Activity Report 2017

Team RANDOPT

Randomized Optimization

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).

Table of contents

Team RANDOPT

Creation of the Team: 2016 December 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. Personnel

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

Post-Doctoral Fellow
    Asma Atamna [Inria, from Feb 2017]

PhD Students
    Cheikh Saliou Toure [Inria, from Oct 2017]
    Konstantinos Varelas [Thales, from Dec 2017]

Technical staff
    Asma Atamna [Inria, until Jan 2017]
    Umut Batu [Inria, from Jul 2017]
    Dejan Tusar [Inria, from Nov 2017]

Interns
    Shan Wolf [Inria, from Jun 2017 until Sep 2017]
    Yiyang Yu [Inria, from Apr 2017 until Jul 2017]

Administrative Assistants
    Natalia Alves [Inria]
    Olga Mwana Mobulakani [Inria]

2. Overall Objectives

2.1. Scientific Context

Important problems in various scientific domains like biology, physics, medicine or in industry critically rely on the resolution of difficult numerical optimization problems. Often those problems depend on noisy data or are the outcome of complex numerical simulations such that derivatives are not available or not useful 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—a single function evaluation might 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 has proven to be one key component for the design of robust global numerical optimization algorithms.
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 sought 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 [28], geoscience [23], or robotics [25].

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 [27]. 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 (see Section on Software and 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 [20] while the stochastic derivative-free counterpart, CMA-ES dates from the early 2000s [21]. 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, as in the unconstrained case, has been clearly stated [13].

- 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 where 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 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 medium scale optimization 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 algorithms based on Gaussian processes. In this domain, no standard implementations exist and performance across implementations can vary significantly as particularly different algorithm components are used. For instance, there is no common agreement on which initial design to use, which bandwidth for the Gaussian kernel to take, and which strategy to follow to optimize the expected improvement.

2.2. Overall Objectives

In the context of black-box numerical optimization previously described, the main ambition of the RandOpt team is **to design and implement novel methods in subdomains with a strong practical demand**. Those methods should become future standards that allow to solve important challenging applications in industry or academia. For this, we believe that (i) **theory** can greatly help for **algorithm design**; (ii) the **development and implementation of proper scientific experimentation methodology** is crucial and (iii) it is decisive to provide **parameter-less implementations** of the methods through open-source software packages. This shapes four main scientific goals for our proposed team:

1. **develop novel theoretical frameworks** for guiding (a) the design of novel 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 **novel 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.

All the above relate to 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. Developing Novel Theoretical Frameworks for Analyzing and Designing Adaptive Stochastic Algorithms

The lines of research of the RandOpt team are organized along four axis namely developing novel theoretical framework, developing novel algorithms, setting novel standards in scientific experimentation and benchmarking and applications.

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: not only derivatives (gradients) are not exploited, but often the methods are so-called comparison-based which means that the algorithm will only rely on the ranking of the candidate solutions' function values. This renders those methods more robust as they are invariant to strictly increasing transformations of the objective function but at the same time the theoretical analysis becomes more difficult as **we cannot exploit a well defined framework using (strong) properties of the function like convexity or smoothness**.

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, the **critical points of the ODEs are at the boundary of the domain**.

Last, since we aim at developing theory that one 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 to not 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 to analyze one of the core properties of adaptive stochastic methods, namely **linear convergence** on the widest possible class of functions.

We are planning on approaching 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 [15]. 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 [26], 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 [16], [17], [24]. 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. 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.

3.2. Developing Novel Adaptive Stochastic Algorithms

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

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 [18]. However, the penalization coefficient is a sensitive parameter that needs to be adapted in order to achieve a robust and general method [19]. 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 [19]. Also, it is particularly only recently that it was pointed out that **linear convergence properties should be preserved** when addressing constraint problems [13].

Promising approaches though, rely on using augmented Lagrangians [13], [14]. 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. 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 [12], [11]. Yet there are still open problems related to being able to learn both short and long axes in the models.

Another direction, we want to pursue, is exploring the use of large-scale variants of CMA-ES to solve reinforcement learning problems [27].

Last, we are interested to investigate the very-large-scale setting. One approach consist 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[1] 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. Though performing

---

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

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 [22].

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 (see also the software section below) 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. Applications

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 to identify novel difficulties or set priorities in terms of algorithm design.

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

- With EDF R&D through the design and placement of bi-facial photovoltaic panels for the postdoc of Asma Atamna funded by the PGMO project NumBER.
- With Thales for the PhD thesis of Konstantinos Varelas (DGA-CIFRE thesis) related to applications in the defense domain.
- With Storengy, a subsidiary of Engie specialized in gas storage, for the PhD thesis of Cheikh Touré.

Another type of application we want to focus on comes from reinforcement learning. The problems addressed in [27] seem to be particularly suited for large-scale variants of CMA-ES.

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 our COCO platform. This is a recurrent demand by researchers in optimization.

5. Highlights of the Year

5.1. Highlights of the Year

- A Auger has been (re)-elected member of the ACM-SIGEVO executive board.

6. New Results

6.1. Theory

Participants: Anne Auger, Nikolaus Hansen.

The paper "Information-Geometric Optimization Algorithms: A Unifying Picture via Invariance Principles" has finally been published in the JMLR journal [3]. In this paper in collaboration with Yann Ollivier in particular, we lay the ground of stochastic optimization by means of information geometry. We provide a unified framework for stochastic optimization on arbitrary search spaces that allow to recover well-known algorithms on continuous or discrete search spaces and put them under the same umbrella of Information Geometric Optimization.

When analyzing the stability of Markov chains stemming from comparison-based stochastic algorithms, we are facing difficulties due to the fact that the Markov chains have the following form $\Phi_{t+1} = F(\Phi_t, U_{t+1})$ where $\{U_t : t \geq 0\}$ are i.i.d. random vectors and F is a discontinuous function. The discontinuity comes from the comparison-based property of the algorithms. If F were C^∞ or C^1 we could prove easily stability properties like irreducibility and show that compact are small sets by investigating the underlying control model and showing that it has globally attracting states where controllability conditions hold using results developed by Sean Meyn and co-authors.

In the paper [2], we found that we can actually generalize to a great extent the results by Meyn to the case where $\Phi_{t+1} = F(\Phi_t, \alpha(\Phi_t, U_{t+1}))$ where F is C^1 and α is discontinuous but such that $\alpha(x, U)$ admits a lower-semi continuity density. We have proposed verifiable conditions for the irreducibility and aperiodicity and shown that compact sets are small sets.

The development of evolution strategies has been greatly driven by so-called progress rate or quality gain analysis where simplification assumptions are made to obtain quantitative estimate of progress in one step and deduce from it how to set different parameters like recombination weights, learning rates ...

This theory while very useful often relied on approximations that were not always well appraised, justified or clearly stated. We have been in the past rigorously deriving different progress rate results and related them to bounds on convergence rates. We have investigated rigorously the quality gain (that is progress measured in terms of objective function) on general convex quadratic functions using weighted recombination. This allowed to derive the dependency of the convergence rate of evolution strategies with respect to the eigenspectrum of the Hessian matrix of convex-quadratic function as well as give hints on how to set learning rate [4] and [9].

6.2. Novel Constraint Handling

Participants: Asma Atamna, Anne Auger, Nikolaus Hansen.

In the context of constrained optimization, we have investigated to use augmented Lagrangian approaches to handle constraints. The originality of the approach is that the parameters of the augmented Lagrangian are adapted online. We have shown sufficient conditions for linear convergence of the ensuing methods with linear constraints [5]. Those sufficient conditions rely on finding a Markov chain candidate to be stable. This Markov chain derives from invariance properties of the algorithm. At the same time we have proposed an algorithm variant for the $(\mu/\mu, \lambda)$-CMA-ES and an arbitrary number of constraints.

In [10], we have investigated the linear convergence question on the point of view of invariance. We have analyzed the invariances of adaptive algorithms handling constraints with augmented Lagrangian: we have shown that invariance to monotonic transformation of the objective functions is lost but that a subclass of invariance can and should be preserved, namely affine transformation of the objective function and scaling of the constraint by a positive constant.

6.3. Benchmarking

Participants: Anne Auger, Dimo Brockhoff, Nikolaus Hansen, Umut Batu, Dejan Tusar.

In his thesis, Ouassim AitElHara has been investigating the benchmarking of algorithms in large dimensions [1]. In this context, the first steps for a testbed of the COCO platform in large dimension have been done. Particularly, the methodology for building a large-scale testbed has been defined: it consists in replacing the usual orthogonal transformation by block-diagonal orthogonal matrices multiplied to the left and to the right by permutation matrices. While still under testing, we expect to be able to release the large-scale testbed in the coming year.

The population size is one of the few parameters, a user is supposed to touch in the state-of-the-art optimizer CMA-ES. In [7], a new approach to also adapt the population size in CMA-ES is proposed and benchmarked on the bbob test suite of our COCO platform. The method is based on tracking the non-decrease of the median of the objective function values in each slot of S successive iterations to decide whether to increase or decrease or keep the population size in the next slot of S iterations. The experimental results show the efficiency of our approach on some multi-modal functions with adequate global structure.

Benchmarking budget-dependent algorithms (for which parameters might depend on the given budget of function evaluations) is typically done for a fixed (set of) budget(s). This, however, has the disadvantage that the reported function values at this budget are difficult to interpret. Furthermore, assessing performance in this way does not give any hints how an algorithm would behave for other budgets. Instead, we proposed in [8] a new way to do "Anytime Benchmarking of Budget-Dependent Algorithms" and implemented this functionality in our COCO platform. The idea is to run several experiments for varying budgets and report target-based runtimes in the form of empirical cumulative distribution functions (aka data profiles) as in the case of anytime algorithms.

6.4. Performance Assessment in Multiobjective Optimization

In the context of performance assessment in multiojective optimization, two contributions have been made in 2017. First, we proposed a new visualization method to quantitatively assess the performance of multiobjective optimizers (for 2-objective problems) in the form of average runtime attainment functions [6]. The main idea is to display, for each point in objective space, when (in terms of the average runtime) it has been attained or in other words when it has been dominated by the algorithm. Second, we continued our effort towards automated benchmarking via our COCO platform and described a generic test suite generator that can produce test suites like the previous `bbob-biobj` test suite for an arbitrary number of objectives[2].

6.5. Comparing Continous Optimizers Platform

Participants: Anne Auger, Dimo Brockhoff, Nikolaus Hansen, Umut Batu, Dejan Tusar.

Thanks to the ADT support for Dejan Tušar (since November, previously supported by ESA) and Umut Batu (since July), as well as due to an increased effort from the core development team, we could progress on several aspects regarding our Comparing Continuous Optimizers platform (COCO, https://github.com/numbbo/coco) in 2017.

Most notably, we provide the new functionality of data archives which allows to access the available data of 200+ algorithms much easier. We also made significant progress towards a first constrained test suite—in particular did we add logging support for constrained problems. The postprocessing module is finally python 3 compatible and zip files are supported as input files. The reference worst f-values-of-interest are exposed to the (multiobjective) solver, algorithms can now be displayed in the background, and simplified example experiment scripts (in python) are available (for both anytime, and budget-dependend algorithms, see also [8]). We also improved our continuous integration support, now using also CircleCI and AppVeyor in addition to Inria's Jenkins system. Version 2.0, released in January 2017, saw new functionality of reference algorithms for the multiobjective test suite, a new format of reference algorithms that allow to use any existing data set as reference, improved HTML output and navigation, the COCO version number being part of the plots now, and new regression tests for all provided test suites.

COCO facts for 2017

- 218 issues closed
- major release 2.0 in January plus three additional releases, version 2.2 planned for January 2018
- 10 new contributors outside the main development team
- 14 new algorithm data sets made public (total: 233)

Currently, we are working on an entire rewrite of the postprocessing (ADT COCOpost project of Umut Batu), an improved cocoex module for proposing test suites, functions, data loggers etc. in python (ADT COCOpysuites of Dejan Tušar), a first constrained test suite (in particular Asma Atamna via the PGMO project NumBER), and a large-scale test suite (part of Konstantinos Varelas' PhD thesis, based on the PhD work of Ouassim AitElHara).

Finally, we continued to use COCO also for teaching, in particular for the group project ("controle continue") of our Introduction to Optimization (about 40 Master students) and the Derivative-Free Optimization lectures at Université Paris-Sud (about 30 Master students).

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Grants and Contracts with Industry

- CIFRE-DGA with Thales, for the PhD of Konstantinos Varelas (2017—2020)

---

[2]Paper to be submitted to the Evolutionary Computation journal and to arXiv/HAL in January 2018.

- contract with Storengy to finance a part of the PhD of Cheikh Touré (2017—2020)

8. Partnerships and Cooperations

8.1. Regional Initiatives

- PGMO project "NumBER: Numerical Black Box Optimization for Energy Applications", in collaboration with EDF, financing the postdoc of Asma Atamna, project length: 2 years (2016–2018), PI: Anne Auger
- PGMO project "AESOP: Algorithms Expensive Simulation-Basd Optimization Problems", a project involving several researchers from CentraleSupelec, Ecole des Mines de St.-Etienne, INRA Toulouse, JSI (Slovenia), Safran, Ruhr-Universität Bochum (Germany), and TU Dortmund University (Germany), project length: 2 years (2017–2019), PI: Dimo Brockhoff

8.2. National Initiatives

8.2.1. ANR

- ANR project "NumBBO: Analysis, Improvement and Evaluation of Numerical Blackbox Optimizers", with partners DOLPHIN team (till 2016), Ecole des Mines de St.-Etienne and TU Dortmund University (Germany), Anne Auger was PI of this project which had a total budget of 660kEUR (2012–2017)
- ANR project "Big Multiobjective Optimization (BigMO)", Dimo Brockhoff participates in this project through the Inria team BONUS in Lille (2017–2020)

8.3. International Initiatives

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

8.3.1.1. s3-bbo

Title: Threefold Scalability in Any-objective Black-Box Optimization

International Partner (Institution - Laboratory - Researcher):

Shinshu (Japan) - Tanaka-Hernan-Akimoto Laboratory - Hernan Aguirre

Start year: 2015

See also: http://francejapan.gforge.inria.fr/doku.php?id=associateteam

This associate team brings together researchers from the TAO and Dolphin Inria teams with researchers from Shinshu university in Japan. Additionally, researchers from the University of Calais are external collaborators to the team. The common interest is on black-box single and multi-objective optimization with complementary expertises ranging from theoretical and fundamental aspects over algorithm design to solving industrial applications. The work that we want to pursue in the context of the associate team is focused on black-box optimization of problems with a large number of decision variables and one or several functions to evaluate solutions, employing distributed and parallel computing resources. The objective is to theoretically derive, analyze, design, and develop scalable black-box stochastic algorithms including evolutionary algorithms for large-scale optimization considering three different axes of scalability: (i) decision space, (ii) objective space, and (iii) availability of distributed and parallel computing resources.

We foresee that the associate team will make easier the collaboration already existing through a proposal funded by Japan and open-up a long term fruitful collaboration between Inria and Shinshu university. The collaboration will be through exchanging researchers and Ph.D. students and co-organization of workshops.

8.3.2. Inria International Partners

8.3.2.1. Declared Inria International Partners

- We are collaborating with Shinshu University and particularly Youhei Akimoto through our joint associate team.

8.3.2.2. Informal International Partners

- We are collaborating with Tea Tušar from the Josef-Stefan Institute in Ljubljana, Slovenia for extending and maintaining our COCO platform and on benchmarking in general.
- We are collaborating with Jun.-Prof. Tobias Glasmachers from the Ruhr-Universität Bochum in Germany on runtime analysis of adaptive stochastic algorithms.

8.4. International Research Visitors

8.4.1. Visits of International Scientists

- Filip Matzner from Charles University Prague (Czech Republic) - Visit of one month in November 2017 to work on Evolution Strategies for reinforcement learning and classification problems.
- Prof. Dr. Youhei Akimoto from Shinhu University (Japan) - Visit of one month in November 2017 to work on several projects related to theory and algorithm design for large-scale optimization.
- Dr. Alexandre Chotard from KTU (Sweden) - Visit of one month in November 2017 to work on adaptive MCMC.
- Dr. Tea Tušar from the Josef-Stefan Institute (Slovenia) - Visit of one week in November 2017 to work on our projects around (multiobjective) blackbox optimization benchmarking.

8.4.2. Visits to International Teams

8.4.2.1. Research Stays Abroad

- Anne Auger and Dimo Brockhoff visited Jun.-Prof. Tobias Glasmachers and Prof. Günter Rudolph in Dortmund from April 10 till April 14, 2017

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events Organisation

9.1.1.1. General Chair, Scientific Chair

- A. Auger, program chair of the PPSN 2018 conference Coimbra, Portugal

9.1.1.2. Member of the Organizing Committees

- Anne Auger, Dimo Brockhoff, Nikolaus Hansen, and Dejan Tušar, co-organizer of the ACM-GECCO-2017 workshop on Black Box Optimization Benchmarking, together with Tea Tušar
- 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

9.1.2. Scientific Events Selection

9.1.2.1. Chair of Conference Program Committees

- Anne Auger: theory track chair for the ACM-GECCO conference 2018, Kyoto, Japan
- Nikolaus Hansen: co-track chair at ACM-GECCO-2018 for the "Evolutionary Numerical Optimization" track, Kyoto, Japan

- Nikolaus Hansen: co-track chair at ACM-GECCO-2017 for the "Evolutionary Numerical Optimization" track, Berlin, Germany

9.1.2.2. Member of the Conference Program Committees or Reviewer

- Dimo Brockhoff reviewed for ACM-GECCO
- Anne Auger is reviewer for ACM-GECCO, ACM-FOGA , NIPS, ICML

9.1.3. Journal

9.1.3.1. Member of the Editorial Boards

- Anne Auger and Nikolaus Hansen, members of the editorial board of the Evolutionary Computation Journal
- Dimo Brockhoff, co-guest editor of a special issue on Evolutionary Multiobjective Optimization in the Computers & Operations Research journal (issue 79), together with Joshua Knowles, Boris Naujoks, and Karthik Sindhya

9.1.3.2. Reviewer - Reviewing Activities

- Dimo Brockhoff reviewed in 2017 for IEEE Transactions on Evolutionary Computation, the Evolutionary Computation Journal, Natural Computing, PLoS One, Algorithmica, and Optimal Control, Applications and Methods
- Anne Auger reviewed in 2017 for IEEE Transactions on Evolutionary Computation, the Evolutionary Computation Journal, Algorithmica, SIAM Journal on Optimization

9.1.4. Invited Talks and Tutorials

- Dimo Brockhoff: invited tutorial on benchmarking (multiobjective) optimizers at the Symposium on Search-based Software Engineering (SSBSE'2017) in September 2017 in Paderborn, Germany
- Dimo Brockhoff: two invited talks (one on Evolutionary Multiobjective Optimization, one on Benchmarking) at the Mascot-Num conference in March 2017 in Paris, France
- Anne Auger and Nikolaus Hansen, tutorial on *Introduction to randomized continuous optimization* at the ACM-GECCO conference, Berlin, Germany
- Dimo Brockhoff: introductory tutorial on Evolutionary Multiobjective Optimization at the ACM-GECCO conference, Berlin, Germany
- Nikolaus Hansen: tutorial "A Practical Guide to Benchmarking and Experimentation" at the ACM-GECCO conference, Berlin, Germany
- Nikolaus Hansen: tutorial "CMA-ES and Advanced Adaptation Mechanisms" at the ACM-GECCO conference, Berlin, Germany, together with Youhei Akimoto

9.1.5. Leadership within the Scientific Community

- Since 2011, Anne Auger **Elected** member of the **ACM-SIGEVO executive board**, re-elected in 2017.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Master: Dimo Brockhoff, "Introduction to Optimization", 31.5h ETD, M2, Université Paris-Sud, France

Master: Anne Auger and Dimo Brockhoff, "Advanced Optimization", 31.5h ETD, M2, Université Paris-Sud, France

Master: Anne Auger, "Derivative-free Optimization", Paris-Saclay, Optimization Master

Master: Anne Auger : Anne Auger (Introduction to Machine Learning, Advanced Machine Learning), Ecole Polytechnique, c.a. 50h

Summer school: Anne Auger and Dimo Brockhoff, July 3-7, 2017. CEA-EDF-Inria **summer school** on *Design and optimization under uncertainty of large-scale numerical models*. Course on *Introduction to Randomized Black-Box Numerical Optimization and CMA-ES*, Paris.

9.2.2. Supervision

PhD in progress: Cheikh Touré, topic: multiobjective optimization, started in October 2017, supervised by Anne Auger and Dimo Brockhoff

PhD in progress: Konstantinos Varelas, topic: constrained and expensive optimization, started in December 2017, supervised by Anne Auger and Dimo Brockhoff

9.2.3. Juries

- A. Auger in the PhD jury of Paul Feliot (defense July 2017)

10. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] O. AIT ELHARA. *Stochastic Black-Box Optimization and Benchmarking in Large Dimensions*, Université Paris-Saclay, July 2017, https://tel.archives-ouvertes.fr/tel-01615829

Articles in International Peer-Reviewed Journals

[2] A. CHOTARD, A. AUGER. *Verifiable Conditions for the Irreducibility and Aperiodicity of Markov Chains by Analyzing Underlying Deterministic Models*, in "Bernoulli", 2017, https://arxiv.org/abs/1508.01644 , forthcoming, https://hal.inria.fr/hal-01222222

[3] Y. OLLIVIER, L. ARNOLD, A. AUGER, N. HANSEN. *Information-Geometric Optimization Algorithms: A Unifying Picture via Invariance Principles*, in "Journal of Machine Learning Research", 2017, vol. 18, n° 18, pp. 1-65, https://hal.inria.fr/hal-01515898

International Conferences with Proceedings

[4] Y. AKIMOTO, A. AUGER, N. HANSEN. *Quality Gain Analysis of the Weighted Recombination Evolution Strategy on General Convex Quadratic Functions*, in "Proceedings of the 14th ACM/SIGEVO Conference on Foundations of Genetic Algorithms, FOGA", Copenhagen, Denmark, January 2017, pp. 111-126 [*DOI :* 10.1145/3040718.3040720], https://hal.inria.fr/hal-01516326

[5] A. ATAMNA, A. AUGER, N. HANSEN. *Linearly Convergent Evolution Strategies via Augmented Lagrangian Constraint Handling*, in "The 14th ACM/SIGEVO Workshop on Foundations of Genetic Algorithms (FOGA XIV)", Copenhagen, Denmark, January 2017, pp. 149 - 161 [*DOI :* 10.1145/3040718.3040732], https://hal.inria.fr/hal-01455379

[6] D. BROCKHOFF, A. AUGER, N. HANSEN, T. TUŠAR. *Quantitative Performance Assessment of Multiobjective Optimizers: The Average Runtime Attainment Function*, in "Evolutionary Multi-Criterion Optimization (EMO 2017)", Münster, Germany, LNCS, March 2017, vol. 10173, pp. 103-119 [*DOI :* 10.1007/978-3-319-54157-0_8], https://hal.inria.fr/hal-01591151

[7] D. M. Nguyen, N. Hansen. *Benchmarking CMAES-APOP on the BBOB noiseless testbed*, in "Proceedings of the 2017 Genetic and Evolutionary Computation Conference Companion (GECCO '17 Companion)", Berlin, Germany, July 2017, pp. 1756 - 1763 [*DOI :* 10.1145/2908812.2908864], https://hal.inria.fr/hal-01591423

[8] T. Tušar, N. Hansen, D. Brockhoff. *Anytime Benchmarking of Budget-Dependent Algorithms with the COCO Platform*, in "IS 2017 - International multiconference Information Society", Ljubljana, Slovenia, October 2017, pp. 1-4, https://hal.inria.fr/hal-01629087

Other Publications

[9] Y. Akimoto, A. Auger, N. Hansen. *Quality Gain Analysis of the Weighted Recombination Evolution Strategy on General Convex Quadratic Functions*, December 2017, Submitted to Journal of Theoretical Computer Science, https://hal.inria.fr/hal-01662568

[10] A. Atamna, A. Auger, N. Hansen. *On Invariance and Linear Convergence of Evolution Strategies with Augmented Lagrangian Constraint Handling*, December 2017, working paper or preprint, https://hal.inria.fr/hal-01660728

References in notes

[11] 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

[12] 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

[13] 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

[14] A. Atamna, A. Auger, N. Hansen. *Linearly Convergent Evolution Strategies via Augmented Lagrangian Constraint Handling*, in "Foundation of Genetic Algorithms (FOGA)", 2017

[15] 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° 3, pp. 1589-1624

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

[17] 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° 2

[18] 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

[19] 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

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

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

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

[23] 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$^{\text{o}}$ 1, 127 p.

[24] 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

[25] 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

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

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

[28] 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$^{\text{o}}$ 35, pp. 14271–14276