Comparison-Based Algorithms for One-Dimensional Stochastic Convex Optimization

Xi Chen∗, Qihang Lin† and Zizhuo Wang‡

1Stern School of Business, New York University
2Tippie College of Business, University of Iowa
3Department of Industrial and Systems Engineering, University of Minnesota

Abstract

Stochastic optimization finds a wide range of applications in operations research and management science. However, existing stochastic optimization techniques often require the information of random samples (e.g., demands in newsvendor problem) or the objective values at the sampled points (e.g., the lost sales cost), which might not be available in practice. In this paper, we consider a new setup for stochastic optimization, in which the decision maker can only access to comparison information between a random sample and two chosen points in each iteration. We propose a comparison-based algorithm (CBA) to solve such problems in single dimension with convex objective functions. Particularly, the CBA properly chooses the two points in each iteration and constructs an unbiased gradient estimate for the original problem. We show that the CBA achieves the same convergence rate as the optimal stochastic gradient methods (with the samples observed). We also consider extensions of our approach to high dimensional quadratic problems as well as problems with non-convex objective functions. Numerical experiments show that the CBA performs well in test problems.

1. Introduction

In this paper, we consider the following stochastic optimization problem:

$$\min_{\ell \leq x \leq u} H(x) = E_{\xi} [h(x, \xi)]$$

where $-\infty \leq \ell \leq u \leq +\infty$. This problem has many applications and is fundamental for stochastic optimization. For example,

1. If $h(x, \xi) = (x - \xi)^2$ with $\ell = -\infty$ and $u = +\infty$, then the problem is to find the expectation of $\xi$;

* xchen3@stern.nyu.edu
† qihang-lin@uiowa.edu
‡ zwang@umn.edu

1 Throughout the note, when $\ell = -\infty$ ($u = +\infty$, resp.), the notation $\ell \leq x$ ($x \leq u$, resp.) will be interpreted as $x > -\infty$ ($x < +\infty$, resp.).
2. If \( h(x, \xi) = h \cdot (x - \xi)^+ + b \cdot (\xi - x)^+ \), then the problem is the classical newsvendor problem with unit holding cost \( h \) and unit backorder cost \( b \). It can also be viewed as the problem of finding the \( \frac{b}{h + b} \)-th quantile of \( \xi \) when \( \ell = -\infty \) and \( u = +\infty \). Furthermore, one can consider a more general version of this problem in which

\[
h(x, \xi) = \begin{cases} 
  h_+(x, \xi) & \text{if } x \geq \xi \\
  h_-(x, \xi) & \text{if } x < \xi.
\end{cases}
\]

This problem can be viewed as a newsvendor problem with general holding and backorder costs (see, e.g., Halman et al. 2012 and references therein for discussions of this problem, where \( h_+(x, \xi) \) is a general holding cost function and \( h_-(x, \xi) \) is a general backorder cost function). It can also be viewed as a single period appointment scheduling problem with general waiting and overtime costs (see, e.g., Gupta and Denton 2008) or a staffing problem with general underage and overage costs (see, e.g., Kolker 2017).

3. If \( h(x, \xi) = -x \cdot 1(\xi \geq x) \), then the problem can be viewed as an optimal pricing problem where \( x \) is the price set by the seller, \( \xi \) is the valuation of each customer, and \( h(x, \xi) \) is the negative of the revenue obtained from the customer (a customer purchases at price \( x \) if and only if his/her valuation \( \xi \) is greater than or equal to \( x \)).

In many practical settings, the distribution of \( \xi \) (whose c.d.f. will be denoted by \( F(\cdot) \)) is unknown a priori. To solve (1), existing stochastic optimization techniques often require knowing either the random sample \( \xi \) or the objective value \( h(x, \xi) \) at a decision \( x \) and a sampled point \( \xi \). However, such information may not be available in practice (see Examples 1-3 below). Instead, the comparison relationship between a chosen decision variable and a random sample is sometimes accessible. This motivates us to study stochastic optimization with only the presence of comparison information. Specifically, given a decision \( x \), a sample \( \xi \) is drawn from the underlying distribution, and we assume that we only have information about whether \( \xi \) is greater than or less than (or equals to) \( x \). In addition, after knowing the relationship between \( x \) and \( \xi \), we further assume that we can choose another point \( z \) and obtain information about whether \( \xi \) is greater than or less than (or equals to) \( z \). (Such a \( z \) is not treated as a decision variable.) In the following, we list several practical scenarios that fit this setting:

**Example 1.** Suppose \( x \) represents a certain feature of a product (e.g., size, taste, etc) and \( \xi \) is the preference of each customer about that feature, and the firm selling this product would like to find out the average preference of the customers (or equivalently, to find the optimal offering to minimize the expected customer dissatisfaction which is measured by \( h(x, \xi) = (x - \xi)^2 \)). Such a firm faces a stochastic optimization problem described in the first example above. In many cases, it is hard for a customer to give an exact value for his/her preference (i.e., the exact value of his/her \( \xi \)). However, it is quite plausible that the customer can report comparison relationship between his/her preferred value of the feature and the actual value of the feature of the product presented to him/her (e.g., whether the product should be larger or whether the taste should be saltier). Furthermore, in many cases, it is possible to ask two such questions to a single customer, for example, by giving each customer two samples. Moreover, the second sample may be given in a customer satisfaction survey, and the customer will not count the second sample toward its (dis)satisfaction value. Therefore, such a scenario fits the setting described above.

**Example 2.** In a newsvendor problem, it is sometimes hard to observe the exact demand in each period due to demand censorship. In such situations, one does not have direct access...
to the sample point (the demand) nor does one have access to the cost in the corresponding period (the lost sales cost). However, the seller often has comparison information between the realized demand and the chosen inventory level (e.g., by observing if there is a stock out or a leftover). Moreover, by allowing the seller to make a one-time additional ordering in each time period (this ability is sometimes called the quick response ability for the seller, see e.g., Cachon and Swinney 2011), it is possible that one can obtain such information at two points. In such cases, the firm will face a newsvendor problem as described in the second example earlier and thus it will correspond to the setting in our problem.

Example 3. In a revenue management problem, by offering a price to each customer, the seller can observe whether the customer purchased the product or not, and the seller faces a stochastic optimization problem described in the third example above. In practice, it is hard to ask the customer to report his true valuation of the product. However, it is possible to ask the customer whether he will purchase the product at a different price (e.g., for a customer who didn’t purchase, the seller can offer him a discount immediately and ask whether he would like to purchase or not; for a customer who did purchase, the seller can ask whether he is still willing to purchase should the price be a little higher). Such an example can also be extended to the divisible product case where a customer can buy a continuous number of a product with a maximum amount of 1. In this case, the $h$ function can be redefined as $h(x, \xi) = -x \min\{1, g(x, \xi)\}$ where $x$ is the offered price, $g(x, \xi)$ is the unconstrained purchase amount of the customer, and $\xi$ is the maximum price this customer is willing to buy the full amount of this product (i.e., $g(x, \xi)$ is decreasing in $x$ with $g(x, \xi) > 1$ when $x < \xi$ and $g(x, \xi) < 1$ when $x > \xi$). Such a price behavior can be explained by a quadratic utility function of the customer, which is often used in the literature (see e.g., Candogan et al. 2012). For the seller, by observing whether the customer buys the full amount of the product, he can infer whether an offered price is greater than or less than the $\xi$ value of this customer.

In this paper, we propose an efficient algorithm to solve the above-described stochastic optimization problem. More precisely, we propose a stochastic approximation algorithm that only utilizes comparison information between each sample point and two chosen points in each iteration. We show that by properly choosing the two points (one point has to be chosen randomly according to a specifically designed distribution), we can obtain unbiased gradient estimates for the original problem. The unbiased gradient estimates will in turn give rise to efficient algorithms based on standard stochastic gradient method (we will review the related literature shortly). Under some mild conditions, we show that if the original problem is convex, then our algorithm will achieve a convergence rate of $O(1/\sqrt{T})$ for the objective value (where $T$ is the number of iterations); if the original problem is strongly convex, then the convergence rate can be improved to $O(1/T)$. These results can also be translated into $O(\sqrt{T})$ and $O(\log T)$ regrets in an online learning setting. Moreover, the information at two points is necessary in this setting as we show that only knowing comparison information between the sample and one point in each iteration is insufficient for any algorithm to converge to the optimal solution (see Example 4). We also perform a few numerical experiments using our algorithm. The experiment results show that our algorithms are indeed efficient, with convergence speed in the same

---

2 There has been a vast literature on newsvendor/inventory problems with censored demand. For some recent references, we refer the readers to Ding et al. (2002), Bensoussan et al. (2007) and Besbes and Muharremoglu (2013).

3 If $h(x, \xi)$ is piecewise linear with two pieces, e.g., $h(x, \xi) = h \cdot (x - \xi)^+ + b \cdot (\xi - x)^+$, only comparing $x$ and $\xi$ may be sufficient to compute the stochastic gradient $h'(x, \xi)$ (that equals $h$ or $-b$).

4 For a review of the setup in an online learning problem we refer the readers to Shalev-Shwartz (2012).
order compared to the case when one has direct observations of the samples. We also extend our algorithm to a multi-dimensional setting with quadratic objective function, a setting with non-convex objective function and a setting in which multiple comparisons can be conducted in each iteration.

**Literature Review.** Broadly speaking, our work falls into the area of stochastic optimization. There has been a vast literature on stochastic optimization. For a comprehensive review of this literature, we refer the readers to Shapiro et al. (2014). In particular, in this literature, it is usually assumed that one has access to random samples (or alternatively, the objective values at the sampled points). Based on the samples, two main types of algorithms have been proposed, namely the stochastic approximation methods (e.g., Robbins and Monro 1951, Kiefer and Wolfowitz 1952) and the sample average approximation method (e.g., Shapiro et al. 2014). In the stochastic approximation methods (e.g., the stochastic gradient methods), in each iteration, a new sample (or new samples) is drawn and a new iterate is computed by using the new sample(s); while in the sample average approximation method, all samples are drawn at once and will be used in all iterations. Due to its low computational cost and memory requirement, the stochastic approximation method has received much attention recently in optimization and machine learning literature, especially for solving problems with big data. Our work also falls into this category. In the following, we shall focus our literature review on the stochastic approximation methods.

If the objective function is convex, then various stochastic gradient methods can guarantee, under slightly different assumptions, that the objective value of the iterates converges to the optimal value in a rate of $O(1/\sqrt{T})$ after $T$ iterations (see, e.g., Nemirovski et al. 2009, Duchi and Singer 2009, Hu et al. 2009, Xiao 2010, Lin et al. 2011, Chen et al. 2012, Lan 2012, Lan and Ghadimi 2012, Rakhlin et al. 2012, Lan and Ghadimi 2013, Shamir and Zhang 2013, Hazan and Kale 2014). Furthermore, this convergence rate is known to be optimal (Nemirovski and Yudin, 1983). When the objective function is strongly convex, some stochastic gradient methods can obtain an improved convergence rate of $O(\log T/T)$ (Duchi and Singer, 2009, Xiao, 2010). More recently, several papers have further improved the convergence rate to $O(1/T)$. Among those papers, there are three different methods used: (a) accelerated stochastic gradient method with auxiliary iterates besides the main iterate (Hu et al., 2009, Lin et al., 2011, Lan, 2012, Lan and Ghadimi, 2012, Chen et al., 2012, Lan and Ghadimi, 2013); (b) averaging the historical solutions (Rakhlin et al., 2012, Shamir and Zhang, 2013) and (c) multi-stage stochastic gradient method that periodically restarts (Hazan and Kale, 2014). Again, the convergence rate of $O(1/T)$ has been shown to be optimal by Nemirovski and Yudin (1983) for strongly convex problems. Thus the recent results have matched the best possible convergence rate.

Apart from the setting of minimizing a single objective function, stochastic gradient methods can also be applied to online learning problems (see Shalev-Shwartz 2012 for a comprehensive review) where a sequence of functions is presented to a decision maker who needs to provide a solution sequentially to each function with the goal of minimizing the total regret. It is known that the stochastic gradient methods can obtain a regret of $O(\sqrt{T})$ after $T$ decisions if the functions presented are convex. Moreover, this regret has been shown to be optimal (Cesa-Bianchi and Lugosi, 2006). When the functions are strongly convex, Zinkevich (2003), Duchi and Singer (2009), Duchi et al. (2011), and Xiao (2010) show that the regret can be further improved to $O(\log T)$.

To distinguish our work from the above, we note that all of the above works have assumed that either the sample is directly accessible (one can observe the value of each sample) or the
objective value corresponding to the decision variable and the current sample is accessible. In either case, it is easy to obtain an estimate of the gradient of the objective function. In contrast, in our case, we do not have access to the sample or the objective value. Instead, we only have comparison information between each sample and two chosen points. Indeed, as we shall discuss in the next section, one of the main challenges in our problem is to use this very limited information to construct an unbiased gradient for the original problem and then further use it to find the optimal solution. Our contribution is to show that the same order of convergence rate can still be achieved under this setting with less information.

2. Main Results

In this paper, we make the following assumptions:

Assumption 1.

(A1) The random variable $\xi$ follows a continuous distribution.

(A2) For each $\xi$, $h(x, \xi)$ is continuously differentiable with respect to $x$ on $[\ell, \xi]$ and $(\xi, u]$ with the derivative denoted by $h'_x(x, \xi)$. Furthermore, for any $x \in [\ell, u]$, $h'_-(x) := \lim_{z \to x^-} h'_x(x, z)$ and $h'_+(x) := \lim_{z \to x^+} h'_x(x, z)$ exist and are finite.

(A3) For any $x \in [\ell, u]$, $x \neq \xi$, $h''_{x, \xi}(x, \xi) = \frac{\partial^2 h(x, \xi)}{\partial \xi \partial x}$ exists.

(A4) $H(x)$ in (1) is differentiable and $\mu$-convex on $[\ell, u]$ for some $\mu \geq 0$, namely,

$$H(x_2) \geq H(x_1) + H'(x_1)(x_2 - x_1) + \frac{\mu}{2}(x_2 - x_1)^2, \quad \forall x_1, x_2 \in [\ell, u]. \quad (2)$$

Moreover, $\mathbb{E}_\xi(h'_x(x, \xi)) = H'(x)$ for all $x \in [\ell, u]$.

(A5) Either of the following statements is true:

a. There exists a constant $K_1$ such that $\mathbb{E}_\xi(h'_x(x, \xi))^2 \leq K_1^2$ for any $x \in [\ell, u]$;

b. $H'(x)$ is $L$-Lipschitz continuous on $[\ell, u]$. Furthermore, there exists a constant $K_2$ such that

$$\mathbb{E}_\xi(h'_x(x, \xi) - H'(x))^2 \leq K_2^2, \quad \forall x \in [\ell, u].$$

Now we make a few comments on the above assumptions. The first assumption that $\xi$ is continuously distributed is mainly for the ease of discussion. In fact, all of our results will continue to hold as long as with probability 1, for all iterates $x$ in our algorithm, $\mathbb{P}(\xi = x) = 0$. We shall revisit this assumption in Section 6. Assumptions A2-A4 are some regularity assumptions on the functions $h$ and $H$. Particularly, the last point of Assumption A4 is satisfied under many cases, for example, when $h'_x(x, \xi)$ is continuous in $\xi$ and $\xi$ is supported on a finite set (Widder 1990), or when $h(x, \xi)$ is convex in $x$ for each $\xi$ (by monotone convergence theorem). When (2) holds with $\mu > 0$, we call $H$ a $\mu$-strongly convex function. The last assumption states that the partial derivative $h'_x(x, \xi)$ has uniformly bounded second-order moment or variance. This is used to guarantee that the step in each iteration in our algorithm has bounded variance, which is a common assumption in stochastic approximation literature (see, e.g., Nemirovski et al. 2009, Duchi and Singer 2009, Lan and Ghadimi 2013).
In addition, Assumption 1 is not hard to satisfy in our examples mentioned earlier. Specifically, for Example 1, it satisfies Assumption 1 when ξ is continuously distributed and has finite variance. For Example 2, it satisfies Assumption 1 when ξ is continuously distributed and the cost functions are linear. When the cost functions are nonlinear (h_1 and h_- respectively), it satisfies Assumption 1 if both h_1 and h_- are second-order continuously differentiable on their respective domains, have bounded first-order derivatives (for example, when x and ξ are restricted to finite intervals), and h is convex in x. For example 3, it satisfies Assumption 1 under the divisible case when ξ is a continuous random variable and the expected revenue function $x\mathbb{E}_x \min\{g(x, ξ), 1\}$ is concave on [ℓ, u], which holds, for example, when g(x, ξ) is a piecewise linear function and when the range [ℓ, u] is small.5

In the following, we propose a comparison-based algorithm (CBA) to solve (1). Let Ξ denote the support of ξ with $-\infty \leq \underline{s} := \inf\{Ξ\} \leq \bar{s} := \sup\{Ξ\} \leq +\infty$. The algorithm requires specification of two functions, $f_-(x, z)$ and $f_+(x, z)$, which need to satisfy the following conditions.

- (C1) $f_-(x, z) = 0$ for all $z \geq x$ and $f_-(x, z) > 0$ for all $\underline{s} \leq z < x$. In addition, for all x, we have $\int_{\underline{s}}^x f_-(x, z)dz = 1$.
- (C2) $f_+(x, z) = 0$ for all $z \leq x$ and $f_+(x, z) > 0$ for all $\bar{s} \geq z > x$. In addition, for all x, we have $\int_{\bar{s}}^x f_+(x, z)dz = 1$.
- (C3) There exists a constant $K_3$ such that $\int_{\underline{s}}^x \frac{F(z)(h''_+(x, z))^2}{f_-(x, z)}dz \leq K_3$ and $\int_{\bar{s}}^x \frac{(1-F(z))(h''_-(x, z))^2}{f_+(x, z)}dz \leq K_3$ for all $x \in [\ell, u]$, where $F(\cdot)$ is the c.d.f. of ξ.

Note that, for any given $x \in [\ell, u]$, $f_-(x, z)$ and $f_+(x, z)$ essentially define two density functions of z on $(-\infty, x]$ and $[x, +\infty)$. (We will discuss how to choose $f_-(\cdot, \cdot)$ and $f_+(\cdot, \cdot)$ in Section 3.) Also, we note that $\underline{s}$ and $\bar{s}$ need not to be known in advance. If one is unsure about $\underline{s}$ ($\bar{s}$, resp.), then one can choose a sufficiently small (large, resp.) value, or just choose $\underline{s} = -\infty$ ($\bar{s} = +\infty$, resp.). Condition (C3) is a technical condition and may not be straightforward to verify at the first glance. However, in Section 3, we show that under mild conditions (e.g., ξ has a light tail and $h''_{x,z}(x, z)$ is uniformly bounded), it is not hard to choose the functions $f_-(x, z)$ and $f_+(x, z)$ such that condition (C3) is satisfied (we will leave the detailed discussions in Section 3). Next, in Algorithm 1, we describe the detailed procedure of the CBA.

In iteration t of CBA, the solution $x_t$ will be updated based on two random samples. First, a sample $\xi_t$ is drawn from the distribution of ξ. In contrast to existing stochastic gradient methods, CBA does not require exactly observing $\xi_t$ but only needs to know whether $\xi_t < x_t$ or $\xi_t > x_t$. Based on the result of the comparison between $\xi_t$ and $x_t$, a second sample $z_t$ is drawn from the density function $f_+(x_t, z)$ or $f_-(x_t, z)$. An unbiased stochastic gradient, $g(x_t, \xi_t, z_t)$, of $H(x)$ is then constructed and used to update $x_t$ with the standard gradient descent step.

We note that the output of Algorithm 1 is the average of the historical solutions $\bar{x}_T = \frac{1}{T} \sum_{t=1}^T x_t$. This is because the convergence of objective value is established based on $\bar{x}_T$.

However, Algorithm 1 can be applied to the online learning setting where one can use the solution $x_t$ as the decision in each stage t and obtain the desired expected total regret (see Theorems 1-3). We have the following proposition about the stochastic gradient $g(x_t, \xi_t, z_t)$ in CBA.

---

5The indivisible product case with $h(x, \xi) = -x \cdot 1(\xi \geq x)$ does not satisfy Assumption (A4). In particular, it does not satisfy $\mathbb{E}_\xi(h'_+(x, \xi)) = H'(x)$ (it does satisfy all the other assumptions under mild conditions though). In order to satisfy $\mathbb{E}_\xi(h'_+(x, \xi)) = H'(x)$, it is sufficient that $h(x, \xi)$ is continuous in x, which holds in the divisible product case.
Algorithm 1 Comparison-Based Algorithm (CBA):

1. **Initialization.** Set $t = 1$, $x_1 \in [\ell, u]$. Define $\eta_t$ for all $t \geq 1$. Set the maximum number of iterations $T$. Choose functions $f_-(x, z)$ and $f_+(x, z)$ that satisfy (C1)-(C3).

2. **Main iteration.** Sample $\xi_t$ from the distribution of $\xi$. If $\xi_t = x_t$, then resample $\xi_t$ until it does not equal $x_t$. (This step will always terminate in a finite number of steps as long as $\xi$ is not deterministic.)

   (a) If $\xi_t < x_t$, then generate $z_t$ from a distribution on $(-\infty, x_t]$ with p.d.f. $f_-(x_t, z_t)$. Set
   \[
   g(x_t, \xi_t, z_t) = \begin{cases} 
   h'_-(x_t), & \text{if } z_t < \xi_t, \\
   h'_-(x_t) \frac{h''_-(x_t, z_t)}{f_-(x_t, z_t)} & \text{if } z_t \geq \xi_t.
   \end{cases}
   \]

   (b) If $\xi_t > x_t$, then generate $z_t$ from a distribution on $[x_t, +\infty)$ with p.d.f. $f_+(x_t, z_t)$. Set
   \[
   g(x_t, \xi_t, z_t) = \begin{cases} 
   h'_+(x_t), & \text{if } z_t > \xi_t, \\
   h'_+(x_t) + \frac{h''_+(x_t, z_t)}{f_+(x_t, z_t)} & \text{if } z_t \leq \xi_t.
   \end{cases}
   \]

   Let
   \[
   x_{t+1} = \operatorname{Proj}_{[\ell, u]}(x_t - \eta_t g(x_t, \xi_t, z_t)) = \max(\ell, \min(u, x_t - \eta_t g(x_t, \xi_t, z_t))).
   \]

3. **Termination.** Stop when $t \geq T$. Otherwise, let $t \leftarrow t + 1$ and go back to Step 2.

4. **Output.** $\text{CBA}(x_1, T, \{\eta_t\}_{t=1}^T) = \bar{x}_T = \frac{1}{T} \sum_{t=1}^T x_t.$

**Proposition 1.** Suppose $f_-(x, z)$ and $f_+(x, z)$ satisfy (C1)-(C3) and Assumption 1 holds. Then

1. $\mathbb{E}_z g(x, \xi, z) = h'_-(x, \xi)$, for all $x \in [\ell, u]$, $x \neq \xi$.
2. $\mathbb{E}_z g(x, \xi, z) = H'(x)$, for all $x \in [\ell, u]$.
3. If Assumption A5(a) holds, then $\mathbb{E}_z \xi(g(x, \xi, z))^2 \leq G^2 := K_1^2 + 2K_3$. If Assumption A5(b) holds, then $\mathbb{E}_z \xi(g(x, \xi, z) - H'(x))^2 \leq \sigma^2 := K_2^2 + 2K_3$.

**Proof of Proposition 1.** First, we consider the case when $\xi < x$. We have
\[
\mathbb{E}_z g(x, \xi, z) = h'_-(x) - \int_{\xi}^{x-} h''_{x,z}(x, z) dz = h'_-(x, \xi).
\]
Similarly, when $\xi > x$,
\[
\mathbb{E}_z g(x, \xi, z) = h'_+(x) + \int_{x+}^{\xi} h''_{x,z}(x, z) dz = h'_+(x, \xi).
\]
Thus the first conclusion of the proposition is proved. The second conclusion of the proposition follows from Assumption A1 (which ensures $\xi = x$ is a zero-measure event) and Assumption A4.
Next, we show the first part of the third conclusion when Assumption A5(a) is true. If \( \xi < x \), then we have
\[
\mathbb{E}_z(g(x, \xi, z))^2 = \int_{-\infty}^{x-} (h'_-(x))^2 f_- (x, z) \, dz + \int_{\xi}^{x-} \left( -2h'_-(x) \frac{h''_{x,z}(x, z)}{f_-(x, z)} + \left(\frac{h''_{x,z}(x, z)}{f_-(x, z)}\right)^2 \right) f_-(x, z) \, dz \\
= (h'_-(x))^2 - 2h'_-(x)(h'_-(x) - h'_{x}(x, \xi)) + \int_{\xi}^{x-} \frac{(h''_{x,z}(x, z))^2}{f_-(x, z)} \, dz \\
\leq (h'_x(x, \xi))^2 + \int_{\xi}^{x-} \frac{(h''_{x,z}(x, z))^2}{f_-(x, z)} \, dz.
\]
where the last inequality is because \( a^2 + b^2 \geq 2ab \) for any \( a, b \). By similar arguments, if \( \xi > x \), then
\[
\mathbb{E}_z(g(x, \xi, z))^2 \leq (h'_x(x, \xi))^2 + \int_{x+}^{\xi} \frac{(h''_{x,z}(x, z))^2}{f_+(x, z)} \, dz.
\]
These two inequalities and Assumption A5(a) further imply
\[
\mathbb{E}_{x,\xi}(g(x, \xi, z))^2 \leq K_1^2 + \int_{\xi}^{x-} \left( \int_{\xi}^{x-} \frac{(h''_{x,z}(x, z))^2}{f_-(x, z)} \, dz \right) dF(\xi) + \int_{x+}^{\xi} \left( \int_{x+}^{\xi} \frac{(h''_{x,z}(x, z))^2}{f_+(x, z)} \, dz \right) dF(\xi) \\
= K_1^2 + \int_{\xi}^{x-} F(z)(h''_{x,z}(x, z))^2 \, dz + \int_{x+}^{\xi} (1 - F(z))(h''_{x,z}(x, z))^2 \, dz \\
\leq K_1^2 + 2K_3,
\]
where the interchanging of integrals in the equality is justified by Tonelli’s theorem and the last inequality is due to (C3).

Next, we show the second part of the third conclusion when Assumption A5(b) is true. If \( \xi < x \), then following the similar analysis as in (6), we have
\[
\mathbb{E}_z(g(x, \xi, z) - h'_x(x, \xi))^2 = \mathbb{E}_z(g(x, \xi, z))^2 - (h'_x(x, \xi))^2 \leq \int_{\xi}^{x-} (h''_{x,z}(x, z))^2 \, dz.
\]
Similarly, if \( \xi > x \), then
\[
\mathbb{E}_z(g(x, \xi, z) - h'_x(x, \xi))^2 \leq \int_{x+}^{\xi} \frac{(h''_{x,z}(x, z))^2}{f_+(x, z)} \, dz.
\]
By using the same argument as in (7), we have
\[
\mathbb{E}_{x,\xi}(g(x, \xi, z) - h'_x(x, \xi))^2 \leq 2K_3.
\]
Finally, we note that,
\[
\mathbb{E}_{x,\xi}(g(x, \xi, z) - H'(x))^2 = \mathbb{E}_{x,\xi}(g(x, \xi, z) - h'_x(x, \xi))^2 + \mathbb{E}_{\xi}(h'_x(x, \xi) - H'(x))^2.
\]
Therefore, when Assumption A5(b) holds, we have \( \mathbb{E}_{x,\xi}(g(x, \xi, z) - H'(x))^2 \leq K_2^2 + 2K_3 \). Thus the proposition holds. \( \square \)

Proposition 1 shows that in the CBA, the gradient estimate \( g(x, \xi, z) \) is an unbiased estimate of the true gradient at \( x \) and can be utilized as a stochastic gradient of \( H(x) \). Note that such
an unbiased gradient is generated without accessing the sample $\xi$ itself nor the value of the objective function $h(x, \xi)$ at the sampled point. The only information used in generating the unbiased gradient is comparison information between the sample and two points.

Using the unbiased stochastic gradient, we can characterize the convergence results of the CBA in the following theorems.

**Theorem 1.** Suppose $\mu = 0$. Let $G^2$ and $\sigma^2$ be defined as in Proposition 1 and $x^*$ be any optimal solution to (1).

- If Assumption A5(a) holds, then by choosing $\eta_t = \frac{1}{\sqrt{T}}$, the CBA ensures that
  \[
  \mathbb{E}(H(\bar{x}_T) - H(x^*)) \leq \frac{(x_1 - x^*)^2}{2\sqrt{T}} + \frac{G^2}{2\sqrt{T}} \quad \text{and} \quad \sum_{t=1}^{T} \mathbb{E}(H(x_t) - H(x^*)) \leq \frac{\sqrt{T}}{2} \frac{(x_1 - x^*)^2}{2} + \frac{2}{2} G^2.
  \]

  If, in addition, $u$ and $\ell$ are finite, then by choosing $\eta_t = \frac{1}{\sqrt{T}}$, the CBA ensures that
  \[
  \mathbb{E}(H(\bar{x}_T) - H(x^*)) \leq \frac{(u - \ell)^2}{2\sqrt{T}} + \frac{G^2}{\sqrt{T}} \quad \text{and} \quad \sum_{t=1}^{T} \mathbb{E}(H(x_t) - H(x^*)) \leq \frac{\sqrt{T}}{2} (u - \ell)^2 + \sqrt{T} G^2.
  \]

- If Assumption A5(b) holds, then by choosing $\eta_t = \frac{1}{L+\sqrt{T}}$, the CBA ensures that
  \[
  \mathbb{E}(H(\bar{x}_T) - H(x^*)) \leq \frac{L + \sqrt{T}}{2T} (x_1 - x^*)^2 + \frac{H(x_1) - H(x^*)}{T} + \frac{\sigma^2}{2\sqrt{T}}, \quad \text{and} \quad \sum_{t=1}^{T} \mathbb{E}(H(x_t) - H(x^*)) \leq \frac{L + \sqrt{T}}{2} (x_1 - x^*)^2 + H(x_1) - H(x^*) + \frac{\sqrt{T} \sigma^2}{2}.
  \]

  If, in addition, $u$ and $\ell$ are finite, then by choosing $\eta_t = \frac{1}{L+\sqrt{T}}$, the CBA ensures that
  \[
  \mathbb{E}(H(\bar{x}_T) - H(x^*)) \leq \frac{(u - \ell)^2}{2\sqrt{T}} + \frac{H(x_1) - H(x^*)}{T} + \frac{L(x_1 - x^*)^2}{2T} + \frac{\sigma^2}{\sqrt{T}}, \quad \text{and} \quad \sum_{t=1}^{T} \mathbb{E}(H(x_t) - H(x^*)) \leq \frac{\sqrt{T}}{2} (u - \ell)^2 + H(x_1) - H(x^*) + \frac{L}{2} (x_1 - x^*)^2 + \sqrt{T} \sigma^2.
  \]

**Theorem 2.** Suppose $\mu > 0$. Let $G^2$ and $\sigma^2$ be defined as in Proposition 1 and $x^*$ be any optimal solution to (1).

- If Assumption A5(a) holds, then by choosing $\eta_t = \frac{1}{\mu T}$, the CBA ensures that
  \[
  \mathbb{E}(H(\bar{x}_T) - H(x^*)) \leq \frac{G^2 \log T + 1}{2\mu T} \quad \text{and} \quad \sum_{t=1}^{T} \mathbb{E}(H(x_t) - H(x^*)) \leq \frac{G^2}{2\mu} \log T + 1.
  \]

- If Assumption A5(b) holds, then by choosing $\eta_t = \frac{1}{\mu T + L}$, the CBA ensures that
  \[
  \mathbb{E}(H(\bar{x}_T) - H(x^*)) \leq \frac{\sigma^2 \log T + 1}{2\mu T} + \frac{H(x_1) - H(x^*)}{T} + \frac{L(x_1 - x^*)^2}{2T}, \quad \text{and} \quad \sum_{t=1}^{T} \mathbb{E}(H(x_t) - H(x^*)) \leq \frac{\sigma^2}{2\mu} \log T + 1 + H(x_1) - H(x^*) + \frac{L(x_1 - x^*)^2}{2}.
  \]
Theorems 1 and 2 give the performance of the CBA under different assumptions (i.e., whether $\mu = 0$ or $\mu > 0$ and whether Assumption A5(a) or A5(b) holds). The convergence rate of $O(\log T/T)$ when $\mu > 0$ is better than the convergence rate of $O\left(\frac{1}{\sqrt{T}}\right)$ when $\mu = 0$. When $\mu = 0$ and $u$ and/or $\ell$ are infinite, the stepsize $\eta_t$ is chosen to be a constant depending on the total number of iterations $T$ in order to achieve the optimal convergence rate (i.e., $\eta_t = \frac{1}{\sqrt{T}}$ when Assumption A5(a) holds and $\eta_t = \frac{1}{L + \sqrt{T}}$ when Assumption A5(b) holds). Therefore, one needs to determine the total number of iterations $T$ before running the optimization algorithm for the computation of the stepsize. When $\mu = 0$ and $u$ and $\ell$ are finite, the stepsize $\eta_t$ can be chosen as a decreasing sequence in $t$ (i.e., $\eta_t = \frac{1}{\sqrt{t}}$ when Assumption A5(a) holds and $\eta_t = \frac{1}{L + \sqrt{t}}$ when Assumption A5(b) holds). In such a case, one does not need to pre-specify the total number of iterations $T$. The convergence result when Assumption A5(a) holds is proved by Nemirovski et al. (2009) and Duchi and Singer (2009) for regular stochastic gradient. Also, the convergence result when Assumption A5(b) holds is shown by Lan (2012) for regular stochastic gradient. The detailed proofs of the theorems are given in the online Appendix (Section 7.1) for completeness.

According to Theorem 2, when $\mu > 0$, the convergence rate of CBA is $O(\log T/T)$. This convergence rate can be improved to $O(1/T)$ when $\mu > 0$ using a restarting method first proposed by Hazan and Kale (2014) or using a smart weighted averaging of solutions by Lacoste-Julien et al. (2012). In Hazan and Kale (2014), the authors show that the restarting method works when Assumption A5(a) holds. In this paper, we extend the result by showing that the restarting method can also obtain the $O(1/T)$ rate if Assumption A5(b) holds. According to Nemirovski and Yudin (1983), no algorithm can achieve convergence rate better than $O(1/T)$, thus we have obtained the best possible convergence rates in those settings. We now describe the restarting method in Algorithm 2, which we will later refer to as the multi-stage comparison-based algorithm (MCBA).

**Algorithm 2** Multi-stage Comparison-Based Algorithm (MCBA)

1. Initialize the number of stages $K \geq 1$, the starting solution $\hat{x}^1$. Set $k = 1$.
2. Let $T_k$ be the number of iterations in stage $k$ and $\eta_t^k$ be the step length in iteration $t$ of CBA in stage $k$ for $t = 1, 2, \ldots, T_k$.
3. Let $\hat{x}^{k+1} = \text{CBA}(\hat{x}^k, T_k, \{\eta_t^k\}_{t=1}^{T_k})$.
4. Stop when $k \geq K$. Otherwise, let $k \leftarrow k + 1$ and go back to step 2.
5. Output $\hat{x}^{K+1}$.

We have the following theorem about the performance of Algorithm 2. The proof of the theorem is given in the Appendix.

**Theorem 3.** Suppose $\mu > 0$. Let $G^2$ and $\sigma^2$ be defined as in Proposition 1, $x^*$ be any optimal solution to (1), and $T = \sum_{k=1}^{K} T_k$ with $T_k$ defined in MCBA.

- If Assumption A5(a) holds, then by choosing $\eta_t^k = \frac{1}{2^{k+1} \mu}$ and $T_k = 2^{k+3}$, the MCBA ensures that

$$\mathbb{E}(H(\hat{x}_{K+1}) - H(x^*)) \leq \frac{16(H(\hat{x}_1) - H(x^*) + G^2/\mu)}{T}.$$
• If Assumption A5(b) holds, then by choosing $\eta_k = \frac{1}{2^{k+1}T}$ and $T_k = 2^{k+3} + 4$, the MCBA ensures that
  \[
  \mathbb{E}(H(\hat{x}_{K+1}) - H(x^*)) \leq \frac{32(H(\hat{x}_1) - H(x^*) + L(\hat{x}_1 - x^*)^2/2 + \sigma^2/\mu)}{T}.
  \]

Both Theorems 2 and 3 give the convergence rates of CBA and MCBA for strongly convex problems (i.e., $\mu > 0$). The convergence rate of MCBA is $O(1/T)$ which improves the $O(\log T/T)$ convergence rate of CBA by a factor of $\log(T)$. The intuition for this difference is that the output $\tilde{x}_T$ of CBA is the average of all historical solutions so its quality is reduced by the earlier solutions (i.e., $x_t$ with a small $t$) which are far from the optimal solution. On the contrary, MCBA restarts CBA periodically with a better initial solution (i.e., $\bar{x}_k$) for each restart. As a result, the output of MCBA is the average of historical solutions only in the last ($K$th) call of CBA which does not involve the earlier solutions and thus has a higher quality. This is the main reason for MCBA to have a better solution after the same number of iterations, or equivalently, a better convergence rate than CBA. However, CBA is easier to implement as it does not require periodic restart as needed in MCBA. Moreover, the theoretical convergence of MCBA requires strong convexity in the problem while CBA converges without strong convexity requirement (see Theorem 1).

By Theorems 1-3, we have shown that under some mild assumptions (Assumption 1), if one has access to comparison information between each sample $\xi_t$ and two points, then one can still find the optimal solution to (1), and the convergence speed is in the same order as when one can observe the actual value of the sample (or the objective value at the sampled point). One natural question is whether the same convergence result can be achieved by only having comparison information between each sample $\xi_t$ and one point. The next example gives a negative answer to this question, showing that it is impossible to always find the optimal solution in this case, even if one allows the algorithm to be a randomized one. Thus it verifies the necessity of having comparison information at two points in each iteration (for each sample).

**Example 4.** Let $h(x, \xi) = (x - \xi)^2$, $\ell = -1$ and $u = 1$. In this case, the optimization problem (1) is to find the projection of the expected value of $\xi$ onto the interval $[-1, 1]$. Suppose there are two underlying distributions for $\xi$. In the first case, $\xi$ follows a uniform distribution on $[-3, -2]$ or $[2, 3]$, each with probability 0.5. In the second case, $\xi$ follows a uniform distribution on $[-3, -2]$ or $[3, 4]$, each with probability 0.5. In the following, we denote the distributions corresponding to the first and second cases by $F_1(\cdot)$ and $F_2(\cdot)$, respectively. It is easy to verify that in the first case, the optimal solution to (1) is $x^* = 0$, while in the second case, the optimal solution to (1) is $x^* = 0.25$. And it is also easy to verify that the above settings (both cases) satisfy Assumption 1.

Now we consider any algorithm that only utilizes the comparison information between $\xi_t$ and one decision point $x_t$ in each iteration (however, the point has to be chosen between $[\ell, u]$ since we can modify $h(x, \xi)$ such that it is undefined or $\infty$ on $x \notin [\ell, u]$). Suppose the algorithm maps $x_t$ and the comparison information between $\xi_t$ and the chosen point to a distribution of $x_{t+1}$ (thus we allow randomized algorithm). Note that for any $x \in [\ell, u]$, $F_1(x) = F_2(x) = 0.5$. In other words, there is a 0.5 probability that $\xi > x$ and a 0.5 probability that $\xi < x$ no matter whether $\xi$ is drawn from $F_1(\cdot)$ or $F_2(\cdot)$. For any algorithm, the distribution of each $x_t$ will be the same under either case. Thus, no algorithm can return the optimal solution in both cases. In other words, no algorithm can guarantee to solve (1) with comparison information between each sample and only one point, even under Assumption 1.
3. Choice of $f_-$ and $f_+$

In the CBA, one important step is the specification of the two sets of density functions $f_-(x, z)$ and $f_+(x, z)$. In the last section, we only said that $f_-$ and $f_+$ need to satisfy conditions (C1)-(C3) but did not give any specific examples. Nor did we discuss what are good choices of $f_-$ and $f_+$. In this section, we address this issue by first showing a few examples of $f_-$ and $f_+$ which could be useful in practice and then discussing the effect of choices of $f_-$ and $f_+$ on the efficiency of the algorithms. We start with the following examples of choices of $f_-$ and $f_+$.

**Example 5** (Uniform Sampling Distribution). Suppose the support of $\xi$ is known to be contained in a finite interval $[s, s]$ and the optimal decision $x^*$ is known to be within a finite interval $[\ell, u]$. (Without loss of generality, we assume $[s, s] \subset [\ell, u]$. Otherwise, we can expand $[\ell, u]$ to contain $[s, s]$.) And we assume $h''_{x,z}(x, z)$ is uniformly bounded on $[\ell, u] \times [s, s], x \neq z$. Then we can set both $f_-(x, z)$ and $f_+(x, z)$ to be uniformly distributed, i.e., for $x \in (\ell, u)$,

$$f_-(x, z) = \begin{cases} \frac{1}{\ell - s} & \ell \leq z < x \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad f_+(x, z) = \begin{cases} \frac{1}{u - s} & x < z \leq u \\ 0 & \text{otherwise} \end{cases}.$$  

When $x = \ell$, we can set $f_-$ to be a uniform distribution on $[\ell - 1, \ell]$; and when $x = u$, we can set $f_+$ to be a uniform distribution on $[u, u + 1]$. It is not hard to verify that this set of choices satisfies conditions (C1)-(C3).

**Example 6** (Exponential Sampling Distribution). Suppose the support of $\xi$ is $\mathbb{R}$ or unknown, and $\xi$ follows a light tail distribution (more precisely, there exists a constant $\lambda > 0$ such that $\lim_{t \to \infty} \lambda^t \mathbb{P}(|\xi| > t) = 0$). Moreover, $x$ is constrained on a finite interval $[\ell, u]$ and we assume $h''_{x,z}(x, z)$ is uniformly bounded on $[\ell, u] \times \mathbb{R}, x \neq z$. Then we can choose $f_-$ and $f_+$ to be exponential distributions. More precisely, we can choose

$$f_-(x, z) = \begin{cases} 0 & z \geq x \\ \lambda_- \exp(-\lambda_-(x - z)) & z < x \end{cases} \quad \text{and} \quad f_+(x, z) = \begin{cases} \lambda_+ \exp(-\lambda_+(z - x)) & z > x \\ 0 & z \leq x \end{cases}$$

where $0 < \lambda_-, \lambda_+ < \tilde{\lambda}$ are two parameters one can adjust. Apparently under this choice of $f_-$ and $f_+$, conditions (C1)-(C2) are satisfied. For (C3), we note that by the light tail assumption, there exists a constant $C$ such that $e^{-\lambda t}F(t) \leq C$ and $e^{\lambda t}(1 - F(t)) \leq C$ for all $t$. Therefore, we have

$$\int_{-\infty}^{\ell-} \frac{F(z)}{f_-(x, z)} dz = \frac{e^{\lambda_-(x)}}{\lambda_-} \int_{-\infty}^{\ell} e^{-\lambda_- z} F(z) dz \leq \frac{Ce^{\lambda_- x}}{(\lambda - \lambda_-)\lambda_-} \int_{-\infty}^{\ell} e^{(\lambda - \lambda_-)z} dz \leq \frac{Ce^{\lambda x}}{(\lambda - \lambda_-)\lambda_-},$$

$$\int_{x}^{\infty} \frac{1 - F(z)}{f_+(x, z)} dz = \frac{e^{-\lambda_+ x}}{\lambda_+} \int_{x}^{\infty} e^{\lambda_+ z} (1 - F(z)) dz \leq \frac{Ce^{-\lambda_+ x}}{(\lambda - \lambda_+)\lambda_+} \int_{x}^{\infty} e^{-(\lambda - \lambda_+)z} dz \leq \frac{Ce^{-\lambda x}}{(\lambda - \lambda_+)\lambda_+}.$$  

Combined with the uniform boundedness of $h''_{x,z}(x, z)$, condition (C3) also holds in this case.

Next we discuss the effect of choosing different $f_-$ and $f_+$ on the efficiency of the algorithm and the optimal choices of $f_-$ and $f_+$. First we note that by Theorems 1-3, the choice of $f_-$ and $f_+$ does not affect the asymptotic convergence rate of the algorithms as long as they satisfy conditions (C1)-(C3). All what they affect is the constant in the convergence results, which depends on $\mathbb{E}_{x, z}(g(x, \xi, z))^2$ or $\mathbb{E}_{x, z}(g(x, \xi, z) - H'(x))^2$ (depending on whether Assumption A5(a) or A5(b).
is a uniform distribution on $\Xi$. However, by Proposition 1, $E_{x,\xi}(g(x,\xi,z) - H'(x))^2 = E_{x,\xi}(g(x,\xi,z))^2 - (H'(x))^2$, i.e., the two terms only differ by a constant which does not depend on the choice of $h$. Therefore, in what follows, we focus on choosing $f_-$ and $f_+$ to minimize $E_{x,\xi}(g(x,\xi,z))^2$.

By (6), for any $\xi < x$, we have

$$E_x(g(x,\xi,z))^2 = 2h'_-(x)h'_z(x,\xi) - (h'_-(x))^2 + \int_{\xi}^{x^-} \frac{(h''_{x,z}(x,z))^2}{f_-(x,z)} dz.$$  

By a similar argument, for $\xi > x$, we have

$$E_x(g(x,\xi,z))^2 = 2h'_-(x)h'_z(x,\xi) - (h'_-(x))^2 + \int_{x^+}^{\xi} \frac{(h''_{x,z}(x,z))^2}{f_+(x,z)} dz.$$  

Further taking expectation over $\xi$, we have

$$E_{x,\xi}(g(x,\xi,z))^2 = 2h'_-(x)H'(x) - (h'_-(x))^2 + \int_{-\infty}^{x^-} \left( \int_{\xi}^{x^-} \frac{(h''_{x,z}(x,z))^2}{f_-(x,z)} dz \right) dF(\xi) + \int_{x^+}^{\infty} \left( \int_{x^+}^{\xi} \frac{(h''_{x,z}(x,z))^2}{f_+(x,z)} dz \right) dF(\xi).$$

By Cauchy-Schwarz inequality and condition (C1), we have

$$\int_{-\infty}^{x^-} \frac{F(z)(h''_{x,z}(x,z))^2}{f_-(x,z)} dz = \int_{-\infty}^{x^-} \frac{F(z)(h''_{x,z}(x,z))^2}{f_-(x,z)} dz \cdot \int_{-\infty}^{x^-} f_-(x,z) dz \geq \left( \int_{-\infty}^{x^-} \sqrt{F(z)h''_{x,z}(x,z)} dz \right)^2.$$  

And the equality holds only if $f_-(x,z) = C_-\sqrt{F(z)}h''_{x,z}(x,z)$, for all $z < x$ for some $C_- > 0$. Therefore, if $\sqrt{F(z)h''_{x,z}(x,z)}$ is integrable on $(-\infty, x]$, then the optimal choice for $f_-(x,z)$ is

$$f_-(x,z) = \frac{\sqrt{F(z)h''_{x,z}(x,z)}}{\int_{-\infty}^{x^-} \sqrt{F(z)h''_{x,z}(x,z)} dz}, \quad \forall z < x. \quad (8)$$

Similarly, if $\sqrt{1 - F(z)h''_{x,z}(x,z)}$ is integrable on $[x, \infty)$, then the optimal choice for $f_+(x,z)$ is

$$f_+(x,z) = \frac{\sqrt{1 - F(z)h''_{x,z}(x,z)}}{\int_{x}^{\infty} \sqrt{1 - F(z)h''_{x,z}(x,z)} dz}, \quad \forall z > x. \quad (9)$$

Now we use an example to illustrate the above results. Suppose $h(x,\xi) = (x-\xi)^2$ and $F(z)$ is a uniform distribution on $[a,b]$. Then the optimal choice of $f_-$ and $f_+$ are

$$f_-(x,z) = \frac{3(z-a)^{1/2}}{2(x-a)^{3/2}}, \quad \forall a \leq z < x \leq b \quad \text{and} \quad f_+(x,z) = \frac{3(b-z)^{1/2}}{2(b-x)^{3/2}}, \quad \forall a \leq x < z \leq b.$$  

Similarly, when $h(x,\xi) = (x-\xi)^2$ and $F(z)$ is a normal distribution $\mathcal{N}(a,b^2)$, the optimal choice of $f_-$ and $f_+$ are (it is easy to show that the integrals on the bottom are finite)

$$f_-(x,z) = \frac{\sqrt{\Phi\left(\frac{z-a}{b}\right)}}{\int_{-\infty}^{x} \sqrt{\Phi\left(\frac{z-a}{b}\right)} dz}, \quad \forall z < x \quad \text{and} \quad f_+(x,z) = \frac{\sqrt{1 - \Phi\left(\frac{z-a}{b}\right)}}{\int_{x}^{\infty} \sqrt{1 - \Phi\left(\frac{z-a}{b}\right)} dz}, \quad \forall z > x.$$
where $\Phi(\cdot)$ is the c.d.f. of a standard normal distribution.

However, in many cases, the optimal choice of $f_-$ and $f_+$ may not exist due to either 1) $\sqrt{F(z)}|h''_{x,z}(x,z)|$ or $\sqrt{1-F(z)}|h''_{x,z}(x,z)|$ is not integrable, or 2) the integration is 0 (e.g., in the case when $h(x,\xi)$ is piecewise linear in $x$ and $\xi$). In those cases, either the optimal $f_-$ and $f_+$ are not attainable, or the choice of $f_-$ and $f_+$ does not matter (e.g., in the piecewise linear case$^6$). Moreover, finding the optimal $f_-$ and $f_+$ essentially needs the knowledge of the distribution of $\xi$, which is not known in advance. Therefore, one can only use an approximate (or prior) distribution of $\xi$ to calculate the distribution.$^7$ In addition, sampling from the distributions described above often involves much more computational efforts than sampling from a uniform distribution or an exponential distribution, the overhead of which may well overshadow the improvement of the convergence speed. Therefore, in practice, choosing a heuristic sampling distribution $f_-$ and $f_+$ may be more preferable, such as the uniform or exponential distributions described earlier in this section. Indeed, as we will see in the next section, using uniform or exponential distributions lead to efficient solutions in our test cases.

4. Extensions

In this section, we discuss a few extensions of our comparison-based algorithms. In particular, in Section 4.1, we extend our discussions to high dimensional problems with quadratic objective functions. In Section 4.2, we consider the case in which the objective function is not a convex function. In Section 4.3, we consider the case in which multiple samples can be drawn and multiple comparisons can be conducted in each iteration. We also consider a case in which categorical results instead of binary results can be obtained from each comparison in the Appendix. We show that our proposed ideas can still be applied in those settings (with some variations).

4.1 Multi-Dimensional Convex Quadratic Problem

In this section, we extend our model to a multi-dimensional convex quadratic problem and propose a stochastic optimization algorithm for such a setting based on comparison information. Specifically, we consider the following stochastic convex optimization problem

$$\min_{x \in \mathcal{X}} H(x) = \mathbb{E}_\xi \left[ h(x,\xi) := \frac{1}{2} (x-\xi)^\top Q (x-\xi) \right], \quad (10)$$

where $x \in \mathbb{R}^d$, $\xi \in \mathbb{R}^d$ is a random variable, $Q$ is a positive definite matrix and $\mathcal{X}$ is a closed convex set in $\mathbb{R}^d$. We denote the gradient of $H$ by $\nabla H(x) := \mathbb{E} Q (x-\xi)$, the directional derivative of $h(x,\xi)$ along a direction $u \in \mathbb{R}^d$ by $\nabla_u h(x,\xi) := u^\top Q (x-\xi)$, and the gradient of $h(x,\xi)$ with respect to $x$ by $\nabla h(x,\xi) := Q (x-\xi)$. The following assumption is made in this section:

**Assumption 2.** There exists a constant $K_4$ such that $\mathbb{E} \|\xi - \mathbb{E} \xi\|_2^2 \leq K_4^2$.

This assumption simply requires that $\xi$ has a finite variance, which is not hard to satisfy in practice.

---

$^6$In the piecewise linear case, the comparison information between $x$ and $\xi$ will imply the knowledge of the stochastic gradient at the current sample point, and the gradient does not depend on the choice of $f_-$ and $f_+$ functions.

$^7$Such issues are common in variance reduction problems in simulation. The optimal choice to reduce variance often relies on the knowledge of the underlying distribution. See, e.g., Asmussen and Glynn (2007).
Suppose we can generate a random vector \( u \) in \( \mathbb{R}^d \) that satisfies \( \mathbb{E}(uu^\top) = I_d \), where \( I_d \) is the \( d \times d \) identity matrix. We will have \( \mathbb{E}_u u \nabla_h (x, \xi) = \mathbb{E}_u uu^\top Q(x - \xi) = \nabla h(x, \xi) \). Hence, to construct an unbiased stochastic gradient for \( H(x) \), we only need to construct an unbiased stochastic estimation for \( u^\top Q(x - \xi) \) and multiply it to \( u \). In the following, we show that this can be done by first comparing \( x + zu \) and \( x - zu \) (in the value of \( h(\cdot, \xi) \)) with a random positive number \( z \) and then comparing the better one between \( x + zu \) and \( x - zu \) with \( x \). In other word, we can still construct a stochastic gradient for \( H(x) \) in (10) using two comparisons.

Similar to Example 1, \( x \) may represent \( d \) features (e.g., size, taste, etc.) of a product while \( \xi \) is the preference of a random customer on all features. The firm would like to find the optimal features to minimize the expected customer dissatisfaction which is measured by \( h(x, \xi) \) in (10). Suppose the current product is \( x \). To implement the two comparisons above, the firm can generate a random change \( u \) and a random level \( z \) and then ask a customer for his/her preference between two new products \( x + zu \) and \( x - zu \) and his/her preference between the better new product and the current product. We call this method comparison-based algorithm for quadratic problem (CBA-QP) and provide its details in Algorithm 3. In CBA-QP, we need to specify a density function \( f(z) \) on \([0, +\infty)\), which needs to satisfy the following condition.

- **(C4)** There exists a constant \( K_5 \) such that for any \( x \in \mathcal{X} \),

\[
\int_0^{+\infty} \frac{\text{Prob} \left( \frac{\lambda_{\text{max}}(Q) \|x - \xi\|_2}{\lambda_{\text{min}}(Q) \sqrt{d}} \geq z \right)}{f(z)} dz \int_0^{\lambda_{\text{max}}(Q) \|x - \xi\|_2} \frac{1}{f(z)} dz dF(\xi) \leq K_5
\]

where \( \lambda_{\text{max}}(Q) \) and \( \lambda_{\text{min}}(Q) \) are the largest and smallest eigenvalues of \( Q \), respectively.

Below we give two examples of choices of \( f(z) \) such that it satisfies condition (C4).

**Example 7** (Uniform Sampling Distribution). Suppose \( \Xi \) (the support of \( \xi \)) and the feasible set \( \mathcal{X} \) are both bounded. The quantity \( R := \frac{\lambda_{\text{max}}(Q)}{\lambda_{\text{min}}(Q) \sqrt{d}} \max_{x \in \mathcal{X}, \xi \in \Xi} \|x - \xi\|_2 \) is finite. We can set \( f(z) \) to be the density function of a uniform distribution on \([0, R] \), i.e.,

\[
f(z) = \begin{cases} \frac{1}{R} & 0 \leq z \leq R \\ 0 & \text{otherwise.} \end{cases}
\]

With this choice, we have

\[
\int_0^{\lambda_{\text{max}}(Q) \|x - \xi\|_2} \frac{1}{f(z)} dz dF(\xi) \leq R^2,
\]

so Assumption 2 is satisfied with \( K_5 = R^2 \).

**Example 8** (Exponential Sampling Distribution). Suppose \( \xi \) follows a light tail distribution with \( \mathbb{E}\exp(\|\xi\|_2) \leq \bar{\sigma} \) for some \( \bar{\sigma} > 0 \). Moreover, suppose the feasible set \( \mathcal{X} \) is bounded. Then we can choose \( f(z) \) to be an exponential distribution, i.e.,

\[
f(z) = \begin{cases} 0 & z < 0 \\ \lambda \exp(-\lambda z) & z \geq 0, \end{cases}
\]

15
thus Assumption 2 is satisfied with $K_5 = \frac{\lambda}{\sigma x_2}$.

By the light tail assumption and the concavity of $\mathcal{CBA}$ for Quadratic Problem (CBA-QP):

\[\begin{align*}
\int_0^{\lambda_{\max}(Q)x - \lambda_{\min}(Q)v} & \frac{1}{f(z)} dz \\
& \leq \frac{1}{\lambda^2} \exp \left( \frac{\lambda_{\max}(Q)\|x - \xi\|_2^2}{\lambda_{\min}(Q)^2} \right) \leq \frac{1}{\lambda^2} \exp \left( \frac{\lambda_{\max}(Q)\|x\|_2^2}{\lambda_{\min}(Q)^2} \right) \exp \left( \frac{\lambda\|\xi\|_2^2}{\sigma} \right).
\end{align*}\]

thus Assumption 2 is satisfied with $K_5 = \frac{\lambda}{\sigma x_2} \exp \left( \frac{\lambda_{\max}(Q)x}{\lambda_{\min}(Q)^2} \right)$.

**Algorithm 3 CBA for Quadratic Problem (CBA-QP):**

1. **Initialization.** Set $t = 1$, $x_1 \in \mathcal{X} \subset \mathbb{R}^d$. Define $\eta_t$ for all $t \geq 1$. Set the maximum number of iterations $T$. Choose a density function $f(z)$ on $[0, +\infty)$. Let $\mathcal{Q}$ be the uniform distribution on a sphere in $\mathbb{R}^d$ with radius $\sqrt{d}$. Note that $\mathbb{E}uu^T = I_d$ for $u$ following distribution $\mathcal{Q}$.

2. **Main iteration.** Sample $u_t$ from $\mathcal{Q}$. Sample $z_t$ from $f(z)$
   
   (a) If $h(x_t + z_t u_t, \xi_t) < h(x_t - z_t u_t, \xi_t)$, set
   \[g(x_t, \xi_t, u_t, z_t) = \begin{cases} 0, & \text{if } h(x_t + z_t u_t, \xi_t) > h(x_t, \xi_t), \\ -\frac{1}{2} \frac{u_t^\top Qu_t}{f(z_t)} u_t, & \text{if } h(x_t + z_t u_t, \xi_t) \leq h(x_t, \xi_t). \end{cases}\]
   (b) If $h(x_t + z_t u_t, \xi_t) \geq h(x_t - z_t u_t, \xi_t)$, set
   \[g(x_t, \xi_t, u_t, z_t) = \begin{cases} 0, & \text{if } h(x_t - z_t u_t, \xi_t) > h(x_t, \xi_t), \\ \frac{1}{2} \frac{u_t^\top Qu_t}{f(z_t)} u_t, & \text{if } h(x_t - z_t u_t, \xi_t) \leq h(x_t, \xi_t). \end{cases}\]

   Let
   \[x_{t+1} = \text{Proj}_\mathcal{X} \left( x_t - \eta_t g(x_t, \xi_t, u_t, z_t) \right). \quad (11)\]

3. **Termination.** Stop when $t \geq T$. Otherwise, let $t \leftarrow t + 1$ and go back to Step 2.

4. **Output.** $\mathcal{CBA-QP}(x_1, T, \{\eta_t\}_{t=1}^T) = \bar{x}_T = \frac{1}{T} \sum_{t=1}^T x_t$.

The proposition below gives some properties of the stochastic gradient $g(x_t, \xi_t, u_t, z_t)$ in CBA-QP (see Algorithm 3):

**Proposition 2.** Let $z$, $u$, and $g$ be defined in Algorithm 3. Then the following properties hold.

1. $\mathbb{E}_{z,u} g(x, \xi, u, z) = \nabla h(x, \xi)$, for all $x \in \mathcal{X}$.
2. \( \mathbb{E}_{z,u} g(x, \xi, u, z) = \nabla H(x), \) for all \( x \in \mathcal{X} \).

3. \( \mathbb{E}_{z,u,\xi} \| g(x, \xi, u, z) - \nabla H(x) \|^2 \leq \sigma^2 \triangleq \lambda_{\max}(Q)^2 K_2^2 + \frac{\lambda_{\max}(Q)^2 d^3 K_5}{4} \).

**Proof of Proposition 2.** First, we consider the case when \( h(x + zu, \xi) < h(x - zu, \xi) \), which is equivalent to \( u^T Q(x - \xi) < 0 \) by the definition of \( h \) and the non-negativity of \( z \). Note that \( h(x + zu, \xi) \leq h(x, \xi) \) if and only if \( zu^T Q(x - \xi) + \frac{z^2}{2} u^T Qu \leq 0 \), or equivalently, \( 0 \leq z \leq -\frac{2u^T Q(x-\xi)}{u^T Qu} \). It then follows from the definition of \( g \) that

\[
\mathbb{E}_z g(x, \xi, u, z) = \int_0^{\frac{2u^T Q(x-\xi)}{u^T Qu}} 2\frac{u^T Qu}{u^T Qu} \cdot u d\sigma = u^T Q(x - \xi) u = \nabla_u h(x, \xi) u.
\]

Similarly, the second case \( h(x + zu, \xi) \geq h(x - zu, \xi) \) occurs when \( u^T Q(x - \xi) \geq 0 \). The inequality \( h(x - zu, \xi) \leq h(x, \xi) \) holds if and only if \(-zu^T Q(x - \xi) + \frac{z^2}{2}u^T Qu \leq 0 \), or equivalently, \( 0 \leq z \leq \frac{2u^T Q(x-\xi)}{u^T Qu} \). It then follows again from the definition of \( g \) that

\[
\mathbb{E}_z g(x, \xi, u, z) = \int_0^{\frac{2u^T Q(x-\xi)}{u^T Qu}} \frac{1}{2} u^T Qu \cdot u d\sigma = u^T Q(x - \xi) u = \nabla_u h(x, \xi) u.
\]

Therefore, in both cases, we have \( \mathbb{E}_z g(x, \xi, u, z) = \nabla_u h(x, \xi) u \). Since \( \mathbb{E}(uu^T) = I_d \), further taking expectation over \( u \) on both sides of this equality gives the first conclusion of the proposition.

The second conclusion can be obtained by taking expectation over \( \xi \) on both sides of the first conclusion, namely, \( \mathbb{E}_{z,u,\xi} g(x, \xi, u, z) = \mathbb{E}_\xi \mathbb{E}_{z,u} g(x, \xi, u, z) = \mathbb{E}_\xi \nabla h(x, \xi) = \nabla H(x) \). (This holds because \( h \) is a convex function, thus one can apply the monotone convergence theorem.)

Next, we prove the third conclusion. The first conclusion implies that

\[
\mathbb{E}_{u,z} \| g(x, \xi, u, z) - \nabla h(x, \xi) \|^2 = \mathbb{E}_{u,z} \| g(x, \xi, u, z) \|^2 - \| \nabla h(x, \xi) \|^2.
\]

If \( h(x + zu, \xi) < h(x - zu, \xi) \), by the definition of \( g \), we can show that

\[
\mathbb{E}_z \| g(x, \xi, u, z) \|^2 = \int_0^{\frac{2u^T Q(x-\xi)}{u^T Qu}} \frac{(u^T Qu)^2}{4f(z)} \| u \|^2 d\sigma \leq \int_0^{\lambda_{\max}(Q) \| u \|_2 \| x-\xi \|_2} \frac{\lambda_{\max}(Q)^2 \| u \|_2^4}{4f(z)} d\sigma \leq \frac{\lambda_{\max}(Q)^2 d^3}{4} \int_0^{\lambda_{\min}(Q) \| x-\xi \|_2} \frac{1}{f(z)} d\sigma,
\]

where the inequality is by Cauchy-Schwarz inequality and the second equality is because \( \| u \| = \sqrt{d} \). Similarly, if \( h(x + zu, \xi) \geq h(x - zu, \xi) \), then we can also show that

\[
\mathbb{E}_z \| g(x, \xi, u, z) - \nabla h(x, \xi) \|^2 \leq \frac{\lambda_{\max}(Q)^2 d^3}{4} \int_0^{\lambda_{\max}(Q) \| x-\xi \|_2} \frac{1}{f(z)} d\sigma.
\]

As a result, we have

\[
\mathbb{E}_{u,z} \| g(x, \xi, u, z) - \nabla h(x, \xi) \|^2 \leq \frac{\lambda_{\max}(Q)^2 d^3}{4} \int_0^{\lambda_{\max}(Q) \| x-\xi \|_2} \frac{1}{f(z)} d\sigma.
\]
Therefore,

\[ \mathbb{E}_z, u, \xi \| g(x, \xi, u, z) - \nabla h(x, \xi) \|^2 \leq \frac{\lambda_{\max}(Q)^2 d^3}{4} \int \int \frac{\lambda_{\min}(Q) \| x - \xi \|^2}{f(z) d z d F(\xi)} \leq \frac{\lambda_{\max}(Q)^2 d^3 k_5}{4}. \]

In addition, by Assumption 2, we have

\[ \mathbb{E}_\xi \| \nabla h(x, \xi) - \nabla H(x) \|^2 = \mathbb{E}_\xi \| Q(x - \xi) - Q(x - \mathbb{E}(\xi)) \|^2 \leq \lambda_{\max}(Q)^2 \mathbb{E}_\xi \| \xi - \mathbb{E}(\xi) \|^2 \leq \lambda_{\max}(Q)^2 k_4. \]

Finally, we note that

\[ \mathbb{E}_z, u, \xi \| g(x, \xi, u, z) - \nabla h(x, \xi) \|^2 = \mathbb{E}_z, u, \xi \| g(x, \xi, u, z) - \nabla h(x, \xi) \|^2 + \mathbb{E}_\xi \| \nabla h(x, \xi) - \nabla H(x) \|^2. \]

Therefore, \( \mathbb{E}_z, u, \xi \| g(x, \xi, u, z) - \nabla H(x) \|^2 \leq \sigma^2 := \lambda_{\max}(Q)^2 k_4^2 + \frac{\lambda_{\max}(Q)^2 d^3 k_5}{4}. \) Thus the proposition holds.

Based on Proposition 2, we have the following theorem about the performance of CBA-QP.

**Theorem 4.** Let \( \sigma^2 \) be defined as in Proposition 2. Let \( x^* \) be any optimal solution to (10) and \( \mu > 0 \) and \( L > 0 \) be the smallest and largest eigenvalues of \( Q \), respectively. Then by choosing \( \eta_t = \frac{1}{\mu t + L} \), the CBA-QP ensures that

\[ \mathbb{E}(H(\hat{x}_T) - H(x^*)) \leq \frac{\sigma^2}{2\mu} \log T + 1 - \frac{H(x(1)) - H(x^*))}{T} + \frac{L \| x(1) - x^* \|^2}{2T}, \] and

\[ \sum_{t=1}^{T} \mathbb{E}(H(x_t) - H(x^*)) \leq \frac{\sigma^2}{2\mu} (\log T + 1) + H(x(1)) - H(x^*) + \frac{L \| x(1) - x^* \|^2}{2}. \]

Similar to CBA, we can further improve the theoretical performance of CBA-QP using a restarting strategy as described in MCBA. We denote the resulting restarted algorithm by MCBA-QP.

**Algorithm 4** Multi-stage Comparison-Based Algorithm for Quadratic Problem (MCBA-QP)

1. Initialize the number of stages \( K \geq 1 \), the starting solution \( \hat{x}_1 \). Set \( k = 1 \).
2. Let \( T_k \) be the number of iterations in stage \( k \) and \( \eta_k \) be the step length in iteration \( t \) of CBA-QP in stage \( k \) for \( t = 1, 2, \ldots, T_k \).
3. Let \( \hat{x}_{k+1} = \text{CBA-QP}(\hat{x}_k, T_k, \{\eta_k\}_{t=1}^{T_k}) \).
4. Stop when \( k \geq K \). Otherwise, let \( k \leftarrow k + 1 \) and go back to step 2.
5. Output \( \hat{x}_{K+1} \).

**Theorem 5.** Let \( \sigma^2 \) be defined as in Proposition 2, \( \mu > 0 \) and \( L > 0 \) be defined as in Theorem 4, and \( T = \sum_{k=1}^{K} T_k \) with \( T_k \) defined in MCBA-QP. By choosing \( \eta_k = \frac{1}{2k+\mu + L} \) and \( T_k = 2^{k+3} + 4 \), the MCBA-QP ensures that

\[ \mathbb{E}(H(\hat{x}_{K+1}) - H(x^*)) \leq \frac{32(H(\hat{x}_1) - H(x^*) + L \| \hat{x}_1 - x^* \|^2/2 + \sigma^2/\mu)}{T}. \]
We omit the proofs of Theorems 4 and 5 because they are very similar to those of the second part of Theorem 2 and the second part of Theorem 3, respectively.

Although we mainly focus on a special quadratic problem for the high-dimensional case, it is easy to see that our approach can also be applied to the high dimensional cases where the objective function is separable with respect to each dimension while the constraint may not be separable. In particular, our method can be applied to the problem like

\[
\min_{x \in \mathcal{X}} H(x) = \mathbb{E}_\xi \left[ h(x, \xi) := \sum_{i=1}^d h_i(x_i, \xi_i) \right]
\]

where \( x = (x_1, \ldots, x_d)^\top \in \mathbb{R}^d, \) \( \xi = (\xi_1, \ldots, \xi_d)^\top \in \mathbb{R}^d, \) \( \mathcal{X} \) is a convex closed set in \( \mathbb{R}^d, \) and \( h_i(x_i, \xi_i) \) is a function satisfying Assumption 1. In such cases, one can apply the idea of CBA to construct an unbiased stochastic gradient of each \( h_i(x_i, \xi_i) \) based on comparisons and then concatenate them into an unbiased stochastic gradient of \( h(x, \xi) \) in order to apply the projected gradient step (11). The resulting algorithm will have the same convergence rates as CBA under each setting in Theorem 1-3. Note that, we do not assume \( \mathcal{X} \) is separable so we still cannot solve the problem above as \( d \) independent problems.

### 4.2 Non-Convex Problem

The focus of our study in this paper is to construct an unbiased stochastic gradient based on comparative information. Once the construction is done, one can also apply the stochastic gradient for non-convex stochastic optimization problem. Although the stochastic gradient method can no longer guarantee a global optimal solution for a non-convex problem, the recent result by Davis and Drusvyatskiy (2018) shows that the iterative solution generated by stochastic gradient method still converges to a nearly stationary point under some conditions.

In this section, we still consider one-dimensional problem (1) but \( H(x) \) is no longer convex. More specifically, we still assume Assumption 1 holds except that Assumption (A4) is replaced by

\[(A4') \ H(x) \text{ in (1) is differentiable and } \rho\text{-weakly convex on } [\ell, u] \text{ for some } \rho > 0, \text{ namely}, \]

\[
H(x_2) \geq H(x_1) + H'(x_1)(x_2 - x_1) - \frac{\rho}{2} (x_2 - x_1)^2, \quad \forall x_1, x_2 \in [\ell, u].
\]

Moreover, \( \mathbb{E}_\xi(h'_i(x, \xi)) = H'(x) \) for all \( x \in [\ell, u]. \)

For any \( \lambda > 0, \) the Moreau envelope for (1) is defined as a function

\[
H_\lambda(x) := \min_{\ell \leq y \leq u} \left\{ H(y) + \frac{1}{2\lambda} (x - y)^2 \right\}.
\]

By definition, as long as \( \lambda < \frac{1}{\rho}, \) the minimization problem (13) has a strongly convex objective function and has a unique solution, denoted by \( \hat{x}. \) Moreover, the function \( H_\lambda(x) \) is continuously differentiable with the gradient given by \( H'_\lambda(x) := \frac{1}{\lambda}(x - \hat{x}). \) According to Davis and Drusvyatskiy (2018), the value of \( |H'_\lambda(x)| \) measures the near stationarity of a solution \( x \in [\ell, u] \) because

\[
|x - \hat{x}| = \lambda |H'_\lambda(x)| \quad \text{and} \quad \text{dist}(0; \partial H(\hat{x})) \leq |H'_\lambda(x)|
\]

where \( \partial H(\hat{x}) \) the subdifferential of \( H \) at \( x, \) i.e., the set of all \( v \) satisfying \( H(y) \geq H(x) + v(y - x) + o(\|y - x\|) \) as \( y \to x. \) This means that if \( |H'_\lambda(x)| = \frac{1}{\lambda}|x - \hat{x}| \) is small, then the solution
Suppose Assumption (A4) is replaced by Assumption (A4') and the optimal value of (1) is finite. Let $G^2$ be defined as in Proposition 1 and $t^*$ be a random index such that $\mathbb{P}(t^* = t) = \frac{\eta_t}{\sum_{s=1}^{S} \eta_s}$ for $t = 1, 2, \ldots, T$. Then, we have $\mathbb{E}[H'_{\lambda}(x_{t^*})]$ with $\lambda = \frac{1}{2\rho}$ converges to zero in a rate of $O(\sqrt{T})$. We state this result formally as follows.

**Theorem 6** (Corollary 2.2 by Davis and Drusvyatskiy 2018). Suppose Assumption (A4) is replaced by Assumption (A4') and the optimal value of (1) is finite. Let $G^2$ be defined as in Proposition 1 and $t^*$ be a random index such that $\mathbb{P}(t^* = t) = \frac{\eta_t}{\sum_{s=1}^{S} \eta_s}$ for $t = 1, 2, \ldots, T$. By choosing $\eta_t = \frac{1}{\sqrt{T}}$, the CBA ensures that

$$\mathbb{E}[H'_{\lambda}(x_{t^*})]^2 \leq 2\frac{(H_{\lambda}(x_1) - \min_{\ell \leq u \leq T} H(x)) + \rho G^2}{\sqrt{T}}. $$

Therefore, the comparison-based algorithm can be partly applied to non-convex problems.

### 4.3 Mini-Batch Method with Additional Comparisons

In this section, we consider the scenario where multiple comparisons can be conducted in each iteration. In such cases, a mini-batch technique can be implemented in CBA and MCBA to reduce the noise in stochastic gradient and improve the performance of the algorithms. In particular, we still compare $\xi_t$ and $x_t$ in iteration $t$ in Algorithm 1. In case (a) where $\xi_t < x_t$, we generate $S$ independent samples, denoted by $z^s_t$ for $s = 1, 2, \ldots, S$, from a distribution on $(-\infty, x_t]$ with p.d.f. $f_-(x_t, z)$ and construct a stochastic gradient $\tilde{g}(x_t, \xi_t, z^s_t)$ as in (3) with $z_t$ replaced by $z^s_t$. Similarly, in case (b) where $\xi_t > x_t$, we generate $z^s_t$ from a distribution on $[x_t, +\infty)$ with p.d.f. $f_+(x_t, z)$ and construct $g(x_t, \xi_t, z^s_t)$ as in (4) with $z_t$ replaced by $z^s_t$. After obtaining $g(x_t, \xi_t, z^s_t)$ in either case, we replace (5) in CBA with the following two steps

$$\tilde{g}_t = \frac{1}{S} \sum_{s=1}^{S} g(x_t, \xi_t, z^s_t)$$

$$x_{t+1} = \text{Proj}_{[\ell, u]}(x_t - \eta_t \tilde{g}_t) = \max (\ell, \min (u, x_t - \eta_t g_t))$$

Here, $\tilde{g}_t$ is the average gradient constructed by a mini-batch, which satisfies $\mathbb{E}[z^s_t, s = 1, \ldots, S, \xi_t, \tilde{g}_t] = H'(x_t)$ by conclusion 2 in Proposition 1 and has a smaller noise than $g(x_t, \xi_t, z_t)$. MCBA can also benefit from this technique by calling CBA after the aforementioned modification.

This mini-batch technique will not improve the asymptotic convergence rate of CBA and MCBA in the dependency on the number of iterations $T$ because it will not completely eliminate the noise in the stochastic gradient. However, by reducing the noise, it will improve the algorithm’s performance in practice as we demonstrate in Section 5.3.8

---

8Note that in the mini-batch method, the multiple samples are drawn simultaneously. It is worth noting that it is also possible to draw multiple samples sequentially, each one depending on the results of all previous ones. However, that mechanism will be quite complex. Moreover, the asymptotic performance will not improve because it will not surpass the asymptotic performance when the samples can be directly observed (which is already achieved by our current algorithm with two comparisons). Therefore, we here only choose to present the mini-batch approach.
5. Numerical Tests

In this section, we conduct numerical experiments to show that although much less information is used, the proposed algorithms based only on comparison information converge at the same rate as the stochastic gradient methods. We will also investigate the impact of choices of \( f_- \) and \( f_+ \) and the impact of mini-batch on the performances of the proposed methods. We implemented all algorithms in Matlab running on a 64-bit Microsoft Windows 10 machine with a 2.70 Ghz Intel Core i7-6820HQ CPU and 8GB of memory.

5.1 Convergence of Objective Value

In this section, we conduct numerical experiments to test the performance of the CBA and the MCBA. We consider two objective functions:

1) \( h_1(x, \xi) = (x - \xi)^2 \) and 2) \( h_2(x, \xi) = \begin{cases} \frac{(x - \xi)^2 + (x - \xi)}{(x - \xi)^2 + 2(\xi - x)} & \text{if } \xi < x \\ \frac{2(x - \xi)^2 + 2(\xi - x)}{(x - \xi)^2 + 2(\xi - x)} & \text{if } \xi \geq x \end{cases} \)

Here the two objective functions correspond to the two examples we described in the beginning with \( h_1 \) being smooth but not \( h_2 \). For each choice of \( h(x, \xi) \), we consider two distributions of \( \xi \), a uniform distribution \( U[50, 150] \) and a normal distribution \( N(100, 100) \). Thus we have four cases in total. It is easy to see that for the first objective function, the optimal solution under either underlying distribution is \( x^* = 100 \). For the second objective function, the optimal solution under the uniform distribution is \( x^* = 108.66 \) while the optimal solution under the normal distribution is \( x^* = 102.82 \). In all experiments, we choose the feasible set to be \([\ell, u] = [50, 150] \).

For the cases where \( \xi \sim U[50, 150] \), we choose \( f_- \) and \( f_+ \) to be uniform distributions as in Example 5 with \( \ell = 50 \) and \( u = 150 \). For the cases where \( \xi \sim N(100, 100) \), we choose \( f_- \) and \( f_+ \) to be exponential distributions as in Example 6 with \( \lambda_+ = \lambda_- = 2^{-4} = 0.0625 \). (Later in Section 5.2 we will test the effect of choosing different \( f_- \) and \( f_+ \) on the convergence speed of the algorithms.)

In each of the four settings above, Assumption 1 holds with both A5(a) and A5(b) satisfied, and \( H(x) \) is strongly convex. To compare the performance of the CBA with or without utilizing the strong convexity of the problem, we test both step lengths \( \eta_t = 1/\sqrt{T} \) and \( \eta_t = 1/(\mu t) \) for the CBA as suggested by Theorem 1 and Theorem 2. For the MCBA, we use the step sizes \( \eta^k_t = 1/(2^{k+1} \mu) \) and \( T_k = 2^{k+3} \) as suggested by Theorem 3. We choose \( \mu = 0.5 \) in all settings.

In the following, we compare the CBA and the MCBA with the standard stochastic gradient descent (SGD) method (see Nemirovski et al. 2009, Duchi and Singer 2009). In the standard SGD method, it is assumed that \( \xi_t \) can be observed directly and the stochastic gradient update step \( x_{t+1} = \text{Proj}_{[\ell, u]}(x_t - \eta_t h'_x(x_t, \xi_t)) \) is performed in each iteration. In the experiments, we apply the same step lengths in the SGD method as in the CBA. In all tests, we start from a random initial point \( x_1 \sim U[50, 150] \) and run each algorithm for \( T = 500 \) iterations and we report the average relative optimality gap \( \delta_t = \frac{H(x_t) - H(x^*)}{H(x^*)} \) over 2000 independent trials for each \( t \). The results are reported in Figure 1.

In Figure 1, the \( x \)-axis represents the number of iterations and the \( y \)-axis represents the average relative optimality gap for each algorithm. The curves \textbf{CBA} and \textbf{SGD} represent the results of CBA and SGD with \( \eta_t = 1/\sqrt{T} \) respectively while the curves \textbf{CBAstc} and \textbf{SGDstc} represent the results of CBA and SGD with \( \eta_t = 1/(\mu t) \) respectively. The curve \textbf{MCBA} represents the results of the MCBA algorithm. In each of the curves, the bar at each point represents the standard error of the corresponding \( \delta_t \). As one can see, the standard errors are...
Figure 1: The convergence of the average relative optimality gap for different algorithms for the four instances. First row: \( h_1(x, \xi) \); Second row: \( h_2(x, \xi) \). First column: \( \xi \sim U[50, 150] \); Second column: \( \xi \sim N(100, 100) \).

fairly small. Thus the test results are quite stable in these numerical experiments. We also present the average computation time of all algorithms in Table 1.

| Algorithm | \( h_1(x, \xi) \) | \( h_2(x, \xi) \) |
|-----------|-----------------|-----------------|
| \( \xi \sim U[50, 150] \) | \( \xi \sim N(100, 100) \) | \( \xi \sim U[50, 150] \) | \( \xi \sim N(100, 100) \) |
| SGD       | 0.008           | 0.007           | 0.010           | 0.040           |
| SGDstc    | 0.007           | 0.007           | 0.009           | 0.038           |
| CBA       | 0.011           | 0.018           | 0.013           | 0.049           |
| CBAstc    | 0.011           | 0.017           | 0.013           | 0.045           |
| MCBA      | 0.013           | 0.021           | 0.017           | 0.055           |

Table 1: The computation time (in seconds) of different algorithms for 500 iterations for the four instances.

From the results shown in Figure 1, we can see that all of CBA, CBAstc and MCBA converge quite fast in these problems. Even though they use much less information than the SGD and the SGDstc methods, it takes only about twice as many iterations to get the same accuracy. As shown in Table 1, the computation time for 500 iterations is less than 0.1 seconds in each algorithm and the time is not much different across different algorithms. (This short runtime is because of the low dimensionality of the problems.) Moreover, in our tests, CBA, CBAstc
and MCBA have quite similar performance despite their different theoretical guarantees.\footnote{Note that in Figure 1, the CBA and the CBAsct sometimes perform even better than the MCBA despite the worse theoretical guarantee. In fact, this is common in convex optimization literature for such types of (restarting) methods. For example, Chen et al. (2012) proposed a stochastic gradient method called MORDA, which improves ORDA in the same paper in theoretical convergence rate using a similar restarting technique to our MCBA method. However, in the fourth column of Table 2 in Chen et al. (2012), the objective value in MORDA is higher than that in ORDA. Similarly, Lan and Ghadimi (2012) developed a stochastic gradient method called Multistage AC-SA, which improves AC-SA in theoretical convergence rate but not necessarily in numerical performance (see Table 4.3 in Lan and Ghadimi 2012).}

### 5.2 Choices of $f_-$ and $f_+$

In this section, we perform some additional tests to study the impact of different choices of $f_-$ and $f_+$ on the performance of the CBA. We still consider the two objective functions considered in the last section, but we focus on the case in which $\xi \sim \mathcal{N}(100, 100)$. First we keep $f_-$ and $f_+$ to be exponential distributions (as in Example 6) and see how the performance is affected by different values of $\lambda_- = \lambda_+ = \lambda$. The results are shown in Figure 2.

![Figure 2: The impact of $\lambda$ in $f_+$ and $f_-$ to different algorithms, measured by the average relative optimality gap after 500 iterations. Left: $h_1(x, \xi)$; Right: $h_2(x, \xi)$. In both cases, $\xi \sim \mathcal{N}(100, 100)$.](image)

In Figure 2, the $x$-axis represents the value of $\log_2 \lambda$ while the $y$-axis represents the relative optimality gap $\delta_T$ after 500 iterations evaluated as the average value of 2000 independent trials (again the bars show the standard error in these trials). The influences of different values of $\lambda$ in $f_-$ and $f_+$ are presented for the CBA, the CBAsct and the MCBA algorithms. We can see that the value of $\lambda$ does influence the convergence speed of the algorithms. Particularly, in both figures in Figure 2, the optimality gap after $T = 500$ iterations decreases first as $\lambda$ increases but starts to increase when $\lambda$ is large. Moreover, the influence is relatively small when $\lambda$ is small but is large when $\lambda$ is large. And the influence is more pronounced for the CBAsct algorithm. In our setting, the best performance is obtained around $\lambda = 2^{-4} = 0.0625$ for all algorithms.

In Section 3, we provided the optimal choice of $f_-$ and $f_+$ in (8) and (9). In Figure 3, we present the difference in the performances of the CBA when $f_-$ and $f_+$ are chosen optimally versus when they are chosen as exponential distributions with $\lambda_- = \lambda_+ = 0.0625$. In this experiment, we choose $h(x, \xi) = h_1(x, \xi)$ and run the CBA for 500 iterations and compute the average relative optimality gap $\delta_{500}$ over 2000 independent trials. The results are plotted in Figure 3.
Figure 3: The convergence of the average relative optimality gap in CBA when choosing $f_+$ and $f_-$ to be the exponential distribution and the optimal distribution in (8) and (9). Left: $\xi \sim U[50, 150]$; Right: $\xi \sim \mathcal{N}(100, 100)$. In both cases, $h(x, \xi) = (x - \xi)^2$.

In Figure 3, we can see that the optimal choice of $f_-$ and $f_+$ does improve the performance of the CBA, which confirms our analysis in Section 3. However, the improvement is not essential yet generating samples from the optimal distribution is much more time consuming. For example, when $\xi \sim \mathcal{N}(100, 100)$, the computational time is less than 0.05 seconds when using exponential distribution but about 1 second when using the optimal distribution. Therefore, as we discussed in the end of Section 3, one can just choose a simple distribution in practice.

5.3 Mini-Batch Method with Additional Comparisons

In this section, we numerically test how the performance of CBA depends on the sample size $S$ in the mini-batch technique described in Section 4.3. The instances and the choice of parameters are all identical to Section 5.1. We present the convergence of CBA with $S = 1, 2, 5, 10, 100$ in Figure 4 and the associated runtimes in Table 2. According to the figures, one additional comparison (i.e. $S = 2$) with $z$ can improve the convergence of Algorithm 1 in all four instances. Although increasing $S$ can still improve the performance further, the effect diminishes quickly. This is because the noise in the stochastic gradient is generated from the sample noise of both $\xi$ and $z$. The mini-batch technique for sampling $z$ does not help reduce the noise due to $\xi$, which eventually dominates the noise due to $z$ when $S$ is large enough. Although a similar mini-batch technique can be applied to $\xi$, it might not be practical when applying our algorithm in practice. For example, when $\xi$ represents the ideal product of a customer, creating a mini-batch for $\xi$ means asking different customers’ preference without updating the solution, which is not consistent with the setting of online optimization where the solution is updated after the visit of each customer.

6. Concluding Remarks

In this paper, we considered a stochastic optimization problem when neither the underlying uncertain parameters nor the objective value at the sampled point can be observed. Instead, the decision maker can only access to comparison information between the sample point and two chosen points in each iteration. We proposed an algorithm that gives unbiased gradient estimates for this problem, which achieves the same asymptotic convergence rate as standard
Figure 4: The convergence of the average relative optimality gap in CBA when using different numbers of comparisons \((S)\). First row: \(h_1(x, \xi)\); Second row: \(h_2(x, \xi)\). First column: \(\xi \sim \mathcal{U}[50, 150]\); Second column: \(\xi \sim \mathcal{N}(100, 100)\).

Table 2: The computation time (in seconds) of CBA for 500 iterations when using different numbers of comparisons \((S)\) per iteration.

| \(S\) | \(\xi \sim \mathcal{U}[50, 150]\) | \(\xi \sim \mathcal{N}(100, 100)\) | \(\xi \sim \mathcal{U}[50, 150]\) | \(\xi \sim \mathcal{N}(100, 100)\) |
|---|---|---|---|---|
| 1 | 0.011 | 0.018 | 0.013 | 0.049 |
| 2 | 0.015 | 0.022 | 0.020 | 0.053 |
| 5 | 0.021 | 0.025 | 0.022 | 0.057 |
| 10 | 0.026 | 0.033 | 0.029 | 0.064 |
| 100 | 0.117 | 0.158 | 0.125 | 0.175 |

stochastic gradient methods. Numerical experiments demonstrate that our proposed algorithm is efficient.

There is one remark we would like to make. In this paper, we assumed that \(\xi\) follows a continuous distribution. However, we only need that in each iteration, the probability that \(\xi_t = x_t\) is 0. This can be guaranteed by only requiring \(\mathbb{P}(\xi = \ell) = \mathbb{P}(\xi = u) = 0\), and then in the CBA, adding a small and decaying random perturbation to \(g(x_t, \xi_t, z_t)\) (for example, a uniform distribution on \([-1/2^t, 1/2^t]\) in iteration \(t\)). By doing this, one can still use the same
There are several future directions of research. First, for multi-dimensional problems, we only considered convex quadratic problems in this paper (and as mentioned in the previous paragraph, we can also generalize our method to separable objective function cases). It would be of interest to see whether the ideas and techniques can be generalized to more general high-dimensional settings. Second, in this paper, we assumed that the distribution of $\xi$ is stationary over time, and the comparison information is always reported accurately. However, in practice, the distribution of $\xi$ may change over time or the comparison information may be reported in a noisy fashion. It would be interesting to see whether we could extend our discussions to consider such situations. Finally, our paper only considers a continuous decision setting, i.e., we assumed that all the decision variables as well as the test variables can take any continuous values. In many practical situations, the decision variables may only be chosen from a finite set. It is worth further research to see whether a similar idea can be applied in such settings.

References

Asmussen, S., P. W. Glynn. 2007. *Stochastic Simulation*. Springer.

Bensoussan, A., M. Çakanyildirim, S. P. Sethi. 2007. A multiperiod newsvendor problem with partially observed demand. *Mathematics of Operations Research* 32(2) 322–344.

Besbes, O., A. Muharremoglu. 2013. On implications of demand censoring in the newsvendor problem. *Management Science* 59(6) 1407–1424.

Cachon, G., R. Swinney. 2011. The value of fast fashion: Quick response, enhanced design, and strategic consumer behavior. *Management Science* 57(4) 579–595.

Candogan, O., K. Bimpikis, A. Ozdaglar. 2012. Optimal pricing in networks with externalities. *Operations Research* 60(4) 883–905.

Cesa-Bianchi, N., G. Lugosi. 2006. *Prediction, Learning, and Games*. Cambridge University Press.

Chen, X., Q. Lin, J. Pena. 2012. Optimal regularized dual averaging methods for stochastic optimization. *Advances in Neural Information Processing Systems* 25. 395–403.

Davis, D., D. Drusvyatskiy. 2018. Stochastic subgradient method converges at the rate $O(k^{-1/4})$ on weakly convex functions. *arXiv preprint arXiv:1802.02988*.

Ding, X., M. Puterman, A. Bisi. 2002. The censored newsvendor and the optimal acquisition of information. *Operations Research* 50(3) 517–527.

Duchi, J., E. Hazan, Y. Singer. 2011. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research* 12 2121–2159.

Duchi, J., Y. Singer. 2009. Efficient online and batch learning using forward-backward splitting. *Journal of Machine Learning Research* 10 2873–2898.

Gupta, D., B. Denton. 2008. Appointment scheduling in health care: Challenges and opportunities. *IIE Transactions* 40 800–819.
Halman, N., J. Orlin, D. Simchi-Levi. 2012. Approximating the nonlinear newsvendor and single-item stochastic lot-sizing problems when data is given by an oracle. *Operations Research* 60(2) 429–446.

Hazan, E., S. Kale. 2014. Beyond the regret minimization barrier: Optimal algorithms for stochastic strongly-convex optimization. *Journal of Machine Learning Research* 15 2489–2512.

Hu, C., J. T. Kwok, W. Pan. 2009. Accelerated gradient methods for stochastic optimization and online learning. *Advances in Neural Information Processing Systems* 22, 781–789.

Kiefer, J., J. Wolfowitz. 1952. Stochastic estimation of the maximum of a regression function. *The Annals of Mathematical Statistics* 23 (3) 462–466.

Kolker, A. 2017. *The Optimal Workforce Staffing Solutions With Random Patient Demand in Healthcare Settings*, chap. 322. IGI Global Information Science, 3711 – 3724.

Lacoste-Julien, Simon, Mark Schmidt, Francis Bach. 2012. A simpler approach to obtaining an $o\left(\frac{1}{t}\right)$ convergence rate for the projected stochastic subgradient method. *arXiv preprint arXiv:1212.2002*.

Lan, G. 2012. An optimal method for stochastic composite optimization. *Mathematical Programming* 133(1-2) 365–397.

Lan, G., S. Ghadimi. 2012. Optimal stochastic approximation algorithms for strongly convex stochastic composite optimization, Part I: A generic algorithmic framework. *SIAM Journal on Optimization* 22(4) 1469–1492.

Lan, G., S. Ghadimi. 2013. Optimal stochastic approximation algorithms for strongly convex stochastic composite optimization, Part II: Shrinking procedures and optimal algorithms. *SIAM Journal on Optimization* 23(4) 2061–2089.

Lin, Q., X. Chen, J. Pena. 2011. A sparsity preserving stochastic gradient method for composite optimization. Tech. rep., Carnegie Mellon University.

Nemirovski, A., A. Juditsky, G. Lan, A. Shapiro. 2009. Robust stochastic approximation approach to stochastic programming. *SIAM Journal on Optimization* 19(4) 1574–1609.

Nemirovski, A., D. Yudin. 1983. *Problem Complexity and Method Efficiency in Optimization*. Weily, New York.

Rakhlin, A., O. Shamir, K. Sridharan. 2012. Making gradient descent optimal for strongly convex stochastic optimization. *Proceedings of the 29th International Conference on Machine Learning*. 449–456.

Robbins, H., S. Monro. 1951. A stochastic approximation method. *The Annals of Mathematical Statistics* 22(3) 400–407.

Shalev-Shwartz, S. 2012. Online learning and online convex optimization. *Foundations and Trends in Machine Learning* 4(2) 107–194.

Shamir, O., T. Zhang. 2013. Stochastic gradient descent for non-smooth optimization: Convergence results and optimal averaging schemes. *Proceedings of the 30th International Conference on Machine Learning*. 71–79.
Shapiro, A., D. Dentcheva, A. Ruszczynski. 2014. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM.

Widder, D. V. 1990. *Advanced Calculus*. Dover Publication Inc.

Xiao, L. 2010. Dual averaging methods for regularized stochastic learning and online optimization. *Journal of Machine Learning Research* **11** 2543–2596.

Zinkevich, M. 2003. Online convex programming and generalized infinitesimal gradient ascent. *Proceedings of the 20th International Conference on Machine Learning*. 928–936.
7. Appendix

In this section, we first prove theoretical convergence properties of the proposed algorithms, and then provide the extension of CBA when the comparison result is categorical instead of just binary.

7.1 Proof of Theorems 1, 2 and 3

**Proof of Theorem 1.** We prove the theorem by considering the case when Assumption A5(a) or A5(b) holds respectively. For the ease of notation, we shall use $E_t$ to denote the conditional expectation taken over $\xi_t$ and $z_t$ conditioning on $\xi_1, z_1, \xi_2, z_2, \ldots, \xi_{t-1}, z_{t-1}$.

1. When Assumption A5(a) holds (based on Nemirovski et al. 2009, Duchi and Singer 2009):

   According to (5), we have
   
   \[
   (x_{t+1} - x^*)^2 = (\text{Proj}[\xi, u](x_t - \eta_t g(x_t, \xi_t, z_t)) - \text{Proj}[\xi, u](x^*))^2 \\
   \leq (x_t - \eta_t g(x_t, \xi_t, z_t) - x^*)^2 \\
   = (x_t - x^*)^2 - 2\eta_t g(x_t, \xi_t, z_t)(x_t - x^*) + \eta_t^2 (g(x_t, \xi_t, z_t))^2.
   \]

   Taking expectation of (14) over $\xi_t$ and $z_t$, we have
   
   \[
   E_t(x_{t+1} - x^*)^2 = (x_t - x^*)^2 - 2\eta_t H'(x_t)(x_t - x^*) + \eta_t^2 E_t(g(x_t, \xi_t, z_t))^2 \\
   \leq (1 - \eta_t \mu)(x_t - x^*)^2 - 2\eta_t (H(x_t) - H(x^*)) + \eta_t^2 E_t(g(x_t, \xi_t, z_t))^2,
   \]

   where the first equality is because of Proposition 1 and the last inequality is because of (2) ($H(\cdot)$ is $\mu$-convex). According to the third statement in Proposition 1, (15) implies that

   \[
   H(x_t) - H(x^*) \leq \frac{1 - \eta_t \mu}{2\eta_t}(x_t - x^*)^2 - \frac{1}{2\eta_t} E_t(x_{t+1} - x^*)^2 + \frac{\eta_t G^2}{2}.
   \]  

   If $\mu = 0$ and we choose $\eta_t = \frac{1}{\sqrt{T}}$, then summing (16) for $t = 1, 2, \ldots, T$ and taking expectation give

   \[
   T \mathbb{E}(H(\bar{x}_T) - H(x^*)) \leq \sum_{t=1}^{T} \mathbb{E}(H(x_t) - H(x^*)) \leq \frac{\sqrt{T}}{2} (x_1 - x^*)^2 + \frac{\sqrt{T} G^2}{2}.
   \]

   The desired result for this part is obtained by dividing this inequality by $T$. In addition, if both $u$ and $\ell$ are finite and we choose $\eta_t = \frac{1}{\sqrt{t}}$, then summing (16) for $t = 1, 2, \ldots, T$ and taking expectation give

   \[
   T \mathbb{E}(H(\bar{x}_T) - H(x^*)) \leq \sum_{t=1}^{T} \mathbb{E}(H(x_t) - H(x^*)) \leq \sum_{t=1}^{T} \frac{\sqrt{T - t} - 1}{2} (x_t - x^*)^2 + \sum_{t=1}^{T} \frac{G^2}{2\sqrt{t}}.
   \]

   Note that $(x_t - x^*)^2 \leq (u - \ell)^2$ and $\sum_{t=1}^{T} \frac{1}{\sqrt{t}} \leq \int_{0}^{T} \frac{1}{\sqrt{x}} dx = 2\sqrt{T}$ so that the above inequality implies

   \[
   T \mathbb{E}(H(\bar{x}_T) - H(x^*)) \leq \sum_{t=1}^{T} \mathbb{E}(H(x_t) - H(x^*)) \leq \frac{\sqrt{T}}{2} (u - \ell)^2 + \sqrt{T} G^2.
   \]
The desired result for this part is obtained by dividing this inequality by $T$.

2. When Assumption A5(b) holds (based on Lan 2012):

The $\mu$-convexity property (2) of $H(\cdot)$ implies that

$$H(x^*) \geq H(x_t) + H'(x_t)(x^* - x_t) + \frac{\mu}{2}(x^* - x_t)^2$$

$$= H(x_t) + \mathbb{E}_t[g(x_t, \xi_t, z_t)(x^* - x_t)] + \frac{\mu}{2}(x^* - x_t)^2$$

$$= H(x_t) + \mathbb{E}_t[(g(x_t, \xi_t, z_t) - H'(x_t))(x_{t+1} - x_t)] + H'(x_t)(x_{t+1} - x_t)$$

$$+ \mathbb{E}_t[g(x_t, \xi_t, z_t)(x^* - x_{t+1})] + \frac{\mu}{2}(x^* - x_t)^2,$$

where the first equality is because of Proposition 1. By Assumption A5(b), $H'(x)$ is $L$-Lipschitz continuous, thus

$$H(x_{t+1}) \leq H(x_t) + H'(x_t)(x_{t+1} - x_t) + \frac{L}{2}(x_{t+1} - x_t)^2$$

which, together with (17), implies that

$$H(x^*) \geq \mathbb{E}_t H(x_{t+1}) + \mathbb{E}_t[(g(x_t, \xi_t, z_t) - H'(x_t))(x_{t+1} - x_t)] - \frac{L}{2}\mathbb{E}_t(x_{t+1} - x_t)^2$$

$$+ \mathbb{E}_t[g(x_t, \xi_t, z_t)(x^* - x_{t+1})] + \frac{\mu}{2}(x^* - x_t)^2$$

$$\geq \mathbb{E}_t H(x_{t+1}) - \frac{1}{2a_t}\mathbb{E}_t(g(x_t, \xi_t, z_t) - H'(x_t))^2 - \frac{L + a_t}{2}\mathbb{E}_t(x_{t+1} - x_t)^2$$

$$+ \mathbb{E}_t[g(x_t, \xi_t, z_t)(x^* - x_{t+1})] + \frac{\mu}{2}(x^* - x_t)^2$$

$$\geq \mathbb{E}_t H(x_{t+1}) - \frac{\sigma^2}{2a_t} - \frac{L + a_t}{2}\mathbb{E}_t(x_{t+1} - x_t)^2 + \mathbb{E}_t[g(x_t, \xi_t, z_t)(x^* - x_{t+1})] + \frac{\mu}{2}(x^* - x_t)^2,$$

where $a_t$ is a positive constant, the second inequality is due to Young’s inequality, namely, $xy \leq \frac{x^2}{2a} + \frac{a^2}{2}$ for any $a > 0$, and the last equality is due to the third statement of Proposition 1. We will determine the value of $a_t$ later but we always ensure that $a_t$ and $\eta_t$ satisfy

$$\frac{L + a_t}{2} - \frac{1}{2\eta_t} \leq 0.$$  

By the optimality of $x_{t+1}$ as a solution to the projection problem (5), we have

$$(x_{t+1} - x_t + \eta_t g(x_t, \xi_t, z_t))(x^* - x_{t+1}) \geq 0.$$ 

Thus we have

$$H(x^*) \geq \mathbb{E}_t H(x_{t+1}) - \frac{\sigma^2}{2a_t} - \frac{L + a_t}{2}\mathbb{E}_t(x_{t+1} - x_t)^2 + \frac{1}{\eta_t}\mathbb{E}_t[(x_{t+1} - x_t)(x_{t+1} - x^*)] + \frac{\mu}{2}(x^* - x_t)^2$$

$$= \mathbb{E}_t H(x_{t+1}) - \frac{\sigma^2}{2a_t} - \left(\frac{L + a_t}{2} - \frac{1}{2\eta_t}\right)\mathbb{E}_t(x_{t+1} - x_t)^2$$

$$+ \frac{1}{2\eta_t}\mathbb{E}_t(x_{t+1} - x^*)^2 - \left(\frac{1}{2\eta_t} - \frac{\mu}{2}\right)\mathbb{E}_t(x_t - x^*)^2$$

$$\geq \mathbb{E}_t H(x_{t+1}) - \frac{\sigma^2}{2a_t} + \frac{1}{2\eta_t}\mathbb{E}_t(x_{t+1} - x^*)^2 - \left(\frac{1}{2\eta_t} - \frac{\mu}{2}\right)\mathbb{E}_t(x_t - x^*)^2,$$  

(19)
where the second inequality is from (18).

If \( \mu = 0 \) and we choose \( \eta_t = \frac{1}{L + \sqrt{t}} \) and \( a_t = \sqrt{T} \) so that (18) is satisfied. Summing (19) for \( t = 1, 2, \ldots, T - 1 \) and organizing terms give

\[
T \mathbb{E}(H(\tilde{x}_T) - H(x^*)) \leq \sum_{t=1}^{T} \mathbb{E}(H(x_t) - H(x^*)) \leq \frac{L+\sqrt{T}}{2}(x_1 - x^*)^2 + H(x_1) - H(x^*) + \frac{\sqrt{T} \sigma^2}{2}.
\]

The desired result for this part is obtained by dividing this inequality by \( T \). In addition, if both \( u \) and \( \ell \) are finite and we choose \( \eta_t = \frac{1}{L + \sqrt{t}} \) and \( a_t = \sqrt{t} \), then summing (19) for \( t = 1, 2, \ldots, T - 1 \) and taking expectation give

\[
T \mathbb{E}(H(\tilde{x}_T) - H(x^*)) \leq \sum_{t=1}^{T} \mathbb{E}(H(x_t) - H(x^*)) \leq \frac{L}{2}(x_1 - x^*)^2 + \sum_{t=1}^{T-1} \frac{\sigma^2}{2t}.
\]

Note that \((x_t - x^*)^2 \leq (u - \ell)^2\) and \( \sum_{t=1}^{T-1} \frac{1}{\sqrt{t}} \leq \int_0^T \frac{1}{\sqrt{x}} \, dx = 2\sqrt{T} \) so that the above inequality implies

\[
T \mathbb{E}(H(\tilde{x}_T) - H(x^*)) \leq \sum_{t=1}^{T} \mathbb{E}(H(x_t) - H(x^*)) \leq \frac{T}{2}(u - \ell)^2 + H(x_1) - H(x^*) + \frac{L}{2}(x_1 - x^*)^2 + \sqrt{T} \sigma^2.
\]

The desired result for this part is obtained by dividing this inequality by \( T \).

\[\square\]

**Proof of Theorem 2.** Similar to the proof of Theorem 1, we consider the case when Assumption A5(a) or A5(b) holds respectively, and use \( \mathbb{E}_t \) to denote the conditional expectation taken over \( \xi_t \) and \( z_t \) conditioning on \( \xi_1, z_1, \xi_2, z_2, \ldots, \xi_{t-1}, z_{t-1} \).

1. When Assumption A5(a) holds: We can still show (16) using exactly the same argument in the proof of Theorem 1.

   If \( \mu > 0 \) and we choose \( \eta_t = \frac{1}{\mu t} \), summing (16) for \( t = 1, 2, \ldots, T \) gives

   \[
   T \mathbb{E}(H(\tilde{x}_T) - H(x^*)) \leq \sum_{t=1}^{T} \mathbb{E}(H(x_t) - H(x^*)) \leq \sum_{t=1}^{T} \frac{C^2}{2\mu t}.
   \]

   The desired result for this part is obtained by dividing this inequality by \( T \) and using the fact that \( \sum_{t=1}^{T} \frac{1}{t} \leq \log T + 1 \).

2. When Assumption A5(b) holds: Using exactly the same argument in the proof of Theorem 1, we can show that (19) holds for any positive constant \( a_t \) that satisfies (18).

   If \( \mu > 0 \) and we choose \( \eta_t = \frac{1}{\mu t + L} \) and \( a_t = \mu t \) so that (18) is satisfied. Summing (19) for \( t = 1, 2, \ldots, T - 1 \) gives

   \[
   T \mathbb{E}(H(\tilde{x}_T) - H(x^*)) \leq \sum_{t=1}^{T} \mathbb{E}(H(x_t) - H(x^*)) \leq \sum_{t=1}^{T} \frac{\sigma^2}{2\mu t} + H(x_1) - H(x^*) + \frac{L(x_1 - x^*)^2}{2}.
   \]

   The desired result for this part is obtained by dividing this inequality by \( T \) and using the fact that \( \sum_{t=1}^{T} \frac{1}{t} \leq \log T + 1 \). \[\square\]

Before we prove Theorem 3, we first introduce the following lemma:
Lemma 1. If Assumption A5(a) holds, then by choosing $\eta_t = \eta \in (0, +\infty)$, the CBA ensures that
\[
\mathbb{E}(H(\hat{x}_T) - H(x^*)) \leq \frac{(x^*_1 - x_1^*)^2}{2\eta T} + \frac{\eta G^2}{2}.
\]
If Assumption A5(b) holds, then by choosing $\eta_t = \eta \in (0, \frac{1}{T})$ and $a_t = \frac{1}{\eta_t} - L$, the CBA ensures that
\[
\mathbb{E}(H(\hat{x}_T) - H(x^*)) \leq \frac{(x^*_1 - x_1^*)^2}{2\eta T} + \frac{H(x_1^*) - H(x^*)}{T} + \frac{\sigma^2}{1/\eta - L}.
\]

Proof of Lemma 1. When Assumption A5(a) holds, by choosing $\eta_t = \eta \in (0, \infty)$ and summing (16) over $t = 1, 2, \ldots, T$, we have
\[
TE(\hat{x}_{\pi_T} - H(x^*)) \leq \sum_{t=1}^{T} E(H(x_t) - H(x^*)) \leq \frac{(x^*_1 - x_1^*)^2}{2\eta} + \frac{T\eta G^2}{2}.
\]
The first conclusion is obtained by dividing this inequality by $T$.

When Assumption A5(b) holds, by setting $\eta_t = \eta \in (0, \frac{1}{T})$, $a_t = \frac{1}{\eta_t} - L$ and summing (19) over $t = 1, 2, \ldots, T-1$, we have
\[
TE(\hat{x}_{\pi_T} - H(x^*)) \leq \frac{(x^*_1 - x_1^*)^2}{2\eta} + \frac{H(x_1^*) - H(x^*)}{T} + \frac{T\sigma^2}{1/\eta - L}.
\]
The second conclusion is obtained by dividing this inequality by $T$. □

Proof of Theorem 3. The optimality of $x^*$ and the $\mu$-convexity property (2) of $H(\cdot)$ imply
\[
\frac{\mu}{2}(\hat{x}^k - x^*)^2 \leq H(\hat{x}_k) - H(x^*).
\]

With a slight abuse of notation, let $E_k$ be the conditional expectation conditioning on $\hat{x}^1, \hat{x}^2, \ldots, \hat{x}^k$.

1. When Assumption A5(a) holds (based on Hazan and Kale 2014):

Define $\Delta_k = H(\hat{x}_k) - H(x^*)$ for $k \geq 1$. In the following, we use induction to show $E\Delta_k \leq \frac{\Delta_1 + G^2/\mu}{2^k}$ for $k \geq 1$. Note that this statement holds trivially when $k = 1$. Suppose $E\Delta_k \leq \frac{\Delta_1 + G^2/\mu}{2^k}$. Now we consider $E\Delta_{k+1}$.

By Lemma 1 and $\eta_t^k = \frac{1}{2^k/T}$, we have
\[
E\Delta_{k+1} = E_k(H(\hat{x}^{k+1}) - H(x^*)) \leq \frac{2^k \mu (\hat{x}^k - x^*)^2}{T_k} + \frac{G^2}{2^k + 2 \mu} \leq \frac{2^{k+1} \Delta_k}{T_k} + \frac{G^2}{2^k + 2 \mu},
\]
where the second inequality is due to (20). Taking expectation over $\hat{x}^1, \hat{x}^2, \ldots, \hat{x}^k$ and applying the induction assumption $E\Delta_k = \frac{\Delta_1 + G^2/\mu}{2^k}$, we have
\[
E\Delta_{k+1} \leq \frac{2^{k+1} \Delta_1 + G^2}{T_k 2^{k-1}} + \frac{G^2}{2^k + 2 \mu} \leq \frac{\Delta_1}{2^k + 1} + \frac{G^2}{2^{k+1} \mu} + \frac{G^2}{2^k + 2 \mu} \leq \frac{\Delta_1 + G^2/\mu}{2^k},
\]
where we use the facts that $T_k = 2^{k+3}$. By induction assumption, $E\Delta_k \leq \frac{\Delta_1 + G^2/\mu}{2^k}$ for $k \geq 1$. Since $T = \sum_{k=1}^{K} T_k = \sum_{k=1}^{K} 2^{k+3} \leq 2^{K+4}$, we have $K \geq \log_2(T/16)$. Let $k = K$ in (21), we have
\[
E(H(\hat{x}^{K+1}) - H(x^*)) = E\Delta_{K+1} \leq \frac{\Delta_1 + G^2/\mu}{2^k} \leq \frac{16(\Delta_1 + G^2/\mu)}{T}.
\]
2. When Assumption A5(b) holds:

Define \( \Delta_k = H(\hat{x}^k) - H(x^*) + \frac{1}{2}(\hat{x}^k - x^*)^2 \) for \( k \geq 1 \). In the following, we use induction to show \( \mathbb{E}\Delta_k \leq \frac{\Delta_1 + \sigma^2/\mu}{2^{k+1}} \) for \( k \geq 1 \). Note that this statement holds trivially when \( k = 1 \). Suppose \( \mathbb{E}\Delta_k \leq \frac{\Delta_1 + \sigma^2/\mu}{2^{k+1}} \). Now we consider \( \mathbb{E}\Delta_{k+1} \).

By Lemma 1 and \( \eta_k^i = \frac{1}{2^{k+1}} (1 + \eta_k^i) \in (0, \frac{1}{2}) \), we have

\[
\mathbb{E}\Delta_{k+1} = \mathbb{E}_k(H(\hat{x}^{k+1}) - H(x^*)) \leq \frac{(2^{k+1} + L)(\hat{x}^k - x^*)^2}{2T_k} + \frac{\Delta_k}{T_k} \leq \frac{(2^{k+1} + 1)(\Delta_1 + \sigma^2/\mu)}{2^{k+1}} + \frac{\sigma^2}{2^{k+1} \mu},
\]

where the second inequality is due to (20). Taking expectation over \( \hat{x}^1, \hat{x}^2, \ldots, \hat{x}^k \) and applying the induction assumption \( \mathbb{E}\Delta_k = \frac{\Delta_1 + \sigma^2/\mu}{2^{k+1}} \), we have

\[
\mathbb{E}\Delta_{k+1} \leq \left(2^{k+1} + 1\right) \frac{\Delta_1 + \sigma^2/\mu}{2^{k+1}} + \frac{\sigma^2}{2^{k+1} \mu} \leq \frac{\Delta_1 + \sigma^2/\mu}{2^{k+1}} + \frac{\sigma^2}{2^{k+1} \mu} \leq \frac{\Delta_1 + \sigma^2/\mu}{2^k},
\]

where we use the facts that \( T_k = 2^{k+1} + 4 \). Thus, \( \mathbb{E}\Delta_k \leq \frac{\Delta_1 + \sigma^2/\mu}{2^k} \) for \( k \geq 1 \). Since \( T = \sum_{k=1}^{K} T_k = \sum_{k=1}^{K} (2^{k+1} + 4) \leq 2^{K+5} \), we have \( K \geq \log_2(T/32) \). Let \( k = K \) in (22), we have

\[
\mathbb{E}(H(\hat{x}^{K+1}) - H(x^*)) = \mathbb{E}\Delta_{K+1} \leq \frac{\Delta_1 + \sigma^2/\mu}{2^K} = \frac{32(\Delta_1 + \sigma^2/\mu)}{T}.
\]

Thus the theorem is proved. \( \square \)

### 7.2 CBA with Categorical Results from Comparison

In CBA given in Algorithm 1, the results of the comparison between \( x \) and \( \xi \) is binary, i.e., either \( \xi > x \) or \( \xi < x \). In this section, we extend CBA to allow the comparison result to be categorical and depend on the gap between \( \xi \) and \( x \). In particular, we assume there exist \( m + 1 \) non-negative quantities \( \theta_0, \theta_1, \ldots, \theta_m \) and \( \theta_m := +\infty \) such that, after presenting a solution \( x \), we know whether \( \xi \in [x + \theta_i, x + \theta_i + 1] \) or \( \xi \in [x - \theta_i, x - \theta_i + 1] \) for any \( i = 0, 1, \ldots, m - 1 \). Using Example 1 as an example, this type of comparison result corresponds to the case where the customer reports a coarse level of the difference between his/her preferred value of the feature and the actual value of the feature of the product presented to him/her (e.g., whether the size of the product is too small, a little small, a little large or too large). Note that the binary comparison result is a special case of the categorical result with \( m = 1 \).

To facilitate the development of algorithm, we need to replace Assumption A2 with the following assumption:

(A2') For each \( \xi \), \( h(x, \xi) \) is continuously differentiable with respect to \( x \) on \([\ell, \xi] \) and \( (\xi, u) \) with the derivative denoted by \( h'_x(x, \xi) \). Furthermore, for any \( x \in [\ell, u] \), \( h'_+ (x, x - \theta_i) := \lim_{z \to x - \theta_i} h'_x(x, z) \) and \( h'_- (x, x + \theta_i) := \lim_{z \to x + \theta_i} h'_x(x, z) \) exist and are finite for any \( i = 0, 1, \ldots, m - 1 \).

With this comparison information, we propose a comparison-based algorithm with categorical comparison result (CBA-C) to solve (1). The algorithm requires specification of \( 2m \) functions, \( f^1_i(x, z) \) and \( f^2_i(x, z) \) for \( i = 0, 1, \ldots, m - 1 \), which need to satisfy the following conditions.
• (C1’$ f^i_-(x, z) = 0 \text{ for all } z \notin [x - \theta_{i+1}, x - \theta_i] \text{ and } f^i_-(x, z) > 0 \text{ for all } \max\{z, x - \theta_{i+1}\} \leq z \leq x - \theta_i$. In addition, for all $x$, we have $\int_{x - \theta_{i+1}}^{x - \theta_i} f^i_-(x, z)dz = 1.$

• (C2’$ f^i_+(x, z) = 0 \text{ for all } z \notin (x + \theta_i, x + \theta_{i+1}) \text{ and } f^i_+(x, z) > 0 \text{ for all } \min\{\bar{s}, x + \theta_{i+1}\} \geq z > x + \theta_i$. In addition, for all $x$, we have $\int_{x + \theta_i}^{x + \theta_{i+1}} f^i_+(x, z)dz = 1.$

• (C3’) There exists a constant $K^i_3$ such that $\int_{x - \theta_{i+1}}^{x - \theta_i} \frac{(F(z) - F(x - \theta_{i+1}))(h^i_{x,z}(x, z))^2}{f^i_-(x, z)}dz \leq K^i_3$ and $\int_{x - \theta_{i+1}}^{x - \theta_i} \frac{(F(x + \theta_{i+1}) - F(z))(h^i_{x,z}(x, z))^2}{f^i_+(x, z)}dz \leq K^i_3$ for all $x \in [\ell, u]$, where $F(\cdot)$ is the c.d.f. of $\xi$.

Note that $f^i_-(x, z)$ and $f^i_+(x, z)$ essentially define the density function on $[x - \theta_{i+1}, x - \theta_i]$ and $(x + \theta_i, x + \theta_{i+1}]$, respectively, for any given $x \in [\ell, u]$. The functions $f^i_-(x, z)$ and $f^i_+(x, z)$ satisfying C1’-C3’ can be constructed in a similar way as in Example 5 and 6 and their optimal choices are similar to (8) and (9). Next, in Algorithm 5, we describe the detailed procedure of the CBA-C.

Algorithm 5 Comparison-Based Algorithm with Categorical Comparison Result (CBA-C):

1. **Initialization.** Set $t = 1$, $x_1 \in [\ell, u]$. Define $\eta_t$ for all $t \geq 1$. Set the maximum number of iterations $T$. Choose functions $f^i_-(x, z)$ and $f^i_+(x, z)$ for $i = 0, 1, \ldots, m - 1$ that satisfy (C1’)-(C3’).

2. **Main iteration.** Sample $\xi_t$ from the distribution of $\xi$. If $\xi_t = x_t$, then resample $\xi_t$ until it does not equal $x_t$. (This step will always terminate in a finite number of steps as long as $\xi$ is not deterministic.)

   (a) If $\xi_t \in [x_t - \theta_{i+1}, x_t - \theta_i]$, then generate $z_t$ from a distribution on $[x_t - \theta_{i+1}, x_t - \theta_i]$ with p.d.f. $f^i_-(x_t, z_t)$. Set
   
   \[ g(x_t, \xi_t, z_t) = \begin{cases} 
   h^i_-(x_t, x_t - \theta_i), & \text{if } z_t < \xi_t, \\
   h^i_-(x_t, x_t - \theta_i) - \frac{h^i_{x,z}(x_t, z_t)}{f^i_-(x_t, z_t)}, & \text{if } z_t \geq \xi_t.
   \end{cases} \]  
   \[ (23) \]

   (b) If $\xi_t \in (x_t + \theta_i, x_t + \theta_{i+1}]$, then generate $z_t$ from a distribution on $(x_t + \theta_i, x_t + \theta_{i+1}]$ with p.d.f. $f^i_+(x_t, z_t)$. Set
   
   \[ g(x_t, \xi_t, z_t) = \begin{cases} 
   h^i_+(x_t, x_t + \theta_i), & \text{if } z_t > \xi_t, \\
   h^i_+(x_t, x_t + \theta_i) + \frac{h^i_{x,z}(x_t, z_t)}{f^i_+(x_t, z_t)}, & \text{if } z_t \leq \xi_t.
   \end{cases} \]  
   \[ (24) \]

   Let
   
   \[ x_{t+1} = \text{Proj}_{[\ell, u]}(x_t - \eta_t g(x_t, \xi_t, z_t)) = \max(\ell, \min(u, x_t - \eta_t g(x_t, \xi_t, z_t))). \]  
   \[ (25) \]

3. **Termination.** Stop when $t \geq T$. Otherwise, let $t \leftarrow t + 1$ and go back to Step 2.

4. **Output.** CBA-C$(x_1, T, \eta_1^T) = \bar{x}_T = \frac{1}{T} \sum_{t=1}^{T} x_t$.

We have the following proposition about the comparison-based algorithm with categorical result.
Proposition 3. Suppose \( f^i_-(x, z) \) and \( f^i_+(x, z) \) satisfy (C1')-(C3') and Assumption 1 holds with (A2) replaced by (A2'). Then

1. \( \mathbb{E}_z g(x, \xi, z) = h'_x(x, \xi) \), for all \( x \in [\ell, u] \), \( x \neq \xi \).
2. \( \mathbb{E}_z g(x, \xi, z) = H'(x) \), for all \( x \in [\ell, u] \).
3. If Assumption A5(a) holds, then \( \mathbb{E}_{z, \xi}(g(x, \xi, z))^2 \leq \sigma^2 := K^2_2 + 2 \sum_{i=0}^{m-1} K^i_3 \). If Assumption A5(b) holds, then \( \mathbb{E}_{z, \xi}(g(x, \xi, z) - H'(x))^2 \leq \sigma^2 := K^2_2 + 2 \sum_{i=0}^{m-1} K^i_3 \).

Proof of Proposition 1. First, we consider the case when \( \xi \in [x - \theta_{i+1}, x - \theta_i) \). We have

\[
\mathbb{E}_z g(x, \xi, z) = h'_-(x, x - \theta_i) - \int_{\xi}^{x-\theta_i} h''_{x,z}(x, z) dz = h'_x(x, \xi).
\]

Similarly, when \( \xi \in (x + \theta_i, x + \theta_{i+1}] \),

\[
\mathbb{E}_z g(x, \xi, z) = h'_+(x, x + \theta_i) + \int_{x+\theta_i}^{\xi} h''_{x,z}(x, z) dz = h'_x(x, \xi).
\]

Thus the first conclusion of the proposition is proved. The second conclusion of the proposition follows from Assumption A1 (which ensures \( \xi = x \) is a zero-measure event) and Assumption A4.

Next, we show the first part of the third conclusion when Assumption A5(a) is true. If \( \xi \in [x - \theta_{i+1}, x - \theta_i) \), then we have

\[
\mathbb{E}_z (g(x, \xi, z))^2 = (h'_-(x, x - \theta_i))^2 + \int_{\xi}^{x-\theta_i} \left( -2h'_-(x, x - \theta_i) \frac{h''_{x,z}(x, z)}{f^i_-(x, z)} + \left( \frac{h''_{x,z}(x, z)}{f^i_-(x, z)} \right)^2 \right) f^i_-(x, z) dz
\]

\[
= (h'_-(x, x - \theta_i))^2 - 2h'_-(x, x - \theta_i)(h'_-(x, x - \theta_i) - h'_x(x, \xi)) + \int_{\xi}^{x-\theta_i} \left( \frac{h''_{x,z}(x, z)}{f^i_-(x, z)} \right)^2 f^i_-(x, z) dz.
\]

where the last inequality is because \( a^2 + b^2 \geq 2ab \) for any \( a, b \). By similar arguments, if \( \xi \in (x + \theta_i, x + \theta_{i+1}] \), then

\[
\mathbb{E}_z (g(x, \xi, z))^2 \leq (h'_x(x, \xi))^2 + \int_{x+\theta_i}^{\xi} \left( \frac{h''_{x,z}(x, z)}{f^i_+(x, z)} \right)^2 f^i_+(x, z) dz.
\]

These two inequalities and Assumption A5(a) further imply

\[
\mathbb{E}_{z, \xi}(g(x, \xi, z))^2 \leq K^2_2 + \sum_{i=0}^{m-1} \int_{x-\theta_i}^{x-\theta_{i+1}} \left( \int_{\xi}^{x-\theta_i} \left( \frac{h''_{x,z}(x, z)}{f^i_-(x, z)} \right)^2 f^i_-(x, z) \right) dF(\xi) + \sum_{i=0}^{m-1} \int_{x+\theta_i}^{x+\theta_{i+1}} \left( \int_{x+\theta_i}^{x+\theta_{i+1}} \left( \frac{h''_{x,z}(x, z)}{f^i_+(x, z)} \right)^2 f^i_+(x, z) \right) dF(\xi)
\]

\[
= K^2_2 + \sum_{i=0}^{m-1} \int_{x-\theta_i}^{x-\theta_{i+1}} (F(z) - F(x - \theta_{i+1})) (h''_{x,z}(x, z))^2 f^i_-(x, z) dz + \sum_{i=0}^{m-1} \int_{x+\theta_i}^{x+\theta_{i+1}} (F(x + \theta_{i+1}) - F(x)) (h''_{x,z}(x, z))^2 f^i_+(x, z) dz
\]

\[
\leq K^2_2 + 2 \sum_{i=0}^{m-1} K^i_3,
\]

where the interchanging of integrals in the equality is justified by Tonelli’s theorem and the last inequality is due to (C3').
Next, we show the second part of the third conclusion when Assumption A5(b) is true. If \( \xi \in [x - \theta_i, x - \theta_{i+1}) \), then following the similar analysis as in (26), we have

\[
\mathbb{E}_z(g(x, \xi, z) - h'_x(x, \xi))^2 = \mathbb{E}_z(g(x, \xi, z))^2 - (h'_x(x, \xi))^2 \leq \int_{\xi}^{x-\theta_i} \frac{(h''_{x,z}(x, z))^2}{f_+(x, z)} d\xi.
\]

Similarly, if \( \xi \in (x, x + \theta_i, x + \theta_{i+1}] \), then

\[
\mathbb{E}_z(g(x, \xi, z) - h'_x(x, \xi))^2 \leq \int_{x+\theta_i}^{\xi} \frac{(h''_{x,z}(x, z))^2}{f_+(x, z)} d\xi.
\]

By using the same argument as in (27), we have

\[
\mathbb{E}_{z,\xi}(g(x, \xi, z) - h'_x(x, \xi))^2 \leq 2 \sum_{i=0}^{m-1} K_i^3.
\]

Finally, we note that,

\[
\mathbb{E}_{z,\xi}(g(x, \xi, z) - H'(x))^2 = \mathbb{E}_{z,\xi}(g(x, \xi, z) - h'_x(x, \xi))^2 + \mathbb{E}_{\xi}(h'_x(x, \xi) - H'(x))^2.
\]

Therefore, when Assumption A5(b) holds, we have \( \mathbb{E}_{z,\xi}(g(x, \xi, z) - H'(x))^2 \leq K^2 + 2 \sum_{i=0}^{m-1} K_i^3 \).

Thus the proposition holds. \( \square \)

Proposition 3 shows that in the CBA-C, the gradient estimate \( g(x, \xi, z) \) is an unbiased estimate of the true gradient at \( x \) and can be utilized as a stochastic gradient of \( H(x) \). As a result, we can also prove that the convergence rates for CBA-C is exactly the same as those of CBA given in Theorems 1 and 2 except that \( G^2 \) and \( \sigma^2 \) are replaced by \( G'^2 \) and \( \sigma'^2 \).