Activity Report 2018

Team RANDOPT

Randomized Optimisation

Inria teams are typically groups of researchers working on the definition of a common project, and objectives, with the goal to arrive at the creation of a project-team. Such project-teams may include other partners (universities or research institutions).

RESEARCH CENTER
Saclay - Île-de-France

THEME
Optimization, machine learning and statistical methods

Table of contents

Team RANDOPT

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

Keywords:

Computer Science and Digital Science:

A6. - Modeling, simulation and control
A6.2. - Scientific computing, Numerical Analysis & Optimization
A6.2.2. - Numerical probability
A6.2.3. - Probabilistic methods
A6.2.4. - Statistical methods
A6.2.6. - Optimization
A8.2.2. - Evolutionary algorithms
A8.9. - Performance evaluation

Other Research Topics and Application Domains:

B4. - Energy

1. Team, Visitors, External Collaborators

Research Scientists

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

Post-Doctoral Fellow

Asma Atamna [Inria, until Jun 2018]

PhD Students

Abdelbari Bouzarkouna [Inria, from Nov 2018]
Paul Dufosse [Inria, from Oct 2018]
Cheikh Saliou Toure [Inria]
Konstantinos Varelas [Thales]

Technical staff

Umut Batu [Inria]
Dejan Tusar [Inria, until Apr 2018]

Interns

Eugenie Marescaux [Inria, from Mar 2018 until Jul 2018]
Xudong Zhang [Inria, from Mar 2018 until Aug 2018]

Administrative Assistant

Natalia Alves [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 such that *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, non-smooth* and 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 [35], [24].

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 [38], geoscience [30], or robotics [33]).

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 [36]. [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 those 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 [23] while the stochastic derivative-free counterpart, CMA-ES dates from the early 2000s [25]. 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 $i = 1, ..., m$. Only recently, the fundamental requirement of linear convergence [2], as in the unconstrained case, has been clearly stated [14].

[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 an important factor for the important 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.

- 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 needed. 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 efficiently problems with one or two orders of magnitude more variables is still largely open.
- 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 [29], SMAC [27], Spearmint [37], or TPE [17] 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 tested. 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."* [28]

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.

2.2. Overall Objectives

In the context of black-box numerical optimization previously described, the scientific positioning of RandOpt 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, multi-objective 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 a *non*-property—*non*-convex, *non*-linear, *non*-smooth. 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 it is also our firm belief that parameter tuning is part of the designer's task.

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

3.1. Introduction

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.

3.2. 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 $x_1, ..., x_\lambda$ 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:

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

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

$$m \longleftarrow \frac{1}{\lambda} \sum_{i=1}^{\lambda/2} x_{i:\lambda} \ .$$

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 [16]. 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 [34], 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 [19], [20], [31]. 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.

3.3. 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 [35], [24]. 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".

3.3.1. 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 [21]. However, the penalization coefficient is a sensitive parameter that needs to be adapted in order to achieve a robust and general method [22]. Yet, **the question of how to handle properly constraints is largely unsolved**. The latest constraints handling for CMA-ES is an ad-hoc technique driven by many heuristics [22]. Also, it is particularly only recently that it was pointed out that **linear convergence properties should be preserved** when addressing constraint problems [14].

Promising approaches though, rely on using augmented Lagrangians [14], [15]. 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.

3.3.2. 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 [13], [12]. Yet so far they fail to optimize functions whose Hessian matrix has some small eigenvalues (say around 10^{-4}) some eigenvalues equal to 1 and some very large eigenvalue (say around 10^4), that is functions whose level sets have short and long axis.

Another direction, we want to pursue, is exploring the use of large-scale variants of CMA-ES to solve reinforcement learning problems [36].

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.

3.3.3. 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

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

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.

3.3.4. 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.

3.4. 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. 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 [26].

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 200+ 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

4.1. Application Domains

Applications of black-box algorithms occur in various domains. Industry but also researchers in other academic domains have therefore a great need to apply black-box algorithms on a daily basis. We see this as a great source of motivation to design better methods. Applications not only allow us to backup our methods and understand what are the relevant features to solve a real-world problem but also help identify novel difficulties or set priorities in terms of algorithm design.

Asides from the two applications to Machine Learning that we detail below, we however do not target a specific application domain and we are interested in possible black-box applications stemming from various origins. This is for us intrinsic to the nature of the methods we develop that are general purpose algorithms. Hence our strategy with respect to applications can be seen as opportunistic and our main selection criteria when approached by colleagues who want to develop a collaboration around an application is whether we judge the application interesting: that is the application brings new challenges and/or gives us the opportunity to work on topics we already intended to work on.

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

- With EDF R&D through the design and placement of bi-facial photovoltaic panel for the postdoc of Asma Atamna funded by the PGMO project.
- With Thales for the thesis of Konstantinos Varelas (DGA-CIFRE thesis) related to the design of radars (shape optimization of the wave form). This thesis investigates the development of large-scale variants of CMA-ES.
- With Storengy, a subsidiary of Engie specialized in gas storage for the thesis of Cheikh Touré. Different multi-objective applications are considered in this context but the primary motivation of Storengy is to get at their disposal a better multi-objective variant of CMA-ES which is the main objective of the developments within the thesis.

Additionally, there are two specific types of applications stemming from Machine Learning we would like to focus on: problems with non-differentiable loss that can occur in reinforcement learning and hyperparameter tuning problems. For the first class of problems the motivation comes from the paper [36] where different reinforcement learning problems are addressed and the weights of neural networks are adjusted using evolution strategies. Those problems are large-scale (in [36] up to 10^6 weights are adjusted), and the large-scale variants of CMA-ES we want to investigate might be relevant in this case. For the second class of problems (hyperparameter tuning problems), standard approches to handle those problems are Bayesian optimization algorithms but despite the tremendous effort for developing Bayesian optimization techniques and having implementations of Bayesian optimization algorithms within libraries, pure random search is still often used for training neural networks. One reason is that pure random search is intrinsically parallel [18]. This suggests that methods like CMA-ES—that are also intrinsically parallel—can be also advantageously used for hyperparameter tuning: this was demonstrated to tune deep neural networks in [32]. One limitation though of the CMA-ES algorithm is that it cannot deal with categorical/integer and continuous variables at the same time. This motivates us to investigate the development of CMA-ES variants that are able to deal with mixed variables.

When dealing with single applications, the results observed are difficult to generalize: typically not many methods are tested on a single application as tests are often time consuming and performed in restrictive settings. Yet, if one circumvent the problem of confidentiality of data and of criticality for companies to publish their applications, real-world problems could become benchmarks as any other analytical function. This would allow to test wider ranges of methods on the problems and to find out whether analytical benchmarks properly capture real-world problem difficulties. We will thus seek to incorporate real-world problems within the COCO platform. This is a recurrent demand by researchers in optimization. As far as confidentiality of data are concerned, our preliminary discussions with industrials allow us to be optimistic that we can convince industrials to propose real-world problems with anonymized (and uncritical) data that still capture the essence of the underlying real-world problem.

5. Highlights of the Year

5.1. Highlights of the Year

A. Auger appointed general chair of the ACM GECCO 2019 conference (GECCO being the largest most selective conference in EC)

6. New Software and Platforms

6.1. COCO

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, (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 blackbox 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, 160+ algorithms and algorithm variants by contributors from all over the world have been benchmarked on the platform's three supported test suites so far. The most recent extension towards bi-objective problems has been used for the BBOB-2016 workshop at GECCO and we are currently developing new test suites around large-scale and constrained optimization.

- Participants: Anne Auger, Asma Atamna, Dejan Tusar, Dimo Brockhoff, Marc Schoenauer, Nikolaus Hansen, Ouassim Ait Elhara, Raymond Ros, Tea Tusar, Thanh-Do Tran and Umut Batu
- Partners: TU Dortmund University - Charles University Prague - Jozef Stefan Institute (JSI)
- Contact: Dimo Brockhoff
- URL: https://github.com/numbbo/coco

6.2. CMA-ES

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.

- Participant: Nikolaus Hansen
- Contact: Nikolaus Hansen
- URL: http://cma.gforge.inria.fr/cmaes_sourcecode_page.html

6.3. Platforms

6.3.1. New developments around COCO

There were two public releases of the COCO software this year including quite some new features that have also been used for the Blackbox Optimization Benchmarking workshop (BBOB) which was held in Kyoto, Japan during GECCO-2019.

The most important new features are updated, streamlined plots, a Python 3 compatible postprocessing module with a corresponding restructuring of the postprocessing code, the support for zip files in the postprocessing, a simplified example experiment script for beginners and a non-anytime example experiment for benchmarking budget-dependent algorithms, improved coverage of the continuous integration testing via CircleCI and AppVeyor, and finally and most-important from a practical perspective an archive with automatized download from all 200+ algorithm data sets available in the COCO data archive. Of these, 17 algorithm data sets have been made newly available in 2018 with four scientific papers being presented at the BBOB-2019 workshop.

In the background, there have been additional (preparational) activities, in particular due to the two Inria ADT projects “COCOpysuites” and “COCOpost”. The “COCOpysuites” project aimed at a rewriting of the experimental part of COCO in python to allow for an easier development, testing, and implementation of new test suites. The “COCOpost” project aimed at a complete rewrite of the python postprocessing with a focus on new, interactive plots and a clearer structure for improved maintenance. In addition, new test suites have been developed and implemented for large-scale, constrained, multiobjective, and mixed-integer optimization. All those extensions will be made available step-by-step to the scientific community after proper alpha- and beta-testing in the coming planned releases.

6.3.2. Developments within the CMA-ES library

The `pycma` library has not seen major changes, but overall 39 commits pushed for maintenance, bug-fixes and smaller improvements (roughly 1000 lines of code). An as of yet unpublished development has been the modularization of the data logger. A surrogate fitness model module with 969 lines of code has been developed and is already operative but has also not yet been released.

7. New Results

7.1. Analysis of adaptive Stochastic Optimizers

7.1.1. New ODE method for proving the geometric convergence of adaptive Stochastic Optimizers

The ODE method is a standard technique to analyze the convergence of stochastic algorithms defined as a stochastic approximation of an ODE. In a nutshell, the convergence of the algorithms derives from the stability of the ODE and the control of the error between the solution of the ODE and the trajectory of the stochastic algorithm. We have been developing a new ODE method to be able to prove the geometric convergence of stochastic approximation algorithms that derive from the family of adaptive stochastic optimization algorithms. Standard theory did not apply in this context as the state variable adapted typically converge to the boundary of the state-space domain where an infinite number of points are equilibrium points for the ODE [7].

7.1.2. Convergence and convergence rate analysis of the (1+1)-ES with one-fifth success rule

When analyzing adaptive stochastic optimizers, one is typically interested to prove the linear convergence and investigate the dependency of the convergence rate with respect to the dimension. We have greatly simplified the analysis of the convergence and convergence rate of the (1+1)-ES with one-fifth success rule on the sphere function. We have shown that the analysis derives from applying a simple "drift" theorem and consequently shown a hitting time to reach an ϵ-ball of the optimum of $\Theta(\frac{1}{d}\log(1/\epsilon))$ akin to linear convergence with a convergence rate scaling linearly with the dimension [4].

7.1.3. Quality-gain Analysis on Convex-quadratic functions

We have analyzed the expected function value decrease (related to the convergence rate) of Evolution Strategies with weighed recombination on convex-quadratic functions. We have derive different bounds and limit expression that allow to derive optimal recombination weights and the optimal step-size, and found that the optimal recombination weights are independent of the Hessian of the objective function. We have moreover shown the dependencies of the optimal parameters in the dimension and population size [1].

7.2. Benchmarking Methodology

7.2.1. Single-Objective Benchmarking

Benchmarking optimization algorithms seems trivial at first sight but is quite involved in practice and little decisions on the experimental setup can have a large effect on the displayed algorithm performance.

We have investigated some of these effects in the context of the Black-Box Optimization Benchmarking (`bbob`) test suite of the COCO platform and the well-known quasi-Newton BFGS algorithm, default in MATLAB's `fminunc` and in Python's `scipy.optimize` module [5]. We realized in particular that the instance instantiation in the COCO platform has little impact while the initial search point has a larger one. The largest performance differences, however, stem from implementation details that are typically not documented and not exposed to the user via internal algorithm parameters. For example is the MATLAB implementation of the BFGS algorithm significantly worse than the Python implementation and the MATLAB 2017 version is worse than the MATLAB 2009 implementation.

Additionally, Nikolaus Hansen gave a hands-on tutorial on good benchmarking practice at the GECCO-2018 conference in Kyoto [11].

7.2.2. Multi-Objective Benchmarking

In terms of multiobjective benchmarking, our contributions are two-fold. Firstly, we wrote a scientific article on the scientific methodology for defining our new multiobjective benchmark suite. In [10] we introduce two new bi-objective test suites on the basis of the above mentioned, well-known 24 `bbob` test functions and propose a generic test suite generator for an arbitrary number of objectives. The former are implemented in our COCO platform and extensively documented in terms of search and objective space plots for each function.

Secondly, we realized with the proposal of the biobjective `bbob` test suites that there is a need for more theoretical analyses of simple test functions that still test for practical challenges such as ill-conditioning or search space rotations. In our upcoming EMO conference paper [3] we therefore characterize theoretically Pareto sets and Pareto fronts of combinations of two convex quadratic functions with arbitrary search space dimension. Based on this theoretical analysis, we suggest a wide set of new biobjective test functions.

7.3. Large-scale Optimization

We have been studying different large-scale variants of the CMA-ES algorithm and tested them thoroughly empirically on a set of scalable large-scale testbed. The study includes comparison with the large-scale quasi-Newton algorithm, namely L-BFGS [6].

7.4. Constrained Optimization

In the context of constrained optimization, A. Atamna studied invariance properties of Augmented Lagrangian approaches and showed the relation between invariance to strictly increasing affine transformations of the objective function and the scaling of the constraints and linear convergence [2]. Progress were made towards a methodology to define scalable constrained problems with control optimum and the implementation of a new constrained testbed within the COCO platform. In her internship, E. Marescaux studied the connection between augmented Lagrangian approaches and a previously proposed adaptive constrained handling mechanism. She also studied a new idea to turn a constrained problem into an unconstrained one.

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)

9. Partnerships and Cooperations

9.1. Regional Initiatives

- PGMO/FMJH project “AESOP: Algorithms for Expensive Simulation-Based Optimization”, 7kEUR, 2017–2019

9.2. National Initiatives

9.2.1. ANR

- ANR project “Big Multiobjective Optimization (BigMO)”, Dimo Brockhoff participates in this project through the Inria team BONUS in Lille (2017–2020)

9.3. International Research Visitors

9.3.1. Visitors to RandOpt

- Filip Matzner, October 15–19 and December 10-14, 2018

9.3.2. Internships

- Eugenie Marescaux, March–July 2018
- Xudong Zhang, March–August 2018

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

- Anne Auger is the General Chair of the forthcoming ACM GECCO 2019 conference, Prague CZ (largest and most prestigious conference in the Evolutionary Computation domain).

10.1.2. Member of the Organizing Committees

- Anne Auger, Dimo Brockhoff and Nikolaus Hansen, co-organizer of the ACM-GECCO-2018 workshop on Black Box Optimization Benchmarking, together with Julien Bect, Rodolphe Le Riche, Victor Picheny, and Tea Tušar
- Anne Auger, Dimo Brockhoff, Nikolaus Hansen, and Konstantinos Varelas, co-organizer of the ACM-GECCO-2019 workshop on Black Box Optimization Benchmarking, together with Tea Tušar
- Dimo Brockhoff: co-organization of the Lorentz Center workshop on Many-Criteria Optimization (MACODA), September 2019, with Boris Naujoks, Michael Emmerich, and Robin Purshouse

10.1.3. Scientific Events Selection

10.1.3.1. Chair of Conference Program Committees

- Anne Auger has been program chair of the PPSN 2018 conference, Coimbra, Portugal.
- Anne Auger was Theory track chair for the ACM GECCO 2018 conference, Kyoto, Japan.
- Nikolaus Hansen was ENUM track chair for the ACM GECCO 2018 conference, Kyoto, Japan.

10.1.3.2. Member of the Conference Program Committees

- Dimo Brockhoff and Nikolaus Hansen were members of the program committee of the PPSN 2018 conference.

10.1.3.3. Reviewer

- Dimo Brockhoff: GECCO’2018, GECCO’2018 student workshop, PPSN’2018, LeGO 2018, EMO’2019, and FOGA’2019

10.1.4. Journal

10.1.4.1. Member of the Editorial Boards

- Anne Auger and Nikolaus Hansen members of the editorial board of the Evolutionary Computation journal.

10.1.4.2. Reviewer - Reviewing Activities

- The three permanent members are frequent reviewers for the main two journals on Evolutionary Computation: IEEE transaction on Evolutionary Computation, Evolutionary Computation.
- Anne Auger is guest editor of Algorithmica special issue of papers selected from GECCO theory tracks 2018
- Anne Auger is guest editor of IEEE Transactions on Evolutionary Computation special issue on Theoretical Foundations of Evolutionary Computation 2018/2019

10.1.5. Invited Talks

- Dimo Brockhoff: “Benchmarking multiobjective optimizers: An algorithmic jam session of recent results”, December 2018, Centre for Informatics and Systems, University of Coimbra, Portugal

10.1.6. Leadership within the Scientific Community

- Anne Auger, Elected Member of the ACM-SIGEVO executive board
- Dimo Brockhoff, member of the International Advisory Committee for EMO-2019 in East Lansing, USA

10.1.7. Scientific Expertise

- Anne Auger scientific expert for an audit on Artificial Intelligence of a large French industrial consortium.
- Dimo Brockhoff, external reviewer for the Luxembourg National Research Fund (FNR) in the CORE 2018 call

10.1.8. Research Administration

- Anne Auger, member of the conseil de laboratoire of the CMAP, Ecole Polytechnique.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

- Master: Anne Auger, Course on Derivative-free Optimization of AMS and Optimization masters, 22.5 hours M2, Paris-Saclay University
- Master: Anne Auger, exercises for the courses Introduction to Machine Learning (MAP 534), M2 X/HEC and Advanced Machine Learning (MAP 541), 30 hours, Ecole Polytechnique, France
- Master: Anne Auger, exercises for the courses Advanced Machine Learning (MAP 541), M2 X/HEC and Advanced Machine Learning (MAP 541), 30 hours, Ecole Polytechnique, France
- Master: Anne Auger, 2nd year students of Ecole Polytechnique, Datacamp (solving practical machine learning problems) MAP 583, 20 hours, Ecole Polytechnique, France
- Master: Anne Auger, “Advanced Optimization”, 9h ETD, M2, Université Paris-Sud, France (joint course with D. Brockhoff)
- Master: Dimo Brockhoff, “Introduction to Optimization”, 31.5h ETD, M2, Université Paris-Sud, France
- Master: Dimo Brockhoff, “Advanced Optimization”, 22.5h ETD, M2, Université Paris-Sud, France (joint course with A. Auger)

10.2.2. Tutorials

- Dimo Brockhoff gave the tutorial “Evolutionary Multiobjective Optimization” at the GECCO-2018 conference in Kyoto, Japan [9]
- Nikolaus Hansen gave the tutorials “CMA-ES and Advanced Adaptation Mechanisms” [8] and “A Practical Guide to Experimentation (and Benchmarking)” [11] at the GECCO-2018 conference in Kyoto, Japan.

10.2.3. Supervision

- PhD in progress: Konstantinos Varelas, “Large-Scale Optimization, CMA-ES and Radar Applications”, Dec. 2017, Anne Auger and Dimo Brockhoff
- PhD in progress: Cheikh Touré, “Linearly Convergent Multi-objective Stochastic Optimizers”, Dec. 2017, Anne Auger and Dimo Brockhoff
- PhD in progress: Paul Dufossé, “Constrained Optimization and Radar Applications”, Oct. 2018, Nikolaus Hansen
- PhD in progress: Marie-Ange Dahito, “Mixed-Integer Blackbox Optimization for Multiobjective Problems in the Automotive Industry”, Jan 2019, Dimo Brockhoff and Nikolaus Hansen

10.2.4. Juries

- Anne Auger, member of Hiring committee for assistant professor positions for Toulouse 1 Capitole University, MCF 0225, section CNU 27
- Anne Auger, PhD jury of Asmaa Ghoumari, defense in Dec. 2018
- Dimo Brockhoff, PhD jury of Lukas Bajer, Charles University Prague, Czech Republic, defense in June 2018
- Dimo Brockhoff, PhD jury of Andreia P. Guerrero, University of Coimbra, Portugal, defene in December 2018

10.3. Popularization

10.3.1. Interventions

- National events: Cheikh Touré made an intervention on Poker for "college" students at the Fête de la Science (explaining relation between the game and probability)

11. Bibliography

Publications of the year

Articles in International Peer-Reviewed Journals

[1] Y. Akimoto, A. Auger, N. Hansen. *Quality Gain Analysis of the Weighted Recombination Evolution Strategy on General Convex Quadratic Functions*, in "Theoretical Computer Science", 2018 [*DOI :* 10.1016/j.tcs.2018.05.015], https://hal.inria.fr/hal-01662568

[2] A. Atamna, A. Auger, N. Hansen. *On Invariance and Linear Convergence of Evolution Strategies with Augmented Lagrangian Constraint Handling*, in "Theoretical Computer Science", November 2018, https://hal.inria.fr/hal-01660728

Invited Conferences

[3] C. Touré, A. Auger, D. Brockhoff, N. Hansen. *On Bi-Objective convex-quadratic problems*, in "10th International Conference on Evolutionary Multi-Criterion Optimization", East Lansing, Michigan, United States, March 2019, https://arxiv.org/abs/1812.00289 , https://hal.inria.fr/hal-01942159

International Conferences with Proceedings

[4] Y. AKIMOTO, A. AUGER, T. GLASMACHERS. *Drift Theory in Continuous Search Spaces: Expected Hitting Time of the (1+1)-ES with 1/5 Success Rule*, in "Proceedings of the GECCO 2018 Conference", Kyoto, Japan, 2018, https://arxiv.org/abs/1802.03209 , https://hal.inria.fr/hal-01778116

[5] A. BLELLY, M. FELIPE-GOMES, A. AUGER, D. BROCKHOFF. *Stopping Criteria, Initialization, and Implementations of BFGS and their Effect on the BBOB Test Suite*, in "GECCO '18 Companion", Kyoto, Japan, July 2018 [*DOI :* 10.1145/3205651.3208303], https://hal.inria.fr/hal-01811588

[6] K. VARELAS, A. AUGER, D. BROCKHOFF, N. HANSEN, O. A. ELHARA, Y. SEMET, R. KASSAB, F. BARBARESCO. *A Comparative Study of Large-scale Variants of CMA-ES*, in "PPSN XV 2018 - 15th International Conference on Parallel Problem Solving from Nature", Coimbra, Portugal, LNCS, September 2018, vol. 11101, pp. 3-15 [*DOI :* 10.1007/978-3-319-99253-2_1], https://hal.inria.fr/hal-01881454

Other Publications

[7] Y. AKIMOTO, A. AUGER, N. HANSEN. *An ODE Method to Prove the Geometric Convergence of Adaptive Stochastic Algorithms*, November 2018, https://arxiv.org/abs/1811.06703 - working paper or preprint, https://hal.inria.fr/hal-01926472

[8] Y. AKIMOTO, N. HANSEN. *CMA-ES and Advanced Adaptation Mechanisms*, July 2018, GECCO '18 Companion: Proceedings of the Genetic and Evolutionary Computation Conference Companion, https://hal.inria.fr/hal-01959479

[9] D. BROCKHOFF. *GECCO 2018 tutorial on evolutionary multiobjective optimization*, July 2018, GECCO '18 Companion: Proceedings of the Genetic and Evolutionary Computation Conference Companion, https://hal.inria.fr/hal-01943586

[10] D. BROCKHOFF, T. TUSAR, A. AUGER, N. HANSEN. *Using Well-Understood Single-Objective Functions in Multiobjective Black-Box Optimization Test Suites*, January 2019, https://arxiv.org/abs/1604.00359 - ArXiv e-prints, arXiv:1604.00359, https://hal.inria.fr/hal-01296987

[11] N. HANSEN. *A Practical Guide to Experimentation (and Benchmarking)*, July 2018, GECCO '18 Companion: Proceedings of the Genetic and Evolutionary Computation Conference Companion, https://hal.inria.fr/hal-01959453

References in notes

[12] Y. AKIMOTO, N. HANSEN. *Online model selection for restricted covariance matrix adaptation*, in "International Conference on Parallel Problem Solving from Nature", Springer, 2016, pp. 3–13

[13] Y. AKIMOTO, 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

[14] D. V. ARNOLD, 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

[15] A. ATAMNA, A. AUGER, N. HANSEN. *Linearly Convergent Evolution Strategies via Augmented Lagrangian Constraint Handling*, in "Foundation of Genetic Algorithms (FOGA)", 2017

[16] A. AUGER, N. HANSEN. *Linear Convergence of Comparison-based Step-size Adaptive Randomized Search via Stability of Markov Chains*, in "SIAM Journal on Optimization", 2016, vol. 26, n^{o} 3, pp. 1589-1624

[17] J. BERGSTRA, R. BARDENET, Y. BENGIO, B. KÉGL. *Algorithms for Hyper-Parameter Optimization*, in "Neural Information Processing Systems (NIPS 2011)", 2011, https://hal.inria.fr/hal-00642998/file/draft1.pdf

[18] J. BERGSTRA, Y. BENGIO. *Random search for hyper-parameter optimization*, in "Journal of Machine Learning Research", 2012, vol. 13, pp. 281–305

[19] V. S. BORKAR. *Stochastic approximation: a dynamical systems viewpoint*, 2008, Cambridge University Press

[20] V. BORKAR, S. MEYN. *The O.D.E. Method for Convergence of Stochastic Approximation and Reinforcement Learning*, in "SIAM Journal on Control and Optimization", January 2000, vol. 38, n^{o} 2

[21] 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

[22] G. COLLANGE, S. REYNAUD, N. HANSEN. *Covariance Matrix Adaptation Evolution Strategy for Multidisciplinary Optimization of Expendable Launcher Families*, in "13th AIAA/ISSMO Multidisciplinary Analysis Optimization Conference, Proceedings", 2010

[23] J. E. DENNIS, R. B. SCHNABEL. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, NJ, 1983

[24] N. HANSEN, 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

[25] N. HANSEN, A. OSTERMEIER. *Completely Derandomized Self-Adaptation in Evolution Strategies*, in "Evolutionary Computation", 2001, vol. 9, n^{o} 2, pp. 159–195

[26] J. N. HOOKER. *Testing heuristics: We have it all wrong*, in "Journal of heuristics", 1995, vol. 1, n^{o} 1, pp. 33–42

[27] F. HUTTER, H. HOOS, K. LEYTON-BROWN. *An Evaluation of Sequential Model-based Optimization for Expensive Blackbox Functions*, in "GECCO (Companion) 2013", ACM, 2013, pp. 1209–1216

[28] 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", 2002, vol. 59, pp. 215–250

[29] D. R. JONES, M. SCHONLAU, W. J. WELCH. *Efficient global optimization of expensive black-box functions*, in "Journal of Global optimization", 1998, vol. 13, n^{o} 4, pp. 455–492

[30] I. KRIEST, V. SAUERLAND, S. KHATIWALA, A. SRIVASTAV, A. OSCHLIES. *Calibrating a global three-dimensional biogeochemical ocean model (MOPS-1.0)*, in "Geoscientific Model Development", 2017, vol. 10, n^{o} 1, 127 p.

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

[32] I. LOSHCHILOV, F. HUTTER. *CMA-ES for hyperparameter optimization of deep neural networks*, in "arXiv preprint arXiv:1604.07269", 2016

[33] P. MACALPINE, S. BARRETT, D. URIELI, V. VU, 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)", July 2012

[34] S. MEYN, R. TWEEDIE. *Markov Chains and Stochastic Stability*, Springer-Verlag, New York, 1993

[35] Y. OLLIVIER, L. ARNOLD, A. AUGER, N. HANSEN. *Information-geometric optimization algorithms: A unifying picture via invariance principles*, in "Journal Of Machine Learning Research", 2016, accepted

[36] T. SALIMANS, J. HO, X. CHEN, I. SUTSKEVER. *Evolution strategies as a scalable alternative to reinforcement learning*, in "arXiv preprint arXiv:1703.03864", 2017

[37] J. SNOEK, H. LAROCHELLE, R. P. ADAMS. *Practical bayesian optimization of machine learning algorithms*, in "Neural Information Processing Systems (NIPS 2012)", 2012, pp. 2951–2959

[38] J. UHLENDORF, A. MIERMONT, T. DELAVEAU, G. CHARVIN, F. FAGES, S. BOTTANI, G. BATT, 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", 2012, vol. 109, n^{o} 35, pp. 14271–14276