INRIA

Team ALEA

Advanced Learning Evolutionary Algorithms

Bordeaux - Sud-Ouest

Theme : Stochastic Methods and Models

Activity Report

2009

Table of contents

The team ALEA has been created on March the 1st, 2009.

1. Team

Research Scientist

Pierre Del Moral [Team leader, Research Director (DR) INRIA, HdR]
François Caron [Research Associate (CR) INRIA]

Faculty Member

Pierrick Legrand [University Bordeaux 2]
Arnaud Doucet [Associate Member, University of British Columbia]
Josselin Garnier [Associate Member, University Paris 7]

Technical Staff

Frédéric Proia [Development Engineer]
Adrien Todeschini [Development Engineer]

PhD Student

Michele Pace [INRIA]
Aurélie Le Cain [CEA]
François Giraud [CEA]
Hu Peng [Bourse Ministère]

Post-Doctoral Fellow

Emilia Tantar [INRIA]
Alexandru-Adrian Tantar [INRIA + Région Aquitaine]

Visiting Scientist

Emmanuel Rio [University Versailles (Until July 2009)]
Ming Lei [MétéoFrance]
Shulan Hu [Wuhan University]

Administrative Assistant

Marie-Christine Echegoyen [Team Assistant]

2. Overall Objectives

2.1. Overall Objectives

In recent years, a new generation of numerical algorithms have begun to spread through the scientific community. Surprisingly enough, up to a few exceptions, many of these modern ideas do not really come from physics, but from biology and ethology. In a growing number of scientific disciplines, the researchers are now interpreting real world processes and engineering type systems less like purely deterministic and crude clockwork mechanisms, but much more like random and sophisticated biology inspired processes. This new generation of engineering models is based on stochastic ideas and natural principles like : Chance, randomness, interactions, reinforcement strategies, exploration rules, biology-inspired adaptation and selection transitions, learning, reproduction, birth and death, ancestral lines and historical processes, genealogical tree evolutions, as well as self-organization principles, and many others.

These biology-inspired stochastic algorithms are often presented as natural heuristic simulation schemes without any mathematical foundations, nor a single performance analysis ensuring their convergence, nor even a single theoretical discussion that clarifies the applicability of these models. An important aspect of our project is to create a concrete bridge between pure and applied probability, statistics, biology, stochastic engineering and computer sciences. This fundamental bridging effort is probably one of the most important key to turn real nature's processes into engineering devices and stochastic algorithms, by learning what can be abstracted, copied or adapted. In the reverse angle, we can mention that these abstracted models adapting nature mechanisms and biological capabilities also provides a better understanding of the real processes.

By essence, *the team-project is not a single application-driven research project* . The reasons are three folds: Firstly, the same stochastic algorithm is very often used in a variety of application areas. On the other hand every application domain area offers a series of different perspectives that can be used to improve the design and the performances of the algorithms. Last but not least, concrete industrial applications, as well as most of concrete problems arising in biology, physics and chemistry, require a specific attention. In general, we do not use a single class of stochastic algorithm but a broader set of stochastic search algorithms that incorporates facets of nature inspired strategies.

Our research project is centered on two central problems in advanced stochastic engineering: *Bayesian inference and rare event simulation* and more particularly *unsupervised learning, multi-target tracking, spike sorting, data assimilation and forecasting, as well as infection spreads inference.* These important and natural research directions have emerged as logical parts of the team project combined with interdisciplinary approaches well-represented at Bordeaux university campus.

The fundamental and the theoretical aspects of our research project are essentially concerned with the stochastic analysis of the following three classes of biology inspired stochastic algorithms: *branching and interacting particle systems, reinforced random walks and self-interacting processes, random tree based models.* One of our prospective research project is to apply the Bayesian learning methodology and the recent particle filter technology to the design of a new generation of interactive `evolutionary computation` and `stochastic art composition` models.

3. Scientific Foundations

3.1. Description

This idea of analyzing nature systems and transferring the underlying principles into stochastic algorithms and technical implementations is one of the central component of the ALEA team project. Adapting nature mechanisms and biological capabilities clearly provides a better understanding of the real processes, and it also improves the performance and the power of engineers devices. Our project is centered on both the understanding of biological processes in terms of mathematical, physical and chemical models, and on the other hand, on the use of these biology inspired stochastic algorithms to solve complex engineering problems.

There is a huge series of virtual interfaces, robotic devices, numerical schemes and stochastic algorithms which were invented mimicking biological processes or simulating natural mechanisms. The terminology *"mimicking or simulating"* doesn't really mean to find an exact copy of natural processes, but *to elaborate the mathematical principles so that they can be abstracted from the original biological or physical model.* In our context, the whole series of evolutionary type principles discussed in previous sections can be abstracted into only three different and natural classes of stochastic algorithms, depending on the nature of the biology-inspired interaction mechanism used in the stochastic evolution model. These three stochastic search models are listed below :

1) *Branching and interacting particle systems (birth and death chains, spatial branching processes, mean-field interaction between generations):*

The first generation of adaptive branching-selection algorithms is very often built on the same genetic type paradigm: When exploring a state space with many particles, we duplicate better fitted individuals at the expense of light particles with poor fitness die. From a computational point of view, we generate a large number of random problem solvers. Each one is then rated according to a fitness or performance function defined by the developer. Mimicking natural selection, an evolutionary algorithm selects the best solvers in each generation and breeds them.

2) *Reinforced random walks and self-interacting chains (reinforced learning strategies, interaction processes with respect to the occupation measure of the past visited sites):*

This type of reinforcement is observed frequently in nature and society, where "beneficial" interactions with the past history tend to be repeated. A new class of historical mean field type interpretation models of reinforced processes were developed by the team project leader in a pair of articles [36], [35]. Self interaction gives the opportunity to build new stochastic search algorithms with the ability to, in a sense, re-initialized their exploration from the past, re-starting from some better fitted initial value already met in the past [28], [29].

3) *Random tree based stochastic exploration models (coalescent and genealogical tree search explorations techniques on path space):*

The last generation of stochastic random tree models is concerned with biology-inspired algorithms on paths and excursions spaces. These genealogical adaptive search algorithms coincide with genetic type particle models in excursion spaces. They have been applied with success in generating the excursion distributions of Markov processes evolving in critical and rare event regimes, as well as in path estimation and related smoothing problems arising in advanced signal processing (cf. [37] and references therein). We underline the fact that the complete mathematical analysis of these random tree models, including their long time behavior, their propagations of chaos properties, as well as their combinatorial structures are far from being completed. This class of genealogical tree based models has been introduced in [34] for solving smoothing problems and more generally Feynman-Kac semigroups on path spaces, see also [27], [37], and references therein.

4. Application Domains

4.1. Panorama

This short section is only concerned with the list of concrete application domains developed by our team project on Bayesian inference and unsupervised learning, nonlinear filtering and rare event analysis. Most of these application areas result from fruitful collaborations with other national institutes through a series of four recently selected ANR research projects and one INRIA-INRA joint research project.

Three application domains are directly related to evolutionary computing, particle filtering and Bayesian inference are currently investigated by our team project, two of them are 2008's ANR research projects :

1. **Multi-target tracking**. Multi-target tracking deals with the task of estimating the states of a set of moving targets from a set of measurements obtained sequentially. These measurements may either arise from one of the targets or from clutter and the measurement-to-target association is generally unknown. This problem can then be recast as a dynamic clustering one where the clusters are the clutter and the different targets. The targets actually move over time, some targets may appear/disappear over time and the number of targets is generally unknown and time-varying. We are running this research project with the DCNS-SIS division in Toulon.

2. **Forecasting and Data assimilation :** This new application domain concerns the application of the particle filter technology and more general sequential Monte Carlo methods to data assimilation problems arising in forecasting. The ALEA team project is involved in the `ANR 2008 selected project PREVASSEMBLE` with Météo France Toulouse, the INRIA Rennes and the LMD in Paris.

3. **Virtual prairie:** This application domain of evolutionary computing is concerned with the design of ecological systems, mixed-species models and prairial ecosystems. For more details, we refer the reader to the web site of the `Virtual Prairie project` The ALEA project is a partner of the 2008 ANR SYSCOM project named MODECOL.

Three other application domains are directed related to rare event analysis using particle stochastic simulations techniques. These projects are currently investigated by our team project, two of them are 2008's ANR research projects :

1. **Watermarking of digital contents:**

The terminology watermarking refers to a set of techniques for imbedding/hiding information in a digital audio or video file, such that the change is not noticed, and very hard to remove. In order to be used in an application, a watermarking technique must be reliable. Protection false alarms and failures of traceability codes are practically not achievable without using rare event analysis. This application domain area of particle rare event technology is the subject of joint ANR 2007 research project with the IRISA-INRIA in Rennes and the LIS INPG in Grenoble. For more details, we refer the reader to the web site of the Nebbiano project. Security and Reliability in Digital Watermarking

2. **Epidemic propagations analysis:**

   This project aims at developing stochastic mathematical models for the spread of transmissible infectious diseases, together with dedicated statistical methodologies, with the intent to deliver efficient diagnostic/prediction tools for epidemiologists. This application domain area of particle rare event technology is the subject of joint ANR 2008 research project with Telcom Paristech, the Laboratoire Paul Painlevé in Lille 1 and the University of Paris 5 (cf. Programme Systèmes Complexes et Modélisation Mathématique, list of 2008 selected projects).

3. **Statistical eco-microbiology predictions:**

   This project aims at developing stochastic models and algorithms for the analysis of bacteriology ecosystems, especially in food safety. The objective is to predict and control critical risk of proliferations. This is the subject of joint research project with the INRA of Paris and Montpellier (Appel d'Offre INRIA-INRA 2008 : Systèmes Complexes).

5. New Results

5.1. Bayesian Nonparametrics

Participants: François Caron, Arnaud Doucet, Christopher Holmes, Jim Griffin, Dave Stephens.

Bayesian Nonparametric Models on Decomposable Graphs

Over recent years Dirichlet processes and the associated Chinese restaurant process (CRP) have found many applications in clustering while the Indian buffet process (IBP) is increasingly used to describe latent feature models. These models are attractive because they ensure exchangeability (over samples). We propose here extensions of these models where the dependency between samples is given by a known decomposable graph. These models have appealing properties and can be easily learned using Markov Chain Monte Carlo and Sequential Monte Carlo techniques.

This work has been presented as an invited talk at the 7th workshop on Bayesian nonparametrics, Turin, Italy, and has been accepted at the NIPS international conference [8].

Bayesian Nonparametric Models for two sample hypothesis testing

In this work [23] we describe Bayesian nonparametric procedures for two-sample hypothesis testing. That is, given two sets of samples $\mathbf{y}^{(1)} \underset{\text{iid}}{\sim} F^{(1)}$ and $\mathbf{y}^{(2)} \underset{\text{iid}}{\sim} F^{(2)}$, with $F^{(1)}, F^{(2)}$ unknown, we wish to evaluate the evidence for the null hypothesis $H_0 : F^{(1)} \equiv F^{(2)}$ verses the alternative $H_1 : F^{(1)} \neq F^{(2)}$. Our method is based upon a nonparametric Polya tree prior centered either *a priori* or empirically. We show that the Polya tree prior allows us to calculate an analytic expression for the marginal likelihood under the two hypotheses and hence provide an explicit measure of $P(H_0|\{\mathbf{y}^{(1)}, \mathbf{y}^{(2)}\})$.

5.2. Estimation of Hölderian regularity using Genetic Programming

Participants: Leonardo Trujillo, Pierrick Legrand, Jacques Levy-Vehel, Gustavo Olague.

The concept of Hölderian regularity allows us to characterize the singular structures contained within a signal [32]. A quantitative measure of the regularity of a signal can be obtained from measuring Hölder exponents, either within a local region of the signal or at each point. In our work, we are interested in measuring the pointwise Hölder exponent which is defined below.

Definition Let $f : \mathfrak{R} \to \mathfrak{R}$, $s \in \mathfrak{R}^{+*} \smallsetminus \mathbf{N}$ and $x_0 \in \mathfrak{R}$. If $f \in C^s(x_0) \Leftrightarrow \exists \eta \in \mathfrak{R}^{+*}$, and a polynom P of degree $< s$ and a constant c exist such that

$$\forall x \in B(x_0, \eta), |f(x) - P(x - x_0)| \leq c|x - x_0|^s \ , \tag{1}$$

then the pointwise Hölder exponent of f at x_0 is $\alpha_p = \sup_s \{f \in C^s(x_0)\}$.

However, this exponent is very costly to estimate in the common framework (oscillations or wavelets). Therefore, we will propose a new estimator of this exponent.

Over the past two decades, Genetic Programming (GP) has proven to be a powerful paradigm for the development of computer systems that that are able to solve complex problems without the need for substantial amounts of a priori knowledge. Indeed, GP solves two tasks simultaneously because it can search for the desired functionality and also determine the structure that the final solutions possess. The former is achieved by using the basic principles of artificial evolution, a stochastic search process that is guided by fitness and which uses simple variation operators. On the other hand, the latter is produced by eliminating strict structural restrictions on the evolving population that most evolutionary techniques require, as well as other black-box methods such as neural nets [30], [31]. The overall flexibility of GP has allowed researchers to apply it to a very diverse set of problems from different disciplines [38]. Therefore, we have chosen GP as our search and optimization algorithm to produce novel estimators of Hölderian regularity for digital images.

We obtained some new results listed below. 1) We studied the problem of estimating a measure of Hölderian regularity using a GP-based search, which is the first such study. 2) We have successfully produced several operators that are capable of extracting good estimations of the pointwise Hölder exponent when compared with a canonical estimator. 3) The evolved estimators do not require parameter tuning or design choices, because this is implicitly carried out during the optimization process. Hence, our evolved operators provide a *simpler*, in this sense, estimation method.4) We have applied our evolved operators to the problem of local image description, and we show that several of them achieve a comparable performance when compared with a canonical estimation method [39].

5.3. Local regularity and multifractal methods for image and signal analysis

Participant: Pierrick Legrand.

We reviewed some of the important and recent applications of local regularity and multifractal analysis to signal/image processing [13]. (Multi)Fractal processing of signals and images is indeed now present in numerous applications. We tried to explain in a very concrete manner how tools developed for the study of irregular functions may be applied to solve typical problems of signal processing. These problems are denoising, biomedical signal analysis, segmentation, edge detection, change detection in sequences of images, image reconstruction.

5.4. Interacting Markov chain Monte Carlo methods

Participants: Bernard Bercu, Anthony Brockwell, Pierre Del Moral, Arnaud Doucet.

This new line of research is mainly concerned with the design and the analysis of a new class of interacting stochastic algorithms for sampling complex distributions including Boltzmann-Gibbs measures and Feynman-Kac path integral semigroups arising in physics, in biology and in advanced stochastic engineering science. These interacting sampling methods can be described as adaptive and dynamic simulation algorithms which take advantage of the information carried by the past history to increase the quality of the next series of samples. One critical aspect of this technique as opposed to standard Markov chain Monte Carlo methods is that it provides a natural adaptation and reinforced learning strategy of the physical or engineering evolution equation at hand. This type of reinforcement with the past is observed frequently in nature and society, where beneficial interactions with the past history tend to be repeated. Moreover, in contrast to more traditional mean field type particle models and related sequential Monte Carlo techniques, these stochastic algorithms can increase the precision and performance of the numerical approximations iteratively. The origins of these interacting sampling methods can be traced back to a pair of articles [35], [36] by P. Del Moral and L. Miclo. These studies are concerned with biology-inspired self-interacting Markov chain models with applications to genetic type algorithms involving a competition between the natural reinforcement mechanisms and the potential attraction of a given exploration landscape.

In 2008-2009, these lines of research have been developed in three different directions :

- In [4], we introduce a novel methodology for sampling from a sequence of probability distributions of increasing dimension and estimating their normalizing constants. These problems are usually addressed using Sequential Monte Carlo (SMC) methods. The alternative Sequentially Interacting Markov Chain Monte Carlo (SIMCMC) scheme proposed here works by generating interacting non-Markovian sequences which behave asymptotically like independent Metropolis-Hastings (MH) Markov chains with the desired limiting distributions. Contrary to SMC methods, this scheme allows us to iteratively improve our estimates in an MCMC-like fashion. We establish convergence of the algorithm under realistic verifiable assumptions and demonstrate its performance on several examples arising in Bayesian time series analysis.

- In [5], We present a new class of interacting Markov chain Monte Carlo algorithms for solving numerically discrete-time measure-valued equations. The associated stochastic processes belong to the class of self-interacting Markov chains. In contrast to traditional Markov chains, their time evolutions depend on the occupation measure of their past values. This general methodology allows us to provide a natural way to sample from a sequence of target probability measures of increasing complexity. We develop an original theoretical analysis to analyze the behaviour of these iterative algorithms which relies on measure-valued processes and semigroup techniques. We establish a variety of convergence results including exponential estimates and a uniform convergence theorem with respect to the number of target distributions. We also illustrate these algorithms in the context of Feynman-Kac distribution flows.

- In [3], we present a functional central limit theorem for a new class of interaction Markov chain Monte Carlo interpretations of discrete generation measure valued equations. We provide an original stochastic analysis based on semigroup techniques on distribution spaces and fluctuation theorems for self-interaction random fields. Besides the fluctuation analysis of these models, we also present a series of sharp $\mathbb{L}_m$-mean error bounds in terms of the semigroup associated with the first order expansion of the limiting measure valued process, yielding what seems to be the first results of this type for this class of interacting processes. We illustrate these results in the context of Feynman-Kac integration semigroups arising in physics, biology and stochastic engineering science.

5.5. Mean field interacting particle systems

Participants: Ajay Jasra, Frédéric Patras, Pierre Del Moral, Arnaud Doucet, Sylvain Rubenthaler.

The design and the mathematical analysis of genetic type and branching particle interpretations of Feynman-Kac-Schroedinger type semigroups (and vice versa) has been developed by group of researchers since the beginning of the 90's. In Bayesian statistics this sampling technology is also called sequential Monte Carlo

methods. For further details, we refer to the books [33], [37], [26], and references therein. This Feynman-Kac particle methodology is increasingly identified with emerging subjects of physics, biology, and engineering science. This new theory on genetic type branching and interacting particle systems has led to spectacular results in signal processing and in quantum chemistry with precise estimates of the top eigenvalues, and the ground states of Schrodinger operators. It offers a rigorous and unifying mathematical framework to analyze the convergence of a variety of heuristic-like algorithms currently used in biology, physics and engineering literature since the beginning of the 1950's. It applies to any stochastic engineering problem which can be translated in terms of functional Feynman-Kac type measures. During the last two decades the range of application of this modern approach to Feynman-Kac models has increased revealing unexpected applications in a number of scientific disciplines including in : The analysis of Dirichlet problems with boundary conditions, financial mathematics, molecular analysis, rare events and directed polymers simulation, genetic algorithms, Metropolis-Hastings type models, as well as filtering problems and hidden Markov chains.

In the period 2008-2009, these lines of research have been developed in three different directions. To develop a concrete peer to peer interaction between applied mathematics, engineering and computer sciences, we have written two review and pedagogical book chapters on stochastic particle algorithms and their applications ([12], [17]). Our second line of research is concerned with the foundations and the mathematical analysis of mean field particle models. In 2009, we have developed a series of new important results such as exact propagation of chaos expansions of the law of particle blocks ([7], [21]), sharp and non asymptotic exponential concentration inequalities [22], and the refined stability analysis of neutral type genetic models [6]. The third line of our research is concerned with the design of new classes of stochastic particle algorithms, including particle approximate Bayesian computation [18], backward Feynman-Kac particle models [19], particle approximations of fixed parameters in hidden Markov chain models [20], and particle rare event simulation of static probability laws [16]. The details of these contributions are provided below.

5.5.1. *Review book chapters*

- In [12], we present a review on the mean field particle theory for the numerical approximation of Feynman-Kac path integrals in the context of nonlinear filtering. We show that the conditional distribution of the signal paths given a series of noisy and partial observation data is approximated by the occupation measure of a genealogical tree model associated with mean field interacting particle model.

- In [17], we provide a pedagogical introduction to the stochastic modeling and the theoretical analysis of stochastic particle algorithms. These lectures notes were written for the Machine Learning Summer School 08. We also illustrate these methods through several applications including random walk confinements, particle absorption models, nonlinear filtering, stochastic optimization, combinatorial counting and directed polymer models.

5.5.2. *Stochastic analysis*

- In [7], [21], We design a theoretic tree-based functional representation of a class of discrete and continuous time Feynman-Kac particle distributions, including an extension of the Wick product formula to interacting particle systems. These weak expansions rely on an original combinatorial, and permutation group analysis of a special class of forests. They provide refined non asymptotic propagation of chaos type properties, as well as sharp $\mathbb{L}_p$-mean error bounds, and laws of large numbers for U-statistics. Applications to particle interpretations of the top eigenvalues, and the ground states of Schroedinger semigroups are also discussed.

- The article [6], is concerned with the long time behavior of neutral genetic population models, with fixed population size. We design an explicit, finite, exact, genealogical tree based representation of stationary populations that holds both for finite and infinite types (or alleles) models. We then analyze the decays to the equilibrium of finite populations in terms of the convergence to stationarity of their first common ancestor. We estimate the Lyapunov exponent of the distribution flows with respect to the total variation norm. We give bounds on these exponents only depending on the stability with

respect to mutation of a single individual; they are inversely proportional to the population size parameter.

- The article [22] is concerned with the fluctuations and the concentration properties of a general class of discrete generation and mean field particle interpretations of non linear measure valued processes. We combine an original stochastic perturbation analysis with a concentration analysis for triangular arrays of conditionally independent random sequences, which may be of independent interest. Under some additional stability properties of the limiting measure valued processes, uniform concentration properties with respect to the time parameter are also derived. The concentration inequalities presented here generalize the classical Hoeffding, Bernstein and Bennett inequalities for independent random sequences to interacting particle systems, yielding very new results for this class of models. We illustrate these results in the context of McKean Vlasov type diffusion models, McKean collision type models of gases, and of a class of Feynman-Kac distribution flows arising in stochastic engineering sciences and in molecular chemistry.

5.5.3. *New class of particle algorithms*

- The article [18] is concerned with Approximate Bayesian computation (ABC) in inference problems where the likelihood function is intractable, or expensive to calculate. To improve over Markov chain Monte Carlo (MCMC) implementations of ABC, the use of sequential Monte Carlo (SMC) methods has recently been suggested. Effective SMC algorithms that are currently available for ABC have a computational complexity that is quadratic in the number of Monte Carlo samples and require the careful choice of simulation parame- ters. In this article an adaptive SMC algorithm is proposed which admits a computational complexity that is linear in the number of samples and determines on-the-fly the simulation parameters. We demonstrate our algorithm on a toy example and a population genetics example.

- [19], [20], we design a particle interpretation of Feynman-Kac measures on path spaces based on a backward Markovian representation combined with a traditional mean field particle interpretation of the flow of their final time marginals. In contrast to traditional genealogical tree based models, these new particle algorithms can be used to compute normalized additive functionals on-the-fly as well as their limiting occupation measures with a given precision degree that does not depend on the final time horizon. We provide uniform convergence results w.r.t. the time horizon parameter as well as functional central limit theorems and exponential concentration estimates, yielding what seems to be the first results of this type for this class of models. We also illustrate these results in the context of fixed parameter estimation in hidden Markov chain problems ([20]), as well as in computational physics and imaginary time Schroedinger type partial differential equations, with a special interest in the numerical approximation of the invariant measure associated to h-processes ([19]).

- The paper [16] discusses the rare event simulation for a fixed probability law. The motivation comes from problems occurring in watermarking and fingerprinting of digital contents, which is a new application of rare event simulation techniques. We provide two versions of our algorithm, and discuss the convergence properties and implementation issues. A discussion on recent related works is also provided. Finally, we give some numerical results in watermarking context. The pair of articles [25], [24] are concerned with the the analysis of the Fisher information matrix-based nonlinear system conversion for state estimation problems.

6. Contracts and Grants with Industry

6.1. EDF

Participants: Bernard Bercu, François Caron, Marie Chavent, Pierre Del Moral, François Dufour, Delphine Féral, Anne Gégout-Petit, Frédéric Proïa, Jérôme Saracco.

The objective of this two-year contract (140kE, 2009-2011) between the teams ALEA and CQFD and EDF - ICAME, is to develop algorithms for the recursive prediction of the electricity consumption.

6.2. DCNS

Participants: François Caron, Pierre Del Moral, Michele Pace.

The objective of this contract (20kE, 2008-2009) is to give an overview of the particle Probability Hypothesis Filter for multi-target tracking in cluttered environments, and to compare several implementations of these methods on benchmarks.

6.3. CEA

Participants: Pierre Del Moral, Alexandru-Adrian Tantar, Emilia Tantar.

The objective of this contract (10kE, 2009) is to implement interacting particle algorithms for the optimization of networks of sparse antenna arrays.

Sparse antenna arrays represent a topic of major interest in the electromagnetic measures domain, communications, etc., offering cost and space efficient solutions. From a formal point of view, the optimization of a sparse antenna array, with respect to various constraints, can be modeled over a set of continuous functions, e.g. describing directivity, lobes. Nonetheless, as a result of the non-convex and highly multi-modal nature of the functions to be optimized, classical algorithms are generally ineffective. Extending previous approaches, a Kullback-Leibler cross-entropy based stochastic paradigm has been first considered for the study, the algorithm relying on iterative adaptive changes of the probability density functions in order to explore the search space.

As a second part of the study, an extension has been proposed by adopting an evolutionary based approach. Different designs have been considered ranging from simple direct local search methods to highly complex hybrid constructions, e.g. relying on island-based models of differential evolution and evolutionary algorithms. A significant improvement of the formerly obtained results was attained, superseding the cross-entropy based approaches, previously addressed in the project. Compared to the initial solutions, the newly obtained arrays provided a higher directivity and a reduced coupling with the enclosing environment, i.e. the objectives to be attained.

Furthermore a multi-objective formulation was introduced, in order to provide a set of good compromise (Pareto) approximate solutions. In this context a new interacting particle approach was proposed and experimentally tested. Its performance guarantees (the convergence and quality of the offered solutions) remain to be theoretically addressed.

7. Other Grants and Activities

7.1. National Actions

7.1.1. ANR

7.1.1.1. ANR PREVASSEMBLE

This new application domain concerns the application of the particle filter technology and more general sequential Monte Carlo methods to data assimilation problems arising in forecasting. The ALEA team project is involved in the ANR 2008 selected project PREVASSEMBLE with Météo France Toulouse, the INRIA Rennes and the LMD in Paris.

7.1.1.2. ANR MODECOL

This application domain of evolutionary computing is concerned with the design of ecological systems, mixed-species models and prairial ecosystems. For more details, we refer the reader to the web site of the Virtual Prairie project The ALEA project is a partner of the 2008 ANR SYSCOM project named MODECOL.

7.1.1.3. ANR Nebbiano

The terminology watermarking refers to a set of techniques for imbedding/hiding information in a digital audio or video file, such that the change is not noticed, and very hard to remove. In order to be used in an application, a watermarking technique must be reliable. Protection false alarms and failures of traceability codes are practically not achievable without using rare event analysis. This application domain area of particle rare event technology is the subject of joint ANR 2007 research project with the IRISA-INRIA in Rennes and the LIS INPG in Grenoble. For more details, we refer the reader to the website of the `Nebbiano project. Security and Reliability in Digital Watermarking`

7.1.1.4. ANR Viroscopy

This project aims at developing stochastic mathematical models for the spread of transmissible infectious diseases, together with dedicated statistical methodologies, with the intent to deliver efficient diagnostic/prediction tools for epidemiologists. This application domain area of particle rare event technology is the subject of joint ANR 2008 research project with Telcom Paristech, the Laboratoire Paul Painlevé in Lille 1 and the University of Paris 5 (cf. `Programme Systèmes Complexes et Modélisation Mathématique, list of 2008 selected projects`).

7.1.1.5. ANR Propagation

To combat dramatic event such as happened in Bombay last year (coming from the sea, a terrorist commando killed more than 200 peoples in Bombay city), authorities are decided to deploy efficient sea surveillance system to protect coastal zone including sensitive infrastructures often in vicinity of important cities.

Regulation on frequencies allocation and on coastal constructions is strong constraint to be taken into account to install technical capabilities to permanently survey vulnerable littoral zones. For example, new active sensor shall be frequencies compatible within numerous existing ones in inhabited region. In this context to perform coastal surveillance, attractive solution is to deploy passive sensors networks because:

- Not necessarily compatible within existing active sensors network.
- Provide large possibilities to install the passive sensors, because, it is not needed to be on the shoreline, but can be deployed inside the territory. Such as facility offers more potential sites and then, to optimise the deployment for optimal coverage of the sensitive zone.
- Is totally undetectable by external technical means in hand of structured criminal organisations.

For these objectives, the PROPAGATION project will study, develop and experiment a demonstrator to carry out maritime traffic picture from a set of passive sensors: passive radar, AIS and optronic cameras deployed over a coastal site. This is a joint ANR project with DCNS, Thalès, Ecomer and Exavision, accepted in 2009.

7.1.2. ARC INRIA

7.1.2.1. ARC EPS

This project aims at developing stochastic models and algorithms for the analysis of bacteriology ecosystems, especially in food safety. The objective is to predict and control critical risk of proliferations. This is the subject of joint research project with the INRA of Paris and Montpellier, the University of Bretagne and ENV Alford.

7.1.2.2. ARC M2A3PC

The spread of a pathogen within a strongly anthropized perennial vegetal cover depends on many parameters acting at contrasted spatio-temporal scales. This is of paramount importance for vine and apple trees and their airborne obligated parasites (powdery and downy mildew, scab) which strongly rely on the susceptibility and status of their hosts. Both crop systems require from 17 to 20 fungicide treatments per year. Sustainable crop management goes through a better understanding of the dynamics of these epidemic diseases at various spatio-temporal scales. With this in mind we aimed at developing a plant-pathogen methodology at several spatio-temporal scales. Dedicated teams from INRIA, INRA, CIRAD and CNRS research institutes joined together to form a task force with well-known abilities on :

- the overall dynamics of vine and apple tree systems studied at the desired spatio-temporal scales, in biology, epidemiology and agronomy at the plant and lanscape scales.

- modelling and 3D vizualisation of perenial plant growth coupled to development of pathogens,

- mathematical modelling and numerical simulations of epidemic diseases propagation within a growing host population, including spatial spread of pathogens at different scales.

7.1.3. *Organization of conferences*

Pierrick Legrand participated to the organization of the Ecole d'été EA, Porquerolles, 2009 and the Journées évolutionnaires trimestrielles (JET 19).

Pierre Del Moral and François Caron were in the organizing committee of the 41èmes Journées de Statistiques in Bordeaux (May 2009, 400 participants).

7.2. International Actions

M. Pace, P. Del Moral and F. Caron organized an international workshop on Multi-target tracking in Bordeaux (25 participants).

P. Del Moral organized an international workshop on Stochastic Algorithms Analysis in Bordeaux (25 participants).

F. Caron organized an international workshop on Bayesian nonparametrics methods in Whistler, Canada (100 participants).

P. Legrand was in the organizing committee of the international conference EA 2009, in Strasbourg (50 participants).

Pierre Del Moral was the coordinator of an associate team with the University of Wuhan, China.

F. Caron gave an invited talk at the 7th workshop on Bayesian nonparametrics, Turin, Italy, 2009.

P. Del Moral in an Associate Editor of the following journals

- Applied Mathematics and Optimization,

- Stochastic Processes and their Applications (2006-2009),

- Stochastic Analysis and Applications,

- Revista de Matematica: Teoria y aplicaciones.

P. Del Moral is chief editor of the ESAIM Proceedings and guest editor of ESAIM M2AN.

7.3. Exterior research visitors

The following researchers visited the Team ALEA during 2009: Andreas Greven (U. Erlangen), Arnaud Doucet (UBC), Sumeetpal S. Singh (U. Cambridge), N. Whitteley (U. Bristol), B.N. Vo (U. Melbourne), D. Laneuville (DCNS), Emmanuel Rio (U. Versailles), Jean-Pierre Vila (INRA Montpellier), Jean-Pierre Gauchi (INRA Jouy-en-Josas), Vlada Limic (CNRS), Laurent Miclo (CNRS), Arnaud Guyader (U. Rennes), Frédéric Cérou (INRIA Rennes), Pierre Minvielle (CEA), Nicolas Champagnat (INRIA), Fabien Campillo (INRIA), Anthony Lee (U. Oxford), Shulan Hu (Wuhan University), Wu Li Ming (U. Clermont-Ferrand), Eduardo Rodriguez Tello (CINESTAV), Aline Tabet (UBC), Zhengliang Zhang (Wuhan University), Yao Nian (Wuhan University), Arnaud Guilin (U. Clermont-Ferrand, Fuqing Gao (Wuhan University).

7.4. Teaching

P. Del Moral gives two courses in the Master MIMSE on Stochastic Algorithms.

F. Caron gives one course in the Master MIMSE on Unsupervised Learning.

P. Legrand is teaching the following courses (238 hours)

- Analyse Licence 1 SDV : 32H

- Mathématiques générales Licence 1 SCIMS : 72H

- Informatique pour les mathématiques Licence 1 SCIMS : 60H

- Complément d'algèbre Licence 2 SCIMS : 72H

8. Bibliography

Year Publications

Doctoral Dissertations and Habilitation Theses

[1] A.-A. TANTAR. *Métaheuristiques parallèles hybrides pour le docking moléculaire sur grilles de calcul*, Université des Sciences et Technologie de Lille, 2009, Ph. D. Thesis.

[2] E. TANTAR. *Analyse du paysage en optimisation combinatoire multi-objectif*, Université des Sciences et Technologie de Lille, 2009, Ph. D. Thesis.

Articles in International Peer-Reviewed Journal

[3] B. BERCU, P. DEL MORAL, A. DOUCET. *A functional central limit theorem for a class of interacting Markov chain Monte Carlo models*, in "Electronic Journal of Probability", vol. 14, 2009, p. 2130-2155, http://hal.inria.fr/inria-00227534/en/CAJP.

[4] A. BROCKWELL, P. DEL MORAL, A. DOUCET. *Sequentially INteracting Markov chain Monte Carlo*, in "Annals of Statistics", 2009, http://www.math.u-bordeaux.fr/~delmoral/simcmc.pdf, To appear US CA JP .

[5] P. DEL MORAL, A. DOUCET. *Interacting Markov chain Monte Carlo methods for solving nonlinear measure-valued equations*, in "Annals of Applied Probability", 2009, http://hal.inria.fr/inria-00227508/en/, To appearCAJP.

[6] P. DEL MORAL, L. MICLO, F. PATRAS, S. RUBENTHALER. *The convergence to equilibrium of neutral genetic models*, in "Stochastic Analysis and Applications", 2009, http://hal.inria.fr/inria-00239259/en/, To appear.

[7] P. DEL MORAL, F. PATRAS, S. RUBENTHALER. *Coalescent tree based functional representations for some Feynman-Kac particle models*, in "Annals of Applied Probability", vol. 19, 2009, p. 1-50, http://arxiv.org/pdf/math/0607453.

International Peer-Reviewed Conference/Proceedings

[8] F. CARON, A. DOUCET. *Bayesian nonparametric models on decomposable graphs*, in "Neural Information Processing Systems (NIPS)", 2009, http://books.nips.cc/papers/files/nips22/NIPS2009_0562.pdf CA .

[9] D. HORVATH, L. BRILLET, S. ROY, S. CONILLEAU, A.-A. TANTAR, J.-C. BOISSON, N. MELAB, E.-G. TALBI. *Local vs. global search strategies in evolutionary GRID-based conformational sampling & docking*, in "Evolutionary Computation in Bioinformatics and Computational Biology, Evolutionary Computation, 2009. CEC '09. IEEE Congress on", May 2009, p. 247–254.

[10] M. PACE. *How a Genetic Algorithm Learns to Play Traveler's Dilemma by Choosing Dominated Strategies to Achieve Greater Payoffs*, in "IEEE Symposium on Computational Intelligence and Games", 2009.

Scientific Books (or Scientific Book chapters)

[11] P. COLLET, P. LEGRAND, C. BOURGEOIS-REPUBLIQUE, V. PEAN, B. FRACHET. *Using interactive evolutionary algorithms to help fit cochlear implants*, in "Optimization in Signal and Image Processing", P. SIARRY (editor), vol. ISBN: 9781848210448 - 2009, John Wiley and Sons,Iste, June 2009.

[12] P. DEL MORAL, F. PATRAS, S. RUBENTHALER. *A mean field theory of nonlinear filtering*, in "In D. Crisan and B. Rosovsky, editors, Oxford Handbook of Nonlinear Filtering", Oxford University Press, 2009, http://hal.inria.fr/inria-00238398/en/.

[13] P. LEGRAND. *Local regularity and multifractal methods for image and signal analysis*, in "Scaling, Fractals and Wavelets", P. ABRY, P. GONÇALVES, J. LEVY-VEHEL (editors), Digital signal and image processing series, Iste-Wiley, January 2009.

[14] A.-A. TANTAR, N. MELAB, E.-G. TALBI. *An Analysis of Dynamic Mutation Operators for Conformational Sampling*, in "Biologically-Inspired Optimisation Methods", A. LEWIS, S. MOSTAGHIM, M. RANDALL (editors), Studies in computational intelligence, Vol. 210, Springer, 2009, p. 291–323, http://www.springerlink.com/content/284p361vnk753048.

[15] E. TANTAR, A.-A. TANTAR, N. MELAB, E.-G. TALBI. *Landscape Analysis in Adaptive Metaheuristics for Grid Computing*, in "Parallel Programming, Models and Applications in Grid and P2P Systems", F. XHAFA (editor), Advances in Parallel Computing, IOS Press, 2009, p. 313–344.

Research Reports

[16] F. CEROU, P. DEL MORAL, A. GUYADER. *Rare event simulation for a static distribution*, INRIA, 2009, http://hal.inria.fr/inria-00350762/fr/, Technical report.

[17] P. DEL MORAL, A. DOUCET. *Particle Methods: An introduction with applications*, INRIA, 2009, http://hal.inria.fr/inria-00403917/en/, Technical reportCAJP.

[18] P. DEL MORAL, A. DOUCET, A. JASRA. *An adaptive sequential Monte Carlo method for approximate bayesian computation*, INRIA, 2009, http://people.cs.ubc.ca/~arnaud/delmoral_doucet_jasra_smcabc.pdf, Technical report CA JP .

[19] P. DEL MORAL, A. DOUCET, S. SINGH. *A Backward Particle Interpretation of Feynman-Kac Formulae*, INRIA, 2009, http://hal.inria.fr/inria-00410165/en/, Technical reportCAJPGB.

[20] P. DEL MORAL, A. DOUCET, S. SINGH. *Forward Smoothing Using Sequential Monte Carlo*, Engineering Department, Cambridge University, 2009, http://www-sigproc.eng.cam.ac.uk/~sss40/newerwork.htm, Technical report CA JP GB .

[21] P. DEL MORAL, F. PATRAS, S. RUBENTHALER. *Convergence of U-statistics for interacting particle systems*, INRIA, 2009, http://hal.inria.fr/inria-00397366/en/, Technical report.

[22] P. DEL MORAL, E. RIO. *Concentration Inequalities for Mean Field Particle Models*, INRIA, 2009, http://hal.inria.fr/inria-00375134/fr/, Technical report.

[23] C. HOLMES, F. CARON, J. GRIFFIN, D. STEPHENS. *Two-sample Bayesian nonparametric hypothesis testing*, Oxford Man Institute, 2009, http://arxiv.org/abs/0910.5060, Technical report GB CA .

[24] M. LEI, P. DEL MORAL, C. BAEHR. *Analysis of Approximated PCRLBs for Nonlinear Dynamics Using Different Moments of State Estimate*, INRIA, 2009, http://www.math.univ-toulouse.fr/~baehr/papers/error_of_pcrlb.pdf, Technical report.

[25] M. LEI, P. DEL MORAL, C. BAEHR. *Fisher Information Matrix-based Nonlinear System Conversion for State Estimation*, INRIA, 2009, http://www.math.univ-toulouse.fr/~baehr/papers/pcrlb_trans.pdf, Technical report.

References in notes

[26] A. DOUCET, N. DE FREITAS, N. GORDON (editors). *Sequential Monte Carlo Methods in practice*, Springer-Verlag, 2001.

[27] N. BARTOLI, P. D. MORAL. *Simulation and Stochastic Algorithms (An introduction with applications)*, Cépaduès Editions, 2001.

[28] A. JASRA, C. ANDRIEU, A. DOUCET, P. D. MORAL. *Non-Linear Markov chain Monte Carlo via Self Interacting Approximations*, 2007, Univ. Bristol Preprints.

[29] A. JASRA, C. ANDRIEU, A. DOUCET, P. D. MORAL. *A Note on Convergence of the Equi-energy Sampler*, in "Journal Stochastic Analysis and its Applications", 2008.

[30] J. R. KOZA. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, MIT Press, Cambridge, MA, USA, 1992.

[31] W. B. LANGDON, R. POLI. *Foundations of Genetic Programming*, Springer-Verlag, New York, New York, 2002.

[32] S. MALLAT. *A wavelet tour of signal processing*, 2nd, Elsevier, San Diego, CA, 1999.

[33] P. D. MORAL, L. MICLO. *Branching and Interacting Particle Systems Approximations of Feynman-Kac Formulae with Applications to Non-Linear Filtering*, in "Séminaire de Probabilités XXXIV", J. AZÉMA, M. EMERY, M. LEDOUX, M. YOR (editors), Lecture Notes in Mathematics, vol. 1729, Springer-Verlag Berlin, 2000, p. 1-145.

[34] P. D. MORAL, L. MICLO. *Genealogies and Increasing Propagations of Chaos for Feynman-Kac and Genetic Models*, in "Annals of Applied Probability", vol. 11, n° 4, 2001, p. 1166-1198.

[35] P. D. MORAL, L. MICLO. *On convergence of chains with time empirical self-interactions*, in "Proc. Royal Soc. Lond. A.", vol. 460, 2003, p. 325–346.

[36] P. D. MORAL, L. MICLO. *Self Interacting Markov chains*, in "Stochastic Analysis and Applications", vol. 24, n° 3, 2006, p. 615–660.

[37] P. D. MORAL. *Feynman-Kac formulae. Genealogical and interacting particle systems, with applications*, Springer Verlag New York, Series: Probability and its Applications, 2004.

[38] R. POLI, W. B. LANGDON, N. F. MCPHEE. *A field guide to genetic programming*, 2008, http://www.gp-field-guide.org.uk, (With contributions by J. R. Koza).

[39] L. TRUJILLO, G. OLAGUE, P. LEGRAND, E. LUTTON. *Regularity based descriptor computed from local image oscillations*, in "Optics Express", vol. 15, 2007, p. 6140-6145.