# Probabilistic unitary synthesis with optimal accuracy 

SEISEKI AKIBUE, NTT Communication Science Laboratories, NTT Corporation, Japan<br>GO KATO, Advanced ICT Research Institute, NICT, Japan<br>SEIICHIRO TANI, Department of Mathematics, Waseda University, Japan


#### Abstract

The purpose of unitary synthesis is to find a gate sequence that optimally approximates a target unitary transformation. A new synthesis approach, called probabilistic synthesis, has been introduced, and its superiority has been demonstrated over traditional deterministic approaches with respect to approximation error and gate length. However, the optimality of current probabilistic synthesis algorithms is unknown. We obtain the tight lower bound on the approximation error obtained by the optimal probabilistic synthesis, which guarantees the sub-optimality of current algorithms. We also show its tight upper bound, which improves and unifies current upper bounds depending on the class of target unitaries. These two bounds reveal the fundamental relationship of approximation error between probabilistic approximation and deterministic approximation of unitary transformations. From a computational point of view, we show that the optimal probability distribution can be computed by the semidefinite program (SDP) we construct. We also construct an efficient probabilistic synthesis algorithm for single-qubit unitaries, rigorously estimate its time complexity, and show that it reduces the approximation error quadratically compared with deterministic algorithms.


CCS Concepts: • Theory of computation $\rightarrow$ Quantum complexity theory; Quantum information theory.
Additional Key Words and Phrases: quantum gate synthesis, convex approximation, unitary gate decomposition

## 1 INTRODUCTION

In quantum simulation and quantum computation, a global unitary transformation on a many-body quantum system is obtained as a sequence of unitary transformations on a fixed-size system, e.g., those obtained by nearest-neighbor interactions. To guarantee and increase the accuracy of obtaining such transformations, rather than controlling their continuous parameters, each unitary transformation on the fixed-size system is realized as a sequence of gates chosen from a finite gate set $\left\{g_{i}\right\}_{i}$, where each $g_{i}$ results in a fixed unitary transformation with negligible error thanks to the sophisticated calibration, quantum error correction [28] or the nature of the system [17]. If $\left\{g_{i}\right\}_{i}$ is universal, arbitrary unitary transformation can be approximated by a unitary transformation $g_{i_{n}} \circ \cdots \circ g_{i_{2}} \circ g_{i_{1}}$ obtained as a gate sequence for an appropriate choice of gate length $n$ depending on the approximate error one wants to achieve. For a given universal gate set such as the set of the Hadamard, controlled-NOT, and $\pi / 8$ gates [24], an algorithm to find a gate sequence for a given unitary transformation and an approximation error bound is called a unitary synthesis algorithm.

To suppress the effect of decoherence or overhead caused by the fault-tolerant implementation of each gate [1, 22], various studies $[3,4,9,13,18,20,21,25]$ have proposed unitary synthesis algorithms for minimizing the length of the output gate sequence. Following the celebrated Solovay-Kitaev algorithm [18], many algorithms are used to find one of the shortest gate sequences that can approximate a target unitary transformation $\Upsilon$ within the desired approximation error. Obviously, the goal can be achieved by brute force search [9]. However, to guarantee their efficiency, many algorithms are designed for synthesizing restricted classes of unitary transformations by using particular gate sets or for finding a nearly shortest gate sequence.

While approximating an $\Upsilon$ by using a single sequence of gates is a natural approach, the advantage of another approach using a probabilistic mixture of unitaries has been demonstrated [6, 14, 19]. Suppose that a synthesis algorithm
produces a gate sequence for implementing a unitary transformation in $\left\{\Upsilon_{\vec{i}}\right\}=\left\{g_{i_{n}} \circ \cdots \circ g_{i_{2}} \circ g_{i_{1}}\right\}_{\vec{i}}$ in accordance with the probability distribution $p(\vec{i})$ to approximate an $\Upsilon$. If the algorithm independently samples $\vec{i}$ for each time the $\Upsilon$ is used in the entire circuit, the physical transformation governed by the randomly executed unitary transformation $\Upsilon_{i}$ in accordance with the $p(\vec{i})$ is described by a probabilistic mixture $\sum_{\vec{i}} p(\vec{i}) \Upsilon_{\vec{i}}$ of unitaries. In this case, the approximation error should be measured by the distance between the $\Upsilon$ and $\sum_{\vec{i}} p(\vec{i}) \Upsilon_{\vec{i}}$.

Campbell [6] and Vadym et al. [19] constructed algorithms to compute a probability distribution $\{p(\vec{i})\}_{\vec{i}}$ for a given $\Upsilon$ and a set $\left\{\Upsilon_{\vec{i}}\right\}_{\vec{i}}$ of unitary transformations implemented as a gate sequence such that the approximation error of $\sum_{\vec{i}} p(\vec{i}) \Upsilon_{\vec{i}}$ against $\Upsilon$ is almost quadratically better than that of a single optimal unitary transformation in $\left\{\Upsilon_{\vec{i}}\right\}_{\vec{i}}$. More precisely, $\| \Upsilon-\sum_{\vec{i}} p\left(\vec{i} \Upsilon_{\vec{i}} \|_{\diamond}=O\left(\epsilon^{2}\right)\right.$ for the worst approximation error $\epsilon=\max _{\Upsilon} \min _{\vec{i}} \frac{1}{2}\left\|\Upsilon-\Upsilon_{\vec{i}}\right\|_{\diamond}$ caused by deterministic synthesis, where $\|\mathcal{A}-\mathcal{B}\|_{\diamond}$ is the diamond norm [18, 29]. This also indicates that probabilistically executing $\Upsilon_{\vec{i}}$ in accordance with $p(\vec{i})$ can further reduce the length of the shortest gate sequence without increasing the approximation error (if one measures the error by using the above diamond norm) [6]. In general, probabilistic synthesis consists of two procedures; (i) computing a probability distribution $\{p(\vec{i})\}_{\vec{i}}$ and sampling a description $\vec{i}$ of a gate sequence with a classical computer, and (ii) implementing the sampled gate sequence on a quantum computer. In contrast to deterministic synthesis, procedure (ii) may require a quantum computer to rearrange a gate sequence each time it realizes a target unitary transformation. However, such rearrangeability is usually assumed as a standard functionality of a quantum computer.

However, procedure (i) should be meticulously designed to construct a practical synthesis algorithm. This is because the number of possible gate sequences grows exponentially with respect to the length $n$ of the sequence, resulting in a large degree of freedom in choosing $\{p(\vec{i})\}_{\vec{i}}$. While a probabilistic synthesis algorithm with guaranteed time complexity exists for single-qubit unitary transformations that correspond to axial rotations [19], no such algorithm was known even for general single-qubit unitary transformations. Furthermore, a fundamental question remained open regarding the optimality of existing synthesis algorithms in comparison to the minimum approximation error $\min _{p}\left\|\Upsilon-\sum_{\vec{i}} p(\vec{i}) \Upsilon_{\overparen{i}}\right\|_{\Delta}$. Minimax optimization makes it difficult to investigate the minimum approximation error from an analytical perspective except for a few specific $\Upsilon$ and sets $\left\{\Upsilon_{\vec{i}}\right\}_{\vec{i}}$ [27].

### 1.1 Our contribution

We obtain the tight lower bound on $\min _{p}\left\|\Upsilon-\sum_{\vec{i}} p(\vec{i}) \Upsilon_{\vec{i}}\right\|_{\Delta}$, which reveals the fundamental limitation of probabilistic synthesis and indicates the sub-optimality of current algorithms. To obtain the main result, we focus on the analytical relationship between $\min _{p}\left\|\Upsilon-\sum_{\vec{i}} p(\vec{i}) \Upsilon_{\overparen{i}}\right\|_{\diamond}$ and $\min _{\vec{i}}\left\|\Upsilon-\Upsilon_{\vec{i}}\right\|_{\diamond}$, which represent the minimum approximation error obtained by probabilistic synthesis and that by deterministic synthesis, respectively. To be mathematically comprehensive, we also obtain the tight upper bound on $\min _{p}\left\|\Upsilon-\sum_{\vec{i}} p(\vec{i}) \Upsilon_{\vec{i}}\right\|_{\circlearrowleft}$, which essentially unifies various upper bounds $[6,14,19]$ depending on the class of target unitary transformations. More precisely, the two bounds are given as the following theorem.

THEOREM 4.3. (simplified version) For an integer $d \geq 2$ specified below, let $\Upsilon$ and $\left\{\Upsilon_{\vec{i}}\right\}_{\vec{i}}$ be a target unitary transformation and a finite set of unitary transformations on the d-dimensional Hilbert space, respectively. It then holds that

$$
\frac{4 \delta}{d}\left(1-\frac{\delta}{d}\right) \leq \max _{\Upsilon} \min _{p} \frac{1}{2}\left\|\Upsilon-\sum_{\vec{i}} p(\vec{i}) \Upsilon_{\vec{i}}\right\|_{\diamond} \leq \epsilon^{2} \text { with }\left\{\begin{array}{l}
\delta=1-\sqrt{1-\epsilon^{2}} \text { and }  \tag{1}\\
\epsilon=\max _{\Upsilon} \min _{\vec{i}} \frac{1}{2}\left\|\Upsilon-\Upsilon_{\Upsilon_{i}}\right\|_{\diamond}
\end{array}\right.
$$

This theorem provides bounds on the worst approximation error caused when one probabilistically synthesizes the target unitary transformation that is most difficult to approximate. As shown in Fig. 1, the gap between the upper and lower bounds exists if and only if $d \geq 3$. We can show that the gap is inevitable by constructing $\left\{\left\{_{\vec{i}}\right\}_{\vec{i}}\right.$ for achieving the upper bound and that for achieving the lower bound. That is, Ineq. (1) represents the fundamental relationship of the approximation error between the deterministic approximation of unitary transformations and their probabilistic approximation that depends only on the dimension $d$ of the system.


Fig. 1. Lower and upper bounds on worst approximation error $\max _{\Upsilon} \min _{p} \frac{1}{2}\left\|\Upsilon-\sum_{\vec{i}} p(\vec{i}) \Upsilon_{\vec{i}}\right\|_{\delta}$ caused by probabilistic synthesis with respect to $\max _{\Upsilon} \min _{i} \frac{1}{2}\left\|\Upsilon-\Upsilon_{i}\right\|_{,}$caused by deterministic synthesis for two-qubit systems, i.e., $d=4$. Both lower and upper bounds, represented with thick and thin curves, respectively, are achievable for certain $\left\{\left\{_{\bar{i}}\right\}\right.$.

From a computational point of view, we show that the optimal probability distribution for approximating an $\Upsilon$ can be computed by the semidefinite program (SDP) we construct when the set $\left\{\Upsilon_{\vec{i}}\right\}_{\vec{i}}$ of unitary transformations implemented as a gate sequence is given. (This set is computable with certain synthesis algorithms.) In addition to its optimality, we can rigorously estimate the worst time complexity of our SDP due to established methods for numerically solving SDPs. As the second main result, we construct a probabilistic synthesis algorithm for single-qubit unitary transformations from the following theorem.

THEOREM 5.4. (informal version) For a given gate set, there exists a probabilistic synthesis algorithm for a single-qubit unitary transformation with

INPUT: a target single-qubit unitary transformation $\Upsilon$ and target approximation error $\epsilon \in(0,1)$
OUTPUT: a gate sequence implementing a single-qubit unitary transformation $\Upsilon_{\vec{i}}$ sampled from a set $\left\{\Upsilon_{\vec{i}}\right\}_{\vec{i}}$ in accordance with probability distribution $\hat{p}(\vec{i})$.
such that the algorithm satisfies the following properties:

- Efficiency: All steps of the algorithm take polylog $\left(\frac{1}{\epsilon}\right)$-time,
- Quadratic improvement: The approximation error $\frac{1}{2}\left\|\Upsilon-\sum_{\vec{i}} \hat{p}(\vec{i}) \Upsilon_{\vec{i}}\right\|_{\diamond}$ obtained with this algorithm is upper bounded by $\epsilon^{2}$, whereas the error $\min _{\vec{i}} \frac{1}{2}\left\|\Upsilon-\Upsilon_{\vec{i}}\right\|_{\text {。 }}$ obtained by deterministic synthesis using the unitary transformations in $\left\{\Upsilon_{\vec{i}}\right\}_{\vec{i}}$ is upper bounded by $\epsilon$.

The first property of the algorithm is desirable for fault-tolerant quantum computation (FTQC). The polylog $\left(\frac{1}{\epsilon}\right)$ time overhead due to the synthesis algorithm does not impair a quadratic speedup achieved with a quantum computer over a classical computer since the approximation error of each unitary transformation should satisfy $\frac{1}{\epsilon}=$ poly (n) if a quantum circuit contains a polynomial number of single-qubit unitaries with respect to the problem size $n$. Due to the second property of the algorithm, we can verify that it surpasses current algorithms [6,14] with respect to the approximation error. Our algorithm also surpasses a current algorithm [19] in terms of applicability to a general single-qubit unitary transformation.

### 1.2 Technical outline

Previous studies searched for the mixing probability distribution $\{p(\vec{i})\}_{\vec{i}}$ by using the first-order approximation of unitary operators [6, 14] and obtained the upper bound on the worst approximation error max $\min _{p} \frac{1}{2}\left\|\Upsilon-\sum_{\vec{i}} p(\vec{i}) \Upsilon_{\vec{i}}\right\|_{\diamond}$ caused by probabilistic synthesis. In contrast, we use the strong duality of SDP, essentially equivalent to the minimax theorem, to obtain tight bounds Ineq. (1) obtained by the optimal mixing probability distribution. A similar technique can be found in the analyses of the optimal convex approximation of quantum states by using a restricted set of states [26] and that of unital mappings by using unitary transformations [30]. While inventing tractable upper bounds on the approximation error of a general unital mapping is an open problem [30], we provide an upper bound by exploiting the property of a unitary transformation as a pure unital mapping.

To prove that our single-qubit unitary synthesis algorithm satisfies the expected properties, we show the fact that $\Upsilon_{\vec{i}}$ that is far from $\Upsilon$ is not necessary to be sampled to optimally approximate $\Upsilon$ for single-qubit unitary transformations by exploiting the magic basis [2] representation of single-qubit unitary operators. The magic basis representation enables us to embed the metric space of single-qubit unitary transformations induced by the diamond norm into that of $S^{3}$ with respect to the angle. While numerical simulations indicate the same fact holds for qudit unitary transformations, rigorous proof is a subject for future work.

### 1.3 Organization

This article is organized as follows. Section 2 is devoted to preliminaries, introducing basis notations in quantum information theory and semidefinite programming. In Section 3, we construct an SDP that computes the optimal probability distribution in probabilistic synthesis. The SDP is provided as a primal and dual problem whose solutions coincide due to the strong duality of the SDP. The coincidence plays a crucial role in the proof of the first main theorem about the fundamental limitation on the approximation error shown in Section 4. Section 5 provides an efficient probabilistic synthesis algorithm for single-qubit unitary transformations as the second main theorem. We also provide a simple geometric interpretation of the superiority of probabilistic synthesis by considering single-qubit unitary transformations corresponding to axial rotations in Section 5.2. We present our conclusions in Section 6.

## 2 PRELIMINARIES

In this section, we summarize basic notations used throughout the paper. Note that we consider only finite-dimensional Hilbert spaces. In particular, two-dimensional Hilbert space $\mathbb{C}^{2}$ is called a qubit. The $\mathbf{L}(\mathcal{H})$ and Pos $(\mathcal{H})$ represent the set of linear operators and positive semidefinite operators on Hilbert space $\mathcal{H}$, respectively. $\mathbb{I} \in \operatorname{Pos}(\mathcal{H})$ represents the identity operator, and we sometimes use the subscript to specify the system where $\mathbb{I}$ acts as $\mathbb{I}_{\mathcal{H}}$. For Hermitian operators $A$ and $B$ on $\mathcal{H}, A \geq B$ represents $A-B \in \operatorname{Pos}(\mathcal{H})$, and $A>B$ represents $A-B$ is positive definite. The $\mathbf{S}(\mathcal{H}):=\{\rho \in \operatorname{Pos}(\mathcal{H}): \operatorname{tr}[\rho]=1\}$ and $\mathbf{P}(\mathcal{H}):=\left\{\rho \in \mathbf{S}(\mathcal{H}): \operatorname{tr}\left[\rho^{2}\right]=1\right\}$ represent the set of quantum states and that of pure states, respectively. Pure state $\phi \in \mathbf{P}(\mathcal{H})$ is sometimes alternatively represented by complex unit vector $|\phi\rangle \in \mathcal{H}$ satisfying $\phi=|\phi\rangle\langle\phi|$. Any physical transformation of the quantum state can be represented by a completely positive and trace preserving (CPTP) linear mapping $\Gamma: \mathbf{L}\left(\mathcal{H}_{1}\right) \rightarrow \mathbf{L}\left(\mathcal{H}_{2}\right)$. There exists one-to-one correspondence between a linear mapping $\Xi: \mathbf{L}\left(\mathcal{H}_{1}\right) \rightarrow \mathbf{L}\left(\mathcal{H}_{2}\right)$ and its Choi-Jamiołkowski operator $J(\Xi):=\sum_{i, j}|i\rangle\langle j| \otimes \Xi(|i\rangle\langle j|) \in$ $\mathrm{L}\left(\mathcal{H}_{1} \otimes \mathcal{H}_{2}\right)$.

The trace distance $\|\rho-\sigma\|_{\text {tr }}$ of two quantum states $\rho, \sigma \in \mathbf{S}(\mathcal{H})$ is defined as $\|M\|_{\operatorname{tr}}:=\frac{1}{2} \operatorname{tr}\left[\sqrt{M M^{\dagger}}\right]$ for $M \in \mathbf{L}(\mathcal{H})$. It represents the maximum total variation distance between probability distributions obtained from measurements performed on two quantum states. A similar notion measuring the distinguishability of $\rho$ and $\sigma$ is the fidelity function, defined by $F(\rho, \sigma):=\max \operatorname{tr}\left[\Phi^{\rho} \Phi^{\sigma}\right]$, where $\Phi^{\rho} \in \mathbf{P}\left(\mathcal{H} \otimes \mathcal{H}^{\prime}\right)$ is a purification of $\rho$, i.e., $\rho=\operatorname{tr}_{\mathcal{H}^{\prime}}\left[\Phi^{\rho}\right]$, and the maximization is taken over all the purifications. Fuchs-van de Graaf inequalities [10] provide relationships between the two measures with respect to the distinguishability as follows:

$$
\begin{equation*}
1-\sqrt{F(\rho, \sigma)} \leq\|\rho-\sigma\|_{\mathrm{tr}} \leq \sqrt{1-F(\rho, \sigma)} \tag{2}
\end{equation*}
$$

holds for any state $\rho, \sigma \in \mathbf{S}(\mathcal{H})$, where the equality of the right inequality holds when $\rho$ and $\sigma$ are pure.
The distance measuring the distinguishability of two CPTP mappings $\mathcal{A}, \mathcal{B}: \mathbf{L}\left(\mathcal{H}_{1}\right) \rightarrow \mathbf{L}\left(\mathcal{H}_{2}\right)$ corresponding to the trace distance is the diamond norm $\|\mathcal{A}-\mathcal{B}\|_{\diamond}$ defined by $\frac{1}{2}\|\mathcal{A}-\mathcal{B}\|_{\diamond}:=\max _{\Phi \in \mathbf{P}\left(\mathcal{H}_{1} \otimes \mathcal{H}_{3}\right)}\|((\mathcal{A}-\mathcal{B}) \otimes i d)(\Phi)\|_{\mathrm{tr}}$, where id represents the identity mapping acting on $\mathcal{H}_{3}$.

Let $\Xi: \mathbf{L}\left(\mathcal{H}_{1}\right) \rightarrow \mathbf{L}\left(\mathcal{H}_{2}\right)$ be a linear Hermitian-preserving mapping and $A$ and $B$ be Hermitian operators on $\mathcal{H}_{1}$ and $\mathcal{H}_{2}$, respectively. SDP is an optimization problem formally defined with a triple ( $\Xi, A, B$ ) as follows [29]:

\[

\]

\[

\]

where $\Xi^{\dagger}: \mathbf{L}\left(\mathcal{H}_{2}\right) \rightarrow \mathbf{L}\left(\mathcal{H}_{1}\right)$ is the adjoint of $\Xi$, defined as the linear mapping satisfying $\operatorname{tr}\left[Y^{\dagger} \Xi(X)\right]=\operatorname{tr}\left[\left(\Xi^{\dagger}(Y)\right)^{\dagger} X\right]$ for all $X \in \mathbf{L}\left(\mathcal{H}_{1}\right)$ and $Y \in \mathbf{L}\left(\mathcal{H}_{2}\right)$. We can easily verify that the solution to the primal problem is smaller than or equal to that of the dual problem. The situation when the two solutions coincide is called a strong duality. Slater's theorem states that the strong duality holds if either of the following conditions holds:
(1) The solution to the primal problem is finite, and there exists a Hermitian operator $Y$ on $\mathcal{H}_{2}$ such that $\Xi^{\dagger}(Y)>A$.
(2) The solution to the dual problem is finite, and there exists a positive definite operator $X$ on $\mathcal{H}_{1}$ such that $\Xi(X)=B$.

For a metric space ( $X, d$ ) and two subsets $S, T \subseteq X, S$ is called an $\epsilon$-covering of $T$ if $\sup _{t \in T} \inf _{s \in S} d(s, t) \leq \epsilon$. In this article, we basically assume that $X$ is the set of CPTP mappings, the metric is defined as $d(\mathcal{A}, \mathcal{B})=\frac{1}{2}\|\mathcal{A}-\mathcal{B}\|_{\diamond}, S$ is a finite set of unitary transformations and $T$ is a subset of unitary transformations such as a $2 \epsilon$-ball $\left\{\Upsilon^{\prime}: \frac{1}{2}\left\|\Upsilon^{\prime}-\Upsilon\right\|_{\diamond} \leq 2 \epsilon\right\}$ around a unitary transformation $\Upsilon: \mathbf{L}(\mathcal{H}) \rightarrow \mathbf{L}(\mathcal{H})$.

## 3 SEMIDEFINITE PROGRAMMING FOR COMPUTING OPTIMAL MIXING PROBABILITY

In this section, we construct an SDP for computing the optimal probability distribution that minimizes the diamond norm between the target CPTP mapping $\mathcal{A}$ and a probabilistic mixture of CPTP mappings $\left\{\mathcal{B}_{x}\right\}_{x}$. We can compute the optimal probability distribution in probabilistic unitary synthesis by solving this SDP by restricting $\mathcal{A}$ and $\left\{\mathcal{B}_{x}\right\}_{x}$ as unitary transformations. We also mention the relationship between our SDP and the algorithm proposed by Campbell [6].

Proposition 3.1. Let $\mathcal{A}$ and $\left\{\mathcal{B}_{x}\right\}_{x \in X}$ be a target CPTP mapping and a finite set of CPTP mappings from $\mathbf{L}\left(\mathcal{H}_{1}\right)$ to $\mathbf{L}\left(\mathcal{H}_{2}\right)$, respectively. Then, distance $\min _{p} \frac{1}{2}\left\|\mathcal{A}-\sum_{x \in X} p(x) \mathcal{B}_{x}\right\|_{\diamond}$ and the optimal probability distribution $\{p(x)\}_{x \in X}$, which minimizes the distance, can be computed with the following SDP:

$$
\begin{array}{llll} 
& & & \\
\text { maximal problem } & & \text { Dual problem }  \tag{4}\\
\text { subject to: } & \text { tr }[J(\mathcal{A}) T]-t & \text { minimize: } & r \in \mathbb{R} \\
& \rho \leq T \leq \rho \otimes \mathbb{I}_{\mathcal{H}_{2}}, & \text { subject to: } & S \geq 0 \wedge S \geq J\left(\mathcal{A}-\sum_{x \in X} p(x) \mathcal{B}_{x}\right), \\
& & r \mathbb{I}_{\mathcal{H}_{1} \geq \operatorname{tr}_{\mathcal{H}_{2}}[S],} & \forall x \in X, p(x) \geq 0, \\
& \forall x \in X, \operatorname{tr}\left[J\left(\mathcal{B}_{x}\right) T\right] \leq t . & & \sum_{x \in X} p(x) \leq 1 .
\end{array}
$$

Note that the strong duality holds in this SDP, i.e., the optimum primal and dual values are equal.
Proof. Recall that for two CPTP mapping $\mathcal{A}$ and $\mathcal{B}$ from $\mathbf{L}\left(\mathcal{H}_{1}\right)$ to $\mathbf{L}\left(\mathcal{H}_{2}\right), \frac{1}{2}\|\mathcal{A}-\mathcal{B}\|_{\diamond}$ can be computed by the following SDP:

| Primal problem |  |  | $\frac{\text { Dual problem }}{}$ |
| :--- | :--- | :--- | :--- |
| maximize: | $\operatorname{tr}[J(\mathcal{A}-\mathcal{B}) T]$ | minimize: | $r \in \mathbb{R}$ |
| subject to: | $0 \leq T \leq \rho \otimes \mathbb{I}_{\mathcal{H}_{2}}$, | subject to: | $S \geq 0 \wedge S \geq J(\mathcal{A}-\mathcal{B})$, |
|  | $\rho \in \mathbf{S}\left(\mathcal{H}_{1}\right)$. |  | $r \mathbb{I}_{\mathcal{H}_{1}} \geq \operatorname{tr}_{\mathcal{H}_{2}}[S]$. |

The primal problem can be obtained by observing

$$
\begin{align*}
\frac{1}{2}\|\mathcal{A}-\mathcal{B}\|_{\diamond} & =\max _{\substack{\operatorname{\in }\left(\mathcal{H}_{1} \otimes \mathcal{H}_{3}\right) \\
\Pi \in \operatorname{Proj}\left(\mathcal{H}_{2} \otimes \mathcal{H}_{3}\right)}} \operatorname{tr}[((\mathcal{A}-\mathcal{B}) \otimes i d)(\Phi) \Pi]  \tag{5}\\
& =\max _{T \in \mathbf{T}\left(\mathcal{H}_{1}: \mathcal{H}_{2}\right)} \operatorname{tr}[J(\mathcal{A}-\mathcal{B}) T] \tag{6}
\end{align*}
$$

where $\Pi$ is a Hermitian projector acting on $\mathcal{H}_{2} \otimes \mathcal{H}_{3}, \mathrm{~T}\left(\mathcal{H}_{1}: \mathcal{H}_{2}\right):=\left\{T \in \operatorname{Pos}\left(\mathcal{H}_{1} \otimes \mathcal{H}_{2}\right): \exists \rho \in \mathrm{S}\left(\mathcal{H}_{1}\right), T \leq \rho \otimes \mathbb{I}\right\}$ is called the set of measuring strategies [11] or that of quantum testers [7], and the last equality was shown by Chiribella et al. [7, Theorem 10]. To be self-contained, we provide a proof for the equality in Appendix A, with which the equality can be verified by applying Eq. (55) with fixing $\Xi=\mathcal{A}-\mathcal{B}$. A formal SDP and the verification of the strong duality are provided in Appendix B.

By extending the dual problem of this SDP to include the minimization of probability distribution $\{p(x)\}_{x \in X}$, we obtain Eq. (4). Note that the last condition $\sum_{x \in X} p(x) \leq 1$ in the dual problem is different from the condition $\sum_{x \in X} p(x)=1$ of a probability distribution; however, the optimum dual value can be achieved under the latter condition. Again, a formal SDP and the verification of the strong duality are provided in Appendix B.

For a given $\Upsilon: \mathbf{L}(\mathcal{H}) \rightarrow \mathbf{L}(\mathcal{H})$ and a given set $\left\{\Upsilon_{x}: \mathbf{L}(\mathcal{H}) \rightarrow \mathbf{L}(\mathcal{H})\right\}_{x \in X}$ of unitary transformations implemented as a gate sequence, which forms an $\epsilon$-covering the set of unitary transformations with sufficiently small $\epsilon$, "the convex hull finding algorithm" proposed by Campbell [6] can find a probability distribution $\{\tilde{p}(x)\}_{x \in \tilde{X}}$ such that $\sum_{x \in \tilde{X}} \tilde{p}(x) H_{x}=0$, where $\Upsilon_{x}(\rho)=\Upsilon\left(e^{i H_{x}} \rho e^{-i H_{x}}\right)$ and $H_{x}=O(\epsilon)$ for all $x \in \tilde{X} \subseteq X$. Note that $M=O(\epsilon)$ represents $\|M\|_{\infty}=O(\epsilon)$ as $\epsilon \rightarrow 0$ for a linear operator $M \in \mathrm{~L}(\mathcal{H})$ depending on $\epsilon$. By using the dual problem in Proposition 3.1, we can verify that the distance $\epsilon$, which is achievable by a deterministic unitary synthesis finding the closest $\Upsilon_{x}$ to approximate $\Upsilon$, can be improved into $O\left(\epsilon^{2}\right)$ by mixing unitaries in accordance with the probability distribution $\{\tilde{p}(x)\}_{x \in \tilde{X}}$ as follows. First, by using the dual problem of the SDP to compute the diamond norm between two CPTP mappings, we obtain

$$
\begin{array}{r}
\frac{1}{2}\left\|\Upsilon-\sum_{x \in \tilde{X}} \tilde{p}(x) \Upsilon_{x}\right\|_{\diamond}=\frac{1}{2}\left\|i d-\sum_{x \in \tilde{X}} \tilde{p}(x) \Upsilon^{-1} \circ \Upsilon_{x}\right\|_{\circ} \leq\left\|\operatorname{tr}_{\mathcal{H}^{\prime}}[S]\right\|_{\infty} \\
\text { with } S \geq 0 \wedge S \geq J(i d)-\sum_{x \in \tilde{X}} \tilde{p}(x) J\left(\Upsilon^{-1} \circ \Upsilon_{x}\right), \tag{8}
\end{array}
$$

where $\mathcal{H}^{\prime}$ represents the the output system of $\Upsilon$, which is isomorphic to $\mathcal{H}$. Second, by using the Taylor expansions $e^{i H_{x}}=\mathbb{I}+i H_{x}+R_{x}$, where $R_{x}=O\left(\epsilon^{2}\right)$, we obtain

$$
\begin{align*}
J(i d)-\sum_{x \in \tilde{X}} \tilde{p}(x) J\left(\Upsilon^{-1} \circ \Upsilon_{x}\right)= & \sum_{x \in \tilde{X}} \tilde{p}(x)\left\{-\left(R_{x} J(i d)+J(i d) R_{x}^{\dagger}\right)-i\left(H_{x} J(i d) R_{x}^{\dagger}-R_{x} J(i d) H_{x}\right)\right\}-P  \tag{9}\\
\leq & \sum_{x \in \tilde{X}} \tilde{p}(x)\left\{\left(\frac{1}{\left\|R_{x}\right\|_{\infty}} R_{x} J(i d) R_{x}^{\dagger}+\left\|R_{x}\right\|_{\infty} J(i d)\right)\right. \\
& \left.+\left(\frac{\left\|R_{x}\right\|_{\infty}}{\left\|H_{x}\right\|_{\infty}} H_{x} J(i d) H_{x}+\frac{\left\|H_{x}\right\|_{\infty}}{\left\|R_{x}\right\|_{\infty}} R_{x} J(i d) R_{x}^{\dagger}\right)\right\} \tag{10}
\end{align*}
$$

where $P=\sum_{x \in \tilde{X}} \tilde{p}(x)\left(H_{x} J(i d) H_{x}+R_{x} J(i d) R_{x}^{\dagger}\right) \in \operatorname{Pos}\left(\mathcal{H} \otimes \mathcal{H}^{\prime}\right), H_{x}$ and $R_{x}$ acts on $\mathcal{H}^{\prime}$ and we use the fact that $|\tilde{\phi}\rangle\langle\tilde{\psi}|+|\tilde{\psi}\rangle\langle\tilde{\phi}| \leq \tilde{\phi}+\tilde{\psi}$ with complex vectors $(|\tilde{\phi}\rangle,|\tilde{\psi}\rangle)=\left(\left\|R_{x}\right\|_{\infty}^{-\frac{1}{2}} \sum_{j}\left(\mathbb{I}_{\mathcal{H}} \otimes R_{x}\right)|j j\rangle,\left\|R_{x}\right\|_{\infty}^{\frac{1}{2}} \sum_{j}|j j\rangle\right)$ and $(|\tilde{\phi}\rangle,|\tilde{\psi}\rangle)=$ $\left(\left\|R_{x}\right\|_{\infty}^{\frac{1}{2}}\left\|H_{x}\right\|_{\infty}^{-\frac{1}{2}} \sum_{j}\left(\mathbb{I}_{\mathcal{H}} \otimes H_{x}\right)|j j\rangle, i\left\|R_{x}\right\|_{\infty}^{-\frac{1}{2}}\left\|H_{x}\right\|_{\infty}^{\frac{1}{2}} \sum_{j}\left(\mathbb{I}_{\mathcal{H}} \otimes R_{x}\right)|j j\rangle\right)$ in the inequality. Third, by letting $S$ in Eq. (8) be R.H.S. of Eq. (10), we obtain

$$
\begin{equation*}
\left\|\operatorname{tr}_{\mathcal{H}^{\prime}}[S]\right\|_{\infty}=\left\|\sum_{x \in \tilde{X}} \tilde{p}(x)\left(\frac{\left(R_{x}^{\dagger} R_{x}\right)^{T}}{\left\|R_{x}\right\|_{\infty}}+\left\|R_{x}\right\|_{\infty} \mathbb{I}_{\mathcal{H}}+\frac{\left\|R_{x}\right\|_{\infty}\left(H_{x}^{2}\right)^{T}}{\left\|H_{x}\right\|_{\infty}}+\frac{\left\|H_{x}\right\|_{\infty}\left(R_{x}^{\dagger} R_{x}\right)^{T}}{\left\|R_{x}\right\|_{\infty}}\right)\right\|_{\infty}=O\left(\epsilon^{2}\right) \tag{11}
\end{equation*}
$$

Since the approximation error $\frac{1}{2}\left\|\Upsilon-\sum_{x \in \tilde{X}} \tilde{p}(x) \Upsilon_{x}\right\|_{\diamond}$ is generally worse than the optimal one $\min _{p} \frac{1}{2}\left\|\Upsilon-\sum_{x \in X} p(x) \Upsilon_{x}\right\|_{\diamond}$, we can obtain a better probability distribution and better estimation of the approximation error by numerically solving the SDP shown in Proposition 3.1. The ellipsoid method guarantees that $\{p(x)\}_{x \in X}$ and $r$ in the dual problem such that the difference between $r$ and the optimum dual value is less than $\epsilon$ can be computed in poly $\left(|X| \log \left(\frac{1}{\epsilon}\right)\right)$-time [23].

Note that we assume the dimension of the Hilbert space is constant since the unitary synthesis is usually executed for $\Upsilon$ on a fixed-size system.

## 4 TIGHT BOUNDS ON ERROR OF PROBABILISTIC APPROXIMATION

This section investigates the relationship between the discrete approximation of unitary transformations and the probabilistic approximation for a general $\Upsilon$ and general set $\left\{\Upsilon_{x}\right\}_{x}$. Specifically, we show the tight relationship between $\min _{p}\left\|\Upsilon-\sum_{x} p(x) \Upsilon_{x}\right\|_{\diamond}$ and $\min _{x}\left\|\Upsilon-\Upsilon_{x}\right\|_{\diamond}$, where the former represents the minimum approximation error obtained by probabilistic synthesis and the latter represents that by deterministic synthesis when $\left\{\Upsilon_{x}\right\}_{x}$ is a set of unitary transformations implemented as a gate sequence. The first lemma shows the fundamental limitation of probabilistic synthesis, and the second one shows its superiority over deterministic synthesis.

Lemma 4.1. For an integer $d \geq 2$ specified below, let $\Upsilon: \mathbf{L}\left(\mathbb{C}^{d}\right) \rightarrow \mathbf{L}\left(\mathbb{C}^{d}\right)$ and $\left\{\Upsilon_{x}: \mathbf{L}\left(\mathbb{C}^{d}\right) \rightarrow \mathbf{L}\left(\mathbb{C}^{d}\right)\right\}_{x \in X}$ be a target unitary transformation and finite set of unitary transformations, respectively. Then

$$
\frac{2}{d} \epsilon^{2} \leq \frac{4 \delta}{d}\left(1-\frac{\delta}{d}\right) \leq \min _{p} \frac{1}{2}\left\|\Upsilon-\sum_{x \in X} p(x) \Upsilon_{x}\right\|_{\diamond} \text { with }\left\{\begin{array}{l}
\delta=1-\sqrt{1-\epsilon^{2}} \text { and }  \tag{12}\\
\epsilon=\min _{x \in X} \frac{1}{2}\left\|\Upsilon-\Upsilon_{x}\right\|_{\diamond}
\end{array}\right.
$$

holds, where the minimization of $p$ is taken over probability distributions over $X$.

Proof. The first inequality can be straightforwardly verified as follows:

$$
\begin{equation*}
\frac{2}{d} \epsilon^{2}=\frac{4 \delta}{d}\left(1-\frac{\delta}{2}\right) \leq \frac{4 \delta}{d}\left(1-\frac{\delta}{d}\right) \tag{13}
\end{equation*}
$$

Thus, we prove the second inequality. First, by computing the diamond norm between $\Upsilon$ and $\Upsilon_{x}$, we obtain

$$
\begin{align*}
\frac{1}{2}\left\|\Upsilon-\Upsilon_{x}\right\|_{\diamond} & =\max _{\Phi \in \mathbf{P}\left(\mathbb{C}^{d} \otimes \mathbb{C}^{d}\right)}\left\|\Upsilon \otimes i d_{\mathbb{C}^{d}}(\Phi)-\Upsilon_{x} \otimes i d_{\mathbb{C}^{d}}(\Phi)\right\|_{\operatorname{tr}}  \tag{14}\\
& =\max _{\Phi \in \mathbf{P}\left(\mathbb{C}^{d} \otimes \mathbb{C}^{d}\right)} \sqrt{1-F\left(\Upsilon \otimes i d_{\mathbb{C}^{d}}(\Phi), \Upsilon_{x} \otimes i d_{\mathbb{C}^{d}}(\Phi)\right)}  \tag{15}\\
& =\sqrt{\left.1-\min _{\Phi \in \mathbf{P}\left(\mathbb{C}^{d} \otimes \mathbb{C}^{d}\right)}\left|\langle\Phi| U^{\dagger} U_{x} \otimes \mathbb{I}_{\mathbb{C}^{d}}\right| \Phi\right\rangle\left.\right|^{2}}  \tag{16}\\
& =\sqrt{1-\min _{\rho \in \mathrm{S}\left(\mathbb{C}^{d}\right)}\left|\operatorname{tr}\left[\rho U^{\dagger} U_{x}\right]\right|^{2}}, \tag{17}
\end{align*}
$$

where $\Upsilon(\rho)=U \rho U^{\dagger}$ and $\Upsilon_{x}(\rho)=U_{x} \rho U_{x}^{\dagger}$. This indicates

$$
\begin{equation*}
1-\delta=\max _{x \in X} \min _{\rho \in \mathrm{S}\left(\mathbb{C}^{d}\right)}\left|\operatorname{tr}\left[\rho U^{\dagger} U_{x}\right]\right| \tag{18}
\end{equation*}
$$

Next, by using the primal problem in our SDP in Proposition 3.1, we obtain

$$
\begin{align*}
\min _{p} \frac{1}{2}\left\|\Upsilon-\sum_{x \in X} p(x) \Upsilon_{x}\right\|_{\diamond} & =\max _{T \in \mathbf{T}\left(\mathbb{C}^{d}: \mathbb{C}^{d}\right)}\left(\operatorname{tr}[J(\Upsilon) T]-\max _{x \in X} \operatorname{tr}\left[J\left(\Upsilon_{x}\right) T\right]\right)  \tag{19}\\
& \geq \frac{1}{d^{2}}\left(\operatorname{tr}[J(\Upsilon) J(\Upsilon)]-\max _{x \in X} \operatorname{tr}\left[J\left(\Upsilon_{x}\right) J(\Upsilon)\right]\right)  \tag{20}\\
& =1-\frac{1}{d^{2}} \max _{x \in X}\left|\operatorname{tr}\left[U^{\dagger} U_{x}\right]\right|^{2}, \tag{21}
\end{align*}
$$

where $\mathbf{T}\left(\mathcal{H}_{1}: \mathcal{H}_{2}\right):=\left\{T \in \operatorname{Pos}\left(\mathcal{H}_{1} \otimes \mathcal{H}_{2}\right): \exists \rho \in \mathrm{S}\left(\mathcal{H}_{1}\right), T \leq \rho \otimes \mathbb{I}_{\mathcal{H}_{2}}\right\}$, and we set $T=\frac{1}{d^{2}} J(\Upsilon)\left(\leq \frac{\mathbb{I}_{\mathbb{C}^{d}}}{d} \otimes \mathbb{I}_{\mathbb{C}^{d}}\right)$ to obtain the inequality.

In Eq. (18) and Eq. (21), the same unitary operator $W=U^{\dagger} U_{x}$ appears in the term $\min _{\rho \in \mathrm{S}\left(\mathbb{C}^{d}\right)}|\operatorname{tr}[\rho W]|$ and $|\operatorname{tr}[W]|$, respectively. We can prove the second inequality in Ineq. (12) by establishing a relationship between the two terms as follows. For any unitary operator $W$ on $\mathbb{C}^{d}(d \geq 2)$,

$$
\begin{equation*}
\frac{1}{d}|\operatorname{tr}[W]|=\frac{1}{d}\left|\sum_{i=1}^{d} \lambda_{i}(W)\right| \leq \frac{2}{d} \min _{p}\left|\sum_{i=1}^{d} p(i) \lambda_{i}(W)\right|+\frac{d-2}{d}=\frac{2}{d} \min _{\rho \in \mathrm{S}\left(\mathbb{C}^{d}\right)}|\operatorname{tr}[\rho W]|+\frac{d-2}{d} \tag{22}
\end{equation*}
$$

holds, where $\lambda_{i}(W)$ is the $i$-th eigenvalue of $W$, and in the inequality, we use the following two facts: (i) the minimization is achieved only if $p$ satisfies $\forall i, p(i) \leq \frac{1}{2}$ due to a geometric observation, and (ii) for such $p$ and complex numbers $\lambda_{i} \in\{z \in \mathbb{C}:|z|=1\},\left|\sum_{i} p(i) \lambda_{i}\right| \geq\left|\sum_{i} \frac{1}{2} \lambda_{i}\right|-\left|\sum_{i}\left(\frac{1}{2}-p(i)\right) \lambda_{i}\right| \geq \frac{1}{2}\left|\sum_{i} \lambda_{i}\right|-\sum_{i}\left(\frac{1}{2}-p(i)\right)=\frac{1}{2}\left|\sum_{i} \lambda_{i}\right|-\frac{d-2}{2}$.

To the best of our knowledge, the dependence of the approximation error obtained by probabilistic synthesis on the dimension of the Hilbert space shown in this theorem has never been found. This dependence is inevitable since we can also show the sharpness of this theorem in Appendix C. More precisely, we can show that for any real number $\epsilon \in(0,1]$, any integer $d \geq 2$ and any $\Upsilon$, there exists $\left\{\Upsilon_{x}\right\}_{x \in X}$ achieving the lower bound in Ineq. (12).

In the following lemma, we show the tight upper bound showing that the worst approximation error caused by deterministic synthesis can be reduced by probabilistic synthesis at least quadratically. Our upper bound slightly improves the various existing upper bounds [6, 14], which have been proven for several classes of target unitary transformations and $\epsilon$-coverings $\left\{\Upsilon_{x}\right\}_{x \in X}$ with small $\epsilon$. Using Proposition 5.5 , shown in the next section, we can verify that our upper bound is still tight even if we consider the approximation of axial single-qubit unitary transformations.

Lemma 4.2. For a non-negative real number $\epsilon \geq 0$ and integer $d \geq 2$ specified below, if $\left\{r_{x}: \mathbf{L}\left(\mathbb{C}^{d}\right) \rightarrow \mathbf{L}\left(\mathbb{C}^{d}\right)\right\}_{x \in X}$ is a finite $\epsilon$-covering of the set of unitary transformations, i.e., $\max _{\Upsilon} \min _{x \in X} \frac{1}{2}\left\|\Upsilon-\Upsilon_{x}\right\|_{\circ} \leq \epsilon$, then

$$
\begin{equation*}
\min _{p} \frac{1}{2}\left\|\Upsilon-\sum_{x \in X} p(x) \Upsilon_{x}\right\|_{0} \leq \epsilon^{2} \tag{23}
\end{equation*}
$$

holds for any unitary transformation $\Upsilon$, where the minimization of $p$ are taken over probability distributions over $X$.

Proof. First, by using the primal problem in our SDP in Proposition 3.1, we obtain

$$
\begin{align*}
(\text { L.H.S. }) & =\max _{T \in \mathbf{T}\left(\mathbb{C}^{d}: \mathbb{C}^{d}\right)}\left(\operatorname{tr}[J(\Upsilon) T]-\max _{x \in X} \operatorname{tr}\left[J\left(\Upsilon_{x}\right) T\right]\right)  \tag{24}\\
& =\max _{\substack{\Phi \in \mathbf{P}\left(\mathbb{C}^{d} \otimes \mathcal{H}\right) \\
\Pi \in \operatorname{Proj}\left(\mathbb{C}^{d} \otimes \mathcal{H}\right)}}\left(\operatorname{tr}\left[\left(U \otimes \mathbb{I}_{\mathcal{H}}\right) \Phi\left(U \otimes \mathbb{I}_{\mathcal{H}}\right)^{\dagger} \Pi\right]-\max _{x \in X} \operatorname{tr}\left[\left(U_{x} \otimes \mathbb{I}_{\mathcal{H}}\right) \Phi\left(U_{x} \otimes \mathbb{I}_{\mathcal{H}}\right)^{\dagger} \Pi\right]\right), \tag{25}
\end{align*}
$$

where $\mathrm{T}\left(\mathcal{H}_{1}: \mathcal{H}_{2}\right):=\left\{T \in \operatorname{Pos}\left(\mathcal{H}_{1} \otimes \mathcal{H}_{2}\right): \exists \rho \in \mathrm{S}\left(\mathcal{H}_{1}\right), T \leq \rho \otimes \mathbb{I}_{\mathcal{H}_{2}}\right\}, \Upsilon(\rho)=U \rho U^{\dagger}, \Upsilon_{x}(\rho)=U_{x} \rho U_{x}^{\dagger}, \operatorname{Proj}(\mathcal{H})$ is the set of Hermitian projectors on $\mathcal{H}$, and we use Eq. (55) by taking $\Xi \in\left\{\Upsilon-\Upsilon_{x}\right\}_{x \in X}$ to obtain the last equality.

Let $\hat{\Phi}$ and $\hat{\Pi}$ maximize Eq. (25). We can verify that $\hat{\Pi} U|\hat{\Phi}\rangle=0$ if and only if there exists $x \in X$ such that $\Upsilon_{x}=\Upsilon$. If $\hat{\Pi} U|\hat{\Phi}\rangle \neq 0$, let $\hat{\Psi}$ be the pure state such that $|\hat{\Psi}\rangle \propto \hat{\Pi} U|\hat{\Phi}\rangle$. Then, we can verify that Eq. (25) is still maximized even if we replace $\hat{\Pi}$ with $\hat{\Psi}$. If $\hat{\Pi} U|\hat{\Phi}\rangle=0,\left(\exists x \in X, \Upsilon_{x}=\Upsilon\right)$ indicates that Eq. (25) is still maximized even if we replace $\hat{\Pi}$ and $\hat{\Phi}$ with an arbitrary pure state $\hat{\Psi}$ and $\left(\Upsilon^{-1} \otimes i d_{\mathcal{H}}\right)(\hat{\Psi})$, respectively. Thus, in both cases, $\Pi$ in Eq. (25) can be restricted
as a pure state, i.e., $\Pi=\Psi \in \mathrm{P}\left(\mathbb{C}^{d} \otimes \mathcal{H}\right)$, and we proceed as follows:

$$
\begin{equation*}
\text { Eq. (25) } \left.\left.=\left.\max _{\Phi, \Psi \in \mathbf{P}\left(\mathbb{C}^{d} \otimes \mathcal{H}\right)}\left(\left|\langle\Psi| U \otimes \mathbb{I}_{\mathcal{H}}\right| \Phi\right\rangle\right|^{2}-\max _{x \in X}\left|\langle\Psi| U_{x} \otimes \mathbb{I}_{\mathcal{H}}\right| \Phi\right\rangle\left.\right|^{2}\right) \tag{26}
\end{equation*}
$$

Before proceeding to the next step, we show that the set of mappings $\left.f_{\Phi, \Psi}: U \mapsto\left|\langle\Psi| U \otimes \mathbb{I}_{\mathcal{H}}\right| \Phi\right\rangle \mid$ associated with pure states $\Phi$ and $\Psi$ is equivalent to that of mappings $g_{A}: U \mapsto|\operatorname{tr}[A U]|$ associated with linear operator $A \in \mathbf{L}\left(\mathbb{C}^{d}\right)$ such that $\|A\|_{1} \leq 1$, where $\|A\|_{1}$ is the Schatten 1-norm of $A$. By using decompositions $|\Phi\rangle=\sum_{i, j} \alpha_{i j}|i\rangle|j\rangle$ and $|\Psi\rangle=\sum_{i, j} \beta_{i j}|i\rangle|j\rangle$ with respect to orthonormal bases, we can verify that $g_{A}$ with $A=\sum_{i, j, k} \alpha_{i k} \beta_{j k}^{*}|i\rangle\langle j|$ is equal to $f_{\Phi, \Psi}$ and $\|A\|_{1}=\max _{U} g_{A}(U)=\max _{U} f_{\Phi, \Psi}(U) \leq 1$. On the other hand, by using the singular value decomposition $A=\sum_{i} p_{i}\left|x_{i}\right\rangle\left\langle y_{i}\right|$, where $\|A\|_{1} \leq 1$ indicates $p+\sum_{i} p_{i}=1$ with some $p \geq 0$, we can verify that $f_{\Phi, \Psi}$ with $|\Phi\rangle=$ $\sqrt{p}|0\rangle|\perp\rangle+\sum_{i} \sqrt{p_{i}}\left|x_{i}\right\rangle|i\rangle$ and $|\Psi\rangle=\sqrt{p}|0\rangle\left|\perp^{\prime}\right\rangle+\sum_{i} \sqrt{p_{i}}\left|y_{i}\right\rangle|i\rangle\left(\{|i\rangle\}_{i} \cup\left\{|\perp\rangle,\left|\perp^{\prime}\right\rangle\right\}\right.$ is an orthonormal basis) is equal to $g_{A}$.

By using the equivalent between two sets of mappings, we proceed as follows:

$$
\begin{equation*}
\text { Eq. (26) }=\max _{A:\|A\|_{1} \leq 1}\left(|\operatorname{tr}[A U]|^{2}-\max _{x \in X}\left|\operatorname{tr}\left[A U_{x}\right]\right|^{2}\right)=\max _{V, \rho \in \mathrm{~S}\left(\mathbb{C}^{d}\right)}\left(\left|\operatorname{tr}\left[\rho V^{\dagger} U\right]\right|^{2}-\max _{x \in X}\left|\operatorname{tr}\left[\rho V^{\dagger} U_{x}\right]\right|^{2}\right) \tag{27}
\end{equation*}
$$

where we use the fact that the maximization is achieved when $\|A\|_{1}=1$ and use the polar decomposition $A=\rho V^{\dagger}$ with a unitary operator $V$ acting on $\mathbb{C}^{d}$.

By using Eq. (27), we obtain

$$
\begin{align*}
\max _{\Upsilon} \min _{p} \frac{1}{2}\left\|\Upsilon-\sum_{x \in X} p(x) \Upsilon_{x}\right\|_{\diamond} & =\max _{V, \rho \in \mathrm{~S}\left(\mathbb{C}^{d}\right)}\left(\max _{U}\left|\operatorname{tr}\left[\rho V^{\dagger} U\right]\right|^{2}-\max _{x \in X}\left|\operatorname{tr}\left[\rho V^{\dagger} U_{x}\right]\right|^{2}\right)  \tag{28}\\
& =1-\min _{V, \rho \in \mathrm{~S}\left(\mathbb{C}^{d}\right)} \max _{x \in X}\left|\operatorname{tr}\left[\rho V^{\dagger} U_{x}\right]\right|^{2}  \tag{29}\\
& \leq 1-\min _{V} \max _{x \in X} \min _{\rho \in \mathrm{S}\left(\mathbb{C}^{d}\right)}\left|\operatorname{tr}\left[\rho V^{\dagger} U_{x}\right]\right|^{2}, \tag{30}
\end{align*}
$$

where the maximization of $\Upsilon$ is taken over unitary transformations, and we use the fact that $\max _{x} \min _{y} f(x, y) \leq$ $\min _{y} \max _{x} f(x, y)$ for any $f$ if the maximum and minimum exist in the inequality. Using Eq. (17) completes the proof.

The combination of Lemmas 4.1 and 4.2 can be summarized as the following theorem.
Theorem 4.3. For an integer $d \geq 2$ specified below, let $\Upsilon: \mathbf{L}\left(\mathbb{C}^{d}\right) \rightarrow \mathbf{L}\left(\mathbb{C}^{d}\right)$ and $\left\{\Upsilon_{x}: \mathbf{L}\left(\mathbb{C}^{d}\right) \rightarrow \mathbf{L}\left(\mathbb{C}^{d}\right)\right\}_{x \in X}$ be a target unitary transformation and finite set of unitary transformations, respectively. Then,

$$
\frac{4 \delta_{\Upsilon}}{d}\left(1-\frac{\delta_{\Upsilon}}{d}\right) \leq \min _{p} \frac{1}{2}\left\|\Upsilon-\sum_{x \in X} p(x) \Upsilon_{x}\right\|_{\diamond} \leq \epsilon^{2} \text { with }\left\{\begin{array}{l}
\delta_{\Upsilon}=1-\sqrt{1-\epsilon_{\Upsilon}^{2}}  \tag{31}\\
\epsilon_{\Upsilon}=\min _{x \in X} \frac{1}{2}\left\|\Upsilon-\Upsilon_{x}\right\|_{\diamond} \\
\epsilon=\max _{\Upsilon} \min _{x \in X} \frac{1}{2}\left\|\Upsilon-\Upsilon_{x}\right\|_{\diamond}
\end{array}\right.
$$

holds, where the maximization of $\Upsilon$ and minimization of $p$ are taken over unitary transformations on $\mathbf{L}\left(\mathbb{C}^{d}\right)$ and probability distributions over $X$, respectively.

By maximizing $\Upsilon$ over all the unitary transformations, we obtain Ineq. (1) as a simplified version of this theorem. As mentioned in the introduction, in Appendix C, we show that both the upper and lower bounds in Ineq. (1) are tight, i.e., for any real number $\epsilon \in(0,1]$ and any integer $d \geq 2$, the two bounds are achievable for some $\left\{\Upsilon_{\vec{i}}\right\}_{\vec{i}}$.

## 5 PROBABILISTIC SYNTHESIS FOR SINGLE-QUBIT UNITARY TRANSFORMATION

In this section, we construct a simplified SDP that computes the optimal mixing probability for single-qubit-unitary synthesis. Before discussing that, we first show the special properties of the probabilistic mixture of single-qubit unitaries. In the first subsection, we prove Lemma 5.3, which is a crucial ingredient for constructing the SDP and has a direct application to constructing an efficient probabilistic synthesis algorithm. In the second subsection, we investigate the approximation of single-qubit unitary transformations corresponding to axial rotations to provide a geometric interpretation of the quadratic improvement owing to the probabilistic mixture and confirmation of Lemma 5.3.

We show the first special property of a single-qubit unitary operator in the following Lemma, which essentially shows the equivalence between the set of maximally entangled two-qubit states and a real subspace in the two qubits.

Lemma 5.1. For any finite set $\left\{\Phi_{x} \in \mathbf{P}\left(\mathbb{C}^{2} \otimes \mathbb{C}^{2}\right)\right\}_{x \in X}$ of maximally entangled states and any real numbers $\left\{r_{x} \in\right.$ $\mathbb{R}\}_{x \in X}$, the Hermitian operator $H=\sum_{x \in X} r_{x} \Phi_{x}$ is diagonalizable with respect to maximally entangled eigenstates.

Proof. First, we show the equivalence between the set of two-qubit maximally entangled vectors and a real subspace in the two qubits. Define four vectors representing maximally entangled states:

$$
\begin{array}{ll}
\left|\Psi_{1}\right\rangle=\frac{1}{\sqrt{2}}(|00\rangle+|11\rangle), & \left|\Psi_{2}\right\rangle=\frac{i}{\sqrt{2}}(|00\rangle-|11\rangle), \\
\left|\Psi_{3}\right\rangle=\frac{i}{\sqrt{2}}(|01\rangle+|10\rangle), & \left|\Psi_{4}\right\rangle=\frac{1}{\sqrt{2}}(|01\rangle-|10\rangle) . \tag{32}
\end{array}
$$

Any vector in the real subspace $\mathcal{K}_{M E S}$ spanned by $\left\{\left|\Psi_{i}\right\rangle\right\}_{i=1}^{4}$ can be represented by

$$
\begin{equation*}
\frac{1}{\sqrt{2}}\left(\left(u_{1}+i u_{2}\right)|00\rangle+\left(u_{4}+i u_{3}\right)|01\rangle-\left(u_{4}-i u_{3}\right)|10\rangle+\left(u_{1}-i u_{2}\right)|11\rangle\right) \tag{33}
\end{equation*}
$$

with real numbers $\left\{u_{i} \in \mathbb{R}\right\}_{i=1}^{4}$. On the other hand, any maximally entangled state can be obtained by applying the single-qubit unitary operator represented by $\left(\begin{array}{cc}e^{i \phi_{1}} \cos \theta & e^{i \phi_{2}} \sin \theta \\ -e^{-i \phi_{2}} \sin \theta & e^{-i \phi_{1}} \cos \theta\end{array}\right)$ to $\left|\Psi_{1}\right\rangle$ and can be represented by a vector

$$
\begin{equation*}
\frac{1}{\sqrt{2}}\left(e^{i \phi_{1}} \cos \theta|00\rangle+e^{i \phi_{2}} \sin \theta|01\rangle-e^{-i \phi_{2}} \sin \theta|10\rangle+e^{-i \phi_{1}} \cos \theta|11\rangle\right) . \tag{34}
\end{equation*}
$$

By comparing Eqs. (33) and (34), we can verify that any two-qubit maximally entangled state can be represented as a unit vector in $\mathcal{K}_{M E S}$ and any unit vector in $\mathcal{K}_{M E S}$ represents a maximally entangled state. This equivalence has been indicated in a previous study [2], and the basis defined in Eq. (32) is called the magic basis [15].

Since $H=\sum_{x \in X} r_{x} \Phi_{x}$ is represented as a real symmetric matrix with respect to the basis $\left\{\left|\Psi_{i}\right\rangle\right\}_{i=1}^{4}, H$ is diagonalizable with respect to real eigenvectors, which represents maximally entangled states.

Next, we show a special property of the diamond norm between probabilistic mixtures of single-qubit unitaries in the following Lemma, which essentially shows that the input state in the definition of the diamond norm can be maximally entangled.

Lemma 5.2. For a subset $\left\{\Upsilon_{x}: \mathbf{L}\left(\mathbb{C}^{2}\right) \rightarrow \mathbf{L}\left(\mathbb{C}^{2}\right)\right\}_{x \in X}$ of single-qubit unitary transformations and probability distributions $p$ and $q$ over a finite set $X$, it holds that

$$
\begin{equation*}
\left\|\sum_{x \in X} p(x) \Upsilon_{x}-\sum_{x \in X} q(x) \Upsilon_{x}\right\|_{\diamond}=\left\|\sum_{x \in X}(p(x)-q(x)) J\left(\Upsilon_{x}\right)\right\|_{t r} . \tag{35}
\end{equation*}
$$

Proof. For $d(\geq 2)$-dimensional CPTP maps $\left\{\Upsilon_{x}\right\}_{x \in X}$, it holds that

$$
\begin{equation*}
(\text { L.H.S. })=\max _{\Phi \in \mathbf{P}\left(\mathbb{C}^{d} \otimes \mathbb{C}^{d}\right)} 2\left\|\sum_{x \in X}(p(x)-q(x)) \Upsilon_{x} \otimes i d_{\mathbb{C}^{d}}(\Phi)\right\|_{\mathrm{tr}} \geq \frac{2}{d}\left\|\sum_{x \in X}(p(x)-q(x)) J\left(\Upsilon_{x}\right)\right\|_{\mathrm{tr}} \tag{36}
\end{equation*}
$$

On the other hand, by using the dual problem of the SDP to compute the diamond norm used in the proof of Proposition 3.1, we obtain

$$
\begin{equation*}
(\text { L.H.S. }) \leq 2\left\|\operatorname{tr}_{2}[S]\right\|_{\infty} \text { with }(S \geq 0) \wedge\left(S \geq \sum_{x \in X}(p(x)-q(x)) J\left(\Upsilon_{x}\right)\right) \tag{37}
\end{equation*}
$$

where $\operatorname{tr}_{2}[\cdot]$ represents the partial trace of the second system of $\mathbb{C}^{2} \otimes \mathbb{C}^{2}$. By using Lemma 5.1, we can verify that $\sum_{x \in X}(p(x)-q(x)) J\left(\Upsilon_{x}\right)=\sum_{i=1}^{4} \lambda_{i} \Phi_{i}$ with real numbers $\lambda_{i}$ and a set of orthogonal maximally entangled states $\left\{\Phi_{i}\right\}_{i=1}^{4}$. By setting $S=\sum_{i: \lambda_{i}>0} \lambda_{i} \Phi_{i}$, we obtain

$$
\begin{equation*}
2\left\|\operatorname{tr}_{2}[S]\right\|_{\infty}=2\left\|\sum_{i: \lambda_{i}>0} \lambda_{i} \frac{\mathbb{I}}{2}\right\|_{\infty}=\sum_{i: \lambda_{i}>0} \lambda_{i}=(\text { R.H.S. }) \tag{38}
\end{equation*}
$$

This completes the proof.

### 5.1 Support of optimal probability distribution

To achieve the quadratic improvement owing to the probabilistic approximation of $\Upsilon$ by using $\left\{\Upsilon_{x}\right\}_{x \in X}$, we assume $\left\{\Upsilon_{x}\right\}_{x \in X}$ is an $\epsilon$-covering of the set of unitary transformations in Lemma 4.2. Since $|X|=\Omega\left(\frac{1}{\epsilon^{c}}\right)$ from a volume consideration, the runtime poly $\left(|X| \log \left(\frac{1}{\epsilon}\right)\right)$ of our SDP to compute the optimal probability distribution proposed in Proposition 3.1 increases as poly $\left(\frac{1}{\epsilon}\right)$ at best. However, by using the following lemma, we can construct a much more efficient SDP.

Lemma 5.3. For a non-negative real number $\epsilon \geq 0$, if $\Upsilon$ is a single-qubit unitary transformation and $\left\{\Upsilon_{x}: \mathbf{L}\left(\mathbb{C}^{2}\right) \rightarrow \mathbf{L}\left(\mathbb{C}^{2}\right)\right\}_{x \in X}$ is a finite $\epsilon$-covering of the set of single-qubit unitary transformations, i.e., $\max _{\Upsilon} \min _{x \in X} \frac{1}{2}\left\|\Upsilon-\Upsilon_{x}\right\|_{\diamond} \leq \epsilon$, then

$$
\begin{equation*}
\min _{p}\left\|\Upsilon-\sum_{x \in X} p(x) \Upsilon_{x}\right\|_{\diamond}=\min _{\hat{p}}\left\|\Upsilon-\sum_{x \in \hat{X}} \hat{p}(x) \Upsilon_{x}\right\|_{\diamond} \tag{39}
\end{equation*}
$$

holds, where $\hat{X}:=\left\{x \in X: \frac{1}{2}\left\|\Upsilon-\Upsilon_{x}\right\|_{\diamond} \leq 2 \epsilon\right\}$ and the minimization of $p$ and $\hat{p}$ are taken over probability distributions over $X$ and those over $\hat{X}$, respectively.

Proof. By using Lemma 5.2, we obtain

$$
\begin{equation*}
(\text { L.H.S. })=\min _{p}\left\|J(\Upsilon)-\sum_{x \in X} p(x) J\left(\Upsilon_{x}\right)\right\|_{\operatorname{tr}}=\min _{p}\left\|J(\Upsilon)-\sum_{x \in X} p(x) J\left(\Upsilon_{x}\right)\right\|_{\infty} \tag{40}
\end{equation*}
$$

where we use the dimension of the eigenspace of $J(\Upsilon)-\sum_{x \in X} p(x) J\left(\Upsilon_{x}\right)$ with positive eigenvalues is at most 1 in the last equality. By using Lemma 5.1, we can proceed with the following two ways:

$$
\begin{align*}
& \text { Eq. (40) }=\min _{p} \max _{\rho \in \operatorname{conv}(\mathrm{MES})} \operatorname{tr}\left[\rho\left(J(\Upsilon)-\sum_{x \in X} p(x) J\left(\Upsilon_{x}\right)\right)\right] \text { and }  \tag{41}\\
& \text { Eq. (40) }=\min _{p} \max _{\substack{M \in \operatorname{cone}(\mathrm{MES}) \\
M \leq \mathbb{I}}} \operatorname{tr}\left[M\left(J(\Upsilon)-\sum_{x \in X} p(x) J\left(\Upsilon_{x}\right)\right)\right], \tag{42}
\end{align*}
$$

where conv (MES) and cone(MES) are the convex hull of the set of maximally entangled states $\left\{\Phi \in \mathbf{P}\left(\mathbb{C}^{2} \otimes \mathbb{C}^{2}\right): \operatorname{tr}_{2}[\Phi]=\frac{\mathbb{I}}{2}\right\}$ and the convex cone generated by the set $\{\Phi\}$, respectively. Note that the convex cone generated by a subset X in a vector space is defined as the set of finite linear combinations of X with non-negative coefficients.

Since the domains of $p, \rho$, and $M$ are compact and convex and $f(p, H):=\operatorname{tr}\left[H\left(J(\Upsilon)-\sum_{x \in X} p(x) J\left(\Upsilon_{x}\right)\right)\right]$ is affine with respect to each variable, we can apply the minimax theorem and obtain

$$
\begin{equation*}
\text { Eq. (40) }=\max _{\rho \in \operatorname{conv}(\mathrm{MES})}\left(\operatorname{tr}[\rho J(\Upsilon)]-\max _{x \in X} \operatorname{tr}\left[\rho J\left(\Upsilon_{x}\right)\right]\right)=\max _{\substack{M \in \operatorname{cone(MES)} \\ M \leq \mathbb{I}}}\left(\operatorname{tr}[M J(\Upsilon)]-\max _{x \in X} \operatorname{tr}\left[M J\left(\Upsilon_{x}\right)\right]\right) . \tag{43}
\end{equation*}
$$

When (L.H.S.) $=0$, the theorem holds since there exists $x \in X$ such that $\Upsilon_{x}=\Upsilon$. In the following, we assume (L.H.S.) $>0$. If $\rho$ with $\|\rho\|_{\infty}<1$ maximizes the formula, we can show a contradiction by setting $M=\frac{\rho}{\|\rho\|_{\infty}}$. Thus, $\rho$ that maximizes the formula satisfies $\|\rho\|_{\infty}=1$, i.e., $\rho$ is a (pure) maximally entangled state. Therefore, we obtain

$$
\begin{equation*}
\text { Eq. (43) }=\max _{\Upsilon^{\prime}} \frac{1}{2}\left(\operatorname{tr}\left[J\left(\Upsilon^{\prime}\right) J(\Upsilon)\right]-\max _{x \in X} \operatorname{tr}\left[J\left(\Upsilon^{\prime}\right) J\left(\Upsilon_{x}\right)\right]\right)=\max _{U^{\prime} \in U(2)} \frac{1}{2}\left(\left|\operatorname{tr}\left[U^{\dagger} U^{\prime}\right]\right|^{2}-\max _{x \in X}\left|\operatorname{tr}\left[U_{x}^{\dagger} U^{\prime}\right]\right|^{2}\right), \tag{44}
\end{equation*}
$$

where $\Upsilon(\rho)=U \rho U^{\dagger}, \Upsilon_{x}(\rho)=U_{x} \rho U_{x}^{\dagger}$, the maximization of $\Upsilon^{\prime}$ is taken over single-qubit unitary transformations, and $U(2)$ represents the set of single-qubit unitary operators. By observing that the minimization in Eq. (17) is achieved by $\rho=\frac{\mathbb{I}}{2}$ for single-qubit unitary operators, we obtain

$$
\begin{equation*}
\text { Eq. (44) }=\max _{\Upsilon^{\prime}} \frac{1}{2}\left(\min _{x \in X}\left\|\Upsilon^{\prime}-\Upsilon_{x}\right\|_{\diamond}^{2}-\left\|\Upsilon^{\prime}-\Upsilon\right\|_{\diamond}^{2}\right) . \tag{45}
\end{equation*}
$$

Since so far we did not use the assumption that $\left\{\Upsilon_{x}\right\}_{x \in X}$ is an $\epsilon$-covering, we obtain

$$
\begin{equation*}
\text { (R.H.S.) of Eq. (39) }=\max _{\Upsilon^{\prime}} \frac{1}{2}\left(\min _{x \in \hat{X}}\left\|\Upsilon^{\prime}-\Upsilon_{x}\right\|_{\Delta}^{2}-\left\|\Upsilon^{\prime}-\Upsilon\right\|_{\Delta}^{2}\right) . \tag{46}
\end{equation*}
$$

Note that the maximization in Eq. (45) is achieved by $\Upsilon^{\prime}$ satisfying $\frac{1}{2}\left\|\Upsilon^{\prime}-\Upsilon\right\|_{\diamond} \leq \epsilon \operatorname{since} \min _{x \in X} \frac{1}{2}\left\|\Upsilon^{\prime}-\Upsilon_{x}\right\|_{\diamond} \leq \epsilon$ due to the definition of the $\epsilon$-covering. If we can show that the maximization in Eq. (46) is also achieved by such $\Upsilon^{\prime}$, we can prove the equivalence between Eqs. (45) and (46). For the minimization in Eq. (45) is achieved by $x \in \hat{X}$ owing to the triangle inequality. To complete the proof, we show the following statement: for all $\Upsilon^{\prime}$,

$$
\begin{equation*}
\frac{1}{2}\left\|\Upsilon^{\prime}-\Upsilon\right\|_{\diamond}>\epsilon \Rightarrow \min _{x \in \hat{X}}\left\|\Upsilon^{\prime}-\Upsilon_{x}\right\|_{\diamond} \leq\left\|\Upsilon^{\prime}-\Upsilon\right\|_{\diamond} \tag{47}
\end{equation*}
$$

We assume $\epsilon<1$; otherwise, the statement is trivial. By using the equivalence between the set of two-qubit maximally entangled vectors and a real subspace shown in the proof of Lemma 5.1, there exist unit real vectors $\vec{u}, \vec{u}^{\prime} \in \mathbb{R}^{4}$ such that $\sum_{i, j=1}^{4} u_{i} u_{j}\left|\Psi_{i}\right\rangle\left\langle\Psi_{j}\right|=\frac{1}{2} J(\Upsilon), \sum_{i, j=1}^{4} u_{i}^{\prime} u_{j}^{\prime}\left|\Psi_{i}\right\rangle\left\langle\Psi_{j}\right|=\frac{1}{2} J\left(\Upsilon^{\prime}\right)$ and

$$
\begin{equation*}
0 \leq \cos \theta_{1}:=\vec{u} \cdot \vec{u}^{\prime}<\sqrt{1-\epsilon^{2}}, \tag{48}
\end{equation*}
$$

where $\left\{\left|\Psi_{i}\right\rangle\right\}$ is defined in Eq. (32), $\theta_{1} \in\left[0, \frac{\pi}{2}\right]$, the first inequality can be satisfied by appropriately setting the sign of $\vec{u}$, and the second (strict) inequality is derived from $\frac{1}{2}\left\|\Upsilon^{\prime}-\Upsilon\right\|_{\diamond}>\epsilon$ and Lemma 5.2. In the real subspace spanned by $\left\{\vec{u}, \vec{u}^{\prime}\right\}$, there exists a unique unit real vector $\vec{v} \in \mathbb{R}^{4}$ such that

$$
\begin{equation*}
\cos \theta_{2}:=\vec{u} \cdot \vec{v}=\sqrt{1-\epsilon^{2}} \wedge \vec{u}^{\prime} \cdot \vec{v}=\cos \left(\theta_{1}-\theta_{2}\right) \tag{49}
\end{equation*}
$$

where $\theta_{2} \in\left[0, \frac{\pi}{2}\right]$, as shown in Fig. 2. Note that the unitary transformation $\hat{\Upsilon}$ corresponding to $\vec{v}$, i.e., $\sum_{i, j=1}^{4} v_{i} v_{j}\left|\Psi_{i}\right\rangle\left\langle\Psi_{j}\right|=$ $\frac{1}{2} J(\hat{\Upsilon})$, satisfies $\frac{1}{2}\|\Upsilon-\hat{\Upsilon}\|_{\diamond}=\epsilon$ due to Lemma 5.2. Since there exists $x \in X$ such that $\frac{1}{2}\left\|\Upsilon_{x}-\hat{\Upsilon}\right\|_{\diamond} \leq \epsilon$ and $\frac{1}{2}\left\|\Upsilon_{x}-\Upsilon\right\|_{\diamond} \leq$ $\frac{1}{2}\left\|\Upsilon_{x}-\hat{\Upsilon}\right\|_{\diamond}+\frac{1}{2}\|\Upsilon-\hat{\Upsilon}\|_{\diamond} \leq 2 \epsilon$, we can find a unit real vector $\vec{w} \in \mathbb{R}^{4}$ corresponding to $\Upsilon_{x}$ with $x \in \hat{X}$, i.e., $\sum_{i, j=1}^{4} w_{i} w_{j}\left|\Psi_{i}\right\rangle\left\langle\Psi_{j}\right|=$ $\frac{1}{2} J\left(\Upsilon_{x}\right)$, and satisfying

$$
\begin{equation*}
\cos \theta_{3}:=\vec{w} \cdot \vec{v} \geq \sqrt{1-\epsilon^{2}} \tag{50}
\end{equation*}
$$

where $\theta_{3} \in\left[0, \frac{\pi}{2}\right]$, due to Lemma 5.2. By using Lemma 5.2 again, we obtain

$$
\begin{equation*}
\left\|\Upsilon^{\prime}-\Upsilon_{x}\right\|_{\Delta} \leq\left\|\Upsilon^{\prime}-\Upsilon\right\|_{\Delta} \Leftrightarrow\left|\vec{u}^{\prime} \cdot \vec{u}\right| \leq\left|\vec{u}^{\prime} \cdot \vec{w}\right| . \tag{51}
\end{equation*}
$$

By letting $\cos \theta_{4}:=\vec{u}^{\prime} \cdot \vec{w}$ with $\theta_{4} \in[0, \pi]$ and using the triangle inequality for angles in the three-dimensional subspace spanned by $\left\{\vec{u}, \vec{u}^{\prime}, \vec{w}\right\}$, we obtain

$$
\begin{equation*}
\theta_{4} \leq\left(\theta_{1}-\theta_{2}\right)+\theta_{3} \leq \theta_{1} \tag{52}
\end{equation*}
$$

This completes the proof.


Fig. 2. Three-dimensional subspace spanned by $\left\{\vec{u}, \vec{u}^{\prime}, \vec{w}, \vec{v}\right\}$ in the proof of Lemma 5.3 , where $\vec{v} \in \operatorname{span}\left(\left\{\vec{u}, \vec{u}^{\prime}\right\}\right)$. We apply the triangle inequality for the angle $\theta_{4}$ between $\vec{u}^{\prime}$ and $\vec{w}$, the angle $\theta_{3}$ between $\vec{v}$ and $\vec{w}$ and the angle $\left(\theta_{1}-\theta_{2}\right)$ between $\vec{v}$ and $\vec{u}^{\prime}$.

As an application of Lemma 5.3, we construct an efficient probabilistic synthesis algorithm in the proof of the following theorem.

Theorem 5.4. For a given gate set, there exists a probabilistic synthesis algorithm for a single-qubit unitary transformation with

INPUT: a single-qubit unitary transformation $\Upsilon: \mathbf{L}\left(\mathbb{C}^{2}\right) \rightarrow \mathbf{L}\left(\mathbb{C}^{2}\right)$, an approximation error $\epsilon \in(0,1)$, and precision $\delta>0$ such that $\frac{1}{\delta}=\left(\frac{1}{\epsilon}\right)^{O(1)}$

OUTPUT: a gate sequence for implementing a single-qubit unitary transformation $\Upsilon_{x}$ sampled from a set $\left\{\Upsilon_{x}: \mathbf{L}\left(\mathbb{C}^{2}\right) \rightarrow \mathbf{L}\left(\mathbb{C}^{2}\right)\right\}_{x \in \hat{X}}$ in accordance with probability distribution $\hat{p}(x)$ such that the algorithm satisfies the following properties:

- Efficiency: All steps of the algorithm take polylog $\left(\frac{1}{\epsilon}\right)$-time,
- Quadratic improvement: The approximation error $\frac{1}{2}\left\|\Upsilon-\sum_{x \in \hat{X}} \hat{p}(x) \Upsilon_{x}\right\|_{0}$ obtained by this algorithm is upper bounded by $\epsilon^{2}+\delta$, whereas the error $\min _{x \in \hat{X}} \frac{1}{2}\left\|\Upsilon-\Upsilon_{x}\right\|_{\diamond}$ obtained by deterministic synthesis using the unitary transformations in $\left\{\Upsilon_{x}\right\}_{x \in \hat{X}}$ is upper bounded by $\epsilon$,

Proof. We assume that the algorithm calls an efficient deterministic synthesis algorithm such as the Solovay-Kitaev algorithm as a subroutine, i.e., the subroutine can find a gate sequence for implementing a unitary transformation $\Upsilon^{\prime}$ such that $\frac{1}{2}\left\|\Upsilon-\Upsilon^{\prime}\right\|_{\diamond} \leq \epsilon$ within polylog $\left(\frac{1}{\epsilon}\right)$-time. In the following, we explicitly construct the algorithm:

## Efficient probabilistic synthesis algorithm for single-qubit unitary transformation

(1) Set free parameters $c>0$ and $c^{\prime}>0$ satisfying $c+c^{\prime} \leq 1$.
(2) Generate a list $\left\{\hat{\Upsilon}_{x}\right\}_{x \in \hat{X}}$ of single-qubit unitary transformations such that for any unitary transformation $\hat{\Upsilon}$, $\min _{x \in \hat{X}} \frac{1}{2}\left\|\hat{\Upsilon}-\hat{\Upsilon}_{x}\right\|_{\diamond} \leq c \epsilon$ if $\frac{1}{2}\|\Upsilon-\hat{\Upsilon}\|_{\diamond} \leq 2 \epsilon$. That is, $\left\{\hat{r}_{x}\right\}_{x \in \hat{X}}$ is a $c \epsilon$-covering of the $2 \epsilon$-ball around the target unitary transformation.
(3) Call an efficient deterministic synthesis algorithm to find gate sequences for implementing unitary transformations $\left\{\Upsilon_{x}\right\}_{x \in \hat{X}}$ such that $\frac{1}{2}\left\|\Upsilon_{x}-\hat{\Upsilon}_{x}\right\|_{\nabla} \leq c^{\prime} \epsilon$ for all $x \in \hat{X}$.
(4) Numerically solve our SDP shown in Proposition 3.1 by using $\left\{\Upsilon_{x}\right\}_{x \in \hat{X}}$ as a set of CPTP mappings and obtain a probability distribution $\hat{p}$, which causes the approximation error $\delta$-close to $\min _{p} \frac{1}{2}\left\|\Upsilon-\sum_{x \in \hat{X}} p(x) \Upsilon_{x}\right\|_{\diamond}$.
(5) Sample gate sequences for implementing unitary transformations $\left\{\Upsilon_{x}\right\}_{x \in \hat{X}}$ in accordance with $\hat{p}$.

The two properties can be verified as follows:

- Efficiency: All steps of the algorithm take polylog $\left(\frac{1}{\epsilon}\right)$-time if the size $\hat{X}$ of the list generated in the second step is upper bounded by a constant (independent to $\epsilon$.) We can generate such a constant-size list $\left\{\hat{\Upsilon}_{x}\right\}_{x \in \hat{X}}$ by using the correspondence between a single-qubit unitary operator and unit vector in $\mathbb{R}^{4}$ and Lemma 5.2.
- Quadratic improvement: The approximation error $\frac{1}{2}\left\|\Upsilon-\sum_{x \in \hat{X}} \hat{p}(x) \Upsilon_{x}\right\|_{\diamond}$ obtained by this algorithm is at least $\epsilon^{2}+\delta$ since $\left\{\Upsilon_{x}\right\}_{x \in \hat{X}}$ is a subset of an $\epsilon$-covering $\left\{\Upsilon_{x}\right\}_{x \in \hat{X}} \cup\left\{\Upsilon_{y}^{\prime}\right\}_{y}$ of the set of single-qubit unitary transformations, where $\left\{\Upsilon_{y}^{\prime}\right\}_{y}$ is an $\epsilon$-covering of the complement of the $2 \epsilon$-ball around $\Upsilon$ and $\frac{1}{2}\left\|\Upsilon-\Upsilon_{y}^{\prime}\right\|_{\diamond}>2 \epsilon$ for any $y$, and we can apply Lemmas 4.2 and 5.3.

Note that the quadratic improvement on the approximation error achieved by this algorithm heavily relies on Lemma 5.3. In Appendix D, we perform numerical experiments to confirm that this lemma would hold for qudit unitary transformations, which implies that this synthesis algorithm is applicable to qudit unitary transformations.

### 5.2 Convex-hull approximation for axial rotations

At a glance, the reduction of the approximation error due to probabilistically mixing unitaries seems strange since a unitary transformation is not a probabilistic mixture of any distinct unitary transformations. A simple geometric interpretation of the reduction is given in the following theorem, considering single-qubit unitary transformations corresponding to axial rotations.

We investigate the convex-hull approximation of a single-qubit unitary transformation $\Upsilon_{\hat{\theta}}$ by using unitary transformations $\left\{\Upsilon_{\theta}\right\}_{\theta \in \Theta}$ that rotate Bloch vectors about the same axes as $\Upsilon_{\hat{\theta}}$, where $\Upsilon_{\theta}(\rho):=R(\theta) \rho R^{\dagger}(\theta), R(\theta):=$ $|0\rangle\langle 0|+e^{i \theta}|1\rangle\langle 1|$ with an orthonormal basis $\{|0\rangle,|1\rangle\}$, and $\boldsymbol{\Theta}$ is a finite subset of $[0,2 \pi)$. In this case, every unitary transformation $\Upsilon_{\theta}$ can be represented by a unit complex number $e^{i \theta}$ in the complex plane, as shown in Fig. 3. Furthermore, the following proposition shows that the metric space of probabilistic mixtures of $\Upsilon_{\theta}$ induced by the diamond norm can be identified with a unit disc in the complex plane.

Proposition 5.5. For a finite subset $\boldsymbol{\Theta}$ of $[0,2 \pi)$, let $\left\{\Upsilon_{\theta}\right\}_{\theta \in \Theta}$ be a set of single-qubit unitary transformations that rotate Bloch vectors about a fixed axis, i.e., $\Upsilon_{\theta}(\rho):=R(\theta) \rho R^{\dagger}(\theta)$ with $R(\theta):=|0\rangle\langle 0|+e^{i \theta}|1\rangle\langle 1|$ and an orthonormal basis $\{|0\rangle,|1\rangle\}$. For probability distributions $p$ and $q$ over $\Theta$, it holds that

$$
\begin{equation*}
\left\|\sum_{\theta \in \boldsymbol{\Theta}} p(\theta) \Upsilon_{\theta}-\sum_{\theta \in \boldsymbol{\Theta}} q(\theta) \Upsilon_{\theta}\right\|_{\diamond}=\left|\sum_{\theta \in \boldsymbol{\Theta}} p(\theta) e^{i \theta}-\sum_{\theta \in \boldsymbol{\Theta}} q(\theta) e^{i \theta}\right| \tag{53}
\end{equation*}
$$

Proof. By using Lemma 5.2, we obtain

$$
\begin{equation*}
(\text { L.H.S. })=\left\|\sum_{\theta \in \Theta}(p(\theta)-q(\theta)) J\left(\Upsilon_{\theta}\right)\right\|_{\mathrm{tr}}=(\text { R.H.S. }), \tag{54}
\end{equation*}
$$

where we use the diagonalization of $\sum_{\theta \in \Theta}(p(\theta)-q(\theta)) J\left(\Upsilon_{\theta}\right)$, which can be obtained via a straightforward calculation, in the last equality.

By using this proposition, we can obtain $\frac{1}{2}\left\|\mathrm{r}_{\hat{\theta}}-\sum_{\theta \in \boldsymbol{\Theta}} p(\theta) \Upsilon_{\theta}\right\|_{\Delta}=\frac{1}{2}\left|e^{i \hat{\theta}}-\sum_{\theta \in \boldsymbol{\Theta}} p(\theta) e^{i \theta}\right|$, which indicates that the optimal probability distribution and approximation error in the convex-hull approximation of $\Upsilon_{\hat{\theta}}$ can be computed by finding the closest point in the convex hull of $\left\{e^{i \theta}\right\}_{\theta \in \Theta}$ to the target point $e^{i \hat{\theta}}$. As represented in Fig. 3, the quadratic reduction in approximation error owing to convex-hull approximation over discrete-point approximation can be shown by an elementary geometric observation.


Fig. 3. Corresponding complex numbers to target unitary transformation $\Upsilon_{\frac{\pi}{2}}$ and unitary transformations $\left\{\Upsilon_{\theta}\right\}_{\theta \in \Theta}$ with $\Theta=$ $\left\{0, \frac{\pi}{3}, \frac{2 \pi}{3}, \pi, \frac{4 \pi}{3}, \frac{5 \pi}{3}\right\}$. Convex hull of $\left\{\Upsilon_{\theta}\right\}_{\theta \in \Theta}$ corresponds to shaded region. If we let approximation error obtained by deterministic approximation be $\epsilon=\min _{\theta \in \Theta} \frac{1}{2}\left\|\Upsilon_{\frac{\pi}{2}}-\Upsilon_{\theta}\right\|_{\diamond}=\min _{\theta \in \Theta} \frac{1}{2}\left|i-e^{i \theta}\right|$, that obtained by probabilistic approximation is given by $\epsilon^{2}=$ $\min _{p} \frac{1}{2}\left\|\Upsilon_{\frac{\pi}{2}}-\sum_{\theta \in \Theta} p(\theta) \Upsilon_{\theta}\right\|_{\diamond}=\min _{p} \frac{1}{2}\left|i-\sum_{\theta \in \Theta} p(\theta) e^{i \theta}\right|$, which demonstrates quadratic reduction in error.

## 6 CONCLUSION

We considered the analytical relationship between $\min _{p}\left\|\Upsilon-\sum_{x} p(x) \Upsilon_{x}\right\|_{\diamond}$ and $\min _{x}\left\|\Upsilon-\Upsilon_{x}\right\|_{\diamond}$, which represent the minimum approximation error obtained by probabilistic synthesis and that by deterministic synthesis, respectively. As the main result, we obtained tight upper and lower bounds on $\min _{p}\left\|\Upsilon-\sum_{x} p(x) \Upsilon_{x}\right\|_{\odot}$, which guarantees the suboptimality of the current algorithms as well as suggests the existence of an improved synthesis algorithm. We showed that the optimal probability distribution in the approximation can be computed by an SDP. We also constructed an efficient probabilistic synthesis algorithm for single-qubit unitary transformations and showed that it quadratically reduces approximation error compared with deterministic synthesis and its optimality can be reduced into the choice of unitary transformations close to the target unitary one. While numerical simulations indicate the algorithm works well for qudit unitary transformations, a rigorous proof is a subject for future work.

When we run this algorithm for qudit unitary transformation, the time complexity of the SDP used in the algorithm becomes poly $(d,|\hat{X}|)$ for a fixed desired approximation error $\epsilon$ and precision $\delta$, where $d$ is the dimension of the unitary operators and $|\hat{X}|$ is the size of the list of synthesized unitary transformations. Since $|\hat{X}|$ grows exponentially with respect to $d^{2}$ (to make the list an $\epsilon$-covering of the $2 \epsilon$-ball around the target unitary transformation), the algorithm is not practical for higher dimensional unitary transformations.

There are two ways to make the algorithm more practical. First, restricting a class of target unitary transformations, such as axial rotations, would significantly reduce the size $|\hat{X}|$ but still achieve the guaranteed quadratic improvement. Indeed, Fig. 3 implies that the quadratic improvement can be achieved by mixing only two realizable unitary transformations. Alternatively, we can consider a modified algorithm that probabilistically mixes a randomly sampled small subset of $\hat{X}$. While this modified algorithm does not provide the guaranteed quadratic improvement as the original one, numerical experiments in Appendix D suggest that it still attains such improvement for randomly chosen target unitary transformations.

Similar to the probabilistic mixture of unitary transformations, that of general CPTP mappings implemented by a certain quantum device is relatively easy to implement by classically controlling the quantum device. Such a probabilistic mixture of implementable CPTP mappings is considered a free operation in many quantum resource theories [ $5,8,16]$. To quantify or simulate a target CPTP mapping using the probabilistic mixture (sometimes assisted by a resource state), a mathematical tool is required to analyze the optimal convex approximation of a general CPTP mapping. From the mathematical perspective as well as from the resource theoretical perspective, computing or bounding the approximation error of a unital CPTP mapping by using a probabilistic mixture of unitary transformations plays a crucial role in investigating the asymptotic quantum Birkhoff conjecture [12, 30]. Our SDP shown in Proposition 3.1 and our bounds (or possibly their extension to general CPTP mappings) could be numerical and analytical tools to investigate such problems.

## ACKNOWLEDGMENTS

We thank Yoshihisa Yamamoto, Aram Harrow, Isaac Chuang, Sho Sugiura, Yuki Takeuchi, Yasunari Suzuki, Yasuhiro Takahashi, and Adel Sohbi for their helpful discussions. This work was partially supported by JST Moonshot R\&D MILLENNIA Program (Grant No.JPMJMS2061). SA was partially supported by JST, PRESTO Grant No.JPMJPR2111 and JPMXS0120319794. GK was supported in part by the Grant-in-Aid for Scientific Research (C) No.20K03779, (C) No.21K03388, and (S) No.18H05237 of JSPS, CREST (Japan Science and Technology Agency) Grant No.JPMJCR1671. ST was partially supported by JSPS KAKENHI Grant Numbers JP20H05966 and JP22H00522.

## REFERENCES

[1] Dorit Aharonov and Michael Ben-Or. 2008. Fault-Tolerant Quantum Computation with Constant Error Rate. SIAM 7. Comput. 38, 4 (2008), 12071282. https://doi.org/10.1137/S0097539799359385 arXiv:https://doi.org/10.1137/S0097539799359385
[2] Charles H. Bennett, David P. DiVincenzo, John A. Smolin, and William K. Wootters. 1996. Mixed-state entanglement and quantum error correction. Phys. Rev. A 54 (Nov 1996), 3824-3851. Issue 5. https://doi.org/10.1103/PhysRevA.54.3824
[3] Alex Bocharov, Martin Roetteler, and Krysta M. Svore. 2015. Efficient Synthesis of Universal Repeat-Until-Success Quantum Circuits. Phys. Rev. Lett. 114 (Feb 2015), 080502. Issue 8. https://doi.org/10.1103/PhysRevLett.114.080502
[4] Adam Bouland and Tudor Giurgica-Tiron. 2021. Efficient Universal Quantum Compilation: An Inverse-free Solovay-Kitaev Algorithm. arXiv:2112.02040 [quant-ph]
[5] Fernando G. S. L. Brandão and Gilad Gour. 2015. Reversible Framework for Quantum Resource Theories. Phys. Rev. Lett. 115 (Aug 2015 ), 070503. Issue 7. https://doi.org/10.1103/PhysRevLett.115.070503
[6] Earl Campbell. 2017. Shorter gate sequences for quantum computing by mixing unitaries. Phys. Rev. A 95 (Apr 2017), 042306. Issue 4. https://doi.org/10.1103/PhysRevA. 95.042306
[7] Giulio Chiribella, Giacomo Mauro D'Ariano, and Paolo Perinotti. 2009. Theoretical framework for quantum networks. Phys. Rev. A 80 (Aug 2009), 022339. Issue 2. https://doi.org/10.1103/PhysRevA. 80.022339
[8] Eric Chitambar and Gilad Gour. 2019. Quantum resource theories. Rev. Mod. Phys.s 91,2 (2019), 025001.
[9] Austin G. Fowler. 2011. Constructing Arbitrary Steane Code Single Logical Qubit Fault-Tolerant Gates. Quantum Info. Comput. 11, 9-10 (sep 2011), 867-873.
[10] C.A. Fuchs and J. van de Graaf. 1999. Cryptographic distinguishability measures for quantum-mechanical states. IEEE Trans. Inf. Theory. 45, 4 (1999), 1216-1227. https://doi.org/10.1109/18.761271
[11] Gus Gutoski and John Watrous. 2007. Toward a General Theory of Quantum Games. In Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing (San Diego, California, USA) (STOC '07). Association for Computing Machinery, New York, NY, USA, 565-574. https://doi.org/10.1145/1250790.1250873
[12] Uffe Haagerup and Magdalena Musat. 2011. Factorization and Dilation Problems for Completely Positive Maps on von Neumann Algebras. Commun. Math. Phys. 303, 2 (2011), 555-594. https://doi.org/10.1007/s00220-011-1216-y
[13] Aram W. Harrow, Benjamin Recht, and Isaac L. Chuang. 2002. Efficient discrete approximations of quantum gates. J. Math. Phys. 43, 9 (2002), 4445-4451. https://doi.org/10.1063/1.1495899 arXiv:https://doi.org/10.1063/1.1495899
[14] Matthew B. Hastings. 2017. Turning Gate Synthesis Errors into Incoherent Errors. Quantum Info. Comput. 17, 5-6 (mar 2017), 488-494.
[15] Sam A. Hill and William K. Wootters. 1997. Entanglement of a Pair of Quantum Bits. Phys. Rev. Lett. 78 (Jun 1997), 5022-5025. Issue 26. https://doi.org/10.1103/PhysRevLett.78.5022
[16] Michal Horodecki and Jonathan Oppenheim. 2013. (QUANTUMNESS IN THE CONTEXT OF) RESOURCE THEORIES. Int. 7. Mod. Phys. B 27, 01 n 03 (2013), 1345019. https://doi.org/10.1142/S0217979213450197 arXiv:https://doi.org/10.1142/S0217979213450197
[17] A.Yu. Kitaev. 2003. Fault-tolerant quantum computation by anyons. Ann. Phys. 303, 1 (2003), 2-30. https://doi.org/10.1016/S0003-4916(02)00018-0
[18] A. Yu Kitaev, A. H. Shen, and M. N. Vyalyi. 2002. Classical and Quantum Computation. American Mathematical Society.
[19] Vadym Kliuchnikov, Kristin Lauter, Romy Minko, Adam Paetznick, and Christophe Petit. 2022. Shorter quantum circuits. arXiv:2203.10064 [quant$\mathrm{ph}]$
[20] Vadym Kliuchnikov, Dmitri Maslov, and Michele Mosca. 2013. Asymptotically Optimal Approximation of Single Qubit Unitaries by Clifford and $T$ Circuits Using a Constant Number of Ancillary Qubits. Phys. Rev. Lett. 110 (May 2013), 190502. Issue 19. https://doi.org/10.1103/PhysRevLett.110.190502
[21] Vadym Kliuchnikov, Dmitri Maslov, and Michele Mosca. 2016. Practical Approximation of Single-Qubit Unitaries by Single-Qubit Quantum Clifford and T Circuits. IEEE Trans. Comput. 65, 1 (2016), 161-172. https://doi.org/10.1109/TC.2015.2409842
[22] Emanuel Knill, Raymond Laflamme, and Wojciech H. Zurek. 1998. Resilient quantum computation: error models and thresholds. Proc. R. Soc. Lond. A. 454 (1998), 365-384. https://doi.org/10.1098/rspa.1998.0166
[23] L. Lovász. 2003. Semidefinite Programs and Combinatorial Optimization. Springer New York, New York, NY, 137-194. https://doi.org/10.1007/0-387-22444-0_6
[24] Michael A. Nielsen and Isaac L. Chuang. 2000. Quantum Computation and Quantum Information. Cambridge University Press.
[25] Neil J. Ross. 2015. Optimal Ancilla-Free CLIFFORD+V Approximation of Z-Rotations. Quantum Info. Comput. 15, 11-12 (sep 2015), 932-950.
[26] Massimiliano F. Sacchi. 2017. Optimal convex approximations of quantum states. Phys. Rev. A 96 (Oct 2017), 042325. Issue 4. https://doi.org/10.1103/PhysRevA. 96.042325
[27] Massimiliano F. Sacchi and Tito Sacchi. 2017. Convex approximations of quantum channels. Phys. Rev. A 96 (Sep 2017), 032311. Issue 3. https://doi.org/10.1103/PhysRevA.96.032311
[28] Barbara M. Terhal. 2015. Quantum error correction for quantum memories. Rev. Mod. Phys. 87 (Apr 2015), 307-346. Issue 2. https://doi.org/10.1103/RevModPhys. 87.307
[29] John Watrous. 2018. The Theory of Quantum Information. Cambridge University Press. https://doi.org/10.1017/9781316848142
[30] Nengkun Yu, Runyao Duan, and Quanhua Xu. 2012. Bounds on the distance between a unital quantum channel and the convex hull of unitary channels, with applications to the asymptotic quantum Birkhoff conjecture. arXiv preprint arXiv:1201.1172 (2012).

## A EQUIVALENCE BETWEEN QUANTUM TESTERS AND QUANTUM NETWORKS

Recall that the Choi-Jamiołkowski operator of linear mapping $\Xi: \mathbf{L}\left(\mathcal{H}_{1}\right) \rightarrow \mathbf{L}\left(\mathcal{H}_{2}\right)$ is defined as $J(\Xi):=\sum_{i, j}|i\rangle\langle j| \otimes$ $\Xi(|i\rangle\langle j|) \in \mathrm{L}\left(\mathcal{H}_{1} \otimes \mathcal{H}_{2}\right)$, and the set of quantum testers is defined as $\mathrm{T}\left(\mathcal{H}_{1}: \mathcal{H}_{2}\right):=\left\{T \in \operatorname{Pos}\left(\mathcal{H}_{1} \otimes \mathcal{H}_{2}\right): \exists \rho \in\right.$ $\left.\mathrm{S}\left(\mathcal{H}_{1}\right), T \leq \rho \otimes \mathbb{I}_{\mathcal{H}_{2}}\right\}$. In this section, we show that the set of mappings $f_{T}: \Xi \mapsto \operatorname{tr}[J(\Xi) T]$ associated with quantum testers $T \in \mathbf{T}\left(\mathcal{H}_{1}: \mathcal{H}_{2}\right)$ is equivalent to that of mappings $g_{\Phi, \Pi}: \Xi \mapsto \operatorname{tr}\left[\Xi \otimes i d_{\mathcal{H}_{3}}(\Phi) \Pi\right]$ associated with pure states $\Phi \in \mathrm{P}\left(\mathcal{H}_{1} \otimes \mathcal{H}_{3}\right)$ and Hermitian projectors $\Pi \in \operatorname{Proj}\left(\mathcal{H}_{2} \otimes \mathcal{H}_{3}\right)$ for sufficiently large dimensional Hilbert space $\mathcal{H}_{3}$. This equivalence indicates

$$
\begin{equation*}
\max _{T \in \mathbf{T}\left(\mathcal{H}_{1}: \mathcal{H}_{2}\right)} \min _{\Xi} f_{T}(\Xi)=\max _{\substack{\Phi \in \mathbf{P}\left(\mathcal{H}_{1} \otimes \mathcal{H}_{3}\right) \\ \Pi \in \operatorname{Proj}\left(\mathcal{H}_{2} \otimes \mathcal{H}_{3}\right)}} \min _{\Xi} g_{\Phi, \Pi}(\Xi) \tag{55}
\end{equation*}
$$

where the minimization of $\Xi$ is taken over a compact subset of linear mappings specified in the proofs of Proposition 3.1 and Lemma 4.2. Note that a proof for more general quantum testers is given in [7, Theorem 10].

First, we show that for any $\Phi$ and $\Pi$, there exists $T \in \mathbf{T}\left(\mathcal{H}_{1}: \mathcal{H}_{2}\right)$ such that $f_{T}=g_{\Phi, \Pi}$ as follows. By letting $T=\operatorname{tr}_{3}\left[\left(\Phi^{T_{1}} \otimes \mathbb{I}_{2}\right)\left(\mathbb{I}_{1} \otimes \Pi\right)\right]$, we obtain

$$
\begin{equation*}
g_{\Phi, \Pi}(\Xi)=\operatorname{tr}\left[\Xi \otimes i d_{\mathcal{H}_{3}}(\Phi) \Pi\right]=\operatorname{tr}\left[\left(J(\Xi) \otimes \mathbb{I}_{3}\right)\left(\Phi^{T_{1}} \otimes \mathbb{I}_{2}\right)\left(\mathbb{I}_{1} \otimes \Pi\right)\right]=\operatorname{tr}[J(\Xi) T]=f_{T}(\Xi) \tag{56}
\end{equation*}
$$

where $\Phi^{T_{1}}$ and $\operatorname{tr}_{3}[\cdot]$ represent the partial transpose of $\Phi$ and the partial trace, respectively, and the subscript of the operator denotes the system on which the operator acts. We can also verify that $T \in \mathrm{~T}\left(\mathcal{H}_{1}: \mathcal{H}_{2}\right)$ as follows. Let $X=\sum_{i j} \alpha_{i j}|j\rangle_{3}\left\langle\left. i\right|_{1}\right.$, where $\left.\mid \Phi\right\rangle=\sum_{i j} \alpha_{i j}|i\rangle_{1}|j\rangle_{3}$ with the computational basis $\left\{|i\rangle_{1} \in \mathbf{P}\left(\mathcal{H}_{1}\right)\right\}_{i}$ and $\left\{|j\rangle_{3} \in \mathbf{P}\left(\mathcal{H}_{3}\right)\right\}_{j}$. We then obtain that for any positive semidefinite operator $P \in \operatorname{Pos}\left(\mathcal{H}_{1} \otimes \mathcal{H}_{2}\right)$,

$$
\begin{equation*}
\operatorname{tr}[P T]=\operatorname{tr}\left[\left(P \otimes \mathbb{I}_{3}\right)\left(\Phi^{T_{1}} \otimes \mathbb{I}_{2}\right)\left(\mathbb{I}_{1} \otimes \Pi\right)\right]=\operatorname{tr}\left[\left(X \otimes \mathbb{I}_{2}\right) P\left(X \otimes \mathbb{I}_{2}\right)^{\dagger} \Pi\right] \geq 0 \tag{57}
\end{equation*}
$$

which indicates $T \geq 0$. By letting $\rho=\operatorname{tr}_{3}\left[\Phi^{T_{1}}\right]=\operatorname{tr}_{3}[\Phi]^{T}\left(\in \mathrm{~S}\left(\mathcal{H}_{1}\right)\right)$, we can also verify that

$$
\begin{equation*}
\rho \otimes \mathbb{I}_{2}-T=\operatorname{tr}_{3}\left[\left(\Phi^{T_{1}} \otimes \mathbb{I}_{2}\right)\left(\mathbb{I}_{123}-\mathbb{I}_{1} \otimes \Pi\right)\right]=\operatorname{tr}_{3}\left[\left(\Phi^{T_{1}} \otimes \mathbb{I}_{2}\right)\left(\mathbb{I}_{1} \otimes \Pi_{\perp}\right)\right] \geq 0 \tag{58}
\end{equation*}
$$

where $\Pi_{\perp} \in \operatorname{Proj}\left(\mathcal{H}_{2} \otimes \mathcal{H}_{3}\right)$ satisfies $\Pi+\Pi_{\perp}=\mathbb{I}$, and the last inequality can be verified by the fact that $T \geq 0$.
Next, we show that for any $T \in \mathrm{~T}\left(\mathcal{H}_{1}: \mathcal{H}_{2}\right)$, there exist $\Phi \in \mathbf{P}\left(\mathcal{H}_{1} \otimes \mathcal{H}_{3}\right)$ and $\Pi \in \operatorname{Proj}\left(\mathcal{H}_{2} \otimes \mathcal{H}_{3}\right)$ such that $f_{T}=g_{\Phi, \Pi}$ as follows. Let $T \leq \rho_{1} \otimes \mathbb{I}_{2}, \hat{\Phi} \in \mathbf{P}\left(\mathcal{H}_{1} \otimes \mathcal{H}_{1^{\prime}}\right)$ be a purification of $\rho_{1}^{T}$, its singular value decomposition be $|\hat{\Phi}\rangle=\sum_{i} \sqrt{p(i)}\left|x_{i}\right\rangle_{1}\left|y_{i}\right\rangle_{1^{\prime}}(p(i)>0)$, and $P \in \operatorname{Pos}\left(\mathcal{H}_{2} \otimes \mathcal{H}_{1^{\prime}}\right)$ be $P=X T X^{\dagger}$, where $X=\sum_{i} \frac{1}{\sqrt{p(i)}}\left|y_{i}\right\rangle_{1^{\prime}}\left\langle\left. x_{i}^{*}\right|_{1}\right.$ and $\left.\mid \phi^{*}\right\rangle$ is the complex conjugate of $|\phi\rangle$. We can then verify that

$$
\begin{equation*}
f_{T}(\Xi)=\operatorname{tr}[J(\Xi) T]=\operatorname{tr}\left[\left(J(\Xi) \otimes \mathbb{I}_{1^{\prime}}\right)\left(\hat{\Phi}^{T_{1}} \otimes \mathbb{I}_{2}\right)\left(\mathbb{I}_{1} \otimes P\right)\right]=\operatorname{tr}\left[\Xi \otimes i d_{\mathcal{H}_{1^{\prime}}}(\hat{\Phi}) P\right] \tag{59}
\end{equation*}
$$

Since $P \leq X\left(\rho_{1} \otimes \mathbb{I}_{2}\right) X^{\dagger} \leq \mathbb{I}_{1^{\prime} 2},\{P, \mathbb{I}-P\}$ is a positive operator-valued measure (POVM). Owing to the Naimark's extension, we can embed $\hat{\Phi}$ and $\{P, \mathbb{I}-P\}$ in a larger Hilbert space as a pure state $\Phi$ and a projection-valued measure $(\mathrm{PVM})\left\{\Pi, \Pi_{\perp}\right\}$, respectively, which completes the proof.

## B FORMAL SDPS AND THEIR STRONG DUALITY

A formal SDP to compute $\frac{1}{2}\|\mathcal{A}-\mathcal{B}\|_{\diamond}$ is defined with a triple $(\Xi, A, B)$ such that

$$
A=\left(\begin{array}{ccc}
J(\mathcal{A}-\mathcal{B}) & 0 & 0  \tag{60}\\
0 & 0 & 0 \\
0 & 0 & 0
\end{array}\right), \quad B=\left(\begin{array}{ll}
0 & 0 \\
0 & 1
\end{array}\right), \quad \Xi\left(\left(\begin{array}{ccc}
T & * & * \\
* & T^{\prime} & * \\
* & * & \rho
\end{array}\right)\right)=\left(\begin{array}{cc}
T+T^{\prime}-\rho \otimes \mathbb{I}_{\mathcal{H}_{2}} & 0 \\
0 & \operatorname{tr}[\rho]
\end{array}\right)
$$

holds for any linear operators $T, T^{\prime} \in \mathbf{L}\left(\mathcal{H}_{1} \otimes \mathcal{H}_{2}\right)$ and $\rho \in \mathbf{L}\left(\mathcal{H}_{1}\right)$, where the asterisks in the argument to $\Xi$ represent arbitrary linear operators upon which $\Xi$ does not depend, and we identify a linear operator and its matrix representation with respect to a fixed orthonormal basis. The dual problem is obtained by observing that the adjoint of $\Xi$ satisfies

$$
\Xi^{\dagger}\left(\left(\begin{array}{cc}
S & *  \tag{61}\\
* & r
\end{array}\right)\right)=\left(\begin{array}{ccc}
S & 0 & 0 \\
0 & S & 0 \\
0 & 0 & r \mathbb{I}_{\mathcal{H}_{1}}-\operatorname{tr}_{\mathcal{H}_{2}}[S]
\end{array}\right)
$$

for any linear operator $S \in \mathbf{L}\left(\mathcal{H}_{1} \otimes \mathcal{H}_{2}\right)$ and any complex number $r \in \mathbb{C}$. We can verify the strong duality of this SDP by observing $\Xi\left(\frac{\mathbb{I}_{\mathcal{H}_{1}} \otimes \mathbb{I}_{\mathcal{H}_{2}}}{2 \operatorname{dim} \mathcal{H}_{1}} \oplus \frac{\mathbb{I}_{\mathcal{H}_{1}} \not \mathbb{I}_{\mathcal{H}_{2}}}{2 \operatorname{dim} \mathcal{H}_{1}} \oplus \frac{\mathbb{I}_{\mathcal{H}_{1}}}{\operatorname{dim} \mathcal{H}_{1}}\right)=B$ and applying the Slater's theorem.

A formal SDP shown in Proposition 3.1 is defined with a triple $(\Xi, A, B)$ such that

$$
\begin{align*}
& A=\left(\begin{array}{ccccc}
J(\mathcal{A}) & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -1
\end{array}\right)  \tag{62}\\
& B=\left(\begin{array}{lll}
0 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 0
\end{array}\right)  \tag{63}\\
& \Xi^{\dagger}\left(\left(\begin{array}{ccc}
S & * & * \\
* & r & * \\
* & * & P
\end{array}\right)\right)=\left(\begin{array}{ccccc}
S+\sum_{x \in X} P(x) J\left(\mathcal{B}_{x}\right) & 0 & 0 & 0 & 0 \\
0 & S & 0 & 0 & 0 \\
0 & 0 & r \mathbb{I}_{\mathcal{H}_{1}}-\operatorname{tr}_{\mathcal{H}_{2}}[S] & 0 & 0 \\
0 & 0 & 0 & P & 0 \\
0 & 0 & 0 & 0 & -\operatorname{tr}[P]
\end{array}\right) \tag{64}
\end{align*}
$$

holds for any linear operators $S \in \mathbf{L}\left(\mathcal{H}_{1} \otimes \mathcal{H}_{2}\right), P \in \mathbf{L}\left(\mathbb{C}^{|X|}\right)$ and any complex number $r \in \mathbb{C}$, where $P(x)$ represents a diagonal element $\langle x| P|x\rangle$. The primal problem is obtained by observing that the adjoint of $\Xi^{\dagger}$ satisfies

$$
\Xi\left(\left(\begin{array}{ccccc}
T & * & * & * & *  \tag{65}\\
* & T^{\prime} & * & * & * \\
* & * & \rho & * & * \\
* & * & * & Q & * \\
* & * & * & * & t
\end{array}\right)\right)=\left(\begin{array}{ccc}
T+T^{\prime}-\rho \otimes \mathbb{I}_{\mathcal{H}_{2}} & 0 & 0 \\
0 & \operatorname{tr}[\rho] & 0 \\
0 & 0 & \sum_{x \in X} \operatorname{tr}\left[J\left(\mathcal{B}_{x}\right) T\right]|x\rangle\langle x|+Q-t \mathbb{I}_{\mathbb{C}^{|X|}}
\end{array}\right)
$$

for any linear operators $T, T^{\prime} \in \mathbf{L}\left(\mathcal{H}_{1} \otimes \mathcal{H}_{2}\right), \rho \in \mathbf{L}\left(\mathcal{H}_{1}\right), Q \in \mathbf{L}\left(\mathbb{C}^{|X|}\right)$ and any complex number $t \in \mathbb{C}$. We can verify the strong duality of this SDP by observing $\Xi\left(\frac{\mathbb{I}_{\mathcal{H}_{1}} \otimes \mathbb{I}_{\mathcal{H}_{2}}}{2 \operatorname{dim} \mathcal{H}_{1}} \oplus \frac{\mathbb{I}_{\mathcal{H}_{1}} \mathbb{I}_{\mathcal{H}_{2}}}{2 \operatorname{dim} \mathcal{H}_{1}} \oplus \frac{\mathbb{I}_{\mathcal{H}_{1}}}{\operatorname{dim} \mathcal{H}_{1}} \oplus \frac{\mathbb{I}_{\mathrm{C}}|X|}{2} \oplus 1\right)=B$ and applying the Slater's theorem.

## C SHARPNESS OF APPROXIMATION ERROR BOUNDS

In this section, we make the same assumption $d \geq 2$ as Lemma 4.1 and 4.2.

## C. 1 Lower bounds

To show the sharpness of the lower bounds in Ineqs. (12) and (1), we consider a set $\left\{\Upsilon_{x}\right\}_{x \in X}:=\left\{\Upsilon: \exists W \in \mathbf{W}_{\epsilon}^{(d)}, \Upsilon(\rho)=\right.$ $\left.W \rho W^{\dagger}\right\}$ of unitary transformations, where

$$
\begin{equation*}
\mathbf{W}_{\epsilon}^{(d)}:=\left\{W: W \in U(d) \wedge \min _{z \in \operatorname{conv}(\lambda(W))}|z| \leq \sqrt{1-\epsilon^{2}},\right\} \text { with } \epsilon \in[0,1] \text { and } d \geq 2 \text {, } \tag{66}
\end{equation*}
$$

where $U(d)$ represents the set of unitary operators acting on $\mathbb{C}^{d}, \lambda(W)$ represents the set of eigenvalues of $W$, and conv $(X)$ represents the convex hull of a subset $X$ in a vector space. To be precise, the two lower bounds are not directly applicable to $\left\{\Upsilon_{x}\right\}_{x \in X}$ since the size $|X|$ of the set is infinite. However, the compactness of the set of unitary transformations on a finite-dimensional Hilbert space enables us to extend Ineqs. (12) and (1) for $|X|=\infty$ by replacing $\min _{x \in X} \frac{1}{2}\left\|\Upsilon-\Upsilon_{x}\right\|_{\diamond}$ and $\min _{p} \frac{1}{2}\left\|\Upsilon-\sum_{x \in X} p(x) \Upsilon_{x}\right\|_{\diamond}$ with $\inf _{x \in X} \frac{1}{2}\left\|\Upsilon-\Upsilon_{x}\right\|_{\diamond}$ and $\inf _{\Lambda \in \operatorname{conv}\left(\left\{\Upsilon_{x}\right\}_{x \in X}\right)} \frac{1}{2}\|\Upsilon-\Lambda\|_{\diamond}$, respectively.

We show that this example achieves the lower bounds in the extended inequalities with a target unitary transformation $\Upsilon=i d$ in Ineq. (12). This also indicates that there exists a finite subset $\left\{\Upsilon_{x}\right\}_{x \in \tilde{X}}$ of $\left\{\Upsilon_{x}\right\}_{x \in X}$ such that $\min _{p} \frac{1}{2}\left\|i d-\sum_{x \in \tilde{X}} p(x) \Upsilon_{x}\right\|_{\diamond}$ and $\max _{\Upsilon} \min _{p} \frac{1}{2}\left\|\Upsilon-\sum_{x \in \tilde{X}} p(x) \Upsilon_{x}\right\|_{\diamond}$ are arbitrarily close to their each lower bound in Ineqs. (12) and (1), respectively. For letting $\left\{\Upsilon_{x}\right\}_{x \in \tilde{X}}$ be an $\tilde{\epsilon}$-covering of $\left\{\Upsilon_{x}\right\}_{x \in X}$ with sufficiently small $\tilde{\epsilon}$ is sufficient to show this. Thus, the sharpness of the lower bounds in the extended inequalities indicates that in the original inequalities. Note that we can show the sharpness of Ineq. (12) when an $\Upsilon$ is not the identity transformation by replacing $\left\{\Upsilon_{x}\right\}_{x \in X}$ with $\left\{\Upsilon \circ \Upsilon_{x}\right\}_{x \in X}$.

First, by using Eq. (17), we obtain

$$
\begin{align*}
& \max _{\Upsilon} \inf _{x \in X} \frac{1}{2}\left\|\Upsilon-\Upsilon_{x}\right\|_{\diamond} \geq \inf _{x \in X} \frac{1}{2}\left\|i d-\Upsilon_{x}\right\|_{\diamond} \\
& =\sqrt{1-\sup _{W \in \mathbf{W}_{\epsilon}^{(d)}} \min _{\rho \in \mathrm{S}\left(\mathbb{C}^{d}\right)}|\operatorname{tr}[\rho W]|^{2}}=\sqrt{1-\sup _{W \in \mathbf{W}_{\epsilon}^{(d)}} \min _{z \in \operatorname{conv}(\lambda(W))}|z|^{2}}=\epsilon . \tag{67}
\end{align*}
$$

Second, by using the extended version of Eq. (29), we obtain

$$
\begin{equation*}
\inf _{\Lambda \in \operatorname{conv}\left(\left\{\Upsilon_{x}\right\}_{x \in X}\right)} \frac{1}{2}\|i d-\Lambda\|_{\diamond} \leq \max _{\Upsilon} \inf _{\Lambda \in \operatorname{conv}\left(\left\{\Upsilon_{x}\right\}_{x \in X}\right)} \frac{1}{2}\|\Upsilon-\Lambda\|_{\diamond}=1-\min _{V \in U(d), \rho \in \mathrm{S}\left(\mathbb{C}^{d}\right)} \sup _{W \in \mathbf{W}_{\epsilon}^{(d)}}\left|\operatorname{tr}\left[\rho V^{\dagger} W\right]\right|^{2} . \tag{68}
\end{equation*}
$$

In the following, we show that for any $V \in U(d)$ and $\rho \in \mathrm{S}\left(\mathbb{C}^{d}\right)$,

$$
\begin{equation*}
\sup _{W \in \mathbf{W}_{\epsilon}^{(d)}}\left|\operatorname{tr}\left[\rho V^{\dagger} W\right]\right|^{2} \geq\left(1-\frac{2 \delta}{d}\right)^{2} \text { with } \delta=1-\sqrt{1-\epsilon^{2}} \tag{69}
\end{equation*}
$$

which is sufficient to verify that $\left\{\Upsilon_{x}\right\}_{x \in X}$ achieves lower bounds in the extended Ineqs. (12) and (1).

Let the diagonalization of $V$ be $V=\sum_{i=1}^{d} \lambda_{i}(V)|i\rangle\langle i|$. Since $\sup _{W \in \mathbf{W}_{\epsilon}^{(d)}}\left|\operatorname{tr}\left[\rho V^{\dagger} W\right]\right|^{2}=\left.\operatorname{tr}\left[\rho V^{\dagger} V\right]\right|^{2}=1$ if $\min _{z \in \operatorname{conv}(\lambda(V))}|z| \leq$ $\sqrt{1-\epsilon^{2}}$, we assume $\epsilon>0$ and $\min _{z \in \operatorname{conv}(\lambda(V))}|z|>\sqrt{1-\epsilon^{2}}$. We can then define $\left\{W^{(i j)} \in \mathbf{W}_{\epsilon}^{(d)}\right\}_{1 \leq i<j \leq d}$ as

$$
\begin{align*}
& W^{(i j)}:=\sum_{k \notin\{i, j\}} \lambda_{k}(V)|k\rangle\langle k|+\lambda_{+}^{(i j)}|i\rangle\langle i|+\lambda_{-}^{(i j)}|j\rangle\langle j|,  \tag{70}\\
& \text { where } \lambda_{ \pm}^{(i j)}=\sqrt{1-\epsilon^{2}} \frac{\lambda_{i}(V)+\lambda_{j}(V)}{\left|\lambda_{i}(V)+\lambda_{j}(V)\right|} \pm \epsilon \frac{\lambda_{i}(V)-\lambda_{j}(V)}{\left|\lambda_{i}(V)-\lambda_{j}(V)\right|} \text { if } \lambda_{i}(V) \neq \lambda_{j}(V) \text {, }  \tag{71}\\
& \text { and } \lambda_{ \pm}^{(i j)}=\sqrt{1-\epsilon^{2}} \lambda_{i}(V) \pm i \epsilon \lambda_{i}(V) \quad \text { if } \lambda_{i}(V)=\lambda_{j}(V) . \tag{72}
\end{align*}
$$



Fig. 4. Geometric positions of eigenvalues $\lambda_{i}(V), \lambda_{j}(V)$ and $\lambda_{ \pm}^{(i j)}$ of unitary operators, which lie on unit circle in complex plane. Note that real and imaginary axes are rotated to horizontalize line equidistant from $\lambda_{i}(V)$ and $\lambda_{j}(V)$.
(See geometric positions of eigenvalues in the complex plane shown in Fig. 4.) Note that we can easily verify that $\left|\lambda_{ \pm}^{(i j)}\right|=1$ and $\left|\frac{1}{2}\left(\lambda_{+}^{(i j)}+\lambda_{-}^{(i j)}\right)\right|=\sqrt{1-\epsilon^{2}}$, which guarantees $W^{(i j)} \in \mathbf{W}_{\epsilon}^{(d)}$. Moreover, we can verify that $\lambda\left(V^{\dagger} W^{(i j)}\right)=\left\{1, z^{(i j)}, z^{(i j) *}\right\}$ with a unit complex number $z^{(i j)}$ satisfying $\operatorname{Re}\left[z^{(i j)}\right] \geq \sqrt{1-\epsilon^{2}}$. Then, for any $V \in U(d)$ and $\rho \in \mathrm{S}\left(\mathbb{C}^{d}\right)$, the left hand side of Ineq. (69) can be bounded as

$$
\begin{align*}
\sup _{W \in \mathbf{W}_{\epsilon}^{(d)}}\left|\operatorname{tr}\left[\rho V^{\dagger} W\right]\right|^{2} & \geq \max _{1 \leq i<j \leq d}\left|\operatorname{tr}\left[\rho V^{\dagger} W^{(i j)}\right]\right|^{2} \geq \min _{p} \max _{1 \leq i<j \leq d} \mid \sum_{k \notin\{i, j\}} p(k)+p(i) z^{(i j)}+p(j) z^{\left.(i j) *\right|^{2}}  \tag{73}\\
& \geq \min _{p} \max _{1 \leq i<j \leq d}\left\{\sum_{k \notin\{i, j\}} p(k)+(p(i)+p(j)) \operatorname{Re}\left[z^{(i j)}\right]\right\}^{2}  \tag{74}\\
& \geq \min _{p} \max _{1 \leq i<j \leq d}\left\{\sum_{k \notin\{i, j\}} p(k)+(p(i)+p(j)) \sqrt{1-\epsilon^{2}}\right\}^{2}  \tag{75}\\
& \geq \min _{p} \max _{1 \leq i<j \leq d}\{1-\delta(p(i)+p(j))\}^{2} \geq\left(1-\frac{2 \delta}{d}\right)^{2} \tag{76}
\end{align*}
$$

This completes the proof.

## C. 2 Upper bound

We show the sharpness of the upper bound in Ineq. (1), We consider a set $\left\{\Upsilon_{x}\right\}_{x \in X}:=\left\{\Upsilon: \exists V \in \mathbf{V}_{\epsilon}^{(d)}, \Upsilon(\rho)=V \rho V^{\dagger}\right\}$ of unitary transformations, where

$$
\begin{align*}
\mathbf{V}_{\epsilon}^{(d)} & :=\left\{\left(\begin{array}{cc}
1 & 0 \\
0 & V_{1}
\end{array}\right)\left(\begin{array}{cc}
W & 0 \\
0 & \mathbb{I}_{d-2}
\end{array}\right)\left(\begin{array}{cc}
1 & 0 \\
0 & V_{2}
\end{array}\right): V_{1}, V_{2} \in U(d-1), W \in \mathbf{R}_{\epsilon}\right\}  \tag{77}\\
\mathbf{R}_{\epsilon} & :=\left\{\left(\begin{array}{cc}
\cos \theta & -\sin \theta \\
\sin \theta & \cos \theta
\end{array}\right): 0 \leq \theta \leq \arccos (\epsilon)\right\} \text { with } \epsilon \in[0,1] \text { and } d \geq 2 . \tag{78}
\end{align*}
$$

Here $\mathbb{I}_{d}$ represents the $d \times d$ identity matrix, and we identify a unitary operator and its matrix representation with respect to a fixed orthonormal basis $\{|i\rangle\}_{i=0}^{d-1}$. Since $|X|=\infty$, we show the sharpness of the upper bound in the extended Ineq. (1), which is defined in the proof of the sharpness of the lower bounds. Note that

$$
\begin{equation*}
\forall U \in U(d), \exists \alpha \in \mathbb{R}, \exists V \in \mathbf{V}_{0}^{(d)}, U=e^{i \alpha} V \tag{79}
\end{equation*}
$$

holds. This can be verified from the following three observations: First, by letting $U|i\rangle=\left|e_{i}\right\rangle$, there exists $\tilde{V}_{1}, \tilde{V}_{2} \in$ $U(d-1)$ and $\tilde{W} \in U(2)$ such that $\left(\begin{array}{cc}\tilde{W}^{\dagger} & 0 \\ 0 & \mathbb{I}_{d-2}\end{array}\right)\left(\begin{array}{cc}1 & 0 \\ 0 & \tilde{V}_{1}^{\dagger}\end{array}\right)\left|e_{0}\right\rangle=|0\rangle$ and $\left(\begin{array}{cc}1 & 0 \\ 0 & \tilde{V}_{2}^{\dagger}\end{array}\right)\left(\begin{array}{cc}\tilde{W}^{\dagger} & 0 \\ 0 & \mathbb{I}_{d-2}\end{array}\right)\left(\begin{array}{cc}1 & 0 \\ 0 & \tilde{V}_{1}^{\dagger}\end{array}\right)\left|e_{i}\right\rangle=|i\rangle$ for all i. Second, for any $\tilde{W} \in U(2)$, there exists $\alpha, \beta, \gamma \in \mathbb{R}$ and $W \in \mathbf{R}_{0}$ such that $\tilde{W}=e^{i \alpha}\left(\begin{array}{cc}1 & 0 \\ 0 & e^{i \beta}\end{array}\right) W\left(\begin{array}{cc}1 & 0 \\ 0 & e^{i \gamma}\end{array}\right)$. Third, by letting $V_{1}=\tilde{V}_{1}\left(\begin{array}{cc}e^{i \beta} & 0 \\ 0 & e^{-i \alpha \mathbb{I}_{d-2}}\end{array}\right)$ and $V_{2}=\left(\begin{array}{cc}e^{i \gamma} & 0 \\ 0 & \mathbb{I}_{d-2}\end{array}\right) \tilde{V}_{2}$, we can verify $U=e^{i \alpha}\left(\begin{array}{cc}1 & 0 \\ 0 & V_{1}\end{array}\right)\left(\begin{array}{cc}W & 0 \\ 0 & \mathbb{I}_{d-2}\end{array}\right)\left(\begin{array}{cc}1 & 0 \\ 0 & V_{2}\end{array}\right)$.

First, by using Eq. (17), we obtain

$$
\begin{equation*}
\max _{\Upsilon} \inf _{x \in X} \frac{1}{2}\left\|\Upsilon-\Upsilon_{x}\right\|_{\diamond}=\sqrt{1-\min _{U \in U(d)} \sup _{V \in \mathrm{~V}_{\epsilon}^{(d)}} \min _{\rho \in \mathrm{S}\left(\mathbb{C}^{d}\right)}\left|\operatorname{tr}\left[\rho U^{\dagger} V\right]\right|^{2}} \tag{80}
\end{equation*}
$$

$$
=\sqrt{1-\min _{W \in \mathbf{R}_{0}} \sup _{V \in \mathrm{~V}_{\epsilon}^{(d)}} \min _{\rho \in \mathrm{S}\left(\mathbb{C}^{d}\right)}\left|\operatorname{tr}\left[\rho\left(\begin{array}{cc}
W^{\dagger} & 0  \tag{81}\\
0 & \mathbb{I}_{d-2}
\end{array}\right) V\right]\right|^{2}}
$$

$$
\leq \sqrt{1-\min _{W \in \mathbf{R}_{0}} \sup _{W^{\prime} \in \mathbf{R}_{\epsilon}} \min _{\rho \in \mathrm{S}\left(\mathbb{C}^{d}\right)}\left|\operatorname{tr}\left[\rho\left(\begin{array}{cc}
W^{\dagger} W^{\prime} & 0  \tag{82}\\
0 & \mathbb{I}_{d-2}
\end{array}\right)\right]\right|^{2}}
$$

$$
=\sqrt{1-\min _{0 \leq \theta \leq \frac{\pi}{2}} \sup _{0 \leq \theta^{\prime} \leq \arccos (\epsilon)} \min _{\rho \in \mathrm{S}\left(\mathbb{C}^{d}\right)}\left|\operatorname{tr}\left[\rho\left(\begin{array}{ccc}
\cos \left(\theta^{\prime}-\theta\right) & -\sin \left(\theta^{\prime}-\theta\right) & 0  \tag{83}\\
\sin \left(\theta^{\prime}-\theta\right) & \cos \left(\theta^{\prime}-\theta\right) & 0 \\
0 & 0 & \mathbb{I}_{d-2}
\end{array}\right]\right)\right|^{2}}
$$

$$
\begin{equation*}
=\sqrt{1-\min _{0 \leq \theta \leq \frac{\pi}{2}} \sup _{0 \leq \theta^{\prime} \leq \arccos (\epsilon)} \cos ^{2}\left(\theta^{\prime}-\theta\right)} \tag{84}
\end{equation*}
$$

$$
\begin{equation*}
=\max _{0 \leq \theta \leq \frac{\pi}{2}} \inf _{0 \leq \theta^{\prime} \leq \arccos (\epsilon)}\left|\sin \left(\theta^{\prime}-\theta\right)\right|=\epsilon \tag{85}
\end{equation*}
$$

where we use Eq. (79) in the second equality and use $\lambda\left(\left(\begin{array}{ccc}\cos \left(\theta^{\prime}-\theta\right) & -\sin \left(\theta^{\prime}-\theta\right) & 0 \\ \sin \left(\theta^{\prime}-\theta\right) & \cos \left(\theta^{\prime}-\theta\right) & 0 \\ 0 & 0 & \mathbb{I}_{d-2}\end{array}\right)\right)=\left\{1, e^{ \pm i\left(\theta^{\prime}-\theta\right)}\right\}$ in the fourth equality.

Second, by using the definition of the diamond norm, we obtain

$$
\begin{align*}
& \max _{\Upsilon} \inf _{\Lambda \in \operatorname{conv}\left(\left\{\Upsilon_{x}\right\}_{x \in X}\right)} \frac{1}{2}\|\Upsilon-\Lambda\|_{\diamond} \geq \max _{\Upsilon} \inf _{\Lambda \in \operatorname{conv}\left(\left\{\Upsilon_{x}\right\}_{x \in X}\right)}\|\Upsilon(|0\rangle\langle 0|)-\Lambda(|0\rangle\langle 0|)\|_{\text {tr }}  \tag{86}\\
& \geq 1-\min _{\Upsilon} \sup _{\Lambda \in \operatorname{conv}\left(\left\{\Upsilon_{x}\right\}_{x \in X}\right)} F(\Upsilon(|0\rangle\langle 0|), \Lambda(|0\rangle\langle 0|))  \tag{87}\\
& =1-\min _{\Upsilon} \sup _{x \in X} F\left(\Upsilon(|0\rangle\langle 0|), \Upsilon_{x}(|0\rangle\langle 0|)\right)  \tag{88}\\
& \left.=1-\min _{U \in U(d)} \sup _{V \in \mathbf{V}_{\epsilon}^{(d)}}\left|\langle 0| U^{\dagger} V\right| 0\right\rangle\left.\right|^{2}  \tag{89}\\
& \left.=1-\min _{W \in \mathbf{R}_{0}} \sup _{\substack{W^{\prime} \in \mathbf{R}_{\epsilon} \\
V \in U(d-1)}}\left|\langle 0|\left(\begin{array}{cc}
W^{\dagger} & 0 \\
0 & \mathbb{I}_{d-2}
\end{array}\right)\left(\begin{array}{cc}
1 & 0 \\
0 & V
\end{array}\right)\left(\begin{array}{cc}
W^{\prime} & 0 \\
0 & \mathbb{I}_{d-2}
\end{array}\right)\right| 0\right\rangle\left.\right|^{2}  \tag{90}\\
& \left.=1-\min _{0 \leq \theta \leq \frac{\pi}{2}} \sup _{0 \leq \theta^{\prime} \leq \arccos (\epsilon)}\left|\cos \theta \cos \theta^{\prime}+\sin \theta \sin \theta^{\prime}\langle 1| V\right| 1\right\rangle\left.\right|^{2}  \tag{91}\\
& V \in U(d-1) \\
& =\max _{0 \leq \theta \leq \frac{\pi}{2}} \inf _{0 \leq \theta^{\prime} \leq \arccos (\epsilon)} \sin ^{2}\left(\theta^{\prime}-\theta\right)=\epsilon^{2}, \tag{92}
\end{align*}
$$

where we use $\|\phi-\rho\|_{\text {tr }}=\max _{\Pi \in \operatorname{Proj}(\mathcal{H})} \operatorname{tr}[\Pi(\phi-\rho)] \geq 1-\operatorname{tr}[\phi \rho]$ in the second inequality and use Eq. (79) in the third equality. This and the extended upper bound in Ineq. (1) complete the proof.

## D NUMERICAL EXPERIMENT FOR ACTUAL APPROXIMATION ERRORS

Recall that Theorem 4.3 has established the relationship between the actual approximation error $\epsilon_{\mathrm{r}}^{\text {prob }}$ caused by the probabilistic approximation, that $\epsilon_{\Upsilon}$ caused by the deterministic approximation, and the worst approximation error $\epsilon$, where each approximation error is defined by

$$
\begin{align*}
\epsilon_{\Upsilon}^{p r o b} & :=\min _{p} \frac{1}{2}\left\|\Upsilon-\sum_{x \in X} p(x) \Upsilon_{x}\right\|_{\diamond}  \tag{93}\\
\epsilon_{\Upsilon} & :=\min _{x \in X} \frac{1}{2}\left\|\Upsilon-\Upsilon_{x}\right\|_{\diamond}  \tag{94}\\
\epsilon & :=\max _{\Upsilon} \min _{x \in X} \frac{1}{2}\left\|\Upsilon-\Upsilon_{x}\right\|_{\diamond} . \tag{95}
\end{align*}
$$

By using these notations, we can rewrite the relationship established in Theorem 4.3 as follows:

$$
\begin{equation*}
\frac{2 \epsilon_{\Upsilon}^{2}}{d} \simeq \frac{4 \delta_{\Upsilon}}{d}\left(1-\frac{\delta_{\Upsilon}}{d}\right) \leq \epsilon_{\Upsilon}^{p r o b} \leq \epsilon^{2}, \tag{96}
\end{equation*}
$$

where $\delta_{\Upsilon}=1-\sqrt{1-\epsilon_{\Upsilon}^{2}}$ and the approximation becomes tighter when $\epsilon_{\Upsilon} \rightarrow 0$.
While these inequalities are tight, it is helpful to know how small $\epsilon_{\Upsilon}^{\text {prob }}$ can be realized compared to $\epsilon_{\Upsilon}^{2}$. In this appendix, we perform numerical experiments to demonstrate that $\epsilon_{\mathrm{r}}^{\text {prob }}$ is comparable to $\epsilon_{\mathrm{r}}^{2}$ for randomly chosen target unitary transformations $\Upsilon$. Moreover, we demonstrate that $\epsilon_{\Upsilon}^{p r o b}$ becomes smaller than $\epsilon_{\Upsilon}^{2}$ for high-dimensional unitary transformations, i.e., the probabilistic approximation reduces the approximation error more than quadratically.

Additionally, the experiments have another purpose: to provide pieces of evidence supporting that Lemma 5.3 holds for qudit unitary transformations, which is crucial in constructing our probabilistic synthesis algorithm. Recall that it
states that

$$
\begin{equation*}
\min _{p}\left\|\Upsilon-\sum_{x \in X} p(x) \Upsilon_{x}\right\|_{\diamond}=\min _{\hat{p}}\left\|\Upsilon-\sum_{x \in \hat{X}} \hat{p}(x) \Upsilon_{x}\right\|_{\diamond} \tag{97}
\end{equation*}
$$

where $\hat{X}:=\left\{x \in X: \frac{1}{2}\left\|\Upsilon-\Upsilon_{x}\right\|_{\diamond} \leq 2 \epsilon\right\}$. Our numerical experiments support this statement for randomly sampled target unitary transformations $\Upsilon: \mathbf{L}\left(\mathbb{C}^{d}\right) \rightarrow \mathbf{L}\left(\mathbb{C}^{d}\right)$, randomly constructed $\epsilon$-coverings $\left\{\Upsilon_{x}: \mathbf{L}\left(\mathbb{C}^{d}\right) \rightarrow \mathbf{L}\left(\mathbb{C}^{d}\right)\right\}_{x \in X}$, and $d=3,4$.

## Setting of numerical experiments

First, we construct $\epsilon$-coverings $\left\{\Upsilon_{x}: \mathbf{L}\left(\mathbb{C}^{d}\right) \rightarrow \mathbf{L}\left(\mathbb{C}^{d}\right)\right\}_{x \in X}$ by randomly choosing $|X|=10^{5},|X|=10^{6}$ and $|X|=$ $10^{7}$ unitary operators for $d=2, d=3$, and $d=4$, respectively. Note that the random sampling of unitary operators means the sampling probability distribution is induced by the Haar measure on $U(d)$. We compute a lower bound on $\epsilon$ as $\max _{i} \min _{x \in X} \frac{1}{2}\left\|\Upsilon_{i}-\Upsilon_{x}\right\|_{\text {。 }}$ by using 30 randomly chosen target unitary transformations $\left\{\Upsilon_{i}\right\}_{i=1}^{30}$. We interpret $\left\{\mathrm{Y}_{x}\right\}_{x \in X}$ as the set of available unitary transformations in probabilistic and deterministic approximation.

Next, we randomly choose a target unitary transformation $\Upsilon$ and compute $\epsilon_{\Upsilon}$. We define the actual approximation error caused by probabilistically mixing restricted available unitary transformations as

$$
\begin{equation*}
\epsilon_{\Upsilon}^{p r o b}\left(\epsilon^{\prime}\right):=\min _{p} \frac{1}{2}\left\|\Upsilon-\sum_{x \in \hat{X}\left(\epsilon^{\prime}\right)} p(x) \Upsilon_{x}\right\|_{\diamond} \tag{98}
\end{equation*}
$$

where $\hat{X}\left(\epsilon^{\prime}\right):=\left\{x \in X: \frac{1}{2}\left\|\Upsilon-\Upsilon_{x}\right\|_{\diamond} \leq \epsilon^{\prime}\right\}$. Note that $\epsilon_{\Upsilon}^{p r o b}\left(\epsilon^{\prime}\right)$ is a monotonically decreasing function. Moreover, if Lemma 5.3 holds for qudit unitary transformations, $\epsilon_{\mathrm{Y}}^{p r o b}(2 \epsilon)=\epsilon_{\mathrm{\Upsilon}}^{p r o b}(1)=\epsilon_{\mathrm{\Upsilon}}^{\text {prob }}$.

Third, we compute the actual approximation error caused by probabilistically mixing more restricted available unitary transformations as

$$
\begin{equation*}
\epsilon_{\Upsilon}^{p r o b}(N):=\min _{p} \frac{1}{2}\left\|\Upsilon-\sum_{x=1}^{N} p(x) \Upsilon_{x}\right\|_{\diamond}, \tag{99}
\end{equation*}
$$

where $\left\{\Upsilon_{x}\right\}_{x=1}^{N}$ is a randomly sampled subset of $\hat{X}\left(\epsilon^{\prime}\right)$ and $\epsilon^{\prime}$ is chosen large enough for $\epsilon_{\Upsilon}^{p r o b}\left(\epsilon^{\prime}\right)$ to converge.

## Results of numerical experiments

In Fig. 5, we draw the graphs of $\epsilon_{\Upsilon}^{\text {prob }}\left(\epsilon^{\prime}\right)$ for 10 randomly chosen $\Upsilon$ by using different colors corresponding to $\Upsilon$. We can observe that $\epsilon_{\Upsilon}^{p r o b}\left(\epsilon^{\prime}\right)$ is saturated when $\epsilon^{\prime} \geq 1.4 \epsilon$ and $\epsilon_{\Upsilon}^{p r o b}$ is comparable or smaller than $\epsilon_{\Upsilon}^{2} \operatorname{since} \frac{\log \left(\epsilon_{\mathrm{r}}^{\text {prob }}\left(\epsilon^{\prime}\right)\right)}{\log \left(\epsilon_{\mathrm{Y}}\right)} \geq$ $2 \Leftrightarrow \epsilon_{\mathrm{\Upsilon}}^{\text {prob }}\left(\epsilon^{\prime}\right) \leq \epsilon_{\mathrm{\Upsilon}}^{2}$ and $\epsilon_{\mathrm{\Upsilon}}^{\text {prob }} \leq \epsilon_{\mathrm{\Upsilon}}^{\text {prob }}\left(\epsilon^{\prime}\right)$ by definition.

In Fig. 6, we draw the graph of $\epsilon_{\Upsilon}^{p r o b}(N)$ for a randomly chosen $\Upsilon: \mathbf{L}\left(\mathbb{C}^{4}\right) \rightarrow \mathbf{L}\left(\mathbb{C}^{4}\right)$. Note that we plot its empirical variance and average since $\epsilon_{\mathrm{r}}^{p r o b}(N)$ is a random variable depending on the choice of a subset of $\hat{X}\left(\epsilon^{\prime}\right)$. We can observe that the approximation error $\epsilon_{\Upsilon}^{p r o b}(N)$ rapidly converges to its minimum $\epsilon_{\Upsilon}^{p r o b}\left(\epsilon^{\prime}\right)$. Therefore, we can expect that our probabilistic synthesis algorithm proposed in Theorem 5.4 can be made more efficient while still attaining an approximation error that is nearly optimal.

## Environment of numerical experiments

We performed numerical experiments using Mathematica 14.0.0.0 on a MacBook Pro equipped with a 2.4 GHz Intel Core i9 processor and 64 GB of memory.


Fig. 5. The comparison between the actual approximation error $\epsilon_{\Upsilon}^{p r o b}$ caused by the probabilistic approximation and that $\epsilon_{\Upsilon}$ caused by the deterministic approximation for 10 randomly sampled target unitary transformations $\Upsilon$. For both approximations, we use the set of available unitary transformations induced by an $\epsilon$-covering of $U(d)$. For each $\Upsilon$, we compute the actual approximation errors $\epsilon_{\Upsilon}^{p r o b}\left(\epsilon^{\prime}\right)$ caused by probabilistically mixing the set of available unitary transformations $\Upsilon_{x}$ such that $\frac{1}{2}\left\|\Upsilon-\Upsilon_{x}\right\|_{\diamond} \leq \epsilon^{\prime}$. The gray vertical lines represent lower bounds on $\epsilon$ and 1.4 .


Fig. 6. The approximation error caused by probabilistically mixing restricted subsets $\left\{\Upsilon_{x}\right\}_{x=1}^{N}$ of available unitary transformations. First, we randomly choose a target unitary transformation $\Upsilon$. Next, we randomly choose 8 such subsets from the set $\left\{\Upsilon_{x}: \frac{1}{2}\left\|\Upsilon-\Upsilon_{x}\right\|_{\diamond} \leq \epsilon^{\prime}\right\}$ of available unitary transformations that seem sufficient to attain the optimal approximation error caused by probabilistically mixing all the available unitary transformations. The blue graph represents the average of the approximation error and the gray region represents its variance.

