformly bounded by 1.

Recalling that

$$d(\mu, \Phi_{[0,n]}(\mu)) = 0 \Longleftrightarrow \mu = \eta_{[0,n]} \overset{\text{def.}}{=} \eta_0 \otimes \cdots \otimes \eta_n$$

we see that $B_{n,\epsilon}$ is an open neighborhood of the product measure $\eta_{[0,n]}$. Recalling the usual inequality

$$d(\mu, \nu) \leq \|\mu - \nu\|_{tv} \leq (2 \, I(\mu|\nu))^{\frac{1}{2}}$$

where

$$\|\mu - \nu\|_{tv} = \sup \{|\mu(f) - \nu(f)| \; ; \; f \in \mathcal{B}_b(E) \; \|f\| \leq 1\}$$

denotes the total variation norm, the large deviation upper bound implies that

$$\limsup_{N \to \infty} \frac{1}{N} \log \mathbb{P} \left(\eta_{[0,n]}^N \in B_{n,\epsilon}^c \right) \leq -\frac{\epsilon^2}{2}$$

It follows that for any $\epsilon > 0$ there exists an integer $N(\epsilon) \geq 1$ such that for any $N \geq N(\epsilon)$

$$\log \mathbb{P} \left(d(\eta_{[0,n]}^N, \Phi_{[0,n]}(\eta_{[0,n]}^N)) > \epsilon \right) \leq e^{-\frac{N \, \epsilon^2}{4}}$$

from which one concludes that for any $F \in \mathcal{C}_b(\mathcal{P}(E^{n+1}))$

$$\lim_{N \to \infty} \mathbb{E} \left(F(\eta_{[0,n]}^N) \right) = F(\eta_{[0,n]})$$

Remark 6.4:

When the state space E is finite, condition $(\mathcal{H})$ holds as soon as for any time $n \geq 1$ there exists a reference probability measure $\lambda_n \in \mathcal{P}(E)$ and a positive number $\epsilon_n > 0$ such that $\Phi_n(\mu) \sim \lambda_n, \forall \mu \in \mathcal{P}(E)$ and

$$\epsilon_n \leq \frac{d\Phi_n(\mu)}{d\lambda_n}(x) \leq \frac{1}{\epsilon_n} \tag{31}$$

for any $\mu \in \mathcal{P}(E)$ and $n \geq 1$ and $x \in E$.

One can also check that the previous condition is met as soon as for any $\mu \in \mathcal{P}(E)$ and $n \geq 1$ and $x \in E$

$$\Phi_n(\mu)(\{x\}) > 0$$

One way to remove assumption $(\mathcal{H})$ is to work with the law P_n^N of the density profiles

$$\eta_n^N \overset{\text{def.}}{=} \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi_n^i}$$

Theorem 6.5 ([2]) *Assume that the continuous functions $\{\Phi_n \; ; \; n \geq 1\}$ satisfy the following condition*
($\mathcal{H}'$) **For any $n \geq 1$, $\epsilon > 0$ and for any Markov transition M on E, there exist a Markov kernel $\tilde{M}$ and $0 < \delta \leq \epsilon$ such that**

$$\mu\tilde{M}(A^c) < \delta \Longrightarrow \Phi_n(\mu)M(A^c) < \epsilon$$

for any $\mu \in \mathcal{P}(E)$ and for any compact set $A \subset E$.
Then, for any $n \geq 0$, $\{P_n^{(N)} \; : \; N \geq 1\}$ obeys a LDP with convex good rate function H_n given by

$$\begin{cases} H_n(\mu) & = \displaystyle\sup_{V \in \mathcal{C}_b(E)} \left(\mu(V) + \inf_{\nu \in \mathcal{P}(E)} \left(H_{n-1}(\nu) - \log\left(\Phi_n(\nu)(e^V)\right) \right) \right), & n \geq 1 \\ H_0(\mu) & = I(\mu|\eta_0) \end{cases}$$

In addition $H_n(\mu) = 0$ iff $\mu = \eta_n$, for any $n \geq 1$.

When the state space E is finite, condition $(\mathcal{H}')$ holds whenever there exists a sequence of transition probability kernels $\{K_n \; ; \; n \geq 1\}$ and a sequence of positive numbers $\{\epsilon_n \; ; \; n \geq 1\}$ such that $\Phi_n(\mu) \sim \mu K_n$, $\forall \mu \in \mathcal{P}(E)$ and $n \geq 1$ and

$$\epsilon_n \leq \frac{d\Phi_n(\mu)}{d\mu K_n}(x) \leq \frac{1}{\epsilon_n} \tag{32}$$

for any $\mu \in \mathcal{P}(E)$ and $n \geq 1$ and $x \in E$.
To see that this condition relaxes condition (31) we simply notice that under condition (31) the desired inequality (32) holds with

$$K_n(x, dy) = \lambda_n(dy)$$

7   Genetic-type Interacting Particle Systems

The research literature abounds with genetic type stochastic models. We believe that many of them could benefit by a more thorough approach along the IPS lines presented here.

To illustrate the modeling impact of our IPS methodology we have chosen to restrict our attention to 4 generic models, namely

- The simple (selection/mutation) GA.

- GAs with interacting mutation.

- GAs with cross-over transitions.

- GAs with randomly ordered selection/mutation.

The interested reader will find that, although we restrict ourselves to these generic models in illustrating the value of the IPS modeling presented in section 6, virtually all the asymptotic results presented in the previous section apply across more complex GAs.

Let us briefly present the three main ingredients we will use to define these generic GAs.

1. The selection transitions are related to a collection of time-inhomogeneous and continuous fitness functions

$$g_n : E \longrightarrow]0, +\infty[, \qquad n \geq 0.$$

Although it is not essential, to avoid technical difficulties we assume that for any $n \geq 0$ there exists some positive constant $a_n > 0$ such that

$$\forall x, y \in E \qquad \frac{1}{a_n} \leq \frac{g_n(x)}{g_n(y)} \leq a_n \tag{33}$$

2. The mutation transitions are described in terms of a sequence of Markov transition kernels $K_n(x, dz)$, $n \geq 1$, from E into itself. More generally interacting mutations will be defined in terms of a collection of Markov transition kernels

$$\{K_{n,\eta} ; \ n \geq 1, \ \eta \in \mathcal{P}(E)\}$$

indexed by the set of all probability measures $\mathcal{P}(E)$ and by the time parameter $n \geq 1$.
We also assume that K_n and $K_{n,\eta}$ are Feller transitions, in the sense that

$$K_n \left(\mathcal{C}_b(E) \right) \subset \mathcal{C}_b(E) \qquad K_{n,\eta} \left(\mathcal{C}_b(E) \right) \subset \mathcal{C}_b(E)$$

3. The one point cross-over transition will be regarded as a (not necessarily time-homogeneous) random transition from $E^2 = E \times E$ into E It will be defined in terms of transitions probability kernels

$$C_n^{(2)}((x_1, x_2), dy), \qquad n \geq 1.$$

More generally a $p-$point cross-over, $p \geq 1$, will be regarded as a (not necessarily time-homogeneous) random transition from

$$E^p = \underbrace{E \times \ldots \times E}_{p-\text{times}}$$

into E. It will be defined in terms of transitions probability kernels

$$C_n^{(p)}(((x_1, \ldots, x_p), dz), \qquad n \geq 1.$$

We also assume that $C_n^{(p)}$ are Feller, that is

$$C_n^{(p)}(\mathcal{C}_b(E)) \subset \mathcal{C}_b(E^p) \tag{34}$$

To describe precisely our generic models it is convenient to introduce some additional notations.

Let $(E_1, \mathcal{E}_1)$ and $(E_2, \mathcal{E}_2)$ be two measurable spaces and let $K_1(x_1, dx_2)$ be a transition probability measure from E_1 into E_2. It generates two operators, one acting on bounded $\mathcal{E}_2$-measurable functions $f \in \mathcal{B}_b(E_2)$ and the other on measures $\mu \in \mathcal{P}(E_1)$

$$(K_1 f)(x_1) \stackrel{\text{def.}}{=} \int_{E_2} K_1(x_1, dx_2) \; f(x_2) \qquad (\mu K_1)(dx_2) \stackrel{\text{def.}}{=} \int_{E_1} \mu(dx_1) \; K_1(x_1, dx_2)$$

As usual, for any $\mu_1 \in \mathcal{P}(E_1)$ and $\mu_2 \in \mathcal{P}(E_2)$ we denote by $\mu_1 \otimes \mu_2$ the tensor product measure on the product space $(E_1 \times E_2, E_1 \otimes E_2)$ defined by setting

$$(\mu_1 \otimes \mu_2)(dx_1, dx_2) = \mu_1(dx_1) \; \mu_2(dx_2)$$

If $K_2(x_2, dx_3)$ is a transition probability kernel from $(E_2, \mathcal{E}_2)$ into an auxiliary measurable space $(E_3, \mathcal{E}_3)$ then we write $K_1 K_2(x_1, dx_3)$ the composite transition probability from $(E_1, \mathcal{E}_1)$ into $(E_3, \mathcal{E}_3)$ defined by

$$K_1 K_2(x_1, dx_3) \stackrel{\text{def.}}{=} \int_{E_2} K_1(x_1, dx_2) \; K_2(x_2, dx_3)$$

and we write $K_1 \times K_2$ the transition probability measure from E_1 into $E_2 \times E_3$ defined by

$$(K_1 \times K_2)(x_1, d(x_2, x_3)) = K_1(x_1, dx_2) \; K_2(x_2, dx_3)$$

Finally, for any probability measure $\mu \in \mathcal{P}(E_1)$ we write $(\mu \times K_1)$ the probability measure on $(E_1 \times E_2)$

$$(\mu \times K_1)(dx_1, dx_2) \stackrel{\text{def.}}{=} \mu(dx_1) \; K_1(x_1, dx_2)$$

and if $K_1'(x_1, dx_3)$ is transition probability measure from E_1 into E_3 we write $K_1 \otimes K_1'$ the transition probability measure from E_1 into $E_2 \times E_3$ defined by

$$(K_1 \otimes K_1')(x_1, d(x_2, x_3)) = K_1(x_1, dx_2) \; K_1'(x_1, dx_3)$$

7.1   The Simple Genetic Algorithm

7.1.1   Description of the Model

The simple GA can be defined as the N-IPS approximating scheme associated with the nonlinear measure valued process

$$\eta_n = \Phi_n(\eta_{n-1}) \qquad \forall n \geq 1 \tag{35}$$

where for all $n \geq 1$, $\Phi_n : \mathcal{P}(E) \rightarrow \mathcal{P}(E)$ is the mapping defined by

$$\begin{aligned} \Phi_n(\eta) &= \Psi_n(\eta) K_n \\ \forall f \in \mathcal{B}_b(E), \qquad \Psi_n(\eta)(f) &= \frac{\eta(g_n f)}{\eta(g_n)} \end{aligned} \tag{36}$$

The transition probabilities of the N-IPS model $\xi_n = (\xi_n^1, \ldots, \xi_n^N)$, $n \geq 0$, associated with the limiting process (35) here take the form

$$P\left(\xi_n \in dz \,|\, \xi_{n-1} = x\right) \;\; = \;\; \prod_{p=1}^{N} \left(\Psi_n(m(x))K_n\right)(dz^p)$$

where $dz \overset{\text{def}}{=} dz^1 \times \cdots \times dz^N$ is an infinitesimal neighborhood of the point $z = (z^1, \ldots, z^N) \in E^N$ and $x = (x^1, \ldots, x^N) \in E^N$. We notice that this transition can be rewritten in the form

$$P\left(\xi_n \in dz \,|\, \xi_{n-1} = x\right) \;\; = \;\; \int_{E^N} \left(\prod_{p=1}^{N} \Psi_n(m(x))(dy^p)\right) \left(\prod_{p=1}^{N} K_n(y^p, dz^p)\right)$$

$$= \;\; \int_{E^N} S(x, dy) \; M(y, dz)$$

where $dy \overset{\text{def}}{=} dy^1 \times \cdots \times dy^N$ is again an infinitesimal neighborhood of the point $y = (y^1, \ldots, y^N) \in E^N$ and S and M are respectively the selection and mutation transitions on E^N defined by

$$S(x, dy) \overset{\text{def.}}{=} \prod_{p=1}^{N} \Psi_n(m(x))(dy^p) \qquad M(y, dz) \overset{\text{def.}}{=} \prod_{p=1}^{N} K_n(y^p, dz^p)$$

Using the fact that

$$\Psi_n\left(\frac{1}{N} \sum_{i=1}^{N} \delta_{x^i}\right) = \sum_{i=1}^{N} \frac{g_n(x^i)}{\sum_{j=1}^{N} g_n(x^j)} \, \delta_{x^i} \,,$$

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

$$\xi_{n-1} \xrightarrow{\;\;\text{Selection}\;\;} \widehat{\xi}_{n-1} \xrightarrow{\;\;\text{Mutation}\;\;} \xi_n$$

with

$$P\big(\widehat{\xi}_{n-1} \in dz \,|\, \xi_{n-1} = x\big) = S(x, dz) \quad \text{and} \quad P\big(\xi_n \in dz \,|\, \widehat{\xi}_{n-1} = x\big) = M(x, dz)$$

More precisely, at each time $n \geq 1$, each particle examines the system of particles $\xi_{n-1} = \big(\xi_{n-1}^1, \ldots, \xi_{n-1}^N\big)$ and chooses randomly a site ξ_{n-1}^i, $1 \leq i \leq N$, with a probability which depends on the entire configuration ξ_{n-1} and given by

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

This mechanism is called the Selection transition as the particles are selected for reproduction, the most fit individuals being more likely to be selected. In other words, this transition allows particles to give birth to some particles at the expense of other, low

fitness, particles which die.

The second mechanism is called Mutation since at this step each particle evolves randomly according to a given transition probability kernel.

The preceding scheme is clearly a system of interacting particles undergoing adaptation in a time inhomogeneous environment represented by the fitness functions $\{g_n; n \geq 1\}$. Roughly speaking the natural idea is to approximate the two step transitions

$$\eta_{n-1} \xrightarrow{\text{Updating}} \widehat{\eta}_{n-1} \overset{\text{def}}{=} \psi_n(\eta_{n-1}) \xrightarrow{\text{Prediction}} \eta_n = \widehat{\eta}_{n-1} K_n \qquad (37)$$

of the system (35) 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, we have that

$$\eta_{n-1}^N \overset{\text{def.}}{=} \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi_{n-1}^i} \xrightarrow{\text{Selection}} \widehat{\eta}_{n-1}^N \overset{\text{def.}}{=} \frac{1}{N} \sum_{i=1}^{N} \delta_{\widehat{\xi_{n-1}^i}} \xrightarrow{\text{Mutation}} \eta_n^N = \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi_n^i}.$$

Remark 7.1:

In nonlinear filtering settings the choice of the transition probability kernels $\{K_n \; ; \; n \geq 1\}$ and the choice of the fitness functions $\{g_n \; ; \; n \geq 0\}$ is dictated by the filtering problem (see [7] for specific nonlinear filtering examples).

In numerical function optimization settings the choice of the mutation transitions (and also the choice of the fitness function) is more flexible. For example the transitions K_n may involve a cross-over mechanism. For instance K_n may be given by

$$K_n(x, dz) = \int_{E^p} K_n^{(1)}(x, dy_1) \; \ldots \; K_n^{(p)}(x, dy_p) \; C_n^{(p)}((y_1, \ldots, y_p), dz)$$

for some $p \geq 2$ and some transitions $K_n^{(1)}, \ldots, K_n^{(p)}$ on the state space E.

7.1.2 Feynman-Kac Representation

In [7] the simple GA is regarded as a natural N-IPS approximating model for Feynman-Kac type formulae. This connection simply comes from the fact that the solution of the limiting process and infinite population model (35) can be explicitly solved in terms of Feynman-Kac type formulae.

Indeed, let $X = \{X_n \; ; \; n \geq 0\}$ be a time-inhomogeneous Markov chain taking values in E with initial distribution η_0 and whose transition probability kernels are defined by

$$\mathbb{P}\left(X_n \in dy \,|\, X_{n-1} = x\right) = K_n\left(x, dy\right)$$

(here dx stands for an infinitesimal neighborhood of a point $x \in E$ and $y \in E$) and let $\{\eta_n \; ; \; n \geq 0\}$ be the Feynman-Kac type distribution flow given for any $f \in \mathcal{B}_b(E)$ and $n \geq 0$ by

$$\eta_n(f) = \frac{\gamma_n(f)}{\gamma_n(1)} \qquad \text{with} \quad \gamma_n(f) = \mathbb{E}\left(f(X_n) \prod_{m=1}^{n} g_m(X_{m-1}) \right) \qquad (38)$$

(and the usual convention $\prod_\emptyset = 1$). Using the Markov property of X one can check that

$$\eta_n(f) = \frac{\gamma_{n-1}(\ g_n \ (K_nf) \ )}{\gamma_{n-1}(\ g_n \ )}$$

from which one concludes easily that

$$\eta_n(f) \quad = \quad \frac{\gamma_{n-1}(\ g_n \ (K_nf) \ )/\gamma_{n-1}(\ 1 \ )}{\gamma_{n-1}(\ g_n \ )/\gamma_{n-1}(\ 1 \ )} = \frac{\eta_{n-1}(\ g_n \ (K_nf) \ )}{\eta_{n-1}(\ g_n \ )}$$

and therefore one gets the desired result

$$\eta_n = \Psi_n(\eta_{n-1})K_n, \qquad n \geq 1.$$

In nonlinear filtering settings the distributions

$$\eta_n \qquad \text{and} \qquad \widehat{\eta}_{n-1} \overset{\text{def.}}{=} \Psi_n(\eta_{n-1})$$

represent the one-step predictor conditional distributions and the optimal filter conditional filter. These distributions also arise in physics and they can also be regarded as the conditional distribution of a killed particle given non-extinction.

7.1.3   Asymptotic Behavior

Since the mappings Φ_n are continuous Theorem 6.1 can be be used to guarantee the convergence of the N-IPS approximating model to the limiting model as $N \to \infty$.
To check that assumption (28) of Theorem 6.2 is satisfied we first use the decomposition

$$\begin{aligned}
&\Phi_n(\mu)(f) - \Phi_n(\nu)(f) \\
&= \frac{1}{\nu(g_n)} \left[(\mu(g_n \ K_nf) - \nu(g_n \ K_nf)) + \Phi_n(\mu)(f) \, (\nu(g_n) - \mu(g_n)) \right]
\end{aligned}$$

to prove that

$$\begin{aligned}
&|\Phi_n(\mu)(f) - \Phi_n(\nu)(f)| \\
&\leq \ a_n \ [\ |\mu(g_n \ K_nf) - \nu(g_n \ K_nf)| \ + \ \|f\| \ |\mu(g_n) - \nu(g_n)| \ ] \\
&\leq \ a_n \ \max{(1, \|f\|)} \ [\ |\mu(g_n \ K_nf) - \nu(g_n \ K_nf)| \ + \ |\mu(g_n) - \nu(g_n)| \ ]
\end{aligned}$$

¿From the above inequality one concludes that assumption (28) holds. More precisely, for any finite subset $\mathcal{F} \subset \mathcal{B}_b(E)$ one gets

$$\|\Phi_n(\mu) - \Phi_n(\nu)\|_{\mathcal{F}} \leq \ C_n(\mathcal{F}) \ \|\mu - \nu\|_{\Theta_n(\mathcal{F})}.$$

with

$$C_n(\mathcal{F}) = 2a_n \max{(1, \sup_{f \in \mathcal{F}} \|f\|)} \quad \text{and} \quad \Theta_n(\mathcal{F}) = \{g_n \ (K_nf) \, ; \ f \in \mathcal{F} \cup \{1\}\}$$

Thus, Theorem 6.2 can be used to obtain $\mathbb{L}^p$-mean errors as well as exponential rates. More precise estimates including Glivenko-Cantelli Theorem, Donsker and Central Limit Theorems can be founded in [7].

We will not describe in details all these limit theorems but we would like to mention a fluctuation result which gives precise estimates for the previous $\mathbb{L}^p$-mean errors stated in Theorem 6.2.

Let $\{Q_{p,n} \; ; \; 0 \leq p \leq n\}$ be the time inhomogeneous semi-group defined by

$$Q_{n-1,n}(x, dy) = g_n(x) \; K_n(x, dy), \qquad \forall n \geq 1$$

and the inductive formula

$$Q_{p,n}(x, dz) = \int_E Q_{p,p+1}(x, dy) \; Q_{p,n}(y, dz), \qquad \forall 0 \leq p \leq n,$$

with the convention $Q_{n,n} = Id$. For any $f \in \mathcal{B}_b(E)$ and $\mu \in \mathcal{P}(E)$ we denote by $Q_{p,n}f$ the bounded measurable function

$$(Q_{p,n}f)(x) = \int_E Q_{p,n}(x, dz) \; f(z)$$

and by $\mu Q_{p,n} \in \mathcal{P}(E)$ the distribution given by

$$(\mu Q_{p,n})(dy) = \int_E \mu(dx) \; Q_{p,n}(x, dy)$$

If $\{\overline{Q}_{p,n} \; ; \; 0 \leq p \leq n\}$ denotes the time inhomogeneous semi-group defined for any $f \in \mathcal{B}_b(E)$ and $0 \leq p \leq n$ and $x \in E$ by setting

$$(\overline{Q}_{p,n}f)(x) \stackrel{\text{def.}}{=} \frac{(Q_{p,n}f)(x)}{\eta_p(Q_{p,n}1)}$$

then we have the following Central Limit Theorem.

Theorem 7.2 ([7])
For any time $n \geq 0$, the sequence of random fields $\{W_n^N(f) \; ; \; f \in \mathcal{B}_b(E)\}$ where

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

converges in law as $N \to \infty$ to a centered Gaussian field $\{W_n(f) \; ; \; f \in \mathcal{B}_b(E)\}$ satisfying

$$\mathbb{E}\left(W_n(f)W_n(h)\right) = \sum_{p=0}^{n} \eta_p \left[\overline{Q}_{p,n}\left(f - \eta_n f\right) \overline{Q}_{p,n}\left(h - \eta_n h\right) \right]$$

for any $f, h \in \mathcal{B}_b(E)$, in the sense of convergence of finite dimensional distributions.

One consequence of the previous Central Limit Theorem is that for any $f \in \mathcal{B}_b(E)$ we have the following stochastic expansion

$$\eta_n^N(f) = \eta_n(f) + \frac{1}{\sqrt{N}} \; W_n(f) + o_p\left(1/\sqrt{N}\right)$$

where $o_p\left(1/\sqrt{N}\right)$ denotes the order in probability, that is for any $\epsilon > 0$

$$\lim_{N\to\infty} \mathbb{P}\left(\left|\sqrt{N}\, o_p\left(1/\sqrt{N}\right)\right| > \epsilon\right) = 0$$

Next we present how the large deviations principles given in section 6.2.3 translate in this situation.

Corollary 7.3
Assume that E is a Polish state space and the functions $\{\Phi_n\ ;\ n \geq 1\}$ are given by (35) and the transitions probability kernels $\{K_n\ ;\ n \geq 1\}$ are Feller and satisfy the following conditions:
For any time $1 \leq k \leq n$ there exists a reference probability measure $\lambda_k \in \mathcal{P}(E)$ such that $K_k(x, .) \sim \lambda_k$ and

- *For any time $1 \leq k \leq n$ the function*

$$z \mapsto \log \frac{dK_k(x, .)}{d\lambda_k}(z)$$

  is Lipschitz, uniformly on the parameter $x \in E$, and for any $z \in E$ the map

$$x \mapsto \frac{dK_k(x, .)}{d\lambda_k}(z)$$

  is continuous.

- *There exist a $\mathbf{B}(E)$-measurable function φ and constants $\alpha > 1$ and $\epsilon > 0$ such that for any time $1 \leq k \leq n$*

$$\left|\log \frac{dK_k(x, .)}{d\lambda_k}(z)\right| \leq \varphi(z) \qquad \text{and} \qquad \int \exp\left(\alpha\varphi^{1+\epsilon}\right) d\lambda_k < \infty$$

Then, $\{Q_n^{(N)}\ :\ N \geq 1\}$ satisfies a LDP with good rate function J_n.

Remark 7.4:
In view of the remark 6.4 when the state space E is finite the previous large deviation principle holds as soon as for any $n \geq 1$ and $x, y \in E$

$$K_n(x, y) > 0$$

To see that $(\mathcal{H}')$ holds if the functions $\{\Phi_n\ ;\ n \geq 1\}$ are given by (35) we simply notice that, for any $n \geq 1$ and for any compact set $A \subset E$,

$$\forall \eta \in \mathcal{P}(E), \qquad \Phi_n(\eta)(A^c) \leq a_n\, \eta K_n(A^c)$$

Corollary 7.5 *If the functions $\{\Phi_n\ ;\ n \geq 1\}$ are given by (35) and the transitions probability kernels $\{K_n\ ;\ n \geq 1\}$ are Feller then for any $n \geq 0$, $\{P_n^N\ :\ N \geq 1\}$ obeys a LDP with convex good rate function H_n.*

7.1.4   Asymptotic Stability and Uniform Convergence.

Because of its importance in practice we conclude this section with a brief discussion on the asymptotic behavior of the Feynman-Kac type dynamical system (35). More details can be founded in [3] and [7, 9].

Let $\|\mu - \nu\|_{tv}$ be the usual total variation distance between probability measures $\mu, \nu \in \mathcal{P}(E)$. If $\{\Phi_{p,n} ; 0 \le p \le n\}$ denote the composite mappings

$$\Phi_{p,n} = \Phi_n \circ \Phi_{n-1} \circ \dots \circ \Phi_{p+1}$$

with the convention $\Phi_{n,n} = Id$ then we have the following asymptotic stability result

Theorem 7.6 ([3, 4])
Assume that the transition probability kernels $\{K_n ; n \ge 1\}$ satisfy the following condition
$(\mathcal{K})$ **For any time $n \ge 1$, there exist a reference probability measure $\lambda_n \in \mathcal{P}(E)$ and a positive number $\epsilon_n \in (0,1]$ so that $K_n(x, \bullet) \sim \lambda_n$ for any $x \in E$ and**

$$\epsilon_n \le \frac{dK_n(x, \bullet)}{d\lambda_n} \le \frac{1}{\epsilon_n}$$

Then, we have for any $\mu, \nu \in \mathcal{P}(E)$

$$\sum_{n \ge 1} \epsilon_n^2 = \infty \quad \Longrightarrow \quad \lim_{n \to \infty} \|\Phi_{0,n}(\mu) - \Phi_{0,n}(\nu)\|_{tv} = 0$$

$$\lim_{n \to \infty} \frac{1}{n} \sum_{p=1}^n \epsilon_p^2 \overset{\text{def.}}{=} \epsilon^2 > 0 \quad \Longrightarrow \quad \limsup_{n \to \infty} \frac{1}{n} \log \|\Phi_{0,n}(\mu) - \Phi_{0,n}(\nu)\|_{tv} \le -\epsilon^2 < 0$$

$$\inf_{n \ge 1} \varepsilon_n \overset{\text{def.}}{=} \epsilon > 0 \quad \Longrightarrow \quad \sup_{p \ge 0} \|\Phi_{p,p+T}(\mu) - \Phi_{p,p+T}(\nu)\|_{tv} < \exp -(\epsilon^2 \, T)$$

The one step mappings $\{\Phi_n ; n \ge 1\}$ usually fail to be contractive (see for instance [3] for a simple example in the case where $E = \{0,1\}$) and the classical tools of dynamical systems theory cannot be used to study the asymptotic stability of our limiting systems.

Some authors have made the sanguine assumption that the one step mappings are locally contractive (see for instance [13], page 401). This assumption is very restrictive and the resulting stability results do not apply to many situations of interest. In addition (as noticed by the author in [13] page 398) the long time time behavior of such limiting systems strongly depends on its initial condition.

In contrast to the latter our assumption only depends on the mutation transitions $\{K_n ; n \ge 1\}$ and further, the one step mappings $\{\Phi_n ; n \ge 1\}$ are not necessarily contractive for the total variation norm. Our asymptotic stability Theorem 7.6 also implies that the long time behavior of the system does not depend on its initial condition.

<u>Remark</u> 7.7:

- When the state space E is countable the total variation distance coincide with the $\mathbb{L}^1$-norm, that is

$$\|\mu - \nu\|_{tv} = \frac{1}{2} \sum_{x \in E} |\mu(x) - \nu(x)|$$

When the state space $E = \mathbb{R}$ is the real line and the probability measures μ and ν have a probability density (with respect to Lebesgue measure dx) $p(x)$ and $q(x)$ we also have that

$$\|\mu - \nu\|_{tv} = \frac{1}{2} \int_{-\infty}^{+\infty} |p(x) - q(x)| \ dx$$

- In time homogeneous settings (that is $K_n = K$, $g_n = g$) the mapping $\Phi_n = \Phi$ is time homogeneous. If the state space E is compact then $\mathcal{P}(E)$ is again a compact set (for the weak topology) and Brouwer's fixed point Theorem tells us that there exists some fixed point

$$\mu = \Phi(\mu) \in \mathcal{P}(E).$$

Theorem 7.6 gives several conditions under which this fixed point μ is unique and any solution $\{\eta_n \ ; \ n \geq 0\}$ of the time homogeneous dynamical system associated with Φ converges exponentially fast to μ as the time parameter $n \to \infty$. For instance if $(\mathcal{K})$ holds with $\epsilon_n = \epsilon > 0$ then we clearly have that

$$\forall \eta \in \mathcal{P}(E) \qquad \sup_{p \geq 0} \|\Phi_{p,p+T}(\eta) - \mu\|_{tv} \leq e^{-\epsilon^2 \cdot T}$$

- When the state space E is finite condition $(\mathcal{K})$ is satisfied whenever for any $n \geq 1$ and $x, y \in E$

$$K_n(x, y) > 0$$

with the uniform distribution λ_n on E.

- In [3] we propose two different ways to relax the assumptions of theorem 7.6. First we prove that for time homogeneous Markov kernels $(K_n = K)$ the asymptotic stability of (35) also holds when some iterate K^m, $m \geq 1$ satisfies the mixing condition $(\mathcal{K})$. We also relax the assumptions of theorem 25 so that to apply to Gaussian transition probability kernels such as those arising in nonlinear filtering settings.

In [3, 4] and [7, 6, 9] the authors use the asymptotic stability properties of the limiting system (35) to obtain uniform convergence results with respect to the time parameter for the N-IPS approximating model.

Theorem 7.8 ([3, 7]) *Assume that the fitness functions* $\{g_n \ ; \ n \geq 0\}$ *satisfy (33) for some sequence of positive numbers* $\{a_n \ ; \ n \geq 0\}$ *such that*

$$\sup_{n \geq 1} a_n \overset{\text{def.}}{=} a < \infty$$

If the limiting measure valued system (35) is asymptotically stable in the sense that

$$\lim_{T\to\infty} \sup_{\mu,\nu\in\mathcal{P}(E)} \sup_{p\geq 0} \|\Phi_{p,p+T}(\mu) - \Phi_{p,p+T}(\nu)\|_{tv} = 0$$

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

$$\lim_{N\to\infty} \sup_{n\geq 0} \mathbb{E}\left(|\eta_n^N(f) - \eta_n(f)|\right) = 0$$

In addition if the limiting system 35) is exponentially asymptotically stable in the sense that

$$\sup_{\mu,\nu\in\mathcal{P}(E)} \sup_{p\geq 0} \|\Phi_{p,p+T}(\mu) - \Phi_{p,p+T}(\nu)\|_{tv} \leq \exp -\gamma.T$$

then for any $f \in \mathcal{B}_b(E)$ and $p \geq 1$ we have that

$$\sup_{n\geq 0} \mathbb{E}\left(|\eta_n^N(f) - \eta_n(f)|^p\right)^{\frac{1}{p}} \leq \frac{C_p\, a^2}{N^{\frac{\beta}{2}}}\, \|f\|$$

where C_p is a universal constant which only depends on $p \geq 1$ and β is given by

$$\beta = \frac{\gamma}{\gamma' + \gamma}, \qquad \gamma' = 1 + 2\,\log a$$

Finally if the transitions probability kernels $\{K_n \,;\, n \geq 1\}$ satisfy the mixing condition $(\mathcal{K})$ (stated in Theorem 7.6) for some sequence of positive numbers $\{\epsilon_n \,;\, n \geq 1\}$ such that

$$\inf_{n\geq 1} \epsilon_n \stackrel{\text{def.}}{=} \epsilon > 0$$

then, for any $f \in \mathcal{B}_b(E)$ and $p \geq 1$ we have that

$$\sup_{n\geq 0} \mathbb{E}\left(|\eta_n^N(f) - \eta_n(f)|^p\right)^{\frac{1}{p}} \leq \frac{C_p}{\sqrt{N}}\, \frac{a^2}{\epsilon^6}\, \|f\|$$

where C_p is a universal constant which only depends on $p \geq 1$.

This Theorem is in fact a simpler version of the uniform convergence result for Zolotarev's type semi-norms presented in [6, 7]. A full discussion of the empirical process techniques involved in [6, 7] would be too great digression here.

7.2 Genetic Algorithm with Interacting Mutation

7.2.1 Description of the Model

Let $\{\mathcal{K}_{n,\eta} \,;\, n \geq 1, \,\, \eta \in \mathcal{P}(E)\}$ be a collection of transition probability kernels and let $\{\eta_n \,;\, n \geq 0\}$ be the solution of the limiting measure valued process

$$\eta_n = \Phi_n(\eta_{n-1}), \qquad n \geq 1, \tag{39}$$

where the one step mappings $\Phi_n : \mathcal{P}(E) \to \mathcal{P}(E)$ are given for any $\eta \in \mathcal{P}(E)$ by

$$\Phi_n(\eta) = \Psi_n(\eta)\mathcal{K}_{n,\eta} \tag{40}$$

In connection with the Feynman-Kac representation (38) of the limiting model associated with the simple GA described in section 7.1, it is interesting to note that the limiting process (39) can also be explicitly solved using a similar Feynman-Kac formulae.

More precisely, let $\{\eta_n ; n \geq 0\}$ be the solution of (39) and let $\mathcal{X} = \{\mathcal{X}_n ; n \geq 0\}$ be a time in-homogeneous Markov chain taking values in E and whose transition probability kernels are defined by

$$\mathbb{P}\left(\mathcal{X}_n \in dy \,|\, \mathcal{X}_{n-1} = x\right) = \mathcal{K}_{n,\eta_{n-1}}(x, dy)$$

(here again the notation dx stands for an infinitesimal neighborhood of a point $x \in E$ and $y \in E$) and initial law η_0.

Arguing as before one can check (by induction) that the flow of distributions $\{\eta_n ; n \geq 0\}$ can be rewritten as follows

$$\eta_n(f) = \frac{\mathbb{E}\left(f(\mathcal{X}_n)\prod_{m=1}^{n} g_m(\mathcal{X}_{m-1})\right)}{\mathbb{E}\left(\prod_{m=1}^{n} g_m(\mathcal{X}_{m-1})\right)}$$

7.2.2   The Genetic Algorithm

As usual the N-IPS approximating scheme of (40) is the E^N-valued Markov chain $\xi_n = (\xi_n^1, \dots, \xi_n^N)$, $n \geq 0$, with transition probability kernels

$$P\left(\xi_n \in dz \,|\, \xi_{n-1} = x\right) = \prod_{p=1}^{N} \left(\Psi_n(m(x))\mathcal{K}_{n,m(x)}\right)(dz^p)$$

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

As in the preceding section the latter transition can be rewritten as follows

$$\begin{aligned}
P\left(\xi_n \in dz \,|\, \xi_{n-1} = x\right) &= \int_{E^N} (\prod_{p=1}^{N} \Psi_n(m(x))(dy^p)) \, (\prod_{p=1}^{N} \mathcal{K}_{n,m(x)}(y^p, dz^p)) \\
&= \int_{E^N} S(x, dy) \, M'((x,y), dz) \tag{41}
\end{aligned}$$

where $dy \stackrel{\text{def}}{=} dy^1 \times \cdots \times dy^N$ is again an infinitesimal neighborhood of the point $y = (y^1, \dots, y^N) \in E^N$ and S and M' are respectively the selection and mutation transitions from E^N (respectively $E^N \times E^N$) into E^N defined by

$$S(x, dy) \stackrel{\text{def.}}{=} \prod_{p=1}^{N} \Psi_n(m(x))(dy^p) \qquad M'((x,y), dz) \stackrel{\text{def.}}{=} \prod_{p=1}^{N} \mathcal{K}_{n,m(x)}(y^p, dz^p)$$

According to 41 the transition from ξ_{n-1} to ξ_n is decomposed into two mechanisms

$$\xi_{n-1} \in E^N \xrightarrow{\text{Selection}} (\xi_{n-1}, \widehat{\xi}_{n-1}) \in E^N \times E^N \xrightarrow{\text{InteractingMutation}} \xi_n \in E^N$$

with transitions:

$$
\begin{aligned}
P\big(\widehat{\xi}_{n-1} \in dy \,|\, \xi_{n-1} = x\big) &= S(x, dy) \\
P\big(\xi_n \in dz \,|\, \xi_{n-1} = x, \widehat{\xi}_{n-1} = y\big) &= M'((x,y), dz).
\end{aligned}
$$

In contrast to the simple GA the mutation transition from the individual $\widehat{\xi}^i_{n-1}$ to ξ^i_n depends on the whole configuration $\xi_{n-1} = (\xi^1_{n-1}, \ldots, \xi^N_{n-1})$. More precisely, given the configurations $(\xi_{n-1}, \widehat{\xi}_{n-1})$ the next configuration $\xi_n = (\xi^1_n, \ldots, \xi^N_n)$ consists of N-conditionally independent random variables with distributions

$$\mathcal{K}_{n, m(\xi_{n-1})}(\widehat{\xi}^i_{n-1}, .), \qquad 1 \leq i \leq N. \tag{42}$$

As a guide to their usage let us work out some simple examples of interacting mutation kernels.

Example 1 *The simplest example of interacting mutation to describe is the transition probability kernel with cross-over component and given by*

$$\mathcal{K}_{n, \eta}(x, dz) = \int_E \eta(dy) \, C_n^{(2)}((x,y), dz)$$

It is clear that the sampling of a transition $x \to z$ with distribution $\mathcal{K}_{n,\eta}(x, dz)$ consists in crossing over the initial individual x with a randomly chosen individual y with distribution η.

Note that the distributions (42) take the form

$$\mathcal{K}_{n, m(\xi_{n-1})}(\widehat{\xi}^i_{n-1}, \cdot) = \frac{1}{N} \sum_{j=1}^N C_n^{(2)}((\xi^j_{n-1}, \widehat{\xi}^i_{n-1}), \cdot)$$

for each $1 \leq i \leq N$ and each interacting mutation

$$\widehat{\xi}^i_{n-1} \longrightarrow \xi^i_n, \qquad 1 \leq i \leq N,$$

is decomposed into two stages

- *One chooses randomly and uniformly a site ξ^j_{n-1} in the configuration $\xi_{n-1} = (\xi^1_{n-1}, \ldots, \xi^N_{n-1})$.*

- *When the individual ξ^j_{n-1} is chosen, we cross-over the individuals $(\xi^j_{n-1}, \widehat{\xi}^i_{n-1})$ according to the transition $C_n^{(2)}((\xi^j_{n-1}, \widehat{\xi}^i_{n-1}), .)$, that is*

$$(\xi^j_{n-1}, \widehat{\xi}^i_{n-1}) \longrightarrow \xi^i_n \sim C_n^{(2)}((\widehat{\xi}^i_{n-1}, \xi^j_{n-1}), .)$$

Example 2 *The second example*

$$\mathcal{K}_{n,\eta}(x, du) = \int_{E\times E} \eta(dy)\ C_n^{(2)}((x,y), dz)\ K_n(z, du)$$

is essentially the same as the first one except that the cross-over mechanism is followed by a simple mutation transition.

Example 3 *In previous examples the cross-over operates on randomly chosen individual with common law η_n. If we replace η by $\Psi_n(\eta)$ one gets the following interacting mutation*

$$\mathcal{K}_{n,\eta}(x, dz) = \int_{E\times E} \Psi_n(\eta)(dy)\ C_n^{(2)}((x,y), dz)$$

Note that in this situation the cross-over mechanism acts on randomly chosen but selected individuals with common law $\Psi_n(\eta)$.
One can also combine the selection Ψ_n with the mutation transition K_n and define

$$\mathcal{K}_{n,\eta}(x, dz) = \int_{E\times E} (\Psi_n(\eta)K_n)\,(dy)\ C_n^{(2)}((x,y), dz)$$

7.2.3   Asymptotic Behavior

In order to apply Theorem 6.1 we need to ensure that the one step mappings $\{\Phi_n\ ;\ n \geq 1\}$ are continuous (for the weak topology on $\mathcal{P}(E)$). Next we propose sufficient continuity conditions in terms of the interacting transitions $\{\mathcal{K}_{n,\eta}\ ;\ n \geq 1,\ \eta \in \mathcal{P}(E)\}$, namely

- For any $\eta \in \mathcal{P}(E)$ and for any $n \geq 1$, $\mathcal{K}_{n,\eta}$ is a Feller transition kernel.

- For any $f \in \mathcal{C}_b(E)$ and $n \geq 1$ the mapping

$$\eta \in \mathcal{P}(E)\ \longrightarrow\ \mathcal{K}_{n,\eta}f(x)$$

  is uniformly continuous with respect to $x \in E$.

Under these conditions and when the state space E is compact Theorem 6.1 implies that the density profiles

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

converge weakly to the desired distribution η_n as $N \to \infty$.
In order to be able to verify the Lipschitz condition (28) of Theorem 6.2 we need to strengthen the previous continuity properties.
Let us assume that for any $n \geq 1$ and $\eta \in \mathcal{P}(E)$ and for any finite subset $\mathcal{F} \subset \mathcal{B}_b(E)$ there exists a finite constant $C_{n,\eta}^{(1)}(\mathcal{F}) < \infty$ and a finite subset $\Theta_{n,\eta}^{(1)}(\mathcal{F}) \subset \mathcal{B}_b(E)$ such that for any $\mu \in \mathcal{P}(E)$ and $x \in E$

$$\|\mathcal{K}_{n,\mu}(x, .) - \mathcal{K}_{n,\eta}(x, .)\|_{\mathcal{F}} \leq\ C_{n,\eta}^{(1)}(\mathcal{F})\ \|\mu - \eta\|_{\Theta_{n,\eta}^{(1)}(\mathcal{F})} \tag{43}$$

Arguing as in section 7.1.3 one can check that the condition (28) of Theorem 6.2 holds with

$$C_{n,\eta}(\mathcal{F}) = C_{n,\eta}^{(1)}(\mathcal{F}) + C_{n,\eta}^{(2)}(\mathcal{F}) \quad \text{and} \quad \Theta_{n,\eta}(\mathcal{F}) = \Theta_{n,\eta}^{(1)}(\mathcal{F}) \cup \Theta_{n,\eta}^{(2)}(\mathcal{F})$$

where

$$C_{n,\eta}^{(2)}(\mathcal{F}) = 2a_n \max\left(1, \sup_{f\in\mathcal{F}} \|f\|\right) \quad \text{and} \quad \Theta_{n,\eta}^{(2)}(\mathcal{F}) = \{g_n \ (\mathcal{K}_{n,\eta}f) \ ; \ f \in \mathcal{F} \cup \{1\}\}$$

from which one concludes that under (43) the $\mathbb{L}^p$-mean errors and the exponential rates stated in Theorem 6.2 hold true. When the state space E is finite the inequality (43) is met for the three examples of interacting mutations given in the end of the previous section.

Next we propose a mean field interacting mutation which also fits into our framework.

Example 4 *Suppose that $E = \mathbb{R}$ and $\mathcal{K}_{n,\eta}$ are given by*

$$\mathcal{K}_{n,\eta}(x, dz) = \frac{\alpha_n}{2} \ e^{-\alpha_n|z-b(x,\eta)|} \ dz$$

where $b_n : \mathbb{R} \times \mathcal{P}(\mathbb{R}) \to \mathbb{R}$ is a bounded measurable function and $\alpha_n > 0$. Using the inequality

$$\left| e^{-x} - e^{-y} \right| \leq |x-y| \ \left| e^{-x} + e^{-y} \right|$$

which is valid for any $x, y \in \mathbb{R}$ one concludes that for any $\mu, \eta \in \mathcal{P}(\mathbb{R})$

$$\left| e^{-\alpha_n|z-b(x,\mu)|} - e^{-\alpha_n|z-b(x,\eta)|} \right| \leq 2 \ e^{\alpha_n\|b_n\|} \ |b(x,\mu) - b(x,\eta)| \ e^{-\alpha_n|z|}$$

with

$$\|b_n\| = \sup\left\{|b(x,\mu)| \ ; \ x \in \mathbb{R}, \mu \in \mathcal{P}(\mathbb{R})\right\}$$

This implies that for any finite subset $\mathcal{F} \subset \mathcal{B}_b(\mathbb{R})$ and for any $x \in \mathbb{R}$

$$\|\mathcal{K}_{n,\mu}(x, \cdot) - \mathcal{K}_{n,\eta}(x, \cdot)\|_{\mathcal{F}} \leq 2 \ \alpha_n \ e^{\alpha_n\|b_n\|} \ \sup_{f\in\mathcal{F}} \|f\| \ |b(x,\mu) - b(x,\eta)|$$

Thus, is we assume that

$$\sup_x |b(x,\mu) - b(x,\eta)| \leq C_n \ \|\mu - \eta\|_{\mathcal{F}_n} \tag{44}$$

for some finite constant and some finite subset $\mathcal{F}_n \subset \mathcal{B}_b(\mathbb{R})$ then the desired inequality (43) holds with

$$C_{n,\eta}^{(1)}(\mathcal{F}) = 2 \ \alpha_n \ C_n \ e^{\alpha_n\|b_n\|} \ \sup_{f\in\mathcal{F}} \|f\|$$

and

$$\Theta_{n,\eta}^{(1)}(\mathcal{F}) = \mathcal{F}_n$$

If the function b_n have the form

$$b(x,\mu) = \sum_{l=1}^{L} a_{l,n}(x) \ \mu(f_{n,l})$$

for some finite L, $\|a_{l,n}\| < \infty$ and some bounded Borel measurable functions $\{f_{n,l}\ ;\ 1 \leq l \leq L\}$ then (44) is clearly satisfied with

$$C_n = \sum_{l=1}^{L} \|a_{l,n}\| \qquad \text{and} \qquad \mathcal{F}_n = \{f_{n,l}\ ;\ 1 \leq l \leq L\}$$

Next we present a sufficient condition for the condition $(\mathcal{H})$ of Theorem 6.3 in terms of the interacting transitions $\{\mathcal{K}_{n,\eta}\ ;\ n \geq 1,\ \eta \in \mathcal{P}(E)\}$. Assume that for any $n \geq 0$ there exists a reference probability measure $\lambda_n \in \mathcal{P}(E)$ such that for any $n \geq 1$, $\eta \in \mathcal{P}(E)$ and $x \in E$

$$\mathcal{K}_{n,\eta}(x, .) \sim \lambda_n$$

and

- $z \in E \rightarrow \frac{d\mathcal{K}_{n,\eta}(x, .)}{d\lambda_n}(z) \in \mathbb{R}$ is uniformly Lipschitz with respect to (x, η).

- $x \in E \rightarrow \frac{d\mathcal{K}_{n,\eta}(x, .)}{d\lambda_n}(z) \in \mathbb{R}$ is continuous.

- $\eta \in \mathcal{P}(E) \rightarrow \frac{d\mathcal{K}_{n,\eta}(x, .)}{d\lambda_n}(z) \in \mathbb{R}$ is continuous uniformly with respect to x.

It is not difficult to check that under these conditions $(\mathcal{H})$ is met. We also notice that when the state space E is finite the first and the second continuity conditions clearly hold and the third and last one reduces to the simple continuity of the mapping

$$\eta \in \Delta^d \rightarrow \frac{d\mathcal{K}_{n,\eta}(x, .)}{d\lambda_n}(z) \in \mathbb{R}$$

where Δ^d is the unit d-simplex $\Delta^d \subset \mathbb{R}^d$ with $d = \mathrm{Card}(E)$.

7.3 Genetic Algorithms with Cross-over

The cross-over transition is often used in GA literature to improve the search procedures. In the present section we single out two rather general GA models with cross-over transitions.
One of the achievements of this section is the identification of some very striking connections between GA with cross-over transitions and the simple GA framework discussed in section 7.1.
We begin with a simple GA model with cross-over which fit into the simple GA framework of section 7.1. We emphasize that all limit theorems for the simple GA can be used to study the long time behavior as well as the fluctuations and the large deviations.
We end this section with a GA variant which does not fit in the previous simple GA framework but which can be studied using the general limit theorems presented in the preliminary section 6.

7.3.1   Description of the Model

We shall now describe a general approach which allows us to model a GA with crossing over transitions as a simple GA. The key idea consists in replacing the state space E by a product state space and to consider mutation transitions involving cross-over mechanisms.

To begin with, let $E^{(2)} = E \times E$ and for any $n \geq 1$ let $\mathcal{K}_n$ be the transition probability kernel from $E^{(2)}$ into itself given by

$$\mathcal{K}_n((x_1, x_2), d(y_1, y_2)) = C_n^{(2)}((x_1, x_2), dy_1) \ K_{n+1}(y_1, dy_2)$$

For each $n \geq 1$, let $\mathcal{G}_n : E^{(2)} \to]0, \infty[$ by a collection of fitness functions on $E^{(2)}$ which only depend on the first component and such that

$$\mathcal{G}_n(x_1, x_2) = g_n(x_1), \qquad g_n : E \to]0, \infty[$$

As in section 7.1 we consider the $\mathcal{P}(E^{(2)})$-valued limiting process

$$\mu_n = \Upsilon_n(\mu_{n-1})\mathcal{K}_n \tag{45}$$

with initial distribution

$$\mu_0 = \eta_0 \times K_1 \in \mathcal{P}(E^{(2)}), \qquad \eta_0 \in \mathcal{P}(E)$$

and where for all $n \geq 1$, $\Upsilon_n : \mathcal{P}(E^{(2)}) \to \mathcal{P}(E^{(2)})$ is the mapping defined for any $f \in \mathcal{B}_b(E^{(2)})$ by setting

$$\Upsilon_n(\mu)(f) = \frac{\mu(\mathcal{G}_n f)}{\mu(\mathcal{G}_n)}$$

It is transparent from our construction that the limiting process (45) has the same form as the limiting process associated with a simple GA except that the state space is now $\mathcal{P}(E^{(2)})$ instead of $\mathcal{P}(E)$.

7.3.2   Feynman-Kac Representation

Arguing as in section 7.1.2, if

$$\mathcal{X} = \{\mathcal{X}_n \ ; \ n \geq 0\}$$

is the time-inhomogeneous Markov chain taking values in $E^{(2)}$ with initial distribution

$$\mu_0(d(x_1, x_2)) = \eta_0(dx_1) \ K_1(x_1, dx_2)$$

and whose transition probability kernels are defined by

$$\begin{aligned} \mathbb{P}\left(\mathcal{X}_n \in d(y_1, y_2) \,|\, \mathcal{X}_{n-1} = (x_1, x_2)\right) &= \mathcal{K}_n((x_1, x_2), d(y_1, y_2)) \\ &= C_n^{(2)}((x_1, x_2), dy_1) \ K_{n+1}(y_1, dy_2) \end{aligned}$$

(here $d(y_1, y_2)$ stands for an infinitesimal neighborhood of a point $(y_1, y_2) \in E \times E$ and $(x_1, x_2) \in E \times E$) then one concludes that the Feynman-Kac distributions flow $\{\mu_n \ ; \ n \geq 0\}$ given for any $f \in \mathcal{B}_b(E \times E)$ and $n \geq 0$ by

$$\mu_n(f) = \frac{\mathbb{E}\left(f(\mathcal{X}_n) \prod_{m=1}^{n} \mathcal{G}_m(\mathcal{X}_{m-1})\right)}{\mathbb{E}\left(\prod_{m=1}^{n} \mathcal{G}_m(\mathcal{X}_{m-1})\right)} \tag{46}$$

(and the usual convention $\prod_\emptyset = 1$) are solutions of (45).

This Feynman-Kac representation of the infinite population model (45) and its N-IPS approximating model clearly enter the framework of the simple GA presented in section 7.1.

7.3.3   The Genetic Algorithm

The particular interest of our model relies on the presence of a cross-over transition in the definition of the mutation transition probability kernel $\mathcal{K}_n$. In our situation the N-IPS approximating model associated with (45) is an $(E^{(2)})^N$-valued Markov chain

$$\xi_n = (\xi_n^1, \ldots, \xi_n^N) \in (E^{(2)})^N, \qquad n \geq 1$$

with transition probability kernels

$$P(\xi_n \in dz \,|\, \xi_{n-1} = x) = \prod_{p=1}^N \left(\Upsilon_n(m(x))\mathcal{K}_n\right)(dz^p) \tag{47}$$

where $dz \overset{\text{def}}{=} dz^1 \times \cdots \times dz^N$ is an infinitesimal neighborhood of the point $z = (z^1, \ldots, z^N) \in (E^{(2)})^N$ and $x = (x^1, \ldots, x^N) \in (E^{(2)})^N$. As usual, the initial particle system

$$\xi_0 = (\xi_0^1, \ldots, \xi_0^N) \in (E^{(2)})^N$$

consists of N-independent and $E^{(2)}$-valued random variables with common law

$$\mu_0(d(x_1, x_2)) = \eta_0(dx_1) \, K_1(x_1, dx_2).$$

Since each individual ξ_n^i, $1 \leq i \leq N$, takes values in the product space

$$E^{(2)} = E \times E$$

it is convenient to use the notations

$$\xi_n^i = (\xi_n^{(1),i}, \xi_n^{(2),i}), \qquad \forall 1 \leq i \leq N.$$

Arguing as in section 7.1.1 the transition probability kernels (47) can be rewritten as follows

$$P\left((\xi_n^{(1)}, \xi_n^{(2)}) \in d(z_1, z_2) \,|\, (\xi_{n-1}^{(1)}, \xi_{n-1}^{(2)}) = (x_1, x_2)\right)$$
$$= \int_{(E^{(2)})^N} \left(\prod_{p=1}^N \Upsilon_n(m((x_1, x_2)))(d(y_1, y_2)^p)\right) \left(\prod_{p=1}^N \mathcal{K}_n((y_1, y_2)^p, d(z_1, z_2)^p)\right)$$
$$= \int_{(E^{(2)})^N} S((x_1, x_2), d(y_1, y_2)) \qquad M((y_1, y_2), d(z_1, z_2))$$

where

$$d(z_1, z_2) \overset{\text{def}}{=} d(z_1, z_2)^1 \times \cdots \times d(z_1, z_2)^N$$

denotes an infinitesimal neighborhood of the point

$$(z_1, z_2) = ((z_1, z_2)^1, \ldots, (z_1, z_2)^N) \in (E^{(2)})^N,$$

and $(x_1, x_2) = ((x_1, x_2)^1, \ldots, (x_1, x_2)^N) \in (E^{(2)})^N$ and S and M are respectively the selection and mutation transitions on $(E^{(2)})^N$ defined by

$$S((x_1, x_2), d(y_1, y_2)) \stackrel{\text{def.}}{=} \prod_{p=1}^N \Upsilon_n(m((x_1, x_2)))(d(y_1, y_2)^p)$$

and

$$M((y_1, y_2), d(z_1, z_2)) \stackrel{\text{def.}}{=} \prod_{p=1}^N \mathcal{K}_n((y_1, y_2)^p, d(z_1, z_2)^p))$$

It follows from the preceding discussion that the transitions (47) are again decomposed into two stages

$$(\xi_{n-1}^{(1)}, \xi_{n-1}^{(2)}) \xrightarrow{\text{Selection}} (\widehat{\xi}_{n-1}^{(1)}, \widehat{\xi}_{n-1}^{(2)}) \xrightarrow{\text{Mutation}} (\xi_n^{(1)}, \xi_n^{(2)})$$

which can be modeled as follows

$$P\left((\widehat{\xi}_{n-1}^{(1)}, \widehat{\xi}_{n-1}^{(2)}) \in d(y_1, y_2) \,|\, (\xi_{n-1}^{(1)}, \xi_{n-1}^{(2)}) = (x_1, x_2) \right) = S((x_1, x_2), d(y_1, y_2))$$

$$P\left((\xi_n^{(1)}, \xi_n^{(2)}) \in d(z_1, z_2) \,|\, (\widehat{\xi}_{n-1}^{(1)}, \widehat{\xi}_{n-1}^{(2)}) = (y_1, y_2) \right) = M((y_1, y_2), d(z_1, z_2))$$

Let us describe more precisely this two stages transition.

- **Selection**
  Given $(\xi_{n-1}^{(1)}, \xi_{n-1}^{(2)}) \in (E \times E)^N$ the selected new configuration

$$\widehat{\xi}_{n-1} = (\widehat{\xi}_{n-1}^{(1)}, \widehat{\xi}_{n-1}^{(2)}) = \left((\widehat{\xi}_{n-1}^{(1),1}, \widehat{\xi}_{n-1}^{(2),1}), \ldots, (\widehat{\xi}_{n-1}^{(1),N}, \widehat{\xi}_{n-1}^{(2),N}) \right)$$

consists of N-(conditionally) independent and $(E \times E)$-random variables with common distribution law

$$\Upsilon_n(m((\xi_{n-1}^{(1)}, \xi_{n-1}^{(2)})), .) = \sum_{i=1}^N \frac{g_n(\xi_{n-1}^{(1),i})}{\sum_{j=1}^N g_n(\xi_{n-1}^{(1),j})} \, \delta_{(\xi_{n-1}^{(1),i}, \xi_{n-1}^{(2),i})}(.)$$

Note that the fitness function only depends on the first component so that the selection of the individuals

$$(\xi_{n-1}^{(1),i}, \xi_{n-1}^{(2),i}), \qquad 1 \le i \le N$$

only depends on the corresponding first components

$$\xi_{n-1}^{(1),i}, \qquad 1 \le i \le N$$

- **Mutation**
  Given the selected configuration from the previous stage

  $$\widehat{\xi}_{n-1} = (\widehat{\xi}_{n-1}^{(1)}, \widehat{\xi}_{n-1}^{(2)})$$

  Each pair individuals $(\widehat{\xi}_{n-1}^{(1),i}, \widehat{\xi}_{n-1}^{(2),i})$, $1 \leq i \leq N$, cross-over into a new individual $\xi_n^{(1),i}$ according to the transition $C_n^{(2)}((\widehat{\xi}_{n-1}^{(1),i}, \widehat{\xi}_{n-1}^{(2),i}), .)$, that is

  $$(\widehat{\xi}_{n-1}^{(1),i}, \widehat{\xi}_{n-1}^{(2),i}) \longrightarrow \xi_n^{(1),i} \ \sim \ C_n^{(2)}((\widehat{\xi}_{n-1}^{(1),i}, \widehat{\xi}_{n-1}^{(2),i}), .), \qquad 1 \leq i \leq N$$

  Then, independently each individual $\xi_n^{(1),i}$, $1 \leq i \leq N$, evolves randomly according to the transition probability kernels K_{n+1}, that is

  $$\xi_n^{(1),i} \longrightarrow \xi_n^{(2),i} \ \sim \ K_{n+1}\left(\xi_n^{(1),i}, .\right)$$

It is also interesting to note that the limiting distribution flow $\{\mu_n \ ; \ n \geq 0\}$ can also be written in the form

$$\mu_n = \eta_n \times K_{n+1}, \qquad n \geq 0$$

where $\{\eta_n \ ; \ n \geq 0\}$ is the solution of the $\mathcal{P}(E)$-valued dynamical system

$$\eta_n = (\Psi_n(\eta_{n-1}) \times K_n) \, C_n^{(2)}, \qquad n \geq 1 \tag{48}$$

where $\Psi_n : \mathcal{P}(E) \to \mathcal{P}(E)$ is the mapping defined by

$$\forall f \in \mathcal{B}_b(E), \qquad \Psi_n(\eta)(f) = \frac{\eta(g_n f)}{\eta(g_n)}$$

To see this claim it suffices to note that

$$\Upsilon_n(\eta_{n-1} \times K_n) = \Psi_n(\eta_{n-1}) \times K_n$$

and

$$C_n^{(2)} \times K_{n+1} = \mathcal{K}_n$$

imply that

$$\begin{aligned}
\mu_n = \eta_n \times K_{n+1} &= (\Psi_n(\eta_{n-1}) \times K_n) \, C_n^{(2)} \times K_{n+1} \\
&= \Upsilon_n(\eta_{n-1} \times K_n) C_n^{(2)} \times K_{n+1} \\
&= \Upsilon_n(\mu_{n-1}) \mathcal{K}_n
\end{aligned}$$

Thus, one gets the desired recursion

$$\mu_n = \Upsilon_n(\mu_{n-1}) \mathcal{K}_n$$

In this connection it can also be checked directly that the E^N-valued Markov chain $\{\xi_n^{(1)} \; ; \; n \geq 0\}$ defined above coincides with the N-IPS approximating model associated with (48). Indeed, using the above notations we clearly have that

$$\mathbb{P}\left(\xi_n^{(1)} \in dy_1 \,\Big|\, \xi_{n-1}^{(1)} = x_1\right) = \prod_{p=1}^{N} \left(\Psi_n(m(x_1)) \times K_n\right) C_n^{(2)}$$

Remark 7.9:
The previous GA model with pair individual cross-over can also be extended into a GA with p-cross-over transitions. Namely, if we replace in the previous construction the state space $E^{(2)}$ by the product space $E^{(p)}$ and $\mathcal{K}_n$ by a transition probability kernel on $E^{(p)}$ such that

$$\mathcal{K}_n\left((x_1, \ldots, x_p), d(y_1, \ldots, y_p)\right)$$

$$= C_n^{(p)}(((x_1, \ldots, x_p), dy_1) \; K_{n+1}(y_1, dy_2) \; \ldots \; K_{n+(p-1)}(y_{p-1}, dy_p)$$

and the fitness functions $\mathcal{G}_n$ by a fitness function

$$\mathcal{G}_n : (x_1, \ldots, x_p) \in E^{(p)} \longrightarrow \mathcal{G}_n((x_1, \ldots, x_p)) = g_n(x_1)$$

then we end up with a $\mathcal{P}(E^{(p)})$-valued limiting process

$$\mu_n = \Upsilon_n(\mu_{n-1})\mathcal{K}_n$$

which also fits into the simple GA framework of section 7.1.

7.3.4   A Crossover Variant

We conclude this section with a class of GA involving cross-over transitions and which do not fit into the simple GA framework but which can be studied using the N-IPS approach described in section 6.

As usual, one of the simplest way for defining this GA model is as the N-IPS approximating model associated with a suitably chosen limiting process. Here we will consider the $\mathcal{P}(E)$-valued dynamical system defined by

$$\eta_n = (\underbrace{\Phi_n(\eta_{n-1}) \otimes \; \ldots \; \otimes \Phi_n(\eta_{n-1})}_{\text{p-times}})C_n^{(p)} \tag{49}$$

where for all $n \geq 1$, $\Phi_n : \mathcal{P}(E) \to \mathcal{P}(E)$ is the mapping defined by

$$\Phi_n(\eta) = \Psi_n(\eta)K_n$$
$$\forall f \in \mathcal{B}_b(E), \qquad \Psi_n(\eta)(f) = \frac{\eta(g_n f)}{\eta(g_n)} \tag{50}$$

The transition probabilities of the N-IPS approximating model $\xi_n = (\xi_n^1, \ldots, \xi_n^N)$, $n \geq 0$, associated with the limiting process (49) take the form

$$P\left(\xi_n \in dz \,|\, \xi_{n-1} = x\right) = \prod_{q=1}^{N} ((\underbrace{\Phi_n(m(x)) \otimes \; \ldots \; \otimes \Phi_n(m(x))}_{\text{p-times}})C_n^{(p)})(dz^q) \tag{51}$$

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

If $\mathcal{B}_n^{(p)}$ and $\mathcal{C}_n^{(p)}$ denote respectively the transition probability kernel from E^N into $(E^p)^N$ and from $(E^p)^N$ into E^N, defined for any $x \in E^N$ and $\tilde{y} \in (E^p)^N$ by

$$
\mathcal{B}_n^{(p)}(x, d\tilde{y}) = \prod_{q=1}^{N} \underbrace{(\Phi_n(m(x)) \otimes \ldots \otimes \Phi_n(m(x)))}_{\text{p-times}}(d\tilde{y}^q)
$$

$$
\mathcal{C}_n^{(p)}(\tilde{y}, dz) = \prod_{q=1}^{N} C_n^{(p)}(\tilde{y}^q, dz^q))
$$

where

$$
d\tilde{y} = d\tilde{y}^1 \times \ldots \times d\tilde{y}^N \quad \text{and} \quad dz = dz^1 \times \ldots \times dz^N
$$

are the infinitesimal neighborhoods of the points

$$
\tilde{y} = (\tilde{y}^1, \ldots, \tilde{y}^N) \in (E^p)^N \quad \text{and} \quad z = (z^1, \ldots, z^N) \in E^N
$$

then the transition (51) can be rewritten as follows

$$
\begin{aligned}
P\left(\xi_n \in dz \mid \xi_{n-1} = x\right) &= \int_{E^p} \mathcal{B}_n^{(p)}(x, d\tilde{y}) \ \mathcal{C}_n^{(p)}(\tilde{y}, dz)) \\
&= (\mathcal{B}_n^{(p)} \mathcal{C}_n^{(p)})(x, dz)
\end{aligned}
$$

By construction the transitions (51) are once more decomposed into two stages

$$
\xi_{n-1} \in E^N \xrightarrow{\text{Selection/Mutation}} \tilde{\xi}_{n-1} = (\tilde{\xi}_{n-1}^1, \ldots, \tilde{\xi}_{n-1}^N) \in (E^p)^N \xrightarrow{\text{p-Crossover}} \xi_n \in E^N
$$

with

$$
\begin{aligned}
P\left(\tilde{\xi}_{n-1} \in d\tilde{y} \mid \xi_{n-1} = x\right) &= \mathcal{B}_n^{(p)}(x, d\tilde{y}) \\
P\left(\xi_n \in dz \mid \tilde{\xi}_{n-1} = \tilde{y}\right) &= \mathcal{C}_n^{(p)}(\tilde{y}, dz)
\end{aligned}
$$

Let us describe more precisely this two stage transition.

- **Selection/Mutation**
  Given the configuration $\xi_{n-1} \in E^N$, the particle system

  $$
  \tilde{\xi}_{n-1} = \left(\tilde{\xi}_{n-1}^1, \ldots, \tilde{\xi}_{n-1}^N\right) \in (E^p)^N
  $$

  consists of N-conditionally independent and E^p-valued random variables

  $$
  \tilde{\xi}_{n-1}^i = \left(\tilde{\xi}_{n-1,1}^i, \ldots, \tilde{\xi}_{n-1,p}^i\right) \in E^p, \qquad 1 \leq i \leq N
  $$

with common law

$$\left(\underbrace{\Phi_n(m(x)) \otimes \ \ldots \ \otimes \Phi_n(m(x))}\right)$$
$$\text{p-times}$$

In other words each E^p-valued random variable $\tilde{\xi}^i_{n-1}$, $1 \leq i \leq N$, consists in p-conditionally independent with common law $\Phi_n(m(x))$. By definition of the mappings Φ_n, and for each $1 \leq i \leq N$, the d-vector random variables

$$\left(\tilde{\xi}^i_{n-1,1}, \ldots, \tilde{\xi}^i_{n-1,p}\right)$$

can be regarded as the result of d-conditionally independent (selection/mutation)-transitions from the parent individual ξ^i_{n-1}.

- **p-Crossover**
  Given the configuration, $\tilde{\xi}_{n-1}$ each E^p-valued sub-configuration

$$\tilde{\xi}^i_{n-1} = \left(\tilde{\xi}^i_{n-1,1}, \ldots, \tilde{\xi}^i_{n-1,p}\right) \in E^p, \qquad 1 \leq i \leq N$$

cross-over according to the transition probability $C_n^{(p)}$, that is

$$\tilde{\xi}^i_{n-1} \ \longrightarrow \ \xi^i_n \ \sim \ C_n^{(p)}\left(\tilde{\xi}^i_{n-1}, \cdot\right), \qquad 1 \leq i \leq N$$

As noticed this GA do not fit into the simple GA framework of section 7.1. Therefore we need to find another strategy to study its asymptotic behavior. In order to apply Theorem 6.1 we first need to check that the one step mappings of the limiting system are continuous. In what follows the state space E is assumed to be a Polish space.

Let us set

$$\varphi(x_1, \ldots, x_p) = \prod_{k=1}^{p} f_k(x_k), \qquad f_1, \ldots, f_p \in \mathcal{C}_b(E)$$

and assume that $\mu_m \in \mathcal{P}(E)$ is a sequence which weakly converges to $\mu \in \mathcal{P}(E)$ when m tends to infinity. For a test function φ of the previous form we obtain

$$\left(\underbrace{\Phi_n(\mu_m) \otimes \ \ldots \ \otimes \Phi_n(\mu_m)}\right)(\varphi) = \prod_{k=1}^{p} \Phi_n(\mu_m)(f_k)$$
$$\text{p-times}$$

and therefore

$$\lim_{m\to\infty} \left(\underbrace{\Phi_n(\mu_m) \otimes \ \ldots \ \otimes \Phi_n(\mu_m)}\right)(\varphi) = \left(\underbrace{\Phi_n(\mu) \otimes \ \ldots \ \otimes \Phi_n(\mu)}\right)(\varphi)$$
$$\qquad\qquad\text{p-times}\qquad\qquad\qquad\qquad\text{p-times}$$

Since linear combinations of such functions are dense in $\mathcal{C}_b(E^p)$ one can check that the one step mappings of the limiting system (49) are continuous. Now, if E is compact then Theorem 6.1 implies that the N-approximating distributions

$$\eta^N_n = \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi^i_n}, \qquad n \geq 0$$

weakly converge to the desired distribution η_n as N tends to infinity.

Next we discuss the Lipschitz condition (28) of Theorem 6.2 when the state space is finite and for $p = 2$, that is for a pair cross-over GA. The extension to any $p \geq 2$ will be straightforward.

For any $\varphi \in \mathcal{B}_b(E^2)$, $n \geq 1$, $\mu, \eta \in \mathcal{P}(E)$ we clearly have that

$$\left| \Phi_n(\mu)^{\otimes 2}(\varphi) - \Phi_n(\eta)^{\otimes 2}(\varphi) \right|$$

$$\leq \sup_{x_1 \in E} \left| \Phi_n(\mu)(\varphi(x_1, .)) - \Phi_n(\eta)(\varphi(x_1, .)) \right| + \left| \Phi_n(\mu)(\overline{\varphi}_{n,\eta}) - \Phi_n(\eta)(\overline{\varphi}_{n,\eta}) \right|$$

with

$$\overline{\varphi}_{n,\eta}(x_1) \stackrel{\text{def.}}{=} \Phi_n(\mu)(\varphi(x_1, .)) \in \mathbb{R}$$

from which one concludes that for any finite subset $\mathcal{F} \in \mathcal{B}_b(E)$

$$\left\| \Phi_n(\mu)^{\otimes 2} C_n^{(2)} - \Phi_n(\eta)^{\otimes 2} C_n^{(2)} \right\|_{\mathcal{F}} \leq 2 \left\| \Phi_n(\mu) - \Phi_n(\eta) \right\|_{\Gamma_{n,\eta}(\mathcal{F})}$$

with

$$\Gamma_{n,\eta}(\mathcal{F}) \stackrel{\text{def.}}{=} \left\{ C_n^{(2)}(f(x_1, .)) \; ; \; f \in \mathcal{F}, \; x_1 \in E \right\}$$

$$\bigcup \left\{ x_1 \mapsto \int_E \Phi_n(\eta)(dx_2)(C_n^{(2)} f)(x_1, x_2) \; ; \; f \in \mathcal{F} \right\}$$

Since the one step mappings $\{ \Phi_n \; ; \; n \geq 1 \}$ satisfy assumption (28) one concludes that the one step mappings of (49) again satisfy (28).

One can clearly translate conditions presented in section 6.2.3 to get large deviations principles. For instance according to the remark 6.4 condition $(\mathcal{H})$ is satisfied when E is finite and for any $n \geq 1$ and $(x_1, \ldots, x_p) \in E^p$ and $y \in E$

$$C_n^{(2)}((x_1, \ldots, x_p), \{y\}) > 0$$

7.4 A Genetic Model with Random Selection/Mutation

The random selection/mutation GA model presented in this last section is the N-approximating model associated with the limiting $\mathcal{P}(E)$-valued dynamical system given by

$$\eta_n = \alpha \; \eta_{n-1} K_n + (1 - \alpha) \; \Psi_n(\eta_{n-1}) \tag{52}$$

where $\alpha \in]0, 1[$ and as usual

$$\forall f \in \mathcal{B}_b(E), \qquad \Psi_n(\eta)(f) = \frac{\eta(g_n f)}{\eta(g_n)}$$

The transition probabilities of the N-IPS model $\xi_n = (\xi_n^1, \ldots, \xi_n^N)$, $n \geq 0$, associated with the limiting process (52) take the form

$$P\left(\xi_n \in dz \,|\, \xi_{n-1} = x\right) = \prod_{p=1}^{N} \left(\alpha \; m(x)K_n + (1 - \alpha) \; \Psi_n(m(x))\right)(dz^p)$$

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

In other words given the configuration ξ_{n-1}, the next configuration ξ_n consists of N-conditionally random variables with common law

$$\alpha \; \frac{1}{N} \sum_{i=1}^{N} K_n(\xi_{n-1}^i, .) + (1 - \alpha) \; \sum_{i=1}^{N} \frac{g_n(\xi_{n-1}^i)}{\sum_{j=1}^{N} g_n(\xi_{n-1}^j)} \; \delta_{\xi_{n-1}^i}$$

Independently of each other each particle ξ_n^i, $1 \le i \le N$, can be sampled as follows:

- With a probability α we choose randomly and uniformly in the configuration ξ_{n-1} an individual ξ_{n-1}^j, $1 \le j \le N$. Then, the selected individual ξ_{n-1}^j evolves randomly according to the transition K_n.

- With a probability $(1-\alpha)$ we select randomly in the configuration ξ_{n-1} an individual with distribution

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

When the state space E is compact Theorem 6.1 implies that for each $n \ge 0$ the N-approximating measures

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

weakly converge to the desired distributions η_n as $N \to \infty$.

We also notice that the Lipschitz condition of Theorem 6.2 needed to get $\mathbb{L}^p$-rates and exponential bounds is satisfied. To see this claim we simply note that for any $n \ge 1$, $\mu, \eta \in \mathcal{P}(E)$ and $f \in \mathcal{B}_b(E)$

$$|\Phi_n(\mu)(f) - \Phi_n(\eta)(f)|$$

$$\le \alpha \; |\mu K_n(f) - \eta K_n(f)| + (1 - \alpha) \; a_n \; (\mu(g_n f) - \eta(g_n f) + \|f\| \; |\mu(g_n) - \eta(g_n)|)$$

$$\le (\alpha + (1 - \alpha) \; a_n \; \max(1, \|f\|)) \; \|\mu - \eta\|_{\mathcal{F}_n(f)}$$

with

$$\mathcal{F}_n(f) = \{g_n \; f\} \cup \{g_n\} \cup \{K_n f\}$$

This clearly implies that (h1) is satisfied for any finite subset $\mathcal{F} \subset \mathcal{B}_b(E)$ with

$$C_{n,\eta}(\mathcal{F}) = \alpha + (1 - \alpha) \; a_n \; \max\left(1, \max_{f \in \mathcal{F}} \|f\|\right)$$

and

$$\Theta_{n,\eta}(\mathcal{F}) = \bigcup_{f \in \mathcal{F}} \mathcal{F}_n(f)$$

The large deviations principles for the empirical measures on the path space stated in Theorem 6.3 entirely rely on the existence of a collection of reference probability measures $\lambda_n \in \mathcal{P}(E)$ such that for any $\eta \in \mathcal{P}(E)$ and $n \geq 1$

$$\Phi_n(\eta) \sim \lambda_n$$

This condition is not met in our settings since the mappings $\{\Psi_n \; ; \; n \geq 1\}$ are such that for any $\mu \in \mathcal{P}(E)$

$$\Psi_n(\mu) \sim \mu$$

and this contradicts condition $(\mathcal{H})$.

Nevertheless one can use Theorem 6.5 to deduce large deviations principles for the particle density profiles $\{\eta_n^N \; ; \; N \geq 1\}$. Let us check that the one step mappings

$$\Phi_n(\eta) = \alpha \; \eta K_n + (1 - \alpha) \; \Psi_n(\eta)$$

satisfy condition $(\mathcal{H}')$. We observe that for any bounded Borel function $f : E \to [0, \infty[$ we have that

$$\Phi_n(\eta)(f) \leq \; \alpha \; \eta K_n(f) + \; a_n \; (1 - \alpha) \; \eta(f)$$

Therefore, since $a_n \geq 1$ one gets

$$\Phi_n(\eta)(f) \leq \; a_n \; \eta K_n^{(\alpha)} f$$

where $K_n^{(\alpha)}$ is the Markov transition probability kernel on E given by

$$K_n^{(\alpha)} = \alpha \; K_n \; + \; (1 - \alpha) \; Id$$

As noticed in remark 7.4 this implies that $(\mathcal{H}')$ is satisfied.

References

[1] Del Moral, P.; Guionnet, A. (1999) "A central limit theorem for nonlinear filtering using interacting particle systems", *Ann. Appl. Probab.* **9**(2): 275–297.

[2] Del Moral, P.; Guionnet, A. (1998) "Large deviations for interacting particle systems. Applications to nonlinear filtering problems", *Stochastic Processes and their Applications* **78**: 69–95.

[3] Del Moral, P.; Guionnet, A. (1998) "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.

[4] Del Moral, P.; Guionnet, A. (1999) "On the stability of measure valued processes with applications to filtering", *C.R. Acad. Sci., Paris*, t. 328, Série I.

[5] Del Moral, P.; Jacod, J. (1999) "The Monte-Carlo method for filtering with discrete time observations. Central limit theorems", *Publications du Laboratoire de Probabilités*, Paris VI, No 515.

[6] Del Moral, P.; Ledoux, M. (2000) "Convergence of empirical processes for interacting particle systems with applications to nonlinear filtering", *Journal of Theoret. Probability* **13**(1): 225–257.

[7] Del Moral, P.; Miclo, L. (2000) "Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to non-linear filtering", to appear in J. Azéma, M. Emery, M. Ledoux & M. Yor (Eds.) *Séminaire de Probabilités.* Lecture Notes in Mathematics, Vol. 1729, Springer-Verlag, Berlin: 1–145.

[8] Del Moral, P.; Miclo, L. (2000) "A Moran particle system approximation of Feynman-Kac formulae", *Publications du Laboratoire de Statistiques et Probabilités*, No 11-98 (1998). To appear in *Stochastic Processes and their Applications.*

[9] Del Moral, P.; Miclo, L. (1999) "Asymptotic stability of nonlinear semigroups of Feynman-Kac type", *Publications du Laboratoire de Statistiques et Probabilités*, No 04-99 (1999).

[10] Del Moral, P. (1996) "Nonlinear filtering: interacting particle resolution", *Markov Processes and Related Fields* **2**.

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

[12] Shiryaev, A.N. *Probability*, Second edition, Springer, Berlin.

[13] Vose, M.D. (1997) "Logarithmic convergence of random heuristic search", *Evolutionary Computation* **4**(4): 395–404.

[14] Prugel-Bennet, A.; Rogers, A. (1999) "Modelling GA dynamics". To appear in the *Proceedings of the Second Evonet Summerschool on Theoretical Aspects of Evolutionary Computing*, Springer, Berlin.

[15] Ackley, D.H. (1987) *A Connectionist Machine for Genetic Hill-Climbing.* Kluwer, Boston.

[16] Agapie, A. (1997) "Genetic algorithms: minimal conditions for convergence", in: J.-K. Hao, E. Lutton, E. Ronald, M. Schoenauer & D. Snyers (Eds.) *Artificial Evolution'97*, LNCS, Springer-Verlag, Berlin.

[17] Altenberg, L. (1995) "The schema theorem and Price's theorem", in: L. D. Whitley & M. D. Vose (Eds.) *Foundations of Genetic Algorithms 3*, Morgan Kaufmann, San Mateo, CA: 23–49.

[18] Ankenbrandt, C.A. (1991) "An extention to the theory of convergence and a proof of the time complexity of genetic algorithms", in: G. J. E. Rawlins (Ed.) *Foundations of Genetic Algorithms*, Morgan Kaufmann, San Mateo, CA: 53–68.

[19] Antonisse, J. (1989) "A new interpretation of schema notation that overturns the binary encoding constraint", in: J. D. Schaffer (Ed.) *Proceedings of the 3rd International Conference on Genetic Algorithms*, Morgan Kaufmann, San Mateo, CA: 86–91.

[20] Bäck, T. (1992) "The interaction of mutation rate, selection, and self-adaptation in genetic algorithm", in: R. Manner & B. Manderick (Eds.) *Proceedings of the 2nd Conference on Parallel Problems Solving from Nature*, Morgan Kaufmann, San Mateo CA: 85–94.

[21] Bäck, T. (1993) "Optimal mutation rate in genetic search", in: S. Forrest (Ed.) *Proceedings of the 5th International Conference on Genetic Algorithms*, Morgan Kaufmann, San Mateo, CA: 2–8.

[22] Bäck, T. (1995) *Evolutionary Algorithms in Theory and Practice*. Oxford University Press, New York.

[23] Baluja, S. (1995) "An empirical comparison of seven iterative and evolutionary function optimization heuristics", Technical Report CMU-CS-95-193, Carnegie Mellon University.

[24] Berny, A. (1999) *Statistical Machine Learning and Combinatorial Optimization*. To appear in the *Proceedings of the Second Evonet Summerschool on Theoretical Aspects of Evolutionary Computing*, Springer, Berlin.

[25] Beyer, H.G. (1993) "Toward a theory of evolution strategies: Some asymptotical results for the $(1, +\lambda)$-theory", *Evolutionary Computation* **1**(2): 165–188.

[26] Beyer, H.G. (1994) "Toward a theory of evolution strategies: the (μ, λ)-theory. *Evolutionary Computation* **2**(4): 381–407.

[27] Beyer, H.G. (1995) "Toward a theory of evolution strategies: on the benefit of sex - the $(\mu/\mu, \lambda)$-theory", *Evolutionary Computation* **3**(1): 81–111.

[28] Beyer, H.G. (1995) "Toward a theory of evolution strategies: self-adaptation", *Evolutionary Computation* **3**(3): 311–347.

[29] Blumer, M.G. (1980) *The Mathematical Theory of Quantitative Genetics*. Clarendon Press, Oxford.

[30] Booker, L.B. (1993) "Recombination distributions for genetic algorithms", in: L. D. Whitley (Ed.) *Foundations of Genetic Algorithms 2*, Morgan Kaufmann, Palo Alto, CA: 29–44.

[31] Bruce, I.D.; Simpson, R.J. (1999) "Evolution determined by trajectory of expected populations: sufficient conditions with application to crossover", *Evolutionary Computation* **7**(2): 151–171.

[32] Cerf, R. (1994) *Une Théorie Assymptotique des Algorithmes Génétiques*. PhD thesis, Université de Montpellier II.

[33] Cerf, R. (1996) "An asymptotic theory of genetic algorithms", in: J.-M. Alliot, E. Lutton, E. Ronald, M. Schoenauer & D. Snyers (Eds.) *Artificial Evolution*, volume 1063 of *LNCS*. Springer Verlag, Berlin.

[34] Chakraborty, U.; Deb, K.; Chakraborty, M. (1996) "Analysis of selection algorithms: a markov chain approach", *Evolutionary Computation* **4**(2): 133–168.

[35] Darroch, J.N.; Senata, E. (1965) "On quasi-stationary distributions in absorbing discrete time finite markov chains", *Journal of Applied Probability*: 88–100.

[36] Das, R.; Whitley, D. (1991) "The only challenging problems are deceptive: global search by solving order-1 hyperplanes", in: R. K. Belew & L. B. Booker (Eds.) *Proceedings of the 4th International Conference on Genetic Algorithms*, Morgan Kaufmann, Palo Alto CA: 166–173.

[37] Davis, L. (1991) *Towards an Exploration of the Simulated Annealing Exploration Theory onto the Simple Genetic Algorithm*. PhD Thesis, University of Florida, Gainesville, FL.

[38] Davis, T.E.; Principe, J.C. (1991) "A simulated annealing like convergence theory for simple genetic algorithm", in: R. K. Belew & L. B. Booker (Eds.) *Proceedings of the 4th International Conference on Genetic Algorithms*, Morgan Kaufmann, Palo Alto CA: 174–181.

[39] Droste, S.; Jensen, T.; Wegener, I. (1998) "A rigorous complexity analysis of the (1+1) evolutionary algorithm for linear functions with boolean inputs", in: *Proceedings of the Fifth IEEE International Conference on Evolutionary Computation*, IEEE Press.

[40] Fogel, D.B. (1992) "An analysis of evolutionary programming", in: D. B. Fogel & W. Atmar (Eds.) *Proceedings of the 1st Annual Conference on Evolutionary Programming*, Evolutionary Programming Society: 43–51.

[41] Fogel, L.J. (1962) "Autonomous automata", *Industrial Research* **4**: 14–19.

[42] Francois, O. (1999) "An evolutionary strategy for global minimization and its markov chain analysis", *IEEE Transactions on Evolutionary Computation*.

[43] Garnier, J.; Kallel, L. (1998) "Statistical distribution of the convergence time for long k-path problems", Technical Report 378, CMAP, Ecole Polytechnique, Avril 1998; to appear in *IEEE Transactions on Evolutionary Computation*.

[44] Garnier, J.; Kallel, L. (1999) "Statistical distribution of the convergence time for long k-path problems", *IEEE Transactions on Evolutionary Computation*.

[45] Garnier, J.; Kallel, L.; Schoenauer, M. (1999) "Rigourous results of the first hitting times for binary mutations", *Evolutionary Computation* **7**(2): 173–203.

[46] Gieringer, H. (1944) "On the probability theory of linkage in mendelian heridity", *Annals of Math. Stat.* **15**: 25–57.

[47] Goldberg, D.E. (1989) *Genetic Algorithms in Search, Optimization and Machine Learning.* Addison Wesley, Reading MS.

[48] Goldberg, D.E. (1998) "The race, the hurdle and the sweet spot: lessons from genetic algorithms for the automation of design innovation and creativity", Technical Report 98007, Illinois University.

[49] Goldberg, D.E.; Deb, K. (1991) "A comparative analysis of selection schemes used in genetic algorithms", in: G. J. E. Rawlins (Ed.) *Foundations of Genetic Algorithms*, Morgan Kaufmann, Palo Alto CA: 69–93.

[50] Goldberg, D.E.; Richardson, J. (1987) "Genetic algorithms with sharing for multi-modal function optimization", in: J. J. Grefenstette (Ed.) *Proceedings of the 2nd International Conference on Genetic Algorithms*, Lawrence Erlbaum Associates: 41–49.

[51] Goldberg, D.E.; Segrest, P. (1987) "Finite markov chain analysis of genetic algorithms", in: J. J. Grefenstette (Ed.) *Proceedings of the 2nd International Conference on Genetic Algorithms*, Lawrence Erlbaum Associates: 1–8.

[52] Grefenstette, J.J. (1991) "Conditions for implicit parallelism", in: G. J. E. Rawlins (Ed.) *Foundations of Genetic Algorithms*, Morgan Kaufmann, Palo Alto CA: 252–261.

[53] Hartl, H.F.; Belew, R.K. (1990) "A Global Convergence Proof for a Class of Genetic Algorithms", Technical report, Technische Universität Wien.

[54] Holland, J.H. (1962) "Outline for a logical theory of adaptive systems", *Journal of the Association of Computing Machinery* **3**.

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

[56] Horn, J. (1993) "Finite markov chain analysis of genetic algorithms with niching", in: S. Forrest (Ed.) *Proceedings of the 5th International Conference on Genetic Algorithms*, Morgan Kaufmann, Palo Alto CA: 110–117.

[57] Kallel, L.; Naudts, B. (1999) "Longpaths for the genetic algorithm", in: W. Banzhaf & C. Reeves (Eds.) *Foundations of Genetic Algorithms 5*, Morgan Kaufmann, Palo Alto CA: 27–43.

[58] Kimura, M. (1964) "Diffusion models in population genetics", *Applied Probability. Methuen's Review Series in Applied Probability.*

[59] Koza, J.R. (1994) *Genetic Programming: On the Programming of Computers by means of Natural Evolution.* The MIT Press, Cambridge MS.

[60] Mahfoud, S.W. (1993) "Finite markov chain model for an alternative selection strategy for the genetic algorithm", *Complex Systems* **7**(2): 155–170.

[61] Wentzell, A.D.; Freidlin, M.I. (1984) *Random Perturbations of Dynamical Systems.* Springer Verlag, Berlin.

[62] Moran, P.A.P. (1958) "Random processes in genetics", *Proceedings of the Cambridge Philosophical Socity*, **64**: 60–71.

[63] Mühlenbein, H. (1991) "Evolution in space and time", in: G. J. E. Rawlins (Ed.) *Foundations of Genetic Algorithms*, Morgan Kaufmann, Palo Alto CA: 316–338.

[64] Mühlenbein, H. (1992) "How genetic algorithms really work: I. Mutation and hill-climbing", in: R. Manner & B. Manderick (Eds.) *Proceedings of the 2nd Conference on Parallel Problems Solving from Nature*, Morgan Kaufmann, Palo Alto CA: 15–25.

[65] Mühlenbein, H. (1997) "The equation for the response to selection and its use for prediction", *Evolutionary Computation* **5**(3): 303–346.

[66] Nix, A.E.; Vose, M.D. (1992) "Modeling genetic algorithms with markov chains", *Annals of Mathematics and Artificial Intelligence* **5**(1): 79–88.

[67] Prugel-Bennet, A. (1997) "Modelling evolving populations", *Theoretical Biology* **185**: 81–95.

[68] Prugel-Bennet, A.; Shapiro, J.L. (1994) "An analysis of genetic algorithms using statistical mechanics", *Physics Review Letters* **72**(9): 1305–1309.

[69] Qi, X.; Palmieri, F. (1994) "Theoretical analysis of evolutionary algorithms with an infinite population size in continuous space: basic properties of selection and mutation", *IEEE Transactions on Neural Networks* **5**(1): 102–119.

[70] Qi, X.; Palmieri, F. (1994) "Theoretical analysis of evolutionary algorithms with an infinite population size in continuous space: analysis of the diversification role of crossover", *IEEE Transactions on Neural Networks* **5**(1): 120–128.

[71] Radcliffe, N.J. (1992) "Nonlinear genetic representations", in: R. Manner & B. Manderick (Eds.) *Proceedings of the 2nd Conference on Parallel Problems Solving from Nature*, Morgan Kaufmann, Palo Alto CA: 259–268.

[72] Rattay, M. (1996) *Modelling the Dynamics of Genetic Algorithms Using Statistical Mechanics.* PhD Thesis, Manchester University.

[73] Rechenberg, I. (1973) *Evolutionstrategie: Optimierung Technisher Systeme nach Prinzipien des Biologischen Evolution.* Fromman-Holzboog Verlag, Stuttgart.

[74] Rogers, A.; Prugel-Bennet, A. (1998) "Genetic drift in genetic algorithm selection schemes", *IEEE Transactions on Evolutionary Computation.*

[75] Rogers, A.; Prugel-Bennet, A. (1999) "Modelling the dynamics of a steady state genetic algorithm", in: W. Banzhaf & C. Reeves (Eds.) *Foundations of Genetic Algorithms 5*, Morgan Kaufmann, Palo Alto CA: 57–68.

[76] Rogers, A.; Prugel-Bennet, A. (1999) "A solvable model of a hard optimization problem", to appear in the *Proceedings of the second Evonet Summerschool on Theoretical Aspects of Evolutionary Computing.*

[77] Rudolph, G. (1994) "Convergence analysis of canonical genetic algorithm", *IEEE Transactions on Neural Networks* **5**(1): 96–101.

[78] Rudolph, G. (1994) "Convergence of non-elitist strategies", in: Z. Michalewicz, J. D. Schaffer, H.-P. Schwefel, D. B. Fogel & H. Kitano (Eds.) *Proceedings of the First IEEE International Conference on Evolutionary Computation*, IEEE PRess: 63–66.

[79] Rudolph, G. (1996) "How mutation and selection solve long path problems in polynomial expected time", *Evolutionary Computation* **4**(2): 195–205.

[80] Rudolph, G. (1997) "Asymptotical convergence rates of simple evolutionary algorithms under factorizing mutation distributions", in: J.-K. Hao, E. Lutton, E. Ronald, M. Schoenauer & D. Snyers (Eds.) *Artificial Evolution'97*, Springer, Berlin: 223–233.

[81] Rudolph, G. (1997) *Convergence Properties of Evolutionary Algorithms.* Kovac, Hamburg, 1997.

[82] Shapiro, J.; Prugel-Bennet, A. (1997) "Genetic algorithm dynamics in a two well potential", in: R. K. Belew & M. D. Vose (Eds.) *Foundations of Genetic Algorithms 4*, Morgan Kaufmann, Palo Alto CA: 101–115.

[83] Schwefel, H.-P. (1981) *Numerical Optimization of Computer Models.* John Wiley & Sons, New-York. 1995 – 2nd edition.

[84] Stephens, C.; Waelbroeck, H. (1997) "Effective degree of freedom in genetic algorithms and the block hypothesis", in: Th. Bäeck (Ed.) *ICGA97*, Morgan Kaufmann, Palo Alto CA: 34–40.

[85] Stephens, C.; Waelbroeck, H. (1999) "Schemata evolution and building blocks", *Evolutionary Computation* **7**(2): 109–124.

[86] Suzuji, J. (1993) "A markov chain analysis on a genetic algorithm", in: S. Forrest (Ed.) *Proceedings of the 5th International Conference on Genetic Algorithms*, Morgan Kaufmann, Palo Alto CA: 146–153.

[87] Suzuji, J. (1997) "A further result on the markov chain model of genetic algorithms and its application to a simulated annealing-like strategy", in: R. K. Belew & M. D. Vose (Eds.) *Foundations of Genetic Algorithms 4*, Morgan Kaufmann, Palo Alto CA: 53–72.

[88] Thierens, D.; Goldberg, D.E. (1993) "Mixing in genetic algorithms", in: S. Forrest (Ed.) *Proceedings of the 5th International Conference on Genetic Algorithms*, Morgan Kaufmann, Palo Alto CA: 38–55.

[89] Trouvé, A. (1993) *Paralllisation Massive du Recuit Simul.* PhD Thesis, Université Paris XI.

[90] Trouvé, A. (1996) "Cycle decomposition and simulated annealing", *SIAM Journal Control and Optimisation* **34**(3): 966–986.

[91] Laarhoven, P.J. van; Aarts, E.H.L. (1987) *Simulated Annealing: Theory and Applications.* Kluwer Academic Press, Dordrecht.

[92] Vose, M.D.; Liepins, G.E. (1991) "Punctuated equilibria in genetic search", *Complex Systems* **5**(1): 31–44.

[93] Vose, M.D. (1991) "Generalizing the notion of schemata in genetic algorithms", *Artificial Intelligence* **50**(3): 385–396.

[94] Vose, M.D. (1993) "Modeling simple genetic algorithms", in: L. D. Whitley (Ed.) *Foundations of Genetic Algorithms 2*, Morgan Kaufmann, Palo Alto CA: 63–73.

[95] Vose, M.D.; Wright, A.H. (1995) "Stability of vertex fixed points and applications", in: L. D. Whitley & M. D. Vose (Eds.) *Foundations of Genetic Algorithms 3*, Morgan Kaufmann, Palo Alto CA: 103–113.

[96] Vose, M.D. (1999) *The Simple Genetic Algorithm: Foundations and Theory.* The MIT Press, Cambridge MS.

[97] Wilson, S.W. (1991) "GA-easy does not imply steepest-ascent optimizable", in: R. K. Belew & L. B. Booker (Eds.) *Proceedings of the 4th International Conference on Genetic Algorithms*, Morgan Kaufmann, Palo Alto CA.

[98] Zhigljavsky, A.A. (1992) *Theory of Global Random Search.* Mathematics and its Applications. Kluwer, Dordrecht.

[99] Prügel-Bennett, A.; Shapiro, J.L. (1994) "An analysis of genetic algorithms using statistical mechanics", *Phys. Rev. Lett.* **72**(9): 1305–1309.

[100] Prügel-Bennett, A.; Shapiro, J.L. (1997) "The dynamics of a genetic algorithm for simple random ising systems", *Physica D* **104**: 75–114.

[101] Levin, E.; Tishby, N.; Solla, S.A. (1990) "A statistical approach to learning and generalization in layered neural networks", *Proceedings of the IEEE* **78**(10): 1568–1574.

[102] Nimwegen, E. van; Crutchfield, J.P.; Mitchell, M. (1997) "Finite populations induce metastability in evolutionary search", *Physics Letters A* **229**(2): 144–150.

[103] Rowe, J.E. (2000) "Cyclic attractors and quasispecies adaptibility", in: L. Kallel, B. Naudts & A. Rogers (Eds.) *Theoretical Aspects of Evolutionary Computing*, Springer Verlag, Berlin.

[104] Rowe, J.E. (2000) "The dynamical systems model of the simple genetic algorithm", in: L. Kallel, B. Naudts & A. Rogers (Eds.) *Theoretical Aspects of Evolutionary Computing*, Springer Verlag, Berlin.

[105] Rowe, J.E.; Vose, M.D.; Wright, A.H. "Group properties of crossover and mutation". In preparation.

[106] Ronnewinkel, C.; Wilke, C.O.; Martinez, T. (2000) "Genetic algorithms in time-dependent environments", in: L. Kallel, B. Naudts & A. Rogers (Eds.) *Theoretical Aspects of Evolutionary Computing*, Springer Verlag, Berlin.

[107] Rowe, J.E. (1999) "Finding attractors for periodic fitness functions", in: W. Banzhaf, J. Daida, A.E. Eiben, M.H. Garzon, V. Honovar, M. Jakiela & R.E. Smith (Eds.) *Proc. Genetic and Evolutionary Computation Conference (GECCO99)*, Morgan Kaufmann, Palo Alto CA.

[108] Rowe, J.E. (1999) "Population fixed-points for functions of unitation", in: W. Banzhaf & C. Reeves (Eds.) *Foundations of Genetic Algorithms 5*, Morgan Kaufmann, Palo Alto CA: 69–84.