Truncated Variance Reduction: A Unified Approach to Bayesian Optimization and Level-Set Estimation

Ilija Bogunovic¹, Jonathan Scarlett¹, Andreas Krause², Volkan Cevher¹
¹ Laboratory for Information and Inference Systems (LIONS), EPFL
² Learning and Adaptive Systems Group, ETH Zürich
{ilija.bogunovic,jonathan.scarlett,volkan.cevher}@epfl.ch, krausea@ethz.ch

Abstract

We present a new algorithm, truncated variance reduction (TRUVAR), that addresses Bayesian optimization (BO) and level-set estimation (LSE) with Gaussian processes in a unified fashion. The algorithm greedily shrinks a sum of truncated variances within a set of potential maximizers (BO) or unclassified points (LSE), which is updated based on confidence bounds. TRUVAR is effective in several important settings that are typically non-trivial to incorporate into myopic algorithms, including pointwise costs and heteroscedastic noise. We provide a general theoretical guarantee for TRUVAR covering these phenomena, and use it to recover and strengthen existing results. Moreover, we provide a new result for a setting where one can select from a number of noise levels having associated costs. We demonstrate the effectiveness of the algorithm on both synthetic and real-world data sets.

1 Introduction

Bayesian optimization (BO) [1] provides a powerful framework for automating design problems, and finds applications in robotics, environmental monitoring, and automated machine learning, just to name a few. One seeks to find the maximum of an unknown reward function that is expensive to evaluate, based on a sequence of suitably-chosen points and noisy observations. Numerous BO algorithms have been presented previously; see Section 1.1 for an overview.

Level-set estimation (LSE) [2] is closely related to BO, with the added twist that instead of seeking a maximizer, one seeks to classify the domain into points that lie above or below a certain threshold. This is of considerable interest in applications such as environmental monitoring and sensor networks, allowing one to find all “sufficiently good” points rather than the best point alone.

While BO and LSE are closely related, they are typically studied in isolation. In this paper, we provide a unified treatment of the two via a new algorithm, Truncated Variance Reduction (TRUVAR), which enjoys theoretical guarantees, good computational complexity, and the versatility to handle important settings such as pointwise costs, non-constant noise, and multi-task scenarios. The main result of this paper applies to the former two settings, and even the fixed-noise and unit-cost case, we refine existing bounds via a significantly improved dependence on the noise level.

1.1 Previous Work

Three popular myopic techniques for Bayesian optimization are expected improvement (EI), probability of improvement (PI), and Gaussian process upper confidence bound (GP-UCB) [1][3], each of which chooses the point maximizing an acquisition function depending directly on the current posterior mean and variance. In [4], the GP-UCB-PE algorithm was presented for BO, choosing the highest-variance point within a set of potential maximizers that is updated based on confidence bounds. Another relevant BO algorithm is BaMSOO, which also keeps track of potential maximizers, but instead chooses points based on a global optimization technique called simultaneous online
optimization (SOO). An algorithm for level-set estimation with GPs is given in [2], which keeps track of a set of unclassified points. These algorithms are computationally efficient and have theoretical guarantees, but it is usually unclear how best to incorporate phenomena such as pointwise costs and heteroscedastic noise [3]. The same is true for a heuristic for LSE known as Straddle [6].

Entropy search (ES) [7] and its predictive version [8] choose points to reduce the uncertainty of the location of the maximum, doing so via a one-step lookahead of the posterior rather than only the current posterior. While this increases the computation, it also permits versatility with respect to costs [5], heteroscedastic noise [9], and multi-task scenarios [5]. A recent approach called minimum regret search (MRS) [10] similarly performs a look-ahead, but instead chooses points to minimize the regret. To our knowledge, no theoretical guarantees have been provided for these algorithms.

The multi-armed bandit (MAB) [11] literature has developed alongside the BO literature, with the two often bearing similar concepts. The MAB literature is far too extensive to cover here, but we briefly mention some variants relevant to this paper. Extensive attention has been paid to the best-arm identification problem [12], and cost constraints have been incorporated in a variety of forms [13]. Moreover, the concept of “zooming in” to the optimal point has been explored [14]. In general, the assumptions and analysis techniques in the MAB and BO literature are quite different.

1.2 Contributions

We present a unified analysis of Bayesian optimization and level-set estimation via a new algorithm Truncated Variance Reduction (TRUVAR). The algorithm works by keeping track of a set of potential maximizers (BO) or unclassified points (LSE), selecting points that shrink the uncertainty within that set up to a truncation threshold, and updating the set using confidence bounds. Similarly to ES and MRS, the algorithm performs a one-step lookahead that is highly beneficial in terms of versatility. However, unlike these previous works, our lookahead avoids the computationally expensive task of averaging over the posterior distribution and the observations.

Also in contrast with ES and MRS, we provide theoretical bounds for TRUVAR characterizing the cost required to achieve a certain accuracy in finding a near-optimal point (BO) or in classifying each point in the domain (LSE). By applying this to the standard BO setting, we not only recover existing results [2, 4], but we also strengthen them via a significantly improved dependence on the noise level, with better asymptotics in the small noise limit. Moreover, we provide a novel result for a setting in which the algorithm can choose the noise level, each coming with an associated cost.

Finally, we compare our algorithm to previous works on several synthetic and real-world data sets, finding it to perform favorably in a variety of settings.

2 Problem Setup and Proposed Algorithm

Setup: We seek to sequentially optimize an unknown reward function \( f(x) \) over a finite domain \( D \). At time \( t \), we query a single point \( x_t \in D \) and observe a noisy sample \( y_t = f(x_t) + z_t \), where \( z_t \sim N(0, \sigma^2(x_t)) \) for some known noise function \( \sigma^2(\cdot) : D \to \mathbb{R}_+ \). Thus, in general, some points may be noisier than others, in which case we have heteroscedastic noise [9]. We associate with each point a cost according to some known cost function \( c : D \to \mathbb{R}_+ \). If both \( \sigma^2(\cdot) \) and \( c(\cdot) \) are set to be constant, then we recover the standard homoscedastic and unit-cost setting.

We model \( f(x) \) as a Gaussian process (GP) [15] having mean zero and kernel function \( k(x, x') \), normalized so that \( k(x, x) = 1 \) for all \( x \in D \). The posterior distribution of \( f \) given the points and observations up to time \( t \) is again a GP, with the posterior mean and variance given by [9]

\[
\mu_t(x) = k_t(x)^T (K_t + \Sigma_t)^{-1} y_t
\]

\[
\sigma_t(x)^2 = k(x, x) - k_t(x)^T (K_t + \Sigma_t)^{-1} k_t(x),
\]

where \( k_t(x) = [k(x_1, x), \ldots, k(x_t, x)]_t, \) \( K_t = [k(x_t, x)]_t, \) and \( \Sigma_t = \text{diag}(\sigma^2(x_1), \ldots, \sigma^2(x_t)) \). We also let \( \sigma_t^{-1} = \Sigma_t^{-1} + k_t(x) \) denote the posterior variance of \( \bar{x} \) upon observing \( x \) along with \( x_1, \ldots, x_{t-1} \).

---

1 Extensions to continuous domains and non-Bayesian settings are discussed in the supplementary material.
We consider both Bayesian optimization, which consists of finding a point whose function value is as high as possible, and level-set estimation, which consists of classifying the domain according into points that lie above or below a given threshold \( h \). The precise performance criteria for these settings are given in Definition 3.1 below. Essentially, after spending a certain cost we report a point (BO) or a classification (LSE), but there is no preference on the values of \( f(x_t) \) for the points \( x_t \) chosen before coming to such a decision (in contrast with other notions such as cumulative regret).

**TRUVAR Algorithm:** Our algorithm is described in Algorithm 1, making use of the updates described in Algorithm 2. The algorithm keeps track of a sequence of unclassified points \( M_t \), representing potential maximizers for BO or points close to \( h \) for LSE. This set is updated based on the confidence bounds depending on constants \( \beta(i) \). The algorithm proceeds in epochs, where in the \( i \)-th it seeks to bring the confidence \( \beta(i)^{1/2} \sigma_i(x) \) of points within \( M_t \) below a target value \( \eta(i) \). It does this by greedily minimizing the sum of truncated variances \( \sum_{x \in M_{t-1}} \max\{\beta(i)^2 \sigma^2_{i-1}(x), \eta(i)\} \) arising from choosing the point \( x \), along with a normalization and division by \( c(x) \) to favor low-cost points.

The truncation by \( \eta(i) \) in this decision rule means that once the confidence of a point is below the current target value, there is no preference in making it any lower (until the target is decreased). Once the confidence of every point in \( M_t \) is less than a factor \( 1 + \tilde{\alpha} \) above the target value, the target confidence is reduced according to a multiplication by \( r \in (0, 1) \). An illustration of the process is given in Figure 1 with details in the caption.

For level-set estimation, we also keep track of the sets \( H_t \) and \( L_t \), containing points believed to have function values above and below \( h \), respectively. The constraint \( x \in M_{t-1} \) in (5) ensures that \( \{M_t\} \) is non-increasing with respect to inclusion, and \( H_t \) and \( L_t \) are non-decreasing.

### Algorithm 1 Truncated Variance Reduction (TRUVAR)

**Input:** Domain \( D \), GP prior \( \mu_0, \sigma_0, k \), confidence bound parameters \( \tilde{\alpha} > 0, r \in (0, 1) \), \( \{\beta(i)\}_{i \geq 1} \), \( \eta(1) > 0 \), and for LSE, level-set threshold \( h \)

1: Initialize the epoch number \( t = 1 \) and potential maximizers \( M_{(0)} = D \).
2: for \( t = 1, 2, \ldots \) do
3: Choose
   \[
   x_t = \arg \max_{x \in D} \frac{\sum_{x \in M_{t-1}} \max\{\beta(i)^2 \sigma^2_{i-1}(x), \eta(i)\} - \sum_{x \in M_{t-1}} \max\{\beta(i)^2 \sigma^2_{i-1}(x), \eta(i)\}}{c(x)}.
   \] (3)
4: Observe the noisy function sample \( y_t \), and update according to Algorithm 2 to obtain \( M_t \), \( \mu_t, \sigma_t, l_t \) and \( u_t \), as well as \( H_t \) and \( L_t \) in the case of LSE
5: while \( \max_{x \in M_t} \beta(i)^{1/2} \sigma_i(x) \leq (1 + \tilde{\alpha}) \eta(i) \) do
6: Increment \( i \), set \( \eta(i) = r \times \eta(i-1) \).

The choices of \( \beta(i), \tilde{\alpha}, \) and \( r \) are discussed in Section 4. As with previous works, the kernel is assumed known in our theoretical results, whereas in practice it is typically learned from training data [3]. Characterizing the effect of model mismatch or online hyperparameter updates is beyond the scope of this paper, but is an interesting direction for future work.
Algorithm 2 Parameter Updates for TRUVAR

**Input:** Selected points and observations \( \{x'_t\}_{t'=1}^t; \{y'_t\}_{t'=1}^t, \) previous sets \( M_{t-1}, H_{t-1}, L_{t-1}, \) parameter \( \beta_{(i)}^{-1/2}, \) and for LSE, level-set threshold \( h. \)

1: Update \( \mu_t \) and \( \sigma_t \) according to (1)–(2), and form the upper and lower confidence bounds

\[
u_t(x) = \mu_t(x) + \beta_{(i)}^{-1/2} \sigma(x), \quad \ell_t(x) = \mu_t(x) - \beta_{(i)}^{-1/2} \sigma_t(x).\]

2: For BO, set

\[
M_t = \left\{ x \in M_{t-1} : u_t(x) \leq \max_{x \in M_{t-1}} \ell_t(x) \right\},
\]

or for LSE, set

\[
M_t = \left\{ x \in M_{t-1} : u_t(x) \geq h \text{ and } \ell_t(x) \leq h \right\} \quad \text{(6)}
\]

\[
H_t = H_{t-1} \cup \left\{ x \in M_{t-1} : \ell_t(x) > h \right\}, \quad L_t = L_{t-1} \cup \left\{ x \in M_{t-1} : u_t(x) < h \right\}. \quad \text{(7)}
\]

Some variants of our algorithm and theory are discussed in the supplementary material due to lack of space, including pure variance reduction, non-Bayesian settings [3], continuous domains [3], the batch setting [4], and implicit thresholds for level-set estimation [2].

### 3 Theoretical Bounds

In order to state our results for BO and LSE in a unified fashion, we define a notion of \( \epsilon\)-accuracy for the two settings. That is, we define this term differently in the two scenarios, but then we provide theorems that simultaneously apply to both. All proofs are given in the supplementary material.

**Definition 3.1.** After time step \( t \) of TRUVAR, we use the following terminology:

- For BO, the set \( M_t \) is \( \epsilon\)-accurate if it contains all true maxima \( x^* \in \arg \max_x f(x) \), and all of its points satisfy \( f(x^*) - f(x) \leq \epsilon. \)

- For LSE, the triplet \( (M_t, H_t, L_t) \) is \( \epsilon\)-accurate if all points in \( H_t \) satisfy \( f(x) > h \), all points in \( L_t \) satisfy \( f(x) < h \), and all points in \( M_t \) satisfy \( |f(x) - h| \leq \epsilon/2. \)

In both cases, the cumulative cost after time \( t \) is defined as \( C_t = \sum_{t'=1}^t c(x_{t'}). \)

We use \( \frac{\epsilon}{2} \) in the LSE setting instead of \( \epsilon \) since this creates a region of size \( \epsilon \) where the function value lies, which is consistent with the BO setting. Our performance criterion for level-set estimation is slightly different from that of [2], but the two are closely related.

### 3.1 General Result

**Preliminary Definitions:** Suppose that the \( \{\beta_{(i)}\} \) are chosen to ensure valid confidence bounds, i.e., \( l_t(x) \leq f(x) \leq u_t(x) \) with high probability; see Theorem 3.3 and its proof below for such choices. In this case, we have after the \( i\)-th epoch that all points are either already discarded (BO) or classified (LSE), or are known up to the confidence level \( (1 + \delta)\eta_{(i)} \). For the points with such confidence, we have \( u_t(x) - l_t(x) \leq 2(1 + \delta)\eta_{(i)}, \) and hence

\[
u_t(x) \leq l_t(x) + 2(1 + \delta)\eta_{(i)} \leq f(x) + 2(1 + \delta)\eta_{(i)}, \quad \text{(8)}
\]

and similarly \( l_t(x) \geq f(x) - 2(1 + \delta)\eta_{(i)}. \) This means that all points other than those within a gap of width \( 4(1 + \delta)\eta_{(i)} \) must have been discarded or classified:

\[
M_t \subseteq \{ x : f(x) \geq f(x^*) - 4(1 + \delta)\eta_{(i)} \} =: \overline{M}_{(i)} \quad \text{(BO)}
\]

\[
M_t \subseteq \{ x : |f(x) - h| \leq 2(1 + \delta)\eta_{(i)} \} =: \bar{M}_{(i)} \quad \text{(LSE)}
\]

Moreover, we define \( \bar{M}_{(0)} = D. \)
For a collection of points $S = (x'_1, \ldots, x'_k)$, possibly containing duplicates, we write the total cost as $c(S) = \sum_{i=1}^{k} c(x'_i)$. Moreover, we denote the posterior variance upon observing the points up to time $t-1$ and the additional points in $S$ by $\sigma^2_{t-1}\sigma(S)$. Therefore, $c(x) = c\{x\}$ and $\sigma^2_{t-1}\sigma(S) = \sigma^2_{t-1}\sigma(S)$. The minimum cost (respectively, maximum cost) is denoted by $c_{\min} = \min_{x \in D} c(x)$ (respectively, $c_{\max} = \max_{x \in D} c(x)$).

Finally, we introduce the quantity

$$C^*(\xi, M) = \min_S \left\{ c(S) : \max_{x \in M} \sigma_0(S)(x) \leq \xi \right\},$$

representing the minimum cost to achieve a posterior standard deviation of at most $\xi$ within $M$.

**Statement of Theorem:** In all of our results, we make the following assumption.

**Assumption 3.1.** The kernel $k(x, x')$ is such that the variance reduction function

$$\psi_{t,x}(S) = \sigma^2_t(x) - \sigma^2_{t-1}(x)$$

is submodular [16] for any time $t$, and any selected points $(x_1, \ldots, x_i)$ and query point $x$.

This assumption has been used in several previous works based on Gaussian processes, and sufficient conditions for its validity can be found in [17, Sec. 8]. We now state the following general guarantee.

**Theorem 3.1.** Fix $\epsilon > 0$ and $\delta \in (0, 1)$, and suppose there exist values $\{C(i)\}$ and $\{\beta(i)\}$ such that

$$C(i) \geq C^*(\frac{\eta(i)}{\beta(i)}, \frac{M(i-1)}{M(i-1)}|\beta(i)) \log \frac{M(i-1)|\beta(i)}{\delta^2 \eta(i)} + c_{\max},$$

and

$$\beta(i) \geq 2 \log \frac{|D|((\sum_{i'=i}^{\epsilon} C(i'))^2 \pi^2}{6\epsilon c_{\min}}.$$  

Then if TRUVAR is run with these choices of $\beta(i)$ until the cumulative cost reaches

$$C_{\epsilon} = \sum_{i : 4(1+\delta)^i \eta(i) \geq \epsilon} C(i),$$

then with probability at least $1 - \delta$, we have $\epsilon$-accuracy.

While this theorem is somewhat abstract, it captures the fact that the algorithm improves when points having a lower cost and/or lower noise are available, since both of these lead to a smaller value of $C^*(\xi, M)$; the former by directly incurring a smaller cost, and the latter by shrinking the variance more rapidly. Below, we apply this result to some important cases.

### 3.2 Results for Specific Settings

**Homoscedastic and Unit-Cost Setting:** Define the maximum mutual information [3]

$$\gamma_T = \max_{x_1, \ldots, x_T} \frac{1}{2} \log \det (I_T + \sigma^{-2}K_T),$$

and consider the case that $\sigma^2(x) = \sigma^2$ and $c(x) = 1$. In the supplementary material, we provide a theorem with a condition for $\epsilon$-accuracy of the form $T \geq O^*(\frac{C_1 \gamma_T \beta_T}{\epsilon^2} + 1)$ with $C_1 = \frac{1}{\log(1+\sigma^{-2})}$, thus matching [2, 3] up to logarithmic factors. In the following, we present a refined version that has a significantly better dependence on the noise level, thus exemplifying that a more careful analysis of [13] can provide improvements over the standard bounding techniques.

**Corollary 3.1.** Fix $\epsilon > 0$ and $\delta \in (0, 1)$, define $\beta_T = \frac{1}{2} \log \frac{|D|\pi^2}{6\epsilon}$, and set $\eta(1) = 1$ and $r = \frac{1}{2}$. There exist choices of $\beta(i)$ (not depending on the time horizon $T$) such that we have $\epsilon$-accuracy with probability at least $1 - \delta$ once the following condition holds:

$$T \geq \left(2\epsilon^2 \gamma_T \beta_T \frac{96(1+\delta)^2}{\epsilon^2} + C_1 \gamma_T \beta_T \frac{6(1+\delta)^2}{\epsilon^2} + 2 \left[ \log_2 \frac{32(1+\delta)^2}{\epsilon \sigma} \right] \right) \log \frac{16(1+\delta)^2}{\delta^2 \epsilon^2} |D|\beta_T,$$

where $C_1 = \frac{1}{\log(1+\sigma^{-2})}$. Hence, it suffices that $T \geq O^*(\frac{\sigma^2 \gamma_T \beta_T}{\epsilon^2} + \frac{C_1 \gamma_T \beta_T}{\sigma^2} + 1)$. 


The choices $\eta(1) = 1$ and $r = \frac{1}{2}$ are made for mathematical convenience, and a similar result follows for any other choices $\eta(1) > 0$ and $r \in (0, 1)$, possibly with different constant factors.

As $\sigma^2 \to \infty$ (i.e., high noise), both of the above-mentioned bounds have noise dependence $O(\sigma^2)$, since $\log(1 + \alpha^{-1}) = O(\alpha^{-1})$ as $\alpha \to \infty$. On the other hand, as $\sigma^2 \to 0$ (i.e., low noise), $C_1$ is logarithmic, and Corollary 3.1 is significantly better provided that $\epsilon \ll \sigma$.

Choosing the Noise and Cost: Here we consider the setting that there is a domain of points $D_0$ that the reward function depends on, and alongside each point we can choose a noise variance $\sigma^2(k)$ ($k = 1, \ldots, K$). Hence, $D = D_0 \times \{1, \ldots, K\}$. Lower noise variances incur a higher cost according to a cost function $c(k)$.

**Corollary 3.2.** For each $k = 1, \ldots, K$, let $T^*(k)$ denote the smallest value of $T$ such that (17) holds with $\sigma^2(k)$ in place of $\sigma^2$, and with $\beta_T = 2 \log \frac{D^2 T^2 \delta^2 \pi^2}{6 \log_{\min} \max \max}$ . Then, under the preceding setting, there exist choices of $\beta(i)$ (not depending on $T$) such that we have $\epsilon$-accuracy with probability at least $1 - \delta$ once the cumulative cost reaches $\min_k c(k) T^*(k)$.

This result roughly states that we obtain a bound as good as that obtained by sticking to any fixed choice of noise level. In other words, every choice of noise (and corresponding cost) corresponds to a different version of a BO or LSE algorithm (e.g., [2][4]), and our algorithm has a similar performance guarantee to the best among all of those. This is potentially useful in avoiding the need for running an algorithm once per noise level and then choosing the best-performing one. Moreover, we found numerically that beyond matching the best fixed noise strategy, we can strictly improve over it by mixing the noise levels; see Section 4.

## 4 Experimental Results

We evaluate our algorithm in both the level-set estimation and Bayesian optimization settings.

**Parameter choices:** As with previous GP-based algorithms that use confidence bounds, our theoretical choice of $\beta(i)$ in TRuVAR is typically overly conservative. Therefore, instead of using (14) directly, we use a more aggressive variant with similar dependence on the domain size and time: $\beta(i) = a \log(||D||_{t(i)}^2)$, where $t(i)$ is the time at which the epoch starts, and $a$ is a constant. Instead of the choice $a = 2$ dictated by (14), we set $a = 0.5$ for BO to avoid over-exploration. We found exploration to be slightly more beneficial for LSE, and hence set $a = 1$ for this setting. We found TRuVAR to be quite robust with respect to the choices of the remaining parameters, and simply set $\eta(1) = 1, r = 0.1$, and $\delta = 0$ in all experiments (even though our theory requires $\delta > 0$).

**Level-set estimation:** For the LSE experiments, we use a common classification rule in all algorithms, classifying the points according to the posterior mean as $\hat{H}_t = \{x : \mu_t(x) \geq h\}$ and $\hat{L}_t = \{x : \mu_t(x) < h\}$. The classification accuracy is measured by the $F_1$-score (i.e., the harmonic mean of precision and recall) with respect to the true super- and sub-level sets.

We compare TRuVAR against the GP-based LSE algorithm [2], as well as the state-of-the-art straddle (STR) heuristic [6] and maximum variance rule (VAR) [2]. Descriptions of these algorithms can be found in the supplementary material. The LSE algorithm includes an exploration constant $\beta_t$, and we follow the recommendation in [2] of setting $\beta_t^{1/2} = 3$.

**Lake data (unit cost):** We begin with a data set from the domain of environmental monitoring of inland waters, consisting of 2024 in situ measurements of chlorophyll concentration within a vertical transect plane, collected by an autonomous surface vessel in Lake Zürich [18]. As in [2], our goal is to detect regions of high concentration. We evaluate each algorithm on a $50 \times 50$ grid of points, with the corresponding values coming from the GP posterior that was derived using the original data (see Figure 2a). We use the Matérn-5/2 ARD kernel, setting its hyperparameters by maximizing the likelihood on the second (smaller) available dataset. The level-set threshold $h$ is set to 1.5.

In Figure 2a we show the performance of the algorithms averaged over 100 different runs; here the randomness is only with respect to the starting point, as we are in the noiseless setting. We observe that in this unit-cost case, TRuVAR performs similarly to LSE and STR. All three methods outperform VAR, which is good for global exploration but less suited to level-set estimation.
Lake data (varying cost): Next, we modify the above setting by introducing pointwise costs that are a function of the previous sampled point \( x' \), namely, 
\[
c_v(x) = \begin{cases} 
0.25|x_1 - x'_1| + 4|x_2| + 1, & \text{if } x_1 \text{ and } x_2 \text{ are of the same type,} \\
0.5, & \text{otherwise},
\end{cases}
\]
where \( x_1 \) is the vessel position and \( x_2 \) is the depth. Although we did not permit such a dependence on \( x' \) in our original setup, the algorithm itself remains unchanged. Our choice of cost penalizes the distance traveled \( |x_1 - x'_1| \), as well as the depth of the measurement \( |x_2| \). Since incorporating costs into existing algorithms is non-trivial, we only compare against the original version of the LSE algorithm that ignores costs.

In Figure 2(b), we see that TruVaR significantly outperforms the LSE algorithm, achieving a higher F1 score for a significantly smaller cost. The intuition behind this can be seen in Figures 2(c) and 2(f), where we show the points sampled by TruVaR and LSE in one experiment run, connecting all pairs of consecutive points. LSE is designed to pick few points, but since it ignores costs, the distance traveled is large. In contrast, by incorporating costs, TruVaR tends to travel small distances, often even staying in the same \( x_1 \) location to take measurements at multiple depths \( x_2 \).

Synthetic data with multiple noise levels: In this experiment, we demonstrate Corollary 3.2 by considering the setting in which the algorithm can choose the sampling noise variance and incur the associated cost. We use a synthetic function sampled from a GP on a 50 \times 50 grid with an isotropic squared exponential kernel having length scale \( l = 0.1 \) and unit variance, and set \( h = 2.25 \). We use three different noise levels, \( \sigma^2 \in \{10^{-5}, 10^{-3}, 0.05\} \), with corresponding costs \( \{15, 10, 2\} \).

We run the LSE algorithm separately for each of the three noise levels, while running TruVaR as normal and allowing it to mix between the noise levels. The resulting F1-scores are shown in Figure 2(c). The best-performing version of LSE changes throughout the time horizon, while TruVaR is consistently better than all three. A discussion on how TruVaR mixes between the noise levels can be found in the supplementary material.
Bayesian optimization. We now provide the results of two experiments for the BO setting.

Synthetic data: We first conduct a similar experiment as that in [7, 10], generating 200 different test functions defined on $[0, 1]^2$. To generate a single test function, 200 points are chosen uniformly at random from $[0, 1]^2$, their function values are generated from a GP using an isotropic squared exponential kernel with length scale $l = 0.1$ and unit variance, and the resulting posterior mean forms the function on the whole domain $[0, 1]^2$. We subsequently assume that samples of this function are corrupted by Gaussian noise with $\sigma^2 = 10^{-6}$. The extension of TRUVAR to continuous domains is straightforward, and is explained in the supplementary material. For all algorithms considered, we evaluate the performance according to the regret of a single reported point, namely, the one having the highest posterior mean.

We compare the performance of TRUVAR against expected improvement (EI), GP-upper confidence bound (GP-UCB), entropy search (ES) and minimum regret search (MRS), whose acquisition functions are outlined in the supplementary material. We use publicly available code for ES and MRS [19]. The exploration parameter $\beta_t$ in GP-UCB is set according to the recommendation in [3] of dividing the theoretical value by five, and the parameters for ES and MRS are set according to the recommendations given in [10] Section 5.1.

Figure 3a plots the median of the regret, and Figure 3b plots the mean after removing outliers (i.e., the best and worst 5% of the runs). In the earlier rounds, ES and MRS provide the best performance, while TRUVAR improves slowly due to exploration. However, the regret of TRUVAR subsequently drops rapidly, giving the best performance in the later rounds after “zooming in” towards the maximum. GP-UCB generally performs well with the aggressive choice of $\beta_t$, despite previous works’ experiments revealing it to perform poorly with the theoretical value.

Hyperparameter tuning data: In this experiment, we use the SVM on grid dataset, previously used in [20]. A $25 \times 14 \times 4$ grid of hyperparameter configurations resulting in 1400 data points was pre-evaluated, forming the search space. The goal is to find a configuration with small validation error. We use a Matérn-5/2 ARD kernel, and re-learn its hyperparameters by maximizing the likelihood after sampling every 3 points. Since the hyperparameters are not fixed in advance, we replace $M_{t-1}$ by $D$ in (5) to avoid incorrectly ruling points out early on, allowing some removed points to be added again in later steps. Once the estimated hyperparameters stop to vary significantly, the size of the set of potential maximizers decreases almost monotonically. Since we consider the noiseless setting here, we measure performance using the simple regret, i.e., the best point found so far.

We again average over 100 random starting points, and plot the resulting validation error in Figure 3c. Even in this noiseless and unit-cost setting that EI and GP-UCB are suited to, we find that TRUVAR performs slightly better, giving a better validation error with smaller error bars.

5 Conclusion

We highlight the following aspects in which TRUVAR is versatile:

- **Unified optimization and level-set estimation:** These are typically treated separately, whereas TRUVAR and its theoretical guarantees are essentially identical in both cases.
- **Actions with costs:** TRUVAR naturally favors cost-effective points, as this is directly incorporated into the acquisition function.
- **Heteroscedastic noise:** TRUVAR chooses points that effectively shrink the variance of other points, thus directly taking advantage of situations in which some points are noisier than others.
- **Choosing the noise level:** We provided novel theoretical guarantees for the case that the algorithm can choose both a point and a noise level, cf., Corollary 3.2.
- **Multi-task settings:** Although we did not consider it in this paper, TRUVAR extends directly to multi-task settings in the same way as entropy search.

Hence, TRUVAR directly handles several important phenomena that are non-trivial to incorporate into myopic algorithms. Moreover, compared to other BO algorithms that perform a lookahead (e.g., ES and MRS), TRUVAR avoids the computationally expensive task of averaging over the posterior and/or measurements, and comes with rigorous theoretical guarantees.
References


A Variations of the TRUVAR Algorithm

Our algorithm TRUVAR can naturally be adapted to suit various settings, including the following:

- **Non-monotonic** $M_t$: We have defined our sets $M_t$ to become smaller on every time step. However, if $\beta_{(i)}$ is chosen aggressively, it may be preferable to replace $M_{t-1}$ by $D$ in $[5]-[6]$, in which case some removed points may be added back in depending on how the posterior mean changes between steps. We take this approach in the real-world BO example of Section 4 in which the kernel hyperparameters are learned online, so as to avoid incorrectly ruling out points early on due to mismatched hyperparameters.

- **Avoiding computing the acquisition function everywhere**: We found that instead of computing the acquisition function at every point in $D$, limiting the selection to points in $M_{t-1}$ has minimal effect on the performance. To reduce the computation even further, one could adopt a strategy such as that proposed in [21]: Take some relatively small number of points having the top GP-UCB or EI score, and then choose the point in that restricted subset having the highest score according to [5]. In fact, the numerical results in Figure 3 suggest that this may not only reduce the computation, but also improve the performance in the very early rounds by making the algorithm initially behave more like GP-UCB or EI.

- **Pure variance reduction**: Setting $\eta_{(1)} = 0$ yields a pure variance reduction algorithm, which minimizes the total variance within $M_t$ via a one-step lookahead. While our theory does not apply in this case, we found this choice to also work well in practice.

- **Implicit threshold for level-set estimation**: While we have focused on a threshold $h$ for level-set estimation that is fixed in advance, one can easily incorporate the ideas of [2] to allow for an implicit threshold which is equal to some constant multiple of the function’s maximum, which is random and unknown in advance.

- **Anticipating changes in $M_t$**: The acquisition function [3] computes the truncated variance reduction resulting from a one-step lookahead, but still sums over the previous set $M_{t-1}$. In order to make it more preferable to choose points that shrink $M_t$ faster, it may be preferable to instead sum over $M_{t-1,x}$, defined to be the updated set upon adding $x$. The problem here is that such an update depend on the next posterior mean, whose update requires sampling $f$. One solution is to average over the measurement as in [7]; alternatively, a simpler approach is to replace the observed value with its mean when doing this one-step lookahead, and then using the true observed function sample only when $x_1$ is actually chosen.

- **Non-Bayesian Setting**: While we focused on the Bayesian setting, analogous theorems to those in Section 3 apply for finite-domain functions having a low RKHS norm upon suitably changing the choice of $\beta_{(i)}$ in order to ensure valid confidence bounds. The changes are completely analogous to those given in [3]. Specifically, in place of $\beta_T = 2 \log \frac{D \bar{T}^2 \pi^2}{6 \tilde{\gamma}}$ (cf., Corollary 3.1) we have $\beta_T = 2B + 300 \gamma T \log^3 \frac{T}{\delta}$, where $B$ is an upper bound on the function’s RKHS norm. Accordingly, in place of (14) we have the condition

$$\beta_{(i)} \geq 2B + 300 \gamma T_{(i)} \log^3 \frac{T_{(i)}}{\delta}, \quad (18)$$

where $T_{(i)} = \left[ \frac{1}{c_{\min}} \sum_{i' \leq i} C_{(i')} + c_{\max} \right]$.

- **Continuous Domains**: Our algorithm also extends to compact domains such as $[0,1]^d$. The main challenge is that the summations in [3] become integrals that need to be approximated numerically. The simplest way of doing this is to approximate the integrals by summations over a finite number of representer points, e.g., a grid of values that cover the domain sufficiently densely. With this approach, an extension of our theory is also possible for sufficiently regular kernels in the sense of [3] Thm. 2), though we do not provide the details here. The idea is to choose the grid finely enough to ensure that $(1 + \delta) \eta_{(i)}$-confidence on the grid points ensures $(1 + 2\delta) \eta_{(i)}$-confidence (say) on all points. While this may require a very fine grid, this still only amounts to logarithmic factors in the regret bound. It is worth noting that in contrast with [3]...
Thm. 2], this approach does not only introduce the grid as part of the theoretical analysis, but also uses it in the algorithm.

• **Batch Setting:** As we show in the proof of Theorem 3.1, our algorithm can be interpreted as performing the first step of a greedy submodular covering problem at each time step. This leads to a very natural extension to the batch setting, in which multiple points (say, $k$ of them) are chosen at each time step: Simply perform the first $k$ steps of the greedy covering algorithm during each batch.

**B Further Details of Numerical Experiments**

**Other Algorithms Considered:** We outline the algorithms that TRUVAR is compared against; full details can be found in the cited papers. For level-set estimation, we have the following:

- The level-set estimation (LSE) algorithm [2] evaluates, at each iteration, the point that is not yet classified with the largest ambiguity: $x_t = \arg \max_{x \in M_{t-1}} \min \{ u_t(x) - h, h - \ell_t(x) \}$, where $u_t$ and $\ell_t$ are defined as in (4) with a parameter $\beta_t$ replacing $\beta(i)$. Here, similarly to our algorithm, $M_{t-1}$ is the set of points that have not yet been classified as having a value above or below the threshold $h$.
- The straddle (STR) heuristic [6] chooses $x_t = \arg \max_{x \in D} \mu_{t-1}(x) - |\mu_{t-1}(x) - h|$, favoring high-uncertainty points that are expected to have function values closer to $h$.
- The maximum variance rule (VAR) [2] simply chooses $x_t = \arg \max_{x \in D} \sigma_{t-1}(x)$.

For Bayesian optimization, we have the following:

- The expected improvement (EI) algorithm [11] chooses $x_t = \arg \max_{x \in D} \mathbb{E}_t[(f(x) - \xi_t) I \{ f(x) > \xi_t \}]$, where $\mathbb{E}_t[\cdot]$ denotes averaging with respect to the posterior distribution at time $t$, and $\xi_t$ is the best observed value so far. Since the posterior is Gaussian, the expectation can easily be expressed in closed form.
- The Gaussian Process Upper Confidence Bound (GP-UCB) algorithm [3] chooses the points with the highest upper confidence bounds, $x_t = \arg \max_{x \in D} \mu_{t-1}(x) + \beta_t \sigma_{t-1}(x)$, where $\beta_t$ is a parameter controlling the level of exploration performed.
- The Entropy Search (ES) algorithm can be interpreted as approximating the rule $x_t = \arg \min_{x \in D} h(f_{x^*\mid t,x})$, where $h(f) = \int_D f(x) \log \frac{1}{f(x)} dx$ denotes the differential entropy, and $f_{x^*\mid t,x}$ denotes the density function of the optimal action $x^*$ given the observations up to time $t$ along with the additional observation $x$. Intuitively, this rule seeks to minimize the uncertainty of $x^*$. Since its exact evaluation is intractable, it is approximated using Monte Carlo techniques to average with respect to the posterior distribution and the measurements.
- The Minimum Regret Search (MRS) algorithm [10] also resembles ES, but works with the expected regret instead of the differential entropy. Once again, Monte Carlo techniques are used to average with respect to the posterior distribution and the measurements.

**Efficiently Computing the Acquisition Function:** To compute the value of the acquisition function (3) for different $x \in D$, we need to compute $\sigma^2_{t-1|z}(M_{t-1}) \in \mathbb{R}^{|M_{t-1}|}$, i.e., the posterior
Lemma C.1. \textbf{Proof of General Result (Theorem 3.1)}

The acquisition function using DIRECT \cite{22}.

Continuous domains outlined in Appendix A, approximating the integrals over later stages.

\(\text{LSE, with } \sigma \text{ of the experiments, it gradually switches to lower noise levels in order to accurately determine the cost is roughly equally distributed across the three levels. To be more specific, we observed that each noise level by the end of the experiment, again averaged over 100 trials. We see that the function that was used in this experiment. Figure 4b plots the average cost spent by } T \text{RU}

\(\text{Details on LSE Experiment with Multiple Noise Levels: Figure 4a plots the randomly-generated function that was used in this experiment. Figure 4b plots the average cost spent by TRUVAR on each noise level by the end of the experiment, again averaged over 100 trials. We see that the cost is roughly equally distributed across the three levels. To be more specific, we observed that TRUVAR initially chooses high noise levels in order to cheaply explore, and throughout the course of the experiments, it gradually switches to lower noise levels in order to accurately determine the function values around the maximum. This is consistent with the behavior of the three version of LSE, with } \sigma^2 = 0.05 \text{ performing well in the early stages, but } \sigma^2 = 10^{-6} \text{ being preferable in the later stages.}

\textbf{Extension of TRUVAR for Synthetic BO Experiment} \text{We used the extension of TRUVAR to continuous domains outlined in Appendix A} \text{ approximating the integrals over } M_t \text{ by summations that are restricted to points on a uniformly-spaced } 50 \times 50 \text{ grid covering } [0, 1]^2. \text{ We optimized our acquisition function using DIRECT} \cite{22}.

\textbf{C \; Proof of General Result (Theorem 3.1)}

We begin with the following lemma from \cite{3}.

\textbf{Lemma C.1.} \cite{3} For each } t, \text{ define } \beta_t = 2 \log \frac{|D|e^2}{6\delta^2}. \text{ With probability at least } 1 - \delta, \text{ we have for all } x \text{ and } t \text{ that } |f(x) - \mu_t(x)| \leq \beta_t^{1/2}\sigma_t(x).

We conclude that in order for } \mu_t(\cdot) \pm \beta_t^{1/2}\sigma_t(\cdot) \text{ to provide valid confidence bounds, it suffices to ensure that } \beta_{(i)} \geq \beta_t \text{ for all } t \text{ in epoch } i. \text{ From } \cite{14}, \text{ we see that this is true provided that the time taken to reach the end of the } i\text{-th epoch is at most } \frac{1}{c_{\text{min}}} \sum_{i' \leq i} C(i'). \text{ Since } c_{\text{min}} \text{ is the minimum pointwise cost, this holds provided that the cost incurred in epoch } i \text{ is at most } C(i). \text{ We therefore let } \beta_{(i)} \text{ be chosen accordingly (cf. } \cite{14}, \text{ leaving us only to bound } C(i) \text{ itself.}

We connect TRUVAR with the following budgeted submodular covering problem\footnote{Recall that } S \text{ may contain duplicates, and these are counted multiple times accordingly in the definitions of both } c(S) \text{ and } g_t(S). \text{ All of our equations can be cast in terms of standard sets (without duplicates) by expanding } D \text{ to } D \times \{1, \ldots, N\} \text{ for any integer } N \text{ that is larger than the maximum number of points that are chosen throughout the course of the algorithm.}

\begin{equation}
\text{minimize } S \quad c(S) \quad \text{subject to } g_t(S) = g_{t, \text{max}},
\end{equation}

where

\begin{equation}
g_t(S) = \sum_{\mathcal{X} \in M_{t-1}} \max \left\{ \sigma_{t-1}^2(\mathcal{X}), \frac{\eta^2_{(i)}}{\beta_{(i)}} \right\} - \sum_{\mathcal{X} \in M_{t-1}} \max \left\{ \sigma_{t-1|S}^2(\mathcal{X}), \frac{\eta^2_{(i)}}{\beta_{(i)}} \right\},
\end{equation}

\end{equation}

and where $g_{t, \text{max}}$ is the highest possible value of $g_t(S)$ over arbitrarily large $S$, i.e., it is the value obtained once all of the summands in the second summation in (23) have saturated to $\frac{\eta_i^2}{\beta_i}$:

$$g_{t, \text{max}} = \sum_{\mathcal{S} \in M_{t-1}} \left( \max \left\{ \sigma_{t-1}^2(\mathcal{S}), \frac{\eta_i^2}{\beta_i} - \frac{\eta_i^2}{\beta_i} \right\} \right)$$

(24)

$$= \sum_{\mathcal{S} \in M_{t-1}} \max \left\{ 0, \sigma_{t-1}^2(\mathcal{S}) - \frac{\eta_i^2}{\beta_i} \right\}$$

(25)

We thus refer to $g_{t, \text{max}}$ as the \textit{excess variance}; see Figure 5 for an illustration. Note that each time instant $t$ corresponds to a different function $g_t(S)$, and we are considering sets $S$ of an arbitrary size even though our algorithm only chooses one point at each time instant.

By our assumption on the submodularity of the variance reduction function, and the fact that taking the minimum with a constant \cite{16} preserves submodularity, $g_t(S)$ is also submodular. It is also easily seen to be monotonically increasing, and normalized in the sense that $g_t(\emptyset) = 0$.

Our selection rule (3) at time $t$ can now be interpreted as the first step in a greedy algorithm for solving the budgeted submodular optimization problem (22); specifically, the greedy rule optimizes the objective value per unit cost. To obtain performance guarantees, we use Lemma 2 of \cite{23} specialized to $|\mathcal{S}| = 1$, which reads as follows in our own notation:

$$g_t(\{x_t\}) \geq c(x_t) = c(S_t^*),$$

(26)

where $S_t^*$ is an optimal solution to (22), and hence $g_t(S^*) = g_{t, \text{max}}$. Here $x_t$ is the point chosen greedily by our algorithm.

We now consider the behavior of the excess variance $g_{t, \text{max}}$ in a single epoch, i.e., the duration of a single value of $i$ in the algorithm. We claim that for $t$ and $t + 1$ in the same epoch, we have

$$g_{t+1, \text{max}} \leq g_{t, \text{max}} - g_t(\{x_t\}).$$

(27)

To see this, we note from (23)–(24) that this would hold with equality if we were to have $M_t = M_{t-1}$, since by definition we have $\sigma_t^2(\mathcal{S}) = \sigma_{t-1}^2(\mathcal{S})$. We therefore obtain (27) by recalling that $M_t$ is decreasing in $t$ with respect to inclusion, and noting from (25) that any given $g_{t, \text{max}}$ can only decrease when $M_{t-1}$ is smaller.

Combining (26)–(27) gives

$$g_{t+1, \text{max}} \leq \left(1 - \frac{c(x_t)}{c(S_t^*)}\right) g_{t, \text{max}}.$$  

(28)

We also note that $c(S_{t+1}^*) \leq c(S_t^*)$, which follows since $\sigma_t^2$ is decreasing in $t$ and $M_i$ is shrinking in $t$, and therefore at time $t + 1$ a smaller cost is required to ensure that all terms in the second

\footnote{The minimum becomes a maximum after negation.}
summation of \((23)\) have saturated to \(\frac{\eta^2_{(i)}}{\rho_{(i)}}\). Hence, and applying \((28)\) recursively, we obtain for \(t\) and \(t + \ell\) in the same epoch that

\[
\frac{g_{t+\ell,\max}}{g_{t,\max}} \leq \prod_{t'=t+1}^{t+\ell} \left( 1 - \frac{c(x_{t'})}{c(S^i_{t'})} \right) \\
\leq \exp \left( - \frac{\sum_{t'=t+1}^{t+\ell} c(x_{t'})}{c(S^i_{t'})} \right),
\]

where we have applied the inequality \(1 - \alpha \leq e^{-\alpha}\). Moreover, the total cost incurred by choosing these points is precisely \(\sum_{t'=t+1}^{t+\ell} c(x_{t'})\). Thus, letting \(t_0\) be the first time index in the epoch, we find that in order to remove all but a proportion \(\gamma\) of the initial excess variance \(g_{t_0,\max}\), it suffices that the cost \(C_{(i)}\) incurred in epoch \(i\) satisfies

\[
C_{(i)} \geq c(S^*_{t_0}) \log \frac{1}{\gamma}.
\]

Next, we observe that since the posterior variance is upper bounded by one due to the assumption \(k(x, x) = 1\), the initial excess variance \(g_{t_0,\max}\) is upper bounded by \(g_{t_0,\max} \leq |M_{t_0-1}|\), the size of the set of potential maximizers at the start of the epoch. Recalling that \(\overline{M}_{(i-1)} \geq M_{t_0-1}\) (see \((9)\)–\((10)\)), we also have \(g_{t_0,\max} \leq |\overline{M}_{(i-1)}|\). It follows that if we choose

\[
\gamma = \frac{2^2 \eta^2_{(i)}}{\beta_{(i)} |\overline{M}_{(i-1)}|},
\]

then removing all but a proportion \(\gamma\) of \(g_{t_0,\max}\) also implies removing all but \(\frac{2^2 \eta^2_{(i)}}{\beta_{(i)} |\overline{M}_{(i-1)}|}\) of it. In other words, if at time \(t\) we have incurred a cost in epoch \(i\) satisfying \((31)\) with \(\gamma\) as in \((32)\), then we must have \(g_{t,\max} \leq \frac{2^2 \eta^2_{(i)}}{\beta_{(i)} |\overline{M}_{(i-1)}|}\).

Removing all of the excess variance would imply \(\eta_{(i)}\)-confidence at all points in \(M_i\). In the worst case, the remaining excess variance \(\frac{2^2 \eta^2_{(i)}}{\beta_{(i)} |\overline{M}_{(i-1)}|}\) is concentrated entirely on a single point, in which case its confidence is upper bounded by \(\sqrt{1 + \frac{2^2 \eta^2_{(i)}}{\beta_{(i)} |\overline{M}_{(i-1)}|}}\), which is further upper bounded by \((1 + \delta)\eta_{(i)}\) due to the identity \(\sqrt{1 + \alpha^2} \leq 1 + \alpha\).

Combining these observations, we conclude that in the \(i\)-th epoch, upon incurring a cost of at least

\[
C_{(i)} \geq c(S^*_{t_0}) \log \frac{|\overline{M}_{(i-1)}| \beta_{(i)}}{\delta^2 \eta^2_{(i)}},
\]

we are guaranteed to have \((1 + \delta)\eta_{(i)}\)-confidence for all points in \(M_i\). Having such confidence is precisely the condition used in the algorithm to move onto the next epoch, and we conclude that the epoch must end by (or sooner than) the time that \((33)\) holds. By definition, \(c(S^*_{t_0})\) is the smallest possible cost to uniformly shrink the posterior standard deviation within \(M_{t_0-1} \subseteq \overline{M}_{(i-1)}\) down to \(\frac{\eta^2_{(i)}}{\rho_{(i)}}\), thus coinciding with the definition in \((11)\). Therefore, in order for \((33)\) to hold, it suffices that

\[
C_{(i)} \geq C^\ast \left( \frac{\eta_{(i)}}{\beta_{(i)}^{1/2}}, \overline{M}_{(i-1)} \right) \log \frac{|\overline{M}_{(i-1)}| \beta_{(i)}}{\delta^2 \eta^2_{(i)}},
\]

We are now in a position to check the conditions for \(\epsilon\)-accuracy in Definition 3.1. In the case of BO, summing \((34)\) over all of the epochs such that \(4(1 + \delta)\eta_{(i-1)} > \epsilon\) yields \((12)\); recall from \((9)\) that after any epoch \(i\) such that \(4(1 + \delta)\eta_{(i)} \leq \epsilon\), all points are at most \(\epsilon\)-suboptimal. We also note that all true maxima must remain in \(M_i\), due to the fact that we have chosen \(\beta_{(i)}\) to ensure valid confidence bounds with high probability, and we only ever discard points that are deemed suboptimal according
to those bounds. For LSE, a similar conclusion follows from [10] by summing over all epochs such that $2(1 + \tilde{\delta})\eta(i-1) > \frac{\eta}{\tilde{\delta}}$, which is the same as $4(1 + \tilde{\delta})\eta(i-1) > \epsilon$. Once again, all points in $H_t$ and $L_t$ are correct due to the validity of our confidence bounds. In both cases, we add an additional term $\epsilon_{\max}$ to the right-hand side of (34) to account for the fact that once the cost exceeds the right-hand side, it could exceed it by any amount up to $\epsilon_{\max}$.

D  Simplified Result for the Homoscedastic and Unit-Cost Setting

Since we are focusing on unit costs $c(x) = 1$, the cost simply corresponds to the number of rounds $T$. To highlight this fact, we replace $C^*$ in (11) by

$$T^*(\xi, M) = \min_S \left\{ |S| : \max_{\pi \in M} \sigma_{0|S}(\pi) \leq \xi \right\}, \tag{35}$$

and similarly replace (36)–(38) by

$$T_{(i)} \geq T^* \left( \frac{\eta_{(i)}}{\beta_{(i)}^{1/2}}, \frac{M(i-1)}{\delta} \right) + 1 \tag{36}$$

$$\beta_{(i)} \geq 2 \log \frac{M(i-1)}{\delta} \left( \sum_{i'=i}^{M(i)} \pi_{(i')}^2 \right)^{1/2} \tag{37}$$

$$T_\epsilon = \sum_{i: \eta(i) \eta(i-1) > \epsilon} T_{(i)}. \tag{38}$$

In this section, we prove the following as an application of Theorem 3.1.

**Corollary D.1.** Fix $\epsilon > 0$ and $\delta \in (0, 1)$, define $\beta_T = 2 \log \frac{|D|\epsilon^2}{6\delta}$, and set $\eta(1) = 1$ and $r = \frac{1}{2}$. There exist choices of $\beta_{(i)}$ (not depending on the time horizon $T$) such that we have $\epsilon$-accuracy with probability at least $1 - \delta$ once the following condition holds:

$$T \geq \left( C_1 \gamma_T \beta_T \frac{96(1 + \tilde{\delta})^2}{\epsilon^2} + 2 \left\lceil \log_2 \frac{8(1 + \tilde{\delta})}{\epsilon} \right\rceil \right) \log \frac{16(1 + \tilde{\delta})^2 |D| \beta_T}{\delta^2 \epsilon^2} \tag{39}$$

$$= O^* \left( \frac{C_1 \gamma_T \beta_T}{\epsilon^2} + 1 \right), \tag{40}$$

where $C_1 = \frac{1}{\log(1 + \sigma^{-2})}$.

We bound the cardinality of $S$ in (35) by considering a procedure that greedily picks $\arg \max_{x \in M} \sigma_i(x)$. We claim that after selecting $k$ points according to this procedure to construct a set $S_k$, we have

$$\max_x \sigma_{0|S_k}^2(x) \leq C_1 \gamma_k \frac{\eta}{k}, \tag{41}$$

where $C_1 = \frac{1}{\log(1 + \sigma^{-2})}$. This is seen by writing

$$k \max_{x \in M} \sigma_{0|S_k}^2(x) = k \sigma_{0|S_k}^2(x_k) \tag{42}$$

$$\leq \sum_{j=1}^k \sigma_{0|S_j}^2(x_j) \tag{43}$$

$$\leq \frac{1}{\log(1 + \sigma^{-2})} \gamma_k, \tag{44}$$

where we respectively used that $x_k$ maximizes $\sigma_{0|S_k}$, that $\sigma_{0|S_j}(x_i)$ always decreases as more points are chosen, and the bound on the sum of variances of sampled points from [3] Lemma 5.4. Identifying $k$ with $T^*$, and $\max_{x \in M} \sigma_{0|S_k}(x)$ with $\xi = \frac{\eta}{\beta_1^2}$ (for some $\eta$ and $\beta$ to be specified), we obtain from (44) that

$$T^* \left( \frac{\eta}{\beta_1^2}, M \right) = \min \left\{ T^* : T^* \geq \frac{C_1 \gamma_T \beta_T}{\eta^2} \right\}. \tag{45}$$
Consider the value $T^\ast \left( \frac{\eta(i)}{\beta(i)}, \overline{M}_{(i-1)} \right)$ corresponding to the parameters $\eta = \eta(i)$ and $\beta = \beta(i)$ associated with epoch $i$. Letting $T = T_\epsilon$ denote the total time horizon, and using (38), we find that $\beta(i)$ in (37) can be upper bounded by $2 \log \frac{(D^2)_{\frac{1}{-2}}}{\delta^2}$, which is precisely $\beta_T$. By similarly using the monotonicity of $\gamma_i$, we obtain

$$ T^\ast \left( \frac{\eta(i)}{\beta(i)}, \overline{M}_{(i-1)} \right) \leq \frac{C_1 \gamma T \beta_T}{\eta(i)^2} + 1, \quad (46) $$

where the addition of one is to account for possible rounding up to the nearest integer.

Using (46), we find that in order for (50) to hold it suffices that

$$ T(i) \geq \left( \frac{C_1 \gamma T \beta_T}{\eta(i)^2} + 1 \right) \log \frac{|\overline{M}_{(i-1)}| \beta(i)}{\delta^2}, \quad (47) $$

Since we are only considering values of $i$ such that $4(1 + \overline{\})\eta_{(i-1)} > \epsilon$, and since $\overline{M}_{(i-1)} \subseteq D$, we can upper bound the logarithm by $\log \frac{16(1 + \overline{\})^2 |\overline{M}| \beta(i)}{\delta^2} > 1$, and hence in order for (47) to hold it suffices that

$$ T(i) \geq \left( \frac{C_1 \gamma T \beta_T}{\eta(i)^2} + 2 \right) \log \frac{16(1 + \overline{\})^2 |\overline{M}| \beta_T}{\delta^2 \epsilon^2}, \quad (48) $$

We also note that since $\eta(i) = \eta(1) r^{i-1}$, the condition $4(1 + \overline{\})\eta_{(i-1)} > \epsilon$ is equivalent to

$$ 4(1 + \overline{\})\eta(1) r^{i-2} > \epsilon \quad (49) $$

$$ \iff r^{i-2} > \frac{\epsilon}{4(1 + \overline{\})\eta(1)} \quad (50) $$

$$ \iff i < 2 + \log_{1/r} \frac{4(1 + \overline{\})\eta(1)}{\epsilon} \quad (51) $$

$$ \iff i \leq \left\lfloor \log_{1/r} \frac{4(1 + \overline{\})\eta(1)}{\epsilon} \right\rfloor + 1 \quad (52) $$

$$ \iff i \leq \left\lfloor \log_{1/r} \frac{4(1 + \overline{\})\eta(1)}{r \epsilon} \right\rfloor, \quad (53) $$

where in the last line we used $\log_{1/r} \frac{1}{\epsilon} = 1$. Summing (48) over all such $i$ in accordance with (38), we obtain following sufficient condition on the time horizon for $\epsilon$-accuracy:

$$ T \geq \left( \frac{C_1 \gamma T \beta_T}{\eta(1)^2} \sum_{i=1}^{\lfloor \log_{1/r} \frac{4(1 + \overline{\})\eta(1)}{\epsilon} \rfloor + 1} \frac{1}{\eta(i)} + 2 \left\lfloor \log_{1/r} \frac{4(1 + \overline{\})\eta(1)}{r \epsilon} \right\rfloor \right) \log \frac{16(1 + \overline{\抽检})^2 |\overline{M}| \beta_T}{\delta^2 \epsilon^2}. \quad (54) $$

Finally, we weaken this condition by upper bounding the summation as follows:

$$ \sum_{i=1}^{\lfloor \log_{1/r} \frac{4(1 + \overline{\抽检})\eta(1)}{\epsilon} \rfloor + 1} \frac{1}{\eta(i)} = \sum_{i=1}^{\lfloor \log_{1/r} \frac{4(1 + \overline{\抽检})\eta(1)}{\epsilon} \rfloor + 1} \frac{1}{\eta(i)^2} \quad (55) $$

$$ = \sum_{i=0}^{\lfloor \log_{1/r} \frac{4(1 + \overline{\抽检})\eta(1)}{\epsilon} \rfloor} \frac{1}{\eta(i)^2}, \quad (56) $$

$$ \leq \frac{1}{r^2(1 - r^2)} 16(1 + \overline{\抽检})^2 \quad \epsilon^2, \quad (57) $$

where the last line follows from the identity $\sum_{i=0}^{\lfloor \log_{1/r} A \rfloor} \frac{1}{r^2} \leq \frac{1}{r^2(1 - r^2)} A^2$ for $\log_{1/r} A > 0$. Substituting $r = \frac{1}{2}$ and $\eta(1) = 1$ concludes the proof; the former yields $\frac{1}{r^2(1 - r^2)} = \frac{16}{3} \leq 6$. 

16
E Proof of Improved Noise Dependence (Corollary 3.1)

The bound in [44] is based on the inequality [3]Lemma 5.4
\[
\frac{\sigma_i^2}{\sigma^2} \leq C_1 \log \left( 1 + \frac{\sigma_i^2}{\sigma^2} \right) \tag{58}
\]
for \( \sigma_i^2 \in [0, 1] \) (with \( C_1 = \frac{\sigma^2}{\log(1 + \sigma^2)} \)), which can be very loose when \( \sigma_i^2 \) is small. Our starting point to improve the noise dependence is to note that the following holds under the more restrictive condition \( \sigma_i^2 \leq \sigma^2 \):
\[
\sigma_i^2 = \sigma^2 \frac{\sigma_i^2}{\sigma^2} \leq 2 \sigma^2 \log \left( 1 + \frac{\sigma_i^2}{\sigma^2} \right), \tag{60}
\]
where we have used the fact that \( \alpha \leq 2 \log(1 + \alpha) \) for \( \alpha \in [0, 1] \).

The idea now is to use (60) in the epochs that are late enough so that \( \sigma_i^2 \leq \sigma^2 \), and (44) in the earlier epochs. Since \( (1 + \delta) \eta(i) \) is the confidence level obtained after epoch \( i \), and since \( \beta_i^{1/2} \sigma_i \) is the confidence level after time \( t \), we find that in order to ensure \( \sigma_i^2 \leq \sigma^2 \) it suffices that \( \frac{(1 + \delta)^2 \eta(i)}{\sigma} \leq \sigma^2 \). Moreover, our choice of \( \beta_i \) in (14) is always greater than one when \( |D| \geq 2 \) (which is a trivial condition), and hence we can weaken this condition to \( (1 + \delta)^2 \eta(i) \leq \sigma^2 \), and write
\[
\sum_i T(i) \leq \sum_{i : (1 + \delta)^2 \eta(i)} T(C(i)) + \sum_i T(i), \tag{61}
\]
where \( T(C(i)) \) denotes bound on \( T(i) \) in (45) based on (44), and \( T(C(i)) \) denotes the analogous bound based on (60) with \( 2 \sigma^2 \) in place of \( C_1 \). Similarly to (53), the first summation is a subset of the range \( i \leq \lfloor \log_{1/r} \frac{(1 + \delta)^2 \eta(i)}{r \sigma} \rfloor \), and it follows that the condition (54) may be replaced by
\[
T \geq \left( 2 \sigma^2 \gamma_T \beta_T \sum_{i = 1}^{\lfloor \log_{1/r} \frac{(1 + \delta)^2 \eta(i)}{r \sigma} \rfloor + 1} \frac{1}{\eta(i)} \right) + 2 \left( \log_{1/r} \frac{(1 + \delta)^2 \eta(i)}{r \sigma} \right) \log \frac{16(1 + \delta)^2 |D| \beta_T}{\delta^2 \epsilon^2} + \left( C_1 \gamma_T \beta_T \sum_{i = 1}^{\lfloor \log_{1/r} \frac{(1 + \delta)^2 \eta(i)}{r \sigma} \rfloor + 1} \frac{1}{\eta(i)} \right) \log \frac{16(1 + \delta)^2 |D| \beta_T}{\delta^2 \epsilon^2}. \tag{62}
\]
The first summation is handled in the same way as the previous subsection, and the second summation is upper bounded by writing
\[
\sum_{i = 1}^{\lfloor \log_{1/r} \frac{(1 + \delta)^2 \eta(i)}{r \sigma} \rfloor + 1} \frac{1}{\eta(i)} = \sum_{i = 1}^{\lfloor \log_{1/r} \frac{(1 + \delta)^2 \eta(i)}{r \sigma} \rfloor + 1} \frac{1}{\eta(i)^2} \leq \frac{(1 + \delta)^2}{r^2 (1 - r^2) \sigma^2,} \tag{63}
\]
where (63) follows since \( \eta(i) = \eta(i) r^{i-1} \), and (64) follows in the same way as (57). Once again, setting \( r = \frac{1}{2} \) and \( \eta(1) = 1 \) concludes the proof, with the third term in (17) coming from the identity
\[
2 \left( \log_{2} \frac{8(1 + \delta)}{\epsilon} \right) + 2 \left( \log_{2} \frac{2(1 + \delta)^2}{\epsilon \sigma} \right) \leq 2 \left( \log_{2} \frac{32(1 + \delta)^2}{\epsilon \sigma} \right).
\]

F Proof for the Setting of Choosing Noise (Corollary 3.2)

The proof follows the same arguments as those of Appendices D and E, with \( C^* \) being upper bounded in \( K \) different ways, one for each possible noise level. The choice \( \beta_T = 2 \log \frac{|D| \sigma^2}{6 \hat{c}_{min}^2} \) arises as a simple upper bound to the right-hand side of (14) resulting from the fact that \( \sum_{i=1}^{T} c(x_i) \leq c_{max} T \).