On the Complexity of Bandit and Derivative-Free Stochastic Convex Optimization

Ohad Shamir
Microsoft Research
ohadsh@microsoft.com

Abstract

The problem of stochastic convex optimization with bandit feedback (in the learning community) or without knowledge of gradients (in the optimization community) has received much attention in recent years, in the form of algorithms and performance upper bounds. However, much less is known about the inherent complexity of these problems, and there are few lower bounds in the literature, especially for nonlinear functions. In this paper, we investigate the attainable error/regret in the bandit and derivative-free settings, as a function of the dimension d and the available number of queries T. We provide a precise characterization of the attainable performance for strongly-convex and smooth functions, which also imply a non-trivial lower bound for more general problems. Moreover, we prove that in both the bandit and derivative-free setting, the required number of queries must scale at least quadratically with the dimension. Finally, we show that on the natural class of quadratic functions, it is possible to obtain a "fast" $\mathcal{O}(1/T)$ error rate in terms of T, under mild assumptions, even without having access to gradients. To the best of our knowledge, this is the first such rate in a derivative-free stochastic setting, and holds despite previous results which seem to imply the contrary.

1 Introduction

This work considers the following fundamental question: Given an unknown convex function F, and the ability to query for (possibly noisy) realizations of its values at various points, how can we optimize F with as few queries as possible?

This question, under different guises, has played an important role in several communities. In the optimization community, this is usually known as "zeroth-order" or "derivative-free" convex optimization, since we only have access to function values rather than gradients or higher-order information. The goal is to return a point with small optimization error on some convex domain, using a limited number of queries. Derivative-free methods were among the earliest algorithms to numerically solve unconstrained optimization problems, and have recently enjoyed increasing interest, being especial useful in black-box situations where gradient information is hard to compute or does not exist [15, 18]. In a stochastic framework, we can only obtain noisy realizations of the function values (for instance, due to running the optimization process on sampled data). We refer to this setting as *derivative-free SCO* (short for stochastic convex optimization).

In the learning community, these kinds of problems have been closely studied in the context of multi-armed bandits and (more generally) bandit online optimization, which are powerful models for sequential decision making under uncertainty [8, 6]. In a stochastic framework, these settings correspond to repeatedly choosing points in some convex domain, obtaining noisy realizations of some underlying convex function's value. However, rather than minimizing optimization error, our goal is to minimize the (average) regret: roughly speaking, that the average of the function values we obtain is not much larger than the minimal function value. For example, the well-known multi-armed bandit problem corresponds to a linear function over the simplex. We refer to this setting as *bandit SCO*. It can be shown that any algorithm which attains small average regret can be converted to an algorithm with the same optimization error. In other words, bandit SCO is only harder than derivative-free SCO.

When one is given gradient information, the attainable optimization error / average regret is well-known: under mild conditions, it is $\Theta(1/\sqrt{T})$ for convex functions and $\Theta(1/T)$ for strongly-convex functions, where T is the number of queries [19, 12, 16]. Note that these bounds do not explicitly depend on the dimension of the domain.

The inherent complexity of bandit/derivative-free SCO is not as well-understood. An important exception is multi-armed bandits, where the attainable error/regret is known to be exactly $\Theta(\sqrt{d/T})$, where d is the dimension and T is the number of queries[1] [5, 4]. Linear functions over other convex domains has also been explored, with upper bounds on the order of $\mathcal{O}(\sqrt{d/T})$ to $\mathcal{O}(\sqrt{d^2/T})$ (e.g. [1, 7]). For linear functions over general domains, information-theoretic $\Omega(\sqrt{d^2/T})$ lower bounds on the regret has been proven in [9, 10]. However, we emphasize that these lower bounds are on regret, not optimization error, and were shown on *non-convex* domains. This falls outside the scope of stochastic *convex* optimization we consider here, where the convexity generally ensures computational tractability.

When dealing with more general, non-linear functions, much less is known. The problem was originally considered over 30 years ago, in the seminal work by Yudin and Nemirovsky on the complexity of optimization [14]. The authors provided some algorithms and upper bounds, but as they themselves emphasize (cf. pg. 359), the attainable complexity is far from clear. Quite recently, [13] provided an $\Omega(\sqrt{d/T})$ lower bound for strongly-convex functions, which demonstrates that the "fast" $\mathcal{O}(1/T)$ rate in terms of T, that one enjoys with gradient information, is not possible here. In contrast, the current best-known upper bounds are $\mathcal{O}(\sqrt[4]{d^2/T}), \mathcal{O}(\sqrt[3]{d^2/T}), \mathcal{O}(\sqrt{d^2/T})$ for convex, strongly-convex, and strongly-convex-and-smooth functions respectively [11, 2]; And a $\mathcal{O}(\sqrt{d^{32}/T})$ bound for convex functions [3], which is better in terms of dependence on T but very bad in terms of the dimension d.

In this work, we investigate the complexity of bandit and derivative-free stochastic convex optimization, focusing on nonlinear functions, with the following contributions (see also the summary in Table 1):

- We prove that for strongly-convex and smooth functions, the attainable error/regret is exactly $\Theta(\sqrt{d^2/T})$. This has three important ramifications: First of all, it settles the question of attainable performance for such functions, and is the first sharp characterization of complexity for a general nonlinear bandit/derivative-free class of problems. Second, it proves that the required number of queries T in such problems must scale quadratically with the dimension, even in the easier derivative-free setting, and in contrast to the linear case which often allows linear scaling with the dimension. Third, it formally provides a $\Omega(\sqrt{d^2/T})$ lower bound for more general classes of convex problems, which is stronger than the $\Omega(\sqrt{d/T})$ lower bound known so far, e.g. through multi-armed bandits.

- We analyze an important special case of strongly-convex and smooth functions, namely quadratic functions. We show that for such functions, one can (efficiently) attain $\Theta(d^2/T)$ optimization error, and that this rate is sharp. To the best of our knowledge, it is the first general class of nonlinear functions for which one can show a "fast rate" (in terms of T) in a derivative-free stochastic setting. In fact, this may seem to contradict the result in [13], which shows an $\Omega(\sqrt{d/T})$ lower bound on quadratic functions. However, there is no contradiction, since the example establishing the lower bound of [13] imposes an extremely small domain (which actually decays with T). Our d^2/T result holds for a fixed domain, and under the mild assumption that either the minimum of the quadratic function is bounded away from the domain boundary, or that we can query points slightly outside the domain.

- We prove that even for quadratic functions, the attainable average *regret* is exactly $\Theta(\sqrt{d^2/T})$, in contrast to the $\Theta(d^2/T)$ result for optimization error. This shows there is a real gap between what can be obtained for derivative-free SCO and bandit SCO, at least for such functions. Again, this stands in contrast to previously studied settings such as multi-armed bandits, where there is no difference in performance.

[1] In a stochastic setting, a more common bound in the literature is $\mathcal{O}(d\log(T)/T)$, but the $\mathcal{O}$-notation hides a non-trivial dependence on the form of the underlying linear function (in multi-armed bandits terminology, a gap between the expected rewards bounded away from 0). Such assumptions are not natural in a nonlinear bandits SCO setup, and without them, the regret is indeed $\Theta(\sqrt{d/T})$. See for instance [6, Chapter 2] for more details.

	Derivative-Free SCO		Bandit SCO	
Function Type	**Upper Bound**	**Lower Bound**	**Upper Bound**	**Lower Bound**
Quadratic	$\boxed{\mathbf{d^2/T}}$		$\boxed{\sqrt{\mathbf{d^2/T}}}$	
Str. Convex and Smooth	$\boxed{\sqrt{\mathbf{d^2/T}}}$			
Str. Convex	$\min\left\{\sqrt[3]{d^2/T}, \sqrt{d^{32}/T}\right\}$	$\boxed{\sqrt{\mathbf{d^2/T}}}$	$\min\left\{\sqrt[3]{d^2/T}, \sqrt{d^{32}/T}\right\}$	$\boxed{\sqrt{\mathbf{d^2/T}}}$
Convex	$\min\left\{\sqrt[4]{d^2/T}, \sqrt{d^{32}/T}\right\}$	$\boxed{\sqrt{\mathbf{d^2/T}}}$	$\min\left\{\sqrt[4]{d^2/T}, \sqrt{d^{32}/T}\right\}$	$\boxed{\sqrt{\mathbf{d^2/T}}}$

Table 1: A summary of the complexity results for derivative-free stochastic convex optimization (optimization error) and bandit stochastic convex optimization (average regret), for various function classes and in terms of the dimension d and the number of queries T. The boxed results are shown in this work. The upper bounds for the convex and strongly convex case combine results from [11, 2, 3]. The table shows dependence on d, T only and ignores other factors and constants.

2 Outline of Main Results and Proof Ideas

Due to lack of space, we only present a brief sketch of our main results and proof techniques. All details and additional results appear in a separate arXiv technical report [17].

We let $F(\cdot) : \mathcal{W} \mapsto \mathbb{R}$ denote the convex function of interest, where $\mathcal{W} \subseteq \mathbb{R}^d$ is a (closed) convex domain. To prevent trivialities, we consider in this work only "nice" functions, in the sense that their optimum $\mathbf{w}^*$ is known beforehand to lie in some bounded domain (even if $\mathcal{W}$ is large or all of $\mathbb{R}^d$), and the function is Lipschitz in that domain (with a Lipschitz parameter independent of the dimension).

The learning/optimization process proceeds in T rounds. Each round t, we pick and query a point $\mathbf{w}_t \in \mathcal{W}$, obtaining an independent realization of $F(\mathbf{w}) + \xi_{\mathbf{w}}$, where $\xi_{\mathbf{w}}$ is an unknown zero-mean random variable. In the bandit SCO setting, our goal is to minimize the *expected average regret*, namely $\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} F(\mathbf{w}_t) - F(\mathbf{w}^*)\right]$, whereas in the derivative-free SCO setting, our goal is to compute, based on $\mathbf{w}_1, \ldots, \mathbf{w}_T$ and the observed values, some point $\bar{\mathbf{w}} \in \mathcal{W}$, such that the *expected optimization error*, namely $\mathbb{E}\left[F(\bar{\mathbf{w}}) - F(\mathbf{w}^*)\right]$, is as small as possible.

We begin by considering a sub-class of strongly-convex and smooth functions, namely quadratic functions, which have the form

$$F(\mathbf{w}) = \mathbf{w}^\top A\mathbf{w} + \mathbf{b}^\top \mathbf{w} + c$$

where A is positive-definite with bounded eigenvalues. This is a natural and important class of functions, which in learning applications appears, for instance, in the context of least squares and ridge regression.

To achieve our error bound, we need to make the following mild assumption:

Assumption 1. *At least one of the following holds for some fixed $\epsilon \in (0, 1]$:*

- *The quadratic function attains its minimum $\mathbf{w}^*$ in the domain $\mathcal{W}$, and the Euclidean distance of $\mathbf{w}^*$ from the domain boundary is at least ϵ.*
- *We can query not just points in $\mathcal{W}$, but any point whose distance from $\mathcal{W}$ is at most ϵ.*

With strongly-convex functions, the most common case is that $\mathcal{W} = \mathbb{R}^d$, and then both cases actually hold for any value of ϵ. Even in other situations, one of these assumptions virtually always holds. Note that we crucially rely here on the strong-convexity assumption: with (say) linear functions, the domain must always be bounded and the optimum always lies at the boundary of the domain.

With this assumption, we obtain the following bound:

Theorem 1. *Let $F(\mathbf{w}) = \mathbf{w}^\top A\mathbf{w} + \mathbf{b}^\top \mathbf{w} + c$ be a λ-strongly convex function, where $\|A\|_2, \|\mathbf{b}\|, |c|$ are all at most 1, and suppose the optimum $\mathbf{w}^*$ has a norm of at most B. Then under Assumption 1, there is an efficient algorithm which returns $\bar{\mathbf{w}}$ such that*

$$\mathbb{E}\left[F(\bar{\mathbf{w}}) - F(\mathbf{w}^*)\right] \leq \frac{4(4 + 5\log(2))(B + 1)^4}{\lambda\epsilon^2}\frac{d^2}{T}.$$

As discussed earlier, [13] recently proved a $\Omega(\sqrt{d/T})$ lower bound for derivative-free SCO, which actually applies to quadratic functions. This does not contradict our result, since in their example the diameter of $\mathcal{W}$ (and hence also ϵ) decays with T. In contrast, our $\mathcal{O}(d^2/T)$ bound holds for fixed ϵ, which we believe is natural in most applications.

The algorithm we use is based on a well-known 1-point gradient estimate technique, which allows us to get an unbiased estimate of the gradient at any point by randomly querying for a (noisy) value of the function around it (see [14, 11]). Our key insight is that whereas for general functions one must query very close to the point of interest (scaling to 0 with T), quadratic functions have additional structure which allows us to query relatively far away, allowing gradient estimates with much smaller variance. It is interesting to note that due to central-limit considerations, it seems that a gradient-based approach seems essential to obtain $\mathcal{O}(1/T)$ rates (in terms of T), in contrast to other derivative-free optimization techniques based on direct function-value comparisons.

We are also able to obtain a matching lower bound:

Theorem 2. *Let the number of rounds T be fixed. Then for any (possibly randomized) querying strategy, there exists a quadratic function of the form $F(\mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|^2 - \langle \mathbf{e}, \mathbf{w}\rangle$, which is minimized at $\mathbf{e}$ where $\|\mathbf{e}\| \leq 1$, such that the resulting $\bar{\mathbf{w}}$ satisfies*

$$\mathbb{E}[F(\bar{\mathbf{w}}) - F(\mathbf{w}^*)] \geq 0.01 \min\left\{1, \frac{d^2}{T}\right\}.$$

Note that since $\|\mathbf{e}\| \leq 1$, we know in advance that the optimum must lie in the unit Euclidean ball. Despite this, the lower bound holds even if we do not restrict at all the domain in which we are allowed to query - i.e., it can even be all of $\mathbb{R}^d$. The proof is based on choosing $\mathbf{e}$ randomly and reducing the problem to a series of "hypothesis-testing" problem, namely determining the sign of each coordinate e_i.

The results above were for optimization error. Interestingly, it turns out that for regret, one can show a much stronger lower bound:

Theorem 3. *Let the number of rounds T be fixed. Then for any (possibly randomized) querying strategy, there exists a quadratic function of the form $F(\mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|^2 - \langle \mathbf{e}, \mathbf{w}\rangle$, which is minimized at $\mathbf{e}$ where $\|\mathbf{e}\| \leq 1/2$, such that*

$$\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} F(\mathbf{w}_T) - F(\mathbf{w}^*)\right] \geq 0.02 \min\left\{1, \sqrt{\frac{d^2}{T}}\right\}.$$

Note that our lower bound holds even when the domain is unrestricted (the algorithm can pick any point in $\mathbb{R}^d$). Moreover, the lower bound coincides (up to a constant) with the $\mathcal{O}(\sqrt{d^2/T})$ regret upper-bound shown for strongly-convex and smooth functions in [2]. This shows that for strongly-convex and smooth functions, the minimax average regret is $\Theta(\sqrt{d^2/T})$. Also, the lower bound implies that one cannot hope to obtain average regret better than $\sqrt{d^2/T}$ for more general bandit problems, such as strongly-convex or even convex problems. The proof relies on techniques similar to the lower bound of Thm. 2, with a key additional insight. Specifically, in Thm. 2, the lower bound obtained actually depends on the norm of the points $\mathbf{w}_1, \ldots, \mathbf{w}_T$, and the optimal $\mathbf{w}^*$ has a very small norm. In a regret minimization setting the points $\mathbf{w}_1, \ldots, \mathbf{w}_T$ cannot be too far from $\mathbf{w}^*$, and thus must have a small norm as well, leading to a stronger lower bound than that of Thm. 2.

Finally, we prove that for strongly-convex and smooth functions (not necessarily quadratic), even the optimization error cannot be better than $\Omega(\sqrt{d^2/T})$:

Theorem 4. *Let the number of rounds T be fixed. Then for any (possibly randomized) querying strategy, there exists a function F over $\mathbb{R}^d$ which is 0.5-strongly convex and 3.5-smooth; Is 4-Lipschitz over the unit Euclidean ball; has a global minimum in the unit ball; And such that the resulting $\bar{\mathbf{w}}$ satisfies*

$$\mathbb{E}[F(\bar{\mathbf{w}}) - F(\mathbf{w}^*)] \geq 0.004 \min\left\{1, \sqrt{\frac{d^2}{T}}\right\}.$$

The general proof technique is rather similar to that of Thm. 2, but the construction is more intricate. Instead of considering quadratics, we consider functions which behave like quadratics close to the optimum, but become more and more indistinguishable when queried far away from the optimum. This makes the optimization problem, in a sense, as hard as achieving low regret (i.e. there is no use querying far from the optimum), leading to a similar lower bound.

References

[1] Y. Abbasi-Yadkori, D. Pál, and C. Szepesvári. Improved algorithms for linear stochastic bandits. In *NIPS*, 2011.

[2] A. Agarwal, O. Dekel, and L. Xiao. Optimal algorithms for online convex optimization with multi-point bandit feedback. In *COLT*, 2010.

[3] A. Agarwal, D. Foster, D. Hsu, S. Kakade, and A. Rakhlin. Stochastic convex optimization with bandit feedback. In *NIPS*, 2011.

[4] J.-Y. Audibert and S. Bubeck. Minimax policies for adversarial and stochastic bandits. In *COLT*, 2009.

[5] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. Schapire. The nonstochastic multiarmed bandit problem. *SIAM J. Comput.*, 32(1):48–77, 2002.

[6] S. Bubeck and N. Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *CoRR*, abs/1204.5721, 2012.

[7] S. Bubeck, N. Cesa-Bianchi, and S. Kakade. Towards minimax policies for online linear optimization with bandit feedback. In *COLT*, 2012.

[8] N. Cesa-Bianchi and G. Lugosi. *Prediction, learning, and games*. Cambridge University Press, 2006.

[9] V. Dani, T. Hayes, and S. Kakade. The price of bandit information for online optimization. In *NIPS*, 2007.

[10] V. Dani, T. Hayes, and S. Kakade. Stochastic linear optimization under bandit feedback. In *COLT*, 2008.

[11] A. Flaxman, A. Kalai, and B. McMahan. Online convex optimization in the bandit setting: gradient descent without a gradient. In *SODA*, 2005.

[12] E. Hazan and S. Kale. Beyond the regret minimization barrier: an optimal algorithm for stochastic strongly-convex optimization. In *COLT*, 2011.

[13] K. Jamieson, R. Nowak, and B. Recht. Query complexity of derivative-free optimization. *CoRR*, abs/1209.2434, 2012.

[14] A. Nemirovsky and D. Yudin. *Problem Complexity and Method Efficiency in Optimization*. Wiley-Interscience, 1983.

[15] Y. Nesterov. Random gradient-free minimization of convex functions. Technical Report 16, ECORE Discussion Paper, 2011.

[16] A. Rakhlin, O. Shamir, and K. Sridharan. Making gradient descent optimal for strongly convex stochastic optimization. In *ICML*, 2012.

[17] O. Shamir. On the complexity of bandit and derivative-free stochastic convex optimization. *CoRR*, abs/1209.2388, 2012.

[18] S. Stich, C. Müller, and B. Gärtner. Optimization of convex functions with random pursuit. *CoRR*, abs/1111.0194, 2011.

[19] M. Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *ICML*, 2003.