

Upper bound on entanglement distillation via Riemannian optimization

Chengkai Zhu*, Hongyu Mao†, Kun Fang† and Xin Wang*

*Thrust of Artificial Intelligence, Information Hub,

The Hong Kong University of Science and Technology (Guangzhou), Guangdong 511453, China

†School of Data Science, The Chinese University of Hong Kong, Shenzhen, Guangdong, 518172, China

Email: kunfang@cuhk.edu.cn, felixxinwang@hkust-gz.edu.cn

Abstract—Determining the ultimate limits of distillable entanglement of a shared state remains a central challenge in quantum information theory due to the phenomenon of superadditivity. This work develops Riemannian optimization methods to establish computable upper bounds on the one-way distillable entanglement. Our method systematically searches for state extensions that minimize known information-theoretic bounds. We achieve this by parameterizing the space of all possible extensions as a Stiefel manifold, enabling a universal search that overcomes the limitations of previous ad-hoc constructions. Combined with an improved upper bound on the one-way distillable entanglement based on a refined continuity bound on quantum conditional entropy, our approach yields new state-of-the-art upper bounds on the quantum capacity of the qubit depolarizing channel for large values of the depolarizing parameter, strictly improving the previously best-known bounds.

Index Terms—Distillable entanglement, Quantum capacity, Riemannian optimization, Stiefel manifold.

I. INTRODUCTION

Determining the ultimate physical limits on processing and transmitting information encoded in quantum systems is a primary goal of quantum information theory [1]. A vital problem in the field is quantifying the usable entanglement in a bipartite state ρ_{AB} , measured by the *distillable entanglement* $D(\rho_{AB})$ [2], [3]. This quantity represents the ultimate rate at which maximally entangled states can be extracted from independent and identically distributed (i.i.d.) copies of ρ_{AB} through local operations and classical communication (LOCC). When the communication is restricted to one-way transmission from Alice to Bob, we obtain the *one-way distillable entanglement*, denoted as $D_{\rightarrow}(\rho_{AB})$. Despite its fundamental importance, computing $D_{\rightarrow}(\cdot)$ for general states remains notoriously difficult. The quantity is defined via a regularization over many copies of the state:

$$D_{\rightarrow}(\rho_{AB}) = \lim_{n \rightarrow \infty} \frac{1}{n} D_{\rightarrow}^{(1)}(\rho_{AB}^{\otimes n}), \quad (1)$$

where $D_{\rightarrow}^{(1)}(\cdot)$ is the *single-shot* quantity based on the *coherent information*. This regularization is necessary because coherent information is generally *superadditive*, and makes direct computation intractable, necessitating the development of computable upper bounds.

Prominent strategies for bounding distillable entanglement include the Rains bound [4] and methods leveraging the concept of *degradability*. For degradable states, the one-way

distillable entanglement reduces to the single-letter coherent information. Consequently, continuity bounds based on *approximate degradability* have been established to handle general, non-degradable states [5]–[7]. A related approach involves decomposing a state into a convex combination of a degradable and an antidegradable part [6], [8].

Another intuitive and powerful strategy for deriving upper bounds is based on the construction of *state extensions*. This approach relies on the principle that for any extension $\hat{\rho}_{ABF}$ of a state ρ_{AB} (where $\text{Tr}_F \hat{\rho}_{ABF} = \rho_{AB}$), the monotonicity relation holds, i.e., $D_{\rightarrow}(\rho_{AB}) \leq D_{\rightarrow}(\hat{\rho}_{ABF})$. This follows directly from the data processing inequality, as the receiver can freely discard the auxiliary system F . The challenge thus shifts to finding an *optimal extension* that minimizes the distillable entanglement of the extended state. A leading technique for this purpose is the *flag extension* method [6], [9]–[11], which has yielded state-of-the-art (SOTA) upper bounds for highly symmetric states, such as isotropic states. However, the flag extension method faces a fundamental optimization bottleneck: the construction is typically tied to a specific convex decomposition of the state, and the subsequent optimization is performed over a fixed ansatz for the ‘flag’ systems. The efficacy of the method is therefore highly dependent on these initial choices, for which no systematic *a priori* guidelines exist.

In this work, we introduce a framework based on *Riemannian optimization* to tackle these challenges. We reframe the search for optimal state extensions as a tractable optimization problem over the complex Stiefel manifold. This enables a universal search that circumvents the limitations of ad-hoc flag constructions. Furthermore, we incorporate a recently improved continuity bound for conditional entropy [12], [13] to derive tighter bounds on one-way distillable entanglement in terms of approximate degradability.

II. PRELIMINARIES

A quantum system A is modeled by a finite-dimensional Hilbert space $\mathcal{H}_A$ whose dimension is denoted by $|A|$. The sets of linear operators and positive semidefinite operators on $\mathcal{H}_A$ are denoted by $\mathcal{L}(A)$ and $\mathcal{P}(A)$, respectively. The trace norm of X is given by $\|X\|_1 := \text{Tr} \sqrt{X^\dagger X}$, and its Frobenius norm is $\|X\|_F := \sqrt{\text{Tr} X^\dagger X}$. The set of *quantum states* on system A is denoted as $\mathcal{D}(A) := \{\rho \in \mathcal{P}(A) : \text{Tr} \rho = 1\}$. A

quantum channel is a completely positive and trace-preserving (CPTP) linear map from $\mathcal{L}(A)$ to $\mathcal{L}(B)$. A *subchannel* is a completely positive and trace non-increasing linear map from $\mathcal{L}(A)$ to $\mathcal{L}(B)$. We denote by $\text{CP}(A : B)$ the set of completely positive (CP) maps from A to B , and by $\text{CPTP}(A : B)$ the set of all CPTP maps. The *Choi matrix* of a linear map $\mathcal{N}_{A' \rightarrow B}$ is defined as $J_{\mathcal{N}} = (\mathcal{I}_A \otimes \mathcal{N}_{A' \rightarrow B})(|\Gamma\rangle\langle\Gamma|_{A'A})$, where $|\Gamma\rangle_{A'A} = \sum_i |i\rangle_{A'} |i\rangle_A$ is the unnormalized maximally entangled state. For every quantum channel $\mathcal{N}_{A \rightarrow B}$, there exists an isometry $V : \mathcal{H}_A \rightarrow \mathcal{H}_B \otimes \mathcal{H}_E$ such that $\mathcal{N}(\rho) = \text{Tr}_E(V\rho V^\dagger)$ where $\mathcal{H}_E$ is an environment space and V is called a *Stinespring isometry*. The *complementary channel* of $\mathcal{N}$ is denoted by $\mathcal{N}^c \in \text{CPTP}(A : E)$ that acts as $\mathcal{N}^c(\rho) = \text{Tr}_B(V\rho V^\dagger)$.

The von Neumann entropy of a quantum state is defined as $H(\rho) := -\text{Tr}(\rho \log \rho)$ where all logarithms in this work are taken in base two. The coherent information of a bipartite quantum state ρ_{AB} is defined by $I(A)_B)_\rho := H(\rho_B) - H(\rho_{AB})$ and the conditional entropy by $H(A|B)_\rho := -I(A)_B)_\rho$. For $p \in [0, 1]$, the binary entropy is defined by $h(p) := -p \log p - (1-p) \log(1-p)$ and the bosonic entropy is defined by $g(p) := (1+p) h(p/(1+p))$.

III. MAIN RESULTS

a) Extensions of quantum states: A common strategy for constructing extensions of states is referred to as the *flag extension* method, which relies on a convex decomposition of the state [6], [9]–[11]. For a given bipartite quantum state ρ_{AB} , one first decomposes it into $\rho_{AB} = \sum_i p_i \omega_{AB}^{(i)}$ where $\{\omega_{AB}^{(i)}\}_i$ are normalized states. An extension is then constructed by appending an auxiliary *flag* state $\tau_F^{(i)}$ to each $\omega_{AB}^{(i)}$, i.e.,

$$\hat{\rho}_{ABF} := \sum_i p_i \omega_{AB}^{(i)} \otimes \tau_F^{(i)}. \quad (2)$$

By construction, tracing out the flag system F recovers the original state. The central challenge of this method is that the subsequent optimization over the flag state ensemble $\{\tau_F^{(i)}\}_i$ is fundamentally tied to the initial choice of decomposition. To establish a more general framework that avoids this dependency, we begin by formally defining state extensions.

Definition 1 (State extension). *For any given bipartite quantum state ρ_{AB} , a quantum state $\hat{\rho}_{ABF} \in \mathcal{D}(ABF)$ is called an extended state of ρ_{AB} if $\text{Tr}_F \hat{\rho}_{ABF} = \rho_{AB}$.*

We have the following fact about any state extension.

Fact 1. *For any bipartite quantum state ρ_{AB} , $\hat{\rho}_{ABF}$ is an extended state of ρ_{AB} if and only if there is an environment system R and an isometry $V : \mathcal{H}_E \rightarrow \mathcal{H}_R \otimes \mathcal{H}_F$ such that $\hat{\rho}_{ABF} = \text{Tr}_R(V|\phi\rangle\langle\phi|V^\dagger)$ where $|\phi\rangle_{ABE}$ is a purification of ρ_{AB} .*

b) Riemannian optimization for extensions: Based on the fact above, the key insight of our work is that the space of all valid extensions for a given state can be endowed with a *Riemannian manifold* structure. Consequently, the problem of optimizing a chosen figure of merit over this space can be cast as a Riemannian optimization problem. For a given bipartite

quantum state ρ_{AB} , suppose we are interested in minimizing a function $f(\rho_{ABF})$ where ρ_{ABF} is any extension of ρ_{AB} . Then using Fact 1, for any fixed $|F|, |R| \in \mathbb{N}_+$, taking $\rho_{ABF} = \text{Tr}_R(V|\phi\rangle\langle\phi|V^\dagger)$ into the objective, the problem

$$\min_{V \in \text{St}(|FR|, |E|)} f(V) \quad (3)$$

reduces to a constrained optimization problem over the isometries. The feasible region of these isometries is the *complex Stiefel manifold* $\text{St}(|FR|, |E|) := \{V \in \mathbb{C}^{|FR| \times |E|} : V^\dagger V = \mathbb{1}\}$, a well-studied Riemannian manifold [14]. This naturally leads us to the framework of *Riemannian optimization*, which leverages the geometric structure of the manifold to generate iterates that remain within the feasible set [15]. Optimization problems on Riemannian manifolds have appeared in various areas of quantum information sciences (see, e.g., Refs. [16]–[20]).

A. Upper bounds on the one-way distillable entanglement

For the task of entanglement distillation from n i.i.d. copies of a bipartite state ρ_{AB} , the objective is to transform the initial state $\rho_{AB}^{\otimes n}$ into m copies of a Bell state Φ via LOCC, such that the fidelity with the target state $\Phi^{\otimes m}$ approaches one in the asymptotic limit ($n \rightarrow \infty$). When the classical communication is restricted from Alice to Bob, the protocol class is known as *one-way LOCC*. The corresponding optimal rate of distillation is the *one-way distillable entanglement*, denoted $D_{\rightarrow}(\rho_{A|B})$, where the notation $A|B$ emphasizes that the restricted communication direction (only from A to B is possible). A seminal work by Devetak and Winter [21] shows that

$$D_{\rightarrow}(\rho_{A|B}) = \lim_{n \rightarrow \infty} \frac{1}{n} D_{\rightarrow}^{(1)}(\rho_{A|B}^{\otimes n}), \quad (4)$$

where $D_{\rightarrow}^{(1)}(\rho_{A|B})$ can be expressed as $D_{\rightarrow}^{(1)}(\rho_{A|B}) = \max_{\mathcal{T}} I(A')_B M)_{\mathcal{T}(\rho_{AB})}$. Here $\mathcal{T}$ is a quantum instrument, i.e., $\mathcal{T} : A \rightarrow A'M$, $\mathcal{T}(\rho_A) := \sum_{j=1}^m T_j(\rho_A) \otimes |j\rangle\langle j|$, where for each j , $T_j \in \text{CP}(A : A')$ and $\sum_j T_j \in \text{CPTP}(A : A')$, and M is a classical system to store measurement results with $|M| = m$.

A bipartite state ρ_{AB} is termed *degradable* if there exists a quantum channel $\mathcal{M}_{B \rightarrow E}$ such that $\mathcal{M}_{B \rightarrow E}(\rho_{AB}) = \text{Tr}_B |\phi\rangle\langle\phi|_{ABE}$, where $|\phi\rangle_{ABE}$ is a purification of ρ_{AB} [6]. The significance of this property is that the coherent information is additive for any degradable state, which simplifies the regularized expression for one-way distillable entanglement, i.e., $D_{\rightarrow}(\rho_{AB}) = I(A)_B)_\rho$ [6]. In the same work, the approximate degradability was introduced as follows.

Definition 2 (Degradability parameter of bipartite states [6]). *For $\rho \in \mathcal{D}(AB)$ with a purification $|\rho\rangle_{ABE}$, it is called ε -degradable if $\exists \mathcal{M} \in \text{CPTP}(B : E')$ such that $E \cong E'$ and $\frac{1}{2} \|\rho_{AE} - \mathcal{M}_{B \rightarrow E'}(\rho_{AB})\|_1 \leq \varepsilon$. The channel $\mathcal{M}$ is referred to as an ε -degrading channel of ρ_{AB} . The degradability parameter of ρ_{AB} is defined as $\text{dg}(\rho) := \min_{\mathcal{M} \in \text{CPTP}(B : E')} \frac{1}{2} \|\rho_{AE} - \mathcal{M}_{B \rightarrow E'}(\rho_{AB})\|_1$.*

Notably, the degradability parameter of a given state can be efficiently computed via semidefinite programming (SDP) [6, Lemma 2.10]. Based on the approximate degradability, an efficiently computable upper bound on the one-way distillable entanglement was established in [6, Theorem 2.12]. We remark that this upper bound is essentially also based on the continuity bound of the conditional entropy, known as the Alicki-Fannes inequality [22], and was improved by Winter later [23]. Recently, this continuity bound has been further improved with an additional marginal restriction on the considered states in [12, Theorem 5] and [13, Theorem 5].

Lemma 1 (Continuity of the conditional entropy [12], [13]). *For $\rho, \sigma \in \mathcal{D}(AB)$ satisfying $\rho_B = \sigma_B$ and $\frac{1}{2}\|\rho - \sigma\|_1 = \varepsilon$ for some $\varepsilon \in [0, 1]$, it holds that $|H(A|B)_\rho - H(A|B)_\sigma| \leq \varepsilon \log(|A|^2 - 1) + h(\varepsilon)$.*

Building upon the improved continuity bound on conditional entropy, we improve the upper bound on the one-way distillable entanglement in Theorem 1. A parallel result for the upper bound on quantum capacity has been shown in [13, Proposition 10].

Theorem 1 (Improved upper bound on the one-way distillable entanglement). *Let $\rho \in \mathcal{D}(AB)$ be an ε -degradable quantum state with ε -degrading channel $\mathcal{M} \in \text{CPTP}(B : E')$. Let $V : \mathcal{H}_B \rightarrow \mathcal{H}_{E'} \otimes \mathcal{H}_G$ be a Stinespring isometry of $\mathcal{M}$ with $E' \cong E$. It satisfies that*

$$D_{\rightarrow}(\rho_{A|B}) \leq I(A|B)_\rho + 2\varepsilon \log(|E|^2 - 1) + 2h(\varepsilon). \quad (5)$$

The proof closely follows that of [6, Theorem 2.12], leveraging Lemma 1. Based on Theorem 1 and the framework of Riemannian optimization for state extensions, we have the following upper bound on the one-way distillable entanglement.

Corollary 2. *For any bipartite state ρ_{AB} , let $|\phi\rangle_{ABE} = \sum_i \sqrt{p_i} |\psi_i\rangle_{AB} \otimes |e_i\rangle_E$ be the canonical purification of ρ_{AB} where $\rho_{AB} = \sum_i p_i |\psi_i\rangle\langle\psi_i|_{AB}$ is the spectral decomposition,*

$$\begin{aligned} D_{\rightarrow}(\rho_{A|B}) &\leq \inf I(A|BF)_\rho + 2\varepsilon \log(|\hat{E}|^2 - 1) + 2h(\varepsilon) \\ \text{s.t. } V &\in \text{St}(|FR|, |E|), |F|, |R| \in \mathbb{N}_+, \\ \rho_{ABF} &= \text{Tr}_R(V|\phi\rangle\langle\phi|V^\dagger), \varepsilon = \text{dg}(\rho_{A|BF}), \end{aligned} \quad (6)$$

where $\hat{E}$ is the auxiliary system for the purification of ρ_{ABF} .

Proof. The bound directly follows from Fact 1 with Theorem 1 applying to $\rho_{A|BF}$. $\square$

For any fixed $|F|, |R| \in \mathbb{N}_+$, taking $\rho_{ABF} = \text{Tr}_R(V|\phi\rangle\langle\phi|V^\dagger)$ into the objective in Eq. (6), the problem reduces to $\min_{V \in \text{St}(|FR|, |E|)} f(V) := I(A|BF)_{\rho(V)} + 2 \text{dg}(V) \log(\eta^2 - 1) + 2h(\text{dg}(V))$. Here, $I(A|BF)_{\rho(V)}$ and $\text{dg}(V)$ are shorthand for $I(A|BF)_{\rho(V)}$ and $\text{dg}(\rho_{A|BF}(V))$, respectively, and $\eta = |ABF|$. This is because full-rank matrices are dense in $\mathcal{D}(ABF)$, i.e., the probability of an iterative point landing on a full-rank point is 1 under the Lebesgue measure. The primary obstacle to solving this problem is that deriving an analytical expression for the Riemannian gradient is intractable. This difficulty arises because the

degradability parameter, $\text{dg}(\rho_{A|BF})$, and ε -degrading channel $\mathcal{M}$ themselves are the solution to an SDP [6, Lemma 2.10], making Eq. (6) a *bilevel* optimization problem [24]. To obtain an analytical gradient, in principle, one may try to use implicit differentiation of the Karush-Kuhn-Tucker (KKT) conditions [25] of the inner SDP for $\text{dg}(\rho_{A|BF})$, and analyze the gradient of purification function since we also need ρ_{AE} for calculating $\text{dg}(\rho_{A|BF})$. This approach is exceptionally complex to implement and requires solving a large, potentially ill-conditioned linear system at each iteration. For these practical reasons, we adopt a numerical finite difference scheme.

a) Numerical gradient approximation on the manifold: Since our primary goal is to establish a method that can be practically implemented to compute upper bounds for a wide range of quantum states and channels, we prioritize a pragmatic approach to optimization. We utilize an RGD algorithm but replace the intractable analytical gradient with a numerical approximation via a finite-difference scheme, a widely adopted method in derivative-free optimization [26], [27], which has been shown to be effective for some bilevel problems [28]. To apply gradient-based optimization to the objective function in Eq. (6), we will show that the following objective function is actually *locally Lipschitz continuous*.

$$f(V) = I(A|BF)_{\rho(V)} + 2 \text{dg}(V) \log(\eta^2 - 1) + 2h(\text{dg}(V)). \quad (7)$$

For the first term $I(A|B)_\rho$, we note that the von Neumann entropy $H(\rho)$ is continuously differentiable with respect to ρ at any point where ρ is full-rank, i.e., $\nabla H(\rho) = -(\log \rho + \mathbb{1})$. Therefore, $I(A|B)_\rho$ is also continuously differentiable on the full rank density matrices, which means that $\rho \mapsto I(A|B)_\rho$ is locally Lipschitz continuous. The binary entropy term $h(\cdot)$ is also locally Lipschitz. For the second term, the degradability parameter, we have the following results.

Lemma 3. *For any $\rho_{AB}, \sigma_{AB} \in \mathcal{D}(AB)$ such that $\frac{1}{2}\|\rho_{AB} - \sigma_{AB}\|_1 \leq \varepsilon \leq 1$, $|\text{dg}(\rho_{AB}) - \text{dg}(\sigma_{AB})| \leq \sqrt{2\varepsilon} + \varepsilon$.*

Proof. Choose $|\phi^\rho\rangle_{ABE}$ and $|\phi^\sigma\rangle_{ABE}$ to be purifications of ρ_{AB} and σ_{AB} on a common system $\mathcal{H}_E$ with $|E| \geq \dim(\mathcal{H}_A) \dim(\mathcal{H}_B)$, such that they saturate the fidelity bound in Uhlmann's theorem, i.e., $F(\rho_{AB}, \sigma_{AB}) = |\langle\phi^\rho|\phi^\sigma\rangle|^2$, where $F(\rho, \sigma) := \|\sqrt{\rho}\sqrt{\sigma}\|_1^2$ is the fidelity between ρ and σ . The complementary states are given by $\rho_{AE} = \text{Tr}_B |\phi^\rho\rangle\langle\phi^\rho|$ and $\sigma_{AE} = \text{Tr}_B |\phi^\sigma\rangle\langle\phi^\sigma|$. Let us define the function $f_{\mathcal{M}}(\rho_{AB})$ as $f_{\mathcal{M}}(\rho_{AB}) := \frac{1}{2}\|\rho_{AE} - (\mathcal{I}_A \otimes \mathcal{M}_{B \rightarrow E})(\rho_{AB})\|_1$. Consider that

$$\begin{aligned} &|f_{\mathcal{M}}(\rho_{AB}) - f_{\mathcal{M}}(\sigma_{AB})| \\ &\leq \frac{1}{2} \|\rho_{AE} - (\mathcal{I}_A \otimes \mathcal{M})(\rho_{AB}) - \sigma_{AE} + (\mathcal{I}_A \otimes \mathcal{M})(\sigma_{AB})\|_1 \\ &\leq \frac{1}{2} \left(\|\rho_{AE} - \sigma_{AE}\|_1 + \|(\mathcal{I}_A \otimes \mathcal{M})(\rho_{AB} - \sigma_{AB})\|_1 \right) \\ &\leq \frac{1}{2} \left(\|\rho_{AE} - \sigma_{AE}\|_1 + \|\rho_{AB} - \sigma_{AB}\|_1 \right). \end{aligned}$$

Here, we used the triangle inequality for the trace norm in the first and second inequalities, and the last inequality is followed by data processing. Recall the Fuchs–van de Graaf inequalities

$$1 - \sqrt{F(\rho_{AB}, \sigma_{AB})} \leq \frac{1}{2} \|\rho_{AB} - \sigma_{AB}\|_1 \leq \sqrt{1 - F(\rho_{AB}, \sigma_{AB})}. \quad (8)$$

We have that $\|\rho_{AE} - \sigma_{AE}\|_1 \leq 2\sqrt{1 - F(\rho_{AB}, \sigma_{AB})}$, where we used the monotonicity of the trace distance under partial trace and the fact that any pair of pure states saturates the upper bound in Eq. (8). Notice that

$$\begin{aligned} \sqrt{1 - F(\rho_{AB}, \sigma_{AB})} &\leq \sqrt{2(1 - \sqrt{F(\rho_{AB}, \sigma_{AB})})} \\ &\leq \sqrt{\|\rho_{AB} - \sigma_{AB}\|_1}, \end{aligned} \quad (9)$$

where the first inequality is because $F(\rho_{AB}, \sigma_{AB}) \leq 1$ and the second inequality is followed from Eq. (8). It follows that

$$\|\rho_{AE} - \sigma_{AE}\|_1 \leq 2\sqrt{\|\rho_{AB} - \sigma_{AB}\|_1}. \quad (10)$$

Now, let $\mathcal{M}_\rho$ be a CPTP map that achieves the minimum for $\text{dg}(\rho_{AB})$, and let $\mathcal{M}_\sigma$ achieve the minimum for $\text{dg}(\sigma_{AB})$. By definition, we have $\text{dg}(\rho_{AB}) = f_{\mathcal{M}_\rho}(\rho_{AB})$, $\text{dg}(\sigma_{AB}) = f_{\mathcal{M}_\sigma}(\sigma_{AB})$. Since $\mathcal{M}_\rho$ is optimal for ρ_{AB} , we have $\text{dg}(\rho_{AB}) \leq f_{\mathcal{D}}(\rho_{AB})$ for any map $\mathcal{D}$, including $\mathcal{M}_\sigma$. This leads to

$$\begin{aligned} \text{dg}(\rho_{AB}) - \text{dg}(\sigma_{AB}) &\leq f_{\mathcal{M}_\sigma}(\rho_{AB}) - f_{\mathcal{M}_\sigma}(\sigma_{AB}) \\ &\leq |f_{\mathcal{M}_\sigma}(\rho_{AB}) - f_{\mathcal{M}_\sigma}(\sigma_{AB})| \\ &\leq \frac{1}{2} \left(2\sqrt{\|\rho_{AB} - \sigma_{AB}\|_1} + \|\rho_{AB} - \sigma_{AB}\|_1 \right), \end{aligned}$$

By swapping the roles of ρ_{AB} and σ_{AB} , we conclude that

$$|\text{dg}(\rho_{AB}) - \text{dg}(\sigma_{AB})| \leq \frac{1}{2} \left(2\sqrt{\|\rho_{AB} - \sigma_{AB}\|_1} + \|\rho_{AB} - \sigma_{AB}\|_1 \right). \quad \square$$

Lemma 4. For any bipartite state ρ_{AB} , let $|\phi\rangle_{ABE}$ be a purification of ρ_{AB} , $V : \mathcal{H}_E \rightarrow \mathcal{H}_R \otimes \mathcal{H}_F$ be an isometry and denote $\rho_{ABF}(V) = \text{Tr}_R(V|\phi\rangle\langle\phi|V^\dagger)$. It holds that

$$\|\rho_{ABF}(V) - \rho_{ABF}(V')\|_1 \leq 2\|V - V'\|_F. \quad (11)$$

Corollary 5. For any bipartite state ρ_{AB} , let $|\phi\rangle_{ABE}$ be a purification of ρ_{AB} , $V : \mathcal{H}_E \rightarrow \mathcal{H}_R \otimes \mathcal{H}_F$ be an isometry and denote $\rho_{ABF}(V) = \text{Tr}_R(V|\phi\rangle\langle\phi|V^\dagger)$. It holds that

$$|\text{dg}(\rho_{ABF}(V)) - \text{dg}(\rho_{ABF}(V'))| \leq 2\|V - V'\|_F. \quad (12)$$

Lemma 3 establishes the Hölder continuity of the degradability parameter $\text{dg}(\cdot)$ with respect to states. More importantly, Lemma 4 and Corollary 5 demonstrate that the degradability parameter $\text{dg}(\cdot)$ is Lipschitz continuous with respect to isometries V . Consequently, the overall objective function $f(V)$ in Eq. (7) is locally Lipschitz continuous. This property is the theoretical cornerstone of our optimization approach. Specifically, although $f(V)$ may be nondifferentiable, its local Lipschitz nature justifies the use of concepts from nonsmooth optimization, namely the subgradient [29]. In this context, our finite-difference scheme serves as a reliable method to approximate a subgradient, as the local Lipschitz condition guarantees its truncation error is rigorously bounded linearly

by $\mathcal{O}(Lh)$, where L is the local Lipschitz constant and h is the tunable step size [30]. Therefore, our algorithm acts as a practical implementation of a (inexact) Riemannian subgradient method, a principled approach known to converge to the set of Clarke stationary points [31], [32]. Based on the above, we explain the optimization algorithm as follows.

Algorithm 1 Riemannian Gradient Descent (RGD) for upper bounds on $D_{\rightarrow}(\cdot)$

Input: A bipartite state ρ_{AB} and a purification $|\phi\rangle_{ABE}$; Initial guess $V^{(0)} \in \text{St}(n, p)$.

- 1: **while** the stopping criteria are not satisfied **do**
- 2: Approximate $\dot{V}^{(t)} = -\text{grad} f(V_k)$ using the finite-difference scheme.
- 3: Compute a stepsize $s^{(t)}$.
- 4: Update $V^{(t+1)} = R_{V^{(t)}}(s^{(t)}\dot{V}^{(t)})$ by Eq. (13); $t = t + 1$.
- 5: **end while**
- 6: Compute the upper bound d_{upp} in Eq. (6) using the optimal V^* .

Output: V^*, d_{upp} .

b) Riemannian gradient descent: Our RGD algorithm, shown in Algorithm 1, leverages standard tools from the field of Riemannian optimization. To ensure our presentation is self-contained, we briefly recall the essential definitions for the complex Stiefel manifold, which are well-established concepts [33]. The Riemannian gradient, denoted $\text{grad} f(V)$, is a tangent vector in the tangent space $T_V \mathcal{M}$ at a point V on the manifold $\mathcal{M}$. It is defined by the relation $\langle \text{grad} f(V), \dot{V} \rangle_V = Df(V)[\dot{V}]$, $\forall \dot{V} \in T_V \mathcal{M}$, where $\dot{V}$ is any tangent vector, $\langle \cdot, \cdot \rangle_V$ is the Riemannian metric on the *tangent space*, and $Df(V)[\dot{V}]$ is the directional derivative of f at V along $\dot{V}$.

Specifically, for the *complex Stiefel manifold* $\text{St}(n, p) := \{V \in \mathbb{C}^{n \times p} : V^\dagger V = \mathbb{1}\}$, its tangent space is given by $T_V \text{St}(n, p) = \{VA : A \in \mathbb{C}^{n \times n}, A^\dagger + A = 0\}$. The Riemannian metric on $T_V \text{St}(n, p)$ we use is given by $\langle \cdot, \cdot \rangle_V : T_V \text{St}(n, p) \times T_V \text{St}(n, p) \rightarrow \mathbb{R}$, $\langle X, Y \rangle_V = \text{Tr}(X^\dagger Y)$. The projection onto the tangent space at any point $V \in \text{St}(n, p)$ reads $\text{Proj}_V(X) = (\mathbb{1} - VV^\dagger)X + V \frac{V^\dagger X - X^\dagger V}{2}$. The Riemannian gradient is then given by $\text{grad} f(V) = \text{Proj}_V(\nabla \bar{f}(V))$, where $\nabla \bar{f}(V)$ is the gradient of an extension $\bar{f}$ of f on Euclidean space. When doing an optimization on a manifold, in order to move in the direction of a tangent vector while staying on the manifold, we also need a retraction mapping [15, Section 4]. There are different retractions for the Stiefel manifold, a standard one, and the one we used is the retraction based on the QR decomposition, i.e.,

$$R_V(X) := \text{qf}(V + X), \quad (13)$$

where $\text{qf}(X)$ denotes the Q factor of the decomposition of X as $X = QR$ where $Q \in \text{St}(n, p)$ and R is an $p \times p$ upper triangular with nonnegative diagonal entries.

To optimize Eq. (7), since we cannot compute the gradient analytically, we approximate the directional derivative using a

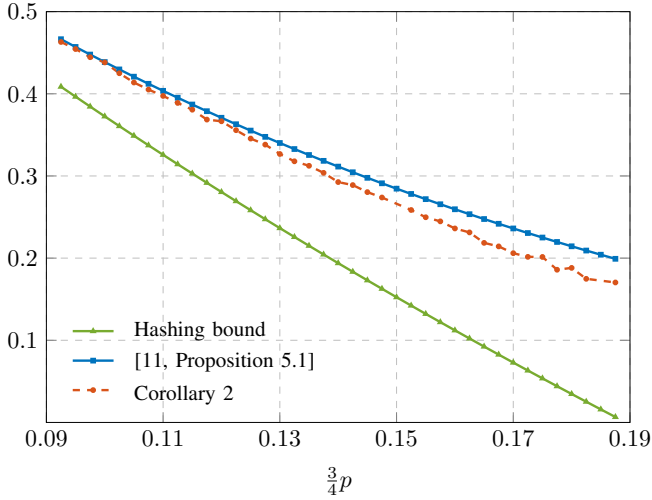


Fig. 1: Bounds on the quantum capacity of the qubit depolarizing channel. The x -axis is the parameter $3/4p = 1 - f \in [0.0925, 0.1875]$ of the qubit depolarizing channel $\mathcal{D}_p$.

first-order finite difference. For a small step t , the directional derivative is approximated as $Df(V)[\dot{V}] \approx \frac{f(R_V(t\dot{V})) - f(V)}{t}$. To compute the full gradient vector $\text{grad } f(V)$, we can compute the following steps:

- i). Define an orthonormal basis $\{\zeta_i\}_{i=1}^d$ for the tangent space $T_V \mathcal{M}$, where $d = \dim(T_V \mathcal{M})$.
- ii). For each basis vector ζ_i , compute the approximate directional derivative using the finite-difference formula $c_i = Df(V)[\zeta_i] \approx \frac{f(R_V(t\zeta_i)) - f(V)}{t}$.
- iii). The Riemannian gradient is then constructed by combining these components $\text{grad } f(V) = \sum_{i=1}^d c_i \zeta_i$.

This process requires $d + 1$ function evaluations to compute a single gradient. While computationally intensive, it bypasses the need for any analytical derivatives.

B. Improved upper bounds on isotropic states and depolarizing channels

In this section, we establish improved upper bounds on the one-way distillable entanglement of isotropic states using our proposed methods. As a consequence, we obtain the SOTA upper bounds on the quantum capacity of qubit depolarizing channels. For a local dimension d , the *isotropic state* is defined by $I_d(f) := f\Phi_+ + \frac{1-f}{d^2-1}(\mathbb{1}_{d^2} - \Phi_+)$, where Φ_+ is the $d \times d$ maximally entangled state and $f \in [0, 1]$. It is also known as the Choi state of the qudit depolarizing channel $\mathcal{D}_p(\cdot) = (1-p)(\cdot) + \frac{p}{d^2-1} \sum_{0 \leq i, j \leq d-1} X^i Z^j (\cdot) (X^i Z^j)^\dagger$, where $p = 1 - f$ and X, Z are the generalized Pauli operators defined by $X|k\rangle := |k+1 \pmod{d}\rangle$, $Z|k\rangle := \omega^k|k\rangle$, $\omega = \exp(2\pi i/d)$.

Notably, the quantum capacity of a teleportation-simulable channel is equal to the one-way distillable entanglement of its Choi state [34]–[36], i.e., $Q(\mathcal{D}_{1-f}) = D_{\rightarrow}(I_d(f))$. Therefore, the upper bound on $D_{\rightarrow}(I_d(f))$ also gives an upper bound on the quantum capacity of depolarizing channels, $Q(\mathcal{D}_{1-f})$. The previously best-known upper bound on $Q(\mathcal{D}_p)$ is given

by the flagged extension method [11, Proposition 5.1 and Figure 1]. Here, based on Corollary 2 and the numerical optimization method on a Stiefel manifold, we obtained SOTA upper bounds on the quantum capacity of qubit depolarizing channels. Specifically, for qubit isotropic states, we have $|E| = 4$ and choose $|R| = 4, |F| = 2$ in our numerical experiments. Thereby, we are performing optimization over $\text{St}(8, 4)$ for the extensions of isotropic states ρ_{ABF} , with a qubit extension system F . The comparison of our bound with the best-known ones is presented in Figure 1. The red dashed line is the upper bound given by the optimized extensions of the isotropic states, according to Corollary 2. The blue solid line corresponds to the upper bound given by [11, Proposition 5.1], using flag extensions. We can observe a notable improvement in the given noise range.

We emphasize that all our numerical experiments leverage high-precision solvers where the primal-dual gap for all semidefinite programs is consistently driven below 10^{-14} , and all other entropy function calculations are performed at standard double-precision machine epsilon ($\approx 10^{-16}$). This high degree of accuracy ensures that the reported gains are significant and reliable. All state extensions as well as corresponding isometries (cf. Fact 1) we have obtained for the qubit isotropic states, which yield the SOTA upper bounds, can be found in [37].

IV. CONCLUDING REMARKS

In this work, we have developed a framework based on Riemannian optimization to derive an improved computable upper bound on the one-way distillable entanglement. We transformed the search for state extensions that can minimize information-theoretic bounds into a tractable optimization over the Stiefel manifold. This method, when combined with a refined continuity bound for conditional entropy, delivers strictly tighter upper bounds than previously known records for large values of the depolarizing parameter [11].

The methods presented here open several avenues for future research. For the upper-bound framework, it would be of great interest to see if the structure of the optimal isometries found via optimization can inspire new analytical insights or closed-form bounds for specific channel families. The computational cost of the finite-difference gradient approximation could be reduced by exploring more advanced non-smooth or bilevel optimization techniques tailored to the Stiefel manifold [28], [38].

REFERENCES

- [1] M. M. Wilde, *Quantum information theory*. Cambridge university press, 2013.
- [2] C. H. Bennett, G. Brassard, S. Popescu, B. Schumacher, J. A. Smolin, and W. K. Wootters, “Purification of Noisy Entanglement and Faithful Teleportation via Noisy Channels,” *Physical Review Letters*, vol. 76, no. 5, pp. 722–725, jan 1996. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.76.722>
- [3] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, “Concentrating partial entanglement by local operations,” *Physical Review A*, vol. 53, no. 4, pp. 2046–2052, apr 1996. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevA.53.2046>

- [4] E. Rains, "A semidefinite program for distillable entanglement," *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 2921–2933, 2001. [Online]. Available: <http://ieeexplore.ieee.org/document/959270/>
- [5] D. Sutter, V. B. Scholz, A. Winter, and R. Renner, "Approximate Degradable Quantum Channels," *IEEE Transactions on Information Theory*, vol. 63, no. 12, pp. 7832–7844, dec 2017. [Online]. Available: <http://arxiv.org/abs/1412.0980> <http://ieeexplore.ieee.org/document/8046086/>
- [6] F. Leditzky, N. Datta, and G. Smith, "Useful states and entanglement distillation," *IEEE Transactions on Information Theory*, vol. 64, no. 7, p. 4689–4708, Jul. 2018. [Online]. Available: <http://dx.doi.org/10.1109/TIT.2017.2776907>
- [7] M. G. Jabbour and N. Datta, "Tightening continuity bounds for entropies and bounds on quantum capacities," *IEEE Journal on Selected Areas in Information Theory*, 2024.
- [8] C. Zhu, C. Zhu, and X. Wang, "Estimate distillable entanglement and quantum capacity by squeezing useless entanglement," *IEEE Journal on Selected Areas in Communications*, vol. 42, no. 7, p. 1850–1860, Jul. 2024. [Online]. Available: <http://dx.doi.org/10.1109/JSAC.2024.3380081>
- [9] X. Wang, "Pursuing the fundamental limits for quantum communication," *IEEE Transactions on Information Theory*, vol. 67, no. 7, pp. 4524–4532, 2021.
- [10] M. Fanizza, F. Kianvash, and V. Giovannetti, "Quantum flags, and new bounds on the quantum capacity of the depolarizing channel," *Physical Review Letters*, vol. 125, no. 2, p. 020503, Jul. 2020.
- [11] F. Kianvash, M. Fanizza, and V. Giovannetti, "Bounding the quantum capacity with flagged extensions," *Quantum*, vol. 6, p. 647, Feb. 2022.
- [12] K. Audenaert, B. Bergh, N. Datta, M. G. Jabbour, Á. Capel, and P. Gondolf, "Continuity bounds for quantum entropies arising from a fundamental entropic inequality," Dec. 2024.
- [13] M. Berta, L. Lami, and M. Tomamichel, "Continuity of entropies via integral representations," *IEEE Transactions on Information Theory*, vol. 71, no. 3, pp. 1896–1908, Mar. 2025.
- [14] M. Atiyah and J. Todd, "On complex stiefel manifolds," in *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 56, no. 4. Cambridge University Press, 1960, pp. 342–353.
- [15] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*. Princeton University Press, 2008.
- [16] M.-C. Hsu, E.-J. Kuo, W.-H. Yu, J.-F. Cai, and M.-H. Hsieh, "Quantum state tomography via nonconvex riemannian gradient descent," *Physical Review Letters*, vol. 132, p. 240804, Jun 2024. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.132.240804>
- [17] X. Zhu, C. Zhang, Z. An, and B. Zeng, "Unified framework for calculating convex roof resource measures," *npj Quantum Information*, vol. 11, no. 1, Apr. 2025. [Online]. Available: <http://dx.doi.org/10.1038/s41534-025-01012-1>
- [18] I. N. M. Le, S. Sun, and C. B. Mendl, "Riemannian quantum circuit optimization based on matrix product operators," *Quantum*, vol. 9, p. 1833, Aug. 2025. [Online]. Available: <https://doi.org/10.22331/q-2025-08-27-1833>
- [19] C. Zhu, R. Peng, B. Gao, and X. Wang, "Riemannian optimization for holevo capacity," 2025. [Online]. Available: <https://arxiv.org/abs/2501.11576>
- [20] Z.-T. Li, X.-L. He, C.-C. Zheng, Y.-Q. Dong, T. Luan, X.-T. Yu, and Z.-C. Zhang, "Quantum comb tomography via learning isometries on stiefel manifold," *Physical Review Letters*, vol. 134, p. 010803, Jan. 2025. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.134.010803>
- [21] I. Devetak and A. Winter, "Distillation of secret key and entanglement from quantum states," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 461, no. 2053, pp. 207–235, Jan. 2005.
- [22] R. Alicki and M. Fannes, "Continuity of quantum conditional information," *Journal of Physics A: Mathematical and General*, vol. 37, no. 5, p. L55–L57, Jan. 2004. [Online]. Available: <http://dx.doi.org/10.1088/0305-4470/37/5/L01>
- [23] A. Winter, "Tight Uniform Continuity Bounds for Quantum Entropies: Conditional Entropy, Relative Entropy Distance and Energy Constraints," *Communications in Mathematical Physics*, vol. 347, no. 1, pp. 291–313, Oct. 2016.
- [24] B. Colson, P. Marcotte, and G. Savard, "An overview of bilevel optimization," *Annals of operations research*, vol. 153, no. 1, pp. 235–256, 2007.
- [25] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM review*, vol. 38, no. 1, pp. 49–95, 1996.
- [26] H. Mizumoto, "A finite-difference method on a riemann surface," *Hiroshima Mathematical Journal*, vol. 3, no. 2, pp. 277–332, 1973.
- [27] T. D. Choi and C. T. Kelley, "Superlinear convergence and implicit filtering," *SIAM Journal on Optimization*, vol. 10, no. 4, pp. 1149–1162, 2000. [Online]. Available: <https://doi.org/10.1137/S1052623499354096>
- [28] D. Sow, K. Ji, and Y. Liang, "On the convergence theory for hessian-free bilevel algorithms," 2022. [Online]. Available: <https://arxiv.org/abs/2110.07004>
- [29] X. Li, S. Chen, Z. Deng, Q. Qu, Z. Zhu, and A. Man-Cho So, "Weakly convex optimization over stiefel manifold using riemannian subgradient-type methods," *SIAM Journal on Optimization*, vol. 31, no. 3, pp. 1605–1634, 2021. [Online]. Available: <https://doi.org/10.1137/20M1321000>
- [30] M. I. Jordan, G. Kornowski, T. Lin, O. Shamir, and M. Zampetakis, "Deterministic nonsmooth nonconvex optimization," 2025. [Online]. Available: <https://arxiv.org/abs/2302.08300>
- [31] J. V. Burke, A. S. Lewis, and M. L. Overton, "A robust gradient sampling algorithm for nonsmooth, nonconvex optimization," *SIAM Journal on Optimization*, vol. 15, no. 3, pp. 751–779, 2005.
- [32] S. Hosseini and A. Uschmajew, "A riemannian gradient sampling algorithm for nonsmooth optimization on manifolds," *SIAM Journal on Optimization*, vol. 27, no. 1, pp. 173–189, 2017. [Online]. Available: <https://doi.org/10.1137/16M1069298>
- [33] N. Boumal, *An introduction to optimization on smooth manifolds*. Cambridge University Press, 2023. [Online]. Available: <https://www.nicolasboumal.net/book>
- [34] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, "Teleporting an unknown quantum state via dual classical and einstein-podolsky-rosen channels," *Physical Review Letters*, vol. 70, pp. 1895–1899, Mar 1993. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.70.1895>
- [35] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, "Mixed-state entanglement and quantum error correction," *Physical Review A*, vol. 54, no. 5, pp. 3824–3851, Nov. 1996.
- [36] S. L. Braunstein and H. J. Kimble, "Teleportation of continuous quantum variables," *Physical Review Letters*, vol. 80, pp. 869–872, Jan 1998. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.80.869>
- [37] C. Zhu, "Geometric optimization for quantum communication," <https://github.com/Chengkai-Zhu/qcapacity-rieopt>, 2025.
- [38] A. Han, B. Mishra, P. Jawanpuria, and A. Takeda, "A framework for bilevel optimization on riemannian manifolds," in *Proceedings of the 38th International Conference on Neural Information Processing Systems*, ser. NIPS '24. Red Hook, NY, USA: Curran Associates Inc., 2025.

ACKNOWLEDGMENT

X. W. and C. Z. acknowledge the support from the National Key R&D Program of China (Grant No. 2024YFB4504004), the National Natural Science Foundation of China (Grant. No. 12447107), the Guangdong Provincial Quantum Science Strategic Initiative (Grant No. GDZX2403008, GDZX2403001), the Guangdong Provincial Key Lab of Integrated Communication, Sensing and Computation for Ubiquitous Internet of Things (Grant No. 2023B1212010007), the Quantum Science Center of Guangdong-Hong Kong-Macao Greater Bay Area, and the Education Bureau of Guangzhou Municipality. K.F. is supported in part by the National Natural Science Foundation of China (Grant No. 12404569 and 92470113), the Shenzhen Science and Technology Program (Grant No. QNXMB20250701091826036 and JCYJ20240813113519025), the Shenzhen Fundamental Research Program (Grant No. JCYJ20241202124023031), the General R&D Projects of 1+1+1 CUHK-CUHK(SZ)-GDST Joint Collaboration Fund (Grant No. GRDP2025-022), and the University Development Fund (Grant No. UDF01003565).