Rate reduction of blind quantum data compression with local approximations based on the unstable structure of quantum states

Kohdai Kuroiwa [1,2,3,4,*] and Debbie Leung [1,3,4,†]

[1]*Institute for Quantum Computing, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1*
[2]*Department of Physics and Astronomy, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1*
[3]*Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1*
[4]*Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada N2L 2Y5*

In this paper we propose a protocol for a data compression task, blind quantum data compression, with finite local approximations. The rate of blind data compression is susceptible to approximations even when the approximations are diminutive. This instability originates from the sensitivity of a structure of quantum states against approximations, which makes the analysis of blind compression in the presence of approximations intractable. In this paper we construct a protocol that takes advantage of the instability to reduce the compression rate substantially. Our protocol shows a significant reduction in rate for specific examples we examine. Moreover, we apply our methods to diagonal states and propose two types of approximation methods in this special case. We perform numerical experiments and observe that one of these two approximation methods performs significantly better than the other. Thus, our analysis makes a step toward general investigation of blind quantum data compression with the allowance of approximations towards further investigation of approximation-rate trade-off of blind quantum data compression.

I. INTRODUCTION

Data compression is a fundamental information processing task in information theory [1]. In this task, the data are drawn from a known distribution identically and independently an asymptotically large number of times, and the goal is to transmit the data as efficiently as possible. The concepts of data, distributions, and data compression have natural generalizations in the quantum setting. Quantum data compression under various setups has been extensively investigated [2–19]. (See Sec. II C for a more detailed discussion.) Among all these setups, we focus on blind quantum data compression, in which the sender does not know the description of the quantum states to be transmitted. This problem has been widely studied [2–7,9,12,18,19], and in particular, in Ref. [12], the structure of a set of quantum states is used to derive an optimal compression rate under the constraint that the error vanishes asymptotically.

More recently, Refs. [18,19] indicated the possibility that the error analysis conducted in Ref. [12] can be further tightened. Indeed, the structure of quantum states used in the analysis is highly sensitive to approximations. Reference [18] introduced a blind compression protocol for special quantum states called classical states. The protocol achieves a considerably lower rate by allowing finite approximations. Previously, such data compression tasks with finite approximations were investigated through quantum rate distortion theories. Here a function called distortion measure is employed to characterize possible approximations between the original and resulting states. A large body of results has been obtained for several setups of data compression and channel coding via rate distortion theories [20–26]. Nevertheless, the approximate compression of general mixed quantum states has been largely unexplored.

In this paper, as a first step of a more thorough error analysis of blind quantum data compression with finite approximations, we propose and study a method to reduce the compression rate that exploits the sensitivity of the structure of quantum states under local approximation. In particular, in previous work [12], the structure of quantum states defines redundant information which need not be transmitted. We investigate how much this structure can be sensitive to approximations, by explicitly exhibiting ensembles that have information that is not exactly redundant, but approximately so. In our examples, this approximately redundant information occupies a large dimension. The previous optimal protocol [12] is defined using the exact redundant information and therefore may have a very high compression rate. We instead propose a protocol that identifies and omits this large approximately redundant information, resulting in a substantial reduction of the compression rate.

We apply our protocol to classical states to show the rate-approximation relation of our protocol. We can interpret the generation of approximately redundant parts as a coarse graining of probability distributions corresponding to the given classical states. More specifically, we propose two methods to generate approximately redundant parts: the arithmetic mean method and the geometric mean method. Comparing these

*kkuroiwa@uwaterloo.ca
†wcleung@uwaterloo.ca

two methods numerically, we gain an understanding of the trade-off between the approximation and the compression rate. Furthermore, we also numerically show the relation between the rate and the system dimension when the allowed approximation depends on the dimension. These numerical simulations show that the arithmetic mean method performs essentially better than the geometric mean method.

In summary, we propose a protocol for blind quantum data compression with finite local approximations. Our examples show the evidence that this protocol can show a substantial reduction in the compression rate. Furthermore, the numerical analysis of classical ensembles implies how the compression rate depends on approximations. These results shed light on a rigorous form of error-rate trade-off of blind data compression, leading to further understanding of efficient quantum data transmission.

The rest of this paper is organized as follows. We first review the basic setup and previous results of blind quantum data compression in Sec. II. Then we show and discuss our results on blind quantum data compression with finite local approximations in Sec. III. We construct a protocol that performs well compared with the previous one that works with asymptotically vanishing errors. We also show numerical experiments for blind compression of two-state classical ensembles. Finally, we summarize and discuss our results in Sec. IV.

II. PRELIMINARIES

In this section we briefly review basic concepts for blind quantum data compression. In Sec. II A we summarize relevant mathematical tools and notation. In Sec. II B we describe a structure of quantum ensembles, which is called Koashi-Imoto decomposition (KI decomposition). In Sec. II C we review the basic setups and previous results on quantum data compression.

A. Mathematical background and notation

Throughout this paper, we use capital letters $A, B, \ldots$ to represent quantum systems. For a quantum system A, we let $\mathcal{H}_A$ denote the corresponding complex Hilbert space. A vector $|\psi\rangle \in \mathcal{H}_A$ is called a pure state if it has unit norm. For convenience, we also say that $\mathcal{H}_A$ is a quantum system. We consider only finite-dimensional quantum systems in this paper. Given a quantum system $\mathcal{H}$, we let $d_{\mathcal{H}}$ denote the dimension of this system.

For a quantum system $\mathcal{H}$, let $L(\mathcal{H})$ denote the set of linear operators on $\mathcal{H}$. In particular, we define $I_{\mathcal{H}} \in L(\mathcal{H})$ as the identity operator on $\mathcal{H}$. Density operators are positive semidefinite operators with unit trace in $L(\mathcal{H})$. Moreover, let $D(\mathcal{H})$ denote the set of density operators on $\mathcal{H}$. When a density operator is rank 1, it can be written as $\rho = |\psi\rangle \langle\psi|$ using some state $|\psi\rangle \in \mathcal{H}$. Such a density operator will also be called a pure state. In general, a density operator has a spectral decomposition which is a convex combination of pure states; therefore, it can be interpreted as a probabilistic mixture of pure states. Therefore, a density operator also represents a quantum state. If $\rho \in D(\mathcal{H})$ is not pure, it is called a mixed state. For a set of states on a quantum system $\mathcal{H}$, if they are

simultaneously diagonalizable in a certain basis $\{|e_i\rangle\}_{i=1}^{d_{\mathcal{H}}}$, we can interpret each state as a distribution over the label i for the basis. We call such a set of states classical. In this case, we may also call each state classical.

Given two quantum systems $\mathcal{H}_A$ and $\mathcal{H}_B$ with $d_{\mathcal{H}_A} \leqslant d_{\mathcal{H}_B}$, we use $U(\mathcal{H}_A, \mathcal{H}_B)$ to denote the set of isometries from $\mathcal{H}_A$ to $\mathcal{H}_B$. That is, for $U \in U(\mathcal{H}_A, \mathcal{H}_B)$, it holds that $UU^\dagger = I_{\mathcal{H}_A}$, where a dagger represents the Hermitian conjugate. A linear map $\mathcal{N} : L(\mathcal{H}_A) \to L(\mathcal{H}_B)$ is called a quantum channel or a quantum operation if it is completely positive and trace preserving. Let $C(\mathcal{H}_A, \mathcal{H}_B)$ be the set of quantum channels from $L(\mathcal{H}_A)$ to $L(\mathcal{H}_B)$. In particular, we let $\mathcal{I}_{\mathcal{H}}$ denote the identity channel on $\mathcal{H}$.

We can introduce a norm on $L(\mathcal{H})$; let $\| \cdot \|_1$ be the trace norm. The trace norm is monotonic under quantum operations, that is, it holds that

$$\|\mathcal{N}(X)\|_1 \leqq \|X\| \tag{1}$$

for all linear operators X and quantum channels $\mathcal{N}$ applicable to X. For a quantum state $\rho \in D(\mathcal{H})$, the quantum entropy of ρ is defined as

$$H(\rho) := -\mathrm{Tr}(\rho \log_2 \rho). \tag{2}$$

It is known that the quantum entropy is upper bounded by the size of the system, that is,

$$H(\rho) \leqslant \log_2 d_{\mathcal{H}}. \tag{3}$$

B. Structure of quantum ensembles

We first review the definition of quantum ensembles, before describing a structure of them.

Definition 1 (quantum ensembles). Let $\mathcal{H}$ be a quantum system and let Σ be an alphabet. A quantum ensemble Φ is a set of pairs of a positive real number and a quantum state

$$\Phi := \{(p_a, \rho_a) \in \mathbb{R} \times D(\mathcal{H}) : a \in \Sigma\}, \tag{4}$$

where $\{p_a : a \in \Sigma\}$ forms a probability distribution, that is,

$$0 \leqslant p_a \leqslant 1 \tag{5}$$

for all $a \in \Sigma$ and

$$\sum_{a \in \Sigma} p_a = 1. \tag{6}$$

We also write

$$\Phi = \{p_a, \rho_a\}_{a \in \Sigma}. \tag{7}$$

In addition, the average state ρ_Φ of a quantum ensemble $\Phi = \{p_a, \rho_a\}_{a \in \Sigma}$ is defined as

$$\rho_\Phi := \sum_{a \in \Sigma} p_a \rho_a. \tag{8}$$

Here we define a special class of quantum ensembles. These are ensembles of classical states.

Definition 2 (classical ensembles). Let $\Phi = \{p_a, \rho_a\}_{a \in \Sigma}$ be a quantum ensemble. The ensemble Φ is said to be a classical ensemble if ρ_a are simultaneously diagonalizable for all $a \in \Sigma$, that is, there exists an orthonormal basis of $\mathcal{H}$ such that each ρ_a is diagonal.

We can also define another special class of quantum ensembles. A quantum ensemble $\{p_a, \rho_a\}_{a \in \Sigma}$ is called a pure-state ensemble if ρ_a is pure for all $a \in \Sigma$. Suppose $\rho_a = |\psi_a\rangle \langle \psi_a|$. We may use the shorthand $\{p_a, |\psi_a\rangle\}_{a \in \Sigma}$ to represent the ensemble.

Suppose we have a quantum ensemble $\Phi = \{p_a, \rho_a\}_{a \in \Sigma}$. Reference [27] gave a structure of a given quantum ensemble, which is called the Koashi-Imoto (KI) decomposition or the KI structure. Intuitively, when we have a quantum ensemble, we can decompose each state in the ensemble into the following three parts: a classical part, a nonredundant quantum part, and a redundant part. The following theorem provides a formal statement.

Theorem 1 [KI decomposition (from [27])]. Let $\mathcal{H}$ be a quantum system. Let $\Phi = \{p_a, \rho_a\}_{a \in \Sigma}$ be an ensemble on the system. Then there exists a decomposition of the quantum system

$$\mathcal{H} := \bigoplus_{l \in \Xi} \mathcal{H}_Q^{(l)} \otimes \mathcal{H}_R^{(l)} \qquad (9)$$

and a corresponding isometry

$$\Gamma_\Phi \in U\left(\mathcal{H}, \bigoplus_{l \in \Xi} \mathcal{H}_Q^{(l)} \otimes \mathcal{H}_R^{(l)}\right) \qquad (10)$$

satisfying the following conditions.

(a) For all $a \in \Sigma$,

$$\Gamma_\Phi \rho_a \Gamma_\Phi^\dagger = \bigoplus_{l \in \Xi} q^{(a,l)} \rho_Q^{(a,l)} \otimes \rho_R^{(l)}. \qquad (11)$$

Here, for each $a \in \Sigma$, $\{q^{(a,l)} : l \in \Xi\}$ forms a probability distribution over labels $l \in \Xi$, $\rho_Q^{(a,l)} \in D(\mathcal{H}_Q^l)$ is a density operator on system $\mathcal{H}_Q^l$, depending on both $a \in \Sigma$ and $l \in \Xi$, and $\rho_R^{(l)} \in D(\mathcal{H}_R^l)$ is a density operator on system $\mathcal{H}_R^l$, which is independent of $a \in \Sigma$.

(b) For each $l \in \Xi$, if a projection operator $P : \mathcal{H}_Q^{(l)} \rightarrow \mathcal{H}_Q^{(l)}$ satisfies

$$P q^{(a,l)} \rho_Q^{(a,l)} = q^{(a,l)} \rho_Q^{(a,l)} P \qquad (12)$$

for all $a \in \Sigma$, then $P = I_{\mathcal{H}_Q^{(l)}}$ or $P = 0$.

(c) For all $l, l' \in \Xi$ such that $l \neq l'$, there exists no isometry $V \in U(\mathcal{H}_Q^{(l)}, \mathcal{H}_Q^{(l')})$ such that

$$V q^{(a,l)} \rho_Q^{(a,l)} = \alpha q^{(a,l')} \rho_Q^{(a,l')} V \qquad (13)$$

with some positive real number α for all $a \in \Sigma$.

The first statement in Theorem 1 shows the form of the KI decomposition. The second and third statements ensure that the decomposition (11) is maximal, that is, we cannot further refine the structure. Indeed, the second one states that we cannot further decompose each block of the KI decomposition; the third one means that we cannot relate a block of the decomposition to another block. In the decomposition shown in Eq. (11), we can observe that all the quantum states in the given ensemble Φ can be decomposed in a block-diagonal structure. Note that $\rho_R^{(l)} \in D(\mathcal{H}_R^{(l)})$ does not depend on $a \in \Sigma$; thus it is called redundant because it does not contain information about the label $a \in \Sigma$. Hereafter, when we specify the KI decomposition of a given ensemble $\Phi = \{p_a, \rho_a\}_{a \in \Sigma}$, we
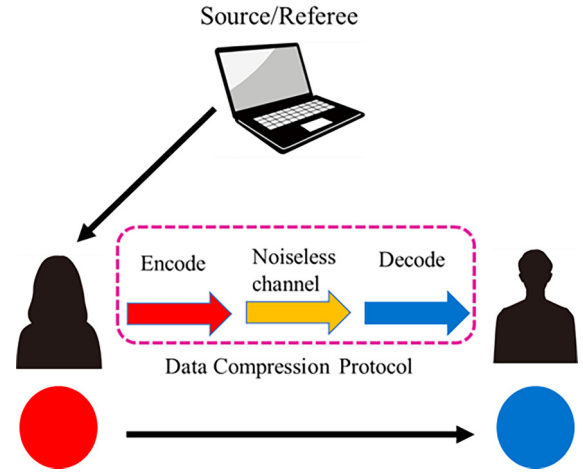

FIG. 1. Schematic diagram showing the setup of quantum data compression. In this task, the sender is given a target state (red circle) from the referee. Then the sender encodes the state and sends it to the receiver via a noiseless quantum channel. The receiver decodes the transmitted state to recover the original state (blue circle).

may omit Γ_Φ and write

$$\rho_a = \bigoplus_{l \in \Xi} q^{(a,l)} \rho_Q^{(a,l)} \otimes \rho_R^{(l)} \qquad (14)$$

for brevity.

Once the KI decomposition of a given ensemble is obtained, we can define two quantum channels that respectively correspond to removing and attaching the ensemble's redundant parts. The proof of this theorem is shown in Appendix A.

Theorem 2. Let $\Phi = \{p_a, \rho_a\}_{a \in \Sigma}$ be an ensemble on a quantum system $\mathcal{H}$. Consider the KI decomposition of Φ,

$$\mathcal{H} := \bigoplus_{l \in \Xi} \mathcal{H}_Q^{(l)} \otimes \mathcal{H}_R^{(l)}, \qquad (15)$$

such that

$$\rho_a = \bigoplus_{l \in \Xi} q^{(a,l)} \rho_Q^{(a,l)} \otimes \rho_R^{(l)}. \qquad (16)$$

Then there exist quantum channels $\mathcal{K}_{\text{off}}$ and $\mathcal{K}_{\text{on}}$ such that

$$\mathcal{K}_{\text{off}}(\rho_a) = \bigoplus_{l \in \Xi} q^{(a,l)} \rho_Q^{(a,l)}, \qquad (17)$$

$$\mathcal{K}_{\text{on}}\left(\bigoplus_{l \in \Xi} q^{(a,l)} \rho_Q^{(a,l)}\right) = \rho_a \qquad (18)$$

for all $a \in \Sigma$. Using $\mathcal{K}_{\text{off}}$ and $\mathcal{K}_{\text{on}}$, we can reversibly remove the redundant parts from the ensemble. In this paper we refer to $\mathcal{K}_{\text{off}}$ and $\mathcal{K}_{\text{on}}$ as the KI operations.

C. Review of quantum data compression problems

We discuss the basic setup for quantum data compression, the variations studied to date, and known results about them.

Quantum data compression is a quantum information processing task between two parties, the sender and the receiver (see Fig. 1). In this task, the sender aims to transmit quantum data to the receiver as efficiently as possible, given the assumption that the data consists of asymptotically large number

of states created independently. The best rate captures fundamental quantum properties of the transmitted states.

We provide a more formal description below. Suppose that $\{p_x, \rho_x\}_{x \in \Sigma}$ is a quantum ensemble on system A. A referee draws $x \in \Sigma$ with probability p_x, writes down the outcome in a system X (a classical system) with Hilbert space $\mathcal{H}_X :=$ $\mathbb{C}^\Sigma$, and prepares ρ_x in system A. The resulting state is given by $\rho_{XA} := \sum_{x \in \Sigma} p_x |x\rangle \langle x| \otimes \rho_x$. Note that $\rho_A = \sum_x p_x \rho_x$ is the average state of the ensemble. If the referee repeats the above process n times independently, then the resulting state is

$$
\begin{aligned}
&\rho_{X_1 A_1} \otimes \rho_{X_2 A_2} \otimes \cdots \otimes \rho_{X_n A_n} \\
&= \sum_{x_1, \ldots, x_n \in \Sigma} p_{x_1} p_{x_2} \cdots p_{x_n} |x_1\rangle \langle x_1| \otimes \cdots \otimes |x_n\rangle \langle x_n| \\
&\quad \otimes \rho_{x_1} \otimes \rho_{x_2} \otimes \cdots \otimes \rho_{x_n},
\end{aligned}
\tag{19}
$$

where $|x_i\rangle\langle x_i|$ is a state on X_i and ρ_{x_i} is a state on A_i. We use the shorthand x^n for $x_1 x_2 \cdots x_n$, p_{x^n} for $p_{x_1} p_{x_2} \cdots p_{x_n}$, ρ_{x^n} for $\rho_{x_1} \otimes \rho_{x_2} \otimes \cdots \otimes \rho_{x_n}$, and $\rho_{X^n A^n}$ for $\rho_{X_1 A_1} \otimes \rho_{X_2 A_2} \otimes \cdots \otimes \rho_{X_n A_n}$. Then the referee gives the systems $A_1 A_2 \cdots A_n$ to the sender. After receiving $A_1 \cdots A_n$ from the referee, the sender applies to them an encoding channel $\mathcal{E}_n \in C(\otimes_{i=1}^n \mathcal{H}_{A_i}, \mathcal{H}_{M_n})$. When the assistance of entanglement is allowed, the sender jointly encodes $A_1 \cdots A_n$ with their share of a given entangled state. Afterward, the sender transmits the system M_n to the receiver through a noiseless quantum channel. The receiver decompresses the state by a decoding channel $\mathcal{D}_n \in C(\mathcal{H}_{M_n}, \otimes_{i=1}^n \mathcal{H}_{A_i'})$, where $A_i' \simeq A_i$ for all i. The final state $\mathcal{I}_{\mathcal{H}_X}^{\otimes n} \otimes (\mathcal{D}_n \circ \mathcal{E}_n)(\rho_{X^n A^n})$ has to resemble the initial state $\rho_{X^n A^n}$ in a certain way. The compression protocol is defined as a pair $(\mathcal{E}_n, \mathcal{D}_n)$ of the encoding and decoding channels, but we also refer to their combined effect $\mathcal{D}_n \circ \mathcal{E}_n$ as the protocol.

The compression rate R_n of the protocol is given by the number of qubits transmitted divided by n:

$$
R_n := \frac{\log_2 |M_n|}{n}.
\tag{20}
$$

In the basic setup shown above, the protocol and compression rate depend crucially on the following: (i) whether the given ensemble consists of pure states or mixed states, (ii) whether the referee tells the sender the label of the state to be compressed, and (iii) what error criterion we adopt between the original state and the one recovered by the receiver. In the second point, if the sender knows the label, the task is called visible compression; otherwise, the task is called blind compression. By definition, a blind compression protocol also works in the visible setting. Correspondingly, the compression rate in the blind setting is no less than that in the visible setting. On the third point, two definitions of error criteria have been considered: global error criterion and local error criterion. Under the global error criterion, a protocol satisfying

$$
\left\| \mathcal{I}_{\mathcal{H}_X}^{\otimes n} \otimes (\mathcal{D}_n \circ \mathcal{E}_n)(\rho_{X^n A^n}) - \rho_{X^n A^n} \right\|_1 \leqslant \epsilon_n
\tag{21}
$$

is said to have global error ϵ_n. In this case, the entire state emerges close to the entire original state. On the other hand, under the local error criterion, a protocol satisfying

$$
\left\| \mathrm{Tr}_{\bar{X}_k \bar{B}_k} \left[\mathcal{I}_{\mathcal{H}_X}^{\otimes n} \otimes (\mathcal{D}_n \circ \mathcal{E}_n)(\rho_{X^n A^n}) \right] - \rho_{XA'} \right\|_1 \leqslant \epsilon_n
\tag{22}
$$

for all integers $1 \leqslant k \leqslant n$ is said to have local error ϵ_n. In this case, the initial and final states should be close letterwise. In these definitions, $\mathrm{Tr}_{(\bar{\cdot})_k}$ denotes the partial trace over all systems other than the kth one. The global error criterion implies the local error criterion, but the converse does not always hold.

We consider an asymptotic scenario where the number n of states is sufficiently large. We call a pair $(\mathcal{E}_n, \mathcal{D}_n)$ with an encoding channel $\mathcal{E}_n$ and a decoding channel $\mathcal{D}_n$ an (n, R_n, ϵ_n) code if the protocol $\mathcal{D}_n \circ \mathcal{E}_n$ yields the rate R_n within error ϵ_n under the error criterion we adopt. Then we say that the rate R is achievable if for any $\epsilon > 0$ and $\delta > 0$ there exists a positive integer n_0 such that for all integers $n \geqslant n_0$ we can construct an $(n, R + \delta, \epsilon)$ code. Similarly, we say that the rate R is achievable within error ϵ if for any $\delta > 0$ there exists a positive integer n_0 such that for all integers $n \geqslant n_0$ we can construct an $(n, R + \delta, \epsilon)$ code.

Next we will summarize the state-of-the-art understanding of quantum data compression in these setups. Quantum data compression was first studied in Refs. [2–4]. These groundbreaking works focused on blind compression of pure-state ensembles under the global error criterion. It was shown that the optimal compression rate is the quantum entropy of the average state of the given pure-state ensemble, that is, letting $\{p_a, |\psi_a\rangle\}_{a \in \Sigma}$ denote a given ensemble, the optimal rate is given by the entropy of the average state

$$
H\left(\sum_{a \in \Sigma} p_a |\psi_a\rangle \langle \psi_a| \right).
\tag{23}
$$

A particularly important compression protocol, named after its inventor as Schumacher compression, was proposed in Ref. [2] and reviewed in various textbooks such as Refs. [28–30].

After the aforementioned initial studies of blind compression of pure states, visible compression was subsequently considered. For a given quantum ensemble $\Phi = \{p_a, \rho_a\}_{a \in \Sigma}$, which is not necessarily pure, the optimal rate is lower bounded by the Holevo information [6,7] defined as

$$
I(\Phi) := H\left(\sum_{a \in \Sigma} p_a \rho_a \right) - \sum_{a \in \Sigma} p_a H(\rho_a).
\tag{24}
$$

Observe that for a pure-state ensemble, the Holevo information is equal to the entropy of the average state since the entropy of a pure state is zero. Therefore, the optimal rate of the visible compression of a pure-state ensemble is also the entropy of the average state. Thus, surprisingly, for compression of a pure-state ensemble under the global error criterion, whether the compression is visible or blind does not make any difference to the optimal rate.

After the extensive investigation of compression of pure-state ensembles, compression of a general mixed-state ensemble became of central interest. Historically, the optimal rate of visible compression is called the effective information; that of blind compression is called the passive information. Letting $I_e(\Phi)$ denote the effective information of a quantum ensemble Φ and $I_p(\Phi)$ denote the passive information of a quantum ensemble Φ, we have the relation

$$
I(\Phi) \leqslant I_e(\Phi) \leqslant I_p(\Phi)
\tag{25}
$$

since visible compression can be regarded as a subclass of blind compression. The quantity $I_d := I_p - I_e$ is called the information defect and it characterizes a difference between the visible and blind compression tasks. In Refs. [6,7], a lower bound of the information defect was given, and Refs. [8,9] showed examples for which the information defect is strictly positive. Moreover, the optimal rate of visible compression was derived in Ref. [10], and the optimal rate is given by the entropy of an extension of a given state. In Ref. [11] the authors gave another representation of the optimal rate of the visible compression. On the other hand, in Ref. [12] the optimal rate of blind compression was studied. The main difficulty of blind compression is that the sender does not know the label of a given state. The authors proposed a protocol in which the sender only sends the essential parts of a given ensemble and the compression rate achieved by the protocol is given by the KI decomposition [27]. This optimal protocol is based on a block coding of mixed states proposed in Ref. [5]. In fact, this compression rate is optimal under both the global and local error criteria.

Note that we mainly discuss quantum data compression without any assistance; however, data compression tasks with assistance, e.g., entanglement and shared randomness, have also been widely investigated [13–19].

III. COMPRESSION PROTOCOL USING FINITE LOCAL APPROXIMATIONS

In this section we show our results on the quantum blind compression with finite local approximations. In particular, we show our protocol and present some examples for which the protocol leads to a large reduction of the compression rate compared to the previous results with asymptotically vanishing errors. Moreover, as a first step in the general understanding of blind compression with finite approximations, we focus on quantum ensembles consisting of two classical states. Through numerical experiments, we reveal the performance of our protocol for classical ensembles.

A. Asymptotic optimal rate of blind quantum data compression

Blind data compression is a quantum information processing task between two parties, the sender and the receiver, in which fundamental quantum properties emerge (see Fig. 2). As also mentioned in Sec. II C, in blind data compression, the sender aims to asymptotically send quantum data, without knowing its actual description, to the receiver as efficiently as possible. Here we describe a formal procedure of blind quantum data compression following the setup shown in Sec. II C. Let $\{p_x, \rho_x\}_{x \in \Sigma}$ be a quantum ensemble the sender wants to send. The referee draws a state from the ensemble n times in an independent and identically distributed manner; let the resulting n-fold state be denoted by ρ_{x^n}. Then the referee gives the state ρ_{x^n} to the sender. What is crucial in this setup is that the referee does not tell the sender the label x^n of the state. After receiving the state ρ_{x^n} from the referee, the sender encodes the state by an encoding channel $\mathcal{E}_n$ and transmits the compressed state to the receiver through a noiseless quantum channel. The receiver decompresses the state by a decoding channel $\mathcal{D}_n$ to recover the initial state ρ_{x_n}. Note that they do

Source/Referee

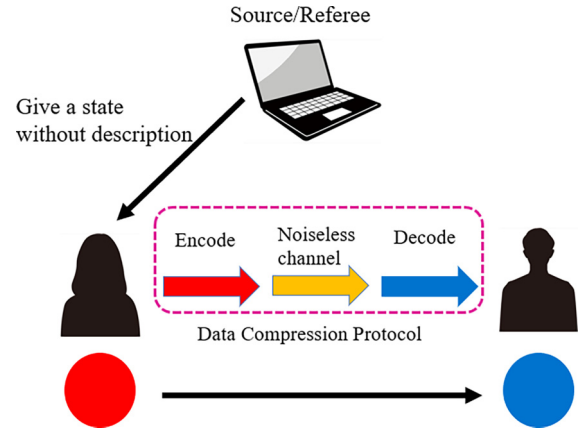Give a state
without description

FIG. 2. Schematic diagram showing the setup of blind quantum data compression. In addition to the basic data compression setup, the sender cannot receive any information about the description or the label of a given state.

not know the description of the given state while the sender has knowledge about the ensemble. Therefore, the compression scheme $(\mathcal{E}_n, \mathcal{D}_n)$ cannot be state specific and must only depend on the ensemble.

Now we provide an overview of the achievable rates of blind data compression. In Ref. [5] an achievable rate of blind state compression with a mixed state ensemble was investigated. It was figured out that the quantum entropy of the average state is an achievable rate even for a mixed state ensemble, that is, for a given ensemble $\{p_x, \rho_x\}_{x \in \Sigma}$,

$$H\left(\sum_{x \in \Sigma} p_x \rho_x\right) \tag{26}$$

is achievable. It had been open whether this rate is optimal or not [5–9] before Ref. [12] conducted a more detailed analysis.

In Ref. [12] the authors applied the KI decomposition (Theorem 1 in Sec. II B) to derive the optimal rate of blind compression. Recall that when an ensemble $\{p_x, \rho_x\}_x$ is given, we have the KI decomposition

$$\rho_x = \bigoplus_{l \in \Xi} q^{(x,l)} \rho_Q^{(x,l)} \otimes \rho_R^{(l)}. \tag{27}$$

Then, as shown in Theorem 2, we can define quantum channels $\mathcal{K}_{\text{off}}$ and $\mathcal{K}_{\text{on}}$ satisfying

$$\mathcal{K}_{\text{off}}(\rho_x) = \bigoplus_{l \in \Xi} q^{(x,l)} \rho_Q^{(x,l)}, \tag{28}$$

$$\mathcal{K}_{\text{on}}\left(\bigoplus_{l \in \Xi} q^{(x,l)} \rho_Q^{(x,l)}\right) = \rho_x \tag{29}$$

for all $x \in \Sigma$. With these operations, the sender does not necessarily send the redundant parts of a target ensemble if the sender and the receiver agree on the ensemble in the scenario of blind data compression. If they both know the description of the ensemble, they also agree on $\mathcal{K}_{\text{off}}$ and $\mathcal{K}_{\text{on}}$, that is, they can freely take off and put on the redundant parts. Therefore, the sender only needs to send the classical and nonredundant quantum parts. By applying the argument of Ref. [5] after removing the redundant parts, we obtain an

achievable rate

$$R^* = H\left(\sum_{x \in \Sigma} p_x \bigoplus_{l \in \Xi} q^{(x,l)} \rho_Q^{(x,l)}\right). \tag{30}$$

In addition, Ref. [12] also proved that Eq. (26) is optimal even under local error criterion. To summarize, we have the following theorem.

Theorem 3 (from [12]). Let $\Phi = \{p_x, \rho_x\}_{x \in \Sigma}$ be a quantum ensemble with KI decomposition

$$\rho_x = \bigoplus_{l \in \Xi} q^{(x,l)} \rho_Q^{(x,l)} \otimes \rho_R^{(l)}. \tag{31}$$

Then

$$R^* = H\left(\sum_{x \in \Sigma} p_x \bigoplus_{l \in \Xi} q^{(x,l)} \rho_Q^{(x,l)}\right) \tag{32}$$

is the optimal rate of blind compression of Φ for asymptotically vanishing errors under both the local and the global error criterion.

To show the optimality of the rate, the authors of Ref. [12] introduced two error functions of compression protocol $\Lambda_n := (\mathcal{E}_n, \mathcal{D}_n)$,

$$f(\Lambda_n) := 1 - \sum_{x^n \in \Sigma^n} p_{x^n} F(\rho_{x^n}, \Lambda_n(\rho_{x^n})), \tag{33}$$

$$g(\Lambda_n) := h_2(\Delta) + \Delta \log_2(d_A - 1), \tag{34}$$

with

$$\Delta := 1 - \sum_{i=1}^{d_A} \lambda_i(\rho) \langle i | \Lambda_n(|i\rangle \langle i|) | i \rangle. \tag{35}$$

Here $F(\cdot, \cdot)$ is the fidelity function and ρ is the average state of the given ensemble with the spectral decomposition

$$\rho = \sum_{i=1}^{d_A} \lambda_i(\rho) |i\rangle \langle i|. \tag{36}$$

Moreover, $h_2(\cdot)$ is the binary entropy function defined as

$$h_2(x) := -x \log_2 x - (1-x) \log_2(1-x) \tag{37}$$

for $x \in [0, 1]$. It was shown that when the protocol Λ_n is given, the compression rate R is

$$R \geqslant R^* - g(\Lambda_n). \tag{38}$$

The authors proved the optimality by showing that when $f(\Lambda_n)$ goes to zero, $g(\Lambda_n)$ also converges to zero, that is, for an infinitesimal error, R^* is the smallest compression rate that can be achievable.

Recently, Ref. [19] further investigated the optimality of the compression rate under the global error criterion. The authors introduced an error function that represents how an error allowed in the protocol affects the compression rate and they revealed properties of this error function to study how the compression rate depends on an error. On the other hand, in this paper we focus on local error criterion to investigate the sensitivity of the rate against approximations.

B. Our protocol

In this section we provide motivation for and describe our protocol for blind quantum data compression. Here we consider local error criterion, that is, we only require the resulting state to be close to the original state letterwise. In addition, we allow finite approximations to the resulting state under this error criterion.

First, we summarize the compression scheme from the preceding section.

(i) After receiving a state ρ_{x^n} from the referee, the sender applies $\mathcal{K}_{\text{off}}$ letterwise to remove the redundant parts of the state and obtain $\mathcal{K}_{\text{off}}^{\otimes n}(\rho_{x^n})$.

(ii) The sender encodes the state $\mathcal{K}_{\text{off}}^{\otimes n}(\rho_{x^n})$ using an encoding channel $\mathcal{E}_n$ to obtain $\mathcal{E}_n \circ \mathcal{K}_{\text{off}}^{\otimes n}(\rho_{x^n})$.

(iii) The sender sends the state $\mathcal{E}_n \circ \mathcal{K}_{\text{off}}^{\otimes n}(\rho_{x^n})$ via a noiseless quantum channel to the receiver.

(iv) The receiver decodes the transmitted state using a decoding channel $\mathcal{D}_n$ and obtains $\mathcal{D}_n \circ \mathcal{E}_n \circ \mathcal{K}_{\text{off}}^{\otimes n}(\rho_{x^n})$.

(v) The receiver put the redundant parts back using the quantum channel $\mathcal{K}_{\text{on}}$; the resulting state is $\mathcal{K}_{\text{on}}^{\otimes n} \circ \mathcal{D}_n \circ \mathcal{E}_n \circ \mathcal{K}_{\text{off}}^{\otimes n}(\rho_{x^n})$.

Our main idea is as follows. We may use quantum channels $\Lambda_s^{(n)}$ and $\Lambda_r^{(n)}$, potentially acting jointly on the entire state, instead of $\mathcal{K}_{\text{off}}^{\otimes n}$ and $\mathcal{K}_{\text{on}}^{\otimes n}$. The encoding and decoding maps $\mathcal{E}_n$ and $\mathcal{D}_n$ may change accordingly. When we allow a finite local approximation of the given state ρ_{x^n} by the resulting state $\Lambda_s^{(n)} \circ \mathcal{D}_n \circ \mathcal{E}_n \circ \Lambda_r^{(n)}(\rho_{x^n})$, a smaller rate may be achieved than the previous optimal rate with a vanishing error. We show that even if $\Lambda_s^{(n)}$ and $\Lambda_r^{(n)}$ are letterwise approximations, it is plausible that a reduction of the compression rate may be observed.

In our protocol, we use the KI operations of a different quantum ensemble on the given ensemble and prove that this approximation procedure indeed performs well and reduces the rate. In more detail, suppose we are given a quantum ensemble $\{p_x, \rho_x\}_{x \in \Sigma}$ and the resulting state can differ letterwise from the original state up to $\epsilon > 0$. We consider all possible approximate ensembles $\{p_x, \tilde{\rho}_x\}_{x \in \Sigma}$ such that $\tilde{\rho}_x$ is close to ρ_x in the original given ensemble for each x. Each such approximate ensemble defines KI operations $\tilde{\mathcal{K}}_{\text{off}}$ and $\tilde{\mathcal{K}}_{\text{on}}$ as in Theorem 2. Now suppose we use these KI operations $\tilde{\mathcal{K}}_{\text{off}}^{\otimes n}$ and $\tilde{\mathcal{K}}_{\text{on}}^{\otimes n}$ as an approximation scheme $\Lambda_s^{(n)}$ and $\Lambda_r^{(n)}$ for $\{p_x, \rho_x\}_{x \in \Sigma}$. How much error do we incur? If the error is small, what rate can be achieved? Before we answer these questions, let us formally define our compression protocol with local approximations.

(0) The sender and receiver preagree on an approximate ensemble $\{p_x, \tilde{\rho}_x\}_{x \in \Sigma}$ of the original given ensemble $\{p_x, \rho_x\}_{x \in \Sigma}$ such that for all $x \in \Sigma$,

$$\|\tilde{\mathcal{K}}_{\text{on}} \circ \tilde{\mathcal{K}}_{\text{off}}(\rho_x) - \rho_x\|_1 \leqslant \epsilon. \tag{39}$$

(1) The referee gives a state ρ_{x^n} generated by n independent and identically distributed draws from $\{p_x, \rho_x\}_{x \in \Sigma}$ to the sender. The sender applies $\tilde{\mathcal{K}}_{\text{off}}^{\otimes n}$ to the given state, which yields $\tilde{\mathcal{K}}_{\text{off}}^{\otimes n}(\rho_{x^n})$.

(2) The sender applies an encoding operation $\mathcal{E}_n$ to obtain $\mathcal{E}_n \circ \tilde{\mathcal{K}}_{\text{off}}^{\otimes n}(\rho_{x^n})$.

(3) The sender transmits the encoded state $\mathcal{E}_n \circ \tilde{\mathcal{K}}_{\text{off}}^{\otimes n}(\rho_{x^n})$ to the receiver via a noiseless quantum channel.

(4) Upon receiving the state, the receiver applies the decoding channel $\mathcal{D}_n$ corresponding to $\mathcal{E}_n$, which yields $\mathcal{D}_n \circ \mathcal{E}_n \circ \tilde{\mathcal{K}}_{\text{off}}^{\otimes n}(\rho_{x^n})$.

(5) Finally, the receiver applies $\tilde{\mathcal{K}}_{\text{on}}^{\otimes n}$ to obtain the resulting state $\tilde{\mathcal{K}}_{\text{on}}^{\otimes n} \circ \mathcal{D}_n \circ \mathcal{E}_n \circ \tilde{\mathcal{K}}_{\text{off}}^{\otimes n}(\rho_{x^n})$.

The resulting state $\tilde{\mathcal{K}}_{\text{on}}^{\otimes n} \circ \mathcal{D}_n \circ \mathcal{E}_n \circ \tilde{\mathcal{K}}_{\text{off}}^{\otimes n}(\rho_{x^n})$ of our protocol should be close to the original state ρ_{x^n}. In step 2 of the protocol, we choose $\mathcal{E}_n$ to achieve blind compression of the ensemble $\{p_x, \tilde{\mathcal{K}}_{\text{off}}(\rho_x)\}_{x \in \Sigma}$ with vanishing error, that is, for any $\delta > 0$, there exists a positive integer n_0 such that for all

integers $n \geqslant n_0$,

$$\left\| \mathcal{D}_n \circ \mathcal{E}_n \circ \tilde{\mathcal{K}}_{\text{off}}^{\otimes n}(\rho_{x^n}) - \tilde{\mathcal{K}}_{\text{off}}^{\otimes n}(\rho_{x^n}) \right\|_1 \leqslant \delta. \tag{40}$$

Together with the condition (39), local error will be $\leqslant \epsilon + \delta$. In the rest of the paper we omit δ, considering that we may take δ arbitrarily small by choosing n appropriately.

In our protocol, the condition (39) is crucial. Here we give a sufficient condition for Eq. (39); we prove that $\tilde{\mathcal{K}}_{\text{off}}$ and $\tilde{\mathcal{K}}_{\text{on}}$ satisfy the local error criterion for $\{p_x, \rho_x\}_{x \in \Sigma}$ if $\{p_x, \tilde{\rho}_x\}_{x \in \Sigma}$ is close enough to the original ensemble. Suppose that $\{p_x, \rho_x\}_{x \in \Sigma}$ is a quantum ensemble and let $\{p_x, \tilde{\rho}_x\}_{x \in \Sigma}$ be a quantum ensemble such that

$$\|\rho_x - \tilde{\rho}_x\|_1 \leqslant \frac{\epsilon}{2} \tag{41}$$

for all x. Then

$$
\begin{aligned}
\|\tilde{\mathcal{K}}_{\text{on}} \circ \tilde{\mathcal{K}}_{\text{off}}(\rho_x) - \rho_x\|_1 &= \|[\tilde{\mathcal{K}}_{\text{on}} \circ \tilde{\mathcal{K}}_{\text{off}}(\rho_x) - \tilde{\mathcal{K}}_{\text{on}} \circ \tilde{\mathcal{K}}_{\text{off}}(\tilde{\rho}_x)] + [\tilde{\mathcal{K}}_{\text{on}} \circ \tilde{\mathcal{K}}_{\text{off}}(\tilde{\rho}_x) - \rho_x]\|_1 \\
&\leqslant \|\tilde{\mathcal{K}}_{\text{on}} \circ \tilde{\mathcal{K}}_{\text{off}}(\rho_x) - \tilde{\mathcal{K}}_{\text{on}} \circ \tilde{\mathcal{K}}_{\text{off}}(\tilde{\rho}_x)\|_1 + \|\tilde{\mathcal{K}}_{\text{on}} \circ \tilde{\mathcal{K}}_{\text{off}}(\tilde{\rho}_x) - \rho_x\|_1 \\
&\leqslant 2\|\tilde{\rho}_x - \rho_x\|_1 \\
&\leqslant \epsilon.
\end{aligned}
$$

The second line follows from the triangle inequality. In the third line, we used the monotonicity of the trace norm and the relation $\tilde{\mathcal{K}}_{\text{on}} \circ \tilde{\mathcal{K}}_{\text{off}}(\tilde{\rho}_x) = \tilde{\rho}_x$. Thus,

$$\|\tilde{\mathcal{K}}_{\text{on}} \circ \tilde{\mathcal{K}}_{\text{off}}(\rho_x) - \rho_x\|_1 \leqslant \epsilon, \tag{42}$$

that is, a quantum ensemble for which each state is $\epsilon/2$-close to the state in the original ensemble works as an approximate ensemble for our protocol.

Second, we analyze the rate performance of our protocol. The compression rate of our protocol is determined by $\mathcal{E}_n$; a rate of

$$R = H\left(\sum_{x \in \Sigma} p_x \tilde{\mathcal{K}}_{\text{off}}(\rho_x)\right) \tag{43}$$

is sufficient to ensure the vanishing error condition (40). In step 0 of the protocol, the sender and the receiver can minimize the above rate by optimizing the approximate ensemble:

$$H\left(\sum_x p_x \tilde{\mathcal{K}}_{\text{off}}(\rho_x)\right) \leqslant H\left(\sum_x p_x \mathcal{K}_{\text{off}}(\rho_x)\right). \tag{44}$$

In this protocol, we need to find a good approximate ensemble to reduce the compression rate. While this optimization may not be intractable, it suffices to find a good enough approximate ensemble to substantially reduce the rate, as we will see in the next section.

C. Reduction of rates with a finite error

Here we present two main examples for which our protocol performs better than the case where finite approximations are not allowed. The first example is over a small-dimensional system and it illustrates the procedure and performance of our protocol in an intuitive way; the second example is over a

growing dimension and exhibits a growing reduction of the compression rate.

First, we consider a four-dimensional two-state ensemble to intuitively understand the protocol.

Example 1. Let $\epsilon > 0$ be a fixed positive number sufficiently smaller than $\frac{1}{2}$. Consider the following two density operators:

$$\rho_1 := \frac{1}{4}\begin{pmatrix} 2 - \frac{\epsilon}{2} & 1 - \epsilon & 0 & 0 \\ 1 - \epsilon & 2 - \frac{3\epsilon}{2} & 0 & 0 \\ 0 & 0 & \frac{3\epsilon}{2} & 0 \\ 0 & 0 & 0 & \frac{\epsilon}{2} \end{pmatrix}, \tag{45}$$

$$\rho_2 := \frac{1}{4}\begin{pmatrix} \epsilon & 0 & 0 & 0 \\ 0 & \epsilon & 0 & 0 \\ 0 & 0 & 2 - \epsilon & 1 \\ 0 & 0 & 1 & 2 - \epsilon \end{pmatrix}. \tag{46}$$

We consider an ensemble $\{p_x, \rho_x\}_{x=1,2}$ where $p_1 = p_2 = \frac{1}{2}$. Note that this ensemble has no redundant parts while ρ_1 and ρ_2 are written in a block-diagonal form. Here we construct an approximate ensemble $\{p_x, \tilde{\rho}_x\}_{x=1,2}$ as

$$\tilde{\rho}_1 := \frac{1}{4}\begin{pmatrix} 2 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{47}$$

$$\tilde{\rho}_2 := \frac{1}{4}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 1 & 2 \end{pmatrix}. \tag{48}$$

Observe that $\tilde{\rho}_1$ and $\tilde{\rho}_2$ can be written as

$$\tilde{\rho}_1 := |0\rangle\langle 0| \otimes \frac{1}{4}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \tag{49}$$

$$\tilde{\rho}_2 := |1\rangle\langle 1| \otimes \frac{1}{4}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}. \tag{50}$$

Therefore,

$$\omega := \frac{1}{4}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \tag{51}$$

is a redundant part of the approximate ensemble. Then we can define the KI operations corresponding to this redundant part as

$$\tilde{\mathcal{K}}_{\text{off}}(\cdot) := \text{Tr}_2(\cdot), \tag{52}$$

$$\tilde{\mathcal{K}}_{\text{on}}(\cdot) := [(|0\rangle\langle 0| \cdot |0\rangle\langle 0|) + (|1\rangle\langle 1| \cdot |1\rangle\langle 1|)] \otimes \omega, \tag{53}$$

where the partial trace Tr_2 is taken over the second qubit system. With these operations,

$$\tilde{\mathcal{K}}_{\text{off}}(\rho_1) = \begin{pmatrix} 1 - \frac{\epsilon}{2} & 0 \\ 0 & \frac{\epsilon}{2} \end{pmatrix} = \left(1 - \frac{\epsilon}{2}\right)|0\rangle\langle 0| + \frac{\epsilon}{2}|1\rangle\langle 1|,$$

$$\tilde{\mathcal{K}}_{\text{off}}(\rho_2) = \begin{pmatrix} \frac{\epsilon}{2} & 0 \\ 0 & 1 - \frac{\epsilon}{2} \end{pmatrix} = \frac{\epsilon}{2}|0\rangle\langle 0| + \left(1 - \frac{\epsilon}{2}\right)|1\rangle\langle 1|.$$

In addition,

$$\tilde{\mathcal{K}}_{\text{on}} \circ \tilde{\mathcal{K}}_{\text{off}}(\rho_1) = \frac{1}{8}\begin{pmatrix} 4 - 2\epsilon & 2 - \epsilon & 0 & 0 \\ 2 - \epsilon & 4 - 2\epsilon & 0 & 0 \\ 0 & 0 & 2\epsilon & \epsilon \\ 0 & 0 & \epsilon & 2\epsilon \end{pmatrix}$$

$$= \left[\left(1 - \frac{\epsilon}{2}\right)|0\rangle\langle 0| + \frac{\epsilon}{2}|1\rangle\langle 1|\right] \otimes \omega,$$

$$\tilde{\mathcal{K}}_{\text{on}} \circ \tilde{\mathcal{K}}_{\text{off}}(\rho_2) = \frac{1}{8}\begin{pmatrix} 2\epsilon & \epsilon & 0 & 0 \\ \epsilon & 2\epsilon & 0 & 0 \\ 0 & 0 & 4 - 2\epsilon & 2 - \epsilon \\ 0 & 0 & 2 - \epsilon & 4 - 2\epsilon \end{pmatrix}$$

$$= \left[\frac{\epsilon}{2}|0\rangle\langle 0| + \left(1 - \frac{\epsilon}{2}\right)|1\rangle\langle 1|\right] \otimes \omega.$$

Hence,

$$\|\tilde{\mathcal{K}}_{\text{on}} \circ \tilde{\mathcal{K}}_{\text{off}}(\rho_1) - \rho_1\|_1 = \left\|\frac{1}{8}\begin{pmatrix} -\epsilon & \epsilon & 0 & 0 \\ \epsilon & \epsilon & 0 & 0 \\ 0 & 0 & -\epsilon & \epsilon \\ 0 & 0 & \epsilon & \epsilon \end{pmatrix}\right\|_1$$

$$= \frac{\sqrt{2}\epsilon}{2} \leqslant \epsilon,$$

$$\|\tilde{\mathcal{K}}_{\text{on}} \circ \tilde{\mathcal{K}}_{\text{off}}(\rho_2) - \rho_2\|_1 = \left\|\frac{1}{8}\begin{pmatrix} 0 & \epsilon & 0 & 0 \\ \epsilon & 0 & 0 & 0 \\ 0 & 0 & 0 & -\epsilon \\ 0 & 0 & -\epsilon & 0 \end{pmatrix}\right\|_1$$

$$= \frac{\epsilon}{2} \leqslant \epsilon.$$

Therefore, the KI operations $\tilde{\mathcal{K}}_{\text{off}}$ and $\tilde{\mathcal{K}}_{\text{on}}$ satisfy the condition (39).

We now see that this example exhibits a reduction of the rate compared to the compression rate under asymptot-

ically vanishing errors. Let R_0 denote the optimal rate for $\{p_x, \rho_x\}_{x \in \Sigma}$; let R denote the rate for $\{p_x, \rho_x\}_{x \in \Sigma}$ obtained by $\tilde{\mathcal{K}}_{\text{off}}$ and $\tilde{\mathcal{K}}_{\text{on}}$. Then we have

$$R_0 = H\left(\frac{1}{2}\rho_1 + \frac{1}{2}\rho_2\right) \approx \log_2 4 = 2, \tag{54}$$

$$R = H\left[\frac{1}{2}\tilde{\mathcal{K}}_{\text{off}}(\rho_1) + \frac{1}{2}\tilde{\mathcal{K}}_{\text{off}}(\rho_2)\right] \approx \log_2 2 = 1. \tag{55}$$

Thus, this approximation reduces the compression rate to about half of the original rate.

Next we show an example for which a finite approximation dramatically changes the KI structure of a given ensemble. For this example, we can see a large reduction of the compression rate compared with the compression rate under asymptotically vanishing errors.

Example 2. Let $0 < \epsilon < \frac{1}{2}$ be a fixed positive number. Let $\omega_a \in \mathcal{D}(\mathcal{H}_a)$ and $\omega_b \in \mathcal{D}(\mathcal{H}_b)$ be density operators. Let N be a positive integer and define $2N$-dimensional density operators

$$\sigma_1 := \frac{1}{4N}\begin{pmatrix} 2 & \epsilon e^{i\alpha} & 0 & \cdots & \epsilon e^{-i\alpha} \\ \epsilon e^{-i\alpha} & 2 & \epsilon e^{i\alpha} & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \epsilon e^{-i\alpha} & 2 & \epsilon e^{i\alpha} \\ \epsilon e^{i\alpha} & 0 & \cdots & \epsilon e^{-i\alpha} & 2 \end{pmatrix},$$

$$\sigma_2 := \frac{1}{4N}\begin{pmatrix} 1 + 2\epsilon & & & & \\ & \ddots & & & \\ & & 1 + 2\epsilon & & \\ & & & 3 - 2\epsilon & \\ & & & & \ddots & \\ & & & & & 3 - 2\epsilon \end{pmatrix},$$

with $0 < \alpha < 1/4N$. Here the off-diagonal elements of σ_2 are all zero and we omit writing these elements in the matrix form above. The states σ_1 and σ_2 do not have redundant parts. We show the proof in Appendix C.

With these density operators, let us define a quantum ensemble $\{p_x, \rho_x\}_{x=1,2}$ with

$$p_1 = p_2 = \frac{1}{2}, \tag{56}$$

$$\rho_1 := \frac{1}{3}\omega_a \oplus \frac{1}{3}\sigma_1 \oplus \frac{1}{3}\omega_b, \tag{57}$$

$$\rho_2 := \frac{1}{6}\omega_a \oplus \frac{1}{3}\sigma_2 \oplus \frac{1}{2}\omega_b, \tag{58}$$

where ω_a and ω_b are density operators so that the KI decomposition of ρ_1 and ρ_2 is given by Eqs. (57) and (58). Note that ρ_1 and ρ_2 already have redundant parts ω_a and ω_b. Now we introduce an approximate ensemble of $\{p_x, \tilde{\rho}_x\}_{x=1,2}$ with

$$\tilde{\rho}_1 := \frac{1}{3}\omega_a \oplus \frac{1}{3}\tilde{\sigma}_1 \oplus \frac{1}{3}\omega_b, \tag{59}$$

$$\tilde{\rho}_2 := \frac{1}{6}\omega_a \oplus \frac{1}{3}\tilde{\sigma}_2 \oplus \frac{1}{2}\omega_b, \tag{60}$$

where

$$\tilde{\sigma}_1 := \frac{1}{2N}\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}, \tag{61}$$

$$\tilde{\sigma}_2 := \frac{1}{4N} \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & 3 & & \\ & & & & \ddots & \\ & & & & & 3 \end{pmatrix}. \tag{62}$$

Then $\tilde{\rho}_1$ and $\tilde{\rho}_2$ can be written as

$$\tilde{\rho}_1 = \tfrac{1}{2}\tilde{\omega}_a \oplus \tfrac{1}{2}\tilde{\omega}_b, \tag{63}$$

$$\tilde{\rho}_2 = \tfrac{1}{4}\tilde{\omega}_a \oplus \tfrac{3}{4}\tilde{\omega}_b, \tag{64}$$

where

$$\tilde{\omega}_a := \frac{2}{3}\left(\omega_a \oplus \frac{1}{2N}I\right), \tag{65}$$

$$\tilde{\omega}_b := \frac{2}{3}\left(\frac{1}{2N}I \oplus \omega_b\right), \tag{66}$$

with the N-dimensional identity operator I. Note that the approximation vastly changes the structure of the KI decomposition. The original KI decomposition of $\{p_x, \rho_x\}_{x=1,2}$ consists of three classical parts. However, after the approximation, the KI decomposition of $\{p_x, \tilde{\rho}_x\}_{x=1,2}$ consists of two large blocks, while the approximation changes the states only by ϵ. Then, considering the KI operations $\tilde{\mathcal{K}}_{\text{off}}$ and $\tilde{\mathcal{K}}_{\text{on}}$ corresponding to this structure, we have

$$\tilde{\mathcal{K}}_{\text{off}}(\rho_1) = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{67}$$

$$\tilde{\mathcal{K}}_{\text{off}}(\rho_2) = \frac{1}{12}\begin{pmatrix} 3+2\epsilon & 0 \\ 0 & 9-2\epsilon \end{pmatrix} \tag{68}$$

and

$$\tilde{\mathcal{K}}_{\text{on}} \circ \tilde{\mathcal{K}}_{\text{off}}(\rho_1) = \frac{1}{2}\tilde{\omega}_a \oplus \frac{1}{2}\tilde{\omega}_b = \tilde{\rho}_1, \tag{69}$$

$$\tilde{\mathcal{K}}_{\text{on}} \circ \tilde{\mathcal{K}}_{\text{off}}(\rho_2) = \left(\frac{1}{4}+\frac{\epsilon}{6}\right)\tilde{\omega}_a \oplus \left(\frac{3}{4}-\frac{\epsilon}{6}\right)\tilde{\omega}_b. \tag{70}$$

Therefore, it holds that

$$\|\tilde{\mathcal{K}}_{\text{on}} \circ \tilde{\mathcal{K}}_{\text{off}}(\rho_1) - \rho_1\|_1 \leqslant \frac{\epsilon}{2N}\sum_{j=0}^{2N-1}\left|\cos\left(\frac{2\pi j}{2N}+\alpha\right)\right| \leqslant \epsilon,$$

$$\|\tilde{\mathcal{K}}_{\text{on}} \circ \tilde{\mathcal{K}}_{\text{off}}(\rho_2) - \rho_2\|_1 = \frac{4\epsilon}{9} \leqslant \epsilon,$$

implying that this pair of KI operations yields a compression protocol within a finite error ϵ. Letting R_0 denote the optimal rate for $\{p_x, \rho_x\}_x$ without any approximations and R denote the rate for $\{p_x, \rho_x\}_x$ obtained by $\tilde{\mathcal{K}}_{\text{off}}$ and $\tilde{\mathcal{K}}_{\text{on}}$, we have

$$R_0 = H\left[\tfrac{1}{2}\mathcal{K}_{\text{off}}(\rho_1) + \tfrac{1}{2}\mathcal{K}_{\text{off}}(\rho_2)\right] \gtrsim \log_2 N, \tag{71}$$

$$R = H\left[\tfrac{1}{2}\tilde{\mathcal{K}}_{\text{off}}(\rho_1) + \tfrac{1}{2}\tilde{\mathcal{K}}_{\text{off}}(\rho_2)\right] \leqslant \log_2 2 = 1, \tag{72}$$

which shows the reduction of the compression rate to a constant rate, which is independent of the size of the system.

Thus, our compression protocol achieves a better compression rate by artificially generating approximate redundant parts. In the first example, we can generate a redundant part ω defined in Eq. (51); in the second example, we can generate large redundant parts $\tilde{\omega}_a$ and $\tilde{\omega}_b$ shown in Eqs. (65) and (66).

Our protocol has a much better rate when a quantum ensemble has an approximate ensemble with large redundant parts. By not sending these approximate redundant parts, we can achieve significantly small compression rates. Remarkably, in Example 2 we can see that even a small approximation leads to a compression rate independent of the dimension of the system.

Despite these examples showing large reductions in rates of blind compression, we note that we do not necessarily find a good approximation. For example, if an allowed error is much smaller than ϵ in Examples 1 and 2, we cannot apply the same approximation anymore. To further advance the study of blind quantum compression with finite approximations, a general investigation of conditions under which we can successfully find a good approximate ensemble of a given ensemble is needed. It would also be interesting to study a more general compression protocol for blind quantum compression with finite approximations.

Remark 1. Let us illustrate the significance of the local error criterion in our examples. In this remark, we consider the fidelity $F(\rho, \sigma) := \|\sqrt{\rho}\sqrt{\sigma}\|_1$ as a measure of accuracy instead of the trace norm to make the analysis easier. Note that these two measures are equivalent in the sense of the Fuchs–van de Graaf inequality [30]

$$1 - \tfrac{1}{2}\|\rho - \sigma\|_1 \leqslant F(\rho, \sigma) \leqslant \sqrt{1 - \tfrac{1}{4}\|\rho - \sigma\|_1^2}; \tag{73}$$

in particular, $\|\rho - \sigma\|_1 = 2$ if and only if $F(\rho, \sigma) = 0$, and $\|\rho - \sigma\|_1 = 0$ if and only if $F(\rho, \sigma) = 1$.

In both of our examples, the resulting states $\tilde{\mathcal{K}}_{\text{on}} \circ \tilde{\mathcal{K}}_{\text{off}}(\rho_1)$ and $\tilde{\mathcal{K}}_{\text{on}} \circ \tilde{\mathcal{K}}_{\text{off}}(\rho_2)$ satisfy

$$F(\rho_i, \tilde{\mathcal{K}}_{\text{on}} \circ \tilde{\mathcal{K}}_{\text{off}}(\rho_i)) \leqslant \sqrt{1 - O(\epsilon^2)}. \tag{74}$$

Since the fidelity is multiplicative to product-state inputs [30], for any $i_1, \ldots, i_n \in \{1, 2\}$,

$$F\big(\rho_{i_1} \otimes \cdots \otimes \rho_{i_n}, \tilde{\mathcal{K}}_{\text{on}} \circ \tilde{\mathcal{K}}_{\text{off}}(\rho_{i_1}) \otimes \cdots \otimes \tilde{\mathcal{K}}_{\text{on}} \circ \tilde{\mathcal{K}}_{\text{off}}(\rho_{i_n})\big)$$

$$= \prod_{k=1}^{n} F\big(\rho_{i_k}, \tilde{\mathcal{K}}_{\text{on}} \circ \tilde{\mathcal{K}}_{\text{off}}(\rho_{i_k})\big)$$

$$\leqslant \prod_{k=1}^{n} \sqrt{1 - O(\epsilon^2)} = [\sqrt{1 - O(\epsilon^2)}]^n \xrightarrow{n \to \infty} 0.$$

Thus, our protocol is asymptotically inaccurate in terms of the global error criterion though the local error is small even when $n \to \infty$. We believe that the rate reduction we achieved is realized by sacrificing the global accuracy. This intuition is closely related to the problem of a strong converse [31]. The strong converse of the blind data compression of mixed states is still largely unexplored, and we do not discuss this topic in this paper.

D. Approximation of classical ensembles

In this section we discuss compression of classical ensembles under finite approximations, aiming to obtain insight for a more general setting. Here, to see properties of blind compression of classical ensembles concisely, we consider classical ensembles consisting of two states. We propose two approximation methods for a two-state classical ensemble,

namely, the arithmetic mean method and the geometric mean method.

As seen in Definition 2, a classical ensemble is associated with a fixed basis in which all the states in the ensemble are diagonal. Indeed, when we approximate a classical ensemble with respect to the fixed basis, we only have to consider an approximation of diagonal elements. (See Appendix B for the details.) Therefore, when a basis is fixed, an approximation of a given classical state is equivalent to an approximation of a probability distribution corresponding to the classical state. To consider an approximation of a probability distribution, binning is a useful method, where we divide a given probability distribution into several groups (bins) and replace each probability with the average value of the bin the probability belongs to.

Definition 3 (binning). Let $\{p_a : a \in \Sigma\}$ be a probability distribution. Consider a partition of an alphabet Σ into m disjoint constituent alphabets Σ_k,

$$\Sigma = \bigcup_{k=1}^{m} \Sigma_k, \tag{75}$$

$$\Sigma_k \cap \Sigma_l = \emptyset \tag{76}$$

for all $k \neq l$. Then we call each Σ_k a bin and m the number of bins. Consider a probability distribution on Σ, $\{p_a : a \in \Sigma\}$. For each bin Σ_k, let $p^{(k)}$ be the average value of the set $\{p_a : a \in \Sigma_k\}$. Then we can construct a new probability distribution $\{p'_a : a \in \Sigma\}$ by replacing p_a with $p'_a = p^{(k)}$ when $a \in \Sigma_k$. This method for generating a probability distribution $\{p'_a : a \in \Sigma\}$ from a given distribution $\{p_a : a \in \Sigma\}$ is called binning. When probabilities within a bin $\{p_a : a \in \Sigma_k\}$ are all close to each other, binning gives an approximation of probability distributions, that is, it is also regarded as a good approximation method of classical states.

We propose two methods to approximate a given ensemble based on binning. Here we further specialize from a general two-state classical ensemble to one in which one of the two states is the flat state, motivated by the setup discussed in Ref. [18].

Suppose $\mathcal{H}$ is a quantum system and consider two classical states $\{\rho, \sigma\}$ where ρ is the flat state. Take an orthonormal basis $\{|i\rangle \in \mathcal{H} : 1 \leqslant i \leqslant d_{\mathcal{H}}\}$ of $\mathcal{H}$ so that we can write

$$\rho = \sum_{i=1}^{d_{\mathcal{H}}} \frac{1}{d_{\mathcal{H}}} |i\rangle\langle i|, \tag{77}$$

$$\sigma = \sum_{i=1}^{d_{\mathcal{H}}} p_i |i\rangle\langle i|, \tag{78}$$

where $\{p_i : 1 \leqslant i \leqslant d_{\mathcal{H}}\}$ forms a probability distribution with $p_1 \geqslant p_2 \geqslant \cdots \geqslant p_{d_{\mathcal{H}}}$. We consider an approximation of $\{p_i : 1 \leqslant i \leqslant d_{\mathcal{H}}\}$ within an allowed error $\epsilon > 0$, that is, a probability distribution $\{p'_i : 1 \leqslant i \leqslant d_{\mathcal{H}}\}$ such that

$$\sum_{i=1}^{d_{\mathcal{H}}} |p_i - p'_i| \leqslant \epsilon. \tag{79}$$

To construct an approximation leading to large redundant parts, we propose the following two methods, namely, the arithmetic mean method and the geometric mean method.

The arithmetic mean method proceeds as follows.

(1) First, for given $\epsilon > 0$, find the largest positive integer k_1 such that

$$|p_1 - p_{k_1}| \leqslant \frac{\epsilon}{d_{\mathcal{H}}}. \tag{80}$$

Define a set of positive integers $I_1 := \{1, \ldots, k_1\}$.

(2) For $i > 1$, find the largest positive integer k_i such that

$$|p_{k_{i-1}+1} - p_{k_i}| \leqslant \frac{\epsilon}{d_{\mathcal{H}}}. \tag{81}$$

Define a set of positive integers $I_i := \{k_{i-1} + 1, \ldots, k_i\}$.

(3) Repeat step 2 until we find a positive integer L such that $k_L = d_{\mathcal{H}}$.

(4) For $j \in I_i$ with some $1 \leqslant i \leqslant L$, replace p_j as

$$p_j \mapsto p_j^{(A)} := \frac{1}{|I_i|} \sum_{m \in I_i} p_m \tag{82}$$

and define

$$\sigma^{(A)} := \sum_{i=1}^{d_{\mathcal{H}}} p_i^{(A)} |i\rangle\langle i| \tag{83}$$

The geometric mean method proceeds as follows.

(1) First, for given $\epsilon > 0$, find the largest positive integer k_1 such that

$$\frac{p_{k_1}}{p_1} \geqslant \frac{1}{1 + \epsilon}. \tag{84}$$

Define a set of positive integers $I_1 := \{1, \ldots, k_1\}$.

(2) For $i > 1$, find the largest positive integer k_i such that

$$\frac{p_{k_i}}{p_{k_{i-1}+1}} \geqslant \frac{1}{1 + \epsilon}. \tag{85}$$

Define a set of positive integers $I_i := \{k_{i-1} + 1, \ldots, k_i\}$.

(3) Repeat step 2 until we find a positive integer L such that $k_L = d_{\mathcal{H}}$.

(4) For $j \in I_i$ with some $1 \leqslant i \leqslant l$, replace p_j as

$$p_j \mapsto p_j^{(G)} := \frac{1}{|I_i|} \sum_{m \in I_i} p_m \tag{86}$$

and define

$$\sigma^{(G)} := \sum_{i=1}^{d_{\mathcal{H}}} p_i^{(G)} |i\rangle\langle i|. \tag{87}$$

We show that these approximation methods are valid, that is, the resulting states $\sigma^{(A)}$ and $\sigma^{(G)}$ are close enough to the original state σ. The following proposition guarantees that $\{\rho, \sigma'\}$ with $\sigma' = \sigma^{(A)}$ or $\sigma^{(G)}$ is considered as an approximation of $\{\rho, \sigma\}$.

Proposition 1. For a positive real number $\epsilon > 0$, let $\sigma^{(A)}$ and $\sigma^{(G)}$ be the approximate states of a given diagonal state σ by the arithmetic mean method and the geometric mean method, respectively. Then it holds that

$$\|\sigma^{(A)} - \sigma\|_1 \leqslant \epsilon, \tag{88}$$

$$\|\sigma^{(G)} - \sigma\|_1 \leqslant \epsilon. \tag{89}$$

Proof. It holds that

$$\|\sigma - \sigma^{(A)}\|_1 = \sum_{i=1}^{d_{\mathcal{H}}} |p_i - p_i^{(A)}|$$

$$= \sum_{i=1}^{L} \sum_{j \in I_i} |p_j - p_j^{(A)}|$$

$$\leqslant \sum_{i=1}^{L} \sum_{j \in I_i} |p_{k_{i-1}+1} - p_{k_i}|$$

$$\leqslant \sum_{i=1}^{L} \sum_{j \in I_i} \frac{\epsilon}{d_{\mathcal{H}}}$$

$$= \epsilon. \tag{90}$$

The first inequality follows because the difference $|p_j - p_j^{(A)}|$ is upper bounded by the difference between the largest value and the smallest value of the set $\{p_i : i \in I_j\}$, which is given by $|p_{k_{i-1}+1} - p_{k_i}|$. The second inequality follows from Eq. (81). On the other hand, it holds that

$$\|\sigma - \sigma^{(G)}\|_1 = \sum_{i=1}^{d_{\mathcal{H}}} |p_i - p_i^{(G)}|$$

$$= \sum_{i=1}^{L} \sum_{j \in I_i} |p_j - p_j^{(G)}|$$

$$= \sum_{i=1}^{L} \sum_{j \in I_i} p_j^{(G)} \left| \frac{p_j}{p_j^{(G)}} - 1 \right|. \tag{91}$$

Here, by Eq. (85) we have

$$\frac{1}{1+\epsilon} \leqslant \frac{p_j}{p_j^{(G)}} \leqslant 1 + \epsilon, \tag{92}$$

which implies that

$$-\epsilon \leqslant \frac{-\epsilon}{1+\epsilon} \leqslant \frac{p_j}{p_j^{(G)}} - 1 \leqslant \epsilon. \tag{93}$$

Combining Eqs. (91) and (93), we have that

$$\|\sigma - \sigma^{(G)}\|_1 = \sum_{i=1}^{L} \sum_{j \in I_i} p_j^{(G)} \left| \frac{p_j}{p_j^{(G)}} - 1 \right|$$

$$\leqslant \sum_{i=1}^{L} \sum_{j \in I_i} p_j^{(G)} \epsilon$$

$$= \epsilon. \tag{94}$$

∎

Consider a quantum ensemble formed by $\{\rho, \sigma'\}$ ($\sigma' = \sigma^{(A)}$ or $\sigma^{(G)}$) and quantum channels $\mathcal{K}_{\text{off}}$ and $\mathcal{K}_{\text{on}}$ with respect to this ensemble. Since

$$\mathcal{K}_{\text{off}}(\rho) = \sum_{i=1}^{L} \frac{|I_i|}{d_{\mathcal{H}}} |i\rangle\langle i|, \tag{95}$$

$$\mathcal{K}_{\text{off}}(\sigma) = \sum_{i=1}^{L} \left(\sum_{m \in I_i} p_m \right) |i\rangle\langle i|, \tag{96}$$

the rate R of our protocol for this ensemble is

$$R = H[p_\rho \mathcal{K}_{\text{off}}(\rho) + p_\sigma \mathcal{K}_{\text{off}}(\sigma)] \leqslant \log_2 L. \tag{97}$$

In addition, by construction,

$$\mathcal{K}_{\text{on}} \circ \mathcal{K}_{\text{off}}(\rho) = \rho, \tag{98}$$

$$\mathcal{K}_{\text{on}} \circ \mathcal{K}_{\text{off}}(\sigma) = \sigma'. \tag{99}$$

Thus, this approximation satisfies the condition (39), namely,

$$\|\mathcal{K}_{\text{on}} \circ \mathcal{K}_{\text{off}}(\rho) - \rho\|_1 = \|\rho - \rho\|_1 = 0 \leqslant \epsilon, \tag{100}$$

$$\|\mathcal{K}_{\text{on}} \circ \mathcal{K}_{\text{off}}(\sigma) - \sigma\|_1 = \|\sigma' - \sigma\|_1 \leqslant \epsilon. \tag{101}$$

While analytical discussion of these two methods can be complicated, in the following section we conduct numerical experiments to investigate the performance of these approximation methods.

E. Numerical simulation for approximating a classical ensemble

In this section we perform numerical simulations to verify and compare the performance of the two methods, the arithmetic mean method and the geometric mean method. In Ref. [18] the authors proposed a protocol for approximately compressing a two-state classical ensemble, which employs the geometric mean. One may wonder if the arithmetic mean yields better results.

We conduct two types of simulations. In the first one, we fix the dimension of the quantum system and vary the allowed approximation ϵ. Then we observe how the compression rates obtained by these approximation methods depend on ϵ. In the second one, we change the dimension of quantum system $d_{\mathcal{H}}$ and we set $\epsilon = 1/\sqrt{d_{\mathcal{H}}}$ for each $d_{\mathcal{H}}$. Then we see the dependence of the compression rates on the dimension of a given system.

1. Numerical simulation for approximation-rate trade-off

Here we evaluate the performance of the two approximation methods with various values of allowed approximation ϵ. In the simulation, we fix the space to $\mathcal{H} = \mathbb{C}^{1024}$, that is, $d_{\mathcal{H}} = 1024 = 2^{10}$. We generate 1000 random diagonal states σ on this space. In more detail, we randomly generate a real number in the range $[0, 1]$ according to the uniform distribution 1024 times to obtain a vector on $\mathbb{R}^{1024}$. Then we normalize the obtained vector to obtain a probability distribution, and we regard this probability distribution as a classical state. For each state, we create the approximate states $\sigma^{(A)}$ and $\sigma^{(G)}$ by using the arithmetic mean method and the geometric mean method.

We compare the rates of the two methods. Figure 3(a) shows that the arithmetic mean method performs better than the geometric mean method on average. In particular, the results show the tendency that the difference between the two methods becomes large as the allowed error increases.

Furthermore, we estimate the curves representing average values of the results in the figure. Observing the graph shown in Fig. 3(a), we adopt the function

$$f(x) = \log_2 d_{\mathcal{H}} - a(1 - e^{-b\sqrt{x}}) \tag{102}$$
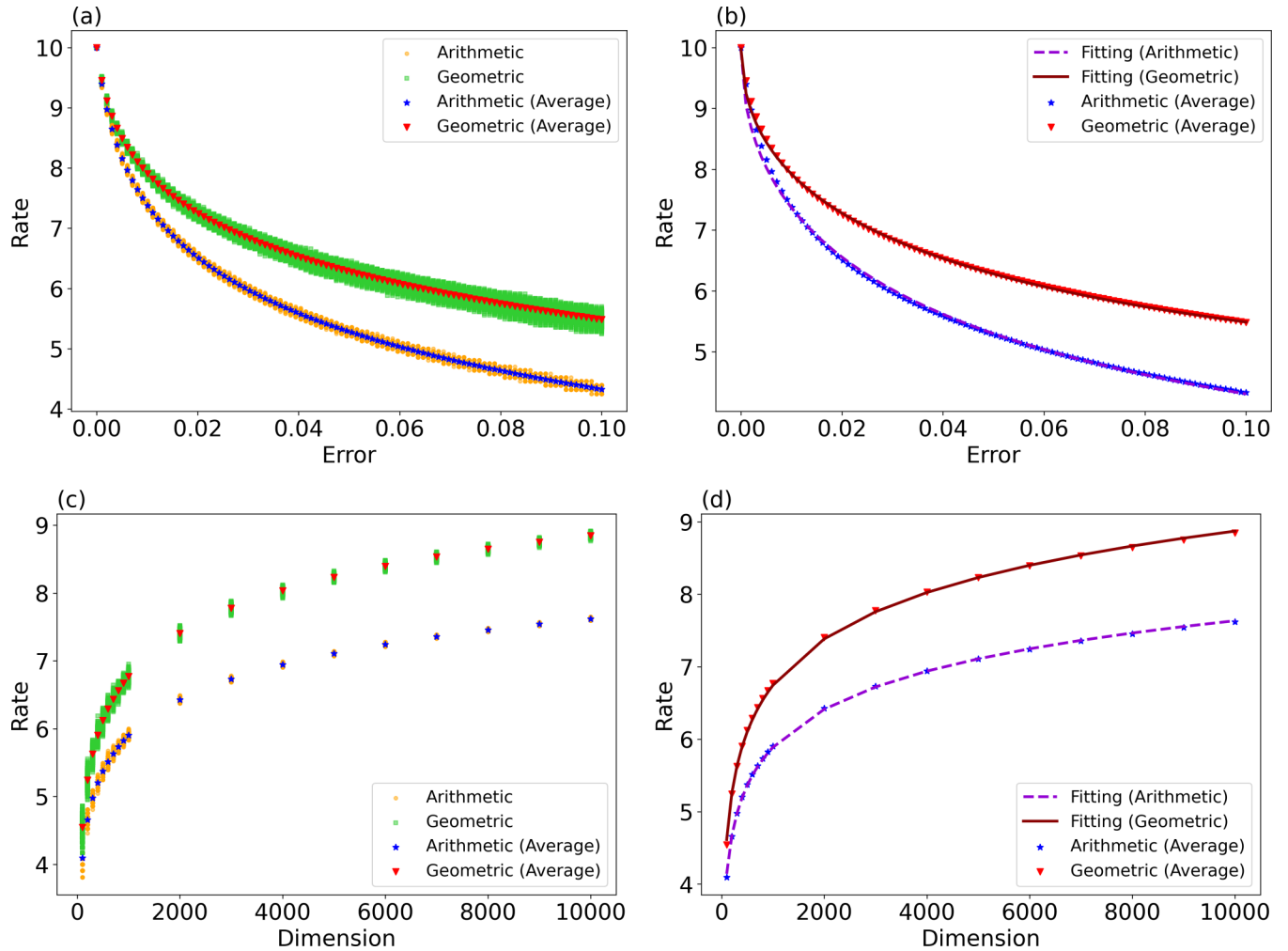
FIG. 3. Results of numerical experiments. (a) Graph of the compression rate as a function of allowed error ϵ shown in Eq. (79). (b) Graph of the fitting curves of the average results of (a). (c) Graph of the compression rate (97) as a function of the dimension D of the system under allowed approximation $1/\sqrt{D}$. (d) Graph of the fitting curves of the average results of (c). In (a) and (c) the yellow circles and green squares represent the compression rates of randomly generated probability distributions obtained by the arithmetic and geometric mean methods, respectively. The blue stars and red triangles represent the averages of the yellow circles and green squares, respectively, that is, the blue stars show the averaged results of the arithmetic mean methods and the red triangles show those of the geometric mean methods. In (b) and (d) the violet dashed lines are the fitting of the blue stars; the brown solid lines are the fitting of the red triangles. In (b) we employ the model function defined in Eq. (102) and we obtain the fitting parameters $a = 7.859(6)$ and $b = 4.077(5)$ for the violet dashed line (the arithmetic mean method) and $a = 6.262(2)$ and $b = 4.030(2)$ for the brown solid line (the geometric mean method). In (d) we use the function defined in Eq. (103) and we obtain the fitting parameters $a = 0.525\,892(4)$ and $b = 0.6477(4)$ for the violet dashed line (the arithmetic mean method) and $a = 0.640\,847(9)$ and $b = 0.358(1)$ for the brown solid line (the geometric mean method). The term "rate" on the y-axes refers to the upper bound shown in Eq. (97). Note that the quantities shown on the graphs are dimensionless.

with parameters a and b as a fitting function. Since the compression rate is strictly upper bounded by $\log_2 d_{\mathcal{H}}$ when we do not allow any approximation, the second term $a(1 - e^{-b\sqrt{x}})$ of Eq. (102) is considered to represent the degree of reduction caused by an approximation. In addition, this function explains the tendency that the rate approaches $\log_2 d_{\mathcal{H}}$ as the error becomes small and it also expresses the sudden decrease around the error $1/d_{\mathcal{H}}$. The results shown in Fig. 3(b) suggest that the fitting function defined in Eq. (102) is the correct function characterizing the compression rate as a function of the error while we have not theoretically and analytically demonstrated it. However, we do not believe that this function explains the full dependence of the rate on the error, because

of the sensitive behavior of the rate against errors when errors are roughly larger than $1/\sqrt{d_{\mathcal{H}}}$. For more general and deep understandings of the dependence, we need further investigations of the compression protocols.

2. Numerical simulation for dimension-rate relation

We investigate the dependence of the performance of the two approximation methods on the dimension of the space. Here we adopt the error $\epsilon = 1/\sqrt{d_{\mathcal{H}}}$, considering the discussion in Ref. [18] stating that this size of error $f(\Lambda_n) \approx 1/\sqrt{d_{\mathcal{H}}}$ [see Eq. (33)] leads to $g(\Lambda_n) \approx \log_2 d$ [see Eq. (34)]. (See Appendix A of Ref. [18] for more details.) Thus, $\epsilon = 1/\sqrt{d_{\mathcal{H}}}$
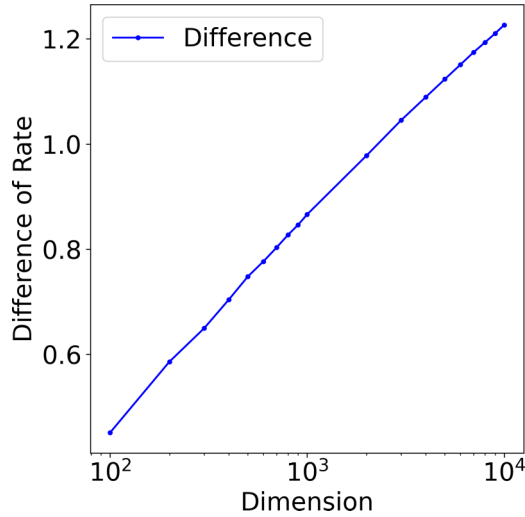
FIG. 4. Graph of the compression rate by the arithmetic mean method minus the compression rate by the geometric mean method as a function of the dimension of system. Here "rate" refers to the upper bound shown in Eq. (97). The horizontal axis is expressed in logarithmic scale. Note that the quantities shown on the graphs are dimensionless.

may well cause a large reduction of the rate. We generate 1000 random diagonal states σ on this space in the same way as in the first simulation. For each state, we create the approximate states $\sigma^{(A)}$ and $\sigma^{(G)}$ by using the arithmetic mean method and the geometric mean method. Then we compare the rates of the two methods. Figure 3(c) shows that the arithmetic mean method also performs better than the geometric mean method on average in this case.

To investigate the dependence of the difference between the two methods on the dimension, we also plotted the difference in Fig. 4. As seen in the graph, the difference becomes large as the dimension gets large and it linearly depends on the logarithm of dimension.

Moreover, we estimate the curves representing average values of the results in the graph. Observing the graphs shown in Figs. 3(c) and 4, adopt the function

$$f(x) = a \log_2 x + b \tag{103}$$

with parameters a and b as a fitting function. The results shown in Fig. 3(d) imply that the fitting function defined in Eq. (103) is appropriate for characterizing the compression rate as a function of the dimension. Since the compression rate is strictly upper bounded by $\log_2 d_{\mathcal{H}}$ when we do not allow any approximation, the coefficient a represents the degree of reduction caused by an approximation.

IV. DISCUSSION AND CONCLUSION

In this paper we have investigated blind compression of quantum ensembles under finite local approximations. In previous research, the optimal rate of blind compression was obtained through the KI decomposition by removing the redundant parts of a given ensemble. However, since KI decomposition is highly sensitive even to small approximations, approximation or error analysis of blind compression has

not been explored much. In this paper we focused on the instability of the KI decomposition against approximations. Taking advantage of the sensitivity of KI decomposition, we constructed a compression protocol that can have a substantial reduction of the compression rate if a finite approximation is allowed. In our protocol, the sender intentionally introduces distortion to a given ensemble by applying KI operations that almost preserve the ensemble. Note that the distortion implemented by the sender plays a major role in a full quantum rate-distortion theory for mixed states [26]. We explicitly showed a reduction of the compression rate by several examples. In particular, in Example 2 we saw a near-maximal reduction of the compression rate allowed by the dimension. Remarkably, the rate reduction of our protocol persists even when approximation ϵ is approximately equal to $1/d_{\mathcal{H}}$, in contrast to a result of Ref. [18] for two classical distributions, where a large rate reduction is seen for $\epsilon \approx 1/\sqrt{d_{\mathcal{H}}}$. We also share the following insight. For any quantum ensemble Φ_1, for any neighborhood of Φ_1 there is an ensemble Φ_2 in this neighborhood with trivial KI decomposition. If the initial ensemble Φ_1 has redundant parts, the two ensembles Φ_1 and Φ_2 will have different compression rates. Our examples can be understood as a manifestation of this generic situation. We want to point out that our simple explicit example with only two states provides a nearly extremal quantitative demonstration. Moreover, we investigated blind compression of classical ensembles to analyze general properties of our compression protocol. We proposed two compression methods for two-state classical ensembles including the flat state, namely, the arithmetic mean method and the geometric mean method. We numerically investigated these compression methods of two-state classical ensembles and discovered that the arithmetic mean method performs better than the geometric mean method.

As a future direction, it would be interesting to analyze our protocol more generally. As noted in the previous sections, our protocol does not necessarily show a large reduction of the compression rate for an arbitrary ensemble. For example, if the allowed approximation is sufficiently small, our protocol might have a rate equal to the optimal compression rate with a vanishing error, obtained in Ref. [12]. The difficulty of the analysis lies in the absence of an approximation of the KI decomposition. While finite approximations can significantly lower the compression rate, as we showed, because of the instability of the KI decomposition, the instability also makes the general analysis intractable. Investigation of compression protocols on general two-state quantum ensembles might be a good first step in this direction. For a two-state ensemble, we may consider that one of the two states is diagonal by choosing an appropriate basis. Fixing the basis would be helpful to study a new approximate compression protocol and the dependence of the compression protocol on allowed errors.

Thus, we have shed light on blind quantum data compression with the allowance of approximations by revealing that the approximation-sensitive nature of KI decomposition can be used to achieve a substantial rate reduction. We believe that our findings will fuel further investigation of blind compression with finite approximations, leading to a more general understanding of the approximation-rate trade-off in this setup.

APPENDIX A: PROOF OF THEOREM 2

We can give the descriptions of $\mathcal{K}_{\text{off}}$ and $\mathcal{K}_{\text{on}}$ explicitly by using Kraus representations. First, $\mathcal{K}_{\text{off}}$ is given by Kraus operators

$$A_{j_l}^{(l)} := I_{\mathcal{H}_Q^{(l)}} \otimes \langle j_l|, \tag{A1}$$

where $\{|j_l\rangle : j_l\}$ is an orthonormal basis of $\mathcal{H}_Q^{(l)}$ for all $l \in \Xi$. Kraus operators $A_{j_l}^{(l)}$ apply to the lth block and $\{A_{j_l}^{(l)} : j_l\}$ forms the partial trace over $\mathcal{H}_R^{(l)}$. Indeed,

$$\sum_{l \in \Xi} \sum_{j_l} A_{j_l}^{(l)} \rho_a \big(A_{j_l}^{(l)}\big)^\dagger = \bigoplus_{l \in \Xi} q^{(a,l)} \rho_Q^{(a,l)} \otimes \left(\sum_{j_l} \langle j_l|\rho_R^{(l)}|j_l\rangle \right)$$

$$= \bigoplus_{l \in \Xi} q^{(a,l)} \rho_Q^{(a,l)} \otimes \text{Tr}\big(\rho_R^{(l)}\big)$$

$$= \bigoplus_{l \in \Xi} q^{(a,l)} \rho_Q^{(a,l)} \tag{A2}$$

for all $a \in \Sigma$. In addition, $\{A_{j_l}^{(l)} : l \in \Xi, j_l\}$ forms a valid Kraus representation because

$$\sum_{l \in \Xi} \sum_{j_l} \big(A_{j_l}^{(l)}\big)^\dagger A_{j_l}^{(l)} = \sum_{l \in \Xi} \sum_{j_l} I_{\mathcal{H}_Q^{(l)}} \otimes |j_l\rangle\langle j_l|$$

$$= \sum_{l \in \Xi} I_{\mathcal{H}_Q^{(l)}} \otimes \left(\sum_{j_l} |j_l\rangle\langle j_l| \right)$$

$$= \sum_{l \in \Xi} I_{\mathcal{H}_Q^{(l)}} \otimes I_{\mathcal{H}_R^{(l)}}$$

$$= I_{\mathcal{H}}. \tag{A3}$$

Therefore, we can construct $\mathcal{K}_{\text{off}}$ by a Kraus representation

$$\big\{A_{j_l}^{(l)} : l \in \Xi, j_l\big\}. \tag{A4}$$

Next $\mathcal{K}_{\text{on}}$ is given by Kraus operators

$$A_{k_l}^{(l)} := I_{\mathcal{H}_Q^{(l)}} \otimes \sqrt{r_{k_l}} |k_l\rangle, \tag{A5}$$

where $\{|k_l\rangle : k_l\}$ is an orthonormal basis of $\mathcal{H}_Q^{(l)}$ for all $l \in \Xi$ corresponding to a spectral decomposition

$$\rho_R^{(l)} := \sum_{k_l} r_{k_l}|k_l\rangle\langle k_l| \tag{A6}$$

with eigenvalues $r_{k_l} \geqslant 0$. Kraus operators $A_{k_l}^{(l)}$ apply to the lth block and $\{A_{k_l}^{(l)} : k_l\}$ forms the construction of $\rho_R^{(l)}$ on system $\mathcal{H}_R^{(l)}$. Indeed, we have that

$$\sum_{l \in \Xi} \sum_{k_l} A_{k_l}^{(l)} \left(\bigoplus_{l \in \Xi} q^{(a,l)} \rho_Q^{(a,l)} \right) \big(A_{k_l}^{(l)}\big)^\dagger$$

$$= \bigoplus_{l \in \Xi} q^{(a,l)} \rho_Q^{(a,l)} \otimes \left(\sum_{k_l} r_{k_l}|k_l\rangle\langle k_l| \right)$$

$$= \bigoplus_{l \in \Xi} q^{(a,l)} \rho_Q^{(a,l)} \otimes \rho_R^{(l)}$$

$$= \rho_a \tag{A7}$$

for all $a \in \Sigma$. In addition, $\{A_{k_l}^{(l)} : l \in \Xi, k_l\}$ forms a valid Kraus representation because it holds that

$$\sum_{l \in \Xi} \sum_{k_l} \big(A_{k_l}^{(l)}\big)^\dagger A_{k_l}^{(l)} = \sum_{l \in \Xi} \sum_{k_l} I_{\mathcal{H}_Q^{(l)}} \otimes r_{k_l} \langle k_l|k_l\rangle$$

$$= \sum_{l \in \Xi} I_{\mathcal{H}_Q^{(l)}} \otimes \left(\sum_{k_l} r_{k_l} \right)$$

$$= \sum_{l \in \Xi} I_{\mathcal{H}_Q^{(l)}} \otimes 1$$

$$= I_{\mathcal{H}_Q^{(l)}}. \tag{A8}$$

Therefore, we can construct $\mathcal{K}_{\text{on}}$ by a Kraus representation $\{A_{k_l}^{(l)} : l \in \Xi, k_l\}$.

APPENDIX B: APPROXIMATION OF A QUANTUM ENSEMBLE IN A FIXED BASIS

A classical ensemble is defined with a fixed basis of a given quantum system in which every state in the ensemble is diagonalized. (See Definition 2.) Here we analyze an approximation of a given classical ensemble in the fixed basis, that is, we consider an approximate ensemble whose structure of the KI decomposition is given in the same fixed basis. When we approximate a classical ensemble with respect to the fixed basis, we only have to consider an approximation of diagonal elements in the sense of the following proposition.

Proposition 2. Let $\mathcal{H}$ be a quantum system. Let $\rho, \sigma \in D(\mathcal{H})$ be quantum states on the system. Suppose that ρ and σ form a classical ensemble, that is, we can write

$$\rho = \sum_{i=1}^{d_{\mathcal{H}}} \rho_i|i\rangle\langle i|, \tag{B1}$$

$$\sigma = \sum_{i=1}^{d_{\mathcal{H}}} \sigma_i|i\rangle\langle i| \tag{B2}$$

for an orthonormal basis $\{|i\rangle \in \mathcal{H} : 1 \leqslant i \leqslant d_{\mathcal{H}}\}$ of $\mathcal{H}$ and probability distributions $\{\rho_i : 1 \leqslant i \leqslant d_{\mathcal{H}}\}$ and $\{\sigma_i : 1 \leqslant i \leqslant d_{\mathcal{H}}\}$. Suppose that ρ and σ can be approximated in the same basis as

$$\left\| \rho - \bigoplus_{l \in \Xi} q^{(\rho,l)} \omega_Q^{(\rho,l)} \otimes \omega_R^{(l)} \right\|_1 \leqslant \epsilon, \tag{B3}$$

$$\left\| \sigma - \bigoplus_{l \in \Xi} q^{(\sigma,l)} \omega_Q^{(\sigma,l)} \otimes \omega_R^{(l)} \right\|_1 \leqslant \epsilon, \tag{B4}$$

where the structure of the decomposition is also given in the basis $\{|i\rangle \in \mathcal{H} : 1 \leqslant i \leqslant d_{\mathcal{H}}\}$. Then there exists an approximation such that all $\omega_Q^{(\rho,l)}$, $\omega_Q^{(\sigma,l)}$, and $\omega_R^{(l)}$ are diagonal.

Proof. Let us take some $l \in \Xi$ and consider $q^{(\rho,l)}\omega_Q^{(\rho,l)} \otimes \omega_R^{(l)}$ and $q^{(\sigma,l)}\omega_Q^{(\sigma,l)} \otimes \omega_R^{(l)}$. Let $\rho^{(l)}$ and $\sigma^{(l)}$ be the corresponding block of ρ and σ, respectively. Let $\Delta \in C(\mathcal{H}, \mathcal{H})$ be the completely dephasing channel with respect to the basis $\{|i\rangle \in \mathcal{H} : 1 \leqslant i \leqslant d_{\mathcal{H}}\}$ defined as

$$\Delta(\rho) = \sum_{i=1}^{d_{\mathcal{H}}} \langle i|\rho|i\rangle \, |i\rangle\langle i| \tag{B5}$$

for $\rho \in D(\mathcal{H})$. Then we have

$$\left\| \rho^{(l)} - \bigoplus_{l \in \Xi} q^{(\rho,l)}\Delta(\omega_Q^{(\rho,l)}) \otimes \Delta(\omega_R^{(l)}) \right\|_1$$

$$= \left\| \Delta(\rho^{(l)}) - \Delta\left(\bigoplus_{l \in \Xi} q^{(\rho,l)}\omega_Q^{(\rho,l)} \otimes \omega_R^{(l)} \right) \right\|_1$$

$$\leqslant \left\| \rho^{(l)} - \bigoplus_{l \in \Xi} q^{(\rho,l)}\omega_Q^{(\rho,l)} \otimes \omega_R^{(l)} \right\|_1. \tag{B6}$$

Similarly,

$$\left\| \sigma^{(l)} - \bigoplus_{l \in \Xi} q^{(\sigma,l)}\Delta(\omega_Q^{(\sigma,l)}) \otimes \Delta(\omega_R^{(l)}) \right\|_1$$

$$\leqslant \left\| \sigma^{(l)} - \bigoplus_{l \in \Xi} q^{(\sigma,l)}\omega_Q^{(\sigma,l)} \otimes \omega_R^{(l)} \right\|_1. \tag{B7}$$

Therefore, diagonal states $\bigoplus_{l \in \Xi} q^{(\rho,l)}\Delta(\omega_Q^{(\rho,l)}) \otimes \Delta(\omega_R^{(l)})$ and $\bigoplus_{l \in \Xi} q^{(\sigma,l)}\Delta(\omega_Q^{(\sigma,l)}) \otimes \Delta(\omega_R^{(l)})$ are also approximations of ρ and σ. ∎

APPENDIX C: PROOF THAT σ_1 AND σ_2 IN EXAMPLE 2 HAVE NO REDUNDANT PARTS

In this Appendix we prove that the two $2N$-dimensional states

$$\sigma_1 := \frac{1}{4N} \begin{pmatrix} 2 & \epsilon e^{i\alpha} & 0 & \cdots & \epsilon e^{-i\alpha} \\ \epsilon e^{-i\alpha} & 2 & \epsilon e^{i\alpha} & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \epsilon e^{-i\alpha} & 2 & \epsilon e^{i\alpha} \\ \epsilon e^{i\alpha} & 0 & \cdots & \epsilon e^{-i\alpha} & 2 \end{pmatrix},$$

$$\sigma_2 := \frac{1}{4N} \begin{pmatrix} 1+2\epsilon & & & & \\ & \ddots & & & \\ & & 1+2\epsilon & & \\ & & & 3-2\epsilon & \\ & & & & \ddots & \\ & & & & & 3-2\epsilon \end{pmatrix},$$

with $0 < \epsilon < \frac{1}{2}$ and $0 < \alpha < 1/4N$, introduced in Example 2, do not have redundant parts.

First, we prove that σ_1 and σ_2 do not have a nontrivial (multiblock) block-diagonal structure. Define normalized

vectors

$$|v_k\rangle = \frac{1}{\sqrt{2N}} \begin{pmatrix} 1 \\ \omega^k \\ \omega^{2k} \\ \omega^{3k} \\ \vdots \\ \omega^{(2N-1)k} \end{pmatrix}, \tag{C1}$$

with

$$\omega = e^{i2\pi/2N}, \tag{C2}$$

for $k = 0, 1, \ldots, 2N - 1$. Note that $\omega^{2N} = 1$. Then

$$\sigma_1 |v_k\rangle = \frac{1 + \epsilon\left[\cos\left(\frac{2\pi k}{2N} + \alpha \right) \right]}{2N} |v_k\rangle \tag{C3}$$

for $k = 0, 1, \ldots, 2N - 1$, that is, $|v_k\rangle$ is an eigenvector of σ_1 with eigenvalue $[\cos(\frac{2\pi k}{2N} + \alpha)]/2N$. Since $0 < \alpha < 1/4N$, there do not exist distinct $k, l \in \{0, 1, 2, \ldots, 2N - 1\}$ such that

$$\cos\left(\frac{2\pi k}{2N} + \alpha \right) = \cos\left(\frac{2\pi l}{2N} + \alpha \right), \tag{C4}$$

that is, σ_1 is not degenerate. On the other hand, eigenvectors of σ_2 can be expressed as

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ a_1 \\ a_2 \\ \vdots \\ a_N \end{pmatrix}, \tag{C5}$$

where $(a_1, a_2, \ldots, a_N) \neq (0, 0, \ldots, 0)$. Without loss of generality, we focus on normalized vectors $(a_1, a_2, \ldots, a_N)$. Indeed,

$$\sigma_2 \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \frac{1 + 2\epsilon}{4N} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{C6}$$

and

$$\sigma_2 \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ a_1 \\ a_2 \\ \vdots \\ a_N \end{pmatrix} = \frac{3 - 2\epsilon}{4N} \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ a_1 \\ a_2 \\ \vdots \\ a_N \end{pmatrix}. \tag{C7}$$

Note that $1 + 2\epsilon < 3 - 2\epsilon$ since $0 < \epsilon < \frac{1}{2}$.

Suppose that σ_1 and σ_2 have a shared nontrivial block-diagonal structure. In this case, there exist a unitary matrix U, nontrivial subspaces $\tilde{\mathcal{H}}$ and $\mathcal{H}'$ with $d_{\tilde{\mathcal{H}}} + d'_{\mathcal{H}} = 2N$, and positive-semidefinite matrices P_1 and P_2 on $\tilde{\mathcal{H}}$ and Q_1 and Q_2 on $\mathcal{H}'$ such that

$$U\sigma_1 U^\dagger = P_1 \oplus Q_1, \tag{C8}$$

$$U\sigma_2 U^\dagger = P_2 \oplus Q_2. \tag{C9}$$

Since σ_1 is not degenerate, from Eq. (C8), vector space $\tilde{\mathcal{H}}$ is spanned by some $d_{\tilde{\mathcal{H}}}$ vectors chosen from $\{U|v_k\rangle\}_{k=1}^{2N}$. Let $\{U|v_{k_j}\rangle\}_{j=1}^{d_{\tilde{\mathcal{H}}}}$ denote such $d_{\tilde{\mathcal{H}}}$ vectors. Similarly, from Eq. (C9), vector space $\tilde{\mathcal{H}}$ is also spanned by some $d_{\tilde{\mathcal{H}}}$ vectors $\{U|u_l\rangle\}_{l=1}^{d_{\tilde{\mathcal{H}}}}$, where $\{|u_l\rangle\}_{l=1}^{2N}$ are linearly independent eigenvectors of σ_2. In particular, since $U|u_1\rangle \in \tilde{\mathcal{H}}$, we may write

$$U|u_1\rangle = \sum_{j=1}^{d_{\tilde{\mathcal{H}}}} c_{k_j} U|v_{k_j}\rangle \tag{C10}$$

with some $c_{k_j} \in \mathbb{C}$. By applying $U^\dagger$ from the left,

$$|u_1\rangle = \sum_{j=1}^{d_{\tilde{\mathcal{H}}}} c_{k_j} |v_{k_j}\rangle. \tag{C11}$$

Recall that

$$|v_{k_j}\rangle = \frac{1}{\sqrt{2N}} \begin{pmatrix} 1 \\ \omega^{k_j} \\ \omega^{2k_j} \\ \omega^{3k_j} \\ \vdots \\ \omega^{(2N-1)k_j} \end{pmatrix}. \tag{C12}$$

Let $\vec{s}_{k_j}$ denote the vector consisting of the top N entries of $|v_{k_j}\rangle$ and $\vec{t}_{k_j}$ denote the vector consisting of the bottom N entries of $|v_{k_j}\rangle$, that is,

$$\vec{s}_{k_j} := \frac{1}{\sqrt{2N}} \begin{pmatrix} 1 \\ \omega^{k_j} \\ \omega^{2k_j} \\ \vdots \\ \omega^{(N-1)k_j} \end{pmatrix}, \tag{C13}$$

$$\vec{t}_{k_j} := \frac{1}{\sqrt{2N}} \begin{pmatrix} \omega^{Nk_j} \\ \omega^{(N+1)2k_j} \\ \omega^{(N+2)2k_j} \\ \vdots \\ \omega^{(2N-1)k_j} \end{pmatrix}, \tag{C14}$$

and $|v_{j_k}\rangle = \vec{s}_{j_k} \oplus \vec{t}_{j_k}$.

When $|u_1\rangle$ is given as

$$|u_1\rangle = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ a_1 \\ a_2 \\ \vdots \\ a_N \end{pmatrix}, \tag{C15}$$

to have Eq. (C10), it is necessary that

$$\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \sum_{j=1}^{d_{\tilde{\mathcal{H}}}} c_{k_j} \vec{s}_{k_j}. \tag{C16}$$

Here we looked at the top N entries of each vector in Eq. (C10). We will prove that $d_{\tilde{\mathcal{H}}} > N$. We first show that for any N distinct vectors $\{|v_{k_l}\rangle\}_{l=1}^{2N}$ chosen from $\{|v_k\rangle\}_{k=1}^{2N}$, $\{\vec{s}_{k_l}\}_{l=1}^{N}$ are linearly independent. For this purpose, it suffices to show that the determinant of matrix

$$M := \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ \omega^{k_1} & \omega^{k_2} & \omega^{k_3} & \cdots & \omega^{k_N} \\ \omega^{2k_1} & \omega^{2k_2} & \omega^{2k_3} & \cdots & \omega^{2k_N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \omega^{(N-1)k_1} & \omega^{(N-1)k_2} & \omega^{(N-1)k_3} & \cdots & \omega^{(N-1)k_N} \end{pmatrix}$$

is not zero. Recalling Vandermonde's determinant, the determinant of M is given as

$$\det M = \prod_{1 \leqslant p < q \leqslant N} (\omega^{k_q} - \omega^{k_p}). \tag{C17}$$

Since $\omega^{k_p} \neq \omega^{k_q}$ for any $1 \leqslant p < q \leqslant N$, $\det M \neq 0$ and thus $\{\vec{s}_{k_l}\}_{l=1}^{N}$ are linearly independent. Now we show that $d_{\tilde{\mathcal{H}}} > N$. By way of contradiction, suppose that $d_{\tilde{\mathcal{H}}} \leqslant N$. From the argument above, the $\{\vec{s}_{k_j}\}_{j=1}^{d_{\tilde{\mathcal{H}}}}$ on the right-hand side of Eq. (C16) are linearly independent. Hence, to have Eq. (C16), it is necessary that $c_{k_1} = c_{k_2} = \cdots = c_{k_N} = 0$. In this case, by Eq. (C10), $a_1 = a_2 = \cdots = a_N = 0$, but this contradicts $(a_1, a_2, \ldots, a_N) \neq (0, 0, \ldots, 0)$. Therefore, $d_{\tilde{\mathcal{H}}} > N$.

On the other hand, when $|u_1\rangle$ is given as

$$|u_1\rangle = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{C18}$$

to have Eq. (C10) it is necessary that

$$\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \sum_{j=1}^{d_{\tilde{\mathcal{H}}}} c_{k_j} \vec{t}_{k_j}. \tag{C19}$$

Here we looked at the bottom N entries of each vector in Eq. (C10). With a similar argument, we also have $d_{\tilde{\mathcal{H}}} > N$ in this case. Indeed, by observing that the determinant of the matrix

$$
\begin{pmatrix}
\omega^{Nk_1} & \omega^{Nk_2} & \omega^{Nk_3} & \cdots & \omega^{Nk_N} \\
\omega^{(N+1)k_1} & \omega^{(N+1)k_2} & \omega^{(N+1)k_3} & \cdots & \omega^{(N+1)k_N} \\
\omega^{(N+2)k_1} & \omega^{(N+2)k_2} & \omega^{(N+2)k_3} & \cdots & \omega^{(N+2)k_N} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\omega^{(2N-1)k_1} & \omega^{(2N-1)k_2} & \omega^{(2N-1)k_3} & \cdots & \omega^{(2N-1)k_N}
\end{pmatrix}
$$

is given as

$$
\omega^{N(k_1+k_2+\cdots+k_N)} \prod_{1 \leqslant p < q \leqslant N} (\omega^{k_q} - \omega^{k_p}) \neq 0, \qquad \text{(C20)}
$$

for any N distinct vectors $\{|v_{k_l}\rangle\}_{l=1}^{N}$ chosen from $\{|v_k\rangle\}_{k=1}^{2N}$, $\{\vec{t}_{k_l}\}_{l=1}^{N}$ are linearly independent, and this leads to $d_{\tilde{\mathcal{H}}} > N$.

By applying the same argument, we also have $d_{\mathcal{H}'} > N$. Hence, $d_{\tilde{\mathcal{H}}} + d_{\mathcal{H}'} > 2N$, which contradicts $d_{\tilde{\mathcal{H}}} + d_{\mathcal{H}'} = 2N$. Therefore, σ_1 and σ_2 cannot be expressed as Eqs. (C8) and (C9), that is, σ_1 and σ_2 do not have a shared nontrivial block-diagonal structure.

We proved that σ_1 and σ_2 may only have a single-block KI decomposition. Now we show that σ_1 and σ_2 do not have a single-block redundant part. By way of contradiction, suppose that we may express

$$
\sigma_1 = U(\tilde{\sigma}_1 \otimes \omega)U^{\dagger}, \qquad \text{(C21)}
$$

$$
\sigma_2 = U(\tilde{\sigma}_2 \otimes \omega)U^{\dagger} \qquad \text{(C22)}
$$

using some unitary matrix U and density matrices $\tilde{\sigma}_1$, $\tilde{\sigma}_2$, and ω. We assume ω is $d \times d$ matrix with $d \geqslant 2$ since otherwise $\omega = 1$ and this will not serve as a redundant part. Let V be the unitary matrix corresponding to the permutation of the first and the second systems in Eqs. (C21) and (C22). We may rewrite Eqs. (C21) and (C22) as

$$
\sigma_1 = UV(\omega \otimes \tilde{\sigma}_1)V^{\dagger}U^{\dagger}, \qquad \text{(C23)}
$$

$$
\sigma_2 = UV(\omega \otimes \tilde{\sigma}_2)V^{\dagger}U^{\dagger}. \qquad \text{(C24)}
$$

Suppose that ω is diagonalized as

$$
\omega = U_{\omega} \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_d \end{pmatrix} U_{\omega}^{\dagger} \qquad \text{(C25)}
$$

using some $d \times d$ unitary matrix U_{ω} and the eigenvalues $\{\lambda_j\}_{j=1}^{d}$ of ω. Combining Eqs. (C23)–(C25), we obtain

$$
\sigma_1 = UV(U_{\omega} \otimes I)\left(\bigoplus_{j=1}^{d} \lambda_j \tilde{\sigma}_1 \right)(U_{\omega} \otimes I)^{\dagger}V^{\dagger}U^{\dagger}, \qquad \text{(C26)}
$$

$$
\sigma_2 = UV(U_{\omega} \otimes I)\left(\bigoplus_{j=1}^{d} \lambda_j \tilde{\sigma}_2 \right)(U_{\omega} \otimes I)^{\dagger}V^{\dagger}U^{\dagger}, \qquad \text{(C27)}
$$

where I is the identity matrix on the system $\tilde{\sigma}_1$ and $\tilde{\sigma}_2$ reside in. Here $UV(U_{\omega} \otimes I)$ is also unitary. Since $d \geqslant 2$, from Eqs. (C26) and (C27), σ_1 and σ_2 have a shared nontrivial block-diagonal structure, which is a contradiction. Therefore, σ_1 and σ_2 cannot be expressed as Eqs. (C21) and (C22) and thus σ_1 and σ_2 do not have a redundant part.

[1] C. E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. **27**, 379 (1948).

[2] B. Schumacher, Quantum coding, Phys. Rev. A **51**, 2738 (1995).

[3] R. Jozsa and B. Schumacher, A new proof of the quantum noiseless coding theorem, J. Mod. Opt. **41**, 2343 (1994).

[4] H. Barnum, C. A. Fuchs, R. Jozsa, and B. Schumacher, General fidelity limit for quantum channels, Phys. Rev. A **54**, 4707 (1996).

[5] H.-K. Lo, Quantum coding theorem for mixed states, Opt. Commun. **119**, 552 (1995).

[6] M. Horodecki, Limits for compression of quantum information carried by ensembles of mixed states, Phys. Rev. A **57**, 3364 (1998).

[7] H. Barnum, C. M. Caves, C. A. Fuchs, R. Jozsa, and B. Schumacher, On quantum coding for ensembles of mixed states, J. Phys. A: Math. Gen. **34**, 6767 (2001).

[8] W. Dür, G. Vidal, and J. I. Cirac, Visible compression of commuting mixed states, Phys. Rev. A **64**, 022308 (2001).

[9] G. Kramer and S. A. Savari, Quantum data compression of ensembles of mixed states with commuting density operators, arXiv:quant-ph/0101119

[10] M. Horodecki, Optimal compression for mixed signal states, Phys. Rev. A **61**, 052309 (2000).

[11] M. Hayashi, Optimal visible compression rate for mixed states is determined by entanglement of purification, Phys. Rev. A **73**, 060301(R) (2006).

[12] M. Koashi and N. Imoto, Compressibility of Quantum Mixed-State Signals, Phys. Rev. Lett. **87**, 017902 (2001).

[13] C. H. Bennett, D. P. DiVincenzo, P. W. Shor, J. A. Smolin, B. M. Terhal, and W. K. Wootters, Remote State Preparation, Phys. Rev. Lett. **87**, 077902 (2001).

[14] C. Bennett, P. Hayden, D. Leung, P. Shor, and A. Winter, Remote preparation of quantum states, IEEE Trans. Inf. Theory **51**, 56 (2005).

[15] A. Abeyesinghe, P. Hayden, G. Smith, and A. Winter, Optimal superdense coding of entangled states, IEEE Trans. Inf. Theory **52**, 3635 (2006).

[16] R. Jain, J. Radhakrishnan, and P. Sen, *Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science, Vancouver, 2002* (IEEE, Piscataway, 2002), pp. 429–438.

[17] C. H. Bennett, I. Devetak, A. W. Harrow, P. W. Shor, and A. Winter, The quantum reverse Shannon theorem and resource tradeoffs for simulating quantum channels, IEEE Trans. Inf. Theory **60**, 2926 (2014).

[18] A. Anshu, D. Leung, and D. Touchette, Incompressibility of classical distributions, IEEE Trans. Inf. Theory **68**, 1758 (2022).

[19] Z. B. Khanian and A. Winter, *Proceedings of the IEEE International Symposium on Information Theory, Los Angeles, 2020* (IEEE, Piscataway, 2020), pp. 1852–1857.

[20] H. Barnum, Quantum rate-distortion coding, Phys. Rev. A **62**, 042309 (2000).

[21] I. Devetak and T. Berger, *Proceedings of the IEEE International Symposium on Information Theory, Washington, DC, 2001* (IEEE, Piscataway, 2001), p. 276.

[22] I. Devetak and T. Berger, Quantum rate-distortion theory for memoryless sources, IEEE Trans. Inf. Theory **48**, 1580 (2002).

[23] N. Datta, M.-H. Hsieh, M. M. Wilde, and A. Winter, Quantum-to-classical rate distortion coding, J. Math. Phys. **54**, 042201 (2013).

[24] N. Datta, M.-H. Hsieh, and M. M. Wilde, Quantum rate distortion, reverse Shannon theorems, and source-channel separation, IEEE Trans. Inf. Theory **59**, 615 (2013).

[25] M. M. Wilde, N. Datta, M.-H. Hsieh, and A. Winter, Quantum rate-distortion coding with auxiliary resources, IEEE Trans. Inf. Theory **59**, 6755 (2013).

[26] Z. B. Khanian, K. Kuroiwa, and D. Leung, Rate-distortion theory for mixed states, arXiv:2208.11698.

[27] M. Koashi and N. Imoto, Operations that do not disturb partially known quantum states, Phys. Rev. A **66**, 022318 (2002).

[28] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, 10th ed. (Cambridge University Press, Cambridge, 2010).

[29] M. M. Wilde, *Quantum Information Theory*, 2nd ed. (Cambridge University Press, Cambridge, 2017).

[30] J. Watrous, *The Theory of Quantum Information* (Cambridge University Press, Cambridge, 2018).

[31] A. Winter, Coding theorems of quantum information theory, arXiv:quant-ph/9907077.