Feasible Action Search for Bandit Linear Programs via Thompson Sampling

Aditya Gangrade [1] **Aldo Pacchiano** [1 2] **Clayton Scott** [3] **Venkatesh Saligrama** [1]

Abstract

We study the 'feasible action search' (FAS) problem for linear bandits, wherein a learner attempts to discover a feasible point for a set of linear constraints $\Phi_* a \geq 0$, without knowledge of the matrix $\Phi_* \in \mathbb{R}^{m\times d}$. A FAS learner selects a sequence of actions a_t, and uses observations of the form $\Phi_* a_t + \text{noise}$ to either find a point with nearly optimal 'safety margin', or detect that the constraints are infeasible, where the safety margin of an action measures its (signed) distance from the constraint boundary. While of interest in its own right, the FAS problem also directly addresses a key deficiency in the extant theory of 'safe linear bandits' (SLBs), by discovering a safe initialisation for low-regret SLB methods.

We propose a novel efficient FAS-learner based on Thompson Sampling (TS), FAST, which applies a *coupled* random perturbation to an estimate of Φ_*, and plays a maximin point of a game induced by this perturbed matrix. We prove that FAST stops in $\widetilde{O}(d^3/\varepsilon^2 M_*^2)$ steps, and incurs $\widetilde{O}(d^3/|M_*|)$ safety cost, to either correctly detect infeasibility, or find a point a_{out} that is at least $(1-\varepsilon)M_*$-safe, where M_* is the *optimal safety margin* of Φ_*. Further, instantiating prior SLB methods with a_{out} yields the first SLB methods that incur $\widetilde{O}(\sqrt{d^3 T/M_*^2})$ regret and $O(1)$ risk without a priori knowledge of a safe action. The main technical novelty lies in the extension of TS to this multiobjective setting, for which we both propose a coupled noise design, and provide an analysis that avoids convexity considerations.

1. Introduction

Linear bandits capture an online approach to the fundamental decision-making paradigm of linear programs (LPs) with unknown objective and noisy measurements of the quality of actions. The generality of this setup lends itself to many applications in, e.g., machine learning, control system, and resource allocation. However, the standard theory assumes that the constraints of these LPs are known, which is unrealistic in a multitude of applications in these domains. This motivates the *safe linear bandit* (SLB) problem, wherein the learner attempts to select actions with high reward while not violating unknown (linear) constraints *at any time* using noisy feedback of the rewards and constraint levels.

The recent literature has described several methods that address this problem in a strong sense, ensuring low regret whilst *never* violating safety, i.e., playing an infeasible action. Such an algorithm *must* be initialised with a safe action, a_{safe}, to begin with, in order to 'seed' the method with an initial safe region to explore over. Moreover, the *safety margin* of a_{safe} (see below) quantitatively affects regret bounds, and so must be large to ensure good performance. However, no prior method discusses *how* to get hold of such a point. This not only limits the applicability of these algorithms, but also hides their true safety costs, since the discovery of such a_{safe} would, with high likelihood, require playing some unsafe actions. We address this question by designing an efficient algorithm for *feasible action search* (FAS).

Concretely, consider the program $\max_{a\in\mathcal{A}} \theta_*^\top a : \Phi_* a \geq \alpha$, where the reward parameters $\theta_* \in \mathbb{R}^d$ and constraint parameters $\Phi_* \in \mathbb{R}^{m\times d}$ are unknown, while $\mathcal{A}$ is a known convex domain. The *safety margin* of an action $a \in \mathcal{A}$, $M(a) := \min_{\lambda\geq 0, \mathbf{1}_m^\top\lambda=1} \lambda^\top \Phi_* a$, measures a 'signed distance' of a from the constraint boundaries. Given $\varepsilon, \delta \in (0,1)$, a FAS learner selects a sequence of actions $\{a_t\} \subset \mathcal{A}$, and accumulates information by observing noisy risk measurements $S_t = \Phi_* a_t + \text{noise}$ in response. This continues until a stopping time, τ, at which point the learner either

- declares infeasibility, i.e., that $\mathcal{A} \cap \{\Phi_* a \geq 0\} = \emptyset$, OR
- outputs an action a_{out}, and a 'certificate' M_{out} such that

$$M(a_{\mathsf{out}}) \geq M_{\mathsf{out}}, \text{ and } M_{\mathsf{out}} \geq (1-\varepsilon)M_*,$$

where $M_* := \max_{a\in\mathcal{A}} M(a)$ is the *optimal safety margin*.

The goal is to minimise τ, while ensuring correctness with probability at least $1-\delta$. Ideal methods should adapt their stopping time to the unknown M_*, incur limited 'safety costs' during exploration, and be computationally efficient.

[1]Boston University [2]Broad Institute of MIT and Harvard [3]University of Michigan. Correspondence to: A. Gangrade <gangrade@bu.edu>.

Proceedings of the 42^{nd} International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025.

Of course, the FAS problem is of interest in its own right, beyond applications to SLBs, with applications to domains such as control, manufacturing, and resource allocation, where practitioners must balance many objectives with one solution. For instance, when designing a manufacturing process, a FAS-learner would find a set of process parameters that balance manifold cost, efficiency, and quality constraints. The parameter ε serves as a dial between the resilience of such parameters to process perturbations, and the cost of identifying the feasible action. In particular, we note that setting $\varepsilon = 1$ recovers a safe action with minimal search costs, but offers no nontrivial safety margin.

Our Contributions. We exploit the design of a recent (intractable) bandit feasibility test for LPs to propose a novel efficient method, 'Feasible Action Search via Thompson Sampling' (FAST), to address the FAS problem for linear constraints. FAST operates by constructing regression estimates of Φ_*, and playing a maximin point of the matrix game induced by randomly perturbing this estimate. This uses $\widetilde{O}(\mathsf{LP\text{-}time})$ computation per round, where $\mathsf{LP\text{-}time}$ is the cost of optimising a linear objective under m linear constraints over $\mathcal{A}$ to constant error. Stoppage is driven by estimates that bound the value of M_*, and we pick the average of played actions as a_{out}.

We show that FAST is reliable, and stops in a near-optimal $\widetilde{O}(\sqrt{d^3/\varepsilon^2 M_*^2})$ rounds if the instance is feasible, or $\widetilde{O}(\sqrt{d^3/|M_*|^2})$ steps if it is infeasible. We further show that FAST accumulates a limited safety cost of $\mathbf{S}_\tau = \widetilde{O}(d^3/|M_*|)$, where $\mathbf{S}_t := \sum_{s\le t}(-\min_i \Phi_* a_t)_+$ measures the extent of constraint violation per round. In each case, the dependence on m/δ is polylogarithmic.

Further, due to its limited safety cost, instantiating prior methods for SLBs with the output of FAST yields the first SLB algorithm that, without a priori knowledge of a safe action, attains $\widetilde{O}(\sqrt{d^3 T/M_*^2})$ regret (see §2), and $\widetilde{O}(d^3/M_*) = O(1)$ safety costs in T rounds. This not only removes the assumption of knowing a_{safe}, but also improves the regret by a factor of $M_*/M(a_{\mathsf{safe}})$, which may be arbitrarily large, but at a (T-independent) safety cost.

Technical Novelty. As with any method based on Thompson Sampling (TS), FAST operates by drawing noisily perturbed estimates of the unknown parameter, $\widetilde{\Phi}_t$, and acting 'greedily' according to a program induced by this $\widetilde{\Phi}_t$. The main challenge is to design an effective distribution on the noise, and to prove the associated bounds.

The influential approach of Abeille & Lazaric (2017) for analysing low-regret linear TS is a natural framework to adopt for this task. However, this approach is rooted in convexity, and further does not consider multiple objectives together. In detail, Abeille & Lazaric (2017) exploit the convexity of the value function of an LP with given constraints, $J(\theta) := \max_{a\in\mathcal{A}} \theta^\top a$, to execute a gradient-based analysis showing that if the noise ensures that the 'global optimism event' $\mathsf{G}_t^{\mathsf{u}} := \{\tilde{\theta}_t : J(\tilde{\theta}_t) \ge J(\theta_*)\}$ occurs at a constant rate, then the regret is bounded. They further exploit the convexity of $J(\theta)$ to show that for certain concrete choices of noise law, this $\mathsf{G}_t^{\mathsf{u}}$ is indeed frequent enough.

In our multiobjective scenario, the role of $J(\theta)$ is instead played by $K(\Phi) := \max_{a\in\mathcal{A}} \min_{\lambda\in\Delta^m} \lambda^\top \Phi a$. However, this K is not convex in Φ, which precludes both the gradient-based approach to showing progress in learning, as well as the convexity-based analysis for the frequency of optimism.

Our analysis of FAST is instead based on a convexity-free approach using the local optimism event

$$\mathsf{L}_t := \{\widetilde{\Phi}_t : \min_{\lambda\in\Delta^m} \lambda^\top \widetilde{\Phi}_t a_* \ge M_*\},$$

where a_* denotes a maximin point of the matrix game induced by Φ_*. In words, L_t is the event that after perturbation, the safety margin of a_* increases. We first argue, via a direct analysis of the value, that as long as the chance of L_t is $\Omega(1)$, a simple test statistic maintains efficient bounds on the true value M_*, enabling fast and reliable stoppage. Next, we provide a generic construction of a coupled noise that ensures that $\mathbb{P}(\mathsf{L}_t|\text{history}) \ge 0.15$. Concretely, this noise design applies *identical* scaled spherical perturbations to each of the rows of an estimate of Φ_*, and we show that L_t is frequent under this noise by directly analysing the behaviour of the perturbed program at a_*. It is interesting to note that in this local analysis, the natural noise design of perturbing each row of an estimate of Φ_* independently is hard to control, and crude analyses only yield ineffective $2^{-\Omega(m)}$ bounds). This local optimism approach may be of independent interest, since it may apply to TS-based algorithms in other scenarios with nonconvex value functions.

1.1. Related Work

Best Arm Identification (BAI). The FAS problem is intimately related to the fixed-confidence BAI problem for linear bandits. Typically, this is formulated for a finite $\mathcal{A}$, and the focus is on identifying the best arm exactly using experimental design based methods (Soare et al., 2014). Several asymptotically optimal (as the confidence parameter $\delta \to 0$) methods have been proposed for this problem (e.g. Fiez et al., 2019; Jedra & Proutiere, 2020; Degenne et al., 2020; Wang et al., 2021). FAS, however, operates over a continuum of actions, which renders the $\Omega(|\mathcal{A}|)$ computation incurred in these methods untenable (note that any reasonable discretisation is of size $2^{\Omega(d)}$). The problem of finding a ε-optimal action in continuum linear bandits has received less attention, but is usually approached by either sampling from a uniform spanner to directly estimate the parameters (Jedra & Proutiere, 2020), or by playing a low-regret algorithm. Our approach is along the latter lines, but with key

differences: we need to deal with finding a near-maximin action for many objectives, our approximation guarantees are multiplicative in the optimal cost rather than additive, and we need to verify the feasibility of the instance. We note that multiplicative approximation is necessary—since M_* is unknown, a ε-safe action for a fixed ε may even be unsafe. Such multiplicative guarantees have been studied for multi-armed bandits (Michel et al., 2022), but we are unaware of a linear bandit treatment. We note the recent work of Li et al. (2024b), which proposes a TS-based sampling strategy for BAI in finite $\mathcal{A}$ linear bandits. While we do not use this approach, adapting it to FAS over continuous $\mathcal{A}$ is an interesting direction.

Best Safe Arm ID (BSAI). Naturally, the FAS problem is intimately related to the problem of discovering a ε-optimal feasible action (e.g. Katz-Samuels & Scott, 2019; Camilleri et al., 2022; Carlsson et al., 2023) for an unknown objective *and* constraint. Again, most of the literature focuses on finite $|\mathcal{A}|$, often in the multi-armed bandit setup, with additive guarantees of both the optimality and safety of actions. Notice that under such an additive guarantee the latter scenario, the selected action can violate the constraints by up to ε. Instead, we study a continuous $\mathcal{A}$, and, with high probability, necessarily output a safe action. Further, we note that the FAS and BSAI problems have distinct structure. For instance, the optimal action in BSAI typically lies close to (if not at) the boundary of the feasible set, and thus simply verifying the safety of such an action requires a huge number of samples. In contrast, in FAS the output lies deep within the feasible set, and solving the FAS problem needs correspondingly fewer samples.

Minimax and Pareto Bandits. FAST operates by exploiting that the FAS problem is equivalent to searching for a point that is near-maximin for the matrix game induced by Φ_*, which naturally relates it to the problem of achieving maximin values in a low-regret sense that has attracted recent attention (O'Donoghue et al., 2021; Maiti et al., 2023; Li et al., 2024a). The extant literature again focuses on the multi-armed bandit setting using UCB-style approaches (with $\Omega(|\mathcal{A}|)$ computation per round), while we study continuous $\mathcal{A}$. Further, the information structure in our study is different from these prior works, reflected in the fact that our bounds scale with $\log(m)$, while theirs scale as $\mathrm{poly}(m)$. Instead of searching for a minimax point, a number of works deal with multiobjectivity by attempting to identify the entire set of Pareto-optimal actions (e.g. Auer et al., 2016). In our notation, this is the set of $a \in \mathcal{A}$ such that $\Phi_* a \not\leq \Phi_* a'$ for all $a' \in \mathcal{A}$. This objective is very different from our setup—we only need to identify a single action, but this must be near-maximin. Note, of course, that the Pareto set may contain actions that are unsafe (and in the worst case, the entire frontier can be composed of unsafe points).

Feasibility Testing. A closely related work to ours is a recent study of a subset of the authors and A. Gopalan (Gangrade et al., 2024b) on feasibility testing of unknown LPs, which is a 'testing version' of our estimation problem. This work uses a frequentist low-regret method based on 'OFUL' (Abbasi-Yadkori et al., 2011) to design a reliable estimate of the sign of M_*, using which they proposed the 'ellipsoidal optimistic-greedy test,' EOGT. Of course, using a version of our Lemma 7, in principle EOGT can also be used for the FAS problem. However, this method is computationally intractable, requiring, in each round, the learner to solve the program

$$\max_{\Phi \in \mathcal{C}} \max_{a \in \mathcal{A}} \min_{\lambda \in \Delta^m} \lambda^\top \Phi a,$$

where $\mathcal{C}$ is a (nonconvex) confidence set for the constraint parameters, structured as a union of polytopes (see §2.1). This problem is in fact NP-hard even if $\mathcal{C}$ were a convex set, and standard L_1-based relaxations (Dani et al., 2008) need the learner to solve $\Omega((2d)^m)$ matrix games in each round. Our procedure FAST both extends the insights of this work to identify safe actions, but also does so efficiently. In the process, we obtain an efficient version of this test as well, which can be obtained by passing $\varepsilon = 1$ to FAST. In passing, we note that this feasibility testing problem is intimately related to the 'minimum threshold testing problem' (Kaufmann et al., 2018), and refer the reader to Gangrade et al. (2024b) for a detailed discussion.

Low-Regret SLBs have attracted attention in the recent literature, and a many low-regret methods no safety violations have been proposed (Amani et al., 2019; Moradipari et al., 2021; Pacchiano et al., 2021; 2024; Hutchinson et al., 2023). These methods fundamentally require the a priori knowledge of a safe action a_{safe} such that $M(a_{\mathsf{safe}}) > 0$, and (under efficient relaxations) the ensuing regret bounds scale as $\widetilde{O}(\sqrt{d^3 T / M(a_{\mathsf{safe}})^2})$. Our solution to the FAS problem not only eliminates this assumption, but also replaces the arbitrary $M(a_{\mathsf{safe}})$ in the regret bounds by the optimal safety margin M_*, improving the resulting regret bounds, albeit at an (unavoidable) $O(1)$ safety cost. The only methods that do not require an assumption of knowing a_{safe} to begin with are known to violate constraints regularly, and only yield $\widetilde{O}(\sqrt{T})$ bounds on safety costs (Gangrade et al., 2024a).

In passing, we note that the above SLB problem is distinct from the 'bandits with knapsacks' problem, wherein learners may violate constraints in each round, but must satisfy long-term aggregate constraints (e.g. Badanidiyuru et al., 2013). An interesting point of contact in this literature lies in the use of primal-dual methods for such problems, which too require the use of a point with nontrivial margin (referred to as a Slater parameter therein). Within this context, Castiglioni et al. (2022, §8) study a related method to identify (or bound) such a Slater parameter, which is related to our FAS problem. We note, however, that this method is

focused on initialising a primal-dual algorithm, and so treats this FAS problem very crudely. In particular, the method does not adapt its stopping behaviour to the size of M_*, nor obtains tight FAS guarantees, unlike our work.

Thompson Sampling (TS). We refer the reader to Russo et al. (2018) for background on TS. For parametric problems, the efficiency of TS has long made this approach attractive, with frequentist regret bounds in the linear setup first demonstrated by Agrawal & Goyal (2013). Abeille & Lazaric (2017) provide a simplification of this analysis that most strongly informs our approach. Of course, these analyses are limited to single objectives, while our focus lies in finding a maximin point of a multiobjective problem. This introduces new challenges arising from the nonconvexity of the value function, and the need for a robust noise design. We address these by developing a systematic convexity-free analysis that directly exploits the saddle-point structure of the optimal noisy value, and further describe a novel coupled noise structure to ensure frequent optimism.

2. Problem Definition and Background

Notation. For $v \in \mathbb{R}^d$, $\|v\|$ denotes the Euclidean 2-norm. For a PSD matrix M, $\|v\|_M := \|M^{1/2}v\|$. An inequality $a \leq b$ is said to hold under an event E if $a\mathbb{1}_{\mathsf{E}} \leq b\mathbb{1}_{\mathsf{E}}$, where $\mathbb{1}_{\mathsf{E}}$ indicates E. Standard Big-O notation is used, and $\widetilde{O}$ hides polylog factors. $\mathbb{S}^d$ is the unit ℓ_2 sphere in $\mathbb{R}^d$, and Δ^d the simplex. For a matrix M, M^i denotes its ith row. Gaussians are denoted $\mathcal{N}$.

Setup. We are concerned with linear programs of the form $\max_{a\in\mathcal{A}}\{\theta_*^\top a : \Phi_* a \geq 0\}$, where $\mathcal{A} \subset \mathbb{R}^d$ is assumed to be a known bounded domain, while $\theta_* \in \mathbb{R}^d, \Phi_* \in \mathbb{R}^{m\times d}$ are unknown reward and constraint parameters. The safe, or feasible, set is denoted $\mathcal{S}_* := \mathcal{A} \cap \{\Phi_* a \geq 0\}$. Note that $\mathcal{S}_*$ may be empty. We define the *safety margin* of an action a as $M(a) := \min_{\lambda\in\Delta^m} \lambda^\top \Phi_* a$. This is equal to $\min_{i\in[1:m]}(\Phi_* a)^i$, and thus captures the smallest 'slack' in the constraints at a if positive, and the greatest violation of them if negative. The *optimal safety margin* is

$$M_* := \max_{a\in\mathcal{A}} M(a) = \max_{a\in\mathcal{A}} \min_{\lambda\in\Delta^m} \lambda^\top \Phi_* a.$$

Note that M_* may be positive or negative. If negative, then $\mathcal{S}_* = \emptyset$, i.e., the instance is infeasible. Throughout, we use (a_*, λ_*) to denote a(ny) saddle point of this game.

Nonzero constraint levels. We note that our setup of $\{\Phi_* a \geq 0\}$ subsumes the general case $\{\Phi_* a \geq \alpha\}$, since we can augment the dimension by one, include α as a column of Φ_*, and include $a^{d+1} = 1$ as a constraint in $\mathcal{A}$.

Play proceeds in rounds indexed by $t \in \mathbb{N}$. At each t, the learner picks an action $a_t \in \mathcal{A}$, and gets the feedback $R_t = \theta_*^\top a_t + w_t^R$, and $S_t = \Phi_* a_t + w_t^S$, where $w_t^R \in \mathbb{R}$ and $w_t^S \in \mathbb{R}^m$ are noise. Let C_t denote randomness available to the learner at round t. We will denote the historical filtration as $\mathfrak{H}_{t-1} = \sigma(\{(a_s, R_s, S_s, C_s)\}_{s<t})$, and $\mathfrak{G}_t := \sigma(\mathfrak{H}_{t-1} \cup \{a_t, C_t\})$. The actions a_t must be adapted to $\sigma(\mathfrak{H}_{t-1} \cup \sigma(\{C_t\}))$.

In the **feasible action search** (FAS) problem, given $\varepsilon, \delta \in (0,1)$, the learner selects actions, and at some point, determines a *stopping time*, τ. Upon stopping, the learner either declares the instance to be infeasible, or outputs an action $a_{\mathsf{out}} \in \mathcal{A}$, along with a certificate M_{out}. We call a FAS learner (ε, δ)-*reliable* if with probability at least $1-\delta$, (a) if $M_* < 0$, the learner declares infeasibility, and (b) if $M_* > 0$, the learner ensures that $M(a_{\mathsf{out}}) \geq M_{\mathsf{out}} \geq (1-\varepsilon)M_*$ w.p. at least $1-\delta$. The goal is to select actions and the stopping rule in a way that maintains reliability, whilst minimising τ (which is the number of samples used). We note that the reward information need not be utilised by a FAS learner. In addition to the stopping time, we will control the *safety cost* of exploration, $\mathbf{S}_\tau$, where for $t \in \mathbb{N}$,

$$\mathbf{S}_t := \sum_{s\leq t} (-\min_{\lambda} \lambda^\top \Phi_* a_t)_+,$$

with $(z)_+ := \max(z, 0)$. Notice that $\mathbf{S}_t$ penalises any excursion outside of the feasible set, but playing within this set does not yield a reduction in $\mathbf{S}_t$.

In the **regret-control** problem, given $\delta \in (0,1)$, the learner's goal is to ensure that w.p. at least $1-\delta$, both the safety cost and the regret are well-controlled, where, with OPT being the optimal value of the program, the *regret* is

$$\mathbf{R}_t := \sum_{s\leq t} (\mathrm{OPT} - \theta_*^\top a_t)_+.$$

When discussing regret-control, we tacitly will assume feasibility, i.e., that $\mathcal{S}_* := \mathcal{A} \cap \{\Phi_* a \leq 0\} \neq \varnothing$.

Standard Assumptions. Throughout, we will assume the following conditions on the instance $(\theta, \Phi, \mathcal{A})$, and noise w. All subsequent results only hold under these assumptions.
•*Boundedness:* it holds that $\max(\|\theta_*\| m \max_i \|\Phi_*^i\|) \leq 1$, and $\mathcal{A} \subset \{a : \|a\| \leq 1\}$.
•*SubGaussian noise:* $w_t := (w_t^R, (w_t^S)^\top)^\top$ is centred and 1-subGaussian with respect to $\mathfrak{G}_t$, i.e., $\mathbb{E}[w_t|\mathfrak{G}_t] = 0$, and $\forall \lambda \in \mathbb{R}^{m+1}, \mathbb{E}[\exp(\lambda^\top w_t)|\mathfrak{G}_t] \leq \exp(\|\lambda^2\|/2)$.

If augmenting the dimension to handle nonzero constraint levels, boundedness only applies to the unknown parts of Φ.

2.1. Background

We include repeatedly used background on online linear regression, and on laws of iterated logarithms.

Confidence Ellipsoids. We estimate the parameters Φ_* given $\mathfrak{H}_{t-1}$ by $\hat{\Phi}_t = \arg\min_{\hat{\Phi}} \sum_{s<t} \|\hat{\Phi} a_s - S_s\|^2 + \sum_i \|\hat{\Phi}^i\|^2$. The standard confidence set for Φ_* is

$$\mathcal{C}_t(\delta) = \{\widetilde{\Phi} : \forall \text{ rows } i, \|\widetilde{\Phi}^i - \hat{\Phi}_t^i\|_{V_t} \leq \omega_t(\delta)\},$$

where $V_t = I + \sum_{s<t} a_s a_s^\top$, and

$$\omega_t(\delta) := 1 + (1/2 \log(2m/\delta) + 1/4 \log\det V_t)^{1/2}.$$

The first key result states that these confidence ellipsoids are *consistent* at all times.

Lemma 1. (Abbasi-Yadkori et al., 2011) *Define the consistency event at time t $\mathsf{Con}_t(\delta) = \{\Phi_* \in \mathcal{C}_t(\delta)\}$, and $\mathsf{Con}(\delta) := \bigcap_{t\geq 1} \mathsf{Con}_t(\delta)$. Under the standard assumptions, for all $\delta \in (0,1)$, $\mathbb{P}(\mathsf{Con}(\delta)) \geq 1-\delta$.*

The following result is termed the *elliptical potential lemma* (Abbasi-Yadkori et al., 2011; Carpentier et al., 2020).

Lemma 2. *For any t, and any sequence of actions $\{a_t\}$ lying in the unit ball, it holds that*

$$\sum_{s\leq t} \|a_s\|_{V_s^{-1}} \leq \sqrt{2t \log\det V_t} \leq \sqrt{2dt \log(1+t/d)}.$$

We will also need the following nonasymptotic *law of iterated logarithms* (LIL) (e.g. Howard et al., 2021).

Lemma 3. *For any filtration $\{\mathfrak{F}_t\}$, if ξ_t is a process adapted to $\mathfrak{F}_t$ that is conditionally centred and 1-subGaussian, then*

$$\forall \delta \in (0,1), \mathbb{P}\Big(\exists t : |Z_t| > \mathrm{LIL}(t,\delta)\Big) \leq \delta, \text{ where}$$

$$Z_t := \sum_{s\leq t} \xi_s, \text{ \& } \mathrm{LIL}(t,\delta) := \sqrt{4t \log(11 \max(\log t, 1)/\delta)}.$$

3. Feasible Action Search via Thompson Sampling

As previously discussed, our algorithm is inspired by the EOGT procedure of Gangrade et al. (2024a). This method uses a low-regret algorithms for linear bandits to construct an approach to estimate the sign of the value of a matrix game. Our strategy is to apply this idea to construct a computationally efficient method by exploiting a randomised approach based on Thompson Sampling.

Our method, FAST is parametrised by a law μ on $\mathbb{R}^{m\times d}$, along with ε, δ. FAST$(\mu, \varepsilon, \delta)$, operates by first drawing a noise matrix $H_t \sim \mu$, independently of $\mathfrak{H}_{t-1}$, and constructing the perturbed constraint matrix

$$\widetilde{\Phi}_t = \Phi(H_t, t, \delta) := \hat{\Phi}_t + \omega_t(\delta) H_t V_t^{1/2}. \tag{1}$$

Observe that the perturbation is aligned with the geometry induced by the historical information via $\omega_t V_t^{1/2}$. Given $\widetilde{\Phi}_t$, we select both an action $a_t \in \mathcal{A}$, as well as a vector $\lambda_t \in \Delta^m$ by solving the game

$$\max_{a\in\mathcal{A}} \min_{\lambda\in\Delta^m} \lambda^\top \widetilde{\Phi}_t a. \tag{2}$$

The action a_t is played, and the ensuing safety feedback is used along with λ_t to construct the main statistic

$$\mathscr{T}_t := \sum_{s\leq t} \lambda_s^\top S_s,$$

Algorithm 1 Feasible Action Search via Thompson Sampling (FAST$(\mu, \varepsilon, \delta)$)

Input: $\varepsilon, \delta \in (0,1)$, μ, a law on $\mathbb{R}^{m\times d}$, and $\mathcal{A}$.
Initialise: $\mathfrak{H}_0 \leftarrow \varnothing$, $\mathscr{T}_0 \leftarrow 0, t \leftarrow 0, \tau \leftarrow -\infty$
while $\tau < 0$ **do**
  Increment t and draw $H_t \sim \mu$ independent of $\mathfrak{H}_{t-1}$.
  $\widetilde{\Phi}_t \leftarrow \hat{\Phi}_t + \omega_t(\delta) H_t V^{-1/2}$.
  Compute λ_t, a_t via (2).
  Play a_t, and observe S_t.
  $\mathscr{T}_t \leftarrow \mathscr{T}_{t-1} + \lambda_t^\top S_t, \langle a\rangle_t \leftarrow \sum_{s\leq t} a_s/t$.
  Update $\mathscr{B}_t(\mu,\delta)$ as defined in Lemma 7.
  $U_t \leftarrow (\mathscr{T}_t + \mathscr{B}_t)/t, L_t \leftarrow (\mathscr{T}_t - \mathscr{B}_t)/t$.
  if $U_t < 0$ or ($L_t > (1-\varepsilon)U_t$ and $U_t > 0$) **then**
    $\tau \leftarrow t$.
if $U_\tau > 0$ **then**
  Return $a_{\mathsf{out}} = \langle a\rangle_\tau$, $M_{\mathsf{out}} = L_\tau$
Return "Instance is infeasible".

which is used to determine stoppage.

There are two important questions to address: how should one select μ, and how should one select the stopping and decision rules to ensure both efficiency and reliability.

To this end, in §3.1, we describe a generic 'concentration-optimism' condition on μ, and show that under this condition, the value of $\mathscr{T}_t$ *tracks* the optimal safety margin in the sense that

$$\mathscr{T}_t \in tM_* \pm \mathscr{B}_t(\mu,\delta),$$

where $\mathscr{B}_t$ is an adapted computable process defined below. Using this, in §3.2 we will describe a simple reliable stopping rule, which either detects that $M_* < 0$, or continues sampling until $\langle a\rangle_t := \sum_{s<t} a_s/t$ is an appropriately feasible action. Next, in §3.3, we provide a generic construction of μs satisfying this CO condition, operationalising this design. Finally, in §3.4, we use these properties to analyse the behaviour of the stopping time and the safety costs of FAST.

3.1. The CO Condition and the Tracking Inequality

As in prior studies of linear TS, the core insight we exploit is that progress is made in learning whenever the perturbation is 'optimistic', i.e., the optimal perturbed value dominates the true value. Thus, it is enough to design a noise that ensures frequent optimism, while limiting blowups in errors due to the scale of these perturbations. For the FAS problem, we encapsulate these two aspects in the following condition.

Definition 4. *Let $B : [0,1] \to \mathbb{R}_{\geq 0}$ be a map, and $\pi \in (0,1]$. Define the* global optimism *event*

$$\mathsf{G}_t(\delta) := \{H_t : K(\widetilde{\Phi}_t) \geq K(\Phi_*)\},$$

where $K(\Phi) := \max_{a\in\mathcal{A}} \min_{\lambda\in\Delta^m} \lambda^\top \Phi a$. Recall the event $\mathsf{Con}_t(\delta) = \{\Phi_* \in \mathcal{C}_t(\delta)\}$ from Lemma 1.

We say that a distribution μ on $\mathbb{R}^{m\times d}$ satisfies a B-concentration condition if $\forall t, \xi \in [0,1]$, it holds that

$$\mu(\{H : \max_i \|H^i\| > B(\xi)\}) \le \xi.$$

Further, we say that μ satisfies a π-optimism condition if

$$\forall \delta, t, \mathbb{1}_{\mathsf{Con}_t(\delta)}\mathbb{E}[\mu(\mathsf{G}_t(\delta))|\mathfrak{H}_{t-1}] \ge \pi \mathbb{1}_{\mathsf{Con}_t(\delta)}.$$

We will succinctly say that μ satisfies a (B,π)-CO condition if both these properties are true.

The concentration property above limits how far the perturbation can move $\bar{\Phi}_t$, and thus, the size of $\widetilde{\Phi}_t - \Phi_*$ under consistency. The event $\mathsf{G}_t(\delta)$ is a 'global optimism' event, in that when $\widetilde{\Phi}_t \in \mathsf{G}_t(\delta)$, the value of the game induced by $\widetilde{\Phi}_t$ dominates the value induced by Φ_* (which equals M_*). Note that $\mathsf{G}_t(\delta)$ dependts on δ through the $\omega_t(\delta)$ term in $\widetilde{\Phi}_t$ in (1). The remainder of this section will derive key consequences of the CO condition, and we leave a generic construction of laws meeting 'good' CO conditions to §3.3.

Roundwise Tracking. The main utility of the CO condition is the following key result.

Lemma 5. *Let $\delta_t = \min(\pi/2, \delta/t(t+1))$, and define the event $\mathsf{Ball}_t(\delta) := \{\max_i \|H_t^i\| \le B(\delta_t)\}$. If μ-satisfies a (B,π)-CO condition, then for all t, under $\mathsf{Ball}_t(\delta)$,*

$$M_* \le \lambda_t^\top \widetilde{\Phi}_t a_t + \frac{4B(\delta_t)\omega_t(\delta)}{\pi}\mathbb{E}[\|a_t\|_{V_t^{-1}}|\mathfrak{H}_{t-1}].$$

Proof Sketch. We defer most details to §A.1, and only highlight the main difference from the prior analysis of Abeille & Lazaric (2017). Using this approach, under $\mathsf{Ball}_t(\delta)$,

$$M_* - K(\widetilde{\Phi}_t) \le \mathbb{E}[K(\widetilde{\Phi}_t) - K(\bar{\Phi}_t)|\mathfrak{H}_{t-1}, \tilde{\mathsf{G}}_t],$$

where $\bar{\Phi}_t$ is the minimiser of $K(\Phi)$ over the set $\mathcal{E}_t = \{\Phi(H,t,\delta) : \max_i \|H^i\| \le B(\delta_t)\}$, and $\tilde{\mathsf{G}}_t = \mathsf{G}_t \cap \mathsf{Ball}_t(\delta)$. At this point, prior analyses use the convexity of the value function in terms of the unknown parameters, which fails for us. Instead, we take the following direct tack.

Let $(\bar{a}_t, \bar{\lambda}_t)$ denote a saddle point of $\bar{\Phi}_t$ such that $\bar{\lambda}_t^\top \bar{\Phi}_t \bar{a}_t = K(\bar{\Phi}_t)$. By the saddle point property, $\bar{\lambda}_t^\top \bar{\Phi}_t a_t \le K(\bar{\Phi}_t)$. By the same coin, $\bar{\lambda}_t^\top \widetilde{\Phi}_t a_t \ge \lambda_t^\top \widetilde{\Phi}_t a_t = K(\widetilde{\Phi}_t)$. So,

$$K(\widetilde{\Phi}_t) - K(\bar{\Phi}_t) \le \bar{\lambda}_t^\top (\widetilde{\Phi}_t - \bar{\Phi}_t) a_t.$$

But, under $\tilde{\mathsf{G}}_t \subset \mathsf{Ball}_t$, each row of $\widetilde{\Phi}_t - \bar{\Phi}_t$ has V_t-norm at most $2B(\delta_t)\omega_t(\delta)$, and thus $\|(\widetilde{\Phi}_t - \bar{\Phi}_t)a_t\|_\infty \le 2B\omega_t\|a_t\|_{V_t^{-1}}$ (see Lemma 11 in §A.1). Since $\|\bar{\lambda}_t\|_1 = 1$, by Hoelder's inequality, we conclude that under Ball_t,

$$M_* - K(\widetilde{\Phi}_t) \le 2B(\delta_t)\omega_t(\delta)\mathbb{E}[\|a_t\|_{V_t^{-1}}|\mathfrak{H}_{t-1}, \tilde{\mathsf{G}}_t].$$

The conclusion follows by using the nonegativity of $\|a_s\|_{V_s^{-1}}$ to see that $\mathbb{E}[\|a_s\|_{V_s^{-1}}|\mathfrak{H}_{t-1}] \ge \mathbb{E}[\mu(\tilde{\mathsf{G}}_t)|\mathfrak{H}_{t-1}]\mathbb{E}[\|a_s\|_{V_s^{-1}}|\mathfrak{H}_{t-1}, \tilde{\mathsf{G}}_t]$, and noting that $\mu(\tilde{\mathsf{G}}_t) \ge \mu(\mathsf{G}_t) - \delta_t$ since $\mu(\mathsf{Ball}_t(\delta)) \ge 1 - \delta_t$. □

In more detail, for the single objective value $J(\theta) = \max_{a\in\mathcal{A}} \theta^\top a$, prior analysis due to Abeille & Lazaric (2017) rely on the convexity of J to bound $J(\tilde{\theta}_t) - J(\bar{\theta}_t)$ by $\nabla J(\tilde{\theta}_t)^\top(\tilde{\theta}_t - \bar{\theta}_t)$, where $\tilde{\theta}_t$ is the perturbed objective, and $\bar{\theta}_t$ is an analogue of $\bar{\Phi}_t$. The final step is to use the fact that $\nabla J(\tilde{\theta}_t) = a_t$ almost surely, which concludes the argument via the Cauchy-Schwarz inequality.

In our case, the map $K(\Phi)$ is nonconvex[1], and we instead approach this question by directly exploiting the saddle point structure of K. This strategy should be useful in other scenarios with unknown constraints by exploiting the saddle point property of a Lagrangian (as long as norms of dual-optimal solutions are controlled).

The above is complemented by a roundwise lower bound.

Lemma 6. *For all t, under $\mathsf{Ball}_t(\delta) \cap \mathsf{Con}_t(\delta)$,*

$$M_* \ge M(a_t) \ge \lambda_t^\top \Phi_* a_t - 2(1 + B(\delta_t))\omega_t(\delta)\|a_t\|_{V_t^{-1}}$$

Proof. Under the event in question, each row of $\widetilde{\Phi}_t - \Phi_*$ has V_t-norm bounded by $\zeta_t = (1 + B(\delta_t))\omega_t(\delta)$. Now, by definition, $M_* \ge M(a_t)$. Further, suppose $M(a_t) = \lambda(a_t)^\top \Phi_* a_t$. Since (a_t, λ_t) is a saddle point for $\widetilde{\Phi}_t$,

$$\lambda(a_t)^\top \widetilde{\Phi}_t a_t \ge \lambda_t^\top \widetilde{\Phi}_t a_t.$$

But, since $\|\lambda(a_t)\|_1 = \|\lambda_t\|_1 = 1$, we further have

$$\lambda(a_t)^\top \Phi_* a_t + \zeta_t\|a_t\|_{V_t^{-1}} \ge \lambda_t^\top \Phi_* a_t - \zeta_t\|a_t\|_{V_t^{-1}}. \quad \square$$

Tracking Inequality. The above roundwise results, coupled with concentration arguments, yield the main consequence of the CO condition for the FAS problem, which relate the behaviour of the statistic $\mathscr{T}$ to M_*. This is shown in §A.1.

Lemma 7. *Define the boundary process*

$$\mathscr{B}_t(\mu,\delta) := \frac{(1 + 5B(\delta_t))\omega_t(\delta)}{\min(\pi, 1/2)}(\mathrm{LIL}(t,\delta) + \sum_{s\le t}\|a_s\|_{V_s^{-1}}).$$

If μ satisfies a (B,π)-CO condition, and $\delta \in (0,1)$, then for $\delta_t := \min(\pi, \delta/t(t+1))$, with probability at least $1 - 4\delta$, $\mathrm{FAST}(\mu,\varepsilon,\delta)$ ensures that simultaneously for all t

$$\mathscr{T}_t + \mathscr{B}_t(\mu,\delta) \ge tM_* \ge \sum_{s\le t} M(a_s) \ge \mathscr{T}_t - \mathscr{B}_t(\mu,\delta).$$

[1] e.g., consider the payoff $\begin{pmatrix} 1-z & z \\ z & z-1 \end{pmatrix}$ over $\mathcal{A} = \Delta^2$ as z varies. Via direct computation (see §A.5.1), we find that $K(z) = z + \min(0, (2z)^{-1} - 1)$, which is nonconvex over any interval containing $z = 1/2$.

3.2. The Design of FAST$(\mu, \varepsilon, \delta)$

The above tracking inequality captures the main thrusts of the design of FAST. The two key processes are

$$U_t := \frac{\mathscr{T}_t + \mathscr{B}_t}{t}, \text{ and } L_t := \frac{\mathscr{T}_t - \mathscr{B}_t}{t}.$$

Under the tracking inequality, these bound the optimal margin M_* as $U_t \geq M_* \geq L_t$. This immediately suggests a reliable stopping criterion: if $U_t < 0$, then $M_* < 0$, and we can declare infeasibility. If instead $L_t > (1-\varepsilon)U_t$ and $U_t > 0$, we can a fortiori conclude that $M_* > 0$ and $L_t > (1-\varepsilon)M_*$. However, yet more is true. Notice that

$$\min_{\lambda} \frac{\lambda^\top \Phi_* \sum_{s\leq t} a_s}{t} \geq \sum_{s\leq t} \min_{\lambda} \frac{\lambda^\top \Phi_* a_s}{t} = \sum_{s\leq t} \frac{M(a_s)}{t}.$$

Thus, we also find that the ergodic average action $\langle a \rangle_t = \sum_{s\leq t} a_s/t$ satisfies $M(\langle a \rangle_t) \geq L_t$. Since $\mathcal{A}$ is convex, this point lies in it, and thus under the second stopping criterion, we a fortiori conclude that $\langle a \rangle_t$ is at least $(1-\varepsilon)M_*$-safe.

This motivates the stopping time of FAST$(\mu, \varepsilon, \delta)$,

$$\tau := \inf\{t : U_t < 0 \text{ OR } L_t > (1-\varepsilon)U_t > 0\}.$$

Upon stopping, we find a reliable decision by declaring infeasiblity if $U_\tau < 0$, and outputting $\langle a \rangle_\tau, L_\tau$ if $U_\tau > 0$.

3.3. Choosing an Effective Noise Distribution

The quantitative effect of the choice of μ on FAST is through the factor $B(\delta_t)/\pi$ appearing in $\mathscr{B}_t(\mu, \delta)$. Since stoppage requires that $\mathscr{T}_t$ dominates $\mathscr{B}_t$, this factor directly scales the costs of FAST, and thus a good μ should control this ratio. We now turn to the problem of designing such μ.

Via a convexity-based analysis, Abeille & Lazaric (2017) show that for single objective TS, optimism is induced if the noise law is anticoncentrated (i.e., the noise vector has large projection along any direction with constant chance) This lies in tension with B-concentration, and a good balance is attained by, e.g., $\mathcal{N}(0, I_d)$, and $\mathrm{Unif}(\sqrt{3d}\mathbb{S}^d)$ (§A.2.1).

A natural idea is to draw each row of H independently from such a law. However, this approach is hard to analyse well: each row of H_t gets an independent shot at reducing the noisy feasible set $\{a : \widetilde{\Phi}_t a \leq 0\}$, and thus at reducing the noisy optimal margin $K(\widetilde{\Phi}_t)$, which suggests that under such noise, π would be of the order $2^{-\Omega(m)}$, where m is the number of unknown constraints. We avoid this issue with a simple fix, *coupling all perturbations*, and via a convexity-free local analysis at a_*, prove the following result in §A.2.

Theorem 8. *For a map $\bar{B} : [0,1] \to \mathbb{R}_{\geq 0}$, and $p \in (0,1]$, let ν be a law on $\zeta \in \mathbb{R}^{d\times 1}$ such that*

$$\nu(\|\zeta\| > \bar{B}(\delta)) \leq \delta, \text{ and } \forall u \in \mathbb{R}^d, \nu(\zeta^\top u \geq \|u\|) \geq p.$$

For $\zeta \sim \nu$, let μ be the law of $H = \mathbf{1}_m \zeta^\top$. Then μ satisfies a (B, π)-CO condition with $\pi = p$, and $B(\xi) = \bar{B}(\xi)$.

The noise design above boils down to setting $H = \mathbf{1}_m \zeta^\top$, where ζ is drawn from a good single-objective law. The core of the analysis shows that under such a noise design, *local optimism at* (a_*, λ_*) is frequent, i.e., that the event

$$\mathsf{L}_t(\delta) := \{\widetilde{\Phi}_t : \min_{\lambda \in \Delta^m} \lambda^\top \widetilde{\Phi}_t a_* \geq M_*\}$$

satisfies $\mathbb{E}[\mu_t(\mathsf{L}_t(\delta))|\mathfrak{H}_{t-1}] \geq p$ under $\mathsf{Con}_t(\delta)$. Notice that in contrast to the global optimism event $\mathsf{G}_t(\delta)$, $\mathsf{L}_t(\delta)$ specifically requires that the value of the perturbed program increases *at a maximal safety margin point* a_*. Of course, $\mathsf{L}_t(\delta) \subset \mathsf{G}_t(\delta)$, so this event ensures global optimism.

The idea of the analysis is as follows. To begin with, consider just the first row of the Φs. Under the event $\mathsf{Con}_t(\delta)$, we know by Lemma 1 and the Cauchy-Schwarz inequality that

$$\hat{\Phi}^1_t a_* \geq \Phi^1_* a_* - \omega_t(\delta)\|a_*\|_{V_t^{-1}}.$$

So, if the noise H^1 satisfies

$$H^1 V_t^{-1/2} a_* \geq \|V_t^{-1/2} a_*\| = \|a_*\|_{V_t^{-1}},$$

then

$$\widetilde{\Phi}^1_t a_* = \hat{\Phi}^1 a_* + \omega_t(\delta) H^1 V_t^{-1/2} a_* \geq \Phi^1_* a_*.$$

But each row $\hat{\Phi}^i_t$ satisfies the *same* property, and so such an H^1 would improve the value of *every* constraint, ergo, the coupled noise $H = \mathbf{1}_m H^1$ would ensure that

$$\begin{aligned}\widetilde{\Phi}_t a_* &= \hat{\Phi}_t a_* + \mathbf{1}_m H^1 V_t^{-1/2} a_* \\ &\geq \Phi_* a_* - \omega_t(\delta)\|a_*\|_{V_t^{-1}} \mathbf{1}_m + \omega_t(\delta)\|a_*\|_{V_t^{-1}} \mathbf{1}_m = \Phi_* a_*,\end{aligned}$$

thus inducing local optimism. The coupled design samples a (column vector) $\zeta \sim \nu$, and then sets $H^1 = \zeta^\top, H = \mathbf{1}_m H^1$. The 'anticoncentration' condition on $\nu(\zeta^\top u \geq \|u\|)$ in Theorem 8 then ensures that no matter the value of V_t, with chance at least p, $H^1(V_t^{-1/2} a_*) \geq \|V_t^{-1/2} a_*\|$. The boundedness of $\zeta \sim \nu$ is of course directly inherited by H^1, thus establishing the mentioned CO condition.

This local approach again bypasses convexity considerations of $K(\Phi)$, which prior proofs of frequent optimism need. We note that even though we only analyse $\mathsf{L}_t \subset \mathsf{G}_t$, the resulting optimism rate we prove is *the same* as these prior analyses, i.e., we derive π-optimism with many unknown constraints under the same conditions under which prior work derives π-optimism for single objectives.

3.4. Stopping Time and Safety Cost Bounds

Instantiating the tracking analysis with our design of effective μ from the previous section, and controlling the growth of $\mathscr{B}_t$ to ${}^B/_\pi \cdot O(\sqrt{d^2 t})$ yields our main result on the stopping time and safety costs of FAST, as shown in §A.3.

Theorem 9. *Let μ be the law of $\mathbf{1}_m\zeta^\top$, where $\zeta \sim \mathrm{Unif}(\sqrt{3d}\cdot\mathbb{S}^d)$, and let $\mathcal{A}$ be convex, and $\log(m/\delta) = o(d)$. Then $\textsc{fast}(\mu, \varepsilon, \delta/5)$ is (ε,δ)-reliable, and w.p. at least $1-\delta$, in the feasible case, the stopping time is bounded as*

$$\tau = \widetilde{O}\left(\frac{d^3 + d^2\log(m/\delta)}{\varepsilon^2 M_*^2}\right),$$

and the excess-risk incurred is bounded as

$$\mathbf{S}_\tau = \widetilde{O}\left(\frac{d^3}{M_*} + \frac{d^{5/2}\sqrt{\log(m/\delta)}}{\varepsilon M_*}\right).$$

If instead the instance is infeasible, then the same bounds hold with all ε above replaced by 1, and M_* by $|M_*|$.

On tightness. FAST yields a point a_{out} that has near-optimal safety margin in time $\widetilde{O}(d^3/\varepsilon^2 M_*^2)$. Note that since M_* is a priori unknown, this procedure adapts to its value. The dependence of τ on ε and M_* is optimal up to polylog factors. Indeed, for $m = 1$, i.e., only a single unknown constraint, the problem of reliably finding a point with safety margin $(1-\varepsilon)M_*$ is equivalent to finding a point that maximises the unknown objective $\Phi_* x$ to within a εM_* error, and it is known that any method solving this problem for linear bandits requires $\Omega(d^2/\varepsilon^2 M_*^2)$ samples in the worst case (Wagenmaker et al., 2022, Thm. 2). The main loss, then, is a factor of d in the stopping time (which in turn inflates the $\mathbf{S}_\tau$). However, such a loss of a factor of d appears in all known efficient algorithms for linear bandits (Dani et al., 2008; Agrawal & Goyal, 2013). Indeed, for continuous $\mathcal{A}$, we are unaware of any polytime algorithm that can identify near-optimal actions with $o(d^3)$ samples.

Computational Costs. The dominating step for FAST is finding a saddle point of the game (2). By a standard reduction (see §A.5.2, and also Boyd & Vandenberghe (2004, Ch. 5)), this can be computed by solving a linear program with $d+1$ variables and m constraints over the set $\mathcal{A}$. We note that some care is needed: typically such methods are only efficient for approximate computation of the optima. However, our tracking analysis is robust to approximation errors that are $o(t^{-0.5})$, and using this approximatplot(mean(tau(ion level, the computational cost needed at round t scales as $O(\mathsf{LP\text{-}time}\cdot\log(t))$ by, e.g., interior-point methods (Boyd & Vandenberghe, 2004). Of course, for convex $\mathcal{A}$, this $\mathsf{LP\text{-}time}$ is scales polynomially in m, d, and appropriate complexity measures of $\mathcal{A}$. In particular, if $\mathcal{A}$ is a polyhedron with k faces, then this is $\mathrm{poly}(d, m+k)$.

We reiterate that this efficiency holds even though $\mathcal{A}$ is continuous. This is distinct from finite-action bandit exploration approaches, where the computation per round grows as $\Omega(|\mathcal{A}|)$. Discretisation of $\mathcal{A}$ at the scale εM_* would require $\Omega((\varepsilon M_*)^{-d})$ points, making these approaches untenable in our scenario. Of course, M_* is unknown, so such discretisation cannot be directly implemented anyway.

Relationship to EOGT. FAST can be seen as an efficient extension of the feasibility test EOGT (Gangrade et al., 2024b). Indeed, if executed with $\varepsilon = 1$, FAST stops as soon as $U_t < 0$ or $L_t > 0$, which yields a reliable detection of the feasibility of the constraints, capturing the testing guarantees of EOGT. Of course, by coupling EOGT with an appropriate version of our Lemma 7, we can also use this to method to solve the FAS problem. Quantitatively, EOGT would need $\tau = \widetilde{O}(d^2/(\varepsilon M_*)^2)$ samples (and the ℓ_1-relaxation proposed by Gangrade et al. (2024b) to implement EOGT would need $\tau = \widetilde{O}(d^3/(\varepsilon M_*)^2)$), with testing guarantees recovered by setting $\varepsilon = 1$. Thus, FAST recovers the same stopping behaviour, except for the loss of the extra factor of d relative to the unrelaxed-EOGT. As previously discussed, such a loss is associated with all known efficient methods for linear bandits. We note that in simulations (§B.2), FAST with appropriately selected noise (see below and §B) is both computationally and *statisically* faster than EOGT even for small m. In particular, in the scenario of §B.2, we have $d = 6, m = 2, M_* = 0.4$, and the stopping time of FAST is $\sim 27\times$ smaller than that of EOGT, and each round is $\sim 100\times$ faster as well.

Mild Dependence on m, δ. Notice that the dependence of τ on m and δ is quite weak. These terms appear logarithmically, and scale with second order factors in the dimension d. Thus, in the practically relevant regime of $m/\delta = \mathrm{poly}(d)$, the statistical effect of large m or small δ is limited.

Mild Dependence of $\mathbf{S}_T$ on ε. Finally, we highlight that the main term of $\mathbf{S}_T$ depends on ε only logarithmically, and the second order term is a $1/\varepsilon\sqrt{d}$ factor smaller (and so is inactive if $\varepsilon = \Omega(d^{-1/2})$). Further, the safety cost scales only inversely with M_*, rather than with its square (as in τ). This limited safety cost can be understood through the tracking behaviour of FAST: since the optimal margin occurs deep inside $\mathcal{S}_*$, and since the margins of the a_t track the optimal margin, FAST must play an $a_t \in \mathcal{S}_*$ in most rounds.

FAS v/s Optimal Margin. Nominally, the result above captures a stronger output than is strictly needed for the FAS problem, in that a_{out} has margin $\geq (1-\varepsilon)M_*$, i.e., close to the optimal margin, while FAS only requires us to output a point with (arbitrarily small) positive margin. We note that our theory actually captures the latter scenario too, by setting $\varepsilon = 1$. Indeed, in this case, the procedure will stop essentially in time required to test feasibility, and provide a point with arbitrarily small positive margin. This arises due to the multiplicative nature of our notion of approximation—in essence, M_* serves as the natural scale of the search problem, and ε simply tunes between the resilience of the solution and the costs of recovery. Note that the resulting stopping time is still tight.

The Zero Optimal Margin Case. Strictly speaking, Theorem 9 only captures settings with $|M_*| > 0$. For com-

pleteness, we note that for $M_* = 0$, no method can have a nontrivial stoppage guarantee. A simple way to see this is to note that the $\Omega(d^2/\varepsilon^2 M_*^2)$ bound above diverges as $M_* \to 0$.

Practical Noise Distributions. It is well-understood that while existing analyses of TS require inflation of the noise variance to demonstrate optimism, in practice the smaller perturbations of the form $\mathcal{N}(0, I_d/\sqrt{d})$ have sufficient optimism rate, and improve regret by a factor of $\sqrt{d}$ by contracting B (e.g. Abeille & Lazaric, 2017). In the case of FAST, this improvement would be stronger still, since the stopping time scales with $(B/\pi)^2$. We recommend that in practice, this method is executed with such a small noise, but without amending the value of π. Simulations supporting this design are presented in §B. Of course, proving that this algorithm is reliable is an open question.

4. Implications for Low-Regret SLBs

Finally, we turn to the implications of the FAS problem, and specifically FAST, for regret control in safe linear bandits (SLBs). Recall that the regret control problem concerns both the reward $\theta_*^\top a$ and the risks $\Phi_* a$ associated with the program $\max_{a \in \mathcal{A}} \theta_*^\top a : \Phi_* a \geq 0$. Using noisy feedback of both the rewards and risks, the 'hard enforcement' problem for SLBs demands an algorithm for which the regret $\mathbf{R}_T$ is small, while ensuring that the actions are always safe, i.e., with high probability, $\mathbf{S}_T = 0$.

As previously discussed, algorithms for SLBs, such as LC-LUCB (Pacchiano et al., 2024) or SAFE-LTS (Moradipari et al., 2021) address this problem, given an initial action a_{safe}, along with a bound M_{safe} such that $M(a_{\mathsf{safe}}) \geq M_{\mathsf{safe}} > 0$. These methods proceed by constructing a sequence of inner approximation of $\mathcal{S}_*$, $\tilde{\mathcal{S}}_t$, and ensure safety by restricting a_t to lie in $\tilde{\mathcal{S}}_t$ for each t. The data $(a_{\mathsf{safe}}, M_{\mathsf{safe}})$ is used to construct the initial $\tilde{\mathcal{S}}_0$, and M_{safe} affects the rate of expansion of $\tilde{\mathcal{S}}_t$, as reflected in the regret bounds, which scale as $\mathbf{R}_T = \widetilde{O}(\sqrt{d^3 T / M_{\mathsf{safe}}^2})$ for efficient methods. Note that as such this M_{safe} is an arbitrary value, and may not be close to M_*. For instance, in settings such as drug design and trialling, while a 'no-dosage' solution is always safe, it may not have a nontrivial margin (since, typically, active compounds must be coupled with auxiliary compounds to manage their side-effects).

Of course, without knowledge of such an a_{safe}, these methods cannot be executed. Indeed, without such an a_{safe}, it is evidently impossible to achieve guarantees such as $\mathbf{S}_T = 0$, since just the initial action may be unsafe with constant chance. Given the development of FAST above, our proposal is natural: we initialise these prior methods with the output of FAST, run with $\varepsilon = 1/2$. This results in the following guarantee (§A.4).

Corollary 10. *For $\zeta \sim \mathrm{Unif}(\sqrt{3d}\mathbb{S}^d)$, let μ be the law of $\mathbf{1}_m \zeta^\top$. For any feasible instance, the two phase method that runs $\mathrm{FAST}(\mu, 1/2, \delta/5)$ until stoppage, and then executes LC-LUCB(δ) initialised with its output ensures that with probability at least $1 - \delta$, for all T,*

$$\mathbf{R}_T = \widetilde{O}\left(\sqrt{\frac{d^3 T}{M_*^2}}\right), \text{ and } \mathbf{S}_T = \widetilde{O}\left(\frac{d^3}{M_*}\right) = O(1)$$

in the regime $\log(m/\delta) = o(d)$.

Comparison to Hard Enforcement Methods. Notice that the regret bound of the two phase method in Corollary 10 improves upon prior results by a factor of M_*/M_{safe}, which can be attributed to the fact that $\mathrm{FAST}(\mu, 1/2, \delta/5)$ finds a point with margin at least $M_*/2$. In the absence of a priori control on M_{safe}, this factor can be arbitrarily large. In greater detail, given that FAST ignores all reward information, our analysis simply incurs a constant cost for $\tau = \widetilde{O}(d^3/M_*^2)$ rounds. Of course, this additive overhead does not grow with T, and so is hidden in the regret bounds. However, even more is true: notice that a bound of $\sqrt{d^3 T / M_{\mathsf{safe}}^2}$ is smaller than T only if $T > d^3/M_{\mathsf{safe}}^2$. Since $M_* \geq M_{\mathsf{safe}}$, this means that by the time prior results give sublinear regret, the above two phase process already recovers the same performance as prior results to within a polylog factor, without being given an initial safe action.

Limited Safety Costs Without A Priori Knowledge. Notice that the bound on $\mathbf{S}_T$ in Cor. 10 does not grow with T, i.e., that the net excess-risk is only $O(1)$. Since the regret performance of this two-stage procedure matches prior results, this is the first hard enforcement method for SLBs that ensures $\widetilde{O}(\sqrt{d^3 T/M_*^2})$ regret at an $O(1)$ cost, without assumed knowledge of an a_{safe} to begin with, and with dependence only on the optimal safety margin M_*.

No Restart Requirement. The two phase method can be executed without resetting the information accumulated over the course of running FAST, since our argument accounts for conditions under which, e.g., LC-LUCB is low-regret. This 'reuse' of information (i.e., the data in $\mathfrak{H}_\tau$) yields a 'warm start', and thus a practical speed up, for regret control.

5. Conclusion

We have presented FAST, a method that efficiently explores a convex action space to either detect the infeasibility of a set of linear constraints, or to find a safe point with safety margin close to optimal. The design and analysis of FAST yield new insights in the theory of TS when applied to multi-objective problems. The stoppage of FAST adapts optimally to the optimal safety margin, and limited safety costs are incurred, which allows this method to be combined with low-regret algorithms for SLBs to yield the first method that, without priori knowledge of a safe action, achieves strong regret performance at only $O(1)$ safety cost.

Impact Statement

This paper develops an efficient algorithm for discovering actions meeting unknown linear constraints, and analyses its theoretical properties and applications to safe bandits. The potential societal consequences of this work depend largely on how and to what domains practitioners apply such algorithms, which we decline to prognosticate. We do not see any ethical concerns with the contents of this paper.

Acknowledgements

AG was supported by the National Science Foundation award CPS-2317079. VS was supported by the Army Research Office Grant W911NF2110246, AFRL Grant FA8650-22-C1039, and the National Science Foundation awards CPS-2317079, CCF-2007350, and CCF-1955981. CS was supported in part by the Department of Defense, Defense Threat Reduction Agency under award HDTRA1-20-2-0002.

References

Abbasi-Yadkori, Y., Pál, D., and Szepesvári, C. Improved algorithms for linear stochastic bandits. *Advances in neural information processing systems*, 24:2312–2320, 2011. 3, 5, 20

Abeille, M. and Lazaric, A. Linear Thompson sampling revisited. *Electronic Journal of Statistics*, 11(2):5165 – 5197, 2017. doi: 10.1214/17-EJS1341SI. URL https://doi.org/10.1214/17-EJS1341SI. 2, 4, 6, 7, 9, 28

Agrawal, S. and Goyal, N. Thompson sampling for contextual bandits with linear payoffs. In *International conference on machine learning*, pp. 127–135. PMLR, 2013. 4, 8

Amani, S., Alizadeh, M., and Thrampoulidis, C. Linear stochastic bandits under safety constraints. *arXiv preprint arXiv:1908.05814*, 2019. 3

Auer, P., Chiang, C.-K., Ortner, R., and Drugan, M. Pareto front identification from stochastic bandit feedback. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, 2016. 3

Badanidiyuru, A., Kleinberg, R., and Slivkins, A. Bandits with knapsacks. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pp. 207–216. IEEE, 2013. 3

Boyd, S. and Vandenberghe, L. *Convex optimization*. Cambridge university press, 2004. 8

Camilleri, R., Wagenmaker, A., Morgenstern, J., Jain, L., and Jamieson, K. Active learning with safety constraints. *arXiv preprint arXiv:2206.11183*, 2022. 3

Carlsson, E., Basu, D., Johansson, F. D., and Dubhashi, D. Pure exploration in bandits with linear constraints. *arXiv preprint arXiv:2306.12774*, 2023. 3

Carpentier, A., Vernade, C., and Abbasi-Yadkori, Y. The elliptical potential lemma revisited. *arXiv preprint arXiv:2010.10182*, 2020. 5

Castiglioni, M., Celli, A., Marchesi, A., Romano, G., and Gatti, N. A unifying framework for online optimization with long-term constraints. *Advances in Neural Information Processing Systems*, 35:33589–33602, 2022. 3

Dani, V., Hayes, T. P., and Kakade, S. M. Stochastic linear optimization under bandit feedback. In *Conference on Learning Theory*, 2008. 3, 8, 21

Degenne, R., Ménard, P., Shang, X., and Valko, M. Gamification of pure exploration for linear bandits. In *International Conference on Machine Learning*, pp. 2432–2442. PMLR, 2020. 2

Fiez, T., Jain, L., Jamieson, K. G., and Ratliff, L. Sequential experimental design for transductive linear bandits. *Advances in neural information processing systems*, 32, 2019. 2

Gangrade, A., Chen, T., and Saligrama, V. Safe linear bandits over unknown polytopes. In *The Thirty Seventh Annual Conference on Learning Theory*, pp. 1755–1795. PMLR, 2024a. 3, 5

Gangrade, A., Gopalan, A., Saligrama, V., and Scott, C. Testing the feasibility of linear programs with bandit feedback. In Salakhutdinov, R., Kolter, Z., Heller, K., Weller, A., Oliver, N., Scarlett, J., and Berkenkamp, F. (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 14513–14539. PMLR, 21–27 Jul 2024b. 3, 8, 27

Howard, S. R., Ramdas, A., McAuliffe, J., and Sekhon, J. Time-uniform, nonparametric, nonasymptotic confidence sequences. *The Annals of Statistics*, 49(2), 2021. 5

Hutchinson, S., Turan, B., and Alizadeh, M. The impact of the geometric properties of the constraint set in safe optimization with bandit feedback. In *Learning for Dynamics and Control Conference*, pp. 497–508. PMLR, 2023. 3

Jedra, Y. and Proutiere, A. Optimal best-arm identification in linear bandits. *Advances in Neural Information Processing Systems*, 33:10007–10017, 2020. 2

Katz-Samuels, J. and Scott, C. Top feasible arm identification. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 1593–1601. PMLR, 2019. 3

Kaufmann, E., Koolen, W. M., and Garivier, A. Sequential test for the lowest mean: From Thompson to Murphy sampling. *Advances in Neural Information Processing Systems*, 31, 2018. 3

Laurent, B. and Massart, P. Adaptive estimation of a quadratic functional by model selection. *Annals of statistics*, pp. 1302–1338, 2000. 15

Li, Y., Liu, L., Pi, W., Liang, H., and Luo, Z.-Q. Optimistic thompson sampling for no-regret learning in unknown games. *arXiv preprint arXiv:2402.09456*, 2024a. 3

Li, Z., Jamieson, K., and Jain, L. Optimal exploration is no harder than thompson sampling. In *International Conference on Artificial Intelligence and Statistics*, pp. 1684–1692. PMLR, 2024b. 3

Maiti, A., Jamieson, K., and Ratliff, L. Instance-dependent sample complexity bounds for zero-sum matrix games. In *International Conference on Artificial Intelligence and Statistics*, pp. 9429–9469. PMLR, 2023. 3

Michel, T., Hajiabolhassan, H., and Ortner, R. Regret bounds for satisficing in multi-armed bandit problems. *Transactions on Machine Learning Research*, 2022. 3

Moradipari, A., Amani, S., Alizadeh, M., and Thrampoulidis, C. Safe linear thompson sampling with side information. *IEEE Transactions on Signal Processing*, 2021. 3, 9, 22

O'Donoghue, B., Lattimore, T., and Osband, I. Matrix games with bandit feedback. In *Uncertainty in Artificial Intelligence*, pp. 279–289. PMLR, 2021. 3

Pacchiano, A., Ghavamzadeh, M., Bartlett, P., and Jiang, H. Stochastic bandits with linear constraints. In *International Conference on Artificial Intelligence and Statistics*, pp. 2827–2835. PMLR, 2021. 3

Pacchiano, A., Ghavamzadeh, M., and Bartlett, P. Contextual bandits with stage-wise constraints. *arXiv preprint arXiv:2401.08016*, 2024. 3, 9, 21, 22

Russo, D. J., Van Roy, B., Kazerouni, A., Osband, I., Wen, Z., et al. A tutorial on thompson sampling. *Foundations and Trends® in Machine Learning*, 11(1):1–96, 2018. 4

Soare, M., Lazaric, A., and Munos, R. Best-arm identification in linear bandits. *Advances in Neural Information Processing Systems*, 27, 2014. 2

Wagenmaker, A. J., Chen, Y., Simchowitz, M., Du, S., and Jamieson, K. Reward-free rl is no harder than reward-aware rl in linear markov decision processes. In *International Conference on Machine Learning*, pp. 22430–22456. PMLR, 2022. 8

Wang, P.-A., Tzeng, R.-C., and Proutiere, A. Fast pure exploration via frank-wolfe. *Advances in Neural Information Processing Systems*, 34:5810–5821, 2021. 2

A. Analysis of FAST

We prove the results determining the reliability of, and quantitative bounds on, the behaviour of FAST. We will repeatedly exploit the following basic observation, which is a consequence of the Cauchy-Schwarz inequality.

Lemma 11. *Let Φ, Ψ be two matrices in $\mathbb{R}^{m\times d}$, and let V be a positive definite matrix in $\mathbb{R}^{d\times d}$. Then for any $a \in \mathbb{R}^d$,*

$$\|(\Phi - \Psi)a\|_\infty \leq (\max_i \|\Phi^i - \Psi^i)\|_V)\|a\|_{V^{-1}}.$$

Proof. By the Cauchy-Schwarz inequality,

$$|(\Phi^i - \Psi^i)a| = |(\Phi^i - \Psi^i)V^{1/2}V^{-1/2}a| \leq \|(\Phi^i - \Psi^i)V^{1/2}\| \cdot \|V^{-1/2}a\|,$$

where we note that $(\Phi^i - \Psi^i)V^{1/2}$ is a row vector. Of course, $\|V^{-1/2}a\|^2 = a^\top V^{-1/2}V^{-1/2}a = \|a\|^2_{V^{-1}}$, and similarly for $\|(\Phi^i - \Psi^i)V^{1/2}\|$. This bounds the absolute value of the ith coordinate of $(\Phi^i - \Psi^i)a$, and maximising over the same bounds the ℓ_∞ norm. □

A.1. Proof of the Tracking Inequality

The main tool required to analyse FAST is the tracking inequality of Lemma 7. In order to prove the same, we begin by completing the proof of the roundwise tracking upper bound of Lemma 5.

Proof of Lemma 5. Recall the event $\mathsf{Ball}_t = \{\max_i \|H_t^i\| \leq B(\delta_t)\}$. In this case, we know that $\widetilde{\Phi}_t \in \mathcal{E}_t := \{\Phi(H,t) : \max_i \|H^i\| \leq B(\delta_t)\}$. Finally, recall the event $\tilde{\mathsf{G}}_t = \mathsf{G}_t(\delta) \cap \mathsf{Ball}_t$.

Now, observe that the constant M_* satisfies

$$M_* = \mathbb{E}[M_*|\mathfrak{H}_{t-1}, \tilde{\mathsf{G}}_t],$$

where we exploit the fact that Ball_t is determined given G_t. Of course, by definition, under $\mathsf{G}_t(\delta) \subset \tilde{\mathsf{G}}_t$, it holds that $M_* = K(\Phi_*) \leq K(\widetilde{\Phi}_t)$, which further yields the upper bound

$$M_*\mathbb{1}_{\mathsf{Ball}_t} \leq \mathbb{E}[K(\widetilde{\Phi}_t)\mathbb{1}_{\mathsf{Ball}_t}|\mathfrak{H}_{t-1}, \tilde{\mathsf{G}}_t].$$

On the flipside, notice that $\lambda_t^\top \widetilde{\Phi}_t a_t = K(\widetilde{\Phi}_t)$. Now, under Ball_t, it holds that $\widetilde{\Phi}_t \in \mathcal{E}_t$, which implies that

$$K(\widetilde{\Phi}_t)\mathbb{1}_{\mathsf{Ball}_t} \geq K(\bar{\Phi}_t)\mathbb{1}_{\mathsf{Ball}_t},$$

where

$$\bar{\Phi}_t \in \operatorname*{arg\,min}_{\Phi\in\mathcal{E}_t} K(\Phi).$$

But notice that $\mathcal{E}_t$ is entirely determined given the historical data (i.e., does not depend on the noise H_t), and thus so is $K(\bar{\Phi}_t)$. As a consequence, $K(\bar{\Phi}_t)$ is independent of $\tilde{\mathsf{G}}_t$ given $\mathfrak{H}_{t-1}$. We can then conclude that

$$K(\widetilde{\Phi}_t)\mathbb{1}_{\mathsf{Ball}_t} \geq K(\bar{\Phi}_t)\mathbb{1}_{\mathsf{Ball}_t} = \mathbb{E}[K(\bar{\Phi}_t)|\mathfrak{H}_{t-1}, \tilde{\mathsf{G}}_t]\mathbb{1}_{\mathsf{Ball}_t}$$

Putting these two together gets us to the point discussed in the main text, that

$$(M_* - K(\widetilde{\Phi}_t)\mathbb{1}_{\mathsf{Ball}_t} \leq \mathbb{E}[K(\widetilde{\Phi}_t) - K(\bar{\Phi}_t)|\mathfrak{H}_{t-1}, \tilde{\mathsf{G}}_t]\mathbb{1}_{\mathsf{Ball}_t},$$

at which point we can continue that analysis. For completeness, we restate this here.

Recall that $K(\widetilde{\Phi}_t) = \lambda_t^\top \widetilde{\Phi}_t a_t$. Since $K(\bar{\Phi}_t)$ is similarly the value of the matrix game over $\mathcal{A}$ induced by the payoffs $\bar{\Phi}_t$, there exists a saddle point $(\bar{\lambda}_t, \bar{a}_t)$ of $\bar{\Phi}_t$ such that $K(\bar{\Phi}_t) = \bar{\lambda}_t^\top \bar{\Phi}_t a_t$.

Note that since λ is selected by the '$\min$'-player, deviating from λ_t while keeping a_t fixed increases the value of $\lambda^\top \widetilde{\Phi}_t a_t$ beyond $K(\widetilde{\Phi}_t)$. Using this for $\lambda = \bar{\lambda}_t$, we find that

$$K(\widetilde{\Phi}_t) = \lambda_t^\top \widetilde{\Phi}_t a_t \leq \bar{\lambda}_t^\top \widetilde{\Phi}_t a_t.$$

By the same coin, deviating from $\bar{a}_t$ while leaving $\bar{\lambda}_t$ fixed decreases the value of the payoff under $\bar{\Phi}_t$, i.e.,

$$\bar{\lambda}_t^\top \bar{\Phi}_t \bar{a}_t \geq \bar{\lambda}_t^\top \bar{\Phi}_t a_t.$$

As a result,

$$K(\widetilde{\Phi}_t) - K(\bar{\Phi}_t) \leq \bar{\lambda}_t^\top (\widetilde{\Phi}_t - \bar{\Phi}_t) a_t.$$

Now, under $\tilde{\mathsf{G}}_t \subset \mathsf{Ball}_t(\delta)$, we further know that $\max_i \|\tilde{\Phi}_t^i - \bar{\Phi}_t^i\|_{V_t} \leq B(\delta_t)\omega_t$, since each lies in $\mathcal{E}_t$. Thus, applying the Cauchy-Schwarz relation of Lemma 11, and using the fact that $\|\bar{\lambda}\|_t = 1$ since it lies in the simplex Δ^m, we conclude that

$$\mathbb{E}[K(\widetilde{\Phi}_t) - K(\bar{\Phi}_t) | \mathfrak{H}_{t-1}, \tilde{\mathsf{G}}_t] \mathbb{1}_{\mathsf{Ball}_t} \leq \mathbb{E}[2B(\delta_t)\omega_t(\delta) \|a_t\|_{V_t^{-1}} \cdot \bar{\lambda}_t^\top | \mathfrak{H}_{t-1}, \tilde{\mathsf{G}}_t] \mathbb{1}_{\mathsf{Ball}_t}.$$

Since $B(\delta_t)$ is deterministic, and $\omega_t(\delta)$ is $\mathfrak{H}_{t-1}$-measurable, this lets us conclude that

$$(M_* - K(\widetilde{\Phi}_t) \mathbb{1}_{\mathsf{Ball}_t} \leq 2B(\delta_t)\omega_t(\delta) \mathbb{E}[\|a_t\|_{V_t^{-1}} | \mathfrak{H}_{t-1}, \tilde{\mathsf{G}}_t] \mathbb{1}_{\mathsf{Ball}_t}.$$

The argument is concluded by observing that for any nonnegative random variable X and event E, $\mathbb{E}[X|\mathsf{E}]\mathbb{P}(\mathsf{E}) \leq \mathbb{E}[X]$, and using the fact that under the π-optimism condition, $\mu(\tilde{\mathsf{G}}_t) \geq \mu(\mathsf{G}_t(\delta)) - \mu(\mathsf{Ball}_t(\delta)) \geq \pi - \pi/2$. □

Proving the Main Tracking Inequality. We recall the roundwise lower bound of Lemma 6, which will also be exploited in the proof of the main tracking inequality.

Proof of Lemma 7. The proof consists of exploiting the roundwise results of Lemmas 5 and 5 to derive basic forms of upper and lower bounds on $\mathscr{T}_t$. These bounds are controlled by a concentration argument to conclude.

Let $\mathsf{Ball}(\delta) := \bigcap_{t \geq 1} \mathsf{Ball}_t(\delta)$, and recall the event $\mathsf{Con}(\delta) := \bigcap_{t \geq 1} \mathsf{Con}_t(\delta)$. Note that the probability of $\mathsf{Ball}(\delta)$ at least $1 - \sum \delta_t = 1 - \delta$.

Lower bound on $\mathscr{T}_t$. Observe that

$$\lambda_s^\top \Phi_* a_s = \lambda_s^\top \widetilde{\Phi}_s a_s + \lambda_s^\top (\widetilde{\Phi}_s - \Phi_*) a_s.$$

Under $\mathsf{Con}(\delta) \cap \mathsf{Ball}(\delta)$, it holds for all s that $\max_i \|\widetilde{\Phi}_s^i - \hat{\Phi}_s^i\|_{V_s} \leq B(\delta_s)\omega_t(\delta)$, and $\max_i \|\Phi_*^i - \hat{\Phi}_s^i\|_{V_s} \leq \omega_s(\delta)$. By the Cauchy-Schwarz relation Lemma 11, and the fact that $\|\lambda_s\|_1 = 1$ since it lies in the simplex Δ^m, we have under $\mathsf{Ball}(\delta) \cap \mathsf{Con}(\delta)$ that

$$\forall s, \lambda_s^\top \Phi_* a_s \geq \lambda_s^\top \widetilde{\Phi}_s a_s - (1 + B(\delta_s))\omega_s(\delta) \|a_s\|_{V_s^{-1}}.$$

Now, using the upper bound of Lemma 5, we further conclude that under $\mathsf{Ball}(\delta) \cap \mathsf{Con}(\delta)$,

$$\lambda_s^\top \Phi_* a_s \geq M_* - (1 + B(\delta_s))\omega_s(\delta) \|a_s\|_{V_s^{-1}} - \frac{4B(\delta_s)\omega_s(\delta)}{\pi} \mathbb{E}[\|a_s\|_{V_s^{-1}} | \mathfrak{H}_{s-1}].$$

Now, let $\zeta_s := \lambda_s^\top w_s^S$ be the projection of the noise in S_s onto λ_s. Notice that since λ_s is measurable given $\mathfrak{G}_s$ (since it can be determinitically determined from the history and H_s), the process $\{\zeta_s\}$ is a centred 1-subGaussian process with respect to the filtration $\mathfrak{G}_s$. We let $Z_t := \sum_{s \leq t} \zeta_s$.

Now, finally, we have

$$\begin{aligned}
\mathscr{T}_t &= \sum_{s \leq t} \lambda_s^\top S_s = \sum_{s \leq t} \lambda_s^\top \Phi_* a_s + \sum_{s \leq t} \zeta_s \\
&\geq Z_t + tM_* - \sum_{s \leq t} (1 + B(\delta_s))\omega_s(\delta) \|a_s\|_{V_s^{-1}} - \sum_{s \leq t} \frac{4B(\delta_s)\omega_s(\delta)}{\pi} \mathbb{E}[\|a_s\|_{V_s^{-1}} | \mathfrak{H}_{s-1}] \\
\implies \mathscr{T}_t &\geq tM_* + Z_t - (1 + B(\delta_t))\omega_t(\delta) \sum_{s \leq t} \|a_s\|_{V_s^{-1}} - \frac{4B(\delta_t)\omega_t(\delta)}{\pi} \sum_{s \leq t} \mathbb{E}[\|a_s\|_{V_s^{-1}} | \mathfrak{H}_{s-1}], \qquad (3)
\end{aligned}$$

where the final line uses the fact that $B(\delta_s)$ and $\omega_s(\delta)$ are nondecreasing in s, and that $\|a_s\|_{V_s^{-1}}$ is a nonnegative quantity.

Upper bound on $\mathscr{T}_t$ Taking a similar approach, we have

$$\mathscr{T}_t = Z_t + \sum_{s\le t} \lambda_s^\top \Phi_* a_s,$$

and applying Lemma 6, under $\mathsf{Con}(\delta) \cap \mathsf{Ball}(\delta)$, we have

$$\mathscr{T}_t \le Z_t + \sum_{s\le t} M(a_s) + 2(1+B(\delta_t))\omega_t(\delta) \sum_{s\le t} \|a_s\|_{V_s^{-1}}. \tag{4}$$

Concentration. There are two quantities in the bounds above that need to be controlled: Z_t, and $\sum_{s\le t} \mathbb{E}[\|a_s\|_{V_s^{-1}} | \mathfrak{H}_{s-1}]$.

Now, notice that since the only randomness in a_s given $\mathfrak{H}_{s-1}$ arises from H_s, which is independent of $\mathfrak{H}_{s-1}$, the fluctuations $\|a_s\|_{V_s^{-1}} - \mathbb{E}[\|a_s\|_{V_s^{-1}} | \mathfrak{H}_{s-1}]$ are conditionally centred. Further, since $\|a_s\|_{V_s^{-1}} \in [0,1]$, which uses the fact that $V_s \succeq I$ and $\mathcal{A} \subset \{\|a\| \le 1\}$, we conclude that these fluctuations are also 1-subGaussian. It follows thus that the process

$$\beta_t := \sum \|a_s\|_{V_s^{-1}} - \mathbb{E}[\|a_s\|_{V_s^{-1}} | \mathfrak{H}_{t-1}]$$

has centered 1-subGaussian increments with respect to $\{\mathfrak{H}_{s-1}\}$. Of course, we have already argued that Z_t is a martingale with centred and 1-subGaussian increments with respect to the filtration $\{\mathfrak{G}_s\}$.

We encapsulate the concentration of these two processes into one event

$$\mathsf{Walk}(\delta) := \{\forall t, |\beta_t| \le \mathrm{LIL}(t,\delta), |Z_t| \le \mathrm{LIL}(t,\delta)\}.$$

Tracking Inequality. But then, applying (4), under $\mathsf{Con}(\delta) \cap \mathsf{Ball}(\delta) \cap \mathsf{Walk}(\delta)$, we have

$$\mathscr{T}_t \ge tM_* - \left(1 + B(\delta_t)) + \frac{4B(\delta_t)}{\pi}\right) \omega_t(\delta) \sum_{s\le t} \|a_s\|_{V_s^{-1}} - \left(1 + \frac{4B(\delta_t)}{\pi}\right) \mathrm{LIL}(t,\delta),$$

and applying (3),

$$\mathscr{T}_t \le \sum_{s\le t} M(a_s) + 2(1+B(\delta_t))\omega_t(\delta) \sum_{s\le t} \|a_s\|_{V_s^{-1}} + \mathrm{LIL}(t,\delta).$$

Now, recall that under $\mathsf{Con}(\delta) \cap \mathsf{Ball}(\delta) \cap \mathsf{Walk}(\delta)$,

$$\mathscr{B}_t(\mu,\delta) = \frac{(1+5B(\delta_t))\omega_t(\delta)}{\min(\pi, 1/2)} \left(\sum_{s\le t} \|a_s\|_{V_s^{-1}} + \mathrm{LIL}(t,\delta)\right).$$

But then it is evident that

$$\mathscr{B}_t(\mu,\delta) \ge 2(1+B(\delta_t))\omega_t(\delta) \sum_{s\le t} \|a_s\|_{V_s^{-1}} + \mathrm{LIL}(t,\delta),$$

and

$$\mathscr{B}_t(\mu,\delta) \ge \left(1 + 5\frac{B(\delta_t)}{\pi}\right) \omega_t(\delta) \sum_{s\le t} \|a_s\|_{V_s^{-1}} + \left(1 + \frac{4B(\delta_t)}{\pi}\right) \omega_t(\delta) \mathrm{LIL}(t,\delta).$$

This lets us conclude that under $\mathsf{Con}(\delta) \cap \mathsf{Ball}(\delta) \cap \mathsf{Walk}(\delta)$, it holds simultaneously for all t that

$$\mathscr{T}_t + \mathscr{B}_t(\mu,\delta) \ge tM_* \text{ and } \mathscr{T}_t - \mathscr{B}_t(\mu,\delta) \le \sum_{s\le t} M(a_s),$$

and the tracking inequality follows from the trivial observation that $\sum_{s\le t} M(a_s) \le tM_*$.

Bookkeeping the probability of violation. Finally, we note from Lemma 1 that the chance of $\mathsf{Con}(\delta)$ is at least $1-\delta$. Via the lil, the chance of $\mathsf{Walk}(\delta)$ is at least $1-2\delta$, using a union bound for the two processes β_t and Z_t. Finally, by a union bound, the chance of $\mathsf{Ball}(\delta)$ is at least $1 - \sum \delta_t \ge 1 - \delta \sum 1/t(t+1) = 1-\delta$. By a union bound, then, these events occur together with chance at least $1-4\delta$. □

A.2. Analysis of the Coupled Noise Design

We move to the crucial argument that our coupled noise design leads to a good FCO condition, which relies on analysing the behaviour of a_* under the perturbations.

Proof of Theorem 8. Notice that the noise distribution μ couples the rows of H such that $H = \mathbf{1}_m \zeta^\top$, where $\zeta \sim \nu$. We will suppress the dependence of $\omega_t, \mathsf{Con}_t(\delta)$ on δ in the following.

Let a_* be *any* maximin point of the matrix game, and fix any t. As stated in §3.1, we will analyse the local optimism event

$$\mathsf{L}_t := \{\widetilde{\Phi}_t : \min_\lambda \lambda^\top \widetilde{\Phi}_t a_* \geq M_* = \lambda_*^\top \Phi_* a_*\}.$$

To this end, recall that

$$\widetilde{\Phi}_t = \hat{\Phi}_t + \omega_t H_t V_t^{-1/2}.$$

Thus, local optimism holds whenever H_t is such that

$$\hat{\Phi}_t a_* + \omega_t H_t V_t^{-1/2} a_* \geq \Phi_* a_*.$$

Now, under $\mathsf{Con}(\delta)$, we know that $\max_i \|\hat{\Phi}_t^i - \Phi_*^i\|_{V_t} \leq \omega_t$. Thus, by the Cauchy-Schwarz relation of Lemma 11, we know that

$$\|(\hat{\Phi}_t - \Phi_*)a_*\|_\infty \leq \omega_t \|a_*\|_{V_t^{-1}} = \omega_t \|V_t^{-1/2} a_*\| \implies \hat{\Phi}_t a_* \geq \Phi_* a_* - \omega_t \|V_t^{-1/2} a_*\| \mathbf{1}_m.$$

But then notice that

$$\hat{\Phi}_t a_* + \omega_t H_t V_t^{-1/2} a_* \geq \Phi_* a_* + \omega_t \left(H_t V_t^{-1/2} a_* - \|V_t^{-1/2} a_*\| \mathbf{1}_m \right).$$

Now, we use the fact that $H_t = \mathbf{1}_m \zeta^\top$ for some $\zeta \sim \nu$ drawn independently of V_t. In this case, we find that

$$H_t V_t^{-1/2} a_* - \|V_t^{-1/2} a_*\| \mathbf{1}_m = (\zeta^\top u_t - \|u_t\|)\mathbf{1}_m,$$

where $u_t = V_t^{-1/2} a_*$.

It immediately follows that $\{\zeta^\top u_t \geq \|u_t\|\} \subset \mathsf{L}_t(\delta)$ under $\mathsf{Con}_t(\delta)$. But since $\zeta \sim \nu$ independently of the history, the chance of this given $\mathfrak{H}_{t-1}$ is at least π. Of course, we can further see that each row of H_t is just a copy of ζ_t, and so automatically inherits the concentration property of ν. □

Let us note that in practice, we expect the optimism rate to be much larger: we only showed that optimism occurs at rate at least π *at* any maximin point a_*. But the global optimism event G_t accounts for any draw where the value of the game induced by $\widetilde{\Phi}_t$ (for which the margin at a_* is only a lower bound). In practice, it is known that in single objective TS, the optimism rate is large even with laws that are poorly anticoncentrated in the sense of ν (e.g., $\mathcal{N}(0, I_d/\sqrt{d})$, which typically only ensures that $\zeta^\top u > \|u\|/\sqrt{d}$.

A.2.1. Bounds for Simple Reference Distributions

We argue that both the standard Gaussian, and the uniform law of the sphere of radius $\sqrt{3d}$ yield effective noise distributions for our coupled design.

For the Gaussian, recall that if $Z \sim \mathcal{N}(0, I_d)$, then $\|Z\|^2$ is distributed as a χ^2-random variable. A classical subexponential concentration argument (e.g. Laurent & Massart, 2000, Lemma 1) yields that for any x,

$$\mathbb{P}(\|Z\|^2 \geq d + 2\sqrt{dx} + 2x) \leq e^{-x}.$$

Note that $(d + 2\sqrt{dx} + 2x) \leq (\sqrt{d} + \sqrt{2x})^2$, and hence taking $x = \log(1/\delta)$ in the above yields that $B(\delta) \leq \sqrt{d} + \sqrt{2\log(1/\delta)}$. Further, due to the isotropicity of Z, $Z^\top u/\|u\| \overset{\text{law}}{=} Z_1 \sim \mathcal{N}(0,1)$, and thus $\pi \geq 1 - \Phi(1) \geq 0.158\ldots$.

Further, notice that if $Z \sim \mathcal{N}(0, I_d)$, then $Y := \sqrt{3d}Z/\|Z\| \sim \mathrm{Unif}(\sqrt{3d} \cdot \mathbb{S}^d)$, and by isotropicity, for any $u : \|u\| = 1, Y^\top u/\|u\| \overset{\text{law}}{=} Y_1$. As a result,

$$\mathbb{P}(Y^\top u/\|u\| \geq 1) = \mathbb{P}(Y_1 \geq 1) = \frac{1}{2}\mathbb{P}(Y_1^2 \geq 1) = \frac{1}{2}\mathbb{P}((3d-1)Z_1^2 \geq \sum_{i=2}^d Z_i^2) \geq \frac{1}{2}\mathbb{P}(Z_1^2 \geq 1) \cdot \mathbb{P}(\sum_{i=2}^d Z_i^2 \geq 3d-1).$$

But notice that $d - 1 + 2\sqrt{(d-1) \cdot d/3} + 2d/3 \leq 3d - 1$, and thus, $\mathbb{P}(\sum_{i=2}^d Z_i^2 \geq 3d-1) \leq \exp(-d/3)$. Invoking the bound on $\mathbb{P}(Z_1 \geq 1) = \frac{1}{2}\mathbb{P}(|Z_1| \geq 1)$ above, we conclude that $\pi \geq 0.15 \cdot (1 - e^{-d/3})$. Of course, $\|Y\| = \sqrt{3d}$ surely, giving the B expression.

We note that while the above analysis only shows a $0.15(1 - e^{-d/3})$ lower bound for $\mathrm{Unif}(\sqrt{3d}\mathbb{S}^d)$, via direct simulation this can easily be seen to exceed 0.28, no matter the d. In this case, the attained ratio B/π behaves roughly as $6\sqrt{d}$.

A.3. Analysis of Stopping Time and Excess-Risk of Exploration

The tracking inequality, along with the development in §3.2 enables the analysis of the stopping time and excess-risk of exploration, which we provide below.

The analysis breaks into three pieces: (i) reliability analysis; (ii) control of the stopping time, and (iii) control of the safety costs. We first separately discuss these aspects through a sequence of Lemmas, throughout working under the event that the tracking inequality holds (and perhaps other auxiliary events), and then prove of Theorem 9 by accounting for the various failure rates of the events. Throughout, we will repeatedly use the notation $\mathsf{Con}(\delta), \mathsf{Ball}(\delta), \mathsf{Walk}(\delta), Z_t$ and β_t as defined in the proof of Lemma 7.

Reliability. We demonstrate reliability under the tracking inequality.

Lemma 12. *If the tracking inequality of Lemma 7 holds, then the procedure* FAST$(\mu, \varepsilon, \delta)$ *is reliable, i.e.,*

$$\begin{aligned} M_* \leq 0 &\implies U_\tau < 0, \text{ and} \\ M_* \geq 0 &\implies \min_\lambda \lambda^\top \Phi_* \langle a \rangle_\tau \geq L_\tau > (1-\varepsilon)M_* \end{aligned}$$

Proof. Assume the tracking inequality. Then for all t,

$$U_t \geq M_* \geq L_t,$$

where $U_t = (\mathscr{T}_t + \mathscr{B}_t(\mu, \delta))/t$ and $L_t = (\mathscr{T}_t - \mathscr{B}_t(\mu, \delta))/t$. Further recall that the stopping time of FAST is

$$\tau = \inf\{t : U_t < 0 \text{ or } L_t > (1-\varepsilon)U_t > 0.$$

But then, if $M_* \leq 0, \implies L_t < 0$ for all t, and so if $U_t > 0$ then $L_t \not\geq (1-\varepsilon)U_t$, and thus upon stopping it must hold that $U_\tau < 0$.

Further, if $M_* > 0$, then it always holds that $U_t > 0$, and so only the second stopping criterion, that $L_\tau > (1-\varepsilon)U_\tau$, can be met, which in turn implies that $L_\tau > (1-\varepsilon)M_*$. Further, as discussed in §3.2, $(\langle a \rangle_\tau, L_\tau)$ form a valid output since

$$\min_\lambda M(\langle a \rangle_\tau) \geq \sum_{s \leq \tau} M(a_s)/\tau \geq L_\tau > (1-\varepsilon)M_*. \qquad \square$$

Stopping Time Analysis. Again, using the tracking inequality, we control the behaviour of τ.

Lemma 13. (Stopping Time Analysis) *Suppose that the tracking inequality holds. Then,*

$$\tau = \widetilde{O}\left(\frac{d^3 + d^2\log(m/\delta)}{\varepsilon^2 M_*^2}\right) \text{ if the instance is feasible, and } \tau = \widetilde{O}\left(\frac{d^3 + d^2\log(m/\delta)}{M_*^2}\right) \text{ otherwise.}$$

Proof. We begin by bounding the boundary $\mathscr{B}_t(\mu, \delta)$. Instantiating B and π with the bounds from §A.2.1 above, we note that $B/\pi = O(\sqrt{d})$ for our choices of laws. Using Lemma 2 to control $\sum \|a_s\|_{V_s^{-1}}$ and ω_t, we conclude that there is a

constant C such that for $t \geq \max(d, 3)$,

$$\begin{aligned}\mathscr{B}_t(\mu, \delta) &\leq C\sqrt{d}\sqrt{t \log\log(t/\delta)} + C\sqrt{d} \cdot \sqrt{d\log(t) + \log(m/\delta)} \cdot \sqrt{dt \log t} \\ &\leq 3C\sqrt{t} \cdot \left(d^{3/2} \log t + d\sqrt{\log(t)\log(m/\delta)}\right).\end{aligned}$$

We now turn to studying the stopping time. Firstly, notice that if the tracking inequality is always true, then

$$U_t = (\mathscr{T}_t + \mathscr{B}_t(\mu, \delta))/t \leq M_* + 2\mathscr{B}_t(\mu, \delta)/t,$$

and

$$L_t = (\mathscr{T}_t - \mathscr{B}_t(\mu, \delta))/t \geq M_* - 2\mathscr{B}_t(\mu, \delta)/t.$$

Now, in the infeasible case, i.e., when $M_* \leq 0$,

$$U_t \leq tM_* + 2\mathscr{B}_t(\mu, \delta) \leq -t|M_*| + 6C\sqrt{t} \cdot \left(d^{3/2} \log t + d\sqrt{\log(t)\log(m/\delta)}\right),$$

meaning that in this case, with probability at least $1 - 4\delta$,

$$\begin{aligned}\tau &\leq \inf\{t : tM_* + 2\mathscr{B}_t(\mu, \delta) < 0\} \\ &= \inf\{t : t|M_*| > 2\mathscr{B}_t(\mu, \delta)\} \\ &\leq \inf\{t \geq \max(d, 3) : t|M_*| > C^{'}\sqrt{t} \cdot \left(d^{3/2} \log t + d\sqrt{\log(t)\log(m/\delta)}\right)\},\end{aligned}$$

with $C' = 6C$. This infimum can be upper bounded using the following result:

Lemma 14. *For $A \geq e$, $\inf\{z \geq e : z \geq A \log z\} \leq 2A \log(2A)$.*

Proof. Notice that the map $f(z) := z/\log z$ is nondecreasing for $z \geq e$, since $f'(z) = \frac{\log z - 1}{\log^2 z} \geq 0$. Now,

$$f(2A\log(2A)) = \frac{2A\log(2A)}{\log(2A) + \log\log(2A)} \geq \frac{2A\log(2A)}{2\log(2A)} \geq A,$$

where we have used that $\log\log(z) \leq \log(z)$ for all $z \geq e$. □

To use the above, notice that $\tau \leq \max(T_1, T_2)$ where

$$\begin{aligned}T_1 &:= \inf\{t \geq \max(d, e^2) : \sqrt{t}/\log(\sqrt{t}) > 4C^{'}d^{3/2}/|M_*|\}, \\ T_2 &:= \inf\{t \geq \max(d, e^2) : \sqrt{t/\log t} \geq 2C^{'}d\sqrt{\log(m/\delta)}/|M_*|\}.\end{aligned}$$

By Lemma 14, if $\sqrt{t} > \frac{4C'd^{3/2}}{|M_*|}\log(d^3/M_*^2)$ and $\sqrt{t} > e$, then $t > T_1$, and similarly, if $t > \frac{4C'^2d^2\log(m/\delta)}{M_*^2} \cdot \log(8C'^2d^2\log(m/\delta)/|M_*|^2)$, then $t > T_2$, yielding the bound that

$$\tau \leq \max\left(d + e^2, \frac{16C'^2d^3}{M_*^2}\log^2\frac{d^3}{M_*^2}, \frac{4C'^2d^2\log(m/\delta)}{M_*^2}\log\frac{4C'^2d^2\log(m/\delta)}{M_*^2}\right).$$

Turning now to the feasible case, where $M_* > 0$, we instead notice that if the tracking inequality holds, $U_t \geq M_* > 0$ always, and thus

$$\begin{aligned}\tau &\leq \inf\{t : M_* - 2\mathscr{B}_t(\mu, \delta)/t \geq (1 - \varepsilon)M_* + (2 - 2\varepsilon)\mathscr{B}_t(\mu, \delta)/t\} \\ &= \inf\{t : \varepsilon t|M_*| \geq 4\mathscr{B}_t(\mu, \delta)\},\end{aligned}$$

where we used $|M_*| = M_* > 0$ in the second line. But this is the same bound as the infeasible case, except with $|M_*|$ replaced by $\varepsilon|M_*|/2$. Thus, we can immediately invoke the analysis above to conclude that in the feasible case,

$$\tau \leq \max\left(d + e^2, \frac{64C'^2d^3}{\varepsilon^2M_*^2}\log^2\frac{4d^3}{\varepsilon^2M_*^2}, \frac{16C'^2d^2\log(m/\delta)}{\varepsilon^2M_*^2}\log\frac{16C'^2d^2\log(m/\delta)}{\varepsilon^2M_*^2}\right).$$ □

Safety Analysis. Finally, we analyse the safety of the actions selected by FAST using both the tracking inequality, the behaviour of τ, as well as a concentration of a sum of $\|a_s\|_{V_s^{-1}}$ on the subset of events where these have large expectation. We leave the analysis of the event Walk' below to the final argument proving Theorem 9.

Lemma 15. *Recall the events* $\mathsf{Con}(\delta), \mathsf{Ball}(\delta)$, *and* $\mathsf{Walk}(\delta)$ *from the proof of Lemma 7. Further, define the process*

$$X_t := \sum_{s \le t} (\mathbb{E}[\|a_s\|_{V_s^{-1}} | \mathfrak{H}_{s-1}] - \|a_s\|_{V_s^{-1}}) \mathbb{1}\{\mathbb{E}[\|a_s\|_{V_s^{-1}} | \mathfrak{H}_{s-1}] > M_*/2\kappa_s\},$$

and the event

$$\mathsf{Walk}'(\delta) := \{\forall t, |X_t| \le \mathrm{LIL}(t, \delta)\}.$$

Under $\mathsf{Con}(\delta) \cap \mathsf{Ball}(\delta) \cap \mathsf{Walk}(\delta) \cap \mathsf{Walk}'(\delta)$, *it holds that*

$$\mathbf{S}_\tau = \widetilde{O}\left(\frac{d^3}{M_*} + \frac{d^{5/2}\sqrt{\log(m/\delta)}}{\varepsilon M_*}\right) \text{ if the instance is feasible, and } \mathbf{S}_\tau = \widetilde{O}\left(\frac{d^3}{|M_*|}\right) \text{ otherwise,}$$

where we assume that $\log(m/\delta) = o(d)$.

Proof. As show in the proof of Lemma 7, the tracking inequality holds under the assumed events. We will exploit further detailed bounds shown in that proof. Throughout, we will suppress the dependence of ω_t on δ until required. Similarly, we will just write B_t instead of $B(\delta_t)$.

Recall that the safety cost at any time t is

$$\mathbf{S}_t = \sum_{s \le t} \left(-\min_{\lambda \in \Delta^m} \lambda^\top \Phi_* a_s\right)_+ = \sum_{s \le t} (-M(a_s))_+.$$

Under the roundwise tracking inequality, which holds under $\mathsf{Con}(\delta) \cap \mathsf{Ball}(\delta)$, we know that

$$\begin{aligned} & M(a_s) \ge \lambda_s^\top \Phi_* a_s - 2(1 + B_t)\omega_t \|a_s\|_{V_s^{-1}} \\ \implies & -M(a_s) \le -\lambda_s^\top \Phi_* a_s + 2(1 + B)\omega_t \|a_s\|_{V_s^{-1}} \end{aligned}$$

Of course, under $\mathsf{Con}(\delta) \cap \mathsf{Ball}(\delta)$, we further know by an application of the Caucy-Schwarz relation that

$$\lambda_s^\top \Phi_* a_s \ge \lambda_s^\top \widetilde{\Phi}_s a_s - (1 + B_t)\omega_t \|a_s\|_{V_s^{-1}}.$$

Further under the roundwise tracking upper bound, we have

$$-\lambda_s \widetilde{\Phi}_s a_s \le M_* + \frac{4B_s\omega_s}{\pi} \sum \mathbb{E}[\|a_s\|_{V_s^{-1}} | \mathfrak{H}_{s-1}].$$

Putting these together, we conclude that

$$-M(a_s) \le 3(1 + B_s)\omega_s \|a_s\|_{V_s^{-1}} + \frac{4B_s\omega_s}{\pi} \mathbb{E}[\|a_s\|_{V_s^{-1}} | \mathfrak{H}_{s-1}] - M_*$$

For convenience we define

$$\kappa_s(\delta) = (3(1 + B_s(\delta_s)) + 4B_s(\delta_s)/\pi)\omega_s(\delta) = O(\sqrt{d^2 \log(t) + d \log(m/\delta)}),$$

where the O only suppresses a universal constant factor, and the bound arises from our choice of $\mu = \mathbf{1}_m \zeta^\top$ for $\zeta \sim \mathrm{Unif}(\sqrt{3d}\mathbb{S}^d)$. We further recall the notation $\beta_t = \sum_{s \le t} \|a_s\|_{V_s^{-1}} - \mathbb{E}[\|a_s\|_{V_s^{-1}} | \mathfrak{H}_{s-1}]$.

Since both $\|a_s\|_{V_s^{-1}}$ and $\mathbb{E}[\|a_s\|_{V_s^{-1}} | hist_{s-1}]$ are nonnegative, with this notation, we can write

$$-M(a_s) \le \kappa_s \left(\|a_s\|_{V_s^{-1}} + \mathbb{E}[\|a_s\|_{V_s^{-1}} | \mathfrak{H}_{s-1}]\right) - M_* \tag{5}$$

Infeasible case. In this case, notice that $-M_* = |M_*| \geq 0$. Since the remaining terms are nonnegative, we end with the bound that under $\mathsf{Con}(\delta) \cap \mathsf{Ball}(\delta)$,

$$\mathbf{S}_t \leq t|M_*| + \kappa_t \sum_{s \leq t} \|a_s\|_{V_s^{-1}} + \kappa_t |\beta_t|$$

where we have used recall that κ_s is nondecreasing as s increases. But, under $\mathsf{Walk}(\delta)$, we further know that $|\beta_t| \leq \mathrm{LIL}(t, \delta))$, which tells us that for all t,

$$\mathbf{S}_t \leq t|M_*| + \kappa_s(\delta)(\sqrt{2dt \log(1 + t/d)} + \mathrm{LIL}(t, \delta)).$$

Recall further that in the infeasible case, under $\mathsf{Con}(\delta) \cap \mathsf{Ball}(\delta) \cap \mathsf{Walk}(\delta)$, it holds that $\tau = \widetilde{O}(M_*^{-2}(d^3 + d^2 \log(m/\delta)))$, as discussed in Lemma 13. Using this, we find that

$$\sqrt{\tau} = \widetilde{O}\left(\frac{d^{3/2} + d\sqrt{\log(m/\delta)}}{|M_*|}\right).$$

Upon accounting for the fact that $\mathrm{LIL}(t, \delta) = O(\sqrt{t \log\log t + t \log(1/\delta)})$ and the bound on κ_t, this induces (by a simple but tedious accounting) that

$$\mathbf{S}_\tau = \widetilde{O}\left(\frac{d^3 + d^{5/2}\sqrt{\log(m/\delta)} + d^2 \log(m/\delta) + d^{3/2} \log(m/\delta)\sqrt{\log(1/\delta)}}{|M_*|}\right).$$

Naturally, in the regime $\log(m/\delta) = o(d)$, the leading term is $d^3/|M_*|$.

Feasible case. Note that in the feasible case, we could always get an upper bound by dropping the $-M_* < 0$ term in (5) and repeating the same analysis as the above. However, this would give an extra $1/\varepsilon$ factor blowup in the main term of the safety costs due to the increase in the bound on τ. To show the stated bound thus requires us to exploit this $-M_*$ term as a resource. To this end, we note using the subadditivity of $(\cdot)_+$ that for any u, v, w, it holds that

$$(u + v - w)_+ \leq (u - w/2)_+ + (v - w/2)_+ = (u - w/2)\mathbb{1}\{u > w/2\} + (v - w/2)\mathbb{1}\{v > w/2\}.$$

Using this with $u = \kappa_s \|a_s\|_{V_s^{-1}}, v = \kappa_s \mathbb{E}[\|a_s\|_{V_s^{-1}} | \mathfrak{H}_{t-1}], w = M_*$, we find upon summing that

$$\begin{aligned}
\forall t, \mathbf{S}_t \leq & \kappa_t \sum_{s \leq t} (\|a_s\|_{V_s^{-1}} - M_*/2\kappa_s)\mathbb{1}\{\|a_s\|_{V_s^{-1}} > M_*/2\kappa_s\} \\
& + \kappa_t \sum_{s \leq t} (\mathbb{E}[\|a_s\|_{V_s^{-1}} | \mathfrak{H}_{s-1}] - M_*/2\kappa_s)\mathbb{1}\{\mathbb{E}[\|a_s\|_{V_s^{-1}} | \mathfrak{H}_{s-1}] > M_*/2\kappa_s\},
\end{aligned}$$

where we again used that $\kappa_t \geq \kappa_s > 0$ for all $s \leq t$.

We will first reduce the analysis of the second sum above to the first using the event Walk' in the statement of the lemma. For the sake of succinctness, define $\ell_s = M_*/2\kappa_s, n_s = \|a_s\|_{V_s^{-1}} - \ell_s, m_s = \mathbb{E}[\|a_s\|_{V_s^{-1}} | \mathfrak{H}_{s-1}] - \ell_s$. Notice that $\|a_s\|_{V_s^{-1}} > \ell_s \iff n_s > 0$ and similarly for m_s.

Now,

$$\begin{aligned}
\sum_{s \leq t} m_s \mathbb{1}\{m_s > 0\} &= \sum_{s \leq t} n_s \mathbb{1}\{m_s > 0\} + \underbrace{\sum_{s \leq t} (m_s - n_s)\mathbb{1}\{m_s > 0\}}_{=: X_t} \\
&= X_t + \sum_{s \leq t} n_s \mathbb{1}\{n_s > 0\} + \underbrace{\sum_{s \leq t} n_s \left(\mathbb{1}\{m_s > 0\} - \mathbb{1}\{n_s > 0\}\right)}_{=: Y_t} \\
&= X_t + Y_t + \sum_{s \leq t} n_s \mathbb{1}\{n_s > 0\}.
\end{aligned}$$

To analyse Y_t, notice that

$$\mathbb{1}\{m_s > 0\} - \mathbb{1}\{n_s > 0\} = \mathbb{1}\{m_s > 0, n_s \leq 0\} - \mathbb{1}\{m_s \leq 0, n_s > 0\}.$$

But then, we have

$$\begin{aligned} Y_t &= \sum n_s \mathbb{1}\{m_s > 0, n_s \leq 0\} - \sum n_s \mathbb{1}\{m_s \leq 0, n_s > 0\} \\ &= \sum \underbrace{n_s \mathbb{1}\{m_s > 0, n_s \leq 0\}}_{\leq 0} + \sum \underbrace{(-n_s)\mathbb{1}\{m_s \leq 0, (-n_s) < 0\}}_{\leq 0} \\ &\leq 0. \end{aligned}$$

Further, moving notation back to terms of $\|a_s\|_{V_s^{-1}}$ instead of n_s, m_s, the process X_t is precisely

$$X_t = \sum_{s \leq t} (\mathbb{E}[\|a_s\|_{V_s^{-1}} | \mathfrak{H}_{s-1} - \|a_s\|_{V_s^{-1}}) \mathbb{1}\{\mathbb{E}[\|a_s\|_{V_s^{-1}} | \mathfrak{H}_{s-1}] > M_*/2\kappa_s\},$$

and thus under $\mathsf{Walk}'(\delta)$, $|X_t| \leq \mathrm{LIL}(t, \delta)$. Let us note that this concentration event is likely: indeed, since M_* and κ_s are deterministic quantities, and $\mathbb{E}[\|a_s\|_{V_s^{-1}} | \mathfrak{H}_{s-1}]$ is $\mathfrak{H}_{s-1}$-measurable, we find that the conditional expectation of

$$c_s := (\mathbb{E}[\|a_s\|_{V_s^{-1}} | \mathfrak{H}_{s-1}] - \|a_s\|_{V_s^{-1}}) \mathbb{1}\{\mathbb{E}[\|a_s\|_{V_s^{-1}} | \mathfrak{H}_{s-1}] > M_*/2\kappa_s\}$$

given $\mathfrak{H}_{s-1}$ is zero. Further, of course, since $0 \leq \|a_s\|_{V_s^{-1}} \leq 1$, the random variable c_s is 1-subGaussian given $\mathfrak{H}_{s-1}$, meaning that the LIL applies and so $\mathsf{Walk}'(\delta)$ has chance at least $1 - \delta$.

Now, moving back, we conclude that

$$\forall t, \mathbf{S}_t \leq \kappa_t X_t + 2\kappa_t \sum_{s \leq t} (\|a_s\|_{V_s^{-1}} - M_*/2\kappa_s) \mathbb{1}\{\|a_s\|_{V_s^{-1}} > M_*/2\kappa_s\}.$$

To address the second term, we note that

$$\begin{aligned} &\sum_{s \leq t} (\|a_s\|_{V_s^{-1}} - M_*/2\kappa_s) \mathbb{1}\{\|a_s\|_{V_s^{-1}} > M_*/2\kappa_s\} \\ &\leq \sum_{s \leq t} \|a_s\|_{V_s^{-1}} \mathbb{1}\{\|a_s\|_{V_s^{-1}} > M_*/2\kappa_s\} \\ &\leq \sum_{s \leq t} \frac{2\kappa_s}{M_*} \|a_s\|^2_{V_s^{-1}} \\ &\leq \frac{2\kappa_t}{M_*} \sum_{s \leq t} \|a_s\|^2_{V_s^{-1}}, \end{aligned}$$

where we use the fact that $\mathbb{1}\{u > v\} \leq u/v$ for nonnegative u, v. At this point, we use a refinement of the Elliptical Potential Lemma (Abbasi-Yadkori et al., 2011, Lemma 11), which states that for any sequence of actions in the unit ball, and any t,

$$\sum_{s \leq t} \|a_s\|^2_{V_s^{-1}} \leq 2 \log \det V_t \leq 2d \log(1 + t/d).$$

Indeed, the statement in Lemma 2 is derived from the above via an application of the Cauchy-Schwarz inequality to write $\sum_{s \leq t} \|a_s\|_{V_s^{-1}} \leq \sqrt{\sum_{s \leq 1} 1} \cdot \sqrt{\sum_{s \leq t} \|a_s\|^2_{V_s^{-1}}}$.

In any case, incorporating all of the above, we find under the event $\mathsf{Con}(\delta) \cap \mathsf{Ball}(\delta) \cap \mathsf{Walk}(\delta) \cap \mathsf{Walk}'(\delta)$, that

$$\mathbf{S}_\tau \leq \frac{4\kappa_\tau(\delta)^2}{M_*}(2d \log(1 + \tau/d)) + \kappa_\tau(\delta)\mathrm{LIL}(\tau, \delta).$$

Further, under this event, we know by Lemma 13 that

$$\tau = \widetilde{O}\left(\frac{d^3 + d^2\log(m/\delta)}{\varepsilon^2 M_*^2}\right).$$

Since $\kappa_\tau^2 = O(d^2\log(\tau) + d\log(m/\delta))$, and since $\mathrm{LIL}(\tau,\delta) = O(\sqrt{\tau\log(1/\delta)}) + \widetilde{O}(\sqrt{\tau})$, we conclude that

$$\mathbf{S}_\tau = \widetilde{O}\left(\frac{d^3 + d^2\log(m/\delta)}{M_*} + \frac{d^{5/2}\sqrt{\log(1/\delta)} + d^2\sqrt{\log(m/\delta)\log(1/\delta)} + d^{3/2}\log(m/\delta)\sqrt{\log(1/\delta)}}{\varepsilon M_*}\right),$$

or, dropping all lower order terms under the regime $\log(m/\delta) = o(d)$,

$$\mathbf{S}_\tau = \widetilde{O}\left(\frac{d^3}{M_*} + \frac{d^{5/2}\sqrt{\log(m/\delta)}}{\varepsilon M_*}\right). \qquad \square$$

Conclusion: proof of the main theorem. With all the pieces in place, the proof of the main reliability and stopping analysis of FAST is a simple matter of putting the arguments together, and arguing the frequency of Walk'.

Proof of Theorem 9. We will first show that the probability of the event $\mathsf{Walk}'(\delta/5)$ defined in the statement of Lemma 15 is at least $1-\delta/5$. Recall that the process under consideration is

$$X_t := \sum_{s\le t} \underbrace{(\|a_s\|_{V_s^{-1}} - \mathbb{E}[\|a_s\|_{V_s^{-1}}|\mathfrak{H}_{s-1}])\mathbb{1}\{\mathbb{E}[\|a_s\|_{V_s^{-1}}|\mathfrak{H}_{s-1}] > M_*/2\kappa_s(\delta)\}}_{=:c_s}.$$

But notice that since $\mathbb{E}[\|a_s\|_{V_s^{-1}}|\mathfrak{H}_{t-1}]$ is $\mathfrak{H}_{t-1}$ measurable, and $M_*/\kappa_s(\delta)$ is not random, $\mathbb{E}[c_s|\mathfrak{H}_{s-1}] = 0$ for all s. Further, since $0 \le \|a_s\|_{V_s^{-1}} \le 1$, we immediately conclude that $c_s \in [-1,1]$ surely. Thus, the process X_t is a martingale with $[-1,1]$-bounded, and thus, conditionally 1-subGaussian increments, meaning the LIL (Lemma 3) applies, and so for all δ, with probability at least $1-\delta/5$, $X_t \le \mathrm{LIL}(t,\delta/5)$, as required.

Now, by a union bound, the probability of the event $\mathsf{Con}(\delta/5)\cap\mathsf{Ball}(\delta/5)\cap\mathsf{Walk}(\delta/5)\cap\mathsf{Walk}'(\delta/5)$, as defined in the proof of Lemma 7 and the statement of Lemma 15, is at least $1-\delta$, and further, as detailed in the proof of Lemma 7, the tracking inequality holds under this event.

Then invoking Lemma 12, we conclude that the probability that the output of $\textsc{fast}(\varepsilon,\delta/5)$ is valid is at least $1-\delta$. Further, under the same event, Lemma 13 and Lemma 15 yield the stated bounds on the stopping time and safety costs. $\square$

A.4. Proof of Regret Bounds for Safe Linear Bandits

Proof of Corollary 10. As discussed in the main text, we execute FAST with $\varepsilon = 1/2$, and δ as specified. We recall that FAST produces a reliable output under the event $\mathsf{E} := \mathsf{Con}(\delta/5)\cap\mathsf{Ball}(\delta/5)\cap\mathsf{Walk}(\delta/5)\cap\mathsf{Walk}'(\delta/5)$ defined in the previous sections. This results in an $a_{\mathsf{out}}, M_{\mathsf{out}}$ such that $M(a_{\mathsf{out}}) \ge M_{\mathsf{out}} \ge M_*/2$.

Now, we instantiate the LC-LUCB method of Pacchiano et al. (2024) with $(a_{\mathsf{out}}, M_{\mathsf{out}})$. We note that since we do not exactly know $M(a_{\mathsf{out}})$, we must run this method without the projection onto the subspace orthogonal to a_{out}, as detailed in their remark 11. Further, to ensure efficiency, we carry out the reward optimisation via an L_1-relaxation, which incurs an extra $\sqrt{d}$ factor relative to their regret bounds (Dani et al., 2008). Invoking Theorem 18 of Pacchiano et al. (2024) with these corrections, and observing that $\theta_*^\top(a_* - a_t) \le 2$, we find

$$\mathbf{R}_T \le 2\min(T,\tau) + \widetilde{O}\left(\sqrt{\frac{d^3(T-\tau)_+}{(M_*/2)^2}}\right),$$

and $\mathbf{S}_T \le \mathbf{S}_\tau$. Of course, using the bounds in Theorem 9, under E, $\tau = \widetilde{O}(d^3/M_*^2)$, and so $\mathbf{R}_T = \widetilde{O}\left(d^3/M_*^2 + \sqrt{d^3T/M_*^2}\right)$, and further $\mathbf{S}_T = \widetilde{O}(d^3/M_*)$. Finally, note that $\sqrt{d^3T}M_*^2 < T \iff d^3/M_*^2 < T$, which allows us to absorb the first term above into just a $\widetilde{O}(\sqrt{d^3T/M_*^2})$ bound.

In further detail, we note the only probabilistic condition needed for Theorem 18 of Pacchiano et al. (2024) is precisely the event $\mathsf{Con}(\delta) \supset \mathsf{E}$, along with a similar condition for a confidence estimate of the parameter θ, which only causes the change that $m \mapsto m+1$ in $\omega_t(\delta)$. Thus, one can 'warm start' LC-LUCB with the information accrued up to time τ and retain these guarantees. An entirely analogous argument holds for any other hard enforcement method, including, e.g., SAFE-LTS (Moradipari et al., 2021). □

A.5. Auxiliary Material on Matrix Games

A.5.1. Nonconvexity of the Value of a Matrix Game in its Payoff Matrix

We explicitly calculate the value of the parametric game presented in the footnote in §3.1.

Recall that the payoff, as a function of a variable $z \in \mathbb{R}$ was $\Phi(z) := \begin{pmatrix} 1-z & z \\ z & z-1 \end{pmatrix}$, and the game was defined over $\mathcal{A} = \Delta^2$. Thus, we have

$$K(z) := K(\Phi(z)) = \max_{a\in[0,1]} \min_{\lambda\in[0,1]} \begin{pmatrix} \lambda & 1-\lambda \end{pmatrix} \begin{pmatrix} 1-z & z \\ z & z-1 \end{pmatrix} \begin{pmatrix} a \\ 1-a \end{pmatrix}.$$

By an explicit computation, the objective above works out to

$$U_z(a,\lambda) = \begin{pmatrix} \lambda & 1-\lambda \end{pmatrix} \begin{pmatrix} a+z-2az \\ -1+a+z \end{pmatrix} = -1+a+z+\lambda-2a\lambda z,$$

and we wish to work out $\max_{a\in[0,1]} \min_{\lambda\in[0,1]} U_z(a,\lambda)$.

Now, the coefficient of λ is $(1-2az)$. There are two cases: if $2az \geq 1$, then the coefficient of λ is negative, making the optimal $\lambda = 1$, and if $2az < 1$, the optimal λ is 0. Thus, we are left with resolving $\max_{a\in[0,1]} V_z(a)$, where

$$V_z(a) = \begin{cases} a+z-2az & 2az \geq 1 \\ a+z-1 & 2az \leq 1 \end{cases}.$$

Now, if $z \leq 0$, then the branch $2az \geq 1$ is never attained over $a \in [0,1]$. In this case, the optimal a is just 1, since the constraint $2az < 1$ is never active. We conclude that $K(z) = z$ if $z \leq 0$.

If instead $z \geq 0$, there are many cases:

- If $2z \leq 1$, then the $2az \geq 1$ branch is infeasible. Again the value is just z.
- If $2z \geq 1$, then both branches are possible. In the $a \geq 1/2z$ branch, the coefficient of a is $1-2z \leq 0$, and so the constraint $a \geq 1/2z$ saturates. For $a \leq 1/2z \leq 1$, the objective increases with a, and so this constraint saturates in the second branch. In either case, the optimal value is attained at $a = 1/2z$, and equals $z + 1/2z - 1$.

Thus, the value is

$$K(z) = z + \begin{cases} 0 & z \leq 1/2 \\ 1/2z - 1 & z > 1/2 \end{cases},$$

which can be succinctly written as $z + \min(0, (2z)^{-1} - 1)$. To see the nonconvexity, explicitly observe that $K(0) = 0, K(1) = \frac{1}{2}$, but $K(1/2) = 1/2 > 1/4$.

A.5.2. Reduction of Solving a Matrix Game to Linear Optimisation

We wish to compute the value of a game of the form

$$\max_{a\in\mathcal{A}} \min_{\lambda\in\Delta^m} \lambda^\top \Phi a.$$

Notice that for any a, and $\lambda \in \Delta^m$, $\lambda^\top(\Phi a) \geq \min_i \Phi^i a$, and this value is attained over Δ^m by placing all of the λ mass on the coordinate corresponding to $\arg\min_i \Phi^i a$. Thus, the value of the game is equal to $\max_a(\min_i \Phi^i a) = \max_{a,v\in\mathbb{R}} v : \min_i \Phi^i a \geq v, a \in \mathcal{A}$. But of course, $\min_i \Phi^i a \geq v \iff \Phi a \geq v\mathbf{1}_m$, and further, for any (a,v) that optimise the same, it holds that $\min_\lambda \lambda^\top \Phi a = v$. Thus, it suffices to solve the program

$$\max_{a,v} v \text{ s.t. } v\mathbf{1}_m - \Phi a \leq 0, a \in \mathcal{A},$$

which is an convex program over $d+1$ variables, with m linear constraints along with those defining $\mathcal{A}$, and the corresponding λ can be computed directly by computing Φa_* and then choosing its smallest coordinate. Since LP-time is polynomial in d, increasing d by 1 leaves the time to solve this LP as $O(\text{LP-time})$, and of course, computing λ after the fact only costs $O(m)$ time. For nonzero constraint level α, the margin is modified to $\min_\lambda \lambda^\top(\Phi^* a - \alpha)$. Carrying out the same analysis, we just need to replace Φa by $\Phi a - \alpha$, which makes the constraint $v\mathbf{1}_m - \Phi a + \alpha \leq 0$.

Let us note further that since M_* is directly defined as $\max_{a\in\mathcal{A}} \min_\lambda \lambda^\top \Phi_* a$, the algorithm only ever solves $\max_{a\in\mathcal{A}} \min_\lambda \lambda^\top \widetilde{\Phi}_t a$. Thus, we do not need to invoke the minimax theorem in the above, and the same analysis persists even if $\mathcal{A}$ were not convex (of course, without the efficiency guarantees). The main thing that breaks in this scenario for FAST is that the cumulative action $\sum_{s\leq t} a_s/t$ is not necessarily an element of $\mathcal{A}$, and the design of an appropriate rounding rule would be required along with the stopping condition to yield a feasible action in this situation.

B. Simulation Study

Our main simulations focus on constructing a practical methodology out of the theoretical study of FAST. These are presented in §B.1 below. Additionally, we do auxiliary simulations that compare the behaviour of FAST and EOGT, which are presented in §B.2.

B.1. Main Simulations: The Behaviour of FAST and A Practical Choice of Noise and Boundary.

We investigate the behaviour of FAST, with the focus being to derive practical recommendations for how to set μ, π in this procedure. Throughout, we set Φ_* to be a certain 9×9 directed adjacency matrix, A, obtained from https://sparse.tamu.edu/vanHeukelum/cage4, which has 49 nonzero entries out of 81. This sets up $d = m = 9$. The rows of this matrix were normalised to have norm 1. Throughout, we work over $\mathcal{A} = \{x : \forall i, 0 \leq x_i \leq 1/\sqrt{9}\}$, i.e., we impose known box constraints that lie within the unit ball.

To generate a variety of M_* in a structured way, we work by varying the constraint level. Specifically, we study the safe sets $\mathcal{S}_*(\phi) := \mathcal{A} \cap \{\Phi_* a \geq -\phi \mathbf{1}_9\}$, where ϕ is a scalar parameter. This has the optimal margin $M_*(\phi) = 0.4849 + \phi$, and is feasible so long as $\phi > -0.4849$. In all cases, the optimal margin is attained by the action $a_* = \mathbf{1}_9/\sqrt{9}$.

Plan of Simulations, and brief summary of results. We study three aspects of the behaviour of FAST in three experiments. The results of each experiment raise further open problems pertinent to the practical use of FAST, which we also discuss.

1. In the first case, we attempt to discover what scales of noise are required for a practically efficient version of FAST. The main impetus is that while theoretical analyses of TS require a large noise of typical standard deviation $\sqrt{d}$, in practice, smaller noise of scale 1 is sufficient to retain good optimism rates, and thus such methods in practice tend to deliver an improvement of $\Omega(\sqrt{d})$ in the regret achieved. Such low-regret TS methods do not use the optimism rate in any way (other than in the analysis), so these prior experiments could just simply be executed with this smaller noise. In FAST, we do not need to know the optimism rate π in order to select actions, but we do need it in the boundary design. Note that an overly pessimistic bound on π would thus hit us with a double whammy: an increased stopping time due to both overestimating $1/\pi$, and also due to choosing a very large B.

   To address this, we execute the action selection procedure of FAST$(\mu_\gamma, 0.1)$ where μ_γ is the law of $\mathbf{1}_m \zeta^\top$ with $\zeta \sim \text{Unif}(\sqrt{\gamma d}\mathbb{S}^d)$, without enforcing stopping for 10^4 steps. The theoretical analyses were performed with $\gamma = 3$. In this case, we investigate $\gamma \in 3^{[-7:1]}$, which yields a wide gamut of these scales, but keep $M_* = 1$ fixed in order to gain a sense of the relative optimism in these problems. Note that large M_* means that the probability of optimism is lowered, so this gives us a reasonably reliable bound on π to use in further simulations.

   The main results here that optimism is frequent even with a moderate noise $\gamma = 1/27$. Since FAST is most effective when B is small, but π is large, the main implication is that $\gamma = 1/27$ would yield the most effective noise for FAST. Much smaller γ leads to a significant drop in the optimism rate, particularly in occasional runs, while much larger γ leads to little increase in π, but a linear in γ increase in the stopping time. Of course, proving that in general such a γ, which essentially corresponds to the noise $\text{Unif}(\sqrt{1/3} \cdot \mathbb{S}^d)$, does retain such good optimism is an open problem (although from the observations of prior TS studies, we do expect this to be true).

2. In the second case, we attempt to understand the behaviour of FAST in response to changing M_* and ε. For this, we select $\gamma = 3^{-3}$ from above as a noise with a good balance of $B/\pi < 0.8$. We use this value of B/π, along with a refinement of the boundary process (see the proof of the tracking inequality in §A.1 for these expressions) to investigate

the behaviour of FAST as we vary M_* and ε. Throughout, we will focus on relatively large values of ε, since this is the main regime in which we intend to apply FAST to the SLB problem.

The main results in this case bear out the broad scaling structures in our results in terms of dependence on ε, M_* of the stopping time and safety costs. We also find that the boundary process *still* tends to be too pessimistic, and thus describe the open problem of designing refinements for the same.

3. Finally, in the third case, we investigate the behaviour of the natural *decoupled* noise, i.e., μ such that each row of H_t is drawn independently from the same distribution. In more detail, we revisit the first experiment with this noise, and attempt to understand the optimism rates and reliability of FAST executed with such noise. We find that such decoupled noise is also effective for FAST, but requires a higher γ than the coupled noise (and thus leads to increase in stopping time). We again note that showing that this decoupled noise attains good performance is an open problem.

B.1.1. Investigating Optimism Rates

We execute FAST on the described instance with $\varepsilon = 0.9$, and ϕ chosen so that $M_* = 1$. As mentioned above, the noise laws investigated are the coupled design, with $\zeta \sim \mathrm{Unif}(\sqrt{\gamma d}\mathbb{S}^d)$, for $\gamma \in \{3^i : i \in [-4, 1]\}$. Note that since $d = 9$, the parameter $\gamma = 3^{-2}$ boils down to choosing the uniform law on $\mathbb{S}^d$.

Since the action selection in FAST is independent of the boundary process, and thus of the value of π, we estimate π by simply executing FAST for 10^4 steps, and averaging the counts of the number of times that the selected action was positive. This experiment is run with a ϕ such that the optimal margin is $M_* = 1$. Note that since this M_* is large, this *lowers* the probability of optimism (given that the signal structure is identical in all of these runs, and the margin is adjusted simply by adjusting the constraint threshold). As such, we expect that the resulting values of π are underestimates for situations with smaller margin. Throughout, we let the feedback noise have variance 1. We also record the probability of 'local optimism' in the same way, i.e., the probability of the event L_t from §3.3. In each case, the experiment is executed 100 times.

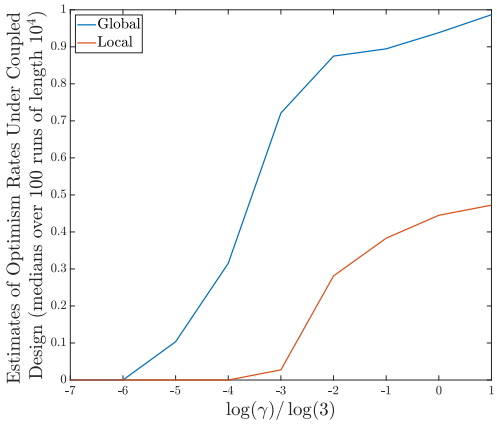

Figure 1. Estimates of the Optimism Rates under the Coupled Design Drawn According to $\mathrm{Unif}(\sqrt{\gamma d}\mathbb{S}^d)$ as γ is varied. *Top*. Medians over 100 runs. Note that the X-axis is log-scale. We observe that global optimism rates are significant for $\gamma = 3^{-3} = 1/3d$, significantly smaller than the theoretically analysed scale of $\gamma = 3$. Similarly, while considerably smaller, local optimism holds for $\gamma \geq 3^{-2}$.

Figure 1 shows medians of these optimism rates as we vary γ The main observation is that the probability of optimism is large even for surprisingly small values of γ, and that this probability is considerably larger than the probability of local optimism (although this too remains nontrivial for large γ). Demonstrating this fact to be true (which is indeed expected from observations regarding TS in single objective low-regret bandits) is a fascinating open problem.

B.1.2. The Behaviour of FAST with respect to M_* and ε

Using the information drawn in the previous section, we now turn to our main simulations on the behaviour of FAST. In order to do so, we run FAST with the noise μ_γ with $\gamma = 3^{-3}$, which, as seen in Figure 1 above provides a regular large probability of optimism. We instantiate the boundary process for FAST with the value of the median global optimism rate estimated above. Note that this has a significant effect: the resulting B/π is roughly 0.8, which makes the boundary process behave as roughly $3\omega \sum \|a_s\|_{V_s^{-1}} + (1 + 2\omega)\mathrm{LIL}$. If we instead ran with the theoretical values of $B/\pi = \sqrt{3d}/0.28 \approx 18$, the boundary process (or refinements of the same, as detailed in §A.1) would instead have a coefficient of ~ 30 for the first term, leading to two orders of magnitude increase in τ in the worst case. Of course, in practice such a π cannot be estimated so simply — resolving this is an important open question for the applicability of methods such as FAST. Our practical recommendation is to simply use a noise of the form $\mathrm{Unif}(\sqrt{c}\mathbb{S}^d)$ for c a not-too-small constant, and set $\pi \sim 0.5$.

We first describe two experimental details before discussing our results.

A Refinement of the Stopping Criterion. In addition to the certificate L_t on the value of M_*, note that our setup yields a running estimate of a potentially feasible action, $\langle a \rangle_t = \sum_{s \leq t} a_s/t$, and using Lemma 1 along with the Cauchy-Schwarz inequality, we can further conclude that with high probability, $\Phi_* \langle a \rangle_t \geq \hat{\Phi}_t \langle a \rangle_t + \omega_t \|\langle a \rangle_t\|_{V_t^{-1}} \mathbf{1}_m$, which yields a further lower bound on the value of the margin, $\min \lambda^\top (\Phi_* \langle a \rangle_t - \alpha) \geq \tilde{L}_t := \min \lambda^\top (\hat{\Phi}_t \langle a \rangle_t - \alpha) - \omega_t \|\langle a \rangle_t\|_{V_t^{-1}}$. In our

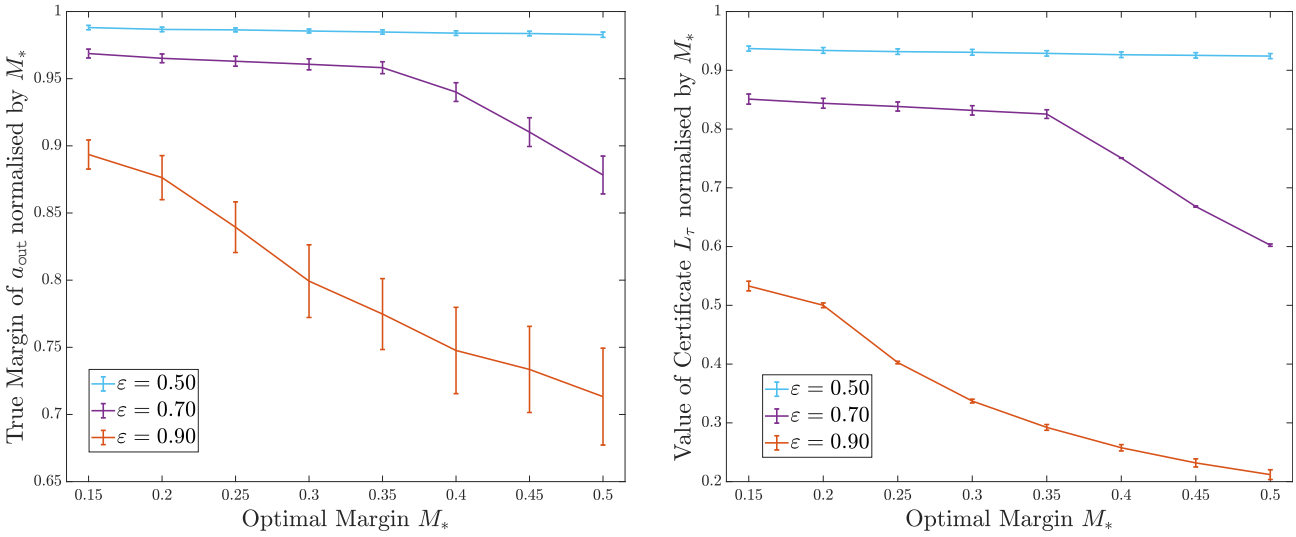

Figure 2. True safety margin of the output of FAST with different ε (left), and the certificate $M_{\text{out}} = \tilde{L}_\tau$ output by it (right), as a fraction of the true margin. Means over 100 runs with one-standard-deviation error bars are presented. Notice that the certified value is typically much larger than $1 - \varepsilon$, and the actual realised value is larger still (always > 0.8), indicating that while reliable, the procedure is too conservative.

experiment, we replace the L_t in the stopping criterion by $\ell_t := \max(L_t, \tilde{L}_t)$. Note that this retains the reliability guarantee. While theoretically this does not improve the behaviour of the stopping time beyond a constant factor, since U_t is left unchanged, we find that in practice, this can significantly improve the stopping time under feasibility, especially if ε and M_* are large.

Reduced feedback noise in simulations. We note that feedback noise is set to be independent Gaussian with standard deviation 0.1, and we use this value to adjust our confidence sets. The main reason for this is for the sake of computational efficiency: the value of the noise standard deviation essentially scales the quantity $\omega_t(\delta)$, and thus decreasing the noise standard deviation by a factor (roughly) shrinks the boundary process $\mathscr{B}_t$ by the same factor. Since the stopping time scale quadratically up to polylog factors in the coefficient of $\sqrt{t}$ in $\mathscr{B}_t$, this change, in essence, reduces the stopping times by a factor of about 100. Since we repeat each of our runs 100 times, this represents a significant computational saving. Indeed, with this change, these simulations took about 3 hours to execute on a midrange laptop computer running a 2022 Ryzen-5U CPU. We note that the relative scaling of all behaviours with M_*, ε is left unchanged by this modification, but, should the actual values be important, the reader should scale them by 100 before interpretation.

Experimental setup. Below, we proceed by setting $M_* \in \{0.15, 0.2, \cdots, 0.5\}$, and $\varepsilon \in \{0.5, 0.7, 0.9\}$, and simulating the behaviour of FAST on the described instances 100 times for each pair of parameters. The choice of M_* is largely to limit computational costs, while the choice of large ε reflects the natural application domain of FAST to SLB problems, as discussed in §4.

Reliability Observations, and the discovery of very safe actions. We find that in *all* of our runs, the resulting feasibility-infeasibility decisions, as well as the actual margin of the output, were always correct, despite the fact that we ran our experiments with the reliability parameter $\delta = 0.1$. In our opinion, this represents an inefficiency in the test design, stemming from the fact that our boundary is too conservative (since a well-defined boundary would see about 10% of decisions being incorrect). We leave the question of finding such a refined boundary to future work. Figure 2 illustrates the same point in the feasible case by showing the actual safety margin of the output $\hat{a}_\tau$ relative to the true margin of the instance.

Certified Margins Figure 2 shows the value of M_{out} output by the method on feasible instances, normalised by M_*. In this case, we again observe that this value tends to be much higher than $1 - \varepsilon$, which bears out the fact that the upper bound U_t decreases too slowly (due to too conservative a $\mathscr{B}_t$)

Stopping Time Behaviour Figure 3 shows the behaviour of the stopping time for both the infeasible case (for which the value of ε is immaterial), as well as for the feasible case with three values of ε. Of these, the runs with $\varepsilon = 0.5$ essentially

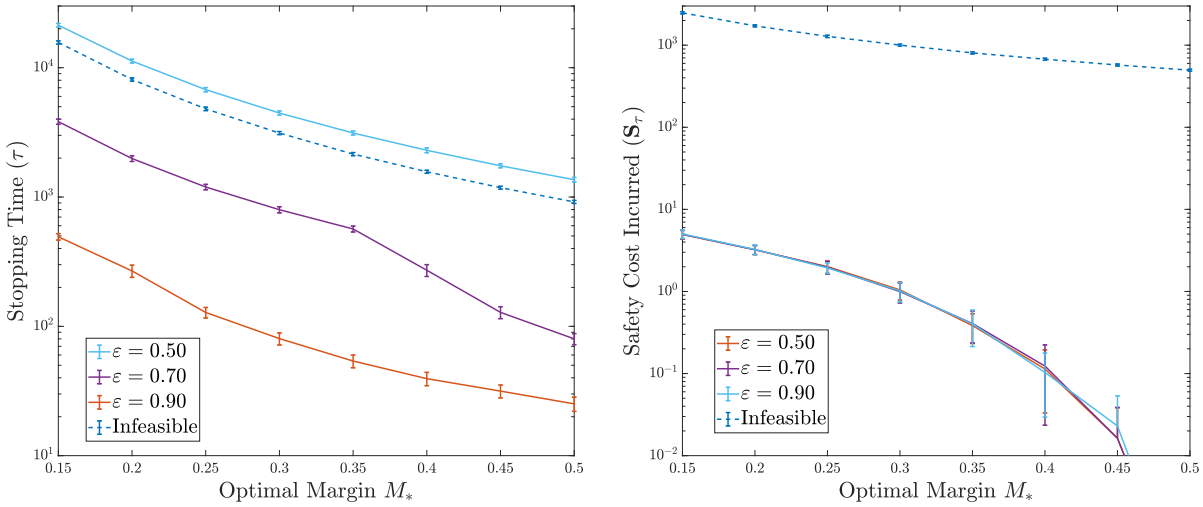

Figure 3. Stopping Time (left) and excess-risk of exploration (right) of FAST in both the infeasible (dashed) case, and the feasible case with different ε. Means over 50 runs with one-standard-deviation error bars are presented. The refinement ℓ_t discussed in this section yields significant improvement in stopping time, especially for large (ε, M_*). Notice also that the excess-risk in feasible scenarios is < 10, much smaller than the corresponding infeasible scenario. We further note that $\mathbf{S}_\tau \in [1.08, 1.11]|M_*|\tau$ in the infeasible case for all $|M_*|$.

correspond to FAST with no refinement of ℓ_t instead of L_t (see below for why), and the stopping time with $\varepsilon = 0.5$ exceeds that of the infeasible case by a factor of about 2. Notice further that the plot clearly demonstrates the inverse quadratic behaviour of τ with M_*.

The early stopping criterion we described helps reduce stopping time mainly when the upper bound U_t takes a long time to decrease. Indeed, in our runs, for small t it essentially held that $U_t = 1, L_t = -1$. Thus, the stopping criterion in these cases boils down to $\ell_t = \tilde{L}_t \geq (1 - \varepsilon)$, and significant gains in stopping time arise when both ε is large, and when M_* is large enough so that $\tilde{L}_t$ grows quickly. This behaviour is demonstrated clearly in Figure 3 by both the sharp drop in stopping time for $M_* = 0.4$, and $\varepsilon = 0.7$, as well as the fact that the stopping time for $\varepsilon = 0.9$ is always $> 50\times$ smaller than that for $\varepsilon = 0.5$ despite the fact that $(0.9/0.5)^2$ is only 3.24.

Safety Costs Figure 3 further shows the behaviour of the safety costs of exploration. There are two qualitatively distinct observations: in the infeasible case, the realised $\mathbf{S}_\tau$ is typically $1.1 \cdot |M_*| \cdot \tau$, bearing out Theorem 9 quite directly (up to the looseness in the τ bounds). In the feasible case, however, we find that the bound of this theorem is extremely loose, and the $\mathbf{S}_\tau$ incurred is very small, being about $100\times$ smaller than the safety costs of exploration in the infeasible case.

B.1.3. Decoupled Noise

Finally, we briefly investigate the behaviour of the natural design of a decoupled noise distribution, i.e., H such that each H^i is drawn uniformly. We study the same set of laws, i.e., each row is drawn from $\mathrm{Unif}(\sqrt{\gamma d}\mathbb{S}^d)$ independently, with $\gamma \in 3^{[-7:1]}$. Figure 4 shows the resulting optimism rates. The main observation is twofold. Firstly, it must be noted that the global and local optimism rates are significant for this noise design, even though our theory does not analyse it. The development of effective analyses for such noise is yet another fascinating open question. The second key observation, however, is that the (global) optimism rates with this decoupled noise design are significantly lower. This raises the values of $\mathscr{B}_t$, since we need a larger $\gamma = 3^{-2}$, and since $\pi \approx 0.5$ is smaller. Roughly, we expect this to cause slowdowns of about $(\sqrt{3} \cdot 0.7/0.5)^2 \approx 6$ times.

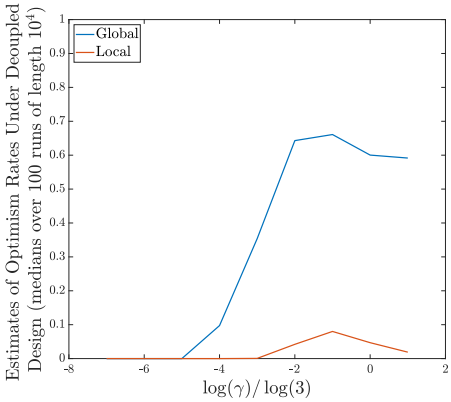

Figure 4. Median-Estimates of the optimism rates with the decoupled noise design with the same setup otherwise as in Figure 1. Observe that the decoupled noise design loses optimism at the same value of γ as the coupled design.

In fact, we observe this behaviour under simulations. These are executed with the above γ (and appropriately adjusted boundary) for the larger values of $M_* \in \{0.35, 0.4, 0.45, 0.50\}$, but with the same values of ε. Figure 5 plots the ratio of the stopping time in this case relative to the stopping times determined for the decoupled design (Figure 3). Notice that in the infeasible and $\varepsilon = 0.5$ feasible cases, the stopping time is indeed about 6 times higher than the coupled design. In the larger ε cases for feasible noise, where early stopping plays a more critical role, the relative loss is smaller, but still significant. We observed the same 6-fold behaviour in the safety cost for the infeasible case, and while again larger, the behaviour was less regular for the feasible case (since $\mathbf{S}_\tau$ is so much smaller, and thus harder to reliably estimate in a multiplicative sense). For this reason, we choose not to present the same data in the case of safety costs.

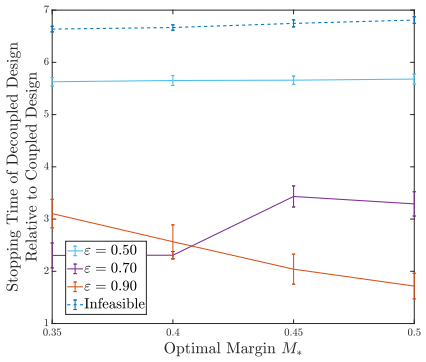

Figure 5. Stopping time of FAST run with decoupled noise, compared to the same with coupled noise. As in Figure 3, means and one-standard-deviation error bars over 100 runs are reported, here on a linear scale. Observe that for small ε in the feasible case, and in the infeasible case, we see a net loss of about a factor of 6. In the case of larger ε, the loss is more limited, although still significant, typically larger than $2\times$.

B.2. Comparison of FAST and EOGT

As a final set of simulations, we investigate the behaviour of EOGT and FAST to compare the merits of the two. Concretely, we will run FAST with the choice of noise and boundary discussed in the previous main section, in particular, setting $\gamma = (9\sqrt{d})^{-1}$, and the B, π values as in the previous section. As such, then, since the domain is significantly different (see below), the success of FAST in this regime serves as a validation of the robustness of this practical design.

Setup. We will implement the relaxed version of EOGT proposed in our prior work (Gangrade et al., 2024b). Recall that this requires us to solve $(2d)^m$ matrix games in each round, and so for practical reasons, only small m can be implemented.

To set the instance, we follow our previous evaluation of EOGT (Gangrade et al., 2024b). Throughout, we work with $m = 2$ constraints. Explicitly, for feasible settings, we impose the constraints

$$a^1 \geq 1/\sqrt{d} - M_*, a^2 \geq 1/\sqrt{d} - M_*,$$

and in the infeasible setting, we impose the constraints

$$a^1 \geq M_*, a^1 \leq -M_*,$$

and work with $\mathcal{A} = [-1/\sqrt{d}, 1/\sqrt{d}]^d$. Note that in both cases, the absolute value of the optimal margin is M_*: in the infeasible case, the point $(0, \cdots, 0)$ has the required margin ($-M_*$, since it is infeasible), and in the feasible case, the point $(1/\sqrt{d}, 1/\sqrt{d}, \cdots)$ has the required margin. This setup is entirely in line with the prior study of EOGT, except that we impose box constraints instead of ball constraints (which allows for faster optimisation).

We investigate the above setup for $d \in [2 : 10]$, and vary $|M_*| \in \{0.2, 0.4, 0.6, 0.8\}$. For the sake of simplicity in presentation, we show the data only for $\varepsilon = 0.5$, although we executed the same for $\varepsilon = 0.7, 0.9$ as well, and saw the same behaviours. As in the previous section, feedback noise is set to Gaussian with standard deviation 0.1.

The main reason for picking box constraints is that the game-solving can be sped up significantly. We do this in a direct brute-force way: since $m = 2$, the parameter λ is effectively one-dimensional. We grid this one dimension at the scale 0.002, and for any Φ, directly compute the values $\lambda^\top \Phi$ at each of these 500 points. For any single value of λ and Φ, the optimal a is simply $\mathrm{sign}(\lambda\Phi)/\sqrt{d}$, and the value of $\max_a \lambda^\top \Phi a$ is thus just $\|\lambda^\top \Phi\|_1$ which is extremely fast to compute. From here, we can compute the minimising λ easily, and then work back to find a. Note that this procedure is actually solving for $\min_\lambda \max_a \lambda^\top \Phi a$, but this is fine due to the minimax theorem. For EOGT, this process is carried out for each Φ lying in the ℓ_1-confidence set (Gangrade et al., 2024b), of which there are $(2d)^m = 4d^2$.

Due to the explicit request of a reviewer of this paper, in these set of experiments we do not implement the early stopping procedure we presented in the previous section. This would largely only affect the large values $\varepsilon = 0.7, 0.9$, which we may expect to be smaller for both methods if this were included.

All instances are simulated 50 times using a MATLAB environment.

Conclusion. The main observation is that while both FAST and EOGT are accurate methods that produce high quality solutions, FAST is strongly preferable because it has both a strong *statisical* and a computational advantage over EOGT, in that it stops *much* sooner (so requiring fewer iterations), and is *much* cheaper to compute per iteration.

Of course, the computational advantage is the main point behind the design of FAST, and so is to be expected, although we note that much of the theoretical advantage survives even the fixed costs expected in practice.

Given the theoretical development in the main text, however, the statistical advantage may be surprising. However, note that we are running FAST with a smaller than analysed scale on noise. Given thi, the observed behaviour is somewhat expected: it is well-established that while analyses of linear-TS are only available for noise of scale $\sqrt{d}$, in typical practical settings, one can execute linear TS with 'non-inflated' constant-sized noise and outperform the regret behaviour of UCB-styles like OFUL (e.g. Abeille & Lazaric, 2017). Within this context, then, EOGT is essentially built upon OFUL, while, of course, FAST exploits a linear-TS style underlying method. The relative advantage is further exacerbated by the fact that stopping times scales roughly quadratically with natural regret metrics, thus giving a very stark benefit.

The remainder of this section presents data supporting the above claim of strong statistical and computational advantage in the two regimes studied.

Fixed d, varying M_*. For our first set of observations, we fix $d = 6$, and vary M_*.

Firstly, we note that in every run, both methods either correctly detected infeasibility, or certified a point with margin at least $M_*/2$. In fact, the quality of the recovered points and the certificates are similar, as seen in Table 1. Of course, observe here that the variance is higher in FAST - this is to be expected, since this is a randomised method, while EOGT is deterministic (given the noise, of course).

Table 1. Comparison of the Quality of a_{out} in the fixed d, varying M_* setting for Feasible Instances, with $\varepsilon = 0.5$. Observe that the quality of solutions is essentially the same, and far exceeds the required 50% in both the certified and realised margin values.

Metric	Method	M_*			
		0.2	0.4	0.6	0.8
Certificate L_τ/M_*	FAST	66.6 ± 0.5	66.4 ± 0.5	66.4 ± 0.6	62.6 ± 0.1
(%age)	EOGT	66.6 ± 0.1	66.6 ± 0.1	66.6 ± 0.1	62.5 ± 0.0
Realised Margin $M(a_{\mathsf{out}})/M_*$	FAST	99.6 ± 0.4	99.4 ± 0.5	99.1 ± 0.7	98.7 ± 1.1
(%age)	EOGT	99.9 ± 0.0	99.9 ± 0.0	99.9 ± 0.0	99.8 ± 0.0

However, the corresponding values of the stopping time are *starkly* different, especially in the feasible case. Indeed, as shown in Table2, the stopping time of FAST in the infeasible case is typically about $30\times$ smaller than that of EOGT, suggesting that FAST is strongly favoured statistically.

Table 2. Comparison of the stopping time of FAST and EOGT for $d = 6, \varepsilon = 0.5$ as M_* varies. The advantage row lists how many times the mean τ for FAST is smaller than that for EOGT. Notice the stark advantage in the feasible case. In the infeasible case, the advantage is small, but still nontrivial.

Metric		$\lvert M_*\rvert$			
		0.2	0.4	0.6	0.8
τ, Feasible Case	FAST	3935 ± 1287	877 ± 245	317 ± 67	138 ± 31
	EOGT	106276 ± 331	24562 ± 105	10292 ± 43	4296 ± 45
	Advantage	$27\times$	$28\times$	$32\times$	$31\times$
τ, Infeasible Case	FAST	731 ± 44	348 ± 14	119 ± 7	48 ± 5
	EOGT	980 ± 15	482 ± 7	263 ± 6	149 ± 4
	Advantage	$1.34\times$	$1.38\times$	$2.21\times$	$3.1\times$

This, of course, is accompanied by a drastic improvement in computation as well. Indeed, for $d = 6$, FAST is $144\times$ faster (theoretically) than EOGT. Practically, we in fact see a $122\times$ speedup per iteration, recovering most of this theoretical gain even for this small d (see below for many more details). The net effect for us is that, e.g., in the $M_* = 0.2$, feasible case, EOGT took about 2700s (with 5-fold parallelization), while FAST took about 1s. Of course, for the infeasible case, FAST remains similarly fast, while EOGT still took ~ 30s.

We note that we see similar behaviour in the safety costs of both methods: for infeasible τ, these grow roughly as $|M_*| \times \tau$, and so are a few percentage higher for EOGT. For the feasible case, both are time ≈ 1 for FAST in the $M_* = 0.2$ case, and ≈ 8 for EOGT for the same - these are miniscule relative to the scale of τ.

Strong Advantage of FAST. This data clearly establishes that FAST is preferable to EOGT even in modest dimensions ($d = 6$)—in the infeasible case, it buys a little bit statistically, and in the feasible case, the statistical gains are massive. This is accompanied by a further strong computational gain (which would be further increased exponentially as m increases, and polynomially as d increases).

Fixed M_*, varying d. To complement the above study, we fix $M_* = 0.4, \varepsilon = 0.5$ and vary d from 2 through 10.

Computation Per-Iteration with d. As discussed in the footnote below, we investigated the time-per-iteration as d grew. Nominally, of course, as d diverges, the relative costs are $4d^2$, but in practice, fixed costs per iteration at finite d tend to reduce in gain. In our experiment, there are two main sources of fixed costs: generating randomness for feedback, and maintaining V_t, and various recording updates to track the data. We find that the time per iteration of FAST does not tend to go below 0.015ms/iteration, even in $d = 2$, suggesting that for very small d these fixed costs begin to be the bottleneck, and reducing the theoretical advantage per iteration. For FAST, in $d = 6$, the costs were about 0.024ms/iteration, while EOGT took about 2.94ms/iteration. For the range investigated, due to the polynomial growth in the time needed for linear programming, the time-per-iteration of FAST never exceeded 0.04ms/iteration, while at $d = 10$, that of EOGT was ≈ 8.5ms/iteration, or $> 200\times$ more.

Quality of Solutions. Again, this is similar: throughout, the certified margin is about $0.66M_*$, while the actual margin of a_{out} is closer to $0.99M_*$. We omit formal presentation of this. Further, safety costs are similar in the feasible case, and scale directly with τ in the infeasible case.

Stopping Time. More importantly, FAST again has a strong advantage in terms of the behaviour of the stopping time, as shown in Figure 6. The relative advantage in stopping behaviour we saw in $d = 6$ persists across dimensions. In particular, for small d, we see about $5 - 10\times$ gain, while for larger d, the gain is always $> 20\times$, hitting a peak of $33\times$ at $d = 7$. Again, coupled with a $> 200\times$ computational advantage per iteration at $d = 10$ (where the statistical advantage is $\approx 22\times$), this makes a strong case that even with modest d and $m = 2$, FAST is strongly preferable to EOGT.

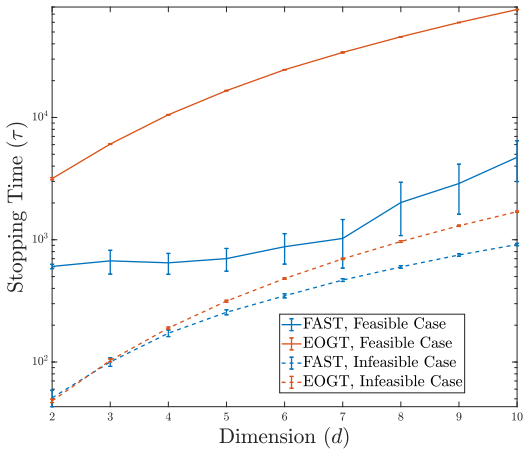

Figure 6. Behaviour of the Stopping Time of FAST and EOGT as d is varied with $m = 2, M_* = 0.4, \varepsilon = 0.5$ in the feasible (solid) and infeasible (dashed) cases. Observe the strong $> 20\times$ advantage in stopping behaviour of FAST relative to EOGT for $d \geq 4$. In this setting, the infeasible case is significantly easier than the feasible, although we still find an advantage of $1.2\times - 1.9\times$ in stopping behaviour.