

Article

# Computing Classical-Quantum Channel Capacity Using Blahut–Arimoto Type Algorithm: A Theoretical and Numerical Analysis <sup>†</sup>

Haobo Li <sup>1,2,3,\*</sup>  and Ning Cai <sup>1,\*</sup><sup>1</sup> School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China<sup>2</sup> Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai 201210, China<sup>3</sup> University of Chinese Academy of Sciences, Beijing 100049, China

\* Correspondence: lihb@shanghaitech.edu.cn (H.L.); ningcai@shanghaitech.edu.cn (N.C.)

<sup>†</sup> This paper is partially presented at 2019 International Symposium on Information Theory.

Received: 15 January 2020; Accepted: 13 February 2020; Published: 16 February 2020



**Abstract:** Based on Arimoto’s work in 1972, we propose an iterative algorithm for computing the capacity of a discrete memoryless classical-quantum channel with a finite input alphabet and a finite dimensional output, which we call the Blahut–Arimoto algorithm for classical-quantum channel, and an input cost constraint is considered. We show that, to reach  $\varepsilon$  accuracy, the iteration complexity of the algorithm is upper bounded by  $\frac{\log n \log \varepsilon}{\varepsilon}$  where  $n$  is the size of the input alphabet. In particular, when the output state  $\{\rho_x\}_{x \in \mathcal{X}}$  is linearly independent in complex matrix space, the algorithm has a geometric convergence. We also show that the algorithm reaches an  $\varepsilon$  accurate solution with a complexity of  $O(\frac{m^3 \log n \log \varepsilon}{\varepsilon})$ , and  $O(m^3 \log \varepsilon \log_{(1-\delta)} \frac{\varepsilon}{D(p^* || p^{N_0})})$  in the special case, where  $m$  is the output dimension,  $D(p^* || p^{N_0})$  is the relative entropy of two distributions, and  $\delta$  is a positive number. Numerical experiments were performed and an approximate solution for the binary two-dimensional case was analysed.

**Keywords:** capacity; classical-quantum channel; Blahut–Arimoto type algorithm; convergence speed

## 1. Introduction

The computation of channel capacity has always been a core problem in information theory. The very well-known Blahut-Arimoto algorithm [1,2] was proposed in 1972 to compute the discrete memoryless classical channel. Inspired by this algorithm, we propose an algorithm of Blahut-Arimoto type to compute the capacity of discrete memoryless classical-quantum channel. The classical-quantum channel [3] can be considered as a mapping  $x \rightarrow \rho_x$  of an input alphabet  $\mathcal{X} = \{1, 2, \dots, |\mathcal{X}|\}$  to a set of quantum states in a finite dimensional Hilbert space  $\mathcal{H}$ . The state of a quantum system is given by a density operator  $\rho$ , which is a positive semi-definite operator with trace equal to one. Let  $\mathcal{D}^m$  denote the set of all density operators acting on a Hilbert space  $\mathcal{H}$  of dimension  $m$ . If the source emits a letter  $x$  with probability  $p_x$ , the output would be  $\rho_x$ , thus the output would form an ensemble:  $\{p_x : \rho_x\}_{x \in \mathcal{X}}$ .

In 1998, Holevo showed [4] that the classical capacity of the classical-quantum channel is the maximization of a quantity called the Holevo information over all input distributions. The Holevo information  $\chi$  of an ensemble  $\{p_x : \rho_x\}_{x \in \mathcal{X}}$  is defined as

$$\chi(\{p_x : \rho_x\}_{x \in \mathcal{X}}) = H(\sum_x p_x \rho_x) - \sum_x p_x H(\rho_x), \quad (1)$$

where  $H(\cdot)$  is the von Neumann entropy, which is defined on positive semidefinite matrices:

$$H(\rho) = -\text{Tr}(\rho \log \rho). \quad (2)$$

Due to the concavity of von Neumann entropy [5], the Holevo information is always non-negative. The Holevo quantity is concave in the input distribution [5], and thus the maximization of Equation (1) over  $p$  is a convex optimization problem. However, it is not a straightforward convex optimization problem. In 2014, Sutter et al. [6] promoted an algorithm based on duality of convex programming and smoothing techniques [7] with a complexity of  $O(\frac{(n \vee m)m^3(\log n)^{1/2}}{\epsilon})$ , where  $n \vee m = \max\{n, m\}$ .

For discrete memoryless classical channels, the capacity can be computed efficiently by using an algorithm called Blahut–Arimoto (BA) algorithm [1,2,8]. In 1998, H. Nagaoka [9] proposed a quantum version of BA algorithm. In his work, he considered the quantum–quantum channel and this problem was proved to be NP-complete in 2008 [10]. Despite the NP-completeness, in [11], an example is given of a qubit quantum channel which requires four inputs to maximize the Holevo capacity. Further research of Nagaoka’s algorithm was presented in [12], where the algorithm was implemented to check the additivity of quantum channels. In [9], Nagaoka mentioned an algorithm concerning classical–quantum channel; however, its speed of convergence was not studied there and the details of the proof were not presented either. In this paper, we show that, with proper manipulations, the BA algorithm can be applied to computing the capacity of classical–quantum channel with an input constraint efficiently. The remainder of this article is structured as follows. In Section 2, we propose the algorithm and show how the algorithm works. In Section 3, we provide the convergence analysis of the algorithm. In Section 4, we show the numerical experiments of BA algorithm to see how well this algorithm performs. In Section 5, we propose an approximate solution for a special case, which is the binary input, two-dimensional output case.

**Notations:** The logarithm with basis 2 is denoted by  $\log(\cdot)$ . The space of all Hermitian operators of dimension  $m$  is denoted by  $H^m$ . The set of all density matrices of dimension  $m$  is denoted by  $\mathcal{D}^m := \{\rho \in H^m : \rho \geq 0, \text{Tr} \rho = 1\}$ . Each letter  $x \in \mathcal{X}$  is mapped to a density matrix  $\rho_x$ , thus the classical–quantum channel can be represented as a set of density matrices  $\{\rho_x\}_{x \in \mathcal{X}}$ . The set of all probability distributions of length  $n$  is denoted by  $\Delta_n := \{p \in \mathbb{R}^n : p_x \geq 0, \sum_{x=1}^n p_x = 1\}$ . The von Neumann entropy of a density matrix  $\rho$  is denoted by  $H(\rho) = -\text{Tr}[\rho \log \rho]$ . The relative entropy between  $p, q \in \Delta_n$ , if  $\text{supp}(p) \subset \text{supp}(q)$ , is denoted by  $D(p||q) = \sum_x p_x(\log p_x - \log q_x)$  and  $+\infty$  otherwise. The relative entropy between  $\rho, \sigma \in \mathcal{D}^m$ , if  $\text{supp}(\rho) \subset \text{supp}(\sigma)$ , is denoted by  $D(\rho||\sigma) = \text{Tr}[\rho(\log \rho - \log \sigma)]$  and  $+\infty$  otherwise.

## 2. Blahut–Arimoto Algorithm for Classical–Quantum Channel

First, we write down the primal optimization problem:

$$\text{Primal : } \begin{cases} \max_p H(\sum_x p_x \rho_x) - \sum_x p_x H(\rho_x), \\ \text{subject to } s^T p \leq S; \\ p \in \Delta_n, \end{cases} \quad (3)$$

where  $\rho_x \in \mathcal{D}^m$ ,  $s \in \mathbb{R}^n$  is a positive real vector,  $S > 0$ . We denote the maximal value of Equation (3) as  $C(S)$ . In this optimization problem, we are to maximize the Holevo quantity with respect to the input distribution  $\{p_x\}_{x \in \mathcal{X}}$ . Practically, the preparation of different signal state  $x$  has different cost, which is represented by  $s = (s_1, s_2, \dots, s_n)$ . We would like to bound the expected cost of the resource within some quantity, which is represented by the inequality constraint in Equation (3).

**Lemma 1.** [6] Let a set  $G$  be defined as  $G := \arg \max_{p \in \Delta_n} \chi(\{p_x : \rho_x\}_{x \in \mathcal{X}})$  and  $S_{max} := \min_{p \in G} s^T p$ . Then, if  $S \geq S_{max}$ , the inequality constraint in the primal problem is inactive; and, if  $S < S_{max}$ , the inequality constraint in the primal problem is equivalent to  $s^T p = S$ .

Now, we assume that  $\min\{s_x\}_{x \in \mathcal{X}} \leq S \leq S_{max}$ . The Lagrange dual problem of Equation (3) is

$$\text{Dual : } \begin{cases} \min_{\lambda \geq 0} \max_p H(\sum_x p_x \rho_x) - \sum_x p_x H(\rho_x) - \lambda(s^T p - S) \\ \text{subject to } p \in \Delta_n. \end{cases} \tag{4}$$

**Lemma 2.** Strong duality holds between Equations (3) and (4).

**Proof.** The lemma follows from standard strong duality result of convex optimization theory ([13], Chapter 5.2.3). □

Define functions. Let

$$f_\lambda(p, p') = \sum_x \text{Tr}\{p_x \rho_x [\log(p'_x \rho_x) - \log(p_x \rho_x)]\} - \lambda s^T p, \tag{5}$$

$$F(\lambda) = \max_p \max_{p'} f(p, p'). \tag{6}$$

where  $p' = \sum_x p'_x \rho_x$ .

**Lemma 3.** For fixed  $p$ ,  $\arg \max_{p'} f_\lambda(p, p') = p$ .

**Proof.** Actually, we can prove a stronger lemma (the following lemma was proposed in [9], but no proof was given, perhaps due to the space limitation). We now restate the lemma in [9] and give the proof. □

**Lemma 4.** For fixed  $\{p_x : \rho_x\}_{x \in \mathcal{X}}$ , we have

$$\begin{aligned} \max_{\{q_x : \sigma_x\}_{x \in \mathcal{X}}} -D(p||q) + \sum_x p_x \text{Tr}\{\rho_x [\log \sigma_x - \log \rho_x]\} &= \sum_x p_x \text{Tr}\{\rho_x [\log \rho_x - \log \rho]\}, \\ \text{i.e., } \arg \max_{\{q_x : \sigma_x\}_{x \in \mathcal{X}}} -D(p||q) + \sum_x p_x \text{Tr}\{\rho_x [\log \sigma_x - \log \rho]\} &= \{p_x : \rho_x\}_{x \in \mathcal{X}}, \end{aligned} \tag{7}$$

where  $p, q \in \Delta_n, \sigma_x \in \mathcal{D}^m$  and  $\rho = \sum_x p_x \rho_x, \sigma = \sum_x q_x \sigma_x$ .

**Proof.** Considering Equation (7), we have

$$\begin{aligned} \text{RHS} - \text{LHS} &= D(p||q) + \sum_x p_x D(\rho_x || \sigma_x) - D(\rho || \sigma) \\ &= D(\rho_{XB} || \sigma_{XB}) - D(\rho || \sigma), \end{aligned}$$

where  $\rho_{XB} = \sum_x p_x |x\rangle\langle x|_X \otimes \rho_x$  and  $\sigma_{XB} = \sum_x q_x |x\rangle\langle x|_X \otimes \sigma_x$  are classical-quantum state [5]. Let the quantum channel  $\mathcal{N}$  be the partial trace channel on  $X$  system; then, by the monotonicity of quantum relative entropy ([5], Theorem 11.8.1), we have

$$D(\rho_{XB} || \sigma_{XB}) \geq D(\mathcal{N}(\rho_{XB}) || \mathcal{N}(\sigma_{XB})) = D(\rho || \sigma).$$

□

Notice that, if we let  $\sigma_x = \rho_x$  in Equation (7), with some calculation, Equation (7) becomes Lemma 3. Thus, Lemma 3 is a straightforward corollary of Lemma 4

**Theorem 1.** The dual problem in Equation (4) is equivalent to

$$\min_{\lambda \geq 0} F(\lambda) + \lambda S. \tag{8}$$

**Proof.** It follows from Equation (5) and Lemma 3 that

$$\max_{p'} f_\lambda(p, p') = f_\lambda(p, p) = H(\rho) - \sum_x p_x H(\rho_x) - \lambda s^T p.$$

Hence,

$$\begin{aligned} & \min_{\lambda \geq 0} \max_p H(\rho) - \sum_x p_x H(\rho_x) - \lambda(s^T p - S) \\ &= \min_{\lambda \geq 0} \max_p \max_{p'} f_\lambda(p, p') + \lambda S \\ &= \min_{\lambda \geq 0} F(\lambda) + \lambda S. \end{aligned}$$

□

The BA algorithm is an alternating optimization algorithm, i.e., to optimize  $f_\lambda(p, p')$ , each iteration step would fix one variable and optimize the function over the other variable. Now, we use BA algorithm to find  $F(\lambda)$ . The iteration procedure is

$$\begin{aligned} p_x^0 &> 0, \\ p_x^{tt} &= p_x^t, \\ p^{t+1} &= \arg \max_p \sum_x \text{Tr}\{p_x \rho_x [\log(p_x^t \rho_x) - \log(p_x \rho^t)]\} - \lambda s^T p, \end{aligned}$$

where  $\rho^t = \sum_x p_x^t \rho_x$ .

To get  $p^{t+1}$ , we can use the Lagrange function:

$$L = \sum_x \text{Tr}\{p_x \rho_x [\log(p_x^t \rho_x) - \log(p_x \rho^t)]\} - \lambda s^T p - \nu(\sum_x p_x - 1),$$

setting the gradient with respect to  $p_x$  to zero. By combining the normalization condition, we can have (taking the natural logarithm for convenience):

$$p_x^{t+1} = \frac{r_x^t}{\sum_x r_x^t}, \tag{9}$$

$$\begin{aligned} \text{where } r_x^t &= \exp(\text{Tr}\{\rho_x [\log(p_x^t \rho_x) - \log \rho^t]\} - s_x \lambda), \\ \rho^t &= \sum_x p_x^t \rho_x. \end{aligned} \tag{10}$$

Thus, we can summarize the Algorithm 1 below.

---

**Algorithm 1** Blahut–Arimoto algorithm for discrete memoryless classical-quantum channel.

---

set  $p_x^0 = \frac{1}{|\mathcal{X}|}$ ,  $x \in \mathcal{X}$ ;

**repeat**

$$p_x^t = p_x^t;$$

$$p_x^{t+1} = \frac{r_x^t}{\sum_x r_x^t}, \text{ where } r_x^t = \exp(\text{Tr}\{\rho_x [\log(p_x^t \rho_x) - \log \rho^t]\} - s_x \lambda);$$

**until** convergence.

---

**Lemma 5.** Let  $p^*(\lambda) = \arg \max_p f(p, p)$  for a given  $\lambda$ ; then,  $s^T p^*(\lambda)$  is a decreasing function of  $\lambda$ .

**Proof.** For convenience, we denote  $\chi(\{p_x : \rho_x\}_{x \in \mathcal{X}})$  as  $\chi(p)$ . Notice that  $f_\lambda(p, p) = \chi(p) - \lambda s^T p$  by definition of  $f(p, p)$ .

For  $\lambda_1 < \lambda_2$ , if  $\langle p^*(\lambda_1) | s \rangle < \langle p^*(\lambda_2) | s \rangle$ , then, by the definition of  $p^*(\lambda)$ , we have:

$$\begin{aligned} \chi(p^*(\lambda_1)) - \lambda_1 s^T p^*(\lambda_1) &\geq \chi(p^*(\lambda_2)) - \lambda_1 s^T p^*(\lambda_2) \\ \implies \chi(p^*(\lambda_2)) - \chi(p^*(\lambda_1)) &\leq \lambda_1 s^T (p^*(\lambda_2) - p^*(\lambda_1)) \\ &< \lambda_2 s^T (p^*(\lambda_2) - p^*(\lambda_1)) \\ \implies \chi(p^*(\lambda_1)) - \lambda_2 s^T p^*(\lambda_1) &> \chi(p^*(\lambda_2)) - \lambda_2 s^T p^*(\lambda_2), \end{aligned}$$

which is a contradiction to the fact that  $p^*(\lambda_2)$  is an optimizer of  $\chi(p) - \lambda_2 s^T p$ . Thus, we must have  $s^T p^*(\lambda_1) \geq s^T p^*(\lambda_2)$  if  $\lambda_1 < \lambda_2$ .  $\square$

We do not need to solve the optimization problem in Equation (8), because from Lemma 1 we can see that the statement “ $p^*$  is an optimal solution” is equivalent to “ $s^T p^* = S$  and  $p^*$  maximizes  $f_\lambda(p, p) + \lambda S = \chi(\{p_x, \rho_x\}_{x \in \mathcal{X}}) - \lambda(s^T p - S)$ ”, which is also equivalent to “ $s^T p^* = S$  and  $p^*$  maximizes  $f_\lambda(p, p)$ ”, thus, if for some  $\lambda \geq 0$ , a  $p$  maximizes  $f_\lambda(p, p)$  and  $s^T p = S$ , then the capacity  $C(S) = F(\lambda) + \lambda S$ , and such  $\lambda$  is easy to find since  $s^T p$  is a decreasing function of  $\lambda$ , and, to reach an  $\epsilon$  accuracy, we need

$$O(\log \epsilon) \tag{11}$$

steps using bisection method.

### 3. Convergence Analysis

Next, we show that the algorithm indeed converges to  $F(\lambda)$  and then provide an analysis of the speed of the convergence.

#### 3.1. The Convergence Is Guaranteed

##### Corollary 1.

$$f_\lambda(p^{t+1}, p^t) = \log \left( \sum_x r_x^t \right).$$

**Proof.**

$$\begin{aligned} f_\lambda(p^{t+1}, p^t) &= - \sum_x \text{Tr} \{ p_x^{t+1} \rho_x \log p_x^{t+1} \} + \sum_x \text{Tr} \{ p_x^{t+1} \rho_x [\log(p_x^t \rho_x) - \log(\rho^t)] \} - \lambda s^T p^{t+1} \\ &= - \sum_x p_x^{t+1} \log p_x^{t+1} + \sum_x p_x^{t+1} \log(r_x^t) \\ &= \sum_x p_x^{t+1} \log \left( \frac{r_x^t}{p_x^{t+1}} \right) \\ &= \log \left( \sum_x r_x^t \right). \end{aligned}$$

The first equality comes from a manipulation of Equation (5). The second equality follows from Equation (10). The last equality follows from Equation (9).  $\square$

**Corollary 2.** For arbitrary distribution  $\{p_x\}_{x \in \mathcal{X}}$ , we have

$$\chi(\{p_x, \rho_x\}_{x \in \mathcal{X}}) - \lambda s^T p - f(p^{t+1}, p^t) \leq \sum_x p_x \log \left( \frac{p_x^{t+1}}{p_x^t(x)} \right).$$

**Proof.** Define  $\rho = \sum_x p_x \rho_x$ . Then, we have

$$\begin{aligned}
 \sum_x p_x \log\left(\frac{p_x^{t+1}}{p_x^t}\right) &= \sum_x p_x \log\left(\frac{1}{p_x^t} \frac{r_x^t}{\sum_{x'} r_{x'}^t}\right) \\
 &= -f_\lambda(p^{t+1}, p^t) + \sum_x p_x \log \frac{r_x^t}{p_x^t} \\
 &= -f_\lambda(p^{t+1}, p^t) + \sum_x p_x \text{Tr}\{\rho_x [\log(p_x^t \rho_x) - \log \rho^t] - s_x \lambda - \log p_x^t\} \\
 &= -f_\lambda(p^{t+1}, p^t) + \sum_x p_x \text{Tr}\{\rho_x [\log \rho_x - \log \rho^t]\} - \lambda s^T p \\
 &= -f_\lambda(p^{t+1}, p^t) + \sum_x p_x \text{Tr}\{\rho_x [\log \rho_x - \log \rho + \log \rho - \log \rho^t]\} - \lambda s^T p \\
 &= -f_\lambda(p^{t+1}, p^t) + \chi(\{p_x, \rho_x\}_{x \in \mathcal{X}}) - \lambda s^T p + D(\rho || \rho^t).
 \end{aligned}
 \tag{12}$$

The first equality follows from Equation (9). The second equality follows from Corollary 1. The third equality follows from Equation (10). The last equality follows from Equation (1). Since the relative entropy  $D(\rho^X || \rho^t)$  is always non-negative [5], we have

$$\chi(\{p_x, \rho_x\}_{x \in \mathcal{X}}) - \lambda s^T p - f_\lambda(p^{t+1}, p^t) \leq \sum_x p_x \log\left(\frac{p_x^{t+1}}{p_x^t(x)}\right).$$

□

**Theorem 2.**  $f_\lambda(p^{t+1}, p^t)$  converges to  $F(\lambda)$  as  $t \rightarrow \infty$ .

**Proof.** Let  $p^*$  be an optimal solution that achieves  $F(\lambda)$ ; then, we have the following inequality

$$\begin{aligned}
 &\sum_{t=0}^N [F(\lambda) - f_\lambda(p^{t+1}, p^t)] \\
 &= \sum_{t=0}^N [\chi(\{p_x^*, \rho_x\}_{x \in \mathcal{X}}) - \lambda s^T p^* - f_\lambda(p^{t+1}, p^t)] \\
 &\leq \sum_{t=0}^N \sum_x p_x^* \log\left(\frac{p_x^{t+1}}{p_x^t}\right) \\
 &= \sum_x p_x^* \sum_{t=0}^N \log\left(\frac{p_x^{t+1}}{p_x^t}\right) \\
 &= \sum_x p_x^* \log\left(\frac{p_x^{N+1}}{p_x^0}\right) \\
 &= \sum_x p_x^* \log\left(\frac{p_x^*}{p_x^0}\right) + \sum_x p_x^* \log\left(\frac{p_x^{N+1}}{p_x^*(x)}\right) \\
 &= D(p^* || p^0) - D(p^* || p^{N+1}) \\
 &\leq D(p^* || p^0).
 \end{aligned}
 \tag{13}$$

The first equality follows from Equations (5), (6), and (1). The first inequality follows from Corollary 2. The last inequality follows from the non-negativity of relative entropy.

Thus, let  $N \rightarrow \infty$  and with  $F(\lambda) - f_\lambda(p^{t+1}, p^t) \geq 0$ , we have

$$0 \leq \sum_{t=0}^{\infty} [F(\lambda) - f_\lambda(p^{t+1}, p^t)] \leq D(p^* || p^0),
 \tag{15}$$

Notice we take the initial  $p^0$  to be uniform distribution, so the right hand side of Equation (15) is finite. Combine with the fact that  $f_\lambda(p^{t+1}, p^t)$  is a non-decreasing sequence, this means  $f_\lambda(p^{t+1}, p^t)$  converges to  $F(\lambda)$ .  $\square$

**Theorem 3.** *The probability distribution  $\{p^t\}_{t=0}^\infty$  also converges.*

**Proof.** Remove the summation over  $t$  in Equations (13) and (14); then, we have

$$0 \leq F(\lambda) - f_\lambda(p^{t+1}, p^t) \leq \sum_x p_x^* \log\left(\frac{p_x^{t+1}}{p_x^t}\right) = D(p^* || p^t) - D(p^* || p^{t+1}). \tag{16}$$

Now that the sequence  $\{p^t\}_{t=0}^\infty$  is a bounded sequence, there exists a subsequence  $\{p^{t_k}\}_{k=0}^\infty$  that converges. Let us say it converges to  $\bar{p}$ . Then, clearly, we have  $f(\bar{p}, \bar{p}) = F(\lambda)$  (or  $f(p^{t+1}, p^t)$  would not converge). Substituting  $p^* = \bar{p}$  in Equation (16), we have

$$0 \leq D(\bar{p} || p^t) - D(\bar{p} || p^{t+1}).$$

Thus, the sequence  $\{D(\bar{p} || p^t)\}_{t=0}^\infty$  is a decreasing sequence and there exists a subsequence  $\{D(\bar{p} || p^{t_k})\}_{k=0}^\infty$  that converges to zero. Therefore, we can conclude that  $\{D(\bar{p} || p^t)\}_{t=0}^\infty$  converges to zero, which means  $\{p^t\}_{t=0}^\infty$  converges to  $\bar{p}$ .  $\square$

### 3.2. The Speed of Convergence

**Theorem 4.** *To reach  $\epsilon$  accuracy to  $F(\lambda)$ , the algorithm needs an iteration complexity less than  $\frac{\log n}{\epsilon}$ .*

**Proof.** From the proof of Theorem 2, we know

$$\begin{aligned} \sum_{t=0}^N [F(\lambda) - f_\lambda(p^{t+1}, p^t)] &\leq D(p^* || p^0) \\ &= \sum_x p_x^* \log\left(\frac{p_x^*}{p_x^0}\right) = \log n - H(p^*) < \log n, \end{aligned}$$

and  $[F(\lambda) - f_\lambda(p^{t+1}, p^t)]$  is non-increasing in  $t$ , thus

$$F(\lambda) - f_\lambda(p^{t+1}, p^t) < \frac{\log n}{t}.$$

$\square$

Next, we show that in some special cases the algorithm has a better convergence performance, which is a geometric speed of convergence.

**Assumption 1.** *The channel matrices  $\{\rho_x\}_{x \in \mathcal{X}}$  are linearly independent, i.e., there does not exist a vector  $c \in \mathbb{R}^n$  such that*

$$\sum_x c_x \rho_x = 0.$$

**Remark 1.** *Assumption 1 is equivalent to: The output state  $\rho = \sum_x p_x \rho_x$  is uniquely determined by the input distribution  $p$ .*

**Theorem 5.** *Under Assumption 1, the optimal solution  $p^*$  is unique.*

**Proof.** Notice that the von Neumann entropy in Equation (2) is strictly concave [14], thus, for distributions  $p \neq p', \rho = \sum_x p_x \rho_x \neq \sum_x p'_x \rho_x = \rho'$ , which is followed from Assumption

refas1. Thus, this means  $H(\rho)$  is strictly concave in  $p$ . Thus, Holevo quantity in Equation (1) is strictly concave in  $p$ , which means the optimal solution  $p^*$  is unique.  $\square$

We also need the following theorem:

**Theorem 6.** [15] *The relative entropy satisfies*

$$D(\rho||\sigma) \geq \frac{1}{2} \text{Tr}(\rho - \sigma)^2.$$

Now, we state the theorem of convergence:

**Theorem 7.** *Suppose start from some initial point  $p^0$ , then, under Assumption 1, the algorithm converges to the optimal point  $p^*$ , and  $p^0$  converges to  $p^*$  at a geometric speed, i.e., there exist  $N_0$  and  $\delta > 0$ , where  $N$  and  $\delta$  are independent, such that, for any  $t > N_0$ , we have*

$$D(p^*||p^t) \leq (1 - \delta)^{t-N_0} D(p^*||p^{N_0}).$$

**Proof.** Define  $d_x = p_x^* - p_x^t$  and the real vector  $|d\rangle = (d_1, d_2, \dots, d_n)^T$ . Using Taylor expansion, we have

$$\begin{aligned} D(p^*||p^t) &= \sum_x p_x^* \log\left(\frac{p_x^*}{p_x^t}\right) = \sum_x -p_x^* \log\left(1 - \frac{d_x}{p_x^*}\right) \\ &= \frac{1}{2} \langle d|P|d\rangle + \sum_x O(d_x^3), \end{aligned}$$

where  $P = \text{diag}(p_1^*, p_2^*, \dots, p_n^*)$ . Now,  $p^t$  converges to  $p^*$ , i.e.,  $|d\rangle$  converges to zero, thus there exists a  $N_0$  such that, for any  $t > N_0$ , we have

$$D(p^*||p^t) \leq \frac{2}{3} \langle d|P|d\rangle. \tag{17}$$

From Theorem 6, we have

$$D(\rho^*||\rho^t) \geq \frac{1}{2} \text{Tr}\left\{\left[\sum_x d_x \rho_x\right]^2\right\} = \frac{1}{2} \langle d|M|d\rangle, \tag{18}$$

where  $M \in \mathbb{R}^{n \times n}$ :

$$M_{ij} = \text{Tr}(\rho_i \rho_j).$$

From Equation (18), we know that, under Assumption 1,  $M$  is positive definite. Thus, there exists a  $\delta > 0$  such that

$$\frac{1}{2} M > \delta \frac{2}{3} P \Rightarrow \frac{1}{2} \langle d|M|d\rangle > \delta \frac{2}{3} \langle d|P|d\rangle.$$

Thus, for any  $t > N_0$ , it follows from Equations (17) and (18) that

$$D(\rho^*||\rho^t) \geq \delta D(p^*||p^t). \tag{19}$$

From Equation (12), we know

$$\begin{aligned} \sum_x p_x^* \log\left(\frac{p_x^{t+1}}{p_x^t}\right) &\geq D(\rho^*||\rho^t) \\ \Rightarrow D(p^*||p^{t+1}) &\leq D(p^*||p^t) - D(\rho^*||\rho^t), \end{aligned}$$

combined with Equation (19), we have

$$\begin{aligned} D(p^*||p^{t+1}) &\leq D(p^*||p^t) - \delta D(p^*||p^t) = (1 - \delta)D(p^*||p^t) \\ \implies D(p^*||p^t) &\leq (1 - \delta)^{t-N_0} D(p^*||p^{N_0}) \end{aligned} \quad (20)$$

for any  $t > N_0$ .  $\square$

**Remark 2.** (Complexity). Denote  $n, m$  as the size of input alphabet and output state dimension, respectively. A closer look at Algorithm 1 reveals that, for each iteration, a matrix logarithm  $\log \rho^t$  needs to be calculated, and the rest are just multiplication of matrices and multiplication of numbers. The matrix logarithm can be done with complexity  $O(m^3)$  [16], thus, by Theorem 4 and Equation (11), the complexity to reach  $\varepsilon$ -close to the true capacity using Algorithm 1 is  $O(\frac{m^3 \log n \log \varepsilon}{\varepsilon})$ . With extra condition of the channel  $\{\rho_x\}_{x \in \mathcal{X}}$ , which is Assumption 1, the complexity to reach an  $\varepsilon$ -close solution (i.e.,  $D(p^*||p^t) < \varepsilon$ ) using Algorithm 1 is  $O(m^3 \log \varepsilon \log_{(1-\delta)} \frac{\varepsilon}{D(p^*||p^{N_0})})$ . Usually, we do not need  $\varepsilon$  to be too small (no smaller than  $10^{-6}$ ), thus, in either case, the complexity is better than  $O(\frac{(n \vee m)m^3(\log n)^{1/2}}{\varepsilon})$  in [6] when  $n \vee m$  is big, where  $n \vee m = \max\{n, m\}$ .

#### 4. Numerical Experiments on BA Algorithm

We only performed experiments on BA algorithm with no input constraint (BA algorithm with input constraint is some combination of BA algorithm with no input constraint.) We studied the relations between iteration complexity and  $n, m$  (i.e., the input size and output dimension) when the algorithm reaches certain accuracy. Since we do not know the true capacity of a certain channel, we used the following theorem to bound the error of the algorithm.

**Theorem 8.** With the iteration procedure in the BA Algorithm 1,  $\max_x \{D(\rho_x||\rho^t) - \lambda s_x\}$  converges to  $F(\lambda)$  from above.

**Proof.** Following from Algorithm 1, Corollary 1, and Theorem 3, we have

$$\lim_{t \rightarrow \infty} \frac{p_x^{t+1}}{p_x^t} = \exp[D(\rho_x||\rho^*) - \lambda s_x - F(\lambda)],$$

where  $\rho^* = \sum_x p_x^* \rho_x$ ,  $p^*$  is an optimal distribution. The limit above is 1 if  $p_x^* > 0$  and does not exceed 1 if  $p_x^* = 0$ . Thus,

$$D(\rho_x||\rho^*) - \lambda s_x \leq F(\lambda)$$

for every  $x \in \mathcal{X}$ , with equality if  $p_x^* > 0$ . This proves

$$\max_x \{D(\rho_x||\rho^t) - \lambda s_x\} \rightarrow F(\lambda).$$

For any  $p^t$  and any optimal distribution  $p^*$ , we have

$$\begin{aligned} \max_x [D(\rho_x||\rho^t) - \lambda s_x] &\geq \sum_x p_x^* [D(\rho_x||\rho^t) - \lambda s_x] \\ &= \sum_x p_x^* D(\rho_x||\rho^*) + D(\rho^*||\rho^t) - \lambda s^T p^* \\ &= F(\lambda) + D(\rho^*||\rho^t) \geq F(\lambda). \end{aligned}$$

The first equality requires some calculation and the second equality follows since  $p^*$  is an optimal distribution. This means  $\max_x \{D(\rho_x||\rho^t) - \lambda s_x\}$  converges to  $F(\lambda)$  from above.  $\square$

Thus, our accuracy criterion was: for a given classical-quantum channel, we ran the BA algorithm (with no input constraint), until  $[\max_x \{D(\rho_x || \rho^t)\} - f(p^t, p^t)]$  was less than  $10^{-k}$ , and recorded the number of iteration. At this time, the accuracy was of order  $10^{-(k+1)}$  at most since  $\max_x \{D(\rho_x || \rho^t)\}$  and  $f(p^t, p^t)$  converged to the true capacity from above and below, respectively.

We performed the following numerical experiments: for given values of input size  $n$ , output dimension  $m$  and accuracy, we generated 200 classical-quantum channels randomly, recorded the numbers of iterations, and then calculated the average number of iterations of these 200 experiments. The results are shown in Figure 1. Note that the accuracy  $10^{-k}$  in Figure 1 means we ran the BA algorithm until  $[\max_x \{D(\rho_x || \rho^t)\} - f(p^t, p^t)]$  was less than  $10^{-k}$ , and the error between the true capacity and the computed value was of order  $10^{-(k+1)}$  at most.

Input size	Output dim.	Accuracy	Ave. # iterations
6	2	e-5	161
		e-6	260
		e-7	399
	6	e-5	141
		e-6	259
		e-7	399
	10	e-5	94
		e-6	151
		e-7	200

Input size	Output dim.	Accuracy	Ave. # iterations
10	2	e-5	153
		e-6	438
		e-7	672
	6	e-5	265
		e-6	469
		e-7	1082
	10	e-5	236
		e-6	426
		e-7	848

Figure 1. The number of iterations needed to reach certain accuracy.

We can see in Figure 1 that the iteration complexity scales better as accuracy and input dimension increase. We can also see for given input size  $n$  and accuracy, the output dimension has vary little influence on iteration complexity, which means the iteration complexity also scales better as the output dimension  $m$  increases. Compared with our theoretical analysis of iteration complexity in Theorem 4: to reach  $\epsilon$  accuracy, we needed  $\frac{\log n}{\epsilon}$  iterations; the numerical experiments showed that the number of iterations was far smaller than  $\frac{\log n}{\epsilon}$  to reach  $\epsilon$  accuracy, whether the output quantum states were independent or not (cases in  $(n, m) = (6, 2), (10, 2)$ ). The reason for this is that the inequalities in the proof of Theorem 4 are quite loose. Thus, Theorem 4 only provides a very loose upper bound on iteration complexity. We can also guess that maybe the relation in Equation (20) holds generally and we just cannot prove it yet.

Next, we needed to see the running time of the BA algorithm. There were three methods to compute the classical-quantum channel capacity: BA algorithm, the duality and smoothing technique [6], and the method created by Hamza Fawzi et al. [17]. In [17], a Matlab code package called CvxQuad is provided which accurately approximates the relative entropy function via semidefinite programming so that many quantum capacity quantities can be computed using a convex optimization tool called CVX. Here, we compared the running time of the above three methods. For different input size  $n$  and output dimension  $m$ , we generated a classical-quantum channel randomly and computed the channel capacity using the above three methods and then recorded the running time of each method. The results are shown in Figure 2.

Input size	Output dim.	Duality & smoothing	CvxQuad	BA algorithm
6	2	1.29s	0.57s	0.076s
	6	7.32s	1.13s	0.10s
	10	17.35s	2.26s	0.082s
10	2	1.23s	0.67s	0.15s
	6	14.56s	1.25s	0.36s
	10	35.77s	2.30s	0.45s
30	10	\	2.5s	3.1s
	20	\	30s	8.9s
	30	\	130s	14s

**Figure 2.** The comparison of the running time of three methods.

In Figure 2, we can see the BA algorithm was the fastest method. The duality and smoothing method was rather slow and we did not record the running time of the duality and smoothing method when  $n = 30$  because it took too long. We can also notice that the running time of the CvxQuad method was extremely sensitive to the output dimension, which is not a surprise because CVX is a second-order method. Thus, our BA algorithm was significantly faster than the other two methods when  $n$  and  $m$  became big.

### 5. An Approximate Solution of $p$ in Binary Two Dimensional Case

In this section, we provide an approximate optimal input distribution for the case of the input size and output dimension are both 2:

$$\{p_1 : \rho_1; p_2 : \rho_2\}, \quad p_1 + p_2 = 1, \rho_1, \rho_2 \in \mathcal{D}^2.$$

#### 5.1. Use Bloch Sphere to Get an Approximate Solution

Any two-dimensional density matrix can be represented as a point in the Bloch sphere [5], as shown in the following:

Any density matrix can be represented as a vector in the Bloch sphere starting from the origin. Suppose  $\rho_1, \rho_2$  can be represented as  $r_1, r_2$  respectively, as shown in Figure 3; then, the two eigenvalues would be  $0.5 \pm r_1/2$  and  $0.5 \pm r_2/2$ , respectively. Extending  $r_1$ , we get two intersections on the surface of the Bloch sphere; then, these two intersections represent the two eigenvectors of  $\rho_1$  (the points on the surface of the sphere represent pure state and the interior points represent mixed states). A probabilistic combination of  $\rho_1, \rho_2$  can be represented as  $p_1\rho_1 + p_2\rho_2 = p_1r_1 + p_2r_2$  ([5] Exercise 4.4.13). Any point on the surface of Bloch sphere can be represented as

$$\cos \frac{\alpha}{2} |0\rangle + \sin \frac{\alpha}{2} e^{i\phi} |1\rangle,$$

where  $\alpha$  is the angle to the Z-axis and  $\phi$  is the angle of the X-axis to the projection of the point on the X-Y plane.

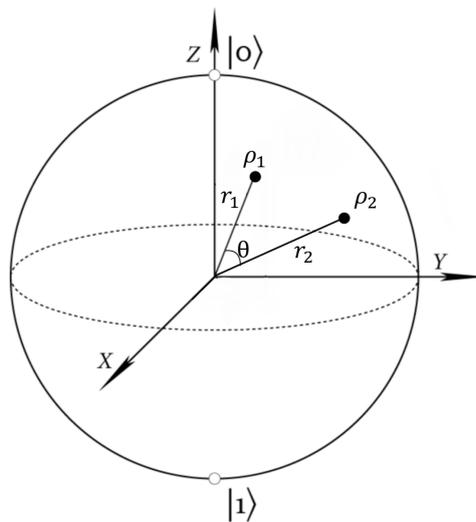


Figure 3. Bloch sphere.

By symmetry, it is obvious that the Holevo quantity is only related to  $r_1, r_2, \theta, p_1$ , where  $\theta$  is the angle between  $r_1, r_2$ . One interesting result is that the angle  $\theta$  has very little influence on  $p^*$ , where  $p^*$  is the optimal distribution that maximizes Holevo quantity. If we know  $\lambda_1, \lambda_2$  (the bigger eigenvalues of  $\rho_1, \rho_2$ , respectively),  $\theta$  and  $p_1$ , then the Holevo quantity can be written as

$$\chi(\lambda_1, \lambda_2, \theta, p_1) = S\left(\frac{1}{2} + \frac{1}{2}\|p_1 r_1 + (1 - p_1)r_2\|_2\right) - [p_1 S\left(\frac{1}{2} + \frac{1}{2}r_1\right) + (1 - p_1)S\left(\frac{1}{2} + \frac{1}{2}r_2\right)], \quad (21)$$

where  $S(\cdot)$  is the binary entropy ( $S(x) = -(x \log x + (1 - x) \log(1 - x))$ ) and  $r_i = 2\lambda_i - 1$ .

Using Cosine Theorem to calculate  $\|p_1 r_1 + (1 - p_1)r_2\|_2$ , the gradient of  $\chi$  with respect to  $p_1$  can be calculated directly, denoted as

$$\nabla_{p_1} \chi(\lambda_1, \lambda_2, \theta, p_1).$$

If we can find a  $p_1$  such that  $\nabla_{p_1} \chi(\lambda_1, \lambda_2, \theta, p_1) = 0$ , then this  $p_1$  is the optimal solution (because  $\chi(\lambda_1, \lambda_2, \theta, p_1)$  is concave in  $p_1$ ). However, we cannot solve the equation  $\nabla_{p_1} \chi(\lambda_1, \lambda_2, \theta, p_1) = 0$  with respect to  $p_1$  when  $\theta \neq 0$ . Now that  $\theta$  has little influence on  $p^*$ , let  $\theta = 0$  (this is actually the classical case), and let

$$\nabla_{p_1} \chi(\lambda_1, \lambda_2, \theta = 0, p_1) = 0,$$

the above equation is easy to solve and we get a solution  $\hat{p}_1$ :

$$\hat{p}_1 = \frac{\frac{1-c}{1+c} - r_2}{r_1 - r_2}, \quad \text{where } c = 2 \frac{S(\lambda_1) - S(\lambda_2)}{(r_1 - r_2)^2}, \quad (22)$$

where we assume  $r_1 \neq r_2$ . (It can be easily seen from the Bloch sphere that if  $r_1 = r_2$ , the optimal distribution would be  $\{\frac{1}{2}, \frac{1}{2}\}$ .)

This  $\hat{p}_1$  can be used as an approximate optimal solution for all  $\theta \in [0, \pi]$ . Next, we need numerical experiments to see how accurate  $\hat{p}_1$  is.

### 5.2. Numerical Experiments on the Approximated Solution $\hat{p}_1$

It is obvious that the maximum of Holevo quantity only depends on  $r_1, r_2$  and  $\theta$ , thus, without loss of generality, we let  $\rho_1$  be on the Z-axis and  $\rho_2$  be on the X-Z plane:

$$\begin{aligned} \rho_1 &= \lambda_1|0\rangle\langle 0| + (1 - \lambda_1)|1\rangle\langle 1|; \\ \rho_2 &= \lambda_2|\psi_0\rangle\langle \psi_0| + (1 - \lambda_2)|\psi_1\rangle\langle \psi_1|, \end{aligned}$$

where

$$\begin{aligned} |\psi_0\rangle &= \cos \frac{\theta}{2}|0\rangle + \sin \frac{\theta}{2}|1\rangle; \\ |\psi_1\rangle &= -\sin \frac{\theta}{2}|0\rangle + \cos \frac{\theta}{2}|1\rangle, \end{aligned}$$

which means the angle between  $\rho_1$  and  $\rho_2$  (i.e.,  $r_1$  and  $r_2$ ) is  $\theta$  [5].

In the numerical experiments, we let  $\lambda_1, \lambda_2$  range from 0.5 to 1, and  $\theta$  range from 0 to  $\pi$ . For each value of  $(\lambda_1, \lambda_2, \theta)$ , we substituted  $(\lambda_1, \lambda_2)$  into Equation (22) to compute  $\hat{p}_1$ . Then, we substituted  $(\lambda_1, \lambda_2, \theta, \hat{p}_1)$  into Equation (21) to get the approximate maximum of Holevo quantity over  $p_1$ :  $\chi(\lambda_1, \lambda_2, \theta, \hat{p}_1)$ . To see how accurate this approximate maximum is, we need a BA algorithm to provide an accurate maximum. The termination criterion for the iteration process of BA algorithm is stopping when  $[\max_x \{D(\rho_x || \rho^t)\} - f(p^t, p^t)]$  is less than  $10^{-6}$ ; then, the BA algorithm outputs a value of Holevo quantity  $\chi_{BA}(\lambda_1, \lambda_2, \theta)$ . We can compute the error of  $\chi(\lambda_1, \lambda_2, \theta, \hat{p}_1)$  then take the maximum over  $\theta \in [0, \pi]$

$$\text{Error}(\lambda_1, \lambda_2) = \max_{\theta \in [0, \pi]} |\chi(\lambda_1, \lambda_2, \theta, \hat{p}_1) - \chi_{BA}(\lambda_1, \lambda_2, \theta)|.$$

Figure 4 is the numerical result, which is a plot of  $(\lambda_1, \lambda_2, \text{Error}(\lambda_1, \lambda_2))$ .

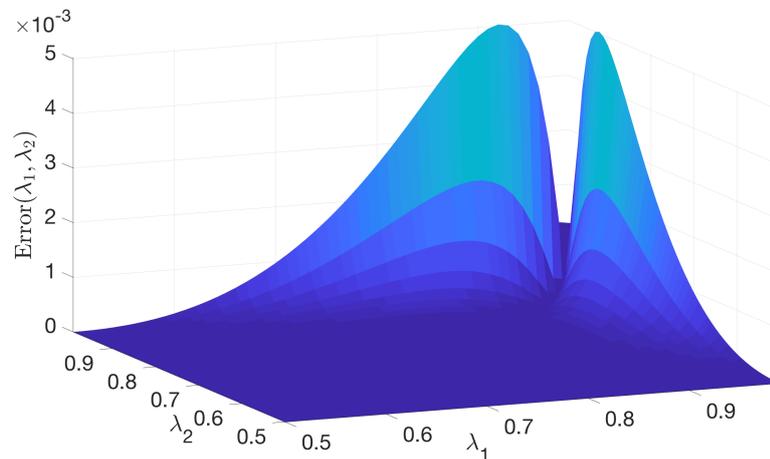
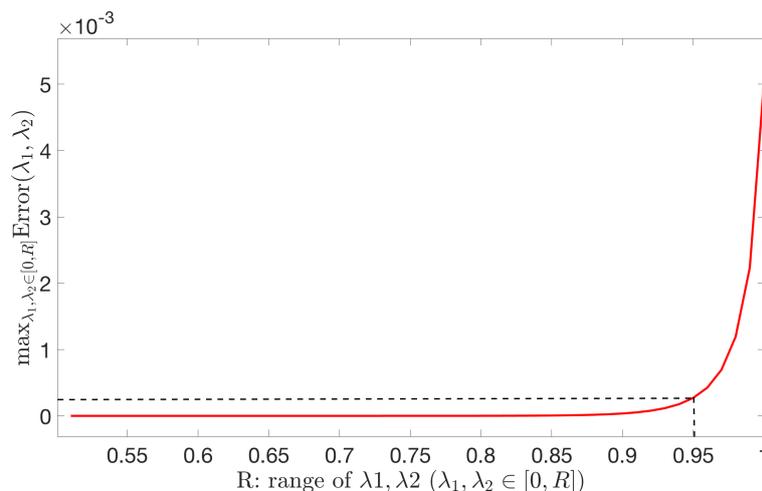


Figure 4. Error of the approximated method.

In Figure 4, we can see that, if  $\lambda_1, \lambda_2$  are not “too big”, the error can be upper bounded by  $10^{-3}$ . To see this more directly, we take the maximum of  $\text{Error}(\lambda_1, \lambda_2)$  for different ranges of  $\lambda_1, \lambda_2$ :

$$\max_{\lambda_1, \lambda_2 \in [0.5, R]} \text{Error}(\lambda_1, \lambda_2).$$

Figure 5 is a plot of  $(R, \max_{\lambda_1, \lambda_2 \in [0.5, R]} \text{Error}(\lambda_1, \lambda_2))$ .



**Figure 5.** Error of the approximate method.

In Figure 5, we can see that, if  $\lambda_1, \lambda_2 < 0.95$ , the error of approximate maximum of Holevo quantity can be upper bounded by  $3 \times 10^{-4}$ . Thus, we can conclude that, when the bigger eigenvalues of  $\rho_1, \rho_2$  are not too big (no bigger than 0.95), Equation (22) can make the error of the maximum of Holevo quantity smaller than  $3 \times 10^{-4}$ .

The approximate solution is an interesting phenomenon. The reason the angle  $\theta$  has such little influence on the maximum of Holevo quantity is unclear.

## 6. Discussion

In this paper, we provide an algorithm which computes the capacity of classical-quantum channel. We analyzed the speed of convergence theoretically and numerical experiments showed that our algorithm outperforms the existing methods [6,17]. We also provide an approximated method to compute the capacity of binary two-dimensional classical-quantum channel, which shows high accuracy. As mentioned in the Introduction, for a general quantum-quantum channel, maximizing Holevo quantity with respect to both the input distribution and output quantum state is NP-complete and this is not a convex optimization problem because Holevo quantity is concave with respect to input distribution and convex with respect to output quantum states. Thus, it remains open whether there exists an efficient algorithm to solve this problem. However, for classical-quantum channel, Holevo quantity has an upper bound, thus one future work would be to maximize Holevo quantity with respect to output states with given input distribution. It remains open if alternating optimization algorithms, in particular of Blahut–Arimoto type, can also be given for other optimization problems in terms of quantum entropy.

**Author Contributions:** Formal analysis, H.L., N.C.; Investigation, H.L.; Project administration, H.L.; Supervision, N.C.; Writing—original draft, H.L.; Writing—review & editing, N.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Arimoto, S. An algorithm for computing the capacity of arbitrary discrete memoryless channels. *IEEE Trans. Inf. Theory* **1972**, *18*, 14–20. [[CrossRef](#)]
2. Blahut, R. Computation of channel capacity and rate-distortion functions. *IEEE Trans. Inf. Theory* **1972**, *18*, 460–473. [[CrossRef](#)]

3. Holevo, A. Problems in the mathematical theory of quantum communication channels. *Rep. Math. Phys.* **1977**, *12*, 273–278. [[CrossRef](#)]
4. Holevo, A.S. The capacity of the quantum channel with general signal states. *IEEE Trans. Inf. Theory* **1998**, *44*, 269–273. [[CrossRef](#)]
5. Wilde, M. *From Classical to Quantum Shannon Theory*; Cambridge University Press: Cambridge, UK, 2016.
6. Sutter, D.; Sutter, T.; Esfahani, P.M.; Renner, R. Efficient Approximation of Quantum Channel Capacities. *IEEE Trans. Inf. Theory* **2016**, *62*, 578–598. [[CrossRef](#)]
7. Nesterov, Y. Smooth minimization of non-smooth functions. *Math. Program.* **2005**, *103*, 127–152. [[CrossRef](#)]
8. Yeung, R.W. *Information Theory and Network Coding*; Springer: Berlin/Heidelberg, Germany, 2008.
9. Nagaoka, H. Algorithms of Arimoto-Blahut type for computing quantum channel capacity. In Proceedings of the IEEE International Symposium on Information Theory (ISIT), Cambridge, MA, USA, 16–21 August 1998; p. 354.
10. Beigi, S.; Shor, P.W. On the Complexity of Computing Zero-Error and Holevo Capacity of Quantum Channels. *arXiv* **2007**, arXiv:0709.2090.
11. Osawa, S.; Nagaoka, H. Numerical Experiments on the Capacity of Quantum Channel with Entangled Input States. *IEICE TRANSACTIONS Fundam. Electron. Commun. Comput.* **2001**, *E84-A*, 2583–2590.
12. Hayashi, M.; Imai, H.; Matsumoto, K.; Ruskai, M.B.; Shimono, T. Qubit Channels Which Require Four Inputs to Achieve Capacity: Implications for Additivity Conjectures. *Quantum Inf. Comput.* **2005**, *5*, 13–31.
13. Boyd, S.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: Cambridge, UK, 2004.
14. Bengtsson, I.; Życzkowski, K. *GEOMETRY OF QUANTUM STATES: An Introduction to Quantum Entanglement*; Cambridge University Press: Cambridge, UK, 2006.
15. Hiai, F.; Petz, D. *Introduction to Matrix Analysis and Applications*; Springer: Berlin/Heidelberg, Germany, 2014.
16. Moler, C.; Van Loan, C. Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later. *SIAM Rev.* **2003**, *45*, 3–49. [[CrossRef](#)]
17. Hamza Fawzi, O.F. Efficient optimization of the quantum relative entropy. *J. Phys. A Math. Theor.* **2018**, *51*, 409601. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).