The Sum Rate of Vector Gaussian Multiple Description Coding With Tree-Structured Covariance Distortion Constraints

Yinfei Xu, *Student Member, IEEE*, Jun Chen, *Senior Member, IEEE*, and Qiao Wang, *Member, IEEE*

Abstract—A single-letter lower bound on the sum rate of multiple description coding with tree-structured distortion constraints is established by generalizing Ozarow's celebrated converse argument through the introduction of auxiliary random variables that form a Markov tree. For the quadratic vector Gaussian case, this lower bound is shown to be achievable by an extended El Gamal-Cover scheme, yielding a complete characterization of the minimum sum rate.

Index Terms—Auxiliary random variable, covariance distortion, Karush-Kuhn-Tucker conditions, Markov structure, multiple description coding, sum rate, vector Gaussian source.

I. INTRODUCTION

IN MULTIPLE description coding, a source is encoded into M descriptions with rates $R_1, \cdots, R_M$, respectively, such that each non-empty subset of these descriptions can be used to produce a (possibly lossy) reconstruction of the source. The fundamental information-theoretic problem here is to characterize the rate region, i.e., the closure of the set of the admissible rate tuples $(R_1, \cdots, R_M)$, subject to the given distortion constraints on the reconstructions. Early work on this problem was mostly devoted to the two-description case. In particular, El Gamal and Cover [1] derived an inner bound of the two-description rate region, which was shown to be tight in the no excess sum-rate case [2] and in the case where one of the descriptions is required to reconstruct a deterministic function of the source [3]; Zhang and Berger [4] obtained an improved inner bound of the two-description rate region by incorporating a common description layer into the El Gamal-Cover scheme. In contrast, recent years have seen extensive research on the general M-description case (see, e.g., [5]–[9]). Although these investigations have revealed many interesting results and considerably deepened our understanding of the subject, a complete characterization of the rate region, even for the two-description case, is still widely considered to be out of reach.

One way to simplify the problem is to consider a special class of sources and distortion measures. In this respect, special attentions have been paid to the Gaussian source and the mean squared error distortion measure (known as the quadratic Gaussian case). In his lauded paper [10], Ozarow proved the tightness of the El Gamal-Cover inner bound for the quadratic Gaussian two-description problem via an ingenious converse argument, in which an auxiliary random variable is introduced to exploit an implicit conditional independence structure in the El Gamal-Cover scheme. Ozarow's work ignited the hope of solving the general quadratic Gaussian M-description problem. Unfortunately, this task turns out to be rather formidable (if not impossible). Nevertheless, the bounding technique in [10] has been extended to obtain conclusive results for some special cases where the reconstruction distortion constraints are only imposed on certain subsets of descriptions, including the quadratic Gaussian M-description problem with individual and central distortion constraints [11] as well as the quadratic Gaussian M-description problem with individual and hierarchical distortion constraints [12]. We go one step further in this work by considering the more general tree-structured distortion constraints. It will been seen that Ozarow's converse argument admits a natural generalization in this context, leading to a complete characterization of the minimum sum rate.

The importance of conditional independence structures in the converse arguments is widely recognized, especially for the distributed source coding problems (see, e.g., [13]–[21]). It should be noted that, in distributed source coding, the conditional independence structures,[1] which are either directly present in the source models or created through the introduction of auxiliary random variables, do not depend on the adopted schemes. In contrast, due to its centralized encoding nature, no such non-trivial scheme-independent conditional independence structures exist in multiple description

Y. Xu was with the School of Information Science and Engineering, Southeast University, Nanjing 210096, China. He is now with the Institute of Network Coding, The Chinese University of Hong Kong, Hong Kong (e-mail: yinfeixu@inc.cuhk.edu.hk).

J. Chen is with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4K1, Canada (e-mail: junchen@ece.mcmaster.ca).

Q. Wang is with the School of Information Science and Engineering, Southeast University, Nanjing 210096, China (e-mail: qiaowang@seu.edu.cn).

[1]For example, in the CEO problem [13], the outputs (as well as the inputs) of different encoders are conditionally independent given the remote source.

Fig. 1.   The multiple description coding system.

Fig. 2.   A perfect binary tree.
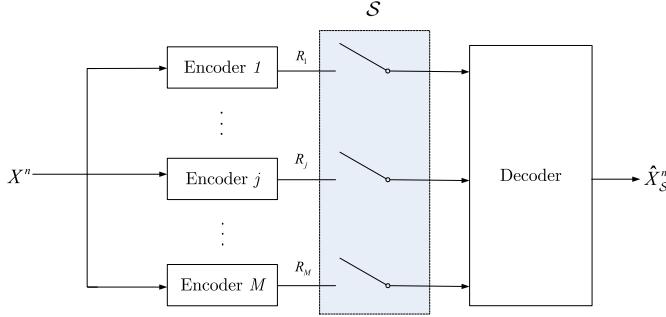
coding. Indeed, Ozarow's converse argument is tailored to an implicit conditional independence structure which is specific to the El Gamal-Cover scheme. In this work we extend Ozarow's method to cope with more sophisticated conditional independence structures in a generalized El Gamal-Cover scheme optimized for tree-structured distortion constraints.

The remainder of this paper is organized as follows. The problem formulation and the main results are stated in Section II. Section III is devoted to characterizing the minimum sum rate of vector Gaussian multiple description coding with tree-structured covariance distortion constraints. We conclude the paper in Section IV.

II. PROBLEM FORMULATION AND MAIN RESULTS

Consider a multiple description coding system (see Fig. 1) with M ($M \geq 2$) encoders, each generating a description of the source, and a decoder which produces a reconstruction of the source based on the received descriptions. We assume that the source is an i.i.d. process $\{X(t)\}_{t=1}^{\infty}$ with marginal distribution $p(x)$ over alphabet $\mathcal{X}$, and let

$$d : \mathcal{X} \times \hat{\mathcal{X}} \to [0, \infty]$$

be a distortion measure, where $\hat{\mathcal{X}}$ is the reconstruction alphabet.

Definition 1: Let $\mathcal{T}$ be a collection of nonempty subsets of $\mathcal{I}_M \triangleq \{1, \cdots, M\}$. We say that a rate tuple $(R_1, \cdots, R_M)$ is achievable subject to distortion constraints $(d_{\mathcal{S}})_{\mathcal{S} \in \mathcal{T}}$ if there exist encoding functions

$$\varphi_j^{(n)} : \mathcal{X}^n \to \mathcal{C}_j^{(n)}, \quad j = 1, \cdots, M,$$

and decoding functions

$$\psi_{\mathcal{S}}^{(n)} : \prod_{j \in \mathcal{S}} \mathcal{C}_j^{(n)} \to \hat{\mathcal{X}}^n, \quad \mathcal{S} \in \mathcal{T},$$

such that

$$\frac{1}{n} \log |\mathcal{C}_j^{(n)}| \leq R_j, \quad j = 1, \cdots, M,$$

$$\frac{1}{n} \sum_{t=1}^{n} \mathbb{E}[d(X(t), \hat{X}_{\mathcal{S}}(t))] \leq d_{\mathcal{S}}, \quad \mathcal{S} \in \mathcal{T},$$

where $\hat{X}_{\mathcal{S}}(t)$ is the t-th component of the reconstruction vector $\hat{X}_{\mathcal{S}}^n \triangleq \psi_{\mathcal{S}}^{(n)}((\varphi_j^{(n)}(X^n))_{j \in \mathcal{S}})$. The rate region $\mathcal{R}((d_{\mathcal{S}})_{\mathcal{S} \in \mathcal{T}})$ is
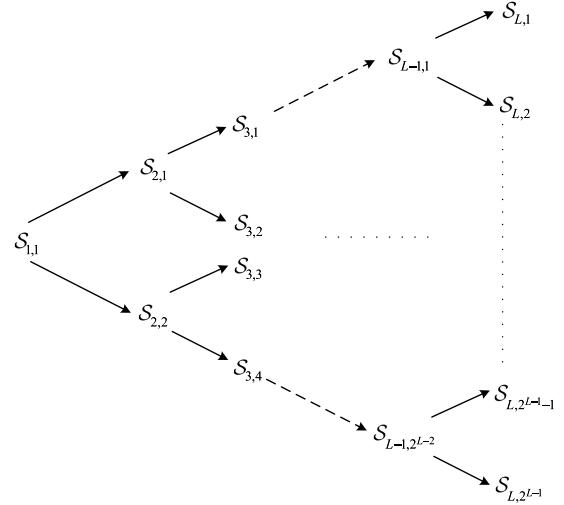
the closure of the set of all such achievable rate tuples, and the minimum sum rate $R((d_{\mathcal{S}})_{\mathcal{S} \in \mathcal{T}})$ is defined as

$$R((d_{\mathcal{S}})_{\mathcal{S} \in \mathcal{T}}) = \min_{(R_1, \cdots, R_M) \in \mathcal{R}((d_{\mathcal{S}})_{\mathcal{S} \in \mathcal{T}})} \sum_{j=1}^{M} R_j.$$

Choosing $\mathcal{T} = 2_+^{\mathcal{I}_M}$ gives the generic form of M-description coding, where $2_+^{\mathcal{I}_M}$ is the collection of all nonempty subsets of $\mathcal{I}_M$. It is clear that every other choice of $\mathcal{T}$ corresponds to a degenerate case of the generic form with $d_{\mathcal{S}} = \infty$ for $\mathcal{S} \in 2_+^{\mathcal{I}_M} \backslash \mathcal{T}$. In this work we consider the case where $\mathcal{T}$ has a tree structure. Specifically, we say that $\mathcal{T}$ has a tree structure if, for any $\mathcal{S}_1, \mathcal{S}_2 \in \mathcal{T}$, one of the following statements is true:

1) $\mathcal{S}_1 \subseteq \mathcal{S}_2$,
2) $\mathcal{S}_2 \subseteq \mathcal{S}_1$,
3) $\mathcal{S}_1 \cap \mathcal{S}_2 = \emptyset$.

The following settings considered in the literature correspond to two special tree structures:

- individual and central distortion constraints [11], i.e., $\mathcal{T} = \{\{1\}, \cdots, \{M\}, \{1, \cdots, M\}\}$,
- individual and hierarchical distortion constraints [12], i.e., $\mathcal{T} = \{\{1\}, \cdots, \{M\}, \{1, 2\}, \cdots, \{1, \cdots, M\}\}$.

Note that $2_+^{\mathcal{I}_M}$ has a tree structure if and only if $M = 2$ (assuming $M \geq 2$); in this sense, the two-description problem is inherently simpler than the general M-description problem.

For notational simplicity, henceforth we shall assume that $M = 2^{L-1}$ for some $L \geq 2$ and that $\mathcal{T}$ has a perfect binary-tree structure (see Fig. 2), i.e.,

$$\mathcal{T} = \{\mathcal{S}_{k,i} : k = 1, \cdots, L; i = 1, \cdots, 2^{k-1}\},$$

where

$$\mathcal{S}_{k,i} = \left\{ j \in \mathbb{N} : \frac{2^L(i-1)}{2^k} < j \leq \frac{2^L i}{2^k} \right\}.$$

It is clear that this assumption incurs no loss of generality since every tree can be converted to a perfect binary tree via inserting dummy nodes (which corresponds to inserting dummy descriptions and imposing redundant distortion constraints in

the multiple description coding system) and relabelling. For example, Fig. 3 illustrates how to convert the following tree-structured distortion constraints

$$d_{\{1\}}, d_{\{2\}}, d_{\{4\}}, d_{\{5\}}, d_{\{3,4,5\}}, d_{\{1,2,3,4,5\}} \quad (1)$$

to a new set of distortion constraints with a perfect binary-tree structure.

We shall establish a single-letter lower bound on the sum rate of multiple description coding for the case where $\mathcal{T}$ has a perfect binary-tree structure. The proof is based on the idea of augmenting the probability space with auxiliary random variables that form a Markov tree. This idea was originated in Ozarow's seminal work [10], in which a single auxiliary random variable is introduced. Later [12], [22], [23] considered a more general construction with multiple auxiliary random variables forming a Markov chain (see also [24] for a related construction). Our Markov-tree construction can be viewed as a further generalization along this line of development.

Let $\mathcal{P}$ denote the set of conditional distributions $p(\underline{z}|x)$ satisfying the binary-Markov-tree condition

$$p(\underline{z}|x) = p(z_{1,1}|x) \prod_{k=1}^{L-2} \prod_{i=1}^{2^{k-1}} p(z_{k+1,2i-1}, z_{k+1,2i}|z_{k,i}), \quad (2)$$

where $\underline{z} = (z_{k,i})_{k=1,\cdots,L-1;i=1,\cdots,2^{k-1}}$. Moreover, let $\mathcal{P}(\underline{d})$ denote the set of conditional distributions $p(\hat{\underline{x}}|x)$ such that the induced $p(x)p(\hat{\underline{x}}|x)$ satisfies

$$\mathbb{E}[d(X, \hat{X}_{\mathcal{S}_{k,i}})] \le d_{\mathcal{S}_{k,i}}, \quad k = 1, \cdots, L; \ i = 1, \cdots, 2^{k-1},$$

where $\underline{d} = (d_{\mathcal{S}_{k,i}})_{k=1,\cdots,L;i=1,\cdots,2^{k-1}}$ and $\hat{\underline{x}} = (\hat{x}_{\mathcal{S}_{k,i}})_{k=1,\cdots,L;i=1,\cdots,2^{k-1}}$. Our single-letter lower bound is obtained by optimizing a certain functional over auxiliary random variables $\underline{Z} \triangleq (Z_{k,i})_{k=1,\cdots,L-1;i=1,\cdots,2^{k-1}}$ and reconstruction variables $\hat{\underline{X}} \triangleq (\hat{X}_{\mathcal{S}_{k,i}})_{k=1,\cdots,L;i=1,\cdots,2^{k-1}}$ constrained by $\mathcal{P}$ and $\mathcal{P}(\underline{d})$, respectively. Specifically, let

$$\eta(p(\hat{x}|x), p(\underline{z}|x)) = I(X; \hat{X}_{\mathcal{S}_{1,1}}|Z_{1,1})$$
$$+ \sum_{k=1}^{L-2} \sum_{i=1}^{2^{k-1}} \Big(I(Z_{k,i}; \hat{X}_{\mathcal{S}_{k+1,2i-1}}|Z_{k+1,2i-1})$$
$$+ I(Z_{k,i}; \hat{X}_{\mathcal{S}_{k+1,2i}}|Z_{k+1,2i}) \Big)$$
$$+ \sum_{i=1}^{2^{L-2}} \Big(I(Z_{L-1,i}; \hat{X}_{\mathcal{S}_{L,2i-1}}) + I(Z_{L-1,i}; \hat{X}_{\mathcal{S}_{L,2i}}) \Big)$$

with $p(\underline{z}, x, \hat{\underline{x}}) = p(\underline{z}|x)p(x)p(\hat{\underline{x}}|x)$, and define

$$r(\underline{d}) = \inf_{p(\hat{x}|x) \in \mathcal{P}(\underline{d})} \sup_{p(\underline{z}|x) \in \mathcal{P}} \eta(p(\hat{x}|x), p(\underline{z}|x)).$$

Theorem 1: $R(\underline{d}) \ge r(\underline{d})$.
Proof: See Appendix A. ∎

In this work, we focus on the case where $\{X(t)\}_{t=1}^{\infty}$ is an i.i.d. process with each $X(t)$ being an $m \times 1$ Gaussian random vector of mean zero and covariance matrix $\mathbf{\Sigma}_X \succ \mathbf{0}$. Moreover, we adopt the covariance distortion constraints

$$\frac{1}{n} \sum_{t=1}^{n} \mathbb{E}[(X(t) - \hat{X}_{\mathcal{S}_{k,i}}(t))(X(t) - \hat{X}_{\mathcal{S}_{k,i}}(t))^T] \preceq \mathbf{D}_{\mathcal{S}_{k,i}},$$
$$k = 1, \cdots, L; \quad i = 1, \cdots, 2^{k-1},$$
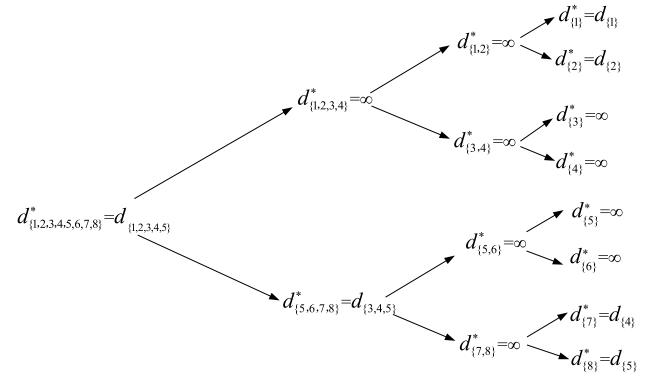
Fig. 3. Convert tree-structured distortion constraints in (1) to a new set of distortion constraints with a perfect binary-tree structure.

and, without loss of generality,[2] assume that $\mathbf{0} \prec \mathbf{D}_{\mathcal{S}_{k,i}} \preceq \mathbf{\Sigma}_X$, $k = 1, \cdots, L; i = 1, \cdots, 2^{k-1}$. The corresponding minimum sum rate will be denoted by $R_G(\underline{\mathbf{D}})$, where $\underline{\mathbf{D}} = ((\mathbf{D}_{\mathcal{S}_{k,i}})_{k=1,\cdots,L;i=1,\cdots,2^{k-1}})$.

The next result provides a computable[3] characterization of $R_G(\underline{\mathbf{D}})$. Define

$$R_G^*(\underline{\mathbf{D}})$$
$$= \max_{\underline{\mathbf{\Theta}}} \frac{1}{2} \log \frac{|\mathbf{\Sigma}_X|}{|\mathbf{D}_{\mathcal{S}_{1,1}}|}$$
$$+ \sum_{k=1}^{L-1} \sum_{i=1}^{2^{k-1}} \Bigg(\frac{1}{2} \log \frac{|\mathbf{\Sigma}_X|}{|\mathbf{D}_{\mathcal{S}_{k+1,2i-1}} \mathbf{\Sigma}_X^{-1}(\mathbf{\Sigma}_X - \mathbf{\Theta}_{k,i}) + \mathbf{\Theta}_{k,i}|}$$
$$+ \frac{1}{2} \log \frac{|\mathbf{D}_{\mathcal{S}_{k,i}} \mathbf{\Sigma}_X^{-1}(\mathbf{\Sigma}_X - \mathbf{\Theta}_{k,i}) + \mathbf{\Theta}_{k,i}|}{|\mathbf{D}_{\mathcal{S}_{k+1,2i}} \mathbf{\Sigma}_X^{-1}(\mathbf{\Sigma}_X - \mathbf{\Theta}_{k,i}) + \mathbf{\Theta}_{k,i}|} \Bigg)$$
$$\text{subject to} \quad \mathbf{\Sigma}_X \succeq \mathbf{\Theta}_{k+1,2i-1}, \mathbf{\Theta}_{k+1,2i} \succeq \mathbf{\Theta}_{k,i} \succeq \mathbf{0},$$
$$k = 1, \cdots, L-2; \quad i = 1, \cdots, 2^{k-1}, \quad (3)$$

Theorem 2: $R_G(\underline{\mathbf{D}}) = R_G^*(\underline{\mathbf{D}})$.
Proof: See Section III. ∎

III. PROOF OF THEOREM 2

Translating Theorem 1 to the quadratic vector Gaussian setting gives

$$R_G(\underline{\mathbf{D}}) \ge r_G(\underline{\mathbf{D}})$$
$$\triangleq \inf_{p(\hat{x}|x) \in \mathcal{P}(\underline{\mathbf{D}})} \sup_{p(\underline{z}|x) \in \mathcal{P}} \eta(p(\hat{x}|x), p(\underline{z}|x)),$$

where $\mathcal{P}(\underline{\mathbf{D}})$ denotes the set of conditional distributions $p(\hat{\underline{x}}|x)$ such that the induced $p(x)p(\hat{\underline{x}}|x)$ satisfies

$$\mathbb{E}[(X - \hat{X}_{\mathcal{S}_{k,i}})(X - \hat{X}_{\mathcal{S}_{k,i}})^T] \preceq \mathbf{D}_{\mathcal{S}_{k,i}},$$
$$k = 1, \cdots, L; \quad i = 1, \cdots, 2^{k-1}.$$

---

[2]Under the covariance distortion constraints, the optimal reconstruction is given by the MMSE estimate of the source based on the received descriptions, for which the resulting distortion must be less than or equal to the source covariance matrix in the semidefinite sense. It is also clear that a distortion constraint is redundant if it is set to be equal to the source covariance matrix.

[3]An expression is said to be computable if it can be evaluated by solving a finite-dimensional optimization problem.

We divide the proof of Theorem 2 into three parts. Section III-A is devoted to establishing the converse part; specifically, by restricting the auxiliary random variables $\underline{Z}$ to form a Gauss-Markov chain with the source variable X, we leverage the worst additive noise lemma to derive a lower bound on $r_G(\underline{\mathbf{D}})$, which turns out to coincide with $R_G^*(\underline{\mathbf{D}})$. In Section III-B, it is shown via the enhancement argument that $R_G^*(\underline{\mathbf{D}})$ can be written in an alternative form by replacing $\underline{\mathbf{D}}$ with a set of (potentially stronger) covariance distortion constraints. The achievability part is presented in Section III-C, where we construct a generalized El Gamal-Cover scheme for this new set of distortion constraints with the sum rate equal to $R_G^*(\underline{\mathbf{D}})$.

A. The Converse Part

Let $W_{k,i}$ be an $m \times 1$ Gaussian random vector of mean zero and covariance matrix $\boldsymbol{\Sigma}_{W_{k,i}} \succ \mathbf{0}$, $k = 1, \cdots, L-1$; $i = 1, \cdots, 2^{k-1}$. We assume that X, $W_{k,i}$, $k = 1, \cdots, L-1$; $i = 1, \cdots, 2^{k-1}$, are mutually independent. Define

$$N_{1,1} = W_{1,1},$$

$$N_{k+1,2i-1} = N_{k,i} + W_{k+1,2i-1},$$
$$k = 1, \cdots, L-2; \quad i = 1, \cdots, 2^{k-1},$$

$$N_{k+1,2i} = N_{k,i} + W_{k+1,2i},$$
$$k = 1, \cdots, L-2; \quad i = 1, \cdots, 2^{k-1}.$$

The covariance matrix of $N_{k,i}$ is denoted by $\boldsymbol{\Sigma}_{N_{k,i}}$, $k = 1, \cdots, L-1$; $i = 1, \cdots, 2^{k-1}$. Let

$$Z_{k,i} = X + N_{k,i}, \quad k = 1, \cdots, L-1; \ i = 1, \cdots, 2^{k-1}.$$

It is clear that $\underline{Z}$ and X form a binary Gauss-Markov tree; as a consequence, we have $p(\underline{z}|x) \in \mathcal{P}$.

Let $\underline{\hat{X}}$ be jointly distributed with X and $\underline{Z}$ such that $p(\underline{\hat{x}}|x) \in \mathcal{P}(\underline{\mathbf{D}})$ and $\underline{Z} \leftrightarrow X \leftrightarrow \underline{\hat{X}}$ form a Markov chain. Note that

$$I(X; \hat{X}_{\mathcal{S}_{1,1}} | Z_{1,1})$$
$$= I(X; \hat{X}_{\mathcal{S}_{1,1}}) - I(Z_{1,1}; \hat{X}_{\mathcal{S}_{1,1}})$$
$$= h(X) - h(X + N_{1,1}) + h(X + N_{1,1}|\hat{X}_{\mathcal{S}_{1,1}}) - h(X|\hat{X}_{\mathcal{S}_{1,1}})$$
$$= h(X) - h(X + N_{1,1}) + I(N_{1,1}; X + N_{1,1}|\hat{X}_{\mathcal{S}_{1,1}})$$
$$= \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X|}{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{N_{1,1}}|} + I(N_{1,1}; X + N_{1,1}|\hat{X}_{\mathcal{S}_{1,1}})$$
$$\geq \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X||\mathbf{D}_{\mathcal{S}_{1,1}} + \boldsymbol{\Sigma}_{N_{1,1}}|}{|\mathbf{D}_{\mathcal{S}_{1,1}}||\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{N_{1,1}}|}, \tag{4}$$

where (4) follows from the conditional version of the worst additive noise lemma [25], [26]. Similarly, we have

$$I(Z_{k,i}; \hat{X}_{\mathcal{S}_{k+1,2i-1}} | Z_{k+1,2i-1})$$
$$\geq \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{N_{k,i}}||\mathbf{D}_{\mathcal{S}_{k+1,2i-1}} + \boldsymbol{\Sigma}_{N_{k+1,2i-1}}|}{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{N_{k+1,2i-1}}||\mathbf{D}_{\mathcal{S}_{k+1,2i-1}} + \boldsymbol{\Sigma}_{N_{k,i}}|},$$
$$k = 1, \cdots, L-2; \quad i = 1, \cdots, 2^{k-1}, \tag{5}$$

$$I(Z_{k,i}; \hat{X}_{\mathcal{S}_{k+1,2i}} | Z_{k+1,2i})$$
$$\geq \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{N_{k,i}}||\mathbf{D}_{\mathcal{S}_{k+1,2i}} + \boldsymbol{\Sigma}_{N_{k+1,2i}}|}{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{N_{k+1,2i}}||\mathbf{D}_{\mathcal{S}_{k+1,2i}} + \boldsymbol{\Sigma}_{N_{k,i}}|},$$
$$k = 1, \cdots, L-2; \quad i = 1, \cdots, 2^{k-1}. \tag{6}$$

Moreover, it is clear that

$$I(Z_{L-1,i}; \hat{X}_{\mathcal{S}_{L,2i-1}}) \geq \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{N_{L-1,i}}|}{|\mathbf{D}_{\mathcal{S}_{L,2i-1}} + \boldsymbol{\Sigma}_{N_{L-1,i}}|},$$
$$i = 1, \cdots, 2^{L-2}, \tag{7}$$

$$I(Z_{L-1,i}; \hat{X}_{\mathcal{S}_{L,2i}}) \geq \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{N_{L-1,i}}|}{|\mathbf{D}_{\mathcal{S}_{L,2i}} + \boldsymbol{\Sigma}_{N_{L-1,i}}|},$$
$$i = 1, \cdots, 2^{L-2}. \tag{8}$$

Combining (4)–(8) proves

$$r_G(\underline{\mathbf{D}})$$
$$\geq \sup_{\boldsymbol{\Sigma}_{N_{k,i}}} \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X||\mathbf{D}_{\mathcal{S}_{1,1}} + \boldsymbol{\Sigma}_{N_{1,1}}|}{|\mathbf{D}_{\mathcal{S}_{1,1}}||\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{N_{1,1}}|}$$
$$+ \sum_{k=1}^{L-2} \sum_{i=1}^{2^{k-1}} \left(\frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{N_{k,i}}||\mathbf{D}_{\mathcal{S}_{k+1,2i-1}} + \boldsymbol{\Sigma}_{N_{k+1,2i-1}}|}{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{N_{k+1,2i-1}}||\mathbf{D}_{\mathcal{S}_{k+1,2i-1}} + \boldsymbol{\Sigma}_{N_{k,i}}|} \right.$$
$$\left. + \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{N_{k,i}}||\mathbf{D}_{\mathcal{S}_{k+1,2i}} + \boldsymbol{\Sigma}_{N_{k+1,2i}}|}{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{N_{k+1,2i}}||\mathbf{D}_{\mathcal{S}_{k+1,2i}} + \boldsymbol{\Sigma}_{N_{k,i}}|} \right)$$
$$+ \sum_{i=1}^{2^{L-2}} \left(\frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{N_{L-1,i}}|}{|\mathbf{D}_{\mathcal{S}_{L,2i-1}} + \boldsymbol{\Sigma}_{N_{L-1,i}}|} \right.$$
$$\left. + \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{N_{L-1,i}}|}{|\mathbf{D}_{\mathcal{S}_{L,2i}} + \boldsymbol{\Sigma}_{N_{L-1,i}}|} \right)$$
$$\text{subject to } \boldsymbol{\Sigma}_{N_{k+1,2i-1}}, \boldsymbol{\Sigma}_{N_{k+1,2i}} \succ \boldsymbol{\Sigma}_{N_{k,i}} \succ \mathbf{0},$$
$$k = 1, \cdots, L-2; i = 1, \cdots, 2^{k-1}. \tag{9}$$

It can be verified that

$$\frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X||\mathbf{D}_{\mathcal{S}_{1,1}} + \boldsymbol{\Sigma}_{N_{1,1}}|}{|\mathbf{D}_{\mathcal{S}_{1,1}}||\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{N_{1,1}}|}$$
$$+ \sum_{k=1}^{L-1} \sum_{i=1}^{2^{k-1}} \left(\frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{N_{k,i}}|}{|\mathbf{D}_{\mathcal{S}_{k+1,2i-1}} + \boldsymbol{\Sigma}_{N_{k,i}}|} \right.$$
$$\left. + \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{N_{k,i}}|}{|\mathbf{D}_{\mathcal{S}_{k+1,2i}} + \boldsymbol{\Sigma}_{N_{k,i}}|} \right)$$
$$- \sum_{k=1}^{L-2} \sum_{i=1}^{2^{k-1}} \left(\frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{N_{k+1,2i-1}}|}{|\mathbf{D}_{\mathcal{S}_{k+1,2i-1}} + \boldsymbol{\Sigma}_{N_{k+1,2i-1}}|} \right.$$
$$\left. + \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{N_{k+1,2i}}|}{|\mathbf{D}_{\mathcal{S}_{k+1,2i}} + \boldsymbol{\Sigma}_{N_{k+1,2i}}|} \right)$$
$$= \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X|}{|\mathbf{D}_{\mathcal{S}_{1,1}}|}$$
$$+ \sum_{k=1}^{L-1} \sum_{i=1}^{2^{k-1}} \left(\frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{N_{k,i}}|}{|\mathbf{D}_{\mathcal{S}_{k+1,2i-1}} + \boldsymbol{\Sigma}_{N_{k,i}}|} \right.$$
$$\left. + \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{N_{k,i}}|}{|\mathbf{D}_{\mathcal{S}_{k+1,2i}} + \boldsymbol{\Sigma}_{N_{k,i}}|} \right)$$
$$- \sum_{k=1}^{L-1} \sum_{i=1}^{2^{k-1}} \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{N_{k,i}}|}{|\mathbf{D}_{\mathcal{S}_{k,i}} + \boldsymbol{\Sigma}_{N_{k,i}}|}$$
$$= \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X|}{|\mathbf{D}_{\mathcal{S}_{1,1}}|}$$
$$+ \sum_{k=1}^{L-1} \sum_{i=1}^{2^{k-1}} \frac{1}{2} \log \frac{|\mathbf{D}_{\mathcal{S}_{k,i}} + \boldsymbol{\Sigma}_{N_{k,i}}||\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{N_{k,i}}|}{|\mathbf{D}_{\mathcal{S}_{k+1,2i-1}} + \boldsymbol{\Sigma}_{N_{k,i}}||\mathbf{D}_{\mathcal{S}_{k+1,2i}} + \boldsymbol{\Sigma}_{N_{k,i}}|}.$$

Now set

$$\boldsymbol{\Theta}_{k,i} = (\boldsymbol{\Sigma}_X^{-1} + \boldsymbol{\Sigma}_{N_{k,i}}^{-1})^{-1},$$
$$k = 1, \cdots, L-1; \quad i = 1, \cdots, 2^{k-1}.$$

Note that

$$\boldsymbol{\Sigma}_{N_{k+1,2i-1}}, \boldsymbol{\Sigma}_{N_{k+1,2i}} \succ \boldsymbol{\Sigma}_{N_{k,i}} \succ \mathbf{0}$$
$$\Leftrightarrow \boldsymbol{\Sigma}_X \succ \boldsymbol{\Theta}_{k+1,2i-1}, \boldsymbol{\Theta}_{k+1,2i} \succ \boldsymbol{\Theta}_{k,i} \succ \mathbf{0},$$
$$k = 1, \cdots, L-2; \quad i = 1, \cdots, 2^{k-1},$$

and

$$\frac{1}{2} \log \frac{|\mathbf{D}_{\mathcal{S}_{k,i}} + \boldsymbol{\Sigma}_{N_{k,i}}||\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{N_{k,i}}|}{|\mathbf{D}_{\mathcal{S}_{k+1,2i-1}} + \boldsymbol{\Sigma}_{N_{k,i}}||\mathbf{D}_{\mathcal{S}_{k+1,2i}} + \boldsymbol{\Sigma}_{N_{k,i}}|}$$
$$= \frac{1}{2} \log \frac{|\mathbf{D}_{\mathcal{S}_{k,i}} + (\boldsymbol{\Theta}_{k,i}^{-1} - \boldsymbol{\Sigma}_X^{-1})^{-1}|}{|\mathbf{D}_{\mathcal{S}_{k+1,2i-1}} + (\boldsymbol{\Theta}_{k,i}^{-1} - \boldsymbol{\Sigma}_X^{-1})^{-1}|}$$
$$+ \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X + (\boldsymbol{\Theta}_{k,i}^{-1} - \boldsymbol{\Sigma}_X^{-1})^{-1}|}{|\mathbf{D}_{\mathcal{S}_{k+1,2i}} + (\boldsymbol{\Theta}_{k,i}^{-1} - \boldsymbol{\Sigma}_X^{-1})^{-1}|}$$
$$= \frac{1}{2} \log \frac{|\mathbf{D}_{\mathcal{S}_{k,i}} (\boldsymbol{\Theta}_{k,i}^{-1} - \boldsymbol{\Sigma}_X^{-1}) + \mathbf{I}_m|}{|\mathbf{D}_{\mathcal{S}_{k+1,2i-1}} (\boldsymbol{\Theta}_{k,i}^{-1} - \boldsymbol{\Sigma}_X^{-1}) + \mathbf{I}_m|}$$
$$+ \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X (\boldsymbol{\Theta}_{k,i}^{-1} - \boldsymbol{\Sigma}_X^{-1}) + \mathbf{I}_m|}{|\mathbf{D}_{\mathcal{S}_{k+1,2i}} (\boldsymbol{\Theta}_{k,i}^{-1} - \boldsymbol{\Sigma}_X^{-1}) + \mathbf{I}_m|}$$
$$= \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X (\mathbf{I}_m - \boldsymbol{\Sigma}_X^{-1} \boldsymbol{\Theta}_{k,i}) + \boldsymbol{\Theta}_{k,i}|}{|\mathbf{D}_{\mathcal{S}_{k+1,2i-1}} (\mathbf{I}_m - \boldsymbol{\Sigma}_X^{-1} \boldsymbol{\Theta}_{k,i}) + \boldsymbol{\Theta}_{k,i}|}$$
$$+ \frac{1}{2} \log \frac{|\mathbf{D}_{\mathcal{S}_{k,i}} (\mathbf{I}_m - \boldsymbol{\Sigma}_X^{-1} \boldsymbol{\Theta}_{k,i}) + \boldsymbol{\Theta}_{k,i}|}{|\mathbf{D}_{\mathcal{S}_{k+1,2i}} (\mathbf{I}_m - \boldsymbol{\Sigma}_X^{-1} \boldsymbol{\Theta}_{k,i}) + \boldsymbol{\Theta}_{k,i}|}$$
$$= \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X|}{|\mathbf{D}_{\mathcal{S}_{k+1,2i-1}} \boldsymbol{\Sigma}_X^{-1} (\boldsymbol{\Sigma}_X - \boldsymbol{\Theta}_{k,i}) + \boldsymbol{\Theta}_{k,i}|}$$
$$+ \frac{1}{2} \log \frac{|\mathbf{D}_{\mathcal{S}_{k,i}} \boldsymbol{\Sigma}_X^{-1} (\boldsymbol{\Sigma}_X - \boldsymbol{\Theta}_{k,i}) + \boldsymbol{\Theta}_{k,i}|}{|\mathbf{D}_{\mathcal{S}_{k+1,2i}} \boldsymbol{\Sigma}_X^{-1} (\boldsymbol{\Sigma}_X - \boldsymbol{\Theta}_{k,i}) + \boldsymbol{\Theta}_{k,i}|},$$
$$k = 1, \cdots, L-1; \quad i = 1, \cdots, 2^{k-1},$$

where $\mathbf{I}_m$ denotes the $m \times m$ identity matrix. Therefore, we can write (9) equivalently as

$$r_G(\mathbf{D})$$
$$\geq \sup_{\boldsymbol{\Theta}} \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X|}{|\mathbf{D}_{\mathcal{S}_{1,1}}|}$$
$$+ \sum_{k=1}^{L-1} \sum_{i=1}^{2^{k-1}} \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X|}{|\mathbf{D}_{\mathcal{S}_{k+1,2i-1}} \boldsymbol{\Sigma}_X^{-1} (\boldsymbol{\Sigma}_X - \boldsymbol{\Theta}_{k,i}) + \boldsymbol{\Theta}_{k,i}|}$$
$$+ \frac{1}{2} \log \frac{|\mathbf{D}_{\mathcal{S}_{k,i}} \boldsymbol{\Sigma}_X^{-1} (\boldsymbol{\Sigma}_X - \boldsymbol{\Theta}_{k,i}) + \boldsymbol{\Theta}_{k,i}|}{|\mathbf{D}_{\mathcal{S}_{k+1,2i}} \boldsymbol{\Sigma}_X^{-1} (\boldsymbol{\Sigma}_X - \boldsymbol{\Theta}_{k,i}) + \boldsymbol{\Theta}_{k,i}|}$$
$$\text{subject to } \boldsymbol{\Sigma}_X \succ \boldsymbol{\Theta}_{k+1,2i-1}, \boldsymbol{\Theta}_{k+1,2i} \succ \boldsymbol{\Theta}_{k,i} \succ \mathbf{0},$$
$$k = 1, \cdots, L-2; \quad i = 1, \cdots, 2^{k-1},$$

from which the desired result follows immediately.

B. An Alternative Form of $R_G^*(\mathbf{D})$

Consider the case where $\mathbf{0} \prec \mathbf{D}_{\mathcal{S}_{k,i}} \prec \boldsymbol{\Sigma}_X$, $k = 1, \cdots, L; i = 1, \cdots, 2^{k-1}$. In this case, we have

$$\boldsymbol{\Sigma}_{\mathcal{S}_{k,i}} \triangleq (\mathbf{D}_{\mathcal{S}_{k,i}}^{-1} - \boldsymbol{\Sigma}_X^{-1})^{-1} \succ \mathbf{0},$$
$$k = 1, \cdots, L; \quad i = 1, \cdots, 2^{k-1}.$$

It can be verified that

$$\frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X|}{|\mathbf{D}_{\mathcal{S}_{1,1}}|} = \frac{1}{2} \log |\boldsymbol{\Sigma}_X (\boldsymbol{\Sigma}_{\mathcal{S}_{1,1}}^{-1} + \boldsymbol{\Sigma}_X^{-1})|$$
$$= \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{1,1}}|}{|\boldsymbol{\Sigma}_{\mathcal{S}_{1,1}}|}. \tag{10}$$

Moreover,

$$\frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X|}{|\mathbf{D}_{\mathcal{S}_{k,i}} \boldsymbol{\Sigma}_X^{-1} (\boldsymbol{\Sigma}_X - \boldsymbol{\Theta}_{k,i}) + \boldsymbol{\Theta}_{k,i}|}$$
$$= \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X|}{|(\boldsymbol{\Sigma}_{\mathcal{S}_{k,i}}^{-1} + \boldsymbol{\Sigma}_X^{-1})^{-1} \boldsymbol{\Sigma}_X^{-1} (\boldsymbol{\Sigma}_X - \boldsymbol{\Theta}_{k,i}) + \boldsymbol{\Theta}_{k,i}|}$$
$$= \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X|}{|\boldsymbol{\Sigma}_{\mathcal{S}_{k,i}} (\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{k,i}})^{-1} (\boldsymbol{\Sigma}_X - \boldsymbol{\Theta}_{k,i}) + \boldsymbol{\Theta}_{k,i}|}$$
$$= \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X|}{\left| \begin{array}{c} \boldsymbol{\Sigma}_{\mathcal{S}_{k,i}} (\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{k,i}})^{-1} (\boldsymbol{\Sigma}_X - \boldsymbol{\Theta}_{k,i}) \\ -(\boldsymbol{\Sigma}_X - \boldsymbol{\Theta}_{k,i}) + \boldsymbol{\Sigma}_X \end{array} \right|}$$
$$= \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X|}{|\boldsymbol{\Sigma}_X - \boldsymbol{\Sigma}_X (\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{k,i}})^{-1} (\boldsymbol{\Sigma}_X - \boldsymbol{\Theta}_{k,i})|}$$
$$= \frac{1}{2} \log \frac{|\mathbf{I}_m|}{|\mathbf{I}_m - (\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{k,i}})^{-1} (\boldsymbol{\Sigma}_X - \boldsymbol{\Theta}_{k,i})|}$$
$$= \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{k,i}}|}{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{k,i}} - \boldsymbol{\Sigma}_X + \boldsymbol{\Theta}_{k,i}|}$$
$$= \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{k,i}}|}{|\boldsymbol{\Theta}_{k,i} + \boldsymbol{\Sigma}_{\mathcal{S}_{k,i}}|},$$
$$k = 1, \cdots, L-1; \quad i = 1, \cdots, 2^{k-1}, \tag{11}$$

and, similarly,

$$\frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X|}{\left| \mathbf{D}_{\mathcal{S}_{k+1,2i-1}} \boldsymbol{\Sigma}_X^{-1} (\boldsymbol{\Sigma}_X - \boldsymbol{\Theta}_{k,i}) + \boldsymbol{\Theta}_{k,i} \right|}$$
$$= \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}}|}{|\boldsymbol{\Theta}_{k,i} + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}}|},$$
$$k = 1, \cdots, L-1; \quad i = 1, \cdots, 2^{k-1}, \tag{12}$$
$$\frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X|}{|\mathbf{D}_{\mathcal{S}_{k+1,2i}} \boldsymbol{\Sigma}_X^{-1} (\boldsymbol{\Sigma}_X - \boldsymbol{\Theta}_{k,i}) + \boldsymbol{\Theta}_{k,i}|}$$
$$= \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i}}|}{|\boldsymbol{\Theta}_{k,i} + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i}}|},$$
$$k = 1, \cdots, L-1; \quad i = 1, \cdots, 2^{k-1}. \tag{13}$$

In view of (10)–(13), the maximization problem in (3) can be expressed equivalently as

$$\max_{\boldsymbol{\Theta}} \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{1,1}}|}{|\boldsymbol{\Sigma}_{\mathcal{S}_{1,1}}|} + \sum_{k=1}^{L-1} \sum_{i=1}^{2^{k-1}} \left(\frac{1}{2} \log \frac{|\boldsymbol{\Theta}_{k,i} + \boldsymbol{\Sigma}_{\mathcal{S}_{k,i}}|}{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{k,i}}|} \right.$$
$$\left. + \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}}|}{|\boldsymbol{\Theta}_{k,i} + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}}|} + \frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i}}|}{|\boldsymbol{\Theta}_{k,i} + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i}}|} \right)$$
$$\text{subject to } \boldsymbol{\Sigma}_X \succeq \boldsymbol{\Theta}_{k+1,2i-1}, \boldsymbol{\Theta}_{k+1,2i} \succeq \boldsymbol{\Theta}_{k,i} \succeq \mathbf{0},$$
$$k = 1, \cdots, L-2; \quad i = 1, \cdots, 2^{k-1}. \tag{14}$$

Let $\underline{\Theta}^* \triangleq (\Theta_{k,i}^*)_{k=1,\cdots,L-1;i=1,\cdots,2^{k-1}}$ be an optimal solution of the above maximization problem. The following lemma provides the Karush-Kuhn-Tucker conditions[4] that $\underline{\Theta}^*$ needs to satisfy. The proof can be found in Appendix B.

Lemma 1: There exist $\mathbf{M}_{k,i}^*$, $k = 1, \cdots, L; i = 1, \cdots, 2^{k-1}$, such that

$$(\Theta_{k,i}^* + \Sigma_{\mathcal{S}_{k+1,2i-1}})^{-1} + \mathbf{M}_{k+1,2i-1}^*$$
$$+ (\Theta_{k,i}^* + \Sigma_{\mathcal{S}_{k+1,2i}})^{-1} + \mathbf{M}_{k+1,2i}^*$$
$$= (\Theta_{k,i}^* + \Sigma_{\mathcal{S}_{k,i}})^{-1} + \mathbf{M}_{k,i}^*,$$
$$k = 1, \cdots, L-1; \quad i = 1, \cdots, 2^{k-1}, \quad (15)$$

$$\mathbf{M}_{1,1}^* \Theta_{1,1}^* = \Theta_{1,1}^* \mathbf{M}_{1,1}^* = \mathbf{0}, \quad (16)$$

$$\mathbf{M}_{k+1,2i-1}^* (\Theta_{k+1,2i-1}^* - \Theta_{k,i}^*)$$
$$= (\Theta_{k+1,2i-1}^* - \Theta_{k,i}^*)\mathbf{M}_{k+1,2i-1}^* = \mathbf{0},$$
$$k = 1, \cdots, L-2; \quad i = 1, \cdots, 2^{k-1}, \quad (17)$$

$$\mathbf{M}_{k+1,2i}^* (\Theta_{k+1,2i}^* - \Theta_{k,i}^*)$$
$$= (\Theta_{k+1,2i}^* - \Theta_{k,i}^*)\mathbf{M}_{k+1,2i}^* = \mathbf{0},$$
$$k = 1, \cdots, L-2; \quad i = 1, \cdots, 2^{k-1}, \quad (18)$$

$$\mathbf{M}_{L,2i-1}^* (\Sigma_X - \Theta_{L-1,i}^*)$$
$$= (\Sigma_X - \Theta_{L-1,i}^*)\mathbf{M}_{L,2i-1}^* = \mathbf{0}, \quad i = 1, \cdots, 2^{L-2}, \quad (19)$$

$$\mathbf{M}_{k,i}^* \succeq \mathbf{0}, \quad k = 1, \cdots, L-1; \; i = 1, \cdots, 2^{k-1},$$
$$\mathbf{M}_{L,2i-1}^* \succeq \mathbf{0}, \quad i = 1, \cdots, 2^{L-2},$$
$$\mathbf{M}_{L,2i}^* = \mathbf{0}, \quad i = 1, \cdots, 2^{L-2}.$$

Inspired by the enhancement argument in [27] (see also [11], [28]), we define $\underline{\tilde{\Sigma}} = (\tilde{\Sigma}_{\mathcal{S}_{k,i}})_{k=1,\cdots,L-1;i=1,\cdots,2^{k-1}}$ with

$$\tilde{\Sigma}_{\mathcal{S}_{k,i}} = ((\Theta_{k,i}^* + \Sigma_{\mathcal{S}_{k,i}})^{-1} + \mathbf{M}_{k,i}^*)^{-1} - \Theta_{k,i}^*,$$
$$k = 1, \cdots, L-1; i = 1, \cdots, 2^{k-1}, \quad (20)$$

$$\tilde{\Sigma}_{\mathcal{S}_{L,i}} = ((\Sigma_X + \Sigma_{\mathcal{S}_{L,i}})^{-1} + \mathbf{M}_{L,i}^*)^{-1} - \Sigma_X,$$
$$i = 1, \cdots, 2^{L-1}. \quad (21)$$

The next lemma collects some useful facts about $\underline{\tilde{\Sigma}}$. The proof is relegated to Appendix C.

Lemma 2: $\underline{\tilde{\Sigma}}$ has the following properties:

$$(\Theta_{k,i}^* + \tilde{\Sigma}_{\mathcal{S}_{k+1,2i-1}})^{-1}$$
$$= (\Theta_{k,i}^* + \Sigma_{\mathcal{S}_{k+1,2i-1}})^{-1} + \mathbf{M}_{k+1,2i-1}^*,$$
$$k = 1, \cdots, L-1; \quad i = 1, \cdots, 2^{k-1}, \quad (22)$$

$$(\Theta_{k,i}^* + \tilde{\Sigma}_{\mathcal{S}_{k+1,2i}})^{-1}$$
$$= (\Theta_{k,i}^* + \Sigma_{\mathcal{S}_{k+1,2i}})^{-1} + \mathbf{M}_{k+1,2i}^*,$$
$$k = 1, \cdots, L-1; \quad i = 1, \cdots, 2^{k-1}, \quad (23)$$

$$(\Theta_{k,i}^* + \tilde{\Sigma}_{\mathcal{S}_{k+1,2i-1}})^{-1} + (\Theta_{k,i}^* + \tilde{\Sigma}_{\mathcal{S}_{k+1,2i}})^{-1}$$
$$= (\Theta_{k,i}^* + \tilde{\Sigma}_{\mathcal{S}_{k,i}})^{-1},$$
$$k = 1, \cdots, L-1; \quad i = 1, \cdots, 2^{k-1}, \quad (24)$$

$$\tilde{\Sigma}_{\mathcal{S}_{k+1,2i-1}}, \tilde{\Sigma}_{\mathcal{S}_{k+1,2i}} \succ \tilde{\Sigma}_{\mathcal{S}_{k,i}} \succ \mathbf{0},$$
$$k = 1, \cdots, L-1; \quad i = 1, \cdots, 2^{k-1}, \quad (25)$$

$$\tilde{\Sigma}_{\mathcal{S}_{1,1}}^{-1}(\Theta_{1,1}^* + \tilde{\Sigma}_{\mathcal{S}_{1,1}}) = \Sigma_{\mathcal{S}_{1,1}}^{-1}(\Theta_{1,1}^* + \Sigma_{\mathcal{S}_{1,1}}), \quad (26)$$

$$(\Theta_{k,i}^* + \tilde{\Sigma}_{\mathcal{S}_{k+1,2i-1}})^{-1}(\Theta_{k+1,2i-1}^* + \tilde{\Sigma}_{\mathcal{S}_{k+1,2i-1}})$$
$$= (\Theta_{k,i}^* + \Sigma_{\mathcal{S}_{k+1,2i-1}})^{-1}(\Theta_{k+1,2i-1}^* + \Sigma_{\mathcal{S}_{k+1,2i-1}})$$
$$k = 1, \cdots, L-2; \quad i = 1, \cdots, 2^{k-1}, \quad (27)$$

$$(\Theta_{k,i}^* + \tilde{\Sigma}_{\mathcal{S}_{k+1,2i}})^{-1}(\Theta_{k+1,2i}^* + \tilde{\Sigma}_{\mathcal{S}_{k+1,2i}})$$
$$= (\Theta_{k,i}^* + \Sigma_{\mathcal{S}_{k+1,2i}})^{-1}(\Theta_{k+1,2i}^* + \Sigma_{\mathcal{S}_{k+1,2i}})$$
$$k = 1, \cdots, L-2; \quad i = 1, \cdots, 2^{k-1}, \quad (28)$$

$$(\Theta_{L-1,i}^* + \tilde{\Sigma}_{\mathcal{S}_{L,2i-1}})^{-1}(\Sigma_X + \tilde{\Sigma}_{\mathcal{S}_{L,2i-1}})$$
$$= (\Theta_{L-1,i}^* + \Sigma_{\mathcal{S}_{L,2i-1}})^{-1}(\Sigma_X + \Sigma_{\mathcal{S}_{L,2i-1}}),$$
$$i = 1, \cdots, 2^{L-2}, \quad (29)$$

$$(\Theta_{L-1,i}^* + \tilde{\Sigma}_{\mathcal{S}_{L,2i}})^{-1}(\Sigma_X + \tilde{\Sigma}_{\mathcal{S}_{L,2i}})$$
$$= (\Theta_{L-1,i}^* + \Sigma_{\mathcal{S}_{L,2i}})^{-1}(\Sigma_X + \Sigma_{\mathcal{S}_{L,2i}}),$$
$$i = 1, \cdots, 2^{L-2}. \quad (30)$$

Lemma 3: We have

$$R_G^*(\mathbf{D})$$
$$= \frac{1}{2}\log\frac{|\Sigma_X + \tilde{\Sigma}_{\mathcal{S}_{1,1}}|}{|\tilde{\Sigma}_{\mathcal{S}_{1,1}}|} + \sum_{k=1}^{L-1}\sum_{i=1}^{2^{k-1}}\left(\frac{1}{2}\log\frac{|\Theta_{k,i}^* + \tilde{\Sigma}_{\mathcal{S}_{k,i}}|}{|\Sigma_X + \tilde{\Sigma}_{\mathcal{S}_{k,i}}|}\right.$$
$$\left. + \frac{1}{2}\log\frac{|\Sigma_X + \tilde{\Sigma}_{\mathcal{S}_{k+1,2i-1}}|}{|\Theta_{k,i}^* + \tilde{\Sigma}_{\mathcal{S}_{k+1,2i-1}}|} + \frac{1}{2}\log\frac{|\Sigma_X + \tilde{\Sigma}_{\mathcal{S}_{k+1,2i}}|}{|\Theta_{k,i}^* + \tilde{\Sigma}_{\mathcal{S}_{k+1,2i}}|}\right).$$

Proof: See Appendix D. ∎

C. The Achievability Part

Let $U_j = X + Q_{L,j}$, $j = 1, \cdots, M$, where $Q_{L,1}, \cdots, Q_{L,M}$ are jointly Gaussian and independent of X. A straightforward extension of the El Gamal-Cover scheme (see, e.g., [12]) shows[5] that

$$R_G(\underline{\mathbf{D}}) \leq -h(U_1, \cdots, U_M|X) + \sum_{j=1}^M h(U_j)$$

if $\text{cov}(X|(U_j)_{j\in\mathcal{S}_{k,i}}) \preceq \mathbf{D}_{\mathcal{S}_{k,i}}$, $k = 1, \cdots, L; i = 1, \cdots, 2^{k-1}$.

Now we shall leverage $\underline{\Theta}^*$ and $\underline{\tilde{\Sigma}}$ to construct the covariance matrix of $(Q_1, \cdots, Q_M)$ such that the resulting scheme satisfies the distortion constraints and attains the minimum sum rate. Define

$$\Lambda_{k,i} = \begin{pmatrix} \tilde{\Sigma}_{\mathcal{S}_{k+1,2i-1}} - \tilde{\Sigma}_{\mathcal{S}_{k,i}} & -\Theta_{k,i}^* - \tilde{\Sigma}_{\mathcal{S}_{k,i}} \\ -\Theta_{k,i}^* - \tilde{\Sigma}_{\mathcal{S}_{k,i}} & \tilde{\Sigma}_{\mathcal{S}_{k+1,2i}} - \tilde{\Sigma}_{\mathcal{S}_{k,i}} \end{pmatrix},$$
$$\Gamma_{k,i} = \begin{pmatrix} \tilde{\Sigma}_{\mathcal{S}_{k+1,2i-1}} & -\Theta_{k,i}^* \\ -\Theta_{k,i}^* & \tilde{\Sigma}_{\mathcal{S}_{k+1,2i}} \end{pmatrix}$$

for $k = 1, \cdots, L-1; i = 1, \cdots, 2^{k-1}$.

Lemma 4: 1) For $k = 1, \cdots, L-1; i = 1, \cdots, 2^{k-1}$,

$$\Lambda_{k,i} \succeq \mathbf{0}. \quad (31)$$

2) For $k = 1, \cdots, L-1; i = 1, \cdots, 2^{k-1}$, there exist $\mathbf{H}_{k+1,2i-1}$ and $\mathbf{H}_{k+1,2i}$ such that

$$\mathbf{H}_{k+1,2i-1} + \mathbf{H}_{k+1,2i} = \mathbf{I}_m, \quad (32)$$

$$(\mathbf{H}_{k+1,2i-1}, \mathbf{H}_{k+1,2i})\Lambda_{k,i}(\mathbf{H}_{k+1,2i-1}, \mathbf{H}_{k+1,2i})^T = \mathbf{0}. \quad (33)$$

Proof: See Appendix E. ∎

---

[4]In fact, it can be shown that these Karush-Kuhn-Tucker conditions are sufficient to guarantee the optimality of $\underline{\Theta}^*$.

[5]Roughly speaking, U_j corresponds to the j-th description, $j = 1, \cdots, M$; moreover, $-h(U_1, \cdots, U_M|X) + \sum_{j=1}^M h(U_j)$ is the sum rate, and $\text{cov}(X|(U_j)_{j\in\mathcal{S}_{k,i}})$ is the reconstruction covariance distortion based on the subset of descriptions specified by $\mathcal{S}_{k,i}$, $k = 1, \cdots, L; i = 1, \cdots, 2^{k-1}$.

Let $\underline{V} \triangleq (V_{k,i})_{k=1,\cdots,L;i=1,\cdots,2^{k-1}}$ be jointly distributed with X such that

$$p(x,\underline{v}) = p(x)p(v_{1,1})\prod_{k=1}^{L-1}\prod_{i=1}^{2^{k-1}} p(v_{k+1,2i-1}, v_{k+1,2i}).$$

Moreover, we assume that $V_{k,i}$, $k = 1,\cdots,L; i = 1,\cdots,2^{k-1}$, are $m \times 1$ zero-mean jointly Gaussian random vectors with[6]

$$\mathrm{cov}(V_{1,1}) = \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{1,1}},$$
$$\mathrm{cov}(V_{k+1,2i-1}, V_{k+1,2i}) = \boldsymbol{\Lambda}_{k,i}, \quad k = 1,\cdots,L-1;$$
$$i = 1,\cdots,2^{k-1}.$$

In this way, the joint distribution of X and $\underline{V}$ is completely specified. Set

$$Q_{1,1} = V_{1,1},$$
$$Q_{k+1,2i-1} = Q_{k,i} + V_{k+1,2i-1},$$
$$k = 1,\cdots,L-1; \quad i = 1,\cdots,2^{k-1},$$
$$Q_{k+1,2i} = Q_{k,i} + V_{k+1,2i},$$
$$k = 1,\cdots,L-1; \quad i = 1,\cdots,2^{k-1}.$$

It is easy to verify that

$$\mathrm{cov}(Q_{k+1,2i-1}, Q_{k+1,2i}) = \boldsymbol{\Gamma}_{k,i},$$
$$k = 1,\cdots,L-1; \quad i = 1,\cdots,2^{k-1}. \quad (34)$$

Note that the constructed $(U_1,\cdots,U_M)$ satisfies the covariance distortion constraints. Indeed,

$$\mathrm{cov}(X|(U_j)_{j\in\mathcal{S}_{k,i}}) = \mathrm{cov}(X|X + Q_{k,i}) \quad (35)$$
$$= (\boldsymbol{\Sigma}_X^{-1} + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k,i}}^{-1})^{-1}$$
$$\preceq (\boldsymbol{\Sigma}_X^{-1} + \boldsymbol{\Sigma}_{\mathcal{S}_{k,i}}^{-1})^{-1}$$
$$= \mathbf{D}_{\mathcal{S}_{k,i}},$$
$$k = 1,\cdots,L; \quad i = 1,\cdots,2^{k-1}, \quad (36)$$

where (35) is due to (32) and (33). Now it remains to show that the resulting sum rate coincides with $R_G^*(\mathbf{D})$. We have

$$-h(U_1,\cdots,U_M|X) + \sum_{j=1}^M h(U_j)$$
$$= I(X;(U_j)_{j\in\mathcal{S}_{1,1}})$$
$$+ \sum_{k=1}^{L-1}\sum_{i=1}^{2^{k-1}} I((U_j)_{j\in\mathcal{S}_{k+1,2i-1}}; (U_j)_{j\in\mathcal{S}_{k+1,2i}}) \quad (37)$$
$$= I(X; X + Q_{1,1})$$
$$+ \sum_{k=1}^{L-1}\sum_{i=1}^{2^{k-1}} I(X + Q_{k+1,2i-1}; X + Q_{k+1,2i}) \quad (38)$$
$$= \frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_X + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{1,1}}|}{|\tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{1,1}}|}$$
$$+ \sum_{k=1}^{L-1}\sum_{i=1}^{2^{k-1}} I(X + Q_{k+1,2i-1}; X + Q_{k+1,2i}), \quad (39)$$

[6]Note that $\tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{1,1}}$ and $\boldsymbol{\Lambda}_{k,i}$, $k = 1,\cdots,L-1; i = 1,\cdots,2^{k-1}$, are valid covariance matrices as ensured by (25) and (31).

where (37) follows by successively applying the fact that

$$\sum_{i=1}^{2^k} h((U_j)_{j\in\mathcal{S}_{k+1,i}}) = \sum_{i=1}^{2^{k-1}} h((U_j)_{j\in\mathcal{S}_{k,i}})$$
$$+ \sum_{i=1}^{2^{k-1}} I((U_j)_{j\in\mathcal{S}_{k+1,2i-1}}; (U_j)_{j\in\mathcal{S}_{k+1,2i}})$$

from $k = L-1$ to $k = 1$, and (38) is due to (32) and (33). Write $X = \tilde{X}_{k,i} + \hat{X}_{k,i}$, where $\tilde{X}_{k,i}$ and $\hat{X}_{k,i}$ are two $m \times 1$ zero-mean Gaussian random vectors with covariance matrices $\boldsymbol{\Sigma}_X - \boldsymbol{\Theta}_{k,i}^*$ and $\boldsymbol{\Theta}_{k,i}^*$, respectively, $k = 1,\cdots,L-1; i = 1,\cdots,2^{k-1}$; moreover, $\tilde{X}_{k,i}$, $\hat{X}_{k,i}$, and $(Q_{k+1,2i-1}, Q_{k+1,2i})$ are mutually independent, $k = 1,\cdots,L-1; i = 1,\cdots,2^{k-1}$. In view of (34),

$$\mathbb{E}[(\hat{X}_{k,i} + Q_{k+1,2i-1})(\hat{X}_{k,i} + Q_{k+1,2i})^T] = \mathbf{0},$$
$$k = 1,\cdots,L-1; \quad i = 1,\cdots,2^{k-1},$$

which implies

$$I(\hat{X}_{k,i} + Q_{k+1,2i-1}; \hat{X}_{k,i} + Q_{k+1,2i}) = 0,$$
$$k = 1,\cdots,L-1; \quad i = 1,\cdots,2^{k-1}. \quad (40)$$

It can be verified that

$$I(X + Q_{k+1,2i-1}; X + Q_{k+1,2i})$$
$$= I(\tilde{X}_{k,i}; X + Q_{k+1,2i-1}) + I(\tilde{X}_{k,i}; X + Q_{k+1,2i})$$
$$- I(\tilde{X}_{k,i}; X + Q_{k+1,2i-1}, X + Q_{k+1,2i-1})$$
$$+ I(X + Q_{k+1,2i-1}; X + Q_{k+1,2i}|\tilde{X}_{k,i})$$
$$= I(\tilde{X}_{k,i}; X + Q_{k+1,2i-1}) + I(\tilde{X}_{k,i}; X + Q_{k+1,2i})$$
$$- I(\tilde{X}_{k,i}; X + Q_{k,i})$$
$$+ I(X + Q_{k+1,2i-1}; X + Q_{k+1,2i}|\tilde{X}_{k,i}) \quad (41)$$
$$= I(\tilde{X}_{k,i}; X + Q_{k+1,2i-1}) + I(\tilde{X}_{k,i}; X + Q_{k+1,2i})$$
$$- I(\tilde{X}_{k,i}; X + Q_{k,i})$$
$$+ I(\hat{X}_{k,i} + Q_{k+1,2i-1}; \hat{X}_{k,i} + Q_{k+1,2i})$$
$$= I(\tilde{X}_{k,i}; X + Q_{k+1,2i-1}) + I(\tilde{X}_{k,i}; X + Q_{k+1,2i})$$
$$- I(\tilde{X}_{k,i}; X + Q_{k,i}) \quad (42)$$
$$= \frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_X + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k+1,2i-1}}|}{|\boldsymbol{\Theta}_{k,i}^* + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k+1,2i-1}}|} + \frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_X + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k+1,2i}}|}{|\boldsymbol{\Theta}_{k,i}^* + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k+1,2i}}|}$$
$$- \frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_X + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k,i}}|}{|\boldsymbol{\Theta}_{k,i}^* + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k,i}}|}$$
$$k = 1,\cdots,L-1; \quad i = 1,\cdots,2^{k-1}, \quad (43)$$

where (41) is due to (32) and (33) while (42) is due to (40). Substituting (43) into (39) and invoking Lemma 3 shows

$$R_G^*(\mathbf{D}) = -h(U_1,\cdots,U_M|X) + \sum_{j=1}^M h(U_j)$$

as desired.

Finally consider the general case where $\mathbf{0} \prec \mathbf{D}_{\mathcal{S}_{k,i}} \preceq \boldsymbol{\Sigma}_X$, $k = 1,\cdots,L; i = 1,\cdots,2^{k-1}$. Define $\underline{\mathbf{D}}(\epsilon) = (\mathbf{D}_{\mathcal{S}_{k,i}} - \epsilon\mathbf{I}_m)_{k=1,\cdots,L;i=1,\cdots,2^{k-1}}$. The preceding argument implies

$$\liminf_{\epsilon\downarrow 0} R_G(\underline{\mathbf{D}}(\epsilon)) = \liminf_{\epsilon\downarrow 0} R_G^*(\underline{\mathbf{D}}(\epsilon)). \quad (44)$$

Since $R_G(\underline{\mathbf{D}}) \leq R_G(\underline{\mathbf{D}}(\epsilon))$ for any $\epsilon > 0$, we have

$$R_G(\underline{\mathbf{D}}) \leq \liminf_{\epsilon \downarrow 0} R_G(\underline{\mathbf{D}}(\epsilon)). \tag{45}$$

Moreover, it follows by the continuity of $R_G(\underline{\mathbf{D}}(\epsilon))$ at $\epsilon = 0$ that

$$\liminf_{\epsilon \downarrow 0} R_G^*(\underline{\mathbf{D}}(\epsilon)) = R_G^*(\underline{\mathbf{D}}). \tag{46}$$

Combining (44)–(46) completes the proof.

Remark: Now we are in a position to make a comparison with the converse argument in Appendix A. Roughly speaking, U_j corresponds to $\varphi_j^n(X^n)$, $j = 1, \cdots, M$, and $\tilde{X}_{k,i}$ can be viewed as the single-letter version of $\mathbb{E}[X^n | Z_{k,i}^n]$ (assuming X^n and $Z_{k,i}^n$ are jointly Gaussian), $k = 1, \cdots, L-1; i = 1, \cdots, 2^{k-1}$. In this converse argument, we obtain a lower bound by dropping the terms $I((\varphi_j^n(X^n))_{j \in \mathcal{S}_{k+1,2i-1}}; (\varphi_j^{(n)}(X^n))_{j \in \mathcal{S}_{k+1,2i}} | Z_{k,i}^n)$, $k = 1, \cdots, L-1; i = 1, \cdots, 2^{k-1}$ (see (50)). For this lower bound to be attainable, one must ensure that the corresponding terms $I((U_j)_{j \in \mathcal{S}_{k+1,2i-1}}; (U_j)_{j \in \mathcal{S}_{k+1,2i}} | \tilde{X}_{k,i})$, $k = 1, \cdots, L-1; i = 1, \cdots, 2^{k-1}$, in the achievability part can be made equal to zero. Although this is not always possible for a generic version of the extended El Gamal-Cover scheme, as we have shown (see (40)), the specific construction based on $\underline{\mathbf{\Theta}}^*$ and $\tilde{\mathbf{\Sigma}}$ indeed possesses the desired conditional independence structures, which can be effectively exploited by Gauss-Markov tree-structured auxiliary random variables to establish a matching converse.

IV. CONCLUSION

We have generalized Ozarow's converse argument to establish a single-letter lower bound on the sum rate of multiple description coding with tree-structured distortion constraints. Moreover, this lower bound is shown to be tight for the quadratic vector Gaussian case. It is worth mentioning that the applicability of Ozarow's method is by no means confined to the setting of tree-structured distortion constraints considered in the present work. Indeed, this method has been successfully used to obtain conclusive results for the multiple description problem with symmetrical distortion constraints [23], [28]. Here it is curious to observe that the projection of the entropy region to the coordinates for which the corresponding subsets form a tree structure and the projected entropy region induced by the symmetric group [29] are both completely characterized by the Shannon type inequalities. This suggests that the difficulty in solving the general M-description problem is potentially related to the lack of explicit characterization of the entropy region, and it might be more fruitful to focus on those special formulations of the multiple description problem for which the corresponding projected entropy regions admit simple characterizations.

APPENDIX A
PROOF OF THEOREM 1

Consider arbitrary encoding functions

$$\varphi_j^{(n)} : \mathcal{X}^n \to \mathcal{C}_j^{(n)}, \quad j = 1, \cdots, 2^{L-1},$$

and decoding functions

$$\psi_{\mathcal{S}_{k,i}}^{(n)} : \prod_{j \in \mathcal{S}_{k,i}} \mathcal{C}_j^{(n)} \to \hat{\mathcal{X}}^n, \quad k = 1, \cdots, L; \ i = 1, \cdots, 2^{k-1},$$

satisfying

$$\frac{1}{n} \sum_{t=1}^n \mathbb{E}[d(X(t), \hat{X}_{\mathcal{S}_{k,i}}(t))] \leq d_{\mathcal{S}_{k,i}},$$

$$k = 1, \cdots, L; \quad i = 1, \cdots, 2^{k-1}.$$

Note that

$$\sum_{j=1}^{2^{L-1}} \log |\mathcal{C}_j^{(n)}| \geq \sum_{j=1}^{2^{L-1}} H(\varphi_j^{(n)}(X^n))$$

$$= \sum_{i=1}^{2^{L-1}} H((\varphi_j^{(n)}(X^n))_{j \in \mathcal{S}_{L,i}}). \tag{47}$$

Successively applying the fact that

$$\sum_{i=1}^{2^k} H((\varphi_j^{(n)}(X^n))_{j \in \mathcal{S}_{k+1,i}})$$

$$= \sum_{i=1}^{2^{k-1}} H((\varphi_j^{(n)}(X^n))_{j \in \mathcal{S}_{k,i}})$$

$$+ \sum_{i=1}^{2^{k-1}} I((\varphi_j^{(n)}(X^n))_{j \in \mathcal{S}_{k+1,2i-1}}; (\varphi_j^{(n)}(X^n))_{j \in \mathcal{S}_{k+1,2i}})$$

from $k = L-1$ to $k = 1$ shows

$$\sum_{i=1}^{2^{L-1}} H((\varphi_j^{(n)}(X^n))_{j \in \mathcal{S}_{L,i}})$$

$$= H((\varphi_j^{(n)}(X^n))_{j \in \mathcal{S}_{1,1}})$$

$$+ \sum_{k=1}^{L-1} \sum_{i=1}^{2^{k-1}} I((\varphi_j^{(n)}(X^n))_{j \in \mathcal{S}_{k+1,2i-1}}; (\varphi_j^{(n)}(X^n))_{j \in \mathcal{S}_{k+1,2i}}). \tag{48}$$

Substituting (48) into (47) and invoking the fact that $H((\varphi_j^{(n)}(X^n))_{j \in \mathcal{S}_{1,1}}) = I(X^n; (\varphi_j^{(n)}(X^n))_{j \in \mathcal{S}_{1,1}})$ gives

$$\sum_{j=1}^{2^{L-1}} \log |\mathcal{C}_j^{(n)}|$$

$$\geq I(X^n; (\varphi_j^{(n)}(X^n))_{j \in \mathcal{S}_{1,1}})$$

$$+ \sum_{k=1}^{L-1} \sum_{i=1}^{2^{k-1}} I((\varphi_j^{(n)}(X^n))_{j \in \mathcal{S}_{k+1,2i-1}}; (\varphi_j^{(n)}(X^n))_{j \in \mathcal{S}_{k+1,2i}}). \tag{49}$$

Now introduce an auxiliary remote vector source $\{(Z_{k,i}(t))_{k=1,\cdots,L-1;i=1,\cdots,2^{k-1}}\}_{t=1}^\infty$, which forms a joint i.i.d. process with $\{X(t)\}_{t=1}^\infty$. We assume that the conditional distribution of $(Z_{k,i}(t))_{k=1,\cdots,L-1;i=1,\cdots,2^{k-1}}$ given $X(t)$

satisfies the binary Markov tree condition (2) for every t. It can be verified that

$$I((\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{k+1,2i-1}}; (\varphi_j^n(X^n))_{j\in\mathcal{S}_{k+1,2i}})$$
$$= I(Z_{k,i}^n; (\varphi_j^n(X^n))_{j\in\mathcal{S}_{k+1,2i-1}})$$
$$\quad + I(Z_{k,i}^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{k+1,2i}})$$
$$\quad - I(Z_{k,i}^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{k,i}})$$
$$\quad + I((\varphi_j^n(X^n))_{j\in\mathcal{S}_{k+1,2i-1}}; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{k+1,2i}}|Z_{k,i}^n)$$
$$\geq I(Z_{k,i}^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{k+1,2i-1}})$$
$$\quad + I(Z_{k,i}^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{k+1,2i}})$$
$$\quad - I(Z_{k,i}^n; (\varphi_j^n(X^n))_{j\in\mathcal{S}_{k,i}}),$$
$$k = 1, \cdots, L-1; \quad i = 1, \cdots, 2^{k-1}. \quad (50)$$

We have

$$\sum_{k=1}^{L-1}\sum_{i=1}^{2^{k-1}} I((\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{k+1,2i-1}}; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{k+1,2i}})$$
$$\geq \sum_{k=1}^{L-1}\sum_{i=1}^{2^{k-1}} \Big(I(Z_{k,i}^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{k+1,2i-1}})$$
$$\qquad\qquad + I(Z_{k,i}^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{k+1,2i}})\Big)$$
$$\quad - \sum_{k=1}^{L-1}\sum_{i=1}^{2^{k-1}} I(Z_{k,i}^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{k,i}})$$
$$= -I(Z_{1,1}^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{1,1}})$$
$$\quad + \sum_{k=1}^{L-1}\sum_{i=1}^{2^{k-1}} \Big(I(Z_{k,i}^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{k+1,2i-1}})$$
$$\qquad\qquad + I(Z_{k,i}^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{k+1,2i}})\Big)$$
$$\quad - \sum_{k=1}^{L-2}\sum_{i=1}^{2^{k-1}} \Big(I(Z_{k+1,2i-1}^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{k+1,2i-1}})$$
$$\qquad\qquad + I(Z_{k+1,2i}^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{k+1,2i}})\Big)$$
$$= -I(Z_{1,1}^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{1,1}})$$
$$\quad + \sum_{k=1}^{L-2}\sum_{i=1}^{2^{k-1}} \Big(I(Z_{k,i}^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{k+1,2i-1}})$$
$$\qquad\qquad - I(Z_{k+1,2i-1}^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{k+1,2i-1}})$$
$$\qquad\qquad + I(Z_{k,i}^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{k+1,2i}})$$
$$\qquad\qquad - I(Z_{k+1,2i}^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{k+1,2i}})\Big)$$
$$\quad + \sum_{i=1}^{2^{L-2}} \Big(I(Z_{L-1,i}^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{L,2i-1}})$$
$$\qquad\qquad + I(Z_{L-1,i}^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{L,2i}})\Big). \quad (51)$$

Substituting (51) into (49) gives

$$\sum_{j=1}^{2^{L-1}} \log|\mathcal{C}_j^{(n)}|$$
$$\geq I(X^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{1,1}}) - I(Z_{1,1}^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{1,1}})$$
$$\quad + \sum_{k=1}^{L-2}\sum_{i=1}^{2^{k-1}} \Big(I(Z_{k,i}^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{k+1,2i-1}})$$
$$\qquad\qquad - I(Z_{k+1,2i-1}^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{k+1,2i-1}})$$
$$\qquad\qquad + I(Z_{k,i}^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{k+1,2i}})$$
$$\qquad\qquad - I(Z_{k+1,2i}^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{k+1,2i}})\Big)$$
$$\quad + \sum_{i=1}^{2^{L-2}} \Big(I(Z_{L-1,i}^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{L,2i-1}})$$
$$\qquad\qquad + I(Z_{L-1,i}^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{L,2i}})\Big). \quad (52)$$

Let T be a random variable uniformly distributed over $\{1, \cdots, n\}$ and independent of everything else. We have

$$I(X^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{1,1}}) - I(Z_{1,1}^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{1,1}})$$
$$= I(Z_{1,1}^n, X^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{1,1}}) - I(Z_{1,1}^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{1,1}})$$
$$= I(X^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{1,1}}|Z_{1,1}^n)$$
$$\geq I(X^n; \hat{X}_{\mathcal{S}_{1,1}}^n|Z_{1,1}^n)$$
$$= \sum_{t=1}^n I(X(t); \hat{X}_{\mathcal{S}_{1,1}}^n|X(1:t-1), Z_{1,1}^n)$$
$$= \sum_{t=1}^n I(X(t); \hat{X}_{\mathcal{S}_{1,1}}^n, X(1:t-1),$$
$$\qquad\qquad Z_{1,1}(1:t-1), Z_{1,1}(t+1:n)|Z_{1,1}(t))$$
$$\geq \sum_{t=1}^n I(X(t); \hat{X}_{\mathcal{S}_{1,1}}(t)|Z_{1,1}(t))$$
$$= nI(X(T); \hat{X}_{\mathcal{S}_{1,1}}(T)|Z_{1,1}(T), T)$$
$$= nI(X(T); \hat{X}_{\mathcal{S}_{1,1}}(T), T|Z_{1,1}(T))$$
$$\geq nI(X(T); \hat{X}_{\mathcal{S}_{1,1}}(T)|Z_{1,1}(T)). \quad (53)$$

Similarly, it can be shown that

$$I(Z_{k,i}^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{k+1,2i-1}})$$
$$\quad - I(Z_{k+1,2i-1}^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{k+1,2i-1}})$$
$$\geq nI(Z_{k,i}(T); \hat{X}_{\mathcal{S}_{k+1,2i-1}}(T)|Z_{k+1,2i-1}(T)),$$
$$k = 1, \cdots, L-2; \quad i = 1, \cdots, 2^{k-1}, \quad (54)$$
$$I(Z_{k,i}^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{k+1,2i}})$$
$$\quad - I(Z_{k+1,2i}^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{k+1,2i}})$$
$$\geq nI(Z_{k,i}(T); \hat{X}_{\mathcal{S}_{k+1,2i}}(T)|Z_{k+1,2i}(T)),$$
$$k = 1, \cdots, L-2; \quad i = 1, \cdots, 2^{k-1}, \quad (55)$$
$$I(Z_{L-1,i}^n; (\varphi_j^{(n)}(X^n))_{j\in\mathcal{S}_{L,2i-1}})$$
$$\geq nI(Z_{L-1,i}(T); \hat{X}_{\mathcal{S}_{L,2i-1}}(T)), \quad i = 1, \cdots, 2^{L-2}, \quad (56)$$

$$I(Z_{L-1,i}^n; (\varphi_j^{(n)}(X^n))_{j \in \mathcal{S}_{L,2i}})$$
$$\geq nI(Z_{L-1,i}(T); \hat{X}_{\mathcal{S}_{L,2i}}(T)), \qquad i = 1, \cdots, 2^{L-2}. \quad (57)$$

Substituting (53)–(57) into (52) gives

$$\sum_{j=1}^{2^{L-1}} \log |\mathcal{C}_j^{(n)}|$$
$$\geq nI(X(T); \hat{X}_{\mathcal{S}_{1,1}}(T)|Z_{1,1}(T))$$
$$+ \sum_{k=1}^{L-2} \sum_{i=1}^{2^{k-1}} \Big(nI(Z_{k,i}(T); \hat{X}_{\mathcal{S}_{k+1,2i-1}}(T)|Z_{k+1,2i-1}(T))$$
$$+ nI(Z_{k,i}(T); \hat{X}_{\mathcal{S}_{k+1,2i}}(T)|Z_{k+1,2i}(T)) \Big)$$
$$+ \sum_{i=1}^{2^{L-2}} \Big(nI(Z_{L-1,i}(T); \hat{X}_{\mathcal{S}_{L,2i-1}}(T))$$
$$+ nI(Z_{L-1,i}(T); \hat{X}_{\mathcal{S}_{L,2i}}(T)) \Big).$$

Note that $(Z_{k,i}(T))_{k=1,\cdots,L-1;i=1,\cdots,2^{k-1}} \leftrightarrow X(T) \leftrightarrow (\hat{X}_{\mathcal{S}_{k,i}}(T))_{k=1,\cdots,L;i=1,\cdots,2^{k-1}}$ form a Markov chain; moreover, the conditional distribution of $(Z_{k,i}(T))_{k=1,\cdots,L-1;i=1,\cdots,2^{k-1}}$ given $X(T)$ is in $\mathcal{P}$ while the conditional distribution of $(\hat{X}_{\mathcal{S}_{k,i}}(T))_{k=1,\cdots,L;i=1,\cdots,2^{k-1}}$ given $X(T)$ is in $\mathcal{P}(\underline{d})$. This completes the proof of Theorem 1.

APPENDIX B
PROOF OF LEMMA 1

The Lagrangian of the maximization problem in (14) is given by

$$\mathcal{L} = \sum_{k=1}^{L-1} \sum_{i=1}^{2^{k-1}} \big(\log |\mathbf{\Theta}_{k,i} + \mathbf{\Sigma}_{\mathcal{S}_{k,i}}| - \log |\mathbf{\Theta}_{k,i} + \mathbf{\Sigma}_{\mathcal{S}_{k+1,2i-1}}|$$
$$- \log |\mathbf{\Theta}_{k,i} + \mathbf{\Sigma}_{\mathcal{S}_{k+1,2i}}| \big)$$
$$+ \text{tr}(\mathbf{\Theta}_{1,1}\mathbf{M}_{1,1})$$
$$+ \sum_{k=1}^{L-2} \sum_{i=1}^{2^{k-1}} \big(\text{tr}((\mathbf{\Theta}_{k+1,2i-1} - \mathbf{\Theta}_{k,i})\mathbf{M}_{k+1,2i-1})$$
$$+ \text{tr}((\mathbf{\Theta}_{k+1,2i} - \mathbf{\Theta}_{k,i})\mathbf{M}_{k+1,2i}) \big)$$
$$+ \sum_{i=1}^{2^{L-2}} \text{tr}((\mathbf{\Sigma}_X - \mathbf{\Theta}_{L-1,i})\mathbf{M}_{L,2i-1}),$$

where $\mathbf{M}_{k,i}$, $k = 1, \cdots, L$; $i = 1, \cdots, 2^{k-1}$, are the Lagrange multipliers[7] satisfying

$$\mathbf{M}_{k,i} \succeq \mathbf{0}, \quad k = 1, \cdots, L-1; \ i = 1, \cdots, 2^{k-1},$$
$$\mathbf{M}_{L,2i-1} \succeq \mathbf{0}, \quad i = 1, \cdots, 2^{L-2},$$
$$\mathbf{M}_{L,2i} = \mathbf{0}, \quad i = 1, \cdots, 2^{L-2}.$$

It can be shown by leveraging [30, p. 327, Proposition 3.3.11] that, for any optimal solution $\underline{\mathbf{\Theta}}^*$, there exists $\underline{\mathbf{M}}^* \triangleq$

[7] Note that $\mathbf{M}_{L,2i}$, $i = 1, \cdots, 2^{L-2}$, do not appear in the Lagrangian. They are introduced to simplify notation, and are set to zero.

$(\mathbf{M}_{k,i}^*)_{k=1,\cdots,L;i=1,\cdots,2^{k-1}}$, such that

$$\nabla_{\mathbf{\Theta}_{k,i}} \mathcal{L}\big|_{\underline{\mathbf{\Theta}}=\underline{\mathbf{\Theta}}^*, \underline{\mathbf{M}}=\underline{\mathbf{M}}^*} = \mathbf{0},$$
$$k = 1, \cdots, L-1; \quad i = 1, \cdots, 2^{k-1}, \quad (58)$$
$$\mathbf{M}_{1,1}^* \mathbf{\Theta}_{1,1}^* = \mathbf{\Theta}_{1,1}^* \mathbf{M}_{1,1}^* = \mathbf{0},$$
$$\mathbf{M}_{k+1,2i-1}^* (\mathbf{\Theta}_{k+1,2i-1}^* - \mathbf{\Theta}_{k,i}^*)$$
$$= (\mathbf{\Theta}_{k+1,2i-1}^* - \mathbf{\Theta}_{k,i}^*)\mathbf{M}_{k+1,2i-1}^* = \mathbf{0},$$
$$k = 1, \cdots, L-2; \quad i = 1, \cdots, 2^{k-1},$$
$$\mathbf{M}_{k+1,2i}^* (\mathbf{\Theta}_{k+1,2i}^* - \mathbf{\Theta}_{k,i}^*)$$
$$= (\mathbf{\Theta}_{k+1,2i}^* - \mathbf{\Theta}_{k,i}^*)\mathbf{M}_{k+1,2i}^* = \mathbf{0},$$
$$k = 1, \cdots, L-2; \quad i = 1, \cdots, 2^{k-1},$$
$$\mathbf{M}_{L,2i-1}^* (\mathbf{\Sigma}_X - \mathbf{\Theta}_{L-1,i}^*)$$
$$= (\mathbf{\Sigma}_X - \mathbf{\Theta}_{L-1,i}^*)\mathbf{M}_{L,2i-1}^* = \mathbf{0}, \quad i = 1, \cdots, 2^{L-2},$$
$$\mathbf{M}_{k,i}^* \succeq \mathbf{0}, \quad k = 1, \cdots, L-1; \quad i = 1, \cdots, 2^{k-1},$$
$$\mathbf{M}_{L,2i-1}^* \succeq \mathbf{0}, \quad i = 1, \cdots, 2^{L-2},$$
$$\mathbf{M}_{L,2i}^* = \mathbf{0}, \quad i = 1, \cdots, 2^{L-2}.$$

Note that

$$\mathcal{L} = \sum_{k=1}^{L-1} \sum_{i=1}^{2^{k-1}} \big(\log |\mathbf{\Theta}_{k,i} + \mathbf{\Sigma}_{\mathcal{S}_{k,i}}| - \log |\mathbf{\Theta}_{k,i} + \mathbf{\Sigma}_{\mathcal{S}_{k+1,2i-1}}|$$
$$- \log |\mathbf{\Theta}_{k,i} + \mathbf{\Sigma}_{\mathcal{S}_{k+1,2i}}| \big)$$
$$+ \text{tr}(\mathbf{\Theta}_{1,1}\mathbf{M}_{1,1})$$
$$+ \sum_{k=1}^{L-2} \sum_{i=1}^{2^{k-1}} \big(\text{tr}(\mathbf{\Theta}_{k+1,2i-1}\mathbf{M}_{k+1,2i-1})$$
$$+ \text{tr}(\mathbf{\Theta}_{k+1,2i}\mathbf{M}_{k+1,2i}) \big)$$
$$- \sum_{k=1}^{L-2} \sum_{i=1}^{2^{k-1}} \text{tr}(\mathbf{\Theta}_{k,i}(\mathbf{M}_{k+1,2i-1} + \mathbf{M}_{k+1,2i}))$$
$$- \sum_{i=1}^{2^{L-2}} \text{tr}(\mathbf{\Theta}_{L-1,i}\mathbf{M}_{L,2i-1}) + \sum_{i=1}^{2^{L-2}} \text{tr}(\mathbf{\Sigma}_X \mathbf{M}_{L,2i-1})$$
$$= \sum_{k=1}^{L-1} \sum_{i=1}^{2^{k-1}} \big(\log |\mathbf{\Theta}_{k,i} + \mathbf{\Sigma}_{\mathcal{S}_{k,i}}| - \log |\mathbf{\Theta}_{k,i} + \mathbf{\Sigma}_{\mathcal{S}_{k+1,2i-1}}|$$
$$- \log |\mathbf{\Theta}_{k,i} + \mathbf{\Sigma}_{\mathcal{S}_{k+1,2i}}| \big)$$
$$+ \sum_{k=1}^{L-1} \sum_{i=1}^{2^{k-1}} \text{tr}(\mathbf{\Theta}_{k,i}\mathbf{M}_{k,i})$$
$$- \sum_{k=1}^{L-1} \sum_{i=1}^{2^{k-1}} \text{tr}(\mathbf{\Theta}_{k,i}(\mathbf{M}_{k+1,2i-1} + \mathbf{M}_{k+1,2i}))$$
$$+ \sum_{i=1}^{2^{L-2}} \text{tr}(\mathbf{\Sigma}_X \mathbf{M}_{L,2i-1})$$
$$= \sum_{k=1}^{L-1} \sum_{i=1}^{2^{k-1}} \big(\log |\mathbf{\Theta}_{k,i} + \mathbf{\Sigma}_{\mathcal{S}_{k,i}}| - \log |\mathbf{\Theta}_{k,i} + \mathbf{\Sigma}_{\mathcal{S}_{k+1,2i-1}}|$$
$$- \log |\mathbf{\Theta}_{k,i} + \mathbf{\Sigma}_{\mathcal{S}_{k+1,2i}}| \big)$$
$$+ \sum_{k=1}^{L-1} \sum_{i=1}^{2^{k-1}} \text{tr}(\mathbf{\Theta}_{k,i}(\mathbf{M}_{k,i} - \mathbf{M}_{k+1,2i-1} - \mathbf{M}_{k+1,2i}))$$
$$+ \sum_{i=1}^{2^{L-2}} \text{tr}(\mathbf{\Sigma}_X \mathbf{M}_{L,2i-1}).$$

Therefore, we have

$$\nabla_{\boldsymbol{\Theta}_{k,i}} \mathcal{L}|_{\underline{\boldsymbol{\Theta}}=\underline{\boldsymbol{\Theta}}^*, \underline{\mathbf{M}}=\underline{\mathbf{M}}^*}$$
$$= (\boldsymbol{\Theta}_{k,i}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{k,i}})^{-1} - (\boldsymbol{\Theta}_{k,i}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}})^{-1}$$
$$- (\boldsymbol{\Theta}_{k,i}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i}})^{-1}$$
$$+ \mathbf{M}_{k,i}^* - \mathbf{M}_{k+1,2i-1}^* - \mathbf{M}_{k+1,2i}^*,$$
$$k = 1, \cdots, L-1; i = 1, \cdots, 2^{k-1}. \quad (59)$$

Substituting (59) into (58) completes the proof of Lemma 1.

APPENDIX C
PROOF OF LEMMA 2

Note that

$$\boldsymbol{\Theta}_{k,i}^* + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k+1,2i-1}}$$
$$= \boldsymbol{\Theta}_{k+1,2i-1}^* + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k+1,2i-1}} - (\boldsymbol{\Theta}_{k+1,2i-1}^* - \boldsymbol{\Theta}_{k,i}^*)$$
$$= ((\boldsymbol{\Theta}_{k+1,2i-1}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}})^{-1} + \mathbf{M}_{k+1,2i-1}^*)^{-1}$$
$$- (\boldsymbol{\Theta}_{k+1,2i-1}^* - \boldsymbol{\Theta}_{k,i}^*) \quad (60)$$
$$= (\mathbf{I}_m + (\boldsymbol{\Theta}_{k+1,2i-1}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}})\mathbf{M}_{k+1,2i-1}^*)^{-1}$$
$$(\boldsymbol{\Theta}_{k+1,2i-1}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}}) - (\boldsymbol{\Theta}_{k+1,2i-1}^* - \boldsymbol{\Theta}_{k,i}^*)$$
$$= (\mathbf{I}_m + (\boldsymbol{\Theta}_{k,i}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}})\mathbf{M}_{k+1,2i-1}^*)^{-1}$$
$$(\boldsymbol{\Theta}_{k+1,2i-1}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}}) - (\boldsymbol{\Theta}_{k+1,2i-1}^* - \boldsymbol{\Theta}_{k,i}^*)$$
$$\quad (61)$$
$$= ((\boldsymbol{\Theta}_{k,i}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}})^{-1} + \mathbf{M}_{k+1,2i-1}^*)^{-1}$$
$$(\boldsymbol{\Theta}_{k,i}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}})^{-1}(\boldsymbol{\Theta}_{k+1,2i-1}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}})$$
$$- (\boldsymbol{\Theta}_{k+1,2i-1}^* - \boldsymbol{\Theta}_{k,i}^*)$$
$$= ((\boldsymbol{\Theta}_{k,i}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}})^{-1} + \mathbf{M}_{k+1,2i-1}^*)^{-1}$$
$$(\boldsymbol{\Theta}_{k,i}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}})^{-1}(\boldsymbol{\Theta}_{k+1,2i-1}^* - \boldsymbol{\Theta}_{k,i}^*)$$
$$+ ((\boldsymbol{\Theta}_{k,i}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}})^{-1} + \mathbf{M}_{k+1,2i-1}^*)^{-1}$$
$$- (\boldsymbol{\Theta}_{k+1,2i-1}^* - \boldsymbol{\Theta}_{k,i}^*)$$
$$= ((\boldsymbol{\Theta}_{k,i}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}})^{-1} + \mathbf{M}_{k+1,2i-1}^*)^{-1}$$
$$((\boldsymbol{\Theta}_{k,i}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}})^{-1} + \mathbf{M}_{k+1,2i-1}^*)$$
$$(\boldsymbol{\Theta}_{k+1,2i-1}^* - \boldsymbol{\Theta}_{k,i}^*)$$
$$+ ((\boldsymbol{\Theta}_{k,i}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}})^{-1} + \mathbf{M}_{k+1,2i-1}^*)^{-1}$$
$$- (\boldsymbol{\Theta}_{k+1,2i-1}^* - \boldsymbol{\Theta}_{k,i}^*) \quad (62)$$
$$= (\boldsymbol{\Theta}_{k+1,2i-1}^* - \boldsymbol{\Theta}_{k,i}^*)$$
$$+ ((\boldsymbol{\Theta}_{k,i}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}})^{-1} + \mathbf{M}_{k+1,2i-1}^*)^{-1}$$
$$- (\boldsymbol{\Theta}_{k+1,2i-1}^* - \boldsymbol{\Theta}_{k,i}^*)$$
$$= ((\boldsymbol{\Theta}_{k,i}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}})^{-1} + \mathbf{M}_{k+1,2i-1}^*)^{-1},$$
$$k = 1, \cdots, L-2; i = 1, \cdots, 2^{k-1}, \quad (63)$$

where (60) is due to (20) while (61) and (62) are due to (17). Moreover,

$$\boldsymbol{\Theta}_{L-1,i}^* + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{L,2i-1}}$$
$$= \boldsymbol{\Sigma}_X + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{L,2i-1}} - (\boldsymbol{\Sigma}_X - \boldsymbol{\Theta}_{L-1,i}^*)$$
$$= ((\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{L,2i-1}})^{-1} + \mathbf{M}_{L,2i-1}^*)^{-1}$$
$$- (\boldsymbol{\Sigma}_X - \boldsymbol{\Theta}_{L-1,i}^*) \quad (64)$$

$$= (\mathbf{I}_m + (\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{L,2i-1}})\mathbf{M}_{L,2i-1}^*)^{-1}(\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{L,2i-1}})$$
$$- (\boldsymbol{\Sigma}_X - \boldsymbol{\Theta}_{L-1,i}^*)$$
$$= (\mathbf{I}_m + (\boldsymbol{\Theta}_{L-1,i}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{L,2i-1}})\mathbf{M}_{L,2i-1}^*)^{-1}(\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{L,2i-1}})$$
$$- (\boldsymbol{\Sigma}_X - \boldsymbol{\Theta}_{L-1,i}^*) \quad (65)$$
$$= ((\boldsymbol{\Theta}_{L-1,i}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{L,2i-1}})^{-1} + \mathbf{M}_{L,2i-1}^*)^{-1}$$
$$(\boldsymbol{\Theta}_{L-1,i}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{L,2i-1}})^{-1}(\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{L,2i-1}})$$
$$- (\boldsymbol{\Sigma}_X - \boldsymbol{\Theta}_{L-1,i}^*)$$
$$= ((\boldsymbol{\Theta}_{L-1,i}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{L,2i-1}})^{-1} + \mathbf{M}_{L,2i-1}^*)^{-1}$$
$$(\boldsymbol{\Theta}_{L-1,i}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{L,2i-1}})^{-1}(\boldsymbol{\Sigma}_X - \boldsymbol{\Theta}_{L-1,i}^*)$$
$$+ ((\boldsymbol{\Theta}_{L-1,i}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{L,2i-1}})^{-1} + \mathbf{M}_{L,2i-1}^*)^{-1}$$
$$- (\boldsymbol{\Sigma}_X - \boldsymbol{\Theta}_{L-1,i}^*)$$
$$= ((\boldsymbol{\Theta}_{L-1,i}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{L,2i-1}})^{-1} + \mathbf{M}_{L,2i-1}^*)^{-1}$$
$$((\boldsymbol{\Theta}_{L-1,i}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{L,2i-1}})^{-1} + \mathbf{M}_{L,2i-1}^*)(\boldsymbol{\Sigma}_X - \boldsymbol{\Theta}_{L-1,i}^*)$$
$$+ ((\boldsymbol{\Theta}_{L-1,i}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{L,2i-1}})^{-1} + \mathbf{M}_{L,2i-1}^*)^{-1}$$
$$- (\boldsymbol{\Sigma}_X - \boldsymbol{\Theta}_{L-1,i}^*) \quad (66)$$
$$= (\boldsymbol{\Sigma}_X - \boldsymbol{\Theta}_{L-1,i}^*) + ((\boldsymbol{\Theta}_{L-1,i}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{L,2i-1}})^{-1} + \mathbf{M}_{L,2i-1}^*)^{-1}$$
$$- (\boldsymbol{\Sigma}_X - \boldsymbol{\Theta}_{L-1,i}^*)$$
$$= ((\boldsymbol{\Theta}_{L-1,i}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{L,2i-1}})^{-1} + \mathbf{M}_{L,2i-1}^*)^{-1},$$
$$i = 1, \cdots, 2^{L-2}, \quad (67)$$

where (64) is due to (21) while (65) and (66) are due to (19). Combining (63) and (67) proves (22). The proof of (23) is similar to that of (22) and is thus omitted.

It is easy to see that (24) is a simple consequence of (15), (20), (22), and (23). In view of (20), (22), and (23), we have

$$(\boldsymbol{\Theta}_{k,i}^* + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k,i}})^{-1} \succ \mathbf{0},$$
$$(\boldsymbol{\Theta}_{k,i}^* + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k+1,2i-1}})^{-1} \succ \mathbf{0},$$
$$(\boldsymbol{\Theta}_{k,i}^* + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k+1,2i}})^{-1} \succ \mathbf{0}$$

for $k = 1, \cdots, L-1; i = 1, \cdots, 2^{k-1}$, which, together with (24), implies

$$\tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k+1,2i-1}}, \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k+1,2i}} \succ \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k,i}},$$
$$k = 1, \cdots, L-1; \quad i = 1, \cdots, 2^{k-1}.$$

Therefore, for the purpose of proving (25), it suffices to show that $\tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{1,1}} \succ \mathbf{0}$. Indeed,

$$\tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{1,1}}$$
$$= ((\boldsymbol{\Theta}_{1,1}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{1,1}})^{-1} + \mathbf{M}_{1,1}^*)^{-1} - \boldsymbol{\Theta}_{1,1}^* \quad (68)$$
$$= (\mathbf{I}_m + (\boldsymbol{\Theta}_{1,1}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{1,1}})\mathbf{M}_{1,1}^*)^{-1}(\boldsymbol{\Theta}_{1,1}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{1,1}}) - \boldsymbol{\Theta}_{1,1}^*$$
$$= (\mathbf{I}_m + \boldsymbol{\Sigma}_{\mathcal{S}_{1,1}}\mathbf{M}_{1,1}^*)^{-1}(\boldsymbol{\Theta}_{1,1}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{1,1}}) - \boldsymbol{\Theta}_{1,1}^* \quad (69)$$
$$= (\boldsymbol{\Sigma}_{\mathcal{S}_{1,1}}^{-1} + \mathbf{M}_{1,1}^*)^{-1}\boldsymbol{\Sigma}_{\mathcal{S}_{1,1}}^{-1}(\boldsymbol{\Theta}_{1,1}^* + \boldsymbol{\Sigma}_{\mathcal{S}_{1,1}}) - \boldsymbol{\Theta}_{1,1}^*$$
$$= (\boldsymbol{\Sigma}_{\mathcal{S}_{1,1}}^{-1} + \mathbf{M}_{1,1}^*)^{-1} + (\boldsymbol{\Sigma}_{\mathcal{S}_{1,1}}^{-1} + \mathbf{M}_{1,1}^*)^{-1}\boldsymbol{\Sigma}_{\mathcal{S}_{1,1}}^{-1}\boldsymbol{\Theta}_{1,1}^* - \boldsymbol{\Theta}_{1,1}^*$$
$$= (\boldsymbol{\Sigma}_{\mathcal{S}_{1,1}}^{-1} + \mathbf{M}_{1,1}^*)^{-1}$$
$$+ (\boldsymbol{\Sigma}_{\mathcal{S}_{1,1}}^{-1} + \mathbf{M}_{1,1}^*)^{-1}(\boldsymbol{\Sigma}_{\mathcal{S}_{1,1}}^{-1} + \mathbf{M}_{1,1}^*)\boldsymbol{\Theta}_{1,1}^* - \boldsymbol{\Theta}_{1,1}^* \quad (70)$$
$$= (\boldsymbol{\Sigma}_{\mathcal{S}_{1,1}}^{-1} + \mathbf{M}_{1,1}^*)^{-1} \succ \mathbf{0},$$

where (68) is due to (20) while (69) and (70) are due to (16).

It can be verified that

$$
\begin{aligned}
(\boldsymbol{\Theta}^*_{1,1} &+ \boldsymbol{\Sigma}_{\mathcal{S}_{1,1}})^{-1}\boldsymbol{\Sigma}_{\mathcal{S}_{1,1}} \\
&= (\boldsymbol{\Theta}^*_{1,1} + \boldsymbol{\Sigma}_{\mathcal{S}_{1,1}})^{-1}(\boldsymbol{\Theta}^*_{1,1} + \boldsymbol{\Sigma}_{\mathcal{S}_{1,1}} - \boldsymbol{\Theta}^*_{1,1}) \\
&= \mathbf{I}_m - (\boldsymbol{\Theta}^*_{1,1} + \boldsymbol{\Sigma}_{\mathcal{S}_{1,1}})^{-1}\boldsymbol{\Theta}^*_{1,1} \\
&= \mathbf{I}_m - ((\boldsymbol{\Theta}^*_{1,1} + \boldsymbol{\Sigma}_{\mathcal{S}_{1,1}})^{-1} + \mathbf{M}^*_{1,1})\boldsymbol{\Theta}^*_{1,1} \quad (71) \\
&= \mathbf{I}_m - (\boldsymbol{\Theta}^*_{1,1} + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{1,1}})^{-1}\boldsymbol{\Theta}^*_{1,1} \quad (72) \\
&= (\boldsymbol{\Theta}^*_{1,1} + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{1,1}})^{-1}\tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{1,1}},
\end{aligned}
$$

where (71) is due to (16), and (72) is due to (20). This proves (26).

Note that

$$
\begin{aligned}
(\boldsymbol{\Theta}^*_{k+1,2i-1} &+ \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}})^{-1}(\boldsymbol{\Theta}^*_{k,i} + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}}) \\
&= (\boldsymbol{\Theta}^*_{k+1,2i-1} + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}})^{-1} \\
&\quad (\boldsymbol{\Theta}^*_{k+1,2i-1} + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}} - (\boldsymbol{\Theta}^*_{k+1,2i-1} - \boldsymbol{\Theta}^*_{k,i})) \\
&= \mathbf{I}_m - (\boldsymbol{\Theta}^*_{k+1,2i-1} + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}})^{-1}(\boldsymbol{\Theta}^*_{k+1,2i-1} - \boldsymbol{\Theta}^*_{k,i}) \\
&= \mathbf{I}_m - ((\boldsymbol{\Theta}^*_{k+1,2i-1} + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}})^{-1} + \mathbf{M}^*_{k+1,2i-1}) \\
&\quad (\boldsymbol{\Theta}^*_{k+1,2i-1} - \boldsymbol{\Theta}^*_{k,i}) \quad (73) \\
&= \mathbf{I}_m - (\boldsymbol{\Theta}^*_{k+1,2i-1} + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k+1,2i-1}})^{-1}(\boldsymbol{\Theta}^*_{k+1,2i-1} - \boldsymbol{\Theta}^*_{k,i}) \\
&\quad\quad\quad (74) \\
&= (\boldsymbol{\Theta}^*_{k+1,2i-1} + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k+1,2i-1}})^{-1}(\boldsymbol{\Theta}^*_{k,i} + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k+1,2i-1}}), \\
&\quad\quad k = 1, \cdots, L-2; \quad i = 1, \cdots, 2^{k-1},
\end{aligned}
$$

where (73) is due to (17), and (74) is due to (20). This proves (27). The proof of (28) is similar to that of (27) and is thus omitted.

Now it remains to prove (29) since (30) is trivially true. Indeed,

$$
\begin{aligned}
(\boldsymbol{\Sigma}_X &+ \boldsymbol{\Sigma}_{\mathcal{S}_{L,2i-1}})^{-1}(\boldsymbol{\Theta}^*_{L-1,i} + \boldsymbol{\Sigma}_{\mathcal{S}_{L,2i-1}}) \\
&= (\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{L,2i-1}})^{-1}(\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{L,2i-1}} - (\boldsymbol{\Sigma}_X - \boldsymbol{\Theta}^*_{L-1,i})) \\
&= \mathbf{I}_m - (\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{L,2i-1}})^{-1}(\boldsymbol{\Sigma}_X - \boldsymbol{\Theta}^*_{L-1,i}) \\
&= \mathbf{I}_m - ((\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{L,2i-1}})^{-1} + \mathbf{M}^*_{L,2i-1})(\boldsymbol{\Sigma}_X - \boldsymbol{\Theta}^*_{L-1,i}) \\
&\quad\quad\quad (75) \\
&= \mathbf{I}_m - (\boldsymbol{\Sigma}_X + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{L,2i-1}})^{-1}(\boldsymbol{\Sigma}_X - \boldsymbol{\Theta}^*_{k,i}) \quad (76) \\
&= (\boldsymbol{\Sigma}_X + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{L,2i-1}})^{-1}(\boldsymbol{\Theta}^*_{L-1,i} + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{L,2i-1}}), \\
&\quad\quad\quad i = 1, \cdots, 2^{L-2},
\end{aligned}
$$

where (75) is due to (19), and (76) is due to (21). This completes the proof of Lemma 2.

APPENDIX D
PROOF OF LEMMA 3

It can be verified that

$$
\begin{aligned}
R^*_G(&\underline{\mathbf{D}}) \\
&= \frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{1,1}}|}{|\boldsymbol{\Sigma}_{\mathcal{S}_{1,1}}|} + \sum_{k=1}^{L-1}\sum_{i=1}^{2^{k-1}}\left(\frac{1}{2}\log\frac{|\boldsymbol{\Theta}^*_{k,i} + \boldsymbol{\Sigma}_{\mathcal{S}_{k,i}}|}{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{k,i}}|}\right. \\
&\quad \left.+ \frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}}|}{|\boldsymbol{\Theta}^*_{k,i} + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}}|} + \frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i}}|}{|\boldsymbol{\Theta}^*_{k,i} + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i}}|}\right) \\
&= \frac{1}{2}\log\frac{|\boldsymbol{\Theta}^*_{1,1} + \boldsymbol{\Sigma}_{\mathcal{S}_{1,1}}|}{|\boldsymbol{\Sigma}_{\mathcal{S}_{1,1}}|} + \sum_{k=2}^{L-1}\sum_{i=1}^{2^{k-1}}\frac{1}{2}\log\frac{|\boldsymbol{\Theta}^*_{k,i} + \boldsymbol{\Sigma}_{\mathcal{S}_{k,i}}|}{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{k,i}}|}
\end{aligned}
$$

$$
\begin{aligned}
&+ \sum_{k=1}^{L-1}\sum_{i=1}^{2^{k-1}}\frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}}||\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i}}|}{|\boldsymbol{\Theta}^*_{k,i} + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}}||\boldsymbol{\Theta}^*_{k,i} + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i}}|} \\
&= \frac{1}{2}\log\frac{|\boldsymbol{\Theta}^*_{1,1} + \boldsymbol{\Sigma}_{\mathcal{S}_{1,1}}|}{|\boldsymbol{\Sigma}_{\mathcal{S}_{1,1}}|} \\
&+ \sum_{k=1}^{L-2}\sum_{i=1}^{2^{k-1}}\left(\frac{1}{2}\log\frac{|\boldsymbol{\Theta}^*_{k+1,2i-1} + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}}|}{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}}|}\right. \\
&\quad \left.+ \frac{1}{2}\log\frac{|\boldsymbol{\Theta}^*_{k+1,2i} + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i}}|}{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i}}|}\right) \\
&+ \sum_{k=1}^{L-1}\sum_{i=1}^{2^{k-1}}\frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}}||\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i}}|}{|\boldsymbol{\Theta}^*_{k,i} + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}}||\boldsymbol{\Theta}^*_{k,i} + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i}}|} \\
&= \frac{1}{2}\log\frac{|\boldsymbol{\Theta}^*_{1,1} + \boldsymbol{\Sigma}_{\mathcal{S}_{1,1}}|}{|\boldsymbol{\Sigma}_{\mathcal{S}_{1,1}}|} \\
&+ \sum_{k=1}^{L-2}\sum_{i=1}^{2^{k-1}}\left(\frac{1}{2}\log\frac{|\boldsymbol{\Theta}^*_{k+1,2i-1} + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}}|}{|\boldsymbol{\Theta}^*_{k,i} + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i-1}}|}\right. \\
&\quad \left.+ \frac{1}{2}\log\frac{|\boldsymbol{\Theta}^*_{k+1,2i} + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i}}|}{|\boldsymbol{\Theta}^*_{k,i} + \boldsymbol{\Sigma}_{\mathcal{S}_{k+1,2i}}|}\right) \\
&+ \sum_{i=1}^{2^{L-2}}\frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{L,2i-1}}||\boldsymbol{\Sigma}_X + \boldsymbol{\Sigma}_{\mathcal{S}_{L,2i}}|}{|\boldsymbol{\Theta}^*_{L-1,i} + \boldsymbol{\Sigma}_{\mathcal{S}_{L,2i-1}}||\boldsymbol{\Theta}^*_{L-1,i} + \boldsymbol{\Sigma}_{\mathcal{S}_{L,2i}}|} \\
&= \frac{1}{2}\log\frac{|\boldsymbol{\Theta}^*_{1,1} + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{1,1}}|}{|\tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{1,1}}|} \\
&+ \sum_{k=1}^{L-2}\sum_{i=1}^{2^{k-1}}\left(\frac{1}{2}\log\frac{|\boldsymbol{\Theta}^*_{k+1,2i-1} + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k+1,2i-1}}|}{|\boldsymbol{\Theta}^*_{k,i} + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k+1,2i-1}}|}\right. \\
&\quad \left.+ \frac{1}{2}\log\frac{|\boldsymbol{\Theta}^*_{k+1,2i} + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k+1,2i}}|}{|\boldsymbol{\Theta}^*_{k,i} + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k+1,2i}}|}\right) \\
&+ \sum_{i=1}^{2^{L-2}}\frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_X + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{L,2i-1}}||\boldsymbol{\Sigma}_X + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{L,2i}}|}{|\boldsymbol{\Theta}^*_{L-1,i} + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{L,2i-1}}||\boldsymbol{\Theta}^*_{L-1,i} + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{L,2i}}|} \quad (77) \\
&= \frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_X + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{1,1}}|}{|\tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{1,1}}|} + \sum_{k=1}^{L-1}\sum_{i=1}^{2^{k-1}}\left(\frac{1}{2}\log\frac{|\boldsymbol{\Theta}^*_{k,i} + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k,i}}|}{|\boldsymbol{\Sigma}_X + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k,i}}|}\right. \\
&\quad \left.+ \frac{1}{2}\log\frac{|\boldsymbol{\Sigma}_X + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k+1,2i-1}}||\boldsymbol{\Sigma}_X + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k+1,2i}}|}{|\boldsymbol{\Theta}^*_{k,i} + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k+1,2i-1}}||\boldsymbol{\Theta}^*_{k,i} + \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k+1,2i}}|}\right),
\end{aligned}
$$

where (77) is due to (26)–(30).

APPENDIX E
PROOF OF LEMMA 4

It follows by (25) that

$$
\tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k+1,2i-1}} - \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k,i}}, \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k+1,2i}} - \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k,i}} \succ \mathbf{0}.
$$

Moreover, it is known [11, eq. (107)] that (24) implies

$$
\begin{aligned}
\boldsymbol{\Theta}^*_{k,i} = (\tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k+1,2i-1}} &- \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k,i}})^{\frac{1}{2}}\left((\tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k+1,2i-1}} - \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k,i}})^{-\frac{1}{2}}\right. \\
&(\tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k+1,2i}} - \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k,i}})(\tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k+1,2i-1}} - \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k,i}})^{-\frac{1}{2}}\Big)^{\frac{1}{2}} \\
&(\tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k+1,2i-1}} - \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k,i}})^{\frac{1}{2}} - \tilde{\boldsymbol{\Sigma}}_{\mathcal{S}_{k,i}},
\end{aligned}
$$

which further implies

$$\tilde{\Sigma}_{\mathcal{S}_{k+1,2i}} - \tilde{\Sigma}_{\mathcal{S}_{k,i}}$$
$$= (\Theta_{k,i}^* - \tilde{\Sigma}_{\mathcal{S}_{k,i}})(\tilde{\Sigma}_{\mathcal{S}_{k+1,2i-1}} - \tilde{\Sigma}_{\mathcal{S}_{k,i}})^{-1}(\Theta_{k,i}^* - \tilde{\Sigma}_{\mathcal{S}_{k,i}}).$$

Therefore, $\Lambda_{k,i}$ is indeed a positive semi-definite matrix.

According to [11, Lemmas 3 and 4],

$$\Gamma_{k,i} \succ \mathbf{0},$$
$$\tilde{\Sigma}_{\mathcal{S}_{k,i}}^{-1} = (\mathbf{I}_m, \mathbf{I}_m)\Gamma_{k,i}^{-1}(\mathbf{I}_m, \mathbf{I}_m)^T. \tag{78}$$

Now set

$$(\mathbf{H}_{k+1,2i-1}, \mathbf{H}_{k+1,2i}) = (\tilde{\Sigma}_{\mathcal{S}_{k,i}}, \tilde{\Sigma}_{\mathcal{S}_{k,i}})\Gamma_{k,i}^{-1}.$$

It can be verified that

$$\begin{aligned}
\mathbf{H}_{k+1,2i-1} + \mathbf{H}_{k+1,2i} &= (\tilde{\Sigma}_{\mathcal{S}_{k,i}}, \tilde{\Sigma}_{\mathcal{S}_{k,i}})\Gamma_{k,i}^{-1}(\mathbf{I}_m, \mathbf{I}_m)^T \\
&= \tilde{\Sigma}_{\mathcal{S}_{k,i}}(\mathbf{I}_m, \mathbf{I}_m)\Gamma_{k,i}^{-1}(\mathbf{I}_m, \mathbf{I}_m)^T \\
&= \tilde{\Sigma}_{\mathcal{S}_{k,i}}\tilde{\Sigma}_{\mathcal{S}_{k,i}}^{-1} \\
&= \mathbf{I}_m,
\end{aligned} \tag{79}$$

where (79) is due to (78). Furthermore,

$$\begin{aligned}
&(\mathbf{H}_{k+1,2i-1}, \mathbf{H}_{k+1,2i})\Lambda_{k,i}(\mathbf{H}_{k+1,2i-1}, \mathbf{H}_{k+1,2i})^T \\
&= (\mathbf{H}_{k+1,2i-1}, \mathbf{H}_{k+1,2i})\Gamma_{k,i}(\mathbf{H}_{k+1,2i-1}, \mathbf{H}_{k+1,2i})^T \\
&\quad - (\mathbf{H}_{k+1,2i-1}, \mathbf{H}_{k+1,2i})\begin{pmatrix} \tilde{\Sigma}_{\mathcal{S}_{k,i}} & \tilde{\Sigma}_{\mathcal{S}_{k,i}} \\ \tilde{\Sigma}_{\mathcal{S}_{k,i}} & \tilde{\Sigma}_{\mathcal{S}_{k,i}} \end{pmatrix} \\
&\quad (\mathbf{H}_{k+1,2i-1}, \mathbf{H}_{k+1,2i})^T \\
&= \tilde{\Sigma}_{\mathcal{S}_{k,i}} - \tilde{\Sigma}_{\mathcal{S}_{k,i}} \\
&= \mathbf{0},
\end{aligned} \tag{80}$$

where (80) is due to (78). This completes the proof of Lemma 4.

REFERENCES

[1] 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.

[2] R. Ahlswede, "The rate-distortion region for multiple descriptions without excess rate," *IEEE Trans. Inf. Theory*, vol. 31, no. 6, pp. 721–726, Nov. 1985.

[3] F.-W. Fu and R. W. Yeung, "On the rate-distortion region for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 48, no. 7, pp. 2012–2021, Jul. 2002.

[4] Z. Zhang and T. Berger, "New results in binary multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 33, no. 4, pp. 502–521, Jul. 1987.

[5] R. Venkataramani, G. Kramer, and V. K. Goyal, "Multiple description coding with many channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 9, pp. 2106–2114, Sep. 2003.

[6] S. S. Pradhan, R. Puri, and K. Ramchandran, "*n*-channel symmetric multiple descriptions—Part I: (n, k) source-channel erasure codes," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 47–61, Jan. 2004.

[7] R. Puri, S. S. Pradhan, and K. Ramchandran, "*n*-channel symmetric multiple descriptions—Part II: An achievable rate-distortion region," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1377–1392, Apr. 2005.

[8] C. Tian and J. Chen, "New coding schemes for the symmetric *k*-descrption problem," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 5344–5365, Oct. 2010.

[9] K. B. Viswanatha, E. Akyol, and K. Rose, "Combinatorial message sharing and a new achievable region for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 62, no. 2, pp. 769–792, Feb. 2016.

[10] L. Ozarow, "On a source-coding problem with two channels and three receivers," *Bell Syst. Tech. J.*, vol. 59, no. 10, pp. 1909–1921, Dec. 1980.

[11] H. Wang and P. Viswanath, "Vector Gaussian multiple description with individual and central receivers," *IEEE Trans. Inf. Theory*, vol. 53, no. 6, pp. 2133–2153, Jun. 2007.

[12] J. Chen, "Rate region of Gaussian multiple description coding with individual and central distortion constraints," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 3991–4005, Sep. 2009.

[13] Y. Oohama, "The rate-distortion function for the quadratic Gaussian CEO problem," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1057–1070, May 1998.

[14] Y. Oohama, "Rate-distortion theory for Gaussian multiterminal source coding systems with several side informations at the decoder," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2577–2593, Jul. 2005.

[15] A. B. 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.

[16] A. B. Wagner, S. Tavildar, and P. Viswanath, "Rate region of the quadratic Gaussian two-encoder source-coding problem," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1938–1961, May 2008.

[17] S. Tavildar, P. Viswanath, and A. B. Wagner, "The Gaussian many-help-one distributed source coding problem," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 564–581, Jan. 2010.

[18] J. Wang, J. Chen, and X. Wu, "On the sum rate of Gaussian multiterminal source coding: New proofs and results," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3946–3960, Aug. 2010.

[19] J. Wang and J. Chen, "Vector Gaussian two-terminal source coding," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3693–3708, Jun. 2013.

[20] J. Wang and J. Chen, "Vector Gaussian multiterminal source coding," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5533–5552, Sep. 2014.

[21] Y. Oohama, "Indirect and direct Gaussian distributed source coding problems," *IEEE Trans. Inf. Theory*, vol. 60, no. 12, pp. 7506–7539, Dec. 2014.

[22] C. Tian, S. Mohajer, and S. N. Diggavi, "Approximating the Gaussian multiple description rate region under symmetric distortion constraints," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3869–3891, Aug. 2009.

[23] L. Song, S. Shuo, and J. Chen, "A lower bound on the sum rate of multiple description coding with symmetric distortion constraints," *IEEE Trans. Inf. Theory*, vol. 60, no. 12, pp. 7547–7567, Dec. 2014.

[24] L. Song, J. Chen, J. Wang, and T. Liu, "Gaussian robust sequential and predictive coding," *IEEE Trans. Inf. Theory*, vol. 59, no. 6, pp. 3635–3652, Jun. 2013.

[25] S. Ihara, "On the capacity of channels with additive non-gaussian noise," *Inf. Control*, vol. 37, no. 1, pp. 34–39, Apr. 1978.

[26] S. N. Diggavi and T. M. Cover, "The worst additive noise under a covariance constraint," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 3072–3081, Nov. 2001.

[27] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936–3964, Sep. 2006.

[28] H. Wang and P. Viswanath, "Vector Gaussian multiple description with two levels of receivers," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 401–410, Jan. 2009.

[29] J. Chen, A. Salimi, T. Liu, and C. Tian, "Orbit-entropy cones and extremal pairwise orbit-entropy inequalities," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Spain, Jul. 2016, pp. 2614–2618.

[30] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA, USA: Athena Scientific, 1999.

Yinfei Xu (S'10) received the B.E. and Ph.D. degrees in 2008 and 2016, respectively, both in Information Engineering, from Southeast University, Nanjing, China.

Since March 2016, he has been with the Institute of Network Coding at The Chinese University of Hong Kong, Hong Kong, where he is currently a Postdoctoral Fellow. He was a visiting student in the Department of Electrical and Computer Engineering at McMaster University, Hamilton, ON, Canada, from July 2014 to January 2015. His research interests include information theory, signal processing and wireless communications.

Jun Chen (S'03–M'06–SM'16) received the B.E. degree with honors in communication engineering from Shanghai Jiao Tong University, Shanghai, China, in 2001 and the M.S. and Ph.D. degrees in electrical and computer engineering from Cornell University, Ithaca, NY, in 2004 and 2006, respectively.

He was a Postdoctoral Research Associate in the Coordinated Science Laboratory at the University of Illinois at Urbana-Champaign, Urbana, IL, from September 2005 to July 2006, and a Postdoctoral Fellow at the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, from July 2006 to August 2007. Since September 2007 he has been with the Department of Electrical and Computer Engineering at McMaster University, Hamilton, ON, Canada, where he is currently an Associate Professor and a Joseph Ip Distinguished Engineering Fellow. His research interests include information theory, wireless communications, and signal processing.

He received several awards for his research, including the Josef Raviv Memorial Postdoctoral Fellowship in 2006, the Early Researcher Award from the Province of Ontario in 2010, and the IBM Faculty Award in 2010. He served as an Associate Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY from 2014 to 2016.

Qiao Wang (M'06) received the B.S., M.S., and Ph.D. degrees in 1988, 1994, and 1997, respectively, all in mathematics, from Wuhan University, China. During 1997 to 1999, he was with the Department of Radio Engineering, Southeast University, Nanjing, China, as a Postdoctoral Fellow, after which, he then joined the Department as an Associate Professor in 1999, and has been a Full Professor since 2001. He was a Visiting Scientist with the Division of Engineering and Applied Science, Harvard University, Cambridge, MA, from Jan. 2003 to Jan. 2004. From June 2007 to June 2017, he was the Director of Laboratory of Signal and Information Processing, Southeast University. In past decades, Dr. Wang led a lot of academic-industrial collaboration projects in signal processing, video analysis and related areas.

Prof. Wang is the associate editor of *China Communications* and *ICT Express*, etc. His research interests include urban data analytics, biomedical data analysis, and information theory.