

# The Dispersion of the Gauss–Markov Source

Peida Tian<sup>✉</sup>, *Student Member, IEEE*, and Victoria Kostina<sup>✉</sup>, *Member, IEEE*

**Abstract**—The Gauss–Markov source produces  $U_i = aU_{i-1} + Z_i$  for  $i \geq 1$ , where  $U_0 = 0$ ,  $|a| < 1$  and  $Z_i \sim \mathcal{N}(0, \sigma^2)$  are i.i.d. Gaussian random variables. We consider lossy compression of a block of  $n$  samples of the Gauss–Markov source under squared error distortion. We obtain the Gaussian approximation for the Gauss–Markov source with excess-distortion criterion for any distortion  $d > 0$ , and we show that the dispersion has a reverse waterfilling representation. This is the *first* finite blocklength result for lossy compression of *sources with memory*. We prove that the finite blocklength rate-distortion function  $R(n, d, \epsilon)$  approaches the rate-distortion function  $\mathbb{R}(d)$  as  $R(n, d, \epsilon) = \mathbb{R}(d) + \sqrt{\frac{V(d)}{n}} Q^{-1}(\epsilon) + o\left(\frac{1}{\sqrt{n}}\right)$ , where  $V(d)$  is the dispersion,  $\epsilon \in (0, 1)$  is the excess-distortion probability, and  $Q^{-1}$  is the inverse  $Q$ -function. We give a reverse waterfilling integral representation for the dispersion  $V(d)$ , which parallels that of the rate-distortion functions for Gaussian processes. Remarkably, for all  $0 < d \leq \frac{\sigma^2}{(1+|a|)^2}$ ,  $R(n, d, \epsilon)$  of the Gauss–Markov source coincides with that of  $Z_i$ , the i.i.d. Gaussian noise driving the process, up to the second-order term. Among novel technical tools developed in this paper is a sharp approximation of the eigenvalues of the covariance matrix of  $n$  samples of the Gauss–Markov source, and a construction of a typical set using the maximum likelihood estimate of the parameter  $a$  based on  $n$  observations.

**Index Terms**—Lossy source coding, Gauss–Markov source, dispersion, finite blocklength regime, rate-distortion theory, sources with memory, achievability, converse, autoregressive processes, covering in probability spaces, parameter estimation.

## I. INTRODUCTION

**I**N RATE-DISTORTION theory [2] [3], a source, modeled as a discrete stochastic process  $\{U_i\}_{i=1}^\infty$ , produces a random vector  $\mathbf{U} \triangleq (U_1, \dots, U_n)^\top$  and the goal is to represent  $\mathbf{U}$  by the minimum number of reproduction vectors  $\mathbf{V}$  such that the distortion is no greater than a given threshold  $d$ . For any such set of reproduction vectors, the associated rate is defined as the ratio between the logarithm of the number of vectors and  $n$ . The rate quantifies the minimum number of bits per symbol needed to describe the source with distortion  $d$ .

Numerous studies have been pursued since the seminal paper [3], where Shannon first proved the rate-distortion theorem for the discrete stationary memoryless sources (DMS)

and then sketched the ideas on generalizing to continuous alphabets and the stationary ergodic sources. Shannon's rate-distortion theorem shows that the minimum rate needed to describe a DMS within distortion  $d$  is given by the rate-distortion function (RDF)  $\mathbb{R}(d)$ , which is computed as a solution to a (single-letter) minimal mutual information convex optimization problem. Goblick [4] proved a coding theorem for a general subclass of strongly mixing stationary sources showing that the RDF is equal to the limit of  $n$ -letter minimal mutual information. That limit has exponential computational complexity in general. Computable expressions for the RDF of sources with memory are known only in the following special cases. Gray [5] showed a closed-form expression for the RDF for a binary symmetric Markov source with bit error rate distortion in a low distortion regime. For higher distortions, Jalali and Weissman [6] recently showed upper and lower bounds allowing one to compute the rate-distortion function in this case with desired accuracy. Gray [7] showed a lower bound to the rate-distortion function of finite-state finite-alphabet Markov sources with a balanced distortion measure, and the lower bound becomes tight when  $d \in (0, d_c]$  for critical distortion  $d_c$ . For the mean squared error distortion measure (MSE), Davisson [8], and also Kolmogorov [9], derived the rate-distortion function for stationary Gaussian processes by applying a unitary transformation to the process to decorrelate it and applying reverse waterfilling to the decorrelated Gaussians [10]. Berger [11] and Gray [5], in separate contributions in the late 60's and early 70's, derived the MSE rate-distortion function for Gaussian autoregressive sources. See [12] for a detailed survey on the development of coding theorems for more general sources.

All of the above mentioned work [2]–[8], [10], [11] apply to the operational regime where the coding length  $n$  grows without bound. Asymptotic coding theorems are important since they set a clear boundary between the achievable and the impossible. However, practical compression schemes are of finite blocklength. A natural, but challenging, question to ask is: for a given coding blocklength  $n$ , what is the minimum rate to compress the source with distortion at most  $d$ ? Answering this question *exactly* is hard. An easier question is that of second-order analysis, which studies the dominating term in the gap between RDF and the finite blocklength minimum rate.

In rate-distortion theorems, the landscape of second-order analyses consists of two criteria: average distortion and excess distortion. The average distortion constraint posits that the average distortion should be at most  $d$ , while excess distortion constraint requires that the probability of distortion exceeding  $d$  be at most  $\epsilon$ . For average distortion criterion, Zhang, Yang and Wei [13] proved that for i.i.d. finite alphabet sources,

Manuscript received May 14, 2018; revised November 17, 2018 and April 11, 2019; accepted May 11, 2019. Date of publication May 29, 2019; date of current version September 13, 2019. This work was supported by the National Science Foundation (NSF) under Grant CCF-1566567 and Grant CCF-1751356. This paper was presented at the 2018 IEEE International Symposium on Information Theory [1].

The authors are with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125 USA (e-mail: ptian@caltech.edu; vkostina@caltech.edu).

Communicated by S. Watanabe, Associate Editor for Shannon Theory.

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIT.2019.2919718

0018-9448 © 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.

See [http://www.ieee.org/publications\\_standards/publications/rights/index.html](http://www.ieee.org/publications_standards/publications/rights/index.html) for more information.

the leading term in the gap  $R(n, d) - \mathbb{R}(d)$  is  $\frac{\log n}{2n}$ <sup>1</sup>, where  $R(n, d)$  denotes the minimum rate compatible with average distortion  $d$  at coding length  $n$ . Later, Yang and Zhang [14] extended the achievability result of [13] to abstract sources.

For lossy compression of i.i.d. sources under excess distortion criterion, the minimum achievable finite blocklength rate admits the following expansion [15], [16], known as the Gaussian approximation:

$$R(n, d, \epsilon) = \mathbb{R}(d) + \sqrt{\frac{V(d)}{n}} Q^{-1}(\epsilon) + O\left(\frac{\log n}{n}\right), \quad (1)$$

where  $V(d)$  is referred to as the source dispersion and  $Q^{-1}(\cdot)$  denotes the inverse  $Q$ -function. Extensions of the result in (1) to joint source-channel coding [17], [18] and multiterminal source coding [19], [20] have also been studied.

The dispersion of lossy compression of *sources with memory* is unknown. In the context of variable-length lossy compression with guaranteed distortion, Kontoyiannis [21, Th. 6, Th. 8] established a connection between the number of bits needed to represent  $n$  given samples produced by an arbitrary source, and the logarithm of the reciprocal of distortion  $d$ -ball probability. Unfortunately computation of that probability has exponential in  $n$  complexity. In contrast, the dispersions of lossless compression of sources with memory and of channel coding over channels with memory are known in some cases. The second-order expansion of the minimum encoded length in the lossless compression of Markov sources is computed in [22] and [23]. Polyanskiy *et al.* found the channel dispersion of the Gilbert-Elliott channel in [24, Th. 4].

In this paper, we derive an expansion of type (1) on  $R(n, d, \epsilon)$  for the Gauss-Markov source, one of the simplest models for sources with memory. We show that the dispersion  $V(d)$  for the Gauss-Markov source is equal to the limiting variance of the  $d$ -tilted information, and has a reverse waterfilling representation. We show that the dispersion  $V(d)$  for low distortions is the same as that of the i.i.d. Gaussian noise driving the process, and becomes smaller for distortions above a critical value  $d_c$ , which extend the corresponding result of Gray [5, Eq. (24)] to the nonasymptotic regime. Section II presents the problem formulation. The main results and proof techniques are presented in Section III. Our proofs of converse and achievability are presented in Sections IV and V, respectively. The converse proof generalizes to the Gaussian autoregressive processes [5], but the achievability proof does not. In proving the converse and achievability, we develop several new tools including a nonasymptotic refinement of Gray's result [5, Eq. (19)] on the eigenvalue distribution of the covariance matrix of the Gauss-Markov source. This refinement relies on a sharp bound on the differences of eigenvalues of two sequences of tridiagonal matrices, proved using the Cauchy interlacing theorem and the Gershgorin circle theorem from matrix theory. In proving achievability, we derive a maximum likelihood estimator of the parameter  $a$  of the Gauss-Markov source and bound the estimation error using the Hanson-Wright inequality [25, Th. 1.1]. Our key tool in the achievability proof is the

construction of a typical set based on the maximum likelihood estimator. Finally, we conclude in Section VI with brief discussions on some open problems. Fig. 4 in Appendix A-A presents a roadmap containing the relations of all theorems, corollaries and lemmas in this paper.

**Notations:** Throughout, lowercase (uppercase) boldface letters denote vectors (random vectors) of length  $n$ . We omit the dimension when there is no ambiguity, i.e.  $\mathbf{u} \equiv \mathbf{u}^n \equiv (u_1, \dots, u_n)^\top$  and  $\mathbf{U} \equiv \mathbf{U}^n \equiv (U_1, \dots, U_n)^\top$ . We write  $U$  for  $U^\infty$ . For a random variable  $X$ , we use  $\mathbb{E}[X]$  and  $\text{Var}[X]$  to denote its mean and variance, respectively. We write matrices using sans serif font, e.g. matrix  $\mathbf{A}$ , and we write  $\|\mathbf{A}\|_F$  and  $\|\mathbf{A}\|$  to denote the Frobenius and operator norms of  $\mathbf{A}$ , respectively. The trace of  $\mathbf{A}$  is denoted by  $\text{tr}(\mathbf{A})$ . For a vector  $\mathbf{v}$ , we denote by  $\|\mathbf{v}\|_p$  the  $\ell_p$ -norm of  $\mathbf{v}$  ( $p = 1$  or  $p = 2$  in this paper). We also denote the sup norm of a function  $F$  by  $\|F\|_\infty \triangleq \sup_{x \in \mathcal{D}} |F(x)|$ , where  $\mathcal{D}$  denotes the domain of  $F$ . We use the standard  $O(\cdot)$ ,  $o(\cdot)$  and  $\Theta(\cdot)$  notations to characterize functions according to their asymptotic growth rates. Namely, let  $f(n)$  and  $g(n)$  be two functions on  $n$ , then  $f(n) = O(g(n))$  if and only if there exists positive real number  $M$  and  $n_0 \in \mathbb{N}$  such that  $|f(n)| \leq M|g(n)|$  for any  $n \geq n_0$ ;  $f(n) = o(g(n))$  if and only if  $\lim_{n \rightarrow \infty} f(n)/g(n) = 0$ ;  $f(n) = \Theta(g(n))$  if and only if there exist positive constants  $c_1, c_2$  and  $n_0 \in \mathbb{N}$  such that  $c_1 g(n) \leq f(n) \leq c_2 g(n)$  for any  $n \geq n_0$ . For any positive integer  $m$ , we denote by  $[m]$  the set of integers  $\{1, 2, \dots, m\}$ . We denote by  $\mathbb{1}\{\cdot\}$  the indicator function. We use  $n!!$  to denote the double factorial of  $n$ . The imaginary unit is denoted by  $j$ . All exponents and logarithms are base  $e$ .

## II. PROBLEM FORMULATION

### A. Operational Definitions

In single-shot lossy compression, we consider source and reproduction alphabets  $\mathcal{X}$  and  $\mathcal{Y}$ , and a given source distribution  $P_X$  over  $\mathcal{X}$ . The distortion measure is a mapping  $d(\cdot, \cdot): \mathcal{X} \times \mathcal{Y} \mapsto [0, +\infty)$ . An encoder  $\mathbf{f}$  is a mapping  $\mathbf{f}: \mathcal{X} \mapsto [M]$ , and a decoder is  $\mathbf{g}: [M] \mapsto \mathcal{Y}$ . The image set of a decoder  $\mathbf{g}$  is referred to as a codebook consisting of  $M$  codewords  $\{\mathbf{g}(i)\}_{i=1}^M$ . Given distortion threshold  $d > 0$  and excess-distortion probability  $\epsilon \in (0, 1)$ , an  $(M, d, \epsilon)$  code consists of an encoder-decoder pair  $(\mathbf{f}, \mathbf{g})$  such that  $\mathbb{P}[d(X, \mathbf{g}(\mathbf{f}(X))) > d] \leq \epsilon$ . The nonasymptotic fundamental limit of lossy compression is the minimum achievable code size for a given distortion threshold  $d$  and an excess-distortion probability  $\epsilon \in (0, 1)$ :

$$M^*(d, \epsilon) \triangleq \min \{M : \exists \text{ an } (M, d, \epsilon) \text{ code}\}. \quad (2)$$

In this paper,  $\mathcal{X} = \mathcal{Y} = \mathbb{R}^n$ , and the distortion measure is the mean squared error (MSE) distortion:  $\forall \mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ ,

$$d(\mathbf{u}, \mathbf{v}) \triangleq \frac{1}{n} \|\mathbf{u} - \mathbf{v}\|_2^2. \quad (3)$$

We refer to the set  $\mathcal{B}(\mathbf{x}, d)$ , defined below, as a distortion  $d$ -ball centered at  $\mathbf{x}$ :

$$\mathcal{B}(\mathbf{x}, d) \triangleq \{\mathbf{x}' \in \mathbb{R}^n : d(\mathbf{x}, \mathbf{x}') \leq d\}. \quad (4)$$

<sup>1</sup>This statement is translated from [13], where the equivalent result was stated in terms of distortion-rate function.



We consider the Gauss-Markov source  $\{U_i\}_{i=1}^{\infty}$ , which satisfies the following difference equation:

$$U_i = aU_{i-1} + Z_i, \quad \forall i \geq 1, \quad (5)$$

and  $U_0 = 0$ . Here,  $a \in [0, 1)$  is the gain,<sup>2</sup> and  $Z_i$ 's are independently and identically distributed (i.i.d.)  $\mathcal{N}(0, \sigma^2)$  that form the *innovation* process. We adopt (5) as the simplest model capturing information sources with memory: gain  $a$  determines how much memory (as well as the growth rate), and the innovation  $Z_i$  represents new randomness being generated at each time step. In statistics, the Gauss-Markov source (5) is also known as the first-order Gaussian autoregressive (AR) process.

For a fixed blocklength  $n \in \mathbb{N}$ , a distortion threshold  $d > 0$  and an excess-distortion probability  $\epsilon \in (0, 1)$ , an  $(n, M, d, \epsilon)$  code consists of an encoder  $\mathbf{f}_n: \mathbb{R}^n \mapsto [M]$  and a decoder  $\mathbf{g}_n: [M] \mapsto \mathbb{R}^n$  such that  $\mathbb{P}[\mathbf{d}(\mathbf{U}, \mathbf{g}_n(\mathbf{f}_n(\mathbf{U}))) \geq d] \leq \epsilon$ , where  $\mathbf{U} = (U_1, \dots, U_n)^T$  denotes the source vector. The rate associated with an  $(n, M, d, \epsilon)$  code is  $R \triangleq \frac{\log M}{n}$ . The nonasymptotic operational fundamental limit, that is, the minimum achievable code size at blocklength  $n$ , distortion  $d > 0$  and excess-distortion probability  $\epsilon \in (0, 1)$ , is

$$M^*(n, d, \epsilon) \triangleq \min \{M : \exists \text{ an } (n, M, d, \epsilon) \text{ code}\}, \quad (6)$$

and the corresponding minimum source coding rate is

$$R(n, d, \epsilon) \triangleq \frac{\log M^*(n, d, \epsilon)}{n}. \quad (7)$$

The objective of this paper is to characterize  $R(n, d, \epsilon)$  for the Gauss-Markov source.

### B. Informational Definitions

The problem of characterizing the operational fundamental limit  $M^*(d, \epsilon)$  in (2) is closely related to the rate-distortion function (RDF)  $\mathbb{R}_X(d)$  of the source  $X$ , which is defined as the solution to the following convex optimization problem [3]:

$$\mathbb{R}_X(d) \triangleq \inf_{P_{Y|X}: \mathbb{E}[\mathbf{d}(X, Y)] \leq d} I(X; Y), \quad (8)$$

where the infimum is over all conditional distributions  $P_{Y|X}: \mathcal{X} \mapsto \mathcal{Y}$  such that the expected distortion is less than or equal  $d$ , and  $I(X; Y)$  denotes the mutual information between  $X$  and  $Y$ . In this paper, we assume that

- 1)  $\mathbb{R}_X(d)$  is differentiable with respect to  $d$ ;
- 2) there exists a minimizer in (8).

The pair  $(X, Y^*)$  is referred to as the *RDF-achieving pair* if  $P_{Y^*|X}$  is the minimizer in (8). For any  $x \in \mathcal{X}$ , the  $\mathbf{d}$ -tilted information  $J_X(x, d)$  in  $x$ , introduced in [15, Definition 6], is

$$J_X(x, d) \triangleq -\lambda^* d - \log \mathbb{E} \exp(-\lambda^* \mathbf{d}(x, Y^*)), \quad (9)$$

where  $\lambda^*$  is the negative slope of the curve  $\mathbb{R}_X(d)$  at distortion  $d$ :

$$\lambda^* \triangleq -\mathbb{R}'_X(d). \quad (10)$$

<sup>2</sup>Note that if  $a \in (-1, 0]$  in (5), then  $\{(-1)^i U_i\}_{i=0}^{\infty}$  is a Gauss-Markov source with nonnegative gain  $-a$  and the same innovation variance. Thus restricting to  $0 \leq a < 1$  is without loss of generality.

The  $\mathbf{d}$ -tilted information  $J_X(X, d)$  has the property that

$$\mathbb{R}_X(d) = \mathbb{E}[J_X(X, d)]. \quad (11)$$

When  $\mathcal{X}$  is a finite set, (11) follows immediately from the Karush-Kuhn-Tucker (KKT) conditions for the optimization problem (8), see [26, Th. 9.4.1] and [11, Eq. (2.5.16)]. Csiszár showed the validity of (11) when  $\mathcal{X}$  is an abstract probability space [27, Corollary, Lemma 1.4, Eqs. (1.15), (1.25), (1.27)–(1.32)], see Appendix A-C for a concise justification. For more properties of the  $\mathbf{d}$ -tilted information, see [15, Eq. (17)–(19)].

Next, we introduce the conditional relative entropy minimization (CREM) problem, which plays a key role in our development. Let  $P_X$  and  $P_Y$  be probability distributions defined on alphabets  $\mathcal{X}$  and  $\mathcal{Y}$ , respectively. For any  $d > 0$ , the CREM problem is defined as

$$\mathbb{R}(X, Y, d) \triangleq \inf_{P_{F|X}: \mathbb{E}[\mathbf{d}(X, F)] \leq d} D(P_{F|X} || P_Y | P_X), \quad (12)$$

where  $F$  is a random variable taking values in  $\mathcal{Y}$ , and  $D(P_{F|X} || P_Y | P_X)$  is the conditional relative entropy:

$$D(P_{F|X} || P_Y | P_X) \triangleq \int D(P_{F|X=x} || P_Y) dP_X(x). \quad (13)$$

A well-known fact in the study of lossy compression is that the CREM problem (12) is related to the RDF (8) as

$$\mathbb{R}_X(d) = \inf_{P_Y} \mathbb{R}(X, Y, d), \quad (14)$$

where the infimization is over all probability distributions  $P_Y$  of the random variables  $Y$  over  $\mathcal{Y}$  that are independent of  $X$ ; and the equality in (14) is achieved when  $P_Y$  is the  $Y^*$ -marginal of the RDF-achieving pair  $(X, Y^*)$ , see [10, Eq. (10.140)] and [28, Th. 4] for the finite alphabets  $\mathcal{X}$ ; [14, Eq. (3.3)] and [21, Eq. (13)] for abstract alphabets  $\mathcal{X}$ . The property (14) is a foundation of the Blahut-Arimoto algorithm, which computes iterative approximations to  $\mathbb{R}_X(d)$  by alternating between inner and outer infimizations in (14). The CREM problem is also important in nonasymptotic analyses of lossy compressors, see [14, Eq. (3.3)], [21, Eq. (13)] and [15, Eq. (27)]. Operationally, it relates to the mismatched-codebooks problem, that is, lossy compression of source  $P_X$  using random codewords drawn from  $P_Y$  [29, Th. 12]. Similar to (9),  $\forall x \in \mathcal{X}$ ,  $\delta > 0$ ,  $d > 0$ , the *generalized tilted information*  $\Lambda_Y(x, \delta, d)$ , defined in [15, Eq. (28)], is

$$\Lambda_Y(x, \delta, d) \triangleq -\delta d - \log \mathbb{E} \exp(-\delta \mathbf{d}(x, Y)). \quad (15)$$

The optimizer  $P_{F^*|X}$  of (12) satisfies the following condition:  $\forall x \in \mathcal{X}, y \in \mathcal{Y}$ ,

$$\log \frac{dP_{F^*|X}(y|x)}{dP_Y(y)} = \Lambda_Y(x, \delta^*, d) - \delta^* \mathbf{d}(x, y) + \delta^* d, \quad (16)$$

where

$$\delta^* \triangleq -\mathbb{R}'(X, Y, d). \quad (17)$$

When  $\mathcal{X}$  and  $\mathcal{Y}$  are discrete, (16) can be verified by the KKT conditions for the optimization problem (12). For abstract

alphabets  $\mathcal{X}$ , see [29, Th. 2] and [14, Property 1] for an exposition. By comparing (9) and (15) and using the relation (14), we see that

$$J_X(x, d) = \Lambda_{Y^*}(x, \lambda^*, d), \quad (18)$$

where  $\lambda^*$  is in (10).

For the Gauss-Markov source defined in (5), its  $n$ -th order rate-distortion function  $\mathbb{R}_U(n, d)$  is defined by replacing  $X$  by  $\mathbf{U}$  in (8) and then normalizing by  $n$ :

$$\mathbb{R}_U(n, d) \triangleq \frac{1}{n} \inf_{P_{\mathbf{V}|\mathbf{U}}: \mathbb{E}[d(\mathbf{U}, \mathbf{V})] \leq d} I(\mathbf{U}; \mathbf{V}). \quad (19)$$

The rate-distortion function  $\mathbb{R}_U(d)$  for the Gauss-Markov source (5) is

$$\mathbb{R}_U(d) \triangleq \limsup_{n \rightarrow \infty} \mathbb{R}_U(n, d). \quad (20)$$

It immediately follows from (11) that

$$\mathbb{R}_U(d) = \limsup_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[J_U(\mathbf{U}, d)], \quad (21)$$

where  $J_U(\mathbf{U}, d)$  is the  $d$ -tilted information random variable defined in (9), that is,

$$J_U(\mathbf{u}, d) = -\lambda^* n d - \log \mathbb{E}[\exp(-\lambda^* n d(\mathbf{u}, \mathbf{V}^*))], \quad (22)$$

where  $(\mathbf{U}, \mathbf{V}^*)$  forms a RDF-achieving pair in (19) and

$$\lambda^* = -\mathbb{R}'_U(n, d). \quad (23)$$

The variance of the  $d$ -tilted information is important in capturing the second-order performance of the best source code. Define

$$\mathbb{V}_U(n, d) \triangleq \text{Var}[J_U(\mathbf{U}, d)], \quad (24)$$

$$\mathbb{V}_U(d) \triangleq \limsup_{n \rightarrow \infty} \frac{1}{n} \mathbb{V}_U(n, d). \quad (25)$$

The quantity  $\mathbb{V}_U(d)$  is referred to as the *informational dispersion*, in contrast to the operational dispersion  $V_U(d)$  defined in the next subsection. Reverse waterfilling solutions for rate-distortion functions of the Gauss-Markov source were well-known, see [5, Eq. (15)] for  $\mathbb{R}_U(n, d)$ , [5, Eq. (22)] for  $\mathbb{R}_U(d)$  and our discussions in Section II-D below. In this paper, we derive similar parametric expressions for both  $\mathbb{V}_U(n, d)$  and  $\mathbb{V}_U(d)$ .

### C. Operational Fundamental Limits

In terms of coding theorems, the equality between  $R(d)$  (the minimum achievable source coding rate under average distortion criterion when the blocklength  $n$  goes to infinity) and  $\mathbb{R}_U(d)$  (the informational rate-distortion function defined in (20)) has been established, e.g. [11, Th. 6.3.4] and [5, Th. 2]. For the Gauss-Markov source, by the Markov inequality, the achievability result under average distortion criterion, e.g. [5, Th. 2], can be converted into an achievability result under the excess distortion criterion. A matching converse follows from Kieffer's strong converse [30, Th. 1] for the

stationary ergodic sources. Therefore, for any  $d > 0$  and  $\epsilon \in (0, 1)$ , we have

$$\lim_{n \rightarrow \infty} R(n, d, \epsilon) = \mathbb{R}_U(d), \quad (26)$$

where  $R(n, d, \epsilon)$  is defined in (7) and  $\mathbb{R}_U(d)$  in (20).

The main result of this paper is the following Gaussian approximation for the minimum achievable rate  $R(n, d, \epsilon)$  in lossy compression of the Gauss-Markov source (5):

$$R(n, d, \epsilon) = \mathbb{R}_U(d) + \sqrt{\frac{\mathbb{V}_U(d)}{n}} Q^{-1}(\epsilon) + o\left(\frac{1}{\sqrt{n}}\right), \quad (27)$$

where  $Q^{-1}(\cdot)$  denotes the inverse  $Q$ -function, and the term  $o(\cdot)$  will be refined in Theorems 7 and 11 in Sections IV and V below. Our main result (27) is a nonasymptotic refinement of (26), implying that the convergence rate in the limit (26) is of order  $\frac{1}{\sqrt{n}}$  with the optimal constant factor given in (27). Formally, the rate-dispersion function  $V_U(d)$ , introduced in [15, Def. 7] and simply referred to as (*operational*) *dispersion*, is

$$V_U(d) \triangleq \lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow \infty} n \left( \frac{R(n, d, \epsilon) - \mathbb{R}_U(d)}{Q^{-1}(\epsilon)} \right)^2. \quad (28)$$

Equivalently, our main result (27) establishes the equality between the operational and informational dispersions for the Gauss-Markov source:

$$V_U(d) = \mathbb{V}_U(d). \quad (29)$$

### D. Related Work

The  $n$ -th order RDF  $\mathbb{R}_U(n, d)$  defined in (19) for the Gauss-Markov source is given by the reverse waterfilling [5, Eq. (17)] and [11, Eq. (6.3.34)–(6.3.36)]:

$$\mathbb{R}_U(n, d) = \frac{1}{n} \sum_{i=1}^n \max\left(0, \frac{1}{2} \log \frac{\sigma_i^2}{\theta_n}\right), \quad (30)$$

$$d = \frac{1}{n} \sum_{i=1}^n \min(\theta_n, \sigma_i^2), \quad (31)$$

where  $\sigma_i^2$ 's are the eigenvalues of the covariance matrix (see the discussions in Section III-A below):

$$\Sigma_U \triangleq \mathbb{E}[\mathbf{U}\mathbf{U}^\top], \quad (32)$$

and  $\theta_n > 0$  is the water level matched to  $d$  at blocklength  $n$ . The rate-distortion function  $\mathbb{R}_U(d)$  for the Gauss-Markov source (5) is obtained by passing to the limit of infinite  $n$  in (30) and (31) via invoking the limiting theorems on the eigenvalues of the covariance matrix  $\Sigma_U$  [5, Eq. (22)] and [11, Th. 6.3.2], given by

$$\mathbb{R}_U(d) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \max\left[0, \frac{1}{2} \log \frac{S(w)}{\theta}\right] dw, \quad (33)$$

$$d = \frac{1}{2\pi} \int_{-\pi}^{\pi} \min[\theta, S(w)] dw, \quad (34)$$

where the power spectrum of the Gauss-Markov source (5) is given by

$$S(w) = \frac{\sigma^2}{g(w)}, \quad (35)$$



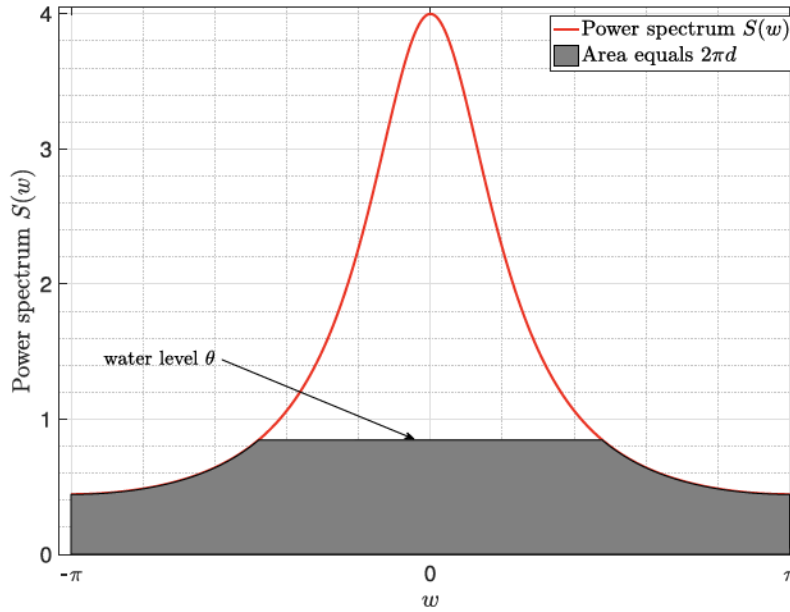


Fig. 1. Reverse waterfilling (34) for  $a = 0.5$ : the water level  $\theta$  is chosen such that the shaded area equals  $2\pi d$ .

and the function  $g$  is defined as

$$g(w) \triangleq 1 + a^2 - 2a \cos(w), \quad \forall w \in [-\pi, \pi]. \quad (36)$$

We refer to (30)-(31) as the  $n$ -th order reverse waterfilling, and to (33)-(34) as the limiting reverse waterfilling. Fig. 1 depicts the limiting reverse waterfilling (34).

Results similar to (30)-(34) hold for the stationary Gaussian processes [9, Eq. (17), (18)], as well as for the higher-order Gaussian AR processes (not necessarily stationary) [5, Eq. (22)]. We discuss the subtle differences between the rate-distortion functions of the (asymptotically) stationary and nonstationary Gaussian AR processes in Section VI below. This paper considers the (asymptotically) stationary Gauss-Markov sources, i.e., (5) with  $|a| < 1$ .

The converse results in this paper extend partly to the higher-order Gaussian AR processes, studied by Gray [5] and Berger [11, Sec. 6.3.2]. The Gaussian AR process is [5, Eq. (1)]

$$U_i = \sum_{\ell=1}^i a_\ell U_{i-\ell} + Z_i, \quad i \geq 1, \quad (37)$$

and  $U_i = 0$  for  $i \leq 0$ , where  $Z_i$ 's are i.i.d.  $\mathcal{N}(0, \sigma^2)$ , and the real constants  $a_\ell$ 's satisfy [5, Eq. (10)]

$$\sum_{\ell=0}^{\infty} |a_\ell| < \infty. \quad (38)$$

The Gauss-Markov source in (5) is a special case of (37) with  $a_1 = a$  and  $a_\ell = 0$  for  $\ell \geq 2$ . The following relation between the rate-distortion functions of the Gaussian AR process  $\{U_i\}_{i=1}^{+\infty}$  in (37) and the i.i.d. Gaussian process  $\{Z_i\}_{i=1}^{+\infty}$  is due to Gray [5, Eq. (24)]:

$$\begin{cases} \mathbb{R}_U(d) = \mathbb{R}_Z(d), & 0 < d \leq d_c, \\ \mathbb{R}_U(d) > \mathbb{R}_Z(d), & d_c < d \leq d_{\max}, \end{cases} \quad (39)$$

where  $d_c$  is referred to as the *critical distortion*, defined as  $d_c \triangleq \theta_{\min}$ , where

$$\theta_{\min} \triangleq \min_{w \in [-\pi, \pi]} S(w). \quad (40)$$

Accordingly, denote the maximum value of  $S(w)$  over the interval  $[-\pi, \pi]$  as

$$\theta_{\max} \triangleq \max_{w \in [-\pi, \pi]} S(w). \quad (41)$$

In (39),  $d_{\max}$  is the *maximum distortion* achievable in (34) (that is, when  $\theta \geq \theta_{\max}$ ):

$$d_{\max} \triangleq \frac{1}{2\pi} \int_{-\pi}^{\pi} S(w) dw, \quad (42)$$

and  $\mathbb{R}_Z(d)$  is the RDF for i.i.d. Gaussian sources derived by Shannon [3, Equation below Fig. 9]:

$$\mathbb{R}_Z(d) = \max \left( 0, \frac{1}{2} \log \frac{\sigma^2}{d} \right). \quad (43)$$

The power spectrum  $S(w)$  of the Gaussian AR process is [5, Eq. (21)]

$$S(w) = \sigma^2 \left| \sum_{\ell=0}^{+\infty} a_\ell e^{-j\ell w} \right|^{-2}, \quad w \in [-\pi, \pi]. \quad (44)$$

Equality in (39) is a deep result stating that in a range of low distortions, the asymptotic rate-distortion tradeoff of a Gaussian AR process and that of its driving innovation process are the same. See Fig. 2 for an illustration of (39) in the special case of a Gauss-Markov source with  $a = 0.5$ .

The critical distortion  $d_c$  and the maximum distortion  $d_{\max}$  can be understood pictorially as follows. In Fig. 1 and equivalently in (34), as the water level  $\theta$  rises from 0 to  $\theta_{\min}$ , the minimum on the right side of (34) equals  $\theta$ , meaning that  $d = \theta$  for  $0 < \theta \leq \theta_{\min}$  (equivalently,  $0 \leq d \leq d_c$ ). As the water level  $\theta$  rises further, lower parts of the spectrum  $S(w)$

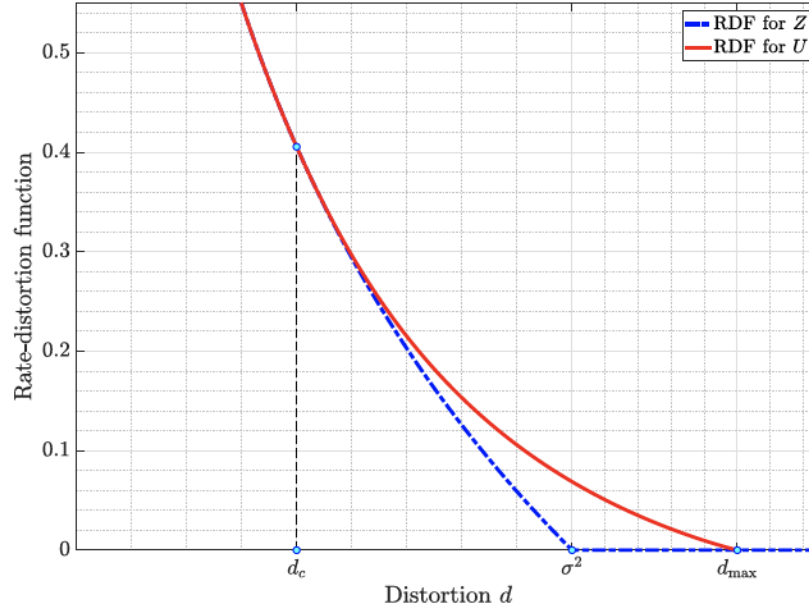


Fig. 2. The rate-distortion functions for the Gauss-Markov source  $\{U_i\}$  with  $a = 0.5$  and  $\sigma^2 = 1$ , and for the innovation process  $\{Z_i\}_{i=1}^\infty$ ,  $Z_i \sim \mathcal{N}(0, 1)$  driving that source.

start to play a role in (34). When the water level  $\theta$  rises above the peak in Fig. 1:  $\theta \geq \theta_{\max}$ , the distortion  $d$  in (34) remains as  $d_{\max}$ . In the case of the Gauss-Markov source, from (35), it is easy to see that  $d_c$  and  $d_{\max}$  are given by

$$d_c = \frac{\sigma^2}{(1 + |a|)^2}, \quad (45)$$

$$d_{\max} = \frac{\sigma^2}{1 - a^2}. \quad (46)$$

Note that  $d_{\max}$  in (46) equals the stationary variance of the source (Appendix B-B), i.e.,

$$d_{\max} = \lim_{n \rightarrow \infty} \text{Var}[U_n]. \quad (47)$$

For the nonstationary Gauss-Markov sources ( $|a| \geq 1$ ),  $d_{\max} = +\infty$ .

### III. MAIN RESULTS

#### A. Preliminary: Decorrelation

We first make a simple but important observation on the equivalence between lossy compression of the Gauss-Markov sources and parallel independent Gaussian sources. For any  $n \in \mathbb{N}$ , the random vector  $\mathbf{U} = (U_1, \dots, U_n)^\top$  generated by the model (5) follows the multivariate Gaussian distribution  $\mathcal{N}(\mathbf{0}, \Sigma_{\mathbf{U}})$ , where  $\Sigma_{\mathbf{U}} = \sigma^2(\mathbf{A}^\top \mathbf{A})^{-1}$  is its covariance matrix and  $\mathbf{A}$  is an  $n \times n$  lower triangular matrix with  $\det \mathbf{A} = 1$ :

$$\mathbf{A} \triangleq \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ -a & 1 & 0 & \dots & 0 \\ 0 & -a & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & -a & 1 \end{bmatrix}. \quad (48)$$

Since (5) can be rewritten as  $\mathbf{Z} = \mathbf{A}\mathbf{U}$ , and  $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ , the covariance matrix of  $\mathbf{U}$  is given by

$$\Sigma_{\mathbf{U}} = \mathbb{E}[\mathbf{A}^{-1} \mathbf{Z} \mathbf{Z}^\top (\mathbf{A}^{-1})^\top] = \sigma^2 (\mathbf{A}^\top \mathbf{A})^{-1}. \quad (49)$$

We refer to the random vector  $\mathbf{X}$  as the *decorrelation* of  $\mathbf{U}$ :

$$\mathbf{X} \triangleq \mathbf{S}^\top \mathbf{U}, \quad (50)$$

where  $\mathbf{S}$  is the unitary matrix in the eigendecomposition of the positive definite matrix  $(\mathbf{A}^\top \mathbf{A})^{-1}$ :

$$(\mathbf{A}^\top \mathbf{A})^{-1} = \mathbf{S} \Lambda \mathbf{S}^\top, \quad (51)$$

$$\Lambda = \text{diag} \left( \frac{1}{\mu_1}, \dots, \frac{1}{\mu_n} \right), \quad (52)$$

where  $0 < \mu_1 \leq \dots \leq \mu_n$  are the eigenvalues of  $\mathbf{A}^\top \mathbf{A}$ . From (49) and (51), it is clear that  $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \Lambda)$ , i.e.,  $X_1, \dots, X_n$  are independent zero-mean Gaussian random variables with variances  $\sigma_i^2$ 's being the eigenvalues of  $\Sigma_{\mathbf{U}}$ :

$$\sigma_i^2 \triangleq \frac{\sigma^2}{\mu_i}, \quad i \in [n]. \quad (53)$$

Since they are related via the unitary transformation  $\mathbf{S}$  which preserves the geometry of the underlying Euclidean space,  $\mathbf{U}$  and  $\mathbf{X}$  are equivalent in terms of their fundamental limits. Indeed, any  $(n, M, d, \epsilon)$  code for  $\mathbf{U}$  (recall the definition in Section II-B above) can be transformed, via  $\mathbf{S}$ , into an  $(n, M, d, \epsilon)$  code for  $\mathbf{X}$ , and vice versa; therefore, the finite blocklength minimum achievable rates  $R(n, d, \epsilon)$  for  $\mathbf{U}$  and  $\mathbf{X}$  are the same. Since  $I(\mathbf{S}\mathbf{X}; \mathbf{S}\mathbf{Y}) = I(\mathbf{X}; \mathbf{Y})$  and  $\mathbb{E} \|\mathbf{S}\mathbf{X} - \mathbf{S}\mathbf{Y}\|_2^2 = \mathbb{E} \|\mathbf{X} - \mathbf{Y}\|_2^2$ , their  $n$ -th order and limiting rate-distortion functions are the same:  $\forall n \in \mathbb{N}$ ,  $d \in (0, d_{\max})$ , we have  $\mathbb{R}_{\mathbf{X}}(n, d) = \mathbb{R}_{\mathbf{U}}(n, d)$ , and hence  $\mathbb{R}_{\mathbf{X}}(d) = \mathbb{R}_{\mathbf{U}}(d)$ . By the same transformation, it is easy to verify that this



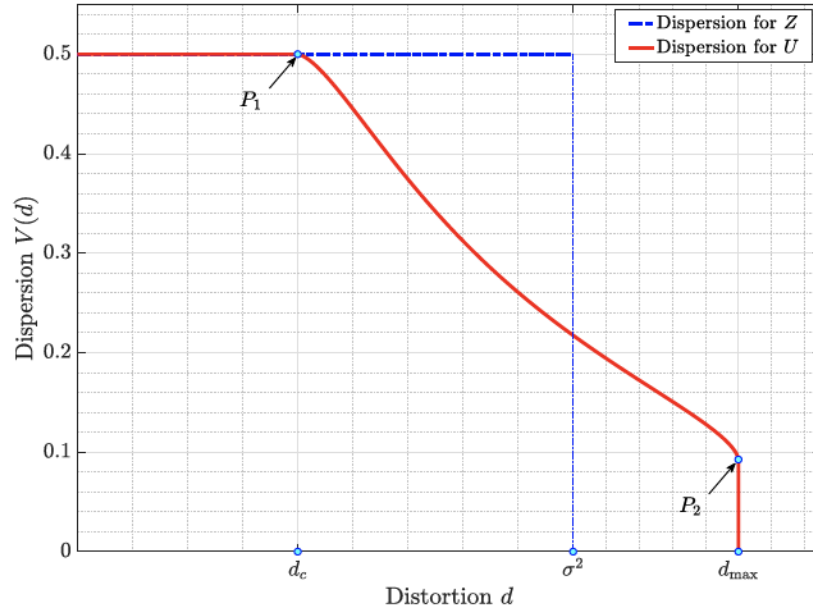


Fig. 3. Dispersion versus distortion: The Gauss-Markov source with  $a = 0$  and  $\sigma^2 = 1$  degenerates to the i.i.d. Gaussian source, i.e., the innovation process  $\{Z_i\}_{i=1}^\infty$ . The dispersion of the Gauss-Markov source with  $a = 0.5$  and  $\sigma^2 = 1$  is given by the solid line. Two corner points on the solid line are labeled as  $P_1$  (corresponding to  $d_c$ ) and  $P_2$  (corresponding to  $d_{\max}$ ).

equivalence also extends to the  $\mathbf{d}$ -tilted information:  $\forall \mathbf{u} \in \mathbb{R}^n$ , let

$$\mathbf{x} \triangleq \mathbf{S}^\top \mathbf{u}, \quad (54)$$

then

$$J_U(\mathbf{u}, d) = J_X(\mathbf{x}, d). \quad (55)$$

Due to the above equivalence, we will refer to both  $\mathbf{U}$  and its decorrelation  $\mathbf{X}$  in our analysis. Decorrelation is a well-known tool, which was used to find the rate-distortion functions for the general Gaussian AR processes [5].

### B. Gaussian Approximations for the Gauss-Markov Sources

We now formally state the main contributions of this paper.

**Theorem 1.** For the Gauss-Markov source in (5) with  $a \in [0, 1)$ , fix any excess-distortion probability  $\epsilon \in (0, 1)$  and distortion threshold  $d \in (0, d_{\max})$ , where  $d_{\max}$  is defined in (46). The minimum achievable source coding rate for the Gauss-Markov source in (5) satisfies

$$R(n, d, \epsilon) = \mathbb{R}_U(d) + \sqrt{\frac{V_U(d)}{n}} Q^{-1}(\epsilon) + o\left(\frac{1}{\sqrt{n}}\right), \quad (56)$$

where  $\mathbb{R}_U(d)$  is the rate-distortion function of the Gauss-Markov source, given in (33); and the operational dispersion  $V_U(d)$ , defined in (28), is given by

$$V_U(d) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \min \left[ 1, \left( \frac{S(w)}{\theta} \right)^2 \right] dw, \quad (57)$$

where  $\theta > 0$  is the water level matched to the distortion  $d$  via (34), and the power spectrum  $S(w)$  is in (35).

The proof of Theorem 1 is in given in Sections IV (converse) and V (achievability).

Pleasingly, the new reverse waterfilling solution for the dispersion in (57) parallels the classical reverse waterfilling representation of the rate-distortion function in (33). Furthermore, just like their rate-distortion functions (recall (39)), the dispersions of the Gauss-Markov source  $U$  in (5) and its innovation process  $Z$  are comparable:

**Corollary 1.** Let  $V_U(d)$  and  $V_Z(d)$  be the dispersions of the Gauss-Markov source (5) and the memoryless Gaussian source  $\{Z_i\}_{i=1}^\infty$ , respectively, then

$$\begin{cases} V_U(d) = V_Z(d), & 0 < d \leq d_c, \\ V_U(d) < V_Z(d), & d_c < d < \sigma^2. \end{cases} \quad (58)$$

*Proof.* From [16, Th. 2] and [15, Th. 40], we know that the dispersion of the memoryless Gaussian source is

$$V_Z(d) = \frac{1}{2}, \quad \forall d \in (0, \sigma^2), \quad (59)$$

which is also shown in Fig. 3. By the definition of  $d_c$  in (45) and the discussion around (45) and (46), we see that (57) satisfies

$$\min \left[ 1, \left( \frac{\sigma^2}{\theta g(w)} \right)^2 \right] \begin{cases} = 1, & \text{if } d \in (0, d_c], \\ < 1, & \text{if } d \in (d_c, d_{\max}), \end{cases} \quad (60)$$

from which Corollary 1 follows. ■

Corollary 1 parallels Gray's result (39) [5, Eq. (24)] for the rate-distortion functions of  $U$  and  $Z$ , and they together imply that for  $d \in (0, d_c]$ , the fundamental limits of lossy compression of the Gauss-Markov source and the i.i.d. Gaussian source  $\{Z_i\}_{i=1}^\infty$  are the same, up to the second-order term. For  $d \in (d_c, \sigma^2)$ , the Gauss-Markov source is harder to compress in the limit of  $n$  going to infinity since  $\mathbb{R}_U(d) > \mathbb{R}_Z(d)$ ,

but the Gauss-Markov source approaches its asymptotic fundamental limit faster since  $V_U(d) < V_Z(d)$ . See the discussions following Theorem 2 below for an intuitive explanation.

The dispersions for  $a = 0$  and  $a = 0.5$  are plotted in Fig. 3, where the dotted line (for  $a = 0, \sigma^2 = 1$ ) recovers the dispersion result (59) in [15] and [16] for the i.i.d. Gaussian sources  $\{Z_i\}_{i=1}^\infty$ , as expected. The solid line (for  $a = 0.5, \sigma^2 = 1$ ) coincides with the dotted line in the region  $d \in (0, d_c]$ , which means that the Gauss-Markov source has the same dispersion as its innovation process in the region of low  $d$ 's. For  $d \in (d_c, \sigma^2)$ , the dispersion of the Gauss-Markov source is smaller than that of its innovation process and decreases with  $d$ , as indicated by Corollary 1.

Using the residue theorem from complex analysis, we also derive the coordinates of the two corner points  $P_1$  and  $P_2$  on the solid line (Appendix B-A):

$$P_1 = (d_c, 1/2), \quad P_2 = \left(d_{\max}, \frac{(1+a^2)(1-a)}{2(1+a)^3}\right). \quad (61)$$

The vertical segment between  $(d_{\max}, 0)$  and  $P_2$  corresponds to the case when the water level  $\theta$  is above the spectrum peak  $\theta_{\max}$  in Fig. 1, and the dispersion  $V_U(d)$  in (57) becomes

$$V_U(d) = \frac{1}{4\pi\theta^2} \int_{-\pi}^{\pi} S(w)^2 dw, \quad (62)$$

which continues decreasing as  $\theta$  increases, even as the distortion  $d$  remains as  $d_{\max}$ , as seen from (34) and (57).

Theorem 2 below gives formulas for the  $d$ -tilted information defined in (22) and informational dispersion defined in (25).

**Theorem 2.** *For the Gauss-Markov source  $U$  in (5), for any  $d \in (0, d_{\max})$  and  $n \geq 1$ , the  $d$ -tilted information is given by*

$$J_U(\mathbf{u}, d) = J_{\mathbf{X}}(\mathbf{x}, d) = \sum_{i=1}^n \frac{\min(\theta_n, \sigma_i^2)}{2\theta_n} \left( \frac{x_i^2}{\sigma_i^2} - 1 \right) + \frac{1}{2} \log \frac{\max(\theta_n, \sigma_i^2)}{\theta_n}, \quad (63)$$

and the informational dispersion satisfies

$$\mathbb{V}_U(d) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \min \left[ 1, \left( \frac{S(w)}{\theta} \right)^2 \right] dw, \quad (64)$$

where  $\theta_n > 0$  is the water level matched to  $d$  via the  $n$ -th order reverse waterfilling (31);  $\theta > 0$  is the water level matched to the distortion  $d$  via the limiting reverse waterfilling (34); and the power spectrum  $S(w)$  is defined in (35).

*Proof.* Appendix C-D. ■

Theorem 2 computes the informational dispersion  $\mathbb{V}_U(d)$  defined in (25). The formula (56) in Theorem 1 is an equivalent reformulation of the definition of the operational dispersion  $V_U(d)$  in (28), while (57) together with Theorem 2 establish the equality  $\mathbb{V}_U(d) = V_U(d)$ . Theorem 1 and Theorem 2 establish that for the Gauss-Markov source, the operational RDF and dispersion are given by (21) and (25), respectively, providing a natural extension to the fact that in lossy compression of i.i.d. sources  $\{X_i\}$ , the mean  $\mathbb{E}[J_{X_1}(X_1, d)]$  and the variance  $\text{Var}[J_{X_1}(X_1, d)]$  of

the single-letter  $d$ -tilted information  $J_{X_1}(X_1, d)$  are equal to the RDF and the dispersion, respectively [15, Th. 12].

Theorem 2 also provides intuition on our result in Corollary 1 for  $d > d_c$ . Since  $(\mathbf{X}, \mathbf{Y}^*)$  forms a RDF-achieving pair in  $\mathbb{R}_{\mathbf{X}}(n, d)$  (recall (19)), it is well-known [10, Th. 10.3.3] that  $\mathbf{Y}^*$  has independent coordinates and  $\forall i \in [n]$ ,

$$Y_i^* \sim \mathcal{N}(0, \max(\sigma_i^2 - \theta_n, 0)), \quad (65)$$

where  $\theta_n > 0$  is the water level matched to the distortion  $d$  in the  $n$ -th order reverse waterfilling over  $\{\sigma_i^2\}_{i=1}^n$  in (31). Since  $d > d_c$ , there are some  $X_i$  in (63) which are "inactive", that is,  $\sigma_i^2 < \theta_n$ , which makes the variance of (63) smaller. Geometrically, since  $\mathbf{X}$  concentrates inside an ellipsoid, we are covering such an ellipsoid by balls of radius  $\sqrt{nd}$ . The centers of these distortion  $d$ -balls lie on another lower dimensional ellipsoid. That lower dimensional ellipsoid is the one on which the random vector  $\mathbf{Y}^*$  concentrates. For  $d > d_c$ , although centered at a lower dimensional ellipsoid (since  $Y_i^* \equiv 0$  for inactive  $X_i$ 's), these  $d$ -balls are large enough to also cover those "inactive" dimensions.

### C. Technical Tools

1) *Eigenvalues of the Covariance Matrices:* Although decorrelation simplifies the problem by transforming a source with memory into a memoryless one, the real challenge is to study the evolution of the variances  $\sigma_i^2$ 's in (53), as  $n$  increases. For finite  $n$ , there is no closed-form expression for the eigenvalues of  $\Sigma_U$  for  $a \in (0, 1)$ .<sup>3</sup> Since the inverse of  $\Sigma_U$  is  $\frac{1}{\sigma^2} \mathbf{A}^\top \mathbf{A}$ , which is almost a Toeplitz matrix except the  $(n, n)$ -th entry, the limiting distribution of the eigenvalues of  $\Sigma_U$  can be deduced from the limiting distribution of eigenvalues of Toeplitz matrices [5, Eq. (19)].

**Theorem 3** (Reformulation of Gray [5, Eq. (19)]). *Fix any  $a \in [0, 1)$ . For any continuous function  $F(t)$  over the interval*

$$t \in [\theta_{\min}, \theta_{\max}], \quad (66)$$

*the eigenvalues  $\sigma_i^2$ 's of  $\Sigma_U$  satisfy*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n F(\sigma_i^2) = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(S(w)) dw, \quad (67)$$

where  $S(w)$  is defined in (35).

There are more general results in the form of Theorem 3, known as Szegő's theorem, see [32, Ch. 5] for Toeplitz forms and [33, Corollary 2.3] for asymptotically Toeplitz matrices. In the context of rate-distortion theory, applying Theorem 3 to (30)-(31) leads to (33)-(34).

Unfortunately, Theorem 3 is insufficient to obtain the fine asymptotics in our Theorem 1. To derive our finite blocklength results, we need to understand the rate of convergence in (67). Towards that end, we develop a nonasymptotic refinement of Theorem 3, presented next.

**Theorem 4** (Nonasymptotic Eigenvalue Distribution of  $\Sigma_U$ ). *Fix any  $a \in [0, 1)$ . For any bounded,  $L$ -Lipschitz and*

<sup>3</sup>A closed-form expression for the eigenvalues of  $\Sigma_U$  is known only for  $a = 1$  [31, Eq. (2)].



non-decreasing function  $F(t)$  over the interval in (66), and for any  $n \geq 1$ , the eigenvalues  $\sigma_i^2$ 's of  $\Sigma_U$  satisfy

$$\left| \frac{1}{n} \sum_{i=1}^n F(\sigma_i^2) - \frac{1}{2\pi} \int_{-\pi}^{\pi} F(S(w)) dw \right| \leq \frac{C_L}{n}, \quad (68)$$

where  $C_L > 0$  is a constant that depends on the Lipschitz constant  $L$  and the sup norm  $\|F\|_\infty$  of  $F$ , and  $S(w)$  is in (35).

*Proof.* Theorem 4 follows from Lemma 1 below and elementary analyses on Riemann sums. See Appendix C-B for details. ■

In the course of the proof of Theorem 4, we obtain the following nonasymptotic bounds on each eigenvalue  $\mu_i$  of  $A^\top A$ , which is of independent interest.

**Lemma 1** (Sharp Approximation of the Eigenvalues of  $A^\top A$ ). *Fix any  $a \in [0, 1]$ . For any  $n \in \mathbb{N}$ , let  $0 < \mu_1 \leq \mu_2 \leq \dots \leq \mu_n$  be the eigenvalues of  $A^\top A$ , and let*

$$\xi_i \triangleq g\left(\frac{i\pi}{n+1}\right), \quad (69)$$

where  $g$  is in (36). Then, we have

$$0 \leq \xi_i - \mu_i \leq \frac{2a\pi}{n}, \quad \forall i \in [n]. \quad (70)$$

*Proof.* The idea in proving Lemma 1 is that  $A^\top A$  is almost a tridiagonal Toeplitz matrix, whose eigenvalues are given by (69). The bound (70) is obtained via the Cauchy interlacing theorem and the Gershgorin circle theorem. See Appendix C-A for details. ■

*Remark 1.* In view of (69) and (70) in Lemma 1, we have  $\forall n \in \mathbb{N}$  and  $\forall i \in [n]$ ,

$$(1-a)^2 \leq \mu_i \leq (1+a)^2. \quad (71)$$

The key difference between the asymptotically stationary case ( $a \in [0, 1)$ ) and the nonstationary case ( $a \geq 1$ ) is that, in the later case,  $\mu_1$  decreases to zero as  $n$  increases to infinity, see [34, Lemma] and [31, Eq. (2)]. In the asymptotically stationary case,  $\mu_1$  is bounded away from zero according to (71).

2) *An Estimation Problem:* Our achievability proof relies on the analysis of the following parameter estimation problem. Given a source sequence  $\mathbf{u} = (u_1, \dots, u_n)^\top$ , drawn from the model (5) with unknown  $a$ , the maximum likelihood estimate (MLE) of the parameter  $a$  is (Appendix F-A)

$$\hat{a}(\mathbf{u}) = \frac{\sum_{i=1}^{n-1} u_i u_{i+1}}{\sum_{i=1}^{n-1} u_i^2}. \quad (72)$$

We show that the estimation error of the MLE decays exponentially in  $n$  for any  $a \in [0, 1]$ .

**Theorem 5.** *Fix  $a \in [0, 1]$ . Let  $\eta \in (0, 1)$ . Then, there exists a universal constant  $c > 0$  and two constants  $c_1, c_2 > 0$  ( $c_1$  and  $c_2$  only depend on  $a$ , see (377) in Appendix F-C below) such that for all  $n$  large enough, the estimation error of the MLE satisfies*

$$\mathbb{P}[|\hat{a}(\mathbf{U}) - a| > \eta] \leq 2 \exp\left[-c \min(c_1 \eta^2 n, c_2 \eta n)\right]. \quad (73)$$

*Proof.* Appendix F-B. ■

Finally, we present a strengthened version of Theorem 5, which is used in our achievability proof. Let  $\alpha > 0$  be a constant. Define  $\eta_n$  as

$$\eta_n \triangleq \sqrt{\frac{\alpha \log \log n}{n}}. \quad (74)$$

**Theorem 6.** *Fix  $a \in [0, 1]$ . Given a constant  $\alpha > 0$ , let  $\eta_n$  be in (74). Then, for all  $n$  large enough, the estimation error of the MLE satisfies*

$$\mathbb{P}[|\hat{a}(\mathbf{U}) - a| > \eta_n] \leq \frac{2}{(\log n)^{\kappa\alpha}}, \quad (75)$$

where  $\kappa$  is a constant given by

$$\kappa \triangleq \frac{c}{8(1-a^2)}, \quad (76)$$

and  $c > 0$  is the constant in Theorem 5.

*Proof.* Appendix F-C. ■

See Section V-B for the construction of a typical set based on  $\hat{a}(\mathbf{u})$ .

#### IV. CONVERSE

**Theorem 7** (Converse). *For the Gauss-Markov source (5) with the constant  $a \in [0, 1]$ , for any excess-distortion probability  $\epsilon \in (0, 1)$ , and for any distortion threshold  $d \in (0, d_{\max})$ , the minimum achievable source coding rate satisfies*

$$R(n, d, \epsilon) \geq \mathbb{R}_U(d) + \sqrt{\frac{\mathbb{V}_U(d)}{n}} Q^{-1}(\epsilon) - \frac{\log n}{2n} + O\left(\frac{1}{n}\right), \quad (77)$$

where  $\mathbb{R}_U(d)$  is the rate-distortion function given in (33), and  $\mathbb{V}_U(d)$  is the informational dispersion, defined in (25) and computed in (64).

We present two converse proofs in the following. The first one is a volumetric argument; while the second one relies on a general converse derived in [15, Th. 7] and a new concentration result on the  $d$ -tilted information of the Gauss-Markov source.

##### A. A Geometric Proof

Geometrically, any  $(n, M, d, \epsilon)$  code induces a covering of  $(\mathbb{R}^n, P_U)$ : the union of  $d$ -balls centered at the codewords have probability mass at least  $1 - \epsilon$ . Converting the underlying probability to  $P_Z$  and using the symmetry of  $\mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ , we obtain the following lower bound on the number of codewords  $M$ . The argument relies on  $\det \mathbf{A} = 1$ , where  $\mathbf{A}$  is in (48).

**Theorem 8.** *Given  $\epsilon \in (0, 1)$  and  $d \in (0, d_{\max})$ , the size of any  $(n, M, d, \epsilon)$  code for the Gauss-Markov source (5) must satisfy*

$$M \geq \left(\frac{r(n, \epsilon)}{d}\right)^{n/2}, \quad (78)$$

where  $r(n, \epsilon)$  is such that

$$\mathbb{P}(G < n \cdot r(n, \epsilon) / \sigma^2) = 1 - \epsilon, \quad (79)$$

and  $G$  is a random variable distributed according to the  $\chi^2$ -distribution with  $n$  degrees of freedom.

*Proof of Theorem 8.* Appendix D-A. ■

**Remark 2.** Theorem 8, which applies to the Gauss-Markov source, parallels [15, Th. 36], which applies to the i.i.d. Gaussian source. Both proofs rely on the volumetric method, though the proof of Theorem 8 requires additional arguments related to linear transformations of the underlying space. Theorem 8 yields the optimal second-order coding rate for the Gauss-Markov source only in the low distortion regime (as we will see in the proof of Theorem 7 below), while an analysis of [15, Th. 36] gives the optimal second-order coding rate for the i.i.d. Gaussian source of any distortion [15, Th. 40].

Equipped with Theorem 8, we are ready to prove the converse in Theorem 7 for  $d \in (0, d_c]$ .

*Proof of Theorem 7 below the critical distortion.* Applying the Berry-Esseen Theorem in Appendix A-B to (79) yields

$$r(n, \epsilon) \geq \sigma^2 \left[ 1 + \sqrt{\frac{2}{n}} Q^{-1} \left( \epsilon + \frac{C_{BE}}{\sqrt{n}} \right) \right]. \quad (80)$$

Plugging (80) into (78) and taking logarithms, we obtain

$$R(n, d, \epsilon) \geq \frac{1}{2} \log \frac{\sigma^2}{d} + \sqrt{\frac{1}{2n}} Q^{-1}(\epsilon) + O\left(\frac{1}{n}\right), \quad (81)$$

where we use the Taylor expansions of  $\log(1+x)$  and the inverse  $Q$ -function. The converse bound (81) holds for any  $\epsilon \in (0, 1)$  and  $d \in (0, d_{\max})$ . By (39) and (58), we see that (81) is the same as (77) for  $d \in (0, d_c]$ , up to the second-order term. In addition, (81) is slightly stronger than (77) in the third-order term. For  $d \in (d_c, d_{\max})$ , (81) is not tight, even in the first order since  $\mathbb{R}_U(d) > \frac{1}{2} \log \frac{\sigma^2}{d}$  for  $d \in (d_c, d_{\max})$ , by (39) and (43). ■

**Remark 3.** The converse (81) holds for the general Gaussian AR processes defined in (37). The proof stays the same, except that the matrix  $\mathbf{A}$  in (48) is replaced by

$$\mathbf{A} \triangleq \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -a_1 & 1 & 0 & \cdots & 0 \\ -a_2 & -a_1 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ -a_{n-1} & \cdots & -a_2 & -a_1 & 1 \end{bmatrix}. \quad (82)$$

### B. Converse Proof

The second proof is based on a general converse by Kostina and Verdú [15], restated here for convenience, and a concentration result which bounds the difference between  $J_{\mathbf{X}}(\mathbf{X}, d)$  and its approximation  $J_{\mathbf{X}}(\mathbf{X}, d_n)$ , for  $d_n$  defined in (86) below.

**Theorem 9** ([15, Th. 7]). Fix  $d \in (0, d_{\max})$ . Any  $(n, M, d, \epsilon)$  code must satisfy

$$\epsilon \geq \sup_{\gamma \geq 0} \mathbb{P}[J_{\mathbf{X}}(\mathbf{X}, d) \geq \log M + \gamma] - \exp(-\gamma). \quad (83)$$

The converse bound in Theorem 9 above provides a lower bound on  $\epsilon$  for any  $(n, M, d, \epsilon)$  code using the  $\mathbf{d}$ -tilted information, and is used to derive a converse result on the dispersion of the stationary memoryless sources in [15, Eq. (103)–(106)]. The key step in the proof of [15, Eq. (103)–(106)] is to write the  $\mathbf{d}$ -tilted information as a sum of  $n$  i.i.d. random variables, to which the Berry-Esseen Theorem is applied.

For the Gauss-Markov source  $\mathbf{X}$ , using (22), (31) and (65), we can write the  $\mathbf{d}$ -tilted information  $J_{\mathbf{X}}(\mathbf{X}, d)$  as a sum of  $n$  independent (but not identical) random variables:

$$J_{\mathbf{X}}(\mathbf{X}, d) = \sum_{i=1}^n J_{X_i}(X_i, \min(\theta_n, \sigma_i^2)), \quad (84)$$

where  $\theta_n$  is given in (31). Indeed, (84) is further simplified to (63) in the proof of Theorem 2. However, it is hard to conduct nonasymptotic analysis using (84) since understanding the evolution of both  $\theta_n$  and  $\sigma_i^2$ 's as  $n$  grows in (84) is challenging. Therefore, we approximate  $J_{\mathbf{X}}(\mathbf{X}, d)$  using

$$J_{\mathbf{X}}(\mathbf{X}, d_n) = \sum_{i=1}^n J_{X_i}(X_i, \min(\theta, \sigma_i^2)), \quad (85)$$

where

$$d_n \triangleq \frac{1}{n} \sum_{i=1}^n \min(\theta, \sigma_i^2), \quad (86)$$

and  $\theta$  is the water level matched to  $d$  via the limiting reverse waterfilling (34). Then,  $\theta$  does not depend on  $n$  in (85). Since our Theorem 4 and Lemma 1 in Section III-C.1 capture the evolution of  $\sigma_i^2$ 's as  $n$  grows, (85) is easier to analyze than (84) in the nonasymptotic regime. Throughout the paper, the relations among a given distortion  $d$ , the water levels  $\theta$ ,  $\theta_n$ , and the distortion  $d_n$  defined in (86), are  $\theta_n \xleftrightarrow{(31)} d \xleftrightarrow{(34)} \theta \xleftrightarrow{(31)} d_n$ . Note that there is no direct reverse waterfilling relation between  $d_n$  in (86) and  $\theta_n$  in (31). As shown by our concentration result Theorem 10 in the following, the approximation  $J_{\mathbf{X}}(\mathbf{X}, d_n)$  stays within a constant from  $J_{\mathbf{X}}(\mathbf{X}, d)$  with probability at least  $1 - O(\frac{1}{n})$ .

**Theorem 10** (Approximation of the  $\mathbf{d}$ -Tilted Information). For any  $d \in (0, d_{\max})$ , let  $\theta > 0$  be the water level matched to  $d$  via the limiting reverse waterfilling (34). Suppose we have a sequence of distortion levels  $d_n \in (0, d_{\max})$  with the property that there exists a constant  $h_1 > 0$  such that for all  $n$  large enough,

$$|d - d_n| \leq \frac{h_1}{n}. \quad (87)$$

Then, there exists a constant  $\tilde{c} \in (0, 1)$  such that for any  $u > \frac{2h_1}{\tilde{c}\theta}$  and all  $n$  large enough, we have

$$\mathbb{P}[|J_{\mathbf{X}}(\mathbf{X}, d) - J_{\mathbf{X}}(\mathbf{X}, d_n)| \leq u] \geq 1 - \frac{1}{n \left( \frac{\tilde{c}\theta u}{2h_1} - 1 \right)^2}. \quad (88)$$

*Proof of Theorem 10.* Appendix D-B. ■

In the rest of this section, we present the detailed proof of Theorem 7 for any  $d \in (0, d_{\max})$ . The  $\mathbf{d}$ -tilted information



$J_{\mathbf{X}}(\mathbf{X}, d)$  is first approximated by  $J_{\mathbf{X}}(\mathbf{X}, d_n)$  defined in (85), which is a sum of independent random variables whose expectations and variances approximate the rate-distortion function  $\mathbb{R}_U(d)$  and the informational dispersion  $\mathbb{V}_U(d)$ , respectively. Combining these approximation bounds and Theorem 9, we obtain the converse in (77). The details follow.

*Proof of Theorem 7.* Fix  $d \in (0, d_{\max})$ . Let  $\theta > 0$  be the water level matched to  $d$  via the limiting reverse waterfilling (34). Notice that  $d_n$ , defined in (86), is the distortion matched to the water level  $\theta$  via the  $n$ -th order reverse waterfilling (31) over  $\sigma_i^2$ 's. Comparing (31) and (86), and applying Theorem 4 to the function  $t \mapsto \min(\theta, t)$ , we deduce that there exists a constant  $C_d > 0$  such that for any  $n \geq 1$ ,

$$|d - d_n| \leq \frac{C_d}{n}. \quad (89)$$

Let  $\bar{\mathbf{Y}}^*$  be the  $n$ -dimensional Gaussian random vector such that  $(\mathbf{X}, \bar{\mathbf{Y}}^*)$  forms a RDF-achieving pair in  $\mathbb{R}_{\mathbf{X}}(n, d_n)$  defined in (19). Note that  $\bar{\mathbf{Y}}^*$  defined here is indeed different from  $\mathbf{Y}^*$  in (65), where  $(\mathbf{X}, \mathbf{Y}^*)$  forms a RDF-achieving pair in  $\mathbb{R}_{\mathbf{X}}(n, d)$ . It is well-known [10, Th. 10.3.3] that  $\bar{\mathbf{Y}}^*$  has independent coordinates and similar to (65),

$$\bar{Y}_i^* \sim \mathcal{N}(0, \max(\sigma_i^2 - \theta, 0)). \quad (90)$$

By the independence of  $\bar{Y}_i^*$ 's, (18) and (22), we have

$$J_{\mathbf{X}}(\mathbf{X}, d_n) = \sum_{i=1}^n \Lambda_{\bar{Y}_i^*}(X_i, \lambda^*, \min(\theta, \sigma_i^2)), \quad (91)$$

where

$$\lambda^* = -\mathbb{R}'_{\mathbf{X}}(n, d_n) = -\mathbb{R}'(\mathbf{X}, \bar{\mathbf{Y}}^*, d_n). \quad (92)$$

Denote by  $\mathbb{E}_i$  and  $\mathbb{V}_i$  the means and the variances of  $\Lambda_{\bar{Y}_i^*}(X_i, \lambda^*, \min(\theta, \sigma_i^2))$  (the summands in (91)). By the same computations leading to (233) and (234) in Appendix C-D, we have

$$\mathbb{E}_i = \max\left(0, \frac{1}{2} \log \frac{\sigma_i^2}{\theta}\right), \quad (93)$$

$$\mathbb{V}_i = \min\left(\frac{1}{2}, \frac{\sigma_i^4}{2\theta^2}\right). \quad (94)$$

We now derive the approximation of  $\mathbb{R}_U(d)$  and  $\mathbb{V}_U(d)$  using the means  $\mathbb{E}_i$ 's and the variances  $\mathbb{V}_i$ 's, respectively. Applying Theorem 4 to the function  $t \mapsto \max(0, \frac{1}{2} \log \frac{t}{\theta})$  in (21) and (93), and to the function  $t \mapsto \min(\frac{1}{2}, \frac{t^2}{2\theta^2})$  in (64) and (94), we conclude that there exist two constants  $c_r, c_v > 0$  (depending on  $d$  only) such that

$$\left| n\mathbb{R}_U(d) - \sum_{i=1}^n \mathbb{E}_i \right| \leq c_r, \quad (95)$$

$$\left| \sqrt{n\mathbb{V}_U(d)} - \sqrt{\sum_{i=1}^n \mathbb{V}_i} \right| \leq c_v. \quad (96)$$

Next, we consider the sequence of distortion levels  $\{d_n\}_{n=1}^{\infty}$ , which satisfies the condition (87) due to (89). Define the event

$$\mathcal{E} \triangleq \left\{ J_{\mathbf{X}}(\mathbf{X}, d) \geq J_{\mathbf{X}}(\mathbf{X}, d_n) - \frac{4C_d}{\tilde{c}\theta} \right\}, \quad (97)$$

where  $\tilde{c} \in (0, 1)$  is the constant in Theorem 10 and  $C_d > 0$  is the constant in (89). Theorem 10 implies that

$$\mathbb{P}[\mathcal{E}] \geq 1 - \frac{1}{n}. \quad (98)$$

Letting  $\gamma = \frac{1}{2} \log n$  in Theorem 9, we see that if an  $(n, M, d, \epsilon')$ -excess-distortion code exists, then

$$\epsilon' \geq \mathbb{P} \left[ J_{\mathbf{X}}(\mathbf{X}, d) \geq \log M + \frac{\log n}{2} \right] - \frac{1}{\sqrt{n}} \quad (99)$$

$$\geq \mathbb{P} \left[ J_{\mathbf{X}}(\mathbf{X}, d) \geq \log M + \frac{\log n}{2} \mid \mathcal{E} \right] \mathbb{P}[\mathcal{E}] - \frac{1}{\sqrt{n}} \quad (100)$$

$$\geq \left(1 - \frac{1}{n}\right) \mathbb{P} \left[ J_{\mathbf{X}}(\mathbf{X}, d_n) \geq \log M + \frac{\log n}{2} + \frac{4C_d}{\tilde{c}\theta} \right] - \frac{1}{\sqrt{n}}, \quad (101)$$

where (101) is due to (98) and (97). For any fixed  $\epsilon \in (0, 1)$ , define  $\epsilon_n$  as

$$\epsilon_n \triangleq \epsilon + \exp(-\gamma) + \frac{C_{\text{BE}}}{\sqrt{n}} + \frac{1}{n}, \quad (102)$$

where  $C_{\text{BE}}$  is the constant in the Berry-Esseen Theorem in Appendix A-B. Then, we have  $\epsilon_n \in (0, 1)$  for all  $n$  large enough. We choose  $M$  as

$$\log M \triangleq n\mathbb{R}_U(d) + \sqrt{n\mathbb{V}_U(d)}Q^{-1}(\epsilon_n) - \gamma - c_r - c_v \left| Q^{-1}(\epsilon_n) \right| - \frac{4C_d}{\tilde{c}\theta}. \quad (103)$$

From (95), (96) and (103), we have

$$\log M \leq \sum_{i=1}^n \mathbb{E}_i + Q^{-1}(\epsilon_n) \sqrt{\sum_{i=1}^n \mathbb{V}_i} - \gamma - \frac{4C_d}{\tilde{c}\theta}. \quad (104)$$

Continuing the inequality in (101), we have

$$\epsilon' \geq \left(1 - \frac{1}{n}\right) \mathbb{P} \left[ \sum_{i=1}^n \Lambda_{\bar{Y}_i^*}(X_i, \lambda^*, \min(\theta, \sigma_i^2)) \geq \right]$$

$$\sum_{i=1}^n \mathbb{E}_i + Q^{-1}(\epsilon_n) \sqrt{\sum_{i=1}^n \mathbb{V}_i} - \frac{1}{\sqrt{n}} \quad (105)$$

$$\geq \left(1 - \frac{1}{n}\right) \left( \epsilon_n - \frac{C_{\text{BE}}}{\sqrt{n}} \right) - \frac{1}{\sqrt{n}} \quad (106)$$

$$\geq \epsilon_n - \frac{1}{n} - \frac{1 + C_{\text{BE}}}{\sqrt{n}} \quad (107)$$

$$= \epsilon, \quad (108)$$

where (105) is by (91) and the bound (104); (106) is by the Berry-Esseen Theorem in Appendix A-B; and (108) is by the choice of  $\epsilon_n$  in (102). Consequently, for all  $n$  large enough, any  $(n, M, d, \epsilon')$ -excess-distortion code must satisfy  $\epsilon' \geq \epsilon$ , so we must have

$$R(n, d, \epsilon) \geq \frac{\log M}{n}. \quad (109)$$

Plugging (103) into (109) and applying the Taylor expansion to  $Q^{-1}(\epsilon_n)$  yields (77). ■

## V. ACHIEVABILITY

**Theorem 11** (Achievability). *Fix any  $\alpha > 0$ . Consider the Gauss-Markov source defined in (5). For any excess-distortion probability  $\epsilon \in (0, 1)$ , and any distortion threshold  $d \in (0, d_{\max})$ , the minimum achievable source coding rate is bounded as*

$$R(n, d, \epsilon) \leq \mathbb{R}_U(d) + \sqrt{\frac{\mathbb{V}_U(d)}{n}} Q^{-1}(\epsilon) + O\left(\frac{1}{(\log n)^{\kappa\alpha} \sqrt{n}}\right), \quad (110)$$

where  $\mathbb{R}_U(d)$  is the rate-distortion function given in (33);  $\mathbb{V}_U(d)$  is the informational dispersion, defined in (25) and computed in (64); and  $\kappa > 0$  is the constant in (76).

This section presents the proof of Theorem 11. We first discuss how bounds on the covering number of  $n$ -dimensional ellipsoids can be converted into achievability results. We then proceed to present our proof of Theorem 11, which relies on random coding [15, Corollary 11], and a lower bound on the probability of a distortion  $d$ -ball using the  $d$ -tilted information  $J_{\mathbf{X}}(\mathbf{X}, d)$ .

### A. Connections to Covering Number

Dumer *et al.* [35] considered the problem of covering an ellipsoid using the minimum number of balls in  $\mathbb{R}^n$ , and derived lower and upper bounds on that number. Although any upper bound on the covering number implies an upper bound on  $R(n, d, \epsilon)$ , the upper bound on covering number in [35] is not tight enough to yield the achievability direction of the Gaussian approximation (56). We proceed to explain how to obtain a bound on  $R(n, d, \epsilon)$  from the results in [35]. An ellipsoid  $E_{\mathbf{r}}^n$  is defined by

$$E_{\mathbf{r}}^n \triangleq \left\{ \mathbf{x} \in \mathbb{R}^n : \sum_{i=1}^n \frac{x_i^2}{r_i^2} \leq 1 \right\}, \quad (111)$$

where  $\mathbf{r} = (r_1, \dots, r_n)$ , and  $r_i > 0$  is one half of the length of the  $i$ -axis of  $E_{\mathbf{r}}^n$ . We say that a subset  $\mathcal{M}_d \subset \mathbb{R}^n$  is a  $d$ -covering<sup>4</sup> of the ellipsoid  $E_{\mathbf{r}}^n$  if

$$E_{\mathbf{r}}^n \subseteq \bigcup_{\mathbf{y} \in \mathcal{M}_d} B(\mathbf{y}, d), \quad (112)$$

where  $B(\mathbf{y}, d)$  is the  $d$ -ball centered at  $\mathbf{y}$ , defined in (4). The covering number  $\mathcal{N}(n, d)$  of an ellipsoid  $E_{\mathbf{r}}^n$  is defined as the size of its minimal  $d$ -covering. The  $d$ -entropy  $H_d(E_{\mathbf{r}}^n)$  is the logarithm of the covering number

$$H_d(E_{\mathbf{r}}^n) \triangleq \log \mathcal{N}(n, d). \quad (113)$$

The result in [35, Th. 2] states that

$$H_d(E_{\mathbf{r}}^n) = K_d + o(K_d), \quad (114)$$

where

$$K_d \triangleq \sum_{i: r_i^2 > nd} \frac{1}{2} \log \frac{r_i^2}{nd}. \quad (115)$$

<sup>4</sup>In [35], the term  $\epsilon$ -covering was used instead of  $d$ -covering used here. They are related by  $\epsilon = \sqrt{nd}$ .

Despite the similarity between (115) and the reverse waterfilling (30), the result in (114) is not strong enough to recover even the asymptotic rate-distortion tradeoff (30) unless  $d \leq d_c$ .

In our problem, let  $\mathbf{X}$  be the decorrelation of  $\mathbf{U}$  in (50), then  $X_1, \dots, X_n$  are independent zero-mean Gaussian distributed with variances being  $\sigma_i^2$  defined in (53). The random vector  $\mathbf{X}$  concentrates around an ellipsoid with probability mass at least  $1 - \epsilon$ . Applying the Berry-Esseen theorem to express  $r_i$ 's in (115), we deduce that for any  $\epsilon \in (0, 0.5)$  and  $d \in (0, d_c]$ ,

$$R(n, d, \epsilon) \leq \frac{1}{2} \log \frac{\sigma^2}{d} + \frac{Q^{-1}(\epsilon)}{\sqrt{2n}} + o(1), \quad (116)$$

where the extra  $o(1)$  term comes from the  $o(K_d)$  term in (114). Due to that  $o(1)$  term, the bound (116) is first-order optimal, but not second-order optimal. Strengthening (114) to  $H_d(E_{\mathbf{r}}^n) = K_d + o(\sqrt{K_d})$  would allow one to replace the  $o(1)$  term in (116) by  $o\left(\frac{1}{\sqrt{n}}\right)$ , yielding the  $\leq$  (achievability) direction of the Gaussian approximation (56) in the regime of  $d \in (0, d_c]$ . We do not pursue this approach here. Instead, we prove (56) via the tilted information.

### B. Outline of the Achievability Proof

We describe the main ideas in our achievability proof and present the details in next subsection. Our proof is inspired by the work of Kostina and Verdú [15, Th. 12], where the same problem was addressed for the stationary memoryless sources. However, the proof there cannot be directly applied to the Gauss-Markov source. The random coding bound, stated next, provides an upper bound on the excess-distortion probability  $\epsilon$  using the probability of the distortion  $d$ -balls.

**Lemma 2** (Random Coding Bound). *Let  $\mathbf{X}$  be the decorrelation of  $\mathbf{U}$  in (50). There exists an  $(n, M, d, \epsilon)$  code with*

$$\epsilon \leq \inf_{P_{\mathbf{Y}}} \mathbb{E}_{\mathbf{X}} \left[ e^{-M P_{\mathbf{Y}}(B(\mathbf{X}, d))} \right], \quad (117)$$

where the infimum is over all pdf's  $P_{\mathbf{Y}}$  on  $\mathbb{R}^n$  with  $\mathbf{Y}$  independent of  $\mathbf{X}$ .

*Proof.* A direct application of [15, Corollary 11] to  $\mathbf{X}$ . ■

The next lemma provides a lower bound on the probability of the distortion  $d$ -balls using the  $d$ -tilted information  $J_{\mathbf{X}}(\mathbf{X}, d)$ .

**Lemma 3** (Lossy AEP for the Gauss-Markov Sources). *Fix any  $\alpha > 0$  and let  $\eta_n$  be in (74) in Section III-C.2 above. For any  $d \in (0, d_{\max})$  and  $\epsilon \in (0, 1)$ , there exists a constant  $K > 0$  such that for all  $n$  large enough,*

$$\begin{aligned} \mathbb{P} \left[ \log \frac{1}{P_{\mathbf{Y}^*}(B(\mathbf{X}, d))} \leq J_{\mathbf{X}}(\mathbf{X}, d) + \beta_1 \log^q n + \beta_2 \right] \\ \geq 1 - \frac{K}{(\log n)^{\kappa\alpha}}, \end{aligned} \quad (118)$$

where  $\mathbf{X}$  is the decorrelation of  $\mathbf{U}$  in (50);  $(\mathbf{X}, \mathbf{Y}^*)$  forms a RDF-achieving pair in  $\mathbb{R}_{\mathbf{X}}(n, d)$ , and  $q > 1, \beta_1 > 0, \beta_2$  are constants, see (271) and (272) in Appendix E-A below. The constant  $\kappa > 0$  is in (76) in Section III-C.2 above.

*Proof.* Appendix E-A. ■



Together with  $\log \frac{1}{P_{Y^*}(\mathcal{B}(x,d))} \geq J_X(x,d)$  in [15, Eq. (26)], obtained by applying Markov's inequality to (22), Lemma 3 establishes the link between the probability of distortion  $d$ -ball and the  $d$ -tilted information:  $\log \frac{1}{P_{Y^*}(\mathcal{B}(x,d))} \approx J_X(\mathbf{X}, d)$  for the Gauss-Markov source. Results of this kind were referred to as lossy asymptotic equipartition property (AEP) in [29, Sec. I.B].

Lemma 3 is the key lemma in our achievability proof for the Gauss-Markov sources. The proof of Lemma 3 is one of the main technical contributions of this paper. An analog of Lemma 3 for the stationary memoryless sources [15, Lemma 2] has been used to prove the non-asymptotic achievability result [15, achievability proof of Th. 12]. Showing a lower bound on the probability of distortion  $d$ -balls in terms of  $J_X(\mathbf{X}, d)$ , that is, in the form of (118), is technical even for i.i.d. sources. To derive such a bound for the Gauss-Markov sources, we rely on fundamentally new ideas, including the maximum likelihood estimator  $\hat{a}(\mathbf{u})$  defined in (72) and analyzed in Theorem 6 in Section III-C.2 above. We proceed to discuss the estimator and its role in the proof of Lemma 3 next.

A major step in proving [15, Lemma 2] for the i.i.d. source  $\{X_i\}$  with  $X_i \sim P_X$  involves the empirical probability distribution  $P_{\hat{\mathbf{X}}}$ : given a source sequence  $\mathbf{x}$ ,  $P_{\hat{\mathbf{X}}}(x) \triangleq \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{x_i = x\}$ . The product of the empirical distributions  $P_{\hat{\mathbf{X}}} \triangleq P_{\hat{\mathbf{X}}} \times \dots \times P_{\hat{\mathbf{X}}}$  was used in the proof of [15, Lemma 2] for the i.i.d. sources [15, Eq. (270)] to form a typical set of source outcomes.

To describe a typical set of outcomes of the Gauss-Markov source, to each source outcome  $\mathbf{x}$  (equivalently,  $\mathbf{u}$ ) we associate a proxy random variable  $\hat{\mathbf{X}}(\mathbf{x})$  as follows. We first estimate the parameter  $a$  in (5) from the source outcome  $\mathbf{u}$  using the maximum likelihood estimator  $\hat{a}(\mathbf{u})$  in (72) in Section III-C.2 above. Then, the proxy random variable  $\hat{\mathbf{X}}(\mathbf{x})$  is defined as a Gaussian random vector with independent (but not identical) coordinates  $\hat{X}_i(\mathbf{x}) \sim \mathcal{N}(0, \hat{\sigma}_i^2(\mathbf{x}))$ , where  $\hat{\sigma}_i^2(\mathbf{x})$ 's are the proxy variances defined using  $\hat{a}(\mathbf{u})$ :

$$\hat{\sigma}_i^2(\mathbf{x}) \triangleq \frac{\sigma^2}{1 + \hat{a}(\mathbf{u})^2 - 2\hat{a}(\mathbf{u}) \cos(i\pi/(n+1))}. \quad (119)$$

Equivalently,  $\hat{\mathbf{X}}(\mathbf{x})$  is a zero-mean Gaussian random vector whose distribution is given by

$$\hat{\mathbf{X}}(\mathbf{x}) \sim \mathcal{N}(\mathbf{0}, \text{diag}(\hat{\sigma}_1^2(\mathbf{x}), \dots, \hat{\sigma}_n^2(\mathbf{x}))). \quad (120)$$

To simplify notations, when there is no ambiguity, we will write  $\hat{\mathbf{X}}$  and  $\hat{\sigma}_i^2$  for  $\hat{\mathbf{X}}(\mathbf{x})$  and  $\hat{\sigma}_i^2(\mathbf{x})$ , respectively. Intuitively, the formula (119) approximates the eigenvalues of the covariance matrix of  $\mathbf{U}$  (or equivalently, that of  $\mathbf{X}$ ) for a typical  $\mathbf{x}$ . Due to Theorem 6, with probability approaching 1, we have  $\hat{a}(\mathbf{u}) \approx a$ , which implies  $\hat{\sigma}_i^2 \approx \sigma_i^2$  and  $\hat{\mathbf{X}} \approx \mathbf{X}$ . The accuracy of these approximations is quantified in Theorem 12 below, which is the main tool in the proof of Lemma 3.

We need a few notations before presenting Theorem 12. First, we particularize the CREM problem (12)–(17) to the Gauss-Markov source. Let  $\mathbf{X}$  be the decorrelation of  $\mathbf{U}$  in (50). For any random vector  $\mathbf{Y}$  with density, replacing  $\mathbf{X}$  by  $\mathbf{X}$

in (12) and normalizing by  $n$ , we define

$$\mathbb{R}(\mathbf{X}, \mathbf{Y}, d) \triangleq \inf_{P_{\mathbf{F}|\mathbf{X}}: \mathbb{E}[d(\mathbf{X}, \mathbf{F})] \leq d} \frac{1}{n} D(P_{\mathbf{F}|\mathbf{X}} \| P_{\mathbf{Y}} | P_{\mathbf{X}}). \quad (121)$$

Properties of the CREM (121) for the two special cases: when (i)  $\mathbf{Y}$  is a Gaussian random vector with independent coordinates and (ii)  $(\mathbf{X}, \mathbf{Y})$  forms a RDF-achieving pair, are presented in Appendix C-C. Let  $\hat{\mathbf{F}}^*$  be the optimizer of  $\mathbb{R}(\hat{\mathbf{X}}, \mathbf{Y}^*, d)$ , where  $\hat{\mathbf{X}}$  is defined in (120) and  $\mathbf{Y}^*$  in (65). For  $\mathbf{x} \in \mathbb{R}^n$ , define  $m_i(\mathbf{x})$  as

$$m_i(\mathbf{x}) \triangleq \mathbb{E}[(\hat{F}_i^* - x_i)^2 | \hat{X}_i = x_i]. \quad (122)$$

**Definition 1** (MLE-Typical Set). Fix any  $d \in (0, d_{\max})$ . Given a constant  $\alpha > 0$ , let  $\eta_n$  be in (74) in Section III-C.2 above. For any constant  $p > 0$  and any  $n \in \mathbb{N}$ , define  $\mathcal{T}(n, \alpha, p)$  as the set of vectors  $\mathbf{u} \in \mathbb{R}^n$  satisfying the following conditions:

$$|\hat{a}(\mathbf{u}) - a| \leq \eta_n, \quad (123)$$

$$\left| \frac{1}{n} \sum_{i=1}^n m_i(\mathbf{x}) - d \right| \leq p\eta_n, \quad (124)$$

$$\left| \frac{1}{n} \sum_{i=1}^n \left( \frac{x_i^2}{\sigma_i^2} \right)^k - (2k-1)!! \right| \leq 2, \text{ for } k = 1, 2, 3, \quad (125)$$

where  $\mathbf{x} = \mathbf{S}^\top \mathbf{u}$ , and  $m_i(\mathbf{x})$ 's are functions of  $\mathbf{x}$  defined in (122) above.

The condition (123) requires that  $\mathbf{u} \in \mathcal{T}(n, \alpha, p)$  should yield a small estimation error, which holds with probability approaching 1 due to Theorem 6. We will explain the condition (124) in Appendix E-D below. To gain insight into the condition (125), note that due to (53), we have  $\frac{X_i}{\sigma_i} \sim \mathcal{N}(0, 1)$  and

$$\mathbb{E} \left[ \left( \frac{X_i^2}{\sigma_i^2} \right)^k \right] = (2k-1)!! \quad (126)$$

Therefore, the condition (125) bounds the variations of  $\mathbf{X}$ , up to its sixth moments, and this condition holds with probability approaching 1 by the Berry-Esseen theorem. Theorem 12 below summarizes the properties of the typical set  $\mathcal{T}(n, \alpha, p)$  used in the proof of Lemma 3.

**Theorem 12** (Properties of the MLE-Typical Set). For any  $d \in (0, d_{\max})$  and any constant  $\alpha > 0$ , let  $\eta_n$  be given in (74) in Section III-C.2 above and  $p$  be a sufficiently large constant (specifically,  $p \geq (298)$  in Appendix E-B below), then we have:

(1). The probability mass of  $\mathcal{T}(n, \alpha, p)$  is large: there exists a constant  $A_1 > 0$  such that for all  $n$  large enough,

$$\mathbb{P}[\mathbf{U} \in \mathcal{T}(n, \alpha, p)] \geq 1 - \frac{A_1}{(\log n)^{\kappa\alpha}}, \quad (127)$$

where the constant  $\kappa$  is defined in (76) in Section III-C.2 above.

(2). The proxy variances are good approximations: there exists a constant  $A_2 > 0$  such that for all  $n$  large enough, for any  $\mathbf{u} \in \mathcal{T}(n, \alpha, p)$ , it holds that

$$|\hat{\sigma}_i^2(\mathbf{x}) - \sigma_i^2| \leq A_2 \eta_n, \quad \forall i \in [n], \quad (128)$$

where  $\hat{\sigma}_i^2(\mathbf{x})$ 's are defined in (119).

- (3). Let  $\theta > 0$  be the water level matched to  $d$  via the limiting reverse waterfilling (34). For all  $n$  large enough, for any  $\mathbf{u} \in T(n, \alpha, p)$ , it holds that

$$\left| \hat{\lambda}^*(\mathbf{x}) - \lambda^* \right| \leq \frac{9A_2}{4\theta^2} \eta_n, \quad (129)$$

where  $\mathbf{x} = \mathbf{S}^\top \mathbf{u}$  with  $\mathbf{S}$  in (51);  $\lambda^*$  is given by (23);

$$\hat{\lambda}^*(\mathbf{x}) = -\mathbb{R}'(\hat{\mathbf{X}}, \mathbf{Y}^*, d); \quad (130)$$

$\mathbf{X}$  is the decorrelation of  $\mathbf{U}$  in (50);  $(\mathbf{X}, \mathbf{Y}^*)$  forms a RDF-achieving pair in  $\mathbb{R}_{\mathbf{X}}(n, d)$ ; and  $\hat{\mathbf{X}}$  is the proxy Gaussian random variable defined in (120).

*Proof.* Appendix E-B.

### C. Achievability Proof

*Proof of Theorem 11.* The proof is based on the random coding bound Lemma 2 and the lower bound in Lemma 3. Fix any  $d \in (0, d_{\max})$  and  $\epsilon \in (0, 1)$ , and let  $\theta > 0$  be the water level matched to  $d$  via the limiting reverse waterfilling (34). We reuse the notations in (89)-(96). Similar to the event  $\mathcal{E}$  in (97), we define the event  $\mathcal{F}$  as

$$\mathcal{F} \triangleq \left\{ J_{\mathbf{X}}(\mathbf{X}, d) \leq J_{\mathbf{X}}(\mathbf{X}, d_n) + \frac{4C_d}{\tilde{c}\theta} \right\}. \quad (131)$$

Theorem 10 implies that

$$\mathbb{P}[\mathcal{F}] \geq 1 - \frac{1}{n}. \quad (132)$$

Define  $\epsilon_n$  as

$$\epsilon_n \triangleq \epsilon - \frac{C_{\text{BE}} + 1}{\sqrt{n}} - \frac{K}{(\log n)^{\kappa\alpha}} - \frac{1}{n}. \quad (133)$$

Since  $\epsilon \in (0, 1)$ , we have  $\epsilon_n \in (0, 1)$  for all  $n$  large enough. Choose  $M$  as

$$\log M \triangleq n\mathbb{R}_U(d) + \sqrt{n\mathbb{V}_U(d)}Q^{-1}(\epsilon_n) + \log \frac{\log n}{2} + \beta_1 \log^q n + \beta_2 + c_r + c_v \left| Q^{-1}(\epsilon_n) \right| + \frac{4C_d}{\tilde{c}\theta}, \quad (134)$$

where  $q > 1$ ,  $\beta_1 > 0$ ,  $\beta_2$  are the constants in Lemma 3; and  $c_r, c_v$  are the positive constants in (95) and (96). Define the random variable  $G_n$  as

$$G_n \triangleq \log M - J_{\mathbf{X}}(\mathbf{X}, d_n) - \beta_1 \log^q n - \beta_2 - \frac{4C_d}{\tilde{c}\theta}, \quad (135)$$

where  $J_{\mathbf{X}}(\mathbf{X}, d_n)$  is in (91). By (95), (96), and (135), we have

$$G_n \geq \sum_{i=1}^n \mathbb{E}_i + Q^{-1}(\epsilon_n) \sqrt{\sum_{i=1}^n \mathbb{V}_i} - J_{\mathbf{X}}(\mathbf{X}, d_n) + \log \frac{\log n}{2}, \quad (136)$$

where  $\mathbb{E}_i$ 's and  $\mathbb{V}_i$ 's are defined in (93) and (94), respectively. Define the event  $\mathcal{G}$  as

$$\mathcal{G} \triangleq \left\{ G_n < \log \frac{\log n}{2} \right\}. \quad (137)$$

By (136), (91) and the Berry-Esseen Theorem, we have

$$\mathbb{P}[\mathcal{G}] \leq \mathbb{P} \left[ J_{\mathbf{X}}(\mathbf{X}, d_n) - \sum_{i=1}^n \mathbb{E}_i > Q^{-1}(\epsilon_n) \sqrt{\sum_{i=1}^n \mathbb{V}_i} \right] \quad (138)$$

$$\leq \epsilon_n + \frac{C_{\text{BE}}}{\sqrt{n}}. \quad (139)$$

Define the event  $\mathcal{L}$  as:

$$\mathcal{L} \triangleq \left\{ \log \frac{1}{P_{\mathbf{Y}^*}(\mathcal{B}(\mathbf{X}, d))} \leq \log M - G_n \right\} \quad (140)$$

$$\blacksquare = \left\{ \log \frac{1}{P_{\mathbf{Y}^*}(\mathcal{B}(\mathbf{X}, d))} \leq J_{\mathbf{X}}(\mathbf{X}, d_n) + \beta_1 \log^q n + \beta_2 + \frac{4C_d}{\tilde{c}\theta} \right\}, \quad (141)$$

where  $\mathbf{Y}^*$  is given in (65). Combining Lemma 3 and (132) yields

$$\mathbb{P}[\mathcal{L}] \geq 1 - \frac{1}{n} - \frac{K}{(\log n)^{\kappa\alpha}}. \quad (142)$$

Indeed, denoting the probability on the left-hand side of (118) by  $\mathbb{P}[\mathcal{H}]$ , we have

$$\mathbb{P}[\mathcal{H}] = \mathbb{P}[\mathcal{H} \cap \mathcal{F}] + \mathbb{P}[\mathcal{H} \cap \mathcal{F}^c] \quad (143)$$

$$\leq \mathbb{P}[\mathcal{L}] + \frac{1}{n}, \quad (144)$$

where (144) holds since  $\mathcal{H} \cap \mathcal{F} \subseteq \mathcal{L}$ .

We have now gathered all the ingredients to prove Theorem 11. Replacing  $\mathbf{Y}$  by  $\mathbf{Y}^*$  in Lemma 2, we conclude that there exists an  $(n, M, d, \epsilon')$  code with

$$\epsilon' \leq \mathbb{E}_{\mathbf{X}} \left[ e^{-M P_{\mathbf{Y}^*}(\mathcal{B}(\mathbf{X}, d))} \right] \quad (145)$$

$$= \mathbb{E}_{\mathbf{X}} \left[ e^{-M P_{\mathbf{Y}^*}(\mathcal{B}(\mathbf{X}, d))} \mathbb{1}[\mathcal{L}] \right] + \mathbb{E}_{\mathbf{X}} \left[ e^{-M P_{\mathbf{Y}^*}(\mathcal{B}(\mathbf{X}, d))} \mathbb{1}[\mathcal{L}^c] \right] \quad (146)$$

$$\leq \mathbb{E}_{\mathbf{X}} \left[ e^{-e^{G_n}} \right] + \frac{K}{(\log n)^{\kappa\alpha}} + \frac{1}{n} \quad (147)$$

$$= \mathbb{E}_{\mathbf{X}} \left[ e^{-e^{G_n}} \mathbb{1}[\mathcal{G}] \right] + \mathbb{E}_{\mathbf{X}} \left[ e^{-e^{G_n}} \mathbb{1}[\mathcal{G}^c] \right] + \frac{K}{(\log n)^{\kappa\alpha}} + \frac{1}{n} \quad (148)$$

$$\leq \mathbb{P}(\mathcal{G}) + \frac{1}{\sqrt{n}} \mathbb{P}(\mathcal{G}^c) + \frac{K}{(\log n)^{\kappa\alpha}} + \frac{1}{n} \quad (149)$$

$$\leq \epsilon_n + \frac{C_{\text{BE}} + 1}{\sqrt{n}} + \frac{K}{(\log n)^{\kappa\alpha}} + \frac{1}{n} \quad (150)$$

$$= \epsilon, \quad (151)$$

where (145) is by weakening (117) using  $\mathbf{Y} = \mathbf{Y}^*$ ; (147) holds by (142) and  $\mathbb{1}[\mathcal{L}] M P_{\mathbf{Y}^*}(\mathcal{B}(\mathbf{X}, d)) \geq e^{G_n}$ ; (149) holds since  $e^{-e^{G_n}} \leq 1$  and  $\mathbb{1}[\mathcal{G}^c] e^{-e^{G_n}} \leq \frac{1}{\sqrt{n}}$ ; (150) is by (139); and (151) is by the choice of  $\epsilon_n$  in (133). Consequently, since there exists an  $(n, M, d, \epsilon')$  code with  $\epsilon' \leq \epsilon$ , we must have

$$R(n, d, \epsilon) \leq \frac{\log M}{n}, \quad (152)$$



where  $\log M$  is given by (134). Similar to the converse proof, plugging (134) into (152) and then using the Taylor expansion of  $Q^{-1}(\epsilon_n)$  yields (110). ■

## VI. CONCLUSION

In this paper, we derived the reverse waterfilling characterization (57) of the dispersion for lossy compression of the Gauss-Markov source (5) with  $|a| < 1$  (Theorem 1). This is the first dispersion result for lossy compression of sources with memory. In doing so, we developed several novel technical tools, which are highlighted below.

- We derived the expression for the limiting variance of the  $d$ -tilted information for the Gauss-Markov source in Theorem 2. Its proof relies on our parametric representation for the  $d$ -tilted information, presented in Lemma 7 in Appendix C-D.
- Theorem 4 presented a nonasymptotic refinement of Gray's result [5] (restated in Theorem 3) on the eigenvalue distribution of the covariance matrix of the random vector  $\mathbf{U}$  from the Gauss-Markov source. The key tool we developed to prove Theorem 4 is Lemma 1 in Section III-C.1, which is a sharp bound relating the eigenvalues of two sequences of symmetric tridiagonal matrices.
- The maximum likelihood estimator  $\hat{a}(\mathbf{u})$ , defined in (72) and analyzed in Theorems 5 and 6, is of independent interest as it allows one to estimate the distribution of  $\mathbf{u}$  drawn from the class of the Gauss-Markov sources with unknown  $a$ . The error bounds in Theorem 5 rely on the Hanson-Wright inequality [25, Th. 1.1]. That inequality applies beyond the case when  $Z_i$ 's are Gaussian, which means that our approach can be applied to other sources with memory.
- To prove achievability, we constructed a typical set in Definition 1 based on the maximum likelihood estimator. This idea of constructing typical sets via estimators could also find its use in other problems.

Finally, we discuss several open problems.

- The dispersion for Gauss-Markov sources with  $|a| \geq 1$  is unknown. This paper treats the asymptotically stationary case, i.e.,  $|a| < 1$ . The case  $|a| \geq 1$  is fundamentally different, since that source is nonstationary. The rate-distortion functions for nonstationary Gaussian autoregressive processes were first derived by Gray [5, Eq. (22)] in 1970, and later in 1980 by Hashimoto and Arimoto [34, Eq.(6)] in an equivalent but distinct form; that equivalence was shown by Gray and Hashimoto [36] in 2008. Gray's reverse waterfilling [5, Eq. (22)] is different from Kolmogorov's reverse waterfilling (33) in the nonstationary case, where the later does not apply. Therefore, in order to characterize the dispersion for the case  $|a| \geq 1$ , one would need to use Gray's reverse waterfilling [5, Eq. (22)] for  $\mathbb{R}_U(d)$ .
- A natural generalization of this work would be to consider the dispersion for the general stationary Gaussian autoregressive processes (37). The geometric converse proof in Section IV already yields a converse bound on

$R(n, d, \epsilon)$ , which is tight in the low distortion regime  $d \in (0, d_c]$  in the first-order term; we conjecture it is also tight in the second-order term. A possible way to show a matching achievability bound for the Gaussian AR processes of order  $m$ , inspired by the estimation idea in this paper, is to analyze an estimator which estimates the vector  $\mathbf{a} = (a_1, \dots, a_m)^T$  in (37) instead of the scalar  $a$ . To deal with large distortions, i.e.  $d > d_c$ , sharp bounds on eigenvalues of  $\mathbf{A}^T \mathbf{A}$  with  $\mathbf{A}$  given by (82) need to be derived, similar to Lemma 1 in Section III-C.1; the tools in Appendix C-A might be useful.

- A formula (analogous to (56)) for the channel dispersion of the Gaussian intersymbol interference (ISI) channels, see [37, Eq. (29)], was presented in [37, Th. 5] without proof. The channel capacity of the Gaussian ISI channel is well-known, e.g. [38, Th.] and [39, Th. 1]. The tools in this paper might be useful in obtaining a proof of the channel dispersion for Gaussian ISI channels in [37, Th. 5].
- A fundamental problem left open is how widely the limiting formula for the dispersion

$$V(d) = \limsup_{n \rightarrow \infty} \frac{1}{n} \text{Var} [J_{\mathbf{X}}(\mathbf{X}, d)] \quad (153)$$

applies. Theorem 1 and Theorem 2 established its validity for the Gauss-Markov source. We conjecture that it continues to apply whenever the central limit theorem type of results can be derived for  $J_{\mathbf{X}}(\mathbf{X}, d)$ .

## APPENDIX A

### A. A Roadmap of the Paper

The relations of our main theorems, lemmas, corollaries are presented in Fig. 4.

### B. Classical Theorems

**Theorem 13** (Berry-Esseen Theorem, e.g. [40, Ch. 16.5]). *Let  $W_1, \dots, W_n$  be a collection of independent zero-mean random variables with variances  $V_i^2 > 0$  and finite third absolute moment  $T_i \triangleq \mathbb{E}[|W_i|^3] < +\infty$ . Define the average variance  $V^2$  and average third absolute moment  $T$  as*

$$V^2 \triangleq \frac{1}{n} \sum_{i=1}^n V_i^2, \quad T \triangleq \frac{1}{n} \sum_{i=1}^n T_i. \quad (154)$$

Then for  $n \in \mathbb{N}$ , we have

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P} \left[ \frac{1}{V\sqrt{n}} \sum_{i=1}^n W_i < t \right] - \Phi(t) \right| \leq \frac{6T}{V^3\sqrt{n}}, \quad (155)$$

where  $\Phi$  is the cdf of the standard normal distribution  $\mathcal{N}(0, 1)$ .

**Remark 4.** Since in this paper, we only consider random variables  $W_i$ 's with bounded  $p$ -th moment for any finite  $p$ , it is easy to check that there exists a constant  $C_{BE} > 0$  such that

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P} \left[ \frac{1}{V\sqrt{n}} \sum_{i=1}^n W_i < t \right] - \Phi(t) \right| \leq \frac{C_{BE}}{\sqrt{n}}. \quad (156)$$

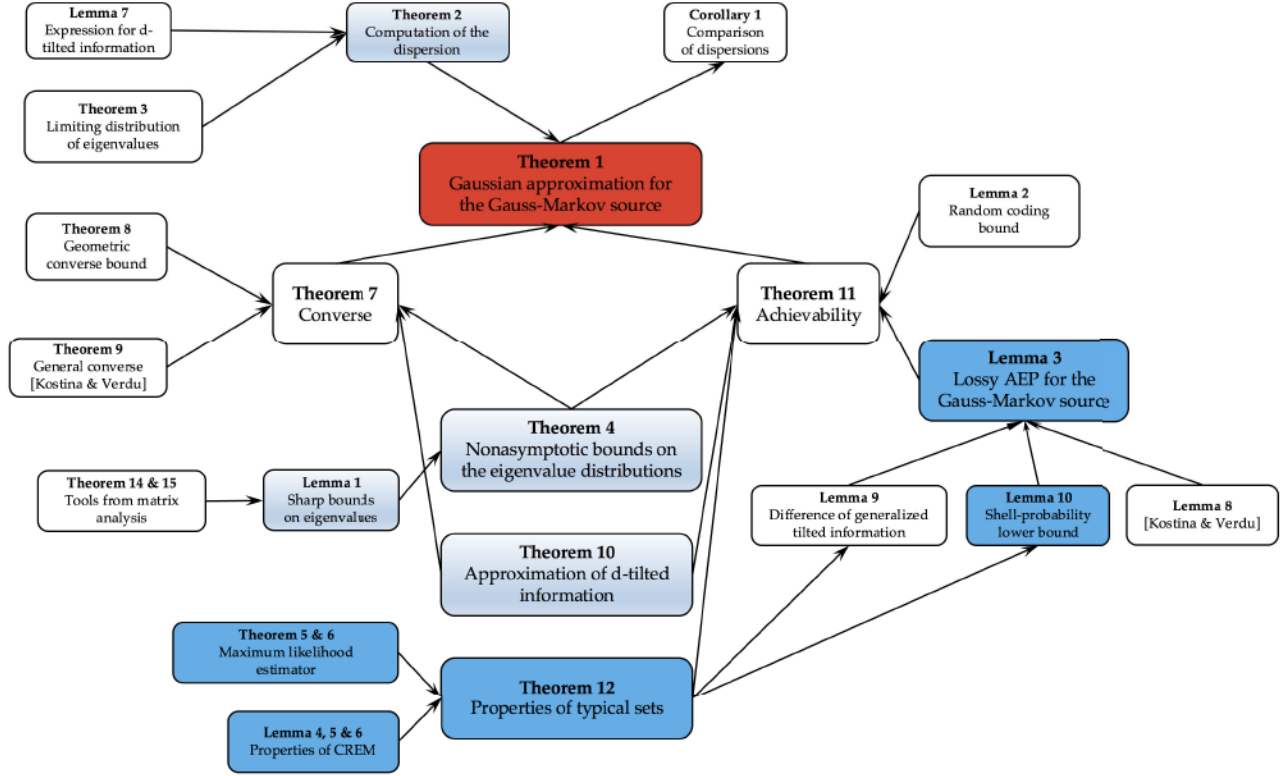


Fig. 4. A roadmap of the paper: an arrow from block A to block B means that the derivation of block B is based on block A. Our main result is Theorem 1. The lightly shaded blocks consist of results that are novel and relatively easier to obtain, while the heavily shaded blocks consist of our main technical contributions.

While the constant  $C_{BE}$  depends on the random variables  $W_i$ 's, to simplify notations, we use  $C_{BE}$  in all applications of the Berry-Esseen Theorem.

### C. Justification of (11)

We provide a short justification of how (11) follows from [27]. We use the same notations in [27]. We denote  $P$  the distribution of the source  $X$ , and  $Q_0$  the optimal reproduction distribution in  $\mathbb{R}_X(d)$ , that is,  $Q_0$  is the  $Y$ -marginal of a minimizer  $\tilde{P}_0$ . First, [27, Corollary, Eq. (1.25)] shows that

$$\mathbb{R}_X(d) = \max_{\alpha(x), s} \mathbb{E}_P [\log \alpha(X)] - sd. \quad (157)$$

Next, the proof of [27, Lemma 1.4] in [27, Eq. (1.27)-(1.32)] shows that the maximizer  $(\alpha^*(\cdot), s^*)$  is given by [27, Eq. (1.15)] with  $s^* = \lambda^*$  due to [27, Eq. (1.12)]. For convenience, we write down [27, Eq. (1.15)]:

$$\alpha^*(x) = \frac{1}{\mathbb{E}_{Q_0}[\exp(-\lambda^* d(x, Y^*))]}, \quad (158)$$

where  $Y^* \sim Q_0$ . Finally, plugging (158) into (157) yields (11). ■

## APPENDIX B PROOFS IN SECTION III

### A. Corner Points on the Dispersion Curve

We derive (61) using the residue theorem from complex analysis [41, Th. 17]. Similar ideas have been applied by

Berger [11, Ch. 6, p. 232] and Gray [36, Eq. (12)] to study the rate-distortion functions for the nonstationary Gaussian AR processes (37). The coordinate of  $P_1$  in Fig. 3 can be easily obtained as follows. The water level matched to  $d_c$  via (34) is  $\theta_{\min}$  in (40). Hence, (57) is simplified as

$$V_U(d_c) = \frac{1}{4\pi} \int_{-\pi}^{\pi} 1 \, dw = \frac{1}{2}. \quad (159)$$

To treat  $P_2$ , note that the water level matched to  $d_{\max}$  via (34) is  $\theta_{\max}$  in (41), which, due to (35), equals,

$$\theta_{\max} = \frac{\sigma^2}{(1-a)^2}. \quad (160)$$

This implies that (57) evaluates as

$$V_U(d_{\max}) = \frac{\sigma^4}{4\pi\theta_{\max}^2} \int_{-\pi}^{\pi} \frac{1}{(g(w))^2} \, dw. \quad (161)$$

Invoking the residue theorem [41, Th. 17], we will obtain the integral

$$\mathcal{I} \triangleq \int_{-\pi}^{\pi} \frac{1}{(g(w))^2} \, dw = \frac{2\pi(1+a^2)}{(1-a^2)^3}, \quad (162)$$

which will complete the derivation. To that end, change variables using  $z = e^{jw}$  and rewrite

$$g(w) = 1 + a^2 - a(z + z^{-1}) \quad (163)$$

$$= (z^{-1} - a)(z - a). \quad (164)$$



The integral  $\mathcal{I}$  is then

$$\mathcal{I} = \oint_{|z|=1} \frac{z}{ja^2(z-a^{-1})^2(z-a)^2} dz \quad (165)$$

$$= 2\pi j \operatorname{Res}_{z=a} \frac{z}{ja^2(z-a^{-1})^2(z-a)^2} \quad (166)$$

$$= \frac{2\pi}{a^2} \lim_{z \rightarrow a} \frac{d}{dz} \frac{z}{(z-a^{-1})^2} \quad (167)$$

$$= \frac{2\pi(1+a^2)}{(1-a^2)^3}, \quad (168)$$

where (165) is by the change of variable  $z = e^{jw}$ ; (166) is due to the residue theorem and  $a \in [0, 1)$ ; and (167) is the standard method of computing residues. ■

### B. Two Interpretations of the Maximum Distortion

We present the computation details of (47) and how (42) leads to (46). Using the same technique as in (165)-(168), we compute (42) as

$$d_{\max} = \frac{\sigma^2}{2\pi} \oint_{|z|=1} \frac{1}{jz(z^{-1}-a)(z-a)} dz \quad (169)$$

$$= \sigma^2 \operatorname{Res}_{z=a} \frac{1}{-a(z-a^{-1})(z-a)} \quad (170)$$

$$= \sigma^2 \lim_{z \rightarrow a} \frac{1}{-a(z-a^{-1})} \quad (171)$$

$$= \frac{\sigma^2}{1-a^2}. \quad (172)$$

To compute the stationary variance, take the variance on both sides of (5),

$$\operatorname{Var}[U_i] = a^2 \operatorname{Var}[U_{i-1}] + \sigma^2, \quad (173)$$

then taking the limit on both sides of (173), we have

$$\lim_{i \rightarrow \infty} \operatorname{Var}[U_i] = a^2 \lim_{i \rightarrow \infty} \operatorname{Var}[U_{i-1}] + \sigma^2, \quad (174)$$

which implies

$$\lim_{i \rightarrow \infty} \operatorname{Var}[U_i] = \frac{\sigma^2}{1-a^2}. \quad (175)$$

## APPENDIX C PROOFS IN SECTION II

### A. Eigenvalues of Nearly Toeplitz Tridiagonal Matrices

For convenience, we record two import results from matrix theory.

**Theorem 14** (Cauchy Interlacing Theorem for Eigenvalues [42, p.59]). *Let  $\mathbf{H}$  be an  $n \times n$  Hermitian matrix partitioned as  $\mathbf{H} = \begin{pmatrix} \mathbf{P} & \star \\ \star & \star \end{pmatrix}$ , where  $\mathbf{P}$  is an  $(n-1) \times (n-1)$  principal submatrix of  $\mathbf{H}$ . Let  $\lambda_1(\mathbf{P}) \leq \lambda_2(\mathbf{P}) \leq \dots \leq \lambda_{n-1}(\mathbf{P})$  be the eigenvalues of  $\mathbf{P}$ , and  $\lambda_1(\mathbf{H}) \leq \lambda_2(\mathbf{H}) \leq \dots \leq \lambda_n(\mathbf{H})$  be the eigenvalues of  $\mathbf{H}$ , then  $\lambda_i(\mathbf{H}) \leq \lambda_i(\mathbf{P}) \leq \lambda_{i+1}(\mathbf{H})$  for  $i = 1, \dots, n-1$ .*

**Theorem 15** (Gershgorin Circle Theorem [43, p.16, Th. 1.11]). *Let  $\mathbf{M}$  be any  $n \times n$  matrix, with entries  $m_{ij}$ .*

*Define  $r_i \triangleq \sum_{j \neq i} |m_{ij}|$ ,  $\forall i \in [n]$ , then for any eigenvalue  $\lambda$  of  $\mathbf{M}$ , there exists  $i \in [n]$  such that  $|\lambda - m_{ii}| \leq r_i$ .*

*Proof of Lemma 1.* To indicate the dimension, denote by  $\mathbf{A}_n$  the matrix  $\mathbf{A}$  defined in (48), and denote

$$\mathbf{B}_n \triangleq \mathbf{A}_n^T \mathbf{A}_n \quad (176)$$

$$= \begin{pmatrix} 1+a^2 & -a & 0 & 0 & \dots & 0 \\ -a & 1+a^2 & -a & 0 & \dots & 0 \\ 0 & -a & 1+a^2 & -a & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 0 & -a & 1+a^2 & -a \\ 0 & \dots & \dots & 0 & -a & 1 \end{pmatrix}. \quad (177)$$

Notice that we obtain a tridiagonal Toeplitz matrix  $\mathbf{W}_n$  if the  $(n, n)$ -th entry of  $\mathbf{B}_n$  is replaced by  $1+a^2$ :

$$\mathbf{W}_n = \begin{pmatrix} 1+a^2 & -a & 0 & 0 & \dots & 0 \\ -a & 1+a^2 & -a & 0 & \dots & 0 \\ 0 & -a & 1+a^2 & -a & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 0 & -a & 1+a^2 & -a \\ 0 & \dots & \dots & 0 & -a & 1+a^2 \end{pmatrix}, \quad (178)$$

whose eigenvalues  $\xi_1^{(n)} \leq \xi_2^{(n)} \leq \dots \leq \xi_n^{(n)}$  are given by (69), see [44, Eq. (4)]. At an intuitive level, we expect  $\xi_i^{(n)}$ 's to approximate  $\mu_i$ 's well since  $\mathbf{B}_n$  and  $\mathbf{W}_n$  differ in only one entry. The first part of the proof applies the Cauchy interlacing theorem (Theorem 14) to show (70) for  $2 \leq i \leq n$ . The bound (70) for  $i = 1$  is proved via the Gershgorin circle theorem (Theorem 15) in the second part.

Applying Theorem 14 by partitioning  $\mathbf{B}_n$  as

$$\mathbf{B}_n = \begin{pmatrix} \mathbf{W}_{n-1} & \star \\ \star & 1 \end{pmatrix}, \quad (179)$$

we obtain

$$\mu_i \leq \xi_i^{(n-1)} \leq \mu_{i+1}, \quad \forall i \in [n-1]. \quad (180)$$

On the other hand, since  $\mathbf{W}_n \geq \mathbf{B}_n$  in the semidefinite order, we have

$$\xi_i^{(n)} \geq \mu_i, \quad \forall i \in [n]. \quad (181)$$

Combining (180) and (181) yields

$$\xi_{i-1}^{(n-1)} \leq \mu_i \leq \xi_i^{(n)}, \quad \forall i = 2, \dots, n. \quad (182)$$

Simple algebraic manipulations using (69) and (182) lead to

$$\xi_i^{(n)} - \mu_i \leq \xi_i^{(n)} - \xi_{i-1}^{(n-1)} \leq \frac{2\pi a}{n}, \quad \forall i = 2, \dots, n. \quad (183)$$

To bound the difference  $\xi_1^{(n)} - \mu_1$ , we apply Theorem 15 to  $\mathbf{B}_n$ . Note that for  $\mathbf{B}_n$ , we have  $r_1 = r_n = a$  and  $r_i = 2a$ ,  $\forall i = 2, \dots, n-1$  (recall  $r_i$ 's defined in Theorem 15). For the eigenvalue  $\mu_1$ , there exists  $j \in [n]$  such that  $|\mu_1 - \mathbf{B}_{jj}| \leq r_j$ . The following analyses lead to  $\mu_1 \geq (1-a)^2$ :

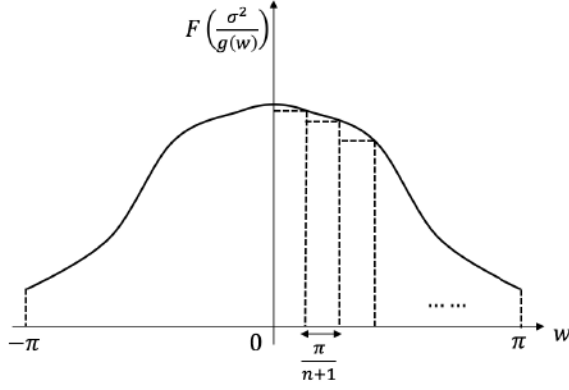


Fig. 5. Bound the integral  $I$  by Riemann sums.

- If  $2 \leq j \leq n-1$ , then  $|\mu_1 - (1+a^2)| \leq 2a$ , which implies that  $\mu_1 \geq 1+a^2-2a$ .
- If  $j=1$ , then  $|\mu_1 - (1+a^2)| \leq a$ , which implies that  $\mu_1 \geq 1+a^2-a \geq 1+a^2-2a$ .
- If  $j=n$ , then  $|\mu_1 - 1| \leq a$ , which implies  $\mu_1 \geq 1-a \geq (1-a)^2$ .

Recall from (69) that  $\xi_1^{(n)} = 1+a^2-2a \cos(\frac{\pi}{n+1})$ . Hence,

$$\xi_1^{(n)} - \mu_1 \leq 2a \left[ 1 - \cos\left(\frac{\pi}{n+1}\right) \right] \quad (184)$$

$$\leq \frac{a\pi^2}{(n+1)^2}, \quad (185)$$

$$\leq \frac{2a\pi}{n} \quad (186)$$

where (185) is by the inequality  $\cos(x) \geq 1 - x^2/2$ . ■

### B. Proof of Theorem 4

*Proof.* Since  $S(w)$  in (35) is even in  $w \in [-\pi, \pi]$ , we have

$$I \triangleq \frac{1}{2\pi} \int_{-\pi}^{\pi} F[S(w)] dw \quad (187)$$

$$= \frac{1}{\pi} \int_0^{\pi} F[S(w)] dw. \quad (188)$$

We bound the integral  $I$  by Riemann sums over intervals of width  $\frac{\pi}{n+1}$ , see Fig. 5. Since  $F[S(w)]$  is a nonincreasing function in  $w \in [0, \pi]$ , we have

$$I \geq \frac{1}{\pi} \sum_{i=1}^n F\left[S\left(\frac{i\pi}{n+1}\right)\right] \frac{\pi}{n+1}. \quad (189)$$

Using Lemma 1, we can further bound (189) from below as

$$I \geq \frac{1}{n+1} \sum_{i=1}^n F\left(\frac{\sigma^2}{\mu_i + 2a\pi/n}\right). \quad (190)$$

Since  $F$  is  $L$ -Lipschitz, we have for  $i \in [n]$ ,

$$F\left(\frac{\sigma^2}{\mu_i + 2a\pi/n}\right) \geq F\left(\frac{\sigma^2}{\mu_i}\right) - L\left(\frac{\sigma^2}{\mu_i} - \frac{\sigma^2}{\mu_i + 2a\pi/n}\right) \quad (191)$$

$$\geq F\left(\frac{\sigma^2}{\mu_i}\right) - \frac{2a\pi L\sigma^2}{n\mu_i^2}. \quad (192)$$

Plugging (192) into (190), we obtain

$$I \geq \frac{1}{n+1} \sum_{i=1}^n F\left(\frac{\sigma^2}{\mu_i}\right) - \frac{2aL\pi\sigma^2}{n(n+1)} \sum_{i=1}^n \frac{1}{\mu_i^2}. \quad (193)$$

From (71), we see that

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{\mu_i^2} \leq \frac{1}{(1-a)^4}. \quad (194)$$

Let  $\|F\|_{\infty}$  be the sup norm of  $F$  over the interval (66), then

$$I \geq \frac{1}{n+1} \sum_{i=1}^n F\left(\frac{\sigma^2}{\mu_i}\right) - \frac{2aL\pi\sigma^2}{(n+1)(1-a)^4} \quad (195)$$

$$\geq \frac{1}{n} \sum_{i=1}^n F\left(\frac{\sigma^2}{\mu_i}\right) - \frac{\|F\|_{\infty} + 2aL\pi\sigma^2/(1-a)^4}{n}. \quad (196)$$

Similarly, we can derive the upper bound

$$I \leq \frac{1}{n} \sum_{i=1}^n F\left(\frac{\sigma^2}{\mu_i}\right) + \frac{2\|F\|_{\infty}}{n}. \quad (197)$$

Therefore, setting

$$C_L \triangleq \max \left\{ \|F\|_{\infty} + \frac{2aL\pi\sigma^2}{(1-a)^4}, 2\|F\|_{\infty} \right\} \quad (198)$$

completes the proof. ■

### C. Properties of the Conditional Relative Entropy Minimization Problem

This section presents three results on the CREM problem (121), all of which are necessary to the proof of Theorem 11.

1) *Gaussian CREM:* The optimization problem (121) is referred to as the Gaussian CREM when  $\mathbf{X}$  and  $\mathbf{Y}$  are Gaussian random vectors with independent coordinates. The optimizer and optimal value of the Gaussian CREM are characterized by the following lemma.

**Lemma 4.** Let  $\mathbf{X}$  and  $\mathbf{Y}$  be Gaussian random vectors with independent coordinates, i.e.,

$$\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \Sigma_X), \text{ where } \Sigma_X = \text{diag}(\alpha_1^2, \dots, \alpha_n^2), \quad (199)$$

$$\mathbf{Y} \sim \mathcal{N}(\mathbf{0}, \Sigma_Y), \text{ where } \Sigma_Y = \text{diag}(\beta_1^2, \dots, \beta_n^2). \quad (200)$$

Then, the optimizer  $P_{\mathbf{F}^*|\mathbf{X}}$  in the Gaussian CREM (121)  $\mathbb{R}(\mathbf{X}, \mathbf{Y}, d)$  is

$$P_{\mathbf{F}^*|\mathbf{X}} = \prod_{i=1}^n P_{F_i^*|X_i}, \quad (201)$$

where for any  $\mathbf{x} \in \mathbb{R}^n$ , the conditional distribution of  $F_i^*$  given  $X_i = x_i$  is<sup>5</sup>

$$F_i^* | \{X_i = x_i\} \sim \mathcal{N}\left(\frac{2\delta^*\beta_i^2 x_i}{1 + 2\delta^*\beta_i^2}, \frac{\beta_i^2}{1 + 2\delta^*\beta_i^2}\right), \quad (202)$$

<sup>5</sup>When  $\beta_i^2 = 0$  for some  $i \in [n]$ , the random variable in (202) degenerates to a deterministic random variable taking value 0, and the notation  $\mathcal{N}(0, 0)$  denotes the Dirac delta function.



and the optimal value is

$$\begin{aligned} \mathbb{R}(\mathbf{X}, \mathbf{Y}, d) &= -\delta^* d + \frac{1}{2n} \sum_{i=1}^n \log \left( 1 + 2\delta^* \beta_i^2 \right) + \\ &\quad \frac{1}{n} \sum_{i=1}^n \frac{\delta^* \alpha_i^2}{1 + 2\delta^* \beta_i^2}, \end{aligned} \quad (203)$$

where  $\delta^*$  is the negative slope defined as

$$\delta^* = -\mathbb{R}'(\mathbf{X}, \mathbf{Y}, d). \quad (204)$$

*Proof.* We particularize (16) to the Gaussian CREM. For any fixed  $\mathbf{x} \in \mathbb{R}^n$ , rearranging (16) yields

$$f_{\mathbf{F}^*|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = f_{\mathbf{Y}}(\mathbf{y}) \exp \left\{ \Lambda_{\mathbf{Y}}(\mathbf{x}, \delta^*, d) - \delta^* n d(\mathbf{x}, \mathbf{y}) + \delta^* n d \right\} \quad (205)$$

$$\propto f_{\mathbf{Y}}(\mathbf{y}) \exp \left\{ -\delta^* n d(\mathbf{x}, \mathbf{y}) \right\} \quad (206)$$

$$\propto \exp \left\{ -\delta^* \sum_{i=1}^n (y_i - x_i)^2 - \sum_{i=1}^n \frac{y_i^2}{2\beta_i^2} \right\} \quad (207)$$

$$= \prod_{i=1}^n \exp \left\{ -\frac{\left( y_i - \frac{2\delta^* \beta_i^2 x_i}{1+2\delta^* \beta_i^2} \right)^2}{\frac{2\beta_i^2}{1+2\delta^* \beta_i^2}} \right\}, \quad (208)$$

where  $p_1 \propto p_2$  means that  $p_1 = c' p_2$  for a positive constant  $c'$ ; (206) is by keeping only terms containing  $\mathbf{y}$  (since  $\mathbf{x}$  is fixed); (207) is by plugging the pdf of  $\mathbf{Y}$  into (206); and (208) is by completing the squares in  $y_i$ . Hence, (201) and (202) follow. Next, the expression (203) is obtained by a direct computation using (16), (201) and (202).

$$\begin{aligned} \mathbb{R}(\mathbf{X}, \mathbf{Y}, d) &= \frac{1}{n} \int_{\mathbb{R}^n} f_{\mathbf{X}}(\mathbf{x}) \int_{\mathbb{R}^n} f_{\mathbf{F}^*|\mathbf{X}}(\mathbf{y}|\mathbf{x}) \left[ \Lambda_{\mathbf{Y}}(\mathbf{x}, \delta^*, d) - \right. \\ &\quad \left. \delta^* n d(\mathbf{x}, \mathbf{y}) + \delta^* n d \right] dy d\mathbf{x} \end{aligned} \quad (209)$$

$$\begin{aligned} &= \frac{1}{n} \int_{\mathbb{R}^n} f_{\mathbf{X}}(\mathbf{x}) \int_{\mathbb{R}^n} f_{\mathbf{F}^*|\mathbf{X}}(\mathbf{y}|\mathbf{x}) \Lambda_{\mathbf{Y}}(\mathbf{x}, \delta^*, d) dy d\mathbf{x} \quad (210) \\ &= -\delta^* d + \frac{1}{2n} \sum_{i=1}^n \log \left( 1 + 2\delta^* \beta_i^2 \right) + \frac{1}{n} \sum_{i=1}^n \frac{\delta^* \alpha_i^2}{1 + 2\delta^* \beta_i^2}, \end{aligned} \quad (211)$$

where (209) follows by substituting (16) into (121); (210) holds since  $\mathbb{E}[d(\mathbf{X}, \mathbf{F}^*)] = d$  by the optimality of  $\mathbf{F}^*$ ; and (211) is by direct integration of (210), which relies on the definition of the generalized tilted information (15) and the well-known formula for the moment generating function (MGF) of a noncentral  $\chi_1^2$ -distribution. ■

2) *Gauss-Markov CREM:* The optimization problem (121) is referred to as the Gauss-Markov CREM if  $\mathbf{X}$  is the decorrelation of  $\mathbf{U}$  in (50), and  $(\mathbf{X}, \mathbf{Y}) = (\mathbf{X}, \mathbf{Y}^*)$  forms a RDF-achieving pair in  $\mathbb{R}_{\mathbf{X}}(n, d)$ . Recall from (53) that  $\mathbf{X} \sim \mathcal{N}(0, \Sigma_{\mathbf{X}})$ , where

$$\Sigma_{\mathbf{X}} = \text{diag}(\sigma_1^2, \dots, \sigma_n^2), \quad (212)$$

and  $\sigma_i^2$ 's are given by (53). Recall from (65) that  $\mathbf{Y}^* \sim \mathcal{N}(0, \Sigma_{\mathbf{Y}^*})$ , where

$$\Sigma_{\mathbf{Y}^*} = \text{diag}(v_1^2, \dots, v_n^2), \quad (213)$$

and we denote  $v_i^2$  as

$$v_i^2 \triangleq \max \left( 0, \sigma_i^2 - \theta_n \right). \quad (214)$$

And  $\theta_n > 0$  is the water level matched to  $d$  via the  $n$ -th order reverse waterfilling (31). From (14), we have

$$\mathbb{R}(\mathbf{X}, \mathbf{Y}^*, d) = \mathbb{R}_{\mathbf{U}}(n, d), \quad (215)$$

and  $\mathbb{R}_{\mathbf{U}}(n, d)$  is given by (30). Lemma 4 is also applicable to the special case of the Gauss-Markov CREM. Furthermore, the next lemma characterizes the negative slope in the Gauss-Markov CREM.

**Lemma 5.** *In the Gauss-Markov CREM, for any  $d \in (0, d_{\max})$  and  $n \in \mathbb{N}$ , let  $\theta_n > 0$  be the water level matched to  $d$  via the  $n$ -th order reverse waterfilling (31), the negative slope  $\lambda^*$  defined in (23) satisfies*

$$\lambda^* = \frac{1}{2\theta_n}. \quad (216)$$

*Proof.* We directly compute the negative slope using the parametric representation (30) and (31). Taking the derivative with respect to  $d$  on both sides of (30) yields

$$\lambda^* = \frac{1}{n} \sum_{i=1}^n \frac{1}{2\theta_n} \frac{d\theta_n}{dd} \mathbb{1} \left\{ \sigma_i^2 > \theta_n \right\}. \quad (217)$$

Differentiating (31), we obtain

$$\frac{dd}{d\theta_n} = \frac{1}{n} \sum_{i=1}^n \mathbb{1} \left\{ \sigma_i^2 > \theta_n \right\}, \quad (218)$$

which is independent of  $i$ . Plugging (218) into (217) yields (216).

To justify the formal differentiation in (218), observe using (31) that  $d$  is a continuous piecewise linear function of  $\theta_n$ , and  $d$  is differentiable with respect to  $\theta_n$  except at the  $n$  points:  $\theta_n = \sigma_i^2$ ,  $i \in [n]$ . The above proof goes through as long as the derivatives at those  $n$  points are understood as the left derivatives. Indeed,  $\mathbb{R}_{\mathbf{U}}(n, d)$  is differentiable w.r.t.  $d$  for any  $d \in (0, d_{\max})$ , e.g. [15, Eq. (16)]. ■

3) *Sensitivity of the Negative Slope:* The following theorem is a perturbation result, which bounds the change in the negative slope when the variances of the input  $\mathbf{X}$  to  $\mathbb{R}(\mathbf{X}, \mathbf{Y}^*, d)$  are perturbed. It is related to lossy compression using mismatched codebook: the codewords are drawn randomly according to the distribution  $P_{\mathbf{Y}^*}$  while the source distribution is  $\hat{\mathbf{X}}$  instead of  $\mathbf{X}$ .

**Lemma 6.** *Let  $\mathbf{X}$  be the decorrelation of  $\mathbf{U}$  in (50), and let  $(\mathbf{X}, \mathbf{Y}^*)$  be a RDF-achieving pair in  $\mathbb{R}_{\mathbf{X}}(n, d)$  (recall (19)). For any fixed distortion  $d \in (0, d_{\max})$ , let  $\theta > 0$  be the water level matched to  $d$  via the limiting reverse waterfilling in (34). For any  $t \in (0, \theta/3)$ , let  $\hat{\sigma}_i^2$ 's be such that*

$$|\hat{\sigma}_i^2 - \sigma_i^2| \leq t, \quad \forall i \in [n]. \quad (219)$$

Let the Gaussian random vector  $\hat{\mathbf{X}}$  be  $\hat{\mathbf{X}} \sim \mathcal{N}(\mathbf{0}, \text{diag}(\hat{\sigma}_1^2, \dots, \hat{\sigma}_n^2))$ , and let  $\hat{\lambda}^*$  be the negative slope of  $\mathbb{R}(\hat{\mathbf{X}}, \mathbf{Y}^*, d)$ . Then, for all  $n$  large enough, the negative slope  $\hat{\lambda}^*$  satisfies

$$|\lambda^* - \hat{\lambda}^*| \leq \frac{9t}{4\theta^2}, \quad (220)$$

where  $\lambda^* = -\mathbb{R}'(\mathbf{X}, \mathbf{Y}^*, d)$  is given by (216).

*Proof.* Consider the Gaussian CREM  $\mathbb{R}(\hat{\mathbf{X}}, \mathbf{Y}^*, d)$ . Let  $\hat{\theta}_n > 0$  be the water level matched to  $d$  via the  $n$ -th order reverse waterfilling (31) over  $\hat{\sigma}_i^2$ 's, and let  $\theta_n > 0$  be the water level matched to  $d$  via the  $n$ -th order reverse waterfilling (31) over  $\sigma_i^2$ 's. In (203), replacing  $(\mathbf{X}, \mathbf{Y})$  by  $(\hat{\mathbf{X}}, \mathbf{Y}^*)$ , and then taking the derivative with respect to  $d$  on both sides yields

$$\begin{aligned} -\hat{\lambda}^* &= -\hat{\lambda}^* + \frac{1}{2n} \sum_{i: \hat{\sigma}_i^2 > \hat{\theta}_n} \frac{-2\hat{\lambda}^*}{1 + 2\hat{\lambda}^* v_i^2} \frac{d\hat{\theta}_n}{dd} + \\ &\quad \frac{1}{n} \sum_{i: \hat{\sigma}_i^2 > \hat{\theta}_n} \frac{2\hat{\sigma}_i^2 \hat{\lambda}^{*2}}{(1 + 2\hat{\lambda}^* v_i^2)^2} \frac{d\hat{\theta}_n}{dd}, \end{aligned} \quad (221)$$

where  $v_i^2$ 's are defined in (214). Rearranging terms yields

$$\hat{\lambda}^* = \sum_{i: \hat{\sigma}_i^2 > \hat{\theta}_n} \frac{1}{(1 + 2\hat{\lambda}^* v_i^2)^2} \bigg/ \sum_{i: \hat{\sigma}_i^2 > \hat{\theta}_n} \frac{2(\hat{\sigma}_i^2 - \sigma_i^2 + \theta_n)}{(1 + 2\hat{\lambda}^* v_i^2)^2}. \quad (222)$$

Substituting the bound (219) into (222), we obtain

$$\hat{\lambda}^* \in \left[ \frac{1}{2(\theta_n + t)}, \frac{1}{2(\theta_n - t)} \right]. \quad (223)$$

Since  $\lim_{n \rightarrow \infty} \theta_n = \theta$ , for all  $n$  large enough, we have

$$\frac{2\theta}{3} \leq \theta_n \leq \frac{4\theta}{3}. \quad (224)$$

Since  $t \in (0, \theta/3)$ , (224) implies that  $0 < t < \theta_n/2$ . From (216), (223) and (224), we see that

$$|\lambda^* - \hat{\lambda}^*| \leq \max \left\{ \left| \frac{1}{2(\theta_n + t)} - \frac{1}{2\theta_n} \right|, \left| \frac{1}{2(\theta_n - t)} - \frac{1}{2\theta_n} \right| \right\} \quad (225)$$

$$\leq \frac{t}{\theta_n^2} \quad (226)$$

$$\leq \frac{9t}{4\theta^2}. \quad (227)$$

#### D. Proof of Theorem 2

Theorem 2 is a direct consequence of the following lemma.

**Lemma 7** (Parametric Representation for the d-Tilted Information). *Let  $\mathbf{X}$  be the decorrelation of  $\mathbf{U}$  (50), and let  $(\mathbf{X}, \mathbf{Y}^*)$  be a RDF-achieving pair in  $\mathbb{R}_{\mathbf{X}}(n, d)$ . For any  $d \in (0, d_{\max})$ , let  $\theta_n > 0$  be the water level matched to  $d$  via the  $n$ -th*

*order reverse waterfilling (31) over  $\sigma_i^2$ ,  $i \in [n]$ . Then, for all  $\mathbf{x} \in \mathbb{R}^n$ ,*

$$\begin{aligned} \Lambda_{Y_i^*}(x_i, \lambda^*, \min(\theta_n, \sigma_i^2)) \\ = \frac{\min(\theta_n, \sigma_i^2)}{2\theta_n} \left( \frac{x_i^2}{\sigma_i^2} - 1 \right) + \frac{1}{2} \log \frac{\max(\theta_n, \sigma_i^2)}{\theta_n}, \end{aligned} \quad (228)$$

where  $\lambda^*$  defined in (23) is given by (216).

*Proof.* The proof relies on the Gaussianity of  $\mathbf{Y}^*$ . For each  $i \in [n]$ , from (15) and (216), we have

$$\begin{aligned} \Lambda_{Y_i^*}(x_i, \lambda^*, \min(\theta_n, \sigma_i^2)) \\ = - \frac{\min(\theta_n, \sigma_i^2)}{2\theta_n} - \log \mathbb{E} \left[ \exp \left\{ -\lambda^* (Y_i^* - x_i)^2 \right\} \right]. \end{aligned} \quad (229)$$

Substituting  $Y_i^* \equiv 0$  a.s. for all  $i$  such that  $\sigma_i^2 \leq \theta_n$  (recall (213)) into (229), we obtain

$$\Lambda_{Y_i^*}(x_i, \lambda^*, \min(\theta_n, \sigma_i^2)) = \frac{x_i^2 - \sigma_i^2}{2\theta_n}. \quad (230)$$

Substituting  $Y_i^* \sim \mathcal{N}(0, \sigma_i^2 - \theta_n)$  for all  $i$  such that  $\sigma_i^2 > \theta_n$  (recall (213)) into (229) and applying the formula for the moment generating function of a noncentral  $\chi^2$ -distribution with one degree of freedom, we obtain

$$\Lambda_{Y_i^*}(x_i, \lambda^*, \min(\theta_n, \sigma_i^2)) = \frac{1}{2} \left( \frac{x_i^2}{\sigma_i^2} - 1 \right) + \frac{1}{2} \log \frac{\sigma_i^2}{\theta_n}. \quad (231)$$

Unifying (230) and (231), we obtain (228). ■

*Proof of Theorem 2.* For any fixed distortion  $d \in (0, d_{\max})$ , let  $\theta > 0$  be the water level matched to  $d$  via the limiting reverse waterfilling (34). By the independence of  $Y_1^*, \dots, Y_n^*$  and (31), we have for any  $\mathbf{x}$ ,

$$J_{\mathbf{X}}(\mathbf{x}, d) = \sum_{i=1}^n \Lambda_{Y_i^*}(x_i, \lambda^*, \min(\theta_n, \sigma_i^2)), \quad (232)$$

where  $\lambda^* = -\mathbb{R}'_{\mathbf{X}}(n, d)$ . Taking the expectation and the variance of (232) using (228) yields<sup>6</sup>

$$\mathbb{E}[J_{\mathbf{X}}(\mathbf{X}, d)] = \sum_{i=1}^n \frac{1}{2} \max \left( 0, \log \frac{\sigma_i^2}{\theta_n} \right), \quad (233)$$

$$\text{Var}[J_{\mathbf{X}}(\mathbf{X}, d)] = \sum_{i=1}^n \frac{1}{2} \min \left( 1, \left( \frac{\sigma_i^2}{\theta_n} \right)^2 \right). \quad (234)$$

An application of Theorem 3 to (233) on the function  $t \mapsto \frac{1}{2} \max(0, \log \frac{t}{\theta})$  yields (21). Similarly, an application of Theorem 3 to (234) on the function  $t \mapsto \frac{1}{2} \min \left[ 1, \left( \frac{t}{\theta} \right)^2 \right]$  yields (64). ■

<sup>6</sup>The result on expectations was implicitly established by Gray [5], which we recover here. The result on variances is new.



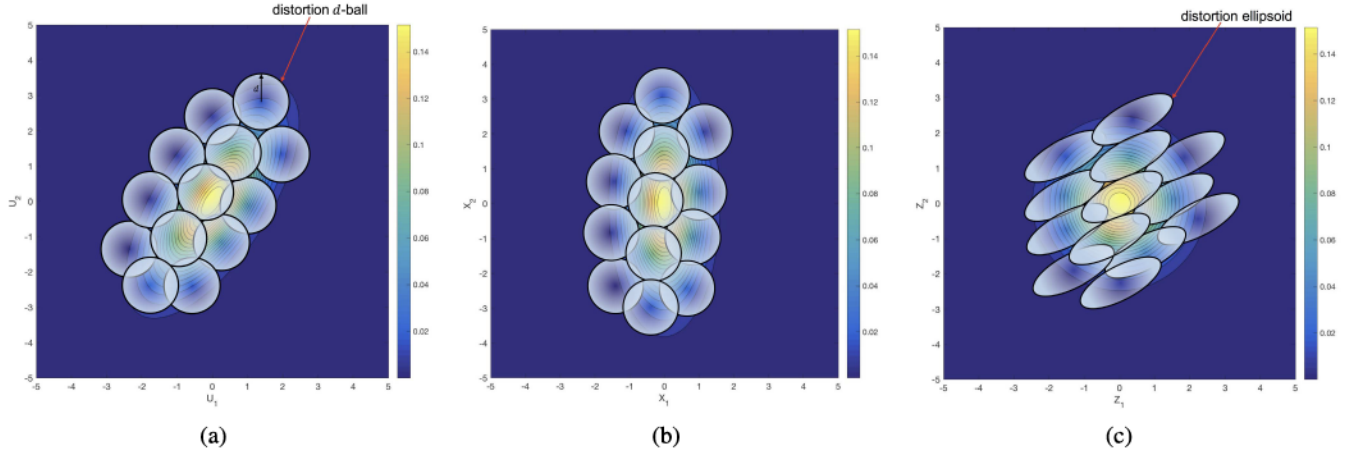


Fig. 6. Converse proof in figures. The contour plot in each figure shows the underlying probability distribution. (a) The  $\mathbf{U}$  space: Given  $d, \epsilon$ , the goal is to cover at least  $1 - \epsilon$  probability mass, under distribution of  $\mathbf{U}$ , using the least number of distortion  $d$ -balls. (b) The  $\mathbf{X}$  space is simply a unitary transformation of the  $\mathbf{U}$  space. (c) The  $\mathbf{Z}$  space: Given  $d, \epsilon$ , the goal is to cover at least  $1 - \epsilon$  probability mass, under distribution of  $\mathbf{Z}$ , using the least number of distortion ellipsoids, each of which has the same volume as the distortion  $d$ -ball since  $\det(\mathbf{A}\mathbf{A}^T) = 1$ .

#### APPENDIX D PROOFS IN SECTION IV

##### A. Proof of Theorem 8

*Proof.* The result follows from a geometric argument, illustrated in Fig. 6. Let  $\mathcal{C} \subset \mathbb{R}^n$  be the set of codewords of an arbitrary  $(n, M, d, \epsilon)$  code, and  $\mathcal{B}(\mathbf{c}, d)$  be the distortion  $d$ -ball centered at a codeword  $\mathbf{c} \in \mathcal{C}$  (recall (4)). By the definition of an  $(n, M, d, \epsilon)$  code, we know that the union of the distortion  $d$ -balls centered at codewords in  $\mathcal{C}$  has probability mass at least  $1 - \epsilon$ :

$$\mathbb{P}[\mathbf{U} \in \mathcal{B}] \geq 1 - \epsilon, \quad (235)$$

where  $\mathcal{B}$  denotes the union of the distortion  $d$ -balls centered at the codewords in  $\mathcal{C}$ :

$$\mathcal{B} \triangleq \bigcup_{\mathbf{c} \in \mathcal{C}} \mathcal{B}(\mathbf{c}, d). \quad (236)$$

For a set  $\mathcal{S} \subseteq \mathbb{R}^n$ , denote by

$$\mathbf{A}\mathcal{S} \triangleq \{\mathbf{A}\mathbf{s} : \mathbf{s} \in \mathcal{S}\} \quad (237)$$

the linear transformation of  $\mathcal{S}$  by the matrix  $\mathbf{A}$ . Recall from (48) that  $\mathbf{A}$  is invertible and the innovation is  $\mathbf{Z} = \mathbf{A}\mathbf{U}$ . Changing variable  $\mathbf{U} = \mathbf{A}^{-1}\mathbf{Z}$  in (235) yields

$$\mathbb{P}[\mathbf{Z} \in \mathbf{A}\mathcal{B}] \geq 1 - \epsilon. \quad (238)$$

Next, we give a geometric interpretation of the set  $\mathbf{A}\mathcal{B}$ . Consider the set  $\mathbf{A}\mathcal{C}$ , that is, the transformation of the codebook  $\mathcal{C}$  by  $\mathbf{A}$ . For any  $\mathbf{x} \in \mathbb{R}^n$ , notice that the set

$$\mathbf{A}\mathcal{B}(\mathbf{A}^{-1}\mathbf{x}, d) = \left\{ \mathbf{x}' \in \mathbb{R}^n : (\mathbf{x}' - \mathbf{x})^T (\mathbf{A}\mathbf{A}^T)^{-1} (\mathbf{x}' - \mathbf{x}) \leq nd \right\} \quad (239)$$

is the set of points bounded by the ellipsoid centered at  $\mathbf{x}$  with principal axes being the eigenvectors of  $\mathbf{A}\mathbf{A}^T$ . It follows that

$$\mathbf{A}\mathcal{B} = \mathbf{A} \bigcup_{\mathbf{c} \in \mathcal{C}} \mathcal{B}(\mathbf{c}, d) \quad (240)$$

$$= \bigcup_{\mathbf{c}' \in \mathbf{A}\mathcal{C}} \mathbf{A}\mathcal{B}(\mathbf{A}^{-1}\mathbf{c}', d), \quad (241)$$

i.e.,  $\mathbf{A}\mathcal{B}$  is the union of ellipsoids centered at transformed codewords. See Fig. 6c for an illustration of the set  $\mathbf{A}\mathcal{B}$ .

Finally, the following volumetric argument completes the proof of Theorem 8. Since the volume of a union of sets is less than or equal to the sum of the sets' volumes, we have

$$M \geq \frac{\text{Vol}(\mathbf{A}\mathcal{B})}{\text{Vol}(\mathbf{A}\mathcal{B}(\mathbf{0}, d))}. \quad (242)$$

Moreover,  $\text{Vol}(\mathbf{A}\mathcal{B}(\mathbf{0}, d)) = \text{Vol}(\mathcal{B}(\mathbf{0}, d))$  due to  $\det \mathbf{A} = 1$ . On the other hand, due to the spherical symmetry of the distribution of  $\mathbf{Z}$ , the ball  $\mathcal{B}(\mathbf{0}, r(n, \epsilon))$ , where  $r(n, \epsilon)$  satisfies (79), has the smallest volume among all sets in  $\mathbb{R}^n$  with probability greater than or equal to  $1 - \epsilon$ , and so

$$\text{Vol}(\mathbf{A}\mathcal{B}) \geq \text{Vol}(\mathcal{B}(\mathbf{0}, r(n, \epsilon))). \quad (243)$$

Therefore, we can weaken (242) as

$$M \geq \frac{\text{Vol}(\mathcal{B}(\mathbf{0}, r(n, \epsilon)))}{\text{Vol}(\mathcal{B}(\mathbf{0}, d))} = \left( \frac{r(n, \epsilon)}{d} \right)^{n/2}. \quad (244)$$

■

##### B. Proof of Theorem 10

*Proof.* The proof is based on Chebyshev's inequality. Fix  $d \in (0, d_{\max})$ . For each fixed  $n \in \mathbb{N}$ , let  $\theta_1, \theta_2 > 0$  be the water levels matched to  $d$  and  $d_n$ , respectively, in the  $n$ -th order reverse waterfilling (31) over  $\sigma_i^2$ ,  $i \in [n]$ , that is,

$$d = \frac{1}{n} \sum_{i=1}^n \min(\theta_1, \sigma_i^2), \quad (245)$$

$$d_n = \frac{1}{n} \sum_{i=1}^n \min(\theta_2, \sigma_i^2). \quad (246)$$

Obviously, both  $\theta_1$  and  $\theta_2$  depend on  $n$ . We now proceed to show that there exists a constant  $h_2 > 0$  such that for all  $n$  large enough,

$$|\theta_1 - \theta_2| \leq \frac{h_2}{n}. \quad (247)$$

Indeed, without loss of generality, assume  $d < d_n$ ,<sup>7</sup> then  $\theta_1 < \theta_2$  by the monotonicity of the reverse waterfilling (31). Define the following index sets

$$I_1 \triangleq \{i \in [n] : \sigma_i^2 \leq \theta_1\}, \quad (248)$$

$$I_2 \triangleq \{i \in [n] : \theta_1 < \sigma_i^2 < \theta_2\}, \quad (249)$$

$$I_3 \triangleq \{i \in [n] : \theta_2 \leq \sigma_i^2\}. \quad (250)$$

Then,

$$d_n - d = \frac{1}{n} \sum_{i=1}^n \left( \min(\theta_2, \sigma_i^2) - \min(\theta_1, \sigma_i^2) \right) \quad (251)$$

$$= \frac{1}{n} \sum_{i \in I_1} 0 + \frac{1}{n} \sum_{i \in I_2} (\sigma_i^2 - \theta_1) + \frac{1}{n} \sum_{i \in I_3} (\theta_2 - \theta_1) \quad (252)$$

$$\geq \frac{|I_3|}{n} (\theta_2 - \theta_1). \quad (253)$$

Since  $d_n < d_{\max}$ , there exists a constant  $\tilde{c} \in (0, 1)$  such that for all  $n$  large enough,  $|I_3| \geq \tilde{c}n$ , hence (247) holds with  $h_2 = \frac{h_1}{\tilde{c}}$ .

Now, let  $G_1, \dots, G_n$  be i.i.d.  $\mathcal{N}(0, 1)$ . To simplify notations, we denote the random variable as

$$\Delta(d, d_n) \triangleq J_{\mathbf{X}}(\mathbf{X}, d) - J_{\mathbf{X}}(\mathbf{X}, d_n). \quad (254)$$

From (232) and (228), we have

$$\begin{aligned} \Delta(d, d_n) &= \sum_{i=1}^n \left\{ \left[ \frac{\min(\theta_1, \sigma_i^2)}{2\theta_1} - \frac{\min(\theta_2, \sigma_i^2)}{2\theta_2} \right] (G_i^2 - 1) \right. \\ &\quad \left. + \frac{1}{2} \log \left[ \frac{\max(\theta_1, \sigma_i^2)}{2\theta_1} \cdot \frac{2\theta_2}{\max(\theta_2, \sigma_i^2)} \right] \right\}. \end{aligned} \quad (255)$$

To apply Chebyshev's inequality, we bound the mean and the variance of  $\Delta(d, d_n)$  as follows.

$$\begin{aligned} \mathbb{E}[\Delta(d, d_n)] &= \sum_{i=1}^n \frac{1}{2} \log \left[ \frac{\max(\theta_1, \sigma_i^2)}{2\theta_1} \cdot \frac{2\theta_2}{\max(\theta_2, \sigma_i^2)} \right] \end{aligned} \quad (256)$$

$$= \sum_{i \in I_1} 0 + \sum_{i \in I_2} \frac{1}{2} \log \frac{\sigma_i^2}{\theta_1} + \sum_{i \in I_3} \frac{1}{2} \log \frac{\theta_2}{\theta_1} \quad (257)$$

$$\leq \frac{h_2}{\theta_1}, \quad (258)$$

where (258) holds since for  $i \in I_2$ , we have  $\frac{1}{2} \log \frac{\sigma_i^2}{\theta_1} \leq \frac{1}{2} \log \frac{\theta_2}{\theta_1}$ , while for  $i \in I_3$ , due to (247), we have

$$\frac{1}{2} \log \frac{\theta_2}{\theta_1} \leq \frac{1}{2} \log \left( 1 + \frac{h_2}{n\theta_1} \right) \leq \frac{h_2}{2n\theta_1}. \quad (259)$$

By a similar argument, we can bound the variance as

$$\text{Var}[\Delta(d, d_n)] \leq \frac{h_2^2}{\theta_1^2 n}. \quad (260)$$

<sup>7</sup>Otherwise, switch  $\theta_1$  and  $\theta_2$  in the rest of the proof.

In conjunction with (258), (260), Chebyshev's inequality yields that for all  $n$  large enough and  $\forall \ell > 0$ ,

$$\mathbb{P}[|\Delta(d, d_n) - \mathbb{E}[\Delta(d, d_n)]| \geq \ell] \leq \frac{h_2^2}{\theta_1^2 n \ell^2}. \quad (261)$$

Choosing  $\ell = \frac{uh_2}{\theta_1}$  in (261) and applying (258) yields that  $\forall u > 0$ ,

$$\mathbb{P}\left[|\Delta(d, d_n)| \geq \frac{(1+u)h_2}{\theta_1}\right] \leq \frac{1}{nu^2}. \quad (262)$$

Let  $\theta > 0$  be the water level matched to  $d$  via the limiting reverse waterfilling (34), then  $\lim_{n \rightarrow \infty} \theta_1 = \theta$  by (245) and (34). Therefore, we have  $\theta_1 \geq \frac{\theta}{2}$  for all  $n$  large enough. Hence, for all  $n$  large enough and  $\forall u > 0$ , we have

$$\mathbb{P}\left[|\Delta(d, d_n)| \geq \frac{2(1+u)h_2}{\theta}\right] \leq \frac{1}{nu^2}. \quad (263)$$

Rearranging terms in (263) completes the proof. ■

## APPENDIX E PROOFS IN SECTION V

### A. Proof of Lemma 3

In addition to new concentration inequalities, shown in Lemma 9 and Lemma 10 below, the proof leverages the following bound, which is a direct application of [15, Lem. 1] to the random vector  $\mathbf{X}$ .

**Lemma 8** (Lower Bound on Probability of Distortion Balls). *Fix  $d \in (0, d_{\max})$ ,  $n \in \mathbb{N}$ , and the distribution  $P_{\mathbf{Y}}$  on  $\mathbb{R}^n$ . Then for any  $\mathbf{x} \in \mathbb{R}^n$ , it holds that*

$$\begin{aligned} P_{\mathbf{Y}}(B(\mathbf{x}, d)) &\geq \sup_{P_{\hat{\mathbf{X}}}, \gamma > 0} \exp \left\{ -\hat{\lambda}^* n \gamma - \Lambda_{\mathbf{Y}}(\mathbf{x}, \hat{\lambda}^*, d) \right\} \times \\ &\quad \mathbb{P} \left[ d - \gamma \leq d(\mathbf{x}, \hat{\mathbf{F}}^*) \leq d \mid \hat{\mathbf{X}} = \mathbf{x} \right], \end{aligned} \quad (264)$$

where the supremum is over all pdfs  $P_{\hat{\mathbf{X}}}$  on  $\mathbb{R}^n$ ;  $\Lambda_{\mathbf{Y}}(\mathbf{x}, \hat{\lambda}^*, d)$  is the generalized tilted information defined in (15) with

$$\hat{\lambda}^* = -\mathbb{R}'(\hat{\mathbf{X}}, \mathbf{Y}, d); \quad (265)$$

and the random variable  $\hat{\mathbf{F}}^*$  achieves  $\mathbb{R}(\hat{\mathbf{X}}, \mathbf{Y}, d)$ .

The high-level idea in proving Lemma 3 is the following. In Lemma 8, we replace  $\mathbf{Y}$  by  $\mathbf{Y}^*$  defined in (65) and (213), and we choose  $\hat{\mathbf{X}}$  to be the proxy Gaussian random variable  $\hat{\mathbf{X}}(\mathbf{x})$  defined in (120). With such choices of  $\hat{\mathbf{X}}$  and  $\mathbf{Y}$ , the next two lemmas provide further lower bounds on the two factors on the right side of (264). The first one is a concentration inequality on the generalized tilted information.

**Lemma 9.** *For any fixed  $d \in (0, d_{\max})$  and excess-distortion probability  $\epsilon \in (0, 1)$ , there exist constants  $C$  and  $C_2 > 0$  such that for all  $n$  large enough,*

$$\mathbb{P} \left[ \Lambda_{\mathbf{Y}^*}(\mathbf{X}, \hat{\lambda}^*(\mathbf{X}), d) \leq \Lambda_{\mathbf{Y}^*}(\mathbf{X}, \lambda^*, d) + C \log n \right] \geq 1 - \frac{C_2}{\sqrt{n}}, \quad (266)$$

where  $\hat{\lambda}^*(\mathbf{x})$  is given by (130) with  $\hat{\mathbf{X}}$  defined in (120), and  $\lambda^*$  is in (216).

*Proof.* Appendix E-C. ■



The second bound, presented in Lemma 10 below, is referred to as the shell-probability lower bound. For any  $\mathbf{x} \in \mathbb{R}^n$  and any  $\gamma \in (0, d)$ , define the shell

$$S(\mathbf{x}, d, \gamma) \triangleq \{\mathbf{x}' \in \mathbb{R}^n : d - \gamma \leq d(\mathbf{x}, \mathbf{x}') \leq d\}. \quad (267)$$

Geometrically, Lemma 8 provides a quantitative connection between the probability of a distortion  $d$ -ball and the probability of its shell, and Lemma 10 below gives a lower bound on the probability of the shell  $S(\mathbf{x}, d, \gamma)$  for “typical” sequences  $\mathbf{x}$ .

**Lemma 10** (Shell-Probability Lower Bound). *Fix any distortion  $d \in (0, d_{\max})$  and any excess-distortion probability  $\epsilon \in (0, 1)$ . For any constant  $\alpha > 0$  and any  $n \in \mathbb{N}$ , consider the set  $T(n, \alpha, p)$  defined in Definition 1, where  $p$  is the constant in Theorem 12. Let*

$$\gamma \triangleq \frac{\log^q n}{n}, \quad (268)$$

where  $q > 1$  is a constant defined in (346) in Appendix E-D below. Then, there exists a constant  $C_1 > 0$  such that for all  $n$  large enough, for any  $\mathbf{u} \in T(n, \alpha, p)$  and  $\mathbf{x} = \mathbf{S}^\top \mathbf{u}$  with  $\mathbf{S}$  in (52), it holds that

$$\mathbb{P}[\hat{\mathbf{F}}^* \in S(\mathbf{x}, d, \gamma) \mid \hat{\mathbf{X}} = \mathbf{x}] \geq \frac{C_1}{\sqrt{n}}, \quad (269)$$

where  $\hat{\mathbf{X}}$  is given in (120).

*Proof.* Appendix E-D. ■

We now present the proof of Lemma 3.

*Proof of Lemma 3.* Let  $\mathbf{X}$  be the decorrelation of  $\mathbf{U}$  in (50). Replace  $\mathbf{Y}$  by  $\mathbf{Y}^*$  in Lemma 8. Let  $T(n, \alpha, p)$  be the set defined in Definition 1, and let  $p$  be the constant in Theorem 12. Let  $C, C_1, C_2, q$  be the constants in Lemmas 9 and 10. Consider any  $n$  that is large enough such that Theorem 12, Lemma 9 and Lemma 10 hold. Let  $\theta > 0$  be the water level matched to  $d$  via the limiting reverse waterfilling (34). Denote the event

$$\mathcal{E} \triangleq \left\{ \log \frac{1}{P_{\mathbf{Y}^*}(\mathcal{B}(\mathbf{X}, d))} > J_{\mathbf{X}}(\mathbf{X}, d) + \beta_1 \log^q n + \beta_2 \right\}, \quad (270)$$

where  $\beta_1$  and  $\beta_2$  are constants defined by

$$\beta_1 \triangleq \frac{1}{2\theta} + C_d + \frac{1}{2} + C, \quad (271)$$

$$\beta_2 \triangleq -\log C_1, \quad (272)$$

and  $C_d > 0$  is a constant such that

$$\left| \hat{\lambda}^*(\mathbf{x}) - \frac{1}{2\theta} \right| \leq C_d \quad (273)$$

for any  $\mathbf{u} \in T(n, \alpha, p)$  and  $\mathbf{x} = \mathbf{S}^\top \mathbf{u}$ . The existence of such  $C_d$  is guaranteed by (129) in Theorem 12 and the fact that  $\lim_{n \rightarrow \infty} \theta_n = \theta$ .

Using elementary probability rules, we write

$$\begin{aligned} & \mathbb{P} \left[ \log \frac{1}{P_{\mathbf{Y}^*}(\mathcal{B}(\mathbf{X}, d))} > J_{\mathbf{X}}(\mathbf{X}, d) + \beta_1 \log^q n + \beta_2 \right] \\ &= \mathbb{P}[\mathcal{E}, \mathbf{U} \in T(n, \alpha, p)] + \mathbb{P}[\mathcal{E}, \mathbf{U} \notin T(n, \alpha, p)] \end{aligned} \quad (274)$$

$$\begin{aligned} & \leq \mathbb{P}[\mathbf{U} \in T(n, \alpha, p), \hat{\lambda}^*(\mathbf{X})n\gamma + \Lambda_{\mathbf{Y}^*}(\mathbf{X}, \hat{\lambda}^*(\mathbf{X}), d) - \\ & \quad \log \mathbb{P}[\hat{\mathbf{F}}^* \in S(\hat{\mathbf{X}}, d, \gamma) \mid \hat{\mathbf{X}} = \mathbf{X}] > \\ & \quad J_{\mathbf{X}}(\mathbf{X}, d) + \beta_1 \log^q n + \beta_2] \\ & \quad + \mathbb{P}[\mathbf{U} \notin T(n, \alpha, p)] \end{aligned} \quad (275)$$

$$\begin{aligned} & \leq \mathbb{P}[\mathbf{U} \in T(n, \alpha, p), \Lambda_{\mathbf{Y}^*}(\mathbf{X}, \hat{\lambda}^*(\mathbf{X}), d) - J_{\mathbf{X}}(\mathbf{X}, d) > \\ & \quad - \hat{\lambda}^*(\mathbf{X})n\gamma + \log \frac{C_1}{\sqrt{n}} + \beta_1 \log^q n + \beta_2] \\ & \quad + \mathbb{P}[\mathbf{U} \notin T(n, \alpha, p)] \end{aligned} \quad (276)$$

$$\begin{aligned} & \leq \mathbb{P}[\mathbf{U} \in T(n, \alpha, p), \Lambda_{\mathbf{Y}^*}(\mathbf{X}, \hat{\lambda}^*(\mathbf{X}), d) - J_{\mathbf{X}}(\mathbf{X}, d) > \\ & \quad C \log n] + \mathbb{P}[\mathbf{U} \notin T(n, \alpha, p)] \end{aligned} \quad (277)$$

$$\leq \frac{K}{(\log n)^{\kappa\alpha}}, \quad (278)$$

where (275) is by Lemma 8; (276) is by (269); (277) is by the choice of  $\gamma$  in (268) and  $q > 1$ ; (278) is by Lemma 9 and (127); and  $K > 0$  is a constant. ■

## B. Proof of Theorem 12

*Proof.* We first prove the property (1). First, Theorem 6 states that for all  $n$  large enough the condition (123) is violated with probability at most  $\frac{2}{(\log n)^{\kappa\alpha}}$ . Second, we bound the probability of violating condition (125). Note that since  $X_i \sim \mathcal{N}(0, \sigma_i^2)$  by (53), we have  $G_i \triangleq \frac{X_i}{\sigma_i} \sim \mathcal{N}(0, 1)$  for all  $i \in [n]$ . For each  $k = 1, 2, 3$ , applying the Berry-Esseen theorem<sup>8</sup> to the zero-mean random variables  $G_i^{2k} - (2k - 1)!!$ , we obtain

$$\mathbb{P} \left[ \left| \frac{1}{n} \sum_{i=1}^n G_i^{2k} - (2k - 1)!! \right| > 2 \right] \leq 2Q \left( \frac{2\sqrt{n}}{r_k} \right) + \frac{12T_k}{r_k^3 \sqrt{n}}, \quad (279)$$

where  $r_k^2$  and  $T_k$  are the variance and the third absolute moment of  $G_i^{2k} - (2k - 1)!!$ , respectively;  $r_k$  and  $T_k$  are both positive constants since  $G_i$ 's have bounded finite-order moments. Therefore, there exists a constant  $A'_1 > 0$  such that for all  $n$  large enough,

$$\mathbb{P} \left[ \left| \frac{1}{n} \sum_{i=1}^n G_i^{2k} - (2k - 1)!! \right| > 2 \right] \leq \frac{A'_1}{\sqrt{n}}. \quad (280)$$

The bound (280) implies that the condition (125) is violated with probability at most  $\frac{3A'_1}{\sqrt{n}}$  by the union bound.

<sup>8</sup>The Berry-Esseen theorem suffices here, though tighter bounds are possible via other concentration inequalities, say Chernoff's bound.

Verifying that the condition (124) is satisfied with high probability is more involved. The high-level procedure is the following. The expressions for  $m_i(\mathbf{x})$ 's in (122) can be directly obtained from (202) in Lemma 4 in Appendix C-C.1. Then we approximate  $m_i(\mathbf{x})$  using carefully crafted  $\bar{m}_i(x_i)$ , for which it is easier to obtain a concentration bound of the form (124). At the end, the approximation gaps between  $m_i(\mathbf{x})$  and  $\bar{m}_i(x_i)$  are shown to be sufficiently small, and (124) ensues.

We now present the details. We start with a closer look at the optimizer  $\hat{\mathbf{F}}^*$  in  $\mathbb{R}(\hat{\mathbf{X}}, \mathbf{Y}^*, d)$ . Recall the distributions of  $\mathbf{X}$  and  $\mathbf{Y}^*$  in (212) and (213), and the distribution of  $\hat{\mathbf{X}}$  in (120). An application of Lemma 4 in Appendix C-C.1 to  $\mathbb{R}(\hat{\mathbf{X}}, \mathbf{Y}^*, d)$  yields that for all  $\mathbf{x} \in \mathbb{R}^n$ ,

$$P_{\hat{\mathbf{F}}^*}[\hat{\mathbf{X}}=\mathbf{x}] = \prod_{i=1}^n P_{\hat{F}_i^*}[\hat{X}_i=x_i], \quad (281)$$

$$\hat{F}_i^* \mid \hat{X}_i = x_i \sim \mathcal{N}\left(\frac{2\hat{\lambda}^*(\mathbf{x})v_i^2x_i}{1+2\hat{\lambda}^*(\mathbf{x})v_i^2}, \frac{v_i^2}{1+2\hat{\lambda}^*(\mathbf{x})v_i^2}\right), \quad (282)$$

where  $\hat{\lambda}^*(\mathbf{x})$  is given by (130) and  $v_i^2$ 's are defined in (214). Then, from (282) and the definition of  $m_i(\mathbf{x})$  in (122), it is straightforward to obtain the expression

$$m_i(\mathbf{x}) = \frac{v_i^2}{1+2\hat{\lambda}^*(\mathbf{x})v_i^2} + \frac{x_i^2}{(1+2\hat{\lambda}^*(\mathbf{x})v_i^2)^2}. \quad (283)$$

The quantity  $m_i(\mathbf{x})$  in the form of (283) is hard to analyze since there is no simple formula for  $\hat{\lambda}^*(\mathbf{x})$ . We instead consider  $\bar{m}_i(x_i)$ 's, defined as

$$\bar{m}_i(x_i) \triangleq \frac{v_i^2}{1+2\lambda^*v_i^2} + \frac{x_i^2}{(1+2\lambda^*v_i^2)^2}, \quad (284)$$

which is obtained from (283) by replacing  $\hat{\lambda}^*(\mathbf{x})$  with  $\lambda^*$ . The random variable  $\bar{m}_i(X_i)$  is much easier to analyze, since  $\lambda^* = \frac{1}{2\theta_n}$  by Lemma 5 in Appendix C-C.2, with which (284) is simplified as

$$\bar{m}_i(x_i) = \frac{(\min(\sigma_i^2, \theta_n))^2}{\sigma_i^2} \left( \frac{x_i^2}{\sigma_i^2} - 1 \right) + \min(\sigma_i^2, \theta_n). \quad (285)$$

We will control the difference between  $m_i(\mathbf{x})$  and  $\bar{m}_i(x_i)$  by bounding  $|\lambda^* - \hat{\lambda}^*(\mathbf{x})|$ . Indeed, a lengthy but elementary calculation, deferred to the end of the proof, shows that there exists a constant  $A_1'' > 0$  (depending only on  $d$ ) such that for all  $n$  large enough,  $\forall \mathbf{x} \in \mathbb{R}^n$  satisfying (123) and (125), we have

$$\left| \frac{1}{n} \sum_{i=1}^n \bar{m}_i(x_i) - \frac{1}{n} \sum_{i=1}^n m_i(\mathbf{x}) \right| \leq A_1'' \eta_n. \quad (286)$$

With (286), we proceed to explain how to apply the Berry-Esseen theorem to obtain the following bound: there exists a constant  $A_1''' > 0$  such that for all  $n$  large enough and  $\forall \omega > 0$ ,

$$\mathbb{P} \left[ \left| \frac{1}{n} \sum_{i=1}^n \bar{m}_i(X_i) - d \right| \geq \omega \sqrt{\frac{\log \log n}{n}} \right] \leq \frac{A_1'''}{(\log n)^{\frac{\omega^2}{2\beta^2}}}, \quad (287)$$

where

$$\beta^2 \triangleq \frac{1}{n} \sum_{i=1}^n \text{Var}[\bar{m}_i(X_i)]. \quad (288)$$

To that end, first note from (285) and (31) that

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E}[\bar{m}_i(X_i)] = d, \quad (289)$$

then an application of the Berry-Esseen theorem to  $\bar{m}_i(X_i) - \min(\sigma_i^2, \theta_n)$  yields

$$\begin{aligned} \mathbb{P} \left[ \left| \frac{1}{n} \sum_{i=1}^n \bar{m}_i(X_i) - d \right| \geq \omega \sqrt{\frac{\log \log n}{n}} \right] \\ \leq 2Q \left( \frac{\omega \sqrt{\log \log n}}{\beta} \right) + \frac{12T}{\beta^3 \sqrt{n}} \end{aligned} \quad (290)$$

$$\leq \frac{2}{(\log n)^{\frac{\omega^2}{2\beta^2}}} + \frac{12T}{\beta^3 \sqrt{n}}, \quad (291)$$

where  $T \triangleq \frac{1}{n} \sum_{i=1}^n \mathbb{E}[|\bar{m}_i(X_i) - \min(\sigma_i^2, \theta_n)|^3]$  is bounded. Using (285), it is easy to check that there exists a constant  $\beta_d > 0$  (depending only on  $d$ ) such that  $0 < \beta_d < \beta \leq \frac{\sqrt{2}\sigma^2}{(1-a)^2}$ . Therefore, (287) follows from (291). Now, we combine (286) and (287) to conclude that the condition (124) is satisfied with high probability. Define the set  $\mathcal{L} \subset \mathbb{R}^n$  as

$$\mathcal{L} \triangleq \{\mathbf{u} \in \mathbb{R}^n : \mathbf{u} \text{ satisfies (123) and (125)}\}. \quad (292)$$

Then, by Theorem 6, (280) and the union bound, we have

$$\mathbb{P}[\mathbf{U} \in \mathcal{L}^c] \leq \frac{2}{(\log n)^{\kappa a}} + \frac{3A_1'}{\sqrt{n}}. \quad (293)$$

Hence, we have

$$\begin{aligned} \mathbb{P} \left[ \left| \frac{1}{n} \sum_{i=1}^n m_i(\mathbf{X}) - d \right| \geq p \eta_n \right] \\ \leq \mathbb{P} \left[ \left| \frac{1}{n} \sum_{i=1}^n \bar{m}_i(X_i) - \frac{1}{n} \sum_{i=1}^n m_i(\mathbf{X}) \right| + \left| \frac{1}{n} \sum_{i=1}^n \bar{m}_i(X_i) - d \right| \geq p \eta_n \right] \end{aligned} \quad (294)$$

$$= \mathbb{P}[\cdot, \mathbf{U} \in \mathcal{L}] + \mathbb{P}[\cdot, \mathbf{U} \in \mathcal{L}^c] \quad (295)$$

$$\leq \mathbb{P} \left[ \left| \frac{1}{n} \sum_{i=1}^n \bar{m}_i(X_i) - d \right| \geq (p - A_1'') \eta_n \right] + \mathbb{P}[\mathbf{U} \in \mathcal{L}^c] \quad (296)$$

$$\leq \frac{A_1'''}{(\log n)^{\frac{(p-A_1'')^2 a}{2\beta^2}}} + \frac{2}{(\log n)^{\kappa a}} + \frac{3A_1'}{\sqrt{n}}, \quad (297)$$

where (294) is due to the triangle inequality; (296) holds by (286); (297) follows from (287) for  $p > A_1''$ . Hence, for any  $p$  such that

$$p \geq A_1'' + \frac{2\sqrt{\kappa}\sigma^2}{(1-a)^2}, \quad (298)$$



we conclude from (297) that there exists a constant  $\tilde{A}_1 > 0$  such that for all  $n$  large enough,

$$\mathbb{P} \left[ \left| \frac{1}{n} \sum_{i=1}^n m_i(\mathbf{X}) - d \right| \geq p\eta_n \right] \leq \frac{\tilde{A}_1}{(\log n)^{\kappa\alpha}}. \quad (299)$$

Therefore, Theorem 6, (280) and (299) altogether imply the property (1) in Theorem 12.

Next, we show property (2) in Theorem 12. By the triangle inequality, we have  $\forall \mathbf{u} \in \mathbb{R}^n$  and  $\forall i \in [n]$ ,

$$\left| \hat{\sigma}_i^2 - \sigma_i^2 \right| \leq \left| \hat{\sigma}_i^2 - \frac{\sigma^2}{\xi_i} \right| + \left| \frac{\sigma^2}{\xi_i} - \sigma_i^2 \right|, \quad (300)$$

where  $\xi_i$  is given in (69). We bound the two terms in (300) separately. From (53), (70) and (71), we have

$$\left| \frac{\sigma^2}{\xi_i} - \sigma_i^2 \right| \leq \frac{2a\pi\sigma^2}{(1-a)^4n}. \quad (301)$$

To simplify notations, let  $\phi_i \triangleq \frac{i\pi}{n+1}$  and denote by  $\varphi(t)$  the function

$$\varphi(t) \triangleq \frac{\sigma^2}{1+t^2-2t\cos\phi_i}. \quad (302)$$

It is easy to see that the derivatives  $\phi'(a)$  and  $\phi''(a)$  are bounded for any fixed  $a \in [0, 1)$ . By the Taylor expansion and the triangle inequality, we have

$$\left| \hat{\sigma}_i^2 - \frac{\sigma^2}{\xi_i} \right| = |\varphi(\hat{\mathbf{u}}) - \varphi(a)| \quad (303)$$

$$\leq |\varphi'(a)| |\hat{\mathbf{u}} - a| + o(|\hat{\mathbf{u}} - a|). \quad (304)$$

Hence, combining (300), (301) and (304), we conclude that there exists a constant  $A_2 > 0$  such that for all  $n$  large enough (128) holds for any  $\mathbf{u} \in \mathcal{T}(n, \alpha, p)$ .

Finally, the bound (129) follows immediately from a direct application of Lemma 6 to (128).

*Calculations to show (286):* From (283) and (284), we have

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \bar{m}_i(x_i) - \frac{1}{n} \sum_{i=1}^n m_i(\mathbf{x}) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{2v_i^4 (\hat{\lambda}^*(\mathbf{x}) - \lambda^*)}{(1 + 2\hat{\lambda}^*(\mathbf{x})v_i^2)(1 + 2\lambda^*v_i^2)} + \\ & \quad \frac{1}{n} \sum_{i=1}^n \frac{2x_i^2v_i^2 (2 + 2\hat{\lambda}^*(\mathbf{x})v_i^2 + 2\lambda^*v_i^2) (\hat{\lambda}^*(\mathbf{x}) - \lambda^*)}{(1 + 2\hat{\lambda}^*(\mathbf{x})v_i^2)^2 (1 + 2\lambda^*v_i^2)^2}. \end{aligned} \quad (306)$$

By (129), for all  $n$  large enough,  $\forall \mathbf{u} \in \mathcal{T}(n, \alpha, p)$  and  $\mathbf{x} = \mathbf{S}^\top \mathbf{u}$ , we have

$$\left| (1 + 2\hat{\lambda}^*(\mathbf{x})v_i^2) - (1 + 2\lambda^*v_i^2) \right| \leq \frac{1 + 2\lambda^*v_i^2}{2}. \quad (307)$$

Using (214) and (216), we deduce that  $1 \leq 1 + 2\lambda^*v_i^2 \leq \frac{\sigma_i^2}{\theta_n}$ . Therefore, (307) implies that

$$\frac{1}{2} \leq 1 + 2\hat{\lambda}^*(\mathbf{x})v_i^2 \leq \frac{3\sigma_i^2}{2\theta_n}. \quad (308)$$

We continue to bound (306) as

$$\left| \frac{1}{n} \sum_{i=1}^n \bar{m}_i(x_i) - \frac{1}{n} \sum_{i=1}^n m_i(\mathbf{x}) \right| \quad (309)$$

$$\leq \frac{1}{n} \sum_{i=1}^n 4v_i^4 |\hat{\lambda}^*(\mathbf{x}) - \lambda^*| + \frac{1}{n} \sum_{i=1}^n \frac{20x_i^2v_i^2\sigma_i^2}{\theta_n} |\hat{\lambda}^*(\mathbf{x}) - \lambda^*| \quad (310)$$

$$\leq \left[ \frac{1}{n} \sum_{i=1}^n 4v_i^4 + \frac{1}{n} \cdot \frac{20\sigma^6}{\theta_n(1-a)^6} \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} \right] |\hat{\lambda}^*(\mathbf{x}) - \lambda^*| \quad (311)$$

$$\leq A_1'' \eta_n, \quad (312)$$

where (310) is by plugging (308) into (306); (311) is by  $v_i^2 \leq \sigma_i^2$  and

$$\sigma_i^6 = \frac{\sigma^6}{\mu_i^3} \leq \frac{\sigma^6}{(1-a)^6}, \quad \forall i \in [n], \quad (313)$$

which is due to (71); and (312) holds for some constant  $A_1'' > 0$  (depending on  $d$  only) by (125), (129) and (224). ■

### C. Proof of Lemma 9

*Proof.* We sketch the proof, which is similar to [15, Lem. 5] except for some slight changes. Since  $Y_1^*, \dots, Y_n^*$  are independent and the distortion measure  $\mathbf{d}(\cdot, \cdot)$  is separable, we use the definition (15), the distribution formula (213) for  $\mathbf{Y}^*$ , and the formula for the moment generating function of a noncentral  $\chi_1^2$ -distributed random variable to obtain for  $\delta > 0$  and  $\mathbf{x} \in \mathbb{R}^n$ ,

$$\Lambda_{\mathbf{Y}^*}(\mathbf{x}, \delta, d) = -n\delta d + \sum_{i=1}^n \frac{\delta x_i^2}{1 + 2\delta v_i^2} + \sum_{i=1}^n \frac{1}{2} \log(1 + 2\delta v_i^2), \quad (314)$$

where  $v_i^2$ 's are in (214). Let  $\hat{\lambda}^*(\mathbf{x})$  be defined in (130). Similar to [15, Eq. (315)-(320)], by the Taylor expansion of  $\Lambda_{\mathbf{Y}^*}(\mathbf{x}, \delta, d)$  in  $\delta$  at the point  $\delta = \lambda^*$ , we have for any  $\mathbf{x} \in \mathbb{R}^n$ ,

$$\Lambda_{\mathbf{Y}^*}(\mathbf{x}, \hat{\lambda}^*(\mathbf{x}), d) - \Lambda_{\mathbf{Y}^*}(\mathbf{x}, \lambda^*, d) \leq \frac{(S'(\mathbf{x}))^2}{2S''(\mathbf{x})}, \quad (315)$$

where we denoted

$$S'(\mathbf{x}) \triangleq \Lambda'_{\mathbf{Y}^*}(\mathbf{x}, \lambda^*, d) = \sum_{i=1}^n \frac{[\min(\theta_n, \sigma_i^2)]^2}{\sigma_i^2} \left( \frac{x_i^2}{\sigma_i^2} - 1 \right), \quad (316)$$

and

$$S''(\mathbf{x}) \triangleq -\Lambda''_{\mathbf{Y}^*}(\mathbf{x}, \lambda^* + \zeta(\mathbf{x}), d), \quad (317)$$

where (316) is by first taking derivatives of (314) with respect to  $\delta$  and then plugging  $\lambda^* = \frac{1}{2\theta_n}$ ;  $\theta_n > 0$  is the water level matched to  $d$  via the  $n$ -th order reverse waterfilling (31) over  $\sigma_i^2$ 's; and

$$\zeta(\mathbf{x}) \triangleq \rho(\hat{\lambda}^*(\mathbf{x}) - \lambda^*), \quad (318)$$

for some  $\rho \in [0, 1]$ . Note from the definition (121) that for any  $\mathbf{x}$ ,  $\mathbb{R}(\hat{\mathbf{X}}(\mathbf{x}), \mathbf{Y}^*, d)$  is a nonincreasing function in  $d$ ,

hence  $\hat{\lambda}^*(\mathbf{x}) \geq 0$ , which, combined with direct computations using (314), implies  $S''(\mathbf{x}) \geq 0$ , see (323) below. This fact was used in deriving the inequality in (315).

Next we show concentration bounds for the two random variables  $S'(\mathbf{X})$  and  $S''(\mathbf{X})$ . From (316), we see that  $S'(\mathbf{X})$  is a sum of  $n$  zero-mean and bounded-variance random variables. Then, by the Berry-Esseen theorem, we have

$$\mathbb{P}\left[|S'(\mathbf{X})| \geq \sqrt{V'n \log n}\right] \leq \frac{K'_1}{\sqrt{n}}, \quad (319)$$

where  $K'_1 > 0$  is a constant and  $V'$  is a constant such that

$$V' \geq \frac{1}{n} \sum_{i=1}^n \frac{2[\min(\theta_n, \sigma_i^2)]^4}{\sigma_i^4}. \quad (320)$$

To treat  $S''(\mathbf{X})$ , consider the following event  $\mathcal{E}$ :

$$\mathcal{E} \triangleq \{\mathbf{x} \in \mathbb{R}^n : |\hat{\lambda}^*(\mathbf{x}) - \lambda^*| \leq n^{-1/4}\}. \quad (321)$$

Using Lemma 6 in Appendix C-C.3, (300)-(304) and Theorem 5 in Section III-C.2, one can show (similar to the proof of Theorem 6) that there exists a constant  $c' > 0$  such that for all  $n$  large enough,

$$\mathbb{P}[\mathcal{E}^c] \leq \exp\{-c'\sqrt{n}\}. \quad (322)$$

Computing the derivatives using (314) yields

$$S''(\mathbf{x}) = \sum_{i=1}^n 4x_i^2 v_i^2 \left[1 + 2v_i^2(\lambda^* + \xi(\mathbf{x}))\right]^{-3} + 2v_i^4 \left[1 + 2v_i^2(\lambda^* + \xi(\mathbf{x}))\right]^{-2}. \quad (323)$$

By conditioning on  $\mathcal{E}$  and  $\mathcal{E}^c$ , we see that for any  $t > 0$ ,

$$\mathbb{P}[S''(\mathbf{X}) \leq nt] = \mathbb{P}[S''(\mathbf{X}) \leq nt, \mathcal{E}] + \mathbb{P}[S''(\mathbf{X}) \leq nt, \mathcal{E}^c] \quad (324)$$

$$\leq \mathbb{P}[S''(\mathbf{X}) \leq nt, \mathcal{E}] + \mathbb{P}[\mathcal{E}^c]. \quad (325)$$

Using (323) and the simple bound  $\frac{1}{x+y} \geq \frac{\exp(-y/x)}{x}$  for any  $x, y > 0$ , we have for any  $\mathbf{x} \in \mathcal{E}$ ,

$$S''(\mathbf{x}) \geq \exp(-3n^{-1/4}/\lambda^*) \times \sum_{i=1}^n \left[4x_i^2 v_i^2 (1 + 2v_i^2 \lambda^*)^{-3} + 2v_i^4 (1 + 2v_i^2 \lambda^*)^{-2}\right] \quad (326)$$

$$= \sum_{i=1}^n A_i x_i^2 + B_i, \quad (327)$$

where  $A_i, B_i \geq 0$  are defined as

$$A_i \triangleq \exp(-3n^{-1/4}/\lambda^*) 4v_i^2 (1 + 2v_i^2 \lambda^*)^{-3}, \quad (328)$$

$$B_i \triangleq \exp(-3n^{-1/4}/\lambda^*) 2v_i^4 (1 + 2v_i^2 \lambda^*)^{-2}. \quad (329)$$

Let  $V_n$  be defined as

$$V_n \triangleq \frac{1}{n} \sum_{i=1}^n 2A_i \sigma_i^4, \quad (330)$$

and choose the constant  $t$  in (325) such that for all  $n$  large enough,

$$0 < t < \frac{1}{n} \sum_{i=1}^n (A_i \sigma_i^2 + B_i) - \sqrt{\frac{V_n \log n}{n}}. \quad (331)$$

By the Berry-Esseen theorem, there exists a constant  $K''_1 > 0$  such that

$$\mathbb{P}[S''(\mathbf{X}) \leq nt, \mathcal{E}] \leq \mathbb{P}\left[\sum_{i=1}^n A_i X_i^2 + B_i \leq nt\right] \leq \frac{K''_1}{\sqrt{n}}. \quad (332)$$

The existence of such a constant  $t$  satisfying (331) is guaranteed since for any  $d < d_{\max}$  and for all  $n$  sufficiently large, there is a constant fraction of strictly positive  $A_i$ 's and  $B_i$ 's, which can be verified by plugging (214) and (216) into (328) and (329). Combining (325) and (332) implies that there exists a constant  $K'''_1 > 0$  such that for all  $n$  large enough

$$\mathbb{P}[S''(\mathbf{X}) \leq nt] \leq \frac{K'''_1}{\sqrt{n}}, \quad (333)$$

where the constant  $t$  satisfies (331). Finally, similar to [15, Eq. (339)], combining (315), (319) and (333) yields (266). ■

#### D. Proof of Lemma 10

*Proof.* Due to Lemma 4 in Appendix C-C.1, for any  $\mathbf{x} \in \mathbb{R}^n$ , we can write the random variable involved in (269) as a sum of independent random variables, to which the Berry-Esseen theorem is applied. The details follow. From (281) and (282), we have

$$d(\mathbf{x}, \hat{\mathbf{F}}^*) = \frac{1}{n} \sum_{i=1}^n M_i^2(\mathbf{x}), \quad (334)$$

where the random variables  $M_i(\mathbf{x})$ 's are defined as

$$M_i(\mathbf{x}) \triangleq (\hat{F}_i^* - x_i). \quad (335)$$

From (282), we know that the conditional distribution satisfies

$$M_i(\mathbf{x}) \mid \hat{\mathbf{X}} = \mathbf{x} \sim \mathcal{N}\left(\frac{-x_i}{1 + 2\hat{\lambda}^*(\mathbf{x})v_i^2}, \frac{v_i^2}{1 + 2\hat{\lambda}^*(\mathbf{x})v_i^2}\right), \quad (336)$$

where  $\hat{\lambda}^*(\mathbf{x})$  is in (130). Hence, conditioned on  $\hat{\mathbf{X}} = \mathbf{x}$ , the random variable  $d(\mathbf{x}, \hat{\mathbf{F}}^*)$  follows the noncentral  $\chi^2$ -distribution with (at most)  $n$  degrees of freedom. Applying the Berry-Esseen theorem to (334) yields that  $\forall \gamma > 0$ ,

$$\begin{aligned} & \mathbb{P}\left[d - \gamma < d(\mathbf{x}, \hat{\mathbf{F}}^*) \leq d \mid \hat{\mathbf{X}} = \mathbf{x}\right] \\ &= \mathbb{P}\left[\frac{nd - n\gamma - \sum_{i=1}^n m_i(\mathbf{x})}{s\sqrt{n}} < \frac{1}{s\sqrt{n}} \sum_{i=1}^n (M_i^2(\mathbf{x}) - m_i(\mathbf{x})) \right. \\ & \quad \left. \leq \frac{nd - \sum_{i=1}^n m_i(\mathbf{x})}{s\sqrt{n}} \mid \hat{\mathbf{X}} = \mathbf{x}\right] \quad (337) \end{aligned}$$

$$\begin{aligned} & \geq \Phi\left(\frac{nd - \sum_{i=1}^n m_i(\mathbf{x})}{s\sqrt{n}}\right) - \Phi\left(\frac{nd - n\gamma - \sum_{i=1}^n m_i(\mathbf{x})}{s\sqrt{n}}\right) \\ & \quad - \frac{12t}{s^3\sqrt{n}}, \quad (338) \end{aligned}$$



where  $m_i(\mathbf{x})$ , defined in (122), is the expectation of  $M_i^2(\mathbf{x})$  conditioned on  $\hat{\mathbf{X}} = \mathbf{x}$ ; and

$$s^2 \triangleq \frac{1}{n} \sum_{i=1}^n \text{Var} \left[ M_i^2(\mathbf{x}) \mid \hat{\mathbf{X}} = \mathbf{x} \right], \quad (339)$$

$$t \triangleq \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[ \left| M_i^2(\mathbf{x}) - m_i(\mathbf{x}) \right|^3 \mid \hat{\mathbf{X}} = \mathbf{x} \right]. \quad (340)$$

By the mean value theorem, (338) equals

$$\frac{n\gamma}{s\sqrt{n}\sqrt{2\pi}} e^{-\frac{b^2}{2}} - \frac{12t}{s^3\sqrt{n}} \quad (341)$$

for some  $b$  satisfying

$$\frac{nd - \sum_{i=1}^n m_i(\mathbf{x}) - n\gamma}{s\sqrt{n}} \leq b \leq \frac{nd - \sum_{i=1}^n m_i(\mathbf{x})}{s\sqrt{n}}. \quad (342)$$

To further lower-bound (341), we bound  $b^2$  as follows.

$$b^2 \leq 2 \left( \frac{nd - \sum_{i=1}^n m_i(\mathbf{x})}{s\sqrt{n}} \right)^2 + 2 \left( \frac{n\gamma}{s\sqrt{n}} \right)^2 \quad (343)$$

$$\leq \frac{2p^2\alpha \log \log n}{s^2} + \frac{2\log^{2q} n}{s^2 n} \quad (344)$$

$$\leq \frac{4p^2\alpha \log \log n}{c_s^2}, \quad (345)$$

where (343) is by (342) and the elementary inequality  $(x + y)^2 \leq 2(x^2 + y^2)$ ; (344) is by (124) and the choice of  $\gamma$  in (268). The constant  $q$  in (268) is chosen to be

$$q \triangleq \frac{2p^2\alpha}{c_s^2} + 1. \quad (346)$$

The constant  $c_s > 0$  is a lower bound of  $s$ , whose existence is justified below at the end of the proof. Finally, (345) holds for all sufficiently large  $n$ . Using (345), we can further lower-bound (341) as

$$\left( \frac{\log^q n}{s\sqrt{2\pi}} (\log n)^{-\frac{2p^2\alpha}{c_s^2}} - \frac{12t}{s^3} \right) \frac{1}{\sqrt{n}} \quad (347)$$

$$= \left( \frac{\log n}{s\sqrt{2\pi}} - \frac{12t}{s^3} \right) \frac{1}{\sqrt{n}} \quad (348)$$

$$\geq \frac{C_1}{\sqrt{n}}, \quad (349)$$

where (347) is by plugging (345) into (341); (348) is by using (346); and (349) holds for all sufficiently large  $n$  and some constant  $C_1 > 0$ . Therefore, (269) follows.

Finally, to justify that  $s$  and  $t$ , defined in (339) and (340), are bounded as we assumed in obtaining (345) and (349), we compute using (336)

$$\begin{aligned} & \text{Var} \left[ M_i^2(\mathbf{x}) \mid \hat{\mathbf{X}} = \mathbf{x} \right] \\ &= \frac{4x_i^2 v_i^2}{(1 + 2\hat{\lambda}^*(\mathbf{x})v_i^2)^3} + \frac{2v_i^4}{(1 + 2\hat{\lambda}^*(\mathbf{x})v_i^2)^2}. \end{aligned} \quad (350)$$

Then, using (308) to bound  $1 + 2\hat{\lambda}^*(\mathbf{x})v_i^2$  and (125) to bound  $x_i^2$ , we can lower- and upper-bound  $s$  by positive constants;  $t$  can be bounded similarly. ■

## APPENDIX F

### A. Derivation of the Maximum Likelihood Estimator

This section presents the details in obtaining (72). The random vector  $(U_1, U_2 - aU_1, \dots, U_n - aU_{n-1})^\top$  is distributed according to  $\mathcal{N}(0, \sigma^2 \mathbf{I})$ . Let  $p_a(\cdot)$  be the probability density function of  $\mathbf{U}$  with parameter  $a$ , then

$$\begin{aligned} \hat{a}(\mathbf{u}) \\ & \triangleq \arg \max_a p_a(\mathbf{u}) \end{aligned} \quad (351)$$

$$= \arg \max_a \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{1}{2\sigma^2}(u_i - au_{i-1})^2} \quad (352)$$

$$= \arg \min_a \sum_{i=2}^n (u_i - au_{i-1})^2 + u_1^2 \quad (353)$$

$$= \arg \min_a \sum_{i=2}^n u_{i-1}^2 a^2 - 2 \sum_{i=2}^n u_{i-1} u_i a + \sum_{i=1}^n u_i^2 \quad (354)$$

$$= \frac{\sum_{i=1}^{n-1} u_i u_{i+1}}{\sum_{i=1}^{n-1} u_i^2}, \quad (355)$$

where (353) is by collecting the terms in the exponent, and (355) is by minimizing the quadratic function of  $a$ .

### B. Proof of Theorem 5

*Proof.* We first bound  $\mathbb{P}[\hat{a}(\mathbf{U}) - a > \eta]$ . Define the random variable  $W(n, \eta)$  as

$$W(n, \eta) \triangleq \sum_{i=1}^{n-1} (U_i Z_{i+1} - \eta U_i^2). \quad (356)$$

Then, from (72) in Section III-C.2 above and (5), it is easy to see that

$$\mathbb{P}[\hat{a}(\mathbf{U}) - a > \eta] = \mathbb{P}[W(n, \eta) > 0]. \quad (357)$$

From the definition of the Gauss-Markov source in (5), we have

$$U_i = \sum_{j=1}^i a^{i-j} Z_j. \quad (358)$$

Plugging (358) into (356), we obtain

$$\begin{aligned} W(n, \eta) &= \sum_{i=2}^n \sum_{j=1}^{i-1} a^{i-j-1} Z_i Z_j - \\ & \quad \frac{\eta}{1-a^2} \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} (a^{|i-j|} - a^{2n-i-j}) Z_i Z_j. \end{aligned} \quad (359)$$

Notice that (359) can be further rewritten as the following quadratic form in the  $n$  i.i.d. random variables  $Z_1, \dots, Z_n$ :

$$W(n, \eta) = \mathbf{Z}^\top \mathbf{Q}(n, \eta) \mathbf{Z}, \quad (360)$$

where  $\mathbf{Q}(n, \eta)$  is an  $n \times n$  symmetric matrix defined as

$$\mathbf{Q}_{i,j}(n, \eta) = \begin{cases} -\eta \frac{1-a^{2(n-i)}}{1-a^2}, & i = j < n \\ 0, & i = j = n \\ \frac{1}{2} a^{|i-j|-1} - \eta \frac{a^{|i-j|} - a^{2n-i-j}}{1-a^2}, & \text{otherwise.} \end{cases} \quad (361)$$

For simplicity, in the rest of the proof, we write  $W$  and  $Q$  for  $W(n, \eta)$  and  $Q(n, \eta)$ , respectively. By the Hanson-Wright inequality [25, Th. 1.1], there exists a universal constant  $c > 0$  such that for any  $t > 0$ <sup>9</sup>

$$\mathbb{P}[W - \mathbb{E}[W] > t] \leq \exp \left[ -c \min \left( \frac{t^2}{\sigma^4 \|Q\|_F^2}, \frac{t}{\sigma^2 \|Q\|} \right) \right], \quad (362)$$

where  $\|Q\|_F$  and  $\|Q\|$  are the Frobenius and operator norms of  $Q$ , respectively. Taking  $t = -\mathbb{E}[W]$  (which is  $> 0$  for all  $n$  large enough, as shown in (366) below) in (362), we can bound (357) as

$$\mathbb{P}[W > 0] \leq \exp \left[ -c \min \left( \frac{(-\mathbb{E}[W])^2}{\sigma^4 \|Q\|_F^2}, \frac{-\mathbb{E}[W]}{\sigma^2 \|Q\|} \right) \right]. \quad (363)$$

It remains to bound  $\mathbb{E}[W]$ ,  $\|Q\|_F^2$  and  $\|Q\|$ . In the following, we show that  $-\mathbb{E}[W] = \Theta(\eta n)$ ,  $\|Q\|_F^2 = \Theta(n)$  and  $\|Q\| = O(1)$ . Plugging these estimates into (363) yields (73) up to constants. The details follow. We first consider  $\mathbb{E}[W]$ . From (360) and (361), we have

$$\mathbb{E}[W] = \sigma^2 \text{tr}(Q) = -\frac{\eta \sigma^2 n}{1-a^2} + \frac{\eta \sigma^2 (1-a^{2n})}{(1-a^2)^2}. \quad (364)$$

Define the constant  $K_1 > 0$  as

$$K_1 \triangleq \frac{1}{2(1-a^2)}. \quad (365)$$

Then, for all  $n$  large enough, we have

$$-\mathbb{E}[W] \geq K_1 \sigma^2 \eta n. \quad (366)$$

We then consider  $\|Q\|_F^2$ . Direct computations using (361) yield

$$\begin{aligned} \|Q\|_F^2 &= \left[ \frac{1}{2(1-a^2)} + \frac{(1+a^2)\eta^2 - 2a(1-a^2)\eta}{(1-a^2)^3} + \frac{(4a\eta^2 - 2(1-a^2)\eta)a^{2n}}{a(1-a^2)^3} \right] \cdot n \\ &\quad + \frac{4a\eta}{(1-a^2)^3} - \frac{1}{2(1-a^2)^2} - \frac{\eta^2(4a^2+1)}{(1-a^2)^4} \\ &\quad + \left[ \frac{4a^2\eta^2}{(1-a^2)^4} + \frac{1}{2(1-a^2)^2} - \frac{4a\eta}{(1-a^2)^3} \right] \cdot a^{2n} \\ &\quad + \frac{\eta^2}{(1-a^2)^4} \cdot a^{4n}. \end{aligned} \quad (367)$$

Define the constant  $K_2 > 0$  as

$$K_2 \triangleq \frac{1}{1-a^2} + \frac{2(1+5a^2)\eta^2}{(1-a^2)^3}. \quad (368)$$

<sup>9</sup>The sub-gaussian norm  $\|Z_i\|_{\psi_2}$  [45, Def. 5.7] of  $Z_i \sim \mathcal{N}(0, \sigma^2)$  satisfies  $\|Z_i\|_{\psi_2} \leq C_1 \sigma$  for a universal constant  $C_1 > 0$ , see [45, Example 5.8]. By [25, Th. 1.1], there exists a universal constant  $C_2 > 0$  such that for any  $t > 0$ ,

$$\mathbb{P}[W - \mathbb{E}[W] > t] \leq \exp \left[ -C_2 \min \left( \frac{t^2}{C_1^4 \sigma^4 \|Q\|_F^2}, \frac{t}{C_1^2 \sigma^2 \|Q\|} \right) \right].$$

In (362), we absorb  $C_1$  and  $C_2$  into a single universal constant  $c$ .

Since  $\eta > 0$  and  $a \in [0, 1)$ , from (367), we have for all  $n$  large enough,

$$\|Q\|_F^2 \leq K_2 n. \quad (369)$$

Finally, we bound  $\|Q\|$ . Using the Gershgorin circle theorem [43, p.16, Th. 1.11], one can easily show that

$$\|Q\| \leq \max_{i \in [n]} \|q_i\|_1, \quad (370)$$

where  $q_i$  denotes the  $i$ -th row in  $Q$ . Direct computations using (361) yield that  $\forall i \in [n]$ ,

$$\|q_i\|_1 \leq K_3, \quad (371)$$

where  $K_3$  is a positive constant given by

$$K_3 \triangleq \frac{1}{1-a} + \frac{\eta}{1-a^2} + \frac{2\eta}{(1-a)^2}. \quad (372)$$

Therefore,  $\forall n \geq 1$ , we have

$$\|Q\| \leq K_3. \quad (373)$$

Plugging the bounds (366), (369) and (373) into (363), we have for all  $n$  large enough

$$\mathbb{P}[W > 0] \leq \exp \left[ -c \min \left( \frac{K_1^2 \eta^2 n}{K_2}, \frac{K_1 \eta n}{K_3} \right) \right]. \quad (374)$$

Notice that all the arguments up to this point are valid for any  $\eta > 0$ . However, the constants  $K_2$  and  $K_3$  in the bound (374) depend on  $\eta$  via (368) and (372), respectively. Since we are interested in small  $\eta$ , in the rest of the proof, we assume  $\eta \in (0, 1)$ . Besides, with the restriction  $\eta \in (0, 1)$ , we can get rid of the dependence of  $K_2$  and  $K_3$  on  $\eta$ , and simplify (374). For any  $\eta \in (0, 1)$ , we can bound  $K_2$  and  $K_3$  as follows:

$$K_2 \leq K'_2 \triangleq \frac{1}{1-a^2} + \frac{2(5+a^2)}{(1-a^2)^3}, \quad (375)$$

$$K_3 \leq K'_3 \triangleq \frac{a+4}{(1-a)^2(1+a)}. \quad (376)$$

Applying the bounds (375) and (376) to (374), and then setting

$$c_1 \triangleq \frac{K_1^2}{K'_2} \quad \text{and} \quad c_2 \triangleq \frac{K_1}{K'_3} \quad (377)$$

yields

$$\mathbb{P}[\hat{a}(U) - a > \eta] \leq \exp \left[ -c \min (c_1 \eta^2 n, c_2 \eta n) \right]. \quad (378)$$

Finally, to bound  $\mathbb{P}[\hat{a}(U) - a < -\eta]$ , in the above proof, we replace the random variable  $W(n, \eta)$  by  $V(n, \eta)$ , defined as

$$V(n, \eta) \triangleq \sum_{i=1}^{n-1} (-U_i Z_{i+1} - \eta U_i^2). \quad (379)$$

In quadratic forms,  $V(n, \eta) = \mathbf{Z}^\top \mathbf{S}(n, \eta) \mathbf{Z}$ , where  $\mathbf{S}(n, \eta)$  is an  $n \times n$  symmetric matrix, defined in a way similar to  $Q(n, \eta)$ :

$$\mathbf{S}_{i,j}(n, \eta) = \begin{cases} -\eta \frac{1-a^{2(n-i)}}{1-a^2}, & i = j < n \\ 0, & i = j = n \\ -\frac{1}{2} a^{|i-j|-1} - \eta \frac{a^{|i-j|} - a^{2n-i-j}}{1-a^2}, & \text{otherwise.} \end{cases} \quad (380)$$



With the same techniques as above, we obtain

$$\mathbb{P}[\hat{a}(\mathbf{U}) - a < -\eta] \leq \exp\left[-c \min\left(c_1 \eta^2 n, c_2 \eta n\right)\right], \quad (381)$$

where  $c$ ,  $c_1$  and  $c_2$  are the same constants as those in (378). ■

### C. Proof of Theorem 6

*Proof.* The proof is similar to that of Theorem 5. In particular, (374) still holds:

$$\mathbb{P}[\hat{a}(\mathbf{U}) - a > \eta_n] \leq \exp\left[-c \min\left(\frac{K_1^2 \eta_n^2 n}{K_2}, \frac{K_1 \eta_n n}{K_3}\right)\right]. \quad (382)$$

Instead of (375) and (376), we bound  $K_2$  and  $K_3$  by

$$K_2 \leq K_2'' \triangleq \frac{2}{1-a^2}, \quad (383)$$

and

$$K_3 \leq K_3'' \triangleq \frac{2}{1-a}, \quad (384)$$

where (383) and (384) hold for all  $n$  large enough in view of (74), (368) and (372). Applying the bounds (383) and (384) to (382) and using the fact that for all  $n$  large enough,

$$\min\left(\frac{K_1^2 \eta_n^2 n}{K_2''}, \frac{K_1 \eta_n n}{K_3''}\right) = \frac{\eta_n^2 n}{8(1-a^2)}, \quad (385)$$

we obtain

$$\mathbb{P}[\hat{a}(\mathbf{U}) - a > \eta_n] \leq \frac{1}{(\log n)^{\kappa a}}, \quad (386)$$

where  $\kappa$  is given in (76). Finally, we can bound  $\mathbb{P}[\hat{a}(\mathbf{U}) - a < -\eta_n]$  in the same way. ■

### ACKNOWLEDGMENT

We would like to thank the associate editor Dr. Shun Watanabe and the anonymous reviewers for their insightful comments that are reflected in the final version.

### REFERENCES

- [1] P. Tian and V. Kostina, "The dispersion of the Gauss-Markov source," in *Proc. IEEE Int. Symp. Inf. Theory*, Vail, CO, USA, Jun. 2018, pp. 1490–1494.
- [2] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul. 1948.
- [3] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE Nat. Conv. Rec.*, vol. 4, no. 1, pp. 142–163, Mar. 1959.
- [4] T. Gollub, "A coding theorem for time-discrete analog data sources," *IEEE Trans. Inf. Theory*, vol. IT-15, no. 3, pp. 401–407, May 1969.
- [5] R. Gray, "Information rates of autoregressive processes," *IEEE Trans. Inf. Theory*, vol. IT-16, no. 4, pp. 412–421, Jul. 1970.
- [6] S. Jalali and T. Weissman, "New bounds on the rate-distortion function of a binary Markov source," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2007, pp. 571–575.
- [7] R. Gray, "Rate distortion functions for finite-state finite-alphabet Markov sources," *IEEE Trans. Inf. Theory*, vol. IT-17, no. 2, pp. 127–134, Mar. 1971.
- [8] L. D. Davisson, "Rate-distortion theory and application," *Proc. IEEE*, vol. 60, no. 7, pp. 800–808, Jul. 1972.
- [9] A. Kolmogorov, "On the Shannon theory of information transmission in the case of continuous signals," *IRE Trans. Inf. Theory*, vol. 2, no. 4, pp. 102–108, Dec. 1956.
- [10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2012.
- [11] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1971.
- [12] T. Berger and J. D. Gibson, "Lossy source coding," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2693–2723, Oct. 1998.
- [13] Z. Zhang, E.-H. Yang, and V. K. Wei, "The redundancy of source coding with a fidelity criterion. 1. Known statistics," *IEEE Trans. Inf. Theory*, vol. 43, no. 1, pp. 71–91, Jan. 1997.
- [14] E.-H. Yang and Z. Zhang, "On the redundancy of lossy source coding with abstract alphabets," *IEEE Trans. Inf. Theory*, vol. 45, no. 4, pp. 1092–1110, May 1999.
- [15] V. Kostina and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3309–3338, Jun. 2012.
- [16] A. Ingber and Y. Kochman, "The dispersion of lossy source coding," in *Proc. Data Compression Conf.*, Snowbird, UT, USA, Mar. 2011, pp. 53–62.
- [17] V. Kostina and S. Verdú, "Lossy joint source-channel coding in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2545–2575, May 2013.
- [18] D. Wang, A. Ingber, and Y. Kochman, "The dispersion of joint source-channel coding," in *Proc. 49th Annu. Allerton Conf. Commun., Control, Comput.*, Sep. 2011, pp. 180–187.
- [19] S. Watanabe, "Second-order region for Gray-Wyner network," *IEEE Trans. Inf. Theory*, vol. 63, no. 2, pp. 1006–1018, Feb. 2017.
- [20] V. Y. F. Tan and O. Kosut, "On the dispersions of three network information theory problems," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 881–903, Feb. 2014.
- [21] I. Kontoyiannis, "Pointwise redundancy in lossy data compression and universal lossy data compression," *IEEE Trans. Inf. Theory*, vol. 46, no. 1, pp. 136–152, Jan. 2000.
- [22] A. A. Yushkevich, "On limit theorems connected with the concept of entropy of Markov chains," (in Russian), *Uspekhi Matematicheskikh Nauk*, vol. 8, no. 5, pp. 177–180, 1953.
- [23] I. Kontoyiannis and S. Verdú, "Optimal lossless data compression: Non-asymptotics and asymptotics," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 777–795, Feb. 2014.
- [24] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Dispersion of the Gilbert-Elliott channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1829–1848, Apr. 2011.
- [25] M. Rudelson and R. Vershynin, "Hanson-Wright inequality and sub-Gaussian concentration," *Electron. Commun. Probab.*, vol. 18, no. 82, pp. 1–10, 2013.
- [26] R. G. Gallager, *Information Theory and Reliable Communication*, vol. 2. New York, NY, USA: Wiley, 1968.
- [27] I. Csiszár, "On an extremum problem of information theory," *Studia Scientiarum Mathematicarum Hungarica*, vol. 9, no. 1, pp. 57–71, 1974.
- [28] R. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 4, pp. 460–473, Jul. 1972.
- [29] A. Dembo and L. Kontoyiannis, "Source coding, large deviations, and approximate pattern matching," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1590–1615, Jun. 2002.
- [30] J. C. Kieffer, "Strong converses in source coding relative to a fidelity criterion," *IEEE Trans. Inf. Theory*, vol. 37, no. 2, pp. 257–262, Mar. 1991.
- [31] T. Berger, "Information rates of Wiener processes," *IEEE Trans. Inf. Theory*, vol. IT-16, no. 2, pp. 134–139, Mar. 1970.
- [32] U. Grenander and G. Szegő, *Toeplitz Forms and Their Applications*. Berkeley, CA, USA: Univ. California Press, 2001, vol. 321.
- [33] R. M. Gray, "Toeplitz and circulant matrices: A review," *Found. Trends Commun. Inf.*, vol. 2, no. 3, pp. 155–239, 2006.
- [34] T. Hashimoto and S. Arimoto, "On the rate-distortion function for the nonstationary Gaussian autoregressive process (Corresp.)," *IEEE Trans. Inf. Theory*, vol. IT-26, no. 4, pp. 478–480, Jul. 1980.
- [35] I. Dumer, M. S. Pinsker, and V. V. Prelov, "On coverings of ellipsoids in Euclidean spaces," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2348–2356, Oct. 2004.
- [36] R. M. Gray and T. Hashimoto, "A note on rate-distortion functions for nonstationary Gaussian autoregressive processes," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 1319–1322, Mar. 2008.
- [37] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Dispersion of Gaussian channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Seoul, South Korea, Jun./Jul. 2009, pp. 2204–2208.

- [38] B. S. Tsybakov, "Capacity of a discrete-time Gaussian channel with a filter," (in Russian), *Problemy Peredachi Informatsii*, vol. 6, no. 3, pp. 78–82, Jul./Sep. 1970.
- [39] W. Hirt and J. L. Massey, "Capacity of the discrete-time Gaussian channel with intersymbol interference," *IEEE Trans. Inf. Theory*, vol. IT-34, no. 3, p. 38, May 1988.
- [40] W. Feller, *An Introduction to Probability Theory and its Applications*, vol. 2. New York, NY, USA: Wiley, 1971.
- [41] L. V. Ahlfors, *Complex Analysis, an Introduction to the Theory of Analytic Functions of One Complex Variable*, 3rd ed. New York, NY, USA: McGraw-Hill, 1979.
- [42] R. Bhatia, *Matrix Analysis*, vol. 169. New York, NY, USA: Springer, 2013.
- [43] R. S. Varga, *Matrix Iterative Analysis*, vol. 27. Berlin, Germany: Springer, 2009.
- [44] S. Noschese, L. Pasquini, and L. Reichel, "Tridiagonal toeplitz matrices: Properties and novel applications," *Numer. Linear Algebra Appl.*, vol. 20, no. 2, pp. 302–326, Mar. 2013.
- [45] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," in *Compressed Sensing: Theory and Applications*, Y. C. Eldar and G. Kutyniok, Eds. New York, NY, USA: Cambridge Univ. Press, 2012, ch. 5, pp. 210–268.

**Peida Tian** is currently a PhD student in the Department of Electrical Engineering at California Institute of Technology. He received a B. Engg. in Information Engineering and a B. Sc. in Mathematics from the Chinese University of Hong Kong (2016), and a M.S. in Electrical Engineering from Caltech (2017). He is interested in optimization and information theory.

**Victoria Kostina** (S'12–M'14) joined Caltech as an Assistant Professor of Electrical Engineering in the fall of 2014. She holds a Bachelor's degree from Moscow Institute of Physics and Technology (2004), where she was affiliated with the Institute for Information Transmission Problems of the Russian Academy of Sciences, a Master's degree from University of Ottawa (2006), and a PhD from Princeton University (2013). She received the Natural Sciences and Engineering Research Council of Canada postgraduate scholarship (2009–2012), the Princeton Electrical Engineering Best Dissertation Award (2013), the Simons-Berkeley research fellowship (2015) and the NSF CAREER award (2017). Kostina's research spans information theory, coding, control and communications.