Bandit Theory and Thompson Sampling-Guided Directed Evolution for Sequence Optimization

Hui Yuan[1], Chengzhuo Ni[2], Huazheng Wang[3], Xuezhou Zhang[4], Le Cong[5], Csaba Szepesvári[6,7], and Mengdi Wang[7,8]

[1,2,4,8]Department of Electrical and Computer Engineering, Princeton University
[3]School of Electrical Engineering and Computer Science, Oregon State University
[5]Department of Pathology and Department of Genetics, Stanford University
[6]Department of Computing Science, University of Alberta
[7]DeepMind *†

Abstract

Directed Evolution (DE), a landmark wet-lab method originated in 1960s, enables discovery of novel protein designs via evolving a population of candidate sequences. Recent advances in biotechnology has made it possible to collect high-throughput data, allowing the use of machine learning to map out a protein's sequence-to-function relation. There is a growing interest in machine learning-assisted DE for accelerating protein optimization. Yet the theoretical understanding of DE, as well as the use of machine learning in DE, remains limited. In this paper, we connect DE with the bandit learning theory and make a first attempt to study regret minimization in DE. We propose a Thompson Sampling-guided Directed Evolution (TS-DE) framework for sequence optimization, where the sequence-to-function mapping is unknown and querying a single value is subject to costly and noisy measurements. TS-DE updates a posterior of the function based on collected measurements. It uses a posterior-sampled function estimate to guide the crossover recombination and mutation steps in DE. In the case of a linear model, we show that TS-DE enjoys a Bayesian regret of order $\widetilde{O}(d^2\sqrt{MT})$, where d is feature dimension, M is population size and T is number of rounds. This regret bound is nearly optimal, confirming that bandit learning can provably accelerate DE. It may have implications for more general sequence optimization and evolutionary algorithms.

1 Introduction

Protein engineering means to design a nucleic acids sequence for maximizing a utility function that measures certain fitness or biochemical/enzymatic properties, i.e., stability, binding affinity, or catalytic activity. Due to the combinatorial sequence space and lack of knowledge about the sequence-to-function map, engineering and identifying optimal protein designs were a quite daunting task. It is only until recently that synthesis of nucleic acid sequences and measurement of protein

*Authors' emails are: {huiyuan, cn10, xz7392, mengdiw}@princeton.edu, huazheng.wang@oregonstate.edu, congle@stanford.edu, szepesva@ualberta.ca.

†Mengdi Wang acknowledges support by NSF grants DMS-1953686, IIS-2107304, CMMI-1653435, and ONR grant 1006977. Le Cong acknowledges support by NIH grant R35-HG011316, and Donald and Delia Baxter Foundation grant, and NSF grant DMS-1953686. Csaba Szepesvári gratefully acknowledges the funding from Natural Sciences and Engineering Research Council (NSERC) of Canada, "Design.R AI-assisted CPS Design" (DARPA) project and the Canada CIFAR AI Chairs Program for Amii.

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

function became reasonably scalable [41, 57], allowing rational optimization or directed evolution of protein designs. Nonetheless, because of the complex landscape of protein functions and the bottleneck of wet-lab experimentation, this remains a very difficult problem.

Directed evolution (DE), one of the top molecular technology breakthroughs in the past century, demonstrates human's ability to engineer proteins at will. DE is a method for exploring new protein designs with properties of interest and maximal utility, by mimicking the natural evolution. It works by artificially evolving a population of variants, via mutation and recombination, while constantly selecting high-potential variants [8, 9, 33, 25, 49, 41]. The development of directed evolution methods was honored in 2018 with the awarding of the Nobel Prize in Chemistry to Frances Arnold for evolution of enzymes, and George Smith and Gregory Winter for phage display [4, 47, 53]. See Figure 1.1 for illustrations of mutation and crossover recombination.

DE practitioners' major considerations center on cost and data quality. First, the ability to synthesize and mutate new biological sequences have been exponentially improved thanks to synthetic chemistry advances. Second, given a population of sequences S, selecting and identifying the set of optimal sequences is straightforward, using low-cost parallel sequencing which works well with pooled selection assays. Third, using pooled measurement to evaluate the average value of protein function (mean fitness) over a population S is generally easy, as such bulk measurements is low-cost and high-quality. Finally, querying $f(x)$ for a given x is often expensive and time-consuming, and the cost adds up quickly if many queries are needed. It can be desirable to perform this procedure in small-scale batches to optimize time and resource consumption.

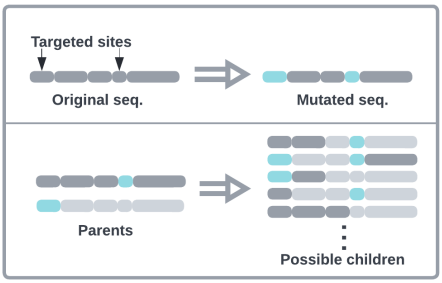

Figure 1.1: Illustration of mutation and crossover recombination. Mutating a sequence means to replace a targeted or random entry (site) by a random or designated value. Recombination involves two or multiple sequences. For example, parent sequences can crossover, exchange subsequences and generate children.

Such difficulties have motivated scientists to apply machine learning approaches to accelerate DE, beginning with Fox et al. [16] and followed by many. Recent development of directed evolution have increasingly utilized *in silico* exploration and machine learning beyond experimental approaches [57, 15, 12, 45, 18, 50, 46]. While these attempts have proved to be successful in simulation and sometimes in real experiments, little is known about the statistical theory of DE.

In this paper, a primary objective is to bridge the directed evolution process with bandit learning theory. In particular, we want to express machine learning-assisted DE as a bandit optimization process, with a theoretical justification. Further, we aim to understand how a machine learning model, as simple as linear, can accelerate DE and reduce the overall cost of evaluation. Specifically, we propose a Bayesian bandit model for DE, namely the Thompson Sampling-guided Directed Evolution framework, which combines posterior model sampling with directed mutation and recombination. The theoretical analysis shows that the crossover selection mimics an optimization iteration, and the optimization progress is proportional to a level of population diversity. In the case of the linear model, we establish a Bayesian regret bound $\widetilde{O}(d^2\sqrt{MT})$[3] that depends polynomially on feature dimension d[4], and optimally in batch size M and time steps T. We finally harmonize our theoretical analysis with a set of simulation and real-world experiment.

Important Remark The scope of this work is to provide a simplified mathematical model and basic theoretical understanding of an evolutionary-based process that is common in directed evolution. We emphasize that our framework is a theoretical simplification, assuming linear objective over a hy-

[3] $\widetilde{O}(\cdot)$ ignores the logarithmic terms.

[4] Our TS-DS regret has two extra $\sqrt{d}$ factors compared to the classic result that the optimal regret for linear bandits is of $\widetilde{O}(d\sqrt{MT})$. One $\sqrt{d}$ comes from our problem setting where the l_2 norm of each action is $O(\sqrt{d})$. Another factor of $\sqrt{d}$ is due to the evolutionary nature of DE. See the remark in §5.2 for more details.

percube. In real-world experimental systems, one needs to consider prior knowledge about the system and applying any machine learning method would require careful calibration and customization.

2 Related work

From a theoretical perspective, our analysis is related to the literature on evolutionary algorithms and linear bandits.

Evolutionary algorithm. The success of DE motivated a large body of works on evolutionary algorithms for optimization. Evolutionary algorithm (EA) [6] is a large class of randomized optimization algorithms, based on the heuristic of mimicking natural evolution. Despite many variants, a typical EA usually maintains a population of solutions and improves the solutions by alternating between reproduction step which produces new offspring solutions, and selection step where solutions are evaluated by the objective function and only the good ones are saved to the next round. Theoretical understandings of EA are focusing on specific EAs, among which the most well-studied setting is $(1+1)$-EA, with parent population size and offspring population size are both 1 to optimize linear objective function on the Boolean space $\{0,1\}^d$, see [14, 24, 27, 28, 35, 55]. EA analysis focuses on optimization and reducing the running time instead of minimizing total regret as in bandit theory. There are other results on population based EAs, such as $(1+\lambda)$-EA [11, 19], $(\mu+1)$-EA [54] and the most general $(\mu+\lambda)$-EA, where μ and λ represent the parent population size and the offspring population size respectively. However, this group of works only adopted mutation. The understanding of the role played by recombination in evolutionary algorithms was left as blank in the $(\mu+\lambda)$-EA framework, while our paper provides a population-based regret minimization analysis with both mutation and recombination.

There are a few works [30, 29, 51, 32] studying EAs with recombination (which are also called genetic algorithms (GAs)). However, their algorithms and analysis are tailored to artificial test objectives and the results are not able to generalize even to linear objectives. Recently, the running time analysis of some natural EAs with recombination has been conducted [39, 40], but still their results are constrained under specific objectives such as ONEMAX and JUMP. We refer readers to the book by [59] for a more comprehensive review of EA.

Linear bandits. Bandit is a powerful framework formulating the sequential decision making process under uncertainties. Under this framework, linear bandits is a central and fruitful branch where in each round a learner makes her decision and receives a noisy reward with its mean value modelled by a linear function of the decision, aiming to maximize her total reward (or minimize total regret equivalently) over multiple rounds [5, 36, 1], with extension to sparse linear bandits [23] and linear MDP [58]. In the same spirit, the process evolving a population of genetic sequences to maximize a linear utility over the evolution trajectory, while getting access to noisy utility values through evaluating sequences along the way, can be mathematically formulated from the perspective of linear bandits. One of the main solutions in linear bandits is the upper confidence bound-based (UCB) strategy represented by LinUCB [36], where the learner makes decision according to upper confidence bounds of the estimated reward and the accumulated regret is proven to be $\widetilde{O}\left(d\sqrt{T}\right)$. A similar strategy is optimism in the face of uncertainty (OFU) principle in Abbasi-Yadkori et al. [1]. The other approach is the Thompson Sampling (TS) strategy, which randomizes actions on the basis of their probabilities to be optimal. Russo and Van Roy [43] proved the Bayesian regret of TS algorithm is also of order $\widetilde{O}\left(d\sqrt{T}\right)$. And there are more results on the regret of TS(-like) algorithms solving linear bandits in the frequentist view [3, 2, 21, 3]. We also refer readers to the book by [34] for a delicate review of bandit theory.

Non-evolutionary methods for protein sequence design. Though our framework applies to an evolution-based DE process, there exist many other methods that are not evolution-based. Protein engineering is a rich field and it is not restricted to methods that are based on mutagenesis and recombination. Protein sequence engineering constantly evolves as new bio-technologies keep emerging. For example, new biotechnology makes it possible to synthesize specific variants and operate on the combinatorial space likewise with high-throughput method, and this allows directly applying a Gaussian process bandit algorithm [42]. See [57] for a high-level survey of this active

area of research, and see [17, 7] for more examples. This active and exciting field brings many new opportunities for machine learning.

Remark. It is important to note that our problem is *not* a multi-armed bandit problem. In bandits, one can choose actions freely from the full action set. However, in biological experiments, it is expensive to synthesize a new protein design sequence out of thin air. Instead, mutation and recombination are used to generate new designs easily at a low cost. Thus our algorithm can only guide the selection step in the DE process. Its regret is not directly comparable with the regret of multi-arm bandits. To the best of our knowledge, this is the first work that studies the bandit theory and regret bound of mutation and recombination-enabled DE.

3 Bandit model for directed evolution

3.1 Process overview

We illustrate the Thompson Sampling-guided Directed Evolution (TS-DE) process in Figure 3.1. A population S_t at time t consists of M candidate sequences. It evolves via mutation, crossover recombination, selection, and function evaluation to the next generation S_{t+1}. The mutation and crossover selection are guided using a learnt function $f_{\tilde{\theta}_t}$, in order to filter out unwanted candidates and keep only a small batch for costly evaluation. Collected data are fed into a Thompson Sampling module for posterior update of $f_{\tilde{\theta}_t}$. Full details of the mutation, crossover selection, and Thompson Sampling modules will be given in Section 4.

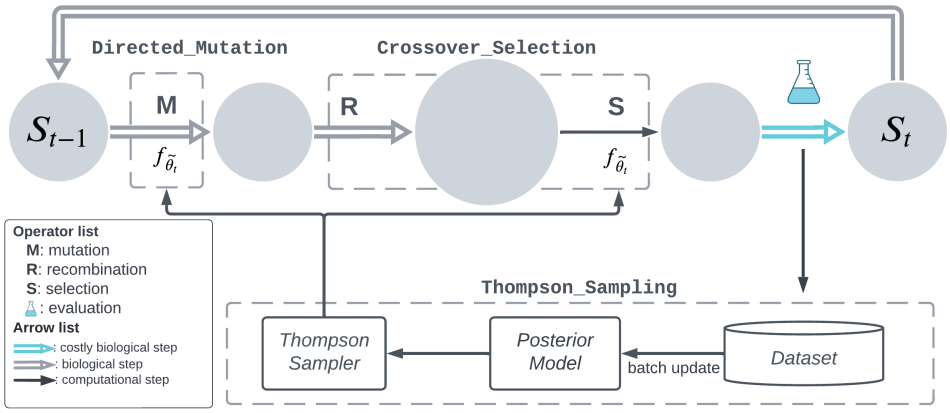

Figure 3.1: **Thompson sampling-guided directed evolution**

3.2 Motif feature, utility model, recombination and mutation operators

A genetic sequence comprises of functional motifs, i.e., functional subsequences that may encode particular features of protein, also known as protein motifs [37, 48, 10]. Such genetic motifs are known to be "evolutionarily conserved", in the sense that they tend to evolve as units, under mutation and recombination.

Suppose a genetic sequence seq is made up of d genetic motifs, given by $seq = (seq_{(1)}, seq_{(2)}, \cdots, seq_{(d)})$. Machine learning models for protein utility prediction are often based on motif features [56, 10, 38]. Let $\mathcal{X}$ be the space of genetic sequences of interest. We assume that a binary motif feature map is given, defined as follows.

Definition 3.1 (Binary Motif Feature Embedding)**.** Let ϕ be the genetic motif feature map given by:

$$\phi : \mathcal{X} \to \{0,1\}^d, \quad \phi(seq) := (\phi_1(seq_{(1)}), \cdots, \phi_d(seq_{(d)})) \tag{3.1}$$

such that at each dimension i, $\phi_i(seq_{(i)})$ is a binary feature of motif $seq_{(i)}$.

The binary motif feature provides a minimalist abstraction for evolutionary processes where $0, 1$ correspond to favorable and nonfavorable directions, respectively, for each motif. Theoretical analysis for evolutionary optimization algorithms made the same assumption and viewed binary sequence optimization as a fundamental problem [14, 24, 27, 28, 35, 55].

Since a protein function is largely determined by its motif, it is common to model the protein utility $f : \mathcal{X} \to \mathbb{R}$ as a function of motif features, i.e., $f(seq) := f_{\theta^\star}(x), x = \phi(seq), \forall seq \in \mathcal{X}$, under a parameterization by $\theta^\star$ [16, 57, 45, 18].

In this work, we study the most elementary Bayesian linear model, where f is a linear model parameterized by θ^* with a Gaussian prior, given as follows.

Assumption 3.2. (Linear Bayesian Utility Model) Assume the utility $f_{\theta^\star}$ is a linear function parameterized by $\theta^\star \in \mathbb{R}^d$, which is sampled from a Gaussian prior, i.e.

$$f_{\theta^\star}(x) = \langle \theta^\star, x \rangle, \qquad \theta^\star \sim \mathcal{N}(\mathbf{0}, \lambda^{-1}\mathbf{I}), \quad \lambda > 0. \tag{3.2}$$

Since motifs tend to mutate and recombine with one another in units, it is often sufficient to focus on recombination and mutation on the motif level, rather than on the entry level. Further, recombination that breaks a motif often result in insignificant low-fitness descendants. Therefore, it suffices to focus on motif-level directed evolution for simplicity of presentation and theory. For theoretical simplicity, we define recombination and mutation operators **on the motif level**:

Definition 3.3 (Directed Mutation Operator)**.** Let x be the motif feature sequence, $\mathcal{I} \subset [d]$ be a collection of targeted sites and $\mu \in (0, 1)$ be a mutation rate. The mutation operator $\texttt{Mut}(x, \mathcal{I}, \mu)$ generates a sequence x' such that while for $\forall j \notin \mathcal{I}, x'_j = x_j$, for $\forall i \in \mathcal{I}, x'_i$ is independently induced to be

$$\begin{cases} x'_i \sim \text{unif}(\{0, 1\}), & \text{w.p. } \mu, \\ x'_i = x_i, & \text{otherwise.} \end{cases} \tag{3.3}$$

Definition 3.4 (Recombination Operator)**.** Let x, y be the motif features associated with two parental genetic sequences. The recombination operator $\texttt{Rcb}(x, y)$ generates a child sequence z such that z_i's are independent and

$$z_i = \begin{cases} x_i & \text{w.p. } \frac{1}{2} \\ y_i & \text{w.p. } \frac{1}{2} \end{cases}, \ \forall i \in [d]. \tag{3.4}$$

We remark that Definitions $3.3, 3.4$ are *mathematical simplifications* of their real-world counterparts. In real world, mutation and recombination can take various forms depending on the context. In our analysis, we define them in a minimalist-style to keep theory generalizable and interpretable.

3.3 Regret minimization problem formulation

Evaluating the protein function for a design sequence x is a most costly and time-consuming step in protein engineering. In the DE process, we consider that regret is incurred only when sequences are evaluated. We also assume that each evaluation is subject to a Gaussian noise with known variance.

Assumption 3.5. (Noisy Feedback) Upon querying the utility of x, we get an independent noisy evaluation given by

$$u(x) \sim \mathcal{N}(f_{\theta^\star}(x), \sigma^2). \tag{3.5}$$

Our goal is to minimize the Bayesian regret, i.e., the cumulative sum of optimality gaps between evaluated sequences and the optimal.

Definition 3.6 (Bayesian Regret)**.** Denote by $f_{\theta^\star}(x^\star)$ the optimal utility value over $\mathcal{X}$, $\{x_{t,i}\}_{i=1}^M$ are the evaluated individuals in each iteration. Throughout T iteration, the accumulated regret is defined as

$$\text{BayesRGT}(T, M) = \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{M}(f_{\theta^\star}(x^\star) - f_{\theta^\star}(x_{t,i}))\right],$$

where M is number of sequences selected for evaluation per timestep, and $\mathbb{E}$ is taken over the prior of $\theta^\star$ and all randomness in the DE process.

4 Thompson Sampling-guided directed evolution (TS-DE)

We restate our goal as to direct a population of genetic sequence to evolve towards higher utility value, until its population-average converges to the optimum $f_{\theta^\star}(x^\star)$. Our knowledge of f is to be learned from noisy evaluations of selected sequences along the way. In this section, by integrating the biological technique - directed evolution - with Thompson Sampling, a Bayesian bandit method, we propose the Thompson Sampling-guided Directed Evolution algorithm (TS-DE) as shown in Alg.1, where in each round Thompson sampling gives an estimate of $\theta^\star$, based on which key operators of DE: mutation, recombination and selection are implemented.

4.1 Crossover-then-selection and directed mutation

Pairwise crossover is a most common type of recombination in natural evolution. Let x, y be a random pair of parents, and let $z = \texttt{Rcb}(x, y)$ be a child. If given a utility function f, we select z only if the child performs better than the parents' average. Module 1 formulates this procedure.

Module 1 `Crossover_Selection`(f, S)

Inputs: utility function $f(x) = \langle \theta, x \rangle$, a population of sequences S
Initialization: $S' \leftarrow \emptyset$
while $|S'| < |S|$ **do**
  Sample x and y from S uniformly with replacement.
  Recombination: $z \leftarrow \texttt{Rcb}(x, y)$ (Definition 3.4).
  Selection: $S' \leftarrow S' \cup \{z\}$ if $f(z) \geq \frac{f(x)+f(y)}{2}$.
end while
Output: S'

Next we turn to designing the strategy for adding directed mutation under a given f as guidance and propose Module 2. An ideal mutation will diversify the population while preserving its fitness level as much as possible. So we add directed mutation to sites where the single site fitness over the population is less than of a uniformly distributed sequence. Formally, we only add mutation to site i if $\frac{1}{M}\sum_{x\in S} \theta_i \cdot x_i \leq \theta_i \cdot \bar{x}_i$, where $\bar{x}_i$ is the mean of uniformly random x_i.

Module 2 `Directed_Mutation`(f, S, μ)

Inputs: utility function $f(x) = \langle \theta, x \rangle$, a population of sequences S, mutation rate μ
Initialization: $\mathcal{I} \leftarrow \emptyset, \mathcal{S}' \leftarrow \emptyset$
for $i \in [d]$ **do**
  if $\frac{1}{M}\sum_{x\in S} \theta_i \cdot x_i \leq \theta_i \cdot \bar{x}_i$ **then**
    $\mathcal{I} \leftarrow \mathcal{I} \cup \{i\}$.
  end if
end for
Directed Mutation: $x' = \texttt{Mut}(x, \mathcal{I}, \mu)$ (Definition 3.3) and $S' \leftarrow S' \cup \{x'\}$ for all $x \in S$.
Output: S'

4.2 Full algorithm

Finally, we are ready to combine all modules and state the full algorithm in Algorithm 1. At each time step t, a posterior distribution is first computed using the data collected in history. Then we sample a $\widetilde{\theta}_t$ from the posterior and do the corresponding directed mutation and crossover selection using this sampled weight, and augment the dataset for the next iteration with the measurements of resulting new population. The procedure is repeated until the time limit T is reached.

5 Main results

In this section, we analyze the performance of TS-DE (Algorithm 1). We will show that the crossover selection module essentially mimics an optimization iteration that strictly improves the population's

Algorithm 1 Thompson Sampling-Guided Directed Evolution (TS-DE)

Inputs: number of rounds T, initial population $S_0 = \{x_{0,i}\}_{i=1}^M$ of size M, mutation rate μ, σ
Initialization: dataset $D_0 \leftarrow \emptyset$, $\Phi_{t-1} = 0$, $U_0 = 0$
for $t = 1$ to T **do**
  Posterior update

$$V_t = \frac{1}{\sigma^2}\Phi_{t-1}^\top \Phi_{t-1} + \lambda I, \qquad \widehat{\theta}_t = \frac{1}{\sigma^2} V_t^{-1} \Phi_{t-1}^\top U_{t-1}. \tag{4.1}$$

  Thompson Sampling $\widetilde{\theta}_t \sim \mathcal{N}(\widehat{\theta}_t, V_t^{-1})$.
  $S'_{t-1} = \texttt{Directed\_Mutation}(f_{\widetilde{\theta}_t}, S_{t-1}, \mu)$ (Module 2).
  $S_t = \texttt{Crossover\_Selection}(f_{\widetilde{\theta}_t}, S'_{t-1})$ (Module 1).
  Evaluation and data collection Evaluate the utilities of all individuals in S_t and $D_t \leftarrow D_{t-1} \cup \{x_{t,i}, u(x_{t,i})\}_{i=1}^M$. Update $\Phi_t^\top \leftarrow \left(\Phi_{t-1}^\top, x_{t,1}, \cdots, x_{t,M}\right)$, $U_t \leftarrow \left(U_{t-1}^\top, u(x_{t,1}), \cdots, u(x_{t,M})\right)^\top$.
  $t \leftarrow t + 1$.
end for

fitness along the designated direction. By using a Bayesian regret analysis, we show the DE modules, when combined with posterior sampling, can effectively optimize towards the best protein design while learning $\theta^\star$.

5.1 Crossover selection as an optimization iteration

Let f by any utility function, and let $F(S) := \text{avg}_{x \in S} f(x)$ denote the population average utility. Our first result states an ascent property showing that `Crossover_Selection` strictly improves the population average.

Theorem 5.1 (Ascent Property of Recombination-then-Selection)**.** Let $f(x) = \langle \theta, x \rangle$ and let S be a set of sequences. Let $S' = \texttt{Crossover\_Selection}(f, S)$, then it satisfies

$$\mathbb{E}\left[F(S')\right] \geq F(S) + \frac{\mathbb{E}_{x,y}\left[\|\theta \cdot (x-y)\|\right]}{2\sqrt{2}} \geq F(S) + \frac{1}{\sqrt{2d}} \sum_i |\theta_i| \,\text{Var}_i(S), \tag{5.1}$$

where $\text{Var}_i(S)$ denotes the variance of x_i when x is uniformly sampled from S.

Proof sketch. See Figure 5.1 for illustration. Given x and y, $z = \texttt{Rcb}(x, y)$ can be represented by $z = \frac{x+y}{2} + \frac{x-y}{2} \cdot e$, where the $\cdot$ denotes the entrywise multiplication between two vectors and $e = (e_i, \cdots, e_d)$ with e_i's being independent Rademacher variables. Then $f(z)$ equals $\frac{f(x)+f(y)}{2} + \frac{1}{2}\sum_{i=1}^d \theta_i \left(x_i - y_i\right) e_i$. After the selection step, the expected amount by which $f(z)$ exceeds its parents' average is at least $\frac{1}{2}\mathbb{E}\left[\left|\sum_{i=1}^d \theta_i \left(x_i - y_i\right) e_i\right|\right]$, which has a tight lower bound of $\frac{1}{2\sqrt{2}}\|\theta \cdot (x-y)\|$ according to Haagerup [20]. The full proof is given in Appendix C.1. ■

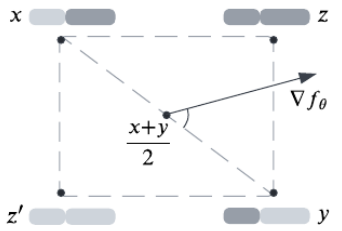

Figure 5.1: Ascent property of crossover recombination

Remark on diversity. Analysis above reveals an intriguing observation: the optimization progress of `Crossover_Selection` scales linearly with $\sum_i \theta_i \text{Var}_i(S)$, i.e., sum of per-motif variances across population S. It measures a level of "diversity" of S with respect to direction θ. More diverse population would enjoy larger progress from crossover selection. This observation is consistent with the natural evolution theory that diversity is key to the adaptability of a population to cope with evolving environment where fitness traits are essential [52].

5.2 Regret bound of TS-DE

Our main result is a Bayesian regret bound for TS-DE. Recall from Definition 3.6 that $\mathrm{BayesRGT}(T, M) = \mathbb{E}[\sum_{t=1}^{T}\sum_{i=1}^{M}(f_{\theta^\star}(x^\star) - f_{\theta^\star}(x_{t,i}))]$.

Theorem 5.2. Under Assumption 3.2 and 3.5, when the population size is sufficient s.t. $M = \Omega\left(\frac{\log(dT)}{\mu^2}\right)$, Alg.1 admits its Bayesian regret s.t.

$$\mathrm{BayesRGT}(T, M) = \widetilde{O}\left(\frac{d}{\mu\sqrt{\lambda}} \cdot d\sqrt{MT}\right). \tag{5.2}$$

If we let $\lambda = 1, \mu = 1/2, \sigma^2 = 1$, the Bayesian regret simplifies to $\widetilde{O}(d^2\sqrt{MT})$.

Remark on regret bound. Regret bound of Theorem 5.2 is optimal in M, T. For comparison, the Bayesian regret of Gaussian linear model is $\widetilde{O}(d\sqrt{T})$ [31], also in contextual linear bandit with batch update, the optimal regret is $\widetilde{O}(d\sqrt{MT})$ [22]. Our TS-DS regret has two extra factors of $\sqrt{d}$. One $\sqrt{d}$ is due to that the l_2 norm of our feature vectors are $\sqrt{d}$, while linear bandit theory often assumes feature to have norm 1. Another factor of $\sqrt{d}$ is due to the evolutionary nature of DE, i.e., TS-DE is not allowed to any possible action but have to select those from the evolving population.

5.3 Proof sketch

Main challenge. Classic bandit method/analysis does not apply to our setting, because each round of DE is limited to actions that are reachable by mutation and recombination based on the current population. It means that we cannot simply explore the optimistic actions that maximize each function estimate $f_{\widetilde{\theta}_t}$. This leads to an optimization gap that complicates the regret proof.

Denote by $x^\star$ and $x_t^\star$ the maximums of $f_{\theta^\star}$ and $f_{\widetilde{\theta}_t}$. Denote by $F_t^\star := f_{\widetilde{\theta}_t}(x_t^\star)$ the maximum value of $f_{\widetilde{\theta}_t}$ and denote by $F_t(S)$ the average value of $f_{\widetilde{\theta}_t}$ over set S.

Step 1: Regret decomposition. With expectation taken over all stochasticity, posterior sampling guarantees $\mathrm{BayesRGT}(T, M) = \sum_{t=1}^{T}\sum_{i=1}^{M}\mathbb{E}\left[f_{\widetilde{\theta}_t}(x_t^\star) - f_{\theta^\star}(x_{t,i})\right]$ since conditioned on data D_{t-1}, $f_{\theta^\star}(x^\star)$ and $f_{\widetilde{\theta}_t}(x_t^\star)$ are identically distributed. Then by breaking $f_{\widetilde{\theta}_t}(x_t^\star) - f_{\theta^\star}(x_{t,i})$ down to the sum of $f_{\widetilde{\theta}_t}(x_t^\star) - f_{\widetilde{\theta}_t}(x_{t,i})$ and $f_{\widetilde{\theta}_t}(x_{t,i}) - f_{\theta^\star}(x_{t,i})$, we decompose the total regret into

$$\mathrm{BayesRGT}(T, M) = M \cdot \underbrace{\mathbb{E}\left[\sum_{t=1}^{T}(F_t^\star - F_t(S_t))\right]}_{H_1} + \underbrace{\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{M}\langle\widetilde{\theta}_t - \theta^\star, x_{t,i}\rangle\right]}_{H_2}. \tag{5.3}$$

Step 2: Bounding H_1 using linear convergence. H_1 is the accumulated optimization error under a time-varying objective $f_{\widetilde{\theta}_t}$. After calling $S'_{t-1} = \texttt{Directed\_Mutation}(f_{\widetilde{\theta}_t}, S_{t-1}, \mu)$ and $S_t = \texttt{Crossover\_Selection}(f_{\widetilde{\theta}_t}, S'_{t-1})$ at step t, the ascent property (5.1) together with property of the mutation module yields a linear convergence towards $F_t^\star$, i.e., $\mathbb{E}\left[F_t^\star - F_t(S_t) \mid S_{t-1}, \widetilde{\theta}_t\right] \leq \gamma(F_t^\star - F_t(S_{t-1}))$ with a modulus of contraction $\gamma \in (0,1)$ s.t. $\frac{1}{1-\gamma} = O\left(\frac{\sqrt{d}}{\mu}\right)$. It follows that

$$F_t^\star - F_t(S_t) \leq \gamma\left[F_{t-1}^\star - F_{t-1}(S_{t-1})\right] + \text{error terms} + e_t,$$

where e_t is a martingale difference. Applying the above recursively to H_1, we get $H_1 \leq$

$$\underbrace{\frac{1}{1-\gamma} \cdot \mathbb{E}\left[F_1^\star - F_1(S_0)\right]}_{O(\frac{1}{1-\gamma})} + \underbrace{\mathbb{E}\left[\sum_{k=2}^{T}\gamma^{T-k+1}F_k^\star - \gamma^{T-1}F_1^\star\right]}_{O(\frac{1}{1-\gamma})} + \underbrace{\mathbb{E}\left[\sum_{t=1}^{T}\sum_{k=1}^{t-1}\gamma^{t-k}\left(F_k(S_k) - F_{k+1}(S_k)\right)\right]}_{\kappa},$$

which is dominated by term κ and $M \cdot \kappa \leq \frac{1}{1-\gamma} \cdot \sum_{t=1}^{T-1}\sum_{i=1}^{M}\left|\langle\widetilde{\theta}_t - \widetilde{\theta}_{t+1}, x_{t,i}\rangle\right| = O\left(\frac{1}{1-\gamma}H_2\right)$.

Step 3: Bounding H_2. H_2 is the accumulated prediction error of $\widetilde{\theta}_t$, which is a classic term to bound in bandit literature and is of $\widetilde{O}\left(d^{1.5}\sqrt{MT}\right)$ by using a batched self-normalization bound. ■

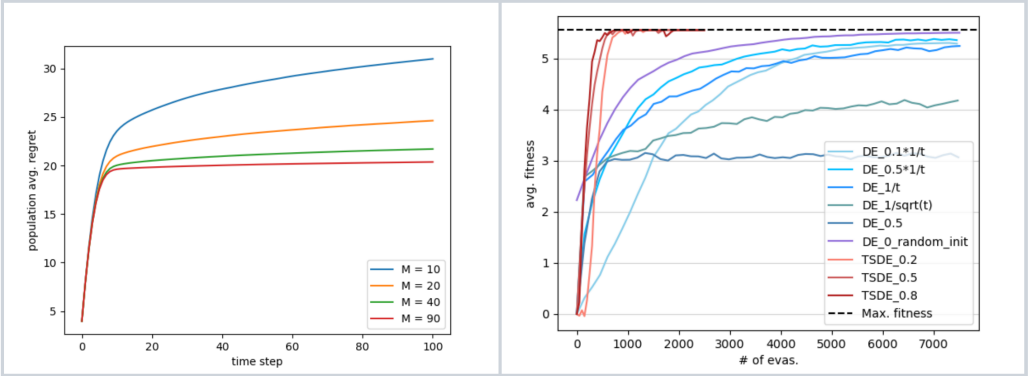

Figure 6.1: **Regret and fitness curves of TS-DE during evolution.** Left: Population-averaged regret with varying population sizes M. Each curve is averaged over 100 trials. Right: Fitness curves of TS-DE with varying values of μ, compared with basic DE with varying mutation rates. (The purple curve plots basic DE without mutation, we modified the initial population to be uniformly distributed in this case to make it non-trivial.)

6 Experiments

6.1 Simulation

We test the TS-DE by simulating the evolution of a population of sequences in $\{0,1\}^d$. We set the initial population to be all zeros, and set $\lambda = 1$, $\sigma = 1$.

Regret and convergence results. Figure 6.1 shows the regret curves and learning curves of TS-DE, with comparison to basic DE. In the left panel of Figure 6.1, we plot the population-averaged Bayesian regret of TS-DE with various values of M, where $d = 10$, $T = 100$ and $\mu = 0.8$. These results confirm our sublinear regret bounds. In the right panel of Figure 6.1, we tested TS-DE using various mutation rates, and compared them with a basic DE approach [5]. The comparison shows that TS-DE converges significantly faster, while the convergence of DE is much slower and very sensitive to mutation scheduling.

Visualizing the evolution of a population. We visualize the evolution trajectory of population S_t in one run of TS-DE, with $d = 40$, $M = 20$ and $\mu = 0.1$. In the left panel of Fig.6.2, we visualize the evolving high-dimensional population S_t by mapping them to 2D (via PCA and KDE density contour plot). In the right panel of Fig.6.2, we plot the fitness distribution of each S_t. These plots illustrate how TS-DE balances the exploration-exploitation trade-off: It guides S_t to "diversify" initially and then quickly approach and concentrate around a maximal solution.

6.2 Real-world experiment validation

Having demonstrated our approach with simulations, we use real-world experiments to showcase the validity and generalizability of our method. The TS-DE method is adapted to work with real-world motif features (continuous-valued instead of binary), linear model and multiple rounds of wet-lab experiments for optimizing a CRISPR design sequence. Our approach together with high-throughput experiment identified a high-performing sequence with 30+ fold improvement in efficiency. Notably, the optimized CRISPR designs generated by our DE approach was experimentally validated in [26] and demonstrated the real-world utility of our method. This technology is used for ex-vivo high-throughput single-cell barcoding with applications in genomics and drug discovery.

We postpone more details about this real-world validation to Appendix B.1 and Figure B.1.

[5] The basic DE approach does not employ any function estimate. It does random mutation with a predefined mutation rate and random crossover recombination. It evaluates every candidate sequence and uses the noisy feedback in replace of $f_{\tilde{\theta}}$ for selection.

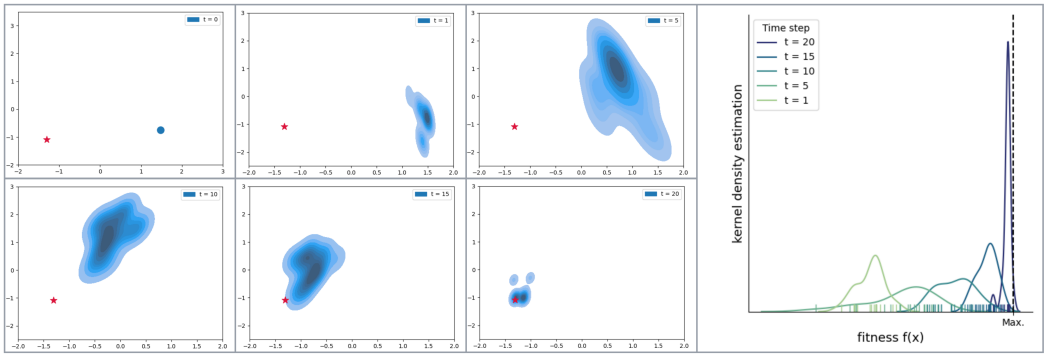

Figure 6.2: **Evolving population of TS-DE and fitness levels.** Left panels: Visualization of population evolution projected in 2D shown, taken at 6 snapshots. Right panel: The population's fitness distribution shifts towards optimal during evolution. ⋆ denotes the optimal solution.

References

[1] Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. *Advances in neural information processing systems*, 24:2312–2320, 2011.

[2] Marc Abeille and Alessandro Lazaric. Linear thompson sampling revisited. In *Artificial Intelligence and Statistics*, pages 176–184. PMLR, 2017.

[3] Shipra Agrawal and Navin Goyal. Thompson sampling for contextual bandits with linear payoffs. In *International conference on machine learning*, pages 127–135. PMLR, 2013.

[4] Frances H Arnold. Design by directed evolution. *Accounts of chemical research*, 31(3):125–131, 1998.

[5] Peter Auer. Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3:397–422, 2002.

[6] Thomas Back. *Evolutionary algorithms in theory and practice: evolution strategies, evolutionary programming, genetic algorithms*. Oxford university press, 1996.

[7] Claire N Bedbrook, Kevin K Yang, Austin J Rice, Viviana Gradinaru, and Frances H Arnold. Machine learning to design integral membrane channelrhodopsins for efficient eukaryotic expression and plasma membrane localization. *PLoS computational biology*, 13(10):e1005786, 2017.

[8] Keqin Chen and Frances H Arnold. Enzyme engineering for nonaqueous solvents: random mutagenesis to enhance activity of subtilisin e in polar organic media. *Bio/Technology*, 9(11): 1073–1077, 1991.

[9] Keqin Chen and Frances H Arnold. Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin e for catalysis in dimethylformamide. *Proceedings of the National Academy of Sciences*, 90(12):5618–5622, 1993.

[10] Joseph M Cunningham, Grigoriy Koytiger, Peter K Sorger, and Mohammed AlQuraishi. Biophysical prediction of protein–peptide interactions and signaling networks using machine learning. *Nature methods*, 17(2):175–183, 2020.

[11] Benjamin Doerr and Marvin Künnemann. Optimizing linear functions with the (1+ λ) evolutionary algorithm—different asymptotic runtimes for different instances. *Theoretical Computer Science*, 561:3–23, 2015.

[12] Janardhan Rao Doppa. Adaptive experimental design for optimizing combinatorial structures. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 4940–4945, 2021.

[13] Jennifer A Doudna and Emmanuelle Charpentier. The new frontier of genome engineering with crispr-cas9. *Science*, 346(6213):1258096, 2014.

[14] Stefan Droste, Thomas Jansen, and Ingo Wegener. On the analysis of the (1+ 1) evolutionary algorithm. *Theoretical Computer Science*, 276(1-2):51–81, 2002.

[15] Clara Fannjiang and Jennifer Listgarten. Autofocused oracles for model-based design. *Advances in Neural Information Processing Systems*, 33:12945–12956, 2020.

[16] Richard Fox, Ajoy Roy, Sridhar Govindarajan, Jeremy Minshull, Claes Gustafsson, Jennifer T Jones, and Robin Emig. Optimizing the search algorithm for protein engineering by directed evolution. *Protein engineering*, 16(8):589–597, 2003.

[17] Richard J Fox, S Christopher Davis, Emily C Mundorff, Lisa M Newman, Vesna Gavrilovic, Steven K Ma, Loleta M Chung, Charlene Ching, Sarena Tam, Sheela Muley, et al. Improving catalytic function by prosar-driven enzyme evolution. *Nature biotechnology*, 25(3):338–344, 2007.

[18] Chase R Freschlin, Sarah A Fahlberg, and Philip A Romero. Machine learning to navigate fitness landscapes for protein engineering. *Current Opinion in Biotechnology*, 75:102713, 2022.

[19] Christian Gießen and Carsten Witt. Optimal mutation rates for the (1+ λ) ea on onemax. In *Proceedings of the Genetic and Evolutionary Computation Conference 2016*, pages 1147–1154, 2016.

[20] Uffe Haagerup. The best constants in the khintchine inequality. *Studia Mathematica*, 70: 231–283, 1981.

[21] Nima Hamidi and Mohsen Bayati. On worst-case regret of linear thompson sampling. *arXiv preprint arXiv:2006.06790*, 2020.

[22] Yanjun Han, Zhengqing Zhou, Zhengyuan Zhou, Jose Blanchet, Peter W Glynn, and Yinyu Ye. Sequential batch learning in finite-action linear contextual bandits. *arXiv preprint arXiv:2004.06321*, 2020.

[23] Botao Hao, Tor Lattimore, and Mengdi Wang. High-dimensional sparse linear bandits. *Advances in Neural Information Processing Systems*, 33:10753–10763, 2020.

[24] Jun He and Xin Yao. A study of drift analysis for estimating computation time of evolutionary algorithms. *Natural Computing*, 3(1):21–35, 2004.

[25] Edward G Hibbert and Paul A Dalby. Directed evolution strategies for improved enzymatic performance. *Microbial Cell Factories*, 4(1):1–6, 2005.

[26] Nicholas W Hughes, Yuanhao Qu, Jiaqi Zhang, Weijing Tang, Justin Pierce, Chengkun Wang, Aditi Agrawal, Maurizio Morri, Norma Neff, Monte M Winslow, et al. Machine-learning-optimized cas12a barcoding enables the recovery of single-cell lineages and transcriptional profiles. *Molecular Cell*, 82(16):3103–3118, 2022.

[27] Jens Jägersküpper. A blend of markov-chain and drift analysis. In *International Conference on Parallel Problem Solving from Nature*, pages 41–51. Springer, 2008.

[28] Jens Jägersküpper. Combining markov-chain analysis and drift analysis. *Algorithmica*, 59(3): 409–424, 2011.

[29] Thomas Jansen and Ingo Wegener. Real royal road functions—where crossover provably is essential. *Discrete applied mathematics*, 149(1-3):111–125, 2005.

[30] Thomas Jansen, Ingo Wegener, et al. The analysis of evolutionary algorithms–a proof that crossover really can help. *Algorithmica*, 34(1):47–66, 2002.

[31] Cem Kalkanlı and Ayfer Özgür. An improved regret bound for thompson sampling in the gaussian linear bandit setting. In *2020 IEEE International Symposium on Information Theory (ISIT)*, pages 2783–2788. IEEE, 2020.

[32] Timo Kötzing, Dirk Sudholt, and Madeleine Theile. How crossover helps in pseudo-boolean optimization. In *Proceedings of the 13th annual conference on Genetic and evolutionary computation*, pages 989–996, 2011.

[33] Olga Kuchner and Frances H Arnold. Directed evolution of enzyme catalysts. *Trends in biotechnology*, 15(12):523–530, 1997.

[34] Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

[35] Per Kristian Lehre and Carsten Witt. Black-box search by unbiased variation. *Algorithmica*, 64 (4):623–642, 2012.

[36] Lihong Li, Wei Chu, John Langford, and Robert E Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pages 661–670. ACM, 2010.

[37] Dan Ofer and Michal Linial. Profet: Feature engineering captures high-level protein functions. *Bioinformatics*, 31(21):3429–3436, 2015.

[38] Dan Ofer, Nadav Brandes, and Michal Linial. The language of proteins: Nlp, machine learning & protein sequences. *Computational and Structural Biotechnology Journal*, 19:1750–1758, 2021.

[39] Pietro S Oliveto and Carsten Witt. Improved time complexity analysis of the simple genetic algorithm. *Theoretical Computer Science*, 605:21–41, 2015.

[40] Pietro S Oliveto, Dirk Sudholt, and Carsten Witt. A tight lower bound on the expected runtime of standard steady state genetic algorithms. In *Proceedings of the 2020 Genetic and Evolutionary Computation Conference*, pages 1323–1331, 2020.

[41] Michael S Packer and David R Liu. Methods for the directed evolution of proteins. *Nature Reviews Genetics*, 16(7):379–394, 2015.

[42] Philip A Romero, Andreas Krause, and Frances H Arnold. Navigating the protein fitness landscape with gaussian processes. *Proceedings of the National Academy of Sciences*, 110(3): E193–E201, 2013.

[43] Daniel Russo and Benjamin Van Roy. Learning to optimize via posterior sampling. *Mathematics of Operations Research*, 39(4):1221–1243, 2014.

[44] Ophir Shalem, Neville E Sanjana, and Feng Zhang. High-throughput functional genomics using crispr–cas9. *Nature Reviews Genetics*, 16(5):299–311, 2015.

[45] Jung-Eun Shin, Adam J Riesselman, Aaron W Kollasch, Conor McMahon, Elana Simon, Chris Sander, Aashish Manglik, Andrew C Kruse, and Debora S Marks. Protein design and variant prediction using autoregressive generative models. *Nature communications*, 12(1):1–11, 2021.

[46] Sam Sinai, Richard Wang, Alexander Whatley, Stewart Slocum, Elina Locane, and Eric D Kelsic. Adalead: A simple and robust adaptive greedy search algorithm for sequence design. *arXiv preprint arXiv:2010.02141*, 2020.

[47] George P Smith and Valery A Petrenko. Phage display. *Chemical reviews*, 97(2):391–410, 1997.

[48] Jérôme Tubiana, Simona Cocco, and Rémi Monasson. Learning protein constitutive motifs from sequence data. *Elife*, 8:e39397, 2019.

[49] Nicholas J Turner. Directed evolution drives the next generation of biocatalysts. *Nature chemical biology*, 5(8):567–573, 2009.

[50] Chenyu Wang, Joseph Kim, Le Cong, and Mengdi Wang. Neural bandits for protein sequence optimization. In *2022 56th Annual Conference on Information Sciences and Systems (CISS)*, pages 188–193. IEEE, 2022.

[51] Richard A Watson and Thomas Jansen. A building-block royal road where crossover is provably essential. In *Proceedings of the 9th annual conference on Genetic and evolutionary computation*, pages 1452–1459, 2007.

[52] Robert H Whittaker. Evolution and measurement of species diversity. *Taxon*, 21(2-3):213–251, 1972.

[53] Greg Winter, Andrew D Griffiths, Robert E Hawkins, and Hennie R Hoogenboom. Making antibodies by phage display technology. *Annual review of immunology*, 12(1):433–455, 1994.

[54] Carsten Witt. Runtime analysis of the (μ+ 1) ea on simple pseudo-boolean functions. *Evolutionary Computation*, 14(1):65–86, 2006.

[55] Carsten Witt. Tight bounds on the optimization time of a randomized search heuristic on linear functions. *Combinatorics, Probability and Computing*, 22(2):294–318, 2013.

[56] Bruce J Wittmann, Kadina E Johnston, Zachary Wu, and Frances H Arnold. Advances in machine learning for directed evolution. *Current opinion in structural biology*, 69:11–18, 2021.

[57] Kevin K Yang, Zachary Wu, and Frances H Arnold. Machine-learning-guided directed evolution for protein engineering. *Nature methods*, 16(8):687–694, 2019.

[58] Lin Yang and Mengdi Wang. Reinforcement learning in feature space: Matrix bandit, kernels, and regret bound. In *International Conference on Machine Learning*, pages 10746–10756. PMLR, 2020.

[59] Zhi-Hua Zhou, Yang Yu, and Chao Qian. *Evolutionary learning: Advances in theories and algorithms*. Springer, 2019.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [N/A]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes]
   (b) Did you include complete proofs of all theoretical results? [Yes]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [No]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [No]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [N/A]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

A Proof of Theorem 5.2

A.1 Notations

We address the following notations that frequently occur throughout the proof section. Denote by f an arbitrary linear fitness function $f(x) := \langle \theta, x \rangle$ parameterized by some $\theta \in \mathbb{R}^d$ and denote by $F^\star$ its maximum. Define $F(S) := \text{avg}_{x \in S} f(x)$, the average fitness under f of population S. While f represents arbitrary fitness function, $\{f_{\widetilde{\theta}_t}(x) := \langle \widetilde{\theta}_t, x \rangle\}_{t \in [T]}$ are the linear function parameterized by $\{\widetilde{\theta}_t\}_{t \in [T]}$ obtained by posterior sampling in each iteration of Alg.1. Corresponding to each $f_{\widetilde{\theta}_t}$, $F_t^\star := f_{\widetilde{\theta}_t}(x_t^\star)$ is its maximum value and $x_t^\star$ is its one maximum point. Denote by $F_t(S)$ the average $f_{\widetilde{\theta}_t}$ value over S. For a clear display, denote by L, an upper bound for the l_2 norm of any $x_{t,i}$ evaluated, i.e. $\|x_{t,i}\| \leq L$ and in our setting, take $L = \sqrt{d}$. Without clarification $\|\cdot\|$ denotes the l_2 norm by default and $\|\cdot\|_A$ denotes the norm normalized by matrix A.

A.2 Routine of Alg.1 and filtrations

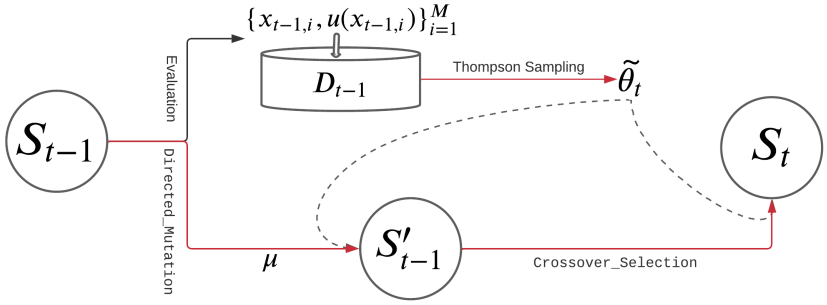

Figure A.1: **Routine of Alg.1.** Red lines represent stochastic steps. Dash lines indicate $\widetilde{\theta}_t$ is involved in those steps.

In Alg.1, there are three steps introducing stochasticity. Two of which are calling Module 2 as $S'_{t-1} = \texttt{Directed\_Mutation}(f_{\widetilde{\theta}_t}, S_{t-1}, \mu)$ and calling as $S_t = \texttt{Crossover\_Selection}(f_{\widetilde{\theta}_t}, S'_{t-1})$. Another one is Thompson sampling step s.t. $\widetilde{\theta}_t$ is sampled from the posterior of $\theta^\star$ given data D_{t-1}. Fig. A.1 illustrates how these three steps are built into the algorithm routine.

There are two other sources of stochasticity inherited from the problem setting: the prior of $\theta^\star$ (Assumption 3.2) and the noisy feedback $\{u(x_{t,i})\}_{i=1}^M$ (Assumption 3.5), which are revealed in the evaluation step. Including all stochasticity, the trajectory of Alg.1 is

$$\theta^\star, \widetilde{\theta}_1, S'_0, S_1, \{u(x_{1,i})\}_{i=1}^M, \cdots, \widetilde{\theta}_{t+1}, S'_t, S_{t+1}, \{u(x_{t+1,i})\}_{i=1}^M, \cdots, \widetilde{\theta}_T, S'_{T-1}, S_T, \{u(x_{T,i})\}_{i=1}^M. \tag{A.1}$$

At the convenience of analysis, we introduce multiple lines of the history up to time step t by carefully partitioning the trajectory (A.1), using $\sigma(\cdot)$ to represent the minimal sigma algebra expanded by $\cdot$.

Definition A.1. Define a filtration $\left\{\mathcal{H}_t^M\right\}_{t=0}^{T-1}$ with $\mathcal{H}_t^M$ be the information accumulated after t rounds of Alg.1 but before the Directed Mutation step in round $t+1$.

$$\begin{aligned}
\mathcal{H}_0^M &:= \sigma\left(\theta^\star, \widetilde{\theta}_1\right), \\
\mathcal{H}_t^M &:= \left(\mathcal{H}_{t-1}^M, \sigma\left(S'_{t-1}, S_t, \{u(x_{t,i})\}_{i=1}^M, \widetilde{\theta}_{t+1}\right)\right), \quad t \in [T-1].
\end{aligned}$$

Definition A.2. Define a filtration $\left\{\mathcal{H}_t^R\right\}_{t=0}^{T-1}$ with $\mathcal{H}_t^R$ be the information accumulated after t rounds of Alg.1 but before the Recombination and Selection step in round $t+1$.

$$\begin{aligned}
\mathcal{H}_0^R &:= \sigma\left(\theta^\star, \widetilde{\theta}_1, S'_0\right), \\
\mathcal{H}_t^R &:= \left(\mathcal{H}_{t-1}^R, \sigma\left(S_t, \{u(x_{t,i})\}_{i=1}^M, \widetilde{\theta}_{t+1}, S'_t\right)\right), \quad t \in [T-1].
\end{aligned}$$

A.3 Property of `Directed_Mutation` (Module 2)

Given a fitness function $f(x) := \langle \theta, x \rangle$, a useful observation is that for the dimension where $\theta_i \geq 0$, feature value 1 is more favorable than 0 in terms of a higher fitness. So in a population S, for each dimension i, the ratio of individuals who are with the favored feature is a key quantity, and we define it formally as follows.

Definition A.3 (Ratio of the Favored Feature)**.** Under a fitness function $f(x) := \langle \theta, x \rangle$, for a population S, define

$$p_i^\theta(S) = \begin{cases} \frac{\#\{x \in S : x_i = 0\}}{|S|} & \theta_i < 0 \\ \frac{\#\{x \in S : x_i = 1\}}{|S|} & \theta_i \geq 0, \end{cases} \qquad \forall i \in [d], \tag{A.2}$$

and we are allowed to omit the superscript θ of $p_i^\theta(S)$ when θ is clear from the context.

We show the following property of `Directed_Mutation`.

Lemma A.4. Suppose $S' = \texttt{Directed\_Mutation}(f, S, \mu)$, then the population-averaged fitness of S' will not decrease compared to that of S in expectation, that is,

$$\mathbb{E}\left[F(S')\right] \geq F(S). \tag{A.3}$$

And for $\forall \delta \in (0,1)$, if $|S| = \Omega\left(\frac{\log(\frac{d}{\delta})}{\mu^2}\right)$, then with probability $1-\delta$,

$$p_i\left(S'\right) \geq \frac{\mu}{4}, \quad \forall i \in [d]. \tag{A.4}$$

Proof. See Appendix D.1. □

A.3.1 High probability events on $\min_i p_i^{\widetilde{\theta}_{t+1}}\left(S'_t\right)$

The Directed Mutation step of Alg.1 ensures S'_t is always sufficient with the feature favored by current $f_{\widetilde{\theta}_{t+1}}$ in each dimension i throughout T rounds, i.e. $\min_i p_i^{\widetilde{\theta}_{t+1}}\left(S'_t\right)$ is lower bounded for $\forall t+1 \in [T]$, recall Definition A.3 for the definition of $p_i^\theta(S)$.

We introduce the following line of events where this sufficiency of S'_t holds and show the intersection of them happens with high probability when the population size M is sufficiently large.

Definition A.5. Define E_t^{DM} to be the event where $\min_i p_i^{\widetilde{\theta}_{s+1}}\left(S'_s\right)$ is lower bounded by $\frac{\mu}{4}$ for $\forall s \leq t$, that is

$$E_t^{\mathrm{DM}} := \{\forall s \leq t, \quad \min_i p_i^{\widetilde{\theta}_{s+1}}\left(S'_s\right) \geq \frac{\mu}{4}\}. \tag{A.5}$$

Also define E_{DM} as the intersection of $\{E_t^{\mathrm{DM}}\}_{t=0}^{T-1}$.

Lemma A.6. For $\forall \delta \in (0,1)$, if the population size is sufficiently large s.t. $M = O\left(\frac{\log(\frac{dT}{\delta})}{\mu^2}\right)$, then

$$\mathbb{P}\left(E_{\mathrm{DM}}\right) \geq 1 - \delta. \tag{A.6}$$

Since event E_{DM} is independent from the realization of $\theta^\star$, thus it still holds with high probability when conditioned on $\theta^\star$. Denote by $E_{\mathrm{DM}}^{\theta^\star}$, the event E_{DM} conditioned on $\theta^\star$, then

$$\mathbb{P}\left(E_{\mathrm{DM}}^{\theta^\star}\right) \geq 1 - \delta.$$

Proof. See Appendix D.2. □

A.4 Linear convergence of `Crossover_Selection` (Module 1)

Continuing from Theorem 5.1, when S is sufficient with the feature favored by f in every dimension, i.e. $\min_i p_i(S)$ is lower bounded, then in expectation, $F(S')$ converges linearly to $F^\star$ with a nontrivial convergence rate.

Lemma A.7 (Linear Convergence)**.** Suppose $S' = \texttt{Crossover\_Selection}(f, S)$, then

$$\mathbb{E}\left[F(S')\right] \geq F(S) + \eta\left(F^\star - F(S)\right), \tag{A.7}$$

with factor $\eta = \frac{\min_i p_i(S)}{\sqrt{2d}}$.

Proof. See Appendix D.3. □

A.5 Thompson Sampling

According to Assumption 3.5, in the dataset $D_t = \left\{\{x_{s,i}, u(x_{s,i})\}_{i=1}^M, s = [t]\right\}$

$$u(x_{s,i}) = f_{\theta^\star}(x_{s,i}) + \xi_{s,i}, \tag{A.8}$$

with $\xi_{s,i}$ i.i.d. sampled from $\mathcal{N}(0, \sigma^2)$ and independent from all other stochasticity.

Therefore, by Bayes' Rule, the posterior of $\theta^\star$ give D_{t-1} is also Gaussian distributed, for $\forall t \in [T]$:

$$\widetilde{\theta}_t \sim \mathcal{N}(\widehat{\theta}_t, V_t^{-1}), \tag{A.9}$$

where

$$V_t = \frac{1}{\sigma^2}\Phi_{t-1}^\top \Phi_{t-1} + \lambda I, \tag{A.10}$$

$$\widehat{\theta}_t = \frac{1}{\sigma^2} V_t^{-1} \Phi_{t-1}^\top U_{t-1}, \tag{A.11}$$

recall from Alg. 1 for the updating rules of Φ_t and U_t.

Given the posterior distribution (A.9), we are able to show $\widetilde{\theta}_t$ concentrates to $\theta^\star$ in term of the normalized distance between them.

A.5.1 High probability events on $\left\|\widetilde{\theta}_t - \theta^\star\right\|_{V_t}$

We introduce two useful lines of high probability events similar to those in Abeille and Lazaric [2], except here these events are defined conditioned on any realization of $\theta^\star$ sampled from its prior. We rephrased the definition to match our notations.

Definition A.8. Conditioned on $\theta^\star$, for any given probability tolerance $\delta \in (0, 1)$, each time step $t \in [T]$ and a line of ($\theta^\star$ dependent) radiuses $\{\beta_t^{\theta^\star}(\delta)\}_{t=1}^T$, we define $\widehat{E}_t^{\theta^\star}$ as the event where $\widehat{\theta}_s$ concentrates around $\theta^\star$ for all previous steps $s \leq t$, i.e.

$$\widehat{E}_t^{\theta^\star}(\delta) = \left\{\forall s \leq t, \left\|\widehat{\theta}_s - \theta^\star\right\|_{V_s} \leq \beta_s^{\theta^\star}(\delta) \middle| \theta^\star\right\}. \tag{A.12}$$

with a line of ($\theta^\star$ independent) radiuses $\{\alpha_t(\delta)\}_{t=1}^T$, we also define $\widetilde{E}_t^{\theta^\star}$ as the event where the sampled parameter $\widetilde{\theta}_s$ concentrates around $\widehat{\theta}_s$ for all steps $s \leq t$, i.e.

$$\widetilde{E}_t^{\theta^\star}(\delta) = \left\{\forall s \leq t, \left\|\widetilde{\theta}_s - \widehat{\theta}_s\right\|_{V_s} \leq \alpha_s(\delta) \middle| \theta^\star\right\}. \tag{A.13}$$

Then under the same δ and $\theta^\star$, which are omitted here, we have $\widehat{E} := \widehat{E}_T \subset \cdots \subset \widehat{E}_1$, $\widetilde{E} := \widetilde{E}_T \subset \cdots \subset \widetilde{E}_1$ and define $E^{\theta^\star}(\delta) := \widehat{E}^{\theta^\star}(\delta) \cap \widetilde{E}^{\theta^\star}(\delta)$.

With appropriate choices of $\{\beta_t\}$ and $\{\alpha_t\}$, event $E^{\theta^\star}(\delta)$ defined above happens with high probability as stated in the following lemma.

Lemma A.9. Under Assumption 3.2 and 3.5, conditioned on any realization of $\theta^\star$ drawn from its prior, for $\forall \delta \in (0, 1)$ and any series of feature vectors $\left(\{x_{1,i}\}_{i=1}^M, \cdots, \{x_{T,i}\}_{i=1}^M\right)$ where each $\|x_{t,i}\| \leq L$, $\mathbb{P}\left(E^{\theta^\star}\left(\frac{\delta}{2}\right)\right) \geq 1 - \delta$ with $\beta_t^{\theta^\star}(\delta)$ and $\alpha_t(\delta)$ specified as

$$\beta_t^{\theta^\star}(\delta) = \sqrt{2\log\left(\frac{1}{\delta}\right) + d\log\left(\frac{\sigma^2\lambda d + tML^2}{\sigma^2\lambda d}\right)} + \sqrt{\lambda}\|\theta^\star\|, \quad \forall t \in [T]. \tag{A.14}$$

$$\alpha_t(\delta) = 2\sqrt{d\log\left(\frac{T}{\delta}\right)} + \sqrt{d}, \quad \forall t \in [T]. \tag{A.15}$$

Proof. See Appendix D.4. □

A.6 Prediction error under batch update

Before regret decomposition, one more preparation to have is a modified concentration on the accumulated prediction error of $\widetilde{\theta}_t$ catering for the batch-data update routine in Alg.1. In the following lemma, we summarize a more general version of this concentration result.

Lemma A.10. Suppose at any timestep a_t lies in a confidence ellipsoid around b_t in the sense that

$$\|a_t - b_t\|_{V_t} \leq \eta_t(\delta), \quad \forall t \in [T],$$

and $\|x_{t,i}\| \leq L, \forall t \in [T], i \in [M]$, then it holds that,

$$\begin{aligned}
&\sum_{t=1}^{T}\sum_{i=1}^{M} |\langle a_t - b_t, x_{t,i}\rangle| \\
\leq &\eta_T(\delta)\sqrt{\frac{2L^2+2\lambda}{\lambda}} \cdot \sqrt{dMT\log\left(\frac{\sigma^2 d\lambda + MTL^2}{\sigma^2 d\lambda}\right)} + \eta_T(\delta)\frac{2L}{\sqrt{\lambda}} \cdot dM\log\left(\frac{\sigma^2 d\lambda + MTL^2}{\sigma^2 d\lambda}\right).
\end{aligned} \tag{A.16}$$

And let us give an alias $\mathrm{RGT}\left(\eta_T(\delta)\right)$ to the RHS of (A.16).

Proof. See Appendix D.6. □

A.7 Regret decomposition

Recall the notation that $x^\star$ is a maximum point of $f_{\theta^\star}$.

By the scheme of posterior sampling, $f_{\theta^\star}(x^\star)$ and $f_{\widetilde{\theta}_t}(x_t^\star)$ are identically distributed conditioned on D_{t-1}, which leads to

$$\mathbb{E}\left[f_{\theta^\star}(x^\star) - f_{\widetilde{\theta}_t}(x_t^\star)\middle| D_{t-1}\right] = 0. \tag{A.17}$$

With expectation taken over all stochasticity, the per-round Bayesian regret is

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^{M}\left(f_{\theta^\star}(x^\star) - f_{\theta^\star}(x_{t,i})\right)\right] &= \sum_{i=1}^{M}\mathbb{E}\left[f_{\theta^\star}(x^\star) - f_{\widetilde{\theta}_t}(x_t^\star)\right] + \sum_{i=1}^{M}\mathbb{E}\left[f_{\widetilde{\theta}_t}(x_t^\star) - f_{\theta^\star}(x_{t,i})\right] \\
&= \sum_{i=1}^{M}\mathbb{E}\left[\mathbb{E}\left[f_{\theta^\star}(x^\star) - f_{\widetilde{\theta}_t}(x_t^\star)\middle| D_{t-1}\right]\right] + \sum_{i=1}^{M}\mathbb{E}\left[f_{\widetilde{\theta}_t}(x_t^\star) - f_{\theta^\star}(x_{t,i})\right] \\
&\overset{(A.17)}{=} \sum_{i=1}^{M}\mathbb{E}\left[f_{\widetilde{\theta}_t}(x_t^\star) - f_{\theta^\star}(x_{t,i})\right] \\
&= \sum_{i=1}^{M}\mathbb{E}\left[f_{\widetilde{\theta}_t}(x_t^\star) - f_{\widetilde{\theta}_t}(x_{t,i}) + f_{\widetilde{\theta}_t}(x_{t,i}) - f_{\theta^\star}(x_{t,i})\right] \\
&= \mathbb{E}\left[M\left(F_t^\star - F_t(S_t)\right)\right] + \mathbb{E}\left[\sum_{i=1}^{M}\langle\widetilde{\theta}_t - \theta^\star, x_{t,i}\rangle\right].
\end{aligned}$$

Then the total Bayesian regret over T rounds sums up to be

$$\begin{aligned}
\mathrm{BayesRGT}(T,M) &= \mathbb{E}\left[M\sum_{t=1}^{T}F_t^\star - F_t(S_t)\right] + \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{M}\langle\widetilde{\theta}_t - \theta^\star, x_{t,i}\rangle\right] \\
&= \mathbb{E}_{\theta^\star\sim\pi}\left[\mathbb{E}_{\theta^\star}\left[M\sum_{t=1}^{T}F_t^\star - F_t(S_t)\right]\right] + \mathbb{E}_{\theta^\star\sim\pi}\left[\mathbb{E}_{\theta^\star}\left[\sum_{t=1}^{T}\sum_{i=1}^{M}\langle\widetilde{\theta}_t - \theta^\star, x_{t,i}\rangle\right]\right],
\end{aligned} \tag{A.18}$$

where $\mathbb{E}_{\theta^\star}[\cdot]$ denotes the conditional expectation on a given $\theta^\star$: $\mathbb{E}[\cdot|\,\theta^\star]$.

Note that under any realization of $\theta^\star$, the regret of each individual at any time step should be no more than the range of $f_{\theta^\star}$ on domain $\mathcal{X}$. For any $\theta \in R^d$ parameterizing the fitness f_θ as $f_\theta(x) = \langle \theta, x \rangle$, denote by B_f^θ an upper bound for the range of f_θ, i.e.

$$B_f^\theta := 2L\|\theta\| \geq \max_x f_\theta(x) - \min_x f_\theta(x). \tag{A.19}$$

For the regret of each individual in each step, it holds that

$$f_{\theta^\star}(x^\star) - f_{\theta^\star}(x_{t,i}) \leq 2\|\theta^\star\|L = B_f^{\theta^\star}. \tag{A.20}$$

Therefore, when bounding the total regret decomposed as (A.18), it is reasonable to truncate terms in the RHS of (A.18) with $B_f^{\theta^\star}$ to derive a tighter bound.

$$\begin{aligned}
\mathrm{BayesRGT}(T, M) \leq & \mathbb{E}_{\theta^\star \sim \pi}\left[M \mathbb{E}_{\theta^\star}\left[\sum_{t=1}^{T} \min\left\{ F_t^\star - F_t(S_t), B_f^{\theta^\star} \right\} \right] \right] && \text{(A.21)} \\
& + \mathbb{E}_{\theta^\star \sim \pi}\left[\mathbb{E}_{\theta^\star}\left[\sum_{t=1}^{T} \sum_{i=1}^{M} \min\left\{ \left| \langle \widetilde{\theta}_t - \theta^\star, x_{t,i} \rangle \right|, B_f^{\theta^\star} \right\} \right] \right] && \text{(A.22)}
\end{aligned}$$

A.8 Bounding the first half (A.21)

A.8.1 After calling $S'_{t-1} = \texttt{Directed\_Mutation}(f_{\widetilde{\theta}_t}, S_{t-1}, \mu)$

As shown in Lemma A.4, the population average of S'_t under $f_{\widetilde{\theta}_{t+1}}$ in not decreasing from that of S_t, that is

$$\mathbb{E}\left[F_{t+1}(S'_t) \middle| \mathcal{H}_t^M \right] \geq F_{t+1}(S_t). \tag{A.23}$$

The other property of `Directed_Mutation` is to ensure that w.h.p. $\min_i p_i^{\widetilde{\theta}_{t+1}}(S'_t)$ is lower bounded for $\forall t+1 \in [T]$, which is stated in the definition of event $E_{\mathrm{DM}}^{\theta^\star}$ (Definition A.5). So from here on, given any realization of $\theta^\star$, our further analysis is conditioned on $E^{\theta^\star} := E^{\theta^\star}\left(\frac{\delta}{2}\right) \cap E_{\mathrm{DM}}^{\theta^\star}$.

Corollary A.11. Given any realization of $\theta^\star$, if $M = \Omega\left(\frac{\log(\frac{dT}{\delta})}{\mu_M^2}\right)$, then $\mathbb{P}\left(E^{\theta^\star}\right) \geq 1 - 2\delta$ for $\forall \delta \in (0,1)$. Conditioned on $E^{\theta^\star} := E^{\theta^\star}\left(\frac{\delta}{2}\right) \cap E_{\mathrm{DM}}^{\theta^\star}$, it is guaranteed that

$$\min_i p_i^{\widetilde{\theta}_{t+1}}(S'_t) \geq C_{S'} := \frac{\mu}{4}, \quad \forall t+1 \in [T], \tag{A.24}$$

$$\|\widetilde{\theta}_t - \theta^\star\|_{V_t} \leq \beta_t^{\theta^\star}\left(\frac{\delta}{2}\right) + \alpha_t\left(\frac{\delta}{2}\right), \quad \forall t \in [T]. \tag{A.25}$$

where recall the definition of $\beta_t^{\theta^\star}$ and α_t from (A.14) and (A.15).

Proof. The proof is directly derived by combining Lemma A.6 and Lemma A.9. □

A.8.2 After calling $S_t = \texttt{Crossover\_Selection}(f_{\widetilde{\theta}_t}, S'_{t-1})$

Conditioned on $\theta^\star$, we are about to decompose $\sum_{t=1}^{T}\left(F_t^\star - F_t(S_t)\right)$ by leveraging the linear convergence property shown in Lemma A.7. Conditionally on $E^{\theta^\star}$, applying Lemma A.7 to each call of $\texttt{Crossover\_Selection}(f_{\widetilde{\theta}_t}, S'_{t-1})$ guarantees for $\forall t+1 \in [T]$,

$$\begin{aligned}
\mathbb{E}_{E^{\theta^\star}}\left[F_{t+1}(S_{t+1}) \middle| \mathcal{H}_t^R \right] &\geq \mathbb{E}_{E^{\theta^\star}}\left[F_{t+1}(S'_t) + \frac{\min_i p_i^{\widetilde{\theta}_{t+1}}(S'_t)}{\sqrt{2d}}\left(F_{t+1}^\star - F_{t+1}(S'_t)\right) \middle| \mathcal{H}_t^R \right] \\
&\overset{(A.24)}{\geq} F_{t+1}(S'_t) + \frac{C_{S'}}{\sqrt{2d}}\left(F_{t+1}^\star - F_{t+1}(S'_t)\right),
\end{aligned} \tag{A.26}$$

where $\mathbb{E}_{E^{\theta^\star}}[\cdot]$ is the conditional expectation on event $E^{\theta^\star}$.

Recall (A.23) that

$$\mathbb{E}\left[F_{t+1}(S'_t)|\,\mathcal{H}_t^M\right] \geq F_{t+1}(S_t). \tag{A.23 revisited}$$

Conditioned on $E^{\theta^\star}$, it still holds that

$$\mathbb{E}_{E^{\theta^\star}}\left[F_{t+1}(S'_t)|\,\mathcal{H}_t^M\right] \geq F_{t+1}(S_t), \tag{A.27}$$

since in $E^{\theta^\star}$, $E^{\theta^\star}\left(\frac{\delta}{2}\right)$ holds independent from the Directed Mutation step $S'_t = \mathrm{DM}(f_{\widetilde{\theta}_{t+1}}, S_t, \mu_M)$, and conditioned on $E^{\theta^\star}_{\mathrm{DM}}$, $F_{t+1}(S'_t)|\,\mathcal{H}_t^M$ tends to be greater then it was unconditionally.

Along with $\mathcal{H}_t^M \subset \mathcal{H}_t^R$, we have

$$\mathbb{E}_{E^{\theta^\star}}\left[F_{t+1}(S_{t+1})|\,\mathcal{H}_t^M\right] \geq \mathbb{E}_{E^{\theta^\star}}\left[\mathbb{E}_{E^{\theta^\star}}\left[F_{t+1}(S_{t+1})|\,\mathcal{H}_t^R\right]\middle|\,\mathcal{H}_t^M\right] \tag{A.28}$$

$$\overset{(A.26)}{\geq} \mathbb{E}_{E^{\theta^\star}}\left[F_{t+1}(S'_t) + \frac{C_{S'}}{\sqrt{2d}}\left(F^\star_{t+1} - F_{t+1}(S'_t)\right)\middle|\,\mathcal{H}_t^M\right] \tag{A.29}$$

$$\overset{(A.27)}{\geq} F_{t+1}(S_t) + \frac{C_{S'}}{\sqrt{2d}}\left(F^\star_{t+1} - F_{t+1}(S_t)\right). \tag{A.30}$$

By introducing the convergence rate $\gamma := 1 - \frac{C_{S'}}{\sqrt{2d}}$ s.t. $\frac{1}{1-\gamma} = O\left(\frac{\sqrt{d}}{\mu}\right)$ and an residual term

$$e_{t+1} := \mathbb{E}_{E^{\theta^\star}}\left[F_{t+1}(S_{t+1})|\,\mathcal{H}_t^M\right] - F_{t+1}(S_{t+1}),$$

we have

$$F^\star_{t+1} - F_{t+1}(S_{t+1}) \leq \gamma(F^\star_{t+1} - F_{t+1}(S_t)) + e_{t+1}, \tag{A.31}$$

where $\{e_t\}_{t=1}^T$ is a martingale difference with

$$\mathbb{E}_{E^{\theta^\star}}[e_{t+1} \mid \mathcal{H}_t^M] = 0. \tag{A.32}$$

Thus,

$$\begin{aligned} F^\star_{t+1} - F_{t+1}(S_{t+1}) &\leq \gamma(F^\star_{t+1} - F_{t+1}(S_t)) + e_{t+1} \\ &= \gamma\left[F^\star_t - F_t(S_t) + F^\star_{t+1} - F^\star_t + F_t(S_t) - F_{t+1}(S_t)\right] + e_{t+1} \\ &= \gamma\left[F^\star_t - F_t(S_t)\right] + \gamma\left[F^\star_{t+1} - F^\star_t + F_t(S_t) - F_{t+1}(S_t)\right] + e_{t+1}. \end{aligned}$$

Therefore we have the recursion that

$$F^\star_t - F_t(S_t) \leq \gamma^t\left(F^\star_1 - F_1(S_0)\right) + \sum_{k=1}^{t-1}\gamma^{t-k}\left(F^\star_{k+1} - F^\star_k\right) + \sum_{k=1}^{t-1}\gamma^{t-k}\left(F_k(S_k) - F_{k+1}(S_k)\right) + \sum_{k=1}^{t}\gamma^{t-k}e_k, \tag{A.33}$$

summing up which from $t = 1$ to T gives

$$\sum_{t=1}^T F^\star_t - F_t(S_t) \leq \sum_{t=1}^T\sum_{k=1}^t \gamma^{t-k}e_k \tag{A.34}$$

$$+ \sum_{t=1}^T \gamma^t\left(F^\star_1 - F_1(S_0)\right) \tag{A.35}$$

$$+ \sum_{t=1}^T\sum_{k=1}^{t-1}\gamma^{t-k}\left(F^\star_{k+1} - F^\star_k\right) \tag{A.36}$$

$$+ \sum_{t=1}^T\sum_{k=1}^{t-1}\gamma^{t-k}\left(F_k(S_k) - F_{k+1}(S_k)\right). \tag{A.37}$$

As it appears in (A.21), what matters in bounding regret is the expected truncated value of $\sum_{t=1}^T F^\star_t - F_t(S_t)$, which is

$$\mathbb{E}_{\theta^\star\sim\pi}\left[M\mathbb{E}_{\theta^\star}\left[\sum_{t=1}^T \min\left\{F^\star_t - F_t(S_t), B_f^{\theta^\star}\right\}\right]\right], \tag{A.21 revisited}$$

and the decomposition of $\sum_{t=1}^{T} F_t^\star - F_t(S_t)$ into four terms as above holds conditionally on $E^{\theta^\star}$. So from here on, we progress with first upper bounding $\mathbb{E}_{\theta^\star}\left[\sum_{t=1}^{T}\min\left\{F_t^\star - F_t(S_t), B_f^{\theta^\star}\right\}\right]$ by

$$\mathbb{E}_{\theta^\star}\left[\sum_{t=1}^{T}\min\left\{F_t^\star - F_t(S_t), B_f^{\theta^\star}\right\}\right] \leq 2\delta T B_f^{\theta^\star} + (1-2\delta)\mathbb{E}_{E^{\theta^\star}}\left[\sum_{t=1}^{T} F_t^\star - F_t(S_t)\right], \quad \text{(A.38)}$$

and then upper bounding $\mathbb{E}_{E^{\theta^\star}}\left[\sum_{t=1}^{T} F_t^\star - F_t(S_t)\right]$ with

$$\begin{aligned}
\mathbb{E}_{E^{\theta^\star}}\left[\sum_{t=1}^{T} F_t^\star - F_t(S_t)\right] \leq & \mathbb{E}_{E^{\theta^\star}}\left[\sum_{t=1}^{T}\sum_{k=1}^{t}\gamma^{t-k}e_k\right] && \text{(A.39)}\\
& + \mathbb{E}_{E^{\theta^\star}}\left[\sum_{t=1}^{T}\gamma^t\left(F_1^\star - F_1(S_0)\right)\right] && \text{(A.40)}\\
& + \mathbb{E}_{E^{\theta^\star}}\left[\sum_{t=1}^{T}\sum_{k=1}^{t-1}\gamma^{t-k}\left(F_{k+1}^\star - F_k^\star\right)\right] && \text{(A.41)}\\
& + \mathbb{E}_{E^{\theta^\star}}\left[\sum_{t=1}^{T}\sum_{k=1}^{t-1}\gamma^{t-k}\left(F_k(S_k) - F_{k+1}(S_k)\right)\right]. && \text{(A.42)}
\end{aligned}$$

A.8.3 Term (A.39)

$\{e_t\}_{t=1}^{T}$ is claimed to be a martingale difference when first being introduced, that is, recall (A.32) that

$$\mathbb{E}_{E^{\theta^\star}}[e_{t+1} \mid \mathcal{H}_t^M] = 0, \quad \forall t+1 \in [T]. \qquad \text{(A.32 revisited)}$$

Thus by the property of martingale difference,

$$\mathbb{E}_{E^{\theta^\star}}\left[e_{t+1}\right] = \mathbb{E}_{E^{\theta^\star}}\left[\mathbb{E}_{E^{\theta^\star}}\left[e_{t+1} \mid \mathcal{H}_t^M\right]\right] = 0, \quad \forall t+1 \in [T].$$

Then by the linearity of $\mathbb{E}_{E^{\theta^\star}}[\cdot]$:

$$\begin{aligned}
\mathbb{E}_{E^{\theta^\star}}\left[\sum_{t=1}^{T}\sum_{k=1}^{t}\gamma^{t-k}e_k\right] &= \sum_{t=1}^{T}\sum_{k=1}^{t}\mathbb{E}_{E^{\theta^\star}}\left[\gamma^{t-k}e_k\right]\\
&= \sum_{t=1}^{T}\sum_{k=1}^{t}\gamma^{t-k}\mathbb{E}_{E^{\theta^\star}}\left[e_k\right]\\
&= 0. \qquad \text{(A.43)}
\end{aligned}$$

A.8.4 Term (A.40)

Before looking into the term (A.40), we first introduce the following lemma upper bounding the expectation of $\widetilde{\theta}_t$'s l_2 norm conditioned on event $E^{\theta^\star}$.

Lemma A.12. For $\forall t \in [T]$, $\mathbb{E}_{E^{\theta^\star}}\left[\|\widetilde{\theta}_t\|\right]$ has the following upper bound.

$$\mathbb{E}_{E^{\theta^\star}}\left[\|\widetilde{\theta}_t\|\right] \leq 2\|\theta^\star\| + 2\sqrt{\frac{d}{\lambda}}. \qquad \text{(A.44)}$$

Proof. See Appendix D.8. □

What is to take expectation in (A.40) is of constant order because

$$\sum_{t=1}^{T}\gamma^{t-1}\left(F_1^\star - F_1(S_0)\right) \leq \frac{1}{1-\gamma}\left|F_1^\star - F_1(S_0)\right| \leq \frac{1}{1-\gamma}B_f^{\widetilde{\theta}_1}, \qquad \text{(A.45)}$$

where $B_f^{\widetilde{\theta}_1} = 2L\|\widetilde{\theta}_1\|$.

Then by taking expectation over both sides of (A.45), we have

$$\mathbb{E}_{E^{\theta^\star}}\left[\sum_{t=1}^{T}\gamma^t\left(F_1^\star - F_1(S_0)\right)\right] \leq \frac{2L}{1-\gamma}\mathbb{E}_{E^{\theta^\star}}\left[\|\widetilde{\theta}_1\|\right] \leq \frac{2}{1-\gamma}\left(B_f^{\theta^\star} + 2\sqrt{\frac{d}{\lambda}}L\right). \tag{A.46}$$

A.8.5 Term (A.41)

Rearrange terms to sum up in (A.41) as

$$\begin{aligned}
\sum_{t=1}^{T}\sum_{k=1}^{t-1}\gamma^{t-k}\left(F_{k+1}^\star - F_k^\star\right) &= \sum_{k=1}^{T-1}\left(F_{k+1}^\star - F_k^\star\right)\sum_{t=k+1}^{T}\gamma^{t-k} \\
&= \sum_{k=1}^{T-1}\frac{\gamma-\gamma^{T-k+1}}{1-\gamma}\left(F_{k+1}^\star - F_k^\star\right) \\
&= \frac{\gamma}{1-\gamma}(F_T^\star - F_1^\star) - \sum_{k=1}^{T-1}\frac{\gamma^{T-k+1}}{1-\gamma}\left(F_{k+1}^\star - F_k^\star\right) \\
&= \frac{\gamma}{1-\gamma}(F_T^\star - F_1^\star) - \frac{\gamma^2}{1-\gamma}F_T^\star + \sum_{k=2}^{T-1}\gamma^{T-k+1}F_k^\star + \frac{\gamma^T}{1-\gamma}F_1^\star \\
&= \sum_{k=2}^{T}\gamma^{T-k+1}F_k^\star - \gamma^{T-1}F_1^\star.
\end{aligned}$$

Thus by taking expectation conditioned on $E^{\theta^\star}$ over the absolute value of RHS, we have

$$\begin{aligned}
\mathbb{E}_{E^{\theta^\star}}\left[\sum_{t=1}^{T}\sum_{k=1}^{t-1}\gamma^{t-k}\left(F_{k+1}^\star - F_k^\star\right)\right] &\leq \mathbb{E}_{E^{\theta^\star}}\left[\sum_{k=2}^{T}\gamma^{T-k+1}\left|F_k^\star\right|\right] + \mathbb{E}_{E^{\theta^\star}}\left[\gamma^{T-1}\left|F_1^\star\right|\right] \\
&\leq L\sum_{k=2}^{T}\gamma^{T-k+1}\cdot\mathbb{E}_{E^{\theta^\star}}\left[\|\widetilde{\theta}_k\|\right] + L\gamma^{T-1}\cdot\mathbb{E}_{E^{\theta^\star}}\left[\|\widetilde{\theta}_1\|\right] \\
&\leq \sum_{k=0}^{T-1}\gamma^k\left(2\|\theta^\star\|L + 2\sqrt{\frac{d}{\lambda}}L\right) \\
&\leq \frac{1}{1-\gamma}\left(B_f^{\theta^\star} + 2\sqrt{\frac{d}{\lambda}}L\right).
\end{aligned} \tag{A.47}$$

A.8.6 Term (A.42)

We start off by rearranging terms in the summation: $\sum_{t=1}^{T}\sum_{k=1}^{t-1}\gamma^{t-k}\left(F_k(S_k) - F_{k+1}(S_k)\right)$.

$$\begin{aligned}
\sum_{t=1}^{T}\sum_{k=1}^{t-1}\gamma^{t-k}\left(F_k(S_k) - F_{k+1}(S_k)\right) &= \sum_{k=1}^{T-1}\left(F_k(S_k) - F_{k+1}(S_k)\right)\sum_{t=k+1}^{T}\gamma^{t-k} \\
&= \sum_{k=1}^{T-1}\frac{\gamma-\gamma^{T-k+1}}{1-\gamma}\left(F_k(S_k) - F_{k+1}(S_k)\right) \\
&= \sum_{k=1}^{T-1}\frac{\gamma-\gamma^{T-k+1}}{1-\gamma}\langle\widetilde{\theta}_k - \widetilde{\theta}_{k+1}, \frac{1}{M}\sum_{x\in S_k}x\rangle \\
&\leq \frac{1}{1-\gamma}\frac{1}{M}\sum_{t=1}^{T-1}\sum_{i=1}^{M}\left|\langle\widetilde{\theta}_t - \widetilde{\theta}_{t+1}, x_{t,i}\rangle\right|.
\end{aligned} \tag{A.48}$$

In the following Corollary A.13, we bound the RHS above by constructing a high probability confidence ellipsoid for $\widetilde{\theta}_{t+1} - \widetilde{\theta}_t$ and then completing with a call of Lemma A.10.

Corollary A.13. For any realization of $\theta^\star$, conditioned on event $E^{\theta^\star}$, it holds that

$$\sum_{t=1}^{T}\sum_{k=1}^{t-1}\gamma^{t-k}\left(F_k(S_k) - F_{k+1}(S_k)\right) \leq \frac{1}{1-\gamma}\frac{1}{M}\,\mathrm{RGT}\left(2\beta_T^{\theta^\star}\left(\frac{\delta}{2}\right) + 2\alpha_T\left(\frac{\delta}{2}\right)\right). \tag{A.49}$$

Proof. See Appendix D.9. □

Therefore, after taking expectation conditioned on $E^{\theta^\star}$, we still have

$$\mathbb{E}_{E^{\theta^\star}}\left[\sum_{t=1}^{T}\sum_{k=1}^{t-1}\gamma^{t-k}\left(F_k(S_k) - F_{k+1}(S_k)\right)\right] \leq \frac{1}{M}\cdot\frac{1}{1-\gamma}\,\mathrm{RGT}\left(2\beta_T^{\theta^\star}\left(\frac{\delta}{2}\right) + 2\alpha_T\left(\frac{\delta}{2}\right)\right). \tag{A.50}$$

A.8.7 Pulling 4 terms into the final bound of the first half (A.21)

Going back to the contribution coming from the first half of the regret decomposition (A.18), plugging (A.43), (A.46), (A.47) and (A.50) into (A.38), it holds that, for $\forall\delta \in (0,1)$

$$\begin{aligned}
M\mathbb{E}_{\theta^\star}\left[\sum_{t=1}^{T}\min\left\{F_t^\star - F_t(S_t), B_f^{\theta^\star}\right\}\right] \leq& 2\delta MT\cdot B_f^{\theta^\star} + (1-2\delta)M\cdot\mathbb{E}_{E^{\theta^\star}}\left[\sum_{t=1}^{T}F_t^\star - F_t(S_t)\right] \\
\leq& 2\delta MT\cdot B_f^{\theta^\star} + (1-2\delta)\cdot\frac{3M}{1-\gamma}\left(B_f^{\theta^\star} + 2\sqrt{\frac{d}{\lambda}}L\right). \\
&+ (1-2\delta)\cdot\frac{1}{1-\gamma}\,\mathrm{RGT}\left(2\beta_T^{\theta^\star}\left(\frac{\delta}{2}\right) + 2\alpha_T\left(\frac{\delta}{2}\right)\right).
\end{aligned} \tag{A.51}$$

Averaging (A.51) over the prior of $\theta^\star$, we have

$$\begin{aligned}
M\mathbb{E}\left[\sum_{t=1}^{T}\min\left\{F_t^\star - F_t(S_t), B_f^{\theta^\star}\right\}\right] =& \mathbb{E}_{\theta^\star\sim\pi}\left[M\mathbb{E}_{\theta^\star}\left[\sum_{t=1}^{T}\min\left\{F_t^\star - F_t(S_t), B_f^{\theta^\star}\right\}\right]\right] \\
\leq& 2\delta\cdot\mathbb{E}_{\theta^\star\sim\pi}\left[B_f^{\theta^\star}\right]\cdot MT \\
&+ \frac{3(1-2\delta)}{1-\gamma}\cdot\left(\mathbb{E}_{\theta^\star\sim\pi}\left[B_f^{\theta^\star}\right] + 2\sqrt{\frac{d}{\lambda}}L\right)\cdot M \\
&+ \frac{1-2\delta}{1-\gamma}\cdot\mathbb{E}_{\theta^\star\sim\pi}\left[\mathrm{RGT}\left(2\beta_T^{\theta^\star}\left(\frac{\delta}{2}\right) + 2\alpha_T\left(\frac{\delta}{2}\right)\right)\right].
\end{aligned} \tag{A.52}$$

A.9 The Second Half of Regret Bound as in (A.22)

Conditioned on $E^{\theta^\star}\left(\frac{\delta}{2}\right)$, which holds with probability $1-\delta$, $\widetilde{\theta}_t$ lies in a confidence ellipsoid around $\theta^\star$ at all times,

$$\|\widetilde{\theta}_t - \theta^\star\|_{V_t} \leq \eta_t(\delta), \quad \forall t \in [T].$$

We wrap up an upper bound for $\sum_{t=1}^{T}\sum_{i=1}^{M}\left|\langle\widetilde{\theta}_t - \theta^\star, x_{t,i}\rangle\right|$ derived by calling Lemma A.10 into the corollary as follows.

Corollary A.14. Conditioned on $E^{\theta^\star}\left(\frac{\delta}{2}\right)$, the part of total regret contributed by the prediction error of TS sampled $\widetilde{\theta}_t$ is upper bounded by

$$\sum_{t=1}^{T}\sum_{i=1}^{M}\left|\langle\widetilde{\theta}_t - \theta^\star, x_{t,i}\rangle\right| \leq \mathrm{RGT}\left(\beta_T^{\theta^\star}\left(\frac{\delta}{2}\right) + \alpha_T\left(\frac{\delta}{2}\right)\right). \tag{A.53}$$

Proof. Lemma A.10 directly applies by customizing the parameter $\eta_t(\delta)$ to be $\beta_T^{\theta^\star}\left(\frac{\delta}{2}\right)+\alpha_T\left(\frac{\delta}{2}\right)$. □

With Corollary A.14 ready, we take expectation first conditioned on $\theta^\star$ and then over the prior of $\theta^\star$, which finally gives an upper bound of (A.22) as

$$\begin{aligned}
&\mathbb{E}_{\theta^\star\sim\pi}\left[\mathbb{E}_{\theta^\star}\left[\sum_{t=1}^{T}\sum_{i=1}^{M}\min\left\{\left|\langle\widetilde{\theta}_t-\theta^\star,x_{t,i}\rangle\right|,B_f^{\theta^\star}\right\}\right]\right]\\
\leq&\mathbb{E}_{\theta^\star\sim\pi}\left[\delta MTB_f^{\theta^\star}+(1-\delta)\,\mathrm{RGT}\left(\beta_T^{\theta^\star}\left(\frac{\delta}{2}\right)+\alpha_T\left(\frac{\delta}{2}\right)\right)\right]\\
=&\delta\cdot\mathbb{E}_{\theta^\star\sim\pi}\left[B_f^{\theta^\star}\right]\cdot MT+(1-\delta)\cdot\mathbb{E}_{\theta^\star\sim\pi}\left[\mathrm{RGT}\left(\beta_T^{\theta^\star}\left(\frac{\delta}{2}\right)+\alpha_T\left(\frac{\delta}{2}\right)\right)\right].
\end{aligned}\tag{A.54}$$

A.10 Final Bound: Combining The Two Halves (A.21) and (A.22)

Pulling two parts (A.52) and (A.54) into the regret decomposition (A.18), for $\forall\delta\in(0,1)$, with γ s.t. $\frac{1}{1-\gamma}=O\left(\frac{\sqrt{d}}{\mu}\right)$, the Bayesian regret of Alg.1 is bounded by

$$\begin{aligned}
\mathrm{BayesRGT}(T,M)\leq&3\delta\cdot\mathbb{E}_{\theta^\star\sim\pi}\left[B_f^{\theta^\star}\right]\cdot MT\\
&+\frac{3(1-2\delta)}{1-\gamma}\cdot\left(\mathbb{E}_{\theta^\star\sim\pi}\left[B_f^{\theta^\star}\right]+2\sqrt{\frac{d}{\lambda}}L\right)\cdot M\\
&+(1-\delta)\cdot\mathbb{E}_{\theta^\star\sim\pi}\left[\mathrm{RGT}\left(\beta_T^{\theta^\star}\left(\frac{\delta}{2}\right)+\alpha_T\left(\frac{\delta}{2}\right)\right)\right]\\
&+\frac{1-2\delta}{1-\gamma}\cdot\mathbb{E}_{\theta^\star\sim\pi}\left[\mathrm{RGT}\left(2\beta_T^{\theta^\star}\left(\frac{\delta}{2}\right)+2\alpha_T\left(\frac{\delta}{2}\right)\right)\right].
\end{aligned}$$

By taking the probability of failure δ to be of $O(\frac{1}{T})$, we finally arrive at

$$\begin{aligned}
\mathrm{BayesRGT}(T,M)\leq&O\left(\frac{1}{1-\gamma}\cdot\left(\mathbb{E}_{\theta^\star\sim\pi}\left[B_f^{\theta^\star}\right]+2\sqrt{\frac{d}{\lambda}}L\right)\cdot M\right.\\
&\left.+\frac{1}{1-\gamma}\cdot\mathbb{E}_{\theta^\star\sim\pi}\left[\mathrm{RGT}\left(2\beta_T^{\theta^\star}\left(\frac{1}{2T}\right)+2\alpha_T\left(\frac{1}{2T}\right)\right)\right]\right),
\end{aligned}\tag{A.55}$$

where $B_f^{\theta^\star}=2L\|\theta^\star\|$ and $\frac{1}{1-\gamma}=O\left(\frac{\sqrt{d}}{\mu}\right)$.

The orders of two expectations in (A.55) is claimed as follows.

Claim A.15. The orders of $\mathbb{E}_{\theta^\star\sim\pi}\left[B_f^{\theta^\star}\right]$ and $\mathbb{E}_{\theta^\star\sim\pi}\left[\mathrm{RGT}\left(2\beta_T^{\theta^\star}\left(\frac{1}{2T}\right)+2\alpha_T\left(\frac{1}{2T}\right)\right)\right]$ are:

- $\mathbb{E}_{\theta^\star\sim\pi}\left[B_f^{\theta^\star}\right]$ is of order
$$O\left(\sqrt{\frac{d}{\lambda}}L\right).\tag{A.56}$$
- $\mathbb{E}_{\theta^\star\sim\pi}\left[\mathrm{RGT}\left(2\beta_T^{\theta^\star}\left(\frac{1}{2T}\right)+2\alpha_T\left(\frac{1}{2T}\right)\right)\right]$ is of order
$$O\left(\frac{L}{\sqrt{\lambda}}d\sqrt{M}(\sqrt{T}+\sqrt{dM})\log\left(\frac{\sigma^2\lambda d+TML^2}{\sigma^2\lambda d}\right)\right).\tag{A.57}$$

Proof. See Appendix D.10. □

Therefore, use $\widetilde{O}$ to hide logarithmic term and lower $O(1)$ order term on T, recall $\frac{1}{1-\gamma}=O\left(\frac{\sqrt{d}}{\mu}\right)$ and $L=\sqrt{d}$, we finally arrived at a Bayesian regret of order

$$\mathrm{BayesRGT}(T,M)=\widetilde{O}\left(\frac{\sqrt{d}}{\mu}\cdot\frac{L}{\sqrt{\lambda}}\cdot d\sqrt{MT}\right)=\widetilde{O}\left(\frac{d}{\mu\sqrt{\lambda}}\cdot d\sqrt{MT}\right).\tag{A.58}$$

B Real-world experiment validation

B.1 Optimizing sequence fitness for CRISPR gene-editing

Our TS-DE method was adapted for use in a gene-editing application in real-world experiments. Briefly, gene-editing, exemplified by technology derived from the Clustered Regularly Interspaced Short Palindromic Repeats, or CRISPR system, is a powerful tool for engineering genetic information in living organisms, and has transformed basic research and human therapeutics [13]. The efficiency and outcome of CRISPR gene-editing is highly dependent on the selection of guideRNA sequences, which form a complex with CRISPR proteins to perform gene-editing [44]. The TS-DE was applied to guide high-throughput CRISPR gene-editing experiments. In particular, we use known genomic motif features and a linear model for modeling the log editing capacity. At the beginning of each round of experiment, we computationally generate a new library of design sequences by randomly generating mutations and recombinations based on the previous population. Then we apply the bandit linear model to select sequences with high predicted fitness, and evaluate their actual editing capacities in the next round of experiments. A total of 14,358 unique guideRNA sequences were measured, and the log capacity improved by ≈ 5. Notably, the optimized CRISPR designs generated by our DE approach is part of another manuscript [26], demonstrating real-world utility of current method. See Fig.B.1 for an illustration of the pipeline. We refer to Hughes et al, 2022 for more details on the experiment and computation.

Remark The above real-world application of bandit DE differs from Algorithm 1 and generalizes it in a number of ways. For example, features used for predicting the gene-editing efficiency are not limited to motif features. Also they are not binary valued. Second, recombination and mutation were not done exactly as in Modules 1 and 2. They were randomized on the basepair level rather than the motif level. Despite these differences, our method was able to guide the experiment and accelerate discovery. This demonstrates the bandit DE method may have broad generalizability and it is not restricted to the abstract mathematical model formulated in this paper.

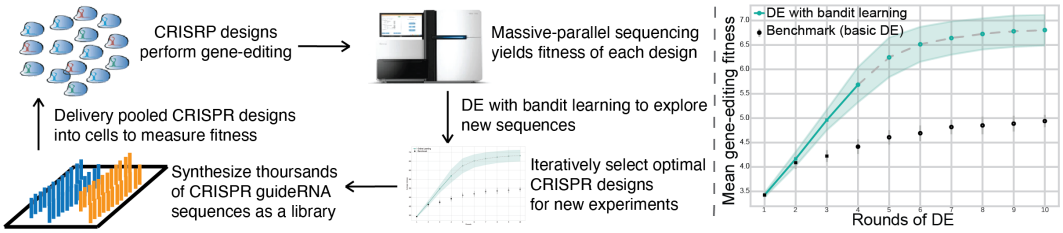

Figure B.1: **Evolving CRISPR sequences using iterative real-world experiments and accelerated DE** Left panels: Workflow overview. Right panel: Fitness distribution showing accelerated optimization using DE with Bandit learning. (Ths figure is borrowed from [26])

C Proof of Theorem 5.1

C.1 Ascent property of `Crossover_Selection`

Proof. Since each $z \in S'$ is generated in the same way independently and $F(S')$ is the fitness averaging over all z's, thus

$$\mathbb{E}\left[F(S')\right] = \mathbb{E}\left[f(z) | \, z \in S'\right],$$

with the expectation taken over the randomness in sampling z's parents x and y and in crossing over x and y. Using notation $\mathbb{E}_{x,y}\left[\cdot\right] := \mathbb{E}\left[\cdot | \, x, y\right]$, the conditional expectation given x and y, rewrite $\mathbb{E}\left[f(z) | \, z \in S'\right]$ as

$$\mathbb{E}\left[f(z) | \, z \in S'\right] = \mathbb{E}\left[\mathbb{E}_{x,y}\left[f(z) \,\middle|\, f(z) \geq \frac{f(x)+f(y)}{2}\right]\right],$$

where the inner expectation is taken over the randomness in the recombination step $z \leftarrow \texttt{Rcb}(x, y)$, and the outer expectation is over sampling x and y.

Given x and y, a recombined child sample z can be represented by

$$z = \frac{x+y}{2} + \frac{x-y}{2} \cdot e, \tag{C.1}$$

where the $\cdot$ operator here multiplies two vector entrywisely into a new vector and e is a vector consisting of d independent Rademacher variables, that is $e = (e_i, \cdots, e_d)^\top$ and

$$e_i \overset{\text{i.i.d.}}{\sim} \text{Rad}\,.$$

Thus $f(z)$ is computed as

$$f(z) = \frac{f(x)+f(y)}{2} + \frac{1}{2}\sum_{i=1}^{d} \theta_i \left(x_i - y_i\right) e_i. \tag{C.2}$$

And then $f(z) \geq \frac{f(x)+f(y)}{2}$ is equivalent to $\sum_{i=1}^{d} \theta_i \left(x_i - y_i\right) e_i \geq 0$, so

$$\begin{aligned}
&\mathbb{E}_{x,y}\left[f(z) \,\middle|\, f(z) \geq \frac{f(x)+f(y)}{2}\right] \\
=&\frac{f(x)+f(y)}{2} + \frac{1}{2}\mathbb{E}\left[\sum_{i=1}^{d} \theta_i \left(x_i - y_i\right) e_i \,\middle|\, \sum_{i=1}^{d} \theta_i \left(x_i - y_i\right) \geq 0\right] \\
&\frac{f(x)+f(y)}{2} + \frac{1}{2}\mathbb{E}\left[\left|\sum_{i=1}^{d} \theta_i \left(x_i - y_i\right) e_i\right|\right] \qquad \text{(C.3)} \\
\geq&\frac{f(x)+f(y)}{2} + \frac{C}{2}\|\theta \cdot (x-y)\,\|,
\end{aligned}$$

where (C.3) holds because $\sum_{i=1}^{d} \theta_i \left(x_i - y_i\right) e_i$ is symmetrically distributed around 0. And in the last line, $\cdot$ is still the entrywise multiplication between vectors and $C \geq \frac{1}{\sqrt{2}}$ according to Haagerup [20].

Thus,

$$\begin{aligned}
\mathbb{E}\left[F(S')\right] &= \mathbb{E}\left[\mathbb{E}_{a,b}\left[f(z) \,\middle|\, f(z) \geq \frac{f(x)+f(y)}{2}\right]\right] \\
&\geq \mathbb{E}\left[\frac{f(x)+f(y)}{2}\right] + \frac{1}{2\sqrt{2}}\mathbb{E}\left[\|\theta \cdot (x-y)\,\|\right] \\
&\geq F(S) + \frac{1}{2\sqrt{2}}\mathbb{E}\left[\|\theta \cdot (x-y)\,\|\right]. \qquad \text{(C.4)}
\end{aligned}$$

By Cauchy-Schwarz, we have

$$\|\theta \cdot (x-y)\,\| \geq \frac{1}{\sqrt{d}}\sum_{i=1}^{d} |\theta_i||x_i - y_i|$$

Thus, by averaging over all x and y sampled from S,

$$\mathbb{E}\left[\|\theta \cdot (x-y)\,\|\right] \geq \frac{1}{\sqrt{d}}\sum_{i=1}^{d} |\theta_i|\mathbb{E}\left[|x_i - y_i|\right]$$

When $\forall i \in [d], x_i, y_i \in \{0,1\}$ for all x and y's in S, then

$$\begin{aligned}
&\mathbb{E}\left[|x_i - y_i|\right] \geq \mathbb{E}\left[\left(x_i - y_i\right)^2\right] = 2Var_i(S), \\
&\mathbb{E}\left[\|\theta \cdot (x-y)\,\|\right] \geq \frac{2}{\sqrt{d}}\sum_{i=1}^{d} |\theta_i| Var_i(S), \qquad \text{(C.5)}
\end{aligned}$$

where $Var_i(S)$ denotes the variance of x_i when x is uniformly sampled from S.

Therefore,

$$\mathbb{E}\left[F(S')\right] \geq F(S) + \frac{1}{\sqrt{2d}}\sum_{i} |\theta_i| Var_i(S).$$

□

D Omitted Proofs in Appendix A

D.1 Proof of Lemma A.4

Proof. For $\forall i \notin \mathcal{I}, \forall x \in S$ is not induced to mutate at site i, thus for $x' = \texttt{Mut}(x, \mathcal{I}, \mu)$, $x'_i = x_i$ and

$$p_i(S') = p_i(S).$$

For $i \in \mathcal{I}$, after the directed mutation formulated as (3.3), $\mathbb{E}\left[\mathbf{I}\left\{x'_i = 1\right\}\right] = (1-\mu)\mathbf{I}\left\{x_i = 1\right\} + \frac{\mu}{2}$.

$$\mathbb{E}\left[p_i(S')\right] = (1-\mu)p_i(S) + \frac{\mu}{2} = p_i(S) + \left(\frac{1}{2} - p_i(S)\right)\mu.$$

Since $i \in \mathcal{I}$ iff $\frac{1}{M}\sum_{x\in S}\theta_i \cdot x_i \leq \theta_i \cdot \bar{x}_i$, which is equivalent to $p_i(S) \leq \frac{1}{2}$, showing that the i-th dimension is not sufficient with the favored feature. Then the directed mutation strictly increases $p_i(S)$ for any insufficient dimension i by boosting it by $\mu\left(\frac{1}{2} - p_i(S)\right) \geq 0$, which resulting in a $|\theta_i|$-increase in the utility value per unit of increase in $p_i(S)$.

Therefore, $\mathbb{E}\left[F(S')\right] \geq F(S)$ and

$$\begin{aligned}
\mathbb{E}\left[p_i(S')\right] &= p_i(S) > \frac{1}{2}, \quad \forall i \notin \mathcal{I}, \\
\mathbb{E}\left[p_i(S')\right] &= p_i(S) + \left(\frac{1}{2} - p_i(S)\right)\mu \geq \frac{\mu}{2}, \quad \forall i \in \mathcal{I}.
\end{aligned}$$

Thus, after calling $S' = \texttt{Directed\_Mutation}(f, S, \mu)$, $\mathbb{E}\left[p_i(S')\right] \geq \frac{\mu}{2}, \forall i \in [d]$. By a standard argument of concentration and a union bound taken over $i \in [d]$, with probability $1-\delta$,

$$p_i(S') \geq \frac{\mu}{4}, \quad \forall i \in [d]$$

when $|S| = \Omega\left(\frac{\log(\frac{d}{\delta})}{\mu^2}\right)$. □

D.2 Proof of Lemma A.6

Proof. Lemma A.6 is derived by taking union bound over $t+1 \in [T]$ upon $\min_i p_i^{\widetilde{\theta}_{t+1}}(S'_t) \geq \frac{\mu}{4}$ obtained by instantiating (A.4) for S'_t and $f_{\widetilde{\theta}_{t+1}}$ in Lemma A.4. □

D.3 Proof of Lemma A.7

Proof. Recall from Theorem 5.1 that

$$\mathbb{E}\left[F(S')\right] \geq F(S) + \frac{1}{\sqrt{2d}}\sum_i |\theta_i| \text{Var}_i(S), \tag{5.1 revisited}$$

where $Var_i(S)$ is the variance of x_i when x is uniformly sampled from S. Using $p_i(S)$ defined in Definition A.3

$$\text{Var}_i(S) = p_i(S)\left(1 - p_i(S)\right). \tag{D.1}$$

Then, it suffices to prove

$$\sum_{i=1}^{d} |\theta_i| p_i(S)\left(1 - p_i(S)\right) \geq \min_i p_i(S) \cdot \left(F^\star - F(S)\right).$$

Taking a closer look at the suboptimality gap $F^\star - F(S)$, it is easily observed that

$$F^\star = \sum_{i:\theta_i \geq 0} \theta_i + \sum_{i:\theta_i < 0} 0, \tag{D.2}$$

$$\begin{aligned}
F(S) &= \sum_{i:\theta_i \geq 0} \theta_i \left[p_i(S) \cdot 1 + (1 - p_i(S)) \cdot 0\right] + \sum_{i:\theta_i < 0} \theta_i \left[p_i(S) \cdot 0 + (1 - p_i(S)) \cdot 1\right] \\
&= \sum_{i:\theta_i \geq 0} \theta_i \cdot p_i(S) + \sum_{i:\theta_i < 0} \theta_i \cdot (1 - p_i(S)).
\end{aligned} \tag{D.3}$$

Plugging in (D.2) and (D.3), we have

$$F^\star - F(S) = \sum_i |\theta_i|(1 - p_i(S)). \tag{D.4}$$

Therefore,

$$\sum_{i=1}^{d} |\theta_i| p_i(S) \left(1 - p_i(S)\right) \geq \min_i p_i(S) \cdot \left(F^\star - F(S)\right).$$

□

D.4 Proof of Lemma A.9

Proof. We finish the proof by lower bounding the probabilities of two events $\widehat{E}^{\theta^\star}\left(\frac{\delta}{2}\right)$ and $\widetilde{E}^{\theta^\star}\left(\frac{\delta}{2}\right)$ by $1 - \frac{\delta}{2}$ separately. Recall that for $\forall t \in [T]$

$$\widetilde{\theta}_t \sim \mathcal{N}(\widehat{\theta}_t, V_t^{-1}), \tag{A.9 revisited}$$

$$V_t = \frac{1}{\sigma^2}\Phi_{t-1}^\top \Phi_{t-1} + \lambda I, \tag{A.10 revisited}$$

$$\widehat{\theta}_t = \frac{1}{\sigma^2} V_t^{-1} \Phi_{t-1}^\top U_{t-1}. \tag{A.11 revisited}$$

Bounding $\mathbb{P}\left(\widehat{E}^{\theta^\star}\left(\frac{\delta}{2}\right)\right)$.

Plugging (A.10) into (A.11), we will see $\widehat{\theta}_t$ is related to the regularized least square estimator (RLS):

$$\widehat{\theta}_t = \frac{1}{\sigma^2} V_t^{-1} \Phi_{t-1}^\top U_{t-1} = \left(\Phi_{t-1}^\top \Phi_{t-1} + \sigma^2 \lambda I\right)^{-1} \Phi_{t-1}^\top U_{t-1}.$$

For any fixed ground truth $\theta^\star$, $\widehat{\theta}_t$ is a RLS estimator of $\theta^\star$ regularized by $\sigma^2\lambda \cdot I$. Conditioned on $\theta^\star$, define a filtration w.r.t. the data $\{(x_{t,i}, u(x_{t,i})), i \in [M], t \in [T-1]\}$ collected along the way.

Definition D.1. Define $\mathcal{F}_t$ be the information accumulated after the t-th batch of data points is collected.

$$\mathcal{F}_0 := \sigma\left(\theta^\star\right) \tag{D.5}$$

$$\mathcal{F}_t := \{\mathcal{F}_{t-1}, \sigma\left(x_{t,1}, u(x_{t,1}), \cdots, x_{t,M}, u(x_{t,M})\right)\}. \tag{D.6}$$

Then we fine grind the filtration $\{F_t\}_{t=0}^{T-1}$ to be

$$\mathcal{F}_0 \subset \mathcal{F}_{0,1} \subset \cdots \subset \mathcal{F}_{0,M} \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_{t-1} \subset \mathcal{F}_{t-1,1} \subset \cdots \mathcal{F}_{t-1,M} \subset \mathcal{F}_{t-1} \subset \cdots \subset \mathcal{F}_{T-1} \tag{D.7}$$

by essentially adding M layers between $\mathcal{F}_{t-1}$ and $\mathcal{F}_t$ and each layer $\mathcal{F}_{t-1,i}$ contains the information obtained after $(x_{t,i}, u(x_{t,i}))$ is added to the dataset.

Under Assumption 3.5, each feedback $u(x_{t,i})$ satisfies

$$u(x_{t,i}) = f_{\theta^\star}(x_{t,i}) + \xi_{t,i}, \tag{A.8 revisited}$$

where

$$\xi_{t,i} | \mathcal{F}_{t,i} \sim \mathcal{N}\left(0, \sigma^2\right).$$

Then bounding $\mathbb{P}\left(\widehat{E}^{\theta^\star}\left(\frac{\delta}{2}\right)\right)$ is a straightforward application of the Theorem 2 in [1], wrapped up into the following proposition.

Proposition D.2. Under Assumption 3.5, for $\forall \delta \in (0,1)$ and any $\mathcal{F}_{t,i}$-adapted data sequence $\left(\{x_{0,i}\}_{i=1}^M, \cdots, \{x_{T-1,i}\}_{i=1}^M\right)$ s.t. $\|x_{t,i}\| \leq L$,

$$\mathbb{P}\left(\exists t \in [T] : \left\|\widehat{\theta}_t - \theta^\star\right\|_{V_t} \geq \beta_t(\delta) \middle| \mathcal{F}_0\right) \leq \delta. \tag{D.8}$$

From the result above, we have

$$\begin{aligned}\mathbb{P}\left(\widehat{E}^{\theta^\star}\left(\frac{\delta}{2}\right)\right) &= \mathbb{P}\left(\left\|\widehat{\theta}_t-\theta^\star\right\|_{V_t}\le\beta_t\left(\frac{\delta}{2}\right),\forall t\in[T]\middle|\theta^\star\right)\\ &= 1-\mathbb{P}\left(\exists t\in[T]:\left\|\widehat{\theta}_t-\theta^\star\right\|_{V_t}\ge\beta_t\left(\frac{\delta}{2}\right)\middle|\theta^\star\right)\\ &\ge 1-\frac{\delta}{2}.\end{aligned}$$

Bounding $\mathbb{P}\left(\widehat{E}^{\theta^\star}\left(\frac{\delta}{2}\right)\right)$. Recall that $\widetilde{\theta}_t$ is sampled from posterior distribution $\mathcal{N}(\widehat{\theta}_t, V_t^{-1})$ independently from $\theta^\star$, then we have

$$\left\|\widetilde{\theta}_t-\widehat{\theta}_t\right\|_{V_t}^2=\left\|V_t^{\frac{1}{2}}(\widetilde{\theta}_t-\widehat{\theta}_t)\right\|^2,\quad\forall t\in[T] \tag{D.9}$$

where $V_t^{\frac{1}{2}}(\widetilde{\theta}_t-\widehat{\theta}_t)\sim\mathcal{N}(0,\mathbf{I})$. Thus $\left\|\widetilde{\theta}_t-\widehat{\theta}_t\right\|_{V_t}^2\sim\chi_d^2$ independently from $\theta^\star$. From the concentration of χ_d^2 random variable, we have

$$\mathbb{P}\left(\chi_d\ge 2\sqrt{d\log\left(\frac{1}{\delta}\right)}+\sqrt{d}\right)\le\delta.$$

Therefore, by taking a union bound over $\forall t\in[T]$, we have

$$\begin{aligned}\mathbb{P}\left(\widetilde{E}\left(\frac{\delta}{2}\right)\middle|\theta^\star\right) &= \mathbb{P}\left(\widetilde{E}\left(\frac{\delta}{2}\right)\right)\\ &= \mathbb{P}\left(\left\|\widetilde{\theta}_t-\widehat{\theta}_t\right\|_{V_t}\le\alpha_t\left(\frac{\delta}{2}\right),\forall t\in[T]\right)\\ &\ge 1-\sum_{t=1}^{T}\mathbb{P}\left(\left\|\widetilde{\theta}_t-\widehat{\theta}_t\right\|_{V_t}\le\alpha_t\left(\frac{\delta}{2}\right)\right)\\ &\ge 1-\sum_{t=1}^{T}\frac{\delta}{2T}=1-\frac{\delta}{2}.\end{aligned}$$

□

D.5 Proof of Proposition D.2

Proof. Use notation $\widetilde{V}_t := \Phi_{t-1}^\top\Phi_{t-1}+\sigma^2\lambda I$, then $V_t=\frac{1}{\sigma^2}\widetilde{V}_t$ and $\widehat{\theta}_t=\widetilde{V}_t^{-1}\Phi_{t-1}^\top U_{t-1}$. According to Theorem 2 in [1], for $\forall\delta\in(0,1)$ and any $\mathcal{F}_{t,i}$-adapted data sequence $\left(\{x_{0,i}\}_{i=1}^M,\cdots,\{x_{T-1,i}\}_{i=1}^M\right)$ s.t. $\|x_{t,i}\|\le L$,

$$\mathbb{P}\left(\exists t\in[T]:\left\|\widehat{\theta}_t-\theta^\star\right\|_{\widetilde{V}_t}\ge\sigma\cdot\beta_t(\delta)\middle|\mathcal{F}_0\right)\le\delta.$$

Therefore, since $\left\|\widehat{\theta}_t-\theta^\star\right\|_{V_t}=\left\|\widehat{\theta}_t-\theta^\star\right\|_{\frac{1}{\sigma^2}\widetilde{V}_t}=\frac{1}{\sigma}\left\|\widehat{\theta}_t-\theta^\star\right\|_{\widetilde{V}_t}$

$$\mathbb{P}\left(\exists t\in[T]:\left\|\widehat{\theta}_t-\theta^\star\right\|_{V_t}\ge\beta_t(\delta)\middle|\mathcal{F}_0\right)\le\delta.$$

□

D.6 Proof of Lemma A.10

Proof. We are about to take a closer look at the incremental increase of the determinant of V_t, define $V_{t,l}=\frac{1}{\sigma^2}\left(\sigma^2\lambda I+\sum_{i=1}^{t-1}\sum_{j=1}^{M}x_{i,j}x_{i,j}^T+\sum_{j=1}^{l}x_{t,j}x_{t,j}^T\right)$ for $\forall t\in[T],l\in[M]$ and thus

$V_{t,0} = V_t$. Mark the time steps where V_t has significant increase in its determinant by $C := \{t \in [T] : \frac{\det(V_{t+1})}{\det(V_t)} > 2\}$. Then the prediction errors in T rounds can be divided into two parts as

$$\sum_{t=1}^{T}\sum_{i=1}^{M} |\langle a_t - b_t, x_{t,i}\rangle| = \sum_{t\notin C}\sum_{i=1}^{M} |\langle a_t - b_t, x_{t,i}\rangle| + \sum_{t\in C}\sum_{i=1}^{M} |\langle a_t - b_t, x_{t,i}\rangle| . \tag{D.10}$$

The first half of (D.10) consists of error accumulated in the rounds where $\det(V_t)$ didn't increased much after having a batch update of size M, so we bound this part in the same spirit of bounding the case where only rank-1 update happens per round. Result is stated in the following claim.

Claim D.3. The first half of (D.10) is bounded by

$$\sum_{t\notin C}\sum_{i=1}^{M} |\langle a_t - b_t, x_{t,i}\rangle| \le \eta_T(\delta)\sqrt{\frac{2L^2+2\lambda}{\lambda}}\sqrt{MT\log\left(\frac{\det(V_{T+1})}{\det(V_1)}\right)}. \tag{D.11}$$

For the second half of (D.10), we are about to bound by showing $|C|$ is small. Notice that

$$\frac{\det(V_{T+1})}{\det(V_1)} = \prod_{t=1}^{T}\frac{\det(V_{t+1})}{\det(V_t)} \ge \prod_{t\in C}\frac{\det(V_{t+1})}{\det(V_t)} \ge 2^{|C|}, \tag{D.12}$$

thus $|C|$ should not be greater than $2\log\left(\frac{\det(V_{T+1})}{\det(V_1)}\right)$. And for $\forall t \in [T], i \in [M]$

$$|\langle a_t - b_t, x_{t,i}\rangle| \le \|a_t - b_t\|_{V_t}\|x_{t,i}\|_{V_t^{-1}} \le \frac{\eta_T(\delta)L}{\sqrt{\lambda}}. \tag{D.13}$$

Putting two parts together, we have

$$\begin{aligned}
\sum_{t=1}^{T}\sum_{i=1}^{M} |\langle a_t - b_t, x_{t,i}\rangle| =& \sum_{t\notin C}\sum_{i=1}^{M} |\langle a_t - b_t, x_{t,i}\rangle| + \sum_{t\in C}\sum_{i=1}^{M} |\langle a_t - b_t, x_{t,i}\rangle| \\
\le& \eta_T(\delta)\sqrt{\frac{2L^2+2\lambda}{\lambda}}\sqrt{MT\log\left(\frac{\det(V_{T+1})}{\det(V_1)}\right)} + \eta_T(\delta)\frac{L}{\sqrt{\lambda}}|C|M \\
\le& \eta_T(\delta)\sqrt{\frac{2L^2+2\lambda}{\lambda}}\sqrt{MT\log\left(\frac{\det(V_{T+1})}{\det(V_1)}\right)} + \eta_T(\delta)\frac{2L}{\sqrt{\lambda}}M\log\left(\frac{\det(V_{T+1})}{\det(V_1)}\right) \\
\le& \eta_T(\delta)\sqrt{\frac{2L^2+2\lambda}{\lambda}}\sqrt{dMT\log\left(\frac{\sigma^2 d\lambda + MTL^2}{\sigma^2 d\lambda}\right)} + \eta_T(\delta)\frac{2L}{\sqrt{\lambda}}dM\log\left(\frac{\sigma^2 d\lambda + MTL^2}{\sigma^2 d\lambda}\right).
\end{aligned}$$

where the final line is referring to the result in [1] that

$$\log\left(\frac{\det(V_{T+1})}{\det(V_1)}\right) \le d\log\left(\frac{\sigma^2 d\lambda + MTL^2}{\sigma^2 d\lambda}\right).$$

□

D.7 Proof of Claim D.3

Proof. With probability $1-\delta$, for $\forall t \in [T]$, normalized by V_t, a_t and b_t concentrate around each other with in a radius of $\eta_t(\delta)$, thus

$$|\langle a_t - b_t, x_{t,i}\rangle| \le \|a_t - b_t\|_{V_t}\|x_{t,i}\|_{V_t^{-1}} \le \eta_t(\delta)\|x_{t,i}\|_{V_t^{-1}}.$$

Summing over all individuals in time steps $t \notin C$, we have

$$\begin{aligned}
\sum_{t\notin C}\sum_{i=1}^{M}|\langle a_t - b_t, x_{t,i}\rangle| &\le \sum_{t\notin C}\sum_{i=1}^{M}\eta_t(\delta)\|x_{t,i}\|_{V_t^{-1}} \\
&\le \eta_T(\delta)\sum_{t\notin C}\sum_{i=1}^{M}\|x_{t,i}\|_{V_t^{-1}} \\
&\le \eta_T(\delta)\sqrt{MT\sum_{t\notin C}\sum_{i=1}^{M}\|x_{t,i}\|^2_{V_t^{-1}}} \\
&\le \eta_T(\delta)\sqrt{\frac{L^2+\lambda}{\lambda}\cdot MT\sum_{t\notin C}\sum_{i=1}^{M}\log\left(1+\|x_{t,i}\|^2_{V_t^{-1}}\right)}, \qquad \text{(D.14)}
\end{aligned}$$

where (D.14) holds because

$$\|x_{t,i}\|^2_{V_t^{-1}} \le \lambda_{\max}(V_t^{-1})\|x_{t,i}\|^2 \le \frac{L^2}{\lambda}.$$

Continuing from (D.14), we can substitute the normalization matrix V_t^{-1} with $V_{t,i}^{-1}$, at the cost of inflating by 2, and then following the classic self-normalized bound on data points. Recall the Lemma 12 in [1]:

$$\frac{\|x\|^2_{\boldsymbol{A}}}{\|x\|^2_{\boldsymbol{B}}} \le \frac{\det(\boldsymbol{A})}{\det(\boldsymbol{B})}, \quad \text{if } \boldsymbol{A}\succeq\boldsymbol{B}. \qquad \text{(D.15)}$$

Substituting V_t^{-1} with $V_{t,i-1}^{-1}$ in $\|x_{t,i}\|^2_{V_t^{-1}}$, noticing $\frac{\det(V_t^{-1})}{\det(V_{t,i-1}^{-1})} = \frac{\det(V_{t,i-1})}{\det(V_t)} \le \frac{\det(V_{t+1})}{\det(V_t)} \le 2$ when $t \notin C$, leads to

$$\begin{aligned}
\|x_{t,i}\|^2_{V_t^{-1}} &\le 2\|x_{t,i}\|^2_{V_{t,i}^{-1}}, \\
\log\left(1+\|x_{t,i}\|^2_{V_t^{-1}}\right) &\le \log\left(1+2\|x_{t,i}\|^2_{V_{t,i-1}^{-1}}\right) \\
&\le 2\log\left(1+\|x_{t,i}\|^2_{V_{t,i-1}^{-1}}\right).
\end{aligned}$$

Then it follows the self-normalized bound in [1] and gives that

$$\begin{aligned}
\sum_{t\notin C}\sum_{i=1}^{M}\log(1+\|x_{t,i}\|^2_{V_t^{-1}}) &\le 2\sum_{t\notin C}\sum_{i=1}^{M}\log\left(1+\|x_{t,i}\|^2_{V_{t,i-1}^{-1}}\right) \\
&\le 2\sum_{t=1}^{T}\sum_{i=1}^{M}\log\left(1+\|x_{t,i}\|^2_{V_{t,i-1}^{-1}}\right) \\
&\le 2\log\left(\frac{\det(V_{T+1})}{\det(V_1)}\right).
\end{aligned}$$

Therefore, the first half of (D.10) is bounded by

$$\sum_{t\notin C}\sum_{i=1}^{M}|\langle a_t - b_t, x_{t,i}\rangle| \le \eta_T(\delta)\sqrt{\frac{2L^2+2\lambda}{\lambda}}\sqrt{MT\log\left(\frac{\det(V_{T+1})}{\det(V_1)}\right)}.$$

□

D.8 Proof of Lemma A.12

Proof. By triangle inequality

$$\mathbb{E}_{E^{\theta^\star}}\left[\|\widetilde{\theta}_t\|\right] \le \|\theta^\star\| + \mathbb{E}_{E^{\theta^\star}}\left[\|\widetilde{\theta}_t - \widehat{\theta}_t\|\right] + \mathbb{E}_{E^{\theta^\star}}\left[\|\widehat{\theta}_t - \theta^\star\|\right].$$

Along with $\lambda_{\min}(V_t) \geq \lambda$ since V_t is regularized with λI in its definition, then we have

$$\mathbb{E}_{E^{\theta^\star}}\left[\|\widetilde{\theta}_t\|\right] \leq \|\theta^\star\| + \frac{1}{\sqrt{\lambda}}\mathbb{E}_{E^{\theta^\star}}\left[\|\widetilde{\theta}_t - \widehat{\theta}_t\|_{V_t}\right] + \frac{1}{\sqrt{\lambda}}\mathbb{E}_{E^{\theta^\star}}\left[\|\widehat{\theta}_t - \theta^\star\|_{V_t}\right].$$

And in event $E^{\theta^\star} := E^{\theta^\star}\left(\frac{\delta}{2}\right) \cap E^{\theta^\star}_{\mathrm{DM}}$, $E^{\theta^\star}_{\mathrm{DM}}$ is independent from the sampling of $\widetilde{\theta}_t$, and conditioned on event $E^{\theta^\star}\left(\frac{\delta}{2}\right)$, both $\|\widetilde{\theta}_t - \widehat{\theta}_t\|_{V_t}$ and $\|\widehat{\theta}_t - \theta^\star\|_{V_t}$ tend to be smaller than it is unconditionally. Thus, we lift the condition on $E^{\theta^\star}$ to get an upper bound as

$$\mathbb{E}_{E^{\theta^\star}}\left[\|\widetilde{\theta}_t\|\right] \leq \|\theta^\star\| + \frac{1}{\sqrt{\lambda}}\mathbb{E}_{\theta^\star}\left[\|\widetilde{\theta}_t - \widehat{\theta}_t\|_{V_t}\right] + \frac{1}{\sqrt{\lambda}}\mathbb{E}_{\theta^\star}\left[\|\widehat{\theta}_t - \theta^\star\|_{V_t}\right]. \tag{D.16}$$

Recall that conditioned on any realization of $\theta^\star$, $\widetilde{\theta}_t$ is sampled from

$$\widetilde{\theta}_t \sim \mathcal{N}(\widehat{\theta}_t, V_t^{-1}) \tag{A.9 revisited}$$

with

$$V_t = \frac{1}{\sigma^2}\Phi_{t-1}^\top \Phi_{t-1} + \lambda I, \tag{A.10 revisited}$$

$$\widehat{\theta}_t = \frac{1}{\sigma^2} V_t^{-1} \Phi_{t-1}^\top U_{t-1}. \tag{A.11 revisited}$$

So $\left\|\widetilde{\theta}_t - \widehat{\theta}_t\right\|^2_{V_t} \sim \chi^2_d$ independent from $\theta^\star$ and thus

$$\mathbb{E}_{\theta^\star}\left[\|\widetilde{\theta}_t - \widehat{\theta}_t\|_{V_t}\right] \leq \sqrt{d}. \tag{D.17}$$

Also, from (A.11), let $U_{t-1} = \Phi_{t-1}\theta^\star + \xi_{t-1}$ and ξ_{t-1} be the corresponding noise vector, then $\widehat{\theta}_t - \theta^\star$ is computed as

$$\begin{aligned}
\widehat{\theta}_t - \theta^\star =& \frac{1}{\sigma^2} V_t^{-1}\Phi_{t-1}^\top U_{t-1} - \theta^\star \\
\overset{A.10}{=}& \left(\Phi_{t-1}^\top \Phi_{t-1} + \sigma^2\lambda I\right)^{-1} \Phi_{t-1}^\top U_{t-1} - \theta^\star \\
=& \left(\Phi_{t-1}^\top \Phi_{t-1} + \sigma^2\lambda I\right)^{-1} \Phi_{t-1}^\top (\Phi_{t-1}\theta^\star + \xi_{t-1}) - \theta^\star \\
=& \left(\Phi_{t-1}^\top \Phi_{t-1} + \sigma^2\lambda I\right)^{-1} \Phi_{t-1}^\top \xi_{t-1} - \sigma^2\lambda\left(\Phi_{t-1}^\top \Phi_{t-1} + \sigma^2\lambda I\right)^{-1}\theta^\star \\
=& \frac{1}{\sigma^2} V_t^{-1}\Phi_{t-1}^\top \xi_{t-1} - \lambda V_t^{-1}\theta^\star.
\end{aligned}$$

Thus,

$$\mathbb{E}_{\theta^\star}\left[\|\widehat{\theta}_t - \theta^\star\|_{V_t}\right] \leq \frac{1}{\sigma^2}\mathbb{E}_{\theta^\star}\left[\|V_t^{-1}\Phi_{t-1}^\top \xi_{t-1}\|_{V_t}\right] + \lambda\mathbb{E}_{\theta^\star}\left[\|V_t^{-1}\theta^\star\|_{V_t}\right], \tag{D.18}$$

where

$$\mathbb{E}_{\theta^\star}\left[\|V_t^{-1}\theta^\star\|_{V_t}\right] = \mathbb{E}_{\theta^\star}\left[\sqrt{\theta^{\star\top} V_t^{-1}\theta^\star}\right] \leq \frac{1}{\sqrt{\lambda}}\|\theta^\star\|, \tag{D.19}$$

and

$$\begin{aligned}
\mathbb{E}_{\theta^\star}\left[\|V_t^{-1}\Phi_{t-1}^\top \xi_{t-1}\|_{V_t}\right] =& \mathbb{E}_{\theta^\star}\left[\sqrt{\xi_{t-1}^\top \Phi_{t-1} V_t^{-1}\Phi_{t-1}^\top \xi_{t-1}}\right] \\
=& \mathbb{E}_{\theta^\star}\left[\mathbb{E}\left[\sqrt{\xi_{t-1}^\top \Phi_{t-1} V_t^{-1}\Phi_{t-1}^\top \xi_{t-1}}\,\middle|\, \Phi_{t-1}\right]\right] \\
\overset{v:=\Phi_{t-1}^\top \xi_{t-1}}{=}& \mathbb{E}_{\theta^\star}\left[\mathbb{E}\left[\sqrt{v^\top V_t^{-1} v}\,\middle|\, \Phi_{t-1}\right]\right],
\end{aligned}$$

with $v \in \mathbb{R}^d$ following the distribution $\mathcal{N}(0, \sigma^2\Phi_{t-1}^\top\Phi_{t-1})$ conditioned on Φ_{t-1} because the noise vector $\xi_{t-1} \mid \Phi_{t-1} \sim \mathcal{N}(0, \sigma^2 I)$. Recall $V_t = \frac{1}{\sigma^2}\Phi_{t-1}^\top\Phi_{t-1} + \lambda I$, therefore

$$\mathbb{E}_{\theta^\star}\left[\|V_t^{-1}\Phi_{t-1}^\top \xi_{t-1}\|_{V_t}\right] \leq \sigma^2\sqrt{d}. \tag{D.20}$$

Plugging (D.19) and (D.20) into (D.18), we have

$$\mathbb{E}_{\theta^\star}\left[\|\widehat{\theta}_t - \theta^\star\|_{V_t}\right] \leq \sqrt{d} + \sqrt{\lambda}\|\theta^\star\|. \tag{D.21}$$

Then plug the inequality above together with (D.17) into (D.16), we finally arrive at

$$\mathbb{E}_{E^{\theta^\star}}\left[\|\widetilde{\theta}_t\|\right] \leq 2\|\theta^\star\| + 2\sqrt{\frac{d}{\lambda}}.$$

□

D.9 Proof of Corollary A.13

Proof. As shown in Lemma A.9, conditioned on event $E^{\theta^\star}\left(\frac{\delta}{2}\right)$, TS estimate $\widetilde{\theta}_t$s are not far away from $\theta^\star$ simultaneously:

$$\|\widetilde{\theta}_t - \theta^\star\|_{V_t} \leq \beta_t^{\theta^\star}\left(\frac{\delta}{2}\right) + \alpha_t\left(\frac{\delta}{2}\right), \quad \forall t \in [T]. \tag{D.22}$$

Thus for $\forall t \in [T-1]$, $\widetilde{\theta}_t$ should not be far away from $\widetilde{\theta}_{t-1}$ with the same high probability. From equation (D.22), we have

$$\begin{aligned}
\|\widetilde{\theta}_t - \theta^\star\|_{V_t} &\leq \beta_t^{\theta^\star}\left(\frac{\delta}{2}\right) + \alpha_t\left(\frac{\delta}{2}\right), \\
\|\widetilde{\theta}_{t+1} - \theta^\star\|_{V_t} \leq \|\widetilde{\theta}_{t+1} - \theta^\star\|_{V_{t+1}} &\leq \beta_{t+1}^{\theta^\star}\left(\frac{\delta}{2}\right) + \alpha_{t+1}\left(\frac{\delta}{2}\right)
\end{aligned}$$

By the triangle inequality of norm $\|\cdot\|_{V_t}$, it holds that

$$\begin{aligned}
\|\widetilde{\theta}_t - \widetilde{\theta}_{t+1}\|_{V_t} &\leq \|\widetilde{\theta}_t - \theta^\star\|_{V_t} + \|\widetilde{\theta}_{t+1} - \theta^\star\|_{V_t} \\
&\leq 2\beta_{t+1}^{\theta^\star}\left(\frac{\delta}{2}\right) + 2\alpha_{t+1}\left(\frac{\delta}{2}\right),
\end{aligned}$$

where the last inequality holds due to the monotonicity in $\{\beta_t^{\theta^\star}\left(\frac{\delta}{2}\right)\}_{t=1}^T$ and $\{\alpha_t\left(\frac{\delta}{2}\right)\}_{t=1}^T$.

Therefore, we have built up the confidence ellipsoid for $\widetilde{\theta}_t - \widetilde{\theta}_{t+1}$ as

$$\|\widetilde{\theta}_t - \widetilde{\theta}_{t+1}\|_{V_t} \leq 2\beta_{t+1}^{\theta^\star}\left(\frac{\delta}{2}\right) + 2\alpha_{t+1}\left(\frac{\delta}{2}\right),$$

which fits into the condition of Lemma A.10 and leads to the result that

$$\sum_{t=1}^{T-1}\sum_{i=1}^{M}\left|\langle\widetilde{\theta}_t - \widetilde{\theta}_{t+1}, x_{t,i}\rangle\right| \leq \mathrm{RGT}\left(2\beta_T^{\theta^\star}\left(\frac{\delta}{2}\right) + 2\alpha_T\left(\frac{\delta}{2}\right)\right).$$

□

D.10 Proof of Claim A.15

Proof. • $\mathbb{E}_{\theta^\star\sim\pi}\left[B_f^{\theta^\star}\right]$.

Recall that $\theta^\star$ is coming from the prior $\mathcal{N}\left(\mathbf{0}, \lambda^{-1}\mathbf{I}\right)$, so $\theta^{\star(i)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \lambda^{-\frac{1}{2}}\right)$, then

$$\begin{aligned}
\mathbb{E}\left[\|\theta^\star\|\right] &\leq \sqrt{\mathbb{E}\left[\|\theta^\star\|^2\right]} = \sqrt{\frac{d}{\lambda}}, \\
\mathbb{E}\left[B_f^{\theta^\star}\right] &= 2L \cdot \mathbb{E}\left[\|\theta^\star\|\right] = O\left(\sqrt{\frac{d}{\lambda}}L\right).
\end{aligned}$$

- $\mathbb{E}_{\theta^\star\sim\pi}\left[\mathrm{RGT}\left(\beta_T^{\theta^\star}\left(\frac{1}{2T}\right)+\alpha_T\left(\frac{1}{2T}\right)\right)\right]$.

  Recall from (A.16)the definition of $\mathrm{RGT}\left(\eta_T(\delta)\right)$ as

  $$\mathrm{RGT}\left(\eta_T(\delta)\right)=\eta_T(\delta)\sqrt{\frac{2L^2+2\lambda}{\lambda}}\cdot\sqrt{dMT\log\left(\frac{\sigma^2 d\lambda+MTL^2}{\sigma^2 d\lambda}\right)}+\eta_T(\delta)\frac{2L}{\sqrt{\lambda}}\cdot dM\log\left(\frac{\sigma^2 d\lambda+MTL^2}{\sigma^2 d\lambda}\right)),$$

  in which only term $\eta_T(\delta)$ is $\theta^\star$ dependent.

  Also recall the definitions of $\beta_T^{\theta^\star}(\delta)$ and $\alpha_T(\delta)$ from (A.14) and (A.15), we have

  $$\mathbb{E}_{\theta^\star\sim\pi}\left[\beta_T^{\theta^\star}\left(\frac{1}{2T}\right)\right]\leq\sqrt{2\log(2T)+d\log\left(\frac{\sigma^2\lambda d+TML^2}{\sigma^2\lambda d}\right)}+\sqrt{d}, \tag{D.23}$$

  $$\mathbb{E}_{\theta^\star\sim\pi}\left[\alpha_T\left(\frac{1}{2T}\right)\right]=2\sqrt{2d\log(2T)}+\sqrt{d}. \tag{D.24}$$

  Plugging into $\eta_T(\delta)=\beta_T^{\theta^\star}(\delta)+\alpha_T(\delta)$, then

  $$\begin{aligned}\mathbb{E}_{\theta^\star\sim\pi}\left[\eta_T\left(\frac{1}{2T}\right)\right]&=\mathbb{E}_{\theta^\star\sim\pi}\left[\beta_T^{\theta^\star}\left(\frac{1}{2T}\right)\right]+\mathbb{E}_{\theta^\star\sim\pi}\left[\alpha_T\left(\frac{1}{2T}\right)\right]\\&\leq\sqrt{2\log(2T)+d\log\left(\frac{\sigma^2\lambda d+TML^2}{\sigma^2\lambda d}\right)}+\sqrt{d}+2\sqrt{2d\log(2T)}+\sqrt{d}\\&=O\left(\sqrt{d\log\left(\frac{\sigma^2\lambda d+TML^2}{\sigma^2\lambda d}\right)}\right).\end{aligned} \tag{D.25}$$

  Therefore, we bound the order of $\mathbb{E}_{\theta^\star\sim\pi}\left[\mathrm{RGT}\left(\beta_T^{\theta^\star}\left(\frac{1}{2T}\right)+\alpha_T\left(\frac{1}{2T}\right)\right)\right]$ by

  $$\begin{aligned}&\mathbb{E}_{\theta^\star\sim\pi}\left[\mathrm{RGT}\left(\beta_T^{\theta^\star}\left(\frac{1}{2T}\right)+\alpha_T\left(\frac{1}{2T}\right)\right)\right]\\=&O\left(\frac{L}{\sqrt{\lambda}}d\sqrt{MT}\log\left(\frac{\sigma^2\lambda d+TML^2}{\sigma^2\lambda d}\right)\right)+O\left(\frac{L}{\sqrt{\lambda}}d^{\frac{3}{2}}M\log\left(\frac{\sigma^2\lambda d+TML^2}{\sigma^2\lambda d}\right)\right).\end{aligned}$$

  □