A Multiple Description CEO Problem with Log-Loss Distortion

Georg Pichler*, Pablo Piantanida†, and Gerald Matz*

*Institute of Telecommunications, Technische Universität Wien, Vienna, Austria
†Laboratoire des Signaux et Systèmes (L2S), CentraleSupélec-CNRS-Université Paris-Sud, Gif-sur-Yvette, France
Email: {georg.pichler,gerald.matz}@nt.tuwien.ac.at; pablo.piantanida@centralesupelec.fr

***Abstract*—This paper investigates the Multiple Description (MD) Chief Executive Officer (CEO) problem under logarithmic-loss distortion. The setup extends previous work of Courtade and Weissman (2014) by requiring the CEO to obtain a useful reconstruction also from a reduced set of descriptions. A single-letter characterization of the achievable region is derived under a suitable conditional independence assumption. Surprisingly, the resulting rate requirement is in general less than that required to ensure successful typicality decoding of the corresponding description.**

I. Introduction

The *Chief Executive Officer* (CEO) problem [1] under logarithmic-loss distortion was introduced and solved by Courtade and Weissman in [2]. We propose and study a *multiple description* (MD) generalization of this CEO problem that is motivated by the following abstract setup. A CEO wants to learn a random vector Y^n using information provided by J agents. The agents do not cooperate. Each agent observes a different random vector X_j^n ($j \in \{1, 2, \dots, J\}$), which is a noisy version of the relevant vector Y^n, and communicates a description of X_j^n to the CEO at a rate less than R_j. The CEO uses these descriptions to determine a soft estimate of Y^n (i.e., a probability distribution). The quality of this estimate is measured by logarithmic-loss distortion. The information-theoretic problem consists of establishing the trade-off between the set of feasible rates and the achievable average distortion.

The CEO's soft estimate of Y^n relies on the full set of descriptions obtained from all agents. If an agent fails to deliver a description, the CEO's estimate may be severely degraded. A similar situation occurs when sensor nodes fail in a sensor network, which can be characterized by the same model (here, the CEO and the agents correspond to a fusion center and the sensor nodes, respectively). We thus propose an extension of the CEO problem that takes into account the possibility that one or more descriptions are not delivered by using MD coding [3] (see [4] for applications). The general MD-CEO setup with two agents was previously investigated in [5]. We will restrict our attention to the log-loss distortion measure, but allow for an arbitrary number of agents, however assuming that $\mathsf{X}_1, \mathsf{X}_2, \dots \mathsf{X}_J$ are independent given Y. Assuming this independence, a single-letter characterization of the achievable region of the CEO problem with log-loss distortion was obtained in [2] using submodularity theory [6]. This seminal result was extended in [7], [8] to address the problem of distributed biclustering of multiple memoryless sources, where redundancy among the descriptions should be preserved. In the present work, we extend the original setup by the requirement that the CEO needs to be able to satisfy distortion constraints given only the message from one agent. Surprisingly, this non-straightforward extension to an MD problem permits a single-letter characterization of the set of all feasible rates and distortions, which is our main result (Theorem 1). In order to highlight some interesting features of the resulting achievable rate-distortion region, we state it separately for the case of $J = 2$ agents (Corollary 1).

This paper is organized as follows. Section II introduces the problem formally and states the main result in Theorem 1. In Section III, the matching inner and outer bounds are presented and proven. Finally, Section IV presents the proof of Theorem 1; auxiliary results are relegated to the Appendix.

Notation

We denote random quantities by uppercase sans-serif letters and their realizations by lowercase letters. Vectors are indicated by bold-face type and sets by calligraphic type. We use $\mathbb{1}_{\mathcal{A}}$ for the indicator function of a set (or event) $\mathcal{A}$. We use $\mathbb{E}[\mathsf{X}]$ and $\mathrm{P}\{\mathcal{A}\}$ for the expectation of the random variable X and the probability of an event $\mathcal{A}$, respectively. Subscripts indicate parts of vectors, e.g., $\boldsymbol{x}_{\mathcal{A}} := (\boldsymbol{x}_i)_{i \in \mathcal{A}}$. We further use the common notation $\boldsymbol{x}_i^j := \boldsymbol{x}_{\{i,\dots,j\}}$, $\boldsymbol{x}^j := \boldsymbol{x}_1^j$. If a vector is already carrying a subscript, it will be separated by a comma, e.g., $\boldsymbol{x}_{3,1}^5 = (\boldsymbol{x}_3)_1^5 = (\boldsymbol{x}_3)^5$. The ith unit vector is denoted by $\boldsymbol{e}_i$ and $\mathbf{0}$ is the all-zeroes vector. Random variables are assumed to be supported on finite sets. We use the same letter for the random variable and for its support set, e.g., Y takes values in $\mathcal{Y}$ and X_3 takes values in $\mathcal{X}_3$. $\mathcal{X}^n$ denotes the nth Cartesian power of a set $\mathcal{X}$. Random vectors comprise n i.i.d. copies of the corresponding random variables, e.g., $(\mathbf{X}, \mathbf{Y}) = (\mathsf{X}, \mathsf{Y})^n$. We use the notation from [9, Chapter 2] for information-theoretic quantities. The notation $\mathsf{X} \multimap \mathsf{Y} \multimap \mathsf{Z}$ indicates that X, Y, and Z form a Markov chain in that order. For a random variable X, we denote the set of (strongly) δ-typical n-sequences [10, Section 2.4] by $\mathcal{T}^n_{[\mathsf{X}]\delta}$ and similarly for conditionally typical sequences. For convenience, we introduce the shorthand notation $[l : k] := \{l, l+1, \dots, k-1, k\}$, $\mathcal{J} := [J]$, $\mathcal{K} := [K]$, and define $K := J + 1$. For a total order $\sqsubset$ of a set $\mathcal{E}$ and $e \in \mathcal{E}$ we will use the notation $\sqsupset e := \{e' \in \mathcal{E} : e' \sqsupset e\}$ and accordingly for $\sqsupseteq$, $\sqsubset$ and $\sqsubseteq$. E.g., given the total order of $\{1, 2, 3\}$ with $3 \sqsubset 1 \sqsubset 2$, we have $\sqsupset 3 = \{1, 2\}$, $\sqsupset 1 = \{2\}$ and $\sqsupset 2 = \varnothing$. We will use $\mathrm{conv}(\mathcal{A})$ and $\overline{\mathcal{A}}$ to denote the convex hull and the topological closure of the set $\mathcal{A}$, respectively.

II. Problem Definition and Main Results

A. Problem Definition

Let $(\mathsf{X}_{\mathcal{J}}, \mathsf{Y})$ be $K = J + 1$ random variables such that $\mathsf{X}_j \multimap \mathsf{Y} \multimap \mathsf{X}_{\mathcal{J} \setminus j}$ form a Markov chain for every $j \in \mathcal{J}$. An

Funding by WWTF Grant ICT12-54 (TINCOIN).

$(n, R_{\mathcal{J}})$ code $f_{\mathcal{J}}$ consists of J functions $f_j : \mathcal{X}_j^n \to \mathcal{M}_j$ with $n^{-1} \log|\mathcal{M}_j| \le R_j$.

Definition 1 (Achievability). *A tuple $(\nu_{\mathcal{K}}, R_{\mathcal{J}})$ is achievable if there exists an $(n, R_{\mathcal{J}})$ code $f_{\mathcal{J}}$ for some $n \in \mathbb{N}$ such that*

$$\nu_K \le \frac{1}{n} \mathrm{I}\big(f_{\mathcal{J}}(\mathbf{X}_{\mathcal{J}}); \mathbf{Y}\big), \tag{1}$$

$$\nu_j \le \frac{1}{n} \mathrm{I}\big(f_j(\mathbf{X}_j); \mathbf{Y}\big) \text{ for all } j \in \mathcal{J}. \tag{2}$$

The set of achievable tuples is denoted by $\mathcal{R}$.

Remark 1. Using a standard time-sharing argument, it can be shown that $\overline{\mathcal{R}}$ is a convex set.

We can define an equivalent region based on logarithmic-loss distortion, defined as

$$\mathrm{d}_{\mathrm{LL}}(\mathrm{p}, z) := -\log \mathrm{p}(z),$$

where $\mathrm{p} \in \mathcal{P}(\mathcal{Z})$ is a probability distribution on some finite set $\mathcal{Z}$ and $z \in \mathcal{Z}$.

Definition 2 (Achievability based on log-loss distortion). *A tuple $(\nu_{\mathcal{K}}, R_{\mathcal{J}})$ is log-loss achievable if for some $n \in \mathbb{N}$ there exists an $(n, R_{\mathcal{J}})$-code $f_{\mathcal{J}}$ and decoding functions $g_j : \mathcal{M}_j \to \mathcal{P}(\mathcal{Y}^n)$, $j \in \mathcal{K}$, ($\mathcal{M}_K := \mathcal{M}_{\mathcal{J}}$) such that*

$$\mathbb{E}\Big[\mathrm{d}_{\mathrm{LL}}\big(g_j \circ f_j(\mathbf{X}_j), \mathbf{Y}\big)\Big] \le \nu_j \text{ for all } j \in \mathcal{J},$$

$$\mathbb{E}\Big[\mathrm{d}_{\mathrm{LL}}\big(g_K \circ f_{\mathcal{J}}(\mathbf{X}_{\mathcal{J}}), \mathbf{Y}\big)\Big] \le \nu_K.$$

The set of all log-loss achievable tuples is denoted by $\mathcal{R}_{\mathrm{LL}}$.

The following proposition follows from [2, Lemma 1], taking into account that equality can be achieved when using $g_j(m_j) = \mathrm{P}\{\mathbf{Y} = \cdot \,|f_j(\mathbf{X}_j) = m_j\}$ and $g_K(m_{\mathcal{J}}) = \mathrm{P}\{\mathbf{Y} = \cdot \,|f_{\mathcal{J}}(\mathbf{X}_{\mathcal{J}}) = m_{\mathcal{J}}\}$.

Proposition 1 (Equivalence with log-loss achievability). *$(\nu_{\mathcal{K}}, R_{\mathcal{J}}) \in \mathcal{R}_{\mathrm{LL}}$ if and only if $(\nu'_{\mathcal{K}}, R_{\mathcal{J}}) \in \mathcal{R}$, where $\nu'_j := \mathrm{H}(\mathsf{Y}) - \nu_j$ for $j \in \mathcal{K}$.*

B. Main Results

We will need the following set of random variables:

$$\mathcal{P}_* := \{\mathsf{U}_{\mathcal{J}}, \mathsf{Q} : \mathsf{Q} \text{ independent of } (\mathsf{X}_{\mathcal{J}}, \mathsf{Y}),$$
$$\mathsf{U}_j \multimap (\mathsf{X}_j, \mathsf{Q}) \multimap (\mathsf{X}_{\mathcal{J}\setminus j}, \mathsf{Y}, \mathsf{U}_{\mathcal{J}\setminus j}) \text{ for all } j \in \mathcal{J}\}.$$

Definition 3 (Rate-distortion region). *For a total order $\sqsubset$ of $\mathcal{J}$ and a set $\mathcal{I} \subseteq \mathcal{J}$, let the region $\mathcal{R}_*^{(\sqsubset,\mathcal{I})}$ be the set of tuples $(\nu_{\mathcal{K}}, R_{\mathcal{J}})$ such that there exist random variables $(\mathsf{U}_{\mathcal{J}}, \varnothing) \in \mathcal{P}_*$ with*

$$R_j \ge \mathrm{I}(\mathsf{U}_j; \mathsf{X}_j | \mathsf{U}_{\sqsupset j}), \qquad j \in \mathcal{J}$$
$$R_j \ge \mathrm{I}(\mathsf{U}_j; \mathsf{X}_j), \qquad j \in \mathcal{I}$$
$$\nu_j \le \mathrm{I}(\mathsf{U}_j; \mathsf{Y} | \mathsf{U}_{\sqsupset j}), \qquad j \notin \mathcal{I} \tag{3}$$
$$\nu_j \le \mathrm{I}(\mathsf{U}_j; \mathsf{Y}), \qquad j \in \mathcal{I} \tag{4}$$
$$\nu_K \le \mathrm{I}(\mathsf{U}_{\mathcal{J}}; \mathsf{Y}). \tag{5}$$

Remark 2. The purpose of the total order $\sqsubset$ is to determine the order of the messages for successive decoding. Equivalently, Definition 3 could be rephrased in terms of a permutation of $\mathcal{J}$ in place of the total order $\sqsubset$.

Now we are able to state our main result, the proof of which is deferred to Section IV.

Theorem 1. *We have $\overline{\mathcal{R}} = \overline{\mathrm{conv}\Big(\bigcup_{\sqsubset,\mathcal{I}} \mathcal{R}_*^{(\sqsubset,\mathcal{I})}\Big)}$, where the union is over all total orders $\sqsubset$ of $\mathcal{J}$ and all sets $\mathcal{I} \subseteq \mathcal{J}$.*

Remark 3. Using standard methods such as the convex cover method [10, Appendix C], one can obtain cardinality bounds for the random variables $\mathsf{U}_{\mathcal{J}}$. This implies that $\overline{\mathcal{R}}$ is computable and can be used to show (using a compactness argument) that $\mathrm{conv}\Big(\bigcup_{\sqsubset,\mathcal{I}} \mathcal{R}_*^{(\sqsubset,\mathcal{I})}\Big)$ is already topologically closed. The reasoning is the same as in [11, Appendix F].

Specializing Theorem 1 to the case $J = 2$ yields the following corollary.

Corollary 1 (Rate region for $J = 2$). *We have $\overline{\mathcal{R}} = \overline{\mathrm{conv}\Big(\bigcup_{i\in\{1,2,3\}} \mathcal{R}_*^{(i)}\Big)}$ for $J = 2$, where $(\nu_{\mathcal{K}}, R_{\mathcal{J}}) \in \mathcal{R}_*^{(i)}$ if and only if, for some $(\mathsf{U}_{\mathcal{J}}, \varnothing) \in \mathcal{P}_*$, the following inequalities are satisfied:*

$\mathcal{R}_*^{(1)}$:

$$R_1 \ge \mathrm{I}(\mathsf{U}_1; \mathsf{X}_1)$$
$$R_2 \ge \mathrm{I}(\mathsf{U}_2; \mathsf{X}_2 | \mathsf{U}_1)$$
$$\nu_1 \le \mathrm{I}(\mathsf{U}_1; \mathsf{Y})$$
$$\nu_2 \le \mathrm{I}(\mathsf{U}_2; \mathsf{Y} | \mathsf{U}_1)$$
$$\nu_3 \le \mathrm{I}(\mathsf{U}_1 \mathsf{U}_2; \mathsf{Y})$$

$\mathcal{R}_*^{(2)}$:

$$R_1 \ge \mathrm{I}(\mathsf{U}_1; \mathsf{X}_1 | \mathsf{U}_2)$$
$$R_2 \ge \mathrm{I}(\mathsf{U}_2; \mathsf{X}_2)$$
$$\nu_1 \le \mathrm{I}(\mathsf{U}_1; \mathsf{Y} | \mathsf{U}_2)$$
$$\nu_2 \le \mathrm{I}(\mathsf{U}_2; \mathsf{Y})$$
$$\nu_3 \le \mathrm{I}(\mathsf{U}_1 \mathsf{U}_2; \mathsf{Y})$$

$\mathcal{R}_*^{(3)}$:

$$R_1 \ge \mathrm{I}(\mathsf{U}_1; \mathsf{X}_1)$$
$$R_2 \ge \mathrm{I}(\mathsf{U}_2; \mathsf{X}_2)$$
$$\nu_1 \le \mathrm{I}(\mathsf{U}_1; \mathsf{Y})$$
$$\nu_2 \le \mathrm{I}(\mathsf{U}_2; \mathsf{Y})$$
$$\nu_3 \le \mathrm{I}(\mathsf{U}_1 \mathsf{U}_2; \mathsf{Y}).$$

Proof. Assuming $1 \sqsubset 2$, we obtain $\mathcal{R}_*^{(\sqsubset,\mathcal{I})} = \mathcal{R}_*^{(2)}$ if $1 \notin \mathcal{I}$ and otherwise $\mathcal{R}_*^{(\sqsubset,\mathcal{I})} = \mathcal{R}_*^{(3)}$. On the other hand, if $2 \sqsubset 1$, we obtain $\mathcal{R}_*^{(\sqsubset,\mathcal{I})} = \mathcal{R}_*^{(1)}$ if $2 \notin \mathcal{I}$ and otherwise also $\mathcal{R}_*^{(\sqsubset,\mathcal{I})} = \mathcal{R}_*^{(3)}$. ∎

Remark 4. Note that the total available rate of encoder 2 is $R_2 = \mathrm{I}(\mathsf{U}_2; \mathsf{X}_2 | \mathsf{U}_1)$ to achieve a point in $\mathcal{R}_*^{(1)}$. Interestingly, this rate is in general less than the rate required to ensure successful typicality decoding of U_2. However, $\nu_2 = \mathrm{I}(\mathsf{U}_2; \mathsf{Y} | \mathsf{U}_1)$ can still be achieved.

Remark 5. On the other hand, fixing the random variables U_1, U_2 in the definition of $\mathcal{R}_*^{(i)}$ shows another interesting feature of this region. The achievable values for ν_1 and ν_2 vary across $i \in \{1, 2, 3\}$ and hence do not only depend on the chosen random variables U_1 and U_2, but also on the specific rates R_1 and R_2.

It is worth mentioning that by setting $\nu_1 = \nu_2 = 0$ the region $\overline{\mathcal{R}}$ reduces to the rate region of the CEO problem with log-loss distortion derived in [2].

Also note that the problem where $\mathsf{X}_1 = \mathsf{X}_2 = \mathsf{Y}$ degenerates. The resulting achievable region $\overline{\mathcal{R}} = \{(\nu_{\mathcal{K}}, R_{\mathcal{J}}) : \nu_1 \le R_1, \nu_2 \le R_2, \nu_3 \le R_1 + R_2, \text{ and } \nu_1, \nu_2, \nu_3 \le \mathrm{H}(\mathsf{Y})\}$ can be obtained from $\mathcal{R}_*^{(3)}$ or from the El Gamal-Cover inner bound [3] using Proposition 1. Thus, the difficulty of the MD-CEO problem results from its distributed nature.

III. MATCHING INNER AND OUTER BOUNDS

A. The Outer Bound

We will prove Theorem 1 by first showing the following outer bound.

	(6)	(7)	(8)	(9)
$\mathcal{A}$	$\varnothing$	$\mathcal{J}$	$\{j_0\}$	$\{j_0\}$
$\mathcal{B}$	$\{j_0\}$	arbitrary	$\{j_0\}$	$\varnothing$
$\nu_{\mathcal{A}}$	0	ν_K	ν_{j_0}	ν_{j_0}

TABLE I: Choices for $\mathcal{A}, \mathcal{B} \subseteq \mathcal{J}$, and $\nu_{\mathcal{A}}$.

Theorem 2 (Outer bound)**.** *If $(\nu_{\mathcal{K}}, R_{\mathcal{J}}) \in \mathcal{R}$ then for any $j_0 \in \mathcal{J}$ and $\mathcal{B} \subseteq \mathcal{J}$,*

$$R_{j_0} \geq 0, \tag{6}$$

$$\sum_{j \in \mathcal{B}} R_j - \nu_K \geq \mathrm{I}(\mathsf{X}_{\mathcal{B}}; \mathsf{U}_{\mathcal{B}} | \mathsf{YQ}) - \mathrm{I}\big(\mathsf{Y}; \mathsf{U}_{\mathcal{J} \setminus \mathcal{B}} \big| \mathsf{Q}\big), \tag{7}$$

$$R_{j_0} - \nu_{j_0} \geq \mathrm{I}(\mathsf{X}_{j_0}; \mathsf{U}_{j_0} | \mathsf{YQ}), \tag{8}$$

$$\nu_{j_0} \leq \mathrm{I}(\mathsf{Y}; \mathsf{U}_{j_0} | \mathsf{Q}), \tag{9}$$

for some random variables $(\mathsf{U}_{\mathcal{J}}, \mathsf{Q}) \in \mathcal{P}_*$.

Proof. Let $\mathsf{W}_j := f_j(\mathbf{X}_j)$ where $f_{\mathcal{J}}$ is an $(n, R_{\mathcal{J}})$ code achieving $(\nu_{\mathcal{K}}, R_{\mathcal{J}}) \in \mathcal{R}$. To show (6)–(9), we choose sets $\mathcal{A}, \mathcal{B} \subseteq \mathcal{J}$ and the value $\nu_{\mathcal{A}}$ according to Table I. With $\mathsf{U}_{j,i} := (\mathsf{W}_j, \mathbf{X}_j^{i-1})$ and $\mathsf{Q}_i := (\mathbf{Y}^{i-1}, \mathbf{Y}_{i+1}^n)$, we obtain

$$\begin{aligned}
n \sum_{j \in \mathcal{B}} R_j &\geq \mathrm{H}(\mathsf{W}_{\mathcal{B}}) \\
&= \mathrm{I}(\mathsf{W}_{\mathcal{B}}; \mathbf{X}_{\mathcal{B}}) \\
&= \mathrm{I}(\mathsf{W}_{\mathcal{B}}; \mathbf{X}_{\mathcal{B}} \mathbf{Y}) \\
&= \mathrm{I}(\mathsf{W}_{\mathcal{B}}; \mathbf{Y}) + \mathrm{I}(\mathsf{W}_{\mathcal{B}}; \mathbf{X}_{\mathcal{B}} | \mathbf{Y}) \\
&= \mathrm{I}(\mathsf{W}_{\mathcal{A}} \mathsf{W}_{\mathcal{B}}; \mathbf{Y}) - \mathrm{I}\big(\mathsf{W}_{\mathcal{A} \setminus \mathcal{B}}; \mathbf{Y} \big| \mathsf{W}_{\mathcal{B}}\big) + \mathrm{I}(\mathsf{W}_{\mathcal{B}}; \mathbf{X}_{\mathcal{B}} | \mathbf{Y}) \\
&\geq n\nu_{\mathcal{A}} - \mathrm{I}\big(\mathsf{W}_{\mathcal{A} \setminus \mathcal{B}}; \mathbf{Y} \big| \mathsf{W}_{\mathcal{B}}\big) + \mathrm{I}(\mathsf{W}_{\mathcal{B}}; \mathbf{X}_{\mathcal{B}} | \mathbf{Y}) && (10) \\
&\geq n\nu_{\mathcal{A}} - \mathrm{I}\big(\mathsf{W}_{\mathcal{A} \setminus \mathcal{B}}; \mathbf{Y}\big) + \mathrm{I}(\mathsf{W}_{\mathcal{B}}; \mathbf{X}_{\mathcal{B}} | \mathbf{Y}) && (11) \\
&= \sum_{i=1}^{n} \big[\nu_{\mathcal{A}} - \mathrm{I}\big(\mathsf{W}_{\mathcal{A} \setminus \mathcal{B}}; \mathbf{Y}_i \big| \mathbf{Y}^{i-1}\big) + \mathrm{I}\big(\mathsf{W}_{\mathcal{B}}; \mathbf{X}_{\mathcal{B},i} \big| \mathbf{Y} \mathbf{X}_{\mathcal{B}}^{i-1}\big)\big] \\
&\geq \sum_{i=1}^{n} \big[\nu_{\mathcal{A}} - \mathrm{I}\big(\mathsf{U}_{\mathcal{A} \setminus \mathcal{B},i}; \mathbf{Y}_i \big| \mathsf{Q}_i\big) + \mathrm{I}\big(\mathsf{W}_{\mathcal{B}}; \mathbf{X}_{\mathcal{B},i} \big| \mathbf{Y} \mathbf{X}_{\mathcal{B}}^{i-1}\big)\big] && (12) \\
&= \sum_{i=1}^{n} \big[\nu_{\mathcal{A}} - \mathrm{I}\big(\mathsf{U}_{\mathcal{A} \setminus \mathcal{B},i}; \mathbf{Y}_i \big| \mathsf{Q}_i\big) + \mathrm{I}(\mathsf{U}_{\mathcal{B},i}; \mathbf{X}_{\mathcal{B},i} | \mathbf{Y}_i \mathsf{Q}_i)\big],
\end{aligned}$$

where (10) follows from the inequality:

$$\mathrm{I}(\mathsf{W}_{\mathcal{A}} \mathsf{W}_{\mathcal{B}}; \mathbf{Y}) \geq \mathrm{I}(\mathsf{W}_{\mathcal{A}}; \mathbf{Y}) \geq n\nu_{\mathcal{A}}$$

by (1) and (2); (11) is a consequence of the Markov chain $\mathsf{W}_{\mathcal{A} \setminus \mathcal{B}} \multimap \mathbf{Y} \multimap \mathsf{W}_{\mathcal{B}}$ and (12) follows from the assumption that the sources are memoryless and the non-negativity of mutual information. The final result follows by a standard time-sharing argument. Note that the required Markov chain and the independence stated in the theorem are satisfied. ■

B. The Matching Inner Bound

Theorem 3 (Inner bound)**.** *For any total order $\sqsubset$ of $\mathcal{J}$ and any $\mathcal{I} \subseteq \mathcal{J}$, $\mathcal{R}_*^{(\sqsubset, \mathcal{I})} \subseteq \overline{\mathcal{R}}$.*

Proof. Let $(\nu_{\mathcal{K}}, R_{\mathcal{J}}) \in \mathcal{R}_*^{(\sqsubset, \mathcal{I})}$ and choose $(\mathsf{U}_{\mathcal{J}}, \varnothing) \in \mathcal{P}_*$, the total order $\sqsubset$ and the set $\mathcal{I}$ according to Definition 3. Selecting $\varepsilon > 0$ and $n \in \mathbb{N}$, we use the following random encoding procedure for each $j \in \mathcal{J}$.

Form a codebook by choosing codewords $\mathbf{U}_j(m_j)$, $m_j \in \big[\mathrm{e}^{n\widetilde{R}_j}\big]$ independently and uniformly from the set of typical sequences $\mathcal{T}^n_{[\mathsf{U}_j]\delta}$ with $\widetilde{R}_j = \mathrm{I}(\mathsf{U}_j; \mathsf{X}_j) + \varepsilon/2$. Denote the typicality encoder $\widetilde{f}_j$, the encoding $\widetilde{\mathsf{W}}_j := \widetilde{f}_j(\mathbf{X}_j)$ and the decoder $\widetilde{g}_j(m_j) = \mathbf{U}_j(m_j)$. We add an additional layer of (deterministic) binning to obtain the encoding function $f'_j := \beta_j \circ \widetilde{f}_j$, where β_j denotes the mapping into $\mathrm{e}^{nR'_j}$ equal-sized bins with $R'_j = \mathrm{I}(\mathsf{U}_j; \mathsf{X}_j | \mathsf{U}_{\sqsupset j}) + \varepsilon$. Denote the random encoding $\mathsf{W}'_j := f'_j(\mathbf{X}_j)$. The final $(n, R_{\mathcal{J}} + \varepsilon)$ code $f_{\mathcal{J}}$ is given by $f_j := \widetilde{f}_j$ for $j \in \mathcal{I}$ and $f_j := f'_j$ for $j \notin \mathcal{I}$. For $\mathcal{A} = \sqsupset j_0$, where $j_0 \in \mathcal{J}$ is arbitrary, we will also need the decoding function $g'_{\mathcal{A}}$, which operates on the bin indices $\mathsf{W}'_{\mathcal{A}}$ and returns the unique codewords in the bins, such that joint typicality holds, i.e., $g'_{\mathcal{A}}(\mathsf{W}'_{\mathcal{A}}) = \mathbf{U}_{\mathcal{A}}(m_{\mathcal{A}}) \in \mathcal{T}^n_{[\mathsf{U}_{\mathcal{A}}]\delta}$ with $m_{\mathcal{A}} \in \beta_{\mathcal{A}}^{-1}(\mathsf{W}'_{\mathcal{A}})$. If this is not possible, $g'_{\mathcal{A}}$ returns arbitrary vectors.

Let the event $\mathcal{S}'_{\mathcal{A}}$ be the success event that joint typicality $\big(\mathbf{Y}, \mathbf{X}_{\mathcal{A}}, g'_{\mathcal{A}}(\mathsf{W}'_{\mathcal{A}})\big) \in \mathcal{T}^n_{[\mathsf{YX}_{\mathcal{A}}\mathsf{U}_{\mathcal{A}}]\delta}$ holds. Also let $\widetilde{\mathcal{S}}_j$ be the event that typicality $\big(\mathbf{Y}, \mathbf{X}_j, \widetilde{g}_j(\widetilde{\mathsf{W}}_j)\big) \in \mathcal{T}^n_{[\mathsf{YX}_j\mathsf{U}_j]\delta}$ holds. We have for $\mathcal{A}' \subseteq \mathcal{A} = \sqsupset j_0$,

$$\begin{aligned}
\sum_{j \in \mathcal{A}'} \mathrm{I}(\mathsf{U}_j; \mathsf{X}_j | \mathsf{U}_{\sqsupset j}) &\geq \sum_{j \in \mathcal{A}'} \mathrm{I}\big(\mathsf{U}_j; \mathsf{X}_{\mathcal{A}'} \big| \mathsf{U}_{\sqsupset j}, \mathsf{U}_{\mathcal{A} \setminus \mathcal{A}'}\big) \\
&= \mathrm{I}\big(\mathsf{U}_{\mathcal{A}'}; \mathsf{X}_{\mathcal{A}'} \big| \mathsf{U}_{\mathcal{A} \setminus \mathcal{A}'}\big). && (13)
\end{aligned}$$

Considering the choice of $R'_{\mathcal{J}}$, the inequality (13) together with standard random coding arguments [11]–[14] is sufficient to show that for n large enough, $\mathrm{P}\{\mathcal{S}'_{\mathcal{A}}\} \geq 1 - \varepsilon$ and also $\mathrm{P}\{\widetilde{\mathcal{S}}_j\} \geq 1 - \varepsilon$ for each $j \in \mathcal{J}$ by the choice of $\widetilde{R}_j$. For further details the reader is referred to [15].

Let us pick an arbitrary[1] $\varepsilon' > 0$. For any $\mathcal{A} = \sqsupset j_0$, provided that n is large enough and ε small enough, we have that

$$\begin{aligned}
\frac{1}{n}\mathrm{I}(\mathbf{Y}; \mathsf{W}_{\mathcal{A}}) &\geq \frac{1}{n}\mathrm{I}(\mathbf{Y}; \mathsf{W}'_{\mathcal{A}}) && (14) \\
&\geq \frac{1}{n}\mathrm{I}\big(\mathbf{Y}; g_{\mathcal{A}}(\mathsf{W}'_{\mathcal{A}})\big) && (15) \\
&= \mathrm{H}(\mathsf{Y}) - \frac{1}{n}\mathrm{H}\big(\mathbf{Y} \big| g'_{\mathcal{A}}(\mathsf{W}'_{\mathcal{A}})\big) \\
&\geq \mathrm{H}(\mathsf{Y}) - \frac{1}{n}\mathrm{H}\big(\mathbf{Y}, \mathbb{1}_{\mathcal{S}'_{\mathcal{A}}} \big| g'_{\mathcal{A}}(\mathsf{W}'_{\mathcal{A}})\big) \\
&= \mathrm{H}(\mathsf{Y}) - \frac{1}{n}\mathrm{H}\big(\mathbb{1}_{\mathcal{S}'_{\mathcal{A}}} \big| g'_{\mathcal{A}}(\mathsf{W}'_{\mathcal{A}})\big) \\
&\qquad - \frac{1}{n}\mathrm{H}\big(\mathbf{Y} \big| g'_{\mathcal{A}}(\mathsf{W}'_{\mathcal{A}}), \mathbb{1}_{\mathcal{S}'_{\mathcal{A}}}\big) \\
&\geq \mathrm{H}(\mathsf{Y}) - \varepsilon' \\
&\qquad - \frac{1}{n}(1 - \varepsilon)\mathrm{H}\big(\mathbf{Y} \big| g'_{\mathcal{A}}(\mathsf{W}'_{\mathcal{A}}), \mathcal{S}'_{\mathcal{A}}\big) - \varepsilon \mathrm{H}(\mathsf{Y}) \\
&\geq \mathrm{H}(\mathsf{Y}) - \varepsilon' - \frac{1}{n}\mathrm{H}\big(\mathbf{Y} \big| g'_{\mathcal{A}}(\mathsf{W}'_{\mathcal{A}}), \mathcal{S}'_{\mathcal{A}}\big) \\
&\geq \mathrm{H}(\mathsf{Y}) - \varepsilon' - \frac{1}{n}\sum_{\boldsymbol{u}_{\mathcal{A}}} \mathrm{P}\{g'_{\mathcal{A}}(\mathsf{W}'_{\mathcal{A}}) = \boldsymbol{u}_{\mathcal{A}} | \mathcal{S}'_{\mathcal{A}}\} \\
&\qquad \times \log\Big|\mathcal{T}^n_{[\mathsf{Y}|\mathsf{U}_{\mathcal{A}}]\delta}(\boldsymbol{u}_{\mathcal{A}})\Big| && (16) \\
&\geq \mathrm{H}(\mathsf{Y}) - \mathrm{H}(\mathsf{Y}|\mathsf{U}_{\mathcal{A}}) - \varepsilon' = \mathrm{I}(\mathsf{U}_{\mathcal{A}}; \mathsf{Y}) - \varepsilon'. && (17)
\end{aligned}$$

Here, (14) and (15) follow from the data processing inequality [9, Theorem 2.8.1] and we applied the entropy bound [9, Theorem 2.6.4] in (16) and used [10, Section 2.5, Property 3] in (17). In particular,

$$\frac{1}{n}\mathrm{I}(\mathbf{Y}; \mathsf{W}_{\mathcal{J}}) \overset{(17)}{\geq} \mathrm{I}(\mathsf{U}_{\mathcal{J}}; \mathsf{Y}) - \varepsilon' \overset{(5)}{\geq} \nu_K - \varepsilon'.$$

[1] In what follows, we will routinely merge expressions that can be made arbitrarily small (for n large and ε sufficiently small) and bound them by ε'.

Also, for $j_0 \in \mathcal{J}$ and $\mathcal{A} = \sqsupset j_0$,

$$
\begin{aligned}
\frac{1}{n}\mathrm{I}(\mathbf{Y}; \mathsf{W}_{j_0}) &\geq \frac{1}{n}\mathrm{I}\big(\mathbf{Y}; \mathsf{W}'_{j_0}\big) \geq \frac{1}{n}\mathrm{I}\big(\mathbf{Y}; \mathsf{W}'_{j_0}\big|\mathsf{W}'_{\mathcal{A}}\big) \\
&= \frac{1}{n}\mathrm{I}\big(\mathbf{Y}; \mathsf{W}'_{j_0}\mathsf{W}'_{\mathcal{A}}\big) - \frac{1}{n}\mathrm{I}\big(\mathbf{Y}; \mathsf{W}'_{\mathcal{A}}\big) \\
&\overset{(17)}{\geq} \mathrm{I}(\mathsf{U}_{\mathcal{A}}\mathsf{U}_{j_0}; \mathsf{Y}) - \varepsilon' - \frac{1}{n}\mathrm{I}\big(\mathbf{Y}; \mathsf{W}'_{\mathcal{A}}\big) \\
&= \mathrm{I}(\mathsf{U}_{\mathcal{A}}\mathsf{U}_{j_0}; \mathsf{Y}) - \varepsilon' - \frac{1}{n}\mathrm{I}\big(\mathbf{X}_{\mathcal{A}}; \mathsf{W}'_{\mathcal{A}}\big) \\
&\qquad + \frac{1}{n}\mathrm{I}\big(\mathbf{X}_{\mathcal{A}}; \mathsf{W}'_{\mathcal{A}}\big|\mathbf{Y}\big) \qquad (18) \\
&\geq \mathrm{I}(\mathsf{U}_{\mathcal{A}}\mathsf{U}_{j_0}; \mathsf{Y}) - \varepsilon' - \mathrm{I}(\mathsf{X}_{\mathcal{A}}; \mathsf{U}_{\mathcal{A}}) \\
&\qquad + \mathrm{H}(\mathsf{X}_{\mathcal{A}}|\mathsf{Y}) - \frac{1}{n}\mathrm{H}\big(\mathbf{X}_{\mathcal{A}}\big|\mathsf{W}'_{\mathcal{A}}, \mathbf{Y}\big) \qquad (19) \\
&\geq \mathrm{I}(\mathsf{U}_{\mathcal{A}}\mathsf{U}_{j_0}; \mathsf{Y}) - \varepsilon' - \mathrm{I}(\mathsf{X}_{\mathcal{A}}; \mathsf{U}_{\mathcal{A}}) \\
&\qquad + \mathrm{H}(\mathsf{X}_{\mathcal{A}}|\mathsf{Y}) - \mathrm{H}(\mathsf{X}_{\mathcal{A}}|\mathsf{U}_{\mathcal{A}}, \mathsf{Y}) \qquad (20) \\
&= \mathrm{I}(\mathsf{U}_{j_0}; \mathsf{Y}|\mathsf{U}_{\mathcal{A}}) - \varepsilon' \overset{(3)}{\geq} \nu_{j_0} - \varepsilon'.
\end{aligned}
$$

Here, (18) follows from the Markov chain $\mathsf{W}'_{\mathcal{A}} \multimap\!- \mathbf{X}_{\mathcal{A}} \multimap\!- \mathbf{Y}$. We need to justify (19) and (20). In (19), we used that for ε small enough and n large enough,

$$
\begin{aligned}
\frac{1}{n}\mathrm{I}(\mathbf{X}_{\mathcal{A}}; \mathsf{W}'_{\mathcal{A}}) &= \frac{1}{n}\mathrm{H}(\mathsf{W}'_{\mathcal{A}}) \leq \frac{1}{n}\sum_{j \in \mathcal{A}} \mathrm{H}\big(\mathsf{W}'_j\big) \qquad (21) \\
&\leq \sum_{j \in \mathcal{A}} \Big(\mathrm{I}(\mathsf{U}_j; \mathsf{X}_j|\mathsf{U}_{\sqsupset j}) + \varepsilon\Big) \leq \mathrm{I}(\mathsf{U}_{\mathcal{A}}; \mathsf{X}_{\mathcal{A}}) + \varepsilon', \qquad (22)
\end{aligned}
$$

where the inequality in (21) follows from the chain rule for entropy [9, Theorem 2.2.1] and the data processing inequality and (22) follows from the entropy bound [9, Theorem 2.6.4] and the definition of R'_j. The inequality (20) follows similar to (17) as for n large enough and ε small enough,

$$
\begin{aligned}
\frac{1}{n}\mathrm{H}(\mathbf{X}_{\mathcal{A}}|\mathsf{W}'_{\mathcal{A}}, \mathbf{Y}) &\leq \frac{1}{n}\mathrm{H}(\mathbf{X}_{\mathcal{A}}|g'_{\mathcal{A}}(\mathsf{W}'_{\mathcal{A}}), \mathbf{Y}) \\
&\leq \frac{1}{n}\mathrm{H}\big(\mathbf{X}_{\mathcal{A}}, \mathbb{1}_{\mathcal{S}'_{\mathcal{A}}}\big|g_{\mathcal{A}}(\mathsf{W}'_{\mathcal{A}}), \mathbf{Y}\big) \\
&\leq \varepsilon' + \frac{1}{n}\mathrm{H}\big(\mathbf{X}_{\mathcal{A}}\big|g'_{\mathcal{A}}(\mathsf{W}'_{\mathcal{A}}), \mathbf{Y}, \mathcal{S}'_{\mathcal{A}}\big) \\
&\leq \varepsilon' + \frac{1}{n}\sum_{\boldsymbol{u}_{\mathcal{A}}, \boldsymbol{y}} \mathrm{P}\big\{g'_{\mathcal{A}}(\mathsf{W}'_{\mathcal{A}}) = \boldsymbol{u}_{\mathcal{A}}, \mathbf{Y} = \boldsymbol{y}\big|\mathcal{S}'_{\mathcal{A}}\big\} \\
&\qquad \times \log\Big|\mathcal{T}^n_{[\mathsf{X}_{\mathcal{A}}|\mathsf{U}_{\mathcal{A}}, \mathsf{Y}]\delta}(\boldsymbol{u}_{\mathcal{A}}, \boldsymbol{y})\Big| \\
&\leq \varepsilon' + \mathrm{H}(\mathsf{X}_{\mathcal{A}}|\mathsf{U}_{\mathcal{A}}, \mathsf{Y}).
\end{aligned}
$$

For $j_0 \in \mathcal{I}$, we have similar to (17) that

$$
\begin{aligned}
\frac{1}{n}\mathrm{I}(\mathbf{Y}; \mathsf{W}_{j_0}) &= \frac{1}{n}\mathrm{I}\big(\mathbf{Y}; \widetilde{\mathsf{W}}_{j_0}\big) \geq \frac{1}{n}\mathrm{I}\big(\mathbf{Y}; \widetilde{g}_{j_0}(\widetilde{\mathsf{W}}_{j_0})\big) \\
&\geq \mathrm{H}(\mathsf{Y}) - \frac{1}{n}\mathrm{H}\big(\mathbf{Y}, \mathbb{1}_{\widetilde{\mathcal{S}}_{j_0}}\big|\widetilde{g}_{j_0}(\widetilde{\mathsf{W}}_{j_0})\big) \\
&\geq \mathrm{H}(\mathsf{Y}) - \frac{1}{n}\mathrm{H}\big(\mathbb{1}_{\widetilde{\mathcal{S}}_{j_0}}\big) - \frac{1}{n}\mathrm{H}\big(\mathbf{Y}\big|\widetilde{g}_{j_0}(\widetilde{\mathsf{W}}_{j_0}), \mathbb{1}_{\widetilde{\mathcal{S}}_{j_0}}\big) \\
&\geq \mathrm{H}(\mathsf{Y}) - \varepsilon' - \frac{1}{n}\mathrm{H}\big(\mathbf{Y}\big|\widetilde{g}_{j_0}(\widetilde{\mathsf{W}}_{j_0}), \widetilde{\mathcal{S}}_{j_0}\big) \\
&\geq \mathrm{H}(\mathsf{Y}) - \varepsilon' \\
&\quad - \frac{1}{n}\sum_{\boldsymbol{u}_{j_0}} \mathrm{P}\Big\{\widetilde{g}_{\mathcal{A}}(\widetilde{\mathsf{W}}_{j_0}) = \boldsymbol{u}_{j_0}\Big|\widetilde{\mathcal{S}}_{j_0}\Big\} \log\Big|\mathcal{T}^n_{[\mathsf{Y}|\mathsf{U}_{j_0}]\delta}(\boldsymbol{u}_{j_0})\Big| \\
&\geq \mathrm{H}(\mathsf{Y}) - \varepsilon' - \mathrm{H}(\mathsf{Y}|\mathsf{U}_{j_0}) \\
&= \mathrm{I}(\mathsf{U}_{j_0}; \mathsf{Y}) - \varepsilon' \overset{(4)}{\geq} \nu_{j_0} - \varepsilon'.
\end{aligned}
$$

■

IV. Proof of Theorem 1

In this proof we will make use of the rather technical Lemma 1, which is proved in Appendix A.

Assume $(\nu_{\mathcal{K}}, R_{\mathcal{J}}) \in \mathcal{R}$. We can then find $(\mathsf{U}_{\mathcal{J}}, \mathsf{Q}) \in \mathcal{P}_*$ such that (6)–(9) hold. We define $\widetilde{\nu}_{\mathcal{K}} := -\nu_{\mathcal{K}}$ to simplify notation. It is straightforward to check that the inequalities (6)–(9) define a sequence of closed convex polyhedra $\mathcal{S}^{(j)}$ in the variables $(R_{\mathcal{J}}, \widetilde{\nu}_K)$ that satisfy assumptions 1 and 2 of Lemma 1. $\mathcal{S}^{(0)}$ is defined by (6) and (7) alone. And for $j \in \mathcal{J}$, the polyhedron $\mathcal{S}^{(j)}$ is given in the $K + j$ variables $(R_{\mathcal{J}}, \widetilde{\nu}_K, \widetilde{\nu}_{[j]})$ by adding constraints (8) and (9) for each $j_0 \in [j]$. The set $\mathcal{S}^{(0)}$ is a supermodular polyhedron [6, Section 2.3] on $(\mathcal{K}, 2^{\mathcal{K}})$ with rank function

$$
f(\mathcal{A}) = \begin{cases} 0, & K \notin \mathcal{A}, \\ \mathrm{I}\big(\mathsf{X}_{\mathcal{A}\setminus K}; \mathsf{U}_{\mathcal{A}\setminus K}\big|\mathsf{Y}\mathsf{Q}\big) - \mathrm{I}\big(\mathsf{Y}; \mathsf{U}_{\mathcal{J}\setminus\mathcal{A}}\big|\mathsf{Q}\big), & K \in \mathcal{A}, \end{cases}
$$

where supermodularity follows via standard information-theoretic arguments. By the extreme point theorem [6, Theorem 3.22], every extreme point of $\mathcal{S}^{(0)}$ is associated with a total order $\sqsubset$ of $\mathcal{K}$. Such an extreme point is given by

$$
\begin{aligned}
R_j^{(\sqsubset)} &= 0 \text{ for } j \sqsubset K, \\
R_j^{(\sqsubset)} &= \mathrm{I}(\mathsf{U}_j; \mathsf{X}_j|\mathsf{U}_{\sqsupset j}\mathsf{Q}) \text{ for } j \sqsupset K, \\
\nu_K^{(\sqsubset)} &= \mathrm{I}(\mathsf{Y}; \mathsf{U}_{\sqsupset K}|\mathsf{Q}) - \mathrm{I}(\mathsf{Y}; \mathsf{U}_{\sqsubset K}|\mathsf{Y}\mathsf{Q}).
\end{aligned}
$$

Assumption 3 of Lemma 1 can now be verified by

$$
R_j^{(\sqsubset)} \leq \mathrm{I}(\mathsf{X}_j; \mathsf{U}_j|\mathsf{Y}\mathsf{Q}) + \mathrm{I}(\mathsf{Y}; \mathsf{U}_j|\mathsf{Q}) = \mathrm{I}(\mathsf{X}_j; \mathsf{U}_j|\mathsf{Q}).
$$

By applying Lemma 1 we find that every extreme point of $\mathcal{S}^{(J)}$ is given by a subset $\mathcal{I} \subseteq \mathcal{J}$ and a total order $\sqsubset$ of $\mathcal{K}$ as

$$
\begin{aligned}
R_j^{(\sqsubset,\mathcal{I})} &= \mathrm{I}(\mathsf{X}_j; \mathsf{U}_j|\mathsf{Q}), && j \in \mathcal{I} && (23) \\
R_j^{(\sqsubset,\mathcal{I})} &= 0, && j \notin \mathcal{I} \wedge j \sqsubset K && (24) \\
R_j^{(\sqsubset,\mathcal{I})} &= \mathrm{I}(\mathsf{U}_j; \mathsf{X}_j|\mathsf{U}_{\sqsupset j}\mathsf{Q}), && j \notin \mathcal{I} \wedge j \sqsupset K && (25) \\
\nu_K^{(\sqsubset,\mathcal{I})} &= \mathrm{I}(\mathsf{Y}; \mathsf{U}_{\sqsupset K}|\mathsf{Q}) - \mathrm{I}(\mathsf{Y}; \mathsf{U}_{\sqsubset K}|\mathsf{Y}\mathsf{Q}) && && (26) \\
\nu_j^{(\sqsubset,\mathcal{I})} &= \mathrm{I}(\mathsf{U}_j; \mathsf{Y}|\mathsf{Q}), && j \in \mathcal{I} && (27) \\
\nu_j^{(\sqsubset,\mathcal{I})} &= -\mathrm{I}(\mathsf{U}_j; \mathsf{X}_j|\mathsf{Y}\mathsf{Q}), && j \notin \mathcal{I} \wedge j \sqsubset K && (28) \\
\nu_j^{(\sqsubset,\mathcal{I})} &= \mathrm{I}(\mathsf{U}_j; \mathsf{Y}|\mathsf{U}_{\sqsupset j}\mathsf{Q}), && j \notin \mathcal{I} \wedge j \sqsupset K. && (29)
\end{aligned}
$$

For each $q \in \mathcal{Q}$ with $\mathrm{P}\{\mathsf{Q} = q\} > 0$ let the point $(\nu_{\mathcal{K}}^{(\sqsubset,\mathcal{I},q)}, R_{\mathcal{J}}^{(\sqsubset,\mathcal{I},q)})$ be given by (23)–(29), but with $\mathsf{Q} = q$. By substituting $\mathsf{U}_j \to \varnothing$ if $j \notin \mathcal{I} \wedge j \sqsubset K$, we see that $(\nu_{\mathcal{K}}^{(\sqsubset,\mathcal{I},q)}, R_{\mathcal{J}}^{(\sqsubset,\mathcal{I},q)}) \in \mathcal{R}_*^{(\sqsubset,\mathcal{I})}$ and consequently $(\nu_{\mathcal{K}}^{(\sqsubset,\mathcal{I})}, R_{\mathcal{J}}^{(\sqsubset,\mathcal{I})}) \in \mathrm{conv}\Big(\mathcal{R}_*^{(\sqsubset,\mathcal{I})}\Big)$. Defining the orthant $\mathcal{O}' := \{(\nu_{\mathcal{K}}, R_{\mathcal{J}}) : \nu_{\mathcal{K}} \leq \mathbf{0}, R_{\mathcal{J}} \geq \mathbf{0}\}$, this implies $(\nu_{\mathcal{K}}, R_{\mathcal{J}}) \in \mathrm{conv}\Big(\bigcup_{\sqsubset,\mathcal{I}} \mathrm{conv}\Big(\mathcal{R}_*^{(\sqsubset,\mathcal{I})}\Big)\Big) + \mathcal{O}' = \mathrm{conv}\Big(\bigcup_{\sqsubset,\mathcal{I}} \mathcal{R}_*^{(\sqsubset,\mathcal{I})}\Big)$. Together with Theorem 3 and the convexity of $\overline{\mathcal{R}}$ (Remark 1) we obtain $\mathcal{R} \subseteq \mathrm{conv}\Big(\bigcup_{\sqsubset,\mathcal{I}} \mathcal{R}_*^{(\sqsubset,\mathcal{I})}\Big) \subseteq \overline{\mathcal{R}}$, finishing the proof.

V. Summary and Discussion

We formulated an extension of the CEO problem with log-loss distortion where multiple description coding compensates for defaulting agents. A single-letter characterization of the set of achievable rates and distortions was provided under a suitable conditional independence assumption. This region has the remarkable feature that it comprises rates that are in general insufficient for successful typicality decoding. The proof used tools from convex analysis and submodularity theory as well as ideas from [2]. Possible future research directions are the incorporation of distortion constraints for arbitrary subsets of agents and the exploration of the achievable region in Theorem 1 and Corollary 1 for specific source distributions.

Acknowledgement

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions.

Appendix

A. A Sequence of Polyhedra

In this section we will use the notation of [16]. In particular, we shall call a closed convex set *line-free* if it does not contain a (straight) line. The *characteristic cone* of a closed convex set $\mathcal{C}$ is denoted $\operatorname{cc}\mathcal{C} := \{\boldsymbol{y} : \boldsymbol{x} + \lambda\boldsymbol{y} \in \mathcal{C} \text{ for all } \lambda \geq 0\}$ ($\boldsymbol{x} \in \mathcal{C}$ arbitrary) and $\operatorname{ext}(\mathcal{C})$ is the set of all *extreme points* of $\mathcal{C}$.

For each $j \in [0:J]$, define the closed convex polyhedron $\mathcal{S}^{(j)} := \{\boldsymbol{x} \in \mathbb{R}^{K+j} : \boldsymbol{A}^{(j)}\boldsymbol{x} \geq \boldsymbol{b}^{(j)}\}$, where $\boldsymbol{A}^{(j)}$ is a matrix and $\boldsymbol{b}^{(j)}$ a vector of appropriate dimension. We make the following three assumptions:

1) $\boldsymbol{A}^{(j)}$ and $\boldsymbol{b}^{(j)}$ are defined recursively as
$$\boldsymbol{A}^{(j)} := \begin{pmatrix} \boldsymbol{A}^{(j-1)} & \boldsymbol{0} \\ \boldsymbol{0}^{\mathrm{T}} & 1 \\ \boldsymbol{e}_j^{\mathrm{T}} & 1 \end{pmatrix}, \quad \boldsymbol{b}^{(j)} = \begin{pmatrix} \boldsymbol{b}^{(j-1)} \\ c_1^{(j)} \\ c_2^{(j)} \end{pmatrix},$$
where $\boldsymbol{e}_j$ is the jth unit vector of appropriate dimension and $c_1^{(j)}$ and $c_2^{(j)}$ are arbitrary reals.
2) Each entry of $\boldsymbol{A}^{(0)}$ equals 0 or 1 and for all $k \in \mathcal{K}$ at least one row of $\boldsymbol{A}^{(0)}$ is equal to $\boldsymbol{e}_k^{\mathrm{T}}$. Due to assumption 1, this also implies that each entry of $\boldsymbol{A}^{(j)}$ is in $\{0,1\}$ and for all $k \in [K+j]$ at least one row of $\boldsymbol{A}^{(j)}$ is equal to $\boldsymbol{e}_k^{\mathrm{T}}$.
3) For any extreme point $\boldsymbol{x} \in \operatorname{ext}(\mathcal{S}^{(0)})$ and any $j \in \mathcal{J}$, assume $\boldsymbol{x}_j \leq c_2^{(j)} - c_1^{(j)}$.

Lemma 1. *Under assumptions 1 to 3, for every $j_0 \in [0:J]$ and every extreme point $\boldsymbol{y} \in \operatorname{ext}(\mathcal{S}^{(j_0)})$ there is an extreme point $\boldsymbol{x} \in \operatorname{ext}(\mathcal{S}^{(0)})$ and a subset $\mathcal{I}_{j_0} \subseteq [j_0]$ such that $\boldsymbol{y}_K = \boldsymbol{x}_K$ and for every $j \in \mathcal{J}$,*
$$\boldsymbol{y}_j = \begin{cases} \boldsymbol{x}_j, & j \notin \mathcal{I}_{j_0}, \\ c_2^{(j)} - c_1^{(j)}, & j \in \mathcal{I}_{j_0}, \end{cases} \tag{30}$$
and for every $j \in [j_0]$,
$$\boldsymbol{y}_{K+j} = \begin{cases} c_2^{(j)} - \boldsymbol{x}_j, & j \notin \mathcal{I}_{j_0}, \\ c_1^{(j)}, & j \in \mathcal{I}_{j_0}. \end{cases} \tag{31}$$

Proof. By assumption 2, for every $j \in \mathcal{J}$, $\mathcal{S}^{(j)}$ is line-free and can be written [16, Lemma 6, p. 25] as $\mathcal{S}^{(j)} = \operatorname{cc}(\mathcal{S}^{(j)}) + \operatorname{conv}\big(\operatorname{ext}(\mathcal{S}^{(j)})\big)$. Assumption 2 also implies $\operatorname{cc}(\mathcal{S}^{(j)}) = \{\boldsymbol{x} : \boldsymbol{x} \geq 0\}$. Let us proceed inductively over $j_0 \in [0:J]$. For $j_0 = 0$ the statement is trivial.

Given any $\boldsymbol{z} \in \operatorname{ext}(\mathcal{S}^{(j_0)})$, we need to obtain $\boldsymbol{x} \in \operatorname{ext}(\mathcal{S}^{(0)})$ and $\mathcal{I}_{j_0}$ such that $\boldsymbol{z}$ is given according to (30) and (31). Let $\boldsymbol{y} = \boldsymbol{z}_1^{K+j_0-1}$ be the truncation of $\boldsymbol{z}$. Exactly $K + j_0$ linear independent inequalities of $\boldsymbol{A}^{(j_0)}\boldsymbol{z} \geq \boldsymbol{b}^{(j_0)}$ are satisfied with equality, which is possible in only two different ways:

- *Construction I:* Exactly $K + j_0 - 1$ linear independent inequalities of $\boldsymbol{A}^{(j_0-1)}\boldsymbol{y} \geq \boldsymbol{b}^{(j_0-1)}$ are satisfied with equality, i.e., $\boldsymbol{y} \in \operatorname{ext}(\mathcal{S})^{(j_0-1)}$, and at least one of
$$\boldsymbol{z}_{K+j_0} \geq c_1^{(j_0)}, \tag{32}$$
$$\boldsymbol{z}_{j_0} + \boldsymbol{z}_{K+j_0} \geq c_2^{(j_0)}, \tag{33}$$
is satisfied with equality.
As $\boldsymbol{y} \in \operatorname{ext}(\mathcal{S})^{(j_0-1)}$, there exists $\boldsymbol{x} \in \operatorname{ext}(\mathcal{S}^{(0)})$ and $\mathcal{I}_{j_0-1}$ such that (30) holds for $j \in \mathcal{J}$ and (31) holds for $j \in [j_0 - 1]$ by the induction hypothesis. In particular $\boldsymbol{z}_{j_0} = \boldsymbol{x}_{j_0}$. Assuming that (33) holds with equality, we have $\boldsymbol{z}_{K+j_0} = c_2^{(j_0)} - \boldsymbol{x}_{j_0}$. Thus, the point $\boldsymbol{x}$ together with setting $\mathcal{I}_{j_0} = \mathcal{I}_{j_0-1}$ yields $\boldsymbol{z}$ from (30) and (31). Equality in (32) implies equality in (33) by assumption 3.
- *Construction II:* Exactly $K + j_0 - 2$ linear independent inequalities of $\boldsymbol{A}^{(j_0-1)}\boldsymbol{y} \geq \boldsymbol{b}^{(j_0-1)}$ are satisfied with equality and (32) and (33) are both satisfied with equality, too. Additionally these $K + j_0$ inequalities together need to be linearly independent. In particular, $\boldsymbol{y}$ is in the relative interior of a 1-dimensional face of $\mathcal{S}^{(j_0-1)}$ which can occur in two different ways.
We could have $\boldsymbol{y} = \lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{x}'$ for $\boldsymbol{x}, \boldsymbol{x}' \in \operatorname{ext}(\mathcal{S}^{(j_0-1)})$, $\boldsymbol{x} \neq \boldsymbol{x}'$ and $\lambda \in (0,1)$. This implies $\boldsymbol{z}_{K+j_0} = c_1^{(j_0)}$ and $\boldsymbol{y}_{j_0} = \lambda\boldsymbol{x}_{j_0} + (1-\lambda)\boldsymbol{x}'_{j_0} = c_2^{(j_0)} - c_1^{(j_0)}$, which by assumption 3 already implies $\boldsymbol{x}_{j_0} = \boldsymbol{x}'_{j_0} = c_2^{(j_0)} - c_1^{(j_0)}$. Thus, (32) and (33) are satisfied with equality for all $\lambda \in (0,1)$ and $\boldsymbol{z}$ cannot be an extreme point.
The second possibility is, that $\boldsymbol{y}$ is on an extreme ray of $\mathcal{S}^{(j_0-1)}$, i.e., $\boldsymbol{y} = \boldsymbol{x} + \lambda\boldsymbol{e}_{j'}$ for some $\boldsymbol{x} \in \operatorname{ext}(\mathcal{S}^{(j_0-1)})$, $\lambda > 0$ and $j' \in [K + j_0 - 1]$. If $j' \neq j_0$, (32) and (33) are satisfied for all $\lambda > 0$ and thus $\boldsymbol{z}$ cannot be an extreme point. As $j' = j_0$, the point $\boldsymbol{x}$ with $\mathcal{I}_{j_0} = \mathcal{I}_{j_0-1} \cup j_0$ yields the desired extreme point. ■

References

[1] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 887–902, May 1996.
[2] T. A. Courtade and T. Weissman, "Multiterminal source coding under logarithmic loss," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 740–761, Jan. 2014.
[3] A. A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 28, no. 6, pp. 851–857, Nov. 1982.
[4] V. K. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Process. Mag.*, vol. 18, no. 5, pp. 74–93, Sep. 2001.
[5] J. Chen and T. Berger, "Robust distributed source coding," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3385–3398, 2008.
[6] S. Fujishige, *Submodular Functions and Optimization*, 2nd ed., ser. Annals of Discrete Mathematics. Elsevier Science, 2005.
[7] G. Pichler, P. Piantanida, and G. Matz, "Distributed information-theoretic biclustering of two memoryless sources," in *Proc. 53rd Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Sep. 2015, pp. 426–433.
[8] ——, "Distributed information-theoretic biclustering," in *Proc. IEEE Int. Symp. on Inform. Theory*, Barcelona, Spain, Jul. 2016, pp. 1083–1087.
[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 2006.
[10] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.
[11] A. Wagner and V. Anantharam, "An improved outer bound for multiterminal source coding," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1919–1937, May 2008.
[12] T. S. Han and K. Kobayashi, "A unified achievable rate region for a general class of multiterminal source coding systems," *IEEE Trans. Inf. Theory*, vol. 26, no. 3, pp. 277–288, May 1980.
[13] J. Chen, X. Zhang, T. Berger, and S. B. Wicker, "An upper bound on the sum-rate distortion function and its corresponding rate allocation schemes for the CEO problem," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 6, pp. 977–987, 2004.
[14] M. Gastpar, "The Wyner-Ziv problem with multiple sources," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2762–2768, 2004.
[15] G. Pichler, P. Piantanida, and G. Matz, "Distributed information-theoretic clustering," *IEEE Trans. Inf. Theory*, 2017, (submitted). [Online]. Available: https://arxiv.org/abs/1602.04605
[16] B. Grünbaum, *Convex Polytopes*. Springer, New York, 2003.