arXiv:2504.21704v3 [quant-ph] 14 Nov 2025

Bounds in Sequential Unambiguous Discrimination of Multiple Pure Quantum States

Jordi Pérez-Guijarro, Alba Pagès-Zamora, and Javier R. Fonollosa

SPCOM Group, Universitat Politècnica de Catalunya, Barcelona, Spain

Sequential methods for quantum hypothesis testing offer significant advantages over fixed-length approaches, which rely on a predefined number of state copies. Despite their potential, these methods remain underexplored for unambiguous discrimination. In this work, we derive performance bounds for such methods when applied to the discrimination of a set of pure states. The performance is evaluated based on the expected number of copies required. We establish a lower bound applicable to any sequential method and an upper bound on the optimal sequential method. The upper bound is derived using a novel and simple non-adaptive method. Importantly, the gap between these bounds is minimal, scaling logarithmically with the number of distinct states.

The problem of identifying a discrete set of states is central to quantum information science. Insights from this task have far-reaching implications, including setting theoretical bounds on quantum error mitigation (QEM) [1], as well as advancements in fields such as quantum cryptography [2] and quantum communication [3].

As this task is so fundamental, it has a rich history of study and results. The study of this task began with binary state discrimination, that is, distinguishing between two states, σ_1 and σ_2. The optimal measurement is given by the Helstrom measurement [4], which yields a probability of error given by $\frac{1}{2}\left(1 - \|\pi_1\sigma_1 - \pi_2\sigma_2\|_1\right)$, where π_i denotes the a priori probability of hypothesis i. With more copies of the states, the probability of error decreases exponentially, motivating the study and characterization of the decay rate of the probability of error. The optimal decay rate is given by the quantum Chernoff distance [5, 6], which generalizes its classical counterpart. For the more general case where $N \geq 2$ states are considered, the optimal decay rate is also well-characterized and given by $\min_{i,j:i\neq j} C_Q(\sigma_i, \sigma_j)$ [7], where $C_Q(\sigma, \rho)$ is the Chernoff distance between states σ and ρ.

A variant of this problem is the unambiguous discrimination of states. In this setting, any decision made must be correct. To guarantee this requirement, the option of choosing not to decide is allowed. That is, for N states, there are $N + 1$ possible results of the discrimination process, with one completely uninformative. However, not every set of states allows for unambiguous discrimination, as certain constraints must be met. For example, for pure states and non-adaptive methods, the states must be linearly independent. This problem has been widely studied for the cases of $N = 2$ [8, 9, 10, 11, 12] and $N = 3$ [13, 14]. In contrast, for $N > 3$, finding an analytical solution for the procedure that minimizes the probability of an inconclusive answer becomes extremely complex.

The results discussed so far are about fixed-length methods, which involve a fixed number of state copies. Yet, other approaches, known as sequential methods, proposed for the task of quantum hypothesis testing in [15], are also studied for both quantum

hypothesis testing and unambiguous discrimination. Sequential methods, unlike fixed-length methods, deal with a random number of state copies. At the cost of this uncertainty, they offer improved performance [16, 17, 18]. Interestingly, for the case of N pure states, as shown in [18], using only simple non-collective measurements, we can achieve unambiguous discrimination using an expected number of copies that scales as $O(N)$. Much like fixed-length methods, the problem of unambiguous discrimination has been scarcely studied for the general case of N states using sequential methods.

In this paper, we analyze the theoretical limits on the expected number of state copies required by sequential methods for the unambiguous discrimination of N pure states. First, we derive a lower bound on the number of copies needed by any sequential method that performs unambiguous discrimination. Next, we establish an upper bound on the performance of the optimal sequential method. The gap between these bounds is of the order $O(\log N)$, which is relatively small.

1 Quantum Hypothesis Testing and Unambiguous Discrimination

Quantum hypothesis testing involves distinguishing between different quantum states or hypotheses based on measurement outcomes. Specifically, given multiple copies of an unknown state σ, the task is to determine which state from the set $\{\sigma_i\}_{i=1}^N$ it corresponds to. Here, we analyze a special case where all the states are pure, i.e., $\sigma_i = |\psi_i\rangle\langle\psi_i|$ for every $i \in [N]$, where $[N] := \{1, 2, \ldots, N\}$. The decision $D \in [N]$ is a random variable (r.v.) that depends on the outcome of the measurements performed on the copies of state σ. The true hypothesis is denoted by S, i.e., $\sigma = \sigma_S$. Importantly, $S \in [N]$ is also a r.v., with probabilities given by the priors $\{\pi_i\}_{i=1}^N$.

Therefore, the probability of error is given by $\mathbb{P}(D \neq S)$ and for simplicity, it is denoted by $\mathbb{P}(\mathcal{E})$. Note that in fixed-length methods, the number of copies used to make a decision, denoted by L, is preassigned. However, in sequential methods, where measurements are performed in an online fashion, the decision is not made until some stopping criterion is met. Consequently, the number of copies may vary depending on the measurement results, so L becomes a random variable.

In unambiguous discrimination, the decision must be either $D = S$, meaning the true hypothesis is identified, or $D = N + 1$, which accounts for cases where the discrimination fails. The value $N+1$ is arbitrary and merely represents this additional outcome associated with the inconclusive response event, denoted by $\mathcal{I}$. For fixed-length methods, the inconclusive probability must always be strictly positive, whereas for sequential methods, it can also be exactly zero. Additionally, sequential methods have another advantage over fixed-length methods. No necessary condition is required for the existence of a sequential method that solves the unambiguous discrimination problem, whereas for fixed-length methods, it is necessary that the states $\{|\psi_i\rangle^{\otimes L}\}_{i=1}^N$ are linearly independent [19]. Throughout the paper, we use the shorthand $\mathbb{P}(\mathcal{I}|s)$ for $\mathbb{P}(\mathcal{I}|S = s)$, and $\mathbb{E}[L|s]$ for $\mathbb{E}[L|S = s]$.

Finally, for the sake of clarity, we introduce the formal definition of a sequential method for quantum unambiguous discrimination.

Definition 1. *A sequential method for quantum unambiguous discrimination is a method that can be expressed in the form of Algorithm 1, and furthermore satisfies* $\mathbb{P}(D \neq S) = 0$.

Algorithm 1

Input: Access to a query function that outputs copies of the unknown state $|\psi\rangle$.
Output: Decision $d \in [N+1]$.
$i \leftarrow 0$
while stopping criterion is not satisfied **do**
• Update the counter: $i \leftarrow i+1$
• Request k_i copies of the unknown state $|\psi\rangle$, where k_i may depend on the data from previous iterations.
• Measure state $|\psi\rangle\langle\psi|^{\otimes k_i} \otimes \tau_{i-1}$ using measurement $\mathcal{L}_i$, obtaining outcome x_i, where τ_{i-1} denotes the collection of all post-measurement states from previous iterations. This detail is included to account for measurements that do not completely collapse the state, such as weak measurements. Similar to k_i, the measurement $\mathcal{L}_i$ may depend on $(\{x_j\}_{j=1}^{i-1}, \{k_j\}_{j=1}^{i-1}, \{\mathcal{L}_j\}_{j=1}^{i-1})$.
• Check whether the stopping criterion is satisfied.
end while
Take a decision $d \in [N+1]$, which is a function of the triplet $(\{x_j\}_{j=1}^{i}, \{k_j\}_{j=1}^{i}, \{\mathcal{L}_j\}_{j=1}^{i})$.

Interestingly, aside from the distinction between fixed-length and sequential methods, additional distinctions can be made to further characterize these methods based on their properties. For instance, if the measurements used do not depend on previous outcomes, the method is said to be non-adaptive; otherwise, it is adaptive.

2 Lower bounds on the expected number of copies

In this section, we first present a lower bound on the expected number of copies used by any sequential method that unambiguously discriminates a set of pure states. Next, we apply the same technique to obtain an analogous bound for quantum hypothesis testing. Specifically, we adopt a worst-case approach, where we bound the expected value $\max_{s\in[N]} \mathbb{E}[L|s]$ rather than $\mathbb{E}[L]$. This approach allows for a more straightforward technical analysis and interpretation.

2.1 Lower Bound in unambiguous discrimination

We start by presenting the main result of this section, which establishes a bound on the expected number of copies required for unambiguous discrimination, and then proceed with its proof. Importantly, the results obtained in this section do not require any particular restriction on the sequential methods used or on the states to be discriminated.

Theorem 1. *For any sequential method that distinguishes unambiguously a set of pure states* $\{|\psi_i\rangle\}_{i=1}^{N}$*, the maximum expected value of state copies* $\max_{s\in[N]} \mathbb{E}[L|s]$ *satisfies*

$$\frac{1-\max_{s\in[N]} \mathbb{P}(\mathcal{I}|s)}{-\log \max_{i,j:i\neq j} |\langle\psi_i|\psi_j\rangle|^2} \left(\frac{\log 3}{3}\right) \leq \left\lceil \max_{s\in[N]} \mathbb{E}[L|s] \right\rceil \tag{1}$$

The most interesting aspect of this lower bound is that it is inversely proportional to the minimum pairwise Chernoff distance. This is because the Chernoff distance can be expressed as $-\log \mathrm{Tr}(\sigma_1\sigma_2)$ when at least one of the two states is pure [20]. That is, the geometry of the states has a significant impact on the performance of the methods. Another interesting point about the bound is its dependence on the inconclusive probability, where a higher probability reduces the required number of copies, as expected.

To prove this theorem, we first need to derive some auxiliary results. In particular, we use Corollary 1 and Lemma 2 to derive the bound for $\mathbb{P}(\mathcal{I}|s) = 0$. Next, using the result for $\mathbb{P}(\mathcal{I}|s) = 0$ and Lemma 3, the final result is obtained. We begin by deriving Corollary 1, which is a specific case of the following lemma.

Lemma 1. *Let $\mathbb{P}_{\text{succ}}(\{\sigma_i\}_{i=1}^N)$ be the optimal probability of success in discriminating the states $\{\sigma_i\}_{i=1}^N$ given a single copy with prior probability of $1/N$ for all the states. Then, for any subset $\mathcal{S} \subseteq [N]$,*

$$\mathbb{P}_{\text{succ}}(\{\sigma_i\}_{i=1}^N) \leq \frac{|\mathcal{S}|}{N}\mathbb{P}_{\text{succ}}(\{\sigma_i\}_{i\in\mathcal{S}}) + \frac{|\overline{\mathcal{S}}|}{N}\mathbb{P}_{\text{succ}}(\{\sigma_i\}_{i\in\overline{\mathcal{S}}}) \tag{2}$$

where $\overline{\mathcal{S}} = [N]\backslash\mathcal{S}$.

Proof. For this proof, we need to express the optimal probability of success as the probability of success given by the optimal positive operator-valued measure (POVM). That is,

$$\begin{aligned}
\mathbb{P}_{\text{succ}}(\{\sigma_i\}_{i=1}^N) &= \frac{1}{N} \sup_{\substack{\{\Lambda_i\succeq 0\}_{i=1}^N \\ \text{s.t.}\sum_{i=1}^N \Lambda_i = I}} \sum_{i=1}^N \text{Tr}(\Lambda_i\sigma_i) \\
&= \sup_{\substack{0\preceq\Theta_1,\Theta_2\preceq I \\ \Theta_1+\Theta_2=I}} \left\{ \frac{1}{N} \sup_{\substack{\{\Lambda_i\succeq 0\}_{i\in\mathcal{S}} \\ \text{s.t.}\sum_{i\in\mathcal{S}}\Lambda_i=\Theta_1}} \sum_{i\in\mathcal{S}} \text{Tr}(\Lambda_i\sigma_i) + \frac{1}{N} \sup_{\substack{\{\Lambda_i\succeq 0\}_{i\in\overline{\mathcal{S}}} \\ \text{s.t.}\sum_{i\in\overline{\mathcal{S}}}\Lambda_i=\Theta_2}} \sum_{i\in\overline{\mathcal{S}}} \text{Tr}(\Lambda_i\sigma_i) \right\} \\
&\leq \frac{1}{N} \sup_{\substack{\{\Lambda_i\succeq 0\}_{i\in\mathcal{S}} \\ \text{s.t.}\sum_{i\in\mathcal{S}}\Lambda_i\preceq I}} \sum_{i\in\mathcal{S}} \text{Tr}(\Lambda_i\sigma_i) + \frac{1}{N} \sup_{\substack{\{\Lambda_i\succeq 0\}_{i\in\overline{\mathcal{S}}} \\ \text{s.t.}\sum_{i\in\overline{\mathcal{S}}}\Lambda_i\preceq I}} \sum_{i\in\overline{\mathcal{S}}} \text{Tr}(\Lambda_i\sigma_i) \\
&= \frac{1}{N} \sup_{\substack{\{\Lambda_i\succeq 0\}_{i\in\mathcal{S}} \\ \text{s.t.}\sum_{i\in\mathcal{S}}\Lambda_i= I}} \sum_{i\in\mathcal{S}} \text{Tr}(\Lambda_i\sigma_i) + \frac{1}{N} \sup_{\substack{\{\Lambda_i\succeq 0\}_{i\in\overline{\mathcal{S}}} \\ \text{s.t.}\sum_{i\in\overline{\mathcal{S}}}\Lambda_i= I}} \sum_{i\in\overline{\mathcal{S}}} \text{Tr}(\Lambda_i\sigma_i) \\
&= \frac{|\mathcal{S}|}{N}\mathbb{P}_{\text{succ}}(\{\sigma_i\}_{i\in\mathcal{S}}) + \frac{|\overline{\mathcal{S}}|}{N}\mathbb{P}_{\text{succ}}(\{\sigma_i\}_{i\in\overline{\mathcal{S}}})
\end{aligned} \tag{3}$$

where the notation $A \succeq 0$ means that A is positive semi-definite. The first equality uses that $0 \preceq \sum_{i\in S}\Lambda_i \preceq I$, and $0 \preceq \sum_{i\in\overline{S}}\Lambda_i \preceq I$, for any POVM, and in particular, for the optimal one. The inequality follows from removing the constraint $\Theta_1 + \Theta_2 = I$, meaning we allow a broader set of matrices Λ_i. Consequently, this modification can only increase the value or leave it unchanged. Finally, the third equality follows from the fact that the supremum is always achieved when $\sum_{i\in S}\Lambda_i = I$. □

From this result, the corollary used in the proof of Theorem 1 follows.

Corollary 1. *Let $\mathbb{P}_{\text{error}}(\{\sigma_i\}_{i=1}^N) = 1 - \mathbb{P}_{\text{succ}}(\{\sigma_i\}_{i=1}^N)$ be the optimal probability of error in discriminating the states $\{\sigma_i\}_{i=1}^N$ given a single copy with prior probability of $1/N$ for all the states. Then,*

$$\mathbb{P}_{\text{error}}(\{\sigma_i\}_{i=1}^N) \geq \frac{1}{N}\left(1 - \frac{1}{2}\min_{i\neq j}\|\sigma_i - \sigma_j\|_1\right) \tag{4}$$

implying that for pure states,

$$\mathbb{P}_{\text{error}}(\{\sigma_i\}_{i=1}^N) \geq \frac{1}{N}\left(1-\sqrt{1-\max_{i\neq j}|\langle\psi_i|\psi_j\rangle|^2}\right) \tag{5}$$

Proof. The proof follows trivially by applying Lemma 1 and performing some algebraic manipulations.

$$\begin{aligned}
\mathbb{P}_{\text{error}}(\{\sigma_i\}_{i=1}^N) &= 1-\mathbb{P}_{\text{succ}}(\{\sigma_i\}_{i=1}^N) \\
&\geq 1-\frac{2}{N}\mathbb{P}_{\text{succ}}(\{\sigma_i\}_{i\in\{p,q\}})-\frac{N-2}{N} \\
&= \frac{2}{N}\left(1-\mathbb{P}_{\text{succ}}(\{\sigma_i\}_{i\in\{p,q\}})\right) \\
&= \frac{2}{N}\left(1-\frac{1}{2}-\frac{1}{4}\|\sigma_p-\sigma_q\|_1\right) \\
&= \frac{1}{N}\left(1-\frac{1}{2}\|\sigma_p-\sigma_q\|_1\right)
\end{aligned} \tag{6}$$

where the inequality follows from Lemma 1, with $\mathcal{S}=\{p,q\}$, and from the fact that $\mathbb{P}_{\text{succ}}(\{\sigma_i\}_{i\in\overline{\mathcal{S}}})$ is upper bounded by 1. The final result is obtained by taking $\{p,q\}$ such that $\|\sigma_p-\sigma_q\|_1=\min_{i\neq j}\|\sigma_i-\sigma_j\|_1$. □

Next, we state the second auxiliary result in Lemma 2.

Lemma 2. *For any sequential method that unambiguously distinguishes a set of pure states* $\{|\psi_i\rangle\}_{i=1}^N$ *with* $\mathbb{P}(\mathcal{I})=0$ *using a maximum expected number of measurements* $\max_{s\in[N]}\mathbb{E}[L|s]$*, there exists at least one fixed-length method that uses* $AK\left\lceil\max_{s\in[N]}\mathbb{E}[L|s]\right\rceil$ *samples and distinguishes the states with a probability of error at most* $1/K^A$*, where* $A,K\in\mathbb{N}$*.*

Proof. See Appendix A. □

Now that Corollary 1 and Lemma 2 have been presented, we can prove Theorem 1 for $\mathbb{P}(\mathcal{I}|s)=0$. To do so, we use the fact that the specific fixed-length method used in Lemma 2 satisfies that its error probability is lower bounded by that of the optimal fixed-length method. Additionally, using Corollary 1, we obtain the following result.

$$\begin{aligned}
\frac{1}{K^A} &\geq \mathbb{P}_{\text{error}}\left(\left\{|\psi_i\rangle\langle\psi_i|^{\otimes AK\left\lceil\max_{s\in[N]}\mathbb{E}[L|s]\right\rceil}\right\}_{i=1}^N\right) \\
&\geq \frac{1}{N}\left(1-\sqrt{1-\max_{i\neq j}|\langle\psi_i|\psi_j\rangle|^{2AK\left\lceil\max_{s\in[N]}\mathbb{E}[L|s]\right\rceil}}\right)
\end{aligned} \tag{7}$$

for any $A\geq 1$. The power $2AK\left\lceil\max_{s\in[N]}\mathbb{E}[L|s]\right\rceil$ appears due to the property $\langle\psi_i|^{\otimes u}|\psi_j\rangle^{\otimes u}=\langle\psi_i|\psi_j\rangle^u$, as we apply Corollary 1 to states $\left\{|\psi_i\rangle^{\otimes AK\left\lceil\max_{s\in[N]}\mathbb{E}[L|s]\right\rceil}\right\}_{i=1}^N$. Therefore,

$$1-\frac{N}{K^A}\leq\sqrt{1-\max_{i\neq j}|\langle\psi_i|\psi_j\rangle|^{2AK\left\lceil\max_{s\in[N]}\mathbb{E}[L|s]\right\rceil}} \tag{8}$$

which implies that for any A such that $(1-N/K^A)\geq 0$, i.e., $A\geq\log_K N$,

$$1-\left(1-\frac{N}{K^A}\right)^2\geq\max_{i\neq j}|\langle\psi_i|\psi_j\rangle|^{2AK\left\lceil\max_{s\in[N]}\mathbb{E}[L|s]\right\rceil} \tag{9}$$

Next, using that $\log \max_{i\neq j} |\langle\psi_i|\psi_j\rangle|^2 < 0$,

$$\begin{aligned}\left\lceil \max_{s\in[N]} \mathbb{E}[L|s] \right\rceil &\geq \frac{\log\left(1-\left(1-\frac{N}{K^A}\right)^2\right)}{AK \log \max_{i\neq j} |\langle\psi_i|\psi_j\rangle|^2} \\ &= \frac{1}{-\log \max_{i\neq j} |\langle\psi_i|\psi_j\rangle|^2} \left(\frac{\log K}{K}\right)\left(\frac{\delta}{(1+\delta)} - \frac{\log\left(2-\frac{1}{N^\delta}\right)}{(1+\delta)\log N}\right) \\ &\geq \frac{1}{-\log \max_{i\neq j} |\langle\psi_i|\psi_j\rangle|^2} \left(\frac{\log K}{K}\right)\left(\frac{\delta-1}{1+\delta}\right) \end{aligned} \tag{10}$$

where the equality follows by substituting $A = \log_K\left(N^{1+\delta}\right)$, being $\delta \in \mathbb{R}^+$. The second inequality uses that $N \geq 2$ and, therefore, $\log\left(2-\frac{1}{N^\delta}\right)/\log N \leq 1$. Finally, by taking the supremum over $\delta \in \mathbb{R}^+$ and $K \in \mathbb{N}$,

$$\frac{1}{-\log \max_{i\neq j} |\langle\psi_i|\psi_j\rangle|^2} \left(\frac{\log 3}{3}\right) \leq \left\lceil \max_{s\in[N]} \mathbb{E}[L|s] \right\rceil \tag{11}$$

for any sequential method that discriminates unambiguously a set of quantum states $\{|\psi_i\rangle\}_{i=1}^N$ with inconclusive probability $\mathbb{P}(\mathcal{I})=0$. Finally, we use the following lemma to conclude the proof.

Lemma 3. *For any sequential method that unambiguously discriminates a set of states $\{|\psi_i\rangle\}_{i=1}^N$ with an inconclusive probability $\mathbb{P}(\mathcal{I}|s) < 1$ and an expected number of copies $\mathbb{E}[L|s]$, there exists another sequential method that unambiguously discriminates the same set of states $\{|\psi_i\rangle\}_{i=1}^N$ with an inconclusive probability $\mathbb{P}(\mathcal{I}|s) = 0$ and an expected number of copies equal to $\mathbb{E}[L|s]/(1-\mathbb{P}(\mathcal{I}|s))$.*

Proof. See Appendix B. □

That is, we can transform a sequential method that unambiguously discriminates a set of states with a strictly positive inconclusive probability into one with zero inconclusive probability by increasing the expected number of samples by a factor of $(1-\mathbb{P}(\mathcal{I}|s))^{-1}$. The theorem follows by applying inequality (11) to this second sequential method.

Finally, we conclude this section by discussing the difficulties involved in improving the lower bound using this proof technique. In particular, small improvements to the one-shot lower bound given in Corollary 1, such as

$$\mathbb{P}_{\text{error}}(\{\sigma_i\}_{i=1}^N) \geq \frac{1}{g(N)}\left(1-\frac{1}{2}\min_{i\neq j}\|\sigma_i-\sigma_j\|_1\right), \tag{12}$$

where $g(N) < N$ is some increasing function, would unfortunately yield the same lower bound when taking $A = \log_K g(N)^{1+\delta}$ and following the same procedure. Therefore, improving Theorem 1 would likely require exploring alternative proof techniques or refining Lemma 2.

2.2 Lower bound in quantum hypothesis testing

Interestingly, the methodology used in the previous section can be applied to the case where a sequential method is used for the standard hypothesis testing task, i.e., there is no inconclusive answer, and the probability of error may be nonzero.

Theorem 2. *For any sequential method that discriminates the set of pure states* $\{|\psi_i\rangle\}_{i=1}^N$ *with a probability of error satisfying* $\max_{s\in[N]} \mathbb{P}(\mathcal{E}|s) \leq \frac{3}{7}$*, then*

$$\frac{1}{-7\log \max_{i,j:i\neq j} |\langle\psi_i|\psi_j\rangle|^2} \left(\frac{1}{2} - \frac{7\max_{s\in[N]} \mathbb{P}(\mathcal{E}|s)}{6}\right)^2 \leq \left\lceil \max_{s\in[N]} \mathbb{E}[L|s] \right\rceil \tag{13}$$

As in Theorem 1, we observe that the lower bound depends inversely on the minimum pairwise Chernoff distance of the different states. Additionally, similar to the dependency of the bound in Theorem 1 on the inconclusive probability, here, as the probability of error increases, the bound decreases. This theorem complements the results in [16], which show that any sequential method solving a binary hypothesis testing problem for mixed states $\{\sigma_1, \sigma_2\}$ satisfies

$$\max_{s\in[2]} \mathbb{E}[L|s] \geq \frac{-\log \mathbb{P}(\mathcal{E})(1-\mathbb{P}(\mathcal{E}))}{\min\{D(\sigma_1,\sigma_2), D(\sigma_2,\sigma_1)\}} + O(1) \tag{14}$$

However, this expression does not hold when $\min\{D(\rho_1,\rho_2), D(\rho_2,\rho_1)\}$ is not finite, as is the case for pure states.

Note that the bounds given in (1) and (13) do not coincide when $\mathbb{P}(\mathcal{E}) = \mathbb{P}(\mathcal{I}) = 0$, as they are derived through slightly different procedures. Specifically, the proof of this theorem is analogous to that of Theorem 1, but it uses Lemma 4 instead of Lemma 2, which is stated below.

Lemma 4. *For any sequential method that distinguishes a set of pure states* $\{|\psi_i\rangle\}_{i=1}^N$ *with a probability of error* $\mathbb{P}(\mathcal{E}|s) < \frac{K-1}{2K}$ *for all* $s \in [N]$*, using an expected number of measurements* $\mathbb{E}[L|s]$*, there exists at least one fixed-length method that uses* $AK\left\lceil \max_{s\in[N]} \mathbb{E}[L|s] \right\rceil$ *samples and distinguishes the states with a probability of error at most*

$$2e^{-A\left(\frac{1}{2} - \frac{K\max_{s\in[N]} \mathbb{P}(\mathcal{E}|s)}{K-1}\right)^2}$$

where $A, K \in \mathbb{N}$ *and* $K \geq 7$*.*

Proof. See Appendix C. □

As before, we use the fact that the specific fixed-length method employed in Lemma 4 has an error probability that is lower bounded by that of the optimal fixed-length method. This, together with Corollary 1, yields

$$\begin{aligned} 2e^{-A\left(\frac{1}{2} - \frac{K\max_{s\in[N]} \mathbb{P}(\mathcal{E}|s)}{K-1}\right)^2} &\geq \mathbb{P}_{\text{error}}\left(\left\{|\psi_i\rangle\langle\psi_i|^{\otimes AK\left\lceil \max_{s\in[N]} \mathbb{E}[L|s]\right\rceil}\right\}_{i=1}^N\right) \\ &\geq \frac{1}{N}\left(1 - \sqrt{1 - \max_{i\neq j} |\langle\psi_i|\psi_j\rangle|^{2AK\left\lceil \max_{s\in[N]} \mathbb{E}[L|s]\right\rceil}}\right) \end{aligned} \tag{15}$$

for $K, A \in \mathbb{N}$, and $K \geq 7$. To obtain the result shown in Theorem 2, we need to substitute

$$A = \frac{\log 2N^{1+\delta}}{\left(\frac{1}{2} - \frac{K\max_{s\in[N]} \mathbb{P}(\mathcal{E}|s)}{K-1}\right)^2} \tag{16}$$

into the previous expression. Consequently,

$$\frac{1}{K\log \max_{i\neq j} |\langle\psi_i|\psi_j\rangle|^2} \frac{\log\left(1 - \left(1 - \frac{1}{N^\delta}\right)^2\right)}{\log(2N^{1+\delta})} \left(\frac{1}{2} - \frac{K\max_{s\in[N]} \mathbb{P}(\mathcal{E}|s)}{K-1}\right)^2 \leq \left\lceil \max_{s\in[N]} \mathbb{E}[L|s] \right\rceil \tag{17}$$

Since the inequality holds for any integer $K \geq 7$ and $\delta \in \mathbb{R}^+$, we choose $K = 7$ and let $\delta \to \infty$ as they maximize the lower bound. Next, for a fixed N,

$$\lim_{\delta \to \infty} \frac{\log\left(1 - \left(1 - \frac{1}{N^\delta}\right)^2\right)}{\log(2N^{1+\delta})} = -1, \tag{18}$$

which leads to the desired result.

3 Upper bound in unambiguous discrimination

In this section, we show an upper bound on the expected number of copies required to unambiguously discriminate a set of pure states of the optimal sequential method. The result is presented in the following theorem.

Theorem 3. *For any $p \in [0,1]$, the optimal sequential method that unambiguously discriminates the set of states $\{|\psi_i\rangle\}_{i=1}^N$, such that $\max_{s\in[N]} \mathbb{P}(\mathcal{I}|s) \leq p$, satisfies the following inequality:*

$$\max_{s\in[N]} \mathbb{E}[L|s] \leq \min_{\delta\in(0,1)} \left\{ \frac{(1-p)\log\left(\frac{N-1}{\delta}\right)}{-\log \max_{i\neq j} |\langle\psi_i|\psi_j\rangle|} \left(\frac{1+\delta}{1-\delta}\right) \right\} \tag{19}$$

For this theorem, we rely on the simple fact that any sequential method, in expectation, uses more copies than the optimal method. Specifically, this upper bound is established using a novel non-adaptive approach, which we now present. As before, we first derive the inequality for the case where the inconclusive probability is zero, and afterwards, through Lemma 6, we generalize it.

The proposed method measures copies of the state $|\psi\rangle$ until a conclusive result is obtained. It proceeds as follows. Each copy is first measured using a two-outcome measurement, and in case the first outcome is obtained, it is then measured with an $N+1$-outcome POVM. The first measurement uses POVM $\{\Lambda^\dagger\Lambda, I - \Lambda^\dagger\Lambda\}$, where matrix Λ is selected such that

(*i*) $0 \preceq \Lambda^\dagger\Lambda \preceq I$,

(*ii*) and $\{\Lambda|\psi_i\rangle\}_{i=1}^N$ is an orthogonal set of vectors.

Without loss of generality, we adopt the convention that if the first outcome is obtained when measuring the state $|\psi\rangle$, the state transforms into $\Lambda|\psi\rangle / \sqrt{\langle\psi|\Lambda^\dagger\Lambda|\psi\rangle}$[1]. Consequently, if the outcome of the POVM corresponds to $\Lambda^\dagger\Lambda$, then the set of post-measurement states becomes an orthonormal set. Hence, the second POVM, given by

$$\left\{ \frac{\Lambda|\psi_1\rangle\langle\psi_1|\Lambda^\dagger}{\langle\psi_1|\Lambda^\dagger\Lambda|\psi_1\rangle}, \frac{\Lambda|\psi_2\rangle\langle\psi_2|\Lambda^\dagger}{\langle\psi_2|\Lambda^\dagger\Lambda|\psi_2\rangle}, \cdots, \frac{\Lambda|\psi_N\rangle\langle\psi_N|\Lambda^\dagger}{\langle\psi_N|\Lambda^\dagger\Lambda|\psi_N\rangle}, I - \sum_{i=1}^N \frac{\Lambda|\psi_i\rangle\langle\psi_i|\Lambda^\dagger}{\langle\psi_i|\Lambda^\dagger\Lambda|\psi_i\rangle} \right\}, \tag{20}$$

identifies the post-measurement state resulting from the first measurement, thereby allowing the correct identification of the initial state $|\psi\rangle$.

Now that we have explained the method, let's propose a specific matrix Λ that satisfies conditions (*i*) and (*ii*) under certain assumptions.

[1] A POVM does not uniquely determine the post-measurement state. However, the post-measurement state of different POVM implementations varies only by a unitary operation, which is irrelevant to our analysis.

Proposition 1. *Iff* $\mathrm{rank}(W) = N$, *where* $W = \sum_{i=1}^{N} |\psi_i\rangle \langle\psi_i|$, *i.e., when the states* $\{|\psi_i\rangle\}_{i=1}^{N}$ *are linearly independent, then matrix*

$$\Lambda = \frac{\sqrt{W^+}}{\left\|\sqrt{W^+}\right\|_2} \tag{21}$$

where $\|M\|_2 := \sup_{u:\|u\|_2=1} \|Mu\|_2 = \sqrt{\lambda_{\max}(M^\dagger M)}$ *and* M^+ *denotes the pseudoinverse of* M, *satisfies that:*

(*i*) $0 \preceq \Lambda^\dagger \Lambda \preceq I$,

(*ii*) *and* $\{\Lambda |\psi_i\rangle\}_{i=1}^{N}$ *is an orthogonal set of vectors.*

Proof. Substituting $\Lambda = \frac{\sqrt{W^+}}{\left\|\sqrt{W^+}\right\|_2}$, we obtain

$$0 \preceq \Lambda^\dagger \Lambda = \frac{\sqrt{W^+}^\dagger \sqrt{W^+}}{\lambda_{\max}(\sqrt{W^+}^\dagger \sqrt{W^+})} = \frac{W^+}{\lambda_{\max}(W^+)} \preceq I \tag{22}$$

where the equality follows from the fact that W is Hermitian. As for the second condition, we prove that $\langle\psi_i| W^+ |\psi_j\rangle = \delta_{i,j}$ when $\mathrm{rank}(W) = N$. To do this, we first define the matrix

$$H = [|\psi_1\rangle, |\psi_2\rangle, \ldots, |\psi_N\rangle]. \tag{23}$$

Matrix W can be written as $HH^\dagger$, and additionally, matrix $M_{i,j} = \langle\psi_i| W^+ |\psi_j\rangle$ is given by $H^\dagger W^+ H$. Since $N = \mathrm{rank}(W) = \mathrm{rank}(H)$, then $H^+ H = I_N$, where I_N is the $N \times N$ identity matrix. Therefore,

$$\begin{aligned} M &= H^\dagger W^+ H = H^\dagger (HH^\dagger)^+ H \\ &= H^\dagger H^{\dagger +} H^+ H = (H^+ H)^\dagger H^+ H = I_N \end{aligned} \tag{24}$$

where the third equality uses that $(AA^\dagger)^+ = A^{\dagger +} A^+$ for any matrix A.

Finally, we prove that the condition $\mathrm{rank}(W) = N$ is a necessary condition. To see this, we first use the fact that any Λ that satisfies the conditions (*i*) and (*ii*) also fulfills

$$\Lambda |\psi_i\rangle \langle\psi_i| \Lambda^\dagger = \alpha_i |v_i\rangle \langle v_i| \tag{25}$$

where $\{|v_i\rangle\}_{i=1}^{N}$ is a set of N orthonormal quantum states. Hence,

$$\Lambda \left(\sum_{i=1}^{N} |\psi_i\rangle \langle\psi_i| \right) \Lambda^\dagger = \sum_{i=1}^{N} \alpha_i |v_i\rangle \langle v_i| \tag{26}$$

which implies that $\mathrm{rank}(\Lambda W \Lambda^\dagger) = N$. Furthermore, using the property

$$\mathrm{rank}(AB) \leq \min\{\mathrm{rank}(A), \mathrm{rank}(B)\}. \tag{27}$$

it follows that $N = \mathrm{rank}(\Lambda W \Lambda^\dagger) \leq \mathrm{rank}(W)$. Since W is the sum of N rank-1 matrices, we have that $\mathrm{rank}(W) \leq N$. Therefore, $\mathrm{rank}(W) = N$. Interestingly, the condition $\mathrm{rank}(W) = N$ is not only necessary for $\Lambda = \frac{\sqrt{W^+}}{\left\|\sqrt{W^+}\right\|_2}$ to satisfy conditions (*i*) and (*ii*), but also necessary for the existence of any matrix Λ that satisfies both conditions.

□

3.1 Performance of the proposed method

In this section, we analyze the performance of the method when using the Λ matrix presented in Proposition 1, which assumes linear independence of the states $\{|\psi_i\rangle\}_{i=1}^N$. Specifically, as in the rest of the paper, the performance is measured based on the expected number of copies used.

Lemma 5. *The maximum expected number of copies used,* $\max_{s\in[N]} \mathbb{E}[L|s]$*, when applying the method introduced in Section 3, is upper bounded as*

$$\max_{s\in[N]} \mathbb{E}[L|s] \leq \frac{\lambda_{\max}(W)}{\lambda^+_{\min}(W)} \tag{28}$$

where $W = \sum_{i=1}^N |\psi_i\rangle \langle\psi_i|$*, and* $\lambda^+_{\min}(W)$ *denotes the smallest positive eigenvalue of* W*.*

Proof. Denoting by $|\psi_s\rangle$ the observed quantum state, then the probability of observing the first outcome with POVM $\{\Lambda^\dagger\Lambda, I - \Lambda^\dagger\Lambda\}$ is $\langle\psi_s| \Lambda^\dagger\Lambda |\psi_s\rangle$. Therefore, the expected number of copies is given by

$$\mathbb{E}[L|s] = \sum_{l=0}^{\infty} l(1 - \langle\psi_s| \Lambda^\dagger\Lambda |\psi_s\rangle)^{l-1} \langle\psi_s| \Lambda^\dagger\Lambda |\psi_s\rangle = \frac{1}{\langle\psi_s| \Lambda^\dagger\Lambda |\psi_s\rangle} \tag{29}$$

where the second equality uses that $\sum_{i=0}^{\infty} i\,\alpha^i = \alpha/(1-\alpha)^2$ for $|\alpha| < 1$. Next, note that $|\psi_s\rangle \perp \mathrm{Ker}(W)$, which leads to the following inequality,

$$\begin{aligned}
\langle\psi_s| \Lambda^\dagger\Lambda |\psi_s\rangle &\geq \lambda^+_{\min}(\Lambda^\dagger\Lambda) \\
&= \lambda^+_{\min}\left(\frac{W^+}{\lambda_{\max}(W^+)}\right) \\
&= \frac{\lambda^+_{\min}(W^+)}{\lambda_{\max}(W^+)} = \frac{\lambda^+_{\min}(W)}{\lambda_{\max}(W)}
\end{aligned} \tag{30}$$

where $\lambda^+_{\min}(\Lambda^\dagger\Lambda)$ denotes the smallest positive eigenvalue of $\Lambda^\dagger\Lambda$. The first equality follows from $\Lambda = \sqrt{W^+}/\left\|\sqrt{W^+}\right\|_2$ and the last equality uses that $\lambda_{\max}(W^+) = 1/\lambda^+_{\min}(W)$ and $\lambda^+_{\min}(W) = 1/\lambda_{\max}(W^+)$. Therefore,

$$\mathbb{E}[L|s] = \frac{1}{\langle\psi_s| \Lambda^\dagger\Lambda |\psi_s\rangle} \leq \frac{\lambda_{\max}(W)}{\lambda^+_{\min}(W)} \tag{31}$$

□

Interestingly, as indicated by Lemma 5, when the states are nearly orthogonal, the required number of copies is quite small. This is because, in such cases, all non-zero eigenvalues of W are close to 1. To illustrate this more clearly, let us analyze an example, where the set of states is chosen such that they satisfy $\| |\psi_i\rangle \langle\psi_i| - |i\rangle \langle i| \|_2 = \epsilon$. Specifically, we set $\epsilon = \frac{1}{2N^2}$. Notably, for any value of ϵ, the method introduced in [18] requires $O(N)$ copies. In contrast, the last method requires fewer than 3 samples on average. To prove this result, we begin by using

$$1 - N^2\epsilon \leq \lambda^+_{\min}(W) \leq \lambda_{\max}(W) \leq 1 + N\epsilon \tag{32}$$

The bound on $\lambda_{\max}(W)$ follows from

$$\begin{aligned}
\lambda_{\max}(W) = \|W\|_2 &\leq \left\|W - \sum_{i=1}^N |i\rangle \langle i|\right\|_2 + \left\|\sum_{i=1}^N |i\rangle \langle i|\right\|_2 \\
&\leq N\epsilon + 1
\end{aligned} \tag{33}$$

For the other bound, we use

$$N = \mathrm{Tr}(W) \leq \lambda^{+}_{\min}(W) + (N-1)\lambda_{\max}(W) \leq \lambda^{+}_{\min}(W) + (N-1)(1+N\epsilon) \tag{34}$$

Therefore, performing some algebraic operations,

$$1 - N^2\epsilon \leq \lambda^{+}_{\min}(W) \tag{35}$$

Consequently,

$$\mathbb{E}[L] \leq \frac{1+N\epsilon}{1-N^2\epsilon} \leq 3 \tag{36}$$

where the second inequality follows from substituting ϵ.

3.2 Removing the requirement for linear independence

The method presented at the beginning of Section 3 is based on Proposition 1 and, therefore, requires the set of states to be linearly independent. However, a sequential method that unambiguously discriminates a set of states does not need to satisfy this condition. In general, the only necessary requirement is that $|\langle\psi_i|\psi_j\rangle| < 1$ for all $i \neq j$. To achieve this, we can modify the proposed method to collectively measure a set of k copies, instead of measuring each copy individually. That is, the problem becomes one of distinguishing the states $\{|\psi_i\rangle^{\otimes k}\}_{i=1}^{N}$. In this case, we can apply the previous method as long as the states $|\psi_i\rangle^{\otimes k}$ are linearly independent. We will now show that a small value of k is generally sufficient to ensure the linear independence of the states. Furthermore, we show also that the ratio between the largest and smallest positive eigenvalues of $W^{(k)} = \sum_i |\psi_i\rangle\langle\psi_i|^{\otimes k}$ is small.

To do this, first, let us define matrix $H^{(k)} = [|\psi_1\rangle^{\otimes k}, |\psi_2\rangle^{\otimes k}, \cdots, |\psi_N\rangle^{\otimes k}]$. Importantly, $H^{(k)\dagger}H^{(k)} \to I_N$ as k increases since $[H^{(k)\dagger}H^{(k)}]_{i,j} = \langle\psi_i|\psi_j\rangle^k$. Therefore, using Gershgorin's circle theorem, the eigenvalues of $H^{(k)\dagger}H^{(k)}$ satisfy

$$1 - (N-1)\max_{i\neq j} |\langle\psi_i|\psi_j\rangle|^k \leq \lambda \leq 1 + (N-1)\max_{i\neq j} |\langle\psi_i|\psi_j\rangle|^k \tag{37}$$

Therefore, for any $\delta \in (0,1)$, if k satisfies

$$k \geq \frac{\log\left(\frac{N-1}{\delta}\right)}{-\log \max_{i\neq j} |\langle\psi_i|\psi_j\rangle|} \tag{38}$$

then the eigenvalues belong to the interval $[1-\delta, 1+\delta]$. Hence, since all eigenvalues are positive, $\mathrm{rank}(H^{(k)\dagger}H^{(k)}) = N$, which is equal to the $\mathrm{rank}(H^{(k)})$. That is, the states are linearly independent. Furthermore, since the singular values satisfy $\sigma(H^{(k)}) = \sigma(H^{(k)\dagger})$, implying that for the non-zero eigenvalues $\lambda(H^{(k)\dagger}H^{(k)}) = \lambda(W^{(k)})$. That is, the non-zero eigenvalues of $W^{(k)}$ fulfill inequality (37). Hence, putting everything together, and using the value of k specified earlier, we conclude that

$$\mathbb{E}[L|s] \leq \min_{\delta\in(0,1)} \left\{ \frac{\log\left(\frac{N-1}{\delta}\right)}{-\log \max_{i\neq j} |\langle\psi_i|\psi_j\rangle|} \left(\frac{1+\delta}{1-\delta}\right) \right\} \tag{39}$$

for any set of N quantum states. Therefore, the modification explained in this section improves efficiency in terms of the expected number of copies, but at the cost of introducing the added complexity of performing collective measurements, which are extremely difficult to implement. Additionally, note that equation (39) proves Theorem 3 for $p = 0$. The result can be generalized for $p \in [0,1]$ using the following lemma:

Lemma 6. *For any $p \in [0,1]$, and any sequential method that unambiguously discriminates a set of states $\{|\psi_i\rangle\}_{i=1}^N$ with an inconclusive probability $\mathbb{P}(\mathcal{I}) = 0$ and an expected number of copies $\mathbb{E}[L|s]$, there exists another sequential method that unambiguously discriminates the same set of states $\{|\psi_i\rangle\}_{i=1}^N$ with an inconclusive probability $\mathbb{P}(\mathcal{I}|s) = p$ and an expected number of copies equal to $\mathbb{E}[L|s](1 - \mathbb{P}(\mathcal{I}|s))$.*

Proof. The idea of the proof is to modify the sequential method slightly to introduce the inconclusive probability. In particular, before anything else, we sample a Bernoulli random variable B with parameter p. If the outcome is 1, then the method terminates and outputs $D = N + 1$. Otherwise, the original sequential method is run. From this, it follows that $\mathbb{P}(\mathcal{I}|s) = p$ for all $s \in [N]$. Let's now move on to analyze the expected value of copies. Clearly, $E[L|s, B = 1] = 0$, and similarly, $E[L|s, B = 0] = \mathbb{E}[L|s]$. Consequently, the expected value of the new sequential method is given by $E[L|s](1 - \mathbb{P}(\mathcal{I}|s))$. □

Interestingly, by examining the results of Theorem 1 and Theorem 3, we observe that the gap between the two bounds is $O(\log N)$. As a result, the optimal method can only improve the performance of the proposed method by a factor of this order. In other words, while our proposed method may not be optimal, it achieves performance that is very close to optimal.

4 Analysis for Mixed States

In this section, we discuss generalizations of some of the main results from the previous sections to the case of mixed states. Interestingly, in Theorem 1, we observe that neither Lemma 2 nor Lemma 3 makes use, at any point, of the restriction to pure states, therefore, they also apply to mixed states. Next, combining (4) with inequality

$$\|\rho - \sigma\|_1 \leq 2\sqrt{1 - F(\rho, \sigma)} \tag{40}$$

yields

$$\mathbb{P}_{\text{error}}(\{\sigma_i\}_{i=1}^N) \geq \frac{1}{N}\left(1 - \sqrt{1 - \max_{i \neq j} F(\sigma_i, \sigma_j)}\right) \tag{41}$$

Consequently, using that $F(\rho^{\otimes K}, \sigma^{\otimes K}) = F(\rho, \sigma)^K$, we obtain an expression analogous to (7). Hence, an analogous lower bound for mixed states holds.

Proposition 2. *For any sequential method that distinguishes unambiguously a set of mixed states $\{\sigma_i\}_{i=1}^N$, the maximum expected value of state copies $\max_{s\in[N]} \mathbb{E}[L|s]$ satisfies*

$$\frac{1 - \max_{s\in[N]} \mathbb{P}(\mathcal{I}|s)}{-\log \max_{i,j:i\neq j} F(\sigma_i, \sigma_j)} \left(\frac{\log 3}{3}\right) \leq \left\lceil \max_{s\in[N]} \mathbb{E}[L|s] \right\rceil \tag{42}$$

Note that, importantly, unlike in the pure case, for some sets of mixed states $\{\sigma_i\}_{i=1}^N$, all sequential methods that distinguish the states unambiguously satisfy $\max_{s\in[N]} \mathbb{P}(\mathcal{I}|s) = 1$. In other words, no sequential method can discriminate these states unambiguously. The necessary and sufficient condition for the existence of a sequential method that satisfies $\max_{s\in[N]} \mathbb{P}(\mathcal{I}|s) < 1$ is stated next.

Proposition 3. *The necessary and sufficient condition for the existence of a sequential method that unambiguously discriminates a set of mixed states $\{\sigma_i\}_{i=1}^N$, such that $\max_{s\in[N]} \mathbb{P}(\mathcal{I} \mid s) < 1$, is $\text{supp}(\sigma_i) \not\subseteq \text{supp}(\sigma_j)$ for all $i \neq j$.*

Proof. The sufficiency is shown in Section 4C of [18]. To show that the previous condition is necessary, we use a proof by contradiction. Specifically, assume that there exists a set of states $\{\sigma_i\}_{i=1}^N$ such that the condition $\text{supp}(\sigma_i) \nsubseteq \text{supp}(\sigma_j)$ for all $i \neq j$ is not satisfied, and that there exists a sequential method A such that $\max_{s\in[N]} \mathbb{P}(\mathcal{I}|s) < 1$. Then, there exist $i, j \in [N]$ such that $\text{supp}(\sigma_i) \subseteq \text{supp}(\sigma_j)$. Let us denote these states as ρ_1 and ρ_2. From [16], we know that any sequential method that discriminates these two states must satisfy

$$\max_{s\in[2]} \mathbb{E}[L|s] \geq \frac{-\log \mathbb{P}(\mathcal{E})(1-\mathbb{P}(\mathcal{E}))}{\min\{D(\rho_1,\rho_2), D(\rho_2,\rho_1)\}} + O(1) \tag{43}$$

Importantly, since $\text{supp}(\rho_1) \subseteq \text{supp}(\rho_2)$, $\min\{D(\rho_1,\rho_2), D(\rho_2,\rho_1)\}$ is finite, and the expected value grows at least as $O(-\log \mathbb{P}(\mathcal{E}))$. However, using Lemma 3, the sequential method A for unambiguous state discrimination can be transformed into a sequential method with $\mathbb{P}(\mathcal{I}) = 0$, or equivalently, $\mathbb{P}(\mathcal{E}) = 0$. Therefore, this transformed method, when applied to the states $\{\rho_1, \rho_2\}$, also satisfies $\mathbb{P}(\mathcal{E}) = 0$ and has a bounded expected value. This leads to a contradiction with (43) and completes the proof. □

Finally, the method described in Section 3 can also be generalized to apply to some sets of mixed states. The procedure starts by measuring with a POVM $\{\Lambda^\dagger\Lambda, I - \Lambda^\dagger\Lambda\}$, where observing the first outcome maps the supports of the states to orthogonal subspaces. Subsequently, a POVM of projectors onto these subspaces is performed, with an additional element projecting onto the complement to guarantee that the sum of all POVM elements equals the identity.

Specifically, a matrix Λ satisfying the previous conditions exists at least when the states $\{\left|v_j^i\right\rangle\}_{i,j}$, formed by the eigenvectors of the different states $\{\sigma_i\}$, are linearly independent, where $\left|v_j^i\right\rangle$ denotes the j^{th} eigenvector of state σ_i. In this scenario, we can take

$$\Lambda = \frac{\sqrt{W^+}}{\left\|\sqrt{W^+}\right\|_2}, \tag{44}$$

where $W = \sum_{i=1}^N \sum_{j=1}^{d_i} \left|v_j^i\right\rangle \left\langle v_j^i\right|$.

In summary, the lower bound presented in Theorem 1 can be fully generalized to mixed states, as shown in Proposition 2. Similarly, the method proposed in Section 3 can also be generalized to mixed states. However, the upper bound derived from this method is not straightforward to obtain for the mixed-state scenario.

5 Conclusions

In summary, by combining the results from both Theorem 1 and Theorem 3, we have demonstrated that the performance of the optimal sequential method for unambiguous discrimination of a set of pure states, as measured by $\max_{s\in[N]} \mathbb{E}[L|s]$, is primarily determined by the inverse of the minimum pairwise Chernoff distance. Furthermore, we have shown that the dependence of $\max_{s\in[N]} \mathbb{E}[L|s]$ on the number of states, N, is at most $O(\log N)$. Despite these findings, an important question remains unresolved: namely, what is the optimal scaling of the expected number of samples as a function of the number of states, N? That is, the optimal method may have no dependency on the number of states N at all, which would show that the lower bound of Theorem 1 is tight. Alternatively, the lower bound could be improved to demonstrate that the $O(\log N)$ factor is necessary, rather than merely an artifact of the proof of Theorem 3.

Apart from these results on optimal sequential schemes, we have also introduced a novel non-adaptive method for the unambiguous discrimination of N pure states and a generalization of Theorem 1 for mixed states. This method achieves a highly competitive performance, as the optimal strategy can improve its performance only by at most a factor of $O(\log N)$. Furthermore, the method is quite simple and has an intuitive geometric interpretation.

6 Acknowledgments

This work has been funded by grants PID2022-137099NB-C41, PID2019-104958RB-C41 funded by MCIN/AEI/10.13039/501100011033 and FSE+ and by grant 2021 SGR 01033 funded by AGAUR, Dept. de Recerca i Universitats de la Generalitat de Catalunya 10.13039/501100002809.

References

[1] Ryuji Takagi, Suguru Endo, Shintaro Minagawa, and Mile Gu. "Fundamental limits of quantum error mitigation". npj Quantum Information **8**, 114 (2022).

[2] Antonio Acín, Joonwoo Bae, E Bagan, M Baig, Ll Masanes, and Ramon Muñoz-Tapia. "Secrecy properties of quantum channels". Physical Review A—Atomic, Molecular, and Optical Physics **73**, 012327 (2006).

[3] Nicolas Gisin and Rob Thew. "Quantum communication". Nature photonics **1**, 165–171 (2007).

[4] C. W. Helstrom. "Quantum detection and estimation". Academic Press. New York (1976).

[5] K. M. R. Audenaert, J. Calsamiglia, R. Muñoz Tapia, E. Bagan, Ll. Masanes, A. Acin, and F. Verstraete. "Discriminating states: The quantum chernoff bound". Phys. Rev. Lett. **98**, 160501 (2007).

[6] M. Nussbaum and A. Szkoła. "The chernoff lower bound for symmetric quantum hypothesis testing". The Annals of Statistics **37**, 1040–1057 (2009).

[7] Ke Li. "Discriminating quantum states: The multiple chernoff distance". Ann. Statist. 44(4): 1661-1679 (2016).

[8] Ulrike Herzog and János A Bergou. "Optimum unambiguous discrimination of two mixed quantum states". Physical Review A—Atomic, Molecular, and Optical Physics **71**, 050301 (2005).

[9] Gregg Jaeger and Abner Shimony. "Optimal distinction between two non-orthogonal quantum states". Physics Letters A **197**, 83–87 (1995).

[10] Igor D Ivanovic. "How to differentiate between non-orthogonal states". Physics Letters A **123**, 257–259 (1987).

[11] Dennis Dieks. "Overlap and distinguishability of quantum states". Physics Letters A **126**, 303–306 (1988).

[12] Asher Peres. "How to differentiate between non-orthogonal states". Physics Letters A **128**, 19 (1988).

[13] János A Bergou, Ulrike Futschik, and Edgar Feldman. "Optimal unambiguous discrimination of pure quantum states". Physical review letters **108**, 250502 (2012).

[14] Gael Sentís, Esteban Martínez-Vargas, and Ramon Muñoz-Tapia. "Online identification of symmetric pure states". Quantum **6**, 658 (2022).

[15] Sergei Slussarenko, Morgan M Weston, Jun-Gang Li, Nicholas Campbell, Howard M Wiseman, and Geoff J Pryde. "Quantum state discrimination using the minimum average number of copies". Physical review letters **118**, 030502 (2017).

[16] Esteban Martínez Vargas, Christoph Hirche, Gael Sentís, Michalis Skotiniotis, Marta Carrizo, Ramon Muñoz-Tapia, and John Calsamiglia. "Quantum sequential hypothesis testing". Physical review letters **126**, 180502 (2021).

[17] Yonglong Li, Vincent YF Tan, and Marco Tomamichel. "Optimal adaptive strategies for sequential quantum hypothesis testing". Communications in Mathematical Physics **392**, 993–1027 (2022).

[18] Jordi Pérez-Guijarro, Alba Pagès-Zamora, and Javier Rodríguez Fonollosa. "Quantum multiple hypothesis testing based on a sequential discarding scheme". IEEE access **10**, 13813–13826 (2022).

[19] Anthony Chefles. "Unambiguous discrimination between linearly independent quantum states". Physics Letters A **239**, 339–347 (1998).

[20] J. Calsamiglia, R. Muñoz Tapia, Ll. Masanes, A. Acin, and E. Bagan. "Quantum chernoff bound as a measure of distinguishability between density matrices: Application to qubit and gaussian states". Phys. Rev. A **77**, 032311 (2008).

Appendices

A Proof of Lemma 2

In this appendix, we provide the proof of Lemma 2, which, for completeness, is restated here.

Lemma 2. *For any sequential method that unambiguously distinguishes a set of pure states* $\{|\psi_i\rangle\}_{i=1}^N$ *with* $\mathbb{P}(\mathcal{I}) = 0$ *using a maximum expected number of measurements* $\max_{s\in[N]} \mathbb{E}[L|s]$*, there exists at least one fixed-length method that uses* $AK \left\lceil \max_{s\in[N]} \mathbb{E}[L|s] \right\rceil$ *samples and distinguishes the states with a probability of error at most* $1/K^A$*, where* $A, K \in \mathbb{N}$*.*

Proof. The fixed-length method is defined as follows:

- Divide the $AK \left\lceil \max_{s\in[N]} \mathbb{E}[L|s] \right\rceil$ samples into A groups.
- Run the sequential method independently for each group.
- The final decision coincides with the one given by any sequential method that has completed the procedure. If the sequential method does not finish for any group, a random decision is made.

To analyze this method, first, let L_s denote the number of copies used by the sequential method when the true state is given by $|\psi_s\rangle$. Then,

$$\mathbb{P}(L_s \geq K\mathbb{E}[L|s]) \leq \frac{1}{K} \tag{45}$$

independently of the method used. This follows from Markov's inequality. Therefore,

$$\mathbb{P}\left(L_s > K\left\lceil \max_{s\in[N]} \mathbb{E}[L|s]\right\rceil\right) \leq \frac{1}{K} \tag{46}$$

holds for all $s \in [N]$. Hence,

$$\begin{aligned}\mathbb{P}(\mathcal{E}) = \mathbb{P}(D \neq S) &= \sum_{s=1}^{N} \pi_s \, \mathbb{P}(D \neq s | S = s) \\ &\leq \sum_{s=1}^{N} \pi_s \, \mathbb{P}\left(L_s > K\left\lceil \max_{s\in[N]} \mathbb{E}[L|s]\right\rceil\right)^A \\ &\leq \frac{1}{K^A}\end{aligned} \tag{47}$$

where the first inequality follows from the fact that an error can only occur if none of the A groups has finished. This happens with probability $\mathbb{P}\left(L_s > K\lceil \max_{s\in[N]} \mathbb{E}[L|s]\rceil\right)^A$, as each group uses $K\lceil \max_{s\in[N]} \mathbb{E}[L|s]\rceil$ states. The last inequality uses (46).

□

B Proof of Lemma 3

In this section we prove the following result.

Lemma 3. *For any sequential method that unambiguously discriminates a set of states $\{|\psi_i\rangle\}_{i=1}^N$ with an inconclusive probability $\mathbb{P}(\mathcal{I}|s) < 1$ and an expected number of copies $\mathbb{E}[L|s]$, there exists another sequential method that unambiguously discriminates the same set of states $\{|\psi_i\rangle\}_{i=1}^N$ with an inconclusive probability $\mathbb{P}(\mathcal{I}|s) = 0$ and an expected number of copies equal to $\mathbb{E}[L|s]/(1 - \mathbb{P}(\mathcal{I}|s))$.*

Proof. The sequential method with zero inconclusive probability is obtained by running the other sequential method until a conclusive answer is obtained. The number of times the method is run is denoted by the r.v. R. Importantly,

$$\mathbb{P}(R = r|s) = \mathbb{P}(\mathcal{I}|s)^{r-1}(1 - \mathbb{P}(\mathcal{I}|s)) \tag{48}$$

Furthermore, for $R = r$, the expected number of copies used is

$$(r-1)\mathbb{E}[L|D = N+1, s] + \mathbb{E}[L|D < N+1, s] \tag{49}$$

where $D = N+1$ is the inconclusive event. Therefore, the expected number of copies used by the new method is

$$\begin{aligned}&\sum_{r=1}^{\infty} \mathbb{P}(R = r|s)\left((r-1)\mathbb{E}[L|D = N+1, s] + \mathbb{E}[L|D < N+1, s]\right) \\ &= \sum_{r=1}^{\infty} \mathbb{P}(\mathcal{I}|s)^{r-1}(1 - \mathbb{P}(\mathcal{I}|s))\left((r-1)\mathbb{E}[L|D = N+1, s] + \mathbb{E}[L|D < N+1, s]\right)\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}[L|D < N+1, s] \sum_{r=1}^{\infty} \mathbb{P}(\mathcal{I}|s)^{r-1}(1-\mathbb{P}(\mathcal{I}|s)) \\
&\qquad + \mathbb{E}[L|D = N+1, s] \sum_{r=1}^{\infty} \mathbb{P}(\mathcal{I}|s)^{r-1}(1-\mathbb{P}(\mathcal{I}|s))(r-1) \\
&= \mathbb{E}[L|D < N+1, s] + \mathbb{E}[L|D = N+1, s] \frac{\mathbb{P}(\mathcal{I}|s)}{1-\mathbb{P}(\mathcal{I}|s)} \\
&= \frac{(1-\mathbb{P}(\mathcal{I}|s))\mathbb{E}[L|D < N+1, s] + \mathbb{P}(\mathcal{I}|s)\mathbb{E}[L|D = N+1, s]}{1-\mathbb{P}(\mathcal{I}|s)} \\
&= \frac{\mathbb{E}[L|s]}{1-\mathbb{P}(\mathcal{I}|s)}
\end{aligned} \tag{50}
$$

where the third equality uses that $\sum_{i=0}^{\infty} \alpha^i = 1/(1-\alpha)$ and $\sum_{i=0}^{\infty} i\,\alpha^i = \alpha/(1-\alpha)^2$ for $|\alpha| < 1$. The last equality uses the identities $1 - \mathbb{P}(\mathcal{I}|s) = P(D < N+1|s)$, and $\mathbb{P}(\mathcal{I}|s) = \mathbb{P}(D = N+1|s)$.

□

C Proof of Lemma 4

In this appendix, we provide the proof of Lemma 4, which, for completeness, is restated here.

Lemma 4. *For any sequential method that distinguishes a set of pure states $\{|\psi_i\rangle\}_{i=1}^N$ with a probability of error $\mathbb{P}(\mathcal{E}|s) < \frac{K-1}{2K}$ for all $s \in [N]$, using an expected number of measurements $\mathbb{E}[L|s]$, there exists at least one fixed-length method that uses $AK \left\lceil \max_{s\in[N]} \mathbb{E}[L|s] \right\rceil$ samples and distinguishes the states with a probability of error at most*

$$
2e^{-A\left(\frac{1}{2} - \frac{K \max_{s\in[N]} \mathbb{P}(\mathcal{E}|s)}{K-1}\right)^2}
$$

where $A, K \in \mathbb{N}$ and $K \geq 7$.

Proof. The method coincides with the one explained in Lemma 2, except in the decision taken. In particular, we decide by a majority vote between the segments that have finished. In case no segment ends the procedure, the decision is taken at random. Now, let's analyze its performance. To do so, first note that

$$
\begin{aligned}
\mathbb{P}(\mathcal{E}|s) &= \mathbb{P}\left(\mathcal{E}|L_s > K\left\lceil \max_{s\in[N]} \mathbb{E}[L|s] \right\rceil, s\right) \mathbb{P}\left(L_s > K\left\lceil \max_{s\in[N]} \mathbb{E}[L|s] \right\rceil\right) \\
&\quad + \mathbb{P}\left(\mathcal{E}|L_s \leq K\left\lceil \max_{s\in[N]} \mathbb{E}[L|s] \right\rceil, s\right) \mathbb{P}\left(L_s \leq K\left\lceil \max_{s\in[N]} \mathbb{E}[L|s] \right\rceil\right) \\
&\geq \left(1 - \frac{1}{K}\right) \mathbb{P}\left(\mathcal{E}|L_s \leq K\left\lceil \max_{s\in[N]} \mathbb{E}[L|s] \right\rceil, s\right)
\end{aligned} \tag{51}
$$

where the inequality uses (46). Therefore, $\mathbb{P}(\mathcal{E}|L_s \leq K \left\lceil \max_{s\in[N]} \mathbb{E}[L|s] \right\rceil, s) \leq \frac{K}{K-1}\mathbb{P}(\mathcal{E}|s) \leq \frac{K}{K-1} \max_{s\in[N]} \mathbb{P}(\mathcal{E}|s)$. Now, we move on to the probability of error associated with the majority vote among the methods that have finished. For the moment, let's assume that the number of segments that have finished is $\tilde{A} \leq A$, and for these segments, we define the following random variables,

$$
X_i^{(s)} = \mathbb{1}\left\{\text{Segment } i \text{ makes the wrong decision given that the correct decision is } s\right\} \tag{52}
$$

Hence, since in these segments $L_s \leq K \left\lceil \max_{s\in[N]} \mathbb{E}[L|s] \right\rceil$,

$$\mathbb{E}\left[X_i^{(s)}\right] = \mathbb{P}\left(\mathcal{E}|L_s \leq K\left\lceil \max_{s\in[N]} \mathbb{E}[L|s]\right\rceil, s\right) \tag{53}$$

Using Hoeffding's inequality,

$$\mathbb{P}\left(\sum_{i=1}^{\tilde{A}} X_i^{(s)} - \mathbb{P}\left(\mathcal{E}|L_s \leq K\left\lceil \max_{s\in[N]} \mathbb{E}[L|s]\right\rceil, s\right)\tilde{A} \geq t\right) \leq e^{-\frac{2t^2}{\tilde{A}}} \tag{54}$$

Taking $t = \left(1/2 - \mathbb{P}(\mathcal{E}|L_s \leq K\left\lceil \max_{s\in[N]} \mathbb{E}[L|s]\right\rceil, s)\right)\tilde{A}$ (as t has to be positive, this implies that to guarantee $t \geq 0$, $\max_{s\in[N]} \mathbb{P}(\mathcal{E}|s) \leq \frac{K-1}{2K}$) the previous inequality becomes,

$$\mathbb{P}\left(\sum_{i=1}^{\tilde{A}} X_i^{(s)} \geq \frac{\tilde{A}}{2}\right) \leq e^{-2\tilde{A}\left(\frac{1}{2} - \frac{K\max_{s\in[N]}\mathbb{P}(\mathcal{E}|s)}{K-1}\right)^2} \tag{55}$$

The left-hand side of the previous inequality is an upper bound on the probability of error of the majority vote. To see this, note that if $\sum_{i=1}^{\tilde{A}} X_i^{(s)} < \tilde{A}/2$, then the majority of the groups that have finished are correct, and therefore, the decision is correct. Therefore, errors can only happen when $\sum_{i=1}^{\tilde{A}} X_i^{(s)} \geq \tilde{A}/2$. Consequently,

$$\mathbb{P}(\mathcal{E}_{\text{fixed}}|s,\tilde{A}) \leq \mathbb{P}\left(\sum_{i=1}^{\tilde{A}} X_i^{(s)} \geq \frac{\tilde{A}}{2}\right) \leq e^{-2\tilde{A}\left(\frac{1}{2} - \frac{K\max_{s\in[N]}\mathbb{P}(\mathcal{E}|s)}{K-1}\right)^2} \tag{56}$$

where $\mathcal{E}_{\text{fixed}}$ denotes the error event of the fixed-length method. Now, let's analyze $\tilde{A}$. First, we denote

$$Y_i^{(s)} = \mathbb{1}\left\{\text{Segment } i \text{ does not terminate}\right\} \tag{57}$$

Hence, $\mathbb{E}[Y_i^{(s)}] = \mathbb{P}(L_s > K\left\lceil \max_{s\in[N]} \mathbb{E}[L|s]\right\rceil)$. Using Hoeffding's inequality,

$$\mathbb{P}\left(\sum_{i=1}^{A} Y_i^{(s)} - A\mathbb{P}\left(L_s > K\left\lceil \max_{s\in[N]} \mathbb{E}[L|s]\right\rceil\right) \geq t\right) \leq e^{\frac{-2t^2}{A}} \tag{58}$$

Taking $t = A\left(\frac{1}{2} - \mathbb{P}(L_s > K\left\lceil \max_{s\in[N]} \mathbb{E}[L|s]\right\rceil)\right)$ (this implies that $K \geq 2$ to guarantee that $t \geq 0$),

$$\mathbb{P}\left(\sum_{i=1}^{A} Y_i^{(s)} \geq \frac{A}{2}\right) \leq e^{-2A\left(\frac{1}{2}-\frac{1}{K}\right)^2} \tag{59}$$

Therefore,

$$\mathbb{P}\left(\tilde{A} \leq \frac{A}{2} \,|\, s\right) \leq e^{-2A\left(\frac{1}{2}-\frac{1}{K}\right)^2} \tag{60}$$

Putting everything together, we have that

$$\begin{aligned}
\mathbb{P}(\mathcal{E}_{\text{fixed}}) &= \sum_{s,\tilde{A}} \mathbb{P}(\mathcal{E}_{\text{fixed}}|s,\tilde{A})\mathbb{P}(\tilde{A},s) \\
&= \sum_{s,\tilde{A}} \mathbb{P}(\mathcal{E}_{\text{fixed}}|s,\tilde{A})\mathbb{P}(\tilde{A}|s)\,\pi_s \\
&= \sum_{s,\tilde{A}>\frac{A}{2}} \mathbb{P}(\mathcal{E}_{\text{fixed}}|s,\tilde{A})\mathbb{P}(\tilde{A}|s)\,\pi_s + \sum_{s,\tilde{A}\leq\frac{A}{2}} \mathbb{P}(\mathcal{E}_{\text{fixed}}|s,\tilde{A})\mathbb{P}(\tilde{A}|s)\pi_s \\
&\leq e^{-A\left(\frac{1}{2} - \frac{K\max_{s\in[N]}\mathbb{P}(\mathcal{E}|s)}{K-1}\right)^2} + e^{-2A\left(\frac{1}{2}-\frac{1}{K}\right)^2}
\end{aligned} \tag{61}$$

Finally, for $K \geq 7$, as

$$\left(\frac{1}{\sqrt{2}} - \frac{\sqrt{2}}{K}\right) \geq \frac{1}{2} \geq \left(\frac{1}{2} - \frac{K \max_{s\in[N]} \mathbb{P}(\mathcal{E}|s)}{K-1}\right) \geq 0 \tag{62}$$

it follows that

$$\mathbb{P}(\mathcal{E}_{\text{fixed}}) \leq 2e^{-A\left(\frac{1}{2} - \frac{K \max_{s\in[N]} \mathbb{P}(\mathcal{E}|s)}{K-1}\right)^2} \tag{63}$$

□