Adaptive A/B Tests and Simultaneous Treatment Parameter Optimization

Yuhang Wu*, Zeyu Zheng†, Guangyu Zhang†, Zuohua Zhang‡, Chu Wang†

* Department of Industrial Engineering and Operations Research, University of California, Berkeley, CA
† Amazon.com Inc, Seattle, Washington

A/B tests, also known as online randomized controlled experiments, are massively used by data-driven enterprises to estimate the expected effect difference between a treatment plan (B) and a control plan (A). Building a valid central limit theorem (CLT) for A/B tests estimators is a crucial yet standard task, in order to provide valid statistical inference such as constructing asymptotically valid confidence intervals. We find that this task of building valid CLT encounters challenges in emerging applications for data-driven enterprises, where the treatment plan is not a single plan, but instead encompasses an infinite continuum of plans indexed by a continuous treatment parameter. As such, the experimenter not only needs to provide valid statistical inference, but also desires to effectively and adaptively find the optimal choice of value for the treatment parameter to use for the treatment plan. However, we find that classical stochastic optimization algorithms, despite of their fast convergence rates under strong convexity assumptions, may not come with a CLT with a vanishing bias. This non-vanishing bias can harm the construction of asymptotically valid confidence intervals. We fix this issue by providing a new stochastic optimization algorithm that on one hand maintains the same fast convergence rate and on the other hand permits the establishment of a valid CLT with vanishing bias. We discuss practical implementations of the proposed algorithm and conduct numerical experiments to illustrate the theoretical findings.

1. Introduction

A/B tests, also referred to as A/B testing or online controlled experiments, play an important role in data-driven enterprises and online platforms to test innovative ideas, evaluate an upgrade or update in products, and support decisions; see Kohavi et al. (2020) for a comprehensive review of A/B tests. Conventionally, the term “A” refers to the control plan, which often describes the current system in use. The term “B” refers to the treatment plan, which often describes a new system that has not been tested online yet. For online platforms, the procedure of an A/B test typically involves an experiment phase and an inference phase. The experiment phase involves randomly assigning a portion of the sequentially arriving samples (customers) to the control plan and the rest to the treatment plan. The inference phase includes deriving an estimator for the average treatment effect (the difference between the expected treatment outcome and the expected control outcome), and establishing a central limit theorem to characterize the asymptotic distribution of the estimator. Gupta et al. (2019) document that A/B tests are massively used at Airbnb, Amazon,
Booking.com, eBay, Facebook, Google, LinkedIn, Lyft, Microsoft, Netflix, Twitter, Uber, among other companies; see also Tang et al. (2010), Kohavi et al. (2013), Johari et al. (2017).

For a range of A/B tests implemented in practice, the treatment plan is not a single plan, but is instead an infinite continuum of plans indexed by a continuous parameter. The different value of such parameter can lead to different expected outcome of the treatment plan. For example, when the treatment plan is a newly designed system that selects what advertisement (ad) to display to a customer, such system may involve a “relevance” threshold parameter $\theta$. That is, the system chooses to block ads that have a relevance score lower than the threshold parameter as part of the ad selection-and-display process, even if those irrelevant ads may generate decent revenue. When the parameter $\theta$ takes continuous value, the treatment plan itself effectively has a continuum infinite number of variants, labeled by each choice of value for $\theta$. We call such parameter that may affect the treatment plan’s expected outcome as treatment parameter.

We consider in this work the problem of adaptive A/B tests with simultaneous treatment parameter optimization. That is, in addition to the standard task of A/B tests to statistically assert whether the treatment plan is better than the control, there is a simultaneous additional task of recommending an optimal choice of value for treatment parameter by the end of experiment. The term “adaptive” means that the experimenter has the ability to sequentially decide, based on adaptively collected information, what choice of value of the treatment parameter to use for the next sample (e.g., customer) that arrives at the experiment.

Technically, the simultaneous additional task of recommending an optimal choice of value for treatment parameter can be formulated as a zero-th order stochastic optimization problem with continuous variable. The term “zero-th order” means that only noisy function value evaluations are available, but no gradient information of the function value is directly available. (One can always choose to use the finite difference method to “indirectly” construct estimators or approximations of the function gradient based on the noisy function value evaluations.) Our considered adaptive A/B tests applications correspond to this zeroth-order indication, because an experimenter can only observe a single random outcome from giving a customer a treatment plan with a certain treatment parameter. For example, such random outcome can be whether the customer clicks on the ad or not. No gradient information or other information is directly available.

Our main results can be summarized as follows.

1. We argue that even though classical zeroth-order simulation optimization or stochastic optimization algorithms can be used to asymptotically find the optimizer at fast rates under a strongly convex assumption on the objective function, they do not come up with a central limit theorem that can be used for valid statistical inference due to a non-vanishing bias.
2. We fix this gap by providing a new optimization algorithm that on one hand maintains the same fast convergence rate and on the other hand permits the establishment of a valid central limit theorem with vanishing bias that can be used to construct asymptotically valid confidence intervals. The new algorithm strategically draws four noisy function value to construct gradient and estimate the objective function value, instead of the classical finite difference gradient estimator based on two draws of noisy function value.

3. We prove under a strong convexity assumption that our proposed algorithm enjoys a convergence rate of $O(T^{-\frac{4}{3}+\varepsilon})$ for any arbitrary small $\varepsilon > 0$ on the estimator for the optimizer, and simultaneously enjoys an almost optimal central limit theorem (CLT) for the estimated optimal objective function value, a CLT that is almost as good as if the optimal treatment parameter were known in advance. We conduct numerical experiments to illustrate the theoretical findings.

1.1. Connections to Literature

Our work is closely related to work on simulation optimization with continuous variable. There is a rich literature on this topic, including L'Ecuyer and Glynn (1994), Hu et al. (2008), Fu et al. (2008), Andradóttir and Prudius (2010), Pasupathy and Ghosh (2013), Hashemi et al. (2014), Zhou et al. (2014), Zhou and Bhatnagar (2017), Fan and Hu (2018), Kiatsupaibul et al. (2018), Hunter et al. (2019), Zhang and Hu (2021), Hong and Zhang (2021). The setting considered by our work, motivated by the A/B tests applications, naturally forms a simulation optimization problem where only noisy function value evaluation is available. In our setting, each sample comes not from a computer simulation but from feeding one of the plans to a customer. Therefore, in our setting, it is not possible to use, for example, infinitesimal perturbation analysis to conveniently obtain a gradient estimator. This is because the outcome observed from each customer can be viewed as an outcome observed from a black-box noisy function oracle. There is not much we can do to look into the black box and obtain more information. Another feature of our setting is that it is difficult, if not impossible, to apply common random numbers. Typically one way to use common random numbers is to keep all randomness the same whereas adjusting the choice of the decision variable to obtain random function output. However, due to the nature of A/B test with customers and causality requirement, we are generally not able to keep all randomness. With customers, using common random numbers is like creating parallel worlds and applying different plans to the same customer in different parallel worlds to observe their outcomes. Creating such “parallel worlds” would be easier on computer simulation compared to on customers.

What is discussed above may be summarized as that: the simulation/stochastic optimization problem in the A/B tests setting is one of the more restrictive cases of the general zeroth-order simulation/stochastic optimization problems, with only noisy function value evaluations available
and without the access to using common random numbers. When constructing a gradient estimator using noisy function value evaluations, one is not able is use common random numbers for these noisy function value evaluations to reduce the variance the gradient estimator. This fact leads the algorithm design in our work to be different from the literature; we use four noisy function value draws to construct the finite-difference gradient estimator, instead of the classical two noisy function value draws. The necessity, as we will discuss in the work, is that the classical two noisy function value draws do not lead to valid statistical inference in the A/B test setting when no common randomness can be used.

Our work is also connected to the stochastic approximation literature; see Robbins and Monro (1951), Kiefer and Wolfowitz (1952), Spall (1992), L’Ecuyer and Yin (1998), Kushner and Yin (2003), Kim et al. (2015), He et al. (2021), Hu et al. (2022) among others. Stochastic approximation problems generally aim to adaptively find the root of an equation in some region of the variable. Stochastic approximation and simulation/stochastic optimization are tightly connected. One way to connect is that some simulation/optimization problems can be formulated into finding the root of the equation that equals the gradient to zero. Our work is especially connected to L’Ecuyer and Yin (1998), who consider constructing central limit theorems for stochastic approximation problems with a given budget. Even though the setting is different, our central limit theorem proof is partially inspired by L’Ecuyer and Yin (1998).

Our work is connected to the literature of stochastic zeroth-order optimization; see Ghadimi and Lan (2013), Duchi et al. (2015), Nesterov and Spokoiny (2017), Liu et al. (2018), Balasubramanian and Ghadimi (2018), Ajallooeian and Stich (2020) among others. The commonality in our work is that we also need to use zeroth-order noisy function evaluations to construct gradient estimator and do gradient search for optimizing the treatment parameter. Our work additionally provides statistical inference results (because they are needed to assess whether the treatment plan is significantly better than the control under recommended value for the treatment parameter) in addition to the optimization results. Our work is also connected to Xie et al. (2016), Wu et al. (2017), Letham et al. (2019), Wang et al. (2020), Balandat et al. (2020); they consider the problem of adaptively optimizing a continuous parameter of the treatment plan, under a Bayesian framework. Our work is different in three folds. Our algorithm takes a frequentist view, constructs gradient information using noisy function value observations, and does gradient search, which is different from the Bayesian approach. Our work develops a central limit theorem (CLT) that can be directly used for large-sample statistical inference, while the above mentioned work does not emphasize on CLT.

**Organization.** In Section 2, we describe the problem setting for *adaptive A/B tests and simultaneous treatment parameter optimization*. Section 3 provides a new algorithm design. In Section
4, we provide theoretical guarantees and central limit theorem for the proposed algorithm. Section 5 presents numerical experiments. Section 6 concludes.

2. Problem Setting

2.1. Adaptive A/B tests

We first describe the problem setting of a simple adaptive A/B test with one treatment plan and one control plan, without treatment parameter optimization. We then state the problem setting for adaptive A/B tests and simultaneous treatment parameter optimization in Section 2.2.

To put the discussion in context, we presume the control plan to be the currently in-use system that decides what type of advertisement (ad) to display for a customer; whereas the treatment plan is a newly designed system that decides what type of advertisement (ad) to display for a customer. The term “adaptive” means that (i) the experimenter have timely observations for the outcomes of treatment/control and (ii) the experimenter can adaptively adjust the assignment of treatment/control. Such adaptivity is generally feasible for online platforms.

To elaborate, the experimenter sets a sampling budget $N$, where $N$ is a positive integer. Customers arrive independently and sequentially. When the $i$-th customer arrives, the experimenter chooses either the treatment plan or the control plan to assign to the customer, through the use of some pre-specified randomization scheme. In the adaptive settings, the experimenter can immediately observe the outcome of the treatment/control, for example, whether the customer clicks on a displayed ad or not. Denote the data collected in the experiment by $\{I_i, y_i\}_{i=1,2,\ldots,N}$, where $I_i$ takes value 1 (or 0) representing the $i$-th customer receiving the treatment (or control), and $y_i$ denotes the outcome observed for the $i$-th customer. When deciding whether to assign the treatment plan or the control plan to the $(i+1)$-th customer, the experimenter has access to the observations of $\{I_j, y_j\}_{j=1,2,\ldots,i}$.

Once the experiment data $\{I_i, y_i\}_{i=1,2,\ldots,N}$ are collected, a standard estimator the average treatment effect (ATE) (i.e., the difference between the treatment’s expected outcome and the control’s expected outcome) is given by

$$\frac{\sum_{i=1}^{N} 1(I_i = 1) \cdot y_i}{\sum_{i=1}^{N} 1(I_i = 1)} - \frac{\sum_{i=1}^{N} 1(I_i = 0) \cdot y_i}{\sum_{i=1}^{N} 1(I_i = 0)},$$

where the notation $1(\cdot)$ represents the indicator function. A core statistical inference task is to build a valid central limit theorem (CLT) and understand the asymptotic distribution for the ATE estimator when the sample size $N$ is large.

2.2. Adaptive A/B tests with simultaneous treatment parameter optimization

This subsection describes the problem setting for adaptive A/B tests and simultaneous treatment parameter optimization. In some practical scenarios, inside the treatment plan is a tunable parameter. When the parameter $\theta$ takes continuous value, the treatment plan itself effectively has a
continuum infinite number of variants, labeled by each choice of value for the treatment parameter $\theta$.

Different from the standard adaptive A/B tests, the experimenter now has two simultaneous goals, which can roughly be described as follows. With the experiment that involves $N$ customers, the experimenter hopes to test whether the treatment plan is better than the control plan; the experimenter also hopes to find a good, or in some sense optimal, choice of value for the tunable treatment parameter $\theta$.

Suppose that the treatment parameter $\theta$ takes value in $[0, 1]^d$ for the ease of analysis, where $d$ is a positive integer. A model that characterizes how the treatment parameter affects the treatment performance is given by

$$Y(\theta) = \mu(\theta) + \sigma(\theta) \cdot \epsilon_\theta, \quad \theta \in \Theta = [0, 1]^d,$$

where $\mu(\theta)$ is the expected treatment outcome given the treatment parameter $\theta$, the term $(\sigma(\theta))^2$ is the variance of the treatment outcome, and $\epsilon_\theta$ is a random variable with mean zero and variance one. For different choices of value for $\theta$, the probability distribution of $\epsilon_\theta$ is allowed to be different. Here is how the model is used to explain the observed experiment data $\{I_i, y_i\}_{i=1,2,\ldots,N}$. For the $i$-th customer, if she is assigned to treatment with treatment parameter $\theta_i$, then the observed outcome $y_i$ is presumed to be an independent realization of the random variable $Y(\theta_i)$.

On the control side (which is a single plan), if the $i$-th customer is assigned to control, then the observed outcome $y_i$ is presumed to be an independent realization of the random variable $\tilde{Y} = \tilde{\mu} + \tilde{\sigma} \cdot \tilde{\epsilon}$, where $\tilde{\mu}$ is the expected control outcome, the term $\tilde{\sigma}^2$ is the variance of the control outcome, and $\tilde{\epsilon}$ is a random variable with mean zero and variance one.

**The experimenter's feasible actions: adaptive treatment/control assignment.** When the $(i+1)$-th customer arrives, the experimenter has access to the treatment/control assignment and outcome information for all previous $i$ customers. The experimental then decides whether to assign the $(i+1)$-th customer to control ($I_{i+1} = 0$) or treatment ($I_{i+1} = 1$); if the decision is treatment, the experimenter chooses a treatment parameter $\theta_{i+1}$. Once the $(i+1)$-th customer is assigned with treatment or control, an outcome $y_{i+1}$ is observed immediately. Let $\mathcal{F}_{i+1} = \{I_j, y_j, \theta_j\}_{j=1,2,\ldots,i+1}$. For all the $j$'s that $I_j = 0$, the value of $\theta_j$ can be set as an arbitrary value.

**The experimenter's tasks.** The experimenter has in mind, before the experiment begins, that she will need to deliver the following tasks after the experiment (that involves $N$ samples) end: (i) Make a recommendation $\hat{\theta}$ as the choice of treatment parameter to use, in case the treatment plan is decided to be used for future customers. (ii) Provide an estimator for $\mu(\hat{\theta}) - \tilde{\mu}$ and prove a central limit theorem (CLT) for the estimator when $N$ is large. The asymptotic distribution in the CLT is then used to construct valid confidence interval and hypothesis test procedures.
3. Aim and Algorithm

In this section, we propose an algorithm that the experimenter can use to the two tasks described in the previous section. To focus the discussion on how an algorithm can adaptively search and find a good choice of treatment parameter, we presume that the experimenter sets upfront that out of the $N$ samples, $qN$ samples are assigned to treatment and $(1 - q)N$ are assigned to control, where $q \in (0, 1)$; we omit the operation of taking integer and view $qN$ as an integer. This presumption for $q$ being constant is made for algorithm’s simplicity. Our focus is on providing algorithm and inference for optimizing the treatment parameter and therefore the sampling decision is on how to wisely spend the $qN$ samples relatively within the treatment parameter choices. In contrast, the choice of $q$ does not affect whether the samples are optimally assigned with the treatment parameter choices, and therefore can be studied independently and orthogonally.

When the optimal $\theta^*$ is not known a priori, there exists an implicit trade-off between optimization and statistical inference. On the one hand, we need to find a good choice of $\theta$, and this goal can generally be achieved by standard zeroth-order simulation/stochastic optimization algorithms. On the other hand, standard algorithms under convexity assumptions largely lead to a good estimator $\hat{\theta}$, but may not yield sufficient representative samples to provide the best estimator for the optimal objective value $\mu(\hat{\theta})$, which is crucially needed for statistical inference.

At the first glance, a straightforward idea to deal with this implicit trade-off is a “search-then-evaluate” strategy, that is, we first use part (say half) of samples to search for a good estimator $\hat{\theta}$ for the optimizer $\theta^*$, then we fix this $\hat{\theta}$ and use the remaining samples to draw noisy observations of $\mu(\hat{\theta})$. This idea works in the sense that it indeed recommends a $\hat{\theta}$ that converges to optimizer and also provides a CLT centered at $\mu(\hat{\theta})$. However, it can suffer from inefficiency and slower rates in both the searching part and evaluation part. As for the searching procedure, since we only use part of all samples, the estimator for the optimizer may not be good enough compared with those algorithms searching with all samples. For the evaluation part, using part of samples instead of all samples can give a larger than optimal variance in the CLT, which then leads to low statistical power for testing and inference. Thus, we accordingly seek for better remedies.

3.1. Our algorithm

In this subsection, we propose an algorithm that adaptively decides how to use the $T = qN$ samples to search over different choices of value for the treatment parameter $\theta$, with the goal to fulfill the dual optimization task and statistical inference task. Our algorithm, as will be proved later, enjoy the following properties:

(i). It uses all $T$ samples to do stochastic optimization and has a fast convergence rate as $\mathbb{E}\|\hat{\theta} - \theta^*\| = O(T^{-\frac{3}{4} + \epsilon})$ for arbitrarily small $\epsilon > 0$. 
(ii). For our recommendation $\hat{\theta}$, it uses all $T$ samples to construct an estimator of $\mu(\hat{\theta})$ and does not waste samples.

(iii.) A CLT holds for our estimator of $\mu(\hat{\theta})$ and is almost as good as the CLT where hypothetically $\theta^*$ was known and we sample $Y(\theta^*)$ for $T$ times to perform inference for $\mu(\theta^*)$.

The flow of our algorithm is as follows. Suppose we have in total $T$ samples, and we begin with an initial point $\theta^0$. In $t$-th iteration, to find a descending direction, we randomly sample a vector $w^t$ uniformly from a unit sphere $B(0, 1)$, and define following four points:

$$
\theta^+ = \theta^t + c_tw^t, \quad \theta^- = \theta^t - c_tw^t,
$$

$$
\theta^{++} = \theta^t + k_c w^t, \quad \theta^{--} = \theta^t - k_c w^t.
$$

We omit the projection operator back into the feasible region $\Theta$ if some of the four points exceed the boundary of $\Theta$. Here, $k > 1$ is an algorithm parameter, and we will discuss how to principally choose this parameter later. The parameters $\{c_t\}_{t=1}^T$ are half lengths of finite difference which will also be determined later.

Note that there is no first-order oracle to draw gradient estimator directly. In order to estimate the gradient through some numerical scheme, we will first estimate $\mu(\theta^+), \mu(\theta^-), \mu(\theta^{++}), \mu(\theta^{--})$. For some predetermined $m_1, m_2 \in \mathbb{N}_+$, define $m = m_1 + m_2$, and without loss of generality assume that $\frac{T}{2m} \in \mathbb{N}$. We randomly generate an $m_1$-element subset $R_t^{(1)}$ from $\{1, \cdots, 2m_1\}$. Note that the randomization design will be predetermined to avoid confounding, and a simple example for the randomization design could be that, we randomly choose an $m_1$-element subset from all $m_1$-element subsets of $\{1, \cdots, 2m_1\}$ with equal probability. Then, for $i = 1, \cdots, 2m_1$, we sample $y_{2mt+i}$ as a realization of $Y(\theta^+)$ if $i \in R_t^{(1)}$, and sample $y_{2mt+i}$ as a realization of $Y(\theta^-)$ if $i \not\in R_t^{(1)}$. This can be equivalently written as:

Sample $y_{2mt+i}$ as a realization of $I_{\{i \in R_t^{(1)}\}} Y(\theta^+) + (1 - I_{\{i \in R_t^{(1)}\}}) Y(\theta^-)$.

Here $I_{\{i \in R_t^{(1)}\}}$ is an indicator function that takes 1 if $i \in R_t^{(1)}$ and takes 0 otherwise. Similarly, we can also generate an $m_2$-element subset $R_t^{(2)}$ from $\{1, \cdots, 2m_2\}$ and sample $y_{2mt+2m_1+i}$ as a realization of $I_{\{i \in R_t^{(2)}\}} Y(\theta^{++}) + (1 - I_{\{i \in R_t^{(2)}\}}) Y(\theta^{--})$ for $i = 1, \cdots, 2m_2$. We estimate $\mu(\theta^+), \mu(\theta^-), \mu(\theta^{++}), \mu(\theta^{--})$ through sample mean as follows:

$$
\bar{\mu}(\theta^+) = \sum_{i \in R_t^{(1)}} \frac{y_{2mt+i}}{m_1},
$$

$$
\bar{\mu}(\theta^-) = \sum_{i \in \{1, \cdots, 2m_1\} \setminus R_t^{(1)}} \frac{y_{2mt+i}}{m_1},
$$

$$
\bar{\mu}(\theta^{++}) = \sum_{i \in R_t^{(2)}} \frac{y_{2mt+2m_1+i}}{m_2},
$$

$$
\bar{\mu}(\theta^{--}) = \sum_{i \in \{1, \cdots, 2m_2\} \setminus R_t^{(2)}} \frac{y_{2mt+2m_1+i}}{m_2}.
$$
Next, the gradient at $\theta^t$ is estimated by:
\[
g^t = \frac{-\bar{\mu}(\theta^{++}) + k^3 \bar{\mu}(\theta^+) - k^3 \bar{\mu}(\theta^-) + \bar{\mu}(\theta^{--})}{2k(k^2 - 1)c_t} w^t, \tag{6}
\]
and an estimator of $\mu(\theta^t)$ is constructed by:
\[
\hat{\mu}_t = \frac{-\bar{\mu}(\theta^{++}) + k^2 \bar{\mu}(\theta^+) + k^2 \bar{\mu}(\theta^-) - \bar{\mu}(\theta^{--})}{2(k^2 - 1)}. \tag{7}
\]
Finally, we update
\[
\theta^{t+1} \leftarrow \theta^t + \alpha_t g^t
\]
for some step size $\alpha_t$ to finish the $t$-th iteration. Here we use the ascent “+” since we are maximizing the objective function. The complete algorithm is provided in Algorithm 1.

Here the gradient estimator (6) and function value estimator (7) may seem not intuitive at first glance. They use four points instead of two points in standard finite difference methods. Also their specific forms read not intuitive. We provide some analysis in the follows. For (7), if we consider $\mu(\cdot)$ to be third-order differentiable, we have for small $h > 0$ that
\[
\mu(\theta + h) + \mu(\theta - h) = 2\mu(\theta) + h^T \nabla^2 \mu(\theta)h + O(||h||^3_2),
\]
then plug in and we obtain:
\[
\mu(\theta^+) + \mu(\theta^-) = 2\mu(\theta) + c_t^2 (w^t)^T \nabla^2 w^t + O(c_t^3),
\]
and
\[
\mu(\theta^{++}) + \mu(\theta^{--}) = 2\mu(\theta) + k^2 c_t^2 (w^t)^T \nabla^2 w^t + O(c_t^3).
\]
From above two equations we have:
\[
\mu(\theta^t) = \frac{-\mu(\theta^{++}) + k^3 \mu(\theta^+) - k^3 \mu(\theta^-) + \mu(\theta^{--})}{2(k^2 - 1)c_t} + O(c_t^3),
\]
which implies that (7) gives an estimator of the function value with third order error. Thanks to this error order, the “not intuitive” construction for this estimator can achieve a high-enough accuracy to later facilitate the establishment of a CLT with vanishing bias.

As for the gradient estimator construction in (6), it uses four points instead of two points to construct the finite difference gradient estimator. We can show that the four-point construction (6) gives a better estimator compared to the standard two-point construction when stronger smoothness condition holds, so it may behave better in practice. In fact, if $\mu(\cdot)$ is fourth-order differentiable, then by fourth-order Taylor expansion we can see that
\[
\nabla \mu(\theta^t)^T w^t = \frac{-\bar{\mu}(\theta^{++}) + k^3 \bar{\mu}(\theta^+) - k^3 \bar{\mu}(\theta^-) + \bar{\mu}(\theta^{--})}{2k(k^2 - 1)c_t} + O(c_t^4),
\]
which gives an estimator with fourth order error.
Algorithm 1 Zeroth-order optimization for A/B tests

1: **Input:** initial point $\theta^0 \in [0,1]^d$, total samples $T$, parameters $k > 1$, $\beta \in (0,1)$, step sizes $\{\alpha_t\}$, subsample sizes $m_1, m_2$, and half lengths of finite difference interval $\{c_t\}$.

2: **for** $t = 0, 1, 2, \ldots, \frac{T}{2m} - 1$ do

3: Sample $w^t$ uniformly from a unit sphere $B(0,1)$.

4: Calculate $\theta^+, \theta^-, \theta^{++}, \theta^{--}$ through (2) and (3).

5: Generate an $m_1$-element subset $R_t^{(1)}$ from $\{1, \ldots, 2m_1\}$ through some predetermined randomization design.

6: **for** $i = 1, \ldots, 2m_1$ do

7: Sample $y_{2(m_1+m_2)t+i}$ as a realization of $I_{\{i \in R_t^{(1)}\}} Y(\theta^+) + (1 - I_{\{i \in R_t^{(1)}\}}) Y(\theta^-)$;

8: end for

9: Generate an $m_2$-element subset $R_t^{(2)}$ from $\{1, \ldots, 2m_2\}$ through some predetermined randomization design.

10: **for** $i = 1, \ldots, 2m_2$ do

11: Sample $y_{2(m_1+m_2)t+2m_1+i}$ as a realization of $I_{\{i \in R_t^{(2)}\}} Y(\theta^+) + (1 - I_{\{i \in R_t^{(2)}\}}) Y(\theta^-)$;

12: end for

13: Compute $\bar{\mu}(\theta^+), \bar{\mu}(\theta^+), \bar{\mu}(\theta^{++}), \bar{\mu}(\theta^{--})$ through (4), (5), respectively.

14: Compute $g^t$ through (6).

15: Compute $\hat{\mu}_t$ through (7).

16: Update $\theta^{t+1} \leftarrow \theta^t + \alpha_t g^t$.

17: end for

18: **Output:** Let $\hat{\theta}_T$ take value of $\frac{2m}{\beta T} \sum_{i=(1-\beta)T/(2m)+1}^{T/(2m)} \theta^i$, $\hat{\mu}(\hat{\theta}_T) \leftarrow \frac{2m}{\beta T} \sum_{i=1}^{T/(2m)} \hat{\mu}_i$.

3.2. **On the gradient estimator**

Algorithm 1 ends up recommending a choice of treatment parameter given by $\hat{\theta}_T$; the algorithm also provides an estimator $\hat{\mu}(\hat{\theta}_T)$ for the (unknown) expected outcome $\mu(\hat{\theta}_T)$ on the recommended choice of treatment parameter. Before establishing a CLT for $\hat{\mu}(\hat{\theta}_T)$, we provide some discussion on why four points are needed to construct the gradient estimator (6), instead of the classical two-points approach.

Classical two-points approach uses the forward finite-difference (FD) or central FD gradient estimator, which are prevalent in zeroth-order simulation/stochastic optimization literature. However, they may not suffice for the dual tasks of optimization and statistical inference. Although it is generally hard to give a lower bound on the convergence of an optimization algorithm, we provide following heuristic arguments and also some numerical examples in Section 5. For the forward FD estimator and central FD estimator, we have $g^t_{\text{FDF}} = \frac{\bar{\mu}(\theta^+) - \bar{\mu}(\theta^-)}{c_t}$ and $g^t_{\text{FDC}} = \frac{\bar{\mu}(\theta^+) - \bar{\mu}(\theta^-)}{2c_t}$. Suppose
we now replace (6) by $g_{\text{FDC}}^t$, and take $c_t = O(t^{-\nu})$ for some $\nu > 0$, then from the literature (e.g. L’Ecuyer and Yin (1998)) we know that the variance of $g_{\text{FDC}}^t$ is $O(t^{2\nu})$ and $\|\theta^t - \theta^*\|^2_2 = O(t^{-1+2\nu})$. Since $|\mu(\theta^t) - \mu(\theta^*)| = O(\|\theta^t - \theta^*\|^2_2)$, then

$$\sqrt{T}\left|\mathbb{E}\frac{2m}{T} \sum_{t=1}^{T/(2m)} \mu(\theta^t) - \mu(\theta^*)\right| = O(T^{-\frac{1}{2}+2\nu}).$$

In addition, now our estimator for $\mu(\theta^t)$ is $\hat{\mu}_{t,\text{FDC}} = \frac{\bar{\mu}(\theta^+) + \bar{\mu}(\theta^-)}{2}$, and under mild conditions, we can show by Taylor expansion that the bias caused by the FD estimator is $\hat{\mu}_{t,\text{FDC}} - \mu(\theta^t) = O(c_t^2) = O(t^{-2\nu})$. Then

$$\sqrt{T}\left|\mathbb{E}\frac{2m}{T} \sum_{t=1}^{T/(2m)} \hat{\mu}_{t,\text{FDC}} - \mathbb{E}\frac{2m}{T} \sum_{t=1}^{T/(2m)} \mu(\theta^t)\right| = O(T^{-\frac{1}{2}-2\nu}).$$

For valid statistical inference, a CLT with following form with vanishing bias is needed:

$$\sqrt{T}\left(\frac{2m}{T} \sum_{t=1}^{T/(2m)} \hat{\mu}_{t,\text{FDC}} - \mu(\theta^*)\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

We therefore will expect $\frac{1}{2} - 2\nu < 0$ and $-\frac{1}{2} + 2\nu < 0$, which is evidently impossible for the classical two-point gradient estimator based approach. The basic idea of above arguments is that, when we are using the central FD estimator, a choice of small $c_t$ will give large variance and thus lead to slow convergence rate, but a choice of large $c_t$ can lead to a bad estimator of $\hat{\mu}(\theta^t)$. This trade-off makes standard two-point FD estimator very difficult to assist the dual tasks. In fact, we provide numerical examples in Section 5 to show that, even if $\mu(\cdot)$ is simply a quadratic function on $\mathbb{R}$, a valid CLT with vanishing bias will not hold for the central FD estimator.

3.3. On the choice of parameters

We now discuss how to choose algorithm parameters $m_1, m_2$ and $k$. First, in practical A/B tests, especially for online platforms, sometimes the treatment parameter value cannot be adjusted too frequently, so the choice of $m = m_1 + m_2$ cannot be too small. Fortunately, the choice of $m$ will not affect our algorithm too much since we can hedge its effect by tuning the step size, that is, when $m$ is large, then we can increase the step size to balance the effect.

To get an intuition on how to choose $m_1, m_2$ for fixed $k$ when $m$ is given, we begin by considering a simple case that $Y(\theta) \sim \mathcal{N}(0,1)$ for all $\theta$. In this case, by some calculations we have $\text{Var}(\hat{\mu}_t) = \frac{1}{2(k^2-1)}\left(\frac{k^2}{m_1} + \frac{1}{m_2}\right)$. To get a good estimator $\hat{\mu}(\hat{\theta}_T)$, recall that $\hat{\mu}(\hat{\theta}_T) = \frac{2m}{T} \sum_{t=1}^{T/(2m)} \hat{\mu}_t$, we would expect that the variance of $\hat{\mu}_t$ can be minimized. Then by Cauchy inequality we can see that the variance is minimized at $m_1 = \frac{k^2+1}{k^2+1}m$ and $m_2 = \frac{1}{k^2+1}m$, and the minimum value is $\frac{k^2+1}{2(k^2-1)m}$. This argument can be easily adapted to the general case and under some assumptions (will be detailed in the next section) we have:
Proposition 1. Suppose $m, k$ are given, under Assumption 1, 2, 3, 4, if $\alpha_t = C_0 t$ for some $C_0 > 0$ and $c_i = \frac{C_i}{t^i}$ for some $C_i > 0$ and $\frac{1}{6} < \nu < \frac{1}{4}$, we have:

$$T \cdot \text{Var}(\hat{\mu}(\hat{\theta}_T)) \rightarrow m \left( \frac{k^2}{k^2 - 1} - \frac{1}{m_2} \right) \sigma(\theta^*)^2$$

(8)

with probability 1 as $T$ tends to infinity.

Proposition 1 suggests that $m_1$ and $m_2$ shall be chosen to minimize the right hand side of (8). Then from our discussions above, we can take $m_1 = \frac{k^2}{k^2 + 1} m$ and $m_2 = \frac{1}{k^2 + 1} m$ in general case.

As for the choice of $k$, since from Proposition 1 we can see that the minimized variance is $\frac{(k^2+1)^2 \sigma(\theta^*)^2}{T(k^2-1)^2}$, while if we know $\theta^*$ in oracle and sample it $T$ times, then the variance of our sample mean estimator is given by $\frac{\sigma(\theta^*)^2}{T}$, so the ratio of the variances of two estimators is $\frac{(k^2+1)^2}{(k^2-1)^2}$. This ratio is monotonically decreasing when $k$ grows, so it seems that a large $k$ could be better. However, when $k$ is too large, the improvement in the variance is small and the initialization phase of the algorithm can become unstable because it needs to sample $\mu(\cdot)$ at $\theta^t + kc_t w_t$ and $\theta^t - kc_t w_t$, which can be quite far away from $\theta^t$. Under this trade-off, a reasonable choice can be to set $k \in [3, 5]$ since when $k$ is 5, the ratio $\frac{(k^2+1)^2}{(k^2-1)^2} \approx 1.17$ is quite close to 1 and the improvement of variance when $k$ grows is small.

4. Theory

In this section, we state assumptions on the model and provide a central limit theorem for $\hat{\mu}(\hat{\theta}_T)$ give by our algorithm. Here $T = qN$, and $k, m_1, m_2, q$ are all treated as deterministic now. We begin by following four assumptions.

Assumption 1. For $\theta \in \Theta$, $-\mu(\theta)$ is third-order continuously differentiable and $-\nabla^2 \mu(\theta) > 0$.

Assumption 2. For $\theta \in \Theta$, $\sigma(\theta)$ is continuous.

Assumption 3. The ordinary differential equation (ODE) $\dot{\theta} = -\nabla \mu(\theta)$ has a unique solution for each initial condition.

Assumption 4. There exists some $a > 0$ and $M_1 > 0$ such that $E|\epsilon_\theta|^{2+a} < M_1$, for all $\theta \in [0, 1]^d$.

Assumption 1, 2 and 3 are relatively standard in the literature of convex simulation optimization and stochastic approximation. For example, they are analogies to Assumption (A3)–(A5) in L’Ecuyer and Yin (1998). Assumption 1 requires that $-\mu(\cdot)$ is smooth enough for the tractability of zeroth-order gradient estimator and is strongly convex around the optimizer, where the strongly convex assumption is used in many zeroth-order optimization literature that renders fast convergence rate, see for example L’Ecuyer and Yin (1998), Shamir (2013), Nesterov and Spokoiny (2017) and Dvurechensky et al. (2021). Relaxing strongly convexity to ordinary convexity would lead to
a different slower convergence rate and require a different procedure about deriving central limit theorems. We are not sure about how to make this relaxation and hope to leave this to future work. Assumption 2 requires that $\sigma(\cdot)$ is continuous. Assumption 3 requires the ODE to be "well behaved" since this ODE is closely related to the behaviour of $\{\theta^t\}$. Finally, Assumption 4 requires a $(2 + a)$th-order moment exists uniformly for all $\epsilon_\theta$, which is a usually made assumption in the derivation of central limit theorem.

We first need to prove following theorem for the convergence rate of Algorithm 1. The proof is displayed in the Appendix.

**Theorem 1.** Under Assumption 1, 2 and 3, if $\alpha_t = \frac{c_0}{t}$ for some $c_0 > 0$, and $\lim_{t \to \infty} t^c c_t = c_1$ for some $c_1 > 0$ and $\frac{1}{6} < \nu < \frac{1}{4}$, then we have $\theta^t \to \theta^*$ almost surely as $t \to \infty$, and

$$E\|\theta^t - \theta^*\|_2^2 = O(t^{-1+2\nu}).$$

**Remark 1.** Theorem 1 gives the convergence rate for the optimizer recommended by our algorithm. By carefully choosing the step sizes $\{\alpha_t\}$ and parameter $\nu$, the convergence rate is fast enough in the sense that $E\|\theta^t - \theta^*\|_2^2 = o(t^{-\frac{1}{2}})$, so that it does not contribute additional asymptotic variances to the CLT for estimated optimal function value. In addition, the choice of parameters are also crucial in the next theorem.

Based on Theorem 1, our central limit theorem for the optimal objective function value estimator $\hat{\mu}(\hat{\theta}_T)$ can be stated as follows:

**Theorem 2.** Assume all the conditions in Theorem 1 hold, and in addition Assumption 4 is satisfied, then we have:

$$\sqrt{T}(\hat{\mu}(\hat{\theta}_T) - \mu(\hat{\theta}_T)) \xrightarrow{d} \frac{(k^2 + 1)\sigma(\theta^*)}{k^2 - 1} N(0, 1)$$
and

$$\sqrt{T}(\hat{\mu}(\hat{\theta}_T) - \mu^{\gamma}(\theta^*)) \xrightarrow{d} \frac{(k^2 + 1)\sigma(\theta^*)}{k^2 - 1} N(0, 1)$$
as $T \to +\infty$, where

$$\hat{\mu}(\hat{\theta}_T) = \frac{2m}{T} \sum_{i=1}^{T/(2m)} \hat{\mu}_i.$$

**Remark 2.** Theorem 2 provides a CLT with desired order and vanishing bias. Here, since we can show that $E|\mu(\hat{\theta}_T) - \mu(\theta^*)| = O(T^{-\gamma})$ with $\gamma > \frac{1}{2}$, the confidence intervals for $\mu(\hat{\theta}_T)$ and $\mu(\theta^*)$ are asymptotically the same. Note that the asymptotic variance is $\frac{(k^2 + 1)^2\sigma(\theta^*)^2}{k^2 - 1}$ and as we discussed in Section 3.3, this is close to $\sigma(\theta^*)^2$ when $k$ is large, while the latter one is the optimal variance, which means the estimator provided by our algorithm behaves almost as well as the case that $\theta^*$ is known in oracle.
Regarding the variance term that is needed in practical use of CLT, we can construct an estimator \( \hat{\sigma}(\theta^*) \) for the unknown \( \sigma(\theta^*) \) through the sample standard deviation of \( y_I, y_{I+1}, \ldots, y_T \) for some \( I \in \mathbb{N}_+ \), then \( \hat{\sigma}(\theta^*) \) is consistent under some additional conditions (details provided in the supplementary material), and an asymptotic 95\% confidence interval can be constructed as

\[
\left[ \hat{\mu}(\hat{\theta}_T) - \frac{1.96(k^2 + 1)\hat{\sigma}(\hat{\theta}_T)}{\sqrt{T(k^2 - 1)}}, \hat{\mu}(\hat{\theta}_T) + \frac{1.96(k^2 + 1)\hat{\sigma}(\hat{\theta}_T)}{\sqrt{T(k^2 - 1)}} \right].
\] (9)

5. Numerical Experiments

In this section, we give some synthetic examples to illustrate the performance of our algorithm and also show that the central finite difference (FD) method cannot give the desired CLT. Because our goal is illustrate the central limit theorem performance, limiting distribution and asymptotic bias, we need to know the true underlying value to quantify the bias etc. That is why synthetic examples are needed. Real data generally only give us realized sample paths but not underlying true value to illustrate CLT. Yet we can use data to calibrate a synthetic model.

We consider a synthetic example where the treatment parameter (a relevance parameter in advertisement recommendation) \( \theta \in [0, 1] \) is one-dimensional, and \( \mu(\theta) \) is calibrated from a set of A/B test data in a e-commerce company as follows:

\[
\mu(\theta) = -0.02125\theta^2 + 0.01825\theta + 0.0105.
\]

Here \( \mu(\theta) \) denotes the click-through rate (CTR) and \( Y(\theta) \) is a Bernoulli random variable. Then \( \sigma(\theta) = \sqrt{\mu(\theta)(1 - \mu(\theta))} \). We hope to note that our work is not intended to propose an algorithm that outperforms all existing algorithms. Instead, we are aiming at one algorithm that on one hand enjoys desirable theoretical convergence rate and on the other hand well facilities statistical inference through establishing a CLT.

We set \( k = 3, m = 50, \nu = \frac{1}{5}, \theta^0 = \frac{1}{2} \), and \( c_t = \frac{1}{\nu}, \alpha_t = \frac{30}{\nu} \), respectively. We repeat our algorithm 1000 times and report the normalized and centralized estimator. Here, the normalized and centralized estimator is defined as:

\[
\frac{\sqrt{T(k^2 - 1)}}{(k^2 + 1)\sigma(\theta^*)}(\hat{\mu}(\hat{\theta}_T) - \mu(\hat{\theta}_T)),
\]

so it should follow a standard normal distribution according to Theorem 2, which has been justified by our results in Section 5.

If our algorithm is not used but the standard two-point central finite difference (FD) method is used to construct gradient estimator and function value estimator, Figure 1(b), 1(d) and 1(f) show the distribution of the normalized and centralized estimator given by central FD method with sample size \( T = 10^5, 10^6, 10^7 \), respectively. From those results in Figure 1(b), 1(d) and 1(f), we can see that the estimators provided by the standard two-point central FD method have a significant non-zero bias (because we know the true underlying optimal value) and are centered around \(-2\)
instead of 0. Because such bias is generally impossible to accurately estimate without knowing the true optimal value, it is not reliable to use the corresponding estimator to provide valid confidence intervals. For example, we would not know how to accurately decide the center of the confidence interval due to the unknown and non-vanishing bias.

In comparison, Figure 1(a), 1(c) and 1(e) show that the estimators given by our Algorithm 1 centered closely to 0 and are nearly normally distributed. In addition, the coverage rates of the 95% confidence interval (CI) we constructed are 93.5%, 93.9% and 94.2% for $T = 10^5$, $10^6$ and $10^7$, respectively, so valid statistical inference is possible for our estimator.

Next we examine the performance of our algorithm under other outcome distributions. We use the same parameters and the same $\mu$, but this time $Y(\theta)$ is a random variable with Pareto distribution, which is known to be heavy-tailed. The shape parameter of this distribution is chosen to be $\alpha = 3$, and the scale parameter is chosen to be $\frac{2\mu(\theta)}{\alpha}$, so $EY(\theta) = \mu(\theta)$. The results are given by Figure 1(g) and 1(h).
(e) Estimator distribution based on Algorithm 1, $T = 10^7$

(f) Estimator distribution based on two-point central FD method, $T = 10^7$

(g) Estimator distribution based on Algorithm 1, Pareto distribution

(h) Estimator distribution based on two-point central FD method, Pareto distribution

This time, the estimator provided by our algorithm still works well and the coverage rate of 95% CI is 94.5%, while the central FD method gives an estimator with bias as large as $-8$ and completely fails. In addition to the Pareto distribution, we also consider the case that the mean of $Y(\theta)$ is $\mu(\theta)$ but the noise is given by a standard t distribution with degree of freedom is 3. The results are given in the appendix.

Finally, we provide the results of a multi-dimensional synthetic numerical experiment. In this experiment we set $d=6$ and $\theta = (\theta_1, \cdots, \theta_6)^T \in [0, 1]^6$. $\mu(\theta)$ is given by following logistic model:

$$\mu(\theta) = \frac{\exp(-\frac{1}{2} \sum_{i=1}^{d} (\theta_i - \frac{1}{3})^2 - 2)}{1 + \exp(-\frac{1}{2} \sum_{i=1}^{d} (\theta_i - \frac{1}{3})^2 - 2)}.$$  

Again we assume that $\mu(\theta)$ denotes the CTR, and set $k = 3, m = 100, \nu = \frac{1}{5}, \theta^0 = (\frac{1}{2}, \cdots, \frac{1}{2})^T, T = 10^7$, and $c_t = \frac{1}{T}, \alpha_t = \frac{20}{7}$, respectively. We repeat our algorithm 200 times and report the normalized estimator as Figure 1(i). The estimator is nearly normal distributed. In addition, the 95% CI has a 94.0% coverage rate.
In this section, the numerical experiments’ implication is that, classical zeroth-order stochastic optimization algorithms using two-point finite difference gradient estimator, despite of its convergence rate guarantee on the optimizer, may fail to provide a valid unbiased central limit theorem for the purpose of valid statistical inference. Our algorithm is proved to fix this issue and is illustrated by the above synthetic experiments.

6. Conclusion

We consider tasks when an A/B test experimenter not only needs to test whether the treatment plan is better than the control, but also needs to recommend an optimal choice of value for the treatment parameter to use for the treatment plan. However, we find that classical optimization algorithms, despite of their fast convergence rates under convexity assumptions, do not come with a central limit theorem that can be used to construct asymptotically valid confidence intervals. We provide a new optimization algorithm that on one hand maintains the same fast convergence rate under convexity assumptions and on the other hand permits the establishment of a valid central limit theorem. Future work may includes the extension to provide valid inference for adaptive A/B tests joint with non-convex optimization.

References

Ajalleian A, Stich SU (2020) On the convergence of SGD with biased gradients. arXiv preprint arXiv:2008.00051.

Andradóttir S, Prudius AA (2010) Adaptive random search for continuous simulation optimization. Naval Research Logistics (NRL) 57(6):583–604.

Balandat M, Karrer B, Jiang D, Daulton S, Letham B, Wilson AG, Bakshy E (2020) Botorch: a framework for efficient monte-carlo bayesian optimization. Advances in neural information processing systems 33:21524–21538.

Balasubramanian K, Ghadimi S (2018) Zeroth-order (non)-convex stochastic optimization via conditional gradient and gradient updates. Advances in Neural Information Processing Systems 31.
Duchi JC, Jordan MI, Wainwright MJ, Wibisono A (2015) Optimal rates for zero-order convex optimization: The power of two function evaluations. IEEE Transactions on Information Theory 61(5):2788–2806.

Dvurechensky P, Gorbunov E, Gasnikov A (2021) An accelerated directional derivative method for smooth stochastic convex optimization. European Journal of Operational Research 290(2):601–621.

Fan Q, Hu J (2018) Surrogate-based promising area search for lipschitz continuous simulation optimization. INFORMS Journal on Computing 30(4):677–693.

Fu MC, Chen CH, Shi L (2008) Some topics for simulation optimization. 2008 Winter Simulation Conference, 27–38 (IEEE).

Ghadimi S, Lan G (2013) Stochastic first-and zeroth-order methods for nonconvex stochastic programming. SIAM Journal on Optimization 23(4):2341–2368.

Gupta S, Kohavi R, Tang D, Xu Y, Andersen R, Bakshy E, Cardin N, Chandran S, Chen N, Coey D, et al. (2019) Top challenges from the first practical online controlled experiments summit. ACM SIGKDD Explorations Newsletter 21(1):20–35.

Hashemi FS, Ghosh S, Pasupathy R (2014) On adaptive sampling rules for stochastic recursions. Proceedings of the Winter Simulation Conference 2014, 3959–3970 (IEEE).

He S, Jiang G, Lam H, Fu MC (2021) Adaptive importance sampling for efficient stochastic root finding and quantile estimation. arXiv preprint arXiv:2102.10631.

Hong LJ, Zhang X (2021) Surrogate-based simulation optimization. Tutorials in Operations Research: Emerging Optimization Methods and Modeling Techniques with Applications, 287–311 (INFORMS).

Hu J, Fu MC, Marcus SI (2008) A model reference adaptive search method for stochastic global optimization. Communications in Information & Systems 8(3):245–276.

Hu J, Peng Y, Zhang G, Zhang Q (2022) A stochastic approximation method for simulation-based quantile optimization. INFORMS Journal on Computing.

Hunter SR, Applegate EA, Arora V, Chong B, Cooper K, Rincón-Guevara O, Vivas-Valencia C (2019) An introduction to multi-objective simulation optimization. ACM Transactions on Modeling and Computer Simulation (TOMACS) 29(1):1–36.

Johari R, Knoen P, Pekelis L, Walsh D (2017) Peeking at a/b tests: Why it matters, and what to do about it. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1517–1525.

Kiatsupaibul S, Smith RL, Zabinsky ZB (2018) Single observation adaptive search for continuous simulation optimization. Operations research 66(6):1713–1727.

Kiefer J, Wolffowitz J (1952) Stochastic estimation of the maximum of a regression function. The Annals of Mathematical Statistics 462–466.

Kim S, Pasupathy R, Henderson SG (2015) A guide to sample average approximation. Handbook of simulation optimization 207–243.
Kohavi R, Deng A, Frasca B, Walker T, Xu Y, Pohlmann N (2013) Online controlled experiments at large scale. Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, 1168–1176.

Kohavi R, Tang D, Xu Y (2020) Trustworthy online controlled experiments: A practical guide to a/b testing (Cambridge University Press).

Kushner H, Yin GG (2003) Stochastic approximation and recursive algorithms and applications, volume 35 (Springer Science & Business Media).

Kushner HJ, Clark DS (2012) Stochastic approximation methods for constrained and unconstrained systems, volume 26 (Springer Science & Business Media).

L'Ecuyer P, Glynn PW (1994) Stochastic optimization by simulation: Convergence proofs for the gi/g/1 queue in steady-state. Management Science 40(11):1562–1578.

L'Ecuyer P, Yin G (1998) Budget-dependent convergence rate of stochastic approximation. SIAM Journal on Optimization 8(1):217–247.

Letham B, Karrer B, Ottoni G, Bakshy E (2019) Constrained bayesian optimization with noisy experiments. Bayesian Analysis 14(2):495–519.

Liu S, Kailkhura B, Chen PY, Ting P, Chang S, Amini L (2018) Zeroth-order stochastic variance reduction for nonconvex optimization. Advances in Neural Information Processing Systems 31.

Nesterov Y, Spokoiny V (2017) Random gradient-free minimization of convex functions. Foundations of Computational Mathematics 17(2):527–566.

Pasupathy R, Ghosh S (2013) Simulation optimization: A concise overview and implementation guide. Theory Driven by Influential Applications 122–150.

Robbins H, Monro S (1951) A stochastic approximation method. The annals of mathematical statistics 400–407.

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

Spall JC (1992) Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. IEEE transactions on automatic control 37(3):332–341.

Tang D, Agarwal A, O'Brien D, Meyer M (2010) Overlapping experiment infrastructure: More, better, faster experimentation. Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, 17–26.

Wang J, Clark SC, Liu E, Frazier PI (2020) Parallel bayesian global optimization of expensive functions. Operations Research 68(6):1850–1865.

Wu J, Poloczek M, Wilson AG, Frazier P (2017) Bayesian optimization with gradients. Advances in neural information processing systems 30.
Xie J, Frazier PI, Chick SE (2016) Bayesian optimization via simulation with pairwise sampling and correlated prior beliefs. *Operations Research* 64(2):542–559.

Zhang Q, Hu J (2021) Actor-critic–like stochastic adaptive search for continuous simulation optimization. *Operations Research*.

Zhou E, Bhatnagar S (2017) Gradient-based adaptive stochastic search for simulation optimization over continuous space. INFORMS Journal on Computing 30(1):154–167.

Zhou E, Bhatnagar S, Chen X (2014) Simulation optimization via gradient-based stochastic search. *Proceedings of the Winter Simulation Conference 2014*, 3869–3879 (IEEE).
Appendix

Proof for Section 3

Proof of Proposition 1.

\[ T \text{Var}(\hat{\mu}(\hat{\theta}_T)) = \frac{4m^2 T/(2m)}{T} \sum_{i=1}^{\frac{m}{2}} \text{Var}(\hat{\mu}_i). \]

For each \( i \), note that we have:

\[ \text{Var}(\hat{\mu}_i) = \frac{\sigma(\theta^+)^2}{m_2} + \frac{k^4 \sigma(\theta^0)^2}{m_1} + \frac{k^4 \sigma(\theta^-)^2}{m_2} + \frac{\sigma(\theta^-)^2}{m_2}. \]

By Theorem 1, \( \theta^t \to \theta^* \) with probability 1, then by the continuity of \( \sigma(\cdot) \), we know \( \text{Var}(\hat{\mu}_i) \to \frac{\sigma(\theta^*)^2}{2(k^2-1)^2} \). Now, plug in and we obtain

\[ T \text{Var}(\hat{\mu}(\hat{\theta}_T)) \to \frac{m \sigma(\theta^*)^2}{(k^2-1)^2}, \]

with probability 1.

Proofs for Section 4

Proof of Theorem 1.

We will follow the idea of Theorem 3.1 of L’Ecuyer and Yin (1998). For simplicity, let’s consider minimizing \( \eta(\theta) = -\mu(\theta) \) instead of maximizing \( \mu(\theta) \) thereafter. Then the updating procedure of \( \theta_t \) is now

\[ \theta_{t+1} = \theta_t - \alpha_t g^t. \]

Note that we can write \( g^t \) as:

\[ g^t = (\nabla \eta(\theta^t)^T w^t) w^t + b_t + (g^t - \mathbb{E}[g^t | w^t, \theta^t]), \]

here \( b_t \) denotes the error of finite difference scheme, the last term is the random error comes from the sampling of \( y_i \), and the \( (\nabla \eta(\theta^t)^T w^t) \) is the directional derivative at \( \theta^t \).

WLOG we assume \( \theta^* = 0 \), then

\[
\begin{align*}
\mathbb{E}[\|\theta^{t+1}\|^2] & - \mathbb{E}[\|\theta^t\|^2] \\
= & \mathbb{E}[\|\theta^{t+1} - \theta^t\|^2] + 2\mathbb{E}(\theta^{t+1} - \theta^t)^T (\theta^t) \\
= & \mathbb{E}[\|\theta^{t+1} - \theta^t\|^2] - 2\alpha_t \mathbb{E}(g^t)^T (\theta^t).
\end{align*}
\]

Since we have

\[ \theta^{t+1} - \theta^t = -\alpha_t g^t, \]
and by Taylor expansion we have:

$$\nabla \eta(\theta^t) = \nabla^2 \eta(\theta^* \theta^t + O(\|\theta^t\|^2)), $$

plug in and we obtain:

$$ \mathbb{E}\|\theta^{t+1}\|^2 - \mathbb{E}\|\theta^t\|^2 
= \mathbb{E}\|\theta^{t+1} - \theta^t\|^2 - 2\alpha_t \mathbb{E}b_t^T \theta^t - 2\alpha_t \mathbb{E}(\nabla \eta(\theta^t) w^t) (w^t)^T (\theta^t) 
= \alpha_t^2 \mathbb{E}\|g^t\|^2 - 2\alpha_t \mathbb{E}b_t^T \theta^t - 2\alpha_t \mathbb{E}((\theta^t)^T \nabla^2 \eta(\theta^t) w^t (w^t)^T \theta^t) 
+ \alpha_t O(\mathbb{E}\|\theta^t\|^3).$$

Note that

$$ \mathbb{E}(\theta^t)^T \nabla^2 \eta(\theta^t) w^t (w^t)^T \theta^t 
= \mathbb{E} \text{tr}(w^t (w^t)^T \theta^t (\theta^t)^T \nabla^2 \eta(\theta^t)) 
= \text{tr}(\mathbb{E}[w^t (w^t)^T] \theta^t (\theta^t)^T \nabla^2 \eta(\theta^t)) 
= \text{tr}(\frac{I_d}{d} \mathbb{E}\theta^t (\theta^t)^T \nabla^2 \eta(\theta^t)) 
\geq \frac{\lambda_{\min}}{d} \mathbb{E}\|\theta^t\|^2,$$

here $d$ is the dimension of $\theta$ and $\lambda_{\min}$ is the minimum eigenvalue of $\nabla^2 \mu(\theta^*)$ and is positive since we assume that $\eta(\cdot)$ is strongly convex. Then we have:

$$ \mathbb{E}\|\theta^{t+1}\|^2 - \mathbb{E}\|\theta^t\|^2 
\leq - \frac{2c_0}{dt} \lambda_{\min} \mathbb{E}\|\theta^t\|^2 + \frac{c_0}{t} O(\mathbb{E}\|\theta^t\|^3) - \frac{2c_0}{t} \mathbb{E}b_t^T \theta^t 
+ \frac{3c_0^2}{t^2} \left( \mathbb{E}(\nabla \eta(\theta^t) w^t)^2 + \mathbb{E}\|b_t\|^2 + \text{Var}(g^t | \theta^t, \theta^t) \right). \quad (10)$$

Now, if we replace (A.3) in the proof of Theorem 3.1 in L’Ecuyer and Yin (1998) by the inequality above, then the proof can be adapted from the proof of Theorem 3.1 in L’Ecuyer and Yin (1998). This proof is given as follows.

By the continuity of $\nabla \eta(\theta)$, we can find a constant $K_\alpha > 0$ s.t.

$$ \mathbb{E}(\nabla \eta(\theta^t) w^t)^2 \leq \mathbb{E}\|\nabla \eta(\theta^t)\|^2 \leq 2K_\alpha (1 + \mathbb{E}\|\theta^t\|^2).$$

We also have

$$ \mathbb{E}\|b_t^T \theta^t\| \leq \mathbb{E}^{1/2}\|b_t\|^2 \mathbb{E}^{1/2}\|\theta^t\|^2.$$ 

By the property of central FD method, there exists $K_\beta > 0$ s.t.

$$ \mathbb{E}^{1/2}\|b_t\|^2 \leq K_\beta t^{-2\nu},$$

then by basic inequality, we have:

$$ \mathbb{E}\|b_t^T \theta^t\| 
\leq K_\beta t^{-2\nu} \left( t^{2\nu} \lambda_{\min} / (2dK_\beta) \mathbb{E}\|\theta^t\|^2 + K_\beta d / (2\lambda_{\min}) t^{-2\nu} \right),$$
which then gives:
\[
\mathbb{E}[\|b_1^T \theta^t \|^2] \leq \frac{\lambda_{\min}}{2d} \mathbb{E}[\|\theta^t\|^2] + K_3^2 d/(2\lambda_{\min}) t^{-4\nu}.
\]

For the \( O(\mathbb{E}[\|\theta^t\|^3]) \) term in (10), since \( \theta \) is bounded, it can easily be controlled by \( K_1(1 + \mathbb{E}[\|\theta^t\|^2]) \) with some \( K_1 > 0 \). Now, plug above results in to (10) and we obtain:
\[
\begin{align*}
\mathbb{E}[\|\theta^{t+1}\|^2] - \mathbb{E}[\|\theta^t\|^2] & \leq - \frac{c_0}{d} \lambda_{\min} \mathbb{E}[\|\theta^t\|^2] + \frac{c_0 K_1}{t} \mathbb{E}[\|\theta^t\|^2] + \frac{c_0 K_2 d}{\lambda_{\min}} t^{-4\nu-1} \\
& \quad + 3c_0 K_3^2 d/(\lambda_{\min}) t^{-4\nu} + 3c_0 K_3^2 K_2 t^{2\nu-2}.
\end{align*}
\]

(11)

Now, take \( 0 < \lambda_0 < c_0 \lambda_{\min}/d \), rearrange terms and take \( t \geq t_0 \) large enough we have:
\[
\begin{align*}
\mathbb{E}[\|\theta^{t+1}\|^2] & \leq (1 - \frac{\lambda_0}{t}) \mathbb{E}[\|\theta^t\|^2] + \frac{2c_0 K_1}{t} \mathbb{E}[\|\theta^t\|^2] + \frac{2c_0 K_3^2 d}{\lambda_{\min}} t^{-4\nu-1} \\
& \quad + 3c_0 K_3^2 K_2 t^{2\nu-2} + K_2 t
\end{align*}
\]

(12)

here \( K_2 > 0 \) is some constant. Note that, this result (12) adapts similar form with (A.8) in L’Ecuyer and Yin (1998) (their parameters \( \gamma, \beta, \delta \) correspond to our \( 1, 2\nu \) and \(-2\nu\)). Apply (12) iteratively and we obtain:
\[
\begin{align*}
\mathbb{E}[\|\theta^{t+1}\|^2] & \leq A_{t_0} \mathbb{E}[\|\theta_{t_0}\|^2] + \sum_{i=t_0}^t \frac{2c_0 K_1}{i} A_{t_i} \mathbb{E}[\|\theta^i\|^2] \\
& \quad + \frac{2c_0 K_3^2 d}{\lambda_{\min}} \sum_{i=t_0}^t A_{t_i} t^{-4\nu-1} \\
& \quad + 3c_0 K_3^2 K_2 \sum_{i=t_0}^t A_{t_i} i^{2\nu-2} + \sum_{i=t_0}^t A_{t_i} K_2 i
\end{align*}
\]

here \( A_{i,j} = \prod_{k=j+1}^i (1 - \lambda_0 k^{-1}) \) if \( j < t \), otherwise \( A_{i,j} = 1 \).

Since we are only interested in the order of \( \mathbb{E}[\|\theta^t\|^2] \), a difference in constant factors will actually not affect our results. Hence, we can calculate the order of those summations of \( A_{t,j} \) exactly the same as follows: for each fixed \( j < t \),
\[
|A_{t,j}| \leq \exp \left(-\lambda_0 \sum_{k=j+1}^t k^{-1} \right) \leq \exp(\lambda_0/t - \lambda_0 \int_j^t x^{-1} dx) \\
= \exp(\lambda_0/j)(j/t)^{\lambda_0}.
\]

We thus have
\[
\sum_{i=t_0}^t |A_{t_i}| i^{-4\nu-1} \leq \exp(\lambda_0/t_0) \sum_{i=t_0}^{t-1} i^{-4\nu-1} (i/t)^{\lambda_0} \\
\leq \exp(\lambda_0/t_0)(\lambda_0 - 4\nu)^{-1} t^{-4\nu}.
\]
Similarly,
\[ \sum_{i=t_0}^{t-1} i^{2\nu-2} |A_{ti}| \leq \exp(\lambda_0/t_0)(\lambda_0 + 2\nu - 1)^{-1} t^{2\nu-1}. \]

In addition, since
\[ A_{tj} - A_{t+1,j} = \lambda_0^{-1} A_{tj}, \]
we have
\[ \sum_{i=t_0+1}^{t} A_{ti} i^{-1} = A_{t} - A_{t_0}. \]
Thus, there exist some \( K_3 > 0 \) and some \( \epsilon_0 > 0 \):
\[ \mathbb{E}\|\theta^{t+1}\|^2 \leq \sum_{i=t_0}^{t} \frac{2c_0 K_1}{i} A_{ti} \mathbb{E}\|\theta^{i}\|^2 + \frac{2c_0 K_2^2 d}{\lambda_{\min}^2 \lambda_0 (1 - \epsilon_0 / \lambda_0)} t^{-2\nu} + \frac{3c_0^2 K_3}{\lambda_0 (1 - 2\epsilon_0 / \lambda_0)} t^{2\nu-1} + K_3 + O(t^{-4\nu - 1} + t^{-2+2\nu}). \]

Now, by Gronwall’s inequality we know that \( \sup_{i} \|\theta^i\|^2 < K_4 \) for some \( K_4 > 0 \). By Assumption 3 and Theorem 2.3.1 of Kushner and Clark (2012), we know \( \theta^t \) converges to \( \theta^* \) w.p.1. Combine these two results together we obtain that \( \mathbb{E}\|\theta^t\|^2 \to 0 = \mathbb{E}\|\theta^*\|^2 \) by our assumption.

Now, to obtain the final result, note that this time we will not replace the \( O(\mathbb{E}\|\theta^t\|^2) \) term in (10) by \( K_1 (1 + \mathbb{E}\|\theta^t\|^2) \), but bound it by \( \frac{\lambda_0}{2c_0} \mathbb{E}\|\theta^t\|^2 \), which is possible for large enough \( t \) based on the previous result \( \mathbb{E}\|\theta^t\|^2 \to 0 \). After that, we proceed exactly the same as before, then \( \sum_{i=t_0}^{t} A_{ti} \mathbb{E}\|\theta^{i}\|^2 \) and \( \sum_{i=t_0}^{t} A_{ti} i^{-1} \) terms will not appear, thus the inequality above will become:
\[ \mathbb{E}\|\theta^{t+1}\|^2 \leq \frac{4c_0 K_2^2 d}{\lambda_{\min} \lambda_0 (1 - 2\epsilon_0 / \lambda_0)} t^{-2\nu} + \frac{6c_0^2 K_3}{\lambda_0 (1 - 2\epsilon_0 / \lambda_0)} t^{2\nu-1} + O(t^{-1+2\nu}). \]

Since \( \frac{1}{6} < \nu < \frac{1}{4} \), then \(-1 + 2\nu > -4\nu \), so \( \mathbb{E}\|\theta^t\|^2 = O(t^{-1+2\nu}) \), which finishes the proof.

**Proof of Theorem 2.** Since \( \mu(\cdot) \) is third-order differentiable, we have
\[ \mu(\theta + h) + \mu(\theta - h) = 2\mu(\theta) + h^T \nabla^2 \mu(\theta) h + O(||h||^3), \]
then plug in and we obtain:
\[ \mu(\theta^+ + \theta^-) = 2\mu(\theta) + c_1^2 (w^i)^T \nabla^2 w^i + O(c_1^3), \]
and
\[ \mu(\theta^{++} + \theta^{--}) = 2\mu(\theta) + k^2 c_1^2 (w^i)^T \nabla^2 w^i + O(c_1^3). \]
From above two equations we have:
\[ \mu(\theta^t) = \frac{-\mu(\theta^{++}) + k^2 \mu(\theta^+) + k^2 \mu(\theta^-) - \mu(\theta^{--})}{2(k^2 - 1)} + O(c_1^3). \] (13)
Now, if we decompose $\hat{\mu}_t$ in the following way:

$$\hat{\mu}_t = (\hat{\mu}_t - \mathbb{E}[\hat{\mu}_t | \theta^{t-1}]) + (\mathbb{E}[\hat{\mu}_t | \theta^{t-1}] - \mathbb{E}[\mu(t') | \theta^{t-1}]) + \mathbb{E}[\mu(t') | \theta^{t-1}],$$

then by (13), we have

$$\hat{\mu}_t = (\hat{\mu}_t - \mathbb{E}[\hat{\mu}_t | \theta^{t-1}]) + \mathbb{E}[\mu(t') | \theta^{t-1}] + O_p(c_1^2).$$

Then

$$\sqrt{T}\left( \frac{1}{T} \sum_{i=1}^{T/(2m)} \hat{\mu}_i - \mu(\theta^*) \right)$$

$$= \frac{1}{\sqrt{T}} \sum_{i=1}^{T/(2m)} (\hat{\mu}_i - \mathbb{E}[\hat{\mu}_i | \theta^{t-1}])$$

$$+ \frac{1}{\sqrt{T}} \left( \sum_{i=1}^{T/(2m)} (\mathbb{E}[\mu(t') | \theta^{t-1}] - \mu(\theta^*)) \right) + O_p(T^{\frac{1}{2} - 3\nu}).$$

As for the first term, from Assumption 4 we know Lindeberg condition is satisfied, so by martingale CLT we obtain a CLT for the first summation. By Proposition 1 we know the asymptotic variance is $(k^2 + 1)^2\sigma(\theta^*)^2$ if we set $m_1, m_2$ as we discussed in Section 3.3. The third term is $o_p(1)$ since $\nu > \frac{1}{6}$. Now we only need to control the second term, note that

$$\mathbb{E}\left| \sum_{i=1}^{T/(2m)} (\mathbb{E}[\mu(t') | \theta^{t-1}] - \mu(\theta^*)) \right|$$

$$\leq \mathbb{E}\sum_{i=1}^{T/(2m)} ||\mathbb{E}[\mu(t') - \mu(\theta^*) || \theta^{t-1}]$$

$$= \sum_{i=1}^{T/(2m)} \mathbb{E}[||\mathbb{E}[\mu(t') - \mu(\theta^*)] ||]$$

$$\leq M \sum_{i=1}^{T/(2m)} \mathbb{E}[||\theta^t - \theta^*||^2].$$

Here $M = \sup_{\theta \in [0, 1]^d} \lambda_{\max}(\nabla^2 \mu(\theta))$, since $\nabla^2 \mu$ is continuous, so $M < +\infty$. From Theorem 1 we know $\mathbb{E}[||\theta^t - \theta^*||^2] = O(t^{-1+2\nu})$, so the second term is $O_p(T^{-\frac{1}{2} + 2\nu}) = o_p(1)$. Combine all above results together we obtain the following CLT:

$$\sqrt{T}(\hat{\mu}_T - \mu(\theta^*)) \xrightarrow{d} \frac{(k^2 + 1)\sigma(\theta^*)}{k^2 - 1} \mathcal{N}(0, 1)$$

as $T \to +\infty$. Similarly we have

$$\sqrt{T}\mathbb{E}[\mu(\hat{\theta}_T) - \mu(\theta^*)] = o(1),$$

so $\sqrt{T}(\hat{\mu}(\hat{\theta}_T) - \mu(\theta^*)) = o_p(1)$, which then gives

$$\sqrt{T}(\hat{\mu}(\hat{\theta}_T) - \mu(\hat{\theta}_T)) \xrightarrow{d} \frac{(k^2 + 1)\sigma(\theta^*)}{k^2 - 1} \mathcal{N}(0, 1)$$
as $T \to +\infty$.

**Consistency of $\hat{\sigma}(\theta^*)$.** We will need an extra condition that the fourth order moment of $\epsilon_\theta$ are uniformly bounded for all $\theta \in [0,1]^d$. The estimator of $\sigma(\theta^*)$ is given by:

$$\hat{\sigma}^2(\theta^*) = \sum_{i=1}^T (y_i - \bar{y})^2 / (T - I + 1),$$

here $\bar{y} = \frac{1}{T-I+1} \sum_{i=1}^T y_i$. Recall that $y_i$ is a realization of $Y(\theta_i)$, we have

$$\hat{\sigma}^2(\theta^*) = \frac{1}{T-I+1} \sum_{i=1}^T \mu(\theta_i) + \sigma(\theta_i) \cdot \epsilon_{\theta_i} - \bar{y})^2.$$  

Expand the square and we have:

$$\hat{\sigma}^2(\theta^*) = \frac{1}{T-I+1} \sum_{i=1}^T \mu^2(\theta_i) - \bar{\mu}^2$$

$$+ \frac{1}{T-I+1} \sum_{i=1}^T \sigma^2(\theta_i) \cdot \epsilon_{\theta_i} - \bar{\sigma}^2 + O_p(\frac{1}{T}),$$

here $\bar{\mu} = \frac{1}{T-I+1} \sum_{i=1}^T \mu(\theta_i)$ and $\bar{\sigma} = \frac{1}{T-I+1} \sum_{i=1}^T \sigma(\theta_i) \cdot \epsilon_{\theta_i}$. Then

$$\hat{\sigma}^2(\theta^*) = \frac{1}{T-I+1} \sum_{i=1}^T \mu^2(\theta_i) - \bar{\mu}^2$$

$$+ \frac{1}{T-I+1} \sum_{i=1}^T \sigma^2(\theta_i) \cdot \epsilon_{\theta_i}^2 - \bar{\sigma}^2 + O_p(\frac{1}{T}).$$

For the first two terms, since $\mu(\theta_i) \to \mu(\theta)^*$ and $\bar{\mu} \to \mu(\theta)^*$, by Stolz theorem we know it’s $o(1)$. For $\sigma^2$, by CLT we know it’s $O_p(\frac{1}{T})$. As for $\frac{1}{T-I+1} \sum_{i=1}^T \sigma^2(\theta_i) \cdot \epsilon_{\theta_i}^2$, by our extra condition it is $\frac{1}{T-I+1} \sum_{i=1}^T \sigma^2(\theta_i)$. Since $\sigma(\theta_i) \to \sigma(\theta^*)$, again by Stolz theorem we know $\frac{1}{T-I+1} \sum_{i=1}^T \sigma^2(\theta_i) = \sigma^2(\theta^*) + o(1)$. Thus, $\hat{\sigma}^2(\theta^*) = \sigma^2(\theta^*) + o_p(1)$, which finishes our proof.

**Additional Numerical Results of t-distribution** The results of t-distribution we mentioned in numerical experiments are given by Figure 1(j) and 1(k).
(j) Result of Algorithm 1, t distribution

(k) Result of two-point central FD method, t distribution