Efficient Computation of the Quantum Rate-Distortion Function

Kerry He[1], James Saunderson[1], and Hamza Fawzi[2]

[1]Department of Electrical and Computer System Engineering, Monash University, Clayton VIC 3800, Australia
[2]Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, United Kingdom

arXiv:2309.15919v3 [quant-ph] 3 Apr 2024

The quantum rate-distortion function plays a fundamental role in quantum information theory, however there is currently no practical algorithm which can efficiently compute this function to high accuracy for moderate channel dimensions. In this paper, we show how symmetry reduction can significantly simplify common instances of the entanglement-assisted quantum rate-distortion problems. This allows us to better understand the properties of the quantum channels which obtain the optimal rate-distortion trade-off, while also allowing for more efficient computation of the quantum rate-distortion function regardless of the numerical algorithm being used. Additionally, we propose an inexact variant of the mirror descent algorithm to compute the quantum rate-distortion function with provable sublinear convergence rates. We show how this mirror descent algorithm is related to Blahut-Arimoto and expectation-maximization methods previously used to solve similar problems in information theory. Using these techniques, we present the first numerical experiments to compute a multi-qubit quantum rate-distortion function, and show that our proposed algorithm solves faster and to higher accuracy when compared to existing methods.

1 Introduction

The rate-distortion function quantifies how much a signal can be compressed by a lossy channel while remaining under a maximum allowable distortion of the signal. This important tool in information theory was originally formulated for the classical setting in Shannon's seminal work [1], and has recently been extended to the quantum setting [2, 3]. More concretely, in the quantum setting, the signal can be represented by an $n \times n$ positive semidefinite Hermitian matrix $\rho \in \mathbb{H}^n_+$ with unit trace, and the distortion can be quantified using a matrix $\Delta \in \mathbb{H}^{mn}_+$. The corresponding entanglement-assisted quantum rate-distortion function $R(\cdot; \rho, \Delta) : \mathbb{R}_+ \to \mathbb{R}$ is then given by

$$R(D; \rho, \Delta) = \min_{\sigma \in \mathbb{H}^{mn}_+} \{I(\sigma; \rho) : \mathrm{tr}[\sigma\Delta] \leq D, \mathrm{tr}_B(\sigma) = \rho\}. \tag{1}$$

Here, $D \geq 0$ represents the maximum allowable distortion of the signal, $\mathrm{tr}_B : \mathbb{H}^{mn} \to \mathbb{H}^n$ is a linear operator called the partial trace (see Section 2.2), and $I(\cdot; \rho)$ is a reparameterization of the quantum mutual information and defined as

$$I(\sigma; \rho) = \mathrm{tr}[\sigma \log(\sigma)] - \mathrm{tr}[\mathrm{tr}_R(\sigma) \log(\mathrm{tr}_R(\sigma))] - \mathrm{tr}[\rho \log(\rho)],$$

where log denotes the matrix logarithm and $\mathrm{tr}_R : \mathbb{H}^{mn} \to \mathbb{H}^m$ is another partial trace operator. We will introduce this problem in more detail in Section 3.

In general, problem (1) does not admit a closed-form solution, and therefore numerical optimization algorithms are required to compute values of the quantum rate-distortion function. However, it is not immediately obvious how to solve such a problem efficiently. Functions involving logarithms often do not have Lipschitz-continuous gradients nor Hessians, which are standard assumptions required to prove global convergence rates of gradient descent-like methods. Therefore, it is not obvious how quickly gradient descent-like methods will converge to the solution, if at all. Additionally, a naïve projected gradient descent approach is not suitable as the algorithm typically produces iterates on the boundary of the feasible set, where the gradient of logarithms are not well-defined.

The Blahut-Arimoto algorithm [4, 5] is one of the most well-known methods in information theory which has been applied to solve various information-theoretic optimization problems, including evaluating the classical rate-distortion function. Recently, we showed in [6] how the Blahut-Arimoto algorithm is related to mirror descent, a well studied type of Bregman proximal method [7, 8, 9, 10], and how relative smoothness analysis [11, 12, 13] could be used to prove convergence rates of the algorithm when applied to various problems in quantum information theory. However, finding an efficient analog of these algorithms for the quantum rate-distortion problem is not straightforward due to the more complicated matrix-valued constraints, and the fact that non-diagonal matrices no longer necessarily commute with each other.

In this paper, we show how we can apply the mirror descent framework to compute the quantum rate-distortion function. Although the mirror descent iterates do not admit closed-form expressions in general when applied to the quantum rate-distortion problem, in this paper we explore how they can still be efficiently computed. The main novelty of our work is showing how we can exploit inherent symmetries of common quantum rate-distortion instances to better understand the structural properties of quantum states and channels obtaining the optimal rate-distortion trade-off. Moreover, we show how this knowledge can be leveraged to gain considerable improvements to the proposed optimization algorithm by restricting the problem to a smaller subspace. A secondary contribution is showing how the mirror descent iterates for solving the quantum rate-distortion problem, which are expressed as an optimization subproblem, can be efficiently computed using a combination of duality and inexact computation of the iterates while retaining convergence guarantees. We note that our resulting algorithm is similar to [14], which proposes similar iterates to solve the quantum rate-distortion problem, but does not show how to implement certain steps of the algorithm. Altogether, we present a fully realizable method for computing the quantum rate-distortion function which outperforms all existing methods.

Related work In [14], it was shown how a method derived using expectation-maximization, which is related to the Blahut-Arimoto algorithm [15], could solve classical and quantum rate-distortion problems by solving a convex subproblem at every iteration. The author shows how one can account for inexact computation of these subproblems for classical rate-distortion problems while retaining convergence guarantees. However, it is not shown how to explicitly implement these same inexact computations for the quantum rate-distortion problem. We fill in this gap between theory and application in this work by proposing an implementable inexact stepping procedure for the quantum rate-distortion problem. In [6], we showed how an instance of the primal-dual hybrid gradient algorithm could be used to compute the classical and quantum rate-distortion functions. However, this method converges at a sublinear rate, and therefore can be slow to converge to high-

accuracy solutions. We show in Appendix C that our algorithm obtains linear convergence for common instances of the quantum-rate distortion problem. Alternatively, we can formulate the quantum rate-distortion problem as an instance of a quantum relative entropy program (see, e.g., [16, 17]), which we can solve using existing software which supports this problem class, such as CvxQuad [18], Hypatia [19], or DDS [20, 21]. However, these methods do not scale well to large problem dimensions. CvxQuad uses large semidefinite approximations of matrix logarithms, while Hypatia and DDS use second-order interior point algorithms. Therefore, they are limited to computing the rate-distortion function of small quantum channels. In contrast, the mirror descent algorithm is a first-order method which is better at scaling to large problem dimensions.

Contributions Our main contributions are threefold.

- First, we show that for certain common distortion measures, the corresponding rate-distortion problems possess symmetries that allow us to better understand the structure of quantum states and channels which obtain the optimal rate-distortion trade-off by using a symmetry reduction technique. For the quantum rate-distortion problem with entanglement fidelity distortion measure and maximally mixed input state, we show in Theorem 4.16 that there are sufficient symmetries which allow us to directly obtain an explicit solution to the rate-distortion function. For some other common quantum-rate distortion instances, we show how symmetry can be exploited to significantly reduce the dimensionality of the optimization problem. This results in improvements to both computational complexity and memory requirements, and is independent of the optimization algorithm used.

- Second, we show that the expectation-maximization approach of [14] can be interpreted as a mirror descent algorithm, and that we can rederive the algorithm and its analysis in an elementary way using tools from convex optimization.

- Third, unlike [14], we present several concrete strategies for how to solve the mirror descent subproblems while retaining convergence guarantees for the quantum rate-distortion problem. We do this primarily by introducing an inexact method, based on [22], which improves on [14] by allowing the tolerance at which we compute each iterate at to converge to zero with each subproblem, and by using an explicitly computable error criterion. This results in a more efficient algorithm which can converge sublinearly to the solution, rather than only converging to a neighborhood around the solution.

Our main algorithm is presented in Algorithm 3, which we provide an implementation of with symmetry reduction techniques at

`https://github.com/kerry-he/efficient-qrd`.

We present numerical experiments on quantum channels of up to 9-qubits to demonstrate the scalability of our approach. To the best of our knowledge, no other works have presented numerical experiments computing the quantum rate-distortion function for more than a single-qubit channel (e.g., [23], which estimates the curve using a sampling-based method). We note that unlike [6, 14], we do not directly account for the distortion constraint. However, it is relatively straightforward to extend most of our analysis and algorithm to account for it.

Throughout the paper, we predominantly study the quantum rate-distortion function using the entanglement fidelity distortion measure. This is generally considered as the

canonical distortion measure first used by the foundational works [24, 2, 3] (see also [25, 26]) as it represents a natural way of measuring the dissimilarity between two quantum states, and possesses nice properties which makes it useful when analyzing the quantum rate-distortion function. We note that although the symmetry reduction in Section 4 is dependent on the choice of distortion measure (we discuss how symmetry reduction can be applied to other common distortion measures in Section 4.4), the inexact mirror descent algorithm and its convergence analysis introduced in Section 5 is independent of this choice. This means the inexact mirror descent method can be applied for any distortion measure, however each iteration may be more expensive to compute if we cannot apply symmetry reduction.

1.1 Outline

The remainder of the paper is structured as follows. In Section 2, we summarize the notation we use, as well as preliminaries about quantum information theory and mirror descent. In Section 3, we formalize the quantum rate-distortion problem, and explain how we can reformulate the optimization problem to make it easier to solve and analyze. In Section 4, we recall symmetry reduction and show how it can be applied to common quantum rate-distortion instances. In Section 5, we introduce the inexact mirror descent algorithm we propose to solve the rate-distortion problem with. In Section 6, we present experimental results comparing our proposed algorithm to existing methods. Finally, we end with concluding remarks in Section 7. We provide some auxiliary results in the appendix, including how our results, analysis and techniques can be applied to the classical rate-distortion in Appendix A, local linear convergence rates in Appendix C, and derivations for bounds on the optimality gap of mirror descent in Appendix D.

2 Preliminaries

2.1 Notation

We denote arbitrary finite-dimensional inner product spaces by $\mathbb{V}$ and $\mathbb{V}'$, the set of $n \times n$ Hermitian matrices as $\mathbb{H}^n$ with inner product $\langle X, Y \rangle = \operatorname{tr}[XY^\dagger]$, the n-dimensional non-negative orthant and $n \times n$ positive semidefinite cone as $\mathbb{R}^n_+$ and $\mathbb{H}^n_+$ respectively, and their respective interiors as $\mathbb{R}^n_{++}$ and $\mathbb{H}^n_{++}$. For $X, Y \in \mathbb{H}^n$, we use $X \succeq Y$ to mean $X - Y \in \mathbb{H}^n_+$, and $X \succ Y$ to mean $X - Y \in \mathbb{H}^n_{++}$. We denote the extended reals by $\bar{\mathbb{R}} = \mathbb{R} \cup \infty$, and the set of positive integers as $\mathbb{N}$. We use dom to denote the domain of a function, int and relint to denote the interior and relative interior of a set, respectively, ker to denote the kernel of a linear operator, $\dim_{\mathbb{C}}$ and $\dim_{\mathbb{R}}$ to denote the dimension of a complex and real vector space respectively, $\mathbb{I}$ to denote the identity matrix, and $\otimes$ to denote the Kronecker product between two matrices. For a complex matrix $X \in \mathbb{C}^{n \times m}$, we use $\overline{X}$ to denote its entry-wise complex conjugate, and $X^\dagger$ to denote its conjugate transpose. For a linear operator $\mathcal{A} : \mathbb{V} \to \mathbb{V}'$, we denote its adjoint by $\mathcal{A}^\dagger : \mathbb{V}' \to \mathbb{V}$. We use $\{e_i\}$ to denote the standard basis for the vector space $\mathbb{C}^n$.

2.2 Quantum information theory

Here, we introduce some preliminary concepts relating to quantum information theory. We refer the reader to e.g., [27, 28, 29] for some standard references for these topics.

States and channels We define $n \times n$ quantum states as $n \times n$ positive semidefinite Hermitian matrices with unit trace, i.e.,

$$\mathcal{D}_n \coloneqq \{\rho \in \mathbb{H}_+^n : \operatorname{tr}[\rho] = 1\}.$$

These matrices are also commonly referred to as *density matrices*. *Quantum channels* are completely positive trace preserving (CPTP) (see e.g., [28, Definitions 4.4.2 and 4.4.3]) linear operators which map $n \times n$ quantum states to $m \times m$ quantum states. We denote the set of these CPTP linear operators as $\Phi_{m,n}$. It turns out that any CPTP linear map $\mathcal{N} : \mathbb{H}^n \to \mathbb{H}^m$ in $\Phi_{m,n}$ can be written in the form

$$\mathcal{N}(\rho) = \sum_{i=1}^{k} V_i \rho V_i^\dagger, \tag{2}$$

where $k \leq mn$ and the $V_i \in \mathbb{C}^{m \times n}$ (for $i = 1, \ldots, k$) are matrices satisfying $\sum_{i=1}^k V_i^\dagger V_i = \mathbb{I}$ [28, Theorem 4.4.1]. See also Definition B.1 for an alternate characterization of these linear maps.

Composite systems An n-dimensional *quantum system* is represented mathematically as an n-dimensional complex vector space $\mathbb{C}^n$, which we can associate with the set of density matrices $\mathcal{D}_n$ which act on this space. A *bipartite system* is obtained by taking the tensor product between two quantum systems, e.g., the tensor product between an n-dimensional system and an m-dimensional system produces an nm-dimensional system, which we can associate with the set of density matrices $\mathcal{D}_{mn}$ which act on this joint system. To aid in keeping track of these composite systems, we will denote quantum systems with letters, e.g., A, B, and the tensor product of these two spaces as, e.g., AB. We will use subscripts to clarify that a density matrix acts on a particular system. For instance, we write ρ_A to indicate that the quantum state acts on system A.

To make the concept of composite quantum systems more concrete, we note that the Kronecker product between any two density matrices $\rho_A \in \mathcal{D}_m$ and $\rho_B \in \mathcal{D}_n$ on systems A and B respectively will produce a density matrix $\rho_A \otimes \rho_B \in \mathcal{D}_{mn}$ on system AB. This operation is analogous to forming the joint distribution of two independent probability distributions. We also consider the "marginalization" operation, called the *partial trace*, which is a linear operator which maps density matrices on the joint system AB to density matrices on the individual systems A or B. This is analogous to finding the marginal distribution from a joint distribution. We denote the partial trace as $\operatorname{tr}_A : \mathbb{H}^{mn} \to \mathbb{H}^n$ and $\operatorname{tr}_B : \mathbb{H}^{mn} \to \mathbb{H}^m$, and define them as the adjoint of the Kronecker product with the identity matrix, i.e., $(\operatorname{tr}_A)^\dagger(\rho_B) = \mathbb{I}_A \otimes \rho_B$ and $(\operatorname{tr}_B)^\dagger(\rho_A) = \rho_A \otimes \mathbb{I}_B$. More concretely, if we represent a bipartite density matrix $\rho_{AB} \in \mathcal{D}_{mn}$ on system AB as a block matrix consisting of an $m \times m$ array of $n \times n$ blocks X_{ij}, then

$$\operatorname{tr}_A(\rho_{AB}) = \sum_{i=1}^{m} X_{ii} \qquad \text{and} \qquad [\operatorname{tr}_B(\rho_{AB})]_{ij} = \operatorname{tr}[X_{ij}].$$

We will also use the tensor product between linear operators. In particular, for an operator $\mathcal{N} \in \Phi_{m,n}$ given by (2), we define $(\mathcal{N} \otimes \mathbb{I}) \in \Phi_{(m+p),(n+p)}$, where $\mathbb{I}$ is of size $p \times p$, as

$$(\mathcal{N} \otimes \mathbb{I})(\rho) = \sum_{i=1}^{k} (V_i \otimes \mathbb{I}) \rho (V_i \otimes \mathbb{I})^\dagger.$$

Purification Consider a quantum state $\rho_A \in \mathcal{D}_n$ with spectral decomposition $\rho_A = \sum_{i=1}^n \lambda_i v_i v_i^\dagger$ defined on an n-dimensional system A, and a system R which is isomorphic to A. A *purification* of ρ_A is a complex unit vector $\psi \in \mathbb{C}^{n^2}$ defined on the joint system AR which satisfies $\mathrm{tr}_R(\psi\psi^\dagger) = \rho_A$. The purification is not unique. However, in this paper we will consider the particular choice of purification

$$\psi = \sum_{i=1}^n \sqrt{\lambda_i}\, v_i \otimes v_i, \tag{3}$$

and refer to this as the purification of ρ_A. With some further abuse of terminology, we will also refer to the quantum state associated with (3), i.e., $\rho_{AR} \coloneqq \psi\psi^\dagger \in \mathcal{D}_{n^2}$, as the purification of ρ_A. Note that this choice of purification satisfies $\rho_R \coloneqq \mathrm{tr}_A(\rho_{AR}) = \mathrm{tr}_R(\rho_{AR}) = \rho_A$.

Entropy For a function $f : \mathbb{R} \to \bar{\mathbb{R}}$ and Hermitian matrix $X \in \mathbb{H}^n$ with spectral decomposition $X = \sum_{i=1}^n \lambda_i v_i v_i^\dagger$, the function f can be extended to matrices X as

$$f(X) \coloneqq \sum_{i=1}^n f(\lambda_i) v_i v_i^\dagger.$$

With some overloading of function notation, we define the von Neumann entropy $S : \mathbb{H}_+^n \to \mathbb{R}$ as

$$S(\rho) \coloneqq -\mathrm{tr}[\rho \log(\rho)], \tag{4}$$

and the quantum relative entropy $S : \mathbb{H}_+^n \times \mathbb{H}_{++}^n \to \mathbb{R}$ as

$$S(\rho \| \sigma) \coloneqq \mathrm{tr}[\rho(\log(\rho) - \log(\sigma)]. \tag{5}$$

For notational convenience, we use the natural base e throughout the paper for logarithms and exponentials. However, all computational experiments are performed in base 2 to obtain results measured in bits, i.e., the standard unit of measurement for information. This is done by rescaling relevant quantities by a factor of $\log_2(e)$.

2.3 Mirror descent

We will focus on constrained convex optimization problems of the form

$$\min_x \quad f(x) \tag{6a}$$
$$\text{subj. to} \quad \mathcal{A}(x) = b, \tag{6b}$$

where $f : \mathbb{V} \to \bar{\mathbb{R}}$ is a convex function which is differentiable on $\operatorname{int} \operatorname{dom} f$, $\mathcal{A} : \mathbb{V} \to \mathbb{V}'$ is a linear operator, and $b \in \mathbb{V}'$. Throughout the rest of the paper, we define

$$\mathcal{C} \coloneqq \{x \in \operatorname{dom} f : \mathcal{A}(x) = b\}, \tag{7}$$

as the intersection of the implicit and explicit constraints of (6).

Before introducing the mirror descent algorithm, we first introduce the class of Legendre functions, which are known to satisfy certain nice properties. We adopt the terminology in [30] for common terms relating to convex functions.

Definition 2.1 ([13, Definition 2.1])**.** *A function $\varphi : \mathbb{V} \to \bar{\mathbb{R}}$ is* Legendre *if it is proper, lower semicontinuous, strictly convex, and essentially smooth.*

Fact 2.2 ([30, Theorem 26.5]). *Let $\varphi : \mathbb{V} \to \bar{\mathbb{R}}$ be a Legendre function.*

(i) The convex conjugate φ^* of φ is also Legendre.

(ii) The gradient $\nabla\varphi : \operatorname{int}\operatorname{dom}\varphi \to \operatorname{int}\operatorname{dom}\varphi^*$ is continuous and invertible, where $(\nabla\varphi)^{-1} = \nabla\varphi^*$.

Given a Legendre function $\varphi : \mathbb{V} \to \bar{\mathbb{R}}$, the Bregman divergence $D_\varphi : \operatorname{dom}\varphi \times \operatorname{int}\operatorname{dom}\varphi \to \mathbb{R}$ is [31]

$$D_\varphi(x\|y) \coloneqq \varphi(x) - (\varphi(y) + \langle \nabla\varphi(y), x - y\rangle). \tag{8}$$

Now consider using a Legendre function φ to construct a local approximation of f at $y \in \operatorname{int}\operatorname{dom}\varphi \cap \operatorname{int}\operatorname{dom} f$ for some scaling factor $t > 0$

$$\min_{x\in\mathcal{C}} \quad f(y) + \langle \nabla f(y), x - y\rangle + \frac{1}{t} D_\varphi(x\|y). \tag{9}$$

If y is close to the optimal solution x^* of (6), and if φ is, in a sense, similar to f, then we expect the solution to (9) to be a good approximation of (6). More formally, we recall from [11, 12, 13] the following definitions for relative smoothness and strong convexity to describe structural similarities between f and φ.

Definition 2.3. *Consider a function $f : \mathbb{V} \to \bar{\mathbb{R}}$, a Legendre function $\varphi : \mathbb{V} \to \bar{\mathbb{R}}$, and convex set $\mathcal{X} \subseteq \mathbb{V}$. Then*

(i) f is L-smooth relative to φ on $\mathcal{X}$ if $L\varphi - f$ is convex on $\operatorname{relint}\mathcal{X}$ for some $L > 0$, and

(ii) f is μ-strongly convex relative to φ on $\mathcal{X}$ if $f - \mu\varphi$ is convex on $\operatorname{relint}\mathcal{X}$ for some $\mu > 0$.

In particular, if f is L-smooth relative to φ on $\mathcal{C}$, then the local approximation used in (9) for $t = 1/L$ is a majorization of f, and therefore gives an upper bound on the optimal value $f(x^*)$. To solve (6), mirror descent uses the approximation (9) to construct the sequence of iterates

$$x^{k+1} = \arg\min_{x\in\mathcal{C}} \left\{ \langle \nabla f(x^k), x\rangle + \frac{1}{t_k} D_\varphi(x\|x^k) \right\}. \tag{10}$$

This is a practical algorithm when (9) is easier to solve compared to the original problem (6). When f satisfies certain smoothness and strong convexity properties relative to φ, mirror descent has the following convergence properties.

Fact 2.4 ([13, Theorem 4.1 and Proposition 4.1]). *Let φ be Legendre, and $\{x^k\}$ be the sequence of iterates generated by mirror descent (10) to solve (6).*

(i) If f is L-smooth relative to φ, then $f(x^k)$ satisfies

$$f(x^k) - f^* \le \frac{L}{k} D_\varphi(x^*\|x^0), \quad \forall k \in \mathbb{N},$$

i.e., $f(x^k)$ converges sublinearly, $O(1/k)$, to the optimal value.

(ii) If f is L-smooth relative to φ and μ-strongly convex relative to φ, then $f(x^k)$ satisfies

$$f(x^k) - f^* \leq \left(1 - \frac{\mu}{L}\right)^k L D_\varphi(x^* \| x^0), \quad \forall k \in \mathbb{N},$$

i.e., $f(x^k)$ converges linearly, $O(\alpha^k)$ for $\alpha \in (0,1)$, to the optimal value.

We will make the following assumptions throughout the rest of the paper to ensure the mirror descent algorithm is well defined and to simplify some of the analysis.

Assumption A. *Consider the convex optimization problem (6) and mirror descent rule (10).*

(i) $f : \mathbb{V} \to \bar{\mathbb{R}}$ *is convex and continuously differentiable on* $\operatorname{int}\operatorname{dom} f$,

(ii) $\varphi : \mathbb{V} \to \bar{\mathbb{R}}$ *is Legendre,*

(iii) $\operatorname{dom} f = \operatorname{dom} \varphi$,

(iv) $\mathcal{C}$ *is compact and non-empty, and*

(v) $\operatorname{dom} \varphi^* = \mathbb{V}$.

Under Assumptions items A(i)–A(iv), the mirror descent iterate (10) always exists and is uniquely defined [11, Lemma 2]. Assumption A(v) may seem restrictive, however it is satisfied by many important examples (see [13, Example 2.1]), including the ones we later use to compute the classical and quantum rate-distortion functions. The motivation behind Assumption A(iii) is that this allows the Bregman divergence to act as a natural barrier for the domain of the objective function, allowing these implicit constraints to be more easily handled. In particular, Legendre functions satisfy the following property.

Fact 2.5 ([32, Lemma 2.2])**.** *Consider an essentially smooth convex function $\varphi : \mathbb{V} \to \bar{\mathbb{R}}$. If φ is minimized at $x \in \operatorname{dom} \varphi$, then $x \in \operatorname{int}\operatorname{dom} \varphi$.*

Therefore, as the sum of an essentially smooth convex function and an affine function is still clearly essentially smooth and convex, Fact 2.5 tells us that the mirror descent iterates (10) always lie in $\operatorname{int}\operatorname{dom} \varphi$.

Dual of mirror descent subproblem It will also be useful to consider the Lagrangian of the problem (9) for some $y \in \operatorname{int}\operatorname{dom} \varphi$

$$\mathcal{L}(x, \nu; y) \coloneqq \langle \nabla f(y), x \rangle + \frac{1}{t} D_\varphi(x \| y) + \langle \nu, \mathcal{A}(x) - b \rangle, \tag{11}$$

and the corresponding dual function

$$g(\nu; y) \coloneqq \inf_x \mathcal{L}(x, \nu; y) = -\frac{1}{t} \varphi^*[\nabla \varphi(y) - t(\nabla f(y) + \mathcal{A}^\dagger(\nu))] - \langle b, \nu \rangle, \tag{12}$$

which under Assumption A(v) has the property $\operatorname{dom} g(\cdot; y) = \mathbb{V}'$. When $y = x^k$, these are the Lagrangian and dual function for the mirror descent subproblem (10) at the k-th mirror descent iteration. We will use $\mathcal{L}^k(x, \nu) \coloneqq \mathcal{L}(x, \nu; x^k)$ and $g^k(\nu) \coloneqq g(\nu; x^k)$ to simplify notation. An alternative characterization of the dual is $g(\nu; y) = \mathcal{L}(\tilde{x}, \nu; y)$ where

$$\tilde{x} = \arg\min_x \mathcal{L}(x, \nu; y) = (\nabla \varphi)^{-1}[\nabla \varphi(y) - t(\nabla f(y) + \mathcal{A}^\dagger(\nu))], \tag{13}$$

which can be used to recover a primal optimal solution from a dual optimal solution assuming $(\nabla\varphi)^{-1}$ is easily computable. Using Fact 2.5, it follows that $\tilde{x} \in \operatorname{int}\operatorname{dom}\varphi$. We can show that the dual problem

$$\max_{\nu} \; g(\nu; y), \tag{14}$$

is well-posed under certain conditions using the following proposition. We begin with a standard definition from convex analysis before introducing the main result.

Definition 2.6 (Coercive function)**.** *A function $f : \mathbb{V} \to \bar{\mathbb{R}}$ is coercive if for every sequence $\{x^k\}$ such that $\|x^k\| \to \infty$, we have $f(x^k) \to \infty$.*

Proposition 2.7. *Consider the dual function $g(\cdot; y) : \mathbb{V}' \to \mathbb{R}$, as defined in (12), for some $y \in \operatorname{int}\operatorname{dom}\varphi$ and where φ is Legendre. Then*

(i) $-g(\cdot; y)$ is strictly convex if and only if $\mathcal{A}^\dagger$ is injective, and

(ii) $-g(\cdot; y)$ is coercive if and only if $\mathcal{A}^\dagger$ is injective and $b \in \{\mathcal{A}(x) : x \in \operatorname{int}\operatorname{dom}\varphi\}$.

Proof. See Appendix E.1. □

Using this result, it is well known that existence of a solution to the dual problem follows from coercivity, and uniqueness of the solution follows from strict convexity [33, Proposition 3.1.1 and 3.2.1].

3 Problem definition

Consider an n-dimensional input system A, an m-dimensional output system B, and a quantum state $\rho_A \in \mathcal{D}_n$ with purification $\rho_{AR} \in \mathcal{D}_{n^2}$. Let $\Delta \in \mathbb{H}_+^{mn}$ be a distortion matrix and $D \geq 0$ be a maximum allowable distortion. The *entanglement-assisted quantum rate-distortion function* [2, 3] is

$$\begin{aligned} R_q(D; \rho_A, \Delta) \quad &:= \quad \min_{\mathcal{N} \in \Phi_{m,n}} \quad I_q(\mathcal{N}; \rho_A) && \text{(15a)} \\ &\qquad\quad \text{subj. to} \quad \langle \Delta, (\mathcal{N} \otimes \mathbb{I})(\rho_{AR}) \rangle \leq D, && \text{(15b)} \end{aligned}$$

where

$$I_q(\mathcal{N}; \rho_A) := S((\mathcal{N} \otimes \mathbb{I})(\rho_{AR}) \| \mathcal{N}(\rho_A) \otimes \rho_A). \tag{16}$$

is the *quantum mutual information*. It follows from joint convexity of quantum relative entropy [28, Corollary 11.9.2] that (15) is a convex optimization problem.

A standard distortion measure used for the quantum rate-distortion function is to use the entanglement fidelity (see, e.g., [24, 2, 3, 25, 26]). This corresponds to the channel where A and B are isomorphic, and the distortion matrix is $\Delta = \mathbb{I} - \rho_{AR}$. This distortion measure possesses symmetries which we will can take advantage of in Section 4. We also briefly discuss the symmetries of other common distortion measures in Section 4.4.

Remark 3.1. *Without loss of generality, we will always assume that the input state of the quantum channel is full-rank, i.e., $\rho_A \succ 0$. This is because it is possible to construct equivalent optimization problems of a lower input dimension to avoid these degeneracies by discarding parts of the channel and distortion measures which correspond to degenerate inputs which have no effect on the problem. See Appendix B.2 for more details.*

3.1 Problem reformulation

In this section, we summarize standard reformulations of the quantum rate-distortion problems that are more convenient for our purposes.

Change of variables We will make a change of variables where we optimize over the bipartite state $\sigma_{BR} \in \mathcal{D}_{mn}$, where $\sigma_{BR} = (\mathcal{N} \otimes \mathbb{I})(\rho_{AR})$. For distortion $D \geq 0$, input state $\rho_A \in \mathcal{D}_n$ and distortion matrix $\Delta \in \mathbb{H}_+^{mn}$, this gives us

$$\begin{aligned} R_q(D; \rho_A, \Delta) \quad = \quad & \min_{\sigma_{BR} \in \mathcal{D}_{mn}} \quad I_q(\sigma_{BR}; \rho_A) && \text{(17a)} \\ & \text{subj. to} \quad \operatorname{tr}_B(\sigma_{BR}) = \rho_A, && \text{(17b)} \\ & \qquad\qquad\ \ \langle \Delta, \sigma_{BR} \rangle \leq D, && \text{(17c)} \end{aligned}$$

where with some slight abuse of notation we redefine quantum mutual information as

$$I_q(\sigma_{BR}; \rho_A) \coloneqq S(\sigma_{BR} \| \operatorname{tr}_R(\sigma_{BR}) \otimes \rho_A). \tag{18}$$

Note that the apparent inconsistency in the quantum systems implied by the subscripts of (17b) is due to the fact that, as discussed in Section 2.2, A and R are isomorphic systems, and that we have defined the purification in a way such that $\rho_R \coloneqq \operatorname{tr}_A(\rho_{AR}) = \rho_A$.

Proposition 3.2. *The quantum rate-distortion problems* (15) *and* (17) *are equivalent, in the sense that the optimal values of the two problems are equal.*

Proof. See Appendix B.3. □

As we made a linear change of variables, these reformulated problem is still convex. We refer the reader to [6, Section 6.2] for the gradient of the reformulated quantum mutual information.

Distortion constraint For rate-distortion algorithms, we are usually interested in characterizing the entire rate-distortion curve rather than computing the rate for a single distortion. Therefore it is standard to parameterize the distortion constraint with its corresponding dual variable $\kappa \geq 0$ instead of D (see e.g., [4]). In this case, the quantum rate-distortion problem for distortion dual variable $\kappa \geq 0$, input state $\rho_A \in \mathcal{D}_n$, and distortion matrix $\Delta \in \mathbb{H}_+^{mn}$ becomes

$$\begin{aligned} \tilde{R}_q(\kappa; \rho_A, \Delta) \quad = \quad & \min_{\sigma_{BR} \in \mathcal{D}_{mn}} \quad I_q(\sigma_{BR}; \rho_A) + \kappa \langle \Delta, \sigma_{BR} \rangle && \text{(QRD.a)} \\ & \text{subj. to} \quad \operatorname{tr}_B(\sigma_{BR}) = \rho_A, && \text{(QRD.b)} \end{aligned}$$

Note that as the distortion constraints are linear, the convexity properties of the objective function does not change with this reformulation. Additionally, we have omitted some additive constant terms in the objective to simplify the expression. Once a solution σ_{BR}^* to (QRD) is found for a given $\kappa \geq 0$, it is straightforward to recover the corresponding optimal rate $R_q(D; \rho_A, \Delta) = I_q(\sigma_{BR}^*; \rho_A)$ for distortion $D = \langle \Delta, \sigma_{BR}^* \rangle$.

4 Symmetry reduction

A practical issue with solving the quantum rate-distortion problem (QRD) is that the problem dimension scales quartically with the quantum system dimension. We will show

that we can restrict this problem to a significantly smaller subspace by taking advantage of symmetry properties of the rate-distortion problem for certain distortion matrices. We will also show that for certain highly symmetric problems, we can obtain explicit expressions for the rate-distortion function.

4.1 Background

We begin by introducing some preliminaries on group and representation theory. We refer the reader to [34, 35] as references for these basic facts. Throughout this section we will denote the general linear group as $GL(\mathbb{V})$, which is the group of invertible linear transforms from $\mathbb{V}$ to $\mathbb{V}$ with the group operation being composition.

Definition 4.1 (Representation)**.** *A* representation *of a group $\mathcal{G}$ is a pair $(\mathbb{V}, \pi)$ where $\mathbb{V}$ is a vector space and $\pi : \mathcal{G} \to GL(\mathbb{V})$ is a group homomorphism.*

Definition 4.2 (Isomorphism)**.** *Let $(\mathbb{V}, \pi)$ and $(\mathbb{V}', \tau)$ be representations of a group $\mathcal{G}$. If there exists an invertible linear map $T : \mathbb{V} \to \mathbb{V}'$ such that*

$$\tau(g) = T \circ \pi(g) \circ T^{-1},$$

for all $g \in \mathcal{G}$, then π and τ are isomorphic, denoted $\pi \cong \tau$.

Definition 4.3 (Fixed-point subspace)**.** *The* fixed point subspace *$\mathcal{V}_\pi$ associated with a representation $(\mathbb{V}, \pi)$ of a group $\mathcal{G}$ is*

$$\mathcal{V}_{\pi|\mathbb{V}} = \{x \in \mathbb{V} : \pi(g)(x) = x,\ \forall g \in \mathcal{G}\}.$$

When the vector space $\mathbb{V}$ is clear from context, we use the abbreviated notation $\mathcal{V}_\pi \coloneqq \mathcal{V}_{\pi|\mathbb{V}}$.

Definition 4.4 (Group average)**.** *For a representation $(\mathbb{V}, \pi)$ of a compact group $\mathcal{G}$, the* group average *is*

$$P_\pi = \int_{g \in \mathcal{G}} \pi(g)\, d\theta(g),$$

where θ is the normalized Haar measure on $\mathcal{G}$ (see e.g., [36]). The group average is the projector onto the fixed point subspace $\mathcal{V}_\pi$.

Lemma 4.5. *Consider a representation $(\mathbb{V}, \pi)$ of a group $\mathcal{G}$. If a convex optimization problem of the form (6) is invariant under π, meaning*

$$f(\pi(g)(x)) = f(x) \qquad \forall g \in \mathcal{G},\ \forall x \in \mathcal{C} \tag{20a}$$

$$\text{and} \quad \pi(g)(x) \in \mathcal{C} \qquad \forall g \in \mathcal{G},\ \forall x \in \mathcal{C}, \tag{20b}$$

then there is an optimal point for (6) in $\mathcal{V}_\pi$.

Proof. Let x^* be a solution to the convex optimization problem and f^* be the optimal value. We can show that

$$f^* \leq f\left(\int_{g \in \mathcal{G}} \pi(g)(x^*)\, d\theta(g)\right) \leq \int_{g \in \mathcal{G}} f(\pi(g)(x^*)))\, d\theta(g) = f^*.$$

The first inequality holds as $P_\pi(x^*) \in \mathcal{C}$ from (20b), and that by definition $f^* \leq f(x)$ for all $x \in \mathcal{C}$. The second inequality uses convexity of f, and the equality uses (20a). Therefore, equality must hold throughout. This implies the projection of the solution onto the fixed-point subspace, $P_\pi(x^*)$, must also be a solution, as desired. □

We will also introduce characters of compact groups and well known results associated with them. For these facts, we will assume that $\mathbb{V}$ is a complex vector space, as some of these results do not hold in general for real vector spaces.

Definition 4.6 (Character)**.** *Let $(\mathbb{V}, \pi)$ be a representation of a compact group $\mathcal{G}$. The* character $\chi_\pi : \mathcal{G} \to \mathbb{C}$ *of this representation is defined as*

$$\chi_\pi(g) \coloneqq \operatorname{tr}[\pi(g)].$$

Given characters χ, ϕ and normalised Haar measure θ of $\mathcal{G}$, we define the inner product

$$\langle \chi, \phi \rangle \coloneqq \int_{g \in \mathcal{G}} \chi(g) \overline{\phi(g)} \, d\theta(g).$$

Lemma 4.7 ([34, Proposition 4.10(v)])**.** *Let $(\mathbb{V}, \pi)$ and $(\mathbb{V}', \tau)$ be representations of a compact group $\mathcal{G}$. Then $\chi_{\pi \oplus \tau} = \chi_\pi + \chi_\tau$, where $\oplus$ denotes the direct sum between representations.*

Lemma 4.8 ([34, Theorem 4.11(iii)])**.** *Characters of irreducible representations of a compact group $\mathcal{G}$ are orthonormal. That is for any irreducible representations $(\mathbb{V}, \pi)$ and $(\mathbb{V}', \tau)$ of $\mathcal{G}$*

$$\langle \chi_\pi, \chi_\tau \rangle = \begin{cases} 1, & \text{if } \pi \cong \tau, \\ 0, & \text{if } \pi \ncong \tau. \end{cases}$$

Lemma 4.9. *Given a complex representation $(\mathbb{C}^n, \tau)$ of a compact group $\mathcal{G}$, define a representation $(\mathbb{C}^{n \times n}, \pi)$ of $\mathcal{G}$ by*

$$\pi(g)(X) = \tau(g) X \tau(g)^\dagger. \tag{21}$$

The dimension (over $\mathbb{C}$) of the fixed-point subspace $\mathcal{V}_\pi$ is

$$\dim_{\mathbb{C}} \mathcal{V}_\pi = \int_{g \in \mathcal{G}} \operatorname{tr}[\tau(g)] \, \overline{\operatorname{tr}[\tau(g)]} \, d\theta(g) = \langle \chi_\tau, \chi_\tau \rangle. \tag{22}$$

Proof. This uses [34, Theorem 4.11(i)] combined with the properties $\operatorname{vec}(AXB^\dagger) = (A \otimes \overline{B}) \operatorname{vec}(X)$ and $\operatorname{tr}[A \otimes B] = \operatorname{tr}[A] \operatorname{tr}[B]$. □

We next consider the analogue of Lemma 4.9 for real representations.

Corollary 4.10. *Let $\mathcal{G}$ be a compact group, let $(\mathbb{C}^n, \tau)$ be a complex representation of $\mathcal{G}$, and let $(\mathbb{H}^n, \pi)$ be the real representation of $\mathcal{G}$ defined by $\pi(g)(X) = \tau(g) X \tau(g)^\dagger$. Then the dimension (over $\mathbb{R}$) of the fixed point subspace is given by $\dim_{\mathbb{R}}(\mathcal{V}_\pi) = \langle \chi_\tau, \chi_\tau \rangle$.*

Proof. We first recognize that by treating $\mathbb{C}^{n \times n}$ as a real vector space, $(\mathbb{C}^{n \times n}, \pi)$ decomposes into two isomorphic real subrepresentations given by the restriction to Hermitian and skew-Hermitian matrices, respectively. The decompositions of these isomorphic representations into irreducibles are, therefore, identical. It follows that the fixed point subspaces in each have the same real dimension. Therefore given that the fixed point subspace $\mathcal{V}_{\pi|\mathbb{C}^{n \times n}}$ has a dimension over the reals of $\dim_{\mathbb{R}}(\mathcal{V}_{\pi|\mathbb{C}^{n \times n}}) = 2 \dim_{\mathbb{C}}(\mathcal{V}_{\pi|\mathbb{C}^{n \times n}})$, restricting to the Hermitian space and using Lemma 4.7 leaves us with the desired real dimension for $\mathcal{V}_{\pi|\mathbb{H}^n}$. □

4.2 Entanglement fidelity distortion

We now restrict our attention to the quantum rate-distortion problem with the entanglement fidelity distortion matrix $\Delta = \mathbb{I} - \rho_{AR}$, where ρ_{AR} is the purification of the input state ρ_A. We will show that for this choice of distortion matrix, there are significant symmetries in the quantum rate-distortion problem (QRD) which can be used to drastically reduce the problem dimension.

We will first introduce the following lemma that will be useful to prove the main result.

Lemma 4.11. *Consider an m-dimensional system A, an n-dimensional B, and a density matrix $\rho_{AB} \in \mathcal{D}_{mn}$ defined on the joint system AB. Let $T_A \in \mathbb{C}^{p\times m}$ be an arbitrary linear operator on A, and $U_B \in \mathbb{C}^{n\times n}$ be an arbitrary unitary operator on B. Then*

$$\operatorname{tr}_B((T_A \otimes U_B)\rho_{AB}(T_A \otimes U_B)^\dagger) = T_A \operatorname{tr}_B(\rho_{AB}) T_A^\dagger.$$

Proof. Recall that the partial trace is defined as the adjoint of the Kronecker product with the identity matrix. Therefore, for all $\sigma_A \in \mathcal{D}_n$ the following equalities must hold

$$\begin{aligned}\langle \sigma_A, \operatorname{tr}_B((T_A \otimes U_B)\rho_{AB}(T_A \otimes U_B)^\dagger)\rangle &= \langle \sigma_A \otimes \mathbb{I}_B, (T_A \otimes U_B)\rho_{AB}(T_A \otimes U_B)^\dagger)\rangle \\ &= \langle (T_A^\dagger \sigma_A T_A) \otimes \mathbb{I}_B, \rho_{AB}\rangle \\ &= \langle \sigma_A, T_A \operatorname{tr}_B(\rho_{AB}) T_A^\dagger\rangle,\end{aligned}$$

where the second equality uses $U_B^\dagger U_B = \mathbb{I}_B$. This produces the desired result. □

We now find a group action that leaves the quantum rate-distortion problem invariant, which we can use to characterize the corresponding fixed-point subspace in which a solution must lie. For simplicity of notation we assume the input state ρ_A is diagonal for the following lemma.

Lemma 4.12. *Consider a diagonal input state $\rho_A \in \mathcal{D}_n$. Let $\mathcal{G}$ be a subgroup of the group of $n \times n$ unitary matrices, such that $g\rho_A g^\dagger = \rho_A$ for all $g \in \mathcal{G}$, and $(\mathbb{H}^{n^2}, \pi)$ be a representation of $\mathcal{G}$ such that*

$$\pi(g)(X) = UXU^\dagger, \quad \forall X \in \mathbb{H}^{n^2}, \quad \text{where} \quad U = g \otimes \overline{g}. \tag{23}$$

Then the quantum rate-distortion problem (QRD) parameterized by input state ρ_A and distortion matrix $\Delta = \mathbb{I} - \rho_{AR}$, where ρ_{AR} is the purification of ρ_A, is invariant under π.

Proof. We will show that the objective function and each constraint of problem (QRD) is invariant under π. First, we recognize that since ρ_A is Hermitian and diagonal, $\rho_A = \overline{\rho_A}$, and therefore $g\rho_A g^\dagger = \rho_A$ implies $\overline{g}\rho_A\overline{g}^\dagger = \rho_A$. Invariance of the objective function (QRD.a) can be shown by first showing quantum mutual information is invariant under π as follows

$$\begin{aligned}I_q(\pi(g)(\sigma_{BR}); \rho_A) &= S(\pi(g)(\sigma_{BR}) \,\|\, \operatorname{tr}_R(\pi(g)(\sigma_{BR})) \otimes \rho_A) \\ &= S(U\sigma_{BR}U^\dagger \,\|\, g \operatorname{tr}_R(\sigma_{BR}) g^\dagger \otimes \overline{g}\rho_A\overline{g}^\dagger) \\ &= S(U\sigma_{BR}U^\dagger \,\|\, U(\operatorname{tr}_R(\sigma_{BR}) \otimes \rho_A)U^\dagger) \\ &= I_q(\sigma_{BR}; \rho_A),\end{aligned}$$

where the second equality uses Lemma 4.11 and $\overline{g}\rho_A\overline{g}^\dagger = \rho_A$, and the third equality uses invariance of quantum relative entropy under unitary transforms. For the linear

component of the objective arising from the distortion constraint, let us first assume ρ_A can be expressed as

$$\rho_A = \sum_{I=1}^{r} \sum_{i \in \mathcal{I}_I} \lambda_I e_i e_i^\top,$$

i.e., ρ_A has r distinct eigenvalues, and $\mathcal{I}_I$ is the set of indices corresponding to eigenvalue λ_I for all $I = 1, \dots, r$. For any $I = 1, \dots, r$ and $i \in \mathcal{I}_I$, we have

$$\rho_A e_i = g \rho_A g^\dagger e_i = \lambda_i e_i \quad \iff \quad \rho_A \, g^\dagger e_i = \lambda_i \, g^\dagger e_i.$$

This implies that $g^\dagger e_i$ must be some vector in the subspace spanned by $\{e_i\}_{i \in \mathcal{I}_I}$, and because $g^\dagger$ is unitary, that $\{g^\dagger e_i\}_{i \in \mathcal{I}_I}$ forms an orthonormal basis of this subspace. Therefore

$$\sum_{i \in \mathcal{I}_I} g^\dagger e_i e_i^\top g = \sum_{i \in \mathcal{I}_I} e_i e_i^\top,$$

i.e., $g^\dagger P_I g = P_I$ where P_I is the orthogonal projector onto the eigenspace corresponding to eigenvalue λ_I. Using this, we can show

$$\begin{aligned}
\pi(g^\dagger)(\rho_{AR}) &= \sum_{ij}^{n} \sqrt{\lambda_i \lambda_j} (g^\dagger e_i e_i^\top g) \otimes \overline{(g^\dagger e_j e_j^\top g)} \\
&= \sum_{I=1}^{r} \sum_{J=1}^{r} \sqrt{\lambda_I \lambda_J} \Big(\sum_{i \in \mathcal{I}_I} g^\dagger e_i e_i^\top g \Big) \otimes \overline{\Big(\sum_{j \in \mathcal{I}_J} g^\dagger e_j e_j^\top g \Big)} \\
&= \sum_{I=1}^{r} \sum_{J=1}^{r} \sqrt{\lambda_I \lambda_J} \Big(\sum_{i \in \mathcal{I}_I} e_i e_i^\top \Big) \otimes \Big(\sum_{j \in \mathcal{I}_J} e_j e_j^\top \Big) \\
&= \rho_{AR}
\end{aligned}$$

Combining this with the fact that $\langle \Delta, \pi(g)(\sigma_{BR}) \rangle = \langle \pi(g^\dagger)(\Delta), \sigma_{BR} \rangle$ and that $\pi(g^\dagger)(\mathbb{I}) = \mathbb{I}$ shows that the linear component of the objective is invariant under π.

Next, we will show invariance of the feasible set. First, note that the set of density matrices is invariant under congruence by unitary transforms. Second, consider an arbitrary σ_{BR} which satisfies (QRD.b). Then we can show that

$$\mathrm{tr}_B(\pi(g)(\sigma_{BR})) = \overline{g} \rho_A \overline{g}^\dagger = \rho_A,$$

where the first equality uses Lemma 4.11. This shows that the constraint set is invariant under π, and concludes the proof. □

The general idea motivating the following theorem is that we want to find an appropriate subgroup $\mathcal{G}$ that satisfies the condition $g \rho_A g^\dagger = \rho_A$, which is sufficiently large so that we can restrict to as small of a fixed-point subspace as possible. In the following theorem, we characterize such a subgroup and its corresponding fixed-point subspace for an arbitrary purification.

Theorem 4.13. *Consider the quantum rate-distortion problem (QRD) parameterized by input state $\rho_A \in \mathcal{D}_n$ with spectral decomposition $\rho_A = \sum_{i=1}^{n} \lambda_i v_i v_i^\dagger$ and distortion matrix $\Delta = \mathbb{I} - \rho_{AR}$, where ρ_{AR} is the purification of ρ_A. A solution to this problem exists in the subspace*

$$\mathcal{V}_\pi := \Big\{ \sum_{i \neq j}^{n} \alpha_{ij} v_i v_i^\dagger \otimes v_j v_j^\dagger + \sum_{ij}^{n} \beta_{ij} v_i v_j^\dagger \otimes v_i v_j^\dagger : \alpha_{ij} \in \mathbb{R} \ \forall i \neq j, \ \beta \in \mathbb{H}^n \Big\}. \tag{24}$$

Proof. We first consider the special case when ρ_A is diagonal, i.e., $v_i = e_i$ for all $i = 1, \ldots, n$. For any of these diagonal inputs, we can show that the unitary subgroup $\mathcal{G}$ consisting of elements

$$\mathcal{G} = \left\{ \sum_{i=1}^{n} z_i e_i e_i^\top : z_i \in \{\pm 1, \pm\sqrt{-1}\} \ \forall i \right\}, \tag{25}$$

satisfies $g\rho_A g^\dagger = \rho_A$ for all $g \in \mathcal{G}$, and therefore Lemma 4.12 tells us that the quantum rate-distortion problem is invariant under representation (23) of this subgroup. It follows from Lemma 4.5 that a solution must live in the fixed-point subspace of this representation. To characterise this subspace, let us first consider representation $(\mathbb{C}^{n^2}, \tau)$ of the group $\mathcal{G}$ such that $\tau(g)(x) = (g \otimes \overline{g})x$. We can show that τ can be decomposed into the following subrepresentations (V_{ij}, τ_{ij}) such that

$$\begin{aligned} V_{ij} &= \{x(e_i \otimes e_j) : x \in \mathbb{C}\} \\ \text{and} \quad \tau_{ij}(g)(x) &= z_i \overline{z_j} (e_i e_i^\top \otimes e_j e_j^\top), \quad \forall i, j = 1, \ldots, n. \end{aligned}$$

where we follow the notation of (25) to define z_i. We recognize that the subrepresentations τ_{ii} for all $i = 1, \ldots, n$ are just the identity map on a one-dimensional subspace, and are therefore all isomorphic to each other. If $i \neq j$, then the τ_{ij} are mutually non-isomorphic irreducible representations. Therefore, using Corollary 4.10 we can compute the dimension over the reals of the fixed point subspace $\mathcal{V}_\pi$ as

$$\dim_{\mathbb{R}} \mathcal{V}_\pi = \langle \chi_\tau, \chi_\tau \rangle = \sum_{ijkl} \langle \chi_{ij}, \chi_{kl} \rangle = \sum_{ij} \langle \chi_{ii}, \chi_{jj} \rangle + \sum_{i \neq j} \langle \chi_{ij}, \chi_{ij} \rangle = 2n^2 - n.$$

where χ_{ij} is the character of subrepresentation τ_{ij}, and the first and second equalities use Lemmas 4.7 and 4.8 respectively. Verifying that the proposed subspace (24) is invariant under representation (23) of group $\mathcal{G}$ and is of the correct dimension gives the desired result for any diagonal input ρ_A. This result generalizes to non-diagonal inputs by using the fact that we can use a change-of-basis to show that the quantum rate-distortion problem for any non-diagonal input state is equivalent to a quantum rate-distortion problem with a diagonal input state. See Appendix B.1 for more details. □

We also introduce the following property of this subspace which can be shown from a straightforward calculation, and will be useful in some following proofs.

Corollary 4.14. *Let $\sigma_{BR} \in \mathcal{V}_\pi$ as defined in (24). The partial traces of σ_{BR} with respect to R and B satisfy*

$$\mathrm{tr}_R(\sigma_{BR}) = \sum_{i=1}^{n} x_i v_i v_i^\dagger \quad \textit{and} \quad \mathrm{tr}_B(\sigma_{BR}) = \sum_{i=1}^{n} y_i v_i v_i^\dagger,$$

for some $x, y \in \mathbb{R}^n$ (i.e., they are simultaneously diagonalizable with the input state corresponding to the fixed point subspace).

The implications of this symmetry reduction are as follows. Originally, the primal variable σ_{BR} was defined on the vector space $\mathbb{H}^{n^2}$, which has a dimension over the reals of n^4. By restricting to the fixed-point subspace $\mathcal{V}_\pi$ as defined in (24), this dimension over the reals reduces to $2n^2 - n$. Moreover, following from Corollary 4.14, the partial trace constraint (QRD.b) is reduced from representing n^2 equations to just n equations. This will be particularly useful for the dualized algorithm we propose in Section 5.

We also recognize that the subspace (24) is isomorphic to a block diagonal matrix with $n^2 - n$ blocks of size 1×1 and one block of size $n \times n$. Without loss of generality, we can change the basis of the data of the problem instance to be in the standard basis, i.e., ρ_A is diagonal, meaning that in practice we can implement the subspace using sparse matrix representations. The block diagonal structure also allows us to perform eigendecompositions, and therefore matrix logarithms and exponentials, very efficiently.

4.3 Uniform input state

Let us continue to consider the distortion matrix $\Delta = \mathbb{I} - \rho_{AR}$. An interesting special case is when the input state is the maximally mixed state $\rho_A = \mathbb{I}/n$. It turns out that for this input state, there exist additional symmetries which further simplify the problem, and which ultimately allow us to obtain explicit expressions for the rate-distortion function.

Corollary 4.15. *A solution to* (QRD) *parameterized by the maximally mixed input state* $\rho_A = \mathbb{I}/n$ *and distortion matrix* $\Delta = \mathbb{I} - \rho_{AR}$*, where* ρ_{AR} *is the purification of* ρ_A*, exists in the subspace*

$$\mathcal{V}_\pi = \{a\mathbb{I} + b\rho_{AR} : a, b \in \mathbb{R}\}. \tag{26}$$

Proof. As ρ_A is a multiple of the identity matrix, $g\rho_A g^\dagger = \rho_A$ is satisfied for all $g \in U(n)$, where $U(n)$ is the entire group of $n \times n$ unitary matrices. Therefore Lemma 4.12 tells us that the quantum rate-distortion problem is invariant under the representation (23) of $U(n)$. It follows from Lemma 4.5 that a solution must live in the fixed-point subspace of this representation. Using Corollary 4.10, the dimension over the reals of the fixed-point subspace $\mathcal{V}_\pi$ is given by

$$\dim_{\mathbb{R}} \mathcal{V}_\pi = \int_{U(n)} \operatorname{tr}[g \otimes \overline{g}] \overline{\operatorname{tr}[g \otimes \overline{g}]} = \int_{U(n)} \operatorname{tr}[g]^2 \overline{\operatorname{tr}[g]^2} = 2,$$

where the last equality comes from [37, Theorem 2.1(a)]. It is straightforward to verify that the proposed two-dimensional subspace (26) is invariant under π, which concludes the proof. □

Given this result, we are able to obtain an explicit expression for the quantum rate-distortion function. In particular, we recover and generalize the result from [3, Theorem 10] to maximally mixed inputs of any dimension.

Theorem 4.16. *The quantum rate-distortion function for the maximally mixed input state* $\rho_A = \mathbb{I}/n$ *and distortion matrix* $\Delta = \mathbb{I} - \rho_{AR}$*, where* ρ_{AR} *is the purification of* ρ_A*, is*

$$R_q(D; \rho_A, \Delta) = \begin{cases} \log(n^2) - H\left(\left[1 - D, \underbrace{\frac{D}{n^2-1}, \ldots, \frac{D}{n^2-1}}_{(n^2-1)\ \text{copies}}\right]\right), & \text{if } 0 \le D \le 1 - \frac{1}{n^2} \\ 0, & \text{if } 1 - \frac{1}{n^2} \le D \le 1. \end{cases} \tag{27}$$

A channel that attains this rate for $0 \le D \le 1 - 1/n^2$ *is the depolarizing channel*

$$\mathcal{N}(\rho) = \frac{D}{n^2 - 1}\mathbb{I} + \left(1 - \frac{n^2 D}{n^2 - 1}\right)\rho. \tag{28}$$

Proof. We will first find the solution to the modified quantum rate-distortion problem (QRD) for the specified problem parameters, then use this solution to recover solutions to the original problem (15). From Corollary 4.15, we can restrict our attention

to consider quantum states of the form $\sigma_{BR} = a\mathbb{I} + b\rho_{AR}$ where $a, b \in \mathbb{R}$. Moreover, we have the relationship $b = 1 - n^2 a$ as σ_{BR} must have unit trace, and so the rate-distortion problem reduces to a one-dimensional problem. Therefore, the solution to (QRD) can be found by finding when the gradient of a univariate objective is zero. Doing this we find that the optimal point of (QRD) for a given $\kappa \geq 0$ is

$$\sigma^*_{BR} = \frac{1}{e^\kappa + n^2 - 1}(\mathbb{I} + (e^\kappa - 1)\rho_{AR}),$$

from which we can easily compute the corresponding rate and distortion as

$$\begin{aligned} R_q &= 2\log(n) + \frac{\kappa e^\kappa}{e^\kappa + n^2 - 1} - \log(e^\kappa + n^2 - 1) \\ D &= \frac{n^2 - 1}{e^\kappa + n^2 - 1}. \end{aligned}$$

It is straightforward to verify that $\sigma^*_{BR}(\kappa)$ is a valid solution as it is a feasible point of (QRD) for all $\kappa \geq 0$. Reparameterizing R_q to be in terms of D instead of κ gives the desired rate-distortion function. Similarly, expressing σ^*_{BR} in terms of D instead of κ, we find

$$\sigma^*_{BR} = \frac{D}{n^2 - 1}\mathbb{I} + \left(1 - \frac{n^2 D}{n^2 - 1}\right)\rho_{AR}. \tag{29}$$

As seen in Appendix B, we can recover the Choi matrix from this solution state as $C_\mathcal{N} = n\sigma^*_{BR}$. Noting that $C_\mathcal{N} = \mathbb{I}/n$ corresponds to the quantum channel that maps all states to the maximally mixed state, and $C_\mathcal{N} = n\rho_{AR}$ corresponds to the identity channel, identifies the desired quantum channel. □

4.4 Other quantum distortion matrices

As noted in [3], there are several other common choices for the quantum distortion matrix Δ which have been studied in the literature. As with the entanglement fidelity distortion, we can use symmetry reduction to gain some additional insight for these other common distortion matrices. Throughout this section, we will assume that $\rho_A = \sum_{i=1}^n \lambda_i v_i v_i^\dagger$, and denote $\{b_i\}$ as an orthonormal basis for system B.

We will omit proofs for these results, as they can be shown using an almost identical method as used in Section 4.2. However, we note that for the quantum rate-distortion problem with input state $\rho_A \in \mathcal{D}_n$ and distortion matrix $\Delta \in \mathbb{H}^{mn}_+$ to be invariant under a representation $(\mathbb{H}^{mn}, \pi)$ of a group $\mathcal{G}$, it suffices that

$$\begin{aligned} \pi(g)(X) &= gXg^\dagger, \quad \forall X \in \mathbb{H}^{mn}, \\ \text{where} \quad g &\in \mathcal{G} \leq \{U_B \otimes U_R : U_B \in U(m), U_R \in U(n)\}, \end{aligned} \tag{30}$$

such that $\mathrm{tr}_B(g)\rho_A = \rho_A \mathrm{tr}_B(g)$ and $g\Delta g^\dagger = \Delta$ for all $g \in \mathcal{G}$, and where we use $\mathcal{G} \leq \mathcal{G}'$ to denote that $\mathcal{G}$ is a subgroup of $\mathcal{G}'$. Note that we have one additional requirement that must be satisfied as compared to Lemma 4.12, as the distortion matrix Δ is no longer necessarily related to the input states.

Example 4.17. *Consider the classical-classical distortion matrix*

$$\Delta_{cc} = \sum_{i=1}^m \sum_{j=1}^n \delta_{ij} b_i b_i^\dagger \otimes v_j v_j^\dagger, \tag{31}$$

where $\delta \in \mathbb{R}_+^{m\times n}$. Consider the representation $(\mathbb{H}^{mn}, \pi_{cc})$ with group homomorphism (30) of the group

$$\mathcal{G}_{cc} = \left\{ \left(\sum_{i=1}^{m} w_i b_i b_i^\dagger \right) \otimes \left(\sum_{j=1}^{n} z_j v_j v_j^\dagger \right) : w_i, z_j \in \{\pm 1, \pm\sqrt{-1}\}, \forall i, j \right\}. \tag{32}$$

It can be shown that (QRD) with the classical-classical distortion matrix is invariant under this representation. Therefore, a solution must lie in the corresponding fixed-point subspace

$$\mathcal{V}_{cc} = \left\{ \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} b_i b_i^\dagger \otimes v_j v_j^\dagger : x \in \mathbb{R}^{m\times n} \right\}. \tag{33}$$

As expected (and commented on in [3]), the problem completely reduces down to diagonal matrices, and is therefore isomorphic to the classical rate-distortion problem (see Appendix A).

Example 4.18. *Consider the quantum-classical distortion matrix studied in [23]*

$$\Delta_{qc} = \sum_{i=1}^{m} b_i b_i^\dagger \otimes \Delta_i, \tag{34}$$

where $\Delta_i \in \mathbb{H}_+^n$ for all $i = 1, \ldots, m$. Consider the representation $(\mathbb{H}^{mn}, \pi_{qc})$ with group homomorphism (30) of the group

$$\mathcal{G}_{qc} = \left\{ \left(\sum_{i=1}^{m} w_i b_i b_i^\dagger \right) \otimes \mathbb{I}_R : w_i \in \{\pm 1, \pm\sqrt{-1}\}, \forall i \right\}. \tag{35}$$

It can be shown that (QRD) with the quantum-classical distortion matrix is invariant under this representation. Therefore, a solution must lie in the corresponding fixed-point subspace

$$\mathcal{V}_{qc} = \left\{ \sum_{i=1}^{m} b_i b_i^\dagger \otimes \rho_i : \rho_i \in \mathbb{H}^n, \forall i = 1, \ldots, m \right\}, \tag{36}$$

which can be interpreted as a block-diagonal structure by changing the basis of $\{b_i\}$ to be in the standard basis.

Example 4.19. *Consider the classical-quantum distortion matrix for a system similar to that studied in [38]*

$$\Delta_{cq} = \sum_{j=1}^{n} \Delta_j \otimes v_j v_j^\dagger, \tag{37}$$

where $\Delta_j \in \mathbb{H}_+^m$ for all $j = 1, \ldots, n$. Consider the representation $(\mathbb{H}^{mn}, \pi_{cq})$ with group homomorphism (30) of the group

$$\mathcal{G}_{cq} = \left\{ \mathbb{I}_B \otimes \left(\sum_{j=1}^{n} z_j v_j v_j^\dagger \right) : z_j \in \{\pm 1, \pm\sqrt{-1}\}, \forall j \right\}. \tag{38}$$

It can be shown that (QRD) with the classical-quantum distortion matrix is invariant under this representation. Moreover, the corresponding fixed-point subspace is

$$\mathcal{V}_{cq} = \left\{ \sum_{j=1}^{n} \rho_j \otimes v_j v_j^\dagger : \rho_j \in \mathbb{H}^m, \forall j = 1, \ldots, n \right\}, \tag{39}$$

which can be interpreted as a block-diagonal structure by changing the basis of $\{v_i\}$ to be in the standard basis.

5 Efficient algorithm

We propose mirror descent to compute the quantum rate-distortion function. This is justified by the following relative smoothness properties of mutual information.

Lemma 5.1 ([6, Theorem 6.4])**.** *Consider any input state $\rho_A \in \mathcal{D}_n$. Quantum mutual information $I_q(\cdot;\rho_A)$ is 1-smooth relative to $-S$ on $\mathcal{D}_{mn}$.*

Therefore, following from Fact 2.4, the sequence generated by mirror descent with unit step size $t_k = 1$ and kernel function $-S$ to compute (QRD) will converge sublinearly to the optimal value. In the special case where we consider the classical-quantum distortion matrix (37) considered in Example 4.19, and if additionally, $\Delta_j = \Delta'$ for all $j = 1, \ldots, m$, then mirror descent updates with unit step size are explicitly given by

$$\sigma_{BR}^{k+1} = \frac{\exp(\log(\mathrm{tr}_R(\sigma_{BR}^k)) + \Delta') \otimes \rho_A}{\mathrm{tr}[\exp(\log(\mathrm{tr}_R(\sigma_{BR}^k)) + \Delta')]}. \tag{40}$$

Unfortunately, for a general distortion matrix, it is not obvious how a similar efficiently computable update rule can be obtained. Rather than solve the mirror descent subproblem (10) directly, it is more convenient to solve the dual problem, which using (12) we can show is

$$\max_{\nu} \quad g_q^k(\nu), \tag{41}$$

where

$$g_q^k(\nu) \coloneqq -\mathrm{tr}[\exp(\log(\mathrm{tr}_R(\sigma_{BR}^k)) \otimes \mathbb{I}_R - \mathbb{I}_B \otimes \nu - \kappa\Delta)] - \langle \rho_A, \nu \rangle, \tag{42}$$

and $\nu \in \mathbb{H}^n$ is the dual variable corresponding to the partial trace constraint (QRD.b). Note that we have ignored the unit trace constraint $\mathrm{tr}[\sigma_{BR}] = 1$ as this is already implied by the partial trace constraint. This dual problem (41) is simpler to solve compared to the original primal subproblem as it is both unconstrained and has a significantly smaller problem dimension of n^2 compared to the original dimension of n^2m^2. Once a solution ν^{k+1} to (41) is found, the primal solution can be recovered using (13) as

$$\sigma_{BR}^{k+1} = \exp(\log(\mathrm{tr}_R(\sigma_{BR}^k)) \otimes \mathbb{I}_R - \mathbb{I}_B \otimes \nu^{k+1} - \kappa\Delta). \tag{43}$$

The overall algorithm is summarised in Algorithm 1.

Algorithm 1 Quantum rate-distortion algorithm

Input: Input state $\rho_A \in \mathcal{D}_n$, distortion matrix $\Delta \in \mathbb{H}_+^{mn}$ and dual variable $\kappa \geq 0$

Initialize: Initial point $\sigma_{BR}^0 \in \mathcal{D}_{mn}$

for $k = 0, 1, \ldots$ **do**

$$\nu^{k+1} = \arg\max_{\nu} -\mathrm{tr}[\exp(\log(\mathrm{tr}_R(\sigma_{BR}^k)) \otimes \mathbb{I}_R - \mathbb{I}_B \otimes \nu - \kappa\Delta)] - \langle \rho_A, \nu \rangle$$

$$\sigma_{BR}^{k+1} = \exp(\log(\mathrm{tr}_R(\sigma_{BR}^k)) \otimes \mathbb{I}_R - \mathbb{I}_B \otimes \nu^{k+1} - \kappa\Delta)$$

end for

Output: Approximate solution σ_{BR}^{k+1}.

There are a number of standard algorithms available which can be used to solve the unconstrained convex optimization problem (41). We will consider methods of the form

$$\nu_{i+1}^{k+1} = \nu_i^{k+1} + t_i G_i \nabla g_q^k(\nu_i^{k+1}), \quad \forall i \in \mathbb{N}, \tag{44}$$

where $G_i = \mathbb{I}$ corresponds to gradient ascent, and $G_i = -[\nabla^2 g_q^k(\nu_i^{k+1})]^{-1}$ corresponds to Newton's method. We also take $t_i > 0$ to be a step size obtained from a backtracking procedure [39, Algorithm 9.2] which guarantees that the objective converges monotonically.

Proposition 5.2. *The dual problem (41) satisfies the following properties:*

(i) There exists a unique solution $\nu_* \in \mathbb{H}^n$ to the dual problem, and;

(ii) Solving the problem using gradient descent or damped Newton's method with backtracking will produce a sequence of iterates $\{\nu_i\}$ which converges to ν_*.

Proof. To show part (i), we first show the dual function (42) is coercive and strictly convex by using Proposition 2.7, recognizing that the tensor product $X \mapsto \mathbb{I} \otimes X$ is injective, and using Remark 3.1 to assume $\rho_A \succ 0$ is full-rank. Existence and uniqueness of a solution then follows from [33, Proposition 3.1.1 and 3.2.1]. To prove part (ii), we use the fact that the superlevel sets of the dual function are compact, and that the dual function is strongly concave and has Lipschitz gradient and Hessian over compacts sets. These properties allow us to use standard convergence results of the two algorithms. We leave the details of the proof for part (ii) in Appendix E.2. □

Note that to use either of these descent methods, we need the gradient and Hessian of g_q^k. The gradient can easily be found using [6, Corollary B.5] as

$$\nabla g_q^k(\nu) = \operatorname{tr}_B[\exp(\log(\operatorname{tr}_R(\sigma_{BR}^k)) \otimes \mathbb{I}_R - \mathbb{I}_B \otimes \nu - \kappa\Delta)] - \rho_A. \tag{45}$$

To find the Hessian, let us define the diagonalization

$$\log(\operatorname{tr}_R(\sigma_{BR}^k)) \otimes \mathbb{I}_R - \mathbb{I}_B \otimes \nu - \kappa\Delta = U\Lambda U^\dagger. \tag{46}$$

We use [40, Theorem 6.6.30] to obtain the directional derivative of ∇g_q^k in direction $V \in \mathbb{H}^n$ as

$$\mathsf{D}\nabla g_q^k(\nu)[V] = -\operatorname{tr}_B[U(f^{[1]}(\Lambda) \odot (U^\dagger(\mathbb{I} \otimes V)U))U^\dagger], \tag{47}$$

where $\odot$ represents the Hadamard or element-wise product, and $f^{[1]}(\Lambda)$ is the first divided differences matrix of Λ corresponding to $f(x) = e^x$, i.e., the matrix whose (i,j)-th entry is given by $f^{[1]}(\lambda_i, \lambda_j)$ where

$$f^{[1]}(\lambda, \mu) = \frac{f(\lambda) - f(\mu)}{\lambda - \mu}, \quad \text{if } \lambda \neq \mu \tag{48a}$$

$$f^{[1]}(\lambda, \lambda) = f'(\lambda). \tag{48b}$$

Equation (47) defines the Hessian as a linear map, which we can use to construct the Hessian matrix by expressing it in coordinates with respect to a choice of basis for Hermitian matrices.

Remark 5.3. *We note that Algorithm 1 is very similar to [14, Algorithm 14], which was derived using an expectation-maximization approach. In fact, if we also dualize the redundant unit trace constraint then solve for the corresponding dual variable using the KKT conditions, we obtain the alternate dual function*

$$\tilde{g}_q^k(\tilde{\nu}) := -\log(\operatorname{tr}[\exp(\log(\operatorname{tr}_R(\sigma_{BR}^k)) \otimes \mathbb{I}_R - \mathbb{I}_B \otimes \tilde{\nu} - \kappa\Delta)]) - \langle \rho_A, \tilde{\nu} \rangle. \tag{49}$$

By recognizing the identity $\log(A \otimes B) = \log(A) \otimes \mathbb{I} + \mathbb{I} \otimes \log(B)$, we see that (49) is identical to the function being minimized in the subproblems of [14, Algorithm 14] up to a simple

translation. The advantage of the dual function (42) which we minimize in Algorithm 1 is that the trace expression in $\tilde{g}_q^k$ is composed with a logarithm, and is therefore no longer strictly convex nor coercive. We can confirm this by showing that $\tilde{g}_q^k(t\mathbb{I})$ is a constant for all $t \in \mathbb{R}$. Therefore, we can no longer use the same arguments as Proposition 5.2 to prove that gradient descent or damped Newton's with backtracking will converge to the solution, as without coercivity the superlevel sets of $\tilde{g}_q^k$ are no longer compact. Additionally, our experimental results in Section 6.2 suggest that (49) is a more difficult function to minimize.

Remark 5.4. *Up to this point, we have established theoretical guarantees that mirror descent applied to the quantum rate-distortion problem will result in sublinear convergence to the solution. However, for rate-distortion problems using the entanglement fidelity distortion, empirical convergence results show that mirror descent converges to the solutions at a linear rate (see Figure 1 in Appendix C). Moreover, this linear rate appears to become faster as the parameter κ increases. One standard way to prove global linear convergence of mirror descent is by using a relative strong convexity argument (see Fact 2.4 and, e.g., [6]). However, we can show that quantum mutual information is* not *relatively strongly convex on the set of density matrices (see Appendix C).*

Instead, it is possible to gain some insight into this linear rate by studying the relative strong convexity properties of the quantum mutual information around a neighborhood of the solution to the rate-distortion problems. In particular, using a priori knowledge about the solutions to particular instances of rate-distortions problem analyzed in Section 4.3, we can characterize the local relative strong convexity parameters of quantum mutual information around the solutions, and use them to better understand the linear convergence behavior of mirror descent as a function of κ. We refer the interested reader to Appendix C for a more in depth discussion about these linear rates.

5.1 Inexact mirror descent

As we use a numerical method to solve the mirror descent subproblem (10) for Algorithm 1, it is important to account for numerical errors that will inevitably arise. Existing works [41, 42, 22] have studied the convergence properties of proximal gradient methods where the update steps are numerically computed to some finite tolerance. Of particular interest is that we do not always need to solve the subproblem to high accuracy to guarantee convergence to the optimal value, which can significantly improve computation speeds. We first present a generalized algorithm, then show how to apply it to the quantum rate-distortion problem. We remind the reader that Assumption A is made throughout this section.

General algorithm To solve equality constrained problems of the form (6) using an inexact method, we propose Algorithm 2. This algorithm is adapted from [22] which is an inexact Bregman proximal point method for arbitrary convex constraints, whereas our algorithm is an inexact mirror descent method for linear equality constraints.

The algorithm works as follows. We first warm-start the dual variable ν from the previous iteration (see Step 0 of Algorithm 2) based on the assumption that consecutive subproblems will have similar solutions. A primal variable $x \in \operatorname{dom}\operatorname{int}\varphi$ is recovered using this dual variable using (13) (see (50) in Algorithm 2). However as the dual variable is not necessarily dual optimal, the primal variable is not necessarily primal feasible. Therefore, we define a pseudo-projection $\operatorname{proj}_{\mathcal{C}} : \operatorname{int}\operatorname{dom}\varphi \to \operatorname{relint}\mathcal{C}$ which is continuous, surjective and idempotent, and compute a feasible primal variable as $\tilde{x} = \operatorname{proj}_{\mathcal{C}}(x) \in \mathcal{C}$ (see (51) in

Algorithm 2). Finally, an error criterion based on [22] is used to check if the subproblem has solved to a sufficient accuracy (see Step 2 of Algorithm 2). If not, then ν is updated in an ascent direction for g^k (using e.g., gradient ascent or Newton's method), and the process repeats until the desired accuracy has been reached.

Algorithm 2 Inexact mirror descent algorithm

Input: Objective function f, Legendre reference function φ where $\operatorname{dom}\varphi = \operatorname{dom} f$, linear equality constraint data $\mathcal{A}, b$, step size $\{t_k\}$, error tolerance $\{\varepsilon_k\}$.

Initialize: Initial primal $x^0 \in \mathcal{C}$ and dual $\nu^0 \in \mathbb{V}'$ variables.

for $k = 0, 1, \ldots$ **do**

Step 0: Initialize dual variable from previous iteration, i.e., $\nu_0^{k+1} = \nu^k$.

for $i = 0, 1, \ldots$ **do**

Step 1: Recover primal variable from dual variable, i.e.,

$$x_i^{k+1} = \nabla\varphi^{-1}[\nabla\varphi(x^k) - t_k(\nabla f(x^k) + \mathcal{A}^\dagger(\nu_i^{k+1}))] \tag{50}$$

$$\tilde{x}_i^{k+1} = \operatorname{proj}_{\mathcal{C}}(x_i^{k+1}), \tag{51}$$

where $\operatorname{proj}_{\mathcal{C}} : \operatorname{int}\operatorname{dom}\varphi \to \operatorname{relint}\mathcal{C}$ is continuous, surjective and idempotent.

Step 2: **if** subproblem has solved to sufficient accuracy, i.e., satisfies

$$D_\varphi(\tilde{x}_i^{k+1} \| x_i^{k+1}) \leq \varepsilon_k, \tag{52}$$

then

$\nu^{k+1} := \nu_i^{k+1}$, $x^{k+1} := x_i^{k+1}$, and $\tilde{x}^{k+1} := \tilde{x}_i^{k+1}$

break

else

Update dual variable, i.e., $\nu_{i+1}^{k+1} = \nu_i^{k+1} + t_i^{k+1}\Delta\nu_i^{k+1}$, where t_i^{k+1} is an appropriate step size and $\Delta\nu_i^{k+1}$ is an ascent direction for g^k.

end if

end for

end for

Output: Approximate solution $\tilde{x}^{k+1}$.

We now present the convergence properties of Algorithm 2. We first show that the exit criterion will always eventually hold.

Proposition 5.5. *Consider the problem (9) for some $y \in \operatorname{relint}\mathcal{C}$ and Legendre function φ. Consider the sequence $\{\nu_i\} \in \mathbb{V}'$, and let $\{x_i\}$ and $\{\tilde{x}_i\}$ be the corresponding sequences generated by (50) and (51) respectively, where $x^k = y$. If $\{\nu_i\}$ converges to the dual optimum ν_* of the dual subproblem (14), then $D_\varphi(\tilde{x}_i \| x_i) \to 0$.*

Proof. Let us denote x_* as the primal optimum of (9). From Fact 2.5, we have $\{x_i\} \in \operatorname{int}\operatorname{dom}\varphi$ and $x_* \in \operatorname{int}\operatorname{dom}\varphi$. Using Fact 2.2(ii), we can show $(\nabla\varphi)^{-1}$ is continuous, and so x_i is a continuous function of ν_i from (50). Therefore, convergence of $\{\nu_i\}$ to ν_* implies $\{x_i\}$ converges to a solution of $\inf_x \mathcal{L}(x, \nu_*; y)$, which from strict convexity of φ must be the unique primal optimum x_* [43, Theorem 12.13]. From continuity of the pseudo-projection operator on $\operatorname{int}\operatorname{dom}\varphi$ and primal feasibility of x_*, it follows that $\tilde{x}_i = \operatorname{proj}_{\mathcal{C}}(x_i)$ must also converge to x_*. The desired result then follows from continuity of D_φ on $\operatorname{int}\operatorname{dom}\varphi \times \operatorname{int}\operatorname{dom}\varphi$, and the property $D_\varphi(x \| x) = 0$ for all $x \in \operatorname{int}\operatorname{dom}\varphi$. □

Remark 5.6. *For the quantum rate-distortion problem, convergence of the iterates $\{\nu_i\}$ to the optimal solution of the subproblem is guaranteed by Proposition 5.2 if we use gradient ascent or Newton's method with a backtracking line search.*

We now move onto proving convergence rates of the inexact mirror descent iterates. We begin with a preliminary result which takes advantage of the fact that we restrict our attention to problems with only linear equality constraints.

Lemma 5.7. *The pair of primal iterates $(x^{k+1}, \tilde{x}^{k+1})$ produced by (50) and (51) always satisfy*

$$0 \in \nabla f(x^k) + N_{\mathcal{A}}(\tilde{x}^{k+1}) + \frac{1}{t_k}(\nabla\varphi(x^{k+1}) - \nabla\varphi(x^k)), \tag{53}$$

where $N_{\mathcal{A}}$ denotes the normal cone of the affine set $\{x \in \mathbb{V} : \mathcal{A}(x) = b\}$.

Proof. As $\tilde{x}^{k+1}$ is primal feasible, it follows that $N_{\mathcal{A}}(\tilde{x}^{k+1})$ is equal to the image of $\mathcal{A}^\dagger$. Rearranging (50) then substituting this in gives

$$\frac{1}{t_k}(\nabla\varphi(x^k) - \nabla\varphi(x^{k+1})) - \nabla f(x^k) = \mathcal{A}^\dagger(\nu^k) \in N_{\mathcal{A}}(\tilde{x}^{k+1}),$$

for all $\nu^k \in \mathbb{V}'$, from which (53) follows. □

The following main convergence result can been seen as a modification of [22, Theorem 3.2].

Theorem 5.8. *Consider Algorithm 2 to solve the convex optimization problem (6). Let φ be Legendre, f^* represent the optimal value of this problem, and x^* be any corresponding optimal point.*

(i) If f is L-smooth relative to φ and $t_k = 1/L$, then the sequence $\{(x^k, \tilde{x}^k)\}$ satisfies

$$f(x^k_{avg}) - f^* \leq \frac{L}{k}\left(D_\varphi(x^* \| x^0) + \sum_{i=0}^{k-1} \varepsilon_k\right), \quad \forall k \in \mathbb{N}, \tag{54}$$

where $x^k_{avg} = \sum_{i=1}^k \tilde{x}^i/k$.

(ii) If f is L-smooth relative to φ, μ-strongly convex relative to φ, and $t_k = 1/L$, then the sequence $\{(x^k, \tilde{x}^k)\}$ satisfies

$$D_\varphi(x^* \| x^k) \leq \left(1 - \frac{\mu}{L}\right)^k \left(D_\varphi(x^* \| x^0) + A\right), \quad \forall k \in \mathbb{N}, \tag{55}$$

where

$$A = \sum_{i=0}^{k-1} \left(1 - \frac{\mu}{L}\right)^{-1-i} \varepsilon_i. \tag{56}$$

Proof. From L-smoothness of f, we have

$$f(\tilde{x}^{k+1}) \leq f(x^k) + \langle \nabla f(x^k), \tilde{x}^{k+1} - x^k \rangle + L D_\varphi(\tilde{x}^{k+1} \| x^k).$$

By substracting $f(u)$ from both sides of the inequality for any $u \in \mathcal{C}$ and using μ-strong convexity of f, we obtain

$$f(\tilde{x}^{k+1}) - f(u) \leq \langle \nabla f(x^k), \tilde{x}^{k+1} - u \rangle + L D_\varphi(\tilde{x}^{k+1} \| x^k) - \mu D_\varphi(u \| x^k). \tag{57}$$

Using Lemma 5.7 and the definition of the normal cone $N_{\mathcal{A}}(\tilde{x}^{k+1})$ of $\{x \in \mathbb{V} : \mathcal{A}(x) = b\}$ at $\tilde{x}^{k+1}$, we can show that

$$\begin{aligned}\langle \nabla f(x^k), \tilde{x}^{k+1} - u\rangle &\leq L\langle \nabla\varphi(x^k) - \nabla\varphi(x^{k+1}), \tilde{x}^{k+1} - u\rangle \\ &= L(D_\varphi(\tilde{x}^{k+1}\|x^{k+1}) + D_\varphi(u\|x^k) - D_\varphi(\tilde{x}^{k+1}\|x^k) - D_\varphi(u\|x^{k+1})) \\ &\leq L(D_\varphi(u\|x^k) - D_\varphi(\tilde{x}^{k+1}\|x^k) - D_\varphi(u\|x^{k+1}) + \varepsilon_k),\end{aligned}$$

for any $u \in \mathcal{C}$, where the equality uses the definition of the Bregman divergence (8), and the second inequality uses (52). Combining this with (57) and letting $u = x^*$ gives

$$f(\tilde{x}^{k+1}) - f^* \leq (L - \mu)D_\varphi(x^*\|x^k) - LD_\varphi(x^*\|x^{k+1}) + L\varepsilon_k. \tag{58}$$

To show the part (i), let $\mu = 0$, i.e., we make no assumption of relative strong convexity. From convexity of f, we find that

$$\begin{aligned}f(x^k_{avg}) - f^* &\leq \frac{1}{k}\sum_{i=1}^{k}(f(\tilde{x}^i) - f^*) \\ &\leq \frac{L}{k}\left(D_\varphi(x^*\|x^0) - D_\varphi(x^*\|x^k) + \sum_{i=0}^{k-1}\varepsilon_k\right),\end{aligned}$$

as the sum telescopes, which recovers the desired result. To show part (ii), we recognize that $f^* \leq f(x)$ for all $x \in \mathcal{C}$, and therefore from (58) we obtain

$$D_\varphi(x^*\|x^{k+1}) \leq \left(1 - \frac{\mu}{L}\right)D_\varphi(x^*\|x^k) + \varepsilon_k.$$

Recursively applying this inequality recovers the desired result. □

Remark 5.9. *If $\varepsilon_k = \varepsilon$ is a constant for all k, then (54) implies ergodic sublinear convergence to an $L\varepsilon$-optimal solution. If f is μ-strongly convex relative to φ, then (55) implies linear convergence to a $L\varepsilon/\mu$-Bregman ball of the solution. If $\{\varepsilon_k\}$ is a summable (i.e., $\lim_{k\to\infty}\sum_{i=0}^{k}\varepsilon_k < \infty$), then (54) implies ergodic sublinear convergence to the optimal solution. If f is μ-strongly convex relative to φ and $\{\varepsilon_k\}$ is a sequence which converges to zero linearly, then (55) implies linear convergence to the solution. See [42, Proposition 3] for a more in-depth discussion on how the linear rate varies depending on how quickly $\{\varepsilon_k\}$ converges. Notably, the best linear convergence behavior occurs when $\{\varepsilon_k\}$ convergences at a similar rate as that of the function values.*

Remark 5.10. *The criterion (52) is equivalent to*

$$\langle \nabla f(x^k), \tilde{x}^{k+1}\rangle + \frac{1}{t_k}D_\varphi(\tilde{x}^{k+1}\|x^k) - g^k(\nu^{k+1}) \leq \frac{\varepsilon_k}{t_k}, \tag{59}$$

i.e., a bound on the primal-dual gap of the mirror descent subproblem (10).

Application to quantum rate-distortion To use Algorithm 2 for the quantum rate-distortion problem, all that remains is to propose a suitable pseudo-projection (51). We propose the following operator

$$\operatorname{proj}_{\mathcal{C}}(\sigma_{BR}) = P\sigma_{BR}P^\dagger \tag{60a}$$

$$\text{for} \quad P = \mathbb{I}_B \otimes (\rho_A^{1/2}\operatorname{tr}_B(\sigma_{BR})^{-1/2}). \tag{60b}$$

It follows from Lemma 4.11, where $X = R$, $Y = B$, $A_X = (\rho_A^{1/2} \operatorname{tr}_B(\sigma_{BR})^{-1/2})$ and $U_Y = \mathbb{I}_B$, that the proposed pseudo-projection maps positive definite matrices to positive definite matrices satisfying (QRD.b). Idempotence follows by recognizing that for any $\sigma_{BR} \in \mathbb{H}_{++}^{mn}$ satisfying (QRD.b), i.e., $\operatorname{tr}_B(\sigma_{BR}) = \rho_A$, the pseudo-projection satisfies $\operatorname{proj}_{\mathcal{C}}(\sigma_{BR}) = \sigma_{BR}$ because the matrix P from (60b) satisfies

$$\begin{aligned} P &= \mathbb{I}_B \otimes (\rho_A^{1/2} \operatorname{tr}_B(\sigma_{BR})^{-1/2}) \\ &= \mathbb{I}_B \otimes (\rho_A^{1/2} \rho_A^{-1/2}) \\ &= \mathbb{I}_B \otimes \mathbb{I}_A. \end{aligned}$$

Continuity of the pseudo-projection follows from the fact that the matrix inverse and square root are continuous on the positive definite cone.

Therefore, (60a) is a suitable pseudo-projection for the quantum rate-distortion problem. Note that the standard Euclidean projection is not suitable as it maps variables to the boundary of the domain, where the gradient of quantum mutual information is not well-defined. Finally, we summarize the inexact mirror descent algorithm applied to the quantum rate-distortion problem in Algorithm 3 and its corresponding convergence guarantees.

Algorithm 3 Inexact quantum rate-distortion algorithm

Input: Input state $\rho_A \in \mathcal{D}_n$, distortion matrix $\Delta \in \mathbb{H}_+^{mn}$, error tolerance $\{\varepsilon_k\}$.
Initialize: Initial primal $\sigma_{BR}^0 \in \mathcal{D}_{mn}$ and dual $\nu^0 \in \mathbb{H}^n$ variables.
for $k = 0, 1, \ldots$ **do**
Step 0: Initialize dual variable from previous iteration, i.e., $\nu_0^{k+1} = \nu^k$.
for $i = 0, 1, \ldots$ **do**
Step 1: Recover primal variable from dual variable, i.e.,

$$\begin{aligned} \sigma_{BR,i}^{k+1} &= \exp(\log(\operatorname{tr}_R(\sigma_{BR}^k)) \otimes \mathbb{I}_R - \mathbb{I}_B \otimes \nu_i^{k+1} - \kappa\Delta) \\ \tilde{\sigma}_{BR,i}^{k+1} &= P\sigma_{BR,i}^{k+1}P^\dagger, \end{aligned}$$

where $P = \mathbb{I}_B \otimes (\rho_A^{1/2} \operatorname{tr}_B(\sigma_{BR,i}^{k+1})^{-1/2})$.
Step 2: **if** subproblem has solved to sufficient accuracy, i.e., satisfies

$$S(\tilde{\sigma}_{BR,i}^{k+1} \| \sigma_{BR,i}^{k+1}) - \operatorname{tr}[\tilde{\sigma}_{BR,i}^{k+1}] + \operatorname{tr}[\sigma_{BR,i}^{k+1}] \leq \varepsilon_k.$$

then
$\nu^{k+1} := \nu_i^{k+1}$, $\sigma_{BR}^{k+1} := \sigma_{BR,i}^{k+1}$, and $\tilde{\sigma}_{BR}^{k+1} := \tilde{\sigma}_{BR,i}^{k+1}$
break
else
Update dual variable ν_{i+1}^{k+1} using (44).
end if
end for
end for
Output: Approximate solution $\tilde{\sigma}_{BR}^{k+1}$.

Corollary 5.11. *Consider the quantum rate-distortion problem (QRD) for some input state $\rho_A \in \mathcal{D}_n$ and distortion matrix $\Delta \in \mathbb{H}_+^{mn}$. Let $\hat{I}^*$ denote its optimal value, and σ_{BR}^* be any corresponding optimal point. The sequence $\{(\sigma_{BR}^k, \tilde{\sigma}_{BR}^k)\}$ generated by Algorithm 3*

satisfies

$$I_q(\sigma_{avg}^k;\rho_A)+\kappa\langle\Delta,\sigma_{avg}^k\rangle-\hat{I}^*\leq\frac{1}{k}\left(S(\sigma_{BR}^*\|\sigma_{BR}^0)+\sum_{i=0}^{k-1}\varepsilon_k\right),\quad\forall k\in\mathbb{N},\tag{61}$$

where $\sigma_{avg}^k=\sum_{i=1}^k\tilde{\sigma}_{BR}^i/k$.

Proof. This follows from Lemma 5.1 and Theorem 5.8. □

6 Numerical experiments

We present numerical experiments to demonstrate the computational performance of Algorithm 3 in solving the quantum rate-distortion problem (QRD). Input states ρ_A are randomly generated using QETLAB [44], and the entanglement fidelity was used to measure distortion by letting $\Delta=\mathbb{I}-\rho_{AR}$, where ρ_{AR} is the purification of ρ_A. All experiments were ran on MATLAB using an Intel i5-11700 CPU with 32GB of RAM.

For all implementations of Algorithm 3, unless otherwise stated, we use an error schedule of

$$\varepsilon_k=\max\{\min\{f(x^k)-f(x^{k-1}),\xi^k,\varepsilon_{k-1}\},10^{-15}\},\tag{62}$$

where $\xi=0.9$ and $\varepsilon_{-1}=10^{-2}$. Based on the discussion from Remarks 5.4 and 5.9, the first term in the minimum aims to match the linear rate of the error sequence to that of the mirror descent iterations, the second term ensures that the error schedule decays linearly, the third term ensures the sequence is monotonic, and the last term accounts for machine precision. For brevity, we will use MD-N to refer to when Newton's method with backtracking is used to compute the inexact mirror descent iterates, and MD-GD to refer to when gradient descent with backtracking is used. We terminate MD-N once the change in objective value drops below 10^{-15}, and MD-GD at 10^{-8} as the sublinear rate of gradient descent makes computing the iterates to higher accuracy impractical. All algorithms are also terminated if the computation time exceeds 3600 s. We use "timeout" to refer to situations when an algorithm is unable to perform a single iteration within this timeframe, and "out of memory" to refer to when there is insufficient RAM to run the algorithm. All backtracking line searches use the implementation described by [39, Algorithm 9.2], where $\alpha=0.1$ and $\beta=0.1$. For gradient descent methods, we start backtracking from $t_0=1000$, while for Newton's method we use $t_0=1$. We report upper bounds on absolute optimality gaps measured in bits, and unless otherwise stated are measured by computing a lower bound of the optimal value using MD-N and the method described in Appendix D. Where relevent, methods are initialized using $\sigma_{BR}^0=\rho_A\otimes\rho_A$ and $\nu^0=-\log(\rho_A)$.

6.1 Comparison between exact and inexact methods

We first present some preliminary results comparing exact and inexact computation of the mirror descent iteration. We compare between variations of Algorithm 3 when the iterates are computed using Newton's method with backtracking exactly, i.e., close to machine precision $\varepsilon_k=10^{-15}$ for all k, and inexactly using (62). We also compare against when iterates are computed using gradient descent with backtracking inexactly using (62), and to a constant error of $\varepsilon_k=10^{-8}$ for all k as proposed in [14, Algorithm 4]. All methods use symmetry reduction to reduce the problem dimension.

We summarize the results in Table 1. We see that using Newton's method to compute the mirror descent iterates yields identical accuracies between exact and inexact implementations, despite the inexact method solving up to 6 times faster. This difference also seems

Table 1: Comparison between various strategies of choosing the tolerance to which mirror descent iterates are computed, when computing the quantum rate-distortion function with entanglement fidelity distortion at channel dimensions n and distortion dual variables κ. We report the time elapsed (s) and upper bounds on the absolute optimality gap (bits) of each method. All methods use symmetry reduction to reduce the problem dimension.

		MD-N ($\varepsilon_k = 10^{-15}$)		MD-N ($\varepsilon_k = (62)$)		MD-GD ($\varepsilon_k = 10^{-8}$)		MD-GD ($\varepsilon_k = (62)$)	
n	κ	Time	Gap	Time	Gap	Time	Gap	Time	Gap
8	1.0	.32	**7e−9**	.17	**7e−9**	.34	1e−7	**.10**	1e−7
	3.0	.10	**4e−9**	**.07**	**4e−9**	.38	1e−8	.14	8e−9
32	3.0	5.52	**6e−8**	1.52	**6e−8**	.90	2e−7	**.45**	2e−7
	5.5	.90	**5e−9**	**.32**	**5e−9**	2.07	1e−8	.81	6e−8
128	7.0	85.70	**3e−8**	**14.40**	**3e−8**	34.75	6e−8	14.97	6e−8
	8.5	31.83	**6e−10**	**8.91**	**6e−10**	57.51	7e−9	16.85	7e−8

more pronounced for lower values of κ. Similarly, both inexact methods using gradient descent solve to similar accuracies, while the proposed decaying error solves approximately 3 times faster compared to solving to a constant tolerance.

6.2 Comparison with other algorithms

We compare Algorithm 3 to existing algorithms which have been used to compute the quantum rate-distortion function. These include the backtracking primal-dual hybrid gradient (PDHG) algorithm from [6, Algorithm 3], the expectation maximization (EM) algorithm from [14, Algorithm 14], and CvxQuad [17, 18] with the SDPT3 [45] solver. We show results with and without symmetry reduction for all of these methods. For PDHG, we use the same backtracking parameters as used in the experiments of [6]. As noted in Remark 5.3, the EM algorithm is very similar to our mirror descent algorithm except that it solves a slight variation of the dual problem. Therefore to allow for a fair comparison, we use the same inexact method as Algorithm 2 using gradient descent to compute these subproblems rather than solving to a constant inexact tolerance as was originally proposed in [14], and which we showed was computationally slower in Section 6.1. Both the PDHG and EM methods were terminated once the change in objective value dropped below 10^{-8}. The default settings for CvxQuad were used, and we report the optimality gap returned by the solver.

Remark 6.1. *Just as our proposed algorithm shares similarities to EM, the inexact mirror descent method in Algorithm 1 also shares similarities to PDHG. In particular, if we use gradient descent to compute the inexact mirror descent iterates, then, like PDHG, we are alternating between primal mirror descent iterates, and dual gradient ascent iterates. The main difference is that we may perform multiple dual steps for every primal step we take.*

We summarize the results without symmetry reduction in Table 2a, and results with symmetry reduction in Table 2b. We see that symmetry reduction results in a significant improvement in both computation time and memory requirements. We emphasize that without symmetry reduction, the quantum rate-distortion for a 9-qubit channel ($n = 512$) needs to optimize over a 70 billion dimensional variable. Symmetry reduction reduces the optimization variable to a dimension of just 523,776.

Overall, we see that our proposed method is able to obtain high-accuracy solutions relatively quickly. In particular, our method generally outperforms all existing benchmarks in both computation time and accuracy simultaneously. We note that it is not entirely

Table 2: Comparison between various algorithms to compute for the quantum rate-distortion function with entanglement fidelity distortion, at channel dimensions n and distortion dual variables κ. We report the time elapsed (seconds) and upper bounds on the absolute optimality gap (bits) of each method. We use "> 3600.00" to refer when the algorithm has not solved to the desired tolerance within an hour, and report the optimality gap of the last completed iteration at this time. We use "Timeout" to refer to when the algorithm has not performed any iterations within an hour.

(a) Without symmetry reduction

		MD-N		MD-GD		PDHG [6]		EM [14]		CvxQuad [17, 18]	
n	κ	Time	Gap	Time	Gap	Time	Gap	Time	Gap	Time	Gap
2	1.0	.19	**4e−9**	.12	2e−8	**.11**	1e−7	.20	2e−5	3.00	8e−9
	3.0	.13	**1e−8**	.10	2e−8	**.05**	2e−6	.22	5e−5	2.94	2e−8
8	1.0	4.88	**2e−8**	2.30	1e−7	**.84**	6e−6	7.56	4e−5	Out of memory	
	3.0	**1.12**	**1e−8**	2.59	2e−8	1.40	9e−7	6.06	3e−5	Out of memory	
32	3.0	> 3600.00	**2e−7**	**332.16**	4e−7	355.38	3e−5	> 3600.00	6e−3	Out of memory	
	5.5	1321.85	**1e−7**	**631.83**	2e−7	1500.75	5e−5	> 3600.00	4e−3	Out of memory	
128	7.0	Timeout		Timeout		$> \mathbf{3600.00}$	**2e−1**	Timeout		Out of memory	
	8.5	Timeout		Timeout		$> \mathbf{3600.00}$	**1e−1**	Timeout		Out of memory	
512	9.5	Out of memory		Out of memory		Out of memory		Out of memory		Out of memory	
	11.0	Out of memory		Out of memory		Out of memory		Out of memory		Out of memory	

(b) With symmetry reduction

		MD-N		MD-GD		PDHG [6]		EM [14]		CvxQuad [17, 18]	
n	κ	Time	Gap	Time	Gap	Time	Gap	Time	Gap	Time	Gap
2	1.0	.06	4e−9	**.05**	2e−8	**.05**	7e−8	.08	2e−8	2.89	**2e−9**
	3.0	.05	1e−8	**.04**	2e−8	.07	7e−8	.07	5e−5	2.61	**2e−9**
8	1.0	.17	**7e−9**	**.10**	1e−7	.12	6e−6	.53	1e−6	454.98	2e−8
	3.0	**.07**	**4e−9**	.16	8e−9	.19	6e−7	.41	5e−8	686.11	6e−8
32	3.0	1.52	**6e−8**	**.45**	2e−7	1.84	2e−5	14.83	2e−5	Out of memory	
	5.5	**.32**	**5e−9**	.81	6e−8	6.51	5e−5	17.25	5e−7	Out of memory	
128	7.0	**14.40**	**3e−8**	14.97	6e−8	47.07	1e−4	> 3600.00	6e−6	Out of memory	
	8.5	**8.91**	**6e−10**	16.84	7e−8	100.55	1e−4	> 3600.00	6e−7	Out of memory	
512	9.5	2174.38	**7e−8**	**140.97**	1e−7	1650.75	3e−4	> 3600.00	9e−5	Out of memory	
	11.0	1216.96	**5e−9**	**575.86**	1e−8	> 3600.00	5e−4	> 3600.00	1e−4	Out of memory	

clear whether Newton's method or gradient descent is better to solve our mirror descent subproblems. At high problem dimensions, gradient descent solves significantly faster to similar accuracy solutions. At low to medium problem dimensions however, Newton's method seems to solve faster for larger values of κ, while gradient descent solves faster for smaller κ. However, Newton's method allows us to solve the mirror descent iterates "exactly", which in turn allows us to efficiently compute an optimality gap using the method from Appendix D.

7 Concluding remarks

We have presented an inexact mirror descent algorithm which is able to efficiently compute the quantum rate-distortion function with provable sublinear convergence rates. Additionally, we have shown how we can exploit symmetries that exist in common quantum rate-distortion settings to significantly reduce the dimensions of the problems we need to solve. It would be interesting to see if our techniques and analysis can be extended to other similar problems, such as the quantum rate-distortion for mixed states [26] and the quantum information bottleneck function [38, 46].

In this work, we have focused on a reformulation of the quantum rate-distortion prob-

lem where we dualize the distortion inequality constraint, as is typically done for these constrained channel capacity-type problems [4]. However, by converting the distortion inequality to an equality constraint using [14, Lemma 21], it is relatively straightforward to extend most of our algorithm and analysis to directly account for the distortion constraint. The only limitation is that there is not an obvious pseudo-projection (51) for this new problem. Therefore, future work to find an efficient pseudo-projection operator will allow us to use our inexact mirror descent algorithm to directly solve the unsimplified rate-distortion problem. Alternatively, we could perform exact computations of the mirror descent steps, which, although slower, circumvents the need for a pseudo-projection.

Computation of the mirror descent subproblems is an important part of inexact mirror descent which significantly impacts how effective the overall algorithm is. In Proposition 5.2, we established qualitative properties of the dual function which allows us to guarantee global convergence when maximizing it using standard techniques. However, we lack a more quantitative understanding of these functions which may allow us to have a better understanding of convergence rates. A more in-depth analysis of the dual function may allow us to understand how to best solve the mirror descent subproblems.

A Classical rate-distortion

In this section, we study the classical rate-distortion function, showing how symmetry reduction and mirror descent can be applied to this problem. The purpose of this is to provide a point of comparison for the concepts we use to study the quantum rate-distortion function, and to show how many of our results still apply to the classical setting.

A.1 Notation

We define n-dimensional classical states as n-dimensional probability distributions, i.e.,

$$\mathcal{P}_n \coloneqq \left\{ p \in \mathbb{R}^n_+ : \sum_{i=1}^n p_i = 1 \right\}.$$

Classical channels are linear functions that map n-dimensional classical states to m-dimensional classical states, and are mathematically represented as $m \times n$ column stochastic matrices

$$\mathcal{Q}_{m,n} \coloneqq \left\{ Q \in \mathbb{R}^{m \times n}_+ : \sum_{i=1}^m Q_{ij} = 1, \forall j = 1, \dots, n \right\}.$$

Probabilities defined on the Cartesian product of two sample spaces are represented by joint probability matrices of the form

$$\mathcal{P}_{n \times m} \coloneqq \left\{ p \in \mathbb{R}^{n \times m}_+ : \sum_{i=1}^n \sum_{j=1}^m p_{ij} = 1 \right\}.$$

With some overloading of function notation, we define the Shannon entropy $H : \mathbb{R}^n_+ \to \mathbb{R}$ as

$$H(x) \coloneqq -\sum_{i=1}^n x_i \log(x_i), \tag{63}$$

and the Kullback-Leibler (KL) divergence $H : \mathbb{R}^n_+ \times \mathbb{R}^n_{++} \to \mathbb{R}$ as

$$H(x \,\|\, y) \coloneqq \sum_{i=1}^n x_i \log(x_i / y_i). \tag{64}$$

Throughout this section we will use $\mathbf{1}$ to denote the all ones column vector.

A.2 Problem definition

For a classical channel with n inputs and m outputs, consider a probability distribution $p \in \mathcal{P}_n$, a distortion matrix $\delta \in \mathbb{R}_+^{m\times n}$ where δ_{ij} represents the distortion of producing the i-th output from the j-th input, and a maximum allowable total distortion $D \geq 0$. The corresponding *classical rate-distortion function* [1] is

$$R_c(D; p, \delta) \quad := \quad \min_{Q\in\mathcal{Q}_{m,n}} \quad I_c(Q; p) \tag{65a}$$

$$\text{subj. to} \quad \sum_{i=1}^{m}\sum_{j=1}^{n} p_j Q_{ij}\delta_{ij} \leq D, \tag{65b}$$

where

$$I_c(Q; p) \coloneqq H(Q\operatorname{diag}(p) \| Qpp^\top), \tag{66}$$

is known as the *classical mutual information*. It follows from joint convexity of the KL divergence [47, Theorem 2.7.2] that (65) is a convex optimization problem.

When there is the same number of inputs and outputs, i.e., $m = n$, a common distortion measure for the classical rate-distortion function is the Hamming distance [47], which corresponds to $\delta = \mathbf{1}\mathbf{1}^\top - \mathbb{I}$ (i.e., the square matrix with zeros on the diagonal and ones everywhere else).

Remark A.1. *Like the quantum analog (see Remark 3.1), without loss of generality, we will always assume that the input distribution of the classical channel is non-degenerate, i.e., $p > 0$. See Appendix B.2 for more details.*

Change of variables We will make a change of variables where we optimize over the joint distribution $P \in \mathcal{P}_{m\times n}$, where $P_{ij} = p_j Q_{ij}$ is the joint probability of obtaining the i-th output from the j-th input. For distortion $D \geq 0$, input distribution $p \in \mathcal{P}_n$, and distortion matrix $\delta \in \mathbb{R}_+^{m\times n}$, this gives us

$$R_c(D; p, \delta) \quad = \quad \min_{P\in\mathcal{P}_{m\times n}} \quad I_c(P; p) \tag{67a}$$

$$\text{subj. to} \quad \sum_{i=1}^{m} P_{ij} = p_j, \qquad j = 1, \ldots, n, \tag{67b}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} P_{ij}\delta_{ij} \leq D, \tag{67c}$$

where with some slight abuse of notation we redefine classical mutual information as

$$I_c(P; p) \coloneqq H(P \| P\mathbf{1}p^\top). \tag{68}$$

Proposition A.2. *The classical rate-distortion problems (65) and (67) are equivalent, in the sense that the optimal values of the two problems are equal.*

Proof. See Appendix B.3. □

Distortion constraint Like the quantum case, we will reparameterize the distortion constraint with the dual variable $\kappa \geq 0$ instead of D. The classical rate-distortion problem

for distortion dual variable $\kappa \geq 0$, input distribution $p \in \mathcal{P}_n$, and distortion matrix $\delta \in \mathbb{R}_+^{m\times n}$ then becomes

$$\tilde{R}_c(\kappa; p, \delta) = \min_{P\in\mathcal{P}_{m\times n}} I_c(P;p) + \kappa \sum_{i=1}^{m}\sum_{j=1}^{n} P_{ij}\delta_{ij} \tag{CRD.a}$$

$$\text{subj. to} \quad \sum_{i=1}^{m} P_{ij} = p_j, \quad j = 1, \ldots, n, \tag{CRD.b}$$

A.3 Symmetry reduction

To show how we can achieve a similar result as Theorem 4.16 for the classical scenario, consider the classical rate-distortion problem with Hamming distortion $\delta = \mathbf{1}\mathbf{1}^\top - \mathbb{I}$ and a uniform input distribution $p = \mathbf{1}/n$. The analytic expression for this setting is well known (see, e.g., [47, Exercise 10.5]), which is typically derived by lower bounding the mutual information using Fano's inequality, then showing achievability of this lower bound. Here, we show how this result can also be derived by using a similar symmetry reduction as we used to derive the quantum rate-distortion function.

Corollary A.3. *Let $\mathcal{G}$ be a compact group, let $(\mathbb{R}^n, \tau)$ be a real representation of $\mathcal{G}$, and let $(\mathbb{R}^{n\times n}, \pi)$ be the real representation of $\mathcal{G}$ defined by $\pi(g)(X) = \tau(g)X\tau(g)^\dagger$. Then the dimension (over $\mathbb{R}$) of the fixed point subspace is given by $\dim_{\mathbb{R}}(\mathcal{V}_\pi) = \langle \chi_\tau, \chi_\tau \rangle$.*

Proof. This follows from a similar argument as the proof for Corollary 4.10. We first extend $(\mathbb{R}^n, \tau)$ to a complex representation in the natural way. We then note that $(\mathbb{C}^{n\times n}, \pi)$ decomposes into isomorphic subrepresentations given by the restriction to real and imaginary subspaces. □

Lemma A.4. *Let us define the group S_n as the group of permutations on $\{1, \ldots, n\}$, and a representation $(\mathbb{R}^{n\times n}, \pi)$ of S_n as*

$$\pi(g)(X) = P_g X P_g^\top, \quad \forall X \in \mathbb{R}^{n^2\times n^2},$$

where P_g is the matrix representation of permutation g. Problem (CRD) for uniform input distribution $p = \mathbf{1}/n$ and Hamming distortion $\delta = \mathbf{1}\mathbf{1}^\top - \mathbb{I}$ is invariant under π.

Proof. Classical mutual information and the constraint (CRD.b) are invariant under column-wise and row-wise permutations. Moreover, as conjugation by permutation matrices always maps diagonal elements to diagonals, and off-diagonal elements to off-diagonals, δ is also invariant under π. □

Corollary A.5. *A solution to (CRD) for input distribution $p = \mathbf{1}/n$ and Hamming distortion $\delta = \mathbf{1}\mathbf{1}^\top - \mathbb{I}$ exists in the subspace*

$$\mathcal{V}_\pi = \{a\mathbb{I} + b\mathbf{1}\mathbf{1}^\top : a, b \in \mathbb{R}\}. \tag{70}$$

Proof. It follows from Lemmas 4.5 and A.4 that a solution must live in the fixed-point subspace of the representation $(\mathbb{R}^{n\times n}, \pi)$ defined in Lemma A.4. It is well known that the matrix representation $(\mathbb{C}^n, \tau)$ of the permutation group S_n (i.e., $\tau(g) = P_g$) can be decomposed as the direct sum of two irreducible subrepresentations $V = \{\mathbf{1}\}$ and $W = \{x \in \mathbb{C}^n : \langle \mathbf{1}, x\rangle = 0\}$. Therefore using Corollary A.3, we can show the dimension over the reals of the corresponding fixed-point subspace is

$$\dim_{\mathbb{R}} \mathcal{V}_\pi = \langle \chi_\tau, \chi_\tau \rangle = \langle \chi_V + \chi_W, \chi_V + \chi_W \rangle = 2,$$

where the second and third equalities use Lemmas 4.7 and 4.8 respectively. It is straightforward to verify that the proposed two-dimensional subspace is invariant under π, which concludes the proof. □

Theorem A.6. *The classical rate-distortion function with uniform input distribution $p_j = \mathbf{1}/n$ and Hamming distortion $\delta = \mathbf{1}\mathbf{1}^\top - \mathbb{I}$ is given by*

$$R_c(D; p, \delta) = \begin{cases} \log(n) - H\left(\left[1-D, \underbrace{\frac{D}{n-1}, \ldots, \frac{D}{n-1}}_{(n-1)\ \text{copies}}\right]\right), & \text{if } 0 \leq D \leq 1 - \frac{1}{n} \\ 0, & \text{if } 1 - \frac{1}{n} \leq D \leq 1. \end{cases} \tag{71}$$

A channel which attains this rate for $0 \leq D \leq 1 - 1/n$ is given by

$$Q = \frac{D}{n-1}\mathbf{1}\mathbf{1}^\top + \left(1 - \frac{nD}{n-1}\right)\mathbb{I}. \tag{72}$$

Proof. This follows from a similar proof to that of Theorem 4.16. □

A.4 Mirror descent

Like the quantum case, we propose mirror descent to compute the classical rate-distortion function. This is justified by the following relative smoothness properties of mutual information.

Lemma A.7 ([6, Theorem 6.4])**.** *Consider any input distribution $p \in \mathcal{P}_n$. Classical mutual information $I_c(\cdot\,; p)$ is 1-smooth relative to $-H$ on $\mathcal{P}_{m\times n}$.*

Therefore, following from Fact 2.4, mirror descent with a unit step size $t_k = 1$ and kernel function $-H$ is guaranteed to converge sublinearly to the optimal value. The mirror descent update (10) for the classical rate-distortion problem can be expressed as the following analytic expression

$$P_{ij}^{k+1} = p_j \frac{q_i^k \exp(-\kappa\delta_{ij})}{\sum_{i=1}^m q_i^k \exp(-\kappa\delta_{ij})} \quad \text{for} \quad i = 1, \ldots, m,\ j = 1, \ldots, n. \tag{73}$$

where $q_i^k = \sum_{j=1}^n P_{ij}^k$ is the marginal probability of obtaining the i-th output. This iteration is precisely the same as the Blahut-Arimoto algorithm [4].

Like the quantum rate-distortion problem with the entanglement fidelity distortion measure, empirical results show that using the mirror descent iterates (73) to compute the classical rate-distortion with Hamming distortion results in a linear convergence rate. More details about this can be found in Appendix C.

B Equivalencies between rate-distortion problems

To prove that two optimization problems are equivalent, it suffices to show that we can use a feasible point of one optimization problem to obtain a feasible point of the other problem with the same objective value, and vice versa.

To show equivalencies between quantum rate-distortion problems, we will first introduce the Choi representation of a quantum channel, which will be a convenient mathematical representation than (2) for us to use to analyze the quantum rate-distortion problem.

Definition B.1 (Choi operator)**.** *Consider a linear map $\mathcal{N} : \mathbb{H}^n \to \mathbb{H}^m$ which maps Hermitian matrices on an n-dimensional input system A to Hermitian matrices on an m-dimensional output system B. The* Choi matrix representation *$C_{\mathcal{N}} \in \mathbb{H}^{mn}$ of $\mathcal{N}$, which acts on the bipartite system BA, is defined as*

$$C_{\mathcal{N}} \coloneqq \sum_{ij}^{n} \mathcal{N}(v_i v_j^\dagger) \otimes v_i v_j^\dagger, \tag{74}$$

where $\{v_i\}$ is any orthonormal basis for A.

A linear operator $\mathcal{N}$ is completely positive if and only if $C_{\mathcal{N}} \succeq 0$ [29, Theorem 2.22], and is trace preserving if and only if $\mathrm{tr}_B(C_{\mathcal{N}}) = \mathbb{I}_A$ [29, Theorem 2.26]. Applying a linear operator on input state $\rho_A \in \mathcal{D}_n$ is equivalent to $\mathcal{N}(\rho_A) = \mathrm{tr}_A(C_{\mathcal{N}}(\mathbb{I}_B \otimes \rho_A))$ [29, Equation (2.66)].

B.1 Basis independence

We will first show that that the quantum rate-distortion problem is independent of the eigenvectors of the input state ρ_A by showing that we can construct an equivalent optimization problem by performing a suitable unitary transform on systems A and R. For a given spectral decomposition $\rho_A = \sum_{i=1} \lambda_i v_i v_i^\dagger$ of ρ_A, let us define the unitary $U = \sum_{i=1}^{n} v_i u_i^\dagger$ where $\{u_i\}$ is any orthonormal basis for A. We can define a new problem with input state $U^\dagger \rho_A U$, purification $(U \otimes U)^\dagger \rho_{AR} (U \otimes U)$ and distortion matrix $(\mathbb{I} \otimes U)^\dagger \Delta (\mathbb{I} \otimes U)$. For any feasible channel $\mathcal{N}$ of the original problem (15), we can construct feasible channel $\mathcal{N}'$ for the new problem with the same objective value by letting

$$C_{\mathcal{N}'} = (\mathbb{I} \otimes U)^\dagger C_{\mathcal{N}} (\mathbb{I} \otimes U).$$

Equivalency in the reverse direction follows from an identical argument. We note that a similar argument can be used to show that rate-distortion problem is independent of the choice of purification, even when R is chosen not to be isomorphic to A.

B.2 Degenerate and non-degenerate inputs

Here, we show that we can always assume that the input states p and ρ_A are non-degenerate (i.e., $p > 0$ and $\rho_A \succ 0$), as it is possible to construct equivalent optimization problems of a lower input dimension to avoid any degeneracies. For the classical rate-distortion function, we can construct a new problem where we simply ignore the j-th element of p_j and column of δ for all j corresponding to $p_j = 0$. Clearly, given any feasible channel Q of the original problem (65) we can construct a feasible channel Q' to the reduced problem with the same objective value by dropping the corresponding columns of Q. In the reverse direction, we just need to populate the columns of Q' we removed with any valid probability distribution.

For the quantum rate-distortion function, let us assume $\{v_i\}$ for $i = 1, \ldots, n'$ are the eigenvectors corresponding to non-zero eigenvalues of the (possibly rank-deficient) matrix ρ_A. Similar to the argument for basis independence, we can define a new n'-dimensional system A' with orthonormal basis $\{u_i\}$ and an isometry $V = \sum_{i=1}^{n'} v_i u_i^\dagger$. We use this to define a new problem with a full-rank input state $V^\dagger \rho_A V$, purification $(V \otimes V)^\dagger \rho_{AR} (V \otimes V)$ and distortion matrix $(\mathbb{I} \otimes V)^\dagger \Delta (\mathbb{I} \otimes V)$. For any feasible channel $\mathcal{N}$ of the original problem (15), we can construct feasible channel $\mathcal{N}'$ for the reduced problem with the same objective value by letting

$$C_{\mathcal{N}'} = (\mathbb{I} \otimes V)^\dagger C_{\mathcal{N}} (\mathbb{I} \otimes V).$$

In the reverse direction, we have

$$C_{\mathcal{N}} = (\mathbb{I} \otimes V) C_{\mathcal{N}'} (\mathbb{I} \otimes V)^\dagger + \sum_{i=n'+1}^{n} \rho_i \otimes v_i v_i^\dagger,$$

where $\rho_i \in \mathcal{D}_m$ is any valid density matrix for all $i = n' + 1, \ldots, n$.

B.3 Optimizing over channel and joint output state

Now given non-degeneracy of p, it is fairly straightforward to show that we can map between feasible channels Q and states P which obtain the same objective values between (65) and (67) using

$$P_{ij} = p_j Q_{ij} \quad \text{and} \quad Q_{ij} = P_{ij}/p_j,$$

for all $i = 1, \ldots, m$, $j = 1, \ldots, n$. Note that this is well defined as $p_j > 0$ for all $j = 1, \ldots, n$. Similarly for the quantum setting, given that ρ_A has spectral decomposition $\rho_A = \sum_{i=1} \lambda_i v_i v_i^\dagger$, without loss of generality we will first represent σ_{BR} and $\mathcal{N}$ as

$$\sigma_{BR} = \sum_{ij}^{n} X_{ij} \otimes v_i v_j^\dagger \quad \text{and} \quad C_{\mathcal{N}} = \sum_{ij}^{n} Y_{ij} \otimes v_i v_j^\dagger,$$

for suitable $X_{ij}, Y_{ij} \in \mathbb{H}^m$. Given this, we can map from feasible channels $\mathcal{N}$ and states σ_{BR} which obtain the same objective values between (15) and (17) using

$$\sigma_{BR} = \sum_{ij}^{n} \sqrt{\lambda_i \lambda_j}\, Y_{ij} \otimes v_i v_j^\dagger \quad \text{and} \quad C_{\mathcal{N}} = \sum_{ij}^{n} \frac{1}{\sqrt{\lambda_i \lambda_j}} X_{ij} \otimes v_i v_j^\dagger.$$

Again, this is well-defined as $\lambda_i > 0$ for all $i = 1, \ldots, n$ from the assumption that ρ_A is full-rank. Note that the first relationship is equivalent to $\sigma_{BR} = (\mathcal{N} \otimes \mathbb{I})(\rho_{AR})$.

$$C_{\mathcal{N}'} = (\mathbb{I} \otimes U_A)^\dagger C_{\mathcal{N}} (\mathbb{I} \otimes U_A).$$

C Linear convergence rates

In Figure 1, we show the empirical convergence behaviors of exact mirror descent on the classical and quantum rate-distortion problems for the Hamming distortion $\delta = \mathbf{1}\mathbf{1}^\top - \mathbb{I}$ and entanglement fidelity distortion $\Delta = \mathbb{I} - \rho_{AR}$ respectively, for various values of κ and randomly generated inputs. We observe linear convergence for all examples, and note that this linear rate appears to become faster as κ increases. These observations were also shared across multiple randomly generated input states.

From Fact 2.4, this suggests that classical and quantum mutual information might be strongly convex relative to Shannon and von Neumann entropy respectively. However, simple counterexamples show that this is false. For a continuously differentiable function f to be strongly convex relative to a Legendre function φ on a set $\mathcal{X}$, there must be a $\mu > 0$ that satisfies

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \mu (D_\varphi(x \| y) + D_\varphi(y \| x)),$$

for all $x, y \in \operatorname{relint} \mathcal{X}$ (see e.g., [12, Proposition 1.1(b)]). For classical mutual information with input distribution $p \in \mathcal{P}_n$, we can show that for $X, Y \in \mathcal{P}_{m \times n}$ with columns $X_j = p_j x$ and $Y_j = p_j y$ for any $x, y \in \mathcal{P}_m$ for all $j = 1, \ldots, n$ that

$$\langle \nabla I_c(X; p) - \nabla I_c(Y; p), X - Y \rangle = 0,$$

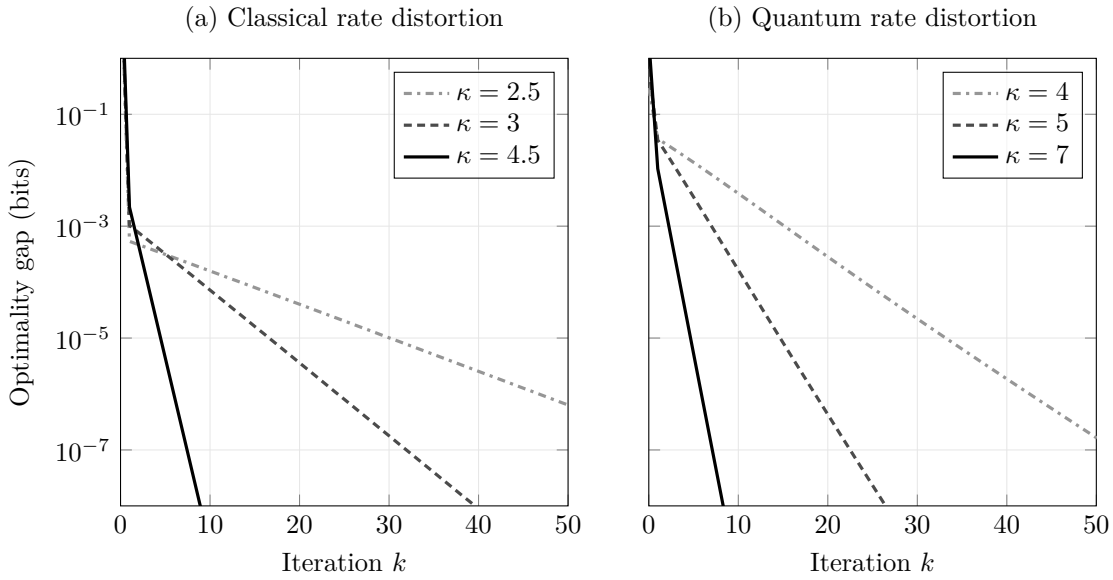

Figure 1: Convergence behaviors of computing the classical and quantum rate-distortion functions using mirror descent for Hamming distortion and entanglement fidelity distortion and for randomly generated inputs of dimension $n = 32$.

Similarly, for the quantum case with input state $\rho_A \in \mathcal{D}_n$, we can show that for $\sigma_{BR} = \sigma_B \otimes \rho_A$ and $\omega_{BR} = \omega_B \otimes \rho_A$ for any $\sigma_B, \omega_B \in \mathcal{D}_m$, that

$$\langle \nabla I_q(\sigma_{BR}; \rho_A) - \nabla I_q(\omega_{BR}; \rho_A), \sigma_{BR} - \omega_{BR} \rangle = 0,$$

by using additivity of quantum relative entropy [28, Property 11.2.1]. Therefore, we conclude that classical and quantum mutual information are not relatively strongly convex on their respective domains. Instead, it is possible to gain insights into this linear convergence behavior through the *local* convexity properties of mutual information at the solutions to the rate-distoriton problems.

C.1 Preliminaries

We will first introduce the notion of local relative strong convexity.

Definition C.1. *Let f and φ be twice continuously differentiable at some x, and let φ be Legendre. Then f is locally μ-strongly convex relative to φ at x if there exists a $\mu > 0$ such that*

$$\nabla^2 f(x) - \mu \nabla^2 \varphi(x) \succeq 0.$$

We also introduce the following additional assumptions on the kernel function φ. The first two are standard assumptions used to prove convergence of the mirror descent iterates to the solution, rather than just convergence of the function values.

Assumption B. *The kernel function $\varphi : \mathbb{V} \to \bar{\mathbb{R}}$ satisfies the following.*

(i) The right partial level sets $\{y \in \operatorname{int} \operatorname{dom} \varphi : D_\varphi(x \| y) \leq \alpha\}$ are bounded for all $x \in \operatorname{dom} \varphi$ and $\alpha \in \mathbb{R}$.

(ii) If a sequence $\{y^k\} \in \operatorname{int} \operatorname{dom} \varphi$ converges to $x \in \operatorname{dom} \varphi$, then $D_\varphi(x \| y^k) \to 0$.

(iii) The function φ is strongly convex.

We now present some results for local linear convergence.

Proposition C.2. *Consider problem (6) where x^* is any corresponding optimal solution, f and φ are twice continuously differentiable at x^*, and φ is Legendre and satisfies Assumption B. If f is locally μ-strongly convex relative to φ at x^*, then the following results hold.*

(i) The optimal solution x^* is the unique optimal solution.

(ii) Let $\{x^k\}$ be the sequence of iterates generated by using the mirror descent iterates (10) to solve (6). If f is L-smooth relative to φ and $t_k = 1/L$ for all k, then there exists a $\mu' \in (0, \mu)$ and $N \in \mathbb{N}$ such that the sequence $\{x^k\}$ satisfies

$$D_\varphi(x^* \| x^k) \leq \left(1 - \frac{\mu'}{L}\right)^k D_\varphi(x^* \| x^N), \tag{75}$$

for all $k > N$.

Proof. As $\mu > 0$ and φ satisfies Assumption B(iii), meaning $\nabla^2\varphi(x) \succ 0$ for all $x \in \operatorname{int}\operatorname{dom}\varphi$, then there exists a $\mu' \in (0, \mu)$ such that

$$\nabla^2 f(x^*) - \mu' \nabla^2 \varphi(x^*) \succ 0.$$

As f and φ are both twice continuously differentiable, then there exists an $\varepsilon > 0$ such that

$$\nabla^2 f(x) - \mu' \nabla^2 \varphi(x) \succeq 0,$$

for all $\|x - x^*\| < \varepsilon$. For part (i), it follows from Assumption B(iii) that f is strictly convex within this neighborhood, in which x^* must therefore be the unique solution. As the set of solutions to (6) must be convex, x^* must be the unique solution on the entire feasible set.

For part (ii), we first note that given L-relative smoothness of f and Assumption B that the mirror descent iterates $\{x^k\}$ converge to the unique solution x^* (this follows from a similar argument as, e.g., [11, Theorem 2(ii)]). Therefore there exists an $N \in \mathbb{N}$ such that $\|x^k - x^*\| < \varepsilon$ for all $k > N$. Linear convergence within this region follows from Fact 2.4. □

Remark C.3. *The kernel functions $\varphi = -H$ and $\varphi = -S$ satisfy Assumption B(i) and Assumption B(ii). Additionally, they satisfy Assumption B(iii) on any bounded set (notably, they are 1-strongly convex with respect to the 1- and trace-norm over $\mathcal{P}_n$ and $\mathcal{D}_n$ respectively [6, Proposition 3.12]).*

C.2 Analysis of rate-distortion problems

If we can characterize the local relative strong convexity parameters of mutual information at the solutions to the rate-distortion problems, then Proposition C.2 allows us to predict the convergence behavior of mirror descent around a neighborhood of the solution. Knowing these local relative strong convexity parameters a priori can be difficult for arbitrary parameterizations of the rate-distortion problems. However, in Section 4.3, we derived explicit expressions for the optimal solutions of the classical rate-distortion problem for uniform input distribution and Hamming distortion (Theorem A.6), and the quantum rate-distortion problem for maximally mixed input and entanglement fidelity distortion (Theorem 4.16). Therefore, we can compute the local relative strong convexity parameters

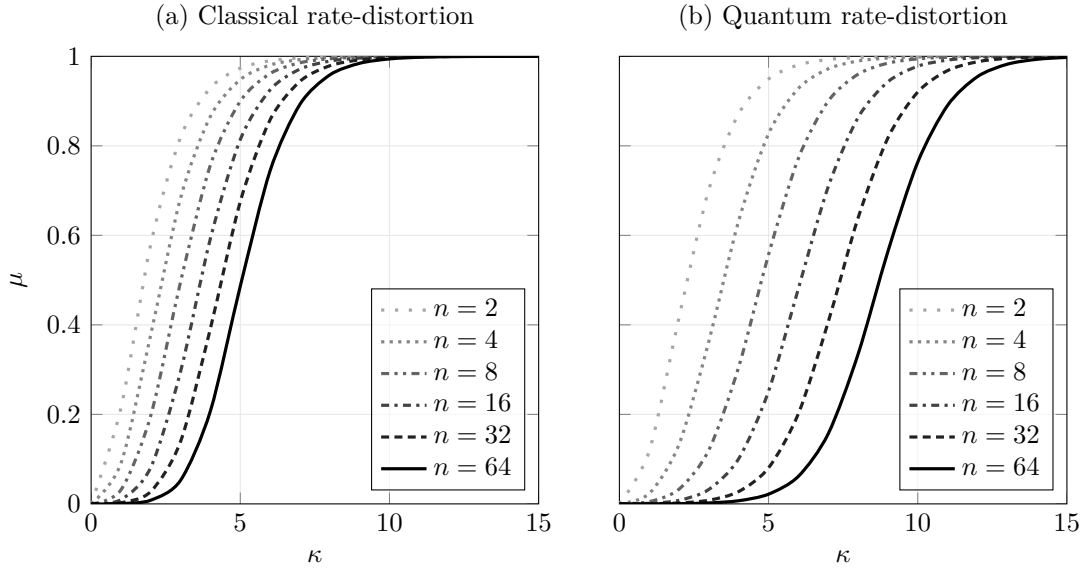

Figure 2: Local relative strong convexity parameters μ at the optimal point of the (a) classical rate-distortion problem with uniform input distribution and Hamming distortion and (b) quantum rate-distortion problem with maximally mixed input state and entanglement fidelity distortion, at different problem dimensions n and distortion dual variables κ.

of mutual information at these solutions by solving what amounts to a generalized eigenvalue problem. If we consider (6) where $\mathbb{V} = \mathbb{R}^n$, this amounts to finding the smallest μ that satisfies

$$(U^\dagger \nabla^2 f(x^*) U)v = \mu (U^\dagger \nabla^2 \varphi(x^*) U)v,$$

for some $v \in \mathbb{R}^m \setminus \{0\}$, where $m = \dim_{\mathbb{R}}(\ker \mathcal{A})$ and $U \in \mathbb{R}^{n \times m}$ is an isometry with columns forming an orthonormal basis of $\ker \mathcal{A}$ such that $UU^\dagger$ is a projection onto this nullspace.

Figure 2 shows these values of the local relative strong convexity parameter μ for various values of distortion dual variables κ and problem dimensions n for both of these specific classical and quantum rate-distortion problems. Notably, we see that for both the classical and quantum scenarios that μ increases monotonically from 0 to 1 as κ increases from 0 to infinity. Combining these observations with Lemma 5.1 and Proposition C.2, we conclude that mirror descent applied to solve these particular instances of rate-distortion problems will converge linearly around a neighborhood around the solution. Moreover, this linear rate becomes arbitrarily fast as κ tends to infinity.

As seen in Figure 1, these observations reflect our empirical findings for the classical rate-distortion problem with Hamming distortion and quantum rate-distortion problem with entanglement fidelity distortion, not only for uniform input distributions and maximally mixed input states, but for *any* input state. We remark that it is possible to rigorously prove that the $\lim_{\kappa \to \infty} \mu(\kappa) = 1$ property seen in Figure 2 holds for any input state.

D Frank-Wolfe lower bound

D.1 Preliminaries

We will first remind the reader about Frank-Wolfe type lower bounds for convex optimization problems over compact sets, then show how these can be applied to the classical and quantum rate-distortion problems. Consider (6) where $\mathcal{C}$ is a compact set. As f is convex,

any linear approximation will always lower bound the function. That is, for any $x, y \in \mathcal{C}$

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle. \tag{76}$$

By finding the infimum of the right hand side over y over $\mathcal{C}$, we can obtain a value which lower bounds the function. As $\mathcal{C}$ is compact, this lower bound will always exist, and therefore we define

$$\begin{aligned} f_{lb}(x) &\coloneqq \min_{y \in \mathcal{C}} \{ f(x) + \langle \nabla f(x), y - x \rangle \} \\ &= f(x) - \langle \nabla f(x), x \rangle + \min_{y \in \mathcal{C}} \{ \langle \nabla f(x), y \rangle \}. \end{aligned} \tag{77}$$

We can use this to define the estimated optimality gap $e(x)$

$$e(x) \coloneqq f(x) - f_{lb}(x) = \max_{y \in \mathcal{C}} \{ \langle \nabla f(x), x - y \rangle \}, \tag{78}$$

which upper bounds the true optimality gap, i.e., $e(x) \geq f(x) - f^*$. We also show that this error bound converges to zero as x converges to an optimal point.

Proposition D.1. *Consider the constrained convex optimization problem (6) where $\mathcal{C}$ is a compact set. At an optimal point x^*, the estimated optimality gap equals zero $e(x^*) = 0$. Moreover, if f is continuously differentiable, for any sequence $\{x^k\}$ that converges to x^*, the sequence $\{e(x^k)\}$ will converge to zero.*

Proof. The first order optimality condition for a constrained minimization problem (6) is

$$-\nabla f(x^*) \in N_{\mathcal{C}}(x^*), \tag{79}$$

where $N_{\mathcal{C}}$ is the normal cone of $\mathcal{C}$ (see e.g. [33]). It follows from the definition of the normal cone and the optimality condition (79) that for all $x \in \mathcal{C}$

$$\langle \nabla f(x^*), x^* - x \rangle \leq 0. \tag{80}$$

Noting that the optimality gap $e(x^*)$ maximizes the left hand side over $x \in \mathcal{C}$ and is non-negative shows that $e(x^*) = 0$ as desired.

To show convergence of the optimality gap to zero, we recognize that $e(x)$ is the sum of a continuous function and the support function of a compact set. As support functions on bounded sets are continuous [48, Proposition 8.6], from continuous differentiability of f and the fact that compositions of continuous functions are continuous, we conclude that $e(x)$ is a continuous function, from which the desired result follows. □

Over certain sets, there are well-known ways to compute the minimization required in (77), see, e.g., [49].

D.2 Application to rate-distortion problems

We now show how the Frank-Wolfe lower bound can be applied to both the classical and quantum rate-distortion problems.

Remark D.2. *Using the mirror descent interpretation of Blahut-Arimoto algorithms [6], the lower bounds derived for Blahut-Arimoto algorithms applied to classical [4] and quantum [50, 51] channel capacities are identical to the Frank-Wolfe lower bound (78).*

Classical rate-distortion For the classical rate-distortion problem (CRD), let us consider the set $\mathcal{P}' = \{P \in \mathcal{P}_{m\times n} : \sum_{i=1}^m P_{ij} = p_j, j = 1, \ldots, n\}$ where $p \in \mathcal{P}_n$. Over this set, we find

$$\min_{y\in\mathcal{P}'}\{\langle \nabla f(x), y\rangle\} = \sum_{j=1}^n p_j \min_i\{\partial_{ij} f(x)\}. \tag{81}$$

where $\partial_{ij} f$ is the partial derivative of f with respect to the (i,j)-th coordinate, by noting the minimization is separable into minimizing each column of P over a scaled probability simplex.

Quantum rate-distortion For the quantum rate-distortion problem (QRD), consider the set $\mathcal{D}' = \{\sigma_{BR} \in \mathcal{D}_{mn} : \mathrm{tr}_B(\sigma_{BR}) = \sigma\}$ where $\sigma \in \mathcal{D}_n$ with spectral decomposition $\sigma = \sum_{i=1}^n \lambda_i v_i v_i^\dagger$. Unfortunately, it is not obvious how to compute the lower bound (77) efficiently over this set. However, if the gradient is of the form

$$\nabla f(x) = \sum_{i=1}^m \sum_{j=1}^n g_{ij} b_i b_i^\dagger \otimes v_j v_j^\dagger, \tag{82}$$

where $g_{ij} \in \mathbb{R}$ for all i, j and $\{b_i\}$ is any orthonormal bases for B, then a symmetry reduction argument shows that the left-hand side of (83) is invariant under the same representation as in Example 4.17, and therefore the optimal value is an element of the subspace (33). Therefore, the minimization is separable in the same way as (81), and

$$\min_{y\in\mathcal{D}'}\{\langle \nabla f(x), y\rangle\} = \sum_{j=1}^n \lambda_j \min_i\{g_{ij}\}. \tag{83}$$

Although (82) this seems like a very limiting assumption, we can show this holds for the quantum rate-distortion problem setting we are most interested in, i.e., where distortion is measured using the entanglement fidelity, and if iterates are generated using Algorithm 1. To see this, consider (QRD) parameterized by input state $\rho_A \in \mathcal{D}_n$ and distortion matrix $\Delta = \mathbb{I} - \rho_{AR}$. By substituting (43) into the gradient of quantum mutual information [6, Equation (61)], we can express the gradient of the objective function as

$$\nabla f(\sigma_{BR}^k) = (\log(\mathrm{tr}_R(\sigma_{BR}^{k-1})) - \log(\mathrm{tr}_R(\sigma_{BR}^k))) \otimes \mathbb{I} - \mathbb{I} \otimes \nu^k. \tag{84}$$

From Theorem 4.13 and Corollary 4.14, we know $\mathrm{tr}_R(\sigma_{BR}^k) = \sum_{i=1}^n \alpha_i v_i v_i^\dagger$ and $\nu^k = \sum_{j=1}^n \beta_j v_j v_j^\dagger$ for some $\alpha, \beta \in \mathbb{R}^n$ for all k, and therefore the gradient is of the desired form (82).

This property of the gradients no longer holds in general if we use the inexact method in Algorithm 3. However, we can still use this lower bounding technique whenever we solve a step exactly, which we can choose to do at the last iteration before terminating the algorithm to obtain an optimality gap for the algorithm's final output.

E Background proofs

E.1 Proof of Proposition 2.7

We first present a simple result about how strict convexity and coercivity are preserved by injective linear maps.

Lemma E.1. *Consider a function $f : \mathbb{V} \to \mathbb{R}$ and an affine operator $\mathcal{A} : \mathbb{V} \to \mathbb{V}'$.*

(i) If f is strictly convex, then $h(x) = f(\mathcal{A}^\dagger(x))$ is strictly convex if and only if $\mathcal{A}^\dagger$ is injective.

(ii) If f is coercive, then $h(x) = f(\mathcal{A}^\dagger(x))$ is coercive if and only if $\mathcal{A}^\dagger$ is injective.

Proof. For part (i), consider any $x, y \in \mathbb{V}'$ where $x \neq y$. If $\mathcal{A}^\dagger$ is injective, then $\mathcal{A}^\dagger(x) \neq \mathcal{A}^\dagger(y)$, and therefore from strict convexity of f we have for $\lambda \in [0,1]$

$$f(\mathcal{A}^\dagger(\lambda x + (1-\lambda)y)) = f(\lambda \mathcal{A}^\dagger(x) + (1-\lambda)\mathcal{A}^\dagger(y)) < \lambda f(\mathcal{A}^\dagger(x)) + (1-\lambda) f(\mathcal{A}^\dagger(y)), \tag{85}$$

which proves strict convexity of h. If $\mathcal{A}^\dagger$ is not injective, then there exists $x, y \in \mathbb{V}$ where $x \neq y$ such that $\mathcal{A}^\dagger(x) = \mathcal{A}^\dagger(y)$. For this example, the inequality (85) must hold with equality throughout, showing that h is not strictly convex.

For part (ii), consider a sequence x^k satisfying $\|x^k\| \to \infty$. Let $a = \min\{\|\mathcal{A}^\dagger(x)\| : \|x\| = 1\}$. Clearly $a > 0$ if and only if $\mathcal{A}^\dagger$ is injective (otherwise $\mathcal{A}^\dagger$ has nontrivial kernel), in which case $a\|x^k\| \leq \|\mathcal{A}^\dagger(x^k)\|$ for all $k \in \mathbb{N}$, and so $\|\mathcal{A}^\dagger(x^k)\| \to \infty$. Therefore, it follows from coercivity of f that h is also coercive. If $\mathcal{A}^\dagger$ is not injective, then there exists a direction $d \in \ker \mathcal{A}^\dagger$ that we can use to construct the sequence $x^k = x^0 + kd$. This satisfies $\|x^k\| \to \infty$, but $h(x^k) = h(x^0)$ for all $k \in \mathbb{N}$, and so h is not coercive. □

We now present the main proof.

Proof of Proposition 2.7. Part (i) follows from Fact 2.2(i) and Lemma E.1(i). For part (ii), we first use [52, Fact 2.11] to show that because φ^* is closed, convex and proper, that the function $x \mapsto \varphi^*(x) - \langle s, x\rangle$ is coercive if and only if $s \in \operatorname{int}\operatorname{dom}\varphi$. Using Lemma E.1(ii), it follows that

$$\nu \mapsto \frac{1}{t}\varphi^*[\nabla\varphi(y) - t(\nabla f(y) + \mathcal{A}^\dagger(\nu))] - \left\langle \frac{s}{t}, \nabla\varphi(y) - t(\nabla f(y) + \mathcal{A}^\dagger(\nu))\right\rangle,$$

is coercive if and only if $s \in \operatorname{int}\operatorname{dom}\varphi$ and $\mathcal{A}^\dagger$ is injective. Letting $b = \mathcal{A}(s)$ and discarding constant terms gives us the negative dual function $-g(\cdot\,; y)$, which concludes the proof. □

E.2 Proof of Proposition 5.2(ii)

We will first show that the dual function has a negative definite Hessian.

Proposition E.2. *The dual function g_q^k, as defined in (42), has negative definite Hessian.*

Proof. From (47) and using the relationship $\mathsf{D}^2 g_q^k(\nu)[V,V] = \langle V, \mathsf{D}\nabla g_q^k(\nu)[V]\rangle$, we have

$$\mathsf{D}^2 g_q^k(\nu)[V,V] = -\sum_{ij}^{mn} f^{[1]}(\Lambda)_{ij}(U^\dagger(\mathbb{I}_B \otimes V)U)_{ij}^2,$$

where U, Λ are given by (46) and $f^{[1]}(\Lambda)$ is the first divided differences matrix, as defined in (48), of Λ corresponding to $f(x) = e^x$. As the exponential function is strictly positive and monotone, all elements of the divided differences matrix $f^{[1]}(\Lambda)$ are strictly positive. Moreover, as the tensor product is injective, $U^\dagger(\mathbb{I}\otimes\nu)U = 0$ if and only if $\nu = 0$. Therefore, we conclude that

$$\mathsf{D}^2 g_q^k(\nu)[V,V] < 0,$$

for all ν and non-zero V, which proves the desired result. □

We will also briefly remind the reader about these fairly standard results which establish Lipschitz-continuity and strong convexity over compact sets.

Lemma E.3. *Consider a continuously differentiable function $f : \mathcal{U} \to \mathbb{V}$ where $\mathcal{U}$ is an open subset of an inner product space $\mathbb{V}'$, and a convex compact set $\mathcal{X} \subset \mathcal{U}$. Then f is Lipschitz on $\mathcal{X}$, meaning that there exists an $L \geq 0$ such that $\|f(x) - f(y)\| \leq L\|x - y\|$ for all $x, y \in \mathcal{X}$.*

Proof. This is a straightfoward consequence of the mean value theorem [53, Theorem X.4.5] and convexity of $\mathcal{X}$. □

Lemma E.4. *Consider a twice continuously differentiable function $f : \mathcal{U} \to \mathbb{R}$ where $\mathcal{U}$ is an open subset of an inner product space $\mathbb{V}$, and a convex compact set $\mathcal{X} \subset \mathcal{U}$. If f has a positive definite Hessian, then f is strongly convex on $\mathcal{X}$, meaning that there exists a $\mu > 0$ such that $\mathsf{D}^2 f(x)[v, v] \geq \mu$ for all $x \in \mathcal{X}$ and $v \in \mathbb{V}$ satisfying $\|v\| = 1$.*

Proof. Given twice continuous differentiability of f, we know $\mathsf{D}^2 f(x)[v, v]$ is a continuous function in both x and v. The desired result then follows by taking the infumum of $\mathsf{D}^2 f(x)[v, v]$ over the product of the compact sets $\mathcal{X}$ and $\{v \in \mathbb{V} : \|v\| = 1\}$. □

We now present the main proof.

Proof of Proposition 5.2(ii). Given that gradient descent and Newton's method with backtracking both result in a monotonic sequence of function values, for some initialization $\nu_0 \in \mathbb{H}^n$, we can constrain our analysis to the superlevel set $\mathcal{S} = \{\nu \in \mathbb{H}^n : g_q^k(\nu) \geq g_q^k(\nu_0)\}$ which is compact from coercivity of $-g_q^k(\nu)$. Over this compact set, from Lemmas E.3 and E.4 we establish that $g_q^k(\nu)$ is strongly concave, and has Lipschitz gradient and Hessian. These properties are sufficient to establish global convergence of the function values to the optimal value by using gradient descent and Newton's method with backtracking [39, Sections 9.3.1 and 9.5.3]. Convergence of the iterates to the unique solution follows from strong convexity and Lipshitz gradient (see e.g., the discussion in [39, Section 9.1.2]). □

References

[1] Claude Elwood Shannon. "A mathematical theory of communication". In: *The Bell System Technical Journal* 27.3 (July 1948), pp. 379–423. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

[2] Nilanjana Datta, Min-Hsiu Hsieh, and Mark M. Wilde. "Quantum rate distortion, reverse Shannon theorems, and source-channel separation". In: *IEEE Transactions on Information Theory* 59.1 (2013), pp. 615–630. DOI: 10.1109/tit.2012.2215575.

[3] Mark M Wilde et al. "Quantum rate-distortion coding with auxiliary resources". In: *IEEE Transactions on Information Theory* 59.10 (Oct. 2013), pp. 6755–6773. DOI: 10.1109/tit.2013.2271772.

[4] Richard Blahut. "Computation of channel capacity and rate-distortion functions". In: *IEEE Transactions on Information Theory* 18.4 (July 1972), pp. 460–473. DOI: 10.1109/tit.1972.1054855.

[5] Suguru Arimoto. "An algorithm for computing the capacity of arbitrary discrete memoryless channels". In: *IEEE Transactions on Information Theory* 18.1 (Jan. 1972), pp. 14–20. DOI: 10.1109/tit.1972.1054753.

[6] Kerry He, James Saunderson, and Hamza Fawzi. *A Bregman proximal perspective on classical and quantum Blahut-Arimoto algorithms*. 2023. arXiv: 2306.04492.

[7] Arkadij Semenovič Nemirovskij and David Borisovich Yudin. *Problem complexity and method efficiency in optimization*. Wiley, 1983. ISBN: 0471103454.

[8] Amir Beck and Marc Teboulle. "Mirror descent and nonlinear projected subgradient methods for convex optimization". In: *Operations Research Letters* 31.3 (May 2003), pp. 167–175. DOI: 10.1016/s0167-6377(02)00231-6.

[9] Paul Tseng. *On accelerated proximal gradient methods for convex-concave optimization*. Tech. rep. 2008. URL: https://pages.cs.wisc.edu/~brecht/cs726docs/Tseng.APG.pdf.

[10] Amir Beck. *First-order methods in optimization*. SIAM, 2017. DOI: 10.1137/1.9781611974997.

[11] Heinz H Bauschke, Jérôme Bolte, and Marc Teboulle. "A descent lemma beyond Lipschitz gradient continuity: First-order methods revisited and applications". In: *Mathematics of Operations Research* 42.2 (May 2017), pp. 330–348. DOI: 10.1287/moor.2016.0817.

[12] Haihao Lu, Robert M Freund, and Yurii Nesterov. "Relatively smooth convex optimization by first-order methods, and applications". In: *SIAM Journal on Optimization* 28.1 (Jan. 2018), pp. 333–354. DOI: 10.1137/16M1099546.

[13] Marc Teboulle. "A simplified view of first order methods for optimization". In: *Mathematical Programming* 170.1 (May 2018), pp. 67–96. DOI: 10.1007/s10107-018-1284-2.

[14] Masahito Hayashi. "Bregman divergence based em algorithm and its application to classical and quantum rate distortion theory". In: *IEEE Transactions on Information Theory* 69.6 (June 2023), pp. 3460–3492. DOI: 10.1109/tit.2023.3239955.

[15] Masahito Hayashi. *Iterative minimization algorithm on mixture family*. 2023. arXiv: 2302.06905.

[16] Venkat Chandrasekaran and Parikshit Shah. "Relative entropy optimization and its applications". In: *Mathematical Programming* 161.1 (Jan. 2017), pp. 1–32. DOI: 10.1007/s10107-016-0998-2.

[17] Hamza Fawzi and Omar Fawzi. "Efficient optimization of the quantum relative entropy". In: *Journal of Physics A: Mathematical and Theoretical* 51.15 (Mar. 2018), p. 154003. DOI: 10.1088/1751-8121/aab285.

[18] Hamza Fawzi, James Saunderson, and Pablo A Parrilo. "Semidefinite approximations of the matrix logarithm". In: *Foundations of Computational Mathematics* 19.2 (Apr. 2019), pp. 259–296. DOI: 10.1007/s10208-018-9385-0.

[19] Chris Coey, Lea Kapelevich, and Juan Pablo Vielma. "Performance enhancements for a generic conic interior point algorithm". In: *Mathematical Programming Computation* 15.1 (Mar. 2023), pp. 53–101. DOI: 10.1007/s12532-022-00226-0.

[20] Mehdi Karimi and Levent Tunçel. "Primal–dual interior-point methods for domain-driven formulations". In: *Mathematics of Operations Research* 45.2 (Feb. 2020), pp. 591–621. DOI: 10.1287/moor.2019.1003.

[21] Mehdi Karimi and Levent Tuncel. *Efficient implementation of interior-point methods for quantum relative entropy*. 2023. arXiv: 2312.07438.

[22] Lei Yang and Kim-Chuan Toh. "Bregman proximal point algorithm revisited: A new inexact version and its inertial variant". In: *SIAM Journal on Optimization* 32.3 (July 2022), pp. 1523–1554. DOI: 10.1137/20M1360748.

[23] Nilanjana Datta et al. "Quantum-to-classical rate distortion coding". In: *Journal of Mathematical Physics* 54.4 (Apr. 2013). DOI: 10.1063/1.4798396.

[24] Howard Barnum. "Quantum rate-distortion coding". In: *Physical Review A* 62.4 (Sept. 2000), p. 042309. DOI: 10.1103/physreva.62.042309.

[25] Zahra Baghali Khanian and Andreas Winter. *A rate-distortion perspective on quantum state redistribution*. 2021. arXiv: 2112.11952.

[26] Zahra Baghali Khanian, Kohdai Kuroiwa, and Debbie Leung. "Rate-distortion theory for mixed states". In: *2023 IEEE International Symposium on Information Theory*. June 2023, pp. 749–754. DOI: 10.1109/isit54713.2023.10206960.

[27] Michael A. Nielsen and Isaac L. Chuang. *Quantum computation and quantum information: 10th anniversary edition*. Cambridge University Press, 2010. DOI: 10.1017/cbo9780511976667.

[28] Mark M. Wilde. *Quantum information theory*. 2nd ed. Cambridge University Press, 2017. DOI: 10.1017/9781316809976.

[29] John Watrous. *The theory of quantum information*. Cambridge University Press, 2018. DOI: 10.1017/9781316848142.

[30] R Tyrrell Rockafellar. *Convex analysis*. Vol. 18. Princeton University Press, 1970. DOI: 10.1007/bfb0110040.

[31] Lev M Bregman. "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming". In: *USSR Computational Mathematics and Mathematical Physics* 7.3 (1967), pp. 200–217. DOI: 10.1016/0041-5553(67)90040-7.

[32] Chris J Maddison et al. "Dual space preconditioning for gradient descent". In: *SIAM Journal on Optimization* 31.1 (Jan. 2021), pp. 991–1016. DOI: 10.1137/19M130858X.

[33] Dimitri Bertsekas. *Convex optimization theory*. Vol. 1. Athena Scientific, 2009. ISBN: 1-886529-31-0.

[34] Theodor Bröcker and Tammo Tom Dieck. *Representations of compact Lie groups*. Vol. 98. Springer Science & Business Media, 2013. DOI: 10.1007/978-3-662-12918-0.

[35] William Fulton and Joe Harris. *Representation theory: A first course*. Vol. 129. Springer Science & Business Media, 2013. DOI: 10.1007/978-1-4612-0979-9.

[36] Glen E Bredon. *Introduction to compact transformation groups*. Academic Press, 1972. DOI: 10.1016/s0079-8169(08)x6007-6.

[37] Persi Diaconis and Steven Evans. "Linear functionals of eigenvalues of random matrices". In: *Transactions of the American Mathematical Society* 353.7 (Mar. 2001), pp. 2615–2633. DOI: 10.1090/S0002-9947-01-02800-8.

[38] Masahito Hayashi and Yuxiang Yang. "Efficient algorithms for quantum information bottleneck". In: *Quantum* 7 (Mar. 2023), p. 936. DOI: 10.22331/q-2023-03-02-936.

[39] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge University Press, 2004. DOI: 10.1017/cbo9780511804441.

[40] Roger A. Horn and Charles R. Johnson. *Topics in matrix analysis*. Cambridge University Press, 1991. DOI: 10.1017/cbo9780511840371.

[41] Mikhail V Solodov and Benar Fux Svaiter. "Error bounds for proximal point subproblems and associated inexact proximal point algorithms". In: *Mathematical Programming* 88.2 (Aug. 2000), pp. 371–389. DOI: 10.1007/s101070050022.

[42] Mark Schmidt, Nicolas Roux, and Francis Bach. "Convergence rates of inexact proximal-gradient methods for convex optimization". In: *Proceedings of the 24th International Conference on Neural Information Processing Systems*. Vol. 24. Dec. 2011, pp. 1458–1466. URL: https://dl.acm.org/doi/10.5555/2986459.2986622.

[43] Jorge Nocedal and Stephen J Wright. *Numerical optimization*. Springer, 1999. DOI: 10.1007/b98874.

[44] Nathaniel Johnston. *QETLAB: A MATLAB toolbox for quantum entanglement, version 0.9*. http://qetlab.com. Jan. 2016. DOI: 10.5281/zenodo.44637.

[45] Kim-Chuan Toh, Michael J Todd, and Reha H Tütüncü. "SDPT3 — A MATLAB software package for semidefinite programming, version 1.3". In: *Optimization Methods and Software* 11.1-4 (1999), pp. 545–581. DOI: 10.1080/10556789908805762.

[46] Masahito Hayashi and Geng Liu. *Generalized quantum Arimoto-Blahut algorithm and its application to quantum information bottleneck*. 2023. arXiv: 2311.11188.

[47] Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. John Wiley & Sons, 1999. DOI: 10.1002/047174882X.

[48] Aram V Arutyunov and Valeri Obukhovskii. *Convex and set-valued analysis*. De Gruyter, 2017. DOI: 10.1515/9783110460308.

[49] Martin Jaggi. "Revisiting Frank-Wolfe: Projection-free sparse convex optimization". In: *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*. JMLR.org, 2013, pp. 427–435. URL: https://dl.acm.org/doi/10.5555/3042817.3042867.

[50] Haobo Li and Ning Cai. "A Blahut-Arimoto type algorithm for computing classical-quantum channel capacity". In: *2019 IEEE International Symposium on Information Theory*. July 2019, pp. 255–259. DOI: 10.1109/isit.2019.8849608.

[51] Navneeth Ramakrishnan et al. "Computing quantum channel capacities". In: *IEEE Transactions on Information Theory* 67.2 (2020), pp. 946–960. DOI: 10.1109/tit.2020.3034471.

[52] Heinz H Bauschke, Jonathan M Borwein, et al. "Legendre functions and the method of random Bregman projections". In: *Journal of Convex Analysis* 4 (1997), pp. 27–67.

[53] Rajendra Bhatia. *Matrix analysis*. Vol. 169. Springer Science & Business Media, 2013. DOI: 10.1007/978-1-4612-0653-8.