A Variance-Adaptive Lower Bound for Simulation Optimization in Continuous Space

Jianzhong Du
School of Management
University of Science and Technology of China
Anhui, China
dujianzhong@ustc.edu.cn

L. Jeff Hong
Dept. of Industrial and Systems Engineering
University of Minnesota
Minneapolis, MN 55455
lhong@umn.edu

Abstract

This paper considers the simulation optimization with continuous decision variables. Under certain reasonable assumptions, we provide a worst-case lower bound on the optimization error for any algorithms. The lower bound can incorporate the noiseless and noisy problems in an unified framework. The result highlights that the optimization error of noisy problems can be very close to that of noiseless problems when the observation's variance is small and the budget is not very large.

1 Introduction

This paper targets the minimization of a continuous objective function under the budget constraint on the total number of observations that we can make. This problem setting is common in operations research and machine learning. For example, we may need simulation to analyze and optimize the performance of complex systems in supply chains and healthcare (Hong and Nelson, 2025). Running simulation is time-consuming and the simulation observation is often stochastic. Another example is the hyper-parameter tuning to optimize the performance of a machine learning model. The model may be very complicated and regarded as a black-box. In the literature, such type of problems is called simulation optimization with continuous decision variables (CSO), global optimization or black-box optimization. In this paper, we call it CSO.

CSO has three distinctive features. First, the objective function may be non-convex and non-smooth. Second, unbiased observations of the gradient are often unavailable. Third, observations of the objective may be noisy. Given the finite total budget of noisy observations, the estimated optimal solution of a CSO algorithm cannot be the true optimal solution with probability one. A common efficiency measure is the optimization error (also called simple regret), which quantifies how far the objective value of the algorithm's estimated optimal solution is from the optimal objective value.

Many popular algorithms have been proposed to solve different types of CSO. Suppose the observations are noisy. Let n denote the total budget. For strongly convex CSO (i.e., the objective is strongly convex), the stochastic approximation (Hong et al., 2020; Hu and Fu, 2024) can achieve the optimization error $O(n^{-1/2})$. When the objective is quadratic, the optimization error can be further reduced to $O(n^{-1})$ (Shamir, 2013). Besides the CSO whose observations are noisy, we may also come across noiseless CSO whose observations are deterministic. The noiseless setting can be considered as the special case of the noisy setting where the variance of observations is zero. Algorithms for noisy CSO may be applied to noiseless CSO. But the optimization error can be significantly smaller if the algorithm is specialized to the noiseless CSO. For example, if the objective is strongly convex, the optimization error can be exponentially small (Munos, 2014).

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: MLxOR: Mathematical Foundations and Operational Integration of Machine Learning for Uncertainty-Aware Decision-Making.

The variance of observations (denoted by σ^2) is an important factor to solve CSO efficiently. Bartlett et al. (2019) proposed a simple algorithm whose optimization error is adaptive to the variance. They found that there exists a threshold such that the optimization error is the polynomial rate $\tilde{O}((\sigma^2/n)^{1/2})$ if the variance is above the threshold but the near-exponential rate $O(\exp(-cn/(\log n)^2))$ for a strictly positive constant c, independent of σ^2, if the variance is smaller than the threshold. The threshold for variance depends on the budget n and tends to decrease to zero as n grows. Thus, as long as the variance is non-zero, the optimization error will finally become the polynomial rate when n is large enough. However, if the variance is small, the optimization error may be the near-exponential rate, same as that for noiseless CSO, before n becomes very large.

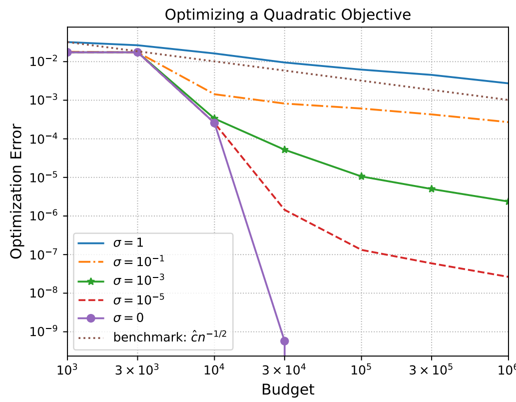

Figure 1: The decrease of optimization error when the observation's variance is different.

In Figure 1, we optimize a quadratic function $y(\mathbf{x})$ whose observation at $\mathbf{x}$ follows the distribution $\mathcal{N}(y(\mathbf{x}), \sigma^2)$ with σ being different values. The main observation from Figure 1 is consistent with the finding in Bartlett et al. (2019): when the variance is small and the budget is not very large, the optimization error of noisy CSO is almost same as that of noiseless CSO ($\sigma = 0$); when the budget is large enough, the optimization error of noisy CSO will finally be larger than that of noiseless CSO. This interesting phenomenon gives the rise to a fundamental question: is the current variance-adaptive rate of optimization error near-optimal? To answer this question, we need the lower bound analysis on the optimization error of CSO.

In this paper, we mainly consider the CSO under a smoothness condition slightly generalized from the strongly convex CSO and leave the analysis when the smoothness condition does not hold to another work. Some lower bounds have been established in the literature under certain smoothness conditions on the objective. For example, Shamir (2013) and Akhavan et al. (2020) considered the strongly convex CSO and found that the lower bound is $\Omega(n^{-1/2})$ and may be improved further if the objective has the higher order smoothness. Locatelli and Carpentier (2018) established the lower bound of non-convex CSO with respect to the budget n. These lower bounds did not analyze the dependency with variance σ^2. Singh (2021) characterized the objective's smoothness in the global domain by a very general function space and provided a variance-adaptive lower bound whose rate matches that of the simple pure exploration algorithm. For example, when the objective is d-dimensional Lipschitz continuous, their lower bound is $\Omega(\max\{(\sigma^2/n)^{-1/(d+2)}, n^{-1/d}\})$, much slower than the above-mentioned polynomial rate $\tilde{O}((\sigma^2/n)^{1/2})$ and near-exponential rate $O(\exp(-cn/(\log n)^2))$. Thus, the result of Singh (2021) cannot answer the research question considered in this paper.

This paper will establish a variance-adaptive lower bound for the optimization error. Together with the conclusion on the algorithms' optimization error in the literature, we confirm that for noisy CSO, the optimal order of optimization error is indeed exponential if the variance is small and the budget is not very large and will be the polynomial rate when the budget is large enough. As a special case, the optimal order for noiseless CSO is exponential.

2 Problem Formulation

We consider the following CSO problem

$$\min_{\mathbf{x}\in\mathcal{X}} y(\mathbf{x})$$

where $\mathbf{x} = (x_1, \ldots, x_d)$ is the vector of decision variables, $\mathcal{X}$ is a d-dimensional hypercube of feasible solutions for $\mathbf{x}$, and $y(\mathbf{x})$ is the objective. The assumptions on the CSO problem are below.

Assumption 1. *The optimal solution to* $\min_{\mathbf{x}\in\mathcal{X}} y(\mathbf{x})$ *is unique.*

Denote the optimal solution by $\mathbf{x}^* \triangleq \arg\min_{\mathbf{x}\in\mathcal{X}} y(\mathbf{x})$. We use the following assumption to characterize the smoothness of $y(\mathbf{x})$.

Assumption 2. *There exists a smoothness parameter* α *with* $0 < \alpha < \infty$ *such that*

$$\tilde{M}\|\mathbf{x} - \mathbf{x}^*\|_\infty^\alpha \leq |y(\mathbf{x}) - y(\mathbf{x}^*)| \leq M\|\mathbf{x} - \mathbf{x}^*\|_\infty^\alpha,\ \forall \mathbf{x} \in \mathcal{X}, \quad (1)$$

with strictly positive constants M *and* $\tilde{M}$ *chosen appropriately.*

The function smoothness in Assumption 2 is defined with respect to the difference $|y(\mathbf{x}) - y(\mathbf{x}^*)|$. Under this assumption, as $\mathbf{x}$ gets closer to $\mathbf{x}^*$, the objective value $y(\mathbf{x})$ will approach the optimal value $y(\mathbf{x}^*)$ due to the upper bound in (1), but $y(\mathbf{x})$ will not be very close to the optimal value due to the lower bound in (1).

The value of parameters α, M and $\tilde{M}$ in Assumption 2 can be arbitrary as long as they are fixed. Many problems including the strongly convex CSO satisfy Assumption 2. Furthermore, the function that is strongly convex around $\mathbf{x}^*$ locally but non-convex in the global domain $\mathcal{X}$ may satisfy Assumption 2, as long as the value of $\tilde{M}$ is small enough to form the lower bound outside the local area of $\mathbf{x}^*$. In addition, Assumption 2 allows $y(\mathbf{x})$ to be discontinuous at $\mathbf{x} \neq \mathbf{x}^*$ only if (1) holds. To find the optimal solution $\mathbf{x}^*$, we can choose to obtain observations at any solution $\mathbf{x}$ in $\mathcal{X}$. The observations of $y(\mathbf{x})$ satisfy the following assumption.

Assumption 3. *Given* $\mathbf{x}$*, the observation* $Y(\mathbf{x})$ *is independent of everything else and* $Y(\mathbf{x}) - y(\mathbf{x})$ *is* σ*-subgaussian with* $\sigma \geq 0$*.*

The independence assumption of Assumption 3 is commonly assumed in the literature. The σ-subgaussian assumption of Assumption 3 permits the distribution of noise $\varepsilon(\mathbf{x}) \triangleq Y(\mathbf{x}) - y(\mathbf{x})$ to be normal, which is also common in the literature. In the special case where $\sigma = 0$, the observation is deterministic.

The CSO algorithm with budget n can decide which solutions to sample (denoted by $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$) and obtain n observations (denoted by $Y(\mathbf{x}_1), Y(\mathbf{x}_2), \ldots, Y(\mathbf{x}_n)$) in total. The algorithm outputs the estimated optimal solution (denoted by $\hat{\mathbf{x}}_n^*$), which is often stochastic due to the randomness of observations. We use the following optimization error as the efficiency measure to quantify how good the estimated optimal solution is:

$$\mathbb{E}[y(\hat{\mathbf{x}}_n^*) - y(\mathbf{x}^*)].$$

An efficient CSO algorithm should have its optimization error as small as possible. We make the following assumption on the algorithm.

Assumption 4. *The algorithm with total budget* n *satisfies the following conditions.*

1. *For* $t = 0$*, the solution* $\mathbf{x}_1$ *is a function of* U_1 *where* U_1 *is an independent random variable.*
2. *For* $t = 1, 2, \ldots, n-1$*, the solution* $\mathbf{x}_{t+1}$ *to be sampled at iteration* $t+1$ *is a function of the historical sampling information* $\mathbf{x}_1, Y(\mathbf{x}_1), \ldots, \mathbf{x}_t, Y(\mathbf{x}_t)$ *and an independent random variable* U_{t+1}*.*
3. *The estimated optimal solution* $\hat{\mathbf{x}}_n^*$ *is a function of the historical sampling information* $\mathbf{x}_1, Y(\mathbf{x}_1), \ldots, \mathbf{x}_n, Y(\mathbf{x}_n)$ *and an independent random variable* U_n^**.*
4. $\hat{\mathbf{x}}_n^* \in \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$*.*

Assumption 4 prescribes how a CSO algorithm decides the solutions for sampling and the estimated optimal solution. Different algorithms correspond to different functions in Assumption 4.1-4.3. If the sampling method or the selection of $\hat{\mathbf{x}}_n^*$ is deterministic given the historical sampling information, then the corresponding function in Assumption 4 is invariant of the random variable U_{t+1} or U_n^*. Many popular algorithms satisfy this assumption, including the tree search algorithms (Munos, 2014), Bayesian optimization (Hong and Zhang, 2021), and random search methods (Fan et al., 2025).

Furthermore, we require in Assumption 4.4 that the estimated optimal solution is one sampled solution. This is just a technical treatment. For an algorithm that may output a solution not sampled before, we may construct a new algorithm with budget $n+1$ whose first n sampled solutions are same as the original algorithm and $(n+1)$-th sampled solution is the original algorithm's estimated optimal solution and analyze the performance of the new algorithm.

3 Main Results

Let $\mathbb{E}_{\mathcal{CSO}}$ denote the expectation taken with respect to the distribution of $\hat{\mathbf{x}}_n^*$ of a given algorithm when solving the CSO problem $\mathcal{CSO}$. The lower bound on the optimization error is below.

Theorem 1. *Let $\mathcal{C}$ denote the set of CSO problems under Assumptions 1-3. For any algorithm satisfying Assumption 4 and n large enough such that $n \geq \max\{b_1\sigma^2, b_2\}$ where b_1 and b_2 are properly defined constants, we have*

$$\sup_{\mathcal{CSO}\in\mathcal{C}} \mathbb{E}_{\mathcal{CSO}}[y(\hat{\mathbf{x}}_n^*) - y(\mathbf{x}^*)] \geq b_3 \max\{(\sigma^2/n)^{\frac{1}{2}}, K^{-b_4 n}\}.$$

where b_3 and b_4 are properly defined constants.

Theorem 1 offers the worst-case (also called minimax) lower bound about the largest optimization error among all CSO problems under Assumptions 1-3. For a specific CSO, the optimization error could still be smaller than the worst-case lower bound. The lower bound is in the form of the maximum of a variance-dependent polynomial term $(\sigma^2/n)^{\frac{1}{2}}$ and a variance-independent exponential term $K^{-b_4 n}$. Which of the two terms is dominating depends on the magnitude of σ^2. If σ^2 is small enough, then the exponential term is the leading one; otherwise, the polynomial term is dominating. However, as long as $\sigma^2 > 0$, when budget n is very large, the polynomial term will always be the dominant term because it decreases to zero relatively slowly as n increases.

For strongly convex CSO with $\alpha = 2$, Theorem 1 implies that depending on the value of σ^2, the fastest rate of optimization error could be either exponential or polynomial and cannot be improved further. It also explains the phenomenon in Figure 1 that given the budget n, the optimization error could be the (nearly) exponential rate when σ^2 is small and n is not very large.

To establish the worst-case lower bounds, we need to construct some CSO problems under Assumptions 1-3 and show that there exists at least one construct CSO problem for whom the optimization error of a given CSO algorithm is greater than the lower bound. The proof relies on three lemmas that connect the optimization result of any two CSO problems. Let $\tilde{\mathcal{X}}$ denote a sub-region of $\mathcal{X}$. The first lemma provides an upper bound on the probability of $\{\hat{\mathbf{x}}_n^* \in \tilde{\mathcal{X}}\}$ for a CSO by that of another CSO and the expected log-likelihood ratio of observations. We can utilize the first lemma to show the variance-dependent lower bound.

The second lemma considers the distribution of the first time t when the sampled solution of an algorithm is in the sub-region $\tilde{\mathcal{X}}$. It shows that the distribution remains the same for two different CSO problems if their observations have the same distribution outside $\tilde{\mathcal{X}}$. The third lemma shows that if there are $2n$ mutually exclusive sub-regions, then there must exist a sub-region such that it does not receive any observation with probability greater than $1/2$ for any algorithm with budget n. We can use the two lemmas to provide the variance-independent lower bound, which is tighter than the variance-dependent lower bound when σ^2 is very small. The maximum of the variance-dependent and variance-independent lower bound is the final lower bound in Theorem 1.

In summary, we provide a variance-adaptive lower bound for CSO under the smoothness condition slightly generalized from the strong convexity. The lower bound confirms that the optimization error of noisy CSO can be an exponential rate, same as that of noiseless CSO, when the variance is small and the budget is not very large.

References

Akhavan, Arya, Massimiliano Pontil, Alexandre Tsybakov. 2020. Exploiting higher order smoothness in derivative-free optimization and continuous bandits. *Advances in Neural Information Processing Systems* **33** 9017–9027.

Bartlett, Peter L, Victor Gabillon, Michal Valko. 2019. A simple parameter-free and adaptive approach to optimization under a minimal local smoothness assumption. *Algorithmic Learning Theory*. PMLR, 184–206.

Fan, Wei-Wei, L Jeff Hong, Guang-Xin Jiang, Jun Luo. 2025. Review of large-scale simulation optimization. *Journal of the Operations Research Society of China* **13** 688–722.

Hong, L Jeff, Chenghuai Li, Jun Luo. 2020. Finite-time regret analysis of kiefer-wolfowitz stochastic approximation algorithm and nonparametric multi-product dynamic pricing with unknown demand. *Naval Research Logistics (NRL)* **67**(5) 368–379.

Hong, L Jeff, Barry L Nelson. 2025. Fifty years of stochastic simulation: Where we are and where we need to go. *European Journal of Operational Research* .

Hong, L Jeff, Xiaowei Zhang. 2021. Surrogate-based simulation optimization. *Tutorials in Operations Research: Emerging Optimization Methods and Modeling Techniques with Applications*. INFORMS, 287–311.

Hu, Jiaqiao, Michael C Fu. 2024. On the convergence rate of stochastic approximation for gradient-based stochastic optimization. *Operations Research* .

Locatelli, Andrea, Alexandra Carpentier. 2018. Adaptivity to smoothness in x-armed bandits. *Conference on Learning Theory*. PMLR, 1463–1492.

Munos, Rémi. 2014. From bandits to monte-carlo tree search: The optimistic principle applied to optimization and planning .

Shamir, Ohad. 2013. On the complexity of bandit and derivative-free stochastic convex optimization. *Conference on Learning Theory*. PMLR, 3–24.

Singh, Shashank. 2021. Continuum-armed bandits: A function space perspective. *International Conference on Artificial Intelligence and Statistics*. PMLR, 2620–2628.

Reviewer Feedback

The paper is clearly interesting but may suffer from a huge competition. It gives interesting bounds for the optimization error in simulation optimization with continuous decision variables but it may be hard to follow for the following reasons (I acknowledge that it is not easy due to space constraints):

- There is no link to a longer version with proofs.
- It is is asserted that the added value with respect to the literature is the dependency with variance, but no comparison with the existing bounds is given to illustrate the gain.
- The impact of the dimension d of decision variables is given and discussed at the end of Page 1, but it seems that due to strong convexity it is not the case anymore. The authors could have gone directly into that assumption to save space, or discuss the impact of d in general on the results (if any).
- Figure 1 is not very neat.