Signal Recovery With Multistage Tests and Without Sparsity Constraints

Yiming Xing, Student Member, IEEE, and Georgios Fellouris, Member, IEEE

Abstract—A signal recovery problem is considered, where the same binary testing problem is posed over multiple, independent data streams. The goal is to identify all signals (resp. noises), i.e., streams with the alternative (resp. null) hypothesis is correct, subject to prescribed bounds on classical or generalized familywise error probabilities of both types. It is not required that the exact number of signals be a priori known, only upper bounds on the numbers of signals and noises are assumed instead. A decentralized formulation is adopted, according to which the sample size and the decision for each testing problem must be based only on observations from the corresponding data stream. A novel multistage testing procedure is proposed for this problem and is shown to enjoy a high-dimensional asymptotic optimality property. Specifically, it achieves the optimal, average over all streams, expected sample size, uniformly in the true number of signals, the maximum possible numbers of signals and noises go to infinity at arbitrary rates, in the class of all sequential tests with the same control of the type-I and type-II error probabilities, as the two error probabilities go to zero as long as they do not do so very asymmetrically. The first goal of the present work is to introduce a multistage test, to which we refer as the General Multistage Test, that is asymptotically optimal under both hypotheses as the two error probabilities go to zero at arbitrary rates. The proposed test adds, if necessary, to the 3-Stage Test in [11, Section 2], [12] opportunities either only to accept or only to reject the null hypothesis. The number of these opportunities is specified explicitly and depends on the relative magnitude of the two user-specified error probabilities.

The second goal of the present work is to apply the proposed multistage test to a high-dimensional signal recovery problem, where a large number of pairs of hypotheses are tested simultaneously and the problem is to correctly identify the data streams in which the alternative (resp. null) hypothesis holds, to which we refer as signals (resp. noises). This problem arises in various scientific and engineering applications, e.g., genetics [13], [14], [16], spectrum sensing in cognitive radio [17], [18], searching for regions of interest (ROI) in an image or other mediums [19], [20].

In the present work, we consider a formulation for the signal recovery problem that generalizes the one adopted in [21]. In the latter, (i) there are multiple, independent data streams, (ii) each of them generates iid data, and (iii) the same binary testing problem is posed for each of them, (iv) the same testing procedure must be applied to each testing problem, using data only from the corresponding data stream, (v) the misclassification probability, i.e., the probability of at least one error of either type, is controlled, and (vi) the exact numbers of signals and noises are assumed to be known a priori. In [21], a multistage test, termed Sequential Thresholding, is introduced and shown to achieve the optimal, average over all streams, expected sample size as the target misclassification probability remains fixed and the number of data streams goes to infinity, under a sparsity condition on the a priori known number of signals.

The signal recovery problem that we consider relaxes features (v) and (vi) in [21]. Specifically, we require control below distinct, user-specified levels of the probabilities of at least one type-I error and at least one type-II error (classical familywise error probabilities) or, more generally, of the...
probabilities of at least $\kappa$ type-I errors and at least $\iota$ type-II errors (generalized familywise error probabilities) \cite{22}, where $\kappa$ and $\iota$ are user-specified positive integers that can be larger than one. Moreover, we do not assume that the number of signals is known a priori, which is a rather unrealistic assumption, especially when the number of streams is large. Instead, we only require that the maximum possible numbers of signals and noises be specified.

In this more general setup, we formulate a novel asymptotic optimality criterion, according to which the optimal, average over all streams, expected sample size, in the class of all sequential tests with the same global error control, is achieved as the number of streams goes to infinity, \textit{uniformly in the true number of signals}. We show that the proposed multitest setup, \textit{i.e.}, the General Multistage Test, enjoys this property as long as the maximum possible numbers of signals and noises go to infinity. On the other hand, Sequential Thresholding (as well as a modification of this test that we introduce in this work) requires an additional sparsity condition on the maximum possible number of signals, whereas the 3-Stage Test in \cite[Section II, revisit the 3-Stage Test of \cite[Section 2]{11} and propose a modified version for it in Section V.}

In Section II, we consider the high-dimensional signal recovery problem in \cite[Section VIII]{} we discuss generalizations of the present work, \textit{i.e.}, the General Multistage Test, enjoys this property as long as the maximum possible numbers of signals and noises be specified.

The theoretical results in this work are supported by two numerical studies, one for the binary testing problem and one for the signal recovery problem. Finally, they are extended to setups with non-iid data or composite hypotheses. Indeed, all results in this work (apart from some that refer to the Sequential Thresholding), as well as their proofs, are shown to remain valid for many testing problems with neither independent nor identically distributed observations. Moreover, we show that for the one-sided testing problem for a one-parameter exponential family, the asymptotic optimality theory developed in this work applies to the optimal \textit{worst-case} expected sample size under each hypothesis.

The remainder of this work is organized as follows: in Sections II-V we focus on the binary testing problem, where we formulate the problem and present some preliminary results in Section II, revisit the 3-Stage Test of \cite[Section II]{} in Section III, introduce and analyze the proposed multitest in Section IV, and revisit the Sequential Thresholding in [21] and propose a modified version for it in Section V. In Section VI we consider the high-dimensional signal recovery problem. In Section VII we present the numerical studies, in Section VIII we discuss generalizations of the present work, and in Section IX we conclude and pose some open problems. The proofs of most results are presented in Appendices A-E.

We end this introductory section with some notations that we use throughout the paper. We denote by $\mathbb{N}$ the set of positive integers and by $\mathbb{R}$ the set of real numbers.

For a sequence of real numbers $\{x_n, n \in \mathbb{N}\}$ and $i, j \in \mathbb{N}$, we make the convention that $\sum_{k=i}^{j} x_k = 0$ if $i > j$. For any $x \in \mathbb{R}$, we denote by $\lfloor x \rfloor$ the smallest integer that is greater than or equal to $x$ and by $\lceil x \rceil$ the greatest integer that is less than or equal to $x$. For any $x, y \in \mathbb{R}$, we set $x \vee y = \max\{x, y\}$ and $x \wedge y = \min\{x, y\}$. For two sequences of positive numbers $\{x_n, n \in \mathbb{N}\}$ and $\{y_n, n \in \mathbb{N}\}$, $x_n \sim y_n$ stands for $\lim(x_n/y_n) = 0$, $x_n \gg y_n$ stands for $\lim(x_n/y_n) = \infty$, $x_n = O(y_n)$ means that there exists $C > 0$ so that $x_n \leq C y_n$ for all $n \in \mathbb{N}$, and $x_n = \Theta(y_n)$ means that there exists $C > 0$ so that $y_n/C \leq x_n \leq C y_n$ for all $n \in \mathbb{N}$. Finally, for a function $g : \mathbb{R} \to \mathbb{R}$ we denote by $g(x +)$ its right limit at $x$ and by $g(x -)$ its left limit at $x$, when they are well-defined.

II. \textsc{Binary Testing}

A. Problem Formulation

We let $X = \{X_n, n \in \mathbb{N}\}$ be a sequence of independent random elements with common density, $f$, with respect to a $\sigma$-finite measure, $\nu$, and consider the following simple versus simple hypothesis testing problem:

$$H_0 : f = f_0 \text{ versus } H_1 : f = f_1.$$ \hspace{1cm} (1)

The only assumption regarding $f_0$ and $f_1$ is that their Kullback-Leibler divergences are positive and finite, \textit{i.e.},

$$I_0 = \int \log \left( \frac{f_0}{f_1} \right) f_0 d\nu \in (0, \infty),$$

$$I_1 = \int \log \left( \frac{f_1}{f_0} \right) f_1 d\nu \in (0, \infty).$$ \hspace{1cm} (2)

We denote by $P$ (resp. $P_i$) the distribution of $X$ and by $E$ (resp. $E_i$) the corresponding expectation when the density is $f$ (resp. $f_i$), where $i \in \{0, 1\}$. For each $n \in \mathbb{N}$, we denote by $\mathcal{F}_n$ the $\sigma$-algebra generated by the first $n$ observations, \textit{i.e.}, $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$, by $\Lambda_n$ the corresponding log-likelihood ratio and by $\tilde{\Lambda}_n$ its average, \textit{i.e.,}

$$\Lambda_n \equiv \sum_{i=1}^{n} \log \left( \frac{f_1(X_i)}{f_0(X_i)} \right), \quad \tilde{\Lambda}_n \equiv \Lambda_n/n. \hspace{1cm} (3)$$

A sequential test, or simply, a test, for this hypothesis testing problem consists of a random time, $T$, that represents the number of observations until a decision is made, and a Bernoulli random variable, $D$, that represents this decision ($H_0$ is rejected if and only if $D = 1$). The determination at each time instant whether to stop sampling and, if so, which hypothesis to select can depend only on the already collected observations. Therefore, we say that $(T, D)$ is a test if

- $T$ is an $\{\mathcal{F}_n, n \in \mathbb{N}\}$-stopping time, \textit{i.e.}, $\{T = n\} \in \mathcal{F}_n$ for every $n \in \mathbb{N}$,

- and $D$ is an $\mathcal{F}_T$-measurable Bernoulli random variable, \textit{i.e.}, $\{T = n, D = i\} \in \mathcal{F}_n$ for every $n \in \mathbb{N}$ and $i \in \{0, 1\}$,

and we denote by $\mathcal{E}$ the family of all tests. For any $\alpha, \beta \in (0, 1)$, we denote by $\mathcal{E}(\alpha, \beta)$ the subfamily of tests whose type-I and type-II error probabilities do not exceed $\alpha$ and $\beta$, respectively, \textit{i.e.,}

$$\mathcal{E}(\alpha, \beta) \equiv \{(T, D) \in \mathcal{E} : P_0(D = 1) \leq \alpha \text{ and } P_1(D = 0) \leq \beta\}, \hspace{1cm} (4)$$

and by $\mathcal{L}_i(\alpha, \beta)$ the optimal expected sample size under $P_i$ in $\mathcal{E}(\alpha, \beta)$, \textit{i.e.,}

$$\mathcal{L}_i(\alpha, \beta) \equiv \inf_{(T, D) \in \mathcal{E}(\alpha, \beta)} E_i[T], \text{ where } i \in \{0, 1\}. \hspace{1cm} (5)$$
We refer to a test \((T, D)\) as fully sequential if its stopping time, \(T\), can take any value in \(\mathbb{N}\), as multistage if \(T\) can only take a small number of values, and as fixed-sample-size if \(T\) is predetermined. The first goal of the present work is to introduce a multistage test, the first in the literature to the best of our knowledge, that achieves both infima in (5) to a first-order asymptotic approximation as \(\alpha, \beta \to 0\) without any assumption on the decay rates of \(\alpha\) and \(\beta\).

**Definition 1:** A family of tests,
\[
\chi^* = \{(T^*(\alpha, \beta), D^*(\alpha, \beta)) \in \mathcal{E}(\alpha, \beta) : \alpha, \beta \in (0, 1)\},
\]
(6)
is asymptotically optimal under the null hypothesis if, as \(\alpha, \beta \to 0\),
\[
\mathbb{E}_0[T^*(\alpha, \beta)] \sim \mathcal{L}_0(\alpha, \beta),
\]
and is asymptotically optimal under the alternative hypothesis if, as \(\alpha, \beta \to 0\),
\[
\mathbb{E}_1[T^*(\alpha, \beta)] \sim \mathcal{L}_1(\alpha, \beta).
\]

**B. The Sequential Probability Ratio Test**

It is well known (see, e.g., [23, Chapter 3.2]) that both infima in (5) are achieved by a fully sequential test, the Sequential Probability Ratio Test (SPRT),
\[
\tilde{T} \equiv \inf\{n \in \mathbb{N} : \Lambda_n \notin (-B, A)\},
\]
\[
\tilde{D} \equiv 1\{\Lambda_T \geq A\},
\]
(7)
when \(A, B\) are selected as functions of \(\alpha\) and \(\beta\) so that the error constraints be satisfied with equality. In what follows, for any \(\alpha, \beta \in (0, 1)\) we denote by \((\tilde{T}(\alpha, \beta), \tilde{D}(\alpha, \beta))\) the test in (7) with
\[
A = |\log \alpha| \quad \text{and} \quad B = |\log \beta|.
\]

It is well known (see, e.g., [23, Chapter 3.1]) that for any \(\alpha, \beta \in (0, 1)\),
\[
(\tilde{T}(\alpha, \beta), \tilde{D}(\alpha, \beta)) \in \mathcal{E}(\alpha, \beta),
\]
and that, as \(\alpha, \beta \to 0\),
\[
\mathbb{E}_0[\tilde{T}(\alpha, \beta)] \sim \mathcal{L}_0(\alpha, \beta) \sim \frac{|\log \beta|}{I_0},
\]
\[
\mathbb{E}_1[\tilde{T}(\alpha, \beta)] \sim \mathcal{L}_1(\alpha, \beta) \sim \frac{|\log \alpha|}{I_1}.
\]
(8)
As a result, according to Definition 1, the family of SPRTs,
\[
\check{\chi} \equiv \{(\tilde{T}(\alpha, \beta), \tilde{D}(\alpha, \beta)) : \alpha, \beta \in (0, 1)\},
\]
(9)
is asymptotically optimal under both hypotheses.

**C. Asymptotic Optimality With Respect to a Mixture**

Generalizing the notation for \(P_i\) and \(\mathcal{L}_i\) with \(i \in \{0, 1\}\), for any \(\pi \in [0, 1]\) we introduce the mixture distribution
\[
P_\pi \equiv (1-\pi)P_0 + \pi P_1,
\]
(10)
and we denote by \(\mathcal{L}_\pi(\alpha, \beta)\) the optimal expected sample size in \(\mathcal{E}(\alpha, \beta)\) under \(P_\pi\), i.e.,
\[
\mathcal{L}_\pi(\alpha, \beta) \equiv \inf_{(T, D) \in \mathcal{E}(\alpha, \beta)} \mathbb{E}_\pi[T],
\]
(11)
where \(\mathbb{E}_\pi\) denotes the expectation under \(P_\pi\). It is easy to see that if a family of tests, \(\chi^*\), defined in (6), is asymptotically optimal under both hypotheses, then, as \(\alpha, \beta \to 0\),
\[
\mathbb{E}_\pi[T^*(\alpha, \beta)] \sim \mathcal{L}_\pi(\alpha, \beta) \sim (1-\pi)\frac{|\log \beta|}{I_0} + \pi \frac{|\log \alpha|}{I_1}
\]
uniformly in \(\pi \in [0, 1]\).
(12)
This will be useful for the formulation and analysis of the signal recovery problem in Section VI.

**D. The Fixed-Sample-Size Test**

The building block for all multistage tests considered in this work is the fixed-sample-size test (FSST) that rejects the null hypothesis if and only if the average log-likelihood ratio at a predetermined time instant exceeds a fixed threshold. For any \(\alpha, \beta \in (0, 1)\), we denote by \(n^*(\alpha, \beta)\) the smallest sample size such a test can have in order to belong to \(\mathcal{E}(\alpha, \beta)\), i.e.,
\[
n^*(\alpha, \beta) \equiv \min\{n \in \mathbb{N} : \exists c \in \mathbb{R} \text{ so that } P_0(\Lambda_n > c) \leq \alpha
\]
\[
\text{and } P_1(\Lambda_n \leq c) \leq \beta\},
\]
(13)
by \(c^*(\alpha, \beta)\) any real number that makes the two inequalities in (13) hold with \(n = n^*(\alpha, \beta)\), and we set
\[
\text{FSST}(\alpha, \beta) \equiv (n^*(\alpha, \beta), c^*(\alpha, \beta)).
\]
(14)
By (2) and the Weak Law of Large Numbers, it follows that, for any \(c \in (-I_0, I_1)\),
\[
P_0(\Lambda_n > c) \to 0 \quad \text{and} \quad P_1(\Lambda_n \leq c) \to 0 \quad \text{as } n \to \infty,
\]
thus, \(n^*(\alpha, \beta)\) is finite for any \(\alpha, \beta \in (0, 1)\).

**Remark 2:** For later use, we denote by \(c^*_0(n, \alpha)\) (resp. \(c^*_1(n, \beta)\)) the least conservative threshold for which the type-I (resp. type-II) error probability of the fixed-sample-size test with sample size \(n\) does not exceed \(\alpha\) (resp. \(\beta\)). Specifically, for any \(n \in \mathbb{N}\) and \(\alpha, \beta \in (0, 1)\) we set
\[
c^*_0(n, \alpha) \equiv \begin{cases} \sup \mathcal{C}_0(n, \beta), & \text{if } \mathcal{C}_0(n, \beta) \text{ is closed} \\ \sup \mathcal{C}_0(n, \beta) - \varepsilon, & \text{otherwise} \end{cases},
\]
\[
c^*_1(n, \alpha) \equiv \inf \mathcal{C}_1(n, \alpha),
\]
(15)
where \(\varepsilon\) is an arbitrarily small positive number and
\[
\mathcal{C}_0(n, \beta) \equiv \{c \in \mathbb{R} : P_1(\Lambda_n \leq c) \leq \beta\},
\]
\[
\mathcal{C}_1(n, \alpha) \equiv \{c \in \mathbb{R} : P_0(\Lambda_n > c) \leq \alpha\}.
\]
Indeed, the set \(\mathcal{C}_1(n, \alpha)\) is always of the form \([c, \infty)\) for some \(c \in \mathbb{R}\), whereas the set \(\mathcal{C}_0(n, \beta)\) can be either of the form
We will need a sharper upper bound, which is expressed in Chernoff information $n^*(\alpha, \beta)$ can be equivalently written as

$$n^*(\alpha, \beta) = \min \{ n \in \mathbb{N} : c_1^*(n, \alpha) \leq c_0^*(n, \beta) \},$$

and $c^*(\alpha, \beta)$ can be selected as any real number in $[c_1^*(n^*(\alpha, \beta), \alpha), c_0^*(n^*(\alpha, \beta), \beta)]$.

Next, we establish a non-asymptotic upper bound on $n^*(\alpha, \beta)$ that we use extensively in the analysis of the multistage tests that we consider in this work. By the Chernoff bound it follows that, for any $c \in (-I_0, I_1)$ and $n \in \mathbb{N}$,

$$P(\Lambda_n > c) \leq \exp(-n \psi_0(c)), \quad \forall \ c \geq -I_0,$$

$$P(\Lambda_n \leq c) \leq \exp(-n \psi_1(c)), \quad \forall \ c \leq I_1,$$  \hspace{1cm} (16)

where

$$\psi_0(c) \equiv \sup_{\theta \geq 0} \left\{ \theta c - \log \left( \int \theta f_{01}^1 f_{00}^{1-\theta} d\nu \right) \right\}, \quad c \geq -I_0,$$

$$\psi_1(c) \equiv \sup_{\theta \leq 0} \left\{ \theta c - \log \left( \int \theta f_{01}^1 f_{00}^{1-\theta} d\nu \right) \right\}, \quad c \leq I_1. \hspace{1cm} (17)$$

A well known (see, e.g., [24, Corollary 3.4.6]) upper bound on $n^*(\alpha, \beta)$ (see (22) below) can be obtained in terms of the Chernoff information:

$$C \equiv \sup_{\theta \geq 0} \left\{ -\log \left( \int \theta f_{01}^1 f_{00}^{1-\theta} d\nu \right) \right\} = \psi_0(0) = \psi_1(0). \hspace{1cm} (18)$$

We will need a sharper upper bound, which is expressed in terms of

$$h_i(\alpha, \beta) \equiv \psi_i \left( g^{-1} \left( \frac{\log \frac{\alpha}{\beta}}{\log \frac{\beta}{\alpha}} \right) \right), \quad i \in \{0, 1\}, \hspace{1cm} (19)$$

where $g^{-1} : (0, \infty) \to (-I_0, I_1)$ is the inverse of

$$g(c) \equiv \frac{\psi_0(c)}{\psi_1(c)}, \quad c \in (-I_0, I_1). \hspace{1cm} (20)$$

Indeed, since (see, e.g., [24, Chapter 2.2 & Chapter 3.4])

- $\psi_0$ (resp. $\psi_1$) is convex and continuous in $[-I_0, \infty)$ (resp. $(-\infty, I_1)$),
- $\psi_0$ is strictly increasing in $[-I_0, \infty)$ with $\psi_0(-I_0) = 0$, $\psi_0(I_1) = I_1$,
- $\psi_1$ is strictly decreasing in $(-\infty, I_1]$ with $\psi_1(-I_0) = I_0$, $\psi_1(I_1) = 0$,
- $g$ is continuous and strictly increasing in $(-I_0, I_1)$ with $g(-I_0) = 0$ and $g(I_1) = \infty$, and, as a result, the inverse $g^{-1} : (0, \infty) \to (-I_0, I_1)$ is well defined.

**Theorem 3:** For any $\alpha, \beta \in (0, 1)$,

$$n^*(\alpha, \beta) \leq \frac{\log \beta}{h_1(\alpha, \beta)} + 1 = \frac{\log \alpha}{h_0(\alpha, \beta)} + 1, \hspace{1cm} (21)$$

and, consequently,

$$n^*(\alpha, \beta) \leq \frac{\log(\alpha \wedge \beta)}{C} + 1. \hspace{1cm} (22)$$

**Proof:** See Appendix A. \hfill \blacksquare

Using (21), we next obtain a generalization of Stein’s lemma (see, e.g., [24, Lemma 3.4.7]), according to which, as $\alpha, \beta \to 0$ so that $|\log \alpha|/|\log \beta|$ goes to 0 (resp. infinity), the fixed-sample-size test is asymptotically optimal under the null (resp. alternative) hypothesis, at the expense of severe performance loss relative to the optimal under the alternative (resp. null) hypothesis.

**Corollary 4:** (i) If $\alpha, \beta \to 0$ so that $|\log \alpha|/|\log \beta|$, then $h_1(\alpha, \beta) \to I_0$ and

$$n^*(\alpha, \beta) \sim \mathcal{L}_0(\alpha, \beta) \gg \mathcal{L}_1(\alpha, \beta) \hspace{1cm} (23)$$

(ii) If $\alpha, \beta \to 0$ so that $|\log \alpha| \gg |\log \beta|$, then $h_0(\alpha, \beta) \to I_1$ and

$$n^*(\alpha, \beta) \sim \mathcal{L}_1(\alpha, \beta) \gg \mathcal{L}_0(\alpha, \beta). \hspace{1cm} (24)$$

**Proof:** See Appendix A. \hfill \blacksquare

### E. A Gaussian Example

We illustrate the above quantities in the case of testing the mean $\mu$ of a Gaussian distribution with unit variance, $N(\mu, 1)$, i.e.,

$$H_0 : \mu = -\eta \quad \text{versus} \quad H_1 : \mu = \eta \hspace{1cm} (25)$$

for some $\eta > 0$. In this case, the Kullback-Leibler divergences and the Chernoff information take the following form:

$$I_0 = I_1 = 2\eta^2 \equiv I, \quad C = \eta^2/2 = I/4,$$

and we have an explicit form for the fixed-sample-size test:

$$n^*(\alpha, \beta) = \left[ \frac{1}{4\eta^2} (z_\alpha + z_\beta)^2 \right],$$

$$c^*(\alpha, \beta) = (z_\alpha - z_\beta) \frac{\eta}{\sqrt{n^*(\alpha, \beta)}}, \hspace{1cm} (26)$$

where $z_\alpha$ is the upper $\alpha$-quantile of the standard Gaussian distribution, and for the quantities in (15):

$$c_0^*(n, \beta) = -2\eta \sqrt{n} + 2\eta^2,$$

$$c_1^*(n, \alpha) = z_\alpha \frac{2\eta}{\sqrt{n}} - 2\eta^2. \hspace{1cm} (27)$$

Moreover, the functions in (17) and (20) take the following form:

$$\psi_0(c) = \frac{1}{4I} (I + c)^2, \quad c \geq -I,$$

$$\psi_1(c) = \frac{1}{4I} (I - c)^2, \quad c \leq I,$$

$$g(c) = \left( \frac{I + c}{I - c} \right)^2, \quad c \in (-I, I). \hspace{1cm} (28)$$

Consequently,

$$\psi_0(g^{-1}(u)) = \frac{u}{(1 + \sqrt{u})^2 I}, \quad u \in \mathbb{R},$$

$$\psi_1(g^{-1}(u)) = (1 + \sqrt{u})^{-2} I, \quad u \in \mathbb{R},$$

and the functions in (19) take the following form:

$$h_0(\alpha, \beta) = I \cdot \left( 1 + \frac{|\log \beta|}{|\log \alpha|} \right)^{-2} \hspace{1cm},$$

$$h_1(\alpha, \beta) = I \cdot \left( 1 + \frac{|\log \alpha|}{|\log \beta|} \right)^{-2}. \hspace{1cm} (29)$$
Remark 5: In this Gaussian mean testing problem, FSST and \( c_0, c_1 \) can be evaluated using the formulas in (26) and (27), which require only the quantile function of the standard Gaussian distribution. In general, they can be computed, approximately, via simulation (see, e.g., [12, Section 6]).

III. 3-Stage Test

In this section, we review the test in [11, Section 2], which provides two opportunities to select \( H_0 \) and two opportunities to select \( H_1 \), and we propose a novel design for it.

A. Description

We denote by \( N \) the maximum possible number of observations and by \( N_i,0 \) the number of observations until the first opportunity to select \( H_i \), where \( N, N_i,0 \) are deterministic integers such that

\[
1 \leq N_i,0 \leq N, \quad i \in \{0, 1\}.
\]

Assuming it has not done so earlier, the test in [11, Section 2] terminates

- after \( N_{0,0} \) observations if \( \bar{\Lambda}_{N_{0,0}} \leq C_{0,0} \), in which case it accepts \( H_0 \),
- after \( N_{1,0} \) observations if \( \bar{\Lambda}_{N_{1,0}} > C_{1,0} \), in which case it rejects \( H_0 \),
- after \( N \) observations, rejecting \( H_0 \) if and only if \( \bar{\Lambda}_N > C \),

where \( C_{0,0}, C_{1,0}, C \) are real-valued thresholds to be specified together with \( N_{0,0}, N_{1,0}, N \). In what follows we refer to this generic test as 3-Stage Test (3ST), since at most 3 stages are needed for its implementation.

Remark 6: When \( N_{0,0} = N_{1,0} \) and \( C_{0,0} > C_{1,0} \), the first opportunity to accept and the first opportunity to reject the null hypothesis occur at the same time and the acceptance and rejection regions intersect. In this case, we make the convention that we first compare the average log-likelihood ratio with \( C_{0,0} \) and then with \( C_{1,0} \), i.e., we first check whether the null hypothesis can be accepted and then whether it can be rejected.

When \( N_i,0 = N \) for some \( i \in \{0, 1\} \), we adopt the convention that \( C_{i,0} \) is ignored and \( C \) is the only effective threshold.

B. Analysis

We continue with the design and asymptotic optimality result for the 3ST in [11, Section 2]. By an application of the union bound it follows that, for any \( \alpha, \beta \in (0, 1) \), 3ST belongs to \( \mathcal{E}(\alpha, \beta) \) when

\[
P_1(\bar{\Lambda}_{N_{0,0}} \leq C_{0,0}) \leq \beta/2, \quad P_0(\bar{\Lambda}_{N_{1,0}} > C_{1,0}) \leq \alpha/2, \quad (31)
\]

\[
P_1(\bar{\Lambda}_N \leq C) \leq \beta/2, \quad P_0(\bar{\Lambda}_N > C) \leq \alpha/2. \quad (32)
\]

Recalling the definition of FSST in (14), it is clear that condition (32) is satisfied when \( N, C \) are selected so that

\[
(N, C) = \text{FSST}(\alpha/2, \beta/2). \quad (33)
\]

By an application of Markov’s inequality, it follows that

\[
P_1(\bar{\Lambda}_{N_{0,0}} \leq C_{0,0}) \leq e^{-N_{0,0}C_{0,0}},
\]

\[
P_0(\bar{\Lambda}_{N_{1,0}} > C_{1,0}) \leq e^{-N_{1,0}C_{1,0}},
\]

thus, condition (31) is satisfied when \( C_{0,0} \) and \( C_{1,0} \) are selected as

\[
C_{0,0} = -\frac{|\log(\beta/2)|}{N_{0,0}}, \quad C_{1,0} = \frac{|\log(\alpha/2)|}{N_{1,0}}, \quad (34)
\]

for any choice of \( N_{0,0} \) and \( N_{1,0} \). The following theorem provides a generic choice for \( N_{0,1} \) and \( N_{1,1} \) that guarantees the asymptotic optimality of the 3ST under both hypotheses when \( \alpha \) and \( \beta \) go to 0 in a relatively symmetric fashion.

**Theorem 7** [11, Section 2]: For any \( \alpha, \beta \in (0, 1) \), let \( \tilde{T}(\alpha, \beta) \) denote the sample size and \( \tilde{D}(\alpha, \beta) \) the decision rule of 3ST when its parameters are selected according to (33)-(34) and

\[
N_{0,0} = \bar{N}_{0,0} \land N, \quad N_{1,0} = \bar{N}_{1,0} \land N,
\]

where \( \bar{N}_{0,0} = \left\lfloor \frac{|\log(\beta/2)|}{(1 - \epsilon_0) I_0} \right\rfloor \), \( \bar{N}_{1,0} = \left\lfloor \frac{|\log(\alpha/2)|}{(1 - \epsilon_1) I_1} \right\rfloor \),

and \( \epsilon_0, \epsilon_1 \) are functions of \( \alpha \) and \( \beta \) that take values in \((0, 1)\) and satisfy

\[
\epsilon_0 \to 0 \quad \text{and} \quad P_0 \left( \bar{\Lambda}_{N_{0,0}} > -(1 - \epsilon_0) I_0 \right) \to 0
\]

\[
\epsilon_1 \to 0 \quad \text{and} \quad P_1 \left( \bar{\Lambda}_{N_{1,0}} \leq (1 - \epsilon_1) I_1 \right) \to 0
\]

as \( \alpha, \beta \to 0 \). Then, the family

\[
\tilde{\chi} = \{(\tilde{T}(\alpha, \beta), \tilde{D}(\alpha, \beta)) : \alpha, \beta \in (0, 1)\}
\]

is asymptotically optimal under both hypotheses as \( \alpha, \beta \to 0 \) so that

\[
|\log(\alpha)| = \Theta(|\log(\beta)|).
\]

**Proof:** See Appendix B.

In the special case of the Gaussian mean testing problem of Subsection II-E, a second-order asymptotic upper bound on the expected sample size of 3ST under \( P_0 \) and \( P_1 \) respectively is stated in [11, Section 2].

**Proposition 8** [11, Section 2]: Consider the Gaussian mean testing problem of Subsection II-E and, for any \( \alpha, \beta \in (0, 1) \), let \( \tilde{T}(\alpha, \beta) \) be defined as in Theorem 7. If \( \epsilon_0 \) and \( \epsilon_1 \) satisfy

\[
\epsilon_0 = \Theta \left( \sqrt{\frac{\log(\log(\beta))}{|\log(\beta)|}} \right) \quad \text{and} \quad \epsilon_1 = \Theta \left( \sqrt{\frac{\log(\log(\alpha))}{|\log(\alpha)|}} \right)
\]

as \( \alpha, \beta \to 0 \), then

\[
E_0[\tilde{T}(\alpha, \beta)] \leq \frac{|\log(\beta)|}{T} \left( 1 + O \left( \sqrt{\frac{\log(\log(\beta))}{|\log(\beta)|}} \right) \right),
\]

\[
E_1[\tilde{T}(\alpha, \beta)] \leq \frac{|\log(\alpha)|}{T} \left( 1 + O \left( \sqrt{\frac{\log(\log(\alpha))}{|\log(\alpha)|}} \right) \right),
\]

as \( \alpha, \beta \to 0 \) so that (37) holds.

**Proof:** See Appendix B.

As can be seen in the proof of Theorem 7, by assumption (2) and the Weak Law of Large Numbers, it is always possible to find \( \epsilon_0 \) and \( \epsilon_1 \) that satisfy (36). However, neither (36) nor (38) provides a concrete selection for \( \epsilon_0 \) and \( \epsilon_1 \), and consequently
for $N_{0,0}$ and $N_{1,0}$. In the next subsection, we propose a concrete selection for these parameters that guarantees less conservative error control and smaller expected sample size under both hypotheses.

C. A Novel Design for the 3-Stage Test

First, we select thresholds $C_{0,0}$ and $C_{1,0}$ that satisfy (31) in a less conservative way than (34). Specifically, for any $\alpha, \beta \in (0, 1)$ and any selection of $N_{0,0}$ (resp. $N_{1,0}$), we select $C_{0,0}$ (resp. $C_{1,0}$) as the largest (resp. smallest) threshold for which the first (resp. second) inequality in (31) is satisfied, i.e., for which the type-II (resp. -I) error probability of the fixed-sample-size test does not exceed $\beta/2$ (resp. $\alpha/2$) when the sample size is $N_{0,0}$ (resp. $N_{1,0}$). Recalling the definition of functions $c_0^*$ and $c_1^*$ in (15), this means selecting $C_{0,0}$ and $C_{1,0}$ in the following way:

$$C_{0,0} = c_0^*(N_{0,0}, \beta/2), \quad C_{1,0} = c_1^*(N_{1,0}, \alpha/2).$$

Second, for any choice of $C_{0,0}$ (resp. $C_{1,0}$) in terms of $N_{0,0}$ (resp. $N_{1,0}$) that satisfies the first (resp. second) inequality in (31), e.g., (34) or (40), we can select $N_{0,0}$ (resp. $N_{1,0}$) in order to minimize, at least approximately, the expected sample size of 3ST under $P_0$ (resp. $P_1$). The upper bounds in the following proposition are useful for this purpose.

Proposition 9: Let $T$ denote the sample size of the 3ST. If its parameters satisfy (31), then

$$N_{0,0}(1 - \alpha/2) + (N - N_{0,0})(P(\Lambda_{N_{0,0}} > C_{0,0}) - \alpha/2) \leq E_0[\bar{T}] \leq N_{0,0} + (N - N_{0,0})P(\Lambda_{N_{0,0}} > C_{0,0}),$$

and

$$N_{1,0}(1 - \beta/2) + (N - N_{1,0})(P(\Lambda_{N_{1,0}} \leq C_{1,0}) - \beta/2) \leq E_1[\bar{T}] \leq N_{1,0} + (N - N_{1,0})P(\Lambda_{N_{1,0}} \leq C_{1,0}),$$

Proof: See Appendix B.

Remark 10: As can be seen in the proof of this proposition, condition (31) is needed only to establish the lower bounds in (41) and (42), i.e., the upper bounds hold for any selection of parameters.

For any $\alpha, \beta \in (0, 1)$, the difference between the upper and the lower bound in (41) and (42) does not exceed $N \cdot (\alpha \vee \beta)/2$, and from the upper bound in (22) it follows that the latter does not exceed

$$\left(\frac{\log((\alpha \vee \beta)/2)}{C} + 1\right) \cdot \frac{\alpha \vee \beta}{2},$$

when $N$ is selected according to (33). Therefore, at least when $\alpha$ and $\beta$ are not very asymmetric, the upper bounds in Proposition 9 are accurate approximations to the actual expected sample sizes of 3ST under $P_0$ and $P_1$. Thus, for any selection of $C_{0,0}$ (resp. $C_{1,0}$) in terms of $N_{0,0}$ (resp. $N_{1,0}$) that satisfies the first (resp. second) inequality in (31), e.g., (34) or (40), we propose selecting $N_{0,0}$ (resp. $N_{1,0}$) as the integer in $\{1, \ldots, N\}$ that minimizes the upper bound in (41) (resp. (42)).

While these modifications to the design of the 3ST in [11, Section 2] lead to less conservative error control and smaller expected sample size under both $P_0$ and $P_1$, they do not imply that the resulting test can achieve asymptotic optimality under $P_0$ and $P_1$ without a symmetry constraint on the decay rates of $\alpha$ and $\beta$, such as (37).

IV. THE GENERAL MULTISTAGE TEST

In this section, we introduce and analyze the multistage test that we propose in this work, which generalizes the 3-Stage Test of the previous section in that it can provide additional opportunities to reject or accept the null hypothesis.

A. Description

As in the previous section, we denote by $N$ the maximum possible number of observations, by $N_{i,0}$ the number of observations until the first opportunity to select $H_i$, where

$$1 \leq N_{i,0} \leq N, \quad i \in \{0, 1\},$$

and by $C, C_{0,0}, i \in \{0, 1\}$ the corresponding thresholds. To describe the proposed test, for each $i \in \{0, 1\}$ we now also need to determine a strictly increasing sequence of positive integers, $\{N_{i,j}, j \geq 1\}$, and a non-negative integer, $K_i$, so that

$$K_i \geq 1, \quad N_{i,0} \leq N_{i,1} < \cdots < N_{i,K_i} < N.$$  

For each $i \in \{0, 1\}$, $K_i$ is the additional (relative to the 3-Stage Test) number of opportunities to select $H_i$, and, when $K_i \geq 1, N_{i,j}$ is the total number of observations until the $(j + 1)^{th}$ additional opportunity to select $H_i$, where $1 \leq j \leq K_i$. Specifically, assuming it has not done so earlier, the proposed test terminates

- after $N_{0,j}$ observations if $\bar{\Lambda}_{N_{j,0}} \leq C_{0,j}$, in which case it accepts $H_0$, where $0 \leq j \leq K_0$,
- after $N_{1,j}$ observations if $\bar{\Lambda}_{N_{j,1}} > C_{1,j}$, in which case it rejects $H_0$, where $0 \leq j \leq K_1$,
- after $N$ observations, rejecting $H_0$ if and only if $\bar{\Lambda}_N > C$, where $C_{1,j}, 1 \leq j \leq K_i, i \in \{0, 1\}$ are additional thresholds to be specified.

This testing procedure can be implemented with at most $3 + K_0 + K_1$ stages and it reduces to the 3-Stage Test of the previous section when $K_0 = K_1 = 0$. Due to its general structure when compared to the 3-Stage Test or the 4-Stage Test in [12], we refer to it as the General Multistage Test (GMT).

Remark 11: For each $i \in \{0, 1\}$, only the first $K_i$ terms of the sequence $\{N_{i,j}, j \geq 1\}$ are used in the implementation of the testing procedure. This over-parametrization has the advantage that it provides a natural upper bound for $K_0$ and $K_1$ in terms of the other parameters. Specifically, by (44) it follows

$$K_i \leq \max\{j \geq 0 : N_{i,j} < N\}, \quad i \in \{0, 1\}.$$  

When, in particular, the terms of the two sequences are selected to be equidistant, i.e.,

$$N_{i,j} = j N_{i,1}, \quad j \in \mathbb{N}, i \in \{0, 1\},$$

(45) takes the following form:

$$K_i \leq \lceil N/N_{i,1}\rceil - 1, \quad i \in \{0, 1\}.$$  

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
Remark 12: By the description of the test it follows that when $N_{i,j} = N_{i,k}$ for some $0 \leq j, k \leq K_i$, only the maximum (resp. minimum) of the corresponding thresholds, $C_{i,j}$ and $C_{i,k}$, is effective when $i = 0$ (resp. $i = 1$).

Moreover, we adopt the convention that when an opportunity to accept and an opportunity to reject the null hypothesis occur at the same time and the acceptance and rejection regions intersect, i.e., $N_{0,j} = N_{1,k}$ and $C_{0,j} > C_{1,k}$ for some $0 \leq j \leq K_0$ and $0 \leq k \leq K_1$, we first compare the average log-likelihood ratio with $C_{0,j}$ and then with $C_{1,k}$, i.e., we first check whether the null hypothesis can be accepted and then whether it can be rejected.

Finally, if $N_{i,j} = N$ for some $0 \leq j \leq K_i$ and $i \in \{0, 1\}$, then we adopt the convention that $C_{i,j}$ is ignored and $C$ is the only effective threshold.

Remark 13: In Algorithm 1, we provide an algorithmic description of the GMT. In it, we denote by $\{N_j : 1 \leq j \leq 2 + K_0 + K_1\}$ the increasingly ordered version of the integers in $\{N_{i,j} : 0 \leq j \leq K_i, i \in \{0, 1\}\}$ and by $\{C_j : 1 \leq j \leq 2 + K_0 + K_1\}$ the corresponding thresholds. Moreover, for each $1 \leq j \leq 2 + K_0 + K_1$, we set $L_j$ equal to 0 (resp. 1) if $(N_j, C_j)$ corresponds to an opportunity to select $H_0$ (resp. $H_1$). We stress, however, that these notations are used only in Algorithm 1.

Algorithm 1 General Multistage Test (GMT)

Input: $K_0, K_1; \{N_j, C_j, L_j\}, 1 \leq j \leq 2 + K_0 + K_1; (N, C)$.

Initialize: $N_0 = 0, j = 1$.

while $j \leq 2 + K_0 + K_1$ do

\begin{itemize}
  \item if $L_j = 0$ and $N_j \leq C_j$ then stop and accept the null
  \item if $L_j = 1$ and $N_j > C_j$ then stop and reject the null
  \item $L_j = 1$, $j = j + 1$
\end{itemize}

end while

if $j = 3 + K_0 + K_1$ then

\begin{itemize}
  \item if $\Lambda N \leq C$ then stop and accept the null
  \item stop and reject the null
\end{itemize}

end if

Remark 1: By (22) it follows that, for any $\alpha, \beta \in (0, 1)$, the sample size of GMT cannot exceed

$$\frac{\log(\alpha \land \beta) + \log 2}{C} + 1$$

if $N$ is selected as in (52), independently of how the other parameters are selected. On the other hand, while the SPRT minimizes the expected sample size under $P_0$ and $P_1$, its expected sample size may be very large when the distribution of $X$ is different from $P_0$ and $P_1$ and $\alpha, \beta$ are small (see, e.g., [23, Chapter 3.1.1.2]). Indeed, if $P$ is a distribution under which $\{\Lambda_n, n \in \mathbb{N}\}$ is a random walk whose increments have zero mean and finite variance $\sigma^2$, then the expected sample size

$$P_1(\tilde{\Lambda}_N \leq C) \leq \beta/2 \quad \text{and} \quad P_0(\tilde{\Lambda}_N > C) \leq \alpha/2,$$
of $\hat{T}(\alpha, \beta)$, defined in (9), is equal to $|\log \alpha| \cdot |\log \beta| / \sigma^2$ if we ignore the overshoot over the boundaries. Therefore, GMT will perform much better than SPRT under such a distribution when $\alpha$ and $\beta$ are small (see Figure 1 for an illustration of this phenomenon).

C. Specification of the Free Parameters

We next specify the free parameters in (51). We start by assuming that $\{N_{i,j}, j \in \mathbb{N}\}, K_i, i \in \{0,1\}$ have already been selected in some way and the only remaining free parameters are $N_{0,0}$ and $N_{1,0}$, i.e., the number of observations at the first opportunities to accept and reject $H_0$, respectively. Similarly to the proposed design for the 3ST, we then select $N_{i,0}$ to minimize a non-asymptotic upper bound on the expected sample size of GMT under $P_i$, where $i \in \{0,1\}$.

**Proposition 16:** Let $\hat{T}$ denote the sample size of the GMT.
For any selection of its parameters, we have:

$$E_0[\hat{T}] \leq N_{0,0} + \sum_{j=1}^{K_0} (N_{0,j} - N_{0,j-1}) P_0 (\hat{\Lambda}_{N_{0,j-1}} > C_{0,j-1}) + (N - N_{0,K_0}) P_0 (\hat{\Lambda}_{N_{0,K_0}} > C_{0,K_0}),$$

(55)

$$E_1[\hat{T}] \leq N_{1,0} + \sum_{j=1}^{K_1} (N_{1,j} - N_{1,j-1}) P_1 (\hat{\Lambda}_{N_{1,j-1}} \leq C_{1,j-1}) + (N - N_{1,K_1}) P_1 (\hat{\Lambda}_{N_{1,K_1}} \leq C_{1,K_1}).$$

(56)

**Proof:** To see (55) (resp. (56)), it suffices to consider only the opportunities to select $H_0$ (resp. $H_1$) and ignore all opportunities to select $H_1$ (resp. $H_0$).


Thus, once all other free parameters have been determined, we propose selecting $N_{0,0}$ as the minimizer of (55) over $\{1, \ldots, N\}$ if $K_0 = 0$
and over $\{1, \ldots, N_{0,1}\}$ if $K_0 \geq 1$,

$N_{1,0}$ as the minimizer of (56) over $\{1, \ldots, N\}$ if $K_1 = 0$
and over $\{1, \ldots, N_{1,1}\}$ if $K_1 \geq 1$.

We continue with the specification, for each $i \in \{0,1\}$, of the first term in $\{N_{i,j}, j \in \mathbb{N}\}$, i.e., $N_{i,1}$, which is by definition an upper bound for $N_{i,0}$ when $K_i > 0$. As we saw in the previous section, in the case of the 3-Stage Test, $N_{i,0}$ should ideally be selected approximately equal to the optimal expected sample size under $P_i$, $L_i(\alpha, \beta)$, for both $i \in \{0,1\}$ (recall Lorden’s design in (35)). For this to be possible in GMT when $K_i \geq 1$ and $N_{i,0}$ is selected according to the above proposal, the upper bound $N_{i,1}$ for $N_{i,0}$ must be larger than $L_i(\alpha, \beta)$ for both $i \in \{0,1\}$. In view of the asymptotic approximations to the optimal expected sample size in (8) and the fact that the Chernoff information, $C$, defined in (18), is smaller than both $I_0$ and $I_1$, a sufficient choice for this purpose is

$$N_{0,1} = \left\lceil \frac{\log(\beta/5)}{C} \right\rceil$$
and
$$N_{1,1} = \left\lceil \frac{\log(\alpha/5)}{C} \right\rceil.$$

(57)

If we also select the terms of the sequence $\{N_{i,j}, j \in \mathbb{N}\}$ to be equidistant for each $i \in \{0,1\}$, i.e., to satisfy (46), then we have specified all free parameters in (51) apart from $K_0$ and $K_1$, for which from (47), (52), and (57) we have the following upper bounds:

$$K_0 \leq K_0 = \left\lceil \frac{n^*(\alpha/2, \beta/2)}{\|\log(\beta/5)/C\|} \right\rceil - 1,$$

$$K_1 \leq K_1 = \left\lceil \frac{n^*(\alpha/2, \beta/2)}{\|\log(\alpha/5)/C\|} \right\rceil - 1.$$

(58)

To sum up, for any choice of $K_0$ and $K_1$ that satisfies (58), we have the following explicit specification, in terms of $\alpha$ and $\beta$, for the remaining free parameters of GMT in (51):

$$N_{0,j} = j \left\lceil \frac{\log(\beta/5)}{C} \right\rceil , \quad N_{1,j} = j \left\lceil \frac{\log(\alpha/5)}{C} \right\rceil , \quad j \in \mathbb{N},$$

(59)

and

$$N_{0,0}$$ is the minimizer of (55)
over $\{1, \ldots, n^*(\alpha/2, \beta/2)\}$ if $K_0 = 0$
and over $\{1, \ldots, \lceil \|\log(\beta/5)/C\| \rceil \}$ if $K_0 \geq 1$,

$N_{1,0}$ is the minimizer of (56)
over $\{1, \ldots, n^*(\alpha/2, \beta/2)\}$ if $K_1 = 0$
and over $\{1, \ldots, \lceil \|\log(\alpha/5)/C\| \rceil \}$ if $K_1 \geq 1$.

(60)

When both $K_0$ and $K_1$ are selected to be zero, which is always a feasible choice for any $\alpha, \beta$, we recover the 3-Stage Test of the previous section with the proposed design in Subsection III-C. However, for the resulting test to achieve asymptotic optimality even if $\alpha$ and $\beta$ go to zero very asymmetrically, we need to select $K_0$ and $K_1$ appropriately, as functions of $\alpha$ and $\beta$.

Specifically, since all other parameters of GMT have been specified (as functions of $K_0$ and $K_1$), we can select $K_0$ (resp. $K_1$) to optimize the upper bound in (55) (resp. (56)) over $\{0, 1, \ldots, K_0\}$ (resp. $\{0, 1, \ldots, K_1\}$), where $K_0$ (resp. $K_1$) is defined in (58). Alternatively, we can simply select $K_0$ and $K_1$ equal to their upper bounds in (58), i.e., as $K_0 = K_0$ and $K_1 = K_1$ or, equivalently, as

$$K_0 = \left\lceil \frac{n^*(\alpha/2, \beta/2)}{\|\log(\beta/5)/C\|} \right\rceil - 1,$$

$$K_1 = \left\lceil \frac{n^*(\alpha/2, \beta/2)}{\|\log(\alpha/5)/C\|} \right\rceil - 1.$$

(61)

As we show in the next subsection, this selection always suffices for asymptotic optimality of GMT under both hypotheses for any decay rates of $\alpha$ and $\beta$. Moreover, even with this selection, at least one of $K_0$ and $K_1$ is always equal to 0, whereas the other one is typically very small unless $\alpha$ and $\beta$ are very asymmetric. Indeed, by the definition of $K_0$ and $K_1$ in (58) and the asymptotic upper bound in (22) it follows that, for any $\alpha, \beta \in (0,1)$,

$$K_0 < \frac{\|\log((\alpha \wedge \beta)/2)/\log(\beta/5)\|}{\|\log((\alpha \wedge \beta)/2)/\log(\alpha/5)\|}, \quad K_1 < \frac{\|\log((\alpha \wedge \beta)/2)/\log(\beta/5)\|}{\|\log((\alpha \wedge \beta)/2)/\log(\alpha/5)\|}.$$

(62)
Consequently, for any $\alpha, \beta \in (0, 1)$ we have
\[
\alpha \geq \frac{2\beta}{5} \Rightarrow \hat{K}_0 = 0, \\
\alpha \leq \frac{5\beta}{2} \Rightarrow \hat{K}_1 = 0.
\]
Therefore, with $K_0$ and $K_1$ selected as in (61), the GMT provides additional opportunities, relative to the ST, to accept (resp. reject) $H_0$ only if $\alpha$ is much smaller (resp. larger) than $\beta$, and it reduces to the 3ST of the previous section when $\alpha, \beta$ are not very different, e.g., when $2\beta/5 \leq \alpha \leq 5\beta/2$.

D. Asymptotic Optimality

We now state the main theoretical result of this work in the context of the binary testing problem, which is the asymptotic optimality of GMT under both hypotheses as $\alpha, \beta \to 0$ at arbitrary rates.

**Theorem 17:** For any $\alpha, \beta \in (0, 1)$, let $\hat{T}(\alpha, \beta)$ denote the sample size and $\hat{D}(\alpha, \beta)$ the decision of GMT when its parameters are selected according to (52)-(54) and (59)-(61). Then, the family
\[
\hat{\chi} = \{ (\hat{T}(\alpha, \beta), \hat{D}(\alpha, \beta)) : \alpha, \beta \in (0, 1) \}
\]
is asymptotically optimal under both hypotheses as $\alpha, \beta \to 0$ at arbitrary rates.

**Proof:** See Appendix C.

In the special case of the Gaussian mean testing problem of Subsection II-E, the same second-order asymptotic upper bounds as in Proposition 8 hold, without the condition (37) on the decay rates of $\alpha$ and $\beta$.

**Proposition 18:** Consider the Gaussian mean testing problem of Subsection II-E and, for any $\alpha, \beta \in (0, 1)$, let $\hat{T}(\alpha, \beta)$ be defined as in Theorem 17. Then, as $\alpha, \beta \to 0$,
\[
\begin{align*}
E_0[\hat{T}(\alpha, \beta)] &\leq \frac{|\log \beta|}{I} \left(1 + O\left(\sqrt{\frac{|\log |\log \beta||}{|\log \beta|}}\right)\right), \\
E_1[\hat{T}(\alpha, \beta)] &\leq \frac{|\log \alpha|}{I} \left(1 + O\left(\sqrt{\frac{|\log |\log \alpha||}{|\log \alpha|}}\right)\right). \quad (63)
\end{align*}
\]

**Proof:** See Appendix C.

**Algorithm 2** Sequential Thresholding (ST) and Modified Sequential Thresholding (Mod-ST)

Input: $K$: $(m_j, b_j)$, $1 \leq j \leq K$
Initialize: $j = 1$
while $j \leq K - 1$
do take $m_j$ samples
if $\Lambda_j' \leq b_j$ then
stop and accept the null
else
$j = j + 1$
end if
end while
if $j = K$ then
take $m_K$ samples
if $\Lambda_K' \leq b_K$ then
stop and accept the null
else
stop and reject the null
end if
end if

V. SEQUENTIAL THRESHOLDING

In this section, we revisit the multistage test that was proposed in [21] and termed Sequential Thresholding (ST). The main features of this test are that (i) it can accept the null hypothesis at every stage, but it can reject it only at the last possible stage, (ii) it discards all previous data at the beginning of every stage.

From the results in [21] it follows that, with an appropriate selection of its maximum number of stages, this test is asymptotically optimal under the null hypothesis as the error probabilities go to zero at arbitrary rates. However, as we show in this section, this comes at the price of severe performance loss under the alternative hypothesis. Motivated by this phenomenon, we introduce a modification of ST, to which we refer as the Modified Sequential Thresholding (mod-ST), which does not discard data from previous stages, and we show that this test achieves substantially better performance than ST under the alternative hypothesis.

A. Description

We describe the two tests, ST and mod-ST, in parallel. For both of them, we denote by $K$ the maximum number of stages and, for each $1 \leq j \leq K$, we denote by $m_j$ the sample size, by $b_j$ the threshold, and by $\Lambda_j'$ the test statistic that is utilized at the $j$th stage, which is
\[
\Lambda_j' = \begin{cases} 
\frac{1}{m_j} (\Lambda_{M_j} - \Lambda_{M_{j-1}}), & \text{for ST} \\
\Lambda_{M_j}, & \text{for mod-ST}, 
\end{cases} \quad (64)
\]
where $M_j = m_1 + \cdots + m_j$, $M_0 \equiv 0$. Thus, $\Lambda_j'$ is the average log-likelihood ratio of the observations collected only during the $j$th stage for ST, whereas it is the average log-likelihood ratio of all observations that have been collected up to and including the $j$th stage for mod-ST. An algorithmic description for both tests is provided in Algorithm 2. Clearly, they both reduce to the fixed-sample-size test when $K = 1$.

B. Selection of Parameters

Given the maximum number of stages, $K$, for each of the two tests there are $2K$ parameters, $(m_j, b_j)$, $1 \leq j \leq K$, that need to be determined. For each test, the events of rejecting and accepting the null hypothesis can be written respectively as
\[
\bigcap_{j=1}^{K} \{ \Lambda_j' > b_j \} \quad \text{and} \quad \bigcup_{j=1}^{K} \{ \Lambda_j' \leq b_j \}.
\]

Therefore, for each of them to belong to $\mathcal{E}(\alpha, \beta)$ it suffices to select its parameters so that
\[
P_0 \left( \bigcap_{j=1}^{K} \{ \Lambda_j' > b_j \} \right) \leq \alpha \quad (65)
\]
and
\[ \sum_{j=1}^{K} P_1 \left( \Lambda_j' \leq b_j \right) \leq \beta. \] (66)

To obtain a concrete specification of the parameters, we require that the type-II error probabilities after the first stage decay exponentially with the number of stages, i.e.,
\[ P_1 \left( \Lambda_j' \leq b_j \right) \leq \left( \beta/2 \right)^j, \quad 2 \leq j \leq K, \] (67)
and that all remaining type-II error probability be assigned to the first stage, i.e.,
\[ P_1 \left( \Lambda_1' \leq b_1 \right) \leq \beta - \sum_{j=2}^{K} \left( \beta/2 \right)^j. \] (68)

Moreover, we require that the type-I error probability be evenly distributed among the \( K \) stages, so that (65) is strengthened to
\[ P_0 \left( \bigcap_{i=1}^{j} \left\{ \Lambda_i' > b_i \right\} \right) \leq \alpha^{j/K}, \quad 1 \leq j \leq K. \] (69)

Inequalities (67)-(69) provide \( 2K \) constraints for the specification of the \( 2K \) parameters of each test. Specifically, from (68) and (69) with \( j = 1 \) we obtain the following specification for the parameters in the first stage:
\[ (m_1, b_1) = \text{FSST} \left( \alpha^{1/K}, \beta - \sum_{j=2}^{K} \left( \beta/2 \right)^j \right), \] (70)
which is common for ST and mod-ST. However, the parameters of the two tests in the remaining stages differ. Indeed, for ST, the test statistics \( \Lambda_j', 1 \leq j \leq K \) are independent, thus, a sufficient condition for (69) is
\[ P_0( \Lambda_j' > b_j ) \leq \alpha^{j/K}, \quad 1 \leq j \leq K, \]
which, combined with (67), implies that the parameters of ST in the remaining stages can be selected as follows:
\[ (m_j, b_j) = \text{FSST} \left( \alpha^{1/K}, \beta/2^j \right), \quad 2 \leq j \leq K. \] (71)

On the other hand, the remaining parameters of mod-ST need to be specified recursively. Specifically, suppose that \( (m_1, b_1), \ldots, (m_{j-1}, b_{j-1}) \) have been specified for some \( 2 \leq j \leq K \) and set \( M_i = m_1 + \ldots + m_i \) for \( 1 \leq i \leq j-1 \). Then, \( m_j \) is the minimum positive integer such that (67) and (69) hold simultaneously, i.e.,
\[ m_j = \min \left\{ n \in \mathbb{N} : \exists b \in \mathbb{R} \text{ so that} \right. \]
\[ P_1 \left( \Lambda_{M_{j-1}+n} \leq b \right) \leq \left( \beta/2 \right)^j \]
and
\[ P_0 \left( \bigcap_{i=1}^{j-1} \left\{ \Lambda_i > b_i \right\}, \Lambda_{M_{j-1}+n} > b \right) \leq \alpha^{j/K} \] (72)
and \( b_j \) is any threshold for which the two probability constraints in (72) are satisfied when \( n = m_j \).

**Remark 19:** To compute \((m_j, b_j)\) in practice, for any \( n \in \mathbb{N} \), let \( b_j(n) \) denote the largest \( b \) so that the first condition in (72) holds, i.e.,
\[ b_j(n) \equiv c_0^* \left( M_{j-1} + n, \left( \beta/2 \right)^j \right), \]
where the function \( c_0^* \) is defined in (15). Then, \( m_j \) is the smallest \( n \in \mathbb{N} \) so that the second condition in (72) holds, i.e.,
\[ m_j = \min \left\{ n \in \mathbb{N} : \right. \]
\[ P_0 \left( \bigcap_{i=1}^{j-1} \left\{ \Lambda_i > b_i \right\}, \Lambda_{M_{j-1}+n} > b_j(n) \right) \leq \alpha^{j/K} \}, \]
and \( b_j = b_j(m_j) \).

**C. The Number of Stages**

To complete the specification of ST and mod-ST, we need to select the maximum number of stages, \( K \). In contrast to 3ST or GMT, this choice has to strike a balance between the relative cost of sampling under the two hypotheses. Indeed, since ST and mod-ST allow for rejecting the null hypothesis only at the last stage, one should clearly select \( K = 1 \), i.e., apply a fixed-sample-size test, when the absolute priority is to have small expected sample size under the alternative hypothesis. On the other hand, a larger value of \( K \) can lead to smaller expected sample size under the null hypothesis at the expense of performance loss under the alternative hypothesis even in comparison to the fixed-sample-size test.

To resolve this trade-off, one may select the largest value of \( K \) for which the increase in the expected sample size under the alternative hypothesis, relative to FSST, can be tolerated. Alternatively, \( K \) can be selected to minimize the expected sample size under a mixture distribution \( P_x \), defined in (10), for some given \( \pi \in [0, 1] \). A natural choice for this \( \pi \) arises in the signal recovery problem of Section VI, as we discuss in Subsection VII-B. We stress, however, that no such external criterion is needed for the design of 3ST or GMT.

**D. Asymptotic Optimality**

We next show that when \( K \) is selected appropriately, ST is asymptotically optimal under the null hypothesis as \( \alpha, \beta \to 0 \) at arbitrary rates. This was shown in [21] under the assumption of a finite second moment on the log-likelihood ratio statistic. Here, we only require finiteness of the first moment, i.e., (2), which is our standing assumption throughout the paper. At the same time, we show that, with this selection of \( K \), the expected sample size of ST under the alternative hypothesis is asymptotically larger than the optimal by a factor that is much larger than \((K+1)/2\). That is, the asymptotic optimality of ST under the null hypothesis comes at the price of severe performance loss under the alternative hypothesis.

Finally, we show that this performance loss is substantially mitigated when using mod-ST instead of ST. Specifically, we show that mod-ST enjoys the same asymptotic optimality property as ST under the null hypothesis, while its expected sample size under the alternative hypothesis is smaller than that of ST by a factor that is not smaller than \((K+1)/2\).

**Theorem 20:** For any \( \alpha, \beta \in (0, 1) \), let \( T'(\alpha, \beta) \) denote the sample size and \( D'(\alpha, \beta) \) the decision of ST when its parameters are selected according to (70) and (71). Moreover, let \( T''(\alpha, \beta) \) denote the sample size and \( D''(\alpha, \beta) \) the decision...
of mod-ST when its parameters are selected according to (70) and (72). If $K$ is selected as a function of $\alpha$ and $\beta$ such that, as $\alpha, \beta \to 0$,
$$\alpha^{1/K} \to 0 \quad \text{and} \quad |\log \alpha^{1/K}| \ll |\log \beta| \quad (73)$$
or, equivalently,
$$|\log \alpha| \ll K \ll |\log \beta|, \quad (74)$$
then the families
$$\chi' = \{ (T'(\alpha, \beta), D'(\alpha, \beta)) : \alpha, \beta \in (0,1) \}$$
$$\chi'' = \{ (T''(\alpha, \beta), D''(\alpha, \beta)) : \alpha, \beta \in (0,1) \}$$
are both asymptotically optimal under the null hypothesis. Moreover, as $\alpha, \beta \to 0$,
$$E_1[T'(\alpha, \beta)] \sim \frac{K(K+1)}{I_0} \frac{|\log \beta|}{2} \gg \frac{K+1}{2} \mathcal{L}_1(\alpha, \beta) \quad (75)$$
$$E_1[T''(\alpha, \beta)] \leq K \frac{|\log \beta|}{I_0} \sim \frac{2}{K+1} E_1[T'(\alpha, \beta)]. \quad (76)$$

**Proof:** See Appendix D.

With a more specific selection of $K$, we obtain the same second-order asymptotic upper bound for the expected sample sizes of ST and mod-ST under the null hypothesis as that for GMT in Proposition 18, in the Gaussian mean testing problem.

**Proposition 21:** Consider the Gaussian mean testing problem of Subsection II-E and for any $\alpha, \beta \in (0,1)$ let $T'(\alpha, \beta)$ and $T''(\alpha, \beta)$ be defined as in Theorem 20 with (73) replaced by
$$\alpha^{1/K} = \Theta\left(1/\sqrt{|\log \beta|}\right) \quad (77)$$
or, equivalently,
$$K = \frac{2|\log \alpha|}{|\log |\log \beta||} + \Theta(1).$$

Then, as $\alpha, \beta \to 0$,
$$E_0[T'(\alpha, \beta)], E_0[T''(\alpha, \beta)] \leq \frac{|\log \beta|}{I_0} \left(1 + O\left(\sqrt{\frac{|\log |\log \beta||}{|\log \beta|}}\right)\right). \quad (78)$$

**Proof:** See Appendix D.

## VI. HIGH-DIMENSIONAL TESTING

In this section, we consider the problem of simultaneously testing $m$ copies of the binary testing problem of Section II in the presence of lower and upper bounds, $l_m$ and $u_m$, on the number of testing problems in which the alternative hypothesis is true, where
$$0 \leq l_m \leq u_m \leq m, \quad u_m > 0, \quad l_m < m. \quad (79)$$

For example, when $l_m = 0, u_m = m$, there is no prior information available, whereas when $0 < l_m = u_m < m$, the exact number of true alternatives is known.

To be specific, we denote by $X_1', \ldots, X_m'$ independent sequences of iid random elements, and we test for each of them whether its density is $f_0$ or $f_1$. As before, our standing assumption regarding $f_0$ and $f_1$ is the finiteness of the Kullback-Leibler divergences, i.e., we assume only that (2) holds. We refer to a stream as noise if its density is $f_0$ and as signal if its density is $f_1$, and we denote by $P_A$ and $E_A$ the probability measure and the expectation when the subset of signals is $A \subseteq [m]$, where here and in what follows we use the following notation:
$$[m] \equiv \{1, 2, \ldots, m\}.$$

We restrict ourselves to multiple testing procedures that apply the same binary test, i.e., a test in $\mathcal{E}$, to each stream. Therefore, in order to specify a multiple testing procedure, it suffices to specify a binary test in $\mathcal{E}$.

For this reason, in what follows we identify the class of multiple testing procedures with the family of binary tests, $\mathcal{E}$.

If $(T, D) \in \mathcal{E}$ is the binary test that is applied to each stream, we denote by $(T_j, D_j)$ its version that is applied to the $j$th stream, where $j \in [m]$, and we note that the expected average sample size over all $m$ streams of the resulting multiple testing procedure depends only on the number of signals. Indeed, for any $A \subseteq [m]$ we have
$$E_A\left[\frac{1}{m} \sum_{j=1}^{m} T_j \right] = \frac{1}{m} \sum_{j \in A} E_0[T_j] + \sum_{j \notin A} E_1[T_j]$$
$$= \left(1 - \frac{|A|}{m}\right) E_0[T] + \frac{|A|}{m} E_1[T]$$
$$= E_{|A|/m}[T], \quad (80)$$

recalling that $E_x$ denotes the expectation under the mixture distribution, $P_x$, defined in (10).

Our goal in this section is to find multiple testing procedures, which apply to each stream a multistage test, that minimize the expected average sample size over all $m$ streams as $m \to \infty$ under certain constraints on global error probabilities of both types. Specifically, we consider classical familywise error control in Subsection VI-A and generalized familywise error control in Subsection VI-B.

**Remark 22:** The setup that is considered in this section generalizes the one in [21] in two ways. First, it does not require the number of signals to be a priori known and, second, it allows for distinct control of type-I and type-II error probabilities.

### A. The Case of Classical Familywise Error Control

For any $(T, D) \in \mathcal{E}$ and $A \subseteq [m]$, we denote by $FWE-I_A(T, D)$ and $FWE-II_A(T, D)$ the familywise type-I and type-II error probabilities of $(T, D)$ when the subset of signals is $A$, i.e.,
$$FWE-I_A(T, D) \equiv P_A(\exists j \notin A : D_j = 1),$$
$$FWE-II_A(T, D) \equiv P_A(\exists j \in A : D_j = 0).$$

For any $m \in \mathbb{N}$, $l_m, u_m$ that satisfy (79), and $\alpha, \beta \in (0,1)$, we denote by $\mathcal{E}_m(\alpha, \beta)$ the family of tests whose familywise...
type-I and type-II error probabilities do not exceed \( \alpha \) and \( \beta \) respectively for any subset of signals whose size is between \( l_m \) and \( u_m \), i.e.,

\[
E_m(\alpha, \beta) \equiv \{(T, D) \in \mathcal{E} : \text{FWE-I}_A(T, D) \leq \alpha \text{ and } \text{FWE-II}_A(T, D) \leq \beta, \forall A \subseteq [m], l_m \leq |A| \leq u_m\}.
\]

This family of tests can be expressed in terms of the one in (4) in the following way:

**Lemma 23:** For any \( m \in \mathbb{N} \) and \( \alpha, \beta \in (0, 1) \),

\[
E_m(\alpha, \beta) = \mathcal{E}(\alpha_m, \beta_m),
\]

where

\[
\alpha_m = 1 - (1 - \alpha)^{(m - l_m)} \quad \text{and} \quad \beta_m = 1 - (1 - \beta)^{1/u_m}.
\]

**Proof:** See Appendix E.

From (80)-(81) it follows that the optimal expected average sample size in \( E_m(\alpha, \beta) \) when the number of signals is \( s \in \{l_m, \ldots, u_m\} \) can be written as

\[
\mathop{\inf}_{(T, D) \in E_m(\alpha, \beta)} E_{s/m}[T] = \mathop{\inf}_{(T, D) \in \mathcal{E}(\alpha_m, \beta_m)} E_{s/m}[T] = \mathcal{L}_{s/m}(\alpha_m, \beta_m).
\]

Our goal in this subsection is to find families of multistage tests that achieve this infimum, uniformly in the possible number of signals, \( s \in \{l_m, \ldots, u_m\} \), as the number of streams, \( m \), and the maximum numbers of signals and noises, \( u_m \) and \( m - l_m \), go to infinity, for any given and fixed \( \alpha, \beta \in (0, 1) \). This is expressed by the following notion of asymptotic optimality.

**Definition 24:** A family of tests, \( \chi^* \), defined as in (6), is asymptotically optimal in the high-dimensional sense if, for any \( \alpha, \beta \in (0, 1) \), as \( u_m \rightarrow \infty \) and \( m \rightarrow \infty \),

\[
\mathop{\inf}_{s \in \{l_m, \ldots, u_m\}} E_{s/m}[T^*(\alpha_m, \beta_m)] \sim \mathcal{L}_{s/m}(\alpha_m, \beta_m)
\]

uniformly in \( s \in \{l_m, \ldots, u_m\} \), (83)

\[
\mathop{\max}_{s \in \{l_m, \ldots, u_m\}} \frac{E_{s/m}[T^*(\alpha_m, \beta_m)]}{\mathcal{L}_{s/m}(\alpha_m, \beta_m)} \rightarrow 1.
\]

We next characterize the optimal asymptotic performance and provide criteria for a family of tests to achieve it.

**Theorem 25:** (i) For any \( \alpha, \beta \in (0, 1) \), as \( u_m \rightarrow \infty \) and \( m - l_m \rightarrow \infty \),

\[
\mathcal{L}_{s/m}(\alpha_m, \beta_m) \sim \left(1 - \frac{s}{m}\right) \frac{\log u_m}{I_0} + \frac{s}{m} \log \left(m - l_m\right) I_1
\]

uniformly in \( s \in \{l_m, \ldots, u_m\} \), (85)

Let \( \chi^* \) be a family of binary tests defined as in (6).

(ii) If, for any \( \alpha, \beta \in (0, 1) \),

\[
E_0[T^*(\alpha_m, \beta_m)] \sim \mathcal{L}_0(\alpha_m, \beta_m),
\]

\[
E_1[T^*(\alpha_m, \beta_m)] \sim \mathcal{L}_1(\alpha_m, \beta_m),
\]

as \( u_m, m - l_m \rightarrow \infty \), then \( \chi^* \) is asymptotically optimal in the high-dimensional sense.

(iii) If, for any \( \alpha, \beta \in (0, 1) \), (86) holds and also

\[
E_1[T^*(\alpha_m, \beta_m)] \ll \frac{(m - u_m) \log u_m}{u_m}
\]

as \( u_m, m - l_m \rightarrow \infty \), then \( \chi^* \) is asymptotically optimal in the high-dimensional sense.

**Proof:** See Appendix E.

Using the previous theorem, we can now establish the asymptotic optimality of SPRT and GMT, without any constraints on the divergence rates of the maximum numbers of signals and noises, \( (u_m) \) and \( (m - l_m) \).

**Corollary 26:** • The family of SPRTs, \( \tilde{\chi} \), defined in (9), is asymptotically optimal in the high-dimensional sense.

• The family of GMTs, \( \tilde{\chi} \), defined in Theorem 17, is asymptotically optimal in the high-dimensional sense.

**Proof:** See Appendix E.

On the other hand, for the asymptotic optimality in the high-dimensional sense of 3ST, we need a symmetry condition on the divergence rates of \( \log(u_m) \) and \( \log(m - l_m) \).

**Corollary 27:** If \( u_m, m - l_m \rightarrow \infty \) so that \( \log(m - l_m) = \Theta(\log u_m) \), then the family of 3STs, \( \tilde{\chi} \), defined in Theorem 7, is asymptotically optimal in the high-dimensional sense.

**Proof:** See Appendix E.

Finally, for the asymptotic optimality in the high-dimensional sense of ST and mod-ST we need the upper bound on the number of signals, \( u_m \), to be much smaller than the total number of streams, \( m \).

**Corollary 28:** If \( u_m \rightarrow \infty \) and \( u_m \ll m \rightarrow \infty \), then,

(i) for any \( \alpha, \beta \in (0, 1) \),

\[
\mathcal{L}_{s/m}(\alpha_m, \beta_m) \sim \frac{\log u_m}{I_0}
\]

uniformly in \( s \in \{l_m, \ldots, u_m\} \), (89)

(ii) the family of STs, \( \chi' \), and the family of mod-STs, \( \chi'' \), defined in Theorem 20, are both asymptotically optimal in the high-dimensional sense.

**Proof:** See Appendix E.

**Remark 29:** In [21, Theorem 3, Corollary 2], the asymptotic optimality of ST in the high-dimensional sense is established

• when the number of signals is a priori known, i.e., \( l_m = u_m \) for every \( m \in \mathbb{N} \),

• when \( u_m \rightarrow \infty \) so that \( u_m \ll (\log m)^2 \),

• assuming a finite second moment for the log-likelihood ratio in (3),

• controlling the probability of at least one misclassification, i.e., an error of either kind.

**Corollary 28** generalizes [21, Theorem 3, Corollary 2] in that it establishes the asymptotic optimality of ST in the high-dimensional sense

• without any assumption on \( l_m \), which can even be equal to 0 for every \( m \in \mathbb{N} \),

• whenever \( u_m \rightarrow \infty \) so that \( u_m \ll m \),

• assuming only (2), i.e., a finite first moment for the log-likelihood ratio in (3),

• allowing to control the probability of at least one type-I error and the probability of at least one type-II error below distinct levels.

**Remark 30:** Corollary 28 shows that under a sparse setup, i.e., when \( u_m \rightarrow \infty \) so that \( u_m \ll m \), the optimal expected average sample size under any true number of signals is of
the order of the logarithm of the maximum possible number of signals. On the other hand, under a setup that is neither sparse nor dense, i.e., when \( u_m = \Theta(m) \) and \( m - l_m = \Theta(m) \) as \( m \to \infty \), which implies \( \log u_m \sim \log(m - l_m) \sim \log m \), from (85) it follows that

\[
\frac{\log m}{I_0 \vee I_1} \leq \mathcal{L}_{s/m}(\alpha_m, \beta_m) \leq \frac{\log m}{I_0 \wedge I_1},
\]

uniformly for \( s \in \{l_m, \ldots, u_m\} \), i.e., the optimal expected average sample size under any true number of signals is of the order of the logarithm of the total number of streams. The difference between these two regimes is illustrated in Figures 4.(a) and 5.(a).

**B. The Case of Generalized Familywise Error Control**

We next generalize the results of the previous subsection by considering multiple testing procedures that control generalized familywise Type-I error probabilities [22]. Thus, in what follows, we let \( \kappa_m \in \mathbb{N} \) and \( \iota_m \in \mathbb{N} \) be such that

\[
1 \leq \iota_m \leq u_m, \quad 1 \leq \kappa_m \leq m - l_m,
\]

where \( \kappa_m \) and \( \iota_m \) are user-specified quantities, similarly to \( l_m \) and \( u_m \) in (79). For any test \( (T, D) \in \mathcal{E} \), we denote by \( \kappa_m \)-GFWE-I \( A(T, D) \) its \( \kappa_m \)-generalized familywise Type-I error probability, that is, the probability of at least \( \kappa_m \) Type-I errors, when the subset of signals is \( A \subseteq [m] \), i.e.,

\[
\kappa_m \text{-GFWE-I}_A(T, D) \equiv P_A(\exists B \subseteq A^c : |B| = \kappa_m \text{ and } D^j = 1 \text{ for all } j \in B),
\]

and by \( \iota_m \)-GFWE-II \( A(T, D) \) its \( \iota_m \)-generalized familywise Type-II error probability, that is, the probability of at least \( \iota_m \) Type-II errors, when the subset of signals is \( A \subseteq [m] \), i.e.,

\[
\iota_m \text{-GFWE-II}_A(T, D) \equiv P_A(\exists B \subseteq A : |B| = \iota_m \text{ and } D^j = 0 \text{ for all } j \in B).
\]

For any \( m \in \mathbb{N} \) and \( \alpha, \beta \in (0, 1) \), we denote by \( \mathcal{E}^G_m(\alpha, \beta) \) the family of tests for which these two error probabilities are bounded above by \( \alpha \) and \( \beta \) for any subset of signals whose size is between \( l_m \) and \( u_m \), i.e.,

\[
\mathcal{E}^G_m(\alpha, \beta) \equiv \{ (T, D) \in \mathcal{E} : \kappa_m \text{-GFWE-I}_A(T, D) \leq \alpha \text{ and } \iota_m \text{-GFWE-II}_A(T, D) \leq \beta, \forall A \subseteq [m], l_m \leq |A| \leq u_m \}.
\]

This family of tests can be expressed in terms of the family in (6) in the following way.

**Lemma 31:** For any \( m \in \mathbb{N} \) and \( \alpha, \beta \in (0, 1) \),

\[
\mathcal{E}^G_m(\alpha, \beta) = \mathcal{E}(\alpha^G_m, \beta^G_m),
\]

where

\[
\alpha^G_m \equiv \sup \{ p \in (0, 1) : B(m - l_m, p; \kappa_m) \leq \alpha \}, \quad \beta^G_m \equiv \sup \{ p \in (0, 1) : B(u_m, p; \iota_m) \leq \beta \},
\]

and \( B(n, p; k) \) is the probability that a Binomial random variable with parameters \( n \in \mathbb{N} \) and \( p \in (0, 1) \) is greater than or equal to \( k \in \{0, 1, \ldots, n\} \).

**Proof:** See Appendix E.

From (80) and (90) it follows that the optimal expected average sample size in \( \mathcal{E}^G_m(\alpha, \beta) \) when the number of signals is equal to \( s \in \{l_m, \ldots, u_m\} \) is

\[
\inf_{(T, D) \in \mathcal{E}^G_m(\alpha, \beta)} \mathbb{E}_{s/m}[T] = \mathcal{L}_{s/m}(\alpha^G_m, \beta^G_m).
\]

This leads us to the following definition of asymptotic optimality.

**Definition 32:** A family of tests, \( \chi^* \), defined as in (6), is asymptotically optimal in the high-dimensional sense under generalized error control if, for any \( \alpha, \beta \in (0, 1/2) \), as \( u_m, m - l_m \to \infty \) so that \( \iota_m \ll u_m, \kappa_m \ll m - l_m \),

\[
\mathbb{E}_{s/m}[T^*(\alpha^G_m, \beta^G_m)] \sim \mathcal{L}_{s/m}(\alpha^G_m, \beta^G_m)
\]

uniformly in \( s \in \{l_m, \ldots, u_m\} \), i.e.,

\[
\max_{s \in \{l_m, \ldots, u_m\}} \frac{\mathbb{E}_{s/m}[T^*(\alpha^G_m, \beta^G_m)]}{\mathcal{L}_{s/m}(\alpha^G_m, \beta^G_m)} \to 1.
\]

We next characterize the optimal asymptotic performance and provide criteria for a family of tests to achieve it.

**Theorem 33:** (i) For any \( \alpha, \beta \in (0, 1/2) \), as \( u_m \to \infty \) and \( m - l_m \to \infty \) so that \( \iota_m \ll u_m, \kappa_m \ll m - l_m \),

\[
\mathcal{L}_{s/m}(\alpha^G_m, \beta^G_m) \sim \left(1 - \frac{s}{m}\right) \frac{\log(u_m/l_m)}{I_0} + \frac{s \log((m - l_m)/\kappa_m)}{I_1}
\]

uniformly in \( s \in \{l_m, \ldots, u_m\} \).

(ii) If, for any \( \alpha, \beta \in (0, 1/2) \), as \( u_m, m - l_m \to \infty \) so that \( \iota_m \ll u_m \) and \( \kappa_m \ll m - l_m \),

\[
\mathbb{E}_0[T^*(\alpha^G_m, \beta^G_m)] \sim \mathbb{E}_0(\alpha^G_m, \beta^G_m), \quad \mathbb{E}_1[T^*(\alpha^G_m, \beta^G_m)] \sim \mathbb{E}_1(\alpha^G_m, \beta^G_m),
\]

then \( \chi^* \) is asymptotically optimal in the high-dimensional sense under generalized error control.

(iii) If, for any \( \alpha, \beta \in (0, 1/2) \), as \( u_m, m - l_m \to \infty \) so that \( \iota_m \ll u_m \) and \( \kappa_m \ll m - l_m \), (94) holds and, also,

\[
\mathbb{E}_1[T^*(\alpha^G_m, \beta^G_m)] \ll \frac{(m - u_m) \log(u_m/l_m)}{u_m},
\]

then \( \chi^* \) is asymptotically optimal in the high-dimensional sense under generalized error control.

**Proof:** Appendix E.

**Remark 34:** The optimal asymptotic performance in (93) reduces to that under classical error control in (85) when \( \iota_m \) and \( \kappa_m \) remain bounded as \( m \to \infty \) or, more generally, when \( \log \iota_m \ll \log u_m \) and \( \log \kappa_m \ll \log(m - l_m) \) as \( m \to \infty \). Moreover, the asymptotic approximation to the optimal performance in (93) suggests that generalized error control essentially reduces the “effective” maximum numbers of signals and noises.

We end this section with the analogues of Corollaries 26, 27 and 28.


Corollary 35: • The family of SPRTs, $\hat{\chi}$, defined in (9), is asymptotically optimal in the high-dimensional sense under generalized error control.
• The family of GMTs, $\chi$, defined in Theorem 17, is asymptotically optimal in the high-dimensional sense under generalized error control.

Proof: See Appendix E.

Corollary 36: If
\[
\log \left(\frac{u_m}{l_m}\right) = \Theta \left(\log \left(\frac{m-l_m}{\kappa_m}\right)\right)
\]
as $u_m, m - l_m \to \infty$ so that $l_m \ll u_m$ and $\kappa_m \ll m - l_m$, then the family of 3STs, $\hat{\chi}$, defined in Theorem 7 is asymptotically optimal in the high-dimensional sense under generalized error control.

Proof: See Appendix E.

Corollary 37: If $u_m \to \infty, u_m \ll m, l_m \ll u_m$ and $\kappa_m \ll m - l_m$ as $m \to \infty$, then
(i) for any $\alpha, \beta \in (0, 1/2)$,
\[
L_{\alpha/m}(\alpha \beta_{m}), \beta_{m}) \sim \frac{\log(u_m/l_m)}{I_0}
\]
uniformly in $s \in \{l_m, \ldots, u_m\}$,
(ii) the family of mod-STs, $\chi''$, defined in Theorem 20, is asymptotically optimal in the high-dimensional sense under generalized error control if also
\[
\log \left(\frac{m}{\kappa_m}\right) \ll \log \left(\frac{u_m}{l_m}\right)
\]
(iii) the family of STs, $\chi'$, defined in Theorem 20, is asymptotically optimal in the high-dimensional sense under generalized error control if also
\[
\log \left(\frac{m}{\kappa_m}\right) \ll \sqrt{\frac{m}{u_m}}.
\]

Proof: See Appendix E.

Remark 38: (i) Unlike the case of classical error control in Corollary 28, the condition $u_m \ll m$ does not suffice for the proof of the asymptotic optimality of ST and mod-ST in the high-dimensional sense under generalized error control unless $l_m \ll \log u_m$.

(ii) Sufficient conditions for (97) and (98), which do not depend on $\kappa_m$ and $l_m$, are
\[
u_m \lesssim \frac{m}{\log m} \quad \text{and} \quad u_m \lesssim \frac{m}{(\log m)^2},
\]
respectively. To see this, it suffices to observe that, since $\kappa_m \geq 1$ and $u_m/l_m \to \infty$,
\[
\frac{\log(m/\kappa_m)}{\log(u_m/l_m)} \ll \log m.
\]

VII. NUMERICAL STUDIES

In this section, we present the results of two numerical studies, one in binary testing and one in multiple testing with classical familywise error control, where we compare the General Multistage Test (GMT) with

- the Sequential Probability Ratio Test (SPRT) and the Fixed-Sample-Size Test (FSST),
- the 3-Stage Test (3ST), the Sequential Thresholding (ST), and the Modified Sequential Thresholding (mod-ST).

A. Binary Testing

In the first study, we consider the binary Gaussian mean testing problem of Subsection II-E, with $\eta = 0.5$, in three cases regarding the target error probabilities, a symmetric one and two asymmetric ones:

(i) $\alpha = 10^{-6}$
(ii) $\alpha = 10^{-12}, \beta = 10^{-2}$
(iii) $\alpha = 10^{-2}, \beta = 10^{-12}$.

The tests are designed as follows:

- FSST is computed according to the formulas in (26). Specifically, FSST$(\alpha, \beta)$ is $(91, 0)$ in case (i), $(88, 0.2509)$ in case (ii) and $(88, -0.2509)$ in case (iii).
- The parameters of 3ST are selected as in Subsection III-C.
- The parameters of GMT are selected as in Theorem 17. Specifically, we have $K_0 = K_1 = 0$ in case (i), $K_0 = 2, K_1 = 0$ in case (ii), and $K_0 = 0, K_1 = 2$ in case (iii). That is, GMT reduces to 3ST in case (i), whereas it adds two more opportunities to select $H_0$ (resp. $H_1$) in case (ii) (resp. (iii)).
- The parameters of ST and mod-ST are selected according to (70)-(72), with the maximum possible number of stages, $K$, selected to match that of GMT. Thus, we have $K = 3$ in case (i) and $K = 5$ in cases (ii) and (iii).
- SPRT, defined in (7), is designed with $A = |\log \alpha|$ and $B = |\log \beta|$.

For each of the three cases, we compute the expected sample size (ESS) of each of the six tests not only when the mean of the Gaussian sequence, $\mu$, is in $\{\pm 0.5, 0.5\}$, but also for 100 equally-spaced values in $[-0.6, 0.6]$. For each value of $\mu$, we compute the ESS of 3ST using the expressions in (103)-(104), whereas we use similar (although more complicated) expressions for the ESS of GMT. Similarly, we compute the ESS of ST and of mod-ST using the expressions in (114). Finally, we compute the ESS of SPRT using $10^5$ Monte-Carlo simulation runs for each value of $\mu$, with the resulting relative standard error well below 1% in all cases.

For each test, the ratios over $n^*(\alpha, \beta)$ of the ESS when $\mu = 0.5$, the ESS when $\mu = -0.5$, and of the maximum

| $\mu$ | 3ST | GMT | ST | mod-ST | SPRT |
|-------|-----|-----|----|--------|------|
| $\mu = -0.5$ | 0.48 | 0.48 | 0.57 | 0.37 | 0.32 |
| worst-case | 1.07 | 1.07 | 2.99 | 2.03 | 2.29 |
| $\mu = 0.5$ | 0.48 | 0.48 | 2.99 | 2.03 | 0.32 |
| worst-case | 1.02 | 1.00 | 3.41 | 2.13 | 2.01 |
| $\mu = -0.5$ | 0.79 | 0.79 | 3.39 | 2.12 | 0.64 |
| worst-case | 1.02 | 1.00 | 3.41 | 2.13 | 2.01 |
ESS with respect to $\mu \in [-0.6, 0.6]$ are presented in Table I. Moreover, the ESS is plotted against $\mu$ in Figure 1.

Based on Table I and Figure 1, we can make the following observations regarding the comparison between GMT, SPRT, and FSST.

- In the symmetric case, the ESS of GMT (resp. SPRT) under both hypotheses is about half (resp. a third) of $n^*(\alpha, \beta)$. In the asymmetric cases, the ESS of GMT (resp. SPRT) is 19% (resp. 12%) of $n^*(\alpha, \beta)$ under one of the two hypotheses and 79% (resp. 64%) of $n^*(\alpha, \beta)$ under the other.
- In all cases, the worst-case ESS of SPRT is more than double of $n^*(\alpha, \beta)$, whereas that of GMT is about the same as $n^*(\alpha, \beta)$.

We continue with the comparison among the multistage tests.

- In the two asymmetric cases, where GMT does not coincide with 3ST, the ESS is the same for the two tests under one of the two hypotheses. Under the other one, the ESS of 3ST is 24% of $n^*(\alpha, \beta)$, while the ESS of GMT is 19% of $n^*(\alpha, \beta)$.
- The ESS of mod-ST is very close to that of ST when $\mu$ is around the null, while being substantially smaller for larger values of $\mu$. Nevertheless, even for mod-ST, the ESS for large values of $\mu$ is much larger than $n^*(\alpha, \beta)$.
- Even when $\mu$ is around the null, ST and mod-ST have somewhat larger expected sample size than even 3ST.

**Remark 39:** For GMT we have selected $K_0$ and $K_1$ as the maximum number of stages implied by the proposed design, i.e., $K_0 = \hat{K}_0$ and $K_1 = \hat{K}_1$, where $\hat{K}_0$ and $\hat{K}_1$ are given by (58). This leads to two additional, with respect to the 3ST, opportunities to accept (resp. reject) the null hypothesis in case (ii) (resp. (iii)). As we saw, this choice leads to a significant improvement over the 3ST under one of the two hypotheses. It turns out that almost the same improvement can be achieved when allowing for only one additional, with respect to the 3ST, opportunity to accept (resp. reject) the null hypothesis in case (ii) (resp. (iii)). This is illustrated in Figure 2.(a), where we plot the ESS of GMT against $\mu$ for all possible choices of $0 \leq K_0 \leq \hat{K}_0$ in case (ii).

**Remark 40:** Increasing the number of stages of ST or mod-ST leads to a further increase of the ESS for larger values of $\mu$, especially in the case of ST, while offering relatively small reduction of the ESS for $\mu$ around the null. This is illustrated in Figure 2.(b) and (c), where we plot the ESS of ST and mod-ST against $\mu$ for $K$ ranging from 1 to 6 in case (ii).

### B. High-Dimensional Testing

In the second study, we consider again the Gaussian mean testing problem of Subsection II-E, with $\eta = 0.5$, but now in the multiple testing context of Section VI with $m = 10^6$ data streams when we control the classical familywise type-I and type-II error probabilities below $\alpha = 0.05$ and $\beta = 0.05$, respectively. Since we consider only one value for $m$, in what follows we simply use $l$ and $u$, instead of $l_m$ and $u_m$, to denote the user-specified lower and upper bound on the number of signals, respectively.

We consider two setups regarding the prior information on the number of signals. In the first one, this number is assumed to be known, i.e., $l = u$, and for each of its possible values, i.e., for each of $1 \leq u \leq m - 1$, we compute the ESS of each of the six tests with respect to the mixture distribution $P_\pi$, defined in (10), with $\pi = u/m$. In the second setup, only an upper bound on the number of signals is assumed to be
known, i.e., \( l = 0 \), and for each possible value of the upper bound, i.e., for each \( 1 \leq u \leq m \), we compute the ESS of each test with respect to the mixture distribution \( P_\pi \), defined in (10), with \( \pi = u/2m \). This corresponds to the average ESS over the \( m \) streams and over the \( u + 1 \) possible values for the number of signals, \( \{0, 1, \ldots, u\} \). Indeed, for any stopping time \( T \) we have

\[
\frac{1}{u+1} \sum_{j=0}^{u} E_{j/m}[T] = \frac{1}{u+1} \sum_{j=0}^{u} \left( \left(1 - \frac{j}{m}\right) E_0[T] + \frac{j}{m} E_1[T] \right) = \left(1 - \frac{u}{2m}\right) E_0[T] + \frac{u}{2m} E_1[T] = E_{u/2m}[T].
\]

(99)

In both setups, all tests are designed in exactly the same way as in the first study, with \( \alpha \) and \( \beta \) replaced by \( \alpha_m \) and \( \beta_m \), respectively, which are defined in (82), i.e.,

\[
1 - (1 - \alpha)^{1/(m-l)} \text{ and } 1 - (1 - \beta)^{1/u}.
\]

(100)

The only difference is that for ST and mod-ST the maximum number of stages, \( K \), is not selected, as in the first study, to match the maximum number of stages of GMT. Instead, for each value of \( u \) we select \( K \) to minimize the criterion of each setup, i.e., the expected sample size under the mixture distribution \( P_\pi \) with \( \pi = u/m \) in the first setup and with \( \pi = u/2m \) in the second.

In Figure 3 we plot the maximum number of stages for GMT, ST and mod-ST against \( u/m \). For GMT, this number is equal to

- 3 when \( u \) and \( m - l \) are close, which is the case when \( u/m \) is around 1/2 (resp. close to 1) in the first (resp. second) setup,
- 5 when \( u \) and \( m - l \) are very different, which is the case when \( u/m \) is close to 0 or 1 (resp. close to 0) in the first (resp. second) setup,
- 4 in all other cases.

Regarding ST and mod-ST, we observe that they both reduce to the fixed-sample-size test when \( u \) is large. Specifically, for ST (resp. mod-ST), this is the case when \( u/m \) is larger than about 0.3 (resp. 0.4) in the first setup and 0.55 (resp. 0.75) in the second. On the other hand, as \( u \) decreases, their maximum number of stages increases up to 9, in comparison to at most 5 for GMT.

For each setup and each possible value of \( u \), the ESS of each test is computed in the same way as in the first study. The results for the first setup are presented in Figure 4 and for the second in Figure 5. Based on these figures we can make the following observations:

- All curves decrease sharply for very small values of \( u \) and the curves for 3ST, GMT and SPRT are flat for medium values of \( u \). This is consistent with the second remark after Corollary 28.
- The ESS of GMT (resp. SPRT) is smaller roughly by a factor of 2 (resp. 4) relative to that of FSST. This does not contradict the results of the previous study, as all these curves correspond to weighted averages of the ESS when \( \mu = \eta \) and when \( \mu = -\eta \), i.e., we do not consider values of \( \mu \) between \( -\eta \) and \( \eta \).
- The ESS of 3ST and of GMT are basically the same except in the case where \( u \) and \( m - l \) are very different, which is the case where \( u \) is very small or large (resp. very small) in the first (resp. second) setup.
The ESS of mod-ST is similar to that of ST when $u$ is very small, but becomes smaller as $u$ increases, until $u/m$ reaches a value of about 0.4 in the first setup and about 0.75 in the second, at which both tests reduce to the FSST.

For very small values of $u$, the ESS of mod-ST, but not that of ST, is slightly smaller than that of GMT, although the difference is too small to be visible. Note, however, that for such values of $u$, mod-ST and ST use 9 stages, whereas GMT uses 5.

VIII. GENERALIZATIONS

For the sake of simplicity and clarity, we have focused on the fundamental problem of testing two simple hypotheses about the distribution of a sequence of iid random elements. However, the methods and results of this paper remain valid with minor modifications for certain testing problems with non-iid data or composite hypotheses.

A. Non-iid Data

Suppose that $X$ is not necessarily an iid sequence and consider the problem of testing two simple hypotheses about its distribution, $P$:

$$H_0 : P = P_0 \text{ versus } H_1 : P = P_1,$$

where $P_0$ and $P_1$ are mutually absolutely continuous when restricted to $\mathcal{F}_n$ for every $n \in \mathbb{N}$. Then, the log-likelihood ratio statistic

$$\Lambda_n \equiv \log \frac{dP_1}{dP_0}(X_1, \ldots, X_n)$$

is not necessarily of the form (3). However, all results in this work still hold (apart from those about ST, which additionally require observations to be independent) as long as there exist (i) positive numbers, $I_0$ and $I_1$, so that

$$P_0(\Lambda_n \rightarrow -I_0) = P_1(\Lambda_n \rightarrow I_1) = 1,$$

where, as before, $\Lambda_n \equiv \Lambda_n / n$,
Fig. 5. $E_u/2n[T(\alpha_m, \beta_m)]$ against $u/m$ when $\alpha = \beta = 0.05, m = 10^6, l = 0$ and $u \in \{1, \ldots, m\}$. (b) is the left 1% of (a).

(ii) and real-valued functions, $\psi_0$ and $\psi_1$, that satisfy the following asymptotic versions of the inequalities in (16):

$$
\lim_{n} \sup_n \frac{1}{n} \log P_0(\Lambda_n > c) \leq -\psi_0(c), \quad \forall c \geq -I_0,
$$

$$
\lim_{n} \sup_n \frac{1}{n} \log P_1(\Lambda_n \leq c) \leq -\psi_1(c), \quad \forall c \leq I_1,
$$

(101)

and the four properties stated after (20).

Indeed, these conditions imply that the asymptotic approximations to the optimal expected sample sizes under the two hypotheses in (8) remain valid (see, e.g., [23, Section 3.4]), and also that the upper bounds (21)-(22) remain valid to a first-order asymptotic approximation as $\alpha, \beta \to 0$. As a result, all proofs in this work remain valid, without essentially any modification.

Remark 41: The above conditions are satisfied, for example, when testing the transition matrix of a finite-state Markov chain, or the correlation coefficient of a first-order autoregression. For more details, we refer to [12].

B. Composite Hypotheses

Let $\xi_\theta(x), x \in \mathbb{R}$ be a density with respect to some $\sigma$-finite measure, $\nu$, on $\mathbb{R}$, and for each $\theta \in \mathbb{R}$, set

$$
b(\theta) \equiv \log \int_{\mathbb{R}} \xi_\theta(x) \exp\{\theta x\} \nu(dx).
$$

Suppose that the effective domain of $b(\cdot)$, $\Theta \equiv \{\theta \in \mathbb{R} : b(\theta) < \infty\}$, is an open interval and that $X \equiv \{X_n : n \in \mathbb{N}\}$ is a sequence of i.i.d random variables with density

$$
\xi_\theta(x) = \xi_0(x) \exp\{\theta x - b(\theta)\}, \quad x \in \mathbb{R}
$$

with respect to $\nu$ for some $\theta \in \Theta$. For any $\theta \in \Theta$, denote by $P_\theta$ and $E_\theta$ the probability measure and expectation of $X$ when the density is $\xi_\theta$.

Given arbitrary $\theta_0, \theta_1 \in \Theta$ such that $\theta_0 < \theta_1$, consider the one-sided testing problem

$$
H_0 : \theta \leq \theta_0 \quad \text{versus} \quad H_1 : \theta \geq \theta_1.
$$

Note that for any $u, v \in \Theta$, $u > v$, and $n \in \mathbb{N}$, the log-likelihood ratio statistic between $P_u$ and $P_v$ based on the first $n$ observations is equal to

$$
\Lambda_n = \sum_{i=1}^{n} \log \frac{\xi_u(X_i)}{\xi_v(X_i)} = (u - v)S_n - n(b(u) - b(v)).
$$

Thus, the SPRT in (7) can be defined using the sum of the observations, $S_n$, in the place of $\Lambda_n$, and all multistage tests in this work can be defined using the average of the observations, $\bar{X}_n = S_n/n$, in the place of $\Lambda_n$.

If we require that

(i) the maximum type-I error probability does not exceed $\alpha$, and

(ii) the maximum type-II error probability does not exceed $\beta$,

for some $\alpha, \beta \in (0, 1)$, then the class of tests of interest becomes

$$
\mathcal{E}(\alpha, \beta) \equiv \left\{ (T, D) \in \mathcal{E} : \sup_{\theta \leq \theta_0} P_\theta(D = 1) \leq \alpha \right. \quad \text{and} \left. \sup_{\theta \geq \theta_1} P_\theta(D = 0) \leq \beta \right\}.
$$

Since $\{P_\theta : \theta \in \Theta\}$ is stochastically monotone and the acceptance / rejection regions of all multistage tests considered in this work are one-sided intervals, all multistage tests considered in this work belong to $\mathcal{E}(\alpha, \beta)$ if they are designed to control the type-I error probability below $\alpha$ when $\theta = \theta_0$ and the type-II error probability below $\beta$ when $\theta = \theta_1$.

Moreover, the asymptotic results in this work remain valid as long as we replace $\mathcal{L}_1$ with $\mathcal{L}_\theta$, for $i \in \{0, 1\}$, where, for each $\theta \in \Theta$, $\mathcal{L}_\theta(\alpha, \beta)$ denotes the smallest expected sample size under $P_\theta$ in $\mathcal{E}(\alpha, \beta)$, i.e.,

$$
\mathcal{L}_\theta(\alpha, \beta) \equiv \inf_{(T, D) \in \mathcal{E}(\alpha, \beta)} E_\theta[T].
$$

In view of the fact (see, e.g., [23, Chapter 5.4]) that, as $\alpha, \beta \to 0$,

$$
\sup_{\theta \leq \theta_0} \mathcal{L}_\theta(\alpha, \beta) \sim \mathcal{L}_{\theta_0}(\alpha, \beta) \sim \frac{\log \beta}{I(\theta_0, \theta_1)},
$$

$$
\sup_{\theta \geq \theta_1} \mathcal{L}_\theta(\alpha, \beta) \sim \mathcal{L}_{\theta_1}(\alpha, \beta) \sim \frac{\log \alpha}{I(\theta_1, \theta_0)},
$$

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
where for any $u, v \in \Theta$ we set
\[
I(u, v) \equiv E_u \left[ \log \frac{\xi_u(X_1)}{\xi_v(X_1)} \right] = (u - v)b'(u) - (b(u) - b(v)),
\]
we conclude that the GMT in this context achieves the optimal worst-case expected sample size under both hypotheses as the error probabilities go to zero at arbitrary rates.

**Remark 42:** If, in addition to (i) we require that
(ii') the maximum expected sample size under $H_0$ does not exceed some $M \in \mathbb{N}$, instead of (ii), then the parameters of each of the above tests will depend only on $\theta_0, \alpha, M$. In this case, all these tests can be implemented not only with “limited knowledge of the alternative distribution” [21, Section V.B], but without any knowledge of the alternative distribution. Indeed, in this formulation, the alternative hypothesis can take the form $H_1 : \theta > \theta_0$. We stress, however, that this is the case for all tests in this work, i.e., this is not a special property of ST or of any other multistage test.

**IX. CONCLUSION AND OPEN PROBLEMS**

In this work, we propose a high-dimensional signal recovery problem that generalizes the one considered in [21]. Specifically, as in the latter work, we consider multiple, independent data streams, each generating iid data, pose the same binary testing problem for each of them, and require that the decision for each of these testing problems be based only on observations from the corresponding data stream. Moreover, we do not assume that the exact numbers of signals and noises are a priori known, but we allow for the incorporation of upper bounds on them. We consider an asymptotic regime in which the maximum numbers of signals and noises go to infinity, while the target classical or generalized familywise error probabilities of both types are fixed.

For this problem, we introduce a novel multistage test that achieves asymptotically the optimal, average over all data streams, expected sample size uniformly in the unknown number of signals. Moreover, we show that the multistage test proposed in [21], as well as a modification of it that we introduce in this work, achieve the same asymptotic optimality property only subject to a sparsity condition on the maximal number of signals and noises. These theoretical results are supported by various simulation studies.

The theoretical results in the high-dimensional setup are based on an asymptotic analysis for the corresponding binary testing problem as the type-1 and type-II error probabilities go to zero. For this problem, we show that GMT achieves the optimal expected sample size under both hypotheses, among all sequential tests with the same error control, as the two error probabilities decay at arbitrary rates. To the best of our knowledge, this is the first multistage test in the literature with this property. On the other hand, ST and its modification are introduced in this work, achieve the same asymptotic optimality property only subject to a sparsity condition on the maximal numbers of signals and noises, and/or they may be more computationally tractable, especially in the case of dependent data.

Finally, one can combine the multiple testing problem we consider here with other error metrics, such as false discovery/non-discovery rates [27], whereas another direction of interest is a “centralized” formulation of the multiple testing problem, which allows using observations from all data streams to decide when to stop sampling for each stream and which hypothesis to select for each testing problem [28], [29], [30], [31], [32], [33]. Even under the assumption of independence among the various data streams, such information can be useful in the presence of non-trivial upper bounds on the numbers of signals and noises or in the case of generalized error control.

**ACKNOWLEDGMENT**

The authors would like to thank two anonymous referees for their constructive feedback.

**APPENDIX A**

**PROOFS FOR THE FIXED-SAMPLE-SIZE TEST**

**Proof of Theorem 3:** For any $\alpha, \beta \in (0, 1)$ and $c \in (-I_0, I_1)$, we define
\[
n^*(\alpha, \beta, c) \equiv \min \{n \in \mathbb{N} : P_0(\bar{\Lambda}_n > c) \leq \alpha \text{ and } P_1(\bar{\Lambda}_n \leq c) \leq \beta \}.
\]
Then,
\[
n^*(\alpha, \beta) \leq \min_{c \in (-I_0, I_1)} n^*(\alpha, \beta, c).
\]
It then suffices to show that, for any $\alpha, \beta \in (0, 1)$ and $c \in (-I_0, I_1)$,
\[
n^*(\alpha, \beta, c) \leq \max \left\{ \frac{\log \beta}{\psi_1(c)}, \frac{\log \alpha}{\psi_0(c)} \right\} + 1. \tag{102}
\]
Indeed, this implies that
\[
n^*(\alpha, \beta) \leq \min_{c \in (-I_0, I_1)} \max \left\{ \frac{\log \alpha}{\psi_0(c)}, \frac{\log \beta}{\psi_1(c)} \right\} + 1.
\]
Since $\psi_0$ (resp. $\psi_1$) is strictly increasing (resp. strictly decreasing) and continuous in $[-I_0, I_1]$, $\psi_0(-I_0) = 0$, $\psi_0(I_1) = 0$.
$I_1 > 0$, $\psi_1(-I_0) = I_0 > 0$, $\psi_1(I_1) = 0$, the minimum is attained when the two terms in the maximum are equal, i.e., at $c = g^{-1}(\log \alpha/|\log \beta|)$, which proves (21). The inequality (22) then follows by noticing that

$$h_0(\alpha, \beta) \lor h_1(\alpha, \beta) \geq \min_{c \in (-I_0, I_0)} \{\psi_0(c) \lor \psi_1(c)\} = C.$$

To prove (102), we fix $\alpha, \beta \in (0, 1)$ and $c \in (-I_0, I_1)$ and, for ease of notation, we write $n^*(\alpha, \beta, c)$ in short as $n^*$. Then, by the definition of minimum,

either $P_1(\bar{A}_{n^* - 1} \leq c) > \beta$ or $P_0(\bar{A}_{n^* - 1} > c) > \alpha$,

which implies

$$1 \leq \max \left\{ \frac{\log \beta}{\log P_1(\bar{A}_{n^* - 1} \leq c)} , \frac{\log \alpha}{\log P_0(\bar{A}_{n^* - 1} > c)} \right\} ,$$

or equivalently

$$n^* - 1 \leq \max \left\{ \frac{\log \beta}{\log P_1(\bar{A}_{n^* - 1} \leq c)/(n^* - 1)} , \frac{\log \alpha}{\log P_0(\bar{A}_{n^* - 1} > c)/(n^* - 1)} \right\} .$$

Applying the inequalities in (16) completes the proof.

**Proof of Corollary 4:** We only prove (i), as the proof of (ii) is similar. By the properties of $\psi_1$ and $g$ it follows that, as $\alpha, \beta \to 0$ so that $|\log \alpha| \ll |\log \beta|$, $h_1(\alpha, \beta) \to \psi_1\left(g^{-1}(0+)) = \psi_1(-I_0+) = I_0.$

Thus, by (21), $n^*(\alpha, \beta) \leq |\log \beta|/I_0.$

**APPENDIX B**

**Proofs for the 3-Stage Test**

**Proof of Proposition 9:** When $N_{0,0} \leq N_{1,0}$,

$$E[\bar{T}] = N_{0,0} + (N_{1,0} - N_{0,0}) P\left(\bar{A}_{N_{0,0}} > C_{0,0}\right) + (N - N_{1,0}) P\left(\bar{A}_{N_{0,0}} > C_{0,0}, \bar{A}_{N_{0,0}} \leq C_{1,0}\right) ,$$

whereas when $N_{0,0} \geq N_{1,0}$,

$$E[\bar{T}] = N_{1,0} + (N_{0,0} - N_{1,0}) P\left(\bar{A}_{N_{1,0}} \leq C_{1,0}\right) + (N - N_{0,0}) P\left(\bar{A}_{N_{1,0}} \leq C_{1,0}, \bar{A}_{N_{1,0}} > C_{0,0}\right) .$$

By the fact that $P(A) - P(B^c) \leq P(A \cap B) \leq P(A)$ for any events $A, B$, one can derive that

$$N_{0,0} \left(1 - P\left(\bar{A}_{N_{0,0}} > C_{1,0}\right)\right) + (N - N_{0,0}) \left(P\left(\bar{A}_{N_{0,0}} > C_{0,0}, \bar{A}_{N_{1,0}} > C_{1,0}\right)\right) \leq E[\bar{T}] \leq N_{0,0} + (N - N_{0,0}) \left(P\left(\bar{A}_{N_{0,0}} > C_{0,0}\right)\right)$$

and

$$N_{1,0} \left(1 - P\left(\bar{A}_{N_{1,0}} \leq C_{0,0}\right)\right) + (N - N_{1,0}) \left(P\left(\bar{A}_{N_{1,0}} \leq C_{1,0}, \bar{A}_{N_{0,0}} \leq C_{0,0}\right)\right) \leq E[\bar{T}] \leq N_{1,0} + (N - N_{1,0}) P\left(\bar{A}_{N_{1,0}} \leq C_{1,0}\right) .$$

Plugging (31) in completes the proof.

**Proof of Theorem 7:** We prove the asymptotic optimality of 3ST only under the null hypothesis, as the corresponding proof under the alternative is similar. For any $\alpha, \beta \in (0, 1)$ and $\epsilon_0, \epsilon \in (0, 1)$, by (110) and the selection of $N_{0,0}$ in (35) we have

$$E_0[\bar{T}(\alpha, \beta)] \leq \begin{cases} N_{0,0} , & \text{if } \tilde{N}_{0,0} \geq N , \\ N_{0,0} + (N - \tilde{N}_{0,0}) P\left(\bar{A}_{N_{0,0}} > \tilde{C}_{0,0}\right) , & \text{otherwise} \end{cases}$$

where

$$\tilde{C}_{0,0} = -\frac{|\log(\beta/2)|}{N_{0,0}} \leq -\frac{|\log(\beta/2)|}{(1 - \epsilon_0) I_0} = -(1 - \epsilon_0) I_0 .$$

Moreover, by (22) we have

$$N = n^*(\alpha/2, \beta/2) \leq \frac{|\log((\alpha \lor \beta)/2)|}{C} + 1 .$$

Thus,

$$E_0[\bar{T}(\alpha, \beta)] \leq \left(\frac{|\log(\beta/2)|}{(1 - \epsilon_0) I_0} + 1\right)$$

$$+ \left(\frac{|\log((\alpha \lor \beta)/2)|}{C} + 1\right) P\left(\bar{A}_{N_{0,0}} > -(1 - \epsilon_0) I_0\right)$$

$$\leq \frac{|\log(\beta/2)|}{I_0} \left(\frac{1}{1 - \epsilon_0}\right) + \frac{|\log((\alpha \lor \beta)/2)|}{|\log(\beta/2)|} \frac{I_0}{C} P\left(\bar{A}_{N_{0,0}} > -(1 - \epsilon_0) I_0\right) + 2 ,$$

and, consequently, $E_0[\bar{T}(\alpha, \beta)] \leq |\log \beta|/I_0$ as $\alpha, \beta \to 0$ so that $|\log \alpha| = \Theta(|\log \beta|)$ as long as $\epsilon_0 = \epsilon_0(\alpha, \beta)$ satisfies (36).

It remains to show that it is possible to find such an $\epsilon_0$ under our standing assumption (2). Indeed, by (16), for any $\epsilon_0 \in (0, 1)$ we have

$$P_0\left(\bar{A}_{N_{0,0}} > -(1 - \epsilon_0) I_0\right)$$

$$\leq \exp\left\{-\tilde{N}_{0,0} \psi_0(-(1 - \epsilon_0) I_0)\right\}$$

$$\leq \exp\left\{-\frac{|\log(\beta/2)|}{I_0} \psi_0(-(1 - \epsilon_0) I_0)\right\} ,$$

thus, it suffices to find an $\epsilon_0 = \epsilon_0(\beta)$ so that $\epsilon_0 \to 0$ and $|\log(\beta/2)| \psi_0(-(1 - \epsilon_0) I_0) \to \infty$ as $\beta \to 0$.

Since $\psi_0$ is strictly increasing in $[-I_0, \infty)$ with $\psi_0(-I_0) = 0$, this is the case, for example, when $\epsilon_0 = \epsilon_0(\beta)$ satisfies $\psi_0(-(1 - \epsilon_0) I_0) = (|\log(\beta/2)|)^{-1/2} \land C$.

**Proof of Theorem 8:** In the Gaussian mean testing problem of Subsection II-E, by the form of $\psi_0$ in (28), (106) translates to

$$E_0[\bar{T}(\alpha, \beta)] \leq \frac{|\log(\beta/2)|}{I_0} \left(\frac{2}{1 - \epsilon_0} + 4\frac{|\log((\alpha \lor \beta)/2)|}{|\log(\beta/2)|} \exp\left\{-\frac{|\log(\beta/2)|\epsilon_0^2}{4(1 - \epsilon_0)}\right\}\right) + 2 .$$
Assuming \( \epsilon_0 < 1/2 \) and \( 4|\log((\alpha \wedge \beta)/2)|/|\log(\beta/2)| \leq C \) as \( \alpha, \beta \to 0 \) where \( C \) is a constant that does not depend on \( \alpha \) or \( \beta \), we have

\[
E_0[\tilde{T}(\alpha, \beta)] \leq \frac{|\log(\beta/2)|}{(1 + 2\epsilon_0 + C \exp \{-\frac{1}{4}|\log(\beta/2)|\epsilon_0^2\})} + 2.
\]

If we select \( \epsilon_0 \) to satisfy (38), then the first line of (39) follows. \( \square \)

**APPENDIX C**

**PROOFS FOR THE GENERAL MULTISTAGE TEST**

**Proof:** [Proof of Theorem 17] We only prove the asymptotic optimality under the null hypothesis as that under the alternative hypothesis can be proved similarly.

By the selection of \( N_{0,0} \) in (60), it suffices to show that there exists a selection of \( N_{0,0} \), as a function of \( \alpha \) and \( \beta \), in the range of \( \{1, \ldots, N_{0,1} \wedge N\} \), where \( N_{0,1} \) and \( N \) are given by (57) and (52), so that (55) is asymptotically upper bounded by \( |\log(\beta)/I_0| \) as \( \alpha, \beta \to 0 \). We will show that

\[
N_{0,0} = \tilde{N}_{0,0} \wedge N_{0,1} \wedge N
\]

is such a selection, where

\[
\tilde{N}_{0,0} = \left\lfloor \frac{|\log(\beta/5)|}{(1 - \epsilon_0)I_0} \right\rfloor,
\]

and \( \epsilon_0 = \epsilon_0(\beta) \) is some function of \( \beta \) that satisfies

\[
\epsilon_0 \to 0 \text{ and } P_0(\tilde{\Lambda}_{N_{0,0}} > -(1 - \epsilon_0)I_0) \to 0 \text{ as } \beta \to 0.
\]

Note that (108) is the same as (36) with \( \tilde{N}_{0,0} \) replaced by \( \tilde{N}_{0,0} \). Similarly as in the proof of Theorem 7, we can always find such an \( \epsilon_0 = \epsilon_0(\beta) \) under assumption (2).

In what follows, we show that the selection of \( N_{0,0} \) in (107) suffices for our purpose. Indeed, (55) is upper bounded by

\[
N_{0,0} + N_{0,1} P_0(\tilde{\Lambda}_{N_{0,0}} > \tilde{C}_{0,0}) + \sum_{j=2}^{K_0} N_{0,K_0} P_0(\tilde{\Lambda}_{N_{0,j}} > C_{0,j}) + N P_0(\tilde{\Lambda}_{N_{0,K_0}} > C_{0,K_0}),
\]

where

\[
\tilde{C}_{0,0} = \epsilon_0^{\frac{\tilde{N}_{0,0}}{\beta/2 - \sum_{j=1}^{K_0} (\beta/5)^j} + 1} \geq \epsilon_0^{\tilde{N}_{0,0}/\beta} \geq \frac{|\log(\beta/5)|}{\tilde{N}_{0,0}} \geq \frac{|\log(\beta/5)|}{|\log(\beta/5)|/(1 - \epsilon_0)I_0} = -(1 - \epsilon_0)I_0.
\]

From (16) it follows that, for any \( n \in \mathbb{N} \),

\[
P_1(\tilde{\Lambda}_n \leq 0) \leq \exp\{-n\psi_0(0)\} = \exp\{-nC\},
\]

\[
P_0(\tilde{\Lambda}_n \geq 0) \leq \exp\{-n\psi_0(0)\} = \exp\{-nC\}.
\]

Thus, for any \( j \in \mathbb{N} \), by the selection of \( N_{0,j} \) in (59) we have

\[
P_1(\tilde{\Lambda}_{N_{0,j}} \leq 0) \leq \exp\{-N_{0,j}C\} \leq (\beta/5)^j.
\]

By this equation and the selection of \( C_{0,j} \) in (53) we have

\[
P_0(\tilde{\Lambda}_{N_{0,j}} > C_{0,j}) \leq P_0(\tilde{\Lambda}_{N_{0,j}} > 0) \leq \exp\{-N_{0,j}C\} \leq (\beta/5)^j.
\]

Moreover, by the selection of \( K_0 \) in (61) it follows that

\[
N \leq N_{0,K_0+1} = (K_0 + 1)(|\log(\beta/5)/C|) \leq (K_0 + 1)(|\log(\beta/5)/C + 1|).
\]

Therefore, (109) can be upper bounded by

\[
\begin{align*}
&\frac{|\log(\beta/5)|}{I_0} \left( \frac{1}{1 - \epsilon_0} \right) + I_0 \left( P_0(\tilde{\Lambda}_{N_{0,0}} > -(1 - \epsilon_0)I_0) + \sum_{j=2}^{\infty} j (\beta/5)^{j-1} \right) \\
&+ \left( 2 + \sum_{j=2}^{\infty} j (\beta/5)^{j-1} \right) \\
&\leq \frac{|\log(\beta/5)|}{I_0} \left( \frac{1}{1 - \epsilon_0} \right) + I_0 \left( P_0(\tilde{\Lambda}_{N_{0,0}} > -(1 - \epsilon_0)I_0 + \beta) \right) + (2 + \beta),
\end{align*}
\]

which is asymptotically upper bounded by \( |\log(\beta)/I_0| \) as \( \alpha, \beta \to 0 \) by (108).

**Proof of Proposition 18:** We only prove the asymptotic upper bound under the null hypothesis, as the corresponding upper bound under the alternative can be proved similarly.

With the same selection of \( N_{0,0} \) in (107), by the form of \( \psi_0 \) in (28), (110) translates to

\[
E_0[\tilde{T}(\alpha, \beta)] \leq \frac{|\log(\beta/5)|}{I_0} \left( 1 + 2\epsilon_0 + 4 \exp \left\{-\frac{1}{4}|\log(\beta/5)|\epsilon_0^2\right\} + \beta \right) + (2 + \beta),
\]

assuming \( \epsilon_0 < 1/2 \). If we select \( \epsilon_0 \) to satisfy (38), then the first line of (63) follows. \( \square \)

**APPENDIX D**

**PROOFS FOR THE SEQUENTIAL THRESHOLDING AND ITS MODIFICATION**

We start with a lemma that provides non-asymptotic upper bounds for the stage sizes of ST and mod-ST, as well as for the expected sample sizes of the two tests under the null hypothesis.

**Lemma 43:** Suppose that the parameters of ST and mod-ST are selected according to (67)-(69).
(i) For the stage sizes of ST we have, for every $1 \leq j \leq K$,
\[
m_j \leq n^* \left( \alpha^{1/K}, (\beta/2)^j \right) \leq j \frac{|\log(\beta/2)|}{h_1(\alpha^{1/K}, \beta/2)} + 1. \tag{111}
\]

(ii) For the stage sizes of mod-ST we have, for every $1 \leq j \leq K$,
\[
m_1 + \ldots + m_j \leq n^* \left( \alpha^{j/K}, (\beta/2)^j \right) \leq j \frac{|\log(\beta/2)|}{h_1(\alpha^{1/K}, \beta/2)} + 1. \tag{112}
\]

(iii) Both $E_0[T'(\alpha, \beta)]$ and $E_0[T''(\alpha, \beta)]$ are bounded above by
\[
\frac{|\log(\beta/2)|}{h_1(\alpha^{1/K}, \beta/2)} \left( 1 - \alpha^{1/K} \right)^{-2} + \left( 1 - \alpha^{1/K} \right)^{-1} \tag{113}
\]

Proof: (i) The second inequality in (111) follows by (21) and the fact that $h_1(\cdot, \cdot)$ decreases in its second argument. For the first inequality, when $j \in \{2, \ldots, K\}$, it holds with equality by the selection of $(m_j, b_j)$ according to (71), and when $j = 1$, it holds by the selection of $(m_1, b_1)$ according to (70) and the fact that
\[
\sum_{j=2}^{\infty} (\beta/2)^j = \frac{(\beta/2)^2}{1 - \beta/2} \leq \beta/2.
\]

(ii) The second inequality in (112) follows by (21) and the fact that
\[
h_1(x^r, y^r) = h_1(x, y), \quad \forall x, y \in (0, 1), r > 0.
\]

To show the first inequality, when $j = 1$, this is the same as in (i). When $j \geq 2$, we use mathematical induction. Suppose that the first inequality in (112) holds for some $1 < j \leq K-1$.

Then by (72) we obtain
\[
M_{j-1} + m_j \leq n^* \left( \alpha^{(j-1)/K}, (\beta/2)^{j-1} \right) \leq n^* \left( \alpha^{j/K}, (\beta/2)^j \right),
\]

where the equality follows from the definition of $n^*(\cdot, \cdot)$ in (13).

(iii) For both ST and mod-ST, the expected sample size under $P$ is of form
\[
m_1 + \sum_{j=2}^{K} m_j \mathbb{P} \left( \bigcap_{i=1}^{j-1} \Lambda'_i > b_i \right). \tag{114}
\]

From (69), when $P = P_0$ this is upper bounded by
\[
m_1 + \sum_{j=2}^{K} m_j \alpha^{(j-1)/K},
\]

and, from (i) and (ii) of the lemma, the latter is further upper bounded by
\[
\frac{|\log(\beta/2)|}{h_1(\alpha^{1/K}, \beta/2)} \sum_{j=1}^{K} \alpha^{(j-1)/K} + \sum_{j=1}^{K} \alpha^{(j-1)/K}. \tag{115}
\]

Replacing the upper limit $K$ in each sum by $\infty$, we obtain (113).

Proof of Theorem 20: When $K$ is selected so that (73) holds, by Corollary 4 and Lemma 43(iii) we conclude that
\[
E_0[T'(\alpha, \beta)], E_0[T''(\alpha, \beta)] \lesssim \frac{|\log \beta|}{I_0},
\]

which proves the asymptotic optimality of both ST and mod-ST under the null hypothesis.

Moreover, since for both tests it is possible to reject the null hypothesis only at the last stage, the expected sample size under the alternative is lower bounded by $(1 - \beta), (m_1 + \ldots + m_K)$ and, as a result, it is equal, to a first-order asymptotic approximation as $\beta \to 0$, to the maximum possible sample size, $m_1 + \ldots + m_K$.

From Lemma 43(ii) with $j = K$, we have, as $\alpha, \beta \to 0$ so that (73) holds, the following asymptotic upper bound on the expected sample size of mod-ST under the alternative:
\[
E_1[T''] \sim \sum_{j=1}^{K} m_j \lesssim K \frac{|\log \beta|}{I_0}. \tag{116}
\]

From Lemma 43(i), we have that, as $\alpha, \beta \to 0$ so that (73) holds,
\[
m_j \lesssim j \frac{|\log \beta|}{I_0}, \quad \forall 1 \leq j \leq K.
\]

Meanwhile, from (70)-(71) it follows that, for any $\alpha, \beta \in (0, 1)$,
\[
m_1 \geq n^*(\alpha^{1/K}, 1) \geq L_0(\alpha^{1/K}, \beta)
\]

and, for every $2 \leq j \leq K$,
\[
m_j = n^*(\alpha^{1/K}, (\beta/2)^j) \geq L_0(\alpha^{1/K}, (\beta/2)^j).
\]

Then, by (8) it follows that, as $\alpha, \beta \to 0$ so that $\alpha^{1/K} \to 0$,
\[
m_j \gtrsim j \frac{|\log \beta|}{I_0}, \quad \forall 1 \leq j \leq K.
\]

Therefore,
\[
E_1[T'] \sim \sum_{j=1}^{K} m_j \sim \frac{|\log \beta|}{I_0} \sum_{j=1}^{K} j \frac{|\log \beta| K(K+1)}{2},
\]

which, in view of (74) and (8), implies
\[
E_1[T'] \gg \frac{K + 1}{2} L_1(\alpha, \beta),
\]
and, in view of (116),
\[ E_1[T'] \lesssim \frac{2}{K+1} E_1[T'(\alpha, \beta)]. \]

**Proof of Proposition 21:** By (29) and (113) we have
\[ E_0[T'], E_0[T'''] \leq \left| \log(\beta/2) \right| \left( 1 + \frac{\left| \log \alpha \right|}{\log(\beta/2)} \right)^2 \left( 1 - \frac{\alpha^{1/K}}{1} \right)^2 + \left( 1 - \frac{\alpha^{1/K}}{1} \right)^{-1}. \]

If \( \alpha^{1/K} \) satisfies (73) as \( \alpha, \beta \to 0 \), then
\[ E_0[T'], E_0[T'''] \lesssim \left| \log(\beta/2) \right| \left( 1 + 2 \sqrt{\frac{\left| \log \alpha \right|}{\log(\beta/2)}} + 2 \alpha^{1/K} \right). \]

If, further, \( \alpha^{1/K} \) is selected as in (77), we obtain (78). □

**APPENDIX E**

**Proofs for the High-Dimensional Signal Recovery Problem**

**Proof of Lemma 23:** Fix \( m \in \mathbb{N} \) and \( \alpha, \beta \in (0,1) \). For any \( A \subseteq [m] \) with \( l_m \leq |A| \leq u_m \), \( j \in [m], i \in \{0,1\} \) and \( (T, D) \in \mathcal{E} \), by the independence of streams and the decentralized nature of the multiple testing procedure, we have
\[ P_A(D^j = i) = \begin{cases} P_1(D = i), & \text{if } j \in A \\ P_0(D = i), & \text{if } j \in A^c, \end{cases} \]
thus,
\[ FWE-I_A(T, D) = P_A \left( \bigcup_{j \in A^c} \{ D^j = 1 \} \right) = 1 - P_A \left( \bigcap_{j \in A^c} \{ D^j = 0 \} \right) = 1 - (1 - P_0(D = 1))^{m-|A|} \leq 1 - (1 - P_0(D = 1))^{m-l_m}, \]
and, similarly,
\[ FWE-I_A(T, D) \leq 1 - (1 - P_1(D = 0))^{u_m}. \]
Therefore, \( (T, D) \in \mathcal{E}(\alpha, \beta) \) if and only if
\[ 1 - (1 - P_0(D = 1))^{m-l_m} \leq \alpha, \quad 1 - (1 - P_1(D = 0))^{u_m} \leq \beta, \]
or, equivalently,
\[ P_0(D = 1) \leq 1 - (1 - \alpha)^{1/(m-l_m)} \equiv \alpha_m, \quad P_1(D = 0) \leq 1 - (1 - \beta)^{1/u_m} \equiv \beta_m, \]
i.e. \( (T, D) \in \mathcal{E}(\alpha_m, \beta_m) \).

To facilitate later usage, we state the following lemma.

**Lemma 44:** Fix \( \alpha, \beta \in (0,1) \) and, for any \( m \in \mathbb{N} \), let \( \alpha_m, \beta_m \) be defined as in (82).

(i) If as \( m \to \infty, u_m \to \infty \), then \( \beta_m = \Theta(\beta/u_m) \) and \( |\log \beta_m| \sim \log u_m \).
(ii) If as \( m \to \infty, m-l_m \to \infty \), then \( \alpha_m = \Theta(\alpha/(m-l_m)) \) and \( |\log \alpha_m| \sim (m-l_m) \).
(iii) If as \( m \to \infty, u_m, m-l_m \to \infty \), then \( L_0(\alpha_m, \beta_m) \sim \log u_m/I_0 \) and \( L_1(\alpha_m, \beta_m) \sim \log (m-l_m)/I_1 \).

**Proof of Lemma 44:** (i) Fix \( \beta \in (0,1) \) and suppose that \( u_m \to \infty \) as \( m \to \infty \). \( |\log \beta_m| \sim \log u_m \) follows from
\[ \beta_m \equiv 1 - (1 - \beta)^{1/u_m} = \Theta(\beta/u_m), \]
i.e., when there are constants \( 0 < c < C \) so that
\[ c\beta/n \leq 1 - (1 - \beta)^{1/n} \leq C\beta/n \]
for all large \( n \in \mathbb{N} \). To show the latter, we observe that the above inequalities are equivalent to
\[ (1 - C\beta/n)^n \leq 1 - \beta \leq (1 - c\beta/n)^n. \]
Since the lower (resp. upper) bound converges to \( e^{-C\beta} \) (resp. \( e^{-c\beta} \)) as \( n \to \infty \), it is clear that we can find \( 0 < c < C \), depending on \( \beta \), so that the above inequality holds for all large \( n \).

(ii) This is analogous to (i).
(iii) This follows from (i), (ii) and (8).

**Proof of Theorem 25:** (i) By Lemma 44, \( u_m, m-l_m \to \infty \) implies \( \alpha_m, \beta_m \to 0 \), and thus (85) follows from (12).
(ii) A sufficient condition for the definition of asymptotic optimality in the high-dimensional sense, i.e., (84), is, for any \( \alpha, \beta \in (0,1) \), as \( u_m, m-l_m \to \infty \),
\[ \sup_{\pi \in [l_m/m, u_m/m]} \frac{E_\pi[T^*(\alpha_m, \beta_m)]}{L(\alpha_m, \beta_m)} \to 1. \]
When (86)-(87) hold, this follows from (12).
(iii) For any \( \pi \) in \([l_m/m, u_m/m]\), we have
\[ \frac{E_\pi[T^*(\alpha_m, \beta_m)]}{L(\alpha_m, \beta_m)} \leq \frac{\left( 1 - \pi \right) E_0[T^*(\alpha_m, \beta_m)] + \pi E_1[T^*(\alpha_m, \beta_m)]}{L(\alpha_m, \beta_m)} \leq \frac{E_0[T^*(\alpha_m, \beta_m)]}{L(\alpha_m, \beta_m)} + \frac{u_m E_1[T^*(\alpha_m, \beta_m)]}{(m-u_m) L(\alpha_m, \beta_m)}. \]

By Lemma 44, \( u_m, m-l_m \to \infty \) implies \( \alpha_m, \beta_m \to 0 \) and \( L(\alpha_m, \beta_m) \sim \log u_m/I_0 \). Therefore, the first term in (117) goes to 1 by (86) and the second term goes to 0 by (88).

**Proof of Corollary 26:** According to (8) (resp. Theorem 17), the family of SPRTs (resp. GMTs) satisfies (86)-(87).

**Proof of Corollary 27:** According to Theorem 7, the family of 3STs satisfies (86)-(87) if also \( |\log \alpha_m| = \Theta(|\log \beta_m|) \) as \( u_m, m-l_m \to \infty \), which, from Lemma 44, is equivalent to \( \log (m-l_m) = \Theta(\log u_m) \). To prove Corollary 28, we need the following lemma.

**Lemma 45:** If \( u_m \ll m \to \infty \), then
\[ \frac{u_m}{\log u_m} \ll \frac{m}{\log m} \text{ as } m \to \infty. \]
Proof: For $u_m > e$ we have
\[
\frac{u_m}{\log{u_m}} = \frac{u_m}{\log{u_m}} \cdot 1\{u_m < \sqrt{m}\} + \frac{u_m}{\log{u_m}} \cdot 1\{u_m \geq \sqrt{m}\}
\leq \sqrt{m} \cdot 1\{u_m < \sqrt{m}\} + \frac{2u_m}{\log{m}} \cdot 1\{u_m \geq \sqrt{m}\}
\leq \max\left\{\sqrt{m}, \frac{2u_m}{\log{m}}\right\},
\]
where $1\{\cdot\}$ is the indicator function. As $m \to \infty$, $\sqrt{m} \ll m/\log m$. If $u_m \ll m$ as $m \to \infty$, then $2u_m/\log m \ll m/\log m$. The proof is complete.

Proof of Corollary 28: (i) When $u_m \to \infty$ and $u_m \ll m$ as $m \to \infty$, by (85) and Lemma 45 we have,
\[
\frac{\log u_m}{I_0} \sim \left(1 - \frac{u_m}{m}\right) \frac{\log u_m}{I_0} \leq \mathcal{L}_{s/m}(\alpha, \beta m)
\leq \frac{\log u_m}{I_0} + \frac{u_m \log m}{I_1} \sim \frac{\log u_m}{I_0}
\]
uniformly in $s \in \{l_1, \ldots, u_n\}$.

(ii) By Theorem 25, it suffices to show that, for any $\alpha, \beta \in (0,1)$, (86) and (88) hold for $\chi'$ and $\chi''$. To this end, we first observe that $u_m \to \infty$ and $u_m \ll m$ as $m \to \infty$ imply that $m - l_m \sim m - \infty$, since $m \geq m - l_m \geq m - u_m \sim m$.

By Theorem 20, (86) is satisfied by $\chi'$ (resp. $\chi''$) as long as for every $m \in \mathbb{N}$, the maximum number of stages, $K_m$ (resp. $K''_m$), can be selected so that as $m \to \infty$,
\[
\frac{\log \alpha_m}{\log \beta_m} \ll K'_m \quad \text{resp.} \quad K''_m \ll \log \alpha_m.
\]
In view of Lemma 44, this is equivalent to
\[
\frac{\log m}{\log u_m} \ll K'_m \quad \text{resp.} \quad K''_m \ll \log m,
\]
which is always feasible when $u_m \to \infty$.

By (117) and Lemma 44 we have
\[
\mathbb{E}[T'(\alpha_m, \beta_m)] \sim \frac{K'_m (K'_m + 1)}{2} \frac{\log \beta_m}{I_0} \sim \frac{K'_m (K'_m + 1) \log u_m}{I_0} \leq K''_m \log u_m.
\]
Thus, (88) is satisfied by $\chi'$ if
\[
K''_m \log u_m \ll \frac{(m - u_m) \log u_m}{u_m} \sim m \log u_m,\]
or, equivalently,
\[
K'_m \ll \sqrt{\frac{m}{u_m}}.
\]
(119)

Similarly, by (118) and Lemma 44 we have
\[
\mathbb{E}[T''(\alpha_m, \beta_m)] \leq K''_m \frac{\log \beta_m}{I_0} \sim K''_m \log u_m,
\]
thus, (88) is satisfied by $\chi''$ if
\[
K''_m \ll \frac{m}{u_m}.
\]
(120)

A condition that guarantees the existence of $(K'_m)$ that satisfies (118) and (119) simultaneously is
\[
\frac{\log m}{\log u_m} \ll \sqrt{\frac{m}{u_m}},
\]
(121)
while a condition that guarantees the existence of $(K''_m)$ that satisfies (118) and (120) simultaneously is
\[
\frac{\log m}{\log u_m} \ll \frac{m}{u_m}.
\]
(122)

By Lemma 45, both (121) and (122) are implied by the condition $u_m \ll m$. The proof is complete.

Proof of Lemma 31: Fix $m \in \mathbb{N}$ and $\alpha, \beta \in (0,1)$. Recall the definition of $\mathcal{B}(\cdot, \cdot, \cdot)$, that is used in the definition of $\alpha^G_m$ and $\beta^G_m$ in (91)-(92), which is increasing in its first argument when the other two arguments are fixed.

For any $A \subseteq [m]$ with $l_m \leq |A| \leq u_m$ and $(T, D) \in \mathcal{E}$ we have
\[
\kappa_m - GFWE-I(A, T, D) = P_A((T, D) \text{ makes at least } \kappa_m \text{ type-I errors})
\]
out of the $|A^c|$ noises
\[
\leq \mathcal{B}(|A^c|, P_0(D = 1); \kappa_m) \leq \mathcal{B}(m - l_m, P_0(D = 1); \kappa_m),
\]
and similarly we obtain
\[
\ell_m - GFWE-II(A, T, D) \leq \mathcal{B}(u_m, P_1(D = 0); \ell_m).
\]
Therefore, $(T, D) \in \mathcal{E}^G(\alpha, \beta)$ if and only if
\[
\mathcal{B}(m - l_m, P_0(D = 1); \kappa_m) \leq \alpha
\]
\[
\mathcal{B}(u_m, P_1(D = 0); \ell_m) \leq \beta,
\]
i.e., $(T, D) \in \mathcal{E}(\alpha^G_m, \beta^G_m)$.

We first establish a pair of lower and upper bounds on $\alpha^G_m$ and $\beta^G_m$, respectively.

Lemma 46: Let $\alpha, \beta \in (0,1/2)$, $\ell_m \leq u_m/2$, and $\kappa_m \leq (m - l_m)/2$, then
\[
\frac{1}{e} \frac{\kappa_m}{m - l_m} \alpha^{1/\kappa_m} \leq \alpha^G_m \leq e^2 \frac{\kappa_m}{m - l_m} \alpha^{1/\kappa_m},
\]
(123)
\[
\frac{1}{e} \frac{\ell_m}{u_m} \beta^{1/\ell_m} \leq \beta^G_m \leq e^2 \frac{\ell_m}{u_m} \beta^{1/\ell_m}.
\]
(124)
Thus, for any $\alpha, \beta \in (0,1/2)$, as $(m - l_m)/\kappa_m \to \infty$ and $u_m/\ell_m \to \infty$,
\[
|\log \alpha^G_m| \sim \log \left(\frac{m - l_m}{\kappa_m}\right)
\]
\[
|\log \beta^G_m| \sim \log \left(\frac{u_m}{\ell_m}\right).
\]
(125)

Proof: We only prove (123), as (124) can be proved similarly. Before starting, we note that for any $p \in (0,1)$,
\[
\left(\frac{m - l_m}{\kappa_m}\right) p^{\kappa_m} (1 - p)^{m - l_m}
\]
\[
\leq \mathcal{B}(m - l_m, p; \kappa_m) \leq \left(\frac{m - l_m}{\kappa_m}\right) p^{\kappa_m}.
\]
Moreover, using the bounds
\[
\left(\frac{n}{k}\right)^k \leq \left(\frac{en}{k}\right)^k, \quad \forall 1 \leq k \leq n,
\]
the previous inequalities can be further strengthened as follows:
\[
\left(\frac{m - l_m}{\kappa_m}\right)^{\kappa_m} (1 - p)^{m - l_m}
\]
\[
\leq \mathcal{B}(m - l_m, p; \kappa_m) \leq \left(\frac{e (m - l_m)}{\kappa_m}\right)^{\kappa_m} p^{\kappa_m}. \]
(126)
We first show the lower bound in (123), for which it suffices to show
\[ B(m - l_m, p_m; \kappa_m) \leq \alpha, \text{ where } p_m = \frac{\kappa_m}{e^{\kappa_m/m} - l_m} \alpha^{1/\kappa_m}. \]

This follows by applying the upper bound in (126).

Next we show the upper bound in (123), for which it suffices to show
\[ B(m - l_m, p_m; \kappa_m) > \alpha, \text{ where } p_m = e^2 \frac{\kappa_m}{e^{\kappa_m/m} - l_m} \alpha^{1/\kappa_m}. \]

Indeed, if \( B(m - l_m, p_m; \kappa_m) \geq 1/2 \), then this holds trivially since we have assumed \( \alpha < 1/2 \). If \( B(m - l_m, p_m; \kappa_m) < 1/2 \), then \( \kappa_m \) has to be strictly greater than the median of \( \text{Binomial}(m - l_m, p) \), which is either \( \lceil (m - l_m)p \rceil \) or \( \lfloor (m - l_m)p \rfloor \), thus, we have \( \kappa_m \geq (m - l_m)p_m \). Since we have also assumed \( \kappa_m \leq (m - l_m)/2 \), we have \( p_m \leq 1/2 \) and, thus, \( 1 - p_m \geq e^{-2p_m} \). Applying the lower bound in (126), we have
\[ B(m - l_m, p_m; \kappa_m) \geq \left( \frac{m - l_m}{\kappa_m} \right)^{\kappa_m} p_m^{\kappa_m} e^{-2\kappa_m} \alpha. \]

Finally, we show (125). Note that the lower and upper bounds in (123) or (124) differ by a constant factor, namely \( e^3 \), and \( \alpha^{1/\kappa_m} \in (\alpha, 1) \). Therefore, as long as \( \kappa_m/(m - l_m) \to 0 \) and \( \ell_m/u_m \to 0 \), we have \( \alpha_m^{G, \beta_m^{G}} \to 0 \) and
\[ \alpha_m^{G} = \Theta \left( \frac{\kappa_m}{m - l_m} \right), \quad \beta_m^{G} = \Theta \left( \frac{\ell_m}{u_m} \right). \]

Thus, (125) follows.

\textbf{Proof of Theorem 33:} This is similar to the proof of Theorem 25 and is omitted.

\textbf{Proof of Corollary 35:} This is similar to the proof of Corollary 26 and is omitted.

\textbf{Proof of Corollary 36:} This is similar to the proof of Corollary 27 and is omitted.

\textbf{Proof of Corollary 37:} (i) This can be proved similarly as Corollary 28.(i).

(ii) & (iii) Now we prove the high-dimensional asymptotic optimality under generalized error control for \( \chi' \) and \( \chi'' \). By Theorem 33, it suffices to show that (94) and (96) hold for \( \chi' \) (resp. \( \chi'' \)) under condition (98) (resp. (97)). Similarly to the proof of Corollary 28.(ii), (94) is satisfied for \( \chi' \) (resp. \( \chi'' \)) if for any \( m \in \mathbb{N} \), the maximum number of stages, \( K_m' \) (resp. \( K_m'' \)), is selected so that as \( m \to \infty \),
\[ \log(m/\kappa_m) \sim \log(u_m/\ell_m) \ll K_m' \text{ (resp. } K_m'') \ll \log(\alpha_m^{G}) \sim \log(m/\kappa_m), \]
which is always feasible given \( u_m/\ell_m \to \infty \).

As in the proof of Corollary 28.(ii), (96) is satisfied by \( \chi' \) if
\[ E_1[T'((\alpha_m^G, \beta_m^G))] \leq K'_m \frac{\log(u_m/\ell_m)}{I_0} \ll \frac{m \log(u_m/\ell_m)}{u_m}, \]
or, equivalently,
\[ K'_m \ll \sqrt{\frac{m}{u_m}}, \tag{128} \]
and (96) is satisfied by \( \chi'' \) if
\[ E_1[T''((\alpha_m^G, \beta_m^G))] \leq K''_m \frac{\log(u_m/\ell_m)}{I_0} \ll \frac{m \log(u_m/\ell_m)}{u_m}, \]
or, equivalently,
\[ K''_m \ll \frac{m}{u_m}. \tag{129} \]

A condition that guarantees the existence of \( (K_m') \) that satisfies both (127) and (128) is (98), while a condition that guarantees the existence of \( (K_m'') \) that satisfies both (127) and (129) is (97).

\textbf{REFERENCES}

[1] Y. Xing and G. Fellouris, “Asymptotically optimal multistage tests for iid data,” in Proc. IEEE Int. Symp. Inform. Theory (ISIT), Jun. 2022, pp. 2985–2990.

[2] H. F. Dodge and H. G. Romig, “A method of sampling inspection,” Bell Syst. Tech. J., vol. 8, no. 4, pp. 613–631, Oct. 1929.

[3] S. J. Pocock, “Group sequential methods in the design and analysis of clinical trials,” Biometrika, vol. 64, no. 2, pp. 191–199, Aug. 1977.

[4] K. K. Gordon Lan and D. L. Demets, “Discrete sequential boundaries for clinical trials,” Biometrika, vol. 70, no. 3, pp. 659–663, 1983.

[5] C. Jennison, “Efficient group sequential tests with unpredictable group sizes,” Biometrika, vol. 74, no. 1, pp. 155–165, 1987.

[6] T. L. Lai and M.-C. Shih, “Power, sample size and adaptation considerations in the design of group sequential clinical trials,” Biometrika, vol. 91, no. 3, pp. 507–528, Sep. 2004.

[7] J. Bartroff and T. L. Lai, “Efficient adaptive designs with mid-course sample size adjustment in clinical trials,” Statist. Med., vol. 27, no. 10, pp. 1593–1611, May 2008.

[8] G. B. Wetherill, Sampling Inspection and Quality Control, vol. 129. Cham, Switzerland: Springer, 2013.

[9] C. Jennison and B. W. Turnbull, Group Sequential Methods With Applications to Clinical Trials. Boca Raton, FL, USA: CRC Press, 1999.

[10] J. Bartroff, T. L. Lai, and M.-C. Shih, Sequential Experimentation in Clinical Trials: Design and Analysis, vol. 298. Cham, Switzerland: Springer, 2012.

[11] G. Lorden, “Asymptotic efficiency of three-stage hypothesis tests,” Ann. Statist., vol. 11, no. 1, pp. 129–140, Mar. 1983.

[12] Y. Xing and G. Fellouris, “Asymptotically optimal multistage tests for non-iid data,” Statistica Sinica, to be published, doi: 10.5705/ss.202022.0235.

[13] J. M. Satagopan and R. C. Elston, “Optimal two-stage genotyping in population-based association studies,” Genetic Epidemiol., Off. Publication Int. Genetic Epidemiol. Soc., vol. 25, no. 2, pp. 149–157, 2003.

[14] S. Zehetmayer, P. Bauer, and M. Posch, “Two-stage designs for experiments with a large number of hypotheses,” Bioinformatics, vol. 21, no. 19, pp. 3771–3777, Oct. 2005.

[15] S. Zehetmayer, P. Bauer, and M. Posch, “Optimized multi-stage designs controlling the false discovery or the family-wise error rate,” Statist. Med., vol. 27, no. 21, pp. 4145–4160, Sep. 2008.

[16] H.-H. Müller, R. Pahl, and H. Schäfer, “Including sampling and phenotyping costs into the optimization of two stage designs for genome wide association studies,” Genetic Epidemiol., vol. 31, no. 8, pp. 844–852, 2007.

[17] A. Tajer, R. M. Castro, and X. Wang, “Adaptive sensing of congested spectrum bands,” IEEE Trans. Inf. Theory, vol. 58, no. 9, pp. 6110–6125, Sep. 2012.

[18] J. Geng, W. Xu, and L. Lai, “Quickest sequential multiband spectrum sensing with mixed observations,” IEEE Trans. Signal Process., vol. 64, no. 22, pp. 5861–5874, Nov. 2016.

[19] E. Bashan, R. Raich, and A. O. Hero, “Optimal two-stage search for sparse targets using convex criteria,” IEEE Trans. Signal Process., vol. 56, no. 11, pp. 5389–5402, Nov. 2008.

[20] E. Bashan, G. Newstadt, and A. O. Hero, “Two-stage multiscale search for sparse targets,” IEEE Trans. Signal Process., vol. 59, no. 5, pp. 2331–2341, May 2011.

[21] M. L. Malloy and R. D. Nowak, “Sequential testing for sparse recovery,” IEEE Trans. Inf. Theory, vol. 60, no. 12, pp. 7862–7873, Dec. 2014.
[22] E. L. Lehmann and J. P. Romano, “Generalizations of the familywise error rate,” in Selected Works of EL Lehmann. Cham, Switzerland: Springer, 2012, pp. 719–735.

[23] A. Tartakovsky, I. Nikiforov, and M. Basseville, Sequential Analysis: Hypothesis Testing and Changepoint Detection. Boca Raton, FL, USA: CRC Press, 2014.

[24] A. Dembo and O. Zeitouni, Large Deviations Techniques and Applications, vol. 38. Cham, Switzerland: Springer, 2009.

[25] K. Kim and D. L. Demets, “Design and analysis of group sequential tests based on the type I error spending rate function,” Biometrika, vol. 74, no. 1, pp. 149–154, 1987.

[26] J. Bartroff, “Asymptotically optimal multistage tests of simple hypotheses,” Ann. Statist., vol. 35, no. 5, pp. 2075–2105, Oct. 2007.

[27] Y. Benjamini and Y. Hochberg, “Controlling the false discovery rate: A practical and powerful approach to multiple testing,” J. Roy. Stat. Soc., B, Methodol., vol. 57, no. 1, pp. 289–300, Jan. 1995.

[28] K. Cohen and Q. Zhao, “Active hypothesis testing for anomaly detection,” IEEE Trans. Inf. Theory, vol. 61, no. 3, pp. 1432–1450, Mar. 2015.

[29] B. Huang, K. Cohen, and Q. Zhao, “Active anomaly detection in heterogeneous processes,” IEEE Trans. Inf. Theory, vol. 65, no. 4, pp. 2284–2301, Apr. 2019.

[30] T. Lambez and K. Cohen, “Anomaly search with multiple plays under delay and switching costs,” IEEE Trans. Signal Process., vol. 70, pp. 174–189, 2022.

[31] Y. Song and G. Fellouris, “Asymptotically optimal, sequential, multiple testing procedures with prior information on the number of signals,” Electron. J. Statist., vol. 11, no. 1, pp. 338–363, Jan. 2017.

[32] Y. Song and G. Fellouris, “Sequential multiple testing with generalized error control: An asymptotic optimality theory,” Ann. Statist., vol. 47, no. 3, pp. 1776–1803, Jun. 2019.

[33] A. Tsopelakos and G. Fellouris, “Sequential anomaly detection under sampling constraints,” IEEE Trans. Inf. Theory, early access, May 23, 2022, doi: 10.1109/TIT.2022.3177142.