

RESEARCH CENTRE

**Inria Centre at Université
Grenoble Alpes**

IN PARTNERSHIP WITH:

Université de Grenoble Alpes, CNRS

2024

ACTIVITY REPORT

Project-Team

POLARIS

**Performance analysis and Optimization of
LARge Infrastructures and Systems**

IN COLLABORATION WITH: Laboratoire d'Informatique de Grenoble (LIG)

DOMAIN

**Networks, Systems and Services,
Distributed Computing**

THEME

**Distributed and High Performance
Computing**

Inria

Contents

| | |
|--|-----------|
| Project-Team POLARIS | 1 |
| 1 Team members, visitors, external collaborators | 2 |
| 2 Overall objectives | 3 |
| 2.1 Context | 3 |
| 2.2 Objectives | 4 |
| 2.3 Contribution to AI/Learning | 4 |
| 3 Research program | 5 |
| 3.1 Performance Evaluation | 5 |
| 3.2 Asymptotic Methods | 7 |
| 3.3 Distributed Online Optimization and Learning in Games | 8 |
| 3.4 Responsible Computer Science | 10 |
| 4 Application domains | 12 |
| 4.1 Large Computing Infrastructures | 12 |
| 4.2 Next-Generation Wireless Networks | 12 |
| 4.3 Energy and Transportation | 13 |
| 4.4 Social Computing Systems | 13 |
| 5 Social and environmental responsibility | 13 |
| 5.1 Footprint of research activities | 13 |
| 5.2 Raising awareness on the climate crisis | 13 |
| 5.3 Impact of research results | 14 |
| 6 Highlights of the year | 14 |
| 6.1 Awards | 14 |
| 7 New software, platforms, open data | 14 |
| 7.1 New software | 14 |
| 7.1.1 SimGrid | 14 |
| 7.1.2 PSI | 15 |
| 7.1.3 marmoteCore | 16 |
| 7.1.4 MarTO | 16 |
| 7.1.5 GameSeer | 16 |
| 7.1.6 rmf_tool | 17 |
| 7.1.7 SCIP | 17 |
| 8 New results | 18 |
| 8.1 Scheduling in Data Centers | 18 |
| 8.2 Performance evaluation of Large Systems (Mean Field) | 20 |
| 8.3 Reinforcement Learning and MDP | 22 |
| 8.4 Optimization Techniques | 23 |
| 8.4.1 Mixed Integer Optimization | 23 |
| 8.4.2 Application to various domains | 24 |
| 8.4.3 Optimization for Machine Learning | 25 |
| 8.5 Learning in games | 26 |
| 8.6 Random matrix analysis and Machine Learning | 28 |
| 8.7 Fairness and equity in digital (recommendation, advertising, persistent storage) systems | 29 |
| 8.8 Training Secondary School Teachers in Computer Science | 31 |
| 9 Bilateral contracts and grants with industry | 31 |

| | |
|--|-----------|
| 10 Partnerships and cooperations | 32 |
| 10.1 International initiatives | 32 |
| 10.1.1 Inria associate team not involved in an IIL or an international program | 32 |
| 10.2 International research visitors | 32 |
| 10.2.1 Visits of international scientists | 32 |
| 10.2.2 Visits to international teams | 32 |
| 10.3 National initiatives | 32 |
| 11 Dissemination | 33 |
| 11.1 Promoting scientific activities | 33 |
| 11.1.1 Scientific events: organisation | 33 |
| 11.1.2 Scientific events: selection | 33 |
| 11.1.3 Journal | 33 |
| 11.1.4 Invited talks | 34 |
| 11.1.5 Leadership within the scientific community | 35 |
| 11.1.6 Scientific expertise | 35 |
| 11.1.7 Research administration | 35 |
| 11.2 Teaching - Supervision - Juries | 36 |
| 11.2.1 Teaching | 36 |
| 11.2.2 Supervision | 37 |
| 11.2.3 Juries | 37 |
| 11.3 Popularization | 37 |
| 11.3.1 Specific official responsibilities in science outreach structures | 37 |
| 11.3.2 Participation in Live events | 37 |
| 12 Scientific production | 38 |
| 12.1 Publications of the year | 38 |
| 12.2 Cited publications | 41 |

Project-Team POLARIS

Creation of the Project-Team: 2018 January 01

Keywords

Computer sciences and digital sciences

- A1.2. – Networks
- A1.3.5. – Cloud
- A1.3.6. – Fog, Edge
- A1.6. – Green Computing
- A3.4. – Machine learning and statistics
- A3.5.2. – Recommendation systems
- A5.2. – Data visualization
- A6. – Modeling, simulation and control
 - A6.2.3. – Probabilistic methods
 - A6.2.4. – Statistical methods
 - A6.2.6. – Optimization
 - A6.2.7. – High performance computing
- A8.2. – Optimization
- A8.9. – Performance evaluation
- A8.11. – Game Theory
- A9.2. – Machine learning
- A9.9. – Distributed AI, Multi-agent

Other research topics and application domains

- B4.4. – Energy delivery
 - B4.4.1. – Smart grids
- B4.5.1. – Green computing
- B6.2. – Network technologies
 - B6.2.1. – Wired technologies
 - B6.2.2. – Radio technology
- B6.4. – Internet of things
- B8.3. – Urbanism and urban planning
- B9.6.7. – Geography
- B9.7.2. – Open data
- B9.8. – Reproducibility

1 Team members, visitors, external collaborators

Research Scientists

- Arnaud Legrand [Team leader, CNRS, Senior Researcher]
- Jonatha Anselmi [INRIA, Researcher]
- Mathieu Besancon [INRIA, Researcher]
- Nicolas Gast [INRIA, Researcher]
- Bruno Gaujal [INRIA, Senior Researcher]
- Panayotis Mertikopoulos [CNRS, Senior Researcher]
- Bary Pradelski [CNRS, Maison Française d'Oxford, Researcher]

Faculty Members

- Romain Couillet [UGA, Professor]
- Vincent Danjean [UGA, Associate Professor]
- Guillaume Huard [UGA, Associate Professor, until Aug 2024]
- Kevin Marquet [UGA, Associate Professor]
- Florence Perronnin [UGA, Associate Professor]
- Jean-Marc Vincent [UGA, Associate Professor]
- Philippe Waille [UGA, Associate Professor]

Post-Doctoral Fellow

- Dheeraj Narasimha [INRIA, Post-Doctoral Fellow]

PhD Students

- Rim Alhajal [CNRS, from Oct 2024]
- Sebastian Allmeier [INPG SA, until Apr 2024]
- Helene Arvis [EDF, CIFRE, from Nov 2024]
- Achille Baucher [UGA]
- Victor Boone [INRIA, from Sep 2024 until Nov 2024]
- Victor Boone [UGA, until Aug 2024]
- Rémi Castera [UGA, until Sep 2024]
- Pierre-Louis Cauvin [CNRS, from May 2024]
- Romain Cravic [INRIA]
- Hugo Lebeau [UGA, from Oct 2024]
- Hugo Lebeau [UGA, until Sep 2024]
- Davide Legacci [UGA]

- Victor Leger [UGA, from Oct 2024 until Nov 2024]
- Victor Leger [UGA, until Sep 2024]
- Charles Sejourne [INRIA, until Mar 2024]
- Hubert Villuendas [UGA, from Oct 2024]

Interns and Apprentices

- Sehane Bel Houari-Durand [ENS DE LYON, Intern, from Feb 2024 until Jul 2024]
- Samuel Bounan [INRIA, Intern, from Oct 2024]
- Joel Charles-Rebuffe [ENS DE LYON, Intern, from Oct 2024 until Oct 2024]
- Joel Charles-Rebuffe [ENS DE LYON, Intern, from Feb 2024 until Jul 2024]
- Arno Delrat [FLORALIS, Intern, until Jul 2024]
- Galaad Langlois [ENS DE LYON, Intern, from Oct 2024]
- Rafael Mascarenhas Dalle Nery [INRIA, Intern, from Jun 2024 until Aug 2024]
- Hugo Poirot [FLORALIS, Intern, from May 2024 until Jul 2024]
- Hugo Strappazzon [INRIA, Intern, from Feb 2024 until Aug 2024]
- Lucas Sube [INRIA, Intern, until Jun 2024]
- Victor Thouvenot [INRIA, Intern, until Aug 2024]
- Hubert Villuendas [UGA, Intern, from Sep 2024 until Sep 2024]
- Hubert Villuendas [INRIA, Intern, from Feb 2024 until Aug 2024]

Administrative Assistants

- Luce Coelho [INRIA, from Oct 2024]
- Annie Simon [INRIA, until Oct 2024]

2 Overall objectives

2.1 Context

Large distributed infrastructures are rampant in our society. Numerical simulations form the basis of computational sciences and high performance computing infrastructures have become scientific instruments with similar roles as those of test tubes or telescopes. Cloud infrastructures are used by companies in such an intense way that even the shortest outage quickly incurs the loss of several millions of dollars. But every citizen also relies on (and interacts with) such infrastructures via complex wireless mobile embedded devices whose nature is constantly evolving. In this way, the advent of digital miniaturization and interconnection has enabled our homes, power stations, cars and bikes to evolve into smart grids and smart transportation systems that should be optimized to fulfill societal expectations.

Our dependence and intense usage of such gigantic systems obviously leads to very high expectations in terms of performance. Indeed, we strive for low-cost and energy-efficient systems that seamlessly adapt to changing environments that can only be accessed through uncertain measurements. Such digital systems also have to take into account both the users' profile and expectations to efficiently and fairly share resources in an online way. Analyzing, designing and provisioning such systems has thus become a real challenge.

Such systems are characterized by their **ever-growing size**, intrinsic **heterogeneity** and **distributedness**, **user-driven** requirements, and an unpredictable variability that renders them essentially **stochastic**. In such contexts, many of the former design and analysis hypotheses (homogeneity, limited hierarchy, omniscient view, optimization carried out by a single entity, open-loop optimization, user outside of the picture) have become obsolete, which calls for radically new approaches. Properly studying such systems requires a drastic rethinking of fundamental aspects regarding the system's **observation** (measure, trace, methodology, design of experiments), **analysis** (modeling, simulation, trace analysis and visualization), and **optimization** (distributed, online, stochastic).

2.2 Objectives

The goal of the POLARIS project is to **contribute to the understanding of the performance of very large scale distributed systems** by applying ideas from diverse research fields and application domains. We believe that studying all these different aspects at once without restricting to specific systems is the key to push forward our understanding of such challenges and to propose innovative solutions. This is why we intend to investigate problems arising from application domains as varied as large computing systems, wireless networks, smart grids and transportation systems.

The members of the POLARIS project cover a very wide spectrum of expertise in performance evaluation and models, distributed optimization, and analysis of HPC middleware. Specifically, POLARIS' members have worked extensively on:

Experiment design: Experimental methodology, measuring/monitoring/tracing tools, experiment control, design of experiments, and reproducible research, especially in the context of large computing infrastructures (such as computing grids, HPC, volunteer computing and embedded systems).

Trace Analysis: Parallel application visualization (paje, triva/viva, framesoc/ocelotl, ...), characterization of failures in large distributed systems, visualization and analysis for geographical information systems, spatio-temporal analysis of media events in RSS flows from newspapers, and others.

Modeling and Simulation: Emulation, discrete event simulation, perfect sampling, Markov chains, Monte Carlo methods, and others.

Optimization: Stochastic approximation, mean field limits, game theory, discrete and continuous optimization, learning and information theory.

2.3 Contribution to AI/Learning

AI and Learning is everywhere now. Let us clarify how our research activities are positioned with respect to this trend.

A first line of research in POLARIS is devoted to the use of statistical learning techniques (Bayesian inference) to model the expected performance of distributed systems, to build aggregated performance views, to feed simulators of such systems, or to detect anomalous behaviours.

In a distributed context it is also essential to design systems that can seamlessly adapt to the workload and to the evolving behaviour of its components (users, resources, network). Obtaining faithful information on the dynamic of the system can be particularly difficult, which is why it is generally more efficient to design systems that dynamically learn the best actions to play through trial and errors. A key characteristic of the work in the POLARIS project is to leverage regularly game-theoretic modeling to handle situations where the resources or the decision is distributed among several agents or even situations where a centralised decision maker has to adapt to strategic users.

An important research direction in POLARIS is thus centered on reinforcement learning (Multi-armed bandits, Q-learning, online learning) and active learning in environments with one or several of the following features:

- Feedback is limited (e.g., gradient or even stochastic gradients are not available, which requires for example to resort to stochastic approximations);
- Multi-agent setting where each agent learns, possibly not in a synchronised way (i.e., decisions may be taken asynchronously, which raises convergence issues);

- Delayed feedback (avoid oscillations and quantify convergence degradation);
- Non stochastic (e.g., adversarial) or non stationary workloads (e.g., in presence of shocks);
- Systems composed of a very large number of entities, that we study through mean field approximation (mean-field games and mean field control).

As a side effect, many of the gained insights can often be used to dramatically improve the scalability and the performance of the implementation of more standard machine or deep learning techniques over supercomputers.

The POLARIS members are thus particularly interested in the design and analysis of adaptive learning algorithms for multi-agent systems, i.e. agents that seek to progressively improve their performance on a specific task. The resulting algorithms should not only learn an efficient (Nash) equilibrium but they should also be capable of doing so quickly (low regret), even when facing the difficulties associated to a distributed context (lack of coordination, uncertain world, information delay, limited feedback, ...)

In the rest of this document, we describe in detail our new results in the above areas.

3 Research program

3.1 Performance Evaluation

Participants: Jonatha Anselmi, Vincent Danjean, Nicolas Gast, Guillaume Huard, Arnaud Legrand, Florence Perronnin, Jean-Marc Vincent.

Project-team positioning

Evaluating the scalability, robustness, energy consumption and performance of large infrastructures such as exascale platforms and clouds raises severe methodological challenges. The complexity of such platforms mandates empirical evaluation but direct experimentation via an application deployment on a real-world testbed is often limited by the few platforms available at hand and is even sometimes impossible (cost, access, early stages of the infrastructure design, etc.). Furthermore, such experiments are costly, difficult to control and therefore difficult to reproduce. Although many of these digital systems have been built by human, they have reached such a complexity level that we are no longer able to study them like artificial systems and have to deal with the same kind of experimental issues as natural sciences. The development of a sound experimental methodology for the evaluation of resource management solutions is among the most important ways to cope with the growing complexity of computing environments. Although computing environments come with their own specific challenges, we believe such general observation problems should be addressed by borrowing good practices and techniques developed in many other domains of science, in particular (1) Predictive Simulation, (2) Trace Analysis and Visualization, and (3) the Design of Experiments.

Scientific achievements

Large computing systems are particularly complex to understand because of the interplay between their discrete nature (originating from deterministic computer programs) and their stochastic nature (emerging from the physical world, long distance interactions, and complex hardware and software stacks). A first line of research in POLARIS is devoted to the design of relatively simple statistical models of key components of distributed systems and their exploitation to feed simulators of such systems, to build aggregated performance views, and to detect anomalous behaviors.

Predictive Simulation Unlike direct experimentation via an application deployment on a real-world testbed, simulation enables fully repeatable and configurable experiments that can often be conducted quickly for arbitrary hypothetical scenarios. In spite of these promises, current simulation practice is often not conducive to obtaining scientifically sound results. To date, most simulation results in the

parallel and distributed computing literature are obtained with simulators that are ad hoc, unavailable, undocumented, and/or no longer maintained. As a result, most published simulation results build on throw-away (short-lived and non validated) simulators that are specifically designed for a particular study, which prevents other researchers from building upon it. There is thus a strong need for recognized simulation frameworks by which simulation results can be reproduced, further analyzed and improved.

Many simulators of MPI applications have been developed by renowned HPC groups (e.g., at SDSC [117], BSC [63], UIUC [125], Sandia Nat. Lab. [123], ORNL [64] or ETH Zürich [93]) but most of them build on restrictive network and application modeling assumptions that generally prevent to faithfully predict execution times, which limits the use of simulation to indication of gross trends at best.

The *SimGrid* simulation toolkit, whose development started more than 20 years ago in UCSD, is a renowned project which gathers more than 1,700 citations and has supported the research of at least 550 articles. The most important contribution of POLARIS to this project in the last years has been to improve the quality of SimGrid to the point where it can be used effectively on a daily basis by practitioners to accurately reproduce the dynamic of real HPC systems. In particular, *SMPI* [71], a simulator based on SimGrid that simulates unmodified MPI applications written in C/C++ or FORTRAN, has now become a very unique tool allowing to faithfully study particularly complex scenario such as legacy Geophysics application that suffers from spatial and temporal load balancing problem [96, 95] or the HPL benchmark [69, 43]. We have shown that the performance (both for time and energy consumption [92]) predicted through our simulations was systematically within a few percents of real experiments, which allows to reliably tune the applications at very low cost. This capacity has also been leveraged to study (through *StarPU-SimGrid*) complex and modern task-based applications running on heterogeneous sets of hybrid (CPUs + GPUs) nodes [54]. The phenomenon studied through this approach would be particularly difficult to study through real experiments but yet allow to address real problems of these applications. Finally, SimGrid is also heavily used through *BatSim*, a batch simulator developed in the DATAMOVE team and which leverages SimGrid, to investigate the performance of machine learning strategies in a batch scheduling context [99, 50].

Trace Analysis and Visualization Many monolithic visualization tools have been developed by renowned HPC groups since decades (e.g., BSC [110], Jülich and TU Dresden [107, 66], UIUC [91, 113, 94] and ANL [124]) but most of these tools build on the classical information visualization [114] that consists in always first presenting an overview of the data, possibly by plotting everything if computing power allows, and then to allow users to zoom and filter, providing details on demand. However in our context, the amount of data comprised in such traces is several orders of magnitude larger than the number of pixels on a screen and displaying even a small fraction of the trace leads to harmful visualization artifacts. Such traces are typically made of events that occur at very different time and space scales and originate from different sources, which hinders classical approaches, especially when the application structure departs from classical MPI programs with a BSP/SPMD structure. In particular, modern HPC applications that build on a task-based runtime and run on hybrid nodes are particularly challenging to analyze. Indeed, the underlying task-graph is dynamically scheduled to avoid spurious synchronizations, which prevents classical visualizations to exploit and reveal the application structure.

In [78], we explain how modern data analytics tools can be used to build, from heterogeneous information sources, custom, reproducible and insightful visualizations of task-based HPC applications at a very low development cost in the *StarVZ* framework. By specifying and validating statistical models of the performance of HPC applications/systems, we manage to identify when their behavior departs from what is expected and detect performance anomalies. This approach has first been applied to state-of-the-art linear algebra libraries in [78] and more recently to a sparse direct solver [47]. In both cases, we have been able to identify and fix several non-trivial anomalies that had not been noticed even by the application and runtime developers. Finally, these models not only allow to reveal when applications depart from what is expected but also to summarize the execution by focusing on the most important features, which is particularly useful when comparing two executions.

Design of Experiments and Reproducibility Part of our work is devoted to the control of experiments on both classical (HPC) and novel (IoT/Fog in a smart home context) infrastructures. To this end, we heavily rely on experimental testbeds such as Grid5000 and FIT-IoT Lab that can be well-controlled but

real experiments are nonetheless quite resource-consuming. *Design of experiments* has been successfully applied in many fields (e.g., agriculture, chemistry, industrial processes) where experiments are considered expensive. Building on concrete use cases, we explore how *Design of Experiments* and *Reproducible Research* techniques can be used to (1) design transparent auto-tuning strategies of scientific computation kernels [65, 57] (2) set up systematic performance non regression tests on Grid5000 (450 nodes for 1.5 year) and detect many abnormal events (related to bios and system upgrades, cooling, faulty memory and power instability) that had a significant effect on the nodes, from subtle performance changes of 1% to much more severe degradation of more than 10%, and had yet been unnoticed by both Grid'5000 technical team and Grid'5000 users (3) design and evaluate the performance of service provisioning strategies [44, 72] in Fog infrastructures.

3.2 Asymptotic Methods

Participants: Jonatha Anselmi, Romain Couillet, Nicolas Gast, Bruno Gaujal, Florence Perronnin, Jean-Marc Vincent.

Project-team positioning

Stochastic models often suffer from the curse of dimensionality: their complexity grows exponentially with the number of dimensions of the system. At the same time, very large stochastic systems are sometimes easier to analyze: it can be shown that some classes of stochastic systems simplify as their dimension goes to infinity because of averaging effects such as the law of large numbers, or the central limit theorem. This forms the basis of what is called an *asymptotic method*, which consists in studying what happens when a system gets large in order to build an approximation that is easier to study or to simulate.

Within the team, the research that we conduct in this axis is to foster the applicability of these asymptotic methods to new application areas. This leads us to work on the application of classical methods to new problems, but also to develop new approximation methods that take into account special features of the systems we study (i.e., moderate number of dimensions, transient behavior, random matrices). Typical applications are mean field method for performance evaluation, application to distributed optimization, and more recently statistical learning. One originality of our work is to quantify precisely what is the error made by such approximations. This allows us to define refinement terms that lead to more accurate approximations.

Scientific achievements

Refined mean field approximation Mean field approximation is a well-known technique in statistical physics, that was originally introduced to study systems composed of a very large number of particles (say $n > 10^{20}$). The idea of this approximation is to assume that objects are independent and only interact between them through an average environment (the *mean* field). Nowadays, variants of this technique are widely applied in many domains: in game theory for instance (with the example of mean field games), but also to quantify the performance of distributed algorithms. Mean field approximation is often justified by showing that a system of n well-mixed interacting objects converges to its deterministic mean field approximation as n goes to infinity. Yet, this does not explain why mean field approximation provides a very accurate approximation of the behavior of systems composed by a few hundreds of objects or less. Until recently, this was essentially an open question.

In [79], we give a partial answer to this question. We show that, for most of the mean field models used for performance evaluation, the error made when using a mean field approximation is a $\Theta(1/n)$. This results greatly improved compared to previous work that showed that the error made by mean field approximation was smaller than $O(1/\sqrt{n})$. On the contrary, we obtain the *exact* rate of accuracy. This result came from the use of Stein's method that allows one to quantify precisely the distance between two stochastic processes. Subsequently, in [83], we show that the constant in the $\Theta(1/n)$ can be computed numerically by a very efficient algorithm. By using this, we define the notion of refined approximation

which consists in adding the $1/n$ -correction term. This methods can also be generalize to higher order extensions or [85, 80].

Design and analysis of distributed control algorithms Mean field approximation is widely used in the performance evaluation community to analyze and design distributed control algorithms. Our contribution in this domain has covered mainly two applications: cache replacement algorithms and load balancing algorithms.

Cache replacement algorithms are widely used in content delivery networks. In [67, 87, 86], we show how mean field and refined mean field approximation can be used to evaluate the performance of list-based cache replacement algorithms. In particular, we show that such policies can outperform the classically used LRU algorithm. A methodological contribution of our work is that, when evaluating precisely the behavior of such a policy, the refined mean field approximation is both faster and more accurate than what could be obtained with a stochastic simulator.

Computing resources are often spread across many machines. An efficient use of such resources requires the design of a good load balancing strategy, to distribute the load among the available machines. In [61, 62, 60], we study two paradigms that we use to design asymptotically optimal load balancing policies where a central broker sends tasks to a set of parallel servers. We show in [61, 60] that combining the classical round-robin allocation plus an evaluation of the tasks sizes can yield a policy that has a zero delay in the large system limit. This policy is interesting because the broker does not need any feedback from the servers. At the same time, this policy needs to estimate or know job durations, which is not always possible. A different approach is used in [62] where we consider a policy that does not need to estimate job durations but that uses some feedback from the servers plus a memory of where jobs where send. We show that this paradigm can also be used to design zero-delay load balancing policies as the system size grows to infinity.

Mean field games Various notions of mean field games have been introduced in the years 2000-2010 in theoretical economics, engineering or game theory. A mean field game is a game in which an individual tries to maximize its utility while evolving in a population of other individuals whose behavior are not directly affected by the individual. An equilibrium is a population dynamics for which a selfish individual would behave as the population. In [74], we develop the notion of discrete space mean field games, that is more amenable to study than the previously introduced notions of mean field games. This leads to two interesting contributions: mean field games are not always the limits of stochastic games as the number of players grow [73], mean field games can be used to study how much vaccination should be subsidized to encourage people to adapt a socially optimal behaviour [51].

3.3 Distributed Online Optimization and Learning in Games

Participants: Nicolas Gast, Romain Couillet, Bruno Gaujal, Arnaud Legrand, Patrick Loiseau, Panayotis Mertikopoulos, Bary Pradelski.

Project-team positioning

Online learning concerns the study of *repeated decision-making in changing environments*. Of course, depending on the context, the words “learning” and “decision-making” may refer to very different things: in economics, this could mean predicting how rational agents react to market drifts; in data networks, it could mean adapting the way packets are routed based on changing traffic conditions; in machine learning and AI applications, it could mean training a neural network or the guidance system of a self-driving car; etc. In particular, the changes in the learner’s environment could be either *exogenous* (that is, independent of the learner’s decisions, such as the weather affecting the time of travel), or *endogenous* (i.e., they could depend on the learner’s decisions, as in a game of poker), or any combination thereof. However, the goal for the learner(s) is always the same: *to make more informed decisions that lead to better rewards over time*.

The study of online learning models and algorithms dates back to the seminal work of Robbins, Nash and Bellman in the 50's, and it has since given rise to a vigorous research field at the interface of game theory, control and optimization, with numerous applications in operations research, machine learning, and data science. In this general context, our team focuses on the asymptotic behavior of online learning and optimization algorithms, both single- and multi-agent: whether they converge, at what speed, and/or what type of non-stationary, off-equilibrium behaviors may arise when they do not.

The focus of POLARIS on game-theoretic and Markovian models of learning covers a set of specific challenges that dovetail in a highly synergistic manner with the work of other learning-oriented teams within Inria (like SCOOOL in Lille, SIERRA in Paris, and THOTH in Grenoble), and it is an important component of Inria's activities and contributions in the field (which includes major industrial stakeholders like Google / DeepMind, Facebook, Microsoft, Amazon, and many others).

Scientific achievements

Our team's work on online learning covers both single- and multi-agent models; in the sequel, we present some highlights of our work structured along these basic axes.

In the single-agent setting, an important problem in the theory of Markov decision processes – i.e., discrete-time control processes with decision-dependent randomness – is the so-called “restless bandit” problem. Here, the learner chooses an action – or “arm” – from a finite set, and the mechanism determining the action's reward changes depending on whether the action was chosen or not (in contrast to standard Markov problems where the activation of an arm does not have this effect). In this general setting, Whittle conjectured – and Weber and Weiss proved – that Whittle's eponymous index policy is asymptotically optimal. However, the result of Weber and Weiss is purely asymptotic, and the rate of this convergence remained elusive for several decades. This gap was finally settled in a series of POLARIS papers [82][45], where we showed that Whittle indices (as well as other index policies) become optimal at a geometric rate under the same technical conditions used by Weber and Weiss to prove Whittle's conjecture, plus a technical requirement on the non-singularity of the fixed point of the mean-field dynamics. We also propose the first sub-cubic algorithm to compute Whittle and Gittins indexes. As for reinforcement learning in Markovian bandits, we have shown that Bayesian and optimistic approaches do not use the structure of Markovian bandits similarly: While Bayesian learning has both a regret and a computational complexity that scales linearly with the number of arms, optimistic approaches all incur an exponential computation time, at least in their current versions [81].

In the multi-agent setting, our work has focused on the following fundamental question:

Does the concurrent use of (possibly optimal) single-agent learning algorithms ensure convergence to Nash equilibrium in multi-agent, game-theoretic environments?

Conventional wisdom might suggest a positive answer to this question because of the following “folk theorem”: under no-regret learning, the agents' empirical frequency of play converges to the game's set of coarse correlated equilibria. However, the actual implications of this result are quite weak: First, it concerns the empirical frequency of play and not the day-to-day sequence of actions employed by the players. Second, it concerns coarse correlated equilibria which may be supported on strictly dominated strategies – and are thus unacceptable in terms of rationalizability. These realizations prompted us to make a clean break with conventional wisdom on this topic, ultimately showing that the answer to the above question is, in general, “no”: specifically, [103, 101] showed that the (optimal) class of “follow-the-regularized-leader” (FTRL) learning algorithms leads to Poincaré recurrence even in simple, 2×2 min-max games, thus precluding convergence to Nash equilibrium in this context.

This negative result generated significant interest in the literature as it contributed in shifting the focus towards identifying *which* Nash equilibria may arise as stable limit points of FTRL algorithms and dynamics. Earlier work by POLARIS on the topic [70, 104, 105] suggested that *strict* Nash equilibria play an important role in this question. This suspicion was recently confirmed in a series of papers [77, 52] where we established a sweeping negative result to the effect that *mixed Nash equilibria are incompatible with no-regret learning*. Specifically, we showed that any Nash equilibrium which is not strict cannot be stable and attracting under the dynamics of FTRL, especially in the presence of randomness and uncertainty. This result has significant implications for predicting the outcome of a multi-agent learning

process because, combined with [104], it establishes the following far-reaching equivalence: *a state is asymptotically stable under no-regret learning if and only if it is a strict Nash equilibrium.*

Going beyond finite games, this further raised the question of what type of non-convergent behaviors can be observed in continuous games – such as the class of stochastic min-max problems that are typically associated to generative adversarial networks (GANs) in machine learning. This question was one of our primary collaboration axes with EPFL, and led to a joint research project focused on the characterization of the convergence properties of zeroth-, first-, and (scalable) second-order methods in non-convex/non-concave problems. In particular, we showed in [53] that these state-of-the-art min-max optimization algorithms may converge with arbitrarily high probability to attractors that are in no way min-max optimal or even stationary – and, in fact, may not even contain a single stationary point (let alone a Nash equilibrium). Spurious convergence phenomena of this type can arise even in two-dimensional problems, a fact which corroborates the empirical evidence surrounding the formidable difficulty of training GANs.

3.4 Responsible Computer Science

Participants: Nicolas Gast, Romain Couillet, Bruno Gaujal, Arnaud Legrand, Patrick Loiseau, Panayotis Mertikopoulos, Bary Pradelski.

Project-team positioning

The topics in this axis emerge from current social and economic questions rather than from a fixed set of mathematical methods. To this end we have identified large trends such as energy efficiency, fairness, privacy, and the growing number of new market places. In addition, COVID has posed new questions that opened new paths of research with strong links to policy making.

Throughout these works, the focus of the team is on modeling aspects of the aforementioned problems, and obtaining strong theoretical results that can give high-level guidelines on the design of markets or of decision-making procedures. Where relevant, we complement those works by measurement studies and audits of existing systems that allow identifying key issues. As this work is driven by topics, rather than methods, it allows for a wide range of collaborations, including with enterprises (e.g., Naverlabs), policy makers, and academics from various fields (economics, policy, epidemiology, etc.).

Other teams at Inria cover some of the societal challenges listed here (e.g., PRIVATICS, COMETE) but rather in isolation. The specificity of POLARIS resides in the breadth of societal topics covered and of the collaborations with non-CS researchers and non-research bodies; as well as in the application of methods such as game theory to those topics.

Scientific achievements

Algorithmic fairness As algorithmic decision-making became increasingly omnipresent in our daily lives (in domains ranging from credits to advertising, hiring, or medicine); it also became increasingly apparent that the outcome of algorithms can be discriminatory for various reasons. Since 2016, the scientific community working on the problem of algorithmic fairness has been exponentially increasing. In this context, in the early days, we worked on better understanding the extent of the problem through measurement in the case of social networks [116]. In particular, in this work, we showed that in advertising platforms, discrimination can occur from multiple different internal processes that cannot be controlled, and we advocate for measuring discrimination on the outcome directly. Then we worked on proposing solutions to guarantee fair representation in online public recommendations (aka trending topics on Twitter) [68]. This is an example of an application in which it was observed that recommendations are typically biased towards some demographic groups. In this work, our proposed solution draws an analogy between recommendation and voting and builds on existing works on fair representation in voting. Finally, in most recent times, we worked on better understanding the sources of discrimination, in the particular simple case of selection problems, and the consequences of fixing it. While most works attribute discrimination to implicit bias of the decision maker [98], we identified a fundamentally different source of discrimination: Even in the absence of implicit bias in a decision maker's estimate of candidates'

quality, the estimates may differ between the different groups in their variance—that is, the decision maker’s ability to precisely estimate a candidate’s quality may depend on the candidate’s group [76]. We show that this differential variance leads to discrimination for two reasonable baseline decision makers (group-oblivious and Bayesian optimal). Then we analyze the consequence on the selection utility of imposing fairness mechanisms such as demographic parity or its generalization; in particular we identify some cases for which imposing fairness can improve utility. In [75], we also study similar questions in the two-stage setting, and derive the optimal selector and the “price of local fairness” one pays in utility by imposing that the interim stage be fair.

Privacy and transparency in social computing system Online services in general, and social networks in particular, collect massive amounts of data about their users (both online and offline). It is critical that (i) the users’ data is protected so that it cannot leak and (ii) users can know what data the service has about them and understand how it is used—this is the transparency requirement. In this context, we did two kinds of work. First, we studied social networks through measurement, in particular using the use case of Facebook. We showed that their advertising platform, through the PII¹-based targeting option, allowed attackers to discover some personal data of users [118]. We also proposed an alternative design—valid for any system that proposed PII-based targeting—and proved that it fixes the problem. We then audited the transparency mechanisms of the Facebook ad platform, specifically the “Ad Preferences” page that shows what interests the platform inferred about a user, and the “Why am I seeing this” button that gives some reasons why the user saw a particular ad. In both cases, we laid the foundation for defining the quality of explanations and we showed that the explanations given were lacking key desirable properties (they were incomplete and misleading, they have since been changed) [59]. A follow-up work shed further light on the typical uses of the platform [58]. In another work, we proposed an innovative protocol based on randomized withdrawal to protect public posts deletion privacy [106]. Finally, in [84], we study an alternative data sharing ecosystem where users can choose the precision of the data they give. We model it as a game and show that, if users are motivated to reveal data by a public good component of the outcome’s precision, then certain basic statistical properties (the optimality of generalized least squares in particular) no longer hold.

Online markets Market design operates at the intersection of computer science and economics and has become increasingly important as many markets are redesigned on digital platforms. Studying markets for commodities, in an ongoing project we evaluate how different fee models alter strategic incentives for both buyers and sellers. We identify two general classes of fees: for one, strategic manipulation becomes infeasible as the market grows large and agents therefore have no incentive to misreport their true valuation. On the other hand, strategic manipulation is possible and we show that in this case agents aim to maximally shade their bids. This has immediate implications for the design of such markets. By contrast, [102] considers a matching market where buyers and sellers have heterogeneous preferences over each other. Traders arrive at random to the market and the market maker, having limited information, aims to optimize when to open the market for a clearing event to take place. There is a tradeoff between thickening the market (to achieve better matches) and matching quickly (to reduce waiting time of traders in the market). The tradeoff is made explicit for a wide range of underlying preferences. These works are adding to an ongoing effort to better understand and design markets [112][46].

COVID The COVID-19 pandemic has put humanity to one of the defining challenges of its generation and as such naturally trans-disciplinary efforts have been necessary to support decision making. In a series of articles [49][109] we proposed Green Zoning. ‘Green zones’—areas where the virus is under control based on a uniform set of conditions—can progressively return to normal economic and social activity levels, and mobility between them is permitted. By contrast, stricter public health measures are in place in ‘red zones’, and mobility between red and green zones is restricted. France and Spain were among the first countries to introduce green zoning in April 2020. The initial success of this proposal opened up the way to a large amount of follow-up work analyzing and proposing various tools to effectively deploy

¹Personally Identifiable Information

different tools to combat the pandemic (e.g., focus-mass testing [111] and a vaccination policy [108]). In a joint work with a group of leading economists, public health researchers and sociologists it was found that countries that opted to aim to eliminate the virus fared better not only for public health, but also for the economy and civil liberties [48]. Overall this work has been characterized by close interactions with policy makers in France, Spain and the European Commission as well as substantial activity in public discourse (via TV, newspapers and radio).

Energy efficiency Our work on energy efficiency spanned multiple different areas and applications such as embedded systems and smart grids. Minimizing the energy consumption of embedded systems with real-time constraints is becoming more important for ecological as well as practical reasons since batteries are becoming standard power supplies. Dynamically changing the speed of the processor is the most common and efficient way to reduce energy consumption [115]. In fact, this is the reason why modern processors are equipped with Dynamic Voltage and Frequency Scaling (DVFS) technology [122]. In a stochastic environment, with random job sizes and arrival times, combining hard deadlines and energy minimization via DVFS-based techniques is difficult because forcing hard deadlines requires considering the worst cases, hardly compatible with random dynamics. Nevertheless, progress have been made to solve these types of problems in a series of papers using constrained Markov decision processes, both on the theoretical side (proving existence of optimal policies and showing their structure [90, 88, 89]) as well as on the experimental side (showing the gains of optimal policies over classical solutions [12]).

In the context of a collaboration with Enedis and Schneider Electric (via the Smart Grid chair of Grenoble-INP), we also study the problem of using smart meters to optimize the behavior of electrical distribution networks. We made three kinds of contributions on this subject: (1) how to design efficient control strategies in such a system [121, 120, 119], (2) how to co-simulate an electrical network and a communication network [97], and (3) what is the performance of the communication protocol (PLC G3) used by the Linky smart meters [100].

4 Application domains

4.1 Large Computing Infrastructures

Supercomputers typically comprise thousands to millions of multi-core CPUs with GPU accelerators interconnected by complex interconnection networks that are typically structured as an intricate hierarchy of network switches. Capacity planning and management of such systems not only raises challenges in term of computing efficiency but also in term of energy consumption. Most legacy (SPMD) applications struggle to benefit from such infrastructure since the slightest failure or load imbalance immediately causes the whole program to stop or at best to waste resources. To scale and handle the stochastic nature of resources, these applications have to rely on dynamic runtimes that schedule computations and communications in an opportunistic way. Such evolution raises challenges not only in terms of programming but also in terms of observation (complexity and dynamicity prevents experiment reproducibility, intrusiveness hinders large scale data collection, ...) and analysis (dynamic and flexible application structures make classical visualization and simulation techniques totally ineffective and require to build on *ad hoc* information on the application structure).

4.2 Next-Generation Wireless Networks

Considerable interest has arisen from the seminal prediction that the use of multiple-input, multiple-output (MIMO) technologies can lead to substantial gains in information throughput in wireless communications, especially when used at a massive level. In particular, by employing multiple inexpensive service antennas, it is possible to exploit spatial multiplexing in the transmission and reception of radio signals, the only physical limit being the number of antennas that can be deployed on a portable device. As a result, the wireless medium can accommodate greater volumes of data traffic without requiring the reallocation (and subsequent re-regulation) of additional frequency bands. In this context, throughput maximization in the presence of interference by neighboring transmitters leads to games with convex action sets (covariance matrices with trace constraints) and individually concave utility functions (each

user's Shannon throughput); developing efficient and distributed optimization protocols for such systems is one of the core objectives of the research theme presented in Section 3.3.

Another major challenge that occurs here is due to the fact that the efficient physical layer optimization of wireless networks relies on perfect (or close to perfect) channel state information (CSI), on both the uplink and the downlink. Due to the vastly increased computational overhead of this feedback – especially in decentralized, small-cell environments – the continued transition to fifth generation (5G) wireless networks is expected to go hand-in-hand with distributed learning and optimization methods that can operate reliably in feedback-starved environments. Accordingly, one of POLARIS' application-driven goals will be to leverage the algorithmic output of Theme 5 into a highly adaptive resource allocation framework for next-generation wireless systems that can effectively "learn in the dark", without requiring crippling amounts of feedback.

4.3 Energy and Transportation

Smart urban transport systems and smart grids are two examples of collective adaptive systems. They consist of a large number of heterogeneous entities with decentralised control and varying degrees of complex autonomous behaviour. We develop an analysis tool to help to reason about such systems. Our work relies on tools from fluid and mean-field approximation to build decentralized algorithms that solve complex optimization problems. We focus on two problems: decentralized control of electric grids and capacity planning in vehicle-sharing systems to improve load balancing.

4.4 Social Computing Systems

Social computing systems are online digital systems that use personal data of their users at their core to deliver personalized services directly to the users. They are omnipresent and include for instance recommendation systems, social networks, online medias, daily apps, etc. Despite their interest and utility for users, these systems pose critical challenges of privacy, security, transparency, and respect of certain ethical constraints such as fairness. Solving these challenges involves a mix of measurement and/or audit to understand and assess issues, and modeling and optimization to propose and calibrate solutions.

5 Social and environmental responsibility

5.1 Footprint of research activities

We try to keep the carbon footprint of the team as low as possible by a stricter laptop renewal policy and by reducing plane travels (e.g., using visioconference or sometimes by avoiding publishing our research in conferences that would take place on the other side of the planet).

Our team does not train heavy ML models requiring important processing power although some of us perform computer science experiments, mostly using the Grid5000 platforms. We keep this usage very reasonable and rely on cheaper alternatives (e.g., simulations) as much as possible.

5.2 Raising awareness on the climate crisis

Romain Couillet has organized several introductory seminars on the Anthropocene, which he has presented to students at the UGA and Grenoble-INP, as well as to associations in Grenoble (FNE, AgirAlternatif). He is also co-responsible of the *Digital Transformation DU*. He has published several articles on the issue of "usability" of artificial intelligence. He is also co-creator of the *sustainable AI* transversal axis of the MIAI project in Grenoble. He connects his professional activity with public action (Lowtechlab de Grenoble, Université Autogérée, Arche des Innovateurs, etc). Finally, he is a trainer for the "Fresque du Climat" and a member of Adrastia and FNE Isère.

5.3 Impact of research results

Jean-Marc Vincent is heavily engaged since several years in the training of computer science teachers at the elementary/middle/high school levels. Among one of his many activities, we can mention his involvement in the design of the *Numérique et Sciences Informatiques, NSI : les fondamentaux* MOOC. See section 11.2.1 and 11.3 for more details.

6 Highlights of the year

6.1 Awards

- Spotlight award at ICML 2024 for D. Legacci, P. Mertikopoulos and B. Pradelski. ‘A geometric decomposition of finite games: Convergence vs. recurrence under exponential weights’. In: ICML 2024 - 41st International Conference on Machine Learning. Vienna, Austria, July 2024. URL: <https://hal.science/hal-04629310>.
- Spotlight award at NeurIPS 2024 for D. Legacci et al. ‘No-regret learning in harmonic games: Extrapolation in the face of conflicting interests’. In: *Proceedings of the 38th International Conference on Neural Information Processing Systems*. NeurIPS 2024 - 38th Conference on Neural Information Processing Systems. Vancouver, Canada, 2024, pp. 1–36. URL: <https://inria.hal.science/hal-04863624>.
- Best student paper at CPAIOR 2024 for G. Mexi et al. ‘Probabilistic Lookahead Strong Branching via a Stochastic Abstract Branching Model’. In: *Integration of Constraint Programming, Artificial Intelligence, and Operations Research*. CPAIOR 2024 - 21st International Conference on Integration of Constraint Programming, Artificial Intelligence, and Operations Research. Vol. 14743. Lecture Notes in Computer Science. Uppsala, Sweden: Springer Nature Switzerland, 25th May 2024, pp. 56–73. DOI: [10.1007/978-3-031-60599-4_4](https://doi.org/10.1007/978-3-031-60599-4_4). URL: <https://hal.science/hal-04792529>.
- Lucas Leandro Nesi’s PhD thesis (L. Leandro Nesi. ‘Strategies for Distributing Task-Based Applications on Heterogeneous Platforms’. PhD thesis. Université Grenoble Alpes, 14th Sept. 2023. URL: <http://www.theses.fr/s300803>) was selected (ranked 2nd ex-æquo) for the PhD thesis award of the GDR RSD in 2024.
- Yu-Guan Hsieh was awarded the Univ. Grenoble Alpes 2024 Academic Thesis Prize for his research work² among PhDs graduating in 2023.

7 New software, platforms, open data

7.1 New software

7.1.1 SimGrid

Keywords: Large-scale Emulators, Grid Computing, Distributed Applications

Scientific Description: SimGrid is a toolkit that provides core functionalities for the simulation of distributed applications in heterogeneous distributed environments. The simulation engine uses algorithmic and implementation techniques toward the fast simulation of large systems on a single machine. The models are theoretically grounded and experimentally validated. The results are reproducible, enabling better scientific practices.

Its models of networks, cpus and disks are adapted to (Data)Grids, P2P, Clouds, Clusters and HPC, allowing multi-domain studies. It can be used either to simulate algorithms and prototypes of applications, or to emulate real MPI applications through the virtualization of their communication, or to formally assess algorithms and applications that can run in the framework.

²Y.-G. Hsieh. ‘Decision-Making in multi-agent systems: delays, adaptivity, and learning in games’. PhD thesis. Université Grenoble Alpes, 7th Nov. 2023. URL: <http://www.theses.fr/s245435>.

The formal verification module explores all possible message interleavings in the application, searching for states violating the provided properties. We recently added the ability to assess liveness properties over arbitrary and legacy codes, thanks to a system-level introspection tool that provides a finely detailed view of the running application to the model checker. This can for example be leveraged to verify both safety or liveness properties, on arbitrary MPI code written in C/C++/Fortran.

Functional Description: SimGrid is a simulation toolkit that provides core functionalities for the simulation of distributed applications in large scale heterogeneous distributed environments.

Release Contributions: The Anne of Brittany release (she became a Duchess 536 years ago).

* Various improvements and unification of the simulation APIs * MC: Enable the verification of Python programs, and of condvars and iprobe calls. * MC: Exhibit the critical transition when a failure is found. * (+ internal refactoring and bug fixes)

News of the Year: There were 2 major releases in 2023. On modeling aspects, we released new plugins simulating chiller, photovoltaic and battery components of Fog/Edge infrastructures, as well as the disk arrays used in desegregated infrastructures. We improved the consistency of the simulation core: the new ActivitySet containers now make it easy to way for the completion of an heterogeneous set of activities (computation, communication, I/O, etc). The simulation of workflow and dataflow applications was also streamlined, with more examples, more documentation and less bugs. A new model of activities mixing disk I/O and network communication was introduced, to efficiently simulate accesses to remote disks. In addition, many efforts were put on the profiling of the software, leading to massive performance gains. We also pursued our efforts to improve the overall framework, through bug fixes, code refactoring and other software quality improvements. In particular, interfaces that were deprecated since almost a decade were removed to ease the maintenance burden on our community.

Many improvement occurred on the model-checker side too. We dropped the old experiments toward stateful verification of liveness properties to boost the development of stateless verification of safety properties. Our tool is simpler internally, and usable on all major operating systems. We modernized the reduction algorithms, implementing several recent algorithms of the literature and paving the way to the introduction of new ones. We also introduced a new module allowing to verify not only distributed applications, but also threaded applications.

URL: <https://simgrid.org/>

Contact: Martin Quinson

Participants: Adrien Lebre, Anne-Cécile Orgerie, Arnaud Legrand, Augustin Degomme, Arnaud Giersch, Emmanuelle Saillard, Frédéric Suter, Jonathan Pastor, Martin Quinson, Samuel Thibault

Partners: CNRS, ENS Rennes

7.1.2 PSI

Name: Perfect Simulator

Keywords: Markov model, Simulation

Functional Description: Perfect simulator is a simulation software of markovian models. It is able to simulate discrete and continuous time models to provide a perfect sampling of the stationary distribution or directly a sampling a functional of this distribution by using coupling from the past. The simulation kernel is based on the CFTP algorithm, and the internal simulation of transitions on the Aliasing method.

URL: <https://gitlab.inria.fr/PSI/psi3/>

Contact: Jean-Marc Vincent

7.1.3 marmoteCore

Name: Markov Modeling Tools and Environments - the Core

Keywords: Modeling, Stochastic models, Markov model

Functional Description: marmoteCore is a C++ environment for modeling with Markov chains. It consists in a reduced set of high-level abstractions for constructing state spaces, transition structures and Markov chains (discrete-time and continuous-time). It provides the ability of constructing hierarchies of Markov models, from the most general to the particular, and equip each level with specifically optimized solution methods.

This software was started within the ANR MARMOTE project: ANR-12-MONU-00019.

URL: <https://gitlab.inria.fr/PSI/marmotecore/>

Publications: [hal-01651940](#), [hal-01276456](#)

Contact: Alain Jean-Marie

Participants: Alain Jean-Marie, Hlib Mykhailenko, Benjamin Briot, Franck Quessette, Issam Rabhi, Jean-Marc Vincent, Jean-Michel Fourneau

Partners: Université de Versailles St-Quentin-en-Yvelines, Université Paris Nanterre

7.1.4 MarTO

Name: Markov Toolkit for Markov models simulation: perfect sampling and Monte Carlo simulation

Keywords: Perfect sampling, Markov model

Functional Description: MarTO is a simulation software of markovian models. It is able to simulate discrete and continuous time models to provide a perfect sampling of the stationary distribution or directly a sampling of functional of this distribution by using coupling from the past. The simulation kernel is based on the CFTP algorithm, and the internal simulation of transitions on the Aliasing method. This software is a rewrite, more efficient and flexible, of PSI

URL: <https://gitlab.inria.fr/MarTo/marto>

Contact: Vincent Danjean

7.1.5 GameSeer

Keyword: Game theory

Functional Description: GameSeer is a tool for students and researchers in game theory that uses Mathematica to generate phase portraits for normal form games under a variety of (user-customizable) evolutionary dynamics. The whole point behind GameSeer is to provide a dynamic graphical interface that allows the user to employ Mathematica's vast numerical capabilities from a simple and intuitive front-end. So, even if you've never used Mathematica before, you should be able to generate fully editable and customizable portraits quickly and painlessly.

URL: <http://polaris.imag.fr/panayotis.mertikopoulos/publications/s01-gameseer/>

Contact: Panayotis Mertikopoulos

7.1.6 `rmf_tool`

Name: A library to Compute (Refined) Mean Field Approximations

Keyword: Mean Field

Functional Description: The tool accepts three model types:

- homogeneous population processes (HomPP)
- density dependent population processes (DDPPs)
- heterogeneous population models (HetPP)

In particular, it provides a numerical algorithm to compute the constant of the refined mean field approximation provided in the paper "A Refined Mean Field Approximation" by N. Gast and B. Van Houdt, SIGMETRICS 2018, and a framework to compute heterogeneous mean field approximations as proposed in "Mean Field and Refined Mean Field Approximations for Heterogeneous Systems: It Works!" by N. Gast and S. Allmeier, SIGMETRICS 2022.

URL: https://github.com/ngast/rmf_tool

Publications: [hal-01622054](#), [tel-02509756](#), [hal-03485044](#)

Contact: Nicolas Gast

7.1.7 `SCIP`

Name: Solving Constraint Integer Programs

Keywords: Linear optimization, Mathematical Optimization, Mixed Integer Programming

Functional Description: SCIP is currently one of the fastest non-commercial solvers for mixed integer programming (MIP) and mixed integer nonlinear programming (MINLP). It is also a framework for constraint integer programming and branch-cut-and-price. It allows for total control of the solution process and the access of detailed information down to the guts of the solver.

Release Contributions: SCIP 9.2.0 Features

added check for absolute and relative gap limits at end of synchronization in concurrent solving mode, in order to terminate earlier, note that if the concurrent solve is stopped due to a gap limit, the "winner" solver will have an interrupted solution status and its primal and dual bounds may not be the best possible ones (use `SCIPgetConcurrentPrimalbound()` and `SCIPgetConcurrentDualbound()` instead) parse pseudoboolean constraint from CIP format (and add linear-"and"-reformulation)

Performance improvements

reoptimization now also stores propagations from propagators if reoptimization/saveconsprop is enabled, the parameter will be renamed to reoptimization/saveprop in a next major release imposed stricter limits on the size of disconnected components which may be solved separately during presolve use individual slack variables also for constraints indicated by a common binary variable to use tighter formulation by default when computing symmetries using Nauty, iteration limits are now available to terminate Nauty early

Fixed bugs

Benders' decomposition subproblems that are always infeasible are correctly handled and the complete problem is declared infeasible skip linear constraint propagation if the residual activity bound cancels the side precision correct bound tracking to make the evaluation of primal-dual-integrals work skip aggregations on fixed variables in milp presolver to avoid errors for earlier versions of PaPILO use indices of negation counterparts and accept fixings when ordering and-resultants of pseudoboolean constraints update locks on model variables before removing pseudoboolean constraints in presolve reformulate soft pseudoboolean constraints with linear constraints if the indicator decomposition is disabled add workaround for recent HiGHS versions resetting the model

status when changing the presolve option after a solve avoid hashmap key error in removal of doubletons and singletons in dual presolve of setppc constraints by updating constraints and corresponding hashmaps after each multi-aggregation

Interface changes New API functions

added SCIPtpiIsAvailable() to check whether a working task processing interface is available (TPI != none) added SCIPtpiGetLibraryName() and SCIPtpiGetLibraryDesc() SCIPdelCons() can now also be called in SCIP_STAGE_TRANSFORMED added SCIPstrcasecmp() and SCIPstrncasecmp() for case-insensitive string comparison added SCIPbendersSubproblemsAreInfeasible() to return if at least one subproblem has been identified as being infeasible prior to performing any variable fixing

New parameters

presolving/milp/abortfaceexhaustive to control the abort threshold for exhaustive presolving in PAPILO presolving/milp/abortfacmedium to control the abort threshold for medium presolving in PAPILO presolving/milp/abortfacfast to control the abort threshold for fast presolving in PAPILO constraints/components/maxcompweight to determine the maximum weight for a disconnected component that is solved during presolve constraints/components/confactor counts the contributing factor of a single continuous variables with respect to the weight limit specified by constraints/components/maxcompweight constraints/indicator/usesameslackvar to decide whether the same slack variable should be used for indicators constraints with common binary variable propagating/symmetry/nautymaxncells and propagating/symmetry/nautymaxnnodes to set iteration limits in Nauty (only available if build with SYM=nauty or SYM=snauty)

Changed parameters

presolving/milp/threads is now only available if PaPILO is built with TBB changed default of numerics/recomputejac to 1e+6 to aim at relative epsilon precision

Build system Cmake

attempted to fix detection of CPLEX library on macOS and Windows systems

Testing

added parameter FILTER for tests/Makefile to run only tests with a specific pattern (ctest with -R FILTER)

Miscellaneous

adjusted Gurobi interface for Gurobi 12 reordered events: BESTSOLFOUND/NODE_FEASIBLE are now processed before the nodes are cut off and before NODE_DELETE events are processed removed #define of getcwd (in case of Windows builds) in scip/def.h the #define of strcasecmp and strncasecmp (in case of Windows builds) in scip/def.h will be removed with SCIP 10, use SCIPstr(n)casecmp() (scip/pub_misc.h) instead

URL: <https://www.scipopt.org/>

Contact: Mathieu Besancon

Partners: TU Darmstadt, RWTH Aachen University, Friedrich-Alexander-Universität Erlangen-Nürnberg, Eindhoven University of Technology, University of Twente, University of Bayreuth, Forschungscampus Modal

8 New results

The new results produced by the team in 2024 can be grouped into the following categories.

8.1 Scheduling in Data Centers

Participants: Jonatha Anselmi, Bruno Gaujal, Nicolas Gast.

Queuing theory is a general modeling framework, originally developed by Erlang to model the system of calls at the Copenhagen Telephone Exchange Company, and which has later been extensively used to optimize telecommunications, traffic, the design of factories and shops, etc. It is particularly suited to the modeling and optimization of the operation of data-centers, clouds or HPC centers and has led to the development of a variety of effective and low-cost scheduling strategies. We contribute to this framework and extend it in relation with recent online learning techniques as well as with characteristics of modern workload.

Dispatching and scheduling multiserver jobs for throughput optimality In [2], we consider the problem of dispatching and scheduling an infinite stream of multiple classes of jobs to a set of single server parallel queues. Each job requires the simultaneous utilization of multiple servers. Our objective is to identify a dispatching algorithm (used by a central dispatcher) and a scheduling discipline (used by each server) that induces throughput optimality, i.e., the mean response time of jobs is finite whenever the system load is less than one. It is known that this problem is not trivial even in the case where all servers share a centralized queue, i.e., no job dispatching. We show that throughput optimality can be obtained by dispatching jobs to queues according to probabilities provided that i) jobs of a given class respect some constraint on the server geometry and ii) the scheduling discipline prioritizes jobs with the largest server need. Due to a connection with the $M/G/1$ priority queue, we provide an exact analytical expression for the mean response time of jobs. Finally, we discuss the connection of our model with redundancy systems, where jobs do not necessarily require the simultaneous utilization of servers.

Load Balancing with Job-Size Testing: Performance Improvement or Degradation? In the context of decision making under explorable uncertainty, scheduling with testing is a powerful technique used in the management of computer systems to improve performance via better job-dispatching decisions. Upon job arrival, a scheduler may run some testing algorithm against the job to extract some information about its structure, e.g., its size, and properly classify it. The acquisition of such knowledge comes with a cost because the testing algorithm delays the dispatching decisions, though this is under control. In [4], we analyze the impact of such extra cost in a load balancing setting by investigating the following questions: does it really pay off to test jobs? If so, under which conditions? Under mild assumptions connecting the information extracted by the testing algorithm in relationship with its running time, we show that whether scheduling with testing brings a performance degradation or improvement strongly depends on the traffic conditions, system size and the coefficient of variation of job sizes. Thus, the general answer to the above questions is non-trivial and some care should be considered when deploying a testing policy. Our results are achieved by proposing a load balancing model for scheduling with testing that we analyze in two limiting regimes. When the number of servers grows to infinity in proportion to the network demand, we show that job-size testing actually degrades performance unless short jobs can be predicted reliably almost instantaneously and the network load is sufficiently high. When the coefficient of variation of job sizes grows to infinity, we construct testing policies inducing an arbitrarily large performance gain with respect to running jobs untested.

Balanced Splitting: A Framework for Achieving Zero-wait in the Multiserver-job Model In [3], we present a new framework for designing nonpreemptive and job-size oblivious scheduling policies in the multiserver-job queueing model. The main requirement is to identify a static and balanced sub-partition of the server set and ensure that the servers in each set of that sub-partition can only handle jobs of a given class and in a first-come first-served order. A job class is determined by the number of servers to which it has exclusive access during its entire execution and the probability distribution of its service time. This approach aims to reduce delays by preventing small jobs from being blocked by larger ones that arrived first, and it is particularly beneficial when the job size variability intra resp. inter classes is small resp. large. In this setting, we propose a new scheduling policy, Balanced-Splitting. In our main results, we provide a sufficient condition for the stability of Balanced-Splitting and show that the resulting

queueing probability, i.e., the probability that an arriving job needs to wait for processing upon arrival, vanishes in both the subcritical (the load is kept fixed to a constant less than one) and critical (the load approaches one from below) many-server limiting regimes. Crucial to our analysis is a connection with the $M/GI/s/s$ queue and Erlang's loss formula, which allows our analysis to rely on fundamental results from queueing theory. Numerical simulations show that the proposed policy performs better than several preemptive/nonpreemptive size-aware/oblivious policies in various practical scenarios. This is also confirmed by simulations running on real traces from High Performance Computing (HPC) workloads. The delays induced by Balanced-Splitting are also competitive with those induced by state-of-the-art policies such as First-Fit-SRPT and ServerFilling-SRPT, though our approach has the advantage of not requiring preemption, nor the knowledge of job sizes.

Asynchronous Load Balancing and Auto-scaling: Mean-Field Limit and Optimal Design In [1], we develop a Markovian framework for load balancing that combines classical algorithms such as Power-of- d with auto-scaling mechanisms that allow the net service capacity to scale up or down in response to the current load on the same timescale as job dynamics. Our framework is inspired by serverless platforms, such as Knative, where servers are software functions that can be flexibly instantiated in milliseconds according to scaling rules defined by the users of the serverless platform. The main question is how to design such scaling rules to minimize user-perceived delay performance while ensuring low energy consumption. For the first time, we investigate this problem when the auto-scaling and load balancing processes operate asynchronously (or proactively), as in Knative. In contrast to the synchronous (or reactive) paradigm, asynchronism brings the advantage that jobs do not necessarily need to wait any time a scale-up decision is taken. In our main result, we find a general condition on the structure of scaling rules able to drive mean-field dynamics to delay and relative energy optimality, i.e., a situation where both the user-perceived delay and the relative energy waste induced by idle servers vanish in the limit where the network demand grows to infinity in proportion to the nominal service capacity. The identified condition suggests to scale up the current net capacity if and only if the mean demand exceeds the rate at which servers become idle and active. Finally, we propose a family of scaling rules that satisfy our optimality condition. Numerical simulations demonstrate that these rules provide better delay performance than existing synchronous auto-scaling schemes while inducing almost the same power consumption.

A Stochastic Approach for Scheduling AI Training Jobs in GPU-based Systems In [10], we optimize the scheduling of Deep Learning (DL) training jobs from the perspective of a Cloud Service Provider running a data center, which efficiently selects resources for the execution of each job to minimize the average energy consumption while satisfying time constraints. To model the problem, we first develop a Mixed-Integer Non-Linear Programming formulation. Unfortunately, the computation of an optimal solution is prohibitively expensive, and to overcome this difficulty, we design a heuristic STochastic Scheduler (STS). Exploiting the probability distribution of early termination, STS determines how to adapt the resource assignment during the execution of the jobs to minimize the expected energy cost while meeting the job due dates. The results of an extensive experimental evaluation show that STS guarantees significantly better results than other methods in the literature, effectively avoiding due date violations and yielding a percentage total cost reduction between 32% and 80% on average. We also prove the applicability of our method in real-world scenarios, as obtaining optimal schedules for systems of up to 100 nodes and 400 concurrent jobs requires less than 5 seconds. Finally, we evaluated the effectiveness of GPU sharing, i.e., running multiple jobs in a single GPU. The obtained results demonstrate that depending on the workload and GPU memory, this further reduces the energy cost by 17-29% on average.

8.2 Performance evaluation of Large Systems (Mean Field)

Participants: Sebastian Allmeier, Nicolas Gast.

As a system of stochastically interacting entities grows large, the size of its state-space quickly explodes (curse of dimensionality) and both its analysis and optimization become intractable. Fortunately, sym-

metries and regularities can be exploited, in particular through the so-called Mean Field approximation, which averages out the state over degrees of freedom. This technique from statistical physics is particularly effective when studying computer systems and over the years, we have refined such approximation, proposed extensions, and developed fine analysis methods.

Mean Field Methods for Heterogeneous and Coupled Systems Mean Field methods have consistently been of interest to the scientific community due to their capacity to approximate a wide range of stochastic population systems. This methodology has proven to be a cornerstone in understanding and predicting the behavior of large-scale, complex systems. The key idea of the mean field approximation is to replace the complex stochastic behavior of systems with a simpler deterministic counterpart. The approximation therefore assumes that individuals become increasingly independent for large system sizes. The behavior of the system is thus obtained by replacing individual interactions with the average state of individuals. Despite its longstanding application and the advancements made in various fields, numerous questions and challenges remain open. In the scope of the PhD thesis of Sebastian Allmeier [29], we present our contributions and advancements for two general types of population models (1) heterogeneous mean field models and (2) mean field models with fast-changing environment. In the first part of the thesis, we focus on stochastic systems with heterogeneous components. We consider two types of heterogeneity, individual-level diversity as well as graph-structured interactions. For both cases, we provide accuracy bounds for the expected state of finite-sized systems. The results are supported through practical examples, including cache replacement policies, supermarket models, load balancing, and bike-sharing systems, highlighting their computational tractability and precision. In the case of individual-level diversity, we further adapt the refined mean field idea and show that the refined approximation significantly reduces the error and provides accurate predictions for small to moderate-sized systems. In the second part of the thesis, we turn our focus to mean field approximation techniques for stochastic systems with a coupled, fast-changing environment. By studying the ‘average’ mean field approximation, we obtain accuracy bounds for the expected system state. Furthermore, we derive a bias refinement term, which increases the accuracy of the approximation. Expanding on these results, we extend the methodology to stochastic approximation with constant step size and state-dependent Markovian noise. We show how to adapt the ideas to obtain accuracy results and a computable bias extension.

Computing the Bias of Constant-step Stochastic Approximation with Markovian Noise This work has been published in several conferences, including [19], where we study stochastic approximation algorithms with Markovian noise and constant step-size α . We develop a method based on infinitesimal generator comparisons to study the bias of the algorithm, which is the expected difference between θ_n — the value at iteration n — and θ^* — the unique equilibrium of the corresponding ODE. We show that, under some smoothness conditions, this bias is of order $O(\alpha)$. Furthermore, we show that the time-averaged bias is equal to $\alpha V + O(\alpha^2)$, where V is a constant characterized by a Lyapunov equation, showing that $E[\bar{\theta}_n] \approx \theta^* + V\alpha + O(\alpha^2)$, where $\bar{\theta}_n = (1/n)\sum_{k=1}^n \theta_k$ is the Polyak-Ruppert average. We also show that $\bar{\theta}_n$ converges with high probability around $\theta^* + \alpha V$. We illustrate how to combine this with Richardson-Romberg extrapolation to derive an iterative scheme with a bias of order $O(\alpha^2)$.

Approximations to Study the Impact of the Service Discipline in Systems with Redundancy Last, we have applied this technique to the study of queuing systems and difficult-to-study scheduling policies in cloud infrastructures, for example in [11]. As job redundancy has been recognized as an effective means to improve performance of large-scale computer systems, queueing systems with redundancy have been studied by various authors. Existing results include methods to compute the queue length distribution and response time but only when the service discipline is First-Come-First-Served (FCFS). For other service disciplines, such as Processor Sharing (PS), or Last-Come-First-Served (LCFS), only the stability conditions are known. In this paper we develop the first methods to approximate the queue length distribution in a queueing system with redundancy under various service disciplines. We focus on a system with exponential job sizes, i.i.d. copies, and a large number of servers. We first derive a mean field approximation that is independent of the scheduling policy. In order to study the impact of service discipline, we then derive refinements of this approximation to specific scheduling policies. In the case of Processor Sharing, we provide a pair and a triplet approximation. The pair approximation can

be regarded as a refinement of the classic mean field approximation and takes the service discipline into account, while the triplet approximation further refines the pair approximation. We also develop a pair approximation for three other service disciplines: First-Come-First-Served, Limited Processor Sharing and Last-Come-First-Served. We present numerical evidence that shows that all the approximations presented in the paper are highly accurate, but that none of them are asymptotically exact (as the number of servers goes to infinity). This makes these approximations suitable to study the impact of the service discipline on the queue length distribution. Our results show that FCFS yields the shortest queue length, and that the differences are more substantial at higher loads.

8.3 Reinforcement Learning and MDP

Participants: Victor Boone, Nicolas Gast, Bruno Gaujal, Chen Yan.

A limitation of the classical queuing theory framework is that policies may require knowledge of the system parameters whereas such parameters are rarely known in advance, hence our interest in reinforcement learning technique and Markov Decision Processes. Over the years, we have developed analysis and modeling techniques specifically suited to queuing systems whose state space quickly explodes (learning in queues), which makes classical RL approaches or results inappropriate.

An MDP-Based Solution for the Energy Minimization of Non-Clairvoyant Hard Real-Time Systems

In [12], we address the problem of scheduling a possibly infinite sequence of hard real-time jobs on a single-core processor, with the dual goal that (1) all the jobs must finish before their deadline, and (2) the energy consumption must be minimized. The decision variable is the speed of the processor at each time instant, to be chosen from a finite set of available speeds. Our goal is to design a speed policy in charge of deciding, at each time instant, the speed of the processor in function of the job characteristics. We focus more specifically on the non-clairvoyant case, meaning that the actual size of the jobs (the amount of work to be done to complete the job) is unknown when they are released, but its probability distribution is known, and of course, the maximal size is known too. In this context, we propose two new speed policies, the optimal solution of a Markov Decision Process (called MOSP), and a heuristic speed policy called Expected Load (EL), obtained by adapting the classical policy Optimal Available (OA) to jobs with random sizes. Our MOSP algorithm is split in two phases: the first phase is offline — it computes the optimal processor speed for each possible system state — while the second phase is online — it retrieves the speed to apply to the processor thanks to a table lookup. Compared with the existing speed policies from the literature, MOSP achieves the optimal energy consumption but at the cost of a significant state space size. In contrast, EL achieves an energy consumption that is, on average, close to the optimal one obtained with MOSP, but at almost no cost in terms of state space.

Learning Optimal Admission Control in Partially Observable Queuing Networks In [5], we present an efficient reinforcement learning algorithm that learns the optimal admission control policy in a partially observable queueing network. Specifically, only the arrival and departure times from the network are observable, and optimality refers to the average holding/rejection cost in infinite horizon. While reinforcement learning in Partially Observable Markov Decision Processes (POMDP) is prohibitively expensive in general, we show that our algorithm has a regret that only depends sub-linearly on the maximal number of jobs in the network, S . In particular, in contrast with existing regret analyses, our regret bound does not depend on the diameter of the underlying Markov Decision Process (MDP), which in most queueing systems is at least exponential in S . The novelty of our approach is to leverage Norton's equivalent theorem for closed product-form queueing networks and an efficient reinforcement learning algorithm for MDPs with the structure of birth-and-death processes.

Reoptimization Nearly Solves Weakly Coupled Markov Decision Processes In [38], we propose a new policy, called the LP-update policy, to solve finite horizon weakly-coupled Markov decision processes. The latter can be seen as multi-constraint multi-action bandits, and generalize the classical restless bandit

problems. Our solution is based on re-solving periodically a relaxed version of the original problem, that can be cast as a linear program (LP). When the problem is made of N statistically identical sub-components, we show that the LP-update policy becomes asymptotically optimal at rate $O(T^2/\sqrt{N})$. This rate can be improved to $O(T/\sqrt{N})$ if the problem satisfies some ergodicity property and to $O(1/N)$ if the problem is non-degenerate. The definition of non-degeneracy extends the same notion for restless bandits. By using this property, we also improve the computational efficiency of the LP-update policy. We illustrate the performance of our policy on randomly generated examples, as well as a generalized applicant screening problem, and show that it outperforms existing heuristics.

Optimal Regrets in Markov Decision Processes Last although *regret* is a common objective in Reinforcement Learning, other criteria are relevant and allow to better understand or discriminate algorithms. The thesis of Victor Boone [30], is dedicated to the theory of learning of Markov decision processes for the average reward criterion under the regret benchmark. Two standard formulations of the learning problem are investigated: the minimax and the model dependent settings. For both, asymptotic lower bounds of the regret are provided together with algorithms reaching them. Beyond the regret benchmark, the local behavior of classical algorithms is further studied and new learning metrics are designed for that matter.

8.4 Optimization Techniques

Participants: Mathieu Besançon, Panayotis Mertikopoulos.

Optimization arises in almost every domain of computer science and requires a variety of techniques, to which we contribute both from a theoretical and practical perspective.

8.4.1 Mixed Integer Optimization

Branching via Cutting Plane Selection: Improving Hybrid Branching Cutting planes and branching are two of the most important algorithms for solving mixed-integer linear programs. For both algorithms, disjunctions play an important role, being used both as branching candidates and as the foundation for some cutting planes. In [42], we relate branching decisions and cutting planes to each other through the underlying disjunctions that they are based on, with a focus on Gomory mixed-integer cuts and their corresponding split disjunctions. We show that selecting branching decisions based on quality measures of Gomory mixed-integer cuts leads to relatively small branch-and-bound trees, and that the result improves when using cuts that more accurately represent the branching decisions. Finally, we show how the history of previously computed Gomory mixed-integer cuts can be used to improve the performance of the state-of-the-art hybrid branching rule of SCIP. Our results show a 4% decrease in solve time, and an 8% decrease in number of nodes over affected instances of MIPLIB 2017.

Probabilistic Lookahead Strong Branching via a Stochastic Abstract Branching Model Strong Branching (SB) is a cornerstone of all modern branching rules used in the Branch-and-Bound (BnB) algorithm, which is at the center of Mixed-Integer Programming solvers. In its full form, SB evaluates all variables to branch on and then selects the one producing the best relaxation, leading to small trees, but high runtimes. State-of-the-art branching rules therefore use SB with working limits to achieve both small enough trees and short run times. So far, these working limits have been established empirically. In [26], we introduce a theoretical approach to guide how much SB to use at each node within the BnB. We first define an abstract stochastic tree model of the BnB algorithm where the geometric mean dual gains of all variables follow a given probability distribution. This model allows us to relate expected dual gains to tree sizes and explicitly compare the cost of sampling an additional SB candidate with the reward in expected tree size reduction. We then leverage the insight from the abstract model to design a new stopping criterion for SB, which fits a distribution to the dual gains and, at each node, dynamically continues or interrupts SB. This algorithm, which we refer to as Probabilistic Lookahead Strong Branching, improves

both the tree size and runtime over MIPLIB instances, providing evidence that the method not only changes the amount of SB, but allocates it better.

The SCIP Optimization Suite 9.0 The SCIP Optimization Suite provides a collection of software packages for mathematical optimization, centered around the constraint integer programming (CIP) framework SCIP. In [34], we discuss the enhancements and extensions included in the SCIP Optimization Suite 9.0. The updates in SCIP 9.0 include improved symmetry handling, additions and improvements of nonlinear handlers and primal heuristics, a new cut generator and two new cut selection schemes, a new branching rule, a new LP interface, and several bug fixes. The SCIP Optimization Suite 9.0 also features new Rust and C++ interfaces for SCIP, new Python interface for SoPlex, along with enhancements to existing interfaces. The SCIP Optimization Suite 9.0 also includes new and improved features in the LP solver SoPlex, the presolving library PaPILO, the parallel framework UG, the decomposition framework GCG, and the SCIP extension SCIP-SDP. These additions and enhancements have resulted in an overall performance improvement of SCIP in terms of solving time, number of nodes in the branch-and-bound tree, as well as the reliability of the solver.

8.4.2 Application to various domains

Solving the Optimal Experiment Design Problem with Mixed-Integer Convex Methods In [28], we tackle the Optimal Experiment Design Problem, which consists of choosing experiments to run or observations to select from a finite set to estimate the parameters of a system. The objective is to maximize some measure of information gained about the system from the observations, leading to a convex integer optimization problem. We leverage Boscia.jl, a recent algorithmic framework, which is based on a nonlinear branch-and-bound algorithm with node relaxations solved to approximate optimality using Frank-Wolfe algorithms. One particular advantage of the method is its efficient utilization of the polytope formed by the original constraints which is preserved by the method, unlike alternative methods relying on epigraph-based formulations. We assess our method against both generic and specialized convex mixed-integer approaches. Computational results highlight the performance of our proposed method, especially on large and challenging instances.

Network Design for the Traffic Assignment Problem with Mixed-Integer Frank-Wolfe In [27], we tackle the network design problem for centralized traffic assignment, which can be cast as a mixed-integer convex optimization (MICO) problem. For this task, we propose different formulations and solution methods in both a deterministic and a stochastic setting in which the demand is unknown in the design phase. We leverage the recently proposed Boscia framework, which can solve MICO problems when the main nonlinearity stems from a differentiable objective function. Boscia tackles these problems by branch-and-bound with continuous relaxations solved approximately with Frank-Wolfe algorithms. We compare different linear relaxations and the corresponding subproblems solved by Frank-Wolfe, and alternative problem formulations to identify the situations in which each performs best. Our experiments evaluate the different approaches on instances from the Transportation Networks library and highlight the suitability of the mixed-integer Frank-Wolfe algorithm for this problem. In particular, we find that the Boscia framework is particularly applicable to this problem and that a mixed-integer linear Frank-Wolfe subproblem performs well for the deterministic case, while a penalty-based approach, with decoupled feasible regions for the design and flow variables, dominates other approaches for stochastic instances with many scenarios.

Optimisation models for the design of multiple self-consumption loops in semi-rural areas Collective electricity self-consumption gains increasing interest in a context where localised consumption of energy is a lever of sustainable development. While easing energy distribution networks, collective self-consumption requires complementary prosumers' profiles. Before determining the proper energy exchanges happening between these prosumers, one must ensure their compatibility in the context of collective self-consumption. Motivated by real use cases from a solar power producer, we propose in [36] network flow-based linear models to solve both the design aspect of the problem and the attribution of distribution flows between the involved prosumers. Two main models are proposed, handling (i) single

collective self-consumption loop creation and (ii) multiple loop creation. One of the objectives of this work is to provide models that can be applied at a strategic level which implies their extension to large time scale and large spatial scale. To do so, associated Benders and Dantzig-Wolfe decompositions are proposed to ensure model scalability along temporal and spatial dimensions. The proposed models are tested on realistic use cases to assess their performances.

8.4.3 Optimization for Machine Learning

Many learning algorithms operate in centralized way, which raises many practical issues in terms of scalability, privacy, hence a high interest for designing efficient distributed and federated machine learning algorithms. Furthermore generally, the optimization space is also quite particular, which calls for specific regularization techniques and optimization algorithms.

What is the long-run distribution of stochastic gradient descent? A large deviations analysis In [20], we examine the long-run distribution of stochastic gradient descent (SGD) in general, non-convex problems. Specifically, we seek to understand which regions of the problem's state space are more likely to be visited by SGD, and by how much. Using an approach based on the theory of large deviations and randomly perturbed dynamical systems, we show that the long-run distribution of SGD resembles the Boltzmann-Gibbs distribution of equilibrium thermodynamics with temperature equal to the method's step-size and energy levels determined by the problem's objective and the statistics of the noise. In particular, we show that, in the long run, (a) the problem's critical region is visited exponentially more often than any non-critical region; (b) the iterates of SGD are exponentially concentrated around the problem's minimum energy state (which does not always coincide with the global minimum of the objective); (c) all other connected components of critical points are visited with frequency that is exponentially proportional to their energy level; and, finally (d) any component of local maximizers or saddle points is "dominated" by a component of local minimizers which is visited exponentially more often.

The rate of convergence of Bregman proximal methods: Local geometry vs. regularity vs. sharpness In [6], we examine the last-iterate convergence rate of Bregman proximal methods - from mirror descent to mirror-prox and its optimistic variants - as a function of the local geometry induced by the prox-mapping defining the method. For generality, we focus on local solutions of constrained, non-monotone variational inequalities, and we show that the convergence rate of a given method depends sharply on its associated Legendre exponent, a notion that measures the growth rate of the underlying Bregman function (Euclidean, entropic, or other) near a solution. In particular, we show that boundary solutions exhibit a stark separation of regimes between methods with a zero and non-zero Legendre exponent: the former converge at a linear rate, while the latter converge, in general, sublinearly. This dichotomy becomes even more pronounced in linearly constrained problems where methods with entropic regularization achieve a linear convergence rate along sharp directions, compared to convergence in a finite number of steps under Euclidean regularization.

Tamed Langevin sampling under weaker conditions Motivated by applications to deep learning which often fail standard Lipschitz smoothness requirements, we examine in [41] the problem of sampling from distributions that are not log-concave and are only weakly dissipative, with log-gradients allowed to grow superlinearly at infinity. In terms of structure, we only assume that the target distribution satisfies either a log-Sobolev or a Poincaré inequality and a local Lipschitz smoothness assumption with modulus growing possibly polynomially at infinity. This set of assumptions greatly exceeds the operational limits of the "vanilla" unadjusted Langevin algorithm (ULA), making sampling from such distributions a highly involved affair. To account for this, we introduce a taming scheme which is tailored to the growth and decay properties of the target distribution, and we provide explicit non-asymptotic guarantees for the proposed sampler in terms of the Kullback-Leibler (KL) divergence, total variation, and Wasserstein distance to the target distribution.

Robust bilevel optimization for near-optimal lower-level solutions Abstract Bilevel optimization problems embed the optimality of a subproblem as a constraint of another optimization problem. We

introduce in [7] the concept of near-optimality robustness for bilevel optimization, protecting the upper-level solution feasibility from limited deviations from the optimal solution at the lower level. General properties and necessary conditions for the existence of solutions are derived for near-optimal robust versions of general bilevel optimization problems. A duality-based solution method is defined when the lower level is convex, leveraging the methodology from the robust and bilevel literature. Numerical results assess the efficiency of exact and heuristic methods and the impact of valid inequalities on the solution time.

The computational complexity of finding second-order stationary points Non-convex minimization problems are universally considered hard, and even guaranteeing that a computed solution is locally minimizing is known to be NP-hard. In this general context, we focus in [21] on the problem of finding stationary points that satisfy an approximate second-order optimality condition, which serves to exclude strict saddles and other non-minimizing stationary points. Our main result is that the problem of finding approximate second-order stationary points (SOSPs) is PLS-complete, i.e., of the same complexity as the problem of finding first-order stationary points (FOSPs), thus resolving an open question in the field. In particular, our results imply that, under the widely believed complexity conjecture that $\text{PLS} \neq \text{FNP}$, finding approximate SOSPs in unconstrained domains is easier than in constrained domains, which is known to be NP-hard. This comes in stark contrast with earlier results which implied that, unless $\text{PLS} = \text{CLS}$, finding approximate FOSPs in unconstrained domains is harder than in constrained domains.

Explicit second-order min-max optimization methods with optimal convergence guarantees In [40], we propose and analyze several inexact regularized Newton-type methods for finding a global saddle point of *convex-concave* unconstrained min-max optimization problems. Compared to first-order methods, our understanding of second-order methods for min-max optimization is relatively limited, as obtaining global rates of convergence with second-order information is much more involved. In this paper, we examine how second-order information can be used to speed up extra-gradient methods, even under inexactness. Specifically, we show that the proposed methods generate iterates that remain within a bounded set and that the averaged iterates converge to an ϵ -saddle point within $O(\epsilon^{-2/3})$ iterations in terms of a restricted gap function. This matched the theoretically established lower bound in this context. We also provide a simple routine for solving the subproblem at each iteration, requiring a single Schur decomposition and $O(\log \log(1/\epsilon))$ calls to a linear system solver in a quasi-upper-triangular system. Thus, our method improves the existing line-search-based second-order min-max optimization methods by shaving off an $O(\log \log(1/\epsilon))$ factor in the required number of Schur decompositions. Finally, we present numerical experiments on synthetic and real data that demonstrate the efficiency of the proposed methods.

8.5 Learning in games

Participants: Davide Legacci, Panayotis Mertikopoulos, Bary Pradelski.

Learning in games naturally occurs in situations where the resources or the decision is distributed among several agents or even in situations where a centralized decision maker has to adapt to strategic users. Yet, it is considerably more difficult than in classical minimization games as the resulting equilibria may be attractive or not and the dynamic often exhibit cyclic behaviors. Understanding and characterizing the geometry of such spaces is thus the key to propose efficient algorithms.

A unified stochastic approximation framework for learning in games In [14], we develop a flexible stochastic approximation framework for analyzing the long-run behavior of learning in games (both continuous and finite). The proposed analysis template incorporates a wide array of popular learning algorithms, including gradient-based methods, the exponential / multiplicative weights algorithm for learning in finite games, optimistic and bandit variants of the above, etc. In addition to providing an integrated view of these algorithms, our framework further allows us to obtain several new convergence

results, both asymptotic and in finite time, in both continuous and finite games. Specifically, we provide a range of criteria for identifying classes of Nash equilibria and sets of action profiles that are attracting with high probability, and we also introduce the notion of coherence, a game-theoretic property that includes strict and sharp equilibria, and which leads to convergence in finite time. Importantly, our analysis applies to both oracle-based and bandit, payoff-based methods — that is, when players only observe their realized payoffs.

On the discrete-time origins of the replicator dynamics: From convergence to instability and chaos

We consider in [37] three distinct discrete-time models of learning and evolution in games: a biological model based on intra-species selective pressure, the dynamics induced by pairwise proportional imitation, and the exponential / multiplicative weights (EW) algorithm for online learning. Even though these models share the same continuous-time limit — the replicator dynamics — we show that second-order effects play a crucial role and may lead to drastically different behaviors in each model, even in very simple, symmetric 2×2 games. Specifically, we study the resulting discrete-time dynamics in a class of parameterized congestion games, and we show that (i) in the biological model of intra-species competition, the dynamics remain convergent for any parameter value; (ii) the dynamics of pairwise proportional imitation exhibit an entire range of behaviors for larger time steps and different equilibrium configurations (stability, instability, and even Li-Yorke chaos); while (iii) in the EW algorithm, increasing the time step (almost) inevitably leads to chaos (again, in the formal, Li-Yorke sense). This divergence of behaviors comes in stark contrast to the globally convergent behavior of the replicator dynamics, and serves to delineate the extent to which the replicator dynamics provide a useful predictor for the long-run behavior of their discrete-time origins.

Nested replicator dynamics, nested logit choice, and similarity-based learning

We consider in [16] a model of learning and evolution in games whose action sets are endowed with a partition-based similarity structure intended to capture exogenous similarities between strategies. In this model, revising agents have a higher probability of comparing their current strategy with other strategies that they deem similar, and they switch to the observed strategy with probability proportional to its payoff excess. Because of this implicit bias toward similar strategies, the resulting dynamics - which we call the nested replicator dynamics - do not satisfy any of the standard monotonicity postulates for imitative game dynamics; nonetheless, we show that they retain the main long-run rationality properties of the replicator dynamics, albeit at quantitatively different rates. We also show that the induced dynamics can be viewed as a stimulus-response model in the spirit of Erev & Roth (1998), with choice probabilities given by the nested logit choice rule of Ben-Akiva (1973) and McFadden (1978). This result generalizes an existing relation between the replicator dynamics and the exponential weights algorithm in online learning, and provides an additional layer of interpretation to our analysis and results.

A geometric decomposition of finite games: Convergence vs. recurrence under exponential weights

In view of the complexity of the dynamics of learning in games, we seek to decompose a game into simpler components where the dynamics' long-run behavior is well understood. A natural starting point for this is Helmholtz's theorem, which decomposes a vector field into a potential and an incompressible component. However, the geometry of game dynamics - and, in particular, the dynamics of exponential / multiplicative weights (EW) schemes - is not compatible with the Euclidean underpinnings of Helmholtz's theorem. This leads us to consider in [16] a specific Riemannian framework based on the so-called Shahshahani metric, and introduce the class of incompressible games, for which we establish the following results: First, in addition to being volume-preserving, the continuous-time EW dynamics in incompressible games admit a constant of motion and are Poincaré recurrent — i.e., almost every trajectory of play comes arbitrarily close to its starting point infinitely often. Second, we establish a deep connection with a well-known decomposition of games into a potential and harmonic component (where the players' objectives are aligned and anti-aligned respectively): a game is incompressible if and only if it is harmonic, implying in turn that the EW dynamics lead to Poincaré recurrence in harmonic games.

No-regret learning in harmonic games: Extrapolation in the face of conflicting interests The long-run behavior of multi-agent learning — and, in particular, no-regret learning — is relatively well-understood

in potential games, where players have aligned interests. By contrast, in harmonic games — the strategic counterpart of potential games, where players have conflicting interests — very little is known outside the narrow subclass of 2-player zero-sum games with a fully-mixed equilibrium. In [23], we seek to partially fill this gap by focusing on the full class of (generalized) harmonic games and examining the convergence properties of follow-the-regularized-leader (FTRL), the most widely studied class of no-regret learning schemes. As a first result, we show that the continuous-time dynamics of FTRL are Poincaré recurrent, that is, they return arbitrarily close to their starting point infinitely often, and hence fail to converge. In discrete time, the standard, "vanilla" implementation of FTRL may lead to even worse outcomes, eventually trapping the players in a perpetual cycle of best-responses. However, if FTRL is augmented with a suitable extrapolation step - which includes as special cases the optimistic and mirror-prox variants of FTRL — we show that learning converges to a Nash equilibrium from any initial condition, and all players are guaranteed at most $O(1)$ regret. These results provide an in-depth understanding of no-regret learning in harmonic games, nesting prior work on 2-player zero-sum games, and showing at a high level that harmonic games are the canonical complement of potential games, not only from a strategic, but also from a dynamic viewpoint.

Accelerated regularized learning in finite N-person games Last, motivated by the success of Nesterov's accelerated gradient algorithm for convex minimization problems, we examine in [25] whether it is possible to achieve similar performance gains in the context of online learning in games. To that end, we introduce a family of accelerated learning methods, which we call "follow the accelerated leader" (FTXL), and which incorporates the use of momentum within the general framework of regularized learning — and, in particular, the exponential/multiplicative weights algorithm and its variants. Drawing inspiration and techniques from the continuous-time analysis of Nesterov's algorithm, we show that FTXL converges locally to strict Nash equilibria at a superlinear rate, achieving in this way an exponential speed-up over vanilla regularized learning methods (which, by comparison, converge to strict equilibria at a geometric, linear rate). Importantly, FTXL maintains its superlinear convergence rate in a broad range of feedback structures, from deterministic, full information models to stochastic, realization-based ones, and even when run with bandit, payoff-based information, where players are only able to observe their individual realized payoffs.

8.6 Random matrix analysis and Machine Learning

Participants: Romain Couillet, Hugo Lebeau, Victor Leger, Charles Sejourne.

Random matrix theory has recently proven to be a very effective tool to understand Machine Learning challenges. In particular, concentration results can be used to derive more efficient and frugal algorithms. This line of work has led to the defense of two PhD thesis in 2024.

Theoretical Guarantees and Improvement of Large Dimensional Multi-Task Semi-Supervised Classification In the field of machine learning, the specific subfield of deep learning has gathered particular interest in the last decade. If deep learning has allowed quick and significant progress in a wide range of fields, this progress has been made at the expense of interpretability, accessibility, robustness and flexibility, not to mention the colossal energy consumption implied by the training of such algorithms. On the contrary, the thesis of Victor Leger [32] aims to open the path to all-encompassing flexible tools for classification problems, buttressed on elementary machine learning notions and rather basic mathematical tools. This thesis conducts a large dimensional study of a simple yet quite versatile classification model, encompassing at once multi-task and semi-supervised learning, and taking into account uncertain labeling. Using tools from random matrix theory, the asymptotics of some key functionals are characterized, which allows on the one hand to predict the performances of the proposed algorithm, and on the other hand to reveal some counter-intuitive guidance on how to use it efficiently. The model, powerful enough to provide good performance guarantees, is also straightforward enough to provide strong insight into its behavior. The resulting algorithm is also compared to an optimal bound derived from statistical physics, which gives a lower bound of the least achievable probability of misclassification

for a given problem. This bound is computed in the extended case of uncertain labeling, and is used to evaluate the performances of the algorithm.

Theoretical performance analysis applied to non-convex and fair algorithms using Random Matrix Theory In the thesis of Charles Sejourne[33], we study theoretically the performance of well-known Machine Learning algorithms (t-SNE and LS-SVM). For the first algorithm, we study an asymptotic characterization of the critical points of the optimization problem of the cousin of t-SNE, Symmetric SNE, in a large-dimensional setting. For the second algorithm, we study theoretically the behavior of LS-SVM when adding fairness constraints aiming at making the algorithm more fair.

Performance Gaps in Multi-view Clustering under the Nested Matrix-Tensor Model We study in [22] the estimation of a planted signal hidden in a recently introduced nested matrix-tensor model, which is an extension of the classical spiked rank-one tensor model, motivated by multi-view clustering. Prior work has theoretically examined the performance of a tensor-based approach, which relies on finding a best rank-one approximation, a problem known to be computationally hard. A tractable alternative approach consists in computing instead the best rank-one (matrix) approximation of an unfolding of the observed tensor data, but its performance was hitherto unknown. We quantify here the performance gap between these two approaches, in particular by deriving the precise algorithmic threshold of the unfolding approach and demonstrating that it exhibits a BBP-type transition behavior. This work is therefore in line with recent contributions which deepen our understanding of why tensor-based methods surpass matrix-based methods in handling structured tensor data.

A Random Matrix Approach to Low-Multilinear-Rank Tensor Approximation In [39], we present a comprehensive understanding of the estimation of a planted low-rank signal from a general spiked tensor model near the computational threshold. Relying on standard tools from the theory of large random matrices, we characterize the large-dimensional spectral behavior of the unfoldings of the data tensor and exhibit relevant signal-to-noise ratios governing the detectability of the principal directions of the signal. These results allow to accurately predict the reconstruction performance of truncated multilinear SVD (MLSVD) in the non-trivial regime. This is particularly important since it serves as an initialization of the higher-order orthogonal iteration (HOOI) scheme, whose convergence to the best low-multilinear-rank approximation depends entirely on its initialization. We give a sufficient condition for the convergence of HOOI and show that the number of iterations before convergence tends to 1 in the large-dimensional limit.

Asymptotic Gaussian Fluctuations of Eigenvectors in Spectral Clustering The performance of spectral clustering relies on the fluctuations of the entries of the eigenvectors of a similarity matrix, which has been left uncharacterized until now. In [13], we show that the signal + noise structure of a general spike random matrix model is transferred to the eigenvectors of the corresponding Gram kernel matrix and the fluctuations of their entries are Gaussian in the large-dimensional regime. This CLT-like result was the last missing piece to precisely predict the classification performance of spectral clustering. The proposed proof is very general and relies solely on the rotational invariance of the noise. Numerical experiments on synthetic and real data illustrate the universality of this phenomenon.

8.7 Fairness and equity in digital (recommendation, advertising, persistent storage) systems

Participants: Rémi Castera, Nicolas Gast, Mathieu Molina, Bary Pradelski.

The general deployment of machine-learning systems in many domains ranging from security to recommendation and advertising to guide strategic decisions leads to an interesting line of research from a game theory perspective. In this context, fairness, discrimination, and privacy are particularly important issues.

Efficiency and Fairness in Matching Problems In the thesis of Rémi Castera [31], we study matching models, i.e., models based on a bipartite graph where the objective is to find a set of edges in the graph (called a matching) that satisfies or maximizes some objective. In the basic setup, the objective is to find the largest possible set of edges. When the vertices of the graph are agents with preferences over the possible partners, we look for matchings that are deemed stable, while trying to maximize the satisfaction of the agents. We assume that vertices on one side of the graph are divided into groups, those groups could represent different demographics, that we would like to be treated fairly when choosing a matching. In this thesis, we look into various definitions of fairness towards those groups (a concept called group fairness), and which parameters of each studied model play a role in the existence or the optimality of a fair matching. In particular, we study the trade-off between fairness and efficiency, defined as the size of the matching when there are no preferences, or the satisfaction of agents when there are preferences. In the setting with no preferences, we propose an original geometric representation of the problem that allows us to give conditions for the existence of a matching that is maximal and fair, and when it does not exist we provide tight bounds on the ratio between the size of the largest matching and the size of the largest fair matching (we call this ratio the Price of Fairness). In the setting with preferences, that we model as a college admission problem with students on one side and colleges on the other side, we study the role of the correlation between colleges' rankings of students. We show that correlation improves the efficiency of the stable matching. However when different groups have different levels of correlation, groups with higher correlation have higher rates of students remaining unassigned, even when each college's individual ranking is completely fair towards each group. We also show that when colleges' rankings are fair, there is no trade-off between efficiency and fairness.

Correlation of Rankings in Matching Markets The latest results have been submitted in [35], where we study the role of correlation in matching, where multiple decision-makers simultaneously face selection problems from the same pool of candidates. We propose a model in which decision-makers have varying information on candidates from different sociodemographic groups when evaluating and ranking them, thus leading to varying correlations among candidates' priority scores. Such differential correlation arises, for example, when the cost of information acquisition, decision-maker preferences, or the prevalence of selection criteria vary across sociodemographic groups. We show that a lower correlation for one of the groups worsens the outcome for all groups, thus leading to efficiency loss. Moreover, the proportion of students from a given group who remain unmatched is increasing in its own correlation level. This implies that it is advantageous to belong to the low-correlation group. Finally, we extend the extent tie-breaking literature to multiple priority classes and intermediate levels of correlation. Overall, our results point to a previously overlooked systemic source of group inequalities in school, university, and job admissions.

Quick or cheap? Breaking points in dynamic markets In [15], we examine two-sided markets where players arrive stochastically over time. The cost of matching a client and provider is heterogeneous, and the distribution of costs — but not their realization — is known. In this way, a social planner is faced with two contending objectives: (a) to reduce the players' waiting time before getting matched; and (b) to reduce matching costs. In this paper, we aim to understand when and how these objectives are incompatible. We identify two regimes dependent on the 'speed of improvement' of the cost of matching with respect to market size. One regime results in a quick or cheap dilemma without 'free lunch': there exists no clearing schedule that is simultaneously optimal along both objectives. In that regime, we identify a unique breaking point signifying a stark reduction in matching cost contrasted by an increase in waiting time. The other regime features a window of opportunity in which free lunch can be achieved. Which scheduling policy is optimal depends on the heterogeneity of match costs. Under limited heterogeneity, e.g., when there is a finite number of possible match costs, greedy scheduling is approximately optimal, in line with the related literature. However, with more heterogeneity greedy scheduling is never optimal. Finally, we analyze a particular model where match costs are exponentially distributed and show that it is at the boundary of the no-free-lunch regime. We then characterize the optimal clearing schedule for varying social planner desiderata.

The representation dynamic and the “normalization” of group differences Abstract Intergroup inequality has been linked to differing norms of economic participation among groups. In [8], we present a theory of endogenous identity-specific norms in which the larger a group’s representation in an economic activity, the more the activity is deemed “normal” or “appropriate” for its members. This representation dynamic can arise from behavioral heuristics or be created by informational technologies such as generative artificial intelligence. Through it, the economic underrepresentation of a group becomes “normalized,” resulting in more severe inequality than in standard models. Equality of opportunity almost never results in equal outcomes, even when groups have the same productivity. Minorities and historically marginalized groups tend to be underrepresented. However, minorities with greater productivity and/or weaker group identification can become overrepresented, and even dominant. When there are multiple career stages, underrepresentation can escalate at senior levels long after “glass ceilings” have disappeared. Underrepresentation disappears as economic returns rise and/or group identification weakens.

Intersectionality: Affirmative Action with Multidimensional Identities Studying the design of affirmative action policies when identities are multidimensional, we provide in [9] a formal demonstration of the importance of intersectionality. Prevailing affirmative action policies are based only on one identity dimension (e.g., race, gender, socioeconomic class). We find that any such nonintersectional policy can almost never achieve a representative outcome. In fact, nonintersectional policies often increase the underrepresentation of underrepresented groups in a manner undetected by standard measures. Examples based on race and gender reveal significant hidden inequality arising from nonintersectional policies. We show how to construct intersectional policies that achieve proportional representation.

8.8 Training Secondary School Teachers in Computer Science

Booléens en programmation : une analyse à destination des enseignants The notion of boolean is fundamental in computer science. Even if it seems simple at first sight, its introduction in the first stages of programming education reveals a number of difficulties related to the fact that it is linked to several general concepts of computer science: data types, truth values, invariants, control structures. . . Starting from this observation, we propose in [18] a reflection on the role of booleans in programming. We elaborate on some points concerning possible “good practices”, especially in the context of computer science teaching. In particular, we focus on boolean as a data type and on the main control structures using booleans. We also discuss the general notion of boolean expression used in programming. Finally through a few examples, we illustrate some typical uses of boolean variables.

Tele-vision: unplugged computing at the service of the concept of information Computer science teaching, from elementary school through to high school, is structured around the “four concepts or facets” of computing and digital objects: information, algorithms, programming languages and computing machines. The aim of the Tele-vision activity analyzed in [17] is to introduce the fundamental concepts of the “information” facet through a sequence of unplugged activities. The activity involves the transmission of images from one student to another, using two-colored tokens. Using the tokens, then the faces of a dice to materialize the information, students agree on a code, encoding, transmitting and decoding the information. They evaluate the entire transmission protocol (correctness, efficiency, etc.). A reflective analysis of experiments carried out in Cycle 3 (6th grade) is proposed to identify the knowledge taught and the skills involved. Finally, the activity is considered from a broader perspective (other levels or other concepts).

9 Bilateral contracts and grants with industry

Participants: Nicolas Gast.

Nicolas Gast participates to the "Defi EDF" and is currently supervising a PhD student (Hélène Arvis) via a "CIFRE" contract.

10 Partnerships and cooperations

10.1 International initiatives

10.1.1 Inria associate team not involved in an IIL or an international program

AIRBA

Title: AI for restless bandits and its application

Duration: 2023 -> 2025

Coordinator: Gupta Manu Kumar (manu.gupta@ms.iitr.ac.in)

Partners:

- Indian Institute of Technology Roorkee Roorkee (Inde)

Inria contact: Nicolas Gast

Summary: Multi-armed restless bandit problems (MARBPs) are Markov decision process models for optimal dynamic priority allocation to a collection of stochastic binary-action (active/passive) projects evolving over time. Typical applications include maintenance problems, in which a collection of agents must be sent to various objects subject to failures, or stochastic scheduling problems. If MARBPs are in general intractable, there exists efficient relaxation when the problems parameters are known. The goal of this project is to build on recent progress on Reinforcement Learning to create new tools to solve this problem when the parameters are unknown. Problems of interests are: how to define online indices and how to learn them; What is the performance of such online policies; Application of these to real-life examples such as machine-repairman problem of dynamic asset allocations.

10.2 International research visitors

10.2.1 Visits of international scientists

Manu Gupta visited Inria for a total duration of one month in June 2024. Deborah Hendrych visited Inria for a duration of one month in October 2024.

10.2.2 Visits to international teams

Research stays abroad Mathieu Besançon spent two weeks in the Zuse Institute Berlin in February to collaborate with Sebastian Pokutta and Timo Berthold.

10.3 National initiatives

Projects indicated with a ★ are projects coordinated by members of the POLARIS team.

ANR

ANR REFINO (JCJC 2020-2025)★ *Refined Mean Field Optimization* [250K€]

REFINO is an ANR starting grant (JCJC) coordinated by Nicolas Gast. The main objective on this project is to provide an innovative framework for optimal control of stochastic distributed agents. Restless bandit allocation is one particular example where the control that can be sent to each arm is restricted to an on/off signal. The originality of this framework is the use of refined mean field approximation to develop control heuristics that are asymptotically optimal as the number of arms

goes to infinity and that also have a better performance than existing heuristics for a moderate number of arms. As an example, we will use this framework in the context of smart grids, to develop control policies for distributed electric appliances.

ANR FAIRPLAY (JCJC 2021-2025)★ *Fair algorithms via game theory and sequential learning* [245K€]

FAIRPLAY is an ANR starting grant (JCJC) coordinated by Patrick Loiseau. Machine learning algorithms are increasingly used to optimize decision making in various areas, but this can result in unacceptable discrimination. The main objective of this project is to propose an innovative framework for the development of learning algorithms that respect fairness constraints. While the literature mostly focuses on idealized settings, the originality of this framework and central focus of this project is the use of game theory and sequential learning methods to account for constraints that appear in practical applications: strategic and decentralized aspects of the decisions and the data provided and absence of knowledge of certain parameters key to the fairness definition.

11 Dissemination

11.1 Promoting scientific activities

11.1.1 Scientific events: organisation

- Arnaud Legrand co-organized with five other colleagues the second **Meeting days of the French network for reproducible research** in Grenoble, France, March 2024, which has attracted about a hundred of researchers. The aim of these days was to provide an overview of the state of reproducibility in scientific research in France, taking into account the diversity of disciplines and practices.

11.1.2 Scientific events: selection

Member of the conference program committees

- Bruno Gaujal was member of the TPC of ICML 2024
- Nicolas Gast is member of the conference program committees of NeurIPS 2024, ACM SIGMETRICS 2024 and 2025, ICLR 2025.
- Panayotis Mertikopoulos served as a senior area chair for NeurIPS 2024, and as an area chair for ICML 2024, and ICLR 2024.
- Mathieu Besançon served as program committee member for CPAIOR 2024 and AAI 2024.

Reviewer

- Jonatha Anselmi was a reviewer for Mathematics of Operation Research, IEEE Transactions on Networking, IEEE Transactions on Parallel and Distributed Systems, Queueing Systems, Performance Evaluation.
- Panayotis Mertikopoulos served as a reviewer for COLT 2024.
- Mathieu Besançon was a reviewer for the INFORMS Journal on Computing.

11.1.3 Journal

Member of the editorial boards

- Nicolas Gast is associate editors of the journals Performance Evaluation and Stochastic Models.
- Panayotis Mertikopoulos is serving as the managing co-editor for the Open Journal of Mathematical Optimization.

- Panayotis Mertikopoulos served as an associate editor for Mathematics of Operations Research, Operations Research Letters, EURO Journal on Computational Optimization, Methodology and Computing in Applied Probability, and the Journal on Dynamics and Games.
- Jean-Marc Vincent is a member of the scientific board of the RADIX journal.

Reviewer - reviewing activities Bruno Gaujal was a reviewer for Questa

11.1.4 Invited talks

- Bruno Gaujal was invited to present his works at the following events:
 - "Machine Learning and Mean Field Games" Seminar series (May, 2024).
 - Recent trends in scheduling Theory (June 2024)
 - scheduling in Aussois (July 2024)
 - Keynote at Conf. Reinforcement Learning in Stochastic Networks (June 2024)
 - Journées MAS (Nov. 2024)
 - Atelier d'eval. de Perf. (table ronde) (Dec. 2024)
- Nicolas Gast was invited to present his work at the "13ème Atelier en Évaluation des Performances" (workshop AEP13, Toulouse), at the conference NetGCOOP 2024, and at the workshop RL4SN (Reinforcement Learning for Stochastic Networks).
- Jonatha Anselmi was invited to present his work on load-balancing and auto-scaling at the "13ème Atelier en Évaluation des Performances" (Toulouse)
- Arnaud Legrand was invited to give lectures and keynotes on Reproducible Research and Open Science on the following occasions:
 - M1 students in computer science at UGA (Dec. 2024): *Reproducible Research and Computer Science*
 - Theory days on Computational tools in physics : performances vs. reproducibility at Toulouse (Nov. 2023): *Numerical/Software Chaos and Reproducibility Issues in HPC*
 - 52nd ORAP Forum on *Reproductibilité et HPC* at Paris (Mar. 2024): Presentation of the *French Reproducible Research Network and Reproducibility Issues and Publication Evolutions (in HPC)*.
 - RSD Winter School at Le Pleyne (Feb. 2024): *Reproducible Research and Computer Science*.
 - L3 students of ENS Lyon at Le Pleyne (Jan. 2024): *Reproducible Research and Computer Science*.
- Panayotis Mertikopoulos was invited to give a tutorial at the following events:
 - EPFL-ETHZ Summer School on Multi-Agent Reinforcement Learning, Lausanne, CH
 - Games and Artificial Intelligence Multidisciplinary Summer School, Metz, FR
- Panayotis Mertikopoulos was invited to give a talk in the following conferences:
 - HMS 2024–The 2024 Conference of the Hellenic Mathematical Society, Olympia, GR
 - One World Optimization Seminar in Vienna, Vienna, AT
 - SOLACE Workshop on Learning in Games, Toulouse, FR
 - ACAC 2024–Athens Colloquium on Algorithms and Complexity, Athens, GR
 - Moroccan Game Theory Days, Rabat, MC
- Panayotis Mertikopoulos was invited to give a talk in the following universities and research institutes:

- Institute of Applied and Computational Mathematics, Herakleion, GR
 - Paris Game Theory Seminar (Institut Henri Poincaré), Paris, FR
 - Algorithmic Learning in Games Seminar, Online
 - GAIA/DATA Seminar, Grenoble, FR
- Mathieu Besançon gave an invited full talk at the Mixed-Integer Programming Workshop (University of Kentucky), and gave a lecture at the Computational Optimization at Work summer school at the Zuse Institute Berlin.

11.1.5 Leadership within the scientific community

- Vincent Danjean, Guillaume Huard, Arnaud Legrand, and Jean-Marc Vincent, are involved in the WP5 (*Performance analysis and prediction*) of the ExaSoft pillar (*High Performance Computing software and tools*) of the PEPR NumPEX (Numérique Hautes Performances pour l'Exascale).
- Panayotis Mertikopoulos is the scientific coordinator of the PEPR IA projet ciblé (acceleration grant) FOUNDRY: *Foundations of robustness and reliability in artificial intelligence*. The project has a total budget of 5M€ and involves five research teams across France (POLARIS in Grenoble, the Inria teams SCOOOL and FAIRPLAY, Dauphine and IMT in Paris, and ENS Lyon).
- Jean-Marc Vincent is a member of the scientific committee of the CIST.
- Jean-Marc Vincent is vice-head of the SIF, adjunct on teaching. In this context he is in charge of the organization of the annual meeting on education in computer science.

11.1.6 Scientific expertise

- Nicolas Gast was co-chair of the ACM SIGMETRICS "Student Research Competition".
- Jean-Marc Vincent is a scientific expert in EFELIA-MIAI.

11.1.7 Research administration

- Mathieu Besançon was a member of the hiring committee for a fixed-term Maître-Assistant at IMT Mines Albi.
- Vincent Danjean is a member of the *Conseil d'Administration* of Grenoble-INP.
- Bruno Gaujal was president of the jury for Mcf hiring in ENSGI (Grenoble).
- Nicolas Gast is vice-head of the Labex *EnergyAlps* that federates the community working on electrical energy in Grenoble.
- Nicolas Gast is vice-head of the *école doctorale MSTII* (the doctoral school managing PhD students in computer science and mathematics at Univ. Grenoble Alpes).
- Arnaud Legrand is a member of the Section 6 of the CoNRS.
- Arnaud Legrand is head of the SRCPR axis of the LIG and a member of LIG bureau.
- Arnaud Legrand is a member of *Comité Scientifique* of the Inria Grenoble.
- Florence Perronin is a member of the QVT team of the LIG.
- Jean-Marc Vincent was a member of the jury for Mcf hiring in Univ. Nancy.
- Jean-Marc Vincent is the head of the evaluation committee for *education in AI* of the MIAI chaire.

11.2 Teaching - Supervision - Juries

11.2.1 Teaching

- Jonatha Anselmi teaches Probabilités et simulation (32h) and Évaluation de performances (32h) at PolyTech Grenoble.
- Mathieu Besançon teaches half of the Advanced Models and Methods of Operations Research at the M2 ORCO (UGA).
- Vincent Danjean teaches the *Operating Systems, Programming Languages, Algorithms, Computer Science and Mediation* lectures in L3, M1 and Polytech Grenoble.
- Vincent Danjean organized with J.M. Vincent a complementary training for high school professors to teach computer science.
- Nicolas Gast teaches the *Reinforcement learning* part of the M2 course *Mathematical foundations of machine learning* at the M2 MOSIG (Grenoble).
- Bruno Gaujal teaches *Optimisation under uncertainties* (18h) at the M2 ORCO (UGA).
- Bruno Gaujal and Nicolas Gast teach *Markov Decision Process and Reinforcement Learning* (32h in total) at the M2 Info (ENS Lyon).
- Nicolas Gast is responsible of the cours "Introduction au Machine Learning" that is an optional course of the third year or the "licence informatique" at Univ. Grenoble Alpes.
- Guillaume Huard is responsible of L3 Info and of the Licence Info.
- Guillaume Huard is responsible of the courses *UNIX & C programming* in the L1 and L3 INFO, of *Object Oriented and Event-Driven Programming* in the L3 INFO, and of the *Objet Oriented Design* in M1 INFO.
- Arnaud Legrand and Jean-Marc Vincent teach the transversal *Scientific Methodology and Empirical Evaluation lecture* (36h) at the M2 MOSIG (UGA).
- The 3rd edition of the MOOC of Arnaud Legrand, K. Hinsin and C. Pouzat on *Reproducible Research: Methodological Principles for a Transparent Science* is still running. Over the 3 editions (Oct.-Dec. 2018, Apr.-June 2019, March 2020 - end of 2024), more than 22,800 persons have followed this MOOC and about 2600 certificates of achievement have been delivered. More than half of participants are PhD students and about 10% are undergraduates.
- The 1st edition of the MOOC of Arnaud Legrand, K. Hinsin and C. Pouzat on *Reproducible Research II: Practices and tools for managing computations and data* has been launched from May 2024 to October 2024. It has attracted 2,000 persons and is the result of more than 3 years of hard work.
- Florence Perronin teaches *Programming Languages* in L1.
- Florence Perronin is a member of the *conseil de perfectionnement* of the Mathematics license.
- Jean-Marc Vincent is in charge of the coordination of the training of high school teachers in computer science (NSI) in Grenoble.
- Jean-Marc Vincent teaches *Algorithms and Probabilities* at the L3, UGA.
- Jean-Marc Vincent teaches *Statistical Models and Litterate Programming* at the L3 MIAGE, UGA.
- Jean-Marc Vincent teaches *Mathematics for Computer Science* at the M1 MOSIG, UGA.
- Jean-Marc Vincent participates to the *Histoire de l'informatique* lecture at the ENSIMAG.
- Panayotis Mertikopoulos taught a graduate course on continuous optimization in the University of Athens.

11.2.2 Supervision

Bruno Gaujal is member of the CSI of Adrienne Tuynman (Univ. Lille) and Vittorio Puricelli (Univ. Toulouse)

11.2.3 Juries

- Nicolas Gast was member of the PhD committee of Younes Ben Mazziane (Univ. Sofia Antopolis and Inria): *Probabilistic analysis for caching*.
- Nicolas Gast was member of the jury for the "Prix de thèse PGM0" 2024.
- Arnaud Legrand was a reviewer for the PhD of Maël Madon (Univ. de Toulouse): *Digital Sufficiency in Data Centers : Studying the Impact of User Behaviors*.
- Arnaud Legrand was president of the PhD committee of Adrien Berthelot (ÉNS Lyon): *L'empreinte environnementale complète d'un usage numérique : contribution à l'analyse de cycle de vie de service numérique*.
- Panayotis Mertikopoulos was an external reviewer for the PhD thesis of Mohammad Reza Karimi (ETH Zürich) on *Stochastic Approximation on Riemannian Manifolds and the Space of Measures*

11.3 Popularization

11.3.1 Specific official responsibilities in science outreach structures

- Jean-Marc Vincent is particularly involved in the teaching of computer science at the national level.
 - He co-animates the national group *infosansordi* (unplugged computer science).
 - He participates to the training of high school computer science teachers (NSI) and he is preparing a new training on programming methods.
 - He also co-organizes series of seminars for high school computer science teachers.
 - He is a member of the national *inter-IREM* group on computer science and writes leaflets for high school computer science teachers.
- Jean-Marc Vincent is a member of the scientific committee of *Territoires de Sciences*, which in charge of popularization in the Grenoble area.
- Jean-Marc Vincent is a member of the committee of the national *Trophée NSI*.
- Jean-Marc Vincent is in charge of relationships on Education between Inria Grenoble and the Grenoble Rectorat.
- Jean-Marc Vincent is in the *comité de pilotage du plan national formation* in computer science for middle schools and high schools.
- Jean-Marc Vincent is in the *comité de pilotage du plan national formation* in unplugged computer science (*informatique débranchée*) for elementary schools.

11.3.2 Participation in Live events

- Jean-Marc Vincent has organized 4 conferences at the *Kateb Yacine* library for general audience: *Le numérique en questions*, with Inria, UGA and the Grenoble city.
- Jean-Marc Vincent has organized 2 days of the conference *MathC2+* toward middle school and high school students with Inria and UGA.

12 Scientific production

12.1 Publications of the year

International journals

- [1] J. Anselmi. ‘Asynchronous Load Balancing and Auto-scaling: Mean-Field Limit and Optimal Design’. In: *IEEE/ACM Transactions on Networking* (2024), pp. 1–12. DOI: [10.1109/TNET.2024.368130](https://doi.org/10.1109/TNET.2024.368130). URL: <https://hal.science/hal-04467913> (cit. on p. 20).
- [2] J. ANSELMi and J. Doncel. ‘Dispatching and scheduling multiserver jobs for throughput optimality’. In: *ACM SIGMETRICS Performance Evaluation Review* (2024), pp. 1–5. URL: <https://hal.science/hal-04792723>. In press (cit. on p. 19).
- [3] J. Anselmi and J. Doncel. ‘Balanced Splitting: A Framework for Achieving Zero-wait in the Multiserver-job Model’. In: *IEEE Transactions on Parallel and Distributed Systems* (27th Sept. 2024), pp. 1–12. DOI: [10.1109/TPDS.2024.3493631](https://doi.org/10.1109/TPDS.2024.3493631). URL: <https://hal.science/hal-04713929>. In press (cit. on p. 19).
- [4] J. Anselmi and J. Doncel. ‘Load Balancing with Job-Size Testing: Performance Improvement or Degradation?’ In: *ACM Transactions on Modeling and Performance Evaluation of Computing Systems* (2024), pp. 1–25. DOI: [10.1145/3651154](https://doi.org/10.1145/3651154). URL: <https://hal.science/hal-04477029> (cit. on p. 19).
- [5] J. Anselmi, B. Gaujal and L.-S. Rebuffi. ‘Learning Optimal Admission Control in Partially Observable Queueing Networks’. In: *Queueing Systems* (29th June 2024), pp. 1–48. DOI: [10.1007/s11134-024-09917-y](https://doi.org/10.1007/s11134-024-09917-y). URL: <https://hal.science/hal-04170992> (cit. on p. 22).
- [6] W. Azizian, F. Iutzeler, J. Malick and P. Mertikopoulos. ‘The rate of convergence of Bregman proximal methods: Local geometry vs. regularity vs. sharpness’. In: *SIAM Journal on Optimization* 34.3 (2024), pp. 2440–2471. DOI: [10.1137/23M1580218](https://doi.org/10.1137/23M1580218). URL: <https://inria.hal.science/hal-04307796> (cit. on p. 25).
- [7] M. Besançon, M. Anjos and L. Brotcorne. ‘Robust bilevel optimization for near-optimal lower-level solutions’. In: *Journal of Global Optimization* 90.4 (30th July 2024), pp. 813–842. DOI: [10.1007/s10898-024-01422-z](https://doi.org/10.1007/s10898-024-01422-z). URL: <https://hal.science/hal-04792534> (cit. on p. 26).
- [8] J.-P. Carvalho and B. Pradelski. ‘The representation dynamic and the “normalization” of group differences’. In: *Journal of Law, Economics, and Organization* (17th June 2024), pp. 1–28. DOI: [10.1093/jleo/ewae014](https://doi.org/10.1093/jleo/ewae014). URL: <https://hal.science/hal-04788154> (cit. on p. 31).
- [9] J.-P. Carvalho, B. Pradelski and C. Williams. ‘Intersectionality: Affirmative Action with Multidimensional Identities’. In: *Management Science* (13th Sept. 2024). DOI: [10.1287/mnsc.2023.03839](https://doi.org/10.1287/mnsc.2023.03839). URL: <https://hal.science/hal-04788142> (cit. on p. 31).
- [10] F. Filippini, J. ANSELMi, D. Ardagna and B. Gaujal. ‘A Stochastic Approach for Scheduling AI Training Jobs in GPU-based Systems’. In: *IEEE Transactions on Cloud Computing* 12.1 (2024), pp. 53–69. DOI: [10.1109/TCC.2023.3336540](https://doi.org/10.1109/TCC.2023.3336540). URL: <https://hal.science/hal-04337856> (cit. on p. 20).
- [11] N. Gast and B. van Houdt. ‘Approximations to Study the Impact of the Service Discipline in Systems with Redundancy’. In: *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 8.1 (Mar. 2024), pp. 1–32. DOI: [10.1145/3639040](https://doi.org/10.1145/3639040). URL: <https://inria.hal.science/hal-04390014> (cit. on p. 21).
- [12] B. Gaujal, A. Girault and S. Plassart. ‘An MDP-Based Solution for the Energy Minimization of Non-Clairvoyant Hard Real-Time Systems’. In: *Real-Time Systems* (5th Dec. 2024), p. 47. DOI: [10.1007/s11241-024-09433-5](https://doi.org/10.1007/s11241-024-09433-5). URL: <https://inria.hal.science/hal-02371742> (cit. on pp. 12, 22).
- [13] H. Lebeau, F. Chatelain and R. Couillet. ‘Asymptotic Gaussian Fluctuations of Eigenvectors in Spectral Clustering’. In: *IEEE Signal Processing Letters* 31 (2024), pp. 1920–1924. DOI: [10.1109/LSP.2024.3422886](https://doi.org/10.1109/LSP.2024.3422886). URL: <https://hal.science/hal-04673319> (cit. on p. 29).

- [14] P. Mertikopoulos, Y.-P. Hsieh and V. Cevher. ‘A unified stochastic approximation framework for learning in games’. In: *Mathematical Programming* 203 (Jan. 2024), pp. 559–609. DOI: [10.1007/s10107-023-02001-y](https://doi.org/10.1007/s10107-023-02001-y). URL: <https://hal.science/hal-03874012> (cit. on p. 26).
- [15] P. Mertikopoulos, H. H. Nax and B. S. R. Pradelski. ‘Quick or cheap? Breaking points in dynamic markets’. In: *Journal of Mathematical Economics* 112 (June 2024), p. 102987. DOI: [10.1016/j.jmateco.2024.102987](https://doi.org/10.1016/j.jmateco.2024.102987). URL: <https://inria.hal.science/hal-04664218> (cit. on p. 30).
- [16] P. Mertikopoulos and W. H. Sandholm. ‘Nested replicator dynamics, nested logit choice, and similarity-based learning’. In: *Journal of Economic Theory* 220 (Sept. 2024), p. 105881. DOI: [10.1016/j.jet.2024.105881](https://doi.org/10.1016/j.jet.2024.105881). URL: <https://hal.science/hal-04733800> (cit. on p. 27).
- [17] J.-M. Vincent and M. Althuser. ‘La Télé-vision : l’informatique débranchée au service du concept d’information’. In: *Petit x* 120 (12th Sept. 2024), pp. 9–38. URL: <https://hal.science/hal-04878673> (cit. on p. 31).

National journals

- [18] S. Alayrangués, O. Baudon, E. Beffara, R. Charpentier, S. Daniel, C. Declercq, A. Grimaud, S. Hoarau, A. Héam, P. Marquet, J.-C. Masseron, A. Meyer, M. More, F. Nény, C. Prouteau, J.-M. Vincent, E. Volte and N. Weibel. ‘Booléens en programmation : une analyse à destination des enseignants’. In: *Recherches et recherches-actions en didactique de l’informatique* 1 (2024), pp. 1–33. URL: <https://hal.science/hal-03812698> (cit. on p. 31).

International peer-reviewed conferences

- [19] S. Allmeier and N. Gast. ‘Computing the Bias of Constant-step Stochastic Approximation with Markovian Noise’. In: *NeurIPS 2024 - 38th Annual Conference on Neural Information Processing Systems*. Vancouver, Canada, 2024, pp. 1–23. URL: <https://inria.hal.science/hal-04837291> (cit. on p. 21).
- [20] W. Azizian, F. Iutzeler, J. Malick and P. Mertikopoulos. ‘What is the long-run distribution of stochastic gradient descent? A large deviations analysis’. In: *ICML 2024 - 41st International Conference on Machine Learning*. Vienna, Austria, July 2024, pp. 1–70. URL: <https://hal.science/hal-04629305> (cit. on p. 25).
- [21] A. Kontogiannis, V. Pollatos, S. Kanellopoulos, P. Mertikopoulos, A. Pagourtzis and I. Panageas. ‘The computational complexity of finding second-order stationary points’. In: *ICML 2024 - 41st International Conference on Machine Learning*. Vienna, Austria, July 2024. URL: <https://hal.science/hal-04629307> (cit. on p. 26).
- [22] H. Lebeau, M. El Amine Seddik and J. H. d. M. Goulart. ‘Performance Gaps in Multi-view Clustering under the Nested Matrix-Tensor Model’. In: *ICLR 2024 - 12th International Conference on Learning Representations*. Wien, Austria, 2024, pp. 1–29. URL: <https://hal.science/hal-04464877> (cit. on p. 29).
- [23] D. Legacci, P. Mertikopoulos, C. Papadimitriou, G. Piliouras and B. S. R. Pradelski. ‘No-regret learning in harmonic games: Extrapolation in the face of conflicting interests’. In: *Proceedings of the 38th International Conference on Neural Information Processing Systems*. NeurIPS 2024 - 38th Conference on Neural Information Processing Systems. Vancouver, Canada, 2024, pp. 1–36. URL: <https://inria.hal.science/hal-04863624> (cit. on pp. 14, 28).
- [24] D. Legacci, P. Mertikopoulos and B. Pradelski. ‘A geometric decomposition of finite games: Convergence vs. recurrence under exponential weights’. In: *ICML 2024 - 41st International Conference on Machine Learning*. Vienna, Austria, July 2024. URL: <https://hal.science/hal-04629310> (cit. on p. 14).
- [25] K. Lotidis, A. Giannou, P. Mertikopoulos and N. Bambos. ‘Accelerated regularized learning in finite N-person games’. In: *Proceedings of the 38th International Conference on Neural Information Processing Systems*. NeurIPS 2024 - 38th Conference on Neural Information Processing Systems. Vancouver, Canada, 2024. URL: <https://inria.hal.science/hal-04863626> (cit. on p. 28).

- [26] G. Mexi, S. Shamsi, M. Besançon and P. Le Bodic. ‘Probabilistic Lookahead Strong Branching via a Stochastic Abstract Branching Model’. In: *Integration of Constraint Programming, Artificial Intelligence, and Operations Research*. CPAIOR 2024 - 21st International Conference on Integration of Constraint Programming, Artificial Intelligence, and Operations Research. Vol. 14743. Lecture Notes in Computer Science. Uppsala, Sweden: Springer Nature Switzerland, 25th May 2024, pp. 56–73. DOI: [10.1007/978-3-031-60599-4_4](https://doi.org/10.1007/978-3-031-60599-4_4). URL: <https://hal.science/hal-04792529> (cit. on pp. 14, 23).

Conferences without proceedings

- [27] K. Sharma, D. Hendrych, M. Besançon and S. Pokutta. ‘Network Design for the Traffic Assignment Problem with Mixed-Integer Frank-Wolfe’. In: INFORMS Optimization Society Conference. Houston, United States, 30th Mar. 2024. URL: <https://hal.science/hal-04453777> (cit. on p. 24).

Edition (books, proceedings, special issue of a journal)

- [28] *Solving the Optimal Experiment Design Problem with Mixed-Integer Convex Methods*. 22nd International Symposium on Experimental Algorithms (SEA 2024). Vol. 301. 301. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024, 16:1–16:22. DOI: [10.4230/LIPIcs.SEA.2024.16](https://doi.org/10.4230/LIPIcs.SEA.2024.16). URL: <https://hal.science/hal-04431320> (cit. on p. 24).

Doctoral dissertations and habilitation theses

- [29] S. Allmeier. ‘Mean Field Methods for Heterogeneous and Coupled Systems’. Université Grenoble Alpes [2020-....], 15th Apr. 2024. URL: <https://theses.hal.science/tel-04756944> (cit. on p. 21).
- [30] V. Boone. ‘Optimal Regrets in Markov Decision Processes’. UGA (Université Grenoble Alpes), 25th Nov. 2024. URL: <https://hal.science/tel-04874467> (cit. on p. 23).
- [31] R. Castera. ‘Efficiency and Fairness in Matching Problems’. UGA (Université Grenoble Alpes), 16th Oct. 2024. URL: <https://hal.science/tel-04874485> (cit. on p. 30).
- [32] V. Leger. ‘Theoretical guarantees and improvement of large dimensional multi-task semi-supervised classification’. UGA (Université Grenoble Alpes), 14th Nov. 2024. URL: <https://hal.science/tel-04874448> (cit. on p. 28).
- [33] C. Sejourne. ‘Theoretical performance analysis applied to non-convex and fair algorithms using Random Matrix Theory’. UGA (Université Grenoble Alpes), 2nd Dec. 2024. URL: <https://hal.science/tel-04874411> (cit. on p. 29).

Reports & preprints

- [34] S. Bolusani, M. Besançon, K. Bestuzheva, A. Chmiela, J. Dionísio, T. Donkiewicz, J. van Doornmalen, L. Eifler, M. Ghannam, A. Gleixner, C. Graczyk, K. Halbig, I. Hedtke, A. Hoen, C. Hojny, R. van der Hulst, D. Kamp, T. Koch, K. Kofler, J. Lentz, J. Manns, G. Mexi, E. Mühmer, M. E. Pfetsch, F. Schlösser, F. Serrano, Y. Shinano, M. Turner, S. Vigerske, D. Weninger and L. Xu. *The SCIP Optimization Suite 9.0*. 27th Feb. 2024. URL: <https://hal.science/hal-04768660> (cit. on p. 24).
- [35] R. Castera, P. Loiseau and B. Pradelski. *Correlation of Rankings in Matching Markets*. Feb. 2024. URL: <https://hal.science/hal-03672270> (cit. on p. 30).
- [36] Y. Chasseray, M. Besançon, E. Petittedemange and X. Lorca. *Optimisation models for the design of multiple self-consumption loops in semi-rural areas*. 2024. DOI: [10.48550/arXiv.2404.04428](https://doi.org/10.48550/arXiv.2404.04428). URL: <https://hal.science/hal-04741974> (cit. on p. 24).
- [37] F. Falniowski and P. Mertikopoulos. *On the discrete-time origins of the replicator dynamics: From convergence to instability and chaos*. Feb. 2024. URL: <https://hal.science/hal-04629311> (cit. on p. 27).

- [38] N. Gast, B. Gaujal and C. Yan. *Reoptimization Nearly Solves Weakly Coupled Markov Decision Processes*. 2024. URL: <https://inria.hal.science/hal-04570177> (cit. on p. 22).
- [39] H. Lebeau, F. Chatelain, R. Couillet and R. Couillet. *A Random Matrix Approach to Low-Multilinear-Rank Tensor Approximation*. 6th June 2024. URL: <https://hal.science/hal-04673321> (cit. on p. 29).
- [40] T. Lin, P. Mertikopoulos and M. I. Jordan. *Explicit second-order min-max optimization methods with optimal convergence guarantees*. Apr. 2024. URL: <https://hal.science/hal-03873989> (cit. on p. 26).
- [41] I. Lytras and P. Mertikopoulos. *Tamed Langevin sampling under weaker conditions*. 27th May 2024. URL: <https://hal.science/hal-04629313> (cit. on p. 25).
- [42] M. Turner, T. Berthold, M. Besançon and T. Koch. *Branching via Cutting Plane Selection: Improving Hybrid Branching*. 1st Feb. 2024. URL: <https://hal.science/hal-04431308> (cit. on p. 23).

12.2 Cited publications

- [43] T. Cornebize and A. Legrand. ‘Simulation-based Optimization and Sensibility Analysis of MPI Applications: Variability Matters’. In: *Journal of Parallel and Distributed Computing* (20th Apr. 2022). DOI: [10.1016/j.jpdc.2022.04.002](https://doi.org/10.1016/j.jpdc.2022.04.002). URL: <https://hal.inria.fr/hal-03141988> (cit. on p. 6).
- [44] B. Donassolo, A. Legrand, P. Mertikopoulos and I. Fajjari. ‘Online Reconfiguration of IoT Applications in the Fog: The Information-Coordination Trade-off’. In: *IEEE Transactions on Parallel and Distributed Systems* 33.5 (2022), pp. 1156–1172. DOI: [10.1109/TPDS.2021.3097281](https://doi.org/10.1109/TPDS.2021.3097281). URL: <https://hal.inria.fr/hal-02636987> (cit. on p. 7).
- [45] N. Gast, B. Gaujal and C. Yan. ‘LP-based policies for restless bandits: necessary and sufficient conditions for (exponentially fast) asymptotic optimality’. In: *Mathematics of Operations Research* (1st Dec. 2023). DOI: [10.1287/moor.2022.0101](https://doi.org/10.1287/moor.2022.0101). URL: <https://inria.hal.science/hal-03262307> (cit. on p. 9).
- [46] J. D. Leshno and B. S. R. Pradelski. ‘The importance of memory for price discovery in decentralized markets’. In: *Games and Economic Behavior* 125 (Jan. 2021), pp. 62–78. DOI: [10.1016/j.geb.2020.11.002](https://doi.org/10.1016/j.geb.2020.11.002). URL: <https://hal.archives-ouvertes.fr/hal-03100097> (cit. on p. 11).
- [47] M. C. Miletto, L. L. Nesi, L. Mello Schnorr and A. Legrand. ‘Performance Analysis of Irregular Task-Based Applications on Hybrid Platforms: Structure Matters’. In: *Future Generation Computer Systems* 135 (1st Oct. 2022). URL: <https://hal.inria.fr/hal-03298021> (cit. on p. 6).
- [48] M. Oliu-Barton, B. Pradelski, P. Aghion, P. Artus, I. Kickbusch, J. Lazarus, D. Sridhar and S. Vander-slott. ‘SARS-CoV-2 elimination, not mitigation, creates best outcomes for health, the economy, and civil liberties’. In: *The Lancet* 397.10291 (June 2021), pp. 2234–2236. DOI: [10.1016/S0140-6736\(21\)00978-8](https://doi.org/10.1016/S0140-6736(21)00978-8). URL: <https://hal.archives-ouvertes.fr/hal-03355143> (cit. on p. 12).
- [49] B. Pradelski and M. Oliu-Barton. ‘Green zoning: An effective policy tool to tackle the Covid-19 pandemic’. In: *Health Policy* 125.8 (Aug. 2021), pp. 981–986. DOI: [10.1016/j.healthpol.2021.06.001](https://doi.org/10.1016/j.healthpol.2021.06.001). URL: <https://hal.archives-ouvertes.fr/hal-03355146> (cit. on p. 11).
- [50] S. Zrigui, R. y. de Camargo, A. Legrand and D. Trystram. ‘Improving the Performance of Batch Schedulers Using Online Job Runtime Classification’. In: *Journal of Parallel and Distributed Computing* 164 (2nd Feb. 2022), pp. 83–95. DOI: [10.1016/j.jpdc.2022.01.003](https://doi.org/10.1016/j.jpdc.2022.01.003). URL: <https://hal.archives-ouvertes.fr/hal-03023222> (cit. on p. 6).
- [51] B. Gaujal, J. Doncel and N. Gast. ‘Vaccination in a Large Population: Mean Field Equilibrium versus Social Optimum’. In: NETGCOOP 2020 - 10th International Conference on NETwork Games, CONTROL and OPTimization. Cargèse, France, 22nd Sept. 2021, pp. 1–9. URL: <https://hal.inria.fr/hal-02938850> (cit. on p. 8).
- [52] A. Giannou, E. V. Vlatakis-Gkaragkounis and P. Mertikopoulos. ‘Survival of the strictest: Stable and unstable equilibria under regularized learning with partial information’. In: COLT 2021 - 34th Annual Conference on Learning Theory. Boulder, United States, 15th Aug. 2021, pp. 1–30. URL: <https://hal.inria.fr/hal-03342404> (cit. on p. 9).

- [53] Y.-P. Hsieh, P. Mertikopoulos and V. Cevher. ‘The limits of min-max optimization algorithms: Convergence to spurious non-critical sets’. In: *ICML 2021 - 38th International Conference on Machine Learning*. Vienna, Austria, July 2021. URL: <https://hal.archives-ouvertes.fr/hal-03043862> (cit. on p. 10).
- [54] L. L. Nesi, A. Legrand and L. Mello Schnorr. ‘Exploiting system level heterogeneity to improve the performance of a GeoStatistics multi-phase task-based application’. In: *ICPP 2021 - 50th International Conference on Parallel Processing*. Lemont, United States, 9th Aug. 2021, pp. 1–10. DOI: [10.1145/3472456.3472516](https://doi.org/10.1145/3472456.3472516). URL: <https://hal.inria.fr/hal-03280459> (cit. on p. 6).
- [55] Y.-G. Hsieh. ‘Decision-Making in multi-agent systems: delays, adaptivity, and learning in games’. PhD thesis. Université Grenoble Alpes, 7th Nov. 2023. URL: <http://www.theses.fr/s245435> (cit. on p. 14).
- [56] L. Leandro Nesi. ‘Strategies for Distributing Task-Based Applications on Heterogeneous Platforms’. PhD thesis. Université Grenoble Alpes, 14th Sept. 2023. URL: <http://www.theses.fr/s300803> (cit. on p. 14).
- [57] P. H. Rocha Bruel. ‘Toward transparent and parsimonious methods for automatic performance tuning’. UGA (Université Grenoble Alpes); USP (Universidade de São Paulo), 9th July 2021. URL: <https://tel.archives-ouvertes.fr/tel-03526483> (cit. on p. 7).
- [58] A. Andreou, M. Silva, F. Benevenuto, O. Goga, P. Loiseau and A. Mislove. ‘Measuring the Facebook Advertising Ecosystem’. In: *NDSS 2019 - Proceedings of the Network and Distributed System Security Symposium*. San Diego, United States, Feb. 2019, pp. 1–15. DOI: [10.14722/ndss.2019.23280](https://doi.org/10.14722/ndss.2019.23280). URL: <https://hal.archives-ouvertes.fr/hal-01959145> (cit. on p. 11).
- [59] A. Andreou, G. Venkatadri, O. Goga, K. P. Gummadi, P. Loiseau and A. Mislove. ‘Investigating Ad Transparency Mechanisms in Social Media: A Case Study of Facebook’s Explanations’. In: *NDSS 2018 - Network and Distributed System Security Symposium*. San Diego, United States, Feb. 2018, pp. 1–15. DOI: [10.14722/ndss.2018.23204](https://doi.org/10.14722/ndss.2018.23204). URL: <https://hal.archives-ouvertes.fr/hal-01955309> (cit. on p. 11).
- [60] J. Anselmi. ‘Combining Size-Based Load Balancing with Round-Robin for Scalable Low Latency’. In: *IEEE Transactions on Parallel and Distributed Systems* (2019), pp. 1–3. DOI: [10.1109/TPDS.2019.2950621](https://doi.org/10.1109/TPDS.2019.2950621). URL: <https://hal.archives-ouvertes.fr/hal-02276789> (cit. on p. 8).
- [61] J. Anselmi and J. Doncel. ‘Asymptotically Optimal Size-Interval Task Assignments’. In: *IEEE Transactions on Parallel and Distributed Systems* 30.11 (2019), pp. 2422–2433. DOI: [10.1109/TPDS.2019.2920121](https://doi.org/10.1109/TPDS.2019.2920121). URL: <https://hal.archives-ouvertes.fr/hal-02318576> (cit. on p. 8).
- [62] J. Anselmi and F. Dufour. ‘Power-of-d-Choices with Memory: Fluid Limit and Optimality’. In: *Mathematics of Operations Research* 45.3 (2020), pp. 862–888. DOI: [10.1287/moor.2019.1014](https://doi.org/10.1287/moor.2019.1014). URL: <https://hal.archives-ouvertes.fr/hal-02394147> (cit. on p. 8).
- [63] R. M. Badia, J. Labarta, J. Giménez and F. Escalé. ‘Dimemas: Predicting MPI Applications Behaviour in Grid Environments’. In: *Proc. of the Workshop on Grid Applications and Programming Tools*. 2003 (cit. on p. 6).
- [64] S. Böhm and C. Engelmann. ‘xSim: The Extreme-Scale Simulator’. In: *HPCS*. Istanbul, Turkey, 2011 (cit. on p. 6).
- [65] P. Bruel, S. Quinito Masnada, B. Videau, A. Legrand, J.-M. Vincent and A. Goldman. ‘Autotuning under Tight Budget Constraints: A Transparent Design of Experiments Approach’. In: *CCGrid 2019 - International Symposium in Cluster, Cloud, and Grid Computing*. Larcana, Cyprus: IEEE, May 2019, pp. 1–10. DOI: [10.1109/CCGRID.2019.00026](https://doi.org/10.1109/CCGRID.2019.00026). URL: <https://hal.inria.fr/hal-02110868> (cit. on p. 7).
- [66] H. Brunst, D. Hackenberg, G. Juckeland and H. Rohling. ‘Comprehensive Performance Tracking with VAMPIR 7’. In: *Tools for High Performance Computing 2009*. Springer Berlin Heidelberg, 2010. DOI: http://dx.doi.org/10.1007/978-3-642-11261-4_2 (cit. on p. 6).
- [67] G. Casale and N. Gast. ‘Performance analysis methods for list-based caches with non-uniform access’. In: *IEEE/ACM Transactions on Networking* (Dec. 2020), pp. 1–18. DOI: [10.1109/TNET.2020.3042869](https://doi.org/10.1109/TNET.2020.3042869). URL: <https://hal.inria.fr/hal-03102188> (cit. on p. 8).

- [68] A. Chakraborty, G. K. Patro, N. Ganguly, K. P. Gummadi and P. Loiseau. ‘Equality of Voice: Towards Fair Representation in Crowdsourced Top-K Recommendations’. In: *FAT* 2019 - ACM Conference on Fairness, Accountability, and Transparency*. Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAT*). Atlanta, United States: ACM, Jan. 2019, pp. 129–138. DOI: [10.1145/3287560.3287570](https://doi.org/10.1145/3287560.3287570). URL: <https://hal.archives-ouvertes.fr/hal-01959135> (cit. on p. 10).
- [69] T. Cornebize, A. Legrand and F. C. Heinrich. ‘Fast and Faithful Performance Prediction of MPI Applications: the HPL Case Study’. In: *2019 IEEE International Conference on Cluster Computing (CLUSTER)*. Albuquerque, United States, Sept. 2019. DOI: [10.1109/CLUSTER.2019.8891011](https://doi.org/10.1109/CLUSTER.2019.8891011). URL: <https://hal.inria.fr/hal-02096571> (cit. on p. 6).
- [70] P. COUCHENEY, B. Gaujal and P. Mertikopoulos. ‘Penalty-Regulated Dynamics and Robust Learning Procedures in Games’. In: *Mathematics of Operations Research* 40.3 (2015), pp. 611–633. DOI: [10.1287/moor.2014.0687](https://doi.org/10.1287/moor.2014.0687). URL: <https://hal.inria.fr/hal-01235243> (cit. on p. 9).
- [71] A. Degomme, A. Legrand, G. Markomanolis, M. Quinson, M. L. Stillwell and F. Suter. ‘Simulating MPI applications: the SMPI approach’. In: *IEEE Transactions on Parallel and Distributed Systems* 28.8 (Aug. 2017), p. 14. DOI: [10.1109/TPDS.2017.2669305](https://doi.org/10.1109/TPDS.2017.2669305). URL: <https://hal.inria.fr/hal-01415484> (cit. on p. 6).
- [72] B. Donassolo, I. Fajjari, A. Legrand and P. Mertikopoulos. ‘Load Aware Provisioning of IoT Services on Fog Computing Platform’. In: *IEEE International Conference on Communications (ICC)*. Shanghai, China: IEEE, May 2019. DOI: [10.1109/ICC.2019.8762052](https://doi.org/10.1109/ICC.2019.8762052). URL: <https://hal.inria.fr/hal-02056743> (cit. on p. 7).
- [73] J. Doncel, N. Gast and B. Gaujal. ‘Are mean-field games the limits of finite stochastic games?’ In: *The 18th Workshop on Mathematical performance Modeling and Analysis*. Nice, France, June 2016. URL: <https://hal.inria.fr/hal-01321020> (cit. on p. 8).
- [74] J. Doncel, N. Gast and B. Gaujal. ‘Discrete Mean Field Games: Existence of Equilibria and Convergence’. In: *Journal of Dynamics and Games* 6.3 (2019), pp. 1–19. DOI: [10.3934/jdg.2019016](https://doi.org/10.3934/jdg.2019016). URL: <https://hal.inria.fr/hal-01277098> (cit. on p. 8).
- [75] V. Emelianov, G. Arvanitakis, N. Gast, K. P. Gummadi and P. Loiseau. ‘The Price of Local Fairness in Multistage Selection’. In: *IJCAI-2019 - Twenty-Eighth International Joint Conference on Artificial Intelligence*. Macao, France: International Joint Conferences on Artificial Intelligence Organization, Aug. 2019, pp. 5836–5842. DOI: [10.24963/ijcai.2019/809](https://doi.org/10.24963/ijcai.2019/809). URL: <https://hal.inria.fr/hal-02145071> (cit. on p. 11).
- [76] V. Emelianov, N. Gast, K. P. Gummadi and P. Loiseau. ‘On Fair Selection in the Presence of Implicit Variance’. In: *EC 2020 - Twenty-First ACM Conference on Economics and Computation*. Budapest, Hungary: ACM, July 2020, pp. 649–675. DOI: [10.1145/3391403.3399482](https://doi.org/10.1145/3391403.3399482). URL: <https://hal.inria.fr/hal-02880578> (cit. on p. 11).
- [77] L. Flokas, E. V. Vlatakis-Gkaragkounis, T. Lianeas, P. Mertikopoulos and G. Piliouras. ‘No-regret learning and mixed Nash equilibria: They do not mix’. In: *NeurIPS 2020 - 34th International Conference on Neural Information Processing Systems*. Vancouver, Canada, Dec. 2020, pp. 1–24. URL: <https://hal.inria.fr/hal-03043763> (cit. on p. 9).
- [78] V. Garcia Pinto, L. M. Schnorr, L. Stanisic, A. Legrand, S. Thibault and V. Danjean. ‘A Visual Performance Analysis Framework for Task-based Parallel Applications running on Hybrid Clusters’. In: *Concurrency and Computation: Practice and Experience* 30.18 (Apr. 2018), pp. 1–31. DOI: [10.1002/cpe.4472](https://doi.org/10.1002/cpe.4472). URL: <https://hal.inria.fr/hal-01616632> (cit. on p. 6).
- [79] N. Gast. ‘Expected Values Estimated via Mean-Field Approximation are $1/N$ -Accurate’. In: *ACM SIGMETRICS / International Conference on Measurement and Modeling of Computer Systems, SIGMETRICS ’17*. Vol. 1. ACM SIGMETRICS / International Conference on Measurement and Modeling of Computer Systems, SIGMETRICS ’17. Urbana-Champaign, United States, June 2017, p. 26. DOI: [10.1145/2896377.2901463](https://doi.org/10.1145/2896377.2901463). URL: <https://hal.inria.fr/hal-01553133> (cit. on p. 7).

- [80] N. Gast, L. Bortolussi and M. Tribastone. ‘Size Expansions of Mean Field Approximation: Transient and Steady-State Analysis’. In: *Performance Evaluation* (2018), pp. 1–15. URL: <https://hal.inria.fr/hal-01891632> (cit. on p. 8).
- [81] N. Gast, B. Gaujal and K. Khun. ‘Learning algorithms for Markovian Bandits: Is Posterior Sampling more Scalable than Optimism?’ In: *Transactions on Machine Learning Research Journal* (Nov. 2022). URL: <https://inria.hal.science/hal-03262006> (cit. on p. 9).
- [82] N. Gast, B. Gaujal and C. Yan. ‘Exponential Convergence Rate for the Asymptotic Optimality of Whittle Index Policy’. Dec. 2020. URL: <https://hal.inria.fr/hal-03041176> (cit. on p. 9).
- [83] N. Gast and B. V. Houdt. ‘A Refined Mean Field Approximation’. In: *ACM SIGMETRICS 2018*. Irvine, United States, June 2018, p. 1. URL: <https://hal.inria.fr/hal-01891642> (cit. on p. 7).
- [84] N. Gast, S. Ioannidis, P. Loiseau and B. Roussillon. ‘Linear Regression from Strategic Data Sources’. In: *ACM Transactions on Economics and Computation* 8.2 (May 2020), pp. 1–24. DOI: [10.1145/3391436](https://doi.org/10.1145/3391436). URL: <https://hal.inria.fr/hal-02982682> (cit. on p. 11).
- [85] N. Gast, D. Latella and M. Massink. ‘A Refined Mean Field Approximation for Synchronous Population Processes’. In: *MAMA 2018 Workshop on Mathematical performance Modeling and Analysis*. Irvine, United States, June 2018, pp. 1–3. URL: <https://hal.inria.fr/hal-01891629> (cit. on p. 8).
- [86] N. Gast and B. Van Houdt. ‘Asymptotically Exact TTL-Approximations of the Cache Replacement Algorithms LRU(m) and h-LRU’. In: *28th International Teletraffic Congress (ITC 28)*. Würzburg, Germany, Sept. 2016. URL: <https://hal.inria.fr/hal-01292269> (cit. on p. 8).
- [87] N. Gast and B. Van Houdt. ‘TTL Approximations of the Cache Replacement Algorithms LRU(m) and h-LRU’. In: *Performance Evaluation* (Sept. 2017). DOI: [10.1016/j.peva.2017.09.002](https://doi.org/10.1016/j.peva.2017.09.002). URL: <https://hal.inria.fr/hal-01622059> (cit. on p. 8).
- [88] B. Gaujal, A. Girault and S. Plassart. ‘A Linear Time Algorithm for Computing Off-line Speed Schedules Minimizing Energy Consumption’. In: *MSR 2019 - 12ème Colloque sur la Modélisation des Systèmes Réactifs*. Angers, France, Nov. 2019, pp. 1–14. URL: <https://hal.archives-ouvertes.fr/hal-02372136> (cit. on p. 12).
- [89] B. Gaujal, A. Girault and S. Plassart. ‘Discrete and Continuous Optimal Control for Energy Minimization in Real-Time Systems’. In: *EBC CSP 2020 - 6th International Conference on Event-Based Control, Communication, and Signal Processing*. Krakow, Poland: IEEE, Sept. 2020, pp. 1–8. DOI: [10.1109/EBC CSP51266.2020.9291353](https://doi.org/10.1109/EBC CSP51266.2020.9291353). URL: <https://hal.archives-ouvertes.fr/hal-03020330> (cit. on p. 12).
- [90] B. Gaujal, A. Girault and S. Plassart. ‘Dynamic Speed Scaling Minimizing Expected Energy Consumption for Real-Time Tasks’. In: *Journal of Scheduling* (July 2020), pp. 1–25. DOI: [10.1007/s10951-020-00660-9](https://doi.org/10.1007/s10951-020-00660-9). URL: <https://hal.inria.fr/hal-02888573> (cit. on p. 12).
- [91] M. Heath and J. Etheridge. ‘Visualizing the performance of parallel programs’. In: *IEEE software* 8.5 (1991) (cit. on p. 6).
- [92] F. C. Heinrich, T. Cornebize, A. Degomme, A. Legrand, A. Carpen-Amarie, S. Hunold, A.-C. Orgerie and M. Quinson. ‘Predicting the Energy Consumption of MPI Applications at Scale Using a Single Node’. In: *Cluster 2017*. IEEE. Hawaii, United States, Sept. 2017. URL: <https://hal.inria.fr/hal-01523608> (cit. on p. 6).
- [93] T. Hoefler, T. Schneider and A. Lumsdaine. ‘LogGOPSIm - Simulating Large-Scale Applications in the LogGOPS Model’. In: *ACM Workshop on Large-Scale System and Application Performance*. 2010 (cit. on p. 6).
- [94] L. V. Kalé, G. Zheng, C. W. Lee and S. Kumar. ‘Scaling applications to massively parallel machines using Projections performance analysis tool’. In: *Future Generation Comp. Syst.* 22.3 (2006) (cit. on p. 6).

- [95] R. Keller Tesser, L. Mello Schnorr, A. Legrand, F. Dupros and P. O. A. Navaux. ‘Using Simulation to Evaluate and Tune the Performance of Dynamic Load Balancing of an Over-decomposed Geophysics Application’. In: *Euro-Par 2017: 23rd International European Conference on Parallel and Distributed Computing*. Santiago de Compostela, Spain, Aug. 2017, p. 15. URL: <https://hal.inria.fr/hal-01567792> (cit. on p. 6).
- [96] R. Keller Tesser, L. Mello Schnorr, A. Legrand, C. Heinrich, F. Dupros and P. O. Alexandre Navaux. ‘Performance Modeling of a Geophysics Application to Accelerate the Tuning of Over-decomposition Parameters through Simulation’. In: *Concurrency and Computation: Practice and Experience* (2018), pp. 1–21. DOI: 10.1002/cpe.5012. URL: <https://hal.inria.fr/hal-01891416> (cit. on p. 6).
- [97] T.-E. Kennouche, F. Cadoux, N. Gast and B. Vinot. ‘ASGridS: Asynchronous Smart-Grids Distributed Simulator’. In: *APPEEC 2019 - 11th IEEE PES Asia-Pacific Power and Energy Engineering Conference*. Macao, Macau SAR China: IEEE, Dec. 2019, pp. 1–5. URL: <https://hal.archives-ouvertes.fr/hal-02384051> (cit. on p. 12).
- [98] J. M. Kleinberg and M. Raghavan. ‘Selection Problems in the Presence of Implicit Bias’. In: *Proceedings of the 9th Innovations in Theoretical Computer Science Conference (ITCS)*. 2018, 33:1–33:17 (cit. on p. 10).
- [99] A. Legrand, D. Trystram and S. Zrigui. ‘Adapting Batch Scheduling to Workload Characteristics: What can we expect From Online Learning?’ In: *IPDPS 2019 - 33rd IEEE International Parallel & Distributed Processing Symposium*. Rio de Janeiro, Brazil: IEEE, May 2019, pp. 686–695. DOI: 10.1109/IPDPS.2019.00077. URL: <https://hal.archives-ouvertes.fr/hal-02044903> (cit. on p. 6).
- [100] M. Mendil, N. Gast and H.-J. Audéoud. ‘Collisions groupées lors du mécanisme d’évitement de collisions de CPL-G3’. In: *CoRes 2020 - Rencontres Francophones sur la Conception de Protocoles, l’Évaluation de Performance et l’Expérimentation des Réseaux de Communication*. Lyon, France, Sept. 2020, pp. 1–4. URL: <https://hal.archives-ouvertes.fr/hal-02879294> (cit. on p. 12).
- [101] P. Mertikopoulos, B. Lecouat, H. Zenati, C.-S. Foo, V. Chandrasekhar and G. Piliouras. ‘Optimistic Mirror Descent in Saddle-Point Problems: Going the Extra (Gradient) Mile’. In: *ICLR 2019 - 7th International Conference on Learning Representations*. New Orleans, United States, May 2019, pp. 1–23. URL: <https://hal.inria.fr/hal-02111937> (cit. on p. 9).
- [102] P. Mertikopoulos, H. Nax and B. Pradelski. ‘Quick or cheap? Breaking points in dynamic markets’. In: *EC 2020 - 21st ACM Conference on Economics and Computation*. Budapest, Hungary, July 2020, pp. 1–32. URL: <https://hal.inria.fr/hal-03043688> (cit. on p. 11).
- [103] P. Mertikopoulos, C. H. Papadimitriou and G. Piliouras. ‘Cycles in adversarial regularized learning’. In: *SODA ’18 - Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*. New Orleans, United States, Jan. 2018, pp. 2703–2717. URL: <https://hal.archives-ouvertes.fr/hal-01643338> (cit. on p. 9).
- [104] P. Mertikopoulos and W. H. Sandholm. ‘Learning in games via reinforcement learning and regularization’. In: *Mathematics of Operations Research* 41.4 (Nov. 2016), pp. 1297–1324. DOI: 10.1287/moor.2016.0778. URL: <https://hal.inria.fr/hal-01073491> (cit. on p. 9, 10).
- [105] P. Mertikopoulos and W. H. Sandholm. ‘Riemannian game dynamics’. In: *Journal of Economic Theory* 177 (Sept. 2018), pp. 315–364. DOI: 10.1016/j.jet.2018.06.002. URL: <https://hal.archives-ouvertes.fr/hal-01382281> (cit. on p. 9).
- [106] M. Minaei, M. Mondal, P. Loiseau, K. P. Gummadi and A. Kate. ‘Forgetting the Forgotten with Lethe: Conceal Content Deletion from Persistent Observers’. In: *PETS 2019 - 19th Privacy Enhancing Technologies Symposium*. Stockholm, Sweden, July 2019, pp. 1–21. URL: <https://hal.archives-ouvertes.fr/hal-01959119> (cit. on p. 11).
- [107] W. Nagel, A. Arnold, M. Weber, H. Hoppe and K. Solchenbach. ‘VAMPIR: Visualization and Analysis of MPI Resources’. In: *Supercomputer* 12.1 (1996) (cit. on p. 6).
- [108] M. Oliu-Barton and B. Pradelski. *A vaccination policy by zones*. Other. Think tank Terra Nova, Oct. 2020. URL: <https://hal.archives-ouvertes.fr/hal-03100256> (cit. on p. 12).

- [109] M. Oliu-Barton and B. Pradelski. ‘Green bridges: Reconnecting Europe to avoid economic disaster’. In: *Europe in the Time of Covid-19*. 2020. URL: <https://hal.archives-ouvertes.fr/hal-03100282> (cit. on p. 11).
- [110] V. Pillet, J. Labarta, T. Cortes and S. Girona. ‘PARAVER: A tool to visualise and analyze parallel code’. In: *Proceedings of Transputer and Occam Developments, WOTUG-18*. Vol. 44. 1995 (cit. on p. 6).
- [111] B. Pradelski and M. Oliu-Barton. *Focus mass testing: How to overcome low test accuracy*. Other. Esade Centre for Economic Policy, Dec. 2020. URL: <https://hal.archives-ouvertes.fr/hal-03100250> (cit. on p. 12).
- [112] B. S. R. Pradelski and H. H. Nax. ‘Market sentiments and convergence dynamics in decentralized assignment economies’. In: *International Journal of Game Theory* 49.1 (Mar. 2020), pp. 275–298. DOI: [10.1007/s00182-019-00694-0](https://doi.org/10.1007/s00182-019-00694-0). URL: <https://hal.archives-ouvertes.fr/hal-03100116> (cit. on p. 11).
- [113] D. Reed, P. Roth, R. Aydt, K. Shields, L. Tavera, R. Noe and B. Schwartz. ‘Scalable performance analysis: the Pablo performance analysis environment’. In: *Scalable Parallel Libraries Conference*. 1993 (cit. on p. 6).
- [114] B. Shneiderman. ‘The eyes have it: A task by data type taxonomy for information visualizations’. In: *IEEE Symposium on Visual Languages*. IEEE. 1996 (cit. on p. 6).
- [115] D. C. Snowdon, S. Ruocco and G. Heiser. ‘Power Management and Dynamic Voltage Scaling: Myths and Facts’. In: *Proceedings of the 2005 Workshop on Power Aware Real-time Computing*. New Jersey, USA, Sept. 2005 (cit. on p. 12).
- [116] T. Speicher, M. Ali, G. Venkatadri, F. Ribeiro, G. Arvanitakis, F. Benevenuto, K. P. Gummadi, P. Loiseau and A. Mislove. ‘Potential for Discrimination in Online Targeted Advertising’. In: *FAT 2018 - Conference on Fairness, Accountability, and Transparency*. Vol. 81. New-York, United States, Feb. 2018, pp. 1–15. URL: <https://hal.archives-ouvertes.fr/hal-01955343> (cit. on p. 10).
- [117] M. Tikir, M. Laurenzano, L. Carrington and A. Snively. ‘PSINS: An Open Source Event Tracer and Execution Simulator for MPI Applications’. In: *Euro-Par*. 2009 (cit. on p. 6).
- [118] G. Venkatadri, A. Andreou, Y. Liu, A. Mislove, K. P. Gummadi, P. Loiseau and O. Goga. ‘Privacy Risks with Facebook’s PII-based Targeting: Auditing a Data Broker’s Advertising Interface’. In: *39th IEEE Symposium on Security and Privacy (S&P)*. Proceedings of the 39th IEEE Symposium on Security and Privacy (S&P). San Francisco, United States, 2018. URL: <https://hal.archives-ouvertes.fr/hal-01955327> (cit. on p. 11).
- [119] B. Vinot, F. Cadoux and R. Héliot. ‘Decentralized Optimization of Energy Exchanges in an Electricity Microgrid’. In: *ACM e-Energy 2016 - 7th ACM International Conference on Future Energy Systems*. Waterloo, Canada, June 2016. DOI: [10.1145/2939912.2939915](https://doi.org/10.1145/2939912.2939915). URL: <https://hal.archives-ouvertes.fr/hal-01340394> (cit. on p. 12).
- [120] B. Vinot, F. Cadoux and R. Héliot. ‘Decentralized optimization of energy exchanges in an electricity microgrid’. In: *2016 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*. Ljubljana, Slovenia: IEEE, Oct. 2016, pp. 1–6. DOI: [10.1109/ISGTEurope.2016.7856319](https://doi.org/10.1109/ISGTEurope.2016.7856319). URL: <https://hal.inria.fr/hal-01784148> (cit. on p. 12).
- [121] B. Vinot, F. Cadoux and N. Gast. ‘Congestion Avoidance in Low-Voltage Networks by using the Advanced Metering Infrastructure’. In: *ePerf 2018 - IFIP WG PERFORMANCE - 36th International Symposium on Computer Performance, Modeling, Measurements and Evaluation*. Toulouse, France, Dec. 2018, pp. 1–3. URL: <https://hal.inria.fr/hal-01953386> (cit. on p. 12).
- [122] M. Weiser, B. Welch, A. Demers and S. Shenker. ‘Scheduling for Reduced CPU Energy’. In: *Proceedings of the 1st USENIX Conference on Operating Systems Design and Implementation*. OSDI ’94. Monterey, California: USENIX Association, 1994, 2–es (cit. on p. 12).
- [123] J. Wilke, K. Sargsyan, J. Kenny, B. Debusschere, H. Najm and G. Hendry. ‘Validation and Uncertainty Assessment of Extreme-Scale HPC Simulation through Bayesian Inference’. In: *Euro-Par*. 2013 (cit. on p. 6).

-
- [124] O. Zaki, E. Lusk, W. Gropp and D. Swider. ‘Toward Scalable Performance Visualization with Jumpshot’. In: *International Journal of High Performance Computing Applications* 13.3 (1999). DOI: <http://dx.doi.org/10.1177/109434209901300310> (cit. on p. 6).
- [125] G. Zheng, G. Kakulapati and L. Kalé. ‘BigSim: A Parallel Simulator for Performance Prediction of Extremely Large Parallel Machines’. In: *IPDPS*. 2004 (cit. on p. 6).