# Quantum Rejection Sampling

by

Ala Shayeghi

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Combinatorics and Optimization (Quantum Information)

Waterloo, Ontario, Canada, 2015

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

## Abstract

Let $\mathcal{H}$ be a finite dimensional Hilbert space and $\rho, \sigma \in \mathsf{D}(\mathcal{H})$ be quantum states in $\mathcal{H}$ such that $\mathrm{S}(\rho||\sigma)$ is finite. In this thesis we consider the following communication task involving two parties Alice and Bob. Suppose that Alice is given a classical description of the states $\rho$ and $\sigma$. Given an unlimited number of copies of an entangled state whose marginal on Bob's side is $\sigma$, Alice's goal is to help Bob output a single copy of the state $\rho$ by sending a single message to Bob in a one-way LOCC (short for local operation and classical communication) protocol. We propose a class of one-way LOCC protocols for this task which we refer to as quantum rejection sampling protocols. Inspired by the classical rejection sampling protocol of Harsha, Jain, McAllester, and Radhakrishnan [25] for a similar classical communication task, we introduce the Greedy Quantum Rejection Sampler. We characterize the expected communication cost of the protocol in terms of max-relative entropy of $\rho$ with respect to $\sigma$, in the case where the state $\rho$ is a pure state and prove that the Greedy Quantum Rejection Sampler is an optimal quantum rejection sampling protocol in this case. We describe an optimal quantum rejection sampling protocol in terms of a semidefinite program and we find general lower bounds and upper bounds on the expected communication cost of the optimal protocol. We propose an LOCC compression protocol based on the Greedy Quantum Rejection Sampler protocol, for lossless compression of an arbitrary pure state quantum source in the visible compression model and we show an upper bound on the average length of this encoding. The upper bound is always less than or equal to the Shannon entropy of the quantum source and the gap between the two quantities can be arbitrary large. Finally, we introduce a high-entanglement deterministic exact protocol for remote preparation of an arbitrary ensemble of quantum states. Our protocol is based on a quantum rejection sampling protocol which uses a prefix-free encoding for communication of the messages. We establish an upper bound on the expected communication cost of this protocol for the worst case choice of the target state in terms of the max-information in Bob's output register at the end of the protocol about Alice's classical input register. Furthermore, this protocol can be used as a non-oblivious universal protocol for exact remote preparation of an arbitrary $d$-dimensional state at an expected communication cost of at most $\lg(d) + \mathcal{O}(\lg(\lg(d)))$.

## Acknowledgements

I would like to express my gratitude to all those who gave me the possibility to complete this thesis. First and foremost, my thanks are due to my supervisor, Ashwin Nayak, for continuous support, lots of fruitful discussions, and his comments on the draft of the thesis. I am also so grateful to the readers of my thesis, Debbie Leung and Levent Tuncel, for their valuable comments. I also thank Jon Tyson for helpful email correspondence.

I am deeply grateful to my parents for their unfailing love and never-ending support. Without you, I would not be the person I am today.

Above all I would like to thank my wife for her love, endless support, and encouragements. Thank you for being my best friend Shima. I owe you everything.

## Dedication

This thesis is dedicated to my wife, Shima, and my parents.

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

## 1.1 Problem description

Let $P, Q : \mathcal{X} \longrightarrow \mathbb{R}$ be probability distributions over a set $\mathcal{X}$ such that their relative entropy, $\mathrm{S}(P||Q)$, defined in Equation 2.2, is finite. Consider the following two-party communication task.

**Task (i):** (Generating one distribution from another) Suppose that Alice is given the description of the probability distributions $P$ and $Q$. Given unlimited samples from the distribution $Q$ as their shared randomness, Alice's goal is to help Bob output an element of $\mathcal{X}$ distributed according to $P$, by sending a single message to Bob in a one-way classical communication protocol.

This communication task was studied by Harsha, Jain, McAllester, and Radhakrishnan [25]. They consider a special class of one-way communication protocols for this task called *rejection sampling* protocols. In a rejection sampling protocol, Alice and Bob share an unlimited sequence $\{x_j\}_{j=1}^{\infty}$ of samples from the distribution $Q$. Alice sends an index $J$ (which is a random index which depends on the shared sequence of samples) to Bob such that the $J$-th sample $x_J$ is distributed according to $P$. Then Bob simply outputs the $J$-th sample. They introduce a rejection sampling protocol for Task (i) which we refer to as the *Greedy Rejection Sampler*. For this protocol they show that the expected communication cost is at most $\mathrm{S}(P||Q) + 2\lg(\mathrm{S}(P||Q) + 1) + \mathcal{O}(1)$. They also prove that the *Greedy Rejection Sampler* is an optimal rejection sampling protocol in the sense that the expected communication cost of any rejection sampling protocol is at least $\mathrm{S}(P||Q)$.

In this thesis we consider a natural extension of Task (i) to its quantum version. Let $\mathcal{H}$

be a finite dimensional Hilbert space and $\rho, \sigma \in \mathsf{D}(\mathcal{H})$ be quantum states in $\mathcal{H}$ such that their relative entropy, $\mathrm{S}(\rho||\sigma)$, defined in Equation 2.2, is finite.

**Task (ii)**: (Generating one quantum state from another) Suppose that Alice is given a classical description of the states $\rho$ and $\sigma$. Given an unlimited number of copies of an entangled state whose marginal on Bob's side is $\sigma$, Alice's goal is to help Bob output a single copy of the state $\rho$ by sending a single message to Bob in a one-way LOCC protocol.

Similar to the classical case we introduce a class of one-way LOCC protocols for Task (ii), which we refer to as *quantum rejection sampling* protocols. We find lower bounds on the expected communication cost of a general quantum rejection sampling protocols. We also introduce a quantum analogue of the classical greedy rejection sampler and we try to find upper bounds on the expected communication cost of the *Greedy Quantum Rejection Sampler* to investigate whether as in the classical case the greedy protocol is an optimal quantum rejection sampling protocol.

## 1.2 Motivation

### 1.2.1 Lossless quantum data compression and quantum variable length codes

One of the fundamental tasks in classical information theory is the compression of information. Perhaps the most important contributions of Claude Shannon to the field of classical information theory are his Source Coding Theorems (lossy and lossless) and his Noisy Channel Coding Theorem [47]. In the most natural and basic model, a memoryless classical source outputs a sequence $X_1, X_2, X_3, \ldots$ of independent, identically distributed random variables taking values in a set of source symbols $\mathcal{X}$. In the lossy compression scenario, a compression scheme with average probability of error at most $\epsilon$ for transmitting a sequence of $n$ source symbols over a noiseless channel consists of an encoding function $e : \mathcal{X}^n \longrightarrow \{0,1\}^m$ and a decoding function $d : \{0,1\}^m \longrightarrow \mathcal{X}^n$ such that $\Pr\left(d\left(e\left(x_1 x_2 \ldots x_n\right)\right) = x_1 x_2 \ldots x_n\right)$ for a random sequence $x_1 x_2 \ldots x_n$ of $n$ source symbols is at least $1-\epsilon$. For any such compression scheme, the compression rate is defined as $R = \frac{m}{n}$. Shannon's Lossy Source Coding Theorem states that if $R > \mathrm{H}(P)$, where $P$ is the common probability distribution of the $X_i$s, then for every $\epsilon > 0$, for large enough $n \in \mathbb{N}$, there exists a compression scheme for a sequence of $n$ source symbols of rate $R$ with average error probability at most $\epsilon$ . The proof is based on the existence of a subset of $\mathcal{X}^n$ referred to as the *typical subset*. The typical subset is a set of relatively small size compared to $\mathcal{X}^n$

with the property that for large enough $n$ the probability that a random sequence of $n$ source symbols belongs to it, can be arbitrarily close to one. The idea is very simple: The strings in the typical subset are each mapped to a unique binary string while other strings are mapped arbitrarily. Shannon's Lossy Source Coding Theorem also establishes that the Shannon entropy of the source is the fundamental limit for the compression rate in this setting. That is, if one compresses at a rate above the Shannon entropy of the source, then it is possible to recover the compressed data with probability approaching 1 in the asymptotic limit, and otherwise, it is not possible to do so.

In lossless source coding, the code-word lengths are allowed to vary for different source symbols. In this setting, a source code $C$ is a mapping from $\mathcal{X}$, the set of source symbols, to $\{0,1\}^*$, the set of all finite-length binary strings. For every $x \in \mathcal{X}$, the length of the codeword $C(x)$ is denoted by $l_C(x)$, and the average length of the code $C$ is defined as $\mathbb{E}(l_C(X)) = \sum_{x\in\mathcal{X}} P(x) l_C(x)$. A source code is uniquely decodable if every finite sequence of source symbols is mapped to a unique binary string. A prefix-free code is a code in which no codeword is a prefix of a longer codeword. Shannon's Lossless Source Coding Theorem states that the average length of the optimal uniquely decodable code for an arbitrary source with distribution $P$ is between $\mathrm{H}(P)$ and $\mathrm{H}(P)+1$. The proof is based on the fact that the set of codeword lengths of a uniquely decodable code $C$ satisfies the Kraft inequality, $\sum_{x\in\mathcal{X}} 2^{-l_C(x)} \leq 1$. In 1972, Huffman invented an optimal prefix-free coding scheme which is referred to as the Huffman code [31]. It is possible to prove that no other code for the same alphabet has a lower expected length than the code constructed by Huffman's algorithm.

The beginning of quantum information theory can be traced back for instance to Holevo's paper in 1973 [28]. With Schumacher's Quantum Lossy Source Coding Theorem [44] one of the most fundamental results of classical information theory could be carried over to the quantum world. A quantum source can be modelled as an ensemble of quantum states $\{P(x), |\psi_x\rangle\}_{x\in\mathcal{X}}$ on a finite dimensional Hilbert space $\mathcal{H}$, i.e., the source outputs the state $|\psi_x\rangle$ with probability $P(x)$ where the states $\{|\psi_x\rangle\}_{x\in\mathcal{X}}$ do not necessarily form an orthonormal basis. Let $\rho = \sum_{x\in\mathcal{X}} P(x)|\psi_x\rangle\langle\psi_x|$. In the lossy compression scenario, a compression scheme with fidelity at least $1-\epsilon$ for transmitting a sequence of $n$ output states of the source over a noiseless quantum channel consists of an encoding channel $\mathcal{E} : \mathcal{H}^{\otimes n} \longrightarrow \mathcal{K}^{\otimes m}$ and a decoding channel $\mathcal{D} : \mathcal{K}^{\otimes m} \longrightarrow \mathcal{H}^{\otimes n}$ where $\mathcal{K} = \mathbb{C}^2$, such that the channel fidelity between the channel $\mathcal{DE}$ and the state $\rho^{\otimes n}$ is at least $1-\epsilon$. Note that channel fidelity is a measure of closeness between the decoded state and the original state which takes into account the fact that the original state may be entangled with another quantum register (the environment). The channel fidelity between the channel $\mathcal{DE}$ and the state $\rho^{\otimes n}$ is defined as $\mathrm{F}_{\mathrm{channel}}(\mathcal{DE}, \rho^{\otimes n}) = \inf \left[\mathrm{F}\left(\xi, (\mathcal{DE} \otimes \mathbb{1}_{\mathsf{L}(\mathcal{M})})(\xi)\right)\right]$, where the infimum is over all finite dimensional Hilbert spaces $\mathcal{M}$ and all $\xi \in \mathsf{D}(\mathcal{H}^{\otimes n} \otimes \mathcal{M})$ satisfying $\mathrm{Tr}_{\mathcal{M}}(\xi) = \rho^{\otimes n}$. Similar to the classical case, the compression rate is defined as $R = \frac{m}{n}$. Schumacher's Quantum Lossy Source Coding Theorem states that if $R > \mathrm{S}(\rho)$, then for every $\epsilon > 0$, there exists a compression scheme of rate $R$ for compressing a sequence of $n$ source output states with fidelity at least $1 - \epsilon$, for large enough $n \in \mathbb{N}$. Similar to the classical case the proof is based on the existence of the *typical subspace* which is a subspace of relatively small size compared to $\mathcal{H}^{\otimes n}$. The probability that a sequence of $n$ source outputs belongs to the typical subspace is arbitrarily close to 1 for large enough $n$. The basic steps of the encoding are performing a typical subspace measurement and an isometry that compresses the typical subspace. The decoder performs the inverse of the isometry and the protocol is successful if the typical subspace measurement successfully projects onto the typical subspace which happens for large enough $n$. Schumacher's Quantum Lossy Source Coding Theorem also states that for a compression rate below $\mathrm{S}(\rho)$, it is impossible to recover the compressed data perfectly in the asymptotic limit. Subsequent work on lossy quantum data compression can be found in Refs. [6, 7, 24, 26, 27, 35, 36, 41, 49].

Extending Shannon's Lossless Source Coding Theorem faces an obvious barrier: If we encode a source using a quantum code with variable lengths, in decoding we need to measure the length of the codewords, which disturbs the message. Hence, as many authors have pointed out, e.g., in Refs. [16, 18, 46], lossless compression of quantum data using quantum variable length source codes is impossible if only the quantum resource is available. In order to overcome this difficulty, different models have been proposed. For instance, Bostrom and Felbinger [16] propose a model for a quantum source with a classical helper, a classical channel used to inform the decoder about the codeword lengths. Later, Ahlswede and Cai [1] proposed a more general model in which the classical channel is used more effectively. Section 1.3.1 contains more details about basic concepts and results on quantum variable-length codes. A different approach to quantum lossless source coding would be classical encoding of a quantum source, i.e., using unlimited access to a classical channel and prior shared entanglement, on average how many bits of communication are required to losslessly encode a quantum source. The extension of the classical Greedy Rejection Sampler gives us an encoding/decoding procedure to encode a quantum source which fits into this model. In Section 3.5.1 we consider this procedure and show an upper bound on the average length of this source coding scheme.

### 1.2.2 Quantum One-shot Reverse Shannon Theorem

A discrete memoryless classical channel $\mathcal{E}$ is defined as a triple $\left(\mathcal{X}, \mathcal{Y}, P_{Y|X}\right)$, where $\mathcal{X}$ and $\mathcal{Y}$ are the set of possible inputs and outputs of the channel, respectively, and $P_{Y|X}$

is the probability transition matrix which specifies the probability of observing output $y$ given that $x$ is sent. Let $X$ and $Y$ denote the random variable corresponding to the channel input and output, respectively. The capacity of discrete memoryless channel $\left(\mathcal{X}, \mathcal{Y}, P_{y|x}\right)$ is defined as $C = \max\limits_{X \sim P} \mathrm{I}(X : Y)$, where the maximum is taken over all possible input distributions $P$. Shannon's Noisy Channel Coding Theorem [47] states that given a noisy discrete memoryless channel with capacity $C$, if $R < C$, then there exist coding schemes of rate $R$ which allow the probability of incorrect decoding at the receiver to be arbitrary small. But if $R > C$, then arbitrarily small probability of error is not achievable, i.e., every code with a rate above the channel capacity has a probability of error greater than a positive minimal error which increases as the rate $R$ increases. In other words, it is possible to transmit information nearly without error at any rate below a limit which is the channel capacity. This result is counter-intuitive in the sense that if the channel introduces errors, any correction process is also subject to error. Shannon also showed that the channel capacity does not increase if one allows the use of shared randomness between the sender and receiver.

Shannon's noisy channel coding theorem establishes the ability of noisy channels to simulate noiseless ones, and gives an operational interpretation to channel capacity as the asymptotic efficiency of this simulation. The reverse problem, of using a noiseless channel to simulate a noisy one, was studied by Bennett, Shor, Smolin, and Thapliyal [13] in the asymptotic case. Let $\mathcal{E}$ be a discrete memoryless classical channel given by $\left(\mathcal{X}, \mathcal{Y}, P_{Y|X}\right)$ of capacity $C$ and let $\epsilon$ be a positive constant. Let $\mathcal{E}^n$ be the extended channel consisting $n$ parallel applications of $\mathcal{E}$, mapping $\mathcal{X}^n$ to $\mathcal{Y}^n$. Bennett *et al.* introduce a one-way communication protocol $\Delta_n$ over a noiseless classical channel which given access to free shared randomness simulates the channel $\mathcal{E}^n$ perfectly, in the sense that for all $n$, the transition matrix corresponding to the protocol is identical to that for $\mathcal{E}^n$. Furthermore, the protocol $\Delta_n$ is asymptotically efficient in the sense that the probability that the protocol uses more than $n(C + \epsilon)$ bits of communication can be made arbitrary small for large enough $n$. So the Asymptotic Classical Reverse Shannon Theorem of Bennett *et al.* establishes that any noisy discrete memoryless classical channel can be simulated by a noiseless channel given access to free shared randomness, and proves that the channel capacity is the asymptotic efficiency of this simulation.

Harsha *et al.* [25], proved the One-shot Classical Reverse Shannon Theorem. Let $\mathcal{E}$ be a noisy discrete memoryless classical channel defined by the triple $\left(\mathcal{X}, \mathcal{Y}, P_{Y|X}\right)$ of capacity $C$. Harsha *et al.* introduced a one-way communication protocol over a noiseless classical channel which given access to free shared randomness simulates $\mathcal{E}$, with an expected worst-case communication of at most $C + 2\lg(C + 1) + \mathcal{O}(1)$ bits. They also show that any such protocol simulating the channel $\mathcal{E}$ has an expected worst-case communication of at least $C$

bits. Their protocol is based on the classical Greedy Rejection Sampler for Task (i) described in Section 1.1. More specifically, let $X$ be the random variable corresponding to the channel input taking values in $\mathcal{X}$. Given $x \in \mathcal{X}$ as the input, the encoder, Alice, needs to help Bob, the decoder, output some $y \in \mathcal{Y}$ according to the distribution $Y|(X = x)$, imposed by the transition matrix $P_{Y|X}$. They show that the Greedy Rejection Sampler can be used to achieve this for every $x \in \mathcal{X}$ with communication of at most $C + 2\lg(C+1) + \mathcal{O}(1)$ bits. This motivates the generalization of the Greedy Rejection Sampler protocol to the quantum case as described in Task (ii) in Section 1.1, in order to investigate whether the same idea can be used to prove a Quantum One-shot Reverse Shannon Theorem.

### 1.2.3 Exact remote state preparation

Quantum *Teleportation* [8] is a protocol which enables a single qubit quantum state to be transmitted from a sender (Alice) to a receiver (Bob) by sharing one maximally entangled state $(|00\rangle + |11\rangle)/\sqrt{2}$ (1 ebit) and communicating 2 classical bits ( 2 cbits). More generally, $\lg(d)$ ebits and $2\lg(d)$ cbits are sufficient for the teleportation of a $d$-dimensional quantum state. It has been shown that these resources are also necessary and neither resource can be traded off against the other. What is interesting about teleportation is that neither Alice nor Bob acquires any classical knowledge on the teleported state. In this sense, the teleportation protocol is said to be oblivious to both Alice and Bob. In 2000, Lo [40] introduced the problem of remote state preparation (RSP) which is a variant of quantum teleportation in which the sender knows the quantum state to be communicated. More precisely, the RSP task involves two parties Alice and Bob. Let $\mathcal{H}$ be a finite dimensional Hilbert space. Alice is given a classical description of a set of quantum states $\{\rho_x\}_{x\in\mathcal{X}} \subseteq \mathsf{D}(\mathcal{H})$. She receives a classical input $x$ according to a probability distribution $P : \mathcal{X} \longrightarrow \mathbb{R}$. Her goal is to help Bob output a single copy of the state $\rho_x$ by engaging in an LOCC protocol. The RSP task may be defined with or without allowing back-communication from Bob to Alice. We call an RSP protocol *universal* if it is capable of remote preparation of an arbitrary finite dimensional state. An RSP protocol is *deterministic* if it always succeeds, and it is *probabilistic*, if it fails with a constant probability and the protocol additionally produces a flag indicating "success" or "failure", accessible to both sender and receiver. An RSP protocol for the ensemble $\{P(x), \rho_x\}_{x\in\mathcal{X}}$ is called an *exact* RSP (ERSP) protocol if Bob's output state, $\tilde{\rho}$, exactly coincides with $\rho_x$, i.e, their fidelity is 1 and it is an *approximate* protocol if the fidelity is greater than $1 - \epsilon$, for a small constant $\epsilon > 0$. Teleportation can be considered as an RSP protocol: Alice can prepare the state $\rho_x$ herself and teleport it to Bob. But complete knowledge of the state $\rho_x$, opens many other possibilities for remote preparation of states. For instance, Alice can send all her classical knowledge of the state $\rho_x$ to Bob, and then Bob constructs the state himself. This simple RSP protocol does not use shared entanglement but requires infinitely many bits of classical communication to encode the state. This also shows the possibility of resource trade off in the RSP task in contrast to quantum teleportation. The motivation for studying the RSP problem is whether the required classical and quantum resources can be reduced given Alice's knowledge on the state to be prepared. Lo [40] and Pati [43] showed that for specific ensembles of pure states (e.g. states on the same latitude on the Bloch sphere), RSP requires less classical communication compared to teleportation, but Lo conjectured that deterministic exact remote preparation of a general pure state of dimension $d$ has the same communication cost as teleportation, i.e., $2\lg(d)$ cbits. Although no counterexamples have been found, Lo's conjecture has only been proved for some restricted cases.

The RSP protocols can be divided into two categories based on the amount of entanglement they require: *Low-entanglement* protocols which require less than $\lg(d)$ ebit to prepare a $d$-dimensional quantum state, and *high-entanglement* protocols which require more than $\lg(d)$ ebits. In 2001, Devetak and Berger [23] and Bennett *et al.* [11] proposed low-entanglement RSP schemes based on teleportation. They introduce two protocols for remote preparation of a sequence of $n$ states, $\rho_{x_1}, \ldots, \rho_{x_n}$, from an ensemble $\{P(x), \rho_x\}_{x\in\mathcal{X}}$ which use similar ideas. Their protocols involve first sending a classical message, giving Bob some information about the target state, hence reducing Bob's uncertainty about it. Then the remaining uncertainty (information) about the target state is compressed into a smaller number of qubits, using Schumacher's source coding scheme. Finally, the encoded states are teleported to Bob and he decompresses the teleported state. Such a protocol requires a smaller amount of entanglement to teleport the Schumacher encoded states at the cost of communicating more cbits. The more classical information is sent about the target states, the less entanglement is needed. Since Schumacher's compression has arbitrarily small error in the limit of large $n$, their protocol is asymptotically exact. The main difference between the protocols given by Bennett *et al.* and Devetak and Berger is the classical information about the target state which is sent by Alice to Bob in the two protocols.

In the same paper [11], Bennett *et al.* also show that at the cost of spending more shared entanglement the communication cost for remote preparation of an arbitrary $d$-dimensional pure state can be reduced to $\lg(d)$ cbits, asymptotically. Their high-entanglement RSP scheme which is referred to as the Column method can be described as follows: Alice and Bob share sufficiently many (say $K$) copies of the maximally entangled state $|\phi\rangle = \frac{1}{\sqrt{d}}\sum_{i=1}^{d}|i\rangle|i\rangle$. Given an arbitrary $d$-dimensional pure state $\psi = |\psi\rangle\langle\psi|$, Alice performs the measurement $\left\{P_0 = \bar{\psi}, P_1 = \mathbb{1} - \bar{\psi}\right\}$ on her marginal state of each of the $K$ copies.

Here $\bar{\psi}$ denotes the complex conjugation of the state $\psi$ with respect to the basis $\{|i\rangle\}_{i=1}^{d}$. In each measurement the outcome is 0 with probability $\frac{1}{d}$, in which case Bob's marginal state is $|\psi\rangle$. If $\lg(K) \geq \lg(d) + \lg(\lg(\epsilon))$ for some $\epsilon > 0$, the probability that the outcome of all $K$ measurements is 1 ("failure") is at most $\epsilon$. If this does not happen, there is at least one measurement with outcome 0, and Alice sends the index of the state using $\lg(K)$ cbits. The required resources for remote preparation of a $d$-dimensional pure state using this protocol are $K\lg(d)$ ebits and $\lg(d)$ cbits, asymptotically.

In Ref. [12], Bennett *et al.* introduce a universal probabilistic exact protocol for remote preparation of arbitrary $d$-dimensional quantum states at an asymptotic cost of $\lg(d)$ ebits and $\lg(d)$ cbits and they use their protocol to obtain an exact trade off curve for the resources required for remote preparation of a quantum state. Their protocol is based on the existence of a sufficiently big set of unitary operators (referred to as randomizing unitaries) on a finite dimensional Hilbert space.

One should note that the above two probabilistic exact RSP protocols can be made deterministic exact by simply using teleportation to send the target state in the case of "failure". Although, these protocols may seem to contradict Lo's conjecture, one should note that they require a probabilistic amount of resources, and in the worst case (in case of failure), they use extra communication to teleport the target state, resulting in communication of more than $2\lg(d)$ cbits. In the same paper [12], Bennett *et al.* also prove that any universal exact RSP protocol requires communication of at least $\lg(d)$ cbits for remote preparation of a $d$-dimensional state.

As mentioned earlier, Lo's conjecture has been proved to be true only for some restricted RSP protocols, such as when besides being deterministic exact and universal, the protocol has deterministic classical communication and entanglement costs. In 2001, Bennett *et al.* [9] proved Lo's conjecture for RSP protocols restricted by the following constraints:

i) Bob is only allowed to perform local unitary operations on his system determined by the message he receives from Alice.

ii) The probability that Alice sends a particular message to Bob does not depend on the target state.

More generally an RSP protocol is called *oblivious* to Bob if:

i) The quantum systems at Bob's side including by-products of the protocol are independent of the target state, so that Bob obtains no more information about the target state than what can be inferred from the single instance he outputs, even if he deviates from the protocol.

ii) The probability that Alice sends a particular message to Bob does not depend on the target state.

Leung and Shor [37] prove Lo's conjecture to be true for any universal deterministic exact RSP protocol for remote preparation of pure states which is oblivious to Bob and uses only forward communication from Alice to Bob.

In the most general model for an RSP protocol without back communication, Alice performs a general quantum operation, which is parameterized by the target state $\rho_x$, on her part of a shared state between the two parties. The output of her operation is a classical message $m$ from a set of messages together with possibly some other classical and quantum outputs. In such a protocol, the probabilities of sending different messages are not generally the same. Hence, by using a variable encoding scheme for compression of the messages we may be able to decrease the expected communication cost. In all of the protocols described above, fixed length coding (block coding) is used for sending the classical messages from Alice to Bob. In this thesis we propose a quantum rejection sampling protocol for remote preparation of an arbitrary ensemble of states $\{P(x), \rho_x\}_{x\in\mathcal{X}}$ on a $d$-dimensional Hilbert space $\mathcal{H}$. Our protocol is a high-entaglement scheme which is deterministic exact and uses a variable length encoding of the set of all messages. Let $\xi_{AB}$ denote the joint quantum state of Alice's classical input register and Bob's output register after the protocol terminates, then $\xi_{AB} = \sum_{x\in\mathcal{X}} P(x)|x\rangle\langle x|^A \otimes \rho_x^B$. The max-information the register $B$ has about the register $A$, is denoted by $\mathrm{I}_{\max}(A:B)_{\xi_{AB}}$. We show that the expected communication cost of our protocol for the worst-case choice of the input $x$ is bounded by

$$\mathrm{I}_{\max}(A:B)_{\xi_{AB}} + 2\lg\left(\mathrm{I}_{\max}(A:B)_{\xi_{AB}}\right) + \mathcal{O}(1) \ ,$$

which is always less than or equal to $\lg(d) + 2\lg(\lg(d)) + \mathcal{O}(1)$. Furthermore, our quantum rejection sampling protocol can be used for remote preparation of a generic quantum state at an expected communication cost of at most $\lg(d) + 2\lg(\lg(d)) + \mathcal{O}(1)$ cbits.

### 1.2.4 Characterization of mutual information

Consider the following communication task between two parties, Alice and Bob. Let $\mathcal{X}$ and $\mathcal{Y}$ be finite, non-empty sets, and let $(X, Y)$ be a pair of (correlated) random variables taking values in $\mathcal{X}\times\mathcal{Y}$. Suppose that Alice is given $x \in \mathcal{X}$ according to the distribution $X$. Alice's goal is to send a message to Bob to help Bob output a value $y \in \mathcal{Y}$ distributed according to $Y|(X = x)$ (i.e., they want the output pair $(x, y)$ to be distributed according to $(X, Y)$). Suppose that Alice and Bob have access to shared randomness. Let $T(X : Y)$ denote the minimum, over all protocols, of the expected number of bits sent from Alice to Bob to achieve this. Harsha *et al.* [25] give an operational interpretation of mutual information by showing that $\mathrm{I}(X:Y) \leq T(X:Y) \leq \mathrm{I}(X:Y)+2\lg(\mathrm{I}(X:Y)+1)+\mathcal{O}(1)$. The lower bound is obtained by a straightforward information-theoretic argument, and they prove the upper bound by introducing a protocol based on the Greedy Rejection Sampler for the described task with expected communication cost of at most $\mathrm{I}(X:Y)+2\lg(\mathrm{I}(X:Y)+1)+\mathcal{O}(1)$. A quantum version of the Greedy Rejection Sampler may give us a similar result in the quantum setting.

### 1.2.5 A direct sum result in quantum communication complexity

Let $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ be finite, non-empty sets and let $f : \mathcal{X} \times \mathcal{Y} \longrightarrow \mathcal{Z}$ be a function. A two-party communication protocol for computing $f$ consists of two parties Alice and Bob, who receive inputs $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, respectively, and communicate with each other to compute $f(x,y) \in \mathcal{Z}$. A *k-round* protocol is a protocol in which the two parties exchange at most $k$ messages. Let $\mu$ be a probability distribution on $\mathcal{X} \times \mathcal{Y}$ and $\epsilon > 0$ be a real number. The $\epsilon$-error, $k$-round distributional communication complexity of $f$ under $\mu$, $\mathrm{D}_{\epsilon}^{\mu,k}(f)$, is defined as the number of bits communicated for the worst-case input by the best deterministic $k$-round protocol for $f$ with average error at most $\epsilon$ under $\mu$. For $n \in \mathbb{N}$, let $f^n : \mathcal{X}^n \times \mathcal{Y}^n \longrightarrow \mathcal{Z}^n$ be defined as $f^n\left((x_1,\ldots,x_n),(y_1,\ldots,y_n)\right) = (f(x_1,y_1),\ldots,f(x_n,y_n))$. A very basic question in communication complexity is whether the communication complexity of $f^n$ is at least $n$ times that of $f$. This question is referred to as the *direct sum* question. Harsha *et al.* use the operational interpretation they obtained for mutual information to prove a direct sum result stronger than the previously known results in the bounded round classical communication complexity setting. More precisely, they prove that for any function $f : \mathcal{X} \times \mathcal{Y} \longrightarrow \mathcal{Z}$, and any product distribution $\mu$ over $\mathcal{X} \times \mathcal{Y}$,

$$\forall \delta > 0, \qquad D_{\epsilon}^{\mu^n,k}(f^n) \quad \geq \quad \frac{n}{2}\left(\delta D_{\epsilon+\delta}^{\mu,k}(f) - \mathcal{O}(k)\right) \ .$$

An extension of their result to the quantum setting may lead to a similar result in the bounded round quantum communication complexity setting.

## 1.3 Related work

### 1.3.1 Quantum variable-length codes

There are several different definitions of quantum variable length codes. Here we use the definition given by Bostrom and Felbinger [16]. Let $\mathcal{H}$ be a Hilbert space of dimension $d$.

Let $\left\{\mathcal{H}^{\otimes l} : l = 1, 2, \ldots, l_{\max}\right\}$ be a set of pairwise orthogonal subspaces of a sufficiently large Hilbert space. Then the direct sum $\mathcal{H}^{\oplus l_{\max}} = \mathcal{H} \oplus \mathcal{H}^{\otimes 2} \oplus \cdots \oplus \mathcal{H}^{\otimes l_{\max}}$, is a Hilbert space of dimension $\displaystyle\sum_{l=1}^{l_{\max}} d^l$. Let $\mathcal{K}$, a Hilbert space of dimension $d'$, be the source space. Then a variable-length encoder of maximum length $l_{\max}$ is defined as an isometry $\mathcal{E} : \mathcal{K} \longrightarrow \mathcal{C}$ from the source space to the code space $\mathcal{C} \subset \mathcal{H}^{\oplus l_{\max}}$ of dimension $d'$. To realize coding schemes, Schumacher and Westmoreland [46] define zero-extended forms. The zero-extended form of $|\psi_l\rangle \in \mathcal{H}^{\otimes l}$ is obtained by appending an ancilla to it as $|\psi_l 0^{l_{\max}-l}\rangle$. The zero-extended form of a superposition of states $\left\{|\psi_{l_i}\rangle \in \mathcal{H}^{\otimes l_i} : \; i \in I \subseteq \{1, \ldots, l_{\max}\}\right\}$ is defined as the superposition of the zero-extended form of the states $|\psi_{l_i}\rangle$. In contrast to the classical case where the codeword lengths are determinate, the length of the codewords in a quantum variable-length code are indeterminate because of superposition. One way to define the length of the codewords is as follows. For every $l \in \{1, \ldots, l_{\max}\}$, let $P_l$ be the orthogonal projector of $\mathcal{H}^{\oplus l_{max}}$ onto $\mathcal{H}^{\otimes l}$ , then the average length of a codeword $|\phi\rangle \in \mathcal{C}$ is defined as $\bar{L}(|\phi\rangle) = \langle\phi|\Lambda|\phi\rangle$, where $\Lambda = \displaystyle\sum_{l=1}^{l_{\max}} l \; P_l$. Schumacher and Westmoreland [46] prove a quantum version of Kraft inequality in this model which states that for any quantum uniquely decodable code with code space $\mathcal{C}$, and $d = \dim(\mathcal{H})$,

$$\sum_{l=1}^{l_{max}} \dim(\mathcal{C} \cap \mathcal{H}^{\otimes l}) d^{-l} \leq 1 \;\; .$$

Similar to the classical case, the quantum Kraft inequality can be used to obtain a lower bound on the expected average length of a uniquely decodable quantum code. Consider a quantum source given by the ensemble $\{P(x), |\psi_x\rangle\}_{x \in \mathcal{X}}$, and let $\rho = \sum_{x \in \mathcal{X}} P(x) |\psi_x\rangle\langle\psi_x|$. Schumacher and Westmoreland [46] prove that for any uniquely decodable quantum variable-length code, the expected length of the code is lower bounded by the von Neumann entropy, $\mathrm{S}(\rho)$, of $\rho$.

The base length of a codeword $|\phi\rangle$ is defined as $L(|\phi\rangle) = \max\{l : \langle\phi|P_l|\phi\rangle > 0\}$[16]. The base length of a codeword is an important parameter since in order to store a codeword of base length $l$, a quantum register of length at least $l$ is required. Hence it is necessary for the decoder to know the base length of the codewords. Furthermore, as stated in Ref. [1], in general, there is no way to measure the base length of unknown codewords without error. So in a lossless encoding of a quantum source in this model we need a way of informing the decoder about the base length of the codewords. To do so, the encoder needs to know the output of the quantum source. This situation is referred to as *visible* encoding, as opposed to *blind* encoding in which the encoder does not know which state is output by the source. Based on the assumption of visible encoding and using a classical channel only for sending the base length of the codewords, Bostrom and Felbinger [16] propose a scheme for compression of a pure state source. In 2004, Ahlswede and Cai [1] proposed a more general model, quantum-classical compression model, in which the classical channel is used more efficiently. They introduce a more general protocol and find a minimum achievable value for the expected average length of a quantum variable length code in the new model. They also show that the von Neumann entropy bound is not a tight bound, in fact they show that there exist quantum sources such that the gap between their lower bound and the von Neumann entropy bound is very large. For more detailed information on the results in this area please refer to the survey by Ahlswede and Cai [2].

### 1.3.2 Asymptotic Quantum Reverse Shannon Theorem

Unlike classical channels, which are adequately characterized by a single capacity, various distinct capacities can be defined for a quantum channel. These include classical capacity for transmitting classical information, quantum capacity for transmitting quantum states, classically-assisted quantum capacity for transmitting quantum states in the presence of a two-way classical side-channel, and entanglement-assisted classical capacity for transmitting classical information with the help of free prior entanglement [13, 29, 22, 45, 39]. In Ref. [13] Bennett, Shor, Smolin, and Thapliyal show that the entanglement-assisted classical capacity $C_{\mathrm{E}}$ of a quantum channel $\mathcal{E}$ is given by the quantum mutual information

$$C_{\mathrm{E}}(\mathcal{E}) \quad = \quad \max_{\rho \in \mathsf{D}(\mathcal{H}_{\mathrm{in}})} \left\{ \mathrm{S}\left(\rho\right) + \mathrm{S}\left(\mathcal{E}(\rho)\right) - \mathrm{S}((\mathcal{E} \otimes \mathbb{1}_{\mathsf{L}(\mathcal{H}_{\mathrm{ref}})}(\psi_\rho))) \right\} \quad ,$$

where $\psi_\rho$ is a pure state over the tensor product of the input Hilbert space $\mathcal{H}_{\mathrm{in}}$ and a reference system $\mathcal{H}_{\mathrm{ref}}$, whose reduced density matrix on the channel's input space is $\rho$, i.e., $\mathrm{Tr}_{\mathcal{H}_{\mathrm{ref}}}(\psi_\rho) = \rho$. The Entanglement-assisted Classical Capacity Theorem is proved by first showing that the quantum mutual information of the channel is an asymptotically achievable rate, by combining superdense coding and Schumacher's compression scheme, and then proving that this rate is optimal by making use of the strong subadditivity of von Neumann entropy and Holevo's formula. In the same paper Bennett *et al.* conjectured an Asymptotic Quantum Reverse Shannon Theorem. Subsequently, Bennett, Devetak, Harrow, Shor and Winter proved the theorem [10], which states that any quantum channel $\mathcal{E}$ can be asymptotically simulated by an unlimited amount of shared entanglement and a rate of classical communication equal to the channel's entanglement assisted classical capacity, $C_{\mathrm{E}}(\mathcal{E})$. In 2011, Berta, Christandl, and Renner gave an alternative proof of the Asymptotic Quantum Reverse Shannon Theorem based on one-shot information theory [15]. Their proof uses a one-shot version of *quantum state merging* [14] to obtain a stronger version of *quantum state splitting.* They use the *post-selection technique* [20] to show that their protocol for quantum state splitting is sufficient to asymptotically simulate the channel $\mathcal{E}$ with a classical communication rate of $C_{\mathrm{E}}(\mathcal{E})$.

### 1.3.3 A closely related problem

Let $P, Q : \mathcal{X} \longrightarrow \mathbb{R}$ be probability distributions over a set $\mathcal{X}$ such that $\mathrm{S}(P||Q)$ is finite. Consider the following two-party communication task which is closely related to Task (i).

**Task (i′)** : Suppose that Alice is given the description of the probability distribution $P$ and Bob is given the description of the probability distribution $Q$. Given access to free shared randomness, Alice's goal is to help Bob output an element of $\mathcal{X}$ according to the distribution $P$, by engaging in a classical communication protocol.

In other words, the goal is to efficiently sample from a distribution $P$ that is only known to Alice, by taking advantage of a distribution $Q$ which is known only to Bob. This communication task was studied by Braverman and Rao [19]. They introduce a two-way communication protocol for a weaker version of Task (i′), in which Bob outputs an element of $\mathcal{X}$ distributed according to a distribution $P'$ which is close to $P$. More precisely, at the end of the protocol, Alice outputs $a \in \mathcal{X}$ according to $P$, and Bob outputs $b \in \mathcal{X}$ such that given that Alice's output is $a = x$, the probability that Bob's output is the same, i.e. $b = x$, is at least $1 - \epsilon$, for some error parameter $\epsilon > 0$. They prove that the expected communication cost of their protocol is given by $\mathrm{S}(P||Q) + \mathcal{O}(\sqrt{\mathrm{S}(P||Q)}) + \lg\left(\frac{1}{\epsilon^2}\right)$. The proof is based on a very clever idea. Alice and Bob interpret their shared random string as a uniformly chosen random sequence $\{(x_i, p_i)\}_{i=1}^{\infty}$ of elements of $\mathcal{X} \times [0,1]$. Let $\mathcal{P} = \{(x,p) \in \mathcal{X} \times [0,1] \ : \ p < P(x)\}$ and for every constant $c \geq 1$, let $c\mathcal{Q} = \{(x,p) \in \mathcal{X} \times [0,1] \ : \ p < cQ(x)\}$. Alice selects the first index $i$ such that $(x_i, p_i) \in \mathcal{P}$, and sends the binary encoding of $k = \lceil \frac{i}{|\mathcal{X}|} \rceil$ to Bob and outputs $a = x_i$. After receiving Alice's message, Bob knows that the index $i$ belongs to the set $\{(k-1)\,|\mathcal{X}| + 1, \ldots, k\,|\mathcal{X}|\}$. They use a different part of their shared random string to obtain a sequence of proper random hash functions. Starting from $j = 0$, in the $j$-th iteration, for properly chosen values $c_j$ and $s_j$ which increase with $j$, Alice sends the values of the first $S_j$ hash functions at $x_i$, which she has not already sent to Bob. He compares these values with the value of the hash functions on $\{x_r \ : \ r \in \{(k-1)\,|\mathcal{X}| + 1, \ldots, k\,|\mathcal{X}|\} \wedge (x_r, p_r) \in C_j\mathcal{Q}\}$. If he finds an index for which all of the hash values are consistent with Alice's hash values, he responds "success" and outputs $b = x_r$ for the smallest such $r$, otherwise he responds "failure", and they go to the $(j+1)$-th iteration. They continue until Bob declares success.

A similar task can be defined in the quantum setting for quantum states $\rho, \sigma \in \mathsf{D}(\mathcal{H})$ in a finite dimensional Hilbert space $\mathcal{H}$ such that $\mathrm{S}(\rho||\sigma)$ is finite.

**Task (ii′)** : Suppose that Alice is given a classical description of the states $\rho$ and Bob is given a classical description of the state $\sigma$. Given access to free shared entanglement, Alice's goal is to help Bob output a single copy of the state $\rho$ by engaging in an LOCC protocol.

The protocol of Braverman and Rao and its applications motivated the study of Task (ii′) to investigate whether it is possible to use a similar idea in the quantum setting. Recently, along with Anshu, Jain, Mukhopadhyay, and Yao[4], we designed a one-way LOCC protocol inspired by the protocol of Braverman and Rao, with a communication cost of at most $\mathcal{O}\left((\mathrm{S}(\rho||\sigma)+1)/\epsilon^4\right)$ bits, in which Bob's output is a quantum state $\rho' \in \mathsf{D}(\mathcal{H})$ such that the fidelity between $\rho$ and $\rho'$ is at least $1-\epsilon$, for some error parameter $\epsilon > 0$. The protocol assumes that the value of $\mathrm{S}(\rho||\sigma)$ is known to both Alice and Bob. Here, for simplicity, we explain a simpler case in which we assume that both parties know the value $c = \mathrm{S}_{\max}(\rho||\sigma)$. Alice and Bob share a sequence $\{|\zeta_j\rangle\}_{j=1}^{\infty}$ of the following state,

$$|\zeta\rangle = \frac{1}{\sqrt{NK}} \sum_{i=1}^{N} |i,i\rangle_{AB} \otimes \sum_{m=1}^{K} |m,m\rangle_{A'B'} \ ,$$

where $\{|1\rangle, \ldots, |N\rangle\}$ is an orthonormal basis for $\mathcal{H}$. Note that similar to the shared random string in the classical protocol the registers $A$ and $B$ serve to sample a maximally mixed state in $\mathcal{H}$ and the registers $A'$ and $B'$ serve to sample uniformly in the interval $[0,1]$, in the limit when $K \longrightarrow \infty$. Let $\rho = \sum_{i=1}^{N} \lambda_i |\psi_i\rangle\langle\psi_i|$ and $\sigma = \sum_{i=1}^{N} \gamma_i |\phi_i\rangle\langle\phi_i|$ be the eigenvalue decomposition of $\rho$ and $\sigma$. For each copy of $|\zeta\rangle$, Alice performs the measurement $\{P_A, \mathbb{1} - P_A\}$ on the registers $AA'$ where

$$P_A = \sum_{i=1}^{N} |\psi_i\rangle\langle\psi_i| \otimes \sum_{m=1}^{K\lambda_i} |m\rangle\langle m| \ .$$

She accepts the index of the first copy in which the output of her measurement successfully corresponds to $P_A$. Similarly, for each copy of $|\zeta\rangle$, Bob performs the measurement $\{P_B, \mathbb{1} - P_B\}$ (for appropriately chosen $\delta$) on the registers $BB'$ where

$$P_B = \sum_{i=1}^{N} |\phi_i\rangle\langle\phi_i| \otimes \sum_{m=1}^{K \min\left(\frac{2^c \gamma_i}{\delta}, 1\right)} |m\rangle\langle m| \ .$$

Bob accepts the index of the first copy in which the output of his measurement successfully corresponds to $P_B$. In the above expressions, $\lambda_i$ and $2^c\gamma_i$ are assumed to be rounded to nearest multiple of $1/K$. The error introduced due to this assumption approaches zero as $K \longrightarrow \infty$. We show that the marginal state in the register $B$ of the first copy in which Alice's measurement is successful is $\rho$. Furthermore, given Alice's success in a copy, Bob's measurement also succeeds with high probability, and hence does not disturb the state in the register $B$ much, conditioned on success. Using a similar technique as in the classical protocol, we argue that Alice can inform Bob about the index of this copy with communication of $\mathcal{O}(c)$ bits (for constant $\epsilon$). For the case where only $\mathrm{S}(\rho||\sigma)$ is known to Alice and Bob, the same protocol is used to construct a state $\rho' \in \mathsf{D}(\mathcal{H})$ such that $\mathrm{F}(\rho, \rho') \geq 1 - \epsilon$ and $\mathrm{S}_{\max}(\rho'||\sigma) \leq \dfrac{\mathrm{S}(\rho||\sigma) + 1}{\epsilon} + \lg(\dfrac{1}{1-\epsilon})$ . The existence of $\rho'$ is guaranteed by the Quantum Substate Theorem [33, 32].

### 1.3.4 Quantum Resampling

Ozols, Roetteler and Roland [42] study a slightly different generalization of Task (i) to the quantum case in the query complexity setting, which they call quantum resampling task. The quantum resampling task can be described as follows. Let $n, d$ be positive integers and $\{|\zeta_i\rangle\}_{i=1}^n \subseteq \mathbb{C}^d$ be a set of normalized quantum states. Let $P, Q \in \mathbb{R}^n$ be two real vectors such that $P(i), Q(i) \geq 0$ for every $i \in \{1, ..., n\}$ and $\sum_{i=1}^n P(i)^2 = \sum_{i=1}^n Q(i)^2 = 1$. Let $O \in \mathsf{U}(\mathbb{C}^{dn})$ be a unitary operator which maps a default state $|\bar{0}\rangle \in \mathbb{C}^d \otimes \mathbb{C}^n$ to the state $|P^\zeta\rangle = \sum_{i=1}^n P(i)\ |\zeta_i\rangle|i\rangle$. Given oracle access to unitary black boxes $O$ and $O^*$, the quantum resampling task is to prepare the state $|Q^\zeta\rangle = \sum_{i=1}^n Q(i)\ |\zeta_i\rangle|i\rangle$. Note that while $P$ and $Q$ are known, the fact that the states $|\zeta_i\rangle$ are unknown makes the problem nontrivial. They give a tight characterization of the query complexity of quantum resampling problem. They prove that the query complexity of the quantum resampling problem for the pair $(P, Q)$ with success probability $p$ is given by $\Theta\left(1/\epsilon_{Q\to P}(p)\right)$, where $\epsilon_{Q\to P}(p)$ is the Euclidean norm of a vector characterizing the amplitudes of the final state prepared by the best algorithm having success probability $p$. The lower bound comes from an extension of the automorphism principle [30, 3], to the framework of quantum state preparation with oracles. The upper bound follows from an algorithm based on amplitude amplification which the authors refer to as quantum rejection sampling.

### 1.3.5 Summary of the results and organization of this thesis

The rest of this thesis is organized as follows. The next chapter contains the mathematical preliminaries and the notations used in this thesis. In chapter 3, we first describe the classical rejection sampling protocol of Harsha *et al.* [25], then we introduce quantum rejection sampling protocols for Task (ii), described in Section 1.1. In Section 3.2.2, we introduce the Greedy Quantum Rejection Sampler which is a natural extension of the protocol of Harsha *et al.* [25] to the quantum setting. For generating a quantum state $\rho$ using purifications of another quantum state $\sigma$, we characterize the expected communication cost of the Greedy Quantum Rejection Sampler in terms of max-relative entropy of $\rho$ with respect to $\sigma$, in the case where $\rho$ is a pure state, and we show that this protocol is an optimal quantum rejection sampling protocol in this case. In Section 3.3.3, we present some of the approaches we took for upper bounding the expected communication cost of the Greedy Quantum Rejection Sampler protocol for general states. In Section 3.4.1, we describe an optimal quantum rejection sampling protocol in terms of an optimization problem and we find general lower bounds and upper bounds on the expected communication cost of the optimal protocol. More precisely, we prove that the optimal value of this optimization problem is bounded below by $\mathrm{S}(\rho||\sigma')$, where $\sigma' = \displaystyle\sum_{x\in\mathcal{X}} \langle\psi_x|\sigma|\psi_x\rangle|\psi_x\rangle\langle\psi_x|$ and $\{|\psi_x\rangle\}_{x\in\mathcal{X}}$ is a set of eigenvectors of $\rho$. We also show two different upper bounds on the expected communication cost of an optimal quantum rejection sampling protocol, one in terms of $\mathrm{S}_{\max}(\rho||\sigma)$ and one in terms of $\sum_{x\in\mathcal{X}} \lambda_x \mathrm{S}_{\max}(|\psi_x\rangle\langle\psi_x|||\sigma)$, where $\lambda_x$ is the eigenvalue corresponding to eigenvector $|\psi_x\rangle$ of $\rho$. In Section3.5.1 , we propose an LOCC compression protocol based on the Greedy Quantum Rejection Sampler protocol, for lossless compression of an arbitrary pure state quantum source in the visible compression model. Let $\mathcal{S}$ be an arbitrary quantum source corresponding to the ensemble $\{P(x), |\psi_x\rangle\}_{x\in\mathcal{X}}$ of pure states in a finite dimensional Hilbert space $\mathcal{H}$. We give an upper bound of

$$\min_{\sigma\in\mathsf{D}(\mathcal{H})} \sum_{x\in\mathcal{X}} P(x)\mathrm{S}_{\max}(|\psi_x\rangle\langle\psi_x|||\sigma) + \mathcal{O}\left(\lg\left(\sum_{x\in\mathcal{X}} P(x)\mathrm{S}_{\max}\left(|\psi_x\rangle\langle\psi_x|||\sigma\right) + 1\right)\right) \ ,$$

on the average length of an optimal lossless LOCC encoding of the source $\mathcal{S}$. In Section 3.5.2, we introduce a high-entanglement deterministic exact RSP protocol for remote preparation of an arbitrary ensemble $\{P(x), \rho_x\}_{x\in\mathcal{X}}$ in a $d$-dimensional Hilbert space, which uses variable-length encoding for communication of the messages. Let $\xi_{AB}$ denote the joint quantum state of Alice's classical input register $A$ and Bob's output register $B$ after the protocol terminates, then $\xi_{AB} = \sum_{x\in\mathcal{X}} P(x)|x\rangle\langle x|^A \otimes \rho_x^B$. We show that the expected communication cost of our protocol for the worst case input $x$ is at most $\mathrm{I}_{\max}(A :$

$B)_{\xi_{AB}} + \mathcal{O}\left(\lg\left(\mathrm{I}_{\max}(A:B)_{\xi_{AB}} + 1\right)\right)$, where $\mathrm{I}_{\max}(A : B)_{\xi_{AB}}$, the max-information in register $B$ about the register $A$ at the end of the protocol, is always less than or equal to $\lg(d)$. Furthermore, we show that our protocol can be used for exact remote preparation of a generic $d$-dimensional quantum state at an expected communication cost of at most $\lg(d) + \mathcal{O}\left(\lg(\lg(d))\right)$ cbits. Finally, an overview of the results is followed by some open questions in Chapter 4.

# Chapter 2

# Preliminaries

## 2.1 Mathematical background

In the current chapter, we introduce the notations and mathematical notions used in this thesis. The reader may find the proofs and details of the properties presented in this chapter in John Watrous's course notes [48].

### Finite dimensional Hilbert spaces

Let $\mathbb{C}$ denote the set of complex numbers and for every $n \in \mathbb{N}$, let $\mathbb{C}^n$ denote the $n$-fold Cartesian product of $\mathbb{C}$. In this thesis we use the Dirac bra-ket notation. We denote the elements of $\mathbb{C}^n$ by the ket notation e.g. $|\psi\rangle$.

For every $n \in \mathbb{N}$, the set $\mathbb{C}^n$ is a vector space over the field of complex numbers with the standard definition for addition and scalar multiplication. Any such vector equipped with standard inner product is referred to as a finite dimensional Hilbert space. We will denote Hilbert spaces by scripted capital letters such as $\mathcal{H}$, $\mathcal{K}$, and $\mathcal{M}$.

### Inner product and Euclidean norm

Let $|\phi\rangle$ and $|\psi\rangle$ be two vectors in a finite dimensional Hilbert space $\mathcal{H}$. We denote $\langle|\psi\rangle, |\phi\rangle\rangle$,the standard inner product of two vectors $|\phi\rangle$ and $|\psi\rangle$, by $\langle\psi|\phi\rangle$.

Two vectors $|\phi\rangle$ and $|\psi\rangle$ are called orthogonal if $\langle\psi|\phi\rangle = 0$.

The Euclidean norm or 2-norm of a vector $|\phi\rangle \in \mathcal{H}$ is defined as

$$\||\phi\rangle\| \quad := \quad \sqrt{\langle\phi|\phi\rangle} \ .$$

A vector $|\phi\rangle$ is called normal or unit vector if $\||\phi\rangle\| = 1$. An orthonormal basis for $\mathcal{H}$ is a set of mutually orthogonal unit vectors spanning $\mathcal{H}$. The *standard basis* of $\mathbb{C}^n$ is the orthonormal basis $\{|i\rangle : i \in \{1, ..., n\}\}$ where $|i\rangle$ corresponds to the unit vector which is equal to one in the $i$-th coordinate.

### Tensor product

Let $\mathcal{H}_1 = \mathbb{C}^{n_1}$, ... ,$\mathcal{H}_k = \mathbb{C}^{n_k}$ . The *tensor product* of $\mathcal{H}_1$ , ... , $\mathcal{H}_k$ is the finite dimensional Hilbert space

$$\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_k = \mathbb{C}^{n_1 \times \ldots \times n_k} \ .$$

For $i_1 = 1, \ldots, n_1$ , ... , $i_k = 1, \ldots, n_k$ , the vector $|i_1\rangle \otimes \ldots \otimes |i_k\rangle \in \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_k$ corresponds to $|\,(i_1, \ldots, i_k)\rangle$, a standard basis element of $\mathbb{C}^{n_1 \times \ldots \times n_k}$.

For $|\psi_1\rangle \in \mathcal{H}_1, \ldots, |\psi_k\rangle \in \mathcal{H}_k$, the vector $|\psi_1\rangle \otimes \cdots \otimes |\psi_k\rangle \in \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_k$ is defined as

$$\langle (|i_1\rangle \otimes \cdots \otimes |i_k\rangle), (|\psi_1\rangle \otimes \cdots \otimes |\psi_k\rangle) \rangle \quad = \quad \langle i_1|\psi\rangle_1 \times \cdots \times \langle i_k|\psi\rangle_k \ .$$

### Linear operators

Let $\mathcal{H}$ and $\mathcal{K}$ be two Hilbert spaces. A mapping $A : \mathcal{H} \longrightarrow \mathcal{K}$ is called *linear* if for every $|\phi\rangle, |\psi\rangle \in \mathcal{H}$ and every scalar $a \in \mathbb{C}$, the following conditions hold:

- $A(|\psi\rangle + |\phi\rangle) = A(|\psi\rangle) + A(|\phi\rangle)$
- $A(a|\phi\rangle) = aA(|\phi\rangle)$ .

We denote the set of all linear mappings from $\mathcal{H}$ to $\mathcal{K}$ by $\mathsf{L}(\mathcal{H}, \mathcal{K})$, and the set of linear mappings from $\mathcal{H}$ to itself is denoted by $\mathsf{L}(\mathcal{H})$. We denote the identity mapping on $\mathcal{H}$ by $\mathbb{1}_{\mathcal{H}}$. Any linear mapping $A$ of the form $A : \mathcal{H} \longrightarrow \mathcal{K}$ can be represented by a matrix $M_A$ defined as $M_A(i,j) := \langle i|A|j\rangle$ for $i = 1, ..., \dim(\mathcal{H})$ and $j = 1, ..., \dim(\mathcal{K})$. For convenience for every linear operator $A \in \mathsf{L}(\mathcal{H}, \mathcal{K})$, we will denote the matrix $M_A$ by $A$.

Let $A \in \mathsf{L}(\mathcal{H}, \mathcal{K})$. $\bar{A} \in \mathsf{L}(\mathcal{H}, \mathcal{K})$, the *conjugate* of $A$ is the mapping given by $\bar{A}(i,j) = \overline{A(i,j)}$. $A^{\mathrm{T}} \in \mathsf{L}(\mathcal{K}, \mathcal{H})$, the *transpose* of $A$ is defined as $A^{\mathrm{T}}(i,j) = A(j,i)$. $A^* \in \mathsf{L}(\mathcal{K}, \mathcal{H})$ the *adjoint* of $A$ is the unique operator satisfying $\langle |\phi\rangle, A|\psi\rangle \rangle = \langle A^*|\phi\rangle, |\psi\rangle \rangle$ for every $|\phi\rangle \in \mathcal{K}$ and $|\psi\rangle \in \mathcal{H}$, and is equal to $\bar{A}^{\mathrm{T}}$.

### Tensor product of linear operators

Let $\mathcal{H}_1, \ldots, \mathcal{H}_n$ and $\mathcal{K}_1, \ldots, \mathcal{K}_n$ be finite dimensional Hilbert spaces, and

$$A_1 \in \mathsf{L}(\mathcal{H}_1, \mathcal{K}_1), \ldots, A_n \in \mathsf{L}(\mathcal{H}_n, \mathcal{K}_n) \ .$$

Then the linear operator $A_1 \otimes \cdots \otimes A_n \in \mathsf{L}(\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n, \mathcal{K}_1 \otimes \cdots \otimes \mathcal{K}_n)$ is defined as the unique operator satisfying

$$(A_1 \otimes \cdots \otimes A_n)\,(|\psi_1\rangle \otimes \cdots \otimes |\psi_n\rangle) \quad = \quad A_1(|\psi_1\rangle) \otimes \cdots \otimes A_n(|\psi_n\rangle) \ ,$$

for every $|\psi_1\rangle \in \mathcal{H}_1$, ... , $|\psi_n\rangle \in \mathcal{H}_n$.

### Eigenvalues and eigenvectors

Let $A \in \mathsf{L}(\mathcal{H})$ be a linear operator on $\mathcal{H}$ and $|\phi\rangle \in \mathcal{H}$ be a nonzero vector such that $A|\phi\rangle = \lambda|\phi\rangle$ for some complex number $\lambda$. The vector $|\phi\rangle$ is called an eigenvector of $A$ and $\lambda$ is referred to as the corresponding eigenvalue of $A$.

### Different classes of linear operators

In this section we introduce important classes of linear operators on a finite dimensional Hilbert space $\mathcal{H}$.

- **Normal Operators:** An operator $A \in \mathsf{L}(\mathcal{H})$ is *normal* if and only if $A^*A = AA^*$.
- **Hermitian operators:** An operator $A \in \mathsf{L}(\mathcal{H})$ is *Hermitian* if and only if $A = A^*$. We denote the set of all Hermitian operators on a Hilbert space $\mathcal{H}$ by $\mathsf{Herm}(\mathcal{H})$.
- **Positive semidefinite operators:** An operator $A \in \mathsf{L}(\mathcal{H})$ is *positive semidefinte* if and only if it is Hermitian and every eigenvalue of $A$ is non-negative. We denote the set of all positive semidefinite operators on a Hilbert space $\mathcal{H}$ by $\mathsf{Psd}(\mathcal{H})$. Alternatively, the notation $A \geq 0$ is used to state that $A$ is a positive semidefinite operator. Also, for Hermitian operators $A, B \in \mathsf{Herm}(\mathcal{H})$, the notation $A \geq B$ means that $A - B \in \mathsf{Psd}(\mathcal{H})$. This partial order on the set of Hermitian operators is referred to as the *Löwner order*.

- **Positive definite operators:** An operator $A \in \mathsf{L}(\mathcal{H})$ is *positive definite* if and only if it is Hermitian and every eigenvalue of $A$ is positive. We denote the set of all positive definite operators on a Hilbert space $\mathcal{H}$ by $\mathsf{Pd}(\mathcal{H})$. The notation $A > 0$ means that $A$ is a positive definite operator.

- **Density operators:** An operator $A \in \mathsf{L}(\mathcal{H})$ is a *density operator* if and only if $A \in \mathsf{Psd}(\mathcal{H})$ and $\mathrm{Tr}(A) = 1$. We denote the set of all density operators on a Hilbert space $\mathcal{H}$ by $\mathsf{D}(\mathcal{H})$.

- **Unitary operators:** An operator $A \in \mathsf{L}(\mathcal{H})$ is *unitary* if and only if it satisfies $A^*A = AA^* = \mathbb{1}_{\mathcal{H}}$. We denote the set of all unitary operators on a Hilbert space $\mathcal{H}$ by $\mathsf{U}(\mathcal{H})$. The Pauli operators, $\{\mathrm{X}, \mathrm{Y}, \mathrm{Z}\} \subset \mathsf{U}(\mathbb{C}^2)$, are a set of three unitary operators defined as

$$\mathrm{X} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \mathrm{Y} = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \mathrm{Z} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} .$$

**Lemma 2.1.1.** *The Pauli operators together with the identity operator,* $\mathbb{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, *span* $\mathsf{L}(\mathbb{C}^2)$*, the vector space of 2 by 2 linear operators.*

- **Projection operators:** An operator $A \in \mathsf{L}(\mathcal{H})$ is a *projection operator* if and only if $A \in \mathsf{Psd}(\mathcal{H})$ and it satisfies $A^2 = A$.

### Eigenvalue decomposition

The following theorem states that any normal operator can be expressed as a linear combination of a set of rank one orthonormal projection operators.

**Theorem 2.1.2.** *Let $\mathcal{H}$ be a Hilbert space and $A \in \mathsf{L}(\mathcal{H})$ be a normal operator with eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n \in \mathbb{C}$. There exists an orthonormal basis $\{|\psi_1\rangle, |\psi_2\rangle, ..., |\psi_n\rangle\}$ of $\mathcal{H}$ such that*

$$A \quad = \quad \sum_{i=1}^{n} \lambda_i |\psi_i\rangle\langle\psi_i| \ .$$

Note that for every $i \in \{1, ..., n\}$, $|\psi_i\rangle$ is an eigenvector of $A$ corresponding to the eigenvalue $\lambda_i$. We will refer to any such decomposition of normal operator $A$ as an eigenvalue decomposition of $A$.

### Functions of normal operators

Every function $f : \mathbb{C} \longrightarrow \mathbb{C}$ can be extended to the set of normal operators on a Hilbert space $\mathcal{H}$, using eigenvalue decomposition. For every normal operator $A \in \mathsf{L}(\mathcal{H})$ with eigenvalue decomposition $A \quad = \quad \sum_{i=1}^{n} \lambda_i |\psi_i\rangle\langle\psi_i|$, $f(A)$ is defined as

$$f(A) \quad := \quad \sum_{i=1}^{n} f(\lambda_i) |\psi_i\rangle\langle\psi_i| \ .$$

### Trace norm

Let $\mathcal{H}$ be a finite dimensional Hilbert space and $A \in \mathsf{L}(\mathcal{H})$. The *trace norm* of $A$ denoted by $\|A\|_{tr}$ is defined as $\|A\|_{\mathrm{tr}} = \mathrm{Tr}\left(\sqrt{A^* A}\right)$.

### Some basic notions of analysis

Let $\mathcal{H}$ be a finite dimensional Hilbert space. The *open ball* of radius $r$ about a vector $|\psi\rangle \in \mathcal{H}$ is defined as

$$\mathcal{B}_r(|\psi\rangle) \quad := \quad \{|\phi\rangle \in \mathcal{H} \ : \ \||\psi\rangle - |\phi\rangle\| < r\} \ .$$

We say that $A \subseteq \mathcal{H}$ is bounded if it is contained in $\mathcal{B}_r(0)$ for some positive real number $r$.

A set $A \subseteq \mathcal{H}$ is *open* with respect to $\mathcal{H}$, if for every $|\phi\rangle \in A$ there exists some $\epsilon > 0$ such that $\mathcal{B}_\epsilon(|\phi\rangle) \subseteq A$. A set $A \subseteq \mathcal{H}$ is *closed* if its complement with respect to $\mathcal{H}$ is open.

A family $\{O_a \ : \ a \in \Sigma\} \subseteq \mathcal{H}$ of open sets is an *open cover* for a set $A \subseteq \mathcal{H}$ if $A \subseteq \cup_{a\in\Sigma} O_a$. A set $A \subseteq \mathcal{H}$ is *compact* in $\mathcal{H}$ if every open cover of $A$ has a finite subcover, i.e. for every open cover $\{O_a \ : \ a \in \Sigma\}$ of $A$ there exists a finite subset $\Gamma \subseteq \Sigma$ such that $A \subseteq \cup_{a\in\Gamma} O_a$. In any finite dimensional Hilbert space $\mathcal{H}$, $A \subseteq \mathcal{H}$ is compact with respect to $\mathcal{H}$ if and only if $A$ is closed with respect to $\mathcal{H}$ and bounded.

**Theorem 2.1.3.** *If $A$ is non-empty and compact and $f \ : \ A \longrightarrow \mathbb{R}$ is continuous on $A$, then $f$ achieves both a maximum and a minimum value on $A$.*

Let $V$ be a vector space over the field of real numbers. A set $A \subseteq V$ is *convex* if for every $u, v \in A$ and every $\lambda \in [0, 1]$, we have $\lambda u + (1 - \lambda)v \in A$. A *convex combination* of vectors in $A$ is a sum of the form $\sum_{i\in\Sigma} P(i)u_i$, where $\Sigma$ is a finite nonempty set, $\{u_i : i \in \Sigma\} \subset A$, and $P : \Sigma \longrightarrow \mathbb{R}^{\Sigma}$ is a probability distribution. The *convex hull* of $A \subseteq V$ is the intersection of all convex sets containing $A$, which is equal to the set of all points in $V$ which can be written as a convex combination of the elements in $A$. Let $A$ be a convex set and $f : A \longrightarrow \mathbb{R}$ be a function. Then $f$ is a *convex function* over $A$ if for every $u, v \in A$ and every $\lambda \in [0,1]$, we have $f(\lambda u + (1-\lambda)v) \leq \lambda f(u) + (1-\lambda) f(v)$. $f$ is a *concave function* if $-f$ is convex.

**Lemma 2.1.4.** *(Jensen's inequality) Let $I \subseteq \mathbb{R}$ be a convex set, $X$ be a random variable taking values in $I$, and $f : I \longrightarrow \mathbb{R}$ be a convex function over $I$. Then we have $\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$.*

### Semidefinite programming

A semidefinite program can be formally defined in many different ways. Here we prefer to use John Watrous's definition for a semidefinite program [48].

A linear mapping $\Phi : \mathsf{L}(\mathcal{H}) \longrightarrow \mathsf{L}(\mathcal{K})$ is *Hermiticity preserving* if and only if for every $\rho \in \mathsf{Herm}(\mathcal{H})$ it holds that $\Phi(\rho) \in \mathsf{Herm}(\mathcal{K})$.

A *semidefinite program* is a triple $(\Phi, A, B)$, where $\Phi : \mathsf{L}(\mathcal{H}) \longrightarrow \mathsf{L}(\mathcal{K})$ is a Hermiticity preserving linear map and $A \in \mathsf{Herm}(\mathcal{H})$ and $B \in \mathsf{Herm}(\mathcal{K})$ are Hermitian operators.

The triple $(\Phi, A, B)$ defines two optimization problems referred to as the *primal* and *dual* problems

Primal Problem

$$\begin{aligned} \text{maximize :} \quad & \langle A, X \rangle \\ \text{subject to :} \quad & \Phi(X) \leq B \\ & X \in \mathsf{Herm}(\mathcal{H}) \end{aligned}$$

Dual problem

$$\begin{aligned} \text{minimize :} \quad & \langle B, Y \rangle \\ \text{subject to :} \quad & \Phi^*(Y) = A \\ & Y \in \mathsf{Psd}(\mathcal{K}) \end{aligned}$$

where $\Phi^* : \mathsf{L}(\mathcal{K}) \longrightarrow \mathsf{L}(\mathcal{H})$ is the adjoint of the linear map $\Phi$, which is the unique linear map satisfying

$$\langle \Phi(X), Y \rangle = \langle X, \Phi^*(Y) \rangle \ ,$$

for every $X \in \mathsf{L}(\mathcal{H})$ and $Y \in \mathsf{L}(\mathcal{K})$.

The linear functions $\langle A, X \rangle$ and $\langle B, Y \rangle$ are referred to as the primal and dual *objective functions*, respectively. The conditions $X \in \mathsf{Herm}(\mathcal{H})$ and $\Phi(X) \leq B$ are called the *primal constraints*, and similarly the conditions $Y \in \mathsf{Psd}(\mathcal{K})$ and $\Phi^*(Y) = A$ are called the *dual constraints.* An operator $X$ is called a *primal (feasible) solution* if it satisfies the *primal constraints.* Similarly, an operator $Y$ is called a *dual (feasible) solution* if it satisfies the *dual constraints.* We denote by $\mathcal{A}$ and $\mathcal{B}$ the set of all primal and dual solutions, respectively. The *primal optimal value* is defined as

$$\mathrm{Opt}_{\mathrm{P}} \quad := \quad \mathrm{Sup}_{X \in \mathcal{A}} \langle A, X \rangle \ ,$$

similarly, the *dual optimal value* is defined as

$$\mathrm{Opt}_{\mathrm{D}} \quad := \quad \mathrm{Inf}_{X \in \mathcal{B}} \langle B, Y \rangle \ .$$

**Proposition 2.1.5.** *(Weak duality) For every semidefinite program $(\Phi, A, B)$ it holds that* $\mathrm{Opt}_{\mathrm{P}} \leq \mathrm{Opt}_{\mathrm{D}}$.

The condition that $\mathrm{Opt}_{\mathrm{P}}$=$\mathrm{Opt}_{\mathrm{D}}$ and at least one of $\mathrm{Opt}_{\mathrm{P}}$ or $\mathrm{Opt}_{\mathrm{D}}$ is achieved is referred to as *strong duality.* Unlike weak duality, strong duality does not necessarily hold for every semidefinite program. The following theorem gives a set of sufficient conditions for strong duality to hold.

**Theorem 2.1.6.** *(Strong duality theorem) For every semidefinite program $(\Phi, A, B)$,*

- *If $\mathcal{A} \neq \emptyset$ (primal is feasible) and there exists $Y > 0$ such that $\Phi^*(Y) = A$ (dual is strictly feasible), then* $\mathrm{Opt}_{\mathrm{P}} = \mathrm{Opt}_{\mathrm{D}}$ *and there exists a primal feasible solution $X \in \mathcal{A}$ such that $\langle A, X \rangle = \mathrm{Opt}_{\mathrm{P}}$.*

- *If $\mathcal{B} \neq \emptyset$ (dual is feasible) and there exists $X \in \mathsf{Herm}(\mathcal{H})$ such that $\Phi(X) < B$ (primal is strictly feasible), then* $\mathrm{Opt}_{\mathrm{P}} = \mathrm{Opt}_{\mathrm{D}}$ *and there exists a dual feasible solution $Y \in \mathcal{B}$ such that $\langle B, Y \rangle = \mathrm{Opt}_{\mathrm{D}}$.*

### Majorization for real vectors

For a vector $v \in \mathbb{R}^n$, let $v^{\downarrow} \in \mathbb{R}^n$ denote the vector with the same elements as $v$, but sorted in descending order. For $u, v \in \mathbb{R}^n$, we say that $u$ *majorizes* $v$, denoted $u \succ v$, if $\sum_{i=1}^n u(i) = \sum_{i=1}^n v(i)$ and for every $k \in \{1, 2, \ldots, n-1\}$ we have

$$\sum_{i=1}^{k} u^{\downarrow}(i) \geq \sum_{i=1}^{k} v^{\downarrow}(i) \ .$$

## 2.2 Basic notions of quantum information theory

In quantum information theory, any physical system which may change over time is modelled as a *register.* We denote registers by capital letters e.g. $A$, $B$, and $X$. To each register we associate a finite dimensional Hilbert space. A finite sequence of registers $X_1, X_2, ..., X_n$ associated with Hilbert spaces $\mathcal{H}_1,\ \mathcal{H}_2, ...,\ \mathcal{H}_n$ can be viewed as a single register $Y = (X_1, X_2, ..., X_n)$ associated with the Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes ... \otimes \mathcal{H}_n$.

### Quantum states

The *quantum state* of a register is represented by density operators in the Hilbert space associated with the register. We will denote quantum states by lower case Greek letters such as $\rho$, $\sigma$, $\zeta$. We use the notation $\rho_{AB}$ to emphasize that $\rho$ is the quantum state of the register $AB$.

A quantum state $\rho$ is said to be *pure* if it is of the form $\rho = |\psi\rangle\langle\psi|$ for some unit vector $|\psi\rangle$, otherwise it is called a *mixed state.* Using eigenvalue decomposition it is possible to write any mixed state as a convex combination of pure states.

A quantum state $\rho \in \mathsf{D}(\mathcal{H} \otimes \mathcal{K})$ is called *separable* on the Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$ if and only if $\rho$ can be written as

$$\rho \quad = \quad \sum_{i=1}^{n} P(i)\, \sigma_i \otimes \zeta_i \ ,$$

for some $\{\sigma_i \ : \ i = 1, ..., n\} \subset \mathsf{D}(\mathcal{H})$ and $\{\zeta_i \ : \ i = 1, ..., n\} \subset \mathsf{D}(\mathcal{K})$. A quantum state is called *entangled* if it is not separable.

### Quantum measurements

Let $X$ be a register and $\mathcal{H}$ be the Hilbert space associated with $X$. A *measurement* on the register $X$ with $n$ classical outcomes is a set of positive semidefinite operators

$$\{E_i \ : \ i = 1, ..., n\} \subset \mathsf{Psd}(\mathcal{H})$$

such that

$$\sum_{i=1}^{n} E_i = \mathbb{1}_{\mathcal{H}} \ \ .$$

Each $E_i$ is called the measurement operator or POVM element corresponding to the outcome $i$ , and if the state of the register $X$ before the measurement is $\rho$, then the outcome of the measurement is $i$ with probability $\Pr(i) = \langle E_i \, , \, \rho \rangle$ .

A special class of measurements are *projective measurements.* A projective measurement is a measurement in which the measurement operators are projection operators onto mutually orthogonal subspaces. For a projective measurement with measurement operators $\{\Pi_i\}_{i=1}^{n}$, if the state of the register $X$ before the measurement is $\rho$, then with probability $\Pr(i) = \langle \Pi_i \, , \, \rho \rangle$ the outcome of the measurement is $i$ and the post-measurement state of $X$ is given by

$$\frac{\Pi_i \rho \Pi_i}{\langle \Pi_i \, , \, \rho \rangle} \ \ .$$

For an orthonormal basis $\{|\psi_i\rangle : i \in \{1, ..., n\}\} \subseteq \mathbb{C}^n$, the projective measurement

$$\{|\psi_i\rangle\langle\psi_i| : i \in \{1, ..., n\}\}$$

is referred to as the measurement with respect to or in the basis $\{|\psi_i\rangle : i \in \{1, ..., n\}\}$.

### Quantum channels

A linear mapping $\Phi \, ; \, \mathsf{L}(\mathcal{H}) \longrightarrow \mathsf{L}(\mathcal{K})$ is *completely positive* if and only if for every choice of Hilbert space $\mathcal{M}$ and every density operator $\rho \in \mathsf{D}(\mathcal{H} \otimes \mathcal{M})$ it holds that

$$\left(\Phi \otimes \mathbb{1}_{\mathsf{L}(\mathcal{M})}\right)(\rho) \in \mathsf{D}(\mathcal{H} \otimes \mathcal{M}) \ \ ,$$

where $\mathbb{1}_{\mathsf{L}(\mathcal{M})}$ denotes the identity transformation on $\mathsf{L}(\mathcal{M})$.

A linear mapping $\Phi \ : \ \mathsf{L}(\mathcal{H}) \longrightarrow \mathsf{L}(\mathcal{K})$ is *trace-preserving* if and only if $\mathrm{Tr}(\Phi(X)) = \mathrm{Tr}(X)$ for every linear operator $X \in \mathsf{L}(\mathcal{H})$.

Any quantum operation can be described as a quantum channel which maps the states of a register into states of another register. A *quantum channel* from a register $X$ associated with the Hilbert space $\mathcal{H}$ to a register $Y$ associated with the Hilbert space $\mathcal{K}$ is a linear mapping $\Phi : \mathsf{L}(\mathcal{H}) \longrightarrow \mathsf{L}(\mathcal{K})$ which is both completely positive and trace-preserving.

The *partial trace* mapping is an example of a quantum channel. The partial trace over the Hilbert space $\mathcal{K}$ denoted $\mathrm{Tr}_{\mathcal{K}} : \mathsf{L}(\mathcal{H} \otimes \mathcal{K}) \longrightarrow \mathsf{L}(\mathcal{H})$ is defined as

$$\mathrm{Tr}_{\mathcal{K}}(\rho_{XY}) \quad := \quad \left(\mathbb{1}_{\mathsf{L}(\mathcal{H})} \otimes \mathrm{Tr}\right)(\rho_{XY}) \ ,$$

for every quantum state $\rho_{XY} \in \mathsf{D}(\mathcal{H} \otimes \mathcal{K})$ of the register $XY$. The notation $\rho_X$ is used to denote $\mathrm{Tr}_{\mathcal{K}}(\rho_{XY})$. Alternatively, we may replace the name of the Hilbert space which is being traced-out by the corresponding register to denote the partial trace mapping, i.e. we may write $\mathrm{Tr}_{\mathrm{Y}}(\rho_{XY})$.

### Reductions, extensions, and purifications

Let $\mathcal{H}$ and $\mathcal{K}$ be the Hilbert spaces associated with registers $X$ and $Y$, respectively, and $\rho_{XY} \in \mathsf{D}(\mathcal{H} \otimes \mathcal{K})$. The state $\rho_X = \mathrm{Tr}_{\mathcal{K}}(\rho_{XY})$ is called the *reduced state* of $\rho_{XY}$ to $\mathcal{H}$ (or $X$).

Conversely, let $\sigma \in \mathsf{D}(\mathcal{H})$ be a quantum state of the register $X$. Any state $\rho_{XY} \in \mathsf{D}(\mathcal{H} \otimes \mathcal{K})$ such that $\mathrm{Tr}_{\mathcal{K}}(\rho_{XY}) = \sigma$ is called an *extension* of $\sigma$. In this case, if $\rho = |\psi\rangle\langle\psi|$ is a pure state it is said that $\rho$ (or $|\psi\rangle$) is a *purification* of $\sigma$.

**Theorem 2.2.1.** *Let $\mathcal{H}$ and $\mathcal{K}$ be the Hilbert spaces associated with registers $X$ and $Y$, and let $\rho \in \mathsf{D}(\mathcal{H})$. A purification $|\psi\rangle \in \mathcal{H} \otimes \mathcal{K}$ of $\rho$ exists if and only if* $\dim(\mathcal{K}) \geq \mathrm{rank}(\rho)$.

**Theorem 2.2.2.** ***(Unitary equivalence of purifications)*** *Let $\mathcal{H}$ and $\mathcal{K}$ be Hilbert spaces, and suppose that $|\phi\rangle$, $|\psi\rangle \in \mathcal{H} \otimes \mathcal{K}$ satisfy*

$$\mathrm{Tr}_{\mathcal{K}}(|\phi\rangle\langle\phi|) = \mathrm{Tr}_{\mathcal{K}}(|\psi\rangle\langle\psi|) \ .$$

*There exists a unitary operator $U \in \mathsf{U}(\mathcal{K})$ such that $|\phi\rangle = (\mathbb{1}_{\mathcal{H}} \otimes U)|\psi\rangle$.*

### The fidelity function

Let $P$, $Q \in \mathsf{Psd}(\mathcal{H})$ be two semidefinite operators on a Hilbert space $\mathcal{H}$. The *fidelity* between $P$ and $Q$ is defined as

$$\mathrm{F}(P, Q) \quad := \quad \mathrm{Tr}\left(\sqrt{\sqrt{P}Q\sqrt{P}}\right) \ .$$

## Some information theoretic quantities

In this thesis we use lg to denote logarithm base 2.

Let $P, Q : \mathcal{X} \longrightarrow \mathbb{R}$ be probability distributions over a set $\mathcal{X}$ . The *Shannon entropy* of $P$ is a measure of uncertainty in a random experiment described by the distribution $P$ which is defined as

$$\mathrm{H}(P) \quad := \quad -\sum_{x \in \mathcal{X}} P(x) \lg(P(x)) \ .$$

The *relative entropy* of $P$ with respect to $Q$ is a non symmetric measure of the distance between $P$ and $Q$ which is defined as

$$\mathrm{S}(P||Q) \quad := \quad \sum_{x \in \mathcal{X}} P(x) \lg \left( \frac{P(x)}{Q(x)} \right) \ .$$

Let $\mathcal{H}$ be a Hilbert space and $\rho, \sigma \in \mathsf{D}(\mathcal{H})$. The *von Neumann entropy* of $\rho$ is defined as

$$\mathrm{S}(\rho) \quad := \quad -\mathrm{Tr}(\rho \lg(\rho)) \ .$$

The *quantum relative entropy* of $\rho$ with respect to $\sigma$ is defined as

$$\mathrm{S}(\rho||\sigma) \quad := \quad \mathrm{Tr}(\rho \lg(\rho)) - \mathrm{Tr}(\rho \lg(\sigma)) \ .$$

The *max-relative entropy* of $\rho$ with respect to $\sigma$ is defined as

$$\mathrm{S}_{\max}(\rho||\sigma) \quad := \quad \min\left\{\lambda \ : \ \rho \leq 2^{\lambda}\sigma\right\} \ .$$

Let $X$ and $Y$ be two quantum registers associated with finite dimensional Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$, respectively. Let $\rho_{XY} \in \mathsf{D}(\mathcal{H} \otimes \mathcal{K})$ be the joint quantum state of the register $XY$. The *max-information* in register $Y$ about register $X$ is defined as

$$\mathrm{I}_{\max}(X:Y)_{\rho_{XY}} \quad := \quad \min_{\sigma \in \mathsf{D}(\mathcal{K})} \mathrm{S}_{\max}(\rho_{XY}||\rho_X \otimes \sigma) \ .$$

**Theorem 2.2.3.** ***(Quantum Substate Theorem[33])*** *Let $\rho, \sigma \in \mathsf{D}(\mathcal{H})$ be quantum states in the Hilbert space $\mathcal{H}$. For any $\epsilon > 0$, there exists $\rho' \in \mathsf{D}(\mathcal{H})$ such that*

$$\mathrm{F}(\rho, \rho') \quad \geq \quad 1 - \epsilon \qquad \textit{and} \qquad \mathrm{S}_{\max}(\rho'||\sigma) \quad \leq \quad \frac{\mathrm{S}(\rho||\sigma) + 1}{\epsilon} + \lg(\frac{1}{1-\epsilon}) \ .$$

### Classical variable-length codes

Let $X$ be a random variable taking values in $\mathcal{X}$. A *variable-length* encoding of the random variable $X$ is a function $C : \mathcal{X} \longrightarrow \{0,1\}^*$, where $\{0,1\}^*$ denotes the set of all finite-length binary strings. For every $x \in \mathcal{X}$, the length of the codeword $C(x)$ is denoted by $l_C(x)$, and the average length of the code $C$ is defined as $\mathbb{E}(l_C(X)) = \sum_{x\in\mathcal{X}} P(x)l_C(x)$. The *extension* $C^*$ of a code $C$ is the mapping from finite-length strings of elements of $\mathcal{X}$ to $\{0,1\}^*$, defined by $C(x_1x_2\cdots x_n) = C(x_1)\,C(x_2)\cdots C(x_n)$, where $C(x_1)\,C(x_2)\cdots C(x_n)$ indicates concatenation of the corresponding codewords. A code is *non-singular* if for every $x_1, x_2 \in \mathcal{X}$, such that $x_1 \neq x_2$ we have $C(x_1) \neq C(x_2)$. A code $C$ is *uniquely decodable* if its extension is non-singular. A *prefix-free code* is a code in which no codeword is a prefix of a longer codeword.

In [38], Li and Vitanyi present a sequence of prefix-free binary encodings $\{E_i\}_{i\in\mathbb{N}}$ of natural numbers. In this sequence, for every $n \in \mathbb{N}$, $E_2(n)$ has length at most $\lg(n) + 2\lg(\lg(n+1)) + \mathcal{O}(1)$, and $E_3(n)$ has length at most $\lg(n) + \lg(\lg(n+1)) + \mathcal{O}(\lg(\lg(1 + \lg(n+1)))) \leq \lg(n) + (1+\epsilon)\lg(\lg(n+1)) + \mathcal{O}(1)$, for any $\epsilon > 0$. We use the encoding function $E_2$ frequently in this thesis.

**Theorem 2.2.4.** *(Shannon's Lossless Source Coding Theorem [47]) The average length of an optimal uniquely decodable code for a random variable $X$ is between* $\mathrm{H}(X)$ *and* $\mathrm{H}(X)+1$.

## 2.3 Communication Protocols and Communication Complexity

### Classical Communication Complexity

Classical communication complexity was first introduced by Yao in 1979 [50]. In this thesis we are only interested in one-way communication protocols. In the following section we describe the model of one-way classical communication complexity, and then we define different measures of randomized communication complexity in this setting.

### One-Way Communication Protocols

Let $\mathcal{X}$ and $\mathcal{Y}$ be arbitrary finite sets, $\mathcal{Z}$ be a set which is not necessarily finite, and $T \subseteq \mathcal{X}\times\mathcal{Y}\times\mathcal{Z}$ be a relation from $\mathcal{X}\times\mathcal{Y}$ to $\mathcal{Z}$. Consider the following communication scenario.

Suppose that two parties, Alice and Bob, are given $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ as their inputs, respectively. They also have access to private random strings $r_A \in \mathcal{R}_A$ (known to Alice) and $r_B \in \mathcal{R}_B$ (known to Bob), and some public random string $r \in \mathcal{R}$ (known to both). Their goal is to find some $z \in \mathcal{Z}$ such that $(x, y, z) \in T$. They have agreed beforehand on a communication protocol but neither knows other player's input. We assume that they both have unlimited computational power, and we are only interested in the amount of communication between the two parties. In one-way communication setting, we assume that only one of the two parties is allowed to communicate with the other one. Let $\mathcal{M}$ denote the set of all possible messages. A one-way communication protocol $\Pi$ consists of two functions $f_A : \mathcal{X} \times \mathcal{R}_A \times \mathcal{R} \to \mathcal{M}$ and $f_B : \mathcal{Y} \times \mathcal{R}_B \times \mathcal{R} \times \mathcal{M} \to \mathcal{Z}$, and the following two steps:

1. Alice computes $m = f_A(x, r_A, r)$ , and sends $m$ to Bob.
2. Bob computes $z = f_B(y, r_B, r, m)$, and outputs $z$.

For $\epsilon \geq 0$, we say that the protocol $\Pi$ computes the relation $T$ with error at most $\epsilon$, if for every pair of inputs $(x, y) \in \mathcal{X} \times \mathcal{Y}$, the probability that the protocol $\Pi$ outputs a $z \in \mathcal{Z}$ such that $(x, y, z) \in T$ is at least $1 - \epsilon$.

### One-Way Classical Communication Complexity

The worst-case randomized classical communication complexity (cost) of a one-way communication protocol $\Pi$ is defined as the maximum length of a message $m$ sent by Alice in the protocol $\Pi$ for the worst-case choice of the inputs and the random strings. The average communication cost of a one-way randomized communication protocol $\Pi$ for the worst case input is defined as the maximum value (over different choices of the inputs) of the average length (over the random strings) of the messages sent in the protocol $\Pi$. Let $\mu$ be a probability distribution over $\mathcal{X} \times \mathcal{Y}$. The average communication cost of a one-way randomized communication protocol $\Pi$ for the input distribution $\mu$ is defined as the average length of the messages sent in the protocol $\Pi$ over input distribution $\mu$ and the random strings of the protocol. Communication complexity measures for relations are defined in a similar way. The $\epsilon$-error worst-case randomized one-way communication complexity of a relation $T \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ is defined as the minimum worst-case communication complexity of a one-way communication protocol which computes the relation $T$ with error at most $\epsilon$.

## Quantum Communication Complexity

In 1993, Yao extended the notion of communication complexity to the quantum setting [51]. In this thesis we are only interested in one-way LOCC protocols, and their associated measures of communication complexity.

### One-Way LOCC Protocols

LOCC (short for local operation and classical communication) protocols form an important class of quantum protocols and they are used widely in quantum computing. Next we define a general one-way LOCC protocol.

Let $\mathcal{X}$ and $\mathcal{Y}$ be arbitrary finite sets, $\mathcal{Z}$ be a set which is not necessarily finite, and $T \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation from $\mathcal{X} \times \mathcal{Y}$ to $\mathcal{Z}$. Alice and Bob get $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ as their inputs respectively, and their goal is to output some $z \in \mathcal{Z}$ such that $(x, y, z) \in T$. In an LOCC protocol, Alice and Bob are allowed to initially share some entangled quantum states. Each player is allowed to perform any quantum operation and quantum measurement on his or her own qubits and send classical messages to the other player. In particular, a one-way LOCC protocol $\Pi$ consists of the following steps:

1. Alice applies a quantum measurement, controlled by her input $x$, on her own qubits and sends the classical outcome $m$ of her measurement to Bob.

2. Bob applies a quantum operation, controlled by his input $y$ and the message $m$ he receives, on his own qubits, and finds the output $z$.

We say that the LOCC protocol $\Pi$ computes the relation $T$ with error at most $\epsilon$, if for every pair of inputs $(x, y) \in \mathcal{X} \times \mathcal{Y}$, the probability that the protocol $\Pi$ outputs a $z \in \mathcal{Z}$ such that $(x, y, z) \in T$ is at least $1 - \epsilon$.

### One-Way Quantum Communication Complexity

The worst-case quantum communication complexity (cost) of a one-way LOCC protocol $\Pi$ is the maximum length of a message $m$ sent by Alice in the protocol $\Pi$ for the worst-case inputs. The average communication cost of a one-way LOCC communication protocol $\Pi$ for the worst case input is defined as the maximum value (over different choices of the inputs) of the average length (over random outcomes of Alice's and Bob's quantum operations) of the messages sent in the protocol $\Pi$. Let $\mu$ be a probability distribution over $\mathcal{X} \times \mathcal{Y}$. The average communication cost of a one-way LOCC communication protocol $\Pi$ for the input distribution $\mu$ is defined as the average length of the messages sent in the protocol $\Pi$, over input distribution $\mu$ and the random outcomes of Alice's and Bob's quantum operations. Quantum communication complexity measures for relations are defined in a similar way. The $\epsilon$-error worst-case one-way communication complexity of a relation $T \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ is defined as the minimum worst-case quantum communication complexity of a one-way LOCC communication protocol which computes the relation $T$ with error at most $\epsilon$.

Next we define a subroutine which we use several times throughout this thesis in different protocols.

**Subroutine** $\Pi(X, \sigma)$**:**

Let $\mathcal{H}$ and $\mathcal{K}$ be finite dimensional Hilbert spaces such that $\dim(\mathcal{K}) \geq \dim(\mathcal{H})$, and $\sigma \in \mathsf{D}(\mathcal{H})$ be a quantum state in $\mathcal{H}$. Let $X \in \mathsf{Psd}(\mathcal{H})$ be a substate of $\sigma$, i.e. $0 \leq X \leq \sigma$, then $\sigma$ can be written as

$$\sigma \quad = \quad X + (\sigma - X) \quad = \quad \mathrm{Tr}(X) X_N + (1 - \mathrm{Tr}(X))(\sigma - X)_N \ ,$$

where $X_N = \frac{X}{\mathrm{Tr}(X)}$ and $(\sigma - X)_N = \frac{\sigma - X}{1 - \mathrm{Tr}(X)}$ are normalized states. Let $|\phi_X\rangle \in \mathcal{H} \otimes \mathcal{K}$ and $|\phi_{\sigma - X}\rangle \in \mathcal{H} \otimes \mathcal{K}$ be arbitrary purifications of $X_N$ and $(\sigma - X)_N$ , respectively. Then

$$|\phi_\sigma\rangle \quad = \quad [\mathrm{Tr}(X)] \ |\phi_X\rangle|1\rangle + [1 - \mathrm{Tr}(X)] \ |\phi_{\sigma - X}\rangle|0\rangle \in \mathcal{H} \otimes \mathcal{K} \otimes \mathbb{C}^2 \ ,$$

is a purification of $\sigma$. The subroutine $\Pi(X, \sigma)$ is defined as follows.

Initially, Alice and Bob share an arbitrary purification, $|\phi\rangle$, of $\sigma$ in $\mathcal{H} \otimes \mathcal{K} \otimes \mathbb{C}^2$ such that Bob's marginal state is $\sigma$ and Alice holds the rest of the state. Alice performs a unitary transformation on her part of the state which maps $|\phi\rangle$ to $|\phi_\sigma\rangle$. The existence of such a unitary operator is guaranteed by Theorem 2.2.2. Then she measures her last qubit in the standard basis. She is successful (her measurement outcome is ' 1') with probability $\mathrm{Tr}(X)$ in which case Bob's marginal state is $X_N$. If her measurement outcome is ' 0', we say the subroutine fails.

# Chapter 3

# Quantum rejection sampling

## 3.1 Generating one distribution from another

Let $P$ and $Q$ be two probability distributions on the set $\mathcal{X}$ such that $\mathrm{S}(P||Q)$ is finite. We consider one-way communication protocols from Alice to Bob. Suppose that Alice knows the description of $P$ and $Q$. Consider the task of generating a sample according to the distribution $P$ by Bob, given samples from the distribution $Q$ as their shared randomness. More precisely, let Alice and Bob share a sequence, $\{x_i : i \in \mathbb{N}\}$, of samples from distribution $Q$. Alice's goal is to send an index $J$ to Bob such that the $J$-th sample, $x_J$, is distributed according to $P$, i.e. $\Pr(x_J = x) = P(x)$ for every $x \in \mathcal{X}$. We call any such protocol a *rejection sampling protocol*.

Consider any rejection sampling protocol as described above. Let $C : \mathbb{N} \to \{0,1\}^*$ be any prefix-free binary encoding of natural numbers. Let $a_j(x)$ be the probability that the index sent by Alice is $J = j$ and the $J$-th sample is $x$, i.e.

$$a_j(x) := \Pr(J = j \ \wedge \ x_J = x) \ .$$

Then we have

$$
\begin{aligned}
\mathbb{E}\left[\, l_C(J) \,\right] &= \sum_x P(x) \mathbb{E}(\, l_C(J) \mid x_J = x) \\
&\geq \sum_x P(x) \mathrm{H}(J | \, x_J = x) \\
&= \sum_x P(x) \sum_j \frac{a_j(x)}{P(x)} \lg \left( \frac{P(x)}{a_j(x)} \right) \\
&\geq \sum_x P(x) \lg \left( \frac{P(x)}{Q(x)} \right) \quad ,
\end{aligned}
$$

where the first inequality holds by Theorem 2.2.4 and the second inequality holds since $a_j(x) \leq \Pr(x_j = x) = Q(x)$ for every $x \in \mathcal{X}$. Hence, $\mathbb{E}(\, l_C(J)\,) \geq \mathrm{S}(P||Q)$ and $\mathrm{S}(P||Q)$ is a lower bound on the communication cost of any rejection sampling protocol for distributions $P$ and $Q$. Harsha *et al.* [25] introduce a rejection sampling protocol which almost achieves this lower bound.

### 3.1.1 A classical rejection sampling protocol

Starting from $j = 1$, Alice examines the $j$-th sample, $x_j$. She either accepts it by sending the index $J = j$ to Bob using the prefix encoding function $E_2$ described in the Preliminaries chapter, or rejects it and moves on to the $(j+1)$-th sample, $x_{j+1}$. The acceptance probability function for each step is defined such that we end up with the correct probability distribution for $x_J$. For every $x \in \mathcal{X}$ and every $j \geq 1$, let $a_j(x)$ be the probability that the index sent by Alice is $J = j$ and the $j$-th sample is $x$, i.e. $a_j(x) := \Pr(J = j \ \wedge \ x_j = x)$. We define the following quantities in terms of $a_j(x)$:

$t_j(x) = \sum_{i=1}^{j} a_j(x)$: The probability that the protocol terminates with $J \leq j$ and $x_J = x$.

$s_j = \sum_{x \in \mathcal{X}} t_j(x)$ : The probability that the protocol terminates within $j$ iterations.

$r_j(x) = P(x) - t_j(x)$ : The probability requirement for $x \in \mathcal{X}$ remaining to be fulfilled at the end of the $j$-th iteration.

Note that $t_j(x) = t_{j-1}(x) + a_j(x)$, $s_j = s_{j-1} + \sum_{x \in \mathcal{X}} a_j(x)$, and $r_j(x) = r_{j-1}(x) - a_j(x)$. For every $x \in \mathcal{X}$, let $t_0(x) = 0$ (hence, $s_0 = 0$ and $r_0(x) = P(x)$). We define $a_j(x)$ (and hence $t_j(x)$, $s_j$, and $r_j(x)$), recursively.

For $j \geq 1$ and $x \in \mathcal{X}$, let

$$a_j(x) \quad = \quad \min\{r_{j-1}(x)\,,\,(1-s_{j-1})Q(x)\} \ .$$

The protocol is given more formally below.

**Gready Rejection Sampler** $(P,Q)$

**Input:** A sequence $\{x_i\}_{i\in\mathbb{N}}$ of independently drawn samples from the distribution $Q$ shared between Alice and Bob.

**A. Initialization:** For every $x \in X$, set $t_0(x) \leftarrow 0$.

**B.** For $j \leftarrow 1$ to $\infty$ do

**Iteration**$(j)$

- **a)** Alice computes $a_j(x)$, $t_j(x)$, $r_j(x)$ for every $x \in X$ and $s_j$.
- **b)** Alice examines sample $x_j$.
- **c)** With probability $\frac{a_j(x_j)}{(1-s_{j-1})Q(x_j)}$, Alice accepts the $j$-th sample, and communicates the index $j$ to Bob by sending the string $E_2(j)$.
- **d)** Bob decodes Alice's message and outputs the $j$-th sample and they stop.

The definition of $a_j(x)$ can be understood as follows. The term $r_{j-1}(x)$ in the definition ensures that the probability that $x_J = x$ never exceeds $P(x)$. Moreover, the probability of going to the $j$-th step is equal to $1-s_{j-1}$and since $\Pr(x_j = x) = Q(x)$, the probability that Alice sends the index $j$ to Bob after examining $x_j = x$ can be at most $(1-s_{j-1})Q(x)$. So with this choice of $a_j(x)$ Alice accepts the $j$-th sample with as much probability as possible under the constraint that $t_j(x) \leq P(x)$ for every $x \in \mathcal{X}$.

In this greedy algorithm in each iteration, we fill the distribution $P$ with the best possible sub-distribution of $Q$, while making sure that the resulting probability distribution never exceeds $P$. Note that since we are doing rejection sampling, we can only get sub-distributions of $Q$ in each iteration. The following claim indicates that $\Pr(x_J = x)$ is indeed equal to $P(x)$ for every $x \in \mathcal{X}$.

**Claim 3.1.1.** *[25] For every $x \in \mathcal{X}$, $\langle r_j(x) \,:\, j \in \mathbb{N}\rangle$ converges to zero.*

The next claim gives an upper bound on the communication cost of the Greedy Rejection Sampler for distributions $P$ and $Q$.

**Claim 3.1.2.** *[25]* $\mathbb{E}\,[\,l_{E_2}(J)\,] \quad \in \quad \mathrm{S}(P||Q) + 2\lg(\mathrm{S}(P||Q)+1) + \mathcal{O}(1) \ .$

## 3.2 Generating one quantum state from another

Let $\mathcal{H}$ be a finite dimensional Hilbert space and $\rho \in \mathsf{D}(\mathcal{H})$ and $\sigma \in \mathsf{D}(\mathcal{H})$ be quantum states in $\mathcal{H}$ such that $\mathrm{S}(\rho||\sigma)$ is finite. Suppose that a (classical) description of $\sigma$ and $\rho$ is known to Alice. Alice and Bob share an unlimited number of copies of an entangled state whose marginal on Bob's side is $\sigma$. The goal is for Bob to output a single copy of the state $\rho$ by engaging in a one-way LOCC protocol in which Alice sends the lone message.

A weaker version of the above task was studied by Jain *et al.* [34] which allows error. More precisely, instead of constructing $\rho$ exactly, Bob's output at the end of the protocol is allowed to be some state $\rho'$ such that the fidelity between $\rho$ and $\rho'$ is at least $1-\epsilon$, for some error parameter $\epsilon > 0$.

The protocol suggested in [34] for the above task is the following.

Let $\rho' \in \mathsf{D}(\mathcal{H})$ be a quantum state such that

$$\mathrm{F}(\rho, \rho') \quad \geq \quad 1-\epsilon \qquad \text{and} \qquad \mathrm{S}_{\max}(\rho'||\sigma) \quad \leq \quad \frac{\mathrm{S}(\rho||\sigma)+1}{\epsilon} + \lg\left(\frac{1}{1-\epsilon}\right) \enspace .$$

The existence of such a state is guaranteed by Theorem 2.2.3 (the Quantum Substate Theorem) . By definition of max-relative entropy we have

$$2^{-\mathrm{S}_{\max}(\rho'||\sigma)}\rho' \quad \leq \quad \sigma \enspace .$$

Let $\mathcal{K}$ be a finite dimensional Hilbert space such that $\dim(\mathcal{K}) \geq \dim(\mathcal{H})$. Alice and Bob initially share a sequence, $\{|\phi\rangle_i\}_{i\in\mathbb{N}}$, of a fixed purification $|\phi\rangle$ of $\sigma$ in $\mathbb{C}^2 \otimes \mathcal{K} \otimes \mathcal{H}$ such that Bob's marginal of each state is $\sigma$ and Alice holds the rest of each state. Starting from $j=1$, Alice performs the subroutine $\Pi(2^{-\mathrm{S}_{\max}(\rho'||\sigma)}\rho', \sigma)$, described in Section 2.3, on $|\phi\rangle_j$. If Alice is successful (which happens with probability $2^{-\mathrm{S}_{\max}(\rho'||\sigma)}$), she sends the index $j$ to Bob, then Bob outputs his part of the $j$-th state and the protocol terminates. Otherwise, Alice moves on to $|\phi\rangle_{j+1}$.

Let $J$ be the random variable corresponding to the index sent by Alice to Bob. At the end of the protocol Bob's marginal of the $J$-th state is $\rho'$. Furthermore, in expectation, Alice succeeds in $2^{\mathrm{S}_{\max}(\rho'||\sigma)}$ iterations, hence the average communication cost of this protocol is at most $\mathrm{S}_{\max}(\rho'||\sigma)$, which is bounded by $\dfrac{\mathrm{S}(\rho||\sigma)+1}{\epsilon} + \lg\left(\dfrac{1}{1-\epsilon}\right)$. Although this is an interesting result, it is not useful for our purposes. More specifically, in applications such as message compression and proving multi-round direct sum results, we do not want to introduce too much error in compression of messages in each round, and since the expected communication cost of this protocol is proportional to $1/\epsilon$, compression of multi-round protocols with sufficiently small error is not efficiently achievable using this protocol. Furthermore, we are interested in applications such as lossless quantum source coding and exact remote state preparation.

### 3.2.1 Quantum rejection sampling protocols

In this section we introduce a special class of LOCC protocols for task (ii) described in Section 1.1, which we refer to as *quantum rejection sampling* protocols. Our definition of a quantum rejection sampling protocol is a natural extension of classical rejection sampling protocols given in [25].

Let $\mathcal{K}$ be a finite dimensional Hilbert space such that $\dim(\mathcal{K}) \geq \dim(\mathcal{H})$. Initially, Alice and Bob share a sequence, $\{|\phi\rangle_i\}_{i\in\mathbb{N}}$, of a fixed purification, $|\phi\rangle$, of $\sigma$ in $\mathbb{C}^2 \otimes \mathcal{K} \otimes \mathcal{H}$ such that Bob's marginal of each state is $\sigma$ and Alice holds the rest of each state. In a one-way LOCC protocol, Alice sends an index $J$ to Bob such that the marginal of the $J$-th state on Bob's side, on average for different choices of $J$, is exactly $\rho$. At the end, Bob outputs his portion of the $J$-th state. We refer to any such protocol for task (ii) as a *quantum rejection sampling* protocol.

We use the following definitions for quantum rejection sampling protocols for every $j \geq 1$.

We denote by $X_j$ the contribution of the $j$-th iteration to Bob's output state. $R_j$ denotes the substate of $\rho$ which still remains to be prepared on Bob's side after $j$ iterations, i.e. $R_j = \rho - \displaystyle\sum_{i=1}^{j} X_i$. Let $s_j$ denote the probability that the protocol terminates within $j$ iterations. We also define $R_0 = \rho$, and $s_0 = 0$.

### 3.2.2 Greedy Quantum Rejection Sampler

In this section we design a quantum rejection sampling protocol which is a natural extension of the classical Greedy Rejection Sampler to achieve task (ii).

Informally, the idea is that in the $j$-th iteration, Alice helps Bob output a simultaneous substate of $\rho$ and $\sigma$, $X_j$, such that it is guaranteed that

$$\sum_{j=1}^{\infty} X_j \quad = \quad \rho \ .$$

Let $R_0 = \rho$, and $s_0 = 0$. Consider the following semidefinite program for every $j \geq 1$.

$$
\begin{array}{lllll}
(P_j): & \text{maximize}: & \mathrm{Tr}(X) & & \\
 & \text{subject to}: & X & \leq & R_{j-1} \\
 & & X & \leq & (1 - s_{j-1})\sigma \\
 & & X & \geq & 0
\end{array}
$$

Let $X_j$ be an optimal solution of $P_j$ (since for every $j$ the trace function is continuous over the feasible region and the feasible region is non-empty and compact, by Theorem 2.1.3, an optimal solution exists). We define $s_j$ and $R_j$ recursively as:

$$
\begin{aligned}
R_j &= R_{j-1} - X_j \ , \\
s_j &= s_{j-1} + \mathrm{Tr}(X_j) \ .
\end{aligned}
$$

The $j$-th iteration of the protocol is defined as follows. Let $\tilde{X}_j = X_j/(1 - s_{j-1})$. Note that $\tilde{X}_j$ is a substate of $\sigma$. Alice uses the subroutine $\Pi(\tilde{X}_j, \sigma)$, described in Section 2.3, on $|\phi\rangle_j$, in order to prepare $\tilde{X}_j$ on Bob's side . Given Alice's success (which happens with probability $\mathrm{Tr}(\tilde{X}_j)$), Bob's marginal of the $j$-th state is $\dfrac{\tilde{X}_j}{\mathrm{Tr}(\tilde{X}_j)}$. Alice sends the index $J = j$ to Bob using the prefix encoding function $E_2$ described in the Preliminaries section, then Bob outputs his marginal of the $j$-th state and the protocol terminates. If Alice fails, she moves on to the $(j + 1)$-th iteration. The protocol is given more formally below.

**Greedy Quantum Rejection Sampler ($\rho$, $\sigma$)**

**Input:** A sequence $\{|\phi\rangle_i\}_{i\in\mathbb{N}}$ of a fixed purification of $|\phi\rangle \in \mathbb{C}^2 \otimes \mathcal{K} \otimes \mathcal{H}$ of $\sigma \in \mathsf{D}(\mathcal{H})$ such that Bob's marginal of each state is $\sigma$ and Alice holds the rest of each state.

**A. Initialization:** Set $R_0 \leftarrow \rho$, $s_0 \leftarrow 0$.

**B.** For $j \leftarrow 1$ to $\infty$ do

**Iteration($j$)**

**a)** Alice computes $X_j$, $R_j$ and $s_j$.

**b)** Alice performs the subroutine $\Pi(\tilde{X}_j, \sigma)$ on $|\phi\rangle_j$, where $\tilde{X}_j = \frac{X_j}{1-s_{j-1}}$.

**c)** If Alice is successful in the subroutine, she communicates the index $j$ to Bob by sending the string $E_2(j)$.

**d)** Bob decodes Alice's message and outputs the $j$-th sample and they stop.

The definition of $X_j$ as the optimal solution of the semidefinite program $(P_j)$ can be understood as follows. The first constraint, $X \leq R_{j-1}$, ensures that $\sum_{i=1}^{j} X_i$ remains a substate of $\rho$ for every $j \in \mathbb{N}$ and the output of the protocol never "exceeds" $\rho$. Moreover, since Alice uses the protocol $\Pi$ to prepare $\tilde{X_j}$ on Bob's side, we need $\tilde{X_j}$ to be a substate of $\sigma$. The second constraint, $X \leq (1 - s_{j-1})\sigma$, ensures that $\tilde{X_j} \leq \sigma$.

## 3.3 Analysis of the Greedy Quantum Rejection Sampler

Next, we prove that the Greedy Rejection Sampling protocol terminates with probability 1, and the state output by Bob is indeed $\rho$. In other words, we show that $s_j$, the probability that the protocol terminates within $j$ iterations, converges to 1 as $j$ goes to infinity, and the series $\sum_{i=1}^{\infty} X_i$ converges to $\rho$. Note that since $R_j = \rho - \sum_{i=1}^{j} X_i$, it is sufficient to show that the sequence $\{R_j\}_{j\in\mathbb{N}}$ converges to zero.

**Theorem 3.3.1.** *The sequence $\{R_j\}_{j\in\mathbb{N}}$ converges to zero.*

**Proof:** First note that $R_j$ is a positive semidefinite operator for every $j \in \mathbb{N}$. So in order to show that the sequence $\{R_j\}_{j\in\mathbb{N}}$ converges to zero, it is sufficient to show that the real sequence $\{r_j\}_{j\in\mathbb{N}}$, where $r_j = \mathrm{Tr}(R_j)$, converges to zero. We show that this sequence is strictly decreasing and bounded below by zero, hence convergent to some $r \geq 0$.

Let $S = \mathrm{S}_{\max}(\rho||\sigma)$. Fix $i \in \mathbb{N}$. If $r_i = 0$ then for every $j \geq i$, $r_j = 0$ and we are done. Otherwise, we show that $r_{i+1} \leq r_i - \frac{r_i^2}{2^S} < r_i$.

In the $(i + 1)$-th iteration, $X_{i+1}$ is an optimal solution of the following semidefinite program.

$$\begin{array}{llrcl} (P_{i+1}) : & \text{maximize :} & \mathrm{Tr}(X) & & \\ & \text{subject to :} & X & \leq & R_i \\ & & X & \leq & (1 - s_i)\sigma \\ & & X & \geq & 0 \end{array}$$

Consider $X = \frac{r_i}{2^S} R_i$. It is straightforward to see that $X$ is a feasible solution of $(P_{i+1})$. The first constraint is satisfied since $\frac{r_i}{2^S} \leq 1$, and we have

$$X \quad \leq \quad \frac{r_i}{2^S}\rho \quad \leq \quad r_i \sigma \quad = \quad (1-s_i)\sigma \ ,$$

where the second inequality follows from the definition of max-relative entropy of $\rho$ and $\sigma$. Hence the second constraint is also satisfied. So

$$r_{i+1} \quad = \quad r_i - \mathrm{Tr}(X_{i+1}) \quad \leq \quad r_i - \frac{r_i^2}{2^S} \ . \tag{3.3.1}$$

Now let $r = \lim_{j\to\infty} r_j$. Towards contradiction, suppose that $r > 0$. Let $\epsilon = \frac{r^2}{2^{S+1}} > 0$. Then by definition, there exists some $k \in \mathbb{N}$ such that for every $j \geq k$ we have

$$0 \quad \leq \quad |r_j - r| \quad = \quad r_j - r \quad \leq \quad \epsilon \ , \tag{3.3.2}$$

or equivalently $r \leq r_j \leq r + \epsilon$. From Equation 3.3.1, we have $r_{k+1} \leq r_k - \frac{r_k^2}{2^S}$. Note that by Equation 3.3.2, $r_k - \frac{r_k^2}{2^S} \leq r_k - \frac{r^2}{2^S} \leq r + \epsilon - \frac{r^2}{2^S} = r - \frac{r^2}{2^{S+1}} < r$, which is in contradiction with Equation 3.3.2 for $j = k+1$. So $\lim_{j\to\infty} r_j = 0$. ■

### 3.3.1 Special Case 1: pure state case

In this section we consider a special case of the problem where $\rho$ is a pure state and $\sigma$ is an arbitrary mixed state such that $\mathrm{S}(\rho||\sigma)$ is finite and we analyse the Greedy Quantum Rejection Sampler. We characterize the expected communication cost of the protocol and show that it is at most $\mathrm{S}_{\max}(\rho||\sigma) + 2\lg(\mathrm{S}_{\max}(\rho||\sigma)+1) + \mathcal{O}(1)$. We also show that our protocol is optimal in the sense that any rejection sampling protocol described in Section 3.2.1 requires at least $\mathrm{S}_{\max}(\rho||\sigma)$ bits of communication in the case where $\rho$ is pure.

Let $\rho = |\psi\rangle\langle\psi| \in \mathsf{D}(\mathcal{H})$ be a pure state and $\sigma \in \mathsf{D}(\mathcal{H})$ such that $\mathrm{S}(\rho||\sigma)$ is finite, hence $|\psi\rangle \in \mathrm{supp}(\sigma)$. For convenience and without loss of generality, let $\mathcal{H}$ be the support of $\sigma$ so that $\sigma$ is a full rank operator on $\mathcal{H}$.

Let $\alpha = \frac{1}{\langle\psi|\sigma^{-1}|\psi\rangle}$. Using induction, for $j = 1, 2, \ldots,$ we show that

$$X_j = \alpha(1-\alpha)^{j-1}\,|\psi\rangle\langle\psi| \ , \ s_j = 1-(1-\alpha)^j \ , \ R_j = (1-\alpha)^j|\psi\rangle\langle\psi| \ .$$

In the first iteration $X_1$ is an optimal solution of the following semidefinite program.

$$\begin{array}{llccl} (P_1): & \text{maximize}: & \mathrm{Tr}(X) & & \\ & \text{subject to}: & X & \leq & |\psi\rangle\langle\psi| \\ & & X & \leq & \sigma \\ & & X & \geq & 0 \end{array}$$

By the first and the last constraint, $0 \leq X_1 \leq |\psi\rangle\langle\psi|$. So $X_1$ is necessarily of the form $X_1 = x_1|\psi\rangle\langle\psi|$, for some $x_1 \in [0,1]$. By the second constraint we have

$$x_1\, \sigma^{-1/2}|\psi\rangle\langle\psi|\sigma^{-1/2} \quad \leq \quad \mathbb{1} \ . \tag{3.3.3}$$

Since $\sigma^{-1/2}|\psi\rangle\langle\psi|\sigma^{-1/2}$ has rank one, inequality 3.3.3 is equivalent to $x_1 \langle\psi|\sigma^{-1}|\psi\rangle \leq 1$. Hence, $x_1 \leq \dfrac{1}{\langle\psi|\sigma^{-1}|\psi\rangle}$. Since $\sigma^{-1} \geq \mathbb{1}$, we have $\langle\psi|\sigma^{-1}|\psi\rangle \geq 1$. So $X_1 = \alpha|\psi\rangle\langle\psi|$, $s_1 = \alpha$, and $R_1 = (1-\alpha)|\psi\rangle\langle\psi|$.

Note that since $|\psi\rangle \in \mathrm{supp}(\sigma)$, the max-relative entropy of $\rho$ and $\sigma$, defined as

$$\mathrm{S}_{\max}(|\psi\rangle\langle\psi|||\sigma) = \min\left\{\lambda \,:\, 2^{-\lambda}|\psi\rangle\langle\psi| \leq \sigma\right\} \ ,$$

is equal to $-\lg(\alpha)$.

Suppose that $s_{j-1} = 1-(1-\alpha)^{j-1}$, and $R_{j-1} = (1-\alpha)^{j-1}|\psi\rangle\langle\psi|$. Then in the $j$-th iteration $X_j$ is the optimal solution of the following semidefinite program.

$$\begin{array}{llccl} (P_j): & \text{maximize}: & \mathrm{Tr}(X) & & \\ & \text{subject to}: & X & \leq & (1-\alpha)^{j-1}|\psi\rangle\langle\psi| \\ & & X & \leq & (1-\alpha)^{j-1}\sigma \\ & & X & \geq & 0 \end{array}$$

Using a similar argument as in the first iteration it is straightforward to show that

$$X_j \quad = \quad \alpha(1-\alpha)^{j-1}|\psi\rangle\langle\psi| \ .$$

Hence,

$$s_j = s_{j-1} + \alpha(1-\alpha)^{j-1} = 1-(1-\alpha)^{j-1} + \alpha(1-\alpha)^{j-1} = 1-(1-\alpha)^j \ ,$$

and

$$R_j \quad = \quad R_{j-1} - X_j \quad = \quad \left[(1-\alpha)^{j-1} - \alpha(1-\alpha)^{j-1}\right] |\psi\rangle\langle\psi| \quad = \quad (1-\alpha)^j \,|\psi\rangle\langle\psi| \ .$$

As we proved in Theorem 3.3.1, the protocol terminates with probability one. Next we show that the protocol outputs $\rho$ as required.

**Claim 3.3.2.** *Given that the protocol terminates, Bob's output state is $\rho = |\psi\rangle\langle\psi|$.*

**Proof:** We need to show that Bob's marginal of the $J$-th state is $\rho = |\psi\rangle\langle\psi|$ in expectation over $J$. For every $j \geq 1$, the probability that the protocol terminates with $J = j$ is

$$\mathrm{Tr}(X_j) \quad = \quad (1-\alpha)^{j-1}\alpha$$

and Bob's state given $J = j$ is $|\psi\rangle\langle\psi|$. Hence Bob's output state will be

$$\sum_{j=1}^{\infty} X_j \quad = \quad \sum_{j=1}^{\infty}(1-\alpha)^{j-1}\alpha|\psi\rangle\langle\psi| \quad = \quad |\psi\rangle\langle\psi| \ .$$

■

Next we characterize the expected communication cost of the protocol and prove the optimality of the protocol in this special case. Before that, we prove the following lemma.

**Lemma 3.3.3.** *Let $I$ be a random variable with values in $\mathbb{N}$. Then we have $\mathbb{E}\left[I\right] = \sum_{i\geq 1} \Pr(I \geq i)$ .*

**Proof:** We have

$$\begin{aligned}
\mathbb{E}\left[I\right] \quad &= \quad \sum_{i\geq 1} \Pr(I = i) i \\
&= \quad \sum_{i\geq 1} \left(\Pr(I \geq i) - \Pr(I \geq i+1)\right) i \\
&= \quad \sum_{i\geq 1} \Pr(I \geq i)(i - (i-1)) \\
&= \quad \sum_{i\geq 1} \Pr(I \geq i) \ .
\end{aligned}$$

■

**Claim 3.3.4.** $\mathbb{E}\left[\, l_{E_2}(J)\,\right] \quad \in \quad \mathrm{S}_{\max}(\rho||\sigma) + 2\lg(\mathrm{S}_{\max}(\rho||\sigma) + 1) + \mathcal{O}(1)$ .

**Proof:** We show that

$$\mathbb{E}\left[\lg(J)\right] \quad \leq \quad \mathrm{S}_{\max}(\rho||\sigma) \ . \tag{3.3.4}$$

Then for some constants $c, c' \in \mathbb{R}$ we have

$$\begin{aligned}
\mathbb{E}\left[\, l_{E_2}(J) \,\right] &= \mathbb{E}\left[\lg(J) + 2\lg(\lg(J+1)) + c\right] \\
&= \mathbb{E}\left[\lg(J)\right] + 2\mathbb{E}\left[\lg(\lg(J+1))\right] + c \\
&\leq \mathbb{E}\left[\lg(J)\right] + 2\lg(\mathbb{E}\left[\lg(J+1)\right]) + c \qquad \text{(By Jensen's inequality)} \\
&\leq \mathbb{E}\left[\lg(J)\right] + 2\lg(\mathbb{E}\left[\lg(J)\right] + 1) + c' \\
&\leq \mathrm{S}_{\max}(\rho||\sigma) + 2\lg(\mathrm{S}_{\max}(\rho||\sigma) + 1) + c' \ .
\end{aligned}$$

Now we return to Inequality 3.3.4 . We have

$$\begin{aligned}
\mathbb{E}\left[\lg(J)\right] &\leq \lg\left(\mathbb{E}[J]\right) \qquad \text{(By Jensen's inequality)} \\
&= \lg\left(\sum_{j=0}^{\infty} \Pr(J > j)\right) \qquad \text{(By Lemma 3.3.3 )} \\
&= \lg\left(\sum_{j=0}^{\infty} (1 - s_j)\right) \\
&= \lg\left(\sum_{j=0}^{\infty} (1 - \alpha)^j\right) \\
&= \lg\left(\alpha^{-1}\right) \\
&= \mathrm{S}_{\max}(\rho||\sigma) \ .
\end{aligned}$$

■

**Claim 3.3.5.** *Let $\mathcal{H}$ be a finite dimensional Hilbert space and $\rho = |\psi\rangle\langle\psi| \in \mathsf{D}(\mathcal{H})$ be a pure state and $\sigma \in \mathsf{D}(\mathcal{H})$ such that $\mathrm{S}(\rho||\sigma)$ is finite. For any quantum rejection sampling protocol described in Section 3.2.1 , and any prefix-free binary encoding of natural numbers, $C : \mathbb{N} \to \{0,1\}^*$, the expected communication for the pair $(\rho, \sigma)$ is bounded as*

$$\mathbb{E}\left[\, l_C(J) \,\right] \quad \geq \quad \mathrm{S}_{\max}(\rho||\sigma) \ .$$

**Proof:** Let $\Gamma$ be an arbitrary quantum rejection sampling protocol. Since $|\psi\rangle \in \mathrm{supp}(\sigma)$, without loss of generality, we assume that $\mathcal{H} = \mathrm{supp}(\sigma)$. Let $X_j$ be the contribution of the $j$-th iteration in Bob's state. Bob's output state is a summation of $X_j$s, thus for every $j \geq 1$, $X_j$ is necessarily a substate of $\rho$. So $X_j$ is of the form $X_j = x_j|\psi\rangle\langle\psi|$ for some $x_j \in [0,1]$. Furthermore, since we are doing rejection sampling given purifications of $\sigma$, $\tilde{X}_j = \dfrac{X_j}{1-s_{j-1}}$ should be a substate of $\sigma$ ( Note that $1-s_{j-1}$ is the probability of reaching the $j$-th iteration). Hence for $j = 1,\, 2,\, \ldots$ we have

$$\Pr(J=j) \quad = \quad \mathrm{Tr}(X_j) \quad = \quad x_j \quad \leq \quad \alpha \ .$$

Thus,

$$\begin{aligned}
\mathbb{E}\left[\, l_C(J) \,\right] \quad &\geq \quad \mathrm{H}(J) \\
&= \quad -\sum_{j=1}^{\infty} \Pr(J=j) \lg\left(\Pr(J=j)\right) \\
&\geq \quad -\lg(\alpha) \\
&= \quad \mathrm{S}_{\max}(\rho||\sigma) \ ,
\end{aligned}$$

where the first inequality follows by Theorem 2.2.4. ■

In Section 3.1, we showed that the average communication cost of constructing a probability distribution $P$ given samples of a distribution $Q$ using any rejection sampling protocol is at least the relative entropy of $P$ and $Q$. One would expect that the same bound holds in the quantum case. The above argument shows that in the quantum setting, unlike the classical case, the expected cost of a rejection sampling protocol for constructing a state $\rho$ using purifications of a state $\sigma$ is at least max-relative entropy of $\rho$ and $\sigma$. The following example shows that the gap between relative entropy and max-relative entropy can be arbitrarily large, even when $\rho$ and $\sigma$ are single qubit states.

Let $\rho = |\psi\rangle\langle\psi|$, where $|\psi\rangle = \sqrt{a}|0\rangle + \sqrt{1-a}|1\rangle$ and $\sigma = b|0\rangle\langle 0| + (1-b)|1\rangle\langle 1|$ be two single qubit states, then we have

$$\begin{aligned}
\mathrm{S}(\rho||\sigma) \quad &= \quad \mathrm{Tr}(\rho \lg(\rho)) - \mathrm{Tr}(\rho \lg(\sigma)) \\
&= \quad 0 - a\lg(b) - (1-a)\lg(1-b) \ ,
\end{aligned}$$

and using a similar argument as in Section 3.3.1 it can be shown that

$$\begin{aligned}
\mathrm{S}_{\max}(\rho||\sigma) \quad &= \quad \lg(\langle\psi|\sigma^{-1}|\psi\rangle) \\
&= \quad \lg\left(\frac{a}{b} + \frac{1-a}{1-b}\right) \ .
\end{aligned}$$

Now it is straightforward to see that for small enough $a$ and $b$ such that $0 \leq b \ll a \ll 1$, the ratio $\dfrac{\mathrm{S}_{\max}(\rho||\sigma)}{\mathrm{S}(\rho||\sigma)}$ can be arbitrarily large.

### 3.3.2 Special Case 2 : two dimensional Hilbert space

In this section we consider another special case of the problem in which $\rho$ and $\sigma$ are single qubit states. We analyze the protocol and show that the expected communication cost of the Greedy Quantum Rejection Sampler is bounded above by $\mathrm{S}_{\max}(\rho||\sigma)+2\lg(\mathrm{S}_{\max}(\rho||\sigma)+1)+\mathcal{O}(1)$ in this case too.

In order to analyze the protocol we introduce a one-to-one correspondence between 2 by 2 Hermitian operators of non-negative trace and balls in $\mathbb{R}^3$, which is a simplified version of a correspondence introduced by Deconinck and Terhal [21]. This correspondence gives us an interesting geometrical interpretation of the Löwner order on qubit states.

As stated in Lemma 2.1.1, it is straightforward to see that any 2 by 2 Hermitian operator $A$ can be written as $A = t\mathbb{1} + x\mathrm{X} + y\mathrm{Y} + z\mathrm{Z}$ for some real numbers $t$, $x$, $y$, and $z$. To the operator $A = t\mathbb{1} + x\mathrm{X} + y\mathrm{Y} + z\mathrm{Z}$, with $t \geq 0$, we associate the ball of radius $t$ about the point $(x, y, z) \in \mathbb{R}^3$. In this picture, as we show in this section, $A \geq 0$ if and only if $t + z \geq 0$, $t - z \geq 0$ and $t^2 \geq x^2 + y^2 + z^2$ which corresponds to a second-order cone. Hence, in the two dimensional case, the constraints of the semidefinite program defining $X_j$ in every iteration correspond to second-order cones and the semidefinite program becomes a second-order cone programming problem. For more information about second-order cone programming please refer to Ref [17]. Next we prove some facts regarding this one-to-one correspondence.

**Claim 3.3.6.** *Any single-qubit pure state corresponds to a ball passing through the origin, of radius* $\frac{1}{2}$ *and vice versa.*

**Proof:** First note that any quantum state $\rho$ is of the form $\frac{1}{2}\mathbb{1} + x\mathrm{X} + y\mathrm{Y} + z\mathrm{Z}$ for some real numbers $x$, $y$, and $z$ which corresponds to a ball of radius $\frac{1}{2}$ centred at $(x, y, z) \in \mathbb{R}^3$.

Furthermore, the quantum state $\rho$ is pure if and only if $\rho^2 = \rho$, i.e.

$$\begin{aligned}
\rho^2 &= \left(\frac{1}{4}\mathbb{1} + \frac{x}{2}\mathrm{X} + \frac{y}{2}\mathrm{Y} + \frac{z}{2}\mathrm{Z}\right) \\
&+ \left(\frac{x}{2}\mathrm{X} + x^2\mathbb{1} + ixy\mathrm{Z} - ixz\mathrm{Y}\right) \\
&+ \left(\frac{y}{2}\mathrm{Y} - ixy\mathrm{Z} + y^2\mathbb{1} + iyz\mathrm{X}\right) \\
&+ \left(\frac{z}{2}\mathrm{Z} + ixz\mathrm{Y} - iyz\mathrm{X} + z^2\mathbb{1}\right) \\
&= \left(\frac{1}{4} + x^2 + y^2 + z^2\right)\mathbb{1} + x\mathrm{X} + y\mathrm{Y} + z\mathrm{Z} \\
&= \rho \ ,
\end{aligned}$$

which is equivalent to $x^2 + y^2 + z^2 = \dfrac{1}{4}$. ■

**Claim 3.3.7.** *A 2 by 2 Hermitian operator is positive semidefinite if and only if the corresponding ball contains the origin.*

**Proof:** We show that any (mixed) quantum state of one qubit corresponds to a ball of radius $\frac{1}{2}$ containing the origin and vice versa. Since normalization corresponds to an isotropic scaling of the ball about the origin, the claim follows.

Any mixed state $\rho$ can be written as a convex combination of a set of pure states each of which corresponds to a ball of radius $\frac{1}{2}$ passing through the origin. The center of the ball $B$ corresponding to $\rho$ is a convex combination of the centers of these balls, hence it belongs to the convex hull of these points which is contained in the ball of radius $\frac{1}{2}$ about the origin. Since the radius of $B$ is $\frac{1}{2}$, it contains the origin.

Conversely, let $B$ be a ball of radius $\frac{1}{2}$ containing the origin centered at the point $c$. Hence $c$ is located somewhere inside the ball of radius $\frac{1}{2}$ about the origin. Let $t_1$ and $t_2$ be the end points of an arbitrary chord of $B$ passing through $c$. Clearly, $c$ can be written as $\alpha t_1 + (1-\alpha)t_2$ for some $\alpha \in [0,1]$. Then $B$ corresponds to the mixed state $\alpha|\psi_1\rangle\langle\psi_1| + (1-\alpha)|\psi_2\rangle\langle\psi_2|$, where $|\psi_1\rangle\langle\psi_1|$ and $|\psi_2\rangle\langle\psi_2|$ are the pure states corresponding to the balls of radius $\frac{1}{2}$ centered at $t_1$ and $t_2$, respectively. ■

**Corollary 3.3.8.** *Let $A_1$ and $A_2$ be 2 by 2 Hermitian operators with non-negative trace, then $A_1 \geq A_2$ if and only if the ball corresponding to $A_1$ contains that of $A_2$.*

**Proof:** Let $A_1$ and $A_2$ be Hermitian operators with non-negative trace. Let $B_1$ and $B_2$ be the balls corresponding to $A_1$ and $A_2$ of radius $t_1$ and $t_2$, centred at the points $c_1$ and $c_2$, respectively. Also let $B$ be the ball corresponding to the operator $A_1 - A_2$.

The following implications hold.

$A_1 \geq A_2 \iff$ The ball corresponding to $A_1 - A_2$ contains the origin $\iff$ The distance between $c_1$ and $c_2$ is smaller than $t_1 - t_2 \iff B_1$ contains $B_2$. Note that $t_1 - t_2$ is equal to the radius of $B$ and the distance between $c_1$ and $c_2$ is equal to the distance between the origin and the centre of $B$. ■

We define the minimum of two single qubit states $\rho$ and $\sigma$, denoted by $\min(\rho, \sigma)$, as an optimal solution of the semidefinite program $P_{\min}$, defined as follows.

$$\begin{array}{rlcl} \text{maximize :} & \mathrm{Tr}(X) & & \\ \text{subject to :} & X & \leq & \rho \\ & X & \leq & \sigma \\ & X & \geq & 0 \end{array}$$

Since the trace function is continuous over the feasible region and the feasible region is non-empty and compact, by Theorem 2.1.3, the optimal value is achieved. Furthermore, as we will show next, the optimal solution is unique, so $\min(\rho, \sigma)$ is well-defined.

Using the correspondence introduced in this section, it is possible to characterize the minimum of two single qubit states.

Let $\rho_1$ and $\rho_2$ be two single qubit states. Let $\mathcal{B}_1$ and $\mathcal{B}_2$ of radius $\frac{1}{2}$ be the balls corresponding to $\rho_1$ and $\rho_2$ centred at the points $c_1 = (x_1, y_1, z_1)$ and $c_2 = (x_2, y_2, z_2)$, respectively. Let $X$ be any feasible solution of the semidefinite program $P_{\min}$. Since $X$ is positive semidefinite, it corresponds to ball $\mathcal{B}$ in $\mathbb{R}^3$ which contains the origin. Furthermore, $X$ is a substate of both $\rho$ and $\sigma$, hence $\mathcal{B}$ is contained in both $\mathcal{B}_1$ and $\mathcal{B}_2$. Note that since both $\rho_1$ and $\rho_2$ are positive semidefinite, the intersection of $\mathcal{B}_1$ and $\mathcal{B}_2$ is non-empty and contains the origin. Now it is clear that $\min(\rho_1, \rho_2)$ corresponds to the biggest ball contained in the intersection of $\mathcal{B}_1$ and $\mathcal{B}_2$ which contains the origin.

Consider $\mathcal{B}^*$ the biggest ball in the intersection of $\mathcal{B}_1$ and $\mathcal{B}_2$. Since both $\mathcal{B}_1$ and $\mathcal{B}_2$ have the same radius the center of $\mathcal{B}^*$ is located in the middle of $c_1$ and $c_2$, i.e. at the point

$$\left(\frac{x_1+x_2}{2}, \frac{y_1+y_2}{2}, \frac{z_1+z_2}{2}\right) ,$$

and the radius of $\mathcal{B}^*$ is equal to

$$\frac{1}{2} - \frac{1}{2}\sqrt{(x_1-x_2)^2+(y_1-y_2)^2+(z_1-z_2)^2} .$$

So $\mathcal{B}^*$ corresponds to the following operator

$$\left(\frac{1}{2} - \frac{1}{2}\sqrt{(x_1-x_2)^2+(y_1-y_2)^2+(z_1-z_2)^2}\right)\mathbb{1}+\left(\frac{x_1+x_2}{2}\right)\mathrm{X}+\left(\frac{y_1+y_2}{2}\right)\mathrm{Y}+\left(\frac{z_1+z_2}{2}\right)\mathrm{Z} ,$$

which can be written as

$$\begin{aligned}&\left(\frac{1}{2}\right)\mathbb{1}+\left(\frac{x_1+x_2}{2}\right)\mathrm{X}+\left(\frac{y_1+y_2}{2}\right)\mathrm{Y}+\left(\frac{z_1+z_2}{2}\right)\mathrm{Z}-\frac{1}{2}\left(\sqrt{(x_1-x_2)^2+(y_1-y_2)^2+(z_1-z_2)^2}\right)\mathbb{1}\\ &= \quad \frac{\rho_1+\rho_2}{2} - \frac{|\rho_1-\rho_2|}{2} ,\end{aligned}$$

as we have

$$\begin{aligned}(\rho_1-\rho_2)^2 &= ((x_1-x_2)\mathrm{X}+(y_1-y_2)\mathrm{Y}+(z_1-z_2)\mathrm{Z})^2\\ &= \left((x_1-x_2)^2+(y_1-y_2)^2+(z_1-z_2)^2\right)\mathbb{1} .\end{aligned}$$

So if $\frac{\rho_1+\rho_2}{2} - \frac{|\rho_1-\rho_2|}{2}$ is a positive semidefinite operator, then

$$\min(\rho_1,\rho_2) \quad = \quad \frac{\rho_1+\rho_2}{2} - \frac{|\rho_1-\rho_2|}{2} .$$

Otherwise, $\mathcal{B}^*$ does not contain the origin, and it is not difficult to see that the biggest ball in the intersection of $\mathcal{B}_1$ and $\mathcal{B}_2$ which contains the origin, is a unique one passing through the origin. The sketch of the proof is as follows. Let $\mathcal{B}$ be the biggest ball in the intersection of $\mathcal{B}_1$ and $\mathcal{B}_2$ which contains the origin, and let $c$ denote the centre of $\mathcal{B}$ and $r$ be the radius of $\mathcal{B}$. First note that $\mathcal{B}$ is tangential to at least one of the two balls $\mathcal{B}_1$ and $\mathcal{B}_2$, since otherwise there exists $\epsilon > 0$ such that the ball of radius $r+\epsilon$ centred at $c$ contains the origin and belongs to the intersection of $\mathcal{B}_1$ and $\mathcal{B}_2$. Moreover, $\mathcal{B}$ is tangential to both $\mathcal{B}_1$ and $\mathcal{B}_2$. In order to prove this, towards contradiction suppose that $\mathcal{B}$ is tangential to $\mathcal{B}_1$ but not to $\mathcal{B}_2$. Let $d$ be the point in which $\mathcal{B}$ is tangential to $\mathcal{B}_1$. Then there exists some $\epsilon' > 0$ such that the ball centred at $d + (1+\epsilon')(c-d)$ of radius $(1+\epsilon')r$ contains $\mathcal{B}$ and belongs to the intersection of $\mathcal{B}_1$ and $\mathcal{B}_2$, which is a contradiction. Now since $\mathcal{B}$ is tangential to both $\mathcal{B}_1$ and $\mathcal{B}_2$, $c$ has to be of the same distance from the circumference of $\mathcal{B}_1$ and $\mathcal{B}_2$, so it belongs to the plane bisecting the line segment connecting $c_1$ and $c_2$ which is perpendicular to it. Also, note that a ball which is inside $\mathcal{B}_1$ and $\mathcal{B}_2$, tangential to both and centred at a point on this perpendicular bisecting plane is completely characterized by its centre and as the centre moves away from the middle point of $c_1$ and $c_2$, its radius decreases. The distance of $c$ from the origin minus the distance of $c$ from the circumference of the two balls is a continuous function of the position of $c$ on the plane. Note that by our assumption this function is positive in $(c_1+c_2)/2$. So the biggest ball in the intersection of $\mathcal{B}_1$ and $\mathcal{B}_2$ which contains the origin is characterized by a point $c$ on this plane in which the above function is equal to zero, i.e., the origin is on the circumference of $\mathcal{B}$. So in this case $\min(\rho_1, \rho_2)$ is a pure state. Note that as long as $\rho_1$ and $\rho_2$ are positive semidefinite operators of the same trace the above argument is valid.

Now we return to the Greedy Quantum Rejection Sampler protocol. Note that $X_j$, the contribution of the $j$-th iteration to Bob's output state, is defined as $X_j = \min\left(R_{j-1}, \mathrm{Tr}(R_{j-1})\sigma\right)$. Next we show an interesting fact which enables us to bound the expected communication cost of the protocol in the case where $\rho$ and $\sigma$ are single qubit states.

**Claim 3.3.9.** *$R_1$, the substate of $\rho$ which still remains to be constructed on Bob's side after the first iteration is always a rank one operator, when $\rho$ and $\sigma$ are single qubit states.*

**Proof:** Let $\mathcal{B}$ be the ball corresponding to $\rho$ centered at the point $c$. Also, let $\mathcal{B}'$ be the ball corresponding to $X_1 = \min(\rho, \sigma)$ centered at $c'$. Whether $\min(\rho, \sigma)$ is equal to $\dfrac{\rho+\sigma}{2} - \dfrac{|\rho-\sigma|}{2}$ or a rank one operator, in both cases $\mathcal{B}'$ is a ball inside $\mathcal{B}$ which is tangent to it at some point $d$, and the radius of $\mathcal{B}$ ending in $d$ passes through $c'$. So the ball corresponding to $R_1 = \rho - X_1$ is a ball centered at the point $c - c'$ and its radius is equal to the difference between the radii of $\mathcal{B}$ and $\mathcal{B}'$, which is equal to the distance of $c$ and $c'$. So the ball corresponding to $R_1$ passes through the origin and the claim follows. ■

**Claim 3.3.10.** *Let $\rho$ and $\sigma$ be single qubit states, then for the expected communication cost of the Greedy Rejection Sampler($\rho$,$\sigma$) we have*

$$\mathbb{E}\left[\, l_{E_2}(J) \,\right] \quad \in \quad \mathrm{S}_{\max}(\rho||\sigma) + 2\lg(\mathrm{S}_{\max}(\rho||\sigma)+1) + \mathcal{O}(1) \ .$$

**Proof:** We show that

$$\mathbb{E}\left[\lg(J)\right] \quad \leq \quad \mathrm{S}_{\max}(\rho||\sigma) + \lg(e) \ .$$

Then the claim follows by an argument as in claim 3.3.4. Suppose that $X_1$, the contribution of the first iteration to Bob's output state has trace $\beta$. By claim 3.3.9, $R_1$ is of the form $R_1 = (1-\beta)|\psi\rangle\langle\psi|$, for some pure state $|\psi\rangle\langle\psi|$ and by definition $s_1 = \beta$. Let $\alpha^{-1} := \langle\psi|\sigma^{-1}|\psi\rangle$. Then similar to the case where $\rho$ is a pure state, it is straightforward to see that for every $j \geq 2$

$$\begin{aligned} X_j &= (1-\beta)(1-\alpha)^{j-2}\alpha|\psi\rangle\langle\psi| \\ R_j &= (1-\beta)(1-\alpha)^{j-1}|\psi\rangle\langle\psi| \\ s_j &= \beta + (1-\beta)\left(1-(1-\alpha)^{j-1}\right) \ . \end{aligned}$$

We have

$$\mathbb{E}\left[\lg(J)\right] \quad = \quad \sum_{j=1}^{\infty} \Pr(J=j)\lg(j) \quad = \quad \sum_{j=1}^{\infty} \mathrm{Tr}(X_j)\lg(j) \ .$$

The term corresponding to $j = 1$ in the above summation is zero. In order to bound $\mathbb{E}\left[\lg(J)\right]$, we bound every $j \geq 2$ as follows. Since $s_j$ is the probability that the protocol terminates within $j$ iterations, we have

$$1 \quad \geq \quad s_j \quad = \quad \sum_{i=1}^{j}\mathrm{Tr}(X_i) \quad = \quad \beta + \sum_{i=2}^{j}(1-s_{i-1})\alpha \quad \geq \quad \beta + (j)(1-s_{j-1})\alpha$$

Thus, $j \leq \dfrac{1-\beta}{(1-s_{j-1})\alpha} \leq \dfrac{1-\beta}{(1-s_{j-1})}(\alpha^{-1})$. Note that this bound also holds for $j = 1$, since

$$\alpha \quad \leq \quad \mathrm{Tr}(R_1) \quad = \quad 1-\beta \ .$$

Using this bound we have

$$\begin{aligned}
\mathbb{E}\left[\lg(J)\right] &= \sum_{j=1}^{\infty} \operatorname{Tr}(X_j)\lg(j) \\
&\leq \sum_{j=1}^{\infty} \operatorname{Tr}(X_j)\lg\left(\frac{1-\beta}{(1-s_{j-1})}(\alpha^{-1})\right) \\
&= \sum_{j=1}^{\infty}(s_j - s_{j-1})\lg\left(\frac{1}{(1-s_{j-1})}\right) \\
&\quad + \sum_{j=1}^{\infty} \operatorname{Tr}(X_j)\lg((1-\beta)\alpha^{-1}) \\
&\leq \int_0^1 \lg\left(\frac{1}{1-s}\right)\mathrm{d}s + \lg((1-\beta)\alpha^{-1}) \qquad (3.3.5) \\
&= \lg(e) + \lg((1-\beta)\alpha^{-1}) \\
&= \lg(e) + \mathrm{S}_{\max}((1-\beta)|\psi\rangle\langle\psi|||\sigma) \\
&\leq \mathrm{S}_{\max}(\rho||\sigma) + \lg(e) \;\; .
\end{aligned}$$

where inequality 3.3.5 holds since the left Riemann sum amounts to an underestimation of the integral $\int_0^1 f(s)\mathrm{d}s$ for the function $f(s) = \lg\left(\frac{1}{1-s}\right)$ which is increasing in the interval $[0,1)$. The last inequality follows from the fact that $(1-\beta)|\psi\rangle\langle\psi|$ is a substate of $\rho$. ■

### 3.3.3 Upper bounding the expected communication cost

In Sections 3.3.1 and 3.3.2, we showed that the average communication of the Greedy Quantum Rejection Sampler for $\rho, \sigma \in \mathsf{D}(\mathcal{H})$ in the case where $\rho$ is a pure state and in the case where $\dim(\mathcal{H}) = 2$ is at most $\mathrm{S}_{\max}(\rho||\sigma) + 2\lg(\mathrm{S}_{\max}(\rho||\sigma)+1) + \mathcal{O}(1)$. This motivates the following question: Is $\mathrm{S}_{\max}(\rho||\sigma) + 2\lg(\mathrm{S}_{\max}(\rho||\sigma) + 1) + \mathcal{O}(1)$ an upper bound for the average communication cost of the Greedy Quantum Rejection Sampler in general? This section contains some of the approaches we took to prove this conjecture. First we introduce a quantum rejection sampling protocol, $\Pi_{\mathrm{S}_{\max}}$, with average communication cost bounded by $\mathrm{S}_{\max}(\rho||\sigma) + 2\lg(\mathrm{S}_{\max}(\rho||\sigma) + 1) + \mathcal{O}(1)$ for arbitrary states $\rho$ and $\sigma$ such that $\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)$.

Let $\rho, \sigma \in \mathsf{D}(\mathcal{H})$ and $\mathcal{K}$ be a finite dimensional Hilbert space such that $\dim(\mathcal{K}) \geq \dim(\mathcal{H})$. Recall that in any quantum rejection sampling protocol, Alice and Bob initially share a sequence, $\{|\phi\rangle_i\}_{i\in\mathbb{N}}$, of a fixed purification, $|\phi\rangle$, of $\sigma$ in $\mathbb{C}^2 \otimes \mathcal{K} \otimes \mathcal{H}$ such that Bob's marginal of each state is $\sigma$ and Alice holds the rest of each state. Let $\alpha = 2^{-\mathrm{S}_{\max}(\rho||\sigma)}$, then by definition $\alpha\rho \leq \sigma$. The protocol $\Pi_{\mathrm{S}_{\max}}$ is defined as follows. In the $j$-th iteration Alice uses the subroutine $\Pi(\alpha\rho, \sigma)$ on $|\phi\rangle_j$ in order to prepare the substate $\alpha\rho$ on Bob's side. If Alice is successful in preparing $\alpha\rho$ she sends the index $J = j$ to Bob using the prefix encoding function $E_2$, then Bob outputs his marginal of the $j$-th state and the protocol terminates. If Alice fails, she proceeds to the $(j+1)$-th iteration.

**Claim 3.3.11.** *The protocol* $\Pi_{\mathrm{S}_{\max}}$ *terminates with probability* $1$ *and Bob's output state is* $\rho$.

**Proof:** The contribution of the $j$-th iteration in Bob's output state, $X_j$, is equal to the probability of reaching the $j$-th iteration times $\alpha\rho$, i.e. $X_j = (1 - s_{j-1})\alpha\rho$. Hence, $s_0 = 1$ and for every $j \geq 1$,

$$s_j \quad = \quad s_{j-1} + (1 - s_{j-1})\alpha \quad = \quad \alpha + (1-\alpha)s_{j-1} \ .$$

It is straightforward to see that for every $j \geq 0$, $s_j = 1 - (1-\alpha)^j$. So we have

$$\sum_{j=1}^{\infty} X_j \quad = \quad \sum_{j=1}^{\infty} (1-\alpha)^j \alpha\rho \quad = \quad \rho \ .$$

■

**Claim 3.3.12.** *The expected communication cost of the protocol* $\Pi_{\mathrm{S}_{\max}}$ *is bounded as*

$$\mathbb{E}\left[\, l_{E_2}(J) \,\right] \quad \in \quad \mathrm{S}_{\max}(\rho||\sigma) + 2\lg(\mathrm{S}_{\max}(\rho||\sigma) + 1) + \mathcal{O}(1) \ .$$

**Proof:** We show that

$$\mathbb{E}\left[\lg(J)\right] \quad \leq \quad \mathrm{S}_{\max}(\rho||\sigma) \ .$$

Then the claim follows by a similar argument as in claim 3.3.4. We have

$$
\begin{aligned}
\mathbb{E}\left[\lg(J)\right] &\leq \lg\left(\mathbb{E}[J]\right) && \text{(By Jensen's inequality)} \\
&= \lg\left(\sum_{j=0}^{\infty} \Pr(J > j)\right) && \text{(By Lemma 3.3.3 )} \\
&= \lg\left(\sum_{j=0}^{\infty} (1 - s_j)\right) \\
&= \lg\left(\sum_{j=0}^{\infty} (1 - \alpha)^j\right) \\
&= \lg\left(\alpha^{-1}\right) \\
&= \mathrm{S}_{\max}(\rho||\sigma) \ .
\end{aligned}
$$

■

Now we return to upper bounding the expected communication cost of the Greedy Quantum Rejection Sampler. Let $P : \mathbb{N} \longrightarrow \mathbb{R}$ be the probability distribution defined for every $j \in \mathbb{N}$ as $P(j) = \mathrm{Tr}(X_j)$, where $X_j$ is the contribution of the $j$-th iteration to Bob's output state in the Greedy Rejection Sampler. Note that in the Greedy Rejection Sampler, for every $j \in \mathbb{N}$, $X_{j+1}$ is a feasible solution to the semidefinite program $(P_j)$ defined in Section 3.2.2 . So $\mathrm{Tr}(X_j)$, the optimal value of $(P_j)$, is more than or equal to $\mathrm{Tr}(X_{j+1})$, for every $j \in \mathbb{N}$, i.e. $P$ is a non-increasing probability distribution over $\mathbb{N}$. Next we introduce a way of proving upper bounds on the expected communication cost of the Greedy Rejection Sampler. We first need to prove the following lemma.

**Lemma 3.3.13.** *Let $P_1, P_2 : \mathbb{N} \longrightarrow \mathbb{R}$ be two probability distributions on $\mathbb{N}$, such that $P_1 \succ P_2$, then*

$$
\sum_{i \in \mathbb{N}} P_1^{\downarrow}(i) \lg(i) \quad \leq \sum_{i \in \mathbb{N}} P_2^{\downarrow}(i) \lg(i) \ .
$$

**Proof:** Let $a_0 = 0$ and for every $i \in \mathbb{N}$, let $a_i = \sum_{j=1}^{i} \left(P_1^{\downarrow}(j) - P_2^{\downarrow}(j)\right)$ and $b_i = \lg(i)$. Then since $P_1 \succ P_2$, for every $i \in \mathbb{N}$, by definition $a_i \geq 0$ . Hence, we have

$$\begin{aligned} 0 &\leq \sum_{i\in\mathbb{N}} a_i(b_{i+1} - b_i) \\ &= \sum_{i\in\mathbb{N}} (a_{i-1} - a_i) b_i \\ &= \sum_{i\in\mathbb{N}} (P_2^{\downarrow}(i) - P_1^{\downarrow}(i)) \lg(i) \ . \end{aligned}$$

■

Note that Lemma 3.3.13 in fact holds if we replace the logarithm function by any function $g : [1, \infty) \longrightarrow \mathbb{R}$ which is non-decreasing over the interval $[1, \infty)$.

Let $Q : \mathbb{N} \longrightarrow \mathbb{R}$ be any non-increasing probability distribution on $\mathbb{N}$ such that $P \succ Q$, then Lemma 3.3.13 states that $\sum_{i\in\mathbb{N}} Q(i)\lg(i)$ is an upper bound on $\mathbb{E}[\lg(J)]$ for the Greedy Quantum Rejection Sampler.

Consider the probability distributions $Q : \mathbb{N} \longrightarrow \mathbb{R}$ defined for every $j \in \mathbb{N}$ as $Q(j) = (1-\alpha)^j \alpha$ , for $\alpha = 2^{-\mathrm{S}_{\max}(\rho||\sigma)}$, where $(1-\alpha)^j\alpha$ is the probability that in the protocol $\Pi_{\mathrm{S}_{\max}}$ terminates with $J = j$. Note that the probability distribution $((1-\alpha)^j\alpha)_{j\in\mathbb{N}}$ is non-increasing. If we could show that $P \succ Q$, then by Lemma 3.3.13 , the upper bound of $\mathrm{S}_{\max}(\rho||\sigma) + 2\lg(\mathrm{S}_{\max}(\rho||\sigma)+1) + \mathcal{O}(1)$ for the expected communication cost of the Greedy Quantum Rejection Sampler could be obtained. It turns out that there are states $\rho$ and $\sigma$ such that $P \not\succ Q$. We simulated the greedy rejection sampler for states $\rho = \frac{5}{6}|+\rangle\langle+| + \frac{1}{6}|-\rangle\langle-|$ and $\sigma = \frac{2}{3}|0\rangle\langle 0| + \frac{1}{3}|1\rangle\langle 1|$, where $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and $|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$, respectively. Table 3.1 contains the results.

A different approach we took for proving an upper bound of $\mathrm{S}_{\max}(\rho||\sigma) + 2\lg(\mathrm{S}_{\max}(\rho||\sigma) + 1) + \mathcal{O}(1)$ is the following. The expected value of $\lg(J)$ in any quantum rejection sampling protocol is given by

$$\mathbb{E}[\lg(J)] \quad = \quad \sum_{j=1}^{\infty} \Pr(J=j)\lg(j) \quad = \quad \sum_{j=1}^{\infty} \mathrm{Tr}(X_j)\lg(j) \ .$$

If we could show that for the Greedy protocol, for every $j \geq 1$,

$$\sum_{i=1}^{j} \mathrm{Tr}(X_i) \quad \geq \quad \sum_{i=1}^{j} (1 - s_{i-1})\alpha \ ,$$

| $j$ | $\mathrm{Tr}(X_j)$ | $\alpha(1-\alpha)^{j-1}$ | $\sum_{i=1}^{j}\mathrm{Tr}(X_i)$ | $\sum_{i=1}^{j}\alpha(1-\alpha)^{i-1}$ |
|---|---|---|---|---|
| 1 | 0.6273 | 0.5193 | 0.6273 | 0.5194 |
| 2 | 0.1441 | 0.2496 | 0.7715 | 0.7690 |
| 3 | 0.0884 | 0.1200 | 0.8599 | 0.8890 |
| 4 | 0.0542 | 0.0577 | 0.9141 | 0.9466 |
| 5 | 0.0332 | 0.0277 | 0.9473 | 0.9744 |
| 6 | 0.0204 | 0.0133 | 0.9677 | 0.9877 |
| 7 | 0.0125 | 0.0064 | 0.9802 | 0.9941 |
| 8 | 0.0077 | 0.0030 | 0.9878 | 0.9972 |
| 9 | 0.0047 | 0.0015 | 0.9925 | 0.9986 |
| 10 | 0.0029 | 0.0007 | 0.9954 | 0.9993 |

Table 3.1: Greedy rejection sampler vs. $\Pi_{\mathrm{S}_{\max}}$ in the first 10 iterations

then using a similar argument as in Claim 3.3.10, we could bound every $j \geq 1$ as follows. Since $s_j$ is the probability that the protocol terminates within $j$ iterations, we have

$$1 \quad \geq \quad s_j \quad = \quad \sum_{i=1}^{j}\mathrm{Tr}(X_i) \quad \geq \quad \sum_{i=1}^{j}(1-s_{i-1})\alpha \quad \geq \quad (j)(1-s_{j-1})\alpha$$

Thus, $j \leq \dfrac{1}{(1-s_{j-1})\alpha} \leq \dfrac{1}{(1-s_{j-1})}(\alpha^{-1})$. Note that this bound also holds for $j = 1$. Similar to Claim 3.3.10

$$\begin{aligned}
\mathbb{E}\left[\lg(J)\right] &= \sum_{j=1}^{\infty}\mathrm{Tr}(X_j)\lg(j) \\
&\leq \sum_{j=1}^{\infty}\mathrm{Tr}(X_j)\lg\left(\frac{1}{(1-s_{j-1})}(\alpha^{-1})\right) \\
&= \sum_{j=1}^{\infty}(s_j - s_{j-1})\lg\left(\frac{1}{(1-s_{j-1})}\right) \\
&\quad + \sum_{j=1}^{\infty}\mathrm{Tr}(X_j)\lg(\alpha^{-1}) \\
&\leq \int_0^1 \lg\left(\frac{1}{1-s}\right)\mathrm{d}s + \lg(\alpha^{-1}) \\
&= \mathrm{S}_{\max}(\rho||\sigma) + \lg(e) \ .
\end{aligned}$$

Note that $\mathrm{Tr}(X_1) \geq \alpha$ since $\alpha\rho$ is a feasible solution of the semidefinite program defining $X_1$. Our first conjecture was that $\mathrm{Tr}(X_j) \geq (1 - s_{j-1})\alpha$, meaning that in the $j$-th iteration at least an $\alpha$ portion of the trace of the remaining state $R_{j-1}$ is removed. This conjecture turned out not to be true. In fact, we could find an instance of the problem for which even the weaker condition $\sum_{i=1}^{j} \mathrm{Tr}(X_i) \geq \sum_{i=1}^{j} (1-s_{i-1})\alpha$ is not satisfied for the Greedy Rejection Sampler. We simulated the greedy rejection sampler for states $\rho = \frac{5}{6}|+\rangle\langle+| + \frac{1}{6}|-\rangle\langle-|$ and $\sigma = \frac{2}{3}|0\rangle\langle 0| + \frac{1}{3}|1\rangle\langle 1|$. Table 3.2 contains the results.

| $j$ | $\mathrm{Tr}(X_j)$ | $(1 - s_{j-1})\alpha$ | $\sum_{i=1}^{j} \mathrm{Tr}(X_i)$ | $\sum_{i=1}^{j}(1 - s_{i-1})\alpha$ |
|---|---|---|---|---|
| 1 | 0.6273 | 0.5194 | 0.6273 | 0.5194 |
| 2 | 0.1441 | 0.1936 | 0.7715 | 0.7129 |
| 3 | 0.0884 | 0.1187 | 0.8599 | 0.8316 |
| 4 | 0.0542 | 0.0728 | 0.9141 | 0.9044 |
| 5 | 0.0332 | 0.0446 | 0.9473 | 0.9490 |
| 6 | 0.0204 | 0.0274 | 0.9677 | 0.9764 |
| 7 | 0.0125 | 0.0168 | 0.9802 | 0.9932 |
| 8 | 0.0077 | 0.0103 | 0.9878 | 1.0035 |
| 9 | 0.0047 | 0.0063 | 0.9925 | 1.0098 |
| 10 | 0.0029 | 0.0039 | 0.9954 | 1.0137 |

Table 3.2: Simulation results for the first 10 iterations of the greedy quantum rejection sampler

What is interesting about this example is that $\rho$ and $\sigma$ are states in $\mathsf{D}(\mathbb{C}^2)$, while in Section 3.3.2 we have shown that the upper bound of $\mathrm{S}_{\max}(\rho||\sigma)+2\lg(\mathrm{S}_{\max}(\rho||\sigma)+1)+\mathcal{O}(1)$ does hold for the Greedy Rejection Sampler in this case. This shows that perhaps we need a stronger approach for proving the upper bound.

An alternative way of upper bounding the expected communication cost of the Greedy Rejection Sampler is to bound $\mathrm{Tr}(X_j)$ for every $j \in \mathbb{N}$. Since in each iteration $\mathrm{Tr}(X_j)$ is the optimal value of a semidefinite program, a natural and more general approach would be using duality theory. Recall that in the Greedy Rejection Sampler, for every $j \in \mathbb{N}$, in the $j$-th iteration, $X_j$ is defined as an optimal solution of the semidefinite program $(P)$

defined as

$$\begin{array}{llrcl} (P): & \text{maximize}: & \operatorname{Tr}(X) & & \\ & \text{subject to}: & X & \leq & A \\ & & X & \leq & B \\ & & X & \geq & 0\ , \end{array}$$

where $A = R_{j-1}$ and $B = \operatorname{Tr}(R_{j-1})\sigma$. We first consider a relaxation $(P')$ of the semidefinite program $(P)$ in which $X$ is free to be any Hermitian operator as follows.

$$\begin{array}{llrcl} (P'): & \text{maximize}: & \operatorname{Tr}(X) & & \\ & \text{subject to}: & X & \leq & A \\ & & X & \leq & B \\ & & X & \in & \mathsf{Herm}(\mathcal{H})\ . \end{array}$$

The dual program of $(P')$ is defined as

$$\begin{array}{llrcl} (D'): & \text{minimize}: & \operatorname{Tr}(AY_1) + \operatorname{Tr}(BY_2) & & \\ & \text{subject to}: & Y_1 + Y_2 & = & \mathbb{1}_{\mathcal{H}} \\ & & Y_1, Y_2 & \geq & 0\ , \end{array}$$

Note that $X = 2^{-(\mathrm{S}_{\max}(A\|B)+1)}A$ and $(Y_1, Y_2) = \left(\frac{\mathbb{1}_{\mathcal{H}}}{2}, \frac{\mathbb{1}_{\mathcal{H}}}{2}\right)$ are strictly feasible solutions for $(P')$ and $(D')$, respectively. So by Theorem 2.1.6 strong duality holds and both $(P')$ and $(D')$ achieve their optimal values. Moreover, $(D')$ is equivalent to

$$\begin{array}{llrcl} (D''): & \text{minimize}: & \operatorname{Tr}(B) + \operatorname{Tr}((A-B)\,Y_1) & & \\ & \text{subject to}: & Y_1 & \geq & 0\ . \end{array}$$

Let $\Pi_-$ be the projector onto the negative eigenspace of $A - B$ and $\Pi_+ = \mathbb{1}_{\mathcal{H}} - \Pi_-$. Then it is straightforward to see that $\Pi_-$ is an optimal solution of $(D'')$ with objective value $\operatorname{Tr}(B) - \frac{1}{2}\|A-B\|_1$. Hence, $(\tilde{Y_1}, \tilde{Y_2}) = (\Pi_-, \Pi_+)$ is an optimal solution of $(D')$. Let $\tilde{X}$ be an optimal solution of $(P')$. By complementary slackness condition, we have $\left(A - \tilde{X}\right)\Pi_- = 0$ and $\left(B - \tilde{X}\right)\Pi_+ = 0$. This implies that

$$\Pi_-\left(A-\tilde{X}\right)\Pi_- = \Pi_+\left(A-\tilde{X}\right)\Pi_- = \Pi_-\left(B-\tilde{X}\right)\Pi_+ = \Pi_+\left(B-\tilde{X}\right)\Pi_+ = 0\ . \tag{3.3.6}$$

Also, by definition of $\Pi_+$ and $\Pi_-$ we have

$$\Pi_+ (A - B) \Pi_- = \Pi_- (A - B) \Pi_+ = 0 \ . \tag{3.3.7}$$

Finally, by (3.3.6) and (3.3.7) we have

$$\begin{aligned}
\tilde{X} &= (\Pi_- + \Pi_+) (\tilde{X}) (\Pi_- + \Pi_+) \\
&= \Pi_- A \Pi_- + \Pi_+ A \Pi_- + \Pi_- B \Pi_+ + \Pi_+ B \Pi_+ \\
&= \Pi_- A \Pi_- + \Pi_+ B \Pi_- + \Pi_- B \Pi_+ + \Pi_+ B \Pi_+ \\
&= \Pi_- A \Pi_- - \Pi_- B \Pi_- + B \\
&= \frac{A - B}{2} - \frac{|A - B|}{2} + B \\
&= \frac{A + B}{2} - \frac{|A - B|}{2} \ .
\end{aligned}$$

Now we return to the original semidefinite program $(P)$. The dual program of $(P)$ is defined as

$$\begin{array}{llrcl}
(D): & \text{minimize :} & \operatorname{Tr}(AY_1) + \operatorname{Tr}(BY_2) & & \\
& \text{subject to :} & Y_1 + Y_2 & \geq & \mathbb{1}_{\mathcal{H}} \\
& & Y_1, Y_2 & \geq & 0 \ .
\end{array}$$

The first difference between the pair $(P, D)$ and $(P', D')$ is that $(P)$ is not necessarily strictly feasible. For example if $\rho$ is not full rank in $\operatorname{supp}(\sigma)$, then no feasible solution of $(P)$ is positive definite. So it is no longer guaranteed that the dual optimal value is achieved. The other difference is that $Y_1 + Y_2 \geq \mathbb{1}_{\mathcal{H}}$ in $(D)$. Since $(D)$ is a minimization problem, one may conjecture that without changing the optimal value of $(D)$ we may replace the constraint $Y_1 + Y_2 \geq \mathbb{1}_{\mathcal{H}}$ with $Y_1 + Y_2 = \mathbb{1}_{\mathcal{H}}$. But this is not generally true since for positive semidefinite operators $A$ and $B$, the operator $\frac{A+B}{2} - \frac{|A-B|}{2}$ is not necessarily positive semidefinite. However, since $(Y_1, Y_2) = (\mathbb{1}_{\mathcal{H}}, \mathbb{1}_{\mathcal{H}})$ is a strictly feasible solution for $(D)$ and $X = 2^{-(\mathrm{S}_{\max}(A,B))}A$ is a feasible solution for $(P)$, by Theorem 2.1.6 strong duality holds and the primal optimal value is achieved. The above observations suggest that the analysis of the dual program $(D)$ is more complicated compared to $(D')$. In order to give a taste of the difficulty in analyzing the dual program $(D)$, next we consider a simple example in which $\rho, \sigma \in \mathsf{D}(\mathbb{C}^2)$ are 2 by 2 density operators and $\rho$ is a pure state.

Let $\sigma = (p) \lvert 0 \rangle\langle 0 \rvert + (1 - p) \lvert 1 \rangle\langle 1 \rvert \in \mathsf{D}(\mathbb{C}^2)$ for $p \in (0, 1)$, and $\rho = \lvert + \rangle\langle + \rvert$. Let $B = \sigma$ and $A = \rho$ in the semidefinite programs $(P)$ and $(D)$. Then $2^{-\mathrm{S}_{\max}(\rho,\sigma)} = 2p(1 - p)$ and the optimal solution of $(P)$ is equal to $\tilde{X} = 2p(1-p)|+\rangle\langle+|$. Let $\epsilon > 0$. As stated in the previous paragraph, the dual program $(D)$ does not necessarily achieve its optimal value but since strong duality holds, there exists a dual feasible solution $(\tilde{Y}_1(\epsilon), \tilde{Y}_2(\epsilon))$ such that

$$0 \quad \leq \quad \mathrm{Tr}(\rho\tilde{Y}_1(\epsilon)) + \mathrm{Tr}(\sigma\tilde{Y}_2(\epsilon)) - \mathrm{Tr}(\tilde{X}) \quad \leq \epsilon \ . \tag{3.3.8}$$

Next we find a dual feasible pair $(\tilde{Y}_1(\epsilon), \tilde{Y}_2(\epsilon))$ satisfying Inequality 3.3.8 . From Inequality 3.3.8 and feasibility of $\tilde{X}$ and $(\tilde{Y}_1(\epsilon), \tilde{Y}_2(\epsilon))$ it follows that

$$\begin{aligned} 0 \quad \leq \quad & \mathrm{Tr}\Big((\rho - \tilde{X})\tilde{Y}_1(\epsilon)\Big) + \mathrm{Tr}\Big((\sigma - \tilde{X})\tilde{Y}_2(\epsilon)\Big) \\ = \quad & \mathrm{Tr}(\rho\tilde{Y}_1(\epsilon)) + \mathrm{Tr}(\sigma\tilde{Y}_2(\epsilon)) - \mathrm{Tr}\Big(\tilde{X}(\tilde{Y}_1(\epsilon) - \tilde{Y}_2(\epsilon))\Big) \\ \leq \quad & \mathrm{Tr}(\rho\tilde{Y}_1(\epsilon)) + \mathrm{Tr}(\sigma\tilde{Y}_2(\epsilon)) - \mathrm{Tr}(\tilde{X}) \quad \leq \epsilon \ , \end{aligned}$$

It is sufficient to have

$$0 \quad \leq \quad \mathrm{Tr}\Big((\rho - \tilde{X})\tilde{Y}_1(\epsilon)\Big) \quad = \quad (1 - 2p(1-p))\ \langle+|\tilde{Y}_1(\epsilon)|+\rangle \quad \leq \quad \epsilon/2$$

and

$$0 \leq \mathrm{Tr}\Big((\sigma - \tilde{X})\tilde{Y}_2(\epsilon)\Big) \leq \epsilon/2 \ . \tag{3.3.9}$$

Let $\zeta = \langle+|\tilde{Y}_1(\epsilon)|+\rangle = \epsilon/2 \leq \frac{\epsilon/2}{(1-2p(1-p))}$. On the other hand, we need to have $\tilde{Y}_1(\epsilon)+\tilde{Y}_2(\epsilon) \geq \mathbb{1}_{\mathbb{C}^2}$, so we choose $\langle+|\tilde{Y}_2(\epsilon)|+\rangle$ to be equal to 1. Let $\tilde{Y}_1(\epsilon)$ and $\tilde{Y}_2(\epsilon)$ be operators of the following form in the $|+\rangle$, $|-\rangle$ basis for real numbers $b, c, d$ and $e$.

$$\tilde{Y}_1(\epsilon) = \begin{pmatrix} \zeta & b \\ b & c \end{pmatrix} \ , \ \tilde{Y}_2(\epsilon) = \begin{pmatrix} 1 & d \\ d & e \end{pmatrix} \ .$$

The constraints $\tilde{Y}_1(\epsilon) \geq 0$ and $\tilde{Y}_2(\epsilon) \geq 0$, are equivalent to $c \geq \frac{b^2}{\zeta}$ and $e \geq d^2$, respectively. Let $e = d^2$. The constraint $\tilde{Y}_1(\epsilon) + \tilde{Y}_2(\epsilon) \geq \mathbb{1}_{\mathbb{C}^2}$ is equivalent to

$$\begin{pmatrix} \zeta & b+d \\ b+d & c+d^2-1 \end{pmatrix} \quad \geq \quad 0 \ ,$$

which holds if $b = -d$ and $c \geq 1 - d^2$. Let $c = \max(1 - d^2\ ,\ \frac{b^2}{\zeta})$. Finally, we need $\tilde{Y}_2(\epsilon)$ to satisfy (3.3.9), i.e.

$$\mathrm{Tr}\left(\begin{pmatrix} \frac{1}{2} - 2p(1-p) & p - \frac{1}{2} \\ p - \frac{1}{2} & \frac{1}{2} \end{pmatrix}\begin{pmatrix} 1 & d \\ d & d^2 \end{pmatrix}\right) \quad \leq \quad \epsilon/2 \ ,$$

or equivalently, $(d + (2p-1))^2 \leq \epsilon$. We choose $d = (1-2p)$. So when $\epsilon \ll 1$ and $p \neq 1/2$, we have $c = \frac{b^2}{\zeta}$, and $\tilde{Y}_1(\epsilon)$ and $\tilde{Y}_2(\epsilon)$ are of the following form.

$$\tilde{Y}_1(\epsilon) = \begin{pmatrix} \zeta & 2p-1 \\ 2p-1 & \frac{(1-2p)^2}{\zeta} \end{pmatrix}, \; \tilde{Y}_2(\epsilon) = \begin{pmatrix} 1 & 1-2p \\ 1-2p & (1-2p)^2 \end{pmatrix} .$$

If $p = 1/2$, then $c = 1$ and we have

$$\tilde{Y}_1(\epsilon) = \begin{pmatrix} \zeta & 0 \\ 0 & 1 \end{pmatrix}, \; \tilde{Y}_2(\epsilon) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} .$$

Note that the dual objective value for the pair $(\tilde{Y}_1(\epsilon), \tilde{Y}_2(\epsilon))$ is equal to $\epsilon/2 + 2p(1-p)$. This example shows that the analysis of the semidefinite programs $(P)$ and $(D)$, even in this simple case, is much more complicated compared to the pair $(P')$ and $(D')$.

## 3.4 General quantum rejection sampling protocols

In this section we take a closer look at general rejection sampling protocols, and give a general lower bound on the average communication cost of a quantum rejection sampling protocol. Also, we describe an optimal rejection sampling protocol in terms of an optimization problem and we find upper bounds on the average communication cost of this protocol.

### 3.4.1 Optimal quantum rejection sampling protocol

According to the definition in Section 3.2.1, any quantum rejection sampling protocol corresponds to a sequence $\{X_j\}_{j\in\mathbb{N}}$, satisfying the following constraints.

- For every $j \in \mathbb{N}$, $X_j$ the contribution of the $j$-th iteration to Bob's output state is a positive semidefinite operator,
- $X_j$s sum up to the state $\rho$, i.e. $\sum_{j\in\mathbb{N}} X_j = \rho$,
- For every $j \in \mathbb{N}$, $X_j$ divided by the probability of reaching the $j$-th iteration is a substate of $\sigma$, in other words $X_1 \leq \sigma$ and $X_j \leq \left(1 - \sum_{i=1}^{j-1} \mathrm{Tr}(X_i)\right) \sigma$, for $j \geq 2$.

Using these constraints, it is possible to define an optimal quantum rejection sampling protocol, $\Pi_{\text{opt}}$, in terms of an optimization problem as follows. Let $C : \mathbb{N} \to \{0,1\}^*$ be an arbitrary prefix-free binary encoding of natural numbers used by Alice to send the index $J$ to Bob, and let $l_C(j)$ denote the length of the codeword for $j \in \mathbb{N}$. Then an optimal solution of the following convex optimization problem corresponds to an optimal quantum rejection sampling protocol.

$$\begin{aligned}
(P_{\text{opt}}) : \qquad \text{minimize :} \quad & \sum_{j\in\mathbb{N}} \operatorname{Tr}(X_j) l_C(j) \\
\text{subject to :} \quad & \sum_{j\in\mathbb{N}} X_j = \rho \\
& X_1 \leq \sigma \\
& X_j \leq \left(1 - \sum_{i=1}^{j-1} \operatorname{Tr}(X_i)\right) \sigma \qquad \forall j \geq 2 \\
& X_j \geq 0 \qquad \forall j \in \mathbb{N} \ .
\end{aligned}$$

### 3.4.2 A lower bound on the expected communication cost

The convex optimization problem $(P_{\text{opt}})$ has infinitely many variables and constraints which makes it more complicated compared to the semidefinite program $(P)$, but the problem is still a convex optimization problem and duality theorems might still be helpful here. Next we prove a lower bound for the communication cost of any quantum rejection sampling protocol using a different approach.

**Theorem 3.4.1.** *Let $C : \mathbb{N} \to \{0,1\}^*$ be an arbitrary prefix-free binary encoding of natural numbers. For any quantum rejection sampling protocol which uses $C$ to encode the index $J$, the expected communication for the pair $(\rho, \sigma)$ is bounded as*

$$\mathbb{E}\left[\, l_C(J) \,\right] \quad \geq \quad \mathrm{S}(\rho || \sigma') \ ,$$

*where* $\sigma' = \sum_{x\in\mathcal{X}} \langle\psi_x|\sigma|\psi_x\rangle |\psi_x\rangle\langle\psi_x|$ .

**Proof:** Let $\rho = \sum_{x\in\mathcal{X}} \lambda_x |\psi_x\rangle\langle\psi_x|$ be the eigenvalue decomposition of the state $\rho$. Suppose that after Alice sends the index $J$ to Bob and Bob outputs his marginal of the $J$-th state, we perform an imaginary measurement $M$ in the eigenbasis of $\rho$. Let $Z_J$ denote the random variable corresponding to the classical outcome of the measurement $M$, taking values in $\mathcal{X}$. Then we have

$$\Pr\left(Z_J = x\right) = \langle\psi_x|\rho|\psi_x\rangle = \lambda_x \ ,$$

and for every $j \in \mathbb{N}$

$$\Pr\left(Z_j = x\right) = \Pr\left(J = j \wedge Z_J = x\right) = \langle\psi_x|X_j|\psi_x\rangle \ .$$

Since for every $j \in \mathbb{N}$, $X_j$ is a feasible solution of $(P_{\mathrm{opt}})$, $X_j$ is a substate of $\sigma$ and we have $\langle\psi_x|X_j|\psi_x\rangle \leq \langle\psi_x|\sigma|\psi_x\rangle$. Hence, the conditional entropy $\mathrm{H}(J|Z_J = x)$ can be bounded as

$$\begin{aligned}
\mathrm{H}(J|Z_J = x) &= -\sum_{j\in\mathbb{N}} \Pr\left(J = j|Z_J = x\right) \lg\left(\Pr\left(J = j|Z_J = x\right)\right) \\
&= \sum_{j\in\mathbb{N}} \frac{\langle\psi_x|X_j|\psi_x\rangle}{\lambda_x} \lg\left(\frac{\lambda_x}{\langle\psi_x|X_j|\psi_x\rangle}\right) \\
&\geq \lg\left(\frac{\lambda_x}{\langle\psi_x|\sigma|\psi_x\rangle}\right) \ .
\end{aligned}$$

Finally, we obtain a lower bound on the average communication as follows.

$$\begin{aligned}
\mathbb{E}\left[\, l_C(J)|\,\right] &\geq \mathrm{H}(J) \\
&= \sum_{x\in\mathcal{X}} \Pr\left(Z_J = x\right) \mathrm{H}(J|Z_J = x) \\
&\geq \sum_{x\in\mathcal{X}} \lambda_x \lg\left(\frac{\lambda_x}{\langle\psi_x|\sigma|\psi_x\rangle}\right) \\
&= \mathrm{S}(\rho||\sigma') \ ,
\end{aligned}$$

for $\sigma' = \sum_{x\in\mathcal{X}} \langle\psi_x|\sigma|\psi_x\rangle|\psi_x\rangle\langle\psi_x|$, where the first inequality follows from Theorem 2.2.4. ■

Note that a key step which allows us to derive the lower bound of $\mathrm{S}(\rho||\sigma')$ on $\mathbb{E}\left[\, l_C(J)\,\right]$ is upper bounding the quantity $\langle\psi_x|X_j|\psi_x\rangle$ and tighter upper bounds on $\langle\psi_x|X_j|\psi_x\rangle$ would give better lower bounds on the expected communication cost of a general quantum rejection sampling protocol. For example, an upper bound of $2^{\langle\psi_x|\lg(\sigma)|\psi_x\rangle}$ on $\langle\psi_x|X_j|\psi_x\rangle$ would imply a lower bound of

$$\sum_{x\in\mathcal{X}} \lambda_x \lg\left(\frac{\lambda_x}{2^{\langle\psi_x|\lg(\sigma)|\psi_x\rangle}}\right) = \mathrm{S}(\rho||\sigma) \ ,$$

on $\mathbb{E}\left[\, l_C(J) \,\right]$ for any quantum rejection sampling protocol and an upper bound of

$$2^{-\mathrm{S}_{\max}(|\psi_x\rangle\langle\psi_x|||\sigma)} = \frac{1}{\langle\psi_x|\sigma^{-1}|\psi_x\rangle}$$

would imply a lower bound of $\sum_{x\in\mathcal{X}}\lambda_x \left(\lg(\lambda_x) + \mathrm{S}_{\max}(|\psi_x\rangle\langle\psi_x|||\sigma)\right)$.

Let $\sigma = \sum_{y\in\mathcal{Y}} \gamma_y|\phi_y\rangle\langle\phi_y|$ be the eigenvalue decomposition of $\sigma$. Then we have

$$\begin{aligned} 2^{-\langle\psi_x|\lg(\sigma)|\psi_x\rangle} &= 2^{\sum_{y\in\mathcal{Y}} -\lg(\gamma_y)|\langle\psi_x|\phi_y\rangle|^2} \\ &\leq \sum_{y\in\mathcal{Y}} \frac{1}{\gamma_y} \left|\langle\psi_x|\phi_y\rangle\right|^2 \\ &= \langle\psi_x|\sigma^{-1}|\psi_x\rangle \ , \end{aligned}$$

also,

$$\begin{aligned} 2^{\langle\psi_x|\lg(\sigma)|\psi_x\rangle} &= 2^{\sum_{y\in\mathcal{Y}} \lg(\gamma_y)|\langle\psi_x|\phi_y\rangle|^2} \\ &\leq \sum_{y\in\mathcal{Y}} \gamma_y \left|\langle\psi_x|\phi_y\rangle\right|^2 \\ &= \langle\psi_x|\sigma|\psi_x\rangle \ . \end{aligned}$$

hence, the following holds.

$$2^{-\mathrm{S}_{\max}(|\psi_x\rangle\langle\psi_x|||\sigma)} \leq 2^{\langle\psi_x|\lg(\sigma)|\psi_x\rangle} \leq \langle\psi_x|\sigma|\psi_x\rangle \ .$$

This in turn implies that

$$\sum_{x\in\mathcal{X}}\lambda_x \left(\lg(\lambda_x) + \mathrm{S}_{\max}(|\psi_x\rangle\langle\psi_x|||\sigma)\right) \geq \mathrm{S}(\rho||\sigma) \geq \mathrm{S}(\rho||\sigma') \ . \tag{3.4.1}$$

Note that all of the quantities in (3.4.1) evaluate to $\mathrm{S}(\rho||\sigma)$ when $\rho$ and $\sigma$ commute, so they are all reasonable candidates to be lower bounds on $\mathbb{E}\left[\, l_C(J) \,\right]$.

To check which quantity is a valid upper bound on $\langle\psi_x|X_j|\psi_x\rangle$, we use the fact that in any quantum rejection sampling protocol, for every $j \in \mathbb{N}$, $X_j$ is constrained to be a positive semidefinite operator which is a simultaneous substate of $\rho$ and $\sigma$. Consider the semidefinite program $Q$ defined as follows.

$$\begin{array}{rlcl} \text{maximize :} & \langle\psi_x|X|\psi_x\rangle & & \\ \text{subject to :} & X & \leq & \rho \\ & X & \leq & \sigma \\ & X & \geq & 0 \ , \end{array}$$

Clearly, the optimal value of this semidefinite program is an upper bound on $\langle\psi_x|X_j|\psi_x\rangle$, for every $j \in \mathbb{N}$. Note that in the case where we impose the additional constraint that every $X_j$ is diagonal in the same basis as $\rho$, i.e. $X_j$ and $\rho$ commute, the optimal value of the above semidefinite program $Q$ is equal to $\min\left(\lambda_x, 2^{-\mathrm{S}_{\max}(|\psi_x\rangle\langle\psi_x|||\sigma)}\right) \leq 2^{-\mathrm{S}_{\max}(|\psi_x\rangle\langle\psi_x|||\sigma)}$, since for every feasible solution $X$ which commutes with $\rho$ we have

$$\begin{array}{rcccl} \langle\psi_x|X|\psi_x\rangle|\psi_x\rangle\langle\psi_x| & \leq & X & \leq & \rho \\ \langle\psi_x|X|\psi_x\rangle|\psi_x\rangle\langle\psi_x| & \leq & X & \leq & \sigma \end{array}$$

so $\langle\psi_x|X|\psi_x\rangle \leq \min\left(\lambda_x, 2^{-\mathrm{S}_{\max}(|\psi_x\rangle\langle\psi_x|||\sigma)}\right)$ and $\min\left(\lambda_x, 2^{-\mathrm{S}_{\max}(|\psi_x\rangle\langle\psi_x|||\sigma)}\right)|\psi_x\rangle\langle\psi_x|$ is a feasible solution.

Next we present a pair $\rho, \sigma \in \mathsf{D}(\mathbb{C}^2)$ for which the optimal value of the semidefinite program $Q$ is as big as $\langle\psi_x|\sigma|\psi_x\rangle$, which shows that this approach fails in giving us better upper bounds on $\langle\psi_x|X_j|\psi_x\rangle$ such as those in (3.4.1).

Let $\sigma = (2/3)\ |0\rangle\langle 0| + (1/3)\ |1\rangle\langle 1|$ and $\rho = a\ |+\rangle\langle +| + (1-a)\ |-\rangle\langle -|$, where $a$ varies in the interval $[0,1]$. Figure 3.1 shows the optimal value of the semidefinite program $Q$ for different values of $a$. Note that $\langle +|\sigma|+\rangle = 0.5$ and $2^{-\mathrm{S}_{\max}(|+\rangle\langle +|||\sigma)} = \frac{4}{9}$.

### 3.4.3 Upper bounds on the expected communication cost

In this section we prove upper bounds on the expected communication cost of the optimal rejection sampling protocol $\Pi_{\text{opt}}$ defined by the convex optimization problem $(P_{\text{opt}})$, where we fix $C : \mathbb{N} \to \{0,1\}^*$ to be the encoding function $E_2$ described in Section 2.2. Since $(P_{\text{opt}})$ is minimization program, any feasible solution of $(P_{\text{opt}})$ (corresponding to a rejection sampling protocol) gives an upper bound on its optimal value. As a result, as shown in Section 3.3.3, the protocol $\Pi_{\mathrm{S}_{\max}}$ gives an upper bound of $\mathrm{S}_{\max}(\rho||\sigma) + 2\lg(\mathrm{S}_{\max}(\rho||\sigma)+1) + \mathcal{O}(1)$ on the optimal value of $(P_{\text{opt}})$. Next we introduce another rejection sampling protocol which

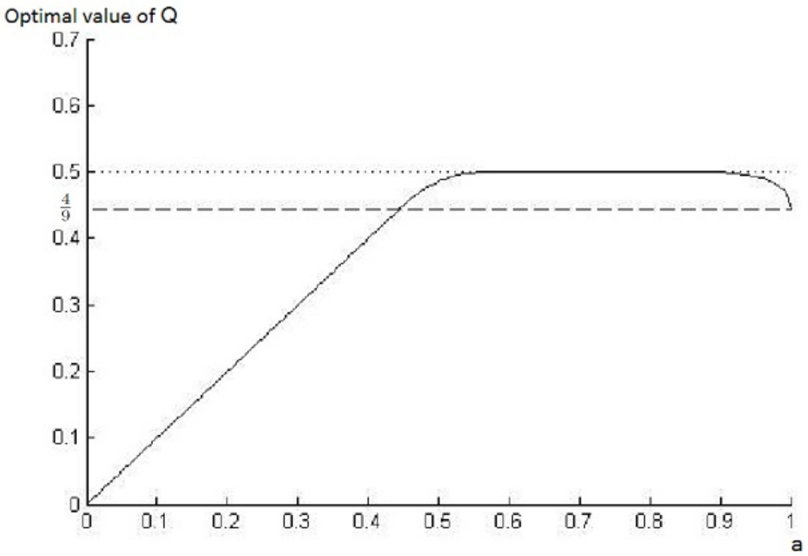

Figure 3.1: Optimal value of the semidefinite program $Q$ versus $a$. The value of $\langle +|X|+\rangle$ can be as big as $\langle +|\sigma|+\rangle$.

gives a different upper bound on the expected communication cost of the optimal rejection sampling protocol.

Let $\rho = \sum_{x\in\mathcal{X}} \lambda_x |\psi_x\rangle\langle\psi_x|$ be the eigenvalue decomposition of the state $\rho$, and let $\alpha_x = 2^{-\mathrm{S}_{\max}(\psi_x||\sigma)}$ for every $x \in \mathcal{X}$. The quantum rejection sampling protocol $\Pi_1$ is defined as follows. In the beginning of the protocol Alice chooses $x \in \mathcal{X}$ with probability $\lambda_x$ and uses the Greedy Rejection Sampler to help Bob output the state $|\psi_x\rangle\langle\psi_x|$ by sending at most $\mathrm{S}_{\max}(|\psi_x\rangle\langle\psi_x|||\sigma) + 2\lg(\mathrm{S}_{\max}(|\psi_x\rangle\langle\psi_x|||\sigma) + 1) + \mathcal{O}(1)$ bits to Bob. Note that Bob's output state is $|\psi_x\rangle\langle\psi_x|$ with probability $\lambda_x$, so it is equal to $\rho$ and the expected communication cost of $\Pi_1$ is bounded as

$$\begin{aligned}\mathbb{E}\left[\,|l(J)|\,\right] &\in \mathbb{E}_x\left[\mathrm{S}_{\max}(|\psi_x\rangle\langle\psi_x|||\sigma) + 2\lg(\mathrm{S}_{\max}(|\psi_x\rangle\langle\psi_x|||\sigma)+1) + \mathcal{O}(1)\right] \\ &= \sum_{x\in\mathcal{X}} \lambda_x \mathrm{S}_{\max}(|\psi_x\rangle\langle\psi_x|||\sigma) + 2\lg(\sum_{x\in\mathcal{X}} \lambda_x \mathrm{S}_{\max}(|\psi_x\rangle\langle\psi_x|||\sigma)+1) + \mathcal{O}(1)\ .\end{aligned}$$

Using the results in Section 3.3.1 it is straightforward to see that the quantum rejection sampling protocol $\Pi_1$ corresponds to the sequence $\left\{\sum_{x\in\mathcal{X}} \lambda_x (1-\alpha_x)^{(j-1)} \alpha_x |\psi_x\rangle\langle\psi_x|\right\}_{j\in\mathbb{N}}$which is a feasible solution to the optimization problem $(P_{\mathrm{opt}})$.

Note that either of $\mathrm{S}_{\max}(\rho||\sigma)$ or $\mathbb{E}_x[\mathrm{S}_{\max}(|\psi_x\rangle\langle\psi_x|||\sigma)]$ may be smaller than the other one. For example, let $\rho, \sigma \in \mathsf{D}(\mathbb{C}^d)$ such that $\sigma = \mathbb{1}_{\mathbb{C}^d}$ and $\rho$ is a state which is close to $\sigma$ with respect to some arbitrary norm. Then $\mathrm{S}_{\max}(\rho||\sigma)$ is close to zero but $\mathbb{E}_x[\mathrm{S}_{\max}(|\psi_x\rangle\langle\psi_x|||\sigma)] = \lg(d)$. Now let $\sigma = \epsilon|0\rangle\langle 0| + (1-\epsilon)|1\rangle\langle 1|$ for $\epsilon \ll 1$ and $\rho = p|0\rangle\langle 0| + (1-p)|1\rangle\langle 1|$ for some $0 < p < 1$. Then it is straightforward to see that $\mathbb{E}_x[\mathrm{S}_{\max}(|\psi_x\rangle\langle\psi_x|||\sigma)]$ in this case is equal to $(1-p)\lg\left(\frac{1}{1-\epsilon}\right) + p\lg\left(\frac{1}{\epsilon}\right)$ and $\mathrm{S}_{\max}(\rho||\sigma) = \lg\left(\frac{p}{\epsilon}\right)$. So in this case, for small enough $\epsilon$, the value of $\mathbb{E}_x[\mathrm{S}_{\max}(|\psi_x\rangle\langle\psi_x||\sigma)]$ can be arbitrarily smaller than $\mathrm{S}_{\max}(\rho||\sigma)$.

## 3.5 Some applications of quantum rejection sampling

### 3.5.1 A scheme for lossless compression of a quantum source

As stated in Section 1.2.1, a quantum source $\mathcal{S}$ can be modelled as an ensemble of quantum states $\{P(x), |\psi_x\rangle\}_{x\in\mathcal{X}}$ on a finite dimensional Hilbert space $\mathcal{H}$, i.e. the source outputs the state $|\psi_x\rangle$ with probability $P(x)$ where the states $\{|\psi_x\rangle\}_{x\in\mathcal{X}}$ do not necessarily form an orthonormal basis. Let $\rho = \sum_{x\in\mathcal{X}} P(x)|\psi_x\rangle\langle\psi_x|$. Variable-length encoding of a quantum source can be studied in different settings. Section 1.3.1 contains some of the known results regarding quantum variable-length codes. Although using a classical side-channel, it is possible to design quantum variable-length coding schemes with lower expected communication cost, the best known general lower and upper bound for the expected length of a quantum variable-length code is given by the Shannon entropy of the source, $\mathrm{H}(P)$, and the von Neumann entropy of the source, $\mathrm{S}(\rho)$, respectively. Note that in general neither bound is tight. In a different model, we may consider classical variable-length encoding of a quantum source. A naive approach for compressing a quantum source by communication of classical bits would be to use the classical data $x$ to encode the source output state by an optimal classical lossless compression scheme like Huffman's coding. Using this strategy, we get an encoding of the source with an average length approximately equal to the Shannon entropy of the source, i.e. $\mathrm{H}(P)$. While this strategy certainly works, it makes no use of the shared entanglement resource.

The Greedy Quantum Rejection Sampler protocol can be used as an LOCC compression scheme for lossless compression of a pure state quantum source in the visible compression model:

Let $\mathcal{S}$ be a pure state source corresponding to the ensemble $\{P(x), |\psi_x\rangle\}_{x\in\mathcal{X}}$, and $\sigma \in \mathsf{D}(\mathcal{H})$ be a quantum state such that for every $x \in \mathcal{X}$, $\langle\psi_x|\sigma|\psi_x\rangle \neq 0$. On source output $|\psi_x\rangle$, Alice uses the Greedy Quantum Rejection Sampler to prepare the state $|\psi_x\rangle$ on Bob's side, using a sequence of purifications of $\sigma$ shared between them. Similar to all previously known quantum variable-length encoding schemes, this protocol is a visible coding scheme. Let $T(\mathcal{S},\sigma)$ denote the average length of this encoding. We have

$$T(\mathcal{S},\sigma) \quad \in \quad \sum_{x\in\mathcal{X}} P(x)\mathrm{S}_{\max}(|\psi_x\rangle\langle\psi_x|||\sigma) + 2\lg\left(\sum_{x\in\mathcal{X}} P(x)\mathrm{S}_{\max}(|\psi_x\rangle\langle\psi_x|||\sigma)+1\right) + \mathcal{O}(1) \ .$$

Hence for the average length of an optimal LOCC encoding of the pure state source $\mathcal{S}$, denoted $T(\mathcal{S})$, we have the following upper bound.

$$T(\mathcal{S}) \quad \leq \quad \min_{\sigma\in\mathsf{D}(\mathcal{H})} T(\mathcal{S},\sigma) \ .$$

Note that $\mathrm{S}_{\max}(|\psi_x\rangle\langle\psi_x|||\rho) \leq \lg\left(1/P(x)\right)$ for every $x\in\mathcal{X}$, hence,

$$\min_{\sigma\in\mathsf{D}(\mathcal{H})} \sum_{x\in\mathcal{X}} P(x)\mathrm{S}_{\max}(|\psi_x\rangle\langle\psi_x|||\sigma) \quad \leq \quad \mathrm{H}(P) \ .$$

Also note that, by Shannon's Lossless Source Coding Theorem 2.2.4, $\mathrm{H}(P)$ is the best achievable value for the average length of an encoding in the case where the states $|\psi_x\rangle$ are mutually orthogonal states (in this case the source can be considered as a classical source). Next we present an example showing the gap between our upper bound and the upper bound of $\mathrm{H}(P)$ can be arbitrary large.

Consider the quantum source corresponding to the ensemble $\left\{\frac{1}{2d}\,,\ \sqrt{\frac{d-1}{d}}|0\rangle \pm \sqrt{\frac{1}{d}}|i\rangle\right\}_{i=1}^{d}$ of pure states in $\mathbb{C}^{d+1}$, where $\{|i\rangle : i=0,\ldots,d\}$ denotes the standard basis for $\mathbb{C}^{d+1}$. For every $i\in\{1,\ldots,d\}$, let $|\phi_i^+\rangle = \sqrt{\frac{d-1}{d}}|0\rangle + \sqrt{\frac{1}{d}}|i\rangle$ and $|\phi_i^-\rangle = \sqrt{\frac{d-1}{d}}|0\rangle - \sqrt{\frac{1}{d}}|i\rangle$. For $\sigma = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2d}\sum_{i=1}^{d}|i\rangle\langle i|$, we have

$$\begin{aligned}
\mathrm{S}_{\max}(|\phi_i^+\rangle\langle\phi_i^+|||\sigma) &= \lg\left(\langle\phi_i^+|\sigma^{-1}|\phi_i^+\rangle\right) \\
&= \lg\left(2(\frac{d-1}{d}) + 2d(\frac{1}{d})\right) \\
&= \lg\left(4-\frac{2}{d}\right) \\
&\leq 1 \ .
\end{aligned}$$

Similarly, we have $\mathrm{S}_{\max}(|\phi_i^-\rangle\langle\phi_i^-|||\sigma) \leq 1$. So the expected communication cost of our protocol is bounded by a constant, while the Shannon entropy of this source is equal to $\lg(2d)$.

### 3.5.2 A protocol for exact remote state preparation

Let $P : \mathcal{X} \longrightarrow \mathbb{R}$ be a probability distribution over a set $\mathcal{X}$, and $\{P(x), \rho_x\}_{x\in\mathcal{X}}$ be an ensemble of quantum states in a $d$-dimensional Hilbert space $\mathcal{H}$. As mentioned earlier in Section 1.2.3, in an RSP protocol for remote preparation of the ensemble $\{P(x), \rho_x\}_{x\in\mathcal{X}}$, the most general action of Alice is to perform a quantum operation, depending on the target state $\rho_x$, on her part of the shared state between the two parties. Since the communication is classical, her quantum operation has a classical output (along with possibly some other classical and quantum by-products), which is the message she sends to Bob. The probabilities of sending different messages are not the same in general. Hence, by using variable encoding schemes for compression of the messages we can save communication in expectation. The quantum rejection sampling protocols introduced in this thesis are all high-entanglement deterministic exact remote state preparation protocols which use prefix-free encoding to communicate messages. In particular, consider the following ERSP protocol: Let $\sigma \in \mathsf{D}(\mathcal{H})$ be a quantum state such that $\mathrm{S}_{\max}(\rho_x||\sigma)$ is finite for every $x \in \mathcal{X}$. On input $x \in \mathcal{X}$, Alice uses the protocol $\Pi_{\mathrm{S}_{\max}}$ introduced in Section 3.3.3 to prepare the state $\rho_x$ on Bob's side. Note that this scheme is a non-oblivious RSP protocol.

The expected communication cost of the described ERSP protocol for the worst case input and for the best choice of the state $\sigma$ is at most

$$\min_{\sigma\in\mathsf{D}(\mathcal{H})} \max_{x\in\mathcal{X}} \mathrm{S}_{\max}(\rho_x||\sigma) + 2\lg(\mathrm{S}_{\max}(\rho_x||\sigma) + 1) + \mathcal{O}(1) \enspace .$$

Let $\xi_{AB}$ denote the joint quantum state of Alice's classical input register and Bob's output register after the protocol terminates, then $\xi_{AB} = \sum_{x\in\mathcal{X}} P(x)|x\rangle\langle x|^A \otimes \rho_x^B$. The max-information the register $B$ has about the register $A$, defined in Section 2.2, is given by

$$\begin{aligned} \mathrm{I}_{\max}(A:B)_{\xi_{AB}} &= \min_{\sigma\in\mathsf{D}(\mathcal{H})} \mathrm{S}_{\max}(\xi_{AB}||\xi_A \otimes \sigma_B) \\ &= \min_{\sigma\in\mathsf{D}(\mathcal{H})} \max_{x\in\mathcal{X}} \mathrm{S}_{\max}(\rho_x||\sigma) \enspace , \end{aligned}$$

where $\xi_A = \mathrm{Tr}_B(\xi_{AB}) = \sum_{x\in\mathcal{X}} P(x)|x\rangle\langle x|$. The second equality follows from the fact that for $\sigma \in \mathsf{D}(\mathcal{H})$ and $\lambda \in \mathbb{R}$,

$$\xi_{AB} \leq 2^{\lambda}\xi_A \otimes \sigma \qquad \text{if and only if} \qquad \rho_x \leq 2^{\lambda}\sigma \quad \text{for every } x \in \mathcal{X} \enspace .$$

So the expected communication cost of the above ERSP protocol for the worst case input is at most $\mathrm{I}_{\max}(A:B)_{\xi_{AB}} + 2\lg\left(\mathrm{I}_{\max}(A:B)_{\xi_{AB}}\right) + \mathcal{O}(1)$. Note that for the maximally mixed state $\sigma = \frac{\mathbb{1}_{\mathcal{H}}}{d}$, the value of $\mathrm{S}_{\max}(\rho||\sigma)$ is at most $\lg(d)$, for every $\rho \in \mathsf{D}(\mathcal{H})$, hence, we always have $\mathrm{I}_{\max}(A:B)_{\xi_{AB}} \leq \lg(d)$. We have the following observations.

- For an ensemble $\{P(x), \rho_x\}_{x \in \mathcal{X}}$ of quantum states, the expected communication cost of our protocol can be strictly less than $\lg(d)$.

- The quantum rejection sampling protocol $\Pi_{\mathrm{S}_{\max}}$ can be used as a universal deterministic ERSP protocol when $\sigma$ is fixed to the maximally mixed state. In this case the expected communication cost is at most $\lg(d) + 2\lg(\lg(d)) + \mathcal{O}(1)$. Note that this does not contradict Lo's conjecture since our protocol uses a probabilistic amount of resources and only the expected communication cost is less than $2\lg(d)$ (for large enough $d$).

- The column method of Bennett *et al.* [9] described in Section 1.2.3 can be modified to a universal ERSP protocol similar to ours as follows. Alice and Bob share infinitely many copies of the maximally entangled state $|\phi\rangle = \frac{1}{\sqrt{d}} \sum_{i=1}^{d} |i\rangle|i\rangle$. Alice sends to Bob the index of the first state in which the outcome of her measurement is "0", by using the prefix-free encoding $E_2$ described in Section 2.2. Since the probability of the "0" outcome in each measurement is equal to $\frac{1}{d}$, the average value of the index sent to Bob is $d$, and the expected communication cost is at most $\lg(d) + 2\lg(\lg(d)) + \mathcal{O}(1)$. Although this protocol has the same upper bound for the expected communication cost as our protocol, it can be used only for pure state ensembles.

- In Ref [5], Bab Hadiashar proves the lower bound of $\mathrm{I}_{\max}(A:B)_{\xi_{AB}}$ for the communication cost of any ERSP protocol which uses fixed length classical messages. The expected communication cost of our protocol achieves this bound. It remains open whether the same lower bound can be extended to the case of variable length messages.

- Characterizing the expected communication cost of the protocol $\Pi_{\mathrm{opt}}$ described in Section 3.4.1, may lead to a better upper bound on the expected communication cost of an optimal ERSP protocol.

# Chapter 4

# Conclusion and outlook

In this thesis we studied the following communication task: Suppose that $\rho$ is a quantum state which is known to Alice but not to Bob. Alice wants to help Bob output a single copy of $\rho$ by taking advantage of having access to an infinite number of copies of a purification of another state $\sigma$, shared between the two parties. Moreover, Alice is only allowed to send a single classical message to Bob. We gave a mathematical definition for a general quantum rejection sampling protocol for this communication task. We introduced the Greedy Quantum Rejection Sampler and we proved an expected communication cost of $\mathrm{S}_{\max}(\rho||\sigma) + \mathcal{O}\left(\lg\left(\mathrm{S}_{\max}(\rho||\sigma) + 1\right)\right)$ for this protocol, in the case where $\rho$ is a pure state. In addition, we showed that the expected communication cost of an optimal quantum rejection sampling protocol is equal to the optimal value of a convex optimization problem. We showed that the optimal value of this problem is bounded below by $\mathrm{S}(\rho||\sigma')$, where $\sigma' = \sum\limits_{x\in\mathcal{X}} \langle\psi_x|\sigma|\psi_x\rangle|\psi_x\rangle\langle\psi_x|$ and $\{|\psi_x\rangle\}_{x\in\mathcal{X}}$ is a set of eigenvectors of $\rho$. We also showed two different upper bounds on the expected communication cost of an optimal quantum rejection sampling protocol, one in terms of $\mathrm{S}_{\max}(\rho||\sigma)$ and one in terms of $\sum_{x\in\mathcal{X}} \lambda_x \mathrm{S}_{\max}(|\psi_x\rangle\langle\psi_x|||\sigma)$, where $\lambda_x$ is the eigenvalue corresponding to eigenvector $|\psi_x\rangle$ of $\rho$. We used the Greedy Quantum Rejection Sampler protocol, as a variable-length LOCC encoding scheme for lossless compression of an arbitrary pure state quantum source. For a quantum source corresponding to the ensemble $\{P(x), |\psi_x\rangle\}_{x\in\mathcal{X}}$ in a finite dimensional Hilbert space $\mathcal{H}$, we showed an upper bound on the average length of this encoding in terms of $\min\limits_{\sigma\in\mathsf{D}(\mathcal{H})} \sum\limits_{x\in\mathcal{X}} \lambda_x \mathrm{S}_{\max}(|\psi_x\rangle\langle\psi_x|||\sigma)$, which is always less than or equal to the Shannon entropy of the quantum source, $\mathrm{H}(P)$. We also showed that the gap between the two quantities can be arbitrarily large for some ensembles. Finally, we introduced a high-entanglement deterministic exact protocol for remote preparation of an arbitrary ensemble of quantum states $\{P(x), \rho_x\}_{x\in\mathcal{X}}$, in a $d$-dimensional Hilbert space. Our protocol is a non-oblivious scheme based on a quantum rejection sampling protocol and uses variable-length encoding for communication of the messages to reduce the expected communication cost. We proved an upper bound of $\mathrm{I}_{\max}(A : B)_{\xi_{AB}} + \mathcal{O}\left(\lg\left(\mathrm{I}_{\max}(A : B)_{\xi_{AB}} + 1\right)\right)$ on the expected communication cost of this protocol for the worst case choice of the state $\rho_x$, where $\xi_{AB}$ is the joint state of Alice's classical input register $A$ and Bob's output register $B$ at the end of the protocol. In addition, we showed that $\mathrm{I}_{\max}(A : B)_{\xi_{AB}}$ is always less than or equal to $\lg(d)$. We also explained how this protocol can be used as a universal deterministic exact protocol for remote preparation of an arbitrary $d$-dimensional state at an expected communication cost of at most $\lg(d) + \mathcal{O}\left(\lg(\lg(d))\right)$.

We conclude this thesis with the following open questions. In the classical setting, as established by Harsha *et al.* [25], the expected communication cost of the greedy rejection sampler can be characterized in terms of relative entropy. In the quantum setting, we showed that the expected communication cost of the Greedy Quantum Rejection Sampler in the case where the target state $\rho$ is a pure state, can be characterized in terms of a different quantity which is max-relative entropy. It remains open whether max-relative entropy is the correct measure for the expected communication cost of the Greedy Quantum Rejection Sampler in general. Furthermore, the question of whether as in the classical case, the Greedy protocol in the quantum setting is an optimal rejection sampling protocol remains to be answered. In the special case where the state $\rho$ is a pure state, we have shown this to be true. Moreover, we characterized the expected communication cost of an optimal quantum rejection sampling protocol in terms of an optimization problem. Tighter bounds on the optimal value of this convex optimization problem can be used to strengthen the results in this thesis and probably to prove a direct sum result in bounded-round quantum communication complexity.

# References

[1] Rudolf Ahlswede and Ning Cai. On lossless quantum data compression with a classical helper. *IEEE Transactions on Information Theory*, 50(6):1208–1219, June 2004.

[2] Rudolf Ahlswede and Ning Cai. On lossless quantum data compression and quantum variable-length codes. In Thomas Beth and Gerd Leuchs, editors, *Quantum Information Processing*, pages 70–82. Wiley-VCH Verlag GmbH & Co. KGaA, 2005.

[3] Andris Ambainis, Loïck Magnin, Martin Roetteler, and Jérémie Roland. Symmetry-assisted adversaries for quantum state generation. In *26th Annual IEEE Conference on Computational Complexity*, pages 167–177. IEEE Computer Society, 2010.

[4] Anurag Anshu, Rahul Jain, Priyanka Mukhopadhyay, Ala Shayeghi, and Penghui Yao. A new operational interpretation of relative entropy and trace distance between quantum states. *arXiv:1404.1366 [quant-ph]*, April 2014.

[5] Shima Bab Hadiashar. *Communication Complexity of Remote State Preparation.* Master's thesis, University of Waterloo, Waterloo, May 2014.

[6] Howard Barnum, Carlton M. Caves, Christopher A. Fuchs, Richard Jozsa, and Benjamin Schumacher. On quantum coding for ensembles of mixed states. *Journal of Physics A: Mathematical and General*, 34(35):6767, September 2001.

[7] Howard Barnum, Christopher A. Fuchs, Richard Jozsa, and Benjamin Schumacher. General fidelity limit for quantum channels. *Physical Review A*, 54(6):4707–4711, December 1996.

[8] Charles Bennett, Gilles Brassard, Claude Crépeau, Richard Jozsa, Asher Peres, and William Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Physical Review Letters*, 70(13):1895–1899, March 1993.

[9] Charles H. Bennett, Gilles Brassard, Richard Jozsa, Dominic Mayers, Asher Peres, Benjamin Schumacher, and William K. Wootters. Reduction of quantum entropy by reversible extraction of classical information. *Journal of Modern Optics*, 41(12):2307–2314, December 1994.

[10] Charles H. Bennett, Igor Devetak, Aram W. Harrow, Peter W. Shor, and Andreas Winter. Quantum reverse Shannon theorem. *IEEE Transactions on Information Theory*, 60(5):2926–2959, May 2014.

[11] Charles H. Bennett, David P. DiVincenzo, Peter W. Shor, John A. Smolin, Barbara M. Terhal, and William K. Wootters. Remote state preparation. *Physical Review Letters*, 87(7), July 2001.

[12] Charles H. Bennett, Patrick Hayden, Debbie W. Leung, Peter W. Shor, and Andreas Winter. Remote preparation of quantum states. *IEEE Transactions on Information Theory*, 51(1):56–74, January 2005.

[13] Charles H. Bennett, Peter W. Shor, John A. Smolin, and Ashish V. Thapliyal. Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem. *IEEE Transactions on Information Theory*, 48(10):2637–2655, October 2002.

[14] Mario Berta. *Single-shot Quantum State Merging.* Diploma thesis, ETH, Zurich, December 2009.

[15] Mario Berta, Matthias Christandl, and Renato Renner. The quantum reverse Shannon theorem based on one-shot information theory. *Communications in Mathematical Physics*, 306(3):579–615, September 2011.

[16] Kim Bostroem and Timo Felbinger. Lossless quantum data compression and variable-length coding. *Physical Review A*, 65(3):032313, February 2002.

[17] Stephen P. Boyd and Lieven Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004.

[18] Samuel L. Braunstein, A. Christopher A. Fuchs, Daniel Gottesman, and Hoi-Kwong Lo. A quantum analog of Huffman coding. In *IEEE International Symposium on Information Theory*, 1998.

[19] Mark Braverman and Anup Rao. Information equals amortized communication. *IEEE Transactions on Information Theory*, 60(10):6058–6069, October 2014.

[20] Matthias Christandl, Robert Konig, and Renato Renner. Postselection technique for quantum channels with applications to quantum cryptography. *Physical Review Letters*, 102(2):020504, January 2009.

[21] Matthieu E. Deconinck and Barbara M. Terhal. Qubit state discrimination. *Phys. Rev. A*, 81:062304, Jun 2010.

[22] Igor Devetak. The private classical capacity and quantum capacity of a quantum channel. *IEEE Transactions on Information Theory*, 51(1):44–55, January 2005.

[23] Igor Devetak and Toby Berger. Low-entanglement remote state preparation. *Physical Review Letters*, 87(19), October 2001.

[24] Igor Devetak and Toby Berger. Quantum rate-distortion theory for memoryless sources. *IEEE Transactions on Information Theory*, 48(6):1580–1589, June 2002.

[25] Prahladh Harsha, Rahul Jain, David McAllester, and Jaikumar Radhakrishnan. The communication complexity of correlation. *IEEE Transactions on Information Theory*, 56(1):438–449, January 2010.

[26] Masahito Hayashi. Exponents of quantum fixed-length pure-state source coding. *Physical Review A*, 66(6):069901, December 2002.

[27] Patrick Hayden, Richard Jozsa, and Andreas Winter. Trading quantum for classical resources in quantum data compression. *Journal of Mathematical Physics*, 43(9):4404–4444, September 2002.

[28] Alexander S. Holevo. Statistical problems in quantum physics. In *Proceedings of the Second Japan-USSR Symposium on Probability Theory*, number 330 in Lecture Notes in Mathematics, pages 104–119. Springer Berlin Heidelberg, January 1973.

[29] Alexander S. Holevo. The capacity of the quantum channel with general signal states. *IEEE Transactions on Information Theory*, 44(1):269–273, January 1998.

[30] Peter Hoyer, Troy Lee, and Robert Spalek. Negative weights make adversaries stronger. In *Proceedings of the Thirty-ninth Annual ACM Symposium on Theory of Computing*, STOC '07, pages 526–535, New York, NY, USA, 2007. ACM.

[31] David Huffman. A method for the construction of minimum-redundancy codes. *Proceedings of the IRE*, 40(9):1098–1101, September 1952.

[32] Rahul Jain and Ashwin Nayak. Short proofs of the quantum substate theorem. *IEEE Transactions on Information Theory*, 58(6):3664–3669, June 2012.

[33] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. Privacy and interaction in quantum communication complexity and a theorem about the relative entropy of quantum states. In *The 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002. Proceedings*, pages 429–438, 2002.

[34] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. Prior entanglement, message compression and privacy in quantum communication. In *The Twentieth Annual IEEE Conference on Computational Complexity, 2005. Proceedings*, pages 285–296, June 2005.

[35] Richard Jozsa and Benjamin Schumacher. A new proof of the quantum noiseless coding theorem. *Journal of Modern Optics*, 41(12):2343–2349, December 1994.

[36] Masato Koashi and Nobuyuki Imoto. Compressibility of quantum mixed-state signals. *Physical Review Letters*, 87(1):017902, June 2001.

[37] Debbie W. Leung and Peter W. Shor. Oblivious remote state preparation. *Physical Review Letters*, 90(12), March 2003.

[38] Ming Li and Paul M. B. Vitanyi. *An Introduction to Kolmogorov Complexity and Its Applications.* Springer, New York, 3rd ed. 2008 edition, November 2008.

[39] Seth Lloyd. Capacity of the noisy quantum channel. *Physical Review A*, 55(3):1613–1622, March 1997.

[40] Hoi-Kwong Lo. Classical communication cost in distributed quantum information processing - a generalization of quantum communication complexity. *Physical Review A*, 62(1), June 2000.

[41] Michael A. Nielsen. *Quantum information theory.* PhD thesis, November 2000.

[42] Maris Ozols, Martin Roetteler, and Jérémie Roland. Quantum rejection sampling. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, ITCS '12, pages 290–308, New York, NY, USA, 2012. ACM.

[43] Arun K. Pati. Minimum cbits for remote preperation and measurement of a qubit. *Physical Review A*, 63(1), December 2000.

[44] Benjamin Schumacher. Quantum coding. *Physical Review A*, 51(4):2738–2747, April 1995.

[45] Benjamin Schumacher and Michael D. Westmoreland. Sending classical information via noisy quantum channels. *Physical Review A*, 56(1):131–138, July 1997.

[46] Benjamin Schumacher and Michael D. Westmoreland. Indeterminate-length quantum coding. *Physical Review A*, 64(4):042304, September 2001.

[47] Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, July 1948.

[48] John Watrous. *Theory of Quantum Information.* Lecture notes, 2013. https://cs.uwaterloo.ca/˜watrous/CS766/.

[49] Andreas Winter. *Coding Theorems of Quantum Information Theory.* PhD thesis, July 1999.

[50] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing(preliminary report). In *Proceedings of the Eleventh Annual ACM Symposium on Theory of Computing*, STOC '79, pages 209–213, New York, NY, USA, 1979. ACM.

[51] Andrew Chi-Chih Yao. Quantum circuit complexity. In *The 34th Annual Symposium on Foundations of Computer Science, 1993. Proceedings*, pages 352–361, November 1993.