Alternating Maximization Algorithm for Mismatch Capacity with Oblivious Relaying

Xinwei Li[1], Lingyi Chen[1], Shitong Wu[1], Huihui Wu[2†], Hao Wu[1], and Wenyi Zhang[3]

[1]Department of Mathematical Sciences, Tsinghua University, Beijing 100084, China
[2]Yangtze Delta Region Institute (Huzhou),
University of Electronic Science and Technology of China, Huzhou, Zhejiang, 313000, China.
[3]Department of Electronic Engineering and Information Science,
University of Science and Technology of China, Hefei, Anhui 230027, China
Email: huihui.wu@ieee.org

Abstract—Reliable communication over a discrete memoryless channel with the help of a relay has aroused interest due to its widespread applications in practical scenarios. By considering the system with a mismatched decoder, previous works have provided optimization models to evaluate the mismatch capacity in these scenarios. The proposed models, however, are difficult due to the complicated structure of the mismatched decoding problem with the information flows in hops given by the relay. Existing methods, such as the grid search, become impractical as they involve finding all roots of a nonlinear system, with the growing size of the alphabet. To address this problem, we reformulate the max-min optimization model as a consistent maximization form, by considering the dual form of the inner minimization problem and the Lagrangian with a fixed multiplier. Based on the proposed formulation, an alternating maximization framework is designed, which provides the closed-form solution with simple iterations in each step by introducing a suitable variable transformation. The effectiveness of the proposed approach is demonstrated by the simulations over practical scenarios, including Quaternary and Gaussian channels. Moreover, the simulation results of the transitional probability also shed light on the promising application attribute to the quantizer design in the relay node.

Index Terms—Mismatch capacity, oblivious relay, information bottleneck, LM rate, alternating maximization.

I. INTRODUCTION

Reliable communication over a discrete memoryless channel (DMC) with a relay has roused great interest [1], due to its widespread applications in practical scenarios [2]–[4]. In communications with relaying, a transmitter sends a message to one or multiple oblivious agencies, and the message is then conveyed to the receiver over links with restricted capacities [5], [6]. The uplink Cloud Radio Access Network (C-RAN) is a typical application of this model and extensive investigations on the capacity of this scenario have been concerning the multi-user communications with base stations [7]–[9].

The capacity of the channels with oblivious relaying is commonly modeled by the Information Bottleneck (IB) framework [10], due to the similarity between the constraints of two optimization problems [5]. Specifically, the compression quality under the IB approach is evaluated by the mutual information between the compressed representation and the bottleneck variable, which inherently matches the channel capacity with oblivious relaying.

In this work, we focus on the channels with oblivious relaying, in the case of mismatched decoding [5], [6]. Recent work [5] derives a random coding capacity of the discrete memoryless information bottleneck channel (DM-IBC) with a mismatched decoder and the corresponding achievable bound of random codes is also analyzed. Specifically, the theoretical capacity of such a channel is proved to be objected by the LM (Lower [bound on the] Mismatch [capacity]) rate [11], [12], which is the tightest achievable rate via a constant-composition codebook ensemble [13]–[15], and constrained by the mutual information with upper bound similar to the IB problem.

The optimization models of the mismatch capacity with oblivious relaying is complicated due to the utilization of the LM rate as the objective function in the IB problem, instead of the mutual information. This implies that commonly employed algorithms for the IB problem [10], [16]–[18], cannot be directly applied. Moreover, the numerical methods for LM rate [12], [19] only provides a minimization solution of the inner problem, which leads to inefficiency when directly applied. To solve this problem, the authors of [5] proposed an algorithm named MMIB, resembling an adaptive grid search, which requires finding all roots of a nonlinear system constrained by mutual information. As the size of the input alphabet increases, this method will become intractable.

To address this problem, this paper develops an alternating maximization algorithm to efficiently compute the mismatch capacity with oblivious relaying. Specifically, to handle the optimization constraint, an IB-Lagrangian model with a fixed multiplier is proposed, by considering the Lagrange relaxation of the original problem. Additionally, we explore a dual formulation of the LM rate [19] from the optimal transport perspective, which transforms the optimization into a consistent maximization problem. To ensure an analytic form in each step of the algorithm, a variable transformation is further

introduced. It is noted that the fixed multiplier contributes to the convergence of the algorithm, as discussed later. Finally, the numerical experiments validate the efficiency and accuracy of the proposed algorithm. The results of the optimized probability distributions shed light on the promising application to the quantizer in practical scenarios.

II. PROBLEM FORMULATION

We consider a 3-node point-to-point communication system with a relay depicted in Fig. 1, referred to as the discrete memoryless information bottleneck channel (DM-IBC) [5].
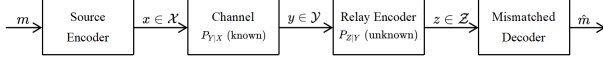

Fig. 1. A two-hop channel with oblivious relaying and a mismatched decoder

Under its model presuppositions, we denote the input alphabet $\mathcal{X} = \{x_i\}_{i=1}^M$, the received alphabet $\mathcal{Y} = \{y_k\}_{k=1}^K$, and the reproduced alphabet $\mathcal{Z} = \{z_j\}_{j=1}^N$. The conditional probability distribution between $\mathcal{X}$ and $\mathcal{Y}$ is denoted by $P_{Y|X}(y|x)$, which is only known by the encoder. With an input distribution $P_X \in \mathcal{P}(\mathcal{X})$, the output distribution at the relay node $P_Y \in \mathcal{P}(\mathcal{Y})$ and the joint distribution $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ are derived from the transition probability $P_{Y|X}$. After compression and transmission though an unknown relay encoder $P_{Z|Y}(z|y)$, the joint distribution $P_{XZ} \in \mathcal{P}(\mathcal{X} \times \mathcal{Z})$ is given by

$$P_{XZ}(x,z) = \sum_{y \in \mathcal{Y}} P_{Z|Y}(z|y)P_{Y|X}(y|x)P_X(x). \quad (1)$$

Given a transmission rate R and a compression rate B, the encoding scheme at the sender is described by a codebook of 2^{nR} n-dimensional sequence $x^n(m) \in \mathcal{X}^n$, while the encoding scheme at the relay source encoder is described by a codebook of 2^{nB} n-dimensional sequence $z^n(w) \in \mathcal{Z}^n$, When a message m is chosen uniformly from the message set $\mathcal{M} = \{1, 2, \cdots, 2^{nR}\}$, the sender transmits the corresponding codeword $x^n(m)$ to the relay source encoder. The source encoder at the relay receives $y^n \in \mathcal{Y}^n$, and then compresses it into $z^n(w)$, where $w \in \mathcal{W} = \{1, 2, \cdots, 2^{nB}\}$ is a representation set. The mismatched decoder then provides an estimate $\hat{m} \in \mathcal{M}$ of the intended transmission message based on the reproduced sequence $z^n(w)$ and the following prescribed decoding rule

$$\hat{m} = \arg\min_{j \in \mathcal{M}} \sum_{i=1}^n d(x_i(j), z_i),$$

where $d : \mathcal{X} \times \mathcal{Z} \to \mathbb{R}$ is called the mismatched metric [13].

With these notations, the random coding capacity of the DM-IBC with a mismatched decoder is defined as [5],

$$C_d(B) \triangleq \max_{P_X} \max_{P_{Z|Y}} \quad I_{\text{LM}}(X;Z)$$
$$\text{s.t.} \quad I(Y;Z) \leq B. \quad (2)$$

Here $I_{\text{LM}}(X;Z)$ is the LM rate defined via distribution P_{XZ},

which is actually an optimization problem

$$I_{\text{LM}}(X;Z)_{P_X,P_{Z|Y}} = \min_{\substack{\widetilde{P}_{XZ} \in \mathcal{P}(\mathcal{X} \times \mathcal{Z}):\widetilde{P}_X = P_X, \widetilde{P}_Z = P_Z \\ \mathbb{E}_{\widetilde{P}_{XZ}}[d(X,Z)] \leq \mathbb{E}_{P_{XZ}}[d(X,Z)]}} I_{\widetilde{P}}(X;Z). \quad (3)$$

The notation $\widetilde{P}$ is the auxiliary probability distribution corresponding to the variables to be optimized and the symbol d in $C_d(B)$ corresponds to the decoding metric $d(x,z)$.

It is noted that the optimization (2) presents the similar structure as the IB problem, with substituting the objective function with the LM rate, which imposes considerable challenge on solving this optimization problem. The LM rate (3) here is different from that in the usual sense, primarily because the distribution P_{XZ} has to be treated as a variable and is also linked via P_X and $P_{Z|Y}$ using (1). Moreover, the problem (2) possesses a max-min structure and it is non-convex. Thus, the solution to problem (2) is intricate.

The MMIB algorithm algorithm is proposed in [5] for solving the problem (2), including a step which entails finding all roots of a system of nonlinear equations. Finding all roots in this system is almost impossible in large-scale cases and even solving for part of the roots selectively still results in large computational costs. Therefore, a more efficient algorithm for this problem is in high demanding.

III. ALTERNATING MAXIMIZATION ALGORITHM

In this section, an efficient algorithm will be proposed for solving (2), with a prior reformulation processing of the model. We simplify the original formulation (2) via the IB-Lagrangian type relaxation and the dual form of the LM rate from the optimal transport perspective, which enables us to solve it in an alternating ascend manner.

A. Relaxation and Dual Form

The compression rate constraint in (2) involves unknown variable $P_{Z|Y}$, leading difficulty to compute directly. We therefore utilize a technique similar to the IB-Lagrangian in the BA algorithm [10], by introducing a fixed Lagrange multiplier λ corresponding to slope of the curve to evaluate the trade-off between the LM rate and the compression rate. This process yields a relaxation of the original problem

$$\max_{P_X} \max_{P_{Z|Y}} (I_{\text{LM}}(X;Z) - \lambda I(Y;Z)). \quad (4)$$

Note that the solution of problem (4) is also not straightforward, since it introduces the term $I_{\text{LM}}(X;Z)$ in the objective function. In the general case, the objective function of an IB problem is the mutual information with an analytic form. However, the LM rate here is an optimization, without an analytic expression.

Consequently, we introduce an optimal transport based dual form of the LM rate $I_{\text{LM}}(X;Z)$ [20], i.e.,

$$I_{\text{LM}}(\boldsymbol{\phi}, \boldsymbol{\psi}, \zeta; P_X, P_{Z|Y}) = \max_{\substack{\boldsymbol{\phi}, \boldsymbol{\psi} > 0, \\ \zeta \geq 0}} \left(-\boldsymbol{\phi}^T \boldsymbol{\Lambda} \boldsymbol{\psi} + \mathbb{E}_{P_{XZ}}(\log \boldsymbol{\Lambda}) \right.$$

$$\left. + \mathbb{E}_{P_X}(\log \boldsymbol{\phi}) + \mathbb{E}_{P_Z}(\log \boldsymbol{\psi}) + H(X) + H(Z) + 1 \right),$$

where $H(\cdot)$ denotes the information entropy, and $\boldsymbol{\Lambda} = e^{-\zeta \boldsymbol{d}}$.

Then, the problem (4) is transformed to a consistent double maximization problem

$$\max_{\substack{P_X, \\ P_{Z|Y}}} \max_{\substack{\phi,\psi>0, \\ \zeta\geq 0}} I_{\mathrm{LM}}(\phi,\psi,\zeta;P_X,P_{Z|Y}) - \lambda\,\mathbb{E}_{P_{YZ}}\!\left(\log\frac{P_{Z|Y}}{P_Z}\right),$$
(5)

where $P_{YZ}(y,z) = P_{Z|Y}(z|y)P_Y(y)$.

B. Alternating Maximization Algorithm

Since the reformulation of the consistent maximization form (5), we could present an algorithm to solve it by an alternating ascending approach. Furthermore, it is highlighted that a closed-form solution can still be obtained at each iteration step, despite the highly complex nature of the problem, which will be addressed in the subsequent discussion.

To simplify the formulations, we introduce the notations

$$p_i = Q_X(x_i), \quad q_k = P_Y(y_k), \quad r_j = P_Z(z_j),$$
$$D_{ij} = d(x_i, z_j), \; \Theta_{ik} = P_{Y|X}(y_k|x_i), \; \Omega_{kj} = P_{Z|Y}(z_j|y_k).$$

The bold lowercase letter, e.g., $\boldsymbol{r}$, denotes the corresponding vector of the probability distribution, while the bold uppercase letter, e.g., $\boldsymbol{\Omega}$, denotes the corresponding metric of the probability distribution. The notion $\odot$ represents the pointwise multiplication. The notion $\boldsymbol{q}\,./\,\boldsymbol{p}$ represents the division of each element of $\boldsymbol{q}$ by the corresponding element of $\boldsymbol{p}$.

The objective function in (5) contains the entropy term $H(Z) = -(\boldsymbol{\Omega}^T\boldsymbol{\Theta}^T\boldsymbol{p})^T\log(\boldsymbol{\Omega}^T\boldsymbol{\Theta}^T\boldsymbol{p})$, making it difficult to obtain a closed-form solution for updating $\boldsymbol{p}$. Thus, we utilize a variable transformation $\widetilde{\boldsymbol{\psi}} = \boldsymbol{\psi}./(\boldsymbol{\Omega}^T\boldsymbol{\Theta}^T\boldsymbol{p})$. Consequently, the relaxation model (5) is equivalent to the following maximization problem

$$\max_{\boldsymbol{p},\boldsymbol{t},\boldsymbol{r}} \max_{\substack{\phi,\widetilde{\psi}, \\ \zeta\geq 0}} -\boldsymbol{p}^T\log\boldsymbol{p} + \boldsymbol{p}^T\log\boldsymbol{J}(\phi,\widetilde{\psi},\zeta;\boldsymbol{\Omega},\boldsymbol{r}) + 1$$
(6)
$$\text{s.t. } \boldsymbol{\Omega}\mathbf{1}_N = \mathbf{1}_K,\; \|\boldsymbol{r}\|_1 = 1,\; \|\boldsymbol{p}\|_1 = 1,$$

where the function $\boldsymbol{J}(\phi,\widetilde{\psi},\zeta;\boldsymbol{\Omega},\boldsymbol{r})$ is defined as

$$\boldsymbol{J} = \phi\odot\exp\left((\boldsymbol{\Theta\Omega})\left[-(\boldsymbol{\Lambda}^T\phi)\odot\widetilde{\psi}+\log\widetilde{\psi}\right]\right.$$
$$\left. -\zeta\left[(\boldsymbol{\Theta\Omega})\odot\boldsymbol{D}\right]\mathbf{1}_N - \lambda\boldsymbol{\Theta}\left[\boldsymbol{\Omega}\odot\left(\log\boldsymbol{\Omega}-\log\left(\mathbf{1}_K\boldsymbol{r}^T\right)\right)\right]\right).$$

The main derivation of the algorithm is presented as follows.

1) Update the input distribution P_X: First, we consider the optimization problem with respect to $\boldsymbol{p}$ only, and construct its Lagrangian:

$$\mathcal{L}_p(\boldsymbol{p};\eta_p) = H(\boldsymbol{p}) + \mathbb{E}_{P_X}\left(\log\boldsymbol{J}(\phi,\widetilde{\psi},\zeta;\boldsymbol{\Omega},\boldsymbol{r})\right) + 1$$
(7)
$$-\eta_p\left(\|\boldsymbol{p}\|_1 - 1\right),$$

where $\eta_p \in \mathbb{R}$ is a dual variable. Following the first-order condition and further analyzing the dual form, we can update $\boldsymbol{p}$ according to $p_i = J_i/\sum_{i'=1}^M J_{i'}$.

2) Update the joint distribution between Y and Z: Considering the first-order condition with respect to $\Omega_{kj} = P_{Z|Y}(z_j|y_k)$, we could obtain an optimal solution. In conjunction with the constraints in (6), variables $\boldsymbol{\Omega}$ can be updated through the normalized condition corresponding to the conditional probability distribution, that is $\Omega_{kj} = \Omega_{kj}^*/\left(\sum_{j'=1}^N \Omega_{kj'}^*\right)$, where

$$\boldsymbol{\Omega}^* = \left(\mathbf{1}_K\left[\boldsymbol{r}\odot(\widetilde{\psi}^{\frac{1}{\lambda}})\odot\exp\left(\frac{1}{\lambda}(-\boldsymbol{\Lambda}^T\phi)\odot\widetilde{\psi}\right)\right]^T\right)$$
$$\odot\exp\left(-\frac{\zeta}{\lambda}(\boldsymbol{\Theta}^T\mathrm{Diag}(\boldsymbol{p})\boldsymbol{D})./(\boldsymbol{\Theta}^T\boldsymbol{p}\mathbf{1}_N^T)\right).$$

In a similar manner, the vector $\boldsymbol{r}$ can be updated according to $\boldsymbol{r} = \boldsymbol{\Omega}^T\boldsymbol{\Theta}^T\boldsymbol{p}$.

3) Update the dual variable of the LM rate: For the fixed input distribution $p_i = P_X(x_i)$ and joint distribution $\Omega_{kj} = \widetilde{P}_{Z|Y}(z_j|y_k)$, this process is equivalent to computing the LM rate between X and Z. Similar to the previous work [19], we update the dual variables alternatively by

$$\phi_i = \frac{p_i}{\sum_{j=1}^N e^{-\zeta D_{ij}}\widetilde{\psi}_j r_j}, \quad \widetilde{\psi}_j = \frac{1}{\sum_{i'=1}^M \phi_{i'}e^{-\zeta D_{i'j}}}.$$

Also, ζ is updated by finding the root of the following monotone function

$$G(\zeta) \triangleq \phi\left(\boldsymbol{D}\odot\boldsymbol{\Lambda}\right)\left(\widetilde{\psi}\odot\left(\boldsymbol{\Omega}^T\boldsymbol{\Theta}^T\boldsymbol{p}\right)\right) - \boldsymbol{p}^T\left((\boldsymbol{\Theta\Omega})\odot\boldsymbol{D}\right)\mathbf{1}_N,$$
(8)

where $\boldsymbol{\Lambda} = e^{-\zeta\boldsymbol{D}} = e^{-\zeta\boldsymbol{d}}$ as defined before.

We summarise the proposed algorithm in the following pseudo-code in Algorithm 1. The derivation details of the AM algorithm are not difficult but is too long to give here. Due to space constraints, we put it in the appendix [1].

Algorithm 1 Alternating Maximization (AM) Algorithm

Input: Decoding metric D_{ij}, Conditional probability distribution Θ_{ik}; The Lagrange multiplier λ; Iteration number max_iter.

1: **Initialize:** $\phi^{(0)} = \mathbf{1}_M$, $\widetilde{\psi}^{(0)} = \mathbf{1}_N$, $\zeta = 1$, $\mu = 1$; **Randomly Initialize:** $\boldsymbol{\Omega}^{(0)}$, $\boldsymbol{r}^{(0)}$.

2: **for** $l = 1 : max\_iter$ **do**

3:      Update $p_i^{(l)} \leftarrow J_i/\left(\sum_{i'=1}^M J_{i'}\right)$

4:      Update $\Omega_{kj}^{(l)} \leftarrow \Omega_{kj}^*/\left(\sum_{j'=1}^N \Omega_{kj'}^*\right)$

5:      Update $r_j^{(l)} \leftarrow \sum_{i=1}^M\sum_{k=1}^K \Omega_{kj}^{(l)}\Theta_{ik}p_i^{(l)}$

6:      Update $\phi_i^{(l)} \leftarrow p_i^{(l)}/\left(\sum_{j=1}^N e^{-\zeta D_{ij}}\widetilde{\psi}_j^{(l-1)}r_j^{(l)}\right)$

7:      Update $\widetilde{\psi}_j^{(l)} \leftarrow 1/\left(\sum_{i=1}^M \phi_i^{(l)}e^{-\zeta D_{ij}}\right)$

8:      Solve $G(\zeta) = 0$ for $\zeta\in\mathbb{R}^+$ with Newton's method

9: **return** C_d

Theorem 1. *The proposed algorithm ensures the convergence of all iteration variables.*

[1]https://github.com/wushitong1234/ADM-IBC

Proof. In every iteration, the variables undergo individual updates based on their first-order conditions respectively. It is noteworthy that the objective function (6) is concave with respect to the variables p, Ω, ϕ, ψ, and ζ separately. Thus, the first-order condition is equivalent to identifying the maximum value of the objective function (5) at each iteration.

The value of the objective function demonstrates a sustained increase through the alternating maximization of individual variables. In the meanwhile, the objective function signifies the LM rate between X and Z, with its upper bound substantiated in [12]. Consequently, the algorithm converges. $\qquad\square$

Remark 1. *An average power constraint like $\mathbb{E}\left[X^2\right] \leq \Gamma$ may be considered under the AWGN channel. In this case, we can also design a similar Alternating Maximization algorithm based on the relaxation form* (5)*, following the analysis of the corresponding Lagrangian.*

IV. NUMERICAL RESULTS

This section applies the proposed AM algorithm to compute the mismatch capacity C_d of the DM-IBC model.

All the experiments are conducted on a PC with 8G RAM and one Intel(R) Core(TM) Gold i5-8265U CPU @1.60GHz.

A. An Example over a Quaternary Channel

First, we use the AM algorithm to compute (6) over a Quaternary channel, and compare the results with the MMIB algorithm of [5]. The transition law and the decoding rule are adopted the same settings with [5, Eq.(8)] and [5, Eq.(9)] respectively. Here the input distribution P_X is assumed to be uniform, and this yields the computational result R_q is an achievable rate. The index q represents the decoding rule $q = e^{-d}$.
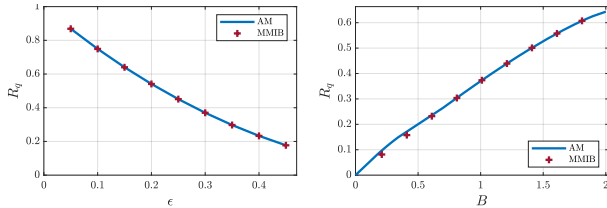

Fig. 2. Left: R_q versus channel parameter ϵ (as defined in [5]) with a fixed compression rate $B = 1.0$. Right: R_q versus compression rate B with a fixed channel parameter $\epsilon = 0.3$.

The achievable rates R_q for different cases of the Quaternary channel are demonstrated in Fig. 2. In addition, Table I compares the computational time of the AM algorithm and the MMIB algorithm. The algorithm terminates when two consecutive iterations yield a residual less than 10^{-5}.

It is obvious that the AM algorithm is superior to the MMIB algorithm, in terms of the computational time. Further, the results of the MMIB algorithm depend on the choice of search parameters and the initialization. However, these issues do not arise in the AM algorithm, which could converge stably with most random initial values.

TABLE I
EFFICIENCY COMPARISON BETWEEN AM AND MMIB ALGORITHMS

(ϵ, B)	Computational time (s)		Speed-up ratio	Feasibility	
	t_{MMIB}	t_{AM}		MMIB	AM
(0.3, 0.41)	78	8.6	9	7e-9	8e-10
(0.3, 0.81)	82	0.5	164	5e-9	1e-10
(0.4, 0.41)	118	1.6	74	7e-9	2e-10
(0.4, 0.81)	134	7.5	18	6e-9	9e-10

Columns 2-3 are the Computational time of two algorithms. Columns 5-6 indicate whether the compression constraint is satisfied.

Remark 2. *In Table I, t_{AM} denotes the total computational time, including the binary search for the multiplier λ, since the AM algorithm requires a fixed multiplier, similar to that in the Blahut-Arimoto (BA) algorithm [16]. This does not affect the efficiency advantage of our algorithm.*

B. Examples over Gaussian Channels with IQ Imbalances

With a further consideration of optimizing the input probability distribution P_X, applying a similar manner of the MMIB algorithm results in a more intricate search structure. Also, as the increasing number of discrete points of the alphabets, it may exceed the computational ability of a computer, due to solving multiple roots of a system of nonlinear equations. Thus, we only adopt the AM algorithm in later experiments.

In this subsection, the channel between X and Y is configured as an additive white Gaussian noise (AWGN) channel with IQ imbalances, modeled by

$$Y = HX + N, \quad N \sim \mathcal{N}(\mathbf{0}, \sigma_N^2 I).$$

The channel metric $H \in \mathbb{R}^{2 \times 2}$ is a combination of rotation and scaling effects as

$$H = \begin{pmatrix} 1 & 0 \\ 0 & \eta \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix},$$

in which the parameter η, respectively, θ denotes the scaling and the degree of the rotation on the signal. Moreover, we set the signal-to-noise ratio $\text{SNR} \triangleq 10\log_{10}\left(1/2\sigma_N^2\right)$.

Under mismatched cases, the values of parameters η and θ are unknown at the decoder. The decoding metric is $d(\boldsymbol{x}, \boldsymbol{z}) = \|\boldsymbol{z} - \hat{H}\boldsymbol{x}\|_2^2$, which is the distance between the expected observation points and the actual reception points. Note that $\hat{H}$ is an estimate of H, and $\hat{H} = I$ in mismatch cases.

This work examines the QPSK and 16QAM modulation schemes. The constellation points (i.e., the channel input alphabet $\mathcal{X}$) of all the examined modulation schemes are normalized as in [19], namely $\sum_{i=1}^{M} \frac{1}{M}\|x_i\|^2 = 1$. It is important to note that, for the input alphabet $\mathcal{X}$, the uniform distribution in the power constraint is solely employed to predetermine the coordinates of the collection points. When computing the mismatch capacity, the input distribution P_X is unknown and is determined during the iterations.

The constellation points for the post-compressed set $\mathcal{Z}$ are configured to be the same positions as those in $\mathcal{Y}$, the points before compression. Note that the alphabet $\mathcal{Y}$ is continuous

on $\mathbb{R}^2$, and the same applies to $\mathcal{Z}$. We restrict it to sufficiently large region $[-8, 8] \times [-8, 8]$, and discretize it by,

$$z_{r\sqrt{N}+s} = (-8 + r\Delta z, -8 + (s-1)\Delta z), \ \Delta z = \frac{16}{\sqrt{N}-1},$$
$$r = 0, 1, \cdots, \sqrt{N}-1, \ s = 1, 2, \cdots, \sqrt{N},$$

where $\{z_j\}_{j=1}^N$ is a set of uniform grid points corresponding to the alphabet $\mathcal{Z}$. The number of discrete points in the alphabet of $\mathcal{Y}$ and $\mathcal{Z}$ are set to $N = 2500$, unless otherwise specified.

First, we examine the convergence property of the proposed algorithm by analyzing the residual errors

$$r_\phi = \sum_{i=1}^M \left| \phi_i \sum_{j=1}^N e^{-\zeta D_{ij}} \widetilde{\psi}_j r_j - p_i \right|, \tag{9a}$$

$$r_{\widetilde{\psi}} = \sum_{j=1}^N \left| \left(\widetilde{\psi}_j \sum_{i=1}^M \phi_i e^{-\zeta D_{ij}} - 1 \right) r_j \right|, \tag{9b}$$

$$r_\zeta = |G(\zeta)|, \quad r_\mu = |F(\mu)|. \tag{9c}$$

Fig. 3 illustrates the convergent trajectories of residual errors versus iteration steps, with the parameters $(\eta, \theta) = (0.9, \pi/18)$, SNR = 10 dB, and a fixed Lagrange multiplier $\lambda = 0.25$. The algorithm converges to 10^{-5} for all cases.
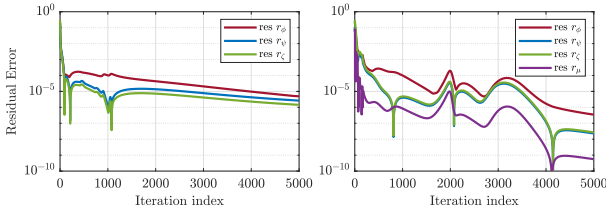

Fig. 3. The convergent trajectories of the residual errors r_ϕ (Red), r_ψ (Blue), r_ζ (Green), and r_μ (Purple). Left: The QPSK modulation scheme. Right: The 16QAM modulation scheme.

To facilitate an intuitive understanding of the relaying process between Y and Z, we plot the values of the optimized conditional probability $P_{Z|Y}$ in Fig. 4, with the discrete points $N = 225$ for clarity and simplity. The vertical and horizontal coordinates correspond to the indices of the discrete points in the alphabets $\mathcal{Y}$ and $\mathcal{Z}$, respectively. The results give the quantizer design at the relay node to some extent.
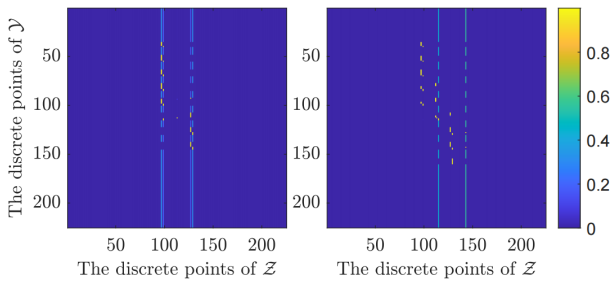

Fig. 4. Values of optimized conditional probability distribution $P_{Z|Y}$. Left: The QPSK modulation scheme. Right: The 16QAM modulation scheme.

Fig. 5 compares the trade-off between the mismatch capacity C_d and $I(Y; Z)$ in two cases, including a fixed parameter $(\eta, \theta) = (0.9, \pi/18)$ with different SNR values, and different mismatch cases $(\eta, \theta) = (0.9, \pi/18), (0.8, \pi/18), (0.9, \pi/12)$ and $(0.8, \pi/12)$ with a fixed SNR = 10 dB. The algorithm terminates after 5000 iterations, or when two consecutive iterations yield a residual less than 10^{-8}.
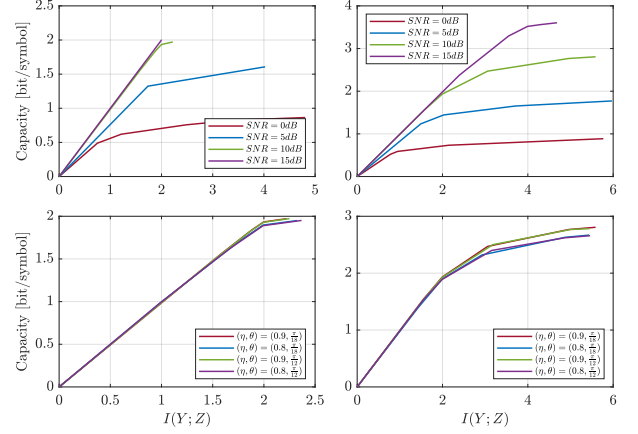

Fig. 5. Mismatch capacity C_d versus compression rate under different cases. Left: The QPSK modulation scheme. Right: The 16QAM modulation scheme.

As illustrated, the channel capacity C_d rises with increasing SNR values at the same compression rate. The observed gap between the capacity of different mismatch cases becomes more noticeable with growing compression rates.

V. CONCLUSION

This work presents an algorithm to evaluate the channel capacity of a 3-node point-to-point communication system with a relay. Such a channel capacity takes the LM rate as the objective function, constrained by an IB-based problem, which makes the corresponding computation complicated. To address the nonlinear constraint in the original optimization problem, an IB-Lagrangian with a fixed multiplier is introduced. We then proceed to examine the unconstrained relaxation form of the original optimization. Further, a dual form of the LM rate from the perspective of optimal transport is utilized, which enables the transformation of the problem into a consistent maximization form. By doing so, the alternating maximization is applicable, where each iteration step could be efficiently solved. Moreover, the convergence property of the proposed algorithm is proved. Finally, simulation results demonstrate the effectiveness of the proposed algorithm, and gives promising application to practical scenarios like the quantizer design.

REFERENCES

[1] E. C. van der Meulen, "Three-terminal communication channels," *Advances in Applied Probability*, vol. 3, pp. 120–154, 1971.

[2] T. Q. Quek, M. Peng, O. Simeone, and W. Yu, *Cloud radio access networks: Principles, technologies, and applications*. Cambridge University Press, 2017.

[3] R. Wang and X. Yuan, "An overview on multiway relay communications: Fundamental issues, recent advances, and new challenges," *arXiv preprint arXiv:2010.04407*, 2020.

[4] G. Dietl, M. Sciora, G. Zeitler, G. Bauch, and J. Widmer, "A quantize-and-forward scheme for future wireless relay networks," in *2011 IEEE Vehicular Technology Conference (VTC Fall)*, 2011, pp. 1–4.

[5] M. Dikshtein, N. Weinberger, and S. Shamai, "On mismatched oblivious relaying," *2023 IEEE International Symposium on Information Theory (ISIT)*, pp. 1687–1692, 2023.

[6] C. Hucher and P. Sadeghi, "Mismatched decoding for the relay channel," in *International Zurich Seminar on Communications - Proceedings*. Eidgenössische Technische Hochschule Zürich, 2010.

[7] I. E. Aguerri, A. Zaidi, G. Caire, and S. Shamai, "On the capacity of cloud radio access networks with oblivious relaying," *IEEE Transactions on Information Theory*, vol. 65, pp. 4575–4596, 2019.

[8] D. Chen and V. Kühn, "Alternating information bottleneck optimization for the compression in the uplink of C-RAN," *2016 IEEE International Conference on Communications*, pp. 1–7, 2016.

[9] H. Xu, T. Yang, G. Caire, and S. Shamai (Shitz), "Information bottleneck for a Rayleigh fading MIMO channel with an oblivious relay," *Information*, vol. 12, no. 4, 2021.

[10] N. Tishby, C. Pereira, and W. Bialek, "The information bottleneck method," *Proceedings of the 37th Allerton Conference on Communication, Control and Computation*, vol. 49, 07 2001.

[11] I. Csiszár and J. Körner, "Graph decomposition: A new key to coding theorems," *IEEE Transactions on Information Theory*, vol. 27, no. 1, pp. 5–12, Jan. 1981.

[12] J. Scarlett, A. Guillèn i Fàbregas, A. Somekh-Baruch, and A. Martinez, "Information-theoretic foundations of mismatched decoding," *Foundations and Trends® in Communications and Information Theory*, vol. 17, no. 2-3, pp. 149–400, 2020.

[13] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai (Shitz), "On information rates for mismatched decoders," *IEEE Transactions on Information Theory*, vol. 40, no. 6, pp. 1953–1967, Nov. 1994.

[14] I. Csiszar and P. Narayan, "Channel Capacity for a Given Decoding Metric," *IEEE Transactions on Information Theory*, vol. 41, no. 1, pp. 35–43, Jan. 1995.

[15] I. Csiszár and J. Körner, *Information theory: Coding theorems for discrete memoryless Systems*. Cambridge, UK: Cambridge University Press, 2011.

[16] R. E. Blahut, "Computation of Channel Capacity and Rate-Distortion Functions," *IEEE Transactions on Information Theory*, vol. 18, no. 4, pp. 460–473, Jan. 1972.

[17] L. Chen, S. Wu, W. Ye, H. Wu, H. Wu, W. Zhang, B. Bai, and Y. Sun, "Information Bottleneck Revisited: Posterior Probability Perspective with Optimal Transport," in *2023 IEEE International Symposium on Information Theory (ISIT)*, Taipei, Taiwan, China, Jun. 2023, pp. 1490–1495.

[18] L. Chen, S. Wu, J. Ye, H. Wu, W. Zhang, and H. Wu, "Efficient and Provably Convergent Computation of Information Bottleneck: A Semi-Relaxed Approach," *IEEE International Conference on Communications (ICC)*, 2024, accepted.

[19] W. Ye, H. Wu, S. Wu, Y. Wang, W. Zhang, H. Wu, and B. Bai, "An optimal transport approach to the computation of the LM rate," in *Proc. 2022 IEEE Global Communications Conference (GLOBECOM)*, Rio de Janeiro, Brazil, Dec. 2022, pp. 239–244.

[20] L. Chen, S. Wu, X. Li, H. Wu, H. Wu, and W. Zhang, "A double maximization approach for optimizing the LM rate of mismatched decoding," *IEEE International Symposium on Information Theory (ISIT)*, 2024, accepted.

The detailed derivation of the AM algorithm will be presented in the following, as mentioned in Sec. III-B.

A. Update the input distribution P_X

Consider the optimization problem (6) with respect to $\boldsymbol{p}$ only, and recall its Lagrangian (7). Taking the partial derivative of the Lagrangian $\mathcal{L}_p(\boldsymbol{p}; \eta_p)$ with respect to $\boldsymbol{p}$, we obtain the optimal solution

$$\widetilde{p}_i = J_i e^{-1-\eta_p}.$$

In conjunction with the constraints $\sum_{i=1}^M p_i = 1$, we can update $\boldsymbol{p}$ according to

$$p_i = \frac{J_i}{\sum_{i'=1}^M J_{i'}}, \tag{10}$$

where the dual variable η_p is vanished.

B. Update the joint distribution between Y and Z

First, consider the optimization problem with respect to $\Omega_{kj} = \widetilde{P}_{Z|Y}(z_j|y_k)$, and construct its Lagrangian

$$
\mathcal{L}_t(\boldsymbol{\Omega}; \boldsymbol{\eta_t}) = -\sum_{i=1}^M p_i \log p_i + \sum_{i=1}^M p_i \log J_i(\boldsymbol{\phi}, \widetilde{\boldsymbol{\psi}}, \zeta; \boldsymbol{\Omega}, \boldsymbol{r}) \\
+ 1 - \sum_{k=1}^K \eta_{t_k} \left(\sum_{j=1}^N \Omega_{kj} - 1 \right), \tag{11}
$$

where $\{\eta_{t_k}\}_{k=1}^K \subset \mathbb{R}$ are dual variables. Taking the partial derivative of the Lagrangian (11) with respect to Ω_{kj}, we obtain the optimal solution

$$
\widetilde{\Omega_{kj}} = r_j * \exp\left(\frac{1}{\lambda} \left(-\sum_{i=1}^M \phi_i e^{-\zeta D_{ij}} \widetilde{\psi_j} + \log \widetilde{\psi_j} \right) \right) \\
* \exp\left(-\frac{\zeta \sum_{i=1}^M D_{ij} \Theta_{ik} p_i}{\lambda \sum_{i=1}^M \Theta_{ik} p_i} \right) * \exp\left(1 - \frac{\eta_{t_k}}{\lambda \sum_{i=1}^M \Theta_{ik} p_i} \right).
$$

In conjunction with the constraints $\sum_{j=1}^N \Omega_{kj} = 1$, we could update Ω_{kj} according to

$$\Omega_{kj} = \frac{\Omega_{kj}^*}{\left(\sum_{j'=1}^N \Omega_{kj'}^* \right)},$$

where Ω_{kj}^* was defined in the main text.

Subsequently, consider the optimization problem with respect to r_j, and construct its Lagrangian

$$
\mathcal{L}_r(\boldsymbol{r}; \eta_r) = -\sum_{i=1}^M p_i \log p_i + \sum_{i=1}^M p_i \log J_i(\boldsymbol{\phi}, \widetilde{\boldsymbol{\psi}}, \zeta; \boldsymbol{\Omega}, \boldsymbol{r}) \\
+ 1 - \eta_r \left(\sum_{j=1}^N r_j - 1 \right), \tag{12}
$$

where $\eta_r \in \mathbb{R}$ is a dual variable. Taking the partial derivative of the Lagrangian (12) with respect to r_j, we obtain the optimal solution

$$\widetilde{r}_j = \frac{\lambda}{\eta_r} \sum_{i=1}^M \sum_{k=1}^K \Omega_{kj} \Theta_{ik} p_i. \tag{13}$$

In conjunction with the constraints in (6), we have

$$\sum_{j=1}^N r_j = 1, \quad \sum_{i=1}^M \sum_{k=1}^K \sum_{j=1}^N \Omega_{kj} \Theta_{ik} p_i = 1. \tag{14}$$

Thus $\eta_r = \lambda$, which yields the updating of r_j

$$r_j = \sum_{i=1}^M \sum_{k=1}^K \Omega_{kj} \Theta_{ik} p_i.$$

C. Update the dual form of the LM rate

Similar to the previous work [19], we analyze the first-order condition of the variables $\boldsymbol{\phi}$, $\widetilde{\boldsymbol{\psi}}$, ζ respectively. One could update the dual variables alternatively according to the following equations

$$\phi_i = \frac{p_i}{\sum_{j=1}^N e^{-\zeta D_{ij}} \widetilde{\psi_j} r_j}, \quad \widetilde{\psi}_j = \frac{1}{\sum_{i'=1}^M \phi_{i'} e^{-\zeta D_{i'j}}}.$$

Also, the variable ζ could be updated by finding the root of a monotone function $G(\zeta)$ with Newton's Method, as defined in (8). Note that $G(0) \leq 0$ indicates that the corresponding constraint is satisfied. Thus, we can simply set $\zeta = 0$ without finding its root when $G(0) \leq 0$.

D. AM algorithm with the power constraint

An average power constraint like $\mathbb{E}\left[X^2\right] \leq \Gamma$ may be considered under the AWGN channel. Accordingly, the optimization problem is turned to be (15), instead of (6) in the main text.

$$
\max_{\boldsymbol{p},\boldsymbol{t},\boldsymbol{r}} \max_{\substack{\boldsymbol{\phi}, \widetilde{\boldsymbol{\psi}}, \\ \zeta \geq 0}} -\boldsymbol{p}^T \log \boldsymbol{p} + \boldsymbol{p}^T \log \boldsymbol{J}(\boldsymbol{\phi}, \widetilde{\boldsymbol{\psi}}, \zeta; \boldsymbol{\Omega}, \boldsymbol{r}) + 1 \\
\text{s.t. } \boldsymbol{\Omega} \boldsymbol{1}_N = \boldsymbol{1}_K, \ \|\boldsymbol{r}\|_1 = 1, \ \|\boldsymbol{p}\|_1 = 1, \mathbb{E}_{\boldsymbol{p}}\|\boldsymbol{x}\| \leq \Gamma. \tag{15}
$$

To solve this problem, the alternating ascend technique is similarly adopted. Actually, the difference only appears in the process of updating the input distribution $\boldsymbol{p}$. The main derivation of the AM algorithm with the power constraint is given below.

Consider the optimization problem (15) with respect to $\boldsymbol{p}$ only, and construct its Lagrangian

$$
\mathcal{L}_p(\boldsymbol{p}; \mu, \eta_p) = -\sum_{i=1}^M p_i \log p_i + \sum_{i=1}^M p_i \log J_i(\boldsymbol{\phi}, \widetilde{\boldsymbol{\psi}}, \zeta; \boldsymbol{\Omega}, \boldsymbol{r}) \\
+ 1 - \mu \left(\sum_{i=1}^M p_i \|x_i\|^2 - \Gamma \right) - \eta_p \left(\sum_{i=1}^M p_i - 1 \right), \tag{16}
$$

where $\mu \in \mathbb{R}^+$ and $\eta_p \in \mathbb{R}$ are dual variables.

Taking the partial derivative of the Lagrangian $\mathcal{L}_p(\boldsymbol{p}; \mu, \eta_p)$ with respect to $\boldsymbol{p}$, we obtain the optimal solution

$$\widetilde{p}_i = J_i e^{-\lambda \|x_i\|^2} e^{-1-\eta_p}.$$

Then the dual form of (15) with respect to $\boldsymbol{p}$ is constructed as

$$\max_{\lambda \geq 0,\ \eta_p} -e^{-1-\eta_p} \left(\sum_{i=1}^{M} J_i e^{-\lambda \|x_i\|^2} \right) - 1 - \lambda\Gamma - \eta_p. \quad (17)$$

By analyzing the dual optimization problem (17), we can update $\boldsymbol{p}$ according to

$$p_i = \frac{J_i e^{-\mu\|x_i\|^2}}{\sum_{i'=1}^{M} J_{i'} e^{-\mu\|x_{i'}\|^2}}, \quad (18)$$

where the dual variable η_p is vanished, and where $\mu \in \mathbb{R}^+$ is updated via finding the root of the following one-dimensional monotonic function

$$F(\mu) \triangleq -\Gamma + \frac{\sum_{i=1}^{M} \|x_i\|^2 J_i e^{-\mu\|x_i\|^2}}{\sum_{i=1}^{M} J_i e^{-\mu\|x_i\|^2}}. \quad (19)$$

When $F(0) \leq 0$, one could simply set $\mu = 0$ without finding its root, since the corresponding constraint has already been satisfied.

The AM algorithm with power constraint is summarized in the pseudo-code in Algorithm 2.

Algorithm 2 AM Algorithm with power constraint

Input: Decoding metric D_{ij}, Conditional probability distribution Θ_{ik}; Power constraint coefficient Γ; The Lagrange multiplier λ; Iteration number max_iter.

1: **Initialize:** $\phi^{(0)} = \mathbf{1}_M$, $\widetilde{\psi}^{(0)} = \mathbf{1}_N$, $\zeta = 1$, $\mu = 1$; **Randomly Initialize:** $\Omega^{(0)}$, $r^{(0)}$.
2: **for** $l = 1 : max\_iter$ **do**
3: Solve $F(\mu) = 0$ for $\mu \in \mathbb{R}^+$ with Newton's method
4: Update $p_i^{(l)} \leftarrow J_i e^{-\mu\|x_i\|^2} / \left(\sum_{i'=1}^{M} J_{i'} e^{-\mu\|x_{i'}\|^2} \right)$
5: Update $\Omega_{kj}^{(l)} \leftarrow \Omega_{kj}^* / \left(\sum_{j'=1}^{N} \Omega_{kj'}^* \right)$
6: Update $r_j^{(l)} \leftarrow \sum_{i=1}^{M} \sum_{k=1}^{K} \Omega_{kj}^{(l)} \Theta_{ik} p_i^{(l)}$
7: Update $\phi_i^{(l)} \leftarrow p_i^{(l)} / \left(\sum_{j=1}^{N} e^{-\zeta D_{ij}} \widetilde{\psi}_j^{(l-1)} r_j^{(l)} \right)$
8: Update $\widetilde{\psi}_j^{(l)} \leftarrow 1 / \left(\sum_{i=1}^{M} \phi_i^{(l)} e^{-\zeta D_{ij}} \right)$
9: Solve $G(\zeta) = 0$ for $\zeta \in \mathbb{R}^+$ with Newton's method
10: **return** C_d

In the numerical simulations of Sec. IV-B, a power constraint $\mathbb{E}\left[X^2\right] \leq \Gamma$ is taken into consideration. And we set $\Gamma = 1$ according to the predetermined coordinate positions of the collection points. Particularly, in the QPSK modulation scheme, each collection point shares the same distance $\|\boldsymbol{x}\| = 1$ according to the initialization. Thus, the power constraint here is equivalent to $\sum_{i=1}^{N} p_i \leq 1$, which always holds by the property of the probability distribution.