HAL
open science

Criteria and Convergence Rates in Noisy Optimization

Sandra Astete-Morales, Marie-Liesse Cauwet, Olivier Teytaud

To cite this version:

Sandra Astete-Morales, Marie-Liesse Cauwet, Olivier Teytaud. Criteria and Convergence Rates in Noisy Optimization. Genetic and Evolutionary Computation Conference (GECCO 2015), Jul 2015, Madrid, Spain. 2015, 10.1145/2739482.2764722 . hal-01217128v2

HAL Id: hal-01217128

https://inria.hal.science/hal-01217128v2

Submitted on 19 Jan 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Criteria and Convergence Rates in Noisy Optimization

Sandra Astete-Morales
TAO, Inria, Lri, Umr Cnrs 8623
Bat. 650, Univ. Paris-Sud
91405 Orsay Cedex, France
sandra-cecilia.astete-morales@inria.fr

Marie-Liesse Cauwet
TAO, Inria, Lri, Umr Cnrs 8623
Bat. 650, Univ. Paris-Sud
91405 Orsay Cedex, France
marie-liesse.cauwet@inria.fr

Olivier Teytaud
TAO, Inria, Lri, Umr Cnrs 8623
Bat. 650, Univ. Paris-Sud
91405 Orsay Cedex, France
olivier.teytaud@inria.fr

ABSTRACT

In an optimization framework, some criteria might be more relevant than others; the internal computational cost of the optimization algorithm might be negligible or not; the quality of intermediate search points might be important or not. For this reason measuring the performance of an algorithm is a delicate task. In addition, the usual criteria are often approximated for the sake of simplicity of the analysis, or for simplifying the design of test beds. This situation makes sense both in noise-free and noisy settings; however it is more often crucial in the latter case. We here discuss and compare several performance criteria published in the literature in the case of noisy optimization. We review existing rates, for various existing criteria, propose new rates, and check if some classically observed criteria are good approximations of sound criteria.

Categories and Subject Descriptors

G.1.6 [**Optimization**]: Unconstrained optimization

Keywords

Noisy optimization

1. FRAMEWORK AND CRITERIA

Given a fitness function $F : D \in \mathbb{R}^d \to \mathbb{R}$, also known as objective function, optimization (minimization) is the search for the optimum point x^* such that $\forall x \in D, F(x^*) \leq F(x)$. The fitness function may be corrupted by noise. A common case of noise is the additive noise case. Given a search point $x \in D$, evaluating F in x results in an altered fitness value $f(x, w)$ as follows:

$$f(x, w) = F(x) + w, \tag{1}$$

where w is an independent random variable of mean zero and variance σ. An optimization algorithm generates $(x_n)_{n\geq 1}$, successive *search points* at which the objective function is evaluated - in a noisy manner. It can also generate $(\tilde{x}_n)_{n\geq 1}$ which are *recommendations* or *approximations of the optimum*. $\tilde{x}_n$ is provided after n fitness evaluations are performed.

[**Simple Regret**] (SR):

$$SR_n = F(\tilde{x}_n) - F(x^*).$$

[**Approximate Simple Regret**] (ASR):

$$ASR_n = \min_{m\leq n} F(x_m) - F(x^*).$$

[**Robust Simple Regret**] (RSR):

$$RSR_n = \min_{k\leq n} \max_{k-g(k)<m\leq k} \left(F(\tilde{x}_m) - F(x^*)\right),$$

where $g(n)$ is a polylogarithmic function of n.

[**Cumulative Regret**] (CR):

$$CR_n = \sum_{i=1}^{n} \left(F(x_i) - F(x^*)\right).$$

[**Average Regret**] (AR):

$$AR_n = \frac{1}{n} CR_n.$$

We define the corresponding "slope" of the various regrets introduced previously by:

$$s(*R) = \limsup_{n\to\infty} \log(*R_n)/\log(n),$$

where $*R$ stands for SR, ASR, RSR, CR or AR and log is the natural logarithm. From now on, $s(.)$ stands for "slope of". The slope is a random variable; however in many cases it is almost surely equal to some constant. We also defined the slope in expectation, as follows:

$$s^{\mathbb{E}}(*R) = \limsup_{n\to\infty} \log(\mathbb{E}(*R_n))/\log(n). \tag{2}$$

2. CONCLUSION

In this paper we analyse several criteria for the performance of algorithms in noisy optimization problems. We provide several rates for each of the criteria and compare them. Some are rigorously proved, others are conjectured. Table 1 summarizes our results.

Simple regret. From theory, it appears that Evolution strategies, when correctly tuned in terms of resamplings, essentially reach half the speed of classical noisy optimization

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

GECCO '15 July 11-15, 2015, Madrid, Spain

© 2016 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-3488-4/15/07.

DOI: http://dx.doi.org/10.1145/2739482.2764722

Algorithm	Slope of the simple regret $s(SR)$	Slope of the approximate simple regret $s(ASR)$	Slope of the robust simple regret $s(RSR)$	Slope of the average regret $s(AR)$
Random search	$\boxed{0}$	$\boxed{-\frac{2}{d}}$	$\boxed{-\frac{2}{d}}$	$\boxed{0}$
ES	0	$-\frac{2}{d}$	$-\frac{2}{d}$	0
ES+resampling	$-\frac{1}{2}^{\sharp}$	$-\frac{1}{2}$	$-\frac{1}{2}^{\sharp}$	$\boxed{-\frac{1}{2}^{\sharp}}$
Modified ES+ resampling	$-\frac{1}{2}^{\sharp}$	$\boxed{-\frac{1}{2}-\frac{2}{d}^{\sharp}}$	$\boxed{-\frac{1}{2}-\frac{2}{d}^{\sharp}}$	$\boxed{-\frac{1}{2}^{\sharp}}$
Shamir (original)	$\boxed{-1^{\dagger}}$	$\boxed{0^{\dagger}}$ and 0^{*}	$\boxed{-1^{\dagger}}$	$\boxed{0}$
Shamir (adapted for ASR)	$\boxed{-1^{\dagger}}$	$\boxed{-1^{\dagger}}$	$\boxed{-1^{\dagger}}$	$\boxed{0}$
Fabian (original)	$\boxed{-(1-e)^{*}}$	$\boxed{-e'^{*}}$	$\boxed{-(1-e)^{*}}$	$\boxed{-e'^{*}}$
Fabian (adapted for ASR)	$\boxed{-(1-e)^{*}}$	$\boxed{-(1-e)^{*}}$	$\boxed{-(1-e)^{*}}$	
Fabian (adapted for AR)				$\boxed{-\frac{1}{2}+e^{*}}$

Table 1: Table of regrets for various algorithms. Rates with $\sharp$ means that the convergence is with high probability; for numbers with $\dagger$, the convergence is in expectation and the convergence is a.s ones with $*$. When nothing is specified, the convergence holds for the 3 different types. Boxed results are proved and others are conjectured.

algorithms in terms of simple regret, on the log-log scale. This corresponds to a squared computation time.

Cumulative regret. In terms of CR, however, they reach the same rate as noisy optimization algorithms, e.g. as Fabian's algorithm with a proper choice of constants ([1] considers mainly simple regret, but their result implies optimal rates for CR as well, for a proper choice of parameters), or the best rate in [2].

Approximations. The ASR is a poor approximation of simple regret since: (i) A fast algorithm for SR can have a bad rate for ASR (see e.g. Fabian and Shamir algorithm). However, this can be solved by modifying the algorithm in order to sample, sometimes, a recommended point. (ii) A slow algorithm for SR can have a fast rate for ASR, and this is a more serious problem. There is no simple patch for that. Test beds using ASR will overestimate algorithms which include random exploration. This is partially, but not totally, solved by RSR. However, the difference decreases with the dimension for most algorithms. Indeed, ASR is a better estimate of AR than an estimate of SR, as in many cases $s(ASR) = s(AR)$, except for algorithms which are specifically adapted for ASR. Incidentally, ASR was proposed in the evolutionary computation community, and AR (or, equivalently, CR) is the criterion for which evolutionary computation performs best in noisy optimization.

We did not come up with a satisfactory criterion, which would be consistent with SR (at least, same slope) and without drawbacks as those presented above, except the simple regret itself. The drawback of SR is that it is not necessarily non-increasing, which is an issue for the concept of "first hitting time". Another further work is the refinement of the theoretical analysis. We have compared slopes, but in some cases we have slopes for almost sure convergence, in other cases slope with high probability, and in others slope in expectation.

3. REFERENCES

[1] V. Fabian. Stochastic Approximation of Minima with Improved Asymptotic Speed. *Annals of Mathematical statistics*, 38:191–200, 1967.

[2] O. Shamir. On the complexity of bandit and derivative-free stochastic convex optimization. In *COLT 2013 - The 26th Annual Conference on Learning Theory, June 12-14, 2013, Princeton University, NJ, USA*, pages 3–24, 2013.