From Parameter Estimation to Dispersion of Nonstationary Gauss-Markov Processes

Peida Tian, Victoria Kostina

***Abstract*—This paper provides a precise error analysis for the maximum likelihood estimate $\hat{a}(\boldsymbol{u})$ of the parameter a given samples $\boldsymbol{u} = (u_1, \ldots, u_n)^\top$ drawn from a nonstationary Gauss-Markov process $U_i = aU_{i-1} + Z_i,\ i \geq 1$, where $a > 1$, $U_0 = 0$, and Z_i's are independent Gaussian random variables with zero mean and variance σ^2. We show a tight nonasymptotic exponentially decaying bound on the tail probability of the estimation error. Unlike previous works, our bound is tight already for a sample size of the order of hundreds. We apply the new estimation bound to find the dispersion for lossy compression of nonstationary Gauss-Markov sources. We show that the dispersion is given by the same integral formula derived in our previous work [1] for the (asymptotically) stationary Gauss-Markov sources, i.e., $|a| < 1$. New ideas in the nonstationary case include a deeper understanding of the scaling of the maximum eigenvalue of the covariance matrix of the source sequence, and new techniques in the derivation of our estimation error bound.**

I. Introduction

A scalar Gauss-Markov process $\{U_i\}_{i=1}^{\infty}$ is a random process defined as

$$U_i = aU_{i-1} + Z_i, \quad i \geq 1, \tag{1}$$

where $U_0 = 0$ and Z_i's are independent Gaussian random variables with zero mean and variance σ^2, $Z_i \sim \mathcal{N}(0, \sigma^2)$. We assume without loss of generality that $a \geq 0$. We make the distinctions between the following three cases: the (asymptotically) stationary case refers to $0 < a < 1$ in (1); the unit-root case to $a = 1$ [1]; and the nonstationary case to $a > 1$. This paper mostly focuses on the nonstationary case.

Our primary motivation for studying the Gauss-Markov process is to understand the role of memory in nonasymptotic rate-distortion theory. The Gauss-Markov process in (1) is one of the simplest models for information sources with memory. The *rate-distortion function* (RDF) [2] captures the rate-distortion tradeoff when the coding length tends to infinity. The central question in nonasymptotic rate-distortion theory is to characterize the rate-distortion tradeoff when the coding length is constrained to be finite, and the *dispersion* is the main quantity of interest. The dispersion of stationary memoryless sources was found in [3, 4]. The dispersion of information sources with memory is largely unknown. Our previous work [1] found the dispersion of the stationary Gauss-Markov source.

P. Tian and V. Kostina are with the Department of Electrical Engineering, California Institute of Technology. (e-mail: {ptian, vkostina}@caltech.edu).
This research was supported in part by the National Science Foundation (NSF) under Grant CCF-1751356.

[1]Technically, the unit-root case is also nonstationary.

One of the key ideas in [1] is to construct a typical set based on $\hat{a}(\boldsymbol{u})$, the maximum likelihood estimate (MLE) of a given samples $\boldsymbol{u} = (u_1, \ldots, u_n)^\top$. For a typical $\boldsymbol{u}$, $\hat{a}(\boldsymbol{u})$ is close to a. The MLE $\hat{a}(\boldsymbol{u})$ of the parameter a can be easily obtained as (see [5, Eq. (352)-(356)] for a derivation):

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

For $0 < a < 1$, our previous work [5, Th. 5] provided a tight bound on the tail probability of the estimation error $|\hat{a}(\boldsymbol{U}) - a|$. Using different tools, for $a > 1$, this paper derives an exponentially decaying upper bound on the tail probability of $\hat{a}(\boldsymbol{U}) - a$. This result complements the large body of works [6–10] studying various aspects of the MLE $\hat{a}(\boldsymbol{u})$. Our bound is nonasymptotic and tighter than existing bounds, see Fig 1 in Section III-A below for a comparison. As an application of the error bound, we find the dispersion for the nonstationary Gauss-Markov source. Although the dispersion is represented by the same formula as the one we derived for the stationary case [1, Th. 1], there is a subtle difference between the analyses of the two scenarios. In fact, after the RDF of the stationary Gauss-Markov source was derived [11] (see also [12, Th. 4.5.3]), it still took several decades to completely understand the RDF of the nonstationary one [13–15]. Throughout the paper, all logarithms and exponents are base e. All detailed proofs are given in the long version [16].

II. Previous Works

A. Parameter estimation

The estimator $\hat{a}(\boldsymbol{u})$ in (2) has been extensively studied in the statistics [6, 7] and economics [17, 18] communities. It is well known [7, 17, 18] that the estimation error $\hat{a}(\boldsymbol{U}) - a$ converges to 0 in probability for any $a \in \mathbb{R}$, as n tends to infinity. To better understand how the error $\hat{a}(\boldsymbol{U}) - a$ scales as n tends to infinity, researchers turned to study the limiting distribution of the normalized estimation error $h(n)(\hat{a}(\boldsymbol{U}) - a)$ for a careful choice of the standardizing function $h(n)$:

$$h(n) \triangleq \begin{cases} \sqrt{\frac{n}{1-a^2}}, & |a| < 1, \\ \frac{n}{\sqrt{2}}, & |a| = 1, \\ \frac{|a|^n}{a^2-1}, & |a| > 1. \end{cases} \tag{3}$$

Mann and Wald [17] and White [6] showed that the distribution of the normalized estimation error $h(n)(\hat{a}(\boldsymbol{U}) - a)$ converges to $\mathcal{N}(0, 1)$ for $|a| < 1$; to the Cauchy distribution with the probability density function $\frac{1}{\pi(1+x^2)}$ for $|a| > 1$; and for $|a| = 1$, to the distribution of $\frac{B^2(1)-1}{2\int_0^1 B^2(t)\ dt}$, where $\{B(t) : t \in [0, 1]\}$ is a Brownian motion.

This paper presents a nonasymptotic fine-grained large deviations analysis of the estimation error. Given an error threshold $\eta > 0$, define the error exponents P_n^+ and P_n^- as

$$P_n^+ \triangleq -\frac{1}{n}\log \mathbb{P}\left[\hat{a}(\boldsymbol{U}) - a > \eta\right], \quad (4)$$

$$P_n^- \triangleq -\frac{1}{n}\log \mathbb{P}\left[\hat{a}(\boldsymbol{U}) - a < -\eta\right]. \quad (5)$$

We also define P_n as

$$P_n \triangleq -\frac{1}{n}\log \mathbb{P}\left[|\hat{a}(\boldsymbol{U}) - a| > \eta\right]. \quad (6)$$

For $0 < a < 1$, Bercu et al. [8] showed that

$$\lim_{n\to\infty} P_n = I_{\mathrm{s}}(\eta), \quad (7)$$

where the rate function $I_{\mathrm{s}}(\eta)$ is given in [8, Prop. 8]. For $a > 1$, Worms [9, Th. 1] proved that

$$\liminf_{n\to\infty} P_n \geq I_{\mathrm{ns}}(\eta), \quad (8)$$

where $I_{\mathrm{ns}}(\eta)$ is specified in [9, Th. 1] as the optimal value of an optimization problem. A bound similar to (8) for the unit-root case was also presented in [9, Th. 1]. To our knowledge, there are two nonasymptotic lower bounds on P_n^+ and P_n^-. For any $a \in \mathbb{R}$, Rantzer [10, Th. 4] showed that

$$P_n^+ \text{ (and } P_n^-) \geq \frac{1}{2}\log(1+\eta^2). \quad (9)$$

Bercu and Touati [19, Cor. 5.2] proved that

$$P_n^+ \text{ (and } P_n^-) \geq \frac{\eta^2}{2(1+y_\eta)}, \quad (10)$$

where y_η is the unique positive solution to $(1+x)\log(1+x) - x - \eta^2 = 0$ in x. Both bounds (9) and (10) do not depend on a and n, and are the same for P_n^+ and P_n^-.

This paper shows tight nonasymptotic bounds on P_n^+, P_n^- and P_n. Our nonasymptotic lower bounds on P_n^+ and P_n^- depend on a and n, and are distinct for P_n^+ and P_n^-. For larger a, our lower bound becomes larger, which suggests that unstable systems are easier to estimate than stable ones, an observation consistent with [20]. The proof is inspired by Rantzer [10, Lem. 5], but our result significantly improves (9) and (10), see Fig. 1 for a comparison.

B. Nonasymptotic rate-distortion theory

Given a distortion threshold $d > 0$, an excess-distortion probability $\epsilon \in (0,1)$ and $M \in \mathbb{N}$, an (n, M, d, ϵ) lossy compression code for a random vector $\boldsymbol{U} = (U_1, \ldots, U_n)^\top$ of length n consists of an encoder $\mathsf{f}_n\colon \mathbb{R}^n \to [M]$, and a decoder $\mathsf{g}_n\colon [M] \to \mathbb{R}^n$, such that $\mathbb{P}\left[\mathsf{d}\left(\boldsymbol{U}, \mathsf{g}_n\left(\mathsf{f}_n(\boldsymbol{U})\right)\right) > d\right] \leq \epsilon$, where $\mathsf{d}(\cdot,\cdot)$ is the distortion measure. In this paper, we consider the mean squared error (MSE) distortion: $\forall\, \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$,

$$\mathsf{d}(\boldsymbol{u}, \boldsymbol{v}) \triangleq \frac{1}{n}\sum_{i=1}^{n}(u_i - v_i)^2. \quad (11)$$

The minimum achievable code size and source coding rate are defined respectively by

$$M^\star(n,d,\epsilon) \triangleq \min\left\{M \in \mathbb{N}\colon \exists\ (n,M,d,\epsilon) \text{ code}\right\}, \quad (12)$$

$$R(n,d,\epsilon) \triangleq \frac{1}{n}\log M^\star(n,d,\epsilon). \quad (13)$$

The core problem in the nonasymptotic rate-distortion theory is to characterize $R(n,d,\epsilon)$. For stationary memoryless sources, Ingber and Kochman [3] (finite-alphabet and Gaussian sources) and Kostina and Verdú [4] (abstract sources) showed that $R(n,d,\epsilon)$ satisfies a Gaussian approximation of the form (42) in Section III-C below. This paper extends our previous analyses [1] on the stationary Gauss-Markov sources to the nonstationary case. One of the key ideas behind that extension is to construct a typical set using the MLE of a, and to use our estimation error bound to probabilistically characterize that set.

III. Main Results

A. Parameter estimation

We first present our nonasymptotic bounds on P_n^+ and P_n^- using two sequences $\{\alpha_\ell\}_{\ell\in\mathbb{N}}$ and $\{\beta_\ell\}_{\ell\in\mathbb{N}}$ defined as follows. Throughout the paper, $\sigma^2 > 0$ and $a > 1$ are fixed. For $\eta > 0$ and a parameter $s > 0$, let $\{\alpha_\ell\}_{\ell\in\mathbb{N}}$ be the following sequence

$$\alpha_1 \triangleq \frac{\sigma^2 s^2 - 2\eta s}{2}, \quad (14)$$

$$\alpha_\ell = \frac{[a^2 + 2\sigma^2 s(a+\eta)]\alpha_{\ell-1} + \alpha_1}{1 - 2\sigma^2\alpha_{\ell-1}}, \quad \forall \ell \geq 2. \quad (15)$$

Similarly, let $\{\beta_\ell\}_{\ell\in\mathbb{N}}$ be the following sequence:

$$\beta_1 \triangleq \frac{\sigma^2 s^2 - 2\eta s}{2}, \quad (16)$$

$$\beta_\ell = \frac{[a^2 + 2\sigma^2 s(-a+\eta)]\beta_{\ell-1} + \beta_1}{1 - 2\sigma^2\beta_{\ell-1}}, \quad \forall \ell \geq 2. \quad (17)$$

Note the slight difference between (15) and (17). We analyze the convergence properties of α_ℓ and β_ℓ in Appendix A-B below. For any $\eta > 0$ and $n \in \mathbb{N}$, we define the following sets

$$\mathcal{S}_n^+ \triangleq \left\{s \in \mathbb{R}\colon s > 0,\ \alpha_\ell < \frac{1}{2\sigma^2},\ \forall \ell \in [n]\right\}, \quad (18)$$

$$\mathcal{S}_n^- \triangleq \left\{s \in \mathbb{R}\colon s > 0,\ \beta_\ell < \frac{1}{2\sigma^2},\ \forall \ell \in [n]\right\}. \quad (19)$$

Theorem 1. *For any constant $\eta > 0$, the estimator* (2) *satisfies for any $n \geq 2$,*

$$P_n^+ \geq \sup_{s\in\mathcal{S}_n^+} \frac{1}{2n}\sum_{\ell=1}^{n-1}\log\left(1 - 2\sigma^2\alpha_\ell\right), \quad (20)$$

$$P_n^- \geq \sup_{s\in\mathcal{S}_n^-} \frac{1}{2n}\sum_{\ell=1}^{n-1}\log\left(1 - 2\sigma^2\beta_\ell\right), \quad (21)$$

where $\{\alpha_\ell\}_{\ell\in\mathbb{N}}$ and $\{\beta_\ell\}_{\ell\in\mathbb{N}}$ are defined in (15) *and* (17)*, respectively, and $\mathcal{S}_n^+$ and $\mathcal{S}_n^-$ are defined in* (18) *and* (19)*, respectively.*

Proof. Appendix A-A. ■

The proof is a detailed analysis of the Chernoff bound using the tower property of conditional expectations. The proof is motivated by [10, Lem. 5], but our analysis is more accurate and the result is significantly tighter, see Fig. 1 for a comparison. Theorem 1 gives the best bound that can be obtained from the Chernoff bound. In view of the Gärtner-Ellis theorem [21, Th. 2.3.6], we conjecture that the bounds (20) and (21) can be reversed in the limit of large n.

The exact characterization of $\mathcal{S}_n^+$ and $\mathcal{S}_n^-$ for each n using σ^2, a and η is involved. However, the limits[2]

$$S_\infty^+ \triangleq \lim_{n\to\infty} S_n^+ = \bigcap_{n\geq 1} S_n^+, \tag{22}$$

$$S_\infty^- \triangleq \lim_{n\to\infty} S_n^- = \bigcap_{n\geq 1} S_n^-, \tag{23}$$

can be characterized in terms of the interval $\mathcal{I}_\eta$:

$$\mathcal{I}_\eta \triangleq \left(0,\ 2\eta/\sigma^2\right). \tag{24}$$

Lemma 1. *For any constant $\eta > 0$, we have for any $n \in \mathbb{N}$,*

$$S_\infty^+ = \mathcal{I}_\eta \cup \{2\eta/\sigma^2\}, \tag{25}$$

$$S_\infty^- \supsetneq \mathcal{I}_\eta \cup \{2\eta/\sigma^2\}. \tag{26}$$

The proof of Lemma 1 can be found in the long version [16]. One recovers Rantzer's lower bound (9) by setting $s = \eta/\sigma^2$ and bounding α_ℓ as $\alpha_\ell \leq \alpha_1$ (due to the monotonicity of α_ℓ, see Appendix A-B) in Theorem 1. Using Lemma 1 and taking limits in Theorem 1, we obtain the following result.

Theorem 2. *For any constant $\eta > 0$, we have*

$$\liminf_{n\to\infty}\ P_n^+ \geq \log(a + 2\eta) \tag{27}$$

$$\liminf_{n\to\infty}\ P_n^- \geq \sup_{s\in\mathcal{I}_\eta}\ \frac{1}{2}\log(1 - 2\sigma^2 t_1), \tag{28}$$

where t_1 is the smaller root of the quadratic equation (59) in Appendix A-B below.

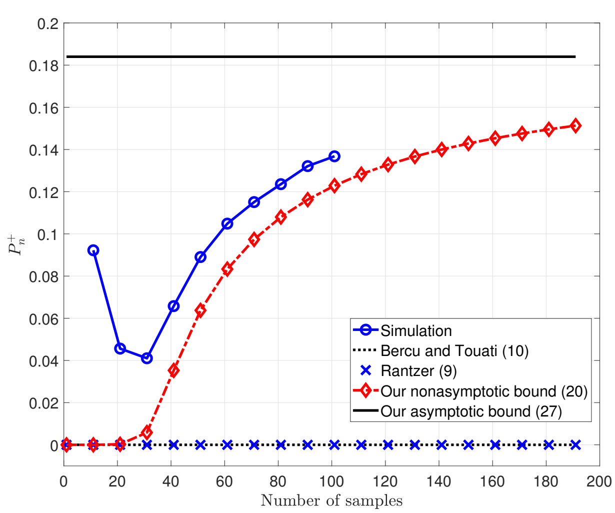

Fig. 1: Numerical simulations for $a = 1.2$ and $\eta = 10^{-3}$. See the full paper for more comparisons.

In the case where $\eta = \eta_n > 0$ is a sequence decreasing to 0, one can show that Theorem 1 still holds. For Theorem 2 to remain valid, we require that the speed of η_n decreasing to zero is no smaller than $\frac{1}{\sqrt{n}}$, which essentially ensures that the right side of (20) and (21) still converges to the right side of (27) and (28), respectively. Let η_n be a positive sequence such that

$$\eta_n = \Omega\left(\frac{1}{\sqrt{n}}\right). \tag{29}$$

Theorem 3. *For $a > 1$, Theorem 1 and Theorem 2 still hold even when $\eta = \eta_n > 0$ is a sequence satisfying (29).*

The following corollary to Theorem 3 is used in Section III-C below to derive the dispersion of nonstationary Gauss-Markov sources.

Corollary 1. *For any $\sigma^2 > 0$ and $a > 1$, there exists a constant $c \geq \frac{1}{2}\log a$ such that for all n large enough,*

$$\mathbb{P}\left[|\hat{a}(\boldsymbol{U}) - a| \geq \sqrt{\frac{\log\log n}{n}}\right] \leq 2e^{-cn}. \tag{30}$$

Remark 1. Theorem 1-3 generalize to the case where Z_i's in (1) are σ-subgaussian. See details in the full paper.

B. Nonasymptotic rate-distortion theory: preliminaries

We review some definitions before we discuss our main results on the dispersion of nonstationary Gauss-Markov processes. For a random process $\{X_i\}_{i=1}^\infty$, the n-th order rate-distortion function $\mathbb{R}_{\boldsymbol{X}}(d)$ is defined as

$$\mathbb{R}_{\boldsymbol{X}}(d) \triangleq \inf_{\substack{P_{\boldsymbol{Y}|\boldsymbol{X}}:\\ \mathbb{E}[\mathsf{d}(\boldsymbol{X},\boldsymbol{Y})]\leq d}} \frac{1}{n} I(\boldsymbol{X};\boldsymbol{Y}), \tag{31}$$

where $\boldsymbol{X} = (X_1, \ldots, X_n)^\top$ is the n-dimensional random source vector. The rate-distortion function $\mathbb{R}_X(d)$ is

$$\mathbb{R}_X(d) \triangleq \limsup_{n\to\infty}\ \mathbb{R}_{\boldsymbol{X}}(d). \tag{32}$$

Closed-form expressions for $\mathbb{R}_{\boldsymbol{X}}(d)$ and $\mathbb{R}_X(d)$ are known only in a few special cases. Specializing Gray's result [13, Eq. (22)] for Gaussian autoregressive processes to our Gauss-Markov source (1), we write down the *n-th order reverse waterfilling* solution for the n-th order informational rate-distortion function $\mathbb{R}_{\boldsymbol{U}}(d)$:

$$\mathbb{R}_{\boldsymbol{U}}(d) = \frac{1}{n}\sum_{i=1}^n \frac{1}{2}\log\max\left(\mu_i, \frac{\sigma^2}{\theta_n}\right), \tag{33}$$

$$d = \frac{1}{n}\sum_{i=1}^n \min\left(\theta_n, \frac{\sigma^2}{\mu_i}\right), \tag{34}$$

where $\theta_n > 0$ is the water level and μ_i's are the eigenvalues of $\mathsf{A}^\top\mathsf{A}$, and A is the $n \times n$ lower triangular matrix:

$$\mathsf{A}_{ij} = \begin{cases} 1, & i = j, \\ -a, & i = j+1, \\ 0, & \text{otherwise.} \end{cases} \tag{35}$$

The rate-distortion function of the Gauss-Markov source is given by the *limiting reverse waterfilling*:

$$\mathbb{R}_U(d) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{2}\log\max\left(g(w),\ \frac{\sigma^2}{\theta}\right)\ dw, \tag{36}$$

$$d = \frac{1}{2\pi}\int_{-\pi}^{\pi} \min\left(\theta,\ \frac{\sigma^2}{g(w)}\right)\ dw, \tag{37}$$

[2]It is easy to see from (18) and (19) that $S_{n+1}^+ \subseteq S_n^+$ and $S_{n+1}^- \subseteq S_n^-$.

where for $w \in [-\pi, \pi]$,

$$g(w) \triangleq 1 + a^2 - 2a\cos(w). \quad (38)$$

Gray [13, Th. 2] showed that

$$\lim_{n\to\infty} R(n,d,\epsilon) = \mathbb{R}_U(d),\ \forall \epsilon \in (0,1). \quad (39)$$

The *operational dispersion* [4, Def. 7] is defined as

$$V_U(d) \triangleq \lim_{\epsilon\to 0} \limsup_{n\to\infty}\ n\left(\frac{R(n,d,\epsilon) - \mathbb{R}_U(d)}{Q^{-1}(\epsilon)}\right)^2, \quad (40)$$

which captures the convergence rate of $R(n,d,\epsilon)$ to $\mathbb{R}_U(d)$.

We refer to the transformed random vector $\boldsymbol{X} \triangleq \mathsf{S}^\top \boldsymbol{U}$ as the *decorrelation* of $\boldsymbol{U}$, where S is the orthornomal matrix that diagonalizes $(\mathsf{A}^\top\mathsf{A})^{-1}$. We can easily see that $\boldsymbol{X}$ has independent coordinates $X_i \sim \mathcal{N}(0, \sigma_i^2)$, where $\sigma_i^2 \triangleq \frac{\sigma^2}{\mu_i}$. The d-*tilted information* [4, Def. 6] for the Gauss-Markov source in $\boldsymbol{u} \in \mathbb{R}^n$ is given by [5, Th. 2]: $\jmath_{\boldsymbol{U}}(\boldsymbol{u}, d) = \jmath_{\boldsymbol{X}}(\boldsymbol{x}, d)$ and

$$\jmath_{\boldsymbol{X}}(\boldsymbol{x}, d) = \sum_{i=1}^n \frac{\min(\theta_n, \sigma_i^2)}{2\theta_n}\left(\frac{x_i^2}{\sigma_i^2} - 1\right) + \frac{1}{2}\sum_{i=1}^n \log\frac{\max(\theta_n, \sigma_i^2)}{\theta_n}, \quad (41)$$

where $\theta_n > 0$ is given by (34) and $\boldsymbol{x} \triangleq \mathsf{S}^\top \boldsymbol{u}$. In lossy compression of i.i.d. sources, the mean and the variance of the d-tilted information are equal to the rate-distortion function and the dispersion, respectively [4, Th. 12]. This paper provides a natural extension of the above fact to nonstationary Gauss-Markov sources.

C. Dispersion of the nonstationary Gauss-Markov process

Theorem 4. *Consider the Gauss-Markov source* (1) *with* $a > 1$. *For any fixed excess-distortion probability* $\epsilon \in (0,1)$ *and distortion threshold* $d > 0$, *the minimum achievable source coding rate* $R(n,d,\epsilon)$ *admits the following Gaussian approximation:*

$$R(n,d,\epsilon) = \mathbb{R}_U(d) + Q^{-1}(\epsilon)\sqrt{\frac{\mathbb{V}_U(d)}{n}} + o\left(\frac{1}{\sqrt{n}}\right), \quad (42)$$

where Q^{-1} *denotes the inverse Q-function;* $\mathbb{R}_U(d)$ *is given in* (36)*; the informational dispersion* $\mathbb{V}_U(d)$ *(defined as the variance of* $\jmath_{\boldsymbol{U}}(U,d)$*) is given by*

$$\mathbb{V}_U(d) = \frac{1}{4\pi}\int_{-\pi}^{\pi} \min\left[1, \left(\frac{\sigma^2}{\theta g(w)}\right)^2\right] dw, \quad (43)$$

and $g(w)$ *is in* (38).

The main ideas in the proof of Theorem 4 are similar to those in proving [5, Th. 1]. In particular, by defining a typical set $\mathcal{T}$ in the form of [5, Def. 1] using the MLE $\hat{a}(\boldsymbol{u})$, we prove a lossy asymptotic equipartition property (AEP) in the form of [5, Lem. 3] for nonstationary Gauss-Markov processes.

One of the key differences between the proofs of the asymptotically stationary ($a \in (0,1)$) and nonstationary ($a > 1$) cases is the scaling of the eigenvalues μ_i's of $\mathsf{A}^\top\mathsf{A}$, where A is defined in (35). For any $a \in (0,1)$, our previous work [5, Lem. 1, Eq. (71)] showed that for all $i = 1, ..., n$,

$$(1-a)^2 \le \mu_i \le (1+a)^2. \quad (44)$$

However, in the case of $a > 1$, we show that (44) holds only for $i = 2, ..., n$, while for $i = 1$, we have

$$2\log a - \frac{c_1}{n} \le -\frac{1}{n}\log\mu_1 \le 2\log a + \frac{c_2}{n}, \quad (45)$$

where $c_1 > 0$ and c_2 are constants. That is, the minimum eigenvalue of $\mathsf{A}^\top\mathsf{A}$ decreases to 0 in the order of $\mu_1 = \Theta(a^{-2n})$ as n tends to infinity. See [16] for the details.

IV. Conclusion

In this paper, we obtain a nonasymptotic bound (Theorem 1) on the estimation error of the maximum likelihood estimator of the parameter a of the nonstationary scalar Gauss-Markov process. An asymptotic bound (Theorem 2) follows immediately. Numerical simulations confirm the tightness of our bounds compared to previous works. As an application of the estimation error bound (Corollary 1), we find the dispersion for lossy compression of the nonstationary Gauss-Markov sources (Theorem 4). As a future work, we are interested in generalizing the estimation error bounds to system identification of vector dynamical systems and in finding the dispersion of the Wiener process ($a = 1$).

Appendix A

A. Proof of Theorem 1

We present the proof of (20). The proof of (21) is similar, which we omit here. For any $n \ge 2$, denote by $\mathcal{F}_n$ the σ-algebra generated by $Z_1, \ldots, Z_n$. For any $s > 0$, $\eta > 0$, and $n \ge 2$, let W_n be the following random variable

$$W_n \triangleq \exp\left\{ s\sum_{i=1}^{n-1}(U_i Z_{i+1} - \eta U_i^2)\right\}. \quad (46)$$

By the Chernoff bound, we have

$$\mathbb{P}\left[\hat{a}(\boldsymbol{U}) - a \ge \eta\right] \le \inf_{s>0}\ \mathbb{E}\left[W_n\right]. \quad (47)$$

To compute $\mathbb{E}[W_n]$, we first condition on $\mathcal{F}_{n-1}$. Since Z_n is the only term in W_n that does not belong to $\mathcal{F}_{n-1}$, we have

$$\begin{aligned} &\mathbb{E}\left[W_n\right] \\ =&\mathbb{E}\left\{W_{n-1}\cdot \mathbb{E}[\exp(s(U_{n-1}Z_n - \eta U_{n-1}^2))|\mathcal{F}_{n-1}]\right\} \quad (48)\\ =&\mathbb{E}\left[W_{n-1}\cdot\exp(\alpha_1 U_{n-1}^2)\right] \quad (49)\end{aligned}$$

where α_1 is the deterministic function of s and η defined in (14) and (49) follows from the moment generating function of Z_n. To obtain a recursion, we condition on $\mathcal{F}_{n-2}$. Since U_{n-1}^2 and $U_{n-2}Z_{n-1}$ are the only two terms in $W_{n-1}\cdot\exp(\alpha_1 U_{n-1}^2)$ that do not belong to $\mathcal{F}_{n-2}$, we use the relation $U_{n-1} = aU_{n-2} + Z_{n-1}$ and we complete squares in Z_{n-1} to obtain

$$\begin{aligned} &W_{n-1}\cdot\exp(\alpha_1 U_{n-1}^2) \\ =&W_{n-2}\cdot\exp\left\{\alpha_1\left(Z_{n-1} + (a + \frac{s}{2\alpha_1})U_{n-2}\right)^2 + \right.\\ &\left.(a^2\alpha_1 - s\eta)U_{n-2}^2 - \alpha_1\left(a + \frac{s}{2\alpha_1}\right)^2 U_{n-2}^2\right\}. \quad (50)\end{aligned}$$

Furthermore, using the formula for the moment generating function of the noncentral χ^2-distributed random variable

$$\left(Z_{n-1} + \left(a + \frac{s}{2\alpha_1}\right) U_{n-2}\right)^2 \tag{51}$$

with 1 degree of freedom, we have

$$\begin{aligned}&\mathbb{E}\left[W_{n-1} \cdot \exp(\alpha_1 U_{n-1}^2)\right] \\ =&\frac{1}{\sqrt{1-2\sigma^2\alpha_1}} \mathbb{E}\left[W_{n-2} \cdot \exp(\alpha_2 U_{n-2}^2)\right],\end{aligned} \tag{52}$$

where α_2 is given by (15). Repeating the above recursion, we obtain

$$\mathbb{E}\left[W_n\right] = \exp\left\{-\frac{1}{2}\sum_{\ell=1}^{n-1} \log(1 - 2\sigma^2\alpha_\ell)\right\}. \tag{53}$$

Finally, if $s \notin \mathcal{S}_n^+$ then $\mathbb{E}\left[W_n\right] = +\infty$. Therefore,

$$\inf_{s>0} \ \mathbb{E}\left[W_n\right] = \inf_{s \in \mathcal{S}_n^+} \ \mathbb{E}\left[W_n\right]. \tag{54}$$

B. Properties of $\{\alpha_\ell\}$ *and* $\{\beta_\ell\}$

We list the essential properties of the sequence α_ℓ, see [16] for the counterparts for β_ℓ and their proofs. We find the two fixed points $r_1 < r_2$ of the recursive relation (15) by solving the following quadratic equation in x:

$$2\sigma^2 x^2 + [a^2 + 2\sigma^2 s(a+\eta) - 1]x + \alpha_1 = 0. \tag{55}$$

1) For any $s > 0$ and $\eta > 0$, (55) always has two distinct solutions r_1 and r_2. Moreover, $r_1 < 0$.
2) For any $\frac{2\eta}{\sigma^2} \neq s > 0$ and $\eta > 0$, the sequence $\{\frac{\alpha_\ell - r_1}{\alpha_\ell - r_2}\}$ is a geometric sequence with common ratio q given by

$$q \triangleq \frac{[a^2 + 2\sigma^2 s(a+\eta)] + 2\sigma^2 r_1}{[a^2 + 2\sigma^2 s(a+\eta)] + 2\sigma^2 r_2}. \tag{56}$$

We show that q is always in $(0, 1)$. Hence, we have the following expression for α_ℓ:

$$\alpha_\ell = r_2 + \frac{r_2 - r_1}{\frac{\alpha_1 - r_1}{\alpha_1 - r_2} q^{\ell-1} - 1} \tag{57}$$

and its limit:

$$\lim_{\ell \to \infty} \alpha_\ell = r_1. \tag{58}$$

3) For any $s \in (0, 2\eta/\sigma^2)$, $\alpha_\ell < 0$ and decreases to r_1 monotonically. For $s = 2\eta/\sigma^2$, we have $\alpha_\ell = 0$, $\forall \ell$. For $s > 2\eta/\sigma^2$, (58) still holds but the convergence is not monotone.
4) For any $\eta > 0$, r_1 is a decreasing function in s.

Finally, in the analysis of $\{\beta_\ell\}$, we denote t_1 and t_2 the two fixed points of the recursive relation (17), which are the two solutions of the following quadratic equation:

$$2\sigma^2 x^2 + [a^2 + 2\sigma^2 s(-a+\eta) - 1]x + \beta_1 = 0. \tag{59}$$

References

[1] P. Tian and V. Kostina, "The dispersion of the Gauss-Markov source," in *Proceedings of 2018 IEEE International Symposium on Information Theory*, Vail, CO, USA, pp. 1490–1494.

[2] 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.

[3] A. Ingber and Y. Kochman, "The dispersion of lossy source coding," in *Proceedings of 2011 IEEE Data Compression Conference*, Snowbird, UT, USA, Mar. 2011, pp. 53–62.

[4] V. Kostina and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3309–3338, Jun. 2012.

[5] P. Tian and V. Kostina, "The dispersion of the Gauss-Markov source," *arXiv preprint arXiv:1804.09418*, Apr. 2018.

[6] J. S. White, "The limiting distribution of the serial correlation coefficient in the explosive case," *The Annals of Mathematical Statistics*, pp. 1188–1197, Dec. 1958.

[7] J. Rissanen and P. Caines, "The strong consistency of maximum likelihood estimators for ARMA processes," *The Annals of Statistics*, pp. 297–315, Mar. 1979.

[8] B. Bercu, F. Gamboa, and A. Rouault, "Large deviations for quadratic forms of stationary Gaussian processes," *Stochastic Processes and their Applications*, vol. 71, no. 1, pp. 75–90, Oct. 1997.

[9] J. Worms, "Large and moderate deviations upper bounds for the Gaussian autoregressive process," *Statistics & probability letters*, vol. 51, no. 3, pp. 235–243, Feb. 2001.

[10] A. Rantzer, "Concentration bounds for single parameter adaptive control," in *Proceedings of 2018 IEEE Annual American Control Conference*, Milwaukee, WI, USA, Jun. 2018, pp. 1862–1866.

[11] A. Kolmogorov, "On the Shannon theory of information transmission in the case of continuous signals," *IRE Transactions on Information Theory*, vol. 2, no. 4, pp. 102–108, Dec. 1956.

[12] T. Berger, "Rate-distortion theory," *Encyclopedia of Telecommunications*, 1971.

[13] R. M. Gray, "Information rates of autoregressive processes," *IEEE Transactions on Information Theory*, vol. 16, no. 4, pp. 412–421, Jul. 1970.

[14] T. Hashimoto and S. Arimoto, "On the rate-distortion function for the nonstationary Gaussian autoregressive process," *IEEE Transactions on Information Theory*, vol. 26, no. 4, pp. 478–480, Jul. 1980.

[15] R. M. Gray and T. Hashimoto, "A note on rate-distortion functions for nonstationary Gaussian autoregressive processes," *IEEE Transactions on Information Theory*, vol. 54, no. 3, pp. 1319–1322, Feb. 2008.

[16] P. Tian and V. Kostina, "From parameter estimation to dispersion of nonstationary Gauss-Markov processes," *Preprint*, 2019.

[17] H. B. Mann and A. Wald, "On the statistical treatment of linear stochastic difference equations," *Econometrica, Journal of the Econometric Society*, pp. 173–220, Jul. 1943.

[18] H. Rubin, "Consistency of maximum likelihood estimates in the explosive case," *Statistical Inference in Dynamic Economic Models*, pp. 356–364, 1950.

[19] B. Bercu and A. Touati, "Exponential inequalities for self-normalized martingales with applications," *The Annals of Applied Probability*, vol. 18, no. 5, pp. 1848–1869, 2008.

[20] M. Simchowitz, H. Mania, S. Tu, M. I. Jordan, and B. Recht, "Learning without mixing: Towards a sharp analysis of linear system identification," *preprint arXiv:1802.08334*, May 2018.

[21] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Springer-Verlag, Berlin, 2010.