Vector Gaussian Successive Refinement With Degraded Side Information

Yinfei Xu, *Member, IEEE*, Xuan Guang, *Member, IEEE*, Jian Lu, *Member, IEEE*, and Jun Chen, *Senior Member, IEEE*

***Abstract*—We investigate the problem of the successive refinement for Wyner-Ziv coding with degraded side information and obtain a complete characterization of the rate region for the quadratic vector Gaussian case. The achievability part is based on the evaluation of the Tian-Diggavi inner bound that involves Gaussian auxiliary random vectors. For the converse part, a matching outer bound is obtained with the aid of a new extremal inequality. Herein, the proof of this extremal inequality depends on the integration of the monotone path argument and the doubling trick as well as information-estimation relations.**

***Index Terms*—Extremal inequality, lossy source coding, mean squared error, rate region, side information, successive refinement, vector Gaussian source, Wyner-Ziv problem.**

I. Introduction

THE research on network source coding can be traced back to the seminal work by Slepian and Wolf [2], where they considered, among other things, the problem of lossless source coding with side information at the decoder. Wyner and Ziv [3] studied the lossy source coding version of this problem (which later bears their names) and characterized its information-theoretic limit. Subsequently, the Wyner-Ziv problem was extended in various ways (see, e.g., [4]–[8]). One particular extension, known as successive refinement for Wyner-Ziv coding with degraded side information, is as follows: A source is encoded and decoded, in a successive manner, to meet different distortion constraints with the aid of progressively enhanced decoder side information. This extended Wyner-Ziv problem was tackled by Steinberg and Merhav [9] for the two-stage case and by Tian and Diggavi [10] for the multi-stage case. Specifically, the computable characterizations of rate regions in the discrete memoryless setting (with a general distortion measure) and in the scalar Gaussian setting (with the quadratic error distortion measure) were obtained accordingly.

In this paper, we consider the same extended Wyner-Ziv problem with a particular attention paid to the vector Gaussian setting (under covariance distortion constraints). The heart of the present paper is a new inequality regarding the optimality of the Gaussian solution to a certain extremal problem. It is well known that extremal inequalities play an important role in characterizing the fundamental limits of Gaussian network source and channel coding problems. Indeed, they are indispensable to the converse argument for the Gaussian broadcast channel coding problem [11]–[20], the Gaussian interference channel coding problem [21]–[23], the Gaussian multi-terminal source coding problem [24]–[30], the secret key generation problem [31], the Gaussian multiple description problem [32]–[35], and others [36], [37].

Basic extremal inequalities that rely on the differential-entropy-maximizing property of the Gaussian distribution can only handle simple situations where the objective functional can be greedily optimized. When there are two or more conflicting terms, Shannon's entropy power inequality is often used to resolve the tension. However, the proportionality condition on the relevant covariance matrices needed for the tightness of the entropy power inequality is quite restrictive, typically only satisfied in scalar source and channel coding problems. As a consequence, more sophisticated extremal inequalities are needed to deal with vector Gaussian sources and channels. The proofs of such extremal inequalities, as well as the proof of the entropy power inequality, are often proved by invoking the monotone path argument or its variants.

The conventional monotone path argument nevertheless appears to have its own limitations. For example, it fails to yield a tight outer bound on the capacity region of the two-user vector Gaussian broadcast channel with private and common messages. The desired result is eventually obtained by Geng and Nair [38] through a different approach involving so-called doubling trick. On the other hand, this approach obscures some useful information regarding the optimal Gaussian solution. Fortunately, this problem can be remedied via a systematic integration of the monotone path argument and the doubling trick, as shown by Wang and Chen [39] in their new proof of Courtade's extremal inequality [40]. In this work, we make

Manuscript received February 17, 2020; revised August 11, 2021; accepted August 16, 2021. Date of publication August 24, 2021; date of current version October 20, 2021. The work of Yinfei Xu was supported in part by the National Natural Science Foundation of China under Grant 61901105 and Grant 52078117 and in part by the Natural Science Foundation of Jiangsu Province under Grant BK20190343. The work of Xuan Guang was supported in part by the Natural Science Foundation of China under Grant 61771259, in part by the Natural Science Foundation of Tianjin, China, under Grant 20JCYBJC01390, in part by the National Key Research and Development Program of China under Grant 2018YFA0704703, and in part by the Fundamental Research Funds for the Central Universities of China (Nankai University). An earlier version of this paper was presented in part at the 2019 IEEE International Symposium on Information Theory [1] [DOI: 10.1109/ISIT.2019.8849425]. *(Corresponding author: Yinfei Xu.)*

Yinfei Xu and Jian Lu are with the School of Information Science and Engineering, Southeast University, Nanjing 210096, China (e-mail: yinfeixu@seu.edu.cn; lujian1980@seu.edu.cn).

Xuan Guang is with the School of Mathematical Sciences and LPMC, Nankai University, Tianjin 300071, China (e-mail: xguang@nankai.edu.cn).

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

Communicated by C. Tian, Associate Editor for Source Coding.

Digital Object Identifier 10.1109/TIT.2021.3107215

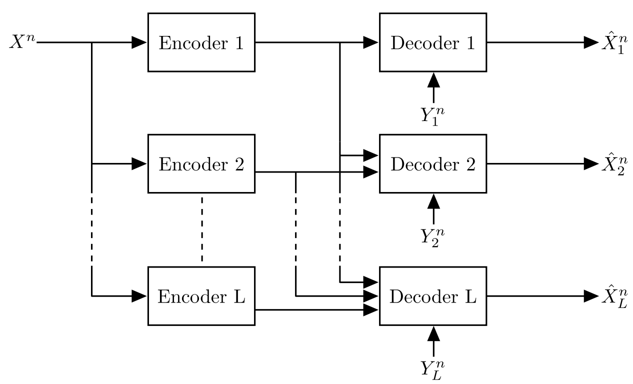

Fig. 1. Successive refinement for Wyner-Ziv coding with degraded side information.

use of this integrated strategy, together with the properties of the minimum mean square error (MMSE) and the Fisher information, to establish a new extremal inequality, which is further leveraged to characterize the rate region of the aforementioned extended Wyner-Ziv problem in the vector Gaussian source setting. It will be seen that the new extremal inequality avoids the comparison of distortion matrices, and thus is particularly handy when dealing with a large number of covariance distortion constraints.

The rest of this paper is organized as follows. We present the problem formulation and the main result in Section II. Section III is devoted to proving a new extremal inequality, which constitutes the main technical part of this paper. The main result is proved in Section IV. We conclude the paper in Section V. During the reviewing process, one anonymous reviewer provided an alternative proof of our main result based on the doubling/rotation method. With his/her kind permission, we include the proof in Appendix D.

II. Problem Statement and Main Result

Let X be a $p \times 1$-dimensional random vector with mean zero and covariance matrix $\boldsymbol{K}_0 \succ \boldsymbol{0}$. Moreover, let

$$Y_i = X + N_i, \qquad i \in [1:L], \tag{1}$$

where N_i is a $p \times 1$-dimensional random vector with mean zero and covariance matrix $\boldsymbol{K}_i \succ \boldsymbol{0}$, $i \in [1:L]$. It is assumed that

$$\boldsymbol{K}_1 \succ \ldots \succ \boldsymbol{K}_{L-1} \succ \boldsymbol{K}_L, \tag{2}$$

and X, $N_i - N_{i+1}$, $i \in [1:L]$, are mutually independent and jointly Gaussian.[1] This assumption implies that

$$X \to Y_L \to Y_{L-1} \to \ldots \to Y_1 \tag{3}$$

forms a Markov chain. Let $(X(t), Y_i(t), i \in [1:L])_{t=1}^{\infty}$ be i.i.d. copies of $(X, Y_i, i \in [1:L])$.

The system model can be described as follows (see also Fig. 1).

[1]Here N_{L+1} is a null random vector with covariance matrix $\boldsymbol{K}_{L+1} = \boldsymbol{0}$.

- L encoding functions $(\phi_i^{(n)}, i \in [1:L])$:

$$\phi_i^{(n)} : \mathcal{X}^n \mapsto \mathcal{M}_i^{(n)}, \qquad i \in [1:L], \tag{4}$$

  where $\phi_i^{(n)}$ maps the source sequence X^n to the codeword $M_i(X^n)$, $i \in [1:L]$.
- L decoding functions $(\varphi_i^{(n)}, i \in [1:L])$:

$$\varphi_i^{(n)} : \prod_{j \in [1:i]} \mathcal{M}_j^{(n)} \times \mathcal{Y}_i^n \mapsto \hat{\mathcal{X}}^n, \qquad i \in [1:L], \tag{5}$$

  where $\varphi_i^{(n)}$ produces the source reconstruction $\hat{X}_i^n(M_j, j \in [1:i], Y_i^n)$ by using codewords $(M_j, j \in [1:i])$ and side information Y_i^n. In particular, under covariance distortion constraints, there is no loss of optimality in assuming that $\varphi_i^{(n)}$ performs MMSE estimation, i.e.,

$$\hat{X}_i^n(M_j, j \in [1:i], Y_i^n) = \mathbb{E}[X^n | M_j, j \in [1:i], Y_i^n].$$

Definition 1: A rate tuple $(R_i, i \in [1:L])$ is said to be achievable subject to covariance distortion constraints $(\boldsymbol{D}_i \succ \boldsymbol{0}, i \in [1:L])$ if there exists a sequence of encoding functions $(\phi_i^{(n)}, i \in [1:L])$ and decoding functions $(\varphi_i^{(n)}, i \in [1:L])$ such that

$$\limsup_{n\to\infty} \frac{1}{n} \log \left| \mathcal{M}_i^{(n)} \right| \le R_i, \qquad i \in [1:L], \tag{6}$$

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \mathbb{E}\left[\left(X(t) - \hat{X}_i(t) \right) \left(X(t) - \hat{X}_i(t) \right)^T \right] \preceq \boldsymbol{D}_i, \quad i \in [1:L]. \tag{7}$$

The rate region $\mathcal{R}^*(\boldsymbol{D}_i, i \in [1:L])$ is defined as the set of all such achievable rate tuples.

The following theorem states a computable characterization of $\mathcal{R}^*(\boldsymbol{D}_i, i \in [1:L])$, which is the main result of this paper.

Theorem 1: $\mathcal{R}^*(\boldsymbol{D}_i, i \in [1:L]) = \mathcal{R}(\boldsymbol{D}_i, i \in [1:L])$, where $\mathcal{R}(\boldsymbol{D}_i, i \in [1:L])$ is the convex hull of the set of $(R_i, i \in [1:L])$ such that

$$R_1 \ge \frac{1}{2} \log \frac{|\boldsymbol{K}_0^{-1} + \boldsymbol{K}_1^{-1} + \boldsymbol{B}_1|}{|\boldsymbol{K}_0^{-1} + \boldsymbol{K}_1^{-1}|}, \tag{8}$$

$$\sum_{j=1}^{i} R_j \ge \frac{1}{2} \log \frac{|\boldsymbol{K}_0^{-1} + \boldsymbol{K}_1^{-1} + \boldsymbol{B}_1|}{|\boldsymbol{K}_0^{-1} + \boldsymbol{K}_1^{-1}|} + \sum_{j=2}^{i} \frac{1}{2} \log \frac{|\boldsymbol{K}_0^{-1} + \boldsymbol{K}_j^{-1} + \sum_{k=1}^{j} \boldsymbol{B}_k|}{|\boldsymbol{K}_0^{-1} + \boldsymbol{K}_j^{-1} + \sum_{k=1}^{j-1} \boldsymbol{B}_k|}, \quad i \in [2:L], \tag{9}$$

for some $(\boldsymbol{B}_i, i \in [1:L])$ satisfying

$$\boldsymbol{B}_i \succeq \boldsymbol{0}, \qquad i \in [1:L], \tag{10}$$

$$\sum_{j=1}^{i} \boldsymbol{B}_j \succeq \boldsymbol{D}_i^{-1} - \boldsymbol{K}_0^{-1} - \boldsymbol{K}_i^{-1}, \qquad i \in [1:L]. \tag{11}$$

The proof of Theorem 1 can be found in Section IV, and it relies critically on the extremal inequality established in Section III.

III. An Extremal Inequality

Theorem 2: Given $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_L \geq 0$, let $(\boldsymbol{B}_i^*, i \in [1:L])$ be any positive semi-definite matrices such that

$$\sum_{j=1}^{i} \boldsymbol{B}_j^* \succeq \boldsymbol{D}_i^{-1} - \boldsymbol{K}_0^{-1} - \boldsymbol{K}_i^{-1}, \qquad i \in [1:L], \tag{12}$$

and

$$\frac{\mu_i}{2}\left(\boldsymbol{K}_0^{-1} + \boldsymbol{K}_i^{-1} + \sum_{j=1}^{i} \boldsymbol{B}_j^*\right)^{-1} - \frac{\mu_{i+1}}{2}\left(\boldsymbol{K}_0^{-1} + \boldsymbol{K}_{i+1}^{-1} + \sum_{j=1}^{i} \boldsymbol{B}_j^*\right)^{-1} = \boldsymbol{\Psi}_i - \boldsymbol{\Psi}_{i+1} + \boldsymbol{\Lambda}_i, \quad i \in [1:L-1], \tag{13}$$

$$\frac{\mu_L}{2}\left(\boldsymbol{K}_0^{-1} + \boldsymbol{K}_L^{-1} + \sum_{j=1}^{L} \boldsymbol{B}_j^*\right)^{-1} = \boldsymbol{\Psi}_L + \boldsymbol{\Lambda}_L, \tag{14}$$

$$\boldsymbol{B}_i^* \boldsymbol{\Psi}_i = \boldsymbol{0}, \quad i \in [1:L], \tag{15}$$

$$\left(\boldsymbol{K}_0^{-1} + \boldsymbol{K}_i^{-1} + \sum_{j=1}^{i} \boldsymbol{B}_j^* - \boldsymbol{D}_i^{-1}\right) \boldsymbol{\Lambda}_i = \boldsymbol{0}, \quad i \in [1:L], \tag{16}$$

for some positive semi-definite matrices $(\boldsymbol{\Psi}_i, i \in [1:L])$ and $(\boldsymbol{\Lambda}_i, i \in [1:L])$. For any random objects $(W_i, i \in [1:L])$ satisfying the Markov chain contraint

$$(W_i, i \in [1:L]) \to X \to Y_L \to Y_{L-1} \to \ldots \to Y_1 \tag{17}$$

and the covariance distortion constraints

$$\mathrm{cov}(X|Y_i, W_j, j \in [1:i]) \preceq \boldsymbol{D}_i, \qquad i \in [1:L], \tag{18}$$

the following extremal inequality holds:

$$\begin{aligned}
&\sum_{i=1}^{L-1} (\mu_i h(Y_i|W_j, j \in [1:i]) - \mu_{i+1} h(Y_{i+1}|W_j, j \in [1:i]) \\
&\qquad -(\mu_i - \mu_{i+1}) h(X|W_j, j \in [1:i])) \\
&\quad + \mu_L h(Y_L|W_j, j \in [1:L]) - \mu_L h(X|W_j, j \in [1:L]) \\
&\geq \sum_{i=1}^{L-1} \left(-\frac{\mu_{i+1}}{2} \log \left| \boldsymbol{K}_{i+1} \left(\boldsymbol{K}_0^{-1} + \boldsymbol{K}_{i+1}^{-1} + \sum_{j=1}^{i} \boldsymbol{B}_j^* \right) \right| \right. \\
&\qquad \left. + \frac{\mu_i}{2} \log \left| \boldsymbol{K}_i \left(\boldsymbol{K}_0^{-1} + \boldsymbol{K}_i^{-1} + \sum_{j=1}^{i} \boldsymbol{B}_j^* \right) \right| \right) \\
&\quad + \frac{\mu_L}{2} \log \left| \boldsymbol{K}_L \left(\boldsymbol{K}_0^{-1} + \boldsymbol{K}_L^{-1} + \sum_{j=1}^{L} \boldsymbol{B}_j^* \right) \right|.
\end{aligned} \tag{19}$$

Remark 1: For the special case $L = 2$, $\boldsymbol{\Lambda}_1 = \boldsymbol{0}$, and $\mu_1 = \mu_2 = 1$, the extremal inequality (19) can be regarded as a variant of [17, Theorem 5], the original proof of which relies on the enhancement argument developed in [41]. However, when $L > 2$, the enhancement argument appears to be inadequate for resolving the difficulty caused by the introduction of $(\boldsymbol{\Psi}_i, i \in [1:L])$ and $(\boldsymbol{\Lambda}_i, i \in [1:L])$. We shall overcome this difficulty via a judicious application of the monotone path argument and the doubling trick.

Remark 2: The doubling trick and the monotone path argument are two widely used approaches for establishing Gaussian extremal inequalities. Inspired by the change measure argument in the proof of Costa's entropy power inequality by Watanabe and Oohama [31] and the monotone path proof of Courtade's strong entropy power inequality in [39], we propose an integrated approach, which appears to be more flexible and informative. Specifically, it will be seen that the doubling trick yields a novel monotone path construction, which enables us to leverage the standard perturbation techniques [26], [42] to prove the optimality of the Gaussian solution.

For notational simplicity, we define

$$\boldsymbol{\Delta}_i^{-1} \triangleq \boldsymbol{K}_0^{-1} + \sum_{j=1}^{i} \boldsymbol{B}_j^*, \qquad i \in [1:L]. \tag{20}$$

The proof of Theorem 2 is divided into four steps.

A. Constructing the Monotone Path

We first construct $3L$ zero-mean Gaussian random vectors

$$X_1^G, \ldots, X_L^G,\ Y_1^G, \ldots, Y_L^G,\ \tilde{Y}_2^G, \ldots, \tilde{Y}_{L+1}^G,$$

which are independent of $(X_i, Y_i, W_i, i \in [1:L])$. Specifically, they are defined as follows.

1) : Let X_L^G, W_i^G, $i \in [2:L]$, be mutually independent Gaussian random vectors with covariance matrices $\boldsymbol{\Delta}_L$, $\boldsymbol{\Delta}_{i-1} - \boldsymbol{\Delta}_i$, $i \in [2:L]$, respectively. We define

$$X_i^G = X_{i+1}^G + W_{i+1}^G, \qquad i \in [1:L-1]. \tag{21}$$

It is easy to see that

$$X_i^G \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Delta}_i), \qquad i \in [1:L]. \tag{22}$$

2) : Let $N_i^G - N_{i+1}^G$, $i \in [1:L]$, be mutually independent Gaussian random vectors with covariance matrices $\boldsymbol{K}_i - \boldsymbol{K}_{i+1}$, $i \in [1:L]$, respectively.[2] We assume that $(N_i^G, i \in [1:L+1])$ is independent of $(X_i^G, i \in [1:L])$. Define

$$Y_i^G = X_i^G + N_i^G, \qquad i \in [1:L], \tag{23}$$

$$\tilde{Y}_i^G = X_{i-1}^G + N_i^G, \qquad i \in [2:L+1]. \tag{24}$$

It is clear that

$$Y_i^G \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Delta}_i + \boldsymbol{K}_i), \qquad i \in [1:L], \tag{25}$$

$$\tilde{Y}_i^G \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Delta}_{i-1} + \boldsymbol{K}_i), \qquad i \in [2:L+1]. \tag{26}$$

Using the covariance preserved transform (see, e.g., [43]), we define

$$X_{i,\gamma} = \sqrt{1-\gamma} X + \sqrt{\gamma} X_i^G, \qquad i \in [1:L], \tag{27}$$

$$Y_{i,\gamma} = \sqrt{1-\gamma} Y_i + \sqrt{\gamma} Y_i^G, \qquad i \in [1:L], \tag{28}$$

$$\tilde{Y}_{i,\gamma} = \sqrt{1-\gamma} Y_i + \sqrt{\gamma} \tilde{Y}_i^G, \qquad i \in [2:L+1],[3] \tag{29}$$

[2]Here N_{L+1}^G is a null random vector with covariance matrix $\boldsymbol{K}_{L+1} = \boldsymbol{0}$.
[3]Here $Y_{L+1} = X$ and $\tilde{Y}_{L+1,\gamma} = X_{L,\gamma}$.

$$Y^*_{i,\gamma} = \sqrt{\gamma} Y_i - \sqrt{1-\gamma} Y_i^G, \qquad i \in [1:L], \tag{30}$$

for any $\gamma \in (0,1)$.

Consider the following function:

$$\begin{aligned} g(\gamma) = \sum_{i=1}^{L-1} \Big(& \mu_i h(Y_{i,\gamma}|Y^*_{i,\gamma}, W_j, j \in [1:i]) \\ & - \mu_{i+1} h(\tilde{Y}_{i+1,\gamma}|Y^*_{i,\gamma}, W_j, j \in [1:i]) \\ & - (\mu_i - \mu_{i+1}) h(X_{i,\gamma}|Y^*_{i,\gamma}, W_j, j \in [1:i]) \Big) \\ & + \mu_L h(Y_{L,\gamma}|Y^*_{L,\gamma}, W_j, j \in [1:L]) \\ & - \mu_L h(X_{L,\gamma}|Y^*_{L,\gamma}, W_j, j \in [1:L]). \end{aligned} \tag{31}$$

Notice that $g(0)$ coincides with the left-hand side of (19) while $g(1)$ coincides with the right-hand side of (19). Therefore, it suffices to show that $g(\gamma)$ decreases monotonically along the path parameterized by γ, i.e.,

$$\frac{d}{d\gamma} g(\gamma) \le 0, \qquad \gamma \in (0,1). \tag{32}$$

Remark 3: The construction of random variable pairs $(Y_{i,\gamma}, Y^*_{i,\gamma})$, $i \in [1:L]$, is inspired by the doubling trick in [38]. Indeed, we have $(Y_{i,\gamma}, Y^*_{i,\gamma}) = (Y_i, -Y_i^G)$ when $\gamma = 0$ and $(Y_{i,\gamma}, Y^*_{i,\gamma}) = (Y_i^G, Y_i)$ when $\gamma = 1$, which implies that $Y^*_{i,\gamma}$ is independent of $Y_{i,\gamma}$ at the starting and ending points of the path. However, different from the doubling trick which only focuses on the tensorization properties at some special points, we consider a continuously parameterized tensorization process, which makes it possible to reveal the convex-like property of the associated optimization problem. The proposed perturbation method is also different from that in [26], [42] as it works in the product probability space instead of the original space. A similar construction can be found in [39].

B. Derivative of $g(\gamma)$

In this step, we utilize a vector generalization of I-MMSE relationship from [44]. First rewrite (31) as

$$\begin{aligned} g(\gamma) = \sum_{i=1}^{L-1} \Big(& \mu_i h(Y_{i,\gamma}, Y^*_{i,\gamma}|W_j, j \in [1:i]) \\ & - \mu_{i+1} h(\tilde{Y}_{i+1,\gamma}, Y^*_{i,\gamma}|W_j, j \in [1:i]) \\ & - (\mu_i - \mu_{i+1}) h(X_{i,\gamma}, Y^*_{i,\gamma}|W_j, j \in [1:i]) \Big) \\ & + \mu_L h(Y_{L,\gamma}, Y^*_{L,\gamma}|W_j, j \in [1:L]) \\ & - \mu_L h(X_{L,\gamma}, Y^*_{L,\gamma}|W_j, j \in [1:L]). \end{aligned} \tag{33}$$

In view of (28) and (30), it can be verified that

$$\begin{aligned} & h\left(Y_{i,\gamma}, Y^*_{i,\gamma} \middle| W_j, j \in [1:i]\right) \\ & = h\Big(\sqrt{1-\gamma} Y_i + \sqrt{\gamma} Y_i^G, \sqrt{\gamma} Y_i - \sqrt{1-\gamma} Y_i^G \Big| W_j, \\ & \qquad j \in [1:i]\Big) \qquad (34) \\ & = h\left(Y_i, Y_i^G \middle| W_j, j \in [1:i]\right), \qquad i \in [1:L]. \qquad (35) \end{aligned}$$

Since Y_i and Y_i^G do not depend on γ, it follows that

$$\frac{d}{d\gamma} h\left(Y_i, Y_i^G \middle| W_j, j \in [1:i]\right) = 0, \qquad i \in [1:L]. \tag{36}$$

Moreover, as shown in Appendices B and C,

$$\begin{aligned} & \frac{d}{d\gamma} h\left(X_{i,\gamma}, Y^*_{i,\gamma} \middle| W_j, j \in [1:i]\right) \\ & = \frac{1}{2(1-\gamma)} \operatorname{tr}\Big\{ (\boldsymbol{\Delta}_i^{-1} + \boldsymbol{K}_i^{-1})^{-1} \\ & \quad \Big(J\left(X_{i,\gamma} \middle| Y^*_{i,\gamma}, W_j, j \in [1:i]\right) - \boldsymbol{\Delta}_i^{-1} \Big) \Big\}, \quad i \in [1:L]. \end{aligned} \tag{37}$$

$$\begin{aligned} & \frac{d}{d\gamma} h\left(\tilde{Y}_{i+1,\gamma}, Y^*_{i,\gamma} \middle| W_j, j \in [1:i]\right) \\ & = \frac{1}{2(1-\gamma)} \operatorname{tr}\Big\{ \Big(\left(\boldsymbol{\Delta}_i^{-1} + \boldsymbol{K}_i^{-1}\right)^{-1} - \left(\boldsymbol{\Delta}_i^{-1} + \boldsymbol{K}_{i+1}^{-1}\right)^{-1} \Big) \\ & \quad \Big(\left(\boldsymbol{\Delta}_i^{-1} + \boldsymbol{K}_{i+1}^{-1}\right) \boldsymbol{K}_{i+1} J\left(\tilde{Y}_{i+1,\gamma} \middle| Y^*_{i,\gamma}, W_j, j \in [1:i]\right) \\ & \quad \boldsymbol{K}_{i+1} \left(\boldsymbol{\Delta}_i^{-1} + \boldsymbol{K}_{i+1}^{-1}\right) - \boldsymbol{\Delta}_i^{-1} \left(\boldsymbol{\Delta}_i + \boldsymbol{K}_{i+1}\right) \boldsymbol{\Delta}_i^{-1} \Big) \Big\}, \\ & \qquad i \in [1:L-1]. \end{aligned} \tag{38}$$

Combining (35), (36), (37), and (38) gives

$$\begin{aligned} & -2(1-\gamma) \frac{d}{d\gamma} g(\gamma) \\ & = \sum_{i=1}^{L-1} \operatorname{tr}\Big\{ \Big(\mu_{i+1} \left(\boldsymbol{\Delta}_i^{-1} + \boldsymbol{K}_i^{-1}\right)^{-1} - \mu_{i+1} \left(\boldsymbol{\Delta}_i^{-1} + \boldsymbol{K}_{i+1}^{-1}\right)^{-1} \Big) \\ & \quad \Big(\left(\boldsymbol{\Delta}_i^{-1} + \boldsymbol{K}_{i+1}^{-1}\right) \boldsymbol{K}_{i+1} J\left(\tilde{Y}_{i+1,\gamma} \middle| Y^*_{i,\gamma}, W_j, j \in [1:i]\right) \\ & \quad \boldsymbol{K}_{i+1} \left(\boldsymbol{\Delta}_i^{-1} + \boldsymbol{K}_{i+1}^{-1}\right) - \boldsymbol{\Delta}_i^{-1} \left(\boldsymbol{\Delta}_i + \boldsymbol{K}_{i+1}\right) \boldsymbol{\Delta}_i^{-1} \Big) \Big\} \\ & \quad + \sum_{i=1}^{L-1} \operatorname{tr}\Big\{ (\mu_i - \mu_{i+1}) (\boldsymbol{\Delta}_i^{-1} + \boldsymbol{K}_i^{-1})^{-1} \\ & \quad \Big(J\left(X_{i,\gamma} \middle| Y^*_{i,\gamma}, W_j, j \in [1:i]\right) - \boldsymbol{\Delta}_i^{-1} \Big) \Big\} \\ & \quad + \operatorname{tr}\Big\{ \mu_L (\boldsymbol{\Delta}_L^{-1} + \boldsymbol{K}_L^{-1}) \\ & \quad \Big(J\left(X_{L,\gamma} \middle| Y^*_{L,\gamma}, W_j, j \in [1:L]\right) - \boldsymbol{\Delta}_L^{-1} \Big) \Big\}. \end{aligned} \tag{39}$$

Hence, for the purpose of proving (32), it suffices to show that (39) is greater than or equal to 0.

C. Lower Bound of (39)

In this step, we establish a lower bound of (39) with the Karush-Kuhn-Tucker (KKT) conditions in (13) and (14) properly incorporated.

Define $L_1(\gamma), L_2(\gamma)$ and $L_3(\gamma)$, as shown in (40a)–(40c), at the bottom of the next page. We aim to show

$$-2(1-\gamma) \frac{d}{d\gamma} g(\gamma) \ge L_1(\gamma) + L_2(\gamma) + L_3(\gamma). \tag{41}$$

First notice that the covariance matrix of random vector

$$\begin{pmatrix} \sqrt{1-\gamma} N_{i+1} + \sqrt{\gamma} N_{i+1}^G \\ \sqrt{\gamma} N_i - \sqrt{1-\gamma} N_i^G \end{pmatrix}$$

is given by

$$\begin{pmatrix} \boldsymbol{K}_{i+1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{K}_i \end{pmatrix}. \tag{42}$$

So $\sqrt{1-\gamma} N_{i+1} + \sqrt{\gamma} N_{i+1}^G$ is independent of $\sqrt{\gamma} N_i - \sqrt{1-\gamma} N_i^G$, which, together with (30), implies that $\sqrt{1-\gamma}N_{i+1}+\sqrt{\gamma}N^G_{i+1}$ is independent of $Y^*_{i,\gamma}$ as well. For $i\in[1:L-1]$, we have

$$\tilde{Y}_{i+1,\gamma}=X_{i,\gamma}+\sqrt{1-\gamma}N_{i+1}+\sqrt{\gamma}N^G_{i+1}. \tag{43}$$

In view of the fact that $\sqrt{1-\gamma}N_{i+1}+\sqrt{\gamma}N^G_{i+1}$ is independent of $X_{i,\gamma}$, the Fisher information inequality (see Lemma 5 in Appendix A) can be invoked to show

$$\begin{aligned}
&\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_{i+1}^{-1}\right)\boldsymbol{K}_{i+1}J\left(\tilde{Y}_{i+1,\gamma}\middle|Y^*_{i,\gamma},W_j,j\in[1:i]\right)\\
&\boldsymbol{K}_{i+1}\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_{i+1}^{-1}\right)-\boldsymbol{\Delta}_i^{-1}\left(\boldsymbol{\Delta}_i+\boldsymbol{K}_{i+1}\right)\boldsymbol{\Delta}_i^{-1}\\
&=\left(\boldsymbol{I}+\boldsymbol{\Delta}_i^{-1}\boldsymbol{K}_{i+1}\right)\\
&\quad J\left(X_{i,\gamma}+\sqrt{1-\gamma}N_{i+1}+\sqrt{\gamma}N^G_{i+1}\middle|Y^*_{i,\gamma},W_j,j\in[1:i]\right)\\
&\quad\left(\boldsymbol{K}_{i+1}\boldsymbol{\Delta}_i^{-1}+\boldsymbol{I}\right)-\boldsymbol{\Delta}_i^{-1}\boldsymbol{K}_{i+1}\boldsymbol{\Delta}_i^{-1}-\boldsymbol{\Delta}_i^{-1} &(44)\\
&\preceq J\left(X_{i,\gamma}\middle|Y^*_{i,\gamma},W_j,j\in[1:i]\right)\\
&\quad+\boldsymbol{\Delta}_i^{-1}\boldsymbol{K}_{i+1}J\left(\sqrt{1-\gamma}N_{i+1}+\sqrt{\gamma}N^G_{i+1}\right)\boldsymbol{K}_{i+1}\boldsymbol{\Delta}_i^{-1}\\
&\quad-\boldsymbol{\Delta}_i^{-1}\boldsymbol{K}_{i+1}\boldsymbol{\Delta}_i^{-1}-\boldsymbol{\Delta}_i^{-1} &(45)\\
&=J\left(X_{i,\gamma}\middle|Y^*_{i,\gamma},W_j,j\in[1:i]\right)-\boldsymbol{\Delta}_i^{-1}. &(46)
\end{aligned}$$

Since $\boldsymbol{K}_i\succ\boldsymbol{K}_{i+1}$, it follows that

$$\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_i^{-1}\right)^{-1}-\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_{i+1}^{-1}\right)^{-1}\succ\boldsymbol{0}. \tag{47}$$

Therefore,

$$\begin{aligned}
&-2(1-\gamma)\frac{d}{d\gamma}g(\gamma)\\
&\geq\sum_{i=1}^{L-1}\operatorname{tr}\Big\{\left(\mu_i\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_i^{-1}\right)^{-1}-\mu_{i+1}\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_{i+1}^{-1}\right)^{-1}\right)\\
&\qquad\Big(\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_{i+1}^{-1}\right)\boldsymbol{K}_{i+1}J\left(\tilde{Y}_{i+1,\gamma}\middle|Y^*_{i,\gamma},W_j,j\in[1:i]\right)\\
&\qquad\boldsymbol{K}_{i+1}\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_{i+1}^{-1}\right)-\boldsymbol{\Delta}_i^{-1}\left(\boldsymbol{\Delta}_i+\boldsymbol{K}_{i+1}\right)\boldsymbol{\Delta}_i^{-1}\Big)\Big\}\\
&\quad-\operatorname{tr}\Big\{\mu_L(\boldsymbol{\Delta}_L^{-1}+\boldsymbol{K}_L^{-1})\Big(J\left(X_{L,\gamma}\middle|Y^*_{L,\gamma},W_j,j\in[1:L]\right)\\
&\qquad-\boldsymbol{\Delta}_L^{-1}\Big)\Big\} &(48)\\
&=\sum_{i=1}^{L-1}\operatorname{tr}\Big\{\left(\boldsymbol{\Psi}_i-\boldsymbol{\Psi}_{i+1}+\boldsymbol{\Lambda}_i\right)\Big(\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_{i+1}^{-1}\right)\boldsymbol{K}_{i+1}\\
&\qquad J\left(\tilde{Y}_{i+1,\gamma}\middle|Y^*_{i,\gamma},W_j,j\in[1:i]\right)\boldsymbol{K}_{i+1}\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_{i+1}^{-1}\right)\\
&\qquad-\boldsymbol{\Delta}_i^{-1}\left(\boldsymbol{\Delta}_i+\boldsymbol{K}_{i+1}\right)\boldsymbol{\Delta}_i^{-1}\Big)\Big\}\\
&\quad+\operatorname{tr}\Big\{\left(\boldsymbol{\Psi}_L+\boldsymbol{\Lambda}_L\right)\Big(J\left(\tilde{Y}_{L+1,\gamma}\middle|Y^*_{L,\gamma},W_j,j\in[1:L]\right)\\
&\qquad-\boldsymbol{\Delta}_L^{-1}\Big)\Big\} &(49)\\
&=\sum_{i=1}^{L-1}\operatorname{tr}\Big\{\left(\boldsymbol{\Psi}_i-\boldsymbol{\Psi}_{i+1}\right)\Big(\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_{i+1}^{-1}\right)\boldsymbol{K}_{i+1}\\
&\qquad J\left(\tilde{Y}_{i+1,\gamma}\middle|Y^*_{i,\gamma},W_j,j\in[1:i]\right)\boldsymbol{K}_{i+1}\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_{i+1}^{-1}\right)\\
&\qquad-\boldsymbol{\Delta}_i^{-1}\left(\boldsymbol{\Delta}_i+\boldsymbol{K}_{i+1}\right)\boldsymbol{\Delta}_i^{-1}\Big)\Big\}\\
&\quad+\operatorname{tr}\Big\{\boldsymbol{\Psi}_L\Big(J\left(\tilde{Y}_{L+1,\gamma}\middle|Y^*_{L,\gamma},W_j,j\in[1:L]\right)-\boldsymbol{\Delta}_L^{-1}\Big)\Big\}\\
&\quad+\sum_{i=1}^{L}\operatorname{tr}\Big\{\boldsymbol{\Lambda}_i\Big(\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_{i+1}^{-1}\right)\boldsymbol{K}_{i+1}\\
&\qquad J\left(\tilde{Y}_{i+1,\gamma}\middle|Y^*_{i,\gamma},W_j,j\in[1:i]\right)\boldsymbol{K}_{i+1}\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_{i+1}^{-1}\right)\\
&\qquad-\boldsymbol{\Delta}_i^{-1}\left(\boldsymbol{\Delta}_i+\boldsymbol{K}_{i+1}\right)\boldsymbol{\Delta}_i^{-1}\Big)\Big\} &(50)\\
&=\operatorname{tr}\Big\{\boldsymbol{\Psi}_1\Big(\left(\boldsymbol{\Delta}_1^{-1}+\boldsymbol{K}_2^{-1}\right)\boldsymbol{K}_2J\left(\tilde{Y}_{2,\gamma}\middle|Y^*_{1,\gamma},W_1\right)\\
&\qquad\boldsymbol{K}_2\left(\boldsymbol{\Delta}_1^{-1}+\boldsymbol{K}_2^{-1}\right)-\boldsymbol{\Delta}_1^{-1}\left(\boldsymbol{\Delta}_1+\boldsymbol{K}_2\right)\boldsymbol{\Delta}_1^{-1}\Big)\Big\}\\
&\quad+\sum_{i=2}^{L}\operatorname{tr}\Big\{\boldsymbol{\Psi}_i\Big(\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_{i+1}^{-1}\right)\boldsymbol{K}_{i+1}\\
&\qquad J\left(\tilde{Y}_{i+1,\gamma}\middle|Y^*_{i,\gamma},W_j,j\in[1:i]\right)\boldsymbol{K}_{i+1}\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_{i+1}^{-1}\right)\\
&\qquad-\left(\boldsymbol{\Delta}_{i-1}^{-1}+\boldsymbol{K}_i^{-1}\right)\boldsymbol{K}_iJ\Big(\tilde{Y}_{i,\gamma}\Big|Y^*_{i-1,\gamma},W_j,j\in[1:i-1]\Big)\\
&\qquad\boldsymbol{K}_i\left(\boldsymbol{\Delta}_{i-1}^{-1}+\boldsymbol{K}_i^{-1}\right)-\boldsymbol{\Delta}_i^{-1}\left(\boldsymbol{\Delta}_i+\boldsymbol{K}_{i+1}\right)\boldsymbol{\Delta}_i^{-1}\\
&\qquad+\boldsymbol{\Delta}_{i-1}^{-1}\left(\boldsymbol{\Delta}_{i-1}+\boldsymbol{K}_i\right)\boldsymbol{\Delta}_{i-1}^{-1}\Big)\Big\}\\
&\quad+\sum_{i=1}^{L}\operatorname{tr}\Big\{\boldsymbol{\Lambda}_i\Big(\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_{i+1}^{-1}\right)\boldsymbol{K}_{i+1}\\
&\qquad J\left(\tilde{Y}_{i+1,\gamma}\middle|Y^*_{i,\gamma},W_j,j\in[1:i]\right)\boldsymbol{K}_{i+1}\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_{i+1}^{-1}\right)\\
&\qquad-\boldsymbol{\Delta}_i^{-1}\left(\boldsymbol{\Delta}_i+\boldsymbol{K}_{i+1}\right)\boldsymbol{\Delta}_i^{-1}\Big)\Big\} &(51)\\
&\geq L_1(\gamma)+L_2(\gamma)+L_3(\gamma). &(52)
\end{aligned}$$

where (49) is due to the KKT properties in (13) and (14), (50) is due to the fact that $\boldsymbol{K}_{L+1}=0$, (51) follows by

$$L_1(\gamma)=\operatorname{tr}\Big\{\boldsymbol{\Psi}_1\Big(\left(\boldsymbol{\Delta}_1^{-1}+\boldsymbol{K}_2^{-1}\right)\boldsymbol{K}_2J\left(\tilde{Y}_{2,\gamma}\middle|Y^*_{1,\gamma},W_1\right)\boldsymbol{K}_2\left(\boldsymbol{\Delta}_1^{-1}+\boldsymbol{K}_2^{-1}\right)-\boldsymbol{\Delta}_1^{-1}\left(\boldsymbol{\Delta}_1+\boldsymbol{K}_2\right)\boldsymbol{\Delta}_1^{-1}\Big)\Big\} \tag{40a}$$

$$\begin{aligned}
L_2(\gamma)=\sum_{i=2}^{L}\operatorname{tr}\Big\{\boldsymbol{\Psi}_i\Big(&\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_{i+1}^{-1}\right)\boldsymbol{K}_{i+1}J\left(\tilde{Y}_{i+1,\gamma}\middle|Y^*_{i,\gamma},W_j,j\in[1:i]\right)\boldsymbol{K}_{i+1}\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_{i+1}^{-1}\right)\\
&-\left(\boldsymbol{\Delta}_{i-1}^{-1}+\boldsymbol{K}_i^{-1}\right)\boldsymbol{K}_iJ\left(\tilde{Y}_{i,\gamma}\middle|Y^*_{i-1,\gamma},W_j,j\in[1:i-1]\right)\boldsymbol{K}_i\left(\boldsymbol{\Delta}_{i-1}^{-1}+\boldsymbol{K}_i^{-1}\right)\\
&-\boldsymbol{\Delta}_i^{-1}\left(\boldsymbol{\Delta}_i+\boldsymbol{K}_{i+1}\right)\boldsymbol{\Delta}_i^{-1}+\boldsymbol{\Delta}_{i-1}^{-1}\left(\boldsymbol{\Delta}_{i-1}+\boldsymbol{K}_i\right)\boldsymbol{\Delta}_{i-1}^{-1}\Big)\Big\}
\end{aligned} \tag{40b}$$

$$\begin{aligned}
L_3(\gamma)=\sum_{i=1}^{L}\operatorname{tr}\Big\{\boldsymbol{\Lambda}_i\Big(&\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_{i+1}^{-1}\right)\boldsymbol{K}_{i+1}J\left(\tilde{Y}_{i+1,\gamma}\middle|Y^*_{i,\gamma},W_j,j\in[1:i]\right)\boldsymbol{K}_{i+1}\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_{i+1}^{-1}\right)\\
&-\boldsymbol{\Delta}_i^{-1}\left(\boldsymbol{\Delta}_i+\boldsymbol{K}_{i+1}\right)\boldsymbol{\Delta}_i^{-1}\Big)\Big\}
\end{aligned} \tag{40c}$$

algebraic manipulations, and (52) is due to the definition of $L_1(\gamma), L_2(\gamma)$ and $L_3(\gamma)$ in (40a)-(40c).

Now it suffices to show that $L_1(\gamma), L_2(\gamma)$ and $L_3(\gamma)$ are all lower bounded by 0.

D. Lower Bound of $L_1(\gamma)$

From (188) in Appendix C,

$$\begin{aligned}&\left(\boldsymbol{\Delta}_1^{-1}+\boldsymbol{K}_2^{-1}\right)\boldsymbol{K}_2 J\left(\tilde{Y}_{2,\gamma}\middle|Y_{1,\gamma}^*,W_1\right)\boldsymbol{K}_2\left(\boldsymbol{\Delta}_1^{-1}+\boldsymbol{K}_2^{-1}\right)\\&-\boldsymbol{\Delta}_1^{-1}\left(\boldsymbol{\Delta}_1+\boldsymbol{K}_2\right)\boldsymbol{\Delta}_1^{-1}\\&=\frac{1-\gamma}{\gamma}\boldsymbol{\Delta}_1^{-1}\left(\boldsymbol{\Delta}_1+\boldsymbol{K}_2\right)\left(\left(\boldsymbol{\Delta}_1+\boldsymbol{K}_2\right)^{-1}+\left(\boldsymbol{K}_1-\boldsymbol{K}_2\right)^{-1}\right)\\&\left(\left(\left(\boldsymbol{\Delta}_1+\boldsymbol{K}_2\right)^{-1}+\left(\boldsymbol{K}_1-\boldsymbol{K}_2\right)^{-1}\right)^{-1}\right.\\&\left.-\frac{1}{\gamma}\operatorname{cov}\left(Y_2\middle|\tilde{Y}_{2,\gamma},Y_{1,\gamma}^*,W_1\right)\right)\\&\left(\left(\boldsymbol{\Delta}_1+\boldsymbol{K}_2\right)^{-1}+\left(\boldsymbol{K}_1-\boldsymbol{K}_2\right)^{-1}\right)\left(\boldsymbol{\Delta}_1+\boldsymbol{K}_2\right)\boldsymbol{\Delta}_1^{-1}.\end{aligned}\tag{53}$$

Combining the data processing inequality for MMSE (see Lemma 8 in Appendix A) and (176) gives

$$\begin{aligned}&\operatorname{cov}\left(Y_2\middle|\tilde{Y}_{2,\gamma},Y_{1,\gamma}^*,W_1\right)\\&\preceq\operatorname{cov}\left(Y_2\middle|\tilde{Y}_{2,\gamma},Y_{1,\gamma}^*\right)\qquad(54)\\&=\left(\left(\boldsymbol{K}_0+\boldsymbol{K}_2\right)^{-1}+\frac{1-\gamma}{\gamma}\left(\boldsymbol{\Delta}_1+\boldsymbol{K}_2\right)^{-1}\right.\\&\qquad\left.+\frac{1}{\gamma}\left(\boldsymbol{K}_1-\boldsymbol{K}_2\right)^{-1}\right)^{-1}.\end{aligned}\tag{55}$$

Substituting (55) into (53) yields the following lower bound:

$$\begin{aligned}&\left(\boldsymbol{\Delta}_1^{-1}+\boldsymbol{K}_2^{-1}\right)\boldsymbol{K}_2 J\left(\tilde{Y}_{2,\gamma}\middle|Y_{1,\gamma}^*,W_1\right)\boldsymbol{K}_2\left(\boldsymbol{\Delta}_1^{-1}+\boldsymbol{K}_2^{-1}\right)\\&-\boldsymbol{\Delta}_1^{-1}\left(\boldsymbol{\Delta}_1+\boldsymbol{K}_2\right)\boldsymbol{\Delta}_1^{-1}\\&\succeq\frac{(1-\gamma)^2}{\gamma^2}\boldsymbol{\Delta}_1^{-1}\left(\boldsymbol{\Delta}_1+\boldsymbol{K}_2\right)\\&\left(\left(\boldsymbol{\Delta}_1+\boldsymbol{K}_2\right)^{-1}+\left(\boldsymbol{K}_1-\boldsymbol{K}_2\right)^{-1}\right)\\&\left(\frac{\gamma}{1-\gamma}\left(\left(\boldsymbol{\Delta}_1+\boldsymbol{K}_2\right)^{-1}+\left(\boldsymbol{K}_1-\boldsymbol{K}_2\right)^{-1}\right)^{-1}\right.\\&-\frac{1}{1-\gamma}\left(\left(\boldsymbol{K}_0+\boldsymbol{K}_2\right)^{-1}+\frac{1-\gamma}{\gamma}\left(\boldsymbol{\Delta}_1+\boldsymbol{K}_2\right)^{-1}\right.\\&\left.\left.+\frac{1}{\gamma}\left(\boldsymbol{K}_1-\boldsymbol{K}_2\right)^{-1}\right)^{-1}\right)\\&\left(\left(\boldsymbol{\Delta}_1+\boldsymbol{K}_2\right)^{-1}+\left(\boldsymbol{K}_1-\boldsymbol{K}_2\right)^{-1}\right)\left(\boldsymbol{\Delta}_1+\boldsymbol{K}_2\right)\boldsymbol{\Delta}_1^{-1}\end{aligned}\tag{56}$$

$$\begin{aligned}&=\frac{1-\gamma}{\gamma}\boldsymbol{\Delta}_1^{-1}\left(\boldsymbol{\Delta}_1+\boldsymbol{K}_2\right)\left(\left(\boldsymbol{\Delta}_1+\boldsymbol{K}_2\right)^{-1}+\left(\boldsymbol{K}_1-\boldsymbol{K}_2\right)^{-1}\right)\\&\left(\left(\boldsymbol{K}_0+\boldsymbol{K}_2\right)^{-1}+\frac{1-\gamma}{\gamma}\left(\boldsymbol{\Delta}_1+\boldsymbol{K}_2\right)^{-1}\right.\\&\left.+\frac{1}{\gamma}\left(\boldsymbol{K}_1-\boldsymbol{K}_2\right)^{-1}\right)^{-1}\left(\left(\boldsymbol{K}_0+\boldsymbol{K}_2\right)^{-1}-\left(\boldsymbol{\Delta}_1+\boldsymbol{K}_2\right)^{-1}\right)\\&\left(\boldsymbol{\Delta}_1+\boldsymbol{K}_2\right)\boldsymbol{\Delta}_1^{-1}\end{aligned}\tag{57}$$

$$\begin{aligned}&=\frac{1-\gamma}{\gamma}\boldsymbol{\Delta}_1^{-1}\left(\boldsymbol{\Delta}_1+\boldsymbol{K}_2\right)\left(\left(\boldsymbol{\Delta}_1+\boldsymbol{K}_2\right)^{-1}+\left(\boldsymbol{K}_1-\boldsymbol{K}_2\right)^{-1}\right)\\&\left(\left(\boldsymbol{K}_0+\boldsymbol{K}_2\right)^{-1}+\frac{1-\gamma}{\gamma}\left(\boldsymbol{\Delta}_1+\boldsymbol{K}_2\right)^{-1}\right.\\&\left.+\frac{1}{\gamma}\left(\boldsymbol{K}_1-\boldsymbol{K}_2\right)^{-1}\right)^{-1}\left(\boldsymbol{K}_0+\boldsymbol{K}_2\right)^{-1}\left(\boldsymbol{\Delta}_1-\boldsymbol{K}_0\right)\boldsymbol{\Delta}_1^{-1}\end{aligned}\tag{58}$$

$$\begin{aligned}&=\frac{1-\gamma}{\gamma}\boldsymbol{\Delta}_1^{-1}\left(\boldsymbol{\Delta}_1+\boldsymbol{K}_2\right)\left(\left(\boldsymbol{\Delta}_1+\boldsymbol{K}_2\right)^{-1}+\left(\boldsymbol{K}_1-\boldsymbol{K}_2\right)^{-1}\right)\\&\left(\left(\boldsymbol{K}_0+\boldsymbol{K}_2\right)^{-1}+\frac{1-\gamma}{\gamma}\left(\boldsymbol{\Delta}_1+\boldsymbol{K}_2\right)^{-1}\right.\\&\left.+\frac{1}{\gamma}\left(\boldsymbol{K}_1-\boldsymbol{K}_2\right)^{-1}\right)^{-1}\left(\boldsymbol{K}_0+\boldsymbol{K}_2\right)^{-1}\boldsymbol{K}_0^{-1}\\&\left(\boldsymbol{K}_0^{-1}-\boldsymbol{\Delta}_1^{-1}\right).\end{aligned}\tag{59}$$

From the complementary slackness condition in (15), i.e.,

$$\boldsymbol{B}_1^*\boldsymbol{\Psi}_1=\left(\boldsymbol{K}_0^{-1}-\boldsymbol{\Delta}_1^{-1}\right)\boldsymbol{\Psi}_1=\boldsymbol{0},\tag{60}$$

we have

$$\begin{aligned}&\operatorname{tr}\left\{\boldsymbol{\Psi}_1\left(\left(\boldsymbol{\Delta}_1^{-1}+\boldsymbol{K}_2^{-1}\right)\boldsymbol{K}_2 J\left(\tilde{Y}_{2,\gamma}\middle|Y_{1,\gamma}^*,W_1\right)\right.\right.\\&\qquad\left.\left.\boldsymbol{K}_2\left(\boldsymbol{\Delta}_1^{-1}+\boldsymbol{K}_2^{-1}\right)-\boldsymbol{\Delta}_1^{-1}\left(\boldsymbol{\Delta}_1+\boldsymbol{K}_2\right)\boldsymbol{\Delta}_1^{-1}\right)\right\}\qquad(61)\\&\geq\operatorname{tr}\left\{\frac{1-\gamma}{\gamma}\boldsymbol{\Delta}_1^{-1}\left(\boldsymbol{\Delta}_1+\boldsymbol{K}_2\right)\right.\\&\qquad\left(\left(\boldsymbol{\Delta}_1+\boldsymbol{K}_2\right)^{-1}+\left(\boldsymbol{K}_1-\boldsymbol{K}_2\right)^{-1}\right)\left(\left(\boldsymbol{K}_0+\boldsymbol{K}_2\right)^{-1}\right.\\&\qquad\left.+\frac{1-\gamma}{\gamma}\left(\boldsymbol{\Delta}_1+\boldsymbol{K}_2\right)^{-1}+\frac{1}{\gamma}\left(\boldsymbol{K}_1-\boldsymbol{K}_2\right)^{-1}\right)^{-1}\\&\qquad\left.\left(\boldsymbol{K}_0+\boldsymbol{K}_2\right)^{-1}\boldsymbol{K}_0^{-1}\left(\boldsymbol{K}_0^{-1}-\boldsymbol{\Delta}_1^{-1}\right)\boldsymbol{\Psi}_1\right\}\\&=0.\end{aligned}\tag{62}$$

This proves that $L_1(\gamma)$ is lower bounded by 0.

E. Lower Bound of $L_2(\gamma)$

To the end of showing that (40b) is lower bounded by 0, we introduce

$$N'_{i+1}\triangleq\sqrt{1-\gamma}\left(N_i-N_{i+1}\right)+\sqrt{\gamma}\left(N_i^G-N_{i+1}^G\right),\quad i\in[1:L].\tag{63}$$

Note that N'_{i+1} is a Gaussian random vector with covariance matrix $\boldsymbol{K}_i-\boldsymbol{K}_{i+1}$ and is independent of $(\tilde{Y}_{i+1,\gamma},Y_{i,\gamma}^*)$. Moreover,

$$\tilde{Y}_{i,\gamma}=\tilde{Y}_{i+1,\gamma}+N'_{i+1},\qquad i\in[2:L].\tag{64}$$

In view of the fact that N'_{i+1} is independent of $Y_{i,\gamma}^*$, we can invoke the Fisher information inequality (see Lemma 5 in Appendix A) to show

$$\begin{aligned}&\left(\boldsymbol{\Delta}_{i-1}^{-1}+\boldsymbol{K}_i^{-1}\right)\boldsymbol{K}_i J\left(\tilde{Y}_{i,\gamma}\middle|Y_{i-1,\gamma}^*,W_j,j\in[1:i-1]\right)\\&\boldsymbol{K}_i\left(\boldsymbol{K}_i^{-1}+\boldsymbol{\Delta}_{i-1}^{-1}\right)\\&=\left(\boldsymbol{\Delta}_{i-1}^{-1}\boldsymbol{K}_i+\boldsymbol{I}\right)J\left(\tilde{Y}_{i+1,\gamma}+N'_{i+1}\middle|Y_{i-1,\gamma}^*,W_j,j\in[1:i-1]\right)\\&\quad\left(\boldsymbol{I}+\boldsymbol{K}_i\boldsymbol{\Delta}_{i-1}^{-1}\right)\\&\preceq\left(\boldsymbol{\Delta}_{i-1}^{-1}\boldsymbol{K}_{i+1}+\boldsymbol{I}\right)J\left(\tilde{Y}_{i+1,\gamma}\middle|Y_{i-1,\gamma}^*,W_j,j\in[1:i-1]\right)\end{aligned}$$

$$\begin{aligned}&\left(\boldsymbol{I}+\boldsymbol{K}_{i+1}\boldsymbol{\Delta}_{i-1}^{-1}\right)+\boldsymbol{\Delta}_{i-1}^{-1}\left(\boldsymbol{K}_i-\boldsymbol{K}_{i+1}\right)\boldsymbol{\Delta}_{i-1}^{-1}\\&\preceq\left(\boldsymbol{\Delta}_{i-1}^{-1}\boldsymbol{K}_{i+1}+\boldsymbol{I}\right)J\left(\tilde{Y}_{i+1,\gamma}\middle|Y_{i,\gamma}^*,W_j,j\in[1:i]\right)\\&\quad\left(\boldsymbol{I}+\boldsymbol{K}_{i+1}\boldsymbol{\Delta}_{i-1}^{-1}\right)+\boldsymbol{\Delta}_{i-1}^{-1}\left(\boldsymbol{K}_i-\boldsymbol{K}_{i+1}\right)\boldsymbol{\Delta}_{i-1}^{-1}\qquad(65)\\&=\left(\boldsymbol{\Delta}_{i-1}^{-1}+\boldsymbol{K}_{i+1}^{-1}\right)\boldsymbol{K}_{i+1}J\left(\tilde{Y}_{i+1,\gamma}\middle|Y_{i,\gamma}^*,W_j,j\in[1:i]\right)\\&\quad\boldsymbol{K}_{i+1}\left(\boldsymbol{K}_{i+1}^{-1}+\boldsymbol{\Delta}_{i-1}^{-1}\right)+\boldsymbol{\Delta}_{i-1}^{-1}\left(\boldsymbol{K}_i-\boldsymbol{K}_{i+1}\right)\boldsymbol{\Delta}_{i-1}^{-1}.\qquad(66)\end{aligned}$$

In particular, (65) can be verified as follows. Let

$$N_i^*=\sqrt{\gamma}(N_{i-1}-N_i)-\sqrt{1-\gamma}(N_{i-1}^G-N_i^G),\qquad i\in[2:L].$$

Notice that N_i^* is a Gaussian random vector with covariance matrix $\boldsymbol{K}_{i-1}-\boldsymbol{K}_i$ and is independent of $(X,\tilde{Y}_{i+1,\gamma},Y_{i,\gamma}^*)$. It is easy to check

$$Y_{i-1,\gamma}^*=Y_{i,\gamma}^*+N_i^*,\qquad i\in[2:L].$$

From the data processing inequality for Fisher information (see Lemma 7 in Appendix A), we have

$$\begin{aligned}&J\left(\tilde{Y}_{i+1,\gamma}\middle|Y_{i-1,\gamma}^*,W_j,j\in[1:i]\right)\\&=J\left(\tilde{Y}_{i+1,\gamma}\middle|Y_{i,\gamma}^*+N_i^*,W_j,j\in[1:i]\right)\\&\preceq J\left(\tilde{Y}_{i+1,\gamma}\middle|Y_{i,\gamma}^*,W_j,j\in[1:i]\right),\end{aligned}$$

which, together with the matrix inequality $U^T(A-B)U\succeq 0$ for $A\succeq B$, gives (65).

Meanwhile, due to the complementary slackness condition in (15), i.e.,

$$\boldsymbol{B}_i^*\boldsymbol{\Psi}_i=\left(\boldsymbol{\Delta}_i^{-1}-\boldsymbol{\Delta}_{i-1}^{-1}\right)\boldsymbol{\Psi}_i=\boldsymbol{0},\qquad i\in[2:L],\qquad(67)$$

we have

$$\begin{aligned}&\operatorname{tr}\Big\{\boldsymbol{\Psi}_i\Big(\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_{i+1}^{-1}\right)\boldsymbol{K}_{i+1}J\left(\tilde{Y}_{i+1,\gamma}\middle|Y_{i,\gamma}^*,W_j,j\in[1:i]\right)\\&\quad\boldsymbol{K}_{i+1}\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_{i+1}^{-1}\right)-\left(\boldsymbol{\Delta}_{i-1}^{-1}+\boldsymbol{K}_i^{-1}\right)\boldsymbol{K}_i\\&\quad J\left(\tilde{Y}_{i,\gamma}\middle|Y_{i-1,\gamma}^*,W_j,j\in[1:i-1]\right)\boldsymbol{K}_i\left(\boldsymbol{\Delta}_{i-1}^{-1}+\boldsymbol{K}_i^{-1}\right)\\&\quad-\boldsymbol{\Delta}_i^{-1}\left(\boldsymbol{\Delta}_i+\boldsymbol{K}_{i+1}\right)\boldsymbol{\Delta}_i^{-1}+\boldsymbol{\Delta}_{i-1}^{-1}\left(\boldsymbol{\Delta}_{i-1}+\boldsymbol{K}_i\right)\boldsymbol{\Delta}_{i-1}^{-1}\Big)\Big\}\\&=\operatorname{tr}\Big\{\boldsymbol{\Psi}_i\Big(\left(\boldsymbol{\Delta}_{i-1}^{-1}+\boldsymbol{K}_{i+1}^{-1}\right)\boldsymbol{K}_{i+1}\\&\quad J\left(\tilde{Y}_{i+1,\gamma}\middle|Y_{i,\gamma}^*,W_j,j\in[1:i]\right)\boldsymbol{K}_{i+1}\left(\boldsymbol{\Delta}_{i-1}^{-1}+\boldsymbol{K}_{i+1}^{-1}\right)\\&\quad-\left(\boldsymbol{\Delta}_{i-1}^{-1}+\boldsymbol{K}_i^{-1}\right)\boldsymbol{K}_iJ\left(\tilde{Y}_{i,\gamma}\middle|Y_{i-1,\gamma}^*,W_j,j\in[1:i-1]\right)\\&\quad\boldsymbol{K}_i\left(\boldsymbol{\Delta}_{i-1}^{-1}+\boldsymbol{K}_i^{-1}\right)+\boldsymbol{\Delta}_{i-1}^{-1}\left(\boldsymbol{K}_i-\boldsymbol{K}_{i+1}\right)\boldsymbol{\Delta}_{i-1}^{-1}\Big)\Big\}\\&\geq 0,\qquad i\in[2:L].\qquad(68)\end{aligned}$$

This proves that $L_2(\gamma)$ is lower bounded by 0.

F. Lower Bound of $L_3(\gamma)$

To the end of showing that $L_3(\gamma)$ is lower bounded by 0, we introduce

$$N''_{i+1}\triangleq\sqrt{\gamma}\left(N_i-N_{i+1}\right)-\sqrt{1-\gamma}\left(N_i^G-N_{i+1}^G\right),\qquad i\in[1:L].\qquad(69)$$

Note that N''_{i+1} is a Gaussian random vector with covariance matrix $\boldsymbol{K}_i-\boldsymbol{K}_{i+1}$ and is independent of $(Y_{i+1},\tilde{Y}_{i+1}^G)$. It can be verified that

$$\begin{aligned}&\operatorname{cov}\left(Y_{i+1}\middle|\tilde{Y}_{i+1,\gamma},Y_{i,\gamma}^*,W_j,j\in[1:i]\right)\\&=\operatorname{cov}\left(Y_{i+1}\middle|\sqrt{1-\gamma}\tilde{Y}_{i+1,\gamma}+\sqrt{\gamma}Y_{i,\gamma}^*,\tilde{Y}_{i+1,\gamma},W_j,j\in[1:i]\right)\qquad(70)\\&=\operatorname{cov}\Big(Y_{i+1}\Big|(1-\gamma)Y_{i+1}+\sqrt{\gamma(1-\gamma)}\tilde{Y}_{i+1}^G+\gamma Y_i\\&\quad-\sqrt{\gamma(1-\gamma)}Y_i^G,\tilde{Y}_{i+1,\gamma},W_j,j\in[1:i]\Big)\qquad(71)\\&=\operatorname{cov}\Big(Y_{i+1}\Big|Y_{i+1}+\sqrt{\gamma}N''_{i+1},Y_{i+1}+\sqrt{\frac{\gamma}{1-\gamma}}\tilde{Y}_{i+1}^G,\\&\quad W_j,j\in[1:i]\Big)\qquad(72)\\&\preceq\operatorname{cov}\Big(Y_{i+1}\Big|\left(\frac{1-\gamma}{\gamma}\left(\boldsymbol{\Delta}_i+\boldsymbol{K}_{i+1}\right)^{-1}+\frac{1}{\gamma}\left(\boldsymbol{K}_i-\boldsymbol{K}_{i+1}\right)^{-1}\right)\\&\quad Y_{i+1}+\sqrt{\frac{1-\gamma}{\gamma}}\left(\boldsymbol{\Delta}_i+\boldsymbol{K}_{i+1}\right)^{-1}\tilde{Y}_{i+1}^G\\&\quad+\sqrt{\frac{1}{\gamma}}\left(\boldsymbol{K}_i-\boldsymbol{K}_{i+1}\right)^{-1}N''_{i+1},W_j,j\in[1:i]\Big).\qquad(73)\end{aligned}$$

where (73) is due to the data processing inequality for MMSE (see Lemma 8 in Appendix A).

For the sake of simplifying notation, we introduce

$$\begin{aligned}\boldsymbol{P}_{i+1}&\triangleq\left(\frac{1-\gamma}{\gamma}\left(\boldsymbol{\Delta}_i+\boldsymbol{K}_{i+1}\right)^{-1}+\frac{1}{\gamma}\left(\boldsymbol{K}_i-\boldsymbol{K}_{i+1}\right)^{-1}\right)^{-1},\\S_{i+1}^G&\triangleq\boldsymbol{P}_{i+1}\left(\sqrt{\frac{1-\gamma}{\gamma}}\left(\boldsymbol{\Delta}_i+\boldsymbol{K}_{i+1}\right)^{-1}\tilde{Y}_{i+1}^G\right.\\&\quad\left.+\sqrt{\frac{1}{\gamma}}\left(\boldsymbol{K}_i-\boldsymbol{K}_{i+1}\right)^{-1}N''_{i+1}\right).\qquad(74)\end{aligned}$$

Now (73) can be rewritten as follows

$$\begin{aligned}&\operatorname{cov}\left(Y_{i+1}\middle|\tilde{Y}_{i+1,\gamma},Y_{i,\gamma}^*,W_j,j\in[1:i]\right)^{-1}\\&\succeq\operatorname{cov}\left(Y_{i+1}\middle|Y_{i+1}+S_{i+1}^G,W_j,j\in[1:i]\right)^{-1}.\end{aligned}$$

By the theory of linear MMSE estimation, it can be verified that

$$N_i-N_{i+1}=S_{i+1}^G+T_{i+1}^G,$$

where T_{i+1}^G is a Gaussian random vector with covariance matrix $\boldsymbol{K}_i-\boldsymbol{K}_{i+1}-\boldsymbol{P}_{i+1}$ and is independent of S_{i+1}^G. We can invoke Lemma 6 in Appendix A to show that

$$\begin{aligned}&\operatorname{cov}\left(Y_{i+1}\middle|Y_{i+1}+S_{i+1}^G,W_j,j\in[1:i]\right)^{-1}\\&\succeq\operatorname{cov}\left(Y_{i+1}\middle|Y_{i+1}+S_{i+1}^G+T_{i+1}^G,W_j,j\in[1:i]\right)^{-1}\\&\quad-\left(\boldsymbol{K}_i-\boldsymbol{K}_{i+1}\right)^{-1}+\boldsymbol{P}_{i+1}^{-1}\\&=\operatorname{cov}\left(Y_{i+1}\middle|Y_i,W_j,j\in[1:i]\right)^{-1}\\&\quad+\frac{1-\gamma}{\gamma}\left(\left(\boldsymbol{\Delta}_i+\boldsymbol{K}_{i+1}\right)^{-1}+\left(\boldsymbol{K}_i-\boldsymbol{K}_{i+1}\right)^{-1}\right),\qquad(75)\end{aligned}$$

where (75) follows by the definition of $\boldsymbol{P}_{i+1}$ in (74). We then bound the two terms in (75) separately.

1) : Note that the following Markov chain condition holds:

$$(W_j, j \in [1:i]) \to X \to Y_{i+1} \to Y_i. \tag{76}$$

Since X, Y_i, and Y_{i+1}) are jointly Gaussian, it follows that

$$\begin{aligned}&\mathbb{E}\left[Y_{i+1}|X, Y_i\right] &(77)\\ &= (\boldsymbol{K}_i - \boldsymbol{K}_{i+1})\boldsymbol{K}_i^{-1}X + \boldsymbol{K}_{i+1}\boldsymbol{K}_i^{-1}Y_i. &(78)\end{aligned}$$

Furthermore, we have

$$Y_{i+1} = (\boldsymbol{K}_i - \boldsymbol{K}_{i+1})\boldsymbol{K}_i^{-1}\left(X + \tilde{N}_{i+1}\right) + \boldsymbol{K}_{i+1}\boldsymbol{K}_i^{-1}Y_i, \tag{79}$$

where $\tilde{N}_{i+1}$ is a zero-mean Gaussian random vector with covariance matrix

$$\tilde{\boldsymbol{K}}_{i+1} = \left(\boldsymbol{K}_{i+1}^{-1} - \boldsymbol{K}_i^{-1}\right)^{-1} \succ \boldsymbol{0}, \tag{80}$$

and is independent of (X, Y_i). Therefore,

$$\begin{aligned}&\operatorname{cov}\left(Y_{i+1}\middle|Y_i, W_j, j \in [1:i]\right)\\ &= \operatorname{cov}\left((\boldsymbol{K}_i - \boldsymbol{K}_{i+1})\boldsymbol{K}_i^{-1}\left(X+\tilde{N}_{i+1}\right)\middle|Y_i, W_j, j \in [1:i]\right)\\ &\preceq (\boldsymbol{K}_i - \boldsymbol{K}_{i+1})\boldsymbol{K}_i^{-1}\left(\boldsymbol{D}_i + \left(\boldsymbol{K}_{i+1}^{-1} - \boldsymbol{K}_i^{-1}\right)^{-1}\right)\boldsymbol{K}_i^{-1}\\ &\quad(\boldsymbol{K}_i - \boldsymbol{K}_{i+1}), &(81)\end{aligned}$$

where (81) is because of covariance distortion constraint $\operatorname{cov}(X|Y_i, W_j, j \in [1:i]) \preceq \boldsymbol{D}_i$ in (18).

2) : It can be verified that

$$\begin{aligned}&\left((\boldsymbol{\Delta}_i + \boldsymbol{K}_{i+1})^{-1} + (\boldsymbol{K}_i - \boldsymbol{K}_{i+1})^{-1}\right)^{-1}\\ &= \left(\boldsymbol{K}_{i+1}^{-1} - (\boldsymbol{K}_{i+1} - \boldsymbol{K}_i)^{-1} - \boldsymbol{K}_{i+1}^{-1} + (\boldsymbol{\Delta}_i+\boldsymbol{K}_{i+1})^{-1}\right)^{-1}\\ &= \boldsymbol{K}_{i+1}\left(\left(\boldsymbol{K}_{i+1}^{-1} - \boldsymbol{K}_i^{-1}\right)^{-1} - \left(\boldsymbol{\Delta}_i^{-1} + \boldsymbol{K}_{i+1}^{-1}\right)^{-1}\right)^{-1}\boldsymbol{K}_{i+1} &(82)\\ &= \boldsymbol{K}_{i+1}\left(\boldsymbol{K}_{i+1}^{-1} - \boldsymbol{K}_i^{-1}\right)\\ &\quad\left(\left(\boldsymbol{\Delta}_i^{-1} + \boldsymbol{K}_i^{-1}\right)^{-1} + \left(\boldsymbol{K}_{i+1}^{-1} - \boldsymbol{K}_i^{-1}\right)^{-1}\right)\\ &\quad\left(\boldsymbol{K}_{i+1}^{-1} - \boldsymbol{K}_i^{-1}\right)\boldsymbol{K}_{i+1} &(83)\\ &\preceq (\boldsymbol{K}_i - \boldsymbol{K}_{i+1})\boldsymbol{K}_i^{-1}\left(\boldsymbol{D}_i + \left(\boldsymbol{K}_{i+1}^{-1} - \boldsymbol{K}_i^{-1}\right)^{-1}\right)\boldsymbol{K}_i^{-1}\\ &\quad(\boldsymbol{K}_i - \boldsymbol{K}_{i+1}), &(84)\end{aligned}$$

where (82) follows by the matrix inversion identities

$$\begin{aligned}\boldsymbol{K}_{i+1}^{-1} - (\boldsymbol{K}_{i+1} - \boldsymbol{K}_i)^{-1} &= \boldsymbol{K}_{i+1}\left(\boldsymbol{K}_{i+1}^{-1} - \boldsymbol{K}_i^{-1}\right)\boldsymbol{K}_{i+1},\\ \boldsymbol{K}_{i+1}^{-1} - (\boldsymbol{\Delta}_i + \boldsymbol{K}_{i+1})^{-1} &= \boldsymbol{K}_{i+1}\left(\boldsymbol{\Delta}_i + \boldsymbol{K}_{i+1}^{-1}\right)\boldsymbol{K}_{i+1},\end{aligned}$$

(83) follows by the matrix inversion identity

$$\begin{aligned}&\left(\left(\boldsymbol{K}_{i+1}^{-1} - \boldsymbol{K}_i^{-1}\right)^{-1} - \left(\boldsymbol{\Delta}_i^{-1} + \boldsymbol{K}_{i+1}^{-1}\right)^{-1}\right)^{-1}\\ &= \left(\boldsymbol{K}_{i+1}^{-1} - \boldsymbol{K}_i^{-1}\right)\left(\left(\boldsymbol{\Delta}_i^{-1} + \boldsymbol{K}_i^{-1}\right)^{-1} + \left(\boldsymbol{K}_{i+1}^{-1} - \boldsymbol{K}_i^{-1}\right)^{-1}\right)\\ &\quad\left(\boldsymbol{K}_{i+1}^{-1} - \boldsymbol{K}_i^{-1}\right),\end{aligned}$$

and (84) is because of $\boldsymbol{\Delta}_i^{-1} + \boldsymbol{K}_i^{-1} \succeq \boldsymbol{D}_i^{-1}$ in (12).

Substituting (81) and (84) into (75) yields

$$\begin{aligned}&\operatorname{cov}\left(Y_{i+1}\middle|\tilde{Y}_{i+1,\gamma}, Y_{i,\gamma}^*, W_j, j \in [1:i]\right)\\ &\preceq \operatorname{cov}\left(Y_{i+1}\middle|Y_{i+1} + S_{i+1}^G, W_j, j \in [1:i]\right)\\ &\preceq \left(\operatorname{cov}\left(Y_{i+1}\middle|Y_i, W_j, j \in [1:i]\right)^{-1}\right.\\ &\quad\left.+\frac{1-\gamma}{\gamma}\left((\boldsymbol{\Delta}_i + \boldsymbol{K}_{i+1})^{-1} + (\boldsymbol{K}_i - \boldsymbol{K}_{i+1})^{-1}\right)\right)^{-1}\\ &\preceq \gamma(\boldsymbol{K}_i - \boldsymbol{K}_{i+1})\boldsymbol{K}_i^{-1}\left(\boldsymbol{D}_i + \left(\boldsymbol{K}_{i+1}^{-1} - \boldsymbol{K}_i^{-1}\right)^{-1}\right)\boldsymbol{K}_i^{-1}\\ &\quad(\boldsymbol{K}_i - \boldsymbol{K}_{i+1}).\end{aligned}$$

In view of (188), we have

$$\begin{aligned}&\left(\boldsymbol{\Delta}_i^{-1} + \boldsymbol{K}_{i+1}^{-1}\right)\boldsymbol{K}_{i+1}J\left(\tilde{Y}_{i+1,\gamma}\middle|Y_{i,\gamma}^*, W_j, j \in [1:i]\right)\\ &\boldsymbol{K}_{i+1}\left(\boldsymbol{\Delta}_i^{-1} + \boldsymbol{K}_{i+1}^{-1}\right) - \boldsymbol{\Delta}_i^{-1}(\boldsymbol{\Delta}_i + \boldsymbol{K}_{i+1})\boldsymbol{\Delta}_i^{-1}\\ &= \boldsymbol{\Delta}_i^{-1}(\boldsymbol{\Delta}_i + \boldsymbol{K}_{i+1})\left(J\left(\tilde{Y}_{i+1,\gamma}\middle|Y_{i,\gamma}^*, W_j, j \in [1:i]\right)\right.\\ &\quad\left.-(\boldsymbol{\Delta}_i + \boldsymbol{K}_{i+1})^{-1}\right)(\boldsymbol{\Delta}_i + \boldsymbol{K}_{i+1})\boldsymbol{\Delta}_i^{-1} &(85)\\ &= \frac{1-\gamma}{\gamma}\boldsymbol{\Delta}_i^{-1}(\boldsymbol{\Delta}_i + \boldsymbol{K}_{i+1})\\ &\quad\left((\boldsymbol{\Delta}_i + \boldsymbol{K}_{i+1})^{-1} + (\boldsymbol{K}_i - \boldsymbol{K}_{i+1})^{-1}\right)\\ &\quad\left(\left((\boldsymbol{\Delta}_i + \boldsymbol{K}_{i+1})^{-1} + (\boldsymbol{K}_i - \boldsymbol{K}_{i+1})^{-1}\right)^{-1}\right.\\ &\quad\left.-\frac{1}{\gamma}\operatorname{cov}\left(Y_{i+1}\middle|\tilde{Y}_{i+1,\gamma}, Y_{i,\gamma}^*, W_j, j \in [1:i]\right)\right)\\ &\quad\left((\boldsymbol{\Delta}_i + \boldsymbol{K}_{i+1})^{-1} + (\boldsymbol{K}_i - \boldsymbol{K}_{i+1})^{-1}\right)\\ &\quad(\boldsymbol{\Delta}_i + \boldsymbol{K}_{i+1})\boldsymbol{\Delta}_i^{-1} &(86)\\ &\succeq \frac{1-\gamma}{\gamma}\left(\boldsymbol{\Delta}_i^{-1} + \boldsymbol{K}_i^{-1}\right)\left(\left(\boldsymbol{\Delta}_i^{-1} + \boldsymbol{K}_i^{-1}\right)^{-1} - \boldsymbol{D}_i\right)\\ &\quad\left(\boldsymbol{\Delta}_i^{-1} + \boldsymbol{K}_i^{-1}\right) &(87)\\ &= \frac{1-\gamma}{\gamma}\left(\boldsymbol{\Delta}_i^{-1} + \boldsymbol{K}_i^{-1}\right)\boldsymbol{D}_i\left(\boldsymbol{D}_i^{-1} - \boldsymbol{\Delta}_i^{-1} - \boldsymbol{K}_i^{-1}\right). &(88)\end{aligned}$$

From the complementary slackness condition in (16), i.e.,

$$\left(\boldsymbol{\Delta}_i^{-1} + \boldsymbol{K}_i^{-1} - \boldsymbol{D}_i^{-1}\right)\boldsymbol{\Lambda}_i = \boldsymbol{0}, \qquad i \in [1:L], \tag{89}$$

we have

$$\begin{aligned}&\operatorname{tr}\left\{\boldsymbol{\Lambda}_i\left(\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_{i+1}^{-1}\right)\boldsymbol{K}_{i+1}J\left(\tilde{Y}_{i+1,\gamma}\middle|Y_{i,\gamma}^*, W_j, j \in [1:i]\right)\right.\right.\\ &\quad\left.\left.\boldsymbol{K}_{i+1}\left(\boldsymbol{\Delta}_i^{-1} + \boldsymbol{K}_{i+1}^{-1}\right) - \frac{1}{\gamma}\boldsymbol{\Delta}_i^{-1}(\boldsymbol{\Delta}_i + \boldsymbol{K}_{i+1})\boldsymbol{\Delta}_i^{-1}\right)\right\}\\ &\geq \operatorname{tr}\left\{\frac{1-\gamma}{\gamma}\boldsymbol{\Lambda}_i\left(\boldsymbol{\Delta}_i^{-1} + \boldsymbol{K}_i^{-1}\right)\boldsymbol{D}_i\left(\boldsymbol{D}_i^{-1} - \boldsymbol{\Delta}_i^{-1} - \boldsymbol{K}_i^{-1}\right)\right\} &(90)\\ &= 0, \qquad i \in [1:L]. &(91)\end{aligned}$$

This proves that $L_3(\gamma)$ is lower bounded by 0.

IV. Proof of Theorem 1

The proof of Theorem 1 is divided into three steps. We first adapt the argument in [9], [10] to show that every rate tuple in $\mathcal{R}(\boldsymbol{D}_i, i \in [1:L])$ is achievable, i.e., $\mathcal{R}(\boldsymbol{D}_i, i \in [1:L]) \subseteq \mathcal{R}^*(\boldsymbol{D}_i, i \in [1:L])$. We then study the supporting hyperplanes of $\mathcal{R}(\boldsymbol{D}_i, i \in [1:L])$ and characterize the optimal solution of the relevant minimization problem via

KKT analysis. Finally we derive a matching converse by leveraging the extremal inequality in Theorem 2.

A. Achievability

It is easy to adapt the achevability argument in [9], [10] to prove the following result.

Lemma 1: $(R_i, i \in [1:L]) \in \mathcal{R}^*(\boldsymbol{D}_i, i \in [1:L])$ if there exist auxiliary random vectors $(W_i, i \in [1:L])$ jointly Gaussian with $(X, Y_i, i \in [1:L])$ satisfying

- the Markov chain constraint

$$(W_i, i \in [1:L]) \to X \to Y_L \to Y_{L-1} \to \ldots \to Y_1, \tag{92}$$

- the rate constraints

$$R_1 \geq I(X;W_1|Y_1), \tag{93}$$

$$\sum_{j=1}^{i} R_j \geq I(X;W_1|Y_1) + \sum_{j=2}^{i} I(X;W_j|W_{j-1},\ldots,W_1,Y_j), \quad i \in [2:L], \tag{94}$$

- the covariance distortion constraints

$$\mathrm{cov}(X|Y_i, W_j, j \in [1:i]) \preceq \boldsymbol{D}_i, \qquad i \in [1:L]. \tag{95}$$

Equipped with Lemma 1, we proceed to show that every rate tuple in $\mathcal{R}(\boldsymbol{D}_i, i \in [1:L])$ is achievable. First choose auxiliary Gaussian random vectors $(W_i, i \in [1:L])$ such that

$$\mathrm{cov}(X|W_j, j \in [1:i]) = \left(\boldsymbol{K}_0^{-1} + \sum_{j=1}^{i} \boldsymbol{B}_j\right)^{-1}, \quad i \in [1:L]. \tag{96}$$

It can be verified that

$$h(X|Y_i, W_j, j \in [1:i]) = \frac{1}{2}\log\left|(2\pi e)\left(\boldsymbol{K}_0^{-1} + \boldsymbol{K}_i^{-1} + \sum_{j=1}^{i} \boldsymbol{B}_j\right)^{-1}\right|, \quad i \in [1:L], \tag{97}$$

$$h(X|Y_{i+1}, W_j, j \in [1:i]) = \frac{1}{2}\log\left|(2\pi e)\left(\boldsymbol{K}_0^{-1} + \boldsymbol{K}_{i+1}^{-1} + \sum_{j=1}^{i} \boldsymbol{B}_j\right)^{-1}\right|, \quad i \in [1:L-1]. \tag{98}$$

Moreover, we have

$$h(X|Y_i) = h(X|X+N_i) = \frac{1}{2}\log\left|(2\pi e)\left(\boldsymbol{K}_0^{-1} + \boldsymbol{K}_i^{-1}\right)^{-1}\right|, \quad i \in [1:L], \tag{99}$$

$$\mathrm{cov}\left(X|Y_i, W_j, j \in [1:i]\right) = \left(\boldsymbol{K}_0^{-1} + \boldsymbol{K}_i^{-1} + \sum_{j=1}^{i} \boldsymbol{B}_j\right)^{-1}, \quad i \in [1:L]. \tag{100}$$

Now one can readily prove $\mathcal{R}(\mathbf{D}_i, i \in [1:L]) \subseteq \mathcal{R}^*(\mathbf{D}_i, i \in [1:L])$ by invoking Lemma 1 and a timesharing argument.

B. Supporting Hyperplane Characterization

Since $\mathcal{R}(\mathbf{D}_i, i \in [1:L])$ is convex, it is completely specified by its supporting hyperplanes. The characterization of the supporting hyperplanes boils down to solving the following optimization problem

$$R^* \triangleq \inf_{(R_1,\ldots,R_L)\in\mathcal{R}(\mathbf{D}_i, i\in[1:L])} \sum_{i=1}^{L} \mu_i R_i, \tag{101}$$

where $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_L \geq 0$. It is clear that

$$R^* = \min_{(\mathbf{B}_i, i\in[1:L])} \frac{\mu_1}{2}\log\frac{|\boldsymbol{K}_0^{-1} + \boldsymbol{K}_1^{-1} + \boldsymbol{B}_1|}{|\boldsymbol{K}_0^{-1} + \boldsymbol{K}_1^{-1}|} + \sum_{i=2}^{L}\frac{\mu_i}{2}\log\frac{|\boldsymbol{K}_0^{-1} + \boldsymbol{K}_i^{-1} + \sum_{j=1}^{i}\boldsymbol{B}_j|}{|\boldsymbol{K}_0^{-1} + \boldsymbol{K}_i^{-1} + \sum_{j=1}^{i-1}\boldsymbol{B}_j|} \tag{102}$$

$$\text{subject to} \quad \boldsymbol{B}_i \succeq \boldsymbol{0}, \qquad i \in [1;L],$$

$$\sum_{j=1}^{i} \boldsymbol{B}_j \succeq \boldsymbol{D}_i^{-1} - \boldsymbol{K}_0^{-1} - \boldsymbol{K}_i^{-1}, \quad i \in [1;L].$$

Theorem 3: The minimizer $(\mathbf{B}_i^*, i \in [1:L])$ of (102) must satisfy

$$\frac{\mu_i}{2}\left(\boldsymbol{K}_0^{-1} + \boldsymbol{K}_i^{-1} + \sum_{j=1}^{i}\boldsymbol{B}_j^*\right)^{-1} - \frac{\mu_{i+1}}{2}\left(\boldsymbol{K}_0^{-1} + \boldsymbol{K}_{i+1}^{-1} + \sum_{j=1}^{i}\boldsymbol{B}_j^*\right)^{-1} = \boldsymbol{\Psi}_i - \boldsymbol{\Psi}_{i+1} + \boldsymbol{\Lambda}_i, \quad i \in [1:L-1], \tag{103}$$

$$\frac{\mu_L}{2}\left(\boldsymbol{K}_0^{-1} + \boldsymbol{K}_L^{-1} + \sum_{j=1}^{L}\boldsymbol{B}_j^*\right)^{-1} = \boldsymbol{\Psi}_L + \boldsymbol{\Lambda}_L, \tag{104}$$

for some positive semi-definite matrices $(\boldsymbol{\Psi}_i, i \in [1:L])$ and $(\boldsymbol{\Lambda}_i, i \in [1:L])$ such that

$$\boldsymbol{B}_i^*\boldsymbol{\Psi}_i = \boldsymbol{0}, \qquad i \in [1:L], \tag{105}$$

$$\left(\boldsymbol{K}_0^{-1} + \boldsymbol{K}_i^{-1} + \sum_{j=1}^{i}\boldsymbol{B}_j^* - \boldsymbol{D}_i^{-1}\right)\boldsymbol{\Lambda}_i = \boldsymbol{0}, \qquad i \in [1:L]. \tag{106}$$

Proof: The Lagrangian of (102) is given by

$$\frac{\mu_1}{2}\log\frac{|\boldsymbol{K}_0^{-1} + \boldsymbol{K}_1^{-1} + \boldsymbol{B}_1|}{|\boldsymbol{K}_0^{-1} + \boldsymbol{K}_1^{-1}|} + \sum_{i=2}^{L}\frac{\mu_i}{2}\log\frac{|\boldsymbol{K}_0^{-1} + \boldsymbol{K}_i^{-1} + \sum_{j=1}^{i}\boldsymbol{B}_j|}{|\boldsymbol{K}_0^{-1} + \boldsymbol{K}_i^{-1} + \sum_{j=1}^{i-1}\boldsymbol{B}_j|} - \sum_{i=1}^{L}\mathrm{tr}\{\boldsymbol{B}_i\boldsymbol{\Psi}_i + (\boldsymbol{K}_0^{-1} + \boldsymbol{K}_i^{-1} - \boldsymbol{D}_i^{-1} + \sum_{j=1}^{i}\boldsymbol{B}_j)\boldsymbol{\Lambda}_i\}, \tag{107}$$

where positive semi-definite matrices $(\mathbf{\Psi}_i, i \in [1:L])$ and $(\mathbf{\Lambda}_i, i \in [1:L])$ serve as Lagrange multipliers. Note that (103)-(106) follow directly form the KKT conditions. The proof is complete by verifying a set of constraint qualifications in [45, Sections 4-5]. ■

Remark 4: It is worth noting that (103)-(106) in Theorem 3 correspond exactly to (13)-(16) in Theorem 2.

C. Converse

It is easy to adapt the converse argument in [9], [10] to prove the following result.

Lemma 2: For any $(R_i, i \in [1:L]) \in \mathcal{R}^*(\boldsymbol{D}_i, i \in [1:L])$ and any $\epsilon > 0$, there exist auxiliary random objects jointly distributed with $(X, Y_i, i \in [1:L])$ satisfying

- the Markov chain constraint

$$(W_i, i \in [1:L]) \to X \to Y_L \to Y_{L-1} \to \ldots \to Y_1, \tag{108}$$

- the rate constraints

$$R_1 + \epsilon \geq I(X; W_1|Y_1), \tag{109}$$

$$\sum_{j=1}^{i}(R_j + \epsilon) \geq I(X; W_1|Y_1) + \sum_{j=2}^{i} I(X; W_j|W_{j-1}, \ldots, W_1, Y_j), \quad i \in [2:L], \tag{110}$$

- the covariance distortion constraints

$$\mathrm{cov}(X|Y_i, W_j, j \in [1:i]) \preceq \boldsymbol{D}_i + \epsilon\mathbf{I}, \qquad i \in [1:L]. \tag{111}$$

Now we proceed to show that $\mathcal{R}^*(\boldsymbol{D}_i, i \in [1:L]) \subseteq \mathcal{R}(\boldsymbol{D}_i, i \in [1:L])$. For any $(R_1, \ldots, R_L) \in \mathcal{R}^*(\mathbf{D}_i, i \in [1:L])$ and any $\epsilon > 0$, it follows by Lemma 2, Theorem 3, and Theorem 2 that

$$\sum_{i=1}^{L}\mu_i(R_i + \epsilon) \geq \mu_1 I(X; W_1|Y_1) + \sum_{i=2}^{L}\mu_i I(X; W_i|W_j, Y_i, j \in [1:i-1]) \tag{112}$$

$$\begin{aligned}
&= \mu_1 h(X|Y_1) + \sum_{i=1}^{L-1}\big(\mu_i h(Y_i|W_j, j \in [1:i]) - \mu_i h(Y_i|X) \\
&\quad -\mu_{i+1}h(Y_{i+1}|W_j, j \in [1:i]) + \mu_{i+1}h(Y_{i+1}|X) \\
&\quad -(\mu_i - \mu_{i+1})h(X|W_j, j \in [1:i])\big) \\
&\quad + \mu_L h(Y_L|W_j, j \in [1:L]) - \mu_L h(Y_L|X) \\
&\quad - \mu_L h(X|W_j, j \in [1:L])
\end{aligned} \tag{113}$$

$$\begin{aligned}
&\geq -\frac{\mu_1}{2}\log\left|(2\pi e)^{-1}\left(\boldsymbol{K}_0^{-1} + \boldsymbol{K}_i^{-1}\right)\right| + \sum_{i=1}^{L-1}\Bigg(-\frac{\mu_{i+1}}{2} \\
&\quad \log\left|(2\pi e)^{-1}\left(\boldsymbol{K}_0^{-1} + \boldsymbol{K}_{i+1}^{-1} + \sum_{j=1}^{i}\boldsymbol{B}_j^*(\epsilon)\right)\right| \\
&\quad + \frac{\mu_i}{2}\log\left|(2\pi e)^{-1}\left(\boldsymbol{K}_0^{-1} + \boldsymbol{K}_i^{-1} + \sum_{j=1}^{i}\boldsymbol{B}_j^*(\epsilon)\right)\right|\Bigg) \\
&\quad + \frac{\mu_L}{2}\log\left|(2\pi e)^{-1}\left(\boldsymbol{K}_0^{-1} + \boldsymbol{K}_L^{-1} + \sum_{j=1}^{L}\boldsymbol{B}_j^*(\epsilon)\right)\right|,
\end{aligned} \tag{114}$$

where $(\mathbf{B}_i^*(\epsilon), i \in [1:L])$ denotes the minimizer of (102) with $(\mathbf{D}_i, i \in [1:L])$ replaced by $(\mathbf{D}_i + \epsilon\mathbf{I}, i \in [1:L])$. Now one can readily show

$$\sum_{i=1}^{L}\mu_i R_i \geq R^* \tag{115}$$

via a simple limiting argument. This completes the proof of Theorem 1.

V. Conclusion

We have studied the problem of successive refinement for Wyner-Ziv coding with degraded side information and obtained a computable characterization of the rate region in the quadratic vector Gaussian setting. From the technical perspective, our main contribution is a new extremal inequality, which is established via a refined monotone path argument inspired by the doubling trick in [38]. The proof of Gaussian optimality also arises in functional inequalities such as the Brascamp-Lieb inequality [46] and Young's inequality [43]. Apart from the doubling trick and the monotone path argument, many other techniques (e.g., rearrangement [47] and optimal transport [48]) can also be used for establishing such inequalities. It is an active research topic to investigate the connections among these inequalities and identify a unifying theme. Moreover, a deeping understanding of the geometric nature of these problems will likely shed light on the feasibility of the relevant techniques.

For the quadratic scalar Gaussian case of the side information scalable source coding problem, Tian and Diggavi [49] proved the optimality of the Gaussian solution even when the side informations at the receivers are not degraded along the same successive coding order (see [8] for a related result regarding a Heegard-Berger problem with two sources and degraded reconstruction sets). Specifically, this is accomplished by ranking the auxiliary random variables through the comparisons of the relevant distortions. However, the vector Gaussian case is considerably more challenging as covariance distortions might not have a linear order, and so far there are only some partial solutions [50]. Our proof technique does not require such comparisons and thus is potentially better suited to the non-degraded side information setting. On the other hand, the absence of a suitable single-letter outer bound for this general setting is a major hurdle for our approach. It is conceivable that one may overcome this difficulty by exploiting certain implicit Markov structures extracted from the KKT conditions of extremal Gaussian solutions for the achievability scheme.

APPENDIX A
PRELIMINARIES ON FISHER INFORMATION AND MMSE

Here is a summary of some basic properties of Fisher information and MMSE, which will be used extensively in the proof of extremal inequality (19).

We begin with the definition of conditional Fisher information matrix and MMSE matrix.

Definition 2: Let (X,U) be a pair of jointly distributed random vectors with differentiable conditional probability density function:

$$f(\boldsymbol{x}|u) \triangleq f(x_i, i \in [1:m]|u). \tag{116}$$

The vector-valued score function is defined as

$$\nabla \log f(\boldsymbol{x}|u) = \left[\frac{\partial \log f(\boldsymbol{x}|u)}{\partial x_1}, \cdots, \frac{\partial \log f(\boldsymbol{x}|u)}{\partial x_m}\right]^T. \tag{117}$$

The conditional Fisher information of X respect to U is given by

$$J(X|U) = \mathbb{E}\left[(\nabla \log f(\boldsymbol{x}|u)) \cdot (\nabla \log f(\boldsymbol{x}|u))^T\right]. \tag{118}$$

Definition 3: Let (X,Y,U) be a set of jointly distributed random vectors. The conditional covariance matrix of X given (Y,U) is defined as

$$\mathrm{cov}(X|Y,U) = \mathbb{E}\left[(X - \mathbb{E}[X|Y,U]) \cdot (X - \mathbb{E}[X|Y,U])^T\right]. \tag{119}$$

Lemma 3 (Matrix Version of de Bruijn's Identity): Let (X,U) be a pair of jointly distributed random vectors, and $N \sim N(\mathbf{0},\mathbf{\Sigma})$ be a Gaussian random vector independent of (X,U). Then

$$\nabla_{\mathbf{\Sigma}} h(X+N|U) = \frac{1}{2} J(X+N|U). \tag{120}$$

Lemma 3 is a conditional version of [51, Theorem 1], which provides a link between differential entropy and Fisher information.

Lemma 4: Let (X,U) be a pair of jointly distributed random vectors, and $N \sim \mathcal{N}(\mathbf{0},\mathbf{\Sigma})$ be a Gaussian random vector independent of (X,U). Then

$$J(X+N|U) + \mathbf{\Sigma}^{-1}\,\mathrm{cov}(X|X+N,U)\mathbf{\Sigma}^{-1} = \mathbf{\Sigma}^{-1}. \tag{121}$$

The complementary identity in Lemma 4 provides a link between Fisher information and MMSE, and its proof can be found in [51, Corollary 1].

Lemma 5: Let (X,Y,U) be a set of jointly distributed random vectors. Assume that X and Y are conditionally independent given U. Then for any square matrix $\boldsymbol{A}$ and $\boldsymbol{B}$,

$$\begin{aligned}&(\boldsymbol{A}+\boldsymbol{B})J(X+Y|U)(\boldsymbol{A}+\boldsymbol{B})^T\\ &\preceq \boldsymbol{A}J(X|U)\boldsymbol{A}^T + \boldsymbol{B}J(Y|U)\boldsymbol{B}^T.\end{aligned} \tag{122}$$

Proof: From the conditional version of matrix Fisher information inequality in [42, Appendix II], we have

$$J(X+Y|U) \preceq \boldsymbol{K}J(X|U)\boldsymbol{K}^T + (\boldsymbol{I}-\boldsymbol{K})J(Y|U)(\boldsymbol{I}-\boldsymbol{K})^T, \tag{123}$$

for any square matrix $\boldsymbol{K}$. Setting

$$\boldsymbol{K} = (\boldsymbol{A}+\boldsymbol{B})^{-1}\boldsymbol{A} \tag{124}$$

proves (122). ■

Lemma 6: Let X be a Gaussian random vector and U be an arbitrary random vector. Let N_1 and N_2 be two zero-mean Gaussian random vectors, independent of (X,U), with covariance matrices $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$, respectively. If

$$\mathbf{\Sigma}_2 \succ \mathbf{\Sigma}_1 \succ \mathbf{0}, \tag{125}$$

then

$$\mathrm{cov}\left(X\middle|X+N_1,U\right)^{-1} - \mathbf{\Sigma}_1^{-1} \succeq \mathrm{cov}\left(X\middle|X+N_2,U\right)^{-1} - \mathbf{\Sigma}_2^{-1}. \tag{126}$$

Lemma 6 can be proved by combining the Cramér-Rao inequality and the complementary identity in Lemma 4. See [39, Lemma 4] for details.

Lemma 7 (Data Processing Inequality for Fisher Information): Let (X,U,V) be a set of jointly distributed random vectors. Assume that $U \to V \to X$ form a Markov chain. Then

$$J(X|U) \preceq J(X|V). \tag{127}$$

Lemma 7 is analogous to [52, Lemma 3], and can be easily proved using the chain rule of Fisher information matrix [52, Lemma 1].

Lemma 8 (Data Processing Inequality for MMSE): Let (X,U,V) be a set of jointly distributed random vectors. Assume $U \to V \to X$ form a Markov chain. Then

$$\mathrm{cov}(X|U) \succeq \mathrm{cov}(X|V). \tag{128}$$

See [53, Proposition 5] for a detailed proof of Lemma 8.

APPENDIX B
DERIVATIVE OF THE BIVARIATE DIFFERENTIAL ENTROPY $h\left(X_{i,\gamma}, Y_{i,\gamma}^*|W_j, j \in [1:i]\right)$

In view of (27) and (30), we have

$$\begin{aligned}&h\left(X_{i,\gamma}, Y_{i,\gamma}^*|W_j, j \in [1:i]\right) &(129)\\ &= h\Big(\sqrt{1-\gamma}X + \sqrt{\gamma}X_i^G, \sqrt{\gamma}Y_i - \sqrt{1-\gamma}Y_i^G|\\ &\quad W_j, j \in [1:i]\Big) &(130)\\ &= h\left(X + \sqrt{\frac{\gamma}{1-\gamma}}X_i^G, Y_i - \sqrt{\frac{1-\gamma}{\gamma}}Y_i^G\middle|W_j, j \in [1:i]\right)\\ &\quad + \frac{n}{2}\log\gamma + \frac{n}{2}\log(1-\gamma). &(131)\end{aligned}$$

Recall from (23) that

$$Y_i^G = X_i^G + N_i^G. \tag{132}$$

The covariance matrix of

$$\begin{pmatrix} \sqrt{\gamma/(1-\gamma)}X_i^G \\ -\sqrt{(1-\gamma)/\gamma}Y_i^G \end{pmatrix}$$

is given by

$$\mathbf{\Sigma}_{i,*} \triangleq \begin{pmatrix} \frac{\gamma}{1-\gamma}\mathbf{\Delta}_i & -\mathbf{\Delta}_i \\ -\mathbf{\Delta}_i & \frac{1-\gamma}{\gamma}(\mathbf{\Delta}_i + \boldsymbol{K}_i) \end{pmatrix}. \tag{133}$$

It is easy to verify that

$$\mathbf{\Sigma}_{i,*}^{-1} = \begin{pmatrix} \frac{1-\gamma}{\gamma}(\mathbf{\Delta}_i^{-1} + \boldsymbol{K}_i^{-1}) & \boldsymbol{K}_i^{-1} \\ \boldsymbol{K}_i^{-1} & \frac{\gamma}{1-\gamma}\boldsymbol{K}_i^{-1} \end{pmatrix} \tag{134}$$

and

$$\nabla_\gamma \mathbf{\Sigma}_{i,*} = \begin{pmatrix} \frac{1}{(1-\gamma)^2}\mathbf{\Delta}_i & \mathbf{0} \\ \mathbf{0} & -\frac{1}{\gamma^2}(\mathbf{\Delta}_i + \boldsymbol{K}_i) \end{pmatrix}. \tag{135}$$

Combining (134) and (135) gives

$$\operatorname{tr}\left\{(\nabla_\gamma \mathbf{\Sigma}_{i,*})\,\mathbf{\Sigma}_{i,*}^{-1}\right\} = 0, \tag{136}$$

$$\mathbf{\Sigma}_{i,*}^{-1}(\nabla_\gamma \mathbf{\Sigma}_{i,*})\mathbf{\Sigma}_{i,*}^{-1} = \begin{pmatrix} -\frac{1}{\gamma^2}(\mathbf{\Delta}_i^{-1} + \boldsymbol{K}_i^{-1}) & \mathbf{0} \\ \mathbf{0} & \frac{1}{(1-\gamma)^2}\boldsymbol{K}_i^{-1} \end{pmatrix}. \tag{137}$$

By invoking the chain rule of matrix calculus and Lemma 3 in Appendix A, we have

$$\begin{aligned} &\frac{d}{d\gamma} h\left(X_{i,\gamma}, Y_{i,\gamma}^* | W_j, j \in [1:i]\right) \\ &= \frac{d}{d\gamma}\left\{ h\left(X + \sqrt{\frac{\gamma}{1-\gamma}}X_i^G, Y_i - \sqrt{\frac{1-\gamma}{\gamma}}Y_i^G \right| \right. \\ &\qquad \left. \left. W_j, j \in [1:i]\right) + \frac{n}{2}\log\gamma + \frac{n}{2}\log(1-\gamma)\right\} \qquad (138) \\ &= \frac{1}{2}\operatorname{tr}\left\{(\nabla_\gamma \mathbf{\Sigma}_{i,*})\, J\left(\left(\sqrt{\tfrac{1}{1-\gamma}}X_{i,\gamma}^T \quad \sqrt{\tfrac{1}{\gamma}}Y_{i,\gamma}^{*\,T}\right)^T \right| \right. \\ &\qquad \left. \left. W_j, j \in [1:i]\right)\right\} + \frac{n}{2}\left(\frac{1}{\gamma} - \frac{1}{1-\gamma}\right). \qquad (139) \end{aligned}$$

It can be verified

$$\begin{aligned} &\operatorname{tr}\left\{ (\nabla_\gamma \mathbf{\Sigma}_{i,*}) \right. \\ &\quad \left. J\left(\left(\sqrt{\tfrac{1}{1-\gamma}}X_{i,\gamma}^T \quad \sqrt{\tfrac{1}{\gamma}}Y_{i,\gamma}^{*\,T}\right)^T \middle| W_j, j \in [1:i]\right)\right\} \\ &= \operatorname{tr}\left\{ (\nabla_\gamma \mathbf{\Sigma}_{i,*})\,\mathbf{\Sigma}_{i,*}^{-1} - \mathbf{\Sigma}_{i,*}^{-1}(\nabla_\gamma \mathbf{\Sigma}_{i,*})\,\mathbf{\Sigma}_{i,*}^{-1} \right. \\ &\qquad \operatorname{cov}\left(\left(X^T \quad Y_i^T\right)^T \middle| X + \sqrt{\frac{\gamma}{1-\gamma}}X_i^G, \right. \\ &\qquad\qquad \left.\left. Y_i - \sqrt{\frac{1-\gamma}{\gamma}}Y_i^G, W_j, j \in [1:i]\right)\right\} \qquad (140) \\ &= \operatorname{tr}\left\{ \begin{pmatrix} -\frac{1}{\gamma^2}(\mathbf{\Delta}_i^{-1} + \boldsymbol{K}_i^{-1}) & \mathbf{0} \\ \mathbf{0} & \frac{1}{(1-\gamma)^2}\boldsymbol{K}_i^{-1} \end{pmatrix} \right. \\ &\qquad \operatorname{cov}\left(\left(X^T \quad Y_i^T\right)^T \middle| X + \sqrt{\frac{\gamma}{1-\gamma}}X_i^G, \right. \\ &\qquad\qquad \left.\left. Y_i - \sqrt{\frac{1-\gamma}{\gamma}}Y_i^G, W_j, j \in [1:i]\right)\right\}, \qquad (141) \end{aligned}$$

where (140) follows by Lemma 4 in Appendix A, and (141) is due to (136) and (137). Notice that

$$\begin{aligned} &\operatorname{cov}\left(\left(X^T \quad Y_i^T\right)^T \middle| X + \sqrt{\frac{\gamma}{1-\gamma}}X_i^G, Y_i - \sqrt{\frac{1-\gamma}{\gamma}}Y_i^G\right) \\ &= \left(\begin{pmatrix} \boldsymbol{K}_0 & \boldsymbol{K}_0 \\ \boldsymbol{K}_0 & \boldsymbol{K}_0 + \boldsymbol{K}_i \end{pmatrix}^{-1} + \mathbf{\Sigma}_{i,*}^{-1}\right)^{-1} \qquad (142) \\ &= \left(\begin{pmatrix} \boldsymbol{K}_0^{-1} + \boldsymbol{K}_i^{-1} & -\boldsymbol{K}_i^{-1} \\ -\boldsymbol{K}_i^{-1} & \boldsymbol{K}_i^{-1} \end{pmatrix} \right. \\ &\qquad \left. + \begin{pmatrix} \frac{1-\gamma}{\gamma}(\mathbf{\Delta}_i^{-1} + \boldsymbol{K}_i^{-1}) & \boldsymbol{K}_i^{-1} \\ \boldsymbol{K}_i^{-1} & \frac{\gamma}{1-\gamma}\boldsymbol{K}_i^{-1} \end{pmatrix}\right)^{-1} \qquad (143) \\ &= \begin{pmatrix} \left(\boldsymbol{K}_0^{-1} + \frac{1-\gamma}{\gamma}\mathbf{\Delta}_i^{-1} + \frac{1}{\gamma}\boldsymbol{K}_i^{-1}\right)^{-1} & \mathbf{0} \\ \mathbf{0} & (1-\gamma)\boldsymbol{K}_i \end{pmatrix}. \qquad (144) \end{aligned}$$

Thus, we have the Markov chain

$$\begin{aligned} &(W_j, j \in [1:i]) \to X \to \\ &\left(X + \sqrt{\frac{\gamma}{1-\gamma}}X_i^G, Y_i - \sqrt{\frac{1-\gamma}{\gamma}}Y_i^G\right) \to Y_i. \qquad (145) \end{aligned}$$

As a consequence,

$$\begin{aligned} &\operatorname{cov}\left(\left(X^T \quad Y_i^T\right)^T \middle| X_{i,\gamma}, Y_{i,\gamma}^*, W_j, j \in [1:i]\right) \\ &= \begin{pmatrix} \operatorname{cov}\left(X \middle| X_{i,\gamma}, Y_{i,\gamma}, W_j, j \in [1:i]\right) & \mathbf{0} \\ \mathbf{0} & (1-\gamma)\boldsymbol{K}_i \end{pmatrix}. \qquad (146) \end{aligned}$$

By combining (139), (141) and (146), we obtain

$$\begin{aligned} &\frac{d}{d\gamma} h\left(X_{i,\gamma}, Y_{i,\gamma}^* \middle| W_j, j \in [1:i]\right) \\ &= \frac{1}{2}\operatorname{tr}\left\{ \begin{pmatrix} -\frac{1}{\gamma^2}(\mathbf{\Delta}_i^{-1} + \boldsymbol{K}_i^{-1}) & \mathbf{0} \\ \mathbf{0} & \frac{1}{(1-\gamma)^2}\boldsymbol{K}_i^{-1} \end{pmatrix} \right. \\ &\qquad \left. \begin{pmatrix} \operatorname{cov}\left(X \middle| X_{i,\gamma}, Y_{i,\gamma}^*, W_j, j \in [1:i]\right) & \mathbf{0} \\ \mathbf{0} & (1-\gamma)\boldsymbol{K}_i \end{pmatrix}\right\} \\ &\quad + \frac{n}{2}\left(\frac{1}{\gamma} - \frac{1}{1-\gamma}\right) \qquad (147) \\ &= -\frac{1}{2\gamma}\operatorname{tr}\left\{ \frac{1}{\gamma}(\mathbf{\Delta}_i^{-1} + \boldsymbol{K}_i^{-1}) \right. \\ &\qquad \left. \operatorname{cov}\left(X \middle| X_{i,\gamma}, Y_{i,\gamma}^*, W_j, j \in [1:i]\right) - \boldsymbol{I}\right\} \qquad (148) \\ &= -\frac{1}{2\gamma}\operatorname{tr}\left\{ (\mathbf{\Delta}_i^{-1} + \boldsymbol{K}_i^{-1}) \right. \\ &\qquad \left(\frac{1}{\gamma}\operatorname{cov}\left(X \middle| X_{i,\gamma}, Y_{i,\gamma}^*, W_j, j \in [1:i]\right) \right. \\ &\qquad \left.\left. - (\mathbf{\Delta}_i^{-1} + \boldsymbol{K}_i^{-1})^{-1}\right)\right\}. \qquad (149) \end{aligned}$$

On the other hand, it follows by the theory of linear MMSE estimation that

$$\begin{aligned} \sqrt{\gamma}X_i^G = &-\sqrt{\gamma(1-\gamma)}\left(\mathbf{\Delta}_i^{-1} + (1-\gamma)\,\boldsymbol{K}_i^{-1}\right)^{-1}\boldsymbol{K}_i^{-1} \\ &\left(\sqrt{\gamma}N_i - \sqrt{1-\gamma}Y_i^G\right) + \sqrt{\gamma}\hat{X}_i^G, \qquad (150) \end{aligned}$$

where $\hat{X}_{i,\gamma}$ is a Gaussian random vector with mean zero and covariance matrix $\left(\mathbf{\Delta}_i^{-1} + (1-\gamma)\,\boldsymbol{K}_i^{-1}\right)^{-1}$, and is independent of $\sqrt{\gamma}N_i - \sqrt{1-\gamma}Y_i^G$. Thus, we have

$$X_{i,\gamma} = \sqrt{1-\gamma}X + \sqrt{\gamma}X_i^G \tag{151}$$

$$= \sqrt{1-\gamma}X - \sqrt{\gamma(1-\gamma)}\left(\boldsymbol{\Delta}_i^{-1} + (1-\gamma)\boldsymbol{K}_i^{-1}\right)^{-1} \boldsymbol{K}_i^{-1}\left(\sqrt{\gamma}N_i - \sqrt{1-\gamma}Y_i^G\right) + \sqrt{\gamma}\hat{X}_i^G \quad (152)$$

$$= \sqrt{1-\gamma}\left(\boldsymbol{\Delta}_i^{-1} + (1-\gamma)\boldsymbol{K}_i^{-1}\right)^{-1}\left(\boldsymbol{\Delta}_i^{-1} + \boldsymbol{K}_i^{-1}\right) X + \sqrt{\gamma}\hat{X}_i^G - \sqrt{\gamma(1-\gamma)}\left(\boldsymbol{\Delta}_i^{-1} + (1-\gamma)\boldsymbol{K}_i^{-1}\right)^{-1}\boldsymbol{K}_i^{-1}Y_{i,\gamma}^*. \quad (153)$$

The complementary Fisher information representation of $\operatorname{cov}\left(X\middle|X_{i,\gamma}, Y_{i,\gamma}^*, W_j, j \in [1:i]\right)$ can thereby be expressed as

$$\operatorname{cov}\left(X\middle|X_{i,\gamma}, Y_{i,\gamma}^*, W_j, j \in [1:i]\right) \quad (154)$$
$$= \operatorname{cov}\left(X\middle|\sqrt{1-\gamma}\left(\boldsymbol{\Delta}_i^{-1}+(1-\gamma)\boldsymbol{K}_i^{-1}\right)^{-1}\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_i^{-1}\right) X + \sqrt{\gamma}\hat{X}_i^G, Y_{i,\gamma}^*, W_j, j \in [1:i]\right) \quad (155)$$
$$= \frac{\gamma}{1-\gamma}\left(\boldsymbol{\Delta}_i^{-1} + \boldsymbol{K}_i^{-1}\right)^{-1}\left(\boldsymbol{\Delta}_i^{-1} + \left(1-\gamma\right)\boldsymbol{K}_i^{-1} - \gamma J\left(X_{i,\gamma}\middle|Y_{i,\gamma}^*, W_j, j \in [1:i]\right)\right)\left(\boldsymbol{\Delta}_i^{-1} + \boldsymbol{K}_i^{-1}\right)^{-1}. \quad (156)$$

Equivalently, we can write

$$\left(\boldsymbol{\Delta}_i^{-1} + \boldsymbol{K}_i^{-1}\right)\left(\frac{1}{\gamma}\operatorname{cov}\left(X\middle|X_{i,\gamma}, \tilde{Y}_{i,\gamma}^*, W_j, j \in [1:L]\right) - \left(\boldsymbol{\Delta}_i^{-1} + \boldsymbol{K}_i^{-1}\right)^{-1}\right)\left(\boldsymbol{\Delta}_i^{-1} + \boldsymbol{K}_i^{-1}\right) \quad (157)$$
$$= \frac{\gamma}{1-\gamma}\boldsymbol{\Delta}_i^{-1} - \frac{\gamma}{1-\gamma}J\left(X_{i,\gamma}\middle|Y_{i,\gamma}^*, W_j, j \in [1:i]\right). \quad (158)$$

Finally, substituting (158) into (149) gives

$$\frac{d}{d\gamma}h\left(X_{i,\gamma}, Y_{i,\gamma}^*\middle|W_j, j \in [1:i]\right) = \frac{1}{2(1-\gamma)}\operatorname{tr}\left\{\left(\boldsymbol{\Delta}_i^{-1} + \boldsymbol{K}_i^{-1}\right)^{-1}\left(J\left(X_{i,\gamma}\middle|Y_{i,\gamma}^*, W_j, j \in [1:i]\right) - \boldsymbol{\Delta}_i^{-1}\right)\right\}. \quad (159)$$

APPENDIX C
DERIVATIVE OF THE BIVARIATE DIFFERENTIAL ENTROPY $h\left(\tilde{Y}_{i+1,\gamma}, Y_{i,\gamma}^*|W_j, j \in [1:i]\right)$

In view of (29) and (30),

$$h\left(\tilde{Y}_{i+1,\gamma}, Y_{i,\gamma}^*|W_j, j \in [1:i]\right) \quad (160)$$
$$= h\left(\sqrt{1-\gamma}Y_{i+1} + \sqrt{\gamma}\tilde{Y}_{i+1}^G, \sqrt{\gamma}Y_i - \sqrt{1-\gamma}Y_i^G\middle| W_j, j \in [1:i]\right) \quad (161)$$
$$= h\left(Y_{i+1} + \sqrt{\frac{\gamma}{1-\gamma}}\tilde{Y}_{i+1}^G, Y_i - \sqrt{\frac{1-\gamma}{\gamma}}Y_i^G\middle| W_j, j \in [1:i]\right) + \frac{n}{2}\log\gamma + \frac{n}{2}\log(1-\gamma). \quad (162)$$

By the definition of Y_i^G and $\tilde{Y}_{i+1}^G$ in (23) and (24) as well as the construction of $\left(N_i^G, i \in [1:L]\right)$, we can write

$$Y_i^G = \tilde{Y}_{i+1}^G + \left(N_i^G - N_{i+1}^G\right), \quad (163)$$

where $N_i^G - N_{i+1}^G$ is a Gaussian random vector with covariance matrix $\boldsymbol{K}_i - \boldsymbol{K}_{i+1}$, and is independent of $\tilde{Y}_{i+1}^G$. Therefore, the covariance matrix of

$$\begin{pmatrix} \sqrt{\gamma/(1-\gamma)}\tilde{Y}_{i+1}^G \\ -\sqrt{(1-\gamma)/\gamma}Y_i^G \end{pmatrix}$$

is given by

$$\tilde{\boldsymbol{\Sigma}}_i \triangleq \begin{pmatrix} \frac{\gamma}{1-\gamma}\left(\boldsymbol{\Delta}_i + \boldsymbol{K}_i\right) & -\left(\boldsymbol{\Delta}_i + \boldsymbol{K}_i\right) \\ -\left(\boldsymbol{\Delta}_i + \boldsymbol{K}_i\right) & \frac{1-\gamma}{\gamma}\left(\boldsymbol{\Delta}_i + \boldsymbol{K}_{i+1}\right) \end{pmatrix}. \quad (164)$$

It can be verified that

$$\tilde{\boldsymbol{\Sigma}}_i^{-1} = \begin{pmatrix} \frac{1-\gamma}{\gamma}\boldsymbol{M}_{11} & \left(\boldsymbol{K}_i - \boldsymbol{K}_{i+1}\right)^{-1} \\ \left(\boldsymbol{K}_i - \boldsymbol{K}_{i+1}\right)^{-1} & \frac{\gamma}{1-\gamma}\left(\boldsymbol{K}_i - \boldsymbol{K}_{i+1}\right)^{-1} \end{pmatrix}, \quad (165)$$

where

$$\boldsymbol{M}_{11} = \left(\boldsymbol{\Delta}_i + \boldsymbol{K}_{i+1}\right)^{-1} + \left(\boldsymbol{K}_i - \boldsymbol{K}_{i+1}\right)^{-1}, \quad (166)$$

and

$$\nabla_\gamma\tilde{\boldsymbol{\Sigma}}_i = \begin{pmatrix} \frac{1}{(1-\gamma)^2}\left(\boldsymbol{\Delta}_i + \boldsymbol{K}_{i+1}\right) & \boldsymbol{0} \\ \boldsymbol{0} & -\frac{1}{\gamma^2}\left(\boldsymbol{\Delta}_i + \boldsymbol{K}_i\right) \end{pmatrix}. \quad (167)$$

Combining (165) and (167) gives

$$\operatorname{tr}\left\{\left(\nabla_\gamma\tilde{\boldsymbol{\Sigma}}_i\right)\tilde{\boldsymbol{\Sigma}}_i^{-1}\right\} = 0, \quad (168)$$

$$\tilde{\boldsymbol{\Sigma}}_i^{-1}\left(\nabla_\gamma\tilde{\boldsymbol{\Sigma}}_i\right)\tilde{\boldsymbol{\Sigma}}_i^{-1} = \begin{pmatrix} -\frac{1}{\gamma^2}\boldsymbol{M}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \frac{1}{(1-\gamma)^2}\left(\boldsymbol{K}_i - \boldsymbol{K}_{i+1}\right)^{-1} \end{pmatrix}. \quad (169)$$

By invoking the chain rule of matrix calculus and Lemma 3 in Appendix A, we have

$$\frac{d}{d\gamma}h\left(\tilde{Y}_{i+1,\gamma}, Y_{i,\gamma}^*|W_j, j \in [1:i]\right) = \frac{d}{d\gamma}\left\{h\left(Y_{i+1} + \sqrt{\frac{\gamma}{1-\gamma}}\tilde{Y}_{i+1}^G, Y_i - \sqrt{\frac{1-\gamma}{\gamma}}Y_i^G\middle| W_j, j \in [1:i]\right) + \frac{n}{2}\log\gamma + \frac{n}{2}\log(1-\gamma)\right\} \quad (170)$$
$$= \frac{1}{2}\operatorname{tr}\left\{\left(\nabla_\gamma\tilde{\boldsymbol{\Sigma}}_i\right)J\left(\left(\sqrt{\tfrac{1}{1-\gamma}}\tilde{Y}_{i+1,\gamma}^T \quad \sqrt{\tfrac{1}{\gamma}}{Y_{i,\gamma}^*}^T\right)^T\middle| W_j, j \in [1:i]\right)\right\} + \frac{n}{2}\left(\frac{1}{\gamma} - \frac{1}{1-\gamma}\right) \quad (171)$$

It can be verified that

$$\operatorname{tr}\left\{\left(\nabla_\gamma\tilde{\boldsymbol{\Sigma}}_i\right) J\left(\left(\sqrt{\tfrac{1}{1-\gamma}}\tilde{Y}_{i+1,\gamma}^T \quad \sqrt{\tfrac{1}{\gamma}}{Y_{i,\gamma}^*}^T\right)^T\middle|W_j, j \in [1:i]\right)\right\}$$
$$= \operatorname{tr}\left\{\left(\nabla_\gamma\tilde{\boldsymbol{\Sigma}}_i\right)\tilde{\boldsymbol{\Sigma}}_i^{-1} - \tilde{\boldsymbol{\Sigma}}_i^{-1}\left(\nabla_\gamma\tilde{\boldsymbol{\Sigma}}_i\right)\tilde{\boldsymbol{\Sigma}}_i^{-1} \operatorname{cov}\left(\left(Y_{i+1}^T \quad Y_i^T\right)^T\middle|Y_{i+1} + \sqrt{\frac{\gamma}{1-\gamma}}\tilde{Y}_{i+1}^G, Y_i - \sqrt{\frac{1-\gamma}{\gamma}}Y_i^G, W_j, j \in [1:i]\right)\right\} \quad (172)$$

$$= \operatorname{tr}\left\{ \begin{pmatrix} -\frac{1}{\gamma^2}\boldsymbol{M}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \frac{1}{(1-\gamma)^2}\left(\boldsymbol{K}_i - \boldsymbol{K}_{i+1}\right)^{-1} \end{pmatrix} \operatorname{cov}\left(\begin{pmatrix} Y_{i+1}^T & Y_i^T \end{pmatrix}^T \middle| Y_{i+1} + \sqrt{\frac{\gamma}{1-\gamma}}\tilde{Y}_{i+1}^G, Y_i - \sqrt{\frac{1-\gamma}{\gamma}}Y_i^G, W_j, j \in [1:i] \right) \right\}, \tag{173}$$

where (172) follows by Lemma 4 in Appendix A, and (173) is due to (168) and (169). Notice that

$$\operatorname{cov}\left(\begin{pmatrix} Y_{i+1}^T & Y_i^T \end{pmatrix}^T \middle| Y_{i+1} + \sqrt{\frac{\gamma}{1-\gamma}}\tilde{Y}_{i+1}^G, Y_i - \sqrt{\frac{1-\gamma}{\gamma}}Y_i^G \right)$$

$$= \left(\begin{pmatrix} \boldsymbol{K}_0 + \boldsymbol{K}_{i+1} & \boldsymbol{K}_0 + \boldsymbol{K}_{i+1} \\ \boldsymbol{K}_0 + \boldsymbol{K}_{i+1} & \boldsymbol{K}_0 + \boldsymbol{K}_i \end{pmatrix}^{-1} + \tilde{\boldsymbol{\Sigma}}_i^{-1} \right)^{-1} \tag{174}$$

$$= \left(\begin{pmatrix} \boldsymbol{P}_{11} & -\left(\boldsymbol{K}_i - \boldsymbol{K}_{i+1}\right)^{-1} \\ -\left(\boldsymbol{K}_i - \boldsymbol{K}_{i+1}\right)^{-1} & \left(\boldsymbol{K}_i - \boldsymbol{K}_{i+1}\right)^{-1} \end{pmatrix} + \begin{pmatrix} \frac{1-\gamma}{\gamma}\boldsymbol{M}_{11} & \left(\boldsymbol{K}_i - \boldsymbol{K}_{i+1}\right)^{-1} \\ \left(\boldsymbol{K}_i - \boldsymbol{K}_{i+1}\right)^{-1} & \frac{\gamma}{1-\gamma}\left(\boldsymbol{K}_i - \boldsymbol{K}_{i+1}\right)^{-1} \end{pmatrix} \right)^{-1} \tag{175}$$

$$= \begin{pmatrix} \boldsymbol{Q}_{11}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & (1-\gamma)\left(\boldsymbol{K}_i - \boldsymbol{K}_{i+1}\right) \end{pmatrix}, \tag{176}$$

where

$$\boldsymbol{P}_{11} = \left(\boldsymbol{K}_0 + \boldsymbol{K}_{i+1}\right)^{-1} + \left(\boldsymbol{K}_i - \boldsymbol{K}_{i+1}\right)^{-1}, \tag{177}$$

$$\boldsymbol{Q}_{11} = \left(\boldsymbol{K}_0 + \boldsymbol{K}_{i+1}\right)^{-1} + \frac{1-\gamma}{\gamma}\left(\boldsymbol{\Delta}_i + \boldsymbol{K}_{i+1}\right)^{-1} + \frac{1}{\gamma}\left(\boldsymbol{K}_i - \boldsymbol{K}_{i+1}\right)^{-1}. \tag{178}$$

Thus, we have the Markov chain

$$(W_j, j \in [1:i]) \rightarrow Y_{i+1} \rightarrow \left(Y_{i+1} + \sqrt{\frac{\gamma}{1-\gamma}}\tilde{Y}_{i+1}^G, Y_i - \sqrt{\frac{1-\gamma}{\gamma}}Y_i^G \right) \rightarrow Y_i. \tag{179}$$

As a consequence,

$$\operatorname{cov}\left(\begin{pmatrix} Y_{i+1}^T & Y_i^T \end{pmatrix}^T \middle| \tilde{Y}_{i+1,\gamma}, Y_{i,\gamma}^*, W_j, j \in [1:i] \right) = \begin{pmatrix} \boldsymbol{T}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & (1-\gamma)(\boldsymbol{K}_i - \boldsymbol{K}_{i+1}) \end{pmatrix}, \tag{180}$$

where

$$\boldsymbol{T}_{11} = \operatorname{cov}\left(Y_{i+1} \middle| \tilde{Y}_{i+1,\gamma}, Y_{i,\gamma}^*, W_j, j \in [1:i] \right). \tag{181}$$

Combining (171), (173) and (180), we obtain

$$\frac{d}{d\gamma} h\left(\tilde{Y}_{i+1,\gamma}, Y_{i,\gamma}^* \middle| W_j, j \in [1:i] \right) = \frac{1}{2}\operatorname{tr}\left\{ \begin{pmatrix} -\frac{1}{\gamma^2}\boldsymbol{M}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \frac{1}{(1-\gamma)^2}\left(\boldsymbol{K}_i - \boldsymbol{K}_{i+1}\right)^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{T}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & (1-\gamma)(\boldsymbol{K}_i - \boldsymbol{K}_{i+1}) \end{pmatrix} \right\} + \frac{n}{2}\left(\frac{1}{\gamma} - \frac{1}{1-\gamma} \right) \tag{182}$$

$$= -\frac{1}{2\gamma}\operatorname{tr}\left\{ \frac{1}{\gamma}\boldsymbol{M}_{11}\boldsymbol{T}_{11} - \boldsymbol{I} \right\}. \tag{183}$$

On the other hand, it follows by the theory of linear MMSE estimation that

$$\sqrt{\gamma}\tilde{Y}_{i+1}^G = -\sqrt{\gamma(1-\gamma)}\left(\left(\boldsymbol{\Delta}_i + \boldsymbol{K}_{i+1}\right)^{-1} + (1-\gamma)\left(\boldsymbol{K}_i - \boldsymbol{K}_{i+1}\right)^{-1} \right)^{-1}\left(\boldsymbol{K}_i - \boldsymbol{K}_{i+1}\right)^{-1} \left(\sqrt{\gamma}N_i - \sqrt{\gamma}N_{i+1} - \sqrt{1-\gamma}Y_i^G \right) + \sqrt{\gamma}\hat{Y}_{i+1}^G, \tag{184}$$

where $\hat{Y}_{i+1}^G$ is a Gaussian random vector with mean zero and covariance matrix $\left(\left(\boldsymbol{\Delta}_i + \boldsymbol{K}_{i+1}\right)^{-1} + (1-\gamma)\left(\boldsymbol{K}_i - \boldsymbol{K}_{i+1}\right)^{-1} \right)^{-1}$, and is independent of $\sqrt{\gamma}\left(N_i - N_{i+1}\right) - \sqrt{1-\gamma}Y_i^G$. Thus, we have

$$\tilde{Y}_{i+1} = \sqrt{1-\gamma}Y_{i+1} + \sqrt{\gamma}\tilde{Y}_{i+1}^G$$

$$= \sqrt{1-\gamma}Y_{i+1} - \sqrt{\gamma(1-\gamma)}\left(\left(\boldsymbol{\Delta}_i + \boldsymbol{K}_{i+1}\right)^{-1} + (1-\gamma)\left(\boldsymbol{K}_i - \boldsymbol{K}_{i+1}\right)^{-1} \right)^{-1}\left(\boldsymbol{K}_i - \boldsymbol{K}_{i+1}\right)^{-1} \left(\sqrt{\gamma}N_i - \sqrt{\gamma}N_{i+1} - \sqrt{1-\gamma}Y_i^G \right) + \sqrt{\gamma}\hat{Y}_{i+1}^G \tag{185}$$

$$= \sqrt{1-\gamma}\left(\left(\boldsymbol{\Delta}_i + \boldsymbol{K}_{i+1}\right)^{-1} + (1-\gamma)\left(\boldsymbol{K}_i - \boldsymbol{K}_{i+1}\right)^{-1} \right)^{-1} \boldsymbol{M}_{11}Y_{i+1} + \sqrt{\gamma}\hat{Y}_{i+1}^G - \sqrt{\gamma(1-\gamma)}\left(\left(\boldsymbol{\Delta}_i + \boldsymbol{K}_{i+1}\right)^{-1} + (1-\gamma)\left(\boldsymbol{K}_i - \boldsymbol{K}_{i+1}\right)^{-1} \right)^{-1}\left(\boldsymbol{K}_i - \boldsymbol{K}_{i+1}\right)^{-1} Y_{i,\gamma}^*. \tag{186}$$

The complementary Fisher information representation of $\operatorname{cov}\left(Y_{i+1} \middle| \tilde{Y}_{i+1,\gamma}, Y_{i,\gamma}^*, W_j, j \in [1:i] \right)$ can be thereby expressed as

$$\operatorname{cov}\left(Y_{i+1} \middle| \tilde{Y}_{i+1,\gamma}, Y_{i,\gamma}^*, W_j, j \in [1:i] \right) = \frac{\gamma}{1-\gamma}\boldsymbol{M}_{11}^{-1}\left(\left(\boldsymbol{\Delta}_i + \boldsymbol{K}_{i+1}\right)^{-1} + (1-\gamma)\left(\boldsymbol{K}_i - \boldsymbol{K}_{i+1}\right)^{-1} - \gamma J\left(\tilde{Y}_{i+1,\gamma} \middle| Y_{i,\gamma}^*, W_j, j \in [1:i] \right) \right)\boldsymbol{M}_{11}^{-1}. \tag{187}$$

Equivalently, we can write

$$\boldsymbol{M}_{11}\left(\frac{1}{\gamma}\operatorname{cov}\left(Y_{i+1} \middle| \tilde{Y}_{i+1,\gamma}, Y_{i,\gamma}^*, W_j, j \in [1:i] \right) - \boldsymbol{M}_{11}^{-1} \right)\boldsymbol{M}_{11} = \frac{\gamma}{1-\gamma}\left(\boldsymbol{\Delta}_i + \boldsymbol{K}_{i+1}\right)^{-1} - \frac{\gamma}{1-\gamma}J\left(\tilde{Y}_{i+1,\gamma} \middle| Y_{i,\gamma}^*, W_j, j \in [1:i] \right). \tag{188}$$

Substituting (188) into (183) gives

$$\frac{d}{d\gamma} h\left(\tilde{Y}_{i+1,\gamma}, Y_{i,\gamma}^* \middle| W_j, j \in [1:i] \right) = \frac{1}{2(1-\gamma)}\operatorname{tr}\left\{ \boldsymbol{M}_{11}^{-1}\left(J\left(\tilde{Y}_{i+1,\gamma} \middle| Y_{i,\gamma}^*, W_j, j \in [1:i] \right) - \left(\boldsymbol{\Delta}_i + \boldsymbol{K}_{i+1}\right)^{-1} \right) \right\}. \tag{189}$$

Furthermore, it follows by the Woodbury matrix inversion lemma that

$$\left(\left(\boldsymbol{\Delta}_i + \boldsymbol{K}_{i+1}\right)^{-1} + \left(\boldsymbol{K}_i - \boldsymbol{K}_{i+1}\right)^{-1} \right)^{-1}$$

$$= \boldsymbol{K}_{i+1}\left(\boldsymbol{K}_{i+1} - \boldsymbol{K}_{i+1}\left(\boldsymbol{K}_{i+1} - \boldsymbol{K}_i\right)^{-1}\boldsymbol{K}_{i+1} - \boldsymbol{K}_{i+1} \right.$$

$$+\boldsymbol{K}_{i+1}\left(\boldsymbol{\Delta}_i+\boldsymbol{K}_{i+1}\right)^{-1}\boldsymbol{K}_{i+1}\Big)^{-1}\boldsymbol{K}_{i+1} \quad (190)$$

$$=\boldsymbol{K}_{i+1}\left(\left(\boldsymbol{K}_{i+1}^{-1}-\boldsymbol{K}_i^{-1}\right)^{-1}-\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_{i+1}^{-1}\right)^{-1}\right)^{-1}\boldsymbol{K}_{i+1} \quad (191)$$

$$=\boldsymbol{K}_{i+1}\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_{i+1}^{-1}\right)\Big(\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_i^{-1}\right)^{-1} -\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_{i+1}^{-1}\right)^{-1}\Big)\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_{i+1}^{-1}\right)\boldsymbol{K}_{i+1}. \quad (192)$$

So we can rewrite (189) as

$$\frac{d}{d\gamma}h\left(\tilde{Y}_{i+1,\gamma},Y_{i,\gamma}^*\middle|W_j,j\in[1:i]\right) =\frac{1}{2(1-\gamma)}\operatorname{tr}\Big\{\Big(\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_i^{-1}\right)^{-1}-\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_{i+1}^{-1}\right)^{-1}\Big) \Big(\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_{i+1}^{-1}\right)\boldsymbol{K}_{i+1}J\left(\tilde{Y}_{i+1,\gamma}\middle|Y_{i,\gamma}^*,W_j,j\in[1:i]\right) \boldsymbol{K}_{i+1}\left(\boldsymbol{\Delta}_i^{-1}+\boldsymbol{K}_{i+1}^{-1}\right)-\boldsymbol{\Delta}_i^{-1}\left(\boldsymbol{\Delta}_i+\boldsymbol{K}_{i+1}\right)\boldsymbol{\Delta}_i^{-1}\Big)\Big\}. \quad (193)$$

APPENDIX D
PROOF OF THEOREM 1 VIA THE DOUBLING TRICK

During the reviewing process, one anonymous reviewer provided an alternative proof of our main result based on the doubling/rotation method, which is included here with his/her kind permission.

A. Definitions

For the sake of simplifying notations, random vector $(X_1, X_2, \ldots, X_i)$ is written as $X_{[i]}$ in this appendix. Let

$$s(W_{[L]}) \triangleq \mu_1 I(X;W_1|Y_1)+\sum_{i=2}^{L}\mu_i I(X;W_i|W_{[i-1]},Y_i) \quad (194)$$

$$=\sum_{i=1}^{L}\left(\mu_i(h(Y_i|W_{[i]})-h(X|W_{[i]})) -\mu_{i+1}(h(Y_{i+1}|W_{[i]})-h(X|W_{[i]}))\right) +\mu_L\left(h(Y_L|W_{[L]})-h(X|W_{[L]})\right). \quad (195)$$

Introduce random variable Q such that

$$(Q,W_{[L]})\to X\to Y_L\to\ldots\to Y_1 \quad (196)$$

form a Markov chain. Similarly to (194), let

$$s(W_{[L]}|Q)\triangleq\mu_1 I(X;W_1|Y_1,Q) +\sum_{i=2}^{L}\mu_i I(X;W_i|W_{[i-1]},Y_i,Q). \quad (197)$$

We further define the lower convex envelop of $s(W_{[L]})$ as

$$S(W_{[L]})\triangleq\inf_{p(q|x,w_{[L]})}s(W_{[L]}|Q). \quad (198)$$

We also define

$$S(W_{[L]}|Q)\triangleq\sum_q p(q)S(W_{[L]}|Q=q) \quad (199)$$

for Q (over a finite alphabet) and its natural extension for an arbitrary Q.

Remark 5: Since $S(W_{[L]})$ is convex in $p(w_{[L]}|x)$, we have

$$S(W_{[L]}|Q)\geq S(W_{[L]}) \quad (200)$$

by Jensen's inequality.

The rate-distortion problem of Theorem 1 can be reformulated as finding the optimal random vectors $W_{[L]}$ for

$$V^*(\boldsymbol{D}_{[L]})\triangleq\inf_{p(w_{[L]}|x)}S(W_{[L]}) \quad (201)$$

$$=\inf_{p(q,w_{[L]}|x)}s(W_{[L]}|Q), \quad (202)$$

where $p(w_{[L]}|x)$ satisfies $\operatorname{cov}(X|Y_i,W_{[i]})\preceq\boldsymbol{D}_i$ for any $i\in[1:L]$.

Lemma 9: There exists a pair of random variables $(W_{*,[L]},Q_*)$ with $\operatorname{cov}(X|Y_i,W_{*,[i]})\preceq\boldsymbol{D}_i$, $i\in[1,L]$, such that

$$V^*(\boldsymbol{D}_{[L]})=s(W_{*,[L]}|Q_*) \quad (203)$$

Proof: We can assume that the conditional law of $(X,Y_{[L]})$ has zero mean for every Q_*. Because the centering condition on each $Q_*=q_*$ does not change the mutual information terms and hence $S(W_{*,[L]}|Q_*)$ remains the same. The existence of a minimizer and the cardinality bound on Q_* follow by the argument in [38, Appendix II.A]. ■

B. The Doubling Trick

Let

$$(W_{a,[L]},W_{b,[L]},X_a,X_b,Y_{a,[L]},Y_{b,[L]})\sim p(w_{a,[L]},x_a,y_{a,[L]})\times p(w_{b,[L]},x_b,y_{b,[L]}) \quad (204)$$

be two i.i.d copies of $(W_{[L]},X,Y_{[L]})$ with

$$(W_{a,[L]})\to X_a\to Y_{a,L}\to Y_{a,L-1}\ldots\to Y_{a,1}, \quad (205)$$

$$(W_{b,[L]})\to X_b\to Y_{b,L}\to Y_{b,L-1}\ldots\to Y_{b,1}. \quad (206)$$

The above Markov chains still hold when conditioned on (Q_a,Q_b) and

$$(Q_a,W_{a,[L]})\to X_a\to Y_{a,L}\to Y_{a,L-1}\ldots\to Y_{a,1}, \quad (207)$$

$$(Q_b,W_{b,[L]})\to X_b\to Y_{b,L}\to Y_{b,L-1}\ldots\to Y_{b,1} \quad (208)$$

are also satisfied.

Given

$$(X_a,X_b)\sim p(x_a)\times p(x_b),$$

we define $s(W_{a,[L]},W_{b,[L]})$, in a similar fashion as above,

$$s(W_{a,[L]},W_{b,[L]}) \triangleq\sum_{i=1}^{L}\Big(\mu_i(h(Y_{a,i},Y_{b,i}|W_{a,[i]},W_{b,[i]}) -h(X_a,X_b|W_{a,[i]},W_{b,[i]})) -\mu_{i+1}(h(Y_{a,i+1},Y_{b,i+1}|W_{a,[i]},W_{b,[i]}) -h(X_a,X_b|W_{a,[i]},W_{b,[i]}))\Big)$$

$$+ \mu_L \left(h(Y_{a,L}, Y_{b,L}|W_{a,[L]}, W_{b,[L]}) - h(X_a, X_b|W_{a,[L]}, W_{b,[L]})\right). \quad (209)$$

We also define the quantities $s(W_{a,[L]}, W_{b,[L]}|Q_a, Q_b)$, $S(W_{a,[L]}, W_{b,[L]})$, $s(W_{a,[L]}, W_{b,[L]}|Q_a, Q_b)$ similarly. The proof of the following lemma can be found in Appendix D-C.

Lemma 10: The following inequality holds for $(X_a, X_b, Y_{a,[L]}, Y_{b,[L]}) \sim p(x_a, y_{a,[L]}) \times p(x_b, y_{b,[L]})$:

$$S(W_{a,[L]}, W_{b,[L]}) \geq S(W_{a,[L]}) + S(W_{b,[L]}). \quad (210)$$

Furthermore, if a particular tuple $(W_{*,[L]}, Q_*)$ satisfies

$$\begin{aligned} &s(W_{a,*,[L]}, W_{b,*,[L]}|Q_{a,*}, Q_{b,*}) \\ &= S(W_{a,*,[L]}, W_{b,*,[L]}) \quad (211) \\ &= S(W_{a,*,[L]}) + S(W_{b,*,[L]}), \quad (212) \end{aligned}$$

the following facts must be true,

1)

$$I(X_{a,*}; X_{b,*}|Y_{a,*,i}, Y_{b,*,i}, W_{a,*,[i]}, W_{b,*,[i]}, Q_{a,*}, Q_{b,*}) = 0, \quad i \in [1:L]; \quad (213)$$

2)

$$\begin{aligned} &S(W_{a,*,[L]}) \\ &= s(W_{a,[L]}|Y_{b,[L]}, W_{b,[L]}, Q_{a,*}, Q_{b,*}) \quad (214) \\ &= s(W_{b,[L]}|Y_{a,[L]}, W_{a,[L]}, Q_{a,*}, Q_{b,*}), \quad (215) \end{aligned}$$

where

$$\begin{aligned} &s(W_{a,[L]}|Y_{b,[L]}, W_{b,[L]}) \\ &\triangleq \sum_{i=1}^{L-1} \Big(\mu_i (h(Y_{a,i}|W_{a,[i]}, Y_{b,i}, W_{b,[i]}) \\ &\quad - h(X_a|W_{a,[i]}), Y_{b,i}, W_{b,[i]}) \\ &\quad - \mu_{i+1}(h(Y_{a,i+1}|W_{a,[i]}, Y_{b,i}, W_{b,[i]}) \\ &\quad - h(X_a|W_{a,[i]}|, Y_{b,i}, W_{b,[i]}))\Big) \\ &\quad + \mu_L(h(Y_{a,L}|W_{a,[L]}, Y_{b,L}, W_{b,[L]})) \\ &\quad - h(X_a|W_{a,[L]}, Y_{b,L}, W_{b,[L]})). \quad (216) \end{aligned}$$

$s(W_{b,[L]}|Y_{a,[L]}, W_{a,[L]})$, $s(W_{a,[L]}|Y_{b,[L]}, W_{b,[L]}, Q_{a,*}, Q_{b,*})$ and $s(W_{b,[L]}|Y_{a,[L]}, W_{a,[L]}, Q_{a,*}, Q_{b,*})$ are defined similarly.

For simplifying notations, we denote

$$Z_+ = \frac{1}{\sqrt{2}}(Z_a + Z_b), \quad Z_- = \frac{1}{\sqrt{2}}(Z_a - Z_b), \quad (217)$$

where (Z_a, Z_b) are two i.i.d copies of random variable Z. In a similar manner, we define $s(W_{+,[L]})$ as

$$s(W_{+,[L]}) \triangleq \mu_1 I(X_+; W_{+,1}|Y_{+,1}) + \sum_{i=2}^{L} \mu_i I(X_{+,1}; W_{+,i}|W_{+,[i-1]}, Y_{+,i}). \quad (218)$$

Furthermore, $s(W_{-,[L]})$, $s(W_{+,[L]}, W_{-,[L]})$, $S(W_{+,[L]})$, $S(W_{-,[L]})$ and $S(W_{+,[L]}, W_{-,[L]})$ can be defined similarly.

The proof of the following lemma can be found in Appendix D-D.

Lemma 11: For $(W_{*,[L]}, Q_*) \sim p_*(w_{[L]}, q)$ that attains $V^*(\boldsymbol{D}_{[L]})$ and $(W_{a,[L]}, W_{b,[L]}, Q_a, Q_b) \sim p_*(w_{a,[L]}, q_a) \times p_*(w_{b,[L]}, q_b)$, the following holds:

1)

$$I(X_+; X_-|Y_{+,i}, Y_{-,i}, W_{+,[i]}, W_{-,[i]}, Q_a, Q_b) = 0, \quad i \in [1:L]; \quad (219)$$

2)

$$V^*(\boldsymbol{D}_{[L]}) = S(W_{+,[L]}|Y_{-,[L]}) = S(W_{-,[L]}|Y_{+,[L]}), \quad (220)$$

where

$$\begin{aligned} &s(W_{+,[L]}|Y_{-,[L]}) \\ &\triangleq \sum_{i=1}^{L-1} \Big(\mu_i (h(Y_{+,i}|W_{+,[i]}, Y_{-,i}) - h(X_+|W_{+,[i]}), Y_{-,i}) \\ &\quad - \mu_{i+1}(h(Y_{+,i+1}|W_{+,[i]}, Y_{-,i}) - h(X_+|W_{+,[i]}|, Y_{-,i}))\Big) \\ &\quad + \mu_L(h(Y_{+,L}|W_{+,[L]}, Y_{-,L})) - h(X_+|W_{+,[L]}, Y_{-,L})), \quad (221) \end{aligned}$$

and $s(W_{-,[L]}|Y_{+,[L]})$ is defined similarly.

Now we are in a position to establish the following result, which will complete the proof of Gaussian optimality in Theorem 1.

Theorem 4: There exist legitimate auxiliary random objects $W_{*,[L]}$ jointly Gaussian with $(X, Y_{[L]})$ such that

$$V^*(\boldsymbol{D}_{[L]}) = s(W_{*,[L]}). \quad (222)$$

Proof: The optimal value V_* defined in (201) can be achieved by a suitable $(W_{*,[L]}, Q_*) \sim p_*(w, q)$ according to Lemma 9. For any pair $(X_a, X_b, Y_{a,[L]}, Y_{b,[L]}) \sim p(x_a, y_{a,[L]}) \times p(x_b, y_{b,[L]})$ satisfying conditions in Lemma 10, $(X_a, Y_{a,[L]})$ and $(X_b, Y_{b,[L]})$ are conditionally independent zero mean random variables given $(W_{*,a,[L]}, W_{*,b,[L]}, Q_{*,a}, Q_{*,b})$. So by Lemma 11, conditioned on $(W_{a,[L]}, W_{b,[L]}, Q_a, Q_b)$ the following Markov chains hold:

$$X_+ \to (Y_{+,j}, Y_{-,j}) \to X_-. \quad (223)$$

Since

$$Y_{+,j} \to (X_+, Y_{-,j}) \to X_-, \quad Y_{-,j} \to (X_-, Y_{+,j}) \to X_+, \quad (224)$$

it follows by double Markovity that

$$(X_+, Y_{+,j}) \to Y_{-,j} \to X_-, \quad (X_-, Y_{-,j}) \to Y_{+,j} \to X_+. \quad (225)$$

Next, invoking double Markovity (conditioned on $(W_{a,[L]}, W_{b,[L]}, Q_a, Q_b)$ with

$$Y_{+,j} \to X_+ \to X_-, \quad Y_{-,j} \to X_- \to X_+, \quad (226)$$

we can deduce that X_+ and X_- are independent conditioned on $(W_{a,[L]}, W_{b,[L]}, Q_a, Q_b)$. According to the property of Gaussian random variables in [38, Theorem 3] and the proof method in [38, Appendix I-A], we can conclude that

$(X|W_{[L]}, Q)$ has a Gaussian distribution. Since Q is arbitrary and the covariance matrix of $(X|W_{[L]}, Q)$ is the same for different Q, it follows that $(X|W_{[L]})$ is Gaussian, which further implies that $W_{[L]}$ can be assumed to be jointly Gaussian with X. This completes the proof. ■

C. Proof of Lemma 10

For any auxiliary random variables Q satisfying (196), (205) and (206), (Q_a, Q_b) is denoted as $\boldsymbol{Q}$ for simplicity. We can expand $S(W_{a,[L]}, W_{b,[L]})$ as

$$\begin{aligned}S(W_{a,[L]}, W_{b,[L]}) &= s(W_{a,[L]}, W_{b,[L]}|\boldsymbol{Q}) &(227)\\ &= \sum_{i=1}^{L-1}\Big((\mu_i(h(Y_{a,i}, Y_{b,i}|W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\qquad - h(X_a, X_b|W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})) &(228)\\ &\quad - \mu_{i+1}(h(Y_{a,i+1}, Y_{b,i+1}|W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\qquad - h(X_a, X_b|W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q}))\Big) &(229)\\ &\quad + \mu_L(h(Y_{a,L}, h(Y_{b,L}|W_{a,[L]}, W_{b,[L]}, \boldsymbol{Q})\\ &\quad - h(X_a, X_b|W_{a,[L]}, W_{b,[L]}, \boldsymbol{Q})). &(230)\end{aligned}$$

First consider the terms in (228):

$$\begin{aligned}&h(Y_{a,i}, Y_{b,i}|W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q}) - h(X_a, X_b|W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &= h(Y_{a,i}|Y_{b,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\quad + h(Y_{b,i}|Y_{a,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\quad + I(Y_{a,i}; Y_{b,i}|W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\quad - h(X_a|W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q}) - h(X_b|W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\quad + I(X_{a,i}; X_{b,i}|W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q}) &(231)\\ &= h(Y_{a,i}|Y_{b,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\quad + h(Y_{b,i}|Y_{a,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\quad - h(X_{a,i}|Y_{b,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\quad - h(X_b|Y_{a,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\quad + I(X_a; X_b|W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\quad + I(Y_{a,i}; Y_{b,i}|W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\quad - I(X_a; Y_{b,i}|W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\quad - I(X_b; Y_{a,i}|W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q}) &(232)\\ &= h(Y_{a,i}|Y_{b,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\quad + h(Y_{b,i}|Y_{a,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\quad - h(X_a|Y_{b,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\quad - h(X_b|Y_{a,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\quad + I(X_a; X_b|Y_{a,i}; Y_{b,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\quad + I(Y_{a,i}; Y_{b,i}|X_b, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\quad - I(X_a; Y_{b,i}|X_b, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\quad - I(Y_{a,i}; X_b|Y_{b,i}, X_a, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q}). &(233)\end{aligned}$$

In view of (205), (206) and the definition of X_a, X_b, the following Markov chains hold (conditioned on $(W_{a,[i]}, W_{b,[i]}, Q_a, Q_b)$):

$$Y_{b,i} \to X_b \to (X_a, Y_{a,i}), \quad Y_{a,i} \to X_a \to (X_b, Y_{b,i}). \qquad (234)$$

Therefore, we have

$$\begin{aligned}&I(Y_{a,i}; Y_{b,i}|X_b, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &= I(X_a; Y_{b,i}|X_b, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q}) &(235)\\ &= I(Y_{a,i}; X_b|Y_{b,i}, X_a, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q}) &(236)\\ &= 0, &(237)\end{aligned}$$

which yields

$$\begin{aligned}&h(Y_{a,i}, Y_{b,i}|W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q}) - h(X_a, X_b|W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &= h(Y_{a,i}|Y_{b,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\quad + h(Y_{b,i}|Y_{a,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\quad - h(X_a|Y_{b,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\quad - h(X_b|Y_{a,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\quad + I(X_a; X_b|Y_{a,i}; Y_{b,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q}). &(238)\end{aligned}$$

Similarly, the terms in (229) can be simplified using the following Markov chains (conditioned on $(W_{a,[i]}, W_{b,[i]}, Q_a, Q_b)$):

$$\begin{aligned}Y_{b,i} \to Y_{b,i+1} \to X_b \to (X_a, Y_{a,i}, Y_{a,i+1}), &\qquad (239)\\ Y_{a,i} \to Y_{a,i+1} \to X_a \to (X_b, Y_{b,i}, Y_{b,i+1}). &\qquad (240)\end{aligned}$$

Specifically, we have

$$\begin{aligned}&h(Y_{a,i+1}, Y_{b,i+1}|W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q}) - h(X_a, X_b|W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &= h(Y_{a,i+1}|W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\quad + h(Y_{b,i+1}|W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\quad - h(X_a|W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q}) - h(X_b|W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\quad + I(X_a; X_b|W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\quad - I(Y_{a,i+1}; Y_{b,i+1}|W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &= h(Y_{a,i+1}|Y_{b,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\quad + h(Y_{b,i+1}|Y_{a,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\quad - h(X_a|Y_{b,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\quad - h(X_b|Y_{a,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\quad + I(X_a; X_b|Y_{a,i}, Y_{b,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\quad - I(Y_{a,i+1}; Y_{b,i+1}|Y_{a,i}, Y_{b,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q}). &(241)\end{aligned}$$

For the terms in (230), it can be shown using the same method that

$$\begin{aligned}&h(Y_{a,L}, Y_{b,L}|W_{a,[L]}, W_{b,[L]}, \boldsymbol{Q}) - h(X_a, X_b|W_{a,[L]}, W_{b,[L]}, \boldsymbol{Q})\\ &= h(Y_{a,L}|W_{a,[L]}, W_{b,[L]}, \boldsymbol{Q}) + h(Y_{b,L}|W_{a,[L]}, W_{b,[L]}, \boldsymbol{Q})\\ &\quad - h(X_a|W_{a,[L]}, W_{b,[L]}, \boldsymbol{Q}) - h(X_b|W_{a,[L]}, W_{b,[L]}, \boldsymbol{Q})\\ &\quad + I(X_a; X_b|Y_{a,L}; Y_{b,L}, W_{a,[L]}, W_{b,[L]}, \boldsymbol{Q}). &(242)\end{aligned}$$

Combining (238), (241), and (242) gives

$$\begin{aligned}&S(W_{a,[L]}, W_{b,[L]})\\ &= \sum_{i=1}^{L-1}\Big(\mu_i(h(Y_{a,i}|Y_{b,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\qquad + h(Y_{b,i}|Y_{a,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\qquad - h(X_a|Y_{b,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\\ &\qquad - h(X_b|Y_{a,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q}))\\ &\quad - \mu_{i+1}(h(Y_{a,i+1}|Y_{b,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\end{aligned}$$

$$+ h(Y_{b,i+1}|Y_{a,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q}) - h(X_a|Y_{b,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q}) - h(X_b|Y_{a,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q}))\Big)$$

$$+ \mu_L(h(Y_{a,L}|Y_{b,L}, W_{a,[L]}, W_{b,[L]}, \boldsymbol{Q}) + h(Y_{b,L}|Y_{a,L}, W_{a,[L]}, W_{b,[L]}, \boldsymbol{Q}) - h(X_a|Y_{b,L}, W_{a,[L]}, W_{b,[L]}, \boldsymbol{Q}) - h(X_b|Y_{a,L}, W_{a,[L]}, W_{b,[L]}, \boldsymbol{Q}))$$

$$+ \sum_{i=1}^{L-1}\Big(\mu_i I(X_a; X_b|Y_{a,i}; Y_{b,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q}) - \mu_{i+1}(I(X_a; X_b|Y_{a,i}, Y_{b,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q}) - I(Y_{a,i+1}; Y_{b,i+1}|Y_{a,i}, Y_{b,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q}))\Big)$$

$$+ \mu_L I(X_a; X_b|Y_{a,L}; Y_{b,L}, W_{a,[L]}, W_{b,[L]}, \boldsymbol{Q}) \quad (243)$$

$$= s(W_{a,[L]}|Y_{b,[L]}, W_{b,[L]}, \boldsymbol{Q}) + s(W_{b,[L]}|Y_{a,[L]}, W_{a,[L]}, \boldsymbol{Q}) + \sum_{i=1}^{L-1}\Big((\mu_i - \mu_{i+1})I(X_a; X_b|Y_{a,i}, Y_{b,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q}) \; \mu_{i+1}I(Y_{a,i+1}; Y_{b,i+1}|Y_{a,i}, Y_{b,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q})\Big) + \mu_L I(X_a; X_b|Y_{a,L}; Y_{b,L}, W_{a,[L]}, W_{b,[L]}, \boldsymbol{Q}) \quad (244)$$

$$\overset{(a)}{\geq} s(W_{a,[L]}|Y_{b,[L]}, W_{b,[L]}, \boldsymbol{Q}) + s(W_{b,[L]}|Y_{a,[L]}, W_{a,[L]}, \boldsymbol{Q}) \quad (245)$$

$$\overset{(b)}{\geq} S(W_{a,[L]}) + S(W_{b,[L]}), \quad (246)$$

where (a) follows from $\mu_i \geq \mu_{i+1}$ and the nonnegativity of mutual information while (b) is due to the fact that $S(W_{[L]})$ is the lower convex envelope of $s(W_{[L]})$ (see (198)).

D. Proof of Lemma 11

In view of the definition of $V^*(\mathbf{D}_{[L]})$ and the assumption that $(W_{*,[L]}, Q_*) \sim p_*(w_{[L]}, q)$ attains $V^*(\boldsymbol{D}_{[L]})$, we have

$$2V^*(\boldsymbol{D}_{[L]}) \overset{(a)}{=} s(W_{a,[L]}|Q_a) + s(W_{b,[L]}|Q_b) \quad (247)$$

$$\overset{(b)}{=} s(W_{a,[L]}, W_{b,[L]}|Q_a, Q_b) \quad (248)$$

$$\overset{(c)}{=} s(W_{+,[L]}, W_{-,[L]}|Q_a, Q_b) \quad (249)$$

$$\overset{(d)}{\geq} S(W_{+,[L]}, W_{-,[L]}) \quad (250)$$

$$\overset{(e)}{\geq} S(W_{+,[L]}|Y_{-,[L]}) + S(W_{-,[L]}|Y_{+,[L]}) \quad (251)$$

$$\overset{(f)}{\geq} S(W_{+,[L]}) + S(W_{-,[L]}) \quad (252)$$

$$\overset{(g)}{\geq} V^*(\boldsymbol{D}_{[L]}) + V^*(\boldsymbol{D}_{[L]}) = 2V^*(\boldsymbol{D}_{[L]}). \quad (253)$$

The justification for each step is given below:

- (a) holds because $p_*(w_{[L]}, q)$ achieves $V^*(\boldsymbol{D}_{[L]})$.
- (b) holds because of the independence of $(X_a, Y_{a,[L]}, W_{a,[L]}, Q_a)$ and $(X_b, Y_{b,[L]}, W_{b,[L]}, Q_b)$.
- (c) holds since differential entropy is invariant under unitary transformation.
- (d) follows from the definition of $S(W_{[L]})$ in (198).
- (e) is a consequence of Lemma 10 and the fact that mutual information is preserved under bijective transformation. It is noticed that

$$S(W_{+,[L]}, W_{-,[L]}) \geq \sum_{i=1}^{L-1}\Big(\mu_i(h(Y_{+,i}|Y_{-,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q}) + h(Y_{-,i}|Y_{+,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q}) - h(X_+|Y_{-,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q}) - h(X_-|Y_{+,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q}))$$
$$- \mu_{i+1}(h(Y_{+,i+1}|Y_{-,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q}) + h(Y_{-,i+1}|Y_{+,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q}) - h(X_+|Y_{-,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q}) - h(X_-|Y_{+,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q}))\Big)$$
$$+ \mu_L(h(Y_{+,L}|Y_{-,L}, W_{a,[L]}, W_{b,[L]}, \boldsymbol{Q}) + h(Y_{-,L}|Y_{+,L}, W_{a,[L]}, W_{b,[L]}, \boldsymbol{Q}) - h(X_+|Y_{-,L}, W_{a,[L]}, W_{b,[L]}, \boldsymbol{Q}) - h(X_-|Y_{+,L}, W_{a,[L]}, W_{b,[L]}, \boldsymbol{Q}))$$
$$= s(W_{+,[L]}|Y_{-,L}, W_{-,[L]}, \boldsymbol{Q}) + s(W_{-,[L]}|Y_{+,L}, W_{+,[L]}, \boldsymbol{Q}) \geq S(W_{+,[L]}) + S(W_{-,[L]}). \quad (254)$$

Defining

$$\widetilde{W}_i^+ = (Y_{-,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q}) \quad i \in [1:L], \quad (255)$$
$$\widetilde{W}_i^- = (Y_{+,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q}) \quad i \in [1:L], \quad (256)$$

we can observe that

$$\mathrm{cov}(X_+|Y_{+,i}, \widetilde{W}_i^+) = \mathrm{cov}(X_+|Y_{+,i}, Y_{-,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q}) \quad (257)$$
$$\preceq \boldsymbol{D}_i, \quad i \in [1:L], \quad (258)$$
$$\mathrm{cov}(X_-|Y_{-,i}, \widetilde{W}_i^-) = \mathrm{cov}(X_-|Y_{-,i}, Y_{+,i}, W_{a,[i]}, W_{b,[i]}, \boldsymbol{Q}) \quad (259)$$
$$\preceq \boldsymbol{D}_i, \quad i \in [1:L]. \quad (260)$$

- (f) follows from Remark 5.
- (g) follows form the definition of $V^*(\boldsymbol{D}_{[L]})$ in (201).

Since the extremes match, all inequalities should be equalities. Therefore, the conditions in Lemma 11 must be satisfied.

Acknowledgment

The authors would like to thank the associate editor and the anonymous reviewers for their valuable comments and suggestions, which helped to improve the readability of the article. In addition, one reviewer provided an alternative proof of our main result based on the doubling trick and kindly permitted us to include it in the appendix. Yinfei Xu is also grateful to Guojun Chen for helpful discussions on perturbation and factorization strategies.

REFERENCES

[1] Y. Xu, X. Guang, and J. Lu, "Vector Gaussian successive refinement with degraded side information," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2019, pp. 1832–1836.

[2] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 4, pp. 471–480, Jul. 1973.

[3] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.

[4] C. Heegard and T. Berger, "Rate distortion when side information may be absent," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 6, pp. 727–734, Nov. 1985.

[5] A. H. Kaspi, "Rate-distortion function when side-information may be present at the decoder," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 2031–2034, Nov. 1994.

[6] R. Timo, T. Chan, and A. Grant, "Rate distortion with side-information at many decoders," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 5240–5257, Aug. 2011.

[7] S. Watanabe, "The rate-distortion function for product of two sources with side-information at decoders," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5678–5691, Sep. 2013.

[8] M. Benammar and A. Zaidi, "Rate-distortion function for a heegard-berger problem with two sources and degraded reconstruction sets," *IEEE Trans. Inf. Theory*, vol. 62, no. 9, pp. 5080–5092, Sep. 2016.

[9] Y. Steinberg and N. Merhav, "On successive refinement for the Wyner-Ziv problem," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1636–1654, Aug. 2004.

[10] C. Tian and S. N. Diggavi, "On multistage successive refinement for Wyner–Ziv source coding with degraded side informations," *IEEE Trans. Inf. Theory*, vol. 53, no. 8, pp. 2946–2960, Aug. 2007.

[11] P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise (Corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 2, pp. 279–280, Mar. 1974.

[12] H. Weingarten, T. Liu, S. Shamai, Y. Steinberg, and P. Viswanath, "The capacity region of the degraded multiple-input multiple-output compound broadcast channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5011–5023, Nov. 2009.

[13] H. D. Ly, T. Liu, and Y. Liang, "Multiple-input multiple-output Gaussian broadcast channels with common and confidential messages," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5477–5487, Nov. 2010.

[14] E. Ekrem and S. Ulukus, "Capacity region of Gaussian MIMO broadcast channels with common and confidential messages," *IEEE Trans. Inf. Theory*, vol. 58, no. 9, pp. 5669–5680, Sep. 2012.

[15] E. Ekrem and S. Ulukus, "Capacity-equivocation region of the Gaussian MIMO wiretap channel," *IEEE Trans. Inf. Theory*, vol. 58, no. 9, pp. 5699–5710, Sep. 2012.

[16] R. Liu, T. Liu, H. V. Poor, and S. Shamai (Shitz), "New results on multiple-input multiple-output broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 59, no. 3, pp. 1346–1359, Mar. 2013.

[17] E. Ekrem and S. Ulukus, "Secure lossy transmission of vector Gaussian sources," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5466–5487, Jun. 2013.

[18] H.-F. Chong and Y.-C. Liang, "The capacity region of the class of three-receiver Gaussian MIMO multilevel broadcast channels with two-degraded message sets," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 42–53, Jan. 2014.

[19] H.-F. Chong and Y.-C. Liang, "An extremal inequality and the capacity region of the degraded compound Gaussian MIMO broadcast channel with multiple users," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6131–6143, Oct. 2014.

[20] A. Khisti and T. Liu, "Private broadcasting over independent parallel channels," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5173–5187, Sep. 2014.

[21] A. S. Motahari and A. K. Khandani, "Capacity bounds for the Gaussian interference channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 620–643, Feb. 2009.

[22] X. Shang, G. Kramer, and B. Chen, "A new outer bound and the noisy-interference sum-rate capacity for Gaussian interference channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 689–699, Feb. 2009.

[23] V. S. Annapureddy and V. V. Veeravalli, "Gaussian interference networks: Sum capacity in the low-interference regime and new outer bounds on the capacity region," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3032–3050, Jul. 2009.

[24] 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.

[25] 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.

[26] Y. Xu and Q. Wang, "A perturbation proof of the vector Gaussian one-help-one problem," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 2013, pp. 1372–1376.

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

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

[29] Y. Xu and Q. Wang, "Rate region of the vector Gaussian CEO problem with the trace distortion constraint," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 1823–1835, Apr. 2016.

[30] Y. Ugur, I. E. Aguerri, and A. Zaidi, "Vector Gaussian CEO problem under logarithmic loss and applications," *IEEE Trans. Inf. Theory*, vol. 66, no. 7, pp. 4183–4202, Jul. 2020.

[31] S. Watanabe and Y. Oohama, "Secret key agreement from vector Gaussian sources by rate limited public communication," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 3, pp. 541–550, Sep. 2011.

[32] 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.

[33] 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.

[34] 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.

[35] Y. Xu, J. Chen, and Q. Wang, "The sum rate of vector Gaussian multiple description coding with tree-structured covariance distortion constraints," *IEEE Trans. Inf. Theory*, vol. 63, no. 10, pp. 6547–6560, Oct. 2017.

[36] 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.

[37] L. Song, J. Chen, and C. Tian, "Broadcasting correlated vector Gaussians," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2465–2477, May 2015.

[38] Y. Geng and C. Nair, "The capacity region of the two-receiver Gaussian vector broadcast channel with private and common messages," *IEEE Trans. Inf. Theory*, vol. 60, no. 4, pp. 2087–2104, Apr. 2014.

[39] J. Wang and J. Chen, "A monotone path proof of an extremal result for long Markov chains," *Entropy*, vol. 21, no. 3, p. 276, Mar. 2019. [Online]. Available: http://www.mdpi.com/1099-4300/21/3/276

[40] T. A. Courtade, "A strong entropy power inequality," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 2173–2192, Apr. 2018.

[41] 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.

[42] T. Liu and P. Viswanath, "An extremal inequality motivated by multiterminal information-theoretic problems," *IEEE Trans. Inf. Theory*, vol. 53, no. 5, pp. 1839–1851, May 2007.

[43] A. Dembo, T. M. Cover, and J. A. Thomas, "Information theoretic inequalities," *IEEE Trans. Inf. Theory*, vol. 37, no. 6, pp. 1501–1518, Nov. 1991.

[44] D. Guo, S. Verdú, and S. Shamai (Shitz), "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.

[45] D. P. Bertsekas, A. Nedić, and A. E. Ozdaglar, *Convex Analysis and Optimization*. Belmont, NY, USA: Athena Scientific, 2003.

[46] V. Anantharam, V. Jog, and C. Nair, "Unifying the brascamp-lieb inequality and the entropy power inequality," 2019, *arXiv:1901.06619*. [Online]. Available: http://arxiv.org/abs/1901.06619

[47] L. Wang and M. Madiman, "Beyond the entropy power inequality, via rearrangements," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5116–5137, Sep. 2014.

[48] O. Rioul, "Yet another proof of the entropy power inequality," *IEEE Trans. Inf. Theory*, vol. 63, no. 6, pp. 3595–3599, Jun. 2017.

[49] C. Tian and S. N. Diggavi, "Side-information scalable source coding," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5591–5608, Dec. 2008.

[50] S. Unal and A. B. Wagner, "Vector Gaussian rate-distortion with variable side information," *IEEE Trans. Inf. Theory*, vol. 63, no. 8, pp. 5162–5178, Aug. 2017.

[51] D. P. Palomar and S. Verdú, “Gradient of mutual information in linear vector Gaussian channels,” *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 141–154, Jan. 2006.
[52] R. Zamir, “A proof of the Fisher information inequality via a data processing argument,” *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1246–1250, May 1998.
[53] O. Rioul, “Information theoretic proofs of entropy power inequalities,” *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 33–55, Jan. 2011.

Yinfei Xu (Member, IEEE) received the B.E. and Ph.D. degrees in information engineering from Southeast University, Nanjing, China, in 2008 and 2016, respectively. He was a Visiting Student with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada, from July 2014 to January 2015. He was a Research Assistant and a Post-Doctoral Fellow with the Institute of Network Coding, The Chinese University of Hong Kong, Hong Kong, from March 2016 to July 2017. Since August 2017, he has been with the School of Information Science and Engineering, Southeast University, where he is currently an Assistant Professor. His research interests include information theory, signal processing, and wireless communications.

Xuan Guang (Member, IEEE) received the Ph.D. degree in mathematics from Nankai University, Tianjin, China, in 2012, and the joint Ph.D. degree from Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA, USA, in August 2012.

He was a Post-Doctoral Fellow under the Hong Kong Scholars Program with the Institute of Network Coding (INC), The Chinese University of Hong Kong, Hong Kong, SAR, China, from November 2015 to November 2018. Since September 2012, he has been with the School of Mathematical Sciences, Nankai University, where he is currently a Professor. His research interests include information theory, the areas of network coding, and network function computation. He was a recipient of the 2018 Chinese Information Theory Young Rising Star Award by China Information Theory Society.

Jian Lu (Member, IEEE) received the B.S. and M.S. degrees in applied mathematics and statistics and the Ph.D. degree in information engineering from Southeast University, Nanjing, China, in 2003, 2006, and 2012, respectively. Since May 2012, he has been with the School of Information Science and Engineering, Southeast University, where he is currently an Assistant Professor. His research interests include information theory and statistical signal processing.

Jun Chen (Senior Member, IEEE) received the B.E. degree 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, USA, in 2004 and 2006, respectively.

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

Dr. Chen was a recipient of the Josef Raviv Memorial Post-Doctoral Fellowship in 2006, the Early Researcher Award from the Province of Ontario in 2010, the IBM Faculty Award in 2010, the ICC Best Paper Award in 2020, and the JSPS Invitational Fellowship in 2021. He held the title of the Barber–Gennum Chair in Information Technology from 2008 to 2013 and the Joseph Ip Distinguished Engineering Fellow from 2016 to 2018. He served as an Editor for the IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING from 2020 to 2021. He is currently an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY.