RESEARCH CENTRE

Saclay - Île-de-France

2020

ACTIVITY REPORT

Project-Team

RANDOPT

Randomized Optimization

IN COLLABORATION WITH: Centre de Mathématiques Appliquées (CMAP)

DOMAIN

Applied Mathematics, Computation and Simulation

THEME

Optimization, machine learning and statistical methods

Contents

Project-Team RANDOPT

Creation of the Team: 2016 December 01, updated into Project-Team: 2019 January 01

Keywords

Computer sciences and digital sciences

A6.2. – Scientific computing, Numerical Analysis & Optimization

A8.2. – Optimization

A8.9. – Performance evaluation

Other research topics and application domains

B4.4. – Energy delivery

B9.5.1. – Computer science

B9.5.2. – Mathematics

B9.8. – Reproducibility

1 Team members, visitors, external collaborators

Research Scientists

- Anne Auger [Team leader, Inria, Senior Researcher, HDR]
- Dimo Brockhoff [Inria, Researcher]
- Nikolaus Hansen [Inria, Senior Researcher, HDR]

PhD Students

- Alann Cheral [École polytechnique]
- Marie Ange Dahito [Groupe PSA, CIFRE]
- Paul Dufosse [Thales, CIFRE]
- Eugenie Marescaux [École polytechnique]
- Cheikh Saliou Toure [Inria]
- Konstantinos Varelas [Thales, CIFRE]

Interns and Apprentices

- Armand Gissler [Ecole normale supérieure Paris-Saclay, from Apr 2020 until Aug 2020]
- Baptiste Plaquevent-Jourdain [Inria, from Mar 2020 until Aug 2020]
- Jingyun Yang [Inria, from Jun 2020 until Sep 2020]

Administrative Assistant

- Marie Enee [Inria]

2 Overall objectives

2.1 Scientific Context

Critical problems of the 21st century like the search for highly energy efficient or even carbon-neutral, and cost-efficient systems, or the design of new molecules against extensively drug-resistant bacteria crucially rely on the resolution of challenging numerical optimization problems. Such problems typically depend on noisy experimental data or involve complex numerical simulations where *derivatives are not useful or not available* and the function is seen as a *black-box*.

Many of those optimization problems are in essence *multiobjective*—one needs to optimize simultaneously several conflicting objectives like minimizing the cost of an energy network and maximizing its reliability—and most of the *challenging* black-box problems are *non-convex* and *non-smooth* and they combine difficulties related to ill-conditioning, non-separability, and ruggedness (a term that characterizes functions that can be non-smooth but also noisy or multi-modal). Additionally, the objective function can be expensive to evaluate, that is one function evaluation can take several minutes to hours (it can involve for instance a CFD simulation).

In this context, the use of randomness combined with proper adaptive mechanisms that particularly satisfy several invariance properties (affine invariance, invariance to monotonic transformations) has proven to be one key component for the design of robust global numerical optimization algorithms [43, 33].

The field of adaptive stochastic optimization algorithms has witnessed some important progress over the past 15 years. On the one hand, subdomains like medium-scale unconstrained optimization may be considered as "solved" (particularly, the CMA-ES algorithm, an instance of *Evolution Strategy* (ES) algorithms, stands out as state-of-the-art method) and considerably better standards have been established in the way benchmarking and experimentation are performed. On the other hand, multiobjective population-based stochastic algorithms became the method of choice to address multiobjective problems when a set of some best possible compromises is thought for. In all cases, the resulting algorithms have been naturally transferred to industry (the CMA-ES algorithm is now regularly used in companies such as Bosch, Total, ALSTOM, ...) or to other academic domains where difficult problems need to be solved such as physics, biology [47], geoscience [39], or robotics [41]).

Very recently, ES algorithms attracted quite some attention in Machine Learning with the OpenAI article *Evolution Strategies as a Scalable Alternative to Reinforcement Learning*. It is shown that the training time for difficult reinforcement learning benchmarks could be reduced from 1 day (with standard RL approaches) to 1 hour using ES [44].[1] A few years ago, another impressive application of CMA-ES, how "Computer Sim Teaches Itself To Walk Upright" (published at the conference SIGGRAPH Asia 2013) was presented in the press in the UK.

Several of these important advances around adaptive stochastic optimization algorithms are relying to a great extent on works initiated or achieved by the founding members of RandOpt, particularly related to the CMA-ES algorithm and to the Comparing Continuous Optimizer (COCO) platform.

Yet, the field of adaptive stochastic algorithms for black-box optimization is relatively young compared to the "classical optimization" field that includes convex and gradient-based optimization. For instance, the state-of-the art algorithms for unconstrained gradient based optimization like quasi-Newton methods (e.g. the BFGS method) date from the 1970s [30] while the stochastic derivative-free counterpart, CMA-ES dates from the early 2000s [31]. Consequently, in some subdomains with *important practical demands*, not even the most fundamental and basic questions are answered:

- This is the case of *constrained* optimization where one needs to find a solution $x^* \in \mathbb{R}^n$ minimizing a numerical function $\min_{x \in \mathbb{R}^n} f(x)$ while respecting a number of constraints m typically formulated as $g_i(x^*) \leq 0$ for $i = 1, \ldots, m$. Only recently, the fundamental requirement of linear convergence[2], as in the unconstrained case, has been clearly stated [22].

- In multiobjective optimization, most of the research so far has been focusing on *how to select candidate solutions from one iteration to the next one*. The difficult question of how to *generate* effectively new solutions is not yet answered in a proper way and we know today that simply applying operators from single-objective optimization may not be effective with the current best selection strategies. As a comparison, in the single-objective case, the question of selection of candidate solutions was already solved in the 1980s and 15 more years were needed to solve the trickier question of an effective adaptive strategy to generate new solutions.

- With the current demand to solve larger and larger optimization problems (e.g. in the domain of deep learning), optimization algorithms that scale linearly (in terms of internal complexity, memory and number of function evaluations to reach an ϵ-ball around the optimum) with the problem dimension are nowadays in increasing demand. Only recently, first proposals of how to reduce the quadratic scaling of CMA-ES have been made without a clear view of what can be achieved in the best case *in practice*. These later variants apply to optimization problems with thousands of variables. The question of designing randomized algorithms capable to handle problems with one or two orders of magnitude more variables effectively and efficiently is still largely open.

[1] The key behind such an improvement is the parallelization of the algorithm (on thousands of CPUs) that is done in such a way that the communication between the different workers is reduced to only exchanging a vector of permutation of small length (typically less than 100) containing the ranking of candidate solutions on the function to be optimized. In contrast, parallelization of backpropagation requires to exchange the gradient vector of the size of the problem (of the order of 10^6). This reduced communication time is a decisive factor for the impressive speedup.

[2] In optimization, linear convergence for an algorithm whose estimate of the optimum x^* of f at iteration t is denoted x_t, refers to a convergence where after a certain time (usually once the initialization is forgotten) the following typically holds: $\|x_{t+1} - x^*\| \leq c\|x_t - x^*\|$ where $c < 1$. This type of convergence is also called geometric. In the case of stochastic algorithms, there exist different definitions of linear convergence (depending on whether we consider the expectation of the sequence or we want a statement that holds with high probability) not strictly equivalent but that always translate the idea that the distance to the optimum at iteration $t+1$ is a fraction of the distance to the optimum at iteration t.

- For expensive optimization, standard methods are so called Bayesian optimization (BO) algorithms often based on Gaussian processes. Commonly used examples of BO algorithms are EGO [38], SMAC [36], Spearmint [45], or TPE [25] which are implemented in different libraries. Yet, our experience with a popular method like EGO is that many important aspects to come up with a good implementation rely on insider knowledge and are not standard across implementations. Two EGO implementations can differ for example in how they perform the initial design, which bandwidth for the Gaussian kernel is used, or which strategy is taken to optimize the expected improvement.

Additionally, the **development of stochastic adaptive methods for black-box optimization has been mainly driven by heuristics and practice**—rather than a general theoretical framework—validated by intensive computational simulations. Undoubtedly, **this has been an asset as the scope of possibilities for design was not restricted by mathematical frameworks** for proving convergence. In effect, powerful stochastic adaptive algorithms for **unconstrained optimization** like the CMA-ES algorithm emerged from this approach. At the same time, naturally, **theory strongly lags behind practice**. For instance, the striking performances of CMA-ES empirically observed contrast with how little is theoretically proven on the method. This situation is clearly not satisfactory. On the one hand, theory generally lifts performance assessment from an empirical level to a conceptual one, rendering results independent from the problem instances where they have been obtained. On the other hand, theory typically provides insights that change perspectives on some algorithm components. Also theoretical guarantees generally increase the trust in the reliability of a method and facilitate the task to make it accepted by wider communities.

Finally, as discussed above, the development of novel black-box algorithms strongly relies on scientific experimentation, and it is quite difficult to conduct proper and meaningful experimental analysis. This is well known for more than two decades now and summarized in this quote from Johnson in 1996

> *"the field of experimental analysis is fraught with pitfalls. In many ways, the implementation of an algorithm is the easy part. The hard part is successfully using that implementation to produce meaningful and valuable (and publishable!) research results."* [37]

Since then, quite some progress has been made to set better standards in conducting scientific experiments and benchmarking. Yet, some domains still suffer from poor benchmarking standards and from the generic problem of the lack of reproducibility of results. For instance, in multiobjective optimization, it is (still) not rare to see comparisons between algorithms made by solely visually inspecting Pareto fronts after a fixed budget. In Bayesian optimization, good performance seems often to be due to insider knowledge not always well described in papers.

In the context of black-box numerical optimization previously described, the scientific positioning of the RandOpt ream is at the intersection between theory, algorithm design, and applications. Our vision is that the field of stochastic black-box optimization should reach the same level of maturity than gradient-based convex mathematical optimization. This entails major algorithmic developments for constrained, multiobjective and large-scale black-box optimization and major theoretical developments for analyzing current methods including the state-of-the-art CMA-ES.

The specificity in black-box optimization is that methods are intended to solve problems characterized by "*non*-properties"—*non*-linear, *non*-convex, *non*-smooth, *non*-Lipschitz. This contrasts with gradient-based optimization and poses on the one hand some challenges when developing theoretical frameworks but also makes it compulsory to complement theory with empirical investigations.

Our ultimate goal is to provide software that is useful for practitioners. We see that theory is a means for this end (rather than an end in itself) and we also firmly belief that parameter tuning is part of the algorithm designer's task.

This shapes, on the one hand, four main scientific objectives for our team:

1. **develop novel theoretical frameworks** for guiding (a) the design of novel black-box methods and (b) their analysis, allowing to

2. provide **proofs of key features** of stochastic adaptive algorithms including the state-of-the-art method CMA-ES: linear convergence and learning of second order information.

3. develop **stochastic numerical black-box algorithms** following a **principled design** in domains with a strong practical need for much better methods namely **constrained, multiobjective, large-scale and expensive optimization**. Implement the methods such that they are easy to use. And finally, to

4. **set new standards in scientific experimentation, performance assessment and benchmarking** both for optimization on continuous or combinatorial search spaces. This should allow in particular to advance the state of **reproducibility of results of scientific papers** in optimization.

On the other hand, the above motivates our objectives with respect to dissemination and transfer:

1. develop software packages that people can directly use to solve their problems. This means having carefully thought out interfaces, generically applicable setting of parameters and termination conditions, proper treatment of numerical errors, catching properly various exceptions, etc.;

2. have direct collaborations with industrials;

3. publish our results both in applied mathematics and computer science bridging the gap between very often disjoint communities.

3 Research program

The lines of research we intend to pursue is organized along four axis namely developing novel theoretical framework, developing novel algorithms, setting novel standards in scientific experimentation and benchmarking and applications.

3.1 Developing Novel Theoretical Frameworks for Analyzing and Designing Adaptive Stochastic Algorithms

Stochastic black-box algorithms typically optimize **non-convex, non-smooth functions**. This is possible because the algorithms rely on weak mathematical properties of the underlying functions: the algorithms do not use the derivatives—hence the function does not need to be differentiable—and, additionally, often do not use the exact function value but instead how the objective function ranks candidate solutions (such methods are sometimes called function-value-free). (To illustrate a comparison-based update, consider an algorithm that samples λ (with λ an even integer) candidate solutions from a multivariate normal distribution. Let $x_1, \ldots, x_\lambda$ in $\mathbb{R}^n$ denote those λ candidate solutions at a given iteration. The solutions are evaluated on the function f to be minimized and ranked from the best to the worse:

$$f(x_{1:\lambda}) \leq \ldots \leq f(x_{\lambda:\lambda}) \enspace .$$

In the previous equation $i\!:\!\lambda$ denotes the index of the sampled solution associated to the i-th best solution. The new mean of the Gaussian vector from which new solutions will be sampled at the next iteration can be updated as

$$m \leftarrow \frac{4}{\lambda} \sum_{i=1}^{\lambda/4} x_{i:\lambda} \enspace .$$

The previous update moves the mean towards the $\lambda/2$ best solutions. Yet the update is only based on the ranking of the candidate solutions such that the update is the same if f is optimized or $g \circ f$ where $g : \mathrm{Im}(f) \to \mathbb{R}$ is strictly increasing. Consequently, such algorithms are invariant with respect to strictly increasing transformations of the objective function. This entails that they are robust and their performances generalize well.)

Additionally, adaptive stochastic optimization algorithms typically have a **complex state space** which encodes the parameters of a probability distribution (e.g. mean and covariance matrix of a Gaussian vector) and other state vectors. This state-space is a **manifold**. While the algorithms are Markov chains, the complexity of the state-space makes that **standard Markov chain theory tools do not directly apply**. The same holds with tools stemming from stochastic approximation theory or Ordinary Differential

Equation (ODE) theory where it is usually assumed that the underlying ODE (obtained by proper averaging and limit for learning rate to zero) has its critical points inside the search space. In contrast, in the cases we are interested in, the **critical points of the ODEs are at the boundary of the domain**.

Last, since we aim at developing theory that on the one hand allows to analyze the main properties of state-of-the-art methods and on the other hand is useful for algorithm design, we need to be careful not to use simplifications that would allow a proof to be done but would not capture the important properties of the algorithms. With that respect one tricky point is to develop **theory that accounts for invariance properties**.

To face those specific challenges, we need to develop novel theoretical frameworks exploiting invariance properties and accounting for peculiar state-spaces. Those frameworks should allow researchers to analyze one of the core properties of adaptive stochastic methods, namely **linear convergence** on the widest possible class of functions.

We are planning to approach the question of linear convergence from three different complementary angles, using three different frameworks:

- the Markov chain framework where the convergence derives from the analysis of the stability of a normalized Markov chain existing on scaling-invariant functions for translation and scale-invariant algorithms [24]. This framework allows for a fine analysis where the exact convergence rate can be given as an implicit function of the invariant measure of the normalized Markov chain. Yet it requires the objective function to be scaling-invariant. The stability analysis can be particularly tricky as the Markov chain that needs to be studied writes as $\Phi_{t+1} = F(\Phi_t, W_{t+1})$ where $\{W_t : t > 0\}$ are independent identically distributed and F is typically discontinuous because the algorithms studied are comparison-based. This implies that practical tools for analyzing a standard property like irreducibility, that rely on investigating the stability of underlying deterministic control models [42, Chapter 7], cannot be used. Additionally, the construction of a drift to prove ergodicity is particularly delicate when the state space includes a (normalized) covariance matrix as it is the case for analyzing the CMA-ES algorithm.

- The stochastic approximation or ODE framework. Those are standard techniques to prove the convergence of stochastic algorithms when an algorithm can be expressed as a stochastic approximation of the solution of a mean field ODE [27, 26, 40]. What is specific and induces difficulties for the algorithms we aim at analyzing is the **non-standard state-space** since the ODE variables correspond to the state-variables of the algorithm (e.g. $\mathbb{R}^n \times \mathbb{R}_{>0}$ for step-size adaptive algorithms, $\mathbb{R}^n \times \mathbb{R}_{>0} \times S^n_{++}$ where S^n_{++} denotes the set of positive definite matrices if a covariance matrix is additionally adapted). Consequently, the ODE can have many critical points at the boundary of its definition domain (e.g. all points corresponding to $\sigma_t = 0$ are critical points of the ODE) which is not typical. Also we aim at proving **linear convergence**, for that it is crucial that the learning rate does not decrease to zero which is non-standard in ODE method.

- The direct framework where we construct a global Lyapunov function for the original algorithm from which we deduce bounds on the hitting time to reach an ϵ-ball of the optimum. For this framework as for the ODE framework, we expect that the class of functions where we can prove linear convergence are composite of $g \circ f$ where f is differentiable and $g : \mathrm{Im}(f) \to \mathbb{R}$ is strictly increasing and that we can show convergence to a local minimum.

We expect those frameworks to be complementary in the sense that the assumptions required are different. Typically, the ODE framework should allow for proofs under the assumptions that learning rates are small enough while it is not needed for the Markov chain framework. Hence this latter framework captures better the real dynamics of the algorithm, yet under the assumption of scaling-invariance of the objective functions. Also, we expect some overlap in terms of function classes that can be studied by the different frameworks (typically convex-quadratic functions should be encompassed in the three frameworks). By studying the different frameworks in parallel, we expect to gain synergies and possibly understand what is the most promising approach for solving the holy grail question of the linear convergence of CMA-ES. We foresee for instance that similar approaches like the use of Foster-Lyapunov drift conditions are needed in all the frameworks and that intuition can be gained on how to establish the conditions from one framework to another one.

3.2 Algorithmic developments

We are planning on developing algorithms in the subdomains with strong practical demand for better methods of constrained, multiobjective, large-scale and expensive optimization.

Many of the algorithm developments, we propose, rely on the CMA-ES method. While this seems to restrict our possibilities, we want to emphasize that CMA-ES became a *family of methods* over the years that nowadays include various techniques and developments from the literature to handle non-standard optimization problems (noisy, large-scale, ...). The core idea of all CMA-ES variants—namely the mechanism to adapt a Gaussian distribution—has furthermore been shown to derive naturally from first principles with only minimal assumptions in the context of derivative-free black-box stochastic optimization [43, 33]. This is a strong justification for relying on the CMA-ES premises while new developments naturally include new techniques typically borrowed from other fields. While CMA-ES is now a full family of methods, for visibility reasons, we continue to refer often to “the CMA-ES algorithm”.

3.2.1 Constrained optimization

Many (real-world) optimization problems have constraints related to technical feasibility, cost, etc. Constraints are classically handled in the black-box setting either via rejection of solutions violating the constraints—which can be quite costly and even lead to quasi-infinite loops—or by penalization with respect to the distance to the feasible domain (if this information can be extracted) or with respect to the constraint function value [28]. However, the penalization coefficient is a sensitive parameter that needs to be adapted in order to achieve a robust and general method [29]. Yet, **the question of how to handle properly constraints is largely unsolved**. Previous constraints handling for CMA-ES were ad-hoc techniques driven by many heuristics [29]. Also, only recently it was pointed out that **linear convergence properties should be preserved** when addressing constraint problems [22].

Promising approaches though, rely on using augmented Lagrangians [22, 23]. The augmented Lagrangian, here, is the objective function optimized by the algorithm. Yet, it depends on coefficients that are adapted online. The adaptation of those coefficients is the difficult part: the algorithm should be stable and the adaptation efficient. We believe that the theoretical frameworks developed (particularly the Markov chain framework) will be useful to understand how to design the adaptation mechanisms. Additionally, the question of invariance will also be at the core of the design of the methods: augmented Lagrangian approaches break the invariance to monotonic transformation of the objective functions, yet understanding the maximal invariance that can be achieved seems to be an important step towards understanding what adaptation rules should satisfy.

3.2.2 Large-scale Optimization

In the large-scale setting, we are interested to optimize problems with the order of 10^3 to 10^4 variables. For one to two orders of magnitude more variables, we will talk about a “very large-scale” setting.

In this context, algorithms with a quadratic scaling (internal and in terms of number of function evaluations needed to optimize the problem) cannot be afforded. In CMA-ES-type algorithms, we typically need to restrict the model of the covariance matrix to have only a linear number of parameters to learn such that the algorithms scale linearly in terms of internal complexity, memory and number of function evaluations to solve the problem. The main challenge is thus to have rich enough models for which we can efficiently design proper adaptation mechanisms. Some first large-scale variants of CMA-ES have been derived. They include the online adaptation of the complexity of the model [21, 20]. Yet, the type of Hessian matrices they can learn is restricted and not fully satisfactory. Different restricted families of distributions are conceivable and it is an open question which can be effectively learned and which are the most promising in practice.

Another direction, we want to pursue, is exploring the use of large-scale variants of CMA-ES to solve reinforcement learning problems [44].

Last, we are interested to investigate the very-large-scale setting. One approach consists in doing optimization in subspaces. This entails the efficient identification of relevant spaces and the restriction of the optimization to those subspaces.

3.2.3 Multiobjective Optimization

Multiobjective optimization, i.e., the simultaneous optimization of multiple objective functions, differs from single-objective optimization in particular in its optimization goal. Instead of aiming at converging to the solution with the best possible function value, in multiobjective optimization, a set of solutions [3] is sought. This set, called Pareto-set, contains all trade-off solutions in the sense of Pareto-optimality—no solution exists that is better in *all* objectives than a Pareto-optimal one. Because converging towards a set differs from converging to a single solution, it is no surprise that we might lose many good convergence properties if we directly apply search operators from single-objective methods. However, this is what has typically been done so far in the literature. Indeed, most of the research in stochastic algorithms for multiobjective optimization focused instead on the so called selection part, that decides which solutions should be kept during the optimization—a question that can be considered as solved for many years in the case of single-objective stochastic adaptive methods.

We therefore aim at rethinking search operators and adaptive mechanisms to improve existing methods. We expect that we can obtain orders of magnitude better convergence rates for certain problem types if we choose the right search operators. We typically see two angles of attack: On the one hand, we will study methods based on scalarizing functions that transform the multiobjective problem into a set of single-objective problems. Those single-objective problems can then be solved with state-of-the-art single-objective algorithms. Classical methods for multiobjective optimization fall into this category, but they all solve multiple single-objective problems subsequently (from scratch) instead of dynamically changing the scalarizing function during the search. On the other hand, we will improve on currently available population-based methods such as the first multiobjective versions of the CMA-ES. Here, research is needed on an even more fundamental level such as trying to understand success probabilities observed during an optimization run or how we can introduce non-elitist selection (the state of the art in single-objective stochastic adaptive algorithms) to increase robustness regarding noisy evaluations or multi-modality. The challenge here, compared to single-objective algorithms, is that the quality of a solution is not anymore independent from other sampled solutions, but can potentially depend on all known solutions (in the case of three or more objective functions), resulting in a more noisy evaluation as the relatively simple function-value-based ranking within single-objective optimizers.

3.2.4 Expensive Optimization

In the so-called expensive optimization scenario, a single function evaluation might take several minutes or even hours in a practical setting. Hence, the available budget in terms of number of function evaluation calls to find a solution is very limited in practice. To tackle such expensive optimization problems, it is needed to exploit the first few function evaluations in the best way. To this end, typical methods couple the learning of a surrogate (or meta-model) of the expensive objective function with traditional optimization algorithms.

In the context of expensive optimization and CMA-ES, which usually shows its full potential when the number n of variables is not too small (say larger than 3) and if the number of available function evaluations is about $100n$ or larger, several research directions emerge. The two main possibilities to integrate meta-models into the search with CMA-ES type algorithms are (i) the successive injection of the minimum of a learned meta-model at each time step into the learning of CMA-ES's covariance matrix and (ii) the use of a meta-model to predict the internal ranking of solutions. While for the latter, first results exist, the former idea is entirely unexplored for now. In both cases, a fundamental question is which type of meta-model (linear, quadratic, Gaussian Process, ...) is the best choice for a given number of function evaluations (as low as one or two function evaluations) and at which time the type of the meta-model shall be switched.

3.3 Setting novel standards in scientific experimentation and benchmarking

Numerical experimentation is needed as a complement to theory to test novel ideas, hypotheses, the stability of an algorithm, and/or to obtain quantitative estimates. Optimally, theory and experimentation go hand in hand, jointly guiding the understanding of the mechanisms underlying optimization algorithms.

[3] Often, this set forms a manifold of dimension one smaller than the number of objectives.

Though performing numerical experimentation on optimization algorithms is crucial and a common task, it is non-trivial and easy to fall in (common) pitfalls as stated by J. N. Hooker in his seminal paper [35].

In the RandOpt team we aim at raising the standards for both scientific experimentation and benchmarking.

On the experimentation aspect, we are convinced that there is common ground over how scientific experimentation should be done across many (sub-)domains of optimization, in particular with respect to the visualization of results, testing extreme scenarios (parameter settings, initial conditions, etc.), how to conduct understandable and small experiments, how to account for invariance properties, performing scaling up experiments and so forth. We therefore want to formalize and generalize these ideas in order to make them known to the entire optimization community with the final aim that they become standards for experimental research.

Extensive numerical benchmarking, on the other hand, is a compulsory task for evaluating and comparing the performance of algorithms. It puts algorithms to a standardized test and allows to make recommendations which algorithms should be used preferably in practice. To ease this part of optimization research, we have been developing the Comparing Continuous Optimizers platform (COCO) since 2007 which allows to automatize the tedious task of benchmarking. It is a game changer in the sense that the freed time can now be spent on the scientific part of algorithm design (instead of implementing the experiments, visualization, statistical tests, etc.) and it opened novel perspectives in algorithm testing. COCO implements a thorough, well-documented methodology that is based on the above mentioned general principles for scientific experimentation.

Also due to the freely available data from 300+ algorithms benchmarked with the platform, COCO became a quasi-standard for single-objective, noiseless optimization benchmarking. It is therefore natural to extend the reach of COCO towards other subdomains (particularly constrained optimization, many-objective optimization) which can benefit greatly from an automated benchmarking methodology and standardized tests without (much) effort. This entails particularly the design of novel test suites and rethinking the methodology for measuring performance and more generally evaluating the algorithms. Particularly challenging is the design of scalable non-trivial testbeds for constrained optimization where one can still control where the solutions lies. Other optimization problem types, we are targeting are expensive problems (and the Bayesian optimization community in particular, see our AESOP project), optimization problems in machine learning (for example parameter tuning in reinforcement learning), and the collection of real-world problems from industry.

Another aspect of our future research on benchmarking is to investigate the large amounts of benchmarking data, we collected with COCO during the years. Extracting information about the influence of algorithms on the best performing portfolio, clustering algorithms of similar performance, or the automated detection of anomalies in terms of good/bad behavior of algorithms on a subset of the functions or dimensions are some of the ideas here.

Last, we want to expand the focus of COCO from automatized (large) benchmarking experiments towards everyday experimentation, for example by allowing the user to visually investigate algorithm internals on the fly or by simplifying the set up of algorithm parameter influence studies.

4 Application domains

Applications of black-box algorithms occur in various domains. Industry but also researchers in other academic domains have a great need to apply black-box algorithms on a daily basis. Generally, we do not target a specific application domain and are interested in black-box applications stemming from various origins. This is to us intrinsic to the nature of the methods we develop that are general purpose algorithms. Hence our strategy with respect to applications can be considered as opportunistic and our main selection criteria when approached by colleagues who want to develop a collaboration around an application is whether we find the application interesting and valuable: that means the application brings new challenges and/or gives us the opportunity to work on topics we already intended to work on, and it brings, in our judgement, an advancement to society in the application domain.

The concrete applications related to industrial collaborations we are currently dealing with are:

- With Thales for the theses of Konstantinos Varelas and Paul Dufossé (DGA-CIFRE theses) related to the design of radars (shape optimization of the wave form). They investigate more specifically the development of large-scale variants of CMA-ES and constrained-handling for CMA-ES, respectively.

- With Storengy, a subsidiary of the ENGIE group, specialized in gas storage for the thesis of Cheikh Touré. Different multiobjective applications are considered in this context but the primary motivation of Storengy is to get at their disposal a better multiobjective variant of CMA-ES which is the main objective of the developments within the thesis.
- With PSA in the context of the OpenLab and the thesis of Marie-Ange Dahito for the design of part of a car body.
- With Onera in the context of the thesis of Alann Cheral related to the optimization of the choice of hyperspectral bandwidth.

5 Highlights of the year

5.1 Awards

Nikolaus Hansen and Anne Auger received the SIGEVO impact award for the article "Comparing Results of 31 Algorithms from the Black-Box Optimization Benchmarking BBOB-2009" that was published in 2010 [34]. This impact award distinguishes work that has been published at the GECCO conference ten years ago and made a significant impact. Through this award the impact of the COCO platform, developed first within the TAO team and then the RandOpt team, has been acknowledged. Details have been presented in the article *"A SIGEVO impact award for a paper arising from the COCO platform: a summary and beyond"* [19].

5.2 Software

We want to highlight the impact and transfer of our work through the following: the main algorithm developed in the team for single-objective optimization, CMA-ES, has been implemented in a lightweight version for Python 3 by Masashi Shibata (not a RANDOPT team member). The implementation is based on publications of the team members and notably on [32]. By the end of the year 2020, the code release on PyPI was being downloaded 150,000 times *per month*.

6 New software and platforms

6.1 New software

6.1.1 COCO

Name: COmparing Continuous Optimizers

Keywords: Benchmarking, Numerical optimization, Black-box optimization, Stochastic optimization

Scientific Description: COmparing Continuous Optimisers (COCO) is a tool for benchmarking algorithms for black-box optimisation. COCO facilitates systematic experimentation in the field of continuous optimization. COCO provides: (1) an experimental framework for testing the algorithms, (2) post-processing facilities for generating publication quality figures and tables, including the easy integration of data from benchmarking experiments of 300+ algorithm variants, (3) LaTeX templates for scientific articles and HTML overview pages which present the figures and tables.

The COCO software is composed of two parts: (i) an interface available in different programming languages (C/C++, Java, Matlab/Octave, Python, external support for R) which allows to run and log experiments on several function test suites (unbounded noisy and noiseless single-objective functions, unbounded noiseless multiobjective problems, constrained problems) are provided (ii) a Python tool for generating figures and tables that can be looked at in every web browser and that can be used in the provided LaTeX templates to write scientific papers.

Functional Description: The Coco platform aims at supporting the numerical benchmarking of black-box optimization algorithms in continuous domains. Benchmarking is a vital part of algorithm engineering and a necessary path to recommend algorithms for practical applications. The Coco platform releases algorithm developers and practitioners alike from (re-)writing test functions, logging, and plotting facilities by providing an easy-to-handle interface in several programming languages. The Coco platform has been developed since 2007 and has been used extensively within the "Blackbox Optimization Benchmarking (BBOB)" workshop series since 2009. Overall, 300+ algorithms and algorithm variants by contributors from all over the world have been benchmarked on the platform's supported test suites so far. The most recent extensions have been towards large-scale as well as mixed-integer problems.

URL: https://github.com/numbbo/coco

Contacts: Anne Auger, Dimo Brockhoff

Participants: Anne Auger, Asma Atamna, Dejan Tusar, Dimo Brockhoff, Marc Schoenauer, Nikolaus HANSEN, Ouassim Ait Elhara, Raymond Ros, Tea Tusar, Thanh-Do Tran, Umut Batu, Konstantinos Varelas

Partners: TU Dortmund University, Charles University Prague, Jozef Stefan Institute (JSI)

6.1.2 CMA-ES

Name: Covariance Matrix Adaptation Evolution Strategy

Keywords: Numerical optimization, Black-box optimization, Stochastic optimization

Scientific Description: The CMA-ES is considered as state-of-the-art in evolutionary computation and has been adopted as one of the standard tools for continuous optimisation in many (probably hundreds of) research labs and industrial environments around the world. The CMA-ES is typically applied to unconstrained or bounded constraint optimization problems, and search space dimensions between three and a hundred. The method should be applied, if derivative based methods, e.g. quasi-Newton BFGS or conjugate gradient, (supposedly) fail due to a rugged search landscape (e.g. discontinuities, sharp bends or ridges, noise, local optima, outliers). If second order derivative based methods are successful, they are usually faster than the CMA-ES: on purely convex-quadratic functions, f(x)=xTHx, BFGS (Matlabs function fminunc) is typically faster by a factor of about ten (in terms of number of objective function evaluations needed to reach a target function value, assuming that gradients are not available). On the most simple quadratic function f(x)=||x||2=xTx BFGS is faster by a factor of about 30.

Functional Description: The CMA-ES is an evolutionary algorithm for difficult non-linear non-convex black-box optimisation problems in continuous domain.

URL: http://cma.gforge.inria.fr/cmaes_sourcecode_page.html

Contacts: Nikolaus HANSEN, Anne Auger

Participant: Nikolaus HANSEN

6.1.3 COMO-CMA-ES

Name: Comma Multi-Objective Covariance Matrix Adaptation Evolution Strategy

Keywords: Black-box optimization, Global optimization, Multi-objective optimisation

Scientific Description: The CMA-ES is considered as state-of-the-art in evolutionary computation and has been adopted as one of the standard tools for continuous optimisation in many (probably hundreds of) research labs and industrial environments around the world. The CMA-ES is typically applied to unconstrained or bounded constraint optimization problems, and search space dimensions between three and a hundred. COMO-CMA-ES is a multi-objective optimization algorithm based on the standard CMA-ES using the Uncrowded Hypervolume Improvement within the so-called Sofomore framework.

Functional Description: The COMO-CMA-ES is an evolutionary algorithm for difficult non-linear non-convex black-box optimisation problems with several (two) objectives in continuous domain.

URL: `https://github.com/CMA-ES/pycomocma`

Contacts: Nikolaus HANSEN, Dimo Brockhoff

6.1.4 MOarchiving

Name: Multiobjective Optimization Archiving Module

Keywords: Mathematical Optimization, Multi-objective optimisation

Scientific Description: Multi-objective optimization relies on the maintenance of a set of non-dominated (and hence incomparable) solutions. Performance indicator computations and in particular the computation of the hypervolume indicator is based on this solution set. The hypervolume computation and the update of the set of non-dominated solutions are generally time critical operations. The module computes the bi-objective hypervolume in linear time and updates the non-dominated solution set in logarithmic time.

Functional Description: The module implements a bi-objective non-dominated archive using a Python list as parent class. The main functionality is heavily based on the bisect module. The class provides easy and fast access to the overall hypervolume, the contributing hypervolume of each element, and to the uncrowded hypervolume improvement of any given point in objective space.

URL: `https://github.com/CMA-ES/moarchiving`

Contacts: Nikolaus HANSEN, Dimo Brockhoff

7 New results

7.1 Analysis of Adaptive Stochastic Search Algorithms

Participants: Anne Auger, Nikolaus Hansen, Cheikh Touré, Armand Gissler
External collaborators: Youhei Akimoto (Tsukuba University), Tobias Glasmachers (Ruhr University, Bochum), Asma Atamna (Telecom Paris)

Central to the design of adaptive Evolution Strategies are the questions of invariance and of linear convergence. In the case of constraints handled with augmented Lagrangians, sufficient conditions for linear convergence have been proposed in [10]. Quantitative estimates of convergence rates on convex-quadratic functions of algorithms using weighted recombination have been published in [8].

We have proven the global linear convergence of the (1+1)-ES with one-fifth success rule as step-size adaptation on a class of functions that embed smooth strongly convex functions and positively homogeneous functions. Because of the invariance to monotonic transformations, the study holds for non-continuous and non-convex functions. Arguably, our study provides the first proof of the linear convergence of an adaptive evolution strategy without modifying its underlying updates (to make a proof work) on such a wide class of functions [16].

Over the past years, we have developed a methodology to analyze the linear convergence of adaptive comparison-based algorithms including Evolution Strategies by studying the stability of underlying Markov chains. This methodology allows to derive convergence on so-called scaling-invariant functions. Yet this class of functions has not been studied in the past such that we needed to derive important mathematical properties that are needed to conduct our convergence studies. In his internship, Armand Gissler derived key results connecting scaling-invariant functions to positively homogeneous ones. Those results complemented results by Cheikh Touré and are now presented in an article [18].

7.2 Large-scale black-box optimization

Participants: Konstantinos Varelas, Anne Auger, Nikolaus Hansen
External collaborators: Youhei Akimoto (Tsukuba University)

One objective of the team is to advance research on black-box stochastic optimization towards large-scale problems. In the context of black-box optimization without derivatives, large-scale starts when the number of variables to be optimized is of the order of a few hundred. We have designed a variant of CMA-ES capable to exploit separability of the functions while keeping the ability to optimize fully non-separable functions. This results in the DD-CMA-ES algorithm [9].

In his thesis, Konstantinos Varelas has explored different approaches to learn sparsity of a problem and increase the learning rates so as to reduce the number of function evaluations to solve a problem. In contrast to many large-scale variants of CMA-ES, the main idea was to not assume a fixed structure for the covariance matrix to be learned. On the one hand, he used graphical Lasso to learn a sparse precision matrix, leading to interesting improvements assuming a proper choice of the penalization coefficient [14]. On the other hand, he also explored thresholding as well as a method used for learning Markov decision processes.

7.3 Applications to Radar

Participants: Konstantinos Varelas, Paul Dufosse, Nikolaus Hansen
External collaborators: THALES and notably Yann Semet, Rami Kassab, Frédéric Barbaresco, Cyrille Enderli

Two of the theses in the team are funded by Thales and DGA interested to improve the resolution of different optimization problems related to the design of antenna and Radar. In this context, the design of phased-array antenna pattern with CMA-ES has been investigated [13].

For the design of new generation digital multi-missions radars, THALES has proposed a framework using in particular CMA-ES as optimization module [15].

In the article [17], different many-objective optimization solvers are empirically compared to approach the problem of finding optimal Pulse Repetion Intervals.

7.4 Multi-objective optimization

Participants: Cheikh Touré, Eugénie Marescaux, Baptiste Plaquevent-Jourdain, Anne Auger, Dimo Brockhoff, Nikolaus Hansen
External collaborators: Youhei Akimoto (Tsukuba University)

A central theme for the team is the design of multi-objective optimization algorithms. We have worked on building algorithms that converge to the entire Pareto-front, both using gradients of the objectives and in a derivative-free mode. Those algorithms are based on the idea to incrementally approximate the Pareto set by single-objective optimization of an improved variant of the hypervolume indicator, called uncrowded hypervolume improvement (UHVI, [46]). One first class of algorithms based on quasi-Newton techniques have been developed, implemented, and experimentally compared on the bi-objective test functions of the COCO platform in the context of the internship of Baptiste Plaquevent-Jourdain. In parallel, we have worked on extending the COMO multi-objective solver. The publications related to those works are in the process to be finalized.

In parallel, Eugénie Marescaux studied the convergence of the ensuing solver and she was able to prove under proper assumptions that they converge towards the entire Pareto front.

Finally, we have released two Python modules related to multi-objective optimization: the MO-archiving module and the module implementing the COMO-CMA-ES algorithm. They are both described in the New Software section.

7.5 Benchmarking: methodology and the Comparing Continuous Optimziers Platform (COCO)

Participants: Anne Auger, Dimo Brockhoff, Nikolaus Hansen, Konstantinos Varelas

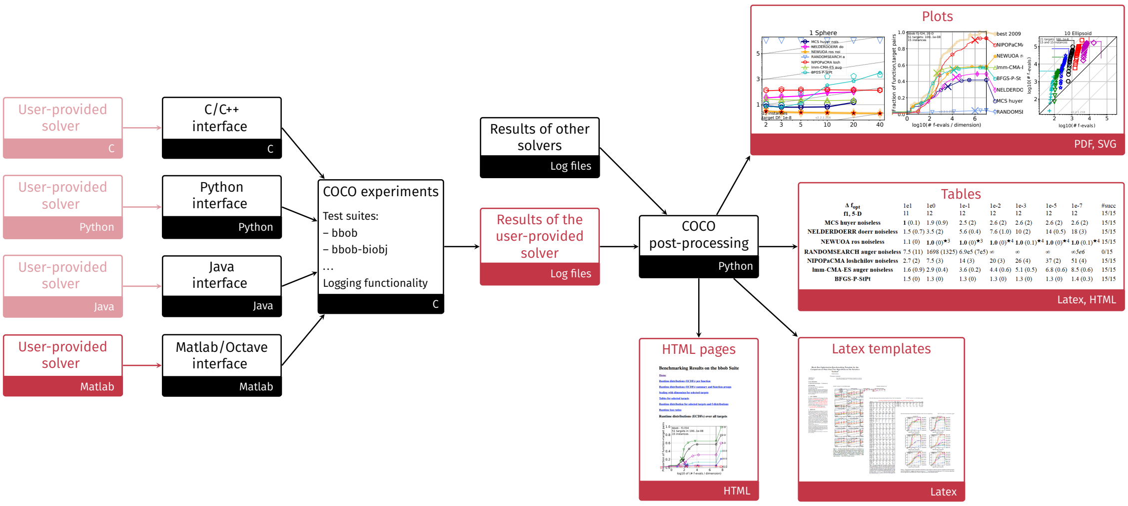

Figure 1: Structural overview of the COCO platform. COCO provides all black parts while users only have to connect their solver to the COCO interface in the language of interest, here for instance Matlab, and to decide on the test suite the solver should run on. The other red components show the output of the experiments (number of function evaluations to reach certain target precisions) and their post-processing and are automatically generated.

External collaborators: Olaf Mersmann (TH Köln), Raymond Ros (U Paris-Saclay), Tea Tušar (Jozef Stefan Institute)

Benchmarking is an important task in optimization in order to assess and compare the performance of algorithms as well as to motivate the design of better solvers. We are leading the benchmarking of derivative free solvers in the context of difficult problems: we have been developing methodologies and testbeds as well as assembled this into a platform automatizing the benchmarking process. This is a continuing effort that we are pursuing in the team.

7.5.1 Methodology

The main innovations and methodological ideas of the platform have been published in the paper [11].

Building on the BBOB testbed of the COCO platform, the bbob-largescale testbed has been finalized, released and the underlying methodology to build the testbed has been published [12].

7.5.2 The COCO platform

The COCO platform, developed at Inria first in the TAO team and then in Randopt since 2007, aims at automatizing numerical benchmarking experiments and the visual presentation of their results. The platform consists of an experimental part to generate benchmarking data (in various programming languages) and a postprocessing module (in python), see Figure 1. At the interface between the two, we provide data sets from numerical experiments of 300+ algorithms and algorithm variants from various fields (quasi-Newton, derivative-free optimization, evolutionary computing, Bayesian optimization) and for various problem characteristics (noiseless/noisy optimization, single-/multi-objective optimization, continuous/mixed-integer, ...).

We have been using the platform in the past to initiate workshop papers during the ACM-GECCO conference as well as to collect algorithm data sets from the entire optimization community. The next workshop in this series is going to take place in 2021[4].

In this context, we constantly improve and extend the software and the year 2020 was no exception. Overall, 57 issues have been closed in 2020. One big change, made in 2020, was to allow for a continuous submission of data sets from benchmarking experiments to our data archive. Since the official start in

[4] See https://numbbo.github.io/workshops/BBOB-2021/

November, we have received five data sets already. New visualizations of our biobjective functions have been made available as well on a new webpage at https://numbbo.github.io/bbob-biobj/vis/.

8 Bilateral contracts and grants with industry

8.1 Bilateral contracts with industry

- Contract with the company Storengy partially funding the PhD thesis of Cheikh Touré (2017–2020)
- Contract with Thales in the context of the CIFRE PhD thesis of Konstantinos Varelas (2017–2020)
- Contract with PSA (now Stellantis) in the context of the CIFRE PhD thesis of Marie-Ange Dahito (2018–2021)
- Contract with Thales for the CIFRE PhD thesis of Paul Dufossé (2018–2021)

9 Partnerships and cooperations

9.1 International initiatives

9.1.1 Inria international partners

Informal international partners

- Tea Tušar, Jozef Stefan Institute (JSI), Ljubljana, Slovenia
- Tobias Glasmachers, Ruhr Universität Bochum, Bochum, Germany
- Youhei Akimoto, University of Tsukuba, Tsukuba, Japan

9.2 National initiatives

ANR

- CIROQUO ("Consortium Industriel de Recherche en Optimisation et QUantification d'incertitudes pour les données Onéreuses"), participation as Inria Saclay/Ecole Polytechnique, together with six other academic and five industrial partners
- ANR project "Big Multiobjective Optimization (BigMO)", Dimo Brockhoff participates in this project through the Inria team BONUS in Lille (2017–2020)

10 Dissemination

10.1 Promoting scientific activities

10.1.1 Scientific events: organisation

- Anne Auger: co-organizer of Dagstuhl seminar on Benchmarking originally planned for 2021 and postponed.
- Anne Auger: co-organizer of Dagstuhl seminar on Theory of Randomized Search Heuristics planned for 2022.

Member of the organizing committees

- Anne Auger, Dimo Brockhoff, and Nikolaus Hansen: co-organizers of the Black-Box Optimization Benchmarking workshop (BBOB) at the ACM-GECCO 2021 conference (together with Peter Bosman, Tobias Glasmachers, Tea Tušar and Petr Pošik).

10.1.2 Scientific events: selection

Reviewer

- Dimo Brockhoff: ACM-GECCO 2020, PPSN 2020, EMO 2021
- Nikolaus Hansen: ACM-GECCO 2020
- Anne Auger: ACM-GECCO 2020, PPSN 2020

10.1.3 Journal

The three permanent members are frequent reviewers for major journals in Evolutionary Computation. Anne Auger is a frequent reviewer of mathematical optimization journal (JOGO, SIAM OPT). We additionally review papers in Machine Learning related to optimization for JMLR, Machine Learning.

Member of the editorial boards

- Anne Auger, Dimo Brockhoff and Nikolaus Hansen: Associate Editor of the ACM Transactions on Evolutionary Learning and Optimization
- Anne Auger and Nikolaus Hansen: Associate Editor of the Evolutionary Computation Journal
- Anne Auger is guest editor of an Algorithmica special issue of papers selected from the ACM-GECCO'2018 theory track
- Anne Auger is guest editor of the IEEE Transactions on Evolutionary Computation special issue on Theoretical Foundations of Evolutionary Computation

10.1.4 Invited talks

- Dimo Brockhoff: "Evolutionary Computation: From Biological to Simulated Evolution and Into Everyday Life", keynote talk at the 18th Shanghai Forum on Software Trade, Shanghai, November 2020, online
- Nikolaus Hansen: "*Ten+ Years of Benchmarking with COCO/BBOB*" at the Lorentz Center workshop Benchmarked: Optimization Meets Machine Learning, November 2020.

10.1.5 Dagstuhl seminar invitations

- Dimo Brockhoff invited at the Dagstuhl Seminar 20031 on "Scalability in Multiobjective Optimization", January 2020

10.1.6 Lorentz Center seminar invitations

- Anne Auger, Dimo Brockhoff, and Nikolaus Hansen invited at the Lorentz Center Seminar "Benchmarked: Optimization Meets Machine Learning", November 2020, held online due to covid-19.

10.1.7 Leadership within the scientific community

- Anne Auger, Elected Member of the ACM-SIGEVO executive board
- Anne Auger is member of the ACM-SIGEVO board

10.1.8 Research administration

- Anne Auger, member of the BCEP of Inria-Saclay.
- Anne Auger, member of the conseil de laboratoire of the CMAP, Ecole Polytechnique.
- Dimo Brockhoff, member of the Commision de Développement Technologique (CDT) at Inria Saclay

10.2 Teaching - Supervision - Juries

10.2.1 Teaching

- Master: Anne Auger, “Optimization without gradients”, 22.5h ETD, niveau M2 (Optimization Master of Paris-Saclay)
- Master: Dimo Brockhoff, “Algorithms and Complexity”, 36h ETD, niveau M1/M2 (joint MSc with ESSEC “Data Sciences & Business Analytics”), CentraleSupelec, France
- Master: Anne Auger and Dimo Brockhoff, “Introduction to Optimization”, 31.5h ETD, niveau M2 (MSc Informatique - Parcours Apprentissage, Information et Contenu (AIC)), U. Paris-Saclay, France
- Master: Anne Auger, “Math for datascience course”, 31.5h ETD, niveau M1 (MSc Informatique - Parcours Apprentissage, Information et Contenu (AIC)), U. Paris-Saclay, France

10.2.2 Supervision

- PhD in progress: Konstantinos Varelas, “Large-Scale Optimization, CMA-ES and Radar Applications” (Dec. 2017–Feb 2021), supervisors: Anne Auger and Dimo Brockhoff
- PhD in progress: Cheikh Touré, “Linearly Convergent Multi-objective Stochastic Optimizers” (Dec. 2017–), supervisors: Anne Auger and Dimo Brockhoff
- PhD in progress: Paul Dufossé, “Constrained Optimization and Radar Applications”, Oct. 2018, supervisor: Nikolaus Hansen
- PhD in progress: Marie-Ange Dahito, “Mixed-Integer Blackbox Optimization for Multiobjective Problems in the Automotive Industry” (Jan 2019–), supervisors: Dimo Brockhoff and Nikolaus Hansen
- PhD in progress: Eugénie Marescaux, Theoretical Analysis of convergence of multi-objective solvers (2019–), supervisor: Anne Auger
- PhD in progress: Alann Cheral, “Black-box optimization for the optimization of hyperspectral bandwidth for anomaly detection” (2019–), supervisor: Anne Auger
- Jingyun Yang, Ecole Polytechnique, since June 2020
- Armand Gissler, Ecole normale supérieure Paris-Saclay, from April 2020 till August 2020
- Baptiste Plaquevent-Jourdain, ENSTA Paris, from March 2020 till August 2020

10.2.3 Juries

- Anne Auger: co-head of the **SIGEVO PhD award** scientific committee.
- Anne Auger member of the comité de selection for professorship in Calais (COS PR 0054).
- Anne Auger: member of the comité de selection for tenure-track professorship in LIMOS Clermont-Ferrand.
- Anne Auger: member of the comité de selection for a tenur-track assistant professor position on Machine Learning in the Applied Math department of Ecole Polytechnique.

11 Scientific production

11.1 Major publications

[1] Y. Akimoto and N. Hansen. 'Diagonal Acceleration for Covariance Matrix Adaptation Evolution Strategies'. In: *Evolutionary Computation* 28.3 (2020), pp. 405–435. DOI: 10.1162/evco_a_00260. URL: https://hal.inria.fr/hal-01995373.

[2] A. Auger and N. Hansen. 'A SIGEVO impact award for a paper arising from the COCO platform'. In: *ACM SIGEVOlution* 13.4 (Jan. 2021), pp. 1–11. DOI: 10.1145/3447929.3447930. URL: https://hal.inria.fr/hal-03132960.

[3] A. Chotard and A. Auger. 'Verifiable Conditions for the Irreducibility and Aperiodicity of Markov Chains by Analyzing Underlying Deterministic Models'. In: *Bernoulli* 25.1 (Dec. 2018), pp. 112–147. DOI: 10.3150/17-BEJ970. URL: https://hal.inria.fr/hal-01222222.

[4] N. Hansen. 'A Global Surrogate Assisted CMA-ES'. In: *GECCO 2019 - The Genetic and Evolutionary Computation Conference*. ACM. Prague, Czech Republic, July 2019, pp. 664–672. DOI: 10.1145/3321707.3321842. URL: https://hal.inria.fr/hal-02143961.

[5] N. Hansen, A. Auger, R. Ros, O. Mersmann, T. Tušar and D. Brockhoff. 'COCO: A Platform for Comparing Continuous Optimizers in a Black-Box Setting'. In: *Optimization Methods and Software* 36.1 (2020). ArXiv e-prints, arXiv:1603.08785, pp. 114–144. DOI: 10.1080/10556788.2020.1808977. URL: https://hal.inria.fr/hal-01294124.

[6] Y. Ollivier, L. Arnold, A. Auger and N. Hansen. 'Information-Geometric Optimization Algorithms: A Unifying Picture via Invariance Principles'. In: *Journal of Machine Learning Research* 18.18 (2017), pp. 1–65. URL: https://hal.inria.fr/hal-01515898.

[7] C. Touré, N. Hansen, A. Auger and D. Brockhoff. 'Uncrowded Hypervolume Improvement: COMO-CMA-ES and the Sofomore framework'. In: *GECCO 2019 - The Genetic and Evolutionary Computation Conference*. Part of this research has been conducted in the context of a research collaboration between Storengy and Inria. Prague, Czech Republic, July 2019. DOI: 10.1145/3321707.3321852. URL: https://hal.inria.fr/hal-02103694.

11.2 Publications of the year

International journals

[8] Y. Akimoto, A. Auger and N. Hansen. 'Quality Gain Analysis of the Weighted Recombination Evolution Strategy on General Convex Quadratic Functions'. In: *Theoretical Computer Science* 832 (2020), pp. 42–67. DOI: 10.1016/j.tcs.2018.05.015. URL: https://hal.inria.fr/hal-01662568.

[9] Y. Akimoto and N. Hansen. 'Diagonal Acceleration for Covariance Matrix Adaptation Evolution Strategies'. In: *Evolutionary Computation* 28.3 (2020), pp. 405–435. DOI: 10.1162/evco_a_00260. URL: https://hal.inria.fr/hal-01995373.

[10] A. Atamna, A. Auger and N. Hansen. 'On Invariance and Linear Convergence of Evolution Strategies with Augmented Lagrangian Constraint Handling'. In: *Theoretical Computer Science* 832 (2020), pp. 68–97. DOI: 10.1016/j.tcs.2018.10.006. URL: https://hal.inria.fr/hal-01660728.

[11] N. Hansen, A. Auger, R. Ros, O. Mersmann, T. Tušar and D. Brockhoff. 'COCO: A Platform for Comparing Continuous Optimizers in a Black-Box Setting'. In: *Optimization Methods and Software* 36.1 (2020), pp. 114–144. DOI: 10.1080/10556788.2020.1808977. URL: https://hal.inria.fr/hal-01294124.

[12] K. Varelas, O. Ait El Hara, D. Brockhoff, N. Hansen, D. M. Nguyen, T. Tušar and A. Auger. 'Benchmarking large-scale continuous optimizers: the bbob-largescale testbed, a COCO software guide and beyond'. In: *Applied Soft Computing* (2020). URL: https://hal.inria.fr/hal-02960253.

International peer-reviewed conferences

[13] P. Dufossé, C. Enderli, L. Savy and N. Hansen. 'Phased-Array Antenna Pattern Optimization with Evolution Strategies'. In: 2020 IEEE International Radar Conference. Fiorenza, Italy, 21st Sept. 2020. URL: https://hal.inria.fr/hal-03031358.

[14] K. Varelas, A. Auger and N. Hansen. 'Sparse Inverse Covariance Learning for CMA-ES with Graphical Lasso'. In: PPSN 2020 - Sixteenth International Conference on Parallel Problem Solving from Nature. Leiden, Netherlands, 5th Sept. 2020. URL: https://hal.inria.fr/hal-02960269.

Conferences without proceedings

[15] M. Klein, T. Carpentier, E. Jeanclaude, R. Kassab, K. Varelas, N. De Bruijn, F. Barbaresco, Y. Briheche, Y. Semet and F. Aligne. 'AI-Augmented Multi Function Radar Engineering with Digital Twin: Towards Proactivity'. In: 2020 IEEE International Radar Conference. Florence, Italy, 21st Sept. 2020. URL: https://hal.archives-ouvertes.fr/hal-02872199.

Reports & preprints

[16] Y. Akimoto, A. Auger, T. Glasmachers and D. Morinaga. *Global Linear Convergence of Evolution Strategies on More Than Smooth Strongly Convex Functions*. 15th Oct. 2020. URL: https://hal.inria.fr/hal-02941429.

[17] P. Dufossé and C. Enderli. *Finding optimal Pulse Repetion Intervals with Many-objective Evolutionary Algorithms*. 30th Nov. 2020. URL: https://hal.inria.fr/hal-03031366.

[18] C. Touré, A. Gissler, A. Auger and N. Hansen. *Scaling-invariant functions versus positively homogeneous functions*. 8th Jan. 2021. URL: https://hal.inria.fr/hal-03104436.

11.3 Other

Scientific popularization

[19] A. Auger and N. Hansen. 'A SIGEVO impact award for a paper arising from the COCO platform'. In: *ACM SIGEVOlution* 13.4 (22nd Jan. 2021), pp. 1–11. DOI: 10.1145/3447929.3447930. URL: https://hal.inria.fr/hal-03132960.

11.4 Cited publications

[20] Y. Akimoto and N. Hansen. 'Online model selection for restricted covariance matrix adaptation'. In: *International Conference on Parallel Problem Solving from Nature*. Springer. 2016, pp. 3–13.

[21] Y. Akimoto and N. Hansen. 'Projection-based restricted covariance matrix adaptation for high dimension'. In: *Proceedings of the 2016 on Genetic and Evolutionary Computation Conference*. ACM. 2016, pp. 197–204.

[22] D. V. Arnold and J. Porter. 'Towards au Augmented Lagrangian Constraint Handling Approach for the $(1+1)$-ES'. In: *Genetic and Evolutionary Computation Conference*. ACM Press, 2015, pp. 249–256.

[23] A. Atamna, A. Auger and N. Hansen. 'Linearly Convergent Evolution Strategies via Augmented Lagrangian Constraint Handling'. In: *Foundation of Genetic Algorithms (FOGA)*. 2017.

[24] A. Auger and N. Hansen. 'Linear Convergence of Comparison-based Step-size Adaptive Randomized Search via Stability of Markov Chains'. In: *SIAM Journal on Optimization* 26.3 (2016), pp. 1589–1624.

[25] J. Bergstra, R. Bardenet, Y. Bengio and B. Kégl. 'Algorithms for Hyper-Parameter Optimization'. In: *Neural Information Processing Systems (NIPS 2011)*. 2011. URL: https://hal.inria.fr/hal-00642998/file/draft1.pdf.

[26] V. Borkar and S. Meyn. 'The O.D.E. Method for Convergence of Stochastic Approximation and Reinforcement Learning'. In: *SIAM Journal on Control and Optimization* 38.2 (Jan. 2000).

[27] V. S. Borkar. *Stochastic approximation: a dynamical systems viewpoint*. Cambridge University Press. 2008.

[28] C. A. Coello Coello. 'Constraint-handling techniques used with evolutionary algorithms'. In: *Proceedings of the 2008 Genetic and Evolutionary Computation Conference*. ACM. 2008, pp. 2445–2466.

[29] G. Collange, S. Reynaud and N. Hansen. 'Covariance Matrix Adaptation Evolution Strategy for Multidisciplinary Optimization of Expendable Launcher Families'. In: *13th AIAA/ISSMO Multidisciplinary Analysis Optimization Conference, Proceedings*. 2010.

[30] J. E. Dennis and R. B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Englewood Cliffs, NJ: Prentice-Hall, 1983.

[31] N. Hansen and A. Ostermeier. 'Completely Derandomized Self-Adaptation in Evolution Strategies'. In: *Evolutionary Computation* 9.2 (2001), pp. 159–195.

[32] N. Hansen. 'The CMA Evolution Strategy: A Tutorial'. In: *CoRR* abs/1604.00772 (2016). URL: http://arxiv.org/abs/1604.00772.

[33] N. Hansen and A. Auger. 'Principled design of continuous stochastic search: From theory to practice'. In: *Theory and principled methods for the design of metaheuristics*. Springer, 2014, pp. 145–180.

[34] N. Hansen, A. Auger, R. Ros, S. Finck and P. Pošík. 'Comparing results of 31 algorithms from the black-box optimization benchmarking BBOB-2009'. In: *Proceedings of the 12th annual conference companion on Genetic and evolutionary computation*. 2010, pp. 1689–1696.

[35] J. N. Hooker. 'Testing heuristics: We have it all wrong'. In: *Journal of heuristics* 1.1 (1995), pp. 33–42.

[36] F. Hutter, H. Hoos and K. Leyton-Brown. 'An Evaluation of Sequential Model-based Optimization for Expensive Blackbox Functions'. In: *GECCO (Companion) 2013*. Amsterdam, The Netherlands: ACM, 2013, pp. 1209–1216.

[37] D. S. Johnson. 'A theoretician's guide to the experimental analysis of algorithms'. In: *Data structures, near neighbor searches, and methodology: fifth and sixth DIMACS implementation challenges* 59 (2002), pp. 215–250.

[38] D. R. Jones, M. Schonlau and W. J. Welch. 'Efficient global optimization of expensive black-box functions'. In: *Journal of Global optimization* 13.4 (1998), pp. 455–492.

[39] I. Kriest, V. Sauerland, S. Khatiwala, A. Srivastav and A. Oschlies. 'Calibrating a global three-dimensional biogeochemical ocean model (MOPS-1.0)'. In: *Geoscientific Model Development* 10.1 (2017), p. 127.

[40] H. J. Kushner and G. Yin. *Stochastic approximation and recursive algorithms and applications*. Applications of mathematics. New York: Springer, 2003. URL: http://opac.inria.fr/record=b1099801.

[41] P. MacAlpine, S. Barrett, D. Urieli, V. Vu and P. Stone. 'Design and Optimization of an Omnidirectional Humanoid Walk: A Winning Approach at the RoboCup 2011 3D Simulation Competition'. In: *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence (AAAI)*. Toronto, Ontario, Canada, July 2012.

[42] S. Meyn and R. Tweedie. *Markov Chains and Stochastic Stability*. New York: Springer-Verlag, 1993.

[43] Y. Ollivier, L. Arnold, A. Auger and N. Hansen. 'Information-geometric optimization algorithms: A unifying picture via invariance principles'. In: *Journal Of Machine Learning Research* (2016). accepted. eprint: 1106.3708.

[44] T. Salimans, J. Ho, X. Chen and I. Sutskever. 'Evolution strategies as a scalable alternative to reinforcement learning'. In: *arXiv preprint arXiv:1703.03864* (2017).

[45] J. Snoek, H. Larochelle and R. P. Adams. 'Practical bayesian optimization of machine learning algorithms'. In: *Neural Information Processing Systems (NIPS 2012)*. 2012, pp. 2951–2959.

[46] C. Touré, N. Hansen, A. Auger and D. Brockhoff. ‘Uncrowded Hypervolume Improvement: COMO-CMA-ES and the Sofomore framework’. In: *GECCO 2019 - The Genetic and Evolutionary Computation Conference*. Part of this research has been conducted in the context of a research collaboration between Storengy and Inria. Prague, Czech Republic, July 2019. DOI: 10.1145/3321707.3321852. URL: https://hal.inria.fr/hal-02103694.

[47] J. Uhlendorf, A. Miermont, T. Delaveau, G. Charvin, F. Fages, S. Bottani, G. Batt and P. Hersen. ‘Long-term model predictive control of gene expression at the population and single-cell levels’. In: *Proceedings of the National Academy of Sciences* 109.35 (2012), pp. 14271–14276.