Near-optimal inference in adaptive linear regression

Koulik Khamaru‡ Yash Deshpande*,
Lester Mackey‡ Martin J. Wainwright†,*

Department of Statistics‡, UC Berkeley
Department of Electrical Engineering and Computer Sciences*, UC Berkeley
Voleon Group* and Microsoft Research‡

July 7, 2021

Abstract

When data is collected in an adaptive manner, even simple methods like ordinary least squares can exhibit non-normal asymptotic behavior. As an undesirable consequence, hypothesis tests and confidence intervals based on asymptotic normality can lead to erroneous results. We propose an online debiasing estimator to correct these distributional anomalies in least squares estimation. Our proposed method takes advantage of the covariance structure present in the dataset and provides sharper estimates in directions for which more information has accrued. We establish an asymptotic normality property for our proposed online debiasing estimator under mild conditions on the data collection process, and provide asymptotically exact confidence intervals. We additionally prove a minimax lower bound for the adaptive linear regression problem, thereby providing a baseline by which to compare estimators. There are various conditions under which our proposed estimator achieves the minimax lower bound up to logarithmic factors. We demonstrate the usefulness of our theory via applications to multi-armed bandit, autoregressive time series estimation, and active learning with exploration.

1 Introduction

Consider a prediction problem in which we observe $n$ datapoints of the form $(\mathbf{x}_i, y_i) \in \mathbb{R}^d \times \mathbb{R}$ with covariate vector $\mathbf{x}_i$ and response $y_i$ linked via the linear model

$$y_i = \mathbf{x}_i^\top \theta^* + \epsilon_i \quad \text{for } i = 1, \ldots, n.$$  

(1)

Here the vector $\theta^* \in \mathbb{R}^d$ is an unknown parameter of interest, and $\epsilon_i$ is additive noise. When the datapoints are generated via some i.i.d. sampling process, this model, and in particular the behavior of the ordinary least squares (OLS) estimate $\hat{\theta}_{\text{LS}}$, is very well-understood. We focus here on a more challenging setting in which the covariate vectors $\{\mathbf{x}_i\}_{i=1}^n$ have been adaptively collected, meaning that the choice of $\mathbf{x}_i$ can depend on the entire set of previous observations $\{\mathbf{x}_j, y_j\}_{j=1}^{i-1}$.

More precisely, given a filtration $\{\mathcal{F}_j\}_{j=1}^n$, assume that the random vector $\mathbf{x}_i$ is $\mathcal{F}_{i-1}$ measurable and that the additive error $\{\epsilon_i\}_{i=1}^n$ is a martingale difference sequence with respect to $\{\mathcal{F}_i\}_{i=1}^n$, so that the random variable $\epsilon_i$ is finite-variance martingale difference sequence, so that it is $\mathcal{F}_i$-measurable, with

$$\mathbb{E}[\epsilon_i | \mathcal{F}_{i-1}] = 0, \quad \text{and} \quad \mathbb{E}[\epsilon_i^2 | \mathcal{F}_{i-1}] = \sigma^2,$$

(2)

for some non-random scalar $\sigma^2 > 0$. We refer to the combination of the linear observation model (1) with such (potentially) adaptive collection procedures as the adaptive linear regression model. Instances of adaptive linear regression arise in a variety of applications, including
multi-armed bandits [21], active learning [11], times series modeling [5], stochastic control [2], and adaptive stochastic approximation schemes [20, 7].

Let us discuss some known results for the OLS estimate $\hat{\theta}_{LS}$. It can be expanded in the form

$$\hat{\theta}_{LS} = S_n^{-1}X_n^Ty_n = \theta^* + S_n^{-1}\sum_{i=1}^{n}x_i\epsilon_i,$$

where $S_n = \sum_{i=1}^{n}x_ix_i^\top$. (3)

This decomposition reveals that the statistical properties of the OLS estimate depend on the martingale transform $\sum_{i=1}^{n}x_i\epsilon_i$, along with the random matrix $S_n^{-1}$. There is a lengthy literature on conditions under which the OLS estimate is consistent [20, 2, 5, 12, 17, 18]. Notably, Lai et al. [20, Thm. 1] show that the OLS estimate is strongly consistent, meaning that $\hat{\theta}_{LS} \overset{\text{a.s.}}{\rightarrow} \theta^*$, whenever

$$\lambda_{\text{min}}(S_n) \overset{\text{a.s.}}{\rightarrow} \infty \quad \text{and} \quad \frac{\log \lambda_{\text{max}}(S_n)}{\lambda_{\text{min}}(S_n)} \overset{\text{a.s.}}{\rightarrow} 0. \quad (4)$$

Arguably, these conditions for consistency are quite mild. In contrast, Lai et al. [20, Theorem 3] also show that asymptotic normality of the least squares estimator in the adaptive linear regression model holds under a stability condition that is substantially more restrictive—namely, the existence of a sequence $\{B_n\}_{n \geq 1}$ of non-random strictly positive definite matrices such that

$$B_n^{-1}S_n \overset{p}{\rightarrow} I. \quad (5)$$

Moreover, Lai et al. [20, Example 3] demonstrate through the example of a unit root autoregressive model, that the OLS estimator fails to be asymptotically normal in absence of the stability property (5). In such cases, confidence intervals and other forms of inference performed using Gaussian limit theory are no longer valid.

Contributions In this paper, we propose and analyze a new family of estimators for the parameter vector $\theta^*$ and its coordinates based on online debiasing techniques. We show that, under mild conditions, our proposed estimators are both asymptotically unbiased and asymptotically normal. The underlying assumptions are less stringent than the stability condition (5) and are satisfied by a large class of models for data generation and protocols for choosing covariate vectors. We provide a detailed discussion of three such example classes in section 4. By deriving minimax lower bounds on the performance of any estimator, we show that the confidence intervals obtained using our estimators are asymptotically near-optimal, in that they match the performance of the best possible estimator up to a logarithmic factor.

Related work The broader literature on bandit algorithms and experimentation focuses mostly on a single statistical objective like minimizing regret or selecting an optimal arm with high probability. In the papers [29, 26], the authors empirically observed that bandit algorithms induce bias, which can be problematic for ex-post inference. Later works [22, 24, 23] characterize the sign and bound the magnitude of this bias. In the paper [13], the authors develop estimators that use propensity scores for the multi-armed bandit setting, a special case of the stochastic regression model (1) in which the covariate vectors $x_i$ are restricted to standard basis vectors. However, it is not clear how to extend this approach to general designs. Also in the bandit setting, Zhang et al. [30] develop a least squares estimator that
exploits an assumed batch structure, meaning that only a fixed, finite number of adaptive decisions are made. This approach, however, does not apply to more general schemes that make adaptive decisions at each round.

There is also a parallel line of work that exploits concentration of measure results (e.g., see the papers [4, 27]) to develop confidence regions that are valid uniformly in time. This approach has its roots in the bandits literature [1, 15] and has been refined in more recent work [14, 16]. An advantage of this approach is that it yields bounds that are uniform in time. On the flip side, it requires very strong exponential tail conditions on the error sequence in contrast to the relatively mild moment conditions that we impose. Overall, we view this line of work as being complementary to our goal of developing corrected estimators that obey asymptotic normality.

This paper builds upon and extends past work, due to a subset of the current authors [7, 8], using online debiasing techniques. We begin with a lower bound, stated in theorem 2, that shows that it is the matrix sequence $S_n^{-1}$ that controls the fundamental difficulty of the problem. This lower bound motivates the particular form of debiasing proposed in this paper. The construction used in past work [7, 8] is based on a non-adaptive upper bound of the form $\lambda_n I$, where the scalar $\lambda_n$ is chosen to be much larger than $\lambda_{\text{max}}(S_n^{-1})$ with high probability. By sharp contrast, our analysis instead makes use of an adaptive upper bound that simultaneously respects the structure of $S_n^{-1}$ and leads to a stable martingale transform; this particular construction and our analysis thereof allows us to obtain sharper guarantees than past work [7, 8].

**Notation** Let us summarize some notation used throughout the remainder of the paper. For a positive integer $n$, we make use of the convenient shorthand $[n] := \{1, 2, \ldots, n\}$. We use $e_j$ to denote the $j$th standard basis vector in $\mathbb{R}^d$. For a matrix $M$, we use the notation $\|M\|_{\text{op}}$ and $\|M\|_F$ to denote the operator norm (maximum singular value) and the Frobenius norm of the matrix $M$, respectively; similarly, we use the notation $\|M\|_{\text{max}}$ to denote the the maximum entry in absolute value. For a square matrix $S$, the quantities $\lambda_{\text{max}}(S)$ and $\lambda_{\text{min}}(S)$ respectively denote the maximum and minimum eigenvalue of the matrix $S$. The quantity $\text{trace}(S)$ denotes the sum of diagonal entries of the square matrix $S$. For a pair of squares matrices $(A, B)$ of compatible dimensions, we use the notation $A \succeq B$ to indicate that the difference matrix $A - B$ is positive semidefinite; we use the notation $A > B$ when the difference matrix $A - B$ is positive definite. The relations $A \preceq B$ and $A < B$ are defined analogously. For a symmetric positive semidefinite matrix $S$, we use $S^\frac{1}{2}$ to denote a symmetric matrix square root of the matrix $S$.

For a sequence of random variables $\{Z_n\}_{n \geq 1}$ and a random variable $Z$, we use the notation $Z_n \xrightarrow{p} Z$ to denote that the sequence of random variables $\{Z_n\}_{n \geq 1}$ converges to $Z$ in probability; the notation $Z_n \xrightarrow{d} Z$ is used to denote convergence in distribution. For a sequence of real-valued random variables $\{Z_n\}_{n \geq 1}$ and a sequence of non-zero real numbers $\{a_n\}_{n \geq 1}$, we say that $Z_n = o_p(a_n)$, if the ratio $Z_n \xrightarrow{p} a_n \rightarrow 0$. We use the notation $Z_n = O_p(a_n)$ to mean that the ratio $Z_n/a_n$ is stochastically bounded. More precisely, for every scalar $\epsilon > 0$, there exists a positive real number $C_\epsilon$ such that $\sup_{n \geq 1} \mathbb{P}[Z_n/a_n > C_\epsilon] < \epsilon$.

2 From ordinary least squares to online debiasing

In this section, we begin by motivating the work by discussing how classical theory about ordinary least squares estimate can break down when data is collected in an adaptive manner.
We then introduce a family of online debiasing estimators for computing alternative estimates.

2.1 Breakdown of the ordinary least squares estimator

Figure 1. Quantitative behavior of the first coordinate $\hat{\theta}_1$ of the ordinary least squares estimator $\hat{\theta}_{LS} := (\hat{\theta}_1, \hat{\theta}_2)$ on a dataset drawn from the $\varepsilon$-greedy two-armed bandit model of section 2.1. The results are obtained with a dataset of size $n = 1000$ and $5000$ independent replications. (a) The distribution of $\frac{\hat{\theta}_2 - \theta^*_2}{\sqrt{(S_n^*)_{22}}}$ is far from standard Gaussian. (b) The bimodal distribution of $(S_n/n)_{22}$ suggests that the scaled covariance matrix $S_n/n$ does not converge to a deterministic matrix.

Let us begin by considering the behavior of the OLS estimator $\hat{\theta}_{LS}$ (3). When the covariates $\{x_i\}_{i \geq 1}$ are either fixed or independently sampled from a fixed distribution, it has several optimality properties. Accordingly, it is natural to ask what the performance of the OLS estimator is when the covariates are drawn in an adaptive manner.

In order to fix ideas, let us consider a two-armed bandit problem [21], a special case of the linear regression model (1) with each $x_i$ chosen to be either $(1, 0)^T$ or $(0, 1)^T$ based on the prior data $\{x_j, y_j \mid j \leq i - 1\}$. In order to generate the covariates $\{x_i\}_{i=1}^n$, suppose that we apply the $\varepsilon$-greedy selection algorithm, a popular choice for tackling bandit problems [21].

A simple simulation helps to reveal some interesting phenomena. We generated linear regression data using $\varepsilon$-greedy selection algorithm with the choices $\varepsilon = 0.1$, $\theta^* = (0.3, 0.3)^T$, and noise variables $\varepsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$. Let $\hat{\theta}_2$ denote the first coordinate of the OLS estimator fit to the bandit data. Figure 1(a) demonstrates that the distribution of $\hat{\theta}_2$, even after proper re-centering and scaling, does not converge to a standard normal distribution. As an undesirable consequence, the confidence intervals for $\theta^*_2$, usually constructed using the quantiles of a standard normal random variable, are not valid.

Let us try to understand why the OLS estimate fails to be asymptotically normal. Figure 1(b) plots a histogram of the $(2, 2)$ entry of the scaled sample covariance matrix $S_n/n$. The bimodal behavior suggests that $S_n/n$ fails to converge to a non-random matrix $B$ and indicates that the stability condition (5) is not satisfied. Indeed, in a recent paper [30], the authors show that when $\theta^*_1 = \theta^*_2$ as in our example, the OLS estimator, after proper centering and scaling, converges to a distribution which is not a standard Gaussian distribution.

It turns out that this distributional anomaly of the OLS estimator is neither specific to the
two-armed bandit problem [21] nor to the ε-greedy algorithm used to simulate the data for Figure 1. The same phenomenon was documented in the time-series and forecasting literature half a century ago, dating back to the works of White [28], Dickey and Fuller [9] and Lai et al. [20]. More recent work [7, 30] has highlighted that a similar phenomenon commonly occurs in multi-armed bandit problems when using popular selection algorithms, including Thompson sampling and the upper confidence bound (UCB) algorithm [21].

In Section 2.2, we rectify the distributional anomaly of the OLS estimator by proposing an estimator based on the online debiasing principles of Deshpande et al. [7] and show that our online debiasing estimator exhibits asymptotic normality even in the absence of the stability condition (5). In Section 4, we demonstrate the usefulness of our theory via applications to the multi-armed bandit problems, autoregressive time series, and active learning problems with exploration.

2.2 Online debiasing estimator

In this section, we propose and analyze an estimator based on an online debiasing technique motivated by the work of Deshpande et al. [7]. At a high-level, the estimator involves a specific perturbation of the ordinary least squares estimator \( \hat{\theta}_{LS} \). This perturbation is constructed via a linear combination of the prediction errors \( \{y_i - x_i^T \hat{\theta}_{LS}\}_{i=1}^n \) along with a carefully chosen sequence of weight vectors \( \{w_i\}_{i=1}^n \). The key property ensured by the construction is that the weight vector \( w_i \) is \( \mathcal{F}_{i-1} \)-measurable for each \( i \in [n] \).

Concretely, for given weight vectors \( \{w_i\}_{i=1}^n \), we compute the online debiasing estimate

\[
\hat{\theta}_{OD} := \hat{\theta}_{LS} + S_n^{-\frac{1}{2}} \sum_{i=1}^n w_i (y_i - x_i^T \hat{\theta}_{LS}).
\]

(6)

Here, the reader should recall our earlier definition \( S_n := \sum_{i=1}^n x_i x_i^T \), and throughout, we assume that the sample covariance \( S_n \) is invertible. The matrix \( S_n^{-\frac{1}{2}} \) denotes a symmetric matrix square root of \( S_n^{-1} \).

Of course, there is an infinite family of estimators of the form (6), and the key question is how to define the weight vectors. In this paper, we propose an estimator in which the sequence \( \{w_i\}_{i=1}^n \) is obtained by solving an optimization problem that takes three inputs:

(i) the original data \( \{(x_i, y_i)\}_{i=1}^n \),

(ii) a non-random scalar \( \gamma_n \in (0, 1] \), and

(iii) a sequence of symmetric positive semidefinite matrices \( \{\Gamma_i\}_{i=1}^n \) such that \( \Gamma_i \in \mathcal{F}_{i-1} \) for each \( i \in [n] \).

In order to simplify notation, we adopt the shorthand \( z_i := \Gamma_i^{-\frac{1}{2}} x_i \). Moreover, for each index \( i \in [n] \), we define the matrices

\[
Z_i := \begin{bmatrix} z_1 & z_2 & \cdots & z_i \end{bmatrix}, \quad \text{and} \quad W_i := \begin{bmatrix} w_1 & w_2 & \cdots & w_i \end{bmatrix}
\]

We also define \( W_0 = 0 \) and \( Z_0 = 0 \). With these definitions the vectors \( \{w_i\}_{i=1}^n \) are obtained recursively by solving the following convex program

\[
\begin{align}
\min_{w \in \mathbb{R}^d} \left\{ \|I - W_{i-1} Z_{i-1} - wz_i^T \|_F^2 + \frac{\gamma_n}{2} \|w\|_2^2 \right\}.
\end{align}
\]

(7a)
Conveniently, this optimization problem has the following explicit solution

\[ w_i = \frac{(I - W_{i-1}Z_{i-1})z_i}{(\gamma_n / 2) + \|z_i\|^2}. \] (7b)

3 Main results

Having motivated and introduced the online debiasing approach, we now turn to some theoretical guarantees that can be given for these methods.

We begin in Section 3.1 by providing sufficient conditions for the online debiasing estimator of section 2.2 to exhibit asymptotically Gaussian behavior (Theorem 1). In Section 3.2, we provide an asymptotically exact confidence region for \( \theta^\ast \) as well as an asymptotically exact confidence interval (Proposition 1) for \( v_j \theta^\ast \), where \( v \) is an arbitrary fixed direction \( v \in \mathbb{R}^d \).

In Section 3.3 (Theorem 2), we complement these results by providing minimax lower bounds on a family of Mahalanobis errors and the length of confidence intervals. These lower bounds apply to any estimator for the stochastic regression model which does not know the true value of the target parameter \( \theta^\ast \) but may have the full knowledge of how the data was collected. Finally, in Section 3.4 we provide general strategies which can be used to verify the conditions of Theorem 1. All of our asymptotic statements assume that the dimension \( d \) is fixed (constant) while the sample size \( n \) grows.

3.1 Asymptotic normality guarantees

The main result of this section is an asymptotic normality guarantee for the proposed estimator (6), where the weight vectors are defined via the recursion (7).

We begin by stating our assumptions and providing some intuition about their role in the theorem.

Assumption A

\( (A1) \) There are positive scalars \( \sigma \) and \( \Delta \) such that the noise sequence \( \{\epsilon_i\}_{i=1}^n \) satisfies the conditions \( \mathbb{E}[\epsilon_i | F_{i-1}] = 0 \) and \( \mathbb{E}[\epsilon_i^2 | F_{i-1}] = \sigma^2 \) for all \( i \in [n] \) and moreover

\[ \max_{i \in [n]} \mathbb{E}[\epsilon_i^{2+\Delta} | F_{i-1}] < \infty. \]

\( (A2) \) The sequence of matrices \( \{S_n\}_{n \geq 1} \) satisfies \( \lambda_{\min}(S_n) \xrightarrow{a.s.} \infty \) and \( \frac{\log \lambda_{\max}(S_n)}{\lambda_{\min}(S_n)} \xrightarrow{a.s.} 0 \).

\( (A3) \) For each \( n \), the scalar \( \gamma_n > 0 \) and positive semidefinite matrices \( \{\Gamma_i\}_{i=1}^n \) with \( \Gamma_i \in F_{i-1} \) are chosen such that:

\( (a) \) Asymptotic negligibility: \( \max_{i \in [n]} \left\{ \frac{1}{\gamma_n} \gamma_n^{-1} \Gamma_i^{-1} \right\} \xrightarrow{p} 0 \) \( \quad \) (8a)

\( (b) \) Vanishing bias: \( \sqrt{\gamma_n \log \lambda_{\max}(S_n)} \cdot \|I - W_n X_n S_n^{-\frac{1}{2}}\|_{\text{op}} \xrightarrow{p} 0 \) \( \quad \) (8b)

\( (c) \) Variance stability: \( \|I - \sum_{i=1}^n w_i x_i^\top \Gamma_i^{-\frac{1}{2}}\|_{\text{op}} \xrightarrow{p} 0. \) \( \quad \) (8c)

6
Let us provide some intuition for the role of each of these assumptions in the theorem. First, Assumption (A1) is quite simple: it imposes relatively mild moment conditions on the noise variables. Second, as discussed in the introduction, Assumption (A2) is standard in guaranteeing the consistency of the least squares estimate. Both Assumptions (A1) and (A2) are viewed as mild conditions in the stochastic linear regression literature and are satisfied by many practical models including those studied in this work [17, 20, 19, 7]. Note that Assumptions (A1) and (A2) concern the regression model itself as opposed to the method: in particular, they do not depend on the algorithm parameters $\gamma_n$ and $\{\Gamma_i\}_{i=1}^n$.

The more subtle requirements for our theorem to apply, which do depend on the algorithm parameters, are stated in Assumption (A3). We discuss the technical role of these conditions in the comments after Theorem 1 to follow. In Section 3.4 to follow, we provide concrete choices of the algorithm parameters $\gamma_n$ and $\{\Gamma_i\}_{i=1}^n$ that ensure that Assumption (A3) holds.

With these preliminaries in place, we are now equipped to state our main theorem on the online debiasing estimator $\hat{\theta}_{OD}$:

**Theorem 1.** Under Assumptions (A1)–(A3) and given any consistent estimator $\hat{\sigma}^2$ of $\sigma^2$, we have

$$\sqrt{\frac{1}{\hat{\sigma}^2}} \cdot S_n^{1/2} (\hat{\theta}_{OD} - \theta^*) \xrightarrow{d} N(0, I).$$ (9)

We prove this theorem in Section 5.1.

A few comments on this theorem are in order. First, it should be noted that needing a consistent estimate for the error variance $\sigma^2$ is a mild requirement; under our conditions, it can be obtained using the training mean squared error of the OLS estimate (see Lemma 3 in the paper [20] for details).

A second important fact is that Assumption (A3) is considerably weaker than the stability condition (5) required for asymptotic normality of the OLS estimate. To reinforce this point, section 4 provides a detailed discussion of three classes of problems for which OLS fails to be asymptotically normal but the guarantee (9) still holds for the online debiasing estimator.

Of all the conditions of theorem 1, verifying the variance stability condition in Assumption (A3) part (c) is the most challenging, and our arguments for doing so vary from problem to problem. In Corollaries 1 and 2, respectively, we verify the variance stability condition for multi-armed bandit problems and autoregressive time series models. In Corollary 3, we verify this condition for a large class of problems satisfying a sufficient exploration condition.

Let us discuss how Assumption (A3) enters the proof of Theorem 1. Our argument is based on the decomposition

$$\sqrt{\gamma_n} \cdot S_n^{1/2} (\hat{\theta}_{OD} - \theta^*) = b_n + v_n,$$ (10a)

where

$$b_n := \sqrt{\gamma_n} \cdot \left( I - W_n X_n S_n^{-1/2} \right) (\hat{\theta}_{LS} - \theta^*)$$ (10b)

and

$$v_n := \sqrt{\gamma_n} \cdot \sum_{i=1}^n w_i \epsilon_i.$$ (10c)

By construction, the term $b_n$ corresponds to the bias in our estimate, a quantity that must be shown to vanish. In order to do so, we first derive an upper bound on the norm $\|b_n\|_2$. 7
The “vanishing bias” condition stated in Assumption (A3)(b) enters in showing that, via our choices of the tuning parameters $\gamma_n$ and $\{\Gamma_i\}_{i=1}^n$, this upper bound converges to zero in probability.

The random vector $v_n$ defines a zero-mean martingale, and our proof controls its behavior via a standard martingale central limit theorem. Doing so requires a Lindeberg type condition on the weight vectors $\{w_i\}_{i=1}^n$, as given in part (a) of Assumption (A3). Moreover, it requires that the conditional covariance of the martingale behave suitably, in which context part (c) of Assumption (A3) enters.

### 3.2 Obtaining confidence regions and intervals

In this section, we use the online debiasing procedure to obtain asymptotically exact confidence regions and intervals.

#### 3.2.1 Confidence region for $\theta^*$

First, consider the problem of finding a confidence region for $\theta^*$—that is, a (random) set $A_{1-\alpha}$ that contains $\theta^*$ with probability at least $1 - \alpha$. We would like a set that is as small as possible, asymptotically exact in the sense that its coverage converges to $1 - \alpha$.

Theorem 1 allows us to construct such a set in the following straightforward way. For any $\alpha \in (0, 1)$, consider the subset of $\mathbb{R}^d$ given by

$$A_{1-\alpha} = \left\{ \theta \in \mathbb{R}^d \mid \frac{\gamma_n}{\sigma} \cdot (\hat{\theta}_{\text{OD}} - \theta)\Sigma_n (\hat{\theta}_{\text{OD}} - \theta)^\top \leq \chi^2_{d, 1-\alpha} \right\}$$

where $\chi^2_{d, 1-\alpha}$ denotes the $(1 - \alpha)$-quantile for a standard chi-squared distribution with degrees of freedom $d$. From the result of Theorem 1, we have the guarantee

$$\lim_{n \to \infty} P(A_{1-\alpha} \ni \theta^*) = 1 - \alpha,$$

In many applications, however, instead of a confidence region for the full vector $\theta^*$, we are instead interested in obtaining a confidence interval for the scalar quantity $v^\top \theta^*$, where $v \in \mathbb{R}^d$ is a fixed direction. It turns out that Theorem 1 no longer provides a straightforward answer to this question. In order to understand why, it is useful to begin by following a naive line of reasoning that is incorrect, and then show how it can be fixed.

#### 3.2.2 An incorrect argument

In order to obtain a confidence interval for $v^\top \theta^*$, it might be tempting to “directly invert” the distributional property (9). In particular, letting $z_{1-\alpha/2} = \Phi^{-1}(1 - \alpha/2)$ denote the $1 - \alpha/2$ quantile of the standard Gaussian distribution, we might claim that the interval

$$\left[ v^\top \hat{\theta}_{\text{OD}} - \frac{\hat{\sigma}}{\sqrt{\gamma_n}} (v^\top S_n^{-1}v)^{1/2} z_{1-\alpha/2}, \quad v^\top \hat{\theta}_{\text{OD}} + \frac{\hat{\sigma}}{\sqrt{\gamma_n}} (v^\top S_n^{-1}v)^{1/2} z_{1-\alpha/2} \right],$$

is an asymptotically exact $1 - \alpha$ confidence interval for $v^\top \theta^*$.

Unfortunately, the conclusion (11) is based on faulty logic, namely that the asymptotic guarantee (9) allows us to write

$$\frac{\sqrt{\gamma_n}}{\sigma} \cdot (\hat{\theta}_{\text{OD}} - \theta^*) \approx \mathcal{N}(0, S_n^{-1}).$$

### 8
This statement is loose in nature; in the absence of the stability condition (5), there is no rigorous and correct form of this statement, since the random matrix \( S_n \) is dependent on the estimate \( \hat{\theta}_{OD} \). Thus, in absence of any further assumptions on the matrix \( S_n \) or direction vector \( v \), the interval (11) is not a valid CI for \( v^T \theta^* \).

Nonetheless, there are certain special cases in which the interval (11) is a valid CI. Concretely, suppose that \( v = e_j \) is one of the standard coordinate basis vectors and that \( S_n \) is diagonal. (This particular case arises in the multi-armed bandit setting studied in Section 4.1, for example.) In this case, the calculations of Section A.1 show that the interval (11) is valid. Thus, given an arbitrary direction \( v \), our strategy is to run a variant of online debiasing that effectively reduces the problem to this favorable case.

### 3.2.3 Correct fixed-direction confidence intervals

Let us now describe the variant of online debiasing that can be used to obtain asymptotically correct confidence intervals for fixed directions. Let \( v \in \mathbb{R}^d \) be the direction of interest; without loss of generality, we assume that \( \|v\| = 1 \). We now form an orthonormal basis of \( \mathbb{R}^d \) with \( v \) as its first element—that is, a collection of orthonormal vectors \( \{v_1, v_2, \ldots, v_d\} \). Let \( V \) be the matrix with \( v_j^T \) as its \( j \)th row. Note that we have \(VV^T = I \) and \( e_1^T V = v \) by construction. Using these two properties, we can write

\[
e_1^T V \theta^* = v^T \theta^* \quad \text{and} \quad y_i = \langle Vx_i, V \theta^* \rangle + \epsilon_i \quad \text{for all } i = 1, \ldots, n.
\]

Consequently, in this new basis, estimating the scalar \( v^T \theta^* \) is same as estimating the first coordinate of transformed vector \( V \theta^* \).

This fact allows us to define a variant of online debiasing that supports asymptotically exact confidence intervals for \( v^T \theta^* \). In particular, let us introduce the notation

\[
x_{v,i} = Vx_i, \quad X_{v,n} = X_nV^T, \quad S_{v,n} = V, S_nV^T, \quad \text{and} \quad D_{v,n} = \text{diag}(VS_n^{-1}V^T)^{-1}.
\]

Now consider the estimator

\[
\hat{\theta}_{v,\text{diagOD}} := \hat{\theta}_{v,\text{LS}} + D_{v,n}^{-\frac{1}{4}} \sum_{i=1}^{n} w_i(y_i - x_{v,i}^TV^T \hat{\theta}_{v,\text{LS}})
\]

where \( \hat{\theta}_{v,\text{LS}} := S_{v,n}^{-1}X_{v,n}y \) is the OLS estimator for \( \{Vx_i\}_{i=1}^n \). We analyze this procedure under the following variant of Assumption (A3).

**Assumption (A3)'**

(A3)' For each \( n \), the scalar \( \gamma_n > 0 \) and positive semidefinite matrices \( \{\Gamma_i\}_{i=1}^n \) with \( \Gamma_i \in \mathcal{F}_{i-1} \) are chosen such that:

\[
\begin{align*}
(a) \text{Asymptotic negligibility:} \quad & \max_{i \in [n]} \left\{ \frac{1}{\gamma_n} x_{v,i}^T \Gamma_i^{-1} x_{v,i} \right\} \overset{p}{\to} 0 \quad (14a) \\
(b) \text{Vanishing bias:} \quad & \sqrt{\gamma_n \log \lambda_{\max}(S_n)} \cdot \left\| D_{v,n}^{\frac{1}{2}} S_{v,n}^{-\frac{1}{2}} - W_nX_{v,n}S_{v,n}^{-\frac{1}{2}} \right\|_{\text{op}} \overset{p}{\to} 0, \quad \text{and} \quad (14b) \\
(c) \text{Variance stability:} \quad & \left\| I - \sum_{i=1}^{n} w_i x_{v,i}^T \Gamma_i^{-\frac{3}{2}} \right\|_{\text{op}} \overset{p}{\to} 0. \quad (14c)
\end{align*}
\]
Proposition 1. Under Assumptions (A1), (A2), and (A3)', given any consistent estimator \( \hat{\sigma}^2 \) of \( \sigma^2 \), the following interval is an asymptotically exact \( 1 - \alpha \) confidence interval for \( v^T \theta^* \)

\[
\left[ e_1^T \hat{\theta}_{v, \text{diagOD}} - \frac{\hat{\sigma}}{\sqrt{n}} (v^T S_n^{-1} v)^{1/2} z_{1-\alpha/2}, \ e_1^T \hat{\theta}_{v, \text{diagOD}} + \frac{\hat{\sigma}}{\sqrt{n}} (v^T S_n^{-1} v)^{1/2} z_{1-\alpha/2} \right].
\]  

(15)

See Section A.1 for the proof of this claim.

A few comments regarding the Proposition 1 and assumption (A3)' are in order. Note that when \( v = e_j \) is one of the standard coordinate basis vectors, a natural choice of the basis matrix \( V \) is \( V = I \). In Lemma 5, we show that the choice \( V = I \) allows us to construct the confidence intervals for all individual coordinates of \( \theta^* \) without constructing a new online debiased estimator \( \hat{\theta}_{v, \text{diagOD}} \) for each \( v = e_j \). Furthermore, when \( v = e_i, V = I \), and the sample covariance matrix \( S_n \) is diagonal, we have \( \hat{\theta}_{v, \text{diagOD}} = \hat{\theta}_{\text{OD}} \); this situation arises in the setting of multi-armed bandit problems which we discuss in Section 4.1.

Finally, we point out that the assumption (A3)' is not significantly stronger than the original assumption (A3). To fix ideas assume \( v = e_j \) and \( V = I \). In that case, assumption (A3)' and (A3) only differs in the vanishing bias condition (b). Assuming \( \sqrt{n} \log \lambda_{\max}(S_n) = o_p(1) \) and condition (A3)(b) holds, we have

\[
\sqrt{n} \cdot \log \lambda_{\max}(S_n) \cdot \| D_n^{-\frac{1}{2}} S_n^{-\frac{1}{2}} - W_n X_n S_n^{-\frac{1}{2}} \|_{op} \\
\leq \sqrt{n} \cdot \log \lambda_{\max}(S_n) \cdot \| I - W_n X_n S_n^{-\frac{1}{2}} \|_{op} \\
+ \sqrt{n} \cdot \log \lambda_{\max}(S_n) \cdot (1 + \| D_n^{-\frac{1}{2}} S_n^{-\frac{1}{2}} \|_{op}) \\
= o_p(1) + \sqrt{d} \cdot o_p(1) \xrightarrow{p} 0.
\]

The last line above uses the vanishing bias condition (A3)(b), the fact that the dimension \( d \) is fixed, and the upper bound \( \| D_n^{-\frac{1}{2}} S_n^{-\frac{1}{2}} \|_{op} \leq \text{trace}(D_n^{-\frac{1}{2}} S_n^{-\frac{1}{2}} D_n^{-\frac{1}{2}}) = d \).

3.3 Minimax lower bounds

Thus far, we have derived two guarantees for online debiasing procedures: asymptotic normality in Theorem 1 along with confidence intervals in Proposition 1. It is natural to wonder in what sense these guarantees are optimal. Accordingly, this section is devoted to lower bounds that apply to the performance of any estimator \( \hat{\theta} \). These bounds are derived within the classical minimax framework and cover two particular risk measures.

Our first risk measure involves the Mahalanobis pseudometric: given an arbitrary positive semidefinite matrix \( M \), possibly random, this pseudometric\(^1\) is given by

\[
\| \hat{\theta} - \theta^* \|_M : = \| M^{\frac{1}{2}} (\hat{\theta} - \theta^*) \|_2,
\]  

(16)

and we provide lower bounds on the squared form of this pseudometric in part (a) of Theorem 2, below. Notably, our analysis allows for the matrix \( M \) to also depend on the dataset \( \{ x_i, y_i \}_{i=1}^n \) itself, so that for example, setting \( M = S_n \) is a valid choice.

Our second risk measure corresponds to the length of a two-sided confidence interval. For a given vector \( v \in \mathbb{R}^d \) and significance level \( \alpha \in (0, 1) \), let \( \mathcal{I}_{\alpha, v} \equiv [\ell_\alpha, u_\alpha] \subseteq \mathbb{R} \) be any level \( \alpha \) confidence interval for the scalar \( v^T \theta^* \), so that by definition, we have

\[
\mathbb{P}_{\theta^* \in \mathcal{I}_{\alpha, v}} \{ \ell_\alpha \leq v^T \theta^* \leq u_\alpha \} \leq \alpha \quad \text{for all } \theta^* \in \mathbb{R}^d.
\]  

(17)

\(^1\)We parameterize the Mahalanobis pseudometric slightly differently than standard definitions, using \( M \) as opposed to its inverse for the quadratic form. This is only for notational ease when \( M \) has a non-trivial null space.
We are interested in finding the smallest such confidence interval, and part (b) of Theorem 2 provides a lower bound on its length $|I_{\alpha,v}| := u_\alpha - \ell_\alpha$.

Our bounds apply to any estimator $\hat{\theta}$, meaning a measurable function of the data as well as the data collection process. The data collection process is summarized by a collection of (potentially randomized) selection algorithms, each of the form $\psi_i : (\mathbb{R} \times \mathbb{R}^d)^{i-1} \rightarrow \mathbb{R}^d$, which take the observed data $\{(x_j, y_j)\}_{j=1}^{i-1}$ up to time $i$ and output a new observation $x_i$. With a slight abuse of notation, we refer to $\Psi_n := (\psi_i)_{i \in [n]}$ as the selection algorithm of the data collection process.

**Theorem 2.** Fix any selection algorithm $\Psi_n$. Under the linear model (1) with i.i.d. Gaussian noise $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ and data collected using $\Psi_n$, the following results hold.

(a) For any choice of $M$ such that $E\text{trace}(S_n^{-1}M)$ is finite, we have

$$\inf_{\theta} \sup_{\theta^* \in \mathbb{R}^d} \mathbb{E}\|\hat{\theta} - \theta^*\|_M^2 \geq \sigma^2 \mathbb{E}\text{trace}(S_n^{-1}M),$$

where the infimum is taken over any estimator $\hat{\theta}$ of $\theta^*$, potentially depending on the selection algorithm $\Psi_n$.

(b) If $E[S_n]$ exists and is invertible, then, for any direction $v \in \mathbb{R}^d$ and $\alpha \in (0, 1/8)$,

$$\inf_{I_{\alpha,v}} \sup_{\theta^* \in \mathbb{R}^d} \mathbb{E}[|I_{\alpha,v}|] \geq \sigma \cdot \left(\frac{1}{2} - \frac{2\alpha}{3}\right) \cdot \left(v^\top (E[S_n])^{-1}v\right)^{1/2},$$

where the infimum is taken over any valid level $1 - \alpha$ confidence interval $I_{\alpha,v}$ for $v^\top \theta^*$ (cf. definition (17)), potentially depending on the selection algorithm $\Psi_n$.

We provide the proofs of Theorem 2(a) and Theorem 2(b) in Sections 5.2.1 and 5.2.2, respectively.

**Comments on the MSE bound**  In order to gain intuition for the MSE bound in part (a), it is helpful to begin with the simplest case—that is, the non-adaptive setting. Let us consider the classical setting of fixed design linear regression, in which the covariates (and hence $S_n$) are viewed as fixed, and the additive noise is zero-mean Gaussian with variance $\sigma^2$. In this case, the standard OLS estimate $\hat{\theta}_{\text{LS}}$ has a Gaussian distribution $\mathcal{N}(\theta^*, \sigma^2 S_n^{-1})$. Consequently, for any fixed matrix $M$, we have

$$\mathbb{E}\|\hat{\theta} - \theta^*\|_M^2 = \sigma^2 \mathbb{E}\text{trace}(S_n^{-1}M),$$

so that the lower bound (18a) is sharp.

Of course, the more substantive content of Theorem 2(a) lies in the fact that it allows for adaptive data collection, along with potentially random choices of $M$. One interesting choice is the random matrix $M = S_n$, for which the bound (18a) guarantees that $\mathbb{E}[\|\hat{\theta} - \theta^*\|_{S_n}^2] \geq \sigma^2 d$. It is worth comparing this lower bound to Theorem 1. From the arguments used to prove this theorem, and under a mildly stronger assumption (A3) where the convergence in distribution conditions are replaced by convergence in $L_1$, it can be shown (see Section 5.1.1) that

$$\lim_{n \rightarrow +\infty} \gamma_n \|\hat{\theta}_{\text{OD}} - \theta^*\|_{S_n}^2 = \sigma^2 d.$$  

(20)
As discussed in the following section (Section 4), in many practical problems of interest, the tuning parameter $\gamma_n$ typically scales logarithmically in the sample size $n$, and also our choice of the tuning parameters ensure that the aforementioned stronger version of assumption (A3) is satisfied. Consequently, the result (20) combined with the lower bound (18a), shows that the online debiasing procedure is minimax optimal up to logarithmic factors.

Another interesting choice is the matrix $\mathbf{M} = \mathbf{e}_1\mathbf{e}_1^T$, for which the lower bound (18a) guarantees that the minimal mean-squared error for estimating the first coordinate $\theta_{1}^*$ is determined by $\sigma^2 \mathbb{E}(\mathbf{S}_n^{-1})_{11}$.

**Comments on the CI bound** Theorem 2(b) provides a lower bound on the length of the confidence interval $v^T \hat{\theta}^*$ which is valid when the data set is collected in an adaptive manner. To the best of our knowledge, this is the first result providing a lower bound on the confidence interval in an adaptive setting.

It is worth comparing this lower bound (18b) to the guarantees provided by the CI construction underlying Proposition 1. First, the pre-factor $c = (\frac{1}{2} - 2\alpha)$ is an artifact of our proof technique. We suspect that it might be possible to remove with a more careful argument. Let us point out a more substantive issue in comparing the two results. From the result (18b), any confidence interval satisfies the lower bound

$$\frac{1}{\sigma^2} \mathbb{E}[|I_{\alpha,v}|^2] \geq v^T (\mathbb{E}[\mathbf{S}_n])^{-1} v.$$  

On the other hand, if we ignore the difference between $\sigma$ and $\hat{\sigma}$ and take a deterministic $\gamma_n$, the CI given by Proposition 1 satisfies

$$\gamma_n \frac{1}{4\sigma^2 z_{\frac{1}{2}-(\alpha/2)}} \mathbb{E}[|I_{\alpha,v}|^2] = v^T \mathbb{E}[\mathbf{S}_n^{-1}] v \overset{(i)}{=} v^T (\mathbb{E}[\mathbf{S}_n])^{-1} v,$$  

where inequality (i) follows from Jensen’s inequality. We suspect that the lower bound (18b) is not sharp, and that the length of optimal CIs should depend on $v^T \mathbb{E}[\mathbf{S}_n^{-1}] v$. However, this conjecture remains to be verified.

### 3.4 Choices of the tuning parameters

Let us now return to the practical issue of choosing the tuning parameters $\gamma_n$ and $\{\Gamma_i\}_{i=1}^n$ of our debiasing procedures. In particular, these parameters must be chosen appropriately so as to ensure that either Assumption (A3), or its variant in Assumption (A3)$'$, is satisfied. In this section, we explore a class of feasible choices of these parameters.

In particular, we analyze choices that are based upon on a deterministic matrix $\mathbf{L}_n$ that acts as a lower bound to the sample covariance matrix $\mathbf{S}_n$. Let $\{\mathbf{L}_n\}_{n \geq 1}$ be a sequence of $d \times d$ diagonal matrices with nonnegative entries satisfying the conditions

$$\|\mathbf{L}_n^{\frac{1}{2}} \text{diag}(\mathbf{S}_n^{-1}) \mathbf{L}_n^{\frac{1}{2}}\|_{op} = O_p(1) \quad \text{and} \quad \lambda_{\min}(\mathbf{L}_n) \overset{a.s.}{\to} \infty. \quad (22)$$

For a given $n$, we define a collection of (diagonal) scaling matrices

$$\Gamma_{i,n} := \max \left\{ \text{diag} \left( \mathbf{S}_i^{-1} \right)^{-1}, \mathbf{L}_n \right\} \quad (23a)$$
where \(\max\{\cdot, \cdot\}\) denotes the element-wise maximum operator.\(^2\) Next, we choose a sequence of tuning parameters \(\{\gamma_i\}_{i=1}^n\) such that
\[
\max_{i \in [n]} \frac{1}{\gamma_n} x_i^\top L_n^{-1} x_i \overset{p}{\to} 0. \tag{23b}
\]

We point out that it is relatively straightforward to find a diagonal matrix \(L_n\) satisfying the condition (22). To fix ideas, let us assume that the covariates satisfy \(\|x_i\|_2 \leq 1\) for all \(i \in [n]\) and that the minimum eigenvalue of the matrix \(S_n\) satisfies \(\lambda_{\min}(S_n) \geq (\log n)^{1 + 2\delta}\) with high probability, for some scalar \(\delta > 0\); the last condition on \(\lambda_{\min}(S_n)\) is only slightly stronger than the minimum eigenvalue condition in Assumption (A2). Then, with the choice \(L_n = (\log n)^{1+2\delta} \cdot I\), the condition (22) is satisfied. Moreover, in this case, the condition (23b) is satisfied for \(\gamma_n = 1/(\log n)^{1+\delta}\).

**Proposition 2.** Consider the solutions \(\{w_i\}_{i=1}^n\) obtained from the optimization problem (7a) using the tuning parameters \(\gamma_n\) and \(\{\Gamma_{i,n}\}_{i=1}^n\) defined in equations (23a)–(23b). Then we have the operator norm bound
\[
\|I - W_n X_n S_n^{-\frac{1}{2}}\|_{op} = O_p(d^2). \tag{24}
\]
In particular, if \(\gamma_n = o_p\left((d^2 \log \lambda_{\max}(S_n))^{-1}\right)\), the vanishing bias and asymptotic negligibility conditions in Assumption (A3) are satisfied.

See appendix A.2 for the proof of this claim.

**Sharper bound for multi-armed bandits** The dimension dependence of the upper bound (24) can be removed in many concrete applications in which we have additional information about the data generating process.

As one concrete example, in the multi-armed bandit model of the sequel (section 4.1), the upper bound can be sharpened to
\[
\|I - W_n X_n S_n^{-\frac{1}{2}}\|_{op} = O_p(1). \tag{25}
\]
See the end of Appendix A.2 for a proof of this claim, and see the proofs of the Corollaries 1, 2, and 3 in the sequel for more details.

### 4 Applications

We next illustrate the concrete consequences of our results in a number of common adaptive learning settings. Sections 4.1 and 4.2 are devoted to multi-armed bandit problems and autoregressive time series models respectively, while section 4.3 discusses active learning with exploration. We end each section with an empirical evaluation of online debiasing. Specifically, we compare the confidence interval (CI) coverage and width of four methods: our online

\(^2\)The choice (23a) of scaling matrix is especially easy to understand for multi-armed bandit problems, where the scaling matrix \(\Gamma_{i,n}\) can be written as \(\Gamma_{i,n} = \max\{S_i, L_n\}\). Assuming that \(S_i = \max\{L_i, S_i\}\) for large value of \(i\), we see that the tuning parameter \(\Gamma_{i,n}\) is the sample covariance matrix up to time \(i\). This assumption indeed holds for Corollaries 1–3 to be presented in the sequel.
debiasing estimator (6), OLS (3) with standard but potentially invalid Gaussian intervals, the W-decorrelation estimator of Deshpande et al. [7], and a valid CI based on the concentration inequality of Abbasi-Yadkori et al. [1]. We highlight that the CIs for OLS are based on the distributional assumption
\[ S_n^{1/2} (\hat{\theta}_{LS} - \theta^*) \sim \mathcal{N}(0, I). \]
This property, while true when covariates are selected \( x_i \) in a non-adaptive manner, need not hold when the covariates \( \{x_i\}_{i=1}^n \) are collected adaptively [7, 30]; as a result, the corresponding CIs need not give the correct coverage. Meanwhile, the valid concentration inequality-based intervals [1] are guaranteed to provide at least the nominal coverage but are often unnecessarily wide.

### 4.1 Multi-armed bandits

Consider a multi-armed bandit with \( d \) arms indexed by the set \([d] := \{1, \ldots, d\}\). At each time \( i \in [n] \), a bandit algorithm selects an arm \( k_i \in [d] \) and observes the reward
\[ y_i = e_k^\top \theta^* + \epsilon_i, \tag{26} \]
where \( e_k \) is the \( k \)th basis vector in dimension \( d \) and \( \theta^* \in \mathbb{R}^d \) is the vector containing the mean rewards of \( d \) arms. We assume that the noise sequence \( \{\epsilon_i\}_{i=1}^n \) satisfies Assumption (A1).

Notably, the multi-armed bandit model (26) is a special case of the adaptive linear regression model (1) with \( x_i = e_{k_i} \) for each \( i \in [n] \).

Since the bandit observation model (26) has a simple linear form, the OLS solution \( \hat{\theta}_{LS} \) is a standard estimate of the reward vector \( \theta^* \in \mathbb{R}^d \). As we mentioned earlier, the behavior of the OLS estimate depends on the stability of the matrix \( S_n \); see the covariance stability condition (5). In the paper [7], the authors conjectured based on empirical evidence that for various popular data selection algorithms, including the Upper Confidence Bound (UCB), Thompson Sampling, and \( \epsilon \)-greedy algorithms (see the book [21]), the stability condition (5) is not satisfied when there are multiple optimal arms. In recent work, Zhang et al. [30] established the validity of this conjecture for the two-armed bandit problem: when the two means are equal, then the OLS estimate fails to have a Gaussian limiting distribution.

In sharp contrast to these negative results for OLS, corollary 1 to follow guarantees that the online debiasing estimator (6) is asymptotically normal under a mild assumption on the minimum number of times that each arm is pulled. More precisely, for each arm \( k \in [d] \) and round \( i \in [n] \), let \( N_{k,i} \) denote the number of times \( k \) is pulled in the first \( i \) rounds, and define the minimum \( N_{\min} := \min_{k \in [d]} N_{k,n} \), and maximum \( N_{\max} := \max_{k \in [d]} N_{k,n} \) arm counts. Then the scaled sample covariance is a \( d \times d \) diagonal matrix, in which the \( k \)th diagonal entry corresponds to the number of times that arm \( k \) is pulled within the first \( i \) rounds:
\[ S_i = \text{diag}(N_{1,i}, \ldots, N_{d,i}). \tag{27} \]

We assume a lower bound on the minimum number of times that each arm is pulled—namely,
\[ N_{\min} \geq (\log n)^{1+2\delta} \text{ for some } \delta > 0. \tag{28} \]
Moreover, we implement the debiasing estimate (6) with the choice of tuning parameters
\[ \gamma_n = (\log n)^{-1+\delta} \quad \text{and} \quad \Gamma_{i,n} = \max \left\{ S_i, (\log n)^{1+2\delta} \cdot I_d \right\}, \tag{29} \]
where \( \max \{\cdot,\cdot\} \) denotes the element-wise maximum operator.
Corollary 1. Suppose the minimum arm pull condition (28) and the moment condition (A1) are valid. Then, given any consistent estimate \( \hat{\sigma}^2 \) of the error variance \( \sigma^2 \), the estimate \( \hat{\theta}_{\text{od}} \) obtained using the tuning parameter choices (29) satisfies

\[
(\hat{\sigma}^2 \cdot (\log n)^{1+\delta})^{-1/2} \cdot S_n^{1/2} \left( \hat{\theta}_{\text{od}} - \theta^* \right) \xrightarrow{d} N(0,1).
\]  

See section 6.1 for the proof of this claim. Corollary 1 also enables us to construct asymptotically exact confidence regions for \( \theta^* \). Moreover, the sample covariance matrix \( S_n \) is diagonal, and as a result, we can also construct confidence intervals of the coordinates \( \theta_{i}^* \); see the comments after Proposition 1. Finally, for a direction \( v \) which is not a standard basis direction, we can obtain an asymptotically exact 1 - \( \alpha \) confidence interval of \( v^T \theta^* \) using Proposition 1; see the comments following Corollary 3 for further details.

4.1.1 Numerical experiment

Fig. 2 illustrates the performance of online debiasing with bandit tuning (29) and \( \delta = 0.05 \). Here we consider a two-armed bandit problem (26) with arm-mean vector \( \theta^* = (0.3, 0.3)^T \) and i.i.d. standard normal error \( \{\epsilon_i\}_{i=1}^n \). The covariates \( \{x_i\}_{i=1}^n \) were generated using the Thompson sampling algorithm [25], and we consider confidence intervals (CIs) for \( \theta_{1}^* \).

We observe first that online debiasing provides appropriate coverage for all confidence levels. Meanwhile, the OLS lower tail interval severely undercovers, and W-decorrelation undercovers for both tails despite having larger widths than online debiasing. Finally, the concentration CI provides 100\% coverage for all confidence levels but yields intervals uniformly larger than the online debiasing CIs. In Appendix C.1, we present analogous results for two other popular multi-armed bandit algorithms, the upper confidence bound (UCB) and \( \varepsilon \)-greedy algorithms.

**Figure 2.** Average coverage and width of confidence intervals for \( \theta_{1}^* \) across 1000 independent replications of a multi-armed bandit experiment (26) with \( \theta^* = (\theta_{1}^*, \theta_{2}^*) = (0.3, 0.3)^T \). The covariates \( \{x_i\}_{i=1}^{1000} \) were selected using Thompson sampling [21], and the error bars represent ±1 standard error. **Left** and **Center**: Coverage of one-sided 1 - \( \alpha \) intervals for \( \theta_{1}^* \). **Right**: Width of two-sided 1 - \( \alpha \) intervals for \( \theta_{1}^* \). See Section 4.1.1 for details.

4.2 Autoregressive time series model

Our next example involves estimating the parameters of an autoregressive time series model. It is well-known that the OLS estimate can exhibit non-Gaussian limit behavior for versions
of such processes that are unstable [20]. In order to focus attention on the key issues, we restrict ourselves here to the simple case of a scalar autoregressive process.

More precisely, given the initial point $y_0 = 0$ and an unknown scalar $\theta^* \in (-1, 1]$, consider a stochastic process generated by the first-order autoregression

$$y_i = \theta^* y_{i-1} + \epsilon_i \quad \text{for } i = 1, \ldots, n. \quad (31)$$

We assume that the noise sequence $\{\epsilon_i\}_{i=1}^n$ consists of i.i.d. standard normal random variables. Note that the autoregression (31) is a special case of the stochastic linear regression model (1), in particular one with $x_i = y_{i-1}$ for all $i \in [n]$. An especially interesting instantiation of the autoregression (31) is obtained by setting $\theta^* = 1$. Such a process is a special case of a unit root autoregression, a class of models that play an important role in econometric time series analysis [5].

With the choice $\theta^* = 1$, the process (31) is a random walk and so has a variance that grows linearly with time. Moreover, by an application of Donsker’s theorem (cf. Example 3 in the paper [20]), we have

$$\frac{1}{n^2} \sum_{i=1}^{n} x_i^2 := \frac{1}{n^2} \sum_{i=1}^{n} y_{i-1}^2 \to \int_0^1 w^2(t)dt, \quad \text{and}$$

$$\sqrt{\sum_{i=1}^{n} y_{i-1}^2 \cdot (\theta_{LS} - \theta^*)} \to \frac{w^2(1) - 1}{2} \int_0^1 w^2(t)dt,$$

where $w(t)$ denotes the standard Wiener process (see the paper [28] for details). Put simply, in the autoregressive time series model (31) with $\theta^* = 1$ the stability condition (5) is not satisfied, and the distribution of the OLS estimate $\theta_{LS}$ is not asymptotically normal.

In contrast to this negative result for the OLS estimate, we can show that the debiasing estimate $\theta_{OD}$, after suitable centering and scaling, does indeed converge in distribution to a standard Gaussian. Our result is based on the tuning parameters and scaling matrices chosen as

$$\gamma_n = 1/(\log n)^{1+\delta}, \quad \text{and} \quad \Gamma_{i,n} = \max \left\{ (\log n)^{1+2\delta} y_{i-1}^2, \sum_{j=1}^{i-1} y_j^2 \right\}. \quad (33)$$

**Corollary 2.** Given a sequence $\{y_i\}_{i=1}^n$ generated from the autoregressive model (31), the estimate $\theta_{OD}$ (6) obtained with the tuning parameters (33) satisfies

$$\sqrt{\sum_{i=1}^{n} y_{i-1}^2} \cdot (\theta_{OD} - \theta^*) \to \mathcal{N}(0, 1). \quad (34)$$

See section 6.2 for the proof of this claim. corollary 2 enables us to construct asymptotically exact confidence intervals for $\theta^*$.

### 4.2.1 Numerical experiment

Fig. 3 illustrates the performance of online debiasing with autoregression tuning (33) and $\delta = 0.05$. Here, our data is generated from the time series model (31) with $\theta^* = 1$. We again find that online debiasing provides appropriate coverage for all confidence levels. Meanwhile, the OLS lower tail interval and the W-decorrelation upper tail interval both exhibit severe undercoverage. Finally, the concentration-based CI again provides 100% coverage for all confidence levels, at the expense of interval lengths that are uniformly longer than the online debiasing CIs.
### 4.3 Active learning with exploration

In our third example, we focus on the case where the covariates \( \{x_i\}_{i=1}^n \) are generated using any algorithm satisfying a sufficient exploration property.

**Definition 1** (Selection algorithms with \( \varepsilon \)-exploration). We say that a selection algorithm \( \Psi_n = \{\psi_i\}_{i=1}^n \) admits a \( \varepsilon \)-exploration property if

\[
x_i := \begin{cases} u_i & \text{with probability } 1 - \varepsilon_i \text{ for some } u_i \in F_{i-1}, \text{ and} \\ v_i & \text{with probability } \varepsilon_i \text{ for some } v_i \text{ independent of } F_{i-1}. \end{cases}
\]  

(35)

Here the exploration probability sequence \( \{\varepsilon_i\}_{i=1}^n \) consists of nonnegative scalars in the interval \((0, 1)\), and the vectors \( \{v_i\}_{i=1}^n \) are i.i.d. random vectors such that

\[
E[v_i v_i^\top] \geq G \quad \text{where} \quad G \succeq 0
\]  

(36)

In words, the selection algorithm \( \psi_i \) behaves as follows: with probability \( 1 - \varepsilon_i \), it chooses vector \( u_i \) based on the previous data points \( \{(x_j, y_j)\}_{j=1}^{i-1} \), and with probability \( \varepsilon_i \), it chooses a random direction \( v_i \), independent of the previous data points.

**Example 1** (\( \varepsilon \)-greedy linear bandits). Let us briefly consider a concrete instance of a selection algorithm \( \{\psi_i\}_{i=1}^n \) that is of the \( \varepsilon \)-greedy type. In the linearly parameterized bandit problem, at each time \( i \in [n] \), an algorithm \( \psi_i \) chooses a context \( x_i \), usually from a bounded set \( A_i \), and obtains a reward \( y_i = x_i^\top \theta^* + \epsilon_i \). A popular and simple strategy for regret minimization is a special case of the \( \varepsilon \)-greedy selection algorithm \([21]\). In the linearly parameterized bandit setting, the selection algorithm \( \psi_i \) makes the following selection

\[
x_i \begin{cases} \in \arg \max_{x \in A_i} x_i^\top \hat{\theta}_{\text{ridge}}^{(i-1)} & \text{with probability } 1 - \varepsilon_i, \\ \sim \text{Unif}(A_i) & \text{with probability } \varepsilon_i. \end{cases}
\]  

(37)

where, \( \hat{\theta}_{\text{ridge}}^{(i-1)} \) denotes a ridge regression estimator which is based on the data up to stage \( i - 1 \), i.e. \( \{x_1, y_1, \ldots, x_{i-1}, y_{i-1}\} \). Put simply, with probability \( 1 - \varepsilon_i \), the selection algorithm \( \psi_i \) chooses an optimal arm given data collected so far (exploitation), and with probability \( \varepsilon_i \)
the algorithm randomizes uniformly amongst its choices (exploration). In the more general setting (35) considered here, it is not necessary to select the optimal arm in the exploitation step. Rather, our result holds also when an arbitrary, \( F_{i-1} \)-measurable choice is made in the first part of eq. (37), as in EXP3 or UCB with exploration [21].

Returning to our general setting (35), we now state a guarantee for selection algorithms with \( \varepsilon \)-exploration. As is standard in the bandit literature, we assume that the covariates are uniformly bounded, so that there exists a scalar \( K \) satisfying

\[
\| x_i \|_2 \leq K \quad \text{for all } i \in [n]. \tag{38a}
\]

See our discussion following the corollary for how this condition can be relaxed. In addition, we impose a sufficient exploration condition, meaning a lower bound on the magnitude of the exploration probabilities, of the form

\[
\sum_{i=1}^{n} \varepsilon_i \geq \frac{\mathbb{E}[\max_{i \in [n]} \| x_i \|^2]}{\lambda_{\min}(G)} (\log n)^{1+2\delta} \quad \text{for some } \delta > 0, \tag{38b}
\]

where the reader should recall that the matrix \( G \) was defined in eq. (36). We implement the debiasing estimate (6) with the choice of tuning parameters

\[
\gamma_n = 1/(\log Kn)^{1+\delta} \quad \text{and} \quad \Gamma_{i,j} = \sum_{j=1}^{n} \varepsilon_j G. \tag{38c}
\]

**Corollary 3.** Suppose that Assumptions (A1) and (A2) hold, the covariates satisfy the bound (38a), and the exploration conditions (36) and (38b) both hold. Then given any consistent estimator \( \hat{\sigma}^2 \) of the error variance \( \sigma^2 \), the estimator \( \hat{\theta}_{\text{OD}} \) with tuning parameters (38c) satisfies

\[
(\log(Kn)^{1+\delta} \cdot \hat{\sigma}^2)^{-1/2} \cdot S_n^{1/2} \left( \hat{\theta}_{\text{OD}} - \theta^* \right) \xrightarrow{d} \mathcal{N}(0, I). \tag{39a}
\]

Moreover, for \( i = 1, \ldots, d \), the following is an asymptotically exact \( 1 - \alpha \) confidence intervals for \( e_i^T \theta^* := \theta_i^* \)

\[
\left[ e_i^T \hat{\theta}_{\text{diagOD}} - \frac{\hat{\sigma}}{\sqrt{\gamma_n}} (e_i^T S^{-1}_n e_i)^{1/2} z_{1-\alpha/2}, \quad e_i^T \hat{\theta}_{\text{diagOD}} + \frac{\hat{\sigma}}{\sqrt{\gamma_n}} (e_i^T S^{-1}_n e_i)^{1/2} z_{1-\alpha/2} \right], \tag{39b}
\]

where the estimator \( \hat{\theta}_{\text{diagOD}} \) is computed via (60) using the scaling matrices (38c) and \( \gamma_n \) satisfying 0 < \( \gamma_n \leq \left( \log(Kn)^{1+\delta} \cdot O_p(\| D_n S_n^{-1/2} D_n \|_{\text{op}}) \right)^{-1}. \)

See section 6.3 for the proof.

A few comments on Corollary 3 are in order. The choice of the tuning parameter \( \gamma_n \) for the second part of the Corollary 3 depends on the quantity the operator norm \( \| D_n^{1/2} S_n^{-1} D_n^{1/2} \|_{\text{op}} \).

While we can calculate the operator norm \( \| D_n^{1/2} S_n^{-1} D_n^{1/2} \|_{\text{op}} \) exactly from the sample covariance matrix \( S_n \), the tuning parameter \( \gamma_n \) is a non-random scalar and hence, it can not depend on the data, and an upper bound on the operator norm \( \| D_n^{1/2} S_n^{-1} D_n^{1/2} \|_{\text{op}} \) is necessary. A
This verifies the conditions of Corollary 3 for the regressors $t$, which equals $d$.

It is worth noting that the bounded covariate condition (38a), while standard in the bandit literature, can be relaxed. For instance, in absence of the condition (38a), one may obtain a result similar to the part (a) of Corollary 3 under the following assumptions:

$$
\gamma_n = o_p(\log \lambda_{\max}(S_n)), \quad \text{and} \quad \sum_{i=1}^{n} \varepsilon_i = \max_{i \in [n]} \mathbb{E}[\|x_i\|_2^2] / \lambda_{\min}(G) \cdot o_p(\gamma_n).
$$

Next, corollary 3 also enables us to construct asymptotically exact confidence region for $\theta^*$ as well as the confidence intervals for $\theta_i^*$, i.e. the $i^{th}$ coordinate of $\theta^*$.

In order to construct a confidence interval for $v^T \theta^*$, where $v \in \mathbb{R}^d$ is an arbitrary unit norm vector, we use the results from Proposition 1. To do so, it suffices to verify that the conditions (36), (38a), and (38b) are satisfied for the new set of regressor vectors $\{Vx_i\}_{i=1}^{n}$. Fortunately, these three conditions are not affected by the change of basis transformation, and are readily satisfied by the regressors $\{Vx_i\}_{i=1}^{n}$.

Indeed, for any orthonormal basis matrix $V$, via linearity of expectation, we have

$$
\mathbb{E}[Vv_i v_i^T V] = V \mathbb{E}[v_i v_i^T] V^\top \geq VGV^\top.
$$

Moreover, for any orthonormal basis matrix $V$ we have

$$
\lambda_{\min}(VGV^\top) = \lambda_{\min}(G) \quad \text{and} \quad \|Vx_i\| = \|x_i\| \leq K.
$$

This verifies the conditions of Corollary 3 for the regressors $\{Vx_i\}_{i=1}^{n}$ with the matrix $G$ replaced by $VGV^\top$. Consequently, invoking Proposition 1 we deduce that for any vector $v \in \mathbb{R}^d$ with $\|v\| = 1$, the following is a asymptotically exact $1 - \alpha$ confidence interval for $v^T \theta^*$

$$
\left[ e_1^T \hat{\theta}_{v,\text{diagOD}} - \frac{\hat{\sigma}}{\sqrt{n}} (v^T S_n^{-1} v)^{1/2} z_{1-\alpha/2}, \quad e_1^T \hat{\theta}_{v,\text{diagOD}} + \frac{\hat{\sigma}}{\sqrt{n}} (v^T S_n^{-1} v)^{1/2} z_{1-\alpha/2} \right],
$$

where, the estimator $\hat{\theta}_{v,\text{diagOD}}$ is constructed using (13).

Finally, as a special case, Corollary 3 and the discussion above also allows us to construct confidence interval for $v^T \theta^*$ for multiarm bandit problems that we discussed in Section 4.1. The condition (38a) is readily satisfied for multiarm bandit problems, but the conditions (36) and (38b) are mildly stronger than the analogous condition (29).

### 4.3.1 Numerical simulation

fig. 4 illustrates the performance of online debiasing with the scaling matrix choice (38c) and $\gamma_n = \frac{1}{2} \cdot \frac{1}{\log(n)^{1/\alpha}}$; recall that $1/\|D_n^{1/2} S_n^{-1/2} D_n^{1/2}\|_{\text{op}} \leq \frac{1}{\alpha} = 1/2$. Here we consider a linear bandits problem with $\theta^* = (0.3, 0.3)^\top$ and i.i.d. standard normal error $\{\varepsilon_i\}_{i=1}^{n}$. The covariates $\{x_i\}_{i=1}^{n}$ were generated using the $\varepsilon$-greedy linear bandits algorithm (37), where, for each stage, the context set $A_i$ consisted of the same 50 vectors drawn and uniformly from the unit sphere in dimension 2. For this problem, the exploration lower bound eq. (36) is satisfied with $G = \frac{1}{|A|} \sum_{a \in A} a a^\top$. In this setting, Abbasi-Yadkori et al. [1] only provide concentration-based CIs based on ridge regression estimators, rather than OLS. Here we report the CIs.

---

*See the proof of Proposition 1 using Lemma 5.*
from ridge regression with regularization parameter $\lambda_{\text{Ridge}} = 0.01$ (which closely approximates the OLS solution) and display analogous results for alternative regularization parameters in Appendix C.2.

We observe once more that online debiasing provides appropriate coverage for all confidence levels, while the OLS lower tail interval severely undercovers. Meanwhile, the concentration CI provides high coverage for all confidence levels but yields intervals (mostly) larger than the online debiasing CIs.

![Figure 4](image_url)

**Figure 4.** Average coverage and width of confidence intervals for $\theta^*_1$ and $\theta^*_2$ across 1000 independent replications of a linear bandits experiment (37) with $\theta^* = (\theta^*_1, \theta^*_2) = (0.3, 0.3)^T$. The covariates $\{x_i\}_{i=1}^{1000}$ were selected using the $\varepsilon$-greedy linear bandits algorithm (37), and the error bars represent $\pm 1$ standard error. **Left** and **Center:** Coverage of one-sided $1 - \alpha$ intervals for $\theta^*_1$ and $\theta^*_2$. **Right:** Width of two-sided $1 - \alpha$ intervals for $\theta^*_1$ and $\theta^*_2$. See section 4.3.1 for details.
5 Proofs of the theorems

In this section, we provide the proofs of our two main results. We prove Theorem 1 in Section 5.1, and Theorem 2 in Section 5.2.

5.1 Proof of Theorem 1

Using the condition \( \lambda_{\min}(S_n) \overset{a.s.}{\rightarrow} \infty \) from assumption (A2), we know that, so that we may assume without loss of generality that \( S_n \) is invertible. We claim that it suffices to show that

\[
\gamma_n \mathbf{S}_n^{-\frac{1}{2}} (\hat{\theta}_{OD} - \theta^*)
\]

converges in distribution to \( N(0, \sigma^2 I_d) \). Indeed, when this claim holds, then since \( \sigma^2 \overset{p}{\rightarrow} \sigma^2 \) by assumption, Slutsky’s theorem implies the claim of the theorem.

Recall from equation (10) that the random vector \( \gamma_n \mathbf{S}_n^{-\frac{1}{2}} (\hat{\theta}_{OD} - \theta^*) \) can be decomposed into the sum \( b_n + v_n \). Based on this decomposition, we see that it is sufficient to prove that \( b_n \overset{p}{\rightarrow} 0 \) and \( v_n \overset{d}{\rightarrow} N(0, \sigma^2 I_d) \). The remainder of our proof is devoted to establishing these two claims.

Analysis of \( b_n \) By definition of the operator norm, we have the upper bound

\[
\|b_n\|_2 \leq \sqrt{\gamma_n} \|I - W_n X_n S_n^{-\frac{1}{2}}\|_{op} \|S_n^{-\frac{1}{2}} (\hat{\theta}_{LS} - \theta^*)\|_2. \tag{40}
\]

Theorem 1 from the paper [20] guarantees that

\[
\|S_n^{-\frac{1}{2}} (\hat{\theta}_{LS} - \theta^*)\|_2 = O \left( \sqrt{\log \lambda_{\max}(S_n)} \right) \overset{a.s.}{\rightarrow} 0. \tag{41a}
\]

On the other hand, the vanishing bias condition from Assumption (A3)(b) guarantees that

\[
\sqrt{\gamma_n \log \lambda_{\max}(S_n)} : \|I - W_n X_n S_n^{-\frac{1}{2}}\|_{op} \overset{p}{\rightarrow} 0. \tag{41b}
\]

Applying the bounds (41a) and (41b) to the right-hand side of the inequality (40) shows that \( \|b_n\|_2 \overset{p}{\rightarrow} 0 \).

Analysis of \( v_n \) In order to control the second term, we seek to apply a classical martingale central limit theorem (cf. Theorem 2.2 in the paper [10]). We begin by observing that \( \{\sqrt{\gamma_n} w_i \epsilon_i\}_{i=1}^n \) is a martingale difference sequence with respect to the sigma-field \( \{F_i\}_{i=1}^n \). Noting that the tuning parameter \( \gamma_n \) is non-random, it follows that the sum \( \sum_{i=1}^n \sqrt{\gamma_n} w_i \epsilon_i \) has zero mean and moreover that

\[
\sum_{i=1}^n \text{Cov} \left[ \sqrt{\gamma_n} w_i \epsilon_i \mid F_{i-1} \right] = \gamma_n \sum_{i=1}^n w_i w_i^\top. \tag{42}
\]

Consequently, in order to apply the martingale CLT so as to obtain the stated claim, we need to show that

\[
\gamma_n \sum_{i=1}^n w_i w_i^\top \overset{p}{\rightarrow} \sigma^2 I_d.
\]

Doing so requires the following auxiliary lemma, which characterizes the behavior of the weight vector sequence \( \{w_i\}_{i=1}^n \) constructed in equation (7b).
Lemma 1. Under the Assumption (A3) parts (a) and (c), the sequence of vectors \( \{w_i\}_{i=1}^n \) obtained from equation (7b) has the following properties:

\[
\begin{align*}
(\text{Stability:}) & \quad \gamma_n \sum_{i=1}^n w_i w_i^\top \xrightarrow{p} I_p, \quad \text{and} \\
(\text{Vanishing norm:}) & \quad \max_{i \in [n]} \sqrt{\gamma_n} \|w_i\|_2 \xrightarrow{p} 0.
\end{align*}
\]

See Appendix B for the proof of this lemma.

With the above lemma in hand, we now apply a standard martingale central limit theorem\(^4\) to conclude that

\[
\sum_{i=1}^n \sqrt{\gamma_n} w_i \epsilon_i \xrightarrow{d} \mathcal{N}(0, \sigma^2 I_d).
\]

Putting together the pieces, we conclude that

\[
\sqrt{\gamma_n} \cdot S_n^{\frac{1}{2}} (\hat{\theta}_{OD} - \theta) = b_n + \sum_{i=1}^n \sqrt{\gamma_n} w_i \epsilon_i \xrightarrow{d} \mathcal{N}(0, \sigma^2 I_d),
\]

which completes the proof of Theorem 1.

5.1.1 Proof of claim (20):

For simplicity, let us assume \( \sigma \) is known. Recalling the decomposition (10) we have

\[
\gamma_n \cdot \|\hat{\theta}_{OD} - \theta^*\|^2_{S_n} = \|b_n\|^2 + 2b_n^\top v_n + \|v_n\|^2 \leq \|b_n\|^2 + 2\|b_n\| \cdot \|v_n\| + \|v_n\|^2
\]

Invoking the condition \( \sqrt{\gamma_n \log \lambda_{\max}(S_n)} \cdot \|I - W_nX_nS_n^{-\frac{1}{2}}\|_{\text{op}} \xrightarrow{L_1} 0 \) we immediately have \( \|b_n\|^2 \xrightarrow{L_1} 0 \). It suffices to show that \( \|v_n\|^2 \leq d \) and \( \mathbb{E}[\|v_n\|^2] \to d \). Observe that

\[
\mathbb{E}[\|v_n\|^2] = \gamma_n \cdot \sum_{i=1}^n \mathbb{E}[\epsilon_i^2 \|w_i\|^2] + \sum_{i \neq j} \gamma_n \cdot \mathbb{E}[\epsilon_i \epsilon_j w_i^\top w_j]
\]

The last line above follows from the assumption \( \mathbb{E}[\epsilon_i^2 | F_{i-1}] = \sigma^2 \) and the fact (by construction) that \( w_i \in F_{i-1} \). Taking trace on both sides of equation (68) we have that \( \gamma_n \cdot \|w_i\|^2 \leq \), and using the stronger \( L_1 \) version of assumption (A3) along with the proof techniques of Lemma 1 we have \( \gamma_n \cdot \sum_{i=1}^n \mathbb{E}[\|w_i\|^2] \to \sigma^2 d \). It remains to show that \( \mathbb{E}[\epsilon_i \epsilon_j w_i^\top w_j] = 0 \) for all \( i \neq j \). Without loss of generality, assume \( i < j \). By construction of \( w_i \) and the martingale assumption (A1) of the noise \( \epsilon_i \), we have \( \{w_i, w_j, \epsilon_i\} \in F_{j-1} \). As a result, we conclude

\[
\mathbb{E}[\epsilon_i \epsilon_j w_i^\top w_j] = \mathbb{E} \left[ \epsilon_i \cdot w_i^\top w_j \mathbb{E}[\epsilon_j | F_{j-1}] \right] = 0
\]

This completes the proof of the claim (20).

\(^4\)Concretely, by applying Theorem 2.2 from the paper [10], we first show that for any unit vector \( u \), the inner product \( \frac{1}{\sqrt{d}} \sum_{i=1}^n \sqrt{\gamma_n} w_i \epsilon_i \) converges to a standard Gaussian.
5.2 Proof of Theorem 2

We prove part (a) of Theorem 2 in Section 5.2.1 and part (b) of Theorem 2 in Section 5.2.2.

5.2.1 Proof of Theorem 2(a)

Throughout the proof, we use \( \hat{\theta} \) to denote a generic estimator for \( \theta^* \). We assume that the estimator \( \hat{\theta} \) is a function only of the \( n \) datapoints \( \{(x_i, y_i)\}_{i=1}^n \) and the \( n \) selection algorithms \( \Psi_n : = (v_i)_{i \in [n]} \) and that \( \hat{\theta} \) does not know the value of the true parameter \( \theta^* \). Consider any positive semidefinite and potentially data-dependent matrix \( M \in \mathbb{R}^{d \times d} \), and define the nonnegative scalar loss function

\[
\ell_M(\hat{\theta}, \theta^*) = (\hat{\theta} - \theta^*)^\top M (\hat{\theta} - \theta^*). \tag{43}
\]

From minimax to Bayes risk

In terms of the above notations, Theorem 2 (a) posits a lower bound on the minimax risk:

\[
\inf_{\hat{\theta}} \sup_{\theta^* \in \mathbb{R}^d} \mathbb{E}[\ell_M(\hat{\theta}, \theta^*) | \theta^*], \tag{44}
\]

where in the above expression, we have taken an expectation of the loss \( \ell_M(\hat{\theta}, \theta^*) \) over the randomness in the data \( (X_n, y_n) \) conditioned on \( \theta^* \). We establish the lower bound Theorem 2(a) on the minimax risk (44) by first lower bounding the minimax risk by the Bayes risk and then providing a lower bound for the Bayes risk. Concretely, we use the inequality

\[
\inf_{\hat{\theta}} \sup_{\theta^* \in \mathbb{R}^d} \mathbb{E}[\ell_M(\hat{\theta}, \theta^*) | \theta^*] \geq \inf_{\hat{\theta}} \mathbb{E}_{\theta^*, X_n, Y_n} \ell_M(\hat{\theta}, \theta^*), \tag{45}
\]

where the expectation \( \mathbb{E}_{\theta^*, X_n, Y_n} \) above is taken with respect the joint distribution of the random variable \( (\theta^*, X_n, Y_n) \); this joint distribution is computed by assigning a prior distribution on the parameter \( \theta^* \).

Main argument

We claim that it suffices to prove that for any estimator \( \hat{\theta} \)

\[
\mathbb{E}[\ell_M(\hat{\theta}, \theta^*) | X_n, Y_n] \geq \sigma^2 \text{trace}(MS_n^{-1}). \tag{46}
\]

where the expectation \( \mathbb{E}[\cdot | X_n, Y_n] \) is taken with respect to the conditional distribution \( \theta^* | X_n, Y_n \). Indeed, taking expectation over \( X_n, Y_n \) yields the desired bound:

\[
\mathbb{E}_{\theta^*, X_n, Y_n} \ell_M(\hat{\theta}, \theta^*) = \mathbb{E}_{X_n, Y_n} \mathbb{E}[\ell_M(\hat{\theta}, \theta^*) | X_n, Y_n]
\geq \sigma^2 \text{trace}(MS_n^{-1}). \tag{47}
\]

It remains to prove the bound (46).

Proof of bound (46)

We complete the proof of this bound by first computing the conditional distribution of the random variable \( \theta^* | X_n, Y_n \) and then providing a lower bound on the conditional expectation of the loss \( \ell_M(\hat{\theta}, \theta^*) \) given the data \( (X_n, Y_n) \). Concretely, we show that under a prior distribution \( \theta^* \sim \mathcal{N}(0, \rho^2 I_d) \), we have:

\[
\theta^* | X_n, Y_n \sim \mathcal{N}(\mu_n, \Sigma_n), \quad \mu_n = \Sigma_n X_n y_n \quad \text{and} \quad \Sigma_n = (S_n/\sigma^2 + I_d/\rho^2)^{-1}. \tag{48}
\]
A simple calculation using the above distributional property yields that for any positive semidefinite matrix \( M \) (which may depend on the data \((X_n, y_n)\)), the expected conditional loss \( \mathbb{E}[\ell_M(\theta, \theta^*) \mid X_n, y_n] \) is minimized \(^5\) when \( \hat{\theta} = \Sigma_n X_n y_n \) with a minimum value of \( \sigma^2 \text{trace}(M \Sigma_n) \). Finally, we are free to choose the value of the prior error variance \( \rho^2 \), and letting \( \rho^2 \to \infty \) yields the claim (46). Now let us prove the claim (48).

Proof of claim (48) We complete the proof by induction on the number of datapoints \( n \).

Base case For \( n = 0 \), we have
\[
\theta^* \mid X_0, y_0 \equiv \theta^* \sim \mathcal{N}(0, \rho^2 I_d).
\]
(49)

The last statement above follows from the prior assumption \( \theta^* \sim \mathcal{N}(0, \rho^2 I_d) \), and the triple \((X_0, y_0, S_0)\) are defined as zeros of respective dimensions. This proves the statement (48) for \( n = 0 \), and \( \mu_0 = 0 \), and \( \Sigma_0 = \rho^2 I_d \).

Inductive step Assume that the claim (48) is true for \( n - 1 \) for some \( n \geq 1 \). We will prove that the statement holds true for \( n \). Recall that, the query algorithm \( \psi_1 : (\mathbb{R} \times \mathbb{R}^d)^{i-1} \to \mathbb{R}^d \) is oblivious to the true value \( \theta^* \); thus, the conditional distribution \( x_n \mid F_n \) is independent of \( \theta^* \) (see the discussion before Theorem 2). Furthermore, from the model (1) we have that the conditional distribution of \( y_n \mid F_{n-1}, x_n, \theta^* \) is \( \mathcal{N}(x_n^T \theta^*, \sigma^2) \), and using the induction hypothesis (48) we conclude that \( \theta^* \mid X_{n-1}, y_{n-1} \sim \mathcal{N}(\mu_{n-1}, \Sigma_{n-1}) \). With the last three observations in hand, an application of the Bayes theorem yields
\[
\frac{d\mathbb{P}(\theta^* \mid X_n, y_n)}{d\mu^d(\theta^*)} \propto \exp \left\{ -\frac{1}{2} (\theta^* - \mu_{n-1})^T \Sigma_n^{-1} (\theta^* - \mu_{n-1}) \right\} \times \exp \left\{ -\frac{1}{2\sigma^2} (y_n - x_n^T \theta^*)^2 \right\} \\
\times \exp \left\{ -\frac{1}{2} (\theta^* - \mu_n)^T \Sigma_n (\theta^* - \mu_n) \right\}
\]
where \( \frac{d\mathbb{P}(\theta^* \mid X_n, y_n)}{d\mu^d(\theta^*)} \) denotes the Radon-Nikodym derivative of \( \theta^* \mid X_n, y_n \) with respect to the Lebesgue measure \( \nu^d(\cdot) \) on \( \mathbb{R}^d \), and the pair \((\mu_n, \Sigma_n)\) satisfies the following equations:
\[
\Sigma_n^{-1} = \Sigma_{n-1}^{-1} + \frac{x_n x_n^T}{\sigma^2} \quad \text{and} \quad \mu_n = \frac{1}{\sigma^2} \Sigma_n \sum_{i=1}^n x_i y_i.
\]
This completes the proof of the inductive step, and putting together the pieces yields the claim of Theorem 2(a).

5.2.2 Proof of Theorem 2(b)

We use \( \mathcal{D}_n = (x_1, y_1, \ldots, x_n, y_n) \) to denote the full data up to time step \( n \), and we use \( \mathbb{P}_{\theta_0}(\mathcal{D}_n) \) and \( \mathbb{P}_{\hat{\theta}_0}(\mathcal{D}_n) \) to denote the marginal distribution of the random variable \( \mathcal{D}_n \) under \( \theta^* = \theta_0 \).

\(^5\)Here, we have assumed that the prior distribution \( \theta^* \sim \mathcal{N}(0, \rho^2) \) and the error variance \( \sigma^2 \) are known to the estimator \( \hat{\theta} \); this assumption is justified since without the knowledge of the prior distribution on \( \theta^* \) and error-variance \( \sigma^2 \), the minimum value of the expected loss \( \mathbb{E}[\ell_M(\hat{\theta}, \theta^*) \mid X_n, y_n] \) can only increase, which yields a (possibly) stronger lower bound.
and $\theta^* = \theta_1$. Throughout, the points $\theta_1, \theta_0$ are two fixed (non-random) points that are chosen to simplify certain calculations in the proof. Finally, the scalars $TV(\cdot, \cdot)$ and $KL(\cdot, \cdot)$, respectively, denote the total variation distance and the Kulback-Liebler distance between two distributions.

Confidence intervals:

We provide a lower bound on the length of $1 - \alpha$ confidence intervals $I_{\alpha, v}$ for $v^\top \theta$, where $v \in \mathbb{R}^d$ is a fixed unit vector, and $\alpha \in (0, \frac{1}{8})$, is the level of the confidence interval. Concretely, the set of valid confidence intervals is

$$I_{\alpha, v}(\theta) = \left\{ I_{\alpha, v} = [l(D_n), u(D_n)] : \inf_{\theta \in \Theta} \mathbb{P}_\theta [l(D_n) \leq v^\top \theta \leq u(D_n)] \geq 1 - \alpha \right\}$$

Here, $\mathbb{P}_\theta(\cdot)$ denotes the conditional distribution of $D_n$ under $\theta$, and our goal is to provide lower bounds on the expected length of the confidence interval defined as

$$|I_{\alpha, v}(\Theta)| := \sup_{\theta \in \Theta} \mathbb{E}_\theta[u(D_n) - l(D_n)]. \quad (50)$$

Main argument

The proof of the theorem is based on the following lemma proved in Section 5.2.3.

**Lemma 2.** Introduce the shorthand $a_+ := \max\{a, 0\}$. Then

$$|I_{\alpha, v}(\theta)| \geq \sup_{\theta_0, \theta_1 \in \Theta} |v^\top (\theta_0 - \theta_1)| \cdot \left(1 - 2\alpha - \left(\frac{(\theta_0 - \theta_1)^\top \mathbb{E}(S_n)(\theta_0 - \theta_1)}{4\sigma^2}\right)^{\frac{1}{2}}\right).$$

Let us complete the proof of Theorem 1(a) using this lemma. We do so by carefully choosing the pair of points $(\theta_0, \theta_1)$ and then using those values in Lemma 2. We choose

$$\theta_0 \in \mathbb{R}^d \quad \text{and} \quad \theta_1 = \theta_0 + \sigma \cdot \frac{\mathbb{E}(S_n)^{-1}v}{\|\mathbb{E}(S_n)^{-\frac{1}{2}}v\|_2} \quad (51)$$

where, $\mathbb{E}(S_n)^{-\frac{1}{2}}$ denotes a positive semidefinite matrix square-root of the matrix $\mathbb{E}(S_n)^{-1}$. Thus, we conclude

$$|I_{\alpha, v}(\theta)| \overset{(i)}{\geq} \sigma \cdot \left(v^\top \mathbb{E}(S_n)^{-1}v\right)^{\frac{1}{2}} \cdot (1 - 2\alpha - 1/2) \overset{(ii)}{\geq} (1/2 - 2\alpha) \cdot \sigma \cdot \left(v^\top \mathbb{E}(S_n)^{-1}v\right)^{\frac{1}{2}}.$$

The inequality (i) above follows by substituting the value of $\theta_0$ and $\theta_1$ in the Lemma 2; inequality (ii) uses the assumption $\alpha \in (0, 1/8)$. It remains to prove the Lemma 2.
5.2.3 Proof of Lemma 2

We claim that it suffices to prove that following two inequalities \(^{6}\)
\[
| \mathcal{I}_{\alpha, \nu}(\Theta) | \geqslant \sup_{\theta_0, \theta_1 \in \Theta} | \mathbf{v}^\top (\theta_0 - \theta_1) | (1 - \alpha - \text{TV}(\mathbb{P}_{\theta_0}(\mathcal{D}_n), \mathbb{P}_{\theta_1}(\mathcal{D}_n)))^+, \quad \text{and} \quad (52a)
\]
\[
\text{KL}(\mathbb{P}_{\theta_0}(\mathcal{D}_n), \mathbb{P}_{\theta_1}(\mathcal{D}_n)) = \frac{1}{2 \sigma^2} (\theta_0 - \theta_1)^\top \mathbb{E}(S_n) (\theta_0 - \theta_1). \quad (52b)
\]

Indeed, with the above two bounds at hand, the proof of Lemma 2 follows by invoking Pinsker’s inequality \(^{27}\):
\[
\text{TV}(\mathbb{P}_{\theta_0}(\mathcal{D}_n), \mathbb{P}_{\theta_1}(\mathcal{D}_n)) \leqslant \frac{1}{\sqrt{2}} \sqrt{\text{KL}(\mathbb{P}_{\theta_0}(\mathcal{D}_n), \mathbb{P}_{\theta_1}(\mathcal{D}_n))}.
\]

It remains to prove relations (52a) and (52b).

**Proof of the bound (52a):** We observe that for each \(\theta \in \Theta\), the interval \(\mathcal{I}_{\alpha, \nu} := [l(\mathcal{D}_n), u(\mathcal{D}_n)]\) is a valid \(1 - \alpha\) confidence interval for \(\mathbf{v}^\top \theta\). In particular, for any fixed pair of points \((\theta_0, \theta_1)\), we have
\[
\mathbb{P}_{\theta_0} \{ \mathbf{v}^\top \theta_0 \in [l(\mathcal{D}_n), u(\mathcal{D}_n)] \} \geqslant 1 - \alpha, \quad \text{and} \quad \mathbb{P}_{\theta_1} \{ \mathbf{v}^\top \theta_1 \in [l(\mathcal{D}_n), u(\mathcal{D}_n)] \} \geqslant 1 - \alpha.
\]

Using properties of the total variance distance \(\text{TV}(\mathbb{P}_{\theta_0}(\mathcal{D}_n), \mathbb{P}_{\theta_1}(\mathcal{D}_n))\) we have
\[
\left| \mathbb{P}_{\theta_0} \{ \mathbf{v}^\top \theta_1 \in [l(\mathcal{D}_n), u(\mathcal{D}_n)] \} - \mathbb{P}_{\theta_1} \{ \mathbf{v}^\top \theta_1 \in [l(\mathcal{D}_n), u(\mathcal{D}_n)] \} \right| \leqslant \text{TV}(\mathbb{P}_{\theta_0}(\mathcal{D}_n), \mathbb{P}_{\theta_1}(\mathcal{D}_n))
\]

Combining the last two inequalities we have
\[
\mathbb{P}_{\theta_0} \{ (\mathbf{v}^\top \theta_0, \mathbf{v}^\top \theta_1) \in [l(\mathcal{D}_n), u(\mathcal{D}_n)] \} \geqslant 1 - 2\alpha - \text{TV}(\mathbb{P}_{\theta_0}(\mathcal{D}_n), \mathbb{P}_{\theta_1}(\mathcal{D}_n)).
\]

Recall that \([l(\mathcal{D}_n), u(\mathcal{D}_n)]\) is an interval in \(\mathbb{R}\), and we have
\[
\mathbb{P}_{\theta_0} \{ u(\mathcal{D}_n) - l(\mathcal{D}_n) \geqslant | \mathbf{v}^\top \theta_0 - \mathbf{v}^\top \theta_1 | \} \geqslant 1 - 2\alpha - \text{TV}(\mathbb{P}_{\theta_0}(\mathcal{D}_n), \mathbb{P}_{\theta_1}(\mathcal{D}_n))
\]

Putting together the pieces and taking supremum over the pair \((\theta_0, \theta_1) \in \Theta\) yields the bound
\[
| \mathcal{I}_{\alpha, \nu}(\Theta) | \geqslant | \mathbf{v}^\top \theta_0 - \mathbf{v}^\top \theta_1 | (1 - 2\alpha - \text{TV}(\mathbb{P}_{\theta_0}(\mathcal{D}_n), \mathbb{P}_{\theta_1}(\mathcal{D}_n)))^+.
\]

**Proof of the relation (52b):** The proof of the bound (52b) is based on a *divergence decomposition lemma* which is well known in the bandits literature \(^{3}\). Throughout, we use the shorthand \(\mathcal{F}_i\) denote the \(\sigma\)-field generated by the data \(\{x_1, y_1, \ldots, x_i, y_i\}\) up to time \(i\). Let \(p_{\theta_0}(y_i \mid x_i)\) and \(p_{\theta_1}(y_i \mid x_i)\) denote the Radon-Nykodim derivatives of the conditional distribution \(y_i \mid x_i\) with respect to the Lebesgue measure \(^{7}\) \(\nu\) on \(\mathbb{R}\), under the \(\theta_0\) and \(\theta_1\) respectively. For \(i = 1, \ldots, n\), let \(p_{\theta_0}(x_i \mid \mathcal{F}_{i-1})\) and \(p_{\theta_1}(x_i \mid \mathcal{F}_{i-1})\), respectively, denote the Radon-Nykodim derivatives of the conditional distribution \(x_i \mid \mathcal{F}_{i-1}\) with respect to

---

\(^{6}\)We point out that the bound (52a) is not new in the literature. A similar inequality can be found in the paper \(^{6}\) (see Lemma 1), where the authors used it to provide lower bounds on the confidence intervals for high dimensional linear regression with an *i.i.d.* dataset. To the best of our knowledge, the application of the bound (52a) to a non-i.i.d. data is novel.

\(^{7}\)Recall that in the model (1) we have \(p_{\theta_0}(y_i \mid x_i) = N(\theta_0^\top x_i, \sigma^2)\) and \(p_{\theta_1}(y_i \mid x_i) = N(\theta_1^\top x_i, \sigma^2)\).
a dominating measure $\lambda$ under $\theta_0$ and $\theta_1$. Clearly, $\mathbb{P}_{\theta_0}(\mathcal{D}_n)$ is dominated by the product measure $\nu^\otimes n \times \lambda^\otimes n$, and we have

$$d\mathbb{P}_{\theta_0}(\mathcal{D}_n) = \prod_{i=1}^{n} p_{\theta_0}(y_i \mid x_i) \cdot p_{\theta_0}(x_i \mid \mathcal{F}_{i-1}) \cdot d\lambda(x_i) \cdot \nu(y_i)$$

where, the step above uses the fact that the distribution of $y_i$ is dependent on the history $\mathcal{F}_{i-1}$ through $x_i$ only. Similarly, we also have:

$$d\mathbb{P}_{\theta_1}(\mathcal{D}_n) = \prod_{i=1}^{n} p_{\theta_1}(y_i \mid x_i) \cdot p_{\theta_1}(x_i \mid \mathcal{F}_{i-1}) \cdot d\lambda(x_i) \cdot \nu(y_i)$$

Now, we assumed that the query algorithm $\psi_i : (\mathbb{R} \times \mathbb{R}^d)^{i-1} \to \mathbb{R}^d$ which generates $x_i$ is oblivious towards the true value of $\theta$; as a result, we have

$$p_{\theta_0}(x_i \mid \mathcal{F}_{i-1}) = p_{\theta_1}(x_i \mid \mathcal{F}_{i-1}) \quad \text{for all} \quad i = 1, \ldots, n.$$ 

With the above observation at hand, the KL-distance $\text{KL}(\mathbb{P}_{\theta_0}(\mathcal{D}_n), \mathbb{P}_{\theta_1}(\mathcal{D}_n))$ can be simplified as follows:

$$\text{KL}(\mathbb{P}_{\theta_0}(\mathcal{D}_n), \mathbb{P}_{\theta_1}(\mathcal{Z}_n)) =: \mathbb{E}_{\theta_0} \left[ \log \frac{d\mathbb{P}_{\theta_0}(\mathcal{D}_n)}{d\mathbb{P}_{\theta_1}(\mathcal{D}_n)} \right]$$

$$= \sum_{i=1}^{n} \mathbb{E}_{\theta_0} \left[ \log \frac{p_{\theta_0}(y_i \mid x_i)}{p_{\theta_1}(y_i \mid x_i)} \right]$$

(54)

Now,

$$\mathbb{E}_{\theta_0} \left[ \log \frac{p_{\theta_0}(y_i \mid x_i)}{p_{\theta_1}(y_i \mid x_i)} \right] = \mathbb{E}_{x_i} \left[ \mathbb{E}_{y_i \mid x_i} \log \frac{p_{\theta_0}(y_i \mid x_i)}{p_{\theta_1}(y_i \mid x_i)} \right]$$

$$= \mathbb{E}_{x_i} \mathbb{E}_{y_i \mid x_i} \left[ -\frac{1}{2\sigma^2} \cdot (y_i - x_i^\top \theta_0)^2 + \frac{1}{2\sigma^2} \cdot (y_i - x_i^\top \theta_1)^2 \right]$$

$$= \mathbb{E}_{x_i} \left[ -\frac{1}{2\sigma^2} \cdot (x_i^\top (\theta_0 - \theta_1))^2 \right]$$

$$= \mathbb{E}_{x_i} \left[ \frac{1}{2\sigma^2} \cdot (x_i^\top (\theta_0 - \theta_1))^2 \right]$$

The second equality above follows since under $\theta_0$ we have $y_i \mid x_i \sim \mathcal{N}(x_i^\top \theta_0, \sigma^2)$. Substituting the last simplification in the KL distance calculation (54) we have

$$\text{KL}(\mathbb{P}_{\theta_0}(\mathcal{D}_n), \mathbb{P}_{\theta_1}(\mathcal{Z}_n)) = \mathbb{E}_{x_i} \left[ \frac{1}{2\sigma^2} \cdot \|x_i(\theta_0 - \theta_1)\|^2 \right]$$

$$= \frac{1}{2\sigma^2} \cdot (\theta_0 - \theta_1)^\top \mathbb{E}(x_i)(\theta_0 - \theta_1).$$

(55)

This completes the proof of equation (52b).

6 Proofs of the Corollaries

We now turn to the proofs of our three corollaries, with Sections 6.1, 6.2, and 6.3 devoted to the proofs of the Corollaries 1, 2 and 3, respectively.
6.1 Proof of Corollary 1

In light of Theorem 1, it suffices to verify Assumptions (A1)–(A3). The assumptions stated in Corollary 1 ensure that the error sequence \( \{\epsilon_i\}_{i=1}^{\infty} \) satisfies Assumption (A1). The growth conditions in Assumption (A2) are satisfied due to the minimum arm-pull assumption (28). It remains to verify the three conditions in Assumption (A3).

Beginning with the asymptotic negligibility condition, we have

\[
\max_{i \in [n]} \frac{1}{\gamma_n} x_i \Gamma_i^{-1} x_i \leq \frac{1}{\gamma_n} \max_{i \in [n]} \| x_i^2 \| = \frac{1}{(\log n)^{1+\delta}} \to 0.
\]

The first inequality above uses the bound \( \Gamma_i^{-1} \preceq \frac{1}{\log(n)^{1+\delta}} \cdot I_d \) (see the definition (29)); the second equality uses \( \| x_i^2 \| = 1 \), and the final step follows by substituting \( \gamma_n = 1/(\log n)^{1+\delta} \).

Turning to the vanishing bias condition in (A3), we invoke the operator norm bound (25) on the matrix \( I_d - W_n X_n S_n^{-\frac{1}{2}} \) to find that

\[
\sqrt{\gamma_n \log \lambda_{\max}(S_n)} \cdot \| I_d - W_n X_n S_n^{-\frac{1}{2}} \|_{op} \leq \sqrt{\gamma_n \log n} \cdot O_p(1)
\]

\[
= \frac{1}{(\log n)^{\delta/2}} \cdot O_p(1) \to 0,
\]

where we have used the bound \( \lambda_{\max}(S_n) \leq \text{trace}(S_n) = n \) in the above derivation.

Finally, we verify the variance stability condition in (A3) with the help of the following lemma

**Lemma 3 (Commutative guarantee).** For any collection of matrices \( \{\Gamma_i^{-\frac{1}{2}} x_i \Gamma_i^{-\frac{1}{2}}\}_{i=1}^{n} \) that commute with each other, we have

\[
\| I - \sum_{i=1}^{n} w_i x_i \Gamma_i^{-\frac{1}{2}} \|_{op} \leq \exp \left( -\frac{\lambda_{\min}(\sum_{i=1}^{n} \Gamma_i^{-\frac{1}{2}} x_i x_i^\top \Gamma_i^{-\frac{1}{2}})}{\gamma_n} \right).
\]

See the end of this subsection for the proof of this claim.

Let us complete the proof of corollary 1 using lemma 3. In the multi-arm bandit setting of Corollary 1, the matrices \( \{\Gamma_i^{-\frac{1}{2}} x_i x_i^\top \Gamma_i^{-\frac{1}{2}}\}_{i=1}^{n} \) are all diagonal, and hence they commute. Thus, invoking the operator norm bound from Lemma 3 yields

\[
\| I_d - \sum_{i=1}^{n} w_i x_i \Gamma_i^{-\frac{1}{2}} \|_{op} \leq \exp \left( -\frac{\lambda_{\min}(\sum_{i=1}^{n} \Gamma_i^{-\frac{1}{2}} x_i x_i^\top \Gamma_i^{-\frac{1}{2}})}{\gamma_n} \right).
\]

Recall that in the bandits model (26), the matrices \( S_n \) and \( x_i x_i^\top \) are diagonal. By construction (29) and the minimum arm-pull condition (28), the tuning matrix \( \Gamma_i \) is also diagonal with diagonal entries upper bounded by the corresponding diagonal entries of the (diagonal) matrix \( S_n \). Combining these two observations we have that \( S_n^{-\frac{1}{2}} x_i x_i^\top S_n^{-\frac{1}{2}} \leq \Gamma_i^{-\frac{1}{2}} x_i x_i^\top \Gamma_i^{-\frac{1}{2}} \). Consequently, we find that

\[
-\lambda_{\min}(\sum_{i=1}^{n} \Gamma_i^{-\frac{1}{2}} x_i x_i^\top \Gamma_i^{-\frac{1}{2}}) \leq -\lambda_{\min}(\sum_{i=1}^{n} S_n^{-\frac{1}{2}} x_i x_i^\top S_n^{-\frac{1}{2}}) = -1,
\]

28
where the final equality follows from the definition of $S_n$. Substituting the value $\gamma_n = \frac{1}{(\log n)^{1+\delta}}$ yields

$$
\|I_d - \sum_{i=1}^{n} w_i x_i^\top \Gamma_i^{-\frac{1}{2}}\|_{op} \leq \exp(-1/\gamma_n) \leq \frac{1}{n} \to 0.
$$

This verifies the variance stability condition from Assumption (A3), and applying Theorem 1 yields Corollary 1.

The only remaining detail is to prove Lemma 3.

**Proof of Lemma 3** For notational convenience, we use the shorthands $z_i := x_i \Gamma_i^{-\frac{1}{2}}$ and $Z_i^\top := [z_1 \cdots z_i]$, as previously introduced in Section 2.2. Substituting the formula for the weight vector $w_i$ from eq. (7b), and performing some algebra yields

$$
(I - W_n Z_n)^\top (I - W_n Z_n) = \prod_{j=1}^{n} (I_d - \frac{Z_{n+1-j} Z_{n+1-j}^\top}{\gamma_n + \|Z_{n+1-j}\|^2}) \prod_{i=1}^{n} (I_d - \frac{z_i z_i^\top}{\gamma_n + \|z_i\|^2})
$$

$$
= \exp \left[ \sum_{j=1}^{n} \log \left( I_d - \frac{Z_{n+1-j} Z_{n+1-j}^\top}{\gamma_n + \|Z_{n+1-j}\|^2} \right) \right] \prod_{i=1}^{n} \log \left( I_d - \frac{z_i z_i^\top}{\gamma_n + \|z_i\|^2} \right)
$$

$$(i) \quad \leq \exp \left( -2 \cdot \sum_{i=1}^{n} \frac{z_i z_i^\top}{\gamma_n} \right),
$$

where step (i) above uses the fact that $\exp(\log(1 - a)) \leq \exp(-a)$ for any scalar $a < 1$ and that the matrices $\{z_i z_i^\top\}_{i \in [n]}$ commute. Via an inductive argument, it can be verified that the entries of the matrix $I_d - \frac{z_i z_i^\top}{\gamma_n + \|z_i\|^2}$ are all upper bounded by 1. Putting together the pieces, we conclude that the operator norm satisfies the bound

$$
\|I - W_n Z_n\|_{op} \leq \exp \left( -\frac{\lambda_{\min}(\sum_{i=1}^{n} z_i z_i^\top)}{\gamma_n} \right),
$$

as claimed.

### 6.2 Proof of Corollary 2

The proof of this claim is similar to that of Corollary 1; in particular, we need to verify Assumptions (A1)–(A3). Recall that the time series model (31) in Corollary 2 is a special case of the stochastic linear regression model (1) with $(x_i, y_i) \equiv (y_{i-1}, y_i)$; thus the covariance term based on the data $\{(x_i, y_i)\}_{i \in [n]}$ is given by $\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i^2$. Here we have used the convention $y_0 = 0$.

The moment condition (A1) is satisfied since the additive noise $\epsilon_i$ in the autoregressive model (31) is assumed to have a standard Gaussian distribution. Before we verify the remaining conditions, it is helpful to deduce a few bounds regarding the sample covariance
term $\sum_{i=1}^{n} y_{i-1}^2$. In particular, we show that for any $\theta^* \in (-1, 1]$, the sample covariance term satisfies the following relations

$$\sum_{i=1}^{n} y_{i-1}^2 \to \infty \quad \text{almost surely},$$

$$\log(\sum_{i=1}^{n} y_{i-1}^2) = O_p(\log n), \quad \text{and} \quad (\log n)^{1+\delta} \sum_{i=1}^{n} y_{i-1}^2 = O_p(\sum_{i=1}^{n} y_{i-1}^2),$$

where $\delta > 0$ is a fixed scalar (typically small). We prove these bounds at the end of this sub-section, but let us complete the proof of the Corollary using these bounds.

First, observe that the condition (A2) follows from the growth condition (56a), and the asymptotic negligibility condition in (A3) is satisfied by noting that

$$\max_{i=\lfloor n \rfloor} \frac{1}{\gamma_n} \cdot \frac{y_{i-1}^2}{\max\{ (\log n)^{1+2\delta} y_{i-1}^2, \sum_{j=1}^{i-1} y_{j}^2 \}} \leq (\log n)^{1+\delta} \to 0.$$

Next, in order to verify the vanishing bias condition in Assumption (A3), doing a calculation similar to Proposition 2 we find that (see the arguments leading up to bounds (63a)–(63b) and their proofs)

$$\left( \frac{\sqrt{n} \log(\sum_{i=1}^{n} y_{i-1}^2)}{1 - \sum_{i=1}^{n} \frac{w_i y_{i-1}}{\Gamma_i^2}} \right)^{(i)} = O_p((\log n)^{-\frac{\delta}{2}}) \cdot O_p\left( 1 + \sqrt{\frac{\Gamma_n}{\sum_{i=1}^{n} y_{i-1}^2}} \right),$$

$$\left( \frac{\sqrt{n} \log(\sum_{i=1}^{n} y_{i-1}^2)}{1 - \sum_{i=1}^{n} \frac{w_i y_{i-1}}{\Gamma_i^2}} \right)^{(ii)} = O_p((\log n)^{-\frac{\delta}{2}}) \cdot O_p(1) \to 0,$$

where step (i) follows by invoking the first part of the bound (56b) and step (ii) uses the second part of eq. (56b).

Finally, we verify the variance stability condition in (A3) with the help of Lemma 3, as previously stated and proved in the proof of Corollary 1. Note that in dimension $d = 1$, the commutativity condition in Lemma 3 holds trivially. Consequently, we may apply Lemma 3 to the one-dimensional autoregressive model (31) so as to obtain the bound

$$\left| 1 - \sum_{i=1}^{n} \frac{w_i y_{i-1}}{\Gamma_i^2} \right| \leq \frac{1}{n}.$$ 

See the calculations following the statement of Lemma 3 in the proof of Corollary 1 for details on this step.

This verifies the variance stability condition from Assumption (A3), and applying Theorem 1 yields Corollary 2.

The only remaining detail is to prove the bounds (56a)–(56b).

**Proofs of the bounds (56a)–(56b)** The proof of the first part of the bound (56b) follows by invoking Theorem 2 part (i) from the paper [20]. Concretely, in the paper [20], the authors showed that when $|\theta^*| \leq 1$, then there is some constant $a > 0$ such that $y_n = O_p(n^a)$. Thus, we have the relation $\sum_{i=1}^{n} y_{i-1}^2 = O_p(n^{2a+1})$, and first part of the bound (56b) follows.

We divide the proof of the remaining bounds into two parts, depending on the value of $\theta^*$. 30
Case 1  First, suppose that \( \theta^* = 1 \). Recall that in eq. (32) we argued that

\[
\frac{1}{n^2} \sum_{i=1}^{n} y_{i-1}^2 \overset{d}{\to} \int_0^1 w^2(t) dt.
\]

In light of the last relation, the growth condition (56a) is immediate. For the remaining bounds, note that \( y_n := \sum_{i=1}^{n} \varepsilon_{i-1} \sim \mathcal{N}(0, n-1) \); thus, for any \( \delta > 0 \), we deduce that

\[
\frac{1}{n^2} \cdot (\log n)^{1+\delta} y_n^2 \overset{P}{\to} 0,
\]

and we conclude that

\[
\log(n)^{1+\delta} \sum_{i=1}^{n} y_n^2 = O_p \left( \sum_{i=1}^{n} y_i^2 \right),
\]

as claimed.

Case 2  Otherwise, we may assume that \(|\theta^*| < 1\), in which case the term \( \sum_{i=1}^{n} y_{i-1}^2 \) stabilizes [20]; concretely, we have

\[
\frac{1}{n} \sum_{i=1}^{n} y_{i-1}^2 \overset{a.s.}{\to} c,
\]

where \( c > 0 \) is a non-random scalar.

The growth condition (56a) follows directly from the above relation. Moreover, we have

\[
y_{n-1} = \sum_{i=1}^{n} \theta^* \varepsilon_{n-1} \sim \mathcal{N} \left( 0, \frac{1}{1-\theta^2} \right).
\]

Putting these two pieces together yields \((\log n)^{1+\delta} \cdot y_{n-1}^2 = O_p \left( \sum_{i=1}^{n} y_{i-1}^2 \right)\), valid for any \( \delta > 0 \).

6.3 Proof of Corollary 3

We obtain the first claim of the Corollary 3 by applying Theorem 1, and the second part of the Corollary 3 follows from lemma 5. We prove these two parts separately.

Proof of claim (39a):  In order to apply Theorem 1 to the setup of Corollary 3 it suffices to verify the Assumption (A3). Recall that our choice of scaling \( \Gamma_i = \sum_{j=1}^{n} \varepsilon_j G \) matrices does not actually vary as a function of the round \( i \). For this reason, we simply write \( \Gamma \) from here onwards.

We begin by verifying the asymptotic negligibility condition in (A3). Observe that

\[
\mathbb{E} \left\{ \max_{i \in [n]} \frac{1}{\gamma_n} x_i^\top \Gamma^{-1} x_i \right\} \leq \frac{1}{\gamma_n} \cdot \mathbb{E} \left[ \max_{i \in [n]} \frac{\|x_i\|_2^2}{\lambda_{\min}(G) \sum_{i=1}^{n} \varepsilon_i} \right] \to 0,
\]

where the first inequality above follows by substituting the value of the scaling matrix \( \Gamma \), and the second step follows by invoking the sufficient exploration condition (38b).

Next, we verify the variance stability and vanishing bias conditions in (A3). In doing, we make use of the following auxiliary result:

Lemma 4. Under the sufficient exploration condition (38b), for any tuning parameter \( \gamma_n \in (0, 1/(\log Kn)^{1+\delta}) \) and a sufficient large sample size \( n \), we have

\[
\mathbb{E} \left[ \|I - \mathbb{W}_n X_n \Gamma^{-\frac{1}{2}} \|_F^2 \right] \leq \frac{d}{Kn}.
\]
See Section 6.3.1 for the proof of this lemma.

Taking lemma 4 as given, we now complete the proof of Corollary 3. Note that the variance stability condition in (A3) follows directly from the Frobenius norm bound in Lemma 4 and by letting the number of datapoints \( n \to \infty \), keeping the dimension \( d \) fixed.

In order to prove the vanishing bias condition in (A3), we first bound the operator norm of the matrix \( \Gamma - W_nX_nS_n^{-\frac{1}{2}} \):

\[
\| \Gamma - W_nX_nS_n^{-\frac{1}{2}} \|_{op} \leq 1 + \| W_nX_n \Gamma^{-\frac{1}{2}} \|_{op} \cdot \| \Gamma^{\frac{1}{2}} S_n^{-\frac{1}{2}} \|_{op} = O_p(1),
\]

where the derivation above uses the Frobenius norm upper bound from Lemma 4 and the fact that \( \| \Gamma^{\frac{1}{2}} S_n^{-\frac{1}{2}} \|_{op} = O_p(1) \) by the choice of the tuning parameter \( \Gamma \); see the bound (36) for instance. Using the last bound on \( \| \Gamma - W_nX_nS_n^{-\frac{1}{2}} \|_{op} \), we then find that

\[
\sqrt{\gamma_n} \log \lambda_{\max}(S_n) \cdot \| \Gamma - W_nX_nS_n^{-\frac{1}{2}} \|_{op} \leq \sqrt{\gamma_n} \log \lambda_{\max}(S_n) \cdot O_p(1) \overset{P}{\to} 0,
\]

where the last step above utilizes the choice \( \gamma_n = o_p(\log Kn) \) and that the maximum eigenvalue of \( S_n \) is upper bounded by \( Kn \); recall the uniform boundedness assumption (38a). All together, we have verified the assumptions of Theorem 1, so that Corollary 3 follows.

It remains to prove Lemma 4.

### 6.3.1 Proof of Lemma 4

Throughout this proof, we use the shorthands \( z_i = \Gamma^{-\frac{1}{2}}x_i, Z_i^\top = [z_1, z_2, \ldots, z_i], W_i = [w_1, \ldots, w_i] \) and \( \Delta_i := \Gamma - W_iZ_i \). Substituting the expression (7b) for the weight vector \( w_i \) we find that

\[
\| \Delta_{i-1} \|_F^2 - \| \Delta_i \|_F^2 = \frac{\gamma_n + \| z_i \|^2}{(\gamma_n/2 + \| z_i \|^2)^2} \text{trace} \{ \Delta_{i-1} z_i z_i^\top \Delta_{i-1}^\top \} \geq \frac{1}{\gamma_n^2 + \| z_i \|^2} \text{trace} \{ \Delta_{i-1} z_i z_i^\top \Delta_{i-1}^\top \}.
\]  

In equation (57), we proved that the random variable \( \frac{1}{\gamma_n} \max_{i \in [n]} \| z_i \|^2 \) converges to zero in probability; consequently, we may assume that

\[
\mathbb{P} \left[ \max_{i \in [n]} \| z_i \|^2 \leq \gamma_n/2 \right] \geq \frac{1}{2}
\]

for all sufficiently large values of the sample size \( n \). Keeping this in mind, taking expectations conditional on the sigma-field \( F_{i-1} \) on both sides in the inequality (59), and using the fact that \( \Delta_i \in F_{i-1} \), we have

\[
\mathbb{E} \left[ \| \Delta_{i-1} \|_F^2 \mid F_{i-1} \right] - \mathbb{E} \left[ \| \Delta_i \|_F^2 \mid F_{i-1} \right] \geq \frac{\varepsilon_i}{2\gamma_n} \sum_{i=1}^{n} \varepsilon_i \mathbb{E} \left[ \| \Delta_{i-1} \|_F^2 \mid F_{i-1} \right].
\]

Rearranging the last inequality and using the upper bound \( (1 - t) \leq \exp(-t) \) for \( t \geq 0 \) we obtain

\[
\mathbb{E} \left[ \| \Delta_i \|_F^2 \mid F_{i-1} \right] \leq \exp \left( \frac{-\varepsilon_i}{2\gamma_n \sum_{i=1}^{n} \varepsilon_i} \| \Delta_{i-1} \|_F^2 \right).
\]
Iterating the last bound $n$ times and removing the conditioning on the sigma field $\mathcal{F}_{i-1}$, we find that

$$
\mathbb{E} \{ \| \Delta_n \|^2_F \} \leq d \exp \left( -\frac{1}{2\gamma_n} \right) \leq d \exp \left( -\frac{1}{2} \cdot (\log Kn)^{1+\delta} \right) \leq \frac{d}{K^n}.
$$

Here step (i) follows by using $\gamma_n \leq \frac{1}{(\log Kn)^{1+\delta}}$, and step (ii) holds since the sample size is assumed $n$ to be sufficiently large. This completes the proof of the claim (39a).

**Proof of claim (39b):** In order to apply Lemma 5, it suffices to verify the vanishing bias condition (A3)'(b) for the choice of tuning parameters $\gamma_n$ and $\{\Gamma_i\}_{i=1}^n$ specified in Corollary 3. We have

$$
\| D_n^\frac{1}{2} S_n^{-\frac{1}{2}} - W_n X_n S_n^{-\frac{1}{2}} \|_{op} \leq \| D_n^\frac{1}{2} S_n^{-\frac{1}{2}} \|_{op} + \| W_n X_n S_n^{-\frac{1}{2}} \|_{op}
$$

$$
= \| D_n^\frac{1}{2} S_n^{-\frac{1}{2}} \|_{op} + O_p(1)
$$

$$
= O_p(\| D_n^\frac{1}{2} S_n^{-\frac{1}{2}} \|_{op})
$$

The second equation above uses the lower bound $\| D_n^\frac{1}{2} S_n^{-\frac{1}{2}} \|_{op} \geq 1$, which is valid by construction of the matrix $D_n$. Finally, invoking the bounded covariate assumption (38a) we have $\sqrt{\log \lambda_{\text{max}}(S_n)} \leq \sqrt{\log(Kn)}$, and using $0 < \gamma_n \leq \left( \log(Kn)^{(1+\delta)} \cdot O_p(\| D_n^\frac{1}{2} S_n^{-\frac{1}{2}} \|_{op}) \right)^{-1}$ we see that the condition (A3)'(b) holds.

7 Discussion

In this paper, we propose a family of online debiasing estimators for adaptive linear regression and analyze their asymptotic properties. We show that the online debiasing estimators admit a Gaussian limit under considerably weaker conditions than the OLS estimator and highlight practical examples from multi-armed bandits, time series modeling, and active learning in which online debiasing yields asymptotic normality while OLS does not. We also prove a minimax lower bound for the adaptive linear regression model and show that the performance of the online debiasing estimators is optimal up to a factor logarithmic in the sample size.

In future work, we would like to more precisely describe the non-asymptotic behavior of the online debiasing estimators; concretely, we would like to investigate the rate of distributional convergence of the online debiasing estimators to the appropriate Gaussian distributions. Finally, the performance of the online debiasing estimators matches the minimax optimal performance up to a logarithmic factor. An open question is whether this logarithmic gap can be closed either by providing a sharper minimax lower bound or by proposing better estimators.

A Proofs of the propositions

In this section, we provide the proofs of our two propositions. The section A.1 is devoted to the proof of Proposition 1, and the Section A.2 is devoted to the proof of Proposition 2.
A.1 Proof of Proposition 1

Let us recall the form of the estimator $\hat{\theta}_{v,\text{diagOD}}$ from definition (13)

$$
\hat{\theta}_{v,\text{diagOD}} := \hat{\theta}_{v,\text{LS}} + D_n^{1/2} \sum_{i=1}^{n} w_i (y_i - x_i^T V^T \hat{\theta}_{v,\text{LS}})
$$

Our proof of Proposition 1 is based on Lemma 5 which is a special case of Proposition 1 for $v = e_1$ — the first basis vector of the standard basis directions. Observe that in this case without loss of generality \(^8\) we can take $V = I$, and the estimator $\hat{\theta}_{v,\text{diagOD}}$ simplifies to the following

$$
\hat{\theta}_{\text{diagOD}} := \hat{\theta}_{\text{LS}} + D_n^{-1/2} \sum_{i=1}^{n} w_i (y_i - x_i^T \hat{\theta}_{\text{LS}}),
$$

where $D_n := \text{diag}(S_n^{-1})^{-1}$.

**Lemma 5.** Under Assumptions (A1),(A2), and (A3)', given any consistent estimator $\hat{\sigma}^2$ of $\sigma^2$, we have

$$
\sqrt{\frac{\gamma_m}{\sigma^2}} \cdot D_n^{1/2} (\hat{\theta}_{\text{diagOD}} - \theta^*) \xrightarrow{d} \mathcal{N}(0,I).
$$

(61a)

Consequently, for any $i = 1, \ldots d$, the following interval is an asymptotically exact $1 - \alpha$ confidence interval for $e_i^T \theta^*$

$$
\left[ e_i^T \hat{\theta}_{\text{diagOD}} - \frac{\hat{\sigma}}{\sqrt{\gamma_m}} (e_i^T D_n^{-1} e_i)^{1/2} z_{1-\alpha/2}, \ e_i^T \hat{\theta}_{\text{diagOD}} + \frac{\hat{\sigma}}{\sqrt{\gamma_m}} (e_i^T D_n^{-1} e_i)^{1/2} z_{1-\alpha/2} \right].
$$

(61b)

With the above Lemma as given, let us complete the proof of Proposition 1 for a general direction $v \in \mathbb{R}^d$ of interest. Let $\{v_1 = v, v_2, \ldots, v_d\}$ form an orthonormal basis of $\mathbb{R}^d$, and $V$ be the matrix with $v_j^T$ as its $j^{th}$ row. Note that we have $VV^T = I$ and $e_j^T V = v_j$ by construction. Using these two properties, we can write

$$
e^T_i V \theta^* = v^T \theta^* \quad \text{and} \quad y_i = \langle Vx_i, V \theta^* \rangle + \epsilon_i \quad \text{for all } i = 1, \ldots, n.
$$

Consequently, in this new basis, estimating the scalar $v^T \theta^*$ is same as estimating the first coordinate of transformed vector $V \theta^*$. Finally, by construction of the matrix $V$, we have

$$
e_j^T D_n^{-1} e_1 = v^T S_n^{-1} v_j,
$$

and invoking Lemma 5 we have that the following is an asymptotically accurate confidence interval for $v^T \theta^*$

$$
\left[ e_j^T \hat{\theta}_{v,\text{diagOD}} - \frac{\hat{\sigma}}{\sqrt{\gamma_m}} (v^T S_n^{-1} v)^{1/2} z_{1-\alpha/2}, \ e_j^T \hat{\theta}_{v,\text{diagOD}} + \frac{\hat{\sigma}}{\sqrt{\gamma_m}} (v^T S_n^{-1} v)^{1/2} z_{1-\alpha/2} \right].
$$

This completes the proof of Proposition 1. It remains to prove Lemma 5.

\(^8\)We point out that the choice of $V = I$ is not necessary in the proof, and the proof can be easily modified for any basis $V$ with it’s first row being equal to $e_1$. The benefit of choosing $V = I$ is that Lemma 2 allows us to simultaneously construct confidence intervals for all individual coordinates of $\theta^*$. 

34
Proof of Lemma 5

The proof of the claim (61a) is very similar to that of Theorem 1. Without loss of generality, we
assume that \( \sigma \) is known, and as a result, it suffices to prove \( \sqrt{n} \cdot D_n^{1/2} (\hat{\theta}_{\text{diagOD}} - \theta^*) \overset{d}{\to} \mathcal{N}(0, \sigma^2 I) \).

Observe that

\[
\sqrt{n} \cdot D_n^{1/2} (\hat{\theta}_{\text{diagOD}} - \theta^*) = \sqrt{n} \cdot \sum_{i=1}^n w_i e_i + \sqrt{n} \cdot \left( D_n^{1/2} S_n^{-1/2} - W_n X_n S_n^{-1/2} \right) S_n^{1/2} (\hat{\theta}_{\text{ls}} - \theta^*)
\]

\[
= v_n + b_n
\]

It remains to prove \( b_n \overset{p}{\to} 0 \) and \( v_n \overset{d}{\to} \mathcal{N}(0, \sigma^2 I) \). Observe that

\[
\|b_n\| \leq \sqrt{n} \cdot \|D_n^{1/2} S_n^{-1/2} - W_n X_n S_n^{-1/2}\|_{\text{op}} \cdot \|S_n^{1/2} (\hat{\theta}_{\text{ls}} - \theta^*)\| \\
\leq \sqrt{n} \log \lambda_{\text{max}}(S_n) \cdot \|D_n^{1/2} S_n^{-1/2} - W_n X_n S_n^{-1/2}\|_{\text{op}} \\
\overset{p}{\to} 0,
\]

where, the second inequality uses Theorem 1 from the paper [20], and the last step uses the vanishing bias condition (A3)\(^{(b)}\).

The analysis of the martingale term \( v_n \) is exactly same as that of Theorem 1 proof. This completes the proof of the claim (61a).

For the proof of claim (61b), observe that the matrix \( D_n \) is diagonal by construction; hence, for any coordinate direction \( e_i \), we have

\[
e_i^T D_n^{1/2} (\hat{\theta}_{\text{OD}} - \theta^*) = \sqrt{(D_n)_{ii}} \cdot ((\hat{\theta}_{\text{OD}})_i - \theta^*_i)
\]

\[
= \sqrt{\frac{1}{(D_n^{-1})_{ii}}} \cdot (e_i^T \hat{\theta}_{\text{OD}} - e_i^T \theta^*)
\]

\[
= \sqrt{\frac{1}{e_i^T S_n^{-1} e_i}} \cdot (e_i^T \hat{\theta}_{\text{OD}} - e_i^T \theta^*),
\]

where the last line follows from the relation \( e_i^T S_n^{-1} e_i = (D_n^{-1})_{ii} \). Thus, from property (9) we deduce

\[
\sqrt{\frac{n}{\sigma^2}} \cdot \sqrt{\frac{1}{e_i^T S_n^{-1} e_i}} \cdot (e_i^T \hat{\theta}_{\text{OD}} - e_i^T \theta^*) \overset{d}{\to} \mathcal{N}(0, 1). \tag{62}
\]

Define, the set \( A_{e_i, 1-\alpha} \subseteq \mathbb{R} \) as

\[
A_{e_i, 1-\alpha} := \left\{ \theta \in \mathbb{R} \mid -z_{1-\alpha/2} \leq \frac{\sqrt{n}}{\sigma^2} \cdot \sqrt{\frac{1}{e_i^T S_n^{-1} e_i}} \cdot (e_i^T \hat{\theta}_{\text{OD}} - \theta) \leq z_{1-\alpha/2} \right\}
\]

\[
= \left\{ e_i^T \hat{\theta}_{\text{OD}} - \frac{\sigma}{\sqrt{n}} (e_i^T S_n^{-1} e_i)^{1/2} z_{1-\alpha/2} \leq \frac{\sigma}{\sqrt{n}} (e_i^T S_n^{-1} e_i)^{1/2} z_{1-\alpha/2} \right\}
\]

From the equation (62) we have \( \lim_{n \to \infty} P(e_i^T \theta^* \in A_{e_i, 1-\alpha} = 1 - \alpha \), i.e., \( A_{e_i, 1-\alpha} \) is an asymptotically exact \( 1 - \alpha \) confidence intervals for \( e_i^T \theta^* \). This completes the proof of the claim (61b).
A.2 Proof of Proposition 2

Recalling that \( \|M\|_{\text{max}} := \max_{i,j} |M_{ij}| \) denotes the maximum absolute entry of a matrix, we claim that it suffices to show that \( \|W_nX_n\Gamma_n^{-\frac{1}{2}}\|_{\text{max}} \leq 4 \). Indeed, when this claim holds, we have

\[
\| I - W_nX_nS_n^{-\frac{1}{2}} \|_{\text{op}} \leq 1 + \| W_nX_n\Gamma_n^{-\frac{1}{2}} \|_{\text{op}} \| \Gamma_n^{-\frac{1}{2}}S_n^{-\frac{1}{2}} \|_{\text{op}} \\
\leq 1 + O_p(\sqrt{d}) \| W_nX_n\Gamma_n^{-\frac{1}{2}} \|_{\text{op}} \\
\leq 1 + O_p(d^2) \| W_nX_n\Gamma_n^{-\frac{1}{2}} \|_{\text{max}}
\]

The second last inequality above follows by noting that the diagonal entries of the matrix \( \Gamma_n^{-\frac{1}{2}}S_n^{-1}\Gamma_n^{-\frac{1}{2}} \) is of the order \( O_p(1) \); this bound uses the expression of the scaling matrix \( \Gamma_n \) from the definition (23a), and the operator-norm bound \( \| L_d^\frac{1}{2}\text{diag}(S_n^{-1})L_d^\frac{1}{2} \|_{\text{op}} = O_p(1) \) from assumption (22). The last inequality above follows from the fact that \( \| A \|_{\text{op}} \leq d^\frac{1}{2}\| A \|_{\text{max}} \), for any \( d \)-dimensional matrix \( A \). This completes the proof of Proposition 2. The remainder of the proof is devoted to establishing an upper-bound on the max-norm of the matrix \( W_nX_n\Gamma_n^{-\frac{1}{2}} \).

We do so by proving the following upper bounds

\[
\| \sum_{i=1}^{k} w_i x_i^\top \Gamma_i^{\frac{1}{2}} \|_{\text{max}} \leq 2 \quad \text{for} \quad k = 1, \ldots, n, \quad \text{and} \quad (63a)
\]

\[
\| \sum_{i=1}^{n} w_i x_i^\top \Gamma_i^{\frac{1}{2}} (I - \Gamma_i^{\frac{1}{2}}\Gamma_i^{-\frac{1}{2}}) \|_{\text{max}} \leq 2. \quad (63b)
\]

Note that a combination of these two bounds implies that \( \| W_nX_n\Gamma_n^{-\frac{1}{2}}\|_{\text{max}} \leq 4 \).

Accordingly, the remainder of our proof is devoted to establishing the bounds (63a) and (63b).

**Proof of bound (63a)** Using the expression for the weight vector \( w_i \) from equation (7b), we have

\[
I - \sum_{i=1}^{k} w_i x_i^\top \Gamma_i^{\frac{1}{2}} = \prod_{i=1}^{k} \left( I - \frac{\Gamma_i^{-\frac{1}{2}}x_i x_i^\top \Gamma_i^{-\frac{1}{2}}}{\gamma_n/2 + \| \Gamma_i^{-\frac{1}{2}}x_i \|^2} \right).
\]

Invoking the lower bound assumption (23b) and doing simple algebra, we find that

\[
\| I - \frac{\Gamma_i^{-\frac{1}{2}}x_i x_i^\top \Gamma_i^{-\frac{1}{2}}}{\gamma_n/2 + \| \Gamma_i^{-\frac{1}{2}}x_i \|^2} \|_{\text{op}} \leq 1 \quad \text{for all} \quad i \in [n].
\]

Consequently, for all \( k \in [n] \) we have the bound

\[
\| \sum_{i=1}^{k} w_i x_i^\top \Gamma_i^{\frac{1}{2}} \|_{\text{max}} \leq \| \sum_{i=1}^{k} w_i x_i^\top \Gamma_i^{\frac{1}{2}} \|_{\text{op}} \leq 2, \quad (64)
\]

where, in the last derivation we used the fact that the max-norm of a matrix is upper bounded by the operator norm of that matrix. This completes the proof of the bound (63a).
Proof of bound (63b) The proof is this bound exploits the following auxiliary lemma:

Lemma 6. Consider non-increasing nonnegative real numbers \( \{\delta_i\}_{i=1}^n \) and of real numbers \( \{a_i\}_{i=1}^n \) satisfying \( \max_{k \in [n]} |\sum_{i=1}^k a_i| \leq C \) for some constant \( C \). Then, we have

\[
|\sum_{i=1}^n a_i \delta_i| \leq C \delta_1. \tag{65}
\]

We prove this lemma at the end of this subsection.

Taking lemma 6 as given, let us prove the bound (63b). The bounds (63a) guarantee that

\[
\| \sum_{i=1}^k w_i x_i^T \Gamma_i^{-\frac{1}{2}} \|_{\max} \leq 2 \quad \text{for each } k \in [n].
\]

Moreover, by construction, the diagonal entries of the matrix \( (I - \Gamma_i^{-\frac{1}{2}} \Gamma_n^{-\frac{1}{2}}) \), for \( i = 1, \ldots, n \), are positive and non-increasing. Thus, we can apply Lemma 6 with the sequence \( \{a_i\}_{i=1}^n \) as the entries of the matrix \( w_i x_i^T \Gamma_i^{-\frac{1}{2}} \) and \( \delta_i \) as the diagonal entries of the (diagonal) matrix \( I - \Gamma_i^{\frac{1}{2}} \Gamma_n^{-\frac{1}{2}} \). Invoking Lemma 6 yields

\[
\| \sum_{i=1}^n w_i x_i^T \Gamma_i^{-\frac{1}{2}} (I - \Gamma_i^{\frac{1}{2}} \Gamma_n^{\frac{1}{2}}) \|_{\max} \leq 2 \cdot \| I - \Gamma_i^{\frac{1}{2}} \Gamma_n^{-\frac{1}{2}} \|_{\max} \leq 2,
\]

where, the last inequality above uses the property that the diagonal matrices \( \Gamma_1 \) and \( \Gamma_n \), by construction, satisfy a positive semidefinite ordering \( \Gamma_1 \leq \Gamma_n \). This concludes the proof of bound (63b).

It remains to prove the Lemma 6.

Proof of Lemma 6 Let \( s_k := \sum_{i=1}^k a_i \) denote the \( k^{th} \) partial sum of the sequence \( \{a_i\}_{i=1}^n \). The sum \( \sum_{i=1}^n a_i \delta_i \) can be represented in terms of these partial sums as

\[
|\sum_{i=1}^n a_i \delta_i| = |\sum_{i=1}^{n-1} q(\delta_{n-i} - \delta_{n-i+1})s_{n-i} + \delta_n s_n|
\]

\[
\leq C \cdot [\delta_n + \sum_{i=1}^{n-1} (\delta_{n-i} - \delta_{n-i+1})]
\]

\[
= C \delta_1,
\]

where inequality (i) uses the bound \( |s_{n-i}| \leq C \) and the ordering \( \delta_{n-i} \geq \delta_{n-i+1} \). This completes the proof of Lemma 6.

Proof of bound (25) Note that in the setting of multi-armed bandits (cf. Section 4.1), the covariance matrix \( S_n \) is diagonal, and consequently, the definition (23a) simplifies to \( \Gamma_i = \max\{S_i, L_n\} \). Moreover, a simple argument, using the method of induction on the
integer index \(i\), reveals that the matrix \(W_nZ_i\) is a diagonal matrix with nonnegative entries. In particular, we have \(\|W_nZ_n\|_{\text{max}} = \|W_nZ_n\|_{\text{op}}\). By combining these facts, we see that

\[
\|I - W_nX_nS_n^{-\frac{1}{2}}\|_{\text{op}} (\text{steps}) \\
\leq 1 + \|W_nX_n\Gamma_n^{-\frac{1}{2}}\|_{\text{op}} \cdot \max\{I, \|L_n^2S_n^{-\frac{1}{2}}L_n\|_{\text{max}}\} \leq 1 + \|W_nX_n\Gamma_n^{-\frac{1}{2}}\|_{\text{max}} \cdot O_p(1),
\]

where step (i) uses the fact that in multi-armed bandit problems the covariance matrix \(S_n\) is diagonal, and the matrix takes the form \(\Gamma_n = \max\{S_n, L_n\}\); and step (ii) follows from assumption (22) on the matrix \(L_n\) and the fact that max-norm equals the operator norm for diagonal matrices.

By combining the bounds (63a) and (63b), we see that \(\|W_nX_n\Gamma_n^{-\frac{1}{2}}\|_{\text{max}} \leq 4\). Combining this bound with inequality (66) yields

\[
\|I - W_nX_nS_n^{-\frac{1}{2}}\|_{\text{op}} = O_p(1),
\]
as claimed in the bound (25).

### B Proof of stability Lemma 1

For notational convenience, we use the shorthand notation

\[
z_i := x_i\Gamma_i^{-\frac{1}{2}}, \quad Z_i^\top = [z_1, \ldots, z_i], \quad \text{and} \quad W_i = [w_1, \ldots, w_i],
\]
as previously introduced in Section 2.2.

**Verifying the stability condition** The proof of the stability condition is based on a recursion relation that connects the terms \(\Delta_i := I - W_iZ_i\) and \(\Delta_{i-1} := I - W_{i-1}Z_{i-1}\). Substituting the expression (7b) for the vector \(w_i\) yields

\[
\Delta_i\Delta_i^\top = (\Delta_{i-1} - w_iZ_i^\top)(\Delta_{i-1} - w_iZ_i^\top)^\top \\
= \Delta_{i-1}\Delta_{i-1}^\top - \Delta_{i-1}(w_iZ_i^\top)^\top - w_iZ_i^\top\Delta_{i-1}^\top + w_iZ_i^\top z_iw_i^\top \\
= \Delta_{i-1}\Delta_{i-1}^\top - (\gamma_n + \|z_i\|^2)w_iw_i^\top,
\]

Summing the last recursion from \(i = 1\) to \(i = n\) and using the initial condition \(W_0 = 0\) yields

\[
I - \sum_{i=1}^n \gamma_n w_iw_i^\top = \sum_{i=1}^n \|z_i\|^2 w_iw_i^\top + (I - W_nZ_n)(I - W_nZ_n)^\top.
\]

Equipped with the last relation, it suffices to verify \(\|A_n\|_{\text{op}} \xrightarrow{P} 0\) and \(\|B_n\|_{\text{op}} \xrightarrow{P} 0\). We begin by observing that

\[
P[\|A_n\|_{\text{op}} > \epsilon] \leq P[\text{trace}(A_n) > \epsilon] \\
\leq P\left[\frac{\max_{i\in[n]} \|z_i\|^2}{\gamma_n} \sum_{i=1}^n \gamma_n \|w_i\|^2 > \epsilon\right].
\]

38
Now from eq. (68), we have the upper bound
\[
\sum_{i=1}^{n} \gamma_n \|w_i\|_2^2 = d - \text{trace}(A_n) - \text{trace}(B_n) \leq d.
\]
Thus, we have
\[
P[\|A_n\|_{\text{op}} > \epsilon] \leq P\left( \frac{\max_{i \in [n]} \|z_i\|_2^2}{\gamma_n} > \frac{\epsilon}{d} \right).
\]
Combined with the asymptotic negligibility assumption in (A3), this bound implies that \(\|A_n\|_{\text{op}} \xrightarrow{p} 0\), as desired.

On the other hand, using the operator-norm bound on the matrix \(I - W_nZ_n\) from the variance stability condition in (A3), we have
\[
\|B_n\|_{\text{op}} = \|I - W_nZ_n\|_{\text{op}}^2 \xrightarrow{p} 0.
\]
Putting together the pieces we conclude \(\gamma_n \sum_{i=1}^{n} w_i w_i^T \xrightarrow{p} I\) as claimed.

Verifying the vanishing norm condition Using the expression (7b) for the weight vector \(w_i\), we find that
\[
\|w_i\|_2^2 \leq \frac{1}{(\gamma_n/2 + \|z_i\|_2^2)^2} \cdot \|I - W_{i-1}Z_{i-1}\|_{\text{op}}^2 \|z_i\|_2^2.
\]
Doing a calculation similar to the derivation (67) we have \(\|\Delta_i\|_{\text{op}}^2 \leq \|\Delta_0\|_{\text{op}}^2 = 1\). Combining the last two observations with the asymptotic negligibility assumption \(\frac{1}{\gamma_n} \max_{i \in [n]} \|z_i\|_2^2 \xrightarrow{p} 0\) yields
\[
\gamma_n \max_{i \in [n]} \|w_i\|_2^2 \leq \frac{4}{\gamma_n} \cdot \max_{i \in [n]} \|z_i\|_2^2 \xrightarrow{p} 0,
\]
as claimed. This completes the proof of Lemma 1.

C Numerical experiment supplement

In this section, we present the results of additional experiments complementing those in section 4.

C.1 Multi-armed bandits:

In this section, we repeat the experiment of Section C.1 using covariates \(\{x_i\}_{i=1}^{n}\) generated by each of three popular multi-armed bandit algorithms:

(a) Thompson sampling algorithm [25]

(b) an \(\varepsilon\)-greedy algorithm [21]

(c) an upper confidence bound (UCB) strategy based on the paper [15]
We observe in Figures 5 and 6 that online debiasing provides appropriate coverage for all confidence levels, all bandit algorithms, and both parameters $\theta_{1}^{x}$ and $\theta_{2}^{x}$. Meanwhile, the OLS lower tail intervals severely undercover for all bandit algorithms and parameters, and W-decorrelation undercovers for several configurations despite having uniformly larger widths than online debiasing in all experiments. Finally, the concentration CI provides 100% coverage for all confidence levels but yields intervals uniformly larger than the online debiasing CIs.

C.2 Linear bandits

In this section, we repeat the experiment of section 4.3.1 with alternative settings of the ridge regression regularization parameter $\lambda_{\text{ridge}} \in \{1, 10\}$ for the concentration inequality CIs. Recall that given a dataset $\{x_i, y_i\}_{i=1}^{n}$ from the model (1), the ridge regression estimate $\hat{\theta}_{\text{ridge}}$ is defined as

$$
\hat{\theta}_{\text{ridge}} \in \arg \max_{\theta} \left\{ \sum_{i=1}^{n} (y_i - \mathbf{x}_i^T \theta)^2 + \lambda_{\text{ridge}} \cdot \|	heta\|_2^2 \right\}
$$

(69)

Here, $\lambda_{\text{ridge}} > 0$ is the regularization parameter for the ridge regression, and $\|	heta\|_2$ denotes the $\ell_2$ norm of the vector $\theta$. In Figure 7, we observe that the concentration based CIs always provide appropriate coverage but are uniformly larger than the online debiasing CIs for both $\lambda_{\text{ridge}} = 1$ and $\lambda_{\text{ridge}} = 10$ and for both parameters $\theta_{1}^{x}$ and $\theta_{2}^{x}$. 
Figure 5. Average coverage and width of confidence intervals for $\theta^*_1$ across 1000 independent replications of a multi-armed bandit experiment (26) with $\theta^* = (0.3, 0.3)$. The covariates $\{x_i\}_{i=1}^{1000}$ were selected using (a) Thompson sampling [25], (b) the $\varepsilon$-greedy algorithm [21], and (c) the upper confidence bound algorithm (UCB) [15]. The error bars represent $\pm 1$ standard error. Left and Center: Coverage of one-sided $1 - \alpha$ intervals for $\theta^*_1$. Right: Width of two-sided $1 - \alpha$ intervals for $\theta^*_1$. See Appendix C.1 for details.
Figure 6. Average coverage and width of confidence intervals for $\theta^*_2$ across 1000 independent replications of a multi-armed bandit experiment (26) with $\theta^* = (0.3, 0.3)^T$. The covariates $\{x_i\}_{i=1}^{1000}$ were selected using (a) Thompson sampling [25], (b) the $\epsilon$-greedy algorithm [21], and (c) the upper confidence bound algorithm (UCB) [15]. The error bars represent ±1 standard error. **Left** and **Center**: Coverage of one-sided $1 - \alpha$ intervals for $\theta^*_2$. **Right**: Width of two-sided $1 - \alpha$ intervals for $\theta^*_2$. See Appendix C.1 for details.
Figure 7. Average coverage and width of confidence intervals for $\theta_1^*$ and $\theta_2^*$ across 1000 independent replications of linear bandits experiment (26) with $\theta^* = (\theta_1^*, \theta_2^*) = (0.3, 0.3)^T$. The covariates $\{x_j\}_{j=1}^{1000}$ were selected using the $\varepsilon$-greedy linear bandits algorithm (37), and the error bars represent $\pm 1$ standard error. **Left** and **Center**: Coverage of one-sided $1 - \alpha$ intervals. **Right**: Width of two-sided $1 - \alpha$ intervals. See Appendix C.2 for details.
Acknowledgments

This work was partially supported by the Microsoft-Berkeley BAIR collaboration.

References

[1] Yasin Abbasi-Yadkori, David Pál, and Csaba Szepesvári. Online least squares estimation with self-normalized processes: An application to bandit problems. *arXiv preprint arXiv:1102.2670*, 2011.

[2] Karl J Åström. *Introduction to stochastic control theory*. Courier Corporation, 2012.

[3] P. Auer, N. Cesa-Bianchi, Y. Freund, and R.E. Schapire. Gambling in a rigged casino: The adversarial multi-armed bandit problem. In *Proceedings of IEEE 36th Annual Foundations of Computer Science*, pages 322–331, 1995. doi: 10.1109/SFCS.1995.492488.

[4] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford University Press, Oxford, UK, 2013.

[5] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[6] T Tony Cai, Zijian Guo, et al. Confidence intervals for high-dimensional linear regression: Minimax rates and adaptivity. *The Annals of statistics*, 45(2):615–646, 2017.

[7] Yash Deshpande, Lester Mackey, Vasilis Syrgkanis, and Matt Taddy. Accurate inference for adaptive linear models. In *International Conference on Machine Learning*, volume 35, pages 1194–1203, 2018.

[8] Yash Deshpande, Adel Javanmard, and Mohammad Mehrabi. Online debiasing for adaptively collected high-dimensional data. *arXiv preprint arXiv:1911.01040*, 2019.

[9] David A Dickey and Wayne A Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a): 427–431, 1979.

[10] Aryeh Dvoretzky et al. Asymptotic normality for sums of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Probability Theory*. The Regents of the University of California, 1972.

[11] Xavier Fontaine, Pierre Perrault, Michal Valko, and Vianney Perchet. Online a-optimal design and active linear regression. *arXiv preprint arXiv:1906.08509*, 2019.

[12] GC Goodwin and RL Payne. *Dynamic system identification experiment design and data analysis*. Mathematics in Science and Engineering, Volume 136. Elsevier, 1977.

[13] Vitor Hadad, David A Hirshberg, Ruohan Zhan, Stefan Wager, and Susan Athey. Confidence intervals for policy evaluation in adaptive experiments. *arXiv preprint arXiv:1911.02768*, 2019.

[14] Steven R Howard, Aaditya Ramdas, Jon McAuliffe, and Jasjeet Sekhon. Time-uniform, nonparametric, non-asymptotic confidence sequences. *The Annals of Statistics*, page To appear.
[15] Kevin Jamieson, Matthew Malloy, Robert Nowak, and Sébastien Bubeck. lil\text{ucb}: An optimal exploration algorithm for multi-armed bandits. In Conference on Learning Theory, volume 27, pages 423–439, 2014.

[16] Emilie Kaufmann and Wouter Koolen. Mixture martingales revisited with applications to sequential tests and confidence intervals. arXiv preprint arXiv:1811.11419, 2018.

[17] T L Lai and Herbert Robbins. Adaptive design and stochastic approximation. The Annals of Statistics, 7:1196–1221, 1979.

[18] T L etc Lai, Herbert Robbins, and C Zi Wei. Strong consistency of least squares estimates in multiple regression ii. Journal of Multivariate Analysis, 9(3):343–361, 1979.

[19] Tze Leung Lai. Asymptotic properties of nonlinear least squares estimates in stochastic regression models. The Annals of Statistics, 22(4):1917–1930, 1994.

[20] Tze Leung Lai, Ching Zong Wei, et al. Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. The Annals of Statistics, 10(1):154–166, 1982.

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

[22] Xinkun Nie, Xiaoying Tian, Jonathan Taylor, and James Zou. Why adaptively collected data have negative bias and how to correct for it. In International Conference on Artificial Intelligence and Statistics, volume 84, pages 1261–1269, 2018.

[23] Jaehyeok Shin, Aaditya Ramdas, and Alessandro Rinaldo. On the bias, risk and consistency of sample means in multi-armed bandits. arXiv preprint arXiv:1902.00746, 2019.

[24] Jaehyeok Shin, Aaditya Ramdas, and Alessandro Rinaldo. Are sample means in multi-armed bandits positively or negatively biased? In Advances in Neural Information Processing Systems, volume 32, pages 7102–7111, 2019.

[25] William R Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. Biometrika, 25(3/4):285–294, 1933.

[26] Sofía S Villar, Jack Bowden, and James Wason. Multi-armed bandit models for the optimal design of clinical trials: benefits and challenges. Statistical science: a review journal of the Institute of Mathematical Statistics, 30(2):199, 2015.

[27] Martin J Wainwright. High-dimensional statistics: A non-asymptotic viewpoint. Cambridge University Press, 2019.

[28] John S White. The limiting distribution of the serial correlation coefficient in the explosive case. The Annals of Mathematical Statistics, pages 1188–1197, 1958.

[29] Min Xu, Tao Qin, and Tie-Yan Liu. Estimation bias in multi-armed bandit algorithms for search advertising. In Advances in Neural Information Processing Systems, volume 26, pages 2400–2408, 2013.

[30] Kelly W Zhang, Lucas Janson, and Susan A Murphy. Inference for batched bandits. arXiv preprint arXiv:2002.03217, 2020.