

The logo for Inria, featuring the word "Inria" in a stylized, cursive red font.

IN PARTNERSHIP WITH:
CNRS

Université de Bordeaux

Activity Report 2019

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. Team, Visitors, External Collaborators	1
2. Overall Objectives	2
3. Research Program	3
3.1. Introduction	3
3.2. Main research topics	3
4. Application Domains	5
5. New Software and Platforms	5
5.1. biips	5
5.2. PCAmixdata	5
5.3. QuantifQuantile	6
6. New Results	6
6.1. Power-of-d-Choices with Memory: Fluid Limit and Optimality	6
6.2. Hamilton-Jacobi-Bellman Inequality for the Average Control of Piecewise Deterministic Markov Processes	6
6.3. Approximation of discounted minimax Markov control problems and zero-sum Markov games using Hausdorff and Wasserstein distances	7
6.4. Combining clustering of variables and feature selection using random forests: the CoV/VSURF procedure	7
6.5. Statistical model choice including variable selection based on variable importance: A relevant way for biomarkers selection to predict meat tenderness	7
6.6. Genome sequencing for rightward hemispheric language dominance	8
6.7. An Original Methodology for the Selection of Biomarkers of Tenderness in Five Different Muscles	8
6.8. New Approach Studying Interactions Regarding Trade-Off between Beef Performances and Meat Qualities	8
6.9. Impact of Speller Size on a Visual P300 Brain-Computer Interface (BCI) System under Two Conditions of Constraint for Eye Movement	9
6.10. High-Dimensional Multi-Block Analysis of Factors Associated with Thrombin Generation Potential	9
6.11. Multiple-output quantile regression through optimal quantization	9
6.12. Artificial evolution, fractal analysis and applications	10
6.13. Self-affinity of an Aircraft Pilot's Gaze Direction as a Marker of Visual Tunneling	12
6.14. Filtering-based Analysis Comparing the DFA with the CDFA for Wide Sense Stationary Processes	13
6.15. 2D Fourier Transform Based Analysis Comparing the DFA with the DMA	13
6.16. Interpréter les Fonctions de Fluctuation du DFA et du DMA comme le Résultat d'un Filtrage	13
6.17. A perturbation analysis of stochastic matrix Riccati diffusions	14
6.18. On the stability of matrix-valued Riccati diffusions	14
6.19. A variational approach to nonlinear and interacting diffusions	14
6.20. An explicit Floquet-type representation of Riccati aperiodic exponential semigroups	15
6.21. Uniform propagation of chaos and creation of chaos for a class of nonlinear diffusions	15
6.22. Stability Properties of Systems of Linear Stochastic Differential Equations with Random Coefficients	15
6.23. On One-Dimensional Riccati Diffusions	15
6.24. Adaptive Approximate Bayesian Computational Particle Filters for Underwater Terrain-Aided Navigation.	16
6.25. Inference for conditioned Galton-Watson trees from their Harris path	16
7. Bilateral Contracts and Grants with Industry	16
7.1.1. Naval Group	16

7.1.2.	Thales Optronique	16
7.1.3.	Case Law Analytics	16
8.	Partnerships and Cooperations	17
8.1.	National Initiatives	17
8.1.1.	QuAMProcs of the program Project Blanc of the ANR	17
8.1.2.	Chaire Stress Test of the Ecole Polytechnique	17
8.1.3.	ANR StocMC (2014-2018) of the program Project Blanc of the ANR	17
8.1.4.	ANR BNPSI: Bayesian Non Parametric methods for Signal and Image Processing	18
8.1.5.	Gaspard Monge Program for Optimisation and Operational Research (2017-2019)	18
8.1.6.	Mission pour les initiatives transverses et interdisciplinaires, Défi Modélisation du Vivant, projet MISGIVING	18
8.2.	European Initiatives	18
8.3.	International Initiatives	19
8.4.	International Research Visitors	19
8.4.1.	Visits of International Scientists	19
8.4.2.	Visits to International Teams	19
9.	Dissemination	20
9.1.	Promoting Scientific Activities	20
9.1.1.	Member of the Organizing Committees	20
9.1.2.	Chair of Conference Program Committees	20
9.1.3.	Journal	20
9.1.3.1.	Member of the Editorial Boards	20
9.1.3.2.	Reviewer - Reviewing Activities	20
9.1.4.	Scientific Expertise	20
9.1.5.	Research Administration	20
9.2.	Teaching - Supervision - Juries	21
9.2.1.	Teaching	21
9.2.2.	Supervision	21
9.3.	Popularization	22
10.	Bibliography	22

Project-Team CQFD

Creation of the Project-Team: 2009 January 01

Keywords:

Computer Science and Digital Science:

- A3.3. - Data and knowledge analysis
- A3.4.1. - Supervised learning
- A3.4.2. - Unsupervised learning
- A3.4.5. - Bayesian methods
- A3.4.6. - Neural networks
- A3.4.7. - Kernel methods
- A5.9.2. - Estimation, modeling
- A5.9.6. - Optimization tools
- A6.1.2. - Stochastic Modeling
- A6.1.3. - Discrete Modeling (multi-agent, people centered)
- A6.2.2. - Numerical probability
- A6.2.3. - Probabilistic methods
- A6.2.4. - Statistical methods
- A6.2.6. - Optimization
- A6.4.2. - Stochastic control
- A6.4.6. - Optimal control
- A9.2. - Machine learning
- A9.6. - Decision support

Other Research Topics and Application Domains:

- B1.2.2. - Cognitive science
- B2.2.4. - Infectious diseases, Virology
- B2.6.1. - Brain imaging
- B5.9. - Industrial maintenance
- B6.2. - Network technologies
- B6.3.3. - Network Management
- B6.5. - Information systems
- B9.2.3. - Video games
- B9.5.2. - Mathematics
- B9.5.3. - Physics
- B9.5.6. - Data science
- B9.11. - Risk management
- B9.11.1. - Environmental risks
- B9.11.2. - Financial risks

1. Team, Visitors, External Collaborators

Research Scientist

Pierre Del Moral [Inria, Senior Researcher, HDR]

Faculty Members

Francois Dufour [Team leader, Institut National Polytechnique de Bordeaux, Professor, HDR]

Marie Chavent [Univ de Bordeaux, Associate Professor, HDR]

Alexandre Genadot [Univ de Bordeaux, Associate Professor]

Pierrick Legrand [Univ de Bordeaux, Associate Professor, HDR]

Jerome Saracco [Institut National Polytechnique de Bordeaux, Professor, HDR]

Huilong Zhang [Univ de Bordeaux, Associate Professor]

PhD Students

Bastien Berthelot [Thales, PhD Student, granted by CIFRE]

Jimmy Bondu [Thales, PhD Student, granted by CIFRE]

Tiffany Cherchi [Thales, PhD Student, granted by CIFRE]

Alexandre Conanec [Inra, PhD Student]

Loic Labache [CEA, PhD Student, granted by CIFRE]

Corentin Lallier [Lectra, PhD Student]

Alex Mourer [Safran, PhD Student, from Feb 2019, granted by CIFRE]

Camille Palmier [ONERA, PhD Student]

Nathanael Randriamihamison [INRA, PhD Student]

Administrative Assistant

Catherine Cattaert Megrat [Inria, Administrative Assistant]

Visiting Scientists

Oswaldo Luiz Do Valle Costa [Universidade de São Paulo, from Dec 2019]

Anna Jaskiewicz [Politechnika Wroclawska, May 2019]

Tomas Prieto-Rumeau [Uned Madrid, September 2019]

External Collaborator

Benoite de Saporta [Univ Montpellier II (sciences et techniques du Languedoc), HDR]

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 dynamic and stochastic mathematical representations with discrete and/or continuous variables. Their 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 taking into account various probabilistic phenomena;
- *Estimation*: developing theoretical and computational methods in order to estimate the parameters of the model and to evaluate the performance of the system;
- *Control*: developing theoretical and numerical control tools to optimize the performance.

These three approaches are strongly connected and the most important feature of the team is to consider these topics as a whole. This enables the team to deal with real industrial problems in several contexts such as biology, production planning, trajectory generation and tracking, performance and reliability.

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. The team CQFD focuses on two complementary types of approaches. The first approach is based on mathematical representations built upon physical models where the dynamic of the real system is described by *stochastic processes*. The second one consists in studying the modeling issue in an abstract framework where the real system is considered as black-box. In this context, the outputs of the system are related to its inputs through a *statistical model*. Regarding stochastic processes, the team studies Piecewise Deterministic Markov Processes (PDMPs) and Markov Decision Processes (MDPs). These two classes of Markov processes form general families of controlled stochastic models suitable for the design of sequential decision-making problems. They appear in many fields such as biology, engineering, computer science, economics, operations research and provide powerful classes of processes for the modeling of complex systems. Our contribution to this topic consists in expressing real-life industrial problems into these mathematical frameworks. Regarding statistical methods, the team works on dimension reduction models. They provide a way to understand and visualize the structure of complex data sets. Furthermore, they are important tools in several different areas such as data analysis and machine learning, and appear in many applications such as biology, genetics, environment and recommendation systems. Our contribution to this topic consists in studying semiparametric modeling which combines the advantages of parametric and nonparametric models.

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. 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 [37] and Duan and Li [35].

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 [36].

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

The main objective is to develop *approximation techniques* to provide quasi-optimal feasible solutions and to derive *optimality results* for control problems related to MDPs and PDMPs:

- *Approximation techniques.* The analysis and the resolution of such decision models mainly rely on the maximum principle and/or the dynamic/linear programming techniques together with their various extensions such as the value iteration (VIA) and the policy iteration (PIA) algorithm. However, it is well known that these approaches are hardly applicable in practice and suffer from the so-called *curse of dimensionality*. Hence, solving numerically a PDMP or an MDP is a difficult and important challenge. Our goal is to obtain results which are both consistent from a theoretical point of view and computationally tractable and accurate from an application standpoint. It is important to emphasize that these research objectives were not planned in our initial 2009 program.

Our objective is to propose approximation techniques to efficiently compute the optimal value function and to get quasi-optimal controls for different classes of constrained and unconstrained MDPs with general state/action spaces, and possibly unbounded cost function. Our approach is based on combining the linear programming formulation of an MDP with probabilistic approximation techniques related to quantization techniques and the theory of empirical processes. An other aim is to apply our methods to specific industrial applications in collaboration with industrial partners such as Airbus Defence & Space, Naval Group and Thales.

Asymptotic approximations are also developed in the context of queueing networks, a class of models where the decision policy of the underlying MDP is in some sense fixed a priori, and our main goal is to study the transient or stationary behavior of the induced Markov process. Even though the decision policy is fixed, these models usually remain intractable to solve. Given this complexity, the team has developed analyses in some limiting regime of practical interest, i.e., queueing models in the large-network, heavy-traffic, fluid or mean-field limit. This approach is helpful to obtain a simpler mathematical description of the system under investigation, which is often given in terms of ordinary differential equations or convex optimization problems.

- *Optimality results.* Our aim is to investigate new important classes of optimal stochastic control problems including constraints and combining continuous and impulse actions for MDPs and PDMPs. In this framework, our objective is to obtain different types of optimality results. For

example, we intend to provide conditions to guarantee the existence and uniqueness of the optimality equation for the problem under consideration and to ensure existence of an optimal (and ϵ -optimal) control strategy. We also plan to analyze the structural properties of the optimal strategies as well as to study the associated infinite dimensional linear programming problem. These results can be seen as a first step toward the development of numerical approximation techniques in the sense described above.

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 is the collaboration that started in September 2014 with THALES Optronique. The goal of this project is the optimization of the maintenance of an onboard system equipped with a HUMS (Health Unit Monitoring Systems). 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.

The spectrum of applications of the topics that the team can address is large and can 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. New Software and Platforms

5.1. biips

Bayesian Inference with Interacting Particle Systems

KEYWORD: Bayesian estimation

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: Adrien Todeschini, François Caron, Pierre Del Moral and Pierrick Legrand
- Contact: Adrien Todeschini
- URL: <http://biips.gforge.inria.fr>

5.2. PCAmixdata

KEYWORD: Statistic analysis

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.

- Contact: Marie Chavent
- URL: <https://cran.r-project.org/web/packages/PCAmixdata/index.html>

5.3. QuantifQuantile

KEYWORD: Regression

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.

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

6. New Results

6.1. Power-of-d-Choices with Memory: Fluid Limit and Optimality

Abstract: In multi-server distributed queueing systems, the access of stochastically arriving jobs to resources is often regulated by a dispatcher, also known as load balancer. A fundamental problem consists in designing a load balancing algorithm that minimizes the delays experienced by jobs. During the last twodecades, the power-of-d-choice algorithm, based on the idea of dispatching each job to the least loaded server out of d servers randomly sampled at the arrival of the job itself, has emerged as a breakthrough in the foundations of this area due to its versatility and appealing asymptotic properties. In this paper, we consider the power-of-d-choice algorithm with the addition of a local memory that keeps track of the latest observations collected over time on the sampled servers. Then, each job is sent to a server with the lowest observation. We show that this algorithm is asymptotically optimal in the sense that the load balancer can always assign each job to an idle server in the large-system limit. Our results quantify and highlight the importance of using memory as a means to enhance performance in randomized load balancing.

Authors: J. Anselmi (CQFD); F. Dufour (CQFD).

6.2. Hamilton-Jacobi-Bellman Inequality for the Average Control of Piecewise Deterministic Markov Processes

Abstract : The main goal of this work is to study the infinite-horizon long run average continuous-time optimal control problem of piecewise deterministic Markov processes (PDMPs) with the control acting continuously on the jump intensity λ and on the transition measure Q of the process.

Authors : O.L.V. Costa; F. Dufour (CQFD)

6.3. Approximation of discounted minimax Markov control problems and zero-sum Markov games using Hausdorff and Wasserstein distances

Abstract : This work is concerned with a minimax control problem (also known as a robust Markov Decision Process (MDP) or a game against nature) with general state and action spaces under the discounted cost optimality criterion. We are interested in approximating numerically the value function and an optimal strategy of this general discounted minimax control problem. To this end, we derive structural Lipschitz continuity properties of the solution of this robust MDP by imposing suitable conditions on the model, including Lipschitz continuity of the elements of the model and absolute continuity of the Markov transition kernel with respect to some probability measure μ . Then, we are able to provide an approximating minimax control model with finite state and action spaces, and hence computationally tractable, by combining these structural properties with a suitable discretization procedure of the state space (related to a probabilistic criterion) and the action spaces (associated to a geometric criterion). Finally, it is shown that the corresponding approximation errors for the value function and the optimal strategy can be controlled in terms of the discretization parameters. These results are also extended to a two-player zero-sum Markov game.

Authors : F. Dufour (CQFD); T. Prieto-Rumeau

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

Abstract : Standard approaches to tackle high-dimensional supervised classification problem often include variable selection and dimension reduction procedures. The novel methodology proposed in this paper combines clustering of variables and feature selection. More precisely, hierarchical clustering of variables procedure allows to build groups of correlated variables in order to reduce the redundancy of information and summarizes each group by a synthetic numerical variable. Originality is that the groups of variables (and the number of groups) are unknown a priori. Moreover the clustering approach used can deal with both numerical and categorical variables (i.e. mixed dataset). Among all the possible partitions resulting from dendrogram cuts, the most relevant synthetic variables (i.e. groups of variables) are selected with a variable selection procedure using random forests. Numerical performances of the proposed approach are compared with direct applications of random forests and variable selection using random forests on the original p variables. Improvements obtained with the proposed methodology are illustrated on two simulated mixed datasets (cases $n > p$ and $n < p$, where n is the sample size) and on a real proteomic dataset. Via the selection of groups of variables (based on the synthetic variables), interpretability of the results becomes easier.

Authors : Marie Chavent (CQFD), Robin Genuer (SISTM), Jerome Saracco (CQFD)

6.5. Statistical model choice including variable selection based on variable importance: A relevant way for biomarkers selection to predict meat tenderness

Abstract : In this work, we describe a new computational methodology to select the best regression model to predict a numerical variable of interest Y and to select simultaneously the most interesting numerical explanatory variables strongly linked to Y . Three regression models (parametric, semi-parametric and non-parametric) are considered and estimated by multiple linear regression, sliced inverse regression and random forests. Both the variables selection and the model choice are computational. A measure of importance based on random perturbations is calculated for each covariate. The variables above a threshold are selected. Then a learning/test samples approach is used to estimate the Mean Square Error and to determine which model (including variable selection) is the most accurate. The R package `modvarsel` (MODEl and VARIABLE SElection) implements this computational approach and applies to any regression datasets. After checking the good behavior of the methodology on simulated data, the R package is used to select the proteins predictive of meat tenderness among a pool of 21 candidate proteins assayed in semitendinosus muscle from 71 young bulls. The biomarkers were selected by linear regression (the best regression model) to predict meat tenderness. These biomarkers, we confirm the predominant role of heat shock proteins and metabolic ones.

Authors : Marie-Pierre Ellies-Oury, Marie Chavent, Alexandre Conanec, Jérôme Saracco

6.6. Genome sequencing for rightward hemispheric language dominance

Abstract : Most people have left-hemisphere dominance for various aspects of language processing, but only roughly 1 % of the adult population has atypically reversed, rightward hemispheric language dominance (RHLD). The genetic-developmental program that underlies leftward language laterality is unknown, as are the causes of atypical variation. We performed an exploratory whole-genome-sequencing study, with the hypothesis that strongly penetrant, rare genetic mutations might sometimes be involved in RHLD. This was by analogy with situs inversus of the visceral organs (left-right mirror reversal of the heart, lungs etc.), which is sometimes due to monogenic mutations. The genomes of 33 subjects with RHLD were sequenced, and analysed with reference to large population-genetic datasets, as well as thirty-four subjects (14 left-handed) with typical language laterality. The sample was powered to detect rare, highly penetrant, monogenic effects if they would be present in at least 10 of the 33 RHLD cases and no controls, but no individual genes had mutations in more than 5 RHLD cases while being un-mutated in controls. A hypothesis derived from invertebrate mechanisms of left-right axis formation led to the detection of an increased mutation load, in RHLD subjects, within genes involved with the actin cytoskeleton. The latter finding offers a first, tentative insight into molecular genetic influences on hemispheric language dominance.

Authors : Amaia Carrion-castillo, Lise van der Haegen, Nathalie Tzourio-mazoyer, Tulya Kavaklioglu, Solveig Badillo, Marie Chavent, Jérôme Saracco, Marc Brysbaert, Simon Fisher, Bernard Mazoyer, Clyde Francks

6.7. An Original Methodology for the Selection of Biomarkers of Tenderness in Five Different Muscles

Abstract : For several years, studies conducted for discovering tenderness biomarkers have proposed a list of 20 candidates. The aim of the present work was to develop an innovative methodology to select the most predictive among this list. The relative abundance of the proteins was evaluated on five muscles of 10 Holstein cows: gluteobiceps, semimembranosus, semitendinosus, Triceps brachii and Vastus lateralis. To select the most predictive biomarkers, a multi-block model was used: The Data-Driven Sparse Partial Least Square. Semimembranosus and Vastus lateralis muscles tenderness could be well predicted ($R^2= 0.95$ and 0.94 respectively) with a total of 7 out of the 5 times 20 biomarkers analyzed. An original result is that the predictive proteins were the same for these two muscles: μ -calpain, m-calpain, h2afx and Hsp40 measured in m. gluteobiceps and μ -calpain, m-calpain and Hsp70-8 measured in m. Triceps brachii. Thus, this method is well adapted to this set of data, making it possible to propose robust candidate biomarkers of tenderness that need to be validated on a larger population.

Authors : Ellies-Oury, M.-P., Lorenzo, H., Denoyelle, C., Conanec, A., Saracco, J., Picard B.

6.8. New Approach Studying Interactions Regarding Trade-Off between Beef Performances and Meat Qualities

Abstract : The beef cattle industry is facing multiple problems, from the unequal distribution of added value to the poor matching of its product with fast-changing demand. Therefore, the aim of this study was to examine the interactions between the main variables, evaluating the nutritional and organoleptic properties of meat and cattle performances, including carcass properties, to assess a new method of managing the trade-off between these four performance goals. For this purpose, each variable evaluating the parameters of interest has been statistically modeled and based on data collected on 30 Blonde d'Aquitaine heifers. The variables were obtained after a statistical pre-treatment (clustering of variables) to reduce the redundancy of the 62 initial variables. The sensitivity analysis evaluated the importance of each independent variable in the models, and a graphical approach completed the analysis of the relationships between the variables. Then, the models were used to generate virtual animals and study the relationships between the nutritional and organoleptic quality. No apparent link between the nutritional and organoleptic properties of meat ($r = -0.17$)

was established, indicating that no important trade-off between these two qualities was needed. The 30 best and worst profiles were selected based on nutritional and organoleptic expectations set by a group of experts from the INRA (French National Institute for Agricultural Research) and Institut de l'Élevage (French Livestock Institute). The comparison between the two extreme profiles showed that heavier and fatter carcasses led to low nutritional and organoleptic quality.

Authors : Conanec ,A., Picard, B., Cantalapiedra-Hijar, G., Chavent, M., Denoyelle, C., Gruffat, D., Normand, J., Saracco, J., Ellies-Oury M.P.

6.9. Impact of Speller Size on a Visual P300 Brain-Computer Interface (BCI) System under Two Conditions of Constraint for Eye Movement

Abstract : The vast majority of P300-based brain-computer interface (BCI) systems are based on the well-known P300 speller presented by Farwell and Donchin for communication purposes and an alternative to people with neuromuscular disabilities, such as impaired eye movement. The purpose of the present work is to study the effect of speller size on P300-based BCI usability, measured in terms of effectiveness, efficiency, and satisfaction under overt and covert attention conditions. To this end, twelve participants used three speller sizes under both attentional conditions to spell 12 symbols. The results indicated that the speller size had, in both attentional conditions, a significant influence on performance. In both conditions (covert and overt), the best performances were obtained with the small and medium speller sizes, both being the most effective. The speller size did not significantly affect workload on the three speller sizes. In contrast, covert attention condition produced very high workload due to the increased resources expended to complete the task. Regarding users' preferences, significant differences were obtained between speller sizes. The small speller size was considered as the most complex, the most stressful, the less comfortable, and the most tiring. The medium speller size was always considered in the medium rank, which is the speller size that was evaluated less frequently and, for each dimension, the worst one. In this sense, the medium and the large speller sizes were considered as the most satisfactory. Finally, the medium speller size was the one to which the three standard dimensions were collected: high effectiveness, high efficiency, and high satisfaction. This work demonstrates that the speller size is an important parameter to consider in improving the usability of P300 BCI for communication purposes. The obtained results showed that using the proposed medium speller size, performance and satisfaction could be improved.

Authors :Ron-Angevin, R., Garcia, L., Fernandez-Rodriguez, A., Saracco, J., André, J.-M., Lespinet-Najib, V.

6.10. High-Dimensional Multi-Block Analysis of Factors Associated with Thrombin Generation Potential

Abstract : The identification of novel biological factors associated with thrombin generation, a key biomarker of the coagulation process, remains a relevant strategy to disentangle pathophysiological mechanisms underlying the risk of venous thrombosis (VT). As part of the MARseille THrombosis Association Study (MARTHA), we measured whole blood DNA methylation levels, plasma levels of 300 proteins, 3 thrombin generation biomarkers (endogenous thrombin potential, peak and lagtime), clinical and genetic data in 700 patients with VT. The application of a novel high-dimensional multi-levels statistical methodology we recently developed, the data driven sparse Partial Least Square method (ddsPLS), on the MARTHA datasets enabled us 1/ to confirm the role of a known mutation of the variability of endogenous thrombin potential and peak, 2/ to identify a new signature of 7 proteins strongly associated with lagtime.

Authors : Lorenzo, H., Razzaq, M., Odeberg, J., Saracco, J., Tregouet, D.-A., Thiébaud, R.

6.11. Multiple-output quantile regression through optimal quantization

Abstract : A new nonparametric quantile regression method based on the concept of optimal quantization was developed recently and was showed to provide estimators that often dominate their classical, kernel-type, competitors. In the present work, we extend this method to multiple-output regression problems. We show

how quantization allows approximating population multiple-output regression quantiles based on halfspace depth. We prove that this approximation becomes arbitrarily accurate as the size of the quantization grid goes to infinity. We also derive a weak consistency result for a sample version of the proposed regression quantiles. Through simulations, we compare the performances of our estimators with (local constant and local bilinear) kernel competitors. The results reveal that the proposed quantization-based estimators, which are local constant in nature, outperform their kernel counterparts and even often dominate their local bilinear kernel competitors. The various approaches are also compared on artificial and real data.

Authors : Charlier, I., Paidaveine, D., Saracco, J.

6.12. Artificial evolution, fractal analysis and applications

Abstract :

This document contains a selection of research works to which I have contributed. It is structured around two themes, artificial evolution and signal regularity analysis and consists of three main parts: Part I: Artificial evolution, Part II: Estimation of signal regularity and Part III: Applications, combination of signal processing, fractal analysis and artificial evolution. In order to set the context and explain the coherence of the rest of the document, this manuscript begins with an introduction, Chapter 1, providing a list of collaborators and of the research projects carried out. Theoretical contributions focus on two areas: evolutionary algorithms and the measurement of signal regularity and are presented in Part I and Part II respectively. These two themes are then exploited and applied to real problems in Part III. Part I, Artificial Evolution, consists of 8 chapters. Chapter 2 contains a brief presentation of various types of evolutionary algorithms (genetic algorithms, evolutionary strategies and genetic programming) and presents some contributions in this area, which will be detailed later in the document. Chapter 3, entitled Prediction of Expected Performance for a Genetic Programming Classifier proposes a method to predict the expected performance for a genetic programming (GP) classifier without having to run the program or sample potential solutions in the research space. For a given classification problem, a pre-processing step to simplify the feature extraction process is proposed. Then the step of extracting the characteristics of the problem is performed. Finally, a PEP (prediction of expected performance) model is used, which takes the characteristics of the problem as input and produces the predicted classification error on the test set as output. To build the PEP model, a supervised learning method with a GP is used. Then, to refine this work, an approach using several PEP models is developed, each now becoming a specialized predictors of expected performance (SPEP) specialized for a particular group of problems. It appears that the PEP and SPEP models were able to accurately predict the performance of a GP-classifier and that the SPEP approach gave the best results. Chapter 4, entitled A comparison of fitness-case sampling methods for genetic programming presents an extensive comparative study of four fitness-case sampling methods, namely: Interleaved Sampling, Random Interleaved Sampling, Lexicase Selection and the proposed Keep-Worst Interleaved Sampling. The algorithms are compared on 11 symbolic regression problems and 11 supervised classification problems, using 10 synthetic benchmarks and 12 real-world datasets. They are evaluated based on test performance, overfitting and average program size, comparing them with a standard GP search. The experimental results suggest that fitness-case sampling methods are particularly useful for difficult real-world symbolic regression problems, improving performance, reducing overfitting and limiting code growth. On the other hand, it seems that fitness-case sampling cannot improve upon GP performance when considering supervised binary classification. Chapter 5, entitled Evolving Genetic Programming Classifiers with Novelty Search, deals with a new and unique approach towards search and optimization, the Novelty Search (NS), where an explicit objective function is replaced by a measure of solution novelty. This chapter proposes a NS-based 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. 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. This chapter also discusses the effects of NS on GP search dynamics and code growth.

The 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. In Chapter 6, entitled Evaluating the Effects of Local Search in Genetic Programming, a memetic GP that incorporates a local search (LS) strategy to refine GP individuals expressed as syntax trees is studied in the context of symbolic regression. A simple parametrization for GP trees is proposed, by weighting each function with a parameter (unique for each function used in the construction of a tree). These parameters are then optimized using a trust region optimization algorithm which is therefore used here as a local search method. Then different heuristic methods are tested over several benchmark and real-world problems to determine which individuals from the tree population should be subjected to a LS. The results show that the best performances (in term of both quality of the solution and bloat control) was achieved when LS is applied to all of the solutions or to random individuals chosen from the top percentile (with respect to fitness) of the population. Chapter 7, entitled A Local Search Approach to Genetic Programming for Binary Classification, proposes a memetic GP, tailored for binary classification problems, extending the work on symbolic regression presented in the previous chapter. In particular, a small linear subtree is added on the top of the root node of the original tree and each node in a tree is weighted by a real-valued parameter, which is then numerically optimized using the trust-region algorithm used as a local search method. Experimental results show that potential classifiers produced by GP are improved by the local searcher, and hence the overall search is improved achieving substantial performance gains. Application on well-known benchmarks provided results competitive with state-of-the-art. Chapter 8, entitled RANSAC-GP: Dealing with Outliers in Symbolic Regression with Genetic Programming, presents a hybrid methodology based on the RANdom SAMpling Consensus (RANSAC) algorithm and GP, called RANSAC-GP. RANSAC is an approach to deal with outliers in parameter estimation problems, widely used in computer vision and related fields. This work presents the first application of RANSAC to symbolic regression with GP. The proposed algorithm is able to deal with extreme amounts of contamination in the training set, evolving highly accurate models even when the amount of outliers reaches 90%. Part II, Estimation of signal regularity consists of 3 chapters. Chapter 9, entitled Hölderian Regularity, provides some reminders and some theoretical contributions on the estimation of Hölderian regularity. Some details are given on the estimation of the Hölder exponent using oscillation method or a wavelet transform. These approaches and improved versions are compared on synthetic signals. The FracLab software, where all the above methods have been integrated, is also presented at the end of this chapter. The work proposed in Chapter 10, entitled Theoretical comparison of the DFA and variants for the estimation of the Hurst exponent, involves a theoretical and numerical comparison between the Detrended Fluctuation Analysis (DFA) and its variants, namely DMA, AFA, RDFA and the proposed Continuous DFA method, in which the trend is constrained to be continuous. The DFA is a well-established method to detect long-range correlations in time series. It has been used in a wide range of applications, from biomedical applications to signal denoising. It allows the Hurst exponent of a pure mono-fractal time series to be estimated. It operates as follows: after integration, the signal is split into segments. Using a least-squares criterion, local trends are deduced. The resulting piecewise linear trend is then subtracted to the whole signal. The power of the residual is computed for different segment lengths and its log-log representation allows the Hurst exponent to be deduced. The comparison performed in this chapter is based on a new common matrix writing formalism of the square of the fluctuation function from the instantaneous correlation function of the process for all these methods. In the case where the process under study is stationary in the broad sense, the statistical mean of the square of the fluctuation function is thus expressed as a weighted sum of the terms of the autocorrelation function, and this without any approximation. More precisely, the mathematical expectation of the square of the fluctuation function can be seen for each method as the autocorrelation function of the output of a filter dependent on this method and calculated for a lag equal to zero, i.e. the power of the filter output. In the general case, this analytical framework provides a means of comparing the DFA and its variants that is different from a traditional synthetic signal performance study, and explains the different behaviours of these regularity estimation methods, using the proposed filter analysis. Chapter 11 contains two patents with THALES AVS related to the work presented in the previous chapter. Part III of this manuscript, Applications, combination of signal processing, fractal analysis and artificial evolution, contains contributions combining the tools previously mentioned in order to develop new tools such as in the Chapters 12 and 13 or contributions on the resolution of real problems in the biomedical field, such as in the Chapters 14, 15 and 16. Chapter 12, entitled "The Estimation of Hölderian Regularity using Genetic

Programming", presents a GP approach to synthesize estimators for the pointwise Hölder exponent in 2D signals. The optimization problem to solve is to minimize the error between a prescribed regularity and the estimated regularity given by an image operator. The search for optimal estimators is then carried out using a GP algorithm. Experiments confirm that the GP operators produce a good estimation of the Hölder exponent in images of multifractional Brownian motions. In fact, the evolved estimators significantly outperform a traditional method by as much as one order of magnitude. These results provide further empirical evidence that GP can solve difficult problems of applied mathematics. In Chapter 13, entitled "Optimization of the Hölder Image Descriptor using a Genetic Algorithm", a local descriptor based on the Hölder exponent is studied. The proposal is to find an optimal number of dimensions for the descriptor using a genetic algorithm (GA). To guide the GA search, fitness is computed based on the performance of the descriptor when applied to standard region matching problems. This criterion is quantified using the F-Measure, derived from recall and precision analysis. Results show that it is possible to reduce the size of the canonical Hölder descriptor without degrading the quality of its performance. In fact, the best descriptor found through the GA search is nearly 70% performance on standard tests. Chapter 14, entitled "Interactive evolution for cochlear implants fitting", presents a study that intends to make cochlear implants more adaptable to environment and to simplify the process of fitting, by designing and using a specific interactive evolutionary algorithm combined with signal processing. Real experiments on volunteer implanted patients are presented, that show the efficiency of interactive evolution for this purpose. In Chapter 15, entitled "Feature extraction and classification of EEG signals. The use of a genetic algorithm for an application on alertness prediction", the development of computer systems for the automatic analysis and classification of mental states of vigilance; i.e., a person's state of alertness is studied. Such a task is relevant to diverse domains, where a person is expected or required to be in a particular state. For instance, pilots, security personnel or medical staffs are expected to be in a highly alert state, and a brain computer interface could help confirm this or detect possible problems. In this chapter, a combination of an evolutionary algorithm and signal processing is used. The purpose of this algorithm was to select an electrode and a frequency range to use in order to discriminate between the two states of vigilance. This approach determined the most useful electrode for the classification task. Using the recording of this electrode, the prediction obtained has a reliability rate of 89.33

In Chapter 16, entitled "Regularity and Matching Pursuit Feature Extraction for the Detection of Epileptic Seizures", a novel methodology for feature extraction on EEG signals that allows to perform a highly accurate classification of epileptic states is presented. 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% on a well known database, reveals that the proposal achieves state-of-the-art performance. The experimental results suggest that using a feature extraction methodology composed of regularity analysis, a Matching Pursuit algorithm and time-domain statistic measures together with a classifier produces a system that can predict epileptic states with competitive performance that matches or even surpasses other novel methods. Finally the last chapter concludes this manuscript and provides perspectives for future work.

Authors : Pierrick Legrand

6.13. Self-affinity of an Aircraft Pilot's Gaze Direction as a Marker of Visual Tunneling

Abstract : For the last few years, a great deal of interest has been paid to crew monitoring systems in order to tackle potential safety problems during a flight. They aim at detecting any degraded physiological and/or cognitive state of an aircraft pilot, such as attentional tunneling or excessive focalization. Indeed, they might have a negative impact on his performance to pursue his mission with adequate flight safety levels. One of the usual approaches consists in using sensors to collect physiological signals which are analyzed in real-time. Two main families exist to process the signals. The first one combines feature extraction and machine learning whereas the second is based on deep-learning approaches but may require a large amount

of labelled data. Here, we focused on the first family. In this case, various features can be deduced from the data by different approaches: spectrum analysis, a priori modelling and nonlinear dynamical system analysis techniques including the estimation of the self-affinity of the signals. In this paper, our purpose was to analyze whether the self-affinity of the pilot gaze direction can be related to his cognitive state. To this end, an experiment was carried out on 18 subjects in a representative aircraft environment based on a modified version of the software MATB-II. The scenarii were designed to elicit different levels of mental workload eventually associated to attentional tunneling. A database to train the machine learning step was first created by recording the directions of gaze of the subjects with an eye-tracker. The self-affinities of these signals were extracted with the Detrended Fluctuation Analysis method. They constituted the inputs of the classifier based on a Support Vector Machine. Then, new signals were analyzed and classified. Preliminary results showed promising abilities to detect attentional tunnelling episodes for different levels of mental workload.

Authors : Bastien Berthelot, Patrick Mazoyer, Sarah Egea, Jean-Marc André, Eric Grivel

6.14. Filtering-based Analysis Comparing the DFA with the CDFA for Wide Sense Stationary Processes

Abstract : The detrended fluctuation analysis (DFA) is widely used to estimate the Hurst exponent. Although it can be outperformed by wavelet based approaches, it remains popular because it does not require a strong expertise in signal processing. Recently, some studies were dedicated to its theoretical analysis and its limits. More particularly, some authors focused on the so-called fluctuation function by searching a relation with an estimation of the normalized covariance function under some assumptions. This paper is complementary to these works. We first show that the square of the fluctuation function can be expressed in a similar matrix form for the DFA and the variant we propose, called Continuous-DFA (CDFA), where the global trend is constrained to be continuous. Then, using the above representation for wide-sense-stationary processes, the statistical mean of the square of the fluctuation function can be expressed from the correlation function of the signal and consequently from its power spectral density, without any approximation. The differences between both methods can be highlighted. It also confirms that they can be seen as ad hoc wavelet based techniques.

Authors : Bastien Berthelot, Eric Grivel, Pierrick Legrand, Jean-Marc André, Patrick Mazoyer, et al

6.15. 2D Fourier Transform Based Analysis Comparing the DFA with the DMA

Abstract : Even if they can be outperformed by other methods, the detrended fluctuation analysis (DFA) and the detrended moving average (DMA) are widely used to estimate the Hurst exponent because they are based on basic notions of signal processing. For the last years, a great deal of interest has been paid to compare them and to better understand their behaviors from a mathematical point of view. In this paper, our contribution is the following: we first propose to express the square of the so-called fluctuation function as a 2D Fourier transform (2D-FT) of the product of two matrices. The first one is defined from the instantaneous correlations of the signal while the second, called the weighting matrix, is representative of each method. Therefore, the 2D-FT of the weighting matrix is analyzed in each case. In this study, differences between the DFA and the DMA are pointed out when the approaches are applied on non-stationary processes

Authors : Bastien Berthelot, Eric Grivel, Pierrick Legrand, Marc Donias, Jean-Marc André, et al.

6.16. Interpréter les Fonctions de Fluctuation du DFA et du DMA comme le Résultat d'un Filtrage

Abstract : The detrended fluctuation analysis (DFA) and the detrending moving average (DMA) are often used to estimate the regularity of the signal, since they do not require a strong expertise in the field of signal processing while providing good results. In this paper, our contribution is twofold. We propose a framework that allows these approaches to be compared. It is based on a matrix form of the square of the fluctuation function. Using

the above representation for wide-sense-stationary processes, we show that the statistical mean of the square of the fluctuation function can be expressed from the correlation function of the signal and consequently from its power spectral density, without any approximation. The differences between both methods can be highlighted. It also confirms that they can be seen as ad hoc wavelet based techniques to estimate the Hurst exponent.

Authors : Bastien Berthelot, Eric Grivel, Pierrick Legrand, Jean-Marc Andre, Patrick Mazoyer, et al.

6.17. A perturbation analysis of stochastic matrix Riccati diffusions

Abstract : Matrix differential Riccati equations are central in filtering and optimal control theory. The purpose of this article is to develop a perturbation theory for a class of stochastic matrix Riccati diffusions. Diffusions of this type arise, for example, in the analysis of ensemble Kalman-Bucy filters since they describe the flow of certain sample covariance estimates. In this context, the random perturbations come from the fluctuations of a mean field particle interpretation of a class of nonlinear diffusions equipped with an interacting sample covariance matrix functional. The main purpose of this article is to derive non-asymptotic Taylor-type expansions of stochastic matrix Riccati flows with respect to some perturbation parameter. These expansions rely on an original combination of stochastic differential analysis and nonlinear semigroup techniques on matrix spaces. The results here quantify the fluctuation of the stochastic flow around the limiting deterministic Riccati equation, at any order. The convergence of the interacting sample covariance matrices to the deterministic Riccati flow is proven as the number of particles tends to infinity. Also presented are refined moment estimates and sharp bias and variance estimates. These expansions are also used to deduce a functional central limit theorem at the level of the diffusion process in matrix spaces.

Authors : Adrian N. Bishop, Pierre Del Moral, Angele Niclas

6.18. On the stability of matrix-valued Riccati diffusions

Abstract : The stability properties of matrix-valued Riccati diffusions are investigated. The matrix-valued Riccati diffusion processes considered in this work are of interest in their own right, as a rather prototypical model of a matrix-valued quadratic stochastic process. Under rather natural observability and controllability conditions, we derive time-uniform moment and fluctuation estimates and exponential contraction inequalities. Our approach combines spectral theory with nonlinear semigroup methods and stochastic matrix calculus. This analysis seems to be the first of its kind for this class of matrix-valued stochastic differential equation. This class of stochastic models arise in signal processing and data assimilation, and more particularly in ensemble Kalman-Bucy filtering theory. In this context, the Riccati diffusion represents the flow of the sample covariance matrices associated with McKean-Vlasov-type interacting Kalman-Bucy filters. The analysis developed here applies to filtering problems with unstable signals.

Authors : Adrian N. Bishop, Pierre Del Moral

6.19. A variational approach to nonlinear and interacting diffusions

Abstract : The article presents a novel variational calculus to analyze the stability and the propagation of chaos properties of nonlinear and interacting diffusions. This differential methodology combines gradient flow estimates with backward stochastic interpolations, Lyapunov linearization techniques as well as spectral theory. This framework applies to a large class of stochastic models including non homogeneous diffusions, as well as stochastic processes evolving on differentiable manifolds, such as constraint-type embedded manifolds on Euclidian spaces and manifolds equipped with some Riemannian metric. We derive uniform as well as almost sure exponential contraction inequalities at the level of the nonlinear diffusion flow, yielding what seems to be the first result of this type for this class of models. Uniform propagation of chaos properties w.r.t. the time parameter are also provided. Illustrations are provided in the context of a class of gradient flow diffusions arising in fluid mechanics and granular media literature. The extended versions of these nonlinear Langevin-type diffusions on Riemannian manifolds are also discussed.

Authors : Marc Arnaudon (IMB), Pierre Del Moral (CMAP, CQFD)

6.20. An explicit Floquet-type representation of Riccati aperiodic exponential semigroups

Abstract : The article presents a rather surprising Floquet-type representation of time-varying transition matrices associated with a class of nonlinear matrix differential Riccati equations. The main difference with conventional Floquet theory comes from the fact that the underlying flow of the solution matrix is aperiodic. The monodromy matrix associated with this Floquet representation coincides with the exponential (fundamental) matrix associated with the stabilizing fixed point of the Riccati equation. The second part of this article is dedicated to the application of this representation to the stability of matrix differential Riccati equations. We provide refined global and local contraction inequalities for the Riccati exponential semigroup that depend linearly on the spectral norm of the initial condition. These refinements improve upon existing results and are a direct consequence of the Floquet-type representation, yielding what seems to be the first results of this type for this class of models.

Authors : Adrian N. Bishop, Pierre Del Moral

6.21. Uniform propagation of chaos and creation of chaos for a class of nonlinear diffusions

Abstract : We are interested in nonlinear diffusions in which the own law intervenes in the drift. This kind of diffusions corresponds to the hydrodynamical limit of some particle system. One also talks about propagation of chaos. It is well-known, for McKean-Vlasov diffusions, that such a propagation of chaos holds on finite-time interval. We here aim to establish a uniform propagation of chaos even if the external force is not convex, with a diffusion coefficient sufficiently large. The idea consists in combining the propagation of chaos on a finite-time interval with a functional inequality, already used by Bolley, Gentil and Guillin, see [BGG12a, BGG12b]. Here, we also deal with a case in which the system at time $t = 0$ is not chaotic and we show under easily checked assumptions that the system becomes chaotic as the number of particles goes to infinity together with the time. This yields the first result of this type for mean field particle diffusion models as far as we know.

Authors : Pierre Del Moral and Julian Tugaut

6.22. Stability Properties of Systems of Linear Stochastic Differential Equations with Random Coefficients

Abstract : This work is concerned with the stability properties of linear stochastic differential equations with random (drift and diffusion) coefficient matrices, and the stability of a corresponding random transition matrix (or exponential semigroup). We consider a class of random matrix drift coefficients that involves random perturbations of an exponentially stable flow of deterministic (time-varying) drift matrices. In contrast with more conventional studies, our analysis is not based on the existence of Lyapunov functions, and it does not rely on any ergodic properties. These approaches are often difficult to apply in practice when the drift/diffusion coefficients are random. We present rather weak and easily checked perturbation-type conditions for the asymptotic stability of time-varying and random linear stochastic differential equations. We provide new log-Lyapunov estimates and exponential contraction inequalities on any time horizon as soon as the fluctuation parameter is sufficiently small. These seem to be the first results of this type for this class of linear stochastic differential equations with random coefficient matrices.

Authors : Adrian N. Bishop, Pierre Del Moral

6.23. On One-Dimensional Riccati Diffusions

Abstract : This article is concerned with the fluctuation analysis and the stability properties of a class of one-dimensional Riccati diffusions. These one-dimensional stochastic differential equations exhibit a quadratic drift function and a non-Lipschitz continuous diffusion function. We present a novel approach, combining tangent process techniques, Feynman-Kac path integration, and exponential change of measures, to derive

sharp exponential decays to equilibrium. We also provide uniform estimates with respect to the time horizon, quantifying with some precision the fluctuations of these diffusions around a limiting deterministic Riccati differential equation. These results provide a stronger and almost sure version of the conventional central limit theorem. We illustrate these results in the context of ensemble Kalman-Bucy filtering. To the best of our knowledge, the exponential stability and the fluctuation analysis developed in this work are the first results of this kind for this class of nonlinear diffusions.

Authors: Adrian N. Bishop, Pierre Del Moral, Kengo Kamatani, Bruno Remillard

6.24. Adaptive Approximate Bayesian Computational Particle Filters for Underwater Terrain-Aided Navigation.

Authors : C. Palmier, K. Dahia, N. Merlinge, P. Del Moral, D. Laneuville & C. Musso

6.25. Inference for conditioned Galton-Watson trees from their Harris path

Tree-structured data naturally appear in various fields, particularly in biology where plants and blood vessels may be described by trees, but also in computer science because XML documents form a tree structure. This paper is devoted to the estimation of the relative scale parameter of conditioned Galton-Watson trees. New estimators are introduced and their consistency is stated. A comparison is made with an existing approach of the literature. A simulation study shows the good behavior of our procedure on finite-sample sizes and from missing or noisy data. An application to the analysis of revisions of Wikipedia articles is also considered through real data.

Authors: Romain Azaïs, Alexandre Genadot and Benoit Henry

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

7.1.1. Naval Group

Participants: Huilong Zhang, François Dufour, Dann Laneuville, Alexandre Genadot.

The increasing complexity of warfare submarine missions has led Naval Group to study new tactical help functions for underwater combat management systems. In this context, the objective is to find optimal trajectories according to the current mission type by taking into account sensors, environment and surrounding targets. This problem has been modeled as a discrete-time Markov decision process with finite horizon. A quantization technique has been applied to discretize the problem in order to get a finite MDP for which standard methods such as the dynamic and/or the linear programming approaches can be applied. Different kind of scenarios have been considered and studied.

7.1.2. Thales Optronique

Participants: Benoîte de Saporta, François Dufour, Tiffany Cerchi.

Maintenance, optimization, fleet of industrial equipments The topic of this collaboration with Université de Montpellier and Thales Optronique is the application of Markov decision processes to the maintenance optimization of a fleet of industrial equipments.

7.1.3. Case Law Analytics

Pierrick Legrand is a consultant for the startup Case Law Analytics. The object of the consulting is confidential.

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. *QuAMProcs of the program Project Blanc of the ANR*

The mathematical analysis of metastable processes started 75 years ago with the seminal works of Kramers on Fokker-Planck equation. Although the original motivation of Kramers was to « elucidate some points in the theory of the velocity of chemical reactions », it turns out that Kramers' law is observed to hold in many scientific fields: molecular biology (molecular dynamics), economics (modelization of financial bubbles), climate modeling, etc. Moreover, several widely used efficient numerical methods are justified by the mathematical description of this phenomenon.

Recently, the theory has witnessed some spectacular progress thanks to the insight of new tools coming from Spectral and Partial Differential Equations theory.

Semiclassical methods together with spectral analysis of Witten Laplacian gave very precise results on reversible processes. From a theoretical point of view, the semiclassical approach allowed to prove a complete asymptotic expansion of the small eigenvalues of Witten Laplacian in various situations (global problems, boundary problems, degenerate diffusions, etc.). The interest in the analysis of boundary problems was rejuvenated by recent works establishing links between the Dirichlet problem on a bounded domain and the analysis of exit event of the domain. These results open numerous perspectives of applications. Recent progress also occurred on the analysis of irreversible processes (e.g. on overdamped Langevin equation in irreversible context or full (inertial) Langevin equation).

The above progresses pave the way for several research tracks motivating our project: overdamped Langevin equations in degenerate situations, general boundary problems in reversible and irreversible case, non-local problems, etc.

8.1.2. *Chaire Stress Test of the Ecole Polytechnique*

The Chaire "Stress Testing" is a specific research program between Ecole Polytechnique, BNP Paribas, Fondation de l'Ecole Polytechnique, and is hosted at Polytechnique by the Center of Applied Mathematics. This research project is part of an in-depth reflection on the increasingly sophisticated issues surrounding stress tests (under the impulse of the upcoming European Banking regulation). Simulation of extreme adverse scenarios is an important topic to better understand which critical configurations can lead to financial and systemic crises. These scenarios may depend on complex phenomena, for which we partially lack information, making the modeling incomplete and uncertain. Last, the data are multivariate and reflect the dependency between driving variables. From the above observations, different lines of research are considered:

1. the generation of stress test and meta-modeling scenarios using machine learning;
2. the quantification of uncertainties in risk metrics;
3. modeling and estimation of multidimensional dependencies.

8.1.3. *ANR StocMC (2014-2018) of the program Project Blanc of the ANR*

The involved research groups are Inria Rennes/IRISA Team SUMO; Inria Rocquencourt Team Lifeware; LIAFA University Paris 7; Bordeaux University.

The aim of this research project is to develop scalable model checking techniques that can handle large stochastic systems. Large stochastic systems arise naturally in many different contexts, from network systems to system biology. A key stochastic model we will consider is from the biological pathway of apoptosis, the programmed cell death.

8.1.4. 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.

8.1.5. Gaspard Monge Program for Optimisation and Operational Research (2017-2019)

The involved research groups are Inria Bordeaux Sud-Ouest Team CQFD and Thales Optronique. This new collaboration with Thales Optronique that started in October 2017 is funded by the Fondation Mathématique Jacques Hadamard. This is the continuation of the PhD Thesis of A. Geeraert. The objective of this project is to optimize the maintenance of a multi-component equipment that can break down randomly. The underlying problem is to choose the best dates to repair or replace components in order to minimize a cost criterion that takes into account costs of maintenance but also the cost associated to the unavailability of the system for the customer. In the PhD thesis of A. Geeraert, the model under consideration was rather simple and only a numerical approximation of the value function was provided. Here, our objective is more ambitious. A more realistic model will be considered and our aim is to provide a tractable quasi-optimal control strategy that can be applied in practice to optimize the maintenance of such equipments.

8.1.6. Mission pour les initiatives transverses et interdisciplinaires, Défi Modélisation du Vivant, projet MISGIVING

The aim of MISGIVING (Mathematical Secrets penGuins dIVING) is to use mathematical models to understand the complexity of the multiscale decision process conditioning not only the optimal duration of a dive but also the diving behaviour of a penguin inside a bout. A bout is a sequence of successive dives where the penguin is chasing prey. The interplay between the chasing period (dives) and the resting period due to the physiological cost of a dive (the time spent at the surface) requires some kind of optimization.

8.2. European Initiatives

8.2.1. 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 Francois 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

8.3. International Initiatives

8.3.1. 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.

8.4. International Research Visitors

8.4.1. 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 december 2019.

Tomas Prieto-Rumeau (Department of Statistics and Operations Research, UNED, Madrid, Spain) visited the team during one week in 2019. The main subject of the collaboration is the approximation of Markov Decision Processes

Anna Jaskiewicz (Politechnika Wroclawska) visited the team during one week in 2019. The main subject of the collaboration is the approximation of Markov Decision Processes

8.4.2. Visits to International Teams

Pierrick Legrand visited the Instituto Tecnológico de Tijuana from 08/12/2019 to 17/12/2019.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Member of the Organizing Committees

Pierrick Legrand is co-organisor of the international conference EA 2019 in Mulhouse <https://ea2019.inria.fr/>.

9.1.2. Chair of Conference Program Committees

F. Dufour is the chair of the Program Committee of the SIAM Conference on Control and Its Applications (CT19) in Pittsburgh, USA, 2019.

9.1.3. Journal

9.1.3.1. Member of the Editorial Boards

P. Del Moral is an associate editor for the journal Stochastic Analysis and Applications since 2001.

P. Del Moral is an associate editor for the journal Annals of Applied Probability since 2018.

F. Dufour is corresponding editor of the SIAM Journal of Control and Optimization since 2018.

F. Dufour is associate editor of the journal Applied Mathematics & Optimization (AMO) since 2018.

F. Dufour is associate editor of the journal Stochastics: An International Journal of Probability and Stochastic Processes since 2018.

F. Dufour is the representative of the SIAM activity group in control and system theory for the journal SIAM News since 2014.

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

9.1.3.2. Reviewer - Reviewing Activities

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

9.1.4. Scientific Expertise

J. Saracco is elected member of the council of the Société Française de Statistique (SFdS, French Statistical Society). A. Genadot is member of the scientific council of the mathematical institute of Bordeaux.

P. Del Moral is a member of the Data Science Foundation of the American Institute of Mathematical Sciences.

9.1.5. Research Administration

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 and Pierrick Legrand are members of the council of the Institut de Mathématique de Bordeaux.

F. Dufour has been the coordinator for the Inria evaluation of the theme "Stochastic Approaches"

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

- Licence : P. Legrand, Algèbre, 129h, L1, Université de Bordeaux, France.
- Licence : P. Legrand, Espaces Euclidiens, 46,5h, L2, Université de Bordeaux, France.
- Licence : P. Legrand, Informatique pour les mathématiques, 30h, L2, Université de Bordeaux, France.
- DU : P. Legrand, Evolution Artificielle, Big data, 8h, DU, Bordeaux INP, France.
- Licence : A. Genadot, Bases en Probabilités, 18h, L1, Université de Bordeaux, France.
- Licence : A. Genadot, Projet Professionnel de l'étudiant, 8h, L1, Université de Bordeaux, France.
- Licence : A. Genadot, Probabilité, 30h, L2, Université de Bordeaux, France.
- Licence : A. Genadot, Techniques d'Enquêtes, 10h, L2, Université de Bordeaux, France.
- Licence : A. Genadot, Modélisation Statistiques, 16.5h, L3, Université de Bordeaux, France.
- Licence : A. Genadot, Préparation Stage, 15h, L3, Université de Bordeaux, France.
- Licence : A. Genadot, TER, 5h, L3, Université de Bordeaux, France.
- Licence : A. Genadot, Processus, 16.5h, L3, Université de Bordeaux, France.
- Licence : A. Genadot, Statistiques, 20h, L3, Bordeaux INP, France.
- Master : A. Genadot, Savoirs Mathématiques, 81h, M1, Université de Bordeaux et ESPE, France.
- Master : A. Genadot, Martingales, 29h, M1, Université de Bordeaux, France.
- Licence : F. Dufour, Probabilités et statistiques, 70h, first year of école ENSEIRB-MATMECA, Institut Polytechnique de Bordeaux, France.
- Master : F. Dufour, Approche probabiliste et methode de Monte Carlo, 24h, third year of école ENSEIRB-MATMECA, Institut Polytechnique de Bordeaux, France.

9.2.2. Supervision

- PhD: Hadrien Lorenzo, "Supervised analysis of high dimensional multi block data", supervised by Jérôme Saracco (CQFD) and Rodolphe Thebaut (Inserm), thesis defense: 27/11/19 in Bordeaux.
- PhD in progress: Alex Mourer, "Variables importance in clustering", CIFRE Safran Aircraft Engines, supervised by Jérôme Lacaille (Safran), Madalina Olteanu (SAMM, Paris1), Alex Mourer (doctorant), Marie Chavent (CQFD).
- PhD in progress: de Nathanaël Randriamihison, "Contiguity Constrained Hierarchical Agglomerative Clustering for Hi-C data analysis", supervised by Nathalie Vialaneix (MIAT, INRA Toulouse), Pierre Neuvial (IMT, CNRS), Marie Chavent (CQFD) .
- PhD in progress: Alexandre Conanec, "Modulation et optimisation statistique de données multi-tableaux : modélisation des facteurs de variations dans la gestion des compromis entre différents jeux de données", supervised by Marie Chavent(CQFD), Jérôme Saracco (CQFD), Marie-Pierre Ellies (INRA).
- PhD in progress: Loic Labache , "Création d'un atlas cérébral évolutif de régions fonctionnelles définies à partir d'une cohorte de 297 sujets ayant effectués 20 tâches cognitives en IRMf ", supervised by Jérôme Saracco (CQFD), Marc Joliot (CEA).
- PhD in progress: Tiffany Cherchi, "Automated optimal fleet management policy for airborne equipment", Montpellier University, since 2017, supervised by B. De Saporta and F. Dufour.
- PhD in progress: Bastien Berthelot, "Algorithmes de traitement du signal pour l'extraction de signatures robustes sur des bio-signaux", CIFRE THALES, supervised by P. Legrand.
- PhD in progress: Jimmy Bondu, "Classification de trajectoires d'objets par apprentissage.", CIFRE THALES, supervised by P. Legrand.

PhD in progress: Camille Palmier , “Nouvelles approches de fusion multi-capteurs par filtrage particulière pour le recalage de navigation inertielle par corrélation de cartes”, CIFRE, supervised by P. Del Moral, Dann Laneuville (NavalGroup) and Karim Dahia (ONERA)

9.3. Popularization

9.3.1. Internal or external Inria responsibilities

Alexandre Genadot is a member of the Commission des emplois de recherche Inria BSO. Pierrick Legrand is a member of the commission consultative section 26 of the IMB.

10. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

- [1] P. LEGRAND. *Artificial evolution, fractal analysis and applications*, Université de Bordeaux, November 2019, Habilitation à diriger des recherches, <https://hal.inria.fr/tel-02429815>

Articles in International Peer-Reviewed Journals

- [2] J. ANSEMI. *Combining Size-Based Load Balancing with Round-Robin for Scalable Low Latency*, in "IEEE Transactions on Parallel and Distributed Systems", 2019, pp. 1-3, forthcoming [DOI : 10.1109/TPDS.2019.2950621], <https://hal.archives-ouvertes.fr/hal-02276789>
- [3] J. ANSEMI, J. DONCEL. *Asymptotically Optimal Size-Interval Task Assignments*, in "IEEE Transactions on Parallel and Distributed Systems", 2019, vol. 30, n^o 11, pp. 2422-2433 [DOI : 10.1109/TPDS.2019.2920121], <https://hal.archives-ouvertes.fr/hal-02318576>
- [4] J. ANSEMI, F. DUFOUR. *Power-of-d-Choices with Memory: Fluid Limit and Optimality*, in "Mathematics of Operations Research", 2019, pp. 1-31, forthcoming, <https://hal.archives-ouvertes.fr/hal-02394147>
- [5] M. ARNAUDON, P. DEL MORAL. *A variational approach to nonlinear and interacting diffusions*, in "Stochastic Analysis and Applications", 2019, vol. 37, n^o 5, pp. 717-748 [DOI : 10.1080/07362994.2019.1609985], <https://hal.archives-ouvertes.fr/hal-02429162>
- [6] R. AZAÏS, A. GENADOT, B. HENRY. *Inference for conditioned Galton-Watson trees from their Harris path*, in "ALEA : Latin American Journal of Probability and Mathematical Statistics", 2019, vol. 16, n^o 1, pp. 1-45, <https://arxiv.org/abs/1609.01057> [DOI : 10.30757/ALEA.v16-21], <https://hal.archives-ouvertes.fr/hal-01360650>
- [7] A. N. BISHOP, P. DEL MORAL. *An explicit Floquet-type representation of Riccati aperiodic exponential semi-groups*, in "International Journal of Control", 2019, pp. 1-9 [DOI : 10.1080/00207179.2019.1590647], <https://hal.archives-ouvertes.fr/hal-02429300>
- [8] A. N. BISHOP, P. DEL MORAL. *On the stability of matrix-valued Riccati diffusions*, in "Electronic Journal of Probability", 2019, vol. 24 [DOI : 10.1214/19-EJP342], <https://hal.archives-ouvertes.fr/hal-02429272>

- [9] A. N. BISHOP, P. DEL MORAL. *Stability Properties of Systems of Linear Stochastic Differential Equations with Random Coefficients*, in "SIAM Journal on Control and Optimization", 2019, vol. 57, n^o 2, pp. 1023-1042 [DOI : 10.1137/18M1182759], <https://hal.archives-ouvertes.fr/hal-02429241>
- [10] A. N. BISHOP, P. DEL MORAL, K. KAMATANI, B. RÉMILLARD. *On one-dimensional Riccati diffusions*, in "Annals of Applied Probability", 2019, vol. 29, n^o 2, pp. 1127-1187 [DOI : 10.1214/18-AAP1431], <https://hal.archives-ouvertes.fr/hal-02429264>
- [11] A. CARRION-CASTILLO, L. VAN DER HAEGEN, N. TZOURIO-MAZOYER, T. KAVAKLIOGLU, S. BADILLO, M. CHAVENT, J. SARACCO, M. BRYSSBAERT, S. FISHER, B. MAZOYER, C. FRANCKX. *Genome sequencing for rightward hemispheric language dominance*, in "Genes, Brain and Behavior", April 2019, e12572 p. [DOI : 10.1111/GBB.12572], <https://hal.archives-ouvertes.fr/hal-02093544>
- [12] I. CHARLIER, D. PAINDAVEINE, J. SARACCO. *Multiple-output quantile regression through optimal quantization*, in "Scandinavian Journal of Statistics", December 2019 [DOI : 10.1111/SJOS.12426], <https://hal.archives-ouvertes.fr/hal-02429263>
- [13] M. CHAVENT, R. GENUER, J. SARACCO. *Combining clustering of variables and feature selection using random forests*, in "Communications in Statistics - Simulation and Computation", February 2019, pp. 1-20 [DOI : 10.1080/03610918.2018.1563145], <https://hal.archives-ouvertes.fr/hal-02013631>
- [14] A. CONANEC, B. PICARD, D. DURAND, G. CANTALAPIEDRA-HIJAR, M. CHAVENT, C. DENOYELLE, D. GRUFFAT, J. NORMAND, J. SARACCO, M.-P. ELLIES-OURY. *New Approach Studying Interactions Regarding Trade-Off between Beef Performances and Meat Qualities*, in "Foods", June 2019, vol. 8, n^o 6, 197 p. [DOI : 10.3390/FOODS8060197], <https://hal.archives-ouvertes.fr/hal-02429287>
- [15] O. L. D. V. COSTA, F. DUFOUR. *Hamilton-Jacobi-Bellman inequality for the average control of piecewise deterministic Markov processes*, in "Stochastics: An International Journal of Probability and Stochastic Processes", 2019, vol. 91, n^o 6, pp. 817-835 [DOI : 10.1080/17442508.2018.1546305], <https://hal.archives-ouvertes.fr/hal-02414337>
- [16] P. DEL MORAL, J. TUGAUT. *Uniform propagation of chaos and creation of chaos for a class of nonlinear diffusions*, in "Stochastic Analysis and Applications", 2019, vol. 37, n^o 6, pp. 909-935 [DOI : 10.1080/07362994.2019.1622426], <https://hal.archives-ouvertes.fr/hal-02429140>
- [17] F. DUFOUR, T. PRIETO-RUMEAU. *Approximation of Discounted Minimax Markov Control Problems and Zero-Sum Markov Games Using Hausdorff and Wasserstein Distances*, in "Dynamic Games and Applications", 2019, vol. 9, n^o 1, pp. 68-102 [DOI : 10.1007/s13235-018-0253-Y], <https://hal.inria.fr/hal-01953983>
- [18] M. P. ELLIES-OURY, M. CHAVENT, A. CONANEC, M. BONNET, B. PICARD, J. SARACCO. *Statistical model choice including variable selection based on variable importance: A relevant way for biomarkers selection to predict meat tenderness*, in "Scientific Reports", December 2019, vol. 9, n^o 1 [DOI : 10.1038/s41598-019-46202-Y], <https://hal.archives-ouvertes.fr/hal-02429345>
- [19] M.-P. ELLIES-OURY, H. LORENZO, C. DENOYELLE, J. SARACCO, B. PICARD. *An Original Methodology for the Selection of Biomarkers of Tenderness in Five Different Muscles*, in "Foods", June 2019, vol. 8, n^o 6, 206 p. [DOI : 10.3390/FOODS8060206], <https://hal.archives-ouvertes.fr/hal-02429282>

- [20] R.-A. RICARDO, L. GARCIA, Á. FERNÁNDEZ-RODRÍGUEZ, J. SARACCO, J.-M. ANDRÉ, V. LESPINET-NAJIB. *Impact of Speller Size on a Visual P300 Brain-Computer Interface (BCI) System under Two Conditions of Constraint for Eye Movement*, in "Computational Intelligence and Neuroscience", July 2019, vol. 2019, pp. 1-16 [DOI : 10.1155/2019/7876248], <https://hal.archives-ouvertes.fr/hal-02429293>

Invited Conferences

- [21] B. DE SAPORTA, F. DUFOUR, H. ZHANG. *Numerical Approximation of Optimal Strategies for Impulse Control of Piecewise Deterministic Markov Processes. Application to Maintenance Optimisation*, in "SIAM Conference on Control and Its Applications CT19", Chengdu, China, June 2019, <https://hal.archives-ouvertes.fr/hal-02161718>

International Conferences with Proceedings

- [22] B. BERTHELOT, E. GRIVEL, P. LEGRAND, J.-M. ANDRÉ, P. MAZOYER, T. FERREIRA. *Filtering-based Analysis Comparing the DFA with the CDFA for Wide Sense Stationary Processes*, in "EUSIPCO 2019 - 27th European Signal Processing Conference", A Coruna, Spain, September 2019, <https://hal.archives-ouvertes.fr/hal-02147655>
- [23] B. BERTHELOT, E. GRIVEL, P. LEGRAND, M. DONIAS, J.-M. ANDRÉ, P. MAZOYER, T. FERREIRA. *2D Fourier Transform Based Analysis Comparing the DFA with the DMA*, in "EUSIPCO 2019 - 27th European Signal Processing Conference", A Coruna, Spain, September 2019, <https://hal.archives-ouvertes.fr/hal-02147663>

Conferences without Proceedings

- [24] T. CHERCHI, C. BAYSSE, B. DE SAPORTA, F. DUFOUR. *Optimal predictive maintenance policy for multi-component systems*, in "JdS 2019 - 51èmes Journées de Statistique", Nancy, France, June 2019, <https://hal.archives-ouvertes.fr/hal-02376646>
- [25] B. DE SAPORTA, F. DUFOUR, H. ZHANG. *Maintenance optimization with piecewise deterministic Markov processes*, in "Journées IMT-IMAG", Toulouse, France, April 2019, <https://hal.archives-ouvertes.fr/hal-02376724>
- [26] H. LORENZO, R. MISBAH, J. ODEBER, P.-E. MORANGE, J. SARACCO, D.-A. TRÉGOUËT, R. THIÉBAUT. *High-Dimensional Multi-Block Analysis of Factors Associated with Thrombin Generation Potential*, in "CBMS 2019 - 32nd IEEE International Symposium on Computer-Based Medical Systems (CBMS)", Cordoba, Spain, IEEE, June 2019, pp. 453-458 [DOI : 10.1109/CBMS.2019.00094], <https://hal.archives-ouvertes.fr/hal-02429302>

Scientific Books (or Scientific Book chapters)

- [27] F. DUFOUR, T. PRIETO-RUMEAU. *Numerical Approximations for Discounted Continuous Time Markov Decision Processes*, in "Modeling, Stochastic Control, Optimization, and Applications", G. YIN, Q. ZHANG (editors), Springer, July 2019, pp. 147-171 [DOI : 10.1007/978-3-030-25498-8_7], <https://hal.archives-ouvertes.fr/hal-02414376>

Research Reports

- [28] B. D. O. ANDERSON, A. N. BISHOP, P. DEL MORAL, C. PALMIER. *Backward Nonlinear Smoothing Diffusions*, ANU; UTS; Inria; IMB., October 2019, <https://arxiv.org/abs/1910.14511> , <https://hal.archives-ouvertes.fr/hal-02342600>
- [29] P. DEL MORAL, S. S. SINGH. *A forward-backward stochastic analysis of diffusion flows*, Inria, June 2019, <https://arxiv.org/abs/1906.09145> , <https://hal.inria.fr/hal-02161914>

Other Publications

- [30] M. ARNAUDON, P. DEL MORAL. *A second order analysis of McKean-Vlasov semigroups*, December 2019, <https://arxiv.org/abs/1906.05140> - working paper or preprint, <https://hal.archives-ouvertes.fr/hal-02151808>
- [31] M. ARNAUDON, P. DEL MORAL. *A variational approach to nonlinear and interacting diffusions*, January 2019, <https://arxiv.org/abs/1812.04269> - working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01950673>
- [32] B. DELYON, B. DE SAPORTA, N. KRELL, L. ROBERT. *Investigation of asymmetry in E. coli growth rate*, October 2019, <https://arxiv.org/abs/1509.05226> - working paper or preprint, <https://hal.archives-ouvertes.fr/hal-02327095>
- [33] F. DUFOUR, A. GENADOT. *A convex programming approach for discrete-time Markov decision processes under the expected total reward criterion*, April 2019, <https://arxiv.org/abs/1903.08853> - working paper or preprint, <https://hal.archives-ouvertes.fr/hal-02071036>
- [34] N. RANDRIAMIHAMISON, N. VIALANEIX, P. NEUVIAL. *Applicability and Interpretability of Hierarchical Agglomerative Clustering With or Without Contiguity Constraints*, September 2019, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-02294847>

References in notes

- [35] 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>
- [36] R. DUDA, P. HART, D. STORK. *Pattern Classification*, John Wiley, 2001
- [37] 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