



IN PARTNERSHIP WITH:
CNRS

Université de Bordeaux

Activity Report 2016

Project-Team CQFD

Quality control and dynamic reliability

IN COLLABORATION WITH: Institut de Mathématiques de Bordeaux (IMB)

RESEARCH CENTER
Bordeaux - Sud-Ouest

THEME
Stochastic approaches

Table of contents

| | |
|--|-----------|
| 1. Members | 1 |
| 2. Overall Objectives | 2 |
| 3. Research Program | 3 |
| 3.1. Introduction | 3 |
| 3.2. Main research topics | 3 |
| 4. Application Domains | 5 |
| 5. Highlights of the Year | 6 |
| 6. New Software and Platforms | 6 |
| 6.1. Package divclust | 6 |
| 6.2. Package ClustGeo | 6 |
| 6.3. PCAmixdata | 6 |
| 6.4. QuantifQuantile | 7 |
| 6.5. biips | 7 |
| 6.6. VCN: Software for analysis of VCN | 7 |
| 6.7. EMGView: Software for visualisation and time-frequency analysis of bio signals | 7 |
| 7. New Results | 7 |
| 7.1. Computable approximations for continuous-time Markov decision processes on Borel spaces based on empirical measures | 7 |
| 7.2. Decentralized Proportional Load Balancing | 8 |
| 7.3. Constrained and Unconstrained Optimal Discounted Control of Piecewise Deterministic Markov Processes | 8 |
| 7.4. Approximate Kalman-Bucy filter for continuous-time semi-Markov jump linear systems | 8 |
| 7.5. Investigation of asymmetry in E. coli growth rate | 9 |
| 7.6. Impulsive Control for Continuous-Time Markov Decision Processes: A Linear Programming Approach | 9 |
| 7.7. Conditions for the Solvability of the Linear Programming Formulation for Constrained Discounted Markov Decision Processes | 9 |
| 7.8. Spatio-temporal averaging for a class of hybrid systems and application to conductance-based neuron models | 9 |
| 7.9. A comparison of fitness-case sampling methods for genetic programming | 10 |
| 7.10. Prediction of Expected Performance for a Genetic Programming Classifier | 10 |
| 7.11. Evolving Genetic Programming Classifiers with Novelty Search | 10 |
| 7.12. Regularity and Matching Pursuit Feature Extraction for the Detection of Epileptic Seizures | 11 |
| 7.13. Probabilistic safety analysis of the collision between a space debris and a satellite with an island particle algorithm | 11 |
| 7.14. Particle association measures and multiple target tracking | 11 |
| 7.15. Exponential mixing properties for time inhomogeneous diffusion processes with killing | 12 |
| 7.16. On particle Gibbs Markov chain Monte Carlo models | 12 |
| 7.17. Sequential Monte Carlo with Highly Informative Observations | 12 |
| 7.18. A duality formula for Feynman-Kac path particle models | 13 |
| 7.19. Non-Asymptotic Analysis of Adaptive and Annealed Feynman-Kac Particle Models | 13 |
| 7.20. Uniform stability of a particle approximation of the optimal filter derivative | 14 |
| 7.21. Combining clustering of variables and feature selection using random forests: the CoV/VSURF procedure | 14 |
| 7.22. An innovative approach combining animal performances, nutritional value and sensory quality of meat | 14 |
| 7.23. BIG-SIR: a Sliced Inverse Regression approach for massive data | 15 |
| 8. Bilateral Contracts and Grants with Industry | 15 |
| 8.1.1. Airbus | 15 |

| | |
|--|-----------|
| 8.1.2. Thales Optronique | 15 |
| 8.1.3. DCNS | 16 |
| 9. Partnerships and Cooperations | 16 |
| 9.1. Regional Initiatives | 16 |
| 9.1.1. MATCHABLE project | 16 |
| 9.1.2. project LabEx CPU TIMIC | 16 |
| 9.2. National Initiatives | 17 |
| 9.2.1. ANR Piece | 17 |
| 9.2.2. ANR BNPSI “Bayesian Non Parametric methods for Signal and Image Processing” | 17 |
| 9.3. European Initiatives | 17 |
| 9.3.1. FP7 & H2020 Projects | 17 |
| 9.3.2. Collaborations in European Programs, Except FP7 & H2020 | 18 |
| 9.4. International Initiatives | 18 |
| 9.4.1. Inria Associate Teams Not Involved in an Inria International Labs | 18 |
| 9.4.2. Inria International Partners | 19 |
| 9.5. International Research Visitors | 19 |
| 10. Dissemination | 19 |
| 10.1. Promoting Scientific Activities | 19 |
| 10.1.1. Scientific Events Selection | 19 |
| 10.1.2. Journal | 20 |
| 10.1.2.1. Member of the Editorial Boards | 20 |
| 10.1.2.2. Reviewer - Reviewing Activities | 20 |
| 10.1.3. Invited Talks | 20 |
| 10.1.4. Research Administration | 21 |
| 10.2. Teaching - Supervision - Juries | 21 |
| 10.2.1. Teaching | 21 |
| 10.2.2. Supervision | 22 |
| 10.2.3. Juries | 22 |
| 11. Bibliography | 22 |

Project-Team CQFD

Creation of the Project-Team: 2009 January 01

Keywords:

Computer Science and Digital Science:

- 1.1.6. - Cloud
- 1.2.4. - QoS, performance evaluation
- 1.3. - Distributed Systems
- 3.3. - Data and knowledge analysis
- 3.4.1. - Supervised learning
- 3.4.2. - Unsupervised learning
- 3.4.5. - Bayesian methods
- 3.4.6. - Neural networks
- 3.4.7. - Kernel methods
- 5.9.2. - Estimation, modeling
- 5.9.6. - Optimization tools
- 6.1.2. - Stochastic Modeling (SPDE, SDE)
- 6.1.3. - Discrete Modeling (multi-agent, people centered)
- 6.2.2. - Numerical probability
- 6.2.3. - Probabilistic methods
- 6.2.4. - Statistical methods
- 6.2.6. - Optimization
- 6.4.2. - Stochastic control
- 7.14. - Game Theory
- 8.2. - Machine learning
- 8.6. - Decision support

Other Research Topics and Application Domains:

- 2.2.4. - Infectious diseases, Virology
- 2.6.1. - Brain imaging
- 5.9. - Industrial maintenance
- 6.2. - Network technologies
- 6.3.3. - Network Management
- 6.5. - Information systems
- 9.2.3. - Video games
- 9.4.2. - Mathematics

1. Members

Research Scientists

Jonatha Anselmi [Inria, Researcher]
Pierre Del Moral [Inria, Senior Researcher, HDR]

Faculty Members

Francois Dufour [Team leader, INP Bordeaux, Professor, HDR]
Marie Chavent [Univ. Bordeaux, Associate Professor, HDR]
Alexandre Genadot [Univ. Bordeaux, Associate Professor]
Pierrick Legrand [Univ. Bordeaux, Associate Professor]
Jerome Saracco [INP Bordeaux, Professor, HDR]
Huilong Zhang [Univ. Bordeaux, Associate Professor]

PhD Students

Alizée Geeraert [Thales]
Christophe Nivot [Inria, until Aug 2016, granted by AIRBUS DEFENCE AND SPACE]
Chloe Pasin [Univ. Bordeaux]
Jessica Sodjo [Inria]
Adrien Todeschini [Inria, until november 2016, granted by EVOLLIS SAS]

Post-Doctoral Fellow

Solveig Badillo [Univ. Bordeaux, until Feb 2016]

Visiting Scientists

Oswaldo Luiz Do Valle Costa [Polytechnic School of the University of São Paulo]
Victor Raul Lopez Lopez [ITT (Instituto tecnologico de Tijuana), from Sep 2016 until Oct 2016]
Alexei Piunovskiy [University of Liverpool, until Jul 2016]

Administrative Assistant

Chrystel Plumejeau [Inria]

Others

Benoite de Saporta [Univ. Montpellier II, Professor, HDR]
Jose Enrique Hernandez Beltran [ITT (Instituto tecnologico de Tijuana), from Apr 2016 until Jul 2016]
Uriel Lopez Islas [Centro de Investigación y Desarrollo de Tecnología Digital (CITEDI), from Apr 2016 until Jul 2016]

2. Overall Objectives

2.1. Presentation

The core component of our scientific agenda focuses on the development of statistical and probabilistic methods for the modeling and the optimization of complex systems. These systems require mathematical representations which are in essence dynamic and stochastic with discrete and/or continuous variables. This increasing complexity poses genuine scientific challenges that can be addressed through complementary approaches and methodologies:

- Modeling: design and analysis of realistic and tractable models for such complex real-life systems and various probabilistic phenomena;
- Estimation: developing theoretical and computational procedures in order to estimate and evaluate the parameters and the performance of the system;
- Optimization: developing theoretical and numerical control tools to optimize the performance of complex systems such as computer systems and communication networks.

3. Research Program

3.1. Introduction

The scientific objectives of the team are to provide mathematical tools for modeling and optimization of complex systems. These systems require mathematical representations which are in essence dynamic, multi-model and stochastic. This increasing complexity poses genuine scientific challenges in the domain of modeling and optimization. More precisely, our research activities are focused on stochastic optimization and (parametric, semi-parametric, multidimensional) statistics which are complementary and interlinked topics. It is essential to develop simultaneously statistical methods for the estimation and control methods for the optimization of the models.

3.2. Main research topics

- Stochastic modeling: Markov chain, Piecewise Deterministic Markov Processes (PDMP), Markov Decision Processes (MDP).

The mathematical representation of complex systems is a preliminary step to our final goal corresponding to the optimization of its performance. For example, in order to optimize the predictive maintenance of a system, it is necessary to choose the adequate model for its representation. The step of modeling is crucial before any estimation or computation of quantities related to its optimization. For this we have to represent all the different regimes of the system and the behavior of the physical variables under each of these regimes. Moreover, we must also select the dynamic variables which have a potential effect on the physical variable and the quantities of interest. The team CQFD works on the theory of Piecewise Deterministic Markov Processes (PDMP's) and on Markov Decision Processes (MDP's). These two classes of systems form general families of controlled stochastic processes suitable for the modeling of sequential decision-making problems in the continuous-time (PDMPs) and discrete-time (MDP's) context. They appear in many fields such as engineering, computer science, economics, operations research and constitute powerful class of processes for the modeling of complex system.

- Estimation methods: estimation for PDMP; estimation in non- and semi parametric regression modeling.

To the best of our knowledge, there does not exist any general theory for the problems of estimating parameters of PDMPs although there already exist a large number of tools for sub-classes of PDMPs such as point processes and marked point processes. However, to fill the gap between these specific models and the general class of PDMPs, new theoretical and mathematical developments will be on the agenda of the whole team. In the framework of non-parametric regression or quantile regression, we focus on kernel estimators or kernel local linear estimators for complete data or censored data. New strategies for estimating semi-parametric models via recursive estimation procedures have also received an increasing interest recently. The advantage of the recursive estimation approach is to take into account the successive arrivals of the information and to refine, step after step, the implemented estimation algorithms. These recursive methods do require restarting calculation of parameter estimation from scratch when new data are added to the base. The idea is to use only the previous estimations and the new data to refresh the estimation. The gain in time could be very interesting and there are many applications of such approaches.

- Dimension reduction: dimension-reduction via SIR and related methods, dimension-reduction via multidimensional and classification methods.

Most of the dimension reduction approaches seek for lower dimensional subspaces minimizing the loss of some statistical information. This can be achieved in modeling framework or in exploratory data analysis context.

In modeling framework we focus our attention on semi-parametric models in order to conjugate the advantages of parametric and nonparametric modeling. On the one hand, the parametric part of the

model allows a suitable interpretation for the user. On the other hand, the functional part of the model offers a lot of flexibility. In this project, we are especially interested in the semi-parametric regression model $Y = f(X'\theta) + \varepsilon$, the unknown parameter θ belongs to \mathbb{R}^p for a single index model, or is such that $\theta = [\theta_1, \dots, \theta_d]$ (where each θ_k belongs to \mathbb{R}^p and $d \leq p$ for a multiple indices model), the noise ε is a random error with unknown distribution, and the link function f is an unknown real valued function. Another way to see this model is the following: the variables X and Y are independent given $X'\theta$. In our semi-parametric framework, the main objectives are to estimate the parametric part θ as well as the nonparametric part which can be the link function f , the conditional distribution function of Y given X or the conditional quantile q_α . In order to estimate the dimension reduction parameter θ we focus on the Sliced Inverse Regression (SIR) method which has been introduced by Li [44] and Duan and Li [42].

Methods of dimension reduction are also important tools in the field of data analysis, data mining and machine learning. They provide a way to understand and visualize the structure of complex data sets. Traditional methods among others are principal component analysis for quantitative variables or multiple component analysis for qualitative variables. New techniques have also been proposed to address these challenging tasks involving many irrelevant and redundant variables and often comparably few observation units. In this context, we focus on the problem of synthetic variables construction, whose goals include increasing the predictor performance and building more compact variables subsets. Clustering of variables is used for feature construction. The idea is to replace a group of "similar" variables by a cluster centroid, which becomes a feature. The most popular algorithms include K-means and hierarchical clustering. For a review, see, e.g., the textbook of Duda [43].

- Stochastic optimal control: optimal stopping, impulse control, continuous control, linear programming.

The first objective is to focus on the development of computational methods.

- In the continuous-time context, stochastic control theory has from the numerical point of view, been mainly concerned with Stochastic Differential Equations (SDEs in short). From the practical and theoretical point of view, the numerical developments for this class of processes are extensive and largely complete. It capitalizes on the connection between SDEs and second order partial differential equations (PDEs in short) and the fact that the properties of the latter equations are very well understood. It is, however, hard to deny that the development of computational methods for the control of PDMPs has received little attention. One of the main reasons is that the role played by the familiar PDEs in the diffusion models is here played by certain systems of integro-differential equations for which there is not (and cannot be) a unified theory such as for PDEs as emphasized by M.H.A. Davis in his book. To the best knowledge of the team, there is only one attempt to tackle this difficult problem by O.L.V. Costa and M.H.A. Davis. The originality of our project consists in studying this unexplored area. It is very important to stress the fact that these numerical developments will give rise to a lot of theoretical issues such as type of approximations, convergence results, rates of convergence,....
- Theory for MDP's has reached a rather high degree of maturity, although the classical tools such as value iteration, policy iteration and linear programming, and their various extensions, are not applicable in practice. We believe that the theoretical progress of MDP's must be in parallel with the corresponding numerical developments. Therefore, solving MDP's numerically is an awkward and important problem both from the theoretical and practical point of view. In order to meet this challenge, the fields of neural networks, neurodynamic programming and approximate dynamic programming became recently an active area of research. Such methods found their roots in heuristic approaches, but theoretical results for convergence results are mainly obtained in the context of finite MDP's. Hence, an ambitious challenge is to investigate such numerical problems but for models with general state and action spaces. Our motivation is to develop theoretically consistent

computational approaches for approximating optimal value functions and finding optimal policies.

- An effort has been devoted to the development of efficient computational methods in the setting of communication networks. These are complex dynamical systems composed of several interacting nodes that exhibit important congestion phenomena as their level of interaction grows. The dynamics of such systems are affected by the randomness of their underlying events (e.g., arrivals of http requests to a web-server) and are described stochastically in terms of queueing network models. These are mathematical tools that allow one to predict the performance achievable by the system, to optimize the network configuration, to perform capacity-planning studies, etc. These objectives are usually difficult to achieve without a mathematical model because Internet systems are huge in size. However, because of the exponential growth of their state spaces, an exact analysis of queueing network models is generally difficult to obtain. Given this complexity, we have developed analyses in some limiting regime of practical interest (e.g., systems size grows to infinity). This approach is helpful to obtain a simpler mathematical description of the system under investigation, which leads to the direct definition of efficient, though approximate, computational methods and also allows to investigate other aspects such as Nash equilibria.

The second objective of the team is to study some theoretical aspects related to MDPs such as convex analytical methods and singular perturbation. Analysis of various problems arising in MDPs leads to a large variety of interesting mathematical problems.

4. Application Domains

4.1. Dependability and safety

Our abilities in probability and statistics apply naturally to industry in particular in studies of dependability and safety.

An illustrative example which gathers several topics of team is a collaboration started in September 2013 with Airbus Defence & Space. The goal of this project is the optimization of the assembly line of the future European launcher, taking into account several kinds of economical and technical constraints. We have started with a simplified model with five components to be assembled in workshops liable to breakdowns. We have modeled the problem using the Markov Decision Processes (MDP) framework and built a simulator of the process in order to run a simulation-based optimization procedure.

A second example concerns the optimization of the maintenance of a on board system equipped with a HUMS (Health Unit Monitoring Systems) in collaboration with THALES Optronique. The physical system under consideration is modeled by a piecewise deterministic Markov process. In the context of impulse control, we propose a dynamic maintenance policy, adapted to the state of the system and taking into account both random failures and those related to the degradation phenomenon.

However the spectrum of applications of the topics of the team is larger and may concern many other fields. Indeed non parametric and semi-parametric regression methods can be used in biometry, econometrics or engineering for instance. Gene selection from microarray data and text categorization are two typical application domains of dimension reduction among others. We had for instance the opportunity via the scientific program PRIMEQUAL to work on air quality data and to use dimension reduction techniques as principal component analysis (PCA) or positive matrix factorization (PMF) for pollution sources identification and quantization.

5. Highlights of the Year

5.1. Highlights of the Year

Publication of the book: *Stochastic Processes. From Applications to Theory* written by P. Del Moral and S. Penev, CRC Press, 1290 pages, Jan 2017.

Pierre del Moral has been invited to the IMS World Congress in Toronto to give a Medallion lectures in May 2016.

6. New Software and Platforms

6.1. Package divclust

FUNCTIONAL DESCRIPTION

DIVCLUS-T is a divisive hierarchical clustering algorithm based on a monothetic bipartitional approach allowing the dendrogram of the hierarchy to be read as a decision tree. It is designed for numerical, categorical (ordered or not) or mixed data. Like the Ward agglomerative hierarchical clustering algorithm and the k-means partitioning algorithm, it is based on the minimization of the inertia criterion. However, it provides a simple and natural monothetic interpretation of the clusters. Indeed, each cluster is described by set of binary questions. The inertia criterion is calculated on all the principal components of PCAmix (and then on standardized data in the numerical case).

- Participants: Marie Chavent, Marc Fuentes
- Contact: Marie Chavent
- URL: <https://github.com/chavent/divclust>

6.2. Package ClustGeo

FUNCTIONAL DESCRIPTION

This R package is dedicated to the clustering of objects with geographical positions. The clustering method implemented in this package allows the geographical constraints of proximity to be taken into account within the ascendant hierarchical clustering.

- Marie Chavent, Amaury Labenne, Vanessa Kuentz, Jérôme Saracco
- Contact: Amaury Labenne
- URL: <https://cran.r-project.org/web/packages/ClustGeo/index.html>

6.3. PCAmixdata

FUNCTIONAL DESCRIPTION

Mixed data type arise when observations are described by a mixture of numerical and categorical variables. The R package PCAmixdata extends standard multivariate analysis methods to incorporate this type of data. The key techniques included in the package are PCAmix (PCA of a mixture of numerical and categorical variables), PCARot (rotation in PCAmix) and MFAmix (multiple factor analysis with mixed data within a dataset). The MFAmix procedure handles a mixture of numerical and categorical variables within a group - something which was not possible in the standard MFA procedure. We also included techniques to project new observations onto the principal components of the three methods in the new version of the package.

- Participants: Marie Chavent, Amaury Labenne, Jérôme Saracco
- Contact: Marie Chavent
- URL: <https://cran.r-project.org/web/packages/PCAmixdata/index.html>

6.4. QuantifQuantile

FUNCTIONAL DESCRIPTION

QuantifQuantile is an R package that allows to perform quantization-based quantile regression. The different functions of the package allow the user to construct an optimal grid of N quantizers and to estimate conditional quantiles. This estimation requires a data driven selection of the size N of the grid that is implemented in the functions. Illustration of the selection of N is available, and graphical output of the resulting estimated curves or surfaces (depending on the dimension of the covariate) is directly provided via the plot function.

- Isabelle Charlier, Jérôme Saracco
- Contact: Isabelle Charlier
- URL: <https://cran.r-project.org/web/packages/QuantifQuantile/index.html>

6.5. biips

Bayesian Inference with Interacting Particle Systems

FUNCTIONAL DESCRIPTION

Biips is a software platform for automatic Bayesian inference with interacting particle systems. Biips allows users to define their statistical model in the probabilistic programming BUGS language, as well as to add custom functions or samplers within this language. Then it runs sequential Monte Carlo based algorithms (particle filters, particle independent Metropolis-Hastings, particle marginal Metropolis-Hastings) in a black-box manner so that to approximate the posterior distribution of interest as well as the marginal likelihood. The software is developed in C++ with interfaces with the softwares R, Matlab and Octave.

- Participants: Francois Caron and Adrien Todeschini
- Contact: Adrien Todeschini
- URL: <http://biips.gforge.inria.fr>

6.6. VCN: Software for analysis of VCN

FUNCTIONAL DESCRIPTION

VCN is a software for the analysis of the vigilance of the patient based on the analysis of the EEG signals. The code is written in Matlab and provides an interface easy to use for someone without informatics skills.

- Participants: Pierrick Legrand, Julien Clauzel, Laurent Vezard, Charlotte Rodriguez, Borjan Geshkovski.
- Contact: Pierrick Legrand

6.7. EMGView: Software for visualisation and time-frequency analysis of bio signals

FUNCTIONAL DESCRIPTION

EMGView is a software for the visualisation and the analysis of bio-signals. The code is written in Matlab and provides an interface easy to use for someone without informatics skills.

- Participants: Luis Herrera, Eric Grivel, Pierrick Legrand, Gregory Barriere
- Contact: Pierrick Legrand

7. New Results

7.1. Computable approximations for continuous-time Markov decision processes on Borel spaces based on empirical measures

The following result has been obtained by J. Anselmi (Inria CQFD), F. Dufour (Inria CQFD) and T. Prieto-Rumeau.

We propose an approach for approximating the value function and computing an ε -optimal policy of a continuous-time Markov decision processes with Borel state and action spaces, with possibly unbounded cost and transition rates, under the total expected discounted cost optimality criterion. Under the assumptions that the controlled process satisfies a Lyapunov type condition and the transition rate has a density function with respect to a reference measure, together with piecewise Lipschitz continuity of the elements of the control model, one can approximate the original controlled process by a sequence of models that are computationally solvable. Convergence of the approximations takes place at an exponential rate in probability.

7.2. Decentralized Proportional Load Balancing

The following result has been obtained by J. Anselmi (Inria CQFD) and N. Walton.

Load balancing is a powerful technique commonly used in communication and computer networks to improve system performance, robustness and fairness. In this paper, we consider a general model capturing the performance of communication and computer networks, and on top of it we propose a decentralized algorithm for balancing load among multiple network paths. The proposed algorithm is inspired by the modus operandi of the processor-sharing queue and on each network entry point operates as follows: every time a unit of load completes its service on a path, it increases by one unit the load of that path and decreases by one unit the load of a path selected at random with probability proportional to the amount of load on each of the available paths. We develop a dynamical system to argue that our load-balancer achieves a desirable network-wide utility optimization.

7.3. Constrained and Unconstrained Optimal Discounted Control of Piecewise Deterministic Markov Processes

The following result has been obtained by O. Costa, F. Dufour (Inria CQFD), and A. B. Piunovskiy.

The main goal of this paper is to study the infinite-horizon expected discounted continuous-time optimal control problem of piecewise deterministic Markov processes with the control acting continuously on the jump intensity λ and on the transition measure Q of the process but not on the deterministic flow ϕ . The contributions of the paper are for the unconstrained as well as the constrained cases. The set of admissible control strategies is assumed to be formed by policies, possibly randomized and depending on the history of the process, taking values in a set valued action space. For the unconstrained case we provide sufficient conditions based on the three local characteristics of the process ϕ , λ , Q and the semicontinuity properties of the set valued action space, to guarantee the existence and uniqueness of the integro-differential optimality equation (the so-called Bellman–Hamilton–Jacobi equation) as well as the existence of an optimal (and δ -optimal, as well) deterministic stationary control strategy for the problem. For the constrained case we show that the values of the constrained control problem and an associated infinite dimensional linear programming (LP) problem are the same, and moreover we provide sufficient conditions for the solvability of the LP problem as well as for the existence of an optimal feasible randomized stationary control strategy for the constrained problem.

7.4. Approximate Kalman-Bucy filter for continuous-time semi-Markov jump linear systems

The following result has been obtained by B. Saporta and E. F. Costa.

The aim of this paper is to propose a new numerical approximation of the Kalman-Bucy filter for semi-Markov jump linear systems. This approximation is based on the selection of typical trajectories of the driving semi-Markov chain of the process by using an optimal quantization technique. The main advantage of this approach is that it makes pre-computations possible. We derive a Lipschitz property for the solution of the Riccati equation and a general result on the convergence of perturbed solutions of semi-Markov switching Riccati equations when the perturbation comes from the driving semi-Markov chain. Based on these results, we prove the convergence of our approximation scheme in a general infinite countable state space framework and derive

an error bound in terms of the quantization error and time discretization step. We employ the proposed filter in a magnetic levitation example with Markovian failures and compare its performance with both the Kalman-Bucy filter and the Markovian linear minimum mean squares estimator.

7.5. Investigation of asymmetry in E. coli growth rate

The following result has been obtained by B. Saporta in collaboration with B. Delyon, N. Krell and Lydia Robert.

The data we analyze derives from the observation of numerous cells of the bacterium *Escherichia coli* (*E. coli*) growing and dividing. Single cells grow and divide to give birth to two daughter cells, that in turn grow and divide. Thus, a colony of cells from a single ancestor is structured as a binary genealogical tree. At each node the measured data is the growth rate of the bacterium. In this paper, we study two different data sets. One set corresponds to small complete trees, whereas the other one corresponds to long specific sub-trees. Our aim is to compare both sets. This paper is accessible to post graduate students and readers with advanced knowledge in statistics.

7.6. Impulsive Control for Continuous-Time Markov Decision Processes: A Linear Programming Approach

The following result has been obtained by F. Dufour (Inria CQFD) and A. B. Piunovskiy.

In this paper, we investigate an optimization problem for continuous-time Markov decision processes with both impulsive and continuous controls. We consider the so-called constrained problem where the objective of the controller is to minimize a total expected discounted optimality criterion associated with a cost rate function while keeping other performance criteria of the same form, but associated with different cost rate functions, below some given bounds. Our model allows multiple impulses at the same time moment. The main objective of this work is to study the associated linear program defined on a space of measures including the occupation measures of the controlled process and to provide sufficient conditions to ensure the existence of an optimal control.

7.7. Conditions for the Solvability of the Linear Programming Formulation for Constrained Discounted Markov Decision Processes

The following result has been obtained by F. Dufour (Inria CQFD) and T. Prieto-Rumeau.

We consider a discrete-time constrained discounted Markov decision process (MDP) with Borel state and action spaces, compact action sets, and lower semi-continuous cost functions. We introduce a set of hypotheses related to a positive weight function which allow us to consider cost functions that might not be bounded below by a constant, and which imply the solvability of the linear programming formulation of the constrained MDP. In particular, we establish the existence of a constrained optimal stationary policy. Our results are illustrated with an application to a fishery management problem.

7.8. Spatio-temporal averaging for a class of hybrid systems and application to conductance-based neuron models

The following result has been obtained by A. Genadot (Inria CQFD).

We obtain a limit theorem endowed with quantitative estimates for a general class of infinite dimensional hybrid processes with intrinsically two different time scales and including a population. As an application, we consider a large class of conductance-based neuron models describing the nerve impulse propagation along a neural cell at the scales of ion channels.

7.9. A comparison of fitness-case sampling methods for genetic programming

The following result has been obtained by Pierrick Legrand (Inria CQFD) in collaboration with Y. Martinez, E. Naredo, L. Trujillo, U. Lopez.

The canonical approach towards fitness evaluation in Genetic Programming (GP) is to use a static training set to determine fitness, based on a cost function averaged over all fitness-cases. However, motivated by different goals, researchers have recently proposed several techniques that focus selective pressure on a subset of fitness-cases at each generation. These approaches can be described as fitness-case sampling techniques, where the training set is sampled, in some way, to determine fitness. This paper shows a comprehensive evaluation of some of the most recent sampling methods, using benchmark and real-world problems for symbolic regression. The algorithms considered here are Interleaved Sampling, Random Interleaved Sampling, Lexicase Selection and a new sampling technique is proposed called Keep-Worst Interleaved Sampling (KW-IS). The algorithms are extensively evaluated based on test performance, overfitting and bloat. Results suggest that sampling techniques can improve performance compared with standard GP. While on synthetic benchmarks the difference is slight or none at all, on real-world problems the differences are substantial. Some of the best results were achieved by Lexicase Selection and KeepWorse-Interleaved Sampling. Results also show that on real-world problems overfitting correlates strongly with bloating. Furthermore, the sampling techniques provide efficiency, since they reduce the number of fitness-case evaluations required over an entire run.

7.10. Prediction of Expected Performance for a Genetic Programming Classifier

The following result has been obtained by Pierrick Legrand (Inria CQFD) in collaboration with Y. Martínez, L. Trujillo and E. Galván-López.

The estimation of problem difficulty is an open issue in genetic programming (GP). The goal of this work is to generate models that predict the expected performance of a GP-based classifier when it is applied to an unseen task. Classification problems are described using domain-specific features, some of which are proposed in this work, and these features are given as input to the predictive models. These models are referred to as predictors of expected performance. We extend this approach by using an ensemble of specialized predictors (SPEP), dividing classification problems into groups and choosing the corresponding SPEP. The proposed predictors are trained using 2D synthetic classification problems with balanced datasets. The models are then used to predict the performance of the GP classifier on unseen real-world datasets that are multidimensional and imbalanced. This work is the first to provide a performance prediction of a GP system on test data, while previous works focused on predicting training performance. Accurate predictive models are generated by posing a symbolic regression task and solving it with GP. These results are achieved by using highly descriptive features and including a dimensionality reduction stage that simplifies the learning and testing process. The proposed approach could be extended to other classification algorithms and used as the basis of an expert system for algorithm selection.

7.11. Evolving Genetic Programming Classifiers with Novelty Search

The following result has been obtained by Pierrick Legrand (Inria CQFD) in collaboration with E. Naredo, L. Trujillo, S. Silvac and Luis Muñoz.

Novelty Search (NS) is a unique approach towards search and optimization, where an explicit objective function is replaced by a measure of solution novelty. However, NS has been mostly used in evolutionary robotics while its usefulness in classic machine learning problems has not been explored. This work presents a NS-based genetic programming (GP) algorithm for supervised classification. Results show that NS can solve real-world classification tasks, the algorithm is validated on real-world benchmarks for binary and multiclass problems. These results are made possible by using a domain-specific behavior descriptor. Moreover, two new versions of the NS algorithm are proposed, Probabilistic NS (PNS) and a variant of Minimal Criteria NS (MCNS). The former models the behavior of each solution as a random vector and eliminates all of the

original NS parameters while reducing the computational overhead of the NS algorithm. The latter uses a standard objective function to constrain and bias the search towards high performance solutions. The paper also discusses the effects of NS on GP search dynamics and code growth. Results show that NS can be used as a realistic alternative for supervised classification, and specifically for binary problems the NS algorithm exhibits an implicit bloat control ability.

7.12. Regularity and Matching Pursuit Feature Extraction for the Detection of Epileptic Seizures

The following result has been obtained by Pierrick Legrand (Inria CQFD) in collaboration with E. Z-Flores, L. Trujillo, A. Sotelo and L. N. Coriaa.

The neurological disorder known as epilepsy is characterized by involuntary recurrent seizures that diminish a patient's quality of life. Automatic seizure detection can help improve a patient's interaction with her/his environment, and while many approaches have been proposed the problem is still not trivially solved.

In this work, we present a novel methodology for feature extraction on EEG signals that allows us to perform a highly accurate classification of epileptic states. Specifically, Hölderian regularity and the Matching Pursuit algorithm are used as the main feature extraction techniques, and are combined with basic statistical features to construct the final feature sets. These sets are then delivered to a Random Forests classification algorithm to differentiate between epileptic and non-epileptic readings.

Several versions of the basic problem are tested and statistically validated producing perfect accuracy in most problems and 97.6% accuracy on the most difficult case. Comparison with existing methods: A comparison with recent literature, using a well known database, reveals that our proposal achieves state-of-the-art performance. The experimental results show that epileptic states can be accurately detected by combining features extracted through regularity analysis, the Matching Pursuit algorithm and simple time-domain statistical analysis. Therefore, the proposed method should be considered as a promising approach for automatic EEG analysis.

7.13. Probabilistic safety analysis of the collision between a space debris and a satellite with an island particle algorithm

The following result has been obtained by P. Del Moral (Inria CQFD) in collaboration with C. Verge, J. Morio and J.C Dolado Perez.

Collision between satellites and space debris seldom happens, but the loss of a satellite by collision may have catastrophic consequences both for the satellite mission and for the space environment. To support the decision to trigger off a collision avoidance manoeuvre, an adapted tool is the determination of the collision probability between debris and satellite. This probability estimation can be performed with rare event simulation techniques when Monte Carlo techniques are not enough accurate. In this chapter, we focus on analyzing the influence of different simulation parameters (such as the drag coefficient) that are set for to simplify the simulation, on the collision probability estimation. A bad estimation of these simulation parameters can strongly modify rare event probability estimations. We design here a new island particle Markov chain Monte Carlo algorithm to determine the parameters that, in case of bad estimation, tend to increase the collision probability value. This algorithm also gives an estimate of the collision probability maximum taking into account the likelihood of the parameters. The principles of this statistical technique are described throughout this chapter.

7.14. Particle association measures and multiple target tracking

The following result has been obtained by P. Del Moral (Inria CQFD) in collaboration with J. Houssineau.

In the last decade, the area of multiple target tracking has witnessed the introduction of important concepts and methods, aiming at establishing principled approaches for dealing with the estimation of multiple objects in an efficient way. One of the most successful classes of multi-object filters that have been derived out of these new grounds includes all the variants of the Probability Hypothesis Density (phd) filter. In spite of the attention that these methods have attracted, their theoretical performances are still not fully understood. In this chapter, we first focus on the different ways of establishing the equations of the phd filter, using a consistent set of notations. The objective is then to introduce the idea of observation path, upon which association measures are defined. We will see how these concepts highlight the structure of the first moment of the multi-object distributions in time, and how they allow for devising solutions to practical estimation problems.

7.15. Exponential mixing properties for time inhomogeneous diffusion processes with killing

The following result has been obtained by P. Del Moral (Inria CQFD) in collaboration with D. Villemonais.

We consider an elliptic and time-inhomogeneous diffusion process with time-periodic coefficients evolving in a bounded domain of \mathbb{R}^d with a smooth boundary. The process is killed when it hits the boundary of the domain (hard killing) or after an exponential time (soft killing) associated with some bounded rate function. The branching particle interpretation of the non absorbed diffusion again behaves as a set of interacting particles evolving in an absorbing medium. Between absorption times, the particles evolve independently one from each other according to the diffusion semigroup; when a particle is absorbed, another selected particle splits into two offsprings. This article is concerned with the stability properties of these non absorbed processes. Under some classical ellipticity properties on the diffusion process and some mild regularity properties of the hard obstacle boundaries, we prove a uniform exponential strong mixing property of the process conditioned to not be killed. We also provide uniform estimates w.r.t. the time horizon for the interacting particle interpretation of these non-absorbed processes, yielding what seems to be the first result of this type for this class of diffusion processes evolving in soft and hard obstacles, both in homogeneous and non-homogeneous time settings.

7.16. On particle Gibbs Markov chain Monte Carlo models

The following result has been obtained by P. Del Moral (Inria CQFD) in collaboration with R. Kohn and F. Patras.

This result analyses a new class of advanced particle Markov chain Monte Carlo algorithms recently introduced by Andrieu, Doucet, and Holenstein (2010). We present a natural interpretation of these methods in terms of well known unbiasedness properties of Feynman-Kac particle measures, and a new duality with Feynman-Kac models. This perspective sheds new light on the foundations and the mathematical analysis of this class of methods. A key consequence is their equivalence with the Gibbs sampling of a (many-body) Feynman-Kac target distribution. Our approach also presents a new stochastic differential calculus based on geometric combinatorial techniques to derive non-asymptotic Taylor type series for the semigroup of a class of particle Markov chain Monte Carlo models around their invariant measures with respect to the population size of the auxiliary particle sampler. These results provide sharp quantitative estimates of the convergence rate of the models with respect to the time horizon and the size of the systems. We illustrate the direct implication of these results with sharp estimates of the contraction coefficient and the Lyapunov exponent of the corresponding samplers, and explicit and non-asymptotic L^p -mean error decompositions of the law of the random states around the limiting invariant measure. The abstract framework developed in the article also allows the design of natural extensions to island (also called SMC²) type particle methodologies. We illustrate this general framework and results in the context of nonlinear filtering, hidden Markov chain problems with fixed unknown parameters, and Feynman-Kac path-integration models arising in computational physics and chemistry.

7.17. Sequential Monte Carlo with Highly Informative Observations

The following result has been obtained by P. Del Moral (Inria CQFD) in collaboration with L. Murray.

We propose sequential Monte Carlo (SMC) methods for sampling the posterior distribution of state-space models under highly informative observation regimes, a situation in which standard SMC methods can perform poorly. A special case is simulating bridges between given initial and final values. The basic idea is to introduce a schedule of intermediate weighting and resampling times between observation times, which guide particles towards the final state. This can always be done for continuous-time models, and may be done for discrete-time models under sparse observation regimes; our main focus is on continuous-time diffusion processes. The methods are broadly applicable in that they support multivariate models with partial observation, do not require simulation of the backward transition (which is often unavailable), and, where possible, avoid pointwise evaluation of the forward transition. When simulating bridges, the last cannot be avoided entirely without concessions, and we suggest an epsilon-ball approach (reminiscent of Approximate Bayesian Computation) as a workaround. Compared to the bootstrap particle filter, the new methods deliver substantially reduced mean squared error in normalising constant estimates, even after accounting for execution time. The methods are demonstrated for state estimation with two toy examples, and for parameter estimation (within a particle marginal Metropolis–Hastings sampler) with three applied examples in econometrics, epidemiology and marine biogeochemistry.

7.18. A duality formula for Feynman-Kac path particle models

The following result has been obtained by P. Del Moral (Inria CQFD) in collaboration with R. Kohn and F. Patras.

This result presents a new duality formula between genetic type genealogical tree based particle models and Feynman–Kac measures on path spaces. Among others, this formula allows us to design reversible Gibbs–Glauber Markov chains for Feynman–Kac integration on path spaces. Our approach yields new Taylor series expansions of the particle Gibbs–Glauber semigroup around its equilibrium measure w.r.t. the size of the particle system, generalizing the recent work of Andrieu, Doucet, and Holenstein [1]. We analyze the rate of convergence to equilibrium in terms of the ratio of the length of the trajectories to the number of particles. The analysis relies on a tree-based functional and combinatorial representation of a class of Feynman–Kac particle models with a frozen ancestral line. We illustrate the impact of these results in the context of Quantum and Diffusion Monte Carlo methods.

7.19. Non-Asymptotic Analysis of Adaptive and Annealed Feynman-Kac Particle Models

The following result has been obtained by P. Del Moral (Inria CQFD) in collaboration with F. Giraud.

Sequential and quantum Monte Carlo methods, as well as genetic type search algorithms can be interpreted as a mean field and interacting particle approximations of Feynman-Kac models in distribution spaces. The performance of these population Monte Carlo algorithms is strongly related to the stability properties of nonlinear Feynman-Kac semigroups. In this paper, we analyze these models in terms of Dobrushin ergodic coefficients of the reference Markov transitions and the oscillations of the potential functions. Sufficient conditions for uniform concentration inequalities w.r.t. time are expressed explicitly in terms of these two quantities. We provide an original perturbation analysis that applies to annealed and adaptive Feynman-Kac models, yielding what seems to be the first results of this kind for these types of models. Special attention is devoted to the particular case of Boltzmann-Gibbs measures' sampling. In this context, we design an explicit way of tuning the number of Markov chain Monte Carlo iterations with temperature schedule. We also design an alternative interacting particle method based on an adaptive strategy to define the temperature increments. The theoretical analysis of the performance of this adaptive model is much more involved as both the potential functions and the reference Markov transitions now depend on the random evolution on the particle model. The nonasymptotic analysis of these complex adaptive models is an open research problem. We initiate this study with the concentration analysis of a simplified adaptive models based on reference Markov transitions that coincide with the limiting quantities, as the number of particles tends to infinity.

7.20. Uniform stability of a particle approximation of the optimal filter derivative

The following result has been obtained by P. Del Moral (Inria CQFD) in collaboration with A. Doucet and S.S. Singh.

Particle methods, also known as Sequential Monte Carlo methods, are a principled set of algorithms used to approximate numerically the optimal filter in nonlinear non-Gaussian state-space models. However, when performing maximum likelihood parameter inference in state-space models, it is also necessary to approximate the derivative of the optimal filter with respect to the parameter of the model. References [G. Poyiadjis, A. Doucet, and S. S. Singh, Particle methods for optimal filter derivative: Application to parameter estimation, in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 5, Philadelphia, 2005, pp. 925–928 and G. Poyiadjis, A. Doucet, and S. S. Singh, Biometrika, 98 (2011), pp. 65–80] present an original particle method to approximate this derivative, and it was shown in numerical examples to be numerically stable in the sense that it did not deteriorate over time. In this paper we theoretically substantiate this claim. \mathbb{L}_p bounds and a central limit theorem for this particle approximation are presented. Under mixing conditions these \mathbb{L}_p bounds and the asymptotic variance are uniformly bounded with respect to the time index.

7.21. Combining clustering of variables and feature selection using random forests: the CoV/VSURF procedure

The following result has been obtained by M. Chavent, and J. Saracco in collaboration with R. Genuer.

High-dimensional data classification is a challenging problem. A standard approach to tackle this problem is to perform variables selection, e.g. using stepwise procedures or LASSO approaches. Another standard way is to perform dimension reduction, e.g. by Principal Component Analysis (PCA) or Partial Least Square (PLS) procedures. The approach proposed in this paper combines both dimension reduction and variables selection. First, a procedure of clustering of variables (CoV) is used to built groups of correlated variables in order to reduce the redundancy of information. This dimension reduction step relies on the R package ClustOfVar which can deal with both numerical and categorical variables. Secondly, the most relevant synthetic variables (which are numerical variables summarizing the groups obtained in the first step) are selected with a procedure of variable selection using random forests (VSURF), implemented in the R package VSURF. Numerical performances of the proposed methodology called CoV/VSURF are compared with direct applications of VSURF or random forests (RF) on the original p variables. Improvements obtained with the CoV/VSURF procedure are illustrated on two simulated mixed datasets (cases $n > p$ and $n \ll p$) and on a real proteomic dataset.

7.22. An innovative approach combining animal performances, nutritional value and sensory quality of meat

The following result has been obtained by M. Chavent and J. Saracco in collaboration with M.P. Ellies.

This work sets out a methodological approach to assess how to simultaneously control together animal performances, nutritional value, sensory quality of meat. Seventy-one young bulls were characterized by 97 variables. Variables of each element were arranged into either 5 homogeneous Intermediate Scores (IS) or 2 Global Indices (GI) via a clustering of variables and analysed together by Principal Component Analysis (PCA). These 3 pools of 5 IS (or 2 GI) were analysed together by PCA to established the links existing among the triptych. Classification on IS showed no opposition between animal performances and nutritional value of meat, as it seemed possible to identify animals with a high butcher value and intramuscular fat relatively rich in polyunsaturated fatty acids. Concerning GI, the classification indicated that animal performances were negatively correlated with sensory quality. This method appeared to be a useful contribution to the management of animal breeding for an optimal trade-off between the three elements of the triptych.

7.23. BIG-SIR: a Sliced Inverse Regression approach for massive data

The following result has been obtained by J. Saracco (Inria CQFD) in collaboration with B. Lique.

In a massive data setting, we focus on a semiparametric regression model involving a real dependent variable Y and a p -dimensional covariable X . This model includes a dimension reduction of X via an index $X'\beta$. The Effective Dimension Reduction (EDR) direction β cannot be directly estimated by the Sliced Inverse Regression (SIR) method due to the large volume of the data. To deal with the main challenges of analysing massive datasets which are the storage and computational efficiency, we propose a new SIR estimator of the EDR direction by following the “divide and conquer” strategy. The data is divided into subsets. EDR directions are estimated in each subset which is a small dataset. The recombination step is based on the optimisation of a criterion which assesses the proximity between the EDR directions of each subset. Computations are run in parallel with no communication among them. The consistency of our estimator is established and its asymptotic distribution is given. Extensions to multiple indices models, q -dimensional response variable and/or SIR α -based methods are also discussed. A simulation study using our `edrGraphicalTools` R package shows that our approach enables us to reduce the computation time and conquer the memory constraint problem posed by massive datasets. A combination of `foreach` and `bigmemory` R packages are exploited to offer efficiency of execution in both speed and memory. Finally, results are visualised using the `bin-summarise-smooth` approach through the `bigvis` R package

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. Airbus

Participants: Benoîte de Saporta, François Dufour, Christophe Nivot.

We are interested in the optimization of a launcher integration process. It comprises several steps from the production of the subassemblies to the final launch. The four subassemblies go through various types of operations such as preparation, integration, control and storage. These operations are split up into three workshops. Due to possible breakdowns or staff issues, the time spent in each workshop is supposed random. So is the time needed to deliver the subassemblies, for similar reasons including e.g. shipping delays. We also have to deal with constraints related to the architecture of the assembly process itself. Indeed, we have to take into account waiting policies between workshops. The workshops may work in parallel but can be blocked if their output is not transferred to the next workshop in line. Storage capacity of output products is limited.

Our goal is finding the best rates of delivery of the subassemblies, the best choice of architecture (regarding stock capacities) and the best times when to stop and restart the workshops to be able to carry out twelve launches a year according to a predetermined schedule at minimal cost. To solve this problem, we choose a mathematical model particularly suitable for optimization with randomness: Markov decision processes (MDPs).

We have implemented a numerical simulator of the process based on the MDP model. It provides the fullest information possible on the process at any time. The simulator has first been validated with deterministic histories. Random histories have then been run with exponentially distributed delivery times for the subassemblies and several families of random laws for the time spent in each workshop. Using Monte Carlo simulations, we obtain the distribution of the launch times. Preliminary optimization results allow choosing stock capacities and delivery rates that satisfy the launch schedule.

In this context, the PhD Thesis of Christophe Nivot (2013-2016) is funded by Chaire Inria-Astrium-EADS IW-Conseil régional d'Aquitaine.

8.1.2. Thales Optronique

Participants: Benoîte de Saporta, François Dufour, Alizée Geeraert.

Integrated maintenance, failure intensity, optimisation.

As part of optimizing the reliability, Thales Optronics includes systems that examine the state of their equipment. This function is performed by HUMS (Health Unit Monitoring Systems). The collaboration is the subject of the PhD of Alize Geeraert (CIFRE). The aim of this thesis is to implement in the HUMS a program based on observations that can determine the state of the system, optimize maintenance operations and evaluate the failure risk of a mission.

8.1.3. DCNS

Participants: Huilong Zhang, Jonatha Anselmi, François Dufour, Dann Laneuville.

This contract is with DCNS, a French industrial group specialized in naval defense and energy. In particular, DCNS designs and builds submarines and surface combatants, develops associated systems and infrastructure, and offers a full range of services to naval bases and shipyards, together with a focus into marine renewable energy. The main objective is to have robust algorithms able to build an accurate picture of the objects that are around a submarine by only using “passive sonar” information. This means that no information is transmitted by the submarine, which just listens to acoustic waves coming in, to the target. We estimate the position and the velocity of moving targets through noisy observations and a Kalman-type filter. Estimates become accurate depending on the type and the number of maneuvers done by the submarine. Our goal is to combine the filter that is currently used in DCNS with a Markov decision process. This provides a systematic framework to compute the best sequence of submarine maneuvers that allows the system to determine, as soon as possible, accurate target position and velocity. The current technological transfer to DCNS stands in a stochastic optimization framework developed in Matlab that operates under the hypothesis that the target follows a uniform linear motion with constant velocity or zero acceleration. The case where targets move in a more complex manner gives concrete perspectives for further transfers to DCNS.

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. *MATCHABLE project*

Matchable is a startup incubated at IRA (Incubateur Régional d’Aquitaine) since Mars 2014. This startup predicts how players will behave, who is likely to spend money, who you should target with promotions/product placement, and who the developer has to pay attention to in order to prevent churners. The members of CQFD have supervised two masters internships and a postdoctoral researcher, granted by two PEPS contracts from AMIES.

9.1.2. *project LabEx CPU TIMIC*

The topic of the project is TIMIC is the multivariate treatment of human brain imaging and its application to the analysis of cerebral connectivity graph during rest. The project focuses on the analysis of variability of cerebral organisation on a large population using several methods of supervised and unsupervised classification. The volume of data and the iterative aspect of the methods will lead to implement the classification process on infrastructure of distributed computing.

Alexandre Laurent has been hired as full time research engineer this project for 12 months in 2016.

9.2. National Initiatives

9.2.1. ANR Piece

ANR Piece (2013-2016) of the program *Jeunes chercheuses et jeunes chercheurs* of the French National Agency of Research (ANR), lead by F. Malrieu (Univ. Tours). The Piecewise Deterministic Markov Processes (PDMP) are non-diffusive stochastic processes which naturally appear in many areas of applications as communication networks, neuron activities, biological populations or reliability of complex systems. Their mathematical study has been intensively carried out in the past two decades but many challenging problems remain completely open. This project aims at federating a group of experts with different backgrounds (probability, statistics, analysis, partial derivative equations, modeling) in order to pool everyone's knowledge and create new tools to study PDMPs. The main lines of the project relate to estimation, simulation and asymptotic behaviors (long time, large populations, multi-scale problems) in the various contexts of application.

9.2.2. ANR BNPSI “Bayesian Non Parametric methods for Signal and Image Processing”

Statistical methods have become more and more popular in signal and image processing over the past decades. These methods have been able to tackle various applications such as speech recognition, object tracking, image segmentation or restoration, classification, clustering, etc. We propose here to investigate the use of Bayesian nonparametric methods in statistical signal and image processing. Similarly to Bayesian parametric methods, this set of methods is concerned with the elicitation of prior and computation of posterior distributions, but now on infinite-dimensional parameter spaces. Although these methods have become very popular in statistics and machine learning over the last 15 years, their potential is largely underexploited in signal and image processing. The aim of the overall project, which gathers researchers in applied probabilities, statistics, machine learning and signal and image processing, is to develop a new framework for the statistical signal and image processing communities. Based on results from statistics and machine learning we aim at defining new models, methods and algorithms for statistical signal and image processing. Applications to hyperspectral image analysis, image segmentation, GPS localization, image restoration or space-time tomographic reconstruction will allow various concrete illustrations of the theoretical advances and validation on real data coming from realistic contexts.

9.3. European Initiatives

9.3.1. FP7 & H2020 Projects

IRSES ACOBSEC

Project reference: 612689 Funded under: FP7-PEOPLE

Coordinator : Pierrick Legrand

Participants UNIVERSITE VICTOR SEGALEN BORDEAUX II Participation ended

UNIVERSITE DE BORDEAUX

FUNDACAO DA FACULDADE DE CIENCIAS DA UNIVERSIDADE DE LISBOA Portugal

UNIVERSIDAD DE EXTREMADURA Spain

INESC ID - INSTITUTO DE ENGENHARIA DE SISTEMAS E COMPUTADORES, INVESTIGACAO E DESENVOLVIMENTO EM LISBOA Participation ended

Over the last decade, Human-Computer Interaction (HCI) has grown and matured as a field. Gone are the days when only a mouse and keyboard could be used to interact with a computer. The most ambitious of such interfaces are Brain-Computer Interaction (BCI) systems. BCI's goal is to allow a person to interact with an artificial system using brain activity. A common approach towards BCI is to analyze, categorize and interpret Electroencephalography (EEG) signals in such a way that they alter the state of a computer. ACoBSEC's objective is to study the development of computer systems for the automatic analysis and classification of mental states of vigilance; i.e., a person's state of alertness. Such a task is relevant to diverse domains, where a

person is required to be in a particular state. This problem is not a trivial one. In fact, EEG signals are known to be noisy, irregular and tend to vary from person to person, making the development of general techniques a very difficult scientific endeavor. Our aim is to develop new search and optimization strategies, based on evolutionary computation (EC) and genetic programming (GP) for the automatic induction of efficient and accurate classifiers. EC and GP are search techniques that can reach good solutions in multi-modal, non-differentiable and discontinuous spaces; and such is the case for the problem addressed here. This project combines the expertise of research partners from five converging fields: Classification, Neurosciences, Signal Processing, Evolutionary Computation and Parallel Computing in Europe (France Inria, Portugal INESC-ID, Spain UNEX, Bordeaux university, Sciences University of Lisbon) and South America (Mexico ITT, CICESE). The exchange program goals and milestones give a comprehensive strategy for the strengthening of current scientific relations amongst partners, as well as for the construction of long-lasting scientific relationships that produce high quality theoretical and applied research.

9.3.2. Collaborations in European Programs, Except FP7 & H2020

Program: Direccion General de Investigacion Cientifica y Tecnica, Gobierno de Espana

Project acronym: GAMECONAPX

Project title: Numerical approximations for Markov decision processes and Markov games

Duration: 01/2017 - 12/2019

Coordinator: Tomas Prieto-Rumeau, Department of Statistics and Operations Research, UNED (Spain)

Abstract:

This project is funded by the Gobierno de Espana, Direccion General de Investigacion Cientifica y Tecnica (reference number: MTM2016-75497-P) for three years to support the scientific collaboration between Tomas Prieto-Rumeau, Jonatha Anselmi and François Dufour. This research project is concerned with numerical approximations for Markov decision processes and Markov games. Our goal is to propose techniques allowing to approximate numerically the optimal value function and the optimal strategies of such problems. Although such decision models have been widely studied theoretically and, in general, it is well known how to characterize their optimal value function and their optimal strategies, the explicit calculation of these optimal solutions is not possible except for a few particular cases. This shows the need for numerical procedures to estimate or to approximate the optimal solutions of Markov decision processes and Markov games, so that the decision maker can really have at hand some approximation of his optimal strategies and his optimal value function. This project will explore areas of research that have been, so far, very little investigated. In this sense, we expect our techniques to be a breakthrough in the field of numerical methods for continuous-time Markov decision processes, but particularly in the area of numerical methods for Markov game models. Our techniques herein will cover a wide range of models, including discrete- and continuous-time models, problems with unbounded cost and transition rates, even allowing for discontinuities of these rate functions. Our research results will combine, on one hand, mathematical rigor (with the application of advanced tools from probability and measure theory) and, on the other hand, computational efficiency (providing accurate and ?applicable? numerical methods). In this sense, particular attention will be paid to models of practical interest, including population dynamics, queueing systems, or birth-and-death processes, among others. So, we expect to develop a generic and robust methodology in which, by suitably specifying the data of the decision problem, an algorithm will provide the approximations of the value function and the optimal strategies. Therefore, the results that we intend to obtain in this research project will be of interest for researchers in the fields of Markov decision processes and Markov games, both for the theoretical and the applied or practitioners communities

9.4. International Initiatives

9.4.1. Inria Associate Teams Not Involved in an Inria International Labs

9.4.1.1. CDSS

Title: Control of Dynamic Systems Subject to Stochastic Jumps

International Partner (Institution - Laboratory - Researcher):

Universidade de São Paulo (Brazil) - Departamento de Matemática Aplicada e Estatística (ICMC) - Costa Eduardo

Start year: 2014

See also: <https://team.inria.fr/cdss/fr/>

The main goals of this joint team CDSS is to study the control of dynamic systems subject to stochastic jumps. Three topics will be considered throughout the next 3 years. In the first topic we study the control problem of piecewise-deterministic Markov processes (PDMP's) considering constraints. In this case the main goal is to obtain a theoretical formulation for the equivalence between the original optimal control of PDMP's with constraints and an infinite dimensional static linear optimization problem over a space of occupation measures of the controlled process. F. Dufour (CQFD, Inria) and O. Costa (Escola Politécnica da Universidade de São Paulo, Brazil) mainly carry out this topic. In the second topic we focus on numerical methods for solving control and filtering problems related to Markov jump linear systems (MJLS). This project will allow a first cooperation between B. de Saporta (Univ. Montpellier II) and E. Costa (Universidade de São Paulo, Brazil). The third research subject is focused on quantum control by using Lyapunov-like stochastic methods conducted by P. Rouchon (Ecole des Mines de Paris) and P. Pereira da Silva (Escola Politécnica da Universidade de São Paulo, Brazil).

9.4.2. Inria International Partners

9.4.2.1. Declared Inria International Partners

Tree-Lab, ITT. TREE-LAB is part of the Cybernetics research line within the Engineering Science graduate program offered by the Department of Electric and Electronic Engineering at Tijuana's Institute of Technology (ITT), in Tijuana Mexico. TREE-LAB is mainly focused on scientific and engineering research within the intersection of broad scientific fields, particularly Computer Science, Heuristic Optimization and Pattern Analysis. In particular, specific domains studied at TREE-LAB include Genetic Programming, Classification, Feature Based Recognition, Bio-Medical signal analysis and Behavior-Based Robotics. Currently, TREE-LAB incorporates the collaboration of several top researchers, as well as the participation of graduate (doctoral and masters) and undergraduate students, from ITT. Moreover, TREE-LAB is actively collaborating with top researchers from around the world, including Mexico, France, Spain, Portugal and USA.

9.5. International Research Visitors

9.5.1. Visits of International Scientists

Oswaldo Costa (Escola Politécnica da Universidade de São Paulo, Brazil) collaborate with the team on the theoretical aspects of continuous control of piecewise-deterministic Markov processes. He visited the team during two weeks in 2016 supported by the Associate Team Inria: CDSS.

Alexey Piunovskiy (University of Liverpool) visited the team during 2 weeks in 2016. The main subject of the collaboration is the linear programming approach for Markov Decision Processes. This research was supported by the Clusters d'excellence CPU.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Selection

10.1.1.1. Member of the Conference Program Committees

J. Anselmi has been a member of the TPC of the international conferences: VALUETOOLS-2016, ASMTA-2016 and ECQT-2016.

F. Dufour is a member of the organizing committee of the international SIAM conference on Control & its Application, SIAM CT17.

M. Chavent has been vice-president of the program committee of the 5èmes Rencontres R in Toulouse in 2016.

10.1.2. Journal

10.1.2.1. Member of the Editorial Boards

F. Dufour is associate editor of the journal: SIAM Journal of Control and Optimization since 2009.

J. Saracco is an associate editor of the journal Case Studies in Business, Industry and Government Statistics (CSBIGS) since 2006.

10.1.2.2. Reviewer - Reviewing Activities

All the members of CQFD are regular reviewers for several international journals and conferences in applied probability, statistics and operations research.

10.1.3. Invited Talks

F. Dufour gave the following invited talks:

- *Unconstrained and Constrained Optimal Control of Piecewise Deterministic Markov Processes*, Workshop on switching dynamics & verification, Institut Henri Poincaré, Paris, France, January 28-29, 2016.
- *Stability of piecewise deterministic Markov processes*, Department of Statistics, Oxford University, United Kingdom, October 11, 2016.
- *Numerical Approximations for Average Cost Markov Decision Processes*, Inria Team TAO Seminar, February 9, 2016.

P. del Moral gave the following invited lectures:

- *An introduction to Feynman-Kac integration and genealogical tree based particle models*, Thematic Cycle on Monte-Carlo techniques, Labex Louis Bachelier, Institut Henri Poincaré, January 15, 22, 29, and February 12, 2016.
- *Mean Field Particle Samplers In Statistical Learning and Rare Event Analysis*, CFM-Imperial Distinguished Lecture Series, Imperial College, United Kingdom, October 18, 25, and November 1, 8, 2016.

A. Genadot gave the following talks:

- *Moyennisation à la mode de T. G. Kurtz pour des processus déterministes par morceaux*, Université de Lorraine, Nancy, January 14, 2016.
- *Averaging for some simple constrained Markov process*, Journées MAS, université Grenoble-Alpes, August 29, 30 and 31, 2016.

J. Saracco gave the following talks:

- *Analyse de la variance : une vision de type modèle linéaire gaussien ou comment expliquer une variable quantitative par un ou plusieurs facteurs qualitatifs*, University of Monastir (Tunisia), April 2016
- *Un exemple de régression semiparamétrique : l'approche SIR (sliced inverse regression)*, University of Monastir (Tunisia), April 2016
- *La régression par quantile non-paramétrique et semi-paramétrique*, "Les jeudis de Santé Publique", Paris, November 2016

J. Saracco was an invited professor at University of Monastir (Tunisia) in november 2016 and gave a course on Multidimensional Statistics.

Marie Chavent gave the following invited lecture :

- *Multivariate analysis of mixed data: The PCAmixdata R package*, CMStatistics, Seville, December 2016.

10.1.4. Research Administration

F. Dufour is member of the *Bureau du comité des projets*, Inria Bordeaux Sud-Ouest.

J. Saracco is deputy director of IMB (Institut de Mathématiques de Bordeaux, UMR CNRS 5251) since 2015.

M. Chavent is member of the national evaluation committee of Inria.

M. Chavent is member of the council of the Institut de Mathématique de Bordeaux.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

- Licence : J. Anselmi, Probabilités et statistiques, 13 heures, Institut Polytechnique de Bordeaux, école ENSEIRB-MATMECA, filiere telecom, France.
- Licence : J. Anselmi, Probabilités et statistiques, 13 heures, Institut Polytechnique de Bordeaux, école ENSEIRB-MATMECA, filiere electronique, France.
- Licence: M. Chavent, Analyse des données, 15 ETD, L3, Bordeaux university, France
- License: M. Chavent, Modélisation statistique, 15 ETD, niveau L3, Bordeaux university, France
- Master : M. Chavent, Apprentissage automatique, 50 ETD, niveau M2, Bordeaux university, France
- Licence : F. Dufour, Probabilités et statistiques, 16 heures, niveau L3, Institut Polytechnique de Bordeaux, école ENSEIRB-MATMECA, France.
- Master : F. Dufour, Méthodes numériques pour la fiabilité, 24 heures, niveau M1, Institut Polytechnique de Bordeaux, école ENSEIRB-MATMECA, France.
- Master : F. Dufour, Probabilités, 20 heures, niveau M1, Institut Polytechnique de Bordeaux, école ENSEIRB-MATMECA, France.
- P. Legrand, Algèbre (responsable de l'UE), Licence 1 SCIMS (108 heures)
- P. Legrand, Informatique pour les mathématiques (responsable de l'UE), Licence 1 et Licence 2 (36 heures)
- P. Legrand, Espaces Euclidiens. (responsable de l'UE), Licence 2 SCIMS (54 heures)
- P. Legrand, Formation Matlab pour le personnel CNRS (responsable de l'UE), (24 heures)
- Licence: J. Saracco, Probability and Descriptive statistics, 27h, L3, First year of ENSC - Bordeaux INP, France
- Licence: J. Saracco, Mathematical statistics, 20h, L3, First year of ENSC - Bordeaux INP, France
- Licence: J. Saracco, Data analysis (multidimensional statistics), 20h, L3, First year of ENSC - Bordeaux INP, France
- Master: J. Saracco, Statistical modeling, 27h, M1, Second year of ENSC - Bordeaux INP, France
- Master: J. Saracco, Applied probability and Statistics, 40h, M1, Second year of ENSCBP - Bordeaux INP, France
- Master: J. Saracco, Probability and Statistics, 12h, M2, Science Po Bordeaux, France
- A. Genadot, Probabilités de bases (18h), Licence MIASHS première année, Université de Bordeaux.
- A. Genadot, Statistiques de bases (18h), Licence MIASHS première année, Université de Bordeaux.
- A. Genadot, Probabilités (36h), Licence MIASHS deuxième année, Université de Bordeaux.
- A. Genadot, Processus (18h), Licence MIASHS troisième année, Université de Bordeaux.
- A. Genadot, Modélisation statistique (18h), Licence MIASHS troisième année, Université de Bordeaux.
- A. Genadot, Martingales (25h), Master MIMSE première année, Université de Bordeaux.
- A. Genadot, Probabilités (20h), Master MEEF première année, Université de Bordeaux.

10.2.2. Supervision

- PhD completed: Adrien Todeschini, Elaboration et validation d'un système de recommandation bayésien, supervised by F. Caron and M. Chavent.
- PhD completed: Christophe Nivot, Optimisation de la chaîne de montage du futur lanceur européen, May 2016, B. supervised by B. de Saporta and F. Dufour.
- PhD in progress : Alizé Geeraert, Contrôle optimal des processus Markoviens déterministes par morceaux et application à la maintenance, University of Bordeaux, supervised by B. de Saporta and F. Dufour (defense scheduled in June 2017).
- PhD in progress : Ines Jlassi, Contributions à la régression inverse par tranches et à l'estimation non para métrique des quantiles conditionnels, University of Monastir (Tunisia), September 2013, supervised by J. Saracco and L. Ben Abdelghani Bouraoui.
- PhD in progress : Hadrien Lorenzo, Analyses de données longitudinales de grandes dimensions appliquées aux essais vaccinaux contre le VIH et Ebola, University of Bordeaux, September 2016, supervised by J. Saracco and R. Thiebaut.

10.2.3. Juries

J. Saracco is vice president of the french statistical society (SFdS).

11. Bibliography

Publications of the year

Articles in International Peer-Reviewed Journals

- [1] J. ANSELMi, F. DUFOUR, T. PRIETO-RUMEAU. *Computable approximations for continuous-time Markov decision processes on Borel spaces based on empirical measures*, in "Journal of Mathematical Analysis and Applications", 2016, vol. 443, n^o 2, pp. 1323 - 1361 [DOI : 10.1016/j.jmaa.2016.05.055], <https://hal.archives-ouvertes.fr/hal-01412615>
- [2] J. ANSELMi, N. S. WALTON. *Decentralized Proportional Load Balancing*, in "SIAM Journal on Applied Mathematics", 2016, vol. 76, n^o 1, pp. 391-410 [DOI : 10.1137/140969361], <https://hal.archives-ouvertes.fr/hal-01415856>
- [3] O. COSTA, F. DUFOUR, A. B. PIUNOVSKIY. *Constrained and Unconstrained Optimal Discounted Control of Piecewise Deterministic Markov Processes*, in "SIAM Journal on Control and Optimization", 2016, vol. 54, n^o 3, pp. 1444 - 1474 [DOI : 10.1137/140996380], <https://hal.archives-ouvertes.fr/hal-01412604>
- [4] B. DE SAPORTA, E. COSTA. *Approximate Kalman-Bucy filter for continuous-time semi-Markov jump linear systems*, in "IEEE Transactions on Automatic Control", 2016, vol. 61, n^o 8, pp. 2035 - 2048 [DOI : 10.1109/TAC.2015.2495578], <https://hal.archives-ouvertes.fr/hal-01062618>
- [5] B. DE SAPORTA, F. DUFOUR, A. GEERAERT. *Optimal strategies for impulse control of piecewise deterministic Markov processes*, in "Automatica", 2017 [DOI : 10.1016/j.automatica.2016.11.039], <https://hal.archives-ouvertes.fr/hal-01294286>
- [6] B. DE SAPORTA, F. DUFOUR, C. NIVOT. *Partially observed optimal stopping problem for discrete-time Markov processes*, in "4OR: A Quarterly Journal of Operations Research", 2017 [DOI : 10.1007/s10288-016-0337-8], <https://hal.archives-ouvertes.fr/hal-01274645>

- [7] P. DEL MORAL, R. KOHN, F. PATRAS. *On particle Gibbs samplers*, in "Annales de l'IHP - Probabilités et Statistiques", 2016, <https://hal.archives-ouvertes.fr/hal-01312953>
- [8] B. DELYON, B. DE SAPORTA, N. KRELL, L. ROBERT. *Investigation of asymmetry in E. coli growth rate*, in "CSBIGS (Case Studies in Business, Industry and Government Statistics)", 2016, forthcoming, <https://hal.inria.fr/hal-01201923>
- [9] F. DUFOUR, A. B. PIUNOVSKIY. *Impulsive Control for Continuous-Time Markov Decision Processes: A Linear Programming Approach*, in "Applied Mathematics and Optimization", 2016, vol. 74, n^o 1, pp. 129 - 161 [DOI : 10.1007/s00245-015-9310-8], <https://hal.archives-ouvertes.fr/hal-01246229>
- [10] F. DUFOUR, T. PRIETO-RUMEAU. *Conditions for the Solvability of the Linear Programming Formulation for Constrained Discounted Markov Decision Processes*, in "Applied Mathematics and Optimization", 2016, vol. 74, n^o 1, pp. 27 - 51 [DOI : 10.1007/s00245-015-9307-3], <https://hal.archives-ouvertes.fr/hal-01246228>
- [11] M. ELLIES-OURY, G. CANTALAPIEDRA-HIJAR, D. DURAND, D. GRUFFAT, A. LISTRAT, D. MICOL, I. ORTIGUES-MARTY, J. HOCQUETTE, M. CHAVENT, J. SARACCO, B. PICARD. *An innovative approach combining Animal Performances, nutritional value and sensory quality of meat*, in "Meat Science", 2016, vol. 122, pp. 163 - 172 [DOI : 10.1016/J.MEATSCI.2016.08.004], <https://hal.archives-ouvertes.fr/hal-01417538>
- [12] A. GENADOT. *Spatio-temporal averaging for a class of hybrid systems and application to conductance-based neuron models*, in "Nonlinear Analysis: Hybrid Systems", November 2016, vol. 22, pp. 176-190 [DOI : 10.1016/J.NAHS.2016.03.003], <https://hal.archives-ouvertes.fr/hal-01414107>
- [13] B. LIQUET, J. SARACCO. *BIG-SIR: a Sliced Inverse Regression approach for massive data*, in "Statistics and its interfaces", 2016, vol. 9, n^o 4, pp. 509-520, <https://hal.archives-ouvertes.fr/hal-01417425>
- [14] Y. MARTINEZ, E. NAREDO, L. TRUJILLO, P. LEGRAND, U. LOPEZ. *A comparison of fitness-case sampling methods for genetic programming*, in "Journal of Experimental and Theoretical Artificial Intelligence", 2017, <https://hal.inria.fr/hal-01389047>
- [15] Y. MARTINEZ, L. TRUJILLO, P. LEGRAND, E. GALVAN-LOPEZ. *Prediction of Expected Performance for a Genetic Programming Classifier*, in "Genetic Programming and Evolvable Machines", 2016, vol. 17, n^o 4, pp. 409–449 [DOI : 10.1007/s10710-016-9265-9], <https://hal.inria.fr/hal-01252141>
- [16] E. NAREDO, L. TRUJILLO, P. LEGRAND, S. SILVA, L. MUNOZ. *Evolving Genetic Programming Classifiers with Novelty Search*, in "Information Sciences", November 2016, vol. 369, pp. 347–367 [DOI : 10.1016/J.INS.2016.06.044], <https://hal.inria.fr/hal-01389049>
- [17] E. Z-FLORES, L. TRUJILLO, A. SOTELO, P. LEGRAND, L. CORIA. *Regularity and Matching Pursuit Feature Extraction for the Detection of Epileptic Seizures*, in "Journal of Neuroscience Methods", 2016, <https://hal.inria.fr/hal-01389051>

Invited Conferences

- [18] M. CHAVENT. *Multivariate analysis of mixed data: The PCAmixdata R package*, in "9th International Conference of the ERCIM WG on Computational and Methodological Statistics (CMStatistics 2016)", Seville, Spain, December 2016, <https://hal.archives-ouvertes.fr/hal-01416742>

- [19] A. GEERAERT, B. DE SAPORTA, F. DUFOUR. *Impulse control of piecewise deterministic processes*, in "28th European Conference on Operational Research", Poznan, Poland, 2016, <https://hal.archives-ouvertes.fr/hal-01336314>
- [20] J. SARACCO, I. JLIASSI. *Variable importance assessment in sliced inverse regression for variable selection*, in "9th International Conference of the ERCIM WG on Computational and Methodological Statistics (CM-Statistics 2016)", Séville, Spain, December 2016, <https://hal.archives-ouvertes.fr/hal-01417436>

International Conferences with Proceedings

- [21] E. COSTA, B. DE SAPORTA. *Precomputable Kalman-based filter for Markov Jump Linear Systems*, in "3rd International Conference on Control and Fault-Tolerant Systems", Barcelone, Spain, 2016, pp. 387-392, <https://hal.archives-ouvertes.fr/hal-01336597>
- [22] J. E. HERNANDEZ-BELTRAN, V. H. DÍAZ-RAMÍREZ, L. TRUJILLO, P. LEGRAND. *Restoration of degraded images using genetic programming*, in "Optics and photonics for information processing X", San Diego, United States, August 2016, vol. 9970, <https://hal.inria.fr/hal-01389064>
- [23] V. R. LÓPEZ-LÓPEZ, L. TRUJILLO, P. LEGRAND, G. OLAGUE. *Genetic Programming: From design to improved implementation*, in "Gecco 2016", Denver, United States, June 2016, <https://hal.inria.fr/hal-01389066>

Conferences without Proceedings

- [24] V. SAUTRON, M. CHAVENT, N. VIGUERIE, N. VILLA-VIALANEIX. *Multiway-SIR for longitudinal multi-table data integration*, in "22nd International Conference on Computational Statistics (COMPSTAT)", Oviedo, Spain, August 2016, <https://hal.archives-ouvertes.fr/hal-01416735>

Scientific Books (or Scientific Book chapters)

- [25] S. GIRARD, J. SARACCO. *Supervised and unsupervised classification using mixture models*, in "Statistics for Astrophysics: Clustering and Classification", D. FRAIX-BURNET, S. GIRARD (editors), EAS Publications Series, EDP Sciences, May 2016, vol. 77, pp. 69-90, <https://hal.archives-ouvertes.fr/hal-01417514>
- [26] T. LEONARDO, E. Z-FLORES, P. S. JUAREZ SMITH, P. LEGRAND, S. SILVA, M. CASTELLI, L. VAN- NESCHI, O. SCHÜTZE, L. MUNOZ. *Local Search is Underused in Genetic Programming*, in "Genetic Programming Theory and Practice XIV", A. ARBOR (editor), Springer, 2017, <https://hal.inria.fr/hal-01388426>
- [27] J. SARACCO, M. CHAVENT. *Clustering of Variables for Mixed Data*, in "Statistics for Astrophysics: Clustering and Classification", EAS Publications Series, EDP Sciences, 2016, vol. 77, pp. 91-119, <https://hal.archives-ouvertes.fr/hal-01417442>

Books or Proceedings Editing

- [28] S. BONNEVAY, P. LEGRAND, N. MONMARCHÉ, E. LUTTON, M. SCHOENAUER (editors). *Artificial Evolution 2015*, LNCS - Lecture Notes in Computer Science, Springer, Lyon, France, 2016, vol. 9554 [DOI : 10.1007/978-3-319-31471-6], <https://hal.inria.fr/hal-01389072>
- [29] O. SCHUETZE, L. TRUJILLO, P. LEGRAND, Y. MALDONADO (editors). *NEO 2015 - Numerical and Evolutionary Optimization: Results of the Numerical and Evolutionary Optimization Workshop NEO 2015*

held at September 23-25 2015 in Tijuana, Mexico, Studies in Computational Intelligence, Springer, Tijuana, Mexico, 2016, <https://hal.inria.fr/hal-01389071>

Other Publications

- [30] R. AZAÏS, A. GENADOT, B. HENRY. *Inference for conditioned Galton-Watson trees from their Harris path*, September 2016, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01360650>
- [31] L. CHAARI, S. BADILLO, T. VINCENT, G. DEHAENE-LAMBERTZ, F. FORBES, P. CIUCIU. *Subject-level Joint Parcellation-Detection-Estimation in fMRI*, January 2016, working paper or preprint, <https://hal.inria.fr/hal-01255465>
- [32] I. CHARLIER, D. PAINDAVEINE, J. SARACCO. *Multiple-output quantile regression through optimal quantization*, 2016, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01417544>
- [33] M. CHAVENT, R. GENUER, J. SARACCO. *Combining clustering of variables and feature selection using random forests: the CoV/VSURF procedure*, July 2016, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01345840>
- [34] P. DEL MORAL, A. KURTZMANN, J. TUGAUT. *On the stability and the exponential concentration of Extended Kalman-Bucy filters*, June 2016, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01337709>
- [35] P. DEL MORAL, A. KURTZMANN, J. TUGAUT. *On the stability and the uniform propagation of chaos of Extended Ensemble Kalman-Bucy filters*, June 2016, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01337716>
- [36] M. ELLIES-OURY, M. GAGAOUA, J. SARACCO, M. CHAVENT, B. PICARD. *Biomarker abundance in two beef muscles depending on animal breeding practices and carcass characteristics*, 2016, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01417561>
- [37] A. GENADOT. *Averaging for some simple constrained Markov processes*, September 2016, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01336752>
- [38] I. JLASSI, J. SARACCO. *A Smooth Nonparametric Estimator of a Conditional Quantile*, 2016, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01417554>
- [39] I. JLASSI, J. SARACCO. *Variable importance assessment in sliced inverse regression for variable selection*, 2016, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01417552>
- [40] C. NIVOT, B. DE SAPORTA, F. DUFOUR, D. BÉRARD-BERGERY, C. ELEGBEDE. *Optimization of a launcher integration process: a Markov decision process approach **, 2016, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01269541>
- [41] A. TODESCHINI, F. CARON. *Exchangeable Random Measures for Sparse and Modular Graphs with Overlapping Communities*, February 2016, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01270854>

References in notes

- [42] N. DUAN, K.-C. LI. *Slicing regression: a link-free regression method*, in "Ann. Statist.", 1991, vol. 19, n^o 2, pp. 505–530, <http://dx.doi.org/10.1214/aos/1176348109>
- [43] R. DUDA, P. HART, D. STORK. *Pattern Classification*, John Wiley, 2001
- [44] K.-C. LI. *Sliced inverse regression for dimension reduction*, in "J. Amer. Statist. Assoc.", 1991, vol. 86, n^o 414, pp. 316–342, With discussion and a rejoinder by the author