Finite-Time Nonanticipative Rate Distortion Function for Time-Varying Scalar-Valued Gauss-Markov Sources

Photios A. Stavrou, Themistoklis Charalambous, and Charalambos D. Charalambous

***Abstract*—We derive the finite-time horizon nonanticipative rate distortion function (NRDF) of time-varying scalar Gauss-Markov sources under an *average* mean squared-error (MSE) distortion fidelity. Further, we show that a conditionally Gaussian reproduction process realizes the optimal reproduction distribution, and this is determined from the solution of a dynamic reverse-waterfilling optimization problem. We provide an iterative algorithm that approximates the solution of the dynamic reverse-waterfilling problem. From the above results, we also obtain, as a special case, the NRDF under a *per-letter or pointwise* MSE distortion fidelity, and we draw connections to the classical RDF of Gaussian processes. Our results are corroborated with illustrative examples.**

***Index Terms*—Nonanticipative rate distortion function, finite-time, scalar Gauss-Markov processes, mean squared-error distortion, dynamic reverse-waterfilling.**

I. Introduction

THE finite-time horizon nonanticipative rate distortion function (NRDF) and its per unit time asymptotic limit, are information theoretic variants of the finite-time classical RDF and its per unit time asymptotic limit (i.e., the classical RDF [1]), with the additional property that the reproduction distribution depends causally on past data.

The finite-time NRDF was first introduced by Gorbunov and Pinsker in [2], [3], to address issues related to real-time processing of data in communication applications. The authors of [2] also characterized the finite-time NRDF for time-varying and stationary scalar and vector-valued Gauss-Markov processes with pointwise or per-letter mean squared-error (MSE) distortion fidelity. For scalar-valued Gaussian processes, they gave the expression of finite-time NRDF in terms of a reverse-waterfilling at each time instant. Tatikonda *et al.* in [4] applied the asymptotic NRDF of time-invariant scalar and vector-valued Gauss-Markov processes with pointwise MSE distortion fidelity to identify fundamental limitations of controling a linear quadratic control system over a limited rate communication channel. Derpich and Østergaard in [5] characterized many variants of the asymptotic NRDF of stationary stable scalar-valued Gaussian autoregressive models with pointwise MSE distortion fidelity. For Gaussian autoregressive sources with unit memory, the authors in [5], showed that, in the asymptotic regime, the NRDF with average or pointwise MSE distortion fidelity coincide. Tanaka *et al.* in [6] showed that the characterization of the finite-time and stationary NRDF of vector-valued Gauss-Markov processes with pointwise MSE distortion fidelity is semidefinite representable (under certain assumptions), and, thus, it can be computed numerically. Stavrou *et al.* in [7] considered the time-varying vector-valued Gauss-Markov source under an average MSE distortion fidelity, applied a realization scheme to compute the finite-time NRDF, and proposed a method to solve the resulting reverse-waterfilling algorithm. Although, the realization scheme given in [7] is correct, the reverse-waterfilling of the finite-time horizon NRDF of both vector-valued and scalar-valued Gauss-Markov sources under an average MSE distortion fidelity (see [7, Theorem 2]) is suboptimal; for scalar-valued Gauss-Markov sources, it is optimal in the asymptotic regime.

The renewed interest in NRDF, stems from the fact that it provides a tighter lower bound on the optimal performance theoretically attainable (OPTA) by causal [8] and zero-delay codes [5], [9] compared to the OPTA by noncausal codes (i.e., the classical RDF). For this reason, as demonstrated in [4], the NRDF is often applied to identify fundamental performance limitations of real-time communication schemes in feedback control systems.

In this paper, we revisit the finite-time horizon NRDF for time-varying scalar-valued Gauss-Markov sources subject to an average MSE distortion constraint. The contributions of this work are as follows: *(i)* we derive the realization of the optimal reproduction distribution and we show this depends on the current value of the source symbol with unit memory on the values of past reproduction symbols, *(ii)* we obtain the expression of finite-time horizon NRDF using sequential optimization, and we characterize it by a dynamic reverse-waterfilling solution, that is solved numerically via an iterative algorithm, and *(iii)* we use the proposed framework to recover the analytical expression of the Gaussian NRDF under pointwise MSE distortion constraint. Further, we draw connections to the classical RDF of a Gaussian random process. We illustrate examples to verify our framework and to show that the asymptotic limit of the finite-time NRDF coincides with that of pointwise MSE fidelity.

Our results illustrate fundamental differences between finite-time Gaussian NRDF under average MSE and pointwise

P. A. Stavrou is with the Department of Electronic Systems, Aalborg University, Denmark, *email:* fos@es.aau.dk.

T. Charalambous is with the Department of Electrical Engineering and Automation, School of Electrical Engineering, Aalto University, Finland, *email:* themistoklis.charalambous@aalto.fi.

C. D. Charalambous is with the Department of Electrical and Computer Engineering, University of Cyprus, Cyprus, *email:* chadcha@ucy.ac.cy

MSE distortion fidelity, and the analogous classical RDF. In classical finite-time RDF the reverse-waterfilling algorithm reveals waterlevels which are constant over time, whereas in finite-time NRDF, the waterlevels are determined sequentially in time, from a dynamic optimization problem.

II. Problems Statement and Preliminaries

We use the following notation. $\mathbb{R}$ is the set of real numbers, $\mathbb{N}_0$ the set of nonnegative integers, and $\mathbb{N}_0^n \triangleq \{0,\ldots,n\}, n \in \mathbb{N}_0$. The distribution of a random variable (RV) X with realizations $X=x\in\mathbb{X}$ is denoted by $\mathbf{P}_X(dx)\equiv\mathbf{P}(dx)$. The set of all distributions on the measurable set $\mathbb{X}$ is denoted by $\mathcal{M}(\mathbb{X})$. Given another RV Y with realizations $Y=y\in\mathbb{Y}$, we denote the conditional distribution of Y given $X=x$ by $\mathbf{P}_{Y|X}(dy|x)\equiv\mathbf{P}(dy|x)$. We denote by $\mathbf{E}_x$ the expectation operator for a fixed realization $X=x\in\mathbb{X}$. We denote sequences of RVs by $X_t^j \triangleq (X_t, X_{t+1},\ldots,X_j), (t,j)\in\mathbb{N}_0\times\mathbb{N}_0, j\geq t$, and their values by $x_t^j\in\mathbb{X}_t^j\triangleq \times_{k=t}^j \mathbb{X}_k$, with $\mathbb{X}_k=\mathbb{X}$, for simplicity. If $t=0$, we use the notation $X_0^j=X^j$. Given two distributions $\mathbf{P}(dx)$ and $\mathbf{P}(dy|x)$, we write $\mathbf{P}(dy)=\int_{\mathbb{X}}\mathbf{P}(dy|x)\otimes\mathbf{P}(dx)$ to denote the marginal distribution is a compound probability distribution, and similarly for the rest.

A. Problems Formulation

We consider a source that randomly generates symbols $X^n=x^n\in\mathbb{X}^n$, that we wish to reproduce or reconstruct by $Y^n=y^n\in\mathbb{Y}^n$, subject to a distortion or a fidelity of reconstruction, defined by $d_{0,n}(x^n,y^n)\triangleq\sum_{t=0}^n|x_t-y_t|^2$, where $S=s\in\mathbb{S}$ is the initial data of the reproduction.

Data Source. The sequences $X^n=x^n\equiv(x_0,\ldots,x_n), n\in\mathbb{N}_0$, are generated from the distributions

$$\mathbf{P}_{X_t|X^{t-1},Y^{t-1},S}=\mathbf{P}_{X_t|X_{t-1}}\equiv P_t(dx_t|x_{t-1}),\ \forall t\in\mathbb{N}_0^n. \quad (1)$$

At time $t=0$, we assume $\mathbf{P}_{X_0|X^{-1},Y^{-1},S}=P_0(dx_0)$. Thus,

$$\mathbf{P}_{X^n}\equiv P_{0,n}(dx^n)\triangleq\otimes_{t=0}^n P_t(dx_t|x_{t-1}).$$

Reproduction or "test-channel". Suppose the reproduction sequence $Y^n=y^n\equiv(y_0,\ldots,y_n), n\in\mathbb{N}_0$ of $x^n\triangleq(x_0,\ldots,x_n)$ is generated, from the collection of conditional distributions, known as test-channels, defined by

$$\mathbf{P}_{Y_t|Y^{t-1},X^t,S}\triangleq Q_t(dy_t|y^{t-1},x^t,s),\quad t\in\mathbb{N}_0^n. \quad (2)$$

At $n=0$, we assume $\mathbf{P}_{Y_0|Y^{-1},X^0,S}=Q_0(dy_0|x_0,s)$, where $s\in\mathbb{S}$ is the initial state. If the initial state carries no information, then the RV S generates the trivial $\sigma-$algebra, $\sigma\{S\}=\{\Omega,\emptyset\}$.

By [10], we can formally define the distributions on $\mathbb{Y}^n$ parametrized by $(x^n,s)\in\mathbb{X}^n\times\mathbb{S}$, via

$$\mathbf{P}_{Y^n|X^n,S}\equiv\overrightarrow{Q}_{0,n}(y^n|x^n,s)\triangleq\otimes_{t=0}^n Q_t(dy_t|y^{t-1},x^t,s).$$

By (1) and (2), we can uniquely define the joint distribution of (X^n,Y^n) conditioned on $S=s$ by

$$\mathbf{P}_{X^n,Y^n|S}\triangleq P_{0,n}(dx^n)\otimes\overrightarrow{Q}_{0,n}(dy^n|x^n,s). \quad (3)$$

From the joint distribution (3), we can uniquely define the $\mathbb{Y}^n-$marginal distribution conditioned on $S=s$ by

$$\mathbf{P}_{Y^n|S}\equiv V_{0,n}(dy^n|s)\triangleq\int_{\mathbb{X}^n}P_{0,n}(dx^n)\otimes\overrightarrow{Q}_{0,n}(dy^n|x^n,s),$$

and the conditional distributions

$$\mathbf{P}_{Y_t|Y^{t-1},S}\equiv V_t(dy_t|y^{t-1},s),\quad\forall t\in\mathbb{N}_0^n.$$

Finite Time Horizon NRDF**.** Given the above construction of distributions, we define the information measure of NRDF from the definition of relative entropy $\mathbb{D}(\cdot||\cdot)$ between $\mathbf{P}_{X^n,Y^n|S}$ and the product distribution $\mathbf{P}_{Y^n|S}\times\mathbf{P}_{X^n}$, as follows:

$$\begin{aligned}
I(X^n\rightarrow Y^n|s)&\triangleq\mathbb{D}(\mathbf{P}_{X^n,Y^n|S}||\mathbf{P}_{Y^n|S}\times\mathbf{P}_{X^n})\in[0,\infty]\\
&\stackrel{(a)}{=}\int_{\mathbb{X}^n\times\mathbb{Y}^n}\log\left(\frac{\overrightarrow{Q}_{0,n}(\cdot|x^n,s)}{V_{0,n}(\cdot|s)}(y^n)\right)P_{0,n}(dx^n)\otimes\\
&\qquad\overrightarrow{Q}_{0,n}(y^n|x^n,s)\equiv\mathbb{I}(P_{0,n},\overrightarrow{Q}_{0,n}) \quad (4a)\\
&\stackrel{(b)}{=}\mathbf{E}_s\left\{\sum_{t=0}^n\log\left(\frac{Q_t(\cdot|Y^{t-1},X^t,S)}{V_t(\cdot|Y^{t-1},S)}(y_t)\right)\right\} \quad (4b)\\
&\stackrel{(c)}{=}\sum_{t=0}^n I(X^t;Y_t|Y^{t-1},s), \quad (4c)
\end{aligned}$$

where (a) is due to the Radon-Nikodym derivative theorem [11], (b) due to chain rule of relative entropy, (c) by definition. $\mathbb{I}(P_{0,n},\overrightarrow{Q}_{0,n})$ indicates the dependence of $I(X^n;Y^n|s)$ on $\{P_{0,n},\overrightarrow{Q}_{0,n}\}$ conditioned on $s\in\mathbb{S}$. Often, we use either (4a) or (4c). It should be noted that due to (1), (4c) is a special form of directed information from X^n to Y^n conditioned on $S=s$ (see, e.g., [10]).

Next, we state the problems of interest.

Problem 1: (Finite-time horizon NRDF of Gaussian sources with average MSE distortion fidelity) Let X_t be a time-varying scalar-valued Gauss-Markov process

$$X_{t+1}=a_tX_t+W_t,\quad X_0=x,\ \forall t\in\mathbb{N}_0^{n-1}, \quad (5)$$

where a_t is non-random, $X_0\sim\mathcal{N}(0;\sigma_{X_0}^2)$, i.e., Gaussian, and $W_t\sim\mathcal{N}(0;\sigma_{W_t}^2)$, is an independent Gaussian noise process independent of (X_0,S). The finite-time horizon NRDF of the Gaussian process (5) with average MSE distortion fidelity is defined as

$$R_{0,n}^{\mathrm{na,P1}}(D)\triangleq\frac{1}{n+1}\inf_{\mathcal{Q}_{0,n}(D)}I(X^n\rightarrow Y^n|s), \quad (6)$$

where $\mathcal{Q}_{0,n}(D)$ is the distortion fidelity set defined by

$$\begin{aligned}\mathcal{Q}_{0,n}(D)\triangleq\Big\{&Q_t(dy_t|y^{t-1},x^t,s),\ \forall t\in\mathbb{N}_0^n:\\
&\frac{1}{n+1}\mathbf{E}_s\left\{\sum_{t=0}^n|X_t-Y_t|^2\right\}\leq D\Big\}, \quad (7)\end{aligned}$$

for some $D\in[D_{0,n}^{\min},D_{0,n}^{\max}]\subseteq[0,\infty)$.

Problem 2: (Finite-time horizon NRDF of Gaussian sources with pointwise MSE distortion fidelity) This is defined similar to Problem 1, with $R_{0,n}^{\mathrm{na},P1}(D)$ replaced by

$$R_{0,n}^{\mathrm{na,P2}}(D)\triangleq\frac{1}{n+1}\inf_{\mathcal{Q}_{0,n}(D_0,\ldots,D_n)}I(X^n\rightarrow Y^n|s), \quad (8)$$

and the fidelity set (7) replaced by

$$\mathcal{Q}_{0,n}(D_0,\ldots,D_n) \triangleq \Big\{Q_t(dy_t|y^{t-1},x^t,s), \forall t\in\mathbb{N}_0^n: \mathbf{E}_s\left\{|X_t-Y_t|^2\right\}\leq D_t, \forall t\in\mathbb{N}_0^n\Big\}, \tag{9}$$

for some $D_t\in[D_t^{\min},D_t^{\max}]\subseteq[0,\infty),\ \forall t\in\mathbb{N}_0^n$.

B. The Finite-Time Horizon Classical Gaussian RDF

The finite-time horizon classical RDF of a stationary Gaussian process X^n with distortion fidelity set

$$\mathcal{Q}_{0,n}^{cl}(D)\triangleq\{\mathbf{P}_{Y^n|X^n,S}: \frac{1}{n+1}\mathbf{E}_s\left\{\sum_{t=0}^n|X_t-Y_t|^2\right\}\leq D\},$$

is (see [1, Chapter 4.5.2])

$$R_{0,n}(D)\triangleq\frac{1}{n+1}\inf_{\mathcal{Q}_{0,n}^{cl}(D)} I(X^n;Y^n|s) = \frac{1}{n+1}\sum_{t=0}^n\log\left(\frac{\lambda_t}{\delta_t}\right),$$

where $\{\lambda_t:\ t\in\mathbb{N}_0^n\}$ are the eigenvalues of the covariance matrix, $\Phi\in\mathbb{R}^{n\times n}$, and $\{\delta_t:\ t\in\mathbb{N}_0^n\}$ are the distortion levels allocated based on the reverse-waterfilling algorithm:

$$\delta_t=\begin{cases}\xi & \text{if} \quad \xi\leq\lambda_t\\ \lambda_t & \text{if} \quad \xi>\lambda_t\end{cases},\quad \sum_{t=0}^n\delta_t=D_{\text{tot}}\triangleq(n+1)D, \tag{10}$$

with $\xi=\frac{1}{2\theta}>0$, and $D\in[D_{0,n}^{\min},D_{0,n}^{\max}]\subseteq[0,\infty)$. The above expression $R_{0,n}(D)$ was first shown in [12], for nonstationary Gaussian processes. However, its asymptotic per unit time limit $R(D)=\lim_{n\longrightarrow\infty}R_{0,n}(D)$, for nonstationary unstable Gaussian processes, was only understood recently in [13].

C. Properties of Problems 1, 2

Now, we state certain properties of the finite-time NRDF of Problems 1, 2, which are special cases of the ones in [7], [10], and analogous to those of the classical RDFs (see, e.g., [14]).

Consider Problem 1. Then, the following statements hold.
(1) The set of distributions $\overrightarrow{Q}_{0,n}(\cdot|x^n,s)\in\mathcal{M}(\mathbb{Y}^n)$ is convex.
(2) The set $\mathcal{Q}_{0,n}(D)$ is convex in $\overrightarrow{Q}_{0,n}(\cdot|x^n,s)\in\mathcal{M}(\mathbb{Y}^n)$.
(3) The functional $\mathbb{I}(P_{0,n},\cdot)$ is convex in $\overrightarrow{Q}_{0,n}(\cdot|x^n,s)\in\mathcal{M}(\mathbb{Y}^n)$.
(4) Suppose $R_{0,n}^{\text{na,P1}}(D)<\infty$. Then, $R_{0,n}^{\text{na,P1}}(D)$ is a convex, non-increasing function of $D\in[0,\infty)$.
Statements (1)-(4) hold for Problem 2 as well.

The next structural result follows from [7, Theorem 1]. Consider either Problem 1 or Problem 2. Then, the following property holds.
(5) For a fixed $S=s$, the optimal reproduction conditional distributions of Problems 1 or 2 is Markov with respect to X^n, i.e., is of the form $Q_t^*(y_t|y^{t-1},x^t,s)=Q_t^*(dy_t|y^{t-1},x_t,s),\ \forall t\in\mathbb{N}_0^n$, and the corresponding joint process (X^n,Y^n) is jointly Gaussian.

III. Main Results

In this section we give the structure of the optimal realization for Problems 1 and 2, and the parametric solution to Problem 1 in terms of a dynamic reverse-waterfilling algorithm. Using this result, we obtain the closed form expression of Problem 2 as a special case, and we discuss connections with classical Gaussian RDF.

Lemma 1: (Realization of optimal reproduction distribution) Consider Problem 1. Then, the following statements hold.
(a) Any candidate of the optimal reproduction distribution $\{Q_t(dy_t|y^{t-1},x_t,s):\ t\in\mathbb{N}_0^n\}$ is realized by the recursion

$$Y_t=h_t\left(X_t-\widehat{X}_{t|t-1}\right)+\widehat{X}_{t|t-1}+V_t,\quad \forall t\in\mathbb{N}_0^n \tag{11}$$

where $\widehat{X}_{t|t-1}\triangleq\mathbf{E}_s\{X_t|Y^{t-1}\}$, and $\{V_t\sim N(0,\sigma_{V_t}^2):t=t\in\mathbb{N}_0^n\}$ is an independent Gaussian process independent of $\{W_t: t\in\mathbb{N}_0^{n-1}\}$ and (X_0,S), and $\{h_t: t\in\mathbb{N}_0^n\}$ are time-varying deterministic functions.
Moreover, define the process $\{\nu_t:\ t\in\mathbb{N}_0^n\}$

$$\nu_t\triangleq Y_t-\mathbf{E}_s\left\{Y_t|Y^{t-1}\right\} = Y_t-\widehat{X}_{t|t-1}=h_t\left(X_t-\widehat{X}_{t|t-1}\right)+V_t. \tag{12}$$

Then, $\{\nu_t\ :\ t\in\mathbb{N}_0^n\}$ is the innovations process of $\{Y_t\ :\ t\in\mathbb{N}_0^n\}$, i.e., it is an orthogonal process with $\nu_t\sim N(0;\sigma_{\nu_t}^2)$, $\sigma_{\nu_t}^2=h_t^2\Sigma_{t|t-1}+\sigma_{V_t}^2$ and $\Sigma_{t|t-1}\triangleq\mathbf{E}_s\left\{(X_t-\widehat{X}_{t|t-1})^2|Y^{t-1}\right\}$.
(b) Let $\widehat{X}_{t|t}\triangleq\mathbf{E}_s\{X_t|Y^t\}$ and $\Sigma_{t|t}\triangleq\mathbf{E}_s\left\{(X_t-\widehat{X}_{t|t})^2|Y^t\right\}$. Then, $\{\widehat{X}_{t|t-1},\Sigma_{t|t-1}\ :\ t\in\mathbb{N}_0^n\}$ satisfy the following scalar-valued filtering recursions:

$$\widehat{X}_{t|t-1}=a_{t-1}\widehat{X}_{t-1|t-1}, \tag{13a}$$
$$\Sigma_{t|t-1}=a_{t-1}^2\Sigma_{t-1|t-1}+\sigma_{W_{t-1}}^2, \tag{13b}$$
$$\widehat{X}_{t|t}=\widehat{X}_{t|t-1}+K_t\nu_t, \tag{13c}$$
$$K_t=\Sigma_{t|t-1}h_t(\sigma_{\nu_t}^2)^{-1}, \tag{13d}$$
$$\Sigma_{t|t}=\Sigma_{t|t-1}-\Sigma_{t|t-1}^2h_t^2(\sigma_{\nu_t}^2)^{-1}. \tag{13e}$$

(c) $R_{0,n}^{\text{na,P1}}(D)$ is given by

$$R_{0,n}^{\text{na,P1}}(D)=\inf_{\substack{h_t\in\mathbb{R},\\ \sigma_{V_t}^2\geq0,\ t\in\mathbb{N}_0^n}}\frac{1}{2}\frac{1}{n+1}\sum_{t=0}^n\left[\log\left(\frac{\Sigma_{t|t-1}}{\Sigma_{t|t}}\right)\right]^+, \tag{14a}$$

$$\text{s.t.}\quad \frac{1}{n+1}\sum_{t=0}^n\left((1-h_t)^2\Sigma_{t|t-1}+\sigma_{V_t}^2\right)\leq D \tag{14b}$$
$$\Sigma_{t|t-1}=a_{t-1}^2\Sigma_{t-1|t-1}+\sigma_{W_{t-1}}^2 \tag{14c}$$
$$\Sigma_{t|t}=\Sigma_{t|t-1}-\Sigma_{t|t-1}^2h_t^2(h_t^2\Sigma_{t|t-1}+\sigma_{V_t}^2)^{-1} \tag{14d}$$

for $D\in[D_{0,n}^{\min},D_{0,n}^{\max}]\subseteq[0,\infty)$, where $[x]^+\triangleq\max\{0,x\}$.

Proof: (a) By § II-C property (5), for a fixed $S=s$, we have the following orthogonal realization $Y_t=h_tX_t+g_t(Y^{t-1})+V_t$, $g_t(Y^{t-1})=\Gamma_tY^{t-1},\forall t\in\mathbb{N}_0^n$, where $\{V_t\ :\ t\in\mathbb{N}_0^n\}$ is an independent Gaussian process. Moreover, since (2) holds, and $\{W_t\ :\ t\in\mathbb{N}_0^n\}$ is an independent process, then $\{V_t\ :\ t\in\mathbb{N}_0^n\}$ is independent of $\{W_t\ :\ t\in\mathbb{N}_0^n\}$, X_0, and S. For such a realization,

$I(X^n \rightarrow Y^n|s) = \sum_{t=0}^n I(X_t;Y_t|Y^{t-1},s)$ does not depend on $g_t(\cdot)$, $\forall t \in \mathbb{N}_0^n$. Since $\mathbf{E}_s\left\{\sum_{t=0}^n |X_t - Y_t|^2\right\} = \mathbf{E}_s\left\{\sum_{t=0}^n |(1-h_t)X_t - g_t(Y^{t-1})|^2\right\} + \sum_{t=0}^n \sigma_{V_t}^2$, then, by mean-square estimation theory, a smaller average distortion occurs when $g_t(Y^{t-1}) = (1-h_t)\widehat{X}_{t|t-1}$, $\forall t \in \mathbb{N}_0^n$. This completes the derivation of (11). The second part of (a) follows from well-known properties of innovations process. (b) This follows from the Kalman filter equations of conditionally Gaussian models. (c) This follows directly from (a), (b), and (4c), i.e., $I(X^n \rightarrow Y^n|s) = \sum_{t=0}^n I(X_t;Y_t|Y^{t-1},s) \equiv \sum_{t=0}^n \mathbf{E}_s\left\{\log\left(\frac{\mathbf{P}_{X_t|Y^t,S}}{\mathbf{P}_{X_t|Y^{t-1},S}}\right)\right\}$. This completes the proof. ■

Statements (a) and (b) in Lemma 1 are precisely the same for Problem 2. However, in statement (c) we need to replace average MSE distortion with pointwise MSE distortion.

The next theorem shows that the optimal reproduction distribution is of the form $Q_t(dy_t|y_{t-1},x_t,s), \forall t \in \mathbb{N}_0^n$.

Theorem 1: (Characterization of Problem 1) Define $\lambda_t \triangleq \Sigma_{t|t-1}$ and $\delta_t \triangleq \Sigma_{t|t}$. Then the following hold.
(a) The optimal reproduction of (14) is realized by

$$Y_t = h_t X_t + (1-h_t)a_{t-1}Y_{t-1} + V_t, \quad Y_{-1} = s, \tag{15}$$

$$h_t \triangleq 1 - \frac{\delta_t}{\lambda_t}, \quad \sigma_{V_t}^2 \triangleq h_t\delta_t \geq 0, \ \delta_t \geq 0, \ \lambda_t \geq 0, \tag{16}$$

$$\lambda_t = \alpha_{t-1}^2\delta_{t-1} + \sigma_{W_{t-1}}^2, \ \lambda_0 = \sigma_{X_0}^2, \ \forall t \in \mathbb{N}_1^n. \tag{17}$$

Moreover, for the above realization the following hold.

$$\widehat{X}_{t|t} = Y_t, \ \ \widehat{X}_{t|t-1} = a_{t-1}Y_{t-1}, \tag{18}$$

$$\mathbf{P}_{Y_t|Y^{t-1},X^t,S} = \mathbf{P}_{Y_t|Y_{t-1},X_t,S}, \ \ \forall t \in \mathbb{N}_0^n, \tag{19}$$

$$I(X^n \rightarrow Y^n|s) = \sum_{t=0}^n I(X_t;Y_t|Y_{t-1},s). \tag{20}$$

(b) The characterization of $R_{0,n}^{na,\text{P1}}(D)$ is

$$R_{0,n}^{\text{na,P1}}(D) = \min_{\delta_t \geq 0, \ t \in \mathbb{N}_0^n} \frac{1}{2}\frac{1}{n+1}\sum_{t=0}^n \left[\log\left(\frac{\lambda_t}{\delta_t}\right)\right]^+, \tag{21a}$$

$$\text{s.t.} \ \ \lambda_t \ \text{satisfies} \ \ (17), \tag{21b}$$

$$\sum_{t=0}^n \delta_t \leq D_{\text{tot}}, \quad D_{\text{tot}} \triangleq (n+1)D \tag{21c}$$

where $D \in [D_{0,n}^{\min}, D_{0,n}^{\max}] \subseteq [0,\infty)$.

Proof: (a) By Lemma 1, (14c) and (14d), we get

$$\delta_t = \lambda_t - \frac{h_t^2\lambda_t^2}{h_t^2\lambda_t + \sigma_{V_t}^2}, \ \ \forall t \in \mathbb{N}_0^n. \tag{22}$$

Note that in (22), if at time instant t, $\lambda_t = 0$, then, $\delta_t = 0$ and the rate at that time instant will be zero (trivial). In order to decouple the dependence of δ_t from the dynamics of λ_t in (14d), and hence freely optimize over δ_t, we choose $(h_t, \sigma_{V_t}^2)$ according to (16). It can be verified that the choice (16), when substituted into (13d), yields the Kalman gain $K_t = 1$, which then implies (18). Hence, by substituting in (11) we obtain (15). From (15) we deduce (19), which implies (20). An alternative derivation is the following. The inequality $\mathbf{E}_s\left(\sum_{t=0}^n |X_t - Y_t|^2\right) \geq \mathbf{E}_s\left\{\sum_{t=0}^n |X_t - \mathbf{E}_s\{X_t|Y^t\}|^2\right\}$, holds $\forall Y_t$, $t \in \mathbb{N}_0^n$, i.e., for all $(h_t, \sigma_{V_t}^2)$, $t \in \mathbb{N}_0^n$, and it is achieved if $\mathbf{E}_s\{X_t|Y^t\} = Y_t$. Similarly as above, the choice (16) achieves the lower bound, i.e., a smaller distortion is achieved for a given rate. (b) This follows from (a). This completes the proof. ■

It should be noted that the choice of h_t in (16) requires that $\delta_t \leq \lambda_t$, $\forall t$, but such a constraint is always satisfied (cf. eq. (22)), and hence can be omitted from (21). For the expression $R_{0,n}^{\text{na,P1}}(D)$ given in Theorem 1, δ_t, is chosen to control $\lambda_t = \alpha_{t-1}^2\delta_{t-1} + \sigma_{W_{t-1}}^2, t \in \mathbb{N}_0^n$. This is fundamentally different from the classical RDF (see §II-B), in which the analogous expression of λ_t is independent of δ_{t-1}. This point is illustrated in the next theorem.

Theorem 2: (Parametric solution to Problem 1) The parametric solution of Problem 1 is the following.

$$R_{0,n}^{\text{na,P1}}(D) = \frac{1}{2}\frac{1}{n+1}\sum_{t=0}^n \log\left(\frac{\lambda_t}{\delta_t}\right), \tag{23}$$

where λ_t satisfies recursion (17) and δ_t is evaluated based on the following dynamic reverse-waterfilling algorithm:

$$\delta_t \triangleq \begin{cases} \xi_t & \text{if} \quad \xi_t \leq \lambda_t \\ \lambda_t & \text{if} \quad \xi_t > \lambda_t \end{cases}, \ \forall t, \tag{24}$$

with $\sum_{t=0}^n \delta_t = D_{\text{tot}} \triangleq (n+1)D$, and

$$\xi_t = \begin{cases} \frac{1}{2\beta_t^2}\left(\sqrt{1 + \frac{2\beta_t^2}{\theta}} - 1\right), \ \forall t \in \mathbb{N}_0^{n-1} \\ \frac{1}{2\theta}, \ t = n \end{cases}, \tag{25}$$

where $\theta > 0$, $\beta_t^2 \triangleq \frac{\alpha_t^2}{\sigma_{W_t}^2}$, and $D \in [D_{0,n}^{\min}, D_{0,n}^{\max}] \subseteq (0,\infty)$.

Proof: We solve the optimization problem of (21) by employing Karush-Kuhn-Tucker (KKT) conditions [15, Chapter 5.5.3]. First, notice that the objective function (21a) can be reformulated (without the normalization with respect to the total number of time steps) as follows:

$$\frac{1}{2}\sum_{t=0}^n \log\left(\frac{\lambda_t}{\delta_t}\right) = \frac{1}{2}\sum_{t=0}^n (\log\lambda_t - \log\delta_t)$$
$$= \frac{1}{2}\left\{\underbrace{\log(\lambda_0)}_{\text{initial step}} + \sum_{t=0}^{n-1}\log\left(\alpha_t^2 + \frac{\sigma_{W_t}^2}{\delta_t}\right) - \underbrace{\log\delta_n}_{\text{final step}}\right\}. \tag{26}$$

Next, we introduce the augmented Lagrange functional (assuming the initial step is fixed) as follows:

$$J(\{\delta_t\}_{t=0}^n, \theta, \{\mu_t\}_{t=0}^n) = \frac{1}{2}\sum_{t=0}^{n-1}\log\left(\alpha_t^2 + \frac{\sigma_t^2}{\delta_t}\right) - \log(\delta_n)$$
$$+ \theta\left(\sum_{t=0}^n \delta_t - D_{\text{tot}}\right) - \sum_{t=0}^n \mu_t\delta_t, \ \ \theta \geq 0, \ \ \mu_t \geq 0, \tag{27}$$

where δ_t is the primal variable and the pair (θ, μ_t) are dual variables. From (27) we observe that the KKT conditions are necessary and sufficient conditions of optimality, since the problem is convex and differentiable. Using (27), we can differentiate for $\{\delta_t:\ t\in\mathbb{N}_0^n\}$. This gives

$$\frac{\partial J(\{\delta_t\}_{t=0}^n,\theta,\{\mu_t\}_{t=0}^n)}{\partial\delta_t}=\frac{1}{2}\left(-\frac{\sigma_{W_t}^2}{2\delta_t(\alpha_t^2\delta_t+\sigma_{W_t}^2)}\right)+\theta$$
$$-\mu_t=0\Longrightarrow\theta-\mu_t=\frac{\sigma_{W_t}^2}{2\delta_t(\alpha_t^2\delta_t+\sigma_{W_t}^2)},\ \ \forall t\in\mathbb{N}_0^{n-1},\tag{28}$$

$$\frac{\partial J(\{\delta_t\}_{t=0}^n,\theta,\{\mu_t\}_{t=0}^n)}{\partial\delta_n}=\frac{1}{2}\left(-\frac{1}{\delta_n}\right)+\theta-\mu_n=0$$
$$\Longrightarrow\theta-\mu_n=\frac{1}{2\delta_n},\ \text{(final time step } n).\tag{29}$$

From the complementary slackness conditions, we have

$$\theta\left(\sum_{t=0}^n\delta_t-D_{\text{tot}}\right)=0,\ \ \theta\geq 0,\tag{30}$$
$$\mu_t\delta_t=0,\ \ \mu_t\geq 0.\tag{31}$$

Since the problem is convex, constraint (21c) holds with equality and, therefore, $\theta>0$. Since $\delta_t>0\ \forall t\in\mathbb{N}_0^n$ (by definition), then by (31), $\mu_t=0$. By solving (28), (29) with respect to $\{\delta_t:\ t\in\mathbb{N}_0^n\}$, we obtain the solution of a quadratic equation as follows:

$$\delta_t=-\frac{1}{2\beta_t^2}\pm\frac{1}{2\beta_t^2}\sqrt{1+\frac{2\beta_t^2}{\theta}},\ \ \ \beta_t^2\triangleq\frac{\alpha_t^2}{\sigma_{W_t}^2},\tag{32}$$

for all $t\in\mathbb{N}_0^{n-1}$, with $\delta_n=\frac{1}{2\theta}$ for the final time step n. However, the negative solution in (32) is excluded, since $\delta_t>0,\ \forall t$, leaving only the positive solution. Hence, we obtain (25). The previous analysis gives an θ^*, $\mu_t^*=0$ and $\xi_t^*,\forall t$. The problem is solved once we ensure that the objective function in (21) is non-negative, that is, by checking that $\delta_t=\min\{\xi_t,\lambda_t\},\forall t$. This completes the proof. ■

In the next remark, we draw connections with finite-time classical RDF of Gaussian processes.

Remark 1: (Comparison to the classical Gaussian RDF) Comparing the NRDF of Theorem 2 to the classical RDF given in §II-B, there is a fundamental difference; ξ_t given by (25) is different at each time t, whereas in (10), $\xi=\frac{1}{2\theta}$ is a constant, and these coincide only at $t=n$.

In the next remark, we give the region in which D is well-defined, i.e., $D\in[D_{0,n}^{\min},D_{0,n}^{\max}]$.

Remark 2: ($D_{0,n}^{\min}$ and $D_{0,n}^{\max}$) From (25) it can be seen that as s grows large, $\xi_t\to 0$. Hence, for $\theta\to\infty$, $D_{0,n}^{\min}\to 0^+$. As $D_{0,n}^{\min}\to 0^+$, then $\sigma_{V_t}^2\to 0^+$, $h_t\to 1^+$, $\forall\ t\in\mathbb{N}_0^n$, and hence, it can be verified from (15) that $Y_t\to X_t$. Additionally, as $\theta\to 0^+$, then $\xi_t\to\infty$ and hence $\delta_t=\lambda_t,\ \ \forall t\in\mathbb{N}_0^n$. Therefore,

$$D_{0,n}^{\max}=\frac{1}{n+1}\sum_{t=0}^n\lambda_t=\frac{1}{n+1}\sum_{t=0}^n D_t^{\max},\tag{33}$$

where $\lambda_t=\alpha_{t-1}^2\lambda_{t-1}+\sigma_{W_{t-1}}^2$, and $\lambda_0=\sigma_{X_0}^2$. In order to find the minimum θ for which $D\leq D_{0,n}^{\max}$, we compute the maximum θ for each t for which there is no rate (i.e., $\delta_t=\lambda_t$), denoted by $\theta_t^{\max}$, in (28), (29). The minimum θ is the minimum among all $\theta_t^{\max}$, i.e., $\theta^{\min}=\min_t\{\theta_t^{\max}\}$.

A. Connections to Problem 2

Theorem 1 can be solved in closed form because by the KKT conditions of the solution occurs on the boundary, i.e., $\delta_t=D_t,\ \forall t\in\mathbb{N}_0^n$. Next, we give the solution to Problem 2, that was first derived in [2].

Corollary 1: (Optimal solution of Problem 2) Consider (14). Then, by choosing $h_t,\sigma_{V_t}^2$ as in (16) and assuming pointwise MSE distortion fidelity, then the optimization problem of (21) is simplified as follows:

$$R_{0,n}^{\text{na,P2}}(D)=\frac{1}{2(n+1)}\sum_{t=0}^n\left[\log\left(\frac{\lambda_t}{D_t}\right)\right]^+,\tag{34}$$

where $\lambda_t=a_{t-1}^2D_{t-1}+\sigma_{W_{t-1}}^2\geq D_t,\ \forall t$ with $\lambda_0=\sigma_{X_0}^2$.

Proof: Special case of Theorem 2. ■

Remark 3: ($D_t^{\min}$ and $D_t^{\max}$) By employing the KKT conditions for the pointwise MSE distortion fidelity, we obtain similar equations to (25), but there is a different Lagrange multiplier θ_t for each time t. Hence, similar to Remark 2, as θ_t grows large, $\xi_t\to 0^+$. Hence, for $\theta_t\to\infty$, the minimum distortion, satifies $D_t^{\min}\to 0^+$. When $\theta_t\to 0^+$, then, $\xi_t\to\infty$ and hence $D_t^{\max}=\lambda_t$. Note that when $D_t\leq D_t^{\max}$, $\theta_t\geq\theta_t^{\max}$.

IV. Algorithm

We propose an iterative method (see Algorithm 1) that solves numerically the dynamic reverse-waterfilling (24), (25).

Algorithm 1 Dynamic Reverse-Waterfilling algorithm

Initialize: number of time-steps n; distortion level D ($D\leq D_{0,n}^{\max}$); error tolerance ϵ; initial variance $\lambda_0=\sigma_{X_0}^2$ of the initial state X_0, values a_t and $\sigma_{W_t}^2$ of (5).
Set $\theta=1/2D$; flag $=0$.
while flag $=0$ **do**
  Compute $\delta_t\ \forall\ t$ as follows:
  for $t=0:n$ **do**
    Compute ξ_t according to (25).
    Compute δ_t according to (24).
    if $t<n$ **then**
      Compute λ_{t+1} according to (17).
    end if
  end for
  if $|\frac{1}{n+1}\sum_{t=0}^n\delta_t-D|\leq\epsilon$ **then**
    flag $\leftarrow 1$
  else
    Re-adjust θ as follows:

$$\theta\leftarrow\max\left\{\theta^{\min},\theta-\gamma(D-\frac{1}{n+1}\sum_{t=0}^n\delta_t)\right\},\tag{35}$$

    where $\gamma\in(0,1]$ is a proportionality gain; its choice affects the rate of convergence.
  end if
end while

V. Numerical Examples

Example 1: (Non-asymptotic regime)
For this example, we choose a time-horizon $n = 100$ for which we pick random values of $\{(a_t,\ \sigma^2_{W_t}) :\ t = 0, \ldots, 100\}$ in the range $(0, 1)$. We choose the initial value of the variance of (5) to be $\sigma^2_{X_0} = 1$ (hence, $\lambda_0 = \sigma_{X_0} = 1$). For these initial conditions, we evaluate $D^{\max}_{0,100} = 0.5468$ and $\theta^{\min} = 0.0020$ (see Remark 2). We choose the distortion level $D = 0.2$. We run Algorithm 1 for error tolerance $\epsilon = 10^{-3}$ and an initial $\theta = \theta_0$ to start our iterations (a good starting point is $\theta_0 = \frac{1}{2D}$). Then, we proceed as follows:

1) At $t = 0$, using (25), we evaluate ξ_0. Then, we use (24) to compute δ_0, i.e., $\delta_0 = \min\{\xi_0, \lambda_0\}$. Next, from (17), $\lambda_1 = \alpha_0^2\delta_0 + \sigma_0^2$.
2) At $t = 1$, using (25), we evaluate ξ_1 and subsequently, δ_1 and λ_2 are computed.
3) Similarly, the procedure is repeated until $t = 100$.
4) At the end, for the given value of θ, we check if $|\frac{1}{n+1}\sum_{t=0}^{n}\delta_t - D| \leq \epsilon$. If it does, we stop the iterations and the last evaluated value of θ is the solution of the ξ_t's which are the desired water-levels.
5) If not, we update θ using (35); in this example $\gamma = 0.1$. We repeat the previous procedure (steps 1) – 4)) with the new value of θ for all t.

The final reverse-waterfilling is found after 762 iterations and it is shown in Figure 1. By (23) we obtain:

$$R^{\text{na,P1}}_{0,100}(D) = \frac{1}{2}\frac{1}{101}\sum_{t=0}^{100}\log\left(\frac{\lambda_t}{\delta_t}\right) = 0.3749 \text{ bits}.$$

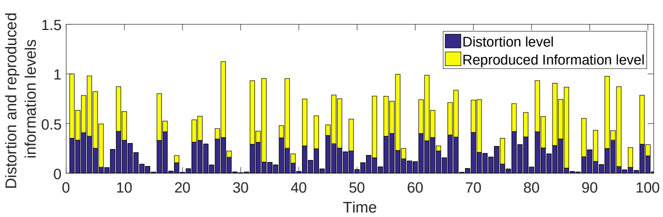

Fig. 1: Dynamic reverse-waterfilling in time-domain for $n = 100$ time units.

Example 2: (Asymptotic regime of finite-time NRDF)
Consider the time-invariant version of (5), i.e., $a_t \equiv a$, $\sigma^2_{W_t} \equiv \sigma^2_W$, $X_0 \sim N(0; \sigma^2_{X_0})$, with distortion level $D = 1.4$ and $(a, \sigma^2_W) = (1.5, 1)$. We run Algorithm 1 following the procedure described in Example 1, for $n = 100000$ with error tolerance $\epsilon = 10^{-3}$ and an initial $\theta = \frac{1}{2D}$. By (23) we obtain:

$$R^{\text{na,P1}}_{0,100000}(D) = \frac{1}{2}\frac{1}{100001}\sum_{t=0}^{100000}\log\left(\frac{\lambda_t}{\delta_t}\right) = 0.7838 \text{ bits}.$$

Note that for n large enough, then, ξ_t in (25) converges to $\xi = \frac{1}{2\beta^2}\left(\sqrt{1 + \frac{2\beta^2}{\theta}} - 1\right)$, and λ_t to $\lambda = \alpha^2\delta + \sigma^2_W$, hence $R^{\text{na,P1}}_{0,n}(D)$ converges to $R^{\text{na,P1}}(D) = \frac{1}{2}\log\left(\frac{\lambda}{\delta}\right)$, where $\delta = D$. Thus, $R^{\text{na,P1}}(D)$ is equivalent to the asymptotic NRDF with pointwise MSE found in [2, Example 1] and [4, Eq. (14)]. This means that for time-invariant scalar-valued Gauss-Markov sources, the asymptotic NRDF with average and pointwise MSE coincide, generalizing the result derived in [5, Theorem 3] for stationary stable sources. Compared to [4, Eq. (14)], we can additionally use Remark 2 to find that $D^{\max}_{0,100000} = 4.15$ bits and $\theta^{\min} = 0.0117$.

VI. Conclusions and Future Directions

We derived the realization of the optimal reproduction distribution, and the parametric solution of finite-time NRDF for time-varying scalar Gauss-Markov sources subject to an average MSE distortion which includes a dynamic reverse-waterfilling algorithm. The parametric solution indicates the solution at the final step is different from the other steps. We solved the reverse-waterfilling algorithm via an iterative algorithm which allocates the NRDF at each time step optimally. Our iterative algorithm is verified via numerical simulations. From the framework developed, we recover the optimal solution of finite-time NRDF subject to a pointwise MSE distortion. In the asymptotic case, the NRDF with average and pointwise MSE distortion coincide.

Generalizations to controlled processes and to time-varying multivariate Gauss-Markov sources will be treated in future works.

References

[1] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ: Prentice-Hall, 1971.
[2] A. K. Gorbunov and M. S. Pinsker, “Prognostic epsilon entropy of a Gaussian message and a Gaussian source,” *Problems Inf. Transmiss.*, vol. 10, no. 2, pp. 93–109, 1974.
[3] ——, “Nonanticipatory and prognostic epsilon entropies and message generation rates,” *Problems Inf. Transmiss.*, vol. 9, no. 3, pp. 184–191, 1973.
[4] S. Tatikonda, A. Sahai, and S. Mitter, “Stochastic linear control over a communication channel,” *IEEE Trans. Autom. Control*, vol. 49, pp. 1549–1561, 2004.
[5] M. S. Derpich and J. Østergaard, “Improved upper bounds to the causal quadratic rate-distortion function for Gaussian stationary sources,” *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 3131–3152, May 2012.
[6] T. Tanaka, K. K. K. Kim, P. A. Parrilo, and S. K. Mitter, “Semidefinite programming approach to Gaussian sequential rate-distortion trade-offs,” *IEEE Trans. Autom. Control*, vol. 62, no. 4, pp. 1896–1910, April 2017.
[7] P. A. Stavrou, T. Charalambous, and C. D. Charalambous, “Filtering with fidelity for time-varying Gauss-Markov processes,” in *Proc. IEEE Conf. Decision Control*, December 2016, pp. 5465–5470.
[8] D. Neuhoff and R. Gilbert, “Causal source codes,” *IEEE Trans. Inf. Theory*, vol. 28, no. 5, pp. 701–713, Sep. 1982.
[9] P. A. Stavrou, J. Østergaard, C. D. Charalambous, and M. S. Derpich, “An upper bound to zero-delay rate distortion via Kalman filtering for vector Gaussian sources,” in *Proc. IEEE Inf. Theory Workshop*, 2017.
[10] C. D. Charalambous and P. A. Stavrou, “Directed information on abstract spaces: Properties and variational equalities,” *IEEE Trans. Inf. Theory*, vol. 62, no. 11, pp. 6019–6052, Nov 2016.
[11] P. Dupuis and R. S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*. John Wiley & Sons, Inc., New York, 1997.
[12] 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, December 1956.
[13] R. M. Gray and T. Hashimoto, “Rate-distortion functions for non-stationary Gaussian autoregressive processes,” in *Data Compression Conference*, March 2008, pp. 53–62.
[14] S. Ihara, *Information theory - for Continuous Systems*. World Scientific, 1993.
[15] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.